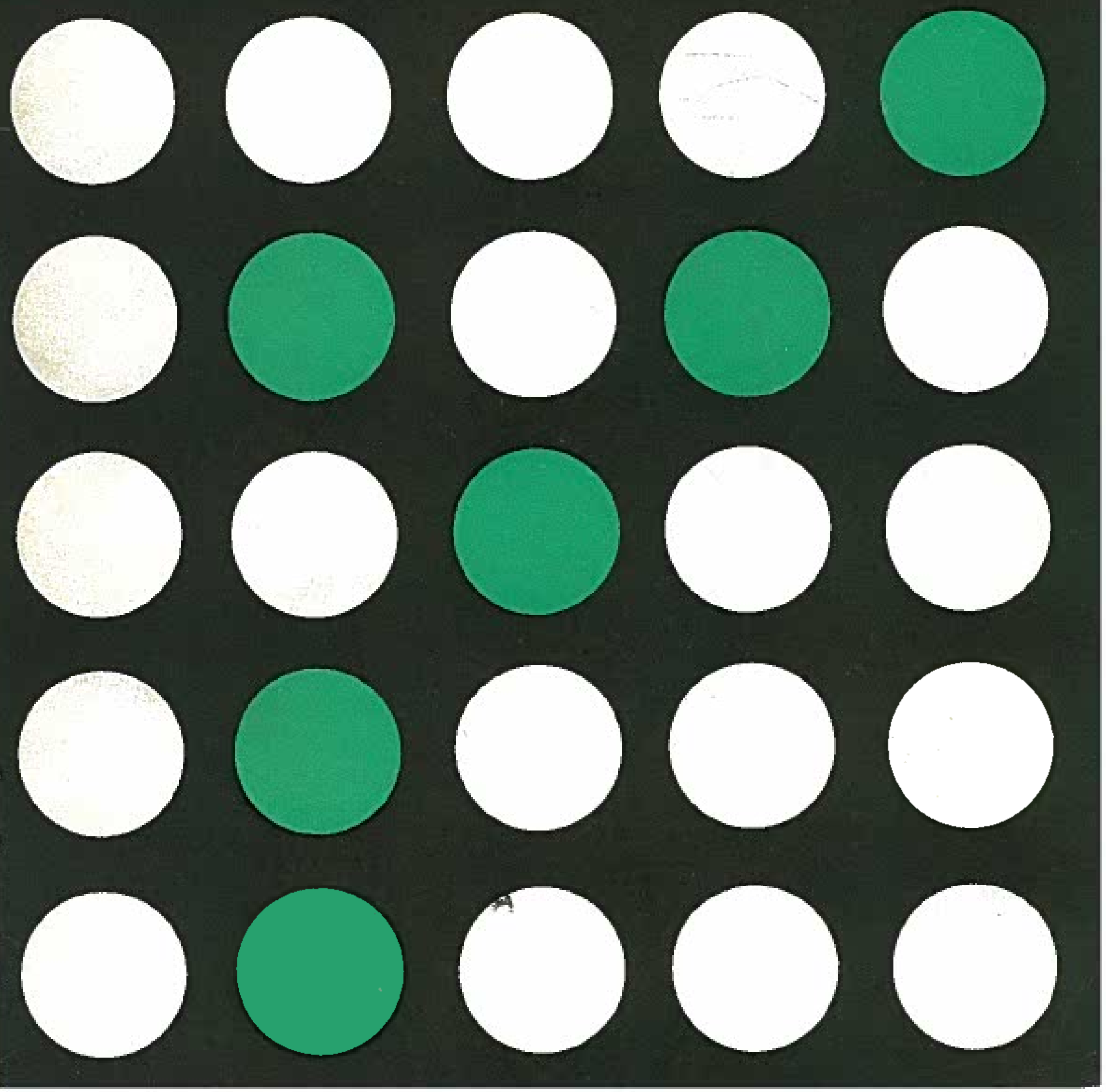
New Zealand	Government Printing Office: Government Bookshops at State Advances Buildings, Rutland Street, P.O. Box 5344, Auckland; 20 Molesworth Street, Private Bag, Wellington; 112 Gloucester Street, P.O. Box 1721, Christchurch; Stock Exchange Building, corner Water and Bond Streets, P.O. Box 1104, Dunedin.
Nigeria	University Bookshop Nigeria Ltd., University College, Ibadan.
Norway	Johan Grundt Tanum Forlag, Karl Johansgt. 43, Oslo.
Pakistan, West	Mirza Book Agency, 65 The Mall, Lahore 3.
Panama	Agencia Internacional de Publicaciones J. Menéndez, Apartado 2052, Panama.
Paraguay	Agencia de Librerías de Salvador Nizza, Calle Pte. Franco No. 39-43, Asunción.
Peru	Librería Internacional del Perú, S.A., Casilla 1417, Lima.
Philippines	The Modern Book Company, 518-520 Rizal Avenue, Manila.
Poland	Ars Polona, Krakowskie Przedmiescie 7, Warsaw.
Portugal	Livraria Bertrand, S.A.R.L., Rua Garrett 73-75, Lisbon.
South Africa	Van Schaik's Book Store Ltd., P.O. Box 724, Pretoria.
Spain	José Bosch Librero, Ronda Universidad 11, Barcelona; Librería Mundi-Prensa, Castelló 37, Madrid; Librería General, S. Miguel 4, Saragossa.
Sweden	C.E. Fritze, Fredsgatan 2, Stockholm 16; Gumperts A.B., Göteborg; Universitetsbokhandel, Sveavägen 166, Stockholm Va.
Switzerland	Librairie Payot, S.A., Lausanne and Geneva; Hans Raunhardt, Kirchgasse 17, Zurich 1.
Syria	Librairie Internationale, B.P. 2456, Damascus.
Taiwan	The World Book Company Ltd., 99 Chungking South Road, Section 1, Taipeh.
Thailand	Requests for FAO publications should be addressed to: FAO Regional Office for Asia and the Far East, Maliwan Mansion, Bangkok.
Turkey	Librairie Hachette, 469 Istaklal Caddesi, Beyoglu, Istanbul.
United Kingdom and Crown Colonies	H.M. Stationery Office, 49 High Holborn, London W.C.1; P.O. Box 569, London S.E.I. *Branches at*: 13a Castle Street, Edinburgh 2; 35 Smallbrook, Ringway, Birmingham 5; 50 Fairfax Street, Bristol 1; 39 King Street, Manchester 2; 109 St. Mary Street, Cardiff; 80 Chichester Street, Belfast.
United States of America	Columbia University Press, International Documents Service, 2960 Broadway, New York 27, New York.
Uruguay	Héctor d'Elía - Editorial Losada Uruguaya S.A., Colonia 1060, Montevideo.
Venezuela	Suma, S.A., Calle Real de Sabana Grande, Caracas; Librería Politécnica, Apartado del Este 4845, Caracas.
Yugoslavia	Drzavno Preduzece, Jugoslovenska Knjiga, Terazije 27/11, Belgrade; Cankarjeva Zalozba, P.O. Box 201 - IV, Ljubljana.
Other countries	Requests from countries where sales agents have not yet been appointed may be sent to: Distribution and Sales Section, Food and Agriculture Organization of the United Nations, Via delle Terme di Caracalla, Rome, Italy.

FAO publications are priced in U.S. dollars and pounds sterling. Payment to FAO sales agents may be made in local currencies.

Price: $6.00 or 30s.

PM33003/9.66/E/1/3800

FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS



QUALITY OF STATISTICAL DATA

by

S. S. Zarkovich
Chief, Methodology Branch
Statistics Division, FAO

FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS
Rome 1966

Printed in Italy

PREFACE

The body of this book consists of the material presented by the author at various seminars and training centers organized by FAO. The aim of these lectures was to spread awareness of the quality problem of statistical data and to promote interest in quality checks as a source of guidance on the adequate uses of data and on the ways and means of improving the methods used.

The composition of the book reflects these aims. First of all, an effort was made to explain in some detail what happens if errors are introduced into the data collected and what problems arise as a result. Afterward examples and experiences were presented in order to illustrate the significance of these problems. Similarly, in the presentation of techniques that might be involved in checking the quality of data, emphasis was laid on the explanation of the logic of the procedures rather than on the analysis of various techniques that were actually used under specific circumstances. At the end of the book a bibliographical list has been added. This list may offer some useful guidance to those who want to continue the study of material covered in the book.

In writing the book an effort was made to give a text that would be as simple as possible. In fact, an elementary course of statistical theory and a basic course on the theory of sample surveys are sufficient for an understanding of the text.

The reader may be surprised to see in this book definitions of certain basic concepts of survey techniques, such as biased and unbiased estimates, mean square error, etc. In spite of the fact that these concepts are known to all statisticians, it was thought necessary to include them in a text that aims at dealing systematically with all the topics that are encountered in a study of the quality of data.

A draft of this book has appeared in mimeographed form under the title *Sampling methods and censuses*, Vol. II: *Quality of statistical data*, FAO, 1963. An earlier and very summarized version of it was presented at the annual meeting of the Yugoslav Statistical Society under the title *Quality problems of statistical data* (in Serbo-Croatian), Ljubljana, 1956.

P.V. SUKHATME
Director, Statistics Division
FAO, Rome

CONTENTS

1. SOME BASIC CONCEPTS

1.1 Definition of errors

Before a survey can be made there are many factors that have to be determined. Such factors are concepts and definitions, methods of collecting data, the units to be used in expressing the response, the tabulation program, the survey program, the wording of questions, etc. We refer to all these factors under the general term of the *adopted system of work*. Accordingly, the adopted system of work shows what data are to be collected in a survey, in what way they are to be collected, etc.

The adopted system of work is shaped according to the aims of the survey. Since it represents a fixed system of concepts, definitions, procedures, and operations that constitute the survey, the specification of the adopted system of work makes it possible to judge whether the action taken is in agreement with the action prescribed. Needless to say, this possibility is sometimes only theoretical.

On the basis of the concept of the adopted system of work the concept of the *true value* can be defined. The true value is simply the result that should be obtained in a particular survey operation if the adopted system of work is carried out correctly. The true value is the *ideal result* of a particular survey operation; it is obtained if the work is done in *absolute conformity* with the adopted system of work.

There are several types of true value. The first is the *individual true value* of a characteristic for a given unit of population. The individual true value follows from the application of the adopted system of work in obtaining the value of a characteristic for a given unit. If in a census of population the age of the head of the household is required in years completed at the last birthday, the true value of this particular item is the number of years the head has in fact completed irrespective of whether he is aware of this value or not, and independently of what he has stated in the census. The true value of the total area of a holding as expressed in hectares would be the sum of the true values of the area of individual fields rounded off to the nearest integer. It is therefore evident that, after the adopted system of work is fixed, the true value becomes a defined quantity.

In some cases it will not be easy to visualize the meaning of true values. By way of example, it is sufficient to remind the reader of "intention" surveys, such as planting intentions. However, the difficulty arising here as well as the practical difficulty of ascertaining true values should not prevent the use of this concept, since, in fact, the real essence of errors in statistics can hardly be described without it.

In addition to individual true values we also speak of true values of totals, averages, proportions, coefficients of correlation, and other statistical measures. The meaning of these concepts is obvious.

In order to define the true value of the population total we use the symbol x_i to designate the true value of a characteristic for the i-th unit of the population. It is assumed that the total number of units of this population is equal to N. The true value of the population total for this characteristic is then defined as

$$X = \sum_i^N x_i \tag{1.1}$$

Definitions of the true values of other statistical measures are obvious.

It is clear that individual true values are not always achieved in survey practice for all the units. The results achieved factually will be called *survey values*. The survey value of the i-th unit of the population for the same characteristics as before, viz. x_i, will be designated by z_i. By analogy with the definitions of true values we distinguish *individual survey values* and survey values of various statistical measures. It is clear that the survey value of the population total must be defined as

$$Z = \sum_i^N z_i \tag{1.2}$$

The survey values of other statistical measures are defined by using the variable z and the well-known formulas from statistical theory.

By means of the true value and the survey value we now define the *error* as the difference between the survey value and the corresponding true value. Thus, the *individual error* of the i-th unit is defined[1] as

$$d_i = z_i - x_i \tag{1.3}$$

Individual errors may be positive and negative. If the survey value is equal to the corresponding true value, i.e., when $z_i = x_i$ or $d_i = 0$, we say that z_i is *accurate*. On the other hand, if $d_i \neq 0$, z_i will be called *inaccurate*.

The following are some additional concepts. From (1.3) we have

$$z_i = x_i + d_i \tag{1.4}$$

and

$$\sum_i^N z_i = \sum_i^N x_i + \sum_i^N d_i$$

or

$$Z = X + D \tag{1.5}$$

The quantity D is called the *bias*. Clearly, if $D = 0$, the survey value of the population total for a given characteristic is equal to the corresponding true value. In this case Z is said to be accurate or *unbiased*. Vice versa, if $D \neq 0$, Z is said to be *biased*.

Using (1.3) or (1.4) it is easy to define biases in other statistical measures.

From a practical point of view great importance is attached to the *frequency distribution of individual errors*. If positive and negative errors are distributed at random around zero, the estimates of totals and averages will be unbiased. In many cases, however, there is some pattern in errors in the sense that either positive or negative errors predominate. In such a case we speak of *systematic errors*. Totals and averages based on data subject to systematic errors will normally be biased. The bias is thus the *net effect* of all the errors.

From (1.5) it will also be clear that biases have their sign. In the case of $Z > X$ the bias is positive, whereas for $Z < X$ it is negative. With positive biases we say that the survey total Z overstates or overestimates the true value of the total. The opposite case is an understatement or underestimation.

The magnitude of the bias and its sign do not have equal importance in all the studies. Users of data will primarily be interested in the magnitude of the bias. In some analyses of errors, however, the sign of the bias may become essential.

---

[1] More details about the definition of errors and reasons for the use of the equation (1.3) will be found in Chapter 3.

The definition of errors presented above refers to simple situations. In subsequent parts of this book the concept of the error will be used in a broader sense. Thus, if there is a discrepancy between the instructions issued according to the adopted system of work and the procedure actually followed, we also speak of an error in spite of the fact that the effects of this error may not take the form of individual errors. For example, if the selection of the sample is not done as prescribed the case will be referred to as an error in the selection of the sample. Data for units included in the sample might be accurate. However, the estimates of totals and other statistical measures based on the sample may easily be biased. The effects of such errors are reflected as biases in final survey results. This is why the procedure might also be referred to as a biased selection procedure.

In Section 1.3 the errors will be classified according to different criteria and each of the established classes will be given a separate name. This will facilitate the understanding of what kind of errors are involved in a concrete case.

1.2 Where do errors appear?

In the preparation of a statistical survey the first stage is usually to establish the program of the survey and to determine the basic concepts and definitions that will be used. At this early stage of the preparations for the survey errors with serious consequences can easily occur: some characteristics may be forgotten which are found later to be important for a proper understanding of the subject; some definitions may be inadequate with the result that some units are left out of the survey; some concepts may be defined in a misleading way, etc.

In subsequent stages of the work there are, of course, many other possibilities for the occurrence of errors. For example, in the preparation of questionnaires the wording of some questions may be misleading; the layout of the material in questionnaires may be difficult to follow; the unwieldy size of the questionnaire may lead to answers being put in incorrect places.

Instructions for field work are another serious source of errors. If some duties are not explained fully, the personnel may easily act on their own initiative. If too lengthy explanations are given, they may be confused and again proceed according to their own ideas. If ambiguities and insufficiently precise statements are tolerated in the instructions, errors in one sense or another will be a likely result.

As regards the cartographic preparations, the delineation of enumeration districts or other area units may be defective in the sense that some units of the population are omitted while others are located on borders so that it is not completely clear where they belong. If descriptions of the enumeration districts' boundaries are used instead of maps the same difficulties may appear.

As regards the enumerators, some may not be properly selected while others may not be adequately trained. In either case errors of various kinds may result. Enumerators may omit some units for a number of subjective and objective reasons; they may enumerate twice or several times some units located on border lines; they may not repeat the call in cases where nobody was found at home at the first call, so that some data are missing; in some cases they may introduce their own ideas and opinions into the work they are doing; sometimes they elicit answers in a direction different from that determined by the adopted system of work; their behavior may create an atmosphere of tension which provokes refusals to answer; they may hurry unduly and neglect the quality of work if paid on a unit basis, etc.

Another source of difficulties lies in the respondent. He also has his own ideas which may be in many respects different from the intentions of the survey. Although precautions are normally taken, an elimination of the influence of the respondent's personal characteristics can never be completely achieved. He is sometimes ashamed, sometimes afraid, sometimes he wants to gain prestige, and therefore he changes his response.

The general conditions under which surveys are made may also favor errors. It is a well-known fact that a survey taken in an atmosphere of widespread belief that data to be collected will be used for nonstatistical purposes may easily contain serious errors. By providing inaccurate information some respondents believe they are protecting their personal integrity and interest.

When the processing of data has started, new possibilities for a large number of errors are opened up. In the editing, many answers in millions of questionnaires have to be checked for quality. In such a mass process the best trained people make errors. A proportion of defective pieces is produced on the best automatic machines. This is also true with other phases of processing such as coding and punching, which may again be compared with mass production in industry.

It is therefore evident that there is no work in statistics which is safe from error. Errors are ubiquitous; they follow a man's work in statistics as inevitably as a shadow follows a man walking in the sun. However, this does not necessarily give cause for pessimism. In general, one could say that errors appear more frequently than was expected some time ago, and therefore their effects are probably more considerable than was once believed. However, the evidence available also shows that a number of measures can be taken to keep errors under control. This is what justifies an optimistic attitude.

As in all other similar fields, efficient action against errors requires a thorough knowledge of the type of errors encountered in statistical work. One should know under what circumstances they occur, what their sources are, what consequences they have on various statistical measures, what tools and techniques could be used in improving data, etc. It is only when such comprehensive knowledge of various aspects of the general error problem is available that efficient measures for action can be prepared.

1.3 Classification of errors[2]

Before we attempt to make the classification of errors at which this book aims, a few remarks will be useful on errors which are committed prior to the commencement of the statistical preparations of the survey. A case in point is the inadequate specification of the survey program, such as an omission of some essential characteristics. The result may be that data collected do not offer the information needed. The same is true if an inadequate program of tabulation is adopted, etc. Although such errors can be very serious they do not basically belong to statistical theory. Setting up programs of statistical surveys as well as the definitions of the population to be surveyed are the joint responsibility of both users of statistical data and statisticians. This class of errors is therefore

---

[2] More details on various errors will be found in Deming, W.E.: On errors in surveys, *American Sociological Review*, Vol. 9, 1944, pp. 359-369; Deming, W.E.: *Some theory of sampling*, John Wiley, New York, 1950; Hansen, M.H., Hurwitz, W.N., and Madow, W.G.: *Sample survey methods and theory*, Vol. I, John Wiley, New York, 1953; Kish, L.: *Survey sampling*, John Wiley, New York, 1965; Mahalanobis, P.C.: Recent experiments in statistical sampling in the Indian Statistical Institute, *Journal of the Royal Statistical Society*, Vol. 109, 1946, pp. 326-378; Yates, F.: *Sampling methods for censuses and surveys*, Third Edition, Charles Griffin, London, 1960.

excluded from this book. For the same reason printing errors are also excluded from our consideration. Good printing is the printing office's responsibility. The errors to be dealt with here are those which are exclusively the statisticians' responsibility. In other words, we shall confine our discussion to errors that statisticians are expected to study and evaluate.

Errors in the broad sense of the word will be classified in three broad groups:

A. Errors resulting from inadequate preparations.
B. Errors committed in the stage of data collecting.
C. Processing errors.

This classification is deficient in a number of ways. For example, there is some overlapping between groups A and B. Thus, if the field worker obtains inaccurate information on a person's income, it may be because the concepts and definitions are inadequate, the instructions incomplete, the interviewing badly handled, etc. In other words, it is not easy to say whether the error belongs to groups A or B or to both. For practical reasons, however, it is useful to keep group A as a separate class. Such a classification reminds statisticians of the need for care because the ground for errors exists even in the first preparatory steps for the survey.

Group A is broken down into *biased procedures* and *biased tools.* If procedures are used which, in their repeated application, cause biases to appear in survey results, they are called biased procedures. An illustration is the selection of the sample on the basis of personal judgment. Three biased procedures are discussed here separately. They are measurement, selection and estimation procedures.

If a tool is used which leads, after repeated application, to biased results even if correctly applied according to the adopted system of work, it is called a biased tool. From among various biased tools that may appear in statistical work tables of random numbers, questionnaires, frames and instructions are discussed here. Measuring instruments may also be biased tools. The use of strings the length of which is different from the one assumed leads to biases in the estimated areas or distances. Our discussion, however, does not extend to measuring instruments as they are not specific statistical tools.

Group B is broken down into *listing errors, missing data*, and *response* or *observational errors.*

Listing errors occur in both the complete enumeration censuses and the sample surveys. In the former case the process of enumeration results in some lists of units constituting the population. If some units are omitted in listing or others are listed twice or several times, the errors committed will be called listing errors. Omitted units are also called omissions; the errors that consist of listing some units twice or several times are often called duplications. Another typical error belonging to this class is the listing of nonexistent units.

Listing errors are also called coverage errors or errors in the completeness of enumeration. The bias in survey estimates that appears as a result of listing errors is called the listing bias.

A special type of errors, called here misclassifications, is often encountered in censuses of population where, in addition to being listed, persons also have to be classified in groups, such as permanent resident, temporary absentee, and temporary resident. A person is misclassified if listed as temporary resident when in fact he or she is a permanent resident. If classes of permanent residents and temporary absentees comprise what is called the resident population or the number of inhabitants in a given area, the importance of misclassification becomes obvious.

Needless to say, misclassifications are based on wrong information about the units concerned.

Another type of listing error in censuses is the listing of units in a wrong enumeration district. In this case the unit is listed but not in the place where it should have been included according to the census instructions. If a person who should be enumerated in enumeration district A were enumerated in district B, we speak of an omission with regard to district A. The listing of this person in district B is called erroneous or wrong inclusion of this particular person in the population of district B.

In sample surveys, however, fresh lists of units are often prepared as a frame for sampling at a later stage. For example, in a sample survey of agricultural holdings a sample of communes or villages might be selected at the first stage of selection. Then the enumerators might be instructed to prepare lists of all the holdings in the selected first-stage units and choose from those lists 10 percent of holdings for interviewing. If in such a listing any errors of the type mentioned above should occur, i.e., omissions or duplications, we also speak of listing errors.

Missing data represent a special type of error encountered primarily in sample surveys. It occurs if the information for a unit included in the sample is not available for some reason. For example, in the food consumption surveys one might be dealing with households as sampling units. It might then happen that in some selected households nobody was found at home during the survey. Data for these units will be missing. Another common example occurs in yield surveys based on crop cutting. Some fields selected in the sample may already have been harvested when the enumerator comes for cutting data. The information for this particular field will again be missing.

Refusals represent a special case of missing data. Refusals appear in interview or mail surveys when the person involved is contacted but does not want to co-operate with the survey.

Missing data are usually dealt with in statistical literature under the name of nonresponse or incomplete samples. The first term corresponds adequately to interview surveys although it may not satisfactorily express the situation in observational surveys.

Broadly speaking, the *response* or *observational errors* refer to the difference between the individual true value and the corresponding survey value irrespective of the reason for the discrepancy. If a holder gives the information that the total area of his holding amounts to 8 ha while cadastral data, assumed accurate, show 7 ha, the answer given by the holder contains a response error. If an enumerator is counting the boats entering a harbor in a given period of time and presents a figure different from the actual number of boats entering, we speak of an observational error.

Clearly, response errors may be positive or negative. If they occur in a systematic way such as the results of weighing using biased scales, the quantities computed, such as totals, will be subject to response bias.

The errors committed in processing data could also be broken down into a number of classes. Examples are errors in *editing*, *coding*, *punching*, *tabulation*, etc. Obviously the number of these classes can differ according to the technique employed and the equipment available for processing data. The classes mentioned here as examples correspond to a standard equipment for mechanical tabulation.

The meaning of processing errors is evident from the above discussion.

1.4 Relative character of errors

Statistical errors have a relative meaning. This is probably their most important characteristic. An answer obtained in an interview survey or a figure recorded in an observational survey can be considered as an error only with regard to the adopted system of work. For example, if the total area of an agricultural holding were 5.4 ha according to the cadastral survey and 5.0 ha according to the census, the census information will be considered as inaccurate if the instructions ask for the response to be accurate to the first decimal place. If, however, the instructions specify that fractions are to be disregarded, there will be no error committed. Similarly, if a person states in a census of population that he was born on 20 April 1933 and was in fact born on 18 May 1931 according to his birth certificate, which is assumed here to be accurate, the answer given is inaccurate from the point of view both of the date of birth and of the number of years completed. For tabulation purposes, however, where age data are classified in five-year age groups, the answer will not be inaccurate if the person, on the basis of the information given, happens to be classified in the group where he belongs according to his birth certificate.

It is useful to keep in mind this feature of statistical errors. In order to reduce the number of errors it might sometimes be possible to adjust the adopted system of work accordingly.

1.5 Biased procedure of estimation

Some of the definitions given need to be broadened to suit particular circumstances met in sample surveys.

It will first be assumed that a sample of n units was drawn as a simple random sample with replacement from the population of N units. It is also assumed that true values of the characteristics under study are available for each of the n units. The true value of the population mean, $\bar{X}$, is then estimated on the basis of sample data by $\bar{x} = \frac{1}{n} \sum_{i}^{n} x_i$.

Because the sample of n units only was used, the estimate $\bar{x}$ will, in general, be different from $\bar{X}$. In fact, if all the possible samples of n units each were selected from the same population and $\bar{x}$ computed from each of them, the estimates $\bar{x}$ would vary around $\bar{X}$ in the form of a normal distribution. An important characteristic of this distribution is that the arithmetic mean of all the possible estimated means is equal to $\bar{X}$. Another way of saying the same thing is: the *expected value* of the estimate $\bar{x}$ is equal to $\bar{X}$ or $E\bar{x} = \bar{X}$. In this particular case we also say that $\bar{x}$ is an *unbiased* estimate of $\bar{X}$. In fact, whenever we have a sample estimate u of the quantity U, irrespective of what statistical measure is concerned, we say that u is an unbiased estimate of U if $Eu = U$.

The advantage of unbiased estimates is clear from what has already been said. The individual unbiased estimates of the mean may be more or less different from $\bar{X}$. However, we know that the average of $\bar{x}$ agrees with $\bar{X}$. The range of variations of $\bar{x}$ around $\bar{X}$ is measured by the quantity called *standard error*, which will be designated by $\sigma_{\bar{x}}$. The square of the standard error, $\sigma_{\bar{x}}^2$, is called the variance of the estimated mean and is defined as $\sigma_{\bar{x}}^2 = E(\bar{x} - \bar{X})^2$. In the case of simple random sampling with replacement we have the fundamental result $\sigma_{\bar{x}}^2 = \sigma_x^2/n$, indicating that the range of the possible variations of the estimate $\bar{x}$ depends, first, upon the variation of x values in the population and, second, upon the size of the sample. If we are sampling from the same population, the estimate $\bar{x}$ will, on an average, be nearer and nearer to $\bar{X}$ as the size of the sample is increased.

The magnitude of the average variations of the estimate $\bar{x}$ around $\bar{X}$ is also referred to by the term *precision*. The measure of the precision is given by the standard error. The smaller the standard error the more precise is the estimate $\bar{x}$ and vice versa.

The same terminology is also used in connection with other statistical measures. For example, if the population variance σ_x^2 is estimated in simple random sampling with replacement by

$$s_x^2 = \frac{1}{n-1} \sum_{i}^{n} (x_i - \bar{x})^2$$

the quantity s_x^2 is called an unbiased estimate of σ_x^2 because $Es_x^2 = \sigma_x^2$. The estimate s_x^2 is also subject to sampling variations and, accordingly, it has its standard error and precision.

All the sample estimates, however, do not fall into the category of unbiased estimates. The following is an illustration. In order to estimate the number of livestock the sample of m enumeration districts (ED) is assumed to be drawn from the population of M districts. In each of the selected m EDs all the agricultural holdings are interviewed and accurate data are obtained on the number of livestock belonging to each holding. In this situation an estimate of the number of livestock per holding, $\bar{\bar{x}}$, might be computed as follows:

$$\bar{\bar{x}} = \frac{1}{m}\sum_{i}^{m}\frac{1}{N_i}\sum_{j}^{N_i} x_{ij} = \frac{1}{m}\sum_{i}^{m}\bar{X}_i \tag{1.6}$$

where the symbols have the following meaning:

x_{ij} = the number of livestock in the j-th holding of the i-th ED,
N_i = number of holdings in the i-th ED, and
$\bar{X}_i$ = the average number of livestock per holding in the i-th ED.

It is evident that (1.6) represents a simple arithmetic mean of the enumeration districts' means.

The equation (1.6) is used with a view to estimating $\bar{X}$ which is defined as

$$\left.\begin{aligned} \bar{X} &= \frac{1}{N}\sum_{i}^{M}\sum_{j}^{N_i} x_{ij} \\ &= \frac{1}{M}\sum_{i}^{M}\frac{N_i}{\bar{N}}\bar{X}_i \end{aligned}\right\} \tag{1.7}$$

where N stands for the total number of holdings in the population while $\bar{N} = N/M$.

The estimate $\bar{\bar{x}}$ is called unbiased if $E\bar{\bar{x}}$ is equal to $\bar{X}$. In fact

$$E\bar{\bar{x}} = \frac{1}{M}\sum_{i}^{M}\bar{X}_i \tag{1.8}$$

It is clear that $E\bar{\bar{x}}$ as defined in (1.8) is not equal to $\bar{X}$ as defined in (1.7). The equation (1.8) is the simple arithmetic mean of district means per holding while (1.7) is the weighted arithmetic mean of district means. In accordance with the terminology used before, we say that $\bar{\bar{x}}$ represents a *biased estimate* of $\bar{X}$.

If we use the symbol $\bar{D}$ for the magnitude of the corresponding bias per holding we have

$$\bar{D} = E\bar{\bar{x}} - \bar{X} \tag{1.9}$$

In this particular case the magnitude of $\bar{D}$ is as follows:

$$\left.\begin{aligned} \bar{D} &= \frac{1}{M}\sum_{i}^{M}\bar{X}_i - \frac{1}{M}\sum_{i}^{M}\frac{N_i}{\bar{N}}\bar{X}_i \\ &= \frac{1}{M}\sum_{i}^{M}\bar{X}_i\left(1 - \frac{N_i}{\bar{N}}\right) \end{aligned}\right\} \tag{1.10}$$

The equation (1.10) shows that the estimate (1.6) is biased if EDs vary in both size and $\bar{X}_i$.

It can be seen that in sample surveys biased estimates may be obtained *even if data available for each individual unit are accurate.*

In connection with the estimate (1.6) it should be added that under the assumptions made earlier one could also use an unbiased estimate $\bar{\bar{x}}'$. It would be defined as

$$\bar{\bar{x}}' = \frac{1}{m\bar{N}}\sum_{i}^{m}\sum_{i}^{N_i} x_{ij} \tag{1.11}$$

with $E\bar{\bar{x}}' = \bar{X}$. If the selected EDs are subsampled and n_i units drawn in the subsample from the i-th selected ED, an alternative estimate of $\bar{X}$ would be

$$\bar{\bar{x}}'' = \frac{1}{m\bar{N}}\sum_{i}^{m}\frac{N_i}{n_i}\sum_{j}^{n_i} x_{ij} \tag{1.12}$$

which is also unbiased.

It is thus seen that biases due to a biased procedure of estimation can be removed in the above cases by using unbiased estimates.[3]

---

[3] With regard to probability statements concerning confidence limits in the case of a biased estimate, the theory of the normal distribution can only be used with particular care. Various problems connected therewith are discussed in Cochran, W.G.: *Sampling techniques*, John Wiley, New York, Second edition, 1963; Hansen, M.H., Hurwitz, W.N., and Madow, W.G.: *Sample survey methods and theory*, John Wiley, New York, 1953, and Kish, L.: *Survey sampling*, John Wiley, New York, 1965.

1.6 Biased versus unbiased estimates

In the preceding section it was assumed that data available for n units of a simple random sample were all accurate. To have a general case, inaccurate data for some units will be allowed in this section. Using our earlier notation and the equation (1.4), the survey value of the mean is estimated as $\bar z$ with

$$\left.\begin{aligned} E\bar z &= \bar Z \\ &= \bar X + \bar D \end{aligned}\right\} \qquad (1.13)$$

and

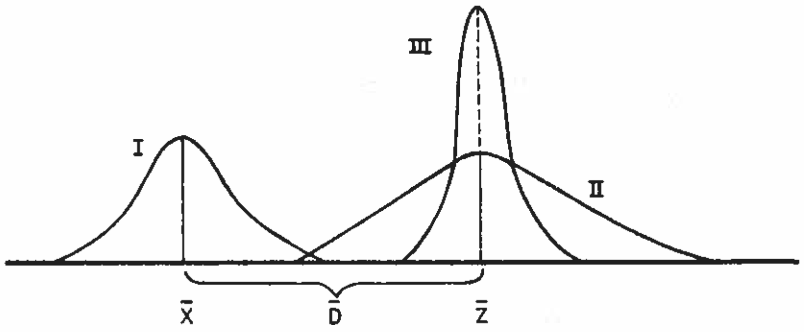
$$\bar D = E\bar z - \bar X$$

The variance of $\bar z$ is, of course,

$$\left.\begin{aligned} \sigma_{\bar z}^2 &= E(\bar z - E\bar z)^2 \\ &= \frac{1}{n}\,(\sigma_x^2 + \sigma_d^2 + 2\,\rho_{xd}\,\sigma_x\,\sigma_d) \end{aligned}\right\} \qquad (1.14)$$

The equation (1.14) is of fundamental importance in the theory of errors. It can be seen from the definition of $\sigma_{\bar z}^2$ that it measures the variation of the estimated mean $\bar z$ around its expected value $\bar Z$. If the bias $\bar D$ is large the situation depicted in Figure 1 might occur in practice.

FIGURE 1. - POSSIBLE RELATIONSHIP BETWEEN BIASED AND UNBIASED ESTIMATES OF THE MEAN IN THE CASE OF INACCURATE DATA



In this figure curve I represents the distribution of the individual true values in the population. The variance of this distribution is σ_x^2. Curve II indicates the corresponding distribution of survey values with the variance σ_z^2. Now, if samples of n units are selected from distribution II, the estimates $\bar z$ will vary around $\bar Z$ according to distribution III. The variance $\sigma_{\bar z}^2$ as defined in (1.14) is a measure of variations of the estimates $\bar z$ around $\bar Z$ in distribution III. It shows the precision of estimates $\bar z$ (with regard to $\bar Z$) but it does not give any indication of the basic aim of the survey, i.e., how far, on an average, our estimate $\bar z$ is from $\bar X$, which is the quantity in which we are interested. *From the magnitude of* $\sigma_{\bar z}^2$ *no conclusion whatsoever can be made about the location of* $\bar X$.

In fact, we would like to know to what extent $\bar z$ varies around $\bar X$. In other words, we need a measure of variations of $\bar z$ around $\bar X$. This new measure will be called *mean square error* and designated by ζ^2.

The definition of ζ^2 is the following:

$$\left.\begin{aligned} \zeta_{\bar z}^2 &= E(\bar z - \bar X)^2 \\ &= E\,[(\bar z - \bar Z) + (\bar Z - \bar X)]^2 \\ &= \sigma_{\bar z}^2 + \bar D^2 \end{aligned}\right\} \qquad (1.15)$$

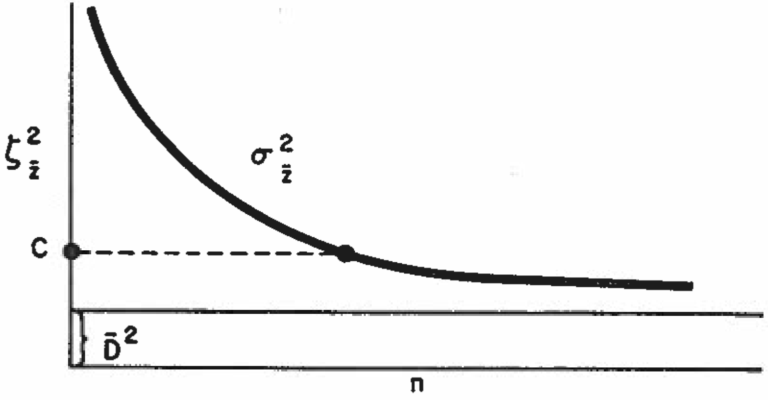
The mean square error as defined in (1.15) is the measure of the variation of the estimate $\bar z$ around $\bar X$. The square root from the mean square error is called *root mean square error*. As opposed to the precision, ζ^2 measures the *accuracy* of the estimate $\bar z$. Thus the term accuracy refers to the true value of the quantity being estimated while the precision refers to the expected value of the estimate. By analogy with the expression already used, we shall say that an estimate $\bar z$ is more accurate than another estimate $\bar z'$ if $\zeta_{\bar z}^2 < \zeta_{\bar z'}^2$, and vice versa. *Precision and accuracy have to be well distinguished.* An estimate might be very precise and at the same time have a very low accuracy. Figure 1 may help to visualize such possibilities.

On the basis of the explanations given the reader will be able to extend the concept of accuracy to other statistical measures.

The equation (1.15) is extremely important for sample survey practice. The magnitude of $\zeta_{\bar z}^2$ comprises two terms, one of which, $\sigma_{\bar z}^2$, depends upon the size of the sample while the other, $\bar D^2$, is independent of the size of the sample. This shows that in attempts to improve the accuracy of the estimate $\bar{z}$ a decision to increase the size of the sample might be very inadequate in some cases. The situation encountered is presented in Figure 2.

The y axis in this figure indicates the magnitude of $\zeta^2_{\bar{z}}$. On the x axis is the size of the sample, n. The curve represents the decreasing value of $\sigma^2_{\bar{z}}$ as the size of the sample increases. If the value of $\zeta^2_{\bar{z}}$, as obtained in a survey, were in point C and we wish to reduce it by, say, 25 percent by reducing $\sigma^2_{\bar{z}}$, the consequence would be such an increase in the size of the sample that it might not be practically feasible. An alternative step might be to use some costly and more refined methods that could give the result $\bar{D} = 0$ or nearly so. This would automatically lead to a reduction of $\zeta^2_{\bar{z}}$. This procedure might be found less costly than the increase of the size of the sample.

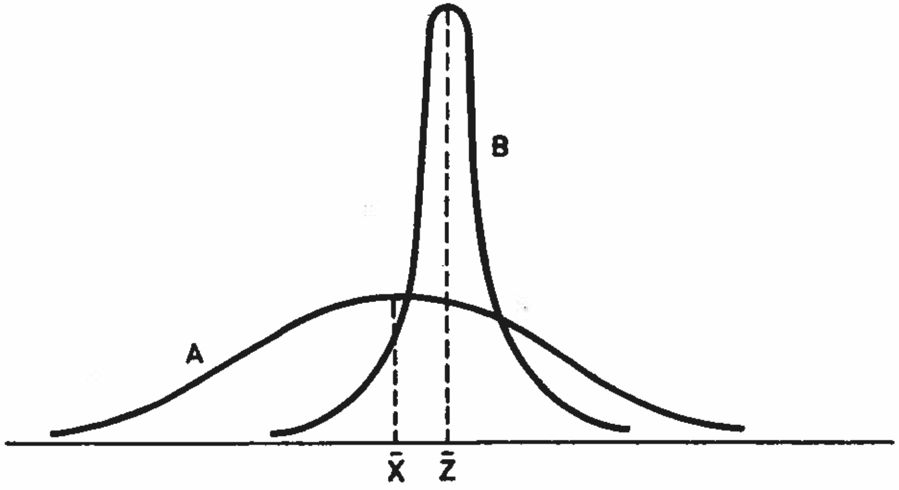
FIGURE 2. - EFFECTS OF INCREASING THE SIZE OF THE SAMPLE ON THE VALUE OF THE MEAN SQUARE ERROR



The relationships established here between the mean square error, the variance, and the bias in the case of inaccurate data are equally valid in the case of accurate data and a biased procedure of estimation.

At this stage it will be clear why in some cases a biased method of estimation is preferred. An illustration of a situation where this might be so is given in Figure 3. Curve A represents the distribution of the unbiased estimates of the arithmetic mean. The arithmetic mean of this distribution is $\bar{X}$. Curve B is the distribution of the biased estimates, $\bar{z}$, with the mean $\bar{Z}$. The magnitude of the bias is $\bar{D} = \bar{Z} - \bar{X}$. It will be seen from the figure that $\sigma^2_{\bar{x}} > \zeta^2_{\bar{z}}$. If this is so we prefer the biased estimate $\bar{z}$ because it gives more information about the location of the true value $\bar{X}$ than the corresponding unbiased estimate $\bar{x}$. Accordingly, if several methods are equal in all other respects, *the one which has the smallest mean square error should be chosen.*

FIGURE 3. - AN ILLUSTRATION OF THE SITUATION WHERE A BIASED ESTIMATE IS PREFERRED TO AN UNBIASED ONE



1.7 Illustrations

At this point some illustrations may be useful. The first is taken from a booklet by Panse.[4] Panse considered two estimates of the rate of yield. One was the unbiased weighted arithmetic mean as in the equation (1.12) and the other was the biased simple arithmetic mean (as in the equation (1.6)). The variances obtained for various crops with the two types of estimate are given in Table 1. It will be seen from this table that biased estimates lead to more accurate data for reasonably large biases. The percentage bias in unweighted mean that would give the same mean square error as weighted mean is equal to $\sqrt{a^2 - b^2}$.

The next example is taken from a quality check conducted in Greece in connection with the Census of Agriculture. In order to check the

---

[4] Panse, V.G.: *Estimation of crop yields*, FAO, Rome, 1954.

TABLE 1. - PERCENTAGE STANDARD ERRORS OF TWO ESTIMATES OF THE RATE OF YIELD[1]

Crop	Weighted arithmetic mean	Simple arithmetic mean
	(*a*)	(*b*)
Wheat (1947-48)	14.0	3.7
(1948-49)	10.0	2.5
Cotton (1944-45)	15.0	5.5
(1945-46)	14.0	6.9

[1] Panse, V.G.: Op. cit.

accuracy of area data as reported in the census a sample of fields was selected and their respective areas were measured. Also, in some areas the reported fields were identified on maps and those omitted were ascertained. The results obtained are presented in Table 2. As before, we use z to designate the value of the characteristics concerned as obtained in the census and x for the result obtained in the check. It can be seen that the estimated mean $\bar{z}$ is subject to a response bias. The estimate of this bias amounts to more than 12 percent in the case of areas and about 32 percent in the case of the number of parcels reported. Both biases are negative which means that reporting as a method of collecting data leads to underestimation of the characteristics involved.

TABLE 2. - SOME RESULTS OBTAINED IN THE CENSUS OF AGRICULTURE IN GREECE AND THE SUBSEQUENT QUALITY CHECK OF AREAS[1]

Measure / Characteristic	$\bar{z}$	$\bar{x}$	$\bar{d}$	σ_z^2	σ_x^2	σ_d^2	ϱ_{dx}
Areas (in stremmas)	37.8	42.2	— 4.4	920	680	507	—0.23
Number of parcels	6.6	8.7	— 2.1	32.5	33.0	18.9	—0.39

[1] National Statistical Service: Farmers reporting at the census, mimeographed report, Athens, 1962.

If the two methods, i.e., the reporting and the measurement, are used to estimate the average size of fields (in stremmas), the accuracy to be achieved for various sizes of the sample is presented in Table 3 (as based on data in Table 2). In computing the accuracy as measured by the root mean square error the equation (1.15) was used. In the case of measurement we assume $\bar{D} = 0$. It is also assumed that the quality of both reporting and measurements does not change with the size of the sample.

TABLE 3. - THE MAGNITUDE OF THE PERCENTAGE ROOT MEAN SQUARE ERROR FOR VARIOUS SIZES OF SAMPLE IN ESTIMATING THE AVERAGE AREA OF FIELDS BY MEANS OF REPORTING AND MEASUREMENT

Method	Size of sample			
	1	100	10 000	Complete census
Reporting	30.6	5.3	4.7	4.4
Measurement	26.1	2.6	0.26	0

Data in Table 3 represent a good illustration of what may happen if errors are present in data. The accuracy of reporting as measured by the root mean square error approaches the limit of 4.4 stremmas, which is the magnitude of the bias. This magnitude of the root mean square error remains even if a complete enumeration survey is taken. It will be seen that a very moderate size of sample with the use of measurements leads to more accurate estimates than a complete survey of all the fields based on reporting.

From data in Table 2 it is also easy to compute what size of sample is needed for an equal root mean square error to be obtained with the two techniques.

If cost elements are known, such as the average cost per field of both reporting and measurement, further possibilities for speculation are opened up. If budget C is available for the survey, c_1 and c_2 being the average cost per field of obtaining data on areas by means of reporting and measurement respectively, it might be desirable to know which of the two methods leads to more accurate data within the budget.

The answer to this problem is obvious: the method selected will be that which, under the assumption presented, leads to a smaller root mean square error. Using the index 1 for reporting and 2 for measurement, the number of units that can be included in the survey with the two methods will be approximately $n_1 = \frac{C}{c_1}$ and $n_2 = \frac{C}{c_2}\cdot$ If these values of n_1 and n_2 are included in the equation (1.15) the answer to the question is achieved.

In other cases the desired accuracy may be fixed and the problem will be to determine which method of collecting data requires a smaller budget. In this situation the value of ζ is fixed. The equation (1.15) is then solved for n as above and the resulting size of the sample is used to establish the survey budget. The method associated with a smaller budget is then chosen. Another illustration of a similar type is found in a paper by David.[5]

1.8 Further aspects of the presence of errors in data

From the equation (1.5) it follows that averages and totals are very sensitive to systematic errors in data. Therefore, if these quantities are estimated in a survey particular care is needed to avoid systematic errors.

The situation will be different with proportions and percentages if units selected in the sample continue to belong to the same classes of population irrespective of the presence of systematic response errors. For example, an estimate of the percentage of holdings growing wheat might be unbiased irrespective of systematic errors in the areas under wheat as reported by selected farmers. A similar result is found in the estimates of various rates if both the numerator and the denominator are subject to the same percentage bias. An example might be the estimation of the rate of yield or changes. This shows that biased totals and averages may sometimes be used to get unbiased estimates of ratios.

The effect of errors on variances is particularly important. It can be seen from the equation (1.14) that the magnitude of σ_z^2 may have very little in common with the true value of variances, σ_x^2. If there is no correlation between x and d we have the result $\sigma_z^2 > \sigma_x^2$ as soon as errors are present in data. The correlation between x and d further complicates the relationships between σ_z^2 and σ_x^2.

It is easy to ascertain that errors in data may substantially change the value of other statistical measures.

From the equation (1.14) it can also be seen how wrong is the attitude that there should not be much concern about errors if they are randomly distributed. Consider the estimate of the total Z, i.e., $Z' = \frac{N}{n}\Sigma z_i$. Under the assumption of random distribution of errors $EZ' = X$ and $\sigma_{Z'}^2 = \frac{N^2}{n}(\sigma_x^2 + \sigma_d^2)$. The quantity σ_d^2 is sometimes very large relative to σ_x^2.

---

[5] David, M.: The validity of income reported by a sample of families who received welfare assistance during 1959, *Journal of the American Statistical Association*, Vol. 57, 1962, pp. 680-685.

This means that σ_d^2 might sometimes make data useless. If in addition we consider what might happen in the case of strong correlation between x and d it becomes obvious that the presence of errors in data makes difficult the interpretation of the results of a survey.

Another source of difficulties is the *integration of errors*. This concept refers to the well-known fact that errors arising on different grounds combine in a joint effect. As a result, we speak of the *total bias* as the net effect of all *separate biases* present in data. For illustration, let us assume that N agricultural holdings have to be enumerated in a census of agriculture. However, census lists have included N' holdings only so that $N' < N$. In the collection of data on characteristic x, systematic response errors have entered into the information obtained. In this case we have to deal with two separate biases in the census total. The first is the listing bias which can be defined as

$$D_1 = X' - X \tag{1.16}$$

with

$$X' = \sum_{i=1}^{N'} x_i$$

In other words, the listing bias reflects the omissions in census listing and is independent of response errors. The next separate bias is due to response errors. It can be defined as

$$D_2 = Z' - X' \tag{1.17}$$

with

$$Z' = \sum_{i=1}^{N'} z_i$$

In other words, the response bias exclusively reflects the effect of response errors. We now define the total bias as

$$\begin{aligned} D &= D_1 + D_2 \qquad (1.18) \\ &= Z' - X \end{aligned}$$

Clearly, D is the net effect and its magnitude will depend upon the relative values of D_1 and D_2. Separate biases are algebraically additive. Considerable separate biases sometimes lead to unbiased totals; small separate biases sometimes produce a large total bias. Total area under some crop as obtained in the census may be accurate while the distribution of this area by size classes of holdings may be considerably biased if listing errors primarily affect some specified class of holdings, such as small holdings.

Variances further complicate the issue. If we estimate the total bias in (1.18) as

$$D' = D_1' + D_2' \tag{1.19}$$

the variance of D' will be

$$\sigma^2_{D'} = \sigma^2_{D'_1} + \sigma^2_{D'_2} + 2\ \mathrm{Cov}\,(D'_1, D'_2) \tag{1.20}$$

Let us assume that D_1 and D_2 are estimated from independent samples. In this case the covariance term in (1.20) vanishes. What remains shows that individual variances are pooled together with the result that even a substantial bias may be found insignificant because of the large magnitude of $\sigma^2_{D'}$.

1.9 Checking the quality of data

As demonstrated above, a large number of different errors may appear in each survey. However, while the work on various phases of a survey is going on, it is difficult to see which of the possible errors are occurring, what is the frequency of their occurrence and what are their effects. It is particularly difficult in the course of the work to make an appraisal of the net effects of the integration of different types of errors. Specially designed measures are normally needed to get this information. These measures constitute what we call *quality checking of data* or *evaluation.* The term *quality appraisal* has the same meaning.

Quality checks are divided into two broad categories. In the first of these we put various types of *post hoc techniques* and in the other the *use of sampling methods.* As the name implies, the post hoc techniques are used after the survey is over and data are already tabulated and published. These techniques can therefore be easily applied to data from very old surveys. Post hoc techniques do not require any field work related to units included in the survey. On the other hand, quality checks conducted by means of sampling methods essentially represent field surveys taken at a time which is not too far removed from the reference period of the main survey. Sample checks thus provide fresh information for a sample of units on characteristics being checked and this information is used as a yardstick to measure the quality of the main survey.

There are some other types of quality checks which are on the border line between the categories presented here. In spite of this, the classification into the two categories specified above is kept as it reflects adequately the main characteristics of many techniques.

2. USE OF POST HOC TECHNIQUES IN CHECKING THE QUALITY OF DATA

Different types of post hoc techniques are available. Their uses depend upon the aims of checking, the characteristics being checked, the facilities available in the country, etc. In this chapter some of the most typical and most frequently used post hoc techniques will be presented.

2.1 Comparisons with data from independent sources

A simple method of checking the quality of data collected in a survey is to compare these data with the results available from some independent source of information. The term "independent" refers to a survey different from the one whose data are being checked.

Here is an illustration of such a comparison. In 1956 the FAO expert in the United Arab Republic conducted a sample survey aiming at the estimation of the rate of yield of cotton. The survey was based on crop-cutting. The estimated rate was released on 28 September 1956 and the estimate was placed at 4.23 kintars per feddan.[1] The cotton produced was afterward sent for ginning. The ginning returns made it possible to prepare final data on the yield of unginned cotton. The yield per feddan was found to be 4.17 kintars. The release of this information was made on 14 April 1957, i.e., six months after the sample estimates were released.[2]

Agreement between the figures compared is not a proof of accuracy. More details on the meaning of such comparisons are found in Section 2.5.

The main problem of such comparisons is to establish the degree to which data from various sources are really comparable. In the above example the comparison is fully valid. The sample survey covers the country as a whole and the survey estimates refer to the total production.

---

[1] 1 kintar = 44.928 kg and 1 feddan = 4,200.8 m^2 = 1.038 acres.

[2] FAO: Report to the Government of the United Arab Republic on the development of sample surveys for the estimation of agricultural production, Koshal, R.S. EPTA Report 2006, 1965.

The same is true of data from cotton ginning mills, since all the production of cotton goes for ginning. In others words, both data refer to the same thing and a comparison between the two is strictly valid.

In many other examples the use of this type of comparison is not fully justified. For example, data obtained in an expenditure survey on expenditures on tobacco in a given period of time may be compared with sales data available from tobacco manufacturers. This comparison is valid if the figures cover the same reference period. This is, however, not easy to achieve as the manufacturers' data may include quantities sold or to be sold in retail business at some time outside the reference period.

An illustration of how difficult it may be to have fully valid comparisons is presented in Table 4. This table represents an attempt at the evaluation of data on jute production as obtained from the sample survey involving crop-cutting and measurements. Data from this survey are compared with trade figures and the official forecasts based on the plot-to-plot enumeration of fields under jute.

What conclusions are possible from data in this table? If we want to evaluate the quality of sample survey data by means of trade figures the latter have to be assumed accurate. If this assumption is justified the comparison is valid. However, if doubts are raised regarding the accuracy of data on the consumption of jute in villages, the consumption from the previous year's stock and the contribution from the production

TABLE 4. - COMPARISON OF OFFICIAL AND SAMPLE SURVEY ESTIMATES OF JUTE PRODUCTION IN BENGAL, INDIA, 1944/45 AND 1945/46[1]

(in thousands of bales)

	1944/45	1945/46
1. Consumption during the season		
in jute mills	6 000	6 308
exports	1 050	2 213
in villages	600	600
2. Total	7 650	9 121
3. Consumed from previous year's stock	324	697
4. Jute crop in other provinces	598	862
5. Balance: Bengal crop: trade figures	*6 728*	*7 562*
6. Complete count: official forecast	*4 895*	*6 304*
7. Sample survey	*6 480*	*7 540*
8. Difference between (6) and (5) (as percent of 5)	— 27.2	— 16.6
9. Difference between (7) and (5) (as percent of 5)	— 3.6	— 0.3

[1] Mahalanobis, P.C. and Lahiri, D.B.: Analysis of errors in censuses and surveys with special reference to experience in India. *Bulletin of the International Statistical Institute*, Vol. 38, Part 2, 1961, pp. 401-433.

in other provinces, then the comparison may be misleading. The agreement between the sample estimates and the trade figures would have no importance and could not be taken as a proof of the accuracy of sample data.

Comparisons of data from two independent surveys are impossible unless both surveys deal with the same population. The case of the identical definition of the population of cotton fields was mentioned above. An estimate of the total production of fruits or vegetables as obtained from a sample survey based on measurements and observations on the spot cannot be compared with trade data, as in each country there will be a part of production consumed directly by farmers.

In addition to the population other concepts and definitions also have to be identical. One cannot compare data on income as collected in two different surveys unless the components of income are identically defined. As already mentioned, identity of the reference period is also necessary. It would be useless to compare for quality checking purposes data on the level of unemployment as collected in two consecutive years.

In many cases satisfactory comparisons of the kind mentioned above are impossible. There might be, however, some possibilities for a comparison of data referring to a part of the country or a specific population group. In some countries there are files showing many characteristics of big farms, agricultural co-operatives, manufacturing plants, etc. If census or survey data are tabulated in such a way that these parts of the population are separated, it might be possible to compare data in files with the results obtained in the survey for the groups concerned.

This type of comparison was applied in a number of censuses of population because population registers, etc. provided data for some groups, such as the male population listed for military purposes, veterans, children under some specified age. It should be noticed, however, that such checks have a restricted value from the point of view of the population as a whole. An accurate subtotal is not a guarantee of the accuracy of the grand total.

Some such comparisons may have a special value if there is a population group which is of particular interest. A case in point is the quality of census totals for children. It is a well-known fact that children are very often underenumerated in censuses of population. An elaborated type of comparison might be possible in some countries in checking the quality of these totals. An illustration of such a check as prepared in connection with the 1950 United States Census of Population is presented in Table 5. It will be seen that census totals are compared with figures resulting from birth and death registration. Column 2 contains the

TABLE 5. - QUALITY OF CENSUS TOTALS FOR CHILDREN IN THE 1950 UNITED STATES CENSUS OF POPULATION[1]
(Population figures in thousands)

Age on 4.1.1950 (years)	Period of birth 4/1 to 4/1	Registered births	Adjustment for under-registration of births	Births adjusted for under-registration (2) × (3)	Deaths from birth to 4.1.50	Net immigration from birth to 4.1.50	Expected population on 4.1.50 (4) — (5) + (6)	1950 census count	Difference (7) — (8)	Difference as percent of expected population
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Under 1	1949–50	3 546	1.024	3 631	99	3	3 535	3 147	388	11.0
1	1948–49	3 526	1.028	3 625	120	7	3 512	3 263	249	7.1
2	1947–48	3 613	1.031	3 725	131	10	3 604	3 513	91	2.5
3	1946–47	3 599	1.035	3 725	137	11	3 599	3 561	38	1.1
4	1945–46	2 711	1.043	2 828	126	11	2 713	2 679	34	1.3
5	1944–45	2 790	1.050	2 930	135	11	2 806	2 712	94	3.4
6	1943–44	2 867	1.056	3 028	146	12	2 894	2 765	129	4.5
7	1942–43	2 904	1.062	3 084	152	12	2 944	2 824	120	4.1
8	1941–42	2 565	1.073	2 752	152	12	2 612	2 554	58	2.2
9	1940–41	2 388	1.082	2 584	154	11	2 441	2 344	97	3.9

---

[1] This table is reproduced from Depoid, P.: Rapport sur le degré de précision des statistiques démographiques, *Bulletin of the International Statistical Institute*, Vol. 35, Part 3, 1957, pp. 119-230.

number of births registered in the period observed. Column 3 contains the adjustment factor to allow for underregistration. Column 5 shows the number of deaths. Column 6 refers to immigration and column 7 shows the expected total population that is comparable with the census count. The last column shows the estimated bias in census totals.

Such a comparison requires a great deal of information. In addition, this information has to be sufficiently accurate; otherwise, meaningful comparisons are impossible.

2.2 Consistency studies

The aim of consistency studies in checking the quality of data is to see how data collected in a survey compare with some generally accepted knowledge about the characteristics involved or their relationships. For example, if one wants to evaluate data on marital condition as collected in a census of population or a demographic survey the following relationships, as listed by Jaffe,[3] could be used as valid in most countries:

(*a*) " The total number of married men should be almost exactly equal to the total number of married women. Only in polygamous and polyandrous societies will the two numbers be unequal.

(*b*) " Females generally marry at a younger age than males; accordingly in the younger age groups one would expect to find more men than women single, and there would be but very few widowed or divorced people at these young ages.

(*c*) " It is also known that the death rates for men tend to be higher than those for women and that the remarriage rate for widowers is higher than that for widows, with the ultimate result that there are more widows than widowers in the older age groups. The older age groups also contain many more divorced persons than do the younger age groups, that is, in those countries where divorce is legally recognized.

(*d*) " Ordinarily the age of the husband does not differ too radically from that of the wife. The average size of this difference probably varies from culture to culture. Accordingly, a cross-classification of the two should not reveal many unusual combinations."

---

[3] Jaffe, A.J.: *Handbook of statistical methods for demographers*, U.S. Bureau of the Census, Washington, D.C., 1951.

TABLE 6. - DATA ON ILLITERACY IN FRANCE AS OBTAINED FROM DIFFERENT SOURCES[1]

Year	Percentage of illiteracy				
	According to marriage records		According to military records	According to census of population	
	Grooms	Brides	Recruits	Males 10 years old and over	Females 10 years old and over
1832	...	...	53.3	...	...
1841	...	...	45.6	...	...
1854	31.0	46.0	37.3	...	...
1861	29.0	44.0	32.0	...	...
1872	23.0	35.0	21.5	...	...
1881	15.0	23.0	16.1	...	...
1891	8.0	13.0	9.4	...	...
1901	4.0	6.0	5.6	13.5	19.4
1911	1.8	3.0	4.3	9.7	14.1
1926	0.8	1.1	9.4	5.3	6.9
1931	0.5	0.7	8.5	4.8	5.7
1936	...	...	6.2	3.5	4.3
1946	...	...	...	3.2	3.6

[1] Unesco; *World illiteracy at mid-century*, Paris, 1957.

If the data of a population census are tabulated according to marital condition, age and sex, the consistency of the data collected can be checked in the light of the four points listed above. The discrepancies between data and the above assumptions call for cautious interpretation. In some cases the cultural characteristics of the society or the past history of the population involved will explain some deviations from a general pattern.

In some cases the same information may be available from several sources and the consistency can be studied between the resulting data. An illustration is provided in Table 6 which shows illiteracy in France according to different sources.[4] According to marriage records illiteracy practically disappeared around 1930. Military records offer quite a different picture. They show a high degree of illiteracy for a population group which is probably more literate than others. Censuses of popu-

---

[4] Other examples are available in Depoid, P.: Rapport sur le degré de précision des statistiques démographiques, *Bulletin of the International Statistical Institute*, Vol. 35, Part 3, 1957, pp. 119-230.

lation again show a different picture. The conclusion is that illiteracy data from any of these sources should be treated with extreme caution.

Consistency studies sometimes compare data collected with data resulting from some assumption about the relationships between various characteristics or the pattern of changes of these characteristics. Thus totals of people in different age groups as obtained in a population survey are sometimes compared with the totals computed from an old census by adjusting data according to interim changes. In order to carry out these adjustments some models of population changes are assumed and totals for each age group are computed. Afterward the estimates thus obtained are compared with survey figures. Although conclusions based on these comparisons are subject to many limitations because in fact a population only exceptionally follows closely the pattern of changes assumed, the technique might detect irregularities in data and help in the identification of excessive deviations from the assumed pattern.[5]

Techniques of the above type have been used with success in evaluating age data in advanced age groups. According to common experience old people tend to exaggerate their age and so it is currently held that census totals for age groups after 70 are inaccurate and need to be adjusted. This can be achieved by means of the technique discussed above.[6]

In some special cases satisfactory assumptions about the characteristics to be checked for consistency can easily be made. This will be so in the case of age data tabulated in the form of a single-year age distribution. Some indication of the quality of data presented in this way follows from an observation of the age pyramid. If it shows sharp "sawteeth" this proves the presence of rounding off. A measure of the distortions present in such a case could be obtained by adding up ages ending at each successive digit, i.e., 0, 1, 2, etc., and expressing the total of each of these 10 groups as a percentage of the total population. Low ages are excluded from the check because the age distribution has a somewhat specific character in this domain. Therefore, the calculation starts with, say, 15 years of age. Without any distortion being present, the resulting percentages are supposed to vary around the expected 10 percent which is

---

[5] Further details on this technique and many others, together with examples of their application to data of various censuses, are found in United Nations: *Methods of appraisal of quality of basic data for population estimates*, Population Studies No. 23, New York, 1955 and United Nations: *Population Bulletin*, No. 2, 1952.

[6] Cf. Myers, R.J.: Accuracy of age reporting in the 1950 United States Census, *Journal of the American Statistical Association*, Vol. 49, 1954, pp. 826-831.

arrived at on the assumption of equal probability of each age within a period of ten years.

This method, however, does not lead to entirely valid comparisons because the number of persons of 15 years of age (provided the computation has been started with this year) is normally larger than the number in successive ages. Accordingly, the total number of persons in the group of ages ending in 5 is expected to be larger than the total in the age group ending in 6, etc. This difficulty is avoided in the technique developed by Myers,[7] which consists in starting the count at each of the 10 digits in turn and averaging the totals obtained. Without preferences for digits such as 0 and 5, the total obtained for each digit should be 10 percent of the total population. If deviations from 10 percent are summed up irrespective of their sign, the quantity obtained is called Myers' index of digit preference.

In addition to this there are other indexes developed for similar purposes, such as Whipple's index and Bachi's index. The latter is similar to that developed by Myers but requires more computational work.[8]

These indexes measure a part of the inaccuracy in the age returns, i.e., the digit preference. The errors that do not take the form of accumulation at some digits remain undetected. In spite of this, however, such checks are very useful for comparative studies between countries, within the same country between regions and different strata of the population, between censuses, etc.

With age data available by sex and years of age completed an alternative technique would consist in computing the sex ratio for each year separately. Taking into account the known facts about sex ratios, it is obvious that they could change only slowly in data which do not contain biases. On the other hand, if age reporting is subject to errors, sex ratios might show sharp changes at some critical ages. Unless there is some special explanation for this, such as war casualities, these changes indicate the presence of errors in age returns.

Similar results are arrived at by studying age ratios which are obtained by dividing the number of persons having completed a given age by the arithmetic mean of the number of persons in the two neighboring years and multiplying the result by 100. These ratios should vary around 100.

---

[7] Myers, R.J.: Op. cit.

[8] United Nations: *Methods of appraisal of quality of basic data for population estimates*, New York, 1955; and Bachi, R.: Measurement of the tendency to round off age returns, *Bulletin of the International Statistical Institute*, Vol. 34, Part 3, 1954, pp. 129-138.

However, the variance of these ratios will be smaller in an accurate age distribution than in the case of inaccurate data.

The two ratios mentioned, i.e., sex ratio and age ratio, as obtained in a five-year age distribution, are combined in a technique called the United Nations Secretariat method. This technique leads to an index of the quality of age data which "is affected by differential omission of persons in various age groups from the census count and by tendentious age misstatements as well as by digit-preference."[9]

In many fields of statistics various assumptions are possible which may be more or less useful in studying the consistency of data. For example, if it is assumed that births are equally distributed by hours, this assumption can then be compared with the actual distributions of births over the period of 24 hours. An illustration of such a comparison is given by Golini.[10] As to the distribution of births by months we could assume, in spite of the fact that some "seasonability" in births was found in many countries, that accurate reporting of the date of birth would not deviate excessively from the proportion of total births that corresponds to a particular month according to the number of days in that month. Owing to inaccuracies in reporting, the empirical distributions will be more or less different from what follows from this assumption. For examples and illustrations the reader is referred to Baroni.[11]

More can be learnt about the quality of data if a vast body of statistical information on the topic involved is at hand. For example, if several censuses of population are available along with data on marriages, divorce, separations, remarriages, immigration, emigration and deaths, data on the marital condition of the last census could be checked by comparing census figures with the result obtained from a "balancing equation." In this particular case the equation takes into account all the relevant events that occurred since the penultimate census and gives as a balance the result which is comparable with the last census figure.[12] A similar technique can easily be devised to check the quality of any data, provided sufficiently broad and detailed information is available on changes so that no term is missed in the balancing equation.

---

[9] United Nations: *Methods of appraisal of quality of basic data for population estimates*, New York, 1955.

[10] Golini, A.: L'ora della nascita, *Rivista Italiana di Economia, Demografia e Statistica*, Vol. 18, No. 3-4, 1964, pp. 117-149.

[11] Baroni, U.: La periodicità delle nascite lungo il secolo delle rilevazioni demografiche in Italia (1862-1962), *Rivista Italiana di Economia, Demografia e Statistica*, Vol. 18, No. 3-4, 1964, pp. 151-174.

2.3 Internal consistency

Checking the quality of data by means of a study of their internal consistency involves an investigation of the extent to which the estimates of different characteristics describe the same phenomenon in the same way or can be taken as logically related to each other. Opportunities for such checks arise particularly in surveys with many characteristics on the program.

In an expenditure survey the income in a given period of time taken together with the uses of savings and borrowings represents the upper limit of the total expenditures. In an agricultural survey the sum of areas under various crops cannot exceed the total arable land. Similarly, the sum of totals in all the categories of land utilization cannot exceed the total area of agricultural holdings.[13]

Although studies of internal consistency sometimes refer to final survey data, i.e., estimates of totals and averages, the technique has been much more used in the past to check the quality of data for individual units. The relationships between various characteristics as established in past surveys may be used to check whether the same relationships for the units included in the new survey follow the same pattern. For example, a food consumption survey might give the relationships shown in Table 7. Now, if in the next survey a family is found with a declared annual income of $3,850, 4 members and the per week expenditure on food of $70, the information obtained for this family will be found inconsistent. This information would mean that only a small part of the income is left for expenditures on other items. The expenditure pattern of such a family deviates so much from the established pattern in Table 7 that it would be difficult to accept the information as accurate. Data for such a unit are normally subject to an additional study or even a reinterview.

This type of check is primarily used in editing with the aim of detecting errors in the original material and correcting them before the tabulation starts.

---

[12] Cf. Jaffe, A.J.: *Handbook of statistical methods for demographers*, U.S. Bureau of the Census, Washington, D.C., 1951.

[13] For an example of a study of the internal consistency of data on milk and dairy products as collected in different countries, cf. Louwes, S.L.: Die Statistik der Milcherzeugung und -verarbeitung in den Ländern der Europäischen Wirtschaftsgemeinschaft, *Statistische Information*, 1961, pp. 241-275.

TABLE 7. - FOOD EXPENDITURES OF FAMILIES OF DIFFERENT SIZE [1]

Annual income (dollars) and household size	Average food expenditure per week		
	Per family		Per family member
	Average	Proportion of income (percentages)	
2 000 - 2 999			
2 persons	17.73	33.4	8.65
3 "	21.59	39.8	7.63
4 "	25.14	46.2	6.76
5 and more persons	26.91	45.7	5.24
3 000 - 3 999			
2 persons	21.77	27.7	10.13
3 "	25.83	32.7	8.52
4 "	28.04	38.3	7.34
5 and more persons	31.95	40.3	6.53

[1] Data in this table are taken from: Food consumption of urban families in the United States. *Agricultural Information Bulletin*, No. 132, U.S.D.A., 1954.

2.4 Cohort survival

In checking the quality of listing in censuses of population a study of cohort survival may be very useful. The method is as follows: if a cohort of people as enumerated in a census, i.e., a group of persons belonging to a class of population according to the results of a census, is followed after the census with the aim of establishing all the changes that occur subsequently, such as deaths and migration, the number of persons in the cohort will be known at the time of the next census and can be compared with the size of the cohort as obtained in the new census.

An illustration of the technique is presented in Table 8. Data in this table represent an attempt to estimate the quality of listing of the native white male population in the 1940 United States Census of Population. The expected number of persons in each cohort as presented in column 2 is obtained by adjusting the amount enumerated in 1930 in the light of all the changes that occurred in the period between the two censuses. The difference between the 1940 count and the expected number is given in columns 4 and 5. The penultimate figure in column 4 shows that the persons enumerated exceeded the expected number by 561,700.

TABLE 8. - A CHECK OF THE QUALITY OF LISTING OF THE 1940 UNITED STATES CENSUS OF POPULATION [1]

(All numbers in thousands)

Age (years)	Number in 1930	Expected number in 1940	Enumerated number in 1940	Difference	
				Number (3) — (2)	As percent of enumerated [(4) : (3)] × 100
	(1)	(2)	(3)	(4)	(5)
0 to 4	5 036.9	. . .	. . .	. . .	. . .
0	965.2	. . .	. . .	. . .	. . .
1	960.3	. . .	. . .	. . .	. . .
2	1 023.0	. . .	. . .	. . .	. . .
3	1 046.6	. . .	. . .	. . .	. . .
4	1 041.9	. . .	. . .	. . .	. . .
5 to 9	5 497.3	. . .	. . .	. . .	. . .
10 to 14	5 265.8	4 926.0	5 231.9	+305.9	+ 5.8
15 to 19	4 907.3	5 411.1	5 433.5	+ 22.4	+ 0.4
20 to 24	4 346.9	5 152.0	5 014.7	—137.3	— 2.7
25 to 29	3 731.8	4 769.9	4 698.4	— 71.5	— 1.5
30 to 34	3 408.6	4 208.0	4 230.3	+ 22.3	+ 0.5
35 to 39	3 278.8	3 587.4	3 724.2	+136.8	+ 3.7
40 to 44	2 771.5	3 235.4	3 338.4	+103.0	+ 3.1
45 to 49	2 411.9	3 051.3	3 025.7	— 25.6	— 0.8
50 to 54	2 092.8	2 500.8	2 568.4	+ 67.6	+ 2.6
55 to 59	1 670.6	2 080.9	2 054.2	— 26.7	— 1.3
60 to 64	1 305.3	1 691.4	1 659.2	— 32.2	— 1.9
65 to 69	944.8	1 221.5	1 314.2	+ 92.7	+ 7.1
70 to 74	690.0	818.3	873.2	+ 54.9	+ 6.3
75 to 79	376.6	466.0	487.8	+ 21.8	+ 4.5
80 to 84	165.8	234.2	250.4	+ 16.2	+ 6.5
85 to 89	57.0	73.8	81.8	+ 8.0	+ 9.8
90 to 94	11.7	15.1	17.3	+ 2.2	+ 12.7
95 to 99	1.9	1.9	2.9	+ 1.0	+ 34.5
100 and over	0.3	.1	0.3	+ .2	+ 66.7
Unknown	36.5				
TOTALS:					
All ages	48 010.1	43 445.1	44 006.8	+561.7	+ 1.3
20 years of age and over	. . .	33 108.0	33 341.4	+233.4	+ 0.7

[1] This table is reproduced as an abridged version of Table 8 in Jaffe, A.J.: *Handbook of statistical methods for demographers*, U.S. Bureau of the Census, Washington, D.C., 1951.

The result of this study does not suggest any clear conclusion. The above number of 561,700 persons might be interpreted as a bias in the 1940 Census. Such a conclusion follows from the assumption that the 1930 count was accurate. Another possible conclusion is that the 1940 count was accurate. There are, of course, other possibilities, such as the assumption that neither census is accurate, that vital statistics are inaccurate, etc.

Another illustration of the same technique for a case where data are available for a larger number of consecutive censuses is presented in Table 9. Data in this table refer to censuses of population in Puerto Rico. It can be seen that the size of the cohort at the time of the census has been decreasing continuously, as it should. A glance at the figures for males shows that there is something wrong with the census results for 1920. There is a considerable number of "disappearances" in this census. A good number of them "reappeared" in 1930. The figure for 1920 is not in line with the other data and it indicates the possibility of a serious number of listing errors.

If the cohort survival technique is to be applied, well-established vital statistics are needed to prepare adjustments of census data and compute the expected population at the moment of the new census. Instead of

TABLE 9. - COHORTS OF MALES AND FEMALES, AGED 5 TO 14 YEARS IN 1899, TRACED IN PUERTO RICO THROUGH SUCCESSIVE CENSUSES TO 1940[1]
(All numbers in thousands)

Sex and age	1899	1910	1920	1930	1940
MALES:					
5 to 14 years	138	...	...	...	...
15 to 24 years	...	128	...	...	...
25 to 34 years	...	...	82	...	...
35 to 44 years	...	...	...	83	...
45 to 54 years	...	...	...	...	67
FEMALES:					
5 to 14 years	130	...	...	...	...
15 to 24 years	...	115	...	...	...
25 to 34 years	...	...	92	...	...
35 to 44 years	...	...	...	82	...
45 to 54 years	...	...	...	...	59

---

[1] This table is reproduced from Jaffe, A.J.: *Handbook of statistical methods for demographers*, U.S. Bureau of the Census, Washington, D.C., 1951.

vital statistics life tables might be also used. Neither of these may be available in some countries and this is why the use of this technique is not always easy.

2.5 Drawbacks of post hoc techniques

All the post hoc techniques have some common characteristics that need to be borne in mind while planning their use. They are:

(*i*) If the evaluation consists in a comparison with data from an independent source, the justification of the procedure lies in the assumption of accuracy of the yardstick used. The validity of this assumption is often the real issue.

(*ii*) Every comparison implies the existence of data collected previously on the same topic. This applies even more to techniques used for consistency studies of time series. It restricts the use of all such techniques to countries with a considerable tradition in statistical work and with several surveys made in the past on the same topic.

(*iii*) Most efficient uses of post hoc techniques are restricted to some classes of the population (children included in birth registers, veterans, large farms) or some items on the survey program (area statistics which are compared with cadastral data, etc.). There is therefore no way of extending the evaluation to items where it might be particularly desirable.

(*iv*) The application of post hoc techniques results in impressions on quality rather than in its numerical measures. Since both the users and the statisticians are, for all practical purposes, primarily interested in quantitative measures, the use of the techniques described does not offer much toward meeting the basic requirements with regard to the quality of data.

(*v*) All the post hoc techniques refer to final survey results, such as totals, averages, proportions, etc. Any information on individual errors and circumstances under which they appear is thus lost and the evaluation of data does not furnish any guidance on what has to be changed in the techniques of collecting data in order to achieve better results. Accordingly, the above techniques have little to offer for the improvement of the quality of data in subsequent surveys.

(*vi*) Most of the post hoc techniques have been developed in the field of population statistics where the regular occurrence of various events makes it less difficult to distinguish the deviations from some stable pattern than is the case in other fields. It is sufficient to remember the stability of the sex ratio or the distribution of births over years and dates ending in different digits. In other branches of statistics where characteristics change more irregularly and even in short periods of time, it becomes more difficult to find valid assumptions for studying the consistency of data. The scope of consistency checks is thus greatly reduced, although some opportunities often exist for studying internal consistency in surveys.

2.6 Advantages of sampling methods

If sampling methods are used for checking quality, the difficulties listed in the preceding section are either eliminated or highly reduced. With sampling methods the results of a check are based on data collected for a sample of units. Therefore, quality checking of survey data is possible independently of whether any similar survey has ever been made before. This makes the use of sampling methods suitable for quality studies in all countries, including those that are at the initial stages of their statistical activities.

Moreover, sample checks can be extended to any part of the population involved. For this purpose the sample has to be selected from all parts of the population. If check data are then collected for the units in the sample, by examining the difference between data collected in the check and those obtained for the same units in the survey being evaluated it is possible to estimate the quality of survey results for any part of the population and also for any item on the program.

Sample checks refer to elementary units themselves or small clusters thereof, such as persons, holdings, area segments, etc. If systematic records are kept on various characteristics of units where errors have been detected, it is possible to have a broad knowledge of where errors appear, under what circumstances, on what characteristics, etc. This knowledge is extremely valuable for planning the improvement of techniques for collecting data. In this respect one can go as far as to say that, without sample checks, it is hardly possible to acquire a systematic insight into the basic aspects of errors as well as into the measures that might be useful for the improvement of the quality of data.

An additional advantage of the use of sampling methods for the purpose of quality checking has been pointed out on many occasions. Namely, the existence of a quality check which may be extended to any unit of the population lays a kind of pressure on both the respondents and the enumerators to supply more accurate data.[14] Nobody likes to have errors detected in his work.

Sample checks are also independent of any assumption concerning the relationships of various characteristics, the order of magnitude of quantities being evaluated, etc. Thus sampling methods are a useful tool and sometimes the only one available for checking quality if little or nothing is known about characteristics included in the survey program.

Conclusions drawn from sample checks are based on differences between data collected in the original survey and data obtained in the check survey. This makes it possible to arrive at numerical estimates which are obviously more in accord with the users' interest than the alternative descriptive statements normally resulting from the application of post hoc techniques.

---

[14] An example showing improved quality of work because of the use of sampling methods for checking purposes is found in Hill, D.: The economic incentive provided by sampling inspection, *Applied Statistics*, Vol. 9, 1960, pp. 69-81.

3. MEANING OF QUALITY CHECKING

In this chapter we propose to deal with the meaning of quality checking. Different views are possible on this issue and it may therefore be useful to devote some time to the discussion of this topic.

3.1 Assumption of random response variation

Let us assume the following situation. N fields under some specified crop are listed for a survey aiming at an estimate of the toal area under this particular crop. Each field belongs to a different farmer. The information (called here response) on the area of each field is collected through a mail survey. Before they enter the information in the respective questionnaires farmers are requested to go to the fields concerned and make a visual estimate of the area of their fields. We shall now assume that each farmer has the possibility of repeating the same visual estimate an infinite number of times. For each unit of the population an infinite supply of responses will thus be available. Successive trials in getting a response for the i-th field can then be conceived as random drawings of chips from a container where the supply of responses for this unit is stored. The survey value of the response obtained for the i-th unit at the j-th trial will be designated by z_{ij}. The corresponding true value is x_i. The subscript j is dropped as the true value is constant over trials. The response error of the j-th trial is obviously $d_{ij} = z_{ij} - x_i$.

The model of response variations assumed here makes it possible to consider z as a random variable with regard to both subscripts i and j. Being a random variable, z has its expected value over all the trials for the i-th unit as well as over all the trials for all the units. For the same reason we can also define the variance of z.

Let us now assume that a complete coverage of all the farms was made with a view to ascertaining the total area under the crop. In the light of the model assumed the survey will yield the result:

$$\begin{aligned}\text{total area} &= \sum_{i=1}^{N} z_{ij} \\ &= Z_j \\ &= \text{survey value of the population total as obtained in the } j\text{-th trial.}\end{aligned}$$

It is obvious that Z_j represents a sample estimate of the quantity $\underset{j}{E} Z_j$. The sample consists of one single response selected at random at the j-th trial from an infinite supply of responses available for the i-th unit of the population. Speaking in a general way this means that totals or averages obtained in complete enumeration censuses are conceived as sample estimates. These estimates are subject to sampling errors like all other estimates resulting from sample surveys. The sampling error of Z_j is defined as

$$\sigma^2_{Z_j} = N\sigma^2_{zw} \tag{3.1}$$

with

$$\left.\begin{aligned}\sigma^2_{zw} &= \underset{i,j}{E}\,(z_{ij} - \underset{j}{E}\, z_{ij})^2 \\ &= \underset{i}{E}\,\sigma^2_{z_i} \\ &= \underset{i}{E}\,(\text{individual response variance}) \\ &= \text{response variance.}\end{aligned}\right\} \tag{3.2}$$

Various aspects of the application of this approach to sample surveys are dealt with by Cochran;[1] Fellegi;[2] Hansen, Hurwitz, Marks, and Maul-

---

[1] Cochran, W.G.: *Sampling techniques*, Second edition, John Wiley, New York, 1963.

[2] Fellegi, I. P.: Response variance and its estimation, *Journal of the American Statistical Association*, Vol. 59, 1964, pp. 1016-1041.

din;[3] Hansen, Hurwitz, and Bershad;[4] Hansen, Hurwitz, and Pritzker;[5] Hansen, Pritzker, and Steinberg;[6] Murthy;[7] Pritzker and Hanson,[8] and others.

The advantages of this approach are obvious. The association of sampling errors with the results of a complete enumeration census reminds users of data that another survey, taken at a different time, might lead to results which would be more or less different. The measure of the variations involved is based on the magnitude of the response variance.

As a result of this approach we have $Z_j = X + D_j$ with $D_j = \sum_i^N d_{ij}$. The quantity D_j is the response bias of the total Z_j. Also, D_j is an estimate of the overall response bias, $\bar{D}$, which is defined as

$$\sum_i^N \underset{j}{E}\, d_{ij} = \sum_i^N D_i = \bar{D} \qquad (3.3)$$

with

$$\begin{aligned} \underset{j}{E}\, d_{ij} &= D_i \\ &= \text{individual response bias.} \end{aligned}$$

Quality checking of the response would consist in this case in the estimation of $\bar{D}$. For this purpose a simple random sample of n units is selected and for each of them the true value of the characteristic checked is ascertained. In our example it would mean that for estimation purposes a sample of n primary units is available and a sample of one single response error in each of the selected primary units. This makes it possible to estimate $\bar{D}$ and $\sigma^2_{D'}$ as $\bar{D}'$ and $s^2_{D'}$ respectively. The variance $\sigma^2_{D'}$ is used in testing the significance of the estimated bias. It is also used in defining the root mean square of the total Z_j as

$$\zeta^2_{Z_j} = \sigma^2_{Z_j} + \bar{D}^2 \qquad (3.4)$$

The equation (3.4) completes the aims of quality checking. The question however arises whether quality checking of this type is possible. Can we estimate the bias $\bar{D}_i$, what are the difficulties arising therefrom and what are the consequences of these difficulties? This issue will be discussed in sections 3.4 and 3.5.

Another problem is the response variance. The estimation of $\sigma^2_{Z_j}$ requires the repetition of the response for a sample of units. Such a repetition would make it possible to estimate the individual response variance, as well as σ^2_{zw}. However, we have to examine whether the repetition of the response is feasible. This examination takes us to the problem of the nature of response variations which will be discussed in the next section.

3.2 Unique nature of response

Each particular response must be conceived as the joint effect of a number of factors. Factors governing the response obtained in an interview survey on income will be: the respondent's memory, willingness to cooperate, understanding of the concept of income, the enumerators' ability to present the question if enumerators are used, etc. When the sides of a field are paced with the aim of finding its area, the response will be affected by the slope of the field, its evenness, the meteorological conditions, the pace length of the person carrying out the measurement, etc.

Some factors affecting response have a random character. Random factors are present in any attempt at collecting data. They affect response in a way typical of random deviations from the true value. Namely, the average of the resulting deviations is equal to zero while the variance of these deviations is different from zero and varies in magnitude from variable to variable. Examples of such factors are as follows:

---

[3] Hansen, M.H., Hurwitz, W.N., Marks, E.S. and Mauldin, P.W.: Response errors in surveys, *Journal of the American Statistical Association*, Vol. 46, 1951, pp. 147-190.

[4] Hansen, M.H., Hurwitz, W.N. and Bershad, M.A.: Measurement errors in censuses and surveys, *Bulletin of the International Statistical Institute*, Vol. 38, No. 2, 1961, pp. 359-374.

[5] Hansen, M.H., Hurwitz, W.N. and Pritzker, L.: The estimation and interpretation of gross differences and the simple response variance, *Contributions to Statistics*, Pergamon Press, Oxford, England, 1963.

[6] Hansen, M.H., Pritzker, L. and Steinberg, J.: The evaluation and research program of the 1960 censuses, *Proceedings of the Social Statistics Section, American Statistical Association*, 1959, pp. 172-180.

[7] Murthy, M.N.: Assessment and control of non-sampling errors in censuses and surveys, *Sankhyā*, Vol. 25, Series B, 1963, pp. 263-282.

[8] Pritzker, L. and Hanson, R.H.: Measurement errors in the 1960 Census of Population, *Proceedings of the Social Statistics Section, American Statistical Association*, 1962, pp. 80-90.

(*a*) the mood of the respondent
(*b*) the enumerators' mood
(*c*) fatigue
(*d*) the time of calling
(*e*) the enumerator-respondent interaction
(*f*) the respondent's interest in the survey, etc.

Occasional factors represent another category. They reflect the specific circumstances of each attempt at collecting data. These factors originate either in some measures taken in preparing the survey or in some special events that occurred in the population concerned at the moment of collecting data. Examples of such factors are:

(*a*) the quality of the enumerators selected
(*b*) the effects of the publicity campaign
(*c*) special temporary features of the external circumstances, such as housing problems, psychological tension created as a result of wars, a difficult economic or political situation, etc.

The nature of these factors is such that they appear irregularly in time and are difficult both to predict and control in their effects. Within one single survey these factors produce some pattern in the data. For example, if farmers are afraid that their data about the number of livestock might be used for taxation purposes the tendency to underreport will arise.

The third group consists of cyclical factors which appear in some kind of cycle or affect responses at rates varying in a cyclical way. Examples of such factors are:

(*a*) the season of the year and the resulting variations in respondents' activities, their availability for surveys, their interest, willingness to co-operate, etc.
(*b*) the season of the year and the resulting meteorological conditions (snow, rain, mud, high temperatures)
(*c*) holidays, vacation periods, etc.

In agricultural surveys it may be difficult to get satisfactory co-operation from the farming population during the peak of agricultural activities. In rainy seasons the enumerators might replace the inaccessible households by others more conveniently located. Measuring distances in mud or snow or very low temperatures may give quite different results from those achieved during some other season.

Trend-creating factors constitute another category. These factors produce some trend in the response over time or a situation where successive responses are somehow connected with the preceding ones. The following are examples of factors of this type:

(*a*) Memory. After the response has been given to survey questions, or at least to those which require more thought and work, the respondents will remember the answers given for some time, and any attempt at repeating the survey with the same items on the program will often result in the repetition of the initial response. In other words the conditioning effect is produced. As opposed to this there is the effect of the increasing number of memory lapses with regard to many items on survey programs as time passes and events recede further into the past.
(*b*) Increasing experience and knowledge. Each survey has some educational value and offers respondents a chance of learning about concepts used in statistics, units of measurement, etc.
(*c*) Reluctance to co-operate for some time with any survey after participation in a survey.

As a result of the joint effect of such factors we consider that circumstances under which successive responses are generated change from one trial to another. If the response variations were governed solely and exclusively by random factors, one could accept, as the aim of quality checking, the model and the approach presented in the last section. However, the actual circumstances of statistical surveys are different. The man who has no idea of what income is becomes a different man after he has learned the concept. Once he has taken the trouble to compute his income by reviewing records and going from one component to another, he will not take the same trouble again. In this situation there will be no response variation at all. Moreover he may well refuse to co-operate if asked to provide the same information again. In other words, the first attempt has made other trials impossible. The respondent who has carried out some measurement, although he can theoretically repeat the procedure a large number of times, normally will not do so and will stick to the first result. It appears that some particular play of all the factors that contribute to the process of building up the response cannot be repeated. After the response is formulated the underlying circumstances disappear for ever and the rest is simply the result of memory, refusal to co-operate, stubborn repetition of the same information for prestige and other reasons, etc. Each response must therefore be considered a product of particular circumstances and cannot be regarded as representative of the same population as the responses resulting from other trials. In fact, each response must be considered a product of some response generating system that disappears after the trial has been carried out. Thus, if successive responses are available they only represent members of different populations. The *unique* nature of each response must be stressed, and the fact that in statistical surveys the only thing involved in defining the quality check is the response obtained under *some specific circumstances.* As a consequence, the use of z_{ij} must be discarded and all other concepts that follow therefrom such as individual response bias, response variance, etc. We are left with the unique product of the survey, z_i, which is materially available in survey documents or questionnaires.

In this situation quality checking merely entails a *calibration*[9] of the information available in documents. The response to a survey is treated in the same way as we treat the information on the weight of a product in the acceptance sampling. The calibration or the quality checking then involves the comparison of the information presented with some other information obtained about the same characteristics and the same units.

As regards the estimation procedure that follows from the view adopted here, data collected in the original survey are considered as the supplementary information. From their comparison with check data we establish differences $d_i = z_i - x_i$ and estimate the bias in the census total and the variance of the estimated bias, test the significance of the bias and correct survey results to allow for biases if necessary.

The interpretation of the results yielded by the check follows the same lines. Census totals are regarded here as unique products of circumstances under which data were collected. There is also a general understanding that under the different conditions of another trial they might vary in a greater or lesser degree. In this connection the quality check gives information about the quality of the data being checked. Conclusions drawn from the check offer guidance in judging the quality of other surveys insofar as the underlying circumstances are the same or comparable.

---

[9] Cf. Yates, F.: *Sampling methods for censuses and surveys*, Third edition, Charles Griffin, London, 1960.

3.3 Difficulties of generalization

In the last section a view of the nature of response in statistical surveys was accepted that may encounter difficulties varying in degree according to characteristics and circumstances. For example, if experienced persons, such as people specialized in ground surveys, are asked to make measurements of the areas of their fields, the model of random response variation may approach the true picture. In this case it might be appropriate to think of an infinite supply of responses arising from the same circumstances. Accordingly, a view of quality checking presented in section 3.1 would also be appropriate.

For reasons mentioned in the last section interview surveys would normally call for the views of section 3.2. However, even here the circumstances may be very different thus making different views possible. In countries where people are exposed to many surveys there is probably a tendency on the part of respondents to attach little importance, if any, to an interview, so that repeated trials might be considered as random response variations. A respondent whose attention and feelings are not greatly involved in the survey will not be affected by the interview and at the time of another trial his attitude may be unchanged. However, the situation may be quite different if surveys are taken at long intervals only and preceded by a systematic publicity campaign, as is the case in many countries with government surveys and censuses. These are events that imprint themselves on the memory as they require attention, study of information, preparation of the response, etc. This is a typical situation where the response must be viewed as a unique outcome of some specific circumstances.

Even in fields where there might be some grounds for a different attitude, such as the above-mentioned example of measurements, it may also be necessary to consider that successive responses do not represent the same population. A scientist in a laboratory may be able to carry out several measurements with equal care so that all of them can be taken as resulting from the same circumstances. However, if farmers are requested to measure the areas of their respective fields it may be difficult to consider successive measurements. Even if they agree to repeat the procedure, its outcome will be that of the first trial or some other version of it if they are convinced that there should be some variation. We therefore consider that the attitude taken in the last section corresponds to most practical situations although it may not be fully adequate in some cases.

3.4 Accuracy checks

Another difficulty in defining the aims and purposes of quality checking is connected with the interpretation of data collected in the check itself. These data are used to judge the quality of the original survey. Obviously, before any statement regarding the original survey can be made, a criterion for the judgment must be clearly established. The question therefore arises of the quality of the data collected in the check.

In the ideal case check data would represent true values. In this case the check provides estimates of biases and of the mean square errors. This is what is understood by *accuracy checks* in the strict sense.

Accuracy checks do not represent a practical impossibility as is sometimes claimed. The present author believes that accuracy checks are more frequent than might appear at first glance. There are some fields of statistical work where there is no serious difficulty in getting true values. Examples are the quality control of editing, coding, and punching. The same is true of checks based on documents, such as birth certificates, school certificates, identity cards, cadastral records, etc., and those based on measurements, such as area measurements, crop weighings, etc.

It is obvious that in some of the cases listed here it may be impossible to get true values in the absolutely strict sense of the word. This is why some authors refuse to accept the possibility of accuracy checks. Birth certificates are sometimes inaccurate. However, for practical purposes this is immaterial as some wrong dates of birth in an adequately organized registration system do not change the picture. Similarly, slight errors in measurements can also be disregarded as their effects will be negligible.

In addition to the estimate of the bias, accuracy checks will also provide a measure of the *stability* of response in the original survey. This measure is given by the variance of individual errors, i.e., σ_d^2. For purposes of comparison it may be useful to define the *index of stability* of the response as

$$I = 100 \times \frac{\sigma_{X'}^2}{X'} \tag{3.5}$$

With a simple random sample of n units selected with replacement the

meaning of symbols involved is the following:

$$D' = \frac{N}{n} \Sigma d_i$$

$$X' = Z + D'$$

$$\sigma_{X'}^2 = \frac{N^2}{n} \sigma_d^2$$

These measures of stability reflect the number and the magnitude of individual errors. Sometimes they may represent an important indication of the direction in which efforts for the improvement of work should tend. If σ_d^2 is the result of a large number of small errors, quite different steps may be needed from those required if some units of a particular type were responsible for it.

In designing the accuracy checking considerable attention will be needed in the specification of the method of collecting data. Achieving true values may prove to be a very difficult procedure even in cases where the opposite is expected. In this respect the following two illustrations may be useful.

The first of these illustrations is taken from a paper by Hansen and Pritzker.[10] It refers to the quality check of listing taken in connection with the 1950 United States Census of Population. In this check the best census enumerators were employed for the field work and an attempt was made to obtain some insight into the accuracy of the data they collected. For that purpose some mistakes were purposely introduced into the transcribed census lists with the aim of checking how many of them the enumerators would report. Namely, one person was omitted from the census lists in the case of one out of every 100 households of three or more persons. A total of 165 persons were thus omitted and the check enumerators failed to report 21 of them. In other words, all the efforts made in organizing this check were not sufficient to prevent a specially selected staff from making errors.

---

[10] Hansen, M.H. and Pritzker, L.: The post-enumeration survey of the 1950 Census of Population; some results, evaluation and implications, *Annual Meeting of the Population Association of America*, Ann Arbor, 1956.

In this respect the situation does not differ from methods considered superior, such as the counting of physical objects, the measuring of areas and distances, weighing, etc. Hendricks[11] has presented some interesting cases from this category. The following four cases were reported:

(*a*) "In connection with cotton forecasting studies, boll counts are made on cotton plants in ten-foot row segments in selected fields. During the 1954 season, the accuracy of these counts was tested by stripping bolls from sample plants and counting them again after the 'on-plant' counts were completed. On August 1, the on-plant count was found to be 5.9 percent too low, but this error dropped to only 1.4 percent in a second survey as of September 1.

(*b*) "A similar study was made as of August 1, 1955, except that individual bolls were tagged, instead of being removed from the plants, after being counted to permit tracing the life histories of individual bolls during the growing season. A count of the tags used on each plant served as a verification of the original boll count. The tag count may not be as precise a verification as a strip count because some bolls can be missed in tagging, but it did provide some check on the original count. This study indicated a 2 percent shortage in the original count. As might be expected, the percentage under-count increases as the number of bolls per plant increases.

(*c*) "As part of an objective orange forecasting study in Florida, early-season fruit counts were made recently on sample limbs of orange trees. After one member of a team completed such a count on a tree, the count was repeated by another member. In the first test, counts on the same limbs of 23 trees showed an average of 82.8 oranges per limb counted by the first man and 86.8 by the second man.

(*d*) "Tests of the accuracy of 'on-tree' fruit counts have also been made in California. Here immature fruit was stripped from the trees after counting and counted again. Peach counts in 1952, made while fruit was still on the tree, were about 10 to 12 percent low. The introduction of refinements in counting technique resulted in some improvements the next year, but the 'on-tree' counts were still 8 percent low. Similar studies on pears in 1952 showed 'on-tree' counts to be about 12 percent low."

---

[11] Hendricks, W.A.: Non-sampling errors in agricultural surveys, *Improving the quality of statistical surveys*, American Statistical Association, 1956, pp. 31-39.

This long quotation from Hendricks' paper serves to make it clear that there is no easy way to accurate data.

3.5 Approximation to true values

In the examples mentioned above it was felt that true values existed independently of the survey response and that to achieve them merely required some effort and patience. The situation is quite different in many other cases. In an expenditure survey it will be a sheer practical impossibility for people to provide accurate data. The same is true of food consumption surveys based on either of the methods used in this field. Even if people were interested in keeping accurate records for themselves, they could hardly be expected to remember all of the food consumed and to be able to evaluate accurately its weight and composition.

In dealing with such characteristics it is impossible to carry out accuracy checks as these were defined with regard to true values. In such cases it may be possible to approximate true values with varying degrees of success. Among the factors involved in securing such approximations the first is the requirement that the procedure of obtaining check data be *independent* of the procedure used in the original survey. We call the method used in the check independent if the process of response generation in the check is not influenced by the respondent's earlier exposure to the original survey. Thus, eye estimates are independent of pacing or measurement. If a person is interviewed for information on age in the original survey and the birth certificate is used as the source of information in the check, we call the two procedures of obtaining data independent.

Another device is to direct the check toward the *best source of information*. From the quality point of view all sources of information are not equal. A housewife may be a less reliable source of information on income than the head of the household, while in matters of expenditure on food and food consumption the opposite may be true. In a large-scale survey a systematic orientation on the best source of information may be impossible owing to the complex nature of the work. In quality checks it will normally be possible to designate for each characteristic the best source of information and actually to get data therefrom.

Neither of these two measures will remove difficulties that arise in the interpretation of the data collected in the check. It is obvious that the identical responses in the two surveys will often mean accurate responses. If identical data are obtained from two different persons of the same household on the kind of food consumed on the day preceding the interview this can probably be interpreted as accuracy of that information. In many other cases the identical responses will not be a sign of accuracy. Two different persons may evaluate the rate of yield of a crop in an identical way without approaching accuracy.

The difficulties of interpretation also arise in connection with cases of discrepancy between the check and the original survey. Is this discrepancy a measure of error or how close are the check data to their respective true values? The only way of improving the possibilities of interpretation is to choose for the collection of data in the check a method which is *superior* from the quality point of view to the one used in the original survey. A method is called superior if it has a better chance of achieving or approximating true values than the method used in the main survey. If the original survey used interviewing, a superior method will consist of measurements, counting physical objects, looking up documents, etc.

If the check is using a superior method of collecting data, the estimated difference between, say, the totals in the original survey and the check is the difference between the original survey and other more accurate information. This is the available *substitute* for the bias which can be used in estimating the mean square error. Needless to say, the resulting substitute for the mean square error will no longer be the measure of the accuracy of the original survey. The accurate knowledge of the magnitude of the bias is missing.

Approximations to true values also make it possible to measure the stability of the response in the main survey. Let us use here y_i to designate the response obtained in the check for the i-th unit in the sample. The variable y is used to distinguish this response from true values. The quantity $d_i' = z_i - y_i$ will be called the *individual difference* in order to avoid the use of the concept of errors which was defined by means of true values. The variance needed to measure the stability is then $\sigma^2_{d'}$. It shows the extent to which data of the original survey deviate from the corresponding information of another survey based on an independent and superior method of data collecting. The index of stability is defined by substituting $\sigma^2_{d'}$ for σ^2_d in (3.5).

Clearly, the value of all the conclusions based on $\sigma^2_{d'}$ depends upon the extent to which check data are superior to the original survey.

3.6 Supporting evidence

In the last two sections it was said that conclusions from data collected in the check would be different depending upon whether true values are obtained or some more or less successful approximations to true values. Which of these cases is applicable in a particular check? How do we know what kind of data were collected in a check and what conclusions are justified?

These questions touch the very essence of quality checking and indicate the real difficulties of this work. It would be wrong to think that the design of the check supplies the answer to these questions. Of course, the design is essential for the success of the check. However, the design itself, no matter how good it may be, does not show how fitted data collected in the check may be to act as a yardstick for the measurement of the quality of the original survey. In fact, the implementation of the check may yield quite unexpected results. The staff responsible for measurements may be careless and repeat the information given in the original survey without carrying out the field work. Therefore some special evidence is needed on how the check was carried out, to what extent the instructions were applied, what difficulties occurred in the collection of data on various characteristics, etc. This information is called the *supporting evidence* as it is used in the formulation of the judgment on what was achieved in the check and what can really be measured with the help of check data.

The collecting of the supporting evidence is a special problem of each quality check. In many cases it may take the form of another small-scale survey superimposed on the check survey. The situation which we are facing is therefore as follows: if we want to know how good the data of a survey are, we need a quality check. However, if we want to know the meaning of the quality check and the light it throws on the original survey, we need a check of the basic check or a kind of supercheck.

This is an unfortunate situation, but there is no other way of obtaining an insight into what was achieved in the check. Without such supporting evidence the basic check has no value, as all kinds of unjustified conclusions might be drawn from it. From a check that is less accurate than the original survey it may be concluded that the original survey was grossly inaccurate.

The main bulk of the supporting evidence will be obtained if information is collected on the *general adherence* of check enumerators to their respective instructions. If this information shows that the check enumerators actually did what they were expected to do according to instructions, confidence in the value of the check will increase.

Another important part of the evidence needed comes from the information on what happened to the check at its *critical points*. Each quality check has some critical points which have a considerable bearing on the quality of data. For example, in designing the check various assumptions are sometimes made about ways in which some procedures will be carried out. Thus, the check might provide that (*i*) a sample of households be selected from the lists of names of heads, (*ii*) some characteristics be checked by looking up the documents, (*iii*) economic characteristics such as the branch of industry and monthly salary be taken from the payroll slip, and (*iv*) the number of various kinds of livestock should be obtained by counting.

These provisions have several weaknesses. Firstly, some lists contain several persons with the same name so that enumerators are confused afterward as to the relevant house. Sometimes they go to a wrong house and report values of characteristics quite different from those of the original survey. With other characteristics, such as data of birth or degree of education, relevant documents may not be available for many people. Similarly payroll slips may only be available with large agencies and companies, while counting of livestock may be impossible for several reasons, such as refusal of permission to make a count, the fact that the livestock is grazing or that only a part of it is in the stable. The answer to the question of what can actually be measured by data collected in the check will depend largely upon the information obtained from the supporting evidence on how the work was carried out at such critical points. It might be useful to involve check supervisors in the collecting of information for the supporting evidence. However, the supervisors will not be able to supply this information unless their work is organized accordingly. This means primarily that instructions for supervisors have to specify each step that the supervisors are expected to take in this respect.

Secondly, an adequate system of recording data and observations is needed. Finally, some arrangements for the distribution of supervision will also be necessary, so that the location and duration of the supervisors' work are controlled. In some cases their information will not be useful unless collected from units selected at random. For example, if the check instructions provide for the measurement of areas of fields operated by selected farmers, a quantitative judgment on how well the measurement

was carried out cannot be formulated unless the supervision is itself carried out on a random basis.

According to the view of quality checking adopted in this book the check supervision represents an integral and essential part of quality checks. Without supervision as conceived here the quality check is not complete.

It may also be worthwhile to make the final observation that this systematic collecting of information on the success of the check will be a useful pressure on the check designers. If the report of quality checking is supposed to present evidence on the quality of the data collected, this aim cannot be achieved unless the check survey is analyzed from its elementary phases and a series of steps planned with a view (*i*) to securing in each phase the best possible work, and (*ii*) to obtaining in each phase the information which will show what was achieved. With these requirements imposed designers will be obliged to follow a very systematic approach in the design of the check survey, in the observation of the field work, and in the final analysis of the data collected.

3.7 Other types of quality checks

The use of sampling methods in checking the quality of data has been conceived so far in the form of a separate survey in which it was possible to apply another method and obtain information comparable to that of the original survey. There are some characteristics which do not allow this approach. If a person says in a survey that he would vote for the candidate X it may not be easy to think of another procedure for obtaining the same information that would be independent and more accurate. There is probably no other step available than to go to the same person and ask the same question again. Should the attempt be successful it would yield a repetition of the original response, which has no value as a basis for quality judgments. The position would be the same if it were a question of checking the quality of the statement: "Should I buy a new car next year it will be ...(make)." Other questions of the same kind are frequently found in attitude and preference studies.

How is it possible to check the quality of data in these cases where a check survey is useless?

In such cases we have to accept less ambitious types of checks which do not reveal much about accuracy although they provide information that may sometimes be very useful in judging the quality of the basic survey. In order to understand the nature and characteristics of these checks some properties of accurate data drawn from chapters 1 and 2 are listed below. Although others could also be listed, for our purpose the most important of these properties are:

(*i*) *Zero bias.* This property follows directly from the definition of accuracy. Zero bias is a direct proof of accuracy.

(*ii*) *Stability.* If accurate data on the same characteristics are collected in different surveys and by means of different methods there will be no discrepancy between the resulting estimates. In other words, if two surveys display significant variations in the resulting estimates either one or the other of the two estimates must be inaccurate (assuming the characteristics remain unchanged).

This property is sometimes referred to under the name of *external consistency.*

(*iii*) *Internal consistency.* In a survey covering several related characteristics (such as population, land, livestock, production, etc. in a general agricultural survey or population, income, education, expenditures in family budget surveys) accurate data yield estimates that constitute a coherent whole. A household with a high expenditure on books and concerts in a poor suburban area would not fit the rest of the data from this place.

(*iv*) *Compatibility with existing knowledge.* Accurate data will fit the available knowledge about the subject involved. The variation in time of various characteristics will not lead to drastic deviations of the resulting estimates from the known pattern or generally accepted relationships. It would thus be difficult to accept the information that a person of 17 years of age occupies the position of director of a bank. Available knowledge and experience tell us that a person becomes director of a bank only after a sufficiently long career qualifies him for the post.

(*v*) *Origin in correct procedures.* If individual responses are considered as the final product of a long chain of more or less complex stages of work, it is obvious that accurate data must be preceded by correct work at all the stages as well as by correct application of various procedures included in the survey design. Clearly, totals and some other measurements might represent exceptions from this rule because of compensating effects.

(*vi*) *Product of correct tools.* If in collecting data some tools are used such as compasses in measuring areas or tables of random numbers in selecting random samples, accurate data follow from the use of correct tools and also from their correct application.

From the quality point of view the most important is obviously the first property. The knowledge about this property is the knowledge about accuracy. With other properties the situation is different. Each of them necessarily follows if data collected in a survey are accurate. However, if any of these properties were found to hold in the study of a survey, this does not necessarily mean that the data of that particular survey are accurate. Stable data do not necessarily mean accurate data. The same is true of internally consistent data. The correct use of correct tools is not a guarantee of accuracy.

It is obvious that the most useful type of check is the one that yields information regarding the first property. This aim is achieved with sample checks in which it is possible to get true values or their approximations. Comparisons with data from independent sources sometimes also fall into this category. However, this technique is less universal as subtotals normally lie outside the scope of this approach.

Any information that we might have regarding other properties is merely a more or less valuable substitute for the missing knowledge about the first property. We therefore collect that information if the knowledge about the first property is not available for some reason. Such a reason is the above-mentioned case of characteristics that can hardly be included in a sample check. Another reason might be that the survey was made in the past and quality checking is needed after a long lapse of time.

The information about properties listed from (*ii*) through (*vi*) helps in a negative way in forming a judgment about the accuracy of data. If it is clear that in a survey the sample was selected correctly, the estimation procedure was adequate and the instructions for interviewers covered all the points where confusion might arise, it can be said: "As far as these points are concerned there is no reason to believe that the data are inaccurate." This does not mean that the data are accurate. However, if all the properties listed except the first are checked in this way, several statements of this type can be made. Taken together, these statements strengthen the belief that the survey is reliable. These statements demonstrate that the work, insofar as it was possible to check, was carried out correctly.

It is not difficult to see that the checks we are talking about can be carried out in a number of ways. It has already been pointed out that stability can be checked by accuracy checks in the proper sense as well as by means of checks which do not achieve more than approximation to true values. In some cases several surveys may be available with matching units and a number of common characteristics. For example, we might have the data of a labor force survey and the census of population. Data for matching units can then be used to study the stability. However, if the two surveys are equally accurate or no information is available as to their relative merits in this respect, there will be difficulties with the interpretation of the variance obtained.

Internal consistency and compatibility with existing knowledge are checked by means of the post hoc methods mentioned in the last chapter.

The last two properties, i.e., (*v*) and (*vi*), are checked by means of a separate examination of relevant steps in the work carried out and instructions prepared. The problems involved are discussed in chapters 4 and 5.

It will now be clear that a systematically prepared program of quality checking can hardly be restricted to any single type of check. This is also true in the case of the accuracy check. However careful the preparation of the check, it is never possible to know with absolute certainty what the outcome of the work will be. The extension of the checking program to other types of checks is therefore a reasonable precaution. It is useful to consider different checks as the sources of the *complementary information.*

3.8 Terminology

In this book the term "accuracy" is used to refer to the discrepancy between various estimates and their respective true values. Therefore, an accuracy check aims at providing a measure of accuracy thus defined. A reference to true values is implied. Other types of checks do not provide such a measure. An inspection of the selection procedure used in a survey would end in a qualitative statement such as "correct" or "incorrect." A measure of accuracy is not achieved. We therefore introduce the concept of "quality checks" in order to be able to refer to any type of checking. Accuracy checks thus represent just one possible type of quality check. Similarly, if we make an inspection of the way in which the sample was selected we would say that we are checking the quality of the survey rather than its accuracy.

4. BIASED PROCEDURES

4.1 Definition of biased procedures

In all statistical surveys some procedures are prescribed in order to achieve survey aims. An example is the procedure of selecting the sample, another is the procedure of measuring areas, yet another is the method of estimation, etc. The aim of the procedures used in statistics may be very different: some are intended to obtain data (measuring areas), some are supposed to provide a basis for further stages of work (selection of the sample), some prescribe how the estimates needed are to be obtained from data collected (method of estimation), etc.

Clearly, the aim of each procedure is established upon consideration of the survey as a whole, the circumstances of work, the purposes the survey is supposed to serve, etc. The aim of each particular procedure has to be consistent with the other elements of the survey design. For example, the aim of the estimation procedure should be consistent with the procedure for selection of the sample.

For purposes of discussion, procedures applied in statistics are divided into biased and unbiased procedures. A biased procedure is one whose repeated application results in biased data.[1] It was shown, for example, that in the case of accurate data an estimation procedure is called biased if the expected value of the resulting estimates is not equal to the true value concerned.

The expected value of a procedure can hardly be estimated and procedures can therefore only be qualified as biased upon consideration of how they are formulated and applied, under what circumstances, what experiences and results have been obtained in the past under similar conditions, etc.

---

[1] More details about the difficulties of defining biased procedures are found in Deming, W.E.: *Some theory of sampling*, John Wiley, New York, 1950, pp. 15-16. See also Deming's *Sample design in business research*, John Wiley, New York, 1960 and Uncertainties in statistical data, and their relation to the design and management of statistical surveys and experiments, *Bulletin of the International Statistical Institute*, Vol. 38, Part 4, 1961, pp. 365-383.

A procedure can be biased either because it was *conceived* and *formulated* as such or because its *implementation* was defective. An unbiased procedure may be formulated to select a sample with equal probabilities of the elementary units. If the instructions are not followed properly for some reason the result of the procedure will be a sample with unknown probability structure which may be a source of biases of an unknown magnitude. Through the implementation of the survey plan, unbiased procedures may therefore become biased.

In this chapter three basic groups of biased procedures will be taken into account. They are: (*i*) measurement procedures, (*ii*) procedures of selecting the sample, and (*iii*) estimation procedures. Some other errors and biases which might also be considered as having a procedural character, such as interviewing, are dealt with separately.

4.2 Measurement procedures

The first procedure we wish to discuss is the *pacing*, which is often used in agricultural statistics as a method of area measurement. In order to use the technique the staff are trained to make paces of some standardized length, say 75 cm. The measurement of areas then consists of pacing along the sides of fields and recording the number of paces made along each side.

The quality of the results obtained by means of pacing obviously depends upon how the pacing was carried out. An example of the kind of results that can be obtained by this technique is presented in Table 10. Data in this table were obtained in an experiment conducted jointly by the Uganda Unit of the East African Statistical Department and the Department of Agriculture. The experiment was conducted in two provinces, Nakifuma and Bowa. In each of them a sample of fields was selected and the staff were requested to establish areas under various crops by pacing. The first pacing was carried out at the beginning of the survey, and the second at the end, after the enumerators had finished their work. The purpose of the second pacing was to see how consistent the enumerators were in their work. Finally, the area of the same fields was measured with the help of chains and a compass. Data obtained by means of the latter technique are used as a yardstick to measure the quality of pacing.

In order to make comparison easier, biases obtained by pacing from Table 10 are presented in percentage form in Table 11. Chain and compass measurements are taken as 100.

It will be noticed from Table 11 that the results from the first pacing in Nakifuma are highly biased, thus making the technique useless for any practical purpose unless some improvement can be made in the way it was carried out here. Data from the second pacing are better, although still highly biased. It is not clear in what way the increased experience helped to improve the quality of the data.

It can also be seen from the table that data collected in Bowa are considerably superior. This is assumed to be a result of the difference in the enumerators' training and general knowledge.

TABLE 10. - AREAS UNDER VARIOUS CROPS AS OBTAINED BY THREE DIFFERENT MEASUREMENTS[1]
(Square yards)

Crop	Nakifuma			Bowa		
	Chain and compass	First pacing	Second pacing	Chain and compass	First pacing	Second pacing
Cotton	42 978	56 769	54 190	121 715	125 517	124 133
Coffee	65 198	97 535	78 996	80 422	77 580	76 984
Matoke	70 652	86 543	81 109	127 541	131 345	134 363
Sweet potatoes	5 368	7 321	8 499	12 384	12 488	12 069
Cotton and beans	6 693	9 791	9 024	23 566	26 443	26 364
Cotton and coffee	20 658	25 025	23 435	40 825	41 067	40 699
Matoke and coffee	35 602	39 757	46 278	30 905	30 567	31 480
Matoke and cotton	3 609	5 465	4 956	17 283	19 341	19 248

[1] Data in this table are taken from the publication *Investigation into acreage statistics*, published by the former Uganda Protectorate, East African Statistical Department, Uganda Unit, 1959.

TABLE 11. - PERCENTAGE BIASES IN DATA OBTAINED BY MEANS OF PACING
(Percent)

Crop	Nakifuma		Bowa	
	Error in first pacing	Error in second pacing	Error in first pacing	Error in second pacing
Cotton	32.1	26.1	3.1	2.0
Coffee	49.6	21.2	—3.5	—4.3
Matoke	22.5	14.8	3.0	5.3
Sweet potatoes	36.4	58.3	0.8	—2.5
Cotton and beans	46.3	34.8	12.2	11.9
Cotton and coffee	21.1	13.4	0.6	—0.3
Matoke and coffee	11.7	30.0	—1.1	1.9
Matoke and cotton	51.4	37.3	11.9	11.4

The difference in the quality of data as obtained in Nakifuma and Bowa is the real reason why these data were used for illustration. This difference shows that a varying degree of quality may result from the use of pacing. This makes various precautionary measures necessary if this procedure is to be used. These measures might be: (*i*) adequate training with a check of the success attained, (*ii*) a check of the field work by means of more accurate instruments in order to secure information on the quality of work of the field staff.

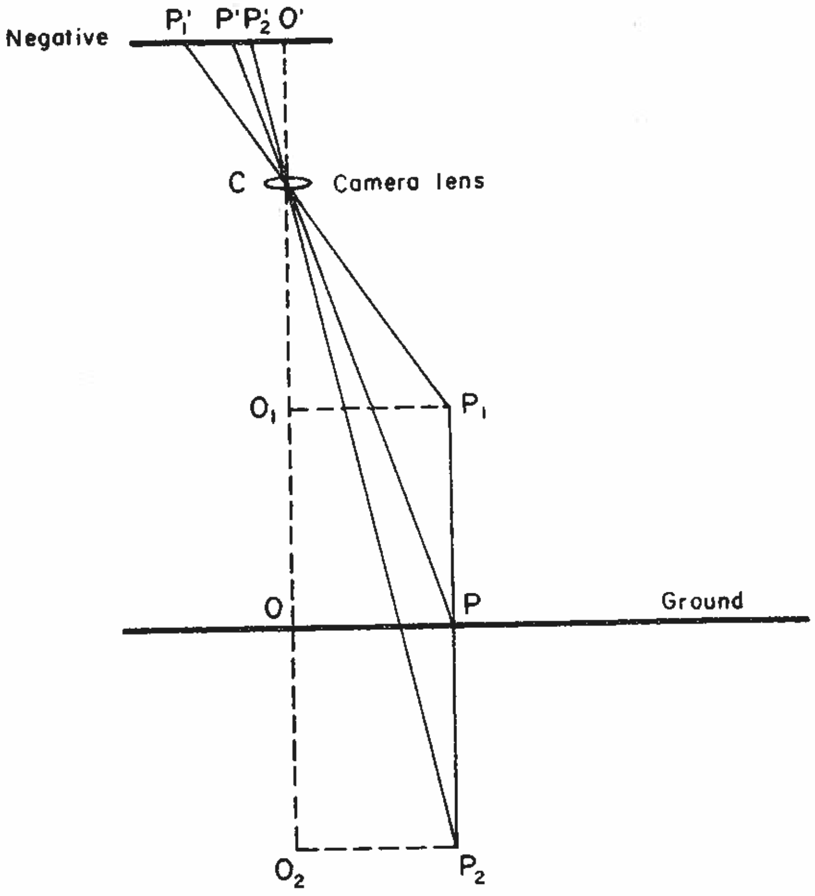
The next example concerns the measurement of areas on *slopes*. If the enumerators are instructed to measure the length of various sides of the fields and disregard slopes, biases may be introduced into the data however accurate the instruments. In agricultural statistics the horizontal projections of areas on slopes are needed. The difference between the actual area on the slope and its horizontal projection will depend upon the slope. Therefore, area measurements may be considerably biased if many of the fields are on hills.

Another procedure which leads to biased results under some circumstances is the measurement of areas on aerial photographs. The procedure will be unbiased if the area photographed is on one level. On the other hand, it will be biased if the topographic relief is not taken into account. The situation will be clear from a study of Figure 4, which shows the displacement of the point on the photograph as a result of its elevation. Point P on ground level is seen as P' on the negative of the photograph. The distance OP becomes $O'P'$ on the photograph. Point P_1, which is raised, is projected as P'_1 on the photograph. As a result, the distance O_1P_1, which is equal to OP, is represented on the photograph as $O'P'_1$, i.e., the distance which is larger than $O'P'$. In other words, all the points which are located higher than ground level and outside the camera axis are displaced on the aerial photograph, radiating outward from the center of the photograph. For the points below ground level the radiation goes inward. It is easy to ascertain that the distance O_2P_2 is represented on the photograph by $O'P'_2$, which is shorter than $O'P'$.

As regards the magnitude of the distortion of the area of fields on aerial photographs Figure 4 shows that

$$P'P'_1 = O'P'_1 \times \frac{PP_1}{CO} \qquad (4.1)$$

FIGURE 4. - DISPLACEMENT OF POINTS ON THE AERIAL PHOTOGRAPH AS A RESULT OF THEIR ELEVATION



In other words, the distortion $P'P'_1$ is directly proportional to the elevation PP_1 and indirectly so to the height of the aircraft above ground level. If we have a circular field photographed at the known elevation PP_1 with the radius O_1P_1 and the center in the camera axis, its factual area will be $\overline{OP}^2\pi$, while the area measured on the photograph will be $\overline{O'P'_1}^2\pi$. In order to adjust the latter according to the distortion resulting from the elevation, $P'P'_1$ can be used from the equation (4.1).

It is for the statistician to decide, upon consideration of the circumstances, how important is it to take into account the bias which may appear here. If hilly areas represent but a small proportion of the total area concerned, it may not be worthwhile to take the trouble of adjusting areas measured on hills. This will also be the case if the elevation of fields on hills is small as compared to the height of the aircraft. The application of the equation (4.1) will show the size of the distortions under given circumstances. It may be very important to take into account such biases in areas if there is some correlation between the height of fields and the crop grown on them. For example, in many countries the hills are covered with vineyards. The area of vineyards as estimated from aerial photographs could then be biased.

Another example of a procedure that might become biased in the course of implementation is the measurement of areas from *cadastral plans* under sharply varying atmospheric conditions. For example, it is known that the length of paper will vary according to moisture. Measurements on a map of the same areas completed at two different times with different degrees of moisture may give different results. Therefore, whenever maps are used for area measurement it must be established how far the biases arising on this basis must be taken into account. In considering this question special attention must be paid to the scale of maps. The smaller the scale the more important the distortions in the size of the paper become.

4.3 Selection procedure

The aim of the selection procedure is to obtain a sample that has a known probability structure. Otherwise no rigorous use of statistical theory can be made. The procedure of selection is normally planned and formulated together with the estimation procedure because they represent two different aspects of the same probability model adopted in the design of the survey. If the selection procedure fails to achieve its aim, the estimation procedure of the type planned may not be worthwhile. The bias resulting from the former will be carried over automatically and the final estimates will be biased unless some means are found to remove biases in the selection of the sample.

An illustration of the consequences of a biased selection of the sample is given in Table 12. Data in this table refer to the 1951 Census of Livestock in Yugoslavia. In each village lists of holdings belonging to the private sector of agriculture were available and copies of them were sent to District Statistical Offices. Using these lists and tables of random numbers, an enumerator in each district was instructed to select a random sample of a given size. The purpose of the sample was to check the quality of data obtained from the selected holdings in the census. Census totals for districts were available for a number of characteristics. In columns 2 and 3 of Table 12 the total number of cattle and poultry is given according to the census. Columns 4 and 6 represent the estimates of the same totals as obtained from the sample selected. Columns 5 and 7 contain sampling errors of the estimated totals in columns 4 and 6.

TABLE 12. - CENSUS DATA AND SAMPLE ESTIMATES OF THE NUMBER OF CATTLE AND POULTRY IN PRIVATE HOLDINGS IN 10 DISTRICTS OF YUGOSLAVIA
(1951 Census of Livestock)

District	Census		Sample estimates			
			Cattle		Poultry	
	Number of cattle	Number of poultry	Number	Percentage error (base: col. 2)	Number	Percentage error (base: col. 3)
(1)	(2)	(3)	(4)	(5)	(6)	(7)
Aleksinacki	7 648	38 435	8 320*	2.0	38 740	3.5
Ariljski	9 702	24 132	10 200	2.3	27 580*	4.2
Azbukovacki	10 516	27 522	11 340*	1.9	30 010	3.1
Belicki	9 924	68 334	11 150*	1.8	71 770*	2.1
Belopalanacki	7 419	35 987	7 360	3.1	35 990	2.9
Beogradski	6 666	43 812	6 580	2.4	44 070	2.0
Boljevacki	10 806	33 508	11 990*	2.3	37 550*	3.9
Jablanicki	25 093	51 915	25 600	1.9	47 480	4.1
Jasenicki	9 762	87 438	10 140	1.6	87 820	2.8
Kosmajski	12 085	70 101	11 980	1.8	75 390*	2.2

A study of these results shows that most of the estimates are larger than the corresponding census results. In the case of cattle four out of ten estimated totals (designated by an asterisk) are larger than census totals for more than three standard errors, while two other estimated totals are larger than corresponding census figures for more than two standard errors. A similar picture is found for poultry. In other words, the resulting sample does not have the desired probability structure. Accordingly, it becomes impossible to apply the estimation procedure adopted in the survey design. However, if it is used irrespective of the biased selection of the sample, which normally remains undetected, the resulting estimates are biased. Table 12 shows the magnitude of the biases that may result.

In this particular case the selection procedure was unbiased but its implementation was faulty.

Another example falling in the same category was reported by Marks.[2] In order to prepare the advanced tabulation of the 1952 Census of Population in Chile it was decided to select a systematic sample of two percent of the households. The clerks responsible for the selection introduced into their work a tendency to avoid large households (probably because they were afraid of a larger volume of work afterward) and selected instead the households immediately preceding or following. No control of the selection procedure was planned and so the bias introduced during the selection passed undetected. Only when data were available from the complete tabulation was it possible to detect the selection bias. The effort, money and time that had been put into the advanced tabulation of census data were lost.

The biases resulting from such an incorrect implementation of a correct procedure are best removed by establishing a strict supervision of the work in all its stages.

Some illustrations will now be presented of selection procedures that were biased. Our first example is taken from Sukhatme.[3] The example refers to the use of sample surveys in estimating the yield of paddy in India by crop-cutting from a sample of plots. For this purpose a sample of villages was first selected. Within the selected villages all the fields under the crop concerned were listed and a sample of them was selected according to their survey numbers, which have to be understood here as cadastral codes for each separate parcel. The instructions issued for the selection were as follows: "Against the name of each selected village are shown three random numbers smaller than the highest survey number in the village. Select the survey numbers corresponding to given random numbers for experiments. If the selected survey number does not grow paddy, select the next highest paddy-growing survey number in its place."

What will be the result of the selection procedure? If the total of survey numbers in a village amounts to 290, the random numbers assigned are 18, 189 and 239, and the fields where paddy is grown are 49 to 88 and 189 to 290, it is obvious that field 49 is selected with the probability of 49/290 and field 189 with a probability (189-88)/290. The remaining paddy fields are selected with equal probability, i.e., 1/290. If the paddy is grown in patches, as is the case here, this procedure is biased in the sense that it gives larger chances of selection to the fields with the smallest survey number in each particular patch. This bias will not affect survey results if there is no difference in yield between the fields with the smallest survey number and the remaining fields. However, if the former are border fields where the crop is more exposed to damage than in the remaining fields, this procedure might lead to negative biases in estimated yield.

This bias could be avoided if field personnel are trained to use tables of random numbers. Instead of using preselected numbers, they would read three digit numbers from the table and reject all those which fall on fields growing crops other than paddy.

Biased procedures of selection are very often introduced when lists of some kind of unit are used for the selection of a sample of a *different type of unit*. For example, lists of dwelling units might be available when a sample of individual persons is needed, or the list of villages in a district when a sample of agricultural holdings is needed. Such uses of frames containing units of one type for the selection of units of a different type are frequent in statistics for purposes of economy. The preparation of lists is always a major expenditure in sample surveys. Therefore, the use of some existing frame is an excellent way of reducing survey costs.

An interesting example of this type of biased selection was discussed by Hansen, Hurwitz and Madow.[4] If one wants to estimate the proportion of school families who own houses, a simple random sample of children might be selected from school lists and an inquiry made whether the corresponding families own houses. The proportion of children from families that own houses would then be the estimate desired.

If the estimation procedure corresponding to simple random sampling is to be applied, the selection procedure is unbiased if each school family has only one child in the school. In this case a simple random sample of children represents at the time a simple random sample of families. However, if some families have two or more children in the school, the selection procedure adopted is confined to the selection of families with

---

[2] Marks, E.S.: Techniques of sampling and statistical evaluation for census work, *Proceedings of the Social Statistics Section, American Statistical Association*, 1961, pp. 49-51.

[3] Sukhatme, P.V.: *Sampling theory of surveys with applications*, The Iowa State College Press, Ames, Iowa, and the Indian Society of Agricultural Statistics, New Delhi, 1954.

[4] Hansen, M.H., Hurwitz, W.N. and Madow, W.G.: *Sample survey methods and theory*, John Wiley, New York, 1953, Vol. I.

probabilities that are proportionate to the number of their schoolchildren. Obviously, the estimation procedure corresponding to simple random sampling cannot be used in this case if an unbiased estimate is desired. Simple random sampling would in fact yield a bias the magnitude of which depends upon the difference with regard to houses owned between families having one schoolchild only and those having several schoolchildren.

The bias arising here can be avoided by a proper weighting of sample results. If x_i is the value of the characteristic under study of the family of the i-th child selected and a_i is the number of children representing that particular family in lists of schoolchildren, an unbiased estimate of the total of the characteristic x is given by

$$X' = \frac{N}{n} \sum_{i}^{n} \frac{x_i}{a_i} \tag{4.2}$$

n and N being the number of children in the sample and in the population respectively. If the survey is aiming at an estimate of the number of families that own houses, x_i is equal to 1 if the family of the i-th child in the sample owns a house. Otherwise, x_i is equal to 0.

Another way of avoiding biases would be to associate family data with one single child, say the oldest. After a sample of n children has been selected in this case, it would be necessary to ascertain whether the selected children are the oldest schoolchildren from the families concerned. If not, they would be disregarded in further stages of the work. In this way each family would have an equal chance of being selected in the sample and the estimation procedure corresponding to simple random sampling would yield unbiased estimates.

Another example of a similar nature is provided by Hansen, Hurwitz, and Jabine.[5] The example refers to the Annual Survey of Manufactures which is conducted by the U.S. Bureau of the Census. The sample used in this survey is selected from the lists of the census of manufactures. " Births," or units which started operating since the last census was taken, are sampled from lists available with the Bureau of Old Age and Survivors' Insurance. In other words, two different types of lists are used

---

[5] Hansen, M.H., Hurwitz, W.N. and Jabine, T.B.: The use of imperfect lists for probability sampling at the U.S. Bureau of the Census, *Bulletin of the International Statistical Institute*, Vol. 40, Part 1, 1964, pp. 497-517.

for the selection of the sample. The procedure, however, leads to many difficulties as the census of manufactures lists the companies while the unit of analysis in the Annual Survey of Manufactures is the individual establishment. This difference of units led to other difficulties. For example, establishments are sometimes transferred from one company to another, companies sometimes merge, etc.

Another interesting case of biased selection was pointed out by Kish.[6] If the frame of households is available and a sample of adult persons is needed, a procedure to get this sample may consist in the selection of households with instructions to enumerate the adult persons available in the selected household at the moment of call. Depending upon circumstances and the time of the call such a procedure may favor some special class of persons. For example, if the aim of the survey is to discover the attitude of the adult population toward some problem and the enumerators call in the morning, the result of the survey will reflect the attitude of housewives.

The bias danger that might appear here is avoided if an objective procedure of selection is introduced which secures adequate chances of selection for each adult member of the sample households. In the procedure developed by Kish, this is achieved in the following way. In the preparation of the frame of households the enumerators will list all the adult members on a special form starting with males in a decreasing order of age and continuing with females. The whole sample of households is then divided into six equal groups or subsamples: A, B, C, D, E and F. These letters are assigned at random to the different groups. The adult persons to be enumerated in each subsample and in each household are then established by looking up their serial number in Table 13. If the table is consulted according to the subsample involved and the number of adults in the particular household, the figure in the body of the table will show the serial number of the person to be interviewed. For example, if a household in subsample A has three adults, the person with the serial number 3 will be selected. In subsample F the first adult listed will be selected. If there is only one adult, then there is no choice. This is why the first column of the table bears 1 in all the rows.

The selection procedure based on Table 13 has some defects. For example, the adults with serial numbers larger than 6, if such households exist,

---

[6] Kish, L.: A procedure for objective respondent selection within the household, *Journal of the American Statistical Association*, Vol. 44, 1949, pp. 380-387.

TABLE 13. - GUIDANCE FOR THE OBJECTIVE SELECTION OF ADULTS WITHIN HOUSEHOLDS[1]

Subsample	Number of adults					
	1	2	3	4	5	6 or more
A	1	1	3	2	5	1
B	1	2	1	3	4	2
C	1	1	2	4	1	3
D	1	2	3	1	2	4
E	1	1	2	1	3	5
F	1	2	1	4	3	6

[1] Kish, L.: A procedure for objective respondent selection within the household, *Journal of the American Statistical Association*, Vol. 44, 1949, pp. 380-387.

have no chance of selection. Another problem is connected with the fact that all the households are selected with equal probability, which means that an adult from a household with two adults has half as much chance of selection as an adult from a household with one adult, etc.

Various improvements of this technique were proposed by Thionet.[7]

Further details on this problem are found in papers by Ferber[8] and Neter and Waksberg.[9]

In a number of other cases the selection procedure adopted becomes biased because of the failure to take adequately into account some *specific features of the population* that is being sampled. For example, a sample of streets might be selected and instructions given to subsample all the households residing in houses numbered 5, 15, 25, etc.[7] Such instructions might be issued on the understanding that the second-stage units will be selected with equal probabilities. In fact, this aim would be achieved if the random start of the second-stage selection is changed from one primary unit to the other. In addition, there is the assumption that houses are numbered consecutively with no houses having the same number. In practice, however, it very often happens that several houses have the

---

[7] Thionet, P.: *Application des méthodes de sondages aux enquêtes statistiques*, Institut national de la statistique et des études économiques, Paris, 1953.

[8] Ferber, R.: On the reliability of responses secured in sample surveys, *Journal of the American Statistical Association*, Vol. 50, 1955, pp. 788-810.

[9] Neter, J. and Waksberg, J.: Effects of interviewing designated respondents in a household survey of home owners' expenditures on alterations and repairs, *Applied Statistics*, Vol. 12, 1963, pp. 46-60.

same number, say 25, 25a, 25b and 25c, while for some other numbers no houses might exist (the house having being demolished, the number being reserved for a house to be built, etc.). If several houses have the same number, the proposed selection procedure will not yield 10 percent of second-stage units. The houses with numbers a, b, c, etc. have no chance at all of being selected in the sample.

The biases resulting from such a situation can easily be avoided by instructing the enumerators to include in the sample all houses having the same basic number. In other words, if the number 25 is selected, houses bearing numbers 25a, 25b, 25c, etc. are also considered as selected and for survey purposes have to be considered as one single house.

This solution may not be satisfactory from the efficiency point of view because a group of houses considered as one single house decreases the precision of estimates. Therefore, a more efficient alternative will be to instruct enumerators to prepare a fresh list of houses as the first part of their field duties. The selection will then be made by using this list as a frame. This will provide for several houses with the same number as well as for numbers without corresponding houses.

Biased selection may easily take place in sampling with *probabilities proportionate to some measurement of size*. For example, if past census data are used to select some area units with probabilities proportionate to their respective areas under potatoes as stated in the census, the area units which did not grow potatoes at the time of the census will have no chance of being selected. This procedure does not introduce any bias if areas assigned zero probabilities do not grow any potatoes during the survey. On the other hand, if they do so, the survey estimates will be biased because these areas will not be accounted for.

A way of avoiding such biases is to attach the units that did not grow potatoes during the last census to the neighboring units on the list and consider each of these combinations as one single unit. If one of these enlarged units is selected in the sample, the field work will of course be extended over the whole area combined.

Another type of biased selection may take place in *point sampling*. If a cadastral map is available showing individual parcel boundaries, a random sample of points with co-ordinates lying within the area covered by the map will give a random sample of parcels selected with probabilities proportionate to size. If applied in this way and for the purpose of the selection of parcels, the procedure is unbiased. However, a similar procedure is used in some cases to select a sample of agricultural holdings or some other units. The points falling within an area are selected at random and then the holdings located nearest to the points selected are included in the sample. This procedure is clearly biased. The isolated holdings located in a rather large area have a higher chance of selection than others grouped in a smaller area.

Biased selection is often the result of an improper use of the *table of random numbers*. If a simple random sample of agricultural holdings has to be selected from lists containing 275 holdings, the unbiased procedure consists in obtaining the corresponding serial numbers of holdings from a table of random numbers by using three-digit columns and rejecting all the numbers larger than 275 with 000 disregarded.

In some cases the result may be a too high rejection of numbers. It may therefore be decided to use any three-digit number by dividing all the numbers larger than 275 by 275 and taking the remainder as the serial number of the unit selected. If the number read is 578, this gives $578 = 2 \times 275 + 28$. The result indicates that the holding listed under serial number 28 is selected. The remainder 0, of course, means the selection of the unit 275.

Such a selection procedure is biased because it does not give each unit of the population an equal chance of selection. If the number 826 is read from the table, this means that the population unit with the serial number 1 is selected in the sample ($826 = 3 \times 275 + 1$). In other words, this unit is selected whenever any of the following numbers are read from the table of random numbers: 001, 276, 551 and 826. It is easy to ascertain that this chance of selection is given to the 174 units listed first, while the remaining ones, i.e., those listed from 175 through 275, have only three chances of being selected.

If the procedure is to be made unbiased in the sense that all the units are given the same chance of selection, all the numbers larger than 825, including 000, have to be rejected. In other words, the above adjustment of the selection procedure does not completely eliminate the rejections.

Biased selection is sometimes encountered in *systematic sampling*. For example, if a sample of 1 percent of units is needed, the procedure might be to select 10 percent of primary units and apply a fixed random start in the selected primary units and select each tenth unit by starting the counting in each primary unit from 1. The selected units would then be 1, 11, 21, etc. The resulting sample will contain more than one percent of units. The right procedure would be to vary the random start in each primary unit.

An illustration was reported by R.T. Smith.[10] In order to establish railroad statistics it was decided to select at each station a sample of waybills ending in 01. The resulting size of the sample will be biased, the magnitude of the bias depending upon the number of stations and their number of waybills.

Another case of this type was pointed out and analyzed by Yates.[11] In the 1951 Census of Population in Great Britain a sample of one percent of households was selected. For this purpose the enumeration districts were numbered serially and all households with a serial number ending in 25 were selected in odd enumeration districts and those having serial numbers ending in 76 in even districts. Theoretically speaking the procedure is biased as it does not yield one percent of households.[12]

4.4 Control of selection biases

Biased procedures lead to biased data. This is the reason for controlling the selection procedure. In this respect two groups of measures are possible. In the first group are the measures intended to prevent selection biases appearing in data. These are steps taken prior to the actual selection of the sample. In the other group are measures taken after the sample has already been selected and sometimes after the data have been collected.

Among the measures intended to prevent selection biases two steps are useful in most cases. The first is a thorough analysis of the planned procedure with the aim of ascertaining whether, theoretically, it is fully adequate. For example, instructions may be issued to list all the villages in each district in the sample and select in each district all the villages with a serial number ending in 1, i.e., 1, 11, 21, etc. The intention may be to have a simple procedure for selecting a sample of 10 percent of villages. A theoretical analysis of such a procedure will show that it does not yield 10 percent of villages. As a result a change in procedure will be needed if the aim of a 10 percent sample of villages is to be respected.

---

[10] Smith, R.T.: Controlling quality in railroad traffic statistics, *Improving the quality of statistical surveys*, American Statistical Association, 1956, pp. 65-74.

[11] Yates, F.: *Sampling methods for censuses and surveys*, Third edition, Charles Griffin, London, 1960.

[12] Further details on biased selection procedures are to be found in Zarkovich, S.S.: *Sampling methods and censuses*, FAO, 1965.

If properly applied, this is the measure that removes the use of biased procedures.

The next step in the same group is the breakdown of the selection procedure into its basic constituent phases and the organization of a systematic supervision of the operation in each separate phase. With such a detailed program of supervision the staff will find it difficult to deviate from instructions. This will be an important protection against selection biases.

This measure is useful for preventing a correct selection procedure from producing a biased sample through faulty implementation.

It will be remembered from the last section that such a measure would have prevented selection biases from appearing in the results of the 1952 Census of Population in Chile.[13]

Problems become more difficult after the selection has been made and the data have been collected. This is the situation which statisticians sometimes have to face when evaluating the data of a completed survey.

In this case a theoretical study of the procedure applied will also be useful. It will show whether it was correct or not. In the second case it must be seen whether the effect of the selection bias can be estimated and some steps taken to remove the bias. If the procedure is correct it must be seen whether the implementation was also correct. Unfortunately, little can be done in this respect.

Serious possibilities for checking the presence of selection biases in sample data exist only in cases where the sample is selected from a known population. In this situation a test of differences between sample estimates and the corresponding population figures shows whether the differences observed are due to chance variations or to the presence of some biases.

This is how the selection biases were detected in data presented in Table 12. The selection procedure was formulated correctly; instructions on how to carry out the selection in the field were also correct. Since biases were present in the data it was decided to study how the work was carried out in the field. The results found showed that biases appeared because of a number of difficulties in the field work which were not properly provided for. First of all, there was an identification difficulty. Several holders had the same name and it was not always clear to whom the selection referred. This gave the enumerators the chance to follow the easiest course and to replace sample units scattered all over the area of assignment by holdings which were closer and in general richer. After the violation of the selection procedure had started it was difficult to stop it. The weather was cold, the road muddy, and the temptation extremely strong to replace a part of the sample with conveniently located units. On a number of occasions the enumerators could not resist this temptation. Table 12 shows the rest of the story.

If the presence of the selection bias has been ascertained, the question will arise of how to remove the effects of this bias. Two procedures may be involved here. The first is the computational adjustment of sample data and the other is the adjustment of the composition of the sample by adding or taking away some units. If the selection is biased in the sense that some classes of the population are underrepresented and others overrepresented, the computational adjustment is achieved if sample data are weighted by means of the known class frequencies from some other source. The adjustment of the structure of the sample is carried out by changing at random the number of units in various classes of the sample with the aim of obtaining the same frequency distribution as the one in the census.

Details of the procedures involved here are discussed in the author's book *Sampling methods and censuses.*

Special cases of selection biases are dealt with by methods governed by the particular circumstances. Under some circumstances efficient and refined techniques may be applicable. An illustration from laboratory experimental work is to be found in a paper by Blackwell and Hodges.[14] Eklund, in his *Studies of selection bias in applied statistics*,[15] also gives many examples of the adjustment of data from a biased sample. The techniques used by Eklund assume the availability of information about the units involved so that models can be constructed which permit conclusions about the missing part of the population. Eklund's ideas are primarily applicable in research work where sufficient knowledge is available to build up models.

---

[13] Marks, E.S.: Techniques of sampling and statistical evaluation for census work, *Proceedings of the Social Statistics Section, American Statistical Association*, 1961, pp. 49-51.

[14] Blackwell, D. and Hodges, J.L.: Design for the control of selection bias, *The Annals of Mathematical Statistics*, Vol. 28, 1957, pp. 449-460.

[15] Published by Almqvist and Wiksell, Uppsala, Second edition, 1960.

Finally, it may be added that the problem of selection biases is often neglected in survey practice in spite of its great importance. If the selection is not correct, the effort and money spent on a survey are frequently wasted.

4.5 Biased estimation procedure

Biased and unbiased estimation procedures were defined in section 1.5. In section 1.6 general relationships between biased and unbiased estimates were discussed as well as the advantages of using unbiased estimates. It was also shown that the use of biased estimation procedures is of interest as long as these procedures lead to a smaller mean square error than the corresponding unbiased estimates. In other words, interest in biased estimates is built up on efficiency grounds.

The concept of efficiency involved here is to be understood in the broadest sense. It refers primarily to the mean square error. For example, the ratio method of estimation may be used because it has a lower mean square error than some alternative unbiased estimates. In other cases the important factor is the convenience of the methods used, the simplicity of instructions, the cost of computation, etc. For example, computations of variances may be replaced by some approximate method based on the use of data from past surveys, known relationships from work in other countries, etc. In other words, the justification of the procedures used lies in increased speed, more convenient work, etc.

Whenever a biased procedure is used it is necessary to know the mathematical expressions for the magnitude of the bias so that it can be established whether the bias should be a cause for concern, on what quantities it primarily depends and what steps may be useful to reduce or to eliminate it altogether. For example, equation (1.10) shows the bias per elementary unit in the unweighted arithmetic mean of cluster means. This formula shows that the magnitude of the bias depends upon the variation between cluster means per elementary unit and the variation in the size of clusters in terms of the number of elementary units. If such variations do not exist, there is no cause for concern. If they do exist, however, the next measure would be to consider variances of the two alternative estimates. The relative magnitude of variances will indicate the next step to be taken. If the variance of the unweighted mean is considerably smaller, an estimate of the bias will be needed in order to estimate the mean square error which is to be compared with the variance of the weighted mean. This, coupled with a knowledge of survey aims and uses of data, will show how to deal with the bias in this particular case.

In the case of other biased estimates a variety of steps may be possible. In order to provide an illustration in the rest of this section we shall concentrate on the ratio method of estimation. Recent contributions to the theory of this method are very instructive.

The first work in this connection concerns the magnitude of biases of the ratio method. The aim of this work was to provide an evaluation of the bias of the ratio method of estimation at different degrees of approximation in order to provide guidance in situations where the bias might arise. Typical work along these lines was done by Sukhatme,[16] Cochran,[17] Koop,[18] Echimovich,[19] etc. These studies show the conditions under which the ratio method can be considered as the best linear unbiased estimate.[20] If the sample is stratified, Hansen, Hurwitz and Gurney[21] have shown that two different forms of the ratio method are available. One of them consists of separate estimates for each stratum and the other is the combined ratio estimate. The efficiency of the approaches will depend upon the position of the regression line in different strata.

Another useful approach might consist in a study of the empirical evidence with the aim of establishing the magnitude of the biases due to ratio estimates in past surveys. This line was taken in a paper by Kish, Namboodiri and Pillai.[22] They analyzed surveys conducted by the Survey Research Center of the University of Michigan and discovered that the ratio of the bias to standard errors "averages about one percent and seldom appears greater than four percent."

---

[16] Sukhatme, P.V.: *Sampling theory of surveys with applications*, The Iowa State College Press, Ames, and the Indian Society of Agricultural Statistics, New Delhi, 1954.

[17] Cochran, W.G.: The estimation of the yields of cereal experiments by sampling for the ratio of grain to total produce, *Journal of Agricultural Science*, Vol. 30, 1940, pp. 262-275.

[18] Koop, J.C.: A note on the bias of the ratio estimate, *Bulletin of the International Statistical Institute*, Vol. 33, Part 2, 1951, pp. 141-146.

[19] Echimovich, J.: Upotreba razvoja u Taylorov red u statističkoj teoriji uzoraka, *Statistička Revija*, Vol. 10, 1960, pp. 35-43.

[20] David, F.N. and Neyman, J.: Extension of the Markoff theorem of least squares, *Statistical Research Memoirs*, Vol. 2, 1938, pp. 105-116.

[21] Hansen, M.H., Hurwitz, W.N. and Gurney, M.: Problems and methods of the sample survey of business, *Journal of the American Statistical Association*, Vol. 41, 1946, pp. 173-189.

[22] Kish, L., Namboodiri, N.K. and Pillai, R.K.: The ratio bias in surveys, *Journal of the American Statistical Association*, Vol. 57, 1962, pp. 863-876.

In spite of this empirical evidence the theoretical possibility remains that the bias of a ratio estimate assumes a magnitude that cannot be disregarded. The advisability of removing the bias or developing unbiased ratio estimates is therefore obvious.

In this connection it might be useful to point out the technique developed by Quenouille.[23] This technique makes it possible to reduce the magnitude of the bias from $0(n^{-1})$ to $0(n^{-2})$. The essence of the technique consists in the following: if data from a sample of size n are used to compute the estimate u_1 which has the bias $cn^{-1} + 0(n^{-2})$, the total size of the sample is split into two halves and each of these is used to compute the two new estimates u_2 and u_3 based on $n/2$ values each. Afterward u_2 and u_3 are used to compute the estimate

$$u = 2u_1 - 0.5\ (u_2 + u_3) \tag{4.3}$$

The estimate u is subject to a bias of the order n^{-2}.

It is worth pointing out that some survey designs offer an immediate possibility for the application of Quenouille's technique. Such is the case of Deming's replicated samples and Mahalanobis's interpenetrating samples. In fact, Deming provided illustrations of Quenouille's method and suggested its extension to cases where a larger number of u estimates might be available.[24]

Because of the division of the sample the estimator u in (4.3) can be expected to have a larger variance than a corresponding estimator based on all the n observations. Accordingly, it might be considered that in the value of the mean square error the bias term was eliminated at the expense of an increase in the variance. This problem was studied by Durbin[25] who discovered that, instead of the larger variance, the estimator u might lead to a smaller variance in some cases. Such a case appears under circumstances where the relationship between x and y is linear and x is normally distributed.

---

[23] Quenouille, M.H.: Notes on bias in estimation, *Biometrika*, Vol. 43, 1956, pp. 353-360.

[24] Cf. Deming, W.E.: *Sample design in business research*, John Wiley, New York, 1960, and Jones, H.L.: Investigating the properties of a sample means by employing random subsample means, *Journal of the American Statistical Association*, Vol. 51, 1956, pp. 54-83.

[25] Durbin, J.: A note on the application of Quenouille's method of bias reduction to the estimation of ratios, *Biometrika*, Vol. 46, 1959, pp. 477-480.

Another approach consists in defining and adopting a selection procedure that eliminates the bias in the estimated ratio. Such an idea was proposed by Lahiri.[26] According to Lahiri's technique an unbiased estimate of the ratio is obtained by selecting the sample of n units with probability proportional to the aggregate value of the auxiliary variable in the sample, i.e., $\overset{n}{\Sigma} y_i$. In order to have such a sample, n units are first selected out of a population of N units. Using the units in the sample the aggregate Σy_i is then computed. The quantity G, which is equal to the largest possible aggregate of n units, is also established. This is obtained by summing up n largest y values in the frame. Afterward a random number is selected which is not larger than G. If this random number does not exceed Σy_i as computed from the sample, the sample is retained for the survey. If it is larger than the available sample aggregate, the sample selected is rejected and the whole operation repeated until the sample is selected.

The Lahiri procedure leads to unbiased ratio estimates. Another advantage of the technique lies in the fact that it eliminates the necessity for cumulative totals. However, with such a selection of the sample there will be a fairly large proportion of rejections. Another difficulty is that the technique cannot be used unless the auxiliary variable is known for all the units of the population.

Midzuno[27] has considered a similar approach. As regards the procedure of selection, Midzuno used the selection of the first unit with probability proportionate to the value of the auxiliary variable while the remaining $(n - 1)$ units are selected with equal probability without replacement.

For cases where the sample is selected with probability proportionate to the aggregate value of the auxiliary variable, Des Raj[28] has developed the unbiased ratio estimators and their variances in the case of simple random samples, stratified sampling, multistage designs, and double sampling in the case of the estimation of proportions.

Another line of research concentrates on the idea of developing an unbiased estimator of the ratio without selecting the sample with proba-

---

[26] Lahiri, D.B.: A method of sample selection providing unbiased ratio estimates, *Bulletin of the International Statistical Institute*, Vol. 33, Part 2, 1951, pp. 133-140.

[27] Midzuno, H.: On the sampling system with probability proportionate to sum of sizes, *Annals of the Institute of Statistical Mathematics*, Vol. 3, 1952, pp. 99-107; An outline of the theory of sampling systems, *Annals of the Institute of Statistical Mathematics*, Vol. 1, 1950, pp. 149-156.

[28] Des Raj: Ratio estimation in sampling with equal and unequal probabilities, *Journal of the Indian Society of Agricultural Statistics*, Vol. 6, 1954, pp. 127-138.

bilities proportionate to the aggregate of the auxiliary variable. A typical example of such an approach is found in a note by Hartley and Ross.[29] If a sample of n units is selected (with equal probabilities), two different ratio estimators are possible, i.e., $\bar{x}'_r = \frac{\bar{x}}{\bar{y}} \bar{Y}$ and $\bar{x}''_r = \bar{r}\bar{Y}$. The biases of these estimates are

$$E\bar{x}'_r = \bar{X} - \operatorname{Cov}\left(\frac{\bar{x}}{\bar{y}}, \bar{y}\right) \tag{4.4}$$

and

$$E\bar{x}''_r = \bar{X} - \operatorname{Cov}\left(\frac{x}{y}, y\right) \tag{4.5}$$

Now a new ratio estimator can be defined as

$$\bar{x}'''_r = \bar{r}\bar{Y} + \frac{n}{n-1}\,\frac{N-1}{N}\,(\bar{x} - \bar{r}\,\bar{y}) \tag{4.6}$$

with

$$\sigma^2_{\bar{x}'''_r} = \frac{\bar{X}^2}{n}\,(V_x^2 + V_y^2 - 2V_{xy}) \tag{4.7}$$

In the equation (4.6) the bias is removed.

The same problem was taken up later by Goodman and Hartley.[30] The authors compared the variances of the three ratio estimates given above and discovered that the last one is often more efficient from the variance point of view. An unbiased estimate of the variance of the estimator (4.6) was developed by Robson[31] and Olkin.[32]

---

[29] Hartley, H.O. and Ross, A.: Unbiased ratio estimators, *Nature*, Vol. 174, 1954, pp. 270-271.

[30] Goodman, L.A. and Hartley, H.O.: The precision of unbiased ratio-type estimators, *Journal of the American Statistical Association*, Vol. 53, 1958, pp. 491-508.

[31] Robson, D.S.: Applications of multivariate polykeys to the theory of unbiased ratio-type estimation, *Journal of the American Statistical Association*, Vol. 52, 1957, pp. 511-522.

[32] Olkin, I.: Multivariate ratio estimation for finite populations, *Biometrika*, Vol. 45, 1958, pp. 154-165.

The problem of the unbiased ratio estimates was taken up by many other authors, such as Mickey,[33] and de Pascual[34] who developed a "combined" unbiased ratio estimate in the case of sampling from a stratified population. The sampling procedure consists in selecting one unit from each of the L strata and computing $\bar{y}$ and $\bar{x}$ from the sample of size L. The whole sample is then replaced and the procedure is repeated k times, leading to the total sample size kL. These independent samples make it possible to estimate the bias and define an unbiased ratio estimator.

An unbiased ratio-type estimator was also elaborated by B.V. Sukhatme[35] for use in double sampling. Sukhatme's estimator is of the same type as the one developed by Hartley and Ross.

Other contributions dealing with various aspects of this problem have been made by Murthy and Nanjamma,[36] Nanjamma, Murthy and Sethi,[37] Murthy,[38] Konijn,[39] Robson and Vithayasai,[40] Koop,[41] Thionet,[42] Hajek,[43] etc.

The studies reported here are an illustration of the measures that are being developed with the aim of removing biases from the estimation procedure. Work along these lines is a part of the effort that modern statistics is making to improve the quality of data. Awareness of these measures and their consideration at a proper stage of the survey work might in some cases be very useful.

---

[33] Mickey, M.R.: Some finite population unbiased ratio and regression estimators, *Journal of the American Statistical Association*, Vol. 54, 1959, pp. 594-612.

[34] de Pascual, J.N.: Unbiased ratio estimators in stratified sampling, *Journal of the American Statistical Association*, Vol. 56, 1961, pp. 70-87.

[35] Sukhatme, B.V.: Some ratio-type estimators in two-phase sampling, *Journal of the American Statistical Association*, Vol. 57, 1962, pp. 628-632.

[36] Murthy, M.N. and Nanjamma, N.S.: Almost unbiased ratio estimates based on interpenetrating subsample estimates, *Sankhyā*, Vol. 21, 1959, pp. 381-392.

[37] Nanjamma, N.S., Murthy, M.N. and Sethi, V.K.: Some sampling systems providing unbiased ratio estimators, *Sankhyā*, Vol. 21, 1959, pp. 299-314.

[38] Murthy, M.N.: Ordered and unordered estimators in sampling without replacement, *Sankhyā*, Vol. 18, 1957, pp. 379-390.

[39] Konijn, H.S.: Regression analysis in sample surveys, *Journal of the American Statistical Association*, Vol. 57, 1962, pp. 590-606.

[40] Robson, D.S. and Vithayasai, C.: Unbiased componentwise ratio estimation, *Journal of the American Statistical Association*, Vol. 56, 1961, pp. 350-358.

[41] Koop, J.C.: On upper limits to the difference in bias between two ratio estimates, *Metrika*, Vol. 5, 1962, pp. 145-149.

[42] Thionet, P.: Développements récents de la théorie des sondages, *Journal de la Société de Statistique de Paris*, Vol. 100, 1959, pp. 279-296.

[43] Hajek, J.: On the theory of ratio estimates, *Bulletin of the International Statistical Institute*, Vol. 37, Part 2, 1960, pp. 219-226; Some contributions to the theory of probability sampling, *Bulletin of the International Statistical Institute*, Vol. 36, Part 3, 1958, pp. 127-134.

5. BIASED TOOLS

5.1 General remarks

In statistical work various tools are used for measuring distances, weight, volume, etc. In addition to these general tools, which are also used for many purposes outside statistics, there are specific statistical tools, such as tables of random numbers, questionnaires, instructions, etc. Clearly, all the tools can be biased. The use of biased tools introduces various errors in the data collected, and the consequences are in many ways the same as the consequences of errors originating from other sources. Errors in areas of individual fields as measured by means of inaccurate chains might lead to biased estimates of total areas in the same way as errors in eye estimates of the same areas.

There is a very important characteristic of errors resulting from the use of various tools. Namely, these errors tend to have a *systematic* character. If areas of fields are measured with an inaccurate measuring tape or chain, a corresponding error will appear in the areas of the individual fields, thus leading to a more or less considerable bias in the estimated total. The case of questionnaires is very similar. If a question is not properly worded and gives rise to faulty interpretation, the respondents will give the same type of inaccurate answer and the resulting bias may assume considerable magnitude.

The systematic character of errors having their source in biased tools makes it necessary to check the quality of the tools to be used and ensure that they are as accurate as is reasonably possible.

This should be a matter of routine particularly in the case of cheap tools whose characteristics easily change with use or variations in weather conditions. For example, string is often used as a cheap material for measuring distances. Before long tension will stretch it, and its length will also be affected to some extent by rain. Similarly, the spring on balances often used for weighing crops will lose some of its resilience in time, and this may also cause some bias.

For reasons of this nature it is advisable to use the best tools possible. However, this statement needs to be qualified. Better quality of tools increases the cost of the survey, and so the problem is that of striking a balance between the increased cost and the improved quality of the information collected. The subsequent discussion may be instrumental in achieving this balance.

5.2 Random numbers

Tables of random numbers are an important tool in statistical work. The use of tables of random numbers secures an objective procedure in the selection of the sample and eliminates the risk of various biases resulting from the exercise of personal judgment.

Several tables of random numbers are available at present. They are: *Random sampling numbers*, by L.H.C. Tippett, Tracts for Computers No. 15, 1927, Cambridge University Press; *Tables of random sampling numbers*, by M.G. Kendall and B. Babington Smith, Tracts for Computers No. 24, 1939, Cambridge University Press; *Statistical tables for biological, agricultural and medical research*, by R.A. Fisher and F. Yates, Oliver and Boyd, Edinburgh, Third edition, 1948 (contains tables of random numbers in addition to other tables); *A million random digits*, The Rand Corporation, The Free Press, Glencoe, 1955, etc.

Tables of random numbers are produced in different ways. Tippett compiled his tables by taking digits from census returns. Kendall and Babington Smith constructed a special randomizing machine and recorded the digits which appeared by rotating the disk of the machine bearing numbers 0 through 9.

Irrespective of the means used in preparing tables of random numbers, no means are a priori safe and tables need to be tested for randomness. The well-known experiences with dice throwing and roulette tables offer sufficient warning that apart from the possibility of falsification apparently objective procedures may easily result in a sequence of digits which is not random. For this reason the standard tables of random numbers have been tested and are safe for most uses in survey practice. There are, however, some cases in the use of tables of random numbers where biases in the selection of the sample may easily creep in. This is why some consideration of the randomness of tables of random numbers may be useful.

A number of different tests of randomness have been applied in testing tables of random numbers. Kendall and Babington Smith were among the first to consider this question and formulate adequate tests.[1] The following quotation from their paper will show the nature of the tests they proposed:

(*a*) "The first and most obvious test to be applied is that all the digits shall occur an approximately equal number of times. This test we call the *frequency test*;

(*b*) "Secondly, if the series is... random, no digit shall tend to be followed by any other digit. If therefore we form a bivariate table showing the distribution of pairs of digits in the series, arranged in the rows according to the first digit, and in the columns according to the second digit, we should get frequencies which are approximately equal in all the cells. This test we refer to as the *serial test*;

(*c*) "Thirdly, if the digits are arranged in blocks of, say, five, there will be certain expectation of the numbers in which the five digits are all the same, the numbers in which there are four of one kind, and so on. This test we refer to as the *poker test*, from an analogy with the card game;

(*d*) "Finally, there are certain expectations in regard to the gaps occurring between the same digits in the series. For instance, if we take one digit, say, zero, in about one tenth of the cases, the first zero will be followed immediately by a second zero, and there will be no gap. In about nine hundredths of the cases there will be one digit between two zeros. In about eighty-one thousandths of the cases there will be a gap of two digits between successive zeros, and so on. This we call the *gap test*."

Four tests are proposed here because they are more powerful as a combination. In addition, nonrandomness of some series of digits could pass undetected if only one of these tests were applied. For example, the series produced by repeating numbers 0, 1, 2, ...9 would satisfy the frequency test. Some other nonrandom series could also be constructed that would satisfactorily pass the serial test.

Udny Yule[2] applied such a testing procedure to Tippett's random sampling numbers and came to the conclusion that the table was "patchy." In point of fact, if a table of random numbers is to be satisfactorily used it has to satisfy the criteria of randomness as a whole as well as in parts. In other words, parts of the table, which are often used for the selection of the sample, should also have the characteristics of a random series of numbers if the sample, based on these parts, is to be a random sample. In this respect tests carried out by Yule show that there are patches in Tippett's tables that do not completely satisfy the needs of survey work. The procedure often applied in the selection of the sample consists of reproducing a part of the table of random numbers, which is used later in the field for drawing the sample either from one single stratum or from each stratum separately. If the selected part of the table is not random in that it shows, say, a prevalence of small numbers, the resulting sample will be biased in the case of correlation between the serial number of the unit and the value of the characteristic under study. In practical work it is impossible to know whether some such correlation exists. Therefore, the use of a table of random numbers which satisfies criteria of randomness in the whole as well as in its parts is a useful precaution against biased samples. It is also worth noticing that some tables were specially tested for randomness of parts. For example, Kendall and Babington Smith's tables present 100,000 random numbers in groups of 1,000 numbers each. Each of these groups was tested separately and whenever one of them failed to pass the test this was indicated in the text for the user's information.

It is sometimes considered that a rather short table of random numbers is sufficient for most practical purposes. The underlying assumption is that it can be used to read out numbers horizontally and after the table is exhausted, vertically, then diagonally, then in these same directions but starting from the bottom, etc. This is the recommendation made by Karl Pearson in the introduction to Tippett's random numbers. The recommendation is obviously based on the intuitive feeling that a random sequence of digits, by its very definition, should always have this characteristic irrespective of the place where the reading out of the numbers starts and the direction it takes.

In the light of what has been said above, this repeated use of the same part of a table of random numbers might lead to a repetition of biases in the selection of the sample (in each stratum). If parts of the tables have to be used, it is therefore advisable to have a different part for each stratum. For this reason a good table of random numbers should be sufficiently large in terms of the number of digits it contains. A large

---

[1] Kendall, M.G. and Babington Smith, B.: Randomness and random sampling numbers, *Journal of the Royal Statistical Society*, Vol. 101, 1938, pp. 147-166.

[2] Yule, G.U.: A test of Tippett's random sampling numbers, *Journal of the Royal Statistical Society*, Vol. 101, 1938, pp. 167-172.

table makes it possible to avoid using the same set of numbers. Although it does not necessarily result in biased samples, there are theoretical reasons against the repetition of the numbers already used. Kendall and Babington Smith[3] have shown that a set of random numbers obtained by repeating the same table cannot pass the test of randomness if the repetition reaches a certain number. In some situations such repetitions might result in correlated samples where basically independent samples are desired. For this reason there has been a tendency to increase the size of tables of random numbers. Tippett's tables contain 40,000 digits, Kendall and Babington Smith's tables consist of 100,000 digits, while those produced by the Rand Corporation consist of one million digits. Large tables make it possible in most cases to select the necessary sample by reading the numbers in one direction only.

It could also be pointed out that the choice of direction in reading random numbers may not be immaterial. If the table were found to be random in one direction this does not necessarily mean that it would pass tests if checked for randomness in another direction. In this connection Kendall and Babington Smith advise that their tables should be read horizontally because this direction was used in testing the table.

Recent developments in the use of electronic computers have added some new aspects to the problem of the randomness of random numbers. Namely, computers are sometimes used to generate random numbers needed for various purposes and the question arises of how to secure and test the randomness of these numbers, etc. The problems that appear in this connection do not belong to survey techniques. Further discussion of them will be found in Bofinger and Bofinger,[4] Coveyou,[5] Johnson,[6] Peach,[7] Rotenberg,[8] Todd and Todd,[9] and the other literature quoted therein.

---

[3] Op. cit.

[4] Bofinger, E. and Bofinger, V.I.: A periodic property of pseudo-random sequences, *Journal of the Association for Computing Machinery*, Vol. 5, 1958, pp. 261-265.

[5] Coveyou, R.R.: Serial correlation in the generation of pseudo-random numbers, *Journal of the Association for Computing Machinery*, Vol. 7, 1960, pp. 72-74.

[6] Johnson, D.L.: Generating and testing pseudo-random numbers on the IBM type 701, *Mathematical tables and other aids to computation*, Vol. 10, 1956, pp. 8-13.

[7] Peach, P.: Bias in pseudo-random numbers, *Journal of the American Statistical Association*, Vol. 56, 1961, pp. 610-618.

[8] Rotenberg, A.: A new pseudo-random number generator, *Journal of the Association for Computing Machinery*, Vol. 7, 1960, pp. 75-77.

[9] Todd, J. and Todd, O.T.: Generation of pseudo-random numbers. *Symposium on Monte Carlo methods*, John Wiley, New York, 1956, pp. 15-28.

5.3 Questionnaires

The questionnaire can be considered as one of the most important tools in statistical work. Firstly, it is very widely used. Any statistical survey, if it represents more than a small-scale classroom or laboratory experiment, is based on some type of questionnaire. The success of a survey largely depends upon the quality of the questionnaire. For this reason great stress must be laid on the importance of studying all the aspects and details of the questionnaire before it is used in a survey.

The questionnaire is a means of informing respondents about the type of data needed in a survey, as well as on how they should be presented. The questionnaire penetrates all layers of the population and communicates the survey program to people with different levels of education. Accordingly the questionnaire will of necessity provoke a large variety of attitudes, opinions and interpretations that may often deviate from what was intended. The survey's aim, however, is an identical interpretation of the meaning of the information asked for. This ideal cannot be achieved in an absolute way, either because of the statisticians' failure to communicate what they want in an unequivocal manner or because there are no means of preventing differences appearing among people when they are asked to answer statistical questions. The ideal, however, can be more or less approximated and statisticians therefore try to attain as high a degree of approximation as possible under given circumstances. A study of various aspects of problems connected with the design and use of questionnaires in survey work is an extremely useful means of achieving this purpose.

Designing questionnaires is sometimes considered an easy task that can readily and satisfactorily be performed at any moment in the preparations for a survey. In actual fact, the opposite is true. In many cases and particularly in the absence of sufficient experience, the preparation of the questionnaire requires careful study, experimentation and testing of alternative drafts under different conditions. The consideration of all the problems that appear in the preparation of a questionnaire is therefore a very responsible task.

Some general information about the presentation of questions will now be given as it will be a useful basis for an understanding of various sources of biases.

Questions may be presented in the questionnaire in various forms: as *dichotomous* or *two-way* questions, *multiple-choice* questions, or *open* or *free-answer* questions. In the first case the question takes the form of an alternative so that there are two possibilities only for the formulation of the answer, which is most commonly expressed as either "Yes" or "No". Examples would be: "Have you ever worked off your farm during the last year?", "Are you married?", "Did you finish high school?", etc. Such questions are normally followed either by instructions to state the answer in the form of " Yes" or "No" or else by two boxes, one standing for YES and the other for NO, and one of which has to be checked according to the answer. In the multiple-choice or "cafeteria" type of question more than two possibilities are given for expressing the answer. An example of this type of question would be:

What is the legal status of this holding:

(*a*) Civil person
(*b*) Corporation
(*c*) Co-operative
(*d*) Collective farm
(*e*) Government
(*f*) Other.

Another example would be:

Is the person:

(*a*) Single
(*b*) Married
(*c*) Widowed.

With open questions no restriction is imposed on the answer: the respondent has a free choice. This type of question takes two basic forms. In the first a numerical answer is expected and the question is called "open" because any number can be put that fits the question. An example is: "What is the total area operated by this holding?" The answer is free because each holder is supposed to present the number which corresponds to the size of his holding. In the other form of the same type of question a descriptive answer is requested. Such questions are used in case studies and some special small-scale surveys where rather detailed information is needed on a particular problem. The following is an example of this type of question: "What kind of measures, in your opinion, would be the most adequate to improve the economic situation of agricultural holders in this area?" In such a case the respondent is expected to give his opinion in any way he wishes. He is also free to choose any measure he finds useful irrespective of what another person in his position might suggest.[10]

Dichotomous questions are considered simple and clear for respondents, who are in favor of such questions as they can give a prompt answer without losing time in reading the instructions. Such questions increase the respondents' co-operation and are recommended wherever the problem can be stated in this form. In preparing these questions one has to make sure that the alternative is complete in itself. If the question is put as follows: "Are you married?" and the answer is requested as either "Yes" or "No", a widower respondent will be perplexed because he does not belong to either of the two classes. The alternative has to be complete in the sense that the answer of all possible respondents is covered by either "Yes" or "No".

A similar problem exists with multiple-choice questions. Here the respondent is asked to indicate one of the choices presented. Accordingly, care has to be taken to ensure that the list of choices is complete so that each respondent finds it possible to check his own case. It is sometimes difficult to be sure that all possible cases are listed. The procedure followed in such a situation consists in listing the most important cases or those where one wants to have specific answers and in establishing the class "Others," which is supposed to cover less important cases as well as those which may not have been taken into account in the preparation of the questionnaire.[11]

The most important source of errors as far as the questionnaire is concerned is an *improper wording* of questions. The errors appear here for a number of reasons. Very often the person responsible for designing the questionnaire is not sure himself of the meaning of various definitions and concepts used in the survey. Accordingly, the phraseology of some questions may be vague and confusing. Designers of questionnaires often assume that the respondents know everything about the survey and its aims. Sometimes allowances are not made for the fact that the

---

[10] There are also combinations of these basic types of questions. For example, if the holder is asked to show to what type of farming his holding belongs by checking one of the alternatives listed and indicate, at the same time, the percentage of income he draws from the activities connected with the type of farming stated, this would be a combination of a multiple-choice and a free-answer question.

[11] Information on the relative merits of various types of questions is to be found in Payne, S.L.: *The art of asking questions*, Princeton University Press, 1951.

respondents belong to very different social categories and may not always be able to follow or give a proper connotation to the vocabulary used for various purposes in statistics. "The specialist may lose sight of the fact that others have no need for his jargon. He may think that because his associates and his technical books use the same lingo, his brand of gobbledygook should be universal. But what does he himself know of the terminology in some other field beyond his own?"[12] In all such cases of ambiguity in questions, the respondent is left to understand them as best he can. Obviously, this opens the way for errors.

In preparing the questions the most thorough study should be made of the interpretations that might be given to various questions. In some cases *misinterpretation* appears where one would hardly expect it. For example, the question "How many persons live in this household?" seems to have a clear meaning, but in fact this is far from being the case. Census experiences show that the answer to this question may also include people who used to live in this household, those who will live there some time in the future, those who are only present from time to time, etc. It is also a very well-known fact that babies are often excluded. When the enumerator points to the baby he receives the answer: "Oh no, that's the baby." In other words, for statistical purposes a baby is a person and for many respondents "a baby is a baby" and not a person. A similar type of difficulty appears in statistics of the economic characteristics of the population where terms like housewife, director, manager, agricultural activity, etc. can be interpreted in a large number of ways. In other words, much depends upon the phraseology used in preparing questions. There are many examples in statistical practice which show considerable differences in the results obtained on the same subject by using questions worded in different ways. Bancroft and Welch[13] have shown the change in the estimated total number of employed persons in the United States as a result of a change in questions. In 1945 a question asking whether the person selected was at work last week was changed by introducing two questions instead of one. The result of this change was a total of employed persons exceeding by 1,400,000 the number with the former system, which was used parallel to the new one in order to compare the results.

---

[12] Payne, S.L.: Op. cit.

[13] Bancroft, G. and Welch, E.H.: Recent experience with problems of labor force measurement, *Journal of the American Statistical Association*, Vol. 41, 1946, pp. 303-312.

Another example is taken from the 1953 Census of Population in Yugoslavia.[14] In this census there was a question on the type of activity which had to be filled in for each person of 14 years and over. For housewives and persons with several activities, instructions were given to indicate the type of activity to which the person devotes more than 50 percent of his or her time. In a quality check that was conducted immediately after the census enumeration was over, a discrepancy between the census and the check was found in 7.0 percent of the answers in rural areas and 8.5 percent in urban areas. This difference was largely the result of two different procedures used in obtaining data. Census enumerators accepted the information given by respondents. If the respondent was not able to give the answer, they asked for the type of work to which the person was devoting most of his time and on that basis determined the type of activity. Check inspectors, however, applied a series of questions with the aim of detecting all types of activities in which a person was engaged as well as an indication of the time devoted to each activity in a specified period of time. Their decision on the person's type of activity was based on this information. It is therefore obvious that different ways of presenting questions may lead to very different results.

The basic rule of good question wording is to make the meaning of the questions *as clear as possible*. This is normally achieved by making the questions short and by using simple words which are universally known. In this respect dictionaries showing frequencies of various words may be useful. A dictionary is also useful in checking what meanings might be attached to a given word. Much care is needed in using technical terms or some jargon which is restricted in its use to a group of people. If he is not able to understand some terms or questions, the respondent may not be willing to ask for explanations and thereby show what he may consider to be his ignorance. In such a case the answer is likely to be inaccurate. In a study reported by McCord[15] the following questions were asked:

(*i*) Have you ever heard of the word "afrohelia"?
(*ii*) Do you recall that as a good citizen you voted last December in the special election for your state representative?

---

[14] Zarkovich, S.S.: *Population census errors* (in Serbo-Croatian), Studies and Analyses No. 3, Federal Statistical Office, Belgrade. 1954.

[15] McCord, H.: Discovering the "confused" respondent: A possible projective method, *Public Opinion Quarterly*, Vol. 15, 1951, pp. 363-366.

(*iii*) Have you ever heard of the famous writer, John Woodson?
(*iv*) Have you ever heard of the Taft-Johnson-Pepper bill or veterans' housing?
(*v*) Have you ever heard of the Midwestern Life Magazine?

None of the things mentioned here ever existed, but the percentage of "Yes" answers was as follows:

Question	*Percentage of "Yes" answers*
(*i*)	8
(*ii*)	33
(*iii*)	16
(*iv*)	53
(*v*)	25

In other words, the choice of vocabulary has to take into account the respondents' psychology and the resulting possibility of getting inaccurate and even absurd answers.

In order to make questions clear it is also important to avoid the use of *problem words* wherever they might have a somewhat vague meaning. Each language has a number of such words. For purposes of question wording it might be useful to list them and in each particular case to check their use and the different meanings they can give to the questions. Payne[16] has listed some English words which are included in the class of problem words because they have several more or less different meanings. For example, "fair" might be substituted in a given case by any of the following words: average, intermediate, indifferent, medium, middle, moderate, neutral, normal, ordinary, standard, usual. "Any" might mean: every, some, one only. Some other words of this type are "people," "public," "nobody," "sometime," etc.

Another difficulty is connected with the use of *concept words*, such as government, labor force, employment, agricultural holding, etc. Each of these words can be interpreted in many ways, depending upon the respondents' education and social background. For example, "government" might be used for central or federal government as well as for provincial governments. In addition, it might mean anything from the ministers of a country at one extreme to the whole government administrative machinery at the other. If such words are used without a proper definition, all the meanings they have may be reflected in the results of the survey. On the other hand, if they are defined in the questionnaire, the respondents may not like such questions, since they have to make a special effort of memory and thought adjustment. It might be a good principle to avoid the use of such words whenever possible.

A proper use of *commas* will also help in some cases to make the meaning of questions clear. In designing questionnaires it may be useful to keep in mind the old Latin illustration of the importance of a comma, i.e., "*Ibis, redibis, nunquam in bello peribis*" as against "*Ibis, redibis nunquam, in bello peribis.*"[17]

Abbreviations are also sometimes a source of trouble in understanding the meaning of questions. This is particularly so if they are used for some units of measurement which are not commonly known. Here again it may be useful either to spell out the word completely or to find some other way of avoiding abbreviations.

Another problem of question wording is *loaded* or *leading* questions. By these terms a type of question is meant which, by the mere phraseology adopted, directs or leads answer in a given direction. If the aim is to discover something about the frequency of off-farm work the question might be used: "How many days did you work off the holding during the last year?" This question is loaded because it gives the impression to the respondent that his work off the holding has been taken for granted; the only problem is to determine the frequency of this work. With such a question the respondent might think that he is not expected to say "None" and so the wording of the question leads to inaccurate results. In this case it might be better to use two questions. The first could be: "Did you ever work off the holding during the last year?" to be answered by either "Yes" or "No." If the answer is "Yes," the second question should be specific and ask for the number of days worked off the holding.

Loaded questions very often creep into the questionnaire. Many examples could be mentioned as an illustration. In a survey the following question was asked: "What make of car do you own?" This question gives the feeling that owning a car is not subject to any doubt; it gives the impression that the real question is to know the make. In such a situation the respondent might think that his personal prestige is involved.

---

[16] Payne, S.L.: Op. cit.

[17] "You will go, you will return, you will never be killed in the war," as against "You will go, you will never return, you will be killed in the war."

As a result he may easily give an inaccurate answer. Similar effects appear whenever the question is presented in a way which sets in motion the complex machinery of the respondent's emotions and feelings. A survey made with the aim of estimating the number of persons who have read a famous book, attend concerts, etc. might yield results considerably above the respective true values. If names of famous people, politicians or political parties are mixed with the subject matter of the survey, the answers obtained will be influenced by the respondents' attitude toward these "stereotypes."

In this connection it might also be mentioned that the *circumstances* under which a survey is being made lead to the same consequences. If the people do not feel free to express their opinion and suspect some action if their answers disagree with the opinion of the sponsor of the survey, the results will often be a confirmation of the sponsor's ideas. For all these reasons care should be taken in wording questions to avoid any connection with words or topics that might give rise to interpretations of the above type. In surveys dealing with factual characteristics it will be easy to achieve this aim. In studies of opinions, attitudes and preferences it is not so easy to keep off emotional ground because this is often the ground of the survey. In such cases considerable care and skill are needed in order to achieve satisfactory wording of questions.

It might also be noted at this point that loading sometimes takes very complicated forms and is difficult to detect. For example, questions presented at the beginning of the questionnaire might lead the respondent in a given direction when he wants to answer some questions toward the end of the questionnaire. For illustration, a case cited by Hyman[18] might be mentioned. A study was conducted to determine the respondents' opinions on the question of whether the United States was spending too much money on foreign aid. The issue was introduced in two different ways, i.e., by means of two different questionnaires. The first questionnaire was made "internationalist" and the other "isolationist." In the internationalist context 75 percent of the persons interviewed answered that the United States was spending "the right amount" on foreign aid while the corresponding percentage in the isolationist context was only 20. In other words, answers already given, or the general attitude adopted as a result of answering a series of questions tend to influence answers to the remaining questions. Such a situation gives rise to what Crespi[19] called *commitment bias.* The answers tend to be consistent with the line adopted; the desire for consistency might in some cases be strong enough to produce inaccurate answers.[20]

Many other problems have to be kept in mind while designing the questionnaire. One of them is the *number* of questions or the *length* of the questionnaire. It is a well-known fact that a long questionnaire provokes a gradual decrease in interest on the part of the respondent. As a result less attention is paid to the accompanying instructions; the meaning of the questions is less well understood; the respondent's desire to answer the questions adequately declines and the quality of the information decreases. Moreover, a long questionnaire is a source of discouragement to respondents. Willingness to co-operate with enumerators has its limits; there are not many people who are ready to spare a long time for interviews or questionnaires. Refusals to co-operate are often the outcome of a long questionnaire.

No definite rule governs the problem of the permissible length of the questionnaire. A great deal depends upon circumstances. In a country where surveys have been a routine and the general education of the people is sufficiently high to allow quick reading and understanding of a properly prepared questionnaire, even a relatively long questionnaire may not present any serious difficulty. Other factors come into play. If the subject matter is interesting, respondents may be prepared to accept a long questionnaire and try to answer all its questions. The technical appearance of the questionnaire, the layout of the material, the size of the paper, and other aesthetic factors also play a considerable role. An unwieldy questionnaire where it is difficult to find the different instructions is not popular with respondents.

The problem of the *sponsor* is no less important. In many countries there is a widespread opinion that surveys sponsored or made by the government concern problems of common importance. It is also believed that data obtained in such surveys will be useful for the promotion of national goals and interests. In this situation one may, in principle, have a larger number of questions than in surveys sponsored by private bodies or business firms. Therefore in the latter case considerable atten-

---

[18] Hyman, H.H.: Problems in the collection of opinion-research data, *American Journal of Sociology*, Vol. 55, 1950, pp. 362-370.

[19] Crespi, L.P.: The interviewer effect in polling, *Public Opinion Quarterly*, Vol. 12, 1948, pp. 99-111.

[20] See the interesting example on the influence of context in *Gauging public opinion* edited by Hadley Cantril, Princeton University Press, Princeton, New Jersey, 1947.

tion is again necessary for all details in the preparation of the questionnaire.

Another related question is the influence of the *publicity campaign* conducted in connection with a survey. Publicity acts as a pressure on respondents; it helps to combat the common temptation to refuse cooperation. For this reason a survey with a long questionnaire can hardly be made without properly conducted publicity. On the other hand, small-scale surveys and those conducted by private organizations where considerable publicity machinery cannot be set in motion are bound to have a program of reduced scope.

After the program of the survey has been agreed upon and the corresponding questions have been prepared, a decision has to be taken with regard to the *order* of the available questions. A random arrangement of questions will hardly be useful, except in a case where there is no connection whatsoever between various questions. If questions can be classified in a number of groups relating to similar items, such as livestock, farm population, areas, etc., a random order of questions would produce a feeling that the same question had already been answered[21] and that the survey program has not been properly studied. The questions referring to the same topic therefore have to be presented together. For the order of groups or the order of individual questions within groups, a good working rule might be either to follow the order of importance or some kind of logical sequence. In presenting questions on areas of various crops harvested it is inadvisable to start with vegetables if these occupy only a small part of the total area harvested. Similarly, the respondent would be surprised if data on poultry or fur-bearing animals were asked for before information on horses and cattle.

In connection with the problem of the order of questions some experimental results may be pointed out which show that respondents pay more attention to items listed at the beginning and at the end than to those in the middle. An illustration of the point will be taken from Payne.[22] Four ideas, called here A, B, C and D, were presented to respondents in a different order and the respondents were asked to select one of them. The results obtained were as follows:

---

[21] Metzner, H. and Mann, F.: Effects of grouping related questions in questionnaires, *Public Opinion Quarterly*, Vol. 17, 1953, pp. 136-140.

[22] Op. cit.

Idea A was selected by:

27 percent when it appeared at the top of the list,
17 percent when it appeared near the center, and
23 percent when it was put at the bottom of the list.

Idea B was selected by:

11 percent when at the top,
7 percent when near the center, and
7 percent when at the bottom.

Idea C was selected by:

24 percent when at the top,
20 percent when near the center, and
21 percent when at the bottom.

Idea D was selected by:

23 percent when at the top,
16 percent when near the center, and
18 percent when at the bottom.

This shows a first preference for items at the beginning of the list and a second for those at the bottom.

Although it may be useful to take this experiment into account, it is not known to what extent its results may be expected outside laboratory experiments and public opinion surveys. If respondents are confronted with topics entirely unknown to them, the position of the question probably has more influence than in surveys dealing with factual characteristics, such as those in population and agricultural statistics.

It can be seen from this brief review of various problems in connection with the use of questionnaires in statistical work that designing questionnaires is not a simple task requiring no special qualifications or experience. The very opposite is true. It is therefore a good policy for data collecting agencies to have on their staff a person specialized in this matter.

5.4 Frames

By the term "frame" we understand the lists of units. These lists are used in sample surveys for the selection of samples. Since the composition of the sample is directly based on these lists, the frame is considered to be one of the most important tools in statistical work.

Frames may become a source of biases for two different reasons, one being the *composition* of the frame and the other the inadequate *use* of frames. The use of frames will be dealt with in the next section.

The composition of the frame may be accurate or inaccurate. The frame is called accurate if all the units, as resulting from the adopted definition of the population, are listed once and once only. If some unit is listed twice or several times or is omitted from listings, the frame is called inaccurate. The frame is also inaccurate if it contains units which do not exist in the population. In actual survey practice frames normally contain all these defects to a varying degree.

Frames become inaccurate for many different reasons. An important reason is that many populations are subject to continuous changes and frames easily become *out of date*. A list of persons or households based on a census of population easily becomes out of date. In other cases the frames are defective because they are compiled from *inaccurate material*. For example, if lists of establishments in a branch of industry are prepared by using the information available from the last industrial census, the resulting frame will be inaccurate as the census has omitted to register some establishments. *Careless work* in listing is another source of defects in the composition of frames. It appears, however, that out-of-dateness is the most important reason. This is so because of the tendency in sample surveys to avoid the preparation of fresh lists, which is normally an expensive operation.

Examples of frames available in various countries for survey purposes are electoral lists, population registers, lists of villages or enumeration districts as established in censuses, etc.[23]

The different defects in frames are not equally important. As regards the omissions, it is clear that units omitted have no chance of being selected in the sample. In addition, a considerable proportion of omissions will seriously affect the estimated totals as the inflation of sample data is based on the number of units listed in the frame. Duplications offer double chances for the selection of units listed twice. Their effect on the estimated totals is the reverse of that of omissions. However, duplications are theoretically less difficult to deal with as units selected twice in the sample would be detected as such in the course of the field work. Nonexistent units in the frame affect totals but do not cause much trouble if selected in the sample. The field work shows that there is nothing in reality that corresponds to these units in the frame. However the nonexistent units selected affect the precision of the resulting estimates if they reduce the size of the sample.

---

[23] A more detailed description of various files is to be found in Deming, W.E.: *Some theory of sampling*, John Wiley, New York, 1950; Moser, C.A.: The use of sampling in Great Britain, *Journal of the American Statistical Association*, Vol. 44, 1949, pp. 231-259; Yates, F.: *Sampling methods for censuses and surveys*, Third edition, Charles Griffin, London, 1960.

The simplest solution in dealing with the problems arising out of inaccurate frames would be to construct an up-to-date and accurate frame. In most cases, however, such a solution cannot be considered as the establishment of a new frame is an extremely expensive operation. This is why various measures are currently used in sample surveys to make possible the utilization of existing and more or less inaccurate frames. Some of these measures are described here.

It may be useful to point out that inaccuracies in frames normally affect low-ranking units in the hierarchy of sampling units. If the units theoretically involved in a population survey are ranked in the following way: provinces, districts, villages, houses, households, and persons, the danger of inaccuracies in the frame will increase as it passes from provinces to persons. Provinces and districts do not change easily. On the other hand, frequent changes occur in the lists of households and particularly in the lists of persons, these being perpetually subject to change. In this situation a device to avoid inaccurate frames is to sample higher ranking units in the hierarchy of sampling units. In other words, if considerable difficulties are expected with a list of persons the problem may be solved by sampling villages. After a sample of villages is selected the persons in these villages might be listed again or the existing frame brought up to date by means of a new field visit. Obviously, such a solution introduces a new stage of sampling and its advantages have to be weighed against the increased variance.

This measure will provide for omissions, duplications and also for the listing of nonexistent units.

If the frame is arranged geographically a simple procedure for incorporating the units omitted or the "new births" is to provide enumerators with the identification details of the unit selected and of the one that follows on the list. Before the enumerators start collecting data they are instructed to verify whether the adjacent units in the frame are also adjacent in the field. If some units have appeared between the two they will enumerate all of them together with the one originally selected for the sample. This measure is intended to secure adequate selection probabilities for all the existing units irrespective of whether they are listed or not.

A similar approach is discussed by Zarkovich and Krane.[24] Its essence consists in the use of compact cluster sampling. If data from an old census are available by elementary units, such as agricultural holdings, persons, households, etc., clusters of these units are established in the old material along with cluster totals for the characteristics on the survey program. In the new survey all the elementary units are enumerated within the boundaries of the established clusters and fresh cluster totals are prepared for the same characteristics. The new enumeration is supposed to be based on an accurate list of units. The inaccuracies in the frame will affect cluster totals. However, the use of compact clusters will secure high correlation between cluster totals in the old and in the new material. Clustering of units is normally detrimental to efficiency. In this case, however, we have the positive effect of the increased coefficient of correlation that appears as a result of clustering. The question then arises of the optimum size of cluster. This issue is connected with the relationships between the magnitude of the coefficient of correlation and the size of cluster. Some illustrations regarding this problem are given in the paper quoted.

The technique referred to here was used in the quality check of the 1953 Census of Population in Yugoslavia[25] and in the 1956 Census of Agriculture in Canada.[26] In the latter case the selection of the sample was as follows. Each census enumerator was provided with a map of his enumeration district. In the course of enumeration he had to indicate on the map the point where the enumerated farms were located. Some enumeration districts were divided into segments and in these districts the enumerator's map was superimposed on an identical map where the borders of the selected segments were marked. The cluster of three farms was then selected in the sample whose location fell within the boundaries of the segment.

---

[24] Zarkovich, S.S. and Krane, J.: Some efficient uses of compact cluster sampling. Paper presented at the 35th Session of the *International Statistical Institute*, Belgrade, 1965.

[25] Cf. Zarkovich, S.S.: *Population census errors* (in Serbo-Croatian), Federal Statistical Office, Belgrade, 1954.

[26] Dominion Bureau of Statistics: *1956 Census of Agriculture*, Bulletin No. 2, Queen's Printer and Controller of Stationery, Ottawa, 1957.

A similar technique was used in several surveys and quality checks in the United States.[27] The resulting sample was called a "segment list sample" because the segments were used to select listed units. The segment list sample was used to collect data for units listed in the frame. Data for units missed in the frame were collected from an additional area sample where an intensive matching operation was carried out with the aim of identifying the units missed.

Another way of including in the sample the omissions or the units that came into existence since the construction of the frame is to sample the latter part of the population from other up-to-date lists. This is how, in the United States, the basic sample in the annual survey of manufactures is selected from the last quinquennial Census of Manufactures while the new births are taken from the lists kept with the Bureau of Old Age and Survivors' Insurance.[28]

Another illustration can be drawn from the annual survey of industries and labor conditions in India.[29] It was found that in this case the most convenient frame was the list of registered factories as available in the publication *Large industrial establishments.* The book is issued annually but the lists published are about two years old when made available. At the moment of a survey based on a sample selected from such lists some factories will be out of business while some new establishments will have opened. Seal has considered these changes in the population of factories as a continuous stochastic process and has used successive frames to allow for the changes in the population when estimating totals and averages from sample data.

The paper by Gilford and Marks[30] which deals with inaccurate files of corporations might usefully be consulted in this connection.

---

[27] Hansen, M.H., Hurwitz, W.N. and Jabine, T.B.: The use of imperfect lists for probability sampling at the U.S. Bureau of the Census, *Bulletin of the International Statistical Institute*, Vol. 40, Part 1, 1964, pp. 497-517; Jabine, T.B., Hurley, R. and Hurwitz, W.N.: Sample design and estimation procedure for the 1960 sample survey of agriculture in the United States, *Review of the International Statistical Institute*, Vol. 29, 1961, pp. 1-12; Jabine, T.B.: Checking the accuracy of area statistics obtained in the United States censuses of agriculture, *Estimation of areas in agricultural statistics*, FAO, Rome, 1965.

[28] Hansen, M.H., Hurwitz, W.N. and Jabine, T.B.: Op. cit.

[29] Seal, K.C.: Use of out-dated frames in large-scale sample surveys, *Calcutta Statistical Association Bulletin*, Vol. 11, 1962, pp. 68-84.

[30] Gilford, D.M. and Marks, C.L.: Use of a sample survey for estimating an aggregate quarterly financial statement for a population of corporations, *Improving the quality of statistical surveys*, American Statistical Association, Washington, 1956, pp. 15-30.

If the inaccuracy of the frame consists in excessive duplications it is sometimes recommended that units in the frame should be ordered according to different characteristics so that one can ascertain whether the adjacent units in this order represent duplications. For example, if the same holding is listed twice under two different names, the value of other characteristics being equal, the two units will be adjacent if the frame is ordered by total area or some other important characteristic.

Other techniques and illustrations are discussed in Hansen, Hurwitz, and Jabine,[31] Szameitat and Schäffer,[32] Thionet,[33] and Yates.[34]

In most cases none of the techniques mentioned here could be applied without some knowledge of the quality of the frame. This will involve knowing the frequency of various types of inaccuracies in the frame, what group of units is affected, how the inaccuracies are distributed over the territory, etc. The knowledge needed here is collected through quality checks of frames or quality checks of listing. This is the subject matter of our next chapter.

5.5 Inadequate use of frames

The inadequate use of frames is in most cases due to:

(*i*) confusion of units,
(*ii*) confusion of populations,
(*iii*) inaccurate information about units accurately listed, and
(*iv*) wrong assumptions about the structure of the population.

Confusion of units occurs if units of observation are different from listing units or sampling units and no account is taken of this difference. Let us imagine that dwelling units are selected and that an adult person is interviewed in each selected dwelling unit on the cost of transportation on the day preceding the survey. The estimated average per person expenditure on transportation will be biased unless individual data are weighted by the number of adult persons living in the dwelling units selected. If weighting is neglected the differences between dwelling units in the number of people and the expenditure on transportation will not be accounted for.

Frames are frequently used with a view to getting some other type of unit by means of units listed. In all such cases it must be seen that adequate measures are taken to prevent biases in the final estimates from appearing in this field. The measures normally involved are weighting and/or special instructions to enumerators. In order to sample parents from lists of schoolchildren, the enumerators have to collect data on the number of children in the school in each house they visit as the weighting will otherwise be impossible.

A confusion of populations arises if the population described by the frame is not the same as the one with which the survey is concerned, i.e., the target population. An illustration of extreme cases falling into this class are surveys aiming at all adult persons in a city or country while the sample of persons is selected from the telephone directory or from subscribers to a magazine. The population of subscribers to either the telephone or to a magazine is not equivalent to the population of adults. If there is any difference between these populations with regard to the survey program the results of the survey will be biased.

Inaccurate information about listing units in the frame may lead to biased samples if this information is used to establish a subpopulation and a sample therefrom. The following example explains this point. The frame is assumed to consist of the questionnaires filled in a census of population with information available for each person on age, sex, type of activity, etc. The survey to be made refers to females who are 21 years of age and over and who are employed in agriculture. A ten percent sample of such persons is selected by going through the census questionnaires, disregarding all the questionnaires that belong to persons outside the population defined and selecting every tenth person from those who do belong to the population according to the information available. The resulting sample may lead to biases if the information on age and employment contains systematic errors.

Frames are often used in this way and samples frequently become biased as a result. Even if the original frame is accurate, information about each listing unit may be inaccurate for many reasons. Retail shops understate their sales, manufacturing establishments their production and the number of their employees, while individual persons exaggerate or under-

---

[31] Op. cit.

[32] Szameitat, K. and Schäffer, K.A.: Imperfect frames in statistics and the consequences for their use in sampling, *Bulletin of the International Statistical Institute*, Vol. 40, Part 1, 1964, pp. 517-538.

[33] Thionet, P.: *Application des méthodes de sondage aux enquêtes statistiques*, Institut national de la statistique et des études économiques, Paris, 1953.

[34] Yates, F.: *Sampling methods for censuses and surveys*, Third edition, Charles Griffin, London, 1960.

state their incomes, etc. All such inaccuracies are directly reflected in the composition of the resulting populations.

Wrong assumptions about the structure of the population are sometimes the source of biased samples. Let us assume that an accurate frame of certain units is available and an unbiased sample of these units is selected. To return to an example discussed earlier, such a sample is that of streets from the list of streets available in most cities. In order to simplify the selection of the subsample the enumerators might then be instructed to include in the sample each house with a number ending in some specified digit. This is done for the sake of economical listing and on the assumption that each street has houses corresponding to all the serial numbers. If this assumption were justified and a fresh random start made in each street, the sample of houses would be unbiased. We know, however, that no houses exist for some numbers, and that sometimes several houses bear the same number. Biases result unless precautionary measures are adopted.

In another case the sample might consist of dwelling units selected from the corresponding frames. In order to reduce the cost of listing and to simplify instructions, enumerators might be instructed to interview the housewife in each selected dwelling unit regarding expenditure on various food items. The instructions, of course, lead to an unbiased sample under the assumption that in each dwelling unit there is only one household and only one housewife. If several households are living together the result will be a biased sample unless all the households living in the same house are interviewed. On the other hand, if several independent persons live in a dwelling unit there will be no housewife and none of the persons involved will be represented in the sample.

It will be seen that the inadequate use of frames has nothing to do with their accuracy. An accurate frame can be inadequately used and the resulting estimates will be biased. For this reason each use of a frame will raise two different questions. The first will relate to the accuracy of the frame and the second to the adequacy of the proposed use. Both issues have to be considered separately.

5.6 Instructions

Instructions prepared in connection with various surveys are as important as the questionnaires if not more so. Instructions are the means of achieving the standardization of a procedure, such as data collecting, coding, editing, etc.

It is difficult to undertake a general discussion of the preparation of instructions as the decisive and guiding elements vary from one case to another. It is clear that instructions prepared for the field staff for collecting data in a new survey will have to differ according to the degree of training and experience the staff has acquired from the previous work. Bearing this in mind some general principles will be presented which may be useful in preparing the instructions.

The first such principle is to make instructions as *clear* as possible. This aim is achieved by using simple language and words which have an unequivocal meaning. It is also advisable to avoid concepts and terms which do not belong to the vocabulary of a person with an education of an entirely general type. If the text is complicated and not readily understood, the reader may not be willing to make an effort to understand the instructions properly. In all such cases he may proceed according to his own understanding of the situation.

The next requirement is to make the instructions *complete*. The instructions are called complete if they provide guidance on the procedure to be followed for all the problems that may be anticipated. The points which are left open make it possible for the staff to introduce into their work a subjective interpretation of the situation.

To illustrate this point the following example may be useful. In a survey respondents of the age of 14 years and over were requested to answer a question concerning their "occupation." The instructions associated with this question read: "In stating the occupation sufficient detail is needed to make it possible to classify the reply according to the nomenclature of occupations. For example, instead of 'tractor driver,' the details 'tractor driver working for a transport company' should be given, thus making it possible to distinguish this person from another 'tractor driver' working on his own farm."

These instructions are incorrect in a number of respects. First of all they presume that the concept of occupation is known to everybody, which is certainly not the case. They then use the term "sufficient detail" which is more than vague. People will certainly disagree as to what is to be considered sufficient detail. Two other terms are even more difficult to understand for a person who may have but little general education. These terms are "classify" and "nomenclature of occupation." The survey involved was a general population survey and the percentage of persons who would know the meaning of the word "classify" is certainly not large. This applies even more to the concept of the "nomenclature of occupation."

If a person is a student, these instructions do not say whether the reply "student" is sufficient or whether additional specification is required, such as the type of school, the grade, the type of specialization within the school's curriculum, etc. If somebody is employed, he or she will not know whether such a simple reply is sufficient or whether more has to be said, such as the number of years of service, the position the person occupies, the employer's main activity, etc.

Whenever the instructions do not provide complete guidance the result is not only the introduction of subjective elements into the formulation of the answer. Such instructions irritate users. In some cases the effect is to produce complete disregard of the instructions. Refusal to co-operate with the survey is another reaction.

Another principle is that the instructions should be of the *right length.* If they are too short, they are often neither complete nor clear. If they are too long, they become fatiguing and respondents may feel reluctant to look them up for guidance.

Obviously, it is not easy to establish the right length. Much depends upon the work involved. Complicated operations generally need more explanations than simple ones. General respondents, who normally have no interest in statistical surveys, need shorter instructions as they may not have sufficient patience to go through long explanations. For this reason it is a very common practice to reduce the volume of the instructions at the expense of their completeness. Normally, the problem is solved by covering the most important points with short instructions and leaving more complex and infrequent cases to be handled by the professional staff at an appropriate stage of processing or when the questionnaires are collected from the respondents.

Finally, the instructions have to *correspond adequately to the purpose* they are supposed to serve. If this principle is properly obeyed, most of what was said above will be automatically solved in a correct way. For example, if instructions on data collecting have to be written up for a rather small and qualified staff of field workers with several years' experience of data collecting, the purpose might best be served by writing very short instructions referring to points which this staff have not encountered in any previous experience. It will not be difficult to see in this case in what way and what language the instructions have to be presented in order to be clear. It will also be easy to establish what points have to be covered in order to make them complete.

It is necessary to point out the complexity of the work on drafting instructions. This work requires a perfect knowledge of the subject matter of the survey involved, of the machinery that is to be used in achieving the survey aims, of the characteristics of the population to be dealt with, of the psychology of both the respondents and the enumerators, etc. This is again a reason for having experts on the staff of data-collecting agencies to prepare the instructions, working together with ad hoc experts appointed for their specialized knowledge of the topic dealt with in the instructions. For example, in a census of population, in addition to statisticians responsible for the organization, carrying out and processing of the census, members of the committee for drafting instructions might be experts on: educational characteristics of the population, economic characteristics, vital statistics, nomenclature of occupation, labor force, etc. The committee may also include a psychologist, a sociologist, an economist, the representatives of the government agencies responsible for the policy matters in a given field, etc.

6. LISTING ERRORS

6.1 Introduction

In our study of listing errors we shall be using the concept of the listing bias, which was introduced in Chapter 1. The listing bias is also called the *net error* of listing. In addition it will be useful to introduce the concept of the *gross error*. The latter is very useful for presenting a numerical indication of the quality of the enumerators' work. The gross error of listing, G, is defined as

$$\begin{aligned} G &= \sum_i (z_i - x_i)^2 \\ &= \sum_i d_i^2 \end{aligned} \tag{6.1}$$

In this equation $z_i = 1$ if the i-th unit is listed in the survey. Similarly, $z_i = 0$ if the i-th unit is not listed. The quantity x_i is the corresponding true value. The quantity

$$100 \times \left(1 - \frac{G}{N}\right) \tag{6.2}$$

shows the percentage of cases where the enumerators responsible for the listing were able to carry out the listing accurately.

As against the gross error in (6.1) the net error is defined as

$$\begin{aligned} D &= \sum^N (z_i - x_i) \\ &= \sum d_i \end{aligned} \tag{6.3}$$

The magnitude of G depends upon how many errors were committed, while D depends upon whether errors cancel out or not. This is why large G may be associated with $D = 0$.

6.2 Checking the quality of listing: Type I design

Checking the quality of listing can be done by means of post hoc techniques (cf. Chapter 2). This chapter will deal with the use of sampling methods for this purpose.

Quality checks of listing can be designed in a number of different ways depending upon the circumstances. It appears, however, that two basic designs may be sufficient in most practical situations. We call them Type I and Type II designs. In this section we shall concentrate on Type I designs.

Type I designs are based on compact cluster sampling.[1] By compact cluster sampling we understand the selection of clusters of units and the collecting of data for all the units included in the clusters. In other words, the selected clusters are not subsampled.

Let us now see what reasons justify the use of compact cluster sampling.

To be satisfactory, the design for a quality check of listing must be able to detect all kinds of listing errors, i.e., omissions, duplications, and listing of nonexistent units.[2] If we select a sample of elementary units, persons, agricultural holdings, etc. from census lists and check the quality of listing of each of them, we shall be able to estimate the frequency of the inclusion of nonexistent units and their effect on census results. No information will be available about omissions and duplications. In fact, an estimate of errors of all three kinds will be possible if an adequate chance is given to omitted and duplicated units to appear in the sample. This will be done by selecting some area segments and making a fresh study of a sample of such segments. The comparison of the results obtained in the two surveys will offer a chance to estimate the effects of all types of listing errors. Since area segments are conceived here as compact clusters of listing units, this means that compact cluster sampling makes it possible to achieve the aims of a quality check of listing.[3]

---

[1] Cf. Hansen, M.H., Hurwitz, W.N. and Madow, W.G.: *Sample survey methods and theory*, Vol. I, John Wiley, New York, 1953.

[2] Nonexistent units represent primarily the problem of old frames.

[3] In some exceptional cases convenient clusters for checking the quality of listing do not need to be area segments. If listing is carried out according to geographical contiguity one might select clusters of consecutive units from census lists and instruct check enumerators to list all the units they find in the field between the borders established by the first unit in the cluster selected and the first unit of the subsequent cluster. In this way there is also a chance that all types of errors will be detected and adequately estimated.

Compact clusters are often easy to establish. If the clusters needed for checking the quality of listing are the EDs, no extra work or resources are required to prepare the quality check. If EDs are too large there may be some convenient way of splitting them into several segments. In this case a sample of EDs is selected and only these are divided into segments. In the same way the census material related to the selected EDs is also separated into parts which coincide with the segments established. The decision as to whether the large existing EDs are better than the smaller segments which require work should be made on efficiency grounds.

Compact sampling is also satisfactory from the efficiency point of view. In order to show this let us take $z_{ij} = 1$ if the j-th unit of the i-th ED is listed in the census and $z_{ij} = 0$ if it is not listed. A similar definition is used for x_{ij}. The quantity $Z_i = \sum_j z_{ij}$ is the total number of units listed in the census in the i-th ED. The meaning of X_i is analogous. The comparison of the results obtained in the two surveys is based on Z_i and X_i. In fact, the variable z is used as the supplementary information in the estimation of X and of the effect of listing errors on census totals. The efficiency of the resulting estimates will depend upon the correlation between Z_i and X_i.

In considering the correlation between census data and check data it will be useful to remember that we have $z_{ij} = 0$ for all the omitted units, $x_{ij} = 0$ for nonexistent and duplicated units. With a relatively large number of listing errors in the census the correlation between x_{ij} and z_{ij} will thus be poor. As we go toward larger compact clusters the effects of listing errors on the correlation between cluster totals will gradually decrease. At some size of the cluster the correlation between X_i and Z_i will be so high as to make the use of compact clusters an efficient design for the purpose involved.

In each particular use of compact clusters for the purpose of checking the quality of listing, the question of the optimum size of cluster will arise. The correlation between Z_i and X_i will increase with the increase in the size of the cluster. At some size of the cluster the correlation will be so high that further increase in the size of the cluster will not lead to adequate gains in efficiency. Determining the optimum size of cluster is a matter of experimentation involving characteristics, cost elements, quality of the original work, etc.

Some indication of the relationship between the magnitude of the coefficient of correlation and the size of cluster for a number of characteristics is given by Zarkovich and Krane.[4]

6.3 Estimation procedure

The bias in the total number of units as listed in the census is defined as $D = Z - X$ with $D = \sum_i \sum_j d_{ij} = \sum_i D_i$. It is assumed that the population consists of EDs designated here with $i = 1, 2, \ldots, M$. From this population a sample of m EDs is selected and after the check survey has been taken D in (6.3) is estimated as

$$D' = \frac{M}{m} \sum_i^m D_i \tag{6.4}$$

The quantity D_i is the bias in the number of units listed in the i-th ED. D_i is obtained by grouping the units of the i-th ED in the following three groups:

(*i*) units listed in both the census and the check. For these units, $d_{ij} = z_{ij} - x_{ij} = 0$. These units are called "agreements" between the two surveys;

(*ii*) units included in the census and not in the check ($d_{ij} = 1$). These are called "erroneously included" units;

(*iii*) units included in the check and not in the census ($d_{ij} = -1$). These are called "omissions."

After this classification D_i is obtained by summing up the totals in (*ii*) and (*iii*).

After D has been estimated we can also estimate the total numbers of units according to the check. The estimate is $X' = Z - D'$. Before estimating X we should know whether D' is significant. For this purpose we need the variance

$$\sigma^2_{D'} = \frac{M^2}{m} \sigma^2_{D_i}$$

$$= \frac{M}{m} \left[\sum D_i^2 - \frac{1}{M} (\sum D_i)^2 \right] \tag{6.5}$$

---

[4] Zarkovich, S.S. and Krane, J.: Some efficient uses of compact cluster sampling. Paper prepared for the 35th Session of the *International Statistical Institute*, Belgrade, 1965.

In most situations the second term on the right-hand side of (6.5) will be negligible or small. The variance $\sigma^2_{D_i}$ thus depends upon ΣD_i^2 which is the sum of squares of the biases of individual EDs. For practical purposes we therefore use the approximation $\sigma^2_{D_i} \doteq \frac{1}{M} \Sigma D_i^2$ unless it has been found that D' is relatively large.

Clearly, $\sigma^2_{D_i}$ is estimated from the sample as

$$s^2_{D_i} = \frac{1}{m-1}\left[\sum^{m} D_i^2 - \frac{1}{m}\left(\sum^{m} D_i\right)^2\right] \tag{6.6}$$

or

$$s^2_{D_i} \doteq \frac{\Sigma D_i^2}{m-1} \tag{6.7}$$

An approximate test of the significance of D' is obtained from

$$\frac{D'}{s_{D'}} = \frac{\sum^{m} D_i}{\sqrt{\Sigma D_i^2}} \tag{6.8}$$

If the enumerator on the right-hand side of (6.8) is larger than the double magnitude of the denominator, the listing bias may be considered significant.

To estimate the effect of listing errors in other variables the same procedure is used with the actual value of the characteristics involved (z_{ij} and x_{ij}) in the groups *(ii)* and *(iii)* above.

Instead of the difference estimate which was used here one might wish to follow the ratio method.

6.4 Presentation of results

In this section we continue with the assumption that out of M EDs in the population m EDs are selected for checking. After the check is over data obtained are presented in tables called here Tables A and B. Table A shows the effect of listing errors on the number of holdings in various classes of the size classification of holdings. Table B shows the effect of listing errors on various census characteristics broken down by totals of size classes. As an example of such a characteristic we take here the total area of agricultural holdings. Needless to say, it could be any other characteristic.

The stubs of these two tables show what information is to be presented. The absolute numbers which appear in these two tables *refer to units in the sample.* If the estimates for the population available are needed, the figures in the tables are multiplied by M/m. Clearly, this does not apply to percentages as these represent estimates of the corresponding population percentages.

The information included in Table A is the following:

1. Number of units listed in the census (in the selected EDs).
2. Number of units listed in both the census and the check ("Agreement").
3. Number of units listed in the census and not in the check ("Erroneously included units"). The total in lines 2 and 3 should be equal to the figure in line 1.
4. Number of units listed in the check and not in the census ("Omissions").

TABLE A. - EFFECT OF LISTING ERRORS ON SIZE CLASSIFICATION OF HOLDINGS

		Size classification of holdings (in ha)				Total
		0.5 - 5	5 - 10	10 - 20	20 - 50[1]	
1. Census	Number					
2. Agreement	Number Percent					
3. Erroneously included	Number Percent					
4. Omissions	Number Percent					
5. Bias	Number Percent					
6. Squared EDs' biases	Number					
7. Significance test						

[1] Columns in this sequence can be extended as far as is desired.

5. Bias (which is equal to the difference between 3 and 4). The reader is reminded that the table refers to sample data. Therefore, the quantity in this line is not D' from the equation (6.4) but $\overset{m}{\Sigma} D_i$.
6. Squared EDs' biases, i.e., $\overset{m}{\Sigma} D_i^2$. This is the quantity needed for the computation of the variance [according to the equation (6.6)].
7. Significance test, i.e., the quantity $D'/s_{D'}$ in the equation (6.8). If the figure in this line is larger than 2, the bias might be considered significant.

If the approximation to the variance used here is not found satisfactory because of the relatively large bias, one can use the figures in lines 5 and 6 and get a more accurate estimate of the variance by means of the equation (6.6).

In order to obtain data for Table B the classification of units in the stub of Table A is also used. For units falling in each of the classes 2, 3 or 4 the totals are established for the characteristics included in the check. From these totals the bias is computed as well as the remaining figures in lines 6 and 7.

TABLE B. - EFFECT OF LISTING ERRORS ON TOTAL AREA OF AGRICULTURAL HOLDINGS

		Size classification of holdings (in ha)				Total
		0.5 - 5	5 - 10	10 - 20	20 - 50[1]	
1. Census	Amount					
2. Agreement	Amount Percent					
3. Erroneously included	Amount Percent					
4. Omissions	Amount Percent					
5. Bias	Amount Percent					
6. Squared EDs' biases	Amount					
7. Significance test						

[1] Columns in this sequence can be extended as far as is desired.

6.5 Some comments

In connection with Tables A and B the problem arises of the amount of work needed to present the data of a listing check in this form. It must be pointed out here that Table A has to be prepared only once because it shows the effect of listing errors on the frequencies of the size classes. Table B has to be prepared separately for each characteristic that needs to be checked for the effect of listing errors. Of course, the number of such characteristics will depend on the resources and facilities available.

In designing Tables A and B the size classification of holdings was put in the headings. This was done on the understanding that this classification is the most important one in censuses of agriculture. In addition, it appears that this classification represents the most indicative analysis of listing errors. Of course, this does not mean that some other classification should not be used.

It may also be pointed out that the primary intention of Tables A and B is to help statisticians in the analysis of the quality of listing and in planning measures for the improvement of the techniques used. Users of census data are of course also interested in the information available in Tables A and B because it helps them in placing the right amount of reliance on the figures published. In most cases, however, users would be satisfied with much less information, such as an estimate of the bias.

How useful these tables are for the analysis of listing errors will be clear if one remembers that:

(*a*) Data in Table A, line 2, show the percentage of cases where the two listings agree. Data for various size classes also point out what holdings are most affected by disagreements and where attention should be directed in future surveys to ensure higher quality of listing.

(*b*) Line 4 shows the number of omissions in census listing classified by size classes. This information is extremely useful in making an evaluation of the enumerators' ability to secure good coverage. It may also suggest what steps might be taken to improve the coverage.

(*c*) Data on "erroneously excluded" and "erroneously included" holdings also represent a measure of the enumerators' success in interpreting and applying census instructions. If the work is not deemed satisfactory in this respect, this indicates the need for further training, more efficient methods of training, better instructions and working methods, improved mapping, etc.

(*d*) Line 5 is the net effect of errors present in census listing. Data in this row show to what extent errors vary in different size classes. They indicate the overall effect of listing errors on the number of units and offer a final basis for judging the adequacy and appropriate use of various measures for the improvement of listing.

(*e*) The sum of the figures in the respective columns of lines 3 and 4 shows the number of errors committed irrespective of their nature. This information is also extremely useful for the evaluation of the procedures adopted.

It is easy to ascertain that equally useful information for the analysis of errors in censuses is also available in Table B. This table answers many questions about listing errors that are of interest to statisticians. In fact, every statistician responsible for planning censuses would like to know to what extent he should worry about listing errors in connection with various characteristics, whether these errors can be disregarded without running the risk of biases, how different population classes vary with regard to these effects, etc. All such questions can be answered by studying data in Table B for the characteristics involved.

6.6 Type II design

Type I design may not be suitable in the case of primary sampling units which contain a large number of listing units. Such cases appear in censuses of population with EDs that have on an average 1,000 or more persons. Relisting of all the units in this case would be a costly operation. Type II designs help to reduce the cost.

Type II designs are used in a situation where the elementary units are grouped in some kind of natural cluster and listing errors do not only affect elementary units but the clusters as well. Such clusters consist of persons within a household or dwelling units within houses. Listing errors affect persons and dwelling units, and listing errors of clusters are on the whole no less frequent. A common experience in censuses of population is that households or dwelling units are often omitted.

In this situation the essence of the Type II design is as follows. A sample of EDs is first selected. Within the selected EDs a check is carried out of the quality of listing of clusters, say households or dwelling units. This check is carried out as with Type I designs. Dwelling units or households are relisted in the whole territory of the selected EDs and the results are matched with census listing in order to ascertain the listing errors of clusters. All the elementary units (persons) in the dwelling units omitted are enumerated and the basis is thus obtained for the estimation of the listing bias due to listing errors of clusters. In order to estimate the bias due to listing errors of elementary units (persons) within the clusters listed, a sample of clusters is selected from census lists and the check of the quality of listing is effected by considering separately the listing of each elementary unit. Data obtained make it possible to estimate the effect of the listing errors of elementary units. The two components of the listing bias are then combined in one single estimate. The first part of the check is called here *the quality check of the listing of clusters* or briefly *clusters check* while the other is called *the quality check of the listing of elementary units* or *elementary units check.*

It will be seen that in Type II designs clusters of units at two different levels are involved. EDs are used so that all the listing errors concerning households may appear in the sample while households are used as clusters so that listing errors related to persons may appear in the sample.

In practice the clusters check and the elementary units check may be combined in one single field operation. From the population of M EDs m EDs are selected to carry out a clusters check. From the census lists available in the selected EDs a sample of clusters is selected and these are used to check the quality of the listing of the elementary units. The two activities are carried out together. Whenever the enumerator comes to the selected cluster for the purpose of the clusters check he will also carry out the elementary units check.

With Type II designs the estimation procedure is, of course, more complex. Since the bias has two components, we put

$$D = A + B \tag{6.9}$$

A being the contribution to the bias of errors in listing clusters and B the contribution of errors in listing the elementary units.

An estimate of the bias, D', is defined as

$$D' = A' + B' \tag{6.10}$$

where A' and B' are the estimates of A and B respectively.

In order to show how A and B are estimated let us suppose that the quality of listing in a census of population has to be checked. Out of the available M EDs a sample of m EDs is selected and in each of them the quality of the listing of households is checked. In the selected EDs a sample of households is selected from census lists with a view to checking the quality of the listing of persons belonging to the households selected.

The use of compact cluster sampling will have been noticed. In this case households are used as compact clusters of persons. After the check survey has been made the quantities Z_{ij} and X_{ij} are available. Z_{ij} stands for the total number of persons listed in the census in the j-th household of the i-th ED. X_{ij} refers to check data and has the analogous meaning. It is hoped that Z_{ij} and X_{ij} will be highly correlated, thus making possible efficient estimates of biases. If households are too small to have a sufficiently high value of the coefficient of correlation, an alternative type of cluster might be the groups of households. The selection of the optimum size of these clusters is obviously a matter of empirical studies.

If the component B of the bias is estimated by means of the difference estimate, we first find $D_{ij} = Z_{ij} - X_{ij}$ and then put

$$B' = \frac{M}{m} \sum_{i} \frac{N_i}{n_i} \Sigma D_{ij} \tag{6.11}$$

where N_i and n_i are the number of households in the population and in the sample of the i-th ED. Needless to say, $\sigma^2_{B'}$ is available in any textbook on sampling. An unbiased estimate of $\sigma^2_{B'}$ is

$$s^2_{B'} = \frac{M^2}{m\,(m-1)} \Sigma\, (D'_i - \bar{D}')^2 \tag{6.12}$$

with

$$D'_i = \frac{N_i}{n_i} \sum_{j} D_{ij}$$

$$\bar{D}' = \frac{1}{m} \sum_{i} D'_i$$

In the estimation of the component A it is assumed that the check enumerators have identified all the households omitted in the census. All the persons in such households are enumerated in the check. In addition, all the households which should not have been enumerated in the census have also been identified. After the check is over we establish the quantities:

Z_i = the number of persons enumerated in the census in all the households listed in the census in the i-th ED,
X_i = the corresponding total as obtained in the check and
$D_i = Z_i - X_i$.

Afterward A is estimated by means of the equation (6.4) with A' substituted for D'. The variance of A', i.e., $\sigma^2_{A'}$ is obtained from the equation (6.6).

As regards the variance of D' in the equation (6.10) we put

$$\sigma^2_{D'} = \sigma^2_{A'} + \sigma^2_{B'} + 2\, \rho_{A'B'}\, \sigma_{A'}\, \sigma_{B'} \tag{6.13}$$

The last term on the right-hand side expresses the covariance between the two components of the total bias. This term can probably be disregarded in the actual work as there is no reason why the two components of the bias should vary in parallel fashion. Should the experience show that this is not so it might be possible to get independent estimates of A and B. Namely, A could be estimated from a sample of EDs where the cluster check is carried out as a part of the regular duties for which the check supervisors are responsible. Component B is estimated from another sample selected separately either for the purpose of the elementary units check or for this purpose combined with the quality check of the response. In this case the third term on the right-hand side of (6.13) drops out automatically.

Another reason for the separation of the two checks is the following. In population censuses, for example, the main difficulty of any check of listing is connected with population changes. The longer the lapse of time between the census enumeration and the moment of checking, the less chance there is of establishing which persons were omitted in census lists and which were erroneously enumerated.

In order to eliminate or reduce the difficulties arising therefrom, cluster listing checks could be made as a separate survey as soon as the main enumeration is over. As for the elementary units check and the quality check of the response, it might be possible and more convenient to make them at a later date. Such an arrangement might be particularly convenient if the checking of the quality of the response has to be carried out by specially trained staff who are selected from among census supervisors.

TABLE C. - EFFECT OF LISTING ERRORS ON THE NUMBER OF HOUSEHOLDS AND THE NUMBER OF PERSONS

Item	Households		Persons	
	Number	Percent	Number	Percent
A. Cluster check				
1) Census (Number)		—		—
2) Agreements				
3) Erroneously included				
4) Omissions				
5) Bias				
6) Squared EDs' biases		—		—
7) Significance test		—		—
B. Elementary units check				
8) Census	—	—		—
9) Agreements	—	—		
10) Erroneously included	—	—		
11) Omissions	—	—		
12) Bias	—	—		
13) Squared households' biases	—	—		—
14) Significance test	—	—		—
C. Total				
15) Bias				
16) Significance test		—		—

Data obtained in a quality check of listing based on Type II design cannot be presented in Tables A and B. In this case there are two bias sources, i.e., errors in listing households and errors in listing persons. For this reason the effects of the listing errors of each type have to be shown separately. It is also necessary to combine the two components of the bias.

In designing the check it is useful to have a self-weighting sample. Otherwise the results obtained in the check cannot be presented in an easy way. In fact, the weighting needed may sometimes lead to considerable computations.

Let us now assume that a check survey of the quality of listing in a population census was made with EDs selected for the cluster check and households for the elementary units check. The sample is self-weighted. The results obtained may be presented as in Table C. All the absolute numbers presented in parts A and B of this table *refer to units included in the sample.* Thus, the number of households in line 1 is the number listed in the census in the EDs selected in the sample. Similarly, all the persons listed in the census in these households appear in the same line under "Persons." The number of persons appearing in line 8 refers to the persons listed in the census in the households selected in the sample. In the computation of percentages in lines 2 through 5, the figures in line 1 are used as a basis. The percentages in lines 9 through 12 take the figure in line 8 as a basis. The bias in line 5 is equal to $\overset{m}{\Sigma} D_i$. It is obtained by subtracting the figure in line 4 from the figure in line 3. The figure in line 6 is $\overset{m}{\Sigma} D_i^2$, which is needed in the computation of an approximate value of the variance according to the equation (6.7). In line 7 the quantity $\Sigma D_i / \sqrt{\Sigma D_i^2}$ is presented. The bias in line 5 may be considered significant if the figure in line 7 is larger than two.

In part B of the table the figure in line 12 is equal to $\underset{i}{\Sigma} \underset{j}{\Sigma} D_{ij}$ where D_{ij} is the listing bias of the j-th household in the i-th ED. Line 13 shows $\overset{m}{\underset{i}{\Sigma}} (\underset{j}{\Sigma} D_{ij})^2$ while line 14 contains $\dfrac{\underset{i}{\Sigma} \underset{j}{\Sigma} D_{ij}}{\sqrt{\underset{i}{\Sigma} (\underset{j}{\Sigma} D_{ij})^2}}$. The significance of the bias is tested as above in the case of the cluster check.

The total bias in line 15, i.e., D', is estimated as

$$D' = \frac{M}{m} \left(\Sigma D_i + \frac{1}{f_2} \Sigma \Sigma D_{ij} \right) \tag{6.14}$$

where f_2 is the fixed sampling fraction of households in the selected EDs.

Line 16 contains the ratio

$$\frac{D'}{s_{D'}} = \frac{D'm}{M} \left[\overset{m}{\Sigma} D_i^2 + \left(\frac{1}{f_2} \right)^2 \overset{m}{\underset{i}{\Sigma}} (\underset{j}{\Sigma} D_{ij})^2 \right]^{-\frac{1}{2}} \tag{6.15}$$

It is clear that the above table offers the chance to study many aspects of the quality of listing in a census of population.

6.7 Some illustrations

In this section some illustrations of the use of sampling methods for checking the quality of listing in various censuses will be presented as well as some useful experiences.

The simplest examples of listing checks based on Type I design are to be found in the 1953 and 1961 Censuses of Population in Yugoslavia.[5] In both censuses the average size of the EDs was about 250 persons. For this reason a sample of EDs was selected and the supervisors were responsible for relisting all the households as well as the members of the households listed. A special local group was made responsible for matching the census and the check and reconciling the discrepancies. Only a slight inaccuracy of listing was found in both cases.

Another simple case of the use of Type I design is to be found in the 1946 Census of Industrial and Commercial Establishments in France.[6] A considerable listing bias was detected in this census and the results found in the check were useful in preventing many incorrect uses of the information collected in the census.

Other similar listing checks were conducted in connection with the United States Censuses of Agriculture of 1950,[7] 1954,[8] and 1959.[9] The sample used consisted of area segments selected at random from the rural part of the United States. In 1950 there were 1,000 such segments scattered among 270 counties or groups of counties. In 1954 there were 772 segments spread over 319 counties and containing approximately 2,800 farms. The check was conducted by 60 specially selected and trained enumerators

---

[5] For the 1953 census, cf. Zarkovich, S.S.: Sampling methods in the Yugoslav 1953 Census of Population, *Journal of the American Statistical Association*, Vol. 50, 1955, pp. 720-737 and for the 1961 census, Pirochanac, M.: Control of the completeness of the Census of Population of 31.3.1961 (in Serbo-Croatian), *Statistička Revija*, Vol. 12, 1962, pp. 30-40.

[6] Chevry, G.: Control of a general census by means of an area sampling method, *Journal of the American Statistical Association*, Vol. 44, 1949, pp. 373-379.

[7] U.S. Bureau of the Census: *U.S. Census of Agriculture: 1950*, Vol. II, General Report, U.S. Government Printing Office, Washington, D.C., 1952.

[8] U.S. Bureau of the Census: *U.S. Census of Agriculture: 1954*, Vol. III, Part 12, Methods and procedures, U.S. Government Printing Office, Washington, D.C., 1956.

[9] U.S. Bureau of the Census: *U.S. Census of Agriculture: 1959*, Vol. II. General Report, U.S. Government Printing Office, Washington, D.C., 1962.

Table 14. - Estimates of underlisting for number of holdings and selected items for the United States: 1950 and 1954 Censuses of Agriculture

Census and items		Estimated total (000)	Reported in the Census		Estimated underenumeration				
							Sampling error of estimate		
			Amount (000)	Percent of estimated total	Amount (000)	Percent of estimated total	Amount (000)	Percent of estimated total	Percent of estimated under-listing
1950									
Farms	number	5 656	5 382	95.2	274	4.8	36	0.6	13.1
Area of farms	acres	1 209 601	1 159 789	95.9	49 812	4.1	13 917	1.2	27.9
Cropland harvested	acres	363 022	345 528	95.2	17 494	4.8	3 801	1.0	21.7
Corn harvested	acres	83 632	83 351	99.7	281	0.3	750	0.9	267.0
Wheat harvested	acres	74 219	71 161	95.9	3 058	4.1	1 850	2.5	60.5
Cotton harvested	acres	28 646	26 599	92.9	2 047	7.1	1 037	3.6	50.7
1954									
Farms	number	5 201	4 782	91.9	419	8.1	49	0.9	11.7
Area of farms	acres	1 223 891	1 158 192	94.6	65 699	5.4	22 798	1.9	34.7
Cropland harvested	acres	346 580	332 870	96.0	13 710	4.0	3 907	1.1	28.5
Corn harvested	acres	80 886	78 123	96.6	2 763	3.4	950	1.2	34.4
Wheat harvested	acres	54 263	51 362	94.7	2 901	5.3	2 248	4.1	77.5
Cotton harvested	acres	19 026	18 854	99.1	172	0.9	286	1.5	166.0

who had to make independent visits to all the potential units, identify the holdings in the segments selected and collect from them the census information. To help enumerators to fulfill their tasks in the best possible way, aerial photographs were supplied to them of the segments where they had to introduce all the holdings identified together with fields operated by these holdings. The questionnaires filled in during the field visit were sent to the central office to be matched with the census questionnaires. Any major discrepancy was rechecked by a specialized person in order to determine the conditions under which the difference occurred.

The results obtained in these checks are presented in Table 14.[10] The meaning of the figures in this table is clear enough. The first line shows the degree of underlisting in the total number of holdings. The other lines show the effects of listing errors on totals for various selected crops. The figures in the last column help to determine the significance of the estimated bias. The bias may be considered significant if the standard error is less than half the size of the bias. It can be seen that most of the biases are significant.

Data in this table are very instructive. They show that listing errors primarily affect small holdings. Totals for various characteristics in the survey program are less affected than the number of units. This is due to the fact that smaller holdings are more frequently omitted from the census than large ones.

Another interesting illustration is the listing check of the 1956 Census of Agriculture in Canada.[11] For this check a sample of 601 area segments was selected and all the holdings with headquarters within the boundaries of the selected segments were relisted. The check was carried out by 200 enumerators. Matching lists were not made and in the estimation procedure the only data available were the totals established separately in the census and in the check. Because of the high correlation between the quantities involved it was possible to arrive at sufficiently precise estimates of the listing bias. Some results from this check are presented in Table 15. The biases presented in this table represent underlisting, as is the case in most censuses. For the ratio in the last column the reader is referred to the equation (6.8). In judging the significance of the bias it must be remembered that all the biases presented here are negative so that half the area under the normal curve can be used.

TABLE 15. - SOME RESULTS OF THE QUALITY CHECK OF LISTING IN THE 1956 CENSUS OF AGRICULTURE IN CANADA [1]

Item	Listing bias (percent of census total)	Ratio: bias to standard error
Holdings	7.6	5.2
Total area	3.8	0.6
Oats for grain	3.5	1.8
Barley	8.4	3.4
Potatoes	12.8	1.7
Pigs	10.7	3.9
Sheep	3.0	0.5
Cattle	7.1	1.6
Cows, milking	2.2	1.1
Hens and chickens	1.9	0.5

[1] Dominion Bureau of Statistics: Op. cit., p. 58.

More illustrations are available from population censuses where listing checks are widely used because of the importance of obtaining accurate information about the number of persons. The techniques used in checking the quality of listing in population censuses can be classified in two broad groups. In the first group fall checks conducted in the form of a sample survey made after the census enumeration has been completed and with the specific purpose of estimating the quality of census results. This is basically the same procedure as the one in the censuses of agriculture mentioned earlier. Two interesting cases from this group will be presented here. They are checks of listing errors in the 1950 Census of Population in the United States and in the 1951 Census of Population in India. The second group consists of checks based on a comparison of census data with similar information available in surveys conducted independently of the census itself. Examples from this group are drawn from the 1956 Census of Population in Canada and the 1955 Census of Population in Japan. In these two countries current labor force surveys are conducted at frequent intervals. Data from these surveys were compared with figures relating to the population living in the areas selected for matching.

In checking the quality of listing of the 1950 United States Census of Population, Type II design was used. The check was split up into two parts, (*i*) a measure of errors in the listing of dwelling units and (*ii*) a mea-

---

[10] All these tables are reproduced from Bureau of the Census: *U.S. Census of Agriculture: 1954*, Vol. II, General Report, Introduction, U.S. Government Printing Office, Washington, D.C., 1956.

[11] Details about this check are found in *Census of Canada, 1956*: *Administrative Report*, Dominion Bureau of Statistics, Ottawa, 1958.

sure of errors in the listing of persons within the households listed in the census. The design adopted provided for two samples. The first was called a *segment sample* and the other a *list sample*. The segment sample was a stratified multistage area sample. From the United States as a whole a sample of 276 primary units was selected with probabilities proportional to size as expressed in terms of the population in the 1940 Census of Population. Within the selected primary units segments were delineated with an approximate size of six dwelling units in urban areas and ten dwelling units and five farms in rural areas. In the urban zone 2,800 segments were selected in the second stage of selection and about 1,000 in the rural zone. The field staff employed in this check had to relist all the dwelling units in the segments selected. To improve the quality of their work a special map or aerial photograph was prepared for each segment. In addition to this, for most segments the enumerators were given the corresponding census results, i.e., lists of persons, farms and dwelling units.

The list sample was selected from the census lists of dwelling units. In urban areas this sample overlapped with the segment sample; in rural areas it was selected from the same primary units but was independent of the corresponding segment sample.

Some results of this check are presented here in Tables 16 and 17.[12] Data in Table 16 show the magnitude of the bias in census totals for various areas. Table 17 gives an analysis of the errors detected. It provides further information on cases defined as "omitted" and "erroneously included." In this connection the text in the corresponding census publication says: "In interpreting the figures on erroneous omissions and erroneous inclusions it should be recognized that these are defined with respect to the listing for a given enumeration district. For example, some of the 'omitted' cases represent the listing of a person in the wrong census enumeration district rather than his complete omission from the census. Such cases will be included in *both* the estimate of erroneous omissions and the estimate of erroneous inclusions (since such persons enumerated in the wrong census enumeration district are both omitted from the listing where their names should appear and included in a listing where their names should not appear). In the absence of duplicate enu-

---

[12] These two tables are reproduced from U.S. Bureau of the Census: *U.S. Census of Population: 1950*, Vol. II, *Characteristics of the Population*, Part 1, U.S. Summary, U.S. Government Printing Office, Washington, D.C., 1952.

TABLE 16. - LISTING ERRORS FOR PERSONS BY REGIONS AND SIZE OF PLACE IN THE 1950 UNITED STATES CENSUS OF POPULATION [1]

Area	Census population (000)	Estimated total population (000)	Persons omitted[2]		Persons erroneously included[3]		Net error	
			Number (000)	Number per 100 enumerated	Number (000)	Number per 100 enumerated	Number (000)	Percent of estimated total population
United States	150 697	152 788	3 400	2.3	1 309	0.9	2 091	1.4
Northeast	39 478	39 794	732	1.9	416	1.1	316	0.8
North central	44 461	45 064	813	1.8	210	0.5	603	1.3
South	47 197	48 071	1 381	2.9	507	1.1	874	1.8
West	19 562	19 861	476	2.4	177	0.9	299	1.5
Urban	96 468	97 504	1 928	2.0	892	0.9	1 036	1.1
Places of 1,000,000 and over	17 404	17 634	477	2.7	247	1.4	230	1.3
Places of 50,000 to 1,000,000	35 839	36 255	662	1.8	246	0.7	416	1.1
Other	43 226	43 615	789	1.8	400	0.9	389	0.9
Rural	54 230	55 285	1 472	2.7	417	0.8	1 055	1.9

[1] U.S. Bureau of the Census: *U.S. Census of Population: 1950*, Vol. II, *Characteristics of the Population*, Part 1, U.S. Summary, U.S. Government Printing Office, Washington, D.C., 1952.

[2] Includes some persons who were counted elsewhere, at the wrong address.

[3] Includes some persons who were counted only once but at the wrong address.

TABLE 17. - NUMBER OF PERSONS OMITTED FROM, OR ERRONEOUSLY INCLUDED IN, CENSUS ENUMERATION DISTRICT LISTING IN THE 1950 UNITED STATES CENSUS OF POPULATION [1]

Type of error	Total	
	Number (000)	Percent of enumerated population
Persons omitted[2]	3 400	2.3
In omitted households and quasi households	2 416	1.6
In enumerated households and quasi households	984	0.7
Persons erroneously included[3]	1 309	0.9
Persons who should not have been enumerated anywhere:	198	0.1
in households erroneously included	38	
in households properly included	160	0.1
Persons who should have been enumerated in another enumeration district:	1 111	0.7
in households erroneously included	99	0.1
in households properly included	1 012	0.7

[1] U.S. Bureau of the Census: *U.S. Census of Population: 1950*, Vol. II, *Characteristics of the Population*, Part 1, U.S. Summary, U.S. Government Printing Office, Washington, D.C., 1952.

[2] Includes some persons who were counted elsewhere, at the wrong address.

[3] Includes some persons who were counted only once but at the wrong address.

meration these cases do not affect the net error. They do, however, affect the other values estimated..."[13]

It can be seen from this table that a considerable proportion of the persons omitted is to be found in omitted households while the majority of the persons erroneously included are people who should have been enumerated in another ED.

It is clear that the possibility of some people being enumerated in several EDs represents a crucial consideration in such checks. If a person were erroneously included in the population of an ED and not enumerated accurately anywhere else, the case, if detected in the quality check, would represent an error of overlisting according to the procedures adopted

---

[13] Op. cit.

in checking. As such, the case affects the magnitude of the listing bias. If, in addition to this, the same person is found to be omitted in another ED he or she will be included in the check as an omission at that particular place. If the two cases appear jointly, they will not affect the size of the bias. It is therefore useful to know whether persons included in the check at one place are also taken into account in other EDs and in what way.

In order to obtain more information on such cases the persons included in the sample " ... were asked to report all addresses where they might have been enumerated. On the basis of a check against the listings for the census enumeration districts containing the reported addresses, it is estimated that about 400,000 persons were enumerated in the wrong enumeration district. The number is subject to some bias due to incompleteness of the address reports. Some persons were enumerated in the wrong enumeration district simply because the enumerator used the wrong boundary. The estimate of 400,000 may be subtracted from the estimates of erroneous omissions and erroneous inclusions if interest is restricted to those errors which affect the census tabulations for the United States as a whole. Actually all the persons enumerated in the wrong enumeration district were enumerated in the correct region and most were enumerated in the correct state, so that this group of errors has practically no effect on either national or regional tabulations and an extremely small effect on state tabulation."[14]

A very similar procedure was used in the 1960 United States Census of Population.[15] In checking the accuracy of listing in the 1951 Census of Population in India Type II design was used. In rural areas a systematic sample of blocks was selected with a sampling fraction of 1 in 100. At

---

[14] Op. cit. More details on this check are available in Eckler, A.R.: Extent and character of errors in the 1950 census, *The American Statistician*, Vol. 7, No. 5, 1953, pp. 15-21; Eckler, A.R. and Pritzker, L.: Measuring the accuracy of enumerative surveys, *Bulletin of the International Statistical Institute*, Vol. 33, Part 4, 1951, pp. 7-24; Hansen, M.H., Hurwitz, W.N. and Pritzker, L.: The accuracy of census results, *American Sociological Review*, Vol. 18, 1953, pp. 416-423; Marks, E.S., Mauldin, P.W. and Nisselson, H.: The post-enumeration survey of the 1950 census: a case history in survey design, *Journal of the American Statistical Association*, Vol. 48, 1953, pp. 220-243.

[15] Further details on the procedure used and the results obtained are available in Taeuber, C. and Hansen, M.H.: A preliminary evaluation of the 1960 censuses of population and housing, *Proceedings of the Social Statistics Section, American Statistical Association*, 1963, pp. 56-73; Steinberg, J., Gurney, M. and Perkins, W.: The accuracy of the 1960 census count, *Proceedings of the Social Statistics Section, American Statistical Association*, 1962, pp. 76-79; and Hansen, M.H., Pritzker, L. and Steinberg, J.: The evaluation and research program of the 1960 censuses, *Proceedings of the Social Statistics Section, American Statistical Association*, 1959, pp. 172-180.

the second stage every tenth household was selected for checking the listing of persons. In urban areas the corresponding fractions were 1 in 20 and every fiftieth household. The selection of households was based on census lists. Listing errors of persons were checked by matching census results with data obtained in a re-enumeration of all the persons in the households selected. Households omitted were checked by a device whereby the enumerator had to designate the three occupied houses nearest to the household selected. It was later seen whether these houses were included in the census. The check showed a net underlisting of 1.1 percent of the total population.[16]

The same group includes checks conducted in connection with the 1953 Census of Population in Ceylon where an error of underlisting of 0.7 was found.[17]

From the group of checks which was not based on special surveys conducted in the postcensus period our first example is the check carried out in Canada in connection with the 1956 Census of Population.[18] The check was based on data available in the Canadian Labor Force Survey. This survey was made in May 1956 while the census was taken in June of the same year, so that a short time elapsed between them. As before, a sample of census EDs was selected and data of the census and the Labor Force Survey referring to these areas were matched with a view to determining their agreement. If no matching census document were available or a discrepancy appeared between the two surveys in the composition of some matched households, a re-enumeration was undertaken in order to determine whether households or persons without a matching document in either the census or the survey had actually been missed, and if so, for what reason.

The summary of the results obtained is presented in Table 18. It will be seen from this table that omissions prevail again in this case. It will also be noticed that the magnitude of the bias is somewhat larger in urban than in rural areas. It is also interesting to note how large a contribution to the number of persons omitted comes from the households that were missed in the census. In other words, it is of prime importance in censuses of population to secure an accurate listing of households.

---

[16] Cf. Registrar General: Sample verification of the 1951 census count, census of India, Paper No. 1, New Delhi, 1953.

[17] Further details on other checks will be found in: United Nations. *Demographic Yearbook, 1956*. New York, 1956.

[18] Dominion Bureau of Statistics: *Census of Canada, 1956, Administrative Report*. Queen's Printer and Controller of Stationery, Ottawa, 1958.

TABLE 18. - LISTING ERRORS IN THE 1956 CENSUS OF POPULATION IN CANADA [1]

Item	Total		Self-representing urban areas		Other urban areas		Rural areas	
	Number	Percentage	Number	Percentage	Number	Percentage	Number	Percentage
Total number of persons enumerated	144 470	100.0	70 523	100.0	25 866	100.0	48 081	100.0
Persons matched	128 569	89.0	61 454	87.1	23 223	89.8	43 892	91.3
Persons in the survey, not matched in the census	7 318	5.1	4 303	6.1	1 135	4.4	1 880	3.9
Persons in the census, not matched in the survey	8 583	5.9	4 766	6.8	1 508	5.8	2 309	4.8
Persons in households completely omitted in the census (before re-enumeration)	4 188	2.9	2 466	3.5	716	2.8	1 006	2.1
Persons omitted (after re-enumeration)	2 427	1.7	1 438	2.0	382	1.5	607	1.3
Persons omitted in the census in partially matched households (before re-enumeration)	3 130	2.2	1 837	2.6	419	1.6	874	1.8
Persons omitted (after re-enumeration)	1 662	1.2	1 023	1.5	241	0.9	398	0.8
Persons underenumerated								
Males	2 076	1.4	1 225	1.7	313	1.2	538	1.1
Females	2 013	1.4	1 236	1.8	311	1.2	466	1.0
Total	4 089	2.8	2 461	3.5	624	2.4	1 004	2.1
Persons overenumerated								
Males	1 054	0.7	639	0.9	147	0.6	268	0.6
Females	950	0.7	574	0.8	158	0.6	218	0.5
Total	2 004	1.4	1 213	1.7	305	1.2	486	1.0
Correction factor	574	0.4						
Actual overenumeration	1 430	1.0						
Adjustment for known underenumeration	1 093	0.8						
Estimated net underenumeration	1 566	1.1						

[1] Dominion Bureau of Statistics: *Census of Canada, 1956, Administrative Report*. Queen's Printer and Controller of Stationery, Ottawa, 1958.

A very similar check was conducted in Japan in connection with the 1950 Census of Population. This check was also based on data collected in the Monthly Labor Force Survey. Comparisons were made in 85 EDs for persons of 14 years and over who were included in the Labor Force Survey. The census enumerated 4,710 persons, among whom 4,628 were identified in the Labor Force Survey data while 82 were not identified. On the other hand, the Labor Force Survey enumerated another 42 persons, 7 of whom were found in the census reported as being under 14 years of age. This gives a total underlisting error of 0.74 percent.[19]

6.8 Summary of experiences

General experience from checks conducted in the past shows that listing errors primarily appear as a result of *borderline difficulties*, the concept of the border line being used in a broad sense. It is difficult for the enumerators to decide where borderline units such as houses belong, and where their inhabitants should be enumerated. Small agricultural holdings or those which are on the qualifying border line are more frequently missed than the others because they are not easy to distinguish or because it is not clear whether they belong to the population or not. In a population census children, guests, visitors, etc. are often missed and can also be considered as borderline cases. Similar experiences are also found in housing censuses where borderline cases, such as dwelling units that are conversions of nonresidential buildings, are frequently omitted.

For illustration purposes, it may be useful to point out that listing errors primarily affected the following classes of people in the 1950 Census of Population in the United States:

"Non-whites. The over-all rate of net deficiency of non-whites (3.3 percent) is more than two and one-half times the rate for the white population. The rates of net error are greatest for non-whites in rural non-farm households, in the Southern States, in urban areas, and among children under 10 years of age.

Lodgers. The estimated rate of net deficiency for male lodgers was 9.6 percent. The rate for female lodgers was estimated to be 6.2 percent. These rates are among the highest of the classes examined.

---

[19] Morita, Y.: An appraisal of the population census statistics, *Bulletin of the International Statistical Institute*, Vol. 34, Part 3, 1954, pp. 189-201.

Persons in rural non-farm households. The rates of net omissions are estimated to be 9.6 percent for non-whites and 2.3 percent for whites in rural non-farm households.

Persons in low socio-economic classes. The estimates of net underenumeration are 2.0 percent or greater for adults with no formal education, for male 'laborers except farm and mine', and for adults with income under $500."[20]

In the quality check of listing in the 1961 Census of Population in Greece it was found that the largest percentage of households omitted was in one-person households. This is in line with the above experience because these are borderline cases and there may be some doubt as to whether they should be considered as separate households or not. It was also found in the same study that the majority of the persons omitted or incorrectly included in the census lists were people who did not sleep at home during the census night.[21]

Similar results were obtained from a study of the listing of infants in the 1950 United States Census of Population, where it was found that reasons for their omission are: (*i*) death after the census day and prior to census enumeration, (*ii*) temporary absence from parents' house during the enumeration, (*iii*) census information supplied by neighbors, (*iv*) the enumerator's impression that the infant was too young to be enumerated, (*v*) the family's impression that infants should not be enumerated, etc.[22]

Another useful deduction from past studies is that listing errors for some kinds of units result chiefly from erroneous listing of units of a higher order in the hierarchy of units. In the above study of the quality of the listing of infants, 82 percent of omissions are due to the omission of the corresponding household. In studies conducted by the United States Bureau of the Census in 1948 and 1950 it was found that the majority of dwelling unit omissions was due to the whole structure being omitted.[23]

From studies conducted in Greece it was found that the greatest difficulty of area and yield statistics (insofar as areas are concerned) lies in

---

[20] Hansen, M.H. and Pritzker, L.: The post-enumeration survey of the 1950 Census of Population: some results, evaluations and implications, *Annual Meeting of the Population Association of America*, Ann Arbor, 1956.

[21] National Statistical Service: The coverage check of the 1961 Population Census of Greece, Athens, 1962.

[22] U.S. Bureau of the Census: *Infant Enumeration Study, 1950*, Washington, D.C., 1953.

[23] Eckler, A.R. and Pritzker, L.: Measuring the accuracy of enumerative surveys, *Bulletin of the International Statistical Institute*, Vol. 33, Part 4, 1951, pp. 7-24.

the high proportion of omissions in listings of fields based on farmers' reports.[24]

The results of the 1950 Census of Population in the United States in this respect can be seen from the following quotation: "Almost three quarters of the people missed in the census (gross omission) are omitted because the census enumerators failed to find or list dwelling units in which the people resided. If the dwelling unit coverage in the 1950 Census has been complete, the PES estimate of the number of erroneously omitted persons would have been reduced to approximately one million, and the estimated net deficiency in the census count would have been close to zero."[25]

This information is useful in designing surveys because it shows the need to ensure that omission of the units of a higher order is made impossible or nearly so. It can be exploited in a number of other ways. For example, in agricultural statistics, fields are sometimes listed on the basis of farmers' reports and a sample of fields is then selected from such lists and measurements are taken for area and yield data. The experience presented above may show the inadvisability of carrying out expensive measurements if there is a danger of the frame being highly biased.

However, the most general and important conclusion from these experiences is that the quality of listing in population censuses largely depends on the quality of the listing of households or dwelling units. If measures are taken to secure accurate lists of clusters, the effect of erroneous listing of elementary units will normally be negligible.

The next general conclusion to be drawn from listing checks is that units which are subject to listing errors are normally different from the rest of the population. The agricultural holdings omitted are primarily small ones or those which consist of a large proportion of wasteland. In the census of manufacturing establishments, the units omitted have a considerably lower number of employees than the population average.[26] In the listing check of the 1948 United States Census of Business the average 1948 sales of enumerated establishments amounted to $74,000 as against an average of $26,000 for the establishments omitted.[27]

This information is again useful in designing surveys. If census lists

---

[24] National Statistical Service: Farmers reporting at the census, mimeographed report, Athens, 1962.

[25] Hansen, M.H. and Pritzker, L.: Op. cit.

[26] Eckler, A.R. and Pritzker, L.: Op. cit.

[27] Ibid.

are used as a frame it should be understood that listing errors will, in most cases, lead to underestimation.

Another useful experience is that listing errors may be so large as to make some uses of data impossible. For example, Coale has found that the difference in underlisting between whites and nonwhites is large enough to account for approximately two years in the expectation of life at the age of five years.[28]

6.9 Problems in checking the quality of listing

The purpose of this section is to illustrate difficulties that arise in the stages both of designing and of carrying out listing checks. Our illustrations are based on the experience gained in the listing check of the 1953 Census of Population in Yugoslavia.[29] In this census the enumerators were asked to identify the units of the population and classify them in one of the three classes "permanent resident," "temporary absentee," and "temporary resident." The first two classes constitute what is called "resident population." The number of inhabitants to be published for various administrative units was conceived of as the resident population of these units.

For most persons subject to the check it was easy to carry out census definitions. These were called "clear" cases. In any census, however, there are borderline cases which cause difficulties. Students are an example in point. If studying away from home they had to be enumerated as "temporary absentees" in their parents' ED provided they had no job in the place where they were studying. If working, they had to be enumerated as "permanent resident" in the ED where they were employed during the census and not enumerated at all at the address of their parents.

The procedure was adopted for a number of reasons which are immaterial here. Its consequences are what concern us. In this check it was found that parents very often did not know anything about the possibility of their student children having a job. This was particularly so in villages.

---

[28] Coale, A.J.: The population of the United States in 1950, classified by age, sex and color – A revision of census figures, *Journal of the American Statistical Association*, Vol. 50, 1955, pp. 16-54.

[29] This presentation is based on Zarkovich, S.S.: Some remarks on coverage checks in population censuses, *Population Studies*, Vol. 9, 1956, pp. 271-275.

The parents showed a tendency to present their information in a way that classified the students as "temporary absentees" in their parents' EDs.

It was clear on the basis of the first experiences that data collected in the check about students and other persons of similar characteristics were not sufficient from the listing point of view. The check will agree with the census if both are based on the same information. The agreement, however, does not mean that the listing is accurate. The information given by parents may be wrong. In order to have further guidance on the quality of the information collected in the check and the nature of accurate listing, it was decided to ascertain whether persons included in the check were enumerated in the census at some other place and, if so, in what way. If the additional information was consistent with the information obtained originally in the check it would strengthen the conclusions based on the check. Otherwise it would offer an insight into the frequency of doubtful cases and the margin of uncertainty associated with the check.

In order to achieve this aim it was decided to ask each person in the check for the addresses of places where he or she was staying around the time of the census. The census questionnaires corresponding to the addresses given were then looked up with a view to ascertaining whether the persons involved were listed. This additional attempt to improve the quality of the check greatly increased both the volume of work and the costs. The addresses involved in the check normally belonged to a large number of EDs spread over the country as a whole.

Obviously, this supplementary effort does not ensure the accuracy of check data. If a student is found to be enumerated as "temporary absentee" at his parents' home and "temporary resident" at the place of studies, the two responses will be consistent although not necessarily accurate. The student may have a job, which is a fact that calls for a completely different listing. In other words, looking up the census listings at the addresses indicated is not a sufficient measure. In order to improve the quality of the check it may be desirable to see the person involved and ascertain in an interview which is the accurate listing for the case. In other words, in addition to the usual relisting and the study of census material at different addresses, it may be desirable to embark upon special field visits to collect additional useful information.

It should be added that students were not the only type of person who presented difficulties and complications of this kind in the check. The problem was even greater in the case of persons working in one place, having families in another, and commuting from time to time from one place to the other.[30] In addition, there was difficulty resulting from a housing shortage. Rationing of housing space had already been abolished at that time but there was still a tendency to declare an increased number of residents with the aim of using the census as a possible justification of the housing space occupied.

Bearing these difficulties in mind, the following experiment was made. The first stage of the listing check consisted of the usual relisting. After the results of the relisting had been matched with the corresponding census information, the distribution of data was obtained as in Table 19. In fact, this stage of the check consisted of the relisting of 249 EDs.

TABLE 19. - CLASSIFICATION OF THE RESULTS OBTAINED IN RELISTING PERSONS IN A SAMPLE OF ENUMERATION DISTRICTS

Check Census	Permanent resident	Temporary absentee	Temporary resident	Not enumerated
	(α)	(β)	(γ)	(δ)
Permanent resident (A)	$A\alpha$	$A\beta$	$A\gamma$	$A\delta$
Temporary absentee (B)	$B\alpha$	$B\beta$	$B\gamma$	$B\delta$
Temporary resident (C)	$C\alpha$	$C\beta$	$C\gamma$	$C\delta$
Not enumerated (D)	$D\alpha$	$D\beta$	$D\gamma$	—

Any person listed either in the census or in the check will appear in a cell of this table. Some of the combinations presented are unlikely in practice, such a $C\alpha$ or $A\gamma$. The interpretation of the results in such a table is simple if it is assumed that the information obtained in the check

---

[30] The same was found to be a major difficulty in the 1950 census in Germany. Cf. United Nations: Evaluation of quality of demographic statistics, Krieger, K. *Proceedings of the United Nations World Population Conference, 1954*, Vol. 4, New York, pp. 133-147.

is accurate. For interpretation purposes we shall establish the following four classes:

$$\begin{aligned} I &= A\alpha + A\beta + B\alpha + B\beta \\ II &= C\alpha + C\beta + D\alpha + D\beta \\ III &= A\gamma + A\delta + B\gamma + B\delta \\ IV &= C\gamma + C\delta + D\gamma \end{aligned}$$

Class I consists of the resident population according both to the census and to the check. Class II represents the additional resident population in the check that does not appear in the census resident population. This class constitutes the underlisting, and class III the overlisting. Finally, class IV has no bearing on the resident population either in the census or in the check. Assuming now that such a table is available for all the EDs included in the check so that class totals I_i, II_i, etc. are known for each ED, we could put

$$\begin{aligned} X_i &= I_i + II_i \\ Z_i &= I_i + III_i \end{aligned}$$

and estimate the true resident population for the region involved. This would lead to an estimate of the resident population based on relisting.

In order to supplement the information in Table 19 with data resulting from a study of census material, the addresses of persons in a limited number of cells were collected. In fact, the study covered the persons not listed in the census (cells $D\alpha$, $D\beta$ and $D\gamma$), those listed as "permanent resident" in the census and "temporary resident" in the check (cell $A\gamma$), and one single person who was listed as "temporary resident" in the two surveys (cell $C\gamma$), which is a rather strange combination. A check was made in the census material under the addresses thus obtained to see whether these persons were enumerated there and if so, how.

The results obtained are presented in Table 20. Not a single case was found of the same person being enumerated at more than two EDs. A number of combinations was found, however, which gives an indication of how complex this work is and how dangerous it might be to accept as accurate the results of the check as they emerge from relisting. In the cell $D\alpha$, out of 81 cases in the urban stratum 73 were found to be real omissions. Eight persons, however, were found to be enumerated as "permanent residents" elsewhere. Cell $D\gamma$ presents a discouraging example. Twenty-nine cases in the urban stratum are broken down into all the possibilities available.

TABLE 20. - RESULTS OF A CHECK TAKEN TO DETERMINE WHETHER AND, IF SO, HOW 285 PERSONS WERE ENUMERATED IN THE CENSUS IN EDs DIFFERENT FROM THOSE WHERE THEY WERE FOUND IN THE QUALITY CHECK OF LISTING
(1953 Census of Population, Yugoslavia)[1]

Cell and stratum (U = urban R = rural)		Number of cases	Not enumerated elsewhere	Enumerated elsewhere as		
				Permanent resident	Temporary absentee	Temporary resident
Dα	U	81	73	8	—	—
	R	24	21	2	1	—
Dβ	U	34	11	2	3	18
	R	12	2	4	—	6
Dγ	U	29	5	10	11	3
	R	11	2	4	4	1
Aγ	U	78	15	3	50	10
	R	15	2	—	13	—
Cγ	U	1	—	1	—	—
	R	—	—	—	—	—
Total	U	223	104	24	64	31
	R	62	27	10	18	7

[1] This table is reproduced from Zarkovich, S.S.: Some remarks on coverage checks in population censuses, *Population Studies*, Vol. 9, 1956, p. 274.

If results of the type shown in Table 20 were available for more cells, they might be used to improve conclusions based solely on Table 19. Out of 81 persons in the first row of Table 20, 73 are real omissions while the remaining 8 appear in a district which is different from the one where they should have been enumerated according to the check. In estimating the volume of real omissions, i.e., the number of persons not enumerated at all, these 8 persons should be deducted and the remaining 73 only taken into account. A similar thing could be done in other cells.

The results in Table 20 obviously call for a third-stage check, i.e., a new visit to the field with the aim of getting more data on certain persons. If 73 cases in the first row are considered as omissions the remaining 8 persons call for further clarification. They were enumerated as "permanent residents" both by the census and by the check, but in different EDs. One of the two results must be inaccurate. Only a thorough study of these cases on the spot could clarify whether the census or the check information is inaccurate.

Unfortunately, the experiment was discontinued at the stage when the need for a further field visit had become obvious. The work was abandoned because several months had already been spent in the earlier stages of the work. It was therefore considered that the persons involved would not be able to offer the information needed after such a long delay. In addition, because the persons falling in the cells checked were highly mobile, it was thought that in most cases they could not be contacted during the holiday season.

In spite of this, data in Table 20 are useful as a warning against the opinion that it is easy to secure accurate check data.[31]

6.10 Measures for improving the quality of listing

The last topic that we propose to discuss consists in the measures which can be used to reduce the number of listing errors. Several measures will be listed. Of course, this does not mean that some other means cannot be found under given circumstances.[32] These measures are:

1. *Preparing or improving the mapping material.* It is a common experience in many countries that listing errors are often due to difficulties in identifying the borders of EDs. This results either in omissions or in duplications because two or three enumerators may think that a borderline unit forms part of their assignment. Many such errors will be removed if some way is found of preparing maps or sketches where boundaries are marked along some clearly distinguishable natural objects, thus facilitating the identification of the areas they delimit.

Some difficulties may arise owing to the fact that the delineation of EDs on maps usually requires a field visit. Maps go out of date rather quickly, new structures appear, natural objects sometimes disappear and only an inspection on the spot can guarantee that a boundary marked will be clearly recognizable.

Mapping material is therefore an expensive item, requiring large-scale and long-term preparations.

2. *Selection and training of enumerators.* There is hardly any need to point out the importance of adequate training of enumerators. The enumerators have to be given a specific explanation of the types of listing errors which usually appear, the circumstances under which they appear, and the measures they are supposed to take to avoid errors. General statements are of no use in this training. The enumerators have to be shown how to canvass the district to secure the full coverage, where to look for possible omissions when they see a structure, whose help to enlist in ascertaining that no omission is made, how to make sure that everybody has been enumerated in a particular dwelling unit, etc. In planning details for a training program, data from past checks will be extremely useful.

Extreme care has to be taken in preparing the training program. Modern teaching means can profitably be used, such as films, slides, illustrative maps showing typical cases, etc.

3. *Publicity.* Publicity is a means of announcing the proposed survey to the public and of getting its co-operation in order to achieve the best possible results.

In the survey publicity, the public must be given *specific* indications about the co-operation needed. An appeal of the following type may prove useful: "Report to the enumerator every person in your household irrespective of whether you think that this person should be enumerated or not"; "Help the enumerator by giving accurate information on all the persons in your household"; "Help the census by checking with the enumerator (at a given date or in a particular office) that everybody in your neighborhood was enumerated," etc.

---

[31] The difficulties discussed here are, of course, not the only ones. In the reports of listing checks others will also be found. For example, there is the difficulty of matching the units listed in the two surveys. In the case of persons in a census of population, different sources of information might easily give different names for the same persons. If the matching is based exclusively on names, the number of errors reported in the check will be an overestimate. If other characteristics are also used in matching, the likelihood of errors is reduced although not excluded as various discrepancies in the response might lead to wrong decisions. Confusions of this nature are particularly likely in checking the quality of listing in censuses of agriculture. The names of the farm owner, the operator or the shareholder may all appear in the census, thus making the matching more difficult than in population censuses. Various devices are therefore used to make such confusion impossible. For example, both the census and the check enumerators may be asked to show each farm listed on their respective maps or aerial photographs so that geographic location can also be used in matching.

[32] The various steps that can be taken in preparing the ground for satisfactory field operations in censuses and surveys are thoroughly discussed in the *Handbook of Population Census Methods*, Vol. I, United Nations, Studies in Methods, Series F, No. 5, Rev. 1, New York, 1958. The reader who is interested in this problem in a more general way should consult this book.

If the co-operation of the public has been successfully secured the enumerator can enlist the help of persons in his ED who are well informed on people living there, agricultural holdings or business establishments located there, etc. Such people can contribute substantially to improving the coverage.

4. *Legal obligations.* Large-scale government surveys can be considered national goals. This being so, legislation may be prepared requiring, among other things, that citizens help the census by receiving enumerators and giving them the information they need. In census work this legislation is commonly called "census law." The census law might usefully require that each head of a household or farm holder ensure that he himself, his dependents or his holding are properly enumerated. In such a case it might be useful to have an adequate organizational arrangement which offers everybody a chance to co-operate.

In the preparation of a large-scale survey the emphasis is primarily on voluntary co-operation. Legal obligation and legal sanctions are intended to enforce collaboration where it cannot be automatically expected.

Another problem is the extent to which such an obligation can be enforced. The obligation can be made effective if a provision is made for fines or some similar action. Such a provision corresponds to legal tradition in some countries, while in others it would hardly be feasible. Therefore, a decision on the use of legal means to improve the quality of listing has to take into consideration the country's traditions in similar fields.

5. *Supervision.* Large-scale surveys need intensive supervision. In some recent censuses a supervisor was appointed to each 10-15 enumerators. In some more complex sample surveys one supervisor was in charge of four enumerators.

The role of a supervisor of field operations is a dual one. By his presence and inspection on the spot he is supposed to prevent carelessness on the part of enumerators, dishonesty and other types of deviation from the procedures prescribed. He is also supposed to give guidance to his enumerators. For this he must have proper training and specific instructions on how to supervise the listing efficiently in a large number of EDs.

In some cases it may be possible to arrange the supervisors' assignment and reporting on a random basis so that a very early estimate of the quality of listing becomes available prior to the commencement of enumeration.

6. *Short period of enumeration.* The period of time used for the actual process of enumeration has a considerable bearing on the quality of listing. This is especially so in censuses of population where listing normally follows classification of persons in the sense discussed above. If the enumeration can be completed in a few days the interest of the public can be kept alive for such a short time and greater co-operation can be expected.

This interest and the favorable atmosphere developed by census publicity should be exploited to the full before they inevitably vanish after a few days. In the course of their daily activities, the people soon forget about the census, and grow less ready to co-operate. The resulting effect on the quality of the information collected is obvious.

Considerable skill is needed to develop a favorable atmosphere for enumeration and still more to utilize it thoroughly. A short period of enumeration is likely to be an important factor in this respect.

Clearly, the implications of this point are far-reaching. A short period of enumeration means more enumerators, smaller enumeration districts, possibly the use of self-enumeration which reduces the volume of work to be done by enumerators, etc. Each of these measures has in its turn a number of consequences. The problem therefore becomes a very complex one and requires on-the-spot study of the pros and cons of such measures.

7. *Use of sample checks.* It has already been emphasized on several occasions that sample checks encourage enumerators to do their work as carefully as they can. In order to be efficient their application has to be adequately prepared. Through publicity the public should be informed of the existence and nature of sample checks because this will lead to better preparation of the relevant information. The enumerators for their part have to be trained in the various possibilities of checking the quality of their work effectively. This will give them a responsible attitude to supervision.

8. *Other means.* A number of other methods have been used in various countries. For example, in the United States a "missed person" form was published in newspapers and everybody who thought that he might have been omitted in the enumeration was invited to fill in this form and send it to a given address. The forms received were later checked against the census material to ascertain whether the person concerned had in fact been omitted.

Another device was used in the 1940 Census of Population in Brazil. It was announced that any person indicating names omitted in the census would be given a reward in money. According to the report, the technique became very popular, particularly among students.[33]

In the same census, post office workers were asked to inspect the census listings on account of their knowledge of the people living in the area which they served. Another measure, already mentioned, would be to ask the town clerks or the local schoolteachers to inspect the listing.

9. *Studying listing errors.* In planning measures for reducing the inaccuracy of listing there is no substitute for a systematic study of listing errors on the lines discussed above. Such studies reveal the most important sources of errors, the general course to be adopted for their elimination and the specific measures needed.

---

[33] Depoid, P.: Rapport sur le degré de précision des statistiques démographiques, *Bulletin of the International Statistical Institute*, Vol. 35, Part 3, 1957, pp. 119-230.

7. MISSING DATA

7.1 The problem

In sample surveys where the public is under no obligation to co-operate, it may easily happen that data are not collected for all units. This is particularly so if only one visit to the units concerned is planned. Data for a part of the population will be missing and the problem arises as to what conclusions are authorized from such surveys or what measures could be taken to gain some insight into the part of the population left uncovered.

This problem does not usually arise in an acute form with complete enumeration censuses. Clearly, the first reason is the legal obligation imposed. Then comes the census publicity which makes the existence of the census known to everybody so that people have time to prepare themselves for enumeration or establish contact with enumerators. In censuses, moreover, the enumerator is working in a small area and, if a person to be enumerated has not been contacted at the first call, it is easy to repeat the call the same day or even some other day. Finally, censuses are concerned with certain basic facts, usually known to several members of a household. This makes it possible to get the necessary information from a larger body of respondents.

In sample surveys the situation is very different. In most cases legal obligation does not exist. This is particularly so with surveys conducted by private research agencies, where publicity is either minimal or non-existent. As a consequence, respondents are unaware of the existence of the survey and may not be at home when the enumerator calls on them. So the problem of *not-at-homes* arises. Other people, for one reason or another, do not wish to be disturbed and refuse to give information. In statistical literature such cases are called *refusals*. We are going to use the term *missing data* as a common name for all such cases irrespective of the reason for the enumerator's failure to obtain data. In the terminology of some authors the problem is also known by the name of *non-*

response. Others refer to the problem under the name of *incomplete samples.*

In sample surveys the enumerators have a rather tight travel schedule for visiting units which are sometimes far away from each other. If they return to make another attempt to obtain missing data the cost of the survey will be increased, nor is there any guarantee that this new call will be successful. How many attempts should be made to reach the absentees or to persuade the refusers? What increase in the cost of the survey would be justified? What happens if data collected in the first call are used and no additional efforts to get more data are made?

This is the type of question that we propose to discuss in this chapter.

7.2 Consequences of missing data

The first question is as follows: what happens if data collected in the survey are tabulated irrespective of the fact that the information for some units is missing?

In order to answer this question it may be useful to conceive the population to be surveyed as consisting of two strata. The first stratum comprises the units which can be covered at the first call. The next stratum is composed of the remaining units, i.e., the not-at-homes and refusals. The size of the first stratum is N_1 and that of the second N_2, with $N_1 + N_2 = N$. The true value of the characteristic under study will be designated by x and the total size of the sample by n. After the first call n_1 responses will be available for units falling in the first stratum. The arithmetic mean computed from data available will be an estimate of the arithmetic mean of the first stratum, i.e.,

$$\bar{x}_1 = \frac{1}{n_1} \sum_i^{n_1} x_i$$

with $E\bar{x}_1 = \bar{X}_1$. The quantity we want to estimate is

$$\bar{X} = \frac{1}{N}(N_1 \bar{X}_1 + N_2 \bar{X}_2)$$

Accordingly, the bias obtained is equal to

$$\begin{aligned} \bar{D} &= E\bar{x}_1 - \bar{X} \\ &= \bar{X}_1 - \bar{X} \qquad (7.1) \\ &= \frac{N_2}{N}(\bar{X}_1 - \bar{X}_2) \end{aligned}$$

which shows that the procedure leads to biased results whenever $\bar{X}_1 \neq \bar{X}_2$. Therefore, if $\bar{x}_1$ is used as an estimate of the population arithmetic mean there is a risk of bias.[1]

A similar result is obtained with regard to the variance. If we were sampling with replacement and the results collected in the first call used for estimating the population variance by means of

$$s_1^2 = \frac{\sum^{n_1} (x_{1i} - \bar{x}_1)^2}{n_1 - 1}$$

we should have

$$\begin{aligned} Es_1^2 &= \sigma_1^2 \\ &= \frac{1}{N_1} \sum_i^{N_1} (x_{1i} - \bar{X}_1)^2 \end{aligned}$$

which is obviously the variance of the first stratum. The variance of the population, σ^2, can be expressed as

$$\sigma^2 = \frac{1}{N}\left[\sum N_i\, \sigma_i^2 + \sum N_i\, (\bar{X}_i - \bar{X})^2\right]$$

The bias resulting from the use of s_1^2 as the estimate of the population variance is then

$$\begin{aligned} D_{s_1^2} &= Es_1^2 - \sigma^2 \\ &= \frac{N_2}{N}(\sigma_1^2 - \sigma_2^2) - \frac{N_1}{N_2}\bar{D}^2 \qquad (7.2) \end{aligned}$$

---

[1] The symbol $\bar{D}$ is here given a somewhat different meaning from before. It is hoped, however, that no confusion will arise as the equation (7.1) clearly shows what is involved.

σ_2^2 being the variance of the nonresponding stratum. It is seen from (7.2) that s_1^2 is an unbiased estimate of σ^2 if $\sigma_1^2 = \sigma_2^2$ and $\bar{D} = 0$. Since neither $\bar{D}$ nor σ_2^2 is known, it follows that data collected in the first call cannot reasonably be used for estimating the population variance without the risk of some bias.

The problem of the consequences of missing data in sample surveys has been discussed in greater detail by Birnbaum and Sirken.[2]

The question now arises of what to do in the case of a survey with missing data. Sometimes the technique of *substitution* is used in such a situation. It consists in selecting at random more units than the size of the sample originally fixed, i.e., m. These additional units are used in the order of selection until we reach n successful calls.

The aim of the substitution is to prevent the reduction of precision that would result from the reduced size of the sample. It is obvious, however, that the response always comes from the same stratum. The information on units in the nonresponding stratum continues to be missing. Irrespective of the size of the sample from the responding stratum, the bias (7.1) always remains. If the accuracy of such a survey is expressed by means of the mean square error as in the equation (1.15), the substitution will affect the variance term while the bias term will remain unchanged. Thus the substitution does not solve the problem if the responding and the nonresponding stratum are different.

In fact, experience shows that the responding and nonresponding strata may be very different. Some examples may be useful to illustrate the magnitude of these differences. Let us start with examples presented by Hendricks.[3] In the first of these Hendricks shows data collected in a mail survey of 3,241 fruit growers in North Carolina. After the first mailing of the questionnaire only 300 returns were received. The same questionnaire was then sent to the growers who had not answered the first time. This brought in an additional 543 returns. The third mailing resulted in 434 more returns. The average number of fruit trees was computed separately for the returns obtained in each mailing and the results were as follows:

Mailing	*Average number of trees per farm*
1	456
2	382
3	340

In this particular case the true value of the arithmetic mean was known to be 329 trees per farm. If data collected with the first mailing were used and further work on obtaining information discontinued, the estimates would be seriously biased.

In this particular example large farms offered better co-operation. However, this should not be taken as a general rule. Hendricks reports another example showing quite an opposite tendency. A mail survey of 1,189 milk producers was conducted in North Carolina and three successive mailings were used as before. The results in Table 21 were obtained.

TABLE 21. - RESULTS OBTAINED IN THREE SUCCESSIVE MAILINGS OF QUESTIONNAIRES IN A SURVEY OF 1,189 MILK PRODUCERS IN NORTH CAROLINA [1]

Mailing	Questionnaires returned		Average number of cows per reporting farm
	Number	Percent	
1	165	13.9	23.03
2	170	14.3	23.79
3	114	9.6	24.23

[1] Hendricks, W.A.: Op. cit.

In this case the average number of cows per farm was also known. It was 24.27 cows. Here the tendency is a better response rate from smaller holdings (as measured by the number of cows). The use of estimates obtained with the first mailing would again lead to biased results but the bias here would be in the opposite direction.

Gray and Corlett[4] report the results obtained in two interview surveys after a number of successive calls, as shown in Table 22.

Data in Table 22 require the same comments as before. The averages of the first call are substantially different from the others; the need for a careful study of missing data is obvious.

---

[2] Birnbaum, Z.W. and Sirken, M.G.: Bias due to non-availability in sampling surveys, *Journal of the American Statistical Association*, Vol. 45, 1950, pp. 98-111.

[3] Hendricks, W.A.: *The Mathematical Theory of Sampling*, The Scarecrow Press, New Brunswick, N.J., 1956.

[4] Gray, P.G. and Corlett, T.: Sampling for the social survey, *Journal of the Royal Statistical Society*, Vol. 113, Series A, 1950, pp. 150-206.

TABLE 22. - AVERAGES OF TWO CHARACTERISTICS BASED ON DATA COLLECTED IN DIFFERENT CALLS

Call	Average expenditure on meals in catering establishments per adult per week (shillings)	Average number of visits to the cinema per adult per month
1	1.83	2.02
2	3.10	2.44
3	3.88	2.47
4 and more	3.61	2.00
All persons interviewed	2.69	2.25
Adjusted value	2.75	2.26

One more example, shown in Table 23, is very instructive. Corresponding data are taken from a paper by Durbin and Stuart. The table shows the percentage of persons having an attribute as estimated on the basis of data obtained in the first call on the one hand and in the first three calls combined on the other.

These data give an indication of the magnitude of the problems raised by missing data.

TABLE 23. - PERCENTAGES OF PERSONS HAVING AN ATTRIBUTE AS ESTIMATED FROM THE FIRST CALL AND CALLS 1 - 3 COMBINED [1]

Attribute	Estimated percentage of persons	
	First call	1 - 3 calls combined
Women who are housewives	92.6	89.3
Of those with jobs who work full time	79.4	88.0
Workers employed in manufacturing trades	22.9	30.3
Workers whose time to work is less than 20 minutes	54.0	54.0
Who went to pictures in last 7 days	28.7	29.3
Who bet on football pools	36.1	41.8
Who smoke	50.5	57.0
Who read periodicals regularly	51.9	53.6
Who own (or whose family own) a car	23.3	24.2

[1] This table is extracted from data presented in Table 4 of Durbin, J. and Stuart, A.: Call-backs and clustering in sample surveys, *Journal of the Royal Statistical Society*, Vol. 117, Series A, 1954, pp. 387-428.

The above examples show that in survey practice all kinds of situations may arise. Since we never know which of the possible cases will appear in a particular survey, the only safeguard against biases resulting from missing data is to incorporate in the survey design itself some treatment of this problem. Of course, the treatment adopted has to be flexible and provide for successive measures up to the point where further action does not seem to be justified.

7.3 The Hansen and Hurwitz technique[5]

In order better to visualize the various aspects of the theory developed by Hansen and Hurwitz, a file will be assumed containing N addresses of establishments. For a sample survey, a simple random sample of n addresses is selected and a mail questionnaire sent to each of the establishments involved. The population consists of N_1 establishments that answer the questionnaires and of N_2 establishments that do not. Accordingly, out of the total size of the sample, n, there will be n_1 returns and n_2 missing questionnaires. The assumption made also involves $N_1 + N_2 = N$ and $n_1 + n_2 = n$.

To avoid biases due to missing data a subsample of the size n'_2 is selected from n_2 nonresponding units with $n'_2 \leq n_2$. The enumerators are then sent out to interview the n'_2 establishments falling in the subsample. It is assumed that each enumerator is able to make the necessary contacts so that n'_2 responses are collected.

From n_1 data collected in the first stratum the arithmetic mean $\bar{X}_1$ is estimated as

$$\bar{x}_1 = \frac{1}{n_1} \sum_{i}^{n_1} x_{1i} \tag{7.3}$$

In the same way $\bar{X}_2$ is estimated as

$$\bar{x}'_2 = \frac{1}{n'_2} \sum_{i}^{n'_2} x_{2i} \tag{7.4}$$

---

[5] Hansen, M.H. and Hurwitz, W.N.: The problem of non-response in sample surveys, *Journal of the American Statistical Association*, Vol. 41, 1946, pp. 517-529.

Using (7.3) and (7.4) an estimate of $\bar{X}$, i.e., the mean of the whole population, is defined as

$$\bar{x} = \frac{n_1}{n}\,\bar{x}_1 + \frac{n_2}{n}\,\bar{x}_2' \tag{7.5}$$

The estimate (7.5) is unbiased because

$$\begin{aligned} E\left[\frac{n_1}{n}\,\bar{x}_1\right] &= E\left[\frac{n_1}{n}\,E\,\bar{x}_1\right] \\ &= \frac{N_1}{N}\,\bar{X}_1 \end{aligned}$$

and

$$\begin{aligned} E\left[\frac{n_2}{n}\,\bar{x}_2'\right] &= E\left[\frac{n_2}{n}\,E\,\bar{x}_2'\right] \\ &= E\left[\frac{n_2}{n}\,\bar{x}_2\right] \\ &= \frac{N_2}{N}\,\bar{X}_2 \end{aligned}$$

Accordingly,

$$\begin{aligned} E\bar{x} &= \frac{N_1}{N}\,\bar{X}_1 + \frac{N_2}{N}\,\bar{X}_2 \\ &= \bar{X} \end{aligned}$$

If the sampling fraction of the original sample is designated by $f_1 = \frac{n}{N}$ and every k-th unit is subsampled from among n_2 units, we put

$$f_2 = \frac{1}{k} = \frac{n_2'}{n_2}$$

and rewrite (7.5) as follows:

$$\begin{aligned} \bar{x} &= \frac{1}{n}\left(\sum^{n_1} x_{1i} + k\sum^{n_2'} x_{2i}\right) \\ &= \frac{1}{n}\left(\sum^{n_1} x_{1i} + \sum^{n_2} x_{2i} + k\sum^{n_2'} x_{2i} - \sum^{n_2} x_{2i}\right) \end{aligned}$$

$$= \bar{x}' + \frac{n_2}{n}\,(\bar{x}_2' - \bar{x}_2) \tag{7.6}$$

In this equation $\bar{x}'$ stands for an estimate of the population mean that would be obtained had n responses been collected in the initial sample. The variance of $\bar{x}'$ in simple random sampling without replacement is clearly

$$\sigma^2_{\bar{x}'} = \frac{N-n}{N-1}\,\frac{\sigma^2}{n}$$

σ^2 being the per unit variance of the whole population.

To find the variance of the second term in (7.6) we put

$$\begin{aligned} u &= \frac{n_2}{n}\,(\bar{x}_2' - \bar{x}_2) \\ Eu &= 0 \\ \sigma_u^2 &= Eu^2 - (Eu)^2 \\ &= Eu^2 \end{aligned}$$

In the evaluation of Eu^2 Cochran [6] uses the identity

$$(\bar{x}_2' - \bar{X}_2) = (\bar{x}_2' - \bar{x}_2) + (\bar{x}_2 - \bar{X}_2) \tag{7.7}$$

From the equation (7.7) we get

$$E\,(\bar{x}_2' - \bar{x}_2)^2 = E\left[(\bar{x}_2' - \bar{X}_2)^2 - (\bar{x}_2 - \bar{X}_2)^2\right]$$

where the contribution of the cross-products of the two terms on the right vanishes. Taking the expected values of terms in brackets for fixed n_2 we have

$$\left.\begin{aligned} E\,(\bar{x}_2' - \bar{X}_2)^2 &= \frac{N_2 - n_2'}{N_2 - 1}\,\frac{\sigma_2^2}{n_2'} \\ &\text{and} \\ E\,(\bar{x}_2 - \bar{X}_2)^2 &= \frac{N_2 - n_2}{N_2 - 1}\,\frac{\sigma_2^2}{n_2} \end{aligned}\right\} \tag{7.8}$$

---

[6] Cochran, W.G.: *Sampling techniques*, Second edition, John Wiley, New York, 1963.

with

$$\sigma_2^2 = \frac{\sum\limits_{i}^{N_2} (x_{2i} - \bar{X}_2)^2}{N_2}$$

Accordingly,

$$E\left[(\bar{x}_2' - \bar{X}_2)^2 - (\bar{x}_2 - \bar{X}_2)^2\right] = \frac{N_2}{N_2 - 1}\,\sigma_2^2 \left(\frac{1}{n_2'} - \frac{1}{n_2}\right) \tag{7.9}$$

$$= \frac{k-1}{n_2}\,\frac{N_2}{N_2 - 1}\,\sigma_2^2 \tag{7.10}$$

After multiplying (7.10) by $\left(\frac{n_2}{n}\right)^2$ and using $En_2 = n\frac{N_2}{N}$, we get

$$E\left(\frac{n_2}{n}\right)^2 (\bar{x}_2' - \bar{x}_2)^2 = \frac{k-1}{nN}\,\frac{N_2^2}{N_2 - 1}\,\sigma_2^2 \tag{7.11}$$

Collecting the partial results to get $\sigma^2_{\bar{x}}$ by using (7.6) we have

$$\sigma^2_{\bar{x}} = \frac{N-n}{N-1}\,\frac{\sigma^2}{n} + \frac{k-1}{nN}\,\frac{N_2^2}{N_2 - 1}\,\sigma_2^2 \tag{7.12}$$

The equation (7.12) shows that the technique of making a special inquiry in the nonresponding stratum leads to a lesser degree of precision than would have been achieved had a successful simple random sampling been carried out or, in other words, had all the n units selected in the initial sample supplied the necessary data. This decrease in precision is reflected in the second term on the right of (7.12). Among its governing factors k requires special attention. It can be fixed arbitrarily or by considering the resources available, which very often determine the upper limit in this respect. A more rational approach is found by minimizing the variance (7.12) for the fixed survey budget C and obtaining the optimum values of k and n. If the desired average precision is stated in advance, the alternative would be to minimize the budget C for a fixed precision ε and obtain the optimum values of k and n.

To obtain the values of k and n in the latter case the following cost elements will be used:

c_0 = cost of including a unit in the initial sample and checking whether it belongs to stratum 1 or 2;

c_1 = cost of obtaining information from units belonging to stratum 1;

c_2 = cost of obtaining information from a unit in stratum 2 (nonresponding units in the initial survey);

C = total budget of the survey.

Accordingly, the structure of the survey budget is as follows:

$$C = c_0 n + c_1 n_1 + c_2 n_2' \tag{7.13}$$

which has an average

$$EC = n\left(c_0 + c_1\frac{N_1}{N} + c_2\frac{N_2}{kN}\right)$$

Now we use

$$F = EC + \lambda\,(\sigma^2_{\bar{x}} - \varepsilon^2)$$

subject to the condition $\sigma^2_{\bar{x}} - \varepsilon^2 = 0$. Differentiating F with regard to n, setting the result equal to zero and solving for λ one has

$$\lambda = \frac{n^2\left(c_0 + c_1\frac{N_1}{N} + c_2\frac{N_2}{kN}\right)}{\frac{N}{N-1}\,\sigma^2 + \frac{k-1}{N}\,\frac{N_2^2}{N_2 - 1}\,\sigma_2^2} \tag{7.14}$$

Differentiating F with regard to k and putting the result equal to zero one obtains

$$c_2\frac{nN_2}{k^2 N} = \frac{\lambda}{nN}\,\frac{N_2^2}{N_2 - 1}\,\sigma_2^2 \tag{7.15}$$

Substituting the value of λ from (7.14) in (7.15) and solving for k gives the optimum value of k:

$$k = \left[\frac{c_2\frac{N_2}{N}}{c_0 + c_1\frac{N_1}{N}}\,\frac{N^2(N_2 - 1)\,\sigma^2 - N_2^2(N-1)\,\sigma_2^2}{N_2^2(N-1)\,\sigma_2^2}\right]^{\frac{1}{2}} \tag{7.16}$$

Taking $N \doteq N - 1$ and $N_2 \doteq N_2 - 1$, (7.16) reduces to

$$k = \left[\frac{c_2}{Nc_0 + c_1 N_1}\,\frac{N\sigma^2 - N_2\sigma_2^2}{\sigma_2^2}\right]^{\frac{1}{2}} \tag{7.17}$$

The optimum value of the total size of the sample, n, is easily obtained by letting $\sigma^2_{\bar{x}} - \varepsilon^2 = 0$, substituting for $\sigma^2_{\bar{x}}$ from (7.12) and solving for n. The result is

$$n = \frac{\dfrac{N^2}{N-1}\,\sigma^2 + (k-1)\,\dfrac{N_2^2}{N_2-1}\,\sigma_2^2}{N\varepsilon^2 + \dfrac{N}{N-1}\,\sigma^2} \tag{7.18}$$

Again taking $N \doteq N-1$ and $N_2 \doteq N_2 - 1$, (7.18) reduces to

$$n = \frac{N\sigma^2 + (k-1)\,N_2\sigma_2^2}{N\varepsilon^2 + \sigma^2} \tag{7.19}$$

The interpretation of the results obtained is as follows: on the basis of the quantities assumed known in (7.16) the optimum subsampling fraction, $1/k$, is determined and then used in (7.18) to get the total size of the sample. Afterward, a sample of that size is selected and the information on the survey program is requested. After the nonresponding units have been identified, every k-th of them is subsampled for interviewing. The data finally obtained are used to estimate the mean according to (7.5) and its variance as in (7.12). The cost of the survey will be about the minimum for the average precision expressed by ε^2.

It can be seen from (7.17) and (7.18) that the optimum values of k and n assume a knowledge of the response rate prior to the survey. If this condition is not fulfilled it may still be possible to make some guesses on the magnitude of nonresponse from experiences with similar surveys in the past. In cases where there is no guidance whatsoever for such guesses there is still a possibile way of achieving savings. This is to determine the size of the sample for several values of the nonresponse rate and send out the corresponding maximum number of questionnaires. After the returns have been received the subsampling fraction is fixed according to the nonresponse rate found in the sample.[7]

It may also be useful to add that there has been some misunderstanding about the use of the Hansen and Hurwitz technique, in that in some cases it has been interpreted as the solution to the problem of missing data only as at appears in mail surveys. However, the possibility of its extension to general interview surveys is obvious. After the first attempt has been made to interview n units in the sample and the n_2 nonresponding or nonavailable units have been identified, this technique may be used to determine the optimum sampling fraction for selecting a subsample of n_2' units to be covered in the second call.

In such applications, however, the technique probably loses much of its merit and some alternative approach might be more economical. Attention was drawn to this question by Durbin.[8]

Let us use c_1 for the average cost per unit of obtaining the information in the first call and c_2 for the corresponding cost of the second call. Now, if the population is assumed to be large so that finite multipliers can be dropped, the following optimum value for k is obtained

$$k = \left[\frac{c_2}{c_1}\;\frac{\sigma^2 - \dfrac{N_2}{N}\,\sigma_2^2}{\sigma_2^2}\right]^{\frac{1}{2}} \tag{7.20}$$

Assuming further that $\sigma^2 = \sigma_2^2$, (7.20) reduces to

$$k = \left[\frac{N_1}{N}\;\frac{c_2}{c_1}\right]^{\frac{1}{2}} \tag{7.21}$$

For any result with $k \leq 1$ in (7.21) there will be no subsampling at all and the optimum procedure consists in a new attempt to interview all the nonresponding units. This is achieved with $c_2 = 3c_1$ and the response rate at the first call $\dfrac{N_1}{N} = \dfrac{1}{3}$, which is approximately the case with many surveys. Obviously, the crucial point is the cost ratio c_2/c_1. In many situations the application of the Hansen and Hurwitz technique is confined to the standard procedure of re-calling the total nonresponse.

---

[7] For further details on this technique and particularly on its use with stratified samples or with the ratio method, cf. estimation of the original paper by Hansen and Hurwitz.

[8] Durbin, J.: Non-response and call-backs in surveys, *Bulletin of the International Statistical Institute*, Vol. 34, Part 2, 1954, pp. 72-86.

If subsampling of nonresponding units at first call is used, Dalenius[9] has shown the possibility of selecting a subsample of nonresponding units while the survey is going on in the field. With this technique the time lag between the first and the second survey is eliminated.

Dalenius uses Kish's[10] procedure for an objective selection of the respondent within the household. He assumes that, prior to the commencement of the survey, a list will be established of all the potential respondents in the area with an indication against each name of who is present and who is absent at the time of the survey. Afterward, different probabilities can be assigned to persons in each of the two strata in accordance with the subsampling fraction desired of the nonavailable units. For example, persons present will have twice as much chance of selection as those absent if each person present on the list is assigned two consecutive serial numbers and those absent only one number. An alternative is to prepare two separate lists for each class of persons and, during the selection of the sample, apply any sampling fraction found to be convenient.

Obviously, this technique can only be used if lists are prepared immediately before the collection of data. It also assumes an approximate knowledge of the nonresponse rate, or the rate of not-at-homes, from past work. Its advantages are the reduction of the time lag mentioned above as well as the possibility of considerable savings in surveys with moving enumerators. These enumerators select the sample themselves while on the spot. They can also select the subsample of not-at-homes and manage to get information from them so that a second call becomes unnecessary.

An extension of the Hansen and Hurwitz technique was developed by El-Badry.[11] The essence of El-Badry's technique consists in sending out a large number of mail questionnaires as a cheap way of increasing the response. After a number of attempts have been carried out by mail, a sample of the remaining nonresponse is selected and the enumerators are sent out to make personal contacts. Mail questionnaires are used until they cease to be effective. Personal field visits represent the last step to eliminate the bias due to missing data. Estimates are then based on the pooled results of all the attempts. El-Badry has developed the estimator of the average, its variance, and the technique of establishing the optimum sampling fraction for each successive attempt, as well as the optimum number of attempts.

El-Badry's technique is applicable when time is available to allow for the return of the successive batches of questionnaires. In addition, the theory assumes a knowledge of the response rates at each successive attempt.

Foradori[12] also used the Hansen and Hurwitz technique, and has developed the estimators of totals and their variances under the assumption of different designs. He has also considered the problem of the optimum allocation of the sample as well as the generalization of El-Badry's multiphase approach.

7.4 The Politz and Simmons technique

The Hansen and Hurwitz technique was developed primarily to deal with the problem of missing data as it appears in mail surveys where there is no difficulty in collecting the information by interviewing if mail contacts are not successful. In other words, it is based on the assumption that a single interview is sufficient to get data from units in the nonresponding stratum. In some cases this assumption is realistic. For example, if the survey is conducted with the aim of getting information on some characteristics of manufacturing establishments and a list of such establishments is used to select the sample, it can reasonably be expected that nonresponding units will supply the requested data if approached by an able enumerator. Such a situation is encountered in many countries in government surveys.

In a large number of other cases the situation is very different. For example, in public opinion surveys which are generally made by private organizations, the respondents may not co-operate as readily as in the case of government surveys. Difficulties are particularly likely to arise if individual persons are used as sampling units and the information required concerns their attitudes, buying preferences, political opinions, etc. If an attempt is made to reach them by mail questionnaires, they may not take the trouble to fill in and post the questionnaire. The result may then be a high rate of missing data. If approached by enumerators,

---

[9] Dalenius, T.: The problem of not-at-homes, *Statistisk Tidskrift*, Vol. 4, 1955, pp. 208-211.

[10] Kish, L.: A procedure for objective respondent selection within the household, *Journal of the American Statistical Association*, Vol. 44, 1949, pp. 380-387.

[11] El-Badry, M.A.: A sampling procedure for mailed questionnaires, *Journal of the American Statistical Association*, Vol. 51, 1956, pp. 209-227.

[12] Foradori, G.T.: *Some non-response sampling theory for two stage designs*, North Carolina State College. Mimeographed series, Raleigh, 1961.

they may not be at home when the enumerator calls on them. For all these and similar reasons the first attempt to get data, irrespective of what method of collecting information is used, often brings in only a small proportion of the total response. A second attempt will usually improve the situation although a large proportion of nonresponse will still persist. Therefore, the calls are repeated until the nonresponse rate is reduced to a "hard core" of those who can hardly be interviewed at all. After such a limit has been reached further attempts at calling are not likely to reduce the remaining nonresponse.

In view of the importance of successive calls in survey practice an illustration of the success achieved in various calls may usefully be presented.

In a mail pilot survey taken by Social Survey in London, Gray and Corlett[13] report the progress achieved by sending reminders to nonresponding persons after the initial mailing. The results are shown in the following table:

TABLE 24. - PROGRESS IN SUCCESSIVE QUESTIONNAIRE MAILINGS

	Percentage of returns received
Before the first reminder	38.0
After the first reminder	31.7
After the second reminder	7.3
Returned from alternative address	1.4
Information received of death, illness, etc.	2.4
Returned by post office	9.2
No information received	10.0
	100.0

In an interview survey of household income and expenditure Cole[14] reports progress achieved in successive calls, as shown in Table 25.

---

[13] Gray, P.G. and Corlett, T.: Sampling for the Social Survey, *Journal of the Royal Statistical Society*, Vol. 113, Series A, 1950, pp. 150-206.

[14] Cole, D.: Field work in sample surveys of household income and expenditure, *Applied Statistics*, Vol. 5, 1956, pp. 49-61.

TABLE 25. - PROGRESS IN SUCCESSIVE CALLS
(Interview survey of household income and expenditure)

Call	Percentage of households contacted
1	62.0
2	22.3
3	9.0
4	4.1
5	1.6
6 and more	1.0

TABLE 26. - CUMULATIVE PERCENTAGES OF RESPONSE TO INITIAL MAILING AND SUCCESSIVE REMINDERS [1]
(Surveys of U.S. army veterans)

Call	Male separatees (July 1945)	Male separatees (Dec. 1943)	Female separatees	Applicants for educational benefits	Amputees
Initial mailing	55	45	51	54	66
First reminder	76	73	74	77	88
Second reminder	88	87	87	88	94
Number of cases	(1 812)	(3 209)	(1 771)	(14 606)	(1 594)

[1] Data in this table are taken from Clausen, J.A. and Ford, R.N.: Controlling bias in mail questionnaires, *Journal of the American Statistical Association*, Vol. 42, 1947, pp. 497-511.

In mail surveys of United States army veterans, conducted after the second world war, reminders were used after the nonresponse had been identified and the cumulative percentages of response obtained were as shown in Table 26.

In the Canadian Labour Force Survey[15] the first call brings in about 60 percent of the initial number of returns required. After the second call this goes up to about 80 percent.

---

[15] Keyfitz, N. and Robinson, H.L.: The Canadian sample for labour force and other population data, *Population Studies*, Vol. 2, 1949, pp. 427-443.

The importance of repeated calling as a means of eliminating the bias due to missing data is obvious from these examples. However, if successive calling is planned in interview surveys it considerably increases the total cost because of increased subsistence and traveling expenses. The problem thus arises of how to reduce the number of calls and at the same time secure, as far as is feasible, unbiased data. The Politz and Simmons technique[16] was developed in answer to this question.

This technique is as follows. Persons selected in the sample are visited by enumerators at randomly selected points of time. Each person in the sample is visited only once. The person contacted is asked to state how many times in the previous five days he or she was at home at the time when the call was made. Questionnaires of the persons found at home are then classified according to the number of times the persons involved were at home during the specified period at the time of the call. Six groups of questionnaires are thus obtained with the first group representing persons who were at home only the day when the enumerator called on them, the second group being composed of persons who were at home when called upon and one more day in the period of the past five days, etc. In the last group are persons who were at home six times in the specified period. Estimates of population totals, averages and proportions are then obtained by weighting the survey totals of each group by the inverse of the proportion of times the people in the corresponding group were at home at the time of the call. A higher weight is thus given to persons who are less frequently at home.

Let us now assume that the population consists of N units. If the call is made on all the units of the population, the total number of successful interviews will be n. An estimate of the population total will then be

$$X' = 6 \sum_{i}^{n} \frac{x_i}{j} \tag{7.22}$$

---

[16] Details of this technique will be found in the original paper by Politz, A. and Simmons, W.: An attempt to get the "not-at-homes" into the sample without call-backs, *Journal of the American Statistical Association*, Vol. 44, 1949, pp. 9-31 and in: Note on "An attempt to get not-at-homes into the sample without call-backs" by the same authors, published in *Journal of the American Statistical Association*, Vol. 45, 1950, pp. 136-137. Similar ideas were presented earlier by H.O. Hartley in the discussion of F. Yates' paper: A review of recent statistical developments in sampling and sampling surveys, *Journal of the Royal Statistical Society*, Vol. 109, 1946, pp. 12-43. Hartley, however, did not put his ideas in mathematical terms.

where x_i stands for the value of the characteristic for the i-th unit in the sample of successful interviews and j for the number of times the i-th person was at home.

The expected value of X' is

$$\begin{aligned} EX' &= 6\, E \sum_{i}^{n} \frac{x_i}{j} \\ &= 6 \sum_{i}^{N} p_i\, E_i \frac{x_i}{j} \end{aligned} \tag{7.23}$$

where p_i indicates the probability of the i-th person being at home when the interviewer calls. Similarly, $q_i = 1 - p_i$ represents the probability of that person being absent at that time. The term $E_i \frac{x_i}{j}$ stands for conditional expectation of $\frac{x_i}{j}$ knowing that the i-th person is found at home. Let us also use p_{ij} as the probability that the i-th person will be found at home j times out of six calls. We then have

$$p_{ij} = \frac{6\,!}{j\,!\,(6-j)\,!}\, p_i^j\, q_i^{6-j}$$

and

$$\sum_{j=1}^{6} p_{ij} = 1 - q_i^6$$

If the i-th person has already been enumerated the probability of this person being at home during j—1 times out of the remaining five calls is

$$\frac{5\,!}{(j-1)\,!\,(5-j+1)\,!}\, p_i^{j-1}\, q_i^{5-j+1}$$

Afterward we have

$$E_i \frac{x_i}{j} = \sum_{j=1}^{6} \frac{x_i}{j} \frac{5\,!}{(j-1)\,!\,(5-j+1)\,!}\, p_i^{j-1}\, q_i^{5-j+1}$$

and

$$EX' = \sum_{i}^{N} \sum_{j=1}^{6} x_i \frac{6\,!}{j\,!\,(6-j)\,!}\, p_i^j\, q_i^{6-j}$$

$$= \sum_{i}^{N} x_i \, (1 - q_i^6) \tag{7.24}$$

which shows that X' is unbiased estimate of the total value of the variate for all persons that can be found at home at least once during six visits.

Under the assumption that N units of the population are visited the variance of X' is

$$\sigma^2_{X'} = \sum_{i} x_i^2 \left[6 \sum_{j}^{6} \frac{p_{ij}}{j} - (1 - q_i{}^6)^2 \right]$$

For further discussion of this technique and particularly for the estimators of X' and $\sigma^2_{X'}$ the reader is referred to Deming[17] and Cochran.[18]

In some cases the Politz and Simmons technique has important advantages over the alternative approaches. It must be remembered that it was developed with the aim of eliminating the need for re-calls. If re-calling is expensive, the technique might be very efficient in terms of the amount of information obtained per unit of cost. It may also be suitable in cases where re-calls are not likely to get as accurate information as the first call, owing to the lapse of time between subsequent interviews and the event with which the survey is concerned. For example, if the survey is supposed to give information on the expenditure on a number of items during a specified day, re-calling after some days may not be recommended owing to memory shortcomings.

An aspect of the Politz and Simmons technique should be pointed out. This is the probability p_i of finding the i-th person at home when the enumerator comes to call. This probability is the basis of the theory. In the original Politz and Simmons paper it was assumed that interviewers make their call at a time selected at random within the hours of the day when the survey is taking place. The hours selected are generally in the evening when people are normally at home and available for interviewing. If the information obtained on the respondents' presence at home on the past five days is added, the result is a randomly selected systematic sample of six points of time spaced at 24-hour intervals. This sample is used to estimate the probability of finding respondents at home.

---

[17] Deming, W.E.: On a probability mechanism to attain an economic balance between the resultant error of response and the bias of non-response, *Journal of the American Statistical Association*, Vol. 48, 1953, pp. 743-772.

[18] Cochran, W.G.: *Sampling Techniques*, Second edition, John Wiley, New York, 1963.

In this connection it is interesting to consider Durbin's[19] discussion of the Politz and Simmons technique, which discards the above assumption. According to Durbin it is sufficient to secure "that an equal number of calls is made on each of the six evenings of the week, either over the survey as a whole or, more strictly, by each investigator." With this less restricting assumption Durbin arrived at the same basic results as Politz and Simmons.

Durbin also compared the efficiency of the Politz and Simmons technique with the procedure of re-calling and found that, under some assumptions, the latter method is more efficient if the value of the characteristics on the survey program is independent of the probability of finding the respondent at home when the call is made. If there is some correlation, the former method may be more efficient.[20]

7.5 Other contributions

Many other authors have dealt with the problem of missing data. References at the end of this book show how much attention this problem attracts.

Among these additional contributions Deming's study of the problem of missing data should have first mention.[21] Deming recognizes that repeated calling might be the most efficient technique in some situations. From his experience with urban surveys he has constructed Table 27, which shows how the relative bias, Rel ($\bar{D}$), decreases in successive calls. The mean square error also decreases parallel to the bias while the cost of the survey increases. The problem that arises in this situation is the economy of re-calling. Deming shows how to strike the balance between the cost and the accuracy or, in other words, how to establish the optimum subsampling fraction from the nonresponding units and how to determine the optimum number of calls.

Bartholomew[22] has developed a technique of obtaining unbiased estimates and dealing with the problem of missing data in an interview survey consisting of two calls. In this technique the enumerators arrange for

---

[19] Durbin, J.: Non-response and call-backs in surveys, *Bulletin of the International Statistical Institute*, Vol. 34, Part 2, 1954, pp. 72-86.

[20] For further conclusions and details, cf. Durbin: Op. cit.

[21] Deming, W.E.: Op. cit.

[22] Bartholomew, D.J.: A method of allowing for "not-at-home" bias in a sample survey, *Applied Statistics*, Vol. 10, 1961, pp. 52-59.

TABLE 27. - THE MAGNITUDE OF THE RELATIVE BIAS AND THE RELATIVE MEAN SQUARE ERROR AS OBTAINED UNDER SOME SPECIFIED CONDITIONS IN SUCCESSIVE CALLS [1]

Sample size	Number of calls						
	I	I - II	I - III	I - IV	I - V	I - VI	I - VII
Rel ($\bar{D}$)	—.1109	—.0758	—.0573	—.0453	—.0368	—.0304	—.0254
n = 100	.1612	.1411	.1328	.1286	.1262	.1249	.1241
n = 200	.1383	.1133	.1023	.0964	.0930	.0909	.0895
n = 300	.1298	.1023	.0898	.0829	.0788	.0762	.0746
n = 500	.1226	.0926	.0784	.0704	.0654	.0621	.0599
n = 1 000	.1169	.0846	.0687	.0592	.0530	.0489	.0460
n = 2 000	.1139	.0803	.0633	.0527	.0456	.0407	.0372
n = 3 000	.1129	.0788	.0613	.0504	.0429	.0376	.0337
n = 5 000	.1121	.0776	.0598	.0484	.0406	.0349	.0306
Costs at n = 1 000	$3 000	4 125	4 870	5 435	5 896	6 290	6 638

[1] This table is based on data presented in Table 5 of the original paper by Deming.

the second call in the course of the first visit. The arrangement is made in such a way that not-at-homes in the first call have an equal probability of being contacted at the second call. The units interviewed in the second call are then a random subsample of the not-at-homes of the first call.

Other authors have proposed some techniques applicable in specific situations. For example, Kish and Hess[23] propose that a number of addresses of nonresponding units from a recent and similar survey be added to the units selected for the survey in hand. In the course of the field work data are collected from the two parts of the sample and the added units serve as a "replacement" for the nonresponding units in the survey in hand. If similar surveys are repeated frequently and a record of nonresponse is kept, this procedure might represent an economical treatment of the problem of missing data.

In another study the same authors[24] considered the problem of noncoverage, i.e., failure on the part of the enumerators to include certain units in the survey. This failure could also be considered as the problem of missing data if it is known that omitted units exist. Otherwise, it forms part of the problems arising out of inaccurate frames and inaccurate listing.

El-Badry[25] considered the consequences of the enumerators' failure to include 0 (zero) as a response in the appropriate places of the questionnaire, this again being an aspect of the problem of missing data.

The problem of missing data also appears in data processing if entries are not available on all the questions in the questionnaires being processed. In this case field visits as a source of missing information hardly come into play. The techniques that can be applied depend upon the method of processing data and the time available. The use of electronic computers has opened up new possibilities in this field.[26] Developments of this type are too specialized to be dealt with here.

---

[23] Kish, L. and Hess, I.: A "replacement" procedure for reducing the bias of nonresponse, *The American Statistician*, Vol. 13, No. 4, 1959, pp. 17-19.

[24] Kish, L. and Hess, I.: On non-coverage of sample dwellings, *Journal of the American Statistical Association*, Vol. 53, 1958, pp. 509-524.

[25] El-Badry, M.A.: Failure of enumerators to make entries of zero: errors in recording childless cases in population censuses, *Journal of the American Statistical Association*, Vol. 56, 1961, pp. 909-924.

[26] Buck, S.F.: A method of estimation of missing values in multivariate data suitable for use with an electronic computer, *Journal of the Royal Statistical Society*, Vol. 123, Series B, 1960, pp. 302-306. Dalenius, T.: Automatic estimation of missing values in censuses and sample surveys, *Statistisk Tidskrift*, Vol. 11, 1962, pp. 395-400. Nordbotten, S.: On errors and optional allocation in a census, *Skandinavisk aktuarietidskrift*, 1957. Nordbotten, S.: Automatic editing of individual statistical observations, *Conference of European Statisticians*, WG. 9/37, 1962.

Another aspect of the problem of missing data may appear in life testing and medical follow-up. The occurrence of the event under study, such as death, may be prevented for various reasons of an accidental or other nature. For example, in a medical study of various diseases or the consequences of various treatments, data for some persons will be missing because of their death in an accident or in some other way.[27]

7.6 Success of the field work

One of the chief questions in survey practice is the reason for the failure to enumerate units and the characteristics of units where such a failure occurs. A detailed knowledge of this question is of basic importance in planning measures of a practical nature.

The answer to this question certainly depends upon the type of survey. In a survey of areas planted or areas harvested, where the measurements of each individual parcel in a cluster of parcels are used, there is very little likelihood that the problem of missing data will arise. The object of measurement is always available for study. In yield surveys based on crop-cutting, the problem is more serious. The field worker may come to the field selected after the crop has already been harvested or before it is mature. In the former case the information can no longer be obtained, while in the latter an additional visit is needed. This new visit again runs the risk of being a failure.

Further complications arise with interview surveys, where the problem of missing data assumes full importance. After the first call has been made in an interview survey, the outcome of individual attempts, according to Durbin,[28] can be classified according to the following breakdown:

(i) person interviewed
(ii) person known to be away from home during field work period
(iii) not at home (other than those under *ii*)
(iv) refusal to co-operate
(v) unsuitable for interview
(vi) change of address
(vii) person or address could not be traced.

---

[27] Cf.: Kaplan, E.L. and Meier, P.: Nonparametric estimation from incomplete observations, *Journal of the American Statistical Association*, Vol. 53, 1958, pp. 457-481.

[28] Durbin, J.: Non-response and call-backs in surveys, *Bulletin of the International Statistical Institute*, Vol. 34, Part 2, 1954, pp. 72-86.

Of the classes listed, the last two are the result of inaccurate frames. With regard to classes *(ii)* through *(v)* it is interesting to know in each particular survey what people fall into each of these classes, their characteristics and the implications of these characteristics as regards various aspects of the survey design.

The class of persons known to be away from home is composed of people on vacation, students away at school, soldiers, persons on some kind of mission, seasonal workers, sailors, etc. In a modern society there is a percentage of people who move for long periods and there can be no question of changing this situation during the time of the survey. There are, however, some ways of reducing the consequences of population movements on survey results. The first consists of fixing the *date* of the survey for a time when these movements are expected to be at their minimum. If there is no possibility of choosing the date of the survey, it may be possible to consider several *types of units* or *respondents* or *sources of information*. The problem will always be considerable in surveys where some particular persons are asked to supply information. This is the case, for example, in preference, attitude, and opinions studies. However, if households can be used instead of individuals, so that the availability of the information does not depend upon any particular individual, the rate of missing data will be greatly reduced. In this respect the illustrations provided by Stephan and McCarthy[29] may prove interesting.

Another way of reducing the volume of missing data due to persons being absent from home for long periods is obviously to allow the enumerators a longer time in which to catch some or most of these persons when they return home. Obviously this measure is not applicable with surveys where information is urgently needed or in the case of items which cannot be remembered for long.

Returning to our classification, the class listed under *(iii)* comprises persons out of the house for a short time. It includes housewives who are out shopping, persons at work, people out for a walk, at a movie, etc. Just as with class *(ii)* some possibilities for increasing the rate of successful interviewing were found by choosing a convenient period of the year for the survey, so similar effects can be achieved here by selecting a convenient *period of the day* for interviewing. It is noteworthy that

---

[29] Stephan, F.F. and McCarthy, P.J.: *Sampling Opinions*, John Wiley, New York, 1958.

evenings have been found more convenient in many countries. However, this is not an absolute criterion because the most convenient time depends upon the prevailing patterns of life. In a country where the employees go home for lunch, which is the basic meal, there might be a better chance of finding respondents at home during the lunch hour.

Whatever the hours selected for interviewing, there will always be a percentage of not-at-homes. Some people work in the afternoon, some in the evening; somebody else has gone out with a friend, etc. In order to include these people in the sample re-calling will normally be needed.

The success of interviewing also depends upon the quality of the field staff. Experienced and properly trained interviewers who persevere in their efforts to obtain information on not-at-homes from neighbors and to make appointments are usually the most successful. Durbin and Stuart[30] conducted an experimental study in this connection by using inexperienced students as interviewers on the one hand and the field staff of some survey organizations on the other. The latter were able to carry out the field work with a higher response rate than the former.

The success of interviewing also depends upon the characteristics of the people involved. Kemsley[31] has presented some results obtained in surveys conducted by Social Survey in Great Britain. These results show considerable variation in the response rate according to profession.

7.7 Refusals

The most difficult problems arise in the case of refusals, i.e., when persons selected are contacted but do not want to answer. The common reason for refusals is the fact that people do not wish to be bothered with something which does not interest them.

The rate of refusals certainly varies with the program of surveys as well as with the authority and reputation of the agency conducting or sponsoring the survey. Stephan and McCarthy presented data on the rate of refusals as obtained in a number of surveys. There rates range from a minimum of 0.4 percent to a maximum of 13.0 percent.[32] In some exceptional cases the rate even rises to 50 percent.

Refusals constitute a very delicate problem and are often difficult to handle. The aim here is to persuade the people who have decided not to co-operate to reverse their decision. Achieving this aim is far from simple. In order to plan adequate measures for handling refusals, the agencies responsible for conducting interview surveys require information on the persons refusing to give data and the circumstances under which the refusals occur. This will enable them to draw the enumerators' attention to a number of facts which may be useful in reducing the refusal rate. This information is obtained by analyzing the data on refusals. An example of such an analysis is presented by Gray and Corlett,[33] who classified the refusals according to age and sex distribution. Their data show that females refuse to co-operate more frequently than males. They also show that younger people co-operate more readily than older persons. Gray and Corlett's data refer to a particular type of survey and probably cannot be generalized. If, however, enough information of a similar type is available it will show a number of considerations deserving the enumerators' particular attention. Thus, one might point to the need for special care in approaching older people.

It is a very well-known fact that refusals depend largely upon the *time* that the respondents are supposed to spend in answering survey questions. In this connection the following paragraph from the paper by Gray and Corlett may usefully be quoted: "The proportion of people refusing to be interviewed varies with the subject and type of the inquiry. For inquiries where the person is simply to be interviewed and is asked to answer a series of questions the proportion ranges from about 1 percent on the Survey of Sickness to 3 percent. In other inquiries where the person is asked to do more than merely answer questions, the proportion refusing may rise considerably. For example, in a series of inquiries into the nutrition of housewives ...certain groups of housewives, interviewed at home, were asked to volunteer for medical examination at a local clinic: only 40 percent of them agreed to this examination. In a more recent inquiry into the nutrition of schoolboys in families of the lower income groups, the mother was asked to keep a detailed record of the boy's meals

---

[30] Durbin, J. and Stuart, A.: Differences in response rates of experienced and inexperienced interviewers, *Journal of the Royal Statistical Society*, Vol. 114, Series A, 1951, pp. 163-206.

[31] Kemsley, W.F.F.: Some technical aspects of a postal survey into professional earnings, *Applied Statistics*, Vol. 11, 1962, pp. 93-105.

[32] Stephan, F.F. and McCarthy, P.J.: Op. cit.

[33] Gray, P.G. and Corlett, T.: Sampling for the social survey, *Journal of the Royal Statistical Society*, Vol. 113, Series A, 1950, pp. 150-206.

for a period of one week. This involved the weighing and/or measurement of all the food. Only eight percent of the mothers refused to keep such a record, although a further eight percent failed to keep a complete record." Experiences with food consumption surveys are of a similar nature.

There has been a great deal of evidence collected so far showing that refusals often differ considerably from the remaining units in the sample. This makes it particularly important to plan measures for eliminating potential biases arising from this source. The experience of many survey organizations shows that much can be done in this direction by adequately training and selecting the field staff. A successful interviewer can only be a person who knows how to *get co-operation* and who is able to create a favorable and *friendly atmosphere* so that any possible barriers between him and the respondent are removed before he starts introducing his survey.

Another point of primary importance in connection with refusals is the ability of survey organizations and their enumerators to *build up an interest* in the survey on the part of respondents. If the latter are interested in the survey and its results, they will hardly refuse co-operation except in special cases, such as lack of time, indisposition, etc., where refusal to co-operate is mostly limited to a given moment.

The task of building up sufficient interest in the survey will be easier with government-sponsored surveys, which are mostly connected with important social and economic issues. The importance of data for public administration is obvious and, in principle, it will be easy for an enumerator to convince the respondent that his co-operation is needed for the promotion of common goals. An able exploitation of such arguments has led in some cases to a high rate of co-operation in surveys where the respondents had to do considerable work or sacrifice a great deal of their privacy by permitting field staff to stay in their homes and, say, weigh and examine the food they ate.

Building up the respondents' interest in surveys conducted by private survey organizations is a more complex problem. A number of means have been used to this end. In some medical surveys *free medical tests* were offered to respondents if desired; in some food consumption surveys housewives were given *gifts* or *money*. In mail surveys many different techniques have been used to arouse more interest in the survey. One such way is to take special care in preparing the *introductory* letter, and to make it more personal by using the respondents' name rather than the standard formula, "Dear Sir." This gives the respondent the feeling that his personal contribution to the survey is particularly appreciated. Some authors have also pointed out that the use of stamps instead of prepaid postage has similar effects.[34]

Irrespective of the measures taken, some refusals will probably remain in any interview survey. In such cases it may be useful to collect all the available information on persons refusing interviews, such as occupation, social status, composition of the family, etc. This information can sometimes be used in estimating corresponding data for the refusing units. Obviously, the estimates prepared by using more or less relevant information will, on an average, be less accurate than the actual responses if they are available. As a rule, however, these estimates will be more accurate if based on facts than if confined to mere guesses. The knowledge of certain characteristics highly correlated with the variables under study will make it possible to arrive at fairly accurate estimates. It is therefore important to instruct enumerators in each particular survey on what data they are expected to collect about refusals. Gray and Corlett report that in surveys where a high rate of refusals is expected the Social Survey used the practice of starting the interview with some questions which make it possible to classify the case according to certain criteria and allow for biases if the case turns out to be a refusal in subsequent interviewing.

Class (*v*) as listed in Section 7.6 consists of persons who for various reasons are not suitable for interviewing. Sick people fall into this category as well as those who are involved with important problems. Such cases would become refusals if enumerators insisted on getting information. In handling these cases some of the measures listed earlier will be useful, such as repeating the call at a later and more convenient date, collecting relevant data on the units involved, etc.

7.8 Post hoc adjustments for biases due to missing data

In the preceding sections it was shown how important it is to consider the problem of missing data while designing the survey. Several techniques have been presented which can be *built into* the survey design, so that the way of dealing with the problem is to some extent prescribed in ad-

---

[34] A number of similar measures will be found in the literature listed at the end of this study.

vance. The success achieved by applying these techniques will certainly vary according to circumstances. In some cases, after several calls data on a substantial percentage of units may still be missing. Assuming that further field work is discontinued at this point, the problem will arise of finding some post hoc technique for adjusting survey estimates for possible biases. The same problem will also appear if users of data want to adjust the available results of some survey made in the past.

The solution to this problem is simple if the assumption can be made that there is no difference between the strata of responding and nonresponding units. In this case the bias in averages or proportions does not exist. Totals can be estimated either by multiplying averages by the number of units in the population or by using N/n' as the raising factor instead of N/n, n' being the number of responding units in the sample. The effect of missing data is then reduced to a lower precision of estimates than that originally aimed at.

This procedure is risky in cases where the nonavailability of units for interviewing is correlated with the value of the characteristics under study. If some such correlation is assumed to exist the formulation of a technique suitable for the adjustment depends upon the amount and type of information available. Hendricks[35] has shown some examples where simple information was sufficient to achieve an improvement of estimates. The procedure which he used will be illustrated on data referring to the average number of fruit trees per farm as presented in Section 7.2. The data involved were obtained in a mail survey, the population mean being known from the census. The information available for the formulation of the technique consists of the averages per farm known from three successive mailings. The cumulative percentages of returns as well as the average number of trees per farm based on these data are as follows:

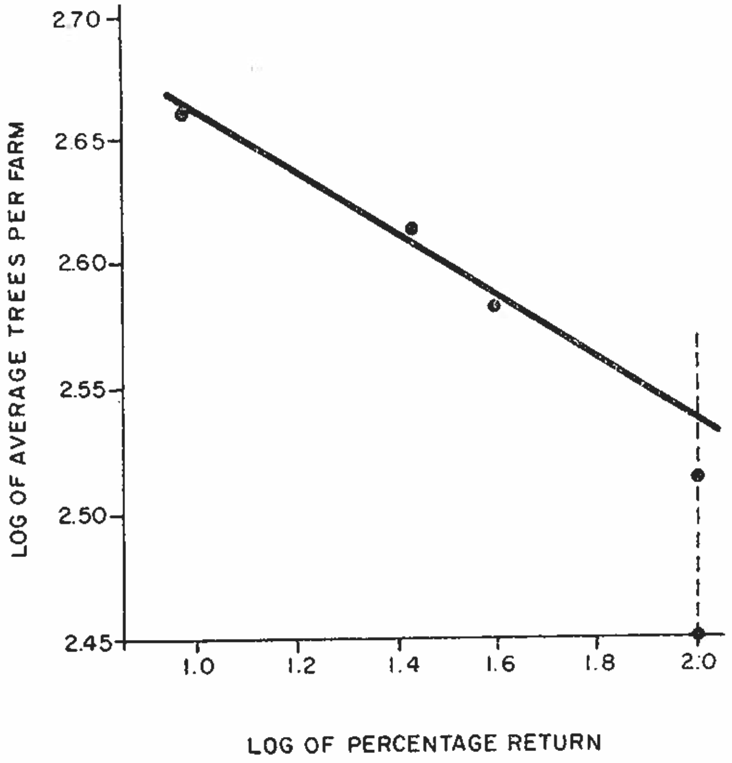
Mailing	*Percentage response*	*Trees per farm*
1	9.3	456
1 + 2	26.0	408
1 + 2 + 3	39.4	385

---

[35] Hendricks, W.A.: *The Mathematical Theory of Sampling*, The Scarecrow Press, New Brunswick, N.J., 1956.

After three successive attempts about 60 percent of responses was still missing. The need to adjust for the bias is therefore obvious. The first method cannot be used because data clearly show the existence of correlation between the average number of trees per farm and the willingness to co-operate with the survey as expressed in the number of attempts needed to get the response. In allowing for the bias this correlation can be taken into account by plotting on the x axis the logarithm of the percentage of response and on the y axis the logarithm of the average number of trees per farm. Figure 5 is obtained in this way. The complete response, i.e., 100 percent, falls on the x scale at point 2.0. At this point a vertical dotted line is drawn. In addition, a straight line

FIGURE 5. - RELATIONSHIPS BETWEEN THE CUMULATIVE PERCENTAGE RESPONSE AND THE AVERAGE NUMBER OF FRUIT TREES PER FARM[1]



---

[1] Hendricks, W.A.: *The Mathematical Theory of Sampling*, The Scarecrow Press, New Brunswick, N.J., 1956.

representing the trend in the three points available is also traced without any computation of its equation. The intersection of the two lines is the mean. In this case it is around 2.54. The antilogarithm of this is 344, which is an estimated average number of trees per farm. The known true value is 329. Therefore the difference still exists but it is smaller than that resulting from the assumption that there is no difference between strata. This assumption gives an average of 385 trees.

The possibilities of allowing for biases due to missing data depend, as a rule, upon the amount of information available on nonresponding or nonavailable units. For this reason it may sometimes be important to request enumerators to obtain the information prescribed in advance on such units by approaching neighbors or by using some other technique. For example, in a survey of areas under wheat, a number of holders may refuse to co-operate with the survey. It may be possible, however, to get data on the total area of the nonresponding holdings by consulting cadastral records, the revenue office or some other source. Data from some past survey are then used to establish the relationships between the total area and the area under wheat. The relationship is then used to adjust results for the nonresponse.

Questions in the survey questionnaire which are used to provide the information needed for adjustments are called *key questions.*

The selection of key questions depends upon the program of the survey and the possibilities of getting information from sources independent of the unit concerned. Several questions of this type may usefully be asked because it is normally not known in advance whether the enumerators will be equally successful with all of them. In addition, the survey program usually consists of a rather large number of characteristics and it is very unlikely that any single variable can be found which is sufficiently correlated with most of the survey characteristics. By collecting data for several items in missing units, the chances of finding a means of adjustment are increased.

In determining the key questions a good deal of subject-matter knowledge is needed in order to choose the most promising items in terms of their expected correlation with other characteristics of the survey program. In a survey on household income and expenditure it may be useful, in the case of missing units, to know the amount of the rent, the composition of the household, the number of children and employed members, the sex and age of each member, the occupational status of the head, etc. The rent may be correlated with the total income, expenditures on food

and clothing may be correlated with the size of the household, the number of employed members may also be correlated with the total income and expenditure for transport, etc. It can be seen that some key questions may themselves be a part of the survey program. The rest are additional questions introduced to increase the possibilities of adjusting data for missing information.

After the information on these questions has been obtained, it is far from simple to use it properly in a survey with many items on the program. Many ways are theoretically possible and a study of the available material is needed in order to find the procedure which is most suitable from the point of view of the purposes, the time and the cost of the work. The essential part of the exploratory study is the estimation of the correlation between key questions and the various survey characteristics. For those characteristics that are not correlated with any of the key questions the only possibility is probably to assume that nonresponding units are the same as the responding units and to adjust accordingly.

In order to illustrate the type of situation that may arise here, let us consider data in Table 28. A family budget survey of 1,000 households is assumed here. Out of the 1,000 households selected, 950 were interviewed, and for another 50 data were not available on account of refusals. The

TABLE 28. - RESULTS OF A HYPOTHETICAL SAMPLE SURVEY OF THE FAMILY BUDGETS OF 1,000 HOUSEHOLDS

Monthly rent (in U.S. $)	Average monthly income (in U.S. $)	Number of successful interviews in the class	Distribution of refusals	
			I	II
(1)	(2)	(3)	(4)	(5)
Less than 10	121	5	0	0
11 - 20	136	20	1	0
21 - 30	203	183	4	1
31 - 40	260	430	27	0
41 - 50	245	274	15	0
51 - 60	302	20	1	2
61 - 70	398	7	0	6
71 - 80	418	5	1	21
81 - 90	410	3	1	13
91 - 100	417	1	0	3
100 and over	950	2	0	4
Totals		950	50	50

information on a number of key questions was available for all the households including refusals. One of these key questions concerns the monthly rent. In Table 28 the first column contains the distribution of monthly rents. Column 2 contains the average monthly income for each of the available monthly rent groups for the 950 responding households. Column 3 is the distribution of the available responses according to this monthly rent. Columns 4 and 5 are two possible distributions of refusals according to the information available on monthly rent.

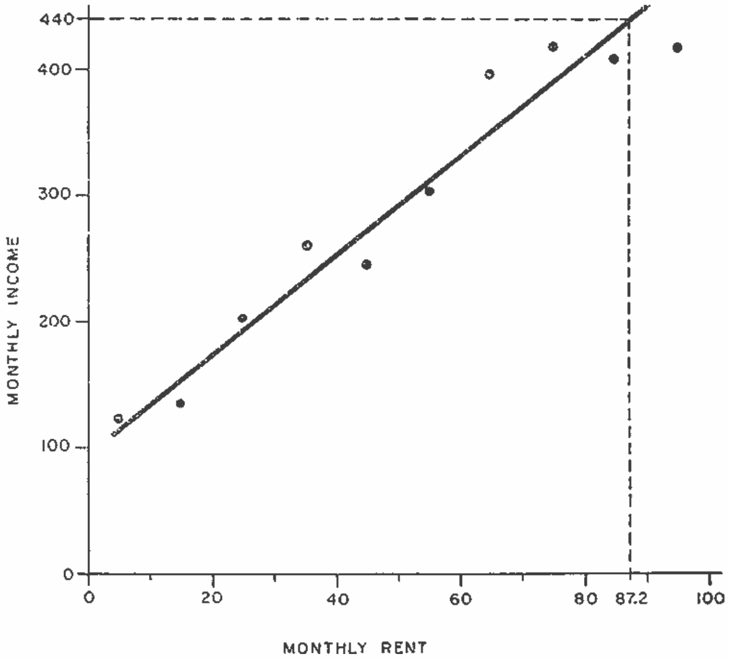
As presented here, monthly rent and income are correlated. Figure 6 shows data in these two columns plotted and the corresponding regression line located freehand. Data in the last row of Table 28 are not included in the chart. The position of the regression line is determined here on the assumption that all the points on the chart had equal weight.

Let us now suppose that an estimate of the monthly income is needed. The first consideration here is whether refusals have to be taken into account or not. The distributions in columns 4 and 5 of Table 28 will be useful in this respect. If the distribution of refusals is as in column 4, it may be considered the same as the distribution of the responding households in column 3. However, if the refusals are distributed as in column 5, it becomes obvious that they are not randomly distributed. The average monthly rent for 50 refusals in this case is as high as $87.2, which is very different from the corresponding average for 950 units, which is $37.8. A simple way of allowing for the bias which is likely to appear here is to draw a vertical line on the x axis at the point 87.2 and take the ordinate of the intersection with the regression line as an estimate of the average monthly income for the 50 missing households. With the regression line positioned as in Figure 6 this estimate is $440. Accordingly, the estimate of the average monthly income per holding for the 1,000 units would be

$$\frac{950 \times 246.2 + 50 \times 440}{1\,000} = \$255.9$$

where $246.2 is the average monthly income per household for the 950 responding households.

FIGURE 6. - RELATIONSHIP BETWEEN THE MONTHLY INCOME AND THE MONTHLY RENT



At this point mention must be made of special attempts at the solution of the question of missing data under conditions where information for several variables is known and some assumptions are possible as to the relationship between the variables concerned. In such cases models can be constructed which reflect the totality of the available facts and hypotheses. These models are then used in dealing with the missing data. G. Eklund, in his *Studies of Selection Bias in Applied Statistics*,[36] gives examples of various models as well as information on their application. It appears however, that such complex approaches can hardly be used outside specialized studies and more delicate research problems where past experiences offer sufficient guidance for building up models and hypotheses about the population.

If adjustments of data are not undertaken for reasons of simplicity it may be useful to publish data on key questions by provinces and various administrative units. This offers a chance to users to make their own adjustments if they so wish.

The final choice of technique depends upon a thorough consideration of all the relevant facts. For example, key questions have no place in a mail survey. In countries just starting their statistical activities, data

---

[36] Published by Almqvist and Wiksell, Uppsala, 1960, Second edition.

from independent sources cannot be used to establish the relationships that might be useful in making allowances for missing data. With several re-calls, collecting data on key questions may be a relatively simple and cheap operation. If electronic computers are available for data processing a larger range of adjustments may be possible. Furthermore, the decision on the type of adjustment to be made will also depend upon the uses of data which may call for anything from very rough estimates on the one hand to high-quality data on the other.

To conclude this chapter it should be added that, wherever applicable, a section in the survey report may usefully be devoted to the problem of missing data. Little attention has been paid to this question in the past and the user of data is often left with very vague ideas as to the incidence of missing data, its effect and its treatment, if any. The inclusion of such a section in the report encourages statisticians to study the question thoroughly and to report on the results achieved.

8. THE RESPONDENT

8.1 Introductory remarks

When the respondents start to answer the questions included in the survey program, an extremely complicated machinery of psychological processes often starts working. Thus, even if they know the accurate replies to the questions, they may deviate from the truth for various reasons and present an inaccurate response. Whenever the questions touch their integrity, authority and dignity, they will automatically tend to formulate the response in such a way as to maintain prestige. It is a well-known fact that expenditure on alcoholic drinks, as reported in interview surveys, falls far below the actual sales figures. On the other hand, expenditure on prestige improving items, such as books, theaters, concerts, etc., is overreported. In other words, in formulating their answers respondents may be guided by many different factors, some of which lead to deviations from the true values.

The present chapter is devoted to a discussion of errors in statistical data resulting from the respondents' characteristics as human and social beings. These errors make up a large part of what is called response errors.

It is obvious that the subject matter of this chapter stands on the border line between statistics and some other sciences, such as psychology and sociology. The knowledge available at present on this type of error is limited; a systematic study of the problems involved has only just started. Some of the results achieved will be presented here.

8.2 Intellectual background

In formulating a reply to survey questions the respondents will use all the elements of their *knowledge*. For example, many people do not know their age. Before they refuse to answer, they will often try to find in their memory certain facts which might be useful in arriving at an approximately correct response. The author, on occasions, has heard answers of the following type: "My mother was telling me that I was born two days before my neighbor, X," or "I was born after my father went to the war against... ." Obviously, the respondents are attempting to locate the relevant events in time.

Similar attempts are frequently met with in statistical practice. The respondent who does not know the yield of his crop may try in a number of ways to arrive at the figure asked for. He may know a friend's yield which he considers the same as his own. So he will give the same answer as his friend. He may have stored his crop in a number of sacks and in answering the question on yield he may simply evaluate the volume of the sacks and then convert the volume into weight. In expenditure surveys a person may evaluate the average price of his lunch and multiply this figure by 30 to obtain the monthly expenditure for lunch. In a study of response errors conducted by the United States Bureau of the Census it was found that some respondents had obtained their figures on crop production in the following way. In a county farmers were asked to report, for experimental purposes, areas harvested as well as bushels harvested (production). After data had been collected the figures on bushels harvested were divided by areas and the following result was obtained: "In 55 percent of cases for oats and 49 percent of cases for soybeans, the quotient was *exactly* an integer (e.g., exactly 32 bushels per acre). Furthermore, 29 percent of the oat reports gave yields of 40, 50 or 60 bushels per acre. Twenty-eight percent of the farms reporting soybeans show yields of exactly 20, 25, or 30 bushels per acre. It is a reasonable inference that, in these cases, production was obtained by multiplying acreage by estimated yield per acre." [1]

Other important factors which also appear in the formulation of the response are the respondent's *own ideas* about the meaning and aims of various survey questions. These ideas inevitably affect the respondent's answers. If he misinterprets the meaning and aims of the questionnaire this may lead to errors unless they are dispelled by the survey instructions.

In any field of statistics many examples can be found which show the importance of this source of errors. A person living in a rural area who, in addition to his work in agriculture, is also exercising a number of other activities to increase his income, might tend to answer that agriculture is his main type of activity even if he is drawing more income from, and devoting more time to, other activities. In his opinion, the fact that he lives in a village among a predominantly agricultural population is the important one and he therefore disregards the additional activities that the survey may need.

In the same way a woman may not take stillbirths into account when answering a question on the number of children she has borne. According to her interpretation this refers only to the children born alive.

This source of error becomes much more important if (*i*) questions are not clearly worded, and (*ii*) the topic is difficult to define in an unambiguous way. In these cases, instead of having outside guidance when answering questions, the respondent is left largely to his own ideas and interpretations.

A typical example of this kind can be found in literacy statistics. If respondents are simply asked whether they are literate, the problem arises as to what constitutes literacy. In the absence of a clear definition the respondents supply their own version of this concept. To remove such arbitrariness a definition has been used in many recent censuses qualifying literacy as "ability both to read with understanding and to write a short statement on everyday life in any language." Although a number of respondents do not care for such definitions and answer immediately according to their own ideas, this definition contributes appreciably to a better understanding of what is meant. However, possibilities for various interpretations still remain because it is not clear how developed this ability must be to qualify a person for the class of literates. Should a person who needs ten minutes to read and understand a sentence be included? For most practical purposes such a person is illiterate. This difficulty leads to a concept of functional literacy which establishes that a person is "... literate when he has acquired the knowledge and skills in reading and writing which enable him to engage effectively in all those activities in which literacy is normally assumed in his culture or group."[2]

Although this definition does not establish a clear border line between illiteracy and literacy, it represents an attempt to elucidate the concept and thus reduce the role of the respondent's personal interpretation.

The respondents' personal ideas are responsible for many other statistical misunderstandings. After finding a considerable difference in the yield of corn as obtained by weighing and by means of farmers' reports, Hendricks says that this "does not necessarily imply that farmers' reports

---

[1] Marks, E.S. and Mauldin, P.W.: Response errors in census research, *Journal of the American Statistical Association*, Vol. 45, 1950, pp. 424-438.

[2] Unesco: *World Illiteracy at Mid-Century*, Paris, 1957.

were incorrect from their own viewpoint. Farmers' concepts of production and yield may differ substantially from statistical definitions. They may contain allowances for wastage during future months, for deficiencies in quality or feeding value, for shrinkage as the crop dries out, or other factors. Unfortunately, little is known about the basis on which farmers report."[3]

In the same paper Hendricks shows the difficulties encountered with the concept of a farm. In a survey conducted in 1954 a farm was defined as "all the land under the control of an 'operator', including land worked by sharecroppers but excluding land operated by other types of tenants." However, the analysis of data has shown that respondents also include land rented to others in area statistics. They obtain a part of the crop from such land and tend to think that these areas belong to their farms in spite of the definition. The difficulty was solved in the next survey by letting farmers report all land, including areas rented out to tenants. In the processing the latter were subtracted from the corresponding totals in order to apply the above definition of the farm.

In another study mentioned in the same paper Hendricks found that farmers, in their reports about areas, sometimes use net areas and sometimes gross areas.[4]

All these examples show that in statistical surveys the respondents draw extensively upon their intellectual background in the formulation of the response. It is useful for statisticians to know the relevant characteristics of this background and the way in which they will influence the formulation of the response. This knowledge will facilitate the preparation of concepts, definitions, instructions, etc.

8.3 Social background

Living in distinct social groups people absorb the opinions, attitudes, and practices of their respective groups, which are called here by the

---

[3] Hendricks, W.A.: Nonsampling errors in agricultural surveys, *Improving the Quality of Statistical Surveys*, Papers contributed as a memorial to Samuel Weiss, American Statistical Association, Washington, 1956, pp. 31-39.

[4] For discrepant definitions and their consequences, cf. Marshall, J.T.: A comparison of some of the census concepts used in Canada and the United States, *Proceedings of the Social Statistics Section, American Statistical Association*, 1961, pp. 81-91 and the National Sample Survey: *Report on Morbidity*, No. 49, Government of India, The Cabinet Secretariat, 1961, where a comparison is made of concepts and definitions used in health surveys in different countries.

common name of "social background." It is a very well-known fact that social background influences many of our reactions. It is also directly responsible for many response errors. A good illustration of such influences is furnished by You Poh Seng.[5] The illustration relates to age data, where one would not expect the social background to play such an important role.

TABLE 29. - PERCENTAGE DISTRIBUTION OF THE POPULATION IN SINGAPORE AS OBTAINED IN SEVEN CENSUSES [1]
(Ages 0-4 and 5-9 years)

Censuses / Age groups	1881	1891	1901	1911	1921	1931	1947
0 - 4	5.4	4.8	5.8	5.5	6.6	8.8	12.2
5 - 9	6.2	5.2	6.1	6.5	7.5	9.4	12.6

[1] You Poh Seng: Op. cit.

The study of the percentage distribution of the population in Singapore for the ages 0 to 4 and 5 to 9 years as obtained in seven censuses gave the results shown in Table 29. A comparison of the percentages in this table shows that the group 5 to 9 years is consistently larger than the group 0 to 4 years, which is contrary to expectation. After having eliminated all other possible explanations for this unexpected phenomenon, You Poh Seng found that it was due to the influence of Chinese cultural background. In fact, the Singapore population contains around 80 percent of Chinese who reckon ages in a different way from that adopted in western civilization. Censuses were always prepared in accordance with the Western way of thinking and age questions were put in the form of "What is your age?", "Age at the last birthday" or "Age according to Western reckoning." The Chinese population apply a different calculation. "According to the traditional method of age counting, a Chinese is one year old at birth, and thereafter becomes a year older at every Chinese New Year. Since the Chinese New Year is based on a form of lunar calendar, New Year's Day varies from year to year, but normally it falls round about February. An extreme case would be the following: a child is born,

---

[5] You Poh Seng: Errors in age reporting in statistically underdeveloped countries, *Population Studies*, Vol. 13, 1959, pp. 164-182.

say, a week before the Chinese New Year. It is one year old at birth, and one week later, on the occasion of the New Year, it becomes two years old, whereas strictly speaking it is only one week old."[6] According to Chinese tradition there can be no child under one year. Consequently, the resulting biases in age data cannot be eliminated without adapting census concepts to this tradition or taking some special steps for converting data based on Chinese reckoning into data based on the Western way of counting years.

The response will also be influenced by the *criteria* and *practices* followed in the respondent's daily life. In our daily life we often speak in terms of various approximations. In referring to prices, salaries, amounts of ready cash, etc., there will be a tendency to neglect small units. In referring to the age of an adult person, months and days are not counted, etc. People often stick to a kind of average, to what is usual in the social group. It has often been observed that in eye estimates of yield there is a tendency to reduce the magnitude of deviations from the average. In an experimental survey conducted by the United States Bureau of the Census the question was asked: "Do you have any fruit or nut trees or grapevines?" Twenty respondents replied "No." The original question was later supplemented by the following question asked of the same 20 farmers: " Do you not even have one or two regardless of whether they are young trees or vines or whether or not they produce any fruit to speak of? Do you have any at all?" It was found that 13 out of 20 changed the original answer and agreed they had some trees. Respondents are not always "literal minded" and their opinion might be that: "You would not be interested in those old trees."[7]

Similarly, certain activities, expenditures, sources of income or persons such as babies and old people are not reported because the respondent assumes that the item concerned has no practical importance.

Rounding off can also be considered as an aspect of the application of the practical criteria of the group in the preparation of the response. Rounding off is very common in population statistics. If age data, as obtained in a census of population, are tabulated according to the last digit of the number of years completed, the frequency of some digits, primarily 0 and 5, will be relatively high. An illustration from United

---

[6] You Poh Seng: Op. cit.

[7] Marks, E.S. and Mauldin, P.W.: Problems of response in enumerative surveys, *American Sociological Review*, Vol. 15, 1950, pp. 649-657.

TABLE 30. - RELATIVE FREQUENCIES OF AGES ENDING IN VARIOUS DIGITS IN THE UNITED STATES CENSUSES OF POPULATION[1]

Digit of age	Census							
	1880	1890	1900	1910	1920	1930	1940	1950
0	16.8	15.1	13.2	13.2	12.4	12.3	11.6	11.2
1	6.7	7.4	8.3	7.7	8.0	8.0	8.5	8.9
2	9.4	9.7	9.8	10.2	10.2	10.3	10.4	10.2
3	8.6	9.1	9.3	9.2	9.4	9.4	9.6	9.7
4	8.8	9.0	9.5	9.4	9.4	9.6	9.7	9.7
5	13.4	12.3	11.3	11.5	11.3	11.2	10.7	10.6
6	9.4	9.6	9.4	9.6	9.7	9.6	9.6	9.8
7	8.5	8.9	9.3	9.1	9.4	9.3	9.6	9.7
8	10.2	10.4	10.2	10.7	10.6	10.5	10.3	10.2
9	8.2	8.5	9.7	9.4	9.6	9.8	10.0	10.1
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

[1] This table is reproduced from Myers, R.J.: Accuracy of age reporting in the 1950 United States Census, *Journal of the American Statistical Association*, Vol. 49, 1954, pp. 826-831.

States censuses of population is given in Table 30. It can be seen from this table that the frequencies of ages ending with different digits vary considerably, although there is a clear tendency toward a decrease in these variations in more recent censuses. A similar result is also obtained if data on the date of birth are tabulated by the last digit.[8]

Rounding off is also found in all other fields of statistics whenever answers to questions are reported in figures. In the paper by Marks and Mauldin quoted above it can be seen that in 29 percent of farms growing oats the yield reported was exactly 40, 50 or 60 bushels per acre. In 22 percent of these farms the production of oats was given in figures ending in 00 and in another 10 percent in figures ending in 000. Altogether, 60 percent of the reports on oats show evidence of rounding off. The corresponding figure for soybeans is 47 percent.

The same tendency to round off figures has been reported in many other fields.[9] Rounding off is a necessity in practical life and is a deeply rooted habit. How could decimal figures be dealt with without rounding off? We are taught to apply the procedure of rounding off from a very early age, in school as well as in the course of life. It is a way of expressing approximate judgments. To report the production of a crop by means of a figure ending in 000 is another way of saying: "I do not know the exact production. In my opinion it should be about ... ."

Rounding off is also governed by established social standards. The digits 0 and 5 have a somewhat special role in social practice. We have banknotes and coins of 100, 50 and 5. This fact has considerable consequences in many aspects of our daily life. The author found some kind of rounding off in data on the degree of education in the 1953 Census of Population in Yugoslavia. There were only three important grades of education: elementary school, secondary school, and university. These three grades were used in classifying employees, in determining salaries, and for a large number of other purposes. Thus, they represented some kind of social standards for measuring the degree of education. As such, they governed the answers to the census questions and the respondents tended to think that no further information was needed. A sample check of census data conducted by an examination of documents and school certificates showed that some people had changed their grade of education by rounding to the nearest socially recognized grade of education. Many people who had completed six years of secondary school reported that they had completed this schooling. Similarly, those who had one or two years of university education rounded off to the completed secondary school.[10] Social standards shape the respondent's way of thinking and of interpreting phenomena and situations. Therefore, the respondent applies them not only in cases where he has to guess at answers but also on some occasions when he can give an accurate answer.

---

[8] Cf. Zarkovich, S.S.: *Population census errors* (in Serbo-Croatian), Federal Statistical Office, Belgrade, 1954.

[9] Cf. Bachi, R.: The tendency to round off age returns: measurement and correction, *Bulletin of the International Statistical Institute*, Vol. 33, Part 4, 1951, pp. 195-222.

8.4 Emotional background

It is well known that people, in whatever they do, normally try to avoid any act that might in any way injure their reputation. Following this characteristic, respondents deliberately change their answers to achieve a particular purpose. These purposes vary according to circumstances and it is becoming increasingly important in modern statistics to learn as much as possible about the circumstances giving rise to such a deliberate supply of inaccurate data, the survey items usually affected, the magnitude of the resulting errors or biases and the means of removing such errors.

Circumstances giving rise to inaccurate data are frequently found in countries where statistical data are used for nonstatistical purposes, such as income *taxation* or an obligation imposed on agricultural holdings for deliveries in kind. Under such circumstances, the respondents consider that it is against their interests to supply accurate information, and so often try to avoid doing so.

The results are similar if respondents are *afraid* of the survey in a general way, without having in mind any particular action that may ensue. The respondents' reaction may then consist of supplying inaccurate data on most survey items. Such cases were reported from colonial or occupied territories where a foreign power was sponsoring or conducting surveys. If opinion or attitude surveys are conducted under such circumstances, in the interests of self-protection the respondents may adapt their answers to the views and wishes of the government or foreign power.

Respondents' emotions also lead to errors if survey questions are connected with *stereotypes*, such as patriotism, communism, fascism, dicta-

---

[10] Zarkovich, S.S.: Op. cit.

torship, religion, names of famous politicians, soldiers, scientists, writers, actors, etc. People often react emotionally if they hear or read the stereotypes and their answers may be influenced accordingly. Menefee[11] has reported some experimental data which show how far the influence of stereotypes may go. In his experiment 742 persons were presented with a number of statements on various social issues and each of them had to answer questions either by "Yes" or "No." Later on, each of these issues was qualified as being typical of patriotism, communism, etc., and the same persons were again asked to give their answers. Considerable changes in the original answers occurred according to the respondent's attitude toward the stereotype involved.

Inaccurate data are also often the result of the respondent's desire to *impress*. Some such desire exists in everybody. People normally do not present facts about themselves which might be harmful to their dignity and provoke unfavorable comment. Their unwillingness to disclose such facts in statistical surveys result in what are known as *prestige errors*.

Many examples of prestige errors are known in statistical literature. It is well known that women do not like to disclose their age and often declare themselves younger, while young people often declare themselves older. Illiterate people report that they are able to write and read. Some people raise the level of their education, others the grade of their occupation. Thus medical assistants become medical practitioners, operators become foremen, bank employees become directors, etc. Similarly, people often exaggerate the salaries they receive and the rents they pay, the price of their food and clothing, the amount of money they spend on books, concerts, theaters and other items showing their cultural interests. These are all cases of prestige errors leading to upward biases. Errors of the opposite type are also well known.

It has already been pointed out that some types of expenditure are not reported fully. "It is a commonplace for expenditure surveys to produce under-estimates of consumers' purchases of alcoholic drink and tobacco. From our own survey it would appear that Cambridgeshire households' expenditure on these two items is for alcoholic drink more than 50 percent, and for tobacco 30 percent below what we estimate to be the national average ... We can be fairly sure that a good deal of this under-estimate results from people's unwillingness to reveal just how much they do spend on these two items."[12]

In some respects prestige errors are different from other errors in statistics. First of all, they may affect all types of items on the survey program, factual items, opinions, attitudes, etc., according to the situation. In addition, they appear both when accurate information is available to the respondent and in cases where answers have to be worked out. It should also be added that they appear rather freakishly; they are difficult to predict and even more difficult to detect in data. If ignorance or fear are the sources of such errors, the statistician's knowledge of the country will give him some guidance as to what to expect. Prestige errors are to a large extent a result of individual idiosyncrasies. In some cases they take ridiculous forms. It has already been mentioned that a considerable proportion of respondents will declare that they have read a nonexistent novel if it is referred to as a famous work by a known author.

8.5 Memory errors

The accuracy of statistical data depends to a large extent upon how exactly past events are remembered. In statistical surveys the holder of an agricultural holding is sometimes asked how many days he worked off his holding during the past year; he is asked for his total income during the year, for his expenditure during some specific period of time, etc. In all these and similar cases the respondent is obliged to recollect the relevant events and build up the answer gradually. The accuracy of his answer obviously depends upon the accuracy of his memory. The errors in data that appear as a result of memory lapses will be referred to by the name of *memory errors* while the resulting bias will be called the *memory bias*.

Clearly, the mechanism of the memory is of great interest to the statistician. Any knowledge on this subject will be useful in planning measures for removing memory errors. Some experimental results that might offer guidance in designing surveys will therefore be presented here.

---

[11] Menefee, S.C.: The effect of stereotyped words on political judgments, *American Sociological Review*, Vol. 1, 1936, pp. 614-621.

[12] Cole, D.E.: Field work in sample surveys of household income and expenditure, *Applied Statistics*, Vol. 5, 1956, pp. 49-61; and Cole, D.E. and Utting, J.E.G.: Estimating expenditure, saving and income from household budgets, *Journal of the Royal Statistical Society*, Vol. 119, Series A, 1956, pp. 371-392.

Some characteristics of the memory are very well known. Firstly, most events fade gradually from the memory so that progressively fewer details can be recalled. Secondly, the speed of the fading process varies from one item to another. Some events remain forever impressed on the memory while others fade fast and disappear completely. People tend to remember forever facts of basic importance in their lives, such as accidents, war injuries and other very pleasant or unpleasant events. On the other hand, they tend to forget events of a routine and recurring nature. For example, they will not remember what they ate for lunch on a particular day unless there was something especially memorable about that lunch.

This type of general knowledge about memory does not help much in statistics. In order to estimate the effect of memory errors we need quantitative data on the speed of the fading process and the part of past events that can be recalled.

Although this question has considerable practical importance few detailed studies are available at present. From among these results some material assembled in the Indian Statistical Institute will be used here for illustration. This material suggests a definite relationship between the distance in time of the events involved and the ability to recall these events. Mahalanobis and Das Gupta[13] drew attention to this question in 1954. In the paper quoted they presented data on the sex ratio of children for different marriage cohorts by taking into account only the children to the fourth birth. These data are shown in Table 31.

Column 1 of this table shows the period when the couples under study got married, and column 2 the number of couples in each cohort. In the last column is the sex ratio as computed on the basis of data supplied by respondents on the sex of their children. The sex ratio of 107 in the most recent periods corresponds to data available in other countries and is considered accurate. In addition, it refers to recent events which are sufficiently well remembered. Since there is no evidence that the sex ratio could change over such a short period to the extent indicated in the table and it is not possible to assume that female children born in more remote periods were reported as males, the trend in the sex ratio was explained as a result of selective sex recall. Namely, in Indian culture

---

[13] Mahalanobis, P.C. and Das Gupta, A.: The use of sample surveys in demographic studies in India, *Proceedings of the United Nations World Population Conference, 1954*, Vol. 6, pp. 363-384.

TABLE 31. - SEX RATIO (NUMBER OF MALE CHILDREN PER 100 FEMALE CHILDREN BORN UP TO THE FOURTH BIRTH) FOR DIFFERENT MARRIAGE COHORTS [1]

Period of marriage	Number of couples	Sex ratio
Up to 1909	910	147
1909 - 1919	945	127
1920 - 1929	1 459	143
1930 - 1939	2 757	108
1940 - 1945	2 204	107
1946 - 1951	1 001	107

[1] Mahalanobis, P.C. and Das Gupta, A.: Op. cit.

TABLE 32. - NUMBER OF CHILDREN DYING IN THE FIRST YEAR OF LIFE PER 1,000 CHILDREN BORN [1]

Period of marriage	Number of couples	Number of infant deaths
Before 1909	2 177	88
1910 - 1919	2 415	102
1920 - 1929	3 612	126
1930 - 1939	4 652	134
1940 - 1945	3 306	133
1946 - 1951	3 714	181

[1] Ibid.

the birth of a male child represents a much more important event than the birth of a female child, and is therefore better remembered. Female children are easily forgotten, particularly if they die very young or are stillborn.

The authors present similar results from a study of infant mortality. The relevant data are shown in Table 32. Here again the tendency to underreport is obvious as one proceeds toward older marriages.

On the basis of such data one could attempt to formulate the relationship between the time lapse and completeness in reporting the events. In fact, for data collected in the fourth round of the National Sample Survey on infant mortality, such relationships were established by Das Gupta, Som, Majumdar and Mitra.[14] The curve showing the relationships between the percentage of underreporting of infant deaths and the age of the marriage

---

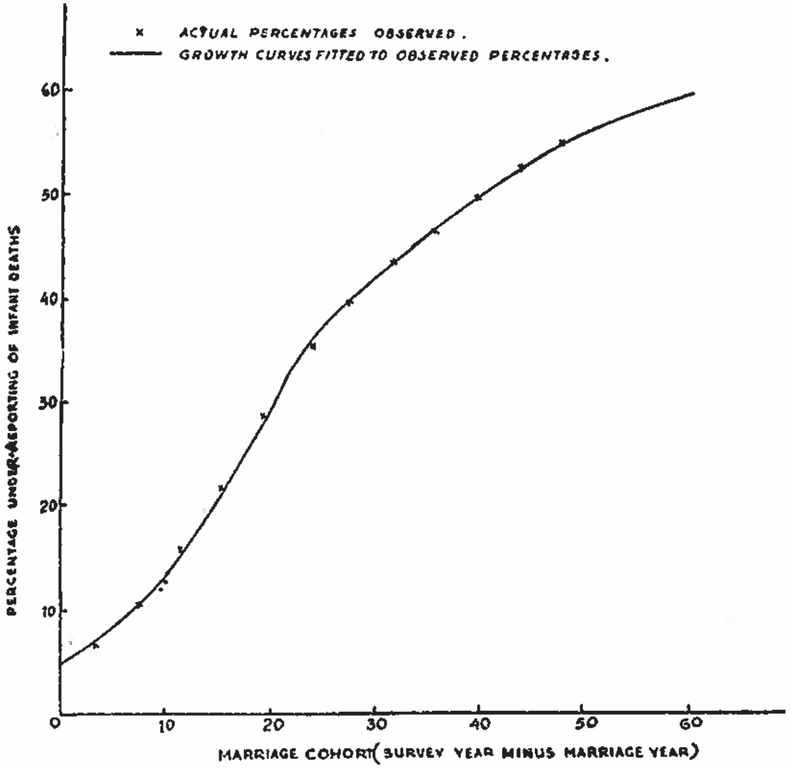
[14] Das Gupta, A., Som, R.K., Majumdar, M. and Mitra, S.N.: *Couple Fertility*, The National Sample Survey, No. 7, Government of India, New Delhi, 1955.

cohorts is given in Figure 7. A good conformity with the data observed was obtained by using two growth curves of the form

$$y = \frac{L}{1 + e^{\alpha + \beta x}}$$

with the common assymptote $L = 0.65$ and the intersecting point at $x = 23.07$. The constants of the curve for the lower part are $\alpha = 2.5025$ and $\beta = -0.1153$. The same constants of the curve for the upper part of the graph are $\alpha = 1.3084$ and $\beta = -0.0635$.

FIGURE 7. - RELATIONSHIPS BETWEEN THE PERCENTAGE OF UNDERREPORTING OF INFANT DEATHS AND MARRIAGE COHORTS [1]



---

[1] This figure is reproduced from Das Gupta, A., Som, R.K., Majumdar, M. and Mitra, S.N.: Op. cit.

In the seventh round of the National Sample Survey a study was made of memory lapses in reporting birth and death cases. Sample households had to report the births and deaths in the year preceding the survey as well as the month when these events occurred. This made it possible to tabulate the data collected according to the lapse of time in months between the survey and the event. Table 33 shows the decreasing percentage of births reported in the year preceding the survey as the time lapse increases.

TABLE 33. - PERCENTAGE DISTRIBUTION OF THE NUMBER OF BIRTHS REPORTED BY THE MONTH OF OCCURRENCE [1]

Month of birth prior to the survey	Percentage
1st	11.25
2nd	9.62
3rd	9.44
4th	9.18
5th	8.97
6th	9.15
7th	8.40
8th	7.84
9th	7.51
10th	6.97
11th	6.43
12th	5.24
	100.00

[1] Cf. Mazumder, M.: *Vital Rates*, The National Sample Survey, No. 54, The Cabinet Secretariat, Government of India, New Delhi, 1962.

Table 34 shows another interesting result of this study. In this table data from the preceding table are tabulated by various zones of India. In some zones the decline of percentages with time is obvious, while in others, such as the east and south, it is not obvious. This calls for further studies and a larger sample. It may well be that recall lapses are not simply governed by the passing of time.

If such data are used for the computation of birth rates, death rates and the growth rate of the population, etc., results will be biased. Som,[15]

---

[15] Som, R.K.: On recall lapse in demographic studies in India, *International Population Conference*, Vienna, 1959.

TABLE 34. - PERCENTAGE DISTRIBUTION OF THE NUMBER OF BIRTHS REPORTED BY THE MONTH OF OCCURRENCE AND ZONE[1]

Zone	Month of occurrence prior to date of survey				Number of births
	1st-3rd	4th-6th	7th-9th	10th-12th	
North	30.6	27.9	27.8	13.7	242
East	27.8	25.1	25.6	21.5	386
South	23.1	31.9	24.9	20.1	157
West	33.4	30.8	20.8	15.0	202
Central	31.3	26.1	20.0	22.6	258
Northwest	40.3	25.6	19.1	15.0	205
All rural India	30.2	27.3	23.8	18.6	1 450

[1] Mazumder, M.: Op. cit.

however, has used the above data from the seventh round of the National Sample Survey to adjust the rates presented for the recall lapse. The essence of his technique is the following:

Restricting the technique to birthrates, b_{ik} will be used to denote the number of births reported in the i-th month of the survey as having occurred in the k-th preceding month with $i = 1, 2, \ldots, 12$ and $k = 1, 2, \ldots, 12$. The purpose of taking the survey period to be one year is to make possible the elimination of seasonal variations. The true value of births is then assumed to be

$$\beta = b_{ik} + s_j + d_k + e_{ik}$$

where the symbols have the following meaning:

s_j = seasonal component
d_k = distortion factor (a function of recall period k)
e_{ik} = random error.

Assuming

$$\sum_j s_j = 0$$

and

$$E(e_{ik}) = 0$$

the total number of births in the k-th month preceding the survey is

$$\sum_i b_{ik} = 12\,\beta - 12\,d_k$$

Similarly, the average number of births in the k-th preceding month is

$$\bar{b}_k = \beta - d_k$$

The values of $\bar{b}_k$ are smoothed by means of a curve of the type

$$y = ce^{-ak^2}$$

Assuming $d_o = 0$, it is then possible to estimate $\bar{b}_o = \beta$ or, in other words, the number of births that would be obtained had the enumerator come to call at the moment of birth, when the process of fading out had not yet started. It is also possible to estimate the magnitude of the memory bias in the k-th month by

$$\begin{aligned} d_k &= \beta - \bar{b}_k \\ &= \bar{b}_o - \bar{b}_k \end{aligned}$$

Table 35 contains the three rates as obtained by means of the above theory. Figures are expressed as indexes with the value at recall period zero taken as 100. The rate of growth as shown is taken as the balance between the birthrate and the death rate. An estimate of the rates which would be free from memory biases is given by their values at recall period zero. In this particular case these values are 40.87, 23.98 and 16.89 respectively.[16]

In another study undertaken by David[17] which might also be considered a study of memory errors, the quality of the reported income from Welfare Assistance was checked by looking up the corresponding information in the files for a sample of families. The check has shown that people report on this component of income according to the equation

$$X = 0.379\ Z + 1074$$

---

[16] Cf. also Som, R.K. and Das, N.C.: On recall lapse in infant death reporting, *Sankhyā*, Vol. 21, 1959, pp. 205-208; "The technical paper on non-sampling errors and biases in retrospective demographic enquiries," prepared for the Seminar on vital statistics, conducted by the Economic Commission for Africa in December, 1964, (Document E/CN.14/CAS.4/VS/3).

[17] David, M.: The validity of income reported by a sample of families who received welfare assistance during 1959, *Journal of the American Statistical Association*, Vol. 57, 1962, pp. 680-685.

TABLE 35. - INDEX OF BIRTH, DEATH AND GROWTH RATES FOR DIFFERENT RECALL PERIODS [1]

Recall period (months)	Index of rates		
	Birth	Death	Growth
0	100	100	100
1	99.8	99.6	100.1
2	99.4	98.5	100.5
3	98.7	97.0	100.1
4	97.8	94.7	102.2
5	96.6	92.5	102.5
6	95.3	89.7	103.3
7	93.7	86.6	103.9
8	92.0	83.2	104.5
9	90.1	79.8	104.9
10	88.1	76.3	105.0
11	86.0	72.8	104.9
12	83.8	69.3	104.4

[1] Som, R.K.: Op. cit.

where X stands for the true income and Z for its survey value. It was also found that poor people report accurately probably because this is the only or the most important part of their income. Richer people underreport this source of income probably because they have many other sources.

8.6 Length of the period of reference

By the term "period of reference" in statistical surveys we understand the period of time to which the data collected refer. This period may vary in length. From data presented in the last section and particularly from Figure 7 it follows that the longer the reference period the more important becomes the effect of memory errors. This is why the *length* of the reference period must be considered as a factor that has an important bearing on the quality of data.

The problem of the period of reference arises whenever data have to be collected for a *period of time* rather than for a given *point of time.* Although the point of time may be preferable in some respects, meaningful statistical information cannot be obtained for some characteristics unless data are requested for a period of time, such as a week, a month or a year. Such characteristics are income, production, use of fertilizers, expenditures, etc.

The problem of the period of reference has two aspects. The first concerns its *length.* The second concerns the *location* in time of the period selected. If a day has been selected as the period of reference, it can be the day of the survey, the preceding day, the day before that, etc. In each survey where the problem of the reference period is involved a decision is needed regarding both aspects.

If the "law" of the memory fading process can be established as in Figure 7, the decision is relatively simple. The longer the period of reference in this situation the more inaccurate are the data collected. In other words, the period of reference should be as short as possible. As to its location, this should be as close as possible to the moment of collecting data because this is the direction in which the magnitude of memory biases decreases.

The situation, however, is not always as simple as this. The choice of the period of time to be covered in a survey is very often restricted for a number of reasons. For example, some events occur in more or less regular *cycles.* In a survey of such events it is impossible to use a period of reference that would be shorter than these cycles. If salaries are paid monthly, a survey of income cannot use a period of reference shorter than a month unless some special design has been prepared. If there are considerable variations in income from one season to another, the most convenient period of reference might be a year. This will be particularly so in nonrepetitive surveys when cost and other considerations do not permit sampling in time. Cost elements often involve long reference periods in the case of characteristics that have to be presented on an annual or some similar basis. For example, the quantity of fertilizers used has to be related to the crop year if the relationships between production and the use of fertilizers are to be studied. Expenditures on heating cannot be properly evaluated unless data are presented on an annual basis. For reasons of economy such data are often collected in a single survey using the past year as a period of reference.

In the course of the preparations for a survey the length of the reference period has to be decided, keeping in mind such relevant circumstances as the survey program, the natural cycles of the characteristics, the respondents' level of education, etc. Instead of relying on guesswork it may be useful to base these decisions on experimental grounds. In fact, such an approach was used in some studies selected here for illustration.

TABLE 36. - EXPENDITURES IN RUPEES PER MONTH PER PERSON WITH TWO REFERENCE PERIODS, A WEEK AND A MONTH [1]

	Period of reference			
	Week		Month	
	Subsample No. 1 (Rs.)	Subsample No. 2 (Rs.)	Subsample No. 1 (Rs.)	Subsample No. 2 (Rs.)
Food grains	9.56	9.47	8.87	8.66
Other food items	5.82	6.21	5.06	5.02
Total food	15.38	15.68	13.93	13.68
Betel leaves Tobacco, intoxicants	0.82	0.82	0.68	0.65
Fuel and light	1.54	1.59	1.22	1.17
Monthly items	1.97	2.34	2.34	2.21
Clothing (annual)	1.62	1.73	1.79	1.65
Other annual items	2.39	2.42	2.42	2.19
TOTAL EXPENDITURE	23.72	24.58	22.38	21.55

[1] Mahalanobis, P.C. and Sen, S.B.: Op. cit.

The first refers to experiments conducted in the National Sample Survey and reported by Mahalanobis and Sen.[18] In the first round of this survey a reference period of a year was used. Later on it was found necessary to abandon a year and use the week preceding the survey for items such as expenditure on food and a month for clothing and certain other articles of a durable nature. In the fourth round these new reference periods were used with a design consisting of interpenetrating samples and the same data were collected with the two periods of reference, a week and a month. The results obtained for some items are presented in Table 36 for each subsample separately. It can be seen from this table that expenditures obtained with a week as a period of reference are higher than corresponding data obtained with a month as the period of reference.

These differences raise the question as to which of these two periods leads to more accurate data. The answer was sought on an experimental basis. A sample of households was selected and divided into two subsamples and the households in one subsample were interviewed on the consumption of clean rice, pulses, sugar and salt, while in the other, data were obtained by actual weighing. For that purpose the households selected were asked to lay in stocks for about ten days and to draw from these stocks whenever necessary. The enumerators weighed the stocks at the start of the survey and later at certain intervals, so that it was possible to get data for a week as well as for a month. Half the households interviewed were asked for weekly consumption and the other half for monthly consumption. The figures obtained are presented in Table 37. The results are shown on a uniform per person and per day basis. The results obtained by measuring are split up into those based on one measurement a week and those where two measurements were made. Despite certain differences, the data in the last column can be considered to agree satisfactorily with the interviewing data with a month reference period. It can be concluded that in the collecting of data on the consumption of foodstuffs, interviewing with a period of reference of one month preceding the survey gives satisfactory results from the accuracy point of view.[19]

---

[18] Mahalanobis, P.C. and Sen, S.B.: On some aspects of the Indian National Sample Survey, *Bulletin of the International Statistical Institute*, Vol. 34, Part 2, 1954, pp. 5-14.

TABLE 37. - CONSUMPTION OF SOME SPECIFIED ITEMS PER PERSON PER WEEK IN OUNCES AS OBTAINED IN TWO PERIODS OF REFERENCE BY INTERVIEWING AND MEASUREMENT [1]

	Interviewing		Measurement	
	Week	Month	Once a week	Twice a week
Rice	16.21	15.14	15.61	14.87
Pulses	1.56	1.19	1.42	1.19
Sugar	0.23	0.16	0.16	0.14
Salt	1.07	0.76	0.86	0.82

[1] Mahalanobis, P.C. and Sen, S.B.: Op. cit.

Ghosh[20] reported the results of another study conducted in several villages in West Bengal with the aim of comparing the quality of data obtained on expenditures on foodstuffs by using three different periods of reference, namely, a day, a week, and a year. For the purpose of

---

[19] Further details will be found in Mahalanobis, P.C. and Sen, S.B.: Op. cit.

[20] Ghosh, A.: Accuracy of family budget data with reference to period of recall, *Calcutta Statistical Association Bulletin*, Vol. 5, 1953, pp. 16-23.

this study investigators were located in each shop in the villages selected. They had to record each sale and classify all the sales into those made by village residents and those made by nonresidents. Thus, the data obtained in these shops represent true values which were compared with the data on food purchases obtained by interviewing. The relevant figures are shown in Table 38. If data in the last column of this table are taken as true values, it can be seen that the reference period of one year leads to the most accurate data, and that of one day to the most inaccurate data.

TABLE 38. - WEEKLY PURCHASES OF SPECIFIED FOODSTUFFS IN RUPEES PER FAMILY BY DIFFERENT PERIODS OF REFERENCE [1]

Item	Period of reference			Daily estimates on shop basis
	Day	Week	Year	
Pulse	0.35	0.31	0.23	0.21
M. oil	0.77	0.62	0.48	0.42
C. oil	0.14	0.19	—	0.14
Salt	0.07	0.11	0.08	0.07
Gur	0.07	0.10	0.07	0.02
Pan-supari	0.07	0.12	0.08	0.03
Tea-leaf	0.07	0.05	0.02	0.01
Tobacco	0.35	0.20	0.13	—
Bidi	0.35	0.37	0.29	0.21

[1] Ghosh, A.: Op. cit.

Interesting differences in data resulting from periods of reference of varying length were also found in the study of expenditure, savings and income conducted by Cole and Utting,[21] as well as in the experiment carried out by Neter and Waksberg,[22] which will be discussed later.

---

[21] Cole, D.E. and Utting, J.E.G.: Estimating expenditure, saving and income from household budgets, *Journal of the Royal Statistical Society*, Vol. 119, Series A, 1956, pp. 371-392.

[22] Cf. Neter, J. and Waksberg, J.: Measurement of nonsampling errors in a survey of home-owners' expenditures for alterations and repairs, *Proceedings of the Social Statistics Section, American Statistical Association*, 1961, pp. 201-210, and the continuation of the same study by the same authors in "Effect of length of recall period on reporting of expenditures in a survey of home-owners' expenditures for alterations and repairs," *Proceedings of the Social Statistics Section, American Statistical Association*, 1962.

8.7 Continued observation

The reader will certainly have noticed some contradiction between data presented in Section 8.5 and those in Section 8.6. If the curve in Figure 7 is taken as a model of how events fade out, data with a one-day reference period should be more accurate than data concerning longer reference periods. However, figures in Tables 37 and 38 suggest that just the opposite is true. Data based on short reference periods are found to be subject to larger biases than data referring to longer periods.

In this connection another factor should be mentioned. This is the *downward tendency* in the value of the characteristics reported if the observation of the same units is continued over a longer period of time. For example, it was found in expenditure surveys that the average expenditure per item per person is usually higher in the first week of the survey than in the second or the third. To illustrate the point, some data from the 1953-54 household expenditure survey conducted in the United Kingdom are presented in Table 39. For most of the items listed the expenditures in the first week are higher than those of subsequent weeks.[23] This difference is presented in the last column of the table. It can be seen that there is a negative sign for only one item. In order to study the magnitude of the differences between the various weeks Kemsley has established the variables y = week 1 less ½ (weeks 2 + 3) and z = week 2 less week 3 and has tested the significance of y and z. The results of the analysis are presented in Table 40. It will be seen that y is significant for most items listed while this is not the case with z.[24]

The difference in the value of characteristics in Table 39 for the first week and successive weeks is of extreme importance for survey techniques. It raises the question of what data should be accepted and how many weeks the survey should be continued in order to obtain accurate estimates. If data for the first week are biased, as they might be on the basis of

---

[23] See also von Hofsten, E.: A budget survey in Sweden, *Family Living Studies*, International Labour Office, Geneva, 1961, pp. 15-35.

[24] Further information on the same problem will be found in the papers by Kemsley, W.F.F.: Designing a budget survey, *Applied Statistics*, Vol. 8, 1959, pp. 114-123; Kemsley, W.F.F. and Nicholson, J.L.: Some experiments in methods of conducting family expenditure surveys, *Journal of the Royal Statistical Society*, Vol. 123, Series A, 1960, pp. 307-328; Prais, S.J.: Some problems in the measurements of price changes with special reference to the cost of living, *Journal of the Royal Statistical Society*, Vol. 121, Series A, 1958, pp. 312-332, and Turner, R.: Interweek variations in expenditure recorded during a two-week survey of family expenditure, *Applied Statistics*, Vol. 10, 1961, pp. 136-146.

TABLE 39. - VARIATIONS IN HOUSEHOLD EXPENDITURE PER HOUSEHOLD PER WEEK FOR SOME SPECIFIED ITEMS[1]

Item	Week 1	Week 2	Week 3	Week 1 less ½ (Weeks 2 + 3)
	Shillings			
Cereal foods	11.99	11.46	11.60	+ 0.46
Meat	18.11	17.45	17.34	+ 0.71
Fish	2.33	2.18	2.15	+ 0.16
Oils and fats	4.16	4.13	4.08	+ 0.06
Dairy produce	13.07	12.58	12.47	+ 0.55
Fruit and vegetables	10.68	10.20	10.37	+ 0.39
Sugar and preserves	6.48	6.16	6.19	+ 0.31
Beverages	4.22	3.89	3.93	+ 0.31
Other foods	10.25	10.24	10.67	— 0.21
All food	81.29	78.29	78.80	+ 2.74
Drink and tobacco	24.06	23.59	23.36	+ 0.59
Clothing and footwear	33.45	27.10	31.11	+ 4.35
Household durables	17.48	16.86	14.23	+ 1.93
Other goods	17.44	16.52	16.36	+ 1.00
Travel	15.30	13.40	12.84	+ 2.18
Services	20.99	19.55	20.29	+ 1.07
Other payments	9.28	8.40	8.31	+ 0.92
Housing and fuel (taken from records)	6.75	5.60	5.74	+ 1.08
All items	226.04	209.31	211.04	+15.86

[1] This table is taken from Kemsley, W.F.F.: The household expenditure enquiry of the Ministry of Labour, *Applied Statistics*, Vol. 10, 1961, pp. 117-135.

Tables 37 and 38, it may be necessary to conduct this type of survey over a longer period of time and discard the information collected in the first week.

There is also the question of the origin of these differences. *Prestige errors* provide a possible explanation. The psychological mechanism would be as follows: the respondents want to impress and this they can easily do when reporting for short reference periods. With longer reference periods their overreporting would be exposed, because it would be clear that their general position (income, social group) was incompatible with the data reported. In other words, with longer periods of reference another prestige reason, i.e., the desire to avoid such exposure, works toward improvement of the quality of data. In the case of data in Table 39, this would mean that respondents reported their expenditure on a rather high level in the first week and were obliged to reduce it after it was clear that the survey would continue.

TABLE 40. - SIGNIFICANCE OF DIFFERENCES IN THE EXPENDITURE FOR SPECIFIED WEEKS[1]

Items	*y*	*S.E. (y)*	*z*	*S.E. (z)*
Cereal foods	[2]+ 0.46	0.10	— 0.14	0.12
Meat	[2]+ 0.71	0.19	+ 0.11	0.24
Fish	[3]+ 0.16	0.05	+ 0.03	0.06
Oils and fats	+ 0.06	0.05	[2]+ 0.05	0.01
Dairy produce	[2]+ 0.55	0.13	+ 0.11	0.17
Fruit and vegetables	[2]+ 0.39	0.11	— 0.17	0.13
Sugar and preserves	[2]+ 0.31	0.09	— 0.03	0.10
Beverages	[2]+ 0.31	0.06	— 0.04	0.08
Other foods	— 0.21	0.19	— 0.43	0.27
All food	[2]+ 2.74	0.38	— 0.51	0.61
Drink and tobacco	[3]+ 0.59	0.32	+ 0.23	0.45
Clothing and footwear[4]	[2]+ 4.15	1.49	[2]— 3.91	1.54
Household durables	+ 1.93	2.16	+ 2.63	2.31
Other goods	+ 1.00	0.78	+ 0.16	0.51
Travel[4]	+ 0.47	0.63	+ 0.42	0.60
Services	+ 1.07	1.13	— 0.74	1.01
Other payments	[2]+ 0.92	0.26	+ 0.09	0.28
Housing and fuel (taken from records)	[2]+ 1.08	0.38	— 0.14	0.42
All items[4,5]	[2]+13.95	3.13	— 1.77	3.29

[1] This table is reproduced from Kemsley, W.F.F.: Op. cit.

[2] Significant at 1 percent level.

[3] Significant at 5 percent level.

[4] Excluding payments to clothing clubs.

[5] Excluding purchases of motor vehicles.

Such an explanation raises some difficulties. If prestige is the source of the biases it should cause a sizable difference in the expenditure as reported in the first week and in the subsequent weeks. Otherwise the respondent's purpose is not achieved. It seems, however, that the difference is only a moderate one. In Table 39 the average difference between payments made in the course of the first week and those in the second and third weeks is about seven percent. This casts some doubt on the possibility of explaining these facts by prestige considerations.

Another weakness in this explanation is that prestige would not affect all the items on the survey program but only those which might be used to impress. An inspection of data in Table 39 shows that it may be difficult to define all the items listed as prestige items. In this respect it would be very useful to see how the characteristics of nonprestige items behave. Further research and more experience will show whether the biases involved here represent a general statistical problem or are typical of certain surveys, such as expenditure surveys, food consumption surveys, family budget surveys, etc.

Another possible explanation of the downward tendency is *fatigue* on the part of respondents and their diminishing willingness to carry out the survey instructions. Thus in the course of the second week the respondents do not co-operate so carefully with the survey and tend to forget items in their record-keeping or reporting.

This explanation also encounters some difficulties. If some such change in the respondents' attitude takes place it should result in larger differences than those reported. Moreover, if diminishing interest in the survey is responsible for the change from the first week to the second, the same process should continue further and lead to even more obvious differences in subsequent weeks. In this connection the data in Table 41 show the nonresponse rate in successive interviews. It is clear that nonresponse does not increase in successive interviews.

TABLE 41. - THE RATE OF MISSING DATA IN FOUR SUCCESSIVE INTERVIEWS ACCORDING TO THE SOURCE OF INFORMATION [1]

Source of information	Nonresponse rate (percent)			
	Interview No. 1	Interview No. 2	Interview No. 3	Interview No. 4
Head	8.9	8.6	7.9	6.7
Wife	7.3	8.6	7.9	6.3
Head and wife	7.9	7.8	7.1	8.1
Unspecified	6.5	6.4	6.3	5.9

[1] This table is taken from Neter, J. and Waksberg, J.: Measurement of non-sampling errors in a survey of home-owners' expenditures for alterations and repairs, *Proceedings of the Social Statistics Section, American Statistical Association*, 1961, pp. 201-210.

The next possible explanation is what is called *end effect*. This issue will be dealt with in the next chapter.

9. THE RESPONDENT (CONTINUED)

9.1 End effect

When people are requested to report on food consumed or expenditures made in the short period of a couple of days in the immediate past they will normally find the task very difficult. This is so at least with regard to some items. Although the period may be precisely defined, the respondents may not be able to remember clearly whether certain events occurred within that period or not. They will therefore transfer out of the reference period, and fail to report, certain events that should be included, and vice versa.

Several processes are involved in the reporting of data for a period of reference. The first is, of course, recollection. The main difficulty in recollection is the increasing indistinctness of past events with the passing of time. The next difficulty is the chronological arrangement of the events recalled according to their time dimension or some kind of time scale, which is necessary if time is to be used as the criterion for the separation of events. Then the cutoff or end points in the sequence of events are located and the part of them belonging to the reference period is established. The accuracy of the data reported clearly depends both upon recollection and on the position given to the end points.

The placing of the end points for events located around the limit of the reference period may prove difficult, since the time dimension of past events is not always accurately known. While the vagueness of our knowledge in this respect has little effect if events far from the end points are concerned, it produces the transference of some events into or out of the reference period when these events are located around the end points. This transference of events, because of its connection with end points, will be referred to under the name of *end effect*.

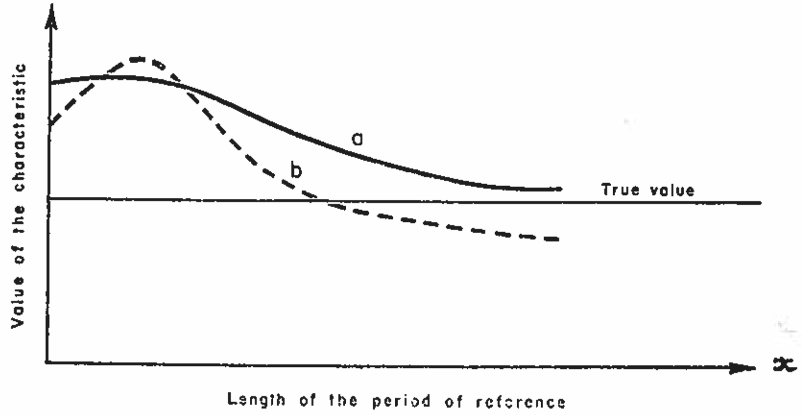
The time dimension of certain events is normally only vaguely known because we do not attach much importance to it. We take coffee several times a day and at different places. When doing so we do not consider the possibility of statisticians coming with a food consumption survey questionnaire and asking us to single out the coffees taken within a specified period of time.

End effect probably varies from item to item. Rare events are better in the memory and there is little likelihood of data being confused. Recurring events, such as food consumption and the purchase of standard products, do not imprint themselves strongly on the memory and therefore transferences may easily occur. Transferences are probably also affected by the relationships between the reference period and what might be called natural cycles of events. Those who receive their salary once a week would normally have a week as their natural spending cycle. In the field of food consumption surveys such cycles are a day, a week, a month or a year. One cannot study food consumption by collecting data for breakfasts only.

It is believed that the transference of events into and out of the reference period is particularly frequent with short reference periods. The relationships between the survey value of the characteristics reported, the corresponding true values and the length of the reference period may be represented as in Figure 8.

The horizontal line in Figure 8 stands for the true value (irrespective of the length of the reference period and the value of the characteristics). The length of the reference period is plotted on x axis. Curves a and b represent two different characteristics. With the characteristic a, short reference periods lead to positive biases (overreporting) which

FIGURE 8. - RELATIONSHIPS BETWEEN SURVEY VALUES, TRUE VALUES AND THE LENGTH OF THE REFERENCE PERIOD



gradually vanish with the lengthening of the reference periods, so that estimates based on long reference periods could be considered unbiased. Such a curve would correspond to data in Tables 37 and 38. Curve b indicates larger positive biases at first in the short reference periods and ends with negative biases in longer reference periods.

Such curves explain data in Tables 37 and 38 and may also provide some explanation for data in Table 39. The first week could be considered as subject to a positive bias resulting from the use of a short reference period. At the end of this period people may include some expenditure actually lying outside it so that a positive bias is produced. After they have been informed that another interview will take place in a week's time, people start paying more attention to their expenditure and a transference of the above type becomes impossible. The same is true of the subsequent week.

The variation in the amount and the effect of transference according to the length of the reference period may be caused by somewhat different approaches that respondents adopt in answering survey questions. There is some indication that with long reference periods and characteristics connected with frequently recurring events (which cannot be remembered separately) the respondents make an effort to establish some kind of average per unit of time and then multiply their average by the number of units of time in the period of reference. For example, if asked for his expenditure on tobacco in the last month, the respondent will try to establish an average expenditure per day and multiply it by 30. In other words, no attempt is made to recall each event separately.

With short reference periods the approach may be different. The respondent may consider each event separately as long as it falls within the reference period. In this case he builds up his response by adding up the values of the characteristics concerned for distinct events. In doing so it appears that respondents transfer into the period of reference more events than they eliminate or forget. As a result, the survey value of the characteristics involved is positively biased in the case of short reference periods.

An interesting parallel can be drawn here between the bias due to the transfer of outside events into the period of reference and the bias in estimating the rate of yield by crop-cutting. Sukhatme[1] has pointed

---

[1] Sukhatme, P.V.: The problem of plot size in large-scale yield surveys, *Journal of the American Statistical Association*, Vol. 42, 1947, pp. 297-310; the same author: *Sampling Theory of Surveys with Applications*, The Iowa State College Press, Ames,

out that field staff tend to include within a frame borderline plants which are really located outside. This results in an overestimation of the rate of yield which is considerable with small plots and vanishes with large plots. It appears that a similar mechanism is active in the case of interview surveys.

The explanation of data in Table 39 (page 204) by means of end effect is possible only on the assumption that the respondents know nothing of the survey before the first interview but are informed that they will be kept in the sample for further interviews. If the respondents are informed in advance of their selection in the sample, their attitude and behavior will be the same in all the weeks and the explanation breaks down. It should also be added that this explanation is possible for interview surveys only. If data are collected by means of record-keeping, the question of memory lapses and reference periods does not arise in the sense presented here. Since data in Table 39 are in fact collected by means of record-keeping the validity of the explanation given is doubtful.

Von Hofsten also reports that a decreasing tendency in estimates was detected in data collected in Sweden by means of accounting books.[2]

Considerable experience from future studies will be needed before a full explanation can be given of the data in Tables 37, 38 and 39 (pages 201, 202, 204). It is not yet known, for example, whether the variations in the estimates illustrated in Table 39 represent an exclusive feature of interview surveys or are possible with other methods as well. Until further information, the attempted explanations must remain mere hypotheses.

9.2 Open and closed reference periods

A reference period is called *open* if both its end points are located in the past so that end effect can be present at both extremities. For example, a person is interviewed on 1 March about the expenditure on various food items in the course of the week starting 11 February and ending 17 February. Both limits are in the past. On the other hand a reference

---

Iowa, and the Indian Society of Agricultural Statistics, New Delhi 1954; Panse, V.G.: *Estimation of Crop Yields*, FAO, Rome, 1954. See also Mahalanobis, P.C.: Recent experiments in statistical sampling in the Indian Statistical Institute, *Journal of the Royal Statistical Society*, Vol. 109, 1946, pp. 326-378.

[2] von Hofsten, E.: A budget survey in Sweden, *Family Living Studies*, ILO, Geneva, 1961, pp. 15-35.

period is called *closed* if its cutoff points are clearly distinguished in the memory and there can be no confusion as to where an event is located. For example, a person settles down in a new city or a new country. If the survey is concerned with the value of the furniture purchased from the moment of arrival in the new place of residence, the reference period will be precisely defined in the memory of the respondent and he will remember clearly which pieces of furniture were purchased during this period. Finally, a *half-open* or *half-closed* reference period is one where one extremity is open and the other closed. In fact, this is the type of reference period predominantly used in statistical surveys. Its position in time is normally such that its near end is located in the present moment. This class includes surveys where information is requested for "yesterday," "last week," "last month" or the day or the week "preceding the interview." In all these cases there is no end effect with regard to the nearer end point.

It will be clear that the terms "open" and "closed" refer to the memory and not to survey instructions or the calendar. With regard to the latter all reference periods are closed because both limits are precisely fixed. This is not so in the case of the memory.

It is obvious from these definitions that closed reference periods are superior from the accuracy point of view. End effect is eliminated and one only has to deal with recollection problems. In order to make use of this advantage one should aim at closing the reference periods. This, however, is not simple. In many surveys it may be possible to close one end only. This is in fact done in many surveys. Closing the far end may be possible with rare and important events that impress themselves on the memory. With recurring events in daily life it becomes practically impossible to find a point of time in the respondent's memory that can serve as an easy border line with regard to all past events.

This can only be achieved partially and more or less successfully by skillfully exploiting the circumstances and by applying various techniques. One measure is the use of the natural cycles of various characteristics, such as a day or week for food consumption, a week or month for income, etc. Since characteristics vary considerably in their cycles, reference periods of *varying* length may be needed in a single survey. Another measure may consist of informing the respondents of the start of the survey so that they can all locate the far end of the reference period. Another measure is to start the survey at a point of time which introduces a new natural period with regard to the characteristics involved. Such a point might be the new calendar year for business establishments, the beginning of the season, the beginning of the scholastic year, etc. Yet another measure may be to list the events concerning the characteristics involved in the period prior to the survey and keep this list for the next interview so that the respondent knows what events are not involved.

Such measures never close the reference period in an absolute sense. They only close it with varying degrees of success. It is believed, however, that the better it is closed, the more accurate are the data. For this reason the designer should spare no effort in finding the means of achieving as complete a closure as possible.

In order to illustrate the importance of closed reference periods some data will be presented from an experiment conducted by Neter and Waksberg.[3] In the experiment house owners were interviewed with the aim of estimating the number and the size of jobs undertaken for the alteration or repair of houses. The design provided for four different reference periods: (*i*) one month open, (*ii*) six months open, (*iii*) one month closed and (*iv*) three months closed. If the reference period was longer than a month, the respondent had to report his data for each month separately. The closing of the period was carried out by recording the jobs and the corresponding expenditure for a specified period. In the next interview the respondent was informed about the response made in the previous interview. This helped the respondent not to confuse the events in the month concerned with events in the more distant past. Some of the data obtained are presented in Tables 42 and 43. It can be seen from Table 42 that the open one-month reference period leads to very high estimates as compared with other types of periods of reference and particularly as compared with a one-month closed period. This result is consistent with data presented in earlier tables. If data in column 4 are considered more accurate than the rest because of the closed reference period and the short time lapse between the occurrence of the events and the moment of reporting, which reduces the incidence of memory errors, then data in column 3 are subject to a positive bias.

Table 43 shows that the estimates of the number of small jobs resulting from both the closed three-month and the open six-month reference periods are subject to negative biases compared with the estimates based on the one-month closed reference period. On the other hand, the estimates relating to large jobs and the open six-month reference period are subject to positive biases.

TABLE 42. - ESTIMATED NUMBER OF JOBS ACCORDING TO TYPE OF REFERENCE PERIOD AND SIZE OF JOB[1]

Size of job (in dollars)	Type of reference period			Percentages	
	Open, six months	Open, one month	Closed, one month	$\frac{(2)}{(4)} \times 100$	$\frac{(3)}{(4)} \times 100$
(1)	(2)	(3)	(4)	(5)	(6)
	...... *Millions of jobs*				
Total	213.9	300.6	215.1	99.4	139.7
1 - 9	89.4	145.1	112.2	79.7	129.3
10 - 19	40.9	47.2	35.5	115.2	133.0
20 - 49	40.7	52.7	34.6	117.6	152.3
50 - 99	17.6	25.6	13.5	130.4	189.6
100 - 499	19.5	25.3	15.7	124.2	161.1
500 and over	5.8	4.6	3.5	165.7	131.4

[1] Neter, J. and Waksberg, J.: Measurement of non-sampling errors in a survey of home-owners' expenditures for alterations and repairs, *Proceedings of the Social Statistics Section, American Statistical Association*, 1961.

Other features of this table are the sharp variations between the estimates referring to various months before the interview. In order to obtain such estimates the respondents were requested to specify the month when each job was carried out. In the processing it was therefore possible to classify jobs by size and month. It can be seen that the estimates referring to the month preceding the interview are positively biased while the others are negatively biased. Data referring to the third month before the interview and the total number of jobs make up only about 50 percent of the number of jobs estimated by means of the one-month closed reference period. This result is in line with the relationships established in the experiments of the Indian Statistical Institute and particularly with those shown in Figure 7 (page 194). It will also be seen that in this respect there are some differences between the small jobs and the large ones. Estimates

---

[3] Neter, J. and Waksberg, J.: Effect of length of recall period on reporting of expenditures in a survey of home-owners' expenditures for alterations and repairs, *Proceedings of the Social Statistics Section, American Statistical Association*, 1962 and "Measurement of nonsampling errors in a survey of home-owners' expenditures for alterations and repairs," *Proceedings of the Social Statistics Section, American Statistical Association*, 1961.

TABLE 43. - ESTIMATED NUMBER OF JOBS ACCORDING TO TYPE AND LENGTH OF REFERENCE PERIOD [1]

Reference period	Size of job			
	Total	Jobs of under $20	Jobs of $20 to $99	Jobs of $110 and over
	Millions of jobs			
Closed, one month	116.3	79.1	25.2	12.0
Closed, three months				
Average	84.1	49.4	22.5	12.1
1 month preceding interview	133.1	82.7	34.1	16.4
2 months preceding interview	70.7	37.7	20.9	12.0
3 months preceding interview	48.4	27.9	12.6	8.0
Open, six months				
Average	87.8	45.8	26.7	15.3
1 month preceding interview	122.5	69.4	35.0	18.0
2 months preceding interview	76.0	34.6	26.9	14.5
3 months preceding interview	65.0	33.3	18.2	13.5

[1] Neter, J. and Waksberg, J.: Op. cit.

of the number of small jobs of under $20 are about the same for both the three-month closed and the six-month open reference periods. It appears that the period of three months is too long to achieve the effects of closing for small jobs. With regard to small jobs these two reference periods are probably equally open. The terms "open" and "closed" probably have a very relative meaning that varies chiefly according to the characteristics, the frequency of events, and the way in which they are impressed on the memory.

It is obvious from the material presented that considerably more information is needed on the effect of closing the reference period and its variation, depending upon different factors. An important line of research will be the study of efficient techniques for closing the reference period in certain specified circumstances. Success in this direction might greatly contribute to improving the quality of survey data.

9.3 Location in time of the reference period

At what point in time should the chosen reference period be located? This is the problem that follows from the last section. A survey with a reference period of a week could be made by collecting data either for the week preceding the day of the interview or for the week falling between two specified dates. The former is called the *moving* reference period and the latter the *fixed* reference period. The moving reference period is half open while the fixed one is open at both ends. This is so because the field work nearly always extends over a longer period of time. For all those who are not interviewed immediately after the fixed reference period is over, both cutoff points will be located in the past so that data collected may be subject to two end effects.

If the field work is extended over a longer period of time, data collected in a survey with a moving reference period do not relate to the same period. This fact does not cause any bias in data if the characteristics involved vary randomly from one day to another within the whole period covered by the field work. However, as soon as there are larger variations within this interval, the application of the moving reference period will normally lead to biases under usual survey conditions. For example, with people who receive their salary once a month, the composition of expenditures may change as the month progresses. Also, the quantity and the quality of the food consumed may change. Now, if the expenditure survey is taken with a moving reference period of a week, it will normally happen that the bulk of the enumeration will be carried out at the beginning of the field work with a decreasing number of interviews per day as the end of the field work approaches. With this distribution of work, which is current in survey practice, data will be biased. The sign of the bias will depend upon where the bulk of the interviews is located. Of course, the bias may be either positive or negative.[4]

Such a bias is eliminated if an equal number of interviews are carried out each day. The day-to-day variations in the characteristics involved are then adequately reflected in the survey estimates. However, this measure is hardly practicable because of not-at-homes and refusals. Therefore, the moving reference period represents a real danger in the case of characteristics subject to considerable variation within the period of the field work.

Another difficulty connected with the moving reference period is the specification of the time to which the data collected refer. With a ref-

---

[4] To get an idea of the various possibilities resulting from the application of the moving reference period to characteristics that are subject to considerable changes, cf. Som, R.K., Das, N.C. and De, A.K.: *Report on Morbidity*, The National Sample Survey, No. 49, Government of India, The Cabinet Secretariat, 1961.

erence period of a week and without variations of the characteristics involved within the period used for the field work, weekly totals or averages might be considered to refer to the whole period of the field work rather than to any specific week. If the characteristics on the program vary, it normally becomes impossible to distinguish the period to which data relate. The only expedient in this situation is to inform users of the data of the progress of the enumeration so that they have some idea of the day or period of time to which the bulk of the data refer. Of course, this measure is only a partial solution. The difficulty of the interpretation of the data or its association with some other information remains unsolved. This is why comparisons of data referring to highly variable material should be made with considerable care.

We can summarize the situation in the following way: the advantage of a moving reference period lies, of course, in the fact that it represents half-closed periods of time so that end effect is made impossible at the nearer cutoff point. The advantage of fixed reference periods is that they raise no difficulty in the interpretation of data. Their drawback is additional end effect.

It will be seen that the problem of the location in time of the reference period requires special consideration in the preparation of the survey. Circumstances may be such that a fixed reference period is needed. This is so in the case of variable characteristics which normally appear on the program of food consumption surveys, expenditure surveys, etc. In order to use the advantages of such periods and avoid their disadvantages some special measures will be needed. Such a measure might be an increase in the number of enumerators so that the field work is carried out in a very short time after the reference period has ended. This would prevent end effect at the nearer limit. Another measure might consist in arranging that all the units selected in the sample be visited prior to the beginning of the reference period and adequately prepared for data collecting. In this visit the purpose of the survey would be explained and the respondents would be requested to pay special attention to the characteristics involved during the specified period. Such a visit would be helpful if it resulted in the closing of the reference period. Obviously, though, it has a disadvantage in that it might disturb the respondents' normal course of life. If aware of observation, they might change the composition of their expenditure or food consumption. The solution of one difficulty would then create another problem.

Another possibility in the use of fixed reference periods is to abandon interviewing and switch to the accounting method. With this method the respondents are requested to record all the events concerned within the fixed period of time. Again, the technique might cause a disturbance in the normal course of their lives. It may also be inapplicable for such reasons as illiteracy.

If the respondents cannot be informed in advance of their inclusion in the sample because of the danger that this might influence their pattern of life, interviewing after the reference period expires may be the only possible approach. If, in addition to this, the material is highly variable, the number of field workers small, and the need for a precisely defined reference period very great, none of the measures mentioned will be applicable. In such a situation sampling in time may be the approach needed. Sampling in time, of course, consists in sampling the households, persons or other relevant units so that estimates for the population as a whole can be obtained. In addition, time is sampled within the year so that estimates for the year as a whole can be made. Such estimates may be necessary in expenditure and food consumption surveys because of considerable qualitative and quantitative changes over time. In connection with food consumption surveys it has been pointed out by Zarkovich, Said and Khamis[5] that data needed for the policy measures in the field of nutrition can hardly be secured by restricting the survey to a week, irrespective of the week selected. Here as well as in the field of expenditure sampling in time will provide the widest range of information.

9.4 Conditioning

By conditioning in statistical surveys we understand those changes in the respondents' reaction to surveys in general and to various questions in the survey program in particular that result from previous exposure to requests for the same or similar material. In statistics conditioning can assume a large variety of aspects. Respondents may consider that their co-operation with a particular type of statistical survey might be harmful to their interests. The result might be either a high rate of refusals or deliberately poor quality response. In the course of increasing experience they will gradually understand that their attitude was wrong. As a result, the rate of refusals will decrease and the quality of the response will improve.

---

[5] Zarkovich, S.S., Said, E.E. and Khamis, S.H.: Statistics of food and nutrition, *Bulletin of the International Statistical Institute*, 1961, Vol. 39, Part 4, 1962, pp. 43-52.

This is one form of conditioning. There are many others. If requested to give information on his income for the past year the respondent may take the trouble to recall each individual item of income and build up the total. In successive surveys he will present the same response as in the first attempt. In other words, the result of the first attempt was decisive for the successive attempts. Another very well-known experience is that the information given by a member of the family is normally communicated to the other members as well. If the survey design provides for asking other members of the same family for the same information, the response will in many cases be a repetition of the answer given by the first respondent. In other words, the members of the family were conditioned by the first response.

In some cases conditioning may exercise a positive influence on the quality of survey data. However, it appears that in most cases it causes a certain amount of bias in data. From among the examples of conditioning that cause biases a case will be chosen that may help to explain relationships very similar to the consequences of end effects and memory lapses. Data for the illustration of the problem are shown in Table 44, which is taken from the experiment by Neter and Waksberg mentioned earlier. It can be seen from the table that data obtained in the second interview are, without exception, higher than the corresponding data collected in the third interview.

So far such differences have been interpreted as the result of memory lapses and end effect. In fact, memory lapses and end effect could supply the explanation of the pattern in Table 44 if the two sets of data had been collected in the same interview and referred to two different months, one preceding the other. In this case, however, the situation is different. Each set of data is collected in a separate interview referring to the "last month" and by means of the one-month closed reference period. Therefore, both sets of data are equally affected by memory lapses. Similarly, if end effect is not completely excluded it is equally present in both series of data. Therefore, the explanation of the data in Table 44 is to be found outside these two factors.

A possible explanation of this difference in data could be found in conditioning in the sense that households gradually lose interest in the survey, co-operate progressively less and gradually lose the ability to recall events with the initial accuracy. In this case the atmosphere would deteriorate from one interview to the next. As a result, a decreasing tendency in data might be obtained as in Table 44.

TABLE 44. - NUMBER OF JOBS AND EXPENDITURES INVOLVED, ESTIMATED FROM SECOND AND THIRD INTERVIEWS BY MEANS OF A ONE-MONTH CLOSED REFERENCE PERIOD [1]

Size of jobs (in dollars)	Number of jobs (in millions)		Expenditures (in $ millions)	
	Second interview	Third interview	Second interview	Third interview
Total	214.8	195.4	8 666	8 203
Under 20	149.8	136.6	930	820
20 - 49	34.0	30.2	1 012	878
50 - 99	13.0	11.7	883	796
100 and over	18.1	17.0	5 839	5 709

It is not easy to say whether conditioning is responsible for differences in Table 44. The design of the experiment was based on a number of subsamples and the interviewing in each subsample took a month. In other words, data used to establish the estimates in both the second and the third interview are equally distributed over the whole period March 1960 to March 1961. The intention was to eliminate the trend, if any, in the number of jobs and the expenditures involved. It appears, however, that the trend could be used as an explanation of the data in Table 44 if it had a downward direction. Estimates referring to the second interview are then higher than the estimates referring to the third interview because the former interview precedes the latter. However, data for the former interviews are not exhibited simultaneously and it is not possible to see how they compare with each other.

Irrespective of whether conditioning is the explanation of the data in Table 44, and if so, to what extent, the purpose of this section is to call attention to this problem. In the designing of surveys it has to be taken into account as it is very possible that some effects of conditioning are present in all the repetitive surveys. In some cases these effects may lead to catastrophic consequences. If monthly surveys are being conducted with the aim of estimating differences, it must be established that the differences in data are really due to changes in the characteristics involved and not to conditioning. If data in Table 44 are attributable to condition-

---

[1] Neter, J. and Waksberg, J.: Effect of length of recall period on reporting of expenditures in a survey of home-owners' expenditures for alterations and repairs, *Proceedings of the Social Statistics Section, American Statistical Association*, 1962.

ing and the statistician forgets this possibility and explains them as a change from one month to another, his conclusions will be very misleading.

9.5 Variance considerations

In deciding about the length of the reference period it is generally understood that the variation of the estimates involved, as it results from the use of various reference periods, also has to be considered. It is believed that these variations will be large with short reference periods and vice versa. If food consumption is considered on a daily basis it is reasonable to expect high variations of the per person consumption because some households will consume various foods which will not be consumed in others. On a weekly basis many of these variations will disappear because those households which did not consume specified items on a given day will do so on some other day. For this reason the estimated per person per day consumption based on data for the whole week will be subject to less variation. It is to be expected that the variation will be further reduced with monthly estimates. It can be concluded that longer reference periods are preferable to shorter ones because they make it possible to achieve a given precision by means of a smaller sample.

This point of view was taken into account by Reisz[6] in his experiments on the methodology of household surveys in Greece. In that part of the survey dealing with expenditures, all the items involved were classified into the following four groups:

(*i*) daily purchases such as food, household supplies, transport, recreation, etc.;
(*ii*) clothing, household utensils, etc.;
(*iii*) items purchased infrequently, such as furniture, major household equipment, etc.;
(*iv*) items for which regular payments were made at fixed but differing intervals, such as rent, electricity, water, etc.

The information available on the frequency of events in various caregories made it necessary to adopt reference periods of varying length.[7] Thus, for the group (*ii*) a period of four weeks was used and for (*iii*) and (*iv*) the whole year. For the first group the alternatives were one week or three weeks. The latter reference period was used in the experimental survey but data were broken down by weeks. The coefficient of variation of the estimated expenditure on food was found to be 5 percent in the first week, 4.7 percent in the second and the third weeks combined, and in the three weeks taken together 4.6 percent. On this basis a one-week reference period was chosen as it was found that an extended reference period would not contribute sufficiently to the improvement of precision to justify the extra effort.

A satisfactory approach to this problem requires that the variance be replaced by the mean square error approach. Different reference periods are associated with different magnitudes of biases which cannot be forgotten because their squares represent the minimum value of the corresponding mean square error. If the bias associated with a particular reference period is large, there may be no justification for considering this particular period at all, as a satisfactory accuracy cannot be achieved in the estimates. Table 38 (page 202) provides an illustration of this. If daily estimates on a shop basis are taken as true values, the one-day reference period in interview surveys leads to a bias of practically 100 percent of the value of total purchases. A one-week reference period does not greatly change the magnitude of the bias. In this situation neither of these reference periods can be used irrespective of the variations of estimates associated with them.

In view of the importance of this problem it will be useful to exhibit some data collected in an experiment carried out in connection with the FAO Training Center on the Methodology of Food Consumption Surveys, held in 1962 in Dubrovnik, Yugoslavia. The aim of the experiment was the comparison of data on food consumption as collected by different methods, i.e., interviewing, accounting and weighing of food. In the subsample assigned to interviewing, a day, a week and a month were used as the reference periods. Accounting was carried out over a period of four weeks, with data available for each week separately. Weighing was carried out for a week, with data available for each day separately. The total sample was broken down into five independent subsamples, three of which were assigned to interviewing and one each to accounting and weighing. Data selected for illustration are presented in Tables 45, 46, 47 and 48.

The first part of Table 45 shows the average per person per day consump-

---

[6] Reisz, A.B.: A budget survey in the urban areas of Greece, *Family Living Studies*, International Labour Office, Geneva, 1961, pp. 67-86.

[7] Cf. also von Hofsten, E.: A budget survey in Sweden, *Family Living Studies*, International Labour Office, Geneva, 1961, p. 15-35.

TABLE 45. - AVERAGE PER PERSON PER DAY CONSUMPTION OF SPECIFIED FOODS BASED ON DATA COLLECTED WITH DIFFERENT METHODS AND DIFFERENT REFERENCE PERIODS

Item Time, method	Bread	Potatoes	Cherries	Beef	Butter
			Grams		
Daily interview 10 June 1962	278	134	140	21	14
” 11 ” ”	260	103	87	5	12
” 12 ” ”	268	95	82	11	13
” 13 ” ”	254	69	104	5	13
” 14 ” ”	251	65	140	19	10
” 15 ” ”	245	83	100	4	11
” 16 ” ”	236	78	101	17	14
Average 10 June - 16 June	256	90	108	12	13
Weekly interview 27.5 - 2.6	265	54	82	12	10
” 3.6 - 9.6	257	78	103	15	9
” 10.6 - 16.6	249	72	88	14	9
” 17.6 - 23.6	254	79	80	16	10
Average 27 May - 23 June	255	71	88	14	9
Monthly interview 27.5 - 23.6	277	72	78	9	7

SOURCE: FAO Training Center on the Methodology of Food Consumption Surveys, Dubrovnik, Yugoslavia, 1962.

tion of some specified food items as collected in seven daily interviews carried out in the period between 10 June and 16 June 1962. Daily averages are followed by an average based on the whole week. The next part shows the corresponding averages resulting from four weekly interview surveys taken for "the last week." These are followed by the average based on combined data for the four weeks. The last line shows the per day per person consumption based on data collected in the interview referring to the last four weeks, i.e., the period 27 May to 23 June.

Table 46 contains the averages based on daily weighing of foods put aside for the preparation of meals. The lower part shows the results of the accounting presented on a weekly basis as it was not possible to separate some quantities of food purchased into daily consumption.

Tables 47 and 48 show the corresponding variances. It will be seen from Table 47 that daily variances are followed by an average daily variance and the variance resulting from the combined data for the week as a whole.

The first interesting result of this experiment is that averages resulting from the application of different methods do not differ greatly. As for the variances, Table 47 shows that they substantially decrease for all the items if daily data for the week as a whole are combined. This tendency of the variance to decrease with larger reference periods is also obvious

TABLE 46. - AVERAGE PER PERSON PER DAY CONSUMPTION OF SPECIFIED FOODS BASED ON DATA COLLECTED WITH DIFFERENT METHODS AND DIFFERENT REFERENCE PERIODS

Item Time, method	Bread	Potatoes	Cherries	Beef	Butter
			Grams		
Daily weighing 10 June 1962	279	125	122	12	13
” 11 ” ”	228	87	63	15	10
” 12 ” ”	280	109	86	13	5
” 13 ” ”	263	69	81	12	6
” 14 ” ”	270	68	78	14	6
” 15 ” ”	263	48	87	2	7
” 16 ” ”	268	48	43	8	7
Average 10 June - 16 June	264	79	80	11	8
Weekly accounting 27.5 - 2.6	273	50	39	13	6
” 3.6 - 9.6	269	86	82	11	8
” 10.6 - 16.6	274	87	92	9	7
” 17.6 - 23.6	262	86	102	12	7
Average 27 May - 23 June	269	77	79	11	7

SOURCE: FAO Training Center on the Methodology of Food Consumption Surveys, Dubrovnik, Yugoslavia, 1962.

from data referring to weekly interviews. If weekly data are put together variances drop substantially. However, the variances resulting from data obtained in the interview for the whole month did not decrease with regard to weekly averages.

The variances in Table 48 also show a tendency to decrease with longer reference periods.

The question that arises is: which technique of collecting data and what period of reference should be used? The costs of an interview survey are hardly affected by the length of the reference period. If it is also assumed that the estimated level of consumption is the same with the three types of interview surveys, the right choice would probably be either weekly or monthly surveys. The choice is governed by variance considerations. Daily interviewing is discarded because of the additional consideration of day-to-day variations in the food consumed. From Tables 45 and 46 it appears that all the days in the week are not the same from the point of view of the food consumption. A larger reference period, such as a week, makes it possible to avoid the resulting difficulties.

The estimated food consumption does not change in this experiment if interviewing is replaced by weighing or accounting. For this reason daily weighing is discarded on the same grounds as daily interviewing.

TABLE 47. - VARIANCES OF THE AVERAGE PER PERSON PER DAY CONSUMPTION OF SPECIFIED FOODS BASED ON DATA COLLECTED WITH DIFFERENT METHODS AND DIFFERENT REFERENCE PERIODS

Items / Time, method	Bread	Potatoes	Cherries	Beef	Butter
	Grams				
Daily interview 10 June 1962	21 308	20 169	28 070	2 880	335
” 11 ” ”	27 684	29 281	22 266	[1]	198
” 12 ” ”	20 442	26 299	25 657	1 311	233
” 13 ” ”	20 984	15 037	30 259	[1]	223
” 14 ” ”	23 976	10 413	30 610	1 734	184
” 15 ” ”	21 982	12 945	25 792	[1]	194
” 16 ” ”	25 558	12 376	17 532	677	358
Average daily variance	23 133	18 074	25 741	1 651	246
Summary 10 June - 16 June	16 092	4 133	10 456	265	146
Weekly interview 27.5 - 2.6	13 605	2 531	6 515	555	103
” 3.6 - 9.6	13 111	4 493	7 888	669	234
” 10.6 - 16.6	12 683	4 041	4 492	407	77
” 17.6 - 23.6	13 449	2 836	6 099	604	105
Average weekly variance	13 212	3 475	6 249	559	130
Summary 27 May - 23 June	10 809	1 634	3 352	341	71
Monthly interview 27.5 - 23.6	13 182	4 287	4 845	1 062	90

SOURCE: FAO Training Center on the Methodology of Food Consumption Surveys, Dubrovnik, Yugoslavia, 1962.

[1] Figure meaningless because of too few data.

Of course, this conclusion is valid under the circumstances of this experiment, carried out in city households. With illiterate people who have no knowledge of units of weight the situation may be different. In this case the alternative might be either interviewing or accounting, with a week as the reference period. If no further considerations are involved, the former possibility may be chosen, as it entails less disturbance for the respondents.

If different methods or different reference periods were associated with considerable biases, as in Table 38 (page 202), the right approach in solving the problem would be the use of the equation (1.15). The value of $\sigma^2_{\bar{x}}$ for $n = 1$ would be variances such as those in Tables 47 and 48. In estimating the bias, the most accurate method should be selected. This is not necessarily the one associated with weighing. After this choice has been made the bias is defined numerically and ζ can be computed for each alternative possibility. This is a comparison of alternatives. With ζ fixed in advance, the equation (1.15) may be solved for n and another comparison is obtained. If the budget is fixed at C and the cost of collecting data by means of the g-th alternative equals c_g per unit, the size of the sample associated with this alternative will be n_g. This makes it possible to estimate the quality of data that could be achieved by various alternatives within the budget C. Other economy considerations are, of course, also relevant.

TABLE 48. - VARIANCES OF THE AVERAGE PER PERSON PER DAY CONSUMPTION OF SPECIFIED FOODS BASED ON DATA COLLECTED BY MEANS OF DIFFERENT METHODS AND DIFFERENT REFERENCE PERIODS

Item / Time, method	Bread	Potatoes	Cherries	Beef	Butter
	Grams				
Daily weighing 10 June 1962	30 411	25 233	29 157	1 439	714
” 11 ” ”	18 974	10 113	10 449	1 693	351
” 12 ” ”	34 848	14 772	23 303	1 471	113
” 13 ” ”	46 165	16 534	21 491	[1]	249
” 14 ” ”	28 979	15 680	14 795	2 266	86
” 15 ” ”	31 974	8 448	19 843	[1]	216
” 16 ” ”	27 086	13 124	8 876	[1]	242
Average daily variance	31 205	14 872	18 273	1 717	282
Summary 10 June - 16 June	18 318	2 390	8 138	455	84
Weekly accounting 27.5 - 2.6	21 756	1 875	7 434	568	69
” 3.6 - 9.6	28 485	5 046	19 273	373	146
” 10.6 - 16.6	35 272	2 854	13 742	478	164
” 17.6 - 23.6	20 054	3 706	21 685	418	97
Average weekly variance	26 392	3 370	15 534	459	119
Summary 27 May - 23 June	22 455	1 407	8 238	308	66

SOURCE: FAO Training Center on the Methodology of Food Consumption Surveys, Dubrovnik, Yugoslavia, 1962.

[1] Figure meaningless because of too few data.

9.6 Integration of errors

It has already been pointed out that it is of greater importance to produce accurate individual data than to remove biases from inaccurate estimates. In order to achieve this aim it is useful to know the sources of various errors and the relative importance of individual sources. This knowledge facilitates the combating of errors in data.

Sufficient knowledge of the individual sources of errors cannot easily be acquired because of the *integration* of errors. In other words, the contributions of individual sources of errors are pooled together and the result is an integrated error, of which the component parts are in most cases indistinguishable. For example, prestige errors represent a deliberate change of data while end effect is an involuntary error. In the final response the contributions from the two sources are pooled together. For practical action against errors it is very important to know which of these two sources is involved, and to what extent. Completely different steps are used in coping with prestige errors on the one hand and with end effect on the other. With regard to age data such as those in Table 30 (page 187), Myers writes: "In the reporting of the age there arise five major forms of errors and bias; first, underreporting of the number of children at ages 0 and 1; second, a tendency to give exact age 21, probably because of its legal significance; third, distinct overstatement among those at very advanced ages; fourth, a general tendency (termed "heaping") for ages to be given as ending in certain digits; and fifth the reporting of some individuals as being of unknown age."[8] Table 30 shows that errors are present in data but the contributions of various sources are not visible.

In this section an example will be given of an attempt to penetrate behind the integrated final result with the aim of detecting the individual sources of errors and, if possible, measuring their relative contributions. It is hoped that further research of this kind will be undertaken, as it can yield very important information.

The study in question was conducted by Das Gupta and Mitra[9] with the aim of obtaining more detailed information on errors in age data.[10] The sources of age data in India are very different from the corresponding sources in western countries. In other words, only relatively few of the answers to the question on age constitute a definite statement of the year of birth. In many more cases ages are reported with reference to certain events or represent guesses based upon the appearance of the person concerned. Clearly, such estimates are made according to very broad age groups. In a special study of the quality of demographic data conducted in 1954 in West Bengal, the results presented in Table 49 were obtained. This shows that more accurate information on ages depends upon the general development of the country and the more frequent use of age data and birth certificates.

---

[8] Myers, R.J.: Errors and bias in the reporting of ages in census data, *Transactions of the Actuarial Society of America*, Vol. 41, 1940, pp. 395-415.

[9] Das Gupta, A. and Mitra, S.N.: *A Technical Note on Age Grouping*, The National Sample Survey, No. 12, The Cabinet Secretariat, Government of India, New Delhi, 1958.

[10] Another source of detailed information on errors in age data is found in Tekse, K.: *Investigations of reliability of age reporting in the 1960 Population Census* (in Hungarian), Central Statistical Office, Budapest, 1964.

TABLE 49. - PERCENTAGE DISTRIBUTION OF INDIVIDUAL AGE DATA BY TYPE OF EVIDENCE AVAILABLE[1]

Age groups	Area	Type of evidence				
		Hearsay, guess or eye estimate	Related with definite or approximate age or events	Definite statement of year of birth	Birth certificate or other document	Total
0 - 6	City	27.1	51.0	21.9	—	100.0
	Other urban	24.6	42.0	31.9	1.5	100.0
	Rural	18.9	34.0	44.2	2.9	100.0
7 - 16	City	41.7	41.7	16.6	—	100.0
	Other urban	33.9	49.3	14.8	2.0	100.0
	Rural	36.7	45.5	15.1	2.7	100.0
17 and over	City	65.8	19.8	13.6	0.8	100.0
	Other urban	53.6	39.0	5.5	1.9	100.0
	Rural	46.2	44.4	6.6	2.8	100.0

[1] Das Gupta, A. and Mitra, S.N.: Op. cit.

The presence of errors in data collected under these circumstances is obvious from the usual heaping up. The following are assumed to be the sources of errors present:

(*a*) digit preference
(*b*) estimation error
(*c*) age bias.

The first of these three classes reflects the tendency on the part of the respondent to use some digits more frequently than others. The second comprises the errors which appear as a result of difficulties in assessing ages, while the third covers all sorts of more or less deliberate changes of age, such as cases classified as prestige errors, changes connected with the fear of some action, etc. These three broad sources of errors have a varying degree of importance according to conditions. In a country where age data are known in most cases, inaccurate information is primarily due to age biases. If they are known only approximately, digit preferences may become more important. If ages are not known, biases arising from difficulties in guessing acquire primary importance.

To get an idea of the magnitude of biases that could be attributed to an isolated effect of digit preference, an experiment was made. Two hundred and twenty employees of the Indian Statistical Institute were asked to fill in three missing middle digits in a group of seven-digit numbers. In such an experiment only digit preference could be expected to operate as a source of bias.

The frequency of the missing digits supplied is given in Table 50. This table shows the preference for middle digits rather than the usual heaping up at 0 and 5. This was called the digit preference of the first order.

Another interesting result found in this study was the preference for consecutively rising digits, which is shown in Table 51. Relatively high frequencies are found for pairs of digits, such as 12, 23, 34, etc., and low frequencies for repeated digits, such as 00, 11, 22, etc. The preference for consecutively rising digits is called the digit preference of the second order. The analysis of the population census data of Bengal led the authors to the conclusion that both the first- and the second-order digit preference had influenced age data.

TABLE 50. - RELATIVE FREQUENCIES OF THE MISSING DIGITS SUPPLIED [1]

Frequency \ Digit	0	1	2	3	4	5	6	7	8	9	Total
Absolute	86	59	133	117	115	134	127	140	94	95	1 100
Percentage	7.8	5.4	12.1	10.6	10.4	12.2	11.6	12.7	8.6	8.6	100.0

[1] Das Gupta, A. and Mitra, S.N.: Op. cit.

TABLE 51. - FREQUENCY DISTRIBUTION OF DIGITS SUPPLIED BY GUESSWORK IN THE FIRST TWO DIGIT BLANKS [1]

First missing digit	Second missing digit										Total
	0	1	2	3	4	5	6	7	8	9	
0	9	7	9	4	6	5	1	2	3	3	49
1	14	9	29	13	10	10	7	7	6	3	108
2	12	8	6	22	16	13	18	12	9	7	123
3	6	4	14	6	33	20	15	18	11	14	141
4	11	4	19	24	5	38	32	13	7	8	161
5	9	5	12	14	13	9	24	17	14	6	123
6	5	5	16	14	7	10	7	40	11	7	122
7	5	5	7	5	7	11	13	7	21	13	94
8	5	3	11	6	9	10	4	16	3	21	88
9	10	9	10	9	9	8	6	8	9	13	91
Total	86	59	133	117	115	134	127	140	94	95	1 100

[1] Das Gupta, A. and Mitra, S.N.: Op. cit.

Another experiment was made to study various characteristics of age guessing or assessment and the biases arising therefrom. For this purpose the same 220 employees were presented with a picture containing five lines of different shape and length. They were asked to estimate the total length of these five lines in terms of a standard line. The relationship between the standard line and the total length of the five lines was such that the second decimal point of the estimate could only be considered pure guesswork. Deviations in the frequency of individual digits from the expected frequencies were considered to show the pattern of errors involved in guessing or eye estimation. In order to check how this pattern compared with the corresponding pattern of age errors, age data for persons of 40 years of age and over from the fourth round of the National Sample Survey were used and tabulated according to the last digit. The two frequency distributions are presented in Table 52. These strikingly similar distributions authorize the conclusion that after a certain age

TABLE 52. - PERCENTAGE FREQUENCY DISTRIBUTION OF THE DIGITS IN THE SECOND DECIMAL PLACE OF THE LENGTH ESTIMATE OF FIVE LINES, AND THE LAST DIGIT OF THE AGE REPORTED FOR PERSONS AGED 40 AND OVER AS OBTAINED IN RURAL AREAS IN THE FOURTH ROUND OF THE NATIONAL SAMPLE SURVEY IN INDIA [1]

Item	Digit										Total
	0	1	2	3	4	5	6	7	8	9	
Second decimal place	37.3	2.3	6.5	4.0	3.4	31.0	5.2	3.7	4.3	2.3	100.0
Last digit of age 40 and over	31.3	4.3	8.8	4.7	5.2	22.7	6.2	5.1	7.8	3.9	100.0

[1] Das Gupta, A. and Mitra, S.N.: Op. cit.

the unit digit of age is supplied by guesswork. The resulting age errors are guessing errors.

According to this study, an essential feature of eye estimation consists in the tendency to overestimate. A highly significant difference was found between the true total length of the five lines and the average estimated length. The same was found to be true for age data. For a number of persons both the estimated and the definite age were known and tabulated according to the sign of the difference between the two. Positive differences predominated. In other words, age data can be expected to contain some upward bias.

This tendency to overestimate was further studied in 1955 when the households that had been visited in the fourth round of the 1952 National Sample Survey were revisited and age data were again collected. Differences in data for the same persons revealed interesting features. First, an average overstatement of 0.18 year was found with reference to age data available from an earlier survey. The overestimation varied, however, according to the age groups. In the youngest group of persons between the ages of 3 and 9 years an underestimate was found; a small overestimate appeared in the range 10-19 and reached its peak for the ages of 40 years and over. This result falls into line with the general experience that advanced ages are often subject to overstatements.

It was also found that the concentration of age returns on some typical digits, such as 0 or 5, varies according to the type of data. For the three different types, namely (*i*) ages guessed, (*ii*) ages reported as approximate and (*iii*) definite statements, the degree of concentration at digit 0 is presented in Table 53. A decreasing degree of rounding off is obvious in these data as one goes from guesses toward definite statements.

TABLE 53. - PERCENTAGE OF AGES REPORTED AS ENDING AT 0 IN THE AGE GROUP OF 23-62 YEARS [1]

Type of data	City	Rural
1. Guess	23.0	26.9
2. Approximate	16.6	21.5
3. Definite	9.8	17.8
Total	18.9	23.1

[1] Das Gupta, A. and Mitra, S.N.: Op. cit.

The next digit of concentration is 5, the high relative frequency of which is a result of both rounding off and digit preference. In the case of estimates or guesses falling between these two digits, 2 and 8 are also used in rounding off.

As regards the age bias, it has some characteristic features and has to be distinguished from other biases. Its possible location and nature can be detected by studying the cultural characteristics of a social group. Some study of age biases in census material can be made by comparing data returned in the census with either the birth and death register or with birth certificates. Neither of these procedures was possible in India and therefore a study was made of the ratios of the number of age answers ending with a particular digit to the total number of returns in a decennial age range. Such ratios, as computed for data collected in the fourth round of the National Sample Survey, are reproduced in Table 54.

These ratios shed some light on several quality aspects of age data. If all the individual answers were accurate the ratios in the table would tend to be equal. Rounding off and digit preferences introduce deviations of varying degree from the expected value. Rounding off increases as the age progresses. In the range 70-79 years more than 50 percent of the returns end in 0. Table 54 may also help to detect deliberate changes of data in the neighborhood of ages having special social importance, such as ages which permit voting, entering the civil service, receiving retirement benefits, etc. If these ages are different from those showing the usual heaping up, the presence of biases will be visible in the magnitude of the ratios. If they end in digits with a considerable concentration,

TABLE 54. - RATIO OF THE NUMBER OF AGE RETURNS ENDING WITH A PARTICULAR DIGIT TO THE TOTAL NUMBER OF RETURNS IN SUCCESSIVE DECENNIAL AGE RANGES[1]

End digit	Decennial age range							
	0 - 9	10 - 19	20 - 29	30 - 39	40 - 49	50 - 59	60 - 69	70 - 79
0	10.6	15.2	18.7	26.6	31.4	38.6	49.2	51.0
1	10.1	8.1	6.1	4.6	4.5	5.0	4.6	3.8
2	10.9	14.4	13.5	11.8	9.5	9.2	8.9	7.7
3	10.8	8.7	6.9	5.8	5.0	4.5	3.8	4.4
4	9.7	9.4	9.1	5.8	5.6	4.6	4.7	3.0
5	10.9	10.4	18.4	20.4	23.1	19.6	18.7	19.7
6	10.3	11.3	8.4	8.3	6.2	5.0	2.7	3.3
7	8.9	5.8	5.4	4.5	4.9	3.8	2.1	2.7
8	10.2	11.7	9.8	8.4	6.8	6.5	3.6	3.0
9	7.6	5.0	3.7	3.8	3.0	3.2	1.7	1.4
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

---

[1] Das Gupta, A. and Mitra, S.N.: Op. cit.

such as 0 and 5, the presence of the bias would be reflected in excessive magnitude of the ratio. Such a case may be present in this table at the age of 60.

Several questions could be raised in connection with certain results of this study. For example, it may be questioned whether the experiment on digit preference is really indicative of the process that takes place in age statistics. The choice of the three missing digits in seven-digit numbers may be guided by what digits are shown. The digits presented in this particular experiment give the impression of having been picked at random, i.e., without any visible preference for any special digit. When requested to fill in the missing digits, the respondent may feel that he is supposed to be consistent with this pattern. In censuses he feels no such limitation, as can be seen, for example, from the data presented earlier on the frequent use of two or three zeros in reporting crop production.

Another question concerns the existence of upward biases in age guesses. Although biases can easily appear in guessing processes, it is not certain that an experiment with lines or stones is applicable in this situation. If lines of a different shape and length had been used, the result might have been a downward bias. It is a well-known fact that eye estimates of the rate of yield show such a bias in the case of exceptionally high yields.[11] Table 55 provides a useful illustration. Data in this table refer to the results of the quality check of the 1953 Census of Population in Yugoslavia. Age data found to be inaccurate in the check are classified here according to whether they represent over- or understatements. The hypothesis of overstatements does not seem to find such obvious support here.

Questions could also be raised in connection with the distinction between digit preference and rounding off. However, the primary reason for pointing out the above results from the study by Das Gupta and Mitra is to emphasize the importance of such an analysis. This study is typical of the recent interest in all the factors affecting the response while the respondents build up their answers to survey questions. It is expected that this type of study will help in the future to shed more light on the "technology of response production" and make it possible to improve survey techniques.

---

[11] Yates, F.: Some examples of biased sampling, *Annals of Eugenics*, Vol. 6, 1935, pp. 202-213.

TABLE 55. - THE RELATIONSHIPS BETWEEN OVER- AND UNDERSTATEMENTS OF AGE (1953 Census of Population, Yugoslavia)[1]

Stratum	Rural stratum			Urban stratum		
	Over-statement	Under-statement	Total	Over-statement	Under-statement	Total
Males	186	176	362	63	59	122
Females	235	216	451	78	94	172
Total	421	392	813	141	153	294

---

[1] Zarkovich, S.S.: *Population Census Errors* (in Serbo-Croatian), Federal Statistical Office, Belgrade, 1954.

9.7 Checking the effects of errors due to respondents

Let us assume that a self-enumeration census of N units was taken. In such a survey each of the N respondents gets the questionnaire and fills it in. The intermediary role of the enumerators is excluded. The resulting response errors are then attributed to respondents. In this situation the response error of the i-th respondent is independent of the response error of the k-th respondent in the answer to the same question.

In order to check the quality of the data resulting from such a situation we select a sample of n units and for each of them get a check response for the characteristics on the check program. The design of the sample may be anything that satisfies efficiency criteria. For each unit in the sample we also establish the response error $d_i = z_i - x_i$ and afterward estimate the response bias and its significance.

Let us assume that the check is based on a simple random sample of n units and the purpose of the check is to examine the effect of response errors on the frequencies of the age distribution. For each person selected from the census lists it is possible to find documents and get accurate information on age. All the units are then classified into age groups and within each group we establish $d_i = z_i - x_i$ where

$$z_i = \begin{cases} 1 \text{ if the } i\text{-th unit belongs to the age group concerned accord-} \\ \quad \text{ing to the response given in the census} \\ 0 \text{ otherwise} \end{cases}$$

with an analogous meaning of x_i. Here again d_i can assume three values only, i.e., 0, 1, and -1. From the units belonging to the groups concerned we estimate the bias in the fequency of that group which is equal to $D' = \frac{N}{n}\Sigma\, d_i$. In the estimation of the variance of the bias we use the equation (6.6) with d_i instead of D_i. The quantity Σd_i^2 will then be the gross error (cf. Section 6.1) in the sample aggregate, while Σd_i will be the bias of the sample aggregate. The estimated bias D' may be considered significant if Σd_i exceeds $2\sqrt{\Sigma d_i^2}$.

An illustration of this procedure is given in Table 56, which refers to the quality check of the response as conducted for the 1950 United States Census of Population. Table 56 is taken from a paper by Ross Eckler and Hurwitz.[12] The first column of this table contains the age classification. Columns 2 and 3 show the census and the check aggregates. Columns 4 and 5 are clear. Column 6 contains class biases, and column 7 the gross number of errors. The last column serves to test the significance of the bias. If the figure in column 6 is more than twice the corresponding figure in column 8, the bias may be considered significant.

If the purpose of the survey were the check of the class totals, such as area, expenditure, etc., the same estimates and procedures could be used. In this case z_i and x_i would be areas collected in the census and the check respectively. However, this is not an efficient method for reasons presented in Section 6.2.

A more realistic design would consist of a two-stage sample. In the first stage, m EDs are selected out of the population of M EDs. In the selected EDs a fixed percentage of respondents will be selected and their response checked. The numerical data on the effect of response errors on class frequencies are then presented as in Table D, which is entitled "Effects of response errors on the frequencies of size classifications of holdings." The effect of response errors on class totals is presented in Table E, called here "Effects of response errors on the total area of holdings in various size classes."

Here again total area is just one of many characteristics that may be of interest.

Data included in both Tables D and E refer to the sample aggregate.

---

[12] Eckler, A.R. and Hurwitz, W.N.: Response variance and biases in censuses and surveys, *Bulletin of the International Statistical Institute*, Vol. 36, Part 2, 1958, pp. 12-35.

TABLE 56. - CHECKING THE QUALITY OF AGE DATA (1950 UNITED STATES CENSUS OF POPULATION) [1]

Age class	Number of persons in the class according to		Number of persons inaccurately included in the class ($d_i = 1$)	Number of persons inaccurately excluded from the class ($d_i = -1$)	Bias (4) — (5) (Σd_i)	Gross number of errors (4) + (5) (Σd_i^2)	$\sqrt{(7)}$
	Census (Σz_i)	Check (Σx_i)					
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Total	26 980	26 980	1 611	1 611	0	3 222	56.8
0 - 4	2 897	2 945	40	88	—48	128	11.3
5 - 9	2 377	2 356	71	50	21	121	11.0
10 - 14	2 023	2 020	65	62	3	127	11.3
15 - 19	1 907	1 876	86	55	31	141	11.9
20 - 24	1 997	1 992	98	93	5	191	13.8
25 - 29	2 160	2 161	123	124	— 1	247	15.7
30 - 34	2 062	2 052	150	140	10	290	17.0
35 - 39	2 010	1 995	147	132	15	279	16.7
40 - 44	1 835	1 810	163	138	25	301	17.3
45 - 49	1 632	1 633	154	155	— 1	309	17.6
50 - 54	1 495	1 490	161	156	5	317	17.8
55 - 59	1 309	1 351	108	150	—42	258	16.1
60 - 64	1 083	1 094	82	93	—11	175	13.2
65 - 69	899	904	76	81	— 5	157	12.5
70 - 74	606	605	57	56	1	113	10.6
75 and over	688	696	30	38	— 8	68	8.2

---

[1] Data in this table are taken from Eckler, A.R. and Hurwitz, W.N.: Response variance and biases in censuses and surveys, *Bulletin of the International Statistical Institute*, Vol. 36, Part 2, 1958, pp. 12-35.

TABLE D. - EFFECT OF RESPONSE ERRORS ON THE FREQUENCIES OF SIZE CLASSIFICATIONS OF HOLDINGS

		Size classification of holdings (in ha)				Total
		0.5 - 5	5 - 10	10 - 20	20 - 50 [1]	
1. Census	Number					
2. Agreement	Number					
	Percent					
3. Erroneously included	Number					
	Percent					
4. Omissions	Number					
	Percent					
5. Bias	Number					
	Percent					
6. Squared EDs' biases	Number					
7. Significance test						

[1] Columns in this sequence can be extended as far as is desired.

The stub of Table D is the same as the one in Table A (page 113). Therefore there is no need for additional explanations. The only remark needed concerns the total column. The bias in this column may come from erroneously included units. It may happen that some units are enumerated because of the inaccurate response on some relevant characteristics given during the listing stage. For example, in the response check it may be detected that the declared size of a holding is inaccurate and that the unit does not need to be enumerated. It should also be added that omissions refer to particular size classes. Because of inaccurate response a unit may be omitted from one class and included in another. Clearly, the omissions do not contribute to the "Total."

Table E is more complex. Data in this table are derived from the same classification of units as that in Table D. Line 1 includes the total area of agricultural holdings for all the holdings falling in a given size class according to the census. Line 2 refers to totals for units where the information on total area as collected in both the census and the check is identical. In considering whether the responses are identical, small discrepancies can be disregarded. The units which fall in the same class according to both surveys and do not have identical response will be distributed over lines 3 and 4 according to the value of d_{ij}. It is obvious that the total number of units used to provide data in lines 2, 3, and 4

TABLE E. - EFFECT OF RESPONSE ERRORS ON THE TOTAL AREA OF HOLDINGS IN VARIOUS SIZE CLASSES

		Size classification of holdings (in ha)				All holdings
		0.5 - 5	5 - 10	10 - 20	20 - 50[1]	
1. Census	Amount					
2. Agreement	Amount					
	Percent					
Discrepancies within classes:						
3. Positive differences	Amount					
	Percent					
4. Negative differences	Amount					
	Percent					
5. Bias	Amount					
	Percent					
6. Erroneously included	Amount					
	Percent					
7. Omissions	Amount					
	Percent					
8. Total bias	Amount					
	Percent					
9. Squared EDs' biases	Amount					
10. Significance test						

[1] Columns in this sequence can be extended as far as is desired.

is equal to the number in line 2 of Table D. The bias in line 5 is the sum of figures in lines 3 and 4. In line 6 we put the total for all the erroneously included units in a particular size class. Line 7 contains similar totals for units omitted from a class. The total bias in line 8 is equal to the sum between the bias in line 5 and the difference between lines 6 and 7. Squared biases of EDs in line 9 and the significance test in line 10 are obtained as in Table D. It will be noticed, however, that here the variable *d* is not a 0,1 variate.

Percentages in various lines of the table refer to the amount shown under "Census."

9.8 Measures for improving the quality of the response

In a self-enumeration survey the quality of the response depends basically upon the respondents.

A great deal can be done to improve the quality of the response by the use of the measures listed in Section 6.10 in connection with the quality of listing. Publicity, legal obligations, supervision, etc., are all important in this respect.

In surveys where the quality of the response depends largely upon the respondent, one of the most important aspects of the preparations for the survey is to develop the respondents' *interest* in the survey as much as possible. If the respondent is interested in the survey he will co-operate and wish to provide as accurate responses as possible. An interested respondent will read the instructions, inquire about facts he may not know and be ready to make an effort to build up the responses which are not readily available.

How can an interest in the survey be developed? This is the problem to be considered in each survey and particularly in large-scale surveys and censuses. Several measures will normally be needed to achieve this aim. Some respondents will appreciate the idea that by helping the survey and collaborating fully they are acting in the public interest. Others are drawn by the possibility that they may benefit from a successful survey. Many people like to help others and will be ready to co-operate if convinced that the survey under preparation is an opportunity to do so. In this respect some respondents will also be interested in the sponsor of the survey. Experience from some countries shows that government agencies are not always good sponsors; universities or other research bodies may exercise more appeal as impartial promoters of the public interest.

To avoid irritating the respondent is another important principle of survey technique. The respondents, however great their interest in the survey, may refuse to co-operate if too many difficulties are involved. For example, if the grouping of questions is illogical, the layout of the material in the questionnaire confusing, the instructions difficult to understand, or incomplete, etc., the respondent, who at first was willing to co-operate, will get irritated, and may withdraw his co-operation. If he is obliged by law to co-operate, the quality of the response will be immaterial to him. Clearly, the survey should not discourage the respondent when he has taken a favorable attitude. The preparations of the survey should secure *undisturbed co-operation.*

In order to achieve undisturbed co-operation with respondents an important step in the preparation of large-scale surveys may consist of a special *study* of various relevant characteristics of the respondents, such as their knowledge of the subject, the way they express this knowledge, the vocabulary and units they use, etc. These details are then used for building up the survey design. Such studies are conducted as pretesting surveys or as a part of some other experimental program. Many relevant facts can also be collected in the course of study programs associated with current statistical activities.

In fixing the details of the survey design an effort should be made to *avoid assumptions* regarding the respondents' knowledge and abilities. Some surveys have failed either partly or entirely because the respondents were assumed to know too much, i.e., everything that the designer had it in mind to achieve. The first question in the questionnaire used in a survey was: "General data" (about the respondent). Obviously, the designer was assuming that all the respondents in this mail survey would know what he intended by this question. The respondents' interpretation of the question might cover: name, sex, age, education, etc. The respondent may easily discard the questionnaire if he does not know what response is wanted.

Another example of possible confusion is connected with the question "Age?", which is often found in questionnaires. The respondents may put their date of birth, the year, the number of years completed at the last birthday, or round off their age under the impression that accuracy is not needed, etc. If accuracy is in fact needed, difficulties will arise. From the processing point of view a difficulty will arise in any case because the variety of responses will call for systematic editing.

"Type of activity," "economic activity," "occupation," "profession," etc., are examples of other questions which cause confusion.

Question wording is another problem that arises. As this matter has already been discussed it will not be treated again here.

With regard to items where the response is not readily available and has to be worked out, it will be useful, prior to taking a decision on the choice and wording of questions, to investigate what approach the respondents will use in arriving at their answer. With this knowledge, the questions can be chosen and worded in a way suited to the respondents' thought processes.

An illustration of this point will be useful. In agricultural statistics questions are often asked on the total area of holdings and areas under various crops. This approach is adequate if holders have this information available for their own economic reasons. This will usually be so in the case of large commercial farms. In the case of small holdings and particularly in subsistence agriculture, holders will probably not be able

to supply this information at once and will have to work it out. As a result the building up of the response may consist of the following stages: (*i*) listing individual pieces of land, (*ii*) associating each of them with the kind of crop grown, (*iii*) associating each field with its area, (*iv*) establishing total area, (*v*) establishing totals for each crop separately. If the respondents' approach is of this type, a more adequate way of obtaining information about areas would be to ask for each separate field, its area and the crop grown. Totals might then be compiled either by the enumerator or in the course of processing. The respondent may find such questioning more natural and easier to answer; it also reduces the danger of his being irritated if faced with something that he does not know, and excludes the possibility of mistakes in his computation.

This raises the problem of the mechanism that governs the respondent's mind in the answering of survey questions. Recent developments have shown that the quality of the response could be greatly improved if more knowledge were available on this mechanism. Sounder techniques for collecting data in certain fields, such as expenditure, income, food consumption, etc., will hardly be possible without further progress in this direction.

Another group of measures for the improvement of the quality of the response is connected with the *general circumstances* of the survey. The impact of fear, doubts and suspicion as to the purposes of surveys is known to everybody. In this respect an analysis of the circumstances will show what measures, if any, can be useful in achieving immediate effects and what program of long-term action might be considered as a basis for general progress in statistics.

In connection with general circumstances the statistician will also have to face problems such as: at what time (month, season) the survey should be made, bearing in mind the variation in the movement of the population, the available leisure time, the climatic difficulties that the field personnel may have to face, the variations in the characteristics of the program, etc.; whether the survey should go on all day or be confined to specified hours (evenings), etc.

10. THE ENUMERATOR

10.1 Reasons for using enumerators [1]

The use of enumerators in collecting data in surveys provides a way of coping with errors that originate in the sources already discussed. The use of enumerators brings to the survey more responses than it is generally possible to obtain with mail questionnaires. If the respondent is not interested in the survey, it is much easier for him to discard the questionnaire which he receives by post than to refuse to co-operate with an enumerator who comes to the house for that purpose. If the respondent is not able to understand the concepts and definitions used, the enumerator is there to put the questions in a language comprehensive to the respondent. If the respondent is not familiar with certain units of measurement or does not know how to express certain quantities in terms of these units, the enumerator's assistance may be necessary to formulate the response. The enumerator will also be useful in eliminating misinterpretation of questions, reminding respondents of certain items that are easily forgotten, etc. Thus, the enumerators make it possible to achieve survey aims more efficiently than if the respondents are left to their own devices and their own interpretation of the survey.

The enumerators can only fulfill these functions if their work is *standardized.* In other words, the enumerators' interpretation of the meaning of survey questions and of the concepts used has to be the same. Otherwise the data obtained will not fit the provisions of the adopted system of work and the results will reflect the enumerators' personal views.

A detailed discussion of purposes that census enumerators are supposed to serve is available in a paper by Bogue.[2]

---

[1] The concept "enumerator" is used here in a broad sense to mean persons collecting data in a survey irrespective of whether they are census enumerators, public opinion survey interviewers or field staff observing the kind of crop grown in selected holdings.

[2] Bogue, D. J.: The pros and cons of "self-enumeration", *Demography*, Vol. 2, 1965, pp. 600-626.

Standardization of the enumerators' procedures and actions is achieved by means of *training.* Ideally, training is intended to remove differences in the enumerators' interpretation of all the problems connected with the field work and bring their action into conformity with the adopted system of work. Obviously, for a number of reasons this ideal can only be more or less approximated in practice. In some cases the preparations for a survey will not be detailed enough and the enumerators will be left to interpret various details for themselves. In other cases the training is not successfully planned and the enumerators are not given a good chance to acquire the knowledge needed. In yet other cases there is no adequate training check to eliminate those enumerators who have not been satisfactorily trained. For these and similar reasons the enumerators become an additional source of error in statistical data. It has been found that they often add something to the data they collect. Changes introduced into data by enumerators will be called *enumerator effect.* The changes introduced reflect the enumerator's personality, his training and education or, as Mahalanobis says, his *personal equation.* Clearly, the enumerator effect will vary in magnitude from one enumerator to another and from one item to another.

10.2 Some illustrations of enumerator effect

The possibilities of training as a means of standardizing the enumerators' work are obviously limited even in the case of a most systematically prepared program of training. The enumerator remains a human being with all the consequences following therefrom. If he is put in front of a field to estimate its area or the rate of yield of the crop grown, there will be some risk of error however good his training. If the enumerator is involved in a survey dealing with important social issues he will hardly be able to prevent his own opinions appearing in the survey results. In this connection the classic example presented by Rice[3] can usefully be quoted. In 1914 a study of certain characteristics of homeless applicants at the New York Municipal Lodging House was made. Among other things, the study aimed at revealing the causes that led applicants to their difficult

---

[3] Rice, S.A.: Contagious bias in the interview, *The American Journal of Sociology*, Vol. 34, 1929, pp. 420-423.

TABLE 57. - DIFFERENCE IN THE RESULTS OBTAINED BY TWO ENUMERATORS IN A STUDY OF APPLICANTS AT THE NEW YORK MUNICIPAL LODGING HOUSE[1]

Enumerator	Percentage of persons ascribing their position to:	
	Liquor	Industrial factors
A	34	43
B	11	60

[1] Rice, S.A.: Op. cit.

social and economic positions. For this purpose a sample of applicants was selected and the enumerators were instructed to obtain the respondents' own description of basic causes. Table 57 shows the results obtained by two enumerators, A and B. The sample was said to be selected by a procedure which makes it difficult to explain the difference in percentages as a result of the selection bias. In fact, the difference was found to be a result of the enumerators' own opinions about the issue involved. Enumerator A was an ardent prohibitionist and believed that liquor was the major cause of the applicants' position. Enumerator B was a socialist who liked to find an explanation of such a situation in the characteristics of society itself. Although they had to obtain the applicants' own explanation, it appears that the enumerators influenced the respondents through their presentation of the survey aims and obtained results in agreement with their own interpretation of the situation.

Other examples of a similar type are available. Hyman[4] reports on two surveys conducted with the aim of studying the differences in the results obtained by enumerators who had different opinions on the issue involved in the survey. In the first of these two surveys a sample of Christian respondents was selected with the aim of obtaining their opinion on the influence of Jews in the business world. In the part of the sample where Christian enumerators were used, 50 percent of the respondents declared that Jews had too much influence in the business world. In the part of the sample interviewed by Jewish enumerators, the same answer was obtained from 22 percent of the respondents only.

---

[4] Hyman, H.H.: Problems in the Collection of Opinion-Research Data, *American Journal of Sociology*, Vol. 55, 1950, pp. 362-370.

In another experimental survey the respondents were asked whether they agreed with the statement: "Prison is too good for sex criminals; they should be publicly whipped or worse." The result of the study was as follows: 61 percent of women respondents agreed with the statement if interviewed by male enumerators; the corresponding percentage obtained by female enumerators was 49.

These examples are in line with common experience. Different persons often see, hear and interpret the same facts in different ways. The author came across a paragraph in a book describing a disagreement between two foreigners living in London on the pronunciation of the English word "such." Each of them thought he was right so they decided to stop the first person in the street and ask him or her to pronounce that word for them. After they had done so the disagreement was not eliminated because each of them was convinced that his point was confirmed. They decided to consult another person and the result was the same.

10.3 Definition of enumerator effect

After the illustration of enumerator effect in the last section let us now define this effect more precisely. Let us assume for this purpose a population consisting of N units. M enumerators are available for the enumeration of this population.[5] Both N and M are large and infinite for sampling purposes. It is also assumed that each enumerator has enumerated each unit of the population. The information obtained by the i-th enumerator for the j-th unit of the population is designated by z_{ij}. The resulting responses could be usefully visualized as representing a population consisting of M primary units and N second-stage units in each primary unit. The mean and the variance of this population are:

$$\left.\begin{aligned} \bar{Z} &= \underset{i}{E}\,\bar{Z}_i \\ &= \underset{i}{E}\,\underset{j}{E}\,z_{ij} \end{aligned}\right\} \qquad (10.1)$$

---

[5] In this chapter we are using $i = 1, 2, \ldots, M$ to designate the enumerators. The symbol i was used earlier for primary units. The individual units of the population are designated by j, which was used earlier for secondary units. It is hoped that this change will not confuse the reader.

and

$$\left.\begin{aligned}\sigma_z^2 &= \underset{i}{E}\,\underset{j}{E}\,(z_{ij}-\bar{Z})^2\\ &= \underset{i}{E}\,\underset{j}{E}\,(z_{ij}-\bar{Z}_i)^2+E\,(\bar{Z}_i-\bar{Z})^2\\ &= \sigma_{zW}^2+\sigma_{zB}^2\end{aligned}\right\} \qquad (10.2)$$

Obviously, σ_{zB}^2 in (10.2) expresses the variation between primary units or between the averages as obtained by different enumerators.

In order to understand what type of errors committed by enumerators leads to enumerator effect four different cases may now be distinguished. In Case I it is assumed that all the enumerators receive accurate data from all the units. In this situation the following relationships hold: $z_{ij} = x_j$, $d_{ij} = 0$, $\bar{Z}_i = \bar{Z} = \bar{X}$. For the same reason the component σ_{zB}^2 vanishes and the population variance reduces to $\sigma_z^2 = \sigma_x^2 = \sigma_{zW}^2 = \underset{j}{E}\,(x_j-\bar{X})^2$.

In Case II the response errors are assumed to appear in such a way that they cancel out within an enumerator's assignment, i.e., $\underset{j}{E}d_{ij} = 0$. Here again the resulting averages are unbiased as we have $\bar{Z}_i = \bar{Z} = \bar{X}$. Also, $\sigma_z^2 = \sigma_{zW}^2$. The presence of errors is reflected in σ_{zW}^2, which, on the basis of (10.2), could be defined as

$$\sigma_{xW}^2 = \sigma_x^2 + \sigma_d^2 + 2\sigma_{xd}$$

with

$$\sigma_d^2 = \underset{i}{E}\,\underset{j}{E}\,d_{ij}^2$$

and

$$\sigma_{xd} = \underset{i}{E}\,\underset{j}{E}\,d_{ij}\,(x_j-\bar{X})$$

Here again $\sigma_{zB}^2 = 0$ while the magnitude of σ_{zW}^2 will be more or less different from σ_x^2 depending upon the characteristics of the errors involved. Although the errors that arise under the assumptions of this case might be the result of the enumerators' work, they do not constitute a specific feature of surveys using enumerators for the collecting of data. Self-enumeration surveys containing errors due entirely to the respondents also fall into this category.

In Case III we assume that all the enumerators collect data yielding equally biased averages, i.e., $\underset{j}{E}d_{ij} = \underset{j}{E}d_{i'j} \neq 0$ and $\bar{Z}_i = \bar{Z} \neq \bar{X}$. In this case we also have $\sigma_{zB}^2 = 0$ and the increased magnitude of σ_{zW}^2 is due to the presence of errors. Neither is this case a specific feature of surveys using enumerators, as the same effect might result from a self-enumeration survey.

Case IV refers to circumstances leading to enumerator effect. This is achieved with $\sigma_{zB}^2 \neq 0$, or if the average of the characteristic under study as obtained by at least one enumerator is different from the population average (or the average taken over all the enumerators). This situation arises when the response errors in data collected by an enumerator tend to go in a given direction. Enumerator effect has nothing to do with the bias in the quantity $\bar{Z}$. Enumerator effect is equally possible with $\bar{Z} = \bar{X}$ and $\bar{Z} \neq \bar{X}$. The essential element of this effect is the existence of a significant variation between averages belonging to different enumerators. This is why enumerator effect is called by some authors the *enumerators' variability* or the *interviewers' variability*.

Let us now assume that N is large enough for $1/N$ to be disregarded, while $N/(N-1)$ is approximately equal to unity. Under this assumption we define the coefficient of intraclass correlation as

$$\rho \doteq \frac{\sigma_{zB}^2}{\sigma_z^2}$$

If enumerator effect is present in the data collected, σ_{zB}^2 is large relative to σ_z^2. This means that the enumerator effect is characterized by a relatively large and positive value of the coefficient of intraclass correlation between individual errors within an enumerator's assignment. As has already been said, such a value of ρ is obtained if errors within an enumerator's assignment tend to go in the same direction. We can write

$$\begin{aligned}\sigma_{zB}^2 &= \underset{i}{E}\,(\bar{Z}_i-\bar{Z})^2\\ &= \underset{i}{E}\,(\bar{D}_i-\bar{D})^2 \qquad (10.3)\\ &= \sigma_{dB}^2\end{aligned}$$

If σ_{dB}^2 is put instead of σ_{zB}^2 in the above definition of the intraclass correlation, the association of enumerator effect with the tendency to produce errors of the same type is clear.

10.4 Reasons for studying enumerator effect

From the preceding section it follows that $\sigma_{zB}^2 > 0$ indicates that the work of the enumerators employed was not successfully standardized, that they introduced their "personal equation" into data and that their training was not sufficient to remove idiosyncrasies from their work. Knowledge of this effect shows whether the "human agency," as Mahalanobis called it, works satisfactorily or needs some improvement. This is the first reason for studying enumerator effect.

The next reason concerns the estimation of errors in sample surveys. In order to show what is meant by this let us assume that a sample of m enumerators is selected at random from a population of M enumerators to enumerate a sample of n units divided into m equal subsamples each of size $\bar{n}$ so that $n = \bar{n}m$. Following the traditional approach which disregards the enumerator effect we would pool together data collected by all the enumerators and would compute $\bar{z}$. Afterward, we would put

$$\sigma_{\bar{z}}^2 = \frac{\sigma_z^2}{m\bar{n}}$$

Because $\bar{Z}_i = \bar{\bar{Z}}$ we have

$$\sigma_{\bar{z}}^2 = \frac{\sigma_{zW}^2}{m\bar{n}}$$

This is the correct way of assessing sampling variation of the estimated mean under the assumption that $\sigma_{zB}^2 = 0$. If enumerator effect is present in data the right procedure is to apply the theory of two-stage sampling and put

$$\sigma_{\bar{\bar{z}}}^2 = \frac{\sigma_{zB}^2}{m} + \frac{\sigma_{zW}^2}{m\bar{n}} \tag{10.4}$$

It can thus be seen that the presence of a significant enumerator effect increases the sampling errors. Neglect of the enumerator effect may thus sometimes lead to a serious misinterpretation of the actual amount of variation in certain sample estimates resulting from data collected by the enumerators.

Following the same argument it could also be said that all the results of a complete enumeration census using enumerators could be considered to represent sample estimates. This follows from the fact that census

enumerators, however numerous they may be, may be regarded as a sample from an infinitely large pool of enumerators. On this assumption the second term in (10.4) will vanish as there is no sampling of the second-stage units, while the first term will reduce to $\dfrac{\sigma_{zB}^2}{M}$ where M is the number of enumerators used. With large M the contribution of this term will be small. One could also say that with large M one does not need to consider the enumerator effect irrespective of the magnitude of σ_{zB}^2.

In a self-enumeration census σ_{zB}^2 would vanish. This is the advantage of surveys using self-enumeration rather than enumerators. However, it would be dangerous to draw the conclusion that self-enumeration is superior from the quality point of view. A systematic discussion of the relative merits of these two approaches is provided by Bogue.[6]

Knowledge of enumerator effect also makes it possible to achieve further rationalization in statistical surveys. This follows clearly from what was said above. If the magnitude of the enumerator effect in data shows the degree of standardization of work reached by the enumerators, the knowledge of this effect in a concrete case might give some guidance in deciding about the amount of training needed, the characteristics of the survey program where improvement in the enumerators' performance would be useful, etc. If the results in complete enumeration censuses are also associated with sampling errors, as follows from the above consideration, it might be useful to reconsider in concrete cases the relative merits of complete censuses on the one hand and sample surveys on the other. If census results, because of the use of inadequately trained enumerators, are associated with a considerable enumerator effect that can be reduced in a comparable sample survey taken by a small number of highly trained enumerators, sample surveys may prove more efficient than complete enumeration censuses.

10.5 Measuring enumerator effect

The presence of enumerator effect can be tested by means of the analysis of variance. Mahalanobis was the first to apply the analysis of variance for this purpose.[7]

---

[6] Bogue, D.J.: Op. cit.

[7] Mahalanobis, P.C.: Recent experiments in statistical sampling in the Indian Statistical Institute, *Journal of the Royal Statistical Society*, Vol. 109, 1946, pp. 326-378.

The following is a simple example of the use of the analysis of variance for this purpose. Let us assume that m enumerators are selected at random from an infinite pool of enumerators. Similarly, from an infinite population m independent subsamples each of size $\bar{n}$ are selected at random. The total size of the sample is $n = m\bar{n}$. Each of the m enumerators is then assigned at random to enumerate the $\bar{n}$ units in one of the m subsamples.

From data collected the analysis of variance in Table 58 is established. It is assumed here that all the enumerators have an equal assignment $\bar{n}$. However, if the number of units collected by each enumerator proves different, equalization might be achieved by rejecting at random data for some units from subsamples having more units than it was decided to use for the analysis. On the other hand, if the variations in the size of samples are not large there is no need for equalization. In that case $\bar{n}$ is taken to represent the average size of the assignment.

After the analysis of variance table has been prepared it is at once possible to judge whether enumerator effect is present. Namely, if A and B are significantly different the effect does exist. If A is found to be larger than B we test the significance of the difference by computing the variance ratio

$$F = \frac{A}{B}$$

and comparing the value of F with the corresponding values of the F table for the selected level of significance and given degrees of freedom. If there is no significant enumerator effect all the mean squares in Table 58 are estimates of the population per unit variance and the sampling error of $\bar{z}$ is found by dividing the total sum of squares by $m\bar{n}\ (m\bar{n} - 1)$.

If the enumerator effect is found to be significant it must be incorporated in the sampling error of $\bar{z}$ according to (10.3).

An estimate of σ^2_{zB} from Table 58 is obtained as

$$s^2_{zB} = A - \frac{B}{\bar{n}} \tag{10.5}$$

TABLE 58. - ANALYSIS OF VARIANCE FOR THE PURPOSE OF ESTIMATING THE ENUMERATOR EFFECT IN SURVEY DATA

Source of variation	Degrees of freedom	Sum of squares	Mean square
Between enumerators	$m - 1$	$\sum_i (\bar{z}_i - \bar{z})^2$	A
Within enumerators between respondents	$m(\bar{n} - 1)$	$\sum_i \sum_j (z_{ij} - \bar{z}_i)^2$	B
Total	$m\bar{n} - 1$	$\sum_i \sum_j (z_{ij} - \bar{z})^2$	

To estimate σ^2_{zW} we use $s^2_{zW} = B$. On this basis the estimate of $\sigma^2_{\bar{z}}$ in (10.4) becomes

$$s^2_{\bar{z}} = \frac{s^2_{zB}}{m} + \frac{s^2_{zW}}{m\bar{n}} \tag{10.6}$$

The equation (10.6) is an estimate of the sampling error that takes the enumerator effect into account. As estimated from this equation the sampling error of $\bar{z}$ is often larger than the corresponding estimate of the variance based on the theory of simple random sampling. The latter is normally an underestimate of the sampling variation in $\bar{z}$ in the case of surveys using enumerators. It is therefore recommended that in sample surveys employing enumerators, the variance of the estimates be determined by the procedure in (10.6), adapted, of course, to the characteristics of the design.

For many purposes it is useful to compute the percentage of the enumerator's contribution to the total variance, i.e., the quantity

$$100\ \frac{s^2_{zB}}{ms^2_{\bar{z}}} \tag{10.7}$$

If computed for different items on the survey program and in different surveys as well this quantity represents a useful basis of comparison.

An example of the application of this procedure can be found in a paper by Stock and Hochstim.[8]

---

[8] Stock, J.S. and Hochstim, J.R.: A method of measuring interviewer variability, *The Public Opinion Quarterly*, Vol. 15, 1951, pp. 322-334.

10.6 Some empirical studies of enumerator effect

In studying empirically the magnitude of the enumerator effect such a simple use of the analysis of variance will not be possible. It will be necessary to adapt the experiment to the characteristics of the design. An example of a more complex analysis of this kind is the Nagpur Labor Inquiry conducted in India in 1942-43.[9] For this study the families to be surveyed were grouped into five zones. Within each zone four subsamples of equal size were established. Each subsample was assigned to a different enumerator. After having finished the work in one zone the same enumerators were transferred to the remaining four zones. The analysis of variance of the results collected is presented in Table 59. The lower part of the table contains the variance ratios. These ratios are computed by dividing the values of the variances in the upper part of the table by the corresponding within-subsample variance.

TABLE 59. - ANALYSIS OF VARIANCE OF THE RESULTS COLLECTED IN THE NAGPUR FAMILY BUDGET INQUIRY, 1943 [1]

Sources of variation	D.f	Total income	Monthly expenditure		
			Total	Food	Cereals
	(a) Values of variances				
Between zones	4	4 439.58	3 707.91	708.41	206.83
Between enumerators	3	85.43	597.08	77.09	3.70
Zones × enumerators	12	382.54	397.28	177.75	49.80
Between subsamples	19	1 189.74	1 127.07	237.58	75.61
Within subsamples	977	401.57	384.71	84.73	24.99
Total	996	424.67	398.87	88.33	25.95
Mean values (in rupees)	—	36.09	34.96	20.09	11.41
Standard deviations	—	18.60	16.12	8.53	4.96
Coefficients of variation	—	51.5	46.1	42.5	43.5
	(b) Ratios of variances				
Between zones	4	[2]11.06	[2]9.64	[2]8.36	[2]8.28
Between enumerators	3	0.21	1.55	0.91	0.15
Zones × enumerators	12	0.95	1.03	[3]2.10	[3]2.00
Between subsamples	19	[2]2.96	[2]2.93	[2]2.80	[2]3.02
Within subsamples	977	—	—	—	—
Total	996	—	—	—	—

[1] This table is reproduced from Mahalanobis. P.C.: Op. cit.
[2] Significant at 1 percent level.
[3] Significant at 5 percent level.

In this study zones were made purposely as different as possible. As a consequence, all the between zones variance ratios were found to be significant. The same was true of ratios concerning "between subsamples" variances. On the other hand, no "between enumerators" ratio was found to be significant. This shows that there was no significant enumerator effect on data.

In other studies significant "between enumerators" or interaction ratios were obtained. In some cases the analysis of significant ratios was carried a stage further by examining the corresponding data by zones and by enumerators. In the Jagaddal Labor Inquiry, conducted in 1941-42, such an analysis showed that significant variance ratio originated from abnormally high values collected by one single enumerator in one single zone. Such data can be discarded.

For rational survey designing an important problem is to find the type of questions which normally lead to high enumerator effect. The paper by Stock and Hochstim gives a number of interesting examples in this connection. The first refers to a study conducted in the Baltimore area with the aim of measuring the enumerator effect in a survey to assess the percentage of "dilapidated" houses. To standardize the enumerators' work a film was shown in addition to verbal training. In spite of this, it was found that the enumerator effect amounted to 12 percent of the total variance.

This experience is in agreement with similar experiences from other surveys. Whenever the enumerators' judgment is allowed to play a considerable part in the response, their effect on data becomes substantial. In this connection Table 60 may be instructive. It contains data on the percentage of the enumerator effect in the total variance as obtained in different surveys for various questions classified into three groups: factual questions, opinion and information questions and judgment questions. It will be seen from the table that the enumerator effect is smallest in the case of factual questions and largest in judgment questions.

The percentages in Table 60 may appear rather low. This is because the enumerator effect is not presented here by means of (10.7). Instead, the percentages in this table represent the quantity $100\,\sigma^2_{zB}/\sigma^2_z$. In fact, the enumerator effect found here would be very considerable if presented

---

[9] Mahalanobis, P.C.: Op. cit.

TABLE 60. - ENUMERATOR EFFECT AS THE PERCENTAGE OF THE TOTAL VARIANCE [1]

Type of questions	Percentage contribution to the total variance
Factual:	
Is respondent paid on hourly rate?	.06
Union to which respondent belongs	1.87
Does he own a car?	.62
Opinion and information:	
Is city X a good place to work?	3.90
Is electric company important to city?	2.00
Is store A owned by local people?	.00
Does local store benefit shoppers more?	5.44
Is good variety main reason for buying at store A?	1.43
Are trucks big enough?	0.27
Does respondent know of law limiting size of trucks?	.58
Does he favor less government control of business?	.00
Does he expect less prosperity next year?	1.42
Do farmers get right amount for their products?	.48
Judgment:	
Lower socio-economic status (survey A)	7.55
Lower socio-economic status (survey B)	7.17
Dilapidation	11.30

[1] Stock, J.S. and Hochstim, J.R.: Op. cit.

as a part of the variance of a sample estimate in a concrete design, as in (10.7). In the latter case σ^2_{zB} is divided by m, the number of the enumerators used, while σ^2_{zW} is divided by the total number of ultimate units, i.e., $m\bar{n}$. If a census is carried out in an area with five enumerators ($m = 5$), an assignment of 100 units per enumerator ($\bar{n} = 100$), $\sigma^2_{zB} = 1$ and $\sigma^2_{zW} = 100$, the variance of the estimated mean would be

$$\sigma^2_{\bar{z}} = \frac{1}{5} + \frac{100}{5 \times 100}$$

$$= 0.2 + 0.2$$

If expressed by means of (10.7) the enumerator effect would amount to 50 percent.

In Table 61 the analysis of variance of data in Table 60 is made according to the types of questions. The total variance is broken down into two components, between types of questions and within types of questions. The variance ratio observed amounts to $F = 16.19$ while the value of F at one percent significance level amounts to 6.70. In other words, questions are significantly different.

Additional examples and illustrations of enumerator effect in a large variety of characteristics will be found in the literature at the end of this book. A review of basic results obtained in many studies was presented by Kish.[10] The same paper also gives an account of the work done in this field in the Survey Research Center of the University of Michigan.

At this stage the importance of enumerator effect should again be stressed. Highly personal enumerators' contributions may lose any practical importance in a survey where, for a given area, many enumerators are employed. The smaller the number of enumerators in the area, the more important is the problem of enumerator effect. This fact is of primary importance in the preparation of the tabulation and publication program. Before it is decided to tabulate and publish data by small administrative units it is advisable to study the errors of sample estimates or census results in the light of the considerations presented. If the enumerator effect is too great at the level of the average size of the smallest administrative units planned in the publication program, the original plan may have to be revised in the sense that units covered by a larger number of enumerators are used instead.

TABLE 61. - ANALYSIS OF VARIANCE OF DATA IN TABLE 60 [1]

Source of variation	Degrees of freedom	Sum of squares	Mean square
Between types of questions	2	121.81	60.90
Within types of questions	13	48.78	3.75
Total	15	170.59	

[1] Stock, J.S. and Hochstim, J.R.: Op. cit.

10.7 General theory

It was assumed in previous sections that only one response for each unit is available. In some surveys, however, the enumerators may be requested to obtain more than one response for the units assigned. For

---

[10] Kish, L.: Studies of interviewer variance for attitudinal variables, *Journal of the American Statistical Association*, Vol. 57, 1962, pp. 92-115.

example, they may be requested to perform several measurements on each of the units assigned. In this case a more general theory will be needed. Such a theory was developed independently by Hansen, Hurwitz, Marks and Mauldin[11] on the one hand, and by Sukhatme and Seth[12] on the other. The theory assumes a random sample of m enumerators selected out of a large pool of M potential enumerators. A random sample of n units is selected from the large population of N units. Sample units are assigned at random to the m enumerators. As regards the response structure, the following model is used

$$z_{ij} = x_j + a_i + b_{ij} \tag{10.8}$$

with

$x_j =$ the true value of the j-th unit

$a_i =$ the effect of the i-th enumerator on the response

$b_{ij} =$ the random component of the response with

$\underset{i}{E}\,\underset{j}{E}\, b_{ij} = 0$ and

$\underset{i}{E}\,\underset{j}{E}\, b_{ij}^2 = \sigma_b^2$

In addition, the following symbols will be used:

$r_{ij} =$ the number of responses collected for the j-th unit by the i-th enumerator

$r_{j\cdot} = \sum_i r_{ij} =$ the number of responses for the j-th unit collected by all the enumerators

$r_{i\cdot} = \sum_j r_{ij} =$ the number of responses collected by the i-th enumerator

$r = \sum_i \sum_j r_{ij} =$ the total number of responses.

---

[11] Response errors in surveys, *Journal of the American Statistical Association*, Vol. 46, 1951, pp. 147-190.

[12] Non-sampling errors in surveys, *Journal of the Indian Society of Agricultural Statistics*, Vol. 4, 1952, pp. 5-41. See also: Sukhatme, P.V.: *Sampling Theory of Surveys with Applications*, The Iowa State College Press, Ames, and The Indian Society of Agricultural Statistics, New Delhi, 1954; the same author: Measurement of Observational Errors in Surveys, *Review of the International Statistical Institute*, Vol. 20, 1952, pp. 121-134. The present review follows the derivations given in Sukhatme's book.

It will also be assumed that each enumerator has collected an equal number of responses $\bar{r} = r/m$. Moreover, the number of responses available for each unit is equal, i.e., $s = r/n$.

Using (10.8) and the symbols introduced, the estimated average per unit has the following composition:

$$\bar{z} = \frac{1}{n}\sum_j^n x_j + \frac{1}{m}\sum_i a_i + \frac{1}{r}\sum_i\sum_j b_{ij}\, r_{ij} \tag{10.9}$$

with

$$\left.\begin{aligned} E\bar{z} &= \bar{X} + \frac{1}{M}\sum_i a_i \\ &= \bar{X} + \bar{A} \end{aligned}\right\} \tag{10.10}$$

This equation shows that the estimated mean from such a design is biased unless a_i's cancel out.

As for the variance of $\bar{z}$ we have

$$\sigma^2_{\bar{z}} = E(\bar{z} - E\bar{z})^2$$

$$= E\left[\left(\frac{1}{n}\sum_j^n x_j - \bar{X}\right)^2 + \left(\frac{1}{m}\sum_i^m a_i - \bar{A}\right)^2 + \left(\frac{1}{r}\sum_i\sum_j b_{ij}\, r_{ij}\right)^2\right]$$

$$= \frac{\sigma_x^2}{n} + \frac{\sigma_a^2}{m} + \frac{\sigma_b^2}{r} \tag{10.11}$$

If one response only is available for each unit, as will often be the case, we have $r = n$ so that (10.11) reduces to

$$\left.\begin{aligned} \sigma^2_{\bar{z}} &= \frac{\sigma_x^2 + \sigma_b^2}{n} + \frac{\sigma_a^2}{m} \\ \text{or} \quad \sigma^2_{\bar{z}} &= \frac{\sigma_x^2 + \sigma_a^2 + \sigma_b^2}{n} + \sigma_a^2\,\frac{n-m}{nm} \\ &= \frac{\sigma_z^2}{n} + \frac{n-m}{nm}\,\sigma_a^2 \end{aligned}\right\} \tag{10.12}$$

The first right-hand side term of the equation (10.12) is the usual variance of the mean in the case of simple random sampling from a large population. The second term is the addition to the variance due to enumerator effect. Accordingly, the usual formula for the variance is biased, although the bias will not be serious if the enumerators' work is highly standardized, so that σ_a^2 is small.

The equation (10.12) also shows that $\sigma^2_{\bar z}$ has the smallest magnitude if $n = m$ or, in other words, when each unit is enumerated by a separate enumerator. Although such an arrangement is not possible in practice, it is obvious from (10.12) that enumerator effect on the total variance diminishes as the number of enumerators increases.

An important alternative expression for (10.12) is obtained by Hansen, Hurwitz, Marks and Mauldin.[13] They introduce the covariance between the responses obtained by the same enumerator, or

$$\sigma_{zE} = \underset{i}{E}\,\underset{j}{E}\,(z_{ij} - \underset{i}{E}\,z_{ij})\,(z_{ij}' - \underset{i}{E}\,z_{ij}')$$

Using (10.8) it can be shown that for the assumed population with large N and M one has $\sigma_{zE} = \sigma_a^2$. Accordingly, instead of (10.12) we can write the alternative expression

$$\sigma^2_{\bar z} = \frac{\sigma_z^2}{m} + \frac{n - m}{nm}\,\sigma_{zE} \tag{10.13}$$

The equation (10.13) clearly shows the source of the additional term in the variance $\sigma^2_{\bar z}$. The covariance σ_{zE} is zero if data collected by the same enumerator are not correlated. However, if the enumerators introduce into their work some common element, such as overestimation or underestimation, σ_{zE} will not be zero.

For uncorrelated data the second term on the right-hand side of (10.13) vanishes and we are left with the usual expression of the variance as given by the first term of this equation.

To estimate the variance in (10.12) or (10.13) from data collected in a sample, we first estimate the mean for each enumerator employed. Clearly, this estimate is composed as follows:

---

[13] Response errors in surveys, *Journal of the American Statistical Association*, Vol. 46, 1951, pp. 147-190.

$$\bar z_i = \frac{1}{\bar r}\sum_j x_j\, r_{ij} + a_i + \frac{1}{\bar r}\sum_j b_{ij}\, r_{ij} \tag{10.14}$$

with the variance

$$\sigma^2_{\bar z_i} = \frac{\sigma_x^2}{\bar r} + \sigma_a^2 + \frac{\sigma_b^2}{\bar r} \tag{10.15}$$

On the basis of (10.14) $\sigma^2_{\bar z}$ is estimated by means of

$$s^2_{\bar z} = \frac{\sum_i^m (\bar z_i - \bar z)^2}{m\,(m - 1)} \tag{10.16}$$

with

$$Es^2_{\bar z} = \sigma^2_{\bar z} \tag{10.17}$$

For the estimation of the components of the variance in (10.13) the reader is referred to the original paper by Sukhatme and Seth or Sukhatme's book.

The theory presented here is handicapped by the fact that units are assigned at random to the enumerator selected. Each enumerator is supposed to travel over the whole area where the population is located. To solve this difficulty Sukhatme recommends the stratification of the population and the use of the above theory within each stratum separately.

A similar assumption is made by Hansen, Hurwitz, Marks and Mauldin. They assume that the population of N units is divided into L groups with N_h units in the h-th group. So are the M enumerators. As a result M_h enumerators are available to enumerate N_h units. To reduce traveling costs the M_h enumerators reside in the area where the N_h units are located. In the survey n units are selected at random from the whole population out of which n_h are from the h-th population group. Accordingly, n_h is a random variable. In addition, each enumerator has $\bar n$ units in his assignment. It is also assumed that there is no correlation between the data collected by different enumerators for different units. With such a design the mean is estimated[14] as

---

[14] The derivation of all the formulas that follow is found in Hansen, M.H., Hurwitz, W.N., Marks, E.S. and Mauldin, P.W.: Op. cit.

$$\bar{z} = \frac{1}{\bar{n}} \sum_h \sum_i \sum_j z_{hij} \tag{10.18}$$

The variance of $\bar{z}$ is given in (10.13). An unbiased estimate of $\sigma^2_{\bar{z}}$ is equal to

$$s^2_{\bar{z}} = \frac{\sum\limits_h \frac{n_h - 1}{m_h - 1} \sum\limits_i^{m_h} (\bar{z}_{hi} - \bar{z})^2}{m\,(n-1)} + \frac{\sum\limits_h n_h\,(\bar{z}_h - \bar{z})^2}{n\,(n-1)} \tag{10.19}$$

If the contribution of the enumerator effect to the variance is to be shown separately, we use

$$s^2_{\bar{z}} = \frac{\sum\limits_h \sum\limits_i \sum\limits_j (z_{hij} - \bar{z})^2}{n\,(n-1)} + \frac{n-m}{m(n-1)}\, s_{zE} \tag{10.20}$$

with

$$s_{zE} = \frac{\sum\limits_h \frac{m_h}{m_h - 1} \sum\limits_i (\bar{z}_{hi} - \bar{z}_h)^2}{m} - \frac{\sum\limits_h \sum\limits_i \sum\limits_j (z_{hij} - \bar{z}_{hi})^2}{n\,(\bar{n} - 1)} \tag{10.21}$$

In the equation (10.20) the first term on the right-hand side is the variance based on simple random sampling without taking into account the enumerator effect. The second term is the enumerators' contribution to the total variance.

To test the significance of the enumerator effect the following quantities are computed

$$A = \frac{1}{m} \sum_h^L \frac{m_h}{m_h - 1} \sum_i^{m_h} (\bar{z}_{hi} - \bar{z}_h)^2 \tag{10.22}$$

$$B = \frac{1}{n-m} \sum_h \sum_i \sum_j (z_{hij} - \bar{z}_{hi})^2 \tag{10.23}$$

Afterward the ratio $F = \frac{A}{B}$ is tested as above. The averages used in these equations are defined as follows:

$$\bar{z}_{hi} = \frac{\sum\limits_j^{\bar{n}} z_{hij}}{\bar{n}}$$

$$\bar{z}_h = \frac{\sum\limits_i^{m_h} \bar{z}_{hi}}{m_h}$$

It is clear that the model developed by Hansen, Hurwitz, Marks and Mauldin includes the theory presented earlier as a special case when the number of groups is reduced to $L = 1$.

Another problem is the determining of the optimum value of m and n in the equation (10.12). A simple solution is obtained in the following way. The budget of the survey is C, the average cost of getting information for an elementary unit is c_1 and the average cost per enumerator (traveling and subsistence expenses) is c_2. The simplest cost function is then

$$C = c_1 n + c_2 m \tag{10.24}$$

To get the optimum values of m and n we put

$$F = \frac{1}{n}(\sigma_z^2 - \sigma_a^2) + \frac{\sigma_a^2}{m} + \lambda\,(c_1 n + c_2 m - C)$$

From $\frac{\delta F}{\delta m} = 0$ and $\frac{\delta F}{\delta n} = 0$

we get $\frac{c_1\,\sigma_a^2}{m^2} = \frac{c_2}{n^2}(\sigma_z^2 - \sigma_a^2)$

and

$$m = n \sqrt{\frac{c_1}{c_2}\,\frac{\sigma_a^2}{\sigma_z^2 - \sigma_a^2}} = nQ \tag{10.25}$$

This gives the optimum m. The optimum n is obtained if we substitute for m from the equation (10.24) and solve for n. This gives

$$n = \frac{C}{c_1 + Qc_2}$$

In most surveys c_1 will be small compared to c_2. For a moderately large area the relationships might be $c_1 = 0.01\, c_2$. Moreover σ_z^2 will normally be much larger than σ_a^2.

It must be pointed out that this calculation of the optimum values of n and m is based on the assumption of independence in the variations in m and σ_a^2. In other words, it is assumed that the number of enumerators may be increased as desired without affecting σ_a^2. In fact, this assumption is probably valid for small changes in m. However, if the number of enumerators is to be increased substantially, it will be necessary to hire progressively less qualified personnel. This might be expected to result in a larger value of σ_a^2. At present very little information is available about the relationships between m and σ_a^2.

10.8 Interpenetrating or replicated subsamples

Interpenetrating or replicated subsamples are obtained when two or more samples are selected from the same population by the same process of selection.[15] All the subsamples together are then referred to under the name of the sample.

The selection procedure may be such that the resulting subsamples are either independent or dependent. In this section we are concerned with independent subsamples. They are obtained if the selection is carried out with replacement and in accordance with some probability system, such as sampling with equal probabilities or with probabilities proportional to a measure of size. For example, if ten independent subsamples of households are desired in a two-stage design with villages as first-stage sampling units, the aim is achieved by selecting the first village and recording its name, say, on the list representing the first subsample. The unit is then replaced and the first village of the second subsample is selected in the same way. After recording, the unit is replaced again and the selection continues in the same way until the lists of all the subsamples contain the requisite number of villages. Within the selected villages any procedure may be used in selecting the second-stage units that follows some adopted probability system. For example, households may be selected with or without replacement.

The subsamples resulting from this technique of selection are all spread over the population as a whole; the units belonging to different subsamples are interpenetrating. Each subsample is an independent replication of the same information.

The first statistician to use interpenetrating subsamples was Mahalanobis, who in 1937 designed a pilot survey of the area under jute in Bengal in the form of two interpenetrating subsamples. Since that time independent and dependent (linked or overlapping) subsamples are a standard feature of sample surveys carried out by the Indian Statistical Institute.[16] Outside India the technique has also been used by many statisticians, such as Deming,[17], Dalenius,[18] Tsukibayashi,[19] Flores,[20] Kohler,[21] etc.

A basic property of interpenetrating subsamples is the fact that each of them provides an equally valid estimate of the value being estimated.[22] If we have m subsamples of $\bar{n}$ units each with i used to designate the subsamples and j the units, we compute for each subsample the average

$$\bar{z}_i = \frac{1}{\bar{n}} \sum_{j}^{\bar{n}} z_{ij} \tag{10.26}$$

Since each subsample is a random sample from the population involved, each $\bar{z}_i$ is an equally valid and unbiased estimate of the population average

---

[15] Cf.: Kendall, M.G. and Buckland, W.R.: *A Dictionary of Statistical Terms*, International Statistical Institute, 1957; Lahiri, D.B.: *The National Sample Survey*, Report No. 5, Technical paper on some aspects of the development of the sample design, *Sankhyā*, Vol. 14, 1954, pp. 264-313; Koop, J.C.: On theoretical questions underlying the technique of replicated or interpenetrating samples, *Proceedings of the Social Statistics Sections, American Statistical Association*, 1960, pp. 196-205.

[16] Lahiri, D.B.: Observations on the use of interpenetrating samples in India, *Bulletin of the International Statistical Institute*, Vol. 36, Part 3, 1958, pp. 144-152; Mahalanobis, P.C.: Recent experiments in statistical sampling in the Indian Statistical Institute, *Journal of the Royal Statistical Society*, Vol. 109, 1946, pp. 326-378; Mahalanobis, P.C.: On large-scale sample surveys, *Philosophical Transactions of the Royal Society*, Vol. 231B, No. 584, 1944, pp. 329-451.

[17] Deming, W.E.: *Some theory of sampling*, John Wiley, New York, 1950; *Sample Design in Business Research*, John Wiley, 1960; On simplifications of sampling design through replication with equal probabilities and without stages, *Journal of the American Statistical Association*, Vol. 51, 1956, pp. 24-53.

[18] Dalenius, T.: The use of sampling methods for the estimation of areas, *Estimation of areas in agricultural statistics*, FAO, Rome, 1965.

[19] Federation of Malaya, *Census of Agriculture, 1960*. Technical report on the sample design, Kuala Lumpur, 1960.

[20] Flores, A.M.: The theory of duplicated samples and its use in Mexico, *Bulletin of the International Statistical Institute*, Vol. 36, Part 3, 1958, pp. 120-126.

[21] Kohler, S.: On the problems of replicated sampling in German governmental statistics, *Bulletin of the International Statistical Institute*, Vol. 36, Part 3, 1958, pp. 135-143.

[22] Cf. Cochran, W.G.: *Sampling Techniques*, Second edition, John Wiley, New York, 1963; Koop, J.C.: Op. cit.

$\bar{Z}$. The averages of independent subsamples are assembled in a linear combination by means of

$$\bar{z} = \frac{1}{m} \sum_{i}^{m} \bar{z}_i \tag{10.27}$$

which is the estimate of $\bar{Z}$ based on the whole sample. The estimate $\bar{z}$ is again an unbiased estimate of $\bar{Z}$. Because subsamples are independent, the variance of $\bar{z}$ is equal to

$$\sigma^2_{\bar{z}} = \frac{1}{m} \sigma^2_{\bar{z}_i} \tag{10.28}$$

An unbiased estimate of $\sigma^2_{\bar{z}_i}$ is

$$s^2_{\bar{z}_i} = \frac{1}{m-1} \sum_{i}^{m} (\bar{z}_i - \bar{z})^2 \tag{10.29}$$

Accordingly, an unbiased estimate of $\sigma^2_{\bar{z}}$ is

$$s^2_{\bar{z}} = \frac{1}{m(m-1)} \sum_{i}^{m} (\bar{z}_i - \bar{z})^2 \tag{10.30}$$

It will be noticed that $\bar{z}$ in (10.27) is based on $(m - 1)$ degrees of freedom. The comparable estimate of the average in the case of simple random sampling would be based on $(m\bar{n} - 1)$ degrees of freedom. If L strata are available each of them will contribute $(m - 1)$ degrees of freedom. Even with $m = 2$, which is often the case in the practice of the Indian Statistical Institute, $\bar{z}$ is thus based on L degrees of freedom.

The equation (10.30) shows that interpenetrating subsamples make it possible to estimate easily the variance resulting from various sample surveys. The computation involved is only a small fraction of what is needed if individual data are evaluated using a classical approach. With an interpenetrating sample design, a small number of averages will normally be involved. Computation is particularly simplified if there are only two subsamples resulting in $\bar{z}_1$ and $\bar{z}_2$. In this case the standard error in (10.30) reduces to

$$s_{\bar{z}} = 0.5 \left| \bar{z}_1 - \bar{z}_2 \right| \tag{10.31}$$

which means that the error problem is solved without even using calculating machines. This property of interpenetrating subsamples has in recent years attracted the attention of many survey designers, particularly in countries where computational facilities are not available.

In this connection it should also be noted that interpenetrating samples make it possible to compute easily sampling errors of some estimates where the expressions for errors are so complicated that traditional computations are out of the question. An illustration of this was given by Koop.[23] If the price index is estimated from prices collected by means of sampling methods, the computation of the variance by the usual methods would be a practical impossibility. However, with a design consisting of two interpenetrating subsamples the problem is easily solved. The two subsamples provide two independent estimates of the index, i.e., T_1 and T_2. Afterward the standard error of the combined index follows from (10.31) with T_1 and T_2 substituted for $\bar{z}_1$ and $\bar{z}_2$.

The use of interpenetrating subsamples therefore offers an opportunity of estimating the precision of certain quantities for which it was formerly impossible to establish the magnitude of sampling fluctuations.[24]

Another equally important property of interpenetrating subsamples is the fact that they make it easy to incorporate the enumerator effect into the magnitude of the variance. It will be seen that the equation (10.30) is identical to (10.16), which was presented earlier as a means of combining the enumerator effect and the sampling error in the magnitude of the variance. The only difference between the two is that m in (10.16) refers to the number of enumerators while in (10.30) it stands for the number of subsamples. However, reconciliation between the two is obtained if each subsample is assigned a separate enumerator. Clearly, the design of the survey is made in the form of interpenetrating subsamples so that the organization of the field work itself makes independent estimates possible. In such cases the use of interpenetrating subsamples is a fast means of obtaining estimates of variances that also take enumerator effect into account.

Credit goes to Mahalanobis for having been the first to see this possibility and to Deming who was the first statistician outside India to make wide use of this technique.

Interpenetrating subsamples have been used recently for different experimental purposes. In the United States Bureau of the Census an

---

[23] Op. cit.

[24] Cf. Kish, L.: *Survey Sampling*, John Wiley, New York, 1965.

experiment based on interpenetrating subsamples was carried out to study enumerator effect on a large number of census questions.[25] Similarly, interpenetrating subsamples can also be used for other experimental purposes. For example, in the case of establishing whether there is any difference in data as a result of different questionnaires, the training of enumerators, their education and experience, etc., the use of interpenetrating subsamples may be an efficient way of obtaining the necessary experimental evidence. The solution consists in assigning alternative procedures to different subsamples.

There are many other ways of using interpenetrating subsamples. Aggarwal used such subsamples in the preparation of the advanced estimates of the 1961 Census of Agriculture in Peru.[26] So many difficulties were encountered in the complete tabulation of the returns that there was doubt as to when or even whether it would be terminated. In this situation Aggarwal selected five independent subsamples and started processing the first of them immediately with the limited facilities available. This gave the first and most indispensable advance estimates. In addition, it was a guarantee that the effort put into the census would not be lost if complete tabulation proved impossible. It was also planned to start the processing of further subsamples if the problem of complete tabulation had not been solved in the meantime. This made it possible gradually to improve the estimates to the extent needed or to discontinue the improvement if it became certain that complete tabulation would make this unnecessary.

Such an approach may be useful in any situation where there is any likelihood of setbacks either immediately before the survey starts or in the course of implementation. If the survey is designed in such cases in the form of interpenetrating subsamples, this makes it possible to reduce the extent of operations if need be or to resume and expand them if conditions improve. This is achieved by changing the number of subsamples. Such an approach may be used even in complete enumeration censuses since they can also be designed in the form of interpenetrating samples.

---

[25] Cf. Hanson, R.H. and Marks, E.S.: Influence of the interviewer on the accuracy of survey results, *Journal of the American Statistical Association*, Vol. 53, 1958, pp. 635-655; Hansen, M.H., Hurwitz, W.N. and Bershad, M.A.: Measurement errors in censuses and surveys, *Bulletin of the International Statistical Institute*, Vol. 38, Part 2, 1961, pp. 359-374; Pritzker, L. and Hanson, R.H.: Measurement errors in the 1960 Census of Population, *Proceedings of the Social Statistics Section, American Statistical Association*, 1962, pp. 80-90.

[26] Republica del Perú, Instituto Nacional de Planificación: *Resultados Preliminares, Primer Censo Nacional Agropecuario, 1961*, Lima, 1963, mimeographed.

Fractile graphical analysis as developed by Mahalanobis[27] is another interesting way of using interpenetrating subsamples. In an application of this technique to the analysis of aggregate household expenditure on food two subsamples were used. In each of them households were first ranked in ascending order of expenditure in rupees per person. In the second step all the households were grouped into 20 equal fractile groups each containing five percent of the households. For each of the groups thus obtained averages were then calculated for a number of characteristics, such as expenditure on cereals, price of cereals, etc. The averages were computed for each subsample separately and for the combined sample as well. The data obtained were then plotted with the values of the characteristics on y axis and the cumulative percentages (of per caput consumer expenditure) on x axis. The values of the averages plotted for each group were then joined by a line. As a result, there are two lines on the graph representing subsamples and another line corresponding to the combined sample. The area between the subsample lines is hatched. This area is an indication of the magnitude of the error associated with the corresponding estimates on the combined sample line.

The advantages of the technique are: (*i*) it offers an immediate picture of the relationships between various variables over the whole range of households; (*ii*) it makes possible comparisons between the same fractile groups for different zones (rural or urban) and different years (if the survey is repeated and analyzed accordingly); (*iii*) the approximate magnitude of sampling errors is directly available on the chart; (*iv*) the approximate test of significance is available on the chart if it contains a simultaneous analysis of data collected at two different points of time or in two different areas.

The technique has also been used sometimes as an automatic check of the quality of the work of the field staff. If two subsamples are used and the field staff are grouped into two parties so that each party is assigned a subsample, discrepancies between the two subsamples can then be tested for significance.

Sukhatme[28] has pointed out in this connection that such tests may be sensitive enough for the country as a whole. They do not help, however,

---

[27] Mahalanobis, P.C.: A method of fractile graphical analysis, *Econometrica*, Vol. 28, 1960, pp. 325-351; the same author: A preliminary note on the consumption of cereals in India, *Bulletin of the International Statistical Institute*, Vol. 39, Part 4, 1962, pp. 53-76.

[28] Cf. *Sampling Theory of Surveys with Applications*, The Iowa State College Press, Ames, Iowa and the Indian Society of Agricultural Statistics, New Delhi, India, 1954.

in identifying the source of discrepancies. For this purpose tests by strata or other small areas are needed and these will normally be impossible because of the smallness of the samples: a significant number of discrepancies will not be found. In this respect Lahiri[29] presented the view that rigorous significance tests are not necessary; interpenetrating subsamples are useful to detect gross errors in survey work. It seems, however, that gross errors can be detected with some success without interpenetrating subsamples by careful inspection of data and a comparison of averages or percentages by various administrative units.

If all the enumerators to be used in a survey are divided into two groups, each with one of the two subsamples, the achievements of the two parties will, under all normal circumstances, be very much the same. Therefore, there will be no discrepancy in the estimates for the different subsamples. This will be even more so if the enumerators are assigned at random to their respective groups. This agreement, however, has nothing to do with the accuracy of the estimates. If the two parties produce equally biased data the agreement will be perfect although the final estimates are biased.[30]

It can be seen from this review that the use of interpenetrating subsamples for sample surveys has some advantages and also some disadvantages. A study of the problem in each particular case will show which of these prevail. If it is known that enumerator effect can be discounted and other means are available for obtaining fast estimates of precision, there would probably be little justification for the use of interpenetrating subsamples. If the opposite is the case, particularly in countries with no facilities for computations, interpenetrating subsamples will be an attractive device for solving problems that might otherwise be very difficult to solve. In the latter case it is useful to examine the cost effect of the application of interpenetrating subsamples. Mokashi[31] has dealt with the question of the increased cost of surveys as a result of the use of interpenetrating subsamples. The increased cost follows necessarily from the fact that, in the case of interpenetrating subsamples, area A has to be canvassed by at least two enumerators, while it is covered by only one in the alternative designs. Mahalanobis found that the interpenetrating subsamples only increase the survey budget by an average of 10 percent.[32] It is obvious, however, that the difference in the budget depends upon many factors and is probably highly variable from one survey to another. It therefore remains to be seen in each particular case how the gains that might be achieved by the use of interpenetrating subsamples compare with the losses.

---

[29] Lahiri, D.B.: Observations on the use of interpenetrating samples in India, *Bulletin of the International Statistical Institute*, Vol. 36, Part 3, 1958, pp. 144-152.

[30] Cf. Nandi, H.K.: Report on the Bihar Crop Survey: Rubi Season 1943-44, *Calcutta Statistical Association Bulletin*, Vol. 1, 1947, pp. 34-37; Ghosh, B.: Interpenetrating (networks of) samples, *Calcutta Statistical Association Bulletin*, Vol. 2, 1949, pp. 108-119; Sukhatme, P.V.: Op. cit.; Yates, F.: *Sampling Methods for Censuses and Surveys*, Third edition, Charles Griffin, London, 1960.

[31] Mokashi, V.K.: A note on interpenetrating samples, *Journal of the Indian Society of Agricultural Statistics*, Vol. 2, 1950, pp. 189-195.

10.9 Census applications

It has been pointed out on several occasions that the enumerators' contribution to the variance remains even in the results of complete enumeration censuses [cf. the equation (10.12)]. The question therefore arises as to whether complete enumeration censuses are justified under conditions where a large enumerator effect is expected. In other words, is it not preferable to take a sample survey and employ a small number of qualified enumerators rather than to embark upon a complete enumeration census with a large number of enumerators and the possibility of a considerable enumerator effect on data for small areas? Another question that arises in these circumstances is the following: is it not preferable to eliminate from the program of a complete enumeration census all those items which are associated with a large enumerator effect and put them on the program of a separate or a supplementary sample survey?

For questions of this type, the United States Bureau of the Census has conducted several experimental studies with the aim of establishing whether there are some items which could advantageously be withdrawn from the regular complete enumeration censuses and included on the program of supplementary sample surveys. The field work of such a study was carried out during the 1950 Census of Population in 24 counties of Ohio and Michigan.[33] The purpose was to estimate enumerator effect under realistic census conditions. Under normal census conditions data are collected in such a way that each enumerator is assigned one ED or sometimes more.

---

[32] Mahalanobis, P.C.: Cost and accuracy of results in sampling and complete enumeration, *Bulletin of the International Statistical Institute*, Vol. 32, Part 2, 1950, pp. 210-213.

[33] Details on this work are to be found in Eckler, A.R. and Hurwitz, W.N.: Response variance and biases in censuses and surveys, *Bulletin of the International Statistical Institute*, Vol. 36, Part 2, 1958, pp. 12-35 and Hanson, R.H. and Marks, E.S.: Influence of the interviewer on the accuracy of survey results, *Journal of the American Statistical Association*, Vol. 53, 1958, pp. 635-655.

Therefore, in a study which aims at an estimation of enumerator effect an arrangement is needed which will make it possible to estimate the mean square between EDs within an enumerator's assignment and the mean square between enumerators' assignments. Independent estimates of these two mean squares were obtained from the following design. First, 125 strata in the 24 counties were established. In these strata the EDs were made half as large as in other parts of the country. The average size of a stratum was about 6,500 persons. Within each stratum the EDs were paired at random and then each pair was assigned to a randomly selected enumerator. It was assumed that the enumerator assigned to do the work in a stratum would be able to travel over the whole stratum.

Without enumerator effect the expected value of the census total for EDs would be the same. The expected value of the two mean squares is also the same. On the other hand, significant enumerator effect is found if the mean square expressing the variations in data for EDs collected by different enumerators is significantly larger than the mean square relating to variations between pairs of EDs where data are collected by the same enumerators. F test can be applied within each stratum if the EDs' totals are normally distributed. The sum of squares is then distributed approximately as χ^2. Inclusion of several strata into the experiment does not make it difficult to use the F test provided strata mean squares are the same. With unequal strata mean squares it is necessary to reduce the number of degrees of freedom.[34]

In the study reported here it was necessary to discard the data of enumerators who did not finish the work in at least two EDs. Thus the final sample consisted of 105 enumerators and 1,489 EDs. The enumerator effect was studied for more than 100 census items so that a sufficiently detailed picture was obtained of what items are mostly affected. Hanson and Marks' paper contains a table of F values for all the items studied. It can easily be ascertained from this table that the enumerator effect is significant for many items, particularly those concerning income, school attendance, occupation, migration, "not reported," etc.

An interesting result of this study, which answers our original question, is to be found in Table 62. This shows the magnitude of error in the complete enumeration census data as resulting from the enumerator effect and the magnitude of the total error of the corresponding sample estimates based on 25 percent of the households. The latter takes into

account both components, i.e., the errors due to sampling and those due to the enumerator effect. It will be seen that the two are of approximately the same order of magnitude. The reader will also notice large errors in complete enumeration census data on income and occupation items.

A comparison of the magnitude of errors associated with the two surveys provides a strong argument for collecting on a sample basis information on many of the items in Table 62. With equally good enumerators there will be little difference between the precision obtained in the two surveys. However, both the collecting and processing of data should be less costly for the 25-percent sample of households.

Better comparisons of the two methods are secured if root mean square errors are computed. To obtain a root mean square error of census data, the information obtained in the Current Population Survey (CPS) is used. The latter information is collected by experienced enumerators who are subject to many checks in the course of their service with CPS. In addition, they are selected and trained for CPS because of their proved ability. Therefore, the difference between census and CPS data is considered for our purposes as an estimate of the bias in census information. Table 63 shows such estimates of biases for specified items.

It can be seen from this table that the "bias" varies from one item to another, both in direction and in magnitude. However, it was concluded from this table that an average absolute value of the "bias" could be placed at 7 percent. This is the estimate based on Table 63, which refers to the United States as a whole. Corresponding "biases" for smaller areas, such as individual states or some other regions, would probably be larger because some errors cancel out in data for the whole country.

Tables 62 and 63 give a basis for a comparison of the root mean square errors associated with complete enumeration censuses on the one hand and with 25-percent sample surveys on the other. If the error component in the complete enumeration census resulting from the enumerator effect, s_B^2, is approximately equal to the component due to sampling in the 25-percent sample of households, as was found to be the case in this study, the computation of the root mean square errors, on the basis of data in Tables 62 and 63, is simple. In computing root mean square errors ζ_c (c for census) and ζ_s (s for sample survey) for the number of persons falling in a population class the procedure is the following:

$$\zeta_c = [D'^2 + s_B^2]^{1/2} \qquad (10.32)$$

$$\zeta_s = [D'^2 + s_B^2 + s_s^2]^{1/2} \qquad (10.33)$$

---

[34] Hanson, R.H. and Marks, E.S.: Op. cit.

TABLE 62. - STANDARD ERRORS IN A COMPLETE CENSUS AND IN A 25-PERCENT SAMPLE SURVEY, FOR AN AREA WITH A POPULATION OF 6,500 CANVASSED BY SEVEN ENUMERATORS[1]

Characteristic	Number of persons	Percent of total population	Standard error in results of			
			Complete census		25-percent sample	
			Number	Percent of class (4) / (2)	Number	Percent of class (6) / (2)
(1)	(2)	(3)	(4)	(5)	(6)	(7)
Total population	6 500	100.0	—	—	—	—
Nativity						
Native white	5 939	90.5	62	1.04	74	1.25
Foreign-born white	239	3.6	20	8.5	33	13.9
Residence 1 year earlier						
Same house	5 264	80.3	79	1.5	97	1.8
Different house, same county	750	11.4	64	8.5	78	10.4
Different county or abroad	307	4.7	34	11.0	45	14.6
Age, males						
Under 5 years	376	5.7	10	2.7	34	9.1
15 and over	2 314	35.3	22	0.96	71	3.1
35 and over	1 360	20.7	24	1.8	62	4.6
55 and over	554	8.4	9	1.6	40	7.2
Highest grade of school attended						
Grade 5 or over	3 539	54.0	19	0.53	72	2.0
Grade 9 or over	2 103	32.1	53	2.5	84	4.0
Grade 13 or over	446	6.8	46	10.4	58	13.1
Income						
Wage and salary						
None	2 284	34.8	78	3.4	103	4.5
Under $2,500	1 209	18.4	59	4.9	80	6.7
$2,500 and over	1 081	16.5	36	3.3	63	5.8
Not reported	242	3.7	92	38.1	96	39.6
From own business						
None	4 172	63.6	96	2.3	117	2.8
Under $2,500	265	4.0	31	11.7	42	15.7
$2,500 and over	128	2.0	18	13.8	26	20.6
Not reported	252	3.8	92	36.6	96	38.2
Other						
None	3 777	57.6	106	2.8	127	3.4
Under $2,500	751	11.4	77	10.2	89	11.8
$2,500 and over	30	.5	5	15.4	10	34.9
Not reported	258	3.9	92	35.5	96	37.0
Major occupation group						
Craftsmen, foremen, etc., male	400	6.1	20	4.9	39	9.7
Farmers and farm managers, male	146	2.2	40	27.3	45	30.8
Farm, unpaid family workers, male	18	.3	14	79.2	16	89.4
Farm laborers, paid, male	30	.5	3	11.4	10	33.3
Occupation not reported, both sexes	25	.4	5	21.8	10	40.6
Industry group						
Manufacturing	1 000	15.2	44	4.4	67	6.7

[1] This table is reproduced from Eckler, A.R. and Hurwitz, W.N.: Response variance and biases in censuses and surveys, *Bulletin of the International Statistical Institute*, Vol. 36, Part 2, 1958, pp. 12-35.

with

D' = estimated bias
s_B^2 = estimated component of the variance resulting from the enumerator effect
s_s^2 = estimated component of the variance resulting from sampling

$$= (1 - f)\,\frac{Q}{nP}\,[1 + \rho\,(\bar{n} - 1)] \qquad (10.34)$$

with

n = the total number of persons in the sample
$\bar{n}$ = the average size of households
P = the proportion of persons falling in the class involved
$Q = 1 - P$
ρ = the intraclass correlation
f = the sampling fraction

Using $\bar{n} = 3.5$ and $\rho = 0.1$, which are approximately the values of these quantities that correspond to the situation found in this study, the data in Table 64 are computed. The data in the body of this table are computed for several different assumptions. Column 1 shows that four different percentage biases were used, i.e., 6, 3, 15 and 4 percent. Column 2 contains different ratios of s_B^2 to s_s^2. Column 3 contains the percentages of units falling in a class that have to be estimated in the survey. In the remaining columns the magnitude of ζ in percentage points is shown for areas with a population of 2,500, 10,000 and 50,000 persons. While for a population of 50,000 persons and over there is practically no difference in the magnitude of the root mean square error for a complete census and a 25-percent sample survey of households, these differences grow more important as one goes from larger areas toward smaller ones.

Guidance as to which combination of the bias and the variance ratio is to be used in evaluating the magnitude of the root mean square error for an item is obtained from data in Tables 62 and 63. In the case of data for which there is a relatively small difference between the 1950 Census and the CPS in Table 64 and a small s_B^2 in Table 62, the values of ζ could be found at $D' = 3$ percent and $(s_B^2/s_s^2) = 1/2$. In this particular study such data are those on occupation, industry, employment status, etc. Various income and school attainment items require a more unfavorable combination of D' and the variance ratio.

TABLE 63. - BIAS IN CENSUS DATA AS ESTIMATED BY MEANS OF CPS[1]
(in millions)

Persons of 14 years old and over	1950 Census	CPS, April 1950	Percent difference (Census-CPS)/CPS	Estimated standard error of CPS estimate
Employment status				
Total civilian noninstitutional population	109.9	109.2	.6	0
Civilian labor force	59.1	62.2	— 5.0	.32
Employed	56.2	58.7	— 4.3	.31
Unemployed	2.8	3.5	—20.0	.13
Not in labor force	50.9	47.0	8.3	.32
Keeping house	32.2	33.1	— 2.7	.29
Unable to work	4.6	2.3	100.0	.09
Other and not reported	14.1	11.6	21.6	.20
Employed persons, by major occupation group				
Total employed	56.2	58.7	— 4.3	.31
Professional, technical, and kindred workers	4.9	4.9	0.0	.13
Farmers and farm managers	4.3	4.6	— 6.5	.17
Managers, officials, and proprietors, except farm	5.0	6.2	—19.4	.15
Clerical and kindred workers	6.9	7.3	— 5.5	.16
Sales workers	3.9	4.1	— 4.9	.12
Craftsmen, foremen, and kindred workers	7.8	7.7	1.3	.16
Operatives and kindred workers	11.1	11.6	— 4.3	.20
Private household workers	1.4	1.9	—26.3	.08
Service workers, except private household	4.3	4.7	— 8.5	.13
Farm laborers and foremen	2.4	2.4	0.0	.13
Laborers, except farm and mine	3.4	3.2	6.2	.10
Not reported	.7	...	...	...
Employed persons, by major industry group				
Total employed	56.2	58.7	— 4.3	.31
Agriculture, forestry, and fisheries	7.0	7.3	— 4.1	.21
Mining	.9	.8	12.5	.05
Construction	3.4	3.4	0.0	.11
Manufacturing	14.6	14.8	— 1.4	.22
Transportation, communications, and other public utilities	4.4	4.4	0.0	.13
Wholesale and retail trade	10.6	11.9	—10.9	.20
Finance, insurance, and real estate	1.9	2.0	— 5.0	.09
Business and repair services	1.4	1.3	7.7	.07
Personal services	3.5	4.2	—16.7	.12
Entertainment and recreation services	.6	.6	0.0	.05
Professional and related services	4.7	4.9	— 4.1	.13
Public administration	2.5	2.9	—13.8	.10
Industry not reported	.8	...	...	...

---

[1] This table is reproduced from the U.S. Bureau of the Census: *The accuracy of census statistics with and without sampling*, Technical Paper No. 2, Washington, D.C., 1960.

TABLE 64. - ROOT MEAN SQUARE ERRORS OF CENSUS DATA AND THE ESTIMATES BASED ON A 25-PERCENT SAMPLE OF HOUSEHOLDS[1]

Relative response bias (percent)	$\frac{s_B^2}{s_s^2}$	Percentage to be estimated	Root mean square error in percentage points for an area with a population of:					
			2,500		10,000		50,000	
			Complete census	25-percent sample	Complete census	25-percent sample	Complete census	25-percent sample
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Case I 6	1	0.5	.27	.39	.14	.20	.07	.09
		2	.56	.78	.30	.40	.17	.21
		5	.90	1.23	.52	.67	.35	.40
		20	2.0	2.5	1.4	1.6	1.2	1.3
		50	3.6	4.1	3.2	3.3	3.0	3.1
Case II 3	½	0.5	.19	.33	.10	.17	.04	.08
		2	.39	.67	.20	.34	.10	.16
		5	.62	1.0	.33	.54	.20	.28
		20	1.2	2.0	.81	1.12	.65	.73
		50	2.0	2.8	1.6	1.9	1.5	1.6
Case III 15	1 ½	0.5	.34	.44	.18	.23	.11	.12
		2	.73	.91	.45	.52	.33	.36
		5	1.3	1.5	.91	1.00	.78	.81
		20	3.6	3.9	3.1	3.2	3.0	3.0
		50	7.9	8.1	7.6	7.7	7.5	7.5
Case IV 4	2	0.5	.39	.47	.19	.24	.09	.11
		2	.77	.94	.39	.48	.19	.22
		5	1.2	1.5	.63	.76	.33	.38
		20	2.3	2.8	1.4	1.6	.94	1.00
		50	3.4	3.9	2.4	2.6	2.1	2.1

[1] This table is reproduced from the U.S. Bureau of the Census: *The accuracy of census statistics with and without sampling*, Bureau of the Census, Technical Paper No. 2, Washington, D.C., 1960.

Obviously, the comparison between the root mean square errors in Table 64 speaks in favor of using sampling techniques for collecting data on items where the difference in ζ is negligible for data resulting from the two different methods of collecting information. If the census does not considerably reduce the error as compared with the sample survey, the latter is obviously advantageous in view of the reduced effort, time and costs. This is why it was decided in the 1960 United States Census of Population to collect information on employment status (employment and unemployment), occupation, industry, class of worker, and certain housing items from a sample of households only. The gains from such a change in the procedure are: (*i*) reduction of the delay between the collecting and publication of data; (*ii*) reduction of costs; (*iii*) the possibility of using the resources thus made available for improving certain census operations, such as listing.[35]

It should be pointed out that the relative magnitude of the two root mean squares also depends upon the method of collecting data. If the regular census enumerators are used for collecting data on the program of the supplementary sample survey, the bias in the two surveys as well as the enumerator effect will be approximately the same. In such a case the difference in the values of the root mean squares will be larger than if sample data were collected by highly trained enumerators. These might be able to collect unbiased data with no enumerator effect. On this assumption the cost of collecting data on the supplementary program increases and the whole problem becomes more complicated. However, irrespective of the relative value of the two root mean squares, from certain cases it is obvious that the study of the enumerator effect has helped to shed new light on the problems of census techniques and to indicate new directions for further rationalization of these techniques.

Another study of the enumerator effect that might usefully be studied by the reader was made in connection with the 1960 United States Census of Population and Housing. Details of the procedure used and a discussion of the main results obtained are given by Powell and Pritzker.[36]

---

[35] For further details cf. U.S. Bureau of the Census: Proposal for extension of sampling and related changes to accelerate completion and reduce costs in 1960 Censuses of Population and Housing, 1957, mimeographed.

[36] Powell, B.A. and Pritzker, L.: Effects of variation in field personnel on census results, *Demography*, Vol. 2, 1965, pp. 8-32.

10.10 Some comments

A significant enumerator effect is not an indication of biases in estimated totals, percentages, etc. Therefore, studies of the enumerator effect should not be taken as a substitute for quality checks. In fact, these studies supplement quality checks and offer information about an additional source of variation in the estimates resulting from surveys employing enumerators. Quality checks remain the only source of information about discrepancies between survey values and their respective true values.

Studies of the enumerator effect offer an insight into the degree of standardization achieved in the enumerators' work. As such these studies are particularly recommended as an experimental tool, to test the success of enumerators who are new and not yet adequately experienced in the work. In such cases, studies of the enumerator effect provide an indication of the direction in which improvements or more attention are needed. For example, questions may be singled out where the enumerators were not successful. Better instructions, additional training or improved wording may then be the remedy.

There are cases where the study of the enumerator effect must precede any decision on whether the survey is to be taken or not and, if so, in what way. For example, in collecting price data enumerators are sometimes instructed on what quality of goods they are supposed to take into account. Owing to the large number of goods on the market it is difficult to ensure that all enumerators observe the prices of goods that are equal from the quality point of view. Each enumerator may have his own interpretation of the quality that is to be selected. A study of the enumerator effect in such a situation may show whether the enumerators' criteria in the choice of the quality can be standardized and to what extent. It would also show whether the survey can be undertaken on this basis or if some other approach is needed in the selection of the goods to be surveyed.

In addition to this it seems that studies of the enumerator effect are essential in surveys devoted to political problems, social issues, preferences, attitudes, etc. This is so for two reasons. The first is that in surveys of this type quality checks are normally impossible. The only possibility is to study the enumerator effect and establish whether the enumerators influenced data or collected the information correctly according to the prescriptions of the survey. In surveys of this type the effect of supervision is also very limited and the quality of the results of the survey primarily depends upon how the enumerators carry out their work. The magnitude of the enumerator effect is an indication of the success achieved.

The second reason has already been pointed out. Namely, from the experience gained so far in this matter, enumerator effect is less likely with factual questions, such as the number of persons living in an apartment, the number of children a woman has borne, etc. With such questions the enumerator feels dependent upon the facts that he wants to establish; he is guided by the respondent and does not attempt to impose his personal interpretations of the meaning of the questions. The situation is quite different in opinion and attitude studies. In this field it may therefore be useful to design the survey in the form of replicated or interpenetrating subsamples.

The question arises here as to the treatment of enumerator effect in large-scale surveys which are currently part of government statistical activities, such as population surveys, general agricultural surveys, housing surveys, etc.

It may be too early to attempt any policy formulation in this connection as too little is known yet to permit generalizations. It seems, however, that enumerator effect in the fields mentioned here need not cause much concern provided the survey is kept under sufficient control and data are tabulated by reasonably large units. With surveys of this type there seems to be no need for a design in the form of replicated subsamples unless simplification of the variance computations strongly calls for it.

This attitude seems to tally with research results. Some of the studies presented in previous sections indicate that substantial changes in the magnitude of variances have been found after allowance had been made for enumerator effect. On the other hand, there are surveys and experimental studies where it has not been possible to find significant changes of variances as a result of enumerator effect. Some examples of this kind were presented by Mahalanobis[37] who found them in surveys of very varying types. In an experimental study conducted by Durbin and Stuart,[38] 96 values of F ratio were computed and only seven of them were found to be significant. Two of these were attributed to a vague question where the respondents were not given proper guidance. The

---

[37] Mahalanobis, P.C.: Recent experiments in statistical sampling in the Indian Statistical Institute, *Journal of the Royal Statistical Society*, Vol. 109, 1946, pp. 326-378.

[38] Durbin, J. and Stuart, A.: Callbacks and clustering in sample surveys: An experimental study, *Journal of the Royal Statistical Society*, Vol. 117, Series A, 1954, pp. 387-428.

remaining five are just the number that could be expected on the basis of random variations. This led the authors to the conclusion that enumerator effect did not appear to be a problem at least for factual characteristics.

In another experimental study carried out by Gales and Kendall,[39] the design adopted was such that it was possible to compare enumerators, various interviewing organizations, different methods of briefing, and different questionnaire designs. Although some significant differences were found in this study, it was not easy to interpret their meaning. In the discussion which took place after the paper was read it was pointed out that most of the differences were probably due to ambiguous question wording. It is likely that many of them would disappear with proper wording. This might be interpreted as a confirmation of the findings in the preceding paper. It should also be mentioned that Kish has found that enumerator effect can be either greatly reduced or eliminated from attitude questions even if they are highly ambiguous and emotionally loaded provided the necessary care is taken in selecting and training the enumerators.

This subject has also been treated in papers by Booker and David,[40] Gray,[41] Horvitz,[42] Kish and Slater,[43] Kish,[44] Price and Searles[45] and in other studies listed in the bibliography at the end of this book.

More experience is needed in order to gain a more precise view of this problem.

At this stage the difficulties that arise in the application of the theory presented in this chapter should also be pointed out. It will be remembered that the formulas for the estimation of enumerator effect were developed under the assumption of random selection of enumerators from a large pool of enumerators available for an assignment. The actual survey practice of many countries does not satisfactorily fulfill this assumption. An example in point is the case of countries where the enumerators are recruited locally. In this situation there will be neither random selection nor a pool of enumerators as the choice is often restricted to one single person in the village who is qualified enough to carry out the survey. Each enumerator is associated here with one single assignment and these are clearly conditions which make the theory of enumerator effect inapplicable.

Bearing this in mind it can be said that experiments and pretests are types of surveys that fill the conditions for the application of the theory involved. On such occasions several enumerators are normally available as a sample from a larger pool. Similarly, their random selection as well as the random distribution of assignments can easily be secured. The estimates of enumerator effect are then valid within the conditions of this experiment.

It should also be pointed out that the estimation of enumerator effect under conditions where the theory is not applicable cannot normally be achieved by using the estimates of σ_B^2 from a pilot survey and applying them in another survey by inflating the variances accordingly. If an enumerator is used to do the work in the area of his residence because of his knowledge of local conditions, σ_B^2 under such arrangements might be quite different from the results obtained with the same enumerator in a pilot survey. In the latter case the enumerator is unfamiliar with the conditions and people involved and the resulting variance per unit and its components may have a magnitude quite different from that of the results obtained in the former case.

In other conditions it may be possible to assign enumerators to various areas without prejudicing the quality of the work. This is a case where the theory of enumerator effect can be applied. Variations in the conditions may therefore make the theory of enumerator effect more or less applicable. This is why careful consideration of this matter in actual survey work is necessary.[46]

---

[39] Gales, K. and Kendall, M.G.: An inquiry concerning interviewer variability, *Journal of the Royal Statistical Society*, Vol. 120, Series A, 1957, pp. 121-147.

[40] Booker, H.S. and David, S.T.: Differences in results obtained by experienced and inexperienced interviewers, *Journal of the Royal Statistical Society*, Vol. 115, Series A, 1952, pp. 232-257.

[41] Gray, P.G.: Examples of interviewer variability taken from two sample surveys, *Applied Statistics*, Vol. 5, 1956, pp. 73-85.

[42] Horvitz, D.G.: Sampling and field procedures of the Pittsburgh Morbidity Survey, *Public Health Reports*, Vol. 67, 1952, pp. 1003-1012.

[43] Kish, L. and Slater, C.W.: Two studies of interviewer variance of socio-phsychological variables, *Proceedings of the Social Statistics Section, American Statistical Association*, 1960.

[44] Kish, L.: Studies of interviewer variance for attitudinal variables, *Journal of the American Statistical Association*, Vol. 57, 1962, pp. 92-115.

[45] Price, D.O. and Searles, R.: Some effects of interviewer-respondent interaction on response in a survey situation, *Proceedings of the Social Statistics Section, American Statistical Association*, 1961, pp. 211-221.

[46] These changes of conditions make any simple transference of knowledge in this field from one item to another or from one survey to another very difficult. One might also say that a question could be subject to a considerable enumerator effect in one survey without necessarily being so in another. The reader will find more details on this type of problem in Hyman, H. H., Cobb, W.J., Feldman, J.J., Hart, C.W. and Stember, C.H.: *Interviewing in Social Research*, The University of Chicago Press, 1954. This book is a very valuable source of reference for anybody interested in the analysis of interviewing problems.

There is another problem that deserves mention. In the development of the theory in this chapter common principles of the theory of sampling were used. It is not certain, however, that these principles are always applicable here. To clarify the point, let us start with the following example. A sample of fields under some crop is selected and the enumerator is requested to visit each of them and make an eye estimate of the yield of each parcel. The result of the enumerator's work may then easily depend upon the order of units he is using in visiting the fields. If the poor-yield fields are visited first, the enumerator may be struck by the yield seen with the result that the yield of all the other fields may be either underestimated or overestimated.

Facts of this type are very well known from psychology. The result of an evaluation of the temperature of water where the hand is submerged will depend upon the temperature to which the hand has previously been exposed. Food containing a usual quantity of salt will give the impression of not being sufficiently salty if we have been previously exposed to something excessively salty. The result of the evaluation therefore depends upon the order of the samples submitted for testing.

The first to call attention to this problem were Hansen, Hurwitz and Bershad.[47] In taking the expected values the classical theory of sampling disregards the order of units. In view of the above facts it seems necessary to adjust the theory of interview surveys in such a way as to allow for the order of units as well as its consequences. This is not an easy undertaking in our present state of knowledge.

10.11 Checking the quality of data collected by enumerators

The discussion of the quality check of data collected by enumerators follows the line of previous chapters. Here again we disregard the possibility of a repetition of the survey and are interested exclusively in the quality of data as collected in the survey being checked. The main problem is again the data collected in the check itself or the information that is to be used as a yardstick in the measurement of the quality of the original survey.

In checking the quality of data collected by enumerators it is assumed

---

[47] Hansen, M.H., Hurwitz, W.N. and Bershad, M.A.: Measurement errors in censuses and surveys, *Bulletin of the International Statistical Institute*, Vol. 38, Part 2, 1961, pp. 359-374. See also Cochran, W.G.: *Sampling Techniques*, Second edition, John Wiley, New York, 1963.

that the country is divided into M EDs with an enumerator assigned to each district. A sample of m EDs is selected as the sample of primary sampling units. The selected EDs are then subsampled and the units selected in the subsample are checked for the quality of the response. For each unit checked the variable d is established. The procedure for the estimation of the bias and its variance follows the theory of two-stage sampling.

The result of such a check will be an estimate of the response bias and its variance. In the estimated response bias all types of response errors will be taken into account. The same is true of the variance. Such a check design will not make it possible to estimate enumerator effect. This is because a positive intraclass correlation between response errors within the enumerators' assignment does not necessarily mean the presence of enumerator effect. In point of fact, if response errors are correlated with the true values and the areas assigned to different enumerators consist of units which are homogeneous within the assignments and heterogeneous between the assignments, the resulting response errors may show a positive intraclass correlation between errors within assignments and a large σ^2_{dB} without enumerators being responsible for it. The same situation may also arise as a result of the enumerators' work. Therefore, the significant magnitude of the variance ratio based on response errors does not show whether the data involved contain a significant enumerator effect or not. From the point of view of the relative magnitude of the components of the population variance of d values, a self-enumeration survey and a survey using enumerators for the collection of data might yield the same result. Therefore, for the estimation of enumerator effect the replicated or interpenetrating subsamples remain the solution.

The above design for checking the quality of the response will also be suitable under conditions different from these assumed. Namely, if a single enumerator is responsible for several primary sampling units or several enumerators work in the same primary unit, the same design will also yield a valid estimate of the bias and its variance. In such cases, however, the meaning of the "between" and "within" components of variation will be somewhat different. For example, if several enumerators carry out the enumeration in the same primary unit, σ^2_{dB} will under no circumstances express differences in the enumerators' performances.

For the presentation of the results obtained in the quality check, the reader is referred to section 10.7 as the same principles are applicable in this case.

10.12 Methods of improving the enumerators' work

A number of measures may be used to improve the quality of the enumerators' work. These measures vary, of course, in efficiency and in the facilities needed.

The preceding discussion has already shown that an important measure is the *selection* of enumerators. Once the enumerators have been selected the survey results are conditioned by the degree of success achieved by the enumerators.

In selecting the enumerators some criteria should be established to ensure the recruitment of suitable candidates. Obviously, these criteria reflect the survey circumstances in the broad sense of the word. Before such criteria are established it has to be decided what personal qualities are needed for an enumerator to be able to carry out his work successfully. The qualities normally needed are: general education, resourcefulness, intelligence, tact, etc. Sometimes age is also very important. Young enumerators may not be tolerated by the respondents in some matters which are of a delicate nature. The same is true of sex. The enumerator may have to stay in the kitchen with the housewife in a food consumption survey in order to weigh the foods before they are used for the preparation of meals, and in many countries male enumerators may not even be allowed to enter the house for this purpose. Moreover, a male enumerator may make the whole food consumption survey ridiculous by talking to housewives about food, preparation of meals, quantities of various foodstuffs used, etc.

In the preparation of the criteria for the selection of enumerators the following points may be taken into account:

(*i*) size of the survey (which is a decisive factor in the establishment of the number of enumerators);

(*ii*) sponsorship (in privately sponsored surveys which often deal with delicate matters, the enumerators must be very able in order to eliminate refusals and develop willingness to co-operate with the survey. In many cases they need to apply a "personal approach" in dealing with respondents, which is not easy unless they are very able. On the other hand, government-sponsored surveys deal primarily with factual characteristics and are preceded by a systematic publicity campaign);

(*iii*) available pay (for the field work and its rate as compared with the possibilities of alternative employment);

(*iv*) availability of candidates (with the qualifications needed at the moment of selection).

If conditions are such that supply exceeds demand, various tests can easily be devised for an objective classification of the applicants and the recruitment of the most suitable to make up the requisite number. If the methods of collecting data are complicated it may be useful to make conditions of employment attractive for the better candidates. In exceptionally difficult cases such as government-sponsored surveys the release of qualified officials in the government machinery may be needed. Examples are available of successful censuses and large-scale surveys which would have been impossible had the governments not been ready to release for the purpose their permanently employed staff, such as schoolteachers, extension work personnel, etc. Government help and co-operation will be particularly necessary in the case of specialized surveys, such as health surveys, where knowledge of the subject matter is essential.

The next measure is the training of enumerators. Training may vary in length and comprehensiveness. A training program reflects the criteria set for the selection of enumerators. With longer, more costly and more efficient training, it may be possible to develop the necessary skills even among those who are initially less able.

In the first training stage enumerators are usually made to read some relevant literature. This literature varies, of course, from case to case. In censuses, where enumerators encounter problems of enumeration, they are normally given an "Enumerator's Manual" to read. The manual contains general information on the purpose of the survey as well as the requisite details on the work the enumerators are expected to do. When they have finished this reading, a discussion takes place where questions can be raised and the more difficult problems thrashed out. An examination may then follow to eliminate those candidates who did not reach a sufficiently high level. Those remaining are then taken to the field where they are given the chance of trying out their knowledge by applying it to concrete cases. This type of enumeration is conducted in the presence of experienced enumerators who take notes of all the points that may be useful to the trainees.

The main difficulty with such training is its relative inefficiency and meager control over the enumerators' actual achievement. For this reason various measures have been introduced recently to improve efficiency. One such measure is the use of intelligence tests to eliminate, prior to the beginning of the training, those candidates who are considered incapable of acquiring the knowledge needed within the time available. Other means are getting progressively more popular, such as films, the use of tape recorders, etc.

Another useful means of training consists in the use of "role playing." Some typical enumeration difficulties are selected and introduced in an imaginary interview with the aim of testing the enumerators' reactions. The role of the respondent is played by a person responsible for training in one way or another. This person purposely gives answers leading the enumerator into difficulty and vague situations. The trainees who are not performing the enumeration are given the chance to follow the interview, consider what their solution to the difficulty would be and discuss their impressions and proposals afterward.

A common deficiency in training courses is the absence of a systematically prepared program of work. Such programs must be prepared in advance in sufficient detail so that it is always known what subject will be dealt with, what examples used, what difficulties pointed out, etc. Instead, persons responsible for training sometimes carry through the program established ad hoc. The same is true of the tests of knowledge acquired.

The details of the training program have to be arranged in advance by persons who are most experienced in various fields of survey operations. The drawing up of this program must be considered an essential phase of survey preparations. Each piece of information imparted to the enumerators has to be carefully selected in order to achieve a particular aim; the instructors must make sure that all enumerators are familiar with what is considered essential knowledge for carrying out the work successfully. Without such a detailed and centrally established program the quality of the enumerators cannot be ascertained.

Another means of improving the response is supervision. This measure may be more efficient in detecting systematic errors in data as produced by enumerators than in detecting the random errors that appear in self-enumeration surveys. If the supervisor is present at several interviews taken by the same enumerator he will often be able to detect systematic errors and correct them.

There has been considerable discussion recently regarding the intensity or volume of supervision. It appears from some factual studies that increased supervision is a good investment. In some recent surveys a supervisor was used for every 4-7 enumerators.

Another problem is the organization of the supervision. There are several possibilities, starting on the one hand with complete freedom being given to supervisors to organize their work and travel as they think best. On the other hand there is the predetermined pattern of supervision. The latter is achieved by selecting a sample of enumerators or EDs which have to be supervised, say, in the order of their selection. Of course, the enumerators are not informed of the selection.

The advantage of the former method is in the fact that it can be directed toward areas or enumerators that are expected to be less successful than the rest. The advantage of the latter system lies in the fact that it offers scientific generalization from the first results as to the quality of the field work, its problems and the resulting quality of survey data. It also offers a chance for action while the enumeration is in the field.

A training program for supervisors as well as detailed instructions on how to carry out supervision in the field, at what place and time, in what way, etc., have to be prepared in advance.

Apart from such general measures, various surveys will always offer opportunities for specific measures for the improvement of the enumerators' work.

11. SOME PROBLEMS OF QUALITY CHECKING

At this stage of our discussion it might be useful to devote some attention to a number of general problems involved in quality checking. So far there has been no opportunity to deal with such general issues as various specific aspects of quality checking had to be considered.

11.1 Reinterviewing

Reinterviewing is an important source of information in many check surveys. The methods and techniques of reinterviewing should therefore be discussed in more detail.

The simplest method of reinterviewing is the *repetition* of the interview taken in the original survey. Clearly, however, a simple repetition of the original interview will not be an adequate technique for collecting data for the check as the procedure will not be independent of the one used in the main survey. Steps are therefore needed to make repetition of the interview useful and meaningful. In some cases this may be easily achieved. Thus, reinterviewing will probably be adequate in the case of a large program in the main survey or a long delay between the two surveys, since in these cases the respondents will not remember their original response. However, with such exact repetition it is normally very difficult to guarantee the independence of check data.

A better version of this technique can be achieved by changing the *order* of questions as presented in the original survey. For example, in a census of agriculture the questions may be grouped in the following order: (*i*) holder and holding; (*ii*) land utilization and tenure; (*iii*) crops; (*iv*) livestock; (*v*) farm population; (*vi*) employment in agriculture; (*vii*) power and machinery; (*viii*) irrigation; (*ix*) fertilizers and soil dressing. In the check the order of groups may be as follows: *v*, *vi*, *iv*, *vii*, *ix*, *viii*, *iii*, *ii*, and *i*. Such a change of order, with the omission of some of the questions from the original survey, will give the impression that the aims and orientation of the check survey are quite different from those of the original survey. Consequently, the respondent is likely to start an independent train of thought rather than try to recollect the original response.

Yet another version consists in checking a survey by means of *another* survey taken before or after the survey to be checked. In this case the checking program is reduced to some key items. If the program of the new survey is considerably different from the program of the original survey there the respondent has good reason to believe that the two surveys are independent. Therefore, while giving new information he may not make efforts to recall his previous response.

Sometimes an entirely *different approach* can be used to obtain comparable information. The following is an illustration. At the United Nations/FAO Regional Census Training Center for Asia and the Far East, held in Tokyo in 1958, an experimental census of agriculture was taken and questions were asked on the total area of the holding broken down according to the main categories of land utilization and tenure. In the quality check the following procedure was adopted: (*i*) the holders were asked for the total number of separate parcels they operated; (*ii*) all these parcels were listed together with their usual names; (*iii*) the cadastral code number of each plot was obtained; (*iv*) information on the nature of each field, i.e., whether paddy, upland, etc., was obtained; (*v*) information was obtained on the area of each field separately and (*vi*) the areas of each land utilization category were totaled in order to obtain information comparable to that obtained in the census.

In the census the respondent was requested to provide direct information on the total area belonging to various categories of land utilization. The approach used in the check started from individual pieces of land and their areas. If the information given in the two surveys is not accurate a considerable discrepancy in the results is bound to appear.

Another part of the same check dealt with fertilizers. In the census questionnaire the main types of fertilizers were listed and the holders were requested to report on the total quantity used and the different areas treated with each type of fertilizer. This was possible because it was found during preliminary visits to holders that they kept records on all sorts of details regarding their uses of fertilizers. The following modified procedure was adopted in the check questionnaire. In the stub the most important crops were listed including a row for those unspecified. Holders were then requested to report on the quantity of various fertilizers used for the crops listed. Totals for the various types of fertilizers applied to different crops provide a basis of comparison with the corresponding information in the main survey.

Another procedure consists in breaking down the response process into its elements, and asking questions relating to each element separately. In this way the information requested in one question in the main survey is obtained in the check by a *series* of questions which help the respondent to recall details better. A version of this technique was discussed by Marks, Mauldin and Nisselson.[1] The example shows an attempt to check the quality of income data as collected in the 1950 Census of Population in the United States. In the census the information was obtained by using the "one-question" approach. In the check it was decided to use the "job history" approach. "The essential feature of this approach is accounting for all periods of employment or unemployment during the year. For each job held, a series of 'check' questions on income are asked and for each period of unemployment, questions are asked about unemployment compensation. The job history approach makes for a considerably longer interview..." and "... detects income which may be omitted when other techniques are used."[2]

11.2 Timing check surveys

When should check surveys be made, before or after the main survey?

In recent statistical practice check surveys have usually been conducted after the original survey has already been made. This is why they are known under the name of *postenumeration surveys* (PES). However, this does not mean that this is necessarily the case.

If the results collected in the quality check are to serve as the basis for conclusions concerning the quality of various aspects of the main survey, the sample of units used in quality checking has to be representative of the main survey as a whole. This is why a check conducted prior to the main survey may not be adequate. This will be the case particularly if the collecting of check data affects the subsequent response in the main survey. For example, if data on the total area operated by a holding as well as the areas harvested for various crops are obtained by measurements in a check survey made prior to the census, the holder may repeat this information in the subsequent census. Since measurement data are more accurate, census data for the units checked will be different from the response for the remaining units. In other words, the check has influenced the census information and so does not provide the expected quality measure of census data obtained for units not included in the check.

Obviously, this difficulty is avoided by taking the check survey as a PES. After the main survey is over the response given cannot be changed and the field staff responsible for the check are free to collect check data in any way found appropriate.

This advantage is accompanied in practice by some difficulties which are primarily related to the time lapse between the original enumeration and the check survey. For a number of reasons of an organizational nature the check survey may be made some months after the main survey. By this time many characteristics may be quite different, and it may be impossible for the check staff to establish the state of the characteristics as of the moment of the main survey. For example, in yield statistics the crop may be harvested so that no check becomes possible.

Memory lapses are another problem. These occur for many items and sometimes cause less accurate data in the check than in the original survey. If the quality of listing in a census of population is being checked, even after a short delay the unavoidable mobility of the population will result in the absence of some persons who were on the spot during the census and vice versa. Since a large number of listing errors are precisely connected with mobile persons, it is obvious that check enumerators will not be able to do a good job. Some other characteristics, such as labor force, employment and unemployment, expenditure data, food consumption, production, etc., vary from one day to another and some time after the original survey the respondents may have no idea of the state of these characteristics on the day of the main enumeration.

The necessity for reducing the time lapse between the main survey and the check survey is therefore obvious. A practical problem in this connection consists in establishing the extent to which this lapse can be reduced. Census questionnaires are often used to select the sample of units for checking. Therefore, nothing can be done until they become available. Next comes the problem of the field work of the check enumerators. Because of the need to work with a smaller number of good enumerators,

---

[1] Marks, E.S., Mauldin, P.W. and Nisselson, H.: The post-enumeration survey of the 1950 census: a case study in survey design. *Journal of the American Statistical Association*, Vol. 48, 1953, pp. 220-243.

[2] Ibid.

enumeration is often extended over a long period of a couple of weeks or months. Therefore, the delay becomes unavoidably long at least for a part of the units to be checked. Finally, in some cases the enumerators of the main survey are used for conducting the check survey. If so, time is needed to train and prepare them for their new duties.

The following is an account of the situation affecting the PES in the 1950 Census of Population and Housing in United States.[3] "The effectiveness of the PES as an evaluation of the 1950 censuses was reduced by the length of the interval between the two surveys. About 90 percent of the census enumeration was completed in April 1950, about 99 percent by the end of May, and virtually all by the end of June. The PES field work was not begun until late July, and most of it was done in August and September. Thus the average interval between the original census enumeration and the PES was between four and five months. This delay was felt necessary in order to avoid the risk of biasing the census results by identifying the PES sample areas or units before the census was entirely completed. The time lag undoubtedly affected the ability of the PES interviewers to obtain an accurate count of the persons present at the time of the original enumeration, and accurate information for characteristics difficult to recall after the passage of time. For this reason, the PES was not used to evaluate census data on labor force status where the time delay would be particularly important. Unquestionably the usefulness of the PES results for some other characteristics, and particularly for coverage, was affected by this time lag."

The timing of quality checks is therefore a question to be decided after consideration of the particular circumstances involved. In solving it two main principles should be borne in mind. They are:

(*i*) the check should not influence the main survey, and
(*ii*) the time of the check must be selected in such a way as to make it possible for the check to achieve its purpose.

With regard to the first principle the check survey may be made either before or after the main survey. However, if the method of collecting data in both surveys is interviewing or some other procedure where the response depends upon the respondent, there will be no alternative to the postenumeration checking. This is why check surveys in most censuses take the form of PES. On the other hand, if the check aims can be achieved by concentrating on certain sources of information independent of the respondent, there is no reason why pre-enumeration checks should not be used. In some countries and for some characteristics it may be possible to have recourse to files, school certificates, population registers, etc., so that neither the respondents nor the enumerators of the main survey are aware of the existence of the check. Similarly, the measurement of areas could also sometimes start prior to the main survey.

The second principle is decisive in determining the delay between the two surveys. If the survey program is such that the characteristics involved change fast, the delay must be short. If it cannot be made so for practical reasons the program of checking may usefully be reduced or the check abandoned altogether. Checking data on food expenditure for a given day, to be at all feasible, must immediately follow the collecting of the original response. This is also the case with checking the quality of listing. If the latter is considered necessary, it may call for some steps in the design of the check survey that would not be needed otherwise. For example, the quality check of a census of population may be spilt up into two different surveys: one for checking the quality of the listing to be made immediately after the original listing and the other to check the quality of the response to be made later.

The delay between the two surveys can also be reduced sometimes by a skillful use of various devices, such as using alternative sources of information. Thus, heads of households may be used as the source of information in the main survey and housewives in the check survey or vice versa. Clearly, this device can only be used if the characteristics involved are equally known to both members of the household. This is so with basic demographic data. In other fields, such as expenditure, consumption, agricultural production, etc., the two sources are not equally good.

Other devices are based on a skillful use of different techniques for obtaining data. Data on food consumption as obtained by any of the main methods of collecting data could be checked by a parallel collecting of information on expenditure. Another possibility is to ask for the menu of each meal. Staff conversant with the preparation of food will know what ingredients are needed for different dishes and this will provide a basis for the evaluation of the information presented in the main survey.

It can be seen that many facts must be considered before a decision is taken with regard to the timing of the check.

---

[3] U.S. Bureau of the Census: *The Post-Enumeration Survey: 1950*, Technical Paper No. 4, Washington, D.C., 1960.

11.3 Advance knowledge of information being checked

Should the check enumerators be provided with the information they are checking?

In some checks the enumerators are provided with the information collected in the main survey while in others they collect check data without knowing what response was given in the main check. Combinations of the two procedures are also possible. In the check of the field work of the current Population Survey in the United States, supervisors are assigned segments in which to carry out the check of both the listing and the response. In the check of the listing they are provided with the lists already prepared. In the check of the response they make an independent reinterview. After the reinterview has been completed they compare the response with data obtained in the original survey and make a list of discrepancies which are then reconciled in the presence of the person concerned. An additional visit to the field is thus needed for reconciliation purposes.[4]

Another useful combination would be to divide the enumerators' assignments or EDs into random subsamples and use a different procedure in each subsample. After the check has been completed it should be possible to compare the results. There may therefore be two subsamples, in one of which the enumerators know the response collected in the main survey, while in the other they conduct an entirely fresh survey. The comparison will then show whether any significant difference was obtained.

There is obviously some merit in both procedures. Supplying enumerators with the data being checked is a way of providing useful guidance; the main work is by then basically done and the enumerators can easily concentrate on detecting the shortcomings of what was done. In addition, if there is any disagreement between the two surveys it can be reconciled immediately on the spot by carrying on the investigation until sufficient information is obtained for a final decision. If some characteristics of cases subject to disagreement have to be collected for the purpose of subsequent research on census techniques, this can be done as soon as a case is detected. In other words, the procedure saves expense and effort.

If the enumerators collect independent and fresh information, they will partly be repeating work which has already been done. In addition, after they have carried out their independent work they have to go to the commune or some other office where the census material is located. The matching of data will be done in the presence of the local census agent and a list of disagreements compiled. The reconciliation of disagreements then becomes the subject matter of a new field visit, where the check enumerator is accompanied by the local census agent and possibly by the corresponding census enumerator who, in many cases, may be available on the spot. In fact, this was the procedure applied in the quality check of the 1953 Census of Population in Yugoslavia.

A serious disadvantage of the latter approach is the new field visit for the purpose of reconciliation and the additional time lost by the whole team. The procedure is sometimes adopted because it has been found that enumerators, irrespective of the degree of their training, tend to repeat the information available. The experiment described by Hansen and Pritzker in section 3.4 gives strong evidence of this tendency.

In addition, the use of an external and impartial judge in matching and reconciling the results of two listings is also considered useful because it puts pressure on check enumerators to do as good a job as possible. With such an arrangement they cannot simply reject the results obtained by census enumerators as they might tend to do for prestige reasons if they were completely independent in their decisions. It is therefore believed that two independent census listings combined with matching and reconciliation will have considerably more chance of providing accurate lists. However, this procedure is definitely more costly.

A decision as to which of these two procedures should be used in a specified case will depend largely upon the organization of statistics and local facilities. If local census agents are available and can be used for matching and reconciliation without seriously affecting the cost of the check, there may be some advantage in requesting the enumerators to do all the work again. Of course, other factors are also important. The most efficient use of highly reliable enumerators may be to give them a chance to concentrate directly on discrepancies. Moreover, if the check is carried out a long time after the census, there can be hardly any possibility of counting on the services of the census organization. It will already be dissolved by that time.

---

[4] For further details see Hansen, M.H. and Steinberg, J.: Control of errors in surveys, *Biometrics*, Vol. 12, 1956, pp. 462-474.

11.4 Origin of errors

It has already been pointed out that quality checks have two purposes. The first is to provide users of data with some criterion of the quality of the information they are using. The other is to provide information about the deficiencies of the methods used and suggestions for possible improvements.

The most important knowledge that can be obtained from quality checks for the purpose of an appraisal of the adequacy of methods of work and for designing measures for the improvement of the work is information about the *origin of errors*. The information needed in this connection will cover (*i*) the *characteristics* of units affected by errors and (*ii*) the *causes* of various types of errors. By the characteristics of units affected by listing errors in a census of agriculture we understand the size of the holdings, the area of various land utilization categories, the tenure of land, the type of farming, the number of various kinds of livestock, the value of produce, etc. Similarly, in a census of population it is useful to know various characteristics of the persons omitted in census listings, such as their age, sex, occupation, the size of the household they live in, whether the household concerned was listed or not, etc. With this information the statistician has quite precise indications on the directions where action for the improvement of the quality of listing is needed as well as on the measures that may be considered relevant.

With regard to causes it is useful to know *why* and *how* the particular error occurred. If children or infants are relatively often omitted in census listings, it is very useful to know why this happens. Is it because people consider that censuses have nothing to do with children or is there some other reason? If the former is the case, emphasis on this point in the census publicity might be a solution. If people understate their income and it is found that this is so because they do not take pension and social insurance benefits into account, the instructions might be improved so that the components of income become clear to each respondent.

Information on the characteristics of the units affected and the causes of errors is complementary. The former helps to locate the segment of the population where measures for the improvement of survey techniques have to be directed; the latter shows what measures may be considered as a means of improvement.

The knowledge needed here cannot be obtained unless the check survey is designed accordingly. In order to secure sufficiently broad information on the characteristics of the units affected by various types of errors it is necessary to prepare in advance a list of the characteristics involved so that they can be incorporated in the questionnaire of the check survey. Once the field work is over it becomes impossible to broaden the list of these characteristics without considerable additional effort.

Collecting data on the causes is normally more difficult, because in the case of many errors several causes are possible and there is no means of distinguishing which one is involved. For example, an error in age might be due to recording, misunderstanding, ignorance on the part of the respondent, deliberate deviation from the true value to achieve some effect, etc. This is why the easiest approach in collecting data on the causes of errors consists in requesting the check enumerators to present their impressions on the possible causes of the errors detected. These impressions are normally collected while a reconciliation is being made of the discrepancies between the check and the survey. This is an opportunity for the check enumerators to obtain additional information about the units affected and to have a discussion with the people concerned. Of course, the resulting data on the causes of errors contain many subjective elements.

The collecting of data on causes also needs to be planned in advance. On the basis of previous experience a list of likely causes is prepared for each characteristic. These lists are then introduced into the questionnaire so that the check enumerators can indicate those that, in their opinion, correspond to the case involved. Causes which are not specified in advance can be described. However, description cannot easily be taken into account in processing.

Such studies on the causes of errors have been undertaken by the United States Bureau of the Census,[5] Hansen and Steinberg,[6] and the present author.[7]

11.5 Joint effect of listing and response errors

According to their purposes, quality checks are normally broken down operationally into separate parts. For example, in checking the quality

---

[5] U.S. Bureau of the Census: *The Sample Survey of Retail Stores, A report on methodology*, Technical Paper No. 1, Washington D.C., 1953.

[6] Hansen, M.H. and Steinberg, J.: Control of errors in surveys, *Biometrics*, Vol. 12, 1956, pp. 462-474.

[7] Zarkovich, S.S.: *Population Census Errors* (in Serbo-Croatian), Federal Statistical Office, Belgrade, 1954.

of a complete enumeration census three separate checks may be planned: (*i*) the quality of listing, (*ii*) the quality of the response and (*iii*) the quality of the processing. Of course, additional checks can also be added, such as those based on techniques other than sampling. The problem then arises of how to combine the results of the separate parts in order to form a general picture of the quality of the data.

The problem of combining data from a sample check and a check based on techniques other than sampling does not arise. If available, such results are normally presented parallel to each other and represent two complementary pictures of the quality of the data. The problem that arises here is the combination of the results of separate quality checks carried out by means of sampling methods. In other words, if sample checks of listing and response are made, the problem is to see how their results can be assembled in a composite quality estimate.

Obviously, the solution of this problem has to reflect the underlying circumstances. For this reason the principles for solving the problem will be emphasized here on the understanding that the details of the design must be fixed in each case separately after consideration of the relevant conditions.

In presenting a possible solution to the problem of combining the data of a listing check and a response check it is first assumed that a sample of EDs was selected. Assuming further that the check refers to the data of a census of agriculture, the sample of EDs selected is used to check the quality of the listing of agricultural holdings along the lines described in Chapter 6. From the holdings listed in the census in the selected EDs (or in another sample of EDs) a subsample of holdings is selected and the data of the holdings involved are checked for the quality of the response. For analytical and study purposes, data from the two checks are assumed to be available in the form of the Tables A, B, D and E given in chapters 6 and 9.

In fact, combining data from the two checks consists in putting together the data in Tables A, B, D and E. The combined effect of listing and response errors on the frequencies of the size classification of holdings is obtained by pooling data in Tables A and D, while the combined effect on the total area of agricultural holdings is obtained by putting together data in Tables B and E.

In estimating the total bias in the number of holdings belonging to a size class we put

$$D' = D_1' + D_2' \tag{11.1}$$

D_1' being the contribution of listing errors and D_2' the contribution of response errors. D_1' is obtained from the equation (6.4) while D_2' is based on a two-stage sample and follows from the equation (6.11) using d_{ij} instead of D_{ij}. Accordingly, if an equal second-stage sampling fraction f_2 is used in all the EDs the equation (11.1) could be written as

$$D' = \frac{M}{m}\left(\sum_i^m D_i + \frac{1}{f_2}\sum_i^m \sum_j^{n_i} d_{ij}\right) \tag{11.2}$$

The structure of the variance of D' is the same as the variance of D' in the equation (6.13). In this case we also neglect the covariance term on the assumption that the bias due to listing errors is independent of the bias due to response errors. The variance of D' is thus reduced to two components, $\sigma^2_{D_1'}$ and $\sigma^2_{D_2'}$. Approximate estimates of these components are given in equations (6.5), (6.7) and (6.12). With an equal second-stage sampling fraction the two components could be put together as follows:

$$s^2_{D'} \doteq \frac{M^2}{m\,(m-1)}\left[\sum_i^m D_i^2 + \left(\frac{1}{f_2}\right)^2 \sum_i^m \left(\sum_j^{n_i} d_{ij}\right)^2\right] \tag{11.3}$$

In testing the significance of the bias D' we again use the same type of approximation as in the equation (6.8). The corresponding ratio will be

$$\frac{D'}{s_{D'}} \doteq \frac{\sum_i^m D_i + \frac{1}{f_2}\sum_i^m \sum_j^{n_i} d_{ij}}{\left[\sum_i^m D_i^2 + \left(\frac{1}{f_2}\right)^2 \sum_i^m \left(\sum_j^{n_i} d_{ij}\right)^2\right]^{\frac{1}{2}}} \tag{11.4}$$

The combined effect of listing errors and response errors can be presented as in Table F.

Data for the lower part of Table F are obtained in a similar way although they are based on the information available in Tables B and E.

For discussion of many details that arise in the pooling of data from checks of listing and checks of response the reader is referred to the United States Bureau of the Census: *The Post-Enumeration Survey, 1950; An Evaluation Study of the 1950 Census of Population and Housing*, Technical Paper No. 4, Washington, D.C., 1960.

TABLE F. - COMBINED EFFECT OF LISTING ERRORS AND RESPONSE ERRORS ON VARIOUS CENSUS CHARACTERISTICS

			Size classification of holdings (in ha)				Total
			0.5 - 5	5 - 10	10 - 20	20 - 50[1]	
Number of holdings	Census	Number					
	Bias	Number					
		Percent					
	Significance test						
Total area	Census	Number					
	Bias	Number					
		Percent					
	Significance test						

[1] Columns in this sequence can be extended as far as is desired.

11.6 Report of the check

When the check is over and the data collected have been tabulated and analyzed, the results achieved have to be presented. The persons interested in these results are users of data on the one hand, statisticians on the other. The data users are interested in the check as a source of information about the quality of the data in the main survey, the extent to which the check supports the main survey data and the adjustments that may be needed in the light of the information gathered in the check. The statisticians are interested in the design of the check, the methods of collecting data, the results achieved by the enumerators, the organization of their training, the supervision of field work, the analysis of errors in various items, and all sorts of empirical information that may be useful in designing quality checks to improve survey techniques. Since these two interests are divergent, it may be reasonable to present the check in such a way that the users of data are not bothered with facts and analyses which do not concern them. This can easily be achieved if the report is split up into two parts, which might also be published separately. The first part could be called *general report of the quality check* and the second *technical report of the quality check*. Of course, in both cases the name of the survey is added as the main title.

The general report could be composed of the following sections:

A. Design of the check (which is intended to give the reader a general description of the check so that he knows enough to understand the background of data presented later). This section will deal with the following issues: aim of the check, program of the check survey, organization of the check, responsibilities, recruitment of field staff and supervisors, their training, payment, etc.; sample, methods of collecting check data.
B. Supporting evidence (regarding the quality of the data collected in the check).
C. Description of the tables presented.
D. Use and interpretation of data (which provides a brief explanation of the concepts of quality, bias, mean square error, etc., and shows what significance is, and how the information on significant biases is used in adjusting the estimates or census figures, etc.).
E. Conclusion (which is supposed to present an overall picture of the work done in the check and in the main survey with an indication of the weak points, the direction of further studies, the improvements needed in survey techniques, etc.).
F. Tables.

As for the technical report, it does not represent an alternative document but a supplement to the information presented in the general report. The information presented in the general report is not repeated in the technical report.

As regards the composition of the technical report, the following might be the items included:

A. Aims of the check (with emphasis on points which are predominantly methodological and of special interest to statisticians. Reference can be made here to earlier studies, the problems they raised, the aspects of errors that need to be studied further, etc.).
B. Sample, its selection, efficiency studies.
C. Methods of collecting data (with a detailed account of various measures taken with the aim of improving the enumerators' work).
D. Organization of the check (the agencies involved, responsibilities, timing of the field operation, its execution, etc.).

E. Supporting evidence on the quality of data collected in the check; how the collecting was organized.
F. Analysis of errors (all the experience and knowledge available on the origin of errors).
G. Suggestions for the improvement of survey techniques (based on facts and experience presented under F).
H. Problems that need further study.
I. Efficiency considerations (costs, variance components, possible alternative designs).
J. Suggestions for the improvement of future quality checks.

The appendix of the report might contain the questionnaire, a more detailed description of certain stages of the work, such as supervisors' or enumerators' training, instructions for the collecting of supporting evidence, instructions for measurements, etc.

Of course, any useful experience not included in the schema presented here should be incorporated in the report.

The adjustment of the main survey data for biases found in the check is not attempted. This is primarily because biases may be successfully estimated for some totals only while indications are normally missing on how to divide the total bias over the various classes established. Proportional apportionment of the total bias, which will be the only possible solution in the absence of further details, cannot be considered because experience has shown that biases are often concentrated in certain classes. For this reason tables containing data from the original survey are left intact and the results collected in the quality check are presented separately.

12. CHECKING THE QUALITY OF PROCESSING

12.1 Introductory remarks

Data processing can be broken down into phases, the number of which depends upon many factors, such as the equipment used for processing, quality requirements, etc. In those censuses and large-scale surveys where conventional punch card equipment is used the number of phases is sometimes relatively large. These phases might be:

(*i*) Field editing, i.e., the first control of data in the field carried out immediately after the enumeration is over and while enumerators are handing over questionnaires either to their respective supervisors or to local census commissions.
(*ii*) Office editing, performed with a view to correcting inconsistencies and filling in missing responses. This might be done either centrally as the first phase of processing or in some district and provincial offices so that the units concerned are near should new contacts prove necessary.
(*iii*) Quality control of office editing.
(*iv*) Coding.
(*v*) Quality control of coding.
(*vi*) Card punching.
(*vii*) Verification of card punching.
(*viii*) Mechanical editing.
(*ix*) Tabulation.

The number of phases can, of course, be increased or decreased according to circumstances.

The phases listed here could be divided into two groups. The first is composed of *data-transforming phases*. As the name implies, these are phases where the data are put into a new form. The phases involved are coding, punching and tabulation. The second group is composed of *error-correcting phases*, the meaning of which is clear.

With more modern processing techniques, such as mark sensing, the number of processing phases is considerably reduced. Questionnaires filled in by means of special pencils or ink are passed through a machine where marks made produce automatic card punching. Accordingly, the basic phases are card punching and mechanical editing.[1] The latest device in the field of data processing known under the name of FOSDIC (Film Optical Sensing Device for Input to Computers) goes even further. Questionnaires filled in with any kind of writing instrument are photographed on 16-millimeter microfilm. Afterward the films are scanned by FOSDIC which reads and records on magnetic tape the information available in the questionnaires. The material thus made available is ready for mechanical editing.[2]

The use of sampling methods in connection with such processing devices is outside the scope of this book. The subject matter of the present chapter is the use of sampling methods for the purpose of checking the quality of various processing phases in complete enumeration censuses tabulated by means of conventional punch card equipment. In many past censuses a complete quality check of office editing, coding and punching was undertaken. Clearly, the cost involved was relatively high. The use of sampling methods for this purpose will reduce the amount of material to be handled and the cost of the operations will be reduced. This chapter contains a review of techniques that have sometimes been used to achieve this aim.

Interest in the use of sampling methods to check the quality of the processing of large-scale surveys seems to have declined in recent years. This is because of the fast-growing use of electronic computers. Modern equipment eliminates the traditional intermediate phases and so the approaches to quality control take somewhat different forms. Among these forms are the automatic identification and correction of errors. Much effort is at present being spent establishing the rational principles of automatic data correction.[3] Work on these problems is in full progress but it will be some time before wide enough empirical information becomes available from the various fields involved.

---

[1] Cf. Statistical Office of the United Nations and Statistics Division, FAO: *Handbook on Data Processing Methods*, Rome, Part I, 1959 and Part II, 1962.

[2] Daly, J.F. and Eckler, R.A.: Application of electronic equipment to statistical data-processing in the U.S. Bureau of the Census, *Bulletin of the International Statistical Institute*, Vol. 39, Part 4, 1962, pp. 319-327.

[3] Cf. Nordbotten, S.: Automatic editing of individual statistical observations, Conference of European Statisticians, WG.9/37, 1962 and the bibliography in this paper. See also Nordbotten, S.: The efficiency of automatic detection and correction of errors in individual observations as compared with other means for improving the quality of statistics, *International Statistical Institute*, Belgrade Session, 1965; Pritzker, L., Ogus, J. and Hansen, M.H.: Computer editing methods; some applications and results, *International Statistical Institute*, Belgrade Session, 1965; Szameitat, K. and Zindler, H.J.: The reduction of errors in statistics by automatic corrections, *International Statistical Institute*, Belgrade Session, 1965.

12.2 Aims of quality checking

What are the aims of quality checks of processing?

One such aim consists in getting information on the quality of a processing phase after it is over. For example, after the process of coding questionnaires is over, it may be desirable to know what quality level has been achieved in the work on various items of the program. This may be the case if a study of the old material is to be undertaken with a view to establishing rates of coding errors. The results of such a study can be used to prepare more complete and better instructions for future coding.

Quality checking connected with such an aim will be called post hoc quality checking of processing errors.

By their very nature the post hoc quality checks cannot be used to influence the quality of processing in progress. In many cases, however, this is just the desired aim. For this reason checking may be carried out to obtain current information about the quality of the work while the process is going on. This information is a practical tool because it makes it possible to discontinue the process if the quality is below a certain standard, or to take some measures to improve the quality level. By analogy with industrial quality control, this type of check is called *process checking*. Its aim is the protection of the quality of the work, made possible on the basis of current information about this work.

12.3 Post hoc checking

The use of sampling methods for the purpose of post hoc checking is relatively simple. Irrespective of the phase of processing that needs to be checked, a convenient sample of the material offering the desired data is selected and all the relevant information is recorded from the documents concerned. These documents may be either original census questionnaires or punch cards if the aim is to study the quality of punching.

The information must be recorded in such a way that the analysis of errors becomes possible. Namely, in addition to the rate of errors in processing data for various questions, it is also useful to know the direction of errors, their magnitude, the characteristics of the units, etc.

The usefulness of post hoc quality checks of processing will be illustrated by means of the results obtained in the processing of the 1953 Census of Population of Yugoslavia.[4] The study was carried out after the processing had already been completed. The aim was to obtain guidance for designing a more efficient processing program for the next census.

The processing of this census was broken down into the nine phases listed at the beginning of this chapter. In each stage the work extended over all census questionnaires or punch cards. The first phase, i.e., the field editing, was carried out by the field staff and no trace of its effect was left on the questionnaires. It was therefore excluded from the study. Fortunately, the effects of the remaining phases were visible on the questionnaires. Since pencils of different colors were used in these phases, it was possible to see all the changes introduced on the questionnaires in the course of the work.

For the purpose of this study it was considered that an error was present whenever a discrepancy was detected between the action taken (survey value) on the one hand, and the information on the questionnaire of the action that was supposed to be taken on the basis of that information on the other hand. In some cases it was not possible to carry out the necessary correction. For example, in the course of editing a female name may be found in combination with the masculine sex. If there is no other information the clerk cannot decide which of the two possibilities is the wrong one. In order to facilitate the clarifying of such cases additional information is requested: letters are sent to field staff who have to revisit the units concerned and clarify the matter.

To study the effects of various phases of processing a sample of 10,882 questionnaires or corresponding punch cards was selected from about 3,900,000 questionnaires from the Republic of Croatia. To check the quality of the office editing the original response was compared with the changes introduced in the stage of office editing. Errors corrected in the original data were classified into those resulting from the study of the questionnaire (inconsistencies, missing data) and those based on the information obtained through subsequent correspondence. Editing errors were also recorded separately. The absolute number of errors of each type as found in the sample of 10,882 questionnaires is presented here in Table 65, which is based on a somewhat different table in the paper by Gjukich-Srdar and Bedenich. The meaning of the various headings in this table is clear. Table 66 shows the number of census errors corrected and the number of errors committed in the phase of office editing. Table 67 contains an additional column showing the number of errors corrected out of those committed in the phase of the quality check of office editing. A similar extension is made in Table 68 for coding errors. Table 69 refers to errors of various phases of the machine processing, i.e., punch cards, verification, and mechanical editing. In the post hoc check it was not possible to separate the effects of various phases of the machine pro-

TABLE 65. - NUMBER OF ERRORS COMMITTED AND CORRECTED IN THE PHASE OF OFFICE EDITING AS OBTAINED FROM A POST HOC CHECK OF THE EDITING OF CERTAIN QUESTIONS BASED ON A SAMPLE OF 10,882 QUESTIONNAIRES SELECTED IN THE REPUBLIC OF CROATIA FROM THE 1953 CENSUS OF POPULATION IN YUGOSLAVIA [1]

Serial number	Item	Number of census errors corrected			Errors committed in this phase
		Without corre-spondence	With corre-spondence	Total	
(1)	(2)	(3)	(4)	(5)	(6)
2	Sex	17	—	17	8
3*a*	Date of birth	2	1	3	20
b	Month of birth	3	3	6	1
c	Year of birth	3	7	10	4
4	Place of birth	95	—	95	193
5	Marital status	63	—	63	7
6	Number of marriages	38	2	40	3
7	Age at first marriage	705	13	718	43
8	Number of children born	11	—	11	7
9	Number of children living	12	—	12	7
10	Citizenship	13	—	13	1
11	Nationality	13	—	13	2
12	Mother tongue	11	—	11	2
13	Religion	39	—	39	4
14	Literacy	48	—	48	12
15	Degree of education	203	—	203	82
16*a*	Active or not	36	6	42	11
b	Dependency status	55	10	65	17
c	Occupation	168	59	227	71
17	Status	235	21	256	21
18	Type of activity	420	18	438	60
19	Sector of economy	383	12	395	25

[1] Tables 65 to 69 inclusive are based on data published in Gjukich-Srdar and Bedenich: Op. cit.

---

[4] All the information presented here on this check is taken from Gjukich-Srdar, M. and Bedenich, B.: The efficiency of data processing in the 1953 Census of Population and its influence on the accuracy of data (in Serbo-Croatian), *Statistička Revija*, Vol. 9, 1959, pp. 297-336. See also: Macura, M. and Balaban, V.: Yugoslav experience in the evaluation of population censuses and sampling, *Bulletin of the International Statistical Institute*, Vol. 38, Part 2, 1961, pp. 375-399.

cessing and for this reason Table 69 contains a summary of the three phases. During the machine processing a part of the errors committed in the census was also corrected (column 3) while some additional errors were introduced (column 4). Column 5 shows the total number of errors corrected out of those which were committed in all the phases.

The reason for introducing these tables here is that such material is only rarely available.

The first useful indication given by these tables is the relatively large number of errors committed in various phases of processing compared with the number of corrected errors. It will be noted that the former is in some cases larger than the latter. For decision purposes it would, of course, be very important to know what kind of errors are dealt with.

TABLE 66. - NUMBER OF ERRORS COMMITTED AND CORRECTED IN THE PHASE OF QUALITY CONTROL OF OFFICE EDITING AS OBTAINED FROM A POST HOC CHECK CARRIED OUT ON CERTAIN QUESTIONS BASED ON A SAMPLE OF 10,882 QUESTIONNAIRES SELECTED IN THE REPUBLIC OF CROATIA FROM THE 1953 CENSUS OF POPULATION IN YUGOSLAVIA

Serial number	Item	Number of errors corrected			Errors committed in this phase
		Census errors	Errors in office editing	Total	
(1)	(2)	(3)	(4)	(5)	(6)
2	Sex	10	—	10	—
3*a*	Date of birth	—	7	7	—
b	Month of birth	—	—	—	—
c	Year of birth	1	1	2	—
4	Place of birth	4	108	112	14
5	Marital status	23	2	25	2
6	Number of marriages	13	—	13	1
7	Age at first marriage	86	9	95	4
8	Number of children born	11	—	11	1
9	Number of children living	12	—	12	1
10	Citizenship	3	—	3	—
11	Nationality	2	1	3	—
12	Mother tongue	2	1	3	—
13	Religion	12	1	13	1
14	Literacy	5	6	11	—
15	Degree of education	6	42	48	13
16*a*	Active or not	7	7	14	—
b	Dependency status	20	6	26	3
c	Occupation	45	21	66	21
17	Status	57	7	64	3
18	Type of activity	92	21	113	25
19	Sector of economy	64	14	78	13

TABLE 67. - NUMBER OF ERRORS COMMITTED AND CORRECTED IN THE PHASE OF CODING AS OBTAINED FROM A POST HOC CHECK CARRIED OUT ON CERTAIN QUESTIONS BASED ON A SAMPLE OF 10,882 QUESTIONNAIRES SELECTED IN THE REPUBLIC OF CROATIA FROM THE 1953 CENSUS OF POPULATION IN YUGOSLAVIA

Serial number	Item	Number of errors corrected				Errors committed in coding
		Census errors	Errors in office editing	Errors in quality control of office editing	Total	
(1)	(2)	(3)	(4)	(5)	(6)	(7)
2	Sex	—	—	—	—	25
3*a*	Date of birth	—	3	—	3	40
b	Month of birth	—	—	—	—	13
c	Year of birth	—	—	—	—	42
4	Place of birth	—	2	2	4	55
5	Marital status	—	—	—	—	34
6	Number of marriages	1	—	—	1	12
7	Age at first marriage	3	1	—	4	41
8	Number of children born	—	—	—	—	20
9	Number of children living	—	—	—	—	18
10	Citizenship	—	1	—	1	3
11	Nationality	—	—	—	—	28
12	Mother tongue	1	—	—	1	33
13	Religion	—	—	—	—	17
14	Literacy	—	—	—	—	23
15	Degree of education	—	2	1	3	98
16*a*	Active or not	—	—	—	—	20
b	Dependency status	1	—	1	2	51
c	Occupation	3	—	—	3	138
17	Status	3	1	—	4	53
18	Type of activity	2	1	6	9	70
19	Sector of economy	4	—	—	4	36

Table 68. - Number of errors committed and corrected in the phase of the quality control of coding as obtained from a post hoc check carried out on certain questions based on a sample of 10,882 questionnaires selected in the Republic of Croatia from the 1953 Census of Population in Yugoslavia

Serial number	Item	Number of errors corrected					Errors committed in this phase
		Census errors	Errors in office editing	Errors in the quality control of office editing	Coding errors	Total	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
2	Sex	3	1	—	17	21	—
3*a*	Date of birth	—	2	—	26	28	—
b	Month of birth	—	1	—	10	11	—
c	Year of birth	—	1	—	33	34	2
4	Place of birth	2	21	10	37	70	1
5	Marital status	4	—	1	29	34	1
6	Number of marriages	2	—	1	12	15	—
7	Age at first marriage	4	2	1	31	38	3
8	Number of children born	2	2	—	11	15	—
9	Number of children living	2	2	—	13	17	—
10	Citizenship	1	—	—	3	4	—
11	Nationality	1	—	—	18	19	—
12	Mother tongue	—	—	—	23	23	—
13	Religion	—	—	1	12	13	1
14	Literacy	3	3	—	19	25	—
15	Degree of education	4	12	4	72	92	12
16*a*	Active or not	1	—	—	15	16	2
b	Dependency status	5	3	1	35	44	2
c	Occupation	6	4	2	82	94	5
17	Status	12	5	1	36	54	2
18	Type of activity	15	7	5	50	77	7
19	Sector of economy	8	—	6	23	37	3

Table 69. - Number of errors committed and corrected in the phases of card punching, verification, and mechanical editing as obtained from a post hoc check carried out on certain questions based on a sample of 10,882 questionnaires selected in the Republic of Croatia from the 1953 Census of Population in Yugoslavia

Serial number	Item	Number of census errors corrected	Number of errors committed in the three phases of processing	Number of errors corrected from all the phases of processing
(1)	(2)	(3)	(4)	(5)
2	Sex	3	—	—
3*a*	Date of birth	—	—	1
b	Month of birth	2	—	2
c	Year of birth	1	2	3
4	Place of birth	—	1	—
5	Marital status	8	3	3
6	Number of marriages	3	1	1
7	Age at first marriage	7	12	3
8	Number of children born	2	—	4
9	Number of children living	2	—	2
10	Citizenship	—	—	—
11	Nationality	—	—	1
12	Mother tongue	—	1	3
13	Religion	—	1	1
14	Literacy	3	2	3
15	Degree of education	6	5	5
16*a*	Active or not	—	1	3
b	Dependency status	10	2	8
c	Occupation	33	17	16
17	Status	3	4	2
18	Type of activity	26	5	15
19	Sector of economy	6	1	6

The report says that only serious errors are taken into account, i.e., those which, for the units concerned, provoke a change of class in the various distributions adopted for tabulation purposes.

Another interesting feature of these data is the indication they give that a repetition of a phase of processing has a relatively poor effect on errors. For example, column 6 of Table 65 shows the number of errors committed in office editing, while column 4 of Table 66 shows how many of these errors were corrected in the quality check of office editing. This number is much smaller than one would expect. The same impression is obtained by comparing figures on errors committed in coding in Table 67 and figures on the correction of these errors in Table 68.

Leaving aside the question of the use of the data presented here, attention should be drawn to the usefulness of the post hoc quality checks of processing as a source of information on the efficiency of the procedures adopted, the problems they raise and the directions where further observations and studies are needed.

12.4 Process checking

As has been pointed out the aim of process checking is to keep under control a phase of data processing while the work is in progress. In this way the amount of control is reduced compared with 100-percent control and savings are achieved. Moreover, current information on the quality level of the work is available and if the evidence shows that it has deteriorated, various measures can be taken to remove the excessive errors either by rejecting the work carried out in a specified period of time or by replacing the operator.

The use of sampling methods cannot produce a sufficiently reliable picture of the quality of the work unless the operators have already reached a relatively stable and satisfactorily high level of achievement.[5] Highly variable material requires large samples and so the advantages of sampling vanish. If stable quality is achieved sampling becomes a device to reduce the amount of checking while providing current information on whether the work still retains the acceptable quality standards.

Before the sampling program of the quality control is started, it is necessary to know the quality of the operators' achievements. This information is normally obtained by carrying out a 100-percent control of the work of each operator and recording the results obtained. Such records will make it possible to detect opportune moments for starting process checks based on sampling. For the operators who did not reach the desired quality level the 100-percent control continues in order to see whether a particular operator will improve and to ensure that his errors do not penetrate into survey results. The same procedure is used for those whose work often changes in quality. Those, however, who have worked for a sufficiently long period on a satisfactory quality level can reasonably be expected to do so in the future. Their work is checked on a sample basis.

---

[5] Deming, W.E. and Geoffrey, L.: On sample inspection in the processing of census returns, *Journal of the American Statistical Association*, Vol. 36, 1941, pp. 351-360.

The ideal process check would be one which shows current changes in the quality of work and makes it possible at the same time to identify the parts of the work found unsatisfactory so that action can be taken to improve the quality of these parts. This aim could be achieved by an appropriate adaptation of the procedure known by the name of acceptance sampling. Statistical data arrive for processing in folders or lots containing the questionnaires of an ED or some other area unit. The technique of acceptance sampling then consists in checking the quality of a sample of questionnaires or some other units from each folder after a phase of processing has been carried out. If this inspection shows that the quality of the sample is satisfactory, the folder is accepted. Otherwise, it is rejected in the sense that the errors committed are removed.

To make this point clearer, let us assume that the phase of processing involved is the punching of cards. A card is called "good" if it contains no error. Otherwise it is considered "bad." An error is committed if, for a given item, the card is punched in the wrong place or is not cut at all (omissions). The quality of punching is then called satisfactory if the percentage of bad cards does not exceed some specified level, such as 1 or 2 percent. If the inspection of a sample of cards in this situation shows that the estimated percentage of bad cards is below "the tolerance limit," the whole lot of cards is accepted. Otherwise, they are rejected in the sense that all the cards in the lot have to undergo complete verification, during which all the bad ones will be corrected.

Such a procedure safeguards the quality of the work. Moreover, it furnishes current information about the performance of each individual operator: if it is found that an operator is producing a succession of rejected lots, he can be removed from the punching work. Another measure might be to call his attention to the poor quality of his work, provide additional training, etc.

A procedure of this type is hardly applicable in the processing of census data. The folders of questionnaires are normally rather small. Since the precision of an estimate depends upon the absolute size of the sample, for small folders this size would be so large that it practically amounts to complete verification. The purpose of using sampling methods is then lost. Assuming an infinite population and a check of punching which produces 2 percent of bad cards, a sample of 196 cards is needed to get the standard deviation of 1 percent. Because of the overhead costs with small folders of about 500 cards, such a sampling would be practically as costly as the complete verification.

The quality control of data processing has often been compared with industrial quality control. In fact, the analogy is not complete and the principles of acceptance sampling may not be appropriate in data processing even if they are very adequate in a comparable situation in industrial production. If a product in the latter case is found defective, either it cannot be used at all or it has to be adjusted. Errors in data processing are different. Some of them may have very little effect on survey results. If the age of a person is punched at 17 instead of 18, the error will hardly have any effect on age classifications. Tolerance limits in industrial quality control have a very definite economic meaning so that any deviations from them can be taken into account in a study of the economy of production. Because of the completely different nature of errors in data processing it is not easy to estimate the consequences of setting tolerance limits at a particular point. Without knowing what types of errors are found and in what way they would affect the final survey data, it may not be appropriate to reject a bunch of punch cards with the number of errors estimated at 3 percent instead of the expected 1 percent. A reasonable aim of process checking might therefore be to use quality control charts which secure information on quality changes in the work process but do not give sufficient guidance to judge the quality of each particular lot. Such a moderate aim can be achieved economically, i.e., by means of rather small samples. It also protects the quality because operators whose work deteriorates are detected.

The organization of process checking is similar to the quality control of industrial production. In order to identify the source of unsatisfactory quality in the whole production, individual machines are checked separately. In the same way individual operators are checked separately to facilitate the quality protective action. In both cases the tool needed is the quality control chart, where data from successive checks are introduced in order to have a synoptical picture of the quality of the work process. One of the aims of industrial quality control is often that of checking the percentage of defective products. In checking the quality of a phase of data processing the aim is to obtain information on the percentage of questionnaires or punch cards where an error was introduced during the phase being checked. Again in both cases samples of work are taken for inspection as the process goes on and the result obtained in the inspection is introduced on the quality control chart.

In order to facilitate the subsequent discussion it will be assumed that the process check refers to the verification of punch cards, which is the most frequent case of the use of sampling methods for the purpose of checking the quality of data processing. No generality is lost through this assumption because any other phase of processing might be taken instead of the verification of punch cards.

It should also be added that for quality checking purposes there is no need to take into account separately the number of errors on individual punch cards. As soon as an error is found the card is called "bad" and continues to be called so irrespective of the number of errors.

In starting the quality control charts it may be a good practice to plot on them the percentage of errors as found by means of the 100-percent verification for each day separately, starting from the first day of work. This gives sufficiently detailed insight into the quality changes which can be expected to take place as experience with the material involved increases. These data will indicate whether some operator should be replaced and used for some other work, such as verification. In the United States Bureau of the Census it was found that poor punchers could often be used as reasonably good verifiers.[6]

After this initial stage where large changes are likely, it may be desirable to pool together the work of several days or even of the whole week and introduce one single point on the chart for the whole period. This is the period of slow changes when the operator is gradually moving toward the quality level that on an average cannot be further improved.

The third stage will be the stabilization of the operator's performance. This period should be long enough to ensure with reasonable certainty that the operator has reached the stage where his work will not change unless some special factors affect it. The percentage of cards containing errors in this period is then the first element necessary to construct the quality control chart.

The quality limits are the next element needed. The lower quality limit is not necessary because the quality of processing is not endangered if the percentage of errors falls below the average. The value of the upper quality limit is equal to $p + 3\sigma_p$. If p is the percentage of bad cards, say 2 percent, the computation of the value of the upper limit is simple after the size of the sample "n" becomes known.

The size of the sample has to be decided after consideration of a number of factors, such as the phase to be checked, the quantity of the work done

---

[6] Hansen, M.H., Hurwitz, W.N. and Madow, W.G.: *Sample Survey Methods and Theory*, John Wiley, New York, 1953, Vol. I, p. 623.

in a unit of time, the feasibility of the checking procedure, the precision of the estimates, etc. If punch cards are verified by means of sampling, the number of cards resulting from an operator's work in a day will be relatively small for sampling purposes and for entering the result of each day's sampling on the control chart. In such cases it may be useful to have a sample of 50 or 100 cards selected from the work for the whole week represented by one single point on the control chart. If the check is of coding or some similar operation where a considerably larger quantity of work is produced each day, it may be possible to select the same size of sample from the process over a shorter period of time.

As regards the feasibility of sampling, one has to take into account the flow of the material. If there is little space available in which to pile up the questionnaires punched or if they have to be taken away for the continuation of punching by some other operator, it may very well be that samples have to be taken from a small quantity of the material. It may thus be necessary to sample from an average number of 500 punch cards. In such cases samples will obviously have to be small. Similarly, if the same person is responsible for checking a large number of operators, it will be impossible to work with large samples.

The size of the sample has, of course, its obvious implications on the precision of estimates. With the 2-percent average of bad punch cards and the sample size of 100 cards checked, the value of the upper 3σ control limit is $2 + 3 \times 1.4 = 6.2$ percent. A sample of 25 cards would raise it to 10.4 percent. Using the probabilities of the normal distribution, the first result means that on an average there is one chance out of 1,000 that more than six bad cards would be found in a sample of 100 cards as a result of random variations. In the case of the second sample the upper limit of the chance variations under the same conditions is 10.4 percent, which means that considerably less information is obtained regarding the quality of the work than in the above case. The question of the size of the sample is thus connected with how large a range of chance variation is allowed.

For the actual selection of the sample several procedures can be used, depending upon the circumstances. In the check reported by Deming and Geoffrey[7] card punching was organized in sections of 20 operators each. The section chief and his assistant were responsible for both punching and verification. It was found that one verifier was able to verify the work of all the punchers on a sample basis if he worked with a sample of 5-percent cards. In selecting cards for verification systematic sampling was used; cards were serially numbered. A random start was made each day so that neither the puncher nor the verifier could know what cards would be selected.

For obvious reasons, the selection of the sample has to be made in such a way that punchers do not know what cards will be checked. This can easily be achieved if the whole of the previous day is kept for checking on the next day. Any selection is then possible. If, however, the work has to be checked as it proceeds, either of the two methods referred to above may be applied or some convenient variation thereof that suits the circumstances and gives the necessary protection. For example, if 5-percent cards have to be checked, five random numbers might be selected for each operator indicating the serial number of five cards out of each hundred which will be verified. After the check of the first hundred is finished the numbers can be rotated between operators so that time is saved in the selection of random numbers.

It may very well happen that in some cases the selection of single cards will not be appropriate. It may then be useful to consider the efficiency of cluster sampling which consists in the selection of a group of cards at a time. The theory of sampling shows how to achieve the optimum size of clusters in such cases.

In recording the results of checking it may be found possible to keep records of the type of error committed in each particular case. Points introduced on the control charts show changes in the quality of work with regard to the number of errors. Some kind of form for recording what survey item each error concerns might give some guidance on the measures to be taken, such as calling the operators' attention to a particular item, giving them more instructions, etc.

The protection of the quality of the work by means of process checking is achieved in the same way as in industrial quality control. For example, if the punching of cards were kept on a quality level of 2 percent of bad cards and the sample check of 100 cards taken from a day's work of 2,000 cards yields nine bad cards, the probability is negligible that so many bad cards could be found in a sample of 100 cards as a result of chance variations. It is highly probable that some nonrandom factors appeared in the processing and caused deterioration of the work. To protect census results from an excessive number of errors the whole lot of 2,000 cards may be subject to 100-percent verification. The need for such action

---

[7] Deming, W.E. and Geoffrey, L.: On sample inspection in the processing of census returns, *Journal of the American Statistical Association*, Vol. 36, 1941, pp. 351-360.

may arise even with excellent operators if they are worried, indisposed, etc. Clearly, steps of this type increase the amount of verification but in return guarantee the quality of the work.

The savings that may be achieved with such a check will vary according to the bulk of the material to be dealt with and the quality of the work. The initial stage of complete checking will normally extend over several weeks. If the quantity of work that could be checked afterward by sampling is not large, it may well be that sample checking is not rewarding in view of the efforts and changes involved. Similarly, if operators can be expected to change the quality of their work often, it may again be better to continue with complete verification.

In considering the usefulness of sampling it is also necessary to keep in mind the fact that the resulting reduction in costs is not the same as the reduction in the volume of checking. With sample checks of punching, more time per card will be needed to make a card available for verification than with complete checking, where one card immediately follows the other. The same is true with questionnaires. On the other hand, there are some advantages to sampling in checking quality that cannot always be expressed in terms of money. For example, with sampling less personnel, space, machines, time, etc., are needed to carry out the processing.

Another advantage of sampling is its flexibility. The following example may usefully illustrate the point. Table 67 shows that in coding data from the population census a relatively large number of errors were committed with the economic characteristics of the population. For this reason the coding might be carried out by groups of two or even three persons, each person being responsible for a part of the questionnaires. The sample check of coding might then consist of, say, two different samples. The first would be composed of 5 percent of the questionnaires where coding of the economic characteristics would be checked. The second sample might consist of one out of five questionnaires in the above sample, and in it the check of the remaining characteristics would be carried out. If both results are entered on two separate charts, the procedure would secure more precise information about the quality of the coding of economic characteristics and less precise indications about other characteristics. In this way the time for checking the coding of noneconomic characteristics is reduced while the critical part of coding is kept under closer control.

It should also be added that the use of sampling methods may encounter special difficulties in some surveys. For example, in population censuses an error committed in some phase normally has as its consequence the classification of the person involved in a wrong class. Each error refers to one person only. In censuses of agriculture an error in punching the total area of a farm or its production of some item might increase the figure involved by 10, 100, 1,000 and even more if wrong columns are punched. If such an error passes undetected into the census results it may change data considerably. This is particularly so if publication areas are relatively small. It is therefore advisable to check all the cards in order to make such errors impossible.

In order to avoid the introduction of errors of this type into survey results some special measures may be necessary. A way of dealing with such errors would be to supplement the sample process check with a mechanical check of consistency carried out over all the cards. It is very likely that such a check would detect errors which substantially increase the original figure. Another possibility would be to compute averages or some proportions by publication areas and study them for their consistency. In the case of inconsistently high or low values individual cards are reviewed and in this way the errors would probably be detected.

Finally it should be added that the use of sampling methods in process checking has to be carefully prepared in order to reflect properly changes in the quality of the work that may appear over time, specific features of the material to be dealt with, etc.

12.5 An illustration

In this section a brief description will be given of the sample quality check of punching for the 1950 United States Census of Population, Housing and Agriculture.[8] The total number of cards involved was about 250 million, and the process check was applied in order to secure reasonable savings as compared with the cost of complete verification. The principles of acceptance sampling were used however to discover, within rather small sample sizes, lots of unusually poor work so that they could be sent for complete verification.

The design adopted for checking consisted of a sample of clusters of units. In checking the punch cards of the census of housing all the cards

---

[8] For more details, cf. Daly, J.F. and Gilford, L.: Sample verification and quality control methods in the 1950 Census, which was published as a case study in Hansen, M.H., Hurwitz, W.N. and Madow, W.G.: *Sample Survey Methods and Theory*, John Wiley and Sons, New York, 1953, Vol. 1.

were checked which were punched for all the units enumerated on every twentieth sheet. Thus a 5-percent sample was used with a fresh random number each day. The work was assigned to operators in bunches of four folders containing an average of 213 cards per folder. The counting of the sheets was continued from one folder to another within the assignment of four folders. The sample of clusters was found to be more efficient than the simple random sample. A study was made to examine the efficiency of a cluster of 2, 5, 10, 15 and 30 units, and it was found that the optimum was around four units.

Another difficulty in preparing an efficient design was due to the variable size of folders. This made it impossible to use equal size samples from each folder unless variable sampling fractions were used. Since this was not found practicable, folders were grouped into large unit lots. The variability of the size of lots was thus reduced with regard to the variability of folders. For purposes of acceptance sampling, the existence of folders within lots was disregarded.

For the purposes of acceptance sampling, Table 70 was used. In the stub of this table the number of cards checked is shown as it results from the above procedure for selection of the sample. The meaning of the two columns is clear from their respective titles.

As regards the construction of tables of this type a number of different sampling plans may be used depending upon various factors. According to Hamaker[9] the following are some of the factors to be considered: (*i*) cost of control; (*ii*) consequences of the acceptance of bad cards; (*iii*) consequences of rejecting a lot; (*iv*) distribution of the percentages of bad cards in lots submitted for inspection; (*v*) size of lots; (*vi*) facilities available for control; (*vii*) the number of errors that might appear on a single card and the possibility of their combining, etc. Several acceptance procedures have been published so far along with corresponding tables for the practical application of these procedures.[10] They are different in the sense that they emphasize the factors listed in varying degrees. Some of these procedures may be directly used in the quality check of data processing provided the conditions are fulfilled which were assumed in working out the procedure involved. For example, it is normally assumed that a simple random sample was selected from the lot submitted to the check, and the acceptance and rejection number is then governed by such factors as the size of the lot, the size of the sample, etc. In the above case the sample of clusters of cards was selected and the acceptance number had to be worked out to reflect this fact.[11]

TABLE 70. - ACCEPTANCE AND REJECTION NUMBERS FOR LOTS [1]

Number verified	Accept lot if the number of errors is as indicated or less	Reject lot if the number of errors is as indicated or more
0—5	*	*
6—15	1	2
16—25	2	3
26—35	3	4
36—45	4	5
46—55	5	6
56—65	6	7
66—75	6	7
76—85	7	8
86—95	8	9
96—105	9	10
106—115	10	11

* If the sample for a lot is less than 6, continue the count to the next lot and treat the combination as one work unit for purposes of verification.

[1] This table is reproduced from Daly, J.F. and Gilford, L.: Op. cit.

A by-product of the plan used was Table 71 which gives guidance for a decision at the end of each week as to whether a particular operator should be allowed to continue punching or not. The average operator was expected to punch about 4,000 cards a week. As indicated in the remark at the bottom of the table, action is not taken if the operator has punched a small number of cards.

---

[9] Hamaker, H.C.: Le contrôle qualitatif sur échantillon, *Revue de Statistique Appliquée*, Vol. 8, 1960, pp. 5-40.

[10] Dodge, H.F. and Romig, H.G.: *Sampling Inspection Tables*, John Wiley, New York, 1944; *Military Standard 105-A*, Sampling procedures and tables for inspection by attributes, U.S. Government Printing Office, Washington, 25, D.C., 1950; Freeman, H.A., Friedman, M. and Wallis, W.A.: *Sampling Inspection*, McGraw-Hill, New York, 1948; Hamaker, H.C., Chabot, J.J.M.T. and Willemze, F.G.: The practical application of sampling inspection plans and tables, *Phillips Technical Review*, Vol. 11, 1950, pp. 362-370; Provtagningstabeller för statistisk felantalls kontrol, *Kungl. Armetygförvaltningen*, Stockholm, 1959; *Vg 95083 - Statistische Qualitätskontrolle für Attributprüfung*, Bundesamt für Wehrtechnik und Beschaffung, 1958.

[11] The theory to be used in developing acceptance sampling plans is described in many books on industrial quality control. It will only be mentioned that the probability of acceptance or rejection of a lot of cards containing c or less bad cards in a sample of n cards follows from the hypergeometric distribution which is currently used if lot sizes are relatively small and sufficiently high probabilities are desired to reject lots of poor quality. With larger lots and a small percentage of bad cards, normal and Poisson distributions will be satisfactory approximations.

Table 71. - Acceptance and rejection numbers for operators[1]

Number of cards verified during week	Operator continues to punch if the number of bad cards *and* the number of omitted cards are as indicated or less		Operator is removed from punching if the number of error cards *or* the number of omitted cards is as indicated or more	
	Errors (including omissions)	Omissions	Errors (including omissions)	Omissions
0— 24	*	*	*	*
25— 49	4	1	5	2
50— 99	7	2	8	3
100—149	10	3	11	4
150—199	14	4	15	5
200—249	17	5	18	6
250—299	20	7	21	8
300—349	24	8	25	9
350—399	27	9	28	10
400—449	30	10	31	11
450—499	34	11	35	12

* If an operator has less than 25 verified cards during any one week, the decision is deferred until the following week and then the records for both weeks are accumulated.

[1] This table is reproduced from Daly, J.F. and Gilford, L.: Op. cit.

As regards the cost of a sample quality check, this is far from being 5 percent of the cost of a comparable complete verification as one might expect because of the 5-percent sample. In view of the additional work needed in sample checking it was found that here the verification cost per card was two to four times as large as in the complete checking. Adding to this the cost of the initial 100-percent checking, the cost of the checking procedure applied was between 25 and 30 percent of the cost of the alternative complete verification.

12.6 Deliberate introduction of errors

Mahalanobis[12] has pointed out one means of obtaining information and even a measure of the quality of the work in various stages of processing. The procedure consists of a deliberate introduction of errors into the material to be processed. After the processing is over, the report on the errors corrected has to include those introduced deliberately. For example, after the punching of a certain number of questionnaires is over, a sample of punch cards is selected and they are repunched in such a way that some error is introduced. The bad cards are then placed in their natural order. After the verification is over, cards with deliberate errors must be among those which are discarded in the verification. The proportion of those left undiscarded may be considered a measure of the operator's standard of work.

The usefulness of this procedure very much depends upon the field of application. Although punch cards are mentioned here to explain the nature of the procedure, there may be another and more appropriate stage of processing for the application of this technique. These are primarily fields where errors cannot be so easily detected as in the verification. This is the case in editing, checking the quality of coding, checking the transcription of data, etc. In all these fields the number of errors reported as well as the inferences made from them depend upon the conscientiousness, loyalty and interest of the staff involved. Only some kind of extra rigorous control can exercise pressure on all of them and compel them to work satisfactorily and report everything about their work that is of importance. The deliberate introduction of errors may help to achieve this result.

It should also be added that the procedure offers other interesting possibilities. In the first place it makes it possible to introduce errors of any type, some of which might be of particular importance for studying the quality of the work.

12.7 Rational planning of data processing

What phases of the work should be introduced and what amount of checking is needed in processing the data of a survey? Tables 65 through 69 show how reasonable it is to raise this question before any decision is taken about the program of data processing. In this particular case some error-correcting phases had a very negligible effect. The elimination of some phases might have contributed substantially toward reducing the total cost of the processing without greatly affecting the quality of the work. It is therefore necessary to draw up a processing program where each phase of the work is introduced and carried out as far as can be justified by the cost and effort involved and the contribution to the

---

[12] Cf. his paper: Recent experiments in statistical sampling in the Indian Statistical Institute, *Journal of the Royal Statistical Society*, Vol. 109, 1946, pp. 326-378.

improvement of the quality of the final survey results. This involves *rational planning of data processing.*

Rational planning of processing is based on a critical examination of any possible step within the complex scheme of processing in order to determine the most efficient way of achieving a tabulation program under given circumstances. In some cases it may lead to the use of sampling methods where a 100-percent operation was carried out earlier. Sometimes it may lead to the elimination of the whole operation from the program of work. In other cases it may call for a more complex procedure, such as different methods of handling different parts of the questionnaire.

Rational planning of data processing does not lose its importance even if the most modern equipment is used. Any program of automatic identification and correction of errors that is fed into an electronic computer will need some of the computer's very expensive time for its implementation. The problem therefore arises of what program should be selected, bearing in mind its effect and cost.

A considerable amount of information is needed before a program of rational data processing can be established. Firstly, information is needed about the quality of the incoming material, namely about listing, response and other errors in the data collected. It is abundantly clear that even perfect processing does not help if the survey has missed a considerable percentage of units. In addition, it is also useful to know the distribution of listing and response errors. The reasons are obvious. If the response errors have a random character and cause no biases, more attention is needed in various stages of processing to avoid errors that might lead to biases than if substantial biases are already present in data as collected in the field. In the latter case errors in the processing stage will not substantially change the already poor quality of the data.

Information about the quality of the work in various stages of processing is also of particular importance. For example, before any decision is taken about the inclusion of office editing in the processing plans it must be known what effect this can be expected to have on individual items in the survey program. This makes it possible to decide whether this phase is needed at all and, if so, whether it should concern all the items or only some of them.

Obviously, data on the cost of various operations by some kind of unit is also needed. With this information available, a rational program of data processing minimizes the processing budget for a given quality level of final survey results. Such an operation may lead to a variety of solutions. Anything between a very elaborate system with many checking operations on the one hand, and a very simple procedure reduced to basic data transformation phases with no checking at all on the other, may prove to be a rational program.

Many difficulties are encountered in any attempt at the rationalization of data processing. For example, the bias or the percentage of errors will vary from one question to another. Therefore, a system of processing that is rational for one item may not be so for another item. This problem is similar to the difficulties encountered in determining the optimum allocation of the sample in the case of many variables with different patterns of variation.

However, the most important difficulty concerns the information needed for drawing up a rational data processing program. This information extends much further than was mentioned above. For example, it may not be sufficient to know the bias in totals referring to the country as a whole. The magnitude of the bias may easily vary from one area or population segment to another, with the result that the program found to be rational on a national level may be far from satisfactory on the level of smaller areas.

Because of the amount of basic information needed, a high degree of rationalization cannot be achieved overnight in any survey. In fact, the rationalization of processing is a cumulative task; a gradual expansion of research, experience and information will make it possible to reach a satisfactory solution after a number of successive attempts.

In the initial stages of the work a realistic approach would consist in ensuring that a reasonable balance be kept between the cost of various phases of processing and their effect on the quality of data. The source of information for this might be the quality check of data. These checks offer information on the quality of incoming data for each item on the survey program. Knowledge of the errors established in the quality check can also be used to obtain an insight into the maximum effect of some phases of the processing. For this purpose data for units used in the check are processed by means of the procedures included in the tentative processing program. Afterward a study is made of the effect of processing. This means that the errors that were corrected in the course of processing are reviewed as well as those that remained undetected. On the assumption that no additional errors will be introduced in processing, such a review offers an idea of the maximum effect of processing by various items on the survey program and makes it possible to outline a plan of processing that will adequately take into account both the quality of the incoming data and the characteristics of the procedures considered.

An illustration of the guidance for rational data processing that can be obtained in this way will be taken from the work carried out by the author in connection with the 1953 Census of Population in Yugoslavia.[13] The check used to obtain information on the quality of the incoming material (listing and response errors) was made immediately after the census enumeration was over in the EDs selected for the check and while the questionnaires were still in the field. Census data used for checking the quality of the response were processed according to the program that had already been established at that time. On this basis it was possible to establish the following groups of errors:

A. errors irrespective of the magnitude or direction (leading to the total number of errors);
B. errors which no amount of editing can detect or correct;
C. errors which can be detected in the editing but not corrected unless additional information is obtained from the field or the person involved is contacted again;
D. errors which can both be detected and corrected in the course of office editing.

Errors for certain specific questions as found in the above check and classified according to these four groups are presented in Table 72.[14]

The first problem that can be clarified with the help of data in Table 72 concerns the collecting of additional information from the field with the aim of filling in the missing data or correcting inconsistent responses. This procedure improves the quality of data. On the other hand, it holds up processing for some time and is costly. Data in Table 72 give guidance as to whether the procedure should be embarked upon. The errors involved are designated by C. They only appear with the last four questions, i.e., with economic characteristics. The quality of the response to the

---

[13] Cf. Zarkovich, S.S.: *Population Census Errors* (in Serbo-Croatian), Federal Statistical Office, Belgrade, 1954.

[14] Before the data in this table are used some comments should be made as to the quality of the data themselves. The total percentage of errors on various questions is always found in row A. Distribution of these errors over types B, C, and D does not always seem accurate. For example, it is doubtful whether all the errors on "Place of birth" are of type B. Missing information or illegibile responses to this question would appear as an error of type C. It may be that such errors were classified as B. The same applies to data on other questions.

TABLE 72. - DISTRIBUTION OF TYPES OF ERRORS ON CERTAIN SPECIFIC CENSUS QUESTIONS [1]
(1953 Census of Population, Yugoslavia)

Item	Type of error	Rural stratum	Urban stratum	Total
		………. *Percentages* ……….		
1 Date of birth	A	32.9	17.4	26.5
	B	14.9	7.7	12.8
	C	—	—	—
	D	18.0	9.7	13.7
2 Place of birth	A	2.6	1.2	2.2
	B	2.6	1.2	2.2
	C	—	—	—
	D	—	—	—
3 Material status	A	2.3	4.0	2.7
	B	1.0	1.4	1.1
	C	—	—	—
	D	1.3	2.6	1.6
4 Number of marriages	A	2.8	4.2	3.2
	B	1.1	1.7	1.3
	C	—	—	—
	D	0.7	2.5	1.9
5 Age at first marriage	A	13.4	12.3	13.1
	B	13.3	12.0	12.9
	C	—	—	—
	D	0.1	0.3	0.2
6 Number of children born	A	1.6	1.5	1.6
	B	1.5	1.3	1.5
	C	—	—	—
	D	0.1	0.2	0.1
7 Number of children living	A	0.5	0.7	0.5
	B	0.3	0.5	0.3
	C	—	—	—
	D	0.2	0.2	0.2
8 Citizenship	A	0.1	0.4	0.2
	B	0.1	0.4	0.2
	C	—	—	—
	D	—	—	—
9 Literacy	A	5.2	5.9	5.4
	B	3.9	2.7	3.6
	C	—	—	—
	D	1.3	3.2	1.8
10 Degree of education	A	12.8	14.8	13.3
	B	8.7	10.5	9.2
	C	—	—	—
	D	4.1	4.3	4.1
11 Occupation	A	11.1	11.2	11.1
	B	9.4	6.9	8.7
	C	0.7	1.1	0.8
	D	1.0	3.2	1.6
12 Status	A	7.8	5.3	7.1
	B	7.1	3.7	6.1
	C	0.2	0.2	0.2
	D	0.5	1.4	0.8
13 Type of activity	A	7.0	8.4	7.4
	B	6.3	5.4	6.0
	C	0.1	0.2	0.2
	D	0.6	2.8	1.2
14 Sector of economy	A	6.8	6.7	6.8
	B	6.0	4.4	5.6
	C	0.0	0.1	0.0
	D	0.8	2.2	1.2

[1] All the percentages in this table are compiled with regard to the total number of persons

remaining questions would not be affected by sending questionnaires or their transcriptions back to the field for completion or improvement. As for the economic characteristics, it will be seen that errors of type B, i.e., those which go undetected into the census data, make up by far the largest part of all the errors. If it is also assumed that errors of type D will be corrected in editing, it can be concluded that the small percentage of errors of type C would not justify holding up the regular flow of processing and obtaining additional information from the field to correct these errors. The price to pay for a small improvement in data would be too high if all the consequences of the action were taken into account.

Table 72 also helps in judging the effect of editing. The errors that can be corrected in editing are D and C if additional information is obtained for all the results affected. Accordingly, the maximum effect of editing, as expressed in the percentage of errors corrected, is equal to C + D. This is the broad program of editing. If no action is taken to collect additional information this maximum is, of sourse, equal to D. This is the narrow program of editing.

Now, it can be seen from Table 72 that for a large part of the characteristics, C errors do not exist. Accordingly, the effect of both the broad and the narrow program of editing is identical. The difference is considerable in economic characteristics only. The effect of editing on some characteristics is either nil or next to nil. Examples are "Place of birth," "Age at first marriage" and "Citizenship." At the other extreme are the questions where the effect of editing is considerable. Examples are "Marital status," "Number of marriages," "Literacy," and "Degree of education."

Data in Table 72 therefore offer useful guidance for building up a rational program of processing. They show the direction where editing is needed. Questions where the effect of editing is negligible may simply be passed over (if editing should be considered at all) and this may lead to appreciable savings because of the simplified instructions, shorter training, and reduced operation. In large-scale surveys, such as censuses, this fact may require particular attention.

With regard to the same question it is interesting to compare data in Table 72 with data in Table 65. They are, of course, not completely comparable because they do not refer to the same area. However, column 4 of Table 65 also shows that the possibility of correcting errors by means of correspondence (additional information) is practically nonexistent except for economic characteristics.

The study reported here did not go beyond editing. However, restricted as it was, it offered important possibilities for establishing a processing program that was more rational than the one used. The latter was guided by a desire to achieve as perfect work as possible. This is why so many phases of checking were introduced. The high percentages of errors in the incoming material presented in Table 72 clearly indicate how unjustified this perfectionism was. Tables 65 through 69 also show how little the quality of data was improved by at least some costly phases. Seventeen million questionnaires passed five times through the hands of the processing staff and the cost of this operation amounted to 30 percent of the total census budget, or to 750,000 working days.[15] Even a small-scale research program like the one reported here could have led to considerable savings in terms of money and effort.

If the information obtained from the quality check were coupled with some knowledge of the efficiency and the quality of the work of the processing personnel, much more guidance would be available for designing a rational program of data processing. Information on the possible achievements of the processing staff in various circumstances was collected in the studies carried out in the Yusoslav census. Such studies cost money and effort. This is an example, however, which shows how useful they can be for programming future work.

If quality checks are not conducted immediately after the enumeration is over, and information of the type shown in Table 72 becomes available after processing has already reached an advanced stage, their results, of course, cannot be used in the above way. A satisfactory substitute might then be either a past quality check conducted in connection with the same type of survey (a past census) or a pilot survey combined with quality checking. Pilot surveys are often conducted in the course of the preparations for large-scale surveys. If they are extended so that information is also collected on the quality of data, the results of such a study will make it possible to consider the efficiency of various alternative programs of data processing.

In this connection stress must be laid on the essential role of quality checks in the rational programming of data processing. In fact, rational programming of data processing is impossible without quality checks. This is an aspect of quality checking that has been either neglected so far

---

[15] Macura, M. and Balaban, V.: Yugoslav experience in evaluation of population censuses and sampling, *Bulletin of the International Statistical Institute*, Vol. 38, Part 2, 1961, pp. 375-399.

or at most inadequately taken into account. It is clear from previous discussion that no program of data processing can be considered rational unless it is built up in such a way that the quality of the incoming material is adequately reflected in it. Otherwise, poor processing may spoil the high quality of the incoming data and vice versa. The results of quality checks prevent any erroneous decision in this direction. Furthermore, if quality checks are extended over various phases of processing they offer additional possibilities for the rationalization of processing as the known error-correcting effect of various phases makes it possible to decide what phases should be included in the program, in what cases sampling could be advantageously used instead of the 100-percent operation, etc.

The importance of quality checking extends much further. The following is an illustration. With the increasing use of electronic computers in data processing, interest in the mechanical editing of data is rapidly growing. In this respect a number of techniques have been developed for the identification of errors and their correction, as well as for the estimation of missing data. How efficient are these techniques and what instruction is to be put into the computer? These questions cannot be satisfactorily answered unless each procedure is considered in the light of its error-correcting effect. This effect, however, cannot be established without relating the data produced by the machine, as a result of the instructions given, to the known errors in the corresponding characteristics of the units involved. Such a comparison shows what percentage of errors was identified and which of the errors detected the machine was able to correct satisfactorily. Without such a comparison various procedures are more or less blindly applied; their inclusion in the program of mechanical editing may result in the same type of inefficient processing as was the case with various phases in the example discussed earlier. The decisive characteristic of each technique or each procedure that is considered for inclusion in the program of mechanical editing is its error-correcting effect.

13. ERRORS AND BIASES IN YIELD STATISTICS

13.1 Introduction

Yield statistics are one of the most important elements of agricultural statistics. Data on yields are used for many purposes, and most attempts at improving agricultural production are aimed at yields. Without adequate data on yields a judgment is hardly possible on the efficiency and usefulness of the measures taken. Coupled with data on areas, information on yields appears to have primary importance in economic planning. The rate of yield multiplied by the area harvested of a given crop represents the production of that particular crop, and production figures represent a basis for the preparation and formulation of many economic measures.

The traditional way of collecting yield statistics is through some kind of reporting service, consisting either of extension work agents or the cultivators themselves. In many countries, however, such an approach is impossible for many reasons. For example, extension work machinery may not be established over the country as a whole; the extension workers may not be interested in this type of work; they may not be experienced enough; cultivators may not be able to evaluate the yields of various crops in any meaningful units, etc. For all these and similar reasons experienced statisticians in the field of yield statistics are growing steadily more interested in the use of objective methods. By means of a sampling procedure small plots of a crop are selected, the crop on these plots is then harvested, threshed, weighed, and the yield is then estimated from data collected. In other words, data are obtained independently of various subjective human factors involved in the alternative eye estimates.[1]

In many countries which do not yet have the facilities needed for establishing reporting services, the use of objective methods, in one form or another, seems to be the only safe way of rapidly establishing a satisfactory system of yield statistics. It should be added, however, that interest in

---

[1] The objective procedure for estimating the rate of yield is dealt with in *Estimation of Crop Yields*, by V.G. Panse, FAO, Rome, 1954.