Designing Survey Data Base

DEN	ResSe	ResAg	State	Count	Head_H	Res_Sta
1	M	38	1	1	1	1
2	M	29	1	2	1	1
3	M	46	1	3	1	1
4	M		2	4	1	1
5	M	83	2	5	1	1
6	M	28	2	6	1	3
7	M	36	3	7	2	2
8	M	44	3	8	1	4
9	M	40	3	9	1	1
10	M	43	3	9	2	4
11	M	35	4	10	1	1
12	M	47	4	11	1	2
13	M	53	4	12	1	2

Objectives

1. Be able to define and design database in Microsoft excel

1. Manage a data base - develop data entry sheet, code, and clean data in Microsoft excel.

Data-base!! What is it?

- A database is a collection of **information** that is organized so that it can easily be accessed, managed, updated and analyzed.

- Databases can be classified according to types of content: bibliographic, full-text, numeric, and images.

Text

Numeric

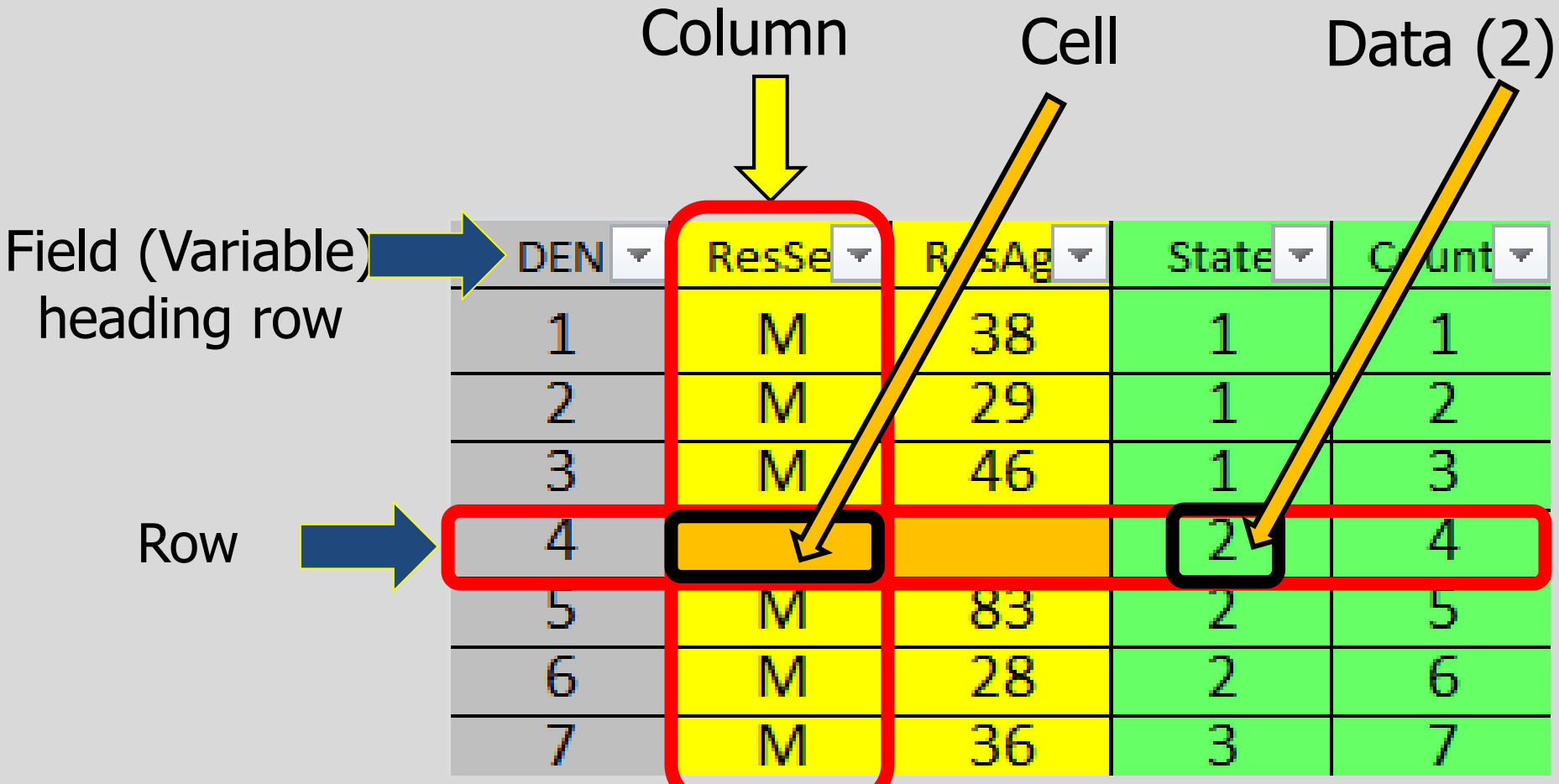
Resp.Name	District	Age	Gender.Res
Andezu Monica	Arua	28	0
Asibazuyo Christe	Arua	20	0
Bezu Gloria	Arua	21	0
Driciru Magret	Arua	41	0
Agotre Onesmas	Arua	65	1

Data-base

- A well designed data base takes care of all types variables collected.
- Should be flexible enough to take care of any introduced or derived variable before and/or during data entry
- There are a number of computer-based programs for handling large volume of data such as those from surveys.
  - Example of computer-based program include Microsoft access and Excel.

Designing data-base using Excel

- Microsoft Excel (Workbook) file consists of sheet(s), and every sheet contains Columns and Rows. The intersections between the column and row are the Cells. Cells are points within a sheet where data are entered.



Guideline in designing data-base

1) **Table:** The basic format for storing data in an Excel database is a table which is made up of rows and Column.

2) **Rows/records:** In a table, data is entered in rows. Each row is known as a RECORD.
   - ✓ A record contain data about only one unit of a sample e.g. a household data.
   - ✓ Therefore, all data in a questionnaire is entered into single row normally corresponding household **data entry number (DEN)**.
     i. A single questionnaire should never be assigned more than one data entry number.
     ii. No two or more questionnaires should have same number assigned to them.

Guideline in designing data-base

2) **Rows/records:** In a table, data is entered in rows. Each row is known as a RECORD.

- ✓ The first row of the list MUST contain a unique name at the top of each column. This first row does not need to be the first row of the worksheet

- ✓ The row containing the column headings must be formatted differently from the rest of the list (i.e. bold, larger font, italicized, etc.)\

- ✓ There should be no blank rows/record in the list (there can be blank cells in a column. The entire row cannot be empty)

General Rules in Designing data-base

3) **Column/fields:** Each column needs a heading to identify the data it contains. These headings are called FIELDS.

- ✓ Variables are normally coded as columns headings . A column is normally used for only one variable e.g.. Sex, Age, Income etc.
- ✓ Data in a column must be in the same format for every row in the column (i.e. numbers can't be spelled out in one row and entered as digits in the rest of the rows in that column).
- ✓ Leave no blank in the field heading

4) **Cells.** One cell for one response.

Defining field (Variable) names

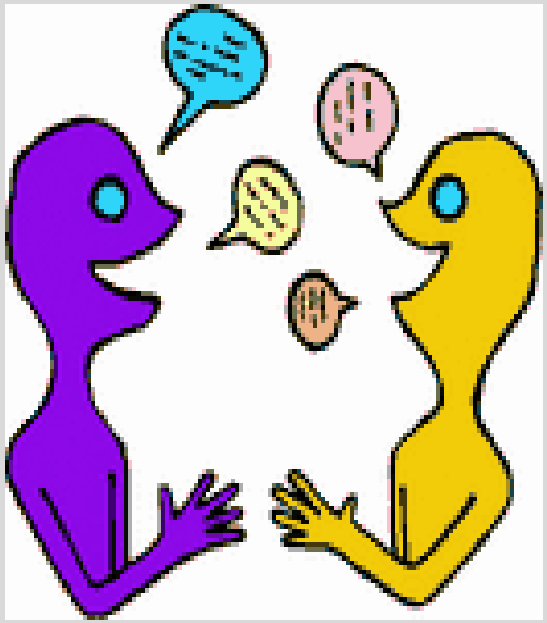
- When defining variable heading, better to use continuous character i.e no space between character
  - **Disirable: Sex or SexRes**
  - **Not desirable: Sex Res.**
- For multiple response questions – e.g. What crops did you plant last season?

Sorghum	**Maize**	**Barley**
Finger millet	**Rice**	**.....**
Groundnut	**Sesame**	**.....**

- Here, each possible response **(e.g. Sorghum)** is a variable which assumes Yes or No value.
  - **Note:** Shorten variable character so that all are visible within the cell. You can use COMMENT to fully define the variable heading for other users

Variables, coding & standardization

- **Discussion:** What is a variable?



- **Types of variables**
  - **Discontinuous** (catergorical, classificatory, descrete) variables: Cannot be divided into fractions or take finite number, e.g. gender (male or female), livestock presence (Yes or No); residential status (Resident, IDP, Refugee, Returnee).

Variables, coding & standardization

- ***Continuous variables:*** Are those that can be divided into fractions or take infinite number of values e.g.
  - Income, temperature, age, area planted, seed quantities, yields.

- Units for measuring continuous variable may vary from individual to individual or from place to place. This must be standardized before or during data analysis.

- **Coding Data.**
- For purpose of easy data entry, discontinuous data are normally given codes such as Yes=1 and No=0 etc.

- Open-ended questions can equally be codded to ease analysis.

- **Qn. How do we treat single and multiple response question?**

Data cell validation

- **Data cell validation** is used to make sure that users (data clerks) enter only certain values into a cell, hence minimizing errors during data entry.

- Number Validation

1. Select a cell
2. On the Data tab, click Data Validation.

On the Settings tab:

3. In the Allow list, click Whole number.
4. In the Data list, click between.
5. Enter the Minimum and Maximum values.

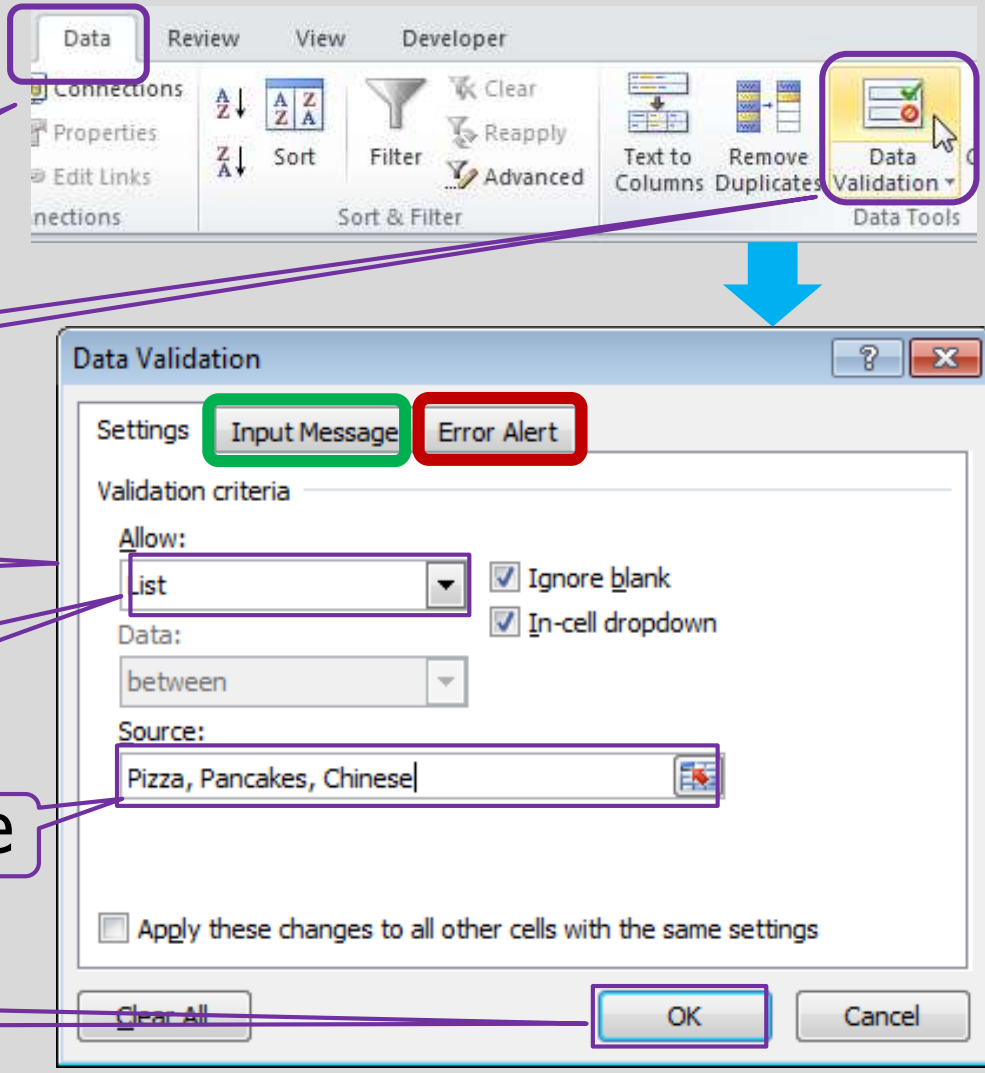
Data cell validation

Drop down list:

1. Select a cell
2. On the Data tab, click Data Validation.

- The 'Data Validation' dialog box appears.

3. In the Allow box, click List.
4. In the Source box, Type the lists separated by a Coma, then OK



Data cell validation

- **Comment:** Comments are used to help others understand your codes and data.
- To insert a comment, execute the following steps.

1) Select a cell.
2) Right click, and then click Insert Comment.
3) Type your comment.

- Excel displays a red triangle in the upper-right corner of the cell.

4) Click outside the comment box.
5) Hover over the cell to view the comment.

- Excel automatically adds your user name. To change this name, execute the following steps.

6) On the File tab, click Excel Options and choose the General category.
7) Change the User name.

Exercise 2.1: Database Design (1 hours 30 minutes)

- You are provided with a hard copy of SSA household questionnaire for West Nile (Uganda). Using your computer please design a data base following the instruction given bellow.
  a) Open a workbook and save the file name as SSA_WN_Uganda-2015_Your Name Initial
  b) On sheet 1 of the workbook, develop the field names for all the variables for section 1-5 and provide comment on the first four field names.
  c) Enter the data entry number (DEN) 1 to 2567 and insert a row between entry No. 1657 and 1658.
  d) Validate the cells of the first two categorical variables (field) with text and the next two with numbers
  e) On three additional sheets, develop a data base for crop A, B and C.
  f) Name the sheet where the data is being entered and protect the sheet.