Modeling Food Insecurity in Bivariate and Regression Analyses

Prepared by the Voices of the Hungry Team

Purpose: These guidelines, prepared by the Voices of the Hungry (VoH) team, are intended to help in the preparation of proposals to the FAO VoH project for access to the Gallup World Poll (GWP) data.

Exception: These guidelines do not apply to proposals in which the primary focus is measurement of food insecurity rather than substantive analysis of food insecurity.

Clarification—individual vs. country-level analyses: These guidelines are specifically about individual-level analyses and measures. For country-level (i.e., ecological) analyses, national prevalence rates to be published by the VoH project (July, 2015) could be used.

I. Character of food-insecurity measurement data elements

Several food-insecurity measurement data elements will be included in the micro-data (individual-level data) provided by VoH. The data elements are all based on responses to the 8 questions in the Food Insecurity Experience Scale (FIES). *(Responses to each of the eight individual questions in the FIES are also in the micro-data, but unless there is a very strong reason to use responses to individual questions, such use is strongly discouraged. The measure based on combining responses to all questions has less measurement error, is more robust, and can be made comparable across countries.)*

The following respondent-level data elements will be included:

- **Respondent severity parameter (on the VoH global reference metric).** This is a linearized, interval-level transformation of the raw score[1]. (Within a country, it is the same for all cases with the same raw score.) It can be used in analyses that require

---

[1] The raw score is the number of affirmative responses, 0 to 8. Raw score is ordinal (probabilistically) with respect to the latent trait of food insecurity. As such, it can be used to classify households within a country as to food insecurity status. (Such classification may not represent equal levels of severity between countries, however.) Raw score is NOT an interval measure, so it should not be used to calculate a mean or in linear regression models.

interval-level measurement except that the value for raw score zero may not be provided, and, even if provided, may not be correct in countries with a large proportion of cases in raw score zero. A valid value of this variable will be provided for respondents with raw score 8[2] based on a standard VoH protocol. In most cases, this value is a reasonable approximation, but in countries with large proportions of cases in raw score 8, errors in these approximations may cause modest distortions in the values of the parameter estimate for raw score 8 and, therefore, in the probabilities of food insecurity and severe food insecurity for raw score 8 (described next). The respondent severity parameter is adjusted to the VoH global standard metric, so that it is intended to represent the same level of severity in all countries. This data element will be missing for cases with missing responses to any of the FIES questions.

- **Probability of moderate or severe food insecurity** (0 to near 1). This functional form of the food insecurity measure takes into account measurement error. It can be thought of as the proportion of people in the population represented by the sampled person whose true food insecurity exceeds a set threshold. The threshold is set at the severity level of the item, "In the past 12 months did you eat less than you thought you should...." on the VoH global standard scale. The value is based on adjustment of respondent severity parameters to the VoH global standard, so is intended to be comparable across countries. The probability of moderate or severe food insecurity for cases with raw score zero will be presented as zero. Values for raw score 8 will be based on VoH standard methods (used to calculate national prevalence rates). This data element will be missing for cases with missing responses to any of the FIES questions.

- **Probability of severe food insecurity** (0 to near 1). This is similar to the probability of moderate or severe food insecurity except that the threshold is more severe—at the severity level of "In the past 12 months did you go a whole day without eating...." on the VoH global reference scale. This data element will be missing for cases with missing responses to any of the FIES questions.

For a detailed description of the use of individual probabilities of food insecurity in regression, see Box 1 in Section III.

II. Representing food insecurity as an independent (right-side-of model) variable

---

[2] The conditional maximum likelihood methods used by VoH to estimate the measurement model cannot provide statistically rigorous severity parameters for respondents who said "no" to all items (raw score 0) or "yes" to all items (raw score 8). However, by imposing some additional assumptions, the VoH protocol provides an approximate respondent severity parameter for raw score 8.

For within-country analysis, or for multi-country analysis if a set of dummy variables is included to identify each country, severity of food insecurity may best be represented by a set of dummy variables based on raw score. This does not impose a functional form on the relationship between severity of food insecurity and the dependent variable, and can, therefore, reveal a non-linear relationship. A full set of 8 dummy variables (with raw-score 0 as reference) can be included, or raw scores may be grouped to reduce the number of categories. (A suggested set is raw score 1 to 3, 4 to 6, and 7 to 8, with raw score 0 as reference. These correspond approximately to some-indication-of food-insecurity/moderate-food-insecurity/severe-food-insecurity.)

For multi-country analysis, respondent severity parameter, probability of moderate-or-severe food insecurity, or probability of severe food insecurity may be used. All three are (putatively) comparable across countries, thus not requiring control for country fixed effects. (Or, if country fixed effects are included, it may be assumed that their coefficients represent primarily factors other than artifacts due to non-equivalent measurement of food insecurity.) The severity parameter is appropriate for use in such models if it is assumed that the association of interest is approximately linear across the range of severity. Raw score zero may be omitted, or, if included, it may be useful to include a dummy variable for raw score zero along with the severity parameter, since the correct value of the parameter for raw score zero is not known with confidence. One or the other of the "probability" measures is appropriate if it is thought that the association of interest is particularly strong at the level of severity represented by that probability. The probability measures largely avoid the problem of the uncertainty of the severity associated with raw score zero, because the thresholds are severe enough that the probability of true severity exceeding the threshold for individuals who said "no" to all 8 items is effectively zero. The "probability" measures have the advantage of rather simple and intuitive communication of results since the coefficient in such an analysis represent the difference between food insecure and food secure individuals (or between severely food insecure individuals and all others).

III. Representing food insecurity as a dependent (left-side-of model) variable

For within-country analysis, or for multi-country analysis with controls for country fixed effects, food insecurity may be represented as a binary variable based on a raw score cutoff. Such models may be estimated using logistic regression, probit regression, or linear probability (OLS with a 0/1 dependent variable) models, depending on constraining factors. It is often helpful to estimate three separate models using different raw-score cutoffs in order to understand what level of severity of food insecurity is most strongly associated with the right-side variables.

Suggested thresholds are:
- Raw score 1 or greater versus raw score 0 (any indication of food insecurity)
- Raw score 4 or greater versus raw score 0-3 (On average, this corresponds approximately to moderately-or-severely food insecure versus secure-or-only-mildly insecure.)
- Raw score 7 or 8 versus raw score 0-6 (On average, this corresponds approximately to severely food insecure versus all others.)

For multi-country analysis, probability of moderate-or-severe food insecurity, or probability of severe food insecurity may be used, as described in the preceding section. These are (putatively) comparable across countries, thus not requiring control for country fixed effects. (Or, if country fixed effects are included, it may be assumed that their coefficients represent primarily factors other than artifacts due to non-equivalent measurement of food insecurity). The severity parameter is not generally appropriate for use as a dependent variable unless cases with raw score zero are omitted, since the correct value of the parameter for raw score zero is not known with confidence.

Box 1. Use of individual probabilities of being food insecure in regression models

There are some characteristics related to probabilities of being food insecure at individual level that must be taken into account, especially when one would like to use them as *dependent* variables in a regression. These attributes have been mentioned along the document, but deserve careful attention, so they will be listed in this box:

1. There is a one-to-one correspondence between individual probabilities of food insecurity and the raw score each subject reports. This means that probabilities are clustered around each individual raw score, and there will be only 9 (from 0 to 8) distinct values for them. In this regard, it is extremely important that probabilities are not used to assign individuals to food security classes, but rather used (as dependent variables) to estimate and interpret regression coefficients for the included independent variables.
2. All subjects with raw score 0 have a probability of being food insecure (at any level) equal to 0. This truncation must be taken into account in the model specification.
3. Distribution of probabilities across subjects is usually asymmetric.

All these characteristics suggest that probabilities of being food insecure should be included as dependent variables in a proper model formulation that, for example, would be along the lines of Tobit models, or at least should indicate to avoid simple linear regression models.