Unit 8 Problem Set

This evaluation uses a dataset called **EG2004**. You are going to work on a database containing information on 300 children <5 y old from a national survey conducted in Equatorial Guinea in 2004. The goal of this exercise is to find out if child stunting is associated with maternal education. If you need to read more about growth indicators, I strongly encourage you to visit the WHO website: <http://www.who.int/nutgrowthdb/about/introduction/en/index2.html>

<http://www.who.int/nutgrowthdb/about/introduction/en/index5.html>

The 300 children were selected using a complex design in which four strata were considered: island-urban, island-rural, continent-urban, and continent-rural. Within strata different villages (clusters) were selected (this is the primary sampling unit (PSU)). Finally, within clusters (villages), some households were selected, and within households, all children <5 y old were selected. You have two variables in the dataset indicating the four different strata (IslaCont2=’Island/continent’ strata, and RUDISE\_O2=’rural/urban strata’) that you will need to include in your strata statements (i.e. strata IslaCont2 RUDISE\_O2;). You also have one variable for the primary sampling unit (n\_congd2). Finally, you have one variable indicating the sampling weights (ponderacion). Please, note that sampling weights do not have to be the same for each PSU. Please, **note that this is NOT the same study that we used in the videos or in the practice, and this one includes stratification on top of clustering.**

For the purposes of this exercise, assume that all values are correct (i.e. there are no errors in the dataset, although there might still be outliers). Follow the instructions and **copy and paste the relevant SAS code and output.**

**Question 1:** Create binary outcomes and report prevalence of the outcomes

**Instructions:**

a) Create binary variables for each continuous z-score (HAZ, WAZ, and WHZ) using the following cutoffs:

|  |  |  |
| --- | --- | --- |
| **Indicator** | **Cutoff** | **Moderate Malnutrition** |
| Height-for-age z-score (HAZ) | <-2 | Stunting |
| Weight-for-age z-score (WAZ) | <-2 | Underweight |
| Weight-for-height z-score (WHZ) | <-2 | Wasting |

b) Report the prevalence of stunting, underweight, and wasting with their corresponding 95% CI assuming this data comes from a complex sampling design. Please, round off to one or two decimal places at most.

c) Report the prevalence of stunting, underweight, and wasting with their corresponding 95% CI, that you would have obtained assuming that these data comes from a simple random sample where all children have the same probability of being selected. Please, round off to one or two decimal places at most.

d) Are there any differences in the prevalence estimates using the different assumptions in b and c? Independently of finding differences or not what would you expect and why?

e) Are there any differences in the 95% CI using the different assumptions in b and c? Please state why or why not differences may exist.

**Question 2:** Bivariate analysis

**Instructions:**

a) Create a descriptive Table 1 with the stunting variable as the outcome (stunted=yes vs. stunted=no) including the variables listed below. Fill in the descriptive Table 1 reporting means and standard errors for continuous variables for stunted and non-stunted groups. Fill in the descriptive Table 1 reporting percent for categorical variables for stunted and non-stunted groups. **Unless otherwise specified binary variables are coded as 1=yes, 0=no. Ignore the fact that some variables have missing values and work with the data you have available.**

Mother's age

Mother BMI

Mother had elementary school

Higher SES

No toilet in the house (**0=there is toilet, 1=there is no toilet**)

Child age in months

Sex of the child (**0=girls, 1=boys**)

Months breastfeeding

Child was bottle fed

Child received vitamin A supplements

Child had diarrhea last month

Child had fever last month

Child had malaria last month

Child had intestinal parasites last month

b) Add a column to your table 1 with p-values and a footnote explaining which statistical test you used for each variable. Assume normality for continuous variables (no need to check)

c) Based on your bivariate analysis (Table 1), is mother’s education associated with stunting? Explain.

**Question 3:** Multivariate analysis

**Instructions:**

In the real world, you should assess for confounding by yourself, but to make your life easier and to make sure that we all work with the same models, we are going to tell you which variables to include in your models. Run two logistic regression models with stunting as the outcome variable and education of the mother as the exposure variable. The first model should be a crude model, and the second model should be an adjusted model, adjusting for age of the child in months, sex, and having a higher SES. (Note: it could be argued that SES is an intermediate variable but given that this is a cross-sectional study let’s assume it is a confounder. Age and sex of the child are not really confounders since they do not affect education of the mother but for the purposes of this exercise we will include them in the model anyway).

a) Create a table to report odds ratios and 95% confidence intervals for the two models described above. **Do not use weights for these models.**

b) Is there clear evidence of confounding by the variables you adjusted for? Explain

c) Estimate and interpret the OR and 95% CI for age of the child in the fully adjusted model NOT for a 1 month change in age BUT for 10 months change in age (you can use a contrast statement or you can do it manually if you prefer. If you do it manually, please show your intermediate calculations. If you use SAS, please, paste your code below)

Extra Credit Challenge Task

If you completed this optional task, please add it here.