

CDIMA December 2021 Stata Test

Part 1:

Data for part 1:

Main dataset, New variables, New observations, and Missing value codes.

Instructions

1. Importing

- a) You have obtained data from a partner organization with extra variables that were not asked in the survey, and so we are interested in adding this to our dataset. However, the file that the organization has sent to you, *New Variables.csv*, is not in .dta format. Import this dataset into Stata.
- b) Our main dataset is called *Main Dataset.dta* – open it and then merge in the *New Variables* dataset to it.
- c) New people have been surveyed today, and their survey responses have been sent to you in the dataset *New Observations.dta*. Add these observations to our main dataset.

2. Cleaning – After collecting all of our data, we want to clean it up so that it is ready to be used for analysis.

- a) Remove personally identifiable information (names of surveyor). Replace these values with numbers.
- b) Refer to the table “Missing Value Codes” and replace missing values (I.e., -997, -999, etc.) with their corresponding extended missing value for the variables *vanalismyn* and *trespassingyn*.

Part 2:

Data for part 2

When working at IFC, you will be requested to manipulate and work with household surveys data ([Guidance on HH surveys](#)). We have provided you with 3 datasets that come from the 2007 round of an Indonesian household level health and education survey. We are including a subset of the questions asked in the survey, structured into 3 datasets: “*household*” contains general information about the household’s characteristics and the date of the survey; “*children*” has data on children born to mothers in the household (not including those that are deceased); and “*deceased*” contains information on the children (born to mothers in the household) that have died.

Instructions

This part of the test is open ended, with the purpose of getting an idea of how you think about data analysis and not necessarily how much knowledge you have on sophisticated econometric and statistical models. The purpose of this exercise is not to end up with a perfect and clean dataset.

We would like to see how you work with the data and use it to tell a story about infant mortality with the resources we provide to you. Please use Stata to complete this exercise.

Data Analysis

A) Use the provided data to describe the households with children that have died (within 28 days or 1 year of birth) in any way you find interesting. We are looking for no more than a simple table, graph or other visualization of your preference. Please write a short paragraph with your results.

B) Use any tools you think are useful (graph, table, tests, text, etc.) to tell an interesting story about infant mortality in this set of households. There is no right answer to this task, we are more interested in how you think about this process than in the story itself. Please be brief.

**Final note: if any point of the instructions is not clear for you, please state that in your answer and provide us with your assumptions to solve the specific task.*