Exercise 8: Advanced Data Management – part 2

Exercise 7 and Exercise 8 cover the materials in Chapter 8 of the module notes.

Launch Stata, open a new do-file and save as *Stata_Exercise8.do*. Add appropriate comments at the beginning of the do-file. Add commands to change the current working directory to the Exercise 7&8. Remember to keep saving the do-file as you go along.

8.1 Repeated Measures Data

Open *fup_sbp.dta*.

- Browse the data to make sure you understand the structure. For the next set of tasks you will need to use `bysort`, the system variables `_n` and `_N`, and subscripts.
- Produce a table showing the distribution of the total number of visits for all the patients, i.e. how many patients had 1 visit, 2 visits, etc.
- Produce a table showing the final visit number for each patient.
- Create a variable containing the change in *sbp* from baseline at each visit for each patient.
- Create a table showing the mean change in *sbp* by visit and treatment group.
- Create a binary variable that takes the value 1 if the patient's *sbp* was ever greater than 140mmHg during the study period.
- Save as *fup_sbp1.dta*.

8.2 Creating Summary Datasets

Open *fup_pot_long1*.

- Browse the data to make sure you understand the structure.
- Create a summary dataset containing the mean and standard error of the mean for *potval* by visit and treatment group.
- Browse the new summary dataset – check that the collapse has worked as required.
- Save as *fup_meanpot.dta*.

8.3 Reshaping Data

Open *fup_hrate.dta*.

- Browse the dataset and make sure you understand the structure of the dataset.
- Could you obtain a pairwise correlation matrix of the *hrate* measurements at each visit.
- Reshape the dataset from long to wide format.
- Look at the output in the Results window. How has the dataset changed?
- Browse the data making sure you understand the new structure of the dataset.
- Obtain a pairwise correlation matrix of the *hrate* measurements at each visit.
- Save as *fup_hrate_wide.dta*.

Open *fup_anthrop.dta*.

- Browse the dataset and make sure you understand the structure of the dataset.
- Obtain a pairwise correlation matrix for *wt* at each visit and then for *wc*. Are there any issues with the data?
- How would you go about generating a change from baseline variable with the data in this format?
- Reshape the dataset from wide to long format.
- Look at the output in the Results Window. How has the dataset changed?
- Browse the data making sure you understand the new structure of the dataset.
- How would you go about generating a change from baseline variable (e.g. change in weight) with the data in this format?
- Save as *fup_anthrop_long.dta*.