

EPIB 613

October 7, 2011

Exercise #6 (chapter 11 of Juul & Frydenbyrg)

The following two datasets contain ankle blood pressure measurements for 2 time periods from the same group of patients. The variables in each dataset are as follows:

ankle-bp-1.dta: patient id, ankle BP measure 1

ankle-bp-2.dta: patient id, ankle BP measure 2

1. Merge the two datasets into one dataset that contains three columns: one column with the patient id and one for each of the BP measures.

(hint: the BP measures are named abp in both datasets, so they need to be renamed in order to be recognized as the two different measures once the datasets are merged).

List the first 5 observations. Drop the variable `_merge` and save the merged dataset as `anklebp-wide.dta`

2. The dataset `anklebp-wide.dta` is in wide format, containing a separate column for each blood pressure measurement:

| id | abp1 | abp2 |
|----|------|------|
| 1 | 105 | 110 |
| 2 | 110 | 110 |
| 4 | 140 | 130 |
| 5 | 90 | 85 |
| 8 | 130 | 130 |

Using `reshape`, set the data into long format like this:

| id | measure | adp |
|----|---------|-----|
| 1 | 1 | 105 |
| 1 | 2 | 110 |
| 2 | 1 | 110 |
| 2 | 2 | 110 |
| 4 | 1 | 140 |
| 4 | 2 | 130 |

List the first 6 observations. Save the dataset as `anklebp-long.dta`.

3. Convert the current dataset back into wide format so you have one row for each patient id. List the first 5 observations.

EPIB 613

October 7, 2011

Exercise #6 (chapter 11 of Juul & Frydenbyrg)

Using the dataset `nlsw88.dta` (a system dataset so use `sysuse`):

4. Is the proportion of married people who are college graduates greater in the South or outside the South? Find a 95% confidence interval for the difference between the two proportions.

5. Generate a histogram of total work experience, adding a kernel density estimate to the plot. Make the y-axis the frequency of individuals for each bar rather than the density.

6. Simultaneously generate two histograms for total work experience, one for college graduates and one for non-graduates. Present the two graphs in one column (i.e., one below the other).

7. Simultaneously generate boxplots for hourly wage across quintiles of total work experience (you need to generate a variable containing the quintiles). Add the following title to the plots: "Hourly wage by quintiles of total work experience".