# Review questions and exercises

## Topic 2

### Before you start

Make sure you are working in the correct directory: `pwd`. Change to another directory if necessary. Also, you may wish to open a command log and a result log:

```
cmdlog using logname.do
log using logname.log
```

### Exploratory Data Analysis

Use the nlsw88 data. It comes shipped with Stata; type `sysuse nlsw88, clear`. The data can also be found in excel format on ELE.

The data set contains labour market information on American women in their 30s and 40s. Your task is to explore the distribution of the hourly wage (variable *wage*, measured in USD).

1. `browse` the data. Make sure you understand which variables are categorical and which are continuous, and what information each variables contains.[1]

[1] "Current grade completed" refers to the years of education that the individual has completed. "SMSA" stands for *Standard Metropolitan Statistical Area*; see `https://en.wikipedia.org/wiki/Metropolitan_statistical_area`.

2. Find the mean, median and standard deviation of *wage*. Is the distribution skewed? What are the smallest and largest values?

3. Select one categorical variable which interests you in the context of the labour market. Find the average hourly wage and the median hourly wage for each of the categories. (Hint: take a look at the `tabstat` command.)

4. Are hourly wage, experience and tenure correlated? Are the results what you expect them to be?

5. Produce two distributional graphs which you find interesting and informative. Comment on your graphs, and why you have chosen them.

6. Do people who work more hours per week have higher hourly wages, on average? Produce a graph to illustrate.

7. The term "operative" in American English means skilled factory worker. Looking at the occupation type "Operative", what is the difference in average hourly wage of union members versus non-union members? Repeat for the occupation type "Transport". Comment on your findings.

8. Perform a test of the null hypothesis that wage is normally distributed.

9. Perform a test of the null hypothesis that the natural logarithm of wage is normally distributed.