

# Review questions and exercises

## Topic 1

### Before you start

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

```
cmdlog using logname.do
```

```
log using logname.log
```

### Basic commands and syntax

Use the auto data: `sysuse auto, clear`

1. Sort the data in order of price. What is the mileage per gallon of the cheapest vehicle? What about the most expensive car?
2. Find out how many foreign vehicles are in the data set.
3. Sort the data in order of miles per gallon. Is the current sort order unique?
4. Generate a new variable called *domestic* which is equal to one if the vehicle is domestic and zero if the vehicle is foreign.
5. Sort the data in order of miles per gallon and price. Generate a new variable called *obsno10* holding the observation number + 10.

### Variable labels

`clear` the memory and use the `pen01aL3` data. Generate the **nominal hourly wage** from the log wage (see the lecture notes, slide 45). Observe that the new variable is different from the others in that it has no label attached (see the Variables

window, or `describe`). Create a label for your variable using `label variable` or the properties window. Help is available at `help label`.

## Value labels

Generate another variable which equals 1 if a person holds a university degree, and 0 otherwise. For this purpose, we define that everyone with 17 or more years of education holds a degree. `tabulate` your new variable. Use `graph hbox` to create a box plot of the hourly wage in the three categories “degree”, “no degree” and “total”.

Looking at this graph, note that your degree variable is labeled 0 and 1. To give more descriptive labels, create a value label which assigns “Degree” to the value 1 and “No degree” to the value 0. Use the properties window to do this. [Hint: highlight your new variable in the variables window. In the properties window, click ‘Lock/Unlock’. Next to the Value Label field (which is empty), click on the three dots and then ‘Create Label’.]

Don’t forget to assign the new value label to the degree variable. (Select your new value label from the drop-down menu.) Repeat the `graph hbox` command and compare to the previous version.

## Stata functions

Stata has numerous built-in functions for manipulating numbers and strings; see `help functions`. Explore some of them, either by using `count`, `display`, or by generating new variables. You can use the expression builder (accessible via the menu for `generate`) to explore.

Find out the following:

1. What is the standard normal density at 0.5?
2. What is the probability that a  $\chi^2$  distributed random variable with 2 degrees of freedom is greater than 5?
3. How to standardise<sup>1</sup> a variable.
4. How to generate a variable holding the sign of another variable.

---

<sup>1</sup>Standardisation refers a transformation of a variable such that it has a mean of zero and a standard deviation of 1.

### Precision puzzle

Try the following in Stata:

```
clear
```

```
set obs 10
```

(This creates an empty data set with 10 observations.)

```
generate x1 = 2
```

```
generate x2 = 0.2
```

```
count if x1==2
```

```
count if x2==0.2
```

Are the results as you expected? (Hint: there is an FAQ entry on this subject; type `search compare precision` in Stata to find it.) How would the second command have to be modified such that `count if x2==0.2` returns the correct result?