

ECN 102: Analysis of Economics Data

Homework 1

Remy Beauregard

Submission rules: homework must be submitted as *one pdf*; please do not submit multiple files. All submissions should include all Stata code and output produced, any figures generated, and answers to questions. For written questions, work must be shown for answers. Code need not be commented.

All datasets can be downloaded in Stata format from the [AED data page](#).

Question 1: Data types

Give an example of the following types of data:

- a) Observational, categorical, time-series data
- b) Observational, numerical, panel data
- c) Experimental, discrete numerical, cross-sectional data
- d) Experimental, categorical, repeated cross-sectional data

What type of data are the following:

- e) Midterm 1 scores for one class of ECN 102
- f) Final grades (%) for ECN 102 from the last 5 years
- g) All homework grades (graded pass/fail) for one student in ECN 102 this quarter
- h) All homework grades (graded pass/fail) for all students in ECN 102 this quarter

Question 2: Univariate data

For the given sample 0, 4, 5, 2, 3, 2, 11, 17, 6:

- Calculate the mean, median, mode, variance, and standard deviation from first principles
- Without using the formula, is this data symmetric or skewed? How do you know? Is it normally distributed?
- How would your above calculations change if we subtracted 7 from each observation? If we multiplied by 2 and *then* subtracted 7? How do these differ?
- [OPTIONAL] Compute z-scores for each observation. What is the mean of these z-scores? What is the standard deviation? Does this surprise you?
- [OPTIONAL] Do you expect these z-scores to be normally distributed? Why or why not?

Question 3: Manual data entry

Use the following code to input *some* data into your Stata browser (it does not have to match my data) and compute summary statistics.

```
clear
input myvar // this can be any name you wish
0
3
7
.
.
.
3
1
4
5
end

summarize myvar
summarize myvar, detail
```

- What is the IQR of your data? What is the skewness and kurtosis? What do these values mean?
- Obtain a table of frequencies for your data using `tabulate [varname]`.
- Give a histogram of the data with bin width one using `histogram [varname], width(1) frequency`.

Question 4: Data in Stata

Download AED_CALELECTRICITY.DTA from the website above and bring it into Stata.

- a) Describe the data using `describe` - what variables are here? What do they correspond to?
- b) Obtain a box plot for both the spot price and one-day ahead forward price of electricity using `graph box [varnames]` and save it with `graph export myboxplot.png, replace`.
- c) Obtain summary statistics for both series. Which has a higher degree of dispersion? Are they symmetric or skewed?
- d) Log transform the spot price variable `ln_niso` using `generate ln_niso = ln(niso)`. Label this new variable “Log of spot price California” using `label variable ln_niso [VARLABEL]`.
- e) Plot the kernel density functions of both `niso` and `ln_niso` (in separate graphs) against a normal density function with `kdensity [varname], normal`. How does the shape of the distribution change? Is this surprising? Which appears more normal?
- f) Plot a histogram of `niso` and `ln_niso` together on the same graph with the code below (you do not need to understand this code). Does this look strange? Why might this be?

```
histogram niso, freq fcolor(red) addplot(hist ln_niso, freq fcolor(blue))
```