McGill University
Department of Economics
Economics 257D: Economic Statistics

Assignment 1: descriptive statistics

J. W. Galbraith

1. Obtain data on the percentage change in US real gross domestic product (GDP) and the civilian unemployment rate from the FRED (Federal Reserve Economic Data) data section of the web site of the Federal Reserve Bank of St. Louis: `https://fred.stlouisfed.org`. If you scroll down on the main page, each of the series will appear (as of the time of writing) as one of the featured series ('at a glance'), so that you can just click on the name of the data series (in blue) and you should see a graph of the series. You can then download the data by clicking on the download button on the right side of your screen. This will transfer a file to your computer in the format you choose. While I normally prefer to keep human-readable copies of small data sets such as these, in ASCII text format, you might prefer to download the simply as .xls files as inputs to Excel or Matlab.

We can label these two data series as $X^a$ and $X^b$. Note that $a$ and $b$ are labels, not exponents.

If you download the files in human-readable ASCII text, you will have to edit the files to eliminate explanatory information and dates before processing with Excel or Matlab. You can use a plain-text editor (e.g. Notepad, OxEdit, etc) to do this.

In your assignment, indicate the day on which you downloaded the data, since the data sets may be updated.

1 a. Compute the sample values of the mean, variance, standard error, coefficients of skewness and kurtosis of each random variable.

1 b. Compute the medians, noting that the series may have an even number of observations depending on when you download.

1 c. Produce histograms of the data with a bin width of 0.3, to start with. Comment on the apparent appropriateness, or otherwise, of this bin width for these data sets and sample size. Describe what would happen if you were to use bin widths of 0.1 or 1.00. Switch to another bin width for either or both of the series, if you think that this would be more revealing of the actual distribution of the data.

1 d. Compare the coefficients of skewness and kurtosis with the values that you would expect in a Normal distribution: 0 and 3 respectively. Do these appear Normal? Of course, we haven't learned yet how to do a formal test, so just make a rough judgment

2. Consider the following random variables derived from $X^b$ : (we'll just use $X^b$ here). $Y = X^b + 2$; $Z = 2X^b$; $W = 4X^b + 3$; $V = W + 2Y$. In each case, give an estimate of the mean and variance of the new random variable (these are estimates because the mean and variance obtained in the first part of the question are estimates; if those were population values, you could compute population values for the new random variables). If you find this non-transparent or have any doubts about your answers, you can actually construct the new variables and do the computations again from scratch. Might be wise in any case.

3. If you are familiar with software that allows you to do this (e.g. Matlab, Python, R, etc.) generate a sample of 200 pseudo-random numbers ($x_i$) from some standard distribution that your software can reproduce, such as the Uniform or Normal, with values mostly in a small interval around 0 (eg within +/- 3 or within 0 to 1, or some such interval).

Alternatively, just work with the unemployment rate data above, and modify the series as described below.

Next compute $\sum_{i=1}^{200}(x_i - \overline{X})^k$, for exponents $k = 2, 4, 6, 8$.
Now consider the following two experiments:

a . multiply all values in the original series by 2, and recompute the results (of course you can figure this out without doing it, as in Q2).

b . replace just one single value in the original series by the number 20 (use 50 if you are using unemployment rate data), and recompute the results.

Which experiment produces the greater change in the statistics, for each value of the exponent $k$? Interpret the result.