BSTAT 3222 - Homework 6

by Craig W. Slinkman

November 12, 2015

# Homework objectives

Because we will need to assess the normality of data in the rest of this course we will need to determine if a set of data is approximately normally distributed. We can do this with normal QQ plots and with the various hypothesis tests.

Our null and alternative hypothesis are stated below:

The null hypothesis is that the i of is normally distributed. The alternative hypothesis is that the null hypothesis is nor normally distributed.

Unfortunately for small samples statistical tests lack the power to detect false null hypotheses. For large samples they often have too much power. This leads to the rejection of the null hypothesis for small deviations of approximately normal distributions.

Approximately normal distributions are what we are looking because in the real world of science and economics there are no true normal distributions. Therefore, we should QQ plots when the sample size is large.

When the sample size is small we should use a hypothesis test because if we are able to declare a distribution to be be non-normal it is very likely to be normal.

# Required

You will have two sets of data to analyze here. For each set if data you must use the following steps:

## Null and Alternative hyothesis in English

For example:  
The null hypothesis is that the scores on homework 5 are normally distributed.  
The alternative hypothesis is not normally distributed.

## Formal Statistcal hypothesis

You must include the formal statistical hypothesis set. You can locate the R-Markdown document used to generate this word document and then copy and paste the hypothesis set in Latex math notation, Your hypothesis should look like

## A discussion of the analysis technique you will use

What methods are you going to use to determine normality or non-normality. Hints will be given in class on how you should proceed. It is often useful to do all the tests and the QQ-plot and compare your results.

## You statisical conclusions

What do you conclude about the distribution of your variable interest. Is it approximately normal or not normal. Use correct English here!

## Subject matter conclusion

Is the distribution normally distributed. Explain your conclusions about the variable of interest. This can be a more technical discussion then when you give a managerial conclusion because managers usually are not interested in normality. An exception is in finance where not-normality can lead to an underestimate of risk.

# The variables

We have two cases that you will analyze. Every student should have different data sets for this homework. If have duplicated data sets and I discover it ***both*** students will earn a zero on the homework.

## Daily stock returns

Load the R-script named and run it. This will create and save a random data set and save it as a data file in the project directory. Then proceed with an analysis of normality of the variable .

## The French Army Draft

Load the R-script named and run the script. This will create a comma separated variable file called which contains the heights of a random sample individuals drafted by the French Army prior to world war I.

Before first world war drafted recruits had their personal physicians give there physical. The French Army suspects the personal physicians are cheating on individuals who are close to the minimum height requirement. They know the the height of draftee's should be normally distributed. Is there evidence that the distribution is non normally distributed?