

Econ C142 - Section 1

Ingrid Haegele and Pablo Muñoz

January 23, 2019

1 Section Syllabus

2 R basics

We prefer to use R via an open-source integrated development environment called RStudio. We recommend that you do the same. You can download the program here: <https://www.rstudio.com/products/rstudio/download/>.

There is a neat way to integrate your code, results, and other text into a document compiled using the document preparation system \LaTeX . \LaTeX is commonly used by academics and learning how to use it -- especially in this stage of your careers -- is well worth the startup costs.¹

Nearly any time you do anything in R, you'll be using functions or datasets contained within a package. The first time you use a package, you will have to download and install it first, you do this with the `install.packages()` function. Once you've installed a package (which you only have to do once), you'll need to tell R to hold its contents in memory for your use. We do this with the `library()` function.

To practice, let's use the R built-in *mtcars* dataset.

1. Open RStudio and create a new .R script.² Install the packages: `haven`, `dplyr`, `ggplot2` and `statar`.
2. Load the *mtcars* dataset and explore the data using the commands: `data()`, `head()`, `nrow()`, `ncol()`, `View()`.
3. Create a smaller dataset that only includes the variables: miles per gallon (*mpg*), weight (*wt*) and horsepower (*hp*). Provide some descriptive statistics of car performance based on *mpg* and plot an histogram of *mpg*.

¹The R package that allows us to compile our code, results, and text into the same document is called *knitr*. Although we won't force you to write your assignments using *knitr*, they will look much better and be much easier for you to prepare.

²Notice that, for good practice, everything you do in R should be contained in a .R script file for the purposes of reproducibility.

4. Create a plot showing the relationship between *mpg* and *hp*. Does it make sense? Make sure your graph is self-explanatory, i.e. include a title and label the axes.
5. Finally, use the `quantile()` function to find 5 quantiles of *mpg*. Then, create a scatter plot showing the mean *hp* against the mean *mpg* in the quantiles. When could a binned scatter plot like this be useful?

3 Hypothesis Testing³

You are hired by the governor to study whether a tax on liquor has decreased average liquor consumption in your state. You are able to obtain, for a sample of individuals selected at random, the difference in liquor consumption (in ounces) for the years before and after the tax. For person i who is randomly sampled from the population, Y_i denotes the change in liquor consumption. Treat these as a random sample from a Normal distribution with mean μ and variance σ^2 , i.e. $N(\mu, \sigma^2)$.

1. The null hypothesis is that there is no change in average liquor consumption. State this formally in terms of μ .
2. The alternative is that there was a change (probably a decline) in liquor consumption; state the alternative in terms of μ .
3. Suppose your sample size is $n = 900$ and you obtain estimates $\bar{y} = -32.8$ and $s = 466.4$. Calculate a test statistic for testing H_0 against H_1 ; what is the distribution of this statistic? Do you reject H_0 at the 5% level? At the 1% level?
4. Construct a 95% confidence interval for μ based on this sample. Would you say that the estimated fall in consumption is large in magnitude? Comment on the practical versus statistical significance of your estimate above. What has been implicitly assumed in your analysis about other determinants of liquor consumption over the two-year period in order to infer causality from the tax change to liquor consumption?

³This exercise is similar to Question C.6 from Wooldridge Appendix C.