**Homework 1 -Math 308 –Sept 1**

* Read Chapters 3, 4, 5.
* Do problems:
	+ 3.36,
	+ 4.41,
	+ 4.37,
	+ **4.51**= use R to do a boxplot, a histogram and a stem-and-leaf plot of the data.
	+ Do the comparative boxplot for the data in problem **5.40** (file with the data can be found in [http://www.albany.edu/~reinhold/m308/IntroStats/Ch05 Cholesterol.xls](http://www.albany.edu/~reinhold/m308/IntroStats/Ch05%20Cholesterol.xls) – copy each column separately – no header!!)

Link is not working. Try this:
Go down one step in the link, that is, go to: [http://www.albany.edu/~reinhold/m308/IntroStats/](http://www.albany.edu/~reinhold/m308/IntroStats/Ch05%20Cholesterol.xls)

You will see tons of files in there. One of them is **Ch05 Cholesterol.xls**
Click in that file. I just tried it and it did work for me.

* **Download R** from: <http://cran.r-project.org>
* **Sing in for the online learning environment** [**www.**coursecompass.com](http://www.coursecompass.com)
	+ Course ID: reinhold76269
	Course Name: Mat 308 Fall 2010
	Course Materials: Intro Stats 3/e
	Description: Topics in Statistical Inference

***R code:***

 to enter data: copy a column of data into the clipboard (highlight and do Control-c)

variablename <- scan()

(paste the data (Control-v) )

If the data consists of just a few numbers, like 3, 4,1,6, you can enter them as a vector. Say I want to call it X, then the command is: X <- c(3,4,1,6)

Plots we can do:

boxplot(variablename)

hist(variablename, probability=T)

you can tell R to do specific breaks in your histogram by giving it a vector with the breaks. So if I want the breaks a1, a2, … ,an then assign the breaks

 breaks=c(a1,a2,…,an)

hist(variablename, probability=T, breaks=c(a1,a2,…,an))

stem(variablename)

To do more than one boxplot, lets say you have data in vectors A, B and C.

Then do

 boxplot(A,B,C)

To save your work: highlight the commands, copy and paste them in a word file. To save the graphs, right click and click on “copy as metafile”, then paste it in the word document.

***Excel Commands:***

When you start a cell with =, excel expects a formula.

To compute functions of data stored in column A from 1 to 30:

|  |  |
| --- | --- |
|  | **Excel fomula** |
| min | =min(A1:A30) |
| Q1 | =percentile(A1:A30,1/4) |
| median | =median(A1:A30) |
| Q3 | =percentile(A1:A30,3/4) |
| max | =max(A1:A30) |
| mean | =average(A1:A30) |
| stdev | =stdev(A1:A30) |
| IQR | =(position of where you stored Q3)-(position of Q1) |

**Determining outliers:** “ 1.5 IQR away from the box”

 Data points that are above **Q3+ 1.5\*IQR** or below **Q1-1.5\*IQR** are considered outliers.

Critical values for normal (top α100 percent) =norminv(1- α, mean, stdev)

For example, in class we did top 5%cut off for standard normal =norminv(.95,0,1)

 top 5% cut off for IQ problem =norminv(.95,100,16)