MES: Quantitative & Qualitative Methods

Week 2, Tuesday Lab Report

Student 1\_Paul Pickell (MES Lab Aide)\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_Student 2

HARDCOPY after lab Thursday, or electronically, as a .docx to: [\\Orca\programs\qqmethods\Workspace\Wk2TueLab](file:///\\Orca\programs\qqmethods\Workspace\Wk2TueLab)

FILENAME: your last names

Today, we will continue the tutorial that we started last week, from [Clarkson University](http://www.clarkson.edu), [Dept. of Mathematics](http://http:/http/www.clarkson.edu/math) : <http://www.cyclismo.org/tutorial/R> . Before you begin, before you even invoke R:

1. Find our fileshare on Orca (qqmethods) and use the Workspace directory where you and your partner store files for the R labs. I suggest you use a different directory for each week’s labs, so create a new directory to use this week (e.g., Week2Lab). Map your network drive so you can get there in one click. This will make your work more efficient.
2. Copy this document (wk2TuesLab) from Handouts to your directory on Workspace. Print a hardcopy, or use it on line. As you work on the lab, complete a hardcopy or fill it out electronically. Aim to complete, and hand in today or Thursday:

* This Lab Report. **Hand in ONE copy per pair**! hardcopy given physically to judy or kathleen, or electronically, i.e., copied (do NOT drag/drop or cut/paste) to \\Orca\programs\qqmethods\Workspace\\_hand\_InWk2Lab, named with your two last names, e.g., SaulCushing.docx.
  + alternatively, you could upload this to the moodle after Thursday’s lab, if you cannot easily upload it to the fileshare from off campus.
* A jpeg file of the plot of your choice, named with your two last names, e.g., [\\Orca\programs\qqmethods\Workspace\\_hand\_In\Wk2TueLabReport\SaulCushing.jpg](file:///\\Orca\programs\qqmethods\Workspace\_hand_In\Wk2TueLabReport\SaulCushing.jpg)

1. Also, remember to list (in a separate sheet of paper, or in a separate file in your Workspace directory) the R commands you learn.

**Now, more fun with R! Using ISwR package/library.**

Invoke R. Be sure the R-console is active and explore the Nav Bar. Be sure you understand how to access the ISwR library, and its associated data files:

1. From the Nav Bar, click on “Packages->Load Package(s)”. A window will appear with a list of packages. If ISwR is on the list (as it should be on all CAL machines), highlight and click OK. At this point, ISwR has been loaded into your workspace. Proceed to 3 below!
2. If ISwR is NOT on the list, you will need to install it. Click on “Install Package(s)” and follow directions. Dalgaard, Appendix A might help. Once the package is installed, load it. On my laptop, files for this package reside on both
   1. C:\Program Files\R\R-2.12.2\library so R would find it, and
   2. in my own directory MES\qqm\R\ISwR\_2.0-5\ISwR\rawdata so I could see it, but you do not have to do this.

Note: You will not have to change your working directory to load the library (but if you want to load one of the files, you will have to use File->Change dir… to set make R point there. Use dir() to check your current directory.

Go back to Step 1.

1. Attach/read the data set of interest from the library into your workspace. There are two kinds of data sets included in the ISwR library! Some are in separate files, in the directory ISwR\rawdata, and you read them into your R workspace just as you have did with the tutorial once you have set your directory to point to that directory (e.g., read.csv to load stroke.csv). Others are included with the library, and you read them into R differently. To do Monday’s assignment, first make sure the ISwR library is loaded. Then type:

* attach(juul)

This will load a data frame named “juul” into your workspace. You can look at in the same way you look at other R variables, e.g.,

* names(juul)
* juul[0]
* juul

etc. But, be aware that juul has about 1330 rows, so “juul” will display many lines on your console!

**Now you are ready to start this week’s R assignment!**

**Your R assignment for this week (Due NEXT MONDAY, April 11, 5PM), to the moodle:**  Create a descriptive statistical summary of the ISwR dataset juul , including captioned plots and graphs. Use the R commands and plot options you have learned thus far to create a two page (maximum) data analysis containing descriptive statistics and graphics that, in your opinion, describe the dataset. Chapter 4 of Dalgaard will help you and feel free to use it, but also be creative! You should do this assignment jointly with your lab partner.

**Once you are sure you understand the assignment and can access the library where you will find the dataset, start today’s lab:**

Start up a browser (Firefox is recommended), and navigate to: <http://www.cyclismo.org/tutorial/R/>. Our object is to finish the following tutorial sections:

* 1. complete Part 5: Plotting. Again, turn in one jpeg file with the result of a plot (of your choice) that you do.
  2. We will start next Tuesday with Parts 6-8 of the tutorial, but feel free to work ahead once you have done a good job on the assignment!
     1. Part 6 linear least squares regression and Part 7 confidence intervals.
     2. Start Part 8 calculate p-values.

1. Answer the questions on Part 5. Questions on Parts 6,7,8 will be given in next week’s lab!

**Questions:**

Do not wait until you complete the tutorial to answer the following questions, i.e., answer the questions for Part 1 before you proceed to Part 2, and Part 2 before going to Part 3, etc.

Questions for Part 5: Plotting

1. Let’s say you include the plot “Leaf BioMass in High CO2 Environment”, with a histogram, box plot, and strip chart in a publication. Write a caption for this plot that explains what the user sees.

> hist(w1$vals,main='Leaf BioMass in High CO2 Environment’, +xlab='BioMass of Leaves',ylim=c(0,16))

> boxplot(w1$vals,horizontal=TRUE,at=16,add=TRUE,axes=FALSE)

> stripchart(w1$vals,add=TRUE,at=15)

**The histogram displays the frequency of leaf biomass values using vertical bars; the boxplot defines the quartiles, minimum value, maximum value and mean; and the strip chart shows the density of the data along the x-axis.**

1. Which of the following two boxplots are best, and in your own words explain why :

* boxplot(tree$STBM~tree$C)
* boxplot(tree$STBM,

main='Stem BioMass in Different CO2 Environments',

ylab='BioMass of Stems')

**> boxplot(tree$STBM~tree$C) most likely, this boxplot best represents how the stem biomass data (STBM) are distributed among different CO2 environments/levels (C).**

**“tree$STBM~tree$C” means “tree$STBM by tree$C”.**

**> boxplot(tree$STBM,main=’Stem BioMass in Different CO2 Environments’,ylab=’BioMass of Stems’) is inaccurate because it is not representing stem biomasses in different CO2 environments, but instead is representing stem biomass data for all CO2 environments - combined. In some (rare) cases, this might be what you want….**

1. When you created a boxplot(tree$STBM~tree$C), how did you annotate your plot with different labels for each level (the tutorial suggests using help(boxplot) to figure out how to do this?

**This was indeed on the boxplot help page! To include labels on a boxplot, use the names parameter/argument to the boxplot() function . names should be set to a character vector.**

**e.g.**

**>boxplot(tree$STBM~tree$C,names=c(‘none’, ‘ambient’, ‘525’, ‘700’))**

1. Given the qqnorm(w1$vals), qqline(w1$vals) add the theoretical line that the data should fall on if they were normally distributed. Say in your own words what that line is.

**qqnorm plots the sample quantiles against the theoretical quantiles of a normal distribution.**

**qqline(w1$vals) will draw a straight line – on which the scatterplot elements would fall if the sample data were normally distributed. In this case, since most of the data fall on or near that line, I would say the leaf biomass is normally distributed.**

1. Extra credit. Another instance of where you might want to draw a “theoretical line” as in the previous qqline example, might be a scatterplot where you suspect a linear relationship as with plot(tree$STBM,tree$LFBM).

answer?

1. Submit a jpeg of your favorite “plot of the week” as a jpeg. Rename the file as YourLastNames.jpg, and copy that to \\Orca\programs\qqmethods\Workspace\\_hand\_In\Wk2LabReport

Remember, to save the graph, use (from the nav bar): file->save as->Jpeg->50% quality

Questions for Part 6: linear least squares regression.

coming Week 3….

Questions for Part 7: confidence intervals.

coming Week 3….

Questions for Part 8: calculating p-values.

coming Week 3….