## 36-617: Applied Linear Models Fall 2020 HW01 – Due Mon Sept 7, 11:59pm

- Please turn the homework in online to GradeScope in our course webspace at canvas.cmu.edu. The easiest way to do this is to go to the Assignments area of our Canvas webspace, then click on HW01, then use the link there to submit to Gradescope. Upload <u>one</u> file per person.
- Please upload only one file (pdf format ONLY) for each homework assignment. If you are an RStudio user, the easiest way to do this is to make an RMarkdown file for your homework solutions, and then "knit" it to pdf. Other approaches such as making an MSWord file or LATEX file and converting it to pdf, or writing things out by hand and scanning to pdf with your phone, etc., are also acceptable.
- Please install the latest version of R on your personal computer using the instructions under the appropriate link in the "Download and Install R" section of https://cran.r-project.org/. (You may also want to install RStudio from https://rstudio.com/). After R (and possibly RStudio) is installed,
  - In your web browser, open http://www.cs.cmu.edu/~10702/R2/Rintro.pdf so you can read it. (This is also available in the hw01 folder in the Files area of our Canvas course site.)
  - If you have never used R before, type in all of the examples from Rintro.pdf into R.
  - If you have used R before, read the examples in Rintro.pdf and try to predict what will happen. Type in any examples where you are not sure.
- For the exercises below, please download the course notes *Using R for Data Analysis and Graphics: Introduction, Code and Commentary*, by J. H. Maindonald, Centre for Mathematics and Its Applications, Australian National University (usingR.pdf), and the accompanying data file (usingR.RData). You can get these in the hw01 folder in the Files area of our Canvas site.

Save the file "usingR.RData" to whatever directory or folder you will work on this assignment in. Then start R (or RStudio), and use "change directory" under the "File" menu in R (in RStudio: "Set working directory" under the "Session" menu), or use the R command setwd(), to change to that directory. All R exercises can be done in either R or RStudio.

Install the "usingR" files with the command

> load("usingR.RData")

and verify that all the files have been installed with the ls() command, as follows:

0			
"ais"	"anesthetic"	"austpop"	"Cars93.summary"
"dewpoint"	"dolphins"	"elasticband"	"florida"
"hills"	"huron"	"islandcities"	"kiwishade"
"leafshape"	"milk"	"moths"	"oddbooks"
"orings"	"possum"	"primates"	"rainforest"
"seedrates"	"tinting"		
	() "ais" "dewpoint" "hills" "leafshape" "orings" "seedrates"	() "ais" "anesthetic" "dewpoint" "dolphins" "hills" "huron" "leafshape" "milk" "orings" "possum" "seedrates" "tinting"	() "ais" "anesthetic" "austpop" "dewpoint" "dolphins" "elasticband" "hills" "huron" "islandcities" "leafshape" "milk" "moths" "orings" "possum" "primates" "seedrates" "tinting"

It is not necessary to do anything else to load individual data files for this assignment, just use the data files as indicated in the MainDonald class notes. If you quit R, save your workspace; R will save your work in a file called ".RData". If you double-click on the ".RData" file to restart R, all your old work, including the above files, should be accessible again.

• For next week, please read Scheather Chapters 1 & 2.

## **Exercises**

- 1. Read & try all of Chapter 1 of Maindonald. You may also need the data file austpop.txt which you can also find in the hw01 folder. Then please do and turn in the following exercises:
  - (a) Chapter 1, #1.
  - (b) Chapter 1, #2.
  - (c) Chapter 1, #3.
- 2. Read & try all of Chapter 2 of Maindonald. You will use the data sets you created for Chapter 1, as well as some of the data sets you loaded with "load(usingR.Rdata)". Then please do and turn in the following exercises:
  - (a) Chapter 2, #1.
  - (b) Chapter 2, #2.
  - (c) Chapter 2, #5. Note that the formula for the volume of a sphere didn't quite come out right in the notes. It is  $4\pi r^3/3$ . What happens when you type "pi" at the R prompt?
  - (d) Chapter 2, #6.
- 3. Read & try all of Chapter 3 of Maindonald. Again you will use the data sets you have created or loaded previously. You can always make a new graphics window with dev.new(), and make the most recent graphics window go away with dev.off(). Then please do and turn in the following exercises:
  - (a) Chapter 3, #1. Notes:
    - For part(b): If you are an Rstudio/rmarkdown user, you will find that identify() does not work inside rmarkdown. Instead, make the plot and run identify() from the console in Rstudio, save the result as an image file, and import it to rmarkdown. Here are two helpful websites for importing figures into rmarkdown:
      - https://stackoverflow.com/questions/46901438/bringing-an-image-into-rmarkdown
      - http://zevross.com/blog/2017/06/19/tips-and-tricks-for-working-with-images-and-figures-in-r-markdowndocuments/
    - For part (c), I personally think lag.plot(huron\$mean.height,do.lines=F) looks better (and easier to interpret!) than just lag.plot(huron\$mean.height).
  - (b) Chapter 3, #3. Note: there is a typo in parb (b) of this question. The code should be stem(possum\$hdlngth) and not stem(qqnorm(possum\$hdlngth)) (i.e. remove the call to qqnorm which doesn't make any sense here).
  - (c) Chapter 3, #5.

If you encounter anything mystifying in the above exercises, please discuss with me and/or the TA.

Some good online sources for R help:

- QuickR: http://www.statmethods.net/
- Cookbook for R: http://www.cookbook-r.com/
- Online course: https://www.datacamp.com/courses/free-introduction-to-r
- Data display using tidyverse and ggplot: https://kieranhealy.org/publications/dataviz/

... and I'm sure you can find more by Googling!