STAT 306 Finding Relationships in Data

Lab 1 - Introduction

David Lee

University of British Columbia
Department of Statistics

January 13/15, 2015

Welcome!

- This course focuses on regression analysis, or more generally relating response variables to explanatory variables. There are also topics on finding low-dimensional structure of a (possibly huge) data set with many variables.
- Course website: http://unixlab.stat.ubc.ca/~stat306/. Do check it often!
- Slides for my labs are available at http://www.stat.ubc.ca/~david.lee/stat306/. I will also make announcements there.

Lab activity

- There are weekly labs for this course. In some of the labs, you will work on problems that serve to consolidate your understanding on the course materials.
- Answers to lab questions are to be submitted on WeBWorK.
- Although individual submission is expected, we encourage discussion during labs. Feel free to ask when you're not sure what to do!

Using Mac Machines

- To log in the Mac machines, the user name is the first 8 characters of your name (first, middle if any, last) and the password is "S" plus the first 7 digits of your student number.
- Remember to log off before you leave!
- You'll notice that there's only one key on your mouse. To achieve the effect of a right click, do "Control + click". You need this to download files from the Internet — simply clicking on the file name will sometimes open it instead.
- The shortcuts for copy and paste are "Command + C" and "Command + V" respectively (*not* Control).

R

- We use R in this course (in fact, in this department). It's a powerful statistical software that is available at no cost at http://cran.r-project.org/. The latest version is 3.1.2 released on October 31, 2014.
- To open R on the Mac machines you are using, click on the magnifying glass at the top right corner (Spotlight) and type `R`. Click on the first search result.
- You may have used R commander before, but it lacks flexibility. To be a real pro you should use the R console.
- If you want to use your own computer, make sure R is installed.

R - useful tips

- Every built-in function you use comes with a help file. Try typing `?plot` or `help(plot)` - it opens the help file for the function `plot` which allows you to plot variables. Knowing how a function works is very important as you will encounter many different functions later on in this course.
- Some specific functions that you may come across are developed in packages that are not yet installed in your machine. If this is the case you have to install them, and load them into R. There is a list of available packages at http://cran.r-project.org/web/packages/available_packages_by_name.html.

R - basic commands

- In this lab, we will go through some simple but important commands in R that you will use frequently throughout the course:
  - Assigning values to variables;
  - Simple arithmetic operations;
  - Creating vectors and matrices, and extracting elements of them;
  - Applying a function to each row/column of a matrix;
  - Matrix manipulations such as matrix multiplication and finding the inverse.

Using R as a calculator

- Try Exercise 1.2 in the Course Notes, using R as a calculator to obtain your answer. The question is reproduced here for your convenience.
- Consider 600 subjects in a clinical trial relating to the level of LDL (bad cholesterol). A 3-way frequency table is given below:

Sex (x_1)	Treatment (x_2)	Frequency		
		Low LDL ($y = 0$)	High LDL ($y = 1$)	Total
0	0	1	134	135
0	1	109	39	148
1	0	42	127	169
1	1	123	25	148
Total		275	325	600

- Find the following probabilities: (a) $\Pr(X_1 = 0, X_2 = 0)$; (b) $\Pr(Y = 1, X_1 = 1, X_2 = 1)$; (c) $\Pr(Y = 1 | X_1 = 0, X_2 = 0)$; (d) $\Pr(X_1 = 1, X_2 = 0 | Y = 1)$.

Reading files into R, and other stuff

- To read files, we typically use the function `read.table`. The `header` argument specifies whether a row of column names is provided in the file. Meanwhile, the argument `sep` specifies the separator (delimiter) used in the file to delimit variables.
  - Type `?read.table` to see how to use this function!
- As a practice, download the file `stock.txt` on the lab webpage and see if you can load it into R.
- Try to play around with the data you have loaded - extract variables, plot histogram (using function `hist`) or scatterplots (using function `plot`.
- You can export matrices, data frames etc. in R to your hard disk, using the command `write.table`. Refer to the help document to see how it can be used.