Seminar Brief Week 1: 16.10.20

BEMM460 STATISTICS FOR BUSINESS ANALYTICS T1 – 2020/21
Instructor: Dr Lalitha Dhamotharan

Please try to install R and R Studio prior to the Seminar by following the instructions below. You can complete Task 1 to Task 4 during the seminar session.

Where to Get R and How to Install R

Before we can work with R, we have to install it. R is provided free of charge, and can be downloaded from the following website:

http://www.r-project.org/

Once you get to this location, note the area in the left-hand margin named "Download." Just below this heading, click on the CRAN link. When you do this, you are presented with the list of hosts called CRAN mirrors (arranged by geographic location) which contain identical R content. Click on the location closest to your own, because the download will typically be faster.

Depending on whether you use Windows or Mac OS, the next steps differ slightly.

Windows. If you are a Windows user:

1. Click on Download R for Windows.
2. When the next screen appears, click on base.
3. Once you see the next page, click on the link Download R 3.2.3 for Windows. (The number 3.2.3 is associated only with the current version of R; this number changes as updated versions of R are made available.)
4. Save the file R-3.2.3-win.exe, double-click, and follow the onscreen instructions. Assuming the installation has been successful, there will be an entry in the All Programs area of the Start Menu. By right-clicking on R 3.2.3, you may place a shortcut R icon on the desktop. To start up R, simply click on the icon in the usual way.

Mac OS X. If you are a Mac user:

1. Click on the Download R for (Mac) OS X link.
2. Click on the latest pkg file. As of this writing, the latest file is R-3.2.3.pkg. (The number will change, however, as updated versions of R are made available.) Once downloaded, the file will appear in the Downloads area.
3. Move R-3.2.3.pkg from the Downloads area to the Applications area and double-click on its entry (icon). Follow the instructions for full installation.
4. The installation process should create an R entry in the Application area. Drag the R entry to the desktop for an R shortcut. To start up R, simply click on the icon in the usual way.

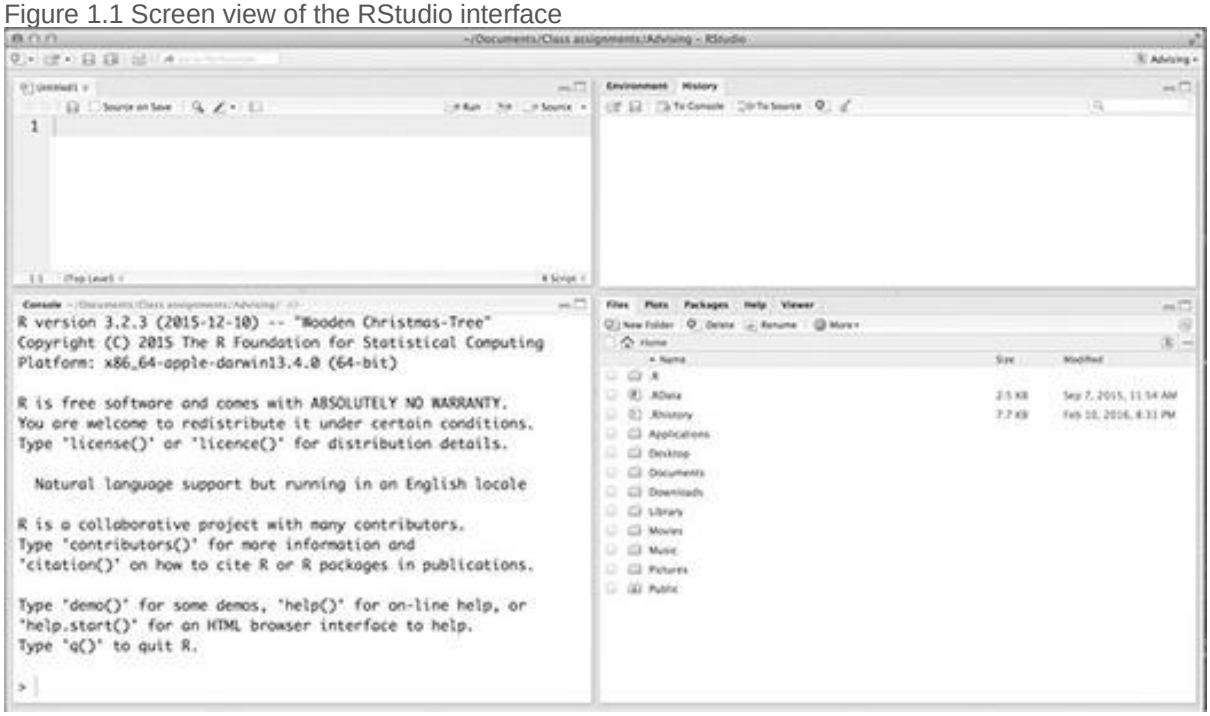
You will also want to install RStudio, an integrated development environment, once you have downloaded the R statistical programming package itself. RStudio can be thought of as a shell in which R operates. It simplifies many basic functions and facilitates the running of R itself, thus making R easier to interact with and more user-friendly. RStudio is provided free of charge and can be downloaded from:

http://rstudio.com

1. When you get to this location, click Download RStudio.
2. When the next screen appears, select Desktop.
3. At the next screen, select Download RStudio Desktop.
4. Depending on whether you use Windows or a Mac, follow the above steps (for installing R). The installation steps are the same for RStudio as they are for R.

Make sure that you have installed R before installing RStudio. RStudio by itself does not include the R program. Once both R and RStudio have been successfully downloaded and installed, a double-click on the RStudio icon launches the program (Figure 1.1) and you are ready to start using R.

Figure 1.1 Screen view of the RStudio interface



There are four panes or windows in the RStudio interface.

1. The Source window, located in the upper left-hand quadrant, is where we write code for our programs. It includes an intuitive, feature-rich text editor that makes it easy to run, save, and access programs.
2. The Console is in the lower left-hand window and is the location where interactive work is done. The RStudio Console is the same as the Console in R.
3. As the name implies, the tabbed Environment/History pane in the upper right-hand area is where the data sets and variables are listed and described (in Environment), and where the history of R commands is displayed (in History).
4. The tabbed Files/Plots/Packages/Help/Viewer pane in the lower right-hand area is a repository of helpful tools which we discuss in subsequent chapters.

To exit RStudio, enter q() at the R prompt and hit the enter key. The R prompt can be seen on the command line of the Console, the lower left-hand pane. It is the greater-than symbol >, and it is the location where we write our R commands when working in an interactive mode.

Task 1

This task is designed to introduce ourselves to each other, i.e. be an ice breaker. It also serves as a practice run for sharing your work with me.

Draw a cartoon of yourself on a piece of paper. If your drawing skills are not great, how about creating a Bitmoji (https://www.bitmoji.com/) or build a Lego character of yourself?

Sharing you work via SharePoint

1. Take a picture of your creation on your mobile phone/tablet

2. On your phone, go this webpage, provide your name, the date, and upload your picture to SharePoint.

3. I will now be able to see your image and share it in class.

4. When your image appears, briefly introduce yourself to the group. Please don't feel pressured to reveal any personal information about yourself that you are not happy to disclose.

Task 2:

On January 23, 2017, the high and low temperatures (Celsius) were recorded for six European cities. Barcelona: 14 and 6; Berlin: 2 and -1; Lisbon: 14 and 3; London: 5 and 0; Paris: 2 and -3; and Rome: 14 and 3. Create a data frame organizing this information for all six cities. Name the data frame E1_1 and the three variables City, High, and Low.

Use the data frame E1_1 to answer the following questions.

a. Use the summary() function to find the mean, median, minimum, maximum, first and third quartiles of the variable High
b. Use the summary() function to find the mean, median, minimum, maximum, first and third quartiles of the variable Low.

Task 3:

Create the object E3_1 consisting of the following elements: 1.50, 1.50, 4.50, 4.50, 11.50, 10.50, 10.50, 10.50, 3.40, 2.00, and 2.00. Using the R functions, find the mean, the median, the mode, the 1st quartile, the 82nd percentile, the range, the interquartile range, the variance, the standard deviation, and the coefficient of variation.

Task 4

a. A useful preliminary step to constructing a cross-tabulation table is to review a small part of the data first. Report the first 10 observations of the mktsurvey data set (found on the companion website) and find the number of observations. As a first step, import the mktsurvey data set into the object E2_2. Hint: use the head() and nrow() functions.
b. Organize the mktsurvey data into a basic cross-tabulation table.
c. Add the row totals to the cross-tabulation of the mktsurvey data. Hint: Apply the rowSums() and cbind() functions to the crosstab object.