IMT 573 Lab: Conditional Probability

Miloni Desai

October 31st, 2019

Collaborators

Objectives

Before beginning this assignment, please ensure you have access to R and RStudio; this can be on your own personal computer or on the IMT 573 R Studio Server.

1. Download the `lab5_conditional_probability.rmd` file from Canvas or save a copy to your local directory on RStudio Server. Open `lab5_conditional_probability.rmd` in RStudio and supply your solutions to the assignment by editing `lab5_conditional_probability.rmd`.
2. Replace the "Insert Your Name Here" text in the `author:` field with your own full name. Any collaborators must be listed on the top of your assignment.
3. Be sure to include well-documented (e.g. commented) code chucks, figures, and clearly written text chunk explanations as necessary. Any figures should be clearly labeled and appropriately referenced within the text. Be sure that each visualization adds value to your written explanation; avoid redundancy – you do no need four different visualizations of the same pattern.
4. Collaboration on problem sets is fun and useful, and we encourage it, but each student must turn in an individual write-up in their own words as well as code/work that is their own. Regardless of whether you work with others, what you turn in must be your own work; this includes code and interpretation of results. The names of all collaborators must be listed on each assignment. Do not copy-and-paste from other students' responses or code.
5. All materials and resources that you use (with the exception of lecture slides) must be appropriately referenced within your assignment.
6. When you have completed the assignment and have **checked** that your code both runs in the Console and knits correctly when you click `Knit PDF`, rename the knitted PDF file to `lab5_YourLastName_YourFirstName.pdf`, and submit the PDF file on Canvas.

Setup

In this lab you will need, at minimum, the following R packages.

```r
# Load standard libraries
library(tidyverse)
```

If a baseball team scores X runs, what is the probability it will win the game?

This is the question we will explore in this lab (ddapted from Decision Science News, 2014). We will use a dataset of baseball game statistics from 2010-2013.

Baseball is a played between two teams who take turns batting and fielding. A run is scored when a player advances around the bases and returns to home plate. More information about the dataset can be found at http://www.retrosheet.org/.

Data files can be found on Canvas in the data folder. Download the files and load them into one data.frame in R as shown below. Comment this code to demonstrate you understand how it works.

Note: More information about the dataset can be found at http://www.retrosheet.org/

```
colNames <- read.csv("/home/imt573/Data/cnames.txt", header = TRUE)
baseballData <- NULL
for (year in seq(2010, 2013, by = 1)){
  mypath <- paste('/home/imt573/Data/GL',year,'.TXT',sep = '')
  baseballData <- rbind(baseballData,read.csv(mypath,
  col.names = colNames$Name))
  baseballData <- tbl_df(baseballData)
}
#View(baseballData)
```

Select the following relevant columns and create a new local data.frame to store the data you will use for your analysis.

- Date
- Home
- Visitor
- HomeLeague
- VisitorLeague
- HomeScore
- VisitorScore

```
df<- data.frame(Date = baseballData$Date, Home =baseballData$Home,Visitor =baseballData$Visitor,HomeLeag
#View(df)
```

Considering only games between two teams in the National League, compute the conditional probability of the team winning given X runs scored, for $X = 0, ..., 10$. Do this separately for Home and Visitor teams. #We first filter out the teams that play in the national league. To find the conditional proability of the winning team, we find the proability for each total number of runs scored.

- Design a visualization that shows your results.
- Discuss what you find. #We can visualize the results of our findings using a line graph, with the proability of winning on the y axis and the runs on the x-axis. We have two lines, one representing the proability of the home team winning and the other representing the proability of the visitor team winning.

Extra Credit: Repeat the above problem, but now consider the probability of winning given the number of hits.