Applied Regression and Time Series Analysis
Lab 1: Probability and Statistics Basics
2016 Summer

May 7, 2016

Instructions

- **Due Date:** Beginning of Week 3 Live Session.
- **Submission:** **Submit 2 files. Missing one of the two files will result in a 50% reduction in grade.**
    1. A report (in pdf format) detailing your answers and all the steps to arrive at your answers
    2. A well-documented R-script, jupyter notebook, or Rmd file detailing all of the codes used to arrive at your answers.
- Late submission will not receive any credit.
- Answers need to be typed. Latex, which you will likely have to use in R markdown, is great for typesetting documents with mathematical symbols.
- All the steps used to arrive at your final answers need to be shown clearly. These steps are as important as the final answer.
- The final answer of each question needs to be very easy identified; the use of bold fonts, highlights, or circling will help.
- This is a group project. Form a group with 3 or 4 people.
- Although this is a group project, we encourage you to attempt all of the exercises before discussing with your teammates. **Do not use the "division-of-labor" approach.** Each of the students in a group is expected to make sufficient contribution to the lab. If any of your teammate does not make sufficient contribution, please contact your instructor.
- DO NOT copy and paste or even leverage on the solutions we gave to the students in previous semesters. Violation of this policy will be reported to the Director of the MIDS program and the Office that oversees UC Berkeley Academic Integrity. In any case, the lab has various subtle changes that make those answers not directly applicable.

Question 1 (10 points)

In a team of 80 data scientists, each of them falls into at least one of the following categories:

- an expert in machine learning
- an expert in statistics
- being awesome

Below includes more information about them.

36 are expert in machine learning, 30 are expert in statistics, and 25 are awesome. 20 are expert in both machine learning and statistics, 15 are expert in machine learning and are awesome, 10 are expert in statistics and are awesome.

Suppose you are in a cocktail party with this group of data scientists and you have an equal probability of meeting any one of them.

1. What is the probability of meeting a data scientist who is an expert in both machine learning and statistics but is not awesome?
2. Suppose the you meet a data scientist who is an expert in machine learning. Given this information, what is the probability that s/he is an expert in either awesome or an expert in machine learning?

Question 2 (10 Points)

Suppose for events A and B, $Pr(A) = p \leq \frac{1}{2}$, $Pr(B) = q$, where $\frac{1}{4} < q < \frac{1}{2}$. These are the only information we have about the events.

1. What are the maximum and minimum possible values for $Pr(A \cup B)$?
2. What are the maximum and minimum possible values for $Pr(A|B)$?

Question 3 (10 Points)

Wooldridge Appendix B Question B.4

Question 4 (10 Points)

Wooldridge Appendix B Question B.5

Question 5 (10 Points)

Wooldridge Appendix B Question B.6

Question 6 (10 Points)

Wooldridge Appendix B Question B.7

Question 7 (10 Points)

Wooldridge Appendix C Question C.6

Question 8 (10 Points)

Wooldridge Appendix C Question C.8

Question 9 (10 Points)

Wooldridge Appendix C Question C.9

Question 10 (10 Points)

Wooldridge Appendix C Question C.10