Section 1: Introduction, Probability Concepts and Decisions

Tengyuan Liang, Chicago Booth

https://tyliang.github.io/BUS41000/

Suggested Reading:
Naked Statistics, Chapters 1, 2, 3, 5, 5.5 and 6
OpenIntro Statistics, Chapters 2 and 3

Getting Started

- Syllabus
- General Expectations
  1. Read the notes
  2. Work on homework assignments
  3. Be on schedule

Course Overview

Section 1: Introduction, Probability Concepts and Decisions

Section 2: Learning from Data: Estimation, Confidence Intervals and Testing Hypothesis

Section 3: Simple Linear Regression

Section 4: Multiple Linear Regression

Section 5: More on MLR, Dummy Variables, Interactions

Section 6: Additional Topics (time permitting)

Course Schedule

Week 1 Section 1

Week 2 Section 1

Week 3 Section 2

Week 4 Section 2

Week 5 Section 3

Week 6 Section 3 and Review

Week 7 Midterm

Week 8 Section 4

Week 9 Section 5

Week 10 Section 6 and Review

Finals Week Final

Let's start with a question...

My entire portfolio is in U.S. equities. How would you describe the potential outcomes for my returns in 2018?

Another question... (Target Marketing)

Suppose you are deciding whether or not to target a customer with a promotion(or an add)...

It will cost you $.80 (eighty cents) to run the promotion and a customer spends $40 if they respond to the promotion.

Should you do it ???

Introduction

Probability and statistics let us talk efficiently about things we are unsure about.

- How likely is Trump to finish a four year term?
- How much will Amazon sell next quarter?
- What will the return of my retirement portfolio be next year?
- How often will users click on a particular Facebook ad?

All of these involve inferring or predicting unknown quantities!!

Random Variables

- *Random Variables* are numbers that we are NOT sure about but we might have some idea of how to describe its potential outcomes.
- Example: Suppose we are about to toss two coins.
  Let X denote the number of heads.

We say that X, is the random variable that stands for the number we are not sure about.

Probability

Probability is a language designed to help us talk and think about aggregate properties of random variables. The key idea is that to each event we will assign a number between 0 and 1 which reflects how likely that event is to occur. For such an immensely useful language, it has only a few basic rules.

1. If an event A is certain to occur, it has probability 1, denoted $P(A) = 1$.
2. $P(\text{not-}A) = 1 - P(A)$.
3. If two events A and B are mutually exclusive (both cannot occur simultaneously), then $P(A \text{ or } B) = P(A) + P(B)$.
4. $P(A \text{ and } B) = P(A)P(B|A) = P(B)P(A|B)$

Probability Distribution

- We describe the behavior of random variables with a Probability Distribution
- Example: If X is the random variable denoting the number of heads in two *independent* coin tosses, we can describe its behavior through the following probability distribution:

$$X = \begin{cases} 0 & \text{with prob.} \quad 0.25 \\ 1 & \text{with prob.} \quad 0.5 \\ 2 & \text{with prob.} \quad 0.25 \end{cases}$$

- X is called a *Discrete Random Variable* as we are able to list all the possible outcomes
- Question: What is $Pr(X = 0)$? How about $Pr(X \geq 1)$?

Pete Rose Hitting Streak

Pete Rose of the Cincinnati Reds set a National League record of hitting safely in 44 consecutive games. . .

- Rose was a 300 hitter.
- Assume he comes to bat 4 times each game.
- Each at bat is assumed to be independent, i.e., the current at bat doesn't affect the outcome of the next.

What probability might reasonably be associated with a hitting streak of that length?

Pete Rose Hitting Streak}

Let A_i denote the event that "Rose hits safely in the i^{th} game"

Then P(Rose Hits Safely in 44 consecutive games) = $P(A_1$ and $A_2 \ldots$ and $A_{44}) = P(A_1)P(A_2)...P(A_{44})$

We now need to find $P(A_i) \ldots$ It is easier to think of the complement of A_i, i.e., $P(A_i) = 1 - P(\text{not } A_i)$

$$\begin{aligned} P(A_i) &= 1 - P(\text{Rose makes 4 outs}) \\ &= 1 - (0.7 \times 0.7 \times 0.7 \times 0.7) \\ &= 1 - (0.7)^4 = 0.76 \end{aligned}$$

So, for the winning streak we have $(0.76)^{44} = 0.0000057!!!$ (Why?)
(btw, Joe DiMaggio's record is 56!!!!)

R Markdown

This is an R Markdown presentation. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see `http://rmarkdown.rstudio.com`.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document.

Slide with Bullets

- Bullet 1
- Bullet 2
- Bullet 3

Slide with R Output

```
summary(cars)
```

```
##      speed           dist       
##  Min.   : 4.0   Min.   :  2.00  
##  1st Qu.:12.0   1st Qu.: 26.00  
##  Median :15.0   Median : 36.00  
##  Mean   :15.4   Mean   : 42.98  
##  3rd Qu.:19.0   3rd Qu.: 56.00  
##  Max.   :25.0   Max.   :120.00
```

Slide with Plot