

# Assignment #2

*Tristan De Alwis*

*10/01/2019*

## Short Answers:

Q: About how long did this assignment take you? Did you feel it was too long, too short, or reasonable? A: ~7 hours

Q: Who, if anyone, did you work with on this assignment? A: Zach Azadian; Haukur Björnsson

Q: What questions do you have relating to any of the material we have covered so far in class? A: None

## Loading Necessary Libraries

```
suppressPackageStartupMessages(library(ggplot2))
```

**Question 1:** 14% of the laptops will be recalled in the next five years. 46% of all laptops are Macbooks. A laptop servicing company notices that 9% of laptops are recalled and not Macbooks.

a. What is the probability that a laptop is a Macbook or recalled?

```
probRecall = 0.14 # P(Recall)
probMacbook = 0.46 # P(MacBook)
probRecallNotMacbook = 0.09 # P(Recalled non MacBooks)

cat("P(Macbook U Recall): ", probMacbook + probRecall - probRecallNotMacbook,
    "\n")
```

```
## P(Macbook U Recall): 0.51
```

b. Given that a laptop is recalled, what is the probability that it is not a Macbook?

```
cat("P(!Macbook|Recall): ", probRecallNotMacbook, "\n")
```

```
## P(!Macbook|Recall): 0.09
```

c. What is the probability that a laptop is a Macbook or not recalled?

```
cat("P(Macbook U !Recall): ", round(probMacbook + (1 - probRecall) - ((1 - probRecall) *
    probMacbook), 2), "\n")
```

```
## P(Macbook U !Recall): 0.92
```

d. What is the probability that a laptop is not a Macbook and not recalled?

```
cat("P(!Macbook & !Recall): ", round((1 - probMacbook) * ((1 - probMacbook) *  
    (1 - probRecall)), 2), "\n")
```

```
## P(!Macbook & !Recall): 0.25
```

e. Are laptop brand and recall status mutually exclusive? Explain.

No, because the data suggests that whether the laptop is a MacBook or not, it can still be recalled

**Question 2:** Consider the fictional company Data Science Supplies, Inc., which produces various types of products. 34% of products are electronics and have probability 0.08 of being defective, in contrast to a probability of 0.03 for the other 66% of non-electronic products.

a. If we pick a product at random, what is the probability that it will not be defective?

```
probElectronic = 0.34  
probDefect = 0.08  
probOther = 0.66  
probDefectOther = 0.03  
  
cat("P(!Defective Product): ", round((probElectronic * (1 - probDefect)) + (probOther *  
    (1 - probDefectOther)), 2), "\n")
```

```
## P(!Defective Product): 0.95
```

b. What is the probability a product is an electronic if it is defective?

```
cat("P(Electronic|Defective): ", round((probElectronic * probDefect)/((probElectronic *  
    probDefect) + (probOther * probDefectOther)), 2), "\n")
```

```
## P(Electronic|Defective): 0.58
```

**Question 3:** Suppose that Las Vegas casinos are introducing a new dice-based table game where players roll a six-sided die two times. Let A be the event that the first roll resulted in a 1, B be the event that the sum of the rolls is 3, and C be the event that the second roll is even. For the following questions, provide a mathematical explanation to justify your answer.

a. Are A and B independent?

No, the probability of B is  $\frac{2}{36}$  and the probability of A is  $\frac{1}{6}$ . The intersection is  $\frac{2}{216}$  but it should be  $\frac{1}{36}$  to be Independent

```

probA = 1/6 # 1 out of 6 possibilities
probB = 2/36 # 2 through 12 are possible sums
probC = 3/6 # 3 out of 6 possibilities

```

**b. Are B and C independent?**

Yes,  $P(B) * P(C) = 1/36$  which is equal to 1 chances of having an even second roll that sums to 3

**c. Are A and C independent?**

Yes, C has no intersection with A.  $P(C) * P(A) = 1/36$  which is equal to the chance of having 1 for the first roll and an even second roll

## Question 4: Two cards are drawn at random from a standard deck of 52 cards.

**a. How many pairs (not ordered) include one Jack or one Ace (or both)?**

```

cat("One Jack: ", 44 * 4, "\n")

```

```

## One Jack: 176

```

```

cat("One Ace: ", 44 * 4, "\n")

```

```

## One Ace: 176

```

```

cat("One Jack & One Ace: ", 4 * 4, "\n")

```

```

## One Jack & One Ace: 16

```

```

cat("Both: ", 176 * 2 + 16, "\n")

```

```

## Both: 368

```

**b. What the probability that the pair includes the Ace of hearts, given the pair includes a Jack and an Ace?**

```

probAceHearts = 1/52
probJack = 4/52
probAce = 4/52
TotalCombs = factorial(52)/(factorial(2) * (factorial(52 - 2)))
cat("P(Ace of Hearts|Jack & Ace): ", (4/TotalCombs)/(16/TotalCombs), "\n")

```

```

## P(Ace of Hearts|Jack & Ace): 0.25

```

## Question 5: The U.S. Senate has 75 male senators and 25 female senators. Suppose a committee of three senators is forming.

**a. Calculate the probability of selecting an all-male committee.**

```

ProbMale = 75/100
ProbFemale = 25/100

```

```
ThreeCombs = factorial(100)/(factorial(3) * (factorial(100 - 3)))
```

```
ThreeMales = factorial(75)/(factorial(3) * (factorial(75 - 3)))
```

```
cat("P(3 Males): ", ThreeMales/ThreeCombs, "\n")
```

```
## P(3 Males): 0.4175943
```

b. Calculate the probability of selecting an all-female committee.

```
ThreeFemales = factorial(25)/(factorial(3) * (factorial(25 - 3)))
```

```
cat("P(3 Females): ", ThreeFemales/ThreeCombs, "\n")
```

```
## P(3 Females): 0.01422387
```

c. Calculate the probability of selecting a committee of two male senators and one female senator.

```
# cat('P(2 Males & 1 Female): ',  
# (factorial(100)/(factorial(2)*(factorial(100-2)))) *  
# (factorial(25)/(factorial(1)*(factorial(25-1)))) / ThreeCombs, '\n')
```

```
cat("P(2 Males & 1 Female): ", (75/100 * 74/99 * 25/98) * 3, "\n")
```

```
## P(2 Males & 1 Female): 0.4290353
```

d. Calculate the probability of selecting a committee of one male senator and two female senators.

```
# cat('P(1 Male & 2 Females): ',  
# (factorial(100)/(factorial(1)*(factorial(100-1)))) *  
# (factorial(25)/(factorial(2)*(factorial(25-2)))) / ThreeCombs, '\n')
```

```
cat("P(1 Male & 2 Female): ", (75/100 * 25/99 * 24/98) * 3, "\n")
```

```
## P(1 Male & 2 Female): 0.1391466
```

e. Calculate the probability of selecting a committee with at least one female senator.

```
cat("P(At least 1 Female): ", 1 - ThreeMales/ThreeCombs, "\n")
```

```
## P(At least 1 Female): 0.5824057
```

f. Calculate the probability of selecting an all-male committee or an all-female committee.

```
cat("P(3 Males or 3 Females): ", (ThreeMales + ThreeFemales)/ThreeCombs, "\n")
```

```
## P(3 Males or 3 Females): 0.4318182
```

**Question 6:** A Class consists of 6 undergraduate student and 4 graduate students.

a. How many different possible rankings are there?

```
factorial(10)
```

```
## [1] 3628800
```

b. What is the probability that graduate students receive the top four scores?

```
FourGrads = factorial(4) * factorial(6)/(factorial(10))
```

```
FourGrads
```

```
## [1] 0.004761905
```

**Question 7:** In the past few years, there has been an outbreak of Ebola in the Republic of Congo. A test is created to help detect Ebola. The test is administered to a group of 84 subjects known to have Ebola. Of this group, 59 test positive. The test is also administered to a group of 428 subjects known to not have Ebola. Of this group, 12 test positive.

a. Present this data in a tabular form similar to the following:

```
s_pos <- 84
```

```
true_pos <- 59
```

```
s_neg <- 428
```

```
false_neg <- 12
```

```
mat <- matrix(NA, nrow = 3, ncol = 3)
```

```
### Set row names and column names
```

```
rownames(mat) <- c("Positive", "Negative", "Total")
```

```
colnames(mat) <- c("Have Ebola", "Do Not Have Ebola", "Total")
```

```
mat[1, 1] = true_pos
```

```
mat[1, 2] = false_neg
```

```
mat[1, 3] = true_pos + false_neg
```

```
mat[2, 1] = s_pos - true_pos
```

```
mat[2, 2] = s_neg - false_neg
```

```
mat[2, 3] = (s_pos - true_pos) + (s_neg - false_neg)
```

```
mat[3, 1] = true_pos + (s_pos - true_pos)
```

```
mat[3, 2] = false_neg + (s_neg - false_neg)
```

```
mat[3, 3] = true_pos + false_neg + (s_pos - true_pos) + (s_neg - false_neg)
```

```
mat
```

	Have Ebola	Do Not Have Ebola	Total
Positive	59	12	71
Negative	25	416	441
Total	84	428	512

b. Calculate the sensitivity and specificity of this test directly from the data.

```
# Sensitivity = P(T+/D)
sensitivity <- 59/84
cat("Sensitivity: ", sensitivity, "\n")
```

```
## Sensitivity: 0.702381
```

```
# Specificity =
specificity = 416/428
cat("specificity: ", specificity, "\n")
```

```
## specificity: 0.9719626
```

c. Assume that the prevalence of the disease is 2.7%. Calculate the NPV and PPV

with this prevalence.

```
0.027 * 512
```

```
## [1] 13.824
```

```
# Negative Predictive Value
NPV <- specificity * (1 - 0.027) / (specificity * (1 - 0.027) + (25/84) * (0.027))
cat("NPV: ", NPV, "\n")
```

```
## NPV: 0.9915747
```

```
# Positive Predictive Value
PPV <- sensitivity * (0.027) / (sensitivity * (0.027) + (12/428) * (1 - 0.027))
cat("PPV: ", PPV, "\n")
```

```
## PPV: 0.410086
```

d. What conclusions can be drawn regarding the effectiveness of this test?

This test is effective at determining those who do not have the disease, but does not give confidence it can detect the disease when it only detects the disease roughly 70% in those that for sure have it. Meaning, ~33% of people are not be accurately diagnosed.

**Question 8:** Recall that the additive rule tells us for events  $A$  and  $B$  that are not mutually exclusive that  $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$ . We can extend this additive rule to more than two events, which gives us the general inclusion-exclusion identity as follows:

a. Explicitly write the inclusion-exclusion identity for  $n = 3$  events (i.e. reduce down so that there aren't summations).

$$P(A \text{ or } B \text{ or } C) = P(A) + P(B) + P(C) - P(A \& B) - P(A \& C) - P(B \& C) - P(A \& B \& C)$$

b. Suppose any integer from 1 to 67 is chosen at random with equal probability. What is the probability this randomly selected integer is divisible by at least one of 5, 7, or 13?

```
Prob5 <- 13/67
Prob7 <- 9/67
Prob13 <- 5/67

cat("P(/5 or /7 or /13: ", Prob5 + Prob7 + Prob13 - 1/67 - 1/67, "\n")

## P(/5 or /7 or /13: 0.3731343
```