

# STAT20 Homework #5

Seth Metcalf

## Table of Contents

|  |   |
|--|---|
| Introduction .....   | 1 |
| Chapter 16 .....   | 1 |
| 1. Ch 16 C2: .....   | 1 |
| 2. Ch 16 C2: .....   | 2 |
| Chapter 17 .....   | 3 |
| 3. Ch 17 B4: .....   | 3 |
| 4. Ch 17 Rev 6: .....  | 4 |
| Chapter 5 .....  | 4 |
| 5. Ch 5 C1: .....  | 4 |
| 6. Ch 5 Rev 10: .....  | 5 |
| Chapter 18 .....   | 5 |
| 7. Based on Ch 18 B3: .....  | 5 |
| 8. Based on Ch 18 Rev 2: .....   | 5 |
| Random Variable Problems .....   | 7 |
| 9. Write down the distribution table and find the expected value and standard error for the number of spades in two draws without replacement from a standard deck of cards (52 cards, 13 spades).....   | 7 |
| Two students flip coins and record the number of heads. They are supposed to flip 200 coins. The first student flips 200 coins, but the second student flips 100 coins and then doubles the result. .... | 7 |

## Introduction

This is Homework #5, which contains questions from Chapters 5, 16, 17, & 18, and contains Random Variable Problems. *Due 28 February 2021.*

## Chapter 16

### 1. Ch 16 C2:

A gambler is going to play roulette 25 times, putting a dollar on a split each time. (A split is two adjacent numbers. For this problem it's not important that they be adjacent, just that

you have 2 possible ways to win.) If either number comes up, the gambler wins \$17. If neither number comes up, he loses a dollar. So a split pays 17 to 1, and there are 2 chances in 38 to win. The gambler's net gain in the 25 plays is like the sum of 25 draws made from one of the following boxes. Which one, and why?

i)

Not this one, this box shows the numbers that the gambler could land on, not the net gain.

ii)

Not this one, there are 38 possible options, excluding the 2 possible winning ones, there are 36 other options, not 34.

iii)

This one, other than the 2 possible options to win, there are 36 other numbers in which the gambler loses, a randomized pull from this box be an accurate representation of the win/loss of the gambler.

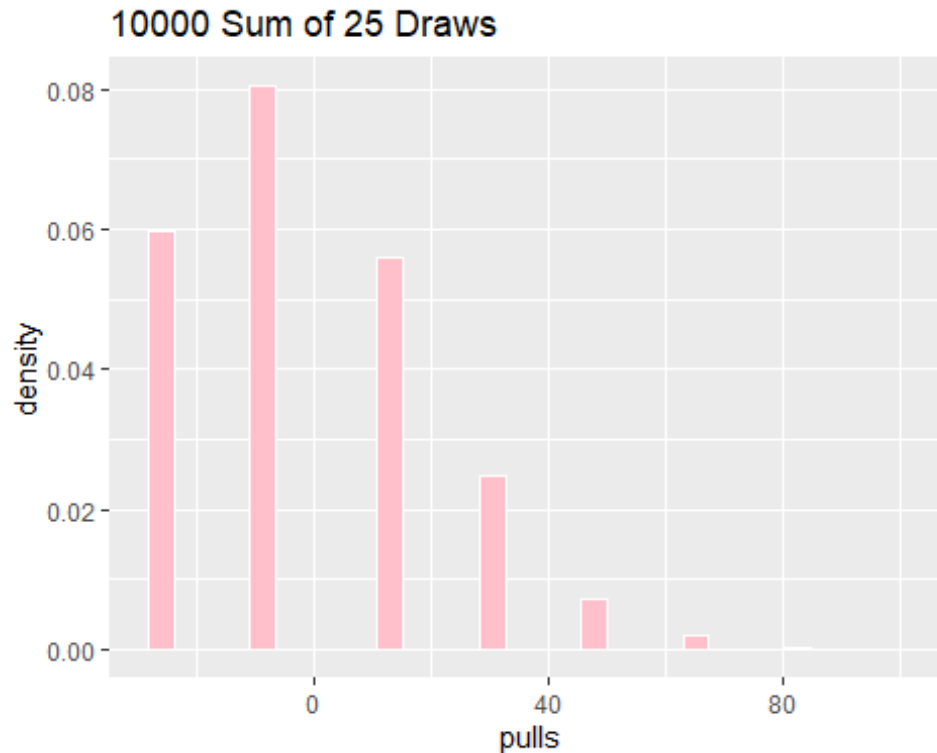
## 2. Ch 16 C2:

write code in R to do a simulation of 1000 gamblers each making 25 plays of roulette according to these rules and the box model you chose. Make a histogram of the results. In your simulation, how many times do the gamblers win money (more than 0 dollars)? This should be done in a code chunk in R Markdown so that your histogram will appear with your submission.

## Code

```
library(ggplot2)
gambleBox = c(-1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1,
-1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1,
1, 17, 17)
gamblePulls = data.frame(pulls = replicate(10000, sum(sample(gambleBox, 25, r
eplace = T))))
ggplot(gamblePulls, aes(x=pulls,y=..density..)) + geom_histogram(col = "white
", fill = "pink") + ggtitle("10000 Sum of 25 Draws")

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



```
sum((sample(gambleBox, 25, replace = T)) > 0)
## [1] 0
```

## Chapter 17

### 3. Ch 17 B4:

Fifty draws are made at random with replacement from the box:

a) Finish the sentence, explain

The sum of the draws turns out to be 157. The expected value for the sum is 150, the observed value is 157, the chance error is 7, and the standard error is 10. Fill in the blanks, and explain briefly.

Since it is told to us that the Sum of the draws turns out to be 157, we know that the observed value of the sum is also 157 as that is what was observed. The expected value is just the sum of the mean of the numbers (3 because  $5!/5 = 3$ ), the chance error is 7 as that is the difference between the observed and expected values, and the standard error is 10 (9.9999041) because you multiply the SD, 1.4142, with the sqrt of the number of draws  $\sqrt{50}$ .

#### 4. Ch 17 Rev 6:

One hundred draws are made at random with replacement from the box: The draws come out with 45 ones, 23 twos, and 32 threes. For each number below, find the phrase that describes it.

#### Matching

Number 12 is the chance error for the sum of the draws, Number 45 is the observed value for the number of ones, 187 is the observed value for the sum of the draws, 25 is the expected value for the number of threes, 50 is the expected value for the number of ones, 175 is the expected value for the sum of the draws, 5 is the standard error for the number of ones, and 32 is the observed value for the number of threes.

## Chapter 5

#### 5. Ch 5 C1:

For the women age 18-24 in a dataset, the average height was about 64.3 inches; the SD was about 2.6 inches. Using the normal curve, estimate the percentage of women with heights-

##### a) below 66 inches

The percentage of women that are below 66 inches in height is approximately 74.34%.

```
suwomen = (66-64.3)/2.6
pnorm(suwomen)
## [1] 0.7433945
```

##### b) between 60 inches and 66 inches.

The percentage of women that are between 60 and 66 inches in height is approximately 69.43%.

```
suwomen66 = (66-64.3)/2.6
suwomen60 = (60-64.3)/2.6
pnorm(suwomen66) - pnorm(suwomen60)
## [1] 0.6943151
```

##### c) above 72 inches.

The percentage of women that are above inches in height is approximately 0.15%.

```
suwomen72 = (72-64.3)/2.6
1 - pnorm(suwomen72)
## [1] 0.001530531
```

## 6. Ch 5 Rev 10:

For women age 25-34 with full time jobs, the average income in 2004 was \$32,000. The SD was \$26,000, and 0.25% had incomes above \$150,000. Was the percentage with incomes in the range from \$32,000 to \$150,000 about 40%, 50%, or 60%? Choose one option and explain briefly.

### a) Explanation

It would be about 50%. Since the range started at the mean, and the mean marks the halfway point for z-scores/normal curve; since the range goes about 5 standard deviations above the mean, it is very close to capturing almost all of the values in that half (above 99.9%), then we can estimate this out to be capturing 100% of everything above the midway point – otherwise known as 50%.

## Chapter 18

### 7. Based on Ch 18 B3:

A fair coin is tossed 100 times. Find the chance of getting exactly 60 heads in two ways: using the `dbinom()` function in R, and using `pnorm()`. Which one gives the exact answer, and which is an approximation? Are the two answers close to each other? Why?

#### a) R Code

```
# using dbinom
dbinom(60, 100, 0.5)

## [1] 0.01084387

# approximation answer P(exactly 60 heads)
sdofbox = (1-0)*sqrt(0.5*0.5)
sesum = sqrt(100) * sdofbox
pnorm((60.5-50)/sesum)-pnorm((59.5-50)/sesum)

## [1] 0.01085214
```

`dbinom` gives the actual answer. `pnorm` gives the approximation. In this scenario, the two answers are close to each other because it follows all the rules when trying to evaluate something through a normal approximations. Furthermore, random probabilities, such as taking coin flips tend to follow the normal curve well.

### 8. Based on Ch 18 Rev 2:

Four hundred draws will be made at random with replacement from the box:

#### a) Estimate the chance that the sum of the draws will be more than 1,500.

Based off of the box, theres about a 98% chance that the sum of the draws will e more than 1,500.

```

boxavg = (1 + 3 + 5 + 7)/4
boxsd = sqrt(((1-4)^2 + (3-4)^2 + (5-4)^2 + (7-4)^2)/4)
ev = 400 * boxavg
se = sqrt(400) * boxsd
1 - pnorm((1500 - ev)/se)

## [1] 0.9873263

```

b) Estimate the chance that the number 3 comes up fewer than 90 times.

The chance that the number 3 comes up fewer than 90 times is about an 89% chance or so.

```

mean(replicate(10000, mean(sample(c(1, 3, 5, 7), 400, replace = T) == 3)) <
90/400)

## [1] 0.11207

```

c) Do a simulation for the chance process in part a), with 10000 sets of 400 draws and get the proportion of times that the sum is more than 1500. It should be close to your answer for the chance.

The chance that the sum of the draws will be more than 1,500 will be about 98.5% or so. It is close to the answer I got in a).

```

drawsTotal = replicate(10000, sum(sample(c(1, 3, 5, 7), 400, replace = T)))
mean(drawsTotal > 1500)

## [1] 0.9854

```

d) Continuing part b), get the answer in R using the pbinom() function and the pnorm() function and compare. Which one is correct, and which one is an approximation?

```

# using pbinom()
pbinom(90, 400, 0.25)

## [1] 0.1358459

# using pnorm()
sdofbox = sqrt(((1-4)^2 + (3-4)^2 + (5-4)^2 + (7-4)^2)/4)
sesum = sqrt(400) * sdofbox
(pnorm((55.5-50)/sesum) - pnorm((44.5-50)/sesum))

## [1] 0.09788003

```

Using pbinom is the correct one while using pnorm is the approximation of the amount of 3's.

## Random Variable Problems

9. Write down the distribution table and find the expected value and standard error for the number of spades in two draws without replacement from a standard deck of cards (52 cards, 13 spades).

Since it is 2 draws without replacement in a deck of cards, the total number of cards goes down by 1 after the first draw.

```
# Chance of drawing 0 cards
chance0 = (39/52) * (38/51)
# Chance of drawing 1 card
chance1 = ((39/52) * (13/51)) + ((13/52) * (39/51))
# Chance of drawing 2 cards
chance2 = (13/52) * (12/51)
# Expected Value
ev = (0 * chance0) + (1 * chance1) + (2 * chance2)
se = sqrt(((0-0.5)^2*(chance0)) + ((1-0.5)^2*(chance1)) + ((2-0.5)^2*(chance2)))
ev

## [1] 0.5

se

## [1] 0.6063391
```

Two students flip coins and record the number of heads. They are supposed to flip 200 coins. The first student flips 200 coins, but the second student flips 100 coins and then doubles the result.

a) Find the EV and SE for the number of heads that the first student gets.

```
ev = 200 * 0.5
se = sqrt(200) * (1-0)*sqrt(0.5*0.5)
ev

## [1] 100

se

## [1] 7.071068
```

b) Find the EV and SE for what the second student gets after doubling.

```
evDoubled = 2 * (100 * 0.5)
seDoubled = 2 * (sqrt(100) * (1-0)*sqrt(0.5*0.5))
evDoubled

## [1] 100

seDoubled

## [1] 10
```

c) Find the EV and SE for what the two students get if they add their results.

```
ev = 200 * 0.5
se = sqrt(200) * (1-0)*sqrt(0.5*0.5)
evDoubled = 2 * (100 * 0.5)
seDoubled = 2 * (sqrt(100) * (1-0)*sqrt(0.5*0.5))
evTotal = ev + evDoubled
seTotal = se + seDoubled
evTotal

## [1] 200

seTotal

## [1] 17.07107
```