>title: "Homework 2"

author: "Hertzbert Casseus"

date: "9/29/2020"

output: html_document

>Group: Stan, Neshma & Fareha

>PP1: After running code for a dice roll, we "rolled" a "1". Which would mean the dice is fair.

dice = c("1", "2", "3", "4", "5", "6")

sample(x = dice, size = 1, replace = TRUE)

[1] "1"

>Next, I will roll my tampered dice and see if my roll is consistent with the code.

>I rolled a "5" initially with my tampered dice, thus this is consistent with the fair dice portion of the protocol.

As the dice is fair, the probability of it being judged unfair is 1/6. As there is only one event in which makes the dice unfair (rolling a 6).

>If the dice were unfair, the probability of them being deemed fair is still a 5/6, because on an initial roll (fair or not), the dice is 6 sided. One event makes it unfair (rolling a 6). Thus, rolling 1-5 on a six-sided die gives a probability of 5/6.

>PP2: When rolling my tampered die, the distribution of events are as follows: 1:1, 2:6, 3:6, 4:4, 5:6, 6:7.

(should be read "number on die:outcomes")

>Rule to judge die's fairness is preliminary 2 rolls of the dice cannot equal 6. Observed in tampered die [2:6]. For the PP, the dice is fair.

>When we run code in R, the values follow rule for judging fairness. If a fair die is rolled 30 times, the expected value of getting a 6 would be 6/30 theoretically, but because of randomness, it is perhaps more common that there will be events in which a 6 can be rolled more or fewer times than expected.

```
dice = c("1", "2", "3", "4", "5", "6")

 sample(x = dice, size = 30, replace = TRUE)
 [1] "4" "4" "3" "6" "3" "5" "3" "6" "3" "1" "1" "6" "3" "2" "1" "3" "4" "1" "4" "1" "4" "3" "4" "6" "6" "1"
"6" "3" "5"
[30] "3"
 outcomes <- c(4,4,3,6,3,5,3,6,3,1,1,6,3,2,1,3,4,1,4,1,4,3,4,6,6,1,6,3,5,3)
 count1 <- length(which(outcomes == 1))
 count1
[1] 6
 count2 <- length(which(outcomes == 2))
 count2
[1] 1
 count3 <- length(which(outcomes == 3))
 count3
[1] 9
 count4 <- length(which(outcomes == 4))
 count4
[1] 6
 count5 <- length(which(outcomes == 5))
 count5
[1] 2
 count6 <- length(which(outcomes == 6))
 count6
[1] 6
```

>The chances of the dice being judged unfair is if any value appears at an alarmingly higher rate that other values. Perhaps one can say if any value appears at 50% of the time, or greater, the dice is unfair.In this scenario, we see that our value that has the highest percentage is 3 at at just about 30% of this distribution.

>PP3: Since our sample size is now greater, we can expect the number of 6's in this sample to be closer to the theoretical probability of 16.7%, which would give us 16-17 number 6's in the sample.

```
 sample(x = dice, size = 100, replace = TRUE)

  [1] "6" "2" "5" "3" "5" "2" "4" "1" "2" "3" "2" "6" "1" "2" "5" "1" "1" "2" "4" "6" "3" "5" "5" "1" "2" "5"
"4" "1" "4"

 [30] "1" "6" "1" "4" "1" "5" "2" "4" "6" "3" "6" "2" "1" "5" "2" "3" "1" "1" "2" "3" "3" "2" "4" "2" "4" "1"
"2" "1" "5"

 [59] "4" "1" "5" "6" "6" "1" "1" "1" "3" "4" "6" "3" "2" "1" "6" "5" "6" "4" "5" "4" "3" "6" "6" "5" "3" "3"
"5" "4" "4"

 [88] "4" "5" "5" "6" "1" "5" "4" "3" "4" "4" "1" "4" "6"

 outcomes <-
c(6,2,5,3,5,2,4,1,2,3,2,6,1,2,5,1,1,2,4,6,3,5,5,1,2,5,4,1,4,1,6,1,4,1,5,2,4,6,3,6,2,1,5,2,3,1,1,2,3,3,2,4,2,4,1,
2,1,5,4,1,5,6,6,1,1,1,3,4,6,3,2,1,6,5,6,4,5,4,3,6,6,5,3,3,5,4,4,4,5,5,6,1,5,4,3,4,4,1,4,6)

 count1 <- length(which(outcomes == 1))

 count1

[1] 21

 count2 <- length(which(outcomes == 2))

 count2

[1] 15

 count3 <- length(which(outcomes == 3))

 count3

[1] 13

 count4 <- length(which(outcomes == 4))

 count4

[1] 19
```

```
count5 <- length(which(outcomes == 5))
count5
[1] 17
count6 <- length(which(outcomes == 6))
count6
[1] 15
a <- table(outcomes)
as.data.frame(table(outcomes))
  outcomes Freq
1        1   21
2        2   15
3        3   13
4        4   19
5        5   17
6        6   15
hist(outcomes)
```

This proves the fairness of the dice, as no single value is appearing at a rate significantly greater than the others.

EP1: "Rolling" a single 100 times, 5 times, essentially with replacement as the rolls are independent of each other. recording the experimental probability of rolling any given number for each sample.

We will then take the average of the experimental probability of the 5 separate trials. For the average of the population, we are willing to accept no more than a 7.5% difference. Meaning that, the average of the experimental probability of all 5 samples must be between 15.46% and 17.95%, or anything between 77 and 90 for any given value on the die's appearances in a total of 500 rolls.

```
dice = c("1", "2", "3", "4", "5", "6")
```

Outcome 1

```
sample(x = dice, size = 100, replace = TRUE)

 [1] "1" "2" "2" "3" "1" "2" "5" "4" "1" "5" "2" "1" "2" "1" "4" "1" "2" "2" "3" "2" "2" "2" "4" "5" "6" "1"
"2" "1" "2"

 [30] "5" "4" "6" "5" "1" "4" "2" "1" "1" "1" "2" "3" "5" "5" "1" "5" "2" "2" "2" "4" "4" "3" "5" "2" "1" "4"
"1" "2" "4"

 [59] "4" "1" "6" "3" "3" "5" "3" "4" "3" "3" "4" "1" "6" "6" "6" "1" "1" "3" "4" "6" "4" "4" "1" "5" "6" "3"
"4" "6" "1"

 [88] "5" "1" "4" "4" "1" "6" "5" "6" "5" "2" "6" "2" "2"

 outcomes1 <-
c(1,2,2,3,1,2,5,4,1,5,2,1,2,1,4,1,2,2,3,2,2,2,4,5,6,1,2,1,2,5,4,6,5,1,4,2,1,1,1,2,3,5,5,1,5,2,2,2,4,4,3,5,2,1,4,
1,2,4,4,1,6,3,3,5,3,4,3,3,4,1,6,6,6,1,1,3,4,6,4,4,1,5,6,3,4,6,1,5,1,4,4,1,6,5,6,5,2,6,2,2)

 a <- table(outcomes1)

 as.data.frame(table(outcomes1))

  outcomes1 Freq
1         1   23
2         2   22
3         3   11
4         4   18
5         5   14
6         6   12

 v1 <- as.data.frame(table(outcomes1))

 sample.frequency1 <- v1 [ ,"Freq"]

 Ttest1 <- t.test(sample.frequency1, mu= 0 )

 View(Ttest1)
```

Outcome 2

```
 sample(x = dice, size = 100, replace = TRUE)

  [1] "1" "1" "6" "5" "6" "1" "5" "1" "2" "4" "4" "6" "3" "3" "4" "2" "5" "6" "5" "6" "2" "3" "6" "1" "1" "5"
"4" "5" "3"

 [30] "1" "6" "1" "3" "2" "5" "2" "4" "6" "2" "2" "1" "3" "4" "4" "2" "6" "4" "1" "4" "5" "2" "5" "4" "6" "1"
"5" "4" "4"
```

[59] "1" "6" "1" "4" "4" "5" "4" "5" "1" "1" "4" "2" "6" "5" "3" "5" "3" "5" "3" "1" "1" "4" "6" "4" "6" "4" "3" "6" "6"

[88] "3" "6" "2" "4" "2" "3" "1" "1" "6" "1" "2" "4" "1"

outcomes2 <- c(1,1,6,5,6,1,5,1,2,4,4,6,3,3,4,2,5,6,5,6,2,3,6,1,1,5,4,5,3,1,6,1,3,2,5,2,4,6,2,2,1,3,4,4,2,6,4,1,4,5,2,5,4,6,1,5,4,4,1,6,1,4,4,5,4,5,1,1,4,2,6,5,3,5,3,5,3,1,1,4,6,4,6,4,3,6,6,3,6,2,4,2,3,1,1,6,1,2,4,1)

a <- table(outcomes2)

as.data.frame(table(outcomes2))

outcomes2 Freq

1 1 21

2 2 13

3 3 12

4 4 21

5 5 15

6 6 18

v2 <- as.data.frame(table(outcomes2))

View(v2)

sample.frequency2 <- v2 [,"Freq"]

Ttest2 <- t.test(sample.frequency2, mu= 0)

Outcome 3

sample(x = dice, size = 100, replace = TRUE)

[1] "6" "2" "3" "6" "4" "4" "1" "6" "1" "6" "5" "5" "1" "3" "6" "4" "1" "5" "4" "6" "2" "3" "6" "3" "2" "4" "2" "6" "2"

[30] "3" "3" "2" "5" "4" "4" "5" "5" "2" "4" "6" "3" "1" "5" "1" "1" "5" "3" "3" "2" "1" "3" "1" "3" "5" "6" "1" "1" "6"

[59] "5" "5" "2" "5" "1" "1" "4" "4" "4" "4" "5" "1" "5" "6" "1" "5" "5" "3" "1" "2" "6" "4" "6" "2" "2" "5" "6" "4" "2"

[88] "3" "4" "2" "1" "3" "5" "6" "1" "4" "4" "5" "1" "5"

outcomes3 <- c(6,2,3,6,4,4,1,6,1,6,5,5,1,3,6,4,1,5,4,6,2,3,6,3,2,4,2,6,2,3,3,2,5,4,4,5,5,2,4,6,3,1,5,1,1,5,3,3,2,1,3,1,3,5,6,1,1,6,5,5,2,5,1,1,4,4,4,4,5,1,5,6,1,5,5,3,1,2,6,4,6,2,2,5,6,4,2,3,4,2,1,3,5,6,1,4,4,5,1,5)

```
a <- table(outcomes3)

as.data.frame(table(outcomes3))

 outcomes3 Freq
1        1   19
2        2   14
3        3   14
4        4   17
5        5   20
6        6   16

v3 <- as.data.frame(table(outcomes3))

sample.frequency3 <- v3 [ ,"Freq"]

Ttest3 <- t.test(sample.frequency3, mu= 0 )

View(Ttest2)

View(Ttest2)

View(Ttest3)
```

Outcome 4

```
sample(x = dice, size = 100, replace = TRUE)

 [1] "3" "1" "6" "6" "2" "2" "6" "5" "3" "4" "2" "2" "1" "5" "6" "6" "2" "3" "1" "2" "6" "1" "3" "2" "6" "1"
"5" "3" "2"

 [30] "3" "2" "2" "1" "3" "4" "4" "2" "1" "2" "3" "2" "1" "5" "4" "1" "1" "2" "6" "2" "2" "4" "6" "5" "5" "2"
"5" "2" "2"

 [59] "2" "2" "5" "1" "4" "1" "2" "1" "6" "4" "6" "3" "6" "5" "2" "1" "6" "1" "6" "4" "6" "5" "4" "6" "2" "4"
"4" "5" "4"

 [88] "2" "6" "4" "3" "1" "3" "4" "6" "3" "6" "4" "4" "4"

outcomes4 <-
c(3,1,6,6,2,2,6,5,3,4,2,2,1,5,6,6,2,3,1,2,6,1,3,2,6,1,5,3,2,3,2,2,1,3,4,4,2,1,2,3,2,1,5,4,1,1,2,6,2,2,4,6,5,5,2,
5,2,2,2,2,5,1,4,1,2,1,6,4,6,3,6,5,2,1,6,1,6,4,6,5,4,6,2,4,4,5,4,2,6,4,3,1,3,4,6,3,6,4,4,4)

a <- table(outcomes4)

as.data.frame(table(outcomes4))
```

outcomes4 Freq

1 1 16

2 2 25

3 3 12

4 4 17

5 5 11

6 6 19

```
v4 <- as.data.frame(table(outcomes4))

sample.frequency4 <- v4 [ ,"Freq"]

Ttest4 <- t.test(sample.frequency4, mu= 0 )
```

Outcome 5

```
sample(x = dice, size = 100, replace = TRUE)

 [1] "2" "1" "1" "5" "2" "3" "3" "4" "1" "1" "6" "1" "3" "6" "3" "4" "6" "3" "4" "5" "6" "3" "5" "6" "4" "5"
"6" "5" "3"

 [30] "2" "4" "4" "5" "5" "4" "4" "5" "3" "3" "5" "6" "4" "4" "3" "3" "1" "6" "5" "2" "4" "3" "2" "4" "1" "2"
"3" "5" "1"

 [59] "4" "3" "2" "1" "1" "4" "6" "5" "6" "3" "2" "4" "3" "4" "5" "2" "6" "3" "4" "1" "5" "6" "2" "2" "3" "1"
"4" "4" "2"

 [88] "3" "2" "5" "3" "5" "2" "3" "6" "6" "2" "2" "6" "3"

 outcomes5 <-
c(2,1,1,5,2,3,3,4,1,1,6,1,3,6,3,4,6,3,4,5,6,3,5,6,4,5,6,5,3,2,4,4,5,5,4,4,5,3,3,5,6,4,4,3,3,1,6,5,2,4,3,2,4,1,2,
3,5,1,4,3,2,1,1,4,6,5,6,3,2,4,3,4,5,2,6,3,4,1,5,6,2,2,3,1,4,4,2,3,2,5,3,5,2,3,6,6,2,2,6,3)

 a <- table(outcomes5)

 as.data.frame(table(outcomes5))
```

outcomes5 Freq

1 1 12

2 2 16

3 3 22

4 4 19

```
5       5  16
6       6  15
```

```
v5 <- as.data.frame(table(outcomes5))
sample.frequency5 <- v5 [ ,"Freq"]
Ttest5 <- t.test(sample.frequency5, mu= 0 )
```

```
PopulationData <- cbind(v1, v2, v3, v4, v5)
colnames(PopulationData )[2]<-"outcomes1_Freq"
colnames(PopulationData )[4]<-"outcomes2_Freq"
colnames(PopulationData )[6]<-"outcomes3_Freq"
colnames(PopulationData )[8]<-"outcomes4_Freq"
colnames(PopulationData )[10]<-"outcomes5_Freq"
PopulationData <- PopulationData[, !duplicated(colnames(PopulationData), fromFirst = TRUE)]
print(PopulationData)
```

```
  outcomes1 outcomes1_Freq outcomes2 outcomes2_Freq outcomes3 outcomes3_Freq outcomes4
outcomes4_Freq outcomes5
1       1         23       1         21       1         19       1         16       1
2       2         22       2         13       2         14       2         25       2
3       3         11       3         12       3         14       3         12       3
4       4         18       4         21       4         17       4         17       4
5       5         14       5         15       5         20       5         11       5
6       6         12       6         18       6         16       6         19       6
  outcomes5_Freq
1          12
2          16
3          22
4          19
5          16
```

6 15

summary(PopulationData)

outcomes1 outcomes1_Freq outcomes2 outcomes2_Freq outcomes3 outcomes3_Freq outcomes4 outcomes4_Freq outcomes5

1:1 Min. :11.00 1:1 Min. :12.00 1:1 Min. :14.00 1:1 Min. :11.00 1:1

2:1 1st Qu.:12.50 2:1 1st Qu.:13.50 2:1 1st Qu.:14.50 2:1 1st Qu.:13.00 2:1

3:1 Median :16.00 3:1 Median :16.50 3:1 Median :16.50 3:1 Median :16.50 3:1

4:1 Mean :16.67 4:1 Mean :16.67 4:1 Mean :16.67 4:1 Mean :16.67 4:1

5:1 3rd Qu.:21.00 5:1 3rd Qu.:20.25 5:1 3rd Qu.:18.50 5:1 3rd Qu.:18.50 5:1

6:1 Max. :23.00 6:1 Max. :21.00 6:1 Max. :20.00 6:1 Max. :25.00 6:1

outcomes5_Freq

Min. :12.00

1st Qu.:15.25

Median :16.00

Mean :16.67

3rd Qu.:18.25

Max. :22.00

PopulationData$mean <- rowMeans(subset(PopulationData, select = c(outcomes1_Freq, outcomes2_Freq, outcomes3_Freq, outcomes4_Freq, outcomes5_Freq)), na.rm = TRUE)

print.listof(PopulationData$mean)

Component 1 :

[1] 18.2

Component 2 :

[1] 18

Component 3 :

[1] 14.2

Component 4 :

[1] 18.4

Component 5 :

[1] 15.2

Component 6 :

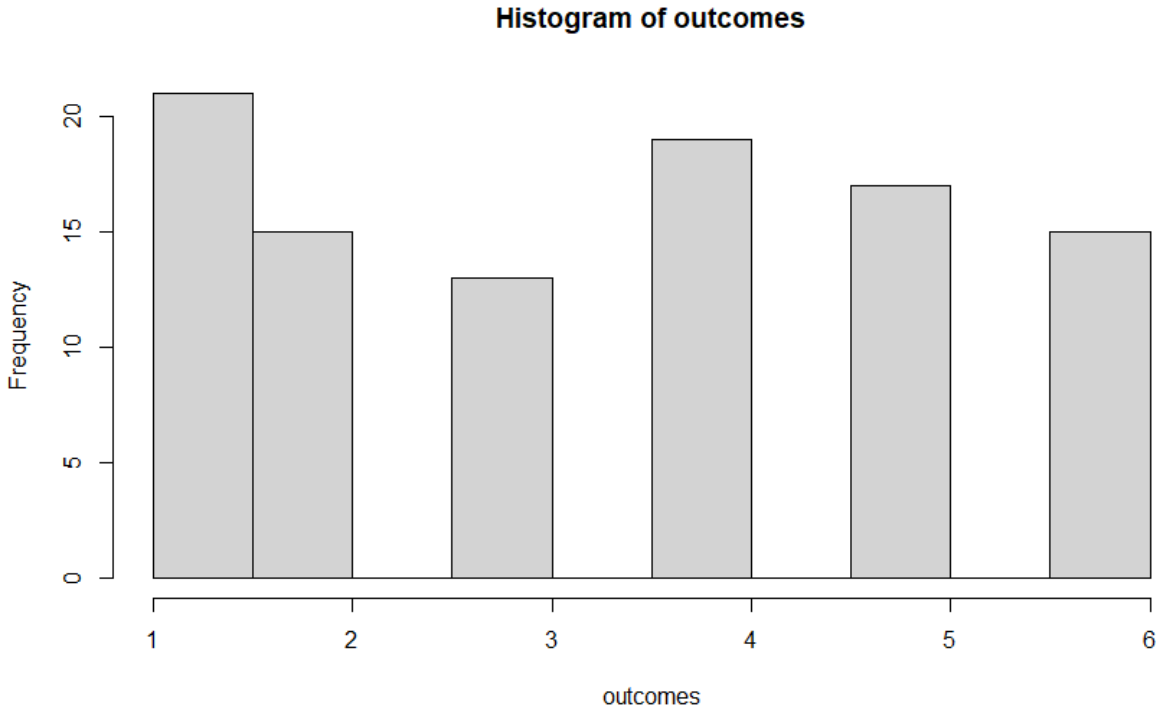
[1] 16

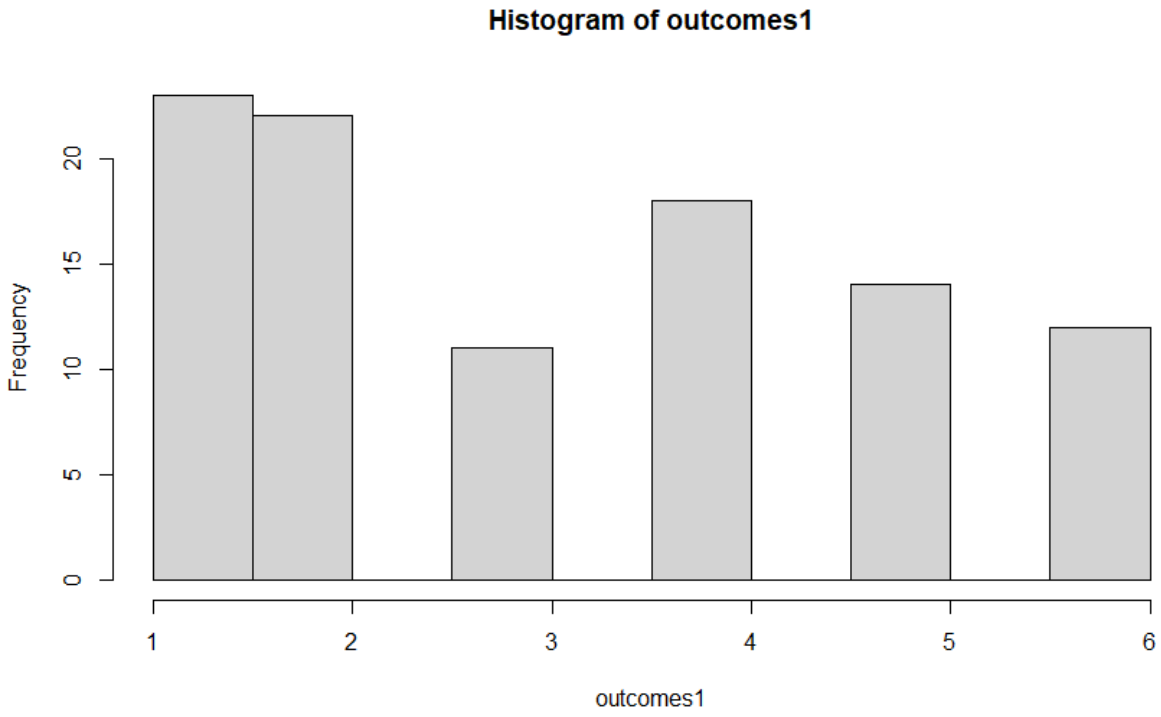
>Probability theory would lead us to believe, for this experiment in particular, that any given number on a die has a 16.7% chance of success and the more trials of rolling a die, the successes average will be closer to that figure.

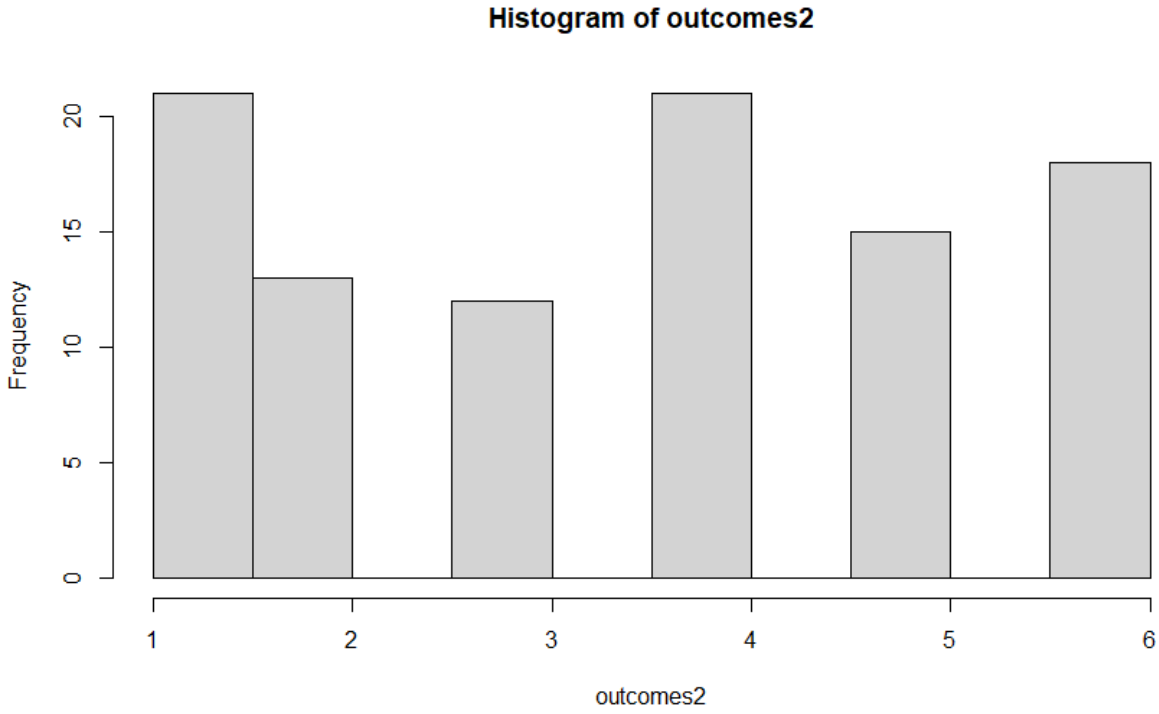
In my experiment, with a 92.5% confidence interval, and 500 trails, it seems tha while my averages were close to the parameters I set, only one value fell into said parameters (the number 6). This could be due to an insufficient amount of trials or the truly unusual confidence interval I chose (I hypothesize that I would have had more success with a 90%).

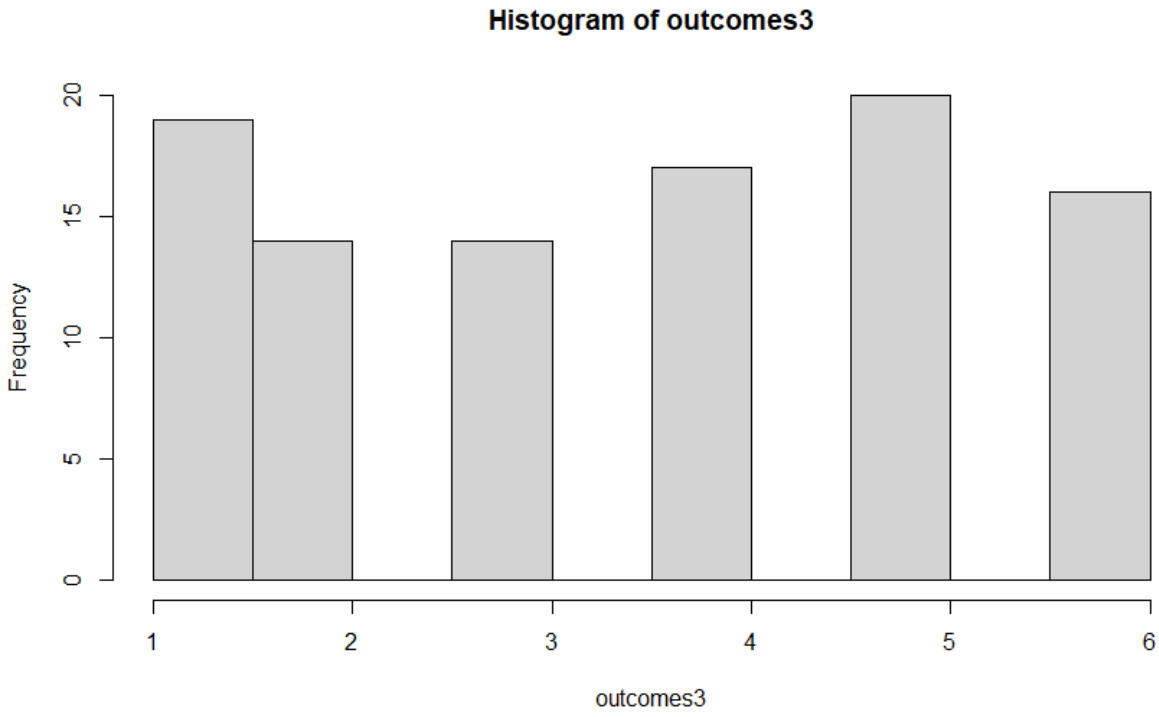
cbind(Ttest1, Ttest2, Ttest3, Ttest4, Ttest5)

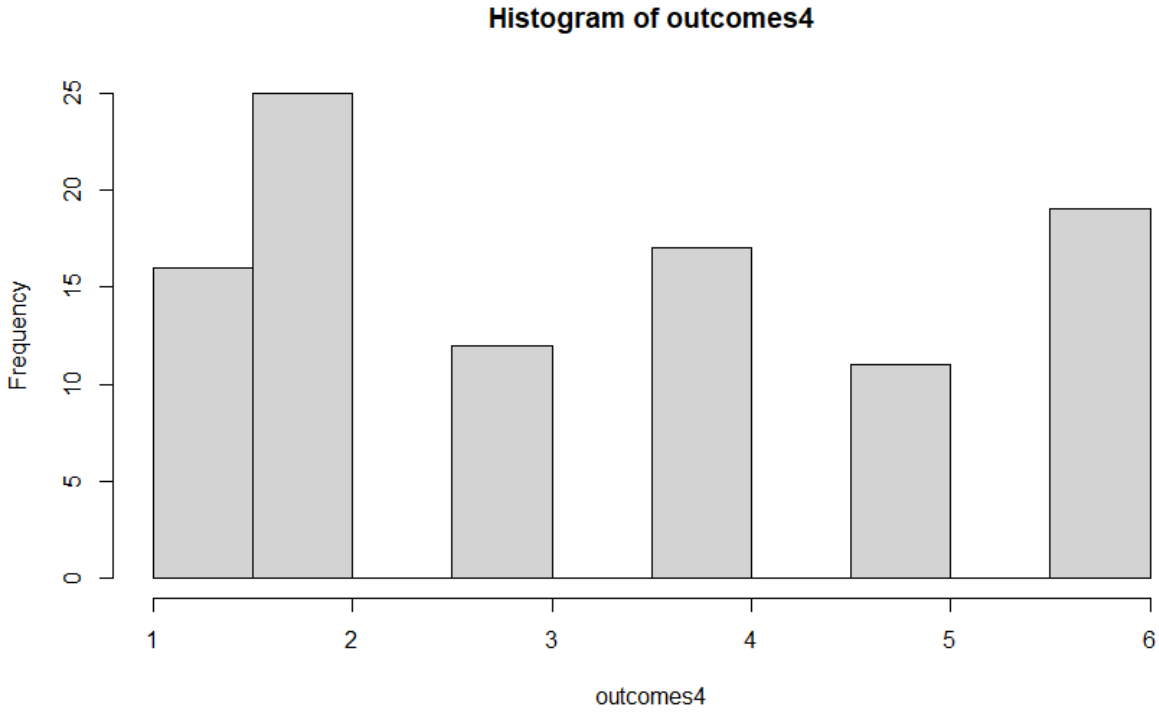
	Ttest1	Ttest2	Ttest3	Ttest4	Ttest5
statistic	7.965662	10.38068	16.3082	8.027016	11.85114
parameter	5	5	5	5	5
p.value	0.0005029459	0.0001428677	1.581011e-05	0.0004851749	7.532365e-05
conf.int	Numeric,2	Numeric,2	Numeric,2	Numeric,2	Numeric,2
estimate	16.66667	16.66667	16.66667	16.66667	16.66667
null.value	0	0	0	0	0
stderr	2.092314	1.605546	1.021981	2.076322	1.406335
alternative	"two.sided"	"two.sided"	"two.sided"	"two.sided"	"two.sided"
method	"One Sample t-test"	"One Sample t-test"	"One Sample t-test"	"One Sample t-test"	"One Sample t-test"
data.name	"sample.frequency1"	"sample.frequency2"	"sample.frequency3"	"sample.frequency4"	"sample.frequency5"

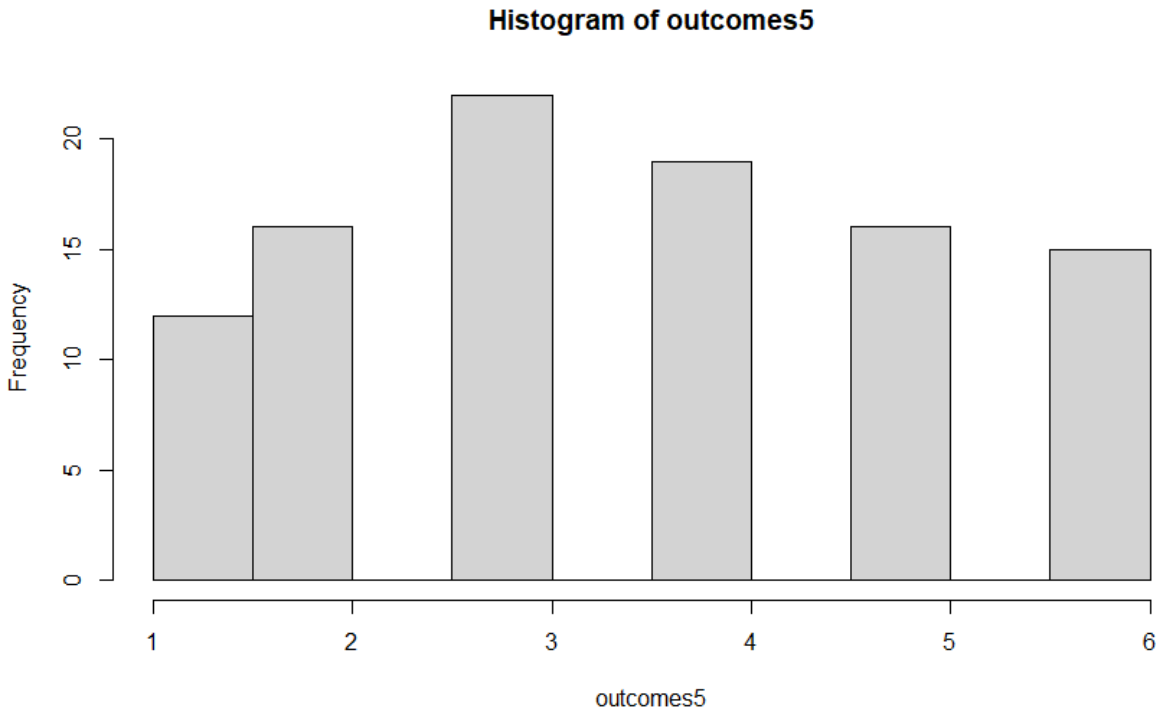

Histogram of outcomes
Frequency
0
5
10
15
20
1
2
3
4
5
6
outcomes


Histogram of outcomes1
Frequency
0
5
10
15
20
1
2
3
4
5
6
outcomes1


Histogram of outcomes2
Frequency
0
5
10
15
20
1
2
3
4
5
6
outcomes2


Histogram of outcomes3
Frequency
0
5
10
15
20
1
2
3
4
5
6
outcomes3


Histogram of outcomes4
Frequency
0
5
10
15
20
25
1
2
3
4
5
6
outcomes4


Histogram of outcomes5
Frequency
0
5
10
15
20
1
2
3
4
5
6
outcomes5