---

title: "Homework#2"

author: "Neshma Simon"

date: "9/28/2020"

output: html\_document

---

```{r setup, include=FALSE}

knitr::opts\_chunk$set(echo = TRUE)

```

## R Markdown

## Lab #1

```

Group members: Fareha, Stan, and Hertz

```

```

Possible Protocol 1 (PP1): roll once; if get 6 then conclude the dice is not fair; if roll any other number then conclude it is fair. Analyze PP1: if the dice were fair, what is the probability it would be judged to be unfair? Oppositely, if the dice were unfair, what is the probability that it would be judged to be fair?

-Once we rolled the dice and ran the code, we "rolled" a "1". This would indicate that the dice is fair.

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

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

a<-sample(1:6,size=1,replace = T)

print(a)

table(a)

[1] "1"

Then, we rolled the tampered dice to see if the roll is consistent with the code.

We rolled a "5" initially with the tampered dice, which makes this consistent with the fair dice portion of the protocol.

As the dice is fair, the probability of it being judged as unfair is 1/6 or 16.67%. Rolling a six is the only way for it to be unfair.

Depending on if the dice was unfair, the probability of them being deemed as fair is still 5/6 or 83.3%, because on an initial roll, regardless of its fairness, the dice is 6 sided. One event makes it unfair (rolling a 6). Thus, rolling 1,2,3,4, or 5 on a 6-sided dice gives a probability of 5/6 or 83.3%.

```

```

PP2: roll the dice 30 times. Group can specify a decision rule to judge that dice is fair or unfair. Consider the stats question: if fair dice are rolled 30 times, what is likely number of 6 resulting? How unusual is it, to get 1 more or less than that? How unusual is it, to get 2 more or less? 3? At least one member of the group should be able to do this with theory; at least one member of the group should be able to write a little program in R to simulate this. Analyze PP2 including the question: if the dice were fair, what is the chance it could be judged as unfair?

-As we rolled the tampered dice, we had a range of different outcomes. The format for the distribution of events are written as "number on die:outcomes". The results for this was 1:1, 2:6, 3:6, 4:4, 5:6, 6:7.

As seen in the tampered dice with the outcome of 2:6, two rolls cannot equal a six which will prove its fairness. 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 or 20% theoretically. Due to unarranged 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)

a<-sample(1:6,size=30,replace = T)

print(a)

table(a)

1 2 3 4 5 6

1 6 7 5 9 2

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

[20] "5" "5" "5" "2" "5" "5" "2" "2" "5" "6" "3"

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

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

1

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

6

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

7

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

5

count5 <- length(which(outcomes == 5))

9

count6 <- length(which(outcomes == 6))

2

If the dice is being judged unfairly,it must be because any value appears at an alarmingly higher rate that other values. For instance, an individual can say if that any value appears at around 50%+ of the time, the dice is unfair. In this scenario, we see that our value that has the highest percentage is 5 at at just about 30% of this distribution. Since it is less than the 50% or more, I can predict that this is fair.

```

```

PP3: roll 100 times and specify decision rules. Some cases are easy: if every roll comes to 6 then might quickly conclude. But what about the edge cases? Is it fair to say that every conclusion has some level of confidence attached? Where do you set boundaries for decisions? Analyze PP3.

-From what we believe,our sample size is now much greater from 30, so we can expect the number of 6's in this sample to be closer to the theoretical probability of 16.67%, which would give us about 16-17 number 6's in the sample.

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

a<-sample(1:6,size=100,replace = T)

print(a)

table(a)

1 2 3 4 5 6

12 17 23 16 15 17

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

[20] "4" "5" "2" "2" "4" "2" "1" "3" "6" "1" "6" "4" "1" "6" "5" "5" "3" "3" "5"

[39] "1" "4" "5" "2" "4" "1" "1" "5" "5" "2" "2" "4" "6" "3" "3" "3" "4" "6" "6"

[58] "4" "2" "6" "6" "2" "6" "3" "3" "3" "6" "1" "5" "5" "3" "6" "1" "3" "3" "1"

[77] "6" "5" "1" "1" "2" "5" "2" "2" "5" "2" "3" "6" "4" "5" "2" "2" "3" "4" "6"

[96] "3" "5" "2" "3" "2"

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

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

12

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

17

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

23

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

16

count5 <- length(which(outcomes == 5))

15

count6 <- length(which(outcomes == 6))

17

Ultimately, this proves the fairness of the dice where not one single number is appearing at a rate significantly greater than the others, it truly varies according to randomness.

```

```

EP1: What is a reasonable number of times to roll your experiment dice? (given time available in class, etc) If you roll the experiment dice a certain number of times and see a particular outcome, then you would conclude it is fair or not. How confident would you be, in any of those conclusions? Note that while previous Possible Protocols emphasized counting just the number of times a 6 came up, you might consider other outcomes such as the average value of the dice when rolled or the distribution of other outcomes (what number is on opposite face from 6? Do you suppose that number might be disproportionately represented if dice were not fair?).

-"Rolling" a single 100 times for 5 rounds, (500 times) with replacements as the rolls are independent of each other. While recording the experimental probability of rolling any given number for each sample, we will be able to tell the fairness.

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 10% difference. Meaning that, the average of the experimental probability of all 5 samples must be between 15.03% and 18.37%, or anything between 75 and 92 for any given value on the die's appearances in a total of 500 rolls.

sample(x = dice, size = 500, replace = TRUE)

a<-sample(1:6,size=500,replace = T)

print(a)

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

[40] 5 6 6 5 3 5 2 1 3 2 4 5 3 6 3 4 3 3 1 2 1 5 5 2 5 5 3 2 1 4 1 1 5 1 5 6 3 4 3

[79] 1 4 3 5 1 4 4 1 5 5 3 3 4 6 2 5 4 5 6 4 1 2 5 4 6 2 1 6 4 1 1 3 4 1 6 4 5 5 3

[118] 4 3 2 5 4 2 3 4 2 5 6 6 3 3 4 1 5 2 3 3 3 4 5 5 1 5 5 5 5 4 2 3 2 3 6 2 6 2 3

[157] 4 6 3 3 2 3 4 4 2 2 2 1 1 5 2 4 3 5 5 4 6 2 2 5 2 2 2 4 5 5 5 3 5 2 1 1 3 4 6

[196] 2 3 5 2 6 6 4 2 1 5 2 2 1 4 3 3 2 5 2 6 3 6 3 4 3 4 6 3 3 2 2 3 5 4 3 3 6 3 2

[235] 1 6 3 2 1 2 3 3 1 3 6 5 1 3 6 1 3 2 6 4 1 1 3 1 5 4 4 3 2 1 1 3 3 1 6 1 1 3 1

[274] 3 4 2 6 2 2 1 2 5 2 1 3 2 5 5 2 4 2 2 6 2 3 2 4 3 1 6 1 5 5 4 4 1 1 5 2 3 5 2

[313] 1 2 3 2 1 4 1 3 3 2 1 6 4 5 6 2 4 6 2 4 2 1 3 5 6 5 3 3 2 2 6 2 5 2 3 4 5 6 2

[352] 2 6 4 6 6 2 4 5 1 1 5 3 4 1 1 3 5 3 1 1 5 5 1 1 1 4 4 4 3 2 5 6 5 3 6 4 6 6 5

[391] 5 5 2 2 5 5 3 5 1 5 4 2 3 3 1 3 2 2 4 2 2 1 4 5 3 5 2 2 2 5 6 2 4 2 4 2 2 5 2

[430] 4 3 4 1 4 2 1 4 5 1 4 6 2 5 6 5 6 4 3 6 6 6 4 4 5 2 1 2 6 3 1 5 5 5 5 3 6 1 2

[469] 3 6 6 2 6 6 5 3 1 6 4 4 4 3 4 2 5 2 1 2 6 3 4 5 1 4 5 2 6 5 1 6

table(a)

1 2 3 4 5 6

75 99 91 76 93 66

library(ggplot2)

qplot(a,binwidth=1)

Graphical user interface, application

Description automatically generated

mean(a)

[1] 3.422

mode(a)

[1] 2

median(a)

[1] 3

var(a)

[1] 2.709335

sqrt(var(a))

[1] 1.646006

summary(a)

Min. 1st Qu. Median Mean 3rd Qu. Max.

1.000 2.000 3.000 3.422 5.000 6.000

sd(a,na.rm=TRUE)

[1] 1.646006]

hist(a)

Graphical user interface, application

Description automatically generated

We see here that the average is 3.422 with a standard deivation of 1.646006. The fairness of the dice is based on not one single number compared to the frequency of other values. The randomness plays a large factor in its fariness.

```