

**title: "Homework#2, Lab#2, Econometrics B2000"**

**author: "Hassan Fayyaz"**

**date:"09/09/2021"**

**#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?

```
dice.roll1 <- sample(x = 1:6, size = 1 , replace = TRUE)
```

```
sample(x = 1:6, size = 1, replace = TRUE)
```

```
#Outcome: [1]"4"
```

**#Results of PP1 experiment:**

#For our experiment, we rolled a 4. Therefore, this concludes dice is fair. The probability of the dice being unfair in this scenario would be  $1/6$ , which is 0.167 and the probability of it being fair would be  $5/6$ , which is 0.83

#In other words: If the dice is rolled and we get a 6 then the probability of the dice being unfair is 16.67% & if we get any other number other than 6 then the dice is fair, and therefore the probability of it being fair is 83.33%.

**#PP2:** roll the dice 20 times. (Each person should have done this beforehand.) Group can specify a decision rule to judge that dice is fair or unfair. Consider the stats question: if fair dice are rolled 20 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? Analyze PP2 including the question: if the dice were fair, what is the chance it could be judged as unfair?

**#Null Hypothesis\*\*( $H_0$ )** - No relationship; the die is generating random numbers and thus is fair.

**#Alternative Hypothesis\*\*( $H_a$ )** - The die is unfair; it has been fixed to generate more 6s on average.

#We are not running the code below in this case because we do not want a random sample every time and would like to focus on the sample we got at the first time, so we wrote the sample in the script.

```
#outcomes <- sample(x = 1:6, size = 20, replace = TRUE)
```

```
#print(outcomes)
```

```
outcomes <- sample(x = 1:6, size = 20 , replace = TRUE)
```

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

```
print(outcomes)
```

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

```
#install.packages("ggplot2")
```

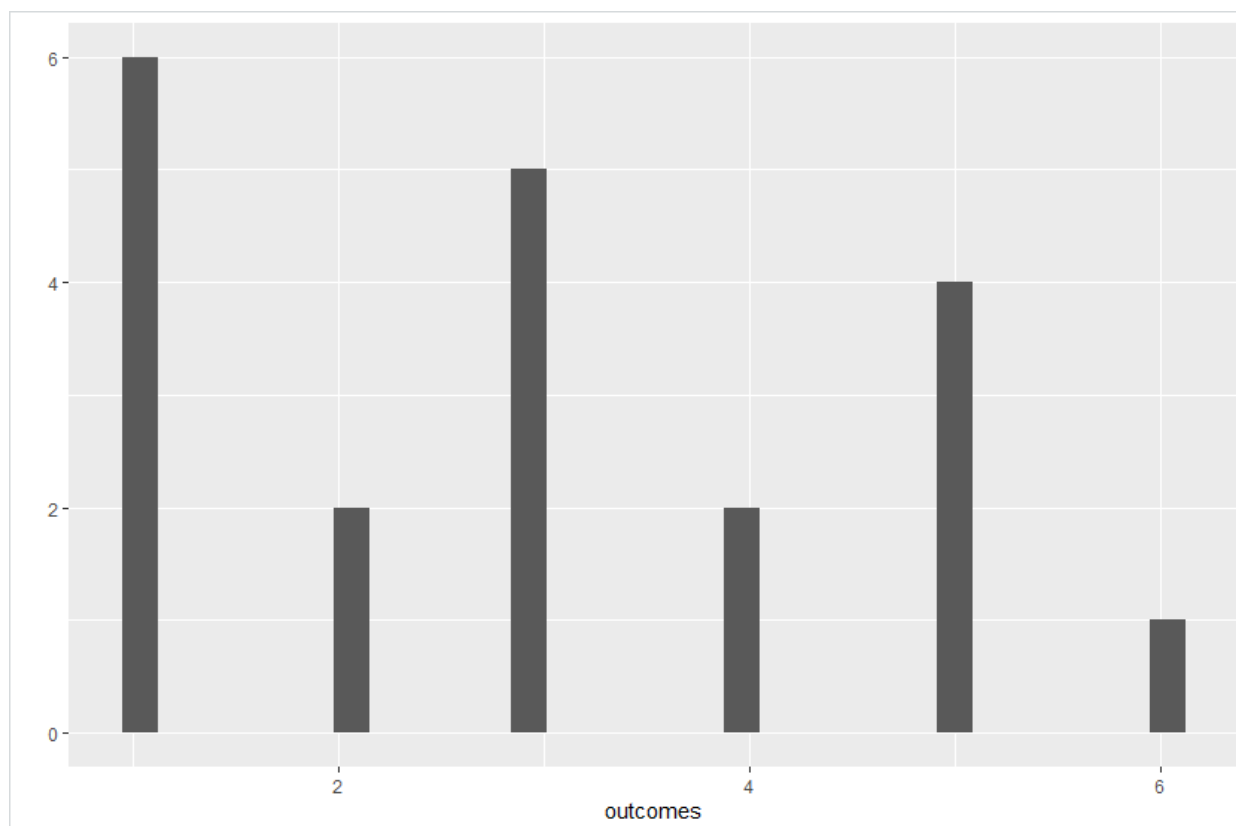
```
library(ggplot2)
```

```
qplot(outcomes, binwidth = 1)
```

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

```
#Warning message:
```

```
# Ignoring unknown parameters: binwidth
```



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

```
count1
```

```
# [1] 6
```

```
# 6/20
```

```
# Probability = 30%
```

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

```
count2
```

```
# [1] 2
```

```
# 2/20
```

```
# Probability = 10%
```

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

```
count3
```

```
# [1] 5
```

```
# 5/20
```

```
# Probability = 25%
```

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

```
count4
```

```
# [1] 2
```

```
# 2/20
```

```
# Probability = 10%
```

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

```
count5
```

```
# [1] 4
```

```
# 4/20
```

```
# Probability = 20%
```

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

```
count6
```

```
# [1] 1
```

```
# 1/20
```

```
# Probability = 5%
```

```
mean(outcomes)
```

```
# [1] 2.95
```

```
median(outcomes)
```

```
# [1] 3
```

```
var(outcomes)
```

```
#[1] 2.786842
```

```
sd(outcomes)
```

```
#[1] 1.669384
```

```
summary(outcomes)
```

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

```
# 1.00   1.00   3.00   2.95   4.25   6.00
```

```
chisq <- chisq.test(table(outcomes))
```

```
chisq
```

```
# Chi-squared test for given probabilities
```

```
# data: table(outcomes)
```

```
# X-squared = 5.8, df = 5, p-value = 0.3262
```

```
chisq$observed
```

```
# outcomes
```

```
# (1, 2, 3, 4, 5, 6)
```

```
# (6, 2, 5, 2, 4, 1)
```

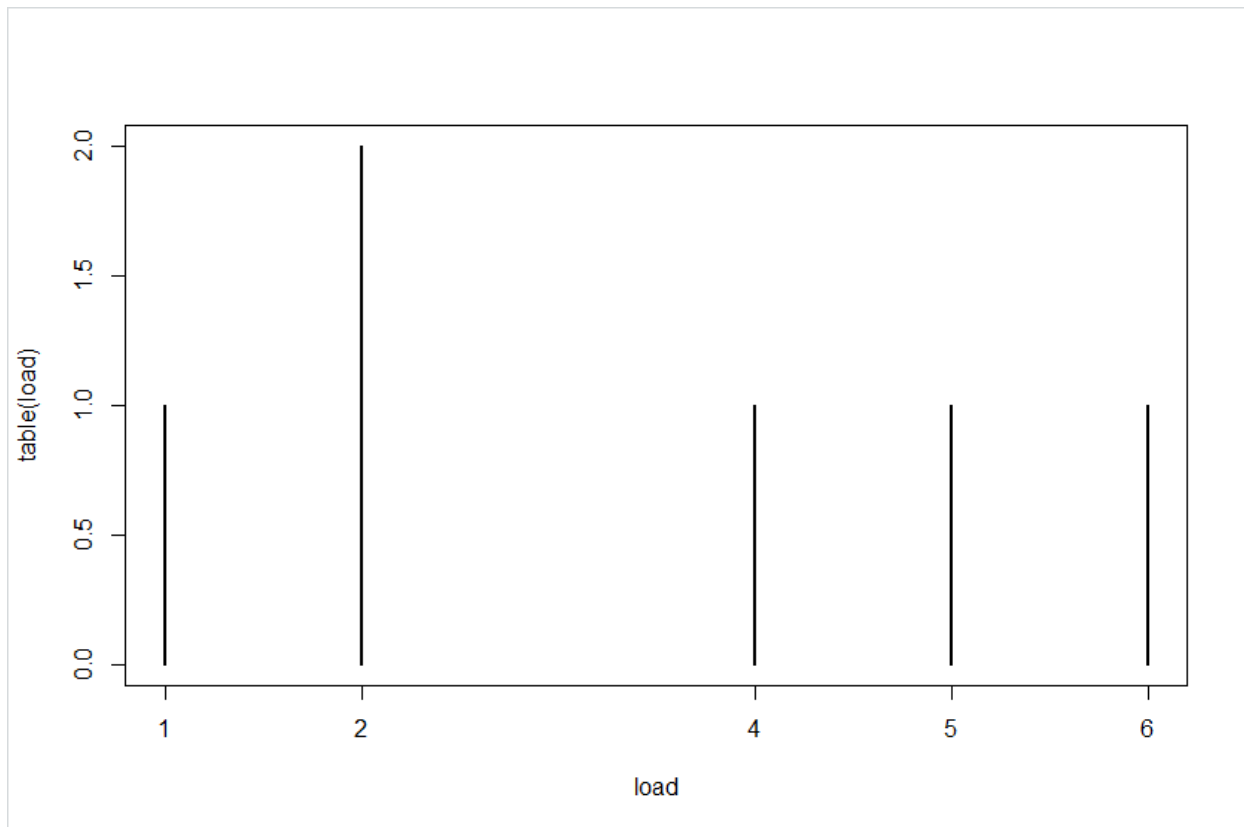
```
round(chisq$expected,2)
```

```
# 1  2  3  4  5  6
```

```
# 3.33 3.33 3.33 3.33 3.33 3.33
```

```
chisq$p.value  
#[1] 0.3261688
```

```
load <- c(6,2,5,2,4,1)  
plot(table(load))
```



### **#PP2\_CONCLUSION:**

#Here at a confidence level of 0.01, the T value is 2.86093

#degree of freedom =  $n-1$ , which is  $20-1 = 19$

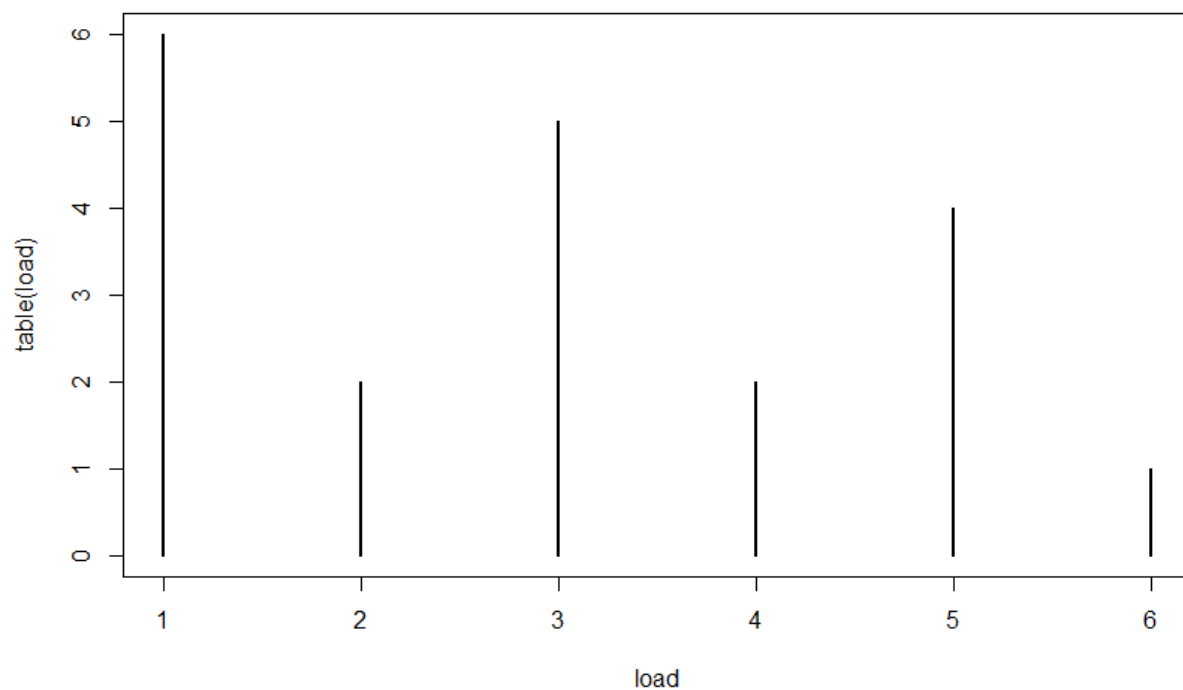
#Therefore in this case, the p value for a t statistics is 0.3262 and thus we fail to reject the null hypothesis as p value is greater than 0.01

#Thus the dice is fair

#The die in this case would be considered unfair if the p value is less than 0.01 and we reject the null hypothesis

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

```
plot(table(load))
```



**#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. What is the chance that fair dice could be judged to be unfair?

```
dice.roll100 <- sample(x = 1:6, size = 100 , replace = TRUE)
```

```
outcomes <- sample(x = 1:6, size = 100, replace = TRUE)
```

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

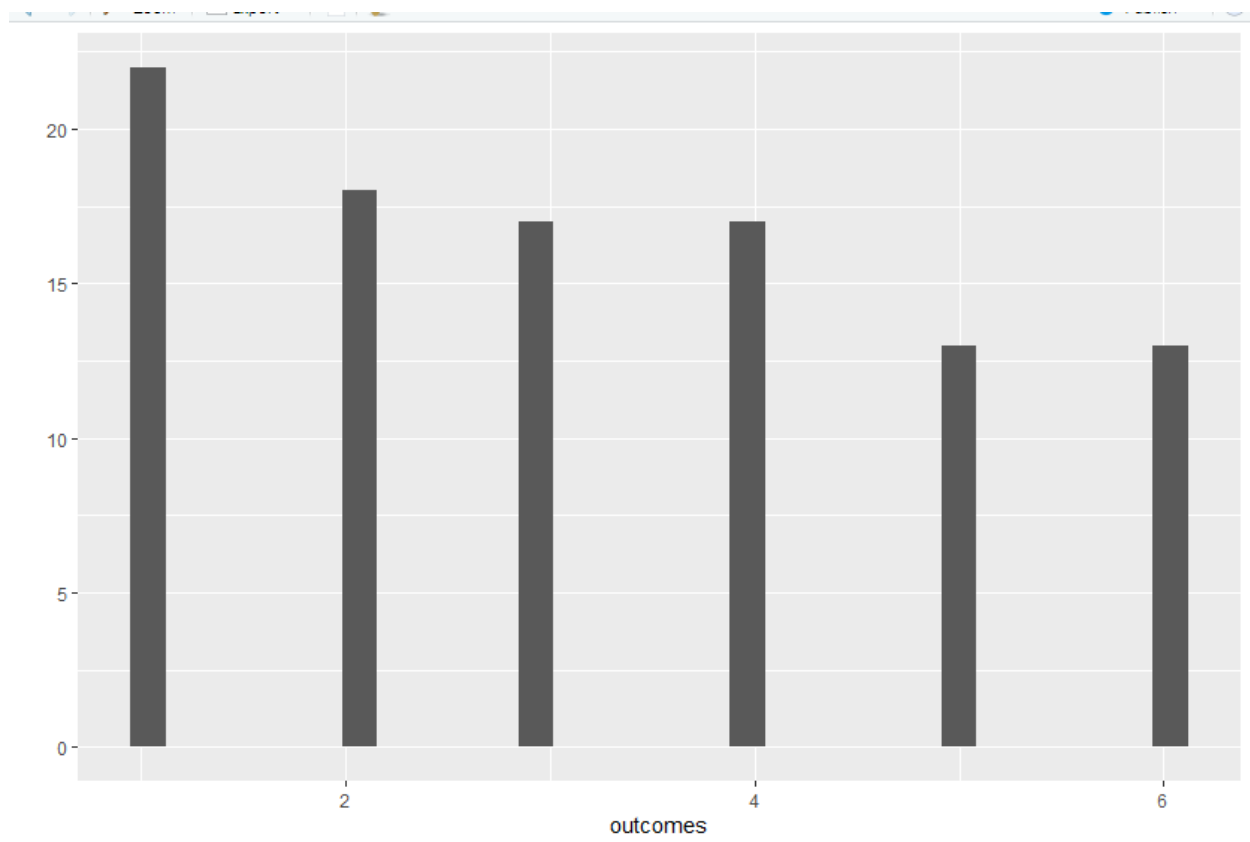
```
print(outcomes)
```

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

```
[61] 2 4 1 2 6 2 3 6 3 5 2 4 5 4 3 3 6 4 4 4 1 3 3 5 2 6 4 1 6 6 1 1 2 1 1 3 6 4 3 5'
```

```
library(ggplot2)
```

```
qplot(outcomes, binwidth = 1)
```



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

```
count1
```



```
# [1] 22
```

```
# 22/100
```

```
# Probability = 22%
```

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

```
count2
```

```
# [1] 18
```

```
# 18/100
```

```
# Probability = 18%
```

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

```
count3
```

```
# [1] 17
```

```
# 17/100
```

```
# Probability = 17%
```

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

```
count4
```

```
# [1] 17
```

```
# 17/100
```

```
# Probability = 17%
```

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

```
count5
```

```
# [1] 13
```

```
# 13/100
```

```
# Probability = 13%
```

```
count6 <- length(which(outcomes == 6))  
count6  
# [1] 13  
# 13/100  
# Probability = 13%  
  
mean(outcomes)  
#[1] 3.2  
  
median(outcomes)  
#[1] 3  
  
var(outcomes)  
#[1] 2.909091  
  
sd(outcomes)  
#[1] 1.705606  
  
summary(outcomes)  
# Min. 1st Qu. Median Mean 3rd Qu. Max.  
# 1.0 2.0 3.0 3.2 5.0 6.0  
  
chisq <- chisq.test(table(outcomes))  
chisq  
  
# Chi-squared test for given probabilities  
# data: table(outcomes)  
# X-squared = 3.44, df = 5, p-value = 0.6325
```

```
chisq$observed
```

```
#outcomes
```

```
# (1, 2, 3, 4, 5, 6)
```

```
# (22, 18, 17, 17, 13, 13)
```

```
round(chisq$expected,2)
```

```
# 1  2  3  4  5  6
```

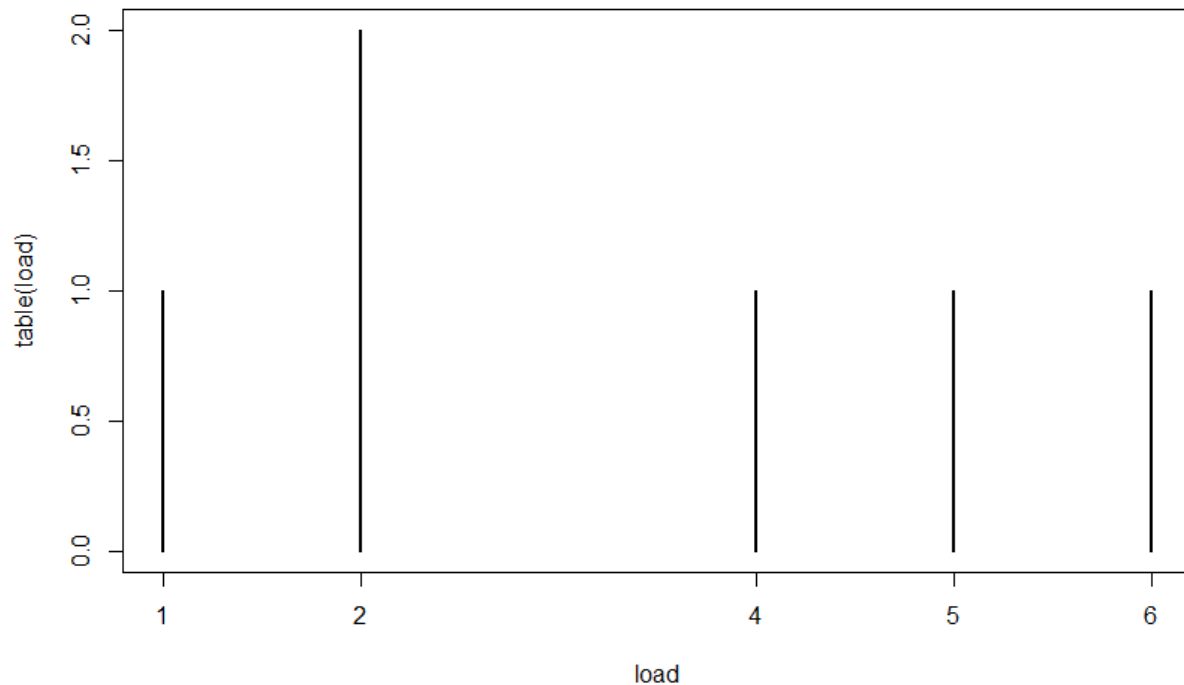
```
# 16.67 16.67 16.67 16.67 16.67 16.67
```

```
chisq$p.value
```

```
#[1] 0.6324852
```

```
load <- c(6,2,5,2,4,1)
```

```
plot(table(load))
```



### #PP3\_Conclusion:

#Here at a confidence level of 0.01, the T value is 2.626405

#degree of freedom =  $n-1$ , which is  $100-1 = 99$

#Therefore in this case, the p value for a t statistics is 0.6325 and thus we fail to reject the null hypothesis as p value is greater than 0.01

#Thus the dice is fair

#The die in this case would be considered unfair if the p value is less than 0.01 and we reject the null hypothesis

### #EP1 & EP2:

#Rolling the dice 100 times and analyzing the data to find out whether the dice is fair or not?

```
outcomes <- sample(x = 1:6, size = 500 , replace = TRUE)
```

```
print(outcomes)
```

#OUTPUT I GOT AFTER RUNNING THE CODE

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

[61] 5 2 1 2 4 2 3 3 4 1 1 1 3 6 1 5 2 6 2 5 4 1 4 5 5 5 1 5 3 6 5 3 2 6 6 1 6 5 5 6 3 6 1 3 1 4 2 6 6
2 1 4 4 2 2 1 2 1 6 6

[121] 1 6 6 3 6 6 5 4 5 2 6 2 2 2 6 3 5 2 5 5 3 1 2 5 6 4 1 5 4 3 5 2 3 3 4 2 3 1 3 5 2 4 2 4 3 5 1 4 3
4 6 1 2 1 6 4 6 3 5 2

[181] 3 2 2 4 1 4 6 6 6 4 2 2 3 5 1 2 3 5 6 1 2 4 6 1 3 3 5 2 1 2 3 5 2 2 1 6 4 4 2 2 2 2 6 6 4 5 4 5 4
3 6 2 6 3 3 3 6 6 2 1

[241] 4 1 4 5 6 5 6 3 4 3 1 6 1 2 6 1 1 4 1 5 5 6 4 5 6 1 1 6 2 5 1 4 2 3 1 2 4 4 6 3 3 5 4 5 3 5 3 6 2
6 6 4 1 6 1 4 5 4 5 3

[301] 2 5 1 5 1 1 2 3 4 5 5 1 3 2 5 4 1 1 6 6 6 5 2 6 2 4 6 3 5 4 5 1 3 2 3 6 4 3 1 4 1 6 5 3 6 3 4 1 6
1 2 1 2 4 5 4 4 5 3 1

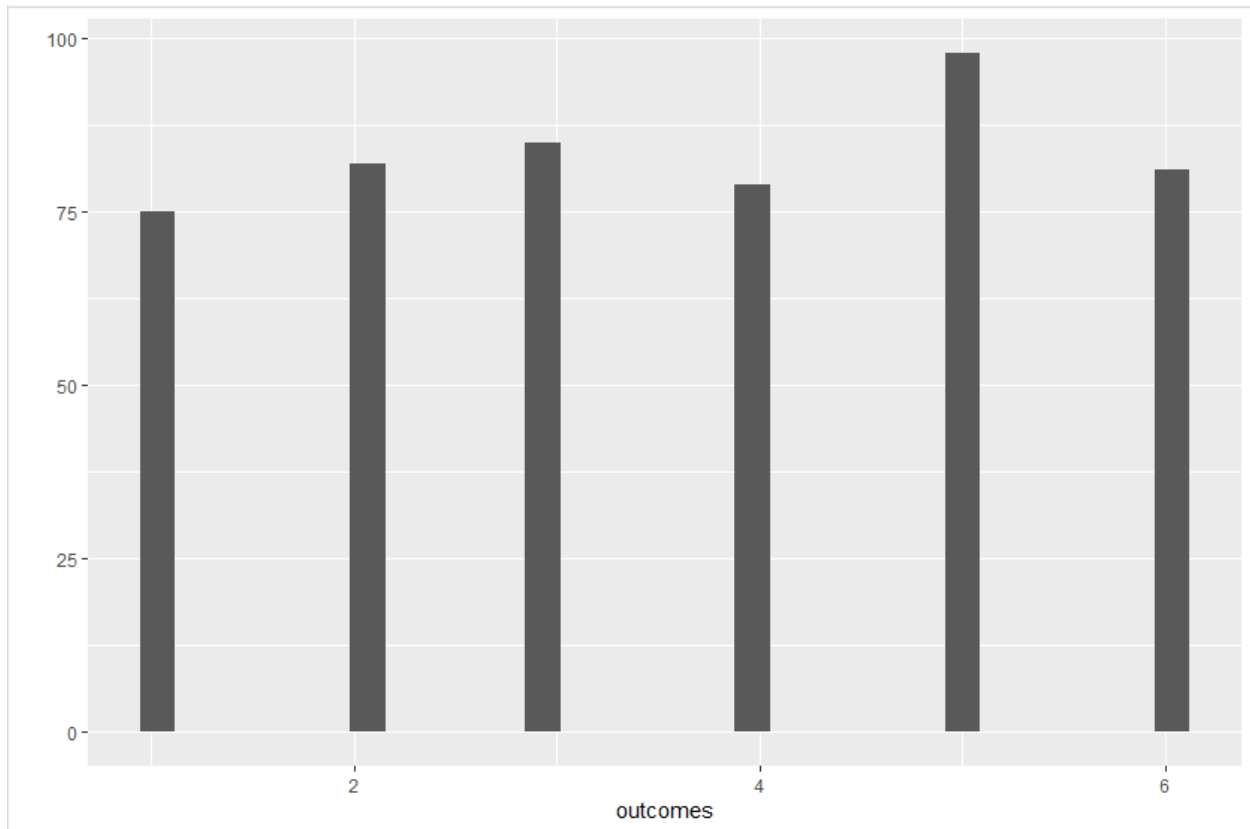
[361] 2 2 5 4 5 3 5 2 5 3 1 1 6 2 3 5 2 3 5 3 4 5 5 4 3 4 1 5 5 4 4 2 3 2 6 6 3 4 2 4 1 5 5 1 5 4 1 6 6
3 6 2 2 5 5 6 3 3 4 5

[421] 3 5 4 6 3 2 2 2 2 6 5 6 5 3 5 5 3 5 1 2 1 3 6 2 3 6 3 3 3 4 2 3 3 5 1 3 1 6 3 4 4 5 5 2 3 4 2 5 5
5 4 6 4 4 3 3 5 4 6 1

[481] 3 4 4 3 5 3 3 1 3 4 1 6 6 5 3 3 3 1 6 5'
```

```
library(ggplot2)
```

```
qplot(outcomes, binwidth = 1)
```



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

```
count1
```

```
# [1] 75
```

```
# 75/500
```

```
# Probability = 15%
```

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

```
count2
```

```
# [1] 82
```

```
# 82/500
```

```
# Probability = 16.4%
```

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

```
count3
```

```
# [1] 85
```

```
# 85/500
```

```
# Probability = 17%
```

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

```
count4
```

```
# [1] 79
```

```
# 79/500
```

```
# Probability = 15.8%
```

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

```
count5
```

```
# [1] 98
```

```
# 98/500
```

```
# Probability = 19.6%
```

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

```
count6
```

```
# [1] 81
```

```
# 81/500
```

```
# Probability = 16.2%
```

```
mean(outcomes)
```

```
# [1] 3.572
```

```
median(outcomes)
```

```
# [1] 4
```

```
var(outcomes)
```

```
#[1] 2.842501
```

```
sd(outcomes)
```

```
#[1] 1.685972
```

```
summary(outcomes)
```

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

```
# 1.000  2.000  4.000  3.572  5.000  6.000
```

```
chisq <- chisq.test(table(outcomes))
```

```
chisq
```

```
'Chi-squared test for given probabilities
```

```
data: table(outcomes)
```

```
X-squared = 3.76, df = 5, p-value = 0.5845'
```

```
chisq$observed
```

```
# outcomes
```

```
# 1 2 3 4 5 6
```

```
# 75 82 85 79 98 81
```

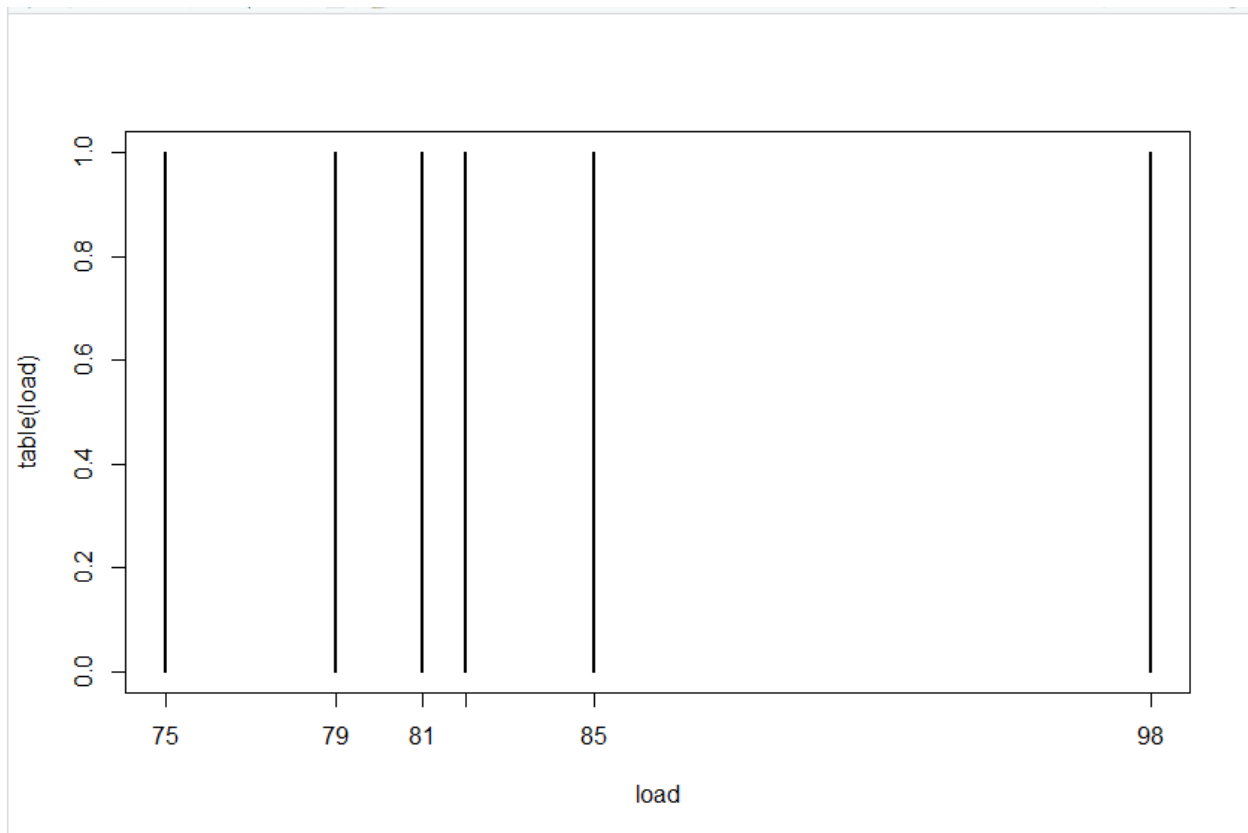
```
round(chisq$expected,2)
```



```
# 1  2  3  4  5  6
# 83.33 83.33 83.33 83.33 83.33 83.33
```

```
chisq$p.value
#[1] 0.5844609
```

```
load <- c(75,82,85,79,98,81)
plot(table(load))
```



**#EP1 & EP2\_Conclusion:**

#Here at a confidence level of 0.01, the T value is 2.585718

#degree of freedom = n-1, which is 500-1 = 499

#Therefore, in this case, the p value for a t statistic is 0.8864 and thus we fail to reject the null hypothesis as p value is greater than 0.01

#Thus, the dice is fair

#The die in this case would be considered unfair if the p value is less than 0.01 and we reject the null hypothesis