Homework 2

Kyle Walter

4/20/2021

Student name:Kyle Walter
Homework number:2
Date due: 2021-04-26

Attribution statement: I did this homework by myself, with help from the book and the professor

---

Exercise 1

Prompt - Flip a fair coin using R 9 times and repeat the expirement 10,000 times.

```
set.seed(1)
table(rbinom(10000, size = 9, prob = 0.5))
```

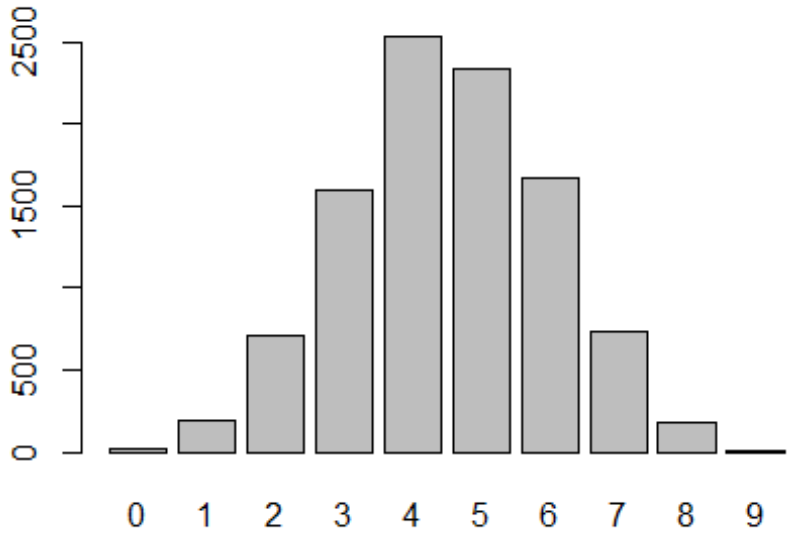
```
##
##    0    1    2    3    4    5    6    7    8    9
##   24  191  706 1593 2533 2341 1678  740  182   12
```

SO utilizing R I've flipped a hypothetical coin 9 times and ran the trials 10 thousand times. Since a coin can be either heads or tails, and with equal odds as it is a fair coin, I would expect to see the coin show up with as many heads as it does tails. Indeed the distribution shows that is basically follows this expectation. When counting the heads the majority show up around 4 or 5 while the less like scenarios of getting 0 heads or all 9 coins coming up as heads is quite small.

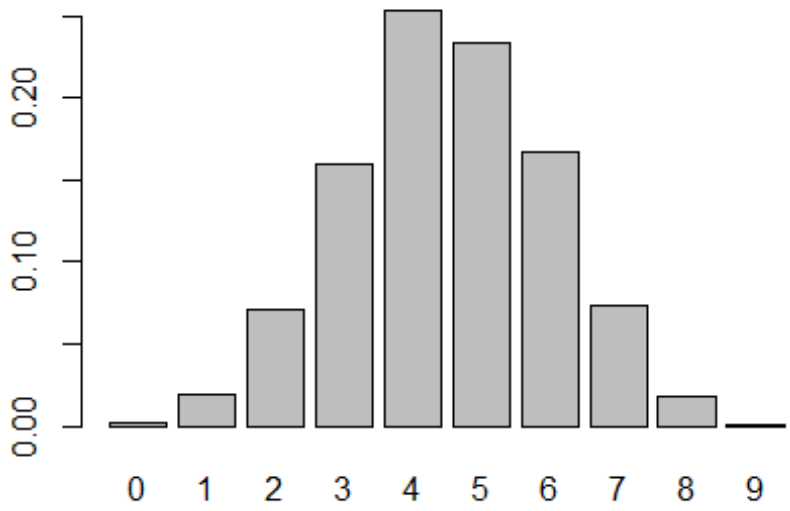
Exercise 2

Prompt - Create barplot of the table from exercise 1 and as well as converting the table to probabilites and generating a barplot. Write a brief analysis of how they are related and why the center is near where it is.

```
set.seed(1)
barplot(table(rbinom(10000, size = 9, prob = 0.5)))
```




```
set.seed(1)
Table <- table(rbinom(10000, 9, prob = .5))
barplot(Table/margin.table(Table))
```



We have two graphs based on the data from the first section. The First graph follows the counts of how many heads show up in trial, while the 2nd graph has transformed those results to probability. Both Charts show that the data centers around 4 and 5 of flips being observered as heads.the shap of the graph is a curve with the aforementioned center around 4 and 5. This shape makes sense as the probability of getting a head on a coin flip is 50%. Which other outcomes such as all heads and no heads are observed as well, they're highly rare as we see with with 9 and 1 respectively.

Exercise 6

Prompt - One hundred college students took a statistics test. 50 of them high school student and 50 college students. 20 failed and 80 passed. While this isn't enough information to draw a table, the additional piece of information offered is that only 3 college students failed. Fill in the table tand comment on why that one piece of information is enough to fill in the table.

```
StudentGrid <- matrix(c(17, 3, 50-17, 50-3), byrow = T, ncol = 2)
rownames(StudentGrid) <- c("Fail","Pass")
colnames(StudentGrid) <- c("High School","College")
StudentGrid
```

```
##      High School College
## Fail          17       3
## Pass          33      47
```

Okay so from the promt we know that that 100 student took the test. 50 High School and 50 College stududnets. 80 of them passed and 3 of the failures, with the additional piece of the information, Since we know that there were for 50 college students and 3 failed that mean

47 college students passed the exam. Additionally since there were 20 students who failed, 17 high student were in the failure category. With that determined, the remaining portion of 50 High School students is 33, which is the number that passed.

From this step we can see what the overall pass percentage of the high school students is;

```
StudentGrid[,1]/margin.table((StudentGrid[,1]))

## Fail Pass
## 0.34 0.66
```

Now focusing in on the population of high school students from the derived numbers, we can see that roughly 1/3rd of the sample has failed. If this is an exam we would expect them to pass it may be time to revisit the curriculum, but there isn't any information about that in the prompt.

Exercise 7

Prompt - In a typicl year 71 out of 100k home in the UK are repossed by the bank due to default. Barclays starts collecting test data. 93,935 house holds pass teh test and 6065 house holds fail the test. Interestint 5996 of the failed households are still doing fine on their mortgage. What percentage of customers both pass and are doing fine on their mortgage?

```
#Calculate the housing Groups
HousingGrid <- matrix(c(6065-5996,round((93935*((6065-
5996)/6065))),5996,round((93935*(5996/6065)))), byrow = T, ncol = 2)
rownames(HousingGrid) <- c("Repossessed", "Good Standing")
colnames(HousingGrid) <- c("Fail Test","Pass Test")
HousingGrid

##                Fail Test Pass Test
## Repossessed           69      1069
## Good Standing       5996     92866
```

In calculating the Barclay's data, we see that 6065 households fail their new credit test. Since 5996 households have failed but are still in good standing, the remaining 69 houses both failed the test and had their homes repossessed. Assuming this ratio holds, I multiplied the total house holds who passed by those who failed and had their home reposed. Since partial households do not make sense, I rounded to the nearest integer and found 1069 households passed the test but with the data period that Barclay collected 1069. that leaves the remaining portion who passed Barclay's test and are in good standing as 92866.

```
#Probabilities as a percent of total
HousingGrid/margin.table(HousingGrid)*100

##                Fail Test Pass Test
## Repossessed        0.069     1.069
## Good Standing      5.996    92.866
```

Transforming the information to a percentage we see that 92.8% of households both are passing the test Established by Barclay's and still in good standing.

Exercise 8

Prompt - Imagine Barclays deploys the screening from the prior exercise and a customer fails the test. Which is the likelihood the cusomter actually fails the test.

```
#Probility a customer will both fail Barclay's Test and have their home repossessed
HousingGrid[1,1]/margin.table(HousingGrid)*100
```

```
## [1] 0.069
```

Using the data from before, there is a 0.07% chance that the customer will actually end up in home being repossessed. As seen in the prior step I calculated the table using the total number of screened customers and those who failed and were still in good standing to fill in the utility table. Since I had actually calculated the percentage for all 4 outcomes in the prior step, I simply highlighted this one and printed the probability as a percentage.