Homework 2

Daniel Caley

7/8/2021

Contents

Questions for this Week

The homework for week two is exercises 1 and 2 on page 35, as well as problems 6, 7, and 8 on page 36.

Question 1

Flip A Coin

Flip a fair coin nine times and write down the number of heads obtained. Write down the results and explain in your own words what they mean.

```
coin_toss_9 <- table(rbinom(n = 1, size = 9,  prob = 0.5))
coin_toss_9
```

```
##
## 5
## 1
```

- The first results of 9 coin flips shows us that the probability of being heads vs tails will hover around 50% but is random in the coin flip resulting in heads over tails. By flipping by hand R can simulate that out of 9 coin flips 5 are heads. That comes out to be 55% which is close to 50%. Not always will we be close to 50% but most of the time.

Flip A Coin 100,000 Times

Now repeat this process 100,000 times. Obviously you don't want to have to do that by hand, so create the necessary lines of R code to do it for you. Hint: You will need both the rbinom() function and the table() function. Write down the results and explain in your own words what they mean.

```
coin_toss_100k <- table(rbinom(n = 100000, size = 9,  prob = 0.5))
coin_toss_100k
```

```
##
##     0     1     2     3     4     5     6     7     8     9
##   194  1708  7015 16390 24866 24269 16521  7025  1834   178
```

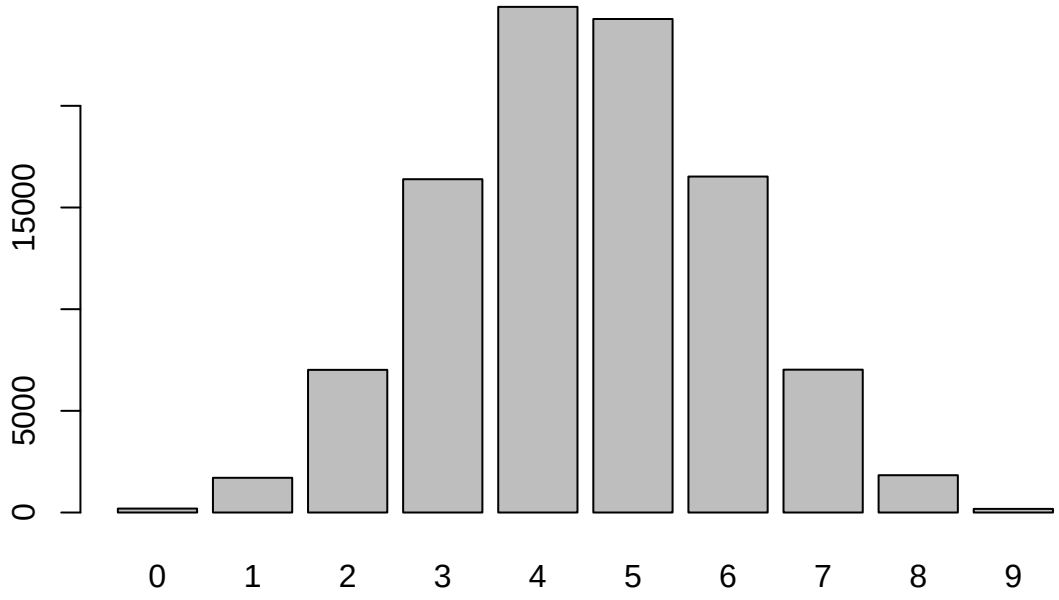
- The second results shows us that when looking at the 100,000 flips we can see that the probability of a 50/50 coin flip is quite high. As we move away from events of 4 and 5 we see the probability begin to drop. So events 0, 1, 2, 3 and 6, 7, 8, 9 the probability reduces greatly from getting a heads or tails.

Question 2

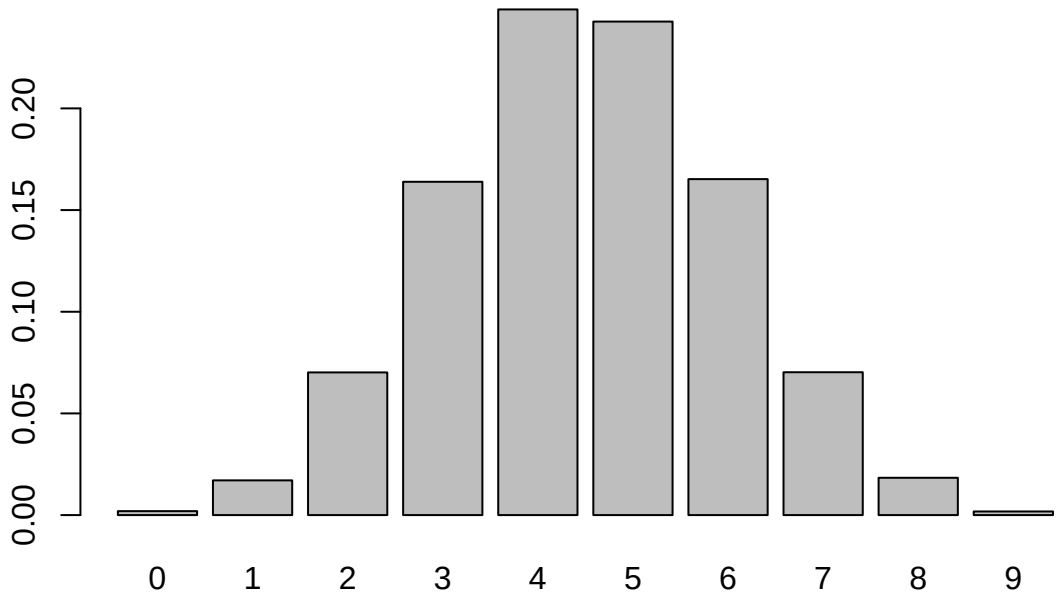
Plot the Results

Using the output from Exercise 1, summarize the results of your 100,000 trials of nine flips each in a bar plot using the appropriate commands in R. Convert the results to probabilities and represent that in a bar plot as well.

```
# Barplot Results
barplot(coin_toss_100k)
```




```
# Barplot probability
barplot(coin_toss_100k/100000)
```



Brief Interpretation

Write a brief interpretive analysis that describes what each of these bar plots signifies and how the two bar plots are related. Make sure to comment on the shape of each bar plot and why you believe that the bar plot has taken that shape. Also make sure to say something about the center of the bar plot and why it is where it is.

- Each of these bar plots signify the probability of the coin being either heads or tails. The first one is in the raw numbers of flips where the second one is as a percentage. The second graph, which is in percentages, is easier to identify the likely hood of the coin either being Head or Tails. By having the percentages we can easily communicate the probability rather than saying raw numbers like 24,986.
- The shape of the bar plot is normally distributed which align with the Law of Large numbers. Also due to having a fair coin our probability is 50%. After performing 100,000 coin flip trials the result will distribute close to our probability of 50%. When taking the average we will get very close to the probability setting.

Question 6

Statistics Test

One hundred students took a statistics test. Fifty of them are high school students and 50 are college students. Eighty students passed and 20 students failed. You now have enough information to create a two-by-two contingency table with all of the marginal totals specified (although the four main cells of the table are still blank). Draw that table and write in the marginal totals.

```
StatsTest1 <- matrix(c(80,20))
StatsTest1 <- rbind(StatsTest1, margin.table(StatsTest1))
colnames(StatsTest1) <- c("Students")
rownames(StatsTest1) <- c("Pass","Fail","Marginal")

StatsTest1
```

```
##          Students
## Pass           80
## Fail           20
## Marginal      100
```

Statistics Test (cont.)

I'm now going to give you one additional piece of information that will fill in one of the four blank cells: only three college students failed the test. With that additional information in place, you should now be able to fill in the remaining cells of the two-by-two table.

```
StatsTest2 <- matrix(c(47,3,33,17), ncol = 2)
colnames(StatsTest2) <- c("College", "High School")
rownames(StatsTest2) <- c("Pass", "Fail")

StatsTest2
```

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

Comments on Statistics Test

Comment on why that one additional piece of information was all you needed in order to figure out all four of the table's main cells.

- That one piece of information told me that 3 college students failed therefore using deductive reasoning the other 47 college students had to pass.
- We also knew that 20 had failed, so if 3 had failed then 17 High School Students had to fail.
- Lastly their were 50 High Students and if 17 had failed 33 had to pass.

Probabilities on Statistics Test

Finally, create a second copy of the complete table, replacing the counts of students with probabilities. What is the pass rate for high school students? In other words, if one focuses only on high school students, what is the probability that a student will pass the test?

```
StatsTest_Prob <- StatsTest2/margin.table(StatsTest2)
StatsTest_Prob
```

```
##      College High School
## Pass    0.47        0.33
## Fail    0.03        0.17
```

```
StatsTest_Prob[,2] / sum(StatsTest_Prob[,2])
```

```
## Pass Fail
## 0.66 0.34
```

- After normalizing the dataset for just High School students we see that 66% would likely pass the test.

Question 7

Barclay Mortgage Default

In a typical year, 71 out of 100,000 homes in the United Kingdom is repossessed by the bank because of mortgage default (the owners did not pay their mortgage for many months). Barclay Bank has developed a screening test that they want to use to predict whether a mortgagee will default. The bank spends a year collecting test data: 93,935 households pass the test and 6,065 households fail the test. Interestingly, 5,996 of those who failed the test were actually households that were doing fine on their mortgage (i.e., they were not defaulting and did not get repossessed). Construct a complete contingency table from this information. Hint: The 5,996 is the only number that goes in a cell; the other numbers are marginal totals.

```
# assuming the other marginal total is 71 for defaults we can then fill out this matrix.
# otherwise this question is too vague or fuzzy and therefore I would clarify this in
# a real world situation.  The question was also asked in a live session, which was confirmed to
# be fuzzy.
Repo_Barclays <- matrix(c(93933, 5996, 2,69), ncol = 2)
colnames(Repo_Barclays) <- c("Not Repo","Repo")
rownames(Repo_Barclays) <- c("Pass","Fail")

Repo_Barclays
```

```
##      Not Repo Repo
## Pass    93933    2
## Fail     5996   69
```

```
addmargins(Repo_Barclays)
```

```
##      Not Repo Repo    Sum
## Pass    93933    2  93935
## Fail     5996   69   6065
## Sum     99929   71 100000
```

Barclay Percentage Not Repossessed

What percentage of customers both pass the test and do not have their homes repossessed?

```
Repo_Prob <- Repo_Barclays / margin.table(Repo_Barclays)
Repo_Prob
```

```
##      Not Repo    Repo
## Pass  0.93933 0.00002
## Fail  0.05996 0.00069
```

```
Repo_Prob[1,1]
```

```
## [1] 0.93933
```

The percentage of customers both pass the test and did not have their homes repossessed was 93.9%.

Question 8

Imagine that Barclays deploys the screening test from Exercise 6 on a new customer and the new customer fails the test. What is the probability that this customer will actually default on his or her mortgage? Show your work and especially show the tables that you set up to help with your reasoning.

Barclay Percentage Reposed

```
Repo_Barclays
```

```
##      Not Repo Repo
## Pass    93933    2
## Fail     5996   69
```

```
failed <- Repo_Prob[2,]/ sum(Repo_Prob[2,])

scales::percent(failed[2], accuracy = .01)
```

```
##    Repo
## "1.14%"
```

Again utilizing the contingency table from Question 7, when someone fails the test only 1% get their house Reposed. Regardless that 6065 people fail the test.