IST 777

Joe Hernandez

7/18/2019

Homework #2

1. Flip fair coin 9 times and write down number of heads obtained after 9 flips, 7 heads were obtained with the following sequence of flips (1 indicates heads, 0 indicates tails).

```
> manflip <- c(1, 1, 0, 1, 1, 1, 1, 0, 1)
> manflip
[1] 1 1 0 1 1 1 1 0 1
```

Repeat process 100,000 times using rbinom() function and table function setting seed for replication.

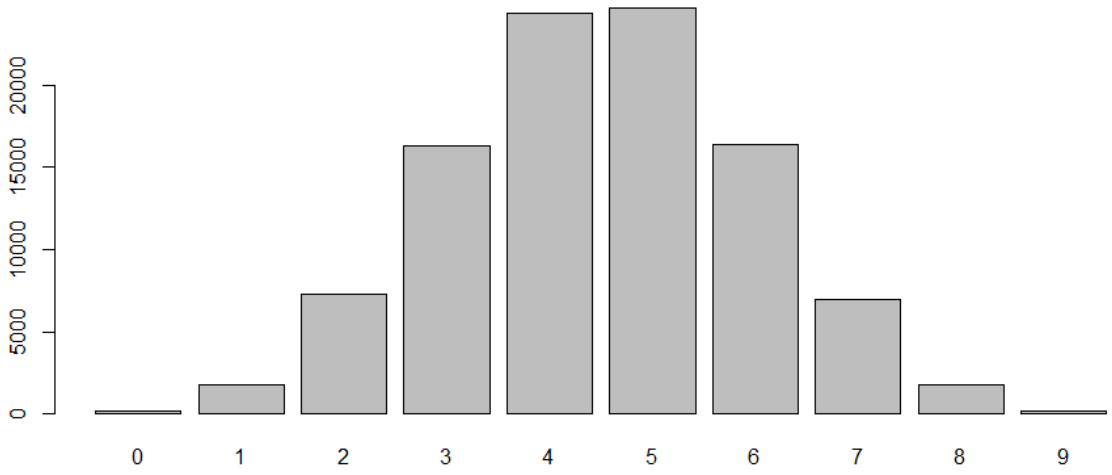
```
> set.seed(188)
> comflip <- table(rbinom(n = 100000, size = 9, prob = 0.5))
> comflip

    0     1     2     3     4     5     6     7     8     9
  205  1757  7290 16327 24394 24670 16431  6980  1769   177
```

The results show that after 100000 trials of 9 flips:
205 resulted in 0 heads
1,757 resulted in 1 head
7,290 resulted in 2 heads
16,327 resulted in 3 heads
24,394 resulted in 4 heads
24,670 resulted in 5 heads
16,431 resulted in 6 heads
6,980 resulted in 7 heads
1,769 resulted in 8 heads
177 resulted in 9 heads

2. Create bar plot of trials

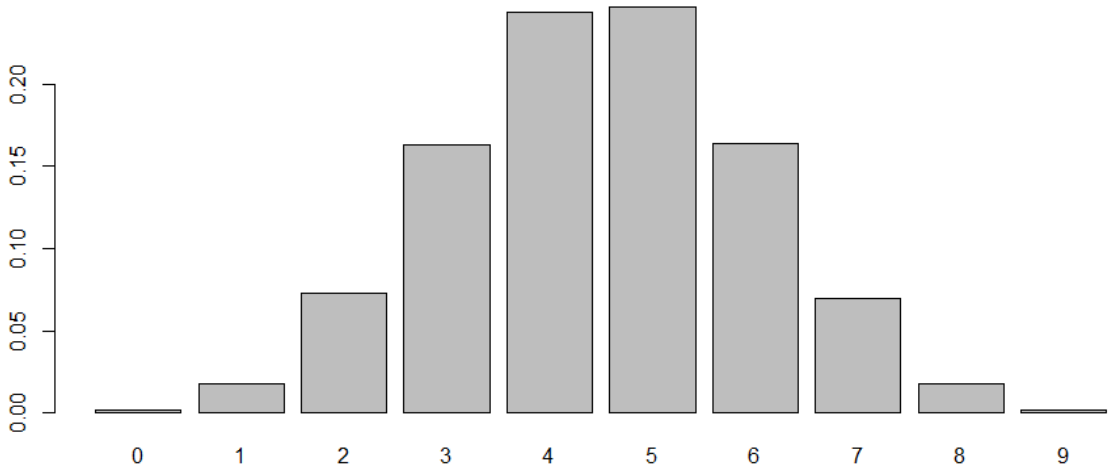
```
> barplot(comflip)
```



Convert to probability and show in bar plot as well

```
> barplot(comflip)
> probcomflip <- comflip/100000
> probcomflip

      0       1       2       3       4       5       6       7
8       9
0.00205 0.01757 0.07290 0.16327 0.24394 0.24670 0.16431 0.06980
0.01769 0.00177
> barplot(probcomflip)
```



The first plot shows the raw data where the x axis shows the number of heads that appeared and the y axis shows the number of trials. The second plot is the probability from this result in a trial resulting in the number of heads. These two plots are related as the first plot sets the foundation in counting the results of each trial and the second takes the results and divides it by the total number of trials, thus, creating a simple probability in predicting a trial result.

The bar plot has taken a normal distribution shape due to the law of large numbers which indicates that the larger the number of observations has a convergence to a normal distribution. This is especially true with a 50/50 probability for a single event.

The center of the plot shows that between 4 and 5 heads are the most frequent results. This is true as there are many combinations to getting 4 or 5 heads than there are in getting extreme results such as 0 heads or 9 heads.

6. Create 2x2 contingency table with marginal totals specified
   Given: Marginal totals of 50 High School, 50 College, 80 Pass, 20 Fail, 3 College Fail
   Populate matrix based on given information (on pen and paper)

```
> stattest <- matrix(c(33, 47, 17, 3), ncol = 2, byrow = TRUE)
> stattest <- as.table(stattest)
> stattest <- rbind(stattest, margin.table(stattest, 2))
> stattest <- cbind(stattest, margin.table(stattest, 1))
> colnames(stattest) <- c("High School", "College", "Total")
> rownames(stattest) <- c("Pass", "Fail", "Total")
> stattest
      High School College Total
Pass           33      47    80
Fail           17       3    20
Total          50      50   100
```

The last piece of information helped shape the contingency table as the marginal totals were given. With the logic that 20 total students failed and 3 were in college the the difference of 17 belonged to high school students. Also, given 3 college students failed the remaining 47 of 50 passed. Lastly, the difference between 80 total passing students and 50 High School students with the newly populated table leaves 33 High School students remaining to complete the contingency table

Convert to probability

```
> stattestProbs <- stattest[1:2, 1:2]/margin.table(stattest[1:2,
1:2])
> stattestProbs <- rbind(stattestProbs,
margin.table(stattestProbs, 2))
```

```
> stattestProbs <- cbind(stattestProbs,
margin.table(stattestProbs, 1))
> colnames(stattestProbs) <- c("High School", "College", "Total")
> rownames(stattestProbs) <- c("Pass", "Fail", "Total")
> stattestProbs
      High School College Total
Pass         0.33    0.47   0.8
Fail         0.17    0.03   0.2
Total        0.50    0.50   1.0
```

Focus only on High School students find the pass probability among them

```
> HSstattestProbs <- stattest[1:2, 1]/sum(stattest[1:2, 1])
> HSstattestProbs
Pass Fail
0.66 0.34
```

The probability of High School Students passing the test is 66%

7. Create 2x2 contingency table with marginal totals specified
   Given: Marginal totals of 50 High School, 50 College, 80 Pass, 20 Fail, 3 College Fail
   Populate matrix based on given information (on pen and paper)

```
> repo <- matrix(c(93933, 2, 5996, 69), ncol = 2, byrow = TRUE)
> repo <- as.table(repo)
> repo <- rbind(repo, margin.table(repo, 2))
> repo <- cbind(repo, margin.table(repo, 1))
> colnames(repo) <- c("No Repo", "Repo", "Total")
> rownames(repo) <- c("Pass", "Fail", "Total")
> repo
      No Repo Repo  Total
Pass    93933    2  93935
Fail     5996   69   6065
Total   99929   71 100000
```

Convert to probability

```
> repoProbs <- repo[1:2, 1:2]/margin.table(repo[1:2, 1:2])
> repoProbs <- rbind(repoProbs, margin.table(repoProbs, 2))
> repoProbs <- cbind(repoProbs, margin.table(repoProbs, 1))
> colnames(repoProbs) <- c("No Repo", "Repo", "Total")
> rownames(repoProbs) <- c("Pass", "Fail", "Total")
> repoProbs
      No Repo    Repo   Total
Pass  0.93933 0.00002 0.93935
Fail  0.05996 0.00069 0.06065
Total 0.99929 0.00071 1.00000
```

~93.9% of homes passed the test and did not have their homes repossessed

8. Find probability where customer fails the test and defaults on mortgage
   Uses Exercise 6 contingency table even though Exercise 7 uses Barclays test

```
> scrntest1 <- matrix(c(33, 47, 17, 3), ncol = 2, byrow = TRUE)
> FscrntestP1 <- scrntest1[2, 1:2]/sum(scrntest1[2, 1:2])
> FscrntestP1
[1] 0.85 0.15
```

   Using the contingency table from Exercise 6, by taking the values of only the customers who failed the test we see a 15% probability of customers defaulting on their mortgage.

   Insurance Exercise 7 Barclays test

```
> scrntest2 <- matrix(c(93933, 2, 5996, 69), ncol = 2, byrow =
TRUE)
> FscrntestP2 <- scrntest2[2, 1:2]/sum(scrntest2[2, 1:2])
> FscrntestP2
[1] 0.98862325 0.01137675
```

   Using the contingency table from Exercise 7, by taking the values of only the customers who failed the test we see a ~1.138% probability of customers defaulting on their mortgage.