Homework 2
Syracuse University
IST 772
Summer 2021

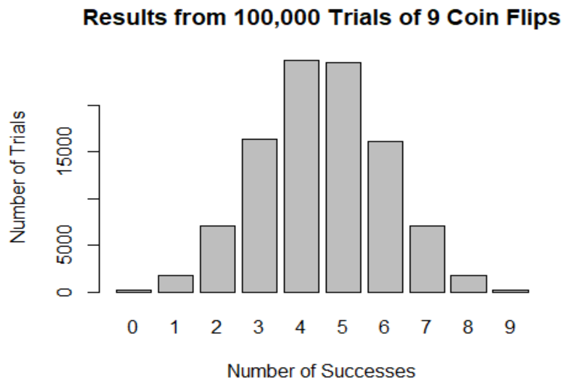
Question 1

```
# flip a fair coin nine times
CoinFlipResults <- rbinom(9, 1, 0.5)

# display the coin flip results
print(CoinFlipResults)

## [1] 0 1 0 0 0 1 1 0 0

# repeat this process 100,000 times
CoinFlipResults <- rbinom(100000, 9, 0.5)

# display the coin flip results
print(table(CoinFlipResults))

## CoinFlipResults
##     0     1     2     3     4     5     6     7     8     9
##   191  1781  7067 16418 24830 24612 16156  7060  1710   175
```

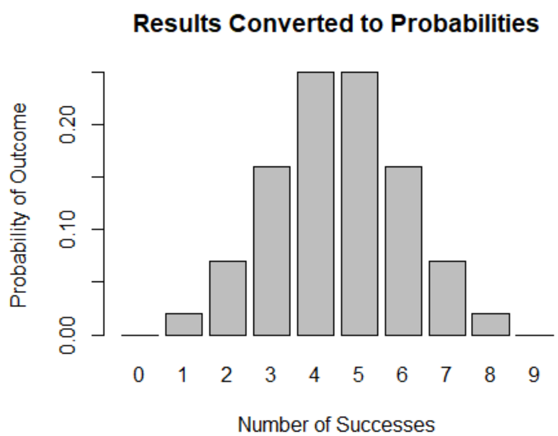
The binomial experiment conducted consisted of flipping a coin 9 times and repeating that 100,000 times. There are 10 possible outcomes that can occur. Each column in the table reflects one possible outcome. For example, the 0 means that there were 0 coins that landed on heads. The 1 means that 1 coin landed on heads, and so on and so forth. Each possible outcome has a corresponding number that reflects the number of times that outcome was observed in the experiment.

Question 2

```
# summarize the results in a barplot
barplot(table(CoinFlipResults),
        main = "Results from 100,000 Trials of 9 Coin Flips",
        xlab = "Number of Successes",
        ylab = "Number of Trials")
```




```
# convert to probabilities and plot again
barplot(round(table(CoinFlipResults) / sum(table(CoinFlipResults)), 2),
        main = "Results Converted to Probabilities",
        xlab = "Number of Successes",
        ylab = "Probability of Outcome")
```



The first barplot represents the number of trials for each of the corresponding outcomes, while the second barplot represents the conversion into probabilities. To convert into probabilities, each value for the number of trials is divided by the total number of trials. The center of the barplot represents the outcomes that are the most likely outcomes, and the further from the middle represents less likely outcomes, with the bars at the end being the least likely outcomes.

Question 6

```
# create the structure of the table
StatsTestResults <- matrix(nrow = 3, ncol = 3)
row.names(StatsTestResults) <- c("Pass", "Fail", "Total")
colnames(StatsTestResults) <- c("College", "HighSchool", "Total")

# enter the marginal totals
```

```r
StatsTestResults[1, 3] <- 80 # number of students that passed
StatsTestResults[2, 3] <- 20 # number of students that failed
StatsTestResults[3, 1] <- 50 # number of college students
StatsTestResults[3, 2] <- 50 # number of high school students
StatsTestResults[3, 3] <- 100 # total number of students

# show the table
print.table(StatsTestResults)

##       College HighSchool Total
## Pass                        80
## Fail                        20
## Total      50         50   100

# 3 college students failed the test
StatsTestResults[2, 1] <- 3

# show the table
print.table(StatsTestResults)

##       College HighSchool Total
## Pass                        80
## Fail        3               20
## Total      50         50   100

# solve for the remaining information
StatsTestResults[1, 1] <- 47 # number of college students that passed
StatsTestResults[1, 2] <- 33 # number of high school studnets that passed
StatsTestResults[2, 2] <- 17 # number of high school students that failed

# show the final table
print.table(StatsTestResults)

##       College HighSchool Total
## Pass       47         33    80
## Fail        3         17    20
## Total      50         50   100

# convert the table to probalities
print.table(StatsTestResults / 100, digits = 2)

##       College HighSchool Total
## Pass     0.47       0.33  0.80
## Fail     0.03       0.17  0.20
## Total    0.50       0.50  1.00
```

By knowing the number of college students that failed, I was able to solve for the number of college students that passed by subtracting from the total. From here, I was able to solve for the number of high school students that passed and failed because I already know the total number of pass and fails. The pass rate for high school students is 33 %.

Question 7

```
# create the structure of the table
Homes <- matrix(nrow = 3, ncol = 3)
row.names(Homes) <- c("Pass", "Fail", "Total")
colnames(Homes) <- c("DefaultYes", "DefaultNo", "Total")

# fill out the information in the table
Homes[3, 3] <- 100000 # total number of homes in the experiment
Homes[1, 3] <- 93935 # number of homes that pass the test
Homes[2, 3] <- 6065 # number of homes that fail the test
Homes[2, 2] <- 5996 # number of homes that fail and are not defaulting
Homes[2, 1] <- 69 # number of homes that fail and are defaulting
Homes[3, 1] <- 71 # typical number of homes that default
Homes[1, 1] <- 2 # number of homes that pass the test and are defaulting
Homes[1, 2] <- 93933 # number of homes that pass the test not defaulting
Homes[3, 2] <- 99929 # total number of homes that are not defaulting

# print out the table
print.table(Homes)

##       DefaultYes DefaultNo  Total
## Pass           2     93933  93935
## Fail          69      5996   6065
## Total         71     99929 100000

# convert to probabilities
print.table(Homes / 100000, digits = 5)

##       DefaultYes DefaultNo   Total
## Pass     0.00002   0.93933 0.93935
## Fail     0.00069   0.05996 0.06065
## Total    0.00071   0.99929 1.00000

# what percent of customers both pass and do not default?
Homes[1, 2] / 1000

## [1] 93.933
```

Question 8

```
# filter down to just the failed test row
NewHomes <- Homes[2, ]

# show the results
NewHomes

## DefaultYes  DefaultNo      Total
##         69       5996       6065
```

```r
# convert to probabilities
NewHomes <- round(NewHomes / NewHomes[3], 2)

# show the results
NewHomes

## DefaultYes  DefaultNo      Total
##       0.01       0.99       1.00

# what is the probability that this customer will default?
NewHomes[1]

## DefaultYes
##       0.01
```

To solve this problem, first I focus in on only those homes that failed the test. From here, I convert to normalized probabilites by dividing by the marginal total. This shows that when knowing ahead of time that the customer failed the test, there is approximately a 1 % chance that the customer will default on their mortgage.