Homework Chapter 2 - Mark Cappiello

1. Flip a fair coin nine times and write down the number of heads obtained. 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.

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

##   
## 0 1 2 3 4 5 6 7 8 9   
## 185 1743 7178 16409 24545 24663 16299 7006 1784 188

Explanation of the Code

Defining Variables in the rbinom function n = 100,000 (this is the number of trials in our binary distribution)

size = 9 (this is the number of events in each trial, in other words we will flip the coin 9 times in each trial)

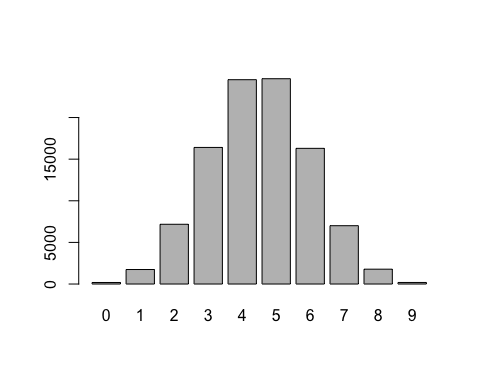
prob = .5 (this is the probability that a fair coin will land on heads)

simulating the trials: The rbinom() function simulates 100,000 trials of flipping a coin 9 times each and counts the number of heads in each trial.

summarizing the results: The table() function creates a frequency table of the number of heads resultin from each trial.

1. 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. 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.

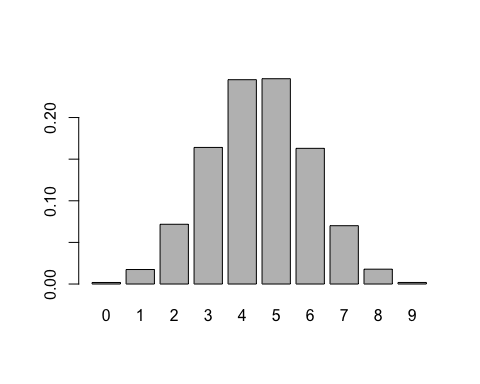
barplot(results)



probabilities <- results/100000  
  
probabilities

##   
## 0 1 2 3 4 5 6 7 8 9   
## 0.00185 0.01743 0.07178 0.16409 0.24545 0.24663 0.16299 0.07006 0.01784 0.00188

barplot(probabilities)



Analysis: Raw Counts Bar Plot The first bar plot shows the event counts of the number of heads recorded in 9 coin flips from our 100,000 trials. The x-axis represents the possible outcomes of 0 to 9 heads, and the y-axis shows the count of each outcome. This bar plot is centered around outcomes of 4 to 5 heads. This indicates that these outcomes occur most frequently.

The shape of the plot is a bell curve that is symmetric and that resembles a normal binomial distribution. This is to be expected because each flip is independent and has a 50 percent chance of being heads. The greatest number of heads outcomes is centered around 4 or 5 heads meaning that is the most likely outcome. Moving away from the center, the counts decrease symmetrically, reflecting the lower probabilities of extreme outcomes of 0 or 9 heads.

Probabilities Bar Plot The second bar plot converts our data raw counts into probabilities by dividing each count by the 100000 or the total number of trials. The x-axis represents the number of heads. The y-axis represents the probability of each outcome.

This plot has the same shape as the raw counts plot from above. But the y-axis is scaled to probabilities of the outcomes. The center of the plot remains around 4 or 5 heads, which have the highest probabilities. The shape is still bell-shaped and symmetric, and it indicates that the probabilities of the outcomes follow a normal binomial distribution.

Relationship Between the Plots The shape and center of both plots are the same because they represent the same underlying distribution. The first plot shows the frequency of each outcome, while the second plot shows the relative probabilities of the same data.

Shape and Center of the Bar Plots The bell-shaped, symmetric form of both plots is characteristic of a binomial distribution. With 9 flips and a fair coin, the expected number of heads is 4.5 (which is the mean of the distribution), and the probabilities are highest near this value. The center of the plots reflects the expected mean outcome, with the number of heads clustering around the middle values (4 and 5), where the combination of flips is most numerous. This central tendency and the symmetrical decrease in probabilities for more extreme outcomes explain the shape of both plots.