Lab 2

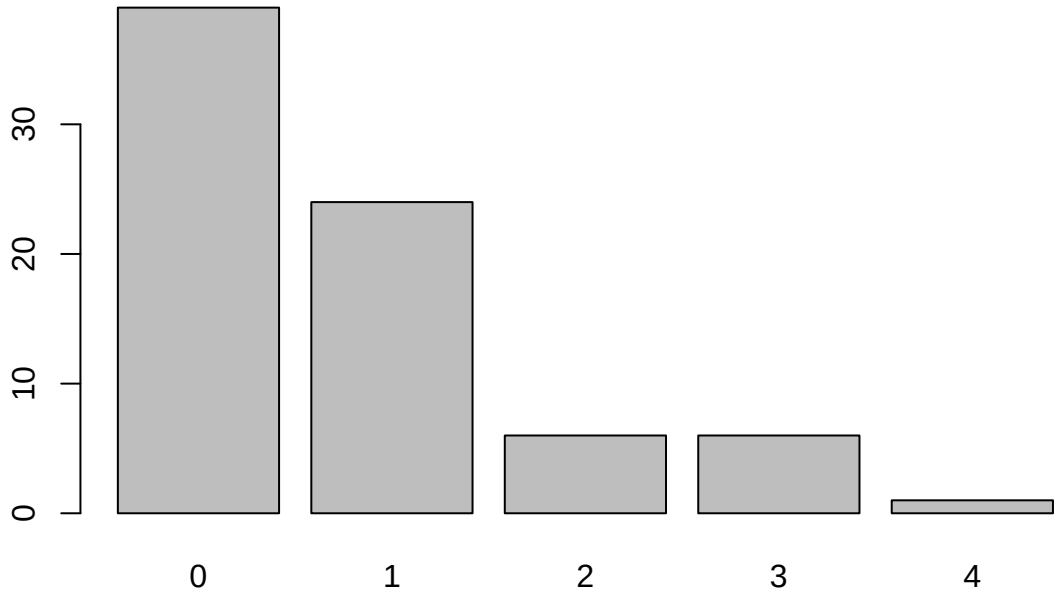
Munkhnaran Gankhuyag

September 13, 2017

```
load("more/kobe.RData")
head(kobe)
```

```
##    vs game quarter time
## 1 ORL    1       1 9:47
## 2 ORL    1       1 9:07
## 3 ORL    1       1 8:11
## 4 ORL    1       1 7:41
## 5 ORL    1       1 7:03
## 6 ORL    1       1 6:01
##                                           description basket
## 1                Kobe Bryant makes 4-foot two point shot      H
## 2                            Kobe Bryant misses jumper      M
## 3                       Kobe Bryant misses 7-foot jumper      M
## 4 Kobe Bryant makes 16-foot jumper (Derek Fisher assists)      H
## 5                        Kobe Bryant makes driving layup      H
## 6                            Kobe Bryant misses jumper      M
```

```
kobe_streak <- calc_streak(kobe$basket)
barplot(table(kobe_streak))
```



1. What does a streak length of 1 mean, i.e. how many hits and misses are in a streak of 1? What about a streak length of 0?

Streak length of 1 means 2 baskets were made in a row.

2. Describe the distribution of Kobe's streak lengths from the 2009 NBA finals. What was his typical streak length? How long was his longest streak of baskets?

Kobe mostly did not make back to back baskets, giving him mostly no streaks. His typical streak length was 1. His longest streak length of baskets was 4.

3. In your simulation of flipping the unfair coin 100 times, how many flips came up heads? Show in New WindowClear OutputExpand/Collapse Output

```
outcomes <- c("heads", "tails")
unfair_coinflip <- sample(outcomes, size = 100, replace = TRUE, prob = c(.20,.80))
table(unfair_coinflip)
```

```
## unfair_coinflip
## heads tails
##    17    83
```

4. What change needs to be made to the `sample` function so that it reflects a shooting percentage of 45%? Make this adjustment, then run a simulation to sample 133 shots. Assign the output of this simulation to a new object called 'sim_basket'.

```
outcomes <- c("H", "M")
sim_basket <- sample(outcomes, size = 133, replace = TRUE, prob = c(.45,.55))
table(sim_basket)
```

```
## sim_basket
##  H  M
## 52 81
```

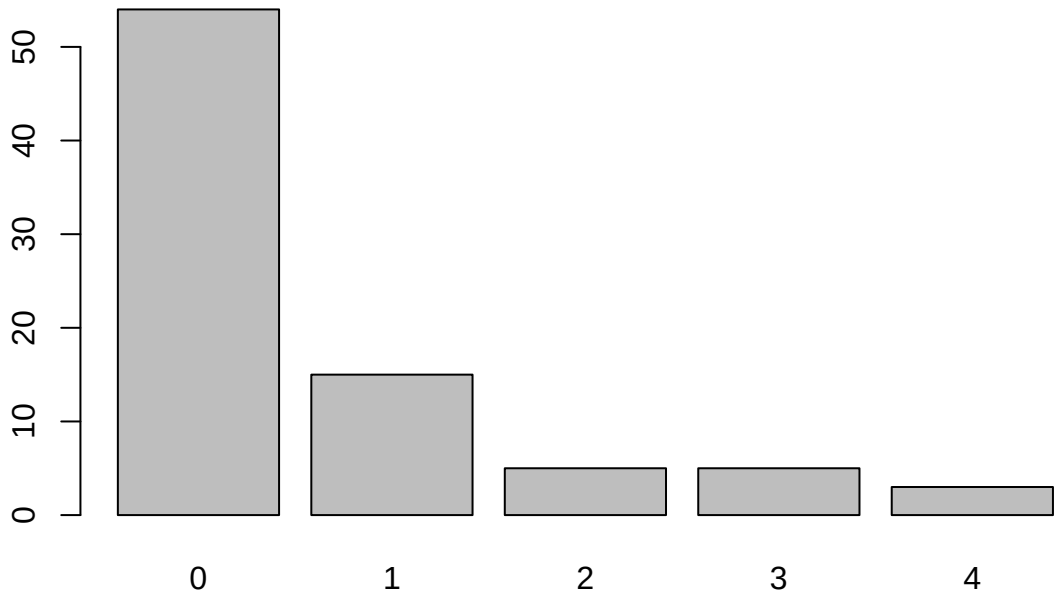
```
kobe$basket
```

```
##   [1] "H" "M" "M" "H" "H" "M" "M" "M" "M" "H" "H" "H" "M" "H" "H" "M" "M"
##  [18] "H" "H" "H" "M" "M" "H" "M" "H" "H" "H" "M" "M" "M" "M" "M" "M" "H"
##  [35] "M" "H" "M" "M" "H" "H" "H" "H" "M" "H" "M" "M" "H" "M" "M" "H" "M"
##  [52] "M" "H" "M" "H" "H" "M" "M" "H" "M" "H" "H" "M" "H" "M" "M" "M" "H"
##  [69] "M" "M" "M" "M" "H" "M" "H" "M" "M" "H" "M" "M" "H" "H" "M" "M" "M"
##  [86] "M" "H" "H" "H" "M" "M" "H" "M" "M" "H" "M" "H" "H" "M" "H" "M" "M"
## [103] "H" "M" "M" "M" "H" "M" "H" "H" "H" "M" "H" "H" "H" "M" "H" "M" "H"
## [120] "M" "M" "M" "M" "M" "M" "H" "M" "H" "M" "M" "M" "M" "H"
```

```
sim_basket
```

```
##   [1] "M" "M" "M" "M" "M" "H" "M" "M" "H" "H" "H" "M" "M" "M" "M" "H" "M"
##  [18] "M" "M" "H" "H" "H" "H" "M" "M" "H" "H" "H" "M" "M" "M" "M" "M" "M"
##  [35] "H" "M" "H" "M" "M" "M" "H" "H" "H" "H" "M" "H" "H" "M" "H" "H" "H"
##  [52] "M" "H" "M" "M" "M" "M" "H" "H" "H" "H" "M" "M" "M" "H" "M" "M" "M"
##  [69] "H" "M" "M" "H" "M" "H" "M" "M" "H" "H" "M" "M" "M" "M" "H" "M" "H"
##  [86] "H" "H" "M" "M" "M" "H" "M" "M" "M" "H" "H" "M" "H" "M" "M" "H" "M"
## [103] "M" "H" "H" "M" "M" "M" "H" "M" "M" "M" "M" "M" "M" "H" "H" "M" "M" "M"
## [120] "H" "M" "M" "M" "H" "H" "H" "M" "M" "M" "M" "M" "M" "M"
```

```
Sim_streak <- calc_streak(sim_basket)
barplot(table(Sim_streak))
```



On your own

Comparing Kobe Bryant to the Independent Shooter

*Since the sample takes a randomizes the simulation everytime, I have captured sample data below to work on the problems

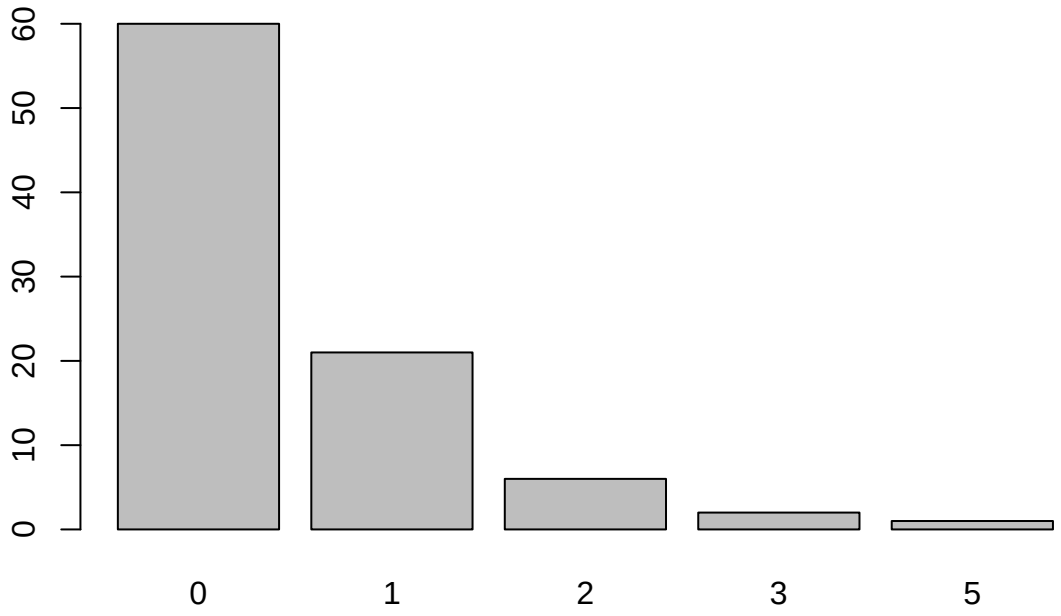
```
Sample_sim <-c("M", "M" ,"M" ,"H" ,"M","H", "M" ,"H" ,"M" ,"M", "M" ,"H" ,"H" ,"M", "M", "M", "H","M" ,
  "M", "H" ,"M", "M", "M", "M", "H" ,"H" ,"M" ,"H" ,"M",
  "M", "M", "H" ,"M", "M" ,"H", "H" ,"M" ,"H" ,"M" ,"M",
  "M", "M", "H","H", "H" ,"M", "H", "M", "M", "M" ,"M", "H" ,"M", "M" ,"M" ,"H", "M", "H", "H", "H", "H
 "M", "M", "H" ,"M" ,"M", "M" ,"M", "H" ,"H", "H", "M",
  "H" ,"M", "H", "H" ,"M", "M" ,"M", "H" ,"M", "H", "M",
 "H" ,"M", "H", "M" ,"M", "H", "M" ,"M" ,"M", "M" ,"M",
 "M" ,"M" ,"M", "H", "M", "M" ,"M", "M", "H", "M", "M",
 "M", "M", "M", "M" ,"M", "M", "H" ,"M", "M", "M", "M",
 "M")

table(Sample_sim)
```

```
## Sample_sim
##  H  M
## 44 89
```

```
calc_streak <- function(x){
  y <- rep(0,length(x))
```

```
  y[x == "H"] <- 1
  y <- c(0, y, 0)
  wz <- which(y == 0)
  streak <- diff(wz) - 1
  return(streak)
}
```

```
Sample_sim_streak <- calc_streak(Sample_sim)
barplot(table(Sample_sim_streak))
```




```
summary(Sample_sim_streak)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0000  0.0000  0.0000  0.4889  1.0000  5.0000
```

```
summary(kobe_streak)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0000  0.0000  0.0000  0.7632  1.0000  4.0000
```

Using `calc_streak`, compute the streak lengths of `sim_basket`.

- Describe the distribution of streak lengths. What is the typical streak length for this simulated independent shooter with a 45% shooting percentage? How long is the player's longest streak of baskets in 133 shots?

  A typical streak for the simulated is 0 and for actual streak is 1, which is Similar to kobe's shots during the 2009 NBA Finals, most weren't streaks. It was different out come everytime I ran it. For the current model, the longest streak is 5.

- If you were to run the simulation of the independent shooter a second time, how would you expect its streak distribution to compare to the distribution from the question above? Exactly the same? Somewhat similar? Totally different? Explain your reasoning.

  I did run it several times. Each time, the distribution of streak length varied. The longest streak sometimes was 7 and sometimes was 3. A better way to compate would be to run it even more times and take the average of the simulations.

- How does Kobe Bryant's distribution of streak lengths compare to the distribution of streak lengths for the simulated shooter? Using this comparison, do you have evidence that the hot hand model fits Kobe's shooting patterns? Explain.

  Compared to Kobe's distribution, the simulation model is higher in streaks of 0 and 1, but when it comes to streaks over 2, the simiulation model is fewer. In the simulation, streaks of 0, 1, 2, 3,and 5 happend about 60,20,6,3,and 1 respectfully. Compared to kobe's data, streaks of 0, 1, 2, 3, and 4 happend about 40, 24, 5, 5, and 1 times.

  Using this model, I can say I do have evidence that the hot hand model fits kobes shooting patterns. Kobe is more likely to have streak of 1. Through the patterns for streak of 2 are similar, probablity of having a streak of 3 is more likely for Kobe than the simulation. When we look at the mean of the two distribution, Kobe's streak length distrubtion is 0.7632, while the similation is 0.4889. Because the simulation is a random outcome, we need to look at more data to further conclude whether the hot hand model is real.