Week 3 Assigment

laknath Gunathilake

January 23, 2020

Loading Kobe Dataset

```
download.file("http://www.openintro.org/stat/data/kobe.RData", destfile = "kobe.RData")
load("kobe.RData")
```

```
      outcomes <- c("H", "M")
      sim_basket <- sample(outcomes, size = 1, replace = TRUE, prob = c(.5, 0.5))
```

- To make a valid comparison between Kobe and our simulated independent shooter, we need to align both their shooting percentage and the number of attempted shots.
- 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 .

```
      set.seed(10)
      sim_basket <- sample(outcomes, size = 133, replace = TRUE, prob = c(.45, 0.55))
```

Note that we've named the new vector `sim_basket` , the same name that we gave to the previous vector reflecting a shooting percentage of 50%. In this situation, R overwrites the old object with the new one, so always make sure that you don't need the information in an old vector before reassigning its name.

With the results of the simulation saved as sim_basket , we have the data necessary to compare Kobe to our independent shooter. We can look at Kobe's data alongside our simulated data.

```
      kobe$basket
```

```
##   [1] "H" "M" "M" "H" "H" "M" "M" "M" "M" "H" "H" "H" "M" "H" "H" "M" "M" "H"
##  [19] "H" "H" "M" "M" "H" "M" "H" "H" "H" "M" "M" "M" "M" "M" "M" "H" "M" "H"
##  [37] "M" "M" "H" "H" "H" "H" "M" "H" "M" "M" "H" "M" "M" "H" "M" "M" "H" "M"
##  [55] "H" "H" "M" "M" "H" "M" "H" "H" "M" "H" "M" "M" "M" "H" "M" "M" "M" "M"
##  [73] "H" "M" "H" "M" "M" "H" "M" "M" "H" "H" "M" "M" "M" "M" "H" "H" "H" "M"
##  [91] "M" "H" "M" "M" "H" "M" "H" "H" "M" "H" "M" "M" "H" "M" "M" "M" "H" "M"
## [109] "H" "H" "H" "M" "H" "H" "H" "M" "H" "M" "H" "M" "M" "M" "M" "M" "M" "H"
## [127] "M" "H" "M" "M" "M" "M" "H"
```

```
      sim_basket
```

```
##   [1] "M" "M" "M" "H" "M" "M" "M" "M" "H" "M" "H" "H" "M" "H" "M" "M" "M" "M"
##  [19] "M" "H" "H" "H" "H" "M" "M" "H" "H" "M" "H" "M" "M" "M" "M" "M" "H" "M" "H"
##  [37] "H" "H" "H" "M" "M" "M" "M" "H" "M" "M" "M" "M" "M" "M" "H" "M" "H" "M" "M"
##  [55] "M" "H" "M" "M" "M" "M" "M" "M" "M" "M" "M" "H" "H" "H" "M" "M" "M" "M" "M"
##  [73] "H" "M" "M" "M" "M" "M" "H" "M" "M" "M" "H" "M" "M" "H" "M" "M" "M" "H" "M" "M" "M"
##  [91] "M" "H" "H" "M" "M" "H" "M" "M" "M" "H" "M" "M" "M" "M" "M" "M" "H" "H" "H" "M"
## [109] "H" "M" "H" "H" "M" "M" "H" "H" "M" "H" "H" "H" "M" "M" "M" "H" "M" "M"
## [127] "H" "M" "H" "H" "H" "M" "H"
```

Both data sets represent the results of 133 shot attempts, each with the same shooting percentage of 45%. We know that our simulated data is from a shooter that has independent shots. That is, we know the simulated shooter does not have a hot hand. Each Shot in an independedt computer generated random number following our Hit probability weight

Compare and Analyze the Two.What is you conclusion????

```
      table(kobe$basket)/1000
```

```
##
##     H     M
## 0.058 0.075
```

```
      table(sim_basket)/1000
```

```
## sim_basket
##     H     M
## 0.050 0.083
```

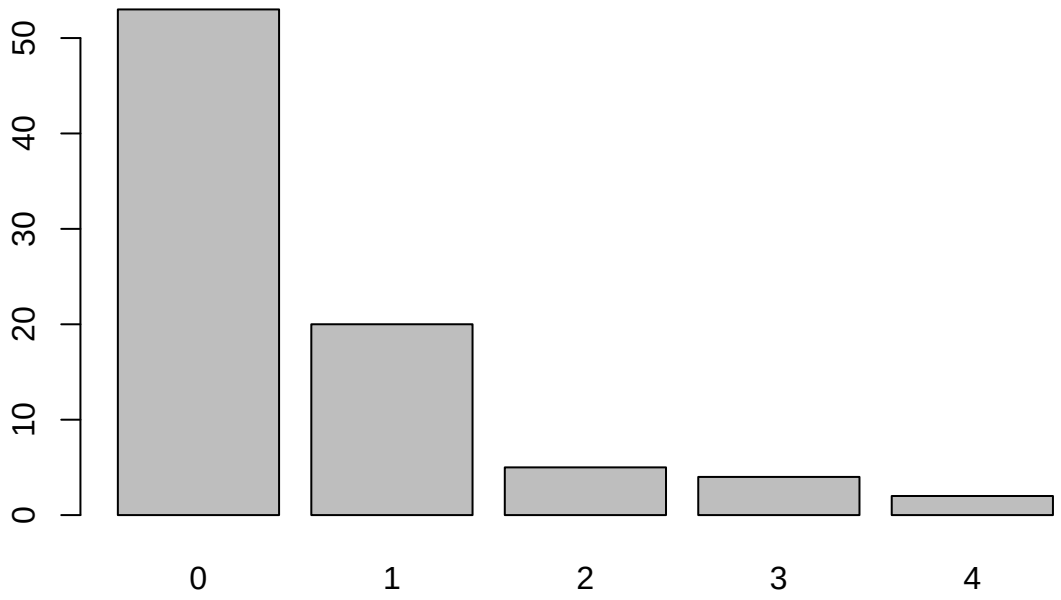
Based on this simulation Kobe has more hits than the simulated player.

Using calc_streak , compute the streak lengths of sim_basket .

```
      table(calc_streak(sim_basket))
```

```
##
##  0  1  2  3  4
## 53 20  5  4  2
```

```
       barplot(table(calc_streak(sim_basket)))
```



1. Describe the distribution of streak lengths. What is the typical streak length for this

- The distribution of the stream length is right skewed. Meaning, there are fewer times a player would have consecutive hits(baskets)
- The typical streak length (mode) for the simulated player is 1

2.simulated independent shooter with a 45% shooting percentage? How long is the player's longest streak of baskets in 133 shots?*

- The simulated players longest streak is four shots.

3.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. (SEEDING THE SIMULATION????)

- If the set.seed function is kept the same we would expect the same results, and consequently, the same distribution. However, if we do not set.seed. That distribution can change because it is a random process

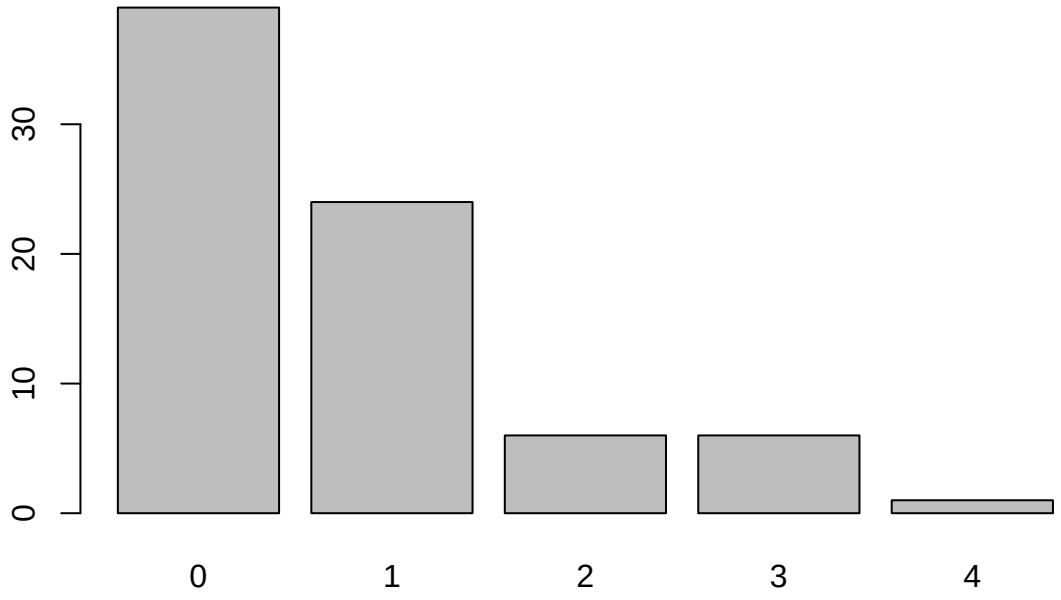
4. How does Kobe Bryant's distribution of streak lengths compare to the distribution of streak lengths for the simulated shooter? Using this comparison, do you have evidence that the hot hand model fits Kobe's shooting patterns? Explain

```
    table(calc_streak(kobe$basket))
```

```
##
```

```
##  0  1  2  3  4
## 39 24  6  6  1
```

```
barplot(table(calc_streak(kobe$basket)))
```



The shape of the streak distribution of the simulated player and kobe seems to have the same shape. There is no direct evidence to say that kobe has a hot hand.