

Data 606 - Lab 3

Cameron Smith

2020-09-10

```
library(tidyverse)
library(openintro)
```

Exercise 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?

A streak of 1 means that one shot was made followed by a miss. A streak of zero means a miss occurred, which would have either been the 1st shot in the sequence or the 1st shot immediately after another miss.

Exercise 2

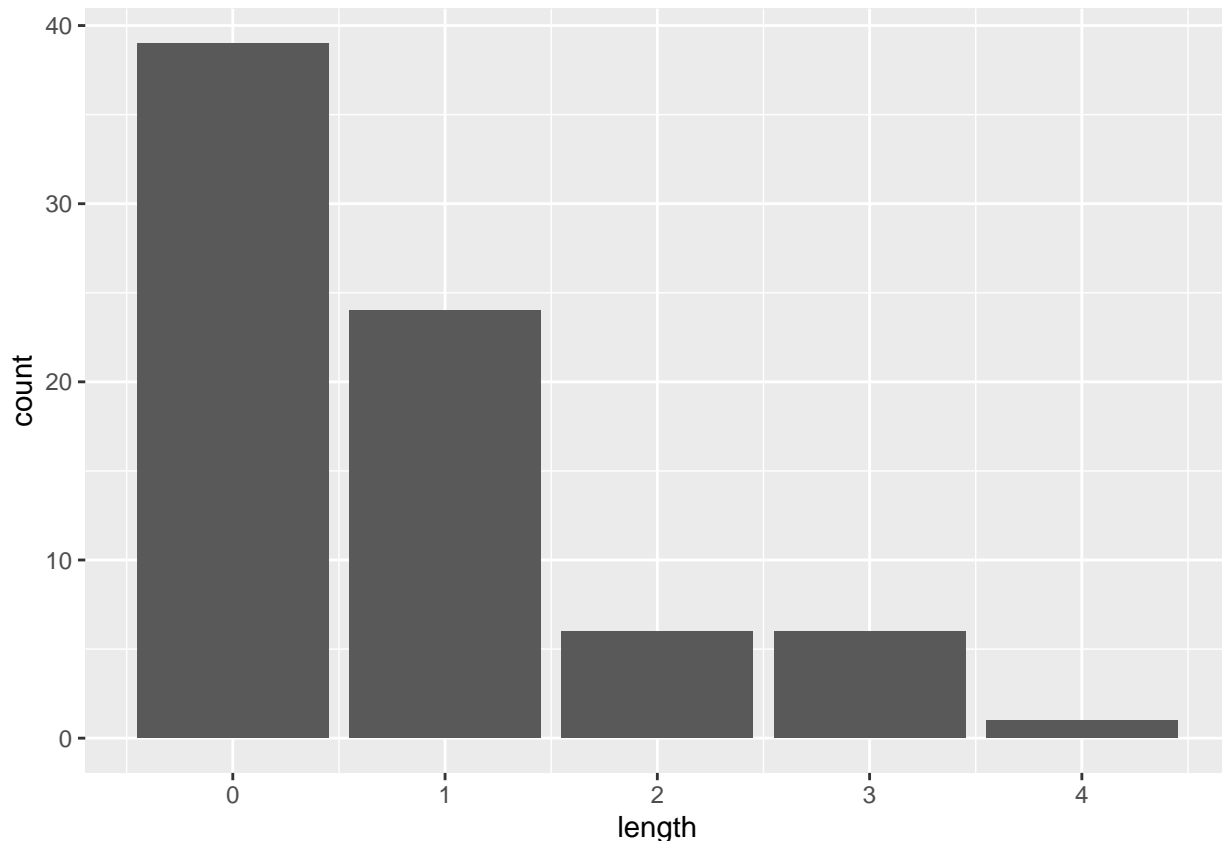
Describe the distribution of Kobe's streak lengths from the 2009 NBA finals. What was his typical streak length? How long was his longest streak of baskets? Make sure to include the accompanying plot in your answer.

It is a geometric distribution with the count of shots declining as the streak length increases (an inverse relationship). Typical streak length depends on the definition of typical, but more often than not he missed. The median is 0, and the mean is .76. The most shots he made in a row was 4.

```
kobe_streak <- calc_streak(kobe_basket$shot)
summary(kobe_streak)
```

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

```
ggplot(data = kobe_streak, aes(x = length)) +
  geom_bar()
```



Exercise 3 In your simulation of flipping the unfair coin 100 times, how many flips came up heads? Include the code for sampling the unfair coin in your response. Since the markdown file will run the code, and generate a new sample each time you Knit it, should also “set a seed” before you sample. Read more about setting a seed below.

The unfair coin came up with heads 18 times and tails 82 times.

```
set.seed(100)
coin_outcomes <- c("heads", "tails")
sim_unfair_coin <- sample(coin_outcomes, size = 100, replace = TRUE,
                          prob = c(0.2, 0.8))
sim_unfair_coin %>% table()
```

```
## .
## heads tails
##    18    82
```

Exercise 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`.

The ‘prob’ argument must be added to the sample function so that a vector is created to indicate the possibility % for each item, i.e. ‘hit’ or ‘miss’.

```
set.seed(100)
shot_outcomes <- c("H", "M")
```

```
sim_basket <- sample(shot_outcomes, size = 133, replace = TRUE, prob = c(0.45, 0.55))
sim_basket %>% table()
```

```
## .
## H M
## 61 72
```

Exercise 5

Using `calc_streak`, compute the streak lengths of `sim_basket`, and save the results in a data frame called `sim_streak`.

```
sim_streak <- calc_streak(sim_basket)
```

Exercise 6

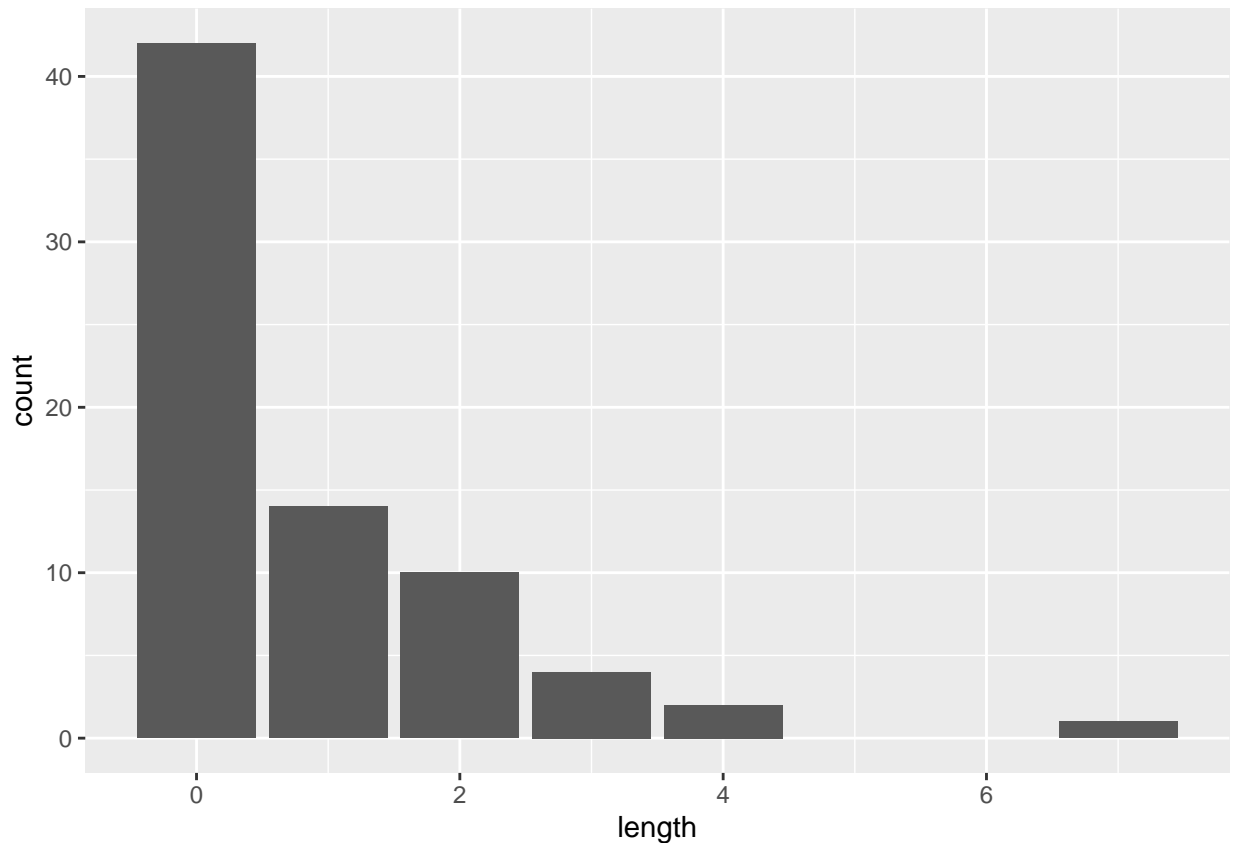
Describe the distribution of streak lengths. What is the typical streak length for this simulated independent shooter with a 45% shooting percentage? How long is the player's longest streak of baskets in 133 shots? Make sure to include a plot in your answer.

The distribution is geometric, with an inverse relationship of # of shots to streak length. The median is 0 (i.e. the independent shooter missed more than he hit), with a mean of .83. The longest streak was 7.

```
summary(sim_streak)
```

```
##      length
## Min.   :0.0000
## 1st Qu.:0.0000
## Median :0.0000
## Mean   :0.8356
## 3rd Qu.:1.0000
## Max.   :7.0000
```

```
ggplot(data = sim_streak, aes(x = length)) +
  geom_bar()
```



Exercise 7

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 would expect it to be a similar distribution, though not (usually) identical, with the assumption that a different random seed was used each time. This is due to random chance, the same logic as a coin with a 50% chance of heads or tails having similar, but slightly different, outcomes when flipped 100 times in multiple trials.

Exercise 8

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.

There is not sufficient evidence that the hot hand model fits Kobe's shooting patterns. The distributions of the simulated shooter and of Kobe are similar but the simulated shooter actually had a longer streak and a higher median. The means (average outcome) were the same for each - zero - meaning that both of them missed more than they hit. Kobe did have 76 total straks in comparison to the simulation's 73, but the difference could just be due to random chance.

```
# Calculate the total number of streaks
nrow(kobe_streak)
```

```
## [1] 76  
nrow(sim_streak)
```

```
## [1] 73  
-End of Lab 3-
```