Data606_Lab3

Code

Sean Connin

2021-02-18

Hide

```
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 length 1 means that Kobe made a shot in which his previous shot missed as well as his following shot. For example, M|HM|

A streak of length 0 suggests that he had a miss that was preceded by a miss. For example, M|M. Each consecutive miss would count as a streak length of 0

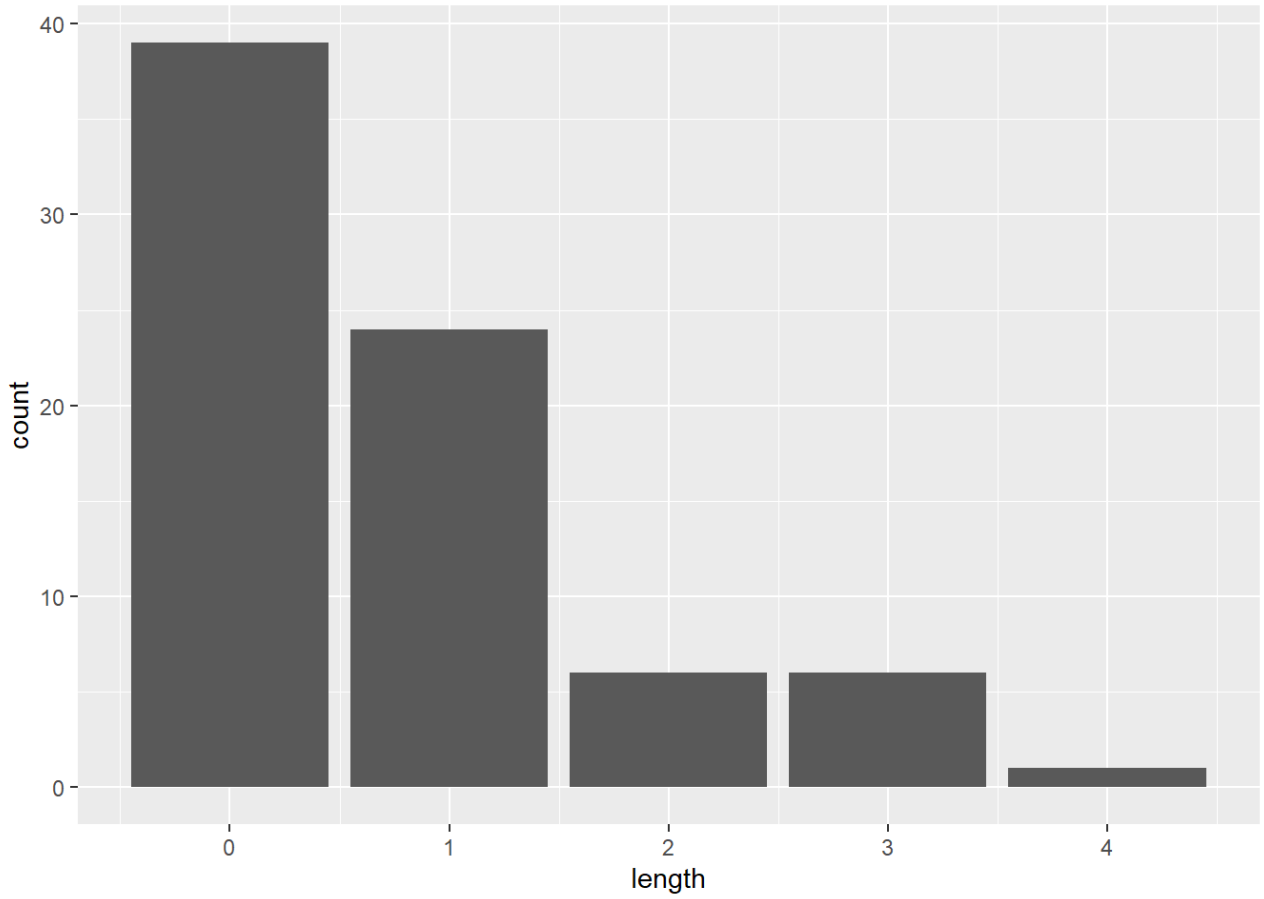
Hide

```
# Evaluation of Kobe's streak lengths

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
```

Hide

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



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.

The distribution of Kobe's streaks is right skewed and unimodal.

Kobe's typical streak length was 0 based on count frequency.

Kobe's longest streak was 4 consecutive shots made in a row.

See previous graph

Exercise 3.

In your simulation of flipping the unfair coin 100 times, how may flips came up heads.

Drawing on the code below: the count for heads after 100 simulations was 23.

Hide

```
coin_outcomes <-c("heads", "tails")
sample(coin_outcomes, size=1, replace=TRUE)
```

```
## [1] "heads"
```

Hide

```
set.seed(1966)
sim_fair_coin<-sample(coin_outcomes, size=100, replace=TRUE, prob = c(0.2,0.8))
sim_fair_coin
```

```
##   [1] "heads" "tails" "tails" "tails" "tails" "tails" "tails" "tails" "tails"
##  [10] "tails" "tails" "tails" "tails" "tails" "tails" "tails" "tails" "heads"
##  [19] "tails" "heads" "tails" "tails" "tails" "tails" "tails" "heads" "heads"
##  [28] "tails" "tails" "tails" "heads" "tails" "heads" "tails" "tails" "tails"
##  [37] "tails" "heads" "heads" "tails" "tails" "tails" "heads" "tails" "tails"
##  [46] "tails" "tails" "tails" "tails" "tails" "heads" "tails" "tails" "heads"
##  [55] "tails" "tails" "tails" "tails" "tails" "heads" "tails" "tails" "tails"
##  [64] "heads" "tails" "tails" "tails" "tails" "tails" "tails" "heads" "tails"
##  [73] "heads" "tails" "tails" "tails" "heads" "tails" "tails" "heads" "tails"
##  [82] "heads" "tails" "tails" "tails" "tails" "heads" "tails" "tails" "tails"
##  [91] "heads" "tails" "tails" "tails" "tails" "tails" "heads" "heads" "tails"
## [100] "tails"
```

Hide

```
table(sim_fair_coin)
```

```
## sim_fair_coin
## heads tails
##    23    77
```

Exercise 4.

Simulating a basketball player who has independent shots uses the same mechanism that you used to simulate a coin flip.

What change needs to be made to the sample function so that it reflects a shooting percentage of 45%?

Answer: we need to include the prob argument in our sample function.

Hide

```
# simulate the independent shooter

shot_outcomes <-c("H","M")
set.seed(1964)
sim_basket <- sample(shot_outcomes, size=133, replace=TRUE)
sim_basket
```

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

Hide

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

```
## sim_basket
##  H  M
## 68 65
```

Hide

```
# simulate Kobe's shooting percentages

shot_outcomes <- c("H", "M")
set.seed(1968)
sim_basket <- sample(shot_outcomes, size = 133, replace = TRUE, prob = c(.45,.55))
sim_basket
```

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

Hide

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

```
## sim_basket
##  H  M
## 60 73
```

Exercise 5.

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

Hide

```
sim_bask<- as.data.frame(sim_basket)%>%rename("shots"=sim_basket)

sim_streak <- calc_streak(sim_bask$shots)

summary(sim_streak)
```

```
##      length
##  Min.   :0.0000
##  1st Qu.:0.0000
##  Median :0.0000
##  Mean   :0.8108
##  3rd Qu.:1.0000
##  Max.   :5.0000
```

Exercise 6.

Describe the distribution of streak lengths.

What is the typical streak length for this simulated independent shooter with a 45% shooting percentage?

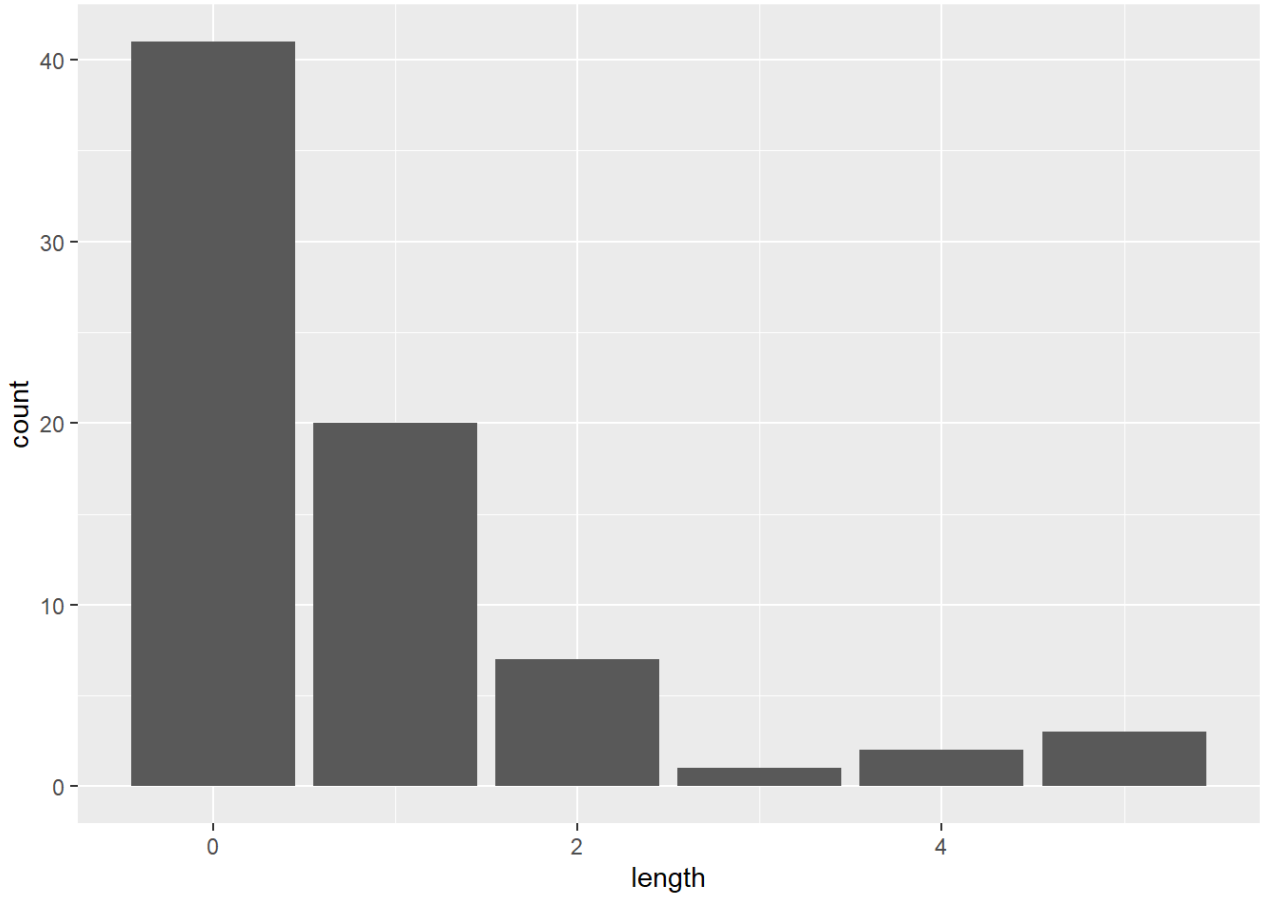
The typical streak length for the simulated independent shooter is 0.

How long is the player's longest streak of baskets in 133 shots?

The simulated player's longest consecutive streak is 6 shots.

Hide

```
#plot simulated players streak lengths

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 expect there to be some variation in the results (length counts) due to the fact that the model is probabilistic.

In other words, the calc_streak function draws on random numbers and, as a result, does not return a single (deterministic) solution.

I expect the overall shape of the distribution would be consistent between runs.

Exercise 8.

How does Kobe Bryant's distribution of streak lengths compare to the distribution of streak lengths for the simulated shooter?

They are very similar in terms of the counts, range of streak lengths, and shape of the distribution.

Using this comparison, do you have evidence that the hot hand model fits Kobe's shooting patterns? Explain.

Based on the comparison here, I do not find evidence that the hot hand model fits Kobe's shooting patterns.

If the model was correct, I would expect that the count & length of shooting streaks would track in a positive manner. I would also expect that Kobe's shooting streaks would, on average, be longer than those produced by the simulated independent shooter.