DA 460 - Lab 2

Marjorie Blanco

July 07, 2018

# Handout 2 R: Probability

## Getting Started

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

# sequence of hits and misses from Kobe's nine shot attempts in the first quarter  
kobe$basket[1:9]

## [1] "H" "M" "M" "H" "H" "M" "M" "M" "M"

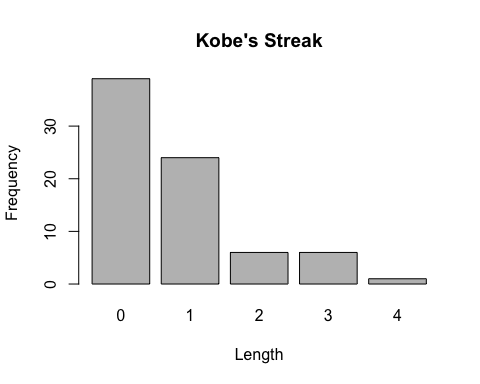
### Exercise 1

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?

“H” “M” | “M” | “H” “H” “M” |“M” | “M” | “M”

A streak length of 1 has a hit followed by a miss. A streak length of 0 has a miss. In the example, the first is streak length of 1 followed by streak length of 0.

# calculate the lengths of all shooting streaks  
kobe\_streak <- calc\_streak(kobe$basket)  
# streak distribution  
barplot(table(kobe\_streak), main="Kobe's Streak",   
 ylab = "Frequency", xlab = "Length")



### 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?

The distribution of Kobe’s streak lengths from the 2009 NBA finals is strongly skewed to the right. The typical streak length is 0. Kobe’s longest streak of baskets is 4.

## Simulations in R

# possible outcome  
outcomes <- c("heads", "tails")  
# random sample of size 1  
sample(outcomes, size = 1, replace = TRUE)

## [1] "heads"

# simulate flipping a fair coin 100 times  
sim\_fair\_coin <- sample(outcomes, size = 100, replace = TRUE)  
  
# view the results of this simulation  
sim\_fair\_coin

## [1] "heads" "heads" "tails" "tails" "heads" "heads" "tails" "tails"  
## [9] "heads" "heads" "heads" "tails" "heads" "tails" "tails" "tails"  
## [17] "tails" "heads" "tails" "tails" "heads" "tails" "heads" "heads"  
## [25] "heads" "heads" "heads" "heads" "tails" "tails" "heads" "heads"  
## [33] "heads" "tails" "heads" "tails" "heads" "heads" "heads" "heads"  
## [41] "tails" "heads" "tails" "heads" "tails" "tails" "heads" "heads"  
## [49] "heads" "heads" "heads" "heads" "heads" "tails" "heads" "heads"  
## [57] "heads" "heads" "tails" "heads" "heads" "tails" "heads" "tails"  
## [65] "tails" "heads" "tails" "heads" "tails" "tails" "tails" "heads"  
## [73] "tails" "heads" "tails" "tails" "heads" "heads" "heads" "tails"  
## [81] "heads" "tails" "tails" "tails" "tails" "tails" "tails" "heads"  
## [89] "heads" "heads" "tails" "tails" "heads" "tails" "heads" "tails"  
## [97] "tails" "tails" "heads" "tails"

table(sim\_fair\_coin)

## sim\_fair\_coin  
## heads tails   
## 54 46

# simulate an unfair coin that we know only lands heads 20% of the time  
sim\_unfair\_coin <- sample(outcomes, size = 100, replace = TRUE, prob = c(0.2, 0.8))  
  
# view the results of this simulation  
sim\_unfair\_coin

## [1] "heads" "tails" "tails" "tails" "tails" "tails" "tails" "tails"  
## [9] "tails" "tails" "tails" "heads" "tails" "tails" "tails" "tails"  
## [17] "tails" "heads" "heads" "tails" "tails" "tails" "tails" "tails"  
## [25] "tails" "tails" "heads" "heads" "tails" "tails" "tails" "tails"  
## [33] "tails" "tails" "tails" "tails" "tails" "heads" "tails" "heads"  
## [41] "tails" "heads" "tails" "heads" "heads" "tails" "tails" "tails"  
## [49] "tails" "tails" "tails" "heads" "tails" "tails" "tails" "heads"  
## [57] "heads" "tails" "tails" "tails" "heads" "tails" "tails" "tails"  
## [65] "tails" "tails" "tails" "tails" "tails" "heads" "tails" "tails"  
## [73] "tails" "heads" "heads" "tails" "tails" "tails" "tails" "heads"  
## [81] "heads" "heads" "tails" "tails" "tails" "tails" "heads" "tails"  
## [89] "tails" "tails" "tails" "tails" "tails" "tails" "tails" "tails"  
## [97] "tails" "tails" "heads" "tails"

table(sim\_unfair\_coin)

## sim\_unfair\_coin  
## heads tails   
## 23 77

### Exercise 3

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

In the simulation of flipping the unfair coin 100 times, 23 flips came up heads. This is consistent with the probality of 20% that a coin would land on heads.

## Simulating the Independent Shooter

# possible outcome H: hit M: missed  
outcomes <- c("H", "M")  
# random sample of size 1  
sim\_basket <- sample(outcomes, size = 1, replace = TRUE)

### 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.

sampleSize = 133  
shooting = 0.45  
sim\_basket <- sample(outcomes, size = 133, replace = TRUE, prob = c(shooting, 1 - shooting))  
table(sim\_basket)

## sim\_basket  
## H M   
## 61 72

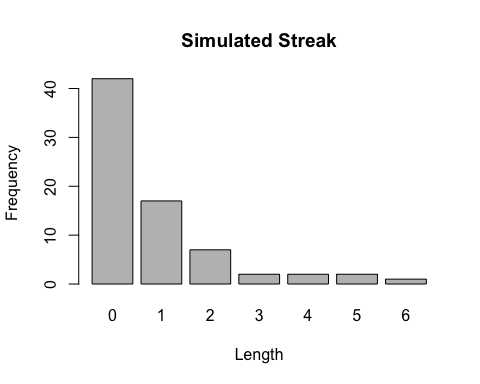
# kobe's shots  
kobe$basket

## [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"

#simulated shooter does not have a hot hand  
sim\_basket

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

# calculate the lengths of all shooting streaks  
sim\_streak <- calc\_streak(sim\_basket)  
# streak distribution  
barplot(table(sim\_streak), main="Simulated Streak",   
 ylab = "Frequency", xlab = "Length")



Suppose there are twelve multiple choice questions in an English class quiz. Each question has five possible answers, and only one of them is correct. Find the probability of having four or less correct answers if a student attempts to answer every question at random.

Solution Since only one out of five possible answers is correct, the probability of answering a question correctly by random is 1/5=0.2. We can find the probability of having exactly 4 correct answers by random attempts as follows.