

Name: Samuel L. Peoples

Collaborators: (Any one you worked with)

## R Lab 2: Probability

Please answer all the Exercises and the questions from the “On Your Own” section. If you use any graphs or charts to justify your answer, please include them.

**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?

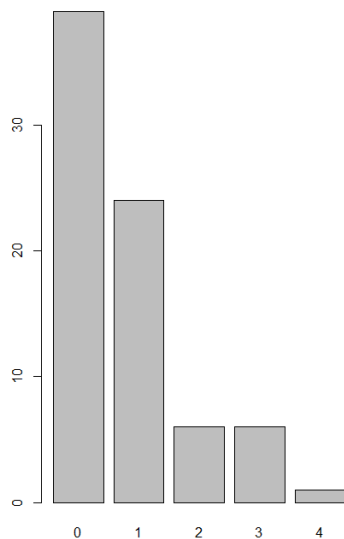
```
kobe$basket[1:9]
```

A streak is how many back-to-back shots were made. A streak of length 1 is “|HM|”, where there is one “hit” followed by a “miss”; similarly, a streak of length 0 is “|M|”, which is a single miss, not preceded by a “H”.

**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?

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

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



Kobe’s distribution is skewed right, with a mode of “0”; the summary shows that 75% of his shots were a streak of 1 or less. Kobe’s longest streak was 4 baskets.

**Exercise 3:** In your simulation of flipping the unfair coin 100 times, how many flips came up heads?

```
> sim_unfair_coin <- sample(outcomes, size = 100,
+replace = TRUE, prob = c(0.2, 0.8))
> sim_unfair_coin
> table(sim_unfair_coin)
```

sim_unfair_coin	
heads	tails
19	81

In the unfair trials, where “heads” was defined with a probability of .2, there were 19/100 results which yielded “heads”.

**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`.

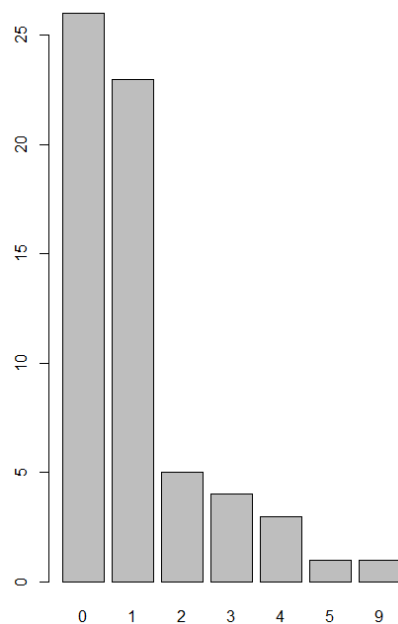
```
> sim_basket <- sample(outcomes, size = 133, replace = TRUE, prob = c(.45, .55))
```

**On Your Own:**

1) 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?

```
> kobe_streak <- calc_streak(sim_basket)
> barplot(table(kobe_streak))
> summary(kobe_streak)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.000	0.000	1.000	1.127	1.000	9.000



The expected streak length is 1, with a maximum streak of 9.

2) 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.

In this situation, I believe that the results would remain skewed to the right, with an expected streak of 1. Because the probability is not dependent on the previous trials, it is uncertain whether the maximum streaks would be similar, but the overall behavior of the simulation should be consistent.

3) 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.

Running the simulation for 500 iterations allows us to form a more confident opinion of the independent shooter, as well as providing us with a more confident summary statistic. In the simulation, we can see that the mean and median are quite similar, settling between streaks of zero and 1.

In the independent simulation, 75% of the streaks were of length one or less, as the distribution having similar shapes and skew. Because the simulation is independent, and the results closely match Kobe's data, we can safely say that Kobe's shots are independent, and he does *not* have a hot hand.

```
> sim_basket <- sample(outcomes, size = 500, replace = TRUE, prob = c(.45, .55))
> sim_basket
> kobe_streak <- calc_streak(sim_basket)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.0000	0.0000	0.0000	0.8285	1.0000	6.0000