Day 15 - Intro to Hypothesis Testing cont.

Learning outcomes and terms to know

Today, we will further develop our intuition with hypothesis testing, motivated by an interesting research question that requires a unique statistic.

Hot Hand: phenomenon or fallacy?



Photo of Curry

To motivate today's activity, let's watch this see this brief clip.

Basketball players who make several shots in succession are described as having a *hot hand*. Fans and players long believed in the hot hand phenomenon, **which suggests that each shot is not independent of the next.** However, a 1985 paper by Gilovich, Vallone, and Tversky collected evidence that contradicted this belief and showed that successive shots are independent events. More recently, Miller and Sanjurjo (2018) revisited this topic and suggested that there may be a small hot hand effect, but it is weaker than people intuitively think.

We are not going to solve the hot hand debate today, however what we *will* do is think about one way to approach questions like this using statistics. The goal for this activity is to understand the steps for conducting a hypothesis test to answer a research question.

Step 1: Frame the research question in terms of hypotheses

1. Write out the null and alternative hypotheses for the hot hand question in words.

> H_0: Curry's shots are independent; whether he makes a shot does not depend on whether he made his previous shots (no hot hand).
> H_a: Curry's shots are not independent; making shots makes him more likely to make his next shot (a hot hand), which would show up as longer streaks of hits than we'd expect by chance.

Step 2: Collect data

We will work with data from the NBA game between the Golden State Warriors and the Atlanta Hawks that took place on February 3, 2024. In particular, we will look at 38 shot attempts by Steph Curry over the course of the game and investigate whether Curry appears to have a hot hand. The data are stored in the `curry` object; we are interested in the `outcome` column (H for "hit" and M for "miss").

```
library(tidyverse)
curry <- read_csv("curry_shots.csv")

head(curry, 10)
```

```
# A tibble: 10 x 6
   date     opponent quarter time   description                          outcome
   <chr>    <chr>    <chr>   <time> <chr>                                <chr>
 1 2/3/2024 ATL      1       08:40  S. Curry misses 27' pullup three     M
 2 2/3/2024 ATL      1       08:19  S. Curry misses 26' three            M
 3 2/3/2024 ATL      1       07:14  S. Curry makes running finger roll   H
 4 2/3/2024 ATL      1       04:48  S. Curry makes 26' three             H
 5 2/3/2024 ATL      1       04:03  S. Curry makes 11' pullup            H
 6 2/3/2024 ATL      2       05:47  S. Curry misses 26' three            M
 7 2/3/2024 ATL      2       05:16  S. Curry makes 27' three             H
 8 2/3/2024 ATL      2       04:04  S. Curry misses driving finger roll  M
 9 2/3/2024 ATL      2       03:47  S. Curry makes cutting finger roll   H
10 2/3/2024 ATL      2       01:59  S. Curry makes 25' pullup three      H
```

2. Look at the data with your group and discuss whether Curry appears to have a hot hand?

> Answers will vary. Curry made 22 of his 38 shots (about 58%). There is one eye-catching stretch: he made 8 shots in a row (shot 21 at the end of the 3rd quarter through shot 28 in the 4th). That *looks* like a hot hand, but it's hard to tell by eye whether a run like that is unusual for someone who makes 58% of his shots.

Looking at the string of hits and misses, it can be difficult to gauge whether or not it seems like Curry was shooting with a hot hand. One method for approaching this question is to consider that hot hand shooters go on shooting streaks. For this activity, we define the length of the shooting streak to be the *number of consecutive baskets made until a miss occurs.*

For example, in these data, the outcomes of the first 11 shots are:

M | M | H H H M | H M | H H M |

In the first 11 shots there are 5 streaks, separated by "|"; the streaks lengths are zero, zero, three, one, and two.

3. What does a streak length of 1 mean? How many hits and misses are in a streak of 1? What about a streak of 0?

> A streak of length 1 is one hit followed by a miss (H M): 1 hit and 1 miss. A streak of length 0 is a miss that is not preceded by a hit, i.e., a miss right after another miss (or a miss on the first shot): 0 hits and 1 miss. Every streak ends with exactly one miss.

4. How long would a streak need to be for you to feel convinced that Curry was shooting with a hot hand during this game?

> Answers will vary. The point is that it's hard to answer without knowing what streak lengths we'd expect *just by chance* from a player who makes 22 of 38 shots. We need a model for what happens under H_0.

5. Calculate the streak lengths using the `calc_streak` function from the `openintro` package. In these data, what is the longest streak before a miss? What is the typical streak length? Make sure to include a plot of all of the streak lengths to support your answer.

```
library(openintro)

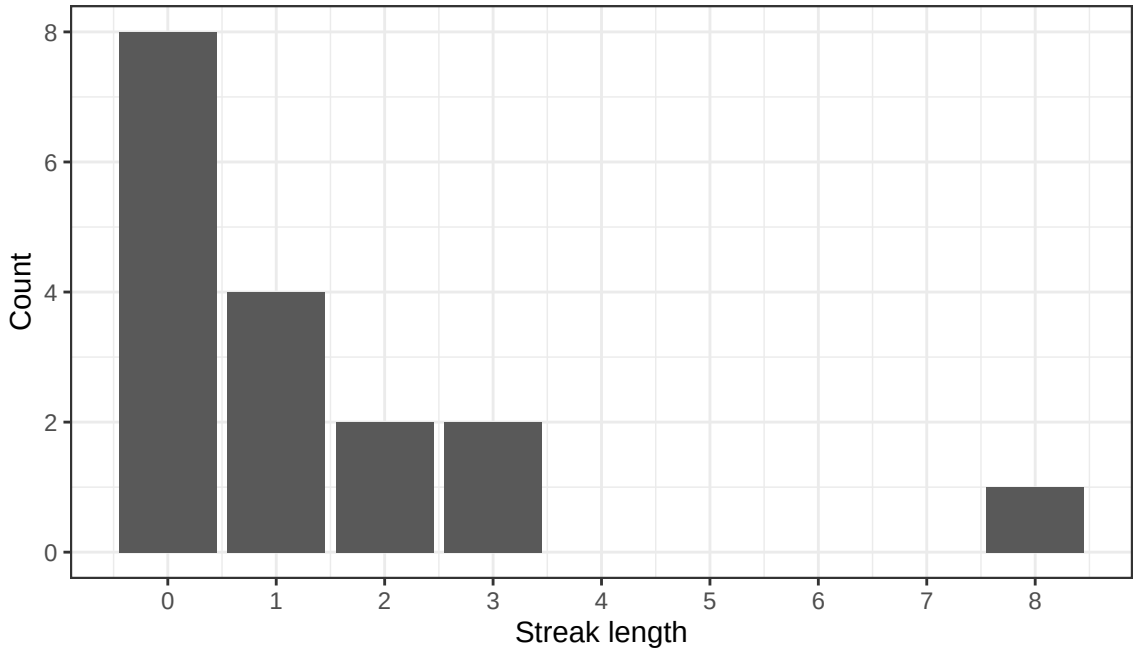
# calculate all streaks - see ?calc_streak
curry_streaks <- calc_streak(curry$outcome)

# longest and typical streak lengths
curry_streaks |>
  summarize(longest = max(length), mean = mean(length), median = median(length))
```

```
  longest     mean median
1       8 1.294118      1
```

```
# plot the streak lengths
curry_streaks |>
  ggplot(aes(x = length)) +
  geom_bar() +
```

```
  scale_x_continuous(breaks = 0:8) +
  labs(x = "Streak length", y = "Count") +
  theme_bw()
```



The longest streak is 8. The typical streak is short: the most common streak length is 0 (8 of the 17 streaks), the median is 1, and the mean is about 1.3. The streak of 8 stands out from the rest of the distribution.

Step 3: Model the randomness that would occur under H_0

6. With your group, discuss the following question: *If the shots are independent (i.e. the null hypothesis is true), does the order of the shots matter?* Why or why not?

   No. If the shots are independent, Curry has the same chance of making each shot regardless of what happened before, so every ordering of his 22 hits and 16 misses is equally plausible. That means we can shuffle the order of the shots to create data that look like what we'd see under H_0, while keeping his overall shooting percentage fixed.

7. With your group, discuss the following code. What does this code do?

```
set.seed(10092025)
newdata <- sample(curry$outcome, length(curry$outcome), replace = F)
newdata
```

```
 [1] "H" "M" "H" "M" "H" "M" "H" "M" "M" "H" "M" "M" "M" "H" "H" "H" "M" "H" "M"
```

```
 [20] "M" "H" "H" "H" "M" "H" "H" "H" "M" "H" "H" "M" "H" "M" "H" "M" "H" "H" "H"
```

> It randomly shuffles (samples without replacement) Curry's 38 shot outcomes. The new sequence has the same 22 hits and 16 misses, just in a random order. This is one simulated game under the null hypothesis that the shots are independent.

8. Calculate the streak lengths for the `newdata`. What is the longest streak that we see in `newdata`?

```
new_streaks <- calc_streak(newdata)
max(new_streaks$length)
```

```
[1] 3
```

> The longest streak in this one shuffled sequence is 3.

9. Does the new streak lengths calculated in question 8 change your answer about whether you think Curry was shooting with a hot hand during this game? Why or why not?

> Maybe a little. The observed longest streak (8) is much longer than the 3 we got from one shuffle, which might suggest a hot hand. But this is a single simulation: another shuffle could give a much longer streak. We need to repeat the shuffle many times to see how much the longest streak varies under H_0.

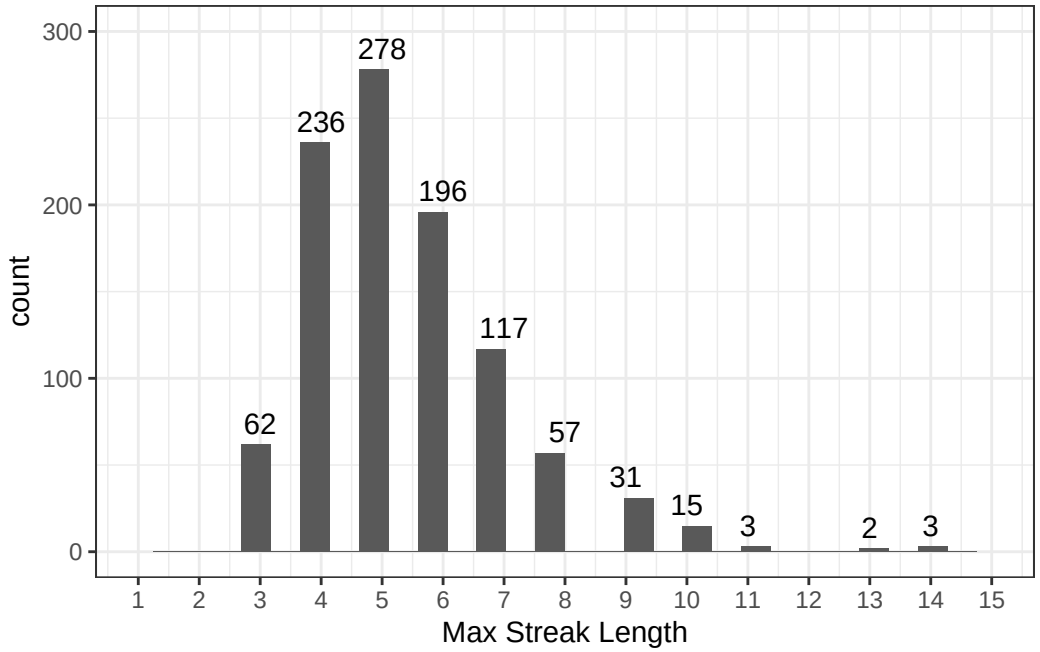
What about if we repeated this many times?

```
nsims <- 1000

max_streak <- matrix(NA, nrow = nsims, ncol = 1)
for(i in 1:nsims){
  newdata <- sample(curry$outcome, length(curry$outcome), replace = F)
  max_streak[i] <- calc_streak(newdata) %>%
    max()
}

max_streak %>%
  data.frame() %>%
  ggplot(aes(x = max_streak)) +
  geom_histogram() +
  scale_x_continuous(limits = c(1, 15),
                     breaks = c(1:15),
                     labels = c(1:15)) +
  geom_text(aes(label = after_stat(count)),
            stat = "count",
            vjust = -0.5,
            size = 4) +
```

```
  ylim(0, 300) +
  labs(x = "Max Streak Length") +
  theme_bw()
```



10. Does the above figure change your answer about whether you think Curry was shooting with a hot hand during this game? Why or why not?

> Yes, it should make us less convinced. Under independence, the longest streak in a 38-shot game typically falls between 4 and 6, but longer streaks are not rare: 8 or more happened in about 11% of the 1,000 shuffles. Curry's streak of 8 is on the high side, but it's the kind of thing that happens fairly often by chance.

11. In the code chunk below I calculate the p-value for you. In words, what does the p-value represent in the context of the problem?

```
# observed max streak length
obs_stat <- curry$outcome |>
  calc_streak() |>
  max()

# p-value calculation
mean(max_streak >= obs_stat)
```

```
[1] 0.111
```

If Curry's shots were independent (no hot hand), there is about a 0.11 probability that the longest streak in a game where he makes 22 of 38 shots would be 8 or longer, just by chance.

12. Is there evidence to suggest Curry's shots in the game are **not** independent?

No, not much. A p-value of about 0.11 provides little evidence against the null hypothesis, so there is insufficient evidence to conclude that Curry's shots in this game were not independent. Note that this does *not* show his shots were independent; we just don't have strong evidence that they weren't.

13. Comment on the scope of inference for this problem.

These data are a single game for a single player, and they are not a random sample of anything, so we can't generalize these results to Curry's shooting in other games, let alone to other players or to the hot hand in general. With only 38 shots, the test also has limited ability to detect a modest hot hand effect. Finally, shuffling assumes every shot had the same chance of going in, but Curry's shots varied a lot (layups vs. 29-foot threes), and a "hot" shooter may take harder shots, which could mask a hot hand.