Testing a Friend’s Claims

**Dataset Used:**

NBA shot logs: <https://www.kaggle.com/dansbecker/nba-shot-logs>

**Data Removed:**

* I removed a handful of columns that were “cluttering” my dataset. Essentially, I removed them because they had no use for my tests.
* The data itself was very clean already, so no further cleaning was necessary.

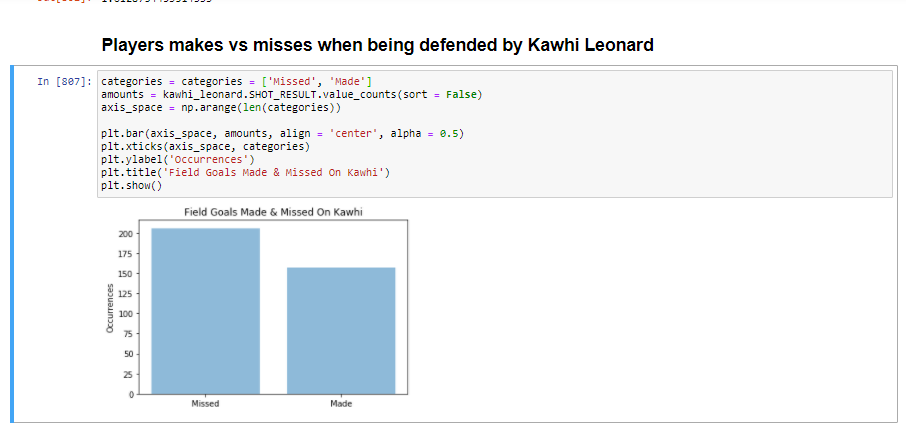
**Alternate Hypothesis Verbiage:**

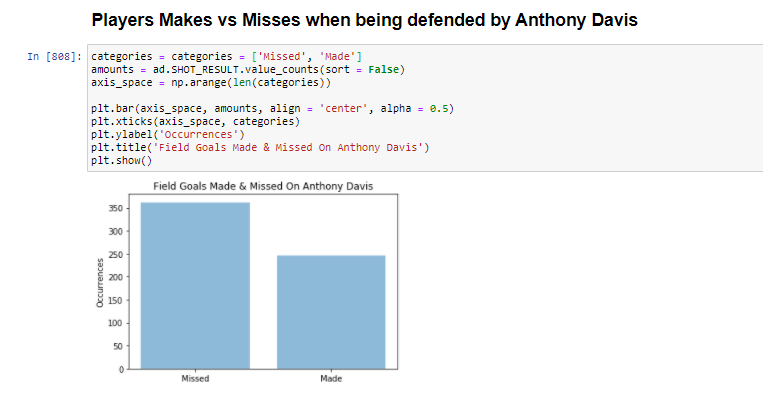
* Claim 1: “In the 2014-2015 NBA season, players defended by Kawhi Leonard shot worse than players defended by Anthony Davis” (Friend).
* Claim 2: “In the 2014-2015 NBA season, Stephen Curry, who won MVP, shot a better field goal percentage than MVP runner-up, James Harden” (Friend).

**Concern: Can my sample set’s conclusions be extended to the population?**

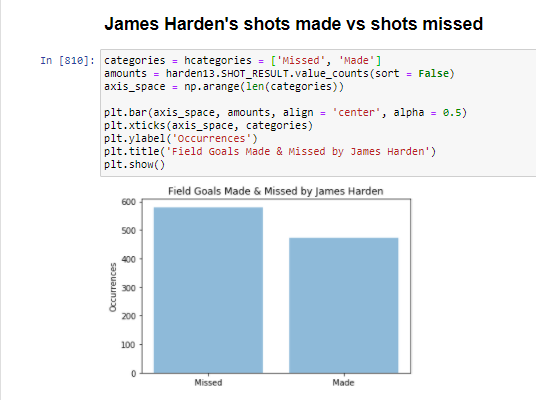
* The dataset I have is the population of NBA stats from the 2014-2015 season, and the samples I am pulling from this population should be effective in generalizing about the two claims I have (since both claims are about performances in the 2014-2015 season, not over the course of careers or anything like that).

**Using Jupyter Notebooks, I will now visualize my variables at hand:**









**Claim re-state:**

**Claim 1:**

“In the 2014-2015 NBA season, players defended by Kawhi Leonard shot worse than players defended by Anthony Davis.”

**Claim 2:**

“In the 2014-2015 NBA season, Stephen Curry shot a better field goal made percentage than James Harden.”

**Claim 1 Hypothesis:**

**Wordy:**

Null Hypothesis: The field goal made percentage of players being defended by Kawhi Leonard is greater than or equal to players being defended by Anthony Davis

Alternative Hypothesis: The field goal made percentage of players being defended by Kawhi Leonard is less than players being defended by Anthony Davis.

**Technical:**

p\_hat1 represents the proportion of field goals made (FGM) by opponents guarded by Anthony Davis. Calculated by FGM/FGA where FGA is equal to “field goal attempted” which considers both makes and misses.

p\_hat2 represents the proportion of field goals made (FGM) by opponents guarded by Kawhi Leonard. Calculated by FGM/FGA where FGA is equal to “field goal attempted” which considers both makes and misses.

Null Hypothesis: p\_hat1 >= p\_hat2

Alternative Hypothesis: p\_hat1 < p\_hat2

Note: Left tailed test

**Claim 2 Hypothesis:**

**Wordy:**

Null Hypothesis: “In the 2014-2015 NBA season, James Harden’s field goal made percentage was at least as good as Stephen Curry’s.”

Alternate Hypothesis: “In the 2014-2015 NBA season, Stephen Curry shot a higher field goal made percentage than James Harden.”

**Technical:**

p\_hat1 represents the proportion of field goals made (FGM) by Stephen Curry. Calculated by FGM/FGA where FGA is equal to “field goal attempted” which considers both makes and misses.

p\_hat2 represents the proportion of field goals made (FGM) by James Harden. Calculated by FGM/FGA where FGA is equal to “field goal attempted” which considers both makes and misses.

Null Hypothesis: p\_hat1 <= p\_hat2

Alternative Hypothesis: p\_hat1 > p\_hat2

Note: right tailed test

**Determining which Test Statistic to Utilized:**

Claim 1:

* Since I have calculated the field goal made percentage of players when being defended by Kawhi, and the field goal made percentage of players when being defended by Anthony Davis, I will use a z-score to calculate the test statistic.

Claim 2:

* Since I have calculated the field goal made percentage of Stephen Curry and James Harden, I will use a z-score to calculate the test statistic.

**Calculating my Test Statistics (T):**

Claim 1:

T = approx. -0.82389 = z (from Jupyter Notebook)

Claim 2:

T = approx. 1.61287 = z (from Jupyter Notebook)

**Calculating my P-value:**

significance level chosen: α = 0.01.

Claim1:

* C.V. = invnorm(0.01, 0, 1) = approx. -2.3263 #critical value method
* p-val = normalcdf(-E99,-0.82389,0,1) = approx. 0.2050 #p-value method

Claim 2:

* C.V. = invnorm(0.99, 0, 1) = approx. 2.3263
* p-val = normalcdf(1.61287,E99,0,1) = approx. 0.05339

**Based on Results, Make Conclusions:**

Claim 1:

* Reminder: left tailed test
* Because T-stat > C.V., we fail to reject the null hypothesis.
* Also: because p-val > significance level = 0.01, we fail to reject null hypothesis

Claim 2:

* Reminder: right tailed test
* Since our T-stat < C.V., we fail to reject the null hypothesis.
* Also: because p-val > significance level = 0.01, we fail to reject null hypothesis.
* Side note: This one was close! Had we chosen a significance level of say 0.06, we would have been able to reject the null hypothesis! But that is why we choose a significance level before-hand.

**Non-Technical Conclusions:**

Claim 1:

* Since we failed to reject the null hypothesis (part f), we conclude that there is not enough evidence to claim Kawhi Leonard caused the players he is defending to shoot worse than players who are defended by Anthony Davis in the 2014-2015 NBA season.

Claim 2:

* Since we failed to reject the null hypothesis (part f), we conclude that there is not enough evidence to claim that Stephen Curry’s field goal made percentage was greater than James Harden’s in the 2014-2015 NBA season.

**Did my Findings Support my Friends Claims?**

* No. I expected to prove that Leonard held players to a lower FGM percentage than Anthony Davis and I expected to prove that Stephen Curry had a better field goal made percentage than James Harden. Better luck next time!