Serena Williams: Aces during US Open versus Australian Open

(name redacted)

Intro to Statistics: Metropolitan State University of Denver

12/14/17

Abstract

In this study I decided to analyze the difference in the number of aces that Serena Williams scores during her US Open matches versus her Australian Open. The data was over a collective of 6 years, ranging from 2012-2017. I compared two sets of data, n= 30 from Australian Open matches from the past 6 years and n=30 from US Open matches. I hypothesized that Serena would score a higher amount of aces in the matches that were inside the United States rather than in Australia. My hypothesis was incorrect, Serena actually scored more aces while she was playing at the Australian Open. The average amount of times that Serena scored aces at the Australian Open was 9.07, while the average for the US Open was 7.53.

Introduction

This study aimed to compare the total aces scored by Serena Williams in matches that were in Australia and the United States. I will be comparing different scores from two professional tennis competitions, the Australian Open and the US Open, both of which take place in their respective countries. Professional matches consist of the best of out of 3 sets. A set is composed of the first to 6 games, and you have to win by 2 games in order to win the set. So for example in one match that Serena plays she will play her opponent until one of them reaches 6 games, and is 2 games ahead of her opponent. Say Serena wins 6 games and her opponent has only won 3 games, then Serena wins the set. They will repeat this process one more time, if Serena wins the 2^{nd} set, she wins the entire match. If Serena losses this set and her opponent wins, then they are tied and a 3^{rd} set must be played. An ace is when a player serves and the opponent does not return the ball, or touch the ball. This is an ace because it is winning a point simply because of a serve, and no actual playing. In this study I wanted to see whether Serena would differ in the amount of aces she gets when she plays in the US versus Australia. I hypothesized that Serena would score more aces in the US since this is where she is from, I assumed that a "home field" advantage would be the reason for this. I choose this subject because I am a tennis player, am knowledgeable about the sport, and because I am an avid Serena Williams fan.

Data Collection Methodology

I collected my data through matchstat.com, a website that has tennis scores, streaming and information. I searched Serena Williams match record, and her entire professional record was there. I then continued and had to find the number of aces that she had in whatever matches I was going to use for my research. Since there was a variety of statistics about everything from aces, serves, unforced errors, fastest serve etc. Since, the website had a wide array of statistics, so this made finding the aces that Serena scored every match wasn't difficult, but was time consuming. In order to achieve randomness I sorted all the information into just statistics on her Australian Open and US Open information. From there I just picked 30 random matches to use for this study. No adjustments were made to the raw date since the raw data was already in the appropriate format required for this study.

Data Analysis

Summary Statistics**:**

Column	n	Mean	Variance	Std. dev.	Std. err.	Median	Range	Min	Max	Q1	Q3
Australian Open	30	9.0666667	16.34023	4.042305	0.73802055	9	15	3	18	6	11
US Open	30	7.5333333	18.188506	4.2647984	0.7786421	7	17	1	18	4	11

Frequency table results for Australian Open:
Count = 30

Australian Open	Frequency	Relative Frequency
3	2	0.066666667
4	3	0.1
5	1	0.033333333
6	3	0.1
7	2	0.066666667
8	2	0.066666667
9	4	0.13333333
10	5	0.16666667
11	1	0.033333333
13	3	0.1
15	2	0.066666667
17	1	0.033333333
18	1	0.033333333

Frequency table results for US Open:
Count = 30

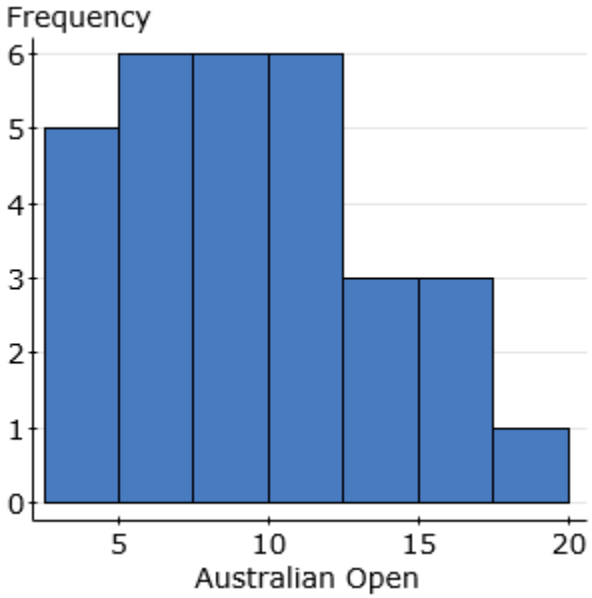
US Open	Frequency	Relative Frequency
1	2	0.066666667
2	1	0.033333333
3	1	0.033333333
4	4	0.13333333
5	2	0.066666667
6	4	0.13333333
7	4	0.13333333
8	2	0.066666667
9	2	0.066666667
11	1	0.033333333
12	3	0.1
13	2	0.066666667
16	1	0.033333333
18	1	0.033333333

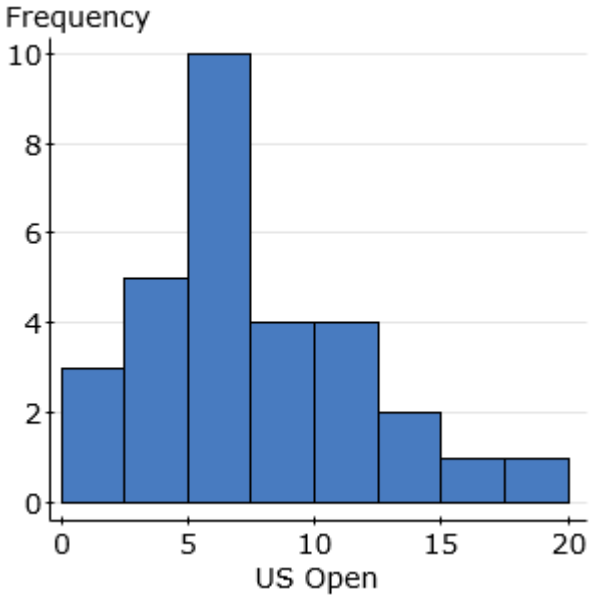
Variable: Australian Open

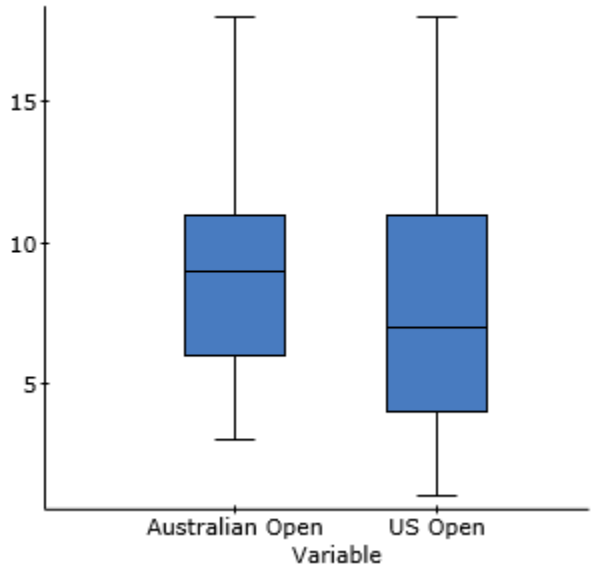
Decimal point is 1 digit(s) to the right of the colon.
Leaf unit = 1
0 : 33444
0 : 566677889999
1 : 000001333
1 : 5578

Variable: US Open

Decimal point is 1 digit(s) to the right of the colon.
Leaf unit = 1
0 : 11234444
0 : 55666677778899
1 : 122233
1 : 68







Percentage of Observations: 1, 2 and 3 Standard Deviations of the Mean

Australian Open $\bar{x} = 9.066$ $s = 4.042305$

-3.06	0.98	5.02	9.06	13.10	17.15	21.19
$\bar{x} - 3s$	$\bar{x} - 2s$	$\bar{x} - s$	$\bar{x}$	$\bar{x} + s$	$\bar{x} + 2s$	$\bar{x} + 3s$

Observations that fall within 1 standard deviation= 26/30 = 86.66%
Observations that fall within 2 standard deviations= 29/30 = 96.66%
Observations that fall within 3 standard deviations= 30/30 = 100%

US Open $\bar{x} = 7.533$ $s = 4.2647984$

-5.26	-0.99	3.26	7.53	11.79	16.05	20.31
$\bar{x} - 3s$	$\bar{x} - 2s$	$\bar{x} - s$	$\bar{x}$	$\bar{x} + s$	$\bar{x} + 2s$	$\bar{x} + 3s$

Observations that fall within 1 standard deviation= 25/30 = 83.33%
Observations that fall within 2 standard deviations= 29/30 = 96.66%
Observations that fall within 3 standard deviations= 30/30 = 100%

Data Analysis

I began this study by hypothesizing that Serena Williams would score a higher amount of aces in the matches that were inside the United States rather than in Australia. While looking at the different histograms I saw that both the Australian Open and the US Open histograms were both right skewed, this is also visible to see when looking at the stem and leaf plots. This simply means that the amount of aces Serena scores decreases as the amount of aces goes up. With all the tests that were done, we found out that the mean for the amount of aces scored during the Australian Open was 9.06, while the mean amount of aces scored during the US Open was 7.53. The summary statistics gave us a lot of information about the random data. The first quartile for the Australian Open was 6, while the first quartile for the US Open was 4. The third quartile for the Australian Open and the US Open was the same, 11. The range for the Australian Open was 15, and 17 during the US Open. The minimum and maximum for the Australian Open was 3 and 18, respectfully. The minimum and maximum for US Open was 1 and 18, respectfully. The relative frequency tables give us some interesting information, based on these tables Serena scored 9 or 10 aces the most frequently during an Australian open match, with a relative frequency =0.133 and 0.166. At the US Open, Serena most frequently scored 4, 6 or 7 aces with a relative frequency = 0.133. All of the data thus far proves discredits my hypothesis

95% Confidence Interval Results:

Variable	Sample Mean	Std. Err.	DF	L. Limit	U. Limit
Australian Open	9.0666667	0.73802055	29	7.5572452	10.576088
US Open	7.5333333	0.7786421	29	5.9408314	9.1258352

We can be 95% confident that the mean number of aces scored by Serena Williams during the Australia Open is between 7.56 and 10.58

We can be 95% confident that the mean number of aces scored by Serena Williams during the US Open is between 5.94 and 9.13.

Two Sample T Hypothesis Test:

μ_1 : Mean of Australian Open

μ_2 : Mean of US Open

$\mu_1 - \mu_2$: Difference between two means

$H_0 : \mu_1 - \mu_2 = 0$

$H_A : \mu_1 - \mu_2 < 0$

(without pooled variances)

Hypothesis test results:

Difference	Sample Diff.	Std. Err.	DF	T-Stat	P-value
$\mu_1 - \mu_2$	1.6873563	1.0811319	56.58566	1.5607312	0.9379

$$x = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{(\frac{s_1^2 + s_2^2}{n_1 + n_2})}}$$

$$t = \frac{9.066 - 7.533}{\sqrt{(\frac{4.04^2 + 4.26^2}{30 + 30})}}$$

I hypothesized that Serena Williams would score a higher amount of aces in the matches that were inside the United States compared to her matches in Australia. Based on a 5% significance level and a p value of 0.9379, since the p value is more than the significance level , I do not reject my null hypothesis $H_0 : \mu_1 - \mu_2 = 0$. There is a 93.79% possibility that the difference between the two means (Australian and US Open) was not due to chance. So at the 5% significance we do reject the null hypothesis, because the data does not provide enough evidence that Serena Williams scored more aces during the US Open in comparison to the Australian Open. The hypothesis yielded the same results as the data analysis did, that my hypothesis was incorrect. All of the information that was found through the data analysis was proven correct in the hypothesis test.

Conclusions

During this study we studied the number of times that Serena Williams scored aces during the Australian compared to during the US Open. When I commenced researching and looking for data, I hypothesized that Serena would score more aces during her US Open matches because she is from the US and could do better for a variety of reasons. She could have “home field” advantage, she feels comfortable on the US courts, or had simply more experience playing in the US. From the beginning my hypothesis was proven wrong. The summary statistics informed us that the mean of both sets of data differed. The mean number of aces scored during the Australian Open for a sample of 30 matches was 9.07 while the mean number of aces scored during the US Open was 7.53. The median for the Australian Open matches was 9 and the median for the US Open was 7. The range for the Australian Open was 15 and the range for the US Open was 17. The relative frequency tables gave us some insight on what the most common amount of aces scored by Serena during the Australian and US Open competitions. During the Australian Open Serena scored 9 or 10 aces the most frequently during an Australian open match, with a relative frequency =0.133 and 0.166. During the US Open, Serena most frequently scored 4, 6 or 7 aces with a relative frequency = 0.133. Both the stem and leaf plots and the histograms both visually show how both sets of data are skewed to the right. Looking at the standard deviation and how many of the data points fell between 1, 2 and 3 standard deviations of the mean gave us more information on where the majority of aces were scored. For the Australian 86.66% of the observations were within 1 standard deviation, the US Open on the other hand only had 83.33% of the observations within 1 standard deviation. During the Australian Open Serena scored between 7.56 and 10.58 aces on average, while she only scored between 5.94 and 9.13 during US matches. The confidence interval part of this study informed us that we could be 955 confident that the mean number of aces scored by Serena Williams during the Australia Open is between 7.56 and 10.58. For the US Open we could be 95% confident that the mean number of aces scored by Serena Williams during the US Open is between 5.94 and 9.13. This, yet again allowed us to see how the mean number of aces scored per match (depending on country) varied. The hypothesis test showed that the 5% significance level we could reject the null hypothesis since there was not enough evidence that Serena Williams scored a higher average of aces during the US Open in comparison to her ace score average during the Australian Open. Overall matches that Serena Williams played in Australia during the Australian Open had a higher number of aces, than those that she played in the United States during the US Open which was proven by a number of tests and statistics that proved my hypothesis wrong.

References

"Serena Williams." *Serena Williams Tennis Stats - H2H Stats · MatchStat,* matchstat.com/tennis/player/Serena%20William.

StatCrunch, www.statcrunch.com/app/index.php.