Gender and Instagram Followers Among High School Students

Andrew Ardle, Star Chen, Alison Pond, Anne-Laure Razat

AP Statistics

Mrs. McMichael

May 2017

Introduction:

The topic of this study is Instagram use among high school seniors. Specifically, it focuses on the relationship between the gender of seniors and the number of followers they have on Instagram. This topic was selected primarily because of the popularity of Instagram among high schoolers; the frequency of its use has led to curiosity of whether or not there is a difference between how people interact with the app based on their gender. The research question guiding this study is: is there a significant difference between the number of Instagram followers that high school senior males and females have? It's important to answer this question because studying how different genders interact with the app could prove to be profitable information for advertisers on Instagram. If they're better able to understand their target audience, connecting with them will be easier. While the sample in this study comes purely from Downingtown West seniors, hopefully this conclusion could be applied to the population of all high school seniors.

Literature Review:

Overall, these three studies reveal that the photos males post on internet, are different from the photos females post. To be more specific, the photos may vary on themes, contents, conceptions, descriptions, and hashtags. There are a lot of factors that lead to the difference of males' and females' photos.

Faces engage us: photos with faces attract more likes and comments on Instagram (Bakhshi)

This study looked at a sample of 1 million randomly selected Instagram images and analyzed the likes and comments as well as content of each. Their results show that

photos with faces are 38% more likely to receive likes and 32% more likely to receive comments, even after controlling for social network reach and activity. They found, however, that the number of faces, their age and gender do not have an effect

How gender-stereotypical are selfies? A content analysis and comparison with magazine adverts (Doring)

This study compared the content of popular Instagram selfies and magazine advertisements. They found that selfies on Instagram reflect traditional gender stereotypes as well as depicting social media specific gender stereotypes, such as poses and angles. They also found that selfies on Instagram are more gender stereotypical than a typical magazine advertisement.

Gender and Instagram Hashtags (Zhang)

This study looked at a randomly selected sample of photos posted in a Malaysian food hashtag and analyzed the gender of the photo's author, and also the other hashtags used in its caption. The results revealed that in general, women tend to use more emotional and positive hashtags while posting photos on Instagram and men use more informative and negative hashtags

The study East Vaco is conducting is more specific, concentrating the followers of males and females (specifically high school seniors) on Instagram. This aspect of studying Instagram was selected because, among all the potential factors of an Instagram account, follower counts

are the most straight-forward and easy to collect quantitative data. This study will add to the topic because of its analysis of followers in relation to gender, while the other studies consider gender of Instagram users, they do not consider amount of followers. For one of the first recorded times, it will help to draw conclusions between gender and follower counts on Instagram.

Methodology:

In order to study the difference in the number of followers between males and females, East Vaco conducted an observational study using a sample of 50 male seniors and 50 female seniors from Downingtown West High School, the sample sizes of 50 ensured that they would be big enough to apply a two mean t-test. In order to obtain this sample, the researchers used a random number generator and a list of seniors' names attending Downingtown West. For both genders, they assigned a number to each senior of that gender. They then randomly picked out seniors based upon the numbers generated. In the end, the sample was made up of 50 male seniors and 50 female seniors.

In terms of collecting the data, the researchers looked up the Instagram profiles of each of the 100 seniors in the study, and recorded the number of followers each senior had. If a particular senior did not have an Instagram profile, they randomly selected another senior in order to achieve a sample size of 50.

Exploratory Data Analysis:

Number of Followers on Male Accounts		
	With Outliers	Without Outliers
Sample Size	50	49
Average (mean)	645.72	508.92
Standard Deviation	495.65	318.90
Min	24	24
Q1	194	189
Median	461.5	453
Q3	822.75	807
Max	7349	1521

Number of Followers on Female Accounts		
	With Outliers	Without Outliers
Sample Size	50	47
Average (mean)	681.06	587.09
Standard Deviation	1043.4	395.14
Min	26	26
Q1	375	360.5
Median	633	617
Q3	811	775
Max	2596	1288
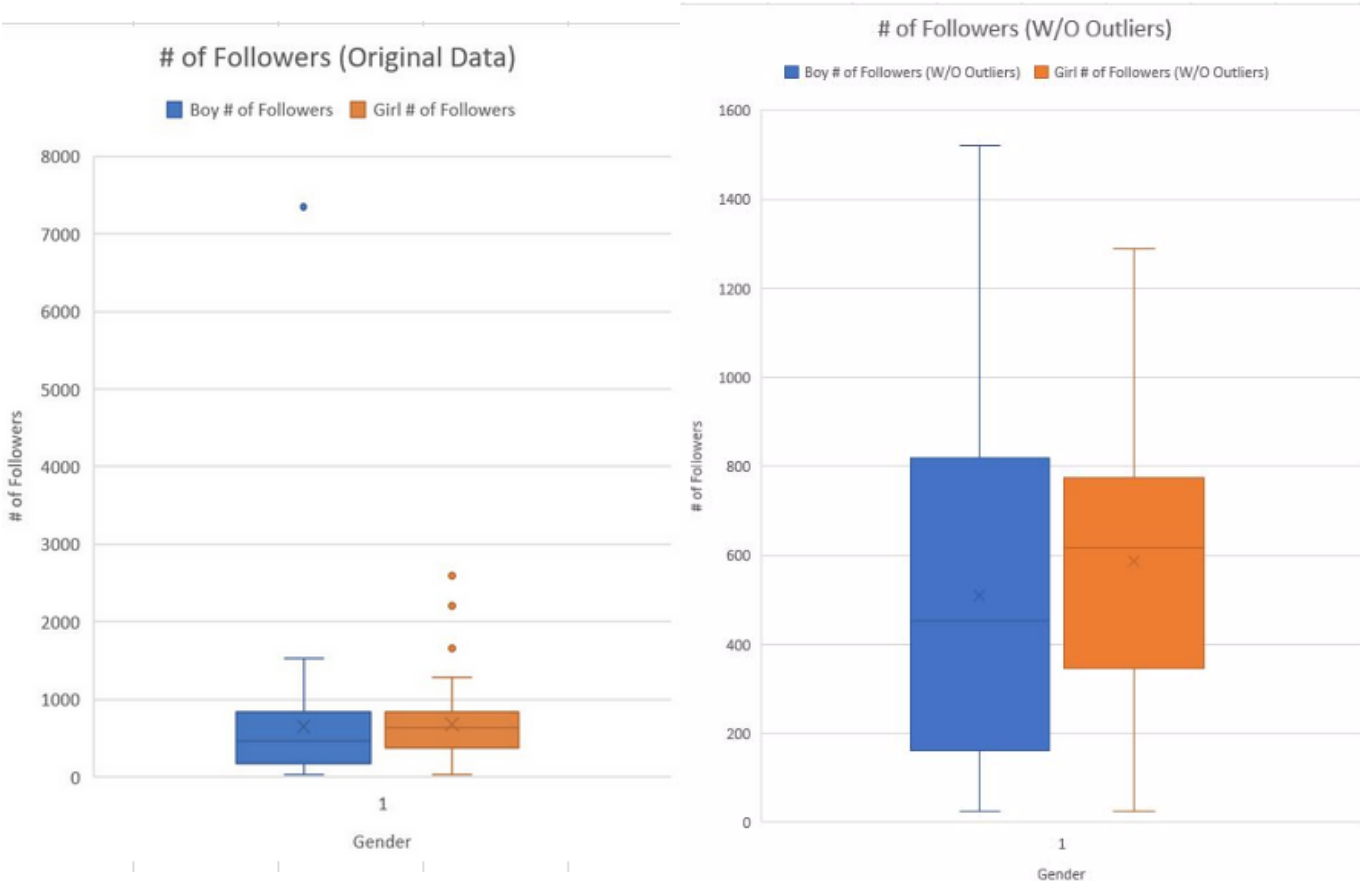
Both data sets, those with and without outliers, are included in this paper, however for their analysis and interpretation, the researchers will only explore the data set without outliers, as they felt it is most representative of all high school senior account follower numbers. They removed outliers of 7,349 followers for the male account data and outliers of 2,596, 1,660, and 2,204 followers from the female account data, as these values all appeared extremely high in comparison to the remainder of the data. Female accounts have a relatively symmetric shape, with seemingly evenly distributed follower counts, while male accounts have a slight right skew, with more values at the lower end of follower counts. The typical amount of followers for a female account is 617 followers as calculated by the median, while the typical amount of followers for a male account is a bit lower, with 453 followers as calculated by the median.

Female accounts have a smaller spread than male accounts, with an IQR of 414.5 followers and a range of 1,262 followers, compared to the larger spread of male accounts with an IQR of 618 followers and a range of 1,497.

Inference:

Conditions

The two-mean t-test passes all required conditions, after removed one high outlier (7349) from the male sample, and three high outliers (2596, 1660, and 2204) from the female sample. Each sample (male and female) contains at least 40 seniors, with the sample of males containing 49 seniors, and the female sample containing 47 seniors. The two samples were selected randomly, so the researchers assumed that the samples of 49 male and 47 female seniors at Downingtown West High School are representative of all seniors in the United States. The number of followers one senior has is independent from the number of followers another senior has. Lastly, the two samples are independent of each other because the number of followers females have does not affect the number of followers males have.

Variables

M = The sample of 49 male seniors that attend Downingtown West High School

F = The sample of 47 female seniors that attend Downingtown West High School

Hypotheses

H_o: There is no difference in the mean amount of followers on Instagram of male and female high school seniors.

$$\mu_F - \mu_M = 0$$

H_A: There is a difference in the mean amount of followers on Instagram of male and female high school seniors.

$$\mu_F - \mu_M \neq 0$$

Test

$$SE = \sqrt{\left(\frac{395.1399^2}{49}\right) + \left(\frac{318.900^2}{47}\right)} = 73.145$$

$$(\mu_F - \mu_M) \sim t_{46}(0,\ 73.145)$$

$$\text{Test Statistic} = \frac{587.085 - 509.918}{73.145} = 1.0687$$

$$\text{P-value} = 2p\,(t_{46} > 1.0687) = 0.291$$

1. The alpha level will be set at .05.

2. Because the conditions passed, the average difference between the two samples is approximately T-distributed with mean of 0, and Standard Error of 73.145.

3. The standard error was found by taking the variance of the male sample (156,136) dividing it by the sample size (49) and then adding that to the variance of the female sample (101,697) divided by the sample size (47). The square root was taken of the entire quantity. The end result was a standard error of 73.145.

4. A T-distribution graph for this sample:
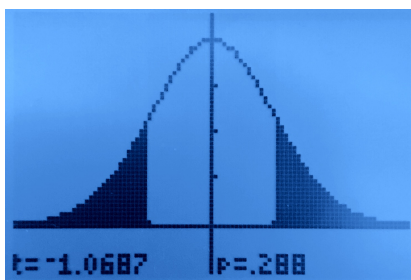
Including Outliers

If the study included the four outliers instead of removing them from the inference, the p-value would be 0.831, which is much larger than the p-value without outliers. It's reasonable because outliers will increase the standard deviation by a large amount. In other words, it is more tolerable to faults. Thus, when calculating the probability to get a sample like this only by chance, it is more possible to get it.

t=-.2163   p=.8294
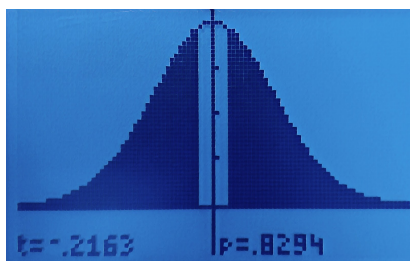
Inference Conclusion

Because the p-value of .291 is greater than the alpha level of 0.05, there is no rejection of the null hypothesis that there is no significant difference between the average number of followers male and female seniors have on Instragram. There is no evidence to suggest there is a significant difference in the number of followers between male and female followers at the 5% significance level. Even when including outliers, a p-value of .8294 is calculated, which is significantly greater than the alpha level of 0.05, and so the same conclusion is drawn and there is no rejection of the null hypothesis. To clearly answer the research question: No, there is not a significant difference in the number of followers between male and female accounts.

Conclusion

To conclude, it cannot be deduced that there is a significant difference in the number of followers between the number of followers male and female high school seniors have on Instagram. There is a greater spread in the number of followers males had in comparison to females. This can likely be explained by the fact that females tend to be more consistent with their Instagram use, as previous studies have shown. Males are more likely to have accounts that have few posts on them and they do not use frequently, correlating with a lower number of followers, and they also are more likely than females to use their accounts for promotional purposes, for example advertising fitness routines or sport talent which can lead to a higher follower account as this type of account may appeal more to a broader audience, rather than simply a collection of mutual friends that appears to compose most of the accounts that we studied. In comparison, the female accounts studied appeared to have a primary purpose of a 'personal' account, where they post pictures with their friends and of activities they do. When viewing accounts, more female accounts appeared to have the 'private' setting, where pictures posted cannot be viewed without explicitly following the account, which can also explain the larger spread in the male accounts as more male accounts had 'public' settings, which typically have higher follower counts in comparison to private accounts.

In this study, there is not a significant difference in the number of followers males and females have on Instagram, however literature sources reveal that the way genders use Instagram differs. This relates more to the types of pictures each gender posts and the hashtags they use. This study suggests relatively similar rates of Instagram activity between genders when analyzing follower counts alone, however analyzing the amount of "likes" each gender gives or

receives may provide a greater understanding for how each gender uses Instagram, relating back to studies previously conducted which indicate certain types of post receive more likes than others, and that women and men tend to post different types of images and use different captions and hashtags in their posts.

It is possible the Eastvaco study failed in finding certain Instagram accounts, marking randomly selected individuals as not having an account when in fact they did. Because there did not appear to be a large number of accounts searched who no Instagram account could be found for and all accounts selected in place were also randomly selected, it is unlikely this potential false exclusion impacted the results.

While the exclusion of outliers seemed necessary in order to conduct the study on a data set that appeared most representative of the population, it is important not to negate these outliers. It is likely any sample of accounts would include a few accounts with follower counts significantly higher than the majority of accounts studies, and so it is possible the outliers expressed in the Eastvaco study are in fact representative of the population of Instagram accounts. The inclusion of outliers does not change the conclusion of the Eastvaco study; there still does not appear to be a significant difference in the number of followers between male and female accounts, and in fact this difference appears to be even smaller when conducting the test with outliers.

To make the Eastvaco study more inclusive, additional age groups should be analyzed rather than only high school seniors. The Downingtown West senior class list provided the most accessible form of data collection, however it does limit the extent to which the Eastvaco conclusion can be extended. Using accounts from all four grades of high school could provide a

stronger analysis and allow the research conclusions to encompass all high schoolers, rather than high school seniors alone.

The use of media as a growing platform for marketing makes studies such as the Eastvaco study increasingly important. To expand upon this study in the future, further gender analysis studies in social media use could be telling of the ways genders differ in their usage of social media and more specifically Instagram. Studies analyzing differences in the number of followers of male and female instagram accounts at the celebrity level could be more telling about gender use and how businesses can appropriately cater to individual Instagram users. Comparing follower makeups in different types of accounts could also be important, for example comparing ratios of males to females following accounts geared towards sports and fitness with that of makeup and fashion accounts. Studying follower counts between genders across different demographics or regions of the world might also find important evidence that could allow for a better understanding of culture and insightful information of how companies can cater towards different genders across the world when using Instagram as a marketing platform.

References:

Bakhshi, S., Gilbert, E., & Shamma, D.A. (2014). Faces engage us: photos with faces attract more likes and comments on Instagram. *CHI*.

Döring, N., Reif, A. & Poeschl, S. (2016). How gender-stereotypical are selfies? A content analysis and comparison with magazine adverts.. *Computers in Human Behavior*, 55, 955-962.

Zhang, Ye (2015) *Gender preferences and Instagram hashtag usage on #Malaysianfood.* Masters thesis, Universiti Teknologi Malaysia, Faculty of Management.

Appendix:

Raw Data:

Note: Individuals marked "X" were individuals chosen by the random number generator, but were not found to have an Instagram account. Individuals highlighted indicate those who were considered outliers in their respective data sets and were removed during data analysis and inference.

MALE ACCOUNTS	# OF FOLLOWERS	FEMALE ACCOUNTS	# OF FOLLOWERS
2	57	12	80
3	663	17	1038
8	1487	20	73
13	348	23	258
14	189	33	704
23	643	35	488
27	254	37	473
30	642	38	952
32	99	40	1192
35	896	43	133
38	7349	46	446
45	838	47	124
47	909	48	1185
48	368	49	2596
51	645	50	968
55	134	52	346
56	51	53	375
62	24	54	530
72	968	60	413

82	828	61	224
93	453	62	706
99	216	63	474
102	248	65	700
109	538	67	775
113	981	73	1070
114	35	75	767
117	107	76	1660
31	X	82	X
120	27	85	519
121	408	86	190
126	965	89	657
41	X	91	X
131	683	92	181
138	502	97	642
42	X	99	X
43	X	103	X
139	877	104	617
146	556	108	450
147	807	109	613
153	133	111	710
156	111	115	2204
160	44	116	26
44	X	118	X
161	368	122	886
68	X	127	X
165	740	130	811
166	239	136	726
173	647	137	284

77	X	140	X
174	36	144	806
94	X	151	X
187	209	154	624
189	354	156	274
191	1388	157	939
193	266	165	433
194	1521	166	1288
199	470	169	757
201	965	173	666
98	X		
103	X		
110	X		
112	X		
118	X		
129	X		
134	X		
137	X		
143	X		
150	X		
152	X		
155	X		
158	X		
163	X		
171	X		
172	X		
184	X		
185	X		
202	X		

203	X		
204	X		