***Statistics Inference Internal – Facebook - by Student X***

**Introduction (Purpose and Question)**

I have been provided with a sample of 94 randomly selected facebook users from the overall population of facebook users worldwide that display all the required information (as of March this year, the overall population was 1.1 billion users). The sample was collected by two year 13 statistics classes at Western Springs College, and the students used the website [www.facebookrandomusers.com](http://www.facebookrandomusers.com) to generate the data. Each student had to get data for 2 random facebook users, but **the user had to have their gender, locale, profile picture type, number of friends, number of photos, and days since last activity on public privacy settings in order to be included in the dataset**. If the randomly generated user had strict privacy settings and the above variables were not available, the user had to be discounted and a new one randomly generated.

 Of the 94 random users collected in the sample, 52 are male, and 42 are female.

My investigative question is: **from this population (all facebook users displaying variables listed above), is there a difference in median number of photos between males and females?**

As mentioned above, the sample of 94 users contained information regarding their locale, profile picture type, number of friends, and days since last activity, but I have chosen to focus on the two variables **gender** and **number of photos**. I have chosen to investigate these two variables specifically, as I am interested to see if one gender has more facebook photos than the other. From observing my own friends activity on facebook, I believe that females appear to upload more photos than males. However, I would like to see if this is the case with all facebook users (**note** that my investigation is limited to facebook users with public settings).

There is also research to suggest that females upload more facebook photos than males, and the reasoning behind this relates to growing image and self-esteem issues particularly prevalent among females. An example of some research is that conducted by University of Buffalo professor Michael A. Stefanone, who wanted to know if Internet behaviors matched traditional norms of men basing their self worth on competition and achievement, and women focusing on appearance.

*“The study found that women who sought approval based on how others' saw them had a much more active social media presence. Specifically, they post a lot more photos of themselves on Facebook. From these findings, Stefanone* [*asserts*](http://www.buffalo.edu/news/12339) *that women perform this behavior to compete for attention with each other”* (<http://www.theatlantic.com/technology/archive/2011/03/study-women-post-more-facebook-photos-to-raise-self-esteem/72150/>).

Stefanone also comments that *“"Perhaps this reflects the distorted value pegged to women's looks throughout the popular culture and in reality programming from 'The Bachelor' to 'Keeping Up with the Kardashians'"”* (<http://news.cnet.com/8301-17852_3-20040699-71.html?part=rss&subj=news&tag=2547-1_3-0-20>).

Stefanone raises a valid question here, and given the mounting media pressure on females to look, weigh, dress, and act a certain way, I think my investigation is particularly poignant. Although my investigation will only focus on the relationship between gender and facebook photos, the information collected from this investigation is relevant to much wider concepts such as gender stereotypes, body-image, and social media as a means of expressing these issues.

**Graphs and Summary Statistics**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Min | 1st Qu. | Median | Mean | 3rd Qu. | Max | Std.dev | Sample Size |
| Female  | 1 | 54.50 | 280.5 | 634.0 | 752 | 4057 | 879.025 | 52 |
| Male | 1 | 12.25 | 73.5 | 145.8 | 186 | 1222 | 226.509 | 42 |





**Analysis**

**Centre -**

The mean number of facebook photos for females is 634 photos, compared with the mean number for males of 145.8 photos (difference of 488.2 photos). The median number of facebook photos for females is 280.5 photos, compared to the median number for males which is 73.5 photos (difference of 207 photos). These mean and median values strongly support the initial hypothesis that females have more facebook photos than males. This claim is also supported by other researchers such as University of Buffalo professor Stefanone (as mentioned in the introduction and referenced below).

(<http://www.theatlantic.com/technology/archive/2011/03/study-women-post-more-facebook-photos-to-raise-self-esteem/72150/>, http://news.cnet.com/8301-17852\_3-20040699-71.html?part=rss&subj=news&tag=2547-1\_3-0-20) .

**Spread –**

The spread of the data is clearly evident in the graphs. The female data is particularly spread out with a minimum value of 1 photo, and a maximum value of 4057 photos. This gives a range of 4056 photos. The male data also has a reasonable spread, but it is not as extreme as the female data. The minimum value for males is 1 photo, and the maximum value is 1222 photos. This gives a range of 1221 photos.

The spread of these two groups is also illustrated by the standard deviation - 226.509 for males, and 879.025 for females.

The inter-quartile range (IQR) is another measure of spread. The IQR for male number of photos is 173.75, and the IQR for female number of photos is 697.5.

All of these measures of spread indicate that the number of photos for female users is more spread out. This means that there is more variation in the number of facebook photos that females have, compared to the number of facebook photos males have. The number of facebook photos for males is more consistent.

**Shape –**

The data for both graphs is skewed left (although the male graph is more significantly skewed), and this indicates that users with lower numbers of facebook photos are more frequently occurring than users with higher numbers of facebook photos (in this sample).

For both groups (but particularly males), the values (number of photos) are clustered in the lower part of the graph (approximately between 1 and 750 photos). The users with higher numbers of photos are more isolated as ‘stand-alone’ values, as shown by the graph tail. This further supports the claim made above that a lot of the users have under 1000 photos on facebook, and those users with over 1000 are more unusual (less common).

There is quite a bit of overlap between the boxplots for each gender group although the size of the graph makes this overlap hard to estimate accurately. The middle 50% for male data almost fits into the lower-middle 25% for female data, although it extends slightly to the left of the females lower quartile. The upper-middle 25% for male data does however fit easily into the lower-middle 25% for female data. The median number of photos for males (73.5) also overlaps with the female data boxplot, fitting into the lower-middle 25%. The median number of photos for female data (280.5) is however outside the inter-quartile range of the male data.

**Unusual Features –**

As mentioned in the ‘spread’ section, the data (number of photos) for females is more scattered than the data for males. There are some unusually high values in the female dataset, which arguably could be viewed as outliers. For example, the maximum value for females is 4057 photos, and this is a significantly larger amount of photos than other users (both male and female). It is possible that someone made a mistake when recording the data, but it is also possible that this one female user just happened to have an excess number of photos. There is no evidence to suggest that this value is invalid, but given these extreme values, I did choose to use the median as my quantity rather than the mean when generating my bootstrapped confidence interval.

Another potentially unusual feature is the peak in the males data at the far left (lower) end of the graph. It seems that there is a few total numbers of photos that are commonly occurring in the male dataset. Research has not uncovered a reason for this, and in a further investigation perhaps this is something that could be investigated. It is possible that this is simply an abnormality in my sample, and if the sample was larger, these results may not occur again (i.e. sampling variability). It would be interesting to know if this peak is also represented in the overall population.

**Inference and Conclusion**

I used iNZight to generate a bootstrap confidence interval. Essentially, iNZight resamples the values from the original dataset (with replacement) to create a new sample, and then this process is repeated 1000 times. From this, I now have several datasets which are able to be representative of the overall population (all facebook users with adequate public settings). A median for each of these samples is also generated, and thus, the confidence interval is created as shown on the graph.

This confidence interval allows me to claim (reasonably confidently) that there is a difference between the number of facebook photos males and females have. The difference in number of photos between males and females in the overall population is very likely to fall between 67 and 430 photos. Given that 0 photos does not fall into this interval, this demonstrates that there is a difference between the two gender groups and the number of facebook photos that they have.

This confidence interval and other data features such as the difference in medians and means between the two gender groups (refer to ‘centre’ for figures), allows me to claim that females generally have more facebook photos than males. (Note: these claims are from the sample and relate to the overall population of facebook users with adequately public settings).

However, it is important to note that this relationship between gender and number of facebook photos may not necessarily be causal. What I mean by this is that being a female or male is not necessarily the only factor influencing the number of photos these users have. For example, the age or ethnicity of the user may have an influence on the number of facebook photos users have. My investigation allows me to claim that there is a difference between the number of facebook photos females and males have, but it does not allow me to claim that gender is what has caused this difference.

Another thing to consider is the reliability or validity of my sample. My results rely on the fact that [www.facebookrandomusers.com](http://www.facebookrandomusers.com) actually has access to all facebook users worldwide and there is an equal chance of each user being selected. My results also rely on the students in Year 13 Statistics recording the data accurately. If either of these two things were not correct, then my sample (and thus my results) is less reliable. It is also important to consider that as with all samples, there is an element of uncertainty, and I can never be 100% confident in my estimates. The bootstrapping method used however, attempts to account for this, and is able to provide a reasonably accurate representation of the overall population.

A limitation of my investigation is that it only allows me to make conclusions regarding facebook users with adequately public profiles. In order to collect the data, the randomly generated user had to have their gender, locale, profile picture type, number of friends, number of photos, and days since last activity available to the public, otherwise there was no variables to record. The issue with this is that a large proportion of facebook users (i.e. those with more secure privacy settings) were excluded, and this could have swayed results. It is possible that people with more private profiles actually have a very different number of photos to those with public settings, and this could mean that my results are not very accurate. However, I am still able to make claims regarding gender and number of photos of facebook users with public settings, I am just not able to generalize these conclusions to include all facebook users.

In conclusion, from my investigation I can say that my sample showed a clear difference in number of facebook photos between males and females, and that this difference is very likely to be representative of the overall population (all facebook users with adequately public profile settings). This conclusion was made using the bootstrapped confidence interval and other data features, and confirms my initial hypothesis. My findings that females tend to have more facebook photos than males supports/is supported by other complex research conducted by universities that give similar results. While in my investigation I have not gone into the reason behind this difference, the other research done by universities do explore possible explanations. It is worth referring back to my introduction to view these links and explore these topics.