|  |  |
| --- | --- |
| **NAME:** | **TEACHER:** |

3

**DATE:**

**Western Springs College 2013**

**AS 91582**

3.10 (v1) Use statistical methods to make a formal inference

*Facebook*

**Credits: 4**

|  |  |  |
| --- | --- | --- |
| **Achievement** | **Achievement with Merit** | **Achievement with Excellence** |
| Use statistical methods to make a formal inference. | Use statistical methods to make a formal inference, with justification. | Use statistical methods to make a formal inference, with statistical insight. |
| Overall level of performance | | |

**Introduction**

The data set Facebook.csv contains a sample of 94 Facebook users. This information was collected by the two Year 13 Statistics classes at Western Springs College. One class was responsible for collecting the male data and the other class was responsible for collecting the female data. Each student obtained information on two random Facebook profiles of their assigned gender by using the site http://www.facebookrandomusers.com/. If a randomly selected user did not have all of the required information publicly available then that user was skipped and a new random profile obtained.

This activity requires you to produce a report describing an investigation that uses statistical methods to make a formal inference related to Facebook users. You will work independently over a period of 2 lessons to pose a comparative investigative question, complete your analysis, make conclusions, and write your report.

The quality of thinking demonstrated in your report and your ability to link the context and population to the different components of the statistical enquiry cycle will determine your overall grade.

**Task:**

You have been provided with a data set of random Facebook users (see Resource A for column heading descriptors).

You have been given an opportunity to research Facebook and its usage to help you understand the variables and develop a purpose for the investigation.

Use the statistical enquiry cycle to conduct your investigation and to write a report describing the investigation.

1. Familiarise yourself with the data set provided. Identify the variables you wish to compare, and write a comparative investigative question.
2. Conduct your investigation and write a report containing:

* your comparative investigative question
* appropriate graphs and summary statistics
* a discussion of the sample distributions
* an appropriate formal statistical inference
* a conclusion communicating your findings, including
  + discussing sampling variability
  + discussing the variability of estimates
  + reflecting on the process that has been used to make the formal inference.

As you write your report, take care to link your discussion to the context and to support your statements with statistical evidence.

Resource A

Facebook is the world’s largest online social networking site. It was founded in February 2004 by Mark Zuckerburg and fellow Harvard students Eduardo Saverin, Andrew McCollum, Dustin Moskovitz and Chris Hughes. By 2008 the site had 100 million users and as of March 2013 there were 1.1 billion Facebook users. Facebook allows users to connect with family and friends around the world by sending messages, chatting, uploading photos and videos, sharing links and joining groups.

Explanation of column headers in the data set:

|  |  |
| --- | --- |
| **Gender** | Male or Female |
| **Locale** | Based on language as chosen by user (not necessarily where they live) |
| **Profile photo type** | Face, full body or other |
| **Number of friends** | Number of Facebook friends the user has |
| **Number of photos** | Number of photos the user has uploaded publicly |
| **Days since last activity** | Days since last activity shown on Timeline (if less than 24 hours this value is 0) |

You should be sourcing relevant contextual knowledge about the situation under investigation from places such as the internet, the school or local library, newspapers and magazines. These sources should be referenced in your report.

Some relevant websites are listed below.

[http://www.statisticbrain.com/Facebook-statistics/](http://www.statisticbrain.com/facebook-statistics/)

<http://www.nzherald.co.nz/business/news/article.cfm?c_id=3&objectid=11127038>

<http://www.bbc.co.uk/news/technology-16832799>

Assessment schedule Mathematics and Statistics 91582 Facebook

|  |  |  |
| --- | --- | --- |
| Evidence/Judgements for Achievement | Evidence/Judgements for Achievement with Merit | Evidence/Judgements for Achievement with Excellence |
| The student has used statistical methods to make a formal inference.  The student has:   * produced a report that shows they have used each component of the statistical enquiry cycle to make a formal inference * posed a comparative investigative question using a given multivariate data set * selected and used appropriate displays and summary statistics * discussed sample distributions * discussed sampling variability, including variability of estimates * made an appropriate formal statistical inference * communicated findings in a conclusion   For example:  **Problem**  **The question is a comparative investigative question that clearly identifies the variables and the population(s).**  e.g. I wonder if females Facebook users have a higher median number of Facebook photos than male users from those with non-private profiles.  **Analysis**  **At least one graph is produced and summary statistics, including the difference between the sample medians, means or quartiles, have been calculated.**  **The sample distributions are discussed and compared in context. This could involve comparing the shift/centre, spread, shape, and unusual features – using features of the displays and the summary statistics.**  e.g. The median number of Facebook photos for males was 73.5 and for females was 280.5 which is 207 photos more than the males. Both sample distributions are right skewed but the females have a longer tail. This is because there are quite a few females who have more than 1000 photos compared with just one male with this many photos. Females have more variation in the number of photos as the IQR is much larger, 697.5 compared to 173.75 for the males.  **Bootstrapping is used to construct a confidence interval. The confidence interval is used to make a formal inference.**  The bootstrap confidence interval does not include 0 so it can be reasonably assumed that the median number of photos for females is higher than for males for all Facebook users who have non-private profiles.  **Conclusion**  **The investigative question is answered in context.**  e.g. We can make the call that female Facebook users have a higher median number of photos than male users from those who have non-private profiles.  **An understanding of sampling variability may be implied in the use of the bootstrapping process.**  **Note: although it may not be included in the comparative question, an understanding that the population consists of Facebook users who have non-private profiles must be evident in the report.** | The student uses statistical methods to make a formal inference, with justification.  The student has:   * produced a report that gives evidence of linking components of the statistical enquiry cycle to the context and/or populations, and referring to evidence such as sample statistics, data values, or features of visual displays in support of statements made * posed a comparative investigative question using a given multivariate data set * selected and used appropriate displays and summary statistics * discussed sample distributions * discussed sampling variability, including variability of estimates * made an appropriate formal statistical inference * communicated findings in a conclusion   For example:  **Problem**  **As for Achievement with an explanation for the choice of variables for the investigation.**  e.g. Based on observing my friends, it seems that my female friends post more photos to Facebook than my male friends. My female friends seem to take photos of every social event and upload them so I think the median number of photos posted by female Facebook users will be higher.  **Analysis**  ***Graphs such as dot plots and box and whisker plots are produced and summary statistics, including the difference between the sample medians, means or quartiles, have been calculated.***  ***The sample distributions are discussed and compared in context. This will involve comparing the shift/centre, spread, shape, and unusual features, with reference to features of the displays and the summary statistics and links to the population or investigative question.***  *e.g. The median number of Facebook photos for females in the sample is 280.5 which is 207 more compared to males (73.5). This is a significant difference and could mean that in the population female Facebook users tend to upload more photos than males.*  *The IQR for the sample of females is 697.5 photos (752-54.5= 697.5) and for the sample of males is 173.75 (186-12.25= 173.75).The IQR for the sample of females is much larger in comparison to the sample of males and this could mean that in the population of Facebook users, females have more variability in the number of photos they upload to Facebook. There are some females who upload a lot of photos (more than 1000) causing their graph to be very right skewed. This could mean that in the population that some females upload a lot of photos as well.*  ***A formal statistical inference is made, using resampling (bootstrapping) to construct a confidence interval.***  *e.g. See screen shot.*  *The confidence interval for the difference of medians shows that the difference between the median number of photos in the population of non-private Facebook users is likely to be between 66 and 429 photos. Since zero does not lie in this interval we can be reasonably sure that females have a higher median number of photos compared to males.*  ***Conclusion***  ***The formal inference is used to answer the investigative question, justifying the call and making links to the context. The conclusion includes an interpretation of the confidence interval.***  ***An understanding of sampling variability is evident.***  e.g. Based on these samples it can be reasonably assumed that the median number of photos for all female Facebook users with non-private profiles is greater than for males. This conclusion is based on observing that the number of photos uploaded by females was shifted up towards the higher numbers compared to the males. Also, the bootstrap confidence interval for the difference in median number of photosd did not include zero. The average difference between the median number of photos in the population is likely to be between about 66 and 429.  Because of sampling variability, a different sample could lead to a different confidence interval but because the lower limit of this interval is clearly above 0 it is unlikely to result in a different conclusion. | The student uses statistical methods to make a formal inference, with statistical insight.  The student has:   * produced a report that gives evidence of integrating statistical and contextual knowledge throughout the statistical enquiry cycle, and may include reflecting about the process and considering other relevant explanations * posed a comparative investigative question using a given multivariate data set * selected and used appropriate displays and summary statistics * discussed sample distributions * discussed sampling variability, including variability of estimates * made an appropriate formal statistical inference * communicated findings in a conclusion   For example:  ***Problem***  **The research is used to develop the purpose for their investigation and the contextual knowledge is used to pose a meaningful comparison investigative question.**  e.g. Based on an infographic on hubspot.com <http://blog.hubspot.com/blog/tabid/6307/bid/6032/Infographic-Men-vs-Women-on-Facebook.aspx> it seems that male and female Facebook users have similar amounts of friends but use Facebook differently. Females tend to provide more information in their profiles about their interests, favourite TV shows, music, etc. University of Buffalo researcher Michael A. Stefanone, PhD, and colleagues found that females who tend to base their self-worth on their appearance tend to post more photos on social media sites <http://www.buffalo.edu/news/releases/2011/03/12339.html>. Since many women seem to be overly concerned with their looks, I wonder of Facebook users with public profiles, if females have a higher median number of photos than males?  **Analysis**  **As for Merit**  **The sample distributions are discussed and compared in context. This includes seeking explanations for features of the data, which have been identified and considering the impact of these on the context or investigative question. Reference to knowledge from the research needs to be included in the discussion.**  *As for Merit with additional discussion e.g.*  *The distribution of the number of photos for both females and males are skewed to the right with most people having less than 1000 photos, however, the males are more clustered towards the lower amounts of photos. There was just one male with more than 1000 photos compared to 9 females with more than 1000. Michael Stefanone’s research into Facebook usage suggests that a cultural difference between genders has resulted in a focus by females on image and appearance. His video* [*http://www.youtube.com/watch?v=T1GQHoLyS5Q&feature=relmfu*](http://www.youtube.com/watch?v=T1GQHoLyS5Q&feature=relmfu) *suggests that women post up to 5 times as many photos as men as a means of enhancing their self-worth. The females in the sample with large amounts of photos could be women with low self-worth using Facebook to attract attention from others as a means of raising their self-worth.*    **A formal statistical inference is made, using resampling (bootstrapping) to construct a confidence interval.**  *e.g. As for Merit*  **Conclusion**  **The formal inference is used to answer the investigative question, justifying the call and linking back to the purpose of the investigation. The conclusion includes an interpretation of the confidence interval and a discussion of sampling variability. Findings are clearly communicated and linked to the context and populations. There is a reflection on the process or other explanations for the findings have been considered which may involve re-examining the data from a different perspective.**  e.g. I have created a bootstrap confidence interval for the difference between the two medians. A bootstrap confidence interval is a range of believable values for the difference in medians of the difference in the number of photos of female and male Facebook users with non-private profiles.  Based on the confidence interval it is reasonably likely that the median number of photos for females is between 66 and 429 more than males. In almost every sample there was a positive difference between the median number of photos posted by females compared to males so it can be reasonably assumed that the median number of photos for all Facebook users with non-private profiles is greater for females than for males. This is consistent with my hypothesis and fits with research done by Michael Stefanone, PhD that suggests that females are more concerned with image appearance and use Facebook as a method of attracting attention by posting lots of photos.  Because of sampling variability, a different sample could lead to a different confidence interval but because the lower limit of this interval is clearly above 0 it is unlikely to result in a different conclusion.  I investigated the difference of medians rather than means because in the female sample there was quite a few females who had posted a lot of photos (more than 1000, highest 4057) which would have pulled the female mean significantly higher so the median was a better average to compare since it is not affected by these extreme values.  When we were collecting our sample of random Facebook users we had to skip any user that did not have all the information on number of friends, photos and date of last activity publicly available. Therefore, it is uncertain that the results I found that females tend to have more photos could be generalised to all Facebook users. Perhaps users with private profiles have different numbers of photos and I may not see the same result of the females having more photos if all Facebook users were able to be included in the investigation. |

Final grades will be decided using professional judgement based on a holistic examination of the evidence provided against the criteria in the Achievement Standard.

Example iNZight dot/box plots and summary stats

and

Example iNZight bootstrap CI

