

# STAT 021 S22 HW 8

Brandon Cramblit

TOTAL POINTS

**10 / 10**

QUESTION 1

**1 Description 3 / 3**

✓ + **3 pts** Correct

QUESTION 2

**2 Identify possible mistakes 3 / 3**

✓ + **3 pts** Complete

+ **0 pts** Incomplete

QUESTION 3

**3 Connection to ASA guidelines 4 / 4**

✓ + **4 pts** Correct

## **I. Describe the statistical analysis and research question**

The study “attempts to test whether online social media allow us to increase the size of our social networks”. The study collects two separate random samples, in which sample 1 consists of 2000 social media users that “make regular use of social media”, and sample 2 consists of 1375 adults who access their social media infrequently. The study performed a chi-square( $X^2$ ) test for homogeneity between the two samples and suggests there was no statistical difference in the distribution of the number of friends on Facebook. The study concludes there is no evidence that frequent use of social media does not lead to an increase in social networks.

### **A. Include the source where you found this analysis referenced as "bad statistics"**

I found this study on this website that summarized 8 research that should be criticized. Second study on the list is the research I will discuss. <https://www.entrepreneur.com/article/275060>

### **B. Also, include the original source whenever possible**

Original study: <https://royalsocietypublishing.org/doi/10.1098/rsos.150292>

## **II. Identify where you think mistakes occurred and why you think they occurred**

### **A. Also reflect on study design and data collection**

First, the mistake could be found in the choose step of the analysis, specifically in the sampling method/data collection. The methods for data collection do not control for potential confounding variables. Based on the research question, we want to compare the network size between the frequent and non-frequent Facebook users. To test this, the study should collect two random samples, in which one consisting of frequent Facebook users, and the other group consisting of non-frequent Facebook users. It is important that the sample is randomly selected, and that the only difference between the two samples should be the frequency of Facebook/social media use. However, to collect sample 2 (non-frequent Facebook users), the study sampled “adults who worked full time”, while sample 1(frequent Facebook users) does not limit to full-time workers. It is clear that there may be a systematic difference between the two groups due to differences in job status. This may have happened because to find non-frequent social media users, one of the easiest ways is to target the sample to those who are busy with full-time jobs. However, this is not an appropriate sampling method to control for all possible confounds. Also, it is questionable that the study measures the response variable(size of networks) through the number of friends on Facebook because other factors contribute to the size of networks, such as the number of contacts, other social media platforms, etc. This may have happened due to a lack of specificity in defining network size; network size could be subjective and defined in various ways. Also, the number of Facebook users was self-reported on a 14 point categorical

1 Description 3 / 3

✓ + 3 pts Correct

## **I. Describe the statistical analysis and research question**

The study “attempts to test whether online social media allow us to increase the size of our social networks”. The study collects two separate random samples, in which sample 1 consists of 2000 social media users that “make regular use of social media”, and sample 2 consists of 1375 adults who access their social media infrequently. The study performed a chi-square( $X^2$ ) test for homogeneity between the two samples and suggests there was no statistical difference in the distribution of the number of friends on Facebook. The study concludes there is no evidence that frequent use of social media does not lead to an increase in social networks.

### **A. Include the source where you found this analysis referenced as "bad statistics"**

I found this study on this website that summarized 8 research that should be criticized. Second study on the list is the research I will discuss. <https://www.entrepreneur.com/article/275060>

### **B. Also, include the original source whenever possible**

Original study: <https://royalsocietypublishing.org/doi/10.1098/rsos.150292>

## **II. Identify where you think mistakes occurred and why you think they occurred**

### **A. Also reflect on study design and data collection**

First, the mistake could be found in the choose step of the analysis, specifically in the sampling method/data collection. The methods for data collection do not control for potential confounding variables. Based on the research question, we want to compare the network size between the frequent and non-frequent Facebook users. To test this, the study should collect two random samples, in which one consisting of frequent Facebook users, and the other group consisting of non-frequent Facebook users. It is important that the sample is randomly selected, and that the only difference between the two samples should be the frequency of Facebook/social media use. However, to collect sample 2 (non-frequent Facebook users), the study sampled “adults who worked full time”, while sample 1(frequent Facebook users) does not limit to full-time workers. It is clear that there may be a systematic difference between the two groups due to differences in job status. This may have happened because to find non-frequent social media users, one of the easiest ways is to target the sample to those who are busy with full-time jobs. However, this is not an appropriate sampling method to control for all possible confounds. Also, it is questionable that the study measures the response variable(size of networks) through the number of friends on Facebook because other factors contribute to the size of networks, such as the number of contacts, other social media platforms, etc. This may have happened due to a lack of specificity in defining network size; network size could be subjective and defined in various ways. Also, the number of Facebook users was self-reported on a 14 point categorical

scale, in which each category is indicated as the following: 0-25, 25-50, 50-75, 75-100, 100-200, 200-300, 400-500, 500-600, 600-700, 700-800, 800-900, and 900-1000+. The scale is not consistent across each category, which could lead to a biased calculation of data, resulting in higher type I or type II errors in hypothesis tests. It is possible that the researcher attempted to manipulate the data by grouping the observations on a categorical scale, in hopes of obtaining a significant p-value.

**B. Think broadly about the steps of modeling: choose, fit, assess, use**

There is also a mistake in the use step of the analysis; the study does not use a correct hypothesis test to answer the research question. The study uses the chi-square test for homogeneity to see if there is a difference in the frequency distribution of the number of Facebook users between the two groups. However, this only tests whether the distribution of Facebook friends in each categorical scale is equal between the two groups, and not whether the frequent Facebook users have more networks than those who do not use Facebook frequently. To test whether the number of Facebook friends is greater for the frequent Facebook user than for non-frequent users, one-tailed unpaired two-sample t-tests will be a more appropriate test. Lastly, this study collected data through surveys instead of controlled experiments, therefore the study cannot answer the research question about causal relationships between social media use and the size of networks.

**III. Connect the analysis to the ASA's guidelines for Ethical Statistical Practice**

**A. Which principles do you think were violated?**

**B. Were these violations intentional? How could you tell?**

I believe this study violates the second principle under Principle A section, which states, "Uses methodology and data that are valid, relevant, and appropriate, without favoritism or prejudice, and in a manner intended to produce valid, interpretable, and reproducible results". As I mentioned previously, the study has an issue with sampling bias, which could lead to invalid data. This may be intentional because to find non-frequent social media users, one of the easiest ways is to target the sample to those who are busy with full-time jobs. Also, measuring network size through the number of friends on Facebook seems to lack validity in measurement. This may be unintentional because network size could be subjective. It could be the case that the researcher believed the number of friends on Facebook would be a good measurement of network size. Chi-square tests for homogeneity are irrelevant to answering the research question because it does not test for whether the network size is greater for one group than the other. I am uncertain whether the inaccurate use of the test was intentional or unintentional, but

## 2 Identify possible mistakes 3 / 3

✓ + 3 pts Complete

+ 0 pts Incomplete



scale, in which each category is indicated as the following: 0-25, 25-50, 50-75, 75-100, 100-200, 200-300, 400-500, 500-600, 600-700, 700-800, 800-900, and 900-1000+. The scale is not consistent across each category, which could lead to a biased calculation of data, resulting in higher type I or type II errors in hypothesis tests. It is possible that the researcher attempted to manipulate the data by grouping the observations on a categorical scale, in hopes of obtaining a significant p-value.

**B. Think broadly about the steps of modeling: choose, fit, assess, use**

There is also a mistake in the use step of the analysis; the study does not use a correct hypothesis test to answer the research question. The study uses the chi-square test for homogeneity to see if there is a difference in the frequency distribution of the number of Facebook users between the two groups. However, this only tests whether the distribution of Facebook friends in each categorical scale is equal between the two groups, and not whether the frequent Facebook users have more networks than those who do not use Facebook frequently. To test whether the number of Facebook friends is greater for the frequent Facebook user than for non-frequent users, one-tailed unpaired two-sample t-tests will be a more appropriate test. Lastly, this study collected data through surveys instead of controlled experiments, therefore the study cannot answer the research question about causal relationships between social media use and the size of networks.

**III. Connect the analysis to the ASA's guidelines for Ethical Statistical Practice**

**A. Which principles do you think were violated?**

**B. Were these violations intentional? How could you tell?**

I believe this study violates the second principle under Principle A section, which states, "Uses methodology and data that are valid, relevant, and appropriate, without favoritism or prejudice, and in a manner intended to produce valid, interpretable, and reproducible results". As I mentioned previously, the study has an issue with sampling bias, which could lead to invalid data. This may be intentional because to find non-frequent social media users, one of the easiest ways is to target the sample to those who are busy with full-time jobs. Also, measuring network size through the number of friends on Facebook seems to lack validity in measurement. This may be unintentional because network size could be subjective. It could be the case that the researcher believed the number of friends on Facebook would be a good measurement of network size. Chi-square tests for homogeneity are irrelevant to answering the research question because it does not test for whether the network size is greater for one group than the other. I am uncertain whether the inaccurate use of the test was intentional or unintentional, but

it is hard to believe that the researcher was unfamiliar with various statistical methods. Lastly, analysis using data collected through surveys can not be used to make an inference about the causal relationship between social media usage and network size, which makes this study methodology not appropriate. Not conducting an experiment could be intentional because it is much more costly and time-consuming relative to survey design.



3 Connection to ASA guidelines 4 / 4

✓ + 4 pts Correct