The Contribution of Social Media in Fuelling Rebellion Among Young Modern Adults in Canada

Joyce Nguyen - 1006873643

September 29, 2022

Contents

I. Part 1

I-1. Goal

In the age of technology, digital media plays a major role in lives of many young adults. The rise of social media networks provides users with the ability to frequently share their curated photos and videos as a way to express themselves. Intentionally or not, constant social interaction among young adults may increase exposure to substance-related content from peers and advertisers and reinforce social norms about using it [1]. In many cases, it even glorifies the use of alcohol and drugs and increase the users' temptation to use these addictive drugs as, potentially, a form of fear of missing out [2].

In a research in the US, by utilizing Poisson regressions, researchers discovered a connection between teen's usage of social media and the tendency of drug use (alcohol, cannabis, vaping) in past month [3]. This requires activities through social media and digital communication for more than three hours per day, rather than the usage of technology itself. Similarly, the same result was also gained from a research in Sweden in which they conclude that there is an association between online socially oriented activities and substance usage, although the effect is not as strong if internet usage was merely for news consumption playing games [4]. Taken together, evidence on the positive association between social media usage and addictive substance consumption is continually emerging in the scientific community.

Noticing this trend, this study aims to investigate the relationship between social media exposure and the amount of drug use in young adults, especially those are locating in Canada. The whole analysis serves to determine if this trend is justified for young adolescents who are living in Canada.

I-2. Procedure

Due to shortage in time, the survey for this study was carried out using a combination of non-probability sampling methods. It was created in Google Forms, with a screening question at the beginning to only let people currently locating in Canada to continue with the list of question. Then the link to the survey was posted to my personal Instagram account for 24 hours using the stories function. Noting that Instagram is one of the most popular social media platforms in the recent years, so it is the best way within my capability to reach out to a wide range of people. I also encouraged my friends to voluntarily participate in the survey. That means, this has a mixture of sample of convenience and sample of volunteers [5]. In the end, the survey got a total of 40 responses, and here is a summary of our types of population: [6]

- The **target population** is the group of all people currently living in Canada.
- The **frame population** consists of all my friends and followers on Instagram.
- The **sample population** includes all 40 people who participated in completing the survey.

A strength of this procedure is that it is both timely and financially efficient [7]. It does not cost much time and money to physically go around asking people to fill out the survey, and it is also convenient for participants in which they just have to click the embedded link to complete the questions. Considering the time constraint and limited scale that can be accessed, probability sampling would be extremely difficult to conduct. Furthermore, conducting a survey on a sample would be much simpler and accessible since we could not gather and idenify all individuals in a target population. [8]

A drawback of this data is that it lacks diversity in demographic information, where only answers from a close community could be recorded. As a result, 75% of the respondents described themselves as coming from East or Southeast Asian countries. Seeing this imbalance in ethnicity distribution, it is possible that cultural perspective might act as a co-founding variable to the result of this study as a result of sampling bias [9]. However, considering my capacity of a college individual, it is difficult to collect data that had diverse representatives on such large scale.

I-3. Showcasing the survey

Here is the link to the survey: https://docs.google.com/forms/d/1_RZbq_5d61DS5e8INgvxaI7EYqT0r4V4E4DzYysDzV4/viewform?edit_requested=true [10]

Note that most of the questions in this survey are set to be mandatory, meaning that participants must finish all of those questions in order to move on to the next section.

In a total of 20 questions listed in the survey, there are three questions that can be helpful for the purpose of this analysis:

- **Question 1**: How many hours do you often spend on social media per day?
- **Question 2**: In general, how often did you consume addictive substances (alcohol, tobacco, nicotine, cannabis, etc.) in the last 6 months?
- **Question 3**: On a scale from 1 to 5, do you agree that the amount of alcohol/drugs related content on your social media changes your frequency of consuming those substances?

II. Part 2

II-1. Data

Data Collection Process

As stated in Part 1, the data was collected by sharing the survey on Instagram story and to my friends. Since the survey was created in Google Forms, all data were recorded in a linked file in Google Sheets so we could easily download it in .csv file type that R can process. A total of 40 observations was gathered with mostly no missing values except for one case-based question and one open question for any suggestions and inquiries at the end. Overall, there were no particular drawbacks or obstacles except for the mentioned imbalance in ethnicity.

Data Cleaning Process

Using the `tidyverse` library, all of the variables are renamed to make it easier to read and conducted in the analysis. Moreover, I removed the column of timestamp that Google Sheets automatically recorded but is not necessary in this study, and also the last column for open suggestions and inquiries which received no responses.

Moreover, there is a scale of frequency written out in words in the survey so respondents could precisely evaluate their personal experience. But this type of data is unexpectedly difficult to incorporate in the analysis, so I changed them into numeric variables as indicated below:

- "Never" = 0
- "Once every few months" = 1
- "Once every few weeks" = 2
- "Weekly" = 3
- "Daily" = 4
- "Multiple times a day" = 5

Data Description

Here are some important variables collected from the survey. Note that the scale of 0-5 is the scale of frequency mentioned in the previous part, and the scale of 1-5 indicates the range of agreement on a statement, with 1 as "Extremely Disagree" and 5 as "Extremely Agree".

- `gender`: The gender that the respondent self-identified with.
- `age`: The age of the respondent.
- `sm_hours`: The number of hours per day the respondent usually spends on social media on the scale of 0-5.
- `see_content`: The respondent's answer on how often they see contents related to addictive substance on social media on the scale of 0-5.
- `temptation_agree`: The respondent's answer on how much they agree that seeing contents related to addictive substance use on social media would increase their temptation of consumption on a scale of 1-5.
- `substance_use`: The respondent's answer on how often they consume addictive substances in general on the scale of 0-5.
- `alcohol`: The respondent's answer on how often they consume alcohol on the scale of 0-5.
- `nic`: The respondent's answer on how often they consume nicotine-based products (e.g. e-cigarettes) on the scale of 0-5.
- `tobacco`: The respondent's answer on how often they consume tobacco-based products (e.g. cigarettes, cigars, etc.) on the scale of 0-5.
- `cannabis`: The respondent's answer on how often they consume cannabis-based products (e.g. weed, edibles, etc.) on the scale of 0-5.
- `purpose`: The respondent's purpose of using addictive substances - recreational, medical reasons or NA if they are non-users.

- `satisfaction`: The respondent's satisfaction when using addictive substances on a scale of 1-5, they can skip if they had no experiences on it.
- `freq_agree`: The respondent's answer on how much they agree that seeing contents related to addictive substance use on social media would increase their frequency of consumption on a scale of 1-5.

Respondents are clearly notified that all of these answers are confidential and strictly be used for the purpose of this assignment.

Data Summary Measures

Here is the table of all the variables, with their respective means and standard deviations. Note that since gender and purpose are categorical variables, they are excluded from this measures table, although we can totally fix it by assigning numeric number to each category).

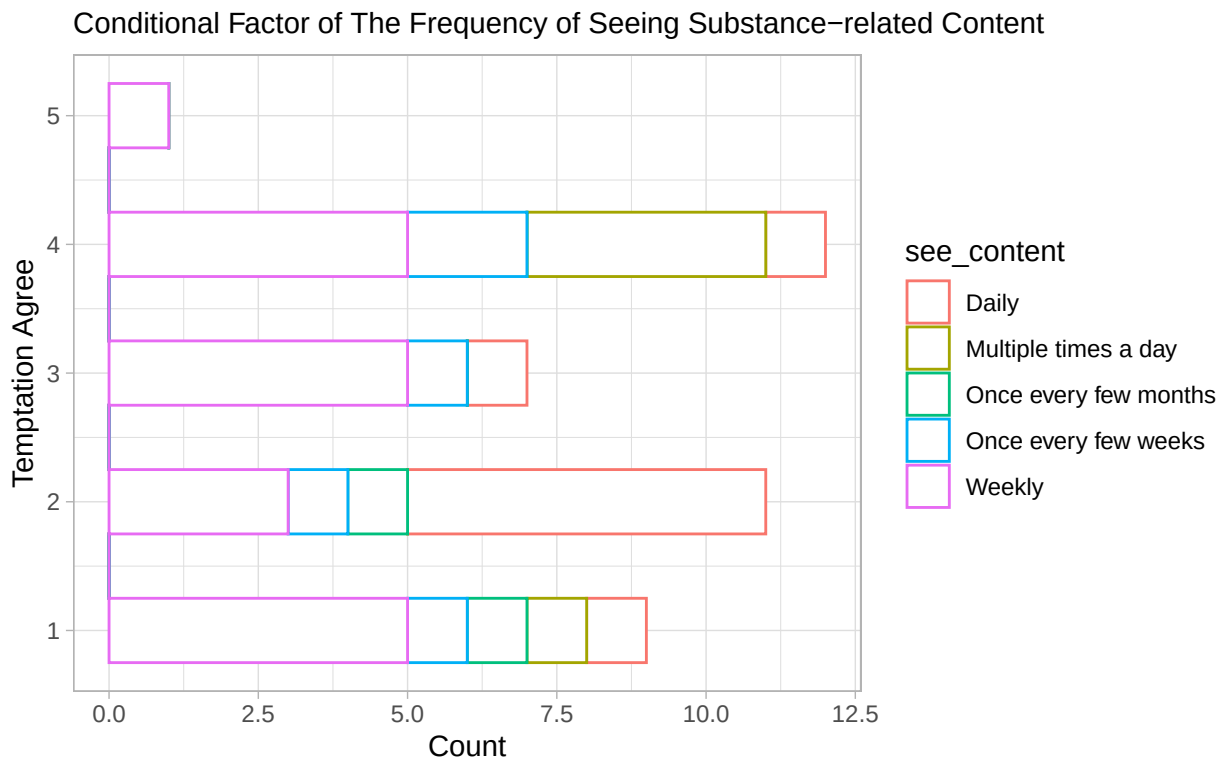
Table 1: Table 1: Means and Standard Deviations of Numeric Variables [11]

Variable	Mean	Standard Deviation
age	20.275	2.309
sm_hours	5.05	2.364
see_content	3.25	1.006
temptation_agree	2.625	1.213
substance_use	1.4	1.533
alcohol	1.375	1.055
nic	0.6	1.336
tobacco	0.275	0.847
cannabis	0.55	1.131
satisfaction	3.32	1.249
freq_agree	2.475	1.24

Here we notice that contents related to addictive substances appeared on respondents' feeds adequately frequently, more than once a week to be more specific, while on average they spent roughly more than 5 hours on social media per day. That means their exposure to these kinds of content is pretty strong. However, it is quite surprising that they don't consume addictive substances that often. The mean for `substance_use` is only 1.4 which means they just use it once every few months on average, but the standard deviation of 1.533 is quite substantial, meaning there is a lot of variations in their frequencies. Finally, in general they thought that the amount of substance related content on social media would not change their temptation or frequency of consuming them.

Now we look at some plots.

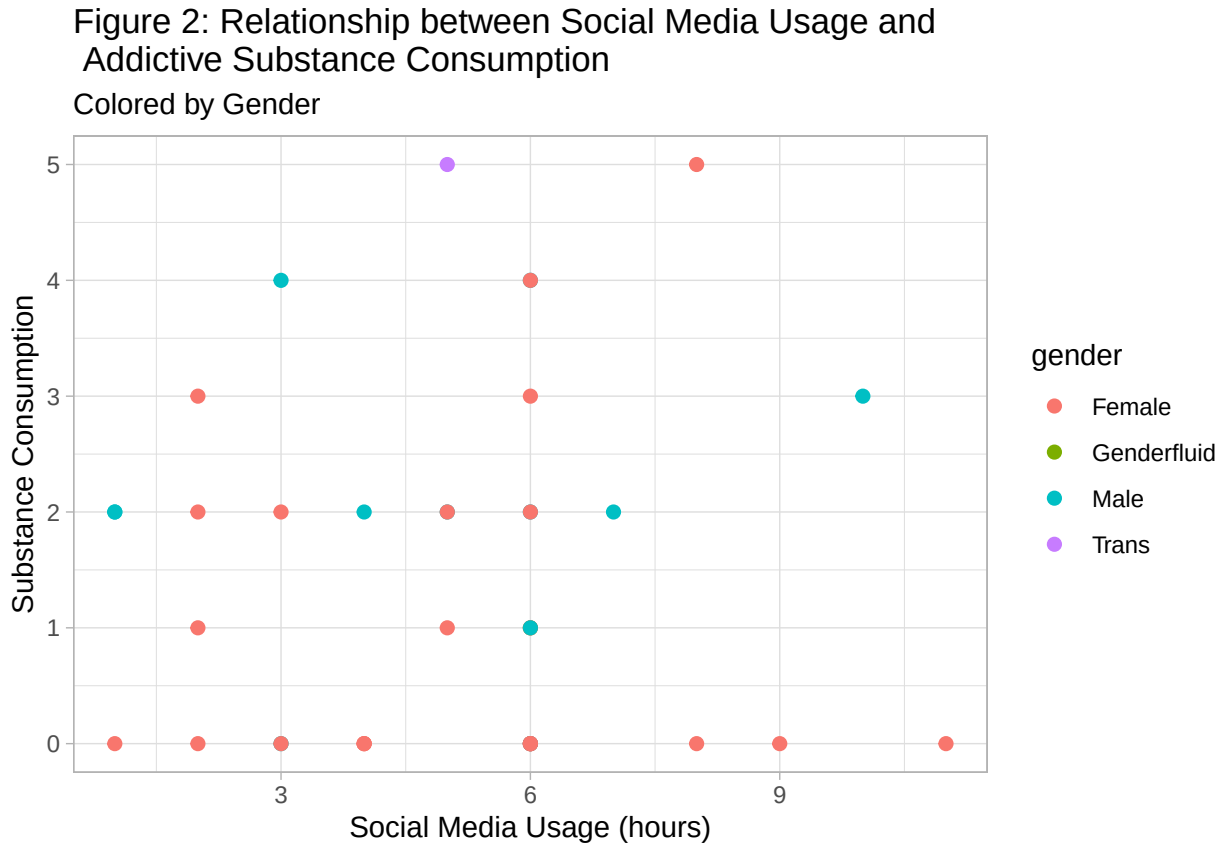
Figure 1: Histogram of Agreement on the Association Between Seeing Substance Content and Temptation



This histogram indicates the distribution of the respondents' agreement on whether they thought seeing contents related to addictive substance usage on social media would change their temptation to consume it, with 1 meaning "Extremely Disagree", and 5 as in "Extremely Agree". It was made with `ggplot()`, and is colored by the frequency of them catching those content. The most popular answer was 4, meaning they do agree that seeing more of those content would increase their temptation to try addictive substances. In addition to that, the frequency they see those content on social media is notably high and quite alerting, where the number of respondents seeing it multiple times a day took the lead and went beyond all the rest.

However, it is interesting that choice number 2 is the second most common answer. That means there were quite some people thought that their temptation were not affected by constant exposure to content related to addictive substances, despite the fact that more than half of them saw those content daily.

Keeping these two numbers (4 and 2) in mind, we have a mean of 3. In the next section, we will examine this variable using hypothesis testing and confidence interval to see if the true population mean differs from 3.



Next, this scatter plot was also created using `ggplot()`, showing the relationship between respondents' number of hours they spent on social media per day anmd their frequency of consuming addictive substances, with different point colors based on gender. At first glance, it is not obvious if there is a linear regression between the two so more rigorous examination is needed in the next section. Moreover, we see that no matter how much time female respondents spent time on social media, a lot of them never consume addictive substances in general. However, this is just an exploratory analysis of data through visualization, more rigorous testings should be done to gain meaningful insights.

Figure 3: Boxplot − Relationship Between Substance Usage Satisfaction and Agreement on Frequency Statement

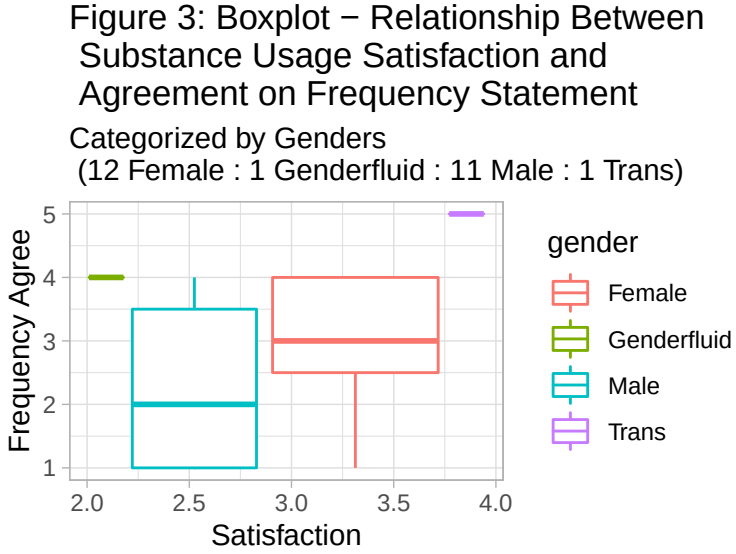


Figure 4: Scatter Plot − Relationship Betweer Substance Usage Satisfaction and Agreement on Frequency Statement

Finally, these two graphs indicate the relationship between how satisfied respondents were with the sensations and experiences of addictive substance usage and whether they agreed with the statement that exposure to drugs related content on social media would increase their frequency of consuming it. Again, these pltos were made using `ggplot()`, and the category is based on respondents' genders. Note that for these visualizations, 15 missing values in satisfaction variable were removed so this plot only includes 25 representatives. Overall, based on the box plot, respondents with more pleasant experiences with drugs use would agree that consuming more contents related to addictive substances would increase their frequency of using it. It is noticeable that female respondents who has tried addictive substances were more satisfied with its effects (females' mean is 3 compared to the mean of 2 for males), and the answers were also more consistent than males with 50% of the responses fell into the range 2.5 to 4 where 50% answers from males were between 1 and 3.5. This graph tells us that aside from, or instead of, the constant exposure to addictive substance related content on social media, their satisfaction with previous drug experiences might be an indirect factor that determines their frequency of alcohol and drugs consumption. In other words, due to the pleasant experience in previous drug consumption, they are motivated to use those addictive substances when being exposed to the mentioned content on social media.

Hence, the scatterplot measuring the same values were introduced to see if there is a relationship between the two variables. At first glance, there is no major association. We will rigorously test this relationship in the next section.

All analysis for this report was programmed using `R version 4.0.2`.

II-2. Methods

- Based on the incentive from **Figure 1**, we will estimate the true mean of the agreement on whether more exposure to addictive substance related content would increase temptation to consume them. Since we do not have full population information available, and we only have one sample from the survey, one-sample t-test would be the most eligible in this case [12]. Specifically, we want to determine whether the unknown population mean is different from the mean of our two most common answers (4 and 2), which is 3, using hypothesis testing and confidence interval. From that, the null hypothesis H_0 is that the true population mean is equal to 3, and the alternative hypothesis H_1 is that the true mean is not equal to 3. These hypotheses can be expressed as:

$$H_0 : \mu = 3$$

$$H_1 : \mu \neq 3$$

  Also, if the confidence interval contains 3 with a 95% confidence level then there is more evidence to believe that the true mean is 3. [13]

- Now looking at **Figure 2**, we would like to investigate whether there exists a linear relationship between social media usage (predictor) and addictive substance consumption (response). Due to the nature of linear regression model, expressed as $y = \beta_0 + \beta_1 x$, we will see if the slope β_1 is different than 0 [14]. So the null hypothesis H_0 is that the slope is equal to 0 meaning there is no relationship between the two variables, and the alternative hypothesis H_1 is that the slope is not equal to 0 and there is an association between them. Here we have the expression for our hypotheses:

$$H_0 : \beta_1 = 0$$

$$H_1 : \beta_1 \neq 0$$

  Continuing with the confidence interval calculation, if our 95% confidence interval includes 0 then there is no linear relationship between social media usage and addictive substance consumption. [15]

- **Figure 3** and **Figure 4** represents the same values on different visualizations so we only need to do one test for both. For this one, we are interested in whether there is a linear relationship between the satisfaction from previous addictive substance usage experiences and whether they agree that constant exposure to these type of content would change their frequency in consuming it. The procedure will be the exact same as the one in Figure 2, but we set the predictor as satisfaction level, and the response is agreement on frequency being affected.

II-3.Results

Results of the tests are indicated in the Table 2 below. The table contains of the estimation of slope, t-value, p-value, and the 95% confidence interval for each mentioned figure. Note that since we used a different method for Figure 1 compared to for Figure 2 and Figure 3&4, no result were available for slope estimation in Figure 1.

Table 2: Summary of Results [16]

Figure	Slope Estimate	t-value	p-value	Confidence Interval
Figure 1	N/A	-2	0.06	(2.24, 3.01)
Figure 2	0.006	0.05	0.959	(-0.207, 0.218)
Figure 3&4	0.308	1.68	0.107	(-0.0722, 0.687)

- By running the one-sample t-test for **Figure 1**, we gained some estimated insights about the true mean of agreement on whether more exposure to substance related content would increase temptation of consuming. We found that the t-statistics was -2 with a degree of freedom of 39. The result p-value is

0.06, which is slightly higher than 0.05 significance level [17]. This implies that we have insufficient evidence to reject the null hypothesis where true mean is 3. After that, when finding the confidence interval, we got a result of (2.24, 3.01) which contains the value 3, and we can be 95% confident that the true mean falls between that range. Hence, it is plausible to say that the mean is equal to 3, or at least slightly lower than 3 since our p-value is vaguely higher than the significance level, implying that people generally have a neutral opinion to slightly disagreement about whether social media triggers their temptation to consume addictive substances. [18]

- Next, we tested whether there exists a linear relationship between social media usage and substance consumption (**Figure 2**), and the result shows that the estimated slope is 0.006, the t-value is 0.05, with a p-value of 0.959. Since the estimated slope is so close to 0, and the p-value is way higher than the significance level of 0.05, we fail to reject the null hypothesis of the slope β_1 being equal to 0 [19]. That implies there is not linear relationship between the effect of social media usage on substance consumption, i.e. an increase in frequency of social media usage would not change the amount of addictive substance to be consumed. Furthermore, we are 95% confident that the slope would lie in the interval (-0.207, 0.218), which obviously consists of 0. This further strengthen the statement that social media usage does not have any influence on the frequency of consuming addictive substances. [20]
- Lastly, shift our attention to the relationship between how satisfied respondents were with the sensations and experiences of addictive substance usage and whether they agreed with the statement that exposure to drugs related content on social media would increase their frequency of consuming it (**Figure 3&4**). After running the hypothesis testing for linear regression model, we received an estimated slope of 0.308, a t-value of 1.68, and a p-value of 0.107. Again, this p-value is too high compared to the 0.05 significance level, so we fail to reject the null hypothesis of having the slope equal to 0 [21]. The 95% confidence interval of (-0.0722, 0.687) also adds more strength into the result that there exists no linear regression between the two mentioned variables. [22]

In conclusion, we saw that there were not enough evidence to conclude that social media usage has an effect on the amount of addictive substance usage among young adults in Canada. However, a reminder that the data in this sample has a very high chance of being biased, so further investigation on this topic is highly encouraged.

Bibliography

1. Cabrera-Nguyen, E. P., Cavazos-Rehg, P., Krauss, M., Bierut, L. J., & Moreno, M. A. (2016, March). *Young adults' exposure to alcohol- and marijuana-related content on Twitter.* Journal of studies on alcohol and drugs. Retrieved October 2, 2022, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4803667/

2. E.W. Hoffman, B.E. Pinkleton, E. Weintraub Austin, W. Reyes-Velázquez. *Exploring college students use of general and alcohol-related social media and their associations with alcohol-related behaviors.* J. Am. Coll. Heal., 62 (2014), pp. 328-335, 10.1080/07448481.2014.902837

3. Kaur, N., Rutherford, C. G., Martins, S. S., & Keyes, K. M. (2020, June 18). Associations between digital technology and substance use among U.S. adolescents: Results from the 2018 Monitoring the Future Survey. Drug and Alcohol Dependence. Retrieved October 2, 2022, from https://www.sciencedirect.com/science/article/pii/S0376871620302891#bib0090

4. Svensson, R., & Johnson, B. (2020, June 24). Internet use and adolescent drinking: Does it matter what young people do online? Drug and Alcohol Dependence. Retrieved October 2, 2022, from https://www.sciencedirect.com/science/article/pii/S0376871620303033?via%3Dihub

5. Survey sampling methods & techniques: Lucid. Cint. (n.d.). Retrieved October 5, 2022, from https://www.cint.com/blog/different-methods-of-survey-sampling

6. Wu, Changbao and Mary E. Thompson, 2020, *Sampling Theory and Practice*, Springer.

7. Wu, Changbao and Mary E. Thompson, 2020, *Sampling Theory and Practice*, Springer.

8. Wu, Changbao and Mary E. Thompson, 2020, *Sampling Theory and Practice*, Springer.

9. Wu, Changbao and Mary E. Thompson, 2020, *Sampling Theory and Practice*, Springer.

10. Nguyen, J. (2022, September 27). STA304 A1 Survey: The Contribution of Social Media in Fuelling Rebellion Among Young Modern Adults in Canada. Google. Retrieved October 6, 2022, from https://docs.google.com/forms/d/1_RZbq_5d61DS5e8INgvxaI7EYqT0r4V4E4DzYysDzV4/viewform?edit_requested=true

11. Allaire, J.J., et. el. *References: Introduction to R Markdown.* RStudio. https://rmarkdown.rstudio.com/docs/. (Last Accessed: May 5, 2021)

12. Dekking, F. M., et al. (2005), *A Modern Introduction to Probability and Statistics: Understanding why and how.* Springer Science & Business Media.

13. Dekking, F. M., et al. (2005), *A Modern Introduction to Probability and Statistics: Understanding why and how.* Springer Science & Business Media.

14. Dekking, F. M., et al. (2005), *A Modern Introduction to Probability and Statistics: Understanding why and how.* Springer Science & Business Media.

15. Dekking, F. M., et al. (2005), *A Modern Introduction to Probability and Statistics: Understanding why and how.* Springer Science & Business Media.

16. Allaire, J.J., et. el. *References: Introduction to R Markdown.* RStudio. https://rmarkdown.rstudio.com/docs/. (Last Accessed: May 5, 2021)

17. Dekking, F. M., et al. (2005), *A Modern Introduction to Probability and Statistics: Understanding why and how.* Springer Science & Business Media.

18. Dekking, F. M., et al. (2005), *A Modern Introduction to Probability and Statistics: Understanding why and how.* Springer Science & Business Media.

19. Dekking, F. M., et al. (2005), *A Modern Introduction to Probability and Statistics: Understanding why and how.* Springer Science & Business Media.

20. Dekking, F. M., et al. (2005), *A Modern Introduction to Probability and Statistics: Understanding why and how.* Springer Science & Business Media.

21. Dekking, F. M., et al. (2005), *A Modern Introduction to Probability and Statistics: Understanding why and how.* Springer Science & Business Media.

22. Dekking, F. M., et al. (2005), *A Modern Introduction to Probability and Statistics: Understanding why and how.* Springer Science & Business Media.

Appendix

Here is a glimpse of the data set simulated/surveyed:

```
## Rows: 40
## Columns: 19
## $ in_canada        <chr> "Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "Yes~
## $ gender           <chr> "Male", "Female", "Female", "Female", "Male", "Female~
## $ age              <int> 31, 19, 18, 19, 20, 18, 20, 20, 21, 18, 20, 20, 20, 2~
## $ ethnicity        <chr> "East and Southeast Asian (Chinese, Korean, Japanese,~
## $ student          <chr> "No", "Yes", "Yes", "Yes", "Yes", "Yes", "Yes", "Yes"~
## $ accounts         <chr> "4 - 6 accounts", "4 - 6 accounts", "4 - 6  accounts"~
## $ sm_hours         <int> 3, 5, 6, 8, 6, 9, 2, 6, 4, 6, 1, 7, 10, 6, 6, 4, 6, 5~
## $ see_content      <chr> "Weekly", "Weekly", "Weekly", "Weekly", "Weekly", "On~
## $ post             <chr> "Weekly", "Once every few months", "Once every few mo~
## $ post_content     <chr> "Once every few months", "Never", "Never", "Never", "~
## $ temptation_agree <int> 1, 3, 1, 2, 3, 3, 2, 1, 1, 3, 1, 4, 4, 1, 4, 2, 4, 4,~
## $ substance_use    <chr> "Never", "Once every few months", "Never", "Never", "~
## $ alcohol          <chr> "Never", "Once every few months", "Once every few mon~
## $ nic              <chr> "Never", "Never", "Never", "Never", "Never", "Never",~
## $ tobacco          <chr> "Never", "Never", "Never", "Never", "Never", "Never",~
## $ cannabis         <chr> "Never", "Never", "Never", "Never", "Daily", "Never",~
## $ purpose          <chr> NA, NA, NA, NA, "Recreational purposes", NA, "Recreat~
## $ satisfaction     <int> 2, NA, NA, NA, 3, NA, 4, NA, NA, 2, 4, 4, 3, NA, 4, 1~
## $ freq_agree       <int> 1, 2, 1, 2, 4, 3, 4, 1, 1, 4, 1, 3, 3, 2, 3, 1, 3, 1,~
```