STA304 - Summer 2021
Assignment 1

David Pham - 1005349053

Part 1

Goal

The goal of this survey is to explore the effects of video games on peoples' mental health, behaviour, as well as performance in school or at work. Video games are a very controversial subject, particularly in the media. For the most part, they have been seen as detrimental to society and are scapegoated for causing negative behaviour in people, such as laziness, decreasing our intelligence, and the age-old favorite: violent tendencies [1]. The whole analysis serves to determine if this cynical press is justified. Additionally, I regularly play video games and would like to see what my friends think on the matter.

Procedure

This survey was implemented using a mixture of non-probability sampling methods due to time-efficiency. In order to conduct a more accurate analysis, I have decided not to simulate my data, but reach out to my friends in the gaming community. This was made easier using Discord, an application that lets you talk and text with people around the world. I put the link to my survey as my status on Discord, and I also encouraged my friends to voluntarily complete my survey. Then, I asked them if they could share the survey with their friends. Altogether, this is a combination of voluntary-based sampling, convenience sampling, and snowball sampling. [2]

Despite the survey focusing on students in college/university, the **target population** is the group of all people who use Discord.
The **frame population** consists of all my friends, as well as friends of friends (due to the snowballing effect). Finally, the **sample population** will have the friends (and friends of friends) that actually completed the survey.

A strength of this sampling procedure is notably the time-efficiency. It would be extremely difficult to use probability sampling in this context, considering the amount of time we had to do this assignment. Furthermore, it is much simpler to analyze the data generated by the sample because specific individuals within a target population cannot be explictly listed or identified. [3]

Finally, a drawback of this sampling procedure comes from the inherent bias of non-probability sampling. My sample will not be representative of the target population (i.e, everybody who uses the Discord platform and plays video games), since the majority of the responses will come from within my local community. [4] This results in sampling bias.

Showcasing the survey.

The link to the survey is here:
https://docs.google.com/forms/d/e/1FAIpQLSfj3YrZKQhx03mZ1ZBAIa_FwhM4z3Q95jewpWYfaI_o6unKHQ/viewform?usp=sf_link [5]

Here are three notable questions that will be considered throughout this assignment:

Question 1: On average, how many hours of video games do you play a week?
A fairly simple and safe question that prompts the respondent to enter a numerical, discrete value. As straightforward as it is, this question will enable us to examine many relationships between other variables, such as hours played and GPA, happiness, and whether or not the amount of hours played affect one's behaviour. The only possible con I can think of is that it may be considered as a loaded question if a respondent does not play video games at all and decides to take this survey.

Question 2: What is your GPA?
This is a question that asks for a numerical, continuous value that will be used in linear regression later on. From the context of this assignment, it is appropriate to ask and will help us generate a conclusion. It was explicitly stated in the survey that all responses will be confidential and cannot be traced back to a name. However, this is a more sensitive question and can make respondents uncomfortable in revealing their actual GPA. It is completely possible that people may have lied when answering, and can lead to inaccurate conclusions.

Question 3: Rate the following statement: "Sometimes, video games negatively affect my mood and can make me violent at times."
Finally, this question personally asks respondents if they think video games can affect their behaviour negatively (can be especially true in a competitive environment). This question relates back once again to our original goal, determining if video games do indeed change the way we act. A potential drawback of this question is that there are negative connotations implied in the question. The question may have inherent bias and can put off respondents, making them more inclined to answer the question differently. [6]

Part 2

Data

As stated in Part 1, the data was collected by sharing the survey around on Discord through my friends, and then having them ask their friends to complete it as well. A total of 29 observations was gathered using Google Forms, and a bit of data wrangling had to be done in order to organize the information. Overall, there were no particular drawbacks or obstacles that occured aside from the fact that not many people completed the survey (I don't have many friends).

Using the `tidyverse` library, I renamed most of the variables to make the data frame more readable and so that it will be easier to incorporate in analysis. Furthermore, I decided to remove the questions where the responses required selecting multiple options because wrangling that information was more difficult than expected. [7]

In summary, these are the most important variables collected from the survey:

- `hours_played`: The average number of hours spent playing video games every week from the respondent.
- `gender`: The gender of the respondent.
- `happiness`: The respondent's answer on how happy playing video games made them from a scale of 1-10.
- `beneficial`: The respondent's answer on how much they thought video games were beneficial to them, on a scale of 1-10.
- `negative_or_violent`: The respondent's answer on how much they thought video games negatively impacted their mood or caused them to be more aggressive, on a scale of 1-10.
- `affect_relationship`: The respondent's answer on how much they thought video games affected their relationships with loved ones, on a scale of 1-10.
- `cognition`: The respondent's answer on how much they thought video games made them smarter or improved their cognitive abilities, on a scale of 1-10.
- `gpa`: The respondent's GPA. [8]

Again, all of these answers are confidential and will be strictly used for this assignment.

Here is a table of all the variables, along with their respective means and standard deviations of the sample:

Table 1: Means and standard deviations of all variables in the survey data. [9]

Variable Name	Mean	Standard Deviation
hours_played	17.034	6.265
gender	NA	NA
happiness	8.241	1.244
beneficial	7.345	0.814
negative_or_violent	3.103	1.676
affect_relationship	2.621	2.043
cognition	8.69	1.168
gpa	3.285	0.455

Note that there is no mean and standard deviation for the respondents' gender, as it is a categorical variable (although we could have fixed this by adding a factor to the gender variable). Furthermore, a notable observation is that the mean for whether or not respondents thought video games made them smarter is 8.69, indicating that gamers do think video games improve cognitive abilities. Finally, for negative/violent

behaviour and whether or not video games affected relationships, the mean for these values were 3.103 and 2.621 respectively (with slightly higher spreads compared to the other variables), showing that people do not think playing games affect their behavior or social circle that much.

Next, we will take a look at some plots.

Figure 1: Histogram for Negative/Violent Behaviour
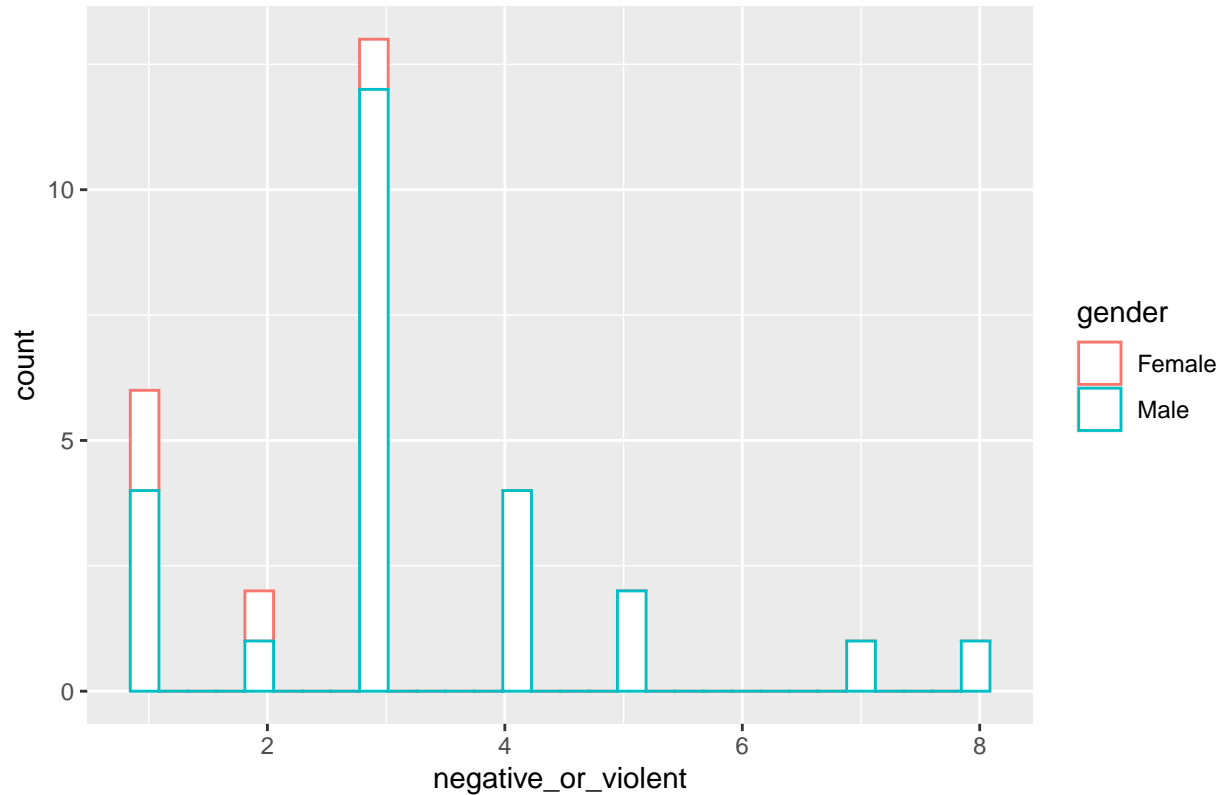
This histogram shows all the respondents' answers to the question asking whether or not they thought video games affected their mood or caused them to act more violently. It was created using ggplot(), and is colored by gender. The most popular answer was a 3, showing that most respondents believe video games do not really impact their mood or cause them to act more aggressively. We will use hypothesis testing and confidence intervals later to see if the true mean differs from 3, proving that this result extends further than the sample.

Figure 2: Hours Spent Playing Video Games and Negative Behaviour
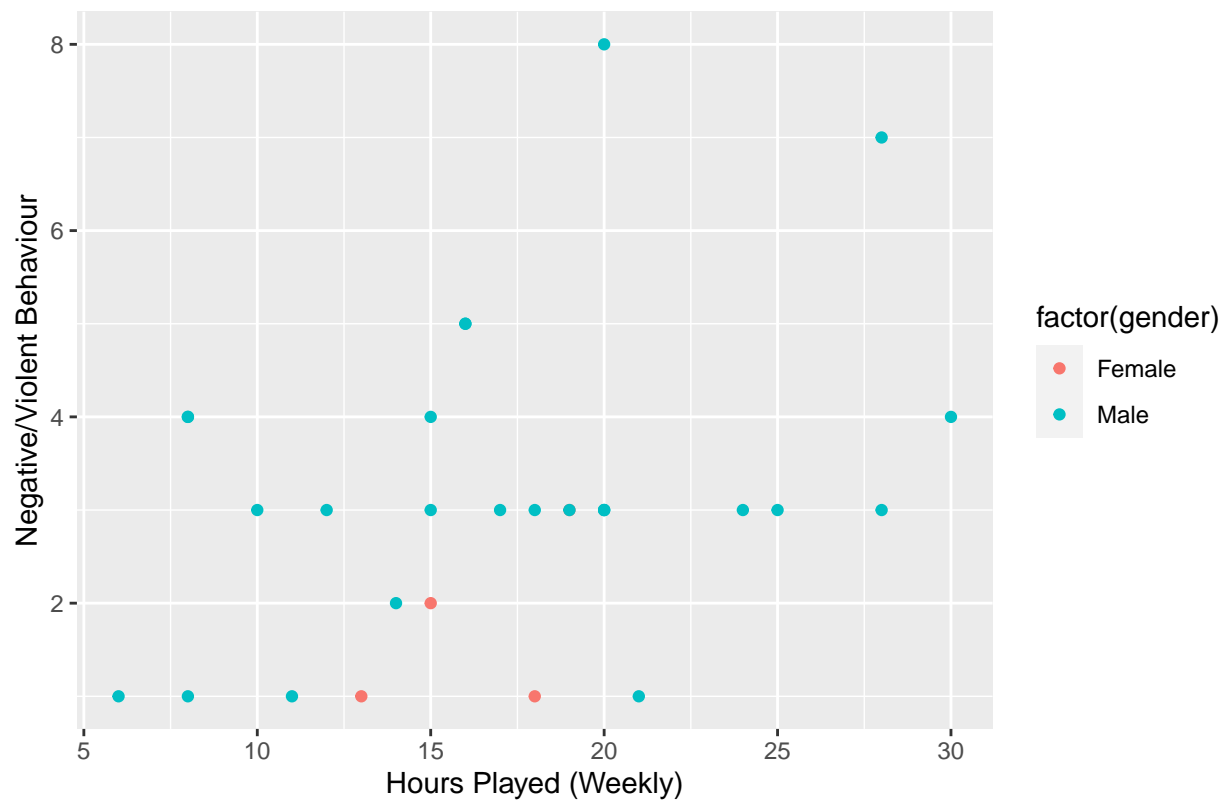
Next, this scatter plot (also made with ggplot()) graphs the respondents' number of hours spent playing video games per week, as well as their answers to the negative behaviour question. Upon first glance, it is not obvious if there is a linear relationship between the two, so we will further examine these two variables in the next section. For the majority of the answers in the dependent variable, they are below 5.

Figure 3: Scatter plot of Hours Spent Playing Video Games and GPA
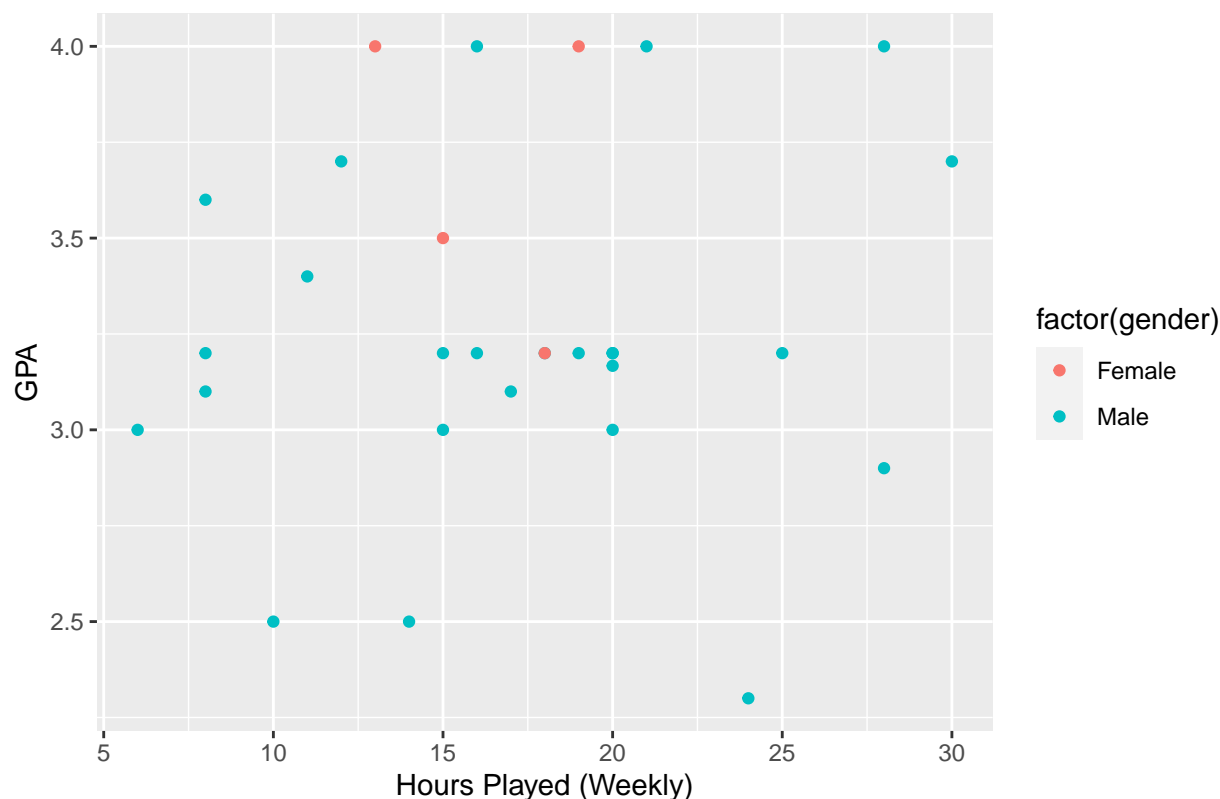
Finally, this scatter plot graphs the respondents' number of hours spent playing video games per week and their GPA. The plot looks completely random and scattered, which indicates that there is no linear relationship between the amount of time spent playing games and your grade point average. This implies that games do not impact school performance. However, this is only a visual inspection, and more rigorous testing needs to be done.

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

Methods

Figure 1
To estimate the true mean of video games causing negative behaviour based on our sample, we will be employing a simple one-sample t-test, along with the associated confidence interval for the true mean. We will use the t.test() function and look at all of its information [10]. The null hypothesis is that the true mean is equal to 3, and the alternative hypothesis is that the true mean is *not* equal to 3. Additionally, if the confidence interval contains 3, then we have reason to believe that the population mean can be equal to 3 as well [11].

Figure 2
Next, we will be using more advanced strategies for hypothesis testing and confidence intervals due to the nature of linear regression. Specifically, we need to test if there is a relationship between playing more video games and more mood changes/violent tendencies. We can look at the slope, the β_1 value to see if it is different from 0 [12]. Hence, our null and alternative hypotheses are:

$$H_0 : \beta_1 = 0$$

$$H_A : \beta_1 \neq 0$$

6

If we fail to reject the null hypothesis, then the slope parameter is 0, implying that there is no evidence of a linear relationship between the two variables. Contrarily, if we reject the null hypothesis, then there is evidence that there is a linear relationship. We will also use confidence intervals for β_1 to see if it contains 0. If it does, then there is no evidence of a linear relationship between the predictor and dependent variables in the population. [13]

Figure 3

Lastly, for figure 3, the procedure will be the exact same as the procedure in Figure 2, but we replace the response for GPA instead of negative/violent behaviour.

Results

Below is a table containing the respective slope estimates, t-values, p-values, and confidence intervals for the three figures/relationships we primarily seek to examine. The interpretation of the results for Figure 1 are different than Figure 2 and 3 because of the data analysis methods used.

Table 2: Results of figures from Data section. [14]

Figure Number	Slope Estimate	t-value	p-value	Confidence Interval
Figure 1	N/A	0.3	0.7	(2.47, 3.74)
Figure 2	0.0936	1.94	0.063	(-0.00535, 0.193)
Figure 3	0.00757	0.54	0.59	(-0.021, 0.0361)

Strictly assessing negative/violent behaviour among those who play video games (**Figure 1**), we saw that the sample mean was 3.1. Upon using a one-sample t-test, we found that the t-statistic was 0.3 with 28 degrees of freedom. The p-value associated with this statistic is 0.7, which is much higher than the 0.05 significance level [15]. Hence, this implies that we fail to reject the null hypothesis where the true mean is equal to 3. Furthermore, if we accompany the hypothesis test with a confidence interval, we see that we can be 95% confident that the true mean lays between 2.47 and 3.74 [16]. It is very plausible that the true mean is equal to 3, implying that people who play video games generally disagree with the fact that games negatively affect their mood and cause them to act violent.

Next, when interpreting the results from the relationship between hours spent playing video games and negative/violent behaviour (**Figure 2**), we have a slope estimate of 0.0936, a t-value of 1.94 and a p-value of 0.063. This is not quite lower than the 0.05 significance level, so we fail to reject our null hypothesis where $\beta_1 = 0$. In words, we do not have evidence to suggest there is a linear relationship between the two variables. As a side note, if the result was statistically significant, then for every additional hour spent on video games, the response to the behavioural question with the 1-10 scale increases by 0.1, on average [16]. We affirm our conclusion by observing the confidence interval for β_1, (-0.00535, 0.193), where the interval contains 0. This implies that there is no evidence of a linear relationship between hours spent playing video games and negative behaviour [17].

Finally, we shift our attention to the relationship between hours spent playing video games and GPA (**Figure 3**). With a slope estimate of 0.00757, t-value of 0.54 and p-value of 0.59, we have a p-value much greater than the 0.05 benchmark significance level, implying that we fail to reject our null hypothesis where $\beta_1 = 0$ [18]. There is no evidence to suggest a linear relationship between the response and predictor. Similarly to the previous conclusion, the confidence interval for this β_1 contains 0, so we can conclude that there is no evidence of a linear relationship between hours spent playing video games and GPA [19].

In conclusion, we saw that we did not have enough evidence to conclude that the number of hours spent playing video games have an effect on your GPA or mood/aggression. Furthermore, by running a one-sample t-test on the results for negative behaviour or violent tendencies, it is plausible that the true mean for this result lays around 3. The reason why people may not completely disagree (i.e, give a score of 1) is because in competitive settings, playing video games can be very stressful and frustrating, which may lead to anger

and behavioural changes. However, claiming that video games are responsable for terroristic acts is simply absurd [20]. Video games should not be villainized and should be seen as a fun hobby that people have to destress, meet new people and improve their cognitive abilities. Games have saved my life numerous times, and I stand firm in believing that a moderate amount of time spent playing them greatly benefit you.

Bibliography

1. Timm, Jane. (2019, August 5) *Fact check: Trump suggests video games to blame for mass shootings.* NBC News. https://www.nbcnews.com/politics/donald-trump/fact-check-trump-suggests-video-games-blame-mass-shootings-n1039411 (Last Accessed: May 14, 2021)

2. Wu, Changbao and Mary E. Thompson, (2020), *Sampling Theory and Practice*, Springer.

3. Wu, Changbao and Mary E. Thompson, (2020), *Sampling Theory and Practice*, Springer.

4. Wu, Changbao and Mary E. Thompson, (2020), *Sampling Theory and Practice*, Springer.

5. Pham, David. (2021, May 14) *STA304 A1 Survey: David Pham.*

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

7. Wickham et al., (2019). *Welcome to the tidyverse.* Journal of Open Source Software, 4(43), 1686, https://doi.org/10.21105/joss.01686

8. Grolemund, G. (2014, July 16) *Introduction to R Markdown.* RStudio. https://rmarkdown.rstudio.com/articles_intro.html. (Last Accessed: May 14, 2021)

9. Allaire, J.J., et. el. *References: Introduction to R Markdown.* RStudio. https://rmarkdown.rstudio.com/docs/. (Last Accessed: May 14, 2021)

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

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

12. Sheather, J. Simon, (2009), *A Modern Approach to Regression with R*, Springer.

13. Sheather, J. Simon, (2009), *A Modern Approach to Regression with R*, Springer.

14. Allaire, J.J., et. el. *References: Introduction to R Markdown.* RStudio. https://rmarkdown.rstudio.com/docs/. (Last Accessed: May 14, 2021)

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

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

17. Sheather, J. Simon, (2009), *A Modern Approach to Regression with R*, Springer.

18. Sheather, J. Simon, (2009), *A Modern Approach to Regression with R*, Springer.

19. Sheather, J. Simon, (2009), *A Modern Approach to Regression with R*, Springer.

20. Timm, Jane. (2019, August 5) *Fact check: Trump suggests video games to blame for mass shootings.* NBC News. https://www.nbcnews.com/politics/donald-trump/fact-check-trump-suggests-video-games-blame-mass-shootings-n1039411 (Last Accessed: May 14, 2021)