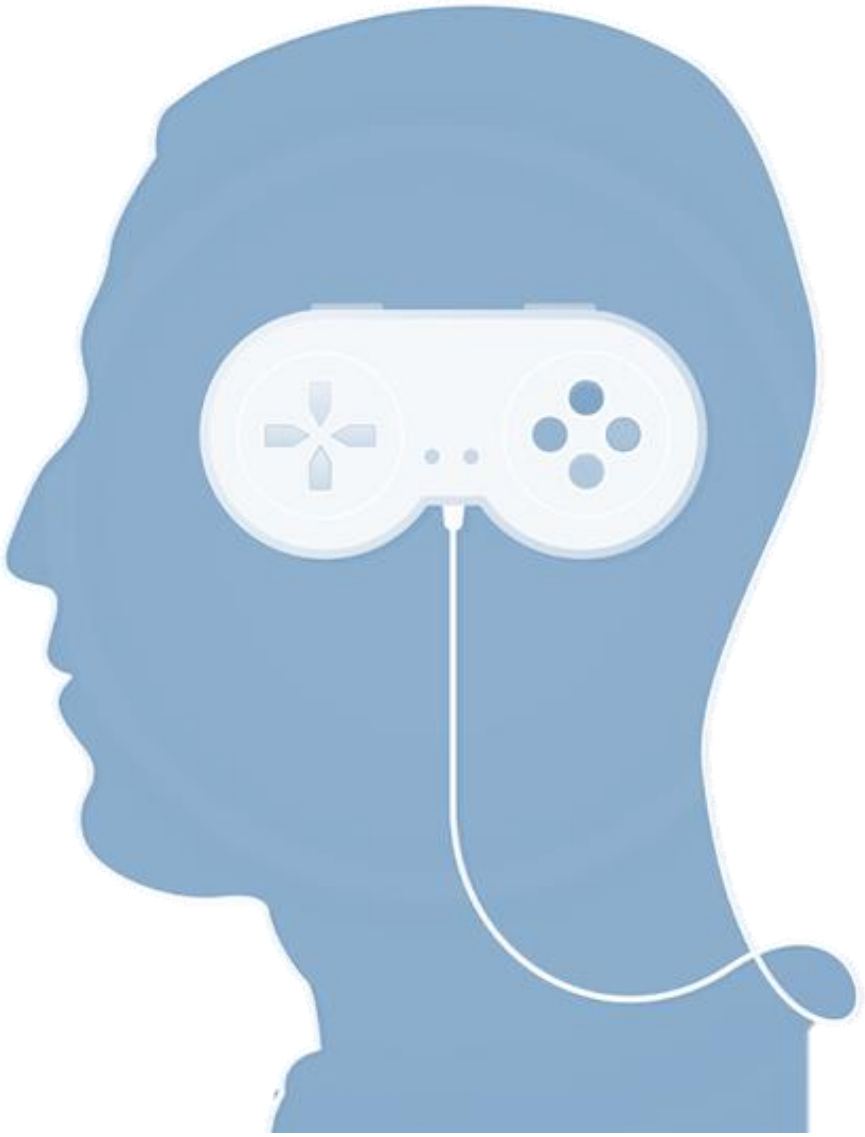


Statistical Interpretation of
ANXIETY AND ONLINE GAMING

Stephen Murphy & Joy Hartsig
Probability & Applied Statistics

Contents

Statistical Interpretation of Anxiety and Online Gaming

Introduction

This data analysis report is based on an "Attitude, Personality & Gaming" survey, the requirements for participants included that participant must be at least 18 years old and must play video games regularly. The survey consisted of multiple questionnaires to gather personal data about the participant such as age, gender, country of birth, country of current residence, etc., as well as questions pertaining to mental health and video game usage. The data set contains complete survey responses from 13,464 participants. The layout of this survey includes open-ended, multiple choice, and interval scale questions.

This data was available in raw form, consisting of a spreadsheet of the participant's answers to each question. In this report, the data will be analyzed, and various problem-solving techniques will be applied to conclude various statistical interpretations of the data.

Chapter 1

1.1 Mean:

From the 13,464 participants that completed the survey, what is the mean age of the participants?

$$\bar{y} = \frac{1}{n}\sum_{i=}^{n} y_i$$

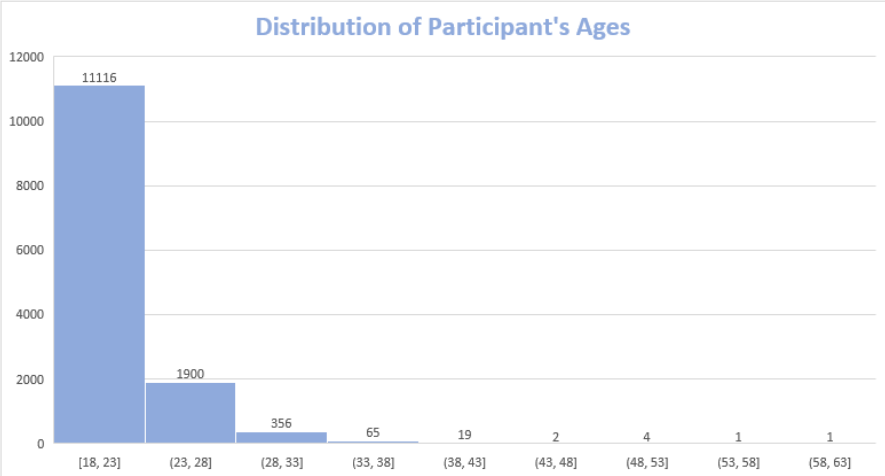
$$mean \ = \ sum \ / \ count$$

$$mean \ = \ 281{,}807 \ / \ 13{,}464$$

$$mean \ = \ 20.9$$

The average age of the participants who completed the survey is about 21 years old.

Distribution of Participant's Ages:



This distribution does not have the "normal" bell shaped curve as the survey results from participants under the age of eighteen were not published.

1.2 Variance:

What is the variance of participant ages?

$$s^2 = \frac{1}{n-1}\sum_{i=}^{n}(y_i - \bar{y})^2$$

$$variance = \frac{sum\ of\ (value - mean)^2}{count - 1}$$

$$variance = \frac{sum\ of\ (value - 20.9)^2}{13{,}464 - 1}$$

$$variance = 10.9$$

The variance of participant ages is 10.9.

1.3 Standard Deviation:

What is the standard deviation of participant ages?

$$s = \sqrt{s^2}$$

$$standard\ deviation = \sqrt{variance}$$

$$standard\ deviation = \sqrt{10.9}$$

$$standard\ deviation = 3.3$$

The standard deviation of participant ages is 3.3.

Chapter 2

2.7 Permutation:

Questionnaire A is an interval scale question that contains seven different questions with five possible responses from "Not at all" to "Nearly every day". How many different ways can all seven questions be answered in which the order of the answers matters?

$$P_r^n = n(n-1)(n-2)\dots(n-r+1) = \frac{n!}{(n-r)!}$$

$$P_5^7 = \frac{7!}{(7-5)!}$$

$$P_5^7 = 2{,}520$$

There are 2,520 ways to answer the seven questions when the order of the answers matters.

2.8 Combination:

Questionnaire A is an interval scale question that contains seven different questions with five possible responses from "Not at all" to "Nearly every day". How many ways can all seven questions be answered?

$$\binom{n}{r} = C_r^n = \frac{P_r^n}{r!} = \frac{n!}{r!\,(n-r)!}$$

$$\binom{7}{5} = \frac{7!}{5!\,(7-5)!}$$

$$\binom{7}{5} = 21$$

There are 21 ways to answer the seven questions when the order of the answers does not matter.

2.9 Conditional Probability:

What is the probability that a participant strongly disagrees with being satisfied with life given that the participant spends 30 or more hours gaming a week?

**Impossible hours per week have been removed from the set (ex: 8000 hours playing video games out of a possible 168 hours a week). **

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$$P(A) = P(strongly\ disagrees\ with\ being\ satisfied\ with\ life) = .07858$$

$$P(B) = P(spends\ 30\ or\ more\ hours\ a\ week\ gaming) = .241162$$

$$P(A|B) = \frac{.07858 * .241162}{.241162}$$

$$P(A|B) = .07858$$

The probability that a participant strongly disagrees with being satisfied with life given that they spend 30 or more hours gaming a week is 7.86%.

2.10 Independent Events:

Prove that strongly disagreeing with being satisfied with life and playing video games for 30 or more hours a week are two independent events.

$P(A) = P(strongly\ disagrees\ with\ being\ satisfied\ with\ life) = .07858$

$P(B) = P(spends\ 30\ or\ more\ hours\ a\ week\ gaming) = .241162$

Only one needs to be satisfied for the events to be independent.

1. P(A|B) = P(A)

$$P(A|B) = \frac{.07858 * .241162}{.241162}$$

$P(A|B) = .07858 = P(A)$ ✓

2. P(B|A) = P(B)

$$P(B|A) = \frac{.241162 * .07858}{.07858}$$

$P(B|A) = .241162 = P(B)$ ✓

3. P(A∩B) = P(A)P(B)

$P(A \cap B) = .01895 = P(A)P(B)$ ✓

The two events are independent.

Chapter 3

3.3 Probability Distribution

Through participant information it was found that 5.3% of participants were female, 94.31% were male, and .39% identified as other. If five participants are randomly chosen from the entire participant list, find the probability distribution for Y, the probability that a female was chosen.

$$p(0\ females\ were\ chosen) = P(Y=0) = \frac{\binom{713}{0}\binom{12751}{5}}{\binom{13464}{5}} = 76.18\%$$

$$p(1\ female\ was\ chosen) = P(Y=1) = \frac{\binom{713}{1}\binom{12751}{4}}{\binom{13464}{5}} = 21.31\%$$

$$p(2\ females\ were\ chosen) = P(Y=2) = \frac{\binom{713}{2}\binom{12751}{3}}{\binom{13464}{5}} = 2.38\%$$

$$p(3\ females\ were\ chosen) = P(Y=3) = \frac{\binom{713}{3}\binom{12751}{2}}{\binom{13464}{5}} = .13\%$$

$$p(4\ females\ were\ chosen) = P(Y=4) = \frac{\binom{713}{4}\binom{12751}{1}}{\binom{13464}{5}} = .003\%$$

$$p(5\ females\ were\ chosen) = P(Y=5) = \frac{\binom{713}{5}\binom{12751}{0}}{\binom{13464}{5}} = .000041\%$$

3.4 Expected Value

What is the expected value of Y, a participant chosen is a female?

$$E(Y) = \sum_{y} yp(y)$$

Number of females chosen (Y)	Probability P(Y)
0	.7618
1	.2131
2	.0238
3	.0013
4	.00003
5	.00000041

$$E(Y) = (0 * .7618) + (1 * .2131) + (2 * .0238) + (3 * .0013) + (4 * .00003) + (5 * .00000041)$$

$$E(Y) = 0.26472205$$

3.5 Variance & Standard Deviation

Find the variance for the random variable Y, a participant chosen is female.

$$V(Y) = E[(Y - \mu)^2]$$

Number of females chosen (Y)	Probability P(Y)	$(Y-\mu)^2$
0	.7618	0.247086
1	.2131	0.002665
2	.0238	0.058043
3	.0013	0.069391
4	.00003	0.070062
5	.00000041	0.070078

$$V(Y) = 0.247086 + 0.002665 + 0.058043 + 0.069391 + 0.070062 + 0.070078$$

$$V(Y) = 0.517325$$

Find the standard deviation for the random variable Y, a participant chosen is female.

$$Standard\ Deviation = \sqrt{V(Y)}$$

$$Standard\ Deviation = 0.71925$$

3.6 Binomial Distribution

A question from the survey asked participants what gaming platform they use to play games. What is the probability that from five trials, exactly five participants chose PC as their main platform?

$$p(y) = \binom{n}{y} p^y q^{n-y}$$

$$p(y = 5) = \binom{5}{5} (0.981803)^5 (1 - 0.981803)^{5-5}$$

$$p(y = 5) = 0.91226659845949$$

The probability that out of five trials exactly five participants chose PC as their main gaming platform is 91.23%.

3.7 Expected Value & Variance

Find the expected value regarding the previous question.

$$\mu = E(Y) = np$$

$$\mu = (5)(0.981803)$$

$$\mu = 4.909015$$

Find the variance regarding the previous question.

$$\sigma^2 = V(Y) = npq$$

$$\sigma^2 = (5)(0.981803)(0.018197)$$

$$\sigma^2 = 0.089329345955$$

3.8 Geometric Probability Distribution

20.306% of the survey participants are employed. If participants are chosen sequentially and are selected at random from the pool of survey participants. Find the probability that the first participant that is employed is found on the fifth interview

$$p(y) = q^{y-1}p$$

$$p(y) = (0.79694)^{5-1}(0.20306)$$

$$p(y) = 0.08190$$

3.8 Expected Value & Variance

Find the expected value regarding the previous question.

$$\mu = E(Y) = \frac{1}{p}$$

$$\mu = \frac{1}{0.20306}$$

$$\mu = 4.9247$$

Find the variance regarding the previous question.

$$\sigma^2 = V(Y) = \frac{1-p}{p^2}$$

$$\sigma^2 = V(Y) = \frac{1-(0.20306)}{(0.20306)^2}$$

$$\sigma^2 = 19.3276$$

3.10 Hypergeometric Probability Distribution

A question from the survey asks participants the county in which they were born. From the total 13,464 participant answers 4,261 participants (or 31.65%) were born in the USA. If 5,000 random participants are chosen and their answers are randomly sampled. What is the probability that 3,000 participants were born in the USA?

N = population size = 13,464 participants

n = sample size = 4,261 participants

r = number of selected items from population size = 5,000 participants

y = random variable = 3,00 participants

$$p(y) = \frac{\binom{r}{y}\binom{N-r}{n-y}}{\binom{N}{n}}$$

$$p(y) = \frac{\binom{5000}{3000}\binom{13464-5000}{4261-3000}}{\binom{13464}{4261}}$$

$$p(y) = \frac{\binom{5000}{3000}\binom{8464}{1261}}{\binom{13464}{4261}}$$

$$p(y) = 1.67817 * 10^{-642}$$

The probability of exactly 3,000 participants being born in the USA out of 5,000 randomly sampled responses is $1.67817 * 10^{-642}$.

3.10 Expected Value & Variance

Find the expected value from the previous question.

$$\mu = E(Y) = \frac{nr}{N} = \frac{(4261)(5000)}{(13464)}$$

$$\mu = 1{,}582.37$$

Find the variance from the previous question.

$$\sigma^2 = V(Y) = n\left(\frac{r}{N}\right)\left(\frac{N-r}{N}\right)\left(\frac{N-n}{N-1}\right)$$

$$\sigma^2 = (4261)\left(\frac{(5000)}{(13464)}\right)\left(\frac{(13464)-(5000)}{(13464)}\right)\left(\frac{(13464)-(4261)}{(13464)-1}\right)$$

$$\sigma^2 = 679.98$$

3.11 Poisson Probability Distribution

Find the Poisson probability for the age of participants being 30 considering the mean, λ, is equal to 21.

$$p(y) = \frac{\lambda^y}{y!} e^{-\lambda}$$

$$p(y) = \frac{21^{30}}{30!} e^{-21}$$

$$p(y) = 0.01326$$

3.11 Expected Value & Variance

Find the expected value of the previous question.

$$\mu = E(Y) = \lambda$$

$$\mu = 21$$

Find the variance of the previous question.

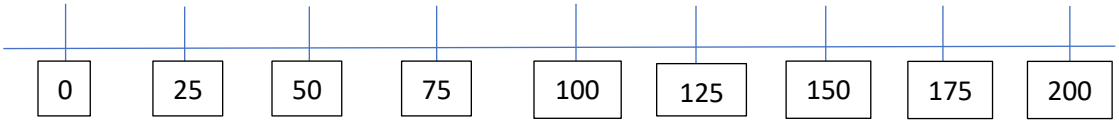
$$\sigma^2 = V(Y) = \lambda$$

$$\sigma^2 = 21$$

Chapter 4

4.4 Uniform Probability Distribution

While reading over the survey a response to the question "How many hours a week do you watch player/tournament streams, coach …etc." an answer, NA, came up within the first 200 responses. Find the probabilities below using this information.



The answer was given within the first 100 responses sent in.

$$\left[0-100:\frac{1}{8}*4=\frac{4}{8}=\frac{1}{2}=50\%\right]$$

Using a number line to break the 200 participants into 8 sections we can see that the answer NA had a 50% chance to be given within the first 100 responses.

The answer was given after the 125th response.

$$\left[125-200:\frac{1}{8}*3=\frac{3}{8}=\ 37.5\%\right]$$

Using the same concept, we break the group up into 8 equal sections and since they want the probability the answer was given after the 125th response leaves us with a probability of 37.5% Meaning there is a 37.5% chance the NA was given after the 125th response.

Given that the NA answer wasn't given within the first 50 responses find the probability that the response was given between the 50th and 75th submitted response

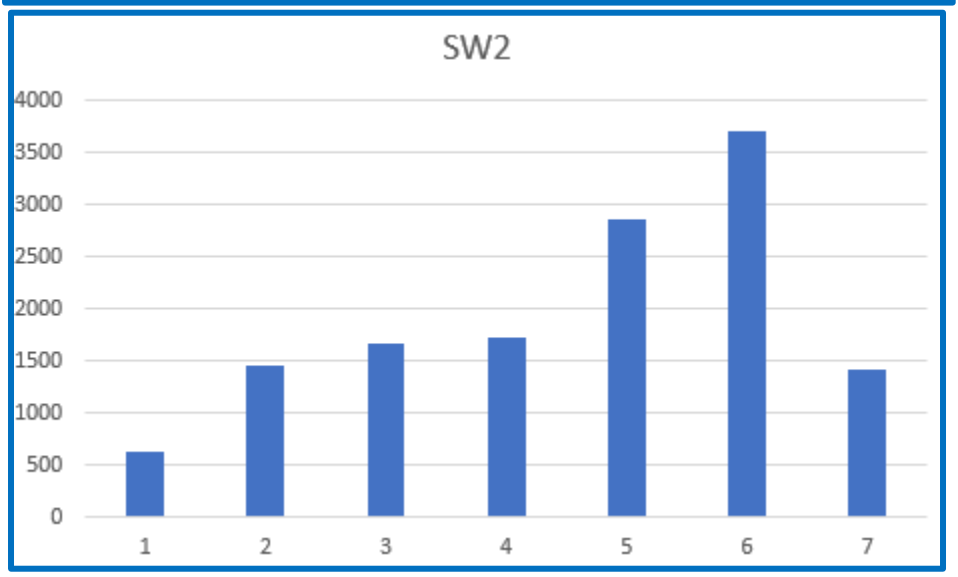
$$\text{c) } P(B|A)=\frac{P(B)P(A)}{P(A)}$$

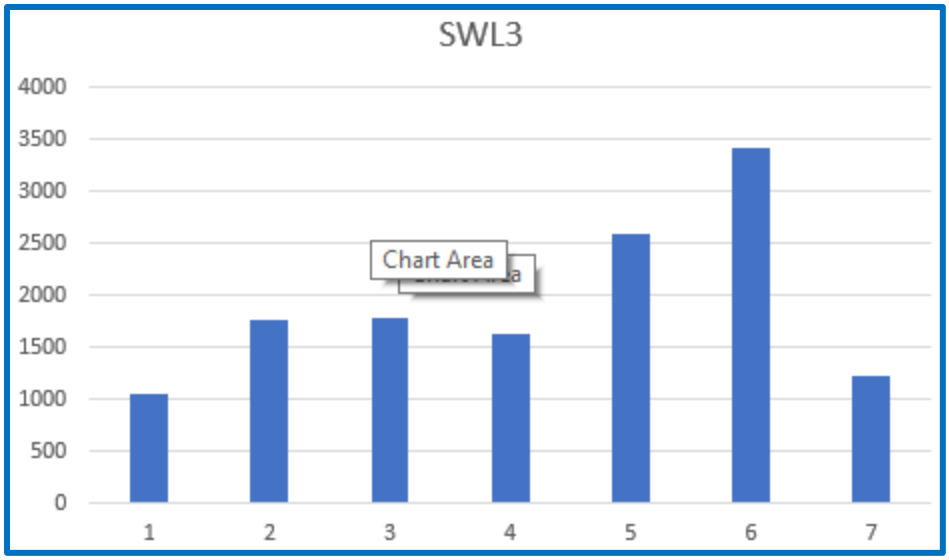
$$\mathrm{P(B|A)}=\frac{\frac{1}{8}}{\frac{6}{8}}=\frac{1}{8}*\frac{8}{6}=\frac{8}{48}=16.67\%$$

This means that there is a 16.67% chance that the NA answer isn't given during the first 50 responses and the NA response was given between the 50th and 75th response.

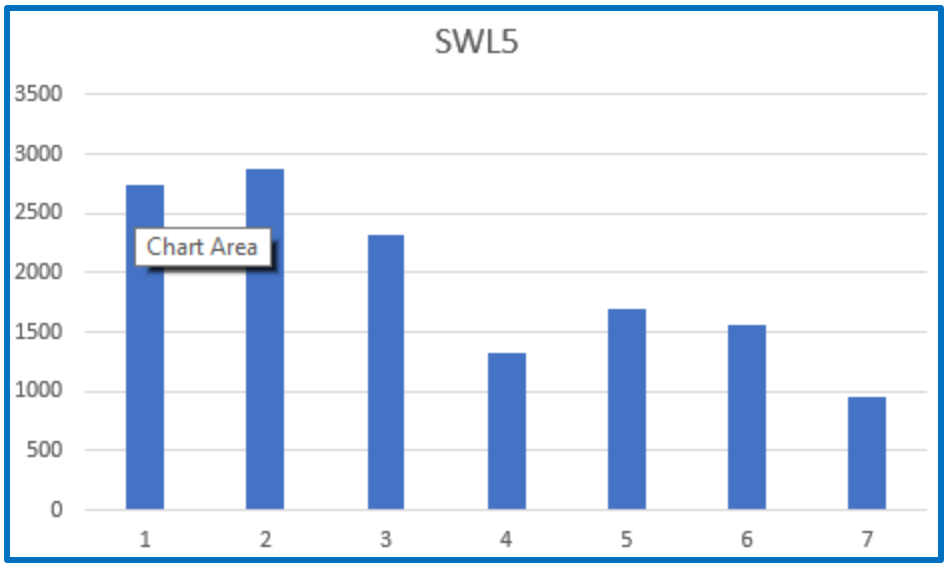
4.5 Normal Probability Distribution

This data set does not provide viable data to create a normal distribution, as the requirements for the survey were very particular. All scale questions did not create a normal distribution in which the mean, media, and mode are all equal. The following are the distributions for the scale questions, none of which create a bell curve or a normal probability distribution.




SWL3
4000
3500
3000
2500
2000
1500
1000
500
0
Chart Area
1
2
3
4
5
6
7

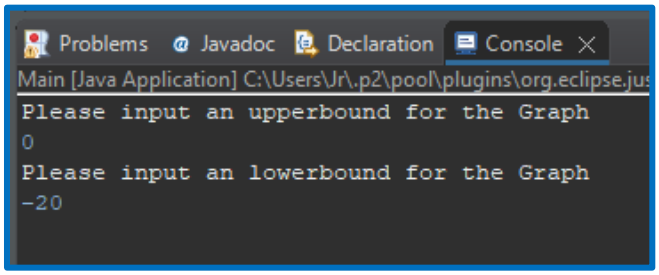

SWL4
3000
2500
2000
1500
1000
500
0
1
2
3
4
5
6
7


SWL5
3500
3000
2500
2000
1500
1000
500
0
Chart Area
1
2
3
4
5
6
7

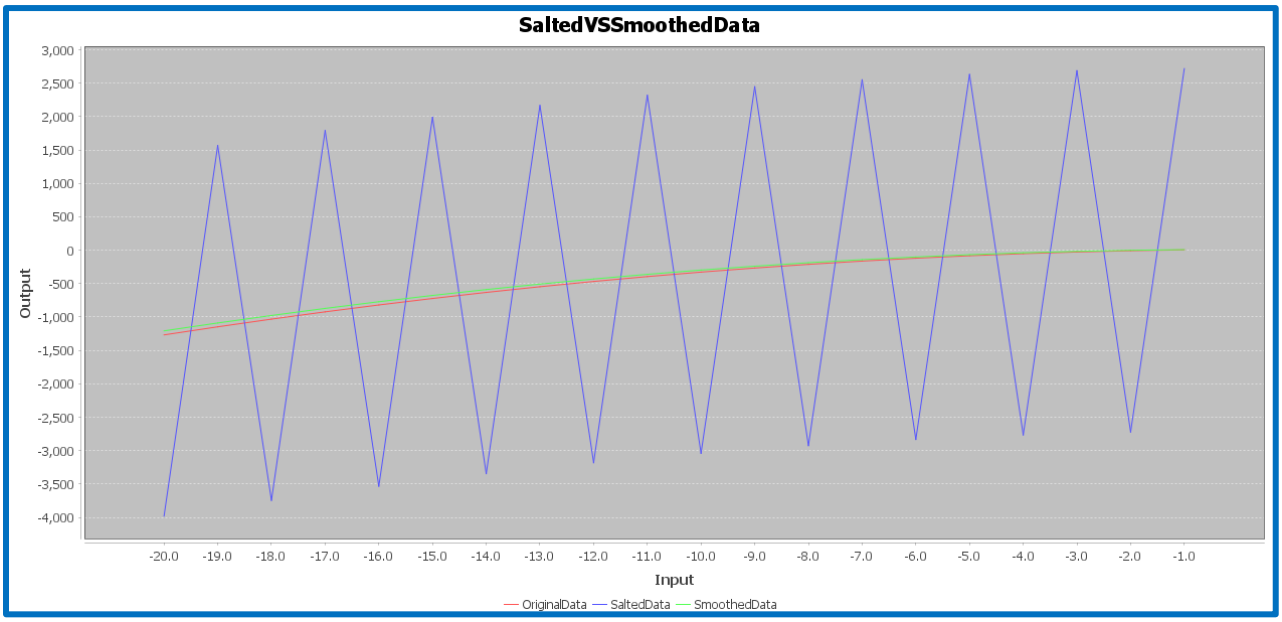
Salter and Smoother

In making the Salter and Smoother we were successful in implementing the JFreeCharts and had great success when we came across any issues with creating the graph and datasets that are displayed within the graph. To get a basic understanding of how we used JFreeCharts, we can look at the basic structure of the graph class.

The graph class is used to create the panel by initializing the JPanel method by setting the size of the window in this case (1280x640), how it should be closed, and by adding the layout of the border and the data and other information that a graph needs. We do this by crating two private classes, createGraph() and createGraphData(). These classes do exactly what their names suggest. The createGraphData() class creates three lines within the dataset that we named data. Each of these lines are referred to as a series, whether it be series1, series2 or series3. These series are used to add the datapoints of the ArrayLists that hold the data that is created when the code is run. Using the addValue method defined in the DefaultCategoryDataset class that comes from JFreeCharts, we can tell it the x and y values of the points as well as what series to add this data point to. In this case instead of adding each datapoint like a lot of the examples suggested when learning how to use JFreeCharts, I used a for loop to add each value found in the ArrayLists. This data is taken and added to the createGraph() class that takes into consideration, the title as well as axis titles. This information is sent up to the graph method and graph is created. When the code is run the user is given a prompt to enter the upper and lower bounds of the x-inputs. It looks like this:



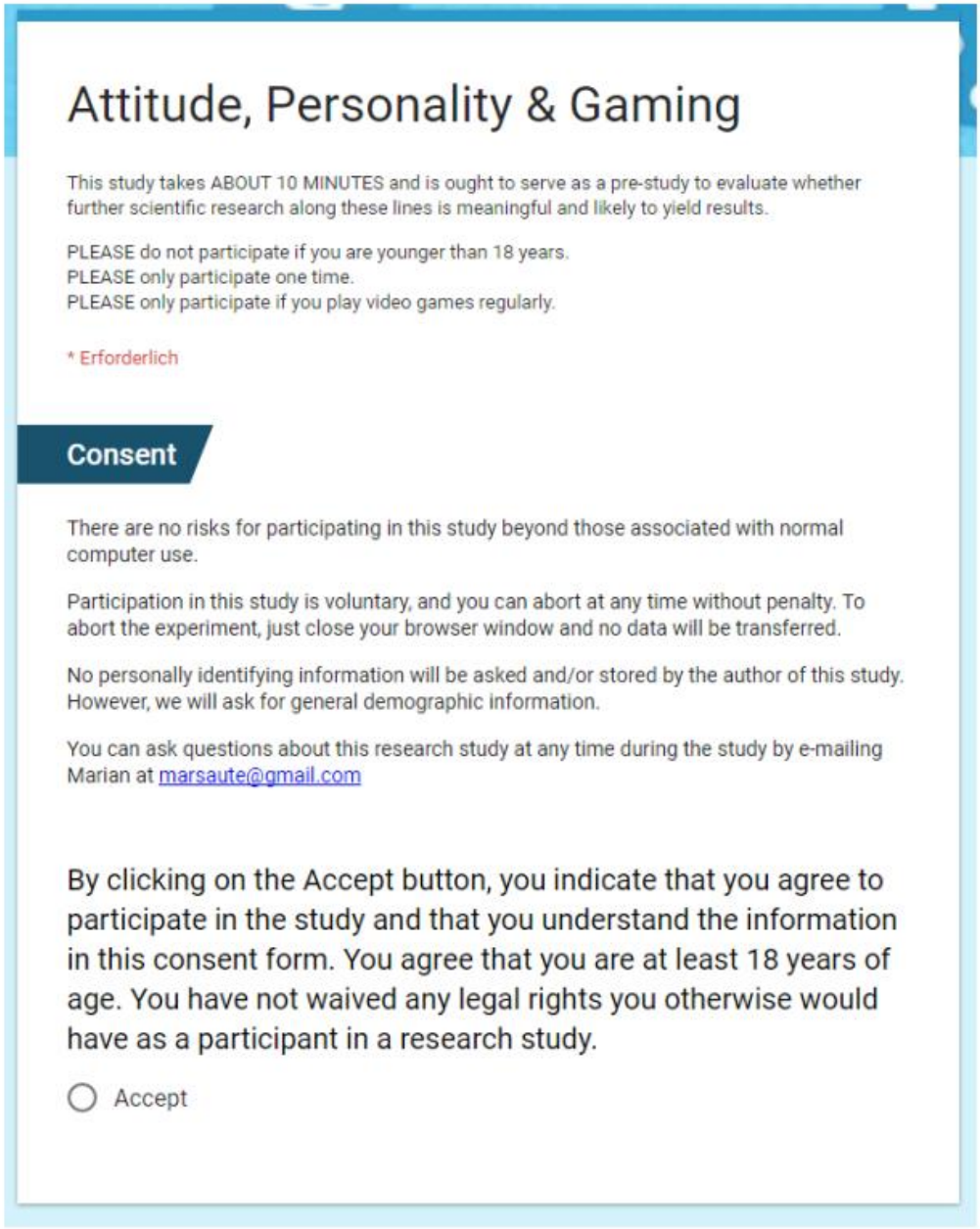
As an example, the inputs of an upper bound of 0 and a lower bound of -20 outputs the following graph:



In this graph we have three different lines, one being the original data points in red, the salted data points in blue, and the smoothed data in green. Not only is the graph displayed but each data ArrayList is output as well. Doing this helped with determining whether or not the code was working as intended and as far as what we wanted it to do it worked well.

Survey Questions Utilized to Obtain Relevant Data

Survey implementation of the questionnaires



Attitude, Personality & Gaming

This study takes ABOUT 10 MINUTES and is ought to serve as a pre-study to evaluate whether further scientific research along these lines is meaningful and likely to yield results.

PLEASE do not participate if you are younger than 18 years.
PLEASE only participate one time.
PLEASE only participate if you play video games regularly.

* Erforderlich

Consent

There are no risks for participating in this study beyond those associated with normal computer use.

Participation in this study is voluntary, and you can abort at any time without penalty. To abort the experiment, just close your browser window and no data will be transferred.

No personally identifying information will be asked and/or stored by the author of this study. However, we will ask for general demographic information.

You can ask questions about this research study at any time during the study by e-mailing Marian at marsaute@gmail.com

By clicking on the Accept button, you indicate that you agree to participate in the study and that you understand the information in this consent form. You agree that you are at least 18 years of age. You have not waived any legal rights you otherwise would have as a participant in a research study.

◯ Accept

1. Questionnaire A

Over the last 2 weeks, how often have you been bothered by the following problems? *

	Not at all	Several days	Over half the days	Nearly every day
1. Feeling nervous, anxious, or on edge	○	○	○	○
2. Not being able to stop or control worrying	○	○	○	○
3. Worrying too much about different things	○	○	○	○
4. Trouble relaxing	○	○	○	○
5. Being so restless that it's hard to sit still	○	○	○	○
6. Becoming easily annoyed or irritable	○	○	○	○
7. Feeling afraid as if something awful might happen	○	○	○	○

If you checked off any problems, how difficult have these made it for you to do your work, take care of things at home, or get along with other people?

- ○ Not difficult at all
- ○ Somewhat difficult
- ○ Very difficult
- ○ Extremely difficult

2. Questionnaire B

DIRECTIONS: Below are five statements with which you may agree or disagree. Using the 1 - 7 scale below, indicate your agreement with each item by crossing the appropriate item. Please be open and honest in your responding.

*

	Strongly Disagree	Disagree	Slightly Disagree	Neither Agree or Disagree	Slightly Agree	Agree	Strongly Agree
1. In most ways my life is close to my ideal.	○	○	○	○	○	○	○
2. The conditions of my life are excellent.	○	○	○	○	○	○	○
3. I am satisfied with life.	○	○	○	○	○	○	○
4. So far I have gotten the important things I want in life.	○	○	○	○	○	○	○
5. If I could live my life over, I would change almost nothing.	○	○	○	○	○	○	○

3. Your game

The next set of questions will focus on the game which you play the most currently. If you play more than one game equally often, please choose one of them.

Which game to you play the most regularly at the moment? *

e.g. "CS:GO", "League of Legends", "Starcraft 2", "FIFA"

Meine Antwort

Which platform do you use to play? *

- ◯ PC
- ◯ Console (PS, Xbox, ...)
- ◯ Smartphone / Tablet

How many hours of the game do you play each week on average? *

e.g. "15"

Meine Antwort

How do you play the most? *

- ◯ Singleplayer
- ◯ Multiplayer - offline (people in the same room)
- ◯ Multiplayer - online - with strangers
- ◯ Multiplayer - online - with online acquaintances or teammates
- ◯ Multiplayer - online - with real life friends
- ◯ Sonstiges:

Is playing this game your hobby or do you make money from it? *

- ◯ I earn a living by playing this game
- ◯ I play mostly for fun but earn a little on the side (tournament winnings, streaming, etc)
- ◯ I play for fun
- ◯ Sonstiges: _____

What is most important to you when playing? *

- ◯ winning
- ◯ improving
- ◯ relaxing
- ◯ having fun
- ◯ Sonstiges: _____

If applicable, which league are you currently in?

Meine Antwort

In addition to playing, how many hours a week do you spend on watching online player/tournament streams / replays / demos, giving or receiving coaching? *

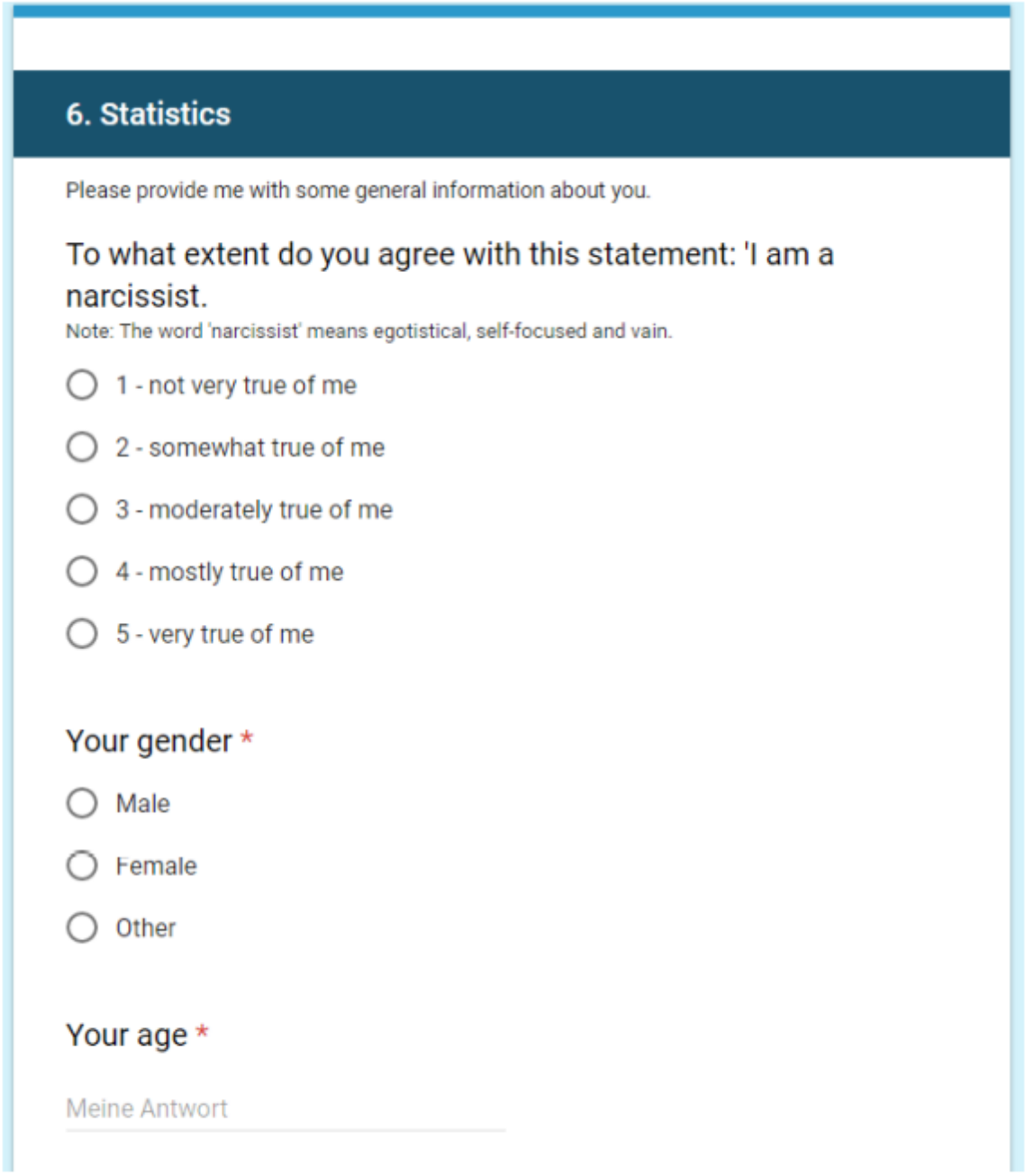
e.g. "15", if you don't, then answer "0"

Meine Antwort

5. Quetionnaire C

Questionnaire C was the Social Phobia Inventory:

Connor, K. M., Davidson, J. R., Churchill, L. E., Sherwood, A., Weisler, R. H., & FOA, E. (2000). Psychometric properties of the social phobia inventory (SPIN). *The British Journal of Psychiatry*, *176*(4), 379-386.



6. Statistics

Please provide me with some general information about you.

To what extent do you agree with this statement: 'I am a narcissist.

Note: The word 'narcissist' means egotistical, self-focused and vain.

- ◯ 1 - not very true of me
- ◯ 2 - somewhat true of me
- ◯ 3 - moderately true of me
- ◯ 4 - mostly true of me
- ◯ 5 - very true of me

Your gender *

- ◯ Male
- ◯ Female
- ◯ Other

Your age *

Meine Antwort

Your work status

- ◯ Student at school
- ◯ Student at college / university
- ◯ Employed
- ◯ Unemployed / between jobs

Your highest degree *

- ◯ None
- ◯ High school diploma (or equivalent)
- ◯ Bachelor (or equivalent)
- ◯ Master (or equivalent)
- ◯ Ph.D., Psy. D., MD (or equivalent)

Your country of birth *

Meine Antwort

Your country of residence *

Meine Antwort

How did you know of this study?

- ◯ TeamLiquid.net
- ◯ Reddit
- ◯ CrowdFlower
- ◯ Other

Sources

Agrawal, D. (2020, August 7). *Online Gaming Anxiety Data*. Kaggle. Retrieved May 2022, from https://www.kaggle.com/datasets/divyansh22/online-gaming-anxiety-data

Wackerly, D. D., Mendenhall, W., & Scheaffer, R. L. (2008). *Mathematical statistics with applications*. Brooks/Cole, Cengage Learning.