Effects of Socialization on Mental Health

STA130 Course Project

Edie Chen Jason Li Zain Mahmoud Rana Nagash
TA Oliver Gatalo
Professor Scott Schwartz

STA130: An Introduction to Statistical Reasoning and Data Science
Department of Statistical Sciences
University of Toronto

Introduction

Social interactions play a pivotal role in shaping individual mental health outcomes. It is becoming increasingly easier, especially for teenagers, to connect with their friends virtually from the comfort of their homes. One may argue that this is harmful for their mental health; is this always the case?

Through this research, we aim to highlight the difference between physically interacting with community members as opposed to virtually connecting with them. Canadian Social Connections Survey (CSCS) to investigate the relationship between various forms of social interactions (physical and non-physical) and how they affect the individuals' mental health states. This presentation outlines the variables we're using, our hypotheses, analyses, key findings, and the conclusions we've drawn from these findings. In this study, we analyze data from the

Our research questions

Question 1

Do the frequency days where an individual spends at least 5 minutes physically socializing lessen an individual's degree of depression?

Question 2

Does playing online games affect how often you feel depressed, and does going outside with friends counteract that?

Question 3

Does video chatting with others make one feel less lonely than text messaging?

Question 1: Variables

Independent variables:

`CONNECTION_social_days_family_p7d_grouped`:
days where individuals spent at least 5 minutes socializing with family.

`CONNECTION_social_days_friends_p7d_grouped`:
days where individuals spent at least 5 minutes socializing with friends.

`CONNECTION_social_days_coworkers_and_classmates_p7d_grouped`:
days where individuals spent at least 5 minutes socializing with co-workers or classmates.

`CONNECTION_social_days_neighbours_p7d_grouped`:
days where individuals spent at least 5 minutes socializing with neighbours.

Dependent variable:
`WELLNESS_phq_score`:
metric used to characterize an individual's level of depression on a scale of 0-6.

Preliminary analysis

After keeping only the columns we're interested in and cleaning the data, we were left with 575 rows and 6 columns.

```
import pandas as pd

# Load the data
file_name = 'Untitled spreadsheet - finalized_data (1).csv'
df = pd.read_csv(file_name)

# Replace empty strings with NaN for easier cleaning
df.replace('', pd.NA, inplace=True)

df =  df.dropna()
# Keep only the relevant columns
columns_to_keep = [
    'CONNECTION_social_days_family_p7d_grouped',
    'CONNECTION_social_days_friends_p7d_grouped',
    'CONNECTION_social_days_coworkers_and_classmates_p7d_grouped',
    'CONNECTION_social_days_neighbours_p7d_grouped',
    'WELLNESS_phq_score_y_n',  # Binary PHQ score
    'WELLNESS_phq_score'       # Continuous PHQ score
]
df_cleaned = df[columns_to_keep]

df_cleaned
df_cleaned.shape
```

```
(575, 6)
```

Figure: 6x575 cleaned dataframe

Preliminary analysis

The independent variables were categorical with 4 categories each:

```
df_cleaned['CONNECTION_social_days_family_p7d_grouped'].unique()
```

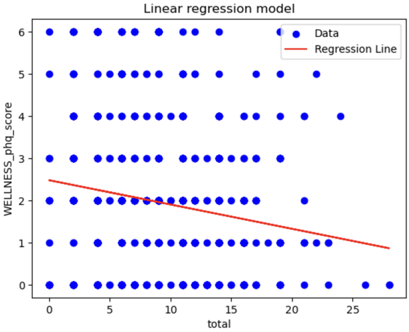
```
array(['None (0 Days)', 'Most days (4 - 6 days)',
       'Some days (1 - 3 days)', 'Every day (7 days)'], dtype=object)
```

Figure: Unique data entries in one of the columns

To better analyze the data, we gave each category a numeric value based on the midpoint of the interval. For example, the 'Most days (4-6)' category was assigned 5 (representing the midpoint of the number of days). Then, we added another column to represent the total number of days where each individual spent at least 5 minutes socializing with any one of the groups above using the numeric values we assigned to each category.

Analysis

First, we examined the relationship between the total column and the numeric PHQ score column. We did this by fitting a simple linear regression through the data.



OLS Regression Results

Dep. Variable:	WELLNESS_phq_score	**R-squared:**	0.026
Model:	OLS	**Adj. R-squared:**	0.025
Method:	Least Squares	**F-statistic:**	15.49
Date:	Sat, 23 Nov 2024	**Prob (F-statistic):**	9.30e-05
Time:	18:39:11	**Log-Likelihood:**	-1153.2
No. Observations:	575	**AIC:**	2310.
Df Residuals:	573	**BIC:**	2319.
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	2.4779	0.158	15.675	0.000	2.167	2.788
total	-0.0576	0.015	-3.936	0.000	-0.086	-0.029

Omnibus:	47.542	**Durbin-Watson:**	1.575
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	53.879
Skew:	0.722	**Prob(JB):**	2.00e-12
Kurtosis:	2.595	**Cond. No.**	22.9

We then created a bootstrapped distribution of model slope coefficients by repeatedly resampling from our original sample and refitting OLS models through the samples. Then, we created a 95% confidence interval of our bootstrapped coefficients for inference.

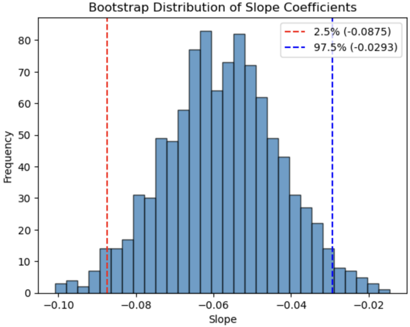


Figure: 95% confidence interval

Summary and conclusion

The confidence interval we constructed only contained negative slopes between -0.0875 and -0.0293 and so we can conclude with 95% confidence that the true value of the slope coefficient lies in that interval. This means that as the number of days where an individual spends at least 5 minutes socializing increases, the average depression score decreases. However, the values of the slopes are very small and so the effect of socializing on depression scores is minuscule (albeit negative).

Question 2: Variables

Independent variables:

CONNECTION_activities_onlinegames_p3m:
how often an individual has played online games in the past 3 months

CONNECTION_activities_walk_p3m:
how often an individual has gone on a walk with friends in the past 3 months

Dependent variable:

WELLNESS_malach_pines_burnout_measure_depressed:
how often an individual feels depressed

UNIVERSITY OF TORONTO

Cleaning data

First, I want to assign numbers to the ordinal categories of how often an individual feels depressed.
I will just use the consecutive numbers 0 (never) through 5 (always).

```python
# Mapping the variables to numeric values
mapping_dict = {
    'WELLNESS_malach_pines_burnout_measure_depressed': {
        'Never': 0,
        'Almost never': 1,
        'Rarely': 2,
        'Sometimes': 3,
        'Very Often': 4,
        'Always': 5
    }
}
```

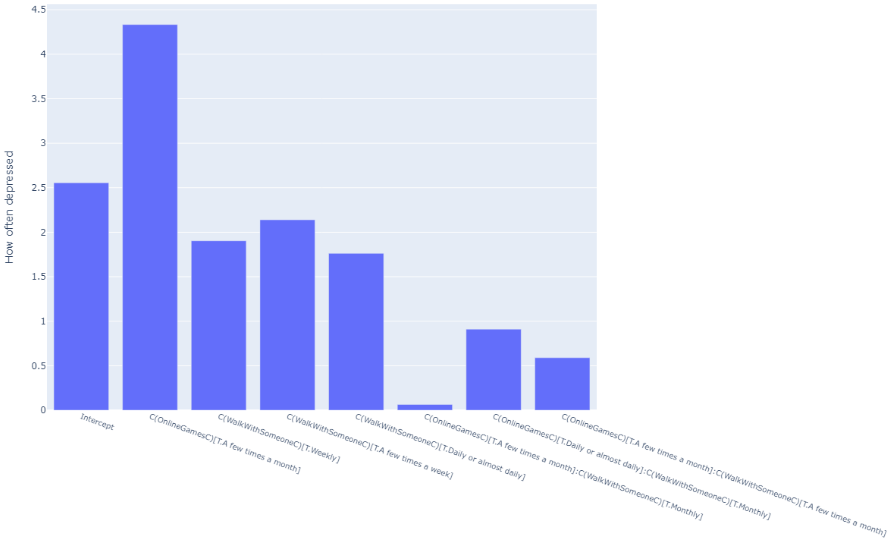
After renaming variables, removing empty values, etc., this is what my DataFrame looks like.

The boolean values True are whenever the categorical values are not never/not in the past three months.

	OnlineGamesC	WalkWithSomeoneC	DepressionC	DepressionN	OnlineGamesB	WalkWithSomeoneB
0	Not in the past three months	Daily or almost daily	Rarely	2.0	False	True
1	Not in the past three months	A few times a week	Almost never	1.0	False	True
2	Not in the past three months	A few times a month	Almost never	1.0	False	True
3	Weekly	Less than monthly	Rarely	2.0	True	True
4	Weekly	Monthly	Almost never	1.0	True	True
...	...	...	...	...	...	...

UNIVERSITY OF TORONTO

Bar plot



How often depressed
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
Intercept
C(OnlineGamesC)[T.A few times a month]
C(WalkWithSomeoneC)[T.Weekly]
C(WalkWithSomeoneC)[T.A few times a week]
C(WalkWithSomeoneC)[T.Daily or almost daily]
C(OnlineGamesC)[T.A few times a month]:C(WalkWithSomeoneC)[T.Monthly]
C(OnlineGamesC)[T.Daily or almost daily]:C(WalkWithSomeoneC)[T.Monthly]
C(OnlineGamesC)[T.A few times a month]:C(WalkWithSomeoneC)[T.A few times a month]

Regression

I will perform a linear regression, with interactions. I will retain only the outcomes with a p-value ≥ 0.05 for simplicity.

```python
# Fit the OLS model using categorical values (OnlineGamesC and SocialFriendsC)
ols_model_c = smf.ols("DepressionN ~ C(OnlineGamesC) * C(WalkWithSomeoneC)", data=df)

fitted_ols_model_c = ols_model_c.fit()
```

```python
summary = fitted_ols_model_c.summary()

# The summary is a text object, so we need to extract the coefficients and p-values from it
# We can use the summary.tables[1] for extracting the coefficient table
summary_lines = summary.tables[1].data

# Convert the summary table into a DataFrame
summary_df = pd.DataFrame(summary_lines[1:], columns=summary_lines[0])

# Convert the p-values to float type and filter
summary_df['P>|t|'] = summary_df['P>|t|'].astype(float)
filtered_summary = summary_df[summary_df['P>|t|'] <= 0.05]
```

```python
coefficients = filtered_summary['coef'].astype(float)
intercept = coefficients[0]

values = intercept + coefficients
values.loc[0] = intercept

values_rounded = values.astype(int)

depression_values = [list(mapping_dict["WELLNESS_malach_pines_burnout_measure_depressed"].keys())[value] for value in val

pd.set_option('display.max_colwidth', None)  # Show full content in each cell

filtered_summary = filtered_summary.assign(value=values, depression_value=depression_values)
filtered_summary
```

		coef	std err	t	P>|t|	[0.025	0.975]	value	depression_value
0	Intercept	2.5546	0.122	21.001	0.000	2.316	2.793	2.5546	Rarely
3	C(OnlineGamesC)[T.A few times a month]	1.7787	0.776	2.293	0.022	0.255	3.302	4.3333	Very Often
10	C(WalkWithSomeoneC)[T.Weekly]	-0.6510	0.190	-3.430	0.001	-1.024	-0.278	1.9036	Almost never
11	C(WalkWithSomeoneC)[T.A few times a week]	-0.4154	0.193	-2.157	0.031	-0.794	-0.037	2.1392	Rarely
12	C(WalkWithSomeoneC)[T.Daily or almost daily]	-0.7927	0.238	-3.328	0.001	-1.260	-0.325	1.7619	Almost never
21	C(OnlineGamesC)[T.A few times a month]:C(WalkWithSomeoneC)[T.Monthly]	-2.4898	0.997	-2.498	0.013	-4.447	-0.533	0.0648	Never
24	C(OnlineGamesC)[T.Daily or almost daily]:C(WalkWithSomeoneC)[T.Monthly]	-1.6440	0.777	-2.115	0.035	-3.170	-0.118	0.9106	Never
27	C(OnlineGamesC)[T.A few times a month]:C(WalkWithSomeoneC)[T.A few times a month]	-1.9651	0.987	-1.991	0.047	-3.903	-0.027	0.5895	Never

UNIVERSITY OF TORONTO

Question 2 Variables

Does video chatting with others make one feel less lonely than text messaging others?

Independant Variables:

Video chatted with friends/family in the past 3 months:

Texted or messaged someone in the past 3 months:

Options:

> "Not in the past three months", "Less than monthly", "Monthly", "A few times a month", "Weekly", "A few times a week", "Daily or almost daily"

Dependant Variables:

How many days felt lonely in the past week:

Options:

> 'None of the time (e.g., 0 days)': 0, 'Rarely (e.g. less than 1 day)': 0.5, 'Some or a little of the time (e.g. 1-2 days)' : 1.5, 'Occasionally or a moderate amount of time (e.g. 3-4 days)': 3.5, 'All of the time (e.g. 5-7 days)': 6

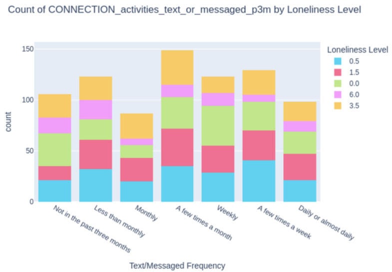
UNIVERSITY OF TORONTO

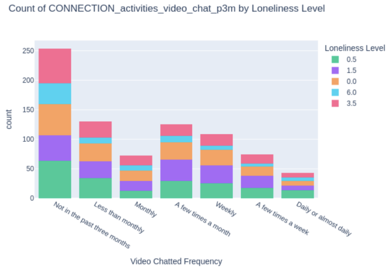
Assumptions

Converting the categorical variable to numerical values makes it not completely accurate since it assumes an exact number of days that they feel lonely

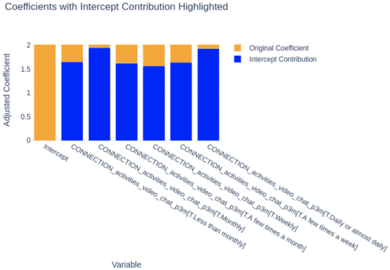
How many days people feel lonely in a week is not reflective of how lonely they feel on average

Visualization of the Raw Data





Simple Linear Regression Data Wrangling



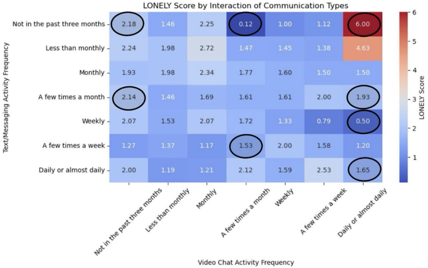
We first tried to analyse the 2 factors separately with simple linear regression

The fitting of the model was poor (R square = 0.01)

So we decide to analyse with a multilinear regression model

Final Result Analysis

Multilinear Regression with Heat Map Visualization.



The circled ones are the final value which got a p value lower than 0.05 (meaningful).
The r-squared is 0.065 for the multilinear regression

Conclusions:

Video chatting daily or almost daily without texting in the past three months leads to high levels of loneliness

Video chatting daily or almost daily and texting weekly leads to low levels of loneliness

Video chatting a few times a month and texting a few times a month also leads to low levels of loneliness

*Reminder that this is not completely accurate due to assumptions mentioned previously

UNIVERSITY OF TORONTO

Conclusion

Connections

Aim to analyze how socializing in different forms affects people's mental health

branched off into 2 directions to find out the effect of interaction when it's offline or online.

Findings

Socializing and Depression:

The average depression score decreases when physical socialization $>= 5$ min/day

Depression levels also drop when taking a walk with others.

Playing video games, however, may have an opposing effect on level depression.

Socializing and Loneliness:

Contacting others a few times a month via video chat and test seems to reduce loneliness the most.

Acknowledgments:
We would like to thank Dr. Kiffer Card and Ms. Christine Ovcaric for providing this valuable opportunity. Special thanks to Prof. Scott, our TA, and all the teaching assistants for their guidance and support throughout the project.