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

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

UNIVERSITY OF TORONTO

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?

Important theorem

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

Question 3

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

UNIVERSITY OF TORONTO

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

UNIVERSITY OF TORONTO

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

UNIVERSITY OF TORONTO

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)
```

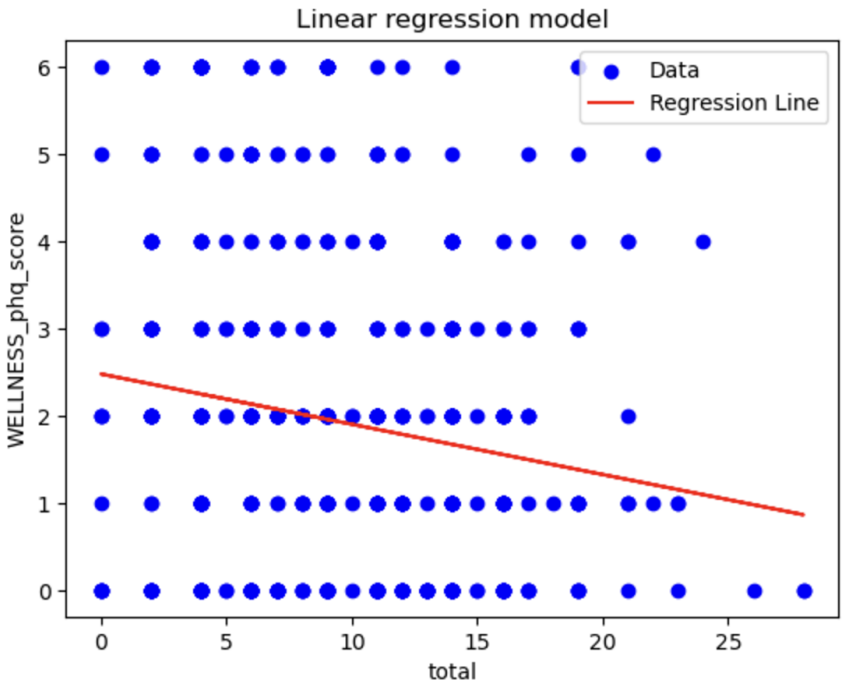
Figure: Unique data entries in one of the columns

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.

UNIVERSITY OF TORONTO

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

Analysis

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.

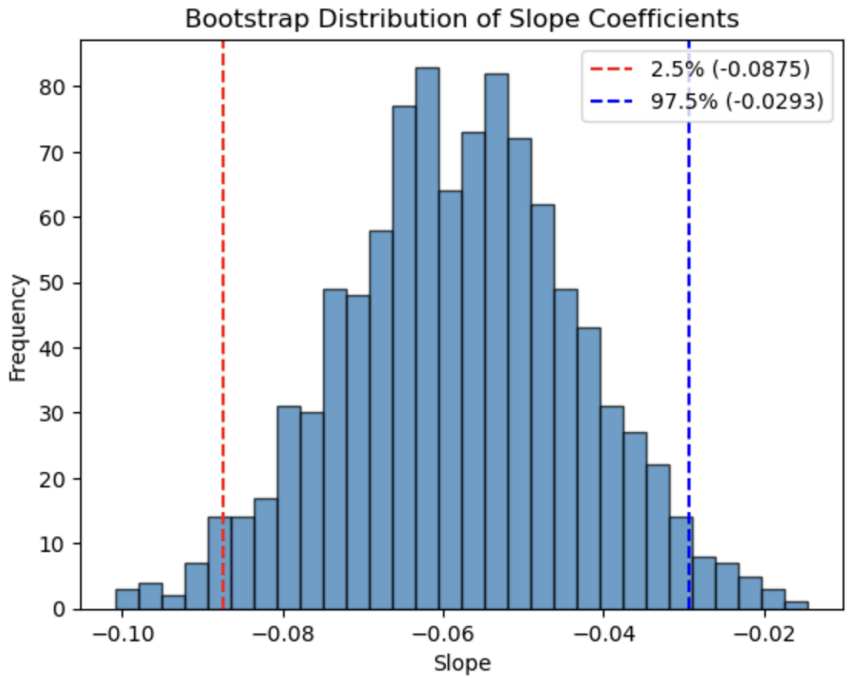


Figure: 95% confidence interval

UNIVERSITY OF TORONTO

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).

UNIVERSITY OF TORONTO

Frame Title

Frame Title