Finding Relationships Between Global Socioeconomic Factors and Mental Illness

Gabriella Weis

Honors Applied Statistics

June 14, 2024

Table of Contents

Introduction and Literature Review

Mental health is an important factor of a country's overall wellbeing. Though culture plays a massive part in the recognition and treatment of mental illness, research suggests there exists a variety of global factors that may be useful in predicting the prevalence of mental illness in a country. Lower mental health could require more medical spending, result in or be a result of lower social cohesion, or be affected by things like education and unemployment. In an attempt to find out if these things are at all true, it's only natural to search for a two way correlation between the socioeconomic factors of a country and its rates of mental illness. Therefore, my research questions are as follows: what social and economic factors influence a country's rates of mental illness, and how does mental health affect the social and economic factors of a country in turn?

My initial hypotheses are as follows:

ANOVA 1:

Null hypothesis - The mean rate of anxiety disorders stays the same across all HDI levels

Alternate hypothesis - At least one mean rate of anxiety disorders is different across HDI levels

ANOVA 2:

Null hypothesis - The mean rate of depression stays the same across all HDI levels

Alternate hypothesis - At least one mean rate of depression is different across HDI levels

Model Utility F-tests:

Null hypothesis - There is no useful relationship between the rate of anxiety disorders, depression, and country GDP

Alternate hypothesis - There exists a useful relationship between the rate of anxiety disorders, depression, and country GDP

Null hypothesis - There is no useful relationship between the rate of anxiety disorders, depression, and country HDI

Alternate hypothesis - There exists a useful relationship between the rate of anxiety disorders, depression, and country HDI

Null hypothesis - There is no useful relationship between the rate of anxiety disorders, depression, and country unemployment

Alternate hypothesis - There exists a useful relationship between the rate of anxiety disorders, depression, and country unemployment

2-Sample T-test:

Null hypothesis - The means of Social Capital Index ratings between countries with high vs. low rates of suicide are equal.

Alternate hypothesis - The means of Social Capital Index ratings between countries with high vs. low rates of suicide are not equal.

In the first article I explored from this field, titled "Social Determinants of Mental Health: Where We Are and Where We Need to Go," the authors spoke of evidence that familial relationships and family structure influence mental health, "unemployment, precarious employment, and employment conditions continue to be routinely linked to increased psychological distress," and "direct and indirect experiences of community violence in adolescence have been significantly associated with elevated depressive, anxiety, and PTSD symptoms." They also noted that mental health has a reciprocal impact on social determinants. According to the article, "the World Health Organization has described how mental health symptoms in each stage of life can negatively impact socioeconomic status and other social determinants in a cumulative and dynamic manner."

The second article, titled "Suicide Rate, Depression and the Human Development Index: An Ecological Study From Mexico," attempts to assess the contribution of depression, the human development index (HDI), and a collection of other factors to suicide rates in Mexico. There was found to be a strong positive relationship between suicide rate and non-family households, and the article writes that "based on the result of this study, it is possible to assume that, as the HDI increases, there is a greater possibility of living alone and having suicidal behavior."

The final article, named "The social determinants of mental health and disorder: evidence, prevention and recommendations," explains how "people exposed to more unfavourable social circumstances are more vulnerable to poor mental health over their life course," and that much evidence for this lays in minority groups, including refugees, asylum seekers and displaced persons, ethnoracial minoritized groups, LGBTQ+ groups, and those living in poverty. The article says that there is very strong evidence that the risk of developing any mental illness is inextricably connected to life circumstances. It also talks about possible mitigation so that poor mental health outcomes may be lessened by manipulating social determinants.

There is increasing evidence for strong correlations between social determinants and mental health, so I would like to create predictive models for one or more of the specific rates of mental illness and a model to predict GDP or HDI of a country using its rates of mental illness. I have not seen them in any of the studies or articles I explored, but by building these models, I can affirm past research and evidence, identifying important factors and discarding less significant ones.

The uses of finding significant relationships between socioeconomic factors and mental health status are widespread. Once important factors are identified, countries may be able to attempt social or economic reform accordingly in order to prevent mental illness before it begins and promote mental health. For example, if unemployment were to be an incredibly reliable predictor of anxiety, countries might consider launching a job support program to raise employment and lessen their anxiety rate. On the other hand, if GDP were to indicate a higher rate of mental illness in a population, an examination of that country's mental health support system to find out where exactly the money's going and how their healthcare system operates may be helpful.

Methods, Process, and Approach

All data sources:

- GDP: https://www.imf.org/external/datamapper/NGDPD@WEO/OEMDC/ADVEC/WEOWORLD
- Social Capital Index: https://solability.com/the-global-sustainable-competitiveness-index/the-index/social-capital
- Mental Health Disorder Rates: https://www.kaggle.com/datasets/thedevastator/uncover-global-trends-in-mental-health-disorder
- HDI: https://hdr.undp.org/data-center/documentation-and-downloads
- Suicide Rates: https://worldpopulationreview.com/country-rankings/suicide-rate-by-country
- Unemployment %: https://www.kaggle.com/datasets/mjshri23/life-expectancy-and-socio-economic-world-bank

Final compiled datasheet: Final Mental Health Dataset
Data separated by test: ANOVAs , F-test 1 , F-test 2 , F-test 3 , T-test

I began with the dataset of global mental health disorder rates because it was the core of my research questions, had the most data, and would serve as a good baseline. Unfortunately, its most recent data was from 2017 (and I wanted to make everything as accurate as possible to present day), so I was forced to accommodate my other data to match. Thankfully GDP, HDI, suicide rates, and unemployment percentage are fairly well-recorded, and I was able to find data for that time frame (with some discrepancies). I collected all of my data from the above sources and compiled it into one google sheet, eliminating countries with no/not enough data. I didn't need to take a random sample because my entire population consists of 193 unique data points.

My first tests are two ANOVAs because they will allow me to determine if there's any statistically significant differences between mental illness rates and country HDI and GDP. The first ANOVA will have the null hypothesis that the mean rate of anxiety disorders stays the same across all HDI levels and the alternate hypothesis that at least one mean rate of anxiety disorders is different across HDI levels. The second ANOVA will have the null hypothesis that the mean rate of depression stays the same across all HDI levels and the alternate hypothesis that at least one mean rate of depression is different across HDI levels. If the alternate hypothesis is proven, it may mean that the rates of mental health issues in certain countries may depend on HDI levels in that country or vice versa (not taking into account other factors or confounding variables).

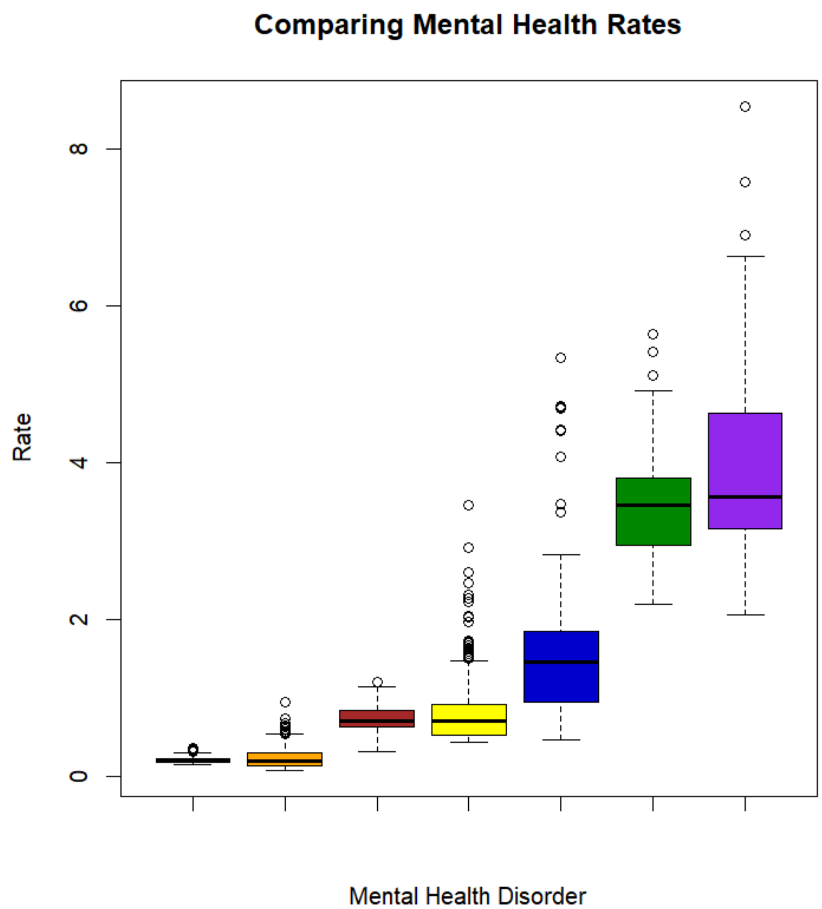
My second group of tests are Model Utility F-tests. The first F-test will compare a country's rates of anxiety disorders and depression with its GDP because I want to see if there's a correlation between a country's wealth and its anxiety and depression. It will have the null hypothesis that there is no useful relationship between the rate of anxiety disorders, depression, and country GDP and the alternate hypothesis that there exists a useful relationship between the rate of anxiety disorders, depression and country GDP. The second F-test will compare a country's rate of anxiety disorders and depression with its HDI because I want to see if there's a correlation between a country's development and its anxiety and depression. It will have the null hypothesis that there is no useful relationship between the rate of anxiety disorders, depression and country HDI and the alternate hypothesis that there exists a useful relationship between the rate of anxiety disorders, depression, and country HDI. The last F-test will compare a country's rate of anxiety disorders and depression with its rate of unemployment because I want to see if there's correlation between a country's unemployment rate and its anxiety and depression. Its hypotheses will follow the same structure as those previous.

My final test will be a 2-Sample T-test comparing the means of social index ratings (rating on people's confidence in their country) between countries with high vs. low rates of suicide because I want to see if there is any affect of societal/country confidence on suicide. It will have the null hypothesis that the means of social index ratings between countries with high vs. low rates of suicide are equal and the alternate hypothesis that the means of social index ratings between countries with high vs. low rates of suicide are not equal.

My predictive models will consist of two linear regression models, one that uses a country's social and economic factors to predict rates of mental illnesses and the other that uses a country's mental illness rates to predict social and economic factors. I want to create these models based off of the results of my test explorations, particularly off of the model utility F-tests, finding a predictable group of socioeconomic factors for mental illness and a predictable group of mental illness factors for country prosperity.

Results, Product, and Findings

First, I created a boxplot comparing all rates of mental health disorders:



[From left to right: Schizophrenia (%), Eating disorders (%), Bipolar disorder (%), Drug use disorders (%), Alcohol use disorders (%), Depression (%), Anxiety disorders (%).]

Because anxiety has the greatest median, IQR, and range, I chose to use it in my ANOVAs. These two tests are meant to determine if there's any statistically significant differences between mental illness rates and country HDI and GDP.

ANOVA 1:

Null hypothesis - The mean rate of anxiety disorders stays the same across all HDI levels
Alternate hypothesis - At least one mean rate of anxiety disorders is different across HDI levels

μ_1= mean rate of anxiety disorders in countries with VH (Very High) HDI

μ_2= mean rate of anxiety disorders in countries with H (High) HDI

μ_3= mean rate of anxiety disorders in countries with M (Medium) HDI

μ_4= mean rate of anxiety disorders in countries with L (Low) HDI

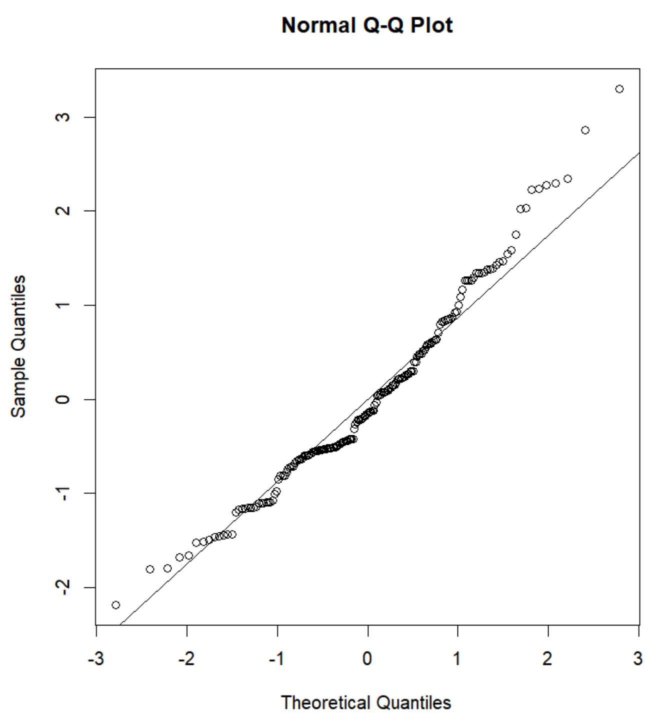
H_o: $\mu_1 = \mu_2 = \mu_3 = \mu_4 = 0$

H_a: At least one of the means is not equal to zero.

$\alpha = 0.05$

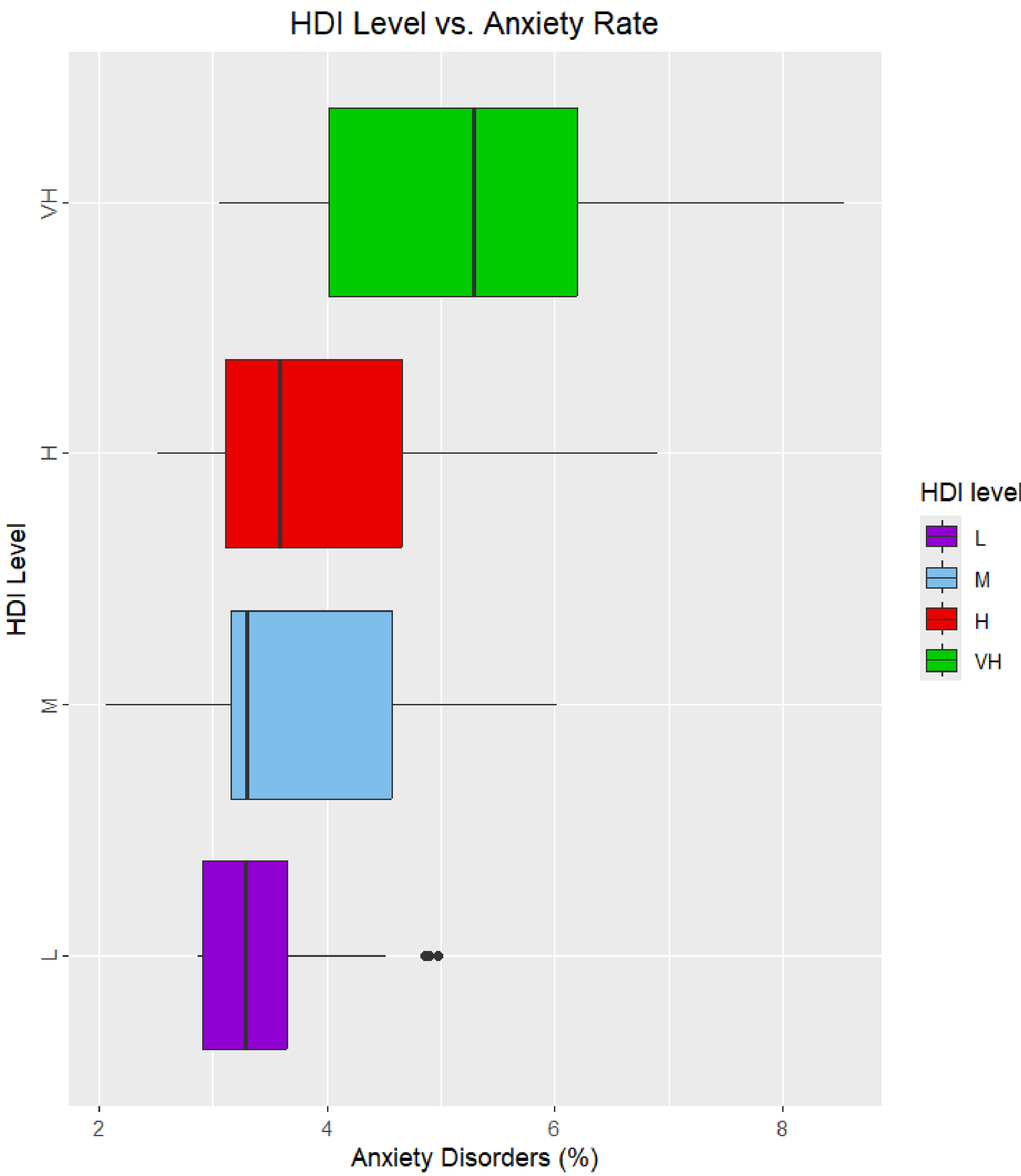
Checks:

- ☑ Normality: The data points on the Normal Probability plot seem to follow the line of best fit fairly closely, except at the tail ends, where it diverges slightly. It appears mostly normally distributed, if slightly skewed right.



☑ Equal Variance: Boxplots appear very varied, and the largest standard deviation VH (Very High) is over two times the size of the smallest L (Low). Proceed with caution.

L	M	H	VH
0.5712062	0.9208133	1.1316126	1.3583070



☑ Independence: Did not use random selection because the data was meant to be the entire population of countries (excluding data points with no data). Proceed with caution.

<u>Calculations:</u>

VH	H	M	L
$\bar{x}_1 = 5.243589$ $s_1 = 1.3583070$ $n_1 = 27$	$\bar{x}_2 = 4.040604$ $s_2 = 1.1316126$ $n_2 = 63$	$\bar{x}_3 = 3.720736$ $s_3 = 0.9208133$ $n_3 = 41$	$\bar{x}_4 = 3.428337$ $s_4 = 0.5712062$ $n_4 = 56$

N = $n_1 + n_2 + n_3 + n_4$

N = 27 + 63 + 41 + 56 = 187

T = $n_1(\bar{x}_1) + n_2(\bar{x}_2) + n_3(\bar{x}_3) + n_4(\bar{x}_4)$

T = 27(5.243589) + 63(4.040604) + 41(3.720736) + 56(3.428337) = 740.672

$\bar{\bar{x}} = \frac{T}{N} = \frac{740.672}{187} = 3.9608$

SSTr = $n_1(\bar{x}_1 - \bar{\bar{x}})^2 + n_2(\bar{x}_2 - \bar{\bar{x}})^2 + n_3(\bar{x}_3 - \bar{\bar{x}})^2 + n_4(\bar{x}_4 - \bar{\bar{x}})^2$

SSTr = $27(5.243589 - 3.9608)^2 + 63(4.040604 - 3.9608)^2 + 41(3.720736 - 3.9608)^2 + 56(3.428337 - 3.9608)^2 = 63.0708$

SSE = $(n_1 - 1)(s_1)^2 + (n_2 - 1)(s_2)^2 + (n_3 - 1)(s_3)^2 + (n_4 - 1)(s_4)^2$

SSE = $26(1.3583070)^2 + 62(1.1316126)^2 + 40(0.9208133)^2 + 55(0.5712062)^2 = 179.225$

MSTr = $\frac{SSTr}{k-1} = \frac{63.0708}{3} = 21.0236$

MSE = $\frac{SSE}{N-k} = \frac{179.225}{183} = 0.97937$

F = $\frac{MSTr}{MSE} = \frac{21.0236}{0.97937} = 21.4664$ w/df = $\frac{k-1}{N-k} = \frac{3}{183} = \frac{1}{61}$

p-value = p(F > 21.4664) = $5.85 \times 10^{-12} \approx 0$

Since the p-value(≈ 0) > α(0.05), we reject the null hypothesis. Therefore, there is convincing evidence to suggest that at least one mean rate of anxiety disorders is different across the HDI levels.

Source	df	SS	MS	F	p-value
HDI level	3	63.0708	21.0236	21.4664	≈ 0
Error	183	179.225	0.97937		
Total	186	242.2958			

<u>R Output:</u>

```
Df Sum Sq Mean Sq F value   Pr(>F)
`HDI level`   3  63.07  21.024   21.47 5.85e-12 ***
Residuals   183 179.22   0.979
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

ANOVA 2:

Null hypothesis - The mean rate of depression stays the same across all HDI levels
Alternate hypothesis - At least one mean rate of depression is different across HDI levels

μ_1= mean rate of depression in countries with VH (Very High) HDI

μ_2= mean rate of depression in countries with H (High) HDI

μ_3= mean rate of depression in countries with M (Medium) HDI

μ_4= mean rate of depression in countries with L (Low) HDI

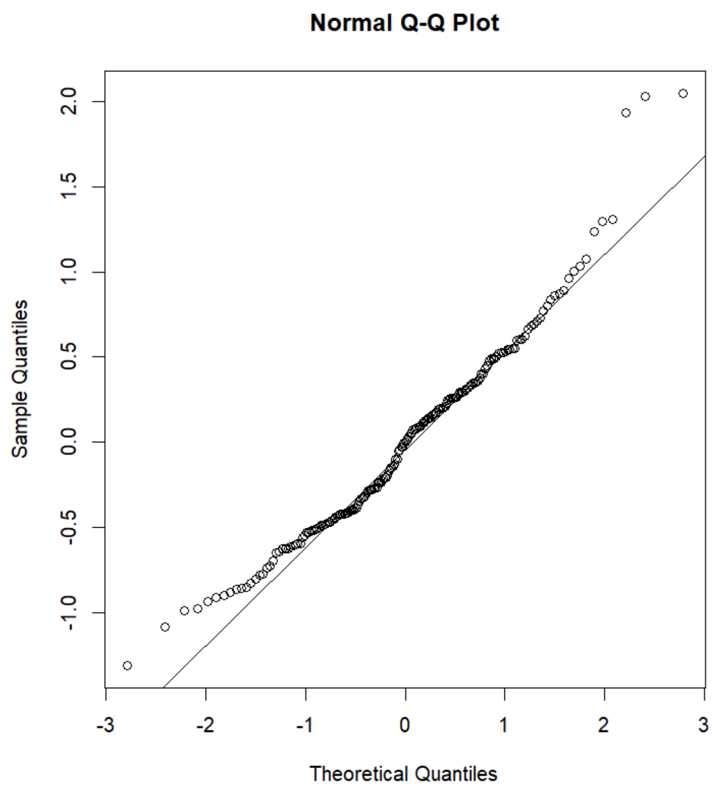
H_o: $\mu_1 = \mu_2 = \mu_3 = \mu_4 = 0$

H_a: At least one of the means is not equal to zero.

$\alpha = 0.05$

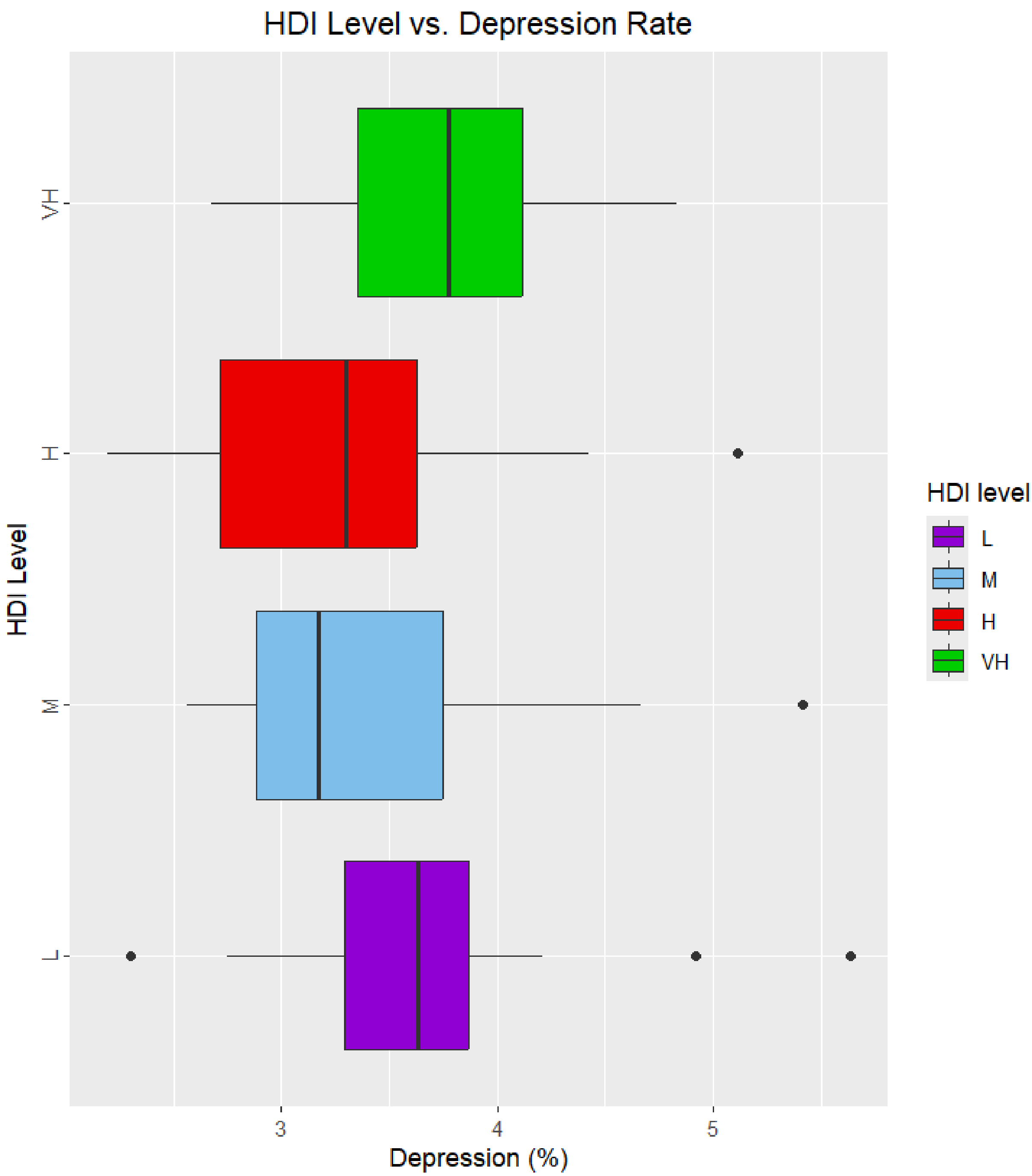
<u>Checks:</u>

- ☑ Normality: The data points on the Normal Probability plot seem to follow the line of best fit fairly closely, except at the tail ends, where it diverges slightly. It appears mostly normally distributed, if slightly skewed right.



☑ Equal Variance: Boxplots have a relatively similar spread, despite several outliers, and the largest standard deviation H (High) is less than two times the size of the smallest L (Low).

L	M	H	VH
0.5124745	0.6137906	0.6280573	0.5780662



☑ Independence: Did not use random selection because the data was meant to be the entire population of countries (excluding data points with no data). Proceed with caution.

Calculations:

VH	H	M	L
$\bar{x}_1 = 3.760700$ $s_1 = 0.5780662$ $n_1 = 27$	$\bar{x}_2 = 3.184104$ $s_2 = 0.6280573$ $n_2 = 63$	$\bar{x}_3 = 3.368812$ $s_3 = 0.6137906$ $n_3 = 41$	$\bar{x}_4 = 3.608863$ $s_4 = 0.5124745$ $n_4 = 56$

$N = n_1 + n_2 + n_3 + n_4$

$N = 27 + 63 + 41 + 56 = 187$

$T = n_1(\bar{x}_1) + n_2(\bar{x}_2) + n_3(\bar{x}_3) + n_4(\bar{x}_4)$

$T = 27(3.760700) + 63(3.184104) + 41(3.368812) + 56(3.608863) = 642.355$

$\bar{\bar{x}} = \frac{T}{N} = \frac{642.355}{187} = 3.4351$

$SSTr = n_1(\bar{x}_1 - \bar{\bar{x}})^2 + n_2(\bar{x}_2 - \bar{\bar{x}})^2 + n_3(\bar{x}_3 - \bar{\bar{x}})^2 + n_4(\bar{x}_4 - \bar{\bar{x}})^2$

$SSTr = 27(3.760700 - 3.4351)^2 + 63(3.184104 - 3.4351)^2 + 41(3.368812 - 3.4351)^2 +$
$56(3.608863 - 3.4351)^2 = 8.7023$

$SSE = (n_1 - 1)(s_1)^2 + (n_2 - 1)(s_2)^2 + (n_3 - 1)(s_3)^2 + (n_4 - 1)(s_4)^2$

$SSE = 26(0.5780662)^2 + 62(0.6280573)^2 + 40(0.6137906)^2 + 55(0.5124745)^2 = 62.6587$

$MSTr = \frac{SSTr}{k-1} = \frac{8.7023}{3} = 2.9008$

$MSE = \frac{SSE}{N-k} = \frac{62.6587}{183} = 0.3424$

$F = \frac{MSTr}{MSE} = \frac{2.9008}{0.3424} = 8.4719$ w/df $= \frac{k-1}{N-k} = \frac{3}{183} = \frac{1}{61}$

p-value $= p(F > 8.4719) = 2.67 \times 10^{-5} \approx 0$

Since the p-value(≈ 0) > α(0.05), we reject the null hypothesis. Therefore, there is convincing evidence to suggest that at least one mean rate of depression is different across the HDI levels.

Source	df	SS	MS	F	p-value
HDI level	3	8.7023	2.9008	8.4719	≈ 0
Error	183	62.6587	0.3424		
Total	186	71.361			

<u>R Output:</u>

```
Df Sum Sq Mean Sq F value   Pr(>F)
`HDI level`   3   8.70  2.9008   8.472 2.67e-05 ***
Residuals   183  62.66  0.3424
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The following three model utility F-tests will be used to search for correlations between a country's GDP, HDI, and unemployment and its rates of anxiety and depression.

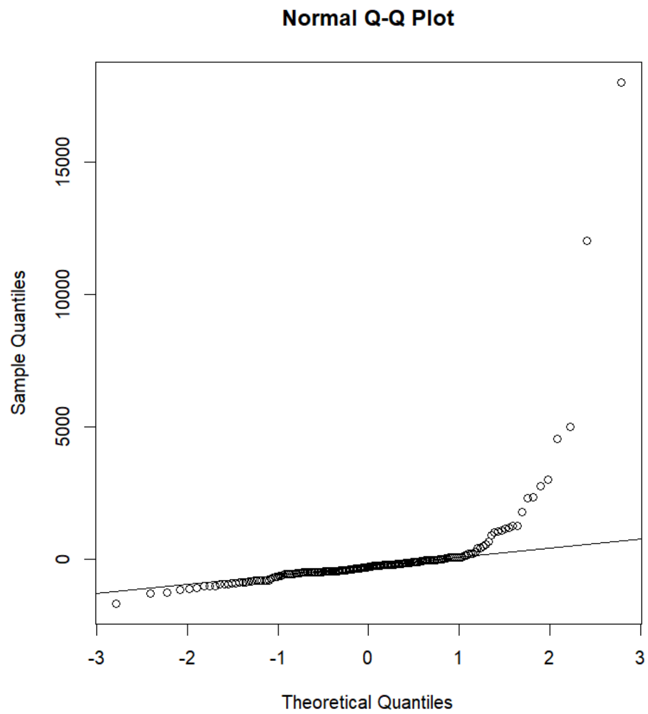
Model Utility F-test 1:

$\hat{y}$ = predicted GDP
x_1 = rate of anxiety disorders
x_2 = rate of depression
β_1 = estimated increase in predicted GDP when the rate of anxiety disorders increases by one unit.
β_2 = estimated increase in predicted GDP when the rate of depression increases by one unit.

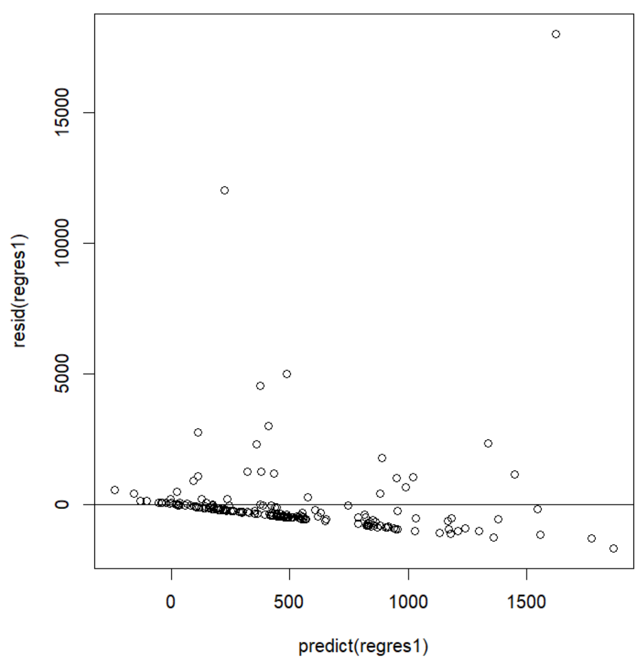
H_o: $\beta_1 = \beta_2 = 0$
H_a: At least one slope $\neq 0$
$\alpha = 0.05$

<u>Checks:</u>

- ☑ Normality: Normal Probability plot is not linear, especially on the ends where there lie extreme outliers. Extreme right skew. Proceed with caution.



☑ Equal Variance: Residual plot does not appear randomly distributed. Proceed with caution.



Calculations:

F = $\frac{R^2/k}{(1-R^2)/(n-(k+1))}$ w/df = $\frac{k}{n-(k+1)}$

F = $\frac{0.04729/2}{(1-0.04729)/(185)}$ w/df = $\frac{2}{185}$

F = 4.5915 w/df = $\frac{2}{185}$

p-value = p(F > 4.5915) = 0.01132

Since the p-value (0.01132) < α (0.05), we reject the null hypothesis. Therefore, there is convincing evidence to suggest a useful linear relationship that uses the rate of anxiety disorders and the rate of depression to predict GDP.

R Output:

```
Call:
lm(formula = `GDP(thousands)` ~ `Anxiety disorders (%)` + `Depression (%)`)

Residuals:
    Min      1Q  Median      3Q     Max 
-1664.6  -496.9  -289.6   -30.2 17989.4 

Coefficients:
                        Estimate Std. Error t value Pr(>|t|)  
(Intercept)              -1549.0      775.4  -1.998   0.0472 *
`Anxiety disorders (%)`    262.9      123.1   2.136   0.0340 *
`Depression (%)`           295.2      227.0   1.300   0.1951  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1799 on 185 degrees of freedom
Multiple R-squared:  0.04729,	Adjusted R-squared:  0.03699 
F-statistic: 4.592 on 2 and 185 DF,  p-value: 0.01132
```

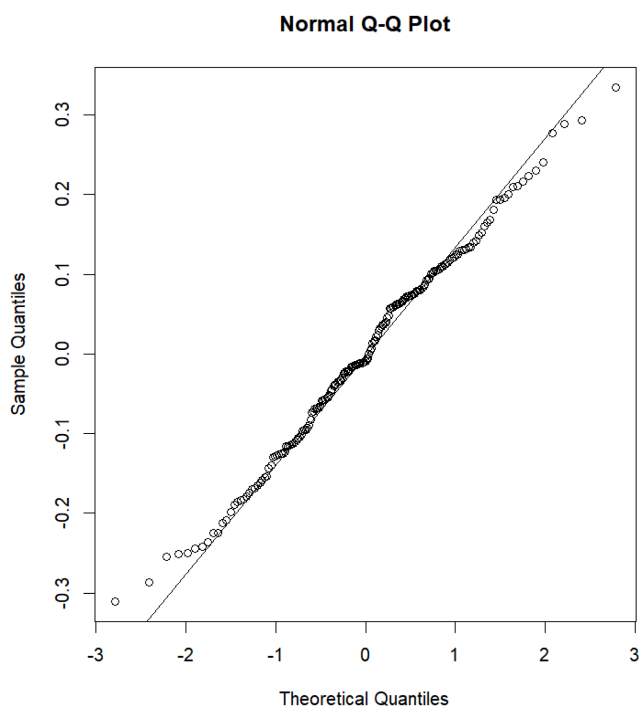
Model Utility F-test 2:

$\hat{y}$ = predicted HDI
x_1 = rate of anxiety disorders
x_2 = rate of depression
β_1 = estimated increase in predicted HDI when the rate of anxiety disorders increases by one unit.
β_2 = estimated increase in predicted HDI when the rate of depression increases by one unit.

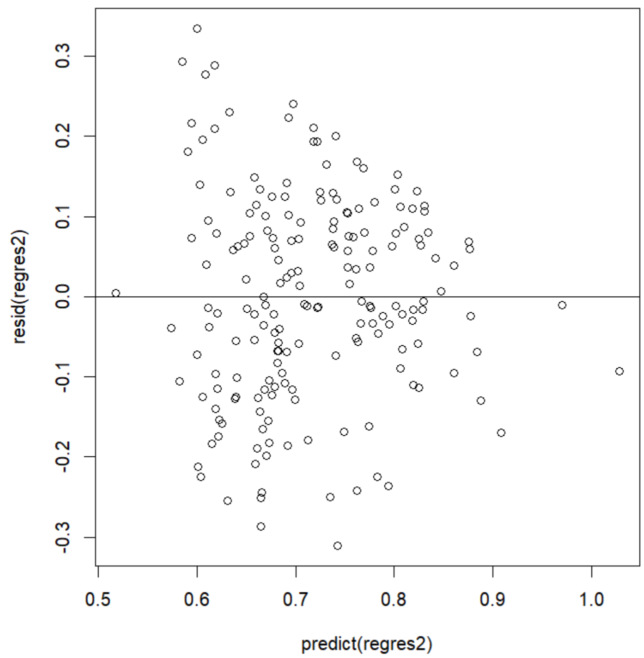
H_o: $\beta_1 = \beta_2 = 0$
H_a: At least one slope $\neq 0$
$\alpha = 0.05$

Checks:

- ☑ Normality: Normal Probability plot appears relatively linear.



- ☑ Equal Variance: Residual plot appears relatively randomly distributed.



Calculations:

$F = \frac{R^2/k}{(1-R^2)/(n-(k+1))}$ w/df $= \frac{k}{n-(k+1)}$

$F = \frac{0.2987/2}{(1-0.2987)/(184)}$ w/df $= \frac{2}{184}$

$F = 39.18$ w/df $= \frac{2}{184}$

p-value = p(F > 39.18) = $6.672\times10^{-15} \approx 0$

Since the p-value (≈ 0) < α (0.05), we reject the null hypothesis. Therefore, there is convincing evidence to suggest a useful linear relationship that uses the rate of anxiety disorders and the rate of depression to predict HDI.

R Output:

```
Call:
lm(formula = hdi_2017 ~ `Anxiety disorders (%)` + `Depression (%)`)

Residuals:
```

```
     Min       1Q   Median       3Q      Max 
-0.31027 -0.09517 -0.00994  0.08956  0.33388 

Coefficients:
                         Estimate Std. Error t value Pr(>|t|)    
(Intercept)              0.663181   0.055318  11.988  < 2e-16 ***
`Anxiety disorders (%)`  0.076848   0.008819   8.714 1.69e-15 ***
`Depression (%)`        -0.073320   0.016251  -4.512 1.14e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1286 on 184 degrees of freedom
Multiple R-squared:  0.2987,   Adjusted R-squared:  0.2911 
F-statistic: 39.18 on 2 and 184 DF,  p-value: 6.672e-15
```

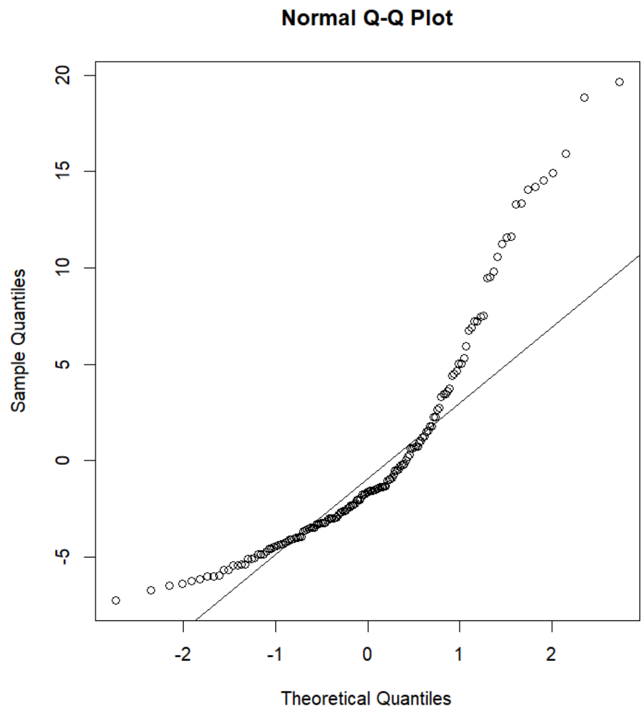
Model Utility F-test 3:

$\hat{y}$ = predicted unemployment (% of labor force)
x_1= rate of anxiety disorders
x_2= rate of depression
β_1= estimated increase in predicted unemployment % when the rate of anxiety disorders increases by one unit.
β_2= estimated increase in predicted unemployment % when the rate of depression increases by one unit.

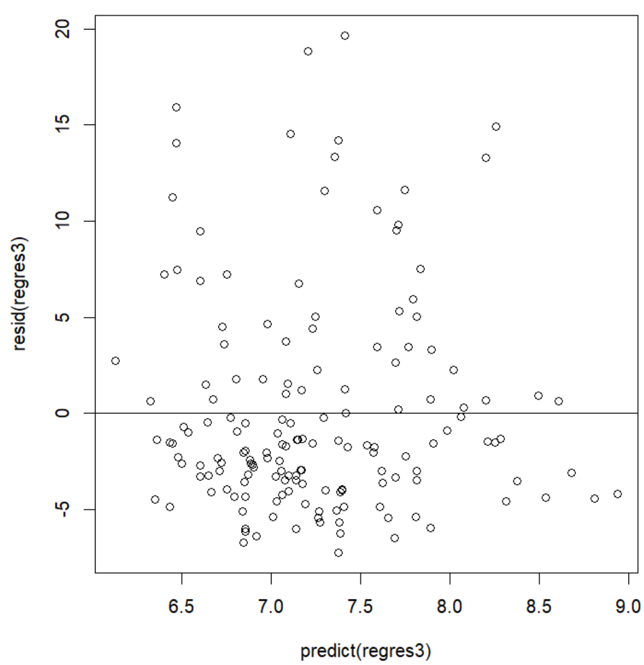
H_o: $\beta_1 = \beta_2 = 0$
H_a: At least one slope $\neq 0$
$\alpha = 0.05$

<u>Checks:</u>

- ☑ Normality: Normal Probability plot is not linear. Extreme right skew and very curved. Proceed with caution.



☑ Equal Variance: Residual plot does not appear randomly distributed. Proceed with caution.



Calculations:

$$F = \frac{R^2/k}{(1-R^2)/(n-(k+1))} \quad \text{w/df} = \frac{k}{n-(k+1)}$$

$$F = \frac{0.0102/2}{(1-0.0102)/(155)} \quad \text{w/df} = \frac{2}{155}$$

$$F = 0.7985 \quad \text{w/df} = \frac{2}{155}$$

p-value = p(F > 0.7985) = 0.4518

Since the p-value (0.4518) > α (0.05), we fail to reject the null hypothesis. Therefore, there is not convincing evidence to suggest a useful linear relationship that uses the rate of anxiety disorders and the rate of depression to predict unemployment (% of the labor force).

R Output:

```
Call:
lm(formula = `Unemployment (% of labor force)` ~ `Anxiety disorders (%)` +
    `Depression (%)`)

Residuals:
   Min     1Q Median     3Q    Max
-7.237 -3.612 -1.638  1.704 19.629

Coefficients:
                        Estimate Std. Error t value Pr(>|t|)
(Intercept)               4.1804     2.6761   1.562    0.120
`Anxiety disorders (%)`   0.3097     0.4092   0.757    0.450
`Depression (%)`          0.5324     0.7938   0.671    0.503

Residual standard error: 5.636 on 155 degrees of freedom
Multiple R-squared:  0.0102,   Adjusted R-squared:  -0.002573
F-statistic: 0.7985 on 2 and 155 DF,  p-value: 0.4518
```

The following 2-Sample T-test is meant to determine whether or not country Social Capital Index (SCI), or societal confidence, has any affect on country suicide rates. It will divide suicide rates into two categories (high and low) and attempt to find a difference between SCI means for those categories. The line for low vs. high suicide rates is based on the World Health Organization's global average, 9.0 per 100,000 population per year.

2-Sample T-test:

H_o: $\mu_1 - \mu_2 = 0$
H_a: $\mu_1 - \mu_2 \neq 0$
$\alpha = 0.05$

μ_1= Average Social Capital Index (SCI) score of countries with a high suicide rate (H)
μ_2= Average Social Capital Index (SCI) score of countries with a low suicide rate (L)

Checks:

- ☑ Random: Not randomly selected because the data was meant to be the entire population of countries (excluding data points with no data). Proceed with caution.
- ☑ Normality:
  - ☑ μ_1: CLT (N > 30, 68 > 30)
  - ☑ μ_2: CLT (N > 30, 102 > 30)
- ☑ Independence: Assume Independence.

Calculations:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{(49.08824 - 41.04020) - (0)}{\sqrt{\frac{10.329360^2}{68} + \frac{7.784858^2}{102}}} = 5.4719$$

p-value = p(t > 5.4719)x2 = $2.594 \times 10^{-7} \approx 0$

Since the p-value ($\approx$ 0) < α (0.05), we reject the null hypothesis. Therefore, there is convincing evidence that the difference between the average Social Capital Index Score of countries with high suicide rates and the average Social Capital Index Score of countries with low suicide rates is not equal to zero.

R Output:

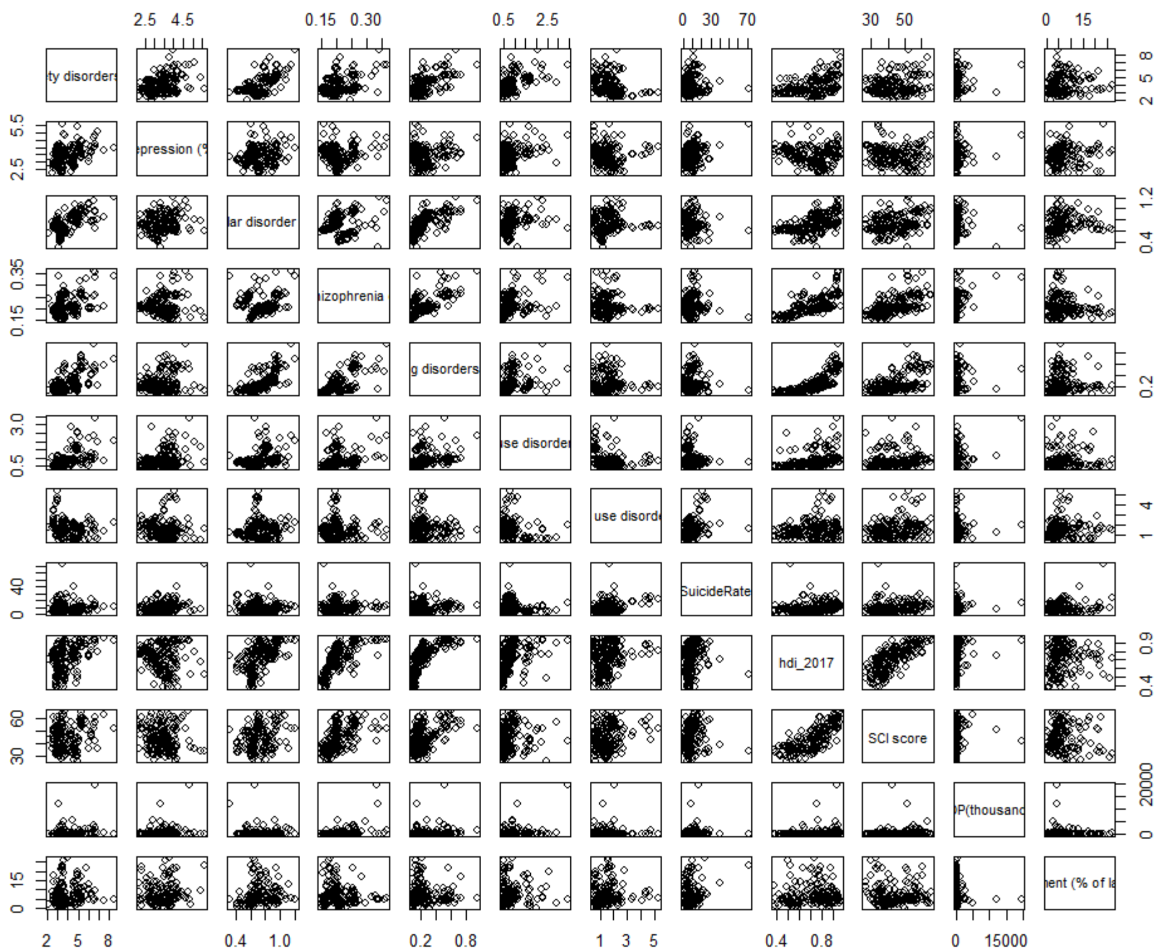
```
Welch Two Sample t-test

data:  SCI score by H vs. L
t = 5.4719, df = 116.29, p-value = 2.594e-07
alternative hypothesis: true difference in means between group H and group L is
not equal to 0
95 percent confidence interval:
```

```
  5.135038 10.961040
sample estimates:
mean in group H mean in group L
       49.08824        41.04020
```

Predictive Models:

Using the data and knowledge from the above tests, I will attempt to construct two predictive linear regression models, one that uses a country's social and economic factors to predict rates of mental illnesses and another that uses a country's mental illness rates to predict social and economic factors. I will start by creating a scatterplot matrix to get a better idea of correlations, and then go from there.



Model #1: Using socioeconomic factors to predict mental health.

Model	Equation	R^2	R^2 (Adj.)	S_e	F-Stat.	p-val.
$\widehat{(y)}$`Anxiety disorders (%)`=`hdi_2017`(x_1)+`GDP(thousands)`(x_2)+`SCI score`(x_3) +`Unemployment (% of labor force) (x_4)	$\hat{y} = 1.437 + 5.393x_1 + 6.414\times10^{-5}x_2 - 3.277\times10^{-2}x_3 + 9.262\times10^{-3}x_4$	0.3021	0.2824	1.021	15.37	$1.822 \times 10^{-10} \approx 0$
$\widehat{(y)}$`Anxiety disorders (%)`=`hdi_2017`(x_1)+`SCI score`(x_2)+`hdi_2017`*`SCI score`(x_3)	$\hat{y} = 10.49827 - 5.87564x_1 - 0.26295x_2 + 0.28210x_3$	0.3387	0.3269	0.96	28.68	$5.025 \times 10^{-15} \approx 0$
$\widehat{(y)}$`Schizophrenia (%)`=`hdi_2017`(x_1)+`GDP(thousands)`(x_2)+`SCI score`(x_3)+`Unemployment (% of labor force)`(x_4)	$\hat{y} = 8.598\times10^{-2} + 1.493\times10^{-1}x_1 + 6.962\times10^{-6}x_2 + 4.774\times10^{-4}x_3 - 1.020\times10^{-3}x_4$	0.591	0.5795	0.02806	51.3	$2.2\times 10^{-16} \approx 0$

I chose the model: $\widehat{(y)}$`Schizophrenia (%)`=`hdi_2017`(x_1)+`GDP(thousands)`(x_2)+`SCI score`(x_3)+`Unemployment (% of labor force)`(x_4), or $\hat{y} = 8.598\times10^{-2} + 1.493\times10^{-1}x_1 + 6.962\times10^{-6}x_2 + 4.774\times10^{-4}x_3 - 1.020\times10^{-3}x_4$, because it has the highest R^2 and adjusted R^2 values, the lowest standard error, the highest F-statistic, and the lowest p-value.

#2: Using mental health to predict socioeconomic factors.

Model	Equation	R^2	R^2 (Adj.)	S_e	F-Stat.	p-val.
$(\hat{y})$`hdi_2017`=`Anxiety disorders (%)`(x_1)+`Depression (%)`(x_2)+`Bipolar disorder (%)`(x_3)+ `Schizophrenia (%)`(x_4)+`Eating disorders (%)`(x_5)+`Drug use disorders (%)`(x_6)+ `Alcohol use disorders (%)`(x_7)+SuicideRate(x_8)	$\hat{y}$ = 0.4309 - 0.00348 x_1 - 0.0755x_2 + 0.1174x_3 + 1.0783x_4 + 0.4228x_5 + 0.0682 x_6 + 0.0352x_7 + 0.0028x_8	0.7667	0.7557	0.07625	69.84	$2.2\times 10^{-16} \approx 0$
$(\hat{y})$`SCI score`=`Anxiety disorders (%)`(x_1)+`Depression (%)`(x_2)+`Bipolar disorder (%)`(x_3)+ `Schizophrenia (%)`(x_4)+`Eating disorders (%)`(x_5)+`Drug use disorders (%)`(x_6)+ `Alcohol use disorders (%)`(x_7)	$\hat{y}$ = 28.0471 - 0.7399 x_1 - 2.5162x_2 + 0.9190x_3 + 83.9400x_4 + 25.6306x_5 - 1.5158 x_6 + 2.6158x_7	0.5066	0.4856	6.965	24.06	$2.2\times 10^{-16} \approx 0$

$(\hat{y})$`hdi_2017`=`Anxiety disorders (%)`(x_1)+`Depression (%)`(x_2)+`Bipolar disorder (%)`(x_3)+ `Schizophrenia (%)`(x_4)+`Eating disorders (%)`(x_5)+`Drug use disorders (%)`(x_6)+ `Alcohol use disorders (%)`(x_7)+SuicideRate(x_8)+ `Schizophrenia (%)`*`Eating disorders (%)`(x_9)	$\hat{y}$ = -0.0183 - 0.0124x_1 - 0.0452x_2 + 0.1248x_3 + 2.7412x_4 + 1.8063x_5 + 0.07396x_6 + 0.0329x_7 + 0.00229x_8 - 5.5561x_9	0.8469	0.8387	0.06196	103.8	$2.2\times 10^{-16} \approx 0$

I chose the model: $(\hat{y})$`hdi_2017`=`Anxiety disorders (%)`(x_1)+`Depression (%)`(x_2)+`Bipolar disorder (%)`(x_3)+ `Schizophrenia (%)`(x_4)+`Eating disorders (%)`(x_5)+`Drug use disorders (%)`(x_6)+ `Alcohol use disorders (%)`(x_7)+SuicideRate(x_8)+ `Schizophrenia (%)`*`Eating disorders (%)`(x_9), or $\hat{y}$ = -0.0183 - 0.0124x_1 - 0.0452x_2 + 0.1248x_3 + 2.7412x_4 + 1.8063x_5 + 0.07396x_6 + 0.0329x_7 + 0.00229x_8 - 5.5561x_9, because it has the highest R^2 and adjusted R^2 values, the lowest standard error, the highest F-statistic, and a low p-value.

Discussion, Analysis, and Evaluation

Summary of all test conclusions and predictive models:

ANOVA 1:
Null hypothesis - The mean rate of anxiety disorders stays the same across all HDI levels.
Alternate hypothesis - At least one mean rate of anxiety disorders is different across HDI levels.

Source	df	SS	MS	F	p-value
HDI level	3	63.0708	21.0236	21.4664	≈ 0
Error	183	179.225	0.97937		
Total	186	242.2958			

Conclusion: Since the p-value(≈ 0) > α(0.05), we reject the null hypothesis. Therefore, there is convincing evidence to suggest that at least one mean rate of anxiety disorders is different across the HDI levels.
Connection: I was using this test to determine whether or not there were statistically significant differences between the means of the rates of anxiety disorders for different levels of country HDI. My initial hypothesis was that countries with higher HDI would have higher levels of anxiety. Together, the boxplots I constructed in my checks and the results of the test support these suspicions.

ANOVA 2:
Null hypothesis - The mean rate of depression stays the same across all HDI levels.
Alternate hypothesis - At least one mean rate of depression is different across HDI levels.

Source	df	SS	MS	F	p-value
HDI level	3	8.7023	2.9008	8.4719	≈ 0
Error	183	62.6587	0.3424		
Total	186	71.361			

Conclusion: Since the p-value(≈ 0) > α(0.05), we reject the null hypothesis. Therefore, there is convincing evidence to suggest that at least one mean rate of depression is different across the HDI levels.

Connection: I was using this test to determine whether or not there were statistically significant differences between the means of the rates of depression for different levels of country HDI. My initial hypothesis was that countries with higher HDI would have higher levels of depression. The boxplots I constructed in my checks did not appear to support this, but the results of the test found a significant difference in some way.

Model Utility F-test 1:
Null hypothesis - There is no useful relationship between the rate of anxiety disorders, depression, and country GDP.
Alternate hypothesis - There exists a useful relationship between the rate of anxiety disorders, depression, and country GDP.
Conclusion: Since the p-value (0.01132) < α (0.05), we reject the null hypothesis. Therefore, there is convincing evidence to suggest a useful linear relationship that uses the rate of anxiety disorders and the rate of depression to predict GDP.
Connection: This F-test was meant to search for a correlation between anxiety, depression, and country GDP. The test suggested a useful relationship between the three variables.

Model Utility F-test 2:
Null hypothesis - There is no useful relationship between the rate of anxiety disorders, depression, and country HDI.
Alternate hypothesis - There exists a useful relationship between the rate of anxiety disorders, depression, and country HDI.
Conclusion: Since the p-value ($\approx$ 0) < α (0.05), we reject the null hypothesis. Therefore, there is convincing evidence to suggest a useful linear relationship that uses the rate of anxiety disorders and the rate of depression to predict HDI.
Connection: This F-test was meant to search for a correlation between anxiety, depression, and country HDI. The test suggested a useful relationship between the three variables.

Model Utility F-test 3:
Null hypothesis - There is no useful relationship between the rate of anxiety disorders, depression, and country unemployment.
Alternate hypothesis - There exists a useful relationship between the rate of anxiety disorders, depression, and country unemployment.
Conclusion: Since the p-value (0.4518) > α (0.05), we fail to reject the null hypothesis. Therefore, there is not convincing evidence to suggest a useful linear relationship that uses the rate of anxiety disorders and the rate of depression to predict unemployment (% of the labor force).
Connection: This F-test was meant to search for a correlation between anxiety, depression, and unemployment. The test did not suggest a useful relationship between the three variables.

2-Sample T-test:
Null hypothesis - The means of Social Capital Index ratings between countries with high vs. low rates of suicide are equal.
Alternate hypothesis - The means of Social Capital Index ratings between countries with high vs. low rates of suicide are not equal.
Conclusion: Since the p-value (≈ 0) < α (0.05), we reject the null hypothesis. Therefore, there is convincing evidence that the difference between the average Social Capital Index Score of countries with high suicide rates and the average Social Capital Index Score of countries with low suicide rates is not equal to zero.
Connection: This test was an attempt to find a difference between the SCI means for two categories (low vs. high suicide rates). The results showed a significant difference.

Model #1: Using socioeconomic factors to predict mental health:
Model: $(\hat{y})$`Schizophrenia (%)`=`hdi_2017`(x_1)+`GDP(thousands)`(x_2)+`SCI score`(x_3)+`Unemployment (% of labor force)`(x_4), or $\hat{y} = 8.598\times10^{-2} + 1.493\times10^{-1}x_1 + 6.962\times10^{-6}x_2 + 4.774\times10^{-4}x_3 - 1.020\times10^{-3}x_4$, because it has the highest R^2 and adjusted R^2 values, the lowest standard error, the highest F-statistic, and the lowest p-value.

R^2	R^2(Adj.)	S_e	F-Stat.	p-val.
0.591	0.5795	0.02806	51.3	$2.2\times10^{-16} \approx 0$

Connection: I constructed this equation in the hopes of creating a model that could use a country's social and economic factors to predict a rate of mental illnesses. My best model had an R^2 of 0.591 and an adjusted R^2 of 0.5795, so it's not that good at predicting schizophrenia. However, its p-value ≈ 0, so it's statistically significant.

#2: Using mental health to predict socioeconomic factors:
Model: $(\hat{y})$`hdi_2017`=`Anxiety disorders (%)`(x_1)+`Depression (%)`(x_2)+`Bipolar disorder (%)`(x_3)+ `Schizophrenia (%)`(x_4)+`Eating disorders (%)`(x_5)+`Drug use disorders (%)`(x_6)+ `Alcohol use disorders (%)`(x_7)+SuicideRate(x_8)+ `Schizophrenia (%)`*`Eating disorders (%)`(x_9), or $\hat{y} = -0.0183 - 0.0124x_1 - 0.0452x_2 + 0.1248x_3 + 2.7412x_4 + 1.8063x_5 + 0.07396x_6 + 0.0329x_7 + 0.00229x_8 - 5.5561x_9$, because it has the highest R^2 and adjusted R^2 values, the lowest standard error, the highest F-statistic, and a low p-value.

R^2	R^2(Adj.)	S_e	F-Stat.	p-val.
0.8469	0.8387	0.06196	103.8	$2.2\times10^{-16} \approx 0$

Connection: I constructed this equation in the hopes of creating a model that could use a country's mental illness rates to predict a social and economic factor. My best model had an R^2 of 0.8469 and an adjusted R^2 of 0.8387, so it's very good at predicting country HDI. Also, its p-value $\approx$ 0, so it's statistically significant.

Implications:
My tests and predictive models suggest a significant relationship between mental health and socioeconomic factors, most specifically between anxiety, depression, GDP, HDI, SCI, and suicide rates. Only one of my tests did not reject the null hypothesis, and I was able to produce two models of varying strength.

Limitations:
I could only use data from 2017 because it was the earliest year I could find rates of mental illness. Since my results are based on data from seven years ago, they may be less applicable today. Also, some countries didn't have data for certain variables, or for any variables, so I had to remove them. Lastly, I wish I had the chance to do some post hoc tests for my ANOVAs to identify the specific groups that differ from each other, and for more socioeconomic factors to compare rates of mental illness to.

The Gap:
By demonstrating that there is a relationship between mental health and socioeconomic factors, my research has demonstrated that more should be done to research the connection between the two. By pinpointing specific socioeconomic factors in countries and finding ways to manipulate them, perhaps the rates of some mental illnesses will decline.

Conclusion and Future Directions

My tests found significant relationships between a country's anxiety, depression, GDP, and HDI, and between a country's SCI and suicide rates. I found nothing significant using unemployment. My models were also significant, especially the one that used rates of mental illnesses to predict country HDI. These conclusions may prove useful in the gap I'm working within, especially if further research is conducted.

I think that my process for collecting data was sound; I didn't need to take a random sample because my data was meant to be the entire population of countries. Unfortunately, I did end up having to remove countries that didn't have enough data. I was also forced to collect data only from the year 2017 because it was the closest year to present day for which extensive mental health data was available. I also failed some of my test conditions. This may make my results different from what is true today, however they may still be relevant. Because of the functionality of my models and the significance of the majority of my tests, I believe that further research on other years (which is what I would have done had I had more time and money) may reveal a solid pattern. This pattern, if tested and found, could be applied to lessening rates of mental illness in a country and improving its socioeconomic health.

First, I would find more mental health data for 2018 up to present day, and then find the corresponding GDP, HDI, SCI, unemployment, and suicide rates. I may also add more socioeconomic factors like educational attainment, financial/food/housing security, and average income. Using this new data, I would run more tests on the data from 2017 and also run tests on the individual years from 2018 to present day. I would expect there to be a lot of discrepancies around the years of the pandemic, and for this to severely affect the overall pattern. Therefore, I would be wary of these results and maybe classify them separately, or use them to demonstrate the ways in which the pandemic affected both the mental health and socioeconomic factors of a country (there will obviously be a lot of confounding variables in this case, so you'd have to be careful in drawing conclusions). Nevertheless, by testing throughout the years, perhaps it will be even easier to create useful models for predicting mental illness rates using socioeconomic factors and vice versa. Maybe it will even be possible to pinpoint a single or several outstanding predictive factors. If these are manipulatable in real life, changing the social and mental landscape of a country could very well be within reach.

Bibliography

Works Cited

Alegria, Margarita. "Social Determinants of Mental Health: Where We Are and Where We Need to Go." *NCBI*, 17 September 2018, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6181118/. Accessed 7 June 2024.

Cabello-Rangel, Héctor, et al. "Suicide Rate, Depression and the Human Development Index: An Ecological Study From Mexico." *NCBI*, 17 November 2020, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7706657/. Accessed 7 June 2024.

"Global Trends in Mental Health Disorder." *Kaggle*, Kaggle, 2023, https://www.kaggle.com/datasets/thedevastator/uncover-global-trends-in-mental-health-disorder. Accessed 7 June 2024.

Human Development Reports. "Documentation and downloads | Human Development Reports." *Human Development Reports*, United Nations Development Programme, 2024, https://hdr.undp.org/data-center/documentation-and-downloads. Accessed 7 June 2024.

International Monetary Fund. "World Economic Outlook (April 2024) - Real GDP growth." *World Economic Outlook (April 2024) - Real GDP growth*, International Monetary Fund, 2021, https://www.imf.org/external/datamapper/NGDPD@WEO/OEMDC/ADVEC/WEOWORLD. Accessed 7 June 2024.

Kirkbride, James B., et al. "The social determinants of mental health and disorder: evidence, prevention and recommendations." *NCBI*, 12 January 2024, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10786006/. Accessed 7 June 2024.

"Life expectancy & Socio-Economic (world bank)." *Kaggle*, Kaggle, September 2023, https://www.kaggle.com/datasets/mjshri23/life-expectancy-and-socio-economic-world-bank. Accessed 7 June 2024.

SOLABILITY. "Global Social Capital Index by country." *SOLABILITY*, https://solability.com/the-global-sustainable-competitiveness-index/the-index/social-capital. Accessed 7 June 2024.

World Health Organization. "Suicide worldwide in 2019: Global Health Estimates." *WHO*, WHO, 2019, https://iris.who.int/bitstream/handle/10665/341728/9789240026643-eng.pdf?sequence=1. Accessed 7 June 2024.

World Population Review. "Suicide Rate by Country 2024." *World Population Review*, World Population Review, 2024, https://worldpopulationreview.com/country-rankings/suicide-rate-by-country. Accessed 7 June 2024.