LUT School of Business and Management

Analysis of Suicide Rates and Its Factors Using Multivariate Econometric Models

Quantitative Data Analysis

Nguyen Tran

11.3.2022

Table of Contents

1 INTRODUCTION

1.1 Purpose and Structure of the Study

The purpose of this study is to analyze the suicide rates of various countries to determine what the social, political, economic, and health factors are of the rate. This study aims to fill the research gaps regarding the relationship between suicide rate and various factors from a more global standpoint with data from over 200 countries being examined. This study will identify influential factors of suicide rate in hopes to provide focus points for reducing the rates and promote future research on these factors that may lead to the suggestion of methods for improvement.

The structure of this study is as follows. The introduction will summarize the main purpose of this study as well as the main concepts that lead to the formulation of research questions. Then, section 2 will explain the hypotheses developed from the research questions, as well as the variables used in this study by providing descriptive analysis results and visuals, and it will explain the data collection process. Section 3 will provide insight into the measure development through factor analysis, and reliability and validity analysis. Section 4 will discuss the explanatory analysis, covering the results of multiple regression and ANOVA analyses. Section 5 will include a penal data analysis and Section 6 will provide a summary conclusion discussing results, validity evaluation, limitations, and notes for future research.

1.2 Main Concepts and Research Questions

About 800,000 people die from suicide each year globally, contributing to 1.3% of death around the world in 2019 (Lee, Roser and Ortiz-Ospina, 2019). There are various factors that may influence the suicide rate of a country. Multiple studies exist discussing isolated economic factors, such as unemployment and gross domestic product, that influence suicide rate in specific countries (Fountoulakis et al., (2014); Ceccherini-Nelli and Priebe, (2010)). Influential social and political factors have been investigated as well (Sedgwick et al., (2019); Girard, (1993)). Another common factor explored in existing literature is health, investigating the relationship between various health indicators and the suicide rate of specific countries (Cooper Smith, (2018); Landberg, (2009)).

There have been various studies conducted that examine isolated factors of suicide for specific countries, however, there seems to be a research gap that examines the relationship between suicide rate and multiple factors on a global scale (Fountoulakis et al., (2014); Emamgholipour, Arab and Shirani, (2021)). The limited research existing about the influential factors of suicide rates on a global scale and how the factors may relate has led to the following research questions:

Q1: What are the dimensions affecting Suicide rate and how they are related?

Q2: What are the effects of social, political, economic and health factors on Suicide Rate?

2 HYPOTHESIS DEVELOPMENT, MODEL STATEMENT AND DATA COLLECTION

2.1 Model and Hypotheses

Based on the existing literature, we have created a model with multiple variables categorized within the factors identified in Section 1.2 to test how they impact suicide rates in countries around the world. The model presented in Figure 1. shows the factor categories, which are classified as independent variables, and identifies the dependent variable, suicide rate. Independent variables are indicated with a box and arrow, while the dependent variable is indicated with a circle in the model.

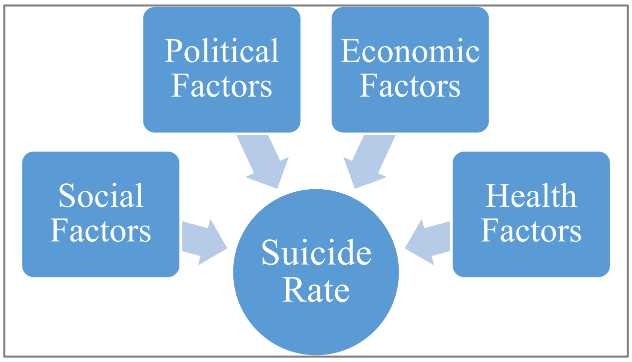


Figure 1. Model of Study

Based on the literature discussed in Section 1.2, it is assumed that political, social, economic, and health factors are influential on countries' suicide rates. The hypotheses that will be tested to validate this assumption are:

H1: Countries with better Political factors (low corruption index, high human right protection index) will have lower suicide rates.

H2: Countries with higher social factors (balance sex ratio at birth, low inequality disparity) will have lower suicide rates.

H3: Countries with higher Economic factors (high GDP per capita, low unemployment rate) are more likely to have lower suicide rates.

H4: Countries with better indications of Health factors are more likely to have lower suicide rates.

2.2 Data Collection & Descriptive Analysis

To understand the factors that influence suicide rate, each factor was broken into multiple variables. To reiterate the factor categories, the independent variables are categorized by factor type: Political, Social, Economic, and Health. Table 1. shows the specific variables of data within each factor category, as well as the variable code used when analyzing the data in the Stata software. Variables were chosen based on existing literature, availability of data in databases, and subjective personal judgement of relevance by the researchers.

Data on each variable was collected from multiple different sources. World Bank Database was used to collect data on the dependent variable, suicide rate (% suicide rate over 100,000 population), as well as independent variables, sex ration at birth (male births per female births), internet use (% of population using the internet), annual healthcare expenditure per capita (in dollars), GDP, PPP per capita (on dollars), extreme poverty (% share population in extreme poverty $1.90 per day), inflation, GDP deflator, GPD growth rate, unemployment rate (%), and labor force participation rate (% of total population ages 15-64). The Our World Database was also used to compile data on anxiety disorders (% share of population with anxiety disorders), depression (% of population with depression), mental health (% of population with mental health disorder), alcohol or drug use disorders (% of population with alcohol or drug use disorder), human right protection (human right protection index scale -4 to 6, with 0 being neutral and higher score meaning less abuse), civil liberties (scale 0-1 captures the extent to which citizens enjoy physical integrity rights, religious freedom, law transparency, etc.), and the human development index (measure of key dimensions of human development: a long and healthy life, a good education, and having a decent standard of living). In addition to these continuous variables, there was also data collected on categorical variables from the World Bank such as area code and income group to further categorize the data for potential insight. It is initially assumed that all independent variables influence suicide rate, however, a factor analysis will be discussed later to eliminate some variables to a reduced number of variables for a more accurate analysis.

Table 1. Factor Categorization of Variables and Variable Codes

Factor	Variable	Variable Code
Dependent Variables		
Suicide Rate	Suicide Rate	var0_suiciderate
Independent Variables		
Political	Human right protection	var1_humanrightprotection
	Civil liberties	var2_civilliberties
	Corruption Perception Index	var3_corruptionperceptionindex
Social	Human Development Index	var4_humandevelopmentindex
	Sex Ratio at birth	var5_sexratioatbirth
	Internet Usage	var6_internetusage

	Life Satisfaction	var7_lifesatisfaction
	Extreme Poverty	var8_extremepoverty
	Gini Index	var9_giniindex
	Share of one person household	var10_shareof1personhousehold
	Crude Marriage Rate (per 1000)	var11_crudemarriagerate
	Crude Divorce Rate (per 1000)	var12_crudedivorcerate
Economic	GDP per capita (PPP)	var13_gdppercapitappp
	Annual GDP growth rate	var14_gdpgrowthrate
	Inflation	var15_inflation
	Unemployment rate	var16_unemploymentrate
	Labor Force Participation rate	var17_laborparticipationrate
	Annual working hours per worker	var18_annualworkinghours
	Annual healthcare exp. per capita	var19_annualhealthexpenditure
Health	Anxiety disorder	var20_anxietydisorder
	Depression	var21_depression
	Mental Health	var22_mentalhealth
	Alcohol and Drug usage	var23_alcoholanddrug

To test the hypotheses identified in section 2.1, data was collected on 203 countries from a ten-year period spanning from the year 2007 to 2017. When comparing the number of participating countries overall for each year it was found that the results for each year were very similar. This means that there was not a significant difference in the amount of data presented for each year. Because the amount of data was similar for each year it was decided to use the year 2017 for further analysis as it is the most recent year, which should provide the most relevant data to today. Table 2. provides a summary of the descriptive statistics for each variable in the year 2017. An important variable to note is the dependent variable, suicide rate, which had a mean rate of 10.50% in 2017. However, the suicide rate does have a large range with a max of 56.6 and min of 2.26 and standard deviation of 7.07, indicating the data is quite volatile.

Table 2. Descriptive Statistics 2017- All Variables

Variable	Obs	Mean	Std. dev.	Min	Max
Suicide Rate	203	10.49717	7.066817	2.26359	56.60839
Human Right Protection	191	0.923155	1.685748	-2.30253	5.311889
Civil Liberties	172	0.679878	0.272369	0.006	0.992
Corruption Perception Index	177	42.90395	19.03904	9	89
Human Development Index	186	0.707059	0.152377	0.354	0.953
Sex Ratio at Birth	162	1.051574	0.01869	1.011	1.13
Internet Usage	189	53.31997	28.41022	1.308907	98.37
Life Satisfaction	144	5.462882	1.148514	2.662	7.788
Extreme Poverty	164	13.24049	20.38566	0	78.2814
Gini Index	146	38.1312	7.79004	24.21931	59.06661
Share of 1-Person Household	34	30.60965	8.416282	9.6	47
Crude Marriage Rate	46	5.206522	1.239068	3.1	7.9
Crude Divorce Rate	41	2.002439	0.59728	0.7	3.4

GDP Per Capita, PPP	165	20620.54	21271.49	773.5729	114985.8
GDP Growth Rate	172	3.410103	3.799996	-6.98758	26.6809
Inflation	172	4.943214	7.017771	-3.20497	60.98697
Unemployment Rate	158	7.278892	5.535805	0.14	26.06
Labor Force Participation Rate	158	68.27797	10.69931	41.7	88.35
Annual Working Hours	64	1850.389	256.7516	1353.887	2455.551
Annual Health Expenditure	188	1445.302	1744.342	44.05	10103.09
Anxiety Disorder	203	4.389515	1.204871	2.128506	8.778921
Depression	201	3.902114	0.910877	1.67966	6.700153
Mental Health	201	13.28705	2.042691	9.547788	19.1111
Alcohol and Drug Use Disorder	203	2.19346	1.131094	0.697173	5.946469

Appendix 4.1 shows a correlation matrix displaying the correlation between all variables. The matrix is color coded to show the spectrum of the strength of the correlation between each variable. Green indicates a positive correlation, red indicates a negative correlation, and white indicates a correlation close to zero or no correlation. The darker the color, the stronger the correlation. Looking at the correlation between the dependent variable and independent variables, we can see that no independent variables are strongly correlated with the dependent variable. The top three variables with the strongest correlation with suicide rate are crude divorce rate (0.4862), mental health (-0.4563), and anxiety disorder (-0.4556).

To further understand the trend of suicide rate across the world, the 203 countries were grouped by area code. The areas are South Asia (SA), Europe and Central Asia (EURCA), Latin America (LA), Sub-Saharan Africa (SSA), North America (NA), East Asia and Pacific (EAP), and Middle East and North Africa (MENA). Figure 2. shows the average suicide rate per area code. It shows East Asia and Pacific (13.22%), Sub-Saharan Africa (13.13%), and Europe and Central Asia (13.06%) have significantly higher average suicide rates compared to the other areas. However, it is worth noting that these areas also have a significantly higher number of countries located in the area. Latin America (8.22%), North America (8.22%), and South Asia (8.93%) have similar moderate average rates compared to the other areas. Meanwhile, the Middel East and North Africa have the lowest average suicide rate (5.29%) compared to all other areas, and it also had the lowest number of countries located in the area at 4 countries.

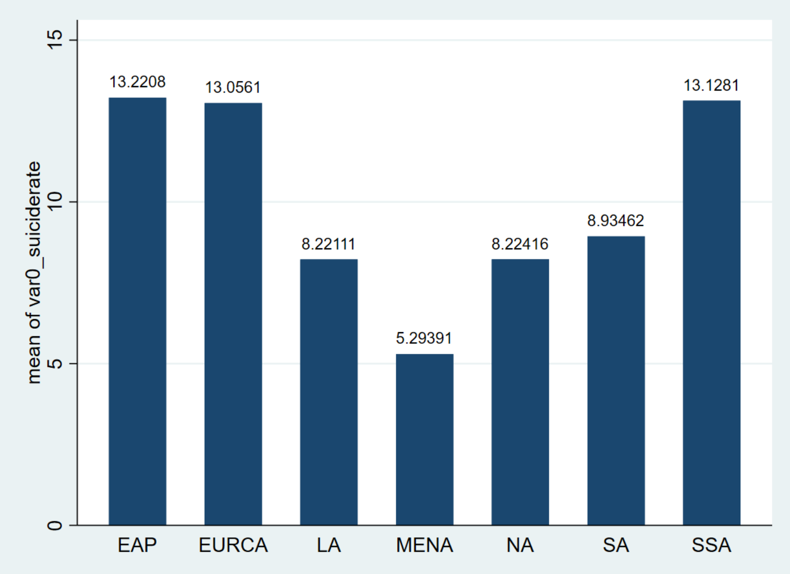


Figure 2. Average Suicide Rate per Area Code

The countries were also grouped by income: high (H), upper-middle (UM), lower-middle (LM), and low (L). Figure 3. shows the average suicide rate for each of these income groups. The results for each group were similar with each average rate being around 10-12%. There was no particular group that stood out as highly significant when compared to the rest. However, it is worth noting that the lower-middle income group had the highest average at 12.07% and the upper-middle income group had the lowest at 10.08%.

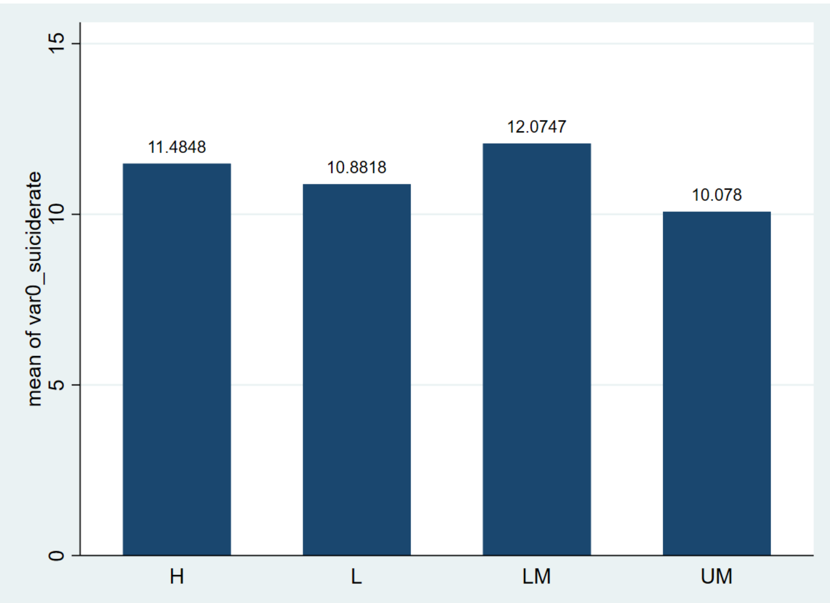


Figure 3. Average Suicide Rate per Income Group

3 MEASURE DEVELOPMENT

In this section, the measures required for later analysis will be developed with factor analysis. The aim of factor analysis is to reduce large number of variables into fewer significant number of factors or also known as latent variables. The goal of factor analysis is to simplify the complexity while maintaining an acceptable trade-off in accuracy.

The extraction method for the factor analysis is principle-component analysis (PCA) for 23 independent variables considering the factors with the minimum eigen value of 1. The year of 2017 is selected with most available data across variables. Due the large number of independent variables and low number of observations, the analysis was divided into two parts with 12 independent variables (var 1 to var12) in Part 1 and 11 independent variables (var13 to var23) in Part 2. The approach to factor analysis is iterative process which we will aim to improve the statistical power of the factors by removing variables with low statistical explanation power.

3.1 Factor analysis 1

In factor analysis 1, the Political variable group (variable 1 to 3) and the Social variable group (variable 4 to 12) with total observations of 1652 will be considered. According to Table 3, there are two factors with Eigen value larger than 1, which has a cumulative variance explanation of over 81%.

Table 3. Factor Analysis 1 (Initial)

```
Factor analysis/correlation                        Number of obs    =        29
    Method: principal factors                      Retained factors =         2
    Rotation: (unrotated)                          Number of params =        23
```

Factor	Eigenvalue	Difference	Proportion	Cumulative
Factor1	6.29505	4.87056	0.6616	0.6616
Factor2	1.42449	0.43537	0.1497	0.8113
Factor3	0.98912	0.52446	0.1040	0.9153
Factor4	0.46465	0.14379	0.0488	0.9641
Factor5	0.32086	0.10429	0.0337	0.9979
Factor6	0.21658	0.10877	0.0228	1.0206
Factor7	0.10780	0.09333	0.0113	1.0320
Factor8	0.01447	0.04499	0.0015	1.0335
Factor9	-0.03052	0.03195	-0.0032	1.0303
Factor10	-0.06247	0.02562	-0.0066	1.0237
Factor11	-0.08809	0.04930	-0.0093	1.0144
Factor12	-0.13739	.	-0.0144	1.0000

LR test: independent vs. saturated: chi2(66) = 311.26 Prob>chi2 = 0.0000

Then the rotated factors were loaded using the orthogonal rotation Varimax and removed low correlations between the variables and factors lower than 0.5 as presented in Table 4 below. The Kaiser-Meyer-Olkin measure was also estimated to calculate the sampling adequacy.

Table 4. Rotated Factor Loading – Factor Analysis 1 (Initial)

Variables	Factor 1.1 Social	Factor 1.2 Political	Uniqueness	KMO
var1_humanrightprotection		0.7450	0.2182	0.7667
var2_civilliberties		0.6639	0.4125	0.6659
var3_corruptionperceptionindex	0.9241		0.0476	0.8606
var4_humandevelopmentindex	0.8657		0.1029	0.7444
var5_sexratioatbirth	-0.6123		0.6012	0.6255
var6_internetusage	0.8037		0.2408	0.8178
var7_lifesatisfaction	0.9101		0.1308	0.8041
var8_extremepoverty	-0.6236		0.5860	0.6111
var9_giniindex		-0.7175	0.3863	0.7219
var10_shareof1personhousehold	0.7079		0.3919	0.8070
var11_crudemarriagerate		-0.6288	0.6005	0.6387
var12_crudedivorcerate	0.6140		0.5617	0.4378
Overall				**0.7356**

With two factors, the degree of informativity can capture 81% of the information and variation from the original variables. By observing the Table X, the Factor 1.1 was identified as Social Factor given their correlation in different social aspect of a society such as share of 1 person household in the society or the sex ratio at birth of a society while Factor 1.2 was identified as Political Factor given its high correlation in human right protection index and GINI index measuring the degree of inequality in a distribution of wealth (which is the main function of a government in author opinion). The overall KMO score is 0.74 presenting a middling result in sampling adequacy (Kaiser, 1974).

However, we were not happy with the results from the initial factor analysis 1 as many variables with middling KMO score in range of 0.6-0.69 so we decided to improve the result by removing variable 10, 11 and 12 due to low number of observations and KMO score. The result of final factor analysis 1 was captured at the Table 5 below.

Table 5. Rotated Factor Loading – Factor Analysis 1 (Final)

Variables	Factor 1.1 Social	Factor 1.2 Equality	Uniqueness	KMO
var1_humanrightprotection	0.8284		0.2730	0.9103
var2_civilliberties	0.7914		03550	0.8458
var3_corruptionperceptionindex	0.9083		0.1412	0.8079
var4_humandevelopmentindex	0.6979	0.6547	0.0843	0.8152
var5_sexratioatbirth		0.7168	0.4848	0.7851
var6_internetusage	0.7020	0.6426	0.0943	0.8778
var7_lifesatisfaction	0.6821		0.3492	0.9293
var8_extremepoverty		-0.7809	0.2592	0.7999
var9_giniindex		-0.5985	0.6048	0.7628
Overall KMO				0.8431
Eigen Value	5.32463	1.02957		

Cumulative proportion	0.8274	0.9874		
(Blanks represent abs(loading) < 0.5)				

We still have two factors with Eigen value larger than 1, which captured up to 98% of the variation from the variables (better than the initial result) and the KMO score was increased to 0.8431 with all KMO score for 9 variables over 0.7. The final Factor 1.1 was identified as Social & Political factor considering its high correlations in human right protection index and corruption perception index while the final Factor 1.2 was identified as Equality factor considering its high correlations in sex ratio at birth and number of extreme poverties in a society.

3.2 Factor analysis 2

In the second factor analysis, we will evaluate 11 variables for Economic and Health variables group and the result was captured in the Table 6 below. At first glance, there are two important factors with a cumulative proportion of variance over 89%. The overall KMO score was high with over 0.7795, indicating a good enough sampling adequacy and most of individual variables had high KMO score as well.

Table 6. Rotated Factor Loading – Factor Analysis 2

Variables	**Factor 2.1 Economic**	**Factor 2.2 Health**	**Uniqueness**	**KMO**
var13_gdppercapitappp	0.8479		0.2319	0.8274
var14_gdpgrowthrate			0.8101	0.7605
var15_inflation			0.8457	0.8040
var16_unemploymentrate		0.6495	0.5720	0.6601
var17_laborparticipationrate	0.6238		0.5994	0.8191
var18_annualworkinghours	-0.6972		0.4297	0.8793
var19_annualhealthexpenditure	0.9040		0.1216	0.8069
var20_anxietydisorder	0.5071	0.6189	0.3598	0.7034
var21_depression		0.7435	0.4166	0.6398
var22_mentalhealth		0.7971	0.1210	0.7123
var23_alcoholanddrug		0.6699	0.4401	0.9370
Overall KMO				0.7795
Eigen Value	3.65969	2.39246		
Cumulative proportion	0.5390	0.8914		
(Blanks represent abs(loading) < 0.5)				

Based on the factor analysis 2, the Factor 2.1 was identified as Economic factor due to its high correlation in GDP per capital (PPP) and annual health expenditure spending while Factor 2.2 was identified as Health factor due to its high correlations in Depression score and Mental Health Score.

It is important to note that variable 14, 15 and 16 for inflation, unemployment rates were not able to form a factor, so they are removed from the factor onward.

In overall, two factor analyses implemented returned the similar results with our variable grouping for Political & Social, Equality, Economic and Health factors that can influence the suicide rate.

3.3 Reliability and validity analysis

From the previous section, there are four factors that were identified for later analyses as presented in table 7. In this section, the reliability and validity of these factors will be evaluated to ensure that these factors are reliable to perform statistical analysis and avoid any non-reliable variables. The method for reliability analysis of these factors is Cronbach's alpha with a threshold of 0.6.

Table 7. Summary of variables in Factors for reliability analysis

Variables	Factor 1.1 Social & Political	Factor 1.2 Equality	Factor 2.1 Economic	Factor 2.2 Health
var1_humanrightprotection	0.8284			
var2_civilliberties	0.7914			
var3_corruptionperceptionindex	0.9083			
var4_humandevelopmentindex	0.6979			
var6_internetusage	0.7020			
var7_lifesatisfaction	0.6821			
var5_sexratioatbirth		0.7168		
var8_extremepoverty		-0.7809		
var9_giniindex		-0.5985		
var13_gdppercapitappp			0.8479	
var17_laborparticipationrate			0.6238	
var18_annualworkinghours			-0.6972	
var19_annualhealthexpenditure			0.9040	
var20_anxietydisorder				0.6189
var21_depression				0.7435
var22_mentalhealth				0.7971
var23_alcoholanddrug				0.6699

For Factor 1.1 Political, there is no negative correlation among variables so there is no need for transformation before taking the analysis. The first result for reliability analysis was low with the Cronbach's alpha below 0.6 so we decided to remove the variable 2 and 4 due to having higher alpha than the test scale. The second round of analysis has the result of 0.61 so we decided to keep the variables and generate a new variable "Factor 1 - Social and Political".

Table 8. Summary of the Cronbach's alpha for Factor 1.1 – Social & Political

Variables	Item-test correlation	Item-rest correlation	Alpha (round 1)	Alpha (round 2)
var1_humanrightprotection	0.6205	0.5996	0.5470	0.6578

var2_civilliberties	0.6343	0.6269	0.5737	
var3_corruptionperceptionindex	0.8831	0.7104	0.1483	0.1573
var4_humandevelopmentindex	0.8309	0.7964	0.5750	
var6_internetusage	0.7078	0.6345	0.2521	0.2714
var7_lifesatisfaction	0.7845	0.7690	0.5468	0.6459
Test scale			0.5536	0.6123

For Factor 1.2 – Equality, the Cronbach's alpha for these variables belong to this factor is surprisingly low with only 0.3, indicating that they are not internally consistent, and the reliability is low. Therefore, the Factor 1.2 – Equality will be removed from further analysis.

For Factor 2.1 – Economic, we evaluated the variables such as GDP per capita (PPP), labor participation rate, annual working hours of employees and the annual health expenditure on the working labors. The Cronbach's alpha of this factor is also low with only 0.20 despite we tried to remove any variable with higher alpha than the test scale. Therefore, the Factor 2.1 – Economic will be removed from further analysis.

For Factor 2.2 – Health, we examined the reliability of the variables for the proportion of population having anxiety disorder, depression, mental health, and alcohol & drug addiction. The test scale Cronbach's alpha is high with 0.69 as presented in Table 9, over the threshold of 0.6 so we decided to create a new variable from this factor "Factor 2 - Health"

Table 9. Summary of the Cronbach's alpha for Factor 2.2 – Health

Variables	**Item-test correlation**	**Item-rest correlation**	**Alpha (round 1)**
var20_anxietydisorder	0.8371	0.7043	0.4907
var21_depression	0.5068	0.3085	0.7153
var22_mentalhealth	0.9581	0.8430	0.3254
var23_alcoholanddrug	0.4714	0.2197	0.7438
Test scale			0.6918

In summary, only two reliable factors (Social & Political, Health) were identified as reliable for further analysis as additional independent variables despite four factors that were identified earlier based on Factor analysis.

4 EXPLANATORY ANALYSIS

In this section of the study, the result of explanatory analysis will be discussed, including the multiple linear regression analysis, binary logistic regression, and the analysis of variance (ANOVA).

4.1 Multiple linear regression analysis

In this multiple linear regression analysis, the relationship between suicide rate and other independent variables from previous section will be examined. The independent variables in this regression analysis will include two factors (Social & Political, Health) from previous factor analysis, plus other independent variables that are not included in these factors as presented in table 10.

Table 10. Descriptive summary of variables considered in the Initial regression model

Variable	Obs	Mean	Std. Dev.	Min	Max
var0 suiciderate	203	10.497	7.067	2.264	56.608
Social Political	198	27.45	14.745	1.858	98.37
Health	203	5.916	1.041	2.272	9.069
var2 civilliberties	172	.68	.272	.006	.992
var4 humandevelopm~x	186	.707	.152	.354	.953
var5 sexratioatbirth	162	1.052	.019	1.011	1.13
var8 extremepoverty	164	13.24	20.386	0	78.281
var9 giniindex	146	38.131	7.79	24.219	59.067
var10 shareof1pers~d	34	30.61	8.416	9.6	47
var11 crudemarriag~e	46	5.207	1.239	3.1	7.9
var12 crudedivorce~e	41	2.002	.597	.7	3.4
var13 gdppercapita~p	165	20620.542	21271.494	773.573	114985.84
var14 gdpgrowthrate	172	3.41	3.8	-6.988	26.681
var15 inflation	172	4.943	7.018	-3.205	60.987
var16 unemployment~e	158	7.279	5.536	.14	26.06
var17 laborpartici~e	158	68.278	10.699	41.7	88.35
var18 annualworkin~s	64	1850.389	256.752	1353.887	2455.551
var19 annualhealth~e	188	1445.302	1744.342	44.05	10103.09

The ordinary least squares (OLS) regression will be considered for the analysis as a classical choice. Then the multiple regression model was estimated with all the variables mentioned above in Table 10. The regression analysis was conducted as an iterative process, meaning the model was optimized by removing any variable with statistical insignificance (the p-value of the variable's coefficient is less than 0.05). The variable with the highest p-value was removed one at a time and the regression model was estimated again until there is no insignificant variable left in the model as the table 11 below.

Table 11. Linear Regression Model Result (Semi-final)

Source	SS	df	MS			
				Number of obs	=	38
				F(4, 33)	=	7.53
Model	516.281885	4	129.070471	Prob > F	=	0.0002
Residual	565.82891	33	17.1463306	R-squared	=	0.4771
				Adj R-squared	=	0.4137
Total	1082.1108	37	29.2462377	Root MSE	=	4.1408

var0_suiciderate	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
Health	-1.970475	.6107234	-3.23	0.003	-3.213002	-.7279493
var5_sexratioatbirth	-132.8969	56.64472	-2.35	0.025	-248.1414	-17.65236
var9_giniindex	-.3344487	.1354174	-2.47	0.019	-.6099574	-.0589399
var12_crudedivorcerate	4.582386	1.138271	4.03	0.000	2.266557	6.898215
_cons	166.1967	63.03663	2.64	0.013	37.94774	294.4457

The linearity between the independent variables and dependent variable was first observed by scatter plot (Appendix 4.1.1). Since the number of suicide rate is measured as number of suicidal cases on 100,000 people while the other independent variables were measured as a percentage, we decided the transform the independent variable with natural logarithm. By transforming the independent variable, it helps to improve some modelling assumption such as linearity or homoscedasticity while the goodness of fit or the distribution of error terms remain unchanged. After transformation, the linearity is more visible in the scatter plot between dependent and independent variables (Appendix 4.1.2). Table 12 below presents all the variables in our final linear regression model.

Table 12. Descriptive summary of variables in final linear regression model

Variable	**Obs**	**Mean**	**Std. Dev.**	**Min**	**Max**
var0_suiciderate (Log)	203	2.169	.605	.817	4.036
Health	203	5.916	1.041	2.272	9.069
var5 sexratioatbirth	162	1.052	.019	1.011	1.13
var9 giniindex	146	38.131	7.79	24.219	59.067
var12 crudedivorce~e	41	2.002	.597	.7	3.4

The result of our final linear regression model is presented in table 13 below. All the independent variables are statistically significant and the goodness of fit using adjusted R-squared is 0.4082, indicating that the regression model can explain 40% of the variation in suicide rate by four mentioned independent variables.

Table 13. Linear Regression Model (Final Result)

Source	SS	df	MS		
				Number of obs =	38
				F(4, 33) =	7.38
Model	4.90943309	4	1.22735827	Prob > F =	0.0002
Residual	5.48893242	33	.166331285	R-squared =	0.4721
				Adj R-squared =	0.4082
Total	10.3983655	37	.281036906	Root MSE =	.40784

var0_suiciderate_log	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
Health	-.1937922	.0601514	-3.22	0.003	-.3161712	-.0714132
var5_sexratioatbirth	-17.80519	5.579057	-3.19	0.003	-29.15586	-6.454512
var9_giniindex	-.03318	.0133375	-2.49	0.018	-.0603155	-.0060446
var12_crudedivorcerate	.384326	.1121107	3.43	0.002	.1562352	.6124168
_cons	22.70262	6.208609	3.66	0.001	10.07111	35.33413

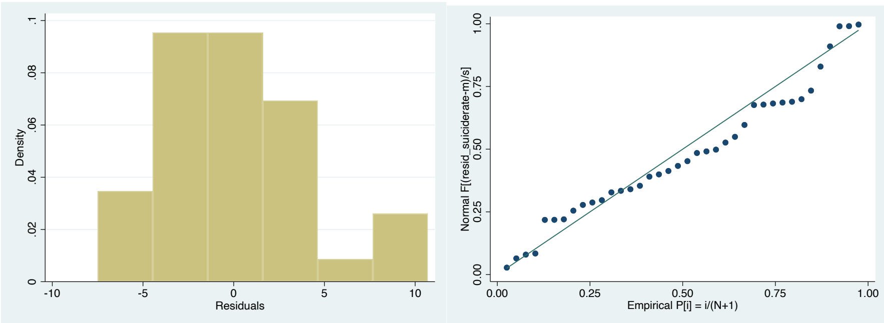
After obtaining the result of the linear regression model, the properties of linear regression model such as existence of linearity, homoscedasticity or autocorrelation will be examined. First, the linearity among the parameter was examined with the Ramsey's RESET test to ensure the correct specification of the model. With the p-value of 0.4, the null hypothesis of the correct linear specification is not rejected, and we confirm the model is specified correctly.

For the assumption of homoscedasticity, the White and Breusch-Pagan tests were employed to test the null hypothesis of homoscedasticity. The p-value of White's test of 0.16 indicates not enough evidence to reject the null hypothesis so the variance of the error terms is constant in the model. The residual-versus-fitted plot was created and presented in Appendix 4.1.4.

For the assumption of autocorrelation, since the observation data is only cross-sectional in year 2017 and non-time-series data so there is no need to employ the Durbin-Watson's test for serial autocorrelation.

For the assumption of no multicollinearity among independent variables, the test of the variance inflation factor (VIF) was employed to examine any perfect correlation existing among independent variables. The overall mean VIF of 1.15 indicates that over 90% of the variation among independent variables is independent from the other variables. Therefore, it is confirmed that there is no multicollinearity in the model.

To test the last assumption for normality in the error terms, the graphical analysis using histogram and normal probability plot was employed and presented as figure 4 below. To confirm, the Shapiro-Wilks test was performed to formally confirm the normal distribution in residuals with p-value of 0.8, indicating that there is not enough evidence to reject the null hypothesis of normal distribution in residuals.



Shapiro-Wilk W test for normal data

Variable	Obs	W	V	z	Prob>z
resid	38	0.98248	0.666	-0.854	0.80343

Figure 4. Histogram and Normal Distribution Plot for Residuals and Shapiro-Wilk test for Normality in Residuals

After confirming that the linear regression model does not violate any assumption, we can look at the result and analyze the meaning of each parameter. Despite all the significant independent variables, it is noticed that most of the dependent variables are near to zero effect (Health factor, Sex ratio at birth, GINI Index), except for the Crude divorce rate having moderate effect. However, due to low number of observations in this cross-sectional data, we unfortunately have to accept this near-zero effect as a limit to our study.

4.2 Binary logistic regression

In this section, the goal is to examine the relationship between suicide rate and other independent variables from another regression angle of binary logistic regression. To perform the analysis, the continuous dependent variable Suicide Rate was transformed into binary categorical variable by assigning the value of 1 for any country with suicide rate higher than average of 10.9 and assigning the value of 0 for any country with suicide rate lower than average.

The data for binary logistic regression analysis is the same cross-sectional data from year 2017 with the dependent variable of binary suicide rate (cut-off at average value) and the independent variables as Health factor, Sex Ratio at birth, GINI index and Crude divorce rates. Those independent variables have been proved to be statistically significant in previous model, so we want to utilize them for this analysis.

Table 14. Binary Logistic Regression

```
Logistic regression                                     Number of obs =     38
                                                        LR chi2(4)    =  10.21
                                                        Prob > chi2   = 0.0370
Log likelihood = -21.233784                             Pseudo R2     = 0.1938
```

var0_suiciderate_binary	Odds ratio	Std. err.	z	P>|z|	[95% conf. interval]	
Health	.4237394	.1625567	-2.24	0.025	.1997825	.898753
var5_sexratioatbirth	1.36e-23	4.73e-22	-1.52	0.129	3.86e-53	4812087
var9_giniindex	.8840647	.0672148	-1.62	0.105	.7616718	1.026125
var12_crudedivorcerate	3.021927	2.090717	1.60	0.110	.7787194	11.727
_cons	2.87e+27	1.11e+29	1.63	0.102	3.43e-06	2.39e+60

Note: **_cons** estimates baseline odds.

Table 14 above presents the result of our binary logistic regression analysis. The number of observations is 38 with the ratio between observation and independent variable is nearly 10:1, which is adequate for conducting the analysis. The overall Prob > chi^2 is 0.03, indicating only 3% chance that all parameters specified are zero so we can conclude that the model is statistically significant and valid. The Pseudo R^2 of 0.1938 explains only 19.38% of the variation in the dependent variable.

When considering the p-values of the independent variables, only Health factor is statistically significant with below 0.05 p-values while the other variables have nearly 0.1 p-value. Therefore, we conclude that the model is not strong enough to generalize any future suicide rate (binary) based on the existing independent variables, but it is important to notice that Health factor still stays significant across the linear regression and binary logistic regression.

We implemented the Hosmer-Lemeshow goodness of fit test and the result is presented in table 15 below. Unlike the Chi^2 test above, the non-significant Chi^2 test in this goodness-of-fit test indicates a good model fit to the data with Chi^2 of 5.7 and p-value of 0.68.

Table 15. Hosmer-Lemeshow Goodness of fit test

```
. estat gof, group(10)
note: obs collapsed on 10 quantiles of estimated probabilities.

Goodness-of-fit test after logistic model
Variable: var0_suiciderate_binary

 Number of observations =     38
       Number of groups =     10
Hosmer-Lemeshow chi2(8) =   5.70
            Prob > chi2 = 0.6805
```

For further assessment of model fitness, the Table 16 presents the classification result of our binary logistic regression model. In total, the model predicted correct 71% of the cases with 13 out of 19 observation (68.4% sensitivity) for suicide and 14 out of 19 (73.7% specificity) for non-suicide.

Table 16. Classification for Suicide Rate

	Actual Suicide	**Actual Non-Suicide**	**Total**
Predicted Suicide	13	5	18
Predicted Non-Suicide	6	14	20
Total	19	19	38
	Sensitivity	Specificity	Correct classification %
	68.42%	73.68%	71.05%

To examine the relationship between sensitivity and specificity, the probability cutoff graph was presented in Appendix 4.2.1. It is interesting to notice that when increasing the cutoff from 0.5 to around 0.6, the sensitivity stays the same while the specificity increases proportionately. The classification was implemented again with the cutoff of 0.6 and the sensitivity stayed at 68% while the specificity increased from 73% to 84% and the overall accuracy improved from 71% to 76%.

4.3 Analysis of variance (ANOVA)

In the previous sections, we have built the multiple linear regression and the binary logistics regression models. Those models illustrate the relationship between explanatory variables and explained variable.

In this section, we will focus on 2 variables: regions and income groups. We will analyze how the suicide rate varies in different regions and among income group level. The predictive powers of income groups and regions factors on the suicide rate are tested in this part.

The ANOVA analysis consists of 3 main parts: One-way ANOVA analysis, two-way ANOVA analyses of categorical variables with and without interaction.

4.3.1 One-way ANOVA

One-way ANOVA analysis is carried out with suicide rate (effect variable) and region (causal categorical variable). This comprises 2 parts: the test of the equality of variances across groups and the test of equal mean.

In the first test, we want to do the comparison of suicide rate in different regions. The descriptive statistic in table 17 summarizes the mean of suicide rate, the standard deviation and frequency in each region. As in this table, the highest mean is of East Asian Pacific (EAP) region (12.69%), followed by Sub-Saharan Africa (SSA) and Europe and Central Asia (EURCA). Means of those groups are closer than the remaining groups. The other 3 groups that can be formed together are Latin America, North America and South Asia region. The lowest mean belongs to Middle East and North Africa (MENA) region with nearly 5%. However, the difference between means is not high (less than 10%).

Table 17. Summary of Suicide rate by regions

Region	Mean	Std. dev.	Freq.
EAP	12.694955	6.8746775	34
EURCA	11.741619	8.2377436	52
LA	8.0689534	6.5214879	34
MENA	4.9896433	3.0589633	22
NA	8.3322651	4.3656829	4
SA	8.431576	4.7313789	11
SSA	12.576976	6.3405664	46
Total	10.497175	7.066817	203

The analysis of variance test is carried out next. The assumption is that variances are equal among regions. Table 18 shows the result of this test. Based on that, sum of squares (SS) between groups are much smaller than within groups. The p-value (Prob>chi2) is statistically significant (p-value = 0.001, less than 0.05). Therefore, we reject the null hypothesis of equal variances and conclude that the variances among groups are not the same.

Table 18. Analysis of variance test result

Source	SS	df	MS	F	Prob > F
Between groups	1377.21141	6	229.535235	5.16	0.0001
Within groups	8710.6488	196	44.4420857		
Total	10087.8602	202	49.939902		

Bartlett's equal-variances test: chi2(6) = 23.7013 Prob>chi2 = 0.001

The second test is implemented to test the mean of suicide rate among regions. The assumption is that they are the same. The test was performed using the multi-comparison test Bonferroni. Bonferroni requests pairwise comparison between regions. Table 19 is the result of this test. Each cell comprises of 2 lines. The first line is difference between means and the second line indicates the significant level for the null hypothesis. Based on it, we can see that the means of most regions are not much different at 5% level. It is the same as what we anticipate in the descriptive statistics table.

There are only 3 pairs that means between them are significantly different: MENA regions compared to EAP, EURCA and SSA.

Table 19. Comparison of Suicide rate by Region (Bonferroni)

	EAP	EURCA	LA	MENA	NA	SA
EURCA	-.953337					
	1.000					
LA	-4.626	-3.67267				
	0.098	0.280				
MENA	-7.70531	-6.75198	-3.07931			
	0.001	0.002	1.000			
NA	-4.36269	-3.40935	.263312	3.34262		
	1.000	1.000	1.000	1.000		
SA	-4.26338	-3.31004	.362623	3.44193	.099311	
	1.000	1.000	1.000	1.000	1.000	
SSA	-.117979	.835357	4.50802	7.58733	4.24471	4.1454
	1.000	1.000	0.066	0.000	1.000	1.000

4.3.2 Two-way ANOVA (without interaction)

In this part, we add another categorical variable “Income group” to the analysis since two-way ANOVA analysis is used for more than 1 group categorical variable. Hence, our explanatory variables are “Income group” and “Region” while our explained variable is remained “Suicide rate”. We first plot the average suicide rate of regions in different income groups to have primary insights (figure 7). In this figure, EURCA region has quite high average suicide rate in most income group, except upper middle income (UM). The rate is also high in SSA region across most income groups. MENA region, in contrast, seems to have the lowest average suicide rate in all income groups.

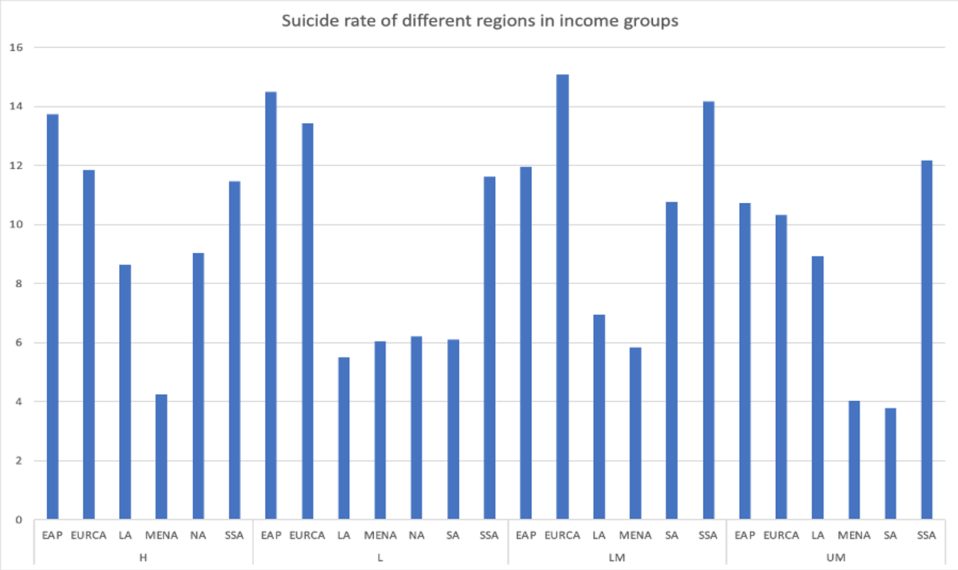


Figure 5. Average suicide rate of regions in different income groups

After gaining primary insights, the two-way ANOVA test is implemented. According to the result of this test (table 20), R-squared and adjusted R-squared are 0.1431 and 0.1031, respectively. This means that the model of the two grouping variables here can explain 14.31% variation of the dependent variable.

Table 20. Two-way ANOVA (without interaction) test result

	F	Prob>F
Model	3.58	0.0004
Region	4.93	0.0001
Income group	0.49	0.6885
Number of observations	203	
R-squared	0.1431	

The F-test for the whole model is similar to the regression model. Since the Prob>F = 0.0004, much less than 0.05, it is statistically significant. It indicates that the regressors here, either one of them or both, have predictive powers on suicide rate.

The F-test of region variable on suicide rate is statistically significant at 5% level (Prob>F = 0.0001, less than 0.05). In contrast, the F-test of income parameter is insignificant (Prob>F = 0.6885, much larger than 0.05). Therefore, "Region" is the factor that has effect on the dependent variable while "Income" seems to have no impact on suicide rate.

This model just told us whether they have impact or not on effect variable. It doesn't tell us how the effects of regions or the income groups differ in terms of suicide rate. Therefore, we need more test to interpret the effect of each region and income group and how they impact the dependent variable.

We run the regress model and show results as in appendix 4.3.1. In this table, Root MSE is 6.6926. This tells us that if we predict suicide rate by only region and income groups, we can explain 6.69% its variation. In the income group categorical variable, given high income group as the base level, only low middle income group has higher average suicide rate. Low income and upper middle-income group have slightly less rate of suicide. However, since all p-values of all groups are larger than 0.05, those differences are not significant. Similarly, in the region variable, EAP is chosen as base level, all other regions seem to have lower suicide rate than EAP but most of the p-values are statistically insignificant; hence, the difference is not much. Only MENA differs EAP significantly in suicide rate because its p-value is nearly 0, less than 0.05. Appendix 4.3.2 shows the means of suicide rate for different income groups in all regions.

We also carried out the test of effect size. The result shows Eta-squared indicators of the model and explanatory variables. According to this result, the model explains 14.3% the dependent variable, whereas "Region" explains 13.29% and "Income group" explains 0.7%. Clearly, region variable has much more predictive power on suicide rate than the income groups.

To check, if the variation of suicide rate is similar in all regions and all income group or not, we applied Breusch–Pagan/Cook–Weisberg test for heteroskedasticity. The final result shows that the Prob>Chi2 value is much less than 0.05; therefore, the null hypothesis is rejected. This means that the variance is not constant in all regions and along the income groups.

4.3.3 Two-way ANOVA (with interaction)

Similar to the previous section, we do two-way ANOVA analysis again with 2 explanatory categorical variables income groups and regions and 1 explained parameter. We added the interaction between "Income group" and "regions". We now have 3 effects: region, income, and their interaction. We run the test again to see whether the differences among regions stay the same in income groups. The result shows that R-squared now is improved from 14.3% to 17.2%.

Table 21. Two-way ANOVA (with interaction) test result

	F	**Prob>F**
Model	1.54	0.0601
Region	4.50	0.0003
Income group	0.60	0.6165
Region#income	0.41	0.9740
Number of observations	203	
R-squared	0.1719	

The p-values of the interaction are higher than 0.05, which is insignificant. It indicates that the differences of regions are not statistically different between income groups. Appendix 4.3.3 clearly illustrates the findings.

The effect size test is also run to see which variable has the most impact on the explained variable. The result shows that based on the Eta-squared value, the model explains 17.1% the variation of suicide rate. "Region" variable explains 13% whereas "income" variable and the interaction explain 1% and 3.3% respectively.

Appendix 4.3.4 put income and region in final comparison and grouping them called Tukey groups. Those margins share the same letter in the group label are not significantly different at 5% level. As we can see in appendix 4.3.4, all categories in the interaction parameter are allocated in the same group (group A). Hence, the means between them are insignificant. Regions EAP, EURCA and SSA are in the same group B while MENA is in group A and LA in group AB. This implies that means in regions EAP, EURCA and SSA are the same whilst means in regions MENA and LA are statistically significant and different from the rest.

5 REGRESSION WITH PANEL DATA

In this part, the report will go through the panel data analysis. While the previous analyses cover data in 2017, this one will study data of period from 2007 to 2017.

During the multiple linear regression model part, we have proven that among all parameters, variables in health factor have the most effect on our dependent variable. This analysis will take all variables used in the Multiple Regression Model section. However, in a long period of time, some variables have missing data leading to bias or unreliability. Therefore, we built regression model again will full data from 2007 to 2017 and do the stepwise regression model until there is no insignificant variable in our dataset in predicting explained variable. The variables chosen for the panel data regression analysis are presented in descriptive statistics table below:

Table 22. Descriptive statistics of explanatory variables and explained variable

Variable	Obs	Mean	Std. dev.	Min	Max
var0_suici~e	2,233	11.13043	7.775275	2.26359	71.76835
var1_human~n	2,097	.8589935	1.606063	-2.415633	5.336182
var2_civil~s	1,892	.6875867	.2659516	.006	.993
var5_sexra~h	1,782	1.052199	.0211604	1.009	1.17
var7_lifes~n	1,413	5.431725	1.130766	2.662	7.971

```
-------------+---------------------------------------------------------
     lngdppc |      1,816    9.199072    1.193053   6.421996   11.86101
var15_infl~n |      1,890     5.22458    8.457166  -30.19965   95.40866
var20_anxi~r |      2,233    4.356732    1.214954   1.976211   8.993181
var21_depr~n |      2,211    3.913354    .9387951   1.640902   7.444226
var22_ment~h |      2,211     13.3014    2.042943   9.467642   19.35467
-------------+---------------------------------------------------------
var23_alco~g |      2,233    2.211706    1.138234   .6745159   5.946469
```

As we can see, all the variables in Multiple Regression Model are included in this list. There are some other variables added because of the rich volume and significance in affecting the dependent variable. Variable 1 (Human right protection index) and 2 (Civil Liberties index) represent the political factor. Variable 5 (Sex ration at birth) and 7 (Life Satisfaction) represent the social factor. Likewise, parameters lngdppc (the logarithm of GDP per capita) and inflation rate represent economic factor and lastly, the rest (variable 20 to 23) are health factor. We want to check our hypothesis in a long period.

The panel data analysis is generated as:

$$y_{it} = b_{0i} + b_1*y_{1it} + b_2*y_{2it}+ b_3*y_{3it}+ b_4*y_{4it}+ b_5*y_{5it}+ b_6*y_{6it}+ b_7*y_{7it} + b_8*y_{8it} + b_9*y_{9it} + b_{10}*y_{10it} + u_{it}$$

The variable i here indicates each country and t is the year in the timeframe. y_{it} is our dependent variable and $x_{1\text{-}10it}$ is the independent variables. u_{it} is the error term.

Next, we will implement the OLS regression. The result is summarized as in table 21. The F-test value is 74.27 and the p-value is less than 0.05, which means that the model is statistically significant. The adjusted R-squared is 0.3686 indicating that this model fits the data well enough. Similar to the multiple regression part above, the sex ratio at birth variable has the most effect on our explained variable. However, while in this model, the explanatory parameter increases by 1 unit leading to the decrease of suicide rate by 44 cases over 100,000 people, in the multiple regression model, the suicide rate decreases less with 22 cases only.

Table 21. Panel data regression: the effects of variable on suicide rate over 100,000 people.

Method	**FE estimations**	**RE estimations**	**Pooled OLS**
Var1_Humanrightprotection	-0.3964045	-0.3723554	0.3989672
	(0.000)	(0.001)	(0.008)
Var2_Civiliberties	-0.3813326	0.9871436	3.336831
	(0.630)*	(0.193)*	(0.000)
Var5_Sexratioatbirth	-53.23107	-42.53704	-44.04323
	(0.000)	(0.000)	(0.000)

Var7_Lifesatisfaction	-0.81864 (0.336)*	-0.0785211 (0.362)	-1.42874 (0.000)
Lngdppc	-2.543406 (0.000)	-2.155106 (0.000)	0.6845363 (0.002)
Var15_Inflation	0.167463 (0.000)	0.0172611 (0.000)	0.0807698 (0.000)
Var20_anxietydisorder	-0.5370459 (0.360)*	-0.415901 (0.376)*	-0.9757123 (0.000)
Var21_depression	1.328431 (0.041)	1.655494 (0.000)	1.9366 (0.000)
Var22_mentalhealth	0.7913175 (0.207)*	0.153448 (0.703)*	-0.7395173 (0.000)
Var23_alcoholanddrug	0.8127228 (0.031)	1.844771 (0.000)	2.265432 (0.000)
Constant	75.6213 (0.000)	64.19634 (0.000)	56.36359 (0.000)
Number of Observation	1256	1256	1256
Number of Countries	140	140	140
R-Squared	0.2407	0.2282	0.3736
Prob>F	0.0000	0.0000	0.0000
Breusch-Pagan test p-value		0.0000	
Hausman test p-value		0.0000	

The fixed effects estimation model is generated and presented in table 5.1. In this model, variable "Sex ratio at birth" also has the largest impact on the dependent variable. The R-squared dropped to 24% and "Civil Liberties", "Life satisfaction" and "Anxiety disorder" and "Mental health" variable are insignificant. However, with the p-value of F test lower than 0.05, this model is statistically significant. This also means that the null hypothesis is rejected and conclude that values of countries are different. Therefore, it implies the fixed effect model is preferred over the OLS regression model.

The RE estimation model is also built and presented in table 5.1. Four variables that are insignificant in this FE model are also statistically insignificant in RE model. The R-squared slightly decreases, 22.82%. However, the Prob>chi2 = 0.0000, much less than 0.05, this model is statistically significant.

In order to see if the countries statistics is significantly different, we run the Breusch and Pagan test. As we can see in table 5.1, the test statistics is quite large (4159.06) and the p-value is much less than 0.05; hence, we reject the null hypothesis that there is no heterogenicity between countries. This means, there are random differences between countries. Therefore, the RE model is preferred over the OLS regression model.

However, we still need to test the consistency of our random estimator. The Hausman test is carried out. Since the test statistics is quite large (61.92) and p-value is much less than 0.05. This implies that the null hypothesis is rejected, and we can conclude our random estimator is not consistent. Therefore, the FE model is considered appropriate and chosen overall. In the FE estimation model, among the significant variables, Human Right Protection has a slight effect on the suicide rate with the coefficient of -0.39. This means if Human Right Protection Index increases 1 unit, the suicide rate decreases 0.39 cases over 100,000 people. Sex ratio at birth, on the other hand, has a huge impact on suicide rate. Increases 1.0 ratio male birth per female birth will lead to decrease 53 cases over 100,000 people. The economic factors including logarithm of GDP per capita and inflation rate have small effect on suicide rate. The health factor, depression rate and alcohol and drug disorder have impacts on suicide rate with coefficient nearly 1. Therefore, we can conclude that during the period between 2007 and 2017, the social factor affects the suicide rate the most, followed by the health factor. The economic and political factors have little to no impact on the suicide rate over 100,000 people.

6 CONCLUSIONS

In summary, the goal of this study is to analyze the factors influencing the suicide rate of a country from the global dataset. We have conducted factor analysis, linear regression analysis, binary logistic regression analysis, analysis of variance (ANOVA) and panel data regression. In this section, the result of the study will be crystalized and presented to answer the hypothesis mentioned earlier.

6.1 Results and Discussion

H1: Countries with better Political factors (low corruption index, high human right protection index) will have lower suicide rates.

Based on analyses, the Political factor has nearly zero effect on influencing the suicide rate of a country due to its statistical insignificance across our models. The mentioned Political factor in this study was factorized from the corruption perception index, the human right protection index and civil liberties index. This rejection of Hypothesis 1 may suggest that the suicide rate is not affected by the selected political independent variables in the scope of this study.

H2: Countries with higher Social factors (balance sex ratio at birth, low inequality disparity) will have lower suicide rates.

The Social factor was constructed on the criteria of social structure such as sex ratio at birth, life expectancy, crude marriage rate, crude divorce rate and the criteria of social advancement such as life satisfaction, human development index, GINI index about inequality distribution.

From the result of our analyses, the Social factor as a latent variable did not have statistical significance on the suicide rate, however, the social independent variables such as sex ratio at birth, crude divorce rate or GINI index have positive effect on the suicide rate. The higher the crude divorce rate is, the more likely the suicide rate increases. The larger the inequality distribution is, the more likely the suicide rate increases. And countries with higher sex ratio at birth, meaning not having equal gender among the population, the more likely the suicide rate is high.

Therefore, we consider accepting the Hypothesis 2 in the scope of this study. The acceptance of the Hypothesis 2 indicates that improving the social aspects in our society would help to reduce the suicide rate in the country.

H3: Countries with higher Economic factors (higher GDP, are more likely to have lower suicide rates.

Regarding to economic factors, our assumption is that the better the economic indicators are, the lower the suicide rate is. The multiple linear regression and binary logistic regression models of year 2017 show no impact of GDP, economic growth rate or any economic factor on suicide rate. The panel data analysis for a period from 2007 to 2017 also points out that economic factor has very little effect on our dependent variable. Moreover, in the ANOVA analysis, the difference in means of suicide rate is not significant across income groups. Therefore, we can conclude, the indicators of good economic factor has very little effect to the suicide rate over 100,000 people.

H4: Countries with better indications of Health factors will be more likely to have lower suicide rates.

Health factors including depression rate, anxiety disorder, mental health, and alcohol & drug disorder parameters, in general, are assumed to have a strong impact on the suicide rate. Our models shows that the statistics is significant, which means that those health factors do have positive influence on the explained variable; the better indicator of health factors, the lower the suicide rate. However, through our analysis, the models show weak impact on the effect variable. This results in the weak acceptance for the fourth hypothesis.

6.2 Evaluation of Validity and Limitations

The validity of the results was tested for each individual analysis performed and reported for each respective model. All assumptions and tests for validity for the regression and ANOVA analyses were accepted and validated since the significant variables were chosen. This includes the assumptions of linearity, homoskedasticity, no autocorrelation, no multicollinearity, and normality for regression and the test of heteroskedasticity for ANOVA.

The study has moderate external validity considering that the aim of the research was to provide results for a more global perspective to understand the influences of suicide rate of many populations across the world. The results provide more generalized results instead of targeting only a small section of a specific population. The sample size was large and diversified. However, a threat to the external validity would be that some countries had less data available and thus less representation. Additionally, the data analyzed was also from one year (2017), which may or may not be transferable to populations in present day.

The variables used in the analysis were unbiased and were based on existing literature related to influences on suicide rate of populations. The measure and analysis of the data was unbiased as well. This indicates that the study can be deemed internally valid.

Several limitations existed in this research that may have impacted the results and validity of the study. As mentioned before, lack of data was a major limitation of the research. Some countries and variables had significantly less available data than others causing a lack of representation of those countries and variables. There was also a lack of data in recent years causing 2017 to be the most recent year with an acceptable amount of data available to conduct the study. The results may have been different had there been more data available. Another limitation regarding the year the data was compiled from is that the results are specific to 2017, and results could vary from year to year and factors change.

6.3 Future Research

As mentioned in Section 6.2 the limitations on the data may have impacted the results of the study, so future research is encouraged to minimize those limitations and provide more results to explain suicide rate. It is suggested that data be collected on more countries in more recent years to update the analysis and results to provide a more recent, relevant, and wider representation of the world's population.

Further investigation of other variables and factors and how they relate to each other, and influence suicide rate is another suggested contribution for future research. After considering the insights provided by this study, it is suggested that there be more research that delves deeper into each factor to formulate suggestions on how to improve the suicide rate across the world. For example, what actions can be taken on specific variables within social factors that is likely to cause an improvement in the suicide rate.

REFERENCES

Ceccherini-Nelli, A. and Priebe, S. (2010). Economic factors and suicide rates: associations over time in four countries. *Social Psychiatry and Psychiatric Epidemiology*, 46(10), pp.975–982.

Cooper Smith (2018). The Link Between Alcohol Use and Suicide - Alcohol Rehab Guide. [online] Alcohol Rehab Guide. Available at: https://www.alcoholrehabguide.org/resources/dual-diagnosis/alcohol-and-suicide/.

Emamgholipour, S., Arab, M. and Shirani, R. (2021). Socioeconomic Determinants of Suicide in Iran: Panel Data Study. Iranian Journal of Public Health.

Fountoulakis, K.N., Kawohl, W., Theodorakis, P.N., Kerkhof, A.J.F.M., Navickas, A., Höschl, C., Lecic-Tosevski, D., Sorel, E., Rancans, E., Palova, E., Juckel, G., Isacsson, G., Jagodic, H.K., Botezat-Antonescu, I., Warnke, I., Rybakowski, J., Azorin, J.M., Cookson, J., Waddington, J. and Pregelj, P. (2014). Relationship of suicide rates to economic variables in Europe: 2000–2011. British Journal of Psychiatry, 205(6), pp.486–496.

Girard, C. (1993). Age, Gender, and Suicide: A Cross-National Analysis. American Sociological Review, 58(4), p.553.

Kaiser, H. F. 1974. An index of factor simplicity. Psychometrika 39: 31–36.

Landberg, J. (2009). Per Capita Alcohol Consumption and Suicide Rates in the U.S., 1950–2002. Suicide and Life-Threatening Behavior, 39(4), pp.452–459.

Lee, L., Roser, M. and Ortiz-Ospina, E. (2019). Suicide. [online] Our World in Data. Available at: https://ourworldindata.org/suicide.

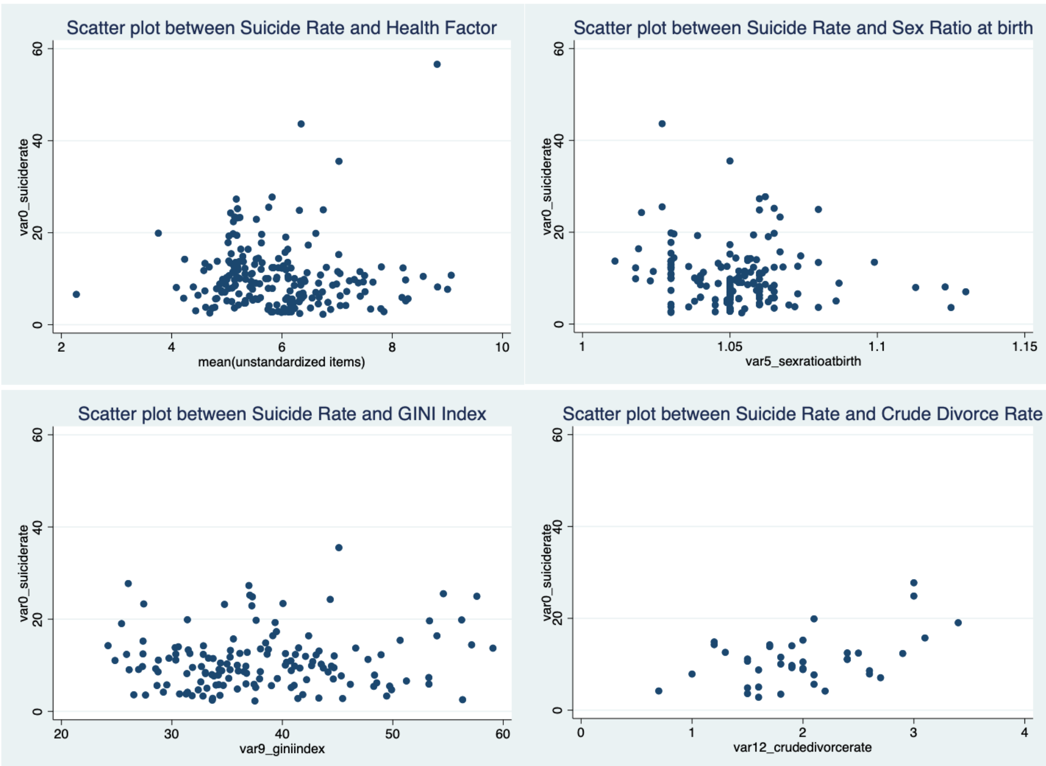
Sedgwick, R., Epstein, S., Dutta, R. and Ougrin, D. (2019). Social media, internet use and suicide attempts in adolescents. Current Opinion in Psychiatry, [online] 32(6), pp.534–541. Available at: https://journals.lww.com/co-psychiatry/Fulltext/2019/11000/Social_media,_internet_use_and_suicide_attempts_in.12.aspx [Accessed 3 Nov. 2019].

APPENDIX 4.1 Appendices for Multiple Linear Regression 1

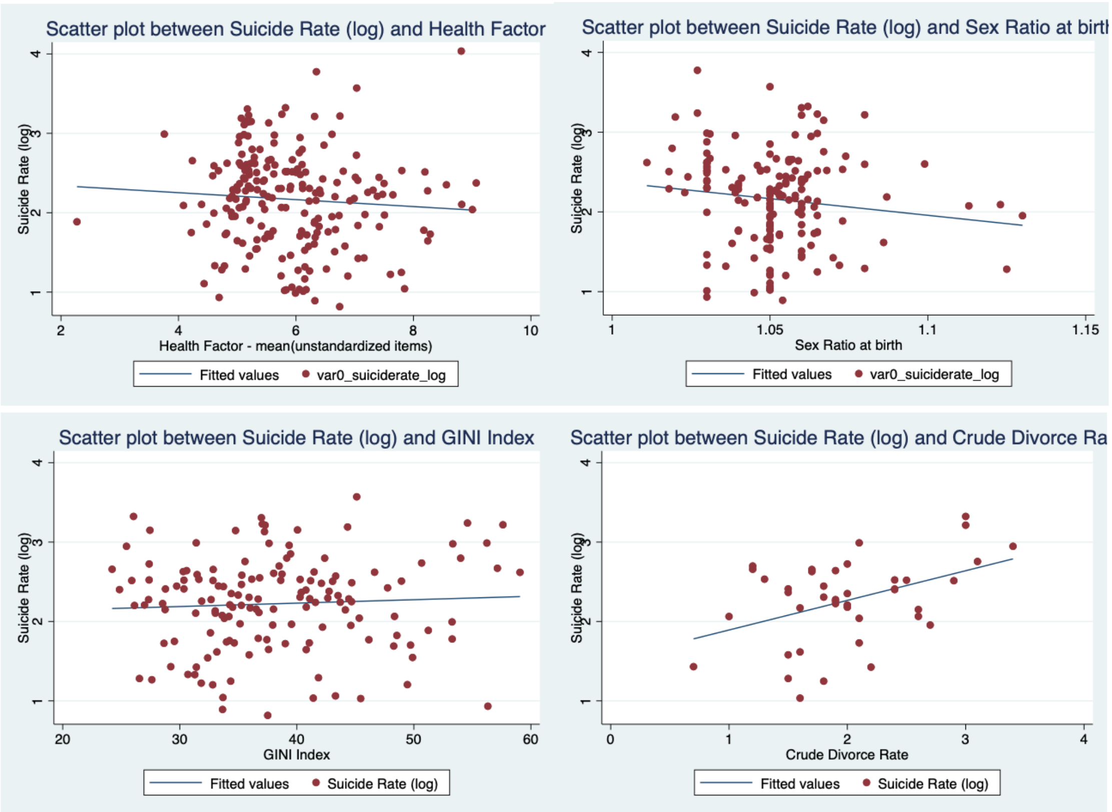
Appendix 4.1.1 Correlation Matrix

	var0_s~e	var1_h~n	var2_c~s	var3_c~x	var4_h~x	var5_s~h	var6_i~e	var7_l~n	var8_e~y	var9_g~x	var10_~d	var11_~e	var12_~e	var13_~	var14_~e	var15_~n	var16_~e	var17_~e	var18_~s	var19_~e	var20_~r	var21_~n	var22_~h	var23_~g
var0_suici~e	1																							
var1_human~n	0.2491	1																						
var2_civil~s	0.3038	0.6756	1																					
var3_corru~x	0.1747	0.6913	0.513	1																				
var4_human~x	0.0603	0.6268	0.5975	0.8994	1																			
var5_sexra~h	-0.1189	-0.1198	0.1742	-0.472	-0.3676	1																		
var6_inter~e	0.1642	0.6876	0.4892	0.9041	0.8406	-0.3297	1																	
var7_lifes~n	0.1226	0.5585	0.3703	0.8951	0.8603	-0.5651	0.8291	1																
var8_extre~y	0.0187	-0.3825	-0.0644	-0.4665	-0.4619	0.1771	-0.5466	-0.3639	1															
var9_ginii~x	-0.1208	-0.6773	-0.4834	-0.4847	-0.5271	-0.0035	-0.5068	-0.4671	0.5558	1														
var10_shar~d	0.4212	0.5715	0.5361	0.6824	0.6516	-0.1216	0.5837	0.6659	-0.1339	-0.4536	1													
var11_crud~e	0.1558	-0.4656	-0.389	-0.2268	-0.3246	0.1068	-0.1639	-0.08	0.2799	0.3778	-0.112	1												
var12_crud~e	0.4862	0.0711	0.2682	0.3828	0.2222	0.0342	0.4076	0.2362	0.0075	0.1911	0.3589	0.3265	1											
var13_gdpp~p	-0.0599	0.6176	0.3146	0.7325	0.6859	-0.5773	0.7389	0.7063	-0.3037	-0.1668	0.2768	-0.2954	0.1301	1										
var14_gdpg~e	-0.1213	-0.355	-0.5351	-0.4645	-0.5602	0.2332	-0.3582	-0.314	0.114	0.0554	-0.3832	0.5515	-0.2984	-0.4765	1									
var15_infl~n	0	-0.4757	-0.8336	-0.4419	-0.5965	-0.0741	-0.3587	-0.3361	0.2073	0.4711	-0.2476	0.4327	-0.0422	-0.2969	0.5472	1								
var16_unem~e	-0.4024	-0.3694	-0.111	-0.4321	-0.2664	0.4298	-0.3755	-0.5082	0.0948	0.2221	-0.4416	-0.1187	-0.0036	-0.3283	-0.168	-0.1184	1							
var17_labo~e	0.3266	0.5629	0.6677	0.714	0.6616	0.0605	0.6876	0.5754	-0.3213	-0.4214	0.7198	-0.1202	0.4376	0.3007	-0.3167	-0.4855	-0.3461	1						
var18_annu~s	-0.0092	-0.5353	-0.2848	-0.7111	-0.6736	0.4493	-0.5943	-0.6722	0.2326	0.2768	-0.5911	0.3748	-0.1382	-0.6048	0.5043	0.1828	0.3267	-0.4276	1					
var19_annu~e	-0.0058	0.351	0.3832	0.7931	0.8537	-0.5755	0.7149	0.7862	-0.25	-0.1477	0.4324	-0.162	0.2276	0.755	-0.5413	-0.4189	-0.3707	0.467	-0.6116	1				
var20_anxi~r	-0.4556	0.2201	0.2433	0.4005	0.4945	-0.0785	0.3563	0.3684	-0.3053	-0.0975	0.052	-0.2396	-0.046	0.4261	-0.3145	-0.3799	0.0756	0.2903	-0.3254	0.5271	1			
var21_depr~n	-0.1303	-0.1264	0.099	0.2081	0.267	-0.0012	0.1988	0.0785	-0.0776	0.203	0.042	-0.0657	0.4567	0.0946	-0.3692	-0.1642	0.5525	0.2062	-0.0411	0.2813	0.4047	1		
var22_ment~h	-0.4563	0.0517	0.1853	0.3826	0.5044	-0.117	0.3822	0.3573	-0.2491	0.0255	-0.0027	-0.2223	0.1208	0.4022	-0.3842	-0.3529	0.3357	0.2503	-0.2662	0.5515	0.875	0.7238	1	
var23_alco~g	0.404	0.2544	0.5535	0.5344	0.4884	-0.3102	0.4914	0.4065	-0.1122	-0.1094	0.2404	-0.1052	0.4348	0.3773	-0.3972	-0.5396	-0.3424	0.4545	-0.1293	0.6137	0.0631	0.1522	0.1328	1

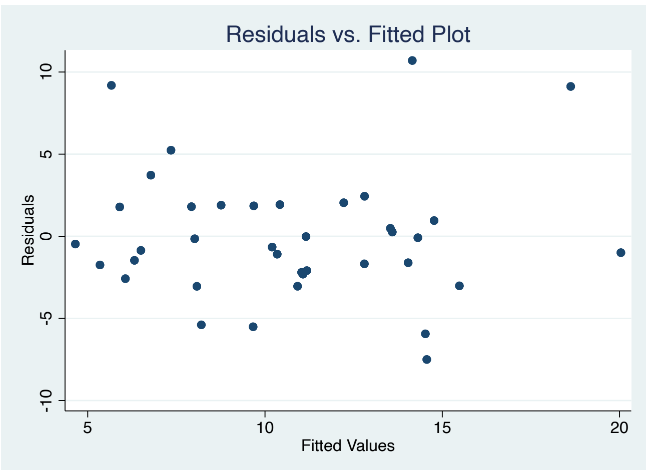
Appendix 4.1.2. Scatter plot between dependent and independent variables



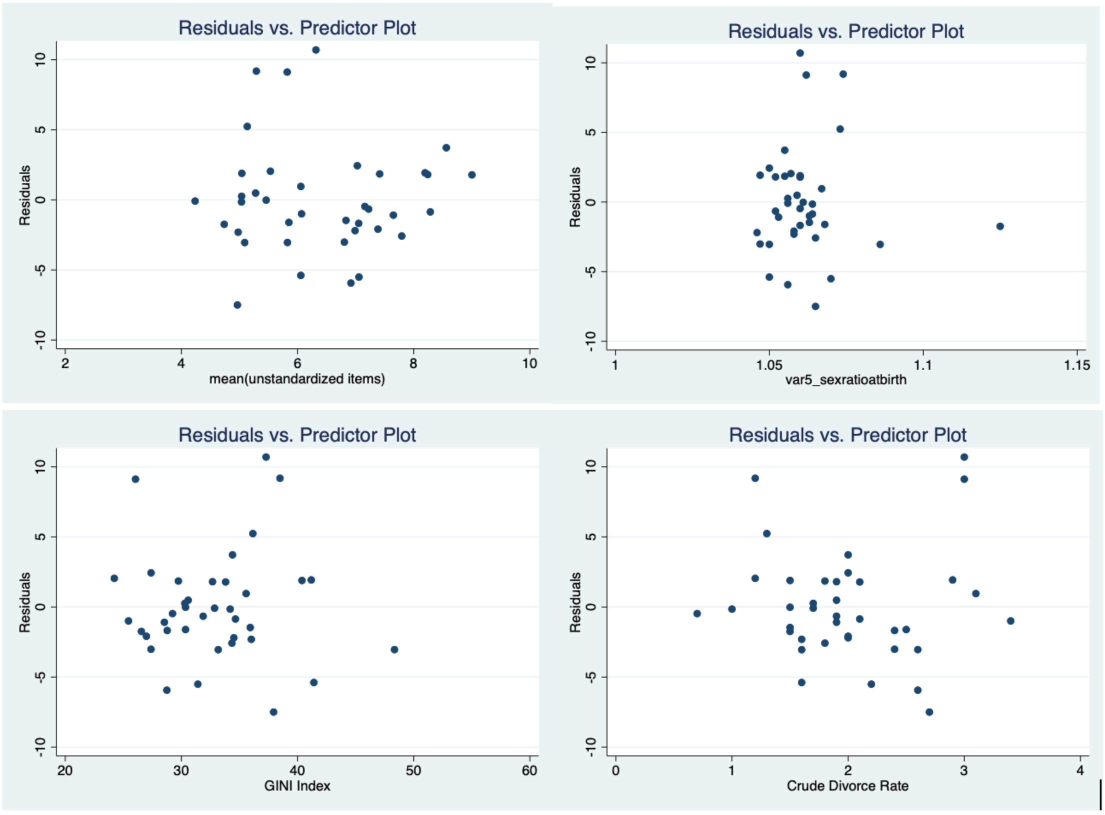
Appendix 4.1.3. Scatter plot between dependent (log) and independent variables



Appendix 4.1.4. Residuals vs Fitted Plots

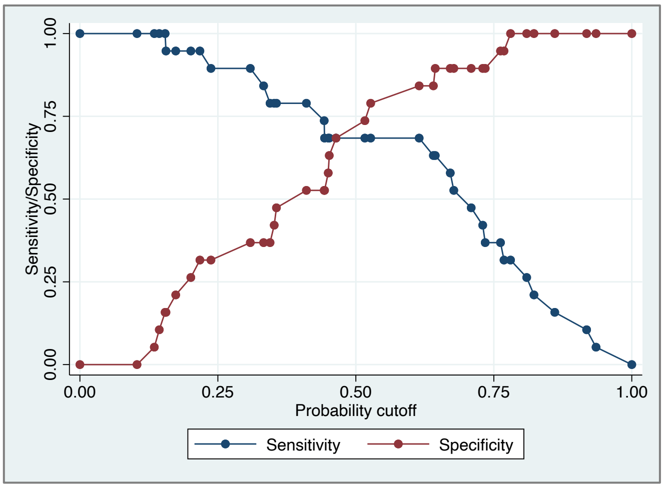


Appendix 4.1.5. Residuals vs Predictor Plots for each independent variable



APPENDIX 4.2 Appendices for Binary Logistic Regression

Appendix 4.2.1 Sensitivity and Specificity

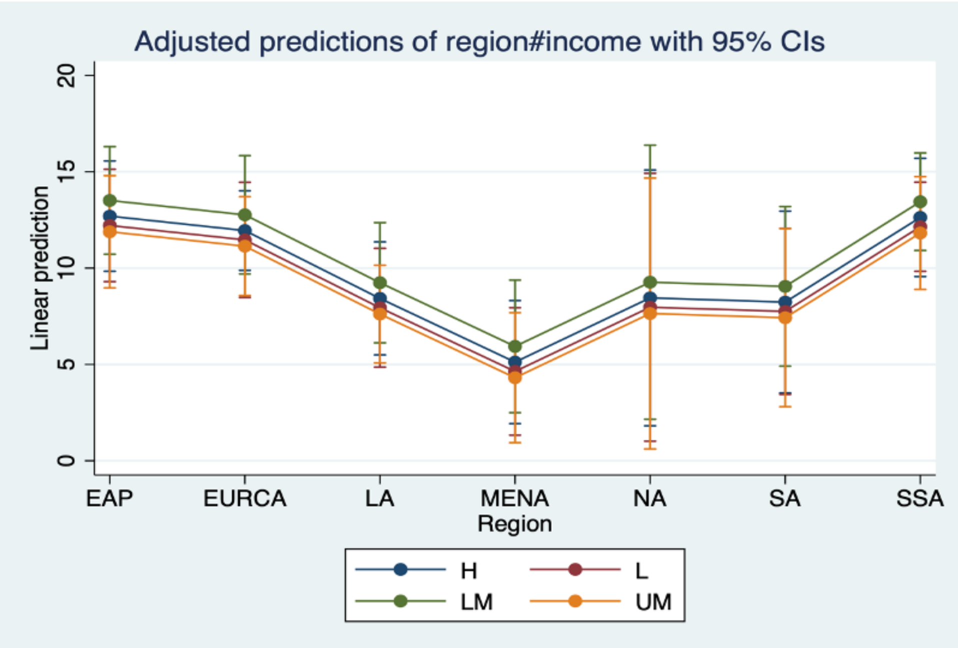


APPENDIX 4.3 Appendices for ANOVA

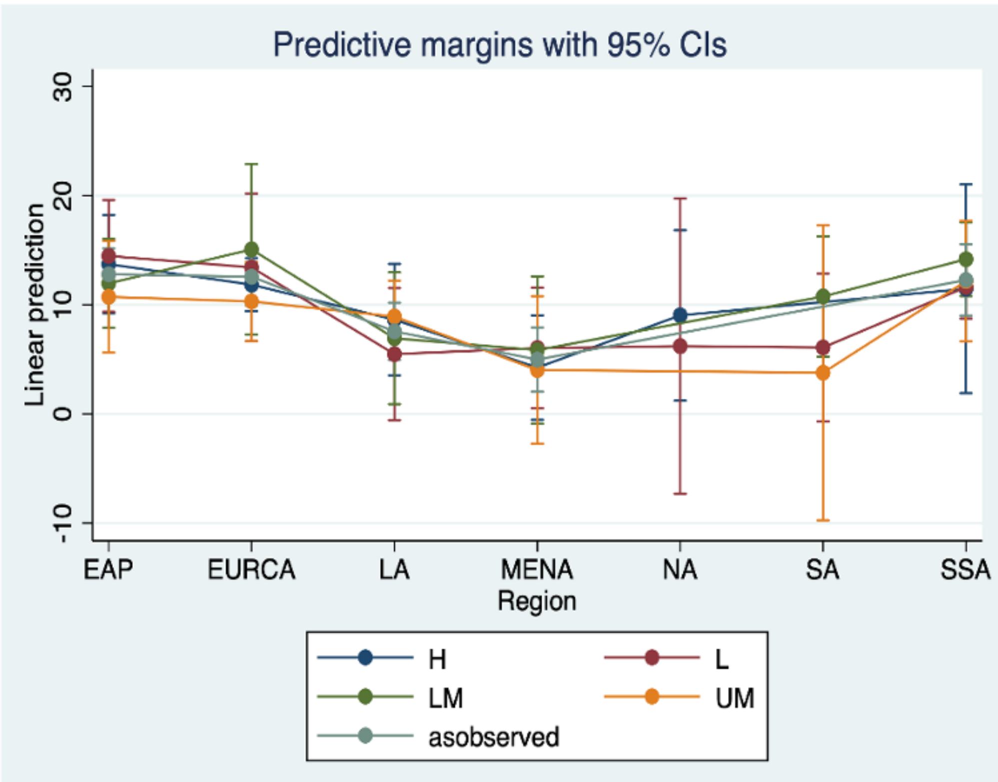
APPENDIX 4.3.1. Regress model, base level test results.

```
      Source |       SS           df       MS      Number of obs   =       203
-------------+----------------------------------   F(9, 193)       =      3.58
       Model |  1443.27038         9  160.363375   Prob > F        =    0.0004
    Residual |  8644.58983       193  44.7906209   R-squared       =    0.1431
-------------+----------------------------------   Adj R-squared   =    0.1031
       Total |  10087.8602       202   49.939902   Root MSE        =    6.6926
------------------------------------------------------------------------------
var0_suici~e | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
      region |
        EAP  |          0  (base)
      EURCA  |  -.7476275   1.557377    -0.48   0.632    -3.819291    2.324036
         LA  |  -4.272088   1.663601    -2.57   0.011     -7.55326   -.9909152
       MENA  |  -7.576888    1.84117    -4.12   0.000    -11.20828   -3.945491
         NA  |  -4.244426   3.593665    -1.18   0.239    -11.33233    2.843474
         SA  |  -4.461194   2.363858    -1.89   0.061    -9.123506    .2011189
        SSA  |  -.0670694   1.562162    -0.04   0.966    -3.148171    3.014033
             |
      income |
          H  |          0  (base)
          L  |  -.4842562   1.482146    -0.33   0.744     -3.40754    2.439027
         LM  |   .8155068   1.519586     0.54   0.592    -2.181621    3.812634
         UM  |  -.8108549   1.359655    -0.60   0.552    -3.492546    1.870836
             |
       _cons |   12.69776   1.452536     8.74   0.000     9.832873    15.56264
------------------------------------------------------------------------------
```

Appendix 4.3.2. Means of suicide rate for different income groups in all regions without interaction



Appendix 4.3.3. Means of suicide rate for different income groups in all regions with interaction



Appendix 4.3.4. Pairwise comparison of means between regions and income groups with interaction.

```
Pairwise comparisons of marginal linear predictions

Margins: asobserved

----------------------------

              |    Number of

              |  comparisons

--------------+-------------

       region |           21

       income |            6

region#income |          378

----------------------------

-----------------------------------------------

              |                          Tukey

              |     Margin   Std. err.   groups

--------------+--------------------------------

       region |

         EAP  |   12.79518   1.202684         B

       EURCA  |   12.57492   1.334126         B

          LA  |   7.570764   1.324635        AB

        MENA  |   4.984025   1.492769         A

          NA  |          .  (not estimable)

          SA  |          .  (not estimable)

         SSA  |   12.27704   1.665562         B

              |

       income |

           H  |          .  (not estimable)

           L  |   10.52588   1.214627

          LM  |          .  (not estimable)

          UM  |          .  (not estimable)

              |

region#income |
```

EAP#H	13.72828	2.283542	A
EAP#L	14.47971	2.589293	A
EAP#LM	11.96226	2.065541	A
EAP#UM	10.73302	2.589293	A
EURCA#H	11.84554	1.230409	A
EURCA#L	13.43014	3.425312	A
EURCA#LM	15.0779	3.95521	A
EURCA#UM	10.31416	1.830907	A
LA#H	8.638091	2.589293	A
LA#L	5.487715	3.063693	A
LA#LM	6.940511	3.063693	A
LA#UM	8.925685	1.661521	A
MENA#H	4.252138	2.422062	A
MENA#L	6.03928	2.796756	A
MENA#LM	5.852033	3.425312	A
MENA#UM	4.027809	3.425312	A
NA#H	9.036822	3.95521	A
NA#L	6.218594	6.850625	A
NA#LM	.	(not estimable)	
NA#UM	.	(not estimable)	
SA#H	.	(not estimable)	
SA#L	6.093401	3.425312	A
SA#LM	10.76522	2.796756	A
SA#UM	3.782434	6.850625	A
SSA#H	11.47066	4.844123	A
SSA#L	11.62452	1.460558	A
SSA#LM	14.17657	1.712656	A
SSA#UM	12.1725	2.796756	A

Note: Margins sharing a letter in the group
label are not significantly different at the 5% level.