Final Project of MA4270 – Design and Analysis of Experiments

Travis Froberg

Comparing Suicide Rates by Age Groupings in Argentina, Finland, and the United States in the Time Period of 1985 to 2015

Section 1 Introduction

The following statistical analysis was done to examine differences in suicide rates among different age groups (e.g. 15-24 years of age) and among three different countries: Argentina, Finland, and the United States. Because suicide rates can be used as one measure of a societies well-being, comparing suicide rates among different countries can help one understand the causes of a societies emotional distress. For example, if one country has a higher suicide rate than another, it is an indication that there are factors in the former country that are contributing to emotional distress for the population. These factors could include economic hardships, a culture that values individualism rather than collectivism, etc. Examining suicide rates through the lens of age-groupings is also important because knowing which age-groups are most effected can help a society determine which interventions would be most helpful to address the problem. For example, if it is found that the people in the age group of 75+ have the highest rates of suicide, a prioritized intervention could be to address chronic loneliness among the elderly.

The data for the following analysis was obtained from kaggle.com and contains data compiled from four different datasets. The full list of datasets is given in the references section at the end of this report under the heading "Original Dataset Retrieved From:" The data was edited by Szamil on kaggle.com and it was this edited dataset that was used for the following analysis. Although the original dataset contained many treatment factors, only two were used in this analysis, and the data for both factors was originally obtained from the World health Organization's data on worldwide suicide rates.

As has been alluded to, the response variable used for this analysis was suicide rate, measured as the number of suicides per 100,000 people. One treatment factor used was the country in which the data was collected with three levels: Argentina, Finland, and the United States. The other treatment factor was the age group of each person/observation in the study, with six levels: 0-14 years, 15-24 years, 25-34, 35-54, 55-74, 75+. The experimental unit and observational unit were the same in this study, namely, an individual person.

There were 1092 observations in the dataset for 18 treatment combination. This means there is was about 61 observations per treatment combination, and that the principle of replication was satisfied. The exact methods of data collection are unknown, however, because the data was originally from the World Health Organization – a reputable organization as far as data collection is concerned – it was taken for granted that proper methods were used to make the observations randomized.

No blocking factors were accounted for in the following analysis. The covariates involved in this study are plentiful but ignored in this analysis. Examples of covariates include: the overall physical health of the population and the difference in living standards among the countries. Another covariate is the time in which the observations were collected. The observations were collected from the time-period 1985 to 2015.

Section 2 Statistical Methods

The R version 4.0.5 (https://www.r-project.org/) was used in the analysis. The overall significance level was set as 0.05.

Section 2.1 Exploratory Data Analysis

Once the data from kaggel.com was in R, the dataset was cleaned so that it only contained the two treatment factors Age Group and Country – and additionally the response variable (suicide rates). A column titled Treatment was then created that contained for each treatment combination. A column containing values 1-1092 in order was also added; this latter column was labeled Possible Ordering and was used to check the independence assumption of the data (this will be explained in detail the Results section.) The reason the column was labeled Possible Ordering instead of Ordering was because the order in which the raw data was collected was unknown. However, after a residual plot to check for independence was created, a clear pattern emerged which suggests that the ordering is in fact correct.

Because the equality of variance assumption appeared to be violated based on the corresponding plot, various transformations on the response variable were tested including: natural log, log base 10, log(suicide rate + 1), and sqrt(suicide rate). The transformations did not fix the problem therefore the untransformed response variable was used for analysis. One benefit of this approach is that the interpretation of the results was easier.

Section 2.2 ANOVA

A two-way complete analysis of variance model was fitted using treatment factors Country and Age Group. This model showed that there was good evidence of an interaction between the two factors. In addition, an interaction plot was created of the two factors which also indicated that there was an interaction. For this reason, a cell-means model was used for further analysis rather than the two-way complete model.

A plot of the standardized residuals against the treatment combinations from the cell-means ANOVA model was then created to check for outliers and to determine if the model was a good fit for the data.

A plot of the standardized residuals against the predicted values from the cell-means ANOVA model was then created to check for equality of variances. In addition, Levene’s Test was used to check the equality of variances assumption.

A Normal Q-Q plot was then constructed to check the normality assumption. In addition, a plot of the standardized residuals against the order of observations was created to check for independence among observations.

After checking model assumptions, the cell-means ANOVA table was analyzed to check if at least one treatment combination differed from one other treatment combination. In addition, the pairwise comparisons for all treatment combinations were tested by analyzing confidence intervals using the Bonferroni, Scheffe, and Tukey methods for multiple comparisons.

Section 3 Results and Conclusions

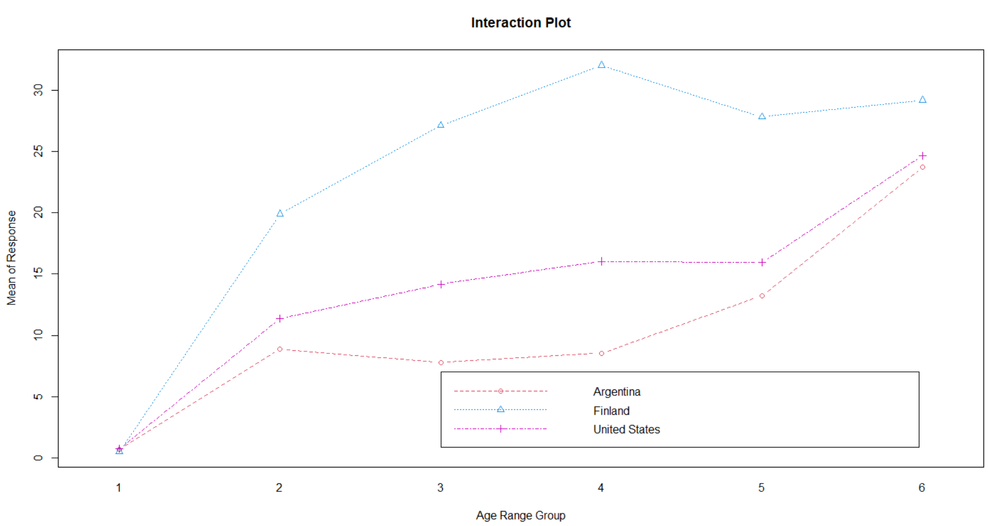
An ANOVA table was created for the two-way complete model (Table 1).

Table 1. ANOVA Table for Two Way Complete Model Using Factors Country and Age Group

Factors and Residuals	Df	Sum of Sqares	Mean of Squares	F value	Pr(>F)
Country	2	28859	14429	84.4790	< 2e-16
Age Group	5	64172	12834	75.1420	< 2e-16
Interaction of Country and Age	10	12048	1205	7.0540	8.37e-11
Residuals	1074	183442	171		

It can be seen from Table 1 that the test statistic for the interaction between Country and Age Group is 7.0540 with a p-value of 8.37 * 10^-11. This means that the null hypothesis, which states that the interaction is negligible, is rejected. This means that there is strong evidence that the interaction between these two factors is significant. To corroborate this finding an interaction plot was created for the factors Country and Age Group.

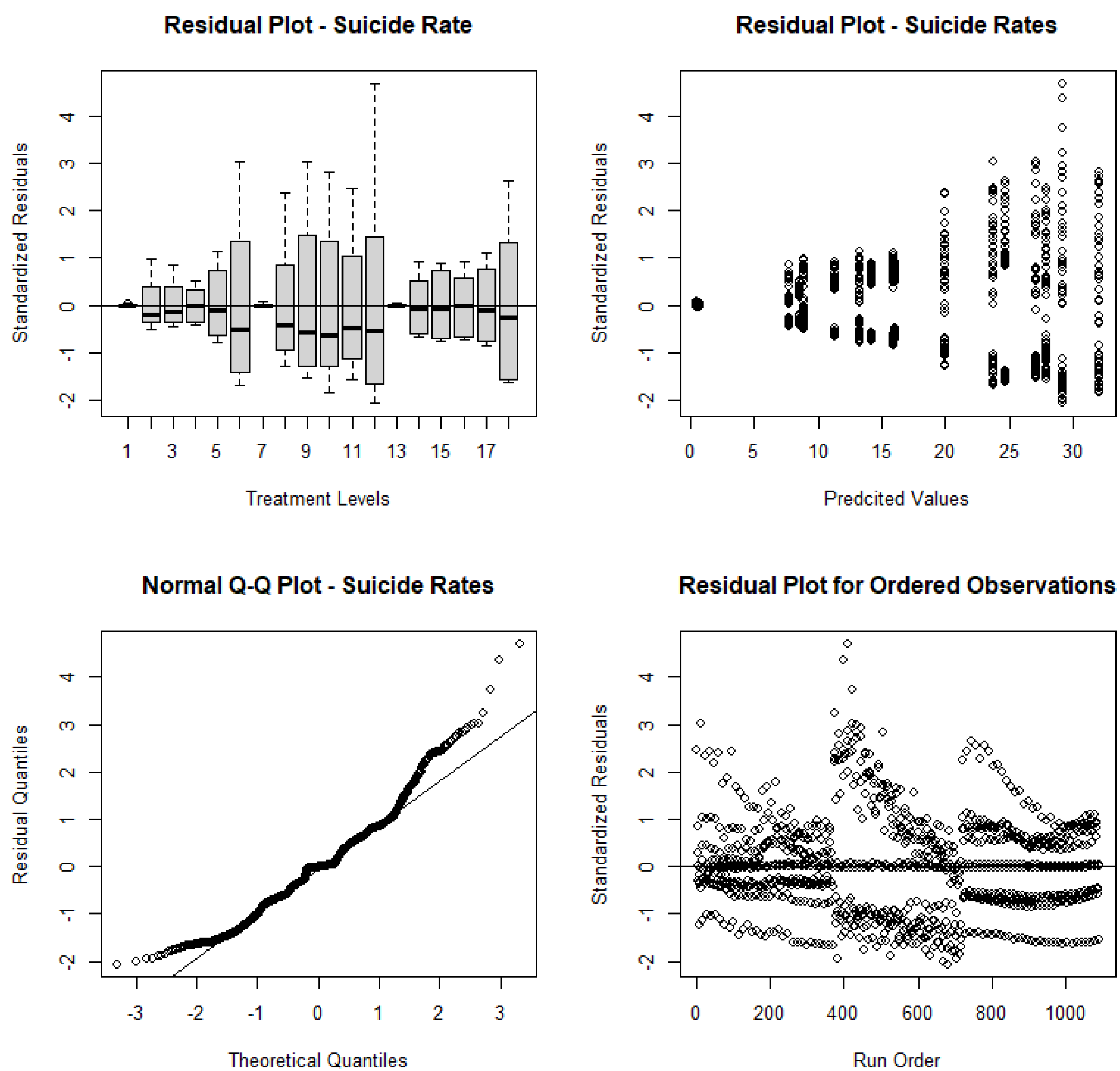
Figure 1. Interaction Plot for Treatment Factors Country and Age Group



From Figure 1, it appears as though there may be a slight interaction between treatment factors Country and Age Group. If there were no interaction, the three lines in the plot would appear parallel. While the lines are partially parallel at some points along the plot and all seem to increase, there are also significant differences at certain points along the graph. Because the plot suggested an interaction in addition to the two-way complete model suggesting the same, a cell-means ANOVA model was used for analysis instead of a two-way complete or main-effects ANOVA model.

Once the cell-means model was fitted, the model assumptions were checked.

Figure 2. Plots used to check model assumptions: (top-left) box plot for standardized residuals and treatment combinations; (top-right) scatter plot for standardized residuals and predicted values; (bottom-left) normal Q-Q plot for standardized residuals. (bottom-right) scatter plot for standardized residuals and observation order



From the top-left plot of Figure 2, it appears as though there are some outliers in the data. An observation could be classified as an outlier if it has a standardized residual value of greater than 3 or less than -3. There are multiple cases were this is the case. For example, for treatment combination 12 there is at least one observation that is greater than 4. However, one explanation for these apparent outliers is that there were a lot of observations in the dataset (1,092 observations). Given this amount of data points, it is likely that some points will be greater than 3 or less than -3. No data points were removed from the data, partly because the exact method of data collection was unknown and therefore it could not be determined if an observation appeared as an outlier due to an experimental error. It was also believed that the possible outliers in the plot would not affect the data to a significant degree.

Looking at the top-right plot of Figure 2, it seems likely that the equality of variances assumption is violated. This is evident by the distinct rightward fanning pattern of the points in the plot. As was stated in the Methods Section, multiple transformations of the response variable (suicide rate) were tested to try to equalize variances, however none of them fixed the problem. For this reason, the data was analyzed further without a transformation of the response variable. Levene's test returned an F-statistic of 104.25 with a p-value of 2.2 * 10^-16. This suggests that the equality of variances assumption is heavily violated.

From the bottom-left plot of Figure 2, it appears that the normality assumption may be violated because the data points stray from the linear line at low and high values of the theoretical quantiles. However, this deviation is not extreme, therefore the normality assumption was determined to hold up and further analysis was performed.

From the bottom-right plot of Figure 2, a clear pattern of dependence between observations can be seen. At the start of the plot, the standardized residual values are large and then proceed to lessen; this suggests a linearly decreasing pattern. This pattern seems to repeat three times throughout the plot. This is because the cycle is the same for each country – and there are three countries. The reason for the dependence among observations is likely due to the different times in which observations were collected. The observations are ordered according to the year in which the observations were collected, and the data for each country spans a time-frame from 1985 to 2015.

With the assumptions of the model checked, the cell-means model was used to test if there was any difference in effect between the treatment combinations. An ANVOA table was created in R, and the F-statistic for the treatment group was 36.1880, with a corresponding p-value of 2.2 * 10^-16. This means that the null hypothesis that there was no difference between treatment combinations in terms of their effects on the response variable was rejected. The conclusion is reached that there is statistically significant evidence that at least one treatment combination differs with at least one other treatment combination in its effect. This was not surprising because there were many treatment combinations and therefore the likelihood that at least two differed in terms of their effect on the response was pretty high.

Confidence intervals for the contrasts of pairwise differences between all treatment combinations were calculated using the cell-means model. The methods of multiple comparison used were the Bonferroni, Scheffe, and Tukey methods, however, the results from the Bonferroni method will be focused on here. The confidence intervals were also constructed without methods of multiple comparison. Due to the large number of pairwise comparisons only the results for three of the comparisons will be displayed here.

Table 2. Confidence Intervals for Pairwise Comparisons of Treatment Combinations of Factors Country and Age Group

Contrast	Point Estimate	SE	Degrees of Freedom	Lower Confidence Interval	Upper Confidence Interval
9 - 15	-12.30	2.39	1074	4.3474	21.559
2 - 8	-11.0647	2.39	1074	-19.6707	-2.459
3 - 15	-6.3829	2.35	1074	-14.8443	2.078

In Table 2, the contrasts can be understood as follows: 9-15 = Finland (25-34 years) subtract United States (25-34 years); 2-8 = Argentina (15-24 years) subtract Finland (15-24 years); 3-15 = Argentina (25-34 years) subtract United States (25-34 years).

One pairwise comparison that was analyzed, shown in table 2, was the suicide rates between treatment combinations 9 and 15. Treatment combination 9 represents the 25-34-year age group in Finland and the treatment combination 15 represents the 25-34-year age group in the United States. The confidence interval for this contrast using the Bonferroni method was (4.3474, 21.559). Because this interval does not include 0, the null hypothesis that the two treatment groups have the same effect on the response is rejected. We therefore can conclude that there is in fact a difference to the 95% confidence level. Because the interval contained only positive values and the contrast was set up as Treatment Combination 9 subtract Treatment Combination 15, we can infer from this confidence interval that the suicide rate in Finland was greater on average than the United States for the time-period from 1985 to 2015. The same conclusions can be drawn for the difference between the 35 to 54-year age group and the 55 to 74 year age group in Finland and the United States (see R output for statistics supporting this claim).

Another pairwise comparison that was analyzed, also shown in table 2, was the suicide rates between treatment combinations 2 and 8. Treatment combination 2 represents the 15-24-year age group in Argentina and the treatment combination 8 represents the 15-24-34-year age group in Finland. The confidence interval for this contrast using the Bonferroni method was (-19.6707, -2.459). Because this interval does not include 0, the null hypothesis that the two treatment groups have the same effect on the response is rejected. We therefore can conclude that there is in fact a difference to the 95% confidence level. Because the interval contained only negative values and the contrast was set up as Treatment Combination 2 subtract Treatment Combination 8, we can infer from this confidence interval that the suicide rate in Finland was greater on average than in Argentina for the time-period from 1985 to 2015. The same conclusions can be drawn for the difference between the 35 to 54-year age group and the 55 to 74 year age group in Argentina and Finland (see R output for statistics supporting this claim).

One more pairwise comparison that was analyzed, shown in table 2, was the suicide rates between treatment combinations 3 and 15. Treatment combination 3 represents the 25-34-year age group in Argentina and the treatment combination 15 represents the 25-34-year age group in the United States. The confidence interval for this contrast using the Bonferroni method was (-14.8443, 2.078). Because this interval includes 0, the null hypothesis that the two treatment groups have the same effect on the response is accepted. We therefore cannot conclude that there is a difference to the 95% confidence level.

Section 4 Discussion

A number of hypothesis were tested in this statistical analysis. A few of the significant findings are discussed here. It was expected at the beginning of the study that Finland would have a lower suicide rate than the United States in the years from 1985 to 2015. This is partially because Finland is known for having many social democratic programs that benefit the population. This includes heavy state investment in education, universal healthcare, etc. In 2021, Finland ranked as the happiest country on earth in the UN Sustainability Development Solution Network's World Happiness Report. For these reasons, it was surprising to see Finland rank higher than both the United States and Argentina in suicide rates.

One possible explanation could be that Finland may have had very high suicide rates in the earliest years for which the data was collected and had lower rates as time progressed. Support for this explanation comes from the World Population Review's analysis of suicide rates which reports that Finland ranked just below the United States in suicide rates in 2021; the U.S. has a suicide rate of 16.1 people per 100,000 people so far in 2021 and Finland has a suicide rate of 15.3 people per 100,000 people so far. However, these figures should be interpreted with caution because the COVID-19 pandemic has changed conditions in both countries as to be abnormal, and it is well-known that the United States' privatized health system performed much worse in dealing with the pandemic than most countries with universal healthcare.

If it is the case that Finland drastically reduced its suicide rate in the years from 1985 to 2021, the United States and Argentina could learn how to produce similar results. As stated above, it is well-known that Finland currently has much more resources invested in social programs than the U.S. and these programs include mental health services which could be effective in lowering the suicide rate. However, this study does not address the reasons Finland has a lower suicide rate than the U.S. now but not previously. Furthermore, this analysis does not address why Finland had a higher suicide rate than the United States. One source of speculation is that because Finland is very far north, the winters are harsh and sunlight is scarcer than in the other two countries. Lack of sunlight is well-known to cause seasonal mood disorders and therefore Finland may have high suicide rates mainly because of natural environmental factors and not because of their type of economic system or social issues. Further research could explore the reasons why Finland had a higher suicide rate than the United States in the time period of this study and has a lower rate than the U.S. in 2021.

Another significant finding from this analysis was that Argentina's suicide rate did not significantly differ from the United States' suicide rate for all age groups. This is significant because American culture, and possibly Western European culture as well, has a tendency for many people to see South America as a much less developed continent than North America or Europe. Some Americans think in black and white about the situation, reasoning that because the United States is the richest country in the world (in terms of GDP) that its citizens must have a better standard of living than most, if not all, South American countries. (The same rationale is sometimes applied to the continent of Africa, the Middle East, and parts of Southeast Asia as well).

As this analysis has shown, these beliefs are incorrect – at least when it comes to suicide rates. Furthermore, the World Population Review's ranking of suicide rates by country in 2021 shows that Argentina's rate is significantly less than in the U.S. or Finland (at 8.4 people per 100,000).

Although Age Group was another treatment factor used, it was investigated less than the Countries factor. One significant finding was that the two factors had an interaction effect. The reasons for this interaction were not investigated in this study. Some speculative reasons for it could be that some countries invest more in education and other youth programs which could lower the suicide rate in those countries for the youth population. Another reason could be that some countries invest more in social security programs for the elderly, which would reduce the suicide rate in the elderly population.

It was shown in the Results section that the equality of variance assumption was violated. The exact reason for this remains unclear but could have to do with the time dependence between observations. The year in which the observations were collected was included in the original data set and this information could be used to account for the dependence. The time (in years) variable could be used as a covariate and the analysis could be adjusted to account for it.