MIDDLE EAST TECHNICAL UNIVERSITY

Department of Statistics–

STAT 364 LINEAR MODELS II PROJECT
"Human Freedom Index"

Contributors

Dilay SEVGİLİ

Ebru ŞAKAR

Gizem Cemile ÇELİK

İrem EKMEKÇİ

TABLE OF CONTENT

1.Abstract

Freedom can be defined differently by individuals. Today, human freedom can be associated with different circumstances such as economic or social. This study analyzes how human freedom can be evaluated globally by various measures. To answer this question, Human Freedom Index is used. This index is a large collection of different measurements of 162 countries. The more details about these countries and their religions are shown in the data description part. The methods of data collection and used statistical test are evaluated in the methodology part. Analysis part includes the following analysis of:

- Human Scores by Regions
- Highest and Lowest Human Freedom Scores of Countries
- Legal Systems and Property Right Scores by Regions
- Correlation Matrix
- Multiple Linear Regression

What can be interpreted as social-economic by looking at analysis results is discussed in the conclusion and discussion part of the study.

2.Introduction

Freedom of human can be explained by different factors with several perceptions. It may have differed for each. To be able to explore how should human freedom evaluate objectively, Human Freedom Index in 2016 is used in this project. Human Freedom Index is the collection of broad measurements about personal and economic freedom with different areas. This large collection is important to understand the relationship between the freedom of individuals and social and economic circumstances. This social and economic phenomenon gathers with 12 main titles in the following:

- Rule of Law
- Security and Safety
- Movement
- Religion
- Association, Assembly, and Civil Society
- Expression and Information

- Identity and Relationships
- Size of Government
- Legal System and Property Rights
- Access to Sound Money
- Freedom to Trade Internationally
- Regulation of Credit, Labour, and Business

Today there are 195 countries according to the United Nations. Human freedom index includes a score of 162 countries about the above areas. Although not all countries are included, these 162 countries give a chance to make an overall evaluation. Further evaluation will be discussed in the analysis part.

3.Literature Review

What is human freedom? Many people think they possess it, but few clearly define it. Throughout history, many philosophers tried to define human freedom. Rousseau, Hobbes, Sartre and John Locke were some of the philosophers who defined human freedom. Some thought free will exists, and some of them did not. For example, the German liberal economist Wilhelm Röpke (1959) said that "Freedom is so valuable that we must be prepared to sacrifice everything for it; even prosperity and opulence when economic freedom constrains us to do so. To our great and undeserved fortune, however, freedom-based economic order which general freedom cannot do without has an incomparable material superiority over an economic order based on force."

Human Freedom Index is a measure of human freedom, understood as the absence of coercive constraint. Human Freedom Index is a degree to which people are free to enjoy the major civil liberties such as freedom of speech, religion, and association and assembly. "The Human Freedom Index measures civil liberties, economic freedom, the rule of law, freedom of movement, women's rights and much more," said Fred McMahon, Dr. Michael A. Walker Research Chair in Economic Freedom at the Fraser Institute and editor of the study.

It is important to measure freedom because it makes more understandable since every people who do not have the same environmental factors as other people will not have the same freedom. These environmental factors can be:

1)Rule of Law

The rule of law is an important condition of freedom that protects the individual from coercion by others. John Locke's emphasis on the importance of law in securing and enlarging freedom is an early formulation of that concept.

2)Security and Safety

The rights to life and safety from physical aggression have long been accepted as fundamental to liberty. The violence of any kind, except in self-defense or the administration of justice, reduces personal freedom and, in the case of violence that results in death, eliminates it.

3)Movement

The freedom to travel may be a basic human right and essential to a free society. Governments that restrict people's movement greatly limit the scope of overall liberty because those limits severely reduce the ability of people to engage in a wide range of peaceful activities of their choosing.

4) Religion

Free societies respect the right to practice a religion of one's selecting. The exercise of religion can be both a supremely private matter involving a person's strongest beliefs and a social affair practiced in an organized way among larger groups.

5)Association, Assembly, and Civil Society

The freedom to associate and assemble with peaceful individuals or organizations of one's choice and to form or join organizations for political, commercial, or other ends is an essential part of individual freedom and a basis of civil society.

6) Expression and Information
**7) Identity and Relationships**
8) Economic Freedom Measures

There is a strong relationship between democracy and freedom. As a result, more democratic countries have more free people. If a country is democratic, it respects people's freedom.

4. Data Description

"The Human Freedom Index" data includes 79 variables from social and economic criteria. These 79 variables are under some main titles which are used in this project. These main titles are defined in the introduction part. To conduct general research, the main titles are used in this project, and others are removed from data. The simpler version of the data includes 17 variables with one response and 162 countries. Also, region proportions of these countries are in Figure 1.
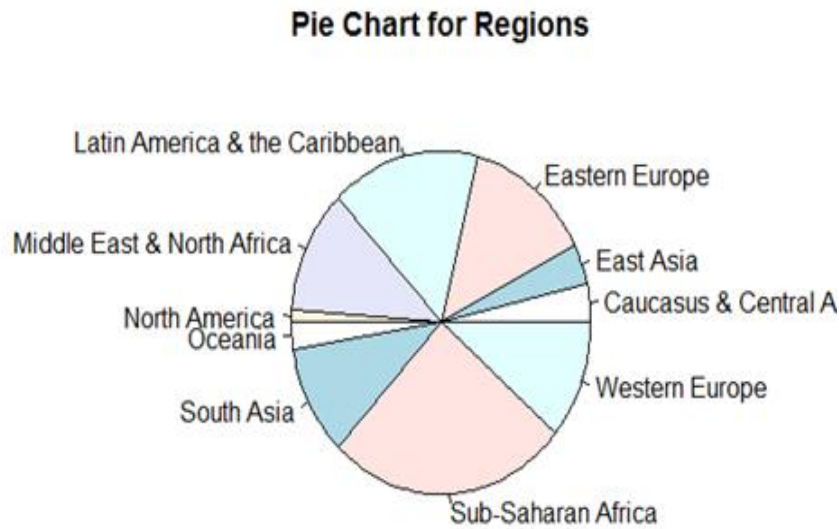
Figure 1: Pie chart for Regions

There are 162 countries with 10 separated regions. To be able to understand these regions pie chart is plotted. These ten regions have different proportions. Sub-Saharan Africa has the largest proportion. The Middle East and North Africa, Latin America and the Caribbean, Western Europe has relatively large proportions. Most of the data values come from these regions. North America has the lowest proportions. Oceania, East Asia, and the Caucasus & Central Asia also has low proportions. This means that the number of countries from this region aren't high.

5.METHODOLOGY

5.1.Data Collection Methods

Researchers focus on the two main criteria when collecting data to make the index reliable and objective. Firstly, the data come from credible external sources such as OECD, World Bank, World Economic Forum. Secondly, the report covers large several countries as possible. When creating the human freedom score, the researchers weight economic and personal freedom in the index equally. Because people's lifestyles are strongly dependent on the economy. Moreover, economic freedom decreases the dependence of individuals on the government which restricts liberty.

To create a personal freedom score, they average the legal protection and security categories and the specific personal freedoms. The legal protection and security made up of Rule of Law and Security and Safety receive half the weight in the personal freedom index. The specific personal freedoms are made up of Movement; Religion; Association, Assembly, and Civil Society; Expression and Information; and Identity and Relationships receive the other half the of weight the personal freedom index. Then, to create an economic freedom score, the weight Size of Government, Legal System and Property Rights, Sound Money, Freedom to Trade Internationally, and Regulation equally. The following table clearly describes the structure of the Human Freedom Index.

PERSONAL FREEDOM	ECONOMIC FREEDOM
1.　　LEGAL PROTECTION AND SECURITY A. Rule of Law B. Security and Safety	A. Size of Government B. Legal System and Property Rights
2.　　SPECIFIC PERSONAL FREEDOMS A. Movement B. Religion C. Association, Assembly, and Civil Society D. Expression and Information E. Identity and Relationships	C. Sound Money D. Freedom to Trade Internationally E. Regulation

Table 1. Structure of the Human Freedom Index

5.2. Data Analysis Methods

1- Correlation Matrix

A correlation matrix is a table showing a correlation between variables. Each coefficient in the table shows the relationship between two variables. Moreover, it is used as a way to summarize data, as an input into a more advanced analysis, and as a diagnostic for advanced analyses. Accordingly, this method is used for investigating the question about the possible relationship between the variables and the level of significance of this relationship in this research.

2-Box Plot

A Box Plot is the visual representation of the statistical five-number summary of a given data set. It is especially useful for indicating whether a distribution is skewed and whether there are any outliers in the data set. Furthermore, the box plot is very effective and easy to read, as they can summarize data from multiple sources and display the results in a single graph. Namely, the box plot is the most appropriate method for the analysis of human scores by regions because of the data includes large numbers of observations, and more than two different categories are being compared.

3-Table

A table is a set of figures arranged in columns and rows which is a very useful way of organizing numerical information or data. Information is displayed as text, using words and numbers, and grid lines may be present or not. Since tables make it easy to compare pairs of related values, the table is the best option for the analysis of the highest and lowest human freedom scores of countries. Another reason for that, it will be used to compare individual values, but not an entire series of values to one another and, precise values are required.

4-Pie Chart

A pie chart is a circle divided into segments that shows the contribution of different categories to an overall total. Each slice of the circle represents percentages. In this manner, the pie chart is used to clarify the data in the region proportions of countries.

5- Shapiro-Wilk test

The Shapiro-Wilk test is a test that tells whether a random sample comes from a normal distribution or not. Similarly, this method is used in the research for checking the normality of the human freedom score, legal system, and property rights score and assumption of multiple linear regressions.

6-One-way ANOVA

The one-way analysis of variance is a method that determines whether there are any statistically significant differences between the means of three or more unrelated groups. In this research, this method is the most appropriate one for analysis of the legal system and property rights score by regions because the model validates all assumptions of ANOVA. The first assumption is that all populations involved follow a normal distribution. The second one is that all populations have the same variance. The last one is that the samples are randomly selected and independent of one another.

7-Bartlett's test

Bartlett's test is used for testing homogeneity of variances in
k samples, where k can be more than two. It's adapted for normally distributed data. Hereby, this method is used for checking assumptions of ANOVA.

8- Tukey HSD test

The Tukey Test (or Tukey procedure), also called Tukey's Honest Significant Difference test, is a post-doc test based on the studentized range distribution. An ANOVA test tells whether or not the results are significant; however, to find out which specific groups' means are different Tukey Test should be used because the test compares all possible pairs of means. Accordingly, this method is used after ANOVA to compare the mean of all regions of the legal system and property right score.

9-Tukey's Ladder of Powers transformation

The Tukey ladder of powers (sometimes called the Bulging Rule) is a way to change the shape of a skewed distribution so that it becomes normal or nearly normal. It can also help to reduce error variability (heteroscedasticity). As human freedom score, which is the response variable has to become normal, this method is used. Other transformation methods did not work; Tukey's Ladder of Powers transformation is the most appropriate one.

10-Multiple Linear Regression

Multiple Linear Regression which is the most common form of linear regression analysis is used to explain the relationship between one continuous dependent variable and two or more independent variables. In this research, since the response variable is continuous and there is more than one variable, multiple linear regression is used.

11-Histogram

The histogram is used for the continuous type of random variables. The histogram is used to find out the shape of the distribution of the variable of interest and to detect the outliers if it exists. In this project, the histogram is used to show the residuals are normally distributed because one of the assumptions of constructing a linear regression model is residuals are normally distributed.

12- Chi-Square Test

The Chi-Square test is used to tell whether there are any relationships between categorical variables. The test is applied to check the independence of the Legal System and Property Right Scores by Regions.

13-Variable Selection

Variable selection is intended to select the "best" subset of predictors. Variable selection can be done in two ways.

- All Possible Regression

Assume the intercept term is in all equations considered. Then, if there are K regressors, there would be 2^K possible regression equations to investigate.

- Stepwise Regression Methods

→Forward Selection: This procedure begins with no regressors in the model other than the intercept. Adding a regressor at a time model conducted.

→Backward Selection: This procedure begins with all regressors in the model. Removing a regressor at a time model conducted.

→Stepwise Regression: This procedure is a modification of forwarding selection.

To find the best model which, explain data in an appropriate way, the variable selection method is used.

6. DATA ANALYSIS

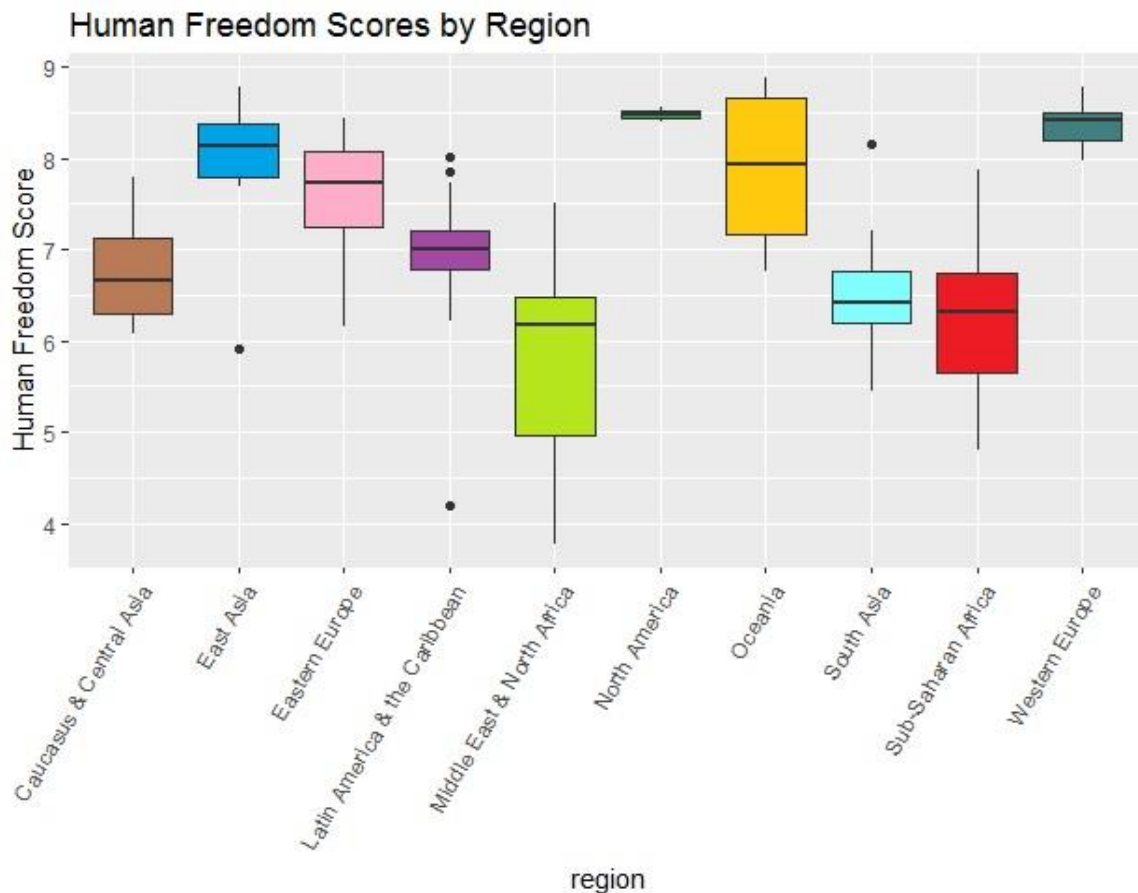## 6.1. Analysis of Human Scores by Regions

Figure 2: Box Plot of Human Freedom Score by Regions

Middle-East and Oceania have the largest spread. The variation in human freedom scores is higher than the others. While Oceania region distributed approximately normal, freedom scores of Middle-East region are distributed left-skewed. Overall freedom rates for Oceania are much higher than Middle-East. Actually, Middle-East overall scores are lower than all other region's rates. North America and Western Europe have the highest rates of freedom while their spread is less because people have approximately the same level of freedom. People have lower freedom scores with higher variety in the east of Europe than West Europe has. If we go through Central Asia to South Asia, the scores are getting lower and lower for most of the Asian people. The distribution of Central Asia countries looks normal and balanced while South Asia countries' distribution looks right-skewed.

6.2. Analysis of Highest and Lowest Human Freedom Scores of Countries

Country	Human Freedom Score
1)New Zealand	8.887
2)Switzerland	8.787
3)Hong Kong	8.776
4)Australia	8.582
5)Canada	8.565
6)Netherlands	8.554
7)Denmark	8.547
8)Ireland	8.504
9)United Kingdom	8.497
10)Finland	8.472

Table 1 The Best 10 Countries For Human Freedom

New Zealand has the highest human freedom score, but there is no considerably big difference between these countries according to table 1.

Country	Human Freedom Score
1)Syria	3.765
2)Venezuela	4.200
3)Yemen, Rep.	4.253
4)Iraq	4.258
5)Libya	4.310
6)Sudan	4.803
7)Egypt	4.807
8)Algeria	5.135
9)Burundi	5.167
10)Iran	5.281

Table 2 The Worst 10 Countries For Human Freedom

There is a little difference between these countries. The lowest score between all countries comes from Syria according to table 2.

6.3. Analysis of Legal Systems and Property Right Scores by Regions

ANOVA is the most appropriate one for analysis of the legal system and property rights score by regions because the model validates all assumptions of ANOVA. Firstly, normality is tested with Shapiro-Wilk test and, it is found that ($w = 0.97885$) $p = 0.01389$ that is greater than 0.01 which shows null hypothesis is not rejected where the null hypothesis is that the legal system and property rights scores are normally distributed. Secondly, the variance is tested with Bartlett test and, it is found that $p = 0.1373$ that is greater than 0.01 which shows null hypothesis is not rejected where the null hypothesis is that the batch variances are all equal. Lastly, the independence of samples tested with the chi-square test and, it is found that $p = 0.4246$ is greater than 0.01(accept Ho). As a result, there are statistically significant differences between group means as determined by one-way ANOVA (F(9,152) = 18.37, p <2e-16).
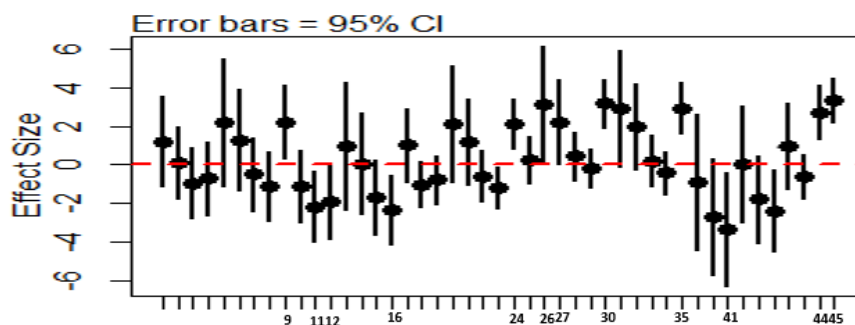

Figure3 Plot of Tukey

9: Western Europe-Caucasus & Central Asia
11: Latin America & Caribbean-East Asia
12: the Middle East & North Africa-East Asia
16: Sub-Saharan Africa-East Asia
24: Western Europe-Eastern Europe
26: North America-Latin America & the Caribbean
27: Oceania-Latin America & the Caribbean
30: Western Europe-Latin America & the Caribbean
35: Western Europe-Middle East & North Africa
38: Sub-Saharan Africa-North America
41: Sub-Saharan Africa-Oceania
44: Western Europe-South Asia
45: Western Europe-Sub-Saharan Africa

After ANOVA, Tukey Test is applied to find out which specific groups' means are different. Although Tukey HSD test indicates that there is no statistically significant difference between legal system and property rights score mean of western Europe, Oceania, East Asia, and North America, Western Europe has the best legal score mean.

Additionally, North America follows western Europe. The third best mean belongs to people of

Oceania and, there is a statistically significant difference between legal system and property rights score mean of Oceania and the Middle East & North Africa, Sub-Saharan Africa, Latin America & the Caribbean. Unfortunately, the worst mean belongs to people of Sub-Saharan Africa. In fact, the test indicates that there is no statistically significant difference between the legal system and property rights score mean of Sub-Saharan Africa and South Asia, Caucasus & Central Asia, Middle East & North Africa, Latin America & the Caribbean.
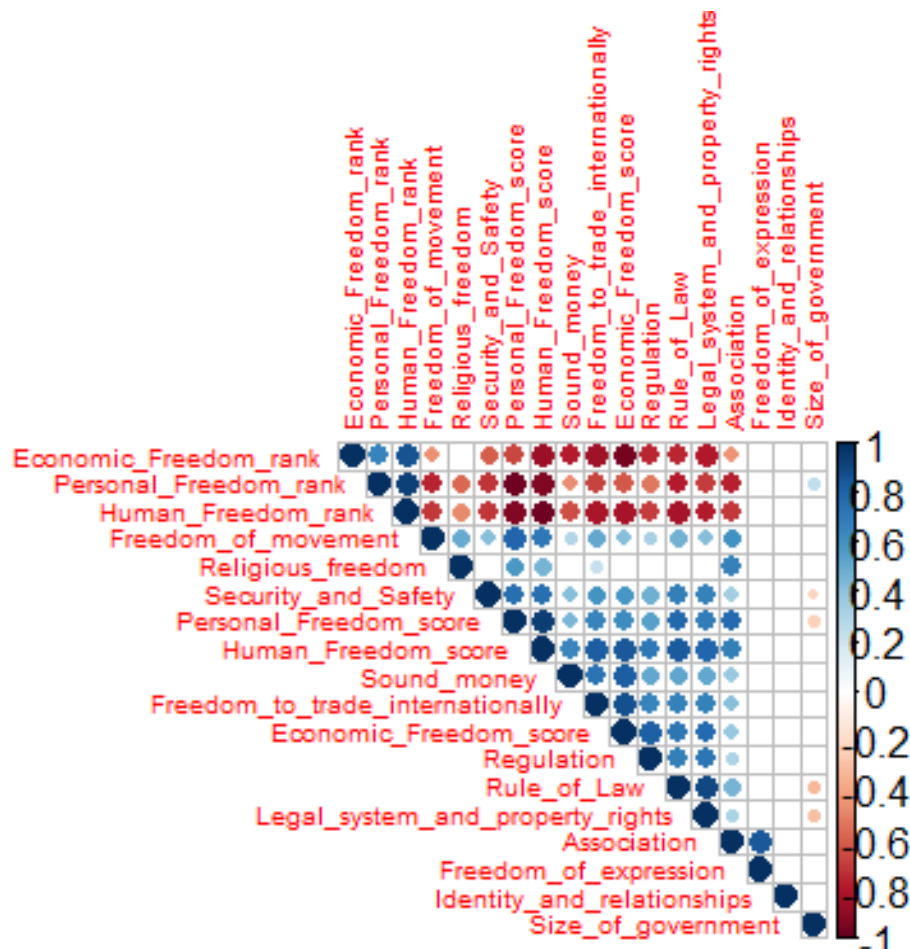
6.4. Analysis of Correlation Matrix

Figure3: Correlation matrix of variables

To see the relationship between variables, correlogram is chosen to use. Positive correlations are displayed in blue and negative correlations in red color. The color intensity and the size of the circle are proportional to the correlation coefficients. In the right side of the correlogram, the legend color shows the correlation coefficients and the corresponding colors. In the above plot, correlations with p-value > 0.01 are considered as insignificant. In this case, the correlation coefficient values are left blank.

6.5. Analysis of Multiple Linear Regression

To conduct regression, all variables are used, and Human Freedom Score is chosen as a response variable. Since the response variable is continuous and there is more than one variable, it is expected to conduct multiple linear regression.

To conduct linear regression below assumptions are checked.

- Normally distributed residuals
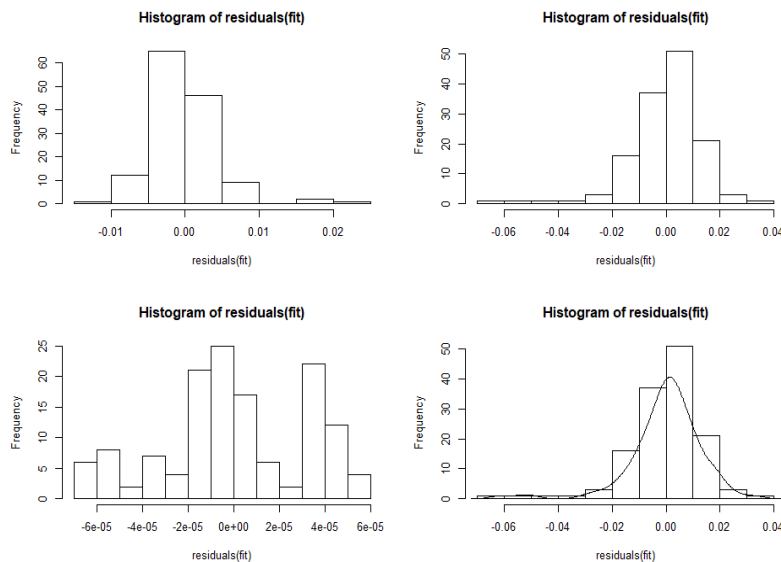
Normally distributed residuals assumption is checked with residuals' histogram. As it is shown in the histogram graph residuals are not distributed normally. To be sure about normality it is tested with the Shapiro-Wilk test. It is found that $p < .001$ that is smaller than 0.01 which shows the null hypothesis is rejected where the null hypothesis is that residuals are normally distributed. To be able to solve the non-normality problem, transformation is required. To figure out what kind of transformation is needed on the response, Tukey's Ladder of Transformation is used. The result of this transformation is that lambda equals 1.5. How to decide using this lambda is shown above and it is taken from R:

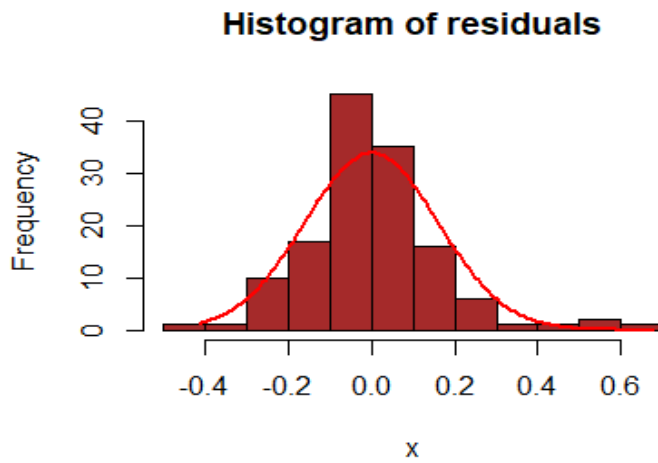if (lambda $>$ 0){TRANS $=$ x $\wedge$ lambda}

if (lambda $==$ 0){TRANS $=$ log(x)}

if (lambda $<$ 0){TRANS $=$ -1 * x $\wedge$ lambda}

Since lambda value is greater than zero, it is decided to use y'=y $\wedge$ (1.5) transformation on response value.

Histogram of residuals

After transformation, residuals fit the normality line at is shown in the graph. Moreover, it is tested with the Shapiro-Wilk test and it is founded that $p=0.002$. Even though it is not satisfied now, after variable selection, it can be better. Therefore, this assumption is acceptable considering selection techniques.

- Non-constant variability in residuals
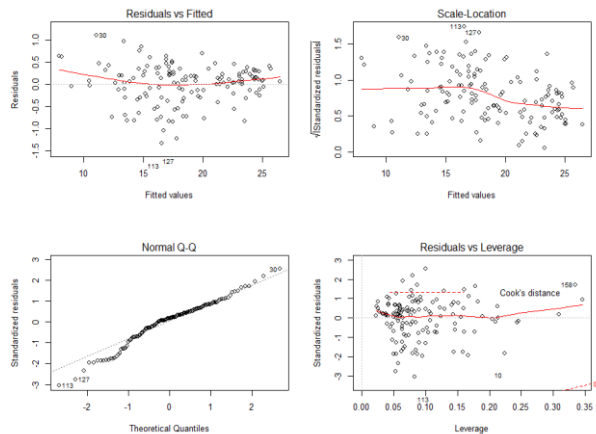
There is no specified pattern in the graph. Thus, it can be said that the non-constant variability assumption is valid.

- Multicollinearity

Since Personal Freedom Score and Economic Freedom score were calculated based on other variables as explained in data collection methods in Methodology, these variables are removed of data to solve the multicollinearity problem. The first VIF values are all greater than 10. It was the reason why Personal Freedom Score and Economic Freedom Score were removed. After these variables removed, all VIF values can be seen below:

Name of Variable	VIF value
Rule of Law	8.180050
Security and Safety	2.560253
Freedom of Movement	2.109872
Religious Freedom	2.124715
Identity and Relationship	1.078688

Freedom of Expression	5.098991
Legal System and Property Rights	7.873718
Sound Money	2.574123
Freedom to Trade Internationally	4.194534
Regulation	2.892001
Association	4.301480
Size of Government	1.465229

Tablo 4 VIF values of variables

As it is seen in the table, all VIF values are less than ten which means there is no multicollinearity problem. Thus, multicollinearity assumption is valid.

Variable Selection

The full model has 12 variables with one response which is Human Freedom Score. To find the best model which, appropriately explain data, variable selections are needed after checking assumptions. In order to find the best model, 3 of selection techniques are used. First, to be able to see which subset the best is in their own, it is applied Best Subset regression. This technique gives the chance to see their C(p), AIC and other values to decide the best model. The results of the best subset regression are below:

Model	R-Square	R-Square Adj	C(p)	AIC	SBIC	SBC	MSEP
1	0.7217	0.7200	3603.6953	722.1031	256.7078	731.3659	4.9896
2	0.8638	0.8618	1501.9319	519.3907	126.2589	531.0413	2.6301
3	0.9304	0.9288	706.2814	430.1416	36.0934	444.7049	1.3651
4	0.9486	0.9470	490.2550	390.9459	-4.0132	408.4218	1.0239
5	0.9612	0.9597	341.4385	354.7733	-40.6736	375.1619	0.7853
6	0.9727	0.9714	205.4247	308.9391	-85.8807	332.2403	0.5611
7	0.9787	0.9776	135.0736	276.9791	-116.8771	303.1930	0.4440
8	0.9832	0.9821	83.5678	246.9611	-145.0570	276.0877	0.3564
9	0.9874	0.9865	34.8358	209.5301	-178.4077	241.5693	0.2710
10	0.9893	0.9885	14.0185	189.2238	-195.3637	224.1756	0.2337
11	0.9897	0.9888	11.3303	186.1498	-197.3571	224.0143	0.2288
12	0.9897	0.9887	13.0000	187.7851	-195.4406	228.5622	0.2319

Tablo 5 Best Subsets Regression Summmary

The 11th model seems the best in all 12 models. Since its C(p) values really close to the number of its parameter, R-Square adjusted is the highest for 11th model and its MSEP, AIC and BIC values are the smallest. It is decided to use the 11th model which has variables following:

- ✓ Rule of Law

- ✓ Security and Safety

- ✓ Freedom of movement

- ✓ Religious freedom

- ✓ Association

- ✓ Freedom of expression

- ✓ Size of government

- ✓ The legal system and property rights

- ✓ Sound money

- ✓ Freedom to trade internationally

- ✓ Regulation

To be sure that the 11th model also the best for other techniques, Forward Selection, and Stepwise regression also applied. Final model results of both techniques are below:

Forward Selection	Stepwise Selection
1 Rule_of_Law	1 Rule_of_Law
2 Freedom_of_movement	2 Freedom_of_movement
3 Freedom_to_trade_internationally	3 Freedom_to_trade_internationally
4 Association	4 Association
5 Security_and_Safety	5 Security_and_Safety
6 Size_of_government	6 Size_of_government
7 Legal_system_and_property_rights	7 Legal_system_and_property_rights
8 Religious_freedom	8 Religious_freedom
9 Sound_money	9 Sound_money
10 Regulation	10 Regulation
11 Freedom_of_expression	11 Freedom_of_expression

Tablo 6 Selected variables

As it is seen above, selected all variables is the same as the 11th model. Thus, it is decided to 11th model is the best. After conducting the 11th model all p-values of coefficients are less than .01 and R-Square is 0.9897 which means %98.97 of Human freedom score can be explained by variables. To understand how well perform our data, it is applied cross-validation.

Cross-Validation

To apply cross validation, data is split into trains set and test set. %70 of data is used for conducting the model as a training set and remained part is used for testing the model. After getting the model from variable selection as the main model, the model is evaluated on the test set. To be able to decide how the model performs well-predicted values from the training set and values from the test set was compared with MSE. *MSE* value is equal to 0.2015 which means very small value. It can be said that the predictive performance of the model is high.

Final Model

A multiple linear regression model is calculated to predict Human Freedom Score based on Rule of Law, Security, and Safety, Freedom of movement, Religious freedom, Association, Freedom of expression, Size of Government, Legal system and property rights, Sound money, Freedom to trade internationally, Regulation. A linear regression was found ($F(df=124)1084$,*p*< 2.2e-16) with an *R-square*=0.9897. All VIF values are less than ten means there is no multicollinearity between variables.

COEFFICIENTS	ESTIMATE
(Intercept)	-8.61361
Rule_of_Law	0.39838
Security_and_Safety	0.61425
Freedom_of_movement	0.24894
Religious_freedom	0.29167
Association	0.25083
Freedom_of_expression	0.12572
Size_of_government	0.33323
Legal_system_and_property_rights	0.55854
Sound_money	0.32270
Freedom_to_trade_internationally	0.43909
Regulation	0.32160

Tablo 7 Coefficients of final model

7. DISCUSSION

It is time to discuss the results of the findings of this research and think about the reasons behind them. First of all, The difference in human freedom scores between the regions which is derived from the circumstances of the regions is observed. There is a strong correlation between the economy and freedom, but there are still some people who aren't free despite having economic power. It is a reality that freedom is also related to the governments' political agenda. For example, Overall rates for middle east are really low in terms of freedom although there is a huge gap between rich and poor people's freedom rates in some region in the middle east. Thus, rich people are economically free in the middle east, they are relatively free according to poor people there, but actually, they aren't free at all comparing by other regions all over the World. Furthermore, citizens' freedom rates are dramatically getting low in the countries face with war, military coup and civil war. These affect countries' economies very bad. Whereat, Other aspects of freedom are also restricted directly. For example, personal freedom issues such as security and safety, the rule of law that protects individual from the coercion of others, the freedom to travel, self-expression and getting information are damaged. Governments that restrict people's movement greatly limit the comprehension of overall freedom in a state of emergency. Another issue that affects freedom is religion. The freedom scores are low in the countries governed by religious rules because governments force people to do whatever the so-called religion says. Even if the religion actually doesn't say so, they use even violence or police to control people. In such countries, it is hard to talk about personal freedoms and human rights. The attitudes of them harm people's identity and the freedom of expressing themselves. In addition, Freedom is under attack in even developed due to some ideologies such as populism that target minorities and support increasing police power. However, The agenda of developed countries are much more stable, so writing optimistic comments are easier because of their powerful economy and strong regal systems which give importance to human rights. The majority of people there reach and share information without coercion more and willing to be part of civil society in a kind way, so these attitudes support freedom. A democratic and peaceful environment is created by the way. People aren't be restricted to move to another place in or outside of their countries by the law. In such a population, people can express themselves freely and show respect to different ideas. People are also not exposed to any religious idea's pressure. Thus, fundamental liberty is satisfied.

8. CONCLUSION

This research helps us to compare the types and levels of freedoms such as economic and personal region by region in the world. Then, we can use the results to detect the barriers to freedom and find a solution to worldwide problems. Mainly, human scores by region, highest and lowest human freedom scores of countries, legal system and property right scores by regions are analyzed. 12 main titles are gathered such as rule of law, security, and safety, movement, etc. In that way, It is revealed that the freedom of individuals and social and economic circumstances. In the literature review part, the definition of the freedom and 12 main titles mentioned before is clarified. Then, The data and methodology are explained. The aspects of data collection and analysis methods that are used in the research is defined clearly. In the data analysis part, The data is visualized and explained in details. There is a huge difference between developed and developing countries' human freedom scores. While the developed countries' region has balanced and higher scores among people, so their plots have less spread, people in developing countries have low scores and does not distribute equally among people. Unfortunately, The same situation is observed when the legal systems and property right scores are compared by regions. Moreover, A correlogram is used to reveal the negative and positive correlation between the variables. By the way, it is easily observed that there is a positive correlation between human freedom score and all of these titles such as sound money, freedom to trade internationally, economic freedom score, regulation, rule of law, legal system and property rights and association at first look. Afterward, multiple linear regression analysis was performed to add depth to the study. At the first non-normal residuals and multicollinearity problems were needed to be solved. In order to solve these problems, transformation and removing variables were applied. To find the best model variable selection model is used, and the model which has all significant variables is needed to be validated. To understand how the model performs well Cross Validation is used. Since the model performs well, it is decided to use as the final model. Finally, The reasons behind the results are discussed in the last part.

REFERENCES

- Fraserinstitute.org. (2019). [online] Available at: https://www.fraserinstitute.org/sites/default/files/human-freedom-index-2016-intl-news-release.pdf [Accessed 2 May 2019].

- Ian Vásquez and Tanja Porčnik, *The Human Freedom Index 2018: A Global Measurement of Personal, Civil, and Economic Freedom* (Washington: Cato Institute, Fraser Institute, and the Friedrich Naumann Foundation for Freedom, 2018)

- Röpke, Wilhelm (1959). Erziehung zur wirtschaftlichen Freiheit. In A. Hunold (ed.), Erziehung zur Freiheit, Zürich.

APPENDIX

R_CODES

```
data <- read_csv("C:/Users/gonul/Desktop/veri364 - Çalışma Sayfası1.csv")
View(data)
data1=round(data[,3:20],digits = 4)
data=cbind(data[,1:2],data1)
```

Analysis of Human Scores by Regions Codes

```
ggplot(data,aes(region,Human_Freedom_score))+geom_boxplot()+labs(title ="Human Freedom Scores by Region",y="Human Freedom Score")+scale_fill_gradient2()+ theme(axis.text.x = element_text(angle = 60, hjust = 1))
```

Analysis of Highest and Lowest Human Freedom Scores of Countries Codes

```
newdata <- data[order(Human_Freedom_score,countries),]
newdata=newdata[,c(1,19)]
newdata[,2]=round(newdata[,2],digits = 1)
newdata=newdata[1:10,]
```

Analysis of Legal System and Property Right Scores by Regions Codes

```
keeps <- c("Legal_system_and_property_rights","region") #independence
master = master_all[keeps]
str(master)
summary(master)
mytable <- table(master)
chi2 <- chisq.test(mytable, correct=TRUE)
shapiro.test(Legal_system_and_property_rights)  # normality check
bartlett.test(Legal_system_and_property_rights ~ region, data=data) #variance check
tukey.test=TukeyHSD(aov.out,conf.level = 0.99)
aov.out = aov(Legal_system_and_property_rights ~ region, data=data)
```

```r
summary(aov.out) #ANOVA
plotTukeyHSD <- function(tukey.out,   #plotTukeysHSD() Code
x.axis.label = "Comparison",
y.axis.label = "Effect Size",
axis.adjust = 0,
adjust.x.spacing = 5){
tukey.out <- as.data.frame(tukey.out[[1]])
means <- tukey.out$diff
categories <- row.names(tukey.out)
groups <- length(categories)
ci.low <- tukey.out$lwr
ci.up  <- tukey.out$upr
n.means <- length(means)
#determine where to plot points along x-axis
x.values <- 1:n.means
x.values <- x.values/adjust.x.spacing
# calculate values for plotting limits
y.max <- max(ci.up) +
max(ci.up)*axis.adjust
y.min <- min(ci.low) -
max(ci.low)*axis.adjust
if(groups == 2){ x.values <- c(0.25, 0.5)}
if(groups == 3){ x.values <- c(0.25, 0.5,0.75)}
x.axis.min <- min(x.values)-0.05
x.axis.max <- max(x.values)+0.05
x.limits <- c(x.axis.min,x.axis.max)
#Plot means
```

```r
plot(means ~ x.values,
xlim = x.limits,
ylim = c(y.min,y.max),
xaxt="n",xlab = "",ylab = "",cex = 1.25,pch = 16)
axis(side = 1,at = x.values,
labels = categories)
#Plot upper error bar
lwd <- 2
arrows(y0 = means,
x0 = x.values,
y1 = ci.up,x1 = x.values,
length = 0,
lwd = lwd.)
#Plot lower error bar
arrows(y0 = means,x0 = x.values,
y1 = ci.low,
x1 = x.values,
length = 0,
lwd = lwd.)
#add reference line at 0
abline(h = 0, col = 2, lwd = 2, lty =2)
mtext(text = x.axis.label,side = 1,line = 1.75)
mtext(axis.text.x = element_text(angle = 90, hjust = 1))
mtext(text = y.axis.label,side = 2,line = 1.95)
mtext(text = "Error bars = 99% CI",side = 3,line = 0,adj = 0)
}
plotTukeysHSD(tukey.test)
```

Analysis of Correlation Matrix Codes

```r
res2 <- rcorr(as.matrix(data))

corrplot(res2$r, type="upper", order="hclust",p.mat = res2$P, sig.level = 0.01, insig = "blank", tl.cex = 0.55/par("cex"),cl.cex = 1/par("cex"))
```

Multiple Linear Regression Codes

```r
T_tuk = transformTukey(Human_Freedom_score,plotit=FALSE)  #transformation
shapiro.test(Human_Freedom_score^(1.5))                    #normality check
fit=lm((Human_Freedom_score^1.5)~Rule_of_Law+Security_and_Safety+Free-
```

```
dom_of_movement+Religious_freedom+Association+Freedom_of_expression+Iden-
tity_and_relationships+Size_of_government+Legal_system_and_prop-
erty_rights+Sound_money+Freedom_to_trade_internationally+Regulation,data=data)
```

summary(fit)

mse=(test.set$Human_Freedom_score-pred)^2

pred=predict(model,test.set)

summary(model)

```
model=lm(Human_Freedom_score~Rule_of_Law+Security_and_Safety+Freedom_of_movement+
Religious_freedom+Association+
Freedom_of_expression+Size_of_government+Legal_system_and_property_rights+Sound_money+Fr
eedom_to_trade_internationally+Regulation,data=train.set)
```

```
fit=lm(Human_Freedom_score^(1.5)~Rule_of_Law+Security_and_Safety+Freedom_of_movement+
Religious_freedom+Association+
Freedom_of_expression+Identity_and_relationships+Personal_Freedom_score+Size_of_government
+Legal_system_and_property_rights+Sound_money+Freedom_to_trade_internationally+Regulation+
Economic_Freedom_score,data=data)
```

summary(fit)

```
fit2=lm(Human_Freedom_score^(1.5)~Rule_of_Law+Security_and_Safety+Freedom_of_movement+
Religious_freedom+Association+
Freedom_of_expression+Identity_and_relationships+Size_of_government+Legal_system_and_proper
ty_rights+Sound_money+Freedom_to_trade_internationally+Regulation,data=data)
```

summary(fit2) after transformation

plot(fit2)

layout(matrix(c(1,2,3,4),2,2)) # optional 4 graphs/page

```
x=residuals(fit2)
        lines(xfit, yfit, col="red", lwd=2)
```

yfit <- yfit*diff(h$mids[1:2])*length(x)

yfit<-dnorm(xfit,mean=mean(x),sd=sd(x))

xfit<-seq(min(x),max(x),length=100)

main="Histogram of residuals")

h<-hist(x, breaks=9, col="brown",)

ols_step_best_subset(fit2)

ols_step_both_p(fit2)

ols_step_forward_p(fit2)

```
fit3=lm(Human_Freedom_score^(1.5)~Rule_of_Law+Security_and_Safety+Freedom_of_movement+
Religious_freedom+Association+
Freedom_of_expression+Size_of_government+Legal_system_and_property_rights+Sound_money+Fr
eedom_to_trade_internationally+Regulation,data=data)      #final model after selection
```

```r
summary(fit3)

#cross validation

test.set=data[trainingRowindex,]

train.set=data[trainingRowindex,]

trainingRow=sample(1:nrow(data),0.7*nrow(data))

data$Human_Freedom_score=data$Human_Freedom_score^(1.5)

createDataPartition(p = 0.8, list = FALSE)

training.samples <- data$Human_Freedom_score^(1.5)

training.sample=data$Human_Freedom_score^(1.5)

MSE=mean(mse, na.rm = T)
```