**A close up of a sign

Description automatically generated**

**INE2002 Statistics in Engineering**

**Term Project Report**

**2020 Summer**

Onur Güzel

September 11, 2020

**Abstract**

With the progress in our century access to almost anything we can image is getting easier day by day. While we think that this good for the sake of our civilization there is certain drawbacks, one of them is drug use. Drug use around the world is in an increasing trend for decades. Governments around the world spending billions of dollars and significant resources to stop this epidemic. While drug use is often associated with crime, poverty, and terror, there is another question that is vital to ask. Is there a relation between the type of regime within a country and drug usage of its population? The some governments trying to combat this epidemic is one of the same reasons behind this drastic increase? This project will be exploring the relationship between drug use disorder within a country and its democracy index.

**Table of Contents**

**Data Collection**………………………………………………………………………………….………………..…………............. 3

**Data Set**………………………………………………………………………………….……………………………..……………............ 3

**Data Analysis**………………………………………………………………………………….…………...………………...………... 4

Scatter Plot………………………………………………………………………………….……………………………..……………............ 4

Box Plot ………………………………………………………………………………….……………………………..………………............. 5

Histogram………………………………………………………………………………….……………………………..……………….......... 6

Q-Q Plot………………………………………………………………………………….……………………………..………………............. 7

**Hypothesis Tests**………………………………………………………………………………….……………………………….... 8

Hypothesis I………………………………………………………………………………….……………………………..………………...... 8

Hypothesis II………………………………………………………………………………….……………………………..……………........ 9

**Goodness of Fits Tests**……………………………………………………………………………………..………………...... 9

Shapiro-Wilk Normality Test…………………………………………………….……………………………..……………….......... 9

Anderson-Darling Test………………………………………………………………………………….……………………….............. 10

**Maximum Likelihood Estimator**…………………………………….……………………………..……….……. 10

Democracy Index…………………………………………………………………….……………………………..………………............ 10

Drug Use………………………………………………………………………………….……………………………..………………........... 11

**Correlation Tests**……………………………………………………………….……………………………..………………...... 12

Pearson’s Correlation Test………………………………………………………….……………………………..………………....... 12

Kendall’s Rank Correlation Test………………………………………………….……………………………..………………..... 13

**Linear Regression**……………………………………………...…………...……………………….…..………...…………….. 13

**Conclusion**………………………………………………………………..…………….………………….…………..……………….. 14

**Bibliography**……………………………………………………….……………………………..……………….............................. 15

# Data Collection:

Dataset used in this project is compilation of two different sources. Democracy Index and Population Drug Disorder Percentage data is compiled in a single excel file and the sample “CountryData” is obtained using “*Stratified Sampling*”. The population is divided in to four stratums and then every 3rd element is picked in order to form the sample.

# Data Set:

In order to run this regression, there are two datasets needed – one including Democracy Dndex data and the other one is the Drug Use data. Both data sets I have chosen are from 2017 for consistency. Compiled dataset consists of 40 countries in total and the complete list of the countries are shown below:

|  |  |  |  |
| --- | --- | --- | --- |
| Country | Regime Type | Democracy Index[[1]](#footnote-1) | Population Drug Disorder Percentage[[2]](#footnote-2) |
| Iceland | Full Democracy | 9,58 | 0,73 |
| New Zealand | Full Democracy | 9,26 | 0,6 |
| Denmark | Full Democracy | 9,22 | 0,9 |
| Canada | Full Democracy | 9,15 | 2,28 |
| Australia | Full Democracy | 9,09 | 2,32 |
| Netherlands | Full Democracy | 8,89 | 0,99 |
| Germany | Full Democracy | 8,61 | 0,88 |
| Austria | Full Democracy | 8,42 | 0,9 |
| Uruguay | Full Democracy | 8,12 | 0,92 |
| Spain | Full Democracy | 8,08 | 1,47 |
| United States of America | Flawed Democracy | 7,98 | 3,45 |
| France | Flawed Democracy | 7,80 | 1,18 |
| Israel | Flawed Democracy | 7,79 | 0,72 |
| Greece | Flawed Democracy | 7,29 | 0,55 |
| India | Flawed Democracy | 7,23 | 0,53 |
| Argentina | Flawed Democracy | 6,96 | 1,01 |
| Brazil | Flawed Democracy | 6,86 | 1,06 |
| Ghana | Flawed Democracy | 6,69 | 0,63 |
| Mexico | Flawed Democracy | 6,41 | 0,82 |
| Singapore | Flawed Democracy | 6,32 | 0,91 |
| Georgia | Hybrid Regime | 5,93 | 0,49 |
| Ukraine | Hybrid Regime | 5,69 | 0,71 |
| Macedonia | Hybrid Regime | 5,57 | 0,5 |
| Nepal | Hybrid Regime | 5,18 | 0,5 |
| Turkey | Hybrid Regime | 4,88 | 1,54 |
| Bosnia and Herzegovina | Hybrid Regime | 4,87 | 0,43 |
| Lebanon | Hybrid Regime | 4,72 | 2,23 |
| Pakistan | Hybrid Regime | 4,26 | 0,66 |
| Armenia | Hybrid Regime | 4,11 | 0,5 |
| Iraq | Hybrid Regime | 4,09 | 1,57 |
| Venezuela | Authoritarian | 3,87 | 0,65 |
| Egypt | Authoritarian | 3,36 | 1,47 |
| Qatar | Authoritarian | 3,19 | 2,02 |
| Russia | Authoritarian | 3,17 | 0,91 |
| China | Authoritarian | 3,10 | 1,11 |
| United Arab Emirates | Authoritarian | 2,69 | 2,92 |
| Iran | Authoritarian | 2,45 | 1,97 |
| Turkmenistan | Authoritarian | 1,72 | 0,52 |
| Saudi Arabia | Authoritarian | 1,93 | 1,71 |
| Syria | Authoritarian | 1,43 | 1,51 |

# Data Analysis:

Throughout R code Data Set will be referred as “CountryData”, “DemocracyIndex” will be referred as “DemocracyIndex” and “Population Drug Disorder Percentage” will be referred as “DrugUse”.

Throughout the analysis confidence level is chosen as 95% to interpret the test results.

* DemocracyIndex<-(CountryData$DemocracyIndex)
* DrugUse<-(CountryData$PopulationDrugDisorderPercentage)
* summary(CountryData)

A screenshot of a cell phone

Description automatically generated

In addition to these values standard deviation and variance calculation can be seen below;

* sdDemocracy<- sd(DemocracyIndex)
* sdDrugUse<- sd(DrugUse)
* varDemocracy<- var(DemocracyIndex)
* varDrugUse<- var(DrugUse)

A screenshot of a cell phone

Description automatically generated

Further analysis about the data at hand will be handled with visuals such as; boxplots, plots, histograms, Q-Q plots etc. After the visualizations various test will be applied to the dataset in order to determine whether there is a correlation between “DemocracyIndex” and “DrugUse”.

**Scatter Plot:**

Points of observations of each data column can be seen below;

A screenshot of a cell phone

Description automatically generated

* plot(DemocracyIndex)

A close up of text on a white background

Description automatically generated

* plot(DrugUse)

A close up of a map

Description automatically generated

* plot(DemocracyIndex, DrugUse)

**BoxPlot:**

* boxplot(DemocracyIndex)

A screenshot of a social media post

Description automatically generated

* boxplot(DrugUse)

A screenshot of a cell phone

Description automatically generated

* boxplot(DemocracyIndex, DrugUse)

A picture containing screenshot

Description automatically generated

**Histogram:**

Separate histogram visualization can be seen below;

* hist(DemocracyIndex)

A screenshot of a cell phone

Description automatically generated

* hist(DrugUse)

A close up of a logo

Description automatically generated

**Q-Q Plot:**

Q-Q plots in R is pretty useful to identify distributions with the lines and firstly with qqnorm() function the data observations are printed and the line seen in the second plot is achieved with qqline() function and as seen below for both of the columns the line of the normal distribution drawn on the table.

**Q-Q Plot**

* qqnorm(DemocracyIndex)

A screenshot of a cell phone

Description automatically generated

* qqnorm(DrugUse)

A close up of a map

Description automatically generated

* qqline(DemocracyIndex)

A close up of a map

Description automatically generated

* qqline(DrugUse)

A close up of a map

Description automatically generated

**Hypothesis Tests**

**Hypothesis I:**

If the chi square test is applied without any other parameters, null hypothesis will assume that samples are not related. Therefore, first hypothesis tests the correlation between the samples assuming that they are not related.

* chisq.test(DemocracyIndex, DrugUse)

A picture containing knife, table

Description automatically generated

Null hypothesis was that to assume that price and calories are not related. P value is greater that 0.05 (0.2522 > 0.05) therefore we fail to reject the null hypothesis. Failing to reject the null indicates that our sample did not provide sufficient evidence to conclude that the effect exists. However, at the same time, that lack of evidence doesn’t prove that the effect does not exist.

**Hypothesis II:**

Goal is to find if lower “DemocracyIndex” (0-6) Countries have equal mean of “DrugUse” compared to higher Democracy Index (6-10) Countries.

* LowDemocracy<-CountryData[CountryData$DemocracyIndex<=5.99,]
* HighDemocracy<-CountryData[CountryData$DemocracyIndex>=6,]
* Low<-LowDemocracy[,"DemocracyIndex"]
* High<-HighDemocracy[,"DemocracyIndex"]
* t.test(High,Low, alternative = "two.sided" ,mu=0,conf.int=0.95)

A picture containing bird, flower

Description automatically generated

P value is significantly lower than 0.05 hence we reject the null hypothesis. Therefore, we conclude that lower “DemocracyIndex” (0-6) Countries don’t have equal mean of “DrugUse” compared to higher Democracy Index (6-10) Countries.

**Goodness of Fits Tests**

**Shapiro-Wilk Normality Test:**

* shapiro.test(DemocracyIndex)

A picture containing knife, bird

Description automatically generated

* shapiro.test(DrugUse)

A picture containing knife, table

Description automatically generated

P value for the “DemocracyIndex” is bigger than 0.05 (0.07018 > 0.05) implying that the distribution is not significantly different from normal distribution however, for the “DrugUse” p value is significantly smaller than 0.05 (0.00007847 > 0.05) is it not possible to assume normality.

**Anderson-Darling Test:**

After installing the required packages;

* ad.test(DemocracyIndex)

A picture containing bird, knife, table

Description automatically generated

* ad.test(DemocracyIndex)

A picture containing knife

Description automatically generated

For the “DemocracyIndex” p value is greater than 0.05 (0.1349 > 0.05) therefore no significant departure from normality was found however, for the “DrugUse” is significantly smaller than 0.05 (0.00003716 < 0.05).Failing the normality test allows us state with 95% confidence the “DrugUse” data does not fit the normal distribution.

**Maximum Likelihood Estimator**

**DemocracyIndex:**

# Checking for normal

* norMLE<-fitdist(DemocracyIndex,"norm",method="mle")

# Checking for gamma

* gamMLE<-fitdist(DemocracyIndex,"norm",method="mle")]

# Checking for exponential

* expMLE<-fitdist(DemocracyIndex,"norm",method="mle")

# Checking for Weibull

* weiMLE<-fitdist(DemocracyIndex,"norm",method="mle")

Graphs from the following codes are shown below;

A close up of a map

Description automatically generatedA close up of a map

Description automatically generated

norMLE gamMLE

A close up of a map

Description automatically generatedA close up of a map

Description automatically generated

expMLE weiMLE

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Test | Normal | Gamma | Exponential | Weibull |
| Likelihood | -91.67096 | -93.29125 | -110.9913 | -91.07928 |
| AIC | 187.3419 | 190.5825 | 223.9826 | 186.1586 |
| BIC | 190.7197g | 193.9603 | 225.6715 | 189.5363 |

From the results we can state that our distribution is either Normal or Weibull and it is slightly closer to Weibull.

**DrugUse:**

# Checking for normal

* nor\_MLE<-fitdist(DrugUse,"norm",method="mle")

# Checking for gamma

* gam\_MLE<-fitdist(DrugUse,"norm",method="mle")]

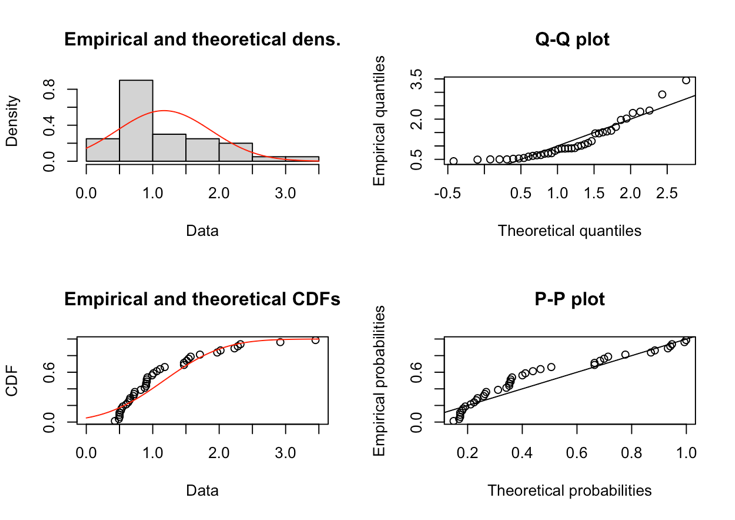
# Checking for exponential

* exp\_MLE<-fitdist(DrugUse,"norm",method="mle")

# Checking for Weibull

* wei\_MLE<-fitdist(DrugUse,"norm",method="mle")

Graphs from the following codes are shown below;



A close up of a map

Description automatically generated

nor\_MLE gam\_MLE

A close up of a map

Description automatically generatedA close up of a map

Description automatically generated

exp\_MLE wei\_MLE

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Test | Normal | Gamma | Exponential | Weibull |
| Likelihood | -43.0316 | -34.63645 | -46.2545 | -36.90217 |
| AIC | 90.06321 | 73.2729 | 94.509 | 77.80434 |
| BIC | 93.44097 | 76.65066 | 96.19788 | 81.1821 |

From the results MLE suggests that compared to the others “DrugUse” sample is most likely to be Gamma distribution.

**Correlation Tests**

A screenshot of a cell phone

Description automatically generated

A close up of a map

Description automatically generated

From the normality plots, we can state that “DemocracyIndex” may come from normal distributions but there “DrugUse” is less likely to be a normal distribution.

**Pearson Correlation Test:**

* res <- cor.test(DemocracyIndex, DrugUse, method = "pearson")
* res

A picture containing bird

Description automatically generated

Since the correlation coefficient is “-0.1042559” we can state that there is a low degree of negative correlation or no correlation at all between the data.

**Kendall’s Rank Correlation Test:**

* res2 <- cor.test(DemocracyIndex, DrugUse, method="kendall")
* res2

A picture containing bird, flower

Description automatically generated

“Tau” value may be used to evaluate the results. Since the correlation coefficient is smaller than 0.1 the test suggests no correlation between the samples.

**Linear Regression**

* model<-lm(DrugUse ~ DemocracyIndex)
* summary(model)

A screenshot of a cell phone

Description automatically generated

We have found our estimated coefficients β0 and β1 through the first column under coefficients. They are, respectively, 1.35155 and -0.03090. Our β1 slope value is negative, suggesting a negative correlation, but it is very small. This shows that given a one unit increase in DemocracyIndex, there would be a 0.03090 decrease in DrugUse.

Our Multiple R-Squared value, 0.01087, explains how much variation in DrugUse can be attributed to variation in DemocracyIndex. In our case, it is not very high. This means we cannot attribute much of the variation in the DrugUse data to that in the DemocracyIndex data.

Our F-statistic value is less than 1, which implies that the Mean Square Error is higher than the Mean Square Model. This tells us the error explains more of the data than the model itself.

Finally, the p-value, 0.522, is quite large compared to our 0.05 α value. This suggests the model is not very significant.

**Conclusion**

Early on the report, scatter plot for “DemocracyIndex vs DrugUse” demonstrates that there is no linear relation between our data and the majority of the test results suggests that there is no correlation between “DemocracyIndex” and “DrugUse” for the countries included in our dataset. More complex analysis might reveal better results when applied to this dataset however, our testing methods and constrained by the scope of this course.

**Bibliography**

"Free speech under attack." A report by The Economist Intelligence Unit Accessed September 11, 2020. https://pages.eiu.com/rs/753-RIQ438/images/Democracy\_Index\_2017.

"Global Burden of Disease Study 2017 (GBD 2017) Results”. Institute for Health Metrics and Evaluation (IHME), 2018. Accessed September 11, 2020. http://ghdx.healthdata.org/gbd-2017.

1. Index, Democracy. "Free speech under attack." A report by The Economist Intelligence Unit available at: https://pages.eiu.com/rs/753-RIQ-438/images/Democracy\_Index\_2017.pdf (2017). (retrieved September 2020). [↑](#footnote-ref-1)
2. "Global Burden of Disease Study 2017 (GBD 2017) Results." Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2018. (retrieved September 2020). [↑](#footnote-ref-2)