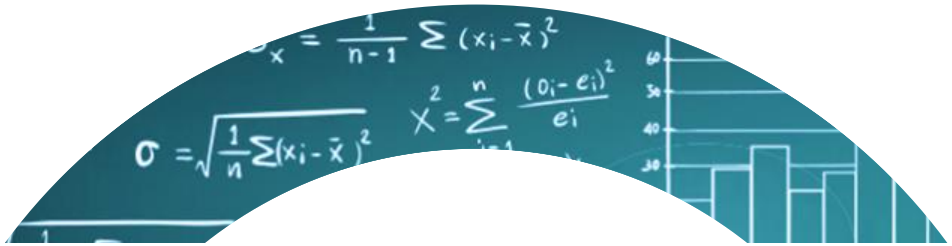

# WEST BENGAL STATE UNIVERSITY

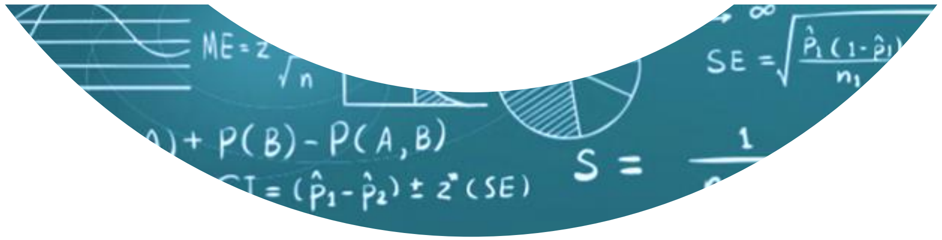
BATCH: 2020-2023
SUBJECT: STATISTICS
PAPER: STSADSE06P
ROLL:*** NO.: ***
REGISTRATION NO.: 1082011400131 OF 2020
COLLEGE ROLL NO.: 20157
SEMESTER-6



A Time Series Analysis and Inferential work on Happiness Index Of The World for around 150 countries over a time period of 10 years consisting of various factors

# STATISTICAL ANALYSIS

## ON
## HAPPINESS INDEX OF THE WORLD



# INTRODUCTION

1972

# INTRODUCTION

World Happiness Council

BHUTAN

Definition of Happiness Index

Bhutanese Gross National Happiness Index

Global Happiness Council

First to Compute Happiness Index

# DESCRIPTION

| Country name | Ladder score | Logged GDP per capita | Social support | Healthy life expectancy | Freedom to make life choices | Generosity | Perceptions of corruption | Ladder score in Dystopia |
|---|---|---|---|---|---|---|---|---|
| Finland | 7.842 | 10.775 | 0.954 | 72.0 | 0.949 | -0.098 | 0.186 | 2.43 |
| Denmark | 7.620 | 10.933 | 0.954 | 72.7 | 0.946 | 0.03 | 0.179 | 2.43 |
| Switzerland | 7.571 | 11.117 | 0.942 | 74.4 | 0.919 | 0.025 | 0.292 | 2.43 |
| Iceland | 7.554 | 10.878 | 0.983 | 73.0 | 0.955 | 0.16 | 0.673 | 2.43 |
| Netherlands | 7.464 | 10.932 | 0.942 | 72.4 | 0.913 | 0.175 | 0.338 | 2.43 |
| Norway | 7.392 | 11.053 | 0.954 | 73.3 | 0.96 | 0.093 | 0.27 | 2.43 |
| Sweden | 7.363 | 10.867 | 0.934 | 72.7 | 0.945 | 0.086 | 0.237 | 2.43 |
| Luxembourg | 7.324 | 11.647 | 0.908 | 72.6 | 0.907 | -0.034 | 0.386 | 2.43 |
| New Zealand | 7.277 | 10.643 | 0.948 | 73.4 | 0.929 | 0.134 | 0.242 | 2.43 |
| Austria | 7.268 | 10.906 | 0.934 | 73.3 | 0.908 | 0.042 | 0.481 | 2.43 |

1

# Ladder Score

Happiness Index is also known as Ladder Score

2

# Dystopia

An Imaginary country which have the least happiness index

In our data it is 1.2

3

# Gross Domestic Product per Capita (GDP per Capita)

A measure that calculates the country's economic output that accounts for the number of people in the country or the country's population.

# Describe the extent to which these factors contribute in evaluating the happiness in each country.

4

## Social Support

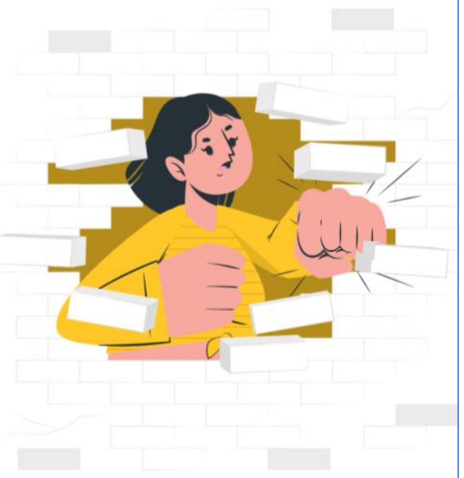
Social support is the perception and actuality that one is cared for, has assistance available from other people, and most popularly, that one is part of a supportive social network

## Life Expectancy

The term "life expectancy" refers to the number of years a person can expect to live. By definition, life expectancy is based on an estimate of the average age that members of a particular population group will be when they die.

## Freedom to make choices

Freedom to make choices describes a individual's opportunity and autonomy to perform an action selected from at least two available options, unconstrained by external



## Generosity

Generosity is the virtue of being liberal in giving, often as gifts

## Perception to corruption

Corruption and corruption perception can be considered as cultural phenomena because they depend on how a society understands the rules and what constitutes a deviation. Indeed, it does not depend only on societies but also on personal values and moral vies.

# PURPOSE

Here our work is to analyse this Happiness index of 146 countries and analyse the impact of the factors affecting happiness of a country and then we analyse the time series data of happiness index of India and try to forecast the happiness index for India

# OBJECTIVE

1

Compare the countries that how much happy they are

2

To infer how the happiness index is dependent upon the factors that may affect happiness index

3

To study happiness index of India for a range of time and try to infer the ladder score for future

# METHODOLOGY

METHODOLOGY
TWO STAGE ANALYSIS

1

Examining the intensity of the factors affecting Happiness Index

**(a)** Checking if there is any change in the intensity of the factors affecting Happiness Index

**(b)** Inferening about how can be it increase (if possible)

2

Time Series Analysis on the Happiness Index

**(a)** Analysing the data on Happiness Index from 2004-2023 of India

**(b)** Trying to forecast about the Happiness Index of India

# ANALYSIS 1 : LINEAR REGRESSION

## Model and Estimates

At first we are preforming **Linear Regression** to check the relationship of the covariates (**GDP per capita, Family Life Expectancy, Generousity, Trust Government Corruption**) on **Ladder Score** and also try to estimate the effects of the covariates (**GDP per capita, Family Life Expectancy, Generousity, Trust Government Corruption**) on **Ladder Score** i.e. **Happiness Index.**

Model :

So, here our model is,

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \beta_5 x_{5i} + \beta_6 x_{6i} + \epsilon_i \text{ , i} = 1(1)\text{n}$$

Assumption : $\epsilon_i \overset{iid}{\sim} \mathcal{N}(0, \sigma^2), where\ \sigma^2 is\ the\ error\ variance$

Hypothesis :

And the testing problem is,

$$H_{0j}: \beta_j = 0 \quad vs \quad H_{1j}: \beta_j \neq 0 \ \forall\ j = 1(1)6$$

Estimate :

Now, we know that by Least Square Method we can obtain estimates of $\beta_j\ \forall\ j = 1(1)6$ i.e. as follows

$$\widehat{\beta_j} = -\frac{R_j}{R_y}\frac{s_y}{s_j}$$

# Model and Estimates

Where R is the correlation matrix. Let $\rho_{ij} = \frac{cov(X_i,X_j)}{\sqrt{Var(X_i)\,Var(X_j)}}$ and $\rho_{yi} = \frac{cov(Y,X_i)}{\sqrt{Var(Y)\,Var(X_i)}}$ then R is as follows, $R = \begin{bmatrix} 1 & \rho_{y1} & \rho_{y2} & \cdots & \rho_{y6} \\ \rho_{y1} & 1 & \rho_{12} & \cdots & \rho_{16} \\ \rho_{y2} & \rho_{12} & 1 & \cdots & \rho_{26} \\ \vdots & \vdots & \vdots & & \vdots \\ \rho_{y6} & \rho_{16} & \rho_{26} & \cdots & 1 \end{bmatrix}^{6\times 6}$

And $\sigma_{ij} = cov(X_i, X_j)$ $and$ $\sigma_{yi} = cov(Y, X_i)$, $\sigma_y^2 = Var(Y)$ $and$ $\sigma_i^2 = Var(X_i)$

First of all we find the correlation

Now the R code is as follows,

```
D = file.choose()    #Choosing data on 2021

S = read.csv(D)

reg = lm(S$Ladder.score~(S$Logged.GDP.per.capita
                         +S$Social.support
                         +S$Healthy.life.expectancy
                         +S$Freedom.to.make.life.choices
                         +S$Generosity
                         +S$Perceptions.of.corruption))

summary(reg)
```

From this code we can obtain the estimates of the regression coefficients which is stated later

## Scatter Matrix

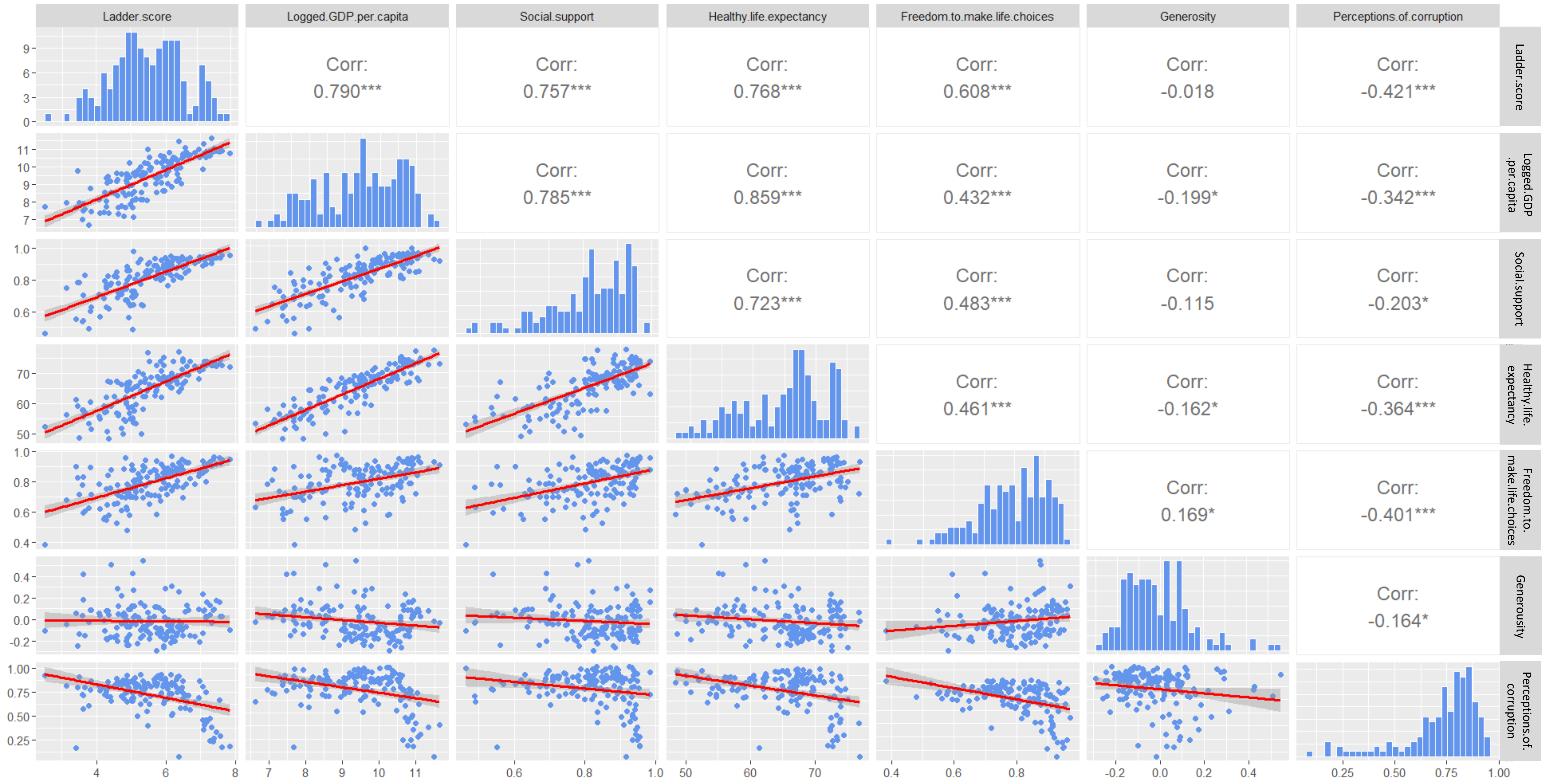
```r
S_1=cor(S[c("Ladder.score", "Logged.GDP.per.capita", "Social.support",
"Healthy.life.expectancy", "Freedom.to.make.life.choices", "Generosity",
"Perceptions.of.corruption")])


# Install and load reshape2 package
install.packages("reshape2")
library(reshape2)

# creating correlation matrix
corr_mat <- round(cor(S_1),2)


# reduce the size of correlation matrix
melted_corr_mat <- melt(corr_mat)
head(melted_corr_mat)

# plotting the correlation heatmap
library(ggplot2)
ggplot(data = melted_corr_mat, aes(x=Var1, y=Var2, fill=value)) +
  geom_tile() +
  geom_text(aes(Var2, Var1, label = value),
            color = "black", size = 4)
```



Ladder.score
Logged.GDP.per.capita
Social.support
Healthy.life.expectancy
Freedom.to.make.life.choices
Generosity
Perceptions.of.corruption
Corr: 0.790***
Corr: 0.757***
Corr: 0.768***
Corr: 0.608***
Corr: -0.018
Corr: -0.421***
Corr: 0.785***
Corr: 0.859***
Corr: 0.432***
Corr: -0.199*
Corr: -0.342***
Corr: 0.723***
Corr: 0.483***
Corr: -0.115
Corr: -0.203*
Corr: 0.461***
Corr: -0.162*
Corr: -0.364***
Corr: 0.169*
Corr: -0.401***
Corr: -0.164*
Ladder.score
Logged.GDP .per.capita
Social.support
Healthy.life. expectancy
Freedom.to. make.life.choices
Generosity
Perceptions.of. corruption

1 2 3 4 5

## Inference

Now let us test the significance of the regression coefficients

**Model :** $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \beta_5 x_{5i} + \beta_6 x_{6i} + \epsilon_i$ , i = 1(1)n

**Assumption :** $\epsilon_i \overset{iid}{\sim} \mathcal{N}(0, \sigma^2) \quad \forall \quad i = 1,2,\dots,n$

**Hypothesis :** Here the testing problem is,

$$H_0: \beta_1 = \beta_2 = \cdots = \beta_6 \quad vs \quad H_1: \text{atleast one inequality in } H_0$$

**Test Statistic :** Define,

$$S_1^2 = min\left\{\sum_{i=1}^{n}(y_i - \beta_0)^2\right\}$$

$$S_2^2 = min\left\{\sum_{i=1}^{n}(y_i - \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \beta_5 x_{5i} + \beta_6 x_{6i})^2\right\}$$

$$F = \frac{S_1^2 - S_2^2}{S_1^2}, \quad \text{where } F \sim F_{6,n-7}$$

**Test :** We reject $H_{0j}$ against $H_{1j}$ at level $\alpha$ iff

$$F > F_{\alpha,6,n-7}$$

In terms of p-value we reject $H_{0j}$ against $H_{1j}$ at level $\alpha$ iff

$$p - value = P_{H_{0j}}\big(|F| \geq observed(F)\big) \leq \alpha$$

Now we will test whether the regression coefficients individually are significant

**Model :** $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \beta_5 x_{5i} + \beta_6 x_{6i} + \epsilon_i$ , i = 1(1)n

**Assumption :** $\epsilon_i \overset{iid}{\sim} \mathcal{N}(0, \sigma^2) \quad \forall \quad i = 1,2,\dots,n$

**Hypothesis :** Here the testing problem is,

$$H_{0j}: \beta_j = 0 \quad vs \quad H_{1j}: \beta_j \neq 0 \,, j = 1(1)6$$

**Test Statistic :** Define,

$$t = \frac{\widehat{\beta_j}}{MSE}, \quad \text{where } t \sim t_{n-7}$$

**Test :** We reject $H_{0j}$ against $H_{1j}$ at level $\alpha$ iff

$$|t| > t_{\frac{\alpha}{2}}$$

In terms of p-value we reject $H_{0j}$ against $H_{1j}$ at level $\alpha$ iff

$$p - value = P_{H_{0j}}\big(|t| \geq observed(t)\big) \leq \alpha$$

So, now the output is,

```
Call:
lm(formula = S$Ladder.score ~ (S$Logged.GDP.per.capita + S$Social.support + 
    S$Healthy.life.expectancy + S$Freedom.to.make.life.choices + 
    S$Generosity + S$Perceptions.of.corruption))

Residuals:
     Min       1Q   Median       3Q      Max 
-1.85049 -0.30026  0.05735  0.33368  1.04878 

Coefficients:
                               Estimate Std. Error t value Pr(>|t|)    
(Intercept)                    -2.23722    0.63049  -3.548 0.000526 ***
S$Logged.GDP.per.capita         0.27953    0.08684   3.219 0.001595 ** 
S$Social.support                2.47621    0.66822   3.706 0.000301 ***
S$Healthy.life.expectancy       0.03031    0.01333   2.274 0.024494 *  
S$Freedom.to.make.life.choices  2.01046    0.49480   4.063 7.98e-05 ***
S$Generosity                    0.36438    0.32121   1.134 0.258541    
S$Perceptions.of.corruption    -0.60509    0.29051  -2.083 0.039058 *  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5417 on 142 degrees of freedom
Multiple R-squared:  0.7558,	Adjusted R-squared:  0.7455 
F-statistic: 73.27 on 6 and 142 DF,  p-value: < 2.2e-16
```

From the p-value of F-statistic we say that it is less than 0.05, so we can say that we reject $H0 : \beta j = 0 \ \forall \ j = 1(1)6$ at level of 0.05. That means all the atleast one of the effects are significant. But from the above data we can also see that, the p-value corresponding to the covariate Generosity is greater than 0.05. So, we accept $H05 : \beta 5 = 0$ at level 0.05 i.e. the effect of Generosity is insignificant to Ladder score, that means change in the value of Generosity more or less does not indicate any change in Ladder score and reject the rest at level 0.05. Now it is necessary to observe the validity of the assumptions for both the testing problems, and hence infer about the situation, or it may lead to a wrong inference.

# Normality of Error Term

Here we have considered that the error term is Normal. Now if we consider the error term is a non normal, then we have some complications in the testing procedure, and we cannot perform any t or F test.

So, let us check whether the error term is normal or not…

For this we first compute the residuals of the fitted model and then check whether they are a random variable following Normal distribution with mean 0 and some specific variance. For this we will perform Kolmogorov-Smirnov test

**Kolmogorov Smirnov Test :**

**Model :** $X_1, X_2, X_3, \dots, X_n \xrightarrow{iid} F(x)$

**Assumption :** Here we assume the F is absolutely continuous distribution function

**Hypothesis :** Here we are interested in testing whether the random sample come from a known distribution with distribution function $F_0$ (say). That is,

$$H_0 : F = F_0 \texttt{ against } H_a : F \neq F_0$$

$F = F_0 \equiv F(x) = F_0(x) \forall\ x \in \mathbb{R}$

$F \neq F_0 \equiv F(x) \neq F_0(x) \forall\ x$ with strict inequality for atleast one $x$

Here we are mainly interested in checking Normality, so, $F_0(x) = \Phi(x)$

So, the hypothesis becomes,

$$H_0 : F = \Phi \texttt{ against } H_a : F \neq \Phi$$

## Normality of Error Term

**Test Statistic :**

For testing $H_0$ against $H_a$ we define the following test statistic ;

$$D_n = \sup_{x\in\mathbb{R}} \left|\widehat{F_n(x)} - \Phi(x)\right|$$

Notice that $\widehat{F_n}$ approximate the true distribution function $\Phi$ . $\widehat{F_n}$ by definition is a step function i.e. the absolute difference measured by $D_n$ provide us the departure of the true situation from the null hypothesis towards the corresponding alternative.

**Distribution Free :**

EDF of $D_n$ :

Here we use the ordered statistics,

$$X_{(1)}, X_{(2)}, \dots, X_{(n)} \text{ . } Further\ we\ denote\ X_{(0)} = -\infty\ and\ X_{(n+1)} = +\infty$$

$$D_n = \sup_{x\,\in\,\mathbb{R}} \left|\widehat{F_n(x)} - \Phi(x)\right|$$

$$= \max_{i\,=\,0,1,2,\dots,n} \left\{ \sup_{X_{(i)} \le x \le X_{(i+1)}} \left|\widehat{F_n(x)} - \Phi(x)\right| \right\} \quad = max\left\{0, \max_{i\,=\,1,2,\dots,n} \left\{ \sup_{X_{(i)} \le x \le X_{(i+1)}} \left|\widehat{F_n(x)} - \Phi(x)\right| \right\}\right\}$$

$$= max\left\{0, \max_{i\,=\,1,2,\dots,n} \left\{ \sup_{X_{(i)} \le x \le X_{(i+1)}} \left|\frac{i}{n} - \Phi(x)\right| \right\}\right\} \quad = max\left\{0, \max_{i\,=\,1,2,\dots,n} \left\{ \left|\frac{i}{n} - \Phi(x_{(i)})\right| \right\}\right\}$$

# Normality of Error Term

Notice that,

$X_1, X_2, \dots, X_n \overset{iid}{\underset{H_0}{\sim}} \Phi$

$\Rightarrow U_i = \Phi(X_i) \overset{iid}{\underset{H_0}{\sim}} U(0,1)$

Since $\Phi$ is absolutely continuous, we can conclude that,

$\Rightarrow U_{(i)} = \Phi(X_{(i)}) \overset{iid}{\underset{H_0}{\sim}} U(0,1)$

That is $D_n$ (under $H_0$) depends on the ordered statistics $\{U_{(1)}, U_{(2)}, \dots, U_{(n)}\}$ from U(0,1). Hence the test is based on $D_n$ is EDF (Exact Distribution Free)

**Test :**

Notice that $D_n$ depends on the empirical distribution function $\widehat{F_n}$ which represents the true distribution function. Thus the directional difference measured by $D_n$ actually indicate the departure of the true situation form the null hypothesis towards the alternative $H_a$ i.e. under $H_a$ : F $\neq \Phi$ , $D_n$ becomes larger than that under $H_0$ . On the hand a small value of $D_n$ indicates the acceptance of $H_0$ . Thus a right tail test based on $D_n$ will be appropriate for testing $H_0 : F = \Phi$ against $H_a : F \neq \Phi$.

We reject $H_0 : F = \Phi$ against $H_a : F \neq \Phi$ at level $\alpha$ iff,

$$D_n > d_\alpha$$

In terms of p-value we can say, we reject $H_0 : F = \Phi$ against $H_a : F \neq \Phi$ at level $\alpha$ iff,

$$p - value = P_{H_0}(D_n \geq observed(\, D_n)) \leq \alpha$$

At first we compute the residuals and then perform the test.

## Normality of Error Term

At first we compute the residuals and then perform the test.

```
e = S$Ladder.score - fitted.values(reg)
ks.test(e, pnorm, mean(e), var(e))
```

Output

```
         One-sample Kolmogorov-Smirnov test

data:  e
D = 0.17886, p-value = 0.0001448
alternative hypothesis: two-sided
```

As the p-value is less than 0.05, we can say that we reject $H_0$ at a level 0.05

## Box Cox Transformation

Now we transform the data to normalize it. For this we use the Box-Cox Transformation.

**Box-Cox Transformation :**

For each real number , the Box–Cox transformation is

$$Y(\lambda) = \begin{cases} \dfrac{Y^{\lambda} - 1}{\lambda} & \lambda \neq 0 \\ \log(Y) & \lambda \neq 0 \end{cases}$$

Where we select the value of $\lambda$ such that the log-likelihood of Y becomes maximum.

The Box Cox transformation is named after statisticians George Box and Sir David Roxbee Cox who collaborated on a 1964 paper and developed the technique.

There are three main reasons for using the Box Cox transformation:

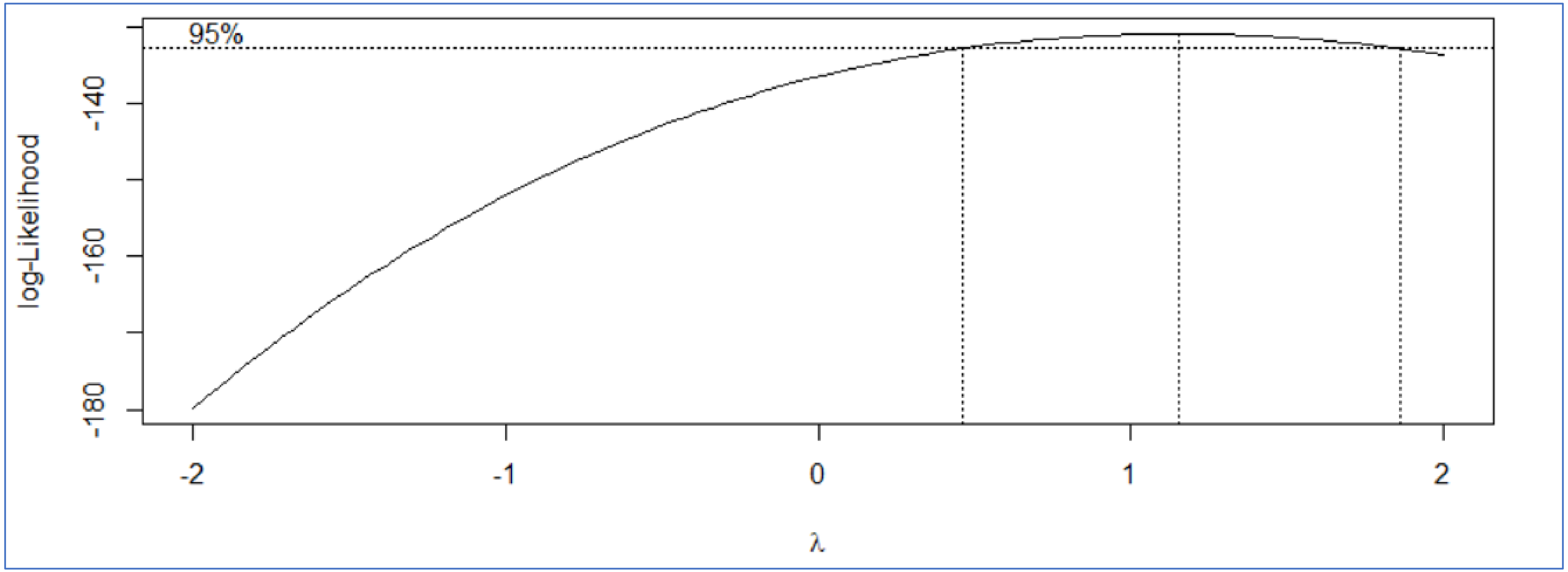
1. To stabilise the variance
2. To improve normality
3. To make patterns in the data more easily recognisable

# Box Cox Transformation

Here is the R code.

```
library(MASS)
u=boxcox(lm(S$Ladder.score~1))
l=u$x[which(u$y==max(u$y))]
y_1 = (((S$Ladder.score)^l) - 1)/l
```

Output




Here the value of $\lambda$ is 1.151515

## Box Cox Transformation

Now we obtain the residuals and again check that if the transformation resulted the normality of the variable. For this we again perform the Kolmogorov-Smirnov Test.

```
reg_y1 = lm(y_1~(S$Logged.GDP.per.capita
                 +S$Social.support
                 +S$Healthy.life.expectancy
                 +S$Freedom.to.make.life.choices
                 +S$Generosity
                 +S$Perceptions.of.corruption))

e_1 = y_1 - fitted.values(reg_y1)

ks.test(e_1, pnorm, mean(e_1), var(e_1))
```

Output

```
          One-sample Kolmogorov-Smirnov test

data:  e_1
D = 0.12269, p-value = 0.02254
alternative hypothesis: two-sided
```

As the p-value is less than 0.05 so we reject $H_0$ at level 0.05 but the data has improved towards normality.

# ANALYSIS 1 : TIME SERIES ANALYSIS

# Analysis of the Time Series

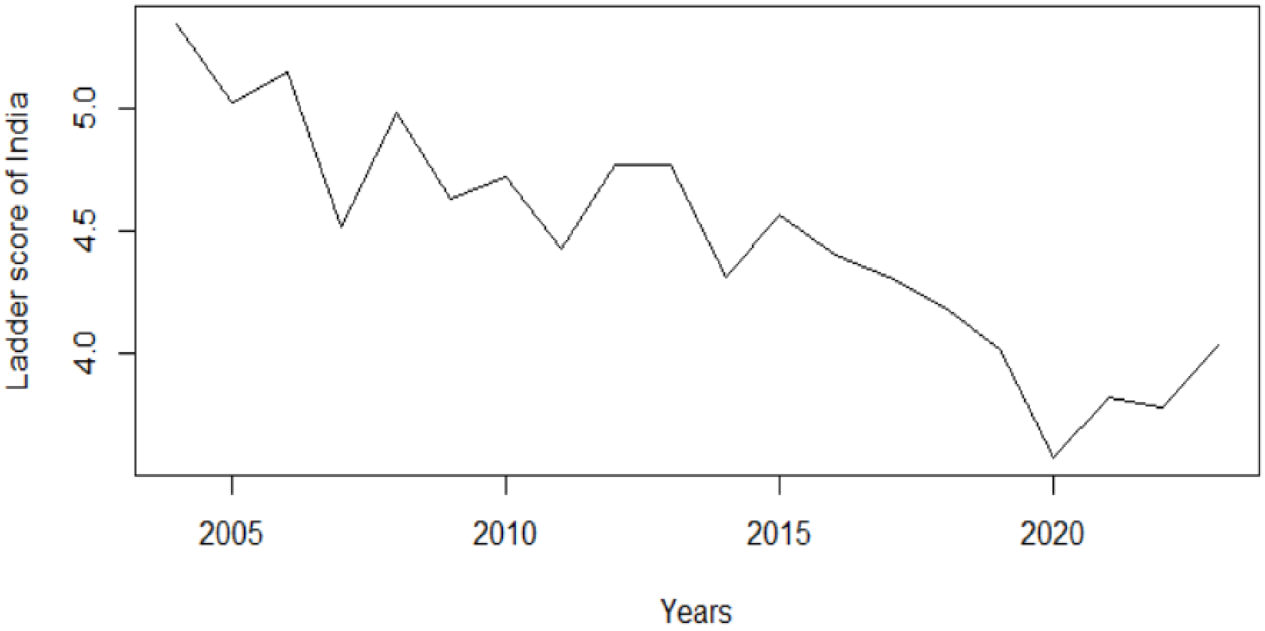
Data set

| Year | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 |
|---|---|---|---|---|---|---|---|---|---|---|
| Ladder Score | 5.350 | 5.030 | 5.150 | 4.520 | 4.990 | 4.630 | 4.720 | 4.43 | 4.772 | 4.770 |

| Year | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 | 2021 | 2022 | 2023 |
|---|---|---|---|---|---|---|---|---|---|---|
| Ladder Score | 4.315 | 4.565 | 4.404 | 4.315 | 4.190 | 4.015 | 3.573 | 3.819 | 3.777 | 4.036 |

Plotting the data set to analyse it

```
data = read.csv(file.choose())
ts = ts(data[,2], start = 2004, end = 2023, frequency = 1)
ts
plot(ts, xlab="Years", ylab="Ladder score of India")
```




From the plot we can say that,

1) There is a decreasing, downward trend in the time series, which seems to be very obvious.
2) The data is yearly data, so, there is no seasonal variation present in the data.
3) In the graph it can be observed that there are many ups and downs, but neither equidistant nor have equal amplitude. From this we can say there is aperiodic cyclical fluctuations.

## Stationary Time Series and Processes

A very special class of stochastic process, called stationary process, is based on the assumption
that the process is in a particular state of statistical equilibrium. A stochastic process is said to be strictly stationary if its properties are unaffected by a change of time origin, if the joint probability distribution associated with m observation Xt1 ,Xt2 , ...,Xtm is same as the associated with m observations Xt1+h,Xt2+h, ...,Xtm+h. From intuitive point of view a time series is said to be stationary if there is no systematic change in mean (trend), if there is no systematic change in variance and strictly periodic variations have been

There are different types of proccess, that are as follows;

1. Purely Random Process
2. Random Walk
3. Moving Average Process
4. Autoregressive Process

1 2 3 4 5

## Unit Root Test

Theory:

Model : Here we employ the model,

$$y_t = \phi y_{t-1} + e_t$$

where et denotes the error term.

Assumptions : To perform the test we consider that the error term is normal, i.e. $e_t \sim iid\ N(0, \sigma^2)$

where σ2 is the error variance.

Hypothesis : We perform the following testing problem,

H0 : ϕ = 1 against Ha : ϕ < 1

Test Statistic : A convenient test statistic is the t ratio of the least-squares (LS) estimated of ϕ under the null hypothesis. Now for the given model the LS estimates a

$$\implies \hat{\phi} = \frac{\sum_{t=1}^{T} y_{t-1} y_t}{\sum_{t=1}^{T} y_{t-1}^2}$$

lo

$$\therefore \hat{\sigma}^2 = \frac{\sum_{t=1}^{T} \left(y_t - \hat{\phi} y_{t-1}\right)^2}{T-1}$$

where y0 = 0 and T is the sample size. Now, the t-ratio is,

$$DF = \frac{\hat{\phi} - 1}{std\left(\hat{\phi}\right)} = \frac{\sum_{t=1}^{T} y_{t-1}(y_t - y_{t-1})}{\hat{\sigma}\sqrt{\sum_{t=1}^{T} y_{t-1}^2}} \overset{H_0}{\sim} t_{T-1}$$

This statistic is commonly referred as Dickey-Fuller (DF) statistic and the test is referred as Dickey-Fuller (DF) test.

Test Rule : We reject H0 against Ha at an level α if $DF < -t_{\alpha;T-1}$

where $t_{\alpha;T-1}$ is such that P($DF < -t_{\alpha;T-1}$) = α. This is also known as upper alpha point.

In terms of p-value we can say we reject H0 against Ha at an level α if

p − value = P (DF ≤ observed(DF)) ≤ α

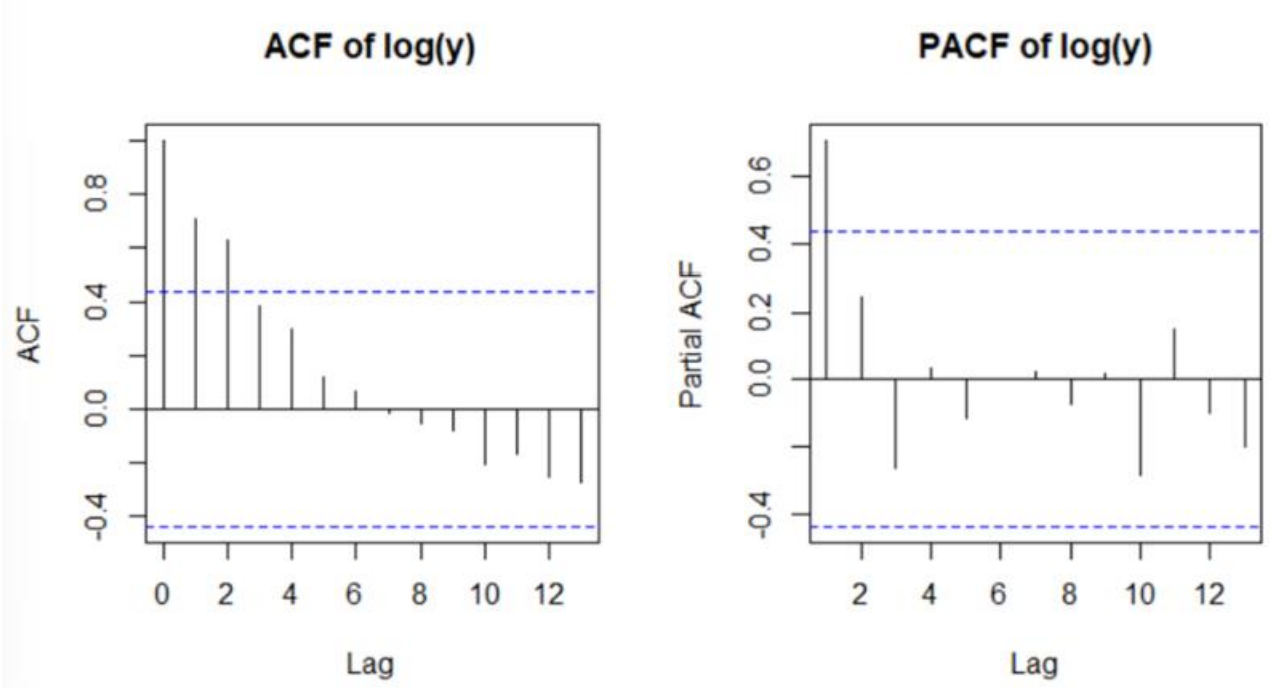
```
adf.test(y_new)
```

By calculation we can see that the value of the test statistic is -2.3359 and p−value is 0.4444, which is greater than 0.05. So we fail to reject $H_0$ at level 0.05.

# Comparing ACF and PACF of the logged and differenced series

## ACF and PACF of log(y)

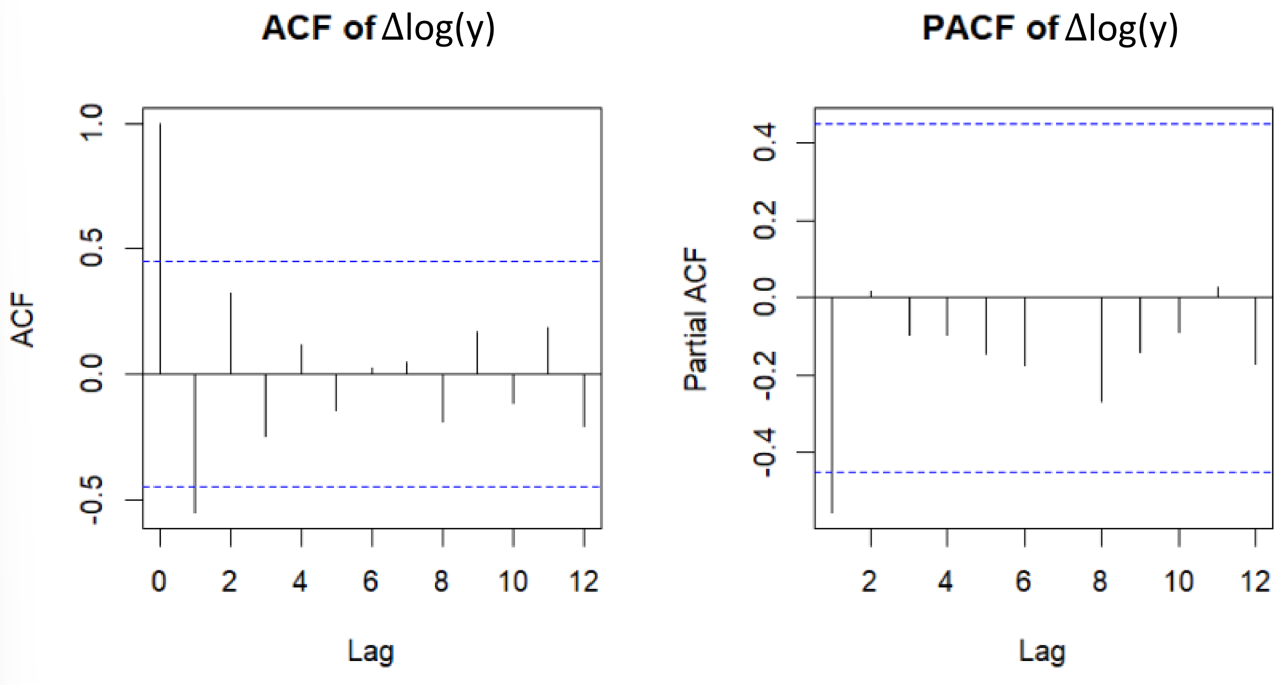
```
acf(y_new)
pacf(y_new)
```



# Comparing ACF and PACF of the logged and differenced series

## ACF and PACF of Δlog(y)

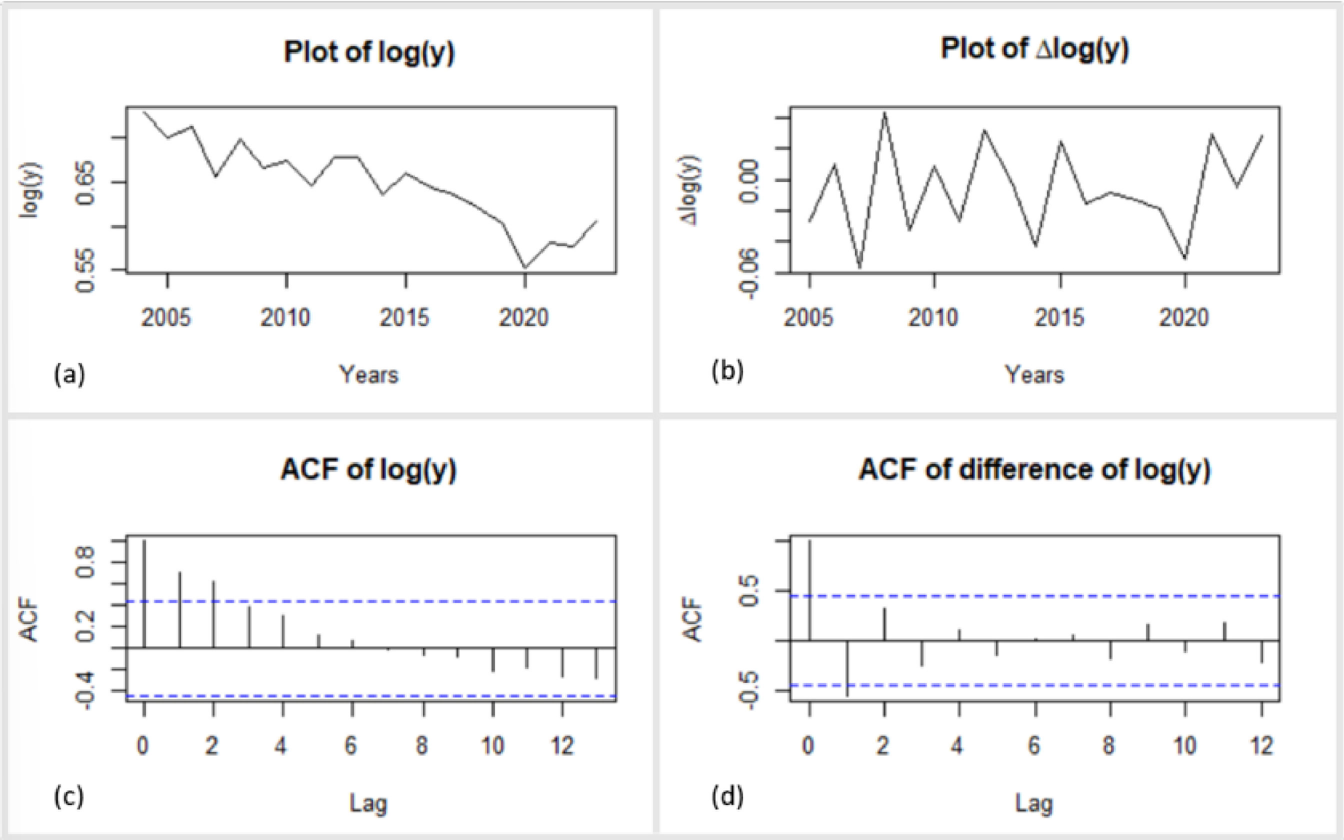
```
acf(diff(y_new))
pacf(diff(y_new))
```



1 2 3 4 5

# Comparing ACF and PACF of the logged and differenced series

Comparison of ACF of log(y) and Δlog(y)



1 2 3 4 5

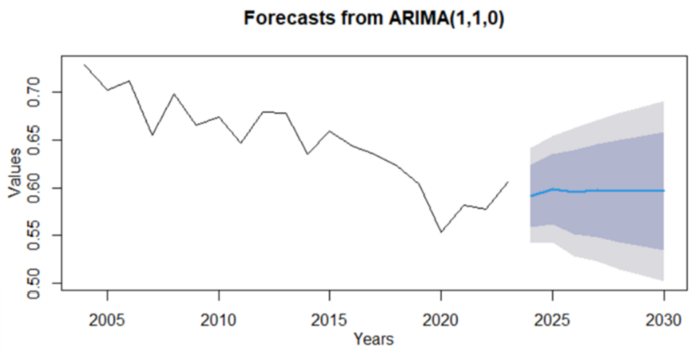
# Model Fitting

## ARIMA(1,1,0)

AIC : -81.37

Multiple R squared : 0.7662135

```
M=arima(y_new, order=c(1, 1, 0))
summary(M)
plot(forecast (M, 7))
(cor(y_new, fitted.values(M)))^2
```

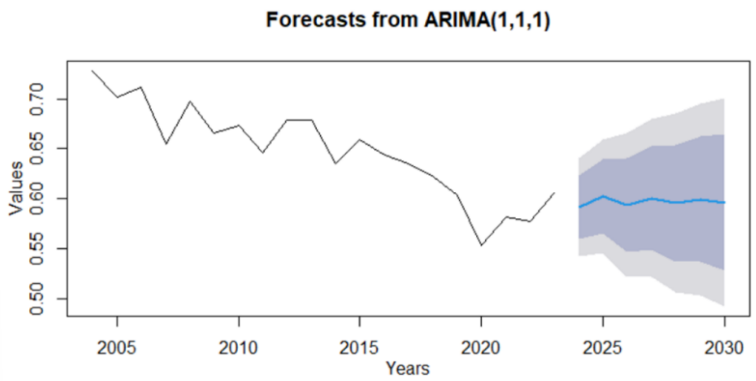



## ARIMA(1,1,1)

AIC : -79.98

Multiple R squared : 0.7687116

```
M1=arima(y_new, order=c(1, 1, 1))
summary(M1)
plot(forecast (M1, 7))
(cor(y_new, fitted.values(M1)))^2
```

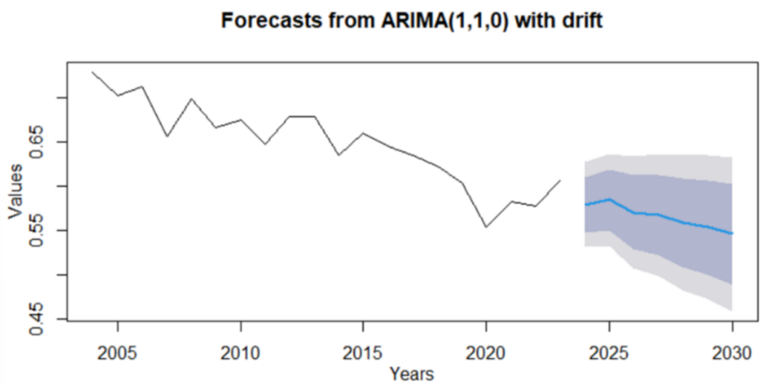



## ARIMA(1,1,0) with a drift

AIC : -82.72

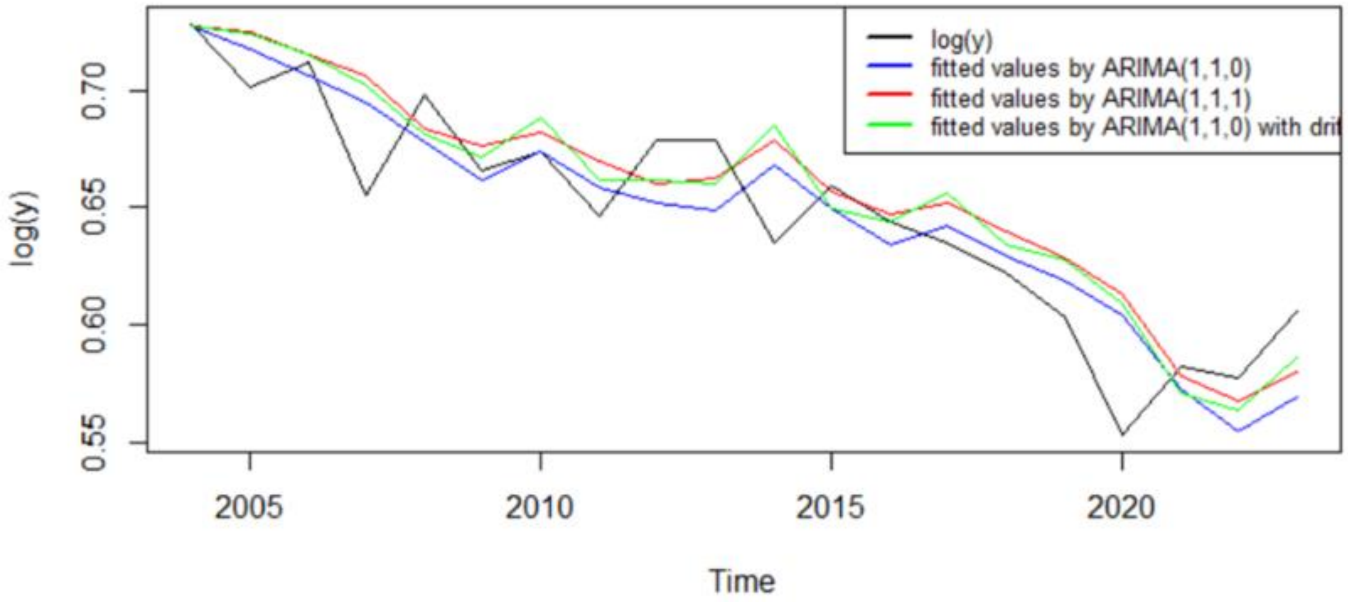
Multiple R squared : 0.7753745

```
M2=auto.arima(y_new)
summary(M2)
plot(forecast (M2, 7))
(cor(y_new, fitted.values(M2)))^2
```



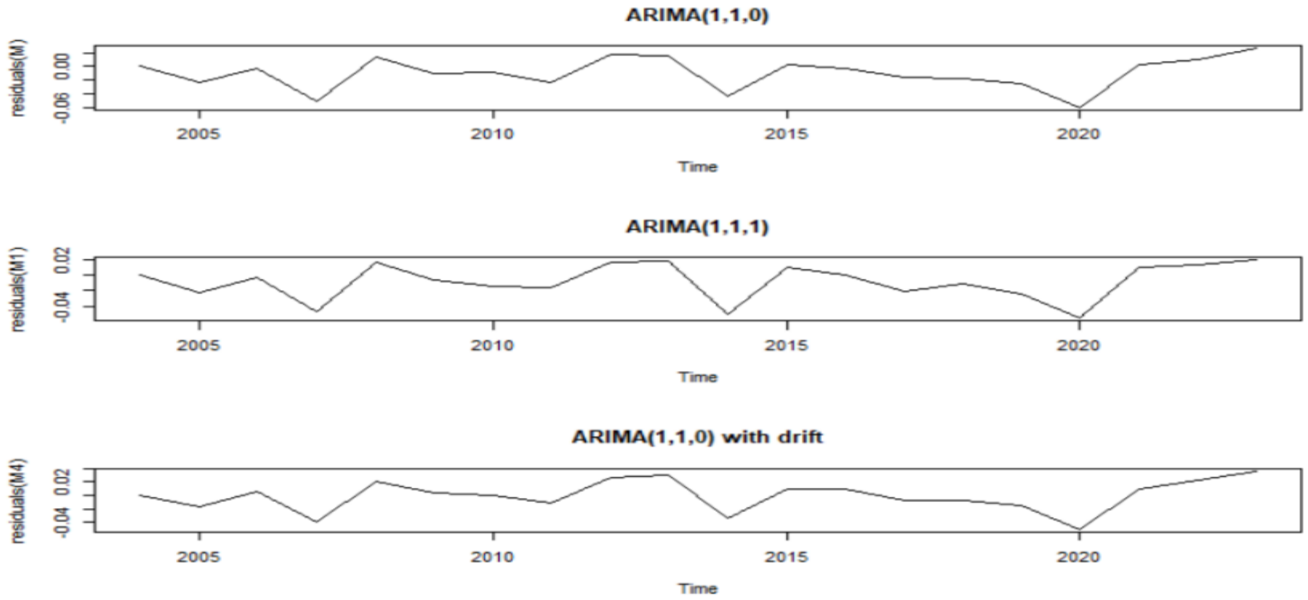
# Comparing the fitted models

```
plot(y_new, type='l', ylab="log(y)")
lines(fitted.values(M4), col="blue")
lines(fitted.values(M), col="red")
lines(fitted.values(M1), col="green")
legend("topright",legend=c('log(y)',
                            'fitted␣values␣by␣ARIMA(1,1,0)',
                           'fitted␣values␣by␣ARIMA(1,1,1)',
                           'fitted␣values␣by␣ARIMA(1,1,0)␣with␣drift'),
        col=c('black', 'blue', 'red','green'),lty=c(1,1,1,1), cex=0.8,lwd=2)
```



# Residuals

```
par(mfrow= c(3,1))
plot(residuals(M), pch = 1, main="ARIMA(1,1,0)")
plot(residuals(M1), pch = 1, main="ARIMA(1,1,1)")
plot(residuals(M4), pch = 1, main="ARIMA(1,1,0) with drift")
```



# Happiness Index in Recent

In a article "Examining India's Falling Rank on the World Happiness Index" written by Aanya Poddar, the writer is enlisting Former Indian President Pranab Mukherjee's comment, "Despite our country's economic progress, India is constantly going downwards in the happiness index. This indicates a lack of a holistic approach towards development." (https://openaxis.in/20 21/04/09/what-we-can-learn-from-the-world-happiness-index/) which is quite obvious and from our analysis we can see that the Happiness index will decrease, though in a low extent. Which also indicates a downward trajectory in the development of India. Again in another article of Governance Now, named "Why India's ranking on Happiness Index has been falling" is had been reported that "Even though India has been one of the fastest developing countries, the happiness score has worsened year by year. Amid the pandemic, happiness has become more elusive than ever. Covid has also taught us how to value immaterial aspects more than anything and the true purpose of a country. These findings suggest that the country needs to focus on intangible aspects and happiness during these challenging times." (https://www.governancenow.com/news/regular-story/why-indias-ranking-on-happiness-index-has-been-falling) which inclines towards the thought to increase the awareness of a wholesome socio-economic development required to increase the score. In the section of linear regression we see that various factors significantly affects the happiness index. So, increase in them can also make an uplift in happiness index. During the Covid period, 2020-2021, we can see that the happiness index got drowned. Again it starts an upward movement. It is increasing but also it need a development in greater extent to make the happiness index more higher.

#### Acknowledgement

I would want to convey my heartfelt gratitude to Prof. Kiranmoy Chatterjee and Prof. Soumyadeep Das, my mentors, for their invaluable advice and assistance in completing my project. They were there to assist me every step of the way, and his motivation is what enabled me to accomplish my task effectively. Their insightful feedback pushed me to sharpen my thinking and brought my work to a higher level. I would also like to thank all of the other supporting personnel who assisted me by supplying the equipment that was essential and vital, without which I would not have been able to perform efficiently on this project. I would also like to thank my tutors, Prof. Debesh Ray, Prof. Arup Kumar Hait, Prof. Suryasish Chatterjee for their valuable guidance throughout my studies. You provided me with the tools that I needed to choose the right direction and successfully complete my dissertation I would also want to thank the West Bengal State University for accepting my project in my desired field of expertise. I would also like to thank my friends and parents for their support and encouragement as I worked on this assignment.