R Codes

**A] Chi square test**

1 ] A=matrix(c(45.53,2.8,19,6.42,1.39,5.03,13.97,1.68,4.19),ncol=3,nrow=3,byrow = TRUE)

chisq.test(A)

2]C=matrix(c(20.9497,28.4916,12.0117,6.4245,3.6312,6.4245,5.5865,1.3966,3.0726,2.5139,1.3966,1.6759,1.9553,1.6759,1.3966,1.39664),ncol=4,nrow=4,byrow = TRUE)

chisq.test(C)

3] B=matrix(c(37.7,29.7,8.37,4.48,8.38,11.46),ncol=2,nrow=3,byrow = TRUE)

chisq.test(B)

**B] Time Series Analysis**

**1]Coal Consumption**

library('readxl')

library('TTR')

library('tseries')

library('forecast')

library('Metrics')

coal\_c <- read\_excel("C:/Users/malla/Downloads/coal-consumption-by-country-terawatt-hours-twh.xlsx")

coalc\_ts=ts(coal\_c$Value,start = c(1965,1), frequency = 1)

plot(coalc\_ts)

cmodel=auto.arima(coalc\_ts,ic="aic",trace = TRUE)

b=forecast(cmodel)

plot(b)

summary(b)

**# Forecast for year 2020**

training\_set = subset(coal\_c,subset = coal\_c$Year>='1965' & coal\_c$Year <='2019')

coal\_cts=ts(training\_set$Value,start = c(1965,1), frequency = 1)

plot(coal\_cts)

cmodel=auto.arima(coal\_cts,ic="aic",trace = TRUE)

b=forecast(cmodel,h=1)

plot(b)

summary(b)

**# New Arima model**

c <- read\_excel("C:/Mallu/Project/c.xlsx")

c1=ts(c$Value,start = c(1965,1), frequency = 1)

plot(c1)

cmodel=auto.arima(c1,ic="aic",trace = TRUE)

b=forecast(cmodel)

plot(b)

summary(b)

**# Residual Analysis**

Box.test(resid(cmodel),type="Ljung",lag=1,fitdf=0)

checkresiduals(cmodel)

**2] Coal Production**

**# Holt Winters Model**

coal\_p<- read\_excel("C:/Mallu/Project/production.xlsx")

training\_set = subset(coal\_p,subset = coal\_p$Year>='1900' & coal\_p$Year<='2000')

test\_set = subset(coal\_p, subset = coal\_p$Year>='2001' & coal\_p$Year <='2020')

coalp\_ts=ts(training\_set$Value,start = c(1900,1), frequency = 1)

plot(coalp\_ts)

model= HoltWinters(coalp\_ts,gamma = FALSE)

plot(model)

a=forecast(model,h=20)

plot(a)

summary(a)

x=as.numeric(unlist(a$mean))

y=as.numeric(unlist(test\_set$Value))

**# RMSE**

rmse(y,x)

**# ARIMA Model**

cmodel=auto.arima(coalp\_ts,ic="aic",trace = TRUE)

b=forecast(cmodel,h=20)

plot(b)

summary(b)

x=as.numeric(unlist(b$mean))

y=as.numeric(unlist(test\_set$Value))

# RMSE

rmse(y,x)

**C] Simple Linear Regression**

**#CO2 And Coal Production**

library(readxl)

coal\_production\_by\_country <- read\_excel("C:/Mallu/Project/New folder/coal-production-by-country.xlsx")

electricity\_coal <- read\_excel("C:/Mallu/Project/New folder/electricity-coal.xlsx")

co2\_emission <- read\_excel("C:/Mallu/Project/New folder/co2 emission.xlsx")

y=co2\_emission$`Annual CO2 emissions`

x=coal\_production\_by\_country$`Coal production (TWh)`

model=lm(y~x)

summary(model)

**# Diagnostic Plots**

par(mfrow=c(2,2))

plot(model)

**#Box Cox Transformation**

library('MASS')

bc=boxcox(y~x)

lamda=bc$x[which.max(bc$y)]

lamda

newmodel=lm(((y^lamda-1)/lamda)~x)

summary(newmodel)

par(mfrow=c(2,2))

plot(newmodel)

**# Electricity Generation vs Coal Production**

y=electricity\_coal$`Electricity from coal (TWh)`

x=coal\_production\_by\_country$`Coal production (TWh)`

model=lm(y~x)

summary(model)

**#Diagnostic Plots**

par(mfrow=c(2,2))

plot(model)

**#Box Cox Transformation**

bc=boxcox(y~x)

lamda=bc$x[which.max(bc$y)]

newmodel=lm(((y^lamda-1)/lamda)~x)

summary(newmodel)

par(mfrow=c(2,2))

plot(newmodel)