UOC LUONG MO HINH

NGUYEN QUANG DONG

febrary 1, 2023

library(sandwich)  
library(car)

## Loading required package: carData

library(lmtest)

## Loading required package: zoo

##   
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':  
##   
## as.Date, as.Date.numeric

library(carData)  
library(ggplot2)

## NHẬP SỐ LIỆU

# setwd("D:/dataR/ch123")  
ch3bt4\_m=read.table("ch3bt4\_m.txt",header=TRUE)  
View(ch3bt4\_m)  
CT=ch3bt4\_m$CT  
TN=ch3bt4\_m$TN  
TS=ch3bt4\_m$TS  
TNP=ch3bt4\_m$TNP

## ƯỚC LƯỢNG HÀM CHI TIÊU CT=F(TN,TS,TNP)

reg1=lm(CT~TN+TS+TNP)  
summary(reg1)

##   
## Call:  
## lm(formula = CT ~ TN + TS + TNP)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -17.354 -10.526 0.716 5.726 31.658   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 55.374810 13.420157 4.126 0.000284 \*\*\*  
## TN 0.780209 0.028218 27.650 < 2e-16 \*\*\*  
## TS 0.007636 0.017628 0.433 0.668079   
## TNP 0.092031 0.193024 0.477 0.637087   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 12.44 on 29 degrees of freedom  
## Multiple R-squared: 0.9995, Adjusted R-squared: 0.9995   
## F-statistic: 2.141e+04 on 3 and 29 DF, p-value: < 2.2e-16

# Kiểm định mô hình (1), hệ số của TS và TNP đều bằng không  
  
  
# Cách 1  
linearHypothesis(reg1,c("TS","TNP"))

## Linear hypothesis test  
##   
## Hypothesis:  
## TS = 0  
## TNP = 0  
##   
## Model 1: restricted model  
## Model 2: CT ~ TN + TS + TNP  
##   
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 31 6895.4   
## 2 29 4485.1 2 2410.3 7.7922 0.001957 \*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

# Cách 2  
reg2=lm(CT~TN)  
summary(reg2)

##   
## Call:  
## lm(formula = CT ~ TN)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -25.041 -11.911 -1.339 12.424 36.273   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 74.478901 7.860251 9.475 1.14e-10 \*\*\*  
## TN 0.853325 0.004038 211.303 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 14.91 on 31 degrees of freedom  
## Multiple R-squared: 0.9993, Adjusted R-squared: 0.9993   
## F-statistic: 4.465e+04 on 1 and 31 DF, p-value: < 2.2e-16

RSS1=sum(resid(reg1)^2)  
RSS2=sum(resid(reg2)^2)  
F=(RSS2-RSS1)\*(length(CT)-4)/(RSS1\*2)  
'F=';F

## [1] "F="

## [1] 7.792217

p\_value=pf(F,2,(length(CT)-4))  
'p\_value=';p\_value

## [1] "p\_value="

## [1] 0.9980429

# Cách 3  
  
R12=summary(reg1)$r.squared  
R12

## [1] 0.9995487

R22=summary(reg2)$r.squared  
R22

## [1] 0.9993062

F=(R12-R22)\*(length(CT)-4)/((1-R12)\*2)  
'F=';F

## [1] "F="

## [1] 7.792217

p\_value=pf(F,2,(length(CT)-4))  
'p\_value='; p\_value

## [1] "p\_value="

## [1] 0.9980429

# Phân tích và kiểm định đa cộng tuyến  
cor(ch3bt4\_m)

## CT TN TNP TS  
## CT 1.0000000 0.9996530 0.9891856 0.9810330  
## TN 0.9996530 1.0000000 0.9870445 0.9781590  
## TNP 0.9891856 0.9870445 1.0000000 0.9979506  
## TS 0.9810330 0.9781590 0.9979506 1.0000000

cor(TS,TNP)

## [1] 0.9979506

HQPhu1=lm(TN~TS+TNP)  
summary(HQPhu1)

##   
## Call:  
## lm(formula = TN ~ TS + TNP)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -166.11 -34.46 13.35 42.44 187.31   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 195.12944 79.18680 2.464 0.0197 \*   
## TS -0.41758 0.08484 -4.922 2.90e-05 \*\*\*  
## TNP 5.60080 0.71704 7.811 1.02e-08 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 80.47 on 30 degrees of freedom  
## Multiple R-squared: 0.9858, Adjusted R-squared: 0.9848   
## F-statistic: 1038 on 2 and 30 DF, p-value: < 2.2e-16

HQPhu2=lm(TNP~TS)  
summary(HQPhu2)

##   
## Call:  
## lm(formula = TNP ~ TS)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -25.163 -15.393 -1.591 11.960 46.504   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -86.33901 12.36749 -6.981 7.8e-08 \*\*\*  
## TS 0.11807 0.00136 86.832 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 20.16 on 31 degrees of freedom  
## Multiple R-squared: 0.9959, Adjusted R-squared: 0.9958   
## F-statistic: 7540 on 1 and 31 DF, p-value: < 2.2e-16

# BỎ BIẾN KHẮC PHỤC ĐA CỘNG TUYẾN

reg3=lm(CT~TN+TS)  
summary(reg3)

##   
## Call:  
## lm(formula = CT ~ TN + TS)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -16.026 -10.504 0.373 5.784 32.274   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 50.606180 8.832144 5.73 2.97e-06 \*\*\*  
## TN 0.791224 0.015991 49.48 < 2e-16 \*\*\*  
## TS 0.015818 0.003984 3.97 0.000414 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 12.27 on 30 degrees of freedom  
## Multiple R-squared: 0.9995, Adjusted R-squared: 0.9995   
## F-statistic: 3.296e+04 on 2 and 30 DF, p-value: < 2.2e-16

reg4=lm(CT~TN+TNP)  
summary(reg4)

##   
## Call:  
## lm(formula = CT ~ TN + TNP)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -18.6212 -10.2061 0.2994 5.7553 31.2154   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 60.1966 7.3949 8.140 4.36e-09 \*\*\*  
## TN 0.7720 0.0207 37.293 < 2e-16 \*\*\*  
## TNP 0.1734 0.0436 3.978 0.000406 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 12.27 on 30 degrees of freedom  
## Multiple R-squared: 0.9995, Adjusted R-squared: 0.9995   
## F-statistic: 3.301e+04 on 2 and 30 DF, p-value: < 2.2e-16

## ĐỔI DẠNG HÀM & RÚT GỌN HÀM

reg5=lm(log(CT)~log(TN)+log(TS))  
summary(reg5)

##   
## Call:  
## lm(formula = log(CT) ~ log(TN) + log(TS))  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.0150566 -0.0065420 -0.0002785 0.0043259 0.0274632   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.08456 0.09528 0.888 0.382   
## log(TN) 0.90739 0.03034 29.903 <2e-16 \*\*\*  
## log(TS) 0.05525 0.03449 1.602 0.120   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.009425 on 30 degrees of freedom  
## Multiple R-squared: 0.9992, Adjusted R-squared: 0.9992   
## F-statistic: 1.926e+04 on 2 and 30 DF, p-value: < 2.2e-16

reg6=lm(log(CT)~log(TN))  
summary(reg6)

##   
## Call:  
## lm(formula = log(CT) ~ log(TN))  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.015858 -0.006890 -0.001609 0.006373 0.025691   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.225653 0.037244 6.059 1.04e-06 \*\*\*  
## log(TN) 0.955372 0.004989 191.501 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.00966 on 31 degrees of freedom  
## Multiple R-squared: 0.9992, Adjusted R-squared: 0.9991   
## F-statistic: 3.667e+04 on 1 and 31 DF, p-value: < 2.2e-16

# Tinh/hien thị cac gia tri Yhat  
LCThat=fitted(reg6)  
# Tinh/hien thi phan dư  
e=resid(reg6)  
# Hien thi cac betahat  
beta=coef(reg6)  
beta

## (Intercept) log(TN)   
## 0.2256527 0.9553722

beta[1]

## (Intercept)   
## 0.2256527

anova(reg6)

## Analysis of Variance Table  
##   
## Response: log(CT)  
## Df Sum Sq Mean Sq F value Pr(>F)   
## log(TN) 1 3.4224 3.4224 36673 < 2.2e-16 \*\*\*  
## Residuals 31 0.0029 0.0001   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

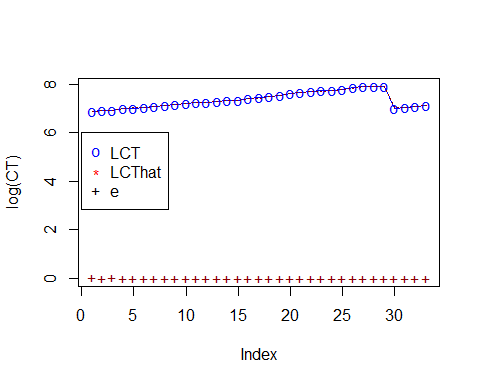
deviance(reg6)

## [1] 0.002893003

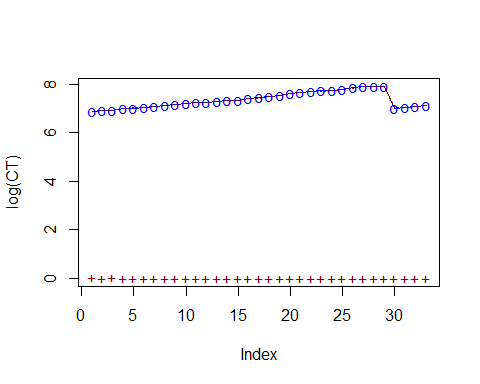
covbetahat=vcov(reg6)  
covbetahat

## (Intercept) log(TN)  
## (Intercept) 0.0013871225 -1.856162e-04  
## log(TN) -0.0001856162 2.488877e-05

plot(log(CT), type="o", col="blue", pch="o", lty=1, ylim=c(0,max(log(CT))))  
lines(LCThat, col="dark red", lty=3)  
points(resid(reg6), col="dark red",pch="+")  
  
legend(0,6,legend=c("LCT","LCThat","e"), col=c("blue","red","black"), pch=c("o","\*","+"),lty=c(0,0,0), ncol=1)



plot(log(CT), type="o", col="blue", pch="o", lty=1, ylim=c(0,max(log(CT))))  
lines(LCThat, col="dark red", lty=3)  
points(resid(reg6), col="dark red",pch="+")



# legend(25,6,legend=c("LCT","LCThat","e"), col=c("blue","red","black"), pch=c("o","\*","+"),lty=c(0,0,0), ncol=1)

## DỰ BÁO

# newdata=data.frame(data)  
TN=c(1200,1300)  
TS=c(8000,9000)  
newdata=cbind(TN,TS)  
newdata

## TN TS  
## [1,] 1200 8000  
## [2,] 1300 9000

newdata=data.frame(newdata)  
# Dự bao diem  
prereg5=predict(reg5,newdata)  
prereg5

## 1 2   
## 7.014612 7.093750

# Du bao khoang  
prereg5=predict(reg5,newdata,interval = 'confidence', level=0.95)  
prereg5

## fit lwr upr  
## 1 7.014612 6.994490 7.034733  
## 2 7.093750 7.070429 7.117071