Dummy Variable-Bai tap 4.9

NGUYEN QUANG DONG

27/2/2023

library(lmtest)

## Loading required package: zoo

##   
## Attaching package: 'zoo'

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

library(zoo)  
library(ggplot2)  
library(car)

## Loading required package: carData

library(carData)  
library(strucchange)

## Loading required package: sandwich

library(sandwich)  
library(urca)  
library(forecast)

## Registered S3 method overwritten by 'quantmod':  
## method from  
## as.zoo.data.frame zoo

## NANG SUAT LAO DONG PHU THUOC VAO KINH NGHIEM, LAO DONG DA TUNG CHUYEN NGHE

# BAI TAP 4.9  
setwd("D:/dataR/ch123")  
ch4bt9\_m=read.table("ch4bt9\_m.txt",header=TRUE)  
View(ch4bt9\_m)  
  
# ns= nang suat; kn= kinh nghiem; td = (td= da tung chuyen nghe, ktd=chua tung chuyen nghe)   
ns=ch4bt9\_m$ns  
kn=ch4bt9\_m$kn  
td=ch4bt9\_m$td  
#TAO BIEN GIA DUMMY, TU KHOA "factors"  
dn = factor(td, labels=c("0", "1"))  
dn

## [1] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1  
## [39] 1 1  
## Levels: 0 1

reg=lm(ns ~ kn+dn)  
summary(reg)

##   
## Call:  
## lm(formula = ns ~ kn + dn)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.51050 -0.86024 -0.04967 0.67546 2.52773   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 9.91727 0.97315 10.191 2.73e-12 \*\*\*  
## kn 0.12176 0.02106 5.780 1.25e-06 \*\*\*  
## dn1 -0.90718 0.38104 -2.381 0.0225 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.166 on 37 degrees of freedom  
## Multiple R-squared: 0.5067, Adjusted R-squared: 0.4801   
## F-statistic: 19 on 2 and 37 DF, p-value: 2.099e-06

## R CO THE TU TAO BIEN GIA

# NEU CO m PHAM TRU, R SE DUA m-1 bien gia

#TU DAT PHAM TRU CO SO THEO TEN “BIEN DINH” + “TEN PHAM TRU CO SO”, O DAY LA “tdtd”

reg\_1=lm(ns~kn+td)  
summary(reg\_1)

##   
## Call:  
## lm(formula = ns ~ kn + td)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.51050 -0.86024 -0.04967 0.67546 2.52773   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 9.91727 0.97315 10.191 2.73e-12 \*\*\*  
## kn 0.12176 0.02106 5.780 1.25e-06 \*\*\*  
## tdtd -0.90718 0.38104 -2.381 0.0225 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.166 on 37 degrees of freedom  
## Multiple R-squared: 0.5067, Adjusted R-squared: 0.4801   
## F-statistic: 19 on 2 and 37 DF, p-value: 2.099e-06

## KIEM DINH CHOW: KIEM DINH SU ON DINH CUA HAM HOI QUY

# Hoi quy voi lao dong chua chuyen nghe

index1=subset(ch4bt9\_m, td=="ktd")  
index1

## dn kn ns td  
## 1 0 38 13.73 ktd  
## 2 0 38 16.43 ktd  
## 3 0 55 16.68 ktd  
## 4 0 35 14.74 ktd  
## 5 0 38 15.31 ktd  
## 6 0 49 15.66 ktd  
## 7 0 54 16.12 ktd  
## 8 0 23 11.58 ktd  
## 9 0 45 15.46 ktd  
## 10 0 38 13.51 ktd  
## 11 0 54 14.94 ktd  
## 12 0 51 17.26 ktd  
## 13 0 45 15.31 ktd  
## 14 0 46 17.69 ktd  
## 15 0 50 14.58 ktd

reg1=lm(ns ~ kn, data=index1)  
summary(reg1)

##   
## Call:  
## lm(formula = ns ~ kn, data = index1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.53489 -0.93453 -0.08469 0.65050 2.17529   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 9.99368 1.59158 6.279 2.84e-05 \*\*\*  
## kn 0.12002 0.03556 3.376 0.00497 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.181 on 13 degrees of freedom  
## Multiple R-squared: 0.4671, Adjusted R-squared: 0.4261   
## F-statistic: 11.39 on 1 and 13 DF, p-value: 0.00497

# Hoi quy voi lao dong da tung chuyen nghe  
  
index2=subset(ch4bt9\_m, td=="td")  
index2

## dn kn ns td  
## 16 1 55 17.15 td  
## 17 1 45 13.63 td  
## 18 1 41 13.04 td  
## 19 1 42 13.20 td  
## 20 1 39 13.70 td  
## 21 1 51 14.88 td  
## 22 1 39 15.19 td  
## 23 1 26 13.36 td  
## 24 1 45 14.84 td  
## 25 1 35 12.95 td  
## 26 1 36 12.79 td  
## 27 1 40 12.00 td  
## 28 1 34 13.98 td  
## 29 1 51 15.58 td  
## 30 1 58 18.60 td  
## 31 1 36 13.88 td  
## 32 1 40 13.84 td  
## 33 1 51 14.31 td  
## 34 1 47 13.87 td  
## 35 1 34 13.15 td  
## 36 1 61 15.63 td  
## 37 1 57 13.44 td  
## 38 1 57 17.81 td  
## 39 1 48 15.50 td  
## 40 1 49 14.94 td

reg2=lm(ns ~ kn, data=index2)  
summary(reg2)

##   
## Call:  
## lm(formula = ns ~ kn, data = index2)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.52257 -0.85968 -0.04064 0.64210 2.51469   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 8.96634 1.21647 7.371 1.7e-07 \*\*\*  
## kn 0.12274 0.02671 4.596 0.000127 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.182 on 23 degrees of freedom  
## Multiple R-squared: 0.4787, Adjusted R-squared: 0.456   
## F-statistic: 21.12 on 1 and 23 DF, p-value: 0.0001275

reg4=lm(ns ~ kn)  
summary(reg4)

##   
## Call:  
## lm(formula = ns ~ kn)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.8246 -0.8753 -0.1244 0.6925 2.7420   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 9.4422 1.0093 9.355 2.1e-11 \*\*\*  
## kn 0.1197 0.0223 5.367 4.2e-06 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.235 on 38 degrees of freedom  
## Multiple R-squared: 0.4312, Adjusted R-squared: 0.4162   
## F-statistic: 28.8 on 1 and 38 DF, p-value: 4.2e-06

# TINH thong ke F=(RSS-(RSS1+RSS2))/k:((RSS1+RSS2)/(n-2k))theo tieu chuan CHOW  
RSS =sum((resid(reg4))^2)  
RSS

## [1] 57.97922

RSS1 =sum((resid(reg1))^2)  
RSS1

## [1] 18.12697

RSS2 =sum((resid(reg2))^2)  
RSS2

## [1] 32.14474

F=(RSS-(RSS1+RSS2))/2:((RSS1+RSS2)/(40-4))  
F

## [1] 3.85375

# TINH GIA TRI P-VALUE CUA PHAN BO F(k, n-2k)  
p\_value= 1-pf(F, 2,36)  
p\_value

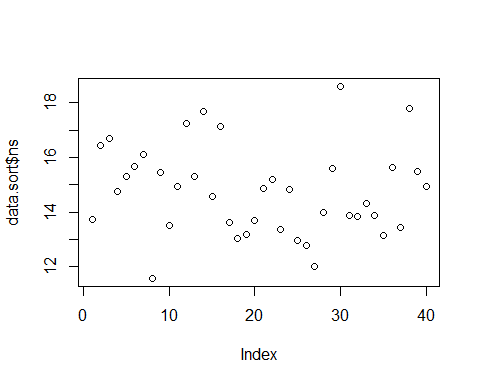
## [1] 0.0304395

# Gia tri toi han 5%  
F\_critical= pf(0.95,2,36)  
F\_critical

## [1] 0.6037774

# Ket luan the nao tu kiem dinh Chow?

## KIEM DINH CHOW IN R  
data.sort= data.frame(ch4bt9\_m[order(dn),])  
View(data.sort)  
plot(data.sort$ns)



sctest(data.sort$ns ~ data.sort$kn, type = "Chow", point = 16)

##   
## Chow test  
##   
## data: data.sort$ns ~ data.sort$kn  
## F = 3.6223, p-value = 0.03687

## KY THUAT BIEN GIA TEST BREAK STRUCTURE

reg5=lm(ns~kn+ dn+ dn\*kn)  
summary(reg5)

##   
## Call:  
## lm(formula = ns ~ kn + dn + dn \* kn)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.52257 -0.86105 -0.04694 0.67044 2.51469   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 9.993684 1.592754 6.274 2.99e-07 \*\*\*  
## kn 0.120022 0.035582 3.373 0.00179 \*\*   
## dn1 -1.027342 2.003854 -0.513 0.61130   
## kn:dn1 0.002718 0.044484 0.061 0.95161   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.182 on 36 degrees of freedom  
## Multiple R-squared: 0.5068, Adjusted R-squared: 0.4657   
## F-statistic: 12.33 on 3 and 36 DF, p-value: 1.064e-05

# KET LUAN NHU THE NAO?  
  
reg6=lm(ns~kn+dn)  
summary(reg6)

##   
## Call:  
## lm(formula = ns ~ kn + dn)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.51050 -0.86024 -0.04967 0.67546 2.52773   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 9.91727 0.97315 10.191 2.73e-12 \*\*\*  
## kn 0.12176 0.02106 5.780 1.25e-06 \*\*\*  
## dn1 -0.90718 0.38104 -2.381 0.0225 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.166 on 37 degrees of freedom  
## Multiple R-squared: 0.5067, Adjusted R-squared: 0.4801   
## F-statistic: 19 on 2 and 37 DF, p-value: 2.099e-06