Ch10bt14

tntrg\_03

2023-09-11

library(foreign)  
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(AER)

## Loading required package: sandwich

## Loading required package: survival

library(caret)#chia tệp số liệu một cách ngẫu nhiên

## Loading required package: ggplot2

## Loading required package: lattice

##   
## Attaching package: 'caret'

## The following object is masked from 'package:survival':  
##   
## cluster

library(caTools)

BAI TAP 10.14.2+3:XAC SUAT VO NO

MO HINH LOGIT: XAC SUAT VO NO. Data: mroz.txt

INPUT DATA & TRAIN.SET, TEST.SET

leha = leha <- read.csv("E:/KTL/KTL\_02/Chương 10/leha.csv", sep=";")  
#summary(leha)  
#View(leha)  
split=sample.split(leha$default\_prin,SplitRatio = 0.8)# chia tệp số liệu theo biến default\_prin  
#split  
train.set=subset(leha,split==TRUE)  
#summary(train.set)  
test.set=subset(leha,split==FALSE)  
#View(test.set)  
#View(train.set)  
dim(test.set)# Cho biết số dòng;số cột của dữ liệu

## [1] 813 25

dim(train.set)

## [1] 3253 25

dim(leha)

## [1] 4066 25

## UOC LUONG LOGIT TREN TRAIN.SET

logitres=glm(default\_int~income2+spending2+experience+age+newcustomer+notran+notran3+wpcity+wpcompany+wpothers+wpmanager+wpspecialist+wpother+cash+home100+liveparent+rent,data=train.set,family=binomial(link=logit))  
summary(logitres)

##   
## Call:  
## glm(formula = default\_int ~ income2 + spending2 + experience +   
## age + newcustomer + notran + notran3 + wpcity + wpcompany +   
## wpothers + wpmanager + wpspecialist + wpother + cash + home100 +   
## liveparent + rent, family = binomial(link = logit), data = train.set)  
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -0.4692151 0.5651662 -0.830 0.40641   
## income2 -0.0449283 0.0173198 -2.594 0.00949 \*\*   
## spending2 0.0140745 0.0120679 1.166 0.24350   
## experience -0.0008309 0.0009318 -0.892 0.37254   
## age -0.0150203 0.0124439 -1.207 0.22742   
## newcustomer 0.7695820 0.1681369 4.577 4.71e-06 \*\*\*  
## notran 1.3182006 0.8000671 1.648 0.09943 .   
## notran3 2.0947116 0.6533916 3.206 0.00135 \*\*   
## wpcity -0.2018221 0.2632329 -0.767 0.44326   
## wpcompany -0.8615588 0.2712975 -3.176 0.00149 \*\*   
## wpothers -0.2467781 0.3556509 -0.694 0.48776   
## wpmanager -0.8068794 0.3180099 -2.537 0.01117 \*   
## wpspecialist -0.6091553 0.3328324 -1.830 0.06722 .   
## wpother -0.3834934 0.2745722 -1.397 0.16251   
## cash -0.1494812 0.2049936 -0.729 0.46588   
## home100 -0.2536385 0.1963464 -1.292 0.19643   
## liveparent 0.0226987 0.1957256 0.116 0.90767   
## rent -0.2049216 0.2348471 -0.873 0.38290   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 2288.1 on 3251 degrees of freedom  
## Residual deviance: 2142.8 on 3234 degrees of freedom  
## (1 observation deleted due to missingness)  
## AIC: 2178.8  
##   
## Number of Fisher Scoring iterations: 5

## DIEU CHINH MO HNH

prob\_L.train=fitted(logitres)   
x=glm(default\_int~income2+age+newcustomer+notran3+wpcompany+wpmanager,data=train.set,family=binomial(link=logit))  
summary(x)

##   
## Call:  
## glm(formula = default\_int ~ income2 + age + newcustomer + notran3 +   
## wpcompany + wpmanager, family = binomial(link = logit), data = train.set)  
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -1.229830 0.332785 -3.696 0.000219 \*\*\*  
## income2 -0.029589 0.013149 -2.250 0.024428 \*   
## age -0.020090 0.007861 -2.556 0.010602 \*   
## newcustomer 0.811399 0.141209 5.746 9.13e-09 \*\*\*  
## notran3 1.954456 0.649583 3.009 0.002623 \*\*   
## wpcompany -0.590385 0.130000 -4.541 5.59e-06 \*\*\*  
## wpmanager -0.386801 0.131479 -2.942 0.003262 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 2288.1 on 3251 degrees of freedom  
## Residual deviance: 2159.6 on 3245 degrees of freedom  
## (1 observation deleted due to missingness)  
## AIC: 2173.6  
##   
## Number of Fisher Scoring iterations: 5

#(prob\_L.train)  
exp(-0.02891)

## [1] 0.9715039

#prob\_L.train.classes <- ifelse(prob\_L.train > 0.5, 1, 0)  
# Proportion of correctly classified observations:  
#x = cbind(prob\_L.test.classes,train.set$default\_prin)  
#x  
#TLDBD\_L.train=mean(prob\_L.train.classes == train.set$default\_int, na.rm=TRUE)  
#print("Ty lệ dự bao đúng trên train.set =")  
#TLDBD\_L.train

Với mức ý nghĩa 5%, trong điều kiện các yếu tố khác không đổi,khi tuổi tăng lên 1 đơn vị thì khả năng trả lãi vay theo log giảm 0.02891 hay khả năng trả lãi vay sẽ tăng theo 0.97

## ƯỚC LƯỢNG MÔ HÌNH TRÊN TEST.SET

#logit.res=glm(default\_int~income2+spending2+experience+age+newcustomer+notran+notran3+wpcity+wpcompany+wpothers+wpmanager+wpspecialist+wpother+cash+home100+liveparent+rent,data=test.set,family=binomial(link=logit))  
#summary(logit.res)  
#prob\_L.test = fitted(logit.res)  
prob\_L.test <- predict(logitres, test.set,type="response") # predicted scores   
length(prob\_L.test)

## [1] 813

prob\_L.test.classes <- ifelse(prob\_L.test > 0.5, 1,0)  
#prob\_L.test.classes  
# Proportion of correctly classified observations:  
TLDBD\_L.test=mean(prob\_L.test.classes == test.set$default\_int, na.rm=TRUE)  
print("Tỷ lệ dự báo đúng trên test.set=")

## [1] "Tỷ lệ dự báo đúng trên test.set="

TLDBD\_L.test

## [1] 0.8806888

## MOT SO PHAN TICH

print('Tính APE- AVERAGE PAARTIAL EFFECTS- ảnh hưởng riêng trung bình của từng biến')

## [1] "Tính APE- AVERAGE PAARTIAL EFFECTS- ảnh hưởng riêng trung bình của từng biến"

Aprob\_L=mean(prob\_L.train)  
APE\_L=coef(logitres)\*Aprob\_L  
print('APE=')

## [1] "APE="

APE\_L

## (Intercept) income2 spending2 experience age   
## -5.280834e-02 -5.056503e-03 1.584035e-03 -9.351507e-05 -1.690482e-03   
## newcustomer notran notran3 wpcity wpcompany   
## 8.661347e-02 1.483584e-01 2.357517e-01 -2.271429e-02 -9.696510e-02   
## wpothers wpmanager wpspecialist wpother cash   
## -2.777392e-02 -9.081115e-02 -6.855807e-02 -4.316070e-02 -1.682353e-02   
## home100 liveparent rent   
## -2.854603e-02 2.554649e-03 -2.306314e-02

print('Tính tỷ số OR, lấy đến 4 chữ số thập phân')

## [1] "Tính tỷ số OR, lấy đến 4 chữ số thập phân"

cbind(Estimate=round(coef(logitres),4),  
 OR=round(exp(coef(logitres)),4))

## Estimate OR  
## (Intercept) -0.4692 0.6255  
## income2 -0.0449 0.9561  
## spending2 0.0141 1.0142  
## experience -0.0008 0.9992  
## age -0.0150 0.9851  
## newcustomer 0.7696 2.1589  
## notran 1.3182 3.7367  
## notran3 2.0947 8.1231  
## wpcity -0.2018 0.8172  
## wpcompany -0.8616 0.4225  
## wpothers -0.2468 0.7813  
## wpmanager -0.8069 0.4462  
## wpspecialist -0.6092 0.5438  
## wpother -0.3835 0.6815  
## cash -0.1495 0.8612  
## home100 -0.2536 0.7760  
## liveparent 0.0227 1.0230  
## rent -0.2049 0.8147

#OR=p/(1-p):cho biết khả năng xảy ra sự kiện Y=1 bằng bao nhiêu lần so với sự kiện Y=0  
print('Khoảng tin cậy cho OR')

## [1] "Khoảng tin cậy cho OR"

exp(cbind(OR = coef(logitres), confint(logitres)))

## Waiting for profiling to be done...

## OR 2.5 % 97.5 %  
## (Intercept) 0.6254930 0.2003531 1.8425637  
## income2 0.9560661 0.9230165 0.9877453  
## spending2 1.0141741 0.9891632 1.0373966  
## experience 0.9991694 0.9970053 1.0003507  
## age 0.9850919 0.9636249 1.0111991  
## newcustomer 2.1588636 1.5604123 3.0182027  
## notran 3.7366915 0.5558396 15.0884534  
## notran3 8.1230982 2.0367617 28.3699856  
## wpcity 0.8172403 0.4947826 1.3941158  
## wpcompany 0.4225030 0.2516399 0.7315955  
## wpothers 0.7813140 0.3949513 1.6036050  
## wpmanager 0.4462484 0.2445787 0.8575329  
## wpspecialist 0.5438100 0.2885846 1.0719977  
## wpother 0.6814765 0.4028718 1.1865099  
## cash 0.8611547 0.5723092 1.2801186  
## home100 0.7759722 0.5312384 1.1487240  
## liveparent 1.0229583 0.7017138 1.5134461  
## rent 0.8147112 0.5142160 1.2935417

print('CIs using profiled log-likelihood')

## [1] "CIs using profiled log-likelihood"

confint(logitres) # khoảng tin cậy cho các hệ số ước lượng

## Waiting for profiling to be done...

## 2.5 % 97.5 %  
## (Intercept) -1.607673965 0.611157900  
## income2 -0.080108209 -0.012330365  
## spending2 -0.010895925 0.036714273  
## experience -0.002999154 0.000350619  
## age -0.037053171 0.011136861  
## newcustomer 0.444950055 1.104661508  
## notran -0.587275597 2.713929775  
## notran3 0.711361161 3.345331743  
## wpcity -0.703636838 0.332260406  
## wpcompany -1.379756087 -0.312527580  
## wpothers -0.928992828 0.472254189  
## wpmanager -1.408218078 -0.153695704  
## wpspecialist -1.242766900 0.069523931  
## wpother -0.909136916 0.171016167  
## cash -0.558075817 0.246952727  
## home100 -0.632544416 0.138651724  
## liveparent -0.354229645 0.414389240  
## rent -0.665111942 0.257383982

print('CIs using standard errors')

## [1] "CIs using standard errors"

confint.default(logitres)

## 2.5 % 97.5 %  
## (Intercept) -1.576920429 0.6384903002  
## income2 -0.078874406 -0.0109821396  
## spending2 -0.009578103 0.0377271796  
## experience -0.002657197 0.0009953879  
## age -0.039409961 0.0093692701  
## newcustomer 0.440039722 1.0991242116  
## notran -0.249902146 2.8863033242  
## notran3 0.814087706 3.3753355551  
## wpcity -0.717749126 0.3141050246  
## wpcompany -1.393292029 -0.3298254898  
## wpothers -0.943840983 0.4502847445  
## wpmanager -1.430167416 -0.1835914588  
## wpspecialist -1.261494900 0.0431843137  
## wpother -0.921644977 0.1546580840  
## cash -0.551261301 0.2522989743  
## home100 -0.638470317 0.1311932742  
## liveparent -0.360916383 0.4063137443  
## rent -0.665213578 0.2553703257

### Uoc luong probit tren train.set

probitres=glm(default\_int~income2+spending2+experience+age+newcustomer+notran+notran3+wpcity+wpcompany+wpothers+wpmanager+wpspecialist+wpother+cash+home100+liveparent+rent,data=train.set,family=binomial(link = probit))  
summary(probitres)

##   
## Call:  
## glm(formula = default\_int ~ income2 + spending2 + experience +   
## age + newcustomer + notran + notran3 + wpcity + wpcompany +   
## wpothers + wpmanager + wpspecialist + wpother + cash + home100 +   
## liveparent + rent, family = binomial(link = probit), data = train.set)  
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -0.3139362 0.3067602 -1.023 0.30612   
## income2 -0.0228837 0.0089082 -2.569 0.01020 \*   
## spending2 0.0072522 0.0064796 1.119 0.26304   
## experience -0.0004841 0.0004899 -0.988 0.32303   
## age -0.0071942 0.0065189 -1.104 0.26977   
## newcustomer 0.3960822 0.0836307 4.736 2.18e-06 \*\*\*  
## notran 0.6985076 0.4387328 1.592 0.11136   
## notran3 1.1641039 0.3962400 2.938 0.00330 \*\*   
## wpcity -0.1161094 0.1500685 -0.774 0.43910   
## wpcompany -0.4684318 0.1516945 -3.088 0.00202 \*\*   
## wpothers -0.1740155 0.2027860 -0.858 0.39082   
## wpmanager -0.4658257 0.1821193 -2.558 0.01053 \*   
## wpspecialist -0.3713405 0.1901123 -1.953 0.05079 .   
## wpother -0.2080909 0.1547226 -1.345 0.17865   
## cash -0.0871477 0.1015879 -0.858 0.39097   
## home100 -0.1328815 0.1061309 -1.252 0.21055   
## liveparent 0.0009043 0.1062456 0.009 0.99321   
## rent -0.1210233 0.1246860 -0.971 0.33174   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 2288.1 on 3251 degrees of freedom  
## Residual deviance: 2140.6 on 3234 degrees of freedom  
## (1 observation deleted due to missingness)  
## AIC: 2176.6  
##   
## Number of Fisher Scoring iterations: 6

#### Điều chỉnh mô hình

durbinWatsonTest(logitres) # kiểm định tự tương quan

## lag Autocorrelation D-W Statistic p-value  
## 1 0.06099306 1.877858 0  
## Alternative hypothesis: rho != 0

library(fBasics)

##   
## Attaching package: 'fBasics'

## The following object is masked from 'package:car':  
##   
## densityPlot

jarqueberaTest(probitres$residuals)

##   
## Title:  
## Jarque - Bera Normalality Test  
##   
## Test Results:  
## STATISTIC:  
## X-squared: 50817.434  
## P VALUE:  
## Asymptotic p Value: < 2.2e-16

#chisq.test(data(train.set),age,income2)

Dieu chinh mo hinh Probit

y=glm(default\_int~income2+newcustomer+notran+wpcompany+wpmanager,data=train.set,family=binomial(link = probit))  
summary(y)

##   
## Call:  
## glm(formula = default\_int ~ income2 + newcustomer + notran +   
## wpcompany + wpmanager, family = binomial(link = probit),   
## data = train.set)  
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -1.096584 0.102290 -10.720 < 2e-16 \*\*\*  
## income2 -0.015606 0.006531 -2.389 0.01688 \*   
## newcustomer 0.406117 0.069574 5.837 5.31e-09 \*\*\*  
## notran 0.596630 0.433370 1.377 0.16860   
## wpcompany -0.336130 0.066517 -5.053 4.34e-07 \*\*\*  
## wpmanager -0.184441 0.070047 -2.633 0.00846 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 2288.4 on 3252 degrees of freedom  
## Residual deviance: 2170.6 on 3247 degrees of freedom  
## AIC: 2182.6  
##   
## Number of Fisher Scoring iterations: 5

### =>hệ số ước lượng icome2 = -0.014063

pnorm(-0.014063)

## [1] 0.4943899

Do đó khi thu thập tăng thêm 1 đơn vị thì khả năng không trả lãi tăng thêm 0,49

f\_train=fitted(probitres)  
predicted\_train.classes=ifelse(f\_train>0.5,1,0)  
TLDBD\_train=mean(predicted\_train.classes==train.set$default\_int,na.rm = TRUE)

## Warning in predicted\_train.classes == train.set$default\_int: longer object  
## length is not a multiple of shorter object length

TLDBD\_train

## [1] 0.8874885

Dự báo trên Test.set

f\_test=predict(probitres,test.set,type = "response")  
predicted\_test.classes=ifelse(f\_test>0.5,1,0)  
TLDBD\_test=mean(predicted\_test.classes==test.set$default\_int,na.rm=TRUE)  
TLDBD\_test

## [1] 0.8806888