

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE CIENCIAS

TAREA 4: RIESGO DE CRÉDITO

Cedeño Ramírez Osvaldo
Felipe Santiago Edson Yair
Morelos Granados José Carlos
Pacheco Tovar Alejandra Citlalli

06/20/222

Ejercicio 3

1. Usando la tabla del archivo “Ejer_anal_discr.xls” (hoja base) que está en la página del curso, ajustar una regresión logística con las variables más significativas (pueden hacer uso de la hoja frecuencia del archivo “logit y disc datos.xls” y utilizando la prueba de Wald para cada variable), pueden usar el archivo “logitprobit.ipynb” para hacer el análisis, obviamente entregar los coeficientes de regresión, la gráfica de datos observados y ajustados, el nivel de significancia (Devianza) del modelo, es decir la prueba de hipótesis de que la Devianza se distribuye como una $X^2_{n-(m+1)}$, y la interpretación de las variables.

Solución

```
1 library(dplyr)
2 incump = "C:\\Users\\citla\\Downloads\\BD.csv"
3 data=read.csv(incump, sep=",", quote="/")
4 print(head(data))
5 ## Cliente Incump Liquid Apalanc AñosCl Rentab Bancos
6 PCP_CT
7 ## 1 1 0 1.418983 0.3678430 5.519115 0.082436996 2
8 0.9578792
9 ## 2 2 0 1.095459 0.6742535 5.365984 0.050120498 3
10 0.7856872
11 ## 3 3 0 1.232216 0.8788182 6.579059 0.047376582 1
12 0.6682592
13 ## 4 4 0 1.112399 0.6111619 3.980561 -0.010216081 2
14 0.9324931
15 ## 5 5 0 1.495808 0.2262790 5.861760 0.002987486 4
16 0.5764472
17 ## 6 6 0 1.106524 0.7106344 6.710337 0.046598640 3
18 0.7644960
19 ## Gop_SD RetrTMP PLP_PT IFAudit UO_VN AñoOper
20 ## 1 0.9619346 1 0.4255506 1 0.16267763 4
21 ## 2 0.8832112 1 0.3777689 0 0.20572504 4
22 ## 3 1.1103575 1 0.5652263 0 0.23805917 4
23 ## 4 1.0153539 1 0.3970641 1 0.17857237 5
24 ## 5 0.9057852 1 0.4313102 0 0.08164283 6
```

```

25 ## 6 1.0523312 0 0.3734093 1 0.24602347 4
26 n<-nrow(data)-1
27 m=ncol(data)
28 x=data[,3:m]
29 y=data[,2]
30 xext=as.matrix(cbind(1,x))
31
32 ylogr = glm(Incump ~ Liquid + Apalanc + AñosCl + Rentab
  ↳ +Bancos
33 + PCP_CT + Gop_SD + RetrTMP + PLP_PT + IFAudit +UO_VN +
  ↳ AñoOper ,
34 family=binomial("logit"), data = data)
35 ## Warning: glm.fit: algorithm did not converge
36 ## Warning: glm.fit: fitted probabilities numerically 0 or
  ↳ 1 occurred
37 summary(ylogr)
38 ##
39 ## Call:
40 ## glm(formula = Incump ~ Liquid + Apalanc + AñosCl +
  ↳ Rentab + Bancos +
41 ## PCP_CT + Gop_SD + RetrTMP + PLP_PT + IFAudit + UO_VN +
  ↳ AñoOper,
42 ## family = binomial("logit"), data = data)
43 ##
44 ## Deviance Residuals:
45 ## Min 1Q Median 3Q Max
46 ## -4.993e-05 -2.100e-08 -2.100e-08 2.100e-08 4.766e-05
47 ##
48 ## Coefficients:
49 ## Estimate Std. Error z value Pr(>|z|)
50 ## (Intercept) 86.097 345041.658 0.000 1.000
51 ## Liquid -35.348 83217.419 0.000 1.000
52 ## Apalanc 63.663 74575.329 0.001 0.999
53 ## AñosCl -7.861 16526.692 0.000 1.000
54 ## Rentab -395.216 893798.551 0.000 1.000
55 ## Bancos 8.519 48731.575 0.000 1.000

```

```

56 ## PCP_CT 47.308 128348.446 0.000 1.000
57 ## Gop_SD -73.761 111142.606 -0.001 0.999
58 ## RetrTMP -9.272 112996.503 0.000 1.000
59 ## PLP_PT -102.053 384818.303 0.000 1.000
60 ## IFAudit 16.038 39775.381 0.000 1.000
61
62 ## UO_VN -124.528 663892.384 0.000 1.000
63 ## AñoOper -3.374 30617.497 0.000 1.000
64 ##
65 ## (Dispersion parameter for binomial family taken to be
66 ↪ 1)
67 ##
68 ## Null deviance: 3.4231e+02 on 246 degrees of freedom
69 ## Residual deviance: 1.3106e-08 on 234 degrees of freedom
70 ## AIC: 26
71 ##
72 ## Number of Fisher Scoring iterations: 25

```

Listing 1: Carga y ajuste de datos

```

71 ## Number of Fisher Scoring iterations: 25
72
73 print(ylogr$coef)
74 ## (Intercept) Liquid Apalanc AñosCl Rentab
75 Bancos
76 ## 86.096523 -35.347612 63.663052 -7.861229 -395.215718
77 8.518809
78 ## PCP_CT Gop_SD RetrTMP PLP_PT IFAudit
79 UO_VN
80 ## 47.308495 -73.760760 -9.272432 -102.053079 16.038037 -
81 124.528160
82 ## AñoOper
83 ## -3.374335

```

Listing 2: Coeficientes de regresión

```

85 Llogit=log(prod(((1+exp(-xext%%as.matrix(ylogr$coef)))^(-1))^y*(1-
86 (1+exp(-xext%%as.matrix(ylogr$coef)))^(-1))^(1-y)))
87 DLogit=-2*Llogit
88 print(paste("Devianza replicada = ",DLogit))
89 ## [1] "Devianza replicada = 1.31061188997963e-08"
90 print(paste("Devianza del modelo lineal generalizado = ",
91   ↪ ylogr$dev))
92 ## [1] "Devianza del modelo lineal generalizado =
93   ↪ 1.3106114410409e-08"
94
95 ptol=0.5
96 yest=(1+exp(-xext%%as.matrix(ylogr$coef)))^(-1)
97 print("Matriz de confusión")
98 ## [1] "Matriz de confusión"
99 print(table(y,ifelse(yest<0.5,0,1)))
100 ##
101 ## y 0 1
102 ## 0 126 0
103 ## 1 0 121

```

Listing 3: Réplica de la devianza y la matriz de confusión

Utilizamos la devianza residual para observar si el modelo de regresión logística proporciona un ajuste adecuado para los datos. Para probar la significancia, podemos encontrar el p-value asociado.

```

107 pvalue = 1 - pchisq(5.9299, 234)
108 pvalue
109 ## [1] 1

```

Listing 4: P-value asociado

Los resultados indican que no hay evidencia para rechazar la hipótesis nula ni variables que sean significativas para predecir el incumplimiento de los clientes. Se va a utilizar un nuevo modelo utilizando la regresión múltiple para mejorar los resultados.

```

103 plot(y,type="l", lty=2, xlab="Observaciones",
104       ylab="Probabilidad", main="Regresión Logística")
105 points(yest, lwd=2, col="red")
106
107 pvalue = 1 - pchisq(5.9299, 234)
108 pvalue
109 ## [1] 1
110
111 ylogr = glm(Incump ~ Liquid + Apalanc + AñosCl + Rentab
112             ↪ +Bancos
113             + PCP_CT + Gop_SD + RetrTMP + PLP_PT + IFAudit +UO_VN +
114             ↪ AñoOper,
115             data = data)
116 summary(ylogr)
117 ##
118 ## Call:
119 ## glm(formula = Incump ~ Liquid + Apalanc + AñosCl +
120 ↪ Rentab + Bancos +
121 ## PCP_CT + Gop_SD + RetrTMP + PLP_PT + IFAudit + UO_VN +
122 ↪ AñoOper,
123 ## data = data)
124 ##
125 ## Deviance Residuals:
126 ## Min 1Q Median 3Q Max
127 ## -0.61744 -0.10534 -0.01249 0.10217 0.47844
128 ##
129 ## Coefficients:
130 ## Estimate Std. Error t value Pr(>|t|)
131 ## (Intercept) 0.311946 0.121002 2.578 0.01055 *
132 ## Liquid -0.071780 0.039473 -1.818 0.07027 .
133 ## Apalanc 0.307901 0.028870 10.665 < 2e-16 ***
134 ## AñosCl -0.015446 0.005509 -2.804 0.00548 **
135 ## Rentab -1.486105 0.301809 -4.924 1.60e-06 ***
136 ## Bancos 0.033695 0.014130 2.385 0.01789 *
137 ## PCP_CT 0.392136 0.042365 9.256 < 2e-16 ***

```

```

135 ## Gop_SD -0.422445 0.052626 -8.027 4.83e-14 ***
136 ## RetrTMP -0.002731 0.024758 -0.110 0.91225
137 ## PLP_PT 0.040351 0.076964 0.524 0.60057
138 ## IFAudit -0.007332 0.022477 -0.326 0.74456
139 ## UD_VN -0.202252 0.208894 -0.968 0.33394
140 ## AñoOper -0.017458 0.009273 -1.883 0.06099 .
141 ## ---
142 ## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ↪ ' ' 1
143 ##
144 ## (Dispersion parameter for gaussian family taken to be
  ↪ 0.02534142)
145 ##
146 ## Null deviance: 61.7247 on 246 degrees of freedom
147 ## Residual deviance: 5.9299 on 234 degrees of freedom
148 ## AIC: -192.2
149 ##
150 ## Number of Fisher Scoring iterations: 2

```

Listing 5: Carga y ajuste de datos

El apalancamiento, la liquidez, la antigüedad, la rentabilidad, el número de bancos, el pasivo corto plazo, la Generación Operativa y los años operando son significativos para el incumplimiento de los clientes. Realizamos un nuevo modelo.

```

153 data2 = data %>%
154   select(Apalanc, AñosCl, Rentab, PCP_CT, Gop_SD)
155 xext2=as.matrix(cbind(1,data2))
156 ylogr2 = glm(Incump ~ Apalanc + AñosCl + Rentab + PCP_CT +
157 Gop_SD, data = data)
158 summary(ylogr2)
159 ##
160 ## Call:
161 ## glm(formula = Incump ~ Apalanc + AñosCl + Rentab +
  ↪ PCP_CT + Gop_SD,
162 ## data = data)

```

```

163 ##
164 ## Deviance Residuals:
165 ## Min 1Q Median 3Q Max
166 ## -0.68280 -0.10883 0.00431 0.11333 0.46782
167 ##
168 ## Coefficients:
169 ## Estimate Std. Error t value Pr(>|t|)
170 ## (Intercept) 0.251781 0.088174 2.856 0.004671 **
171 ## Apalanc 0.330885 0.028521 11.601 < 2e-16 ***
172 ## AñosCl -0.018206 0.005451 -3.340 0.000971 ***
173 ## Rentab -1.718526 0.298939 -5.749 2.71e-08 ***
174 ## PCP_CT 0.414708 0.042167 9.835 < 2e-16 ***
175 ## Gop_SD -0.454613 0.051644 -8.803 2.64e-16 ***
176 ## ---
177 ## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
  ↪ ' ' 1
178 ##
179 ## (Dispersion parameter for gaussian family taken to be
  ↪ 0.02618935)
180 ##
181 ## Null deviance: 61.7247 on 246 degrees of freedom
182 ## Residual deviance: 6.3116 on 241 degrees of freedom
183 ## AIC: -190.79
184 ##
185 ## Number of Fisher Scoring iterations: 2

```

Listing 6: Carga y ajuste de datos

```

187 print(ylogr2$coef)
188 ## (Intercept) Apalanc AñosCl Rentab PCP_CT
189 Gop_SD
190 ## 0.25178090 0.33088544 -0.01820618 -1.71852649
  ↪ 0.41470833 -
191 0.45461331
192 #Replicamos la devianza.
193 Llogit2=log(prod(((1+exp(-xext2%*%as.matrix(ylogr2$coef)))^(-1)))^y*(1-

```



```

194 (1+exp(-xext2%%as.matrix(ylogr2$coef)))^(-1))^ (1-y)))
195 DLogit2=-2*Llogit2
196 print(paste("Devianza replicada = ",DLogit2))
197 ## [1] "Devianza replicada = 261.62491079492
198 print(paste("Devianza del modelo lineal generalizado = ",
  ↪ ylogr2$dev))
199 ## [1] "Devianza del modelo lineal generalizado =
  ↪ 6.31163348561568"

```

Listing 7: Coeficientes de regresión y devianza

```

201 ptol=0.5
202 yest2=(1+exp(-xext2%%as.matrix(ylogr2$coef)))^(-1)
203 print("Matriz de confusión")
204 ## [1] "Matriz de confusión"
205 print(table(y,ifelse(yest2<0.5,0,1)))
206 ##
207 ## y 0 1
208 ## 0 48 78
209 ## 1 0 121

```

Listing 8: Matriz de confusión

Obtenemos la precisión de nuestro modelo utilizando la diagonal de nuestra matriz de confusión.

```

205 print(table(y,ifelse(yest2<0.5,0,1)))
206 ##
207 ## y 0 1
208 ## 0 48 78
209 ## 1 0 121
210
211 plot(y,type="l", lty=2, xlab="Observaciones",
212      ylab="Probabilidad", main="Regresión Multinomial")
213 points(yest2, lwd=2, col="red")
214
215 (48+121)/(48+121+78)

```

```
216 ## [1] 0.6842105
217 pvalue2 = 1 - pchisq(6.3116, 241)
218 pvalue2
219 ## [1] 1
```

Listing 9: Matriz de confusión

Con precisión del 66.3966% de no evidencia para rechazar el modelo.