Texto

Descripción generada automáticamente

RIESGO DE CRÉDITO Y TITULARIZACIÓN DE CARTERA DE FACTORING

DE LAS ENTIDADES DE CRÉDITO EN COLOMBIA

CREDIT RISK AND FACTORING PORTFOLIO SECURITIZATION

OF CREDIT ENTITIES IN COLOMBIA

JULIO ANDRÉS ALZATE SANTA

JUAN GABRIEL MADRIGAL MORENO

Tesis para aspirar al título en Maestría en Administración Financiera

Asesor, docente

Germán Adolfo Peña Higuavita, CFA, FRM, SCR

UNIVERSIDAD EAFIT

ESCUELA DE ECONOMÍA Y FINANZAS

MAESTRÍA EN ADMINISTRACIÓN FINANCIERA - MAF

MEDELLÍN

2023

**Apéndice B. Cálculo probabilidad de prepago**

1. **Escenarios**

##Calculo probabilidad Prepago ##Tesis##

Se considera que existe una relación entre el hecho de que un cliente de factoring no vuelva a hacer operaciones de factoring (PREPAGO: no=1, si=0), los ingresos operacionales, Ebit/Cfo y la modalidad de intereses. Las variables del modelo son:

Prepago (Variable dependiente) = PREPAGO\_BIN

Ingresos operacionales = INGRESOS\_LN\_OP

Ebit/Cfo>0 = Ebit\_Cfo\_Positive

Modalidad de intereses = Modalidad\_int

Ingresos operacionales < COP 3000 = INGRESOS\_OP\_3000

Se requiere generar un modelo en el que a partir de las variables: ingresos operacionales, modalidad de interés, Ebit/Cfo>0 e ingresos operacionales <COP3000, prediga la probabilidad de que un cliente de factoring no vuelva a hacer operaciones de factoring (PREPAGO).

Utilizando “R Versión 4.2.1” como herramienta para realizar los cálculos, a continuación, se detalla el código utilizado:

data\_Default = read\_xlsx("LogitFileBin.xlsx")

View(data\_Default)

data\_Default$PREPAGO\_BIN = as.factor(data\_Default$PREPAGO\_BIN)

data\_Default$INGRESOS\_OP\_3000 = as.factor(data\_Default$INGRESOS\_OP\_3000)

data\_Default$Ebit\_Cfo\_Positive = as.factor(data\_Default$Ebit\_Cfo\_Positive)

1. **Análisis de las observaciones**

Con las tablas de frecuencia y las representaciones gráficas de las observaciones se intuye si las variables independientes escogidas están relacionadas con la variable respuesta y por tanto ser buenos predictores.

La siguiente es la tabla resumen de la segmentación de la variable dependente en función de las variables independientes:

Tabla 1. Tabla de frecuencia variables dependientes e independiente.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  | Ebit\_Cfo\_Positive | | | | Suma |
|  |  |  | NO | | SI | |
|  |  |  | ANT | VEN | ANT | VEN |
| INGRESOS\_OP\_3000 | 0 | 0 | 61 | 28 | 385 | 228 | 702 |
| 1 | 30 | 11 | 275 | 119 | 435 |
| 1 | 0 | 14 | 16 | 65 | 89 | 184 |
| 1 | 1 | 4 | 16 | 23 | 44 |
| Suma | | | 106 | 59 | 741 | 459 | 1365 |

Fuente: Elaboración propia

* 1. **Código en R para el desarrollo del análisis de las observaciones**

Utilizando “R Versión 4.2.1” como herramienta para realizar los cálculos, a continuación, se detalla el código utilizado:

## Analysis of observations

tabla1<- table(data\_Default$PREPAGO\_BIN, data\_Default$Modalidad\_int,

data\_Default$Ebit\_Cfo\_Positive, data\_Default$INGRESOS\_OP\_3000,

dnn = c("PREPAGO" , "MODALIDAD INT" , "Ebit\_Cfo\_Positive" ,

"INGRESOS\_OP\_3000"))

addmargins(tabla1)

# , , Ebit\_Cfo\_Positive = NO, INGRESOS\_OP\_3000 = 0

# MODALIDAD INT

# PREPAGO ANT VEN Sum

# 0 61 28 89

# 1 30 11 41

# Sum 91 39 130

# , , Ebit\_Cfo\_Positive = SI, INGRESOS\_OP\_3000 = 0

# MODALIDAD INT

# PREPAGO ANT VEN Sum

# 0 385 228 613

# 1 275 119 394

# Sum 660 347 1007

# , , Ebit\_Cfo\_Positive = Sum, INGRESOS\_OP\_3000 = 0

# MODALIDAD INT

# PREPAGO ANT VEN Sum

# 0 446 256 702

# 1 305 130 435

# Sum 751 386 1137

# , , Ebit\_Cfo\_Positive = NO, INGRESOS\_OP\_3000 = 1

# MODALIDAD INT

# PREPAGO ANT VEN Sum

# 0 14 16 30

# 1 1 4 5

# Sum 15 20 35

# , , Ebit\_Cfo\_Positive = SI, INGRESOS\_OP\_3000 = 1

# MODALIDAD INT

# PREPAGO ANT VEN Sum

# 0 65 89 154

# 1 16 23 39

# Sum 81 112 193

# , , Ebit\_Cfo\_Positive = Sum, INGRESOS\_OP\_3000 = 1

# MODALIDAD INT

# PREPAGO ANT VEN Sum

# 0 79 105 184

# 1 17 27 44

# Sum 96 132 228

# , , Ebit\_Cfo\_Positive = NO, INGRESOS\_OP\_3000 = Sum

# MODALIDAD INT

# PREPAGO ANT VEN Sum

# 0 75 44 119

# 1 31 15 46

# Sum 106 59 165

# , , Ebit\_Cfo\_Positive = SI, INGRESOS\_OP\_3000 = Sum

# MODALIDAD INT

# PREPAGO ANT VEN Sum

# 0 450 317 767

# 1 291 142 433

# Sum 741 459 1200

# , , Ebit\_Cfo\_Positive = Sum, INGRESOS\_OP\_3000 = Sum

# MODALIDAD INT

# PREPAGO ANT VEN Sum

# 0 525 361 886

# 1 322 157 479

# Sum 847 518 1365

* 1. **Tabla de frecuencias**

tabla\_frecuencias <- prop.table(tabla1)\*100

addmargins(tabla\_frecuencias)

# , , Ebit\_Cfo\_Positive = NO, INGRESOS\_OP\_3000 = 0

# MODALIDAD INT

# PREPAGO ANT VEN Sum

# 0 4.46886447 2.05128205 6.52014652

# 1 2.19780220 0.80586081 3.00366300

# Sum 6.66666667 2.85714286 9.52380952

# , , Ebit\_Cfo\_Positive = SI, INGRESOS\_OP\_3000 = 0

# MODALIDAD INT

# PREPAGO ANT VEN Sum

# 0 28.20512821 16.70329670 44.90842491

# 1 20.14652015 8.71794872 28.86446886

# Sum 48.35164835 25.42124542 73.77289377

# , , Ebit\_Cfo\_Positive = Sum, INGRESOS\_OP\_3000 = 0

# MODALIDAD INT

# PREPAGO ANT VEN Sum

# 0 32.67399267 18.75457875 51.42857143

# 1 22.34432234 9.52380952 31.86813187

# Sum 55.01831502 28.27838828 83.29670330

# , , Ebit\_Cfo\_Positive = NO, INGRESOS\_OP\_3000 = 1

# MODALIDAD INT

# PREPAGO ANT VEN Sum

# 0 1.02564103 1.17216117 2.19780220

# 1 0.07326007 0.29304029 0.36630037

# Sum 1.09890110 1.46520147 2.56410256

# , , Ebit\_Cfo\_Positive = SI, INGRESOS\_OP\_3000 = 1

# MODALIDAD INT

# PREPAGO ANT VEN Sum

# 0 4.76190476 6.52014652 11.28205128

# 1 1.17216117 1.68498168 2.85714286

# Sum 5.93406593 8.20512821 14.13919414

# , , Ebit\_Cfo\_Positive = Sum, INGRESOS\_OP\_3000 = 1

# MODALIDAD INT

# PREPAGO ANT VEN Sum

# 0 5.78754579 7.69230769 13.47985348

# 1 1.24542125 1.97802198 3.22344322

# Sum 7.03296703 9.67032967 16.70329670

# , , Ebit\_Cfo\_Positive = NO, INGRESOS\_OP\_3000 = Sum

# MODALIDAD INT

# PREPAGO ANT VEN Sum

# 0 5.49450549 3.22344322 8.71794872

# 1 2.27106227 1.09890110 3.36996337

# Sum 7.76556777 4.32234432 12.08791209

# , , Ebit\_Cfo\_Positive = SI, INGRESOS\_OP\_3000 = Sum

# MODALIDAD INT

# PREPAGO ANT VEN Sum

# 0 32.96703297 23.22344322 56.19047619

# 1 21.31868132 10.40293040 31.72161172

# Sum 54.28571429 33.62637363 87.91208791

# , , Ebit\_Cfo\_Positive = Sum, INGRESOS\_OP\_3000 = Sum

# MODALIDAD INT

# PREPAGO ANT VEN Sum

# 0 38.46153846 26.44688645 64.90842491

# 1 23.58974359 11.50183150 35.09157509

# Sum 62.05128205 37.94871795 100.00000000

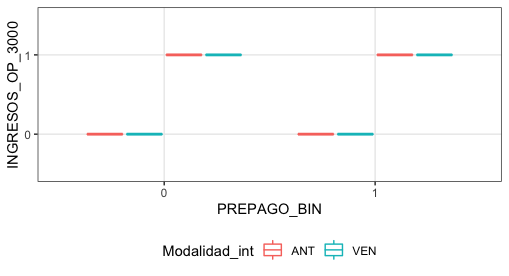
* 1. **Representación gráfica de las observaciones, (INGRESOS\_OP\_3000 Vs. PREPAGO\_BIN, Modalidad\_int)**

ggplot(data = data\_Default, aes(x = PREPAGO\_BIN, y = INGRESOS\_OP\_3000, colour = Modalidad\_int)) +

geom\_boxplot() + theme\_bw() + theme(legend.position = "bottom")

Ilustración 1. Representación gráfica de las observaciones.

(INGRESOS\_OP\_3000 Vs. DEFAULT\_BIN, Modalidad\_int)



Fuente: Elaboración propia Utilizando “R Versión 4.2.1”

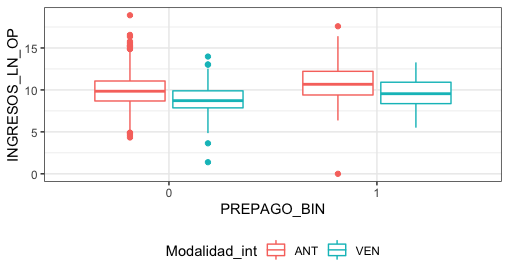
* 1. **Representación gráfica de las observaciones, (INGRESOS\_OP\_LN\_OP Vs. PREPAGO\_BIN, Modalidad\_int)**

ggplot(data = data\_Default, aes(x = PREPAGO\_BIN, y = INGRESOS\_LN\_OP, colour = Modalidad\_int)) +

geom\_boxplot() + theme\_bw() + theme(legend.position = "bottom")

Ilustración 2. Representación gráfica de las observaciones.

(INGRESOS\_LN\_OP Vs. PREPAGO\_BIN, Modalidad\_int)



Fuente: Elaboración propia Utilizando “R Versión 4.2.1”

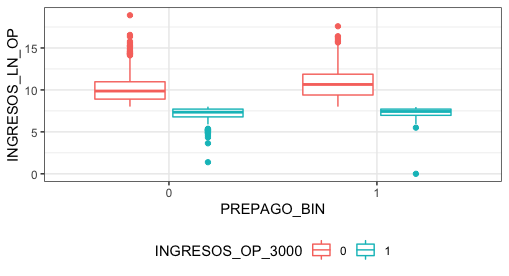
* 1. **Representación gráfica de las observaciones, (INGRESOS\_OP\_LN\_OP Vs. PREPAGO\_BIN, INGRESOS\_OP\_3000)**

ggplot(data = data\_Default, aes(x = PREPAGO\_BIN, y = INGRESOS\_LN\_OP, colour = INGRESOS\_OP\_3000)) +

geom\_boxplot() + theme\_bw() + theme(legend.position = "bottom")

Ilustración 3. Representación gráfica de las observaciones.

(INGRESOS\_LN\_OP Vs. PREPAGO\_BIN, INGRESOS\_OP\_3000)



Fuente: Elaboración propia Utilizando “R Versión 4.2.1”

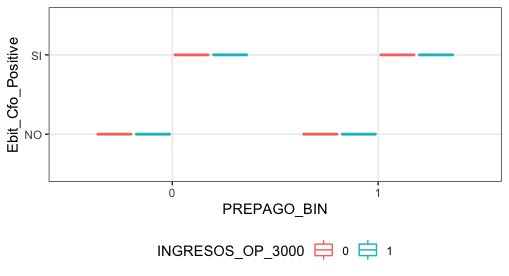
* 1. **Representación gráfica de las observaciones, (Ebit\_Cfo\_Positive Vs. PREPAGO\_BIN, INGRESOS\_OP\_3000)**

ggplot(data = data\_Default, aes(x = PREPAGO\_BIN, y = Ebit\_Cfo\_Positive, colour = INGRESOS\_OP\_3000)) +

geom\_boxplot() + theme\_bw() + theme(legend.position = "bottom")

Ilustración 3. Representación gráfica de las observaciones.

(Ebit\_Cfo\_Positive Vs. PREPAGO\_BIN, INGRESOS\_OP\_3000)



Fuente: Elaboración propia Utilizando “R Versión 4.2.1”

* 1. **Síntesis análisis de las observaciones**

El número de obligaciones de factoring con modalidades de intereses anticipadas es el 62,05% del total, el 37,95% es modalidad de interés vencida (847, 518). Es cerca de tres veces mayor el número de obligaciones en con modalidad vencida (1,098901, 2,857143). En promedio, es mayor el porcentaje de cartera con modalidad anticipada.

1. **Iteración de modelos Logit para encontrar el más ajustado:**

Utilizando “R Versión 4.2.1” como herramienta para realizar los cálculos, a continuación, se detalla el código utilizado:

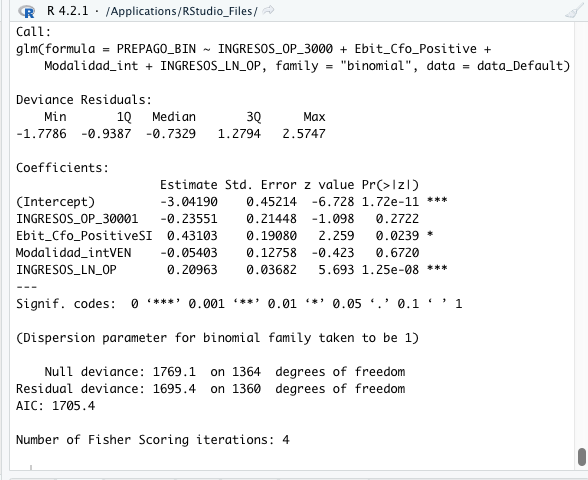
* 1. **Modelo Logit 1**

modeloT2\_glm = glm(PREPAGO\_BIN~INGRESOS\_OP\_3000 + Ebit\_Cfo\_Positive + Modalidad\_int + INGRESOS\_LN\_OP,

data = data\_Default, family = "binomial")

summary(modeloT2\_glm)

Ilustración 4. Modelo Logit 1



Fuente: Elaboración propia Utilizando “R Versión 4.2.1”

* + 1. **Interpretación Modelo Logit 1**

Acorde al modelo, el logaritmo de odds de que una obligación de factoring no vuelva a realizarse, esta negativamente relacionado con los ingresos operacionales inferiores a COP3000 millones (coeficiente parcial ), no es significativa esta relación (P-value = 0,2722). Es significativa la relación positiva entre el logaritmo de odds de que una obligación de factoring sea no vuelva a realizase (PREPAGO\_BIN) y Ebit\_Cfo\_PositiveSI (p-value = 0,0239); en cuanto la Modalidad de intereses, la modalidad vencida presenta una relación negativa pero no significativa (P-value = 0,6720) con respecto a la variable dependiente. Finalmente, los INGRESOS\_LN\_OP presentan una relación positiva y muy significativa (P-Value=1,25e-08).

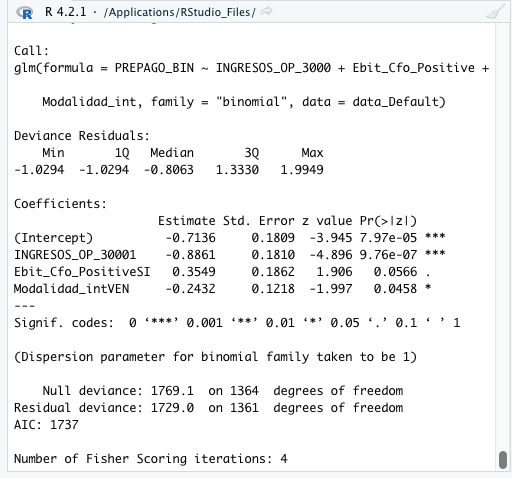
* 1. **Modelo Logit 2**

modeloT3\_glm = glm(PREPAGO\_BIN~INGRESOS\_OP\_3000 + Ebit\_Cfo\_Positive + Modalidad\_int, data = data\_Default,

family = "binomial")

summary(modeloT3\_glm)

Ilustración 5. Modelo Logit 2



Fuente: Elaboración propia Utilizando “R Versión 4.2.1”

* + 1. **Interpretación Modelo Logit 2**

Acorde al modelo, el logaritmo de odds de que una obligación de factoring no vuelva a realizarse, esta negativamente relacionado con los ingresos operacionales inferiores a COP3000 millones (coeficiente parcial ), es significativa esta relación (P-value = 9,76e-07). No es significativa la relación negativa entre el logaritmo de odds de que una obligación de factoring no vuelva a realizarse y Ebit\_Cfo\_PositiveSI (p-value = 0,0566); en cuanto la Modalidad de intereses, la modalidad vencida presenta una relación negativa y significativa (P-value = 0,0458) con respecto a la variable dependiente. En concreto los odds de que la modalidad de intereses vencida es 0.7841147 mayores que los anticipados. (Esto se puede ver gráficamente representado en el modelo para modalidad de intereses anticipada y vencida).

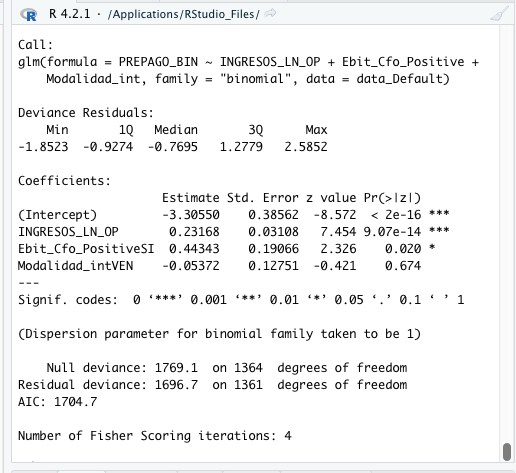
* 1. **Modelo Logit 3**

modeloT4\_glm = glm(PREPAGO\_BIN~INGRESOS\_LN\_OP + Ebit\_Cfo\_Positive + Modalidad\_int, data = data\_Default,

family = "binomial")

summary(modeloT4\_glm)

Ilustración 5. Modelo Logit 3

****

Fuente: Elaboración propia Utilizando “R Versión 4.2.1”

* + 1. **Interpretación Modelo Logit 3**

Acorde al modelo, el logaritmo de odds de que una obligación de factoring no vuelva a realizarse, esta positivamente relacionado con los INGRESOS\_LN\_OP (coeficiente parcial ), es significativa esta relación (P-value = 9,07e-14). Es significativa la relación positiva entre el logaritmo de odds de que una obligación de factoring no vuleva a realizarse y Ebit\_Cfo\_PositiveSI (p-value = 0,020); en cuanto la Modalidad de intereses, la modalidad vencida presenta una relación negativa y no significativa (P-value = 0,674) con respecto a la variable dependiente.

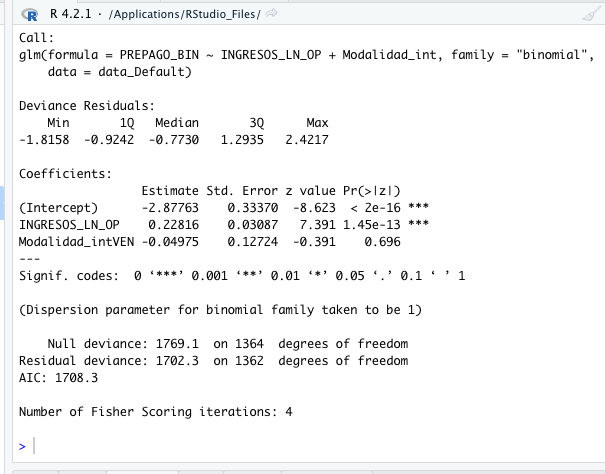
* 1. **Modelo Logit 4**

modeloT5\_glm <- glm(PREPAGO\_BIN~INGRESOS\_LN\_OP + Modalidad\_int, data = data\_Default,

family = "binomial")

summary(modeloT5\_glm)

Ilustración 5. Modelo Logit 4

****

Fuente: Elaboración propia Utilizando “R Versión 4.2.1”

* + 1. **Interpretación Modelo Logit 4**

Acorde al modelo, el logaritmo de odds de que una obligación de factoring no vuelva a realizarse, esta positivamente relacionado con los INGRESOS\_LN\_OP (coeficiente parcial ), es significativa esta relación (P-value = 1,45e-13). En cuanto la Modalidad de intereses, la modalidad vencida presenta una relación negativa y no significativa (P-value = 0,696) con respecto a la variable dependiente.

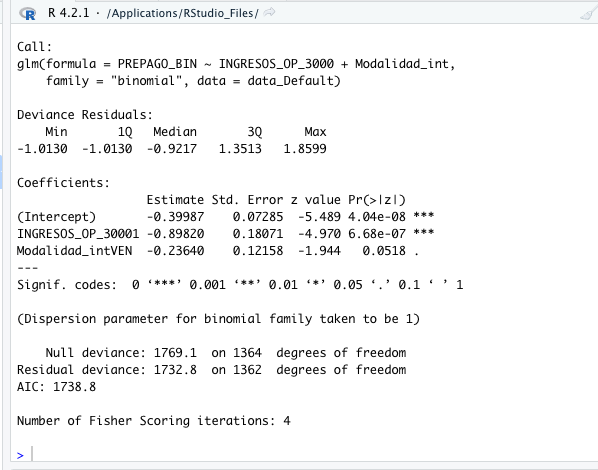
* 1. **Modelo Logit 5**

modeloT6\_glm <- glm(PREPAGO\_BIN~INGRESOS\_OP\_3000 + Modalidad\_int, data = data\_Default,

family = "binomial")

summary(modeloT6\_glm)

Ilustración 6. Modelo Logit 4

****

Fuente: Elaboración propia Utilizando “R Versión 4.2.1”

* + 1. **Interpretación Modelo Logit 5**

Acorde al modelo, el logaritmo de odds de que una obligación de factoring no vuelva a realizarse, esta negativamente relacionado con los INGRESOS\_OP\_30001 (coeficiente parcial ), es significativa esta relación (P-value = 6,68e-07). En cuanto la Modalidad de intereses, la modalidad vencida presenta una relación negativa y no significativa (P-value = 0,0518) con respecto a la variable dependiente.

* 1. **El modelo de regresión logística 6 (Dos Variables independientes)**

##Generar el modelo de regresión logística

## 1. Modelo Logit:

data\_Default$PREPAGO\_BIN <- as.factor(data\_Default$PREPAGO\_BIN)

data\_Default$Modalidad\_int <- as.factor(data\_Default$Modalidad\_int)

levels(data\_Default$Modalidad\_int)=c("ANT","VEN")

table(data\_Default$Modalidad\_int)

# ANT VEN

# 847 518

data\_Default$PREPAGO\_BIN <- relevel(data\_Default$PREPAGO\_BIN,"0")

data\_Default$Modalidad\_int <- relevel(data\_Default$Modalidad\_int,"ANT")

modeloT1\_glm<- glm(PREPAGO\_BIN~INGRESOS\_LN\_OP + Ebit\_Cfo\_Positive, data = data\_Default,

family = "binomial")

summary(modeloT1\_glm)

# [1] "coefficients" "residuals" "fitted.values"

# [4] "effects" "R" "rank"

# [7] "qr" "family" "linear.predictors"

# [10] "deviance" "aic" "null.deviance"

# [13] "iter" "weights" "prior.weights"

# [16] "df.residual" "df.null" "y"

# [19] "converged" "boundary" "model"

# [22] "call" "formula" "terms"

# [25] "data" "offset" "control"

# [28] "method" "contrasts" "xlevels"

# Call:

# glm(formula = PREPAGO\_BIN ~ INGRESOS\_LN\_OP + Ebit\_Cfo\_Positive,

# family = "binomial", data = data\_Default)

# Deviance Residuals:

# Min 1Q Median 3Q Max

# -1.8592 -0.9259 -0.7713 1.2785 2.6071

# Coefficients:

# Estimate Std. Error z value Pr(>|z|)

# (Intercept) -3.36437 0.35987 -9.349 < 2e-16 \*\*\*

# INGRESOS\_LN\_OP 0.23568 0.02963 7.955 1.8e-15 \*\*\*

# Ebit\_Cfo\_PositiveSI 0.44241 0.19063 2.321 0.0203 \*

# ---

# Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

# (Dispersion parameter for binomial family taken to be 1)

# Null deviance: 1769.1 on 1364 degrees of freedom

# Residual deviance: 1696.8 on 1362 degrees of freedom

# AIC: 1702.8

# Number of Fisher Scoring iterations: 4

* + 1. **Interpretación Modelo Logit 6**

Acorde al modelo, el logaritmo de odds de que una obligación de factoring no vuelva a realizarse, esta positivamente relacionado con los ingresos operacionales (coeficiente parcial ), siendo significativa esta relación (P-value = 1.8e-15). Existe una relación significativa y positiva entre el logaritmo de odds de que una obligación de factoring no vuelva a realizarse (PREPAGO\_BIN) y Ebit\_Cfo\_PositiveSI (p-value = 0.0203), siendo, para un mismo ingreso operacional, mayor si Ebit\_Cfo\_Positive es >0 (SI). En concreto los odds de Ebit\_Cfo\_PositiveSI son 1.556454 mayores que los <0 (NO). (Esto se puede ver gráficamente representado en el modelo para Ebit\_Cfo\_Positive >0 (SI) y <0 (NO)).

stargazer(modeloT1\_glm,type = "text")

# ===============================================

# Dependent variable:

# ---------------------------

# PREPAGO\_BIN

# -----------------------------------------------

# INGRESOS\_LN\_OP 0.236\*\*\*

# (0.030)

# Ebit\_Cfo\_PositiveSI 0.442\*\*

# (0.191)

# Constant -3.364\*\*\*

# (0.360)

# -----------------------------------------------

# Observations 1,365

# 3 Log Likelihood -848.418

# Akaike Inf. Crit. 1,702.836

# ===============================================

# Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

* + 1. **Identificación de la Ecuación**

# log(p/1-p) = - B0 + B1 (X1) + B2 (X2)

#log(p/1-p) = - 3.36437 + 0.23568 (X1) + 0.44241 (X2)

* + 1. **Significancia de las variables independientes**
* La variable INGRESOS\_LN\_OP es significativa (<1.8e-15\*\*\*<0,05, se rechaza la hipótesis nula, se concluye que si aporta).
* La variable Ebit\_Cfo\_PositiveSI es significativa (<0.0203 \* <0,05, se rechaza la hipótesis nula, se concluye que si aporta).
  + 1. **Ver que indica el coeficiente**

exp(modeloT1\_glm$coefficients)

# (Intercept) INGRESOS\_LN\_OP Ebit\_Cfo\_PositiveSI

# 0.03458371 1.26576638 1.55645602

Los odd Ratio de las variables independientes son: INGRESOS\_LN\_OP = 1.26576638; Ebit\_Cfo\_PositiveSI = 1.55645602

* + 1. **Interpretación de los coeficientes**

Por cada unidad que aumenta la variable "INGRESOS\_LN\_OP", el odds de que se presente el evento de Prepago aumenta en 1,26 veces.

Por cada unidad que aumenta la variable "Ebit\_Cfo\_PositiveSI", el odds de que se presente el evento de prepago aumenta 1.55645602 veces o aumenta un 1,556%.

Además del valor estimado de los coeficientes parciales de la correlación calculados por el modelo, es conveniente generar sus correspondientes intervalos de confianza. En el caso de la regresión logística, estos intervalos suelen calcularse basados en el profile lokelihood.

confint(modeloT1\_glm)

# 2.5 % 97.5 %

# (Intercept) -4.08213293 -2.6704880

# INGRESOS\_LN\_OP 0.17821711 0.2944343

# Ebit\_Cfo\_PositiveSI 0.07564519 0.8244167

En caso de querer los intervalos basados en el error estándar.

confint.default(modeloT1\_glm)

# 2.5 % 97.5 %

# (Intercept) -4.06970581 -2.6590391

# INGRESOS\_LN\_OP 0.17760899 0.2937466

# Ebit\_Cfo\_PositiveSI 0.06877964 0.8160433

* 1. **Representación gráfica del modelo Logit 6**

Al tratarse de un modelo con dos predictores, no se puede obtener una representación gráfica en 2D en la que se incluyan ambos predictores a la vez. Si es posible representar la curva del modelo logístico cuando se mantiene constante uno de los dos predictores. Por ejemplo, al representar las predicciones del modelo diferenciado entre Ebit\_Cfo\_Positive >0 (SI) y <0 (NO) (fijando el valor predictor de Ebit\_Cfo\_Positive) se aprecia que la curva de Ebit\_Cfo\_Positive >0 (SI) siempre está por encima. Esto se debe a que, como indica el coeficiente parcial de correlación del predictor Ebit\_Cfo\_Positive, para un mismo ingreso operacional, el logaritmo de ODDs de la operación de factoring no vuelva a realizarse es 1,55645602 veces mayor en la Modalidad de intereses anticipada.

##Representación gráfica del modelo

# Para graficar los valores en ggplot junto con la curva, la variable respuesta

# tiene que ser numérica en lugar de factor.

data\_Default$PREPAGO\_BIN <- as.numeric(as.character(data\_Default$PREPAGO\_BIN))

# Se crea un dataframe que contenga la probabilidad de que las operaciones de

# factoring entren en Default dada un determinado valor de Ingresos Operacionales

# y siendo Anticipado el Vector con nuevos valores interpolados en el rango de observaciones.

nuevos\_valores\_ingresos\_LN\_OP <- seq(from = min(data\_Default$INGRESOS\_LN\_OP),

to = max(data\_Default$INGRESOS\_LN\_OP), by = 0.5)

Ebit\_Cfo\_Positive<- as.factor(rep(x = "SI", length(nuevos\_valores\_ingresos\_LN\_OP)))

# Predicciones de los nuevos puntos según el modelo. type = "response" devuelve

# las predicciones en forma de probabilidad en lugar de en log\_ODDs.

prediccionesModeloT1 <- predict(object = modeloT1\_glm,

newdata=data.frame(INGRESOS\_LN\_OP=nuevos\_valores\_ingresos\_LN\_OP,

Ebit\_Cfo\_Positive = Ebit\_Cfo\_Positive),

type = "response")

# Se crea un data frame con los nuevos puntos y sus predicciones para graficar

# la curva.

datos\_curva\_SI <- data.frame(INGRESOS\_LN\_OP = nuevos\_valores\_ingresos\_LN\_OP,

Ebit\_Cfo\_Positive = Ebit\_Cfo\_Positive,

PREPAGO\_BIN = prediccionesModeloT1)

# Mismo proceso para vencido (Modalidad\_int = 0).

nuevos\_valores\_ingresos\_LN\_OP <- seq(from = min(data\_Default$INGRESOS\_LN\_OP),

to = max(data\_Default$INGRESOS\_LN\_OP), by = 0.5)

Ebit\_Cfo\_Positive<- as.factor(rep("NO", length(nuevos\_valores\_ingresos\_LN\_OP)))

prediccionesModeloT1 <- predict(object = modeloT1\_glm,

newdata=data.frame(INGRESOS\_LN\_OP=nuevos\_valores\_ingresos\_LN\_OP,

Ebit\_Cfo\_Positive = Ebit\_Cfo\_Positive),

type = "response")

datos\_curva\_NO <- data.frame(INGRESOS\_LN\_OP = nuevos\_valores\_ingresos\_LN\_OP,

Ebit\_Cfo\_Positive = Ebit\_Cfo\_Positive,

PREPAGO\_BIN = prediccionesModeloT1)

# Se unifican los dos dataframe.

datos\_curvaModeloT1 <- rbind(datos\_curva\_SI,datos\_curva\_NO)

ggplot(data = data\_Default, aes(x = INGRESOS\_LN\_OP, y = as.numeric(PREPAGO\_BIN),

color = Ebit\_Cfo\_Positive)) +

geom\_point() +

geom\_line(data = datos\_curvaModeloT1, aes(y = PREPAGO\_BIN)) +

geom\_line(data = datos\_curvaModeloT1, aes(y = PREPAGO\_BIN)) +

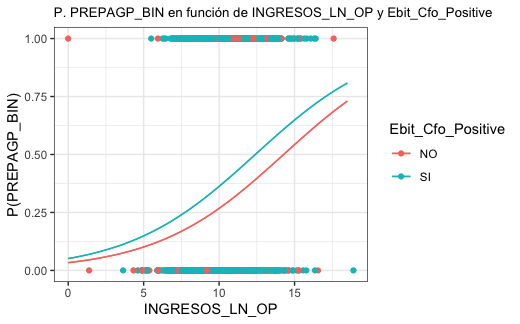
theme\_bw() +

labs(title = "P. PREPAGP\_BIN en función de INGRESOS\_LN\_OP y Ebit\_Cfo\_Positive",

y = "P(PREPAGP\_BIN)") +

theme(plot.title = element\_text(size = 10))

Ilustración 7. Representación gráfica del Modelo Logit 6



Fuente: Elaboración propia Utilizando “R Versión 4.2.1”

* 1. **Evaluación del Modelo Logit 6**

Likelihood ratio:

# Diferencia de residuos

dif\_residuos <- modeloT1\_glm$null.deviance - modeloT1\_glm$deviance

# Grados libertad

df <- modeloT1\_glm$df.null - modeloT1\_glm$df.residual

# p-value

p\_value <- pchisq(q = dif\_residuos,df = df, lower.tail = FALSE)

paste("Diferencia de residuos:", round(dif\_residuos, 4))

## [1] "Diferencia de residuos: 33.7579"

paste("Grados de libertad:", df)

## [1] "Grados de libertad: 2"

paste("p-value:", round(p\_value, 4))

## [1] "p-value: 0"

El modelo en conjunto si es significativo y acorde con los P-values mostrados en el summary(), también es significativa la contribución al modelo de ambos predictores.

* + 1. **Comparación de las predicciones con las observaciones**

Para este estudio de va a emplear un threshold de 0.5. Si la probabilidad predicha de Default es superior a 0,5, se asigna valor de 0 (Prepaga), si es menor se asigna el valor de 1 (No Prepaga).

predicciones <- ifelse(test = modeloT1\_glm$fitted.values > 0.5, yes = 1, no = 0)

matriz\_confusion <- table(modeloT1\_glm$model$PREPAGO\_BIN, predicciones,

dnn = c("observaciones", "predicciones"))

matriz\_confusion

# predicciones

# observaciones 0 1

# 0 827 59

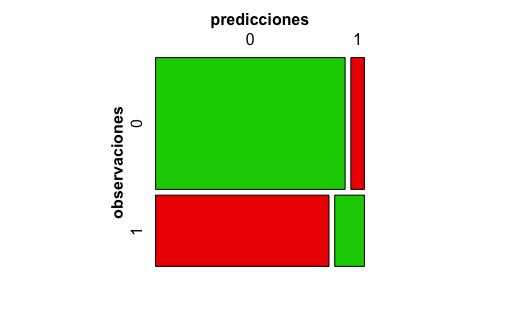
# 1 409 70

library(vcd)

mosaic(matriz\_confusion, shade = T, colorize = T,

gp = gpar(fill = matrix(c("green3", "red2", "red2", "green3"), 2, 2)))

Ilustración 8. Representación gráfica comparación de las predicciones



Fuente: Elaboración propia Utilizando “R Versión 4.2.1”

* + 1. **Interpretación de la evaluación del modelo**

El modelo es capaz de clasificar correctamente de las observaciones de entrenamiento. Si se analiza en detalle cómo se distribuye el error, se aprecia que el modelo ha sido capaz de identificar correctamente el 65,71% de las operaciones de factoring que no vuelva a realizarse (PREPAGO\_BIN). El porcentaje de falsos positivos es 34,29%. Sin embargo, se analizará por el método de threshold para observar la consistencia.

predicciones <- ifelse(test = modeloT1\_glm$fitted.values > 0.45, yes = 1, no = 0)

matriz\_confusion <- table(modeloT1\_glm$model$PREPAGO\_BIN, predicciones,

dnn = c("observaciones", "predicciones"))

matriz\_confusion

# predicciones

# observaciones 0 1

# 0 776 110

# 1 350 129

* + 1. **Conclusión**

El modelo logístico creado para predecir la probabilidad de que una operación de Factoring no vuelva a realizarse a partir de los ingresos operaciones y la Ebit\_Cfo\_Positive es en conjunto significativo acorde al Likelihood ratio (p-value = 0). El p-value de ambos predictores es significativo (INGRESOS\_LN\_OP = 1.8e-15, Ebit\_Cfo\_PositiveSI = 0.0203). La ratio de error obtenido empleando las observaciones con las que se ha entrenado el modelo muestra un porcentaje de falsos negativos bajo.

1. **Calcular las probabilidades:**

View(modeloT1\_glm)

modeloT1\_glm$fitted.values

predict\_modeloT1\_glm <- predict(modeloT1\_glm, type = "response")

View(predict\_modeloT1\_glm)

Nota: BD en Anexo Excel (anexo 7).