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 A. Cálculo probabilidad de default**

1. **Escenarios**

##Calculo probabilidad Default ##Tesis##

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

Default (Variable dependiente) = DEFAULT\_BIN

Ingresos operacionales = INGRESOS\_LN\_OP

Ebit/Cfo>0 = Ebit\_Cfo\_Positive

Modalidad de intereses = Modalidad\_int

Ingresos operacionales <COP3000 = 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 entre en default.

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$DEFAULT\_BIN = as.factor(data\_Default$DEFAULT\_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 | 3 | 4 | 7 | 23 | 37 |
| 1 | 88 | 35 | 653 | 324 | 1100 |
| 1 | 0 | 1 | 1 | 4 | 11 | 17 |
| 1 | 14 | 19 | 77 | 101 | 211 |
| 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$DEFAULT\_BIN, data\_Default$Modalidad\_int,

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

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

"INGRESOS\_OP\_3000"))

> addmargins(tabla1)

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

MODALIDAD INT

DEFAULT ANT VEN Sum

0 3 4 7

1 88 35 123

Sum 91 39 130

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

MODALIDAD INT

DEFAULT ANT VEN Sum

0 7 23 30

1 653 324 977

Sum 660 347 1007

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

MODALIDAD INT

DEFAULT ANT VEN Sum

0 10 27 37

1 741 359 1100

Sum 751 386 1137

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

MODALIDAD INT

DEFAULT ANT VEN Sum

0 1 1 2

1 14 19 33

Sum 15 20 35

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

MODALIDAD INT

DEFAULT ANT VEN Sum

0 4 11 15

1 77 101 178

Sum 81 112 193

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

MODALIDAD INT

DEFAULT ANT VEN Sum

0 5 12 17

1 91 120 211

Sum 96 132 228

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

MODALIDAD INT

DEFAULT ANT VEN Sum

0 4 5 9

1 102 54 156

Sum 106 59 165

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

MODALIDAD INT

DEFAULT ANT VEN Sum

0 11 34 45

1 730 425 1155

Sum 741 459 1200

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

MODALIDAD INT

DEFAULT ANT VEN Sum

0 15 39 54

1 832 479 1311

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

DEFAULT ANT VEN Sum

0 0.21978022 0.29304029 0.51282051

1 6.44688645 2.56410256 9.01098901

Sum 6.66666667 2.85714286 9.52380952

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

MODALIDAD INT

DEFAULT ANT VEN Sum

0 0.51282051 1.68498168 2.19780220

1 47.83882784 23.73626374 71.57509158

Sum 48.35164835 25.42124542 73.77289377

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

MODALIDAD INT

DEFAULT ANT VEN Sum

0 0.73260073 1.97802198 2.71062271

1 54.28571429 26.30036630 80.58608059

Sum 55.01831502 28.27838828 83.29670330

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

MODALIDAD INT

DEFAULT ANT VEN Sum

0 0.07326007 0.07326007 0.14652015

1 1.02564103 1.39194139 2.41758242

Sum 1.09890110 1.46520147 2.56410256

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

MODALIDAD INT

DEFAULT ANT VEN Sum

0 0.29304029 0.80586081 1.09890110

1 5.64102564 7.39926740 13.04029304

Sum 5.93406593 8.20512821 14.13919414

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

MODALIDAD INT

DEFAULT ANT VEN Sum

0 0.36630037 0.87912088 1.24542125

1 6.66666667 8.79120879 15.45787546

Sum 7.03296703 9.67032967 16.70329670

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

MODALIDAD INT

DEFAULT ANT VEN Sum

0 0.29304029 0.36630037 0.65934066

1 7.47252747 3.95604396 11.42857143

Sum 7.76556777 4.32234432 12.08791209

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

MODALIDAD INT

DEFAULT ANT VEN Sum

0 0.80586081 2.49084249 3.29670330

1 53.47985348 31.13553114 84.61538462

Sum 54.28571429 33.62637363 87.91208791

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

MODALIDAD INT

DEFAULT ANT VEN Sum

0 1.09890110 2.85714286 3.95604396

1 60.95238095 35.09157509 96.04395604

Sum 62.05128205 37.94871795 100.00000000

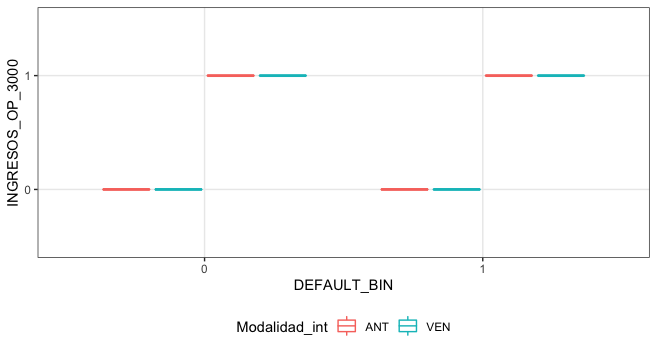
* 1. **Representación gráfica de las observaciones, (INGRESOS\_OP\_3000 Vs. DEFAULT\_BIN, Modalidad\_int)**

ggplot(data = Default, aes(x = DEFAULT\_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”

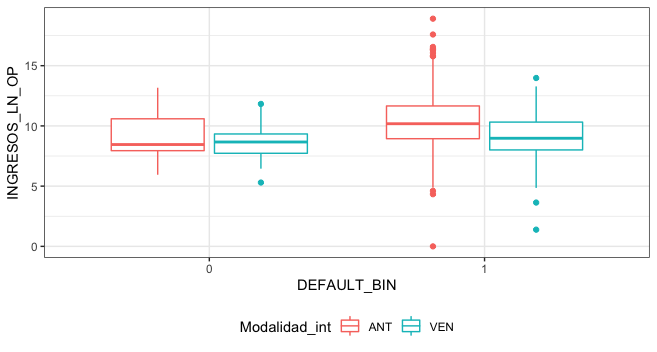
* 1. **Representación grafica de las observaciones, (INGRESOS\_LN\_OP Vs. DEFAULT\_BIN, Modalidad\_int)**

ggplot(data = data\_Default, aes(x = DEFAULT\_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. DEFAULT\_BIN, Modalidad\_int)



Fuente: Elaboración propia Utilizando “R Versión 4.2.1”

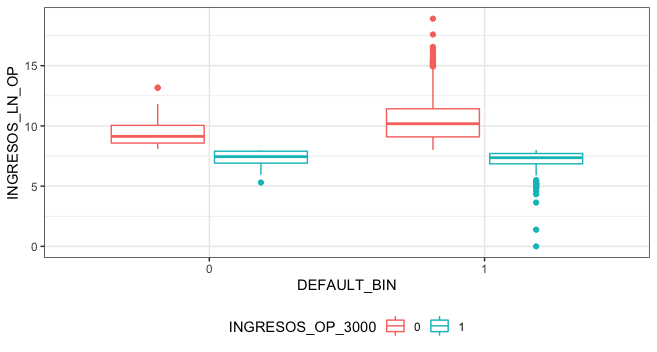
* 1. **Representación grafica de las observaciones, (INGRESOS\_LN\_OP Vs. DEFAULT\_BIN, INGRESOS\_OP\_3000)**

ggplot(data = data\_Default, aes(x = DEFAULT\_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. DEFAULT\_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 (Modalidad int, Ant: 847, Ven: 518). Es cerca de tres veces mayor el número de obligaciones en Default con modalidad vencida (Modalidad int. Ant: 1,098901, Ven: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(DEFAULT\_BIN~INGRESOS\_OP\_3000 + Ebit\_Cfo\_Positive + Modalidad\_int + INGRESOS\_LN\_OP, data = data\_Default,

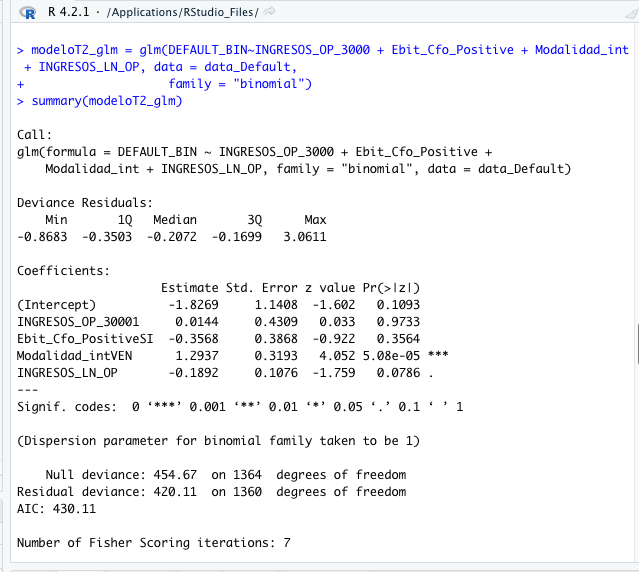
family = "binomial")

summary(modeloT2\_glm)

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

Acorde al modelo, el logaritmo de odds de que una obligación de factoring entre en default, esta positivamente relacionado con los ingresos operacionales inferiores a COP3000 millones (coeficiente parcial ), no es significativa esta relación (P-value = 0,09733). Tampoco es significativa la relación negativa entre el logaritmo de odds de que una obligación de factoring entre en default y Ebit\_Cfo\_PositiveSI (p-value = 0,3564); en cuanto la Modalidad de intereses, la modalidad vencida presenta una relación positiva y significativa (P-value = 5,08e-05) con respecto a la variable dependiente. En concreto los odds de que la modalidad de intereses vencida es 3.646253 mayores que los anticipados. (Esto se puede ver gráficamente representado en el modelo para modalidad de intereses anticipada y vencida). Finalmente, los INGRESOS\_LN\_OP presentan una relación negativa, no significativa (P-value=0,0786).

Ilustración 4. Modelo Logit 1

****

Fuente: Elaboración propia Utilizando “R Versión 4.2.1”

* 1. **Modelo Logit 2**

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

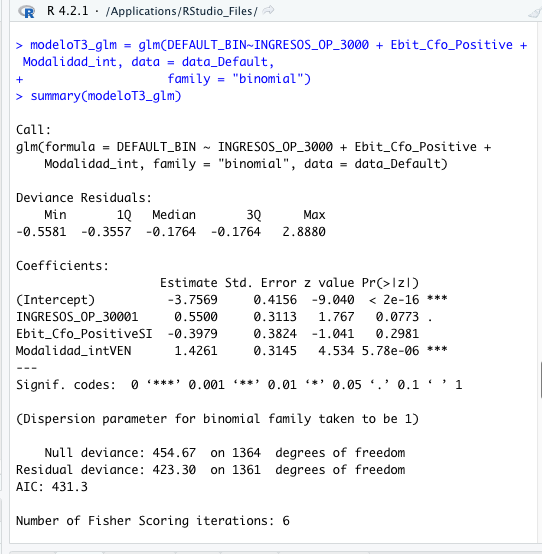
family = "binomial")

summary(modeloT3\_glm)

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

Acorde al modelo, el logaritmo de odds de que una obligación de factoring entre en default, esta positivamente relacionado con los ingresos operacionales inferiores a COP3000 millones (coeficiente parcial ), no es significativa esta relación (P-value = 0,0733). No es significativa la relación negativa entre el logaritmo de odds de que una obligación de factoring entre en default y la Ebit\_Cfo\_PositiveSI (p-value = 0,2981); en cuanto la Modalidad de intereses, la modalidad vencida presenta una relación positiva y significativa (P-value = 5,78e-06) con respecto a la variable dependiente. En concreto, los odds de la modalidad de intereses vencida son 4.162434 mayores que los anticipados. (Esto se puede ver gráficamente representado en el modelo para modalidad de intereses anticipada y vencida).

Ilustración 5. Modelo Logit 2



Fuente: Elaboración propia Utilizando “R Versión 4.2.1”

* 1. **Modelo Logit 3**

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

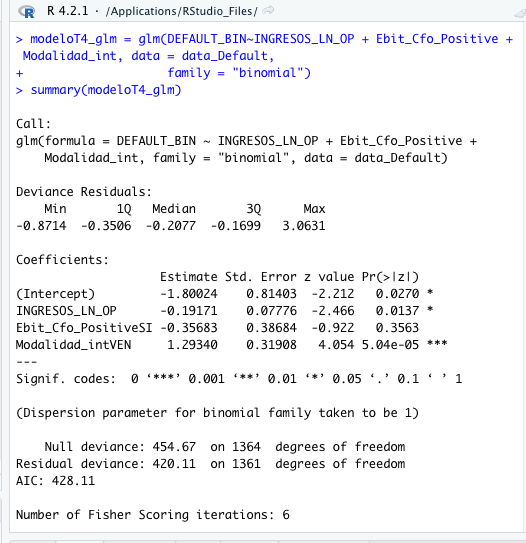
+ family = "binomial")

> summary(modeloT4\_glm)

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

Acorde al modelo, el logaritmo de odds de que una obligación de factoring entre en default, esta negativamente relacionado con los INGRESOS\_LN\_OP (coeficiente parcial ), es significativa esta relación (P-value = 0,0137). No es significativa la relación negativa entre el logaritmo de odds de que una obligación de factoring entre en default y la Ebit\_Cfo\_PositiveSI (p-value = 0,3563); en cuanto la Modalidad de intereses, la modalidad vencida presenta una relación positiva y significativa (P-value = 5,04e-05) con respecto a la variable dependiente. En concreto los odds de la modalidad de intereses vencida son 3.645159 mayores que los anticipados. (Esto se puede ver gráficamente representado en el modelo para modalidad de intereses anticipada y vencida).

Ilustración 5. Modelo Logit 3

****

Fuente: Elaboración propia Utilizando “R Versión 4.2.1”

* 1. **Modelo Logit 4**

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

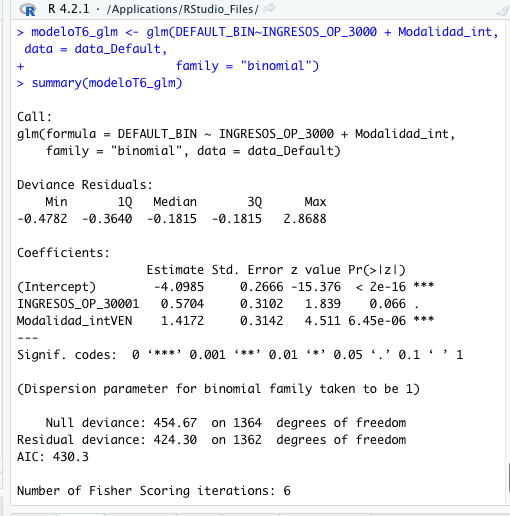
+ family = "binomial")

> summary(modeloT6\_glm)

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

Acorde al modelo, el logaritmo de odds de que una obligación de factoring entre en default, esta positivamente los ingresos operade relacionado con los INGRESOS\_OP\_30001 (coeficiente parcial ), no es significativa esta relación (P-value = 0,066). En cuanto la Modalidad de intereses, la modalidad vencida presenta una relación positiva y significativa (P-value = 5,45e-06) con respecto a la variable dependiente. En concreto, los odds de la modalidad de intereses vencida son 4.125553 mayores que los anticipados. (Esto se puede ver gráficamente representado en el modelo para modalidad de intereses anticipada y vencida).

Ilustración 6. Modelo Logit 4

****

Fuente: Elaboración propia Utilizando “R Versión 4.2.1”

* 1. **El modelo de regresión logística 5 (Dos Variables independientes)**

##Generar el modelo de regresión logística

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

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

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

table(data\_Default$Modalidad\_int)

data\_Default$DEFAULT\_BIN <- relevel(data\_Default$DEFAULT\_BIN,"1")

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

modeloT1\_glm <- glm(DEFAULT\_BIN~INGRESOS\_LN\_OP + Modalidad\_int, data = data\_Default,

family = "binomial")

names(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"

summary(modeloT1\_glm)

Call:

glm(formula = DEFAULT\_BIN ~ INGRESOS\_LN\_OP + Modalidad\_int, family = "binomial",

data = data\_Default)

Deviance Residuals:

Min 1Q Median 3Q Max

-0.7870 -0.3489 -0.2103 -0.1716 3.0569

Coefficients:

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

(Intercept) -2.01820 0.79175 -2.549 0.0108 \*

INGRESOS\_LN\_OP -0.20081 0.07836 -2.563 0.0104 \*

Modalidad\_intVEN 1.28307 0.31853 4.028 5.62e-05 \*\*\*

---

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

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 454.67 on 1364 degrees of freedom

Residual deviance: 420.91 on 1362 degrees of freedom

AIC: 426.91

Number of Fisher Scoring iterations: 6

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

Acorde al modelo, el logaritmo de odds de que una obligación de factoring entre en default, esta negativamente relacionado con los ingresos operacionales (coeficiente parcial ), siendo significativa esta relación (P-value = 0,0104). Existe una relación significativa negativa entre el logaritmo de odds de que una obligación de factoring entre en default y la modalidad de intereses (p-value = 5,62e-05), siendo, para un mismo ingreso operacional, mayor si la modalidad de intereses es vencida. En concreto los odds de que la modalidad de intereses vencida es 3.607698 mayor que los anticipados. (Esto se puede ver gráficamente representado en el modelo para modalidad de intereses anticipada y vencida).

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

=============================================

Dependent variable:

---------------------------

DEFAULT\_BIN

---------------------------------------------

INGRESOS\_LN\_OP -0.201\*\*

(0.078)

Modalidad\_intVEN 1.283\*\*\*

(0.319)

Constant -2.018\*\*

(0.792)

---------------------------------------------

Observations 1,365

Log Likelihood -210.455

Akaike Inf. Crit. 426.909

=============================================

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

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

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

log(p/1-p) = -2.01820 - 0.20081 (X1) + 1.28307 (X2)

* + 1. **Significancia de las variables independientes:**
* La variable INGRESOS\_LN\_OP es significativa (<0.00104 \*\*<0,05, se rechaza la hipótesis nula, se concluye que si aporta).
* la variable Modalidad\_intVEN es significativa (<5.62e-05 \*\*\*<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 Modalidad\_intVEN

# 0.1328938 0.8180660 3.6076808

Los odd Ratio de las variables independientes son: INGRESOS\_LN\_OP = 0.8180660; Modalidad\_intVEN = 3.6076808

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

Por cada unidad que aumenta la variable "INGRESOS\_LN\_OP", el odds de que se presente el evento de Default disminuye en 1,22 veces. (inversa de 0.8180660= 1/0.8180660=1,2224).

Por cada unidad que aumenta la variable "Modalidad\_intVEN", el odds de que se presente el evento aumenta 3.6076808 veces o aumenta un 3,607%.

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, se calcularon basados en el profile lokelihood.

confint(modeloT1\_glm)

2.5 % 97.5 %

(Intercept) -3.6180414 -0.51102255

INGRESOS\_LN\_OP -0.3543402 -0.04711573

Modalidad\_intVEN 0.6790692 1.93645598

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

confint.default(modeloT1\_glm)

2.5 % 97.5 %

(Intercept) -3.5700022 -0.46640774

INGRESOS\_LN\_OP -0.3543863 -0.04723821

Modalidad\_intVEN 0.6587555 1.90737477

* 1. **Representación gráfica del modelo Logit 5**

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 Modalidad de interés Anticipada y Modalidad de interés vencida (fijando el valor predictor de Modalidad de intereses) se aprecia que la curva de Modalidad de intereses Anticipado siempre está por encima. Esto se debe a que, como indica el coeficiente parcial de correlación del predictor Modalidad de intereses, para un mismo ingreso operacional, el logaritmo de ODDs de la operación de factoring entrar en default es 3.6076808 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$DEFAULT\_BIN <- as.numeric(as.character(data\_Default$DEFAULT\_BIN))

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

# factoring entren en Default dada una 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)

Modalidad\_int<- as.factor(rep(x = "ANT", 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,

Modalidad\_int = Modalidad\_int),

type = "response")

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

# la curva.

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

Modalidad\_int = Modalidad\_int,

DEFAULT\_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)

Modalidad\_int<- as.factor(rep("VEN", length(nuevos\_valores\_ingresos\_LN\_OP)))

prediccionesModeloT1 <- predict(object = modeloT1\_glm,

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

Modalidad\_int = Modalidad\_int),

type = "response")

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

Modalidad\_int = Modalidad\_int,

DEFAULT\_BIN = prediccionesModeloT1)

# Se unifican los dos dataframe.

datos\_curvaModeloT1 <- rbind(datos\_curva\_ANT,datos\_curva\_VEN)

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

color = Modalidad\_int)) +

geom\_point() +

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

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

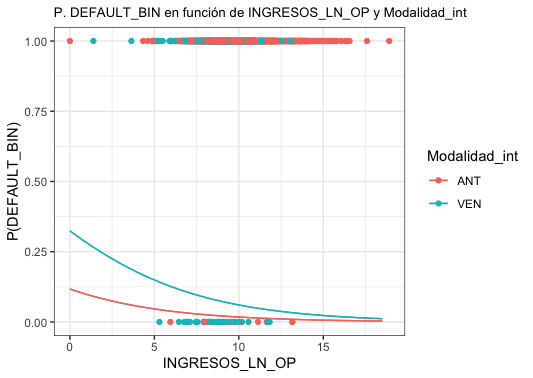
theme\_bw() +

labs(title = "P. DEFAULT\_BIN en función de INGRESOS\_LN\_OP y Modalidad\_int",

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

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

Ilustración 7. Representación gráfica del Modelo Logit 5



Fuente: Elaboración propia Utilizando “R Versión 4.2.1”

* 1. **Evaluación del Modelo**

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

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

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

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

matriz\_confusion

## predicciones

## observaciones 0

## 1 1311

## 0 54

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 todas las operaciones de factoring que no entraron en default. El porcentaje de falsos positivos es 0. 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$model$DEFAULT\_BIN, predicciones,

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

matriz\_confusion

# predicciones

# observaciones 0

# 1 1311

# 0 54

* 1. **Conclusión**

El modelo logístico creado para predecir la probabilidad de que una operación de Factoring entre en default a partir de los ingresos operaciones y la modalidad de intereses es en conjunto significativo acorde al Likelihood ratio (p-value = 0). El p-value de ambos predictores es significativo (INGRESOS\_LN\_OP = 0.0104, Modalidad\_intVEN = 5.62e-05). La ratio de error obtenido empleando las observaciones con las que se ha entrenado el modelo muestra un porcentaje de falsos negativos muy 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)