

# **El Modelo Logit Mixto para la Construcción de un Scoring de Crédito**

**Por:**

**Sandra Moreno Valencia**

**Universidad Nacional de Colombia  
Sede Medellín**

# Agenda

- 1. Objetivos de la investigación**
- 2. Scoring de Crédito y Default**
- 3. Modelos Estadísticos utilizados**
- 4. Modelo Logit Mixto: Aplicación**
- 5. Conclusiones y Recomendaciones**

# 1. Objetivos

- ✓ Estimar un modelo estadístico para el scoring de crédito que tenga un mayor poder de discriminación de los clientes que llegan a default (entidad financiera).
- ✓ Identificar cuales son los factores que determinan el estado de default para los clientes de una entidad financiera del sector cooperativo del departamento de Antioquia.
- ✓ Estimar un modelo logit mixto, como propuesta de un modelo más adecuado para un scoring de crédito, en comparación con los modelos logit y probit.

## 2. Scoring de Crédito y Default

Los modelos tipo scoring son instrumentos de clasificación o puntuación utilizados por las entidades financieras en la decisión de otorgar un crédito.

Para una entidad financiera, el Scoring de Crédito es una herramienta muy poderosa, puesto que permite la estandarización de una parte fundamental del proceso de otorgamiento, la estimación de la probabilidad de incumplimiento total o *default* que puede llegar a tener el cliente.

## 2. Scoring de Crédito y Default

*Default*: Estado en que entra el deudor cuando tiene una mora mayor a 90 días en cualquiera de las obligaciones crediticias que tenga en la entidad (Circular Externa 100, Capítulo II. Superintendencia Financiera de Colombia).

La predicción del *Default* es muy importante en el SARC, ya que permite anticipar el posible comportamiento del cliente desde el inicio de la relación financiera.

## 2. Scoring de Crédito y Default

De acuerdo con los diferentes trabajos empíricos, esta probabilidad se puede estimar considerando las características del individuo (demográficas, sociales, financieras) y del crédito que éste solicita.

- ✓ El modelo logit (frecuentemente utilizado para modelar el scoring de créditos), asume supuestos con respecto a la varianza del error que no son sostenibles.
- ✓ El modelo logit mixto permite levantar esos supuestos al considerar un término de variabilidad adicional.

Autores	Aplicación
Beaver (1966)	Análisis univariado de los indicadores financieros que podrían influir en la insolvencia de una entidad.
Altman (1968)	Análisis discriminante multivariado para el problema de la predicción de la quiebra corporativa.
Orgler (1970)	Análisis de regresión lineal para los créditos comerciales vigentes de una entidad financiera.
Orgler (1971)	Análisis de regresión lineal para los créditos de consumo especiales.
Chandler y Coffman (1979)	Análisis comparativo entre el scoring de crédito y el juicio de un evaluador de créditos. Clasificación en dos grupos.
Ohlson (1980)	Modelo logit para pronosticar la probabilidad de Default de los créditos (metodología estándar).
Abdou (2009) y Greene (1998)	Análisis discriminante y regresiones logística y probit para investigar la eficiencia (clasificación correcta de los deudores buenos y malos), de los modelos tipo scoring para créditos de consumo y comerciales, respectivamente.

### 3. Modelos estadísticos utilizados

- Modelos de elección discreta basados en una función índice (logit y probit). Estos asumen que para cualquier individuo, dado un conjunto de atributos, hay una probabilidad definida de que realmente entrará en *Default* durante la vigencia del crédito.
- En  $n$  deudores, se identifican varias características  $X_1, X_2, \dots, X_p$  de los individuos y del crédito ( $p$  variables).
- *Default*:  $y \in \{0, 1\}$ , donde a los deudores que no entraron en *Default* se les asigna el valor de 0 y los que cayeron en *Default* el valor de 1.



### 3. Modelos estadísticos utilizados

El resultado observado, Default/No-default, se debe a las características y el comportamiento aleatorio de los individuos:

$$D_i^* = \beta' x_i + \varepsilon_i$$

$D_i^*$ : Variable dependiente → “propensión al Default”

$x_i$ : Variables independientes → ingresos mensuales del individuo, historial de crédito, estado civil, etc.

$\varepsilon_i$ : término de error.

### 3. Modelos estadísticos utilizados

La probabilidad de interés es:

$$Prob[D_i = 1|X_i] = Prob[D > 0|X_i] = Prob[\varepsilon_i \leq \beta' x_i|x_i] = F(\beta' x_i)$$

$\varepsilon_i$  se distribuye  $F(\cdot)$  con media  $\mathbf{0}$  y varianza  $\sigma^2$ .

$F(\cdot)$ : función de distribución desconocida de  $x_i$ .

Regla de Predicción  $\rightarrow D_i = 1$  si  $F(\beta' x_i) > c^*$ ,  
donde  $c^*$  es un valor umbral determinado.

### 3. Modelos estadísticos utilizados

Si  $F(\cdot) \sim$  Logística  $\rightarrow$  Modelo Logit:

$$F(\beta'x) = \frac{e^{\beta'x}}{1+e^{\beta'x}} \rightarrow \varepsilon_i \text{ (i. i. d.)} \sim \text{logist.}, E(\varepsilon|x) = 0, \text{Var}(\varepsilon|x) = \pi^2/3$$

Si  $F(\cdot) \sim$  Normal estándar  $\rightarrow$  Modelo Probit:

$$F(\beta'x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\beta'x} e^{-\frac{t^2}{2}} dt \rightarrow \varepsilon_i \text{ (i. i. d.)} \sim \text{Normal}, E(\varepsilon|x) = 0, \text{Var}(\varepsilon|x) = 1$$

### 3. Modelos estadísticos utilizados

#### Limitaciones del modelo logit:

- ✓ Supone parámetros fijos o constantes, por lo tanto no admite variaciones aleatorias en las preferencias o características de los individuos.
- ✓ No admite correlación entre las observaciones.
- ✓ Para el caso multinomial, supone independencia de alternativas irrelevantes, asumiendo que la razón entre probabilidades de dos alternativas no depende de las alternativas restantes.

## 4. Modelo Logit Mixto

$$y_i^* = \mathbf{x}_i' \boldsymbol{\beta}_i + \varepsilon_i$$

Con  $\boldsymbol{\beta}_i = \boldsymbol{\beta} + \boldsymbol{\theta} \mathbf{v}_i$ , donde  $\mathbf{v}_i \sim i.i.d. (\mathbf{0}, \mathbf{I})$

Cada deudor tiene una solvencia no observada  $y_i^*$ , modelada como una variable aleatoria con media condicional  $\mathbf{x}_i' \boldsymbol{\beta}_i$

$$y_i = \begin{cases} 1, & \text{si } y_i^* > c \\ 0, & \text{si } y_i^* \leq c \end{cases}$$

donde  $c$  es un umbral o punto de corte.

Los parámetros  $\boldsymbol{\beta}_i$  se asumen aleatorios y tienen una pdf conjunta  $f(\boldsymbol{\beta}_i | \boldsymbol{\beta}, \boldsymbol{\theta})$ .

## 4. Modelo Logit Mixto:

- $\varepsilon_i$  i.i.d  $\sim$  cdf  $\Lambda(\cdot)$
- Las medias de los parámetros individuales se incluyen especificando  $\boldsymbol{\beta}_i = \boldsymbol{\beta} + \mathbf{z}'_i \boldsymbol{\gamma} + \boldsymbol{\theta} \mathbf{v}_i$

$\boldsymbol{\theta}$ : descomposición de Cholesky de matriz de covarianza  $\mathbf{V}(\boldsymbol{\beta}_i)$ .

- Las distribuciones marginales para cada  $\beta_{i,k}$  pueden ser especificadas libremente.
- Por máxima verosimilitud simulada se hallan las estimaciones de  $\tilde{\boldsymbol{\beta}}$  y  $\tilde{\boldsymbol{\theta}}$  y las probabilidades

$$P(y_i = 1 | \mathbf{x}_i, \tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\theta}}) = \int_{\tilde{\boldsymbol{\beta}}_i} \Lambda(\mathbf{x}'_i \boldsymbol{\beta}_i) f(\boldsymbol{\beta}_i | \tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\theta}}) d\boldsymbol{\beta}_i$$

lo cual es equivalente a  $P(y_i = 1 | \mathbf{x}_i, \boldsymbol{\beta}_i) = \Lambda(\mathbf{x}'_i \boldsymbol{\beta}_i) = \frac{e^{\mathbf{x}'_i \boldsymbol{\beta}_i}}{\sum e^{\mathbf{x}'_i \boldsymbol{\beta}_i}}$

## 4. Modelo Logit Mixto: Aplicación

Autores	Aplicación
Dakovic, Czado y Berg (2009), Altman y Sabato (2005) y Lennox (1999)	Muestran que realizando transformaciones no lineales de las variables independientes y teniendo varianzas heteroscedásticas, se puede mejorar significativamente la predicción de los modelos de crédito para el <i>Default</i> .
Porath (2004) y Lennox (1999)	Afirman que la elección de la familia de distribución del término de error al parecer tiene poco impacto en el desempeño de los modelos.
Jones y Hensher (2004)	Con un enfoque multinomial y ordinal se estima la proporción de individuos que pertenecen a un determinado grupo, en lugar de las probabilidades individuales.
Kukuk y Roennberg (2013)	Comparación entre la aplicación del modelo logit tradicional y el logit mixto para un modelo de default de crédito empresarial. El logit mixto captura la heterogeneidad presente en los datos.

## 4. Modelo Logit Mixto: Aplicación

Para la estimación de los modelos se cuenta con una muestra de 10.841 clientes de créditos de consumo de mediana cuantía concedidos entre julio de 2010 y junio de 2011, en una entidad financiera del sector cooperativo del departamento de Antioquia.

La muestra contiene la variable respuesta DEFAULT y 24 variables que describen los atributos o características demográficas y financieras de los clientes.

	No-default	Default	Total
Número Clientes	10.461	380	10.841
Porcentaje	96,49%	3,51%	100,00%



## 4. Modelo Logit Mixto: Aplicación

Las variables explicativas se dividen en seis factores:

- Factor Crediticio: valor solicitado, plazo y tipo de garantía presentada.
- Factor Financiero: ingreso mensual y nivel de endeudamiento.
- Factor de Comportamiento Crediticio: créditos vigentes, meses último crédito, calificación en central de riesgos, promedio de días mora, etc.
- Factor Laboral: tipo de ocupación, antigüedad laboral y tipo de contrato.
- Factor Socio-demográfico: sexo, edad, estrato, tipo de vivienda, estado civil, nivel de estudios, entre otras.
- Factor Económico: sector económico y zona.

## 4. Modelo Logit Mixto: Aplicación

Variable	No-Default (10.461 observaciones)			Default (380 observaciones)		
	$\bar{X}$	$X_{med}$	$\sigma$	$\bar{X}$	$X_{med}$	$\sigma$
Porc. Endeud.	71,01	81,05	27,78	60,98	49,54	29,17
VlrCrédito (\$)	8.006.000	5.077.000	9.370.007	7.960.000	5.522.000	7.652.626
Plazo (meses)	43,86	48,00	12,20	47,73	48,00	13,37
Ingresos (\$)	1.289.000	865.000	1.602.194	1.175.000	905.700	952.249
Edad (años)	45,17	43,79	14,76	55,33	57,64	14,06
Mora máx. (d)	20,13	13,00	22,56	211,50	144,00	153,08
Prom.mora(d)	4,23	1,57	5,96	36,22	20,37	42,17
Créd.Vigentes	2,23	2,00	1,63	4,04	4,00	1,87
Calific. Central Riesgo	9,01	10,00	2,31	7,70	10,00	3,61

## 4. Modelo Logit Mixto: Aplicación

Variable	No-Default			Default		Test $\chi^2$ (Valor p)
	Número clientes	Porcent.		Número clientes	Porcent.	
<b>Sexo</b>						
Masculino	5.371	51,34%		219	57,63%	5,80 (0,015)
Femenino	5.090	48,66%		161	42,37%	
<b>Pers. Activas</b>						
1	2.945	28,15%		137	36,05%	13,45 (0,003)
2	4.924	47,07%		171	45,00%	
3	1.830	17,50%		51	13,42%	
4 o más	762	7,28%		21	5,53%	
<b>Pers. Cargo</b>						
0	3.537	33,81%		106	27,90%	12,84 (0,024)
1	3.381	32,32%		134	35,26%	
2	2.139	20,45%		93	24,47%	
3 o más	1.404	13,42%		47	12,37%	

## 4. Modelo Logit Mixto: Aplicación

<b>Tipo Vivienda</b>						
<b>Arrendada</b>	1.314	12,56%		66	17,37%	13,57 (0,001)
<b>Familiar</b>	3.703	35,40%		105	27,63%	
<b>Propia</b>	5.444	52,04%		209	55,00%	
<b>Estado Civil</b>						
<b>Casado</b>	4.307	41,17%		178	46,84%	25,72 (3,6 e-07)
<b>Divorciado</b>	657	6,28%		30	7,90%	
<b>Soltero</b>	3.584	34,26%		90	23,68%	
<b>Unión Libre</b>	1.192	11,40%		40	10,53%	
<b>Viudo</b>	721	6,89%		42	11,05%	
<b>Nivel de Estudios</b>						
<b>Universitario</b>	1.571	15,02%		60	15,79%	53,12 (8,0 e-13)
<b>Tecnológico</b>	1.789	17,10%		52	13,69%	
<b>Secundaria</b>	4.968	47,49%		134	35,26%	
<b>Primaria</b>	2.133	20,39%		134	35,26%	

## 4. Modelo Logit Mixto: Aplicación

Modelo: Default ~ variables del factor	Seudo-R <sup>2</sup>
*Factor Crediticio: log (Valor Crédito), Plazo, Garantía	0,0217
*Factor Financiero: log( Ingresos), Endeudamiento	0,0144
*Factor Comportamiento Crediticio: Créditos Vigentes, Meses Último Crédito, Calificación Central de Riesgo, Promedio de Mora, Moras Mayores a 30 días	0,5726
*Factor Laboral: Estado Laboral, Antigüedad Laboral	0,0624
*Factor Socio- demográfico: Sexo, Edad, Estrato, Estado Civil, Nivel de Estudio, Personas Activas, Personas a Cargo, Antigüedad en la Cooperativa, Tipo de Vivienda	0,0914
*Factor Económico: Sector, Zona	0,0747

## 4. Modelo Logit Mixto: Aplicación

Variable	Modelo Logit	Modelo Probit	Modelo Logit Mixto
Intercepto	-9.451 (2.153) p.v.=1.14e-05***	-4.821 (1.150) p.v.=2.78e-5***	-8.790 (2.118) p.v.=3.34e-05 ***
Log(VLR.CREDITO)	-0.080 (0.145) p.v. =0.158	-0.071 (0.076) p.v. = 0.135	-0.094 (0.141) p.v.= 0.150
PLAZO	0.022 (0.009) p.v. =0.020 *	0.016 (0.005) p.v. = 0.001 **	0.023 (0.009) p.v. = 0.013 *
Grupo 2 INGRESOS	-0.563 (0.248) p.v. =0.023 *	-0.291 (0.129) p.v. = 0.024 *	-0.586 (0.247) p.v.=0.018 *
CRED.VIGENT	0.690 (0.067) p.v.=< 2e-16 ***	0.365 (0.035) p.v.=< 2e-16 ***	0.704 (0.067) p.v.< 2e-16 ***
PROM.MORA	0.266 (0.014) p.v.=< 2e-16 ***	0.142 (0.007) p.v.=< 2e-16 ***	0.264 (0.013) p.v.< 2e-16 ***
CENTRAL.RIESGO	-0.117 (0.028) p.v.=2.71e-05***	-0.066 (0.015) p.v.=< 2e-16 ***	-0.115 (0.028) p.v.=2.9e-05 ***
Grupo 2 EDAD	0.323 (0.272) p.v. =0.235	0.006 (0.005) p.v. =0.241	0.250 (ef.alet) St. Dev.=0.286

## 4. Modelo Logit Mixto: Aplicación

Variable	Modelo Logit	Modelo Probit	Modelo Logit Mixto
Grupo3 EDAD	0.759 (0.340) p.v. = 0.026 *	n.i.	0.250 (ef.alet) St. Dev.=0.286
EST.CIVIL: UNIONL	0.583 (0.296) p.v. = 0.048 *	0.295 (0.155) p.v. = 0.057 .	0.566 (0.293) p.v. = 0.053 .
EST.CIVIL: VIUDO	n.i.	-0.330 (0.173) p.v. = 0.056 .	n.i.
NIV.ESTUDIO: PRIMARIA	0.452 (0.227) p.v. = 0.046 *	0.373 (0.120) p.v. = 0.002 **	0.133 (ef.alet) St. Dev.= 0.184
NIV.ESTUDIO: UNIVERSITARIO	0.412 (0.254) p.v. = 0.105	n.i.	0.133 (ef.alet) St. Dev.= 0.184
EST. LABORAL: EMPL- CONTR INDEFIN	-1.207 (0.364) p.v. = 0.0009 ***	n.i.	-0.434 (ef.alet) St. Dev.= 0.496
EST.LABORAL: INDEPENDIENTE	-1.243 (0.608) p.v. = 0.041 *	-0.784 (0.346) p.v. = 0.023 *	-0.434 (ef.alet) St. Dev.= 0.496
EST.LABORAL: PENSIONADO	-0.498 (0.332) p.v. = 0.011	-0.310 (0.174) p.v. = 0.074 .	-0.434 (ef.alet) St. Dev.= 0.496
TIPO.VIV: ARRENDADA	0.548 (0.263) p.v.=0.037 *	0.186 (0.143) p.v. = 0.195	0.512 (0.262) p.v. = 0.051 *

## 4. Modelo Logit Mixto: Aplicación

Variable	Modelo Logit	Modelo Probit	Modelo Logit Mixto
MESES.ULTCR: 0-6	0.929 (0.252) p.v. =0.0002 ***	0.494 (0.115) p.v.=2.04e-5***	0.368 (ef.alet) St. Dev.= (0.353)
MESES.ULTCR: 19 ó más	0.394 (0.376) p.v. =0.294	n.i.	0.368 (ef.alet) St. Dev.= (0.353)
MESES.ULTCR: Nuevo	0.873 (0.307) p.v. =0.004 **	0. (0.153) p.v. = 0.006 **	0.368 (ef.alet) St. Dev.= (0.353)
ZONA: SUR	-0.378 (0.254) p.v. =0.137	-0.197 (0.134) p.v. = 0.142	-0.397 (0.252) p.v. =0.116
SECTOR: GOBIERNO	0.568 (0.266) p.v. =0.032 *	0.398 (0.143) p.v. = 0.005 **	0.699 (0.256) p.v. =0.006 **
SECTOR: TEXTIL	-2.700 (1.052) p.v. =0.010 *	-2.422 (0.981) p.v. = 0.013 *	-2.594 (1.031) p.v. =0.012 *
<b>Desviación Estándar Parámetros de efectos aleatorios (modelo logit mixto)</b>			
Grupo EDAD			$\sigma^2 = 0.08214$
NIV.ESTUDIO			$\sigma^2 = 0.03401$
EST.LABORAL			$\sigma^2 = 0.41550$
MESES.ULTCR			$\sigma^2 = 0.12519$



## 4. Modelo Logit Mixto: Aplicación

Medidas de Validación de los modelos:

Medidas	Modelo Logit	Modelo Probit	Modelo Logit Mixto
Pseudo $-R^2$	0.630	0.655	0.796
AIC	1002.05	915.21	1019.00
BIC (Shwartz)	1199.89	1113.06	1175.00
AUROC	0.946	0.956	0.957

Medidas del Poder predictivo de los modelos:

Medida	Modelo Logit	Modelo Probit	Modelo Logit Mixto
Tasa de aciertos	97.46%	97.46%	97.33%
Tasa de errores	2.54%	2.54%	2.67%
Especificidad	99.66%	99.23%	98.99%
Sensibilidad	40.74%	51.85%	54.32%
Tasa de falsos ceros	2.25%	1.85%	1.76%
Tasa de falsos unos	17.50%	27.59%	32.31%

## 5. Conclusiones

- Los tres modelos estimados: logit tradicional, probit y logit mixto, tienen un buen poder predictivo, reflejado en las altas tasas de aciertos, sobre todo para los clientes morosos.
- El modelo logit mixto resultó ser el de mayor sensibilidad (predicción de los verdaderos positivos), aunque también predijo el mayor número de falsos positivos.

- Las variables que determinan que un cliente llegue a default, con un nivel de significancia de 0.10, son las relacionadas con el factor de comportamiento crediticio, financiero y demográfico, como se esperaba (además, los signos resultaron acordes con la realidad de la entidad financiera).

- Para la entidad financiera es muy importante contar con esta herramienta estadística adecuada para la predicción del comportamiento de los clientes al momento de otorgarles el crédito, puesto que la rentabilidad y los flujos de caja, en gran medida corresponden al correcto pago de las obligaciones crediticias contraídas por parte de los clientes.

- El modelo logit mixto es el más potente en la predicción o detección de los clientes que llegan a estado de default, pero esta condición está asociada a que es un modelo muy estricto en la aceptación de clientes óptimos (no-default), lo que genera un gran porcentaje de rechazo de clientes que en su historial crediticio han pagado bien (Error tipo I). Esto podría ocasionar en el largo plazo un problema de crecimiento de mercado para la entidad financiera.