

Modelos de Riesgo de Crédito con Stata

Economía Financiera | FCE-UBA

Santiago Rodríguez Vázquez

2026

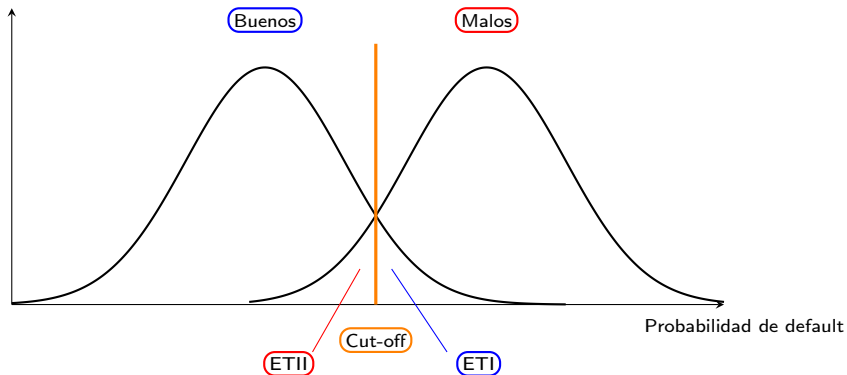
- **Data editor**: permite visualizar y editar cada variable y el valor que toma en cada registro.
- **History**: historial de comandos. Útil para replicar comandos indicados anteriormente.
- **Variables**: ventana con las variables (y sus etiquetas) de la base.
 - Con doble clic las llamamos en la ventana de comandos.
- **Propiedades**: ventana con propiedades de las variables (tipo de almacenamiento, formato, etc.).
 - Botón *lock/unlock* para editar.
- **Command**: ventana donde indicamos los comandos a Stata.

Comandos básicos y de estadística descriptiva

```
generate      // crear nuevas variables
drop          // eliminar variables
clear        // limpia la memoria
describe      // tipo de variable
sum           // estadística descriptiva de la base
sum , detail  // idem, ms detallado
histogram x   // grafica histograma de la variable "x"
tab x         // tabla de frecuencias
```

- `sum , detail` reporta:
 - Tendencia central (media / mediana)
 - Variabilidad
 - Forma (curtosis y asimetría)
- Importación de `.xls` y apertura de base `.dta`.
 - File > Import > Excel

Score y gestión del cut-off



- El problema de gestión: **elegir el cut-off**.
 - Se debe definir la probabilidad de default a partir de la cual se rechaza a un cliente. Mientras menor sea esa probabilidad, más exigente es el criterio.

Matriz de confusión

		Realidad	
		Malo (+)	Bueno (-)
Predicción	Malo (+)	OK	ETI
	Bueno (-)	ETII	OK

- **ETI**: el modelo dice que el sujeto va a defaultear cuando en verdad era buen pagador.
 - Me pierdo de otorgar crédito y no obtengo el rendimiento en intereses.
- **ETII**: el modelo dice que el sujeto es buen pagador cuando en verdad va a defaultear.
 - Al otorgar el crédito pierdo el capital (no considero como pérdida el rendimiento).

- Si reduzco el cut-off mejoro la calidad de la cartera (menos pérdidas por incobrables), pero aumenta la cantidad de individuos que caen en *ETI*.
 - Menos colocaciones $\implies$ menor incobrabilidad. La ganancia operativa contable mejora porque solo capta la menor incobrabilidad; no considera el rendimiento perdido por créditos no otorgados (*ETI*).
 - En el agregado, con la caída de colocaciones la ganancia en el balance podría caer.
- Si aumento el cut-off otorgo más préstamos a individuos que el modelo clasificaba como malos pero eran buenos (cae el *ETI*), pero el modelo comete más *ETII*.
 - Mayores colocaciones $\implies$ mayor incobrabilidad. La ganancia operativa contable desmejora por los *ETII*, y no muestra mejorías adicionales por los nuevos buenos pagadores (estaba previsto que cancelaran).
 - Sin embargo, la ganancia en el balance podría aumentar por el mayor volumen.

El resultado final en el balance dependerá del efecto que pondere más en cada caso.

- En Estadística I atendemos el ETI; aquí dependerá de una decisión de gestión.
- Con mucha liquidez, el banco quizás prefiere aumentar colocaciones aunque suba la incobrabilidad; y viceversa.
- Vínculo con la tasa de transferencia: si el ALM exige una tasa de transferencia pasiva muy baja, hay condiciones de liquidez para reducir el cut-off; y viceversa.

Modelo de Regresión Lineal I

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \dots$$

```
reg varDependiente varInd1 varInd2 varInd3 ...
```

- **Regla del test de significatividad:** si el p -valor ($P > |t| \leq 0,05$), el regresor es significativo.
 - En ocasiones puede interesarnos ver el signo de los coeficientes más allá de su significatividad.
- **Desventajas del modelo lineal:**
 - El modelo lineal implica por construcción un impacto constante de los regresores sobre la variable explicada (e.g. que cada año adicional de edad aporta lo mismo a la probabilidad de default).
 - Podría recurrirse a variables transformadas (e.g. edad al cuadrado).
 - La perturbación puede no ser normal (variable dicotómica) $\Rightarrow$ se pierden propiedades necesarias para los tests.

Modelo de Regresión Lineal II

- Cuando la variable dependiente de nuestra regresión es dicotómica, como es el caso de las dummies de default (1 si el cliente defaultea, 0 si no), usar el modelo lineal resulta en estimaciones sin sentido.
- El siguiente código puede usarse después de correr una regresión para predecir la probabilidad de default de cada observación utilizando los coeficientes estimados.
 - Estima la probabilidad para cada observación y la guarda en la columna default. Veamos su estadística descriptiva:

```
predict default, xb  
summ default
```

Variable	Obs	Mean	Std.Dev.	Min	Max
default	189	.1481481	.0594092	-.0965676	.2666186

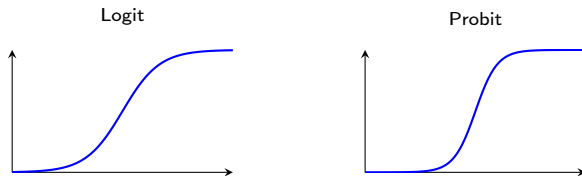
- El modelo lineal predice probabilidades fuera de $[0, 1]$.
 - Corrección: `replace default = 0 if default < 0`. Análogo si hay valores mayores a 1

Modelos de probabilidad no lineal I

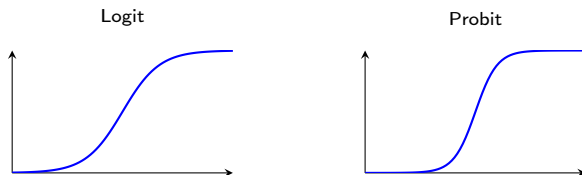
- Dados los problemas del modelo lineal, usamos modelos que sitúen a la variable explicada en el rango $[0; 1]$.
- Las funciones de distribución (continuas en ese rango) lo permiten:

$$Y = f\left(\beta_0 + \sum_{i=1}^n \beta_i X_i\right)$$

- **Logit** y **Probit** son modelos de regresión no lineal, donde la variable explicada es una función que arroja valores en $[0, 1]$ por construcción.
- La constante y los coeficientes (por las explicativas) entran en una función de distribución $\Rightarrow$ relación no lineal: incrementos leves en los extremos y mayores en la parte central.



Modelos de probabilidad no lineal II



- Los **signos** de los parámetros estimados indican la dirección en que se mueve la probabilidad cuando aumenta la variable explicativa.
- Las **magnitudes** de los parámetros **no** coinciden con la magnitud del cambio en la probabilidad.
 - Costos de la no-linealidad: aumentar una variable explicativa no tiene *siempre* el mismo efecto sobre la explicada, sino que depende del nivel en que se evalúe.

Logit (signos) vs. logistic (odds-ratio)

```
logit    varDependiente varInd1 varInd2 varInd3 ...    // SIGNOS
logistic varDependiente varInd1 varInd2 varInd3 ...    // ODDS-RATIO
```

- En logit
 - Los signos no son fácilmente interpretables (más sobre esto en las siguientes slides), aunque el signo nos indica el sentido de la relación.
- En logistic
 - Con un $OR < 1$, la variable reduce las chances de defaultear.
 - E.g. si el OR de una dummy de género (1 si mujer, 0 si hombre) es 0,37, se interpreta que las chances de defaultear siendo mujer multiplican las de los hombres por 0,37 (es decir, son menores).
 - Con un $OR > 1$, la variable aumenta las chances de defaultear.
 - E.g. si la cantidad de cuotas impagas tiene OR 2,3451: una cuota impaga adicional multiplica las chances de default por 2,3451, es decir, más que las duplica.

Efectos marginales I

- Habiendo estimado con logit, podemos estimar el **efecto marginal** de cada variable cuando todas tienen su valor promedio. Esto permite aproximar la interpretación que teníamos en el modelo lineal.
- E.g. si estimamos un modelo logit usando de predictores el ser mujer (fem, variable dummy que vale 1 si la cliente es mujer), la cantidad de cuotas impagas (impagos) y el saldo disponible (disp):

```
margins, dydx(*) atmeans
```

```
At: fem      = .7936508 (mean)
    impagos  = .1957672 (mean)
    disp     = 2944.286 (mean)
```

		Delta-method				
		dy/dx	std.err.	z	P> z	[95% conf.int.]

fem		-.0945053	.0437922	-2.16	0.031	-.18034 -.00867
impagos		.0810433	.0350212	2.31	0.021	.01240 .14968
disp		-.0001038	.0000287	-3.61	0.000	-.00016 -.00005

- Las clientas mujeres tienen una probabilidad de default $\approx 9,5$ p.p. menor que los hombres; una cuota impaga adicional aumenta la probabilidad en $\approx 8,1$ p.p..

Predicción con Logit I

- El mismo comando sirve para usar los coeficientes estimados y predecir la probabilidad de default para una persona puntual.
- Por ejemplo, la probabilidad de default de una mujer ($fem=1$) con saldo ($disp$) de \$2000, con el resto de variables en su valor promedio:

```
margins, at(fem=1 disp=2000) atmeans
```

```
At: fem =    1  
    impagos = .1957672 (mean)  
    disp = 2000
```

		Delta-method				
		Margin	std.err.	z	P> z	[95% conf.int.]

	+					
_cons		.2137173	.0561594	3.81	0.000	.1036468 .3237877

Predicción con Logit II

- Igual que en el modelo lineal, podemos usar el comando `predict` para estimar la probabilidad de default de todas las observaciones.

```
predict defaultlogit, pr  
sum defaultlogit
```

Variable	Obs	Mean	Std.Dev.	Min	Max
defaultlogit	189	.1481482	.1382439	.0087283	.7948883

- La media coincide con la del modelo lineal ($\approx 0,148$), pero ahora **todas** las predicciones caen dentro de $[0, 1]$ (Min = 0,009; Max = 0,795), sin valores negativos.

Criterios de información: estat ic

- AIC y BIC miden la *calidad relativa* de un modelo respecto a un conjunto de datos.
- **Regla de decisión**: a menor magnitud del criterio, mejor el modelo.
 - Si AIC y BIC generan resultados contradictorios, priorizamos el **BIC**.

estat ic

Akaike's and Bayesian information criteria

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
.	189	-79.28236	-66.78924	4	141.5785	154.5455

Matriz de confusión I

estat clas

		True		
Classified		D	~D	Total
+		4	3	7
-		24	158	182
Total		28	161	189
Classified + if predicted $\Pr(D) \geq .5$				
Sensitivity	$\Pr(+ D)$	14.29%		
Specificity	$\Pr(- \sim D)$	98.14%		
Correctly classified				85.71%

- Las columnas dicen si en la realidad el cliente defaultó (D) o no defaultó ($\tilde{D}$).
- En las filas, indica si el modelo lo catalogó como defaulteador (+) o no defaulteador (-).
- Se rechazaron 3 buenos (ETI). Se aceptaron 24 malos (ETII) y 158 buenos.
- **Sensibilidad** (potencia para identificar defaulteadores): $\approx 15\%$.
- **Especificidad** (identificar buenos pagadores): $\approx 98\%$.
- **Exactitud** (correctamente clasificados): $\approx 85\%$.

Matriz de confusión II

		Realidad	
		+	-
Predicción	+	a	b
	-	c	d

$$\text{Sensibilidad} = \frac{a}{a+c} \quad \text{Especificidad} = \frac{d}{b+d} \quad \text{Exactitud} = \frac{a+d}{a+b+c+d}$$

- Hay un *trade-off* entre sensibilidad y especificidad. El cut-off es el instrumento con el cual calibramos ambas.

Costos del ETI y ETII

Cada tipo de error impacta de manera diferente a nivel contable.

- **ETI:** **no** tiene un efecto observable a nivel contable. Son créditos que *podrían* haberse otorgado, recibiendo los pagos en tiempo y forma por los buenos clientes, pero no tuvieron lugar.
- **ETII:** se dieron créditos a malos clientes. Se registra la pérdida del capital (neto del colateral o garantías, si las hubiera).

Supongamos que los resultados previos corresponden a una cartera con promedio de \$100.000 por cliente, otorgados a una TEA del 58 %. Calculando la pérdida:

$$\text{Pérdida Total} = \text{Pérdida ETII} = \underbrace{24}_{\text{clientes}} \times \underbrace{\$100,000}_{\text{capital por cliente}} = \$2,400,000$$

Cambio de cut-off

- Si no especificamos, Stata asume que nuestro cut-off es 50 %, por lo que considera como defaulteador (+) a todas las observaciones con probabilidad estimada mayor a 0,5.
- Podemos evaluar el resultado con distintos niveles de cut-off: a menor cut-off, más exigente es la política de créditos.

```
estat clas, cutoff(0.25)
```

----- True -----			
Classified		D	~D Total
-----+-----+-----			
+		10	21 31
-		18	140 158
-----+-----+-----			
Total		28	161 189
Classified + if predicted Pr(D) >= .25			

Sensitivity	Pr(+ D)		35.71 %
Specificity	Pr(- ~D)		86.96 %

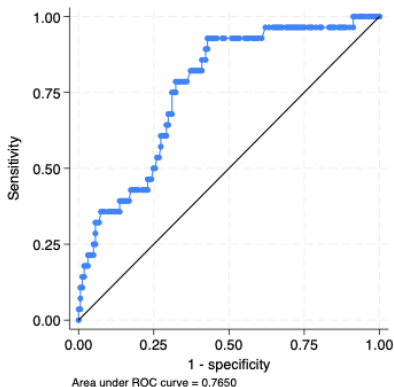
Correctly classified			79.37 %

Al ubicar el cut-off en una posición más exigente:

- La **sensibilidad** aumenta (14,29 % → 35,71 %): identifico más defaulteadores.
- La **especificidad** disminuye (98,14 % → 86,96 %): dejo de otorgar crédito a buenos pagadores que cancelarían.
- Evito otorgar créditos a defaulteadores, a costo de perder buenos clientes a la vez.

Curva ROC

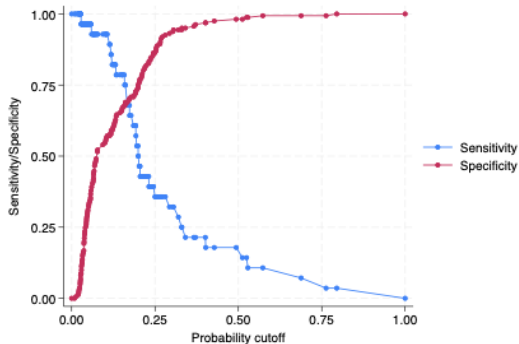
lroc



- El gráfico muestra distintas combinaciones de especificidad y sensibilidad posibles, donde cada punto es el resultado usando un cut-off específico.
- El óptimo está en la esquina superior izquierda. Cuanto más cerca de la línea de 45° , peor el modelo.
 - Una curva más cercana al óptimo tiene mayor área debajo.
 - **Criterio:** aceptar el modelo si $AUROC > 0,8$. Este modelo sería descartado.
- Subir el cut-off para bajar el ETI del 50 % al 25 % tiene como costo de oportunidad pasar de una sensibilidad $\approx 0,9$ a $\approx 0,45$ (aumenta el ETII).
- Es estratégico aprovechar las zonas verticales u horizontales según el cut-off.

Trade-off sensibilidad/especificidad

lsens



- Similar al gráfico previo: grafica el *trade-off* entre sensibilidad y especificidad (eje y) frente al cut-off elegido (eje x).
- Son las distintas combinaciones que obtendríamos si probáramos estat class mientras variamos el cut-off de 0 a 1.
- Permite entender gráficamente cuál es el intervalo de cut-offs considerables.

```
probit d fem impagos disp  
predict probit, pr  
sum probit
```

- No tenemos odds-ratio: solo interpretamos los **signos**.
- Sí podemos usar comandos análogos para entender los efectos marginales de las variables explicativas.
- Pueden compararse los criterios de información (AIC/BIC) y las AUROC del Logit y el Probit.