

UNIVERSIDAD NACIONAL DEL CHIMBORAZO

**MODELOS DE
ELECCIÓN DISCRETA
(LOGÍSTICA)
PROBIT / LOGIT**

Patricia Hernández, PhD

MODELIZACIÓN
ECONÓMICA

RUTA A SEGUIR EN EL CURSO

01

ANÁLISIS EXPLORATORIO /
CONFIRMATORIO DE DATOS

02 MODELOS
ECONOMÉTRICOS

MODELOS MICRO

- Regresión lineal
- **Variable dependiente cualitativa (probit / logit)**
- Variable dependiente categórica (ordinal / multinomial)
- Datos de panel

MODELOS DE ELECCIÓN DISCRETA

MODELOS DE ELECCIÓN DISCRETA VS. MODELOS LINEALES

Y objetivo

Modelo regresión lineal

Valor numérico
(cuantitativa)

Modelo elección discreta
(Probit/logit)

Factor / categoría /
dicotómica(0,1)
(cualitativa)

¿ES POSIBLE ESTIMAR Y CON VALOR 0,1 A TRAVÉS DE UN
MODELO LINEAL?

MODELOS DE ELECCIÓN DISCRETA VS. MODELOS LINEALES

¿ES POSIBLE ESTIMAR “Y” CON VALOR 0,1 A TRAVÉS DE UN
MODELO LINEAL?



NO

- La estimación no garantiza obtener valores entre 0 y 1 que determinan las probabilidades, pues la función que se ajusta a los datos es no lineal
- Los errores no se distribuyen normalmente
- Las varianzas no son homocedásticas

MODELOS DE ELECCIÓN DISCRETA VS. MODELOS LINEALES

¿CÓMO SE RESUELVE?



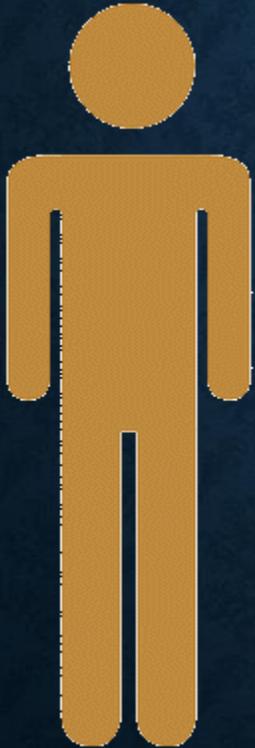
- Estimación de modelos probit y logit, no lineales en parámetros
- Estimación a través de máxima verosimilitud (iterar hasta encontrar la estimación óptima)
- Ya no se minimiza la suma de las diferencias entre valor observado y valor estimado de y (error), ahora el error es la no correcta clasificación en cada grupo

MODELOS DE ELECCIÓN DISCRETA



- Variable objetivo es **dicotómica** (0,1), divide a dos grupos
- Tomará el **valor de 1 de acuerdo con el propósito de la investigación** (por ejemplo, si se quiere estudiar la morosidad tendrá valor 1 si es moroso, 0 si es buen pagador)
- Sn estudios de **corte transversal**
- El objetivo no es predecir, sino **entender comportamientos**

MODELOS DE ELECCIÓN DISCRETA (PROBIT / LOGIT)

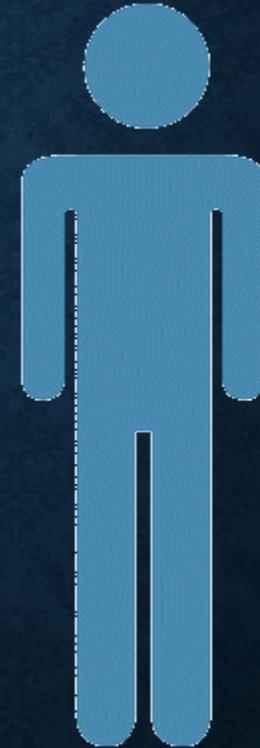


MOROSO
 $Y = 1$



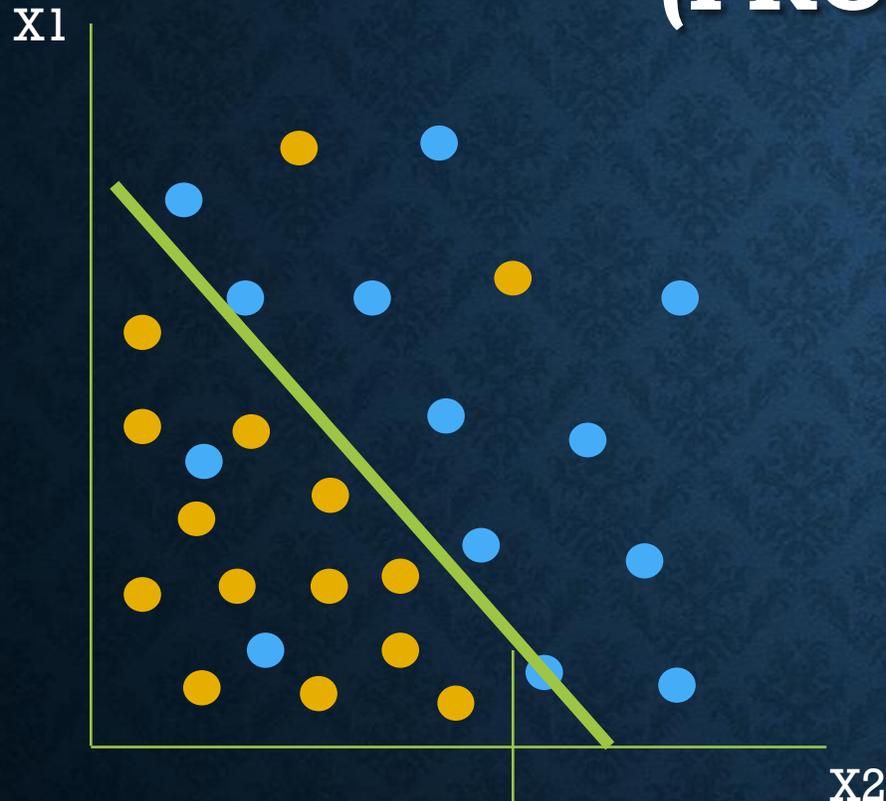
ESTUDIO SOBRE PRÉSTAMOS

DOS GRUPOS



PAGADOR
 $Y = 0$

MODELOS DE ELECCIÓN DISCRETA (PROBIT / LOGIT)



La probabilidad de ser moroso
puede ser el resultado de:
 X_1 = nivel de ingresos
 X_2 = nivel educativo

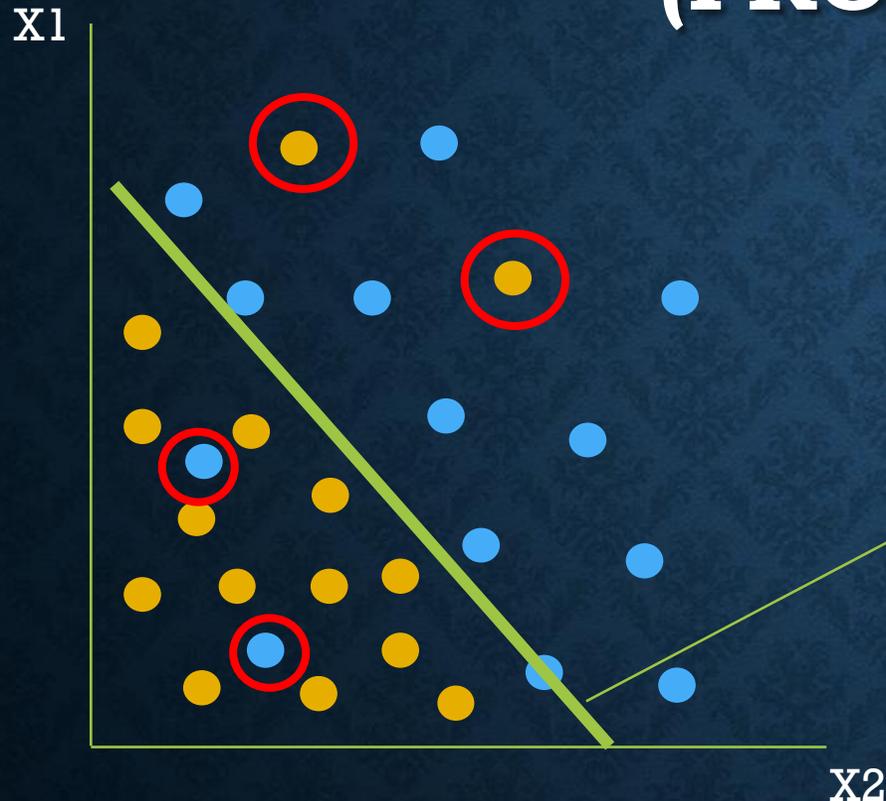
Y {
● Pagador
● Moroso

Regla de decisión – Regresión discreta

Patrón:

X_1 y x_2 bajos = Moroso
 X_1 y X_2 altos = pagador

MODELOS DE ELECCIÓN DISCRETA (PROBIT / LOGIT)



Regla de decisión – Regresión discreta

Tiene error de clasificación (errores de este tipo de estimación)

Se busca minimizar estos errores o
MAXIMIZAR la correcta clasificación
(**ACCURANCY**)

$$\text{MODELO A ESTIMAR: } y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon$$

DIFERENCIA ENTRE “Y” Y LA PROBABILIDAD ESTIMADA



ANA busca pareja
 $Y=1$ si saldría con él
 $Y=0$ si no saldría



Características

Valor de Y

\$ + Tímido

0



\$ + extrovertido

1



\$\$+Intelectual

1



\$\$+Soñador

???

DIFERENCIA ENTRE “Y” Y LA PROBABILIDAD ESTIMADA



Características

Valor de Y

\$ + Tímido

0

Probabilidad rechazo 1/3



\$ + extrovertido

1

Probabilidad aceptar 2/3



\$\$+Intelectual

1



\$\$+Soñador

???

$$E(y/x) = 0 * 1/3 + 1 * 2/3 = 2/3$$

DIFERENCIA ENTRE “Y” Y LA PROBABILIDAD ESTIMADA



Características

Valor de Y

\$ + Tímido

0

Probabilidad rechazo 1/3



\$ + extrovertido

1

Probabilidad aceptar 2/3



\$\$+Intelectual

1



\$\$+Soñador

???

Y es discreta 0,1
Y estimada es continua
Entre 0 y 1 (mide probabilidad)

¿CÓMO SE ESTIMAN LAS PROBABILIDADES?

A TRAVÉS DE LOS ODD RATIOS
(se estiman en el modelo logit/probit)

$$\text{odd ratio} = \frac{P(y = 1)}{P(y = 0)}$$

Si odd ratio = 2
Salir con alguien tiene el doble
de probabilidad de no salir

$$P(y = 1) + P(y = 0) = 1$$

$$2P(y = 0) + P(y = 0) = 1$$

$$3P(y = 0) = 1$$



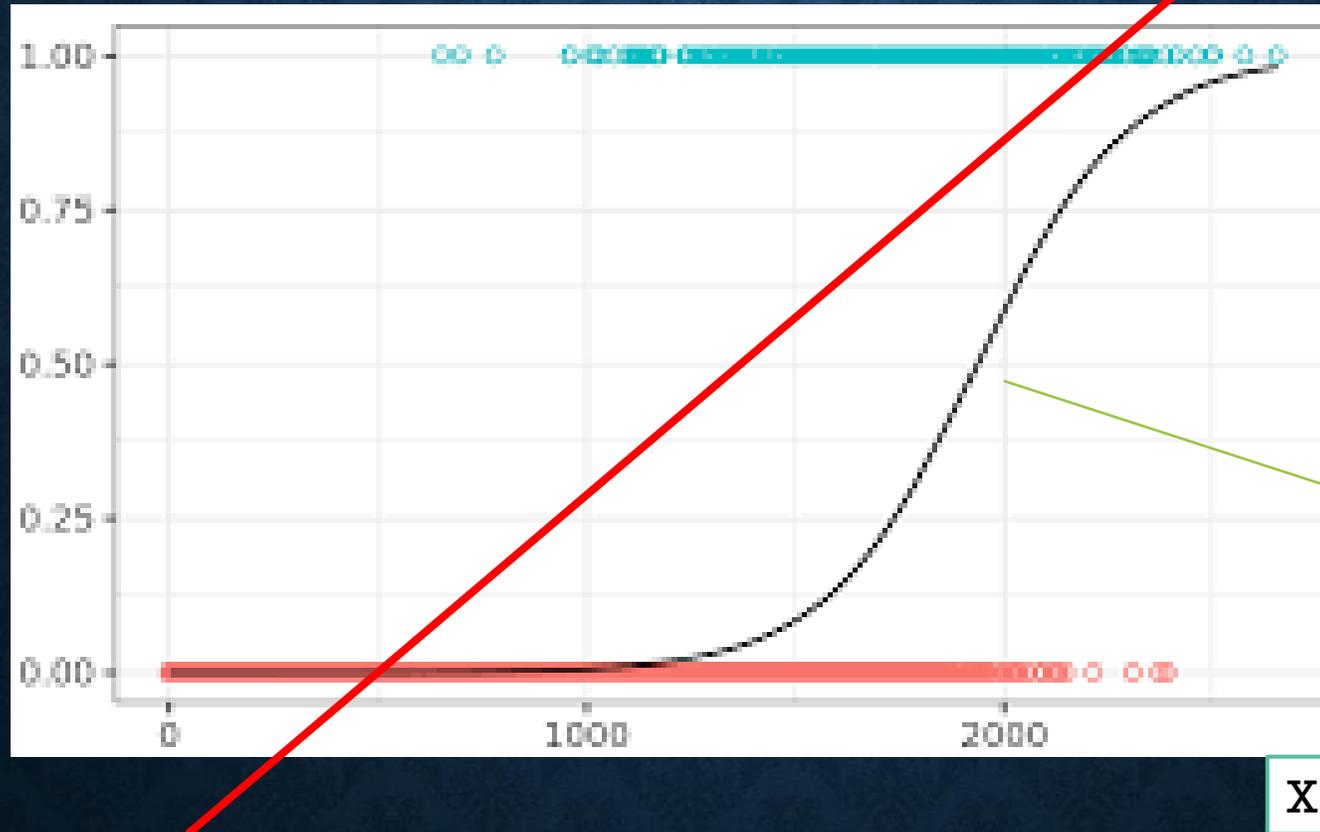
$$P(y = 0) = 1/3$$

$$P(y = 1) = 2/3$$

Probabilidad de no
salir
Probabilidad de
salir

¿CÓMO SE ESTIMA EL MODELO DE ELECCIÓN DISCRETA?

Y
Toma valor 0,1



Regresión lineal
La estimación de y está por encima de 1 y por debajo de 0
NO PUEDE EMPLEARSE

Función sinusoidal "S"
Garantiza que y estimada esté entre 0 y 1, porque son asíntotas

¿CÓMO SE ESTIMA EL MODELO DE ELECCIÓN DISCRETA?

Función sinusoidal "S"

Garantiza que y estimada esté entre 0 y 1, porque son asíntotas

Función **logística acumulada**
LOGIT

$$\frac{e^x}{1 + e^x}$$

Función **normal acumulada**
PROBIT

$$\int \frac{1}{\sigma\sqrt{\pi}} e^{-\frac{(x-\mu)^2}{\sigma^2}} dx$$

¿QUÉ ESCOGER ENTRE PROBIT Y LOGIT?

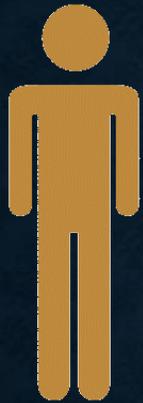
EL QUE TENGA MAYOR
ACCURANCY



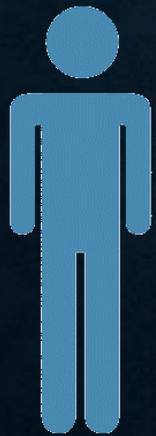
Mayor porcentaje de correcta clasificación del modelo
Mayor precisión

¿QUÉ ES EL ACCURANCY?

Matriz de confusión



MOROSO
 $Y = 1$



PAGADOR
 $Y = 0$

		\hat{y}		
		Pagador	Moroso	
y	Pagador	✓	✗	Correcta clasificación ✓
	Moroso	✗	✓	Incorrecta clasificación ✗

¿QUÉ ES EL ACCURANCY?

ACCURANCY

$$\frac{\Sigma \checkmark}{\text{total}}$$

CAMBIOS MARGINALES EN MODELO DE ELECCIÓN DISCRETA

No son interpretables los coeficientes porque no es una función lineal, deben estimarse cambios marginales (los signos se mantienen)

Cambios marginales

Derivadas

Se usa para variables explicativas discretas

$$\frac{\partial y}{\partial x}$$

$$\frac{\partial y}{\partial x}$$

Si x aumenta la probabilidad se reduce en xx puntos porcentuales

Puede calcularse en promedio o en cada punto (porque no es lineal la función)

Elasticidades

Se usa para variables explicativas continuas

$$e = \frac{\Delta\%y}{\Delta\%x}$$

Si x aumenta la probabilidad es un xx% menor

EVALUACIÓN DEL MODELO

MATRIZ DE CONFUSIÓN

SENSIBILIDAD

Clasificar correctamente a los morosos siendo morosos

ESPECIFICIDAD

Clasificar correctamente a los pagadores siendo pagadores

PROBABILIDAD PARA SEPARAR AMBOS GRUPOS

Probit model for ltp

Classified	True		Total
	D	~D	
+	336	149	485
-	92	176	268
Total	428	325	753

Classified + if predicted $\Pr(D) \geq .5$

True D defined as ltp != 0

Sensitivity	$\Pr(+ D)$	78.50%
Specificity	$\Pr(- \sim D)$	54.15%
Positive predictive value	$\Pr(D +)$	69.28%
Negative predictive value	$\Pr(\sim D -)$	65.67%

False + rate for true ~D	$\Pr(+ \sim D)$	45.85%
False - rate for true D	$\Pr(- D)$	21.50%
False + rate for classified +	$\Pr(\sim D +)$	30.72%
False - rate for classified -	$\Pr(D -)$	34.33%

Correctly classified 67.99%

¿SE PUEDE AJUSTAR LA PROBABILIDAD DE SEPARAR AMBOS GRUPOS?

LA PROBABILIDAD AJUSTADA ES LA QUE HACE CERO LA DIFERENCIA ENTRE LA SENSIBILIDAD Y LA ESPECIFICIDAD

SE PUEDE ESTIMAR LA MATRIZ DE CONFUSIÓN CON EL NUEVO PUNTO DE CORTE

No cambia el accuracy
Cambia especificidad y sensibilidad

```
. estat classification, cutoff(0.5607)
```

```
Probit model for lfp
```

Classified	True		Total
	D	~D	
+	291	104	395
-	137	221	358
Total	428	325	753

```
Classified + if predicted Pr(D) >= .5607
```

```
True D defined as lfp != 0
```

Sensitivity	Pr(+ D)	67.99%
Specificity	Pr(- ~D)	68.00%
Positive predictive value	Pr(D +)	73.67%
Negative predictive value	Pr(~D -)	61.73%

False + rate for true ~D	Pr(+ ~D)	32.00%
False - rate for true D	Pr(- D)	32.01%
False + rate for classified +	Pr(~D +)	26.33%
False - rate for classified -	Pr(D -)	38.27%

```
Correctly classified 67.99%
```

GRÁFICOS DE ANÁLISIS

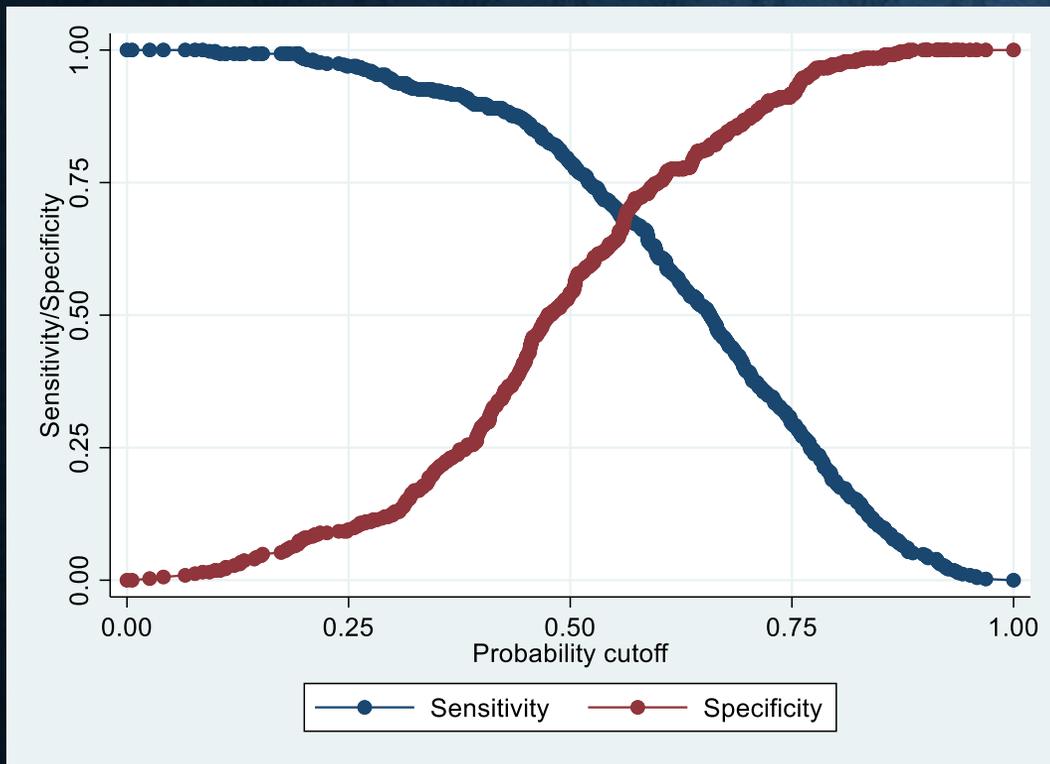


Gráfico sensibilidad – especificidad con la probabilidad estimada

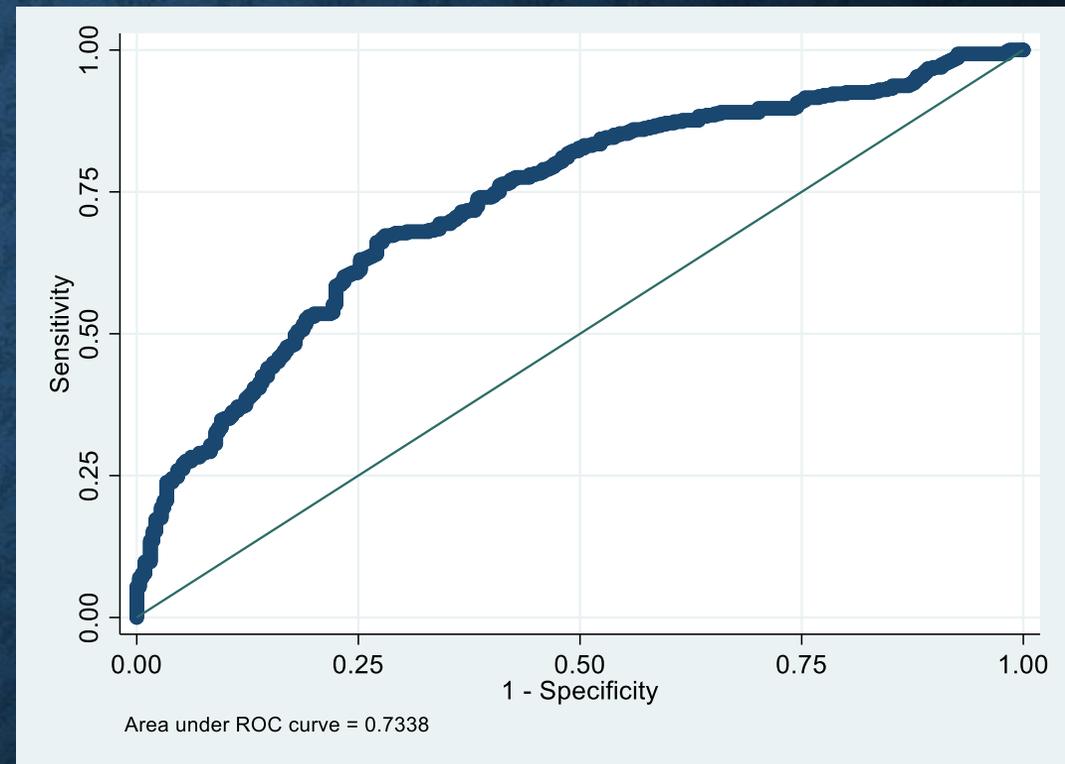


Gráfico de diagnóstico ROC (Característica operativa del receptor)

PROBIT

COMPARACIÓN PROBIT Y LOGIT

LOGIT

Variable	blogit	bprobit
k5	-1.4318036 -7.41	-.85901664 -7.66
age	-.05853213 -5.13	-.0352104 -5.16
wc	.87237348 4.23	.52167811 4.30
lwg	.61568413 4.10	.37316487 4.27
inc	-.03367515 -4.32	-.0202505 -4.45
_cons	2.9019256 5.35	1.7504257 5.44

legend: b/t

COMPARACIÓN PROBIT Y LOGIT

. estat classification, cutoff(0.5607)

Probit model for lfp

Classified	True		Total
	D	~D	
+	291	104	395
-	137	221	358
Total	428	325	753

Classified + if predicted $\Pr(D) \geq .5607$
 True D defined as $lfp \neq 0$

Sensitivity	$\Pr(+ D)$	67.99%
Specificity	$\Pr(- \sim D)$	68.00%
Positive predictive value	$\Pr(D +)$	73.67%
Negative predictive value	$\Pr(\sim D -)$	61.73%

False + rate for true ~D	$\Pr(+ \sim D)$	32.00%
False - rate for true D	$\Pr(- D)$	32.01%
False + rate for classified +	$\Pr(\sim D +)$	26.33%
False - rate for classified -	$\Pr(D -)$	38.27%

Correctly classified 67.99%

Logistic model for lfp

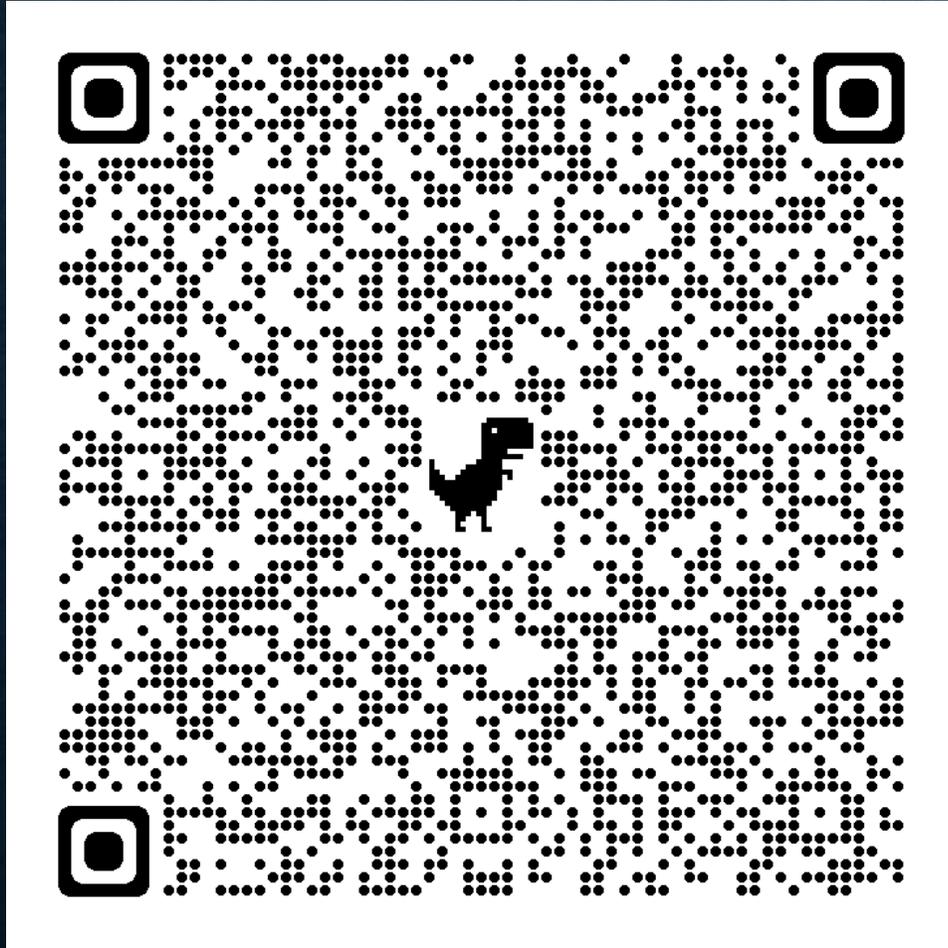
Classified	True		Total
	D	~D	
+	291	105	396
-	137	220	357
Total	428	325	753

Classified + if predicted $\Pr(D) \geq .5601545$
 True D defined as $lfp \neq 0$

Sensitivity	$\Pr(+ D)$	67.99%
Specificity	$\Pr(- \sim D)$	67.69%
Positive predictive value	$\Pr(D +)$	73.48%
Negative predictive value	$\Pr(\sim D -)$	61.62%

False + rate for true ~D	$\Pr(+ \sim D)$	32.31%
False - rate for true D	$\Pr(- D)$	32.01%
False + rate for classified +	$\Pr(\sim D +)$	26.52%
False - rate for classified -	$\Pr(D -)$	38.38%

Correctly classified 67.86%



**APLICACIONES
EN STATA
(EJERCICIOS
3_1 Y 3_2)**