2. SUPERVISED LEARNING

CURSO

Data Science Machine Learning & Deep Learning con Python

Regresión Lineal

Función de Regresión Poblacional

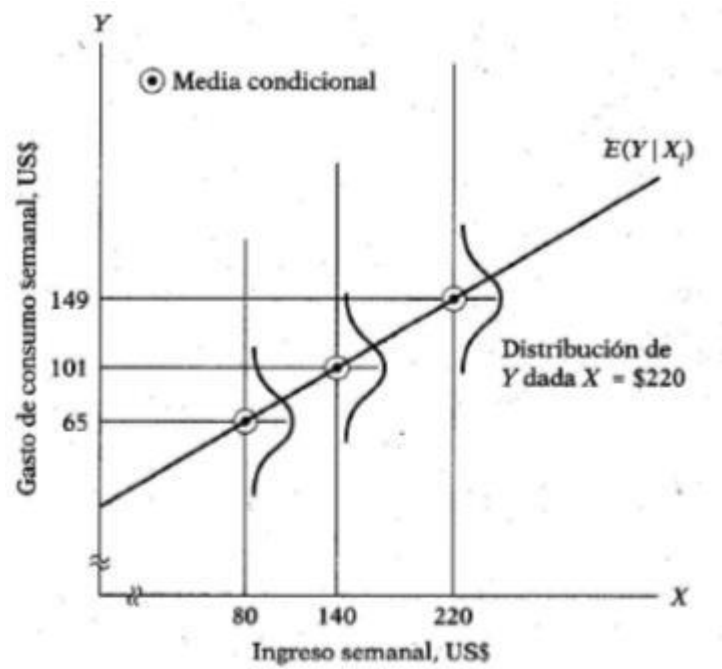
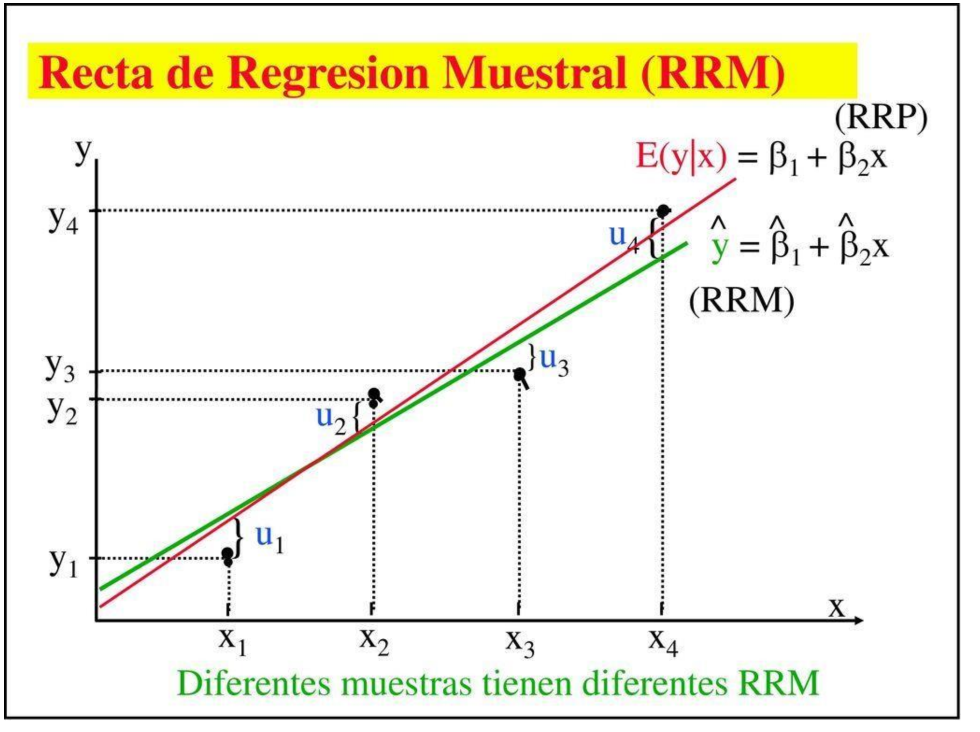


Figura 3: Ingreso semanal y Gasto semanal. Distribución simétrica

FRP:

La recta central objeto de estimación se denomina Función de Regresión Poblacional (FRP) y depende de los coeficientes poblacionales desconocidos α y β. Se trata de la parte sistemática o predecible del modelo y corresponde al comportamiento medio o esperado de la variable a explicar:



FRM:

El resultado final obtenido a partir de la información que ofrece una muestra dada se define como la Función de Regresión Muestral (FRM). Se obtiene una vez que los coeficientes de la regresión hayan sido estimados $(\hat{\alpha}, \hat{\beta})$ y también se conoce como modelo estimado.

Diplomados Perú

Estimación MCO

El objetivo de la estimación por MCO es minimizar la sumatoria de los errores estimados al cuadrado
Error de tipo II

$$min \sum \hat{u}^2$$

$$sujeto\ a\ \{\hat{\beta}\}$$

$$\hat{\beta} = (X'X)^{-1}(X'Y)$$

$$var(\hat{\beta}) = \hat{\sigma}^2(X'X)^{-1}$$

$$\hat{\sigma}^2 = \frac{\hat{u}'\hat{u}}{n-k}$$

Cómo puedo profundizar las matemáticas detrás del MCO?

Libro recomendado:
Econometría (Alfonso Novales Cinca).

Capitulo 3: EL modelo Lineal General

Modelo Lineal Simple

Modelo Lineal Simple

Tenemos que:

$$Y_i = \beta_0 + \beta_1 X_{1i} + u_i$$

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{u}_i$$

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i}$$

$$Y_i - \hat{Y}_i = \hat{u}_i$$

Optimización:

$$min \sum \hat{u}_i{}^2 \quad \Rightarrow \quad min \sum (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i})^2$$

$$sujeto\ a\ \{\hat{\beta}_0\ \hat{\beta}_1\} \qquad sujeto\ a\ \{\hat{\beta}_0\ \hat{\beta}_1\}$$

Diplomados Perú

Modelo Lineal Simple

Derivando:

$$\frac{\partial \sum (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i})^2}{\partial \hat{\beta}_0} = 0$$

$$\frac{\partial \sum (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i})^2}{\partial \hat{\beta}_1} = 0$$

Obtenemos el sistema de ecuaciones:

$$\sum Y_i = \sum \hat{\beta}_0 + \hat{\beta}_1 \sum X_{1i}$$

$$\sum X_{1i} Y_i = \hat{\beta}_0 \sum X_{1i} + \hat{\beta}_1 \sum X_{1i}^2$$

Diplomados Perú

Modelo Lineal General

Tenemos que:

$$Y = X\beta + u$$

$$Y = X\hat{\beta} + \hat{u}$$

$$\hat{Y} = X\hat{\beta}$$

$$Y - \hat{Y} = \hat{u}$$

Optimización:

$$\begin{array}{l} min\ \hat{u}'\hat{u} \\ sujeto\ a\ \ \{\hat{\beta}\} \end{array} \Rightarrow \begin{array}{l} min\ (Y - X\hat{\beta})'(Y - X\hat{\beta}) \\ sujeto\ a\ \ \{\hat{\beta}\} \end{array}$$

Modelo Lineal General

Derivando:

$$\frac{\partial \sum (Y - X\hat{\beta})' (Y - X\hat{\beta})}{\partial \hat{\beta}} = 0$$

Obtenemos:

$$X'Y = X'X\hat{\beta}$$

Finalmente:

$$\hat{\beta} = (X'X)^{-1}(X'Y)$$

Propiedades de un Estimador

- Insesgadez

$$E(\hat{\beta}) = \beta$$

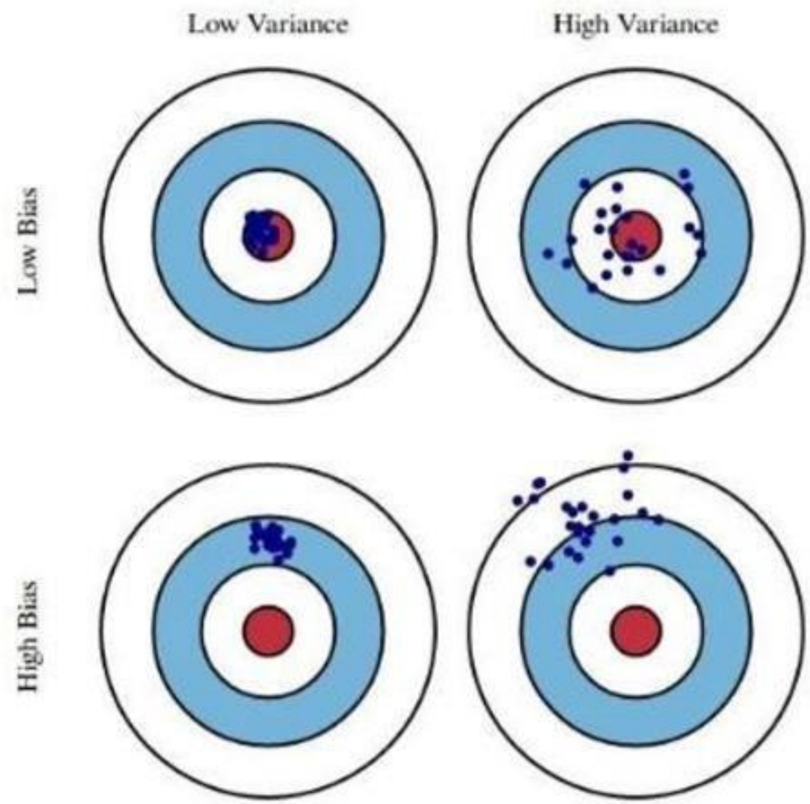
- Eficiencia

$$var(\hat{\beta})\ es\ mínima$$

- Consitencia

$$\underset{n \to \infty}{\text{Plim}}(\hat{\beta}) = \beta$$

Propiedades de un Estimador



Diplomados Perú

Indicadores de performance para modelos de regresión:

Suma de residuos al cuadrado (RSS) :

Sumatoria de la diferencia entre los valores predichos y los valores estimados al cuadrado. (Al cuadrado para mantener el signo positivo).

$$RSS = \sum_{i=1}^{n} \left(Y_i - \hat{Y}_i\right)^2$$

Error estándar residual (RSE):

$$RSE = \sqrt{RSS/(n-p)}$$

Donde p: son los grados de libertad de la predicción -> Número de predictores

Coeficiente de determinación (R^2) :

$$R^2 = 1 - \frac{RSS}{TSS}$$

Donde la varianza muestral:

$$TSS = \sum_{i=1}^{n} \left(Y_i - \bar{Y}\right)^2$$

R2 ajustado:

El coeficiente de determinación es una función creciente al número de variables.
Por tanto el beneficio de incluir variables adicionales es mejorar la bondad de ajuste del modelo, sin embargo el incrementar el número de variables tiene un costo, la pérdida de grados de libertad.
Por ello, el R cuadrado ajustado o corregido, es una medida mas confiable de la bondad de ajuste de un modelo.

El R2 solo es comparable con igual número de variables

$$R^2_{ajustado} = 1 - \frac{(n-1)}{(n-k)}(1 - R^2)$$

$$\uparrow k \Rightarrow \uparrow R^2 \Rightarrow \uparrow R^2_{ajustado} \quad \textit{beneficio}$$

$$\uparrow k \Rightarrow \downarrow (n-k) \Rightarrow \downarrow R^2_{ajustado} \quad \textit{costo}$$

Inferencia MCO

1. Prueba t:

$H_o: \hat{\beta}_i = k$
$H_a: \hat{\beta}_i \neq k$

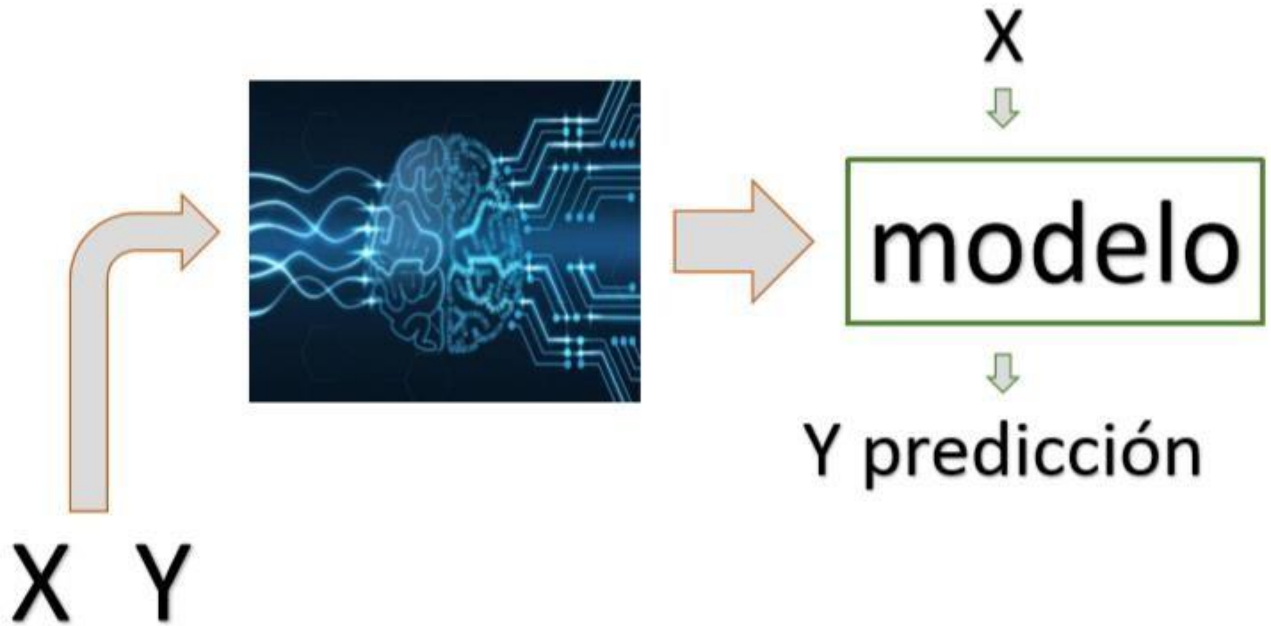
$$t_{calc} = \frac{\hat{\beta} - k}{std(\hat{\beta})}$$

$$t \sim t_{(\frac{\propto}{2}, n-k)}$$

Si: k=0 → Prueba de significancia individual.

Predicción

- Entenderemos por predicción a la aplicación de nuestro modelo.



Anexo

Supuestos MCO

1. Modelo estocástico: (u_i)
2. $\mathrm{E}(u_i) = 0$
3. Homocedasticidad
$$\mathrm{var}(u_i) = \sigma^2 \; \forall \; i = 1,2,3, \ldots, n$$
4. No Autocorrelación
$$\mathrm{cov}(u_i, u_j) = 0 \; \forall \; i \neq j$$
5. Exogeneidad
$$\mathrm{cov}(X_i, u_i) = 0 \; \forall \; i$$

Supuestos MCO

(2) + (3) + (4) : Los errores son ruido blanco.

$$E(uu') = \sigma^2 I$$

(2) + (3) + (4) + (5): Condiciones de Markov

"El estimador de MCO es el Mejor Estimador Lineal Insesgado (MELI)"

Diplomados Perú

Supuestos MCO

6. Modelo es lineal en parámetros.

7. Supuesto de estabilidad estructural.

8. Causalidad unidireccional.

9. No multicolinealidad: Las variables son linealmente independientes.

Supuestos MCO

10. Los valores de X son fijos o determinísticos.

11. Variabilidad en los valores de X.

12. Número de observaciones debe ser mayor al número de incógnitas.

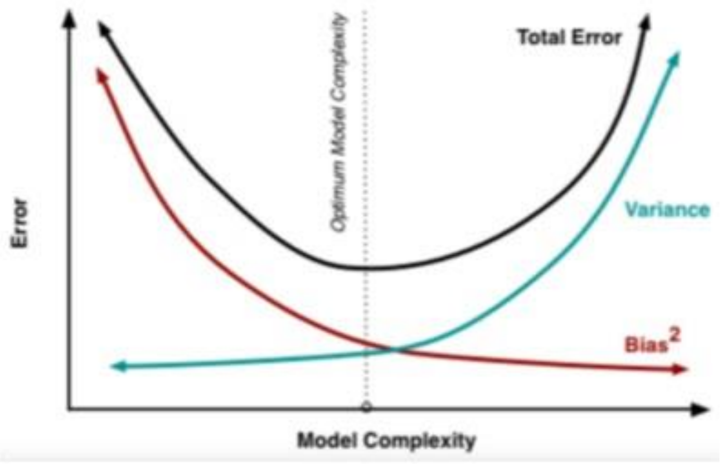
13. Modelos correctamente especificado.

Regularization

Concepto importante en Machine Learning que lidia con el overfitting.

Mientras más complejo sea el modelo, menor es el sesgo pero mayor la varianza y viceversa.

Las técnicas de Regularización tratan de elegir el modelo con la complejidad óptima, teóricamente, minimizando el error total.



Ridge Regression

Penaliza la función objetivo en función de la magnitud de los coeficientes.

Si el parámetro lambda tiende a cero, Ridge Regression tiende a MCO.

$$L_{ridge}(\hat{\beta}) = \sum_{i=1}^{n}(y_i - x_i'\hat{\beta})^2 + \lambda \sum_{j=1}^{m}\hat{\beta}_j^2 = ||y - X\hat{\beta}||^2 + \lambda||\hat{\beta}||^2.$$

Diplomados Perú

Lasso Regression

Penaliza la función objetivo en función de la magnitud de los coeficientes.

Si el parámetro lambda tiende a cero, Ridge Regression tiende a MCO.

$$L_{lasso}(\hat{\beta}) = \sum_{i-1}^{n}(y_i - x_i'\hat{\beta})^2 + \lambda \sum_{j-1}^{m} |\hat{\beta}_j|.$$

Elastic Net

Es una combinación entre Ridge y Lasso.

Si alfa es cero, Elastic Net converge a Ridge.

Si alfa es uno, Elastic Net converge a Lasso.

$$L_{enet}(\hat{\beta}) = \frac{\sum_{i-1}^{n}(y_i - x_i'\hat{\beta})^2}{2n} + \lambda(\frac{1-\alpha}{2}\sum_{j-1}^{m}\hat{\beta}_j^2 + \alpha\sum_{j-1}^{m}|\hat{\beta}_j|),$$

Regresión Logística

Modelos de elección discreta

$$Y_i = a + bX_i + e_i$$

Y_i es una variable discreta, por ejemplo:

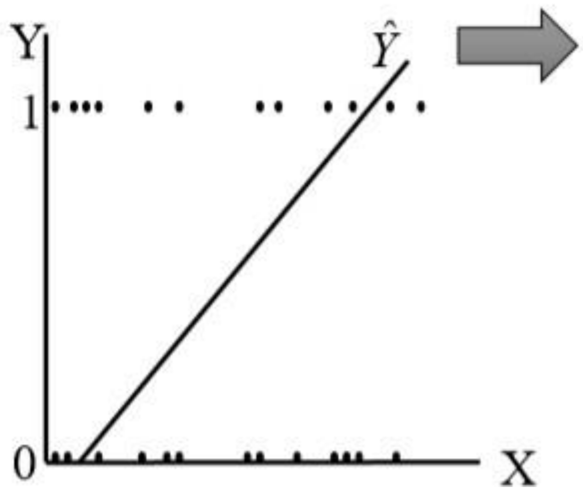
- El color del carro
- Incumplimiento de un pago
- Nivel de educación
- El sexo

Modelos de elección discreta

$$Y_i = a + bX_i + e_i$$

Y_i es una variable discreta dicotómica.

- Y_i =1 si es hombre.
- Y_i =0 si es mujer.



Estimación MCO

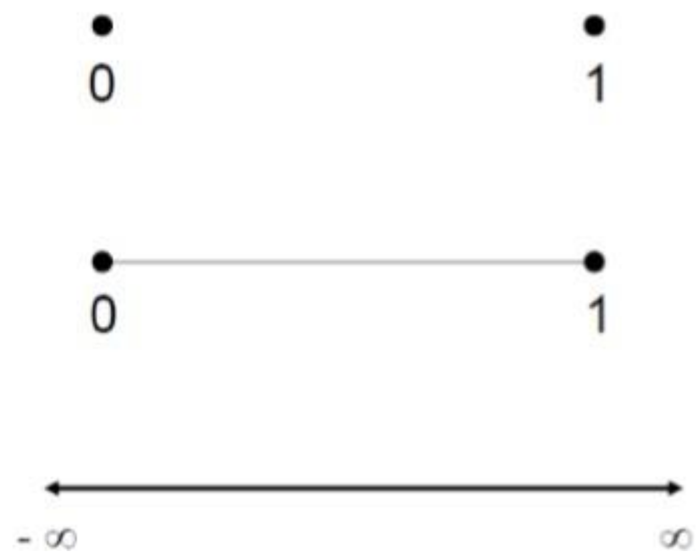
Dadas las características X_i , ¿Cuál es la probabilidad de que el individuo sea hombre o mujer?

Modelos de elección discreta

Variable original

Variable estimada
como probabilidad

Variable estimada
con un modelo lineal



Modelos de elección discreta

Limitaciones MCO:

- La predicción de MCO no necesariamente esta en el intervalo de [0 - 1].
- Existencia de heterocedasticidad.
- Violación del supuesto de normalidad.

Diplomados Perú

Modelos de elección discreta

Probabilidad Condicional

- Recordando: En una regresión lineal.

$E[Y/X] = X\beta \rightarrow \qquad \hat{Y} = X\hat{\beta}$

- Dado que Y es una variable dicotómica, tenemos:

$$E[Y/X] = 0 * P(Y = 0/X) + 1 * P(Y = 1/X)$$

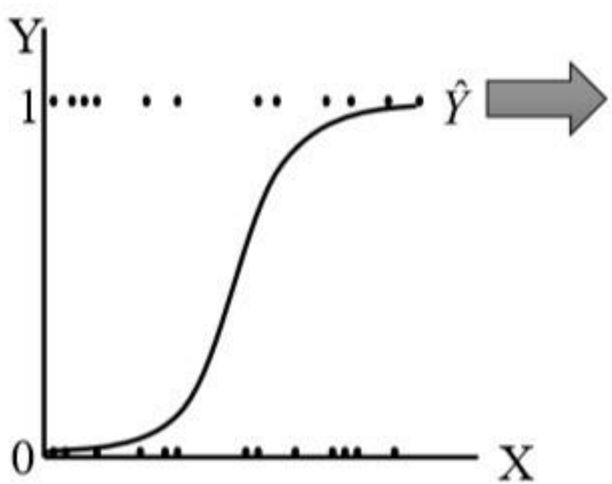
$$E[Y/X] = P(Y = 1/X)$$

$$\hat{Y} = P(Y = 1/X)$$

Modelos de elección discreta

La contribución de Mc Fadden:

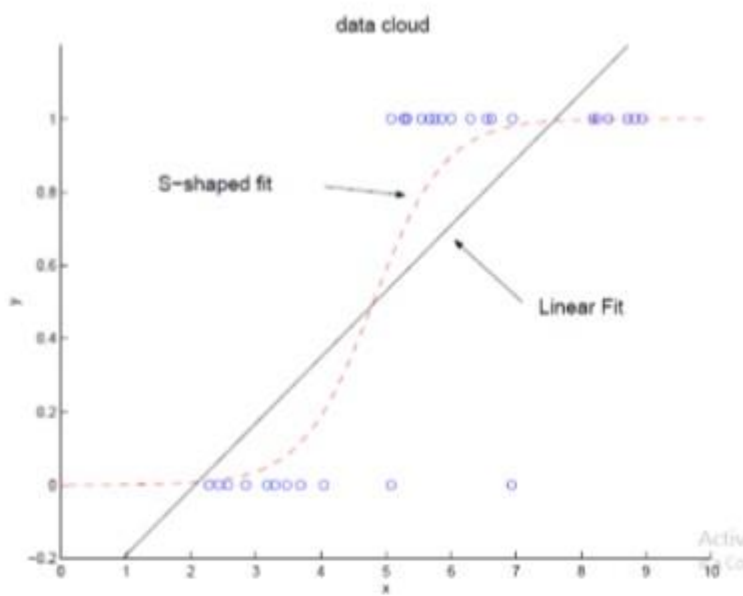
En este tipo de modelos se debe usar una curva cuyos valores estén estrictamente dentro del intervalo [0 – 1].



Estimación propuesta

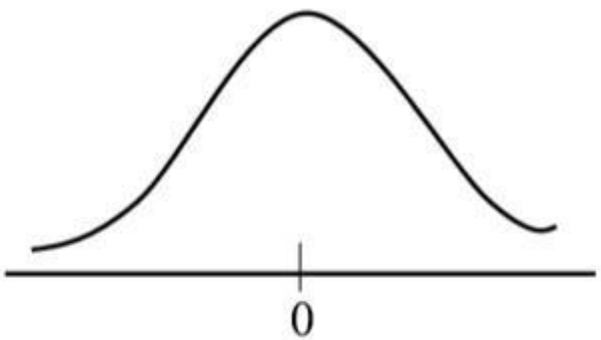
Curva con forma de S

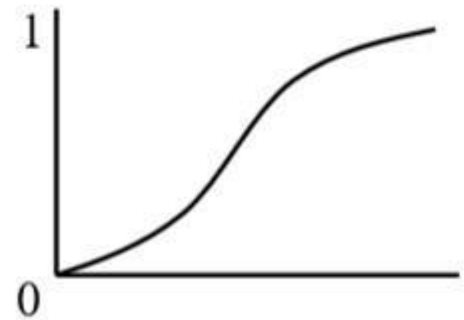
Modelos de elección discreta



Diplomados Perú

Modelos de elección discreta





$$CDF' = PDF$$

$$f(Z) = \frac{1}{\sqrt{2\pi}} e^{[-\frac{1}{2}Z^2]}$$

$$F(Z) = \int_{-\infty}^{Z} f(Z)dZ$$

$$g(Z) = \frac{e^Z}{(1+e^Z)^2}$$

$$G(Z) = \frac{1}{1+e^Z}$$

Modelos de elección discreta

Impactos Marginales

- MCO

$$\frac{\partial Y}{\partial X_i} = \beta_i$$

- PROBIT

$$\frac{\partial Y}{\partial X_i} = f(X\beta) * \beta_i$$

- LOGIT

$$\frac{\partial Y}{\partial X_i} = g(X\beta) * \beta_i$$

Regresión Logística

$$P(Y=1) = \frac{1}{1+e^{-X\beta}} \qquad X\beta = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ \dots + \beta_k X_k +$$

$$\frac{\partial Y}{\partial X_i} = \frac{e^{-X\beta}}{(1+e^{-X\beta})^2} * \beta_i$$

$$\frac{\partial Y}{\partial X_i} = P*(1-P)*\beta_i$$

- El signo del β corresponde al signo del impacto marginal, este debe ser consistente con la teoría económica.
- El impacto marginal no es constante, este varía para cada valor de X_i.
- El impacto marginal puede ser calculado *ceterisparibus*(Las X_i toman su valor promedio).

Ratio Odds (OR)

Se define:

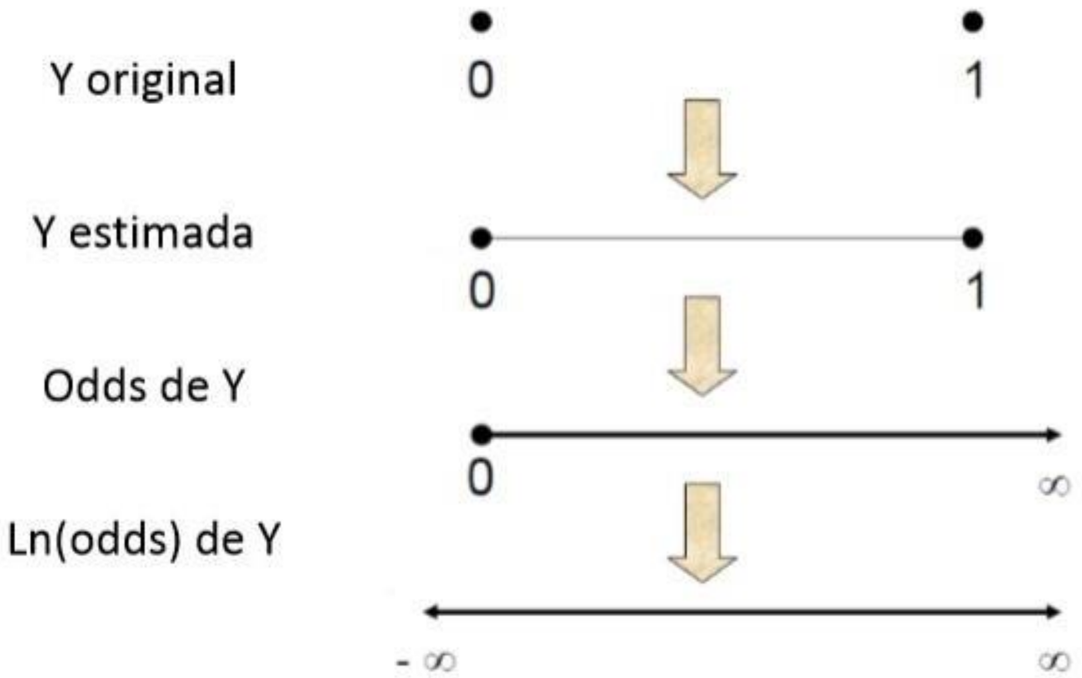
$$OR = \frac{P(evento\ a\ favor)}{P(evento\ en\ contra)}$$

Ejemplo:

EVENTO	PROBABILIDAD
Ganar una apuesta	0.8
Perder una apuesta	0.2

$$OR = \frac{0.8}{0.2} = 4$$

Ratio Odds (OR)



Weight of Evidence (WOE)

Se define:

$$WOE_i = \ln\left(\frac{P(i/evento\ a\ favor)}{P(i/evento\ en\ contra)}\right)$$

Ejemplo:

X=i	Y=0	Y=1	P(X=i/Y=0)	P(X=i/Y=1)	WOE
1	2	1	0.4	0.33	-0.18
2	1	1	0.2	0.33	0.51
3	2	1	0.4	0.33	-0.18
TOTAL	5	3	1	1	-

Diplomados Perú

Information Value (IV)

Se define:

$$\text{IV}=\sum_{i=1}^{k}[P(\text{i/evento } a\ favor) - P(\text{i/evento } en\ contra)] * WOE_i$$

Ejemplo:

X=i	Y=0	Y=1	P(X=i/Y=0)	P(X=i/Y=1)	WOE	Dif.	Prod.
1	2	1	0.4	0.33	-0.18	-0.067	0.0122
2	1	1	0.2	0.33	0.51	0.133	0.0681
3	2	1	0.4	0.33	-0.18	-0.067	0.0122
TOTAL	5	3	1	1	-	-	0.0924

= IV

Tasa de aciertos

- Si $\hat{p} > p*$ entonces $\hat{Y} = 1$
- Si $\hat{p} \leq p*$ entonces $\hat{Y} = 0$

		PREDICCIÓN	
		$\hat{Y} = 1$	$\hat{Y} = 0$
OBSERVADO	$\mathbf{Y = 1}$	n_{11}	n_{10}
	$\mathbf{Y = 0}$	n_{01}	n_{00}

$$n = n_{11} + n_{10} + n_{01} + n_{00}$$

$$(\%)\ Aciertos = \frac{n_{11} + n_{00}}{n}$$

www.diplomadosperu.com.pe