Jaime Rodrigo Roldán Carcelles 77228554M

Tema 2

• Distribuciones condicionadas

$X/_{y=y_j}$ X/y_j Y/x_i

$$f_i^j = \frac{n_{ij}}{n_{\cdot j}} = \frac{f_{ij}}{f_{\cdot j}} \qquad f_j^i = \frac{n_{ij}}{n_{i\cdot}} = \frac{f_{ij}}{f_{i\cdot}}$$

$$f_{ij} = f_i^j \cdot f_{\cdot j} \qquad f_{ij} = f_j^i \cdot f_{i\cdot}$$

• Relación entre variables

• El objetivo de analizar conjuntamente dos variables diferentes es establecer el tipo de relación existente entre ellas.

Hay 3 casos:

- Independientes: No hay relación alguna entre las variables.
- Dependencia funcional: El valor de una variable queda determinado conociendo el valor de la otra variable para esa misma observación a través de una función.
- Dependencia estadística: Una variable proporciona información sobre la otra pero la modalidad de una no queda determinada por la modalidad de la otra.

Jaime Rodrigo Roldán Corcelles 77228554H

- Independencia entre variables

X es independiente de Y si:

$$X/_{Y=y_j} \equiv X$$

Ejemplo:

$X \backslash Y$	C_1	C_2	C_3	C_4	
A	4	6	10	2	22
B	2	3	5	1	11

$Y/X=A$	f_j^i
C_1	2/11
C_2	3/11
C_3	5/11
C_4	1/11

$Y/X=B$	f_j^i
C_1	2/11
C_2	3/11
C_3	5/11
C_4	1/11

Y_j	$f_{\cdot j}$
C_1	2/11
C_2	3/11
C_5	5/11
C_4	1/11

Si X es independiente de Y

$$f_i^j = f_{i\cdot} \quad \forall i \ \forall j \quad f_i^j = \frac{n_{ij}}{n_{\cdot j}} ; \ f_j = \frac{n_i}{N}$$

- Dependencia funcional

$X \backslash Y$	C_1	C_2	C_3	C_4
A_1	4	6	0	0
A_2	0	0	5	7

22

$X/_{Y=C_3}$ sólo toma el valor de A_2

$X/_{Y=C_1}$ " " A_1

$X/_{Y=C_2}$ " " A_1

$X/_{Y=C_4}$ " " A_2

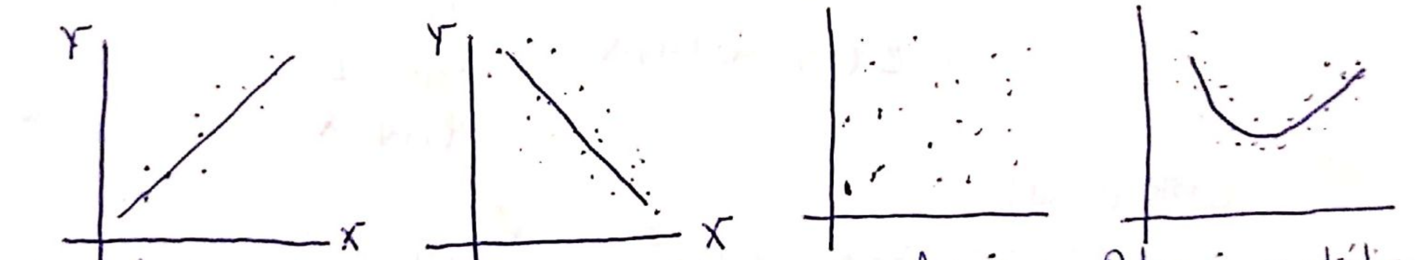
Jaime Rodrigo Roldán Corcelles 77228554M

- Dependencia estadística

Ejemplo:
- Estatura y peso
- Nacionalidad y renta
- Familia por nº de hijos y nº de móviles

+ Pasos a seguir:
  - 1) Nube de puntos
  - 2) Buscar la línea o curva de regresión que mejor se ajuste a la nube de puntos. → Regresión
  - 3) Medir el grado de dependencia entre las variables. → Correlación

Si todos los valores satisfacen la ecuación calculada, se dice que las variables ~~e~~están perfectamente correladas. La ecuación nos permite predecir valores desconocidos.



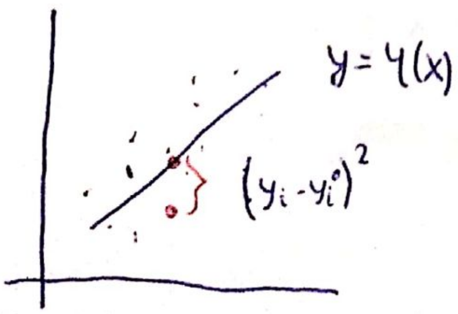
Relación lineal positiva — Relación lineal negativa — No relación — Relación cuadrática

• Regresión Lineal para tipo cuantitativas

• Metodo de minimos cuadrados

sean los datos $\{(x_i, y_i)\}$ para dos v.e. cuantitativas X, Y.

1. La Regresión de Y/X



El objetivo es minimizar

$$\sum_i (y_i - y_i^\circ)^2 = \sum_1 (y_i - \psi(x_i))^2 = F(x)$$

Jaime Rodrigo Roldán Carcelles 77228554M RB

$\text{Min } F \rightarrow \text{Min} \sum_i e_i^2 \rightarrow$ Minimizar el SSE

$y_i^{\circ} = y_i^{est} = \varphi(x_i)$ es el valor de y estimado por la regresión para x_i.

$e_i = y_i - y_i^{est}$ es el error cometido por el ajuste para el iésimo dato.

El tipo de ajuste de mínimo cuadrados viene determinado por el tipo de función $y = \varphi(x)$ elegido.

+ Caso Lineal Generalizado

$$\varphi(x) = a_0 \varphi_0(x) + a_1 \varphi_1(x) + a_2 \varphi_2(x) + \dots$$

Los más usados son:

- Ajuste lineal: $y = \varphi(x) = a + bx$ $\quad a_1 = a_1$, $a_0 = a_0$
  $\varphi(x) = a_0 + a_1 x$ $\quad \varphi_0(x) = 1$, $\varphi_1(x) = x$

- Ajuste parabolico:
  $y = \varphi(x) = a_0 + a_1 x + a_2 x^2$ $\quad \varphi_0(x) = 1$, $\varphi_1(x) = x$, $\varphi_2(x) = x^2$

- Otros ajustes:
  $y = \varphi(x) = a_0 \cdot \cos(x) + a_1 \cdot \text{sen}(x)$ $\quad \varphi_0(x) = \cos x$, $\varphi_1(x) = \text{sen } x$

o Ajuste Lineal. Propiedades

$$\min F = \min \sum_i (y_i - y_i^{est})^2 = \min \sum_i (y_i - \varphi(x_i))^2 =$$

$$= \min_{a,b} \sum (y_i - (a + bx_i))^2$$

Jaime Rodrigo Roldán Carcelles 77228554M RS

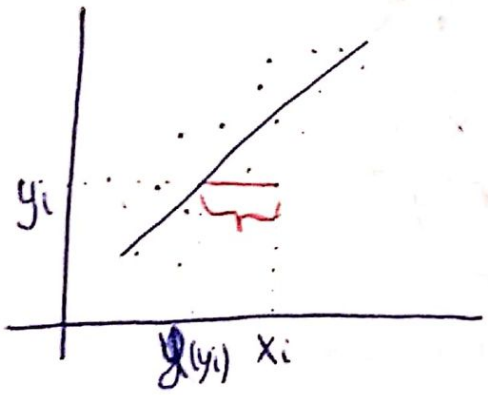
$$\frac{\partial F}{\partial a} = \sum_i 2(y_i - a - bx_i)\cdot(-1) = 0$$

$$\frac{\partial F}{\partial b} = \sum_i 2(y_i - a - bx_i)\cdot(-x_i) = 0$$

$$\left.\begin{aligned} -\sum_i y_i + a\sum_i + b\sum_i x_i = 0 \\ -\sum_i y_i x_i + a\sum x_i + b\sum_i x_i^2 = 0 \end{aligned}\right\} \begin{aligned} a\cdot N + b\sum_i x_i = \sum_i y_i \\ a\cdot\sum_i x_i + b\sum_i x_i^2 = \sum_i x_i\cdot y_i \end{aligned}$$

$$\begin{pmatrix} N & \Sigma x_i \\ \Sigma x_i & \Sigma x_i^2 \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \Sigma y_i \\ \Sigma x_i y_i \end{pmatrix}$$

\+ Regresión de X/Y



$$x = \varphi'(y)$$

$$\text{Min}\sum_i (x_i - \varphi'(y_i))^2 =$$
$$= \text{Min}\sum_i (x_i - x_i^{a})^2 =$$
$$= \text{Min}\sum_i (x_i - x_i^{est})^2$$

\+ Ajuste lineal

$$\varphi'(y) = a' + b'y$$

$$\begin{pmatrix} N & \sum_i y_i \\ \Sigma y_i & \sum_i y_i^2 \end{pmatrix}\begin{pmatrix} a' \\ b' \end{pmatrix} = \begin{pmatrix} \Sigma x_i \\ \Sigma x_i y_i \end{pmatrix}$$

Sist de Ec. normales

Jaime Rodrigo Roldán Carcelles 77228554M

- Ajuste lineal: Propiedades

+ Y/X S. Ecuaciones Normales

$$\left.\begin{array}{l} Na + b\sum x_i = \sum y_i \\ a\sum x_i + b\sum x_i^2 = \sum x_i y_i \end{array}\right\} \quad \left.\begin{array}{l} a + b\frac{\sum x_i}{N} = \frac{\sum y_i}{N} \\ a\frac{\sum x_i}{N} + b\frac{\sum x_i^2}{N} = \frac{\sum x_i y_i}{N} \end{array}\right\}$$

$$\left.\begin{array}{l} a + b\bar{x} = \bar{y} \\ a\bar{x} + b m_{20} = m_{11} \end{array}\right\}$$

① $(\bar{x}, \bar{y})$ pasa por la recta de regresión de Y/X

$a = \bar{y} - b\bar{x} \longrightarrow (\bar{y} - b\bar{x})\bar{x} - b m_{20} = m_{11}$

$\bar{y}\bar{x} - b\bar{x}^2 + b m_{20} = m_{11}$

$b(m_{20} - \bar{x}^2) = m_{11} - \bar{y}\bar{x}$

$$b = \frac{m_{11} - \bar{y}\bar{x}}{m_{20} - \bar{x}^2} = \frac{Cov(x,y)}{\sigma^2_x} = b$$

$$\left.\begin{array}{l} Na' + b'\sum y_i = \sum x_i \\ a'\sum y_i + b'\sum y_i^2 = \sum x_i y_i \end{array}\right\} \quad \begin{array}{l} a' + b'\bar{y} = \bar{x} \\ b' = \frac{Cov(X,Y)}{\sigma^2 y} \end{array}$$

② $(\bar{x}, \bar{y})$ pasa por la recta de regresión de X/Y

$(\bar{x}, \bar{y})$ es el punto de corte de las dos rectas de regresión

③ Signo b = signo b' = signo $cov(X,Y)$

$\sum_i e_i = 0 \qquad e_i = y_i - y_i^* \qquad y_i = a + b x_i$

$$\sum_i e_i = \sum_i (y_i - y_i^*) = \sum_i (y_i - (a + b x_i)) = \sum y_i - a \cdot N - b\sum x_i = 0$$

1ª Ec. del sistema de ec. normales

$\sum e_i x_i = 0$

$$\sum (y_i - y_i^*) x_i = \sum (y_i - a - b x_i) x_i = \sum_i y_i x_i - a\sum_i x_i - b\sum x_i^2 = 0$$

2ª Ec. del SEN

Jaime Rodrigo Roldán Carcelles 77228554 M

Descomposición de la varianza

$$\sigma^2_y = \frac{1}{N}\sum_i (y_i-\bar{y})^2 = \frac{1}{N}\sum_i (y_i - y_i^* + y_i^* - \bar{y})^2 =$$

$$= \frac{1}{N}\sum_i \left((y_i-y_i^*) + (y_i^* - \bar{y}^*)\right)^2 = \frac{1}{N}\left[\sum_i \left((y_i-y_i^*)^2 + (y_i^*-\bar{y}^*)^2 + 2(y_i-y_i^*)(y_i^*-\bar{y})\right)\right] = \frac{1}{N}\sum_i e_i^2 + \sigma_{y^*}^2 + \frac{1}{N}2\sum_i e_i \cdot (y_i^*-\bar{y}) =$$

$$= MSE + \sigma^2_{y^*} + \frac{1}{N}2\sum_i e_i\left(a+bx_i - (a+b\bar{x})\right) = MSE + \sigma^2_{y^*}$$

Varianza residual o varianza de los errores

Sabíamos que: $\sum_i e_i = 0 \Rightarrow \bar{e} = 0$

" " $\sigma^2_y = MSE + \sigma^2_{y^*}$

$$\sigma^2_{resi} = \frac{1}{N}\sum (e_i-\bar{e})^2 = \frac{1}{N}\sum_i e_i^2 - \bar{e}^2 = MSE$$

$$\sigma^2_y = \sigma^2_{res} + \sigma^2_{y^*}$$

¿Quién es $\sigma^2_{y^*}$?

$$\sigma^2_{y^*} = \frac{1}{N}\sum_i (y_i^*-\bar{y}^*)^2 = \frac{1}{N}\sum_i (y_i^*-\bar{y})^2 = \frac{1}{N}\sum_i \left(a+bx_i-(a+b\bar{x})\right)^2 =$$

$$= \frac{1}{N}\sum_i \left(b(x_i-\bar{x})\right)^2 = b^2\frac{1}{N}\sum_i (x_i-\bar{x})^2 = b^2\sigma^2_x$$

Por tanto, $\sigma^2_{res} = \sigma^2_y - b^2\sigma^2_x$

Jaime Rodrigo Roldán Carcelles 77228554M

Coeficiente de correlación lineal de Pearson

Mide el grado de relación lineal entre las variables. Se define como la media geométrica de los coeficientes de regresión de b y b'

$$\boxed{r = 1 = \sqrt{b \cdot b'}}$$

$$b = \frac{cov(x,y)}{\sigma^2_x}$$

$$b' = \frac{cov(x,y)}{\sigma^2_y}$$

signo de b
=
signo de b'
=
$\mu_{11} = cov(x,y)$

$$r = \sqrt{\frac{\mu_{11}^2}{\sigma^2_x \sigma^2_y}} = \frac{\mu_{11}}{\sigma_x \sigma_y} \Rightarrow r = \frac{\mu_{12}}{\sigma_x \cdot \sigma_y}$$

$$r^2 \geqslant 0$$

- b y b' en función de r:

$$b = \frac{r\,\sigma_y}{\sigma_x} \qquad b' = \frac{r\,\sigma_x}{\sigma_y}$$

- Propiedades de r^2
  - Sabiendo que: $\sigma^2_{res} = \sigma^2_y - b^2 \sigma^2_x$

    como $b = \frac{r\,\sigma_y}{\sigma_x} \Rightarrow \sigma^2_{res} = \sigma^2_y - \frac{r^2 \sigma^2_y}{\cancel{\sigma^2_x}} \cdot \cancel{\sigma^2_x}$

    $$\frac{\sigma^2_{res}}{\sigma^2_y} = 1 - r^2 \Rightarrow r^2 = 1 - \frac{\sigma^2_{res}}{\sigma^2_y}$$

  - Si $r^2 = 1$ ($r = 1$ ó $r = -1$

    Correlación lineal es perfecta $\begin{cases} r = 1 \rightarrow \text{directa} \\ r = -1 \rightarrow \text{inversa} \end{cases}$

  - Si $r = 0$

    Variables incorreladas

Cuando más próximo a 1 se encuentre el r^2, mejor será el ajuste lineal.

Jaime Rodrigo Roldán Carcelles 77228554M

¿Cuándo es mejor una recta que otra?

1. Depende de lo que queramos predecir
2. Diremos que y/x es mejor que x/y si

$$\sigma^2_{res}\, y/x \leq \sigma^2_{res}\, x/y$$
$$V^2_y (1-r^2) \leq \sigma^2_x (1-r^2)$$
$$\sigma^2_y \leq \sigma^2_x$$

Relación entre b y b'

$y = a + bx$ $\quad$ $x = a' + b'y$ $\quad$ $y = \frac{x - a'}{b'}$

$m\, y/x = b$ $\quad$ $m\, x/y = \frac{1}{b'}$

$$r^2 = |b \cdot b'| \leq 1 \iff r^2 = |b| \cdot |b'| \leq 1$$
$$|b| \leq \frac{1}{|b'|}$$

$$|m_{y/x}| \leq |m_{x/y}|$$

Escaneado con CamScanner

Jaime Rodrigo Roldán Carcelles 77228554M

Modelo lineal generalizado

Ajustamos la nube de puntos a una funciós del tipo:

$\varphi(x) = a_0 \varphi_0(x) + a_1 \varphi_1(x) + a_2 \varphi_2(x) + \dots$

El objetivo:

- minimizar $\sum_i (y_i - y_i^*)^2$ (x indep)
- minimizar $\sum_i (y_i - \varphi(x_i))^2$ ya que $y_i^* = y_i^{est} = \varphi(x_i)$
- minimizar $\sum e_i^2$

- Considero los vectores:

$y = \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix}$ $x = \begin{pmatrix} x_1 \\ \vdots \\ x_N \end{pmatrix}$ $y^* = \varphi(x) = \begin{pmatrix} y_1^* \\ y_2^* \\ \vdots \\ y_N^* \end{pmatrix} = \begin{pmatrix} \varphi(x_1) \\ \varphi(x_2) \\ \vdots \\ \varphi(x_N) \end{pmatrix}$

$M = \begin{pmatrix} \varphi_0(x_1) & \varphi_1(x_1) & \dots \\ \varphi_0(x_2) & \varphi_1(x_2) & \dots \\ \vdots & \vdots & \\ \varphi_0(x_N) & \varphi_1(x_N) & \dots \end{pmatrix}$ $A = \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \end{pmatrix}$ $e = y - y^* = \begin{pmatrix} y_1 - y_1^* \\ y_2 - y_2^* \\ \vdots \\ y_N - y_N^* \end{pmatrix}$

① $\varphi(x_1) = a_0 \varphi_0(x_1) + a_1 \varphi_1(x_1) + a_2 \varphi_2(x_1) + \dots$

$\varphi(x_2) = a_0 \varphi_0(x_2) + a_1 \varphi_1(x_2) + a_2 \varphi_2(x_2) + \dots$

$\vdots$

$\varphi(x_N) = a_0 \varphi_0(x_N) + a_1 \varphi_1(x_N) + a_2 \varphi_2(x_N) + \dots$

$\varphi(x) = M \cdot A$

② $(y - y^*)^t (y - y^*) = (y_1 - y_1^* \dots y_N - y_N^*) \begin{pmatrix} y_1 - y_1^* \\ \vdots \\ y_N - y_N^* \end{pmatrix} = \sum_i (y_i - y_i^*)^2$

Jaime Rodrigo Roldán Carcelles 77228554 M

$$\min \sum_i e_i^2$$

$$\min_{a_1, a_2, a_3 \dots} \sum_i (y_i - y_i^*)^2$$

$$\min_A (y - y^*)^t (y - y^*)$$

$$\min_A (y - \varphi(x))^t (y - \varphi(x))$$

$$\min_A (y - MA)^t (y - MA)$$

$$\min_A (y^t - A^t M^t)(y - MA)$$

$$\min_A y^t y - y^t MA - A^t M^t y + A^t M^t MA$$

$$f(A) = y^t y - 2 M^t y A + A^t M^t MA$$

$$\frac{\partial f}{\partial A} = 0 \Rightarrow -2M^t y + 2 M^t MA = 0$$

$$\boxed{M^t MA = M^t y} \rightarrow$$ Sistema de Ec. normales

$$\varphi(x) = a + bx \qquad \varphi_0(x) = 1 \qquad \varphi_1(x) = x$$

$$M = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_N \end{pmatrix} \quad A = \begin{pmatrix} a \\ b \end{pmatrix} \quad Y = \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix}$$

$$M^t MA = \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_N \end{pmatrix} \cdot \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_N \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} N & \Sigma x_i \\ \Sigma x_i & \Sigma x_i^2 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix}$$

$$M^t y = \begin{pmatrix} 1 & \dots & 1 \\ x_1 & \dots & x_N \end{pmatrix} \cdot \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix} = \begin{pmatrix} \Sigma y_i \\ \Sigma x_i y_i \end{pmatrix}$$

Jaime Rodrigo Roldán Cáceles 77228554M RS

- **Ajuste mediante un plano**

$z = \varphi(x,y) = a + bx + cy$

$z_i^* = a + bx_i + cy_i$

$$M = \begin{pmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ \vdots & \vdots & \vdots \\ 1 & x_N & y_N \end{pmatrix} \qquad A = \begin{pmatrix} a \\ b \\ c \end{pmatrix}$$

- Ecuaciones normales

$M^t M A = M^t Z \longrightarrow$ Dependiente

$$\begin{pmatrix} N & \Sigma x_i & \Sigma y_i \\ \Sigma x_i & \Sigma x_i^2 & \Sigma x_i y_i \\ \Sigma y_i & \Sigma x_i y_i & \Sigma y_i^2 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} \Sigma z_i \\ \Sigma x_i z_i \\ \Sigma y_i z_i \end{pmatrix}$$

- **Otros ajustes no lineales**

- $y = ae^{bx} \longrightarrow$ Ajuste exponencial
  - Tomamos logaritmos para linealizar el modelo

$Ln\, y = Ln\, a e^{bx}$

$Ln\, y = Ln\, a + Ln\, e^{bx}$

$Ln\, y = Ln\, a + bx$

  - Llamamos $Ln\, y = \hat{y}$ a una nueva variable
    $Ln\, a = \hat{a}$

$\hat{y} = \hat{a} + bx$