

• NOCIONES PRELIMINARES

- Escala de Medida

• Variables Categoricals

- Escala nominal (2 opc) $\rightarrow$ Ej: Sexo
- Escala ordinal (varias) $\rightarrow$ Ej: Aceptación

• Variables Cuantitativas

- Escala por intervalos $\rightarrow$ Ej: Temp.
- Escala por ratios $\rightarrow$ origen = 0

• Variables discretas o continuas

Nº hijos, familia.

- Tabla Frecuencias

Frec Abs	Frec Rel
n_1	$f_i = \frac{n_1}{n}$



ACUMULADA

$$N_i = N_{(i-1)} + n_i$$

$$F_i = F_{(i-1)} + f_i$$

Medidas Centralización (No agrupados)

Mediana

- Impar $\rightarrow M_e = x_{(\frac{n+1}{2})}$

- Par $\rightarrow M_e = \frac{x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}}{2}$

Media

$$\bar{X} = \frac{1}{n} \cdot \sum_{i=1}^n x_i$$

Medidas Dispersión

Varianza

$$S^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

Rango

$$R = x_{(n)} - x_{(1)}$$

Desviación típica

$$S = +\sqrt{S^2}$$

Coefficiente Variación

$$C.V. = \frac{S}{\bar{x}}$$

Quasi-varianza

$$\bar{S}^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{S^2 \cdot n}{n-1}$$

Quasi-desviación típica

$$S = +\sqrt{S^2}$$

Error estándar

$$e.s. = \frac{S}{\sqrt{n-1}} = \frac{S}{\sqrt{n}}$$

- Medidas Forma

- Coeficiente Asimetría

$$\phi_1 = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3}{s^3}$$

< 0 Asimétrica izq.
 > 0 Asimétrica der.
 $= 0$ Simétrica.

- Coeficiente Curtosis

$$\phi_2 = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4}{s^4} - 3$$

< 0 Platicúrtica
 > 0 Leptocúrtica
 $= 0$ Mesocúrtica

- BIVARIANTES (X, Y)

X \ Y	y ₁	y ₂	y ₃	...	y _j	...	y _k	n _{i.}
x ₁	n ₁₁	n ₁₂	n ₁₃		n _{1j}		n _{1k}	n _{1.}
x ₂	n ₂₁	n ₂₂	n ₂₃		n _{2j}		n _{2k}	n _{2.}
x ₃	n ₃₁	n ₃₂	n ₃₃		n _{3j}		n _{3k}	n _{3.}
...								
x _i	n _{i1}	n _{i2}	n _{i3}		n _{ij}			n _{i.}
...								
x _r	n _{r1}	n _{r2}	n _{r3}				n _{rk}	n _{r.}
n _{.j}	n _{.1}	n _{.2}	n _{.3}		n _{.j}		n _{.k}	n

$$n_i = \sum_{j=1}^k n_{ij} \quad n_j = \sum_{i=1}^r n_{ij}$$

$$f_{ij} = \frac{n_{ij}}{n}$$

S. \ E.C.	Soltero	Casado	Viudo	n _{i.}
Hombre	n ₁₁ = 3	n ₁₂ = 4	n ₁₃ = 1	n _{1.} = 8
Mujer	n ₂₁ = 2	n ₂₂ = 3	n ₂₃ = 2	n _{2.} = 7
n _{.j}	n _{.1} = 5	n _{.2} = 7	n _{.3} = 3	n _{..} = 15

- Distribuciones Condicionadas

$$f_{Y/X} = \frac{n_{ij}}{n_{i.}}$$

$$f_{X/Y} = \frac{n_{ij}}{n_{.j}}$$

S. \ E.C.	Soltero	Casado	Viudo	
Hombre	0,376	0,5	0,124	1
Mujer	0,285	0,43	0,285	1

Frecuencias condicionadas del sexo para cada categoría del estado civil.

S. \ E.C.	Soltero	Casado	Viudo	
Hombre	0,6	0,57	0,332	
Mujer	0,4	0,43	0,668	
	1	1	1	

Frecuencias condicionadas del estado civil para cada categoría del sexo.

Medidas de Asociación (Nominal)

S. \ E.C.	Soltero	Casado	Viudo	$n_{i.}$
Hombre	$n_{11} = 3$	$n_{12} = 4$	$n_{13} = 1$	$n_{1.} = 8$
Mujer	$n_{21} = 2$	$n_{22} = 3$	$n_{23} = 2$	$n_{2.} = 7$
$n_{.j}$	$n_{.1} = 5$	$n_{.2} = 7$	$n_{.3} = 3$	$n_{..} = 15$

Frec. abs. esperada en caso de ausencia a asociación

$$f_{e_{ij}} = \frac{n_{i.} \cdot n_{.j}}{n}$$

S. \ E.C.	Soltero	Casado	Viudo	$f_{e_{(i.,j)}}$
Hombre	$f_{e_{(1,1)}} = 2,6$	$f_{e_{(1,2)}} = 3,7$	$f_{e_{(1,3)}} = 1,6$	$f_{e_{(1.,j)}} = 7,9 \approx 8$
Mujer	$f_{e_{(2,1)}} = 2,3$	$f_{e_{(2,2)}} = 3,2$	$f_{e_{(2,3)}} = 1,4$	$f_{e_{(2.,j)}} = 6,9 \approx 7$
$f_{e_{(.,j)}}$	$f_{e_{(.,1)}} = 4,9 \approx 5$	$f_{e_{(.,2)}} = 6,9 \approx 7$	$f_{e_{(.,3)}} = 3$	$f_{e_{(.,j)}} \approx 15$

χ^2 de Pearson

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^k \frac{(e_{ij})^2}{f_{e_{ij}}}$$

S. \ E.C.	Soltero	Casado	Viudo
Hombre	$e_{11} = 0,4$	$e_{12} = 0,3$	$e_{13} = -0,6$
Mujer	$e_{21} = -0,3$	$e_{22} = -0,2$	$e_{23} = 0,6$

$$e_{ij} = n_{ij} - f_{e_{ij}}$$

$$\max \chi = n \cdot \min(r-1, k-1)$$

S. \ E.C.	Soltero	Casado	Viudo
Hombre	$\frac{e_{11}^2}{f_{e_{(1,1)}}} = 0,06$	0,02	0,22
Mujer	0,03	0,01	0,25

Coefficiente C de contingencia

$$C = \sqrt{\frac{\chi^2}{\chi^2 + n}}$$

$$C \in [0, 1)$$

V de Cramer

$$V = \sqrt{\frac{\chi^2}{n \cdot k}}$$

$$V \in [0, 1)$$

Medir Asociación (Ordinal)

S. \ E.C.	Soltero	Casado	Viudo	$n_{i.}$
Hombre	$n_{11} = 3$	$n_{12} = 4$	$n_{13} = 1$	$n_{1.} = 8$
Mujer	$n_{21} = 2$	$n_{22} = 3$	$n_{23} = 2$	$n_{2.} = 7$
$n_{.j}$	$n_{.1} = 5$	$n_{.2} = 7$	$n_{.3} = 3$	$n_{..} = 15$

Coefficiente predictivo (λ)

$$\lambda_{Y/K} = \frac{\sum_{i=0}^r \max_{j=1 \dots k} n_{ij} - \max_{j=1 \dots k} n_{.j}}{n - \max_{j=1 \dots k} n_{.j}}$$

$$\lambda_{X/Y} = \frac{\sum_{j=0}^k \max_{i=1 \dots r} n_{ij} - \max_{i=1 \dots r} n_{i.}}{n - \max_{i=1 \dots r} n_{i.}}$$

$$\lambda_{Y/X} = \frac{4 + 3 - 7}{15 - 7} = 0$$

Indica que la variable X (sexo) no tiene ningún poder predictivo sobre Y (estado civil).

$$\lambda_{X/Y} = \frac{3 + 4 + 2 - 8}{15 - 8} = 0,142$$

Escaso poder predictivo de Y (estado civil) sobre X (sexo).

MAL

AMBOS deben ser ordinales

Medir Asociación (Numeral)

S. \ E.C.	Soltero	Casado	Viudo	$n_{i.}$
Hombre	$n_{11} = 3$	$n_{12} = 4$	$n_{13} = 1$	$n_{1.} = 8$
Mujer	$n_{21} = 2$	$n_{22} = 3$	$n_{23} = 2$	$n_{2.} = 7$
$n_{.j}$	$n_{.1} = 5$	$n_{.2} = 7$	$n_{.3} = 3$	$n_{..} = 15$

Covarianza (S_{xy})

$$S_{xy} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})$$

$$\left(\frac{1}{n} \sum_{i=1}^r \sum_{j=1}^k n_{ij} \cdot x_i \cdot y_j \right) - \bar{x} \bar{y}$$

Coefficiente de Correlación

$$r_{xy} = \frac{S_{xy}}{S_x S_y}$$

$$\in [-1, 1]$$

• REGRESIÓN Y CORRELACIÓN LINEAL

Estudio simultáneo de varias variables

- Analizar relaciones entre ellas
- Establecer modelo matemático que las relacione.

$$\hat{Y} = \beta_0 + \beta_1 X_1 + \varepsilon$$

- ε = error
- $\beta = n^o$ (a, b, ...)
- $\hat{Y}$ = var. dep.
- X = var. indep.

Lineales

- Lineal simple: $\hat{Y} = \beta_0 + \beta_1 X$
- Parabólico: $\hat{Y} = \beta_0 + \beta_2 X + \beta_1 X^2$
- Cúbico: $\hat{Y} = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3$
- Polinómico: $\hat{Y} = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^2 + \dots + \beta_h X^h$

No Lineales

- Exponencial: $\hat{Y} = \beta_0 k^{\beta_1 X}$
- Potencial: $\hat{Y} = \beta_0 X^{\beta_1}$
- Hiperbólico: $\hat{Y} = 1/\beta_0 + \beta_1 X$
- Logístico: $\hat{Y} = 1/e^{-\beta_0 - \beta_1 X}$

- Correlación Simple

$$R_{xy} = \frac{S_{xy}}{S_x \cdot S_y}$$

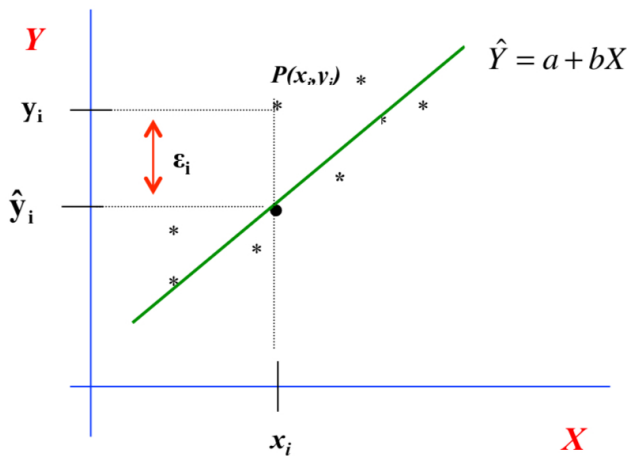
• Matriz correlación $\rightarrow T_{Y_{X_1 X_2}}$

$$\begin{pmatrix} 1 & R_{y x_1} & R_{y x_2} & \dots \\ R_{x_1 y} & 1 & R_{x_1 x_2} & \dots \\ R_{x_2 y} & R_{x_2 x_1} & 1 & \dots \\ \dots & \dots & \dots & \dots \end{pmatrix}$$

$$- R_{xy} = R_{yx}$$

$$- R_{yy} = R_{x_1 x_1} = R_{x_2 x_2} = \boxed{1}$$

Regresión Simple: Línea de Regresión



• $\hat{y}_i \rightarrow$ valor estimado de Y para $X = x_i$

• $\epsilon_i \rightarrow$ residuo:

$$\epsilon_i = y_i - \hat{y}_i$$

Para menor error:

$$H = \sum_{i=1}^n \epsilon^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i - a - b x_i)^2$$

$$\left. \begin{aligned} \bullet \frac{dH}{da} &= 2 \cdot (y_i - a - b x_i) = 0 \\ \bullet \frac{dH}{db} &= 2 \cdot (y_i - a - b x_i) \cdot x_i = 0 \end{aligned} \right\} \text{Ec. Normales}$$

$$a = \bar{y} - b \bar{x}$$

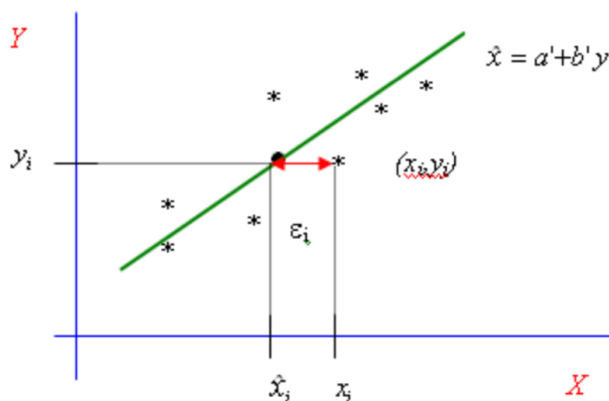
$$b = \frac{S_{xy}}{S_x^2}$$

Usar esto

$$\bar{\epsilon} = 0$$

Recta de regresión de X sobre Y: ($\hat{x} = a' + b'y$)

Estudio análogo.



$$a' = \bar{x} - b' \bar{y}$$

$$b' = \frac{S_{xy}}{S_y^2}$$

Relación pendientes

$$bb' = \frac{S_{xy}^2}{S_x^2 S_y^2} = R_{xy}^2$$

- Regresión Parabólica = caso lineal múltiple

$$\hat{y} = a + b_1 x_i + b_2 x_i^2 \rightarrow \hat{y} = a + b_1 x_i + b_2 z_i$$

$$\left. \begin{aligned} \bullet \frac{dH}{da} &= 2 \cdot (y_i - a - b_1 x_i - b_2 z_i) = 0 \\ \bullet \frac{dH}{db_1} &= 2 \cdot (y_i - a - b_1 x_i - b_2 z_i) x_i = 0 \\ \bullet \frac{dH}{db_2} &= 2 \cdot (y_i - a - b_1 x_i - b_2 z_i) z_i = 0 \end{aligned} \right\} \text{Ec. Normales}$$

$$a = \bar{y} - b_1 \bar{x} - b_2 \bar{z}$$

$$b_i = -\frac{A_{1,i+1}}{A_{1,1}} \rightarrow A_{11} = \text{Adjunta de elemento } a_{11} \text{ de matriz Covariante } (\Sigma) = \Sigma_{\vec{x}}$$

$$\Sigma = \begin{pmatrix} S_y^2 & S_{yx} & S_{yz} \\ S_{xy} & S_x^2 & S_{xz} \\ S_{zy} & S_{zx} & S_z^2 \end{pmatrix}$$

- Correlación Múltiple

$$R_{y\beta} = R_y = \frac{S_{y\beta}}{S_y S_\beta} \in [-1, 1]$$

$$R_y > R_{yx}; \quad \forall i=1, 2, \dots$$

- Varianza Residual

$$V(\varepsilon) = \frac{1}{n} \sum_{i=1}^n (\varepsilon_i - \bar{\varepsilon})^2 = \frac{1}{n} \sum_{i=1}^n (\varepsilon_i)$$

- Coefficiente de Determinación

$$R^2 = R_{xy}^2 = \left(\frac{S_{xy}}{S_x S_y} \right)^2$$

"Tanto % de la S_y^2 explica por el modelo"

- Correlación Parcial

Obj: Eliminar efectos distorsionantes de terceras variables

Dado modelos:

$$X = a + bz + \varepsilon_x; \quad y = a + bz + \varepsilon_y$$

ε_x representa la parte de X que no puede explicar.

- Coefficiente correlación entre x e y , parcial z

$$P_{xy,z} = P_{\varepsilon_x \varepsilon_y}$$

$$\hookrightarrow P_{xy,z} = \frac{P_{xy} - P_{xz} P_{yz}}{\sqrt{1 - P_{xz}^2} \sqrt{1 - P_{yz}^2}} = - \frac{A_{12}}{\sqrt{A_{11} A_{22}}}$$

$$\rho_{xy,z}^2 > \rho_{xy}^2 \Rightarrow \begin{cases} \text{La variable } Z \text{ oculta o amortigua la} \\ \text{dependencia entre } X \text{ e } Y \end{cases}$$

$$\rho_{xy,z}^2 \cong 0 \Rightarrow \begin{cases} \text{La interdependencia entre } X \text{ e } Y \text{ se debe} \\ \text{casi exclusivamente al efecto de } Z \end{cases}$$

$$\rho_{xy,z}^2 < \rho_{xy}^2 \Rightarrow \begin{cases} \text{La interdependencia entre } X \text{ e } Y \text{ se debe} \\ \text{parcialmente a la influencia de } Z \end{cases}$$