

ESTADÍSTICA I · FORMULARIO

Formulario del Tema 4

Distribuciones de frecuencias bidimensionales

NOVA

Cómo usar este formulario

- **Qué entra (plan docente):** 4.1 Distribuciones conjuntas. Independencia · 4.2 Covarianza y correlación · 4.3 Regresión lineal.
- Varianzas y covarianza, siempre entre $n - 1$.

4.1 · La tabla de doble entrada

$X \backslash Y$	y_1	...	y_j	Total
x_1	n_{11}	...	n_{1j}	$n(x_1)$
x_i	n_{i1}	...	n_{ij}	$n(x_i)$
Total	$n(y_1)$	...	$n(y_j)$	n

CONJUNTA

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

MARGINAL

$$n(x_i) \text{ fila} \cdot n(y_j) \text{ columna}$$

CONDICIONADA

$$f_{i/j} = \frac{n_{ij}}{n(y_j)}$$

Entre el total de la condición

💡 «El 20 % de la muestra...» → entre n · «De los que...» → entre el total de esa fila o columna

4.1 · Independencia estadística

EN RELATIVAS

$$f_{ij} = f(x_i) \cdot f(y_j)$$

FRECUENCIA ESPERADA

$$n_{ij} = \frac{n(x_i) \cdot n(y_j)}{n}$$

💡 Tiene que cumplirse en **todas** las casillas: si falla en una, no son independientes

4.2 · Covarianza

DEFINICIÓN

$$S_{XY} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n - 1}$$

PARA CALCULAR MÁS RÁPIDO

$$S_{XY} = \frac{\sum x_i y_i - n \bar{x} \bar{y}}{n - 1}$$

SIGNO

> 0 positiva · < 0 negativa · = 0 sin
relación lineal

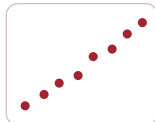
CAMBIO DE ORIGEN Y ESCALA

$$S_{X'Y'} = b d S_{XY}$$

$$X' = a + bX, Y' = c + dY$$

💡 Independientes $\Rightarrow S_{XY} = 0$, pero $S_{XY} = 0$ no implica independencia

4.2 · Coeficiente de correlación lineal de Pearson



$r \approx +1$



$r \approx 0$



$r \approx -1$

CORRELACIÓN

$$r_{XY} = \frac{S_{XY}}{S_X S_Y}$$

RANGO

$$-1 \leq r_{XY} \leq 1$$

Sin unidades

QUÉ MIDE

Intensidad y signo de la relación **lineal**
entre variables cuantitativas

4.2 · Vector de medias y matrices

VECTOR DE MEDIAS

$$m = \begin{pmatrix} \bar{x} \\ \bar{y} \end{pmatrix}$$

VARIANZAS Y
COVARIANZAS

$$S = \begin{pmatrix} S_X^2 & S_{XY} \\ S_{XY} & S_Y^2 \end{pmatrix}$$

CORRELACIONES

$$r = \begin{pmatrix} 1 & r_{XY} \\ r_{XY} & 1 \end{pmatrix}$$

💡 Con los sumatorios: medias, después varianzas y covarianza con las fórmulas rápidas, entre $n - 1$

4.3 · Recta de regresión de Y sobre X **RECTA**

$$\hat{y} = a + b x$$

X explica, Y se predice

PENDIENTE

$$b = \frac{S_{XY}}{S_X^2}$$

Signo de S_{XY} y de r

ORDENADA EN EL ORIGEN

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

Pasa por $(\bar{x}, \bar{y})$

PREDICCIÓN

$$\hat{y}(x_0) = a + b x_0$$

COEF. DE DETERMINACIÓN

$$R^2 = r^2 = \frac{S_{XY}^2}{S_X^2 S_Y^2}$$

De 0 a 1 · % de la variación de Y que explica la recta

💡 No confundas r con R^2 : si $r = 0,86$, la recta explica el $0,86^2 = 74$ %