TEMA 2

ESTADISTICA BIDIMENSIONAL

TABLAS DE DISTRIBUCION DE FRECUENCIAS BIDIMENSIONALES

TAB LA DISTRIBUCION DE FRECUENCIAS

n_{ij} → **Frecuencia absoluta**: número de individuos de la población que presentan la modalidad x_i de X y la modalidad y_j de Y .

Tabla de frecuencias absolutas de una variable bidimensional

							suma
	(X/Y)	y_1	y_2	y_3	…	y_q	$n_{i\cdot}$
	x_1	n_{11}	n_{12}	n_{13}	…	n_{1q}	$n_{1\cdot}$
	x_2	n_{21}	n_{22}	n_{23}	…	n_{2q}	$n_{2\cdot}$
	x_3	n_{31}	n_{32}	n_{33}	…	n_{3q}	$n_{3\cdot}$
	…	…	…	…	…	…	…
	x_p	n_{p1}	n_{p2}	n_{p3}	…	n_{pq}	$n_{p\cdot}$
suma	$n_{\cdot j}$	$n_{\cdot 1}$	$n_{\cdot 2}$	$n_{\cdot 3}$	…	$n_{\cdot q}$	$n_{\cdot\cdot}=N$

Nota: el primer subíndice indica el número de fila y el segundo subíndice indica el número de columna

f_{ij} → **Frecuencia relativa**: proporción de individuos de la población que presentan la modalidad x_i de X y la modalidad y_j de Y. Se obtiene dividiendo la frecuencia absoluta entre el número de elementos de la población (N).

Tabla de frecuencias relativas de una variable bidimensional

							suma
	(X/Y)	y_1	y_2	y_3	…	y_q	$n_{i\cdot}$
	x_1	f_{11}	f_{12}	f_{13}	…	f_{1q}	$f_{1\cdot}$
	x_2	f_{21}	f_{22}	f_{23}	…	f_{2q}	$f_{2\cdot}$
	x_3	f_{31}	f_{32}	f_{33}	…	f_{3q}	$f_{3\cdot}$
	…	…	…	…	…	…	…
	x_p	f_{p1}	f_{p2}	f_{p3}	…	f_{pq}	$f_{p\cdot}$
suma	$f_{\cdot j}$	$f_{\cdot 1}$	$f_{\cdot 2}$	$f_{\cdot 3}$	…	$f_{\cdot q}$	$f_{\cdot\cdot}=1$

DISTRIBUCIONES MARGINALES

Estas distribuciones nos indican cómo se distribuye una variable independientemente de los valores que tome la otra.

Distribución marginal de *X*:

X	$n_{i\cdot}$
x_1	$n_{1\cdot}$
x_2	$n_{2\cdot}$
x_3	$n_{3\cdot}$
…	…
x_p	$n_{p\cdot}$
suma	$n_{\cdot\cdot}=N$

Distribución marginal de *Y*:

Y	$n_{\cdot j}$
y_1	$n_{\cdot 1}$
y_2	$n_{\cdot 2}$
y_3	$n_{\cdot 3}$
…	…
y_q	$n_{\cdot q}$
suma	$n_{\cdot\cdot}=N$

Existe una única distribución marginal de *X* y una única distribución marginal de *Y*.

DISTRIBUCIONES CONDICIONADAS

Son distribuciones unidimensionales obtenidas a partir de las bidimensionales, manteniendo fijo un valor en una de las variables y considerando los valores que toma la otra con sus respectivas frecuencias

Distribución de X condicionada al valor y_j de Y:

$X \mid Y=y_j$	$n_{i\mid j}$
x_1	n_{1j}
x_2	n_{2j}
x_3	n_{3j}
...	...
x_p	n_{pj}
suma	$n_{\cdot j}$

Distribución de Y condicionada al valor x_i de X:

$Y \mid X=x_i$	$n_{j\mid i}$
y_1	n_{i1}
y_2	n_{i2}
y_3	n_{i3}
...	...
y_q	n_{iq}
suma	$n_{i\cdot}$

COVARIANZA

COVARIANZA

$$\sigma_{XY} = Cov(X,Y) = \frac{1}{N}\sum_{j=1}^{q}\sum_{i=1}^{p}(x_i - \bar{x})(y_j - \bar{y})n_{ij} = \frac{1}{N}\sum_{j=1}^{q}\sum_{i=1}^{p}x_i y_j n_{ij} - \overline{x}\overline{y}$$

COVARIANZA CALCULO

i,j i,j i,j

Ejemplo: Calcular la covarianza del Ejemplo 2.2

$X \setminus Y$	1	2	3	4	5	6	$n_{i.}$	$n_{i.}x_i$	$n_{i.}x_i^2$	$\sum_{j=1}^{6} x_i y_j n_{ij}$
1	0	1	0	0	0	3	4	4	4	0+2+0+0+0+18=20
2	1	0	1	0	1	0	3	6	12	2+0+6+0+10+0=18
3	1	0	0	2	1	1	5	15	45	3+24+15+18=60
4	2	3	0	0	0	1	6	24	96	8+24+15+18=56
5	2	1	1	0	0	0	4	20	100	10+10+15=35
6	0	1	0	0	1	0	2	12	72	12+30=42
$n_{.j}$	6	6	2	2	3	5	24	81	329	231
$n_{.j}y_j$	6	12	6	8	15	30	77			
$n_{.j}y_j^2$	6	24	18	32	75	180	335			

$$\sigma_{xy} = \sum_{i,j} x_i y_j f_{ij} - \overline{xy} = \frac{231}{24} - 3{,}3 \times 3{,}2 = -0{,}9$$

Ejemplo: Obtener la covarianza de las variables X e Y

X\Y	150	170	190	210	Suma	$x_i*n_{i\cdot}$	$x_i*y_j*n_{ij}$
50	10	6	2	0	18	900	145000
70	8	12	6	2	28	1960	336000
90	1	8	10	6	25	2250	420300
Suma	19	26	18	8	71	5110	901300
$y_j*n_{\cdot j}$	2850	4420	3420	1680	12370		

Cálculo por filas de $x_i*y_j*n_{ij}$

1. 50*(150*10+170*6+190*2+210*0)=145000
2. 70*(150*8+170*12+190*6+210*2)=336000
3. 90*(150*1+170*8+190*10+210*6)=420300

$$\bar{x} = \frac{5110}{71} = 71{,}9718$$

$$\bar{y} = \frac{12370}{71} = 174{,}2254$$

$$\sigma_{XY} = \frac{901300}{71} - 71{,}9718 * 174{,}2256$$

$$= 155{,}0486$$

REGRESION LINEAL

Recta de regresión de Y respecto X:

$$Y = a + bX$$

b representa la **pendiente de la recta**:

$$b = \frac{\sigma_{xy}}{\sigma_x^2}$$

- Si **b > 0**, representa el **incremento** que se produce en Y al aumentar en una unidad X.
- Si **b < 0**, representa la **disminución** que se produce en Y al aumentar en una unidad X.
- Si **b = 0**, Y no depende linealmente de X (lo que no significa que sean independientes).

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

a, **ordenada en el origen**. Representa el valor que toma Y cuando X = 0.

COEFICIENTE CORRELACION DE PEARSON

$$r = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \quad ; \quad -1 \leq r \leq 1$$

COEFICIENTE DE DETERMINACION

Porcentaje de variabilidad de Y explicada o debida a la regresión

$$R^2 = \frac{\sigma^2_{\text{exp}}}{\sigma^2_y} = 1 - \frac{\sigma^2_{\text{res}}}{\sigma^2_y}, \quad 0 \leq R^2 \leq 1$$

$$R^2 = r^2 = \frac{\sigma^2_{xy}}{\sigma^2_x \sigma^2_y}, \quad 0 \leq R^2 \leq 1$$

- R^2 es una medida de la **bondad de ajuste**, del grado de ajuste de la regresión a los datos.