

5. Estudiar la dependencia o independencia de las variables en cada una de las siguientes distribuciones. Dar, en cada caso, las curvas de regresión y la covarianza de las dos variables.

X\Y	1	2	3	4	5
10	2	4	6	10	8
20	1	2	3	5	4
30	3	6	9	15	12
40	4	8	12	20	16

Se dice que el carácter Y es independiente estadísticamente del carácter X, si las distribuciones de Y condicionadas a cada valor de x, son idénticas para todo $x_i, i=1,2,\dots,k$. Es decir:

$$f_j^i \equiv f_{j/i}$$

Por tanto, se debe cumplir lo siguiente:

$$\frac{n_{1j}}{n_{1.}} = \frac{n_{2j}}{n_{2.}} = \dots = \frac{n_{ij}}{n_{i.}} = \dots = \frac{n_{kj}}{n_{k.}} \quad \forall j = 1, 2, \dots, p$$

En el caso de la tabla que tenemos arriba, vamos a comprobar la independencia de las dos variables. Cabe decir que si resulta que Y sea independiente al carácter X, el recíproco también es cierto.

$$j = 1 \quad \frac{2}{30} = \frac{1}{15} = \frac{3}{45} = \frac{4}{60}$$

$$j = 2 \quad \frac{4}{30} = \frac{2}{15} = \frac{6}{45} = \frac{8}{60}$$

$$j = 3 \quad \frac{6}{30} = \frac{3}{15} = \frac{9}{45} = \frac{12}{60}$$

$$j = 4 \quad \frac{10}{30} = \frac{5}{15} = \frac{15}{45} = \frac{20}{60}$$

$$j = 5 \quad \frac{8}{30} = \frac{4}{15} = \frac{12}{45} = \frac{16}{60}$$

Como se puede observar, Y es independiente del carácter X, así que no tiene sentido estudiar curvas de regresión. Además podemos afirmar:

$$\sigma_{xy} = 0$$

X\Y	1	2	3
-1	0	1	0
0	1	0	1
1	0	1	0

Es evidente que estas dos variables no son independientes, puesto que si algún $n_{ij} = 0$, la igualdad

$$\frac{n_{1j}}{n_{1.}} = \frac{n_{2j}}{n_{2.}} = \dots = \frac{n_{ij}}{n_{i.}} = \dots = \frac{n_{kj}}{n_{k.}} \quad \forall j = 1, 2, \dots, p$$

no se va a dar a menos que $n_{ij} = 0 \quad \forall i, j$, lo cual no tendría sentido porque sería estudiar una variable que no ha presentado ninguna distribución de frecuencias.

Observando ahora la posible dependencia funcional, observamos que X no depende funcionalmente de Y, porque a la modalidad 2 de Y, le corresponden dos posibles modalidades de X (señalado en la tabla).

Procedemos a calcular la covarianza:

$$\sigma_{xy} = m_{11} - \bar{x} \cdot \bar{y}$$

X\Y	1	2	3	$n_{i.} \cdot x_y$	$m_{11} = \frac{1}{n} \sum_i \sum_j n_{ij} \cdot x_i \cdot y_j$
-1	0	1	0	-1	-2
0	1	0	1	0	0
1	0	1	0	1	2
				$\bar{x} = 0$	$m_{11} = 0$

Como la media de x es 0, la covarianza será igual a m_{11} . Resulta que:

$$\sigma_{xy} = 0$$

Además, podemos observar que si las variables son independientes, la covarianza es 0, pero el recíproco no podemos asegurarlo. Aquí vemos un caso en el que las variables son dependientes estadísticamente y la covarianza resulta 0.

Ahora calculamos la curva de regresión de tipo 1. de Y/X, que sabemos que es la curva que pasa por los puntos $(x_i, \bar{y}_i); \quad i = 1, \dots, p$.

Punto 1: $(-1, 2)$ Punto 2: $(0, 2)$ Punto 3: $(1, 2)$

10. Calcular el coeficiente de correlación lineal de dos variables cuyas rectas de regresión son:

$$x + 4y = 1$$

$$x + 5y = 2$$

Supongamos que la primera es la recta para X/Y y la segunda para Y/X.

Además, sabemos que ambas rectas tienen el mismo signo en la pendiente (que lo proporciona la covarianza).

$$(i) \quad x = -4y + 1 \quad (ii) \quad y = -\frac{1}{5}x + \frac{2}{5}$$

$$a_1 = \frac{\sigma_{xy}}{\sigma_y^2} \quad a_2 = \frac{\sigma_{xy}}{\sigma_x^2}$$

$$a_1 \cdot a_2 = r^2 = \frac{\sigma_{xy}^2}{\sigma_x^2 \cdot \sigma_y^2} = \frac{4}{5}$$

Nos piden el coeficiente de correlación lineal, r, que sabemos que tiene que ser negativo por que la covarianza es negativa (en las dos rectas la pendiente es negativa).

$$r = -\sqrt{r^2} = -\frac{\sqrt{4}}{\sqrt{5}} = -\frac{2}{\sqrt{5}}$$

Dicho resultado tiene sentido por que $-1 \leq r \leq 1$.