PROBLEMAS RESUELTOS DE LA RELACIÓN 3

12) Apartado a) Sea (X,Y) un vector aleatorio con función masa de probabilidad. Analizar la dependencia entre las variables y calcular las curvas de regresión. Comentar los resultados.

X \ Y	10	15	20
1	0	2/7	0
2	1/7	0	0
3	0	0	3/7
4	0	1/7	0

En primer lugar, obtenemos las distribuciones marginales sumando filas y columnas

X \ Y	10	15	20	
1	0	2/7	0	2/7
2	1/7	0	0	1/7
3	0	0	3/7	3/7
4	0	1/7	0	1/7
	1/7	3/7	3/7	

Para estudiar la dependencia de X a partir de Y calculamos

- $\mathrm{E}[X/Y]$

  La $\mathrm{E}[X/Y]$ es una variable aleatoria que toma los valores

  $\mathrm{E}[X/Y=10]=2$ si $Y=10$

  $\mathrm{E}[X/Y=15]=1\frac{2}{3}+4\frac{1}{3}=2$ si $Y=15$

  $\mathrm{E}[X/Y=20]=3$ si $Y=20$

  Así la curva de regresión de X sobre Y está determinada por los puntos $(2,\ 10)$, $(2,\ 15)$ y $(3,\ 20)$.

- $\eta^2_{X/Y}=\dfrac{\mathrm{Var}(\mathrm{E}[X/Y])}{\mathrm{Var}(X)}$

  En general, hemos calculado $\eta^2_{X/Y}=1-\dfrac{\mathrm{E}[\mathrm{Var}(X/Y)]}{\mathrm{Var}(X)}$. En este caso, de forma alternativa, calculemos $\mathrm{Var}(\mathrm{E}[X/Y])$.

  $$\mathrm{Var}(\mathrm{E}[X/Y])=\mathrm{E}((\mathrm{E}[X/Y])^2)-(\mathrm{E}(\mathrm{E}[X/Y]))^2=\mathrm{E}((\mathrm{E}[X/Y])^2)-(\mathrm{E}[X])^2$$

  y

  $$\mathrm{E}((\mathrm{E}[X/Y])^2)=(\mathrm{E}[X/Y=10])^2\mathrm{P}[Y=10]+(\mathrm{E}[X/Y=15])^2\mathrm{P}[Y=15]+$$

$$+(\mathrm{E}[X/Y=20])^2\mathrm{P}[Y=20]$$

$$=2^2\frac{1}{7}+2^2\frac{3}{7}+3^2\frac{3}{7}=\frac{43}{7}$$

Dado que

$\mathrm{E}[X]=1\frac{2}{7}+2\frac{1}{7}+3\frac{3}{7}+4\frac{1}{7}=\frac{17}{7}$,

$\mathrm{E}[X^2]=1^2\frac{2}{7}+2^2\frac{1}{7}+3^2\frac{3}{7}+4^2\frac{1}{7}=\frac{49}{7}=7$,

$\mathrm{Var}(X)=7-\left(\frac{17}{7}\right)^2=\frac{54}{49}$,

finalmente

$$\eta^2_{X/Y}=\frac{\mathrm{Var}(\mathrm{E}[X/Y])}{\mathrm{Var}(X)}=\frac{43/7-(17/7)^2}{54/49}=\frac{12}{54}.$$

Para estudiar la dependencia de Y a partir de X calculamos

- $\mathrm{E}[Y/X]$

  La $\mathrm{E}[X/Y]$ es una variable aleatoria que toma los valores

  $\mathrm{E}[Y/X=1]=15$ si $X=1$

  $\mathrm{E}[Y/X=2]=10$ si $X=2$

  $\mathrm{E}[Y/X=3]=20$ si $X=3$

  $\mathrm{E}[Y/X=4]=15$ si $X=4$

  Así la curva de regresión de Y sobre X está determinada por los puntos $(1,\ 15)$, $(2,\ 10)$, $(3,\ 20)$ y $(4,\ 15)$.

- $\eta^2_{Y/X}=\dfrac{\mathrm{Var}(\mathrm{E}[Y/X])}{\mathrm{Var}(Y)}$

$$\mathrm{Var}(\mathrm{E}[Y/X])=\mathrm{E}((\mathrm{E}[Y/X])^2)-(\mathrm{E}[Y])^2$$

  y dado que

$$\mathrm{E}((\mathrm{E}[Y/X])^2)=(\mathrm{E}[Y/X=1])^2\mathrm{P}[X=1]+(\mathrm{E}[Y/X=2])^2\mathrm{P}[X=2]+(\mathrm{E}[Y/X=3])^2\mathrm{P}[X=3]+$$

$$+(\mathrm{E}[Y/X=4])^2\mathrm{P}[X=4]$$

$$= 15^2\frac{2}{7} + 10^2\frac{1}{7} + 20^2\frac{3}{7} + 15^2\frac{1}{7} = \frac{1975}{7}$$

Dado que

$\mathrm{E}[Y] = 10\frac{1}{7} + 15\frac{3}{7} + 20\frac{3}{7} = \frac{115}{7}$,

$\mathrm{E}[Y^2] = 10^2\frac{1}{7} + 15^2\frac{3}{7} + 20^2\frac{3}{7} = \frac{1975}{7}$,

$\mathrm{Var}(Y) = \frac{1975}{7} - \left(\frac{115}{7}\right)^2$,

finalmente

$$\eta^2_{Y/X} = \frac{\mathrm{Var}(\mathrm{E}[Y/X])}{\mathrm{Var}(Y)} = \frac{\frac{1975}{7} - \left(\frac{115}{7}\right)^2}{\frac{1975}{7} - \left(\frac{115}{7}\right)^2} = 1.$$

Se concluye que Y depende funcionalmente de X pero X no depende funcionalmente de Y. De hecho, $\eta^2_{X/Y} = 0.2222$ nos indica que el grado de dependencia de X sobre Y es bajo ya que la proporción de la varianza de la variable X que queda explicada por la regresión es del 22.22 %.

Notemos que la dependencia funcional de Y sobre X se podría haber deducido directamente de la tabla ya que a cada valor de X le corresponde un único valor de Y.

14) Sea X la variable aleatoria número de balanzas e Y el número de dependientes en los puntos de venta de un mercado. Si la función masa de probabilidad de (X, Y) viene dada por

X \ Y	1	2	3	4
1	1/24	2/24	0	0
2	1/24	2/24	3/24	1/24
3	0	1/24	2/24	6/24
4	0	0	2/24	3/24

a) Determinar las rectas de regresión.

En primer lugar, obtenemos las distribuciones marginales sumando filas y columnas

X \ Y	1	2	3	4	
1	1/24	2/24	0	0	3/24
2	1/24	2/24	3/24	1/24	7/24
3	0	1/24	2/24	6/24	9/24
4	0	0	2/24	3/24	5/24
	2/24	5/24	7/24	10/24	

Así, dado que

$$\mathrm{E}[X] = 1\frac{3}{24} + 2\frac{7}{24} + 3\frac{9}{24} + 4\frac{5}{24} = \frac{8}{3},$$

$$\mathrm{E}[X^2] = 1^2\frac{3}{24} + 2^2\frac{7}{24} + 3^2\frac{9}{24} + 4^2\frac{5}{24} = 8,$$

$$\mathrm{Var}(X) = 8 - \left(\frac{8}{3}\right)^2 = \frac{8}{9},$$

$$\mathrm{E}[Y] = 1\frac{2}{24} + 2\frac{5}{24} + 3\frac{7}{24} + 4\frac{10}{24} = \frac{73}{24},$$

$$\mathrm{E}[Y^2] = 1^2\frac{2}{24} + 2^2\frac{5}{24} + 3^2\frac{7}{24} + 4^2\frac{10}{24} = \frac{245}{24},$$

$$\mathrm{Var}(Y) = \frac{245}{24} - \left(\frac{73}{24}\right)^2 = \frac{551}{576},$$

$$\mathrm{E}[X\cdot Y] = 1\cdot 1\frac{1}{24} + 1\cdot 2\frac{2}{24} + 2\cdot 1\frac{1}{24} + 2\cdot 2\frac{2}{24} + 2\cdot 3\frac{3}{24} + 2\cdot 4\frac{1}{24} + 3\cdot 2\frac{1}{24} +$$

$$+3\cdot 3\frac{2}{24} + 3\cdot 4\frac{6}{24} + 4\cdot 3\frac{2}{24} + 4\cdot 4\frac{3}{24} = \frac{209}{24},$$

$$\mathrm{Cov}(X,Y) = \frac{209}{24} - \frac{8}{3}\frac{73}{24} = \frac{43}{72},$$

las rectas de regresión son

$$Y/X \quad y - \frac{73}{24} = \frac{43/72}{8/9}\left(x - \frac{8}{3}\right)$$

$$X/Y \quad y - \frac{8}{3} = \frac{43/72}{551/576}\left(y - \frac{73}{24}\right)$$

b) ¿Es apropiado suponer que existe una relación lineal entre las variables?

Dado que

$$\rho^2_{X,Y} = \frac{(\mathrm{Cov}(X,Y))^2}{\mathrm{Var}(X)\,\mathrm{Var}(Y)} = \frac{\left(\frac{43}{72}\right)^2}{\frac{8}{9}\frac{551}{573}} = 0.417279$$

y

$$\rho_{X,Y} = 0.6459$$

no es apropiado suponer que existe una relación lineal entre las variables.