TEMA 7: Estadística bidimensional

Full de preparació

Aquest full s'ha de lliurà el dia de la prova

Nom: .... Curs: ....

1. Completeu la taula

	1	2	4	7	9	11	
1	1	2		1	0	0	5
3	0			1	1	0	4
4	1	0	2	1	1	3	
5	1	1	3	2	4	0	
6		1	1		1	0	4
7	0	0	0	1	3	1	
	4	5	8	6	10	4	

a) Representeu gràficament
b) Trobeu el coeficient de correlació
c) Cal buscar les rectes de correlació lineal per trobar valors d'una variable en funció de l'altre?. Raoneu la resposta.

2. En un grup de 8 pacients es mesuren pes i edat obtenint els següents resultats:

	Resultat de les mesures							
$X \equiv$ edat	12	8	10	11	7	7	10	14
$Y \equiv$ pes	58	42	51	54	40	39	49	56

a) Hi ha una relació lineal entre ambdues variables?
b) Calculeu la recta de regressió de l'edat en funció del pes i la recta del pes en funció de l'edat;
c) Trobeu la bondat de l'ajustament.

3. Les qualificacions de 40 alumnes en psicologia evolutiva i estadística són els següents:

Psicologia x_i	3	4	5	6	6	7	7	8	10
Estadística y_i	2	5	5	6	7	6	7	9	10
Nº de alumnes f_i	4	6	12	4	5	4	2	1	2

a) Trobeu l'equació de la recta de regressió de les notes d'estadística respecte les de psicologia evolutiva;
b) Quina serà la nota esperada en estadística per un alumne amb un 4,5 en psicologia?

4. Les qualificacions obtingudes per 10 alumnes en matemàtiques i en música són:

Matemàtiques	6	4	8	5	3,5	7	5	10	5	4
Música	6,5	4,5	7	5	4	8	7	10	6	5

a) Calculeu la covariància i el coeficient de correlació;
b) Hi ha correlació entre les dues variables? Justifiqueu la resposta;
c) Quina serà la nota que cal esperar en música per un alumne que hagués obtingut un 8,3 en matemàtiques?

5. Cinc noies de 2, 3, 5, 7 i 8 anys d'edat pesen respectivament 14, 20, 30, 42 i 44 Kg. Calculeu l'equació de la recta de regressió de l'edat sobre el pes. Quin serà el pes aproximat d'una noia de 6 anys?

6. Una associació dedicada a la protecció dels infants estudia la relació entre mortalitat infantil a cada país i el nombre de llits d'hospital per cada mil habitants,

x	50	100	70	60	120	180	200	250	30	90
y	5	2	2,5	3,75	4	1	1,25	0,75	7	3

on x – nombre de llits

y - % de mortalitat.

a) Trobeu les rectes de regressió i el coeficient de correlació lineal;
b) Si es disposen de 175 llits per mil habitants, quin percentatge de mortalitat cal esperar?. L'estimació és fiable?. Raoneu la resposta.

Solució 1:

x \ y	1	2	4	7	9	11	
1	1	2	**1**	1	0	0	5
3	0	**1**	**1**	1	1	0	4
4	1	0	2	1	1	3	**8**
5	1	1	3	2	4	0	**11**
6	**1**	1	1	**0**	1	0	4
7	0	0	0	1	3	1	**5**
	4	5	8	6	10	4	**37**

b) **x** = $\frac{1.5+3.4+4.8+5.11+6.4+7.5}{37}=4{,}405$ **;**

y = $\frac{1\,4+2\,5+4\,8+7.6+9\,10+11\,4}{37}$ = 6

M_{xy} [3] = $\frac{\sum_{ij} x_i y_j n_{ij}}{N}$ = 28,378,

$S_{xy} = M_{xy} - \mathbf{x}.\mathbf{y} = 1{,}948$

$S_x^2 = \frac{1.5+3^2.4+4^2\,8+5^2\,11+6^2\,4+7^2\,5}{37}-(4{,}405)^2=3{,}11$; S_x=1,764

S_y^2 = 47,027 - 36 = 11,027; S_y = 3,321

El coeficient de correlació lineal **r** = $\frac{1{,}948}{(1{,}764)(3{,}321)}$ **= 0,3325 <0,40, correlació baixa.**

m_{yx} = 1,948/3,11=0,626 i m_{xy} = 1,948/11,027=0,177 son els coeficients de regressió.

Las rectes de regressió son:
y - 6 = 0,626 (**x** - 4,405) de Y sobre X, y x - 4,405 = 0,177(**y** - 6) de x sobre y

c) Las rectes de regressió no serveixen per fer prediccions fiables d'una variable respecte l'altre ja que la correlació és baixa (el valor està molt lluny de la unitat).

Solució 2:

$$r = \frac{\mathcal{S}_{XY}}{\mathcal{S}_X\,\mathcal{S}_Y} = \frac{15,2031}{2,3150 \times 6,9631} = 0,9431$$

$$\begin{aligned}
\sum_{i=1}^{8} x_i = 79 &\implies \overline{x} = \frac{79}{8} = 9,875 \text{ años} \\
\sum_{i=1}^{8} y_i = 389 &\implies \overline{y} = \frac{389}{8} = 48,625 \text{ Kg} \\
\sum_{i=1}^{8} x_i^2 = 823 &\implies \mathcal{S}_X^2 = \frac{823}{8} - 9,875^2 = 5,3594 \text{ años}^2 \\
&\implies \mathcal{S}_X = 2,3150 \text{ años} \\
\sum_{i=1}^{8} y_i^2 = 19.303 &\implies \mathcal{S}_Y^2 = \frac{19.303}{8} - 48,625^2 = 48,4844 \text{ Kg}^2 \\
&\implies \mathcal{S}_Y = 6,9631 \text{ Kg} \\
\sum_{i=1}^{8} x_i y_i = 3.963 &\implies \mathcal{S}_{XY} = \frac{3.963}{8} - 9,875 \times 48,625 = 15,2031 \text{ Kg} \cdot \text{año}
\end{aligned}$$

La recta de regressió del pes en funció de la edat es

$$\begin{aligned}
\hat{Y} &= a_1 + b_1 X = 20,6126 + 2,8367 \cdot X \\
a_1 &= \overline{y} - b_1 \overline{x} = 20,6126 \text{ Kg} \\
b_1 &= \frac{\mathcal{S}_{XY}}{\mathcal{S}_X^2} = 2,8367 \text{ Kg/año}
\end{aligned}$$

La recta de regressió de la edat en funció del pes es

$$\begin{aligned}
\hat{X} &= a_2 + b_2 Y = -5,3738 + 0,3136 \cdot Y \\
a_2 &= \overline{x} - b_2 \overline{y} = -5,3738 \text{ años} \\
b_2 &= \frac{\mathcal{S}_{XY}}{\mathcal{S}_Y^2} = 0,3136 \text{ años/Kg}
\end{aligned}$$

La bondat de l' ajust es

$$R^2_{X|Y} = R^2_{Y|X} = r^2 = 0,8894$$

es a dir, el $88,94\%$ de la variabilitat del pes en funció de la edat es explicada per la recta de regressió corresponent. De la mateixa manera es pot dir que hi ha un $100 - 88,94\% = 11,06\%$ per de variància que no s'explica per les rectes de regressió.

La quantitat que varia el pes d'un pacient cada any és, segons la recta de regressió del pes en funció de l'edat, el pendent de la recta, es a dir, b_1=2,8367 Kg/any. Quan dues persones tenen diferent pes, en promig, la diferència d'edat entre totes dues és de b_2=0,3136 anys/Kg de diferència.

Solució 3:

				Mitjana aritmètica		Variància		Covariància
	xi	**yi**	**fi**	**fi . xi**	**fi . yi**	**fi . xi^2**	**fi . yi^2**	**fi . xi . yi**
	3	**2**	**4**	12	8	36	16	24
	4	**5**	**6**	24	30	96	150	120
	5	**5**	**12**	60	60	300	300	300
	6	**6**	**4**	24	24	144	144	144
	6	**7**	**5**	30	35	180	245	210
	7	**6**	**4**	28	24	196	144	168
	7	**7**	**2**	14	14	98	98	98
	8	**9**	**1**	8	9	64	81	72
	9	**10**	**2**	20	20	200	200	200
$\sum$			40	220	224	1314	1378	1336

Medias aritméticas

$\bar{x} = \frac{\sum x_i \cdot fi}{N} = \frac{220}{40} = 55$ $\quad \bar{x} = 5{,}5$ $\qquad \bar{y} = \frac{\sum y_i \cdot fi}{N} = \frac{224}{40} = 5{,}6$ $\quad \bar{y} = 5{,}6$

Varianzas y desviaciones típicas

$\sigma_x^2 = \frac{\sum fi \cdot (x_i)^2}{N} - (\bar{x})^2 = \frac{1314}{40} - (5{,}5)^2 = 2{,}60$ $\qquad \sigma_x = \sqrt{2{,}6} = 1{,}61$

$\sigma_y^2 = \frac{\sum fi \cdot (y_i)^2}{N} - (\bar{y})^2 = \frac{1378}{40} - (5{,}6)^2 = 3{,}09$ $\qquad \sigma_y = \sqrt{3{,}09} = 1{,}76$

Covarianza $\sigma_{xy} \Rightarrow \sigma_{xy} = \frac{\sum fi \cdot xi \cdot yi}{N} - \bar{x} \cdot \bar{y} = \frac{1336}{40} - (5{,}5 \cdot 5{,}6) = 2{,}6$

Correlación lineal de Pearson $r \Rightarrow r = \frac{\sigma_{xy}}{\sigma_x \times \sigma_y} = \frac{2{,}6}{1{,}61 \cdot 1{,}76} = 0{,}92 \Rightarrow r = 0{,}92$

La covarianza es positiva, correlación positiva fuerte.

El valor de r está muy próximo a 1, la estimación realizada estará muy cerca del valor real.

a) Recta de regresión de y sobre x $\Rightarrow \quad y = \bar{y} + \frac{\sigma_{xy}}{\sigma_x^2}(x - \bar{x})$

$y = 5{,}6 + \frac{2{,}6}{2{,}6}(x - 5{,}5) \Rightarrow y = 5{,}6 + 1(x - 5{,}5) \Rightarrow y = x + 0{,}1$

b) Nota esperada en estadística habiendo obtenido un 4,5 en psicología.

$y = 4{,}5 + 0{,}1 = 4{,}6$

Solució 4:

a) Covariància = 3,075. Coeficient de correlació r = 0,92.

b) Correlació positiva forta

c) Recta de regressió: y = 1,6 + 0,817 x La nota esperada de Música = 8, 38

Solució 5:

Equació de la recta de regressió: x = 0,192 y - 0,76

Pes aproximat de una noia de 6 anys: 35,2 kg

Solució 6:

	x_i	y_i	x_i^2	y_i^2	$x_i y_i$
	50	5	2500	25	250
	100	2	10000	4	200
	70	2,5	4900	6,25	170
	60	3,75	3600	14,0625	225
	120	4	14400	16	480
	180	1	32400	1	180
	200	1,25	40000	1,5625	250
	250	0,75	62500	0,5625	187,5
	30	7	900	49	210
	90	3	8100	9	270
å =	1150	30,25	179300	126,4375	2422,5

x =115;   **y**= 3,025%;

$S_x = \sqrt{17930 - 13225} =$ 68,59; $S_y = \sqrt{12,64375 - 9,150625} =$ 1,87 ;

$S_{xy} = 242,25 - (115)(3,025) =$ -105,625

Les rectes de regressió seran

y - 3,025 = -0,022449 (x - 115)
x - 115 = -30,2053 (y - 3,025)

El coeficient de correlació lineal:

$$\mathbf{r} = \frac{-105,625}{(68,59)(1,87)} = -\,0,8235$$

es una correlació inversa alta .

Per l'estimació que nos utilitzarem la recta de regressió de y sobre x.
y= 3,025 - 0,022449(175- 115) = 1,6783 que seria fiable por ser alt el coeficient de correlació.