ANÀLISI DE DADES EN PSICOLOGIA II – PAC 3 – Gemma Pérez Moliner (Aula 1)

SUBJECTE	STAI-E (1) x	HP y	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$	HP estimada $\hat{y} = b + mx_i$	Residu² $(y_i - \hat{y})^2$
1	25	83	-4,15	17,2225	3	9	-12,45	75,7704	52,2674
2	19	69	-10,15	103,0225	-11	121	111,65	69,6553	0,4294
3	25	72	-4,15	17,2225	-8	64	33,2	75,7704	14,2158
4	20	68	-9,15	83,7225	-12	144	109,8	70,6744	7,1527
5	26	75	-3,15	9,9225	-5	25	15,75	76,7896	3,2025
6	28	77	-1,15	1,3225	-3	9	3,45	78,8279	3,3414
7	24	84	-5,15	26,5225	4	16	-20,6	74,7512	85,5404
8	27	71	-2,15	4,6225	-9	81	19,35	77,8088	46,3591
9	29	77	-0,15	0,0225	-3	9	0,45	79,8471	8,1061
10	32	86	2,85	8,1225	6	36	17,1	82,9047	9,5810
11	31	77	1,85	3,4225	-3	9	-5,55	81,8855	23,8681
12	28	80	-1,15	1,3225	0	0	0	78,8279	1,3737
13	33	85	3,85	14,8225	5	25	19,25	83,9239	1,1581
14	31	91	1,85	3,4225	11	121	20,35	81,8855	83,0742
15	32	89	2,85	8,1225	9	81	25,65	82,9047	37,1529
16	37	88	7,85	61,6225	8	64	62,8	88,0006	0,0000
17	29	72	-0,15	0,0225	-8	64	1,2	79,8471	61,5773
18	35	89	5,85	34,2225	9	81	52,65	85,9622	9,2280
19	36	78	6,85	46,9225	-2	4	-13,7	86,9814	80,6660
20	36	89	6,85	46,9225	9	81	61,65	86,9814	4,0747
sumatoris :	583	1600		492,55		1044	502		532,3687

Ens interessa establir un model de predicció de la variable HP, a partir de la variable STAI-E(1). Contesteu a les següents qüestions:

El model ha de predir la variable HP a partir de la variable STAI-E(1), per tant aquesta darrera és la variable explicativa (x); i la variable HP és la variable dependent (y).

càlcul de les mitjanes:

$$\bar{x} = \frac{\sum_i x_i}{n} = \frac{583}{20} = 29{,}15\,; \quad \bar{y} = \frac{\sum_i y_i}{n} = \frac{1600}{20} = 80$$

càlcul de les desviacions std:

$$s_x = \sqrt{\frac{\sum_i (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{492{,}55}{19}} = 5{,}0915\,; \quad s_y = \sqrt{\frac{\sum_i (y_i - \bar{y})^2}{n-1}} = \sqrt{\frac{1044}{19}} = 7{,}4126$$

1.- Calculeu el coeficient de correlació entre les dues variables HP i STAI-E(1).

Calculem el coeficient de correlació (r) per aquestes 20 parelles de valors com:

$$r = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}}$$ **; i usant els valors que hem adjuntat a la taula anterior, el resultat**

és: $$r = \frac{502}{\sqrt{492{,}55 \cdot 1044}} = 0{,}7000$$

2.- Calculeu el pendent i la intersecció de la regressió lineal simple entre HP i STAI-E(1).

Calculem el pendent (m) de la recta de regressió simple com:

$$m = \frac{\sum_i (x_i - \overline{x})(y_i - \overline{y})}{\sum_i (x_i - \overline{x})^2}$$ **; i altre cop usant els valors que hem adjuntat a la taula:**

$$m = \frac{502}{492,55} = 1,0192$$

també : $m = r \cdot \left(\frac{s_y}{s_x} \right) = 0,7 \cdot \left(\frac{7,4126}{5,0915} \right) = 1,0191$ **(la diferència ve de l'arrodoniment de les desviacions standard a 4 decimals).**

Calculem la intersecció (b) com:

$b = \overline{y} - m\overline{x}$; **llavors:** $b = 80 - (1,0192 \cdot 29,15) = 50,290$

3.- Si a partir de la recta de regressió, hem predit a un subjecte una HP igual a 80,86 batecs per minut, quina es la puntuació en STAI-E(1) que presentava?

La recta de regressió que resumeix la relació lineal entre x i y és: $y = mx + b$ **; com que hem calculat abans el pendent (m) i el punt d'intersecció (b). Ara a partir d'una predicció de y (** $\hat{y}$ **) donada, podem trobar la** x_i **que li correspon.**

$x = \frac{\hat{y} - b}{m} = \frac{80,86 - 50,290}{1,0192} = 29,994$ **punts en l'escala STAI-E(1).**

La bondat de l'ajustament és $R^2 = 0,49$ **, el que voldria dir que la recta de regressió s'ajusta als punts en un 49%.**

4.- La relació és significativa? Feu la prova adequada i comenteu-ne el resultat

Volem saber si la relació és significativa o no, és a dir, si "H_1**: la relació és significativa" o si "**H_0**: la relació no és significativa".**
Sabem que en la recta de regressió, si el pendent fos 0, y seria una constant (y = b), i per tant, no hi hauria relació entre y i x. Llavors, les Hipòtesi quedaran així:

$H_0 : \beta_1 = 0$; **on** β_1 **és el pendent del model de la població.**

$H_1 : \beta_1 > 0$;

L'estadístic de contrast és el pendent estimat ($\hat{\beta}_1$ **, que abans hem calculat com a 1,0192) dividit per l'error estàndard del pendent.**

L'error estàndard del pendent és $s_{\hat{\beta}_1} = \dfrac{s}{\sqrt{\sum_{i=1}^{n}(x_1 - \overline{x})^2}}$ **; on s és la desviació estàndard comuna**

així: $s = \sqrt{\dfrac{\sum (y_1 - \hat{y})^2}{n-2}} = \sqrt{\dfrac{532{,}3687}{18}} = 5{,}4384$ **; i llavors** $s_{\hat{\beta}_1} = \dfrac{5{,}4384}{\sqrt{492{,}55}} = 0{,}2450$

Estadístic de contrast : 1,0192 / 0,245 = 4,16

Utilitzem la distribució t amb $\alpha/2 = 0{,}025$ **i 18 graus de llibertat:** $t_{(0{,}025;18)} = 2{,}1009$

Veiem que 2,1009 < 4,16; per tant concluïm que la relació és significativa.

Això ho podem veure també trobant l'interval de confiança del pendent, si l'interval no conté el 0, és a dir si el pendent no pot ser 0 (amb un nivell de confiança), la relació serà significativa:

$\beta_1 = \hat{\beta}_1 \pm t_{(0{,}025;18)} \cdot s_{\hat{\beta}_1} = 1{,}0192 \pm 2{,}1009 \cdot 0{,}245 \Rightarrow (1{,}5339;0{,}5045)$ **l'interval no conté 0 per tant la relació és significativa.**