Problemas resueltos. Tema 13

1- En una tarea de clasificación de patrones que constaba de 10 láminas se obtuvieron los siguientes datos de las diferencias de las distancias logarítmicas del estímulo a clasificar con respecto a los prototipos de las dos clases en que podía ser encuadrado y del número de errores cometidos por los sujetos:

Lámina	1	2	3	4	5	6	7	8	9	10
Diferencia	0,71	0,67	1,98	1,61	0,67	1,48	0,25	1,44	1,06	0,95
Nº errores	12	10	4	2	6	5	16	3	4	8

a) Calcule el coeficiente de correlación de Pearson e interprete el resultado.
b) Determine la recta de regresión que permite predecir el número de errores en función de la diferencia entre las distancias.
c) De acuerdo con el modelo anterior, indique cual tiene que ser la diferencia para que no haya errores.

Solución:

a) Comenzamos calculando media, varianza y desviación típica de ambas variables

$$\overline{Dl} = 1{,}082 \quad S_{Dl}^2 = 0{,}256 \quad S_{Dl} = 0{,}506$$
$$\overline{E} = 7 \quad S_E^2 = 18 \quad S_E = 4{,}243$$

Calculamos la covarianza

$$S_{EDl} = \frac{\sum E \cdot Dl}{n} - \overline{E} \cdot \overline{Dl} = 5{,}794 - 7{,}574 = -1{,}78$$

El coeficiente de correlación será:

$$r = \frac{S_{EDl}}{S_E \cdot S_{Dl}} = \frac{-1{,}78}{4{,}243 \cdot 0{,}506} = -0{,}829$$

El signo negativo del coeficiente de correlación nos indica que la relación entre ambas variables es inversa, es decir que al aumentar la distancia disminuye el número de errores. El valor absoluto nos indica que la relación lineal entre distancia y número de errores es bastante alta, por consiguiente las variaciones en el número de errores en esta tarea se pueden explicar y predecir en gran medida, por la diferencia de las distancias de los estímulos a clasificar.

b) Para determinar la recta de regresión E = a·Dl + b calculamos los valores de los coeficientes a y b mediante las expresiones obtenidas por el método de mínimos cuadrados:

$$a = \frac{S_{EDl}}{S_{Dl}^2} = \frac{-1{,}78}{0{,}256} = -6{,}953$$

$$b \quad = \quad \overline{E} - a\overline{D}l \quad = \quad 7 + 6{,}953 \cdot 1{,}082 \quad = \quad 14{,}523$$

Luego la recta será E = -6,953·Dl + 14,253

c) Haciendo cero el número de errores en la expresión anterior tenemos:
0 = -6,953·Dl + 14,503, despejando obtenemos Dl = -14,503/-6,953 = 2,089
Téngase en cuenta que no se trata de predecir la distancia en función del número de errores, sino de buscar en que punto corta la recta de regresión el eje de abscisas. Es decir para que valor de la distancia se hace cero E.

2- En el mismo trabajo del problema anterior, se calculó también la diferencia de las distancias euclídeas del patrón a clasificar con respecto a los prototipos de ambas clases, obteniéndose el siguiente resultado:

Lámina	1	2	3	4	5	6	7	8	9	10
Diferencia	9,98	9,97	9,93	9,92	9,99	9,99	9,93	9,93	9,97	8,00

Indique que distancia le parece mas adecuada para expresar la dificultad de la tarea y porqué.

Solución:

Calculamos la media, varianza y desviación típica de las diferencias de distancias euclídeas:

$$\overline{D}e = 9{,}761 \quad S_{De}^2 = 0{,}345 \quad S_{De} = 0{,}588$$

Calculamos la covarianza entre el número de errores y esta distancia

$$S_{EDe} \quad = \quad \frac{\sum E \cdot De}{n} - \overline{E} \cdot \overline{D}e \quad = \quad 68{,}146 - 68{,}327 \quad = \quad -0{,}181$$

Por consiguiente el coeficiente de correlación de Pearson entre el número de errores y la diferencia de las distancias euclídeas valdrá:

$$r \quad = \quad \frac{S_{EDe}}{S_E \cdot S_{De}} \quad = \quad \frac{-0{,}181}{4{,}243 \cdot 0{,}588} \quad = \quad -0{,}073$$

Comparando ambos coeficientes de correlación se ve que la diferencia de distancias logarítmicas explica mucho mejor el número de errores que la diferencia de distancias euclídeas.

3- Se ha medido la motivación ante el estudio a 38 sujetos, antes y después de participar en un programa de innovación didáctica. Obteniéndose los siguientes datos:

Sujeto	1	2	3	4	5	6	7	8	9	10	11	12	13
Pre-prueba	55	49	37	40	50	45	35	38	28	56	41	44	44
Post-prueba	65	53	57	51	66	60	51	53	62	57	58	51	48

Sujeto	14	15	16	17	18	19	20	21	22	23	24	25	26
Pre-prueba	38	56	58	38	46	57	45	58	57	62	63	46	60
Post-prueba	48	48	64	67	48	61	59	69	64	69	62	60	61

Sujeto	27	28	29	30	31	32	33	34	35	36	37	38
Pre-prueba	54	56	66	48	46	60	55	57	51	43	56	52
Post-prueba	68	61	55	56	58	65	68	58	49	66	62	65

a) Calcule la recta de regresión que permite predecir los resultados de la post-prueba en función de las puntuaciones en la pre-prueba y descomponga la varianza total de las puntuaciones en la post-prueba en dos componentes.
b) Calcule para cada individuo la diferencia entre el valor pronosticado por la recta obtenida en el apartado a y el valor observado.
c) Calcule la media y varianza de las diferencias obtenidas en el apartado anterior y compare los resultados con lo obtenido en el apartado a.
d) Calcule la recta de regresión que permite deducir las puntuaciones en la pre-prueba a partir del resultado obtenido por los sujetos en la post-prueba.

Solución:

a) Calculamos los estadísticos básicos de ambas puntuaciones:

$$\overline{Pr} = 49{,}737 \quad S^2_{Pr} = 78{,}194 \quad S_{Pr} = 8{,}843$$
$$\overline{Po} = 59{,}026 \quad S^2_{Po} = 42{,}289 \quad S_{Po} = 6{,}503$$

$$S_{Pr\,Po} = \frac{\sum Pr \cdot Po}{n} - \overline{Pr} \cdot \overline{Po} = 2956{,}71 - 2935{,}776 = 20{,}934$$

Por consiguiente los coeficientes de la recta Po = a·Pr + b valdrán:

$$a = \frac{20{,}934}{78{,}194} = 0{,}268 \quad b = 59{,}026 - 0{,}268 \cdot 49{,}737 = 45{,}696$$

Por consiguiente la recta de regresión pedida, tendrá de ecuación:

$$Po = 0{,}268 \cdot Pr + 45{,}696$$

y la varianza de las predicciones será:

$$S^2_{\hat{P}o} \quad = \quad a^2 \cdot S^2_{\mathrm{Pr}} \quad = \quad 0{,}268^2 \cdot 78{,}194 \quad = \quad 5{,}616$$

Por consiguiente, la varianza de los residuos será:

$$S^2_r \quad = \quad S^2_{Po} - S^2_{\hat{P}o} \quad = \quad 42{,}289 - 5{,}616 \quad = \quad 36{,}673$$

b) Sustituyendo las puntuaciones Pre en la recta de regresión, obtenida en el apartado anterior, calculamos las predicciones para las puntuaciones Post. La diferencia entre la puntuación Post y esta predicción constituye el residuo. Repitiendo esta operación para todos los sujetos obtenemos los siguientes datos:

Sujeto	1	2	3	4	5	6	7	8
Predicción	60,436	58,828	55,612	56,416	59,096	57,756	55,076	55,88
Residuo	4,564	-5,828	1,388	-5,416	6,904	2,244	-4,076	-2,88

Sujeto	9	10	11	12	13	14	15	16
Predicción	53,2	60,704	56,684	57,488	57,488	55,88	60,704	61,24
Residuo	8,8	-3,704	1,316	-6,488	-9,488	-7,88	-12,704	2,76

Sujeto	17	18	19	20	21	22	23	24
Predicción	55,88	58,024	60,972	57,756	61,24	60,972	62,312	62,58
Residuo	11,12	-10,024	0,028	1,244	7,76	3,028	6,688	-0,58

Sujeto	25	26	27	28	29	30	31	32
Predicción	58,024	61,776	60,168	60,704	63,384	58,56	58,024	61,776
Residuo	1,976	-0,776	7,832	0,296	-8,384	-2,56	-0,024	3,224

Sujeto	33	34	35	36	37	38
Predicción	60,436	60,972	59,364	57,22	60,704	59,632
Residuo	7,564	-2,972	-10,364	8,78	1.296	5,368

c) Calculando la media y varianza de los residuos obtenemos que la media es 0,0008 y la varianza 36,687. Como vemos la media de los residuos es prácticamente nula y la varianza coincide con la que habíamos deducido teóricamente.

d) Tenemos que calcular los coeficientes de la ecuación Pre = a'·Post + b que serán:

$$a' \quad = \quad \frac{S_{Pr\,Po}}{S^2_{Po}} \quad = \quad \frac{20{,}934}{42{,}289} \quad = \quad 0{,}495$$

$$b' \quad = \quad \overline{Pr} - a' \cdot \overline{Po} \quad = \quad 49{,}737 - 0{,}495 \cdot 59{,}026 \quad = \quad 20{,}519$$

Con lo cual la ecuación resulta ser Pre = 0,495·Post + 20,519