

Curso : Probabilidad y Estadística
Sigla : EYP1113-1
Pauta : Examen
Profesor : Ricardo Olea
Ayudante : Claudia Ortega.

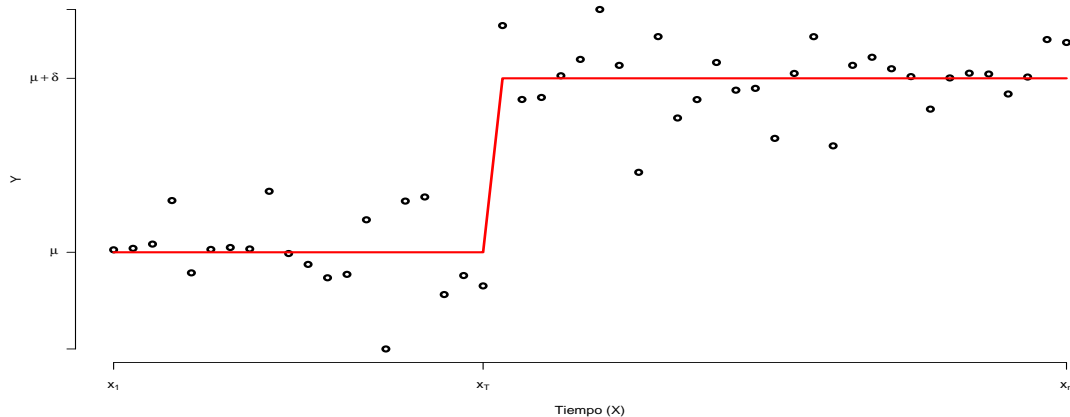
1. (30 %) Un investigador mide n observaciones secuencialmente en el tiempo. Al observar el comportamiento de las observaciones Y con respecto al tiempo X se da cuenta que a partir de la T -ésima observación los datos presentan un cambio brusco de media, pero una ausencia de pendiente. Para modelar este fenómeno el investigador propone lo siguiente:

$$E(Y | X = x) = \mu + \delta g(x),$$

con

$$g(x) = \begin{cases} 0, & \text{si } x \leq x_T \\ 1, & \text{si } x > x_T \end{cases}$$

Asumiendo varianza constante, el resultado del modelo es el siguiente:



Para los valores $(x_1, y_1), \dots, (x_T, y_T), \dots, (x_n, y_n)$ determine una expresión de los estimadores de mínimos cuadrados para los parámetros (μ, δ) .

Solución

Matricialmente los estimadores de mínimos cuadrados de $\beta = (\mu, \delta)'$ están dados por:

$$\hat{\beta} = \begin{bmatrix} \hat{\mu} \\ \hat{\delta} \end{bmatrix} = [X^T X]^{-1} X^T Y \quad \text{[2 \%]}$$

donde

$$Y = \begin{bmatrix} y_1 \\ \vdots \\ y_T \\ y_{T+1} \\ \vdots \\ y_n \end{bmatrix}, \quad X = \begin{bmatrix} 1 & g(x_1) \\ \vdots & \vdots \\ 1 & g(x_T) \\ 1 & g(x_{T+1}) \\ \vdots & \vdots \\ 1 & g(x_n) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & 1 \end{bmatrix} \quad [2\%]$$

Luego, tenemos que

$$\begin{aligned} X^T X &= \begin{bmatrix} n & (n-T) \\ (n-T) & (n-T) \end{bmatrix} \Rightarrow [X^T X]^{-1} = \frac{1}{T(n-T)} \begin{bmatrix} n-T & -(n-T) \\ -(n-T) & n \end{bmatrix} \\ &= \frac{1}{T} \begin{bmatrix} 1 & -1 \\ -1 & n/(n-T) \end{bmatrix} \quad [3\%] \end{aligned}$$

$$X^T Y = \begin{bmatrix} \sum_{i=1}^n y_i \\ \sum_{i=T+1}^n y_i \end{bmatrix} \quad [3\%]$$

Por lo tanto

$$\begin{aligned} \hat{\beta} &= \begin{bmatrix} \hat{\mu} \\ \hat{\delta} \end{bmatrix} = \frac{1}{T} \begin{bmatrix} 1 & -1 \\ -1 & n/(n-T) \end{bmatrix} \cdot \begin{bmatrix} \sum_{i=1}^n y_i \\ \sum_{i=T+1}^n y_i \end{bmatrix} \\ &= \frac{1}{T} \begin{bmatrix} \sum_{i=1}^n y_i & -\sum_{i=T+1}^n y_i \\ -\sum_{i=1}^n y_i & +\frac{n}{n-T} \sum_{i=T+1}^n y_i \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{T} \sum_{i=1}^T y_i \\ \frac{1}{n-T} \sum_{i=T+1}^n y_i - \frac{1}{T} \sum_{i=1}^T y_i \end{bmatrix} \quad [10\%] \\ &\quad [10\%] \end{aligned}$$

2. (30 %) A usted se le pide ajustar para n observaciones el siguiente modelo de regresión lineal:

$$y' = \alpha + \beta x$$

Después de obtener los coeficientes se da cuenta que debía cambiar la escala de la variable X previamente. Si la variable que debía utilizar realmente era $Z = cX$, con c una constante, entonces el modelo correcto está dado por:

$$y' = \alpha^* + \beta^* z$$

- (a) (15 %) ¿Se pueden obtener $\hat{\alpha}^*$ y $\hat{\beta}^*$ a partir de los estimadores de mínimos cuadrados $\hat{\alpha}$ y $\hat{\beta}$?
 (b) (15 %) Compare los coeficientes de determinación r^2 de ambos modelos.

Solución

- (a) Para el modelo: $y' = \alpha + \beta x$, se tiene que

$$\hat{\alpha} = \bar{y} - \hat{\beta} \bar{x}$$

$$\hat{\beta} = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sum_{i=1}^n x_i^2 - n \bar{x}^2}$$

Para el modelo: $y' = \alpha^* + \beta^* z$, con $z = cx$, se tiene que

$$\hat{\beta}^* = \frac{\sum_{i=1}^n z_i y_i - n \bar{z} \bar{y}}{\sum_{i=1}^n z_i^2 - n \bar{z}^2} = \frac{c \cdot \left[\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y} \right]}{c^2 \cdot \left[\sum_{i=1}^n x_i^2 - n \bar{x}^2 \right]} = \frac{1}{c} \hat{\beta} \quad [10 \%]$$

$$\hat{\alpha}^* = \bar{y} - \hat{\beta}^* \bar{z} = \bar{y} - \frac{1}{c} \hat{\beta} c \bar{x} = \bar{y} - \hat{\beta} \bar{x} = \hat{\alpha} \quad [5 \%]$$

- (b) Para el modelo 1: Y vs X , el coeficiente de determinación está dado por

$$r_1^2 = 1 - \frac{s_{Y|x}^2}{s_Y^2}$$

y que

$$\hat{\rho}^2 = 1 - \frac{(n-2)}{(n-1)} \frac{s_{Y|x}^2}{s_Y^2} \quad (1)$$

$$= \hat{\beta}^2 \frac{s_X^2}{s_Y^2} \quad (2)$$

Igualando (1) con (2) se obtiene que

$$\frac{s_{Y|x}^2}{s_Y^2} = \frac{(n-1)}{(n-2)} \left[1 - \hat{\beta}^2 \frac{s_X^2}{s_Y^2} \right] \Rightarrow r_1^2 = 1 - \frac{(n-1)}{(n-2)} \left[1 - \hat{\beta}^2 \frac{s_X^2}{s_Y^2} \right] \quad [5 \%]$$

Análogamente para el modelo 2: Y vs Z su coeficiente de determinación es

$$r_2^2 = 1 - \frac{(n-1)}{(n-2)} \left[1 - (\hat{\beta}^*)^2 \frac{s_Z^2}{s_Y^2} \right] \quad [2 \%]$$

Como

$$s_Z^2 = \sum_{i=1}^n z_i^2 - n \bar{z}^2 = \sum_{i=1}^n c^2 x_i^2 - n c^2 \bar{z}^2 = c^2 s_X^2 \quad [2\%]$$

y

$$\hat{\beta}^* = \frac{1}{c} \hat{\beta}$$

Se tiene que

$$r_2^2 = 1 - \frac{(n-1)}{(n-2)} \left[1 - \frac{1}{c^2} \hat{\beta}^2 \frac{c^2 s_X^2}{s_Y^2} \right] = r_1^2 \quad [6\%]$$

Por lo tanto, los coeficientes de determinación de los dos modelos son iguales.

3. (40 %) Varios vigas de madera se han sometido a diversas cargas P experimentalmente para determinar su deflexión D . Sobre los siguientes resultados

P (Tonelada)	D (cm)
8.4	4.8
6.7	2.9
4.0	2.0
10.2	5.5

- (a) (20 %) Determinar la recta de regresión lineal de la deflexión por carga, la desviación estándar condicional (Asuma que es constante), el coeficiente de determinación y de correlación.
- (b) (10 %) Obtenga el correspondiente intervalo de confianza al 90 % para la ecuación de regresión obtenida en (a).
- (c) (10 %) ¿Cuál sería la deflexión media de la viga bajo una carga $P = 8$ toneladas? Asumiendo normalidad, determine el percentil 75 % de deflexión para esta carga.

Solución

- (a) Tenemos que

$$\sum_{i=1}^4 p_i = 29.3, \quad \sum_{i=1}^4 d_i = 15.2, \quad \sum_{i=1}^4 p_i d_i = 123.85, \quad \sum_{i=1}^4 p_i^2 = 235.49, \quad \sum_{i=1}^4 d_i^2 = 65.7$$

Luego los coeficientes de la recta de regresión D vs P están dados por:

$$\hat{\beta} = \frac{123.85 - 4 \cdot \left(\frac{29.3}{4}\right) \cdot \left(\frac{15.2}{4}\right)}{235.49 - 4 \cdot \left(\frac{29.3}{4}\right)^2} = 0.5994968 \quad [4\%]$$

$$\hat{\alpha} = \left(\frac{15.2}{4}\right) - 0.5994968 \cdot \left(\frac{29.3}{4}\right) = -0.5913142 \quad [4\%]$$

Por lo tanto

$$d' = -0.5913142 + 0.5994968 \cdot p$$

La varianza condicional se obtiene como

$$\begin{aligned} s_{D|p}^2 &= \frac{1}{4-2} \sum_{i=1}^4 (d_i - d'_i)^2 = \frac{1}{4-2} \left[\sum_{i=1}^4 (d_i - \bar{d})^2 - \hat{\beta}^2 \sum_{i=1}^4 (p_i - \bar{p})^2 \right] \\ &= \frac{1}{4-2} \left[\sum_{i=1}^4 d_i^2 - 4 \cdot \bar{d}^2 - \hat{\beta}^2 \left(\sum_{i=1}^4 p_i^2 - 4 \cdot \bar{p}^2 \right) \right] \\ &= \frac{1}{2} \left[65.7 - 4 \cdot \left(\frac{15.2}{4}\right)^2 - (0.5994968)^2 \left(235.49 - 4 \cdot \left(\frac{29.3}{4}\right)^2 \right) \right] \\ &= 0.2201474 \\ s_{D|p} &= 0.4691986 \quad [4\%] \end{aligned}$$

Para el coeficiente de determinación y de correlación necesitamos previamente calcular

$$s_D^2 = \frac{1}{4-1} \sum_{i=1}^4 (d_i - \bar{d})^2 = \frac{1}{4-2} \left[\sum_{i=1}^4 d_i^2 - 4 \cdot \bar{d}^2 \right] = \frac{7.94}{3} = 2.646667 \quad [2\%]$$

Luego

$$r^2 = 1 - \frac{s_{D|p}^2}{s_D^2} = 1 - \frac{0.2201474}{2.646667} = 0.9168209 \quad [3\%]$$

$$\hat{\rho}^2 = 1 - \frac{4-2}{4-1} \frac{s_{D|p}^2}{s_D^2} = 1 - \frac{2}{3} \cdot \frac{0.2201474}{2.646667} = 0.9445473 \Rightarrow \hat{\rho} = 0.9718782 \quad [3\%]$$

(b) La banda de confianza esta dada por

$$\langle \mu_{D|p_i} \rangle_{1-\alpha} = \bar{d}_i \pm t_{(1-\alpha/2), n-2} \cdot s_{D|p} \sqrt{\frac{1}{n} + \frac{(p_i - \bar{p})^2}{\sum_{j=1}^n (p_j - \bar{p})^2}},$$

donde

$$n = 4$$

$$s_{D|p} = 0.4691986$$

$$\bar{d}_i = -0.5913142 + 0.5994968 \cdot p_i$$

$$\bar{p} = 7.325$$

$$\sum_{j=1}^4 (p_j - \bar{p})^2 = \sum_{j=1}^4 p_j^2 - 4 \cdot \bar{p}^2 = 20.8675$$

$$t_{0.95, 2} = 2.920$$

Reemplazando para $p_i = 4.0, 6.7, 8.4$ y 10.2 se tiene que

p_i	$\mu_{D p_i}$	LI	LS
4,0	1,807	0,597	3,017
6,7	3,425	2,715	4,136
8,4	4,444	3,687	5,202
10,2	5,524	4,422	6,625

[10 %]

(c) Para $P = 8$, la deflexión esperada de la viga es

$$E(D | P = 8) = -0.5913142 + 0.5994968 \cdot 8 = 4.20466 \quad [3\%]$$

Bajo Normalidad, se tiene que

$$D | P = p \sim \text{Normal}(-0.5913142 + 0.5994968 \cdot p, 0.4691986) \quad [3\%]$$

Por lo tanto, el percentil 75 % de la deflexión D para $P = 8$ esta dado por:

$$\begin{aligned} d_{0.75} &= \mu_{D|p=8} + s_{D|p} \cdot \Phi^{-1}(0.75) \quad [2\%] \\ &\approx 4.20466 + 0.4691986 \cdot 0.67 \\ &= 4.519023 \quad [2\%] \end{aligned}$$

Tablas de Percentiles p

Distribución Normal Estándar											Distribución t-student $t_p(\nu)$				
Z_p	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	ν	$t_{0.90}$	$t_{0.95}$	$t_{0.975}$	$t_{0.99}$
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359	1	3.078	6.314	12.706	31.821
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753	2	1.886	2.920	4.303	6.965
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141	3	1.638	2.353	3.182	4.541
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517	4	1.533	2.132	2.776	3.747
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879	5	1.476	2.015	2.571	3.365
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224	6	1.440	1.943	2.447	3.143
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549	7	1.415	1.895	2.365	2.998
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852	8	1.397	1.860	2.306	2.896
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133	9	1.383	1.833	2.262	2.821
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389	10	1.372	1.812	2.228	2.764
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621	11	1.363	1.796	2.201	2.718
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830	12	1.356	1.782	2.179	2.681
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015	13	1.350	1.771	2.160	2.650
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177	14	1.345	1.761	2.145	2.624
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319	15	1.341	1.753	2.131	2.602
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441	16	1.337	1.746	2.120	2.583
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545	17	1.333	1.740	2.110	2.567
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633	18	1.330	1.734	2.101	2.552
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706	19	1.328	1.729	2.093	2.539
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767	20	1.325	1.725	2.086	2.528
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817	21	1.323	1.721	2.080	2.518
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857	22	1.321	1.717	2.074	2.508
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890	23	1.319	1.714	2.069	2.500
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916	24	1.318	1.711	2.064	2.492
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936	25	1.316	1.708	2.060	2.485
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952	26	1.315	1.706	2.056	2.479
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964	27	1.314	1.703	2.052	2.473
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974	28	1.313	1.701	2.048	2.467
2.8	0.9974	0.9975	0.9976	0.9977	0.9978	0.9979	0.9980	0.9981	0.9982	0.9983	29	1.311	1.699	2.045	2.462
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986	30	1.310	1.697	2.042	2.457
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990	∞	1.282	1.645	1.960	2.326

Formulario:

- Para el modelo de regresión lineal simple $Y = \alpha + \beta x$ se tiene que

$$\hat{\alpha} = \bar{y} - \hat{\beta} \bar{x}, \quad \hat{\beta} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}, \quad r^2 = 1 - \frac{s_{Y|x}^2}{s_Y^2}, \quad s_Y^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2$$

$$s_{Y|x}^2 = \frac{1}{n-2} \sum_{i=1}^n (y_i - y'_i)^2, \quad \hat{\rho} = \hat{\beta} \frac{s_X}{s_Y} = \sqrt{1 - \frac{(n-2)}{(n-1)} \frac{s_{Y|x}^2}{s_Y^2}}, \quad s_X^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

- Banda de Confianza:

$$\langle \mu_{Y|x_i} \rangle_{1-\alpha} = \bar{y}_i \pm t_{(1-\alpha/2), n-2} \cdot s_{Y|x} \sqrt{\frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum_{j=1}^n (x_j - \bar{x})^2}}, \quad \bar{y}_i = \hat{\alpha} + \hat{\beta} \cdot x_i$$