

## Modelo Lineal

### PRACTICA 3

1. Con los datos del ejercicio 20 de la práctica 1,
  - a) calcule, para cada punto del diseño, el intervalo de confianza de nivel 0.95 para la respuesta.
  - b) calcule, para cada punto del diseño, el intervalo de predicción de nivel 0.95 para la respuesta.
  - c) realice un scatterplot en el que se representen los pares de puntos  $(x,y)$ , la recta de mínimos cuadrados y los límites de los intervalos obtenidos en a) y b) para cada punto del diseño.
  - d) calcule, para cada punto del diseño, el intervalo de confianza para la respuesta de manera que el nivel global de los 25 intervalos obtenidos sea 0.95.
  - e) grafique un scatterplot en el que se representen los pares de puntos  $(x,y)$ , la recta de mínimos cuadrados y las curvas que se obtendrían si se unieran los límites superiores por un lado y los inferiores por otro, de los intervalos de confianza computados en d). Superponga la banda de confianza de nivel total 0.95 para la recta ajustada. ¿Cómo se interpretan estas curvas?

2. Los datos en la Tabla 1 (archivo thanks.txt) corresponden al peso en libras ( $Y$ ) y la edad en semanas ( $X$ ) de 13 pavos de Thanksgiving. De estos pavos 4 provenían del estado de Georgia (G), 4 de Virginia (V) y 5 de Wisconsin (W).

¿Será posible relacionar a  $X$  e  $Y$  mediante un modelo lineal simple sin tener en cuenta el lugar de origen?

- a) Suponga el modelo  $Y = \beta_0 + \beta_1 X + \epsilon$ . Grafique los residuos identificando el lugar de origen. ¿Le parece razonable suponer que la región no afecta al peso de los pavos?
- b) Suponiendo que el lugar de origen sólo afecta el nivel, incorpore al modelo una ó más variables dummies que permitan tener en cuenta los diferentes niveles de la respuesta.
- c) Estime los parámetros del modelo planteado en b) y testeé la hipótesis de que el lugar de origen no afecta el peso de los pavos.

### 3. Test de Paralelismo.

Supongamos que queremos comparar  $k$  rectas de regresión dadas por

$$Y = \alpha_i + \beta_i x + \epsilon \quad i = 1, \dots, k,$$

donde  $E(\epsilon) = 0$  y  $Var(\epsilon) = \sigma^2$ . Para ello tomamos  $n_i$  pares  $(x_{ij}, y_{ij})$ ,  $j = 1, \dots, n_i$  correspondientes a la  $i$ -ésima recta,  $i = 1, \dots, k$ , de manera que

$$Y_{ij} = \alpha_i + \beta_i x_{ij} + \epsilon_{ij}$$

donde los  $\epsilon_{ij}$  son independientes y con distribución  $N(0, \sigma^2)$ .

- a) Encuentre una expresión matricial adecuada para plantear este problema.
- b) Halle los estimadores de mínimos cuadrados de los parámetros.
- c) Supongamos que se desea testear la hipótesis de que las  $k$  rectas son paralelas. Exprese las hipótesis nula y alternativa para este problema y deduzca un test de nivel  $\alpha$  para decidir entre  $H_0$  y  $H_1$ .
- d) Si al realizar el test planteado en c) se rechazara la hipótesis de que las rectas son paralelas tendría sentido tratar de identificar aquellos  $\beta_i$  que son diferentes. ¿Intervalos de confianza para qué combinación de los parámetros serían adecuados para detectar los  $\beta_i$  que difieren? ¿Cuántos intervalos debe plantear? Deduzca los intervalos de confianza de manera tal que tengan un nivel global  $1 - \alpha$ . ¿Qué posibilidades tiene?

#### 4. Test de Coincidencia.

En las mismas condiciones que en el ejercicio anterior suponga que se desea testear que las  $k$  rectas son iguales, es decir que todos los  $\alpha_i$  coinciden y que todos los  $\beta_i$  también coinciden. Exprese las hipótesis nula y alternativa para este problema y deduzca un test de nivel  $\alpha$  para decidir entre  $H_0$  y  $H_1$ .

5. Los datos que se presentan en la Tabla 2 (archivo *catedral.txt*) corresponden a altura y longitud, en pies, de catedrales inglesas medievales. Las catedrales están clasificadas según su estilo arquitectónico: románico (1) o gótico (2). Analizando estos datos, decida si la relación entre la longitud (variable respuesta) y la altura es la misma para ambos estilos. Si no lo es, describa las diferencias.

#### 6. Distancia horizontal entre dos rectas paralelas con pendiente distinta de cero.

Dadas dos rectas paralelas llamemos  $\delta$  a la distancia horizontal entre ellas, con signo. El objetivo de este ejercicio es deducir un estimador de  $\delta$  y un test de hipótesis para testear la hipótesis  $H_0 : \delta = \delta_0$ .

- a) Dadas dos rectas paralelas calcule  $\delta$ . ¿Qué interpretación puede darse al signo de la distancia?
- b) Supongamos que tenemos dos muestras de tamaño  $n_i$  formadas por los pares  $(x_{ij}, y_{ij})$ ,  $j = 1, \dots, n_i$ ,  $i = 1, 2$ , para quienes

$$Y_{ij} = \alpha_i + \beta x_{ij} + \epsilon_{ij} \quad j = 1, \dots, n_i,$$

donde los  $\epsilon_{ij}$  son independientes y con distribución  $N(0, \sigma^2)$ . Deduzca los estimadores de mínimos cuadrados de  $\alpha_1$ ,  $\alpha_2$  y de  $\beta$ . A partir de ellos proponga un estimador de  $\delta$ .

c) Considere el estadístico  $U = \hat{\alpha}_1 - \hat{\alpha}_2 - \delta\hat{\beta}$ . Pruebe que  $E(U) = 0$  y que

$$Var(U) = \sigma^2 \left\{ \frac{1}{n_1} + \frac{1}{n_2} + \frac{(\bar{x}_1 - \bar{x}_2 - \delta)^2}{\sum \sum (x_{ij} - \bar{x}_i)^2} \right\}.$$

Dado que los errores son normales, ¿qué distribución tiene  $U$  ?

(Hint: pruebe que la  $cov(\bar{Y}_i, Y_{ij} - \bar{Y}_i) = 0$ .)

- d) A partir de c) deduzca un test de nivel  $\alpha$  para testear la hipótesis  $H_o : \delta = \delta_o$ .
- e) En la Tabla 3 (archivo distpar.txt) se presenta un conjunto de datos simulados correspondiente al modelo de dos rectas paralelas:

$$Y_{ij} = \alpha_i + \beta x_{ij} + \epsilon_{ij} \quad i = 1, 2 \quad j = 1, \dots, 20$$

- i) halle los estimadores de mínimos cuadrados de los parámetros.
- ii) Sea  $\delta$  la distancia horizontal entre las rectas, testee la hipótesis  $H_o : \delta = -5$ .

### 7. Comparación de la media de $k$ poblaciones

Supongamos que queremos comparar la media de  $k$  poblaciones, para lo cual se toman muestras aleatorias independientes entre sí de tamaño  $J$  de cada una de las poblaciones. Sea  $Y_{ij}$  la  $j$ -ésima observación de la  $i$ -ésima población,  $i = 1, \dots, k$ ,  $j = 1, \dots, J$  y supongamos que  $Y_{ij} \sim N(\mu_i, \sigma^2)$ .

- a) Supongamos que se plantea el modelo

$$Y_{ij} = \mu_i + \epsilon_{ij},$$

con  $\epsilon_{ij} \sim N(0, \sigma^2)$ .

Deduzca los estimadores de mínimos cuadrados de los parámetros y un test de nivel  $\alpha$  para la hipótesis de que las medias poblacionales son iguales.

Deduzca intervalos de confianza para  $\mu_i - \mu_j$ ,  $1 \leq i < j \leq k$  de nivel global  $1 - \alpha$ .

- b) Si se plantea el modelo

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij},$$

con  $\epsilon_{ij} \sim N(0, \sigma^2)$ .

¿Cuál es la matriz de diseño? ¿Qué rango tiene? Dé un ejemplo de una función paramétrica que no sea estimable cuando se utiliza esta parametrización.

Si suponemos que  $\sum_{i=1}^k \alpha_i = 0$ , ¿cómo se interpreta esta restricción? Halle los estimadores de mínimos cuadrados de los parámetros bajo esta restricción.

**8.** (archivo cafeina.txt) Se dice que la cafeína ingerida oralmente es un estimulante. Con el fin de tener alguna idea sobre el efecto físico del consumo de cafeína se realizó el siguiente experimento. Se usaron tres niveles de consumo de cafeína: 0, 100 y 200 mg. y se entrenaron en digitación 30 hombres jóvenes de aproximadamente la misma edad y habilidad física. Una vez que el entrenamiento se completó, 10 hombres fueron asignados aleatoriamente a cada nivel de consumo de cafeína. Ni los evaluadores ni los jóvenes conocían la cantidad de cafeína consumida. Dos horas después de la administración del tratamiento, se requirió a cada uno de los jóvenes un ejercicio de digitación. En la Tabla 4 se muestra el número de digitaciones por minuto de cada uno de los individuos.

- a) Testee la hipótesis de que la cafeína no afecta la digitación al nivel 0.05.
- b) Deduzca intervalos de confianza para la diferencia de las medias con un nivel global 0.95. Interprete los resultados.

**9.** Dado el modelo  $Y_i = \beta_o + \beta_1 x_{i1} + \dots + \beta_{p-1} x_{i,p-1} + \epsilon_i, i = 1, \dots, n$ , pruebe que el estadístico del test de F para testear  $H_o : \beta_1 = \dots = \beta_{p-1} = 0$  puede escribirse como

$$\frac{R^2}{1 - R^2} \frac{(n - p)}{(p - 1)}.$$

**10.** *Consumo de combustible*

Los datos de la Tabla 5 (archivo fuel.txt) corresponden a los 48 estados de EE.UU y se describen a continuación:

TAX: tasa de impuesto al combustible en 1972 (ciento por galón)

DLIC: porcentaje de la población con licencia de conductor en 1971

FUEL: consumo de combustible en 1972 (galones por persona)

- a) Ajuste un modelo lineal para la variable FUEL usando un intercept y como variables predictoras a TAX y DLIC.
- b) Usando el método de Bonferroni realice intervalos de confianza para los coeficientes de DLIC y TAX con un nivel total igual a 0.95 para el modelo ajustado en a).
- c) Calcule el elipsoide de confianza de nivel 0.95 para los coeficientes de DLIC y TAX, ignorando el valor del intercept para el modelo ajustado en a).
- d) (*Opcional*) Realice un gráfico en el que se representen simultáneamente los intervalos hallados en b) y el elipsoide calculado en c). Observe las diferencias.

## 11. Biomasa

En la Tabla 6 (archivo biomasa.txt) se muestran los datos correspondientes al ejemplo de Biomasa presentado en la clase teórica.

- a) Recordando que la variable dependiente es BIO (biomasa), realice las sumas secuenciales entrando a las variables en el siguiente orden: intercept, K, PH, SAL, Zn y SODIO. ¿Qué test está realizando en cada paso? ¿Cuál es el modelo que usa en el denominador de cada test de F?

Usando este orden, ¿cuáles son las variables que usaría como predictoras?

- b) Repita a) usando el orden: intercept, SODIO, PH, SAL, Zn y K.

¿Llega a las mismas conclusiones?

- c) Realice un scatterplot múltiple (usando la instrucción pairs). A partir de este gráfico justifique la diferencia entre las conclusiones de a) y b).

**Tabla 1.** Pavos

| <i>obs</i> | <i>Y</i> | <i>X</i> | <i>origen</i> |
|------------|----------|----------|---------------|
| 1          | 28       | 13.3     | <i>G</i>      |
| 2          | 20       | 8.9      | <i>G</i>      |
| 3          | 32       | 15.1     | <i>G</i>      |
| 4          | 22       | 10.4     | <i>G</i>      |
| 5          | 29       | 13.1     | <i>V</i>      |
| 6          | 27       | 12.4     | <i>V</i>      |
| 7          | 28       | 13.2     | <i>V</i>      |
| 8          | 26       | 11.8     | <i>V</i>      |
| 9          | 21       | 11.5     | <i>W</i>      |
| 10         | 27       | 14.2     | <i>W</i>      |
| 11         | 29       | 15.4     | <i>W</i>      |
| 12         | 23       | 13.1     | <i>W</i>      |
| 13         | 25       | 13.8     | <i>W</i>      |

**Tabla 2.** Catedrales

| <i>obs.</i> | <i>estilo</i> | <i>altura</i> | <i>longitud</i> | <i>obs.</i> | <i>estilo</i> | <i>altura</i> | <i>longitud</i> |
|-------------|---------------|---------------|-----------------|-------------|---------------|---------------|-----------------|
| 1           | 1             | 83            | 542             | 10          | 2             | 71            | 324             |
| 2           | 1             | 71            | 506             | 11          | 2             | 103           | 455             |
| 3           | 1             | 65            | 476             | 12          | 2             | 42            | 198             |
| 4           | 1             | 70            | 490             | 13          | 2             | 47            | 262             |
| 5           | 1             | 63            | 463             | 14          | 2             | 76            | 348             |
| 6           | 1             | 59            | 449             | 15          | 2             | 57            | 284             |
| 7           | 1             | 62            | 492             | 16          | 2             | 69            | 311             |
| 8           | 1             | 88            | 576             | 17          | 2             | 107           | 484             |
| 9           | 1             | 51            | 416             | 18          | 2             | 71            | 328             |
|             |               |               |                 | 19          | 2             | 96            | 442             |
|             |               |               |                 | 20          | 2             | 98            | 396             |
|             |               |               |                 | 21          | 2             | 87            | 425             |
|             |               |               |                 | 22          | 2             | 60            | 323             |
|             |               |               |                 | 23          | 2             | 95            | 397             |
|             |               |               |                 | 24          | 2             | 58            | 290             |
|             |               |               |                 | 25          | 2             | 56            | 282             |

**Tabla 3.** Dos rectas paralelas

| <i>obs.</i> | $x1j$ | $y1j$  | $x2j$ | $y2j$ |
|-------------|-------|--------|-------|-------|
| 1           | 1.86  | 4.53   | 5.19  | 7.48  |
| 2           | -1.65 | -10.88 | 3.13  | 6.25  |
| 3           | 4.26  | 0.93   | 3.67  | 6.73  |
| 4           | 9.43  | 16.35  | -0.15 | 1.09  |
| 5           | 3.09  | -0.13  | 2.39  | 7.56  |
| 6           | 11.19 | 6.2    | 9.07  | 12.15 |
| 7           | 5.12  | -0.93  | 6.44  | 8.32  |
| 8           | 5.04  | 3.76   | 2.3   | 4.92  |
| 9           | 0.69  | -3.1   | 3.27  | 3.53  |
| 10          | 10.88 | 5.8    | 3.33  | 2.64  |
| 11          | -1.16 | 0.81   | 6.19  | 8.6   |
| 12          | 0.96  | -5.5   | 5.3   | 9.69  |
| 13          | -4.77 | -6.05  | 3.59  | 6.06  |
| 14          | 5.78  | 9.36   | 0.61  | 7.78  |
| 15          | 10.58 | 6.41   | 4.28  | 8.93  |
| 16          | 8.57  | 10.46  | 7.32  | 13.29 |
| 17          | 3.67  | -0.36  | 7.63  | 11.43 |
| 18          | 6.25  | 1.59   | 6.75  | 10.37 |
| 19          | 9.6   | 11.59  | 6.78  | 7.93  |
| 20          | 6.57  | 7.29   | 1.77  | 5.63  |

**Tabla 4.** Cafeína

| <i>obs.</i> | nivel de consumo |              |              |
|-------------|------------------|--------------|--------------|
|             | <i>0mg</i>       | <i>100mg</i> | <i>200mg</i> |
| 1           | 242              | 248          | 246          |
| 2           | 245              | 246          | 248          |
| 3           | 244              | 245          | 250          |
| 4           | 248              | 247          | 252          |
| 5           | 247              | 248          | 248          |
| 6           | 248              | 250          | 250          |
| 7           | 242              | 247          | 246          |
| 8           | 244              | 246          | 248          |
| 9           | 246              | 243          | 245          |
| 10          | 242              | 244          | 250          |

**Tabla 5.** Consumo de combustible

| <i>obs</i> | <i>tax</i> | <i>dlic</i> | <i>fuel</i> | <i>obs</i> | <i>tax</i> | <i>dlic</i> | <i>fuel</i> |
|------------|------------|-------------|-------------|------------|------------|-------------|-------------|
| 1          | 9          | 52.5        | 541         | 25         | 8.5        | 55.1        | 460         |
| 2          | 9          | 57.2        | 524         | 26         | 9          | 54.4        | 566         |
| 3          | 9          | 58          | 561         | 27         | 8          | 54.8        | 577         |
| 4          | 7.5        | 52.9        | 414         | 28         | 7.5        | 57.9        | 631         |
| 5          | 8          | 54.4        | 410         | 29         | 8          | 56.3        | 574         |
| 6          | 10         | 57.1        | 457         | 30         | 9          | 49.3        | 534         |
| 7          | 8          | 45.1        | 344         | 31         | 7          | 51.8        | 571         |
| 8          | 8          | 55.3        | 467         | 32         | 7          | 51.3        | 554         |
| 9          | 8          | 52.9        | 464         | 33         | 8          | 57.8        | 577         |
| 10         | 7          | 55.2        | 498         | 34         | 7.5        | 54.7        | 628         |
| 11         | 8          | 53          | 580         | 35         | 8          | 48.7        | 487         |
| 12         | 7.5        | 52.5        | 471         | 36         | 6.58       | 62.9        | 644         |
| 13         | 7          | 57.4        | 525         | 37         | 5          | 56.6        | 640         |
| 14         | 7          | 54.5        | 508         | 38         | 7          | 58.6        | 704         |
| 15         | 7          | 60.8        | 566         | 39         | 8.5        | 66.3        | 648         |
| 16         | 7          | 58.6        | 635         | 40         | 7          | 67.2        | 968         |
| 17         | 7          | 57.2        | 603         | 41         | 7          | 62.6        | 587         |
| 18         | 7          | 54          | 714         | 42         | 7          | 56.3        | 699         |
| 19         | 7          | 72.4        | 865         | 43         | 7          | 60.3        | 632         |
| 20         | 8.5        | 67.7        | 640         | 44         | 7          | 50.8        | 591         |
| 21         | 7          | 66.3        | 649         | 45         | 6          | 67.2        | 782         |
| 22         | 8          | 60.2        | 540         | 46         | 9          | 57.1        | 510         |
| 23         | 9          | 51.1        | 464         | 47         | 7          | 62.3        | 610         |
| 24         | 9          | 51.7        | 547         | 48         | 7          | 59.3        | 524         |

**Tabla 6.** Biomasa

| <i>OBS</i> | <i>K</i> | <i>LOC</i> | <i>PH</i> | <i>SAL</i> | <i>SODIO</i> | <i>TYPE</i> | <i>Zn</i> | <i>BIO</i> |
|------------|----------|------------|-----------|------------|--------------|-------------|-----------|------------|
| 1          | 1441.67  | 1          | 5         | 33         | 35184.5      | 1           | 16.4524   | 676        |
| 2          | 1299.19  | 1          | 4.75      | 35         | 28170.4      | 1           | 13.9852   | 516        |
| 3          | 1154.27  | 1          | 4.2       | 32         | 26455        | 1           | 15.3276   | 1052       |
| 4          | 1045.15  | 1          | 4.4       | 30         | 25072.9      | 1           | 17.3128   | 868        |
| 5          | 521.62   | 1          | 5.55      | 33         | 31664.2      | 1           | 22.3312   | 1008       |
| 6          | 1273.02  | 1          | 5.05      | 33         | 25491.7      | 2           | 12.2778   | 436        |
| 7          | 1346.35  | 1          | 4.25      | 36         | 20877.3      | 2           | 17.8225   | 544        |
| 8          | 1253.88  | 1          | 4.45      | 30         | 25621.3      | 2           | 14.3516   | 680        |
| 9          | 1242.65  | 1          | 4.75      | 38         | 27587.3      | 2           | 13.6826   | 640        |
| 10         | 1282.95  | 1          | 4.6       | 30         | 26511.7      | 2           | 11.7566   | 492        |
| 11         | 553.69   | 1          | 4.1       | 30         | 7886.5       | 3           | 9.882     | 984        |
| 12         | 494.74   | 1          | 3.45      | 37         | 14596        | 3           | 16.6752   | 1400       |
| 13         | 526.97   | 1          | 3.45      | 33         | 9826.8       | 3           | 12.373    | 1276       |
| 14         | 571.14   | 1          | 4.1       | 36         | 11978.4      | 3           | 9.4058    | 1736       |
| 15         | 408.64   | 1          | 3.5       | 30         | 10368.6      | 3           | 14.9302   | 1004       |
| 16         | 646.65   | 2          | 3.25      | 30         | 17307.4      | 1           | 31.2865   | 396        |
| 17         | 514.03   | 2          | 3.35      | 27         | 12822        | 1           | 30.1652   | 352        |
| 18         | 350.73   | 2          | 3.2       | 29         | 8582.6       | 1           | 28.5901   | 328        |
| 19         | 496.29   | 2          | 3.35      | 34         | 12369.5      | 1           | 19.8795   | 392        |
| 20         | 580.92   | 2          | 3.3       | 36         | 14731.9      | 1           | 18.5056   | 236        |
| 21         | 535.82   | 2          | 3.25      | 30         | 15060.6      | 2           | 22.1344   | 392        |
| 22         | 490.34   | 2          | 3.25      | 28         | 11056.3      | 2           | 28.6101   | 268        |
| 23         | 552.39   | 2          | 3.2       | 31         | 8118.9       | 2           | 23.1908   | 252        |
| 24         | 661.32   | 2          | 3.2       | 31         | 13009.5      | 2           | 24.6917   | 236        |
| 25         | 672.15   | 2          | 3.35      | 35         | 15003.7      | 2           | 22.6758   | 340        |
| 26         | 525.65   | 2          | 7.1       | 29         | 10225        | 3           | 0.3729    | 2436       |
| 27         | 563.13   | 2          | 7.35      | 35         | 8024.2       | 3           | 0.2703    | 2216       |
| 28         | 497.96   | 2          | 7.45      | 35         | 10393        | 3           | 0.3205    | 2096       |
| 29         | 458.38   | 2          | 7.45      | 30         | 8711.6       | 3           | 0.2648    | 1660       |
| 30         | 498.25   | 2          | 7.4       | 30         | 10239.6      | 3           | 0.2105    | 2272       |

| <i>OBS</i> | <i>K</i> | <i>LOC</i> | <i>PH</i> | <i>SAL</i> | <i>SODIO</i> | <i>TYPE</i> | <i>Zn</i> | <i>BIO</i> |
|------------|----------|------------|-----------|------------|--------------|-------------|-----------|------------|
| 31         | 936.26   | 3          | 4.85      | 26         | 20436        | 1           | 18.9875   | 824        |
| 32         | 894.79   | 3          | 4.6       | 29         | 12519.9      | 1           | 20.9687   | 1196       |
| 33         | 941.36   | 3          | 5.2       | 25         | 18979        | 1           | 23.9841   | 1960       |
| 34         | 1038.79  | 3          | 4.75      | 26         | 22986.1      | 1           | 19.9727   | 2080       |
| 35         | 898.05   | 3          | 5.2       | 26         | 11704.5      | 1           | 21.3864   | 1764       |
| 36         | 989.87   | 3          | 4.55      | 25         | 17721        | 2           | 23.7063   | 412        |
| 37         | 951.28   | 3          | 3.95      | 26         | 16485.2      | 2           | 30.5589   | 416        |
| 38         | 939.83   | 3          | 3.7       | 26         | 17101.3      | 2           | 26.8415   | 504        |
| 39         | 925.42   | 3          | 3.75      | 27         | 17849        | 2           | 27.7292   | 492        |
| 40         | 954.11   | 3          | 4.15      | 27         | 16949.6      | 2           | 21.5699   | 636        |
| 41         | 720.72   | 3          | 5.6       | 24         | 11344.6      | 3           | 19.6531   | 1756       |
| 42         | 782.09   | 3          | 5.35      | 27         | 14752.4      | 3           | 20.3295   | 1232       |
| 43         | 773.3    | 3          | 5.5       | 26         | 13649.8      | 3           | 19.588    | 1400       |
| 44         | 829.26   | 3          | 5.5       | 28         | 14533        | 3           | 20.1328   | 1620       |
| 45         | 856.96   | 3          | 5.4       | 28         | 16892.2      | 3           | 19.242    | 1560       |