

Durante la producción de celulosa de madera en la industria papelera la lignina de la madera se separa de la celulosa, y ésta conforma las fibras del papel. El licor negro es la combinación del residuo de la lignina con agua y los químicos usados para extracción de la lignina. El vertido de licor negro tiene un pH elevado que contamina los acuíferos. En la tabla siguiente vemos el pH de medidas efectuadas en tres tubos piezométricos, cada uno de los cuales está conectado a un acuífero.

| Tubo (<i>i</i>) | pH | | | | | | $\sum_j y_{ij}$ | \bar{y}_i | s_i^2 |
|-------------------|-----|-----|-----|-----|-----|-----|-----------------|-------------|---------|
| 1 | 7.0 | 7.2 | 7.5 | 7.7 | 8.7 | 7.8 | 45.9 | 7.650 | 0.355 |
| 2 | 6.3 | 6.9 | 7.0 | 6.4 | 6.8 | 6.7 | 40.1 | 6.683 | 0.078 |
| 3 | 8.4 | 7.6 | 7.5 | 7.4 | 9.3 | 9.0 | 49.2 | 8.200 | 0.676 |

a) Con el objetivo de contrastar si el nivel medio de pH es igual en los tres tubos piezométricos, se determina con R la tabla ANOVA para los datos del enunciado. Se reproducen a continuación las correspondientes salidas del programa, de las que se han borrado seis resultados (sustituidos por !!).

```
Datos = read.table("pHtubo.txt", header=TRUE)
pH <- Datos$pH
Tubo = factor(Datos$Tubo)
resultado = aov(pH ~ Tubo)
summary(resultado)
```

```
          Df Sum Sq Mean Sq F value Pr(>F)
Tubo      !!      !!      !!    9.572 0.00209 **
Residuals !!      !!      !!
```

 Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1
 Determinar los valores borrados de la tabla. Contrastar si el nivel medio de pH depende del tubo piezométrico, indicando claramente el contraste de hipótesis realizado.

b) Escribe la expresión del modelo unifactorial y estima los parámetros del modelo.

Solución:

a) Tamaños muestrales $n_1 = n_2 = n_3 = 6$. Media muestral global $\bar{y}_{..} = (45.9 + 40.1 + 49.2)/18 = 7.511$. Suma de cuadrados residual:

$$RSS = \sum_{i=1}^3 (n_i - 1)s_i^2 = 5(0.355 + 0.078 + 0.676) = 5.545$$

Suma de cuadrados explicada por el factor:

$$MSS = \sum_{i=1}^3 n_i(\bar{y}_i - \bar{y}_{..})^2 = 7.074$$

Los grados de libertad debidos al factor y residuales son 3-1=2 y 18-3=15 respectivamente. Los cuadrados medios (Mean Sq) se obtienen dividiendo una suma de cuadrados por sus correspondientes grados de libertad. La tabla ANOVA completa es, pues,

```
          Df Sum Sq Mean Sq F value Pr(>F)
Tubo      2  7.074   3.537   9.572 0.00209 **
Residuals 15  5.543   0.370
```

 Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

El contraste de hipótesis pedido es $H_0 : \beta_1 = \beta_2 = \beta_3$ (el factor no influye sobre la variable respuesta), siendo β_i el valor esperado de la respuesta (pH) para el nivel i del factor (tubo o acuífero). Como el p-valor del contraste es 0.00209, rechazamos H_0 a todos los niveles de significación habituales.

b) Si Y_{ij} representa el j -ésimo pH en el i -ésimo tubo, el modelo unifactorial empleado en (a) es: $Y_{ij} \sim N(\beta_i, \sigma^2)$, $i = 1, 2, 3$, $j = 1, \dots, 6$.

Estimación de los parámetros β_i :

$$\hat{\beta}_1 = \bar{y}_{1.} = 7.650 \quad \hat{\beta}_2 = \bar{y}_{2.} = 6.683 \quad \hat{\beta}_3 = \bar{y}_{3.} = 8.200$$

Estimación de $\sigma^2 = \text{Varianza residual} = s_R^2 = 0.370$.