Modelado de la dependencia entre las concentraciones de estrona y de esteroide

Mariano J. Valderrama y Francisco A. Ocaña

30 de abril de 2009

Introducción

Este ejemplo ilustra los conceptos de inferencia asociados a la modelización lineal de la dependencia entre dos variables aleatorias, cuyas observaciones vienen dadas a través de una muestra aleatoria simple. Además de la estimación puntual del modelo, se llevan a cabo estimaciones por intervalo de distintas características suyas.

Planteamiento

Se ha realizado un experimento para describir la relación entre la concentración de estrona en saliva (X) y la concentración de esteroide libre en plasma (Y). Para disponer de observaciones de ambas variables aleatorias, se seleccionaron aleatoriamente 20 pacientes, recogiéndose sobre ellos observaciones conjuntas de ambas variables aleatorias. Así, se obtuvo un conjunto de datos dado por

$$\{(x_i, y_i) : \; i = 1, \ldots, 20\},$$

el cuál aparece almacenado en dos tipos de ficheros, `estronaesteroide.xls` o `estronaesteroide.ols`. De hecho, todos los cálculos realizados junto con los gráficos propuestos en esta práctica se encuentran disponibles también en los anteriores ficheros.

Para modelar la dependencia entre ambas concentraciones, proponemos el modelo lineal

$$Y \;=\; a \,+\, b\,X \,+\, \epsilon, \qquad (1)$$

siendo a y $b \in \mathbb{R}$, denominados parámetros del modelo, y ϵ una variable aleatoria (residuo), con media 0 y desviación estándar σ_R (desviación típica residual).

Antes de continuar con los cálculos, se propone al lector reflexionar sobre las ventajas prácticas que puede tener considerar el modelo anteriormente propuesto.

Cálculos

A continuación, se enumeran los cálculos y representaciones gráficas realizadas en los ficheros anteriores de Excel y Calc.

1. Para obtener una primera idea acerca de la idoneidad del modelo lineal, obtenemos el diagrama de dispersión de los datos observados $\{(x_i, y_i) : i = 1, \ldots, 20\}$. (Ver hoja `Dispersion`.)

2. Estimamos (puntualmente) los parámetros del modelo (1), dados por a, b y σ_R. Para ello, y para otros cálculos que realizaremos, calculamos previamente estimaciones de la media y la desviación típica de las distribuciones de las vv.aa. X e Y. (Ver hoja `AjustLineal`.)

   Nótese que, además de los parámetros propiamente del modelo, podemos aproximarnos al estudio de los residuos a través de $\hat{\sigma_R}$.

3. La idoneidad del modelo lineal estimado puede ser cuantificada a través del coeficiente de determinación R^2. Una ilustración visual de este hecho viene dada a través de la representación gráfica de los datos observados junto con el modelo lineal ajustado $\hat{a} + \hat{b}x$. (Ver hojas `AjustLineal` y `GrafRecta`.)

4. Podemos incorporar la aleatoriedad subyacente de los datos recogidos de los pacientes a las estimaciones de los parámetros del modelo. Para ello, obtenemos los intervalos de confianza de a y b, para un cierto nivel de confianza. (Ver hoja `AjustLineal`.)

5. Esa aleatoriedad, presente en nuestro problema, se traslada a la propia función (esperanza condicionada) $E[Y/X = x] = a + bx$, pues su estimación depende de los valores de una muestra. Para ilustrar este hecho, obtenemos los intervalos de confianza para la función $E[Y/X = x] =$

$a + bx$. Esta incertidumbre puede visualizarse representando los extremos de los intervalos de confianza así calculados, junto con los datos y el modelo lineal ajustado. (Ver hojas `AjustLineal` y `GrFMedia`.)

6. Una vez estimado el modelo lineal (1) a partir de los datos de la muestra de 20 pacientes, podríamos preguntarnos sobre la idoneidad de dicho modelo teórico para las observaciones recogidas. La validación del modelo lineal propuesto puede ser analizada obteniendo intervalos de confianza para las observaciones de Y, utilizando el modelo probabilístico (1). De nuevo, la representación gráfica de estos intervalos de confianza junto con los datos nos proporcionará una imagen visual que permitirá responder a la cuestión planteada. (Ver hojas `AjustLineal` y `GrICdatos`.)

7. Una de las aplicaciones de los modelos lineales consiste en predecir el valor de Y con sólo conocer el valor de X. Para una secuencia de nuevos valores de X, se obtienen predicciones, puntuales y por intervalos, para el valor que alcanzaría la v.a. Y. La representación gráfica de todos los elementos calculados hasta ahora nos proporcionarán una imagen de conjunto del análisis realizado con las dos magnitudes X e Y. (Ver hojas `AjustLineal` y `GrICdatPred`.)

Referencias

[1] Valderrama Bonnet, Mariano J. (1995), *Modelos Matemáticos en las Ciencias Experimentales.* Ediciones Pirámide: Madrid.