Una Breve Introducción a los Modelos de Regresión

ALBERTO LUCEÑO

Universidad de Cantabria

Autor: ALBERTO LUCEÑO VÁZQUEZ, Catedrático de Universidad.

Título: *Modelos de Regresión y de Series Temporales*. Publicado por la Universidad de Cantabria. Año: 2006.

Dirección: Departamento de Matemática Aplicada y Ciencias de la Computación, Avda de los Castros s/n, 39005 Santander.

Capítulo 1

Modelos de Regresión

1.1. Técnicas de observación

El avance del conocimiento humano se basa en el uso combinado de *métodos deductivos*, basados fundamentalmente en el uso de la razón, y *métodos inductivos*, basados–además–en la información obtenida de la observación de la Realidad. Uno de los objetivos prioritarios de los métodos deductivos es la "construcción de modelos" que sean coherentes desde un punto de vista lógico; ejemplos de modelos[1] estadísticos son los modelos de probabilidad y de distribuciones de probabilidad (en la recta real, o en espacios de 2 o más dimensiones). El objetivo más importantes de los métodos inductivos es la "interpretación de las observaciones".

El método científico surge de la interacción de los métodos deductivos e inductivos. Por una parte, las ciencias experimentales no pueden aceptar modelos teóricos que sean incompatibles con las observaciones de la Realidad. En sentido contrario, para interpretar las observaciones de la Realidad es necesario disponer de modelos teóricos que definan las magnitudes básicas a considerar y estudien las posibles relaciones entre ellas.

Por lo que respecta a las técnicas de observación pueden distinguirse dos grandes grupos de técnicas. El primer grupo está formado por *técnicas de observación inteligente* que son aquellas técnicas de observación que tratan de asegurar que los "sucesos que ocurren de forma natural" y que pueden ser importantes desde un punto de vista científico o técnico no pasan desapercibidos para el científico o ingeniero. A este grupo pertenecen[2], por ejemplo, las

[1] A lo largo de su carrera, el estudiante aprende muchos modelos físicos.

[2] En todo el mundo hay innumerables observatorios dedicados a la recogida de datos cualesquiera que sean el momento y el lugar en que se produzcan dichos datos. Ejemplos importantes son los observatorios astronómicos, meteorológicos o geológicos.

técnicas de control estadístico de procesos (SPC). El segundo grupo está formado por las *técnicas de experimentación dirigida* que tienen por objetivo "provocar la ocurrencia de sucesos" potencialmente importantes desde un punto de vista científico o técnico. A este grupo pertenecen[3] los métodos de diseño de experimentos (DoE).

1.2. Experimentos factoriales

Los experimentos factoriales son el tipo de experimentos de mayor interés para el ingeniero. En un experimento factorial típico hay una respuesta aleatoria cuyos valores dependen de varias covariables o factores. Algunos de estos factores pueden ser cualitativos y otros cuantitativos. Además pueden existen factores ambientales que no se pueden controlar, mientras que hay otros factores controlables. Se trata de analizar la relación que existe entre los valores o niveles de los factores y los valores de la respuesta aleatoria. El objetivo puede ser determinar los niveles de los factores que hacen que la respuesta sea óptima desde un punto de vista ingenieríl y al mismo tiempo dicha respuesta sea insensible a los factores ambientales. Por ejemplo, podemos desear diseñar el motor de un automóvil para que su rendimiento sea alto, su consumo sea bajo, produzca poca contaminación y al mismo tiempo conserve todas estas características en verano, en invierno, en clima seco y en clima lluvioso.

Existen dos maneras de diseñar un experimento factorial, a saber:

- Diseño de un factor cada vez: Con este diseño se elige un nivel de partida para cada uno de los factores y se hacen experimentos con estos niveles y con todas las combinaciones de niveles de los factores que resultan al variar los factores "de uno en uno".

- Diseños factoriales: Con este diseño se hacen experimentos variando todos los factores a la vez de forma que se cubran de forma equilibrada todas las posibles combinaciones de los niveles de los factores.

Es importante conocer que *los diseños factoriales son siempre superiores a los diseños de un factor cada vez*. Entre las ventajas de los diseños factoriales están:

1. Si los factores actúan sobre la respuesta de forma independiente, el diseño factorial proporciona mayor precisión que el diseño de un factor cada vez.

[3]La experimentación dirigida es una herramienta fundamental de las ciencias físicas, químicas y biológicas y en general de todas las ciencias de la Naturaleza.

2. Si los factores interaccionan, el diseño factorial permite detectar y estimar estas interacciones mientras que el diseño de un factor cada vez no lo permite.

Para llevar a cabo un diseño factorial completo, el experimentador tiene que seleccionar un número fijo de niveles para cada uno de los factores y, a continuación, realizar ensayos con todas las posible combinaciones formadas tomando un nivel de cada factor. Si hay k factores, y cada uno de ellos tiene $l_1, \ldots, l_k$ niveles, respectivamente, el número de ensayos (realizaciones o datos) requeridos es $l_1 \times \cdots \times l_k$. Por ejemplo, en el diseño factorial $2 \times 3 \times 5$, se usan 3 factores con 2, 3 y 5 niveles cada uno, siendo el número de ensayos requeridos $2 \times 3 \times 5 = 30$. Cada repetición completa de un diseño factorial básico es una réplica; por ejemplo, un diseño factorial $2 \times 3 \times 5$ replicado 3 veces requiere $30 \times 3 = 90$ ensayos.

La principal restricción al uso de los diseños factoriales es que suelen requerir muchos ensayos. Frecuentemente no se dispone de medios suficientes para realizar todos los ensayos requeridos por un diseño factorial completo. Existen dos maneras de evitar este inconveniente supuesto que el número de factores k es fijo:

1. Seleccionando pocos niveles para cada factor. Así surgen los diseños $2 \times \cdots \times 2$ o diseños 2^k, ya que 2 es el número mínimo de niveles que puede tener un factor. También pueden usarse diseños 3^k, $2^m 3^{k-m}$, etc.

2. Haciendo diseños factoriales fraccionales o incompletos.

Los diseños factoriales deben aleatorizarse asignando al azar las combinaciones de niveles de los factores a las unidades experimentales y sorteando el orden de realización de los ensayos.

1.3. Modelos de regresión

Los modelos de regresión pertenecen al grupo de modelos de Inferencia Estadística. Se usan frecuentemente en situaciones prácticas en las que:

1. Existe una variable (aleatoria), llamada *respuesta*, cuyos valores dependen de los valores de otras variables, llamadas *covariables*, siendo posible observar todas ellas simultáneamente.

2. Pueden asumirse las hipótesis realizadas para el estudio teórico de las propiedades del modelo.

1.4. Modelos de regresión lineal

Supongamos que en n individuos (o unidades experimentales) se observan simultáneamente los valores de una respuesta z_i y de q covariables $w_{1i}, \ldots, w_{qi}$, donde $i = 1, \ldots, n$. La forma general de un modelo de regresión lineal es

$$y_i = \sum_{j=1}^{p} \beta_j x_{ij} + \epsilon_i \quad (i = 1, \ldots, n) \tag{1.1}$$

donde $y_i = h(z_i; w_{1i}, \ldots, w_{qi})$ es una función cualquiera, pero conocida, de la respuesta z_i en la que podrían intervenir una o más covariables $w_{1i}, \ldots, w_{qi}$; $x_{ij} = g_i(w_{1i}, \ldots, w_{qi})$ son funciones cualesquiera, pero conocidas, de las covariables; $\beta_1, \ldots, \beta_p$ son *parámetros* con valores desconocidos; y $\epsilon_1, \ldots \epsilon_n$ son *errores aleatorios*. Además se supone que estos errores aleatorios son variables aleatorias independientes entre sí e igualmente distribuidas (IIDs), cuya distribución de probabilidad común es normal con media cero y varianza σ^2 desconocida.

Llamando μ_i a la media de y_i, el modelo (1.1) puede escribirse

$$\begin{aligned} y_i &= \mu_i + \epsilon_i \\ \mu_i &= \sum_{j=1}^{p} \beta_j x_{ij}. \end{aligned} \tag{1.2}$$

Los siguientes son ejemplos de modelos de regresión lineal:

- El modelo
$$z_i = \beta_1 + \beta_2 w_i + \epsilon_i$$
es lineal, ya que puede ponerse en la forma (1.1) siendo $y_i = z_i$, $p = 2$, $x_{i1} = 1$ y $x_{i2} = w_i$.

  En este ejemplo, la media de la respuesta $\mu_i = \beta_1 + \beta_2 w_i$ es una función lineal de la covariable w_i (suele decirse que se está ajustando una recta).

- El modelo
$$z_i = \beta_1 + \beta_2 w_{1i} + \beta_3 w_{2i} + \epsilon_i$$
es similar al anterior y también es lineal. Puede ponerse en la forma (1.1) siendo $y_i = z_i$, $p = 3$, $x_{i1} = 1$, $x_{i2} = w_{1i}$ y $x_{i3} = w_{2i}$.

  En este ejemplo, la media de la respuesta $\mu_i = \beta_1 + \beta_2 w_{i1} + \beta_3 w_{2i}$ es una función lineal de dos covariables w_{1i} y w_{2i} (suele decirse que se está ajustando un plano a la media de la respuesta).

- También es lineal el modelo

$$\sqrt{z_i} = \beta_1 + \beta_2 w_i + \beta_3 w_i^2 + \epsilon_i,$$

ya que puede ponerse en la forma (1.1) siendo $y_i = \sqrt{z_i}$, $p = 3$, $x_{i1} = 1$, $x_{i2} = w_i$ y $x_{i3} = w_i^2$.

En este ejemplo no es sencillo encontrar la expresión de la media de la respuesta z_i, pero vemos que la media de su raíz cuadrada $y_i = \sqrt{z_i}$ es $\mu_i = \beta_1 + \beta_2 w_i + \beta_3 w_i^2$ una función de segundo grado en w (estamos ajustando una parábola a la media de $y_i = \sqrt{z_i}$).

- Otro ejemplo de modelo lineal es

$$\frac{z_i}{w_{1i}} = \beta_1 + \beta_2 e^{w_{1i}} + \beta_3 w_{2i} + \epsilon_i,$$

ya que puede ponerse en la forma (1.1) siendo ahora $y_i = \frac{z_i}{w_{1i}}$ una función de la respuesta z_i y de la primera covariable w_{1i}, cuya media es $\mu_i = \beta_1 + \beta_2 e^{w_{1i}} + \beta_3 w_{2i}$. Además, $p = 3$, $x_{i1} = 1$, $x_{i2} = e^{w_{1i}}$ y $x_{i3} = w_{2i}$.

Por el contrario, los siguientes modelos no pueden ponerse en la forma (1.1) y, por tanto, no son modelos de regresión lineal:

- El modelo

$$z_i = \beta_1 + \beta_1^2 w_i + \epsilon_i$$

es un modelo de regresión no lineal, a pesar de que la media de la respuesta $\mu_i = \beta_1 + \beta_1^2 w_i$ es una función lineal de la covariable w_i, ya que interviene el cuadrado del parámetro β_1 en la expresión del modelo y no es posible deshacer esta falta de linealidad.

- El siguiente también es un modelo de regresión no lineal

$$z_i = \beta_1 w_i^{\beta_2} + \epsilon_i$$

porque no puede ponerse en la forma (1.1). La media de z_i es una función $\mu_i = \beta_1 w_i^{\beta_2}$ no lineal respecto de β_2.

Sin embargo, el modelo $\ln z_i = \gamma_1 + \beta_2 \ln w_i + \epsilon_i$ sí es lineal. Debe notarse que este modelo no coincide exactamente con el anterior, ya que ahora el modelo proporciona la media de $\ln z_i$, que es una función $\mu_i = \gamma_1 + \beta_2 \ln w_i$ lineal respecto de γ_1 y β_2. Los residuos de ambos modelos no tienen la misma distribución de probabilidad.

- En el contexto de las redes neuronales artificiales se utilizan modelos de regresión no lineal del tipo

$$u_k = \varphi\left(\sum_{j=0}^{m_1} \pi_{kj} w_j\right), \tag{1.3}$$

$$z_i = \varphi\left(\sum_{j=0}^{m_2} \tau_j u_j\right) + \epsilon_i. \tag{1.4}$$

En la expresión (1.3), la k-ésima neurona artificial de la primera capa recibe m_1 señales de entrada $w_1, \ldots, x_{m_1}$ con pesos $\pi_{k1}, \ldots, \pi_{km_1}$ y produce una señal de salida u_k a través de la función de transferencia $\varphi(\cdot)$. Esta señal de salida de la neurona k de la primera capa pasa a ser la señal de entrada de una o más neuronas de la segunda capa, de acuerdo con la expresión (1.3) y así sucesivamente hasta llegar a la última capa en la que la ecuación (1.4) aplica (así pues, el modelo con ecuaciones (1.3) y (1.4) corresponde a dos capas).

La función de transferencia $\varphi(\cdot)$ debe elegirse adecuadamente para proporcionar buenas propiedades al modelo y simplificar su uso tanto como sea posible. Aunque es posible usar una función de transferencia lineal de la forma

$$\varphi(x) = x + \theta$$

(los llamados perceptrones), es más frecuente usar funciones no lineales. Por ejemplo, puede usarse una función de umbral del tipo

$$\varphi(x) = \begin{cases} 0 & \text{si } x < \theta \\ 1 & \text{si } x \geq \theta \end{cases}$$

de forma que θ es un umbral a partir del cual cambia la señal emitida. La función sigmoidal (su nombre procede de su forma en S)

$$\varphi(x) = \frac{1}{1 + e^{-x}},$$

es muy popular porque su derivada se calcula rápidamente usando la relación

$$\frac{d}{dx}\varphi(x) = \varphi(x)[1 - \varphi(x)].$$

La función sigmoidal es un caso particular de la función logística

$$\varphi(x) = \beta_1 \frac{1 + \beta_2 e^{-x/\beta_4}}{1 + \beta_3 e^{-x/\beta_4}}$$

usada por en biología para describir, por ejemplo, el crecimiento de determinados organismos.

1.5. Hipótesis del modelo de regresión

Para poder estimar los parámetros de un modelo de regresión (es decir, para poder ajustar el modelo), hay que realizar observaciones *independientes* de los valores de la respuesta y de las covariables en n individuos (unidades experimentales) elegidos al azar de entre los individuos de la población bajo estudio. La consideración estadística de la respuesta difiere de la de las covariables: la respuesta es siempre una variable aleatoria, cuyos valores se observan en n individuos; por el contrario, las covariables se consideran variables deterministas.

En muchos experimentos diseñados mediante técnicas de DoE, la consideración de las covariables como variables deterministas está plenamente justificada, ya que los valores de dichas covariables se eligen al hacer el diseño (por ejemplo, al fabricar un hormigón puedo elegir el valor de la covariable ´‘porcentaje de cemento”). Sin embargo, el valor de la respuesta nunca puede elegirse (por ejemplo, el valor de la respuesta “resistencia del hormigón” no es conocido, sino que es el resultado del experimento aleatorio de analizar una unidad experimental).

Además de la respuesta y las covariables, en la ecuación (1.1) intervienen los parámetros $\beta_1, \ldots, \beta_p$ y los errores aleatorios $\epsilon_1, \ldots \epsilon_n$. Ninguno de estos valores son observables. Nunca sabremos su valor exacto. A lo que sí podemos aspirar es a estimar los valores de los parámetros y de los errores aleatorios que se han producido en cada uno de los n individuos de la muestra.

La consideración estadística de los parámetros difiere de la de los errores aleatorios: los parámetros son valores desconocidos pero fijos (deterministas); por el contrario, los errores aleatorios son variables aleatorias que toman valores desconocidos en cada uno de los individuos de la muestra.

Las hipótesis del modelo de regresión pueden resumirse como sigue:

1. Independencia: Las respuesta transformadas (y_i) o no transformadas (z_i) son variables aleatorias independientes entre sí.

2. Linealidad: La media μ_i de la respuesta transformada y_i es una función lineal de los parámetros $\beta_1, \ldots, \beta_p$ que además puede depender de una o más covariables.

3. Homoscedasticidad: La varianza σ^2 de la respuesta transformada y_i es constante.

4. Normalidad: La distribución de probabilidad de la respuesta transformada y_i es normal para cualquier combinación de los valores de las covariables.

Si no se verifica la hipótesis 2, pero se verifican las demás, se dice que se trata de un modelo de regresión no lineal. Si no se verifica la hipótesis 1 o la hipótesis 3, puede que no sea posible usar un modelo de regresión (quizás pueda usarse un modelo de series temporales) o que pueda usarse un modelo de regresión ponderada. Si no se verifica la hipótesis 4, se dice que es un modelo de regresión no-normal (los modelos logit o probit son ejemplos clásicos de este tipo).

1.6. Forma matricial del modelo

Supongamos que en n individuos se observan simultáneamente los valores de una respuesta z_i y de q covariables $w_{1i}, \ldots, w_{qi}$, donde $i = 1, \ldots, n$. La expresión matricial de la ecuación (1.1) es

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_{11} & x_{12} & \ldots & x_{1p} \\ x_{21} & x_{22} & \ldots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \ldots & x_{np} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix} \tag{1.5}$$

o brevemente

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}. \tag{1.6}$$

El vector $\mathbf{Y}$ se llama vector respuesta y contiene los valores observados en los n individuos para la respuesta transformada. La matriz $\mathbf{X}$ se llama matriz de diseño y contiene los valores observados para las covariables transformadas en los n individuos. Los vectores $\boldsymbol{\beta}$ y $\boldsymbol{\epsilon}$ son los vectores de parámetros y de errores y no son observables. El vector $\boldsymbol{\epsilon}$ es aleatorio y su distribución de probabilidad conjunta es normal n dimensional con vector de medias nulo y matriz de varianzas-covarianzas $\sigma^2\mathbf{I}_n$ siendo, $\mathbf{I}_n$ la matriz identidad $n \times n$. En consecuencia, antes de ser observado, el vector $\mathbf{Y}$ es aleatorio y su distribución conjunta es normal n dimensional con vector de medias $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$ y matriz de varianzas-covarianzas $\sigma^2\mathbf{I}_n$

1.6.1. Covariables cuantitativas y cualitativas

Las covariables de un modelo de regresión lineal pueden ser cuantitativas si toman valores numéricos o cualitativas si no toman valores numéricos. En un modelo de regresión puede ocurrir que todas las covariables sean cuantitativas, que todas sean cualitativas o que haya covariables de ambos tipos. Para construir la matriz de diseño $\mathbf{X}$ es preciso conocer el tipo de cada covariable como se muestra en los ejemplos siguientes.

Resistencia del hormigón	447	503	522	474	508	556
Porcentaje de cemento	6	6	6	8	8	8
Tiempo de curado	20	25	30	20	25	30

Tabla 1.1: Resistencia alcanzada por un hormigón dependiendo del porcentaje de cemento utilizado y del tiempo de curado.

Ejemplo 1 *La tabla 1.1 muestra datos de la resistencia alcanzada por un hormigón (respuesta z) dependiendo del porcentaje de cemento utilizado (covariable w_1) y del tiempo de curado (covariable w_2). En este ejemplo, ambas covariables son cuantitativas y se trata de un diseño factorial completo 2×3; es decir, hay 2 factores cuantitativos, el primero con 2 niveles (6 y 8) y el segundo con 3 niveles (20, 25 y 30). Suponiendo que se deseara utilizar el modelo de regresión lineal*

$$z_i = \beta_1 + \beta_2 w_{1i} + \beta_3 w_{2i} + \beta_4 w_{1i} w_{2i} + \epsilon_i, \tag{1.7}$$

la ecuación (1.5) *sería*

$$\begin{pmatrix} 447 \\ 503 \\ 522 \\ 474 \\ 508 \\ 556 \end{pmatrix} = \begin{pmatrix} 1 & 6 & 20 & 120 \\ 1 & 6 & 25 & 150 \\ 1 & 6 & 30 & 180 \\ 1 & 8 & 20 & 160 \\ 1 & 8 & 25 & 200 \\ 1 & 8 & 30 & 240 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \\ \epsilon_6 \end{pmatrix}. \tag{1.8}$$

En la sección 1.7 y siguientes se verá la importancia de la matriz $\mathbf{X'X}$ obtenida al multiplicar la matriz traspuesta $\mathbf{X'}$ de la matriz de diseño por la propia matriz de diseño $\mathbf{X}$. En particular, es conveniente que $\mathbf{X'X}$ sea una matriz diagonal o diagonal por bloques. Sin embargo, la ecuación (1.8) *conduce a la siguiente matriz llena a pesar de que se trata de un diseño factorial*

$$\mathbf{X'X} = \begin{pmatrix} 6 & 42 & 150 & 1050 \\ 42 & 300 & 1050 & 7500 \\ 150 & 1050 & 3850 & 26950 \\ 1050 & 7500 & 26950 & 192500 \end{pmatrix}.$$

Por tanto, para aprovechar las ventajas del diseño factorial es conveniente reparametrizar el modelo como sigue

$$z_i = \delta_1 + \delta_2 x_{i1} + \delta_3 x_{i2} + \delta_4 x_{i1} x_{i2} + \epsilon_i, \tag{1.9}$$
$$x_{i1} = (w_{1i} - 7) \ ; \ x_{i2} = \left(\frac{w_{2i} - 25}{5} \right),$$

Resistencia del hormigón	447	503	522	474	508	556
Tipo de cemento	A	A	A	B	B	B
Tipo de aditivo	I	II	III	I	II	III

Tabla 1.2: Resistencia alcanzada por un hormigón dependiendo del tipo de cemento utilizado y del tipo de aditivo empleado.

de forma que

$$\mathbf{X} = \begin{pmatrix} 1 & -1 & -1 & 1 \\ 1 & -1 & 0 & 0 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix}.$$

y, por tanto,

$$\mathbf{X}'\mathbf{X} = \begin{pmatrix} 6 & 0 & 0 & 0 \\ 0 & 6 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 4 \end{pmatrix} \tag{1.10}$$

es una matriz diagonal. En la ecuación (1.9) *del modelo, al término* $\delta_2 x_{i1}$ *se le llama efecto principal del primer factor, al término* $\delta_3 x_{i2}$ *se le llama efecto principal del segundo factor, y al término* $\delta_4 x_{i1} x_{i2}$ *se le llama interacción entre los dos factores. La relación entre los parámetros* β_i *y los* δ_i *se deduce identificando las ecuaciones* (1.7) *y* (1.9).

Ejemplo 2 *La tabla 1.2 muestra datos de la resistencia alcanzada por un hormigón (respuesta* z*) dependiendo del tipo de cemento utilizado (covariable* w_1*) y del tipo de aditivo añadido al cemento (covariable* w_2*). Este ejemplo es muy similar al de la tabla 1.1 (se trata de un diseño factorial completo* 2×3*), pero hay una diferencia muy importante: ahora las dos covariables son cualitativas. Los niveles del primer factor son* A *y* B *y los del segundo factor son* I, II *y* III*. Puesto que no es posible multiplicar un parámetro por un factor cualitativo, no tiene sentido en este caso poner términos del tipo* $\beta_2 w_{1i}$ *en la ecuación del modelo. Sin embargo, pueden formularse modelos de regresión lineal de la forma*

$$z_{ij} = \mu + \alpha_i + \beta_j + \delta_{ij} + \epsilon_{ij}, \tag{1.11}$$

donde $i = 1$ *corresponde al nivel* A *del primer factor,* $i = 2$ *corresponde al nivel* B *del primer factor y, análogamente,* $j = 1, 2$ *y 3 corresponden a los*

niveles I, II y III del segundo factor. El término α_i es el efecto principal del primer factor, el término β_j es el efecto principal del segundo factor, y el término δ_{ij} es la interacción entre los dos factores.

Es conveniente observar que el modelo (1.11) *tiene algunos parámetros redundantes, por lo que dichos parámetros pueden elegirse arbitrariamente. Por ejemplo, o bien α_1 o bien α_2 es redundante, ya que el modelo contiene una media general μ y α_1 representa el incremento sobre μ que es debido al factor A mientras que α_2 representa el incremento sobre μ que es debido al factor B. Podría tomarse $\alpha_2 = 0$ cambiando adecuadamente α_1 y μ. Sin embargo, en lugar de asignar valores nulos a los parámetros redundantes, es más conveniente imponer condiciones de contorno que sean fáciles de interpretar. Por ejemplo, $\sum_{i=1}^{2} \alpha_i = 0$, $\sum_{j=1}^{3} \beta_j = 0$ y $\sum_{i=1}^{2} \delta_{ij} = \sum_{j=1}^{3} \delta_{ij} = 0$ para cada j y para cada i. De esta forma α_1 y α_2 tienen el mismo valor absoluto y signos opuestos.*

La ecuación (1.5) *correspondiente al modelo* (1.11) *es*

$$\left(\begin{array}{c} 447 \\ 503 \\ 522 \\ 474 \\ 508 \\ 556 \end{array}\right) = \left(\begin{array}{c|c|cc|cc} 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & -1 & 1 & 0 & -1 & 0 \\ 1 & -1 & 0 & 1 & 0 & -1 \\ 1 & -1 & -1 & -1 & 1 & 1 \end{array}\right) \left(\begin{array}{c} \mu \\ \hline \alpha_1 \\ \hline \beta_1 \\ \beta_2 \\ \hline \delta_{11} \\ \delta_{12} \end{array}\right) + \left(\begin{array}{c} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{13} \\ \epsilon_{21} \\ \epsilon_{22} \\ \epsilon_{23} \end{array}\right), \quad (1.12)$$

donde se han impuesto las condiciones de contorno $\sum_{i=1}^{2} \alpha_i = 0$, $\sum_{j=1}^{3} \beta_j = 0$ y $\sum_{i=1}^{2} \delta_{ij} = \sum_{j=1}^{3} \delta_{ij} = 0$ para cada j y para cada i. Estas condiciones de contorno eliminan los parámetros redundantes y hacen que la matriz $\mathbf{X}'\mathbf{X}$ no sea singular. Esta matriz es ahora diagonal por bloques

$$\mathbf{X}'\mathbf{X} = \left(\begin{array}{cc|cc|cc} 6 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 6 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 4 & 2 & 0 & 0 \\ 0 & 0 & 2 & 4 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 4 & 2 \\ 0 & 0 & 0 & 0 & 2 & 4 \end{array}\right).$$

El tamaño de los bloques es igual al número de parámetros no redundantes en cada uno de los efectos principales e interacciones. Además hay un bloque de dimensión 1 para la media general.

Ejemplo 3 *La tabla 1.3 muestra datos de la resistencia alcanzada por un hormigón (respuesta z) dependiendo del tipo de cemento utilizado (covariable*

Resistencia del hormigón	447	503	522	474	508	556
Tipo de cemento	A	A	A	B	B	B
Tiempo de curado	20	25	30	20	25	30

Tabla 1.3: Resistencia alcanzada por un hormigón dependiendo del tipo de cemento utilizado y del tiempo de curado.

w_1) y del tiempo de curado (covariable w_2). Este ejemplo es muy similar a los dos anterores (se trata de un diseño factorial completo 2×3), pero ahora un factor es cualitativo y el otro es cuantitativo. Los niveles del primer factor son A y B y los del segundo factor son 20, 25 y 30. Ahora pueden formularse modelos de regresión lineal de la forma

$$\begin{aligned} z_{ij} &= \mu + \alpha_i + \beta x_{j2} + \delta_i x_{j2} + \epsilon_{ij}, \\ x_{j2} &= \left(\frac{w_{2j} - 25}{5} \right), \end{aligned} \tag{1.13}$$

donde $i = 1$ corresponde al nivel A del primer factor, $i = 2$ corresponde al nivel B del primer factor y, análogamente, $w_{21} = 20$, $w_{22} = 25$ y $w_{23} = 30$. El término α_i es el efecto principal del primer factor, el término βx_{j2} es el efecto principal del segundo factor, y el término $\delta_i x_{j2}$ es la interacción entre los dos factores. De acuerdo con el modelo (1.13)*, la media de la respuesta es una función lineal del tiempo de curado, y esta función lineal (esta recta) es diferente para cada tipo de cemento.*

Las condiciones de contorno pueden ser ahora $\sum_{i=1}^{2} \alpha_i = 0$ y $\sum_{j=1}^{2} \delta_i = 0$. La ecuación (1.5) *correspondiente al modelo* (1.13) *es*

$$\left(\begin{array}{c} 447 \\ 503 \\ 522 \\ 474 \\ 508 \\ 556 \end{array} \right) = \left(\begin{array}{c|c|c|c} 1 & 1 & -1 & -1 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & 0 & 0 \\ 1 & -1 & 1 & -1 \end{array} \right) \left(\begin{array}{c} \mu \\ \hline \alpha_1 \\ \hline \beta \\ \hline \delta_1 \end{array} \right) + \left(\begin{array}{c} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \\ \epsilon_6 \end{array} \right). \tag{1.14}$$

La matriz $\mathbf{X}'\mathbf{X}$ coincide con la de la ecuación (1.10)*.*

1.7. Estimación de los parámetros

Los parámetros del modelo de regresión lineal dado por las ecuaciones (1.5) o (1.6) pueden estimarse usando el método de los mínimos cuadrados, el

método de la máxima verosimilitud u otros métodos (por ejemplo, el método de la máxima bondad de ajuste usado en Luceño 2006).

1.7.1. Método de mínimos cuadrados

Denotaremos con $\tilde{\boldsymbol{\beta}}$ el estimador de mínimos cuadrados del vector de parámetros $\boldsymbol{\beta}$ y con $\tilde{\boldsymbol{\mu}}$ el estimador de mínimos cuadrados del vector de medias $\boldsymbol{\mu}$. De acuerdo con la ecuación (1.6), $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$ por lo que $\tilde{\boldsymbol{\mu}} = \mathbf{X}\tilde{\boldsymbol{\beta}}$. Además, el error aleatorio $\boldsymbol{\epsilon}$ puede estimarse con $\mathbf{e} = \mathbf{Y} - \tilde{\boldsymbol{\mu}} = \mathbf{Y} - \mathbf{X}\tilde{\boldsymbol{\beta}}$.

El método de los mínimos cuadrados consiste en minimizar la suma de cuadrados

$$\sum_{i=1}^{n} e_i^2 = \mathbf{e}'\mathbf{e} \tag{1.15}$$

donde $\mathbf{e}'$ indica la traspuesta de $\mathbf{e}$. Puesto que

$$\mathbf{e}'\mathbf{e} = (\mathbf{Y} - \mathbf{X}\tilde{\boldsymbol{\beta}})'(\mathbf{Y} - \mathbf{X}\tilde{\boldsymbol{\beta}}), \tag{1.16}$$

vemos que la suma de cuadrados en (1.15) es una función de segundo grado de los estimadores $\tilde{\boldsymbol{\beta}}$ de los parámetros. Las derivadas primeras de la función (1.16) respecto de cada una de las componentes del vector $\tilde{\boldsymbol{\beta}}$ dan lugar al vector columna $2\mathbf{X}'(\mathbf{Y} - \mathbf{X}\tilde{\boldsymbol{\beta}})$. Igualando a cero las componentes de este vector se obtiene un sistema lineal de p ecuaciones con p incógnitas

$$\mathbf{X}'(\mathbf{Y} - \mathbf{X}\tilde{\boldsymbol{\beta}}) = \mathbf{0}, \tag{1.17}$$

cuya solución es el estimador $\tilde{\boldsymbol{\beta}}$.

El sistema (1.17) puede expresarse en la forma

$$(\mathbf{X}'\mathbf{X})\tilde{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{Y}, \tag{1.18}$$

donde $\mathbf{X}'\mathbf{X}$ es una matriz $p \times p$ de coeficientes y $\mathbf{X}'\mathbf{Y}$ es un vector $p \times 1$ de términos independientes. Por tanto, el sistema tiene solución única si el rango de $\mathbf{X}'\mathbf{X}$ coincide con su dimensión p (para lo cual es condición necesaria y suficiente que las p columnas de $\mathbf{X}$ sean linealmente independientes). En tal caso, la solución del sistema es

$$\tilde{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}. \tag{1.19}$$

En algunas aplicaciones puede ocurrir que $m > 0$ columnas de la matriz de diseño $\mathbf{X}$ sean combinaciones lineales de las demás columnas de $\mathbf{X}$. En tal caso el rango de $\mathbf{X}'\mathbf{X}$ es $p - m$ y el sistema (1.18) no tiene solución única. Sin embargo, este sistema es siempre compatible, ya que el rango de $\mathbf{X}'\mathbf{X}$

coincide con el rango de la matriz ampliada $(\mathbf{X}'\mathbf{X}|\mathbf{X}'\mathbf{Y})$, por lo que existen infinitas soluciones. Por ejemplo, el modelo $y_i = \beta_1 + \beta_2 x_i + \beta_3(2 + 3x_i) + \epsilon_i$ conducirá siempre a un sistema con infinitas soluciones, ya que la tercera covariable $(2 + 3x_i)$ es combinación lineal de las dos primeras: 1 y x_i; puede elegirse una solución única dando un valor arbitrario a β_3 (por ejemplo, $\beta_3 = 0$ con lo que el modelo se reduce a $y_i = \beta_1 + \beta_2 x_i + \epsilon_i$). Siempre es posible eliminar columnas dependientes de $\mathbf{X}$ introduciendo condiciones de contorno, como se ilustra en los ejemplos 1 y 2 de la sección 1.6.1. Salvo que se diga lo contrario, en lo sucesivo supondremos que el rango de $(\mathbf{X}'\mathbf{X})$ es p.

El sistema (1.17) puede escribirse en la forma $\mathbf{X}'\mathbf{e} = \mathbf{0}$, lo que indica que el vector de residuos estimados $\mathbf{e}$ es ortogonal a todas las columnas de la matriz de diseño $\mathbf{X}$. Esto implica a su vez que el vector de medias estimadas $\tilde{\boldsymbol{\mu}} = \mathbf{X}\tilde{\boldsymbol{\beta}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ es la proyección ortogonal del vector de respuestas $\mathbf{Y}$ sobre el espacio vectorial engendrado por las columnas de la matriz de diseño $\mathbf{X}$. A la matriz $\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, que permite proyectar cualquier vector $\mathbf{Y}$ sobre el espacio engendrado por las columnas de la matriz $\mathbf{X}$, se le llama *matriz proyección*. La figura 1.1 representa la geometría del ajuste por mínimos cuadrados.

Es importante señalar que los estimadores $\tilde{\boldsymbol{\beta}}$ son funciones lineales del vector aleatorio $\mathbf{Y}$. Por tanto, mientras que los parámetros $\boldsymbol{\beta}$ son constantes desconocidas, sus estimadores $\tilde{\boldsymbol{\beta}}$ son variables aleatorias, cuyos valores para la muestra observada se obtienen resolviendo el sistema (1.18).

El método de mínimos cuadrados puede usarse independientemente de que la distribución de la respuesta $\mathbf{Y}$ sea normal o no. No obstante, si la distribución de $\mathbf{Y}$ no es normal, el método de la máxima verosimilitud o los métodos que maximizan la bondad de ajuste suelen proporcionar mejores estimadores que el método de los mínimos cuadrados.

1.7.2. Método de la máxima verosimilitud

Este método consiste en maximizar la función de verosimilitud de los parámetros $\boldsymbol{\beta}$ y σ^2 del modelo (1.6), dada la muestra de valores de las covariables $\mathbf{X}$ y correspondientes valores de la respuesta $\mathbf{Y}$. Esta función de verosimilitud depende de la distribución de probabilidad de la respuesta $\mathbf{Y}$. En consecuencia, mientras que los estimadores de mínimos cuadrados no varían con la distribución de probabilidad de $\mathbf{Y}$, los estimadores de máxima verosimilitud (MLE) sí dependen de cuál sea esta distribución de $\mathbf{Y}$.

En el caso de que la distribución conjunta de $\mathbf{Y}$ sea normal n dimensional con vector de medias $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$ y matriz de varianzas-covarianzas $\sigma^2\mathbf{I}_n$, la función logaritmo-verosimilitud (es decir, el logaritmo de la función de

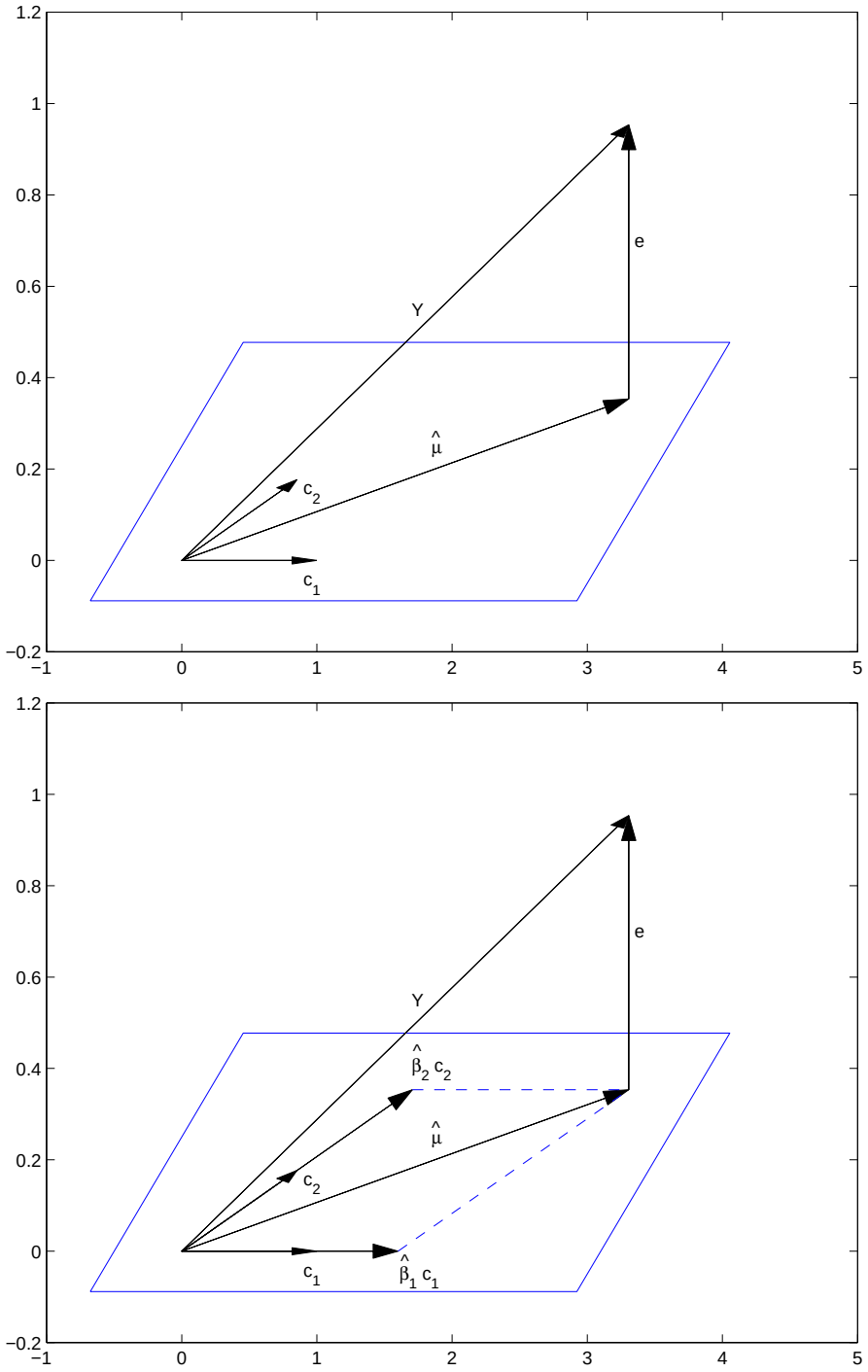


Figura 1.1: Geometría del ajuste por mínimos cuadrados: $\mathbf{Y} = \tilde{\boldsymbol{\mu}} + \mathbf{e}$ siendo $\mathbf{e}$ y $\tilde{\boldsymbol{\mu}}$ ortogonales; si $\mathbf{X}$ tiene columnas $\mathbf{c}_1$ y $\mathbf{c}_2$, se tiene $\tilde{\boldsymbol{\mu}} = \mathbf{c}_1\beta_1 + \mathbf{c}_2\beta_2$ y $\mathbf{e}$ es ortogonal a $\mathbf{c}_1$ y a $\mathbf{c}_2$; tanto $\mathbf{c}_1$ como $\mathbf{c}_2$ pueden representar varias columnas de $\mathbf{X}$. (En este contexto puede usarse $\tilde{\boldsymbol{\mu}}$ o $\hat{\boldsymbol{\mu}}$ indistintamente, aunque en general el símbolo $\hat{}$ indica estimación por maxima verosimilitud.)

verosimilitud) de los parámetros $\boldsymbol{\beta}$ y σ^2 dada la muestra es

$$\ell(\boldsymbol{\beta}, \sigma^2 | \mathbf{X}, \mathbf{Y}) = -\frac{n}{2} \ln\left(2\pi\sigma^2\right) - \frac{(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})}{2\sigma^2}. \tag{1.20}$$

Vemos, por tanto, que maximizar $\ell(\boldsymbol{\beta}, \sigma^2 | \mathbf{X}, \mathbf{Y})$ respecto de $\boldsymbol{\beta}$ es equivalente a minimizar la suma de cuadrados (1.16) respecto de $\boldsymbol{\beta}$. De ahí que el estimador de máxima verosimilitud $\widehat{\boldsymbol{\beta}}$ de $\boldsymbol{\beta}$, correspondiente a la hipótesis de normalidad de $\mathbf{Y}$, coincide con el estimador de mínimos cuadrados de $\boldsymbol{\beta}$ (esto no ocurre para otras distribuciones de $\mathbf{Y}$). Por tanto,

$$(\mathbf{X}'\mathbf{X})\widehat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{Y}. \tag{1.21}$$

Además, maximizando (1.20) respecto de σ^2 resulta el estimador de máxima verosimilitud de σ^2:

$$\widehat{\sigma}^2 = \frac{\mathbf{e}'\mathbf{e}}{n} = \frac{(\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}})'(\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}})}{n}. \tag{1.22}$$

Al numerador de (1.22) se le llama suma de cuadrados residual S_r y verifica

$$S_r = \mathbf{e}'\mathbf{e} = (\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}})'(\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}) = \mathbf{Y}'\mathbf{Y} - \widehat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{Y}. \tag{1.23}$$

1.8. Distribución de los estimadores

Bajo las hipótesis del modelo de regresión lineal, resumidas en la sección 1.5, la distribución de probabilidad del vector de estimadores $\widehat{\boldsymbol{\beta}}$ es normal p dimensional con vector de medias

$$E\left(\widehat{\boldsymbol{\beta}}\right) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E\left(\mathbf{Y}\right) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \boldsymbol{\beta}$$

y matriz de varianzas-covarianzas

$$\mathrm{Var}\left(\widehat{\boldsymbol{\beta}}\right) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathrm{Var}\left(\mathbf{Y}\right)\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = \sigma^2(\mathbf{X}'\mathbf{X})^{-1},$$

recordando que $\mathrm{Var}\left(\mathbf{Y}\right) = \sigma^2\mathbf{I}_n$.

En consecuencia, vemos que los estimadores $\widehat{\boldsymbol{\beta}}$ son centrados para $\boldsymbol{\beta}$. Además, las componentes $\widehat{\beta}_i$ y $\widehat{\beta}_j$ son variables aleatorias incorreladas (e independientes, bajo normalidad) si y solo si el elemento de la fila i y columna j de la matriz $(\mathbf{X}'\mathbf{X})^{-1}$ es nulo. Puesto que el que los estimadores de los parámetros sean incorrelados (independientes) facilita la interpretación de los resultados, es conveniente que la matriz $(\mathbf{X}'\mathbf{X})^{-1}$ sea diagonal o, si ello

no es posible, que sea diagonal por bloques; esto ocurre si y solo si $\mathbf{X}'\mathbf{X}$ es diagonal o diagonal por bloques, respectivamente (como ocurre en los ejemplos de la sección 1.6.1).

También bajo hipótesis de normalidad, la distribución de probabilidad de S_r/σ^2 es χ^2 con $n-p$ grados de libertad. De ahí que $E(S_r/\sigma^2) = n-p$ y $\text{Var}(S_r/\sigma^2) = 2(n-p)$. Por tanto, $E(\widehat{\sigma}^2) = \sigma^2(n-p)/n$ y $\text{Var}(\widehat{\sigma}^2) = 2(n-p)\sigma^4/n^2$. Aunque $\widehat{\sigma}^2$ es el estimador de mínimo error cuadrático medio para σ^2, es frecuente usar en su lugar el estimador centrado

$$s^2 = \frac{\mathbf{e}'\mathbf{e}}{n-p} = \frac{(\mathbf{Y}-\mathbf{X}\widehat{\boldsymbol{\beta}})'(\mathbf{Y}-\mathbf{X}\widehat{\boldsymbol{\beta}})}{n-p} = \frac{n}{n-p}\,\widehat{\sigma}^2, \tag{1.24}$$

cuya esperanza y varianza son $E(s^2) = \sigma^2$ y $\text{Var}(s^2) = 2\sigma^4/(n-p)$.

1.9. Estimación de funciones lineales de los parámetros

En muchas aplicaciones es importante poder estimar funciones lineales de los parámetros incluidos en el vector $\boldsymbol{\beta}$. Definamos

$$\boldsymbol{\xi} = \mathbf{H}\boldsymbol{\beta},$$

donde $\boldsymbol{\xi}$ es un vector columna de dimensión $r \leq p$ y $\mathbf{H}$ es una matriz constante de dimensión $r \times p$ y rango r. Por tanto, el vector $\boldsymbol{\xi}$ contiene r funciones lineales e independientes de los parámetros $\boldsymbol{\beta}$.

El estimador de máxima verosimilitud de $\boldsymbol{\xi}$ es

$$\widehat{\boldsymbol{\xi}} = \mathbf{H}\widehat{\boldsymbol{\beta}}.$$

Bajo hipótesis de normalidad, la distribución de probabilidad de $\widehat{\boldsymbol{\xi}}$ es normal r dimensional con vector de medias $E(\widehat{\boldsymbol{\xi}}) = \boldsymbol{\xi}$ y mariz de varianzas-covarianzas $\text{Var}\left(\widehat{\boldsymbol{\xi}}\right) = \sigma^2\mathbf{H}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{H}'$.

De este resultado se deduce que la variable aleatoria

$$\sigma^{-2}\left(\widehat{\boldsymbol{\xi}}-\boldsymbol{\xi}\right)'\left[\mathbf{H}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{H}'\right]^{-1}\left(\widehat{\boldsymbol{\xi}}-\boldsymbol{\xi}\right)$$

tiene distribución χ^2 con r grados de libertad y es independiente de s^2 (y de $\widehat{\sigma}^2$). Además, la distribución de probabilidad de

$$\frac{1}{rs^2}\left(\widehat{\boldsymbol{\xi}}-\boldsymbol{\xi}\right)'\left[\mathbf{H}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{H}'\right]^{-1}\left(\widehat{\boldsymbol{\xi}}-\boldsymbol{\xi}\right) \tag{1.25}$$

es una F de Snedecor con r y $n-p$ grados de libertad. Esta distribución de probabilidad de la variable aleatoria (1.25) permite obtener intervalos de confianza y regiones de confianza para funciones lineales de los parámetros $\boldsymbol{\beta}$ (y en particular para los propios parámetros $\boldsymbol{\beta}$).

1.10. Regiones e intervalos de confianza

Conociendo la distribución de probabilidad de la variable aleatoria (1.25) es posible construir una región de confianza conjunta para las funciones lineales de los parámetros $\boldsymbol{\beta}$ incluidas en el vector $\boldsymbol{\xi}$, al nivel de confianza $1-\alpha$, para cualquier $0<\alpha<1$. A dicha región de confianza pertenecen todos los puntos $\boldsymbol{\xi}$ que verifican la desigualdad

$$\left(\widehat{\boldsymbol{\xi}}-\boldsymbol{\xi}\right)' \left[\mathbf{H}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{H}'\right]^{-1} \left(\widehat{\boldsymbol{\xi}}-\boldsymbol{\xi}\right) \leq r s^2 F(r, n-p, 1-\alpha), \tag{1.26}$$

donde $F(r, n-p, 1-\alpha)$ es el cuantil $1-\alpha$ de la distribución F de Snedecor con r y $n-p$ grados de libertad en el numerador y denominador, respectivamente.

Puesto que la matriz $[\mathbf{H}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{H}']^{-1}$ tiene todos sus autovalores positivos (supuesto que el rango de $\mathbf{H}$ es r y el rango de $\mathbf{X}$ es p), la región de confianza (1.26) corresponde a un elipsoide en el espacio r dimensional. En el caso particular de que sea $r=1$, dicho elipsoide se reduce a un segmento de la recta real, es decir a un intervalo de confianza. Este intervalo de confianza con nivel de confianza $1-\alpha$ puede expresarse también en la forma

$$\left|\widehat{\boldsymbol{\xi}}-\boldsymbol{\xi}\right| \leq s\, t(n-p, 1-\alpha/2)\sqrt{\mathbf{H}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{H}'}, \tag{1.27}$$

donde $t(n-p, 1-\alpha/2)$ es el cuantil $1-\alpha/2$ de la distribución t de Student con $n-p$ grados de libertad, y $\mathbf{H}$ es un vector fila de dimensión p.

Es interesante observar que si la matriz $\mathbf{H}$ se toma igual a la matriz identidad $\mathbf{I}_p$ de dimensión p, la ecuación (1.26) proporciona una región de confianza conjunta para los parámetros originales $\boldsymbol{\beta}$. Asimismo, si la matriz $\mathbf{H}$ se toma igual a la fila i de $\mathbf{I}_p$, la ecuación (1.27) proporciona un intervalo de confianza para el parámetro β_i.

1.11. Contrastes de hipótesis

Las regiones de confianza de la sección 1.10 pueden usarse para contrastar ciertas hipótesis importantes sobre las funciones lineales de los parámetros $\boldsymbol{\beta}$ incluidas en el vector $\boldsymbol{\xi}$. En particular, dado el vector $\boldsymbol{\xi}_0$ de posibles valores de

las funciones en el vector $\boldsymbol{\xi}$, la región de rechazo con riesgo de primera especie igual a α para la hipótesis nula $H_0 : \boldsymbol{\xi} = \boldsymbol{\xi}_0$ está dada por la desigualdad

$$\left(\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi}_0\right)' \left[\mathbf{H}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{H}'\right]^{-1} \left(\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi}_0\right) > rs^2 F(r, n-p, 1-\alpha). \qquad (1.28)$$

Es decir, la hipótesis nula $H_0 : \boldsymbol{\xi} = \boldsymbol{\xi}_0$ se rechaza con riesgo α si el punto $\boldsymbol{\xi}_0$ no pertenece a la región de confianza para $\boldsymbol{\xi}$ al nivel de confianza $1 - \alpha$.

La desigualdad (1.28) puede expresarse de una forma equivalente, que resulta a veces más cómoda de programar en un ordenador. Esta expresión alternativa es

$$S_{r0} - S_r > rs^2 F(r, n-p, 1-\alpha), \qquad (1.29)$$

donde $s^2 = S_r/(n-p)$, siendo S_r la suma de cuadrados residual del modelo (1.6) y S_{r0} la suma de cuadrados residual del modelo resultante de imponer en el modelo (1.6) la hipótesis nula $\mathbf{H}\boldsymbol{\beta} = \boldsymbol{\xi}_0$. La lógica de la expresión (1.29) es que, si al imponer esta restricción $\mathbf{H}\boldsymbol{\beta} = \boldsymbol{\xi}_0$, la suma de cuadrados residual (es decir, la magnitud de los residuos) del modelo (1.6) aumenta mucho, no es posible aceptar que la hipótesis $\mathbf{H}\boldsymbol{\beta} = \boldsymbol{\xi}_0$ sea cierta. Las expresiones (1.28) y (1.29) son equivalentes porque $S_{r0} - S_r = \left(\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi}_0\right)' \left[\mathbf{H}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{H}'\right]^{-1} \left(\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi}_0\right)$.

1.12. Modelos de regresión no lineal

Supongamos que en n individuos (o unidades experimentales) se observan simultáneamente los valores de una respuesta z_i y de q covariables $w_{1i}, \ldots, w_{qi}$, donde $i = 1, \ldots, n$. La forma general de un modelo de regresión no lineal es similar a la dada en (1.2), pero la media μ_i de la respuesta y_i es una función no lineal de los parámetros $\boldsymbol{\beta}$, es decir

$$\begin{aligned} y_i &= \mu_i + \epsilon_i \\ \mu_i &= G(\beta_1, \ldots, \beta_p; w_{1i}, \ldots, w_{qi}). \end{aligned} \qquad (1.30)$$

donde $y_i = h(z_i; w_{1i}, \ldots, w_{qi})$ es una función cualquiera, pero conocida, de la respuesta z_i en la que podrían intervenir una o más covariables $w_{1i}, \ldots, w_{qi}$; $G(\beta_1, \ldots, \beta_p; w_{1i}, \ldots, w_{qi})$ es una función cualquiera, pero conocida de los parámetros $\beta_1, \ldots, \beta_p$ y de las covariables $w_{1i}, \ldots, w_{qi}$; y $\epsilon_1, \ldots \epsilon_n$ son *errores aleatorios* de los que frecuentemente se supone que son variables aleatorias normales IIDs con media cero y varianza σ^2 desconocida.

La estimación de los parámetros puede hacerse usando los mismos métodos que en el caso lineal (mínimos cuadrados, máxima verosimilitud, máxima

bondad de ajuste, etc.). Las diferencias esenciales entre los modelos lineales y los no lineales pueden clasificarse en numéricas y estadísticas.

Desde el punto de vista numérico, el caso no lineal es más laborioso porque las ecuaciones (1.16) y (1.20) deben reemplazarse respectivamente por

$$\mathbf{e}'\mathbf{e} = \sum_{i=1}^{n} \left[y_i - G(\beta_1, \ldots, \beta_p; w_{1i}, \ldots, w_{qi}) \right]^2 \tag{1.31}$$

y (bajo hipótesis de normalidad e independencia de las respuestas)

$$\ell(\boldsymbol{\beta}, \sigma^2 | \mathbf{X}, \mathbf{Y}) = -\frac{n}{2} \ln \left(2\pi\sigma^2 \right) - \frac{\mathbf{e}'\mathbf{e}}{2\sigma^2} \, , \tag{1.32}$$

que no son funciones de segundo grado respecto de los parámetros $\boldsymbol{\beta}$. Por tanto, los estimadores no pueden obtenerse simplemente resolviendo un sistema lineal de ecuaciones, al contrario de los que ocurre en el caso lineal. En general, hay que minimizar (1.31) o maximizar (1.32) numéricamente.

Desde el punto de vista estadístico, la consecuencia de que los estimadores de los parámetros $\boldsymbol{\beta}$ sean funciones no lineales de las observaciones y_i es que la distribución de dichos estimadores en pequeñas muestras suele ser difícil de calcular y no normal. En muchos casos prácticos, estos estimadores sí son asintóticamente normales. Regiones o intervalos de confianza para funciones (lineales o no) de los parámetros, así como los contrastes de hipótesis sobre los valores de estas funciones, pueden obtenerse con métodos válidos asintóticamente, como por ejemplo, la prueba del cociente de verosimilitudes o las pruebas de Wald y de Rao-Silver.

1.12.1. Prueba del cociente de verosimilitudes

Supongamos que en un modelo de regresión no lineal se desea contrastar hipótesis sobre las funciones lineales en el vector $\boldsymbol{\xi} = \mathbf{H}\boldsymbol{\beta}$, del tipo de las consideradas en la sección 1.11 ($\mathbf{H}$ tiene rango r). Sea $\boldsymbol{\xi}_0$ un vector de posibles valores de $\boldsymbol{\xi}$. La región de rechazo con riesgo de primera especie igual a α para la hipótesis nula $H_0 : \boldsymbol{\xi} = \boldsymbol{\xi}_0$ está dada por la desigualdad

$$2 \left[\ell(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\sigma}}^2 | \mathbf{X}, \mathbf{Y}) - \ell(\widehat{\boldsymbol{\beta}}_0, \widehat{\boldsymbol{\sigma}}_0^2 | \; \boldsymbol{\xi} = \boldsymbol{\xi}_0, \mathbf{X}, \mathbf{Y}) \right] > \chi^2(r, 1-\alpha), \tag{1.33}$$

donde $\ell(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\sigma}}^2 | \mathbf{X}, \mathbf{Y})$ es el máximo de la función logaritmo-verosimilitud (LL) de los parámetros del modelo dada la muestra, $\ell(\widehat{\boldsymbol{\beta}}_0, \widehat{\boldsymbol{\sigma}}_0^2 | \; \boldsymbol{\xi} = \boldsymbol{\xi}_0, \mathbf{X}, \mathbf{Y})$ es el máximo de la misma función calculado en el conjunto de valores de los parámetros que satisfacen la restricción $\boldsymbol{\xi} = \boldsymbol{\xi}_0$, y $\chi^2(r, 1-\alpha)$ es el cuantil $1-\alpha$ de la distribución χ^2 de Pearson con r grados de libertad.

La desigualdad (1.33) sigue siendo válida aunque las restricciones impuestas por la hipótesis nula afecten a funciones no lineales de los parámetros $\boldsymbol{\beta}$. Es decir, aunque sea $H_0 : \mathbf{H}(\boldsymbol{\beta}) = \boldsymbol{\xi}_0$ con $H_i(\boldsymbol{\beta}) = H_i(\beta_1, \ldots, \beta_p)$, $i = 1, \ldots, r$, funciones independientes posiblemente no lineales.

Para un modelo de regresión no lineal, la desigualdad (1.33) es asintóticamente equivalente a

$$S_{r0} > S_r \left[1 + \frac{\chi^2(r, 1-\alpha)}{n-p}\right]. \tag{1.34}$$

La desigualdad $S_{r0} - S_r > rs^2 F(r, n-p, 1-\alpha)$ (dada en la fórmula (1.29)) sigue siendo válida cuando el modelo de regresión es no lineal y es asintóticamente equivalente a (1.33) y (1.34).

1.12.2. Perfil de la función logaritmo verosimilitud

Eligiendo $\mathbf{H}$ igual a la fila i de la matriz identidad $\mathbf{I}_p$, la desigualdad (1.33), proporciona un intervalo de confianza

$$2\left[\ell(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\sigma}}^2|\mathbf{X}, \mathbf{Y}) - \ell(\widehat{\boldsymbol{\beta}}_{-j}, \widehat{\boldsymbol{\sigma}}^2_{-j}|\ \beta_j = \beta_{j0}, \mathbf{X}, \mathbf{Y})\right] \leq \chi^2(1, 1-\alpha), \tag{1.35}$$

para el parámetro β_j al nivel de confianza $1-\alpha$. Obsérvese que en este caso, $\ell(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\sigma}}^2|\mathbf{X}, \mathbf{Y})$ es el máximo de la función LL respecto de todos los parámetros del modelo incluyendo a β_j, mientras que $\ell(\widehat{\boldsymbol{\beta}}_{-j}, \widehat{\boldsymbol{\sigma}}^2_{-j}|\ \beta_j = \beta_{j0}, \mathbf{X}, \mathbf{Y})$ es el máximo de la misma función LL respecto de todos los parámetros salvo β_j, que se mantiene fijo e igual a β_{j0}. El intervalo de confianza para β_j está formado por todos los valores β_{j0} que satisfacen la desigualdad (1.35).

Al máximo condicionado $\ell(\widehat{\boldsymbol{\beta}}_{-j}, \widehat{\boldsymbol{\sigma}}^2_{-j}|\ \beta_j = \beta_{j0}, \mathbf{X}, \mathbf{Y})$, considerado como función de β_{j0}, se le llama *perfil de la función LL (profile log-likelihood) para* β_j. Calcular la función perfil de LL para cada uno de los parámetros es laborioso numéricamente.

Para un modelo de regresión lineal, la desigualdad (1.35) es equivalente a la desigualdad (1.27).

Sin más que cambiar el signo $>$ por el signo $\leq$, la desigualdad (1.33) proporciona una región de confianza para las funciones del vector $\boldsymbol{\xi}$ al nivel de confianza $1-\alpha$. Por tanto, eligiendo como matriz $\mathbf{H}$ a cualquier subconjunto de r filas de $\mathbf{I}_p$, es inmediato generalizar la desigualdad (1.35) para obtener regiones de confianza para r parámetros del vector $\boldsymbol{\beta}$.

Bibliografía

Abramowitz, M. y Stegun, I. A. (1970). *Handbook of Mathematical Functions*. Dover Publications.

Box, G. E. P. y Jenkins, G. M. (1970). *Time Series Analysis, Forecasting and Control*. 2ª ed. Holden-Day.

Box, G. E. P., Jenkins, G. M. y Reinsel G. C. (1994). *Time Series Analysis, Forecasting and Control*. 3ª ed. Prentice-Hall.

Box, G. E. P., Hunter, W. y Hunter J. S. (1973, 2004). *Statistics for Experimenters*. Wiley.

Box, G. E. P. y Luceño, A. (1997). *Statistical Control by Monitoring and Feedback Adjustment*. Wiley.

Brockwell, P. J y Davis, R. A. (1991). *Time Series: Theory and Methods*, 2nd Ed. Springer-Verlag, NY.

Chang, I., Tiao, G.C. y Chen, C. (1988). Estimation of time series parameters in the presence of outliers. *Technometrics* 30, 193–204.

Chen C, Liu LM. (1993). Forecasting time series with outliers. *Journal of Forecasting* 12: 13–35.

Dickey, D. A. and Fuller, W. A. (1981) "Likelihood ratio statistics for autorregresive processes with a unit root". *Econometrica* 49, 1057–72.

Engle, R. F. (1982) "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflations". *Econometrica* 50, 987–1007.

Fox, A. J. (1972) Outliers in time series. *Journal of the Royal Statistical Society B* 34, 350–363.

Hannan, E. J. (1970). *Multiple Time Series*. Wiley. NY.

Hosking, J. R. M. (1981). "Fractional differencing". *Biometrika* 68: 165–176.

Johnson, R. A. (1997). *Probabilidad y Estadística para Ingenieros de Miller y Freund*. 5ª ed. Prentice-Hall Hispanoamericana, S. A. México, Nueva York.

Kalman, R. E. (1960). "A New approach to linear filtering and prediction problems". *Transactions of the ASME, Series D, Journal of Basic Enginee-*

ring 82. 35–45.

Kalman, R. E y Bucy, R. S. (1961). "New results in linear filtering and prediction theory". *Transactions of the ASME, Series D, Journal of Basic Engineering* 83. 95–107.

Ljung, G. M. y Box, G. E. P. (1978). "On a measure of lack of fit in time series models". *Biometrika* 65, 297–303.

Luceño, A. (1996). "A fast likelihood approximation for vector general linear processes with long series: Application to fractional differencing". *Biometrika* 83. 603–614.

Luceño A. (1998). Detecting possibly non-consecutive outliers in industrial time series. *Journal of the Royal Statistical Society B* 60, 295–310.

Luceño, A. (2006). "Fitting the generalized Pareto distribution to data using maximum goodness of fit estimators". *Computational Statistics & Data Analysis*, Vol. 51, No. 2, 904-917. doi:10.1016/j.csda.2005.09.011.

Luceño, A. y González, F. J. (2003). *Métodos Estadísticos para Medir, Describir y Controlar la Variabilidad*. Universidad de Cantabria. (I.S.B.N. 84-607-6773-6).

Luceño, A. y Peña, D. (2007). "ARIMA modelling". *Encyclopedia of Statistics in Quality and Reliability*, Wiley, New York.

Lütkepohl, H. (1993). *Introduction to Multiple Time Series Analysis*. Springer-Verlag, NY.

Mosteller, F. y Tukey, J. W. (1977). *Data Analysis and Regression: a second course in statistics*. Addison-Wesley Publishing Company.

Peña, D. y Rodríguez, J. (2006) "The log of the determinant of the autocorrelation matrix for testing goodness of fit in time series". *Journal of Statistical Planning and Inference* 136, 2706–2708.

Peña, D., Tiao, G.C. y Tsay R. S. (2001). *A Course in Time Series Analysis* Wiley, NY.

Rao, C. R. (1973). *Linear Statistical Inference and its Applications*. Wiley.

Reinsel, G. C. (1993). *Elements of Multivariate Time Series Analysis*. Springer-Verlag, NY.

Shumway, R.H. y Stoffer, D. S. (2000). *Time Series Analysis and its Applications*. Springer-Verlag, NY.

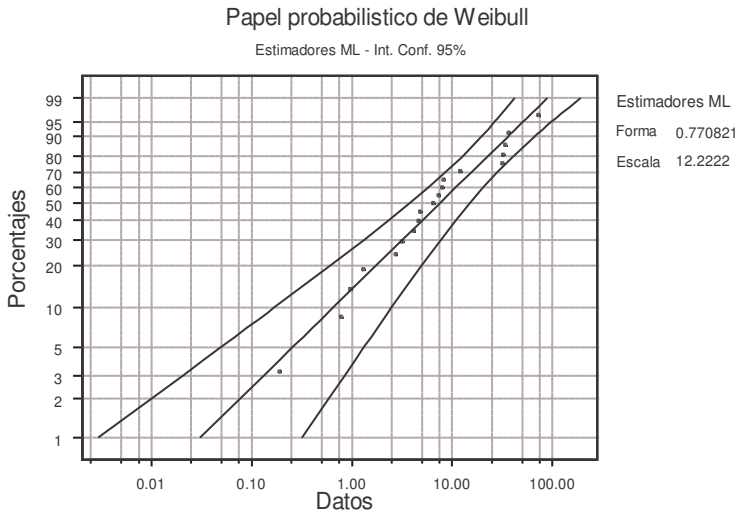
Tong, H. (1990) *Non-linear Time Series: A Dynamical System Approach*. Oxford University Press, Oxford.

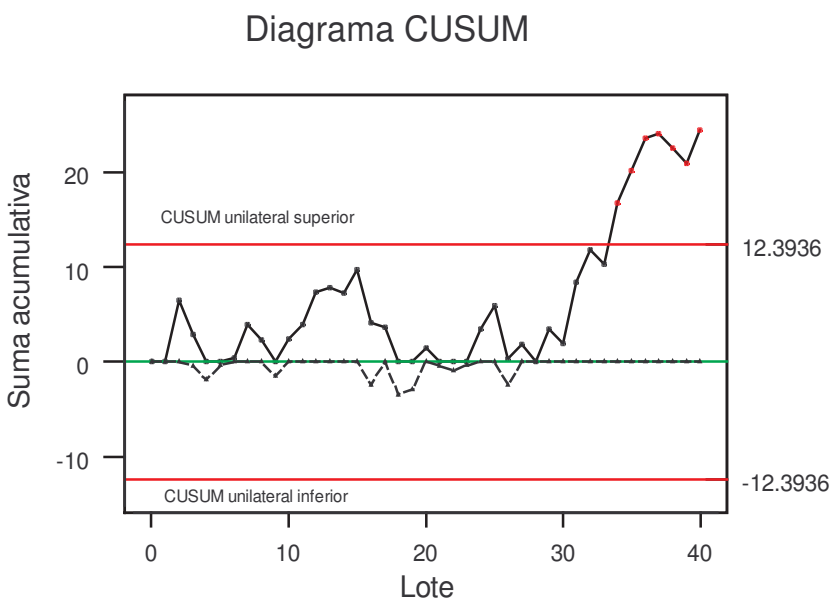
Wei, W. W. S. (1990). *Time Series Analysis*. Addison-Wesley, Redwood City, CA.

Wold, H. O. (1938). *A Study in the Analysis of Stationary Time Series*. Almquist and Wicksell. Uppsala.

Wu L.S.-Y., Ravishanker, N. y Hosking, J. R. M. (1993). Reallocation outliers in time series. *Applied Statistics* 42, 301–313.

Modelos de Regresión





ALBERTO LUCEÑO