

EL MODELO LINEAL

- DEFINICIÓN DE MODELO LINEAL
- UTILIDAD
- ESPECIFICACIÓN DEL MODELO
- HIPOTESIS BÁSICAS DEL MODELO LINEAL
- ESTIMACIÓN DEL MODELO Y PREDICCIÓN
- DISTRIBUCIÓN DE LOS ESTIMADORES DE LOS PARAMETROS DEL MODELO LINEAL
- INFERENCIA SOBRE EL MODELO LINEAL : Inferencia sobre un parámetro. Contraste de hipótesis sobre un regresor. Contrastes lineales sobre un conjunto de regresores. Contraste de significación de los regresores. Contraste de validez del modelo/ significación general/ Anova
- PREDICCIÓN DE UN VALOR EXTRAMUESTRAL

Complementos

- (Notación matricial)(Formas lineales y cuadráticas:-derivación)
- Cuestiones previas sobre la regresión mínimo cuadrática
- Teorema de Cochran

MODELO LINEAL

Puede definirse como un esquema de relación entre una variable Y (EXÓGENA O A EXPLICAR) y otra(s) variable(s) X ($X_1 X_2 \dots X_k$) (endógena(s) o explicativa(s), tal que:

$$Y = F.\text{LINEAL}(X) + \text{PERTURBACIÓN ALEATORIA}$$

(Modelo Lineal Simple)

$$Y = F.\text{LINEAL}(X_1 X_2 \dots X_k) + \text{PERTURBACIÓN ALEATORIA}$$

(Modelo Lineal General)

Las hipótesis (básicas) que se asuman sobre la perturbación aleatoria permitirán realizar el análisis estadístico inferencial

Las razones para la introducción de una perturbación aleatoria, son fundamentalmente:

- 1. efecto de variables no consideradas
- 2. efectos imprevistos (catastrofes, modas, etc.)
- 3. errores de observación o medición.

UTILIDADES DEL MODELO LINEAL:

1. verificar la existencia de la relación lineal.
2. estimar (contrastar) la (una) relación lineal concreta (estructural).- supone actuar sobre los coeficientes de la relación lineal.
3. predecir la variable y en función de x o ($x_1 x_2 \dots x_k$)

ESPECIFICACIÓN DEL MODELO

NOTACIONES ALTERNATIVAS:

La información recogida hace referencia a n periodos de tiempo, o n localizaciones espaciales. Entonces, la relación

(teórica) según el modelo entre las variables sería:

$$\text{M.L.S.: } y_i = a + b x_i + e_i \text{ para } i= 1,2,3,\dots, n$$

$$\text{M.L.G.: } y_i = b_0 + b_1 x_{1i} + b_2 x_{2i} + \dots + b_k x_{ki} + e_i \text{ para } i= 1,2,\dots,n$$

puede expresarse también matricialmente:

$$y = Xb + e \text{ donde:}$$

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ y_n \end{pmatrix} \quad X = \begin{pmatrix} 1 & x_{11} & x_{21} & \dots & x_{k1} \\ 1 & x_{12} & x_{22} & \dots & x_{k2} \\ \cdot & \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \cdot & \dots & \cdot \\ 1 & x_{1n} & x_{2n} & \dots & x_{kn} \end{pmatrix} \quad \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \cdot \\ \beta_k \end{pmatrix} \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdot \\ \varepsilon_n \end{pmatrix}$$

Y EN EL Modelo lineal simple

$$X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \cdot & \cdot \\ 1 & x_n \end{pmatrix} \quad \beta = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$$

HIPÓTESIS BÁSICAS DEL MODELO LINEAL

Sobre el modelo lineal:

M.L.S.:

$$y_i = a + b x_i + e_i \text{ para } i= 1,2,3,\dots, n$$

M.L.G.:

$$y_i = b_0 + b_1 x_{1i} + b_2 x_{2i} + \dots + b_k x_{ki} + e_i \text{ para } i= 1,2,\dots,n$$

que también pueden expresarse matricialmente como: $y = Xb + e$

que establece la relación teórica entre las n observaciones de variable exógena y las de la(s) endógena(s).

Consideramos además las siguientes hipótesis básicas:

1ª " e_i sigue una $N(0, s)$ las perturbaciones aleatorias se distribuyen normalmente con media cero y desviación típica (varianza) constante (homoscedasticidad) en todos los periodos o localizaciones considerados.

2ª " $i \neq j$ $D(e_i, e_j) = 0$ Las n perturbaciones aleatorias están incorrelacionadas (al ser normales, son independientes).

(hipótesis de no autocorrelación)

Como consecuencia de estas dos hipótesis: \mathbf{e} sigue una $N_n(\mathbf{0}; s^2 \mathbf{I})$

3ª La variable x (las variables x_1, x_2, \dots, x_k) es (son) de naturaleza no aleatoria. Además sus valores son independientes de la perturbación aleatoria. Y en el caso del M.L.G. las variables x_1, x_2, \dots, x_k están incorrelacionadas (hipótesis de ausencia de multicolinealidad)

Como consecuencia de estas hipótesis tendremos que el comportamiento teórico de las observaciones de y deberá ser tal que:

y_i sigue una $N(a + bx_i; s)$ con $D(y_i y_j) = 0 \quad i \neq j$ en el M.L.S.

y_i sigue una $N(b_0 + b_1 x_{1i} + \dots + b_k x_{ki}; s)$ con $D(y_i y_j) = 0 \quad i \neq j$ en el M.L.G.

y por lo tanto, para cualquiera de los dos casos la distribución conjunta del vector de observaciones: \mathbf{y} sigue una $N_n(\mathbf{Xb}; s^2 \mathbf{I})$

A partir de estas hipótesis podremos considerar que los valores observados de la variable y son "como si fueran" datos muestrales de una población normal de las características señaladas arriba y podremos plantearnos realizar inferencias sobre los parámetros $(a, b \text{ y } s)$ en el M.L.S. o $(b_0, b_1, \dots, b_k \text{ y } s)$ en el caso del M.L.G.)

ESTIMACIÓN DEL MODELO Y PREDICCIÓN PUNTUAL

ir a modelo lineal

Nos plantemos, en primer lugar construir un estimador del modelo (esto es, del valor teórico de y_i) con la pretensión fundamental de obtener a partir del modelo estimado predictores puntuales de y para situaciones no observadas (supuestos ciertos valores de x).

Vamos buscando un estimador del tipo:

$$\hat{y}_i = \hat{\alpha} + \hat{\beta} x_i \quad (\text{M.L.S.})$$

$$\hat{\mathbf{y}} = \mathbf{X} \hat{\boldsymbol{\beta}} \quad (\text{M.L.G.})$$

Empleando el método de máxima verosimilitud para obtener los E.M.V. de a, b, s en M.L.S. tendremos que:

La función de densidad de cada dato y_i será:

$$f(y_i) = \frac{1}{\sigma \sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2} (y_i - \alpha - \beta x_i)^2\right)$$

y por tanto la f. de verosimilitud para el conjunto de los n datos :

$$L(y_1, y_2, \dots, y_n / \alpha, \beta, \sigma) = (2\pi)^{-n/2} \sigma^{-n} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \alpha - \beta x_i)^2\right]$$

Tomando logaritmos e igualando a cero las derivadas parciales con respecto a a, b, s obtendríamos los tres E.M.V:

$$\hat{\alpha} = \alpha = \bar{y} - \frac{S_{xy}}{S_x^2} \bar{x} = \bar{y} - b \bar{x} \quad \hat{\beta} = b = \frac{S_{xy}}{S_x^2} \quad \hat{\sigma}^2 = S_y^2 = \frac{\sum_{i=1}^n (y_i - \alpha - b x_i)^2}{n}$$

Es decir, se obtiene como resultado de la estimación máximo verosímil que los estimadores de los parámetros son los coeficientes de la regresión mínimo cuadrática y el estimador de la varianza de la perturbación, la varianza residual muestral.

Quedando el modelo estimado como:

$$\hat{y}_i = a + bx_i$$

y el predictor para un periodo extramuestral: $\hat{y}_{n+r} = a + bx_{n+r}$

Puede probarse que el E.M.V. del vector de parámetros b del M.L.G. también coincide con el vector obtenido por ajuste mínimo cuadrático:

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_k \end{pmatrix} = b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_k \end{pmatrix} = (X'X)^{-1} X'y$$

Siendo también la varianza residual muestral el E.M.V. de la varianza de la perturbación

DISTRIBUCIONES DE LOS ESTIMADORES

Con el objeto de poder realizar inferencias (estimaciones por intervalo y contrastes de hipótesis) sobre los parámetros analizaremos las distribuciones que tienen los estimadores obtenidos (E.M.V coincidentes con los E.M.C.O.)

Consideramos los estimadores del M.L.G. que luego particularizaremos para el M.L.S.. Llamemos al estimador del vector (parámetros) b ; b .

El vector b de estimadores (M.C./M.V.) es como sabemos:

$$b = (X'X)^{-1} X'y$$

Esto es, será una función lineal del vector aleatorio n -dimensional y . Como el vector y sigue una $N_n (Xb ; s^2 I)$ la distribución de b será normal y como es un vector formado por $k+1$ (número de variables + 1) estimadores: la distribución de b será normal $(k+1)$ -dimensional.

Su vector de medias vendrá dado por:

$$E(b) = E[(X'X)^{-1} X'y] = (X'X)^{-1} X' E(y) =$$

$= (X'X)^{-1} X' E(Xb + e) = b + E(e) = b$ de modo que los estimadores de los coeficientes son INSESGADOS

Su matriz de varianzas vendrá dada por:

$$D^2(b) = D^2((X'X)^{-1} X'y) = D^2[(X'X)^{-1} X'(Xb + e)] = D^2[(X'X)^{-1} X'X b + (X'X)^{-1} X' e] =$$

$$D^2[(X'X)^{-1} X' e] = (X'X)^{-1} X' D^2(e) [(X'X)^{-1} X']' = (X'X)^{-1} X' . D^2(e) . X(X'X)^{-1} =$$

$$= s^2 . [(X'X)^{-1} X' I X(X'X)^{-1}] = s^2 . (X'X)^{-1}$$

de forma que $b \sim N_{k+1} [b ; s^2 (X'X)^{-1}]$

En el caso del M.L.S. el vector de estimadores era, como sabemos:

$$b = \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \bar{y} - \frac{S_{xy}}{S_x^2} \bar{x} \\ \frac{S_{xy}}{S_x^2} \end{pmatrix}$$

y la matriz $(X'X)^{-1}$ era:

$$(X'X)^{-1} = \frac{1}{nS_x^2} \begin{pmatrix} a_2(x) & -\bar{x} \\ -\bar{x} & 1 \end{pmatrix}$$

De modo que la distribución conjunta de los estimadores a,b será:

$$b = \begin{pmatrix} a \\ b \end{pmatrix} \rightarrow N_2 \left[\begin{pmatrix} \alpha \\ \beta \end{pmatrix}, \frac{\sigma^2}{nS_x^2} \begin{pmatrix} a_2(x) & -\bar{x} \\ -\bar{x} & 1 \end{pmatrix} \right]$$

De donde obtenemos que :

$$a \rightarrow N\left(\alpha, \frac{\sigma}{\sqrt{n}} \sqrt{1 + \frac{\bar{x}}{S_x^2}}\right)$$

$$b \rightarrow N\left(\beta, \frac{\sigma}{S_x \sqrt{n}}\right)$$

Siendo la covarianza entre ambos estimadores:

$$\text{cov}(a,b) = \sigma^2 \frac{\bar{x}}{nS_x^2}$$

Para estudiar la distribución del estimador de s^2 , la varianza residual que en el caso general para k variables (M.L.G.) tomaba la expresión:

$$S_r^2 = 1/n \cdot \mathbf{S} (y_i - b_0 - b_1 x_{i1} - \dots - b_k x_{ik})^2 =$$

$$= 1/n \cdot (y - Xb)'(y - Xb)$$

Consideraremos una variable derivada de este estadístico la variable:

$$nS_r^2 / s^2 = (1/s^2) \mathbf{S} (y_i - b_0 - b_1 x_{i1} - \dots - b_k x_{ik})^2 = (1/s^2) \cdot (y - Xb)'(y - Xb)$$

Tengamos en cuenta que el vector de residuos muestrales $e = y - Xb$ es un vector n-dimensional que se obtiene a través de una transformación lineal de dos vectores normales (y y b): Por tanto su distribución será normal.

No es difícil probar que tiene por vector de medias el vector 0 y por matriz de varianzas la matriz $M = [I - X(X'X)^{-1}X']$ que tiene por rango k+1 y la propiedad de ser idempotente $MM = M'M = M$.

Tampoco es costoso ver que la relación entre el vector de residuos muestrales y el de perturbaciones aleatorias es: $e = Me$

De esta manera que el estadístico nS_r^2 / s^2 puede verse como:

$$nS_r^2 / s^2 = (1/s^2) \cdot (y - Xb)'(y - Xb) = (1/s^2) \cdot e'e = (1/s^2) e'M'e = (1/s^2) e'Me = (1/s^2) (e'e) - (1/s^2) (e'X(X'X)^{-1}X'e)$$

$X'e$ es un vector aleatorio $k+1$ dimensional tal que $X'e \sim N(0, s^2(X'X))$

de forma que, aplicando el teorema de Cochran:

$1/s^2 (e'e)$ sigue una χ^2 con n grados de libertad

y $1/s^2(e'X(X'X)^{-1}X'e)$ sigue una χ^2 con $(k+1)$ grados de libertad

y a partir de este resultado es fácil probar que:

nS_r^2 / s^2 sigue una χ^2 con $(n - k - 1)$ grados de libertad

(En particular para el caso del M.L.S. $k=1$, de forma que:

nS_r^2 / s^2 sigue una χ^2 con $(n - 2)$ grados de libertad)

Este último resultado es, no sólo importante en sí mismo sino que también tiene efectos sobre las inferencias de los estimadores de b .

En efecto, como sabemos b sigue $N_{k+1} [b ; s^2 (X'X)^{-1}]$

de modo que la distribución marginal de cada estimador de cada parámetro b_j , b_j será:

b_j sigue $N [b_j ; s a_{jj}]$

donde a_{jj} es la raíz cuadrada de el elemento $j-j$ de la matriz $(X'X)^{-1}$

Como el parámetro s es desconocido es conveniente encontrar una distribución que no dependa de s .

Es fácil ver que el estadístico:

$$\frac{b_j - \beta_j}{a_{jj} \sqrt{nS_r^2}} \sqrt{n - k - 1} \rightarrow t_{n-k-1}$$

(Al valor que está dividiendo a la diferencia $(b_j - \beta_j)$ se le llama error standard del estimador y se le suele representar por S_{b_j})

Lo que para el M.L.S. supone que:

$$\frac{a - \alpha}{S_r \sqrt{1 + \frac{\bar{x}}{S_x^2}}} \sqrt{n-2} \rightarrow t_{n-2}$$

$$\frac{b - \beta}{\frac{S_r}{S_x}} \sqrt{n-2} \rightarrow t_{n-2}$$

INFERENCIAS SOBRE EL MODELO LINEAL

Inferencia sobre un parámetro.

Contraste de hipótesis sobre un regresor.

Contrastes lineales sobre un conjunto de regresores.

Contraste de significación de los regresores.

Contraste de validez del modelo/ significación general/ Anova

Inferencias sobre un parámetro

Teniendo en cuenta los resultados anteriores resulta sencillo diseñar los métodos para la realización de inferencias sobre uno de los "regresores", b_j .

Intervalo de confianza (1-a) para b_j :

$$\text{teniendo en cuenta que: } \frac{b_j - \beta_j}{\frac{S_r}{S_x} \sqrt{n-2}} \rightarrow t_{n-k-1}$$

si llamamos S_{b_j} al error standard del estimador ,esto es, a todo lo que divide a $(b_j - \beta_j)$ en la expresión anterior; el intervalo de confianza para b_j resulta (para un N.C. de 1-a):

$b_j \pm [t_{a/2} S_{b_j}]$ con (1-a) de confianza

Para el caso particular del M.L.S. los intervalos de confianza para los coeficientes a y b serán respectivamente:

$$\left[a \pm t_{a/2} \frac{S_r \sqrt{1 + \frac{\bar{x}}{S_x^2}}}{\sqrt{n-2}} \right]$$

$$\left[b \pm t_{a/2} \frac{S_r}{S_x \sqrt{n-2}} \right]$$

Ejemplo: En el ejemplo visto con anterioridad: El regresor para la primera variable explicativa (precio por kilo) es $b_2 = -1.44765$. La raíz cuadrada del elemento 2,2 de $(X'X)^{-1}$, $a_{22} = (0.000746)^{1/2} = 0.027313$, la "cuasi-desviación típica residual": $S_r (n/n-k-1)^{1/2} = 8.451406$, de forma que el error standard del estimador es: $S_{b_2} = 0.230912$

Si queremos construir un intervalo de confianza para un nivel del 99%: $t_{0,005} (12-2-1 \text{ g.l.}) = 3.25$ de forma que el intervalo quedará:

$$b_2 \hat{I} [-1.44765 - 3.25 \cdot 0.230912 ; 1.44765 + 3.25 \cdot 0.230912]$$

$$b_2 \hat{I} [-2.198114; 0.697186]$$

Contrastes de hipótesis sobre un regresor.

Basándonos en la distribución de $(b_j - b_j) / S_{b_j}$ podremos igualmente contrastar hipótesis sobre el regresor b_j

Si la hipótesis nula es: $H_0: b_j = b_j^*$ ya sea la alternativa uni o bilateral, el estadístico $(b_j - b_j^*) / S_{b_j}$, supuesta cierta la hipótesis nula tendrá una distribución t de Student (con $n - k - 1$ g.l.) y podremos diseñar el contraste de la manera habitual que ya conocemos.

ejemplo

Contrastes lineales sobre un conjunto de regresores.

En muchos casos prácticos podemos estar interesado en contrastar una hipótesis no ya sobre el valor que toma un regresor, sino sobre cómo se comportan un conjunto de varios regresores. Si la hipótesis a contrastar consiste en que un conjunto de regresores verifican una cierta relación lineal esto puede llevarse a cabo a partir del procedimiento general que exponemos a continuación:

Una o varias relaciones lineales entre un conjunto de regresores puede expresarse de forma general de la siguiente manera:

$$D\mathbf{b} = \mathbf{h}$$

Donde el vector $(k+1)$ dimensional \mathbf{b} es el vector de parámetros (regresores) del modelo; \mathbf{h} es un vector r -dimensional, donde r es el número de relaciones lineales que estamos considerando; y, por último la matriz \mathbf{D} de dimensión $(r \times k+1)$ es una matriz de coeficientes.

Así por ejemplo si queremos expresar sobre un modelo con 4 regresores (b_0, b_1, b_2, b_3) las siguientes 2 relaciones lineales:

$b_1 + b_2 = 1$ y $b_1 - 2b_3 = 0$ podrá hacerse como:

$$\begin{pmatrix} 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & -2 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

Teniendo en cuenta esto es fácil ver que la hipótesis nula de que un conjunto de regresores verifican una o más relaciones lineales puede expresarse como:

$$H_0: D\mathbf{b} = \mathbf{h}$$

Es igualmente sencillo probar que si pre-multiplicamos el vector de estimadores \mathbf{b} por la matriz \mathbf{D} el resultado será un vector aleatorio

r -dimensional cuya distribución será normal ya que es una transformación lineal (forma lineal) del vector \mathbf{b} que seguía una distribución normal:

$$\mathbf{b} \rightarrow N_{k+1}[\mathbf{b}; s^2(\mathbf{X}'\mathbf{X})^{-1}]$$

El vector de medias del vector \mathbf{Db} será $E(\mathbf{Db}) = \mathbf{D}E(\mathbf{b}) = \mathbf{D}\mathbf{b} = \mathbf{h}$ (siempre que la hipótesis sea cierta).

y la matriz de varianzas será

$$\text{Var}(\mathbf{Db}) = \mathbf{D} \cdot \text{Var}(\mathbf{b}) \cdot \mathbf{D}' = s^2 \mathbf{D}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{D}'$$

Así pues, si la hipótesis nula es cierta: $\mathbf{Db} \rightarrow N_r[\mathbf{h}, s^2 \mathbf{D}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{D}']$

Puede apreciarse que la distribución de este estadístico vectorial depende del parámetro desconocido s . Para evitar este problema puede actuarse de la siguiente manera:

Aplicando el teorema de Cochran es inmediato que la forma cuadrática:

$$1/s^2 \cdot (\mathbf{Db} - \mathbf{h})' [\mathbf{D}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{D}']^{-1} (\mathbf{Db} - \mathbf{h}) \text{ sigue una } \mathcal{C}^2 \text{ con } r \text{ grados de libertad}$$

Si consideramos que $1/s^2 \cdot \mathbf{e}'\mathbf{e}$ sigue una \mathcal{C}^2 con $n-k-1$ g.l.

y asumiendo la independencia entre ambas variables aleatorias:

el estadístico:

$$\frac{(\mathbf{Db} - \mathbf{h})' [\mathbf{D}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{D}']^{-1} (\mathbf{Db} - \mathbf{h})}{\frac{\mathbf{e}'\mathbf{e}}{n-k-1}} \rightarrow F_{r, n-k-1}$$

Bajo el supuesto de que la hipótesis nula es cierta.

Así pues evaluando este estadístico y comparandolo con el valor crítico correspondiente para el nivel de significación requerido:

si $F > F_\alpha$ rechazaremos la hipótesis que supone la presencia de esa relación lineal, y en caso contrario, la aceptaremos

ejemplo

Contrastes de significación de los regresores

Un tipo particularmente importante de contrastes de hipótesis sobre los regresores (ya sea sobre 1 o sobre varios de ellos) es el contraste de la hipótesis nula de que el (o los) regresor(es) considerado(s) son cero frente a la alternativa bilateral de que son distintos de cero. En estos casos, si llegamos a rechazar la hipótesis nula quedará establecido que el regresor es significativo queriendo decirse que existirá una relación lineal significativa entre la variable a explicar y la(s) variable(s) explicativa asociada al regresor (o regresores) sujeto(s) a debate.

Suele llamarse contraste de significación a este tipo de contraste. Cuando se trata de un contraste individual se llevará a cabo mediante un contraste t de Student bilateral.

Una medida de la significación o significatividad de un parámetro (regresor) suele darse a través del valor del nivel de significación necesario para rechazar la hipótesis de nulidad; de forma que cuanto más pequeño sea este valor más significativo es el regresor.

Un caso particular importante de los contrastes de significación es aquel en el que se consideran (a la vez) todos los regresores (excepción hecha del término independiente). Este contraste constituye una prueba de la validez global del modelo y recibe el nombre de "contraste de significación general", de validez general (y, también de análisis de la varianza de la regresión)

Contraste de validez general (significación general)(ANOVA)

Como acabamos de comentar se trata de analizar la significación de todos los regresores a la vez, lo que equivale a analizar la validez general del modelo, la fuerza de la relación lineal existente entre la variable y y las variables x, la significación de la correlación lineal múltiple; todo ello es equivalente.

Como también hemos comentado se trata, en el fondo de contrastar la hipótesis:

$$H_0: b_1=b_2=b_3=\dots=b_k=0$$

Podemos plantearlo, exactamente igual que el contraste general de hipótesis lineales sobre un conjunto de regresores.

$$H_0: \mathbf{D}\mathbf{b} = \mathbf{h}$$

Aquí la matriz \mathbf{D} sería la matriz identidad de orden k (con 1 elemento menos que el número total de parámetros k+1) y el vector h sería el vector 0 de orden k.

Pudiéndose expresar la hipótesis como:

$$H_0: \mathbf{I}_k \mathbf{b} = \mathbf{0}_k$$

y si llamamos $\underline{\mathbf{b}}$ al vector de orden k formado por los regresores excluyendo el término independiente podemos expresar la hipótesis como: (igualmente a su estimador lo llamamos $\underline{\mathbf{b}}$ y llamamos $\underline{\mathbf{X}}$ a la matriz de los datos de x excluida la primera columna de unos)

$$H_0: \underline{\mathbf{b}} = \mathbf{0}$$

Al ser la matriz \mathbf{D} la identidad y el vector \mathbf{h} el vector cero, el estadístico del contraste quedaría:

$$\frac{\frac{\underline{\mathbf{b}}' (\underline{\mathbf{X}}' \underline{\mathbf{X}}) \underline{\mathbf{b}}}{k}}{\frac{\mathbf{e}' \mathbf{e}}{n-k-1}} \rightarrow F_{k, n-k-1}$$

Si tenemos en cuenta que la regresión muestral era: $Xb=y^*$ y que pasaba por el centro de gravedad, entonces: $b'X'Xb$ será n veces la varianza debida a la regresión:

$$b'X'Xb = n S_y^2 R^2 \text{ . (Donde } R^2 \text{ es el coeficiente de determinación muestral)}$$

Por otro lado: $e'e$ es n veces la varianza residual:

$$e'e = n S_r^2 = n S_y^2 (1 - R^2).$$

A partir de estos dos hechos es muy fácil ver que el estadístico del contraste queda como:

$$\frac{\frac{n S_y^2 R^2}{k}}{\frac{n S_y^2 (1 - R^2)}{n - k - 1}} = \frac{\frac{R^2}{k}}{\frac{(1 - R^2)}{n - k - 1}}$$

Que bajo el supuesto de la hipótesis nula seguirá una distribución $F_{k, n-k-1}$

Es interesante observar que en el caso del M.L.S. , con $k=1$, y ante la coincidencia del coeficiente de determinación y el cuadrado del coeficiente de correlación, nos encontramos con el estadístico del contraste de incorrelación, cuya distribución no probamos y aquí haya una demostración.

El planteamiento de este contraste admite un esquema ANOVA, en el siguiente sentido: La variación total (muestral) de la y ($n S_y^2$) puede descomponerse en variación debida a la regresión

($b'X'Xb = n S_y^2 R^2$) más la variación residual o no explicada

($e'e = n S_r^2 = n S_y^2 (1 - R^2)$).

PREDICCIÓN DE UN VALOR EXTRAMUESTRAL.-PREDICCIÓN POR INTERVALO

Ya vimos que la predicción puntual de un valor de y extramuestral se llevaba a cabo utilizando como predictor el modelo estimado aplicado sobre los valores futuros de las variables explicativas.

De manera general podríamos expresarlo de la siguiente forma:

$$\hat{y}_0 = x'_0 b$$

Donde x'_0 es el vector fila formado por los valores futuros de las variables explicativas: $x'_0 = (1 \ x_{10} \ x_{20} \ \dots \ x_{k0})$

Este predictor es una transformación lineal del vector aleatorio b de los estimadores del vector de regresores b .Ademas la dimensión del predictor es 1, por lo que será una variable aleatoria Normal.

Ademas su media será: $E(x'_0 b) = x'_0 E(b) = x'_0 b$ esto es, el valor teórico esperado en el futuro para la variable y (el predictor es insesgado).

Por otro lado su varianza será:

$$\text{Var}(x'_0 b) = x'_0 \text{Var}(b) x_0 = s^2 x'_0 (X'X)^{-1} x_0$$

Así pues la distribución del predictor es:

$$\hat{y}_0 = x_0' b \rightarrow N[x_0' \beta; \sigma \sqrt{x_0' (X' X)^{-1} x_0}]$$

Distribución que depende del parámetro desconocido σ . Este problema puede solventarse construyendo el estadístico

$$\frac{(\hat{y}_0 - x_0' \beta)' [x_0' (X' X)^{-1} x_0]^{-1} (\hat{y}_0 - x_0' \beta)}{\frac{e' e}{n-k-1}}$$

Que tendrá una distribución F de Snedecor con 1 y n-k-1 g.l..

Como tanto $x_0' (X' X)^{-1} x_0$; como $x_0' b$; como el predictor son escalares podemos expresar este estadístico como:

$$\frac{(\hat{y}_0 - x_0' \beta)^2}{\frac{e' e}{n-k-1} [x_0' (X' X)^{-1} x_0]}$$

Y como la raíz cuadrada de una variable $F_{1, n-k-1}$ es una variable t de Student con n-k-1 g.l., tendremos que:

$$\sqrt{\frac{(\hat{y}_0 - x_0' \beta)^2}{\frac{e' e}{n-k-1} [x_0' (X' X)^{-1} x_0]}} \rightarrow t_{n-k-1}$$

A partir de este estadístico es sencillo construir un intervalo para la predicción esperada según el modelo, dado un nivel de confianza prefijado (1 - α):

$$x_0' \beta \in \left(\hat{y}_0 \pm t_{\alpha/2} \sqrt{\frac{e' e}{n-k-1} [x_0' (X' X)^{-1} x_0]} \right)$$

Donde $t_{\alpha/2}$ es el correspondiente valor tabulado para n-k-1 g.l.

Es interesante ver cómo quedaría el intervalo de predicción en el caso de un M.L.S.:

Tras realizar las operaciones pertinentes acaba quedando un intervalo para el valor futuro de y (teórico según el modelo):

$$\left(\hat{y}_0 \pm t_{\alpha/2} \frac{s_r \sqrt{1 + \frac{(x_0 - \bar{x})^2}{s_x^2}}}{\sqrt{n-2}} \right)$$

Donde $t_{\alpha/2}$ es el valor tabulado para $n-2$ grados de libertad.

COMPLEMENTOS

ESPECIFICACIÓN DEL MODELO

NOTACIONES ALTERNATIVAS:

La información recogida hace referencia a n periodos de tiempo, o n localizaciones espaciales. Entonces, la relación (teórica) según el modelo entre las variables sería:

$$\text{M.L.S.: } y_i = a + b x_i + e_i \text{ para } i = 1, 2, 3, \dots, n$$

$$\text{M.L.G.: } y_i = b_0 + b_1 x_{1i} + b_2 x_{2i} + \dots + b_k x_{ki} + e_i \text{ para } i = 1, 2, \dots, n$$

puede expresarse también matricialmente:

$$y = Xb + e \text{ donde:}$$

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ y_n \end{pmatrix} \quad X = \begin{pmatrix} 1 & x_{11} & x_{21} & \dots & x_{k1} \\ 1 & x_{12} & x_{22} & \dots & x_{k2} \\ \cdot & \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \cdot & \dots & \cdot \\ 1 & x_{1n} & x_{2n} & \dots & x_{kn} \end{pmatrix} \quad \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \cdot \\ \beta_k \end{pmatrix} \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdot \\ \varepsilon_n \end{pmatrix}$$

Y EN EL Modelo lineal simple

$$X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \cdot & \cdot \\ 1 & x_n \end{pmatrix} \quad \beta = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$$

Formas lineales y cuadráticas ; su derivación.

FORMA LINEAL:

sea \mathbf{a} un vector columna de constantes de \mathbb{R}^n

sea \mathbf{X} un vector columna de variables de \mathbb{R}^n

una forma lineal es entonces $f(\mathbf{X}) = \mathbf{X}' \mathbf{a} = \mathbf{a}' \mathbf{X} = \sum a_i x_i$

FORMA CUADRÁTICA:

Si \mathbf{A} es una matriz cuadrada de orden n

una forma cuadrática es entonces:

$$Q(\mathbf{X}) = \mathbf{X}' \mathbf{A} \mathbf{X} = \sum \sum a_{ij} x_i x_j$$

Tanto f como Q son funciones que hacen corresponder vectores de \mathbb{R}^n con valores (escalares) de \mathbb{R} .

DERIVACIÓN VECTORIAL DE F. LINEALES Y CUADRÁTICAS:

puede probarse que la derivada vectorial de una f . lineal es:

$$\frac{\partial \mathbf{a}' \mathbf{x}}{\partial \mathbf{x}} = \frac{\partial \mathbf{x}' \mathbf{a}}{\partial \mathbf{x}} = \mathbf{a}$$

y que la derivada vectorial de una forma cuadrática es:

$$\frac{\partial \mathbf{x}' \mathbf{A} \mathbf{x}}{\partial \mathbf{x}} = (\mathbf{A} + \mathbf{A}') \mathbf{x}$$

EN EL CASO DE QUE \mathbf{A} SEA UNA MATRIZ SIMÉTRICA:

$$(\mathbf{A}' + \mathbf{A}) = 2\mathbf{A} \text{ Y LA DERIVADA RESULTA SER: } 2\mathbf{A}\mathbf{x}$$

CUESTIONES PREVIAS SOBRE LA REGRESIÓN (MUESTRAL) MÍNIMO-CUADRÁTICA:

Recordemos que el método de ajuste por "mínimos cuadrados" a una

recta $y = a + b x$ se basaba en obtener como coeficientes a, b aquellos valores que :

$$\min_j (a, b) = \min S (y_i - a - b x_i)^2$$

y que el resultado de tal ajuste acababa siendo: $b = \frac{S_{xy}}{S_x^2}$ $a = \bar{y} - b \bar{x}$

Igualmente para el caso multidimensional la ecuación de ajuste mínimo cuadrático $y = b_0 + b_1 x_1 + \dots + b_k x_k$

que matricialmente puede expresarse (si la extendemos para los n datos disponibles) como: $\mathbf{y} = \mathbf{Xb}$

la podremos obtener

$$\min \mathbf{j} (b_0, b_1, \dots, b_k) = \min \mathbf{S} (y_i - b_0 - b_1 x_{i1} - \dots - b_k x_{ik})$$

$$\text{esto es: } \min \mathbf{j} (\mathbf{b}) = \min (\mathbf{y} - \mathbf{Xb})' (\mathbf{y} - \mathbf{Xb})$$

de forma que el vector de coeficientes \mathbf{b} que nos dará la ecuación de ajuste mínimo-cuadrático será tal que igualará a cero la derivada parcial de $\mathbf{j} (\mathbf{b})$ con respecto a \mathbf{b} :

$$\begin{aligned} \mathbf{j} (\mathbf{b}) &= (\mathbf{y} - \mathbf{Xb})' (\mathbf{y} - \mathbf{Xb}) = \mathbf{y}'\mathbf{y} - (\mathbf{Xb})'\mathbf{y} - \mathbf{y}'(\mathbf{Xb}) + (\mathbf{Xb})'(\mathbf{Xb}) = \\ &= \mathbf{y}'\mathbf{y} - \mathbf{b}'\mathbf{X}'\mathbf{y} - \mathbf{y}'\mathbf{X}\mathbf{b} + \mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{y}'\mathbf{y} - 2\mathbf{b}'\mathbf{X}'\mathbf{y} + \mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{b} \end{aligned}$$

por lo que la derivada parcial de $\mathbf{j} (\mathbf{b})$ con respecto a \mathbf{b} acaba siendo:

$-2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\mathbf{b}$ que si la igualamos a cero nos dará la solución del ajuste mínimo-cuadrático:

$$-2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\mathbf{b} = 0 \hat{=} \mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}$$

de forma que el vector de coeficientes que nos da el ajuste a un hiperplano que minimiza la suma de los cuadrados de los residuos es:

$$\mathbf{b} = \begin{pmatrix} b_0 \\ b_1 \\ \vdots \\ b_k \end{pmatrix} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}$$

es facil comprobar que en el caso de la regresión simple los coeficientes de la recta de regresión mínimo-cuadrática verifican también esa relación:

en efecto en el caso de la regresión simple tenemos:

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \quad \mathbf{X} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \quad \mathbf{b} = \begin{pmatrix} a \\ b \end{pmatrix}$$

de manera que $\mathbf{X}'\mathbf{X}$ será:

$$\mathbf{X}'\mathbf{X} = \begin{pmatrix} n & n\bar{x} \\ n\bar{x} & na_2(x) \end{pmatrix}$$

$$\text{y su inversa } (\mathbf{X}'\mathbf{X})^{-1} = (\mathbf{X}'\mathbf{X})^{-1} = \frac{1}{nS_x^2} \begin{pmatrix} a_2(x) & -\bar{x} \\ -\bar{x} & 1 \end{pmatrix}$$

$$\text{y por otro lado } \mathbf{X}'\mathbf{y} \text{ sera: } \mathbf{X}'\mathbf{y} = \begin{pmatrix} n\bar{y} \\ na_{11} \end{pmatrix}$$

por lo que el vector $(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}$ acaba siendo:

$$(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y} = \frac{1}{S_x^2} \begin{pmatrix} a_2(x)\bar{y} - \bar{x}a_{11} \\ a_{11} - \bar{x}\bar{y} \end{pmatrix} = \begin{pmatrix} \bar{y} - \frac{S_{xy}}{S_x^2} \bar{x} \\ \frac{S_{xy}}{S_x^2} \end{pmatrix}$$

es decir los valores de los coeficientes a y b que ya conocemos.

RECORDEMOS, en otro orden de cosas, que la VARIANZA RESIDUAL era uno de los indicadores de la calidad del ajuste tanto en el caso de la REGRESIÓN SIMPLE como de la LA MÚLTIPLE, y se definía como: $S_r^2 = 1/n \cdot \sum (y_i - a - bx_i)^2$ en el caso de la R. simple

$$S_r^2 = 1/n \cdot \sum (y_i - b_0 - b_1x_1 - \dots - b_kx_k)^2$$

en el caso de la R. múltiple

$$\text{y matricialmente } S_r^2 = 1/n \cdot (\mathbf{y} - \mathbf{Xb})'(\mathbf{y} - \mathbf{Xb})$$

En el caso de la regresión simple : $S_r^2 = S_y^2(1 - r^2) = S_y^2 - r^2 \cdot S_y^2$

y en el caso de la regresión múltiple: $S_r^2 = S_y^2(1 - R^2) = S_y^2 - R^2 \cdot S_y^2$

donde r y R son respectivamente el coeficiente de correlación entre x e y y el coeficiente de correlación múltiple entre y, y las x_i (a sus cuadrados se les llama coeficiente de determinación)