Modelo Lineal General

III. MODELO DE REGRESIÓN LINEAL GENERAL

Objetivos

1. Expresar un modelo de regresión lineal de manera desarrollada.
2. Conocer las hipótesis en las que se basa el modelo de regresión lineal general.
3. Realizar un análisis de regresión con dos o más variables económicas de manera satisfactoria.
4. Utilizar el álgebra matricial para expresar el modelo y para recoger expresiones relativas tanto el estimador de mínimos cuadrados ordinarios como a las propiedades de dicho estimador.
5. Conocer las propiedades del estimador de MCO
6. Comprender la lógica de la estimación de los parámetros por máxima verosimilitud.
7. Conocer e interpretar las medidas de bondad del ajuste.
8. Realizar inferencia sobre los parámetros individuales, en concreto, contrastes de hipótesis e intervalos de confianza en el contexto de un modelo de regresión lineal de más de dos variables.

Objetivos (cont.)

9. Realizar contrastes de restricciones lineales sobre los parámetros, con especial incidencia en el contraste conjunto de significación de parámetros.
10. Comprender la lógica del estimador de mínimos cuadrados restringidos.
11. Comprender las propiedades del estimador de mínimos cuadrados restringidos.
12. Conocer las restricciones especiales y sus implicaciones principales.
13. Construir intervalos de confianza para la predicción del valor medio y del valor puntual.
14. Evaluar y validar los resultados obtenidos en un modelo de regresión
15. Interpretar económicamente los resultados de un modelo de regresión de más de dos variables
16. Practicar la estimación, realización de contrastes e interpretación de modelos econométricos de más de dos variables mediante el paquete econométrico EVIEWS.

III. MODELO DE REGRESIÓN LINEAL GENERAL

Índice

1.1. Especificación

Definición: Se llama Modelo de Regresión Lineal General al modelo probabilístico $\boldsymbol{y = \beta_1 + \beta_2 x_2 + \ldots + \beta_k x_k + e}$

- *Para la estimación de la **estructura** –desconocida- del modelo se dispone de un conjunto de n observaciones* $(y_i, x_{2i}, \ldots, x_{ki})$, $i = 1,\ldots,n$, *de modo que para cada una de ellas, es decir para* $i = 1,\ldots, n$, *se tiene*

$$y_i = \beta_1 + \beta_2 x_{2i} + \ldots + \beta_k x_{ki} + e_i ,$$

o bien , haciendo $x_{1i} = 1 \ \forall i$,

$$\boxed{y_i = \sum_{j=1}^{k} \beta_j x_{ji} + e_i}$$

El conjunto de igualdades puede expresarse ***matricialmente****:*

$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{21} & \cdots & x_{k1} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & x_{2n} & \cdots & x_{kn} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} e_1 \\ \vdots \\ e_n \end{bmatrix} \quad \text{o} \quad y = X\beta + e$$

- $\boldsymbol{y = (y_1,\ldots,y_n)'}$ es el *vector columna de las observaciones de la variable endógena.*
- $\boldsymbol{X = (x_{ij})}$ $i = 1,\ldots,n;\ j = 1,\ldots,k$ es la matriz, de orden (n, k), *de las observaciones de las variables exógenas*, considerando $x_{1i} = 1 \ \ \forall i$.
- La *fila i-ésima* de la matriz X, $\boldsymbol{x_i = (1, x_{2i},\ldots, x_{ki})}$ contiene los valores de la *observación i-ésima para las k variables explicativas* del modelo.
- La *columna j-ésima* de X, $\boldsymbol{(x_{j1},\ldots, x_{jn})'}$ contiene los *n valores observados para la variable* x_j.
- $\boldsymbol{\beta = (\beta_1, \beta_2, \ldots, \beta_k)'}$ *es el vector que contiene los parámetros del modelo*
- $\boldsymbol{e = (e_1,\ldots,e_n)'}$ es el vector de los errores o perturbaciones aleatorias.

H1: *El modelo está* ***bien especificado****, es decir, la* ***especificación estocástica*** *de la función de regresión poblacional es* $y_i = \beta_1 + \beta_2 x_{2i} + ... + \beta_k x_{ki} + e_i$ *o, matricialmente,* $y = X\beta + e$

H1.1: *Se incluyen en el modelo todas las variables relevantes.*

H1.2: *El término de error,* e_i*, actúa aditivamente.*

H1.3: *La relación es lineal en los parámetros.*

H1.4: *Los parámetros estructurales (coeficientes de regresión poblacionales) son constantes, no variando ni en el horizonte de observación ni en el de predicción.*

H2: *La matriz* ***X es una matriz no estocástica*** *y toma valores fijos en muestras repetidas.*

También ha de cumplirse que rango(X) = k < n (hipótesis ***de rango pleno****):*

- *más observaciones que parámetros a estimar,*
- *no multicolinealidad.*

H3: $\forall i$ $E[e_i/X] = 0$, *(hipótesis de* ***esperanza nula de los errores****).*

Matricialmente esta condición se expresa como:

$$E[e/X] = \begin{bmatrix} E(e_1/X) \\ \vdots \\ E(e_n/X) \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} = 0_n$$

1.2. Hipótesis básicas del modelo (Cont.)

H4: *$\forall i$ $Var[e_i / X] = \sigma^2 > 0$, hipótesis de* ***homoscedasticidad*** *(igual dispersión) o igualdad de las varianzas condicionadas de los errores.*

H5*: $\forall i \neq j$ $Cov[e_i, e_j / X] = 0$, hipótesis de ausencia de correlación en los errores.(****No autocorrelación****)*

H6*: Cada perturbación e_i sigue una distribución normal y, por tanto, el vector $e=(e_1,\ldots,e_n)'$ sigue una distribución normal multivariante. (Hipótesis de* ***normalidad de los errores****).*

Nota: representaciones adicionales

Las hipótesis H4 y H5 pueden ser agrupadas expresando la matriz de covarianzas del vector de las perturbaciones, ***e****, como*

$$Cov(e) = E\left[(e - E(e))(e - E(e))'\right] = E(ee') = (Cov(e_i, e_j)) = \sigma^2 I_n$$

$$Cov(e) = (Cov(e_i,e_j)) = \begin{bmatrix} Var(e_1) & Cov(e_1,e_2) & \cdots & Cov(e_1,e_n) \\ Cov(e_1,e_2) & Var(e_2) & \cdots & Cov(e_2,e_n) \\ \vdots & \vdots & \vdots & \vdots \\ Cov(e_1,e_n) & \cdots & \cdots & Var(e_n) \end{bmatrix} = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & \sigma^2 \end{bmatrix} = \sigma^2 I_n$$

Nota: representaciones adicionales (cont.)

- *Las perturbaciones que cumplen esta hipótesis (homoscedasticidad y no autocorrelación) son conocidas como **perturbaciones esféricas**.*
- *La hipótesis 3, 4, 5 y 6 pueden resumirse en:* $e / X \approx N(0_n, \sigma^2 I_n)$
- *La hipótesis de normalidad no es necesaria para obtener algunos de los resultados más importantes de la regresión múltiple. Su necesidad estriba, básicamente, en la posibilidad de obtener distribuciones exactas de los estadísticos utilizados.*
- *De acuerdo con las hipótesis establecidas, es inmediato deducir que la **función de regresión poblacional**, desconocida, **es una función lineal** (hiperplano) dada por*

$$E(y_i / x_{2i}, \ldots, x_{ki}) = \beta_1 + \beta_2 x_{2i} + \ldots + \beta_k \; x_{ki},$$

o, matricialmente, $E(y / X) = E(X\beta + e) = X\beta + E(e / X) = X\beta.$

- *La matriz de covarianzas de y viene dada por:*

$$Cov(y / X) = Cov(X\beta + e) = Cov(e) = \sigma^2 I_n$$

- *Por consiguiente, con los resultados anteriores se tiene que* $y \approx N(X\beta ; \sigma^2 I_n)$.

2. Estimación de los parámetros por mínimos cuadrados ordinarios (MCO)

- *Estimación de la función de regresión poblacional*: la *ecuación del hiperplano* más próximo a la nube o conjunto de puntos muestrales.

- MCO➔ el plano más próximo al conjunto de puntos muestrales es aquel para el cual se hace mínima la cantidad:

$$S(\hat{\beta}) = \sum_{i=1}^{n} \hat{e}_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

- La *función minimando* también se puede expresar *matricialmente* en la forma:

$$S(\hat{\beta}) = \Sigma \hat{e}_i^2 = (\hat{e}_1 \quad \cdots \quad \hat{e}_n)\begin{pmatrix} \hat{e}_1 \\ \vdots \\ \hat{e}_n \end{pmatrix} = \hat{e}'\hat{e} = (y - X\hat{\beta})'(y - X\hat{\beta}) = y'y - 2\hat{\beta}X'y + \hat{\beta}'X'X\hat{\beta}$$

- *Para obtener los valores de los parámetros que minimizan la función derivamos respecto de e igualamos a 0 (condición de primer orden) :*

$$\frac{\partial S(\hat{\beta})}{\partial \hat{\beta}} = -2X'y + 2X'X\hat{\beta} = 0 \Leftrightarrow X'y = X'X\hat{\beta}$$

$X'X$ es matriz cuadrada de orden k y rango k ➔ $\exists\ (X'X)^{-1}$,

Premultiplicando las ecuaciones normales por $(X'X)^{-1}$, se obtienen las **estimaciones MCO de β**

$$\hat{\beta} = (X'X)^{-1}X'y$$

3. Propiedades del estimador de *MCO*

- Cuando la expresión $\boldsymbol{b = (X'X)^{-1}X' y}$ se considera función de los valores de la variable y para distintas muestras hablaremos del *estimador* MCO:

$$\boxed{\hat{\beta} = (X'X)^{-1}X'y}$$

→ al tratarse de una v. a.

¿qué propiedades caracterizan su distribución?

1. Es función ***lineal*** de y.

   En efecto, puesto que $\hat{\beta} = (X'X)^{-1}X'y = Ty$

   donde $T = (X'X)^{-1}X'$ es una matriz de valores fijos.

2. Es un *estimador* **insesgado**. $E(\hat{\beta}/X) = E(\hat{\beta}) = \beta$

$$\hat{\beta} = (X'X)^{-1}X'y = (X'X)^{-1}X'(X\beta + e) =$$
$$= (X'X)^{-1}(X'X)\beta + (X'X)^{-1}X'e = \beta + (X'X)^{-1}X'e$$

y, como por hipótesis $E(e) = 0$, se tiene que

$$E(\hat{\beta}) = \beta + (X'X)^{-1}X'E(e) = \beta$$

3. Propiedades del estimador de *MCO* (*Cont.*)

3. Matriz de covarianzas: $\boxed{Cov(\hat{\beta}) = (Cov(\beta_i, \beta_j)) = \sigma^2 (X'X)^{-1}}$

$$Cov(\hat{\beta}) = E((\hat{\beta}-\beta)(\hat{\beta}-\beta)') = E((X'X)^{-1}X'ee'X(X'X)^{-1}) =$$
$$= (X'X)^{-1}X'E(ee')X(X'X)^{-1} == (X'X)^{-1}X'\sigma^2 IX(X'X)^{-1} =$$
$$= \sigma^2(X'X)^{-1}X'X(X'X)^{-1} = \sigma^2(X'X)^{-1}$$

En forma desarrollada:

$$Cov(\hat{\beta}) = (Cov(\hat{\beta}_i, \hat{\beta}_j)) = \begin{pmatrix} Var(\hat{\beta}_1) & Cov(\hat{\beta}_1, \hat{\beta}_2) & \cdots & Cov(\hat{\beta}_1, \hat{\beta}_k) \\ Cov(\hat{\beta}_1, \hat{\beta}_2) & Var(\hat{\beta}_2) & \cdots & Cov(\hat{\beta}_2, \hat{\beta}_k) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(\hat{\beta}_1, \hat{\beta}_k) & \cdots & \cdots & Var(\hat{\beta}_k) \end{pmatrix} = \sigma^2 (X'X)^{-1}$$

Para los parámetros individuales, siendo $(a_{ij}) = (X'X)^{-1}$, se tiene que

$$\boxed{Var(\hat{\beta}_i) = \sigma^2 a_{ii}} \qquad \boxed{Cov(\hat{\beta}_i, \hat{\beta}_j) = \sigma^2 a_{ij}}$$

3. Propiedades del estimador de *MCO* (Cont.)

4. El vector de residuos $\hat{e}$, es ortogonal a la matriz de variables explicativas, X:

$$X'\hat{e} = 0$$

5. La media de los residuos es 0: $\bar{\hat{e}}_i = 0$ y $\sum_i \hat{e}_i = 0$

6. **T*eorema de Gauss-Markov***: El estimador MCO, es óptimo dentro de la clase de los estimadores lineales e insesgados de β

   (MELI: Mejor Estimador Lineal Insesgado)

7. Consistencia: $p \lim \left[\hat{\beta}_j \right] = \beta_j$

8. Eficiencia, eficiencia asintótica y normalidad.

4. Estimación de σ^2

- Como estimador de **la varianza de las distribuciones condicionadas de los errores** se toma la ***varianza residual***

$$\boxed{\hat{\sigma}^2 = \frac{\Sigma \hat{e}_i^2}{n-k} = \frac{\hat{e}'\hat{e}}{n-k}}$$

- La **varianza residual** es un **estimador insesgado** de σ^2

$$\boxed{E(\hat{\sigma}^2) = \sigma^2}$$

5. Estimación de los parámetros por máxima verosimilitud

$y = X\beta + e, \quad e = (e_1, \ldots, e_n)', \quad e / X \approx N(0_n, \sigma^2 I_n)$ → $y \approx N(X\beta, \sigma^2 I_n)$

Se dispone de una muestra aleatoria de *n* observaciones $(y_i, x_{2i}, \ldots, x_{ki})_{i=1,\cdots,n}$ → **Función de densidad del vector de observaciones de la variable endógena,** $y = (y_1, \ldots, y_n)'$, vendrá dada por:

$$f(y; \beta, \sigma^2) = f(y_1, \ldots, y_n; \beta_1, \ldots, \beta_k; \sigma^2) = \prod_{i=1}^{n} f(y_i, \beta, \sigma^2) =$$

$$= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{ -\frac{1}{2\sigma^2}(y_i - \Sigma \beta_j x_{ji})^2 \} =$$

$$= (2\pi\sigma^2)^{-n/2} \exp\{ -\frac{1}{2\sigma^2} \Sigma (y_i - \Sigma \beta_j x_{ji})^2 \} =$$

$$= (2\pi\sigma^2)^{-n/2} \exp\{ -\frac{1}{2\sigma^2} \Sigma e_i^2 \} =$$

$$= (2\pi\sigma^2)^{-n/2} \exp\{ -\frac{1}{2\sigma^2} (y - X\beta)'(y - X\beta) \}$$ →

5. Estimación de los parámetros por máxima verosimilitud (Cont.)

- Función de verosimilitud de la muestra:

$$L(\beta,\sigma^2;y)=(2\pi\sigma^2)^{-n/2}\exp\left(-\frac{1}{2\sigma^2}(y-X\beta)'(y-X\beta)\right)$$

- Tomando logaritmos

$$l(\beta,\sigma^2,y)=\log L(\beta,\sigma^2,y)=-\frac{n}{2}\log 2\pi-\frac{n}{2}\log\sigma^2-\frac{1}{2\sigma^2}(y-X\beta)'(y-X\beta)$$

¿Valores de β y σ^2 *que hacen máxima la función l*?

Se deriva respecto de cada conjunto de parámetros:

$$\frac{\partial l}{\partial\beta}=-\frac{1}{2\sigma^2}(-2X'y+2X'X\beta)=0 \quad\Leftrightarrow\quad X'X\beta=X'y$$

$$\frac{\partial l}{\partial\sigma^2}=-\frac{n}{2\sigma^2}+\frac{1}{\sigma^4}(y-X\beta)'(y-X\beta)=0 \quad\Leftrightarrow\quad \frac{n}{2\sigma^2}=\frac{1}{\sigma^4}(y-X\beta)'(y-X\beta)=0$$

$X'X\beta=X'y$ ➔ sistema de **ecuaciones normales**

$$\boxed{\hat\beta_{MV}=\hat\beta=(X'X)^{-1}X'y}$$

El estimador MCO de β goza de las propiedades del estimador de MV: **consistencia, insesgadez y eficiencia asintóticas.**

$$\frac{n}{2\sigma^2}=\frac{1}{\sigma^4}\hat e'\hat e\Rightarrow\hat\sigma^2_{MV}=\frac{\hat e'\hat e}{n}$$

Estimador **sesgado,** aunque **consistente.**

$$E\left[\hat\sigma^2_{MV}\right]=E\left[\frac{\hat e'\hat e}{n}\right]=\frac{n-k}{n}E\left[\frac{\hat e'\hat e}{n-k}\right]=\frac{n-k}{n}\sigma^2$$

$$SRC = \sum_{i=1}^{n} \hat{e}_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

No es una medida homogénea
se ve afectada por la escala de medida

⇨ Otras medidas

6.1. Descomposición de la variación de la variable dependiente

Sumas de cuadrados: Variación total de la variable dependiente alrededor de su media

$$\boxed{\Sigma(y_i - \bar{y})^2} = \boxed{\Sigma(\hat{y}_i - \bar{y})^2} + \boxed{\Sigma(y_i - \hat{y}_i)^2} \quad \text{o} \quad \Sigma(y_i - \bar{y})^2 = \Sigma(\hat{y}_i - \bar{y})^2 + \Sigma \hat{e}_i^2$$

STC = ***SEC*** + ***SRC***

Total = Explicada + Residual

6.2. Coeficiente de determinación: R^2

% de la variación total de la variable dependiente explicada por el modelo:

$$\boxed{R^2 = \frac{SEC}{STC} = \frac{\Sigma(\hat{y}_i - \bar{y})^2}{\Sigma(y_i - \bar{y})^2} =} \quad \boxed{= 1 - \frac{SRC}{STC} = 1 - \frac{\Sigma(y_i - \hat{y}_i)^2}{\Sigma(y_i - \bar{y})^2}}$$

$$\boxed{R^2} = \frac{\hat{\beta}' X' Y - n\bar{y}^2}{y' y - n\bar{y}^2} = 1 - \frac{y' y - \hat{\beta}' X' y}{y' y - n\bar{y}^2} = 1 - \frac{\hat{e}' \hat{e}}{y' y - n\bar{y}^2} = \boxed{1 - \frac{(n-k)\hat{\sigma}^2}{n s_y^2}}$$

6. Bondad del ajuste. Coeficiente de determinación: R^2

Propiedades:

1. $0 \leq R^2 \leq 1$, (cuando ∃ término independiente)
2. $R^2 = \left(r_{y,\hat{y}}\right)^2$
3. $R^2 = 1$ (ajuste perfecto) ⟶ (STC = SEC o SRC = 0)
4. $R^2 = 0$, (STC = SRC) ⟶ $y_i = \bar{y} + e_i$

6.3. Coeficiente de determinación ajustado

Problemas que presenta $\boldsymbol{R}^2$:

- $\boldsymbol{R}^2$ se incrementa (ó no disminuye) cuando aumenta el número de variables
- No son comparables mediante el $\boldsymbol{R}^2$:
  - Regresiones con la misma variable dependiente, pero distinto nº. de variables
  - Regresiones con diferentes variables dependientes
- Se introduce una nueva medida, *el coeficiente de determinación ajustado* eliminando el efecto que el nº de variables (con la pérdida de grados de libertad) introduce en R^2.

$$\boxed{\bar{R}^2 = 1 - \frac{SRC\ /(n-k)}{STC\ /(n-1)} = 1 - \frac{\hat{\sigma}^2}{s_y^2} =} \quad \boxed{= 1 - (1 - R^2)\frac{n-1}{n-k}}$$

Puede disminuir cuando se incorporan nuevas variables si no compensa la pérdida de grados de libertad (el aumento de su valor al incluir una nueva variable dependerá de su contribución al modelo).

7.1. Distribución de los estimadores de los parámetros individuales

$$\hat{\beta} \approx N(\beta, \sigma^2 (X'X)^{-1})$$
$$Cov(\hat{\beta}, \hat{e}) = 0$$

$\Rightarrow$ $\hat{\beta}$ y $\hat{\boldsymbol{e}}$ son independientes

$$\frac{(\hat{\beta}_i - \beta_i)}{\sigma \sqrt{a_{ii}}} \approx N(0,1)$$

$$\frac{(n-k)\hat{\sigma}^2}{\sigma^2} = \frac{\Sigma \hat{e}_i^2}{\sigma^2} = \frac{e'e}{\sigma^2} \approx \chi^2_{n-k}$$

$$t = \frac{\dfrac{(\hat{\beta}_i - \beta_i)}{\sigma\sqrt{a_{ii}}}}{\sqrt{\dfrac{(n-k)\hat{\sigma}^2}{\sigma^2(n-k)}}} = \frac{(\hat{\beta}_i - \beta_i)}{\hat{\sigma}\sqrt{a_{ii}}} = \frac{(\hat{\beta}_i - \beta_i)}{se(\hat{\beta}_i)} \approx t_{n-k}$$

- Intervalos de confianza
- Contrastes estadísticos

7.2. Intervalos de confianza

Intervalo al (1-α)% de confianza para β_i :

$$I_{(1-\alpha)\%}(\beta_i) = (\hat{\beta}_i - t_{\alpha/2;n-k}\, s(\hat{\beta}_i),\ \hat{\beta}_i + t_{\alpha/2;n-k}\, s(\hat{\beta}_i)) =$$
$$= (\hat{\beta}_i - t_{\alpha/2;n-k}\, \hat{\sigma}\sqrt{a_{ii}},\ \hat{\beta}_i + t_{\alpha/2;n-k}\, \hat{\sigma}\sqrt{a_{ii}})$$

7.3. Contrastes de hipótesis

$$H_0: \{\beta_i = \beta_i^0\} \text{ vs. } H_1: \{\beta_i \neq \beta_i^0\}$$

Estadístico de prueba

$$t = \frac{(\hat{\beta}_i - \beta_i^0)}{s(\hat{\beta}_i)} \xrightarrow{H_0} t_{n-k}$$

Región crítica

$$|t| > t_{\alpha/2,n-k}$$

8.1. Resultados estadísticos previos

Sea X un vector aleatorio n-dimensional:

1. $X \approx N(\mu, \Sigma) \quad y \quad T \text{ matriz fija} \quad \Rightarrow \quad TX \approx N(T\mu, T\Sigma T')$

2. $X \approx N(\mu, \Sigma) \Rightarrow (X-\mu)'\Sigma^{-1}(X-\mu) \approx \chi^2_n$

3. $Z_1 \approx \chi^2_n \;\; y \; Z_2 \approx \chi^2_m \; independientes \Rightarrow \quad \dfrac{Z_1/n}{Z_2/m} = \dfrac{\chi^2_n/n}{\chi^2_m/m} \approx F_{n,m}$

Aplicado a nuestro caso nos lleva a que:

$$Y \approx N(X\beta;\sigma^2 I) \Rightarrow \hat{\beta} = (X'X)^{-1}X'Y \approx N(\beta, \sigma^2(X'X)^{-1}) \Rightarrow \frac{(\hat{\beta}-\beta)'(X'X)(\hat{\beta}-\beta)}{\sigma^2} \approx \chi^2_k$$

$$\frac{(n-k)\hat{\sigma}^2}{\sigma^2} = \frac{\hat{e}'\hat{e}}{\sigma^2} = \chi^2_{n-k} \quad \text{y} \quad \hat{\beta} \text{ y } \hat{e} \quad \text{(independientes)}$$

$$\boxed{\frac{(\hat{\beta}-\beta)'(X'X)(\hat{\beta}-\beta)/k}{\hat{\sigma}^2} = \frac{(\hat{\beta}-\beta)'(X'X)(\hat{\beta}-\beta)/k}{\hat{e}'\hat{e}/(n-k)} \approx F_{k,n-k}}$$

8.2. Contrastes de restricciones lineales sobre β

- Sea R una matriz de restricciones lineales sobre β, siendo $rango(R) = q < k$

  Ejemplo: Restricciones $\beta_2 = 0,\ \beta_3 = \beta_4,\ \beta_1 + 2\beta_5 + 3\beta_6 = 2,\ \beta_1 + \beta_4 + \beta_5 + \beta_6 = 0$

  Matricialmente $$\begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 2 & 3 \\ 1 & 0 & 0 & 1 & 1 & 1 \end{bmatrix} (\beta_1, \beta_2, \beta_3, \beta_4, \beta_5)' = \begin{bmatrix} 0 \\ 0 \\ 2 \\ 0 \end{bmatrix}$$

1) del epígrafe anterior, $R\hat{\beta} \approx N(R\beta, \sigma^2 R(X'X)^{-1}R')$ es un vector q-dimensional, con distribución normal.

De 2) y 3) Si $\boldsymbol{R\beta} = \boldsymbol{r}$ se tiene que

$$\boxed{F = \frac{(R\hat{\beta} - r)'\left(R(X'X)^{-1}R'\right)^{-1}(R\hat{\beta} - r)/q}{\hat{\sigma}^2} = (R\hat{\beta} - r)'\left(\hat{Cov}(R\hat{\beta})\right)^{-1}(R\hat{\beta} - r)/q \approx F_{q,n-k}}$$

Luego:

- Para el contraste $H_0 : \{R\beta = r\}$ *vs.* $H_1 : \{R\beta \neq r\}$
- La región crítica estará dada por $\boxed{F > F_{\alpha;q;n-k}}$

Contrastes de restricciones lineales. 8.3. Contraste de validez general del modelo

- **Hipótesis a contrastar:** $\boxed{H_0:\{\beta_2=0,\beta_3=0,\ldots,\beta_k=0\}\ vs.\ H_1:\{\exists j>1/\ \beta_j\neq 0\}}$

  equivalente a $\boxed{H_0:\{Y_i=\beta_1+e_i\}\ vs.\ H_1:\{y_i=\beta_1+\beta_2x_{2i}+\ldots+\beta_kx_{ki}+e_i\}}$

- **Matriz de restricciones:**

  rango de $R = k - 1$

  (igual al número de restricciones)

$$R=\begin{bmatrix}0 & 1 & 0 & \cdots & 0\\ 0 & 0 & 1 & \cdots & 0\\ \vdots & \vdots & \cdots & \ddots & 0\\ 0 & \cdots & \cdots & 0 & 1\end{bmatrix}=(0:I_{k-1})$$

- **Estadístico de contraste**

$$\boxed{F=\frac{(R\hat{\beta})'(R(X'X)^{-1}R')^{-1}(R\hat{\beta})/(k-1)}{\hat{\sigma}^2}}$$

- **Bajo la hipótesis nula $H_0:\{R\beta=0\}$, es:**

$$F=\frac{(R\hat{\beta})'(X'X)(R\hat{\beta})/(k-1)}{\hat{\sigma}^2}=\frac{SEC/(k-1)}{SRC/(n-k)}=\frac{(SEC/STC)/(k-1)}{(SRC/STC)/(n-k)}=\frac{R^2/(k-1)}{(1-R^2)/(n-k)}\approx F_{k-1,n-k}$$

- **Por tanto, la región crítica es:**

$$\boxed{F>F_{\alpha;k-1,n-k}}$$

8.3. Contraste de validez general del modelo. Cuadro de Análisis de la Varianza

- Para el cálculo del estadístico ***F*** se acostumbra a utilizar el siguiente **cuadro de análisis de la varianza** general.

Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrado Medio	F
Regresión $x_2,\ldots,x_k$	$SEC=\hat{\beta}'X'y-n\bar{y}^2$	$k-1$	$SEC/(k-1)$	$\dfrac{SEC/(k-1)}{SRC/(n-k)} \approx F$
Residual	$SRC=y'y-\hat{\beta}X'y$	$n-k$	$SRC/(n-k)=\hat{\sigma}^2$	
Total	$STC=y'y-n\bar{y}^2$	$n-1$		

Mínimos Cuadrados Restringidos (MCR)

- Supongamos que se tiene información *a priori* sobre el modelo: $\boldsymbol{R\beta = r}$
- Si se incorporan las restricciones al modelo, *reparametrizándolo*, y se lleva cabo la estimación MCO, el método es conocido como ***Mínimos Cuadrados Restringidos***
- Es equivalente a minimizar $S(\beta) = (y - X\beta)'(y - X\beta)$ sujeta a $R\beta = r$

  Que es lo mismo que minimizar $L(\beta) = (y - X\beta)'(y - X\beta) - \lambda'(R\beta - r)$

  donde $\lambda = (\lambda_1, \ldots, \lambda_q)'$ es el vector de multiplicadores de Lagrange
- La solución está dada por el ***Estimador de mínimos cuadrados restringidos***:

$$\boxed{\hat{\beta}_R = \hat{\beta} + (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}(r - R\hat{\beta})}$$

9. Combinación de información muestral y no muestral: estimación restringida (Cont.)

Propiedades del estimador MCR

- $\hat{\beta}_R$ insesgado $\Longleftrightarrow$ $R\beta = r$
- $R\hat{\beta}_R = r$
- $Cov(\hat{\beta}_R) \leq Cov(\hat{\beta})$ aún cuando las restricciones sean falsas

Forma habitual de proceder

- Si las *restricciones* son *válidas* se debe tomar el estimador MCR:
  $\hat{\beta}_R$ será insesgado y de menor varianza
- Si las *restricciones no* son *válidas* se debe tomar el de MCO
  El de MCR será sesgado

9. Combinación de información muestral y no muestral: estimación restringida (Cont.)

- *Se estima el modelo sin restringir*, obteniéndose los residuos no restringidos ê y la suma de los cuadrados de los mismos, $\boldsymbol{SRC_{NR} = \hat{e}'\hat{e}}$.

- *Se estima el modelo restringido* (reparametrizando el modelo original de acuerdo con las restricciones dadas), obteniéndose los residuos $\boldsymbol{\hat{e}_R}$ y la suma de los cuadrados de sus residuos (restringidos): $\boldsymbol{SRC_R = \hat{e}_R'\hat{e}_R}$.

- Si las restricciones $\boldsymbol{R\beta = r}$ son válidas, el estadístico

$$F = \frac{(R\hat{\beta}-r)'(R(X'X)^{-1}R')^{-1}(R\hat{\beta}-r)/q}{\hat{\sigma}^2} = \frac{(\hat{e}'_R\hat{e}_R - \hat{e}'\hat{e})/q}{\hat{e}'\hat{e}/(n-k)} = \frac{(SRC_R - SRC_{NR})/q}{SRC_{NR}/(n-k)} \approx F_{q;n-k}$$

Por tanto:

- Si $\boldsymbol{F > F_{\alpha;q;n-k}}$ rechazamos H_0:´{$\boldsymbol{R\beta = r}$ } tomamos el estimador MCO

- Si $\boldsymbol{F < F_{\alpha;q;n-k}}$ aceptamos H_0:´{$\boldsymbol{R\beta = r}$ } y se toma el estimador restringido

11. Predicción en el modelo lineal general

- Modelo poblacional –desconocido- $y = X\beta + e,$

- Modelo ajustado ó estimado por MCO:

  función de regresión muestral $\hat{y} = X\hat{\beta}$

- El *objetivo de la predicción* puede ser, dados los distintos niveles de las variables explicativas contenidos en el vector $x_{0;}$ $x'_0 = (1, x_{20}, \ldots, x_{k0})$

  - estimar el *valor medio* $E[y / x_0']$

  - estimar el *valor individual* y_0

- El *estimador* propuesto es el mismo

$$\hat{y}_0 = \hat{\beta}_1 + \hat{\beta}_2 x_{20} + \cdots + \hat{\beta}_k x_{k0} = (1, x_{20}, \ldots, x_{k0}) \begin{pmatrix} \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_k \end{pmatrix} = x_0' \hat{\beta}$$

- En ambos caso es el mejor predictor lineal insesgado

- Las varianzas de los estimadores son distintas

Para hacer inferencias sobre las predicciones se debe tener en cuenta que

- Las varianzas de los estimadores son distintas, respectivamente:

$$\boxed{Var\,\hat{E}[y/x_0'] = Var(x_0'(\beta-\hat{\beta})) = \sigma^2(x_0'(X'X)^{-1}x_0)}$$

$$\boxed{Var\,\hat{y}_0 == Var(x_0'(\beta-\hat{\beta})+e_0) = \sigma^2(1+x_0'(X'X)^{-1}x_0)}$$

- Por tanto se tienen distintos intervalos de confianza:

$$I_{(1-\alpha)\%}(E(y_0)) = (x_0'\hat{\beta} - t_{\alpha/2;n-k}\hat{\sigma}\sqrt{x_0'(X'X)^{-1}x_0}\,,\ x_0'\hat{\beta} + t_{\alpha/2;n-k}\hat{\sigma}\sqrt{x_0'(X'X)^{-1}x_0}\,)$$

$$I_{(1-\alpha)\%}(y_0) = (x_0'\hat{\beta} - t_{\alpha/2;n-k}\hat{\sigma}\sqrt{1+x_0'(X'X)^{-1}x_0}\,,\ x_0'\hat{\beta} + t_{\alpha/2;n-k}\hat{\sigma}\sqrt{1+x_0'(X'X)^{-1}x_0}\,)$$

- Y los estadísticos, para contrastes individuales, respectivamente:

$$\boxed{\frac{E(y/x_0)-x_0'\hat{\beta}}{\hat{\sigma}\sqrt{x_0'(X'X)^{-1}x_0}} \approx t_{n-k}} \qquad \boxed{\frac{y_0-x_0'\hat{\beta}}{\hat{\sigma}\sqrt{1+x_0'(X'X)^{-1}x_0}} \approx t_{n-k}}$$

12.1. Criterios para seleccionar un modelo entre distintas especificaciones alternativas, de varias variables exógenas potenciales o de varias formas funcionales posibles

- Coherencia del modelo estimado con los datos
  - ✓ significación conjunta del modelo (coeficientes R^2 o y estadístico F) y
  - ✓ significación de los parámetros individuales (t-ratios)
  - ✓ adecuación a la teoría económica subyacente (signos y magnitudes de los coeficientes, propiedades de largo y corto plazo, etc.)
- Coherencia *a posteriori* con las hipótesis iniciales:
  - ✓ ausencia de correlación, homoscedasticidad, normalidad, exogeneidad de los regresores (ausencia de correlación de las variables explicativas con el término de perturbación), constancia de los parámetros, forma funcional, etc
- Principio de Parsimonia
  - ✓ los modelos elegidos deben ser parsimoniosos (no contener un excesivo número de parámetros) y a ser posible 'anidar' a todos los posibles modelos competidores

12.2. Medidas de precisión de las predicciones y de bondad del ajuste

Basadas en la comparación entre los valores observados y los ajustados

- Varianza residual: $\hat{\sigma}^2 = \frac{1}{n-k}\Sigma(y_i - \hat{y}_i)^2$
- Error absoluto medio $EAM = \frac{1}{n}\Sigma\,|\,y_i - \hat{y}_i\,|$
- Porcentaje medio de error $PME = \frac{1}{n}\Sigma\,|\,\hat{e}_i \,/\, y_i\,| \times 100$
- Raíz cuadrada del error cuadrático medio $RECM = \sqrt{\frac{\Sigma(y_i - \hat{y}_i)^2}{n}}$
- Error cuadrático medio $ECM = RECM^2$.
- El coeficiente de desigualdad de Theil

$$U = \frac{\sqrt{\frac{1}{n}\Sigma(y_i - \hat{y}_i)^2}}{\sqrt{\frac{1}{n}\Sigma y_i^2} + \sqrt{\frac{1}{n}\Sigma \hat{y}_i^2}}$$

$0 \leq U \leq 1$, donde $U = 0$ indicaría un ajuste perfecto.

Para todas las medidas son deseables los valores menores

12.3. Medidas de ajuste basadas en criterios de información

Basados en el valor del logaritmo de la función de verosimilitud en el modelo estimado

$$l = -\frac{n}{2}\left(1 + \log(2\pi) + \log(\hat{e}'\hat{e}/n)\right)$$

- Criterio de información de Akaike (1970,1974): $\boxed{AIC = \dfrac{-2l + 2k}{n}}$
- Criterio Bayesiano de Schwarz (1978): $\boxed{SBC = \dfrac{-2l + k\log(n)}{n}}$
- Criterio de Hannan-Quinn (1979): $\boxed{HQC = \dfrac{-2l + 2k\log(\log(n))}{n}}$

Para todas las medidas *son deseables los valores menores*