CAPíTULO 4

Inferencia estadística

El supuesto de que los datos han sido generados por una determinada ley de probabilidad nos ha permitido establecer las propiedades estadísticas del estimador de mínimos cuadrados: el teorema de Gauss-Markov nos dice que, en el marco del modelo clásico, el estimador de mínimos cuadrados es el estimador más eficiente en la clase de estimadores lineales e insegados.

En este capítulo estudiamos dos problemas clásicos de inferencia estadística: el contraste de hipótesis y la predicción. Mientras que en un contraste de hipótesis nos preguntamos si los datos observados han sido generados por una determinada ley de probabilidad, en un problema de predicción tratamos de averiguar el valor de una variable aleatoria que todavía no se ha observado. El capítulo se organiza del siguiente modo. La sección 1 resume algunos conceptos básicos de la teoría de contraste de hipótesis que se explican en un curso introductorio de inferencia estadística. La sección 2 presenta las distribuciones fundamentales sobre las que se contruyen **los contrastes t y F**, que se describen en las secciones 3 y 4. La sección 5 desarrolla el contraste de la **hipótesis lineal general**, que incluye los contrastes t y F como casos especiales y que sugiere un nuevo método de estimación, sujeto a restricciones, que se deriva en la sección 6. Finalmente, la sección 7 describe la predicción con el modelo lineal general.

4.1. Conceptos básicos

EJEMPLO 2. El problema de contrastar de hipótesis estadísticas guarda ciertas analogías con el proceso de un juicio penal. El acusado se presume inocente hasta que no se demuestre lo contrario. El fallo del jurado puede ser: el acusado es inocente o el acusado es culpable. Dos errores pueden cometerse: declarar inocente al acusado cuando es culpable o declarar culpable al acusado cuando es inocente.

$\triangleleft$

Sea $\{Y_1, Y_2, \ldots, Y_n\}$ una muestra aleatoria de tamaño n de una distribución de probabilidad $p(\mathbf{y}, \theta)$, en donde θ es un parámetro desconocido que pertenece al espacio paramétrico Θ, $\theta \in \Theta \subseteq \Re^m$.

DEFINICIÓN 27. *Una <u>hipótesis paramétrica</u> es una conjetura sobre un parámetro desconocido θ de una distribución probabilidad $p(\mathbf{y}, \theta)$.*

DEFINICIÓN 28. *La hipótesis de interés $H_0 : \theta \in \Theta_0$ se denomina <u>hipótesis nula</u>, y tiene asociada una <u>hipótesis alternativa</u> $H_1 : \theta \in \Theta_1$, cumpliéndose que $\Theta_0 \cup \Theta_1 = \Theta$ y $\Theta_0 \cap \Theta_1 = \emptyset$.*

DEFINICIÓN 29. *La hipótesis nula $H_0 : \theta \in \Theta_0$ es una <u>hipótesis simple</u> si el conjunto Θ_0 incluye sólo un punto. En caso contrario, H_0 es una <u>hipótesis compuesta</u>. Si H_0 es simple, la distribución de probabilidad $p(\mathbf{y}, \theta)$ está completamente especificada bajo H_0.*

En el contexto del modelo clásico gaussiano, la distribución de probabilidad es $\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma_u^2 I_n)$, y el parámetro θ puede ser (1) un elemento del vector $\boldsymbol{\beta}$, (2) un subvector de $\boldsymbol{\beta}$, (3) un conjunto de combinaciones lineales de $\boldsymbol{\beta}$, o (4) la varianza del término de error σ_u^2. Algunos ejemplos de hipótesis que vamos a estudiar en este tema son:

1. Significación individual

$$H_0 : \beta_i = 0 \quad \text{versus} \quad H_1 : \beta_i \neq 0$$

2. Significación conjunta

$$H_0 : \begin{cases} \beta_2 = 0 \\ \beta_3 = 0 \\ \quad\vdots \\ \beta_k = 0 \end{cases} \quad \text{versus} \quad H_1 : \begin{cases} \beta_2 \neq 0 \\ \beta_3 \neq 0 \\ \quad\vdots \\ \beta_k \neq 0 \end{cases}$$

3. Restricciones lineales

$$H_0 : \beta_2 + \beta_3 = 1 \quad \text{versus} \quad H_1 : \beta_2 + \beta_3 \neq 1$$

En estos tres ejemplos, H_0 es una hipótesis simple y H_1 es una hipótesis compuesta. Además, H_1 es una hipótesis compuesta de dos lados (en inglés, *two-sided hypothesis*). Algunos ejemplos de hipótesis compuesta de un lado (en inglés, *one-sided hypothesis*) son $H_0 : \beta_i \geq 0$ y $H_0 : \beta_i < 0$.

La aproximación de Neyman y Pearson (1933) al problema de contrastar hipótesis puede describirse brevemente del siguiente modo. Dada una realización particular $\{y_1, y_2, \ldots, y_n\}$ de la variable aleatoria n-dimensional $\{Y_1, Y_2, \ldots, Y_n\}$, se desea encontrar una **regla de decisión** que indique si la hipótesis nula $H_0 : \theta \in \Theta_0$ se acepta o se rechaza. Parece razonable aceptar H_0 cuando la estimación $\hat{\theta} = \hat{\theta}(y_1, y_2, \ldots, y_n)$ esté próxima a los valores del parámetro θ fijados en H_0. Siguiendo este criterio, el espacio muestral S de todas las posibles realizaciones de la variable aleatoria n-dimensional se puede particionar en dos conjuntos disjuntos, C y su complementario C^c ($S = C \cup C^c$), correspondientes a las dos posibles decisiones: rechazar H_0 o aceptar H_0, repectivamente.

DEFINICIÓN 30. *El conjunto C formado por todas las realizaciones que rechazan H_0 se denomina región crítica.*

Observación 20. *Aceptar $H_0 : \theta \in \Theta_0$ es equivalente a rechazar $H_1 : \theta \notin \Theta_0$, y rechazar H_0 es equivalente a aceptar H_1. Algunos autores prefieren la expresión no rechazar H_0 en lugar de aceptar H_0. Por ejemplo, si no hay pruebas concluyentes para condenar a un acusado, no significa que sea inocente. Aquí se usará la expresión aceptar H_0 con el significado de no rechazar H_0.*

En este problema de decisión se pueden cometer dos errores: (1) rechazar H_0 cuando es cierta, que se denomina *error de tipo I*, y (2) aceptar H_0 cuando es falsa, que se denomina *error de tipo II*. Las probabilidades de estos dos errores se recogen en la función potencia del contraste.

DEFINICIÓN 31. *La potencia o función potencia de un contraste es la probabilidad de rechazar H_0 en cualquier punto del espacio paramétrico*

$$\pi(\theta) = P(\mathbf{y} \in S|\theta) \quad \textit{para } \theta \in \Theta$$

Si la hipótesis nula es compuesta, la función potencia puede evaluarse en cada uno de los puntos de Θ_0. La notación $\pi(\theta)$ para $\theta \in \Theta_0$ indica la probabilidad del error de

Cuadro 1: Errores de tipo I y II

	H_0 es cierta	H_1 es cierta
Aceptar H_0	Correcta	Error de tipo II
Aceptar H_1	Error de tipo I	Correcta

tipo I en cada punto de Θ_0. Análogamente, si la hipótesis alternativa es compuesta, la notación $1 - \pi(\theta)$ para $\theta \in \Theta_1$ indica la probabilidad del error de tipo II en cada punto de Θ_1.

DEFINICIÓN 32. *La máxima probabilidad del error de tipo I se denomina* <u>*nivel de significación*</u> *o tamaño del contraste, y se denota por* α

$$\alpha(\theta) \equiv P(\textit{Error tipo I}) = \sup_{\theta \in \Theta_0} \pi(\theta)$$

El contraste ideal sería aquel que siempre conduce a la decisión correcta: aceptar H_0 cuando es cierta y aceptar H_1 cuando es cierta; en otras palabras, el contraste ideal sería aquel cuya función potencia es

$$\pi(\theta) = \begin{cases} 0, & \forall \theta \in \Theta_0 \\ 1, & \forall \theta \in \Theta_1 \end{cases}$$

Parece razonable pensar que al diseñar un contraste de hipótesis se debería perseguir el doble objetivo de minimizar las probabilidades de los errores de tipo I y II. El problema es que estos dos objetivos no se pueden alcanzar simultáneamente: la probabilidad del error de tipo I sólo puede reducirse aumentando la probabilidad del error de tipo II, y viceversa. La solución de compromiso consiste en fijar la probabilidad del error de tipo I en un nivel α y tratar de minimizar la probabilidad del error de tipo II.

Los contrastes de hipótesis que se estudian en este tema se caracterizan por tener:

1. dos hipótesis plausibles: H_0 y H_1,
2. un estadístico de contraste con distribución conocida bajo H_0,
3. un nivel de significación,
4. y una región crítica.

4.2. Distribuciones asociadas a $\hat{\boldsymbol{\beta}}$ y $\hat{\sigma}_u^2$

PROPOSICIÓN 27. *Los estimadores* $\hat{\boldsymbol{\beta}}$ *y* $\hat{\sigma}_u^2$ *son variables aleatorias independientes.*

DEMOSTRACIÓN. Partiendo de $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}$, se cumple que la forma cuadrática

$$\frac{1}{\sigma_u^2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\mathbf{X}'\mathbf{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \frac{1}{\sigma_u^2}\mathbf{u}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u} = \frac{1}{\sigma_u^2}\mathbf{u}'\mathbf{P}\mathbf{u} \sim \chi_k^2$$

en donde $\mathbf{P}$ es una matriz simétrica e idempotente, cuya traza es igual a k. Por otro lado, tenemos que

$$\frac{(n-k)\hat{\sigma}_u^2}{\sigma_u^2} = \frac{1}{\sigma_u^2}\mathbf{u}'\mathbf{M}\mathbf{u} \sim \chi_{n-k}^2$$

Las distribuciones de las dos formas cuadráticas $\mathbf{u}'\mathbf{P}\mathbf{u}$ y $\mathbf{u}'\mathbf{M}\mathbf{u}$ son independientes porque $\mathbf{PM} = \mathbf{0}$. □

Los resultados que se presentan a continuación se derivan de las distribuciones muestrales de $\hat{\boldsymbol{\beta}}$ y $\hat{\sigma}_u^2$, y son de interés porque permiten construir estadísticos para contrastar hipótesis.

PROPOSICIÓN 28. *Bajo los supuestos básicos, cada elemento $\hat{\beta}_j$ del vector $\hat{\boldsymbol{\beta}}$ tiene una distribución normal con media β_j (el parámetro desconocido) y varianza $\sigma_u^2 a_{jj}$, donde a_{jj} es el j-ésimo elemento de la diagonal principal de la matriz $(\mathbf{X}'\mathbf{X})^{-1}$:*

$$\hat{\beta}_j \sim N(\beta_j, \sigma_u^2 a_{jj})$$

PROPOSICIÓN 29. *Bajo los supuestos básicos, la variable aleatoria $z_j \equiv (\hat{\beta}_j - \beta_j)/\sigma_u\sqrt{a_{jj}}$ tiene una distribución normal estándar*

$$z_j \equiv \frac{\hat{\beta}_j - \beta_j}{\sigma_u\sqrt{a_{jj}}} \sim N(0,1)$$

DEMOSTRACIÓN. z_j es una transformación lineal de $\hat{\beta}_j$ que, por la proposición 28, tiene distribución normal. Así, z_j tendrá una distribución normal con media

$$E(z_j) = E[(\hat{\beta}_j - \beta_j)/\sigma_u\sqrt{a_{jj}}] = (E[\hat{\beta}_j] - \beta_j)/\sigma_u\sqrt{a_{jj}} = 0$$

y varianza

$$V(z_j) = E[(\hat{\beta}_j - \beta_j)/\sigma_u\sqrt{a_{jj}}]^2 = E[\hat{\beta}_j - \beta_j]^2/\sigma_u^2 a_{jj} = 1$$

□

Si, en la proposición anterior, el parámetro desconocido σ_u^2 se sustituye por su estimación $\hat{\sigma}_u^2$, entonces se tiene otra variable aleatoria con una distribución de probabilidad diferente.

PROPOSICIÓN 30. *Bajo los supuestos básicos, la variable aleatoria $\tau_i \equiv (\hat{\beta}_i - \beta_i)/\hat{\sigma}_u\sqrt{a_{ii}}$ tiene una distribución t de Student con $n-k$ grados de libertad*

$$\tau_i \equiv \frac{\hat{\beta}_i - \beta_i}{\hat{\sigma}_u\sqrt{a_{ii}}} \sim t_{n-k}$$

DEMOSTRACIÓN. Se cumple que

$$\frac{\hat{\beta}_i - \beta_i}{\sigma_u\sqrt{a_{ii}}} \sim N(0,1) \quad \text{y} \quad \frac{(n-k)\hat{\sigma}_u^2}{\sigma_u^2} \sim \chi_{n-k}^2$$

Además, $\hat{\beta}_i$ y $\hat{\sigma}_u^2$ son independientes. Por tanto,

$$\frac{\dfrac{\hat{\beta}_i - \beta_i}{\sigma_u\sqrt{a_{ii}}}}{\sqrt{\dfrac{(n-k)\hat{\sigma}_u^2}{\sigma_u^2}/(n-k)}} = \frac{\hat{\beta}_i - \beta_i}{\hat{\sigma}_u\sqrt{a_{ii}}} \sim t_{n-k}$$

□

DEFINICIÓN 33. *Sea z una variable aleatoria con distribución normal estándar y sea y una variable aleatoria con distribución Chi-cuadrado con n grados de libertad, siendo z e y independientes. Entonces la variable aleatoria $x = z/\sqrt{y/n}$ tiene una distribución t de Student con n grados de libertad (Gosset 1908)*

$$x = \frac{z}{\sqrt{y/n}} \equiv \frac{N(0,1)}{\sqrt{\chi_n^2/n}} \sim t_n$$

Proposición 31. *La forma cuadrática*

$$(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})'V(\hat{\boldsymbol{\beta}})^{-1}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}) \equiv \frac{1}{\sigma_u^2}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})'\mathbf{X}'\mathbf{X}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}) \sim \chi_k^2$$

Demostración. En general, si $\mathbf{z}$ es un vector $k \times 1$ de variables aleatorias con distribución normal multivariante $N(\boldsymbol{\mu}, \boldsymbol{\Omega})$, entonces

$$(\mathbf{z}-\boldsymbol{\mu})'\boldsymbol{\Omega}^{-1}(\mathbf{z}-\boldsymbol{\mu}) \sim \chi_k^2$$

Sea la descomposición de Cholesky de la matriz definida positiva $\boldsymbol{\Omega} = \mathbf{C}\mathbf{C}'$, en donde $\mathbf{C}$ es una matriz triangular inferior invertible. Se define la variable $\mathbf{x} = \mathbf{C}^{-1}(\mathbf{z}-\boldsymbol{\mu})$. Entonces $\mathbf{x} \sim N(0, \mathbf{I}_k)$ y $\mathbf{x}'\mathbf{x} \sim \chi_k^2$. En efecto, $\mathbf{x}$ tiene una distribución normal multivariante porque es una combinación lineal de variables normales. Además, $E(\mathbf{x}) = E[\mathbf{C}^{-1}(\mathbf{z}-\boldsymbol{\mu})] = \mathbf{C}^{-1}[E(\mathbf{z})-\boldsymbol{\mu}] = \mathbf{0}$ y $Var(\mathbf{x}) = E(\mathbf{x}\mathbf{x}') = E(\mathbf{C}^{-1}(\mathbf{z}-\boldsymbol{\mu})(\mathbf{z}-\boldsymbol{\mu})'\mathbf{C}'^{-1}) = \mathbf{C}^{-1}E[(\mathbf{z}-\boldsymbol{\mu})(\mathbf{z}-\boldsymbol{\mu})']\mathbf{C}'^{-1} = \mathbf{C}^{-1}\Omega\mathbf{C}'^{-1} = \mathbf{I}_k$. □

Observación 21. La proposición 31 es la generalización al caso multivariante de la proposición 29. Para $k = 1$, se obtiene

$$(\hat{\beta}_j-\beta_j)V(\hat{\beta}_j)^{-1}(\hat{\beta}_j-\beta_j) \equiv \frac{(\hat{\beta}_j-\beta_j)^2}{\sigma_u^2 a_{jj}} \sim \chi_1^2$$

Si, en la proposición anterior, se sustituye el parámetro desconocido σ_u^2 por su estimación $\hat{\sigma}_u^2$, entonces la forma cuadrática ya no tiene una distribución χ_k^2.

Proposición 32. *La forma cuadrática*

$$\frac{1}{k}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})'\hat{V}(\hat{\boldsymbol{\beta}})^{-1}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}) \equiv \frac{1}{k\hat{\sigma}_u^2}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})'\mathbf{X}'\mathbf{X}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}) \sim F_{k,n-k}$$

Demostración. Se cumple que

$$\frac{1}{\sigma_u^2}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})'\mathbf{X}'\mathbf{X}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}) \sim \chi_k^2 \quad \text{y} \quad \frac{(n-k)\hat{\sigma}_u^2}{\sigma_u^2} \sim \chi_{n-k}^2$$

Además, ambas distribuciones son independientes. Por tanto,

$$\frac{\frac{1}{\sigma_u^2}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})'\mathbf{X}'\mathbf{X}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})/k}{\frac{(n-k)\hat{\sigma}_u^2}{\sigma_u^2}/(n-k)} = \frac{1}{k\hat{\sigma}_u^2}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})'\mathbf{X}'\mathbf{X}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}) \sim F_{k,n-k}$$

□

Observación 22. La proposición 32 es la generalización al caso multivariante de la proposición 30. Para $k = 1$, se obtiene

$$(\hat{\beta}_j-\beta_j)\hat{V}(\hat{\beta}_j)^{-1}(\hat{\beta}_j-\beta_j)/1 \equiv \frac{(\hat{\beta}_j-\beta_j)^2}{\hat{\sigma}_u^2 a_{jj}} \sim F_{1,n-k}$$

Definición 34. *Sea U una variable Chi-cuadrado con n grados de libertad y sea V una variable Chi-cuadrado con d grados de libertad independiente de U. Entonces la variable F*

$$F = \frac{U/n}{V/d}$$

tiene una <u>distribución F</u> con n y d grados de libertad, $F \sim F_{n,d}$.

Corolario 9. *Para un subvector $\hat{\boldsymbol{\beta}}_s$ de $\hat{\boldsymbol{\beta}}$, la forma cuadrática*

$$\frac{1}{s}(\hat{\boldsymbol{\beta}}_s-\boldsymbol{\beta}_s)'\hat{V}(\hat{\boldsymbol{\beta}}_s)^{-1}(\hat{\boldsymbol{\beta}}_s-\boldsymbol{\beta}_s)\equiv\frac{1}{s\hat{\sigma}_u^2}(\hat{\boldsymbol{\beta}}_s-\boldsymbol{\beta}_s)'\mathbf{X}_s'\mathbf{M}_r\mathbf{X}_s(\hat{\boldsymbol{\beta}}_s-\boldsymbol{\beta}_s)\sim F_{s,n-k}$$

4.3. El contraste t

Se desea contrastar la hipótesis nula de que un **parámetro individual** β_i es igual a un valor específico β_i^0 frente a la hipótesis alternativa de que dicho parámetro β_i es distinto de β_i^0

$$\begin{aligned} H_0 &: \quad \beta_i=\beta_i^0 \\ H_1 &: \quad \beta_i\neq\beta_i^0 \end{aligned} \tag{4.1}$$

Proposición 33. *En el problema de contraste (4.1), la hipótesis $H_0:\beta_i=\beta_i^0$ se rechaza al nivel de significación α si*

$$t_i\equiv\left|\frac{\hat{\beta}_i-\beta_i^0}{\hat{\sigma}_u\sqrt{a_{ii}}}\right|\equiv\left|\frac{\hat{\beta}_i-\beta_i^0}{\hat{dt}(\hat{\beta}_i)}\right|>c$$

en donde c es el valor crítico para el cual $Prob(t_{n-k}>c)=\alpha/2$.

Definición 35. *Cuando se rechaza H_0, se dice que el contraste es estadísticamente significativo para indicar que hay evidencia estadística en contra de H_0.*

La lógica de este contraste de dos colas es la siguiente. Parece razonable aceptar H_0 cuando la estimación $\hat{\beta}_i$ obtenida para una realización particular $\{y_1, y_2, \ldots, y_n\}$ esté cerca de β_i o, dicho de otro modo, cuando la desviación absoluta $\hat{\beta}_i-\beta_i^0$ sea pequeña en valor absoluto. Sin embargo, como esta desviación absoluta depende de las unidades en que se midan las variables del modelo de regresión, es más apropiado usar la desviación relativa $(\hat{\beta}_i-\beta_i^0)/dt(\hat{\beta}_i)$ que es adimensional. El problema pendiente es decidir cuándo la desviación relativa es grande o pequeña. Ahora bien, como desviación relativa es precisamente el estadístico t_i que, por la proposición 30, sigue una distribución t_{n-k} bajo H_0, la desviación relativa es grande cuando es poco probable, es decir, cuando en valor absoluto es mayor que c.

El gráfico 1 ilustra la elección del valor crítico c para una distribución t con 25 grados de libertad y un nivel de significación $\alpha=0{,}05$. El área bajo la curva es igual a 1 y se divide en dos regiones: una región central (región de aceptación) que tiene un área igual a $1-\alpha=0{,}95$, y una segunda región (región crítica) que comprende las dos colas sombreadas, cada una con un área igual a $\alpha/2=0{,}025$. Como la distribución es simétrica, el valor $-c$ que deja a su izquierda una probabilidad de 0.025 es igual al valor c que deja a su derecha una probabilidad de 0.025. Se puede comprobar en las tablas de la distribución t que c es igual a 2.06.

El nivel de significación α es la probabilidad del error de tipo I, es decir, la probabilidad de rechazar H_0 cuando es cierta. Dada la distribución muestral de $\hat{\beta}_i$ bajo H_0, se tiene que $\alpha=Prob(|t_i|>c)$. En la práctica, el nivel de significación se fija en $\alpha=0{,}05$ ó $\alpha=0{,}01$. Fijado α, el valor crítico c se obtiene de las tablas de la distribución t.

Observación 23. *La elección de un nivel de significación $\alpha=0{,}05$ es arbitaria, y no debe utilizarse mecánicamente. En algunos casos es posible aceptar $H_0:\beta_i=\beta_i^0$ al 5 % de significación y rechazarla a un nivel de significación ligeramente mayor. En estas*

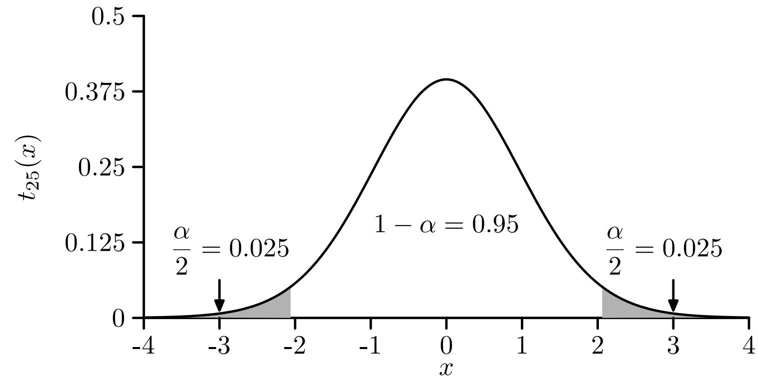

Figura 1: Función de densidad de probabilidad de la distribución t con 25 grados de libertad

situaciones, puede ser aconsejable rechazar H_0 a un nivel de significación mayor que el 5 %.

4.3.1. Contraste de significación individual.

Un caso especial del contraste t de dos colas es el contraste de significación individual:

$$\begin{aligned} H_0 :& \quad \beta_i = 0 \\ H_1 :& \quad \beta_i \neq 0 \end{aligned} \tag{4.2}$$

en donde $H_0 : \beta_i = 0$ supone eliminar la variable explicativa X_i de la ecuación de regresión.

Proposición 34. *En el problema de contraste (4.2), la hipótesis de no significación individual $H_0 : \beta_i = 0$ se rechaza al nivel de significación α si*

$$t_i \equiv \left| \frac{\hat{\beta}_i}{\hat{\sigma}_u \sqrt{a_{ii}}} \right| \equiv \left| \frac{\hat{\beta}_i}{\hat{dt}(\hat{\beta}_i)} \right| > c$$

en donde c es el valor crítico para el cual $Prob(t_{n-k} > c) = \alpha/2$.

Corolario 10. *Cuando el número de datos es grande ($n - k > 30$), la hipótesis nula de no significación individual se rechaza al nivel de significación del 5 % si*

$$t_i \equiv \left| \frac{\hat{\beta}_i}{\hat{\sigma}_u \sqrt{a_{ii}}} \right| \equiv \left| \frac{\hat{\beta}_i}{\hat{dt}(\hat{\beta}_i)} \right| > 2$$

en donde 2 es el valor crítico para el cual $Prob(t_{n-k} > 2) \simeq 0{,}025$ y $n - k > 30$.

Definición 36. *El p-valor del estadístico t_i es la probabilidad del suceso $|t_{n-k}| > |t_i|$, $P(|t_{n-k}| > |t_i|)$. Se interpreta como el nivel de significación mínimo al que se rechaza la H_0. Si el p-valor es mayor que el nivel de significación α, entonces $|t_i| < c$ y se acepta H_0; por el contrario, si el p-valor es menor que el nivel de significación α, entonces $|t_i| > c$ y se rechaza H_0.*

Ejemplo 3. Si el p-valor en el contraste de $H_0 : \beta_i = 0$ frente $H_0 : \beta_i \neq 0$ es 0.003, entonces podemos rechazar H_0 al nivel de significación del 5 %. En cambio, si el p-valor es igual a 0.20, no se rechaza H_0 al nivel de significación del 5 %.

◁

4.3.2. Intervalo de confianza para β_i.

Un método equivalente al contraste t de dos colas es el intervalo de confianza.

Definición 37. *Un intervalo de confianza del* $100(1-\alpha)$ *por ciento para el parámetro* β_i *es el conjunto de hipotesis nulas sobre* β_i *que no se rechazan al nivel de significación* α.

Un intervalo de confianza se construye a partir de la probabilidad del error de tipo I

$$Prob(\left|\frac{\hat{\beta}_i - \beta_i}{\hat{\sigma}_u\sqrt{a_{ii}}}\right| > c) = \alpha$$

que puede escribirse como

$$Prob(-c < \frac{\hat{\beta}_i - \beta_i}{\hat{\sigma}_u\sqrt{a_{ii}}} < c) = 1-\alpha$$

o bien

$$Prob(\hat{\beta}_i - c\hat{\sigma}_u\sqrt{a_{ii}} < \beta_i < \hat{\beta}_i + c\hat{\sigma}_u\sqrt{a_{ii}}) = 1-\alpha$$

Esta ecuación indica la probabilidad de que el valor β_i pertenezca al intervalo aleatorio $(\hat{\beta}_i - c\hat{\sigma}_u\sqrt{a_{ii}},\ \hat{\beta}_i + c\hat{\sigma}_u\sqrt{a_{ii}})$. Dada una realización particular, se obtiene una estimación del intervalo aleatorio, que se denomina intervalo de confianza del $100(1-\alpha)$ por ciento.

Proposición 35. *Cuando el número de observaciones es grande* ($n-k > 30$)*, el intervalo de confianza puede aproximarse por*

$$[\hat{\beta}_i - 2\hat{dt}(\hat{\beta}_i),\ \hat{\beta}_i + 2\hat{dt}(\hat{\beta}_i)]$$

4.3.3. El contrate t de una cola.

Proposición 36. *En el problema de contraste* $H_0 : \beta_i \leq \beta_i^0$ *frente* $H_1 : \beta_i > \beta_i^0$ *, se rechaza* H_0 *al nivel de significación* α *si*

$$t_i \equiv \frac{\hat{\beta}_i - \beta_i^0}{\hat{\sigma}_u\sqrt{a_{ii}}} \equiv \frac{\hat{\beta}_i - \beta_i^0}{\hat{dt}(\hat{\beta}_i)} > c$$

en donde c *es el valor crítico para el cual* $Prob(t_{n-k} > c) = \alpha$.

El intervalo de confianza de un lado equivalente al constraste t de dos colas se construye a partir de la probabilidad de error de tipo I

$$Prob(\frac{\hat{\beta}_i - \beta_i}{\hat{\sigma}_u\sqrt{a_{ii}}} > c) = \alpha$$

que puede escribirse como

$$Prob(\hat{\beta}_i - c \times \hat{\sigma}_u\sqrt{a_{ii}} < \beta_i < \infty) = 1-\alpha$$

Esta ecuación indica la probabilidad de que el parámetro β_i pertenezca al intervalo aleatorio $(\hat{\beta}_i - c \times \hat{\sigma}_u\sqrt{a_{ii}}, \infty)$. La hipótesis $H_0 : \beta_i = \beta_i^0$ se rechaza si β_i^0 no pertenece al intervalo de confianza.

4.4. El contraste F

Se desea contrastar la hipótesis nula de que un subvector de s coeficientes $\boldsymbol{\beta}_s$ es igual a $\boldsymbol{\beta}_s^0$ frente a la hipótesis alternativa de que $\boldsymbol{\beta}_s$ es distinto de $\boldsymbol{\beta}_s^0$

$$\begin{aligned} H_0 &: \quad \boldsymbol{\beta}_s = \boldsymbol{\beta}_s^0 \\ H_1 &: \quad \boldsymbol{\beta}_s \neq \boldsymbol{\beta}_s^0 \end{aligned} \tag{4.3}$$

Proposición 37. *En el problema de contraste (4.3), la hipótesis* $H_0 : \boldsymbol{\beta}_s = \boldsymbol{\beta}_s^0$ *se rechaza al nivel de significación* α *si*

$$F \equiv (\hat{\boldsymbol{\beta}}_s - \boldsymbol{\beta}_s^0)'\hat{V}(\hat{\boldsymbol{\beta}}_s)^{-1}(\hat{\boldsymbol{\beta}}_s - \boldsymbol{\beta}_s^0)/s > c$$

en donde c *es el valor crítico para el cual* $Prob(F_{s,n-k} > c) = \alpha$.

La lógica del contraste F es similar a la del contraste t. Parece razonable aceptar H_0 cuando el subvector de estimaciones $\hat{\boldsymbol{\beta}}_s$ obtenidas en una realización particular $\{y_1, y_2, \ldots, y_n\}$ esté próximo al subvector de parámetros $\boldsymbol{\beta}_s^0$. La distancia euclídea al cuadrado entre los vectores $\hat{\boldsymbol{\beta}}_s$ y $\boldsymbol{\beta}_s^0$ es el producto escalar del vector $(\hat{\boldsymbol{\beta}}_s - \boldsymbol{\beta}_s^0)$, es decir,

$$(\hat{\boldsymbol{\beta}}_s - \boldsymbol{\beta}_s^0)'(\hat{\boldsymbol{\beta}}_s - \boldsymbol{\beta}_s^0)$$

Esta medida de distancia depende de las unidades en que se midan las variables. Una medida de distancia adimensional es

$$(\hat{\boldsymbol{\beta}}_s - \boldsymbol{\beta}_s^0)'\hat{V}(\hat{\boldsymbol{\beta}}_s)^{-1}(\hat{\boldsymbol{\beta}}_s - \boldsymbol{\beta}_s^0)$$

De aquí, se rechazará H_0 cuando esta medida de distancia relativa sea grande. La pregunta que surge ahora es qué se entiende por grande. La respuesta la proporciona la distribución de probabilidad del estadístico. Como bajo H_0 la medida de distancia (dividida por s) sigue una distribución χ_s^2, se considera que la distancia es grande cuando es poco probable; o dicho de otro modo, el estadístico F es grande cuando en valor absoluto es mayor que c.

El gráfico 2 ilustra la elección del valor crítico c para una distribución F con 5 grados de libertad en el numerador, 30 grados de libertad en el denominador y un nivel de significación $\alpha = 0{,}05$. El área bajo la curva es igual a 1 y se divide en dos regiones: una región a la izquierda de c (región de aceptación) que tiene un área igual a $1 - \alpha = 0{,}95$, y una región a la derecha de c (región crítica) un área igual a $\alpha = 0{,}05$. Se puede comprobar en las tablas de la distribución $F_{5,30}$ que c es igual a 2.525.

Ahora bien, dada la distribución muestral de $\hat{\boldsymbol{\beta}}_s$ es probable que $F > c$ incluso cuando $H_0 : \boldsymbol{\beta}_s = \boldsymbol{\beta}_s^0$ es cierta. Por tanto, al utilizar la regla de decisión $F > c$ se pueden cometer dos tipos de error. El error de tipo I consiste en rechazar H_0 cuando es cierta, mientras que el error de tipo II es no rechazar H_0 cuando es falsa. La probabilidad del error de tipo I es $Prob(|F| > c)$, que es la probabilidad de que una variable aleatoria con distribución $F_{s,n-k}$ tome un valor mayor que c. El valor crítico c se elige para que el contraste tenga una probabilidad de error de tipo I o nivel de significación igual a un valor específico α

$$Prob(F_{s,n-k} > c) = \alpha$$

En la práctica el nivel de significación $\alpha = 0{,}05$ o $\alpha = 0{,}01$. Fijado α, el valor crítico c se obtiene de las tablas de la distribución F.

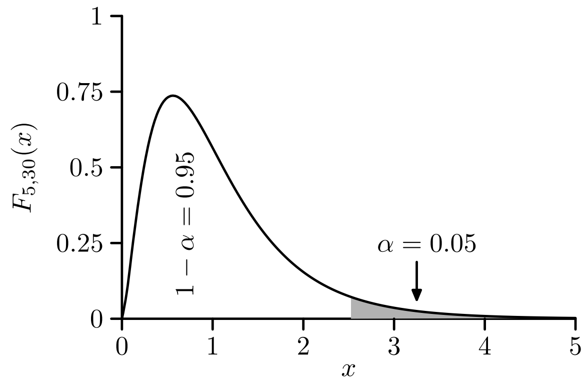

Figura 2: Función de densidad de probabilidad de la distribución F con 5 y 30 grados de libertad

4.4.1. Constrate de significación conjunta.

Un caso especial del contraste F es

$$H_0: \quad \boldsymbol{\beta}_s = \mathbf{0}_s$$

$$H_1: \quad \boldsymbol{\beta}_s \neq \mathbf{0}_s$$

en donde la hipótesis nula $H_0 : \boldsymbol{\beta}_s = \mathbf{0}_s$ conlleva la eliminación de s variables explicativas de la ecuación de regresión.

Proposición 38. *La hipótesis de no significación conjunta $H_0 : \boldsymbol{\beta}_s = \mathbf{0}_s$ se rechaza al nivel de significación α si*

$$F \equiv \hat{\boldsymbol{\beta}}_s' \hat{V}(\hat{\boldsymbol{\beta}}_s)^{-1} \hat{\boldsymbol{\beta}}_s / s > c$$

en donde c es el valor crítico para el cual $Prob(F_{s,n-k} > c) = \alpha$.

Hay otras formas más convenientes de realizar el contraste de significación conjunta.

Proposición 39. *La hipótesis de no significación conjunta $H_0 : \boldsymbol{\beta}_s = \mathbf{0}_s$ se rechaza al nivel de significación α si*

$$F \equiv \frac{(\hat{\mathbf{u}}_r' \hat{\mathbf{u}}_r - \hat{\mathbf{u}}' \hat{\mathbf{u}})/s}{\hat{\mathbf{u}}' \hat{\mathbf{u}}/(n-k)} > c$$

en donde c es el valor crítico para el cual $Prob(F_{s,n-k} > c) = \alpha$, $\hat{\mathbf{u}}'\hat{\mathbf{u}}$ es la suma de cuadrados de los residuos en la regresión de $\mathbf{Y}$ sobre $\mathbf{X}$, y $\hat{\mathbf{u}}_r'\hat{\mathbf{u}}_r$ es la suma de cuadrados de los residuos en la regresión de $\mathbf{Y}$ sobre $\mathbf{X}_r$.

Demostración. Como $V(\hat{\boldsymbol{\beta}}_s) = \sigma_u^2 (\mathbf{X}_s' \mathbf{M}_r \mathbf{X}_s)^{-1}$, el estadístico

$$F \equiv \hat{\boldsymbol{\beta}}_s' \hat{V}(\hat{\boldsymbol{\beta}}_s)^{-1} \hat{\boldsymbol{\beta}}_s / s$$

puede escribirse como

$$F \equiv \frac{1}{s\hat{\sigma}_u^2} \hat{\boldsymbol{\beta}}_s' \mathbf{X}_s' \mathbf{M}_r \mathbf{X}_s \hat{\boldsymbol{\beta}}_s$$

Ahora bien,

$$\hat{\boldsymbol{\beta}}_s' \mathbf{X}_s' \mathbf{M}_r \mathbf{X}_s \hat{\boldsymbol{\beta}}_s = \hat{\mathbf{y}}' \mathbf{M}_r \hat{\mathbf{y}} - \hat{\mathbf{u}}' \hat{\mathbf{u}} = \hat{\mathbf{u}}_r' \hat{\mathbf{u}}_r - \hat{\mathbf{u}}' \hat{\mathbf{u}}$$

y

$$\hat{\sigma}_u^2 = \frac{\hat{\mathbf{u}}' \hat{\mathbf{u}}}{n-k}$$

Luego

$$F \equiv \frac{1}{s\hat{\sigma}_u^2}\hat{\boldsymbol{\beta}}_s'\mathbf{X}_s'\mathbf{M}_r\mathbf{X}_s\hat{\boldsymbol{\beta}}_s = \frac{(\hat{\mathbf{u}}_r'\hat{\mathbf{u}}_r - \hat{\mathbf{u}}'\hat{\mathbf{u}})/s}{\hat{\mathbf{u}}'\hat{\mathbf{u}}/(n-k)}$$

□

Para realizar el contraste de significación conjunta se siguen los siguientes pasos:

1. Estimar el modelo de regresión

$$\mathbf{y} = \mathbf{X}_r\boldsymbol{\beta}_r + \mathbf{X}_s\boldsymbol{\beta}_s + \mathbf{u}$$

y calcular la suma de cuadrados de los residuos, $\hat{\mathbf{u}}'\hat{\mathbf{u}}$.

2. Estimar el modelo de regresión bajo $H_0 : \boldsymbol{\beta}_s = \mathbf{0}_s$

$$\mathbf{y} = \mathbf{X}_r\boldsymbol{\beta}_r + \mathbf{u}_r$$

y calcular la suma de cuadrados de los residuos, $\hat{\mathbf{u}}_r'\hat{\mathbf{u}}_r$.

3. Calcular el estadístico de contraste

$$F = \frac{(\hat{\mathbf{u}}_r'\hat{\mathbf{u}}_r - \hat{\mathbf{u}}'\hat{\mathbf{u}})/s}{\hat{\mathbf{u}}'\hat{\mathbf{u}}/(n-k)}$$

4. Comparar F con el valor crítico c de la distribución $F_{s,n-k}$ al nivel de significación α. Si $F < c$, aceptamos H_0; si $F > c$, rechazamos H_0.

Es conveniente notar que si $H_0 : \boldsymbol{\beta}_s = \mathbf{0}_s$ es cierta, la disminución en la suma de cuadrados de los residuos, $\hat{\mathbf{u}}_r'\hat{\mathbf{u}}_r - \hat{\mathbf{u}}'\hat{\mathbf{u}}$, que se produce al incluir las variables $\mathbf{X}_s$, será pequeña y el estadístico F estará cerca de cero. Por el contrario, si $H_0 : \boldsymbol{\beta}_s = \mathbf{0}_s$ es falsa, la disminución en la suma de cuadrados de los residuos, $\hat{\mathbf{u}}_r'\hat{\mathbf{u}}_r - \hat{\mathbf{u}}'\hat{\mathbf{u}}$, que se produce al incluir las variables $\mathbf{X}_s$, será grande y el estadístico F estará lejos de cero. Note que siempre $\hat{\mathbf{u}}_r'\hat{\mathbf{u}}_r \geq \hat{\mathbf{u}}'\hat{\mathbf{u}}$ y $F \geq 0$, ¿porqué?

4.4.2. Contraste de significación global.

Otro caso especial del contraste F es

$$\begin{aligned} H_0 : \quad & \boldsymbol{\beta}_s = \mathbf{0}_s \\ H_1 : \quad & \boldsymbol{\beta}_s \neq \mathbf{0}_s \end{aligned}$$

en donde el subvector $\boldsymbol{\beta}_s = (\beta_2 \ \beta_3 \ \ldots \ \beta_k)'$ incluye todos los coeficientes del modelo salvo el término constante, y $s = k - 1$.

Los pasos a seguir para contrastar esta hipótesis son los siguientes:

1. Estimar el modelo de regresión

$$Y_i = \beta_1 + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i, \quad i = 1, \ldots, n$$

y calcular la suma de cuadrados de los residuos, $\hat{\mathbf{u}}'\hat{\mathbf{u}} = SCR$.

2. Estimar el modelo de regresión bajo $H_0 : \beta_2 = \cdots = \beta_k = 0$

$$Y_i = \beta_1 + u_i, \quad i = 1, \ldots, n$$

y calcular la suma de cuadrados de los residuos, $\hat{\mathbf{u}}_r'\hat{\mathbf{u}}_r = \sum_{i=1}^n (Y_i - \bar{Y})^2 = SCT$.

3. Calcular el estadístico de contraste

$$F = \frac{(SCT - SCR)/(k-1)}{SCR/(n-k)}$$

4. Comparar F con el valor crítico c de la distribución $F_{k-1,n-k}$ al nivel de significación α. Si $F < c$, aceptamos H_0; si $F > c$, rechazamos H_0.

El contraste de significación global se resume en el cuadro 2, en donde la variación de la variable dependiente (SCT) se descompone en la explicada por la regresión (SCE) y en la no explicada (SCR). Los grados de libertad de estas tres sumas de cuadrados son $n-k$, $k-1$ y $n-k$, respectivamente. A partir de esta información muestral, podemos calcular numerador y el denominador del estadístico F.

Cuadro 2: Análisis de varianza en el modelo lineal general

Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrado medio	Estadístico F
Regresión	$\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2$	$k-1$	$SCE/(k-1)$	$[SCE/(k-1)]/[SCR/(n-k)]$
Residual	$\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$	$n-k$	$SCR/(n-k)$	
Total	$\sum_{i=1}^{n}(Y_i - \bar{Y})^2$	n-1		

Proposición 40. *La hipótesis de no significación global* $H_0 : \boldsymbol{\beta}_s = \mathbf{0}_s$ *y* $s = k-1$ *se rechaza al nivel de significación* α *si*

$$F \equiv \frac{R^2/(k-1)}{(1-R^2)/(n-k)} > c$$

en donde c *es el valor crítico para el cual* $Prob(F_{k-1,n-k} > c) = \alpha$, R^2 *es el coeficiente de determinación en la regresión de* $\mathbf{y}$ *sobre* $\mathbf{X}$.

Demostración. Ahora $\mathbf{X}_r = \mathbf{i}$ es un vector de unos, y $\mathbf{M}_r = \mathbf{M}_i$ es la matriz que transforma un vector de observaciones en un vector de observaciones en desviaciones. Luego,

$$\hat{\mathbf{u}}_r'\hat{\mathbf{u}}_r = \hat{\mathbf{y}}'\mathbf{M}_r\hat{\mathbf{y}} = \hat{\mathbf{y}}'\mathbf{M}_i\hat{\mathbf{y}} = \sum_{i=1}^{n}(y_i - \bar{y})^2$$

es la suma de cuadrados total. Por tanto,

$$F = \frac{(\hat{\mathbf{u}}_r'\hat{\mathbf{u}}_r - \hat{\mathbf{u}}'\hat{\mathbf{u}})/s}{\hat{\mathbf{u}}'\hat{\mathbf{u}}/(n-k)} = \frac{(SCT - SCR)/(k-1)}{SCR/(n-k)} = \frac{(1 - \dfrac{SCR}{SCT})/(k-1)}{\dfrac{SCR}{SCT}/(n-k)} = \frac{R^2/(k-1)}{(1-R^2)/(n-k)}$$

□

Observación 24. No debe confundirse la hipótesis de significación global $H_0 : \beta_2 = \cdots = \beta_k = 0$ *con la hipótesis* $H_0 : \beta_1 = \beta_2 = \cdots = \beta_k = 0$ *que es equivalente a* $H_0 : \bar{Y} = 0$ *porque* $\bar{Y} = \hat{\beta}_1 + \hat{\beta}_2\bar{X}_2 + \cdots + \hat{\beta}_k\bar{X}_k$.

4.4.3. Región de confianza para $\boldsymbol{\beta}_s$.

Un método equivalente al contraste F es la region de confianza.

Definición 38. *Una región de confianza del* $100(1-\alpha)$ *por ciento para el subvector* $\boldsymbol{\beta}_s$ *es el conjunto de hipotesis nulas sobre* $\boldsymbol{\beta}_s$ *que no se rechazan al nivel de significación* α.

Para construir una región de confianza partimos de la probabilidad del error de tipo I

$$Prob\left[(\hat{\boldsymbol{\beta}}_s - \boldsymbol{\beta}_s)'\hat{V}(\hat{\boldsymbol{\beta}}_s)^{-1}(\hat{\boldsymbol{\beta}}_s - \boldsymbol{\beta}_s)/s > c\right] = \alpha$$

que podemos escribir como

$$Prob\left[(\hat{\boldsymbol{\beta}}_s - \boldsymbol{\beta}_s)'\hat{V}(\hat{\boldsymbol{\beta}}_s)^{-1}(\hat{\boldsymbol{\beta}}_s - \boldsymbol{\beta}_s) < sc\right] = 1 - \alpha$$

Esta ecuación indica la probabilidad de que el subvector $\boldsymbol{\beta}$ pertenezca a la región aleatoria

$$(\hat{\boldsymbol{\beta}}_s - \boldsymbol{\beta}_s)'\hat{V}(\hat{\boldsymbol{\beta}}_s)^{-1}(\hat{\boldsymbol{\beta}}_s - \boldsymbol{\beta}_s) < sc$$

Dada una muestra de observaciones, se obtiene una estimación de la región aleatoria, que se denomina región de confianza del $100(1-\alpha)$ por ciento.

4.5. Intervalo de confianza para σ_u^2

Para construir este intervalo de confianza partimos de la probabilidad

$$Prob(c_1 < \frac{(n-k)\hat{\sigma}_u^2}{\sigma_u^2} < c_2) = 1 - \alpha$$

que podemos escribir como

$$Prob(\frac{(n-k)\hat{\sigma}_u^2}{c_2} < \sigma_u^2 < \frac{(n-k)\hat{\sigma}_u^2}{c_1}) = 1 - \alpha$$

en donde los cuantiles c_1 y c_2 son tales que $Prob(\chi^2_{n-k} < c_1) = \alpha/2$ y $Prob(\chi^2_{n-k} > c_2) = \alpha/2$. Esta ecuación indica la probabilidad de que el valor σ_u^2 pertenezca al intervalo aleatorio $[(n-k)\hat{\sigma}_u^2/c_2,\ (n-k)\hat{\sigma}_u^2/c_1]$. Dada una muestra de observaciones, se obtiene una estimación del intervalo aleatorio, que se denomina intervalo de confianza del $100(1-\alpha)$ por ciento.

4.6. La hipótesis lineal general

La hipótesis lineal general especifica un conjunto de relaciones lineales entre los parámetros del modelo de regresión lineal.

DEFINICIÓN 39. *La hipótesis lineal general tiene la forma*

$$\begin{aligned} H_0: &\quad \mathbf{R}\boldsymbol{\beta} = \mathbf{r} \\ H_1: &\quad \mathbf{R}\boldsymbol{\beta} \neq \mathbf{r} \end{aligned}$$

en donde $\mathbf{R}$ *es una matriz conocida de orden* $q \times k$ *y rango* $q \leq k$, *y* $\mathbf{r}$ *es un vector conocido de orden* $q \times 1$.

EJEMPLO 4. En el modelo de regresión múltiple

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + u_i$$

se desea contrastar conjuntamente las siguientes hipótesis

$$\begin{aligned} \beta_1 &= 3 \\ 3\beta_2 + 5\beta_3 &= 8 \\ 2\beta_2 + 8\beta_4 &= 12 \end{aligned}$$

En forma matricial, las tres hipótesis pueden expresarse como

$$\underset{\mathbf{R}}{\begin{pmatrix}1 & 0 & 0 & 0\\ 0 & 3 & 5 & 0\\ 0 & 2 & 0 & 8\end{pmatrix}} \underset{\boldsymbol{\beta}}{\begin{pmatrix}\beta_1\\ \beta_2\\ \beta_3\\ \beta_4\end{pmatrix}} \underset{=}{=} \underset{\mathbf{r}}{\begin{pmatrix}3\\ 8\\ 12\end{pmatrix}}$$

◁

Observación 25. El número de restricciones lineales debe ser menor o igual que el número de coeficientes, $q \leq k$. Si $q > k$ entonces algunas restricciones lineales estarían repetidas y serían redundantes.

La hipótesis lineal general reduce el número de paramétros a estimar de k a $k - q$. En el ejemplo 4, sólo es necesario estimar β_4. Una vez estimado β_4, la tercera restricción nos permite estimar β_2; una vez estimado β_2, la segunda restricción nos permite estimar β_3. La estimación de β_1 está dada por la primera restricción. Esto sugiere particionar la hipótesis lineal general del siguiente modo

$$\mathbf{R}_1\boldsymbol{\beta}_1 + \mathbf{R}_2\boldsymbol{\beta}_2 = r$$

en donde $\mathbf{R}_1$ es una matriz cuadrada $q \times q$ y $\mathbf{R}_2$ es una matriz rectangular $q \times (k - q)$. Si las restricciones son independientes, la matriz $\mathbf{R}_1$ será no singular y podemos expresar $\boldsymbol{\beta}_1$ en términos de $\boldsymbol{\beta}_2$

$$\boldsymbol{\beta}_1 = \mathbf{R}_1^{-1}r - \mathbf{R}_1^{-1}\mathbf{R}_2\boldsymbol{\beta}_2$$

EJEMPLO 5. Las restricciones lineales del ejemplo 4 en forma particionada

$$\begin{pmatrix}1 & 0 & 0\\ 0 & 3 & 5\\ 0 & 2 & 0\end{pmatrix}\begin{pmatrix}\beta_1\\ \beta_2\\ \beta_3\end{pmatrix} + \begin{pmatrix}0\\ 0\\ 8\end{pmatrix}(\beta_4) = \begin{pmatrix}3\\ 8\\ 12\end{pmatrix}$$

implican que

$$\begin{pmatrix}\beta_1\\ \beta_2\\ \beta_3\end{pmatrix} = \begin{pmatrix}3\\ 6\\ -2\end{pmatrix} - \begin{pmatrix}0\\ 4\\ -2{,}4\end{pmatrix}\beta_4$$

◁

4.6.1. Mínimos cuadrados restringidos.

La estimación del modelo clásico sujeto a un conjunto de restricciones lineales puede llevarse a cabo de dos formas equivalentes: (1) incorporando las restricciones en la ecuación y (2) aplicando la fórmula general del estimador de mínimos cuadrados restringidos. Mientras que la forma (1) es útil en aplicaciones prácticas cuando se utiliza un programa de ordenador con capacidad para el análisis de regresión, la forma (2) es interesante para derivar las propiedades estadísticas generales del estimador.

DEFINICIÓN 40. *El modelo que se obtiene al incorporar la la hipótesis lineal H_0 : $\mathbf{R}\boldsymbol{\beta} - \mathbf{r} = \mathbf{0}$ en $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$ se denomina modelo con restricciones o modelo restringido.*

Para incorporar la hipótesis lineal general en el modelo de regresión, usamos la partición

$$\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + u$$

en donde $\boldsymbol{\beta}_1$ es un vector $q \times 1$ de parámetros que pueden obtenerse a partir de los parámetros del vector $(k - q) \times 1$ β_2. Sustituyendo la expresión de $\boldsymbol{\beta}_1$ en el modelo de regresión tenemos que

$$\mathbf{y} - \mathbf{X}_1\mathbf{R}_1^{-1}r = (\mathbf{X}_2 - \mathbf{X}_1\mathbf{R}_1^{-1}\mathbf{R}_2)\boldsymbol{\beta}_2 + u$$

en donde $\mathbf{y} - \mathbf{X}_1\mathbf{R}_1^{-1}r$ es la nueva variable dependiente y $(\mathbf{X}_2 - \mathbf{X}_1\mathbf{R}_1^{-1}\mathbf{R}_2)$ son las nuevas variables explicativas. La estimación de este modelo transformado proporciona las estimaciones con restricciones de β_2, las cuales permiten obtener las estimaciones con restricciones de $\boldsymbol{\beta}_1$.

EJEMPLO 6. En economía la función de producción Cobb-Douglas es utilizada frecuentemente para expresar que el producto es una función del trabajo y del capital

$$\log Y_i = \beta_1 + \beta_2 \log L_i + \beta_3 \log K_i + u_i$$

El supuesto de rendimientos constantes a escala implica que $\beta_2 + \beta_3 = 1$: si el trabajo y el capital aumentan un 5 % entonces el producto aumenta también un 5 %.

Para estimar una función de producción con rendimientos constantes a escala, incorporamos la restricción en la ecuación

$$\log Y_i = \beta_1 + (1 - \beta_3) \log L_i + \beta_3 \log K_i + u_i$$

Reordenando obtenemos

$$\log \frac{Y_i}{L_i} = \beta_1 + \beta_3 \log \frac{K_i}{L_i} + u_i$$

◁

EJEMPLO 7. En el modelo de regresión múltiple

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + u_i$$

si se impone la $H_0 : \beta_2 + \beta_3 = 0$ se obtiene el modelo restringido

$$\begin{aligned} Y_i =& \beta_1 + \beta_2 X_{2i} - \beta_2 X_{3i} + \beta_4 X_{4i} + u_i \\ Y_i =& \beta_1 + \beta_2 (X_{2i} - X_{3i}) + \beta_4 X_{4i} + u_i \end{aligned}$$

◁

PROPOSICIÓN 41. *El estimador de mínimos cuadrados sujeto al conjunto de restricciones lineales* $\mathbf{R}\hat{\boldsymbol{\beta}}_* = \mathbf{r}$ *es*

$$\hat{\boldsymbol{\beta}}_* = \hat{\boldsymbol{\beta}} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})$$

DEMOSTRACIÓN. Siguiendo el método de los multiplicadores de Langrage, especificamos primero el lagrangiano

$$Q = (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_*)'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_*) + 2\boldsymbol{\lambda}'(\mathbf{R}\hat{\boldsymbol{\beta}}_* - \mathbf{r})$$

en donde $\boldsymbol{\lambda}$ es un vector $q \times 1$ de multiplicadores de Lagrange, que aparece multiplicado por 2 para simplificar los desarrollos posteriores. Las condiciones de primer order son

$$\begin{aligned} \frac{\partial Q}{\partial \hat{\boldsymbol{\beta}}_*} =& -2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}}_* - 2\mathbf{R}'\boldsymbol{\lambda} = \mathbf{0}_k \\ \frac{\partial Q}{\partial \boldsymbol{\lambda}} =& 2(\mathbf{R}\hat{\boldsymbol{\beta}}_* - \mathbf{r}) = \mathbf{0}_q \end{aligned}$$

Para encontrar las expresiones de $\hat{\boldsymbol{\beta}}_*$ y $\boldsymbol{\lambda}$ podemos resolver el sistema de ecuaciones matriciales

$$\begin{pmatrix} \mathbf{X}'\mathbf{X} & \mathbf{R}' \\ \mathbf{R} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \hat{\boldsymbol{\beta}}_* \\ \boldsymbol{\lambda} \end{pmatrix} = \begin{pmatrix} \mathbf{X}'\mathbf{y} \\ r \end{pmatrix}$$

premultiplicando por la inversa de la matriz particionada asociada al vector de coeficientes

$$\begin{pmatrix} (\mathbf{X}'\mathbf{X})^{-1} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'(\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}')^{-1}\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1} & (\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'(\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}')^{-1} \\ (\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}')^{-1}\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1} & -(\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}')^{-1} \end{pmatrix}$$

□

PROPOSICIÓN 42. *La suma de cuadrados de los residuos con restricciones es*

$$\hat{\mathbf{u}}_*'\hat{\mathbf{u}}_* = \hat{\mathbf{u}}'\hat{\mathbf{u}} + (\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_*)'\mathbf{X}'\mathbf{X}(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_*)'$$

Demostración. El vector de residuos asociado al estimador $\hat{\boldsymbol{\beta}}_*$ es

$$\hat{\mathbf{u}}_* = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_*$$

Sumando y restando $\mathbf{X}\hat{\boldsymbol{\beta}}$, tenemos

$$\hat{\mathbf{u}}_* = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{X}\hat{\boldsymbol{\beta}}_* = \hat{\mathbf{u}} + (\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_*)\mathbf{X}$$

De aquí, es fácil encontrar la expresión dada para la suma de cuadrados de los residuos, recordando que $\mathbf{X}'\hat{\mathbf{u}} = \mathbf{0}$. □

Corolario 11. *La suma de cuadrados con restricciones será mayor o igual que la suma de cuadrados de los residuos sin restricciones.*

Demostración. Es claro que la forma cuadrática $(\hat{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}}_*)'\mathbf{X}'\mathbf{X}(\hat{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}}_*)$ es semidefinida positiva, por lo que $\hat{\mathbf{u}}_*'\hat{\mathbf{u}}_*$ es igual a $\hat{\mathbf{u}}'\hat{\mathbf{u}}$ más una magnitud no negativa. Intuitivamente, podemos notar que $\hat{\mathbf{u}}'\hat{\mathbf{u}}$ es la menor suma de cuadrados de los residuos que puede alcanzar un estimador lineal, mientras que $\hat{\mathbf{u}}_*'\hat{\mathbf{u}}_*$ es la menor suma de cuadrados de los residuos que puede alcanzar un estimador lineal que cumple las restricciones lineales. □

Proposición 43. *El aumento de la suma de cuadrados de los residuos en la estimación con restricciones es*

$$\hat{\mathbf{u}}_*'\hat{\mathbf{u}}_* - \hat{\mathbf{u}}'\hat{\mathbf{u}} = (\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})$$

Demostración. De la definición de $\hat{\boldsymbol{\beta}}_*$ vemos que

$$\mathbf{X}(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_*) = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})$$

La suma de cuadrados de este vector columna es

$$(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_*)'\mathbf{X}'\mathbf{X}(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_*)$$

o bien

$$(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}\underbrace{\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}}_{\mathbf{I}_q}(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})$$

□

4.6.2. El contraste de la hipótesis lineal general.

Proposición 44. *La hipótesis $H_0 : \mathbf{R}\boldsymbol{\beta} - \mathbf{r} = \mathbf{0}$ se rechaza al nivel de significación α si*

$$F \equiv [\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}]'[\hat{\sigma}_u^2\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}[\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}]/q > c$$

en donde c es el valor crítico para el cual $Prob(F_{q,n-k} > c) = \alpha$.

Observación 26. Definiendo el vector $\mathbf{d} \equiv \mathbf{R}\boldsymbol{\beta} - \mathbf{r}$ de orden $q \times 1$ y su estimador $\hat{\mathbf{d}} \equiv \mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}$, vemos que el estadístico de contraste para la hipótesis $H_0 : \mathbf{d} = \mathbf{0}_q$ es equivalente al discutido para la hipótesis $H_0 : \boldsymbol{\beta}_s = \mathbf{0}$. Es claro, por tanto, que el estadístico de contraste $F \equiv \hat{\mathbf{d}}'\hat{V}(\hat{\mathbf{d}})^{-1}\hat{\mathbf{d}}/q$ tiene una distribución $F_{q,n-k}$. La explicación lógica del contraste de restricciones lineales es similar a la del contraste F.

La proposición 42 sugiere una forma alternativa del contraste de restricciones lineales.

Proposición 45. *En el contraste de la hipótesis* $H_0 : \mathbf{R}\boldsymbol{\beta} - \mathbf{r} = \mathbf{0}$ *frente a* $H_1 : \mathbf{R}\boldsymbol{\beta} - \mathbf{r} \neq \mathbf{0}$, *se rechaza* H_0 *al nivel de significación* α *si*

$$F \equiv \frac{(SCRCR - SCRSR)/(GLCR - GLSR)}{SCRSR/GLSR} > c$$

en donde c *es el valor crítico para el cual* $Prob(F_{GLCR-GLSR,GLSR} > c) = \alpha$, $SCRCR$ *y* $GLCR$ *son la suma de cuadrados de los residuos y los grados de libertad en el modelo con restricciones,* $SCRSR$ *y* $GLSR$ *son los grados de libertad en el modelo sin restricciones.*

Para realizar el contraste de restricciones lineales se siguen los siguientes pasos:

1. Se estima el modelo sin restricciones
$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$$
y se calcula la suma de cuadrados de los residuos, $\hat{\mathbf{u}}'\hat{\mathbf{u}}$, y sus grados de libertad, $n - k$.
2. Se estima el modelo con restricciones
$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}_* \quad \mathbf{R}\boldsymbol{\beta} - \mathbf{r} = \mathbf{0}$$
y se calcula la suma de cuadrados de los residuos, $\hat{\mathbf{u}}_*'\hat{\mathbf{u}}_*$, y sus grados de libertad, $n - (k - q)$.
3. Se calcula el estadístico de contraste
$$F = \frac{(\hat{\mathbf{u}}_*'\hat{\mathbf{u}}_* - \hat{\mathbf{u}}'\hat{\mathbf{u}})/q}{\hat{\mathbf{u}}'\hat{\mathbf{u}}/(n-k)}$$
4. Se compara el valor del estadístico F con el valor crítico c para el cual $Prob(F_{q,n-k} > c) = \alpha$. Si $F < c$, se acepta H_0; si $F > c$, se rechaza H_0.

Dos **casos especiales** del contraste de restricciones lineales son:

1. $H_0 : \beta_i = \beta_i^0$, que corresponde a $R = [0 \ldots, 0, 1, 0, \ldots 0]$, (el 1 en la posición i), y $r = \beta_i^0$,
2. $H_0 : \boldsymbol{\beta}_s = \boldsymbol{\beta}_s^0$, que corresponde a $R = [\mathbf{0}_{q\times r}|\mathbf{I}_{q\times s}]$ y $r = \boldsymbol{\beta}_s^0$.

Ejemplo 8. En el modelo de regresión múltiple

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + u_i$$

la hipótesis de no significación individual $H_0 : \beta_2 = 0$ puede expresarse como

$$\begin{pmatrix} 0 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{pmatrix} = (0)$$

y la hipótesis de no significación gobal $H_0 : \beta_2 = \beta_3 = \beta_4 = 0$

$$\begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

◁

Proposición 46. *El contraste de restricciones lineales rechaza la hipótesis de no significación individual $H_0 : \beta_i = 0$ al nivel de significación α si*

$$F \equiv \frac{\hat{\beta}_i^2}{\hat{\sigma}_u^2 a_{ii}} \equiv \frac{\hat{\beta}_i^2}{\hat{V}(\hat{\beta}_i)} > c$$

donde c es el valor crítico para el cual $Prob(F_{1,n-k} > c) = \alpha$

Demostración. Aquí $\mathbf{R}$ es un vector $1 \times k$ de ceros con un uno en la posición i y r es igual a 0. Así, $\mathbf{R}\boldsymbol{\beta}$ selecciona el elemento i-ésimo de $\boldsymbol{\beta}$ y $\mathbf{R}\boldsymbol{\beta} - r = \beta_i$. Por otro lado, $\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'$ selecciona el elemento i-ésimo de la diagonal principal de $(\mathbf{X}'\mathbf{X})^{-1}$, que denotamos por a_{ii}. Luego,

$$F \equiv [\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}]'[\hat{\sigma}_u^2\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}[\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}]/q = (\hat{\beta}_i)(\hat{\sigma}_u^2 a_{ii})^{-1}(\hat{\beta}_i)/1 = \frac{\hat{\beta}_i^2}{\hat{\sigma}_u^2 a_{ii}} \sim F_{1,n-k}$$

□

De aquí, el contraste de significación individual puede basarse en el estadístico t o en el estadístico F. La siguiente proposición muestra la equivalencia de ambos estadísticos de contraste.

Proposición 47. *El cuadrado de una distribución t con $n-k$ grados de libertad es una distribución F con 1 y $n-k$ grados de libertad.*

Demostración.

$$t_{n-k}^2 = \left(\frac{N(0,1)}{\sqrt{\chi_{n-k}^2/(n-k)}}\right)^2 = \frac{\chi_1^2/1}{\chi_{n-k}^2/(n-k)} \sim F_{1,n-k}$$

□

4.7. Predicción

La **predicción económica** es uno de los principales motivos para construir un modelo econométrico, y adquiere especial relevancia cuando es necesario tomar decisiones en un marco de incertidumbre, por ejemplo, en la gestión de existencias, la planificación y programación de la producción, la planificación financiera, el diseño de políticas económicas, la concesión de hipótecas, la selección de personal, etc.

El problema consiste en estimar el valor de la variable dependiente asociado a determinados valores de las variables explicativas. Cuando el valor a predecir se conoce de antemano, hablamos de **predicción ex-ante**; en caso contrario, de **predicción ex-post**. Las predicciones *ex-ante* son útiles para juzgar la capacidad predictiva del modelo pues nos permiten calcular **errores de predicción**. De hecho, una práctica común en la evaluación de un modelo econométrico consiste en usar una submuestra de observaciones en la estimación y reservar las observaciones restantes para la predicción *ex-ante*.

Podemos considerar cuatro modalidades del problema dependiendo de si los parámetros y las variables explicativas son conocidos o, por el contrario, deben ser estimados. En este capítulo describimos la predicción puntual y por intervalo con parámetros estimados y variables explicativas conocidas, a veces conocida como **predicción incondicional**. Veremos que, bajo determinados supuestos, el *valor ajustado* es la mejor predicción (incondicional) lineal e insesgada. En cambio, las propiedades estadísticas de la **predicción condicional**, basada en estimaciones de parámetros y variables explicativas, no se conocen en muestras finitas.

4.7.1. Predicción de una observación. Nos interesa predecir el valor y_0 de la variable dependiente asociado al vector de valores conocidos $\mathbf{x}_0 = (1\ x_{02}\ \dots\ x_{0k})'$ de las variables explicativas. Parece razonable predecir y_0 como

$$\hat{y}_0 = \mathbf{x}_0'\hat{\boldsymbol{\beta}} \tag{4.4}$$

que se denomina **predicción puntual** de y_0.

Para derivar las propiedades estadísticas de la predicción puntual, necesitamos extender el marco del modelo clásico con los siguientes supuestos sobre la observación a predecir:

1. el valor y_0 es una realización del modelo lineal general, es decir, $y_0 = \mathbf{x}_0'\beta + u_0$;
2. el vector $\mathbf{x}_0 = (1\ x_{02}\ \dots\ x_{0k})'$ asociado a y_0 es conocido;
3. el error u_0 es una variable aleatoria (normal) con media 0 y varianza σ_u^2, siendo independiente de $u_i\,(i = 1, \dots, n)$: $E(u_0) = 0$, $E(u_0^2) = \sigma_u^2$ y $E(u_0 u_i) = 0$ para $i = 1, \dots, n$.

Definición 41. *El error de predicción, denotado por e_0, es la diferencia entre el valor observado y_0 y su pronóstico $\hat{y}_0$*

$$e_0 = y_0 - \hat{y}_0$$

Bajo el supuesto 1, podemos escribir el error de predicción como

$$e_0 = \mathbf{x}_0'\beta + u_0 - \mathbf{x}_0'\hat{\beta} = -\mathbf{x}_0'(\hat{\beta} - \beta) + u_0$$

que es la suma de dos componentes: (1) el error en la estimación de los parámetros $-\mathbf{x}_0'(\hat{\boldsymbol{\beta}} - \beta)$, y (2) el error aleatorio inherente al modelo u_0.

Observación 27. *Si las variables explicativas no se conocen, entonces para predecir la variable dependiente debemos predecir previamente las variables explicativas, surgiendo así una nueva fuente de error.*

Proposición 48. *El error de predicción e_0 sigue una distribución normal con media cero y varianza $\sigma_u^2(1 + \mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0)$.*

Demostración.

1. Normalidad: e_0 es una combinación lineal de $\hat{\boldsymbol{\beta}}$ y u_0 que son variables aleatorias normalmente distribuidas.
2. Media: $E(e_0) = 0$ porque $\hat{\boldsymbol{\beta}}$ es un estimador insesgado y $E(u_0) = 0$ es un supuesto básico.
3. Varianza:

$$\begin{aligned} E(e_0)^2 =& E(-\mathbf{x}_0'(\hat{\boldsymbol{\beta}} - \beta) + u_0)^2 = E[(-\mathbf{x}_0'(\hat{\boldsymbol{\beta}} - \beta) + u_0)(-\mathbf{x}_0'(\hat{\boldsymbol{\beta}} - \beta) + u_0)'] \\ =& \mathbf{x}_0'E[(\hat{\boldsymbol{\beta}} - \beta)(\hat{\boldsymbol{\beta}} - \beta)']\mathbf{x}_0 + E(u_0)^2 - 2\mathbf{x}_0'E[(\hat{\boldsymbol{\beta}} - \beta)u_0] \\ =& \mathbf{x}_0'V(\hat{\boldsymbol{\beta}})\mathbf{x}_0 + \sigma_u^2 = \sigma_u^2(1 + \mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0) \end{aligned}$$

en donde se ha usado el resultado $E(\mathbf{u}u_0) = E(u_1 u_0\ \dots\ u_n u_0)' = \mathbf{0}'$.

□

PROPOSICIÓN 49. *La predicción $\hat{y}_0 = \mathbf{x}_0'\hat{\boldsymbol{\beta}}$ es lineal, insesgada y óptima.*

DEMOSTRACIÓN. La predicción puntual $\hat{y}_0$ puede escribirse como una combinación lineal de las observaciones de la variable dependiente

$$\hat{y}_0 = \mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \sum_{i=1}^{n} c_i y_i$$

en donde los pesos c_i $(i = 1, \dots, n)$ son los elementos del vector fila $\mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. Una predicción se dice insesgada si su error de predicción tiene media nula. Finalmente, la optimalidad significa que la predicción lineal general $\hat{y}_0^* = \sum_{i=1}^{n} \omega_i y_i$, cuando sea insesgada, tendrá asociado un error de predicción con igual o mayor varianza. Veamos que este resultado es un corolario del teorema de Gauss-Markov. Definiendo $\boldsymbol{\omega} = (\omega_1\ \omega_2\ \dots\ \omega_n)'$, la predicción lineal general puede escribirse como

$$\hat{y}_0^* = \boldsymbol{\omega}'\mathbf{y}$$

que contiene como caso especial a $\hat{y}_0$ cuando $\boldsymbol{\omega}' = \mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. El error de predicción asociado

$$\hat{e}_0^* = y_0 - \boldsymbol{\omega}'\mathbf{y} = \mathbf{x}_0'\boldsymbol{\beta} + u_0 - \boldsymbol{\omega}'\mathbf{X}\boldsymbol{\beta} - \boldsymbol{\omega}'\mathbf{u}$$

tendrá media cero cuando $\boldsymbol{\omega}'\mathbf{X} = \mathbf{x}_0'$. Es inmediato comprobar que el vector de ponderaciones $\mathbf{c} = (c_1\ c_2\ \dots\ c_n)'$ de $\hat{y}_0$ cumple esta condición, la cual nos permite descomponer $\boldsymbol{\omega}$ como la suma de dos vectores ortogonales $\boldsymbol{\omega} = \mathbf{c} + \mathbf{d}$. En efecto, premultiplicando por $\mathbf{X}'$ tenemos $\mathbf{X}'\boldsymbol{\omega} = \mathbf{X}'\mathbf{c} + \mathbf{X}'\mathbf{d}$, pero $\mathbf{X}'\boldsymbol{\omega} = \mathbf{X}'\mathbf{c} = \mathbf{x}_0'$ resultando que $\mathbf{d}'\mathbf{X} = 0$ y $\mathbf{d}'\mathbf{c} = \mathbf{d}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0 = 0$. De aquí,

$$\boldsymbol{\omega}'\boldsymbol{\omega} = \mathbf{c}'\mathbf{c} + \mathbf{d}'\mathbf{d}$$

y podemos ver que la varianza del error de la predicción lineal general es igual a la varianza de e_0 más una magnitud no negativa

$$\begin{aligned} V(e_0^*) =& \sigma_u^2(1 + \boldsymbol{\omega}'\boldsymbol{\omega}) = \sigma_u^2(1 + \sum_{i=1}^{n} w_i^2) = \sigma_u^2(1 + \sum_{i=1}^{n} c_i^2) + \sigma_u^2 \sum_{i=1}^{n} d_i^2 \\ =& V(e_0) + \sigma_u^2 \sum_{i=1}^{n} d_i^2 \end{aligned}$$

De aquí, concluimos que, dentro de la clase de predicciones lineales e insesgadas, la predicción $\hat{\mathbf{y}}_0 = \mathbf{x}_0\hat{\boldsymbol{\beta}}$ tiene la menor varianza. □

DEFINICIÓN 42. *El error cuadrático medio de la predicción es igual al cuadrado del sesgo de la predicción más la varianza del error de predicción*

$$E(y_0 - \hat{y}_0)^2 = E[(y_0 - E\hat{y}_0) - (\hat{y}_0 - E\hat{y}_0)]^2 = E(y_0 - E\hat{y}_0)^2 + E(\hat{y}_0 - E\hat{y}_0)^2$$

Observación 28. *Predicción óptima significa predicción de error cuadrático medio mínimo.*

PROPOSICIÓN 50. *La predicción por intervalo o el intervalo de confianza para y_0 de nivel α es*

$$\hat{y}_0 \pm c\sqrt{\hat{V}(e_0)}$$

en donde c es el valor crítico para el cual $Prob(t_{n-k} > c) = \alpha/2$.

Demostración. Análogamente a la derivación del intervalo de confianza para $\hat{\beta}_i$, tenemos que

$$\frac{e_0}{\hat{\sigma}_u\sqrt{1+\mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0}} \sim t_{n-k}$$

De aquí,

$$Prob(-c < \frac{e_0}{\hat{\sigma}_u\sqrt{1+\mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0}} < c) = 1-\alpha$$

que podemos escribir como

$$Prob(-c < \frac{y_0-\hat{y}_0}{\hat{\sigma}_u\sqrt{1+\mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0}} < c) = 1-\alpha$$

o bien

$$Prob(\hat{y}_0 - c\hat{\sigma}_u\sqrt{1+\mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0} < y_0 < \hat{y}_0 + c\hat{\sigma}_u\sqrt{1+\mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0}) = 1-\alpha$$

□

Cuando trabajamos con más de 30 observaciones, podemos aproximar el cuantil del 97,5 % por dos, y construir el intervalo de predicción como

$$\hat{y}_0 \pm 2\sqrt{\hat{V}(e_0)}$$

cuya interpretación es la siguiente: si generamos distintas realizaciones de la variable dependiente, estimamos el modelo en cada realización, y predecimos el valor y_0, entonces el 95 % de las predicciones caerán dentro del intervalo de confianza.

La predicción puntual carece de sentido si no va acompañada de una medida de riesgo (el error estándar) que nos permita calcular la predicción por intervalo. Cuanto mayor sea el error estándar tanto mayor será la amplitud del intervalo y menos confianza tendremos en la predicción.

Observación 29. *Muchos libros de texto describen la predicción puntual y por intervalo del valor esperado* $E(y_0)$. *Debería quedar claro que, generalmente, el valor esperado no sólo es inobservable sino que además depende de la especificación del modelo. En consecuencia no podemos calcular los errores de predicción. En realidad, con esta aproximación se intenta eliminar la componente* u_0 *del error de predicción y reducir artificialmente el riesgo de la predicción puntual.*

Ejemplo 9. *El modelo de regresión ajustado a los datos de las calificaciones predice que un alumno de econometría que estudia 3 horas al día, asiste regularmente a clase y no recibe clases particulares en una academia tendrá una nota:*

$$\hat{y}_0 = \begin{pmatrix} 1 & 3 & 1 & 0 \end{pmatrix} \begin{pmatrix} 0{,}85135135 \\ 1{,}5135135 \\ 1{,}527027 \\ 0{,}10810811 \end{pmatrix} = 6{,}9189188$$

La varianza del error de predicción

$$V(e_0) = 1{,}11712 \times \left[1 + \begin{pmatrix} 1 & 3 & 1 & 0 \end{pmatrix} \begin{pmatrix} 0{,}0324324 & 0{,}0648649 & 0{,}0594595 & -0{,}156757 \\ 0{,}0648649 & 1{,}46306 & 1{,}11892 & -1{,}64685 \\ 0{,}0594595 & 1{,}11892 & 1{,}27568 & -1{,}45405 \\ -0{,}156757 & -1{,}64685 & -1{,}45405 & 2{,}25766 \end{pmatrix} \begin{pmatrix} 1 \\ 3 \\ 1 \\ 0 \end{pmatrix} \right]$$

$$= 1{,}11712 \times (1 + 1{,}60901)$$

De aquí, el intervalo de confianza para la predicción de nivel 5 %

$$(6{,}9189188 \pm 2{,}44691 \times 1{,}70721)$$

que podemos escribir como (2,74153; 11,0963). *Una deficiencia de esta predicción por intervalo es que contiene valores no admisibles mayores que la nota máxima, porque estamos usando una muestra muy pequeña y, además, no tenemos en cuenta el rango de variación de la variable dependiente.*

4.7.2. Predicción de varias observaciones. Los resultados derivados en la sección anterior se extienden fácilmente para predecir conjuntamente varias observaciones. Suponemos que el vector $\mathbf{y}_0$ viene generado por el modelo lineal general

$$\mathbf{y}_0 = \mathbf{X}_0\boldsymbol{\beta} + \mathbf{u}_0$$

en donde

$$\mathbf{y}_0 = \begin{pmatrix} y_{n+1} \\ \vdots \\ y_{n+m} \end{pmatrix}, \quad \mathbf{X}_0 = \begin{pmatrix} 1 & x_{n+1,1} & \dots & x_{n+1,k} \\ \vdots & \vdots & \dots & \vdots \\ 1 & x_{n+m,1} & \dots & x_{n+m,k} \end{pmatrix} \quad \text{y} \quad \mathbf{u}_0 = \begin{pmatrix} u_{n+1} \\ \vdots \\ u_{m+1} \end{pmatrix}$$

Nuestro objetivo es predecir $\mathbf{y}_0$ suponiendo que la matriz $\mathbf{X}_0$ es conocida. Es inmediato derivar las expresiones para la predicción lineal, insesgada y óptima de $\mathbf{y}_0$

$$\hat{\mathbf{y}}_0 = \mathbf{X}_0\hat{\boldsymbol{\beta}}$$

el vector de errores de predicción

$$\mathbf{e}_0 = \mathbf{y}_0 - \mathbf{X}_0\hat{\boldsymbol{\beta}} = \mathbf{u}_0 - \mathbf{X}_0(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$$

y la matriz de varianzas y covarianzas

$$V(\mathbf{e}_0) = \sigma_u^2(\mathbf{I}_m + \mathbf{X}_0(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_0')$$

De la fórmula dada para el predictor podríamos pensar que antes de predecir necesitamos estimar los parámetros del modelo. Vamos a ver, sin embargo, un resultado que demuestra que es posible estimar el vector de coeficientes $\boldsymbol{\beta}$ y predecir el vector $\mathbf{y}_0$ simultáneamente. El modelo lineal general que combina las muestras de observación y predicción viene dado por

$$\begin{pmatrix} \mathbf{y} \\ \mathbf{0}_m \end{pmatrix} = \begin{pmatrix} \mathbf{X} & \mathbf{0} \\ \mathbf{X}_0 & -\mathbf{I}_m \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta} \\ \mathbf{y}_0 \end{pmatrix} + \begin{pmatrix} \mathbf{u} \\ \mathbf{u}_0 \end{pmatrix} \tag{4.5}$$

Aplicando la fórmula del estimador de mínimos cuadrados tenemos

$$\begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{y}}_0 \end{pmatrix} = \begin{pmatrix} \mathbf{X}'\mathbf{X} + \mathbf{X}'_0\mathbf{X}_0 & -\mathbf{X}_0' \\ -\mathbf{X}_0 & \mathbf{I}_m \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{0} \end{pmatrix}$$

y por la fórmula de la inversa de una matriz particionada

$$\begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{y}}_0 \end{pmatrix} = \begin{pmatrix} (\mathbf{X}'\mathbf{X})^{-1} & (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_0' \\ \mathbf{X}_0(\mathbf{X}'\mathbf{X})^{-1} & \mathbf{I}_m + \mathbf{X}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_0 \end{pmatrix} \begin{pmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{0} \end{pmatrix} = \begin{pmatrix} (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\ \mathbf{X}_0(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \end{pmatrix}$$

La matriz de varianzas y covarianzas de este vector de estimadores será

$$\begin{pmatrix} V(\hat{\boldsymbol{\beta}}) & Cov(\hat{\boldsymbol{\beta}}, \hat{\mathbf{y}}_0) \\ Cov(\hat{\boldsymbol{\beta}}, \hat{\mathbf{y}}_0) & V^*(\hat{\mathbf{y}}_0) \end{pmatrix} = \sigma_u^2 \begin{pmatrix} (\mathbf{X}'\mathbf{X})^{-1} & (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_0' \\ \mathbf{X}_0(\mathbf{X}'\mathbf{X})^{-1} & \mathbf{I}_m + \mathbf{X}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_0 \end{pmatrix}$$

en donde el asterisco en $V^*(\hat{\mathbf{y}}_0)$ advierte de que dicha matriz de varianzas y covarianzas no se corresponde con $V(\hat{\mathbf{y}}_0)$ sino con la del error de predicción $\mathbf{e}_0$. Podemos aplicar directamente el teorema de Gauss-Markov para afirmar que $\hat{\mathbf{y}}_0$ es el estimador lineal, insesgado y óptimo de $\mathbf{y}_0$.

Observación 30. *En la aplicación práctica de este procedimiento combinado de estimación y predicción, hay que tener especial cuidado con la interpretación del coeficiente de determinación R^2: la suma de cuadrados total estará distorsionada por la inclusión de m ceros en la variable dependiente. En cambio, la suma de cuadrados de los residuos y la varianza residual en el modelo combinado coinciden con las obtenidas en la muestra de estimación.*

4.7.3. Medidas de acuracidad predictiva.

Supongamos que hemos usado el modelo lineal general estimado

$$y_i = \mathbf{x}_i'\hat{\boldsymbol{\beta}} + u_i, \quad i = 1, \ldots, n$$

para generar m predicciones $\hat{y}_i$ $(i = n+1, \ldots, n+m)$ de los valores y_i $(i = n+1, \ldots, n+m)$. Cuando comparemos estas predicciones con los valores observados, podremos calcular los errores de predicción e_i, $(i = n+1, \ldots, n+m)$, que nos permiten calcular las siguientes medidas de acuracidad predictiva:

1. Error absoluto medio
$$MAE = \frac{1}{m}\sum_{i=1}^{m} |e_{n+i}|$$
2. Error porcentual absoluto medio
$$MAPE = \frac{1}{m}\sum_{i=1}^{m} \frac{|e_{n+i}|}{y_{n+i}}$$
3. Error cuadrático medio
$$MSE = \frac{1}{m}\sum_{i=1}^{m} e_{n+i}^2$$
4. Raíz cuadrada del error cuadrático medio
$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m} e_{n+i}^2}$$

5. Coeficiente de determinación extramuestral o correlación simple entre los valores observados y sus predicciones

$$R_0^2 = 1 - \frac{\sum_{i=1}^m e_{n+i}^2}{\sum_{i=1}^m \left(y_{n+i} - \frac{\sum_{i=1}^m y_{n+i}}{m}\right)^2}$$

Desafortunadamente, la ordenación de métodos alternativos de predicción según su acuracidad es dependiente de la medida utilizada. Así, el mejor método de predicción usando el criterio MAE, puede ocupar el último lugar en el ranking basado en el RMSE. En estos casos, puede ser conveniente utilizar una combinación lineal de las predicciones disponibles con la esperanza de que los errores positivos de un método se compensen con los errores negativos de otros.

Una herramienta fundamental para juzgar la capacidad predictiva es el gráfico de los errores de predicción. La presencia de rachas de errores con el mismo signo (errores sistemáticos) es un indicio de la posible existencia de un error de especificación.

4.8. Resumen

1. Los pasos a seguir para contrastar una hipótesis sobre un coeficiente de regresión ($H_0 : \beta_i = \beta_i^0$) son los siguientes:
   a) Calcular la desviación relativa $t = (\hat{\beta}_i - \beta_i^0)/dt(\hat{\beta}_i)$.
   b) Calcular el p-valor, es decir, la probabilidad de que una variable aletoria distribución t_{n-k} tome un valor mayor que el valor absoluto de la desviación relativa, $Prob(t_{n-k} > |t|)$.
   c) Comparar el p-valor con un nivel de significación determinado α (por ejemplo, $\alpha = 0{,}05$), y rechazar H_0 si $p < \alpha$.
2. El intervalo de confianza del $(1-\alpha)\,\%$ es el conjunto de hipótesis nulas sobre un coeficiente individual de regresión que no se rechazan al nivel de significación α.
3. El contraste de una hipótesis sobre una combinación lineal de coeficientes de regresión es idéntico al contraste de una hipótesis sobre un coeficiente de regresión.
4. Los pasos a seguir para contrastar una hipótesis sobre varios coeficientes de regresión ($H_0 : \boldsymbol{\beta}_s = \boldsymbol{\beta}_s^0$) son los siguientes:
   a) Calcular la distancia relativa $F = (\hat{\boldsymbol{\beta}}_s - \boldsymbol{\beta}_s^0)'[\hat{V}((\hat{\boldsymbol{\beta}}_s)]^{-1}(\hat{\boldsymbol{\beta}}_s - \boldsymbol{\beta}_s^0)'$.
   b) Calcular el p-valor, es decir, la probabilidad de que una variable aleatoria con distribución $F_{s,n-k}$ tome un valor mayor que el valor absoluto de la desviación relativa, $Prob(F_{s,n-k} > |F|)$.
   c) Comparar el p-valor con un nivel de significación determinado α (por ejemplo, $\alpha = 0{,}05$), y rechazar H_0 si $p < \alpha$.
5. El contraste t es un caso especial del contraste F. Sin embargo, los contrastes t individuales no son siempre congruentes con el contraste F correspondiente.
6. El análisis de varianza en el modelo lineal general es un resumen del contraste de significación global.
7. La hipótesis lineal general tiene la forma $\mathbf{R}\boldsymbol{\beta} = \mathbf{r}$.

8. El estimador de mínimos cuadrados restringidos puede obtenerse incorporando las restricciones en el modelo de regresión o minimizando la suma de cuadrados de los residuos sujeta a a las restricciones lineales.
9. La hipótesis $H_0 : \mathbf{R}\boldsymbol{\beta} - \mathbf{r} = \mathbf{0}$ se rechaza al nivel de significación α si

$$F \equiv [\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}]'[\hat{\sigma}_u^2\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}[\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}]/q > c$$

10. La predicción económica es útil para la toma de decisiones bajo incertidumbre.
11. Bajo los supuestos del modelo clásico, el valor ajustado es la predicción (incondicional) lineal, insesgada y óptima.
12. La predicción por intervalo se construye de un modo similar al intervalo de confianza para un coeficiente estimado.
13. El modelo lineal general que combina las muestras de estimación y predicción permite realizar estas dos operaciones simultáneamente.

Palabras clave

Regla de decisión
Nivel de significación
Valor crítico
Contraste t
Intervalo de confianza
Contraste de dos colas
Contraste F
Región de Confianza
Contraste de significación individual
p-valor
Contraste de significación global
Análisis de varianza
Hipótesis lineal general
Restricciones lineales
Modelo restringido
Modelo sin restricciones
Multiplicadores de Lagrange
Mínimos cuadrados restringidos
Contraste basado en sumas de residuos
Predicción puntual
Predicción por intervalo
Error de predicción
Predicción lineal general
Predicción óptima
Predicción incondicional

4.9. Ejercicios

1. Para la función de producción de tipo Cobb-Douglas

$$logY_i = \beta_0 + \beta_1 logL_i + \beta_2 logK_i + u_i$$

use el estadístico t para contrastar la hipótesis de rendimientos constantes a escala

$$H_0 : \quad \beta_1 + \beta_2 = 1 \quad versus \quad H_1 : \quad \beta_1 + \beta_2 \neq 1$$

2. Sea el modelo de regresión particionado

$$\mathbf{y} = \mathbf{X_r}\beta_\mathbf{r} + \mathbf{X_s}\beta_\mathbf{s} + \mathbf{u}$$

donde sabemos que se cumplen todas las hipótesis ideales, y que

$$\hat{\boldsymbol{\beta}}_s = (\mathbf{X_s'M_rX_s})^{-1}\mathbf{X_s'M_ry}$$

Se pide:

a) Exprese el estimador de $\hat{\beta}_\mathbf{s}$ en términos de la perturbación:

$$\hat{\boldsymbol{\beta}}_s = \boldsymbol{\beta_s} + (\mathbf{X_s'M_rX_s})^{\mathbf{-1}}\mathbf{X_s'M_ru}$$

b) Demuestre que $\hat{\beta}_\mathbf{s} \sim \mathbf{N}(\beta_\mathbf{s}, \sigma^\mathbf{2}_\mathbf{u}(\mathbf{X_s'M_rX_s})^{\mathbf{-1}})$

c) Derive la distribución del estadístico $\mathbf{R}\hat{\beta}_{\mathbf{s}} - \mathbf{r}$ donde $\mathbf{R}$ y $\mathbf{r}$ son matrices fijas de órdenes $(q \times s)$ y $(q \times 1)$ respectivamente.

d) Proponga un estadístico de contraste para evaluar la hipótesis nula H_0 : $\mathbf{R}\beta_{\mathbf{s}} - \mathbf{r} = \mathbf{0}$ y explique cuál es el criterio que permite aceptar o rechazarla.

3. Sea $\hat{\mathbf{y}}^*$ el vector de predicciones calculado como $\hat{\mathbf{y}}^* = \mathbf{X}^*\hat{\boldsymbol{\beta}}$. Calcule el vector de medias y la matriz de covarianzas del vector de errores de predicción $\mathbf{e}^* = \mathbf{y}^* - \hat{\mathbf{y}}^*$, suponiendo que $\mathbf{y}^* = \mathbf{X}^*\boldsymbol{\beta} + \mathbf{u}^*$ y $\mathbf{u}^* \sim N(\mathbf{0}, \sigma_u^2\mathbf{I})$.
4. En el modelo estimado
$$\begin{array}{ccc} \hat{C}_t = 2.0 & +0.12Y_t & -0.36P_t \\ (2.9) & (0.01) & (0.07) \end{array}$$
donde C_t, Y_t y P_t representan el consumo, la renta disponible y los precios, respectivamente. ¿Cómo se verán afectadas las estimaciones de los coeficientes y sus desviaciones típicas si escalamos los datos de renta y precios dividiéndolos por 100?
5. Demuestre que el estimador de mínimos cuadrados restringidos es un estimador insesgado si y sólo si las restricciones lineales son ciertas.
6. Demuestre que el estimador de mínimos cuadrados restringidos es más acurado que el estimador de mínimos cuadrados sin restricciones. ¿Es compatible este resultado con el teorema de Gauss-Markov?
7. Sea $\hat{\boldsymbol{\beta}}$ el estimador de mínimos cuadrados ordinarios en la regresión de $\mathbf{y}$ sobre $\mathbf{X}$, y sea $\mathbf{b}$ cualquier estimador alternativo. Demuestre que la diferencia en las dos sumas de cuadros es
$$(\mathbf{y} - \mathbf{X}\mathbf{b})'(\mathbf{y} - \mathbf{X}\mathbf{b}) - (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = (\mathbf{b} - \hat{\boldsymbol{\beta}})'\mathbf{X}'\mathbf{X}(\mathbf{b} - \hat{\boldsymbol{\beta}})$$
8. Comente la siguiente proposición: la imposición de restricciones lineales en la estimación de un modelo de regresión aumenta el R^2.
9. Sea $\mathbf{R}$ una matriz $q \times k$ y $\mathbf{r}$ un vector $q \times 1$ de números conocidos. Demuestre que la variable aleatoria $\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}$ tiene una distribución normal multivariante con vector de medias $\mathbf{R}\boldsymbol{\beta} - \mathbf{r}$ y matriz de varianzas-covarianzas $\sigma_u^2\mathbf{R}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{R}'$.
10. Demuestre que
$$\frac{1}{\sigma_u^2}\left(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}\right)'\left[\mathbf{R}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{R}'\right]^{-1}\left(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}\right) \sim \chi_q^2$$
11. Demuestre que
$$\frac{1}{q\hat{\sigma}_u^2}\left(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}\right)'\left[\mathbf{R}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{R}'\right]^{-1}\left(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}\right) \sim F_{q,n-k}$$
12. Sea el modelo de regresión simple $Y_i = \beta_1 + \beta_2 X_i + u_i$. Obtenga el estimador de mínimos cuadrados de β_2 sujeto a la restricción $\beta_1 = 3$.
13. En el modelo $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + u_i$, ¿son las hipótesis $\beta_2 - \beta_3 = 0$ y $\beta_2 = \beta_3 = 0$ equivalentes? Escriba separadamente ambas hipótesis en la forma $\mathbf{R}\boldsymbol{\beta} = \mathbf{r}$.
14. En el modelo $Y_i = \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$, los coeficientes satisfacen las restricciones
$$\beta_1 + \beta_2 = \alpha \quad \text{y} \quad \beta_1 + \beta_3 = -\alpha$$

Encuentre el estimador de mínimos cuadrados de α y su varianza muestral.

15. Derive la predicción (incondicional) puntual y por intervalo suponiendo que los coeficientes de regresión son conocidos.
16. Demuestre que el *valor ajustado* es la predicción incondicional de mínima varianza.
17. Derive la distribución muestral de la predicción óptima.
18. Demuestre que la varianza del error de predicción puede expresarse como

$$V(e_0) = \sigma^2 \left(1 + \frac{1}{n}\right) + \sum_{i=1}^{k} \sum_{j=1}^{k} (x_{i0} - \bar{x}_i)(x_{j0} - \bar{x}_j) cov(\hat{\beta}_i, \hat{\beta}_j)$$

19. Utilizando la expresión anterior para la varianza del error de predicción, explique qué efectos tienen sobre la amplitud del intervalo de predicción (precisión de la predicción) el tamaño muestal y el uso de valores de las variables explicativas alejados de las correspondientes medias muestrales.