ECONOMETRIA I.
Departamento de Fundamentos del Análisis Económico
Universidad de Alicante. Curso 2011/12

# GUIÓN TEMA 3.

# CONTRASTE DE HIPÓTESIS EN EL MRL

Los procedimientos clásicos de contrastes de hipótesis se basan en la construcción de estadísticos a partir de una muestra aleatoria que permitan decidir, con un nivel de confianza razonable, si los datos de la muestra podrían haber sido generados por una población hipotética. En todos los contrastes tenemos que especificar la hipótesis nula y la hipótesis alternativa, el estadístico de contraste y la región crítica.

En todo este tema vamos a suponer que se verifican las hipótesis básicas del MRL con errores normales.

## 3.1. Contrastes de hipótesis en el MRL

| |
|---|
| Bibliografía apartado 5.1: |
| Greene: 6.6.5, 6.6.6, 7.2 (7.2.1-7.2.5) |
| A. Fernández Gallastegui: 4.4.1 y 4.3 |
| Wooldridge: 4.1, 4.2, 4.4, 4.5 (hasta pag. 164) |
| y Apéndice C.6 (salvo pag. 847 y hasta pag. 851) |

### 3.1.1 Contrastes sobre un único coeficiente

- Supongamos primero que queremos contrastar

$$a) \begin{array}{l} H_0 : \beta_j = \beta_j^0 \\ H_1 : \beta_j \neq \beta_j^0 \end{array} \qquad b) \begin{array}{l} H_0 : \beta_j = \beta_j^0 \\ H_1 : \beta_j > \beta_j^0 \end{array} \qquad c) \begin{array}{l} H_0 : \beta_j = \beta_j^0 \\ H_1 : \beta_j < \beta_j^0 \end{array}$$

El estadístico de contraste es el mismo para los tres casos, lo que varía dependiendo de cuál sea la hipótesis alternativa es la región crítica.

Sabemos que

$$\widehat{\beta}_j \sim N\left[\beta_j, \sigma^2 (X'X)^{-1}_{jj}\right]$$

donde $(X'X)^{-1}_{jj}$ es el elemento $(j, j)$ de la matriz $(X'X)^{-1}$.

Bajo $H_0$

$$\widehat{\beta}_j \sim N\left[\beta_j^0, \sigma^2 (X'X)^{-1}_{jj}\right]$$

Podemos estandarizar $\widehat{\beta}_j$ restando la media y dividiendo por la desviación típica

$$z = \frac{\widehat{\beta}_j - \beta_j^0}{\sqrt{\sigma^2 (X'X)^{-1}_{jj}}} \sim N(0,1) \quad \text{bajo } H_0$$

Nótese que $z$ sólo nos serviría como estadístico de contraste si $\sigma^2$ fuese conocida, lo que no suele ocurrir en la práctica. Si $\sigma^2$ es desconocida, para construir el estadístico de contraste, tenemos que combinar $z$ con otro estadístico que también dependa de $\sigma^2$ de forma que al combinarlos obtengamos un estadístico que no dependa de $\sigma^2$. Puesto que

$$\frac{\widehat{\sigma}^2 (T-k)}{\sigma^2} \sim \chi^2_{(T-k)}$$

y como $\widehat{\sigma}^2$ y $\widehat{\beta}_j$ son independientes definimos el estadístico de contraste:

$$t = \frac{\frac{\widehat{\beta}_j - \beta_j^0}{\sqrt{\sigma^2 (X'X)^{-1}_{jj}}}}{\sqrt{\frac{(T-k)\widehat{\sigma}^2}{\sigma^2 (T-k)}}} = \frac{\widehat{\beta}_j - \beta_j^0}{\sqrt{\widehat{\sigma}^2 (X'X)^{-1}_{jj}}} = \frac{\widehat{\beta}_j - \beta_j^0}{SE(\widehat{\beta}_j)} \sim t_{T-k} \quad \text{bajo } H_0$$

Para un nivel de significación $\alpha$, las regiones críticas son:

$$\begin{array}{l} a)\ \{|t| > t_{T-k,\alpha/2}\} \\ b)\ \{t > t_{T-k,\alpha}\} \\ c)\ \{t < -t_{T-k,\alpha}\} \end{array}$$

Ejemplo 1:

Consideremos un modelo simple que relaciona el número anual de delitos en los campus universitarios ($crime$) con el número de alumnos matriculados ($enroll$)

$$\log(crime) = \beta_1 + \beta_2 \log(enroll) + u$$

Nótese que éste es un modelo de elasticidad constante ya que es lineal en logaritmos. $\beta_2$ mide la elasticidad de los delitos respecto al número de alumnos matriculados.

En base a una muestra de 97 universidades americanas se han obtenido los siguientes resultados

$$\widehat{\log(crime)}_t = \underset{(1.03)}{-6.63} + \underset{(0.11)}{1.27} \log(enroll)_t$$

donde los números entre paréntesis son los errores estándar.

Vamos a contrastar si existe suficiente evidencia para afirmar que un aumento de un 1% en el número de alumnos matriculados supone un aumento de más del 1% en el número de delitos. Primero tenemos que determinar la hipótesis nula y la hipótesis alternativa. Puesto que el incremento porcentual en el número de delitos ante un aumento de un 1% en la matrícula viene dado por la elasticidad $\beta_2$, tenemos que contrastar si hay evidencia suficiente para afirmar que $\beta_2 > 1$. Es decir tenemos que contrastar

$$H_0 : \beta_2 = 1$$
$$H_1 : \beta_2 > 1$$

El estadístico de contraste es

$$t = \frac{\widehat{\beta}_2 - 1}{SE(\widehat{\beta}_2)} \sim t_{95} \quad \text{bajo } H_0$$

El valor del estadístico de contraste para esta muestra es $t = (1.27 - 1)/0.11 = 2.45$. Como $t = 2.45 > t_{95,0.05} = 1.66$, podemos rechazar $H_0$ al 5%. Hay suficiente evidencia para afirmar que un aumento de un 1% en el número de alumnos matriculados supone un aumento de más del 1% en el número de delitos.

**Caso Particular:** Contraste de significatividad individual

$$H_0 : \beta_j = 0$$
$$H_1 : \beta_j \neq 0$$

El estadístico de contraste es:

$$t = \frac{\widehat{\beta}_j}{SE(\widehat{\beta}_j)} \sim t_{T-k} \quad \text{bajo } H_0$$

Este estadístico $t$ se denomina $t - ratio$.

## 3.1.2 Contrastes de una restricción lineal

- Supongamos que ahora queremos contrastar una restricción lineal cualquiera

$$a)\ \begin{array}{l} H_0 : R\beta = r \\ H_1 : R\beta \neq r \end{array} \qquad b)\ \begin{array}{l} H_0 : R\beta = r \\ H_1 : R\beta > r \end{array} \qquad c)\ \begin{array}{l} H_0 : R\beta = r \\ H_1 : R\beta < r \end{array}$$

donde $R$ es un vector $1 \times k$ y $r$ es un escalar.

El estadístico de contraste es el mismo para los tres contrastes, lo que varía dependiendo de cuál sea la hipótesis alternativa es la región crítica.

Dado que

$$\widehat{\beta} \sim N(\beta, \sigma^2 (X'X)^{-1})$$

$$R\widehat{\beta} \sim N(R\beta, \sigma^2 R(X'X)^{-1}R')$$

Bajo $H_0$

$$R\widehat{\beta} \sim N(r, \sigma^2 R(X'X)^{-1}R')$$

y, puesto que $R\widehat{\beta}$ es un escalar, podemos estandarizar su distribución restando la media y dividiendo por la desviación típica

$$z = \frac{R\widehat{\beta} - r}{\sqrt{\sigma^2 R(X'X)^{-1}R'}} \sim N(0,1) \qquad \text{bajo } H_0$$

Nótese que $z$ sólo nos serviría como estadístico de contraste si $\sigma^2$ fuese conocida, lo que no suele ocurrir en la práctica. Si $\sigma^2$ es desconocida, para construir el estadístico de contraste, tenemos que combinar $z$ con otro estadístico que también dependa de $\sigma^2$ de forma que al combinarlos obtengamos un estadístico que no dependa de $\sigma^2$. Puesto que

$$\frac{(T-k)\widehat{\sigma}^2}{\sigma^2} \sim \chi^2_{T-k}$$

y como $\widehat{\sigma}^2$ y $\widehat{\beta}$ son independientes, definimos el estadístico de contraste:

$$t = \frac{\frac{R\widehat{\beta} - R\beta}{\sqrt{\sigma^2 R(X'X)^{-1}R}}}{\sqrt{\frac{(T-k)\widehat{\sigma}^2}{\sigma^2(T-k)}}} = \frac{R\widehat{\beta} - r}{\sqrt{\widehat{\sigma}^2 R(X'X)^{-1}R'}} = \frac{R\widehat{\beta} - r}{\sqrt{R\widehat{var(\widehat{\beta})}R'}} \sim t_{T-k} \quad \text{bajo } H_0$$

Para un nivel de significación $\alpha$, las regiones críticas son:

$$\begin{array}{l} a)\ \{|t| > t_{T-k,\alpha/2}\} \\ b)\ \{t > t_{T-k,\alpha}\} \\ c)\ \{t < -t_{T-k,\alpha}\} \end{array}$$

<u>Ejemplo 2:</u>

Consideremos el modelo para el gasto en vestido y calzado que ya vimos en el Tema 1.

$$gvest = \beta_1 + \beta_2 renta + \beta_3 nad + \beta_4 nhijos + u$$

El modelo estimado en base a una muestra de 7038 hogares españoles es

$$\begin{aligned} \widehat{gvest}_t &= 1.2 + \underset{(0.0033)}{0.064} renta_t + \underset{(0.0419)}{0.132} nad_t + \underset{(0.0391)}{0.159} nhijos_t \\ \widehat{var(\widehat{\beta})} &= \begin{bmatrix} 0.02278446 & & & \\ -0.00014636 & 0.0000112 & & \\ -0.00474617 & -0.00003484 & 0.00175293 & \\ -0.00350562 & 4.471e-07 & 0.00049591 & 0.00152536 \end{bmatrix} \end{aligned}$$

donde $gvest$ es el gasto anual del hogar en vestido y calzado (en miles Euros), $renta$ es la renta anual del hogar (en miles de Euros), $nad$ es el número de adultos en el hogar y $nhijos$ es el número de hijos menores de 18 años. Los números entre paréntesis son los errores estándar. Vamos a contrastar si un aumento en el número de hijos tiene el mismo efecto sobre el gasto en vestido y calzado que un aumento en el número de adultos. Es decir, vamos a contrastar

$$\begin{array}{l} H_0 : \beta_3 = \beta_4 \\ H_1 : \beta_3 \neq \beta_4 \end{array}$$

El vector $R$ es en este caso $R = (0, 0, 1, -1)$ y $r = 0$, por tanto

$$R\widehat{\beta} - r = \widehat{\beta}_3 - \widehat{\beta}_4$$

$$R\widehat{var(\widehat{\beta})}R' = \widehat{var(\widehat{\beta}_3)} + \widehat{var(\widehat{\beta}_4)} - 2 * \widehat{cov(\widehat{\beta}_3, \widehat{\beta}_4)}$$

y el estadístico de contraste es

$$\frac{\widehat{\beta}_3 - \widehat{\beta}_4}{\sqrt{\widehat{var(\widehat{\beta}_3)} + \widehat{var(\widehat{\beta}_4)} - 2 * \widehat{cov(\widehat{\beta}_3, \widehat{\beta}_4)}}} \sim t_{7034} \simeq N(0,1) \qquad \text{bajo } H_0$$

Utilizando $\widehat{var(\widehat{\beta})}$ :

$$\begin{aligned}\widehat{var(\widehat{\beta}_3)} + \widehat{var(\widehat{\beta}_4)} - 2 * \widehat{cov(\widehat{\beta}_3, \widehat{\beta}_4)} &= 0.00175293 + 0.00152536 - 2 * 0.00049591 \\ &= 0.002286\end{aligned}$$

y el valor del estadístico de contraste en la muestra es $t = (0.132 - 0.159)/\sqrt{0.002286} = -0.56$. Puesto que $|t| = 0.56 \not> z_{0.025} = 1.96$, no podemos rechazar $H_0$ al 5% y por tanto concluimos que un aumento en el número de hijos tiene el mismo efecto sobre el gasto en vestido y calzado que un aumento en el número de adultos.

Para realizar este contraste de forma más sencilla podemos reparametrizar el modelo de la siguiente forma. Sea $\theta_3 = \beta_3 - \beta_4$, la hipótesis que tenemos que contrastar es ahora

$$H_0 : \theta_3 = 0$$
$$H_1 : \theta_3 \neq 0$$

Dado que $\beta_3 = \theta_3 + \beta_4$ podemos escribir el modelo como

$$\begin{aligned} gvest &= \beta_1 + \beta_2 renta + (\theta_3 + \beta_4) nad + \beta_4 nhijos + u \\ gvest &= \beta_1 + \beta_2 renta + \theta_3 nad + \beta_4 (nad + nhijos) + u \\ gvest &= \beta_1 + \beta_2 renta + \theta_3 nad + \beta_4 tfam + u \end{aligned}$$

donde $tfam = nad + nhijos$. En base a la misma muestra se obtiene

$$\widehat{gvest}_t = 1.2 + \underset{(0.0033)}{0.064} renta_t - \underset{(0.048)}{0.027} nad_t + \underset{(0.0391)}{0.159} tfam_t$$

El valor del estadístico de contraste es $t = -0.027/0.048 = -0.56$ que (salvo errores de redondeo) coincide con el resultado que obtuvimos antes.

### 3.1.3 Contrastes de un conjunto de restricciones lineales

- Supongamos ahora que queremos contrastar $q$ ($q < k$) restricciones lineales

$$H_0 : R\beta = r$$
$$H_1 : R\beta \neq r$$

donde $R$ es ahora una matriz $q \times k$ de rango $q$ (que la matriz $R$ tenga rango $q$ quiere decir que las restricciones lineales son independientes) y $r$ es un vector $q \times 1$.

Dado que

$$\widehat{\beta} \sim N(\beta, \sigma^2 (X'X)^{-1})$$

Multiplicando por la izquierda por $R$

$$R\widehat{\beta} \sim N(R\beta, \sigma^2 R(X'X)^{-1}R')$$

Bajo $H_0$

$$R\widehat{\beta} \sim N(r, \sigma^2 R(X'X)^{-1}R')$$

dado que ahora $R\widehat{\beta}$ es un vector $q \times 1$ y no un escalar, no podemos estandarizar su distribución como hicimos en el apartado anterior, sin embargo, utilizando el Teorema 6 del Tema 2

$$\chi = (R\widehat{\beta} - r)'(\sigma^2 R(X'X)^{-1}R')^{-1}(R\widehat{\beta} - r) \sim \chi^2_q \quad \text{bajo } H_0$$

Nótese que $\chi$ sólo nos serviría como estadístico de contraste si $\sigma^2$ fuese conocida, lo que no suele ocurrir en la práctica. Si $\sigma^2$ es desconocida, para construir el estadístico de contraste, tenemos que combinar $\chi$ con otro estadístico que también dependa de $\sigma^2$ de forma que al combinarlos obtengamos un estadístico que no dependa de $\sigma^2$. Puesto que

$$\frac{(T-k)\widehat{\sigma}^2}{\sigma^2} \sim \chi^2_{T-k}$$

y como $\widehat{\sigma}^2$ y $\widehat{\beta}$ son independientes, definimos el estadístico de contraste:

$$\begin{aligned} F &= \frac{(R\widehat{\beta}-r)'(\sigma^2 R(X'X)^{-1}R')^{-1}(R\widehat{\beta}-r)/q}{\frac{(T-k)\widehat{\sigma}^2}{\sigma^2(T-k)}} \\ &= (R\widehat{\beta} - r)'(\widehat{\sigma}^2 R(X'X)^{-1}R')^{-1}(R\widehat{\beta} - r)/q \\ &= (R\widehat{\beta} - r)'(R\widehat{var(\beta)}R')^{-1}(R\widehat{\beta} - r)/q \sim F_{q,T-k} \quad \text{bajo } H_0 \end{aligned}$$

Para un nivel de significación $\alpha$, la región crítica es $\{F > F_{q,T-k,\alpha}\}$.

Cuando tenemos más de una restricción lineal, sólo se pueden realizar contrastes bilaterales.

- Ejemplo 3

Consideremos el modelo para los salarios estimado en base a una muestra de 935 individuos que ya vimos en el Tema 1.

$$\begin{aligned}
\widehat{salario}_t &= -7.92 + \underset{(0.056)}{0.605} educ_t + \underset{(1.000)}{0.357} edad_t - \underset{(0.015)}{0.0022} edad2_t \\
\widehat{var(\widehat{\beta})} &= \begin{bmatrix} 271.88 & & & \\ 0.0396283 & 0.00318222 & & \\ -16.470229 & -0.00503106 & 1.0015721 & \\ 0.2467871 & 0.00007604 & -0.0150317 & 0.00022595 \end{bmatrix}
\end{aligned}$$

donde *salario* es el salario mensual en cientos de dolares, *educ* es el nivel de educación en años, *edad* es la edad en años y *edad*2 es la edad al cuadrado. Vamos a contrastar si el efecto marginal de la edad sobre el salario es cero.

Si $\beta_3$ y $\beta_4$ son los coeficientes poblacionales de *edad* y *edad*2, el efecto marginal es

$$\beta_3 + 2 * \beta_4 * edad$$

y el efecto marginal será cero si y solo si

$$\beta_3 + 2 * \beta_4 * edad = 0$$

para todo valor de *edad*, y por tanto tenemos que contrastar:

$$\begin{array}{l} H_0 : \beta_3 = 0, \beta_4 = 0 \\ H_1 : \beta_3 \neq 0 \text{ y/o } \beta_4 \neq 0 \end{array}$$

La matriz $R$ y el vector $r$ son

$$R = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \qquad r = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

El estadístico de contraste es

$$F = (\widehat{\beta}_3, \widehat{\beta}_4) \begin{bmatrix} \widehat{var(\widehat{\beta}_3)} & \widehat{cov(\widehat{\beta}_3, \widehat{\beta}_4)} \\ \widehat{cov(\widehat{\beta}_3, \widehat{\beta}_4)} & \widehat{var(\widehat{\beta}_4)} \end{bmatrix}^{-1} \begin{pmatrix} \widehat{\beta}_3 \\ \widehat{\beta}_4 \end{pmatrix} /2 \sim F_{2,931} \quad \text{bajo } H_0$$

El valor del estadístico de contraste en la muestra es:

$$\begin{aligned}
F &= \frac{1}{2}(0.357, -0.0022)\begin{bmatrix} 1.0015721 & -0.0150317 \\ -0.0150317 & 0.00022595 \end{bmatrix}^{-1}\begin{pmatrix} 0.357 \\ -0.0022 \end{pmatrix} = \\
&= \frac{1}{2}(0.357, -0.0022)\begin{bmatrix} 639.70242 & 42557.269 \\ 42557.269 & 2835618.9 \end{bmatrix}\begin{pmatrix} 0.357 \\ -0.0022 \end{pmatrix} \\
&= \frac{1}{2}28.40 = 14.2
\end{aligned}$$

Puesto que $F = 14.2 > F_{2,931,0.05} = 3$ podemos rechazar $H_0$ al 5% y por tanto concluimos que los salarios dependen de la edad.

**Nota**: Los estadísticos de contraste $t$ y $F$ no varían ante un cambio de unidades de medida en las variables explicativas y/o en la variable dependiente.

# 3.2 Estimación con restricciones lineales

| |
|---|
| Bibliografía apartado 5.2: |
| Greene: 6.6.7, 7.3 y 7.4 |
| A. Fernández Gallastegui: 4.1, 4.4.3, pag 91 y Apéndices 4.B y 4.C |
| Wooldridge: 4.5 |

Si sabemos que los coeficientes del modelo satisfacen una o más restricciones lineales podemos imponer dichas restricciones para mejorar la eficiencia. El método de estimación por mínimos cuadrados imponiendo restricciones lineales en los parámetros del modelo se denomina Estimación de Mínimos Cuadrados Restringidos.

- Consideremos un conjunto de $q$ restricciones lineales $R\beta = r$, donde $R$ es una matriz $q \times k$ y $r$ es un vector $q \times 1$. El estimador de Mínimos Cuadrados Restringidos se obtiene como solución del problema de minimización con restricciones

$$\begin{aligned}
\min_b \sum_{t=1}^{T} e_t^2 &= \min_b \sum_{t=1}^{T} (Y_t - X_t' b)^2 \\
\text{sujeto a } Rb &= r
\end{aligned}$$

La solución de este problema de minimización es

$$\widehat{\beta}_r = \widehat{\beta} - (X'X)^{-1}R'\left[R\left(X'X\right)^{-1}R'\right]^{-1}\left(R\widehat{\beta} - r\right)$$

- No es necesario utilizar la fórmula general para calcular el estimador de Mínimos Cuadrados Restringidos. Lo que se hace en la práctica es imponer las restricciones en el modelo y calcular el estimador MCO del modelo restringido.

  Ejemplo 2 (cont.):

  Consideremos de nuevo el modelo

  $$gvest = \beta_1 + \beta_2 renta + \beta_3 nad + \beta_4 nhijos + u \quad (1)$$

  Si imponemos la restricción que contrastamos anteriormente de que un aumento en el número de hijos tiene el mismo efecto sobre el gasto en vestido y calzado que un aumento en el número de adultos ($\beta_3 = \beta_4$) obtenemos el modelo

  $$gvest = \beta_1 + \beta_2 renta + \beta_3(nad + nhijos) + u \quad (2)$$

  y el estimador de Mínimos Cuadrados Restringidos del modelo (1) imponiendo la restricción $\beta_3 = \beta_4$ es el estimador MCO del modelo (2).

- **Propiedades del estimador $\widehat{\beta}_r$:**

  1. $E[\widehat{\beta}_r] = \beta$ cuando la restricción es cierta

     Demostración

     $$\begin{aligned} E[\widehat{\beta}_r] &= E\left[\widehat{\beta} - (X'X)^{-1}R'\left[R\left(X'X\right)^{-1}R'\right]^{-1}\left(R\widehat{\beta} - r\right)\right] = \\ &\underset{\text{Ya que } E[\widehat{\beta}]=\beta}{=} \beta - (X'X)^{-1}R'\left[R\left(X'X\right)^{-1}R'\right]^{-1}(R\beta - r) \\ &\underset{\text{Si } R\beta=r}{=} \beta \end{aligned}$$

  2. $Var\left[\widehat{\beta}_r\right] = \sigma^2(X'X)^{-1} - \sigma^2(X'X)^{-1}R'\left[R\left(X'X\right)^{-1}R'\right]^{-1}R\left(X'X\right)^{-1}$

     Nótese que $Var\left[\widehat{\beta}_r\right] \leq Var\left[\widehat{\beta}\right]$ con independencia de que las restricciones sean ciertas o no.

- **El contraste $F$:**

  Se puede demostrar que

  $$e_r'e_r = e'e + \left[r - R\widehat{\beta}\right]' \left[R(X'X)^{-1}R'\right]^{-1} \left[r - R\widehat{\beta}\right]$$

  , donde $e_r$ es el vector de residuos del modelo restringido y $e$ es el vector de residuos del modelo sin restringir.

  Por tanto, el estadístico de contraste $F$ para contrastar $R\beta = q$ se puede escribir como:

  $$F = \frac{\left[(R\widehat{\beta} - r)' \left[R\left(X'X\right)^{-1} R'\right]^{-1} (R\widehat{\beta} - r)\right] \Big/ q}{\widehat{\sigma}^2} = \frac{\left[\left(e_r'e_r - e'e\right)/\, q\right]}{\left[\left(e'e\right)/\left(T-k\right)\right]} \sim F_{q,T-k} \quad \text{bajo } H_0$$

  Si no cambia la variable dependiente como consecuencia de imponer la restricción en el modelo, entonces, de la definición de coeficiente de determinación se deduce que:

  $$F = \frac{\left[\left(e_r'e_r - e'e\right)/\, q\right]}{\left[\left(e'e\right)/\left(T-k\right)\right]} = \frac{\left[\left(R^2 - R_r^2\right)/\, q\right]}{\left[\left(1-R^2\right)/\left(T-k\right)\right]} \sim F_{q,T-k} \quad \text{bajo } H_0$$

  Gracias a estas fórmulas es posible calcular los estadísticos de contraste utilizando sumas cuadráticas residuales de modelos estimados por MCO.

  Ejemplo 2 (cont.):

  Consideremos de nuevo el modelo

  $$gvest = \beta_1 + \beta_2 renta + \beta_3 nad + \beta_4 nhijos + u$$

  y la restricción de que un aumento en el número de hijos tiene el mismo efecto sobre el gasto en vestido y calzado que un aumento en el número de adultos ($\beta_3 = \beta_4$)

  Los resultados para la estimación del modelo no restringido y del modelo restringido son

  $$\widehat{gvest}_t = 1.2 + \underset{(0.0033)}{0.064}\, renta_t + \underset{(0.0419)}{0.132}\, nad_t + \underset{(0.0391)}{0.159}\, nhijos_t, \quad SCR = 85138.162$$

  $$\widehat{gvest}_t = 1.2 + \underset{(0.0033)}{0.064}\, renta_t + \underset{(0.0326)}{0.147}\, tfam_t, \quad SCR = 85141.9863$$

donde $tfam = nad + nhijos$ y $SCR$ es la suma cuadrática residual. El valor del estadístico de contraste para esta muestra es $F = (85141.9863 - 85138.162)/(85138.162/7034) = 0.3154$ y como $F = 0.3154 < F_{1,7034}$ no podemos rechazar $H_0$ al 5%. Nótese que este contraste es equivalente al contraste $t$ que vimos anteriormente para este ejemplo ya que, cuando estamos contrastando una única restricción, $t^2 = F$. En este ejemplo obtuvimos $t = -0.56$ y por tanto $t^2 = 0.3136$ que no coincide exactamente con el valor que hemos obtenido para $F$ (salvo por los errores de redondeo).

- Supongamos que en el modelo del Ejemplo 2 queremos contrastar la hipótesis de que un aumento en el número de hijos tiene un efecto mayor sobre el gasto en vestido y calzado que un aumento en el número de adultos. La hipótesis que tenemos que contrastar es ahora

$$\begin{aligned} H_0 &: \beta_3 = \beta_4 \\ H_1 &: \beta_3 < \beta_4 \end{aligned}$$

Supongamos que sólo disponemos de información sobre las sumas cuadráticas residuales del modelo restringido y sin restringir. A partir del estadístico $F$, sólo podemos realizar contrastes bilaterales. Sin embargo, a partir del estadístico $F$ podemos obtener el estadístico $t$, puesto que

$$t = \pm\sqrt{F} = \pm 0.5616$$

Para saber si debemos coger la raíz positiva o negativa, debemos saber el signo del estadístico de contraste $t$. El estadístico $t$ para este contraste es:

$$t = \frac{\hat{\beta}_3 - \hat{\beta}_4}{SE\left(\hat{\beta}_3 - \hat{\beta}_4\right)}$$

y por lo tanto, puesto que el error estándard siempre es positivo, $t$ es positivo si $\hat{\beta}_3 - \hat{\beta}_4 > 0$ y es negativo si $\hat{\beta}_3 - \hat{\beta}_4 < 0$. Dadas las estimaciones del modelo: $\hat{\beta}_3 - \hat{\beta}_4 = 0.132 - 0.159 = -0.0270$ y por lo tanto concluimos que el estadístico $t$ de contraste que debemos utilizar es $t = -0.5616$. La región crítica para este contraste unilateral es:

$$\{t < -t_{7034,0.05}\} \simeq \{t < -z_{0.05}\} = \{t < -1.645\}$$

Puesto que el estadístico de contraste no pertenece a la región crítica, no hay suficiente evidencia para afirmar que un aumento en el número de hijos tiene un efecto mayor sobre el gasto en vestido y calzado que un aumento en el número de adultos.

- Ejemplo 3 (cont.)

  Consideremos de nuevo el modelo para los salarios estimado en base a una muestra de 935 individuos

  $$\widehat{salario}_t = -7.92 + \underset{(0.056)}{0.605} educ_t + \underset{(1.000)}{0.357} edad_t - \underset{(0.015)}{0.0022} edad2_t, \quad R^2 = 0.8690$$

  Vamos a contrastar utilizando ahora el modelo restringido si el efecto marginal de la edad sobre el salario es cero, es decir

  $$\begin{array}{l} H_0 : \beta_3 = 0, \beta_4 = 0 \\ H_1 : \beta_3 \neq 0 \text{ y/o } \beta_4 \neq 0 \end{array}$$

  donde $\beta_3$ y $\beta_4$ son los coeficientes poblacionales de *edad* y *edad*2. Los resultados de la estimación imponiendo las restricciones son:

  $$\widehat{salario}_t = 1.47 + \underset{(0.057)}{0.602} educ_t, \quad R^2 = 0.8651$$

  El valor del estadístico de contraste para esta muestra es

  $$F = [(0.8690 - 0.8651)/2] \, / \, [(1 - 0.8690)/931] = 13.86$$

  Puesto que $F = 13.86 > F_{2,931,0.05}$ podemos rechazar $H_0$ al 5% y por tanto concluimos que los salarios dependen de la edad. Nótese que el valor de $F$ no coincide exactamente con el que calculamos anteriormente para este mismo ejemplo debido a los errores de redondeo.

- **Caso Particular:** Contraste de significatividad global de la regresión.

  Este contraste analiza la hipótesis nula de que todos los coeficientes del modelo, excepto el término constante, son iguales a cero

  $$H_0 : \beta_2 = \beta_3 = ... = \beta_k = 0$$

  Utilizando la expresión que acabamos de ver para el estadístico $F$ en función del $R^2$ tenemos

  $$F = \frac{[(R^2 - R_r^2)/\,q]}{[(1 - R^2)/\,(T - k)]} = \frac{R^2/(k-1)}{(1 - R^2)/(T-k)} \sim F_{k-1,T-k} \quad \text{bajo } H_0$$

ya que el coeficiente de determinación del modelo restringido $R_r^2$ es 0 en este caso, puesto que el modelo restringido es un modelo que sólo tiene una constante como variable explicativa.

## 3.3. Intervalos de confianza

| |
|---|
| Bibliografía apartados: |
| Greene: 4.8 y 6.6.6 |
| A. Fernández Gallastegui: 4.6 |
| Wooldridge: 4.3, Apéndices C.5 (hasta pag. 838) y C.6 (pag. 852) |

- Idea general:
  - La estimación por intervalos consiste en proponer un intervalo cuyos límites sean estadísticos, y que contendrán al verdadero valor del parámetro con una probabilidad determinada.
  - Para una muestra concreta el intervalo contendrá o no el verdadero valor del parámetro y no tiene sentido hablar de la probabilidad de que contenga al verdadero parámetro. Si obtuviésemos sucesivas muestras aleatorias de un determinado tamaño muestral y calculásemos los intervalos cada vez, el porcentaje de intervalos que contendrían el verdadero valor del parámetro sería en torno al $100(1-\alpha)\%$.

- **Intervalos de confianza para $\boldsymbol{\beta}_j$**

  Puesto que
$$\widehat{\beta}_j \sim N\left[\beta_j, \sigma^2(X'X)^{-1}_{jj}\right]$$

  Podemos estandarizar $\widehat{\beta}_j$ restando la media y dividiendo por la desviación tipica
$$z = \frac{\widehat{\beta}_j - \beta_j}{\sqrt{\sigma^2(X'X)^{-1}_{jj}}} \sim N(0,1) \quad \text{bajo } H_0$$

  Sabemos que
$$\frac{\widehat{\sigma}^2(T-k)}{\sigma^2} \sim \chi^2_{(T-k)}$$

y puesto que $\widehat{\sigma}^2$ y $\widehat{\beta}_j$ son independientes

$$t = \frac{\frac{\widehat{\beta}_j - \beta_j}{\sqrt{\sigma^2 (X'X)^{-1}_{jj}}}}{\sqrt{\frac{(T-k)\widehat{\sigma}^2}{\sigma^2 (T-k)}}} = \frac{\widehat{\beta}_j - \beta_j}{\sqrt{\widehat{\sigma}^2 (X'X)^{-1}_{jj}}} = \frac{\widehat{\beta}_j - \beta_j}{SE(\widehat{\beta}_j)} \sim t_{T-k}$$

Por tanto

$$Prob\left(-t_{T-k,\alpha/2} < \frac{\widehat{\beta}_j - \beta_j}{SE(\widehat{\beta}_j)} < t_{T-k,\alpha/2}\right) = 1 - \alpha$$

Manipulando esta expresión tenemos

$$\begin{aligned} Prob\left(-t_{T-k,\alpha/2} SE(\widehat{\beta}_j) < \widehat{\beta}_j - \beta_j < t_{T-k,\alpha/2} SE(\widehat{\beta}_j)\right) &= 1 - \alpha \\ Prob\left(\widehat{\beta}_j - t_{T-k,\alpha/2} SE(\widehat{\beta}_j) < \beta_j < \widehat{\beta}_j + t_{T-k,\alpha/2} SE(\widehat{\beta}_j)\right) &= 1 - \alpha \end{aligned}$$

El intervalo de confianza para $\beta_j$ con nivel de confianza $100(1-\alpha)\%$ es

$$[\widehat{\beta}_j - t_{T-k,\alpha/2} SE(\widehat{\beta}_j), \widehat{\beta}_j + t_{T-k,\alpha/2} SE(\widehat{\beta}_j)]$$

- <u>Ejemplo 1 (cont.)</u>:

  Consideremos de nuevo el modelo que relaciona el número anual de delitos en los campus universitarios ($crime$) con el número de alumnos matriculados ($enroll$)

  $$\log(crime) = \beta_1 + \beta_2 \log(enroll) + u$$

  En base a una muestra de 97 universidades americanas se han obtenido los siguientes resultados

  $$\widehat{\log(crime)}_t = \underset{(1.03)}{-6.63} + \underset{(0.11)}{1.27} \log(enroll)_t$$

  Como $t_{95,0.025} = 1.98$, el intervalo de confianza al 95% para la elasticidad de los delitos repecto al número de alumnos matriculados ($\beta_2$) es

  $$[1.27 - 0.11 * 1.98, 1.27 + 0.11 * 1.98] = [1.0522, 1.4878]$$

- Existe una relación entre los intervalos de confianza y los contraste de hipótesis: Se rechaza a nivel $\alpha$ la hipótesis nula $H_0 : \beta_j = \beta_j^0$ frente a la alternativa $H_1 : \beta_j \neq \beta_j^0$ si y sólo si $\beta_j^0$ no pertenece al intervalo de confianza para $\beta_j$ con nivel de confianza $100(1-\alpha)\%$.

- **Intervalos de confianza para una combinación lineal de los $\boldsymbol{\beta}_j$**

  Sea $R$ un vector $1 \times k$, puesto que

$$\widehat{\beta} \sim N\left[\beta, \sigma^2 (X'X)^{-1}\right]$$

$$R\widehat{\beta} \sim N\left[R\beta, \sigma^2 R(X'X)^{-1}R'\right]$$

  Podemos estandarizar restando la media y dividiendo por la desviación tipica

$$z = \frac{R\widehat{\beta} - R\beta}{\sqrt{\sigma^2 R(X'X)^{-1}R'}} \sim N(0,1) \quad \text{bajo } H_0$$

  Sabemos que

$$\frac{\widehat{\sigma}^2 (T-k)}{\sigma^2} \sim \chi^2_{(T-k)}$$

  y puesto que $\widehat{\sigma}^2$ y $\widehat{\beta}$ son independientes

$$t = \frac{\frac{R\widehat{\beta} - R\beta}{\sqrt{\sigma^2 R(X'X)^{-1}R'}}}{\sqrt{\frac{(T-k)\widehat{\sigma}^2}{\sigma^2 (T-k)}}} = \frac{R\widehat{\beta} - R\beta}{\sqrt{\widehat{\sigma}^2 R(X'X)^{-1}R'}} = \frac{R\widehat{\beta} - R\beta}{SE(R\widehat{\beta})} \sim t_{T-k}$$

  Por tanto

$$Prob\left(-t_{T-k,\alpha/2} < \frac{R\widehat{\beta} - R\beta}{SE(R\widehat{\beta})} < t_{T-k,\alpha/2}\right) = 1-\alpha$$

  Manipulando esta expresión tenemos

$$\begin{aligned} Prob\left(-t_{T-k,\alpha/2} SE(R\widehat{\beta}) < R\widehat{\beta} - R\beta < t_{T-k,\alpha/2} SE(R\widehat{\beta})\right) &= 1-\alpha \\ Prob\left(R\widehat{\beta} - t_{T-k,\alpha/2} SE(R\widehat{\beta}) < R\beta < R\widehat{\beta} + t_{T-k,\alpha/2} SE(R\widehat{\beta})\right) &= 1-\alpha \end{aligned}$$

  El intervalo de confianza para $R\beta$ con nivel de confianza $100(1-\alpha)\%$ es

$$[R\widehat{\beta} - t_{T-k,\alpha/2} SE(R\widehat{\beta}), R\widehat{\beta} + t_{T-k,\alpha/2} SE(R\widehat{\beta})]$$

- Existe una relación entre los intervalos de confianza y los contraste de hipótesis: Se rechaza a nivel $\alpha$ la hipótesis nula $H_0 : R\beta = r$ frente a la alternativa $H_1 : R\beta \neq r$ si y sólo si $r$ no pertenece al intervalo de confianza para $R\beta$ con nivel de confianza $100(1-\alpha)\%$.