Análisis de Regresión Múltiple: Inferencia

Carlos Velasco[1]

[1] Departamento de Economía
Universidad Carlos III de Madrid

Econometría I
Máster en Economía Industrial
Universidad Carlos III de Madrid
Curso 2007/08

Resumen

1. Distribuciones muestrales de los Estimadores MCO
2. Contraste de Hipótesis sobre un único parámetro poblacional: el test de la t
3. Intervalos de Confianza
4. Contraste de hipótesis acerca de una única combinación lineal de parámetros
5. Contraste de múltiples restricciones lineales: el contraste de la F
6. Presentación de resultados de regresión

Contraste de hipótesis sobre los parámetros del modelo de regresión poblacional

- Distribución en el muestreo de los EMCO (normalidad).
- Contraste de hipótesis sobre parámetros individuales.
- Contraste de hipótesis sobre más de un parámetro.
- Contraste de restricciones múltiples (grupos de parámetros).

1. Distribuciones muestrales de los Estimadores MCO
2. Contraste de Hipótesis sobre un único parámetro poblacional: el test de la t
3. Intervalos de Confianza
4. Contraste de hipótesis acerca de una única combinación lineal de parámetros
5. Contraste de múltiples restricciones lineales: el contraste de la F
6. Presentación de resultados de regresión

- Hemos introducido supuestos que garantizan que los EMCO son **insesgados** y el cálculo de sus **varianzas** bajo las condiciones de Gauss-Markov.
- La varianza y el valor esperado de los EMCO son útiles para describir su **precisión.**
- Para hacer inferencia estadística se necesita addemás conocer la **distribución muestral** de los EMCO $\hat{\beta}_j$.
- La distribución de los EMCO depende de la distribución de los errrores.

RLM.6 (Normalidad) El error poblacional u es independiente de las variables explicativas $x_1, x_2, \ldots, x_k$ y se distribuye normalmente con media cero y varianza σ^2, $u \sim Normal\left(0, \sigma^2\right)$.

Distribuciones muestrales de los EMCO

- RLM.6 es un supuesto mucho más fuerte que los anteriores. Como u es independiente de x_j,

$$
\begin{aligned}
E\left(u|x_1, x_2, \ldots, x_k\right) &= E\left(u\right) = 0 \\
Var\left(u|x_1, x_2, \ldots, x_k\right) &= Var\left(u\right) = \sigma^2,
\end{aligned}
$$

  por lo que implica RLM.3 y RLM.5.
- RLM.1-6 se nombran como los **supuestos clásicos** del modelo de regresión lineal (**MLC**), y al modelo que satisface estos supuestos como el **modelo lineal clásico**: satisface las condiciones de Gauss-Markov junto con el supuesto de normalidad.
- Bajo **MLC** los EMCO $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k$ satisfacen una propiedad de eficiencia más fuerte que bajo Gauss-Markov: los EMCO son los estimadores **insesgados de mínima varianza** (ya no se requiere que sean lineales).

Distribuciones muestrales de los EMCO

- Un resumen de las condiciones MLC para el modelo poblacional es
$$y|\mathbf{x} \sim Normal\left(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k, \sigma^2\right),$$
donde $\mathbf{x} = (x_1, x_2, \ldots, x_k)$.
- **Justificación** de la normalidad: como u es el resultado de la suma de muchos factores inobservables que afectan a y, se puede invocar el Teorema Central del Límite.

Distribuciones muestrales de los EMCO (2)

- **Problemas**: cada factor puede tener distribuciones muy distintas en la muestra.
- El TCL asume aditividad de esos factores.
- En la práctica, esta es una cuestión empírica, no hay resultados que digan si $wage|educ, exper, tenure$ es normal.
  Pero si se puede razonar que hay problemas: *wage* no puede ser negativo (ni menor que un salario mínimo), por lo que en principio se puede pensar que la aproximación normal no es un buen supuesto.
  Tampoco lo es $narr86$, número de arrestos en 1986.
- A veces, ciertas transformaciones, como logs, hacen que las distribuciones se aproximen más a la normal.

Teorema 4.1 (Distribuciones muestrales normales). Bajos los supuestos MLC, RLM.1-6, condicional en los valores muestrales de las variables independientes,

$$\hat{\beta}_j \sim Normal\left(\beta_j, Var\left(\hat{\beta}_j\right)\right),$$

donde

$$Var\left(\hat{\beta}_j\right) = \frac{\sigma^2}{SST_j\left(1 - R_j^2\right)}$$

y por tanto

$$\frac{\hat{\beta}_j - \beta_j}{sd\left(\hat{\beta}_j\right)} \sim Normal\,(0, 1)\,.$$

Distribuciones muestrales de los EMCO

- Prueba: inmediata si tenemos en cuenta que

$$\hat{\beta}_j - \beta_j = \sum_{i=1}^{n} w_{ij} u_i$$

donde $w_{ij} = \hat{r}_{ij}/SSR_j$, y $\hat{r}_{ij}$ son los residuos de la regresión MCO de x_j sobre el resto de variables independientes y SSR_j es la suma de cuadrados de los residuos de dicha regresión.
- Por tanto podemos tomar w_{ij} como fijos y usar que la combinación lineal de normales independientes es también normal, con su correspondiente esperanza y varianza.
- La 2^a conclusión es simplemente una estandarización de $\hat{\beta}_j$.
- Además obtenemos que cualquier **combinación lineal** de $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k$ se distribuye también como una normal y que cualquier subconjunto de $\hat{\beta}_j$ también es normal.
- En el Tema 5 veremos que sin RLM.6 los $\hat{\beta}_j$ se distribuyen **aproximadamente como normales en muestras grandes.**

2. Contraste de Hipótesis sobre un único parámetro poblacional: el test de la *t*

Contraste de Hipótesis sobre un único parámetro poblacional: el test de la t

Modelo poblacional, bajo MLC,

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + u.$$

- El problema es hacer contrastes de hipótesis sobre β_j.
- β_j es una propiedad desconocida de la población, y nunca la conoceremos con certeza, pero podemos hacer **hipótesis** sobre su valor, y hacer **inferencia estadística** para contrastar nuestra hipótesis.

Contraste de Hipótesis sobre un único parámetro poblacional: el test de la t

Teorema 4.2 (Distribución del estimador MCO estandarizado). Bajos los supuestos MLC, RLM.1-6,

$$\frac{\hat{\beta}_j - \beta_j}{se\left(\hat{\beta}_j\right)} \sim t_{n-k-1},$$

donde $k+1$ es el número de parámetros desconocidos en el modelo poblacional $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + u$.

<u>Diferencias con el Teorema 4.1:</u>

- En el Teorema 4.1 teníamos que $\left(\hat{\beta}_j - \beta_j\right)/sd\left(\hat{\beta}_j\right)$ era $N(0,1)$, pero ahora se estima σ, por lo que la distribución es t.
- Entonces se puede escribir $\left(\hat{\beta}_j - \beta_j\right)/se\left(\hat{\beta}_j\right)$ como el cociente de una variable $N(0,1)$ y una χ^2_{n-k-1}, independientes.

Contraste de Hipótesis sobre un único parámetro poblacional: el test de la t

- El Teorema 4.2 es importante porque permite contrastar hipótesis sobre β_j.
- En la mayoría de las aplicaciones, el interés primordial es contrastar la **hipótesis nula**

$$H_0 : \beta_j = 0.$$

- Significado: como β_j mide el efecto parcial de x_j sobre (el valor esperado de) y, después de controlar por todas las otras variables independientes, H_0 significa que, una vez que se ha tenido en cuenta $x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_k$, x_j **no tiene efecto** sobre el valor esperado de y.

$$\log\left(wage\right) = \beta_0 + \beta_1 educ + \beta_2\ exper + \beta_3 tenure + u$$

- La hipótesis nula
$$H_0 : \beta_2 = 0$$
significa que, una vez que se ha tenido en cuenta educación y antigüedad, el número de años en la población activa no tiene efecto sobre el salario por hora.
- Esto tiene importantes consecuencias económicas: si esto es cierto, la historia previa de la persona al empleo actual no tiene influencia sobre su salario.
- Si $\beta_2 > 0$ querría decir que la experiencia pasada contribuye a la productividad y por tanto al salario.

- Es similar al contraste de la media de la población.
- Incluye obtener los estimadores de los coeficientes, sus errores estándar, el estadístico de contraste, y los valores de contraste: todo esto lo hace el Eviews.
- El estadístico de contraste se llama **estadístico** t o cociente t (t ratio) de $\hat{\beta}_j$:

$$t_{\hat{\beta}_j} = \frac{\hat{\beta}_j}{se\left(\hat{\beta}_j\right)}.$$

Contraste de Hipótesis: el test de la t

¿Porqué el **estadístico** t es razonable para contrastar H_0?

- $t_{\hat{\beta}_j}$ siempre tiene el signo de $\hat{\beta}_j$, porque $se\left(\hat{\beta}_j\right)$ es positivo.
- Para un valor fijo de $se\left(\hat{\beta}_j\right)$, cuanto mayor $\hat{\beta}_j$, mayor $t_{\hat{\beta}_j}$, en valor absoluto.
- $\hat{\beta}_j$ es un estimador insesgado de β_j, pero nunca será igual a cero, con H_0 cierta o falsa.
- La pregunta clave es, ¿está $\hat{\beta}_j$ cerca o lejos de cero?
- Un valor de $\hat{\beta}_j$ lejano de cero proporciona evidencia en contra de H_0, pero la estimación está sometida a un error de muestreo.
- $t_{se(\hat{\beta}_j)}$ mide cuántas desviaciones estándar estimadas está $\hat{\beta}_j$ de cero.
- La regla de rechazo dependerá de la hipótesis alternativa correspondiente y del nivel de significación elegido.

Contraste de Hipótesis: el test de la t

Regla de rechazo

- Determinar la regla de rechazo dado un nivel de significación (probabilidad de rechazar H_0 cuando es cierta) requiere conocimiento de la distribución muestral de $t_{\hat{\beta}_j}$ cuando H_0 es verdad: t_{n-k-1}.
- Mantenemos el signo de $t_{\hat{\beta}_j}$ para hacer hacer hipótesis unilaterales.
- Clave: contrastamos hipótesis sobre los parámetros poblaciones, no sobre los estimadores obtenidos para una muestra en particular:
- Hipótesis absurdas:

$$H_0 : \hat{\beta}_j = 0. \qquad H_0 : 0{,}434 = 0.$$

Contraste de Hipótesis: el test de la t

Contraste de hipótesis unilaterales

$$H_1 : \beta_j > 0.$$

- Esto significa que no nos preocupamos de alternativas de H_0 para las que $H_1 : \beta_j < 0$. Por alguna razón descartamos que $\beta_j < 0$.
- Regla de decisión: necesitamos un **nivel de significación** o probabilidad de rechazar H_0 cuando es cierta (error de tipo I), por ejemplo 5 %.
- Como $t_{\hat{\beta}_j}$. se distribuye como una t bajo H_0, por lo que tiene media cero, bajo la alternativa $\beta_j > 0$, el valor esperado de $t_{\hat{\beta}_j}$ es positivo: buscamos un valor suficientemente grande (y positivo) de $t_{\hat{\beta}_j}$ para rechazar $H_0 : \beta_j = 0$ en favor de $H_1 : \beta_j > 0$.
- Valores negativos de $\hat{\beta}_j$ o $t_{\hat{\beta}_j}$ no proporcionan evidencia en favor de H_1.

Contraste de Hipótesis: el test de la t

Contraste de hipótesis unilaterales (2)

- La definición de "suficientemente grande", con un 5 % de nivel de significación, es el percentil 95 % en una t_{n-k-1}, c.
- Por tanto la regla de rechazo es que H_0 se rechaza en favor de H_1, al nivel de significación del 5 %, si

$$t_{\hat{\beta}_j} > c.$$

- Por nuestra elección de c, esto ocurrirá el 5 % de todas las muestras cuando H_0 sea cierta.
- Este es un contraste de una sola cola (unilateral): Un valor negativo de $t_{\hat{\beta}_j}$ nunca rechazará H_0.
- Para otros valores se procederá igual: mayor nivel de significación implicará que c es más pequeño, y por tanto se rechazará H_0 más veces cuando es cierta.
- Por tanto, si H_0 se rechaza al 5 %, también se rechazará al 10 %.
- Cuanto más grande es n, más se parece la t a la normal.

$$
\begin{aligned}
\widehat{\log(wage)} &= \underset{(0{,}1042)}{0,2844} + \underset{(0{,}0073)}{0,092}\, educ + \underset{(0{,}0017)}{0,0041}\, exper + \underset{(0{,}0031)}{0,0221}\, tenure \\
n &= 526, \quad R^2 = 0,316 \quad \text{WAGE1.RAW}
\end{aligned}
$$

- Hipótesis: el efecto de *exper*, una vez descontado el efecto de *educ* y *tenure*, es cero en la población. Alternativa: es positivo.
- $$H_0 : \beta_{exper} = 0 \quad \text{contra} \quad H_1 : \beta_{exper} > 0.$$
- $$t_{\hat{\beta}_{exper}} = \frac{0,00412111}{0,0017233} = 2{,}3914,$$

  por tanto $\hat{\beta}_{exper}$, o *exper*, es estadísticamente significativo incluso al 1 por ciento (o $\hat{\beta}_{exper}$ es estadísticamente mayor que cero.)
- El efecto no es muy grande (tres años más de experiencia aportan un salario 1.2 % mayor,) pero ese efecto es positivo.

- Hipótesis alternativa
$$H_1 : \beta_j < 0.$$
- La regla de rechazo es que H_0 se rechaza en favor de H_1, al nivel de significación del 5 %, si
$$t_{\hat{\beta}_j} < -c.$$

- En aplicaciones es frecuente contrastar la hipótesis nula $H_0 : \beta_j = 0$ contra una **alternativa bilateral**,

$$H_1 : \beta_j \neq 0.$$

- Bajo esta alternativa, x_j tiene un efecto ceteris paribus sobre y sin especificar si éste es positivo o negativo. Esta es la alternativa relevante cuando el signo de β_j no está determinado por la la teoría (o el sentido común). Incluso si se se conoce el signo de β_j bajo la alternativa, un test bilateral es en general prudente (ya que por ejemplo no depende del signo de $\hat{\beta}_j$).
- Si no se especifica una alternativa concreta, en general se hace el contraste bilateral, y el 5 % será el nivel de significación por defecto. Eviews y la mayoría de los paquetes hacen este contraste también por defecto para todos los x_j.

- Regla de rechazo: H_0 se rechaza en favor de H_1, al nivel de significación del 5 %, si

$$\left| t_{\hat{\beta}_j} \right| > c,$$

donde c es el valor crítico que deja en cada área 2,5 % de probabilidad en una distribución t_{n-k-1}.

- Si H_1 se **rechaza** en favor de H_1 al 5 % se habla de que x_j es **estadísticamente significativo** (o estadísticamente diferente de cero.
- Si se rechaza a niveles de significación muy pequeños, entonces se habla que x_j es estadísticamente significativo a cualquier nivel de significación *habitual* (convencional).

- En general podemos estar interesados en contrastar

$$H_0 : \beta_j = a_j$$

donde a_j es un valor hipotético para β_j.

- El estadístico t apropiado es

$$t = \frac{\hat{\beta}_j - a_j}{se\left(\hat{\beta}_j\right)}.$$

- t mide el número de desviaciones estándar estimadas que $\hat{\beta}_j$ está del valor hipotético a_j.
- El estadístico t se puede usar para contrastar H_0 en contra de alternativas uni- o bilaterales, con el mismo valor crítico.
- Unilaterales: Si se rechaza H_0 en favor de $H_1 : \beta_j > 1$, se dice que $\hat{\beta}_j$ es estadísticamente mayor que 1 al nivel de significación correspondiente.

- Modelo con elasticidad constante:

$$\log(crime) = \beta_0 + \beta_1 \log(enroll) + u.$$

- En este caso la hipótesis de interés es

$$H_0 : \beta_1 = 1.$$

Si $\beta_1 > 1$ entonces la delincuencia se incrementa más que proporcionalmente con el tamaño del campus, y afecta relativamente más a campus grandes.

- Para comparar las situaciones $\beta_1 <, =, > 1$ en términos absolutos se puede usar el modelo en niveles

$$crime = \exp(\beta_0)\, enroll^{\beta_1} \exp(u).$$

- Ecuación estimada:

$$\begin{aligned} \widehat{\log(crime)} &= \underset{(1{,}03)}{-6,63} + \underset{(0{,}11)}{1{,}27} \log(enroll) \\ n &= 97, \quad R^2 = 0,585 \end{aligned}$$

- $\hat{\beta}_1 > 1$, pero hay evidencia para rechazar $H_0 : \beta_1 = 1$ en contra de

$$H_1 : \beta_1 > 1?$$

- Contraste:

$$t = \frac{\hat{\beta}_j - a_j}{se\left(\hat{\beta}_j\right)} = \frac{1{,}27 - 1}{0{,}11} = \frac{0{,}27}{0{,}11} \approx 2{,}45.$$

El valor crítico de un lado de una $t_{97-2} = t_{95}$ al nivel del 5 % es 1,66 (aproximando con la t_{120}), por lo que rechazamos H_0 claramente en favor de H_1. (Valor crítico al 1 %: 2.37).

- Método tradicional para contrastar hipótesis: elegir un **nivel de significación** y buscar un **valor crítico** para comparar con el t.
- Esto tiene un grado de **arbitrariedad**: la elección del nivel de significación antes del contrate, pero no hay uno correcto o mejor.
- Este procedimiento omite información importante sobre el resultado de un contraste: sólo sabemos que para un nivel de significación particular se acepta o rechaza H_0, pero no que ocurre a cualquier otro.
- Mejor responder: dado el valor del estadístico t, ¿cuál es el *menor* nivel de significación para el que la hipótesis nula se rechaza? El p-valor es una probabilidad $\in (0, 1)$.
- Este valor es el **p-valor** del contraste. Se obtiene usando la distribución de una variable t_{n-k-1} y comparándola con el valor del estadístico t.

- Los paquetes (Eviews) calculan el p-valor para contrastes de hipótesis particulares:

  $$H_0 : \beta_j = 0, \quad \text{en contra de } H_1 : \beta_j \neq 0,$$

  para el que

  $$p\text{-valor} = \Pr\left(|t_{n-k-1}| > |t|\right)$$

  donde t es el valor numérico del contraste.
- El p-valor resume la fuerza o debilidad de la evidencia empírica contra la hipótesis nula.
- El p-valor es la probabilidad de observar un estadístico t tan extremo (o más) como el que obtuvimos si la hipótesis nula fuese cierta:
  - p-valores pequeños son evidencia en contra H_0.
  - p-valores grandes son **poca** evidencia contra H_0.

Si tenemos que los grados de libertad son $n - k - 1 = 40$ y $t = 1,85$, entonces

$$p\!-\!valor = \Pr(|t_{40}| > |1,85|) = 2\Pr(t_{40} > 1,85) = 2{*}0,0359 = 0{,}0718.$$

Esto proporciona cierta evidencia en contra de H_0, pero no se rechazará H_0 en favor de H_1 al 5 %.

- ¿Qué ocurrirá al 10 %?
- ¿Y al 1 %?
- Si obtenemos el p-valor podemos efectuar contrastes clásicos a cualquier nivel de significación:
  - Si α es mayor que el p-valor rechazamos H_0 al nivel de significación α
  - si α es menor que el p-valor no podemos rechazar H_0.

- $$H_0 : \beta_j = 0, \quad \text{en contra de } H_1 : \beta_j > 0.$$
- Si $\hat{\beta}_j < 0$, entonces calcular el p-valor no es importante: es mayor que 0,5, por lo que nunca rechazaremos H_0 en favor de H_1.
- Si $\hat{\beta}_j > 0$ entonces $t > 0$ y

$$p\text{-valor} = \Pr\left(t_{n-k-1} > t\right).$$

- Este p-valor es igual a la mitad del p-valor de la hipótesis bilateral.
- Si

$$H_0 : \beta_j = 0, \quad \text{en contra de } H_1 : \beta_j < 0.$$

- Si $\hat{\beta}_j < 0$ entonces $t < 0$ y

$$p\text{-valor} = \Pr\left(t_{n-k-1} < t\right).$$

- Cuando H_0 no es rechazada hablamos de que "**no podemos rechazar** H_0 al nivel α", en lugar de que "se acepta H_0 al nivel α."
- La **significación estadística** de una variable x_j viene determinada por el tamaño de $t_{\hat{\beta}_j}$.
- La **significación económica** o la significación práctica de una variable está relacionada con el tamaño (y el signo) de $\hat{\beta}_j$.
- $t_{\hat{\beta}_j}$ puede indicar significatividad estadística bien porque $\hat{\beta}_j$ es grande o porque $\mathrm{se}\left(\hat{\beta}_j\right)$ es pequeño (por ejemplo porque n es muy grande): una variable puede ser estadísticamente significativa pero tener un efecto pequeño para explicar y.
- Ejemplo: Planes de pensiones 401(k) y el efecto de *totemp* :

$$prate = \beta_0 + \beta_1 mrate + \beta_2 age + \beta_3 totemp + u.$$

Contraste de Hipótesis: el test de la t

Comentarios (2)

- A veces se sugiere cambiar los niveles de significación con el tamaño muestral.
- Los errores estándar pueden a veces ser muy altos por un problema de multicolinealidad:
  - Deberemos aumentar el tamaño muestral.
  - O cambiar el objetivo del análisis.

- Guía:
  - Comprobar la significación estadística de una variable y luego discutir la significación económica.
  - Si una variable no es significativa a los niveles habituales, todavía nos podemos plantear si tiene el efecto esperado sobre y y si este efecto es grande en términos prácticos. Si es grande, se debe computar el p-valor.
- Es frecuente encontrar variables con estadísticos t pequeños que tienen el signo equivocado: en la práctica estas variables deben ser ignoradas.
  Una variable significativa estadísticamente y con el signo cambiado es mucho más preocupante y difícil de resolver. Hay que repensar el modelo (omisión de variables, forma funcional, etc.).

Intervalos de Confianza

- También estimadores por intervalos: Proporcionan un rango de valores probables para el valor del parámetro poblacional.
- Intervalo al 95 % :

$$\hat{\beta}_j \pm c \cdot se\left(\hat{\beta}_j\right)$$

  donde c es el cuantil 97.5 de la distribución t_{n-k-1}.
- Alternativamente

$$\left(\hat{\beta}_j - c \cdot se\left(\hat{\beta}_j\right), \hat{\beta}_j + c \cdot se\left(\hat{\beta}_j\right)\right).$$

- Significado: si se obtienen sucesivas muestras aleatorias y se computa el intervalo de confianza para cada una de ellas, entonces el valor poblacional β_j (desconocido) estará contenido en el intervalo de confianza para un 95 % de las muestras.
- Para un intervalo de confianza en particular, no hay ninguna seguridad de que contenga el valor β_j.

Intervalos de Confianza (2)

- Para construir el intervalo de confianza hacen falta tres elementos:
  - $\hat{\beta}_j$.
  - c, que depende del cuantil de la t_{n-k-1}. Si n es grande se puede aproximar por la normal y si el nivel de confianza es del 95 % podemos aproximar $c \approx 2$.
  - $se\left(\hat{\beta}_j\right)$.
- Para otros niveles de confianza se opera igual: a mayor nivel de confianza, más amplio es el intervalo.
- Con un intervalo de confianza es inmediato realizar **contrastes de hipótesis bilaterales**:

  Si la hipótesis nula es $H_0 : \beta_j = a_j$, entonces H_0 es rechazada en favor de H_1:$\beta_j \neq a_j$ al nivel de significación del 5 % si y sólo si a_j no está contenida en el intervalo de confianza al 95 %.

$$
\begin{aligned}
\widehat{\log(price)} &= \underset{(1{,}15)}{7{,}46} + \underset{(,184)}{0{,}634} \log(sqrft) - \underset{(,059)}{0{,}066} bdrms + \underset{(,075)}{0{,}158} bthrms \\
n &= 19, \quad R^2 = 0{,}806
\end{aligned}
$$

- La elasticidad respecto a la superficie es .634.
- Usando la distribución $t_{n-k-1} = t_{15}$ podemos obtener el cuantil 97.5 % igual a $c = 2{,}131$.
- Intervalo de confianza al 95 % para $\beta_{\log(sqrft)}$:

$$0{,}634 \pm 2{,}131 \cdot 0{,}184 \quad \equiv \quad (,242, 1{,}026)\,.$$

- Se rechaza $H_0 : \beta_{\log(sqrft)} = 0$ en favor de la altern. bilateral al 5 %.
- El coeficiente de *bdrms* es negativo: pensar interpretación cp.

Intervalos de Confianza

Ejemplo: Modelo para precios hedónicos de viviendas (2)

$$\begin{aligned}\widehat{\log(price)} &= \underset{(1,15)}{7,46} + \underset{(,184)}{0,634}\log(sqrft) - \underset{(,059)}{0,066}bdrms + \underset{(,075)}{0,158}bthrms \\ n &= 19, \quad R^2 = 0,806\end{aligned}$$

- Intervalo de confianza para β_{bdrms} :

$$-0,066 \pm 2,131 \cdot 0,059 \quad \equiv \quad (-,192, ,060).$$

- $H_0 : \beta_{bdrms} = 0$?
- IC para β_{bthrms}:

$$(-0,002, ,318).$$

- $H_0 : \beta_{bthrms} = 0$?

4. Contraste de hipótesis acerca de una única combinación lineal de parámetros

Contraste de hipótesis acerca de una única combinación lineal de parámetros

- En la práctica es frecuente tener que contrastar hipótesis sobre más de un parámetro poblacional (y también múltiples hipótesis).
- EJEMPLO: modelo para comparar el rendimiento de la educación en escuelas universitarias y facultades para una población de trabajadores con el título de bachillerato:

$$\log(wage) = \beta_0 + \beta_1 jc + \beta_2 univ + \beta_3 exper + u$$

- $jc =$ número de años como estudiante en una escuela universitaria.
- $univ =$ número de años en una universidad.
- Hipótesis de interés: un año en una escuela universitaria tiene el mismo valor que un año en la universidad, cp:

$$H_0 : \beta_1 = \beta_2$$

Contraste de hipótesis acerca de una única combinación lineal de parámetros

- Hipótesis alternativa (unilateral): un año en la escuela primaria vale menos que un año en una universidad:

$$H_1 : \beta_1 < \beta_2.$$

- H_0 y H_1 se refieren a dos parámetros, β_1 y β_2 : no es suficiente con los estadísticos t individuales.
- Se reescriben las hipótesis:

$$H_0 : \beta_1 - \beta_2 = 0, \quad H_1 : \beta_1 - \beta_2 < 0.$$

- Nuevo estadístico t, para comparar $\hat{\beta}_1 - \hat{\beta}_2$ con $\beta_1 - \beta_2$ bajo H_0:

$$t = \frac{\hat{\beta}_1 - \hat{\beta}_2}{se\left(\hat{\beta}_1 - \hat{\beta}_2\right)}.$$

- Regla de rechazo: se rechaza H_0 cuando $t < -c$ donde c es un valor crítico de la distribución t.
- Única novedad: cálculo del error estándar.

$$\log(wage) = \beta_0 + \beta_1 jc + \beta_2 univ + \beta_3 exper + u$$

$$\begin{aligned}
\widehat{\log(wage)} &= \underset{(0{,}021)}{1,4723} + \underset{(0{,}0068)}{0,0667} jc + \underset{(0{,}0023)}{0,0769} univ + \underset{(0{,}00016)}{0,0049}\, exper \\
n &= 6763 \quad R^2 = 0,222
\end{aligned}$$

- Se predice una diferencia porcentual de aproximadamente un 1 % :

$$\hat{\beta}_1 - \hat{\beta}_2 = -0{,}0102.$$

Contraste de hipótesis acerca de una única combinación lineal de parámetros

Ejemplo: salarios, TWOYEAR (2)

$$\begin{aligned}
\widehat{\log(wage)} &= \underset{(0{,}021)}{1,4723} + \underset{(0{,}0068)}{0,0667}jc + \underset{(0{,}0023)}{0,0769}univ + \underset{(0{,}00016)}{0,0049}\, exper \\
n &= 6763 \quad R^2 = 0,222
\end{aligned}$$

- No hay información para calcular $se\left(\hat{\beta}_1 - \hat{\beta}_2\right)$, pero

$$Var\left(\hat{\beta}_1 - \hat{\beta}_2\right) = Var\left(\hat{\beta}_1\right) + Var\left(\hat{\beta}_2\right) - 2Cov\left(\hat{\beta}_1, \hat{\beta}_2\right)$$

  y por tanto

$$se\left(\hat{\beta}_1 - \hat{\beta}_2\right) = \left\{ se\left(\hat{\beta}_1\right)^2 + se\left(\hat{\beta}_2\right)^2 - 2\widehat{Cov}\left(\hat{\beta}_1, \hat{\beta}_2\right) \right\}^{1/2}$$

- Eviews proporciona $\widehat{Cov}\left(\hat{\beta}_1, \hat{\beta}_2\right)$.

Contraste de hipótesis acerca de una única combinación lineal de parámetros

Otro método

- Basado en un modelo alternativo: evita tener que calcular $\widehat{Cov}\left(\hat{\beta}_1, \hat{\beta}_2\right)$ y $se\left(\hat{\beta}_1 - \hat{\beta}_2\right)$.
- Se basa en una reparametrización del modelo lineal:

$$\theta_1 = \beta_1 - \beta_2$$

y el contraste se realiza acerca de

$$H_0 : \theta_1 = 0, \quad H_1 : \theta_1 < 0$$

a partir de

$$t = \frac{\hat{\theta}_1}{se\left(\hat{\theta}_1\right)}$$

- θ_1 debe ser el coeficiente de una de las variables explicativas del modelo.

Contraste de hipótesis acerca de una única combinación lineal de parámetros

Otro método (2)

- Usando $\theta_1 = \beta_1 - \beta_2$, y por tanto $\beta_1 = \theta_1 + \beta_2$,

$$\log(wage) = \beta_0 + \overbrace{(\theta_1 + \beta_2)}^{=\beta_1} jc + \beta_2 univ + \beta_3\ exper + u$$

$$\log(wage) = \beta_0 + \theta_1 jc + \beta_2 (jc + univ) + \beta_3\ exper + u$$

- La clave es que el parámetro de interés, θ_1, es ahora el coeficiente de una variable y se puede estimar directamente por MCO.
- Además hay una nueva variable que multiplica a β_3, $jc + univ = totcoll$, en lugar de $univ$,

$$\log(wage) = \beta_0 + \theta_1 jc + \beta_2 totcoll + \beta_3\ exper + u.$$

- El parámetro β_1 ha desaparecido del modelo, mientras que θ_1 aparece explícitamente, e Eviews reportará $se\left(\hat{\theta}_1\right)$.

Contraste de hipótesis acerca de una única combinación lineal de parámetros

Otro método (3)

- Si estimamos por MCO:

$$\begin{aligned} \widehat{\log(wage)} &= \underset{(0{,}02106)}{1,47233} - \underset{(0{,}0069)}{0,0102}jc + \underset{(0{,}00016)}{0,0769}totcoll + \underset{(0{,}00016)}{0,0049}\, expe \\ n &= 6763 \quad R^2 = 0,222 \end{aligned}$$

- El único resultado nuevo es el *se* de $\hat{\theta}_1$, $se\left(\hat{\theta}_1\right) = 0,0069$:

$$t_{\theta_1} = \frac{-0{,}0102}{0{,}0069} = -1{,}478.$$

  El p-valor contra una alternativa unilateral es sobre 0.7, por lo que hay evidencia contra H_0, pero no muy fuerte.
- El coeficiente y el *se* de *exper* deben ser iguales.
- El coeficiente de *totcoll* y su *se* deben ser iguales a los que tenía *univ*.

5. Contraste de múltiples restricciones lineales: el contraste de la F

Contraste de múltiples restricciones lineales: el contraste de la F

Contraste de restricciones de exclusión

- Hasta ahora sólo hemos abordado problemas sobre una sola restricción.
- El primer problema es contrastar si un **grupo** de variables no tiene efecto sobre la variable dependiente, una vez que se ha controlado por otro grupo de variables.

$$\log(salary) = \beta_0 + \beta_1 years + \beta_2 gamsyr + \beta_3 bavg + \beta_4\ hrunsyr + \beta_5 rbisyr + u$$

- *years* : años en la liga.
- *gamesyr* = número medio de partidos jugados al año.
- *bavg* = número medio de bateos.
- *hrunsyr* = número de home runs por año
- *rbisyr* = carreras con bateo por año.

- **Contraste de hipótesis múltiple** o **contraste conjunto**: La hipótesis de interés es que una vez que se ha controlado por el número de años y de partidos por año, las estadísticas que miden el rendimiento, *bavg*, *hrunsyr* y *rbisyr*, no tienen efecto:

$$H_0 : \beta_3 = 0, \beta_4 = 0, \beta_5 = 0.$$

- Son tres **restricciones de exclusión**: si H_0 es cierta, entonces *bavg*, *hrunsyr* y *rbisyr* no deberán estar en el modelo.

Contraste de restricciones de exclusión

- Hipótesis alternativa:

$$H_1 : \quad H_0 \text{ no es cierta.}$$

- H_1 es cierta si al menos uno de β_3, β_4 ó β_5 es diferente de cero.
- El contraste va a detectar cualquier violación de H_1, como que $\beta_3 > 0$ o que $\beta_5 < 0$, pero en ese caso no será el mejor contraste.
- ¿Cómo proceder? Una primera posibilidad es usar los estadísticos individuales de la t para las tres variables para determinar si cada una de las variables puede ser eliminada individualmente.
- Sin embargo esto no es apropiado: un estadístico t no impone restricciones sobre los otros parámetros y no está claro cuántos de ellos deben rechazar para realizar un contraste al 5 % de H_0.

Contraste de restricciones de exclusión

EJEMPLO: salarios de una liga de baseball (2) MLB1

$$\begin{aligned}\log\widehat{(salary)} &= \underset{(0{,}289)}{11,20} + \underset{(0{,}012)}{0,0689}years + \underset{(0{,}0026)}{0,0126}gamsyr \\ &+ \underset{(0{,}0011)}{0,00098}bavg + \underset{(0{,}016)}{0,0144}hrunsyr + \underset{(0{,}0072)}{0,0108}rbisyr \\ n &= 353, SSR = 183{,}186, R^2 = 0,6278\end{aligned}$$

- *years* y *gamsyr* son significativas.
- Ninguna de *bavg*, *hrunsyr* y *rbisyr* es significativa individualmente usando las t: aparentemente no podemos rechazar H_0.
- Sin embargo esta conclusión puede estar equivocada: debemos hacer un contraste conjunto de todas las restricciones.

- El nuevo método va a usar la *SSR* del ajuste (o alternativamente el R^2 en el caso de restricciones de exclusión).
- *SSR* no dice nada acerca de que H_0 sea cierta o no: lo importante es cuánto se incrementa *SSR* cuando eliminamos las variables *bavg*, *hrunsyr* y *rbisyr* del modelo.

Contraste de restricciones de exclusión
EJEMPLO: salarios de una liga de baseball

- **Modelo no restringido**:

$$\log(salary) = \beta_0 + \beta_1 years + \beta_2 gamsyr + \beta_3 bavg + \beta_4\, hrunsyr + \beta_5 rbisy$$

- **Modelo restringido**:

$$\log(salary) = \beta_0 + \beta_1 years + \beta_2 gamsyr + u$$

- Modelo restringido ajustado:

$$\begin{aligned} \widehat{\log(salary)} &= \underset{(0{,}11)}{11,22} + \underset{(0{,}0125)}{0,0713}\, years + \underset{(0{,}0013)}{0,0202}\, gamsyr \\ n &= 353, SSR = 198{,}311, R^2 = 0,5971. \end{aligned}$$

- *SSR* aumenta, como se esperaba, *y* R^2 baja también: hay que decidir si el descenso de *SSR* al pasar del modelo sin restringir al restringido es significativo para que se pueda rechazar H_0.
- Se necesita un estadístico de contraste de distribución conocida y un nivel de significación.

Contraste de restricciones de exclusión: Estadístico de la F

- Modelo no restringido:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + u.$$

- Hipótesis nula (q restricciones):

$$H_0 : \beta_{k-q+1} = 0, \ldots, \beta_k = 0.$$

- Modelo restringido:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_{k-q} x_{k-q} + u.$$

- Estadístico de la F :

$$F = \frac{(SSR_r - SSR_{nr})/q}{SSR_{nr}/(n-k-1)} = \frac{SSR_r - SSR_{nr}}{SSR_{nr}} \frac{n-k-1}{q}$$

  - SSR_{nr} : suma de cuadrados de residuos en el modelo no restringido.
  - SSR_r : suma de cuadrados de residuos en el modelo restringido.

Contraste de restricciones de exclusión: Estadístico de la F

- F nunca es negativo.
- $q =$ grados de libertad del numerador $= gl_r - gl_{nr}$, donde

  $gl =$ número de observaciones $-$ número de parámetros estimados.

- $n - k - 1 =$ grados de libertad del denominador: gl_{nr}.
- El denominador de F es un estimador insesgado de σ^2.
- Ejemplo:

$$\begin{aligned} n &= 353 \\ n - k - 1 &= 353 - 5 - 1 = 347 \\ q &= 3 \end{aligned}$$

- Bajo H_0,
$$F \sim F_{q,n-k-1}.$$
- Rechazaremos H_0 cuando F sea muy grande comparado con la distribución $F_{q,n-k-1}$,
$$F > c.$$
- Si el nivel de significación es del 5 %, eligiremos el cuantil del 5 %.
- En general q será pequeño mientras que $n-k-1$ será grande: cuando $n-k-1 > 120$ la distribución no cambia.
- Si se rechaza H_0 entonces se dice que $x_{k-q+1}, \ldots, x_k$ son **significativos conjuntamente** al nivel de significación apropiado. Pero no sabemos qué variables tienen un efecto parcial sobre y : puede que sean todas o sólo unas de ellas.
- Si no se rechaza H_0: "las variables son **conjuntamente no significativas**"

- $n - k - 1 = 347$; $q = 3$
- $F_{3,347}$: $c = 2{,}60$ al 5 %, $c = 3{,}78$ al 1 %.
- $$F = \frac{(198{,}311 - 183{,}186)}{183{,}186} \frac{347}{3} \approx 9{,}55.$$
- Este valor está por encima del valor crítico al 1 % de la $F_{3,347}$, por lo rechazamos con bastante fuerza la hipótesis nula de que *bavg*, *hrunsyr* y *rbisyr* no tienen efecto sobre el salario.
- Resultado contradictorio con los contrastes *t* individuales.
- El problema es que *hrunsyr* y *rbisyr* están muy correladas y esta **multicolinealidad** hace muy complicado aislar el efecto parcial de cada una de las variables.
- Multicolinealidad es mucho menos relevante para contrastar hipótesis conjuntas acerca de variables muy correladas.
- Intenta eliminar *hrunsyr* o *rbisyr* del modelo

Relación entre el Estadístico F y el Estadístico t

¿Qué ocurre si se usa el estadístico F para contrastar si una sola variable es significativa?

- Sería aplicar el método anterior cuando $q = 1$, por ejemplo

$$H_0 : \beta_k = 0.$$

- Respuesta: ambos métodos son esencialmente iguales:

$$F = t^2,$$

  y como una variable t^2_{n-k-1} tiene una distribución $F_{1,n-k-1}$, los dos métodos proporcionan la misma respuesta.

- El estadístico t es más flexible para contrastar una sola restricción ya que permite dirigir el contraste contra alternativas unilaterales.

Estadístico F en función del R^2

- En muchas ocasiones se puede usar una forma del estadístico F en función de los R^2 de los modelos restringido y no restringido.
- Los R^2 pueden ser más convenientes porque están entre 0 y 1 y no depende de las unidades de medida.
  Además se reporta casi siempre, pero quizá no el SSR.
- La idea es sustituir

$$\begin{aligned} SSR_r &= SST\left(1 - R_r^2\right) \\ SSR_{nr} &= SST\left(1 - R_{nr}^2\right) \end{aligned}$$

- Entonces

$$F = \frac{\left(R_{nr}^2 - R_r^2\right)/q}{\left(1 - R_{nr}^2\right)/(n-k-1)} = \frac{R_{nr}^2 - R_r^2}{1 - R_{nr}^2}\frac{n-k-1}{q}.$$

- Como $R_{nr}^2 > R_r^2$ el estadístico F es siempre positivo.

Estadístico F en función del R^2 (2)

- Ejemplo:

$$F = \frac{(,6278 - ,5971)}{1 - ,6278}\frac{347}{3} \approx 9{,}54.$$

Cálculo de $p-$ valores para contrastes de la F

- **P-valor:** Probabilidad de observar un valor de la F al menos tan grande como el que se ha obtenido, dada que la hipótesis nula es cierta,
$$p - \text{valor} = Pr\left(F_{q,n-k-1} > F\right).$$
- Permite efectuar contrastes a cualquier nivel de significación.
- Los paquetes econométricos tienen funciones que permiten calcular los p-valores para restricciones de exclusión.

El Estadístico F para el contraste de significación global

- Un caso particular muy importante de contraste es aquel en que

$$H_0 : x_1, x_2, \ldots, x_k \text{ no ayudan a explicar } y.$$

- Todos los paquetes lo realizan por defecto.
- Hipótesis nula (k restricciones):

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_k = 0.$$

- Alternativa: al menos uno de los β_j es diferente de cero.
- También

$$H_0 : E(y) = E(y|x_1, x_2, \ldots, x_k).$$

El Estadístico F para el contraste de significación global (2)

- Modelo restringido:

$$
\begin{aligned}
y &= \beta_0 + u, \\
\hat{y} &= \hat{\beta}_0 = \bar{y}, \quad R_r^2 = 0
\end{aligned}
$$

- Si $R^2 = R_{nr}^2$ de la regresión de y sobre $x_1, x_2, \ldots, x_k$,

$$
F = \frac{R^2/k}{(1 - R^2)/(n - k - 1)} = \frac{R^2}{1 - R^2} \frac{n - k - 1}{k}.
$$

- Si no podemos rechazar H_0, entonces no hay evidencia de que ninguna de las variables independientes ayuden a explicar y (debemos buscar otras variables que ayuden en ese caso).

- Ejemplo baseball:

$$F = \frac{0{,}6278}{1 - 0{,}6278} \frac{347}{5} = 117{,}06$$

  y el p-valor es $< 0{,}00001$: se rechaza H_0, y se concluye que las variables independientes conjuntamente contribuyen a explicar y.
- El resultado no depende solamente del tamaño del R^2.

$$\log(price) = \beta_0 + \beta_1 \log(assess) + \beta_2 \log(lotsize) + \beta_3 \log(sqrft) + \beta_4 bdrms + u$$

- *price* = precio de la vivienda.
- *assess* = valor de la casa de tasación
- *lotsize* = tamaño del lote
- *sqrft* = superficie en pies cuadrados
- *bdrms* = número de dormitorios

- Hipótesis nula: ver si el precio de tasación es una predicción racional del precio real de venta:

$$\beta_1 = 1.$$

- Además las otras variables no deberían ayudar a predecir el precio de venta una vez que se tiene en cuenta *assess*.
- 4 restricciones, tres de exclusión

$$H_0 : \beta_1 = 1, \beta_2 = 0, \beta_3 = 0, \beta_4 = 0.$$

- En este caso también se imponen las restricciones sobre el modelo **no restringido:**

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + u$$

para obtener $y = \beta_0 + x_1 + u$.
- El modelo **restringido** a estimar:

$$y - x_1 = \beta_0 + u.$$

- Cálculo del estadístico F : igual que antes:

$$F = \frac{(SSR_r - SSR_{nr})/q}{SSR_{nr}/(n-k-1)} = \frac{SSR_r - SSR_{nr}}{SSR_{nr}} \frac{n-k-1}{q}$$

- NO es posible usar la forma del estadístico F con R^2 porque la variable dependiente en los dos modelos es diferente: SST no se cancelaría!!

- Estimación modelo no restringido: $SSR_{nr} = 1{,}822$.
- Estimación modelo restringido: $SSR_r = 1{,}880$.
- $n = 88$

$$F = \frac{1{,}880 - 1{,}822}{1{,}822}\frac{83}{4} = 0{,}661$$

- El valor crítico al 5 % de una $F_{4,83}$ es $c = 2{,}50$, por lo que la hipótesis nula no se puede rechazar: no hay evidencia en contra de que los valores de tasación no son racionales.

6. Presentación de resultados de regresión

Presentación de resultados de regresión

- Ideas sobre como presentar los resultados y cómo se deben leer.
- Siempre hay que reportar los **estimadores MCO**. Para las variables clave hay que interpretar los coeficientes estimados (es una elasticidad? implicaciones económicas...)
- Siempre hay que acompañar los estimadores de los **errores estándar** (algunos autores prefieren reportar los estadísticos t).
- Mejor los errores estándar:
  - Fuerza a pensar en la hipótesis nula, que no siempre es $\beta_j = 0$.
  - Permite calcular intervalos de confianza fácilmente.
- Siempre se incluye el $\mathbf{R}^2$: es una medida de bondad de ajuste y permite el contraste general.
- También el número de observaciones, **n**.
- Si se estiman sólo dos modelos se pueden presentar en forma de ecuación. Si se estiman más mejor presentar las estimaciones en una tabla.

Para explicar la compensación anual total para un maestro, que incluye salario y todos los beneficios (pensiones, seguro médico, etc.) se postula

$$\log(totcomp) = f\,(productivity\ characteristics,\ other\ factor)\,.$$

Sin embargo:

$$totcomp = salary + benefits = salary\left(1 + \frac{benefits}{salary}\right)$$

y entonces

$$\begin{aligned}\log(totcomp) &= \log(salary) + \log\left(1 + \frac{benefits}{salary}\right)\\ &\approx \log(salary) + \frac{benefits}{salary}\end{aligned}$$

Esto lleva al modelo econométrico:

$$\log(salary) = \beta_0 + \beta_1 (b/s) + other\ \ factors$$

Para contrastar si hay un trade-off salario-beneficios se plantea:

$$H_0 : \beta_1 = -1, \quad H_1 : \beta_1 \neq -1$$

MEAP93: datos a nivel de colegio:

Presentación de resultados de regresión
Ejemplo: salarios y pensiones para maestros (3)

Variable Dependiente: log (*salary*)			
Variables Independientes	(1)	(2)	(3)
b/s	$-,825$ (,20)	$-,605$ (,16)	$-,589$ (,16)
log (*enroll*)	-	,0874 (,0073)	0,0881 (,0073)
log (*staff*)	-	$-,222$ (,050)	$-,218$ (,050)
droprate	-	-	$-,00028$ (,0016)
gradrate	-	-	,00097 (,00066)
intercept	10,523 (,042)	10,884 (,252)	10,738 (,258)
Observations	408	408	408
R^2	0.040	0.353	0.361