ECONOMETRÍA I

Tema 4: El Modelo de Regresión Lineal Múltiple: inferencia y validación

Patricia Moreno
Juan Manuel Rodriguez Poo
Alexandra Soberon
Departamento de Economía

Distribuciones muestrales de los estimadores MCO

- Hasta el momento hemos visto que, gracias a los supuestos de Gauss-Markov, los estimadores MCO son los Mejores Estimadores Lineales Insesgados (MELI).
- La varianza y el valor esperado de los estimadores MCO son útiles para describir su precisión.
- Sin embargo, para hacer inferencia estadística necesitamos conocer también la **distribución muestral** de los estimadores MCO, $\widehat{\beta}_j$.
- La distribución de los estimadores MCO depende de la distribución de los errores.

- **RLM.6:** *Normalidad*. El error poblacional u es independiente de las variables explicativas $X_1, X_2, \cdots, X_k$ y se distribuye normalmente con media cero y varianza σ^2,

$$u \sim \text{Normal}(0, \sigma^2).$$

- Bajo **RLM.6**, u es independiente de X_j, por lo que

$$\begin{aligned} E(u|X_1, X_2, \cdots, X_k) &= E(u) = 0 \\ Var(u|X_1, X_2, \cdots, X_k) &= Var(u) = \sigma^2. \end{aligned}$$

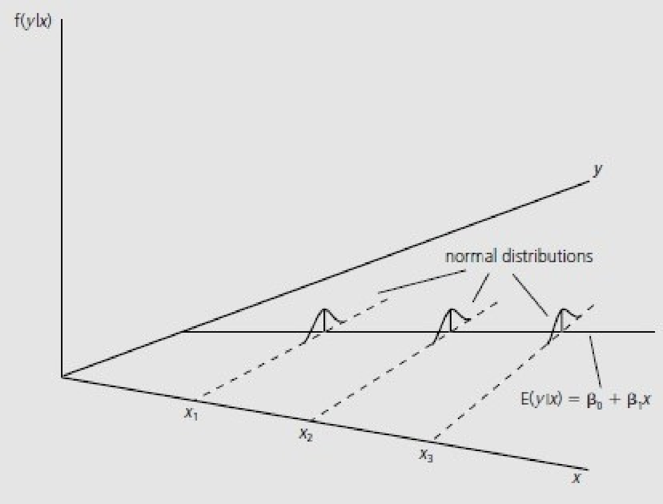
- **RLM.1-6** son conocidos como los supuestos clásicos del modelo de regresión lineal (MLC) y nos permiten resumir los supuestos poblacionales del MLC como

$$Y|X \sim \text{Normal}\left(\beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k, \sigma^2\right).$$

donde $X = (X_1, X_2, \cdots, X_k)$.

- Bajo MLC los estimadores MCO $\widehat{\beta}_0, \widehat{\beta}_1, \cdots, \widehat{\beta}_k$ satisfacen una propiedad de eficiencia más fuerte que bajo Gauss-Markov. Ahora éstos son los estimadores insesgados de mínima varianza (pero ya no necesitamos que sean lineales).

Distribución normal homocedástica con un regresor



Fuente: Wooldridge (2005)

Distribuciones muestrales Normales

Bajo los supuestos MLC, **RLM.1-6**, condicional en los valores muestrales de las variables independientes,

$$\widehat{\beta}_j \sim \mathrm{Normal}\left(0, Var(\widehat{\beta}_j)\right),$$

donde

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

y por lo tanto

$$\frac{\widehat{\beta}_j - \beta_j}{sd(\widehat{\beta}_j)} \sim \mathrm{Normal}(0, 1).$$

Queda demostrado que $\widehat{\beta}_j$ se distribuye como una Normal.

Contraste de hipótesis

- Bajo MLC, el modelo poblacional es

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + u.$$

- β_j es una propiedad desconocida de la población, pero nunca lo podemos conocer con certeza.
- Aunque β_j es desconocido, podemos hacer **hipótesis** sobre su valor y realizar **inferencia estadística** para contrastar esta hipótesis.

Contraste de hipótesis: el test de la t

- Bajo los supuestos MLC, **RLM.1-6**,

$$\frac{\widehat{\beta}_j - \beta_j}{ee(\widehat{\beta}_j)} \sim t_{n-k-1},$$

  donde $k+1$ es el número de parámetros desconocidos en el modelo poblacional y $(n-k-1)$ son los grados de libertad.

- Como σ^2 tiene que ser estimado por $\widehat{\sigma}^2$, este estimador MCO estandarizado es el cociente entre una variable $N(0,1)$ y una χ^2_{n-k-1}, independientes. Así, el estimador MCO estandarizado sigue una distribución t con $n-k-1$ grados de libertad.

- Conocer la distribución muestral del estimador estandarizado nos permite llevar a cabo contrastes de hipótesis.

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

Para realizar un contraste de hipótesis necesitamos:

1) **Hipótesis nula**: hipótesis que queremos contrastar.

$$H_0 : \beta_3 = 0$$

Ej: el número de años de antigüedad no tiene efecto sobre el salario por hora, una vez controlado por la educación y la experiencia.

2) **Hipótesis alternativa**: situación alternativa a la H_0. Puede ser una alternativa bilateral o unilateral.

$$H_1 : \beta_3 \neq 0$$

Ej: los años de antigüedad contribuyen a la productividad y de ahí al salario.

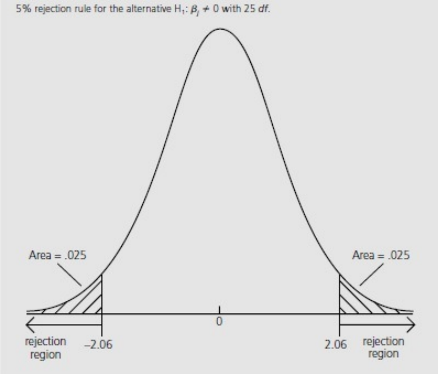
3) **Estadístico de Contraste** (t-ratio):

$$t_{\widehat{\beta}_j} = \frac{\widehat{\beta}_j}{ee(\widehat{\beta}_j)}$$

4) **Nivel de significación**: probabilidad de rechazar H_0 cuando ésta es cierta. Por ejemplo, $\alpha = 5\,\%$.

5) **Regla de decisión**: basándonos en un valor crítico c, determinar en qué situaciones rechazamos la H_0. El valor crítico se calcula como el percentil $(1-\alpha)th$ en una t_{n-k-1}. Así,
   - Rechazamos H_0 si $t_{\widehat{\beta}_j}$ es mayor que el valor crítico.
   - No rechazamos H_0 si $t_{\widehat{\beta}_j}$ es menor que el valor crítico.

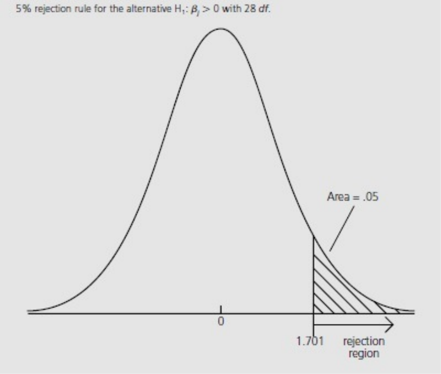
$$H_0 : \quad \beta_j = 0$$
$$H_1 : \quad \beta_j \neq 0$$



Fuente: Wooldridge (2005)

Contraste de hipótesis: contraste unilateral

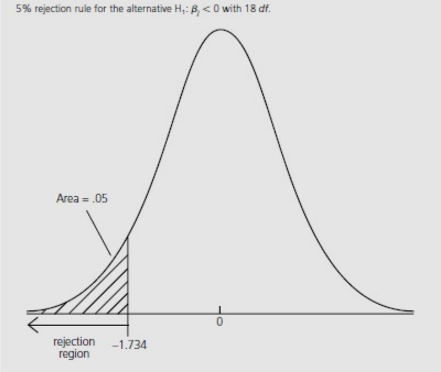
$$H_0 : \quad \beta_j = 0$$
$$H_1 : \quad \beta_j > 0$$



Fuente: Wooldridge (2005)

Contraste de hipótesis: contraste unilateral

$$H_0: \quad \beta_j = 0$$
$$H_1: \quad \beta_j < 0$$



Fuente: Wooldridge (2005)

Contraste de hipótesis: resumen

- **Reglas de rechazo**: Rechazamos H_0
  - **Contraste bilateral** ($H_1 : \beta_1 \neq 0$): si $\left| t_{\widehat{\beta_j}} \right| > c$.
  - **Contraste unilateral** ($H_1 : \beta_1 > 0$): si $t_{\widehat{\beta_j}} > c$.
  - **Contraste unilateral** ($H_1 : \beta_1 < 0$): si $t_{\widehat{\beta_j}} < -c$.
- Si H_0 se rechaza en favor de H_1 al $\alpha\,\%$, X_j es **estadísticamente significativa** (o estadísticamente distinto de cero) al $\alpha\,\%$ de significación.
- Si no rechazamos H_0 al $\alpha\,\%$, X_j es **estadísticamente no-significativa** al $\alpha\,\%$ de significación.

Contraste de hipótesis: otras hipótesis sobre β_j

- En ocasiones queremos probar si β_j toma un valor concreto a_j.
- **Hipótesis nula**: $H_0 : \beta_j = a_j$.
- **Estadístico de contraste**:

$$t_{\widehat{\beta}_j} = \frac{\widehat{\beta}_j - a_j}{ee(\widehat{\beta}_j)},$$

  donde $a_j = 0$ sería para el caso estándar.
- Este estadístico se puede usar para contrastes unilaterales y bilaterales.
- Contraste unilateral: Si se rechaza H_0 en favor de $H_1 : \beta_j > a_j$ tenemos que $\widehat{\beta}_j$ es estadísticamente mayor que a_j al $\alpha\,\%$ de significación.

- Bajo el enfoque convencional elegimos un nivel de significación para el cual obtenemos un valor crítico con el que comparamos nuestro estadístico. Así, aceptamos o rechazamos la H_0 al nivel de significación elegido a priori.
- Enfoque alternativo: "¿cuál es el menor nivel de significación al cual podríamos rechazar la H_0?".
- En este caso primero obtenemos el estadístico y luego buscamos cuál es su percentil correspondiente en la distribución t, i.e.

$$p - valor = Pr\left(\left|t_{n-k-1}\right| > \left|t_{\widehat{\beta}_j}\right|\right).$$

- El p-valor es la probabilidad de observar un esadístico (t) tan extremo como el que obtendríamos si la H_0 fuese cierta.
- Ejemplo: si tenemos $n-k-1=40$ y $t=1{,}85$,

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

- **Regla de rechazo**: cuando obtenemos el p-valor podemos efectuar contrastes clásicos a cualquier nivel de significación
  - Rechazamos H_0 al nivel α de significación cuando α es mayor que el p-valor.
  - No rechazamos H_0 al nivel α de significación cuando el p-valor es mayor que α.

- A la hora de obtener un estimador de los parámetros desconocidos podemos encontrarnos con una serie de dificultades:
  - es improbable que el valor del estimador coincida con el valor del parámetro poblacional.
  - es probable que este valor varíe si se utiliza otra muestra aleatoria.
- Para resolver esta situación, un método alternativo para el contraste de hipótesis se basa en los **Intervalos de Confianza** (rango de posibles valores para el parámetro poblacional).
- Un Intervalo de Confianza del $(1-\alpha)\,\%$ se define como

$$\widehat{\beta}_j \pm c \times se(\widehat{\beta}_j),$$

donde c es el percentil $(1-\alpha/2)$ en una distribución t_{n-k-1}.

- Interpretación Intervalo de Confianza (IC): Si se obtienenen sucesivas muestras aleatorias, y se computa el IC para cada una de ellas, el valor poblacional β_j estará contenido en el IC para un 95 % de las muestras.
- Con un IC es inmediato realizar contrastes de hipótesis bilaterales.
- Regla de rechazo: si $H_0 : \beta_j = a_j$, 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 IC al 95 %.

Contraste de Hipótesis de Combinación Lineal

- En la práctica hay ocasiones en las que tenemos que contrastar hipótesis sobre más de un parámetro poblacional.
- Ejemplo: Utilizando una población de trabajadores con título de bachillerato lo que queremos es determinar si existen diferencias relevantes en los rendimientos educativos por el hecho de haber cursado una carrera universitaria de duración corta o larga.

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

donde

- log(wage)= salarios (logaritmo)
- jc = número de años cursados en una carrera universitaria corta
- univ = número de años cursados en una carrera universitaria larga
- exper = meses transcurridos en la fuerza laboral.

Contraste de Hipótesis de Combinación Lineal

- **Hipótesis nula**: un año más de estudio en una carrera corta y un año más en la carrera larga nos conducen al mismo aumento porcentual en el salario, ceteris paribus.

$$H_0 : \beta_1 = \beta_2$$

- **Hipótesis alternativa**: un año más en la universidad corta tienen menos efecto sobre el salario que un año en la carrera larga, ceteris paribus.

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

- **Estadístico de contraste**: $t = \frac{\widehat{\beta}_1 - \widehat{\beta}_2}{se(\widehat{\beta}_1 - \widehat{\beta}_2)}$.
- **Regla de rechazo:** rechazamos H_0 cuando $t < -c$, donde c es el valor crítico de la distribución t.

Contraste de Hipótesis de Combinación Lineal

- **Problema**: $se(\widehat{\beta}_1 - \widehat{\beta}_2) = \sqrt{Var(\widehat{\beta}_1 - \widehat{\beta}_2)}$, de modo que

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

$$se(\widehat{\beta}_1 - \widehat{\beta}_2) = \left[\left(se(\widehat{\beta}_1)\right)^2 + \left(se(\widehat{\beta}_2)\right)^2 - 2s_{12}\right]^{1/2},$$

donde $s_{12} = Cov(\widehat{\beta}_1, \widehat{\beta}_2)$.

- **Solución**: evitar tener que calcular $Cov(\widehat{\beta}_1, \widehat{\beta}_2)$ y $se(\widehat{\beta}_1 - \widehat{\beta}_2)$. Para ello, reparametrizamos el modelo lineal ($\theta_1 = \beta_1 - \beta_2$) y realizamos el contraste sobre

$$H_0 : \theta_1 = 0, \qquad H_1 : \theta_1 < 0.$$

- **Estadístico t**: $t = \dfrac{\widehat{\theta}_1}{se(\widehat{\theta}_1)}$.

Contraste de múltiples restricciones lineales

- ¿Es posible determinar si un grupo de variables no tiene efecto sobre la dependiente, una vez controlado por otro grupo de variables?
- Si. Contrastando conjuntamente un conjunto de hipótesis múltiples. En concreto, nos interesa determinar si los parámetros de dicho grupo de variables son todos iguales a cero ("restriciones de exclusión").
- **Hipótesis nula**: $H_0 : \beta_1 = 0, \beta_2 = 0, \cdots, \beta_k = 0.$
- **Hipótesis alternativa**: $H_1 : H_0$ no es cierta.

Contraste de múltiples restricciones lineales

- Para resolver este contraste podríamos usar los estadísticos individuales de la t para determinar si cada una de estas variables puede ser eliminada individualmente.
- **Problema**: Un estadístico t no impone restriccione sobre los otros parámetros. Por lo tanto, no queda claro cuántos de ellos han de ser rechazados para realizar un contraste de la H_0.
- **Solución**: estimar un "*modelo restringido*" sin incluir $X_1, X_2, \cdots, X_k$ así como un "*modelo no-restringido*" en el cual incluimos todas las variables explicativas.

- Intuitivamente, lo que queremos es saber si el cambio en la SSR es suficientemente grande para justificar la inclusión de $X_1, X_2, \cdots, X_k$.
- **Estadístico de contraste**:

$$F = \frac{(SSR_R - SSR_{NR})/q}{SSR_{NR}/(n-k-1)},$$

donde R hace referencia al modelo restringido, NR al no-restringido, q es el número de restricciones y $n-k-1$ son los grados de libertad el modelo no-restringido.

- El estadístico F mide el incremento relativo en la SSR cuando pasamos de un modelo no restringido a un modelo restringido. Es siempre es positivo dado que la SSR del modelo restrictivo no debe ser más pequeña que la SSR del modelo no-restringido.

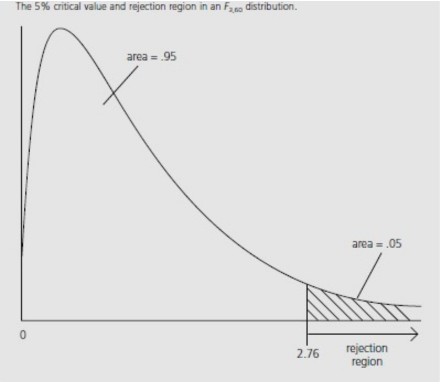
Contraste de múltiples restricciones lineales

- Para decidir si el incremento en la SSR es suficientemente grande y rechazar la H_0 necesitamos conocer la distribución muestral del estadístico F:

$$F \sim F_{q,n-k-1}$$

  donde q son los grados de libertad del numerador de F, mientras que $n-k-1$ refleja los grados de libertad del denominador de F.
- **Regla de rechazo**: rechazamos H_0 cuando $F > c$.
- Si H_0 es rechazado, entonces $X_1, X_2, \cdots, X_k$ con conjuntamente significativas al nivel de significación correspondiente.
- <u>Problema</u>: no sabemos qué variables tienen un efecto parcial sobre Y: puede que sean todas o sólo una de ellas.

Contraste de múltiples restricciones lineales



Fuente: Wooldridge (2005)

- Una de las críticas que tiene este estadístico de F es que depende de las unidades de medida y puede tener valores demasiado altos.
- **Estadístico alternativo**: dado que $SCR = SCT(1 - R^2)$ es cierto para cualquier regresión, podemos reemplazar esta relación tanto en el modelo restringido como en el no-restringido obteniendo

$$F = \frac{(R^2_{NR} - R^2_R)/q}{(1 - R^2_{NR})/(n - k - 1)},$$

donde R hace referencia al modelo restringido y NR al no-restringido.

Contraste de significatividad conjunta

- Un caso especial de restricciones de exclusión es cuando queremos comprobar

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

- Como el R^2 de un modelo de regresión a través de la constante es cero, el estadístico F se simplifica a

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

- Si no podemos rechazar H_0, entonces no hay evidencia de que ninguna de las variables independientes nos ayuden a explicar la dependiente, por lo que deberíamos buscar otras variables.