Notas de Econometría

Pavel Solís

2026

4 Análisis de Regresión Múltiple: Inferencia

Queremos probar hipótesis sobre parámetros del modelo de regresión poblacional

- Una restricción de un parámetro
- Una restricción de varios parámetros
- Varias restricciones
  - Uso común: Evaluar si un grupo de x's se puede omitir del modelo

4.1 Distribuciones muestrales de los estimadores MCO

Con RLM.1-RLM.5 obtuvimos para los estimadores MCO

- $\mathbb{E}\left(\widehat{\beta}_j\right)$: Valor esperado
- $\mathrm{Var}\left(\widehat{\beta}_j\right)$: Varianza (para precisión de los estimadores)

Para hacer inferencia estadística, necesitamos la **distribución muestral** de $\widehat{\beta}_j$

- No definida por supuestos G-M
- Depende de la distribución de los errores (porque condicionamos sobre x's)
- Definimos nuevo supuesto (sobre los errores) para obtenerla

> Supuesto RLM.6. Normalidad del Error
> El error poblacional u es independiente de las variables explicativas $x_1, x_2, \ldots, x_k$ y se distribuye normal con media cero y varianza σ^2:
> $$u \sim \mathrm{N}\left(0, \sigma^2\right)$$

Observaciones sobre el supuesto RLM.6:

- RLM.6 es más fuerte que supuestos anteriores
  - Si u es independiente de x_j

$$\mathbb{E}\left(u \mid x_1, x_2, \ldots, x_k\right) = \mathbb{E}\left(u\right) = 0$$

$$\mathrm{Var}\left(u \mid x_1, x_2, \ldots, x_k\right) = \mathrm{Var}\left(u\right) = \sigma^2$$

- RLM.6 abarca los supuestos RLM.4 y RLM.5
  - Pero escribimos RLM.1-RLM.6 para enfatizar que suponemos más cosas

- En corte transversal, RLM.1-RLM.6 se conocen como los supuestos del **modelo lineal clásico** (MLC)
  - Supuestos MLC = supuestos G-M + supuesto RLM.6
- Estimadores $\widehat{\beta}_j$ (MCO) tienen una propiedad de eficiencia más fuerte con MLC
  - Con G-M, MELI
  - Con MLC, estimadores insesgados de varianza mínima (no solo lineales en y_i)
- Supuestos MLC para la población se pueden resumir como

$$y \mid \vec{x} \sim \mathrm{N}\left(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k,\ \sigma^2\right)$$

  - y condicional en $\vec{x}$, tiene una distribución normal homocedástica
    - Media es lineal en $\vec{x}$
    - Varianza es constante

[Gráfica]

- Justificación para RLM.6 se basa en el teorema del límite central (TLC)
  - u es suma de muchos factores no observados que afectan a y, pero
    - Depende del número factores y qué tan diferentes son sus distribuciones
    - Supone que factores afectan de forma aditiva
- En la práctica, RLM.6 es una cuestión empírica
  - No todas las variables y se distribuyen como una normal
    - Ej. *sueldo* (¿puede ser negativo? ¿hay salario mínimo?)
  - A veces, una transformación de y genera distribución más cercana a normal
    - Ej. $\log(\textit{precio})$ vs *precio*
  - A veces, RLM.6 es claramente falsa
    - Ej. *arrestos* (solo toma valores enteros y generalmente es cero)
- Si RLM.6 no se cumple, no es un problema serio cuando el tamaño de muestra es grande (¿por qué?)
- RLM.6 se traduce en una distribución muestral normal para estimadores MCO

Teorema. Distribuciones Muestrales Normales
Bajo supuestos MLC, RLM.1 a RLM.6, condicional en los valores muestrales de las variables independientes

$$\widehat{\beta}_j \sim \mathrm{N}\left[\beta_j, \mathrm{Var}\left(\widehat{\beta}_j\right)\right]$$

donde $\text{Var}\left(\widehat{\beta}_j\right) = \frac{\sigma^2}{\text{SCT}_j(1-\text{R}_j^2)}$, $j = 1, 2, \ldots, k$, entonces

$$\frac{\widehat{\beta}_j - \beta_j}{\text{desvest}\left(\widehat{\beta}_j\right)} \sim \text{N}\left[0, 1\right]$$

Podemos escribir $\widehat{\beta}_j = \beta_j + \sum_{i=1}^{n} \omega_{ij} u_i$

- $\omega_{ij} = \frac{\widehat{r}_{ij}}{\text{SCR}_j}$ solo depende de las x's (i.e., puede verse como no aleatorio)
  - $\widehat{r}_{ij}$ es el residual i de la regresión de x_j sobre las otras x_j's
  - SCR_j es la SCR de esa regresión
- $\widehat{\beta}_j$ es entonces una combinación lineal de los errores u_i $\forall\, i$, y sabemos que
  - $u_i \overset{\text{iid}}{\sim} \text{N}\left(0, \sigma^2\right)$ por RLM.6
  - Una combinación lineal de normales se distribuye como una normal
- Entonces, $\mathbb{E}\left(\widehat{\beta}_j\right) = \beta_j$ y ya teníamos $Var(\widehat{\beta}_j)$

Estandarización de una variable aleatoria normal da una variable aleatoria normal estándar
Además, para hacer inferencia, ayuda saber que

- Cualquier combinación lineal de $\widehat{\beta}_j$'s también se distribuye normal
- Cualquier subconjunto de $\widehat{\beta}_j$'s tiene una distribución conjunta normal

Normalidad de los estimadores MCO es aproximadamente cierta para muestras grandes

- Aún sin normalidad de los errores

¿Por qué no podemos aplicar el teorema directamente?

4.2 Pruebas de hipótesis sobre un parámetro poblacional: Prueba t

Queremos probar hipótesis para cualquier parámetro de la función de regresión de la población

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

- Suponemos que el modelo poblacional satisface supuestos MLC
- Sabemos que MCO produce estimadores insesgados de β_j
- Queremos hacer pruebas de hipótesis sobre un β_j particular
  - Característica no observada de la población
  - Nunca conoceremos con certeza
- Formulamos hipótesis sobre el valor de β_j y usamos inferencia estadística para probar nuestra hipótesis
- Utilizamos el siguiente resultado para construir pruebas de hipótesis

Teorema. Distribución t para Estimadores Estandarizados
Bajo supuestos MLC, RLM.1 a RLM.6,

$$\frac{\widehat{\beta}_j - \beta_j}{\text{errest}\left(\widehat{\beta}_j\right)} \sim \mathrm{t}_{n-k-1}$$

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

Observaciones sobre el teorema:

- Dos últimos teoremas son diferentes
  - Antes, desvest $\left(\widehat{\beta}_j\right)$ y distribución N
  - Ahora, errest $\left(\widehat{\beta}_j\right)$ y distribución t
- Distribución t
  - Se obtiene porque la constante σ en desvest $\left(\widehat{\beta}_j\right)$ se reemplaza con la variable aleatoria $\widehat{\sigma}$ en errest $\left(\widehat{\beta}_j\right)$
  - Demostración difícil y poco informativa, pero en esencia muestra que dividimos una normal estándar $\mathrm{N}(0,1)$ entre la raíz cuadrada de una χ^2_{n-k-1}
  - Ambas variables aleatorias son independientes por lo que se cumple con la definición de una t_{n-k-1}
- Teorema permite hacer pruebas de hipótesis para β_j

En muchas aplicaciones, la **hipótesis nula** de interés es

$$\mathrm{H}_0 : \beta_j = 0, \quad j = 1, 2, \ldots, k$$

- Significado: Una vez que controlamos por otras x's, x_j no tiene efecto en el valor esperado de y

  Ejemplo. Análisis del salario

  $$\log(salario) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 antig + u$$

  $\mathrm{H}_0 : \beta_2 = 0$ implica que controlando por *educ* y *antig*, *exper* no impacta al *salario*

- Necesitamos estimados, errores estándar, estadístico de prueba y valores críticos

Usamos el **estadístico** t o razón t para evaluar $\mathrm{H}_0 : \beta_j = 0$ y se define como

$$\mathrm{t}_{\widehat{\beta}_j} = \frac{\widehat{\beta}_j}{\text{errest}\left(\widehat{\beta}_j\right)}$$

- Válido contra cualquier **hipótesis alternativa**

- Hay una forma más general del estadístico t (más adelante)
- En aplicaciones, se puede usar el nombre de la variable independiente
  - Por ejemplo, t_{educ} para $\widehat{\beta}_{educ}$
- Reportado por Stata
- Propiedades:
  - $\mathrm{t}_{\widehat{\beta}_j}$ tiene el mismo signo que $\widehat{\beta}_j$ porque errest $\left(\widehat{\beta}_j\right) > 0$
  - Fijando errest $\left(\widehat{\beta}_j\right)$, mayores valores de $\widehat{\beta}_j$ implican mayores valores de $\mathrm{t}_{\widehat{\beta}_j}$

En esencia, contesta la pregunta: ¿qué tan lejos está $\widehat{\beta}_j$ de cero?

- $|\widehat{\beta}_j| >> 0$ puede ser evidencia en contra de $\mathrm{H}_0 : \beta_j = 0$
- Pero hay error de muestreo en estimado $\widehat{\beta}_j$
  - Por eso comparamos el tamaño de $\widehat{\beta}_j$ contra su error muestral
  - errest $\left(\widehat{\beta}_j\right)$ es el estimado de desvest $\left(\widehat{\beta}_j\right)$
- Entonces, $\mathrm{t}_{\widehat{\beta}_j}$ mide cuántas desviaciones estándar $\widehat{\beta}_j$ se aleja de cero
  - Si $|\mathrm{t}_{\widehat{\beta}_j}| >> 0$, rechazamos H_0

Regla de rechazo depende de

- Hipótesis alternativa H_a (un lado o dos lados)
- Nivel de significancia de la prueba α

$$\alpha = \mathbb{P}\left(\text{Rechazar } \mathrm{H}_0 \mid \mathrm{H}_0\right)$$

  - Valores comunes para α: $0.1, 0.05, 0.01$
  - Requiere conocer la distribución muestral de $\mathrm{t}_{\widehat{\beta}_j}$
  - Por teorema, $\mathrm{t}_{\widehat{\beta}_j} \sim \mathrm{t}_{n-k-1}$ cuando H_0 es cierta

Notación:

> Pregunta. ¿Cuál de las opciones siguientes es correcta?
> - $\mathrm{H}_0 : \beta_j = 0$
> - $\mathrm{H}_0 : \widehat{\beta}_j = 0$

- Probamos hipótesis sobre parámetros poblacionales, no sobre estimados

4.2.1 Hipótesis alternativas de un lado

La hipótesis alternativa de un lado tiene la forma $\mathrm{H_a} : \beta_j > 0$

- Para escoger una regla de rechazo, decidimos el nivel de significancia α o $\mathbb{P}\left(\text{Rechazar } \mathrm{H_0} \mid \mathrm{H_0}\right)$
  - Si $\alpha = 0.05$, estamos dispuestos a rechazar $\mathrm{H_0}$ cuando es cierta 5% de las veces
- Bajo $\mathrm{H_0}$, $\mathrm{t}_{\widehat{\beta}_j} \sim \mathrm{t}_{n-k-1}$ y $\mathbb{E}\left(\mathrm{t}_{\widehat{\beta}_j}\right) = 0$ (por teorema) pero bajo $\mathrm{H_a}$, $\mathbb{E}\left(\mathrm{t}_{\widehat{\beta}_j}\right) > 0$
  - Buscamos un valor positivo 'suficientemente grande' de $\mathrm{t}_{\widehat{\beta}_j}$ para rechazar $\mathrm{H_0} : \beta_j = 0$ en favor de $\mathrm{H_a} : \beta_j > 0$
  - Valores negativos de $\mathrm{t}_{\widehat{\beta}_j}$ no dan evidencia a favor de $\mathrm{H_a}$
- 'Suficientemente grande' es el **valor crítico** c dado el nivel de significancia α: c_α
  - Si $\alpha = 0.05$, c es el percentil 95 de una distribución t_{n-k-1}
  - Para obtener c, necesitamos conocer α y los grados de libertad
  - Los valores de c se obtienen de
    - Tablas estadísticas (reportan áreas bajo las distribuciones)
    - Stata: display invttail(gl, α)
- Regla de rechazo (prueba de una cola): Se rechaza $\mathrm{H_0}$ en favor de $\mathrm{H_a}$ con un nivel de significancia $100 \cdot \alpha$ si

$$\mathrm{t}_{\widehat{\beta}_j} > c$$

> Ejemplo. Si $\alpha = 0.05$ y $n - k - 1 = 28$, entonces $c_{0.05} = 1.701$. Si $\mathrm{t}_{\widehat{\beta}_j} \leq 1.701$, no rechazamos $\mathrm{H_0}$ al 5%
>
> [Gráfica]

  - Patrón: Si α baja, c sube
    - Si $\alpha = 0.1$ y $n - k - 1 = 21$, $c_{0.1} = 1.323$
    - Si $\alpha = 0.01$ y $n - k - 1 = 21$, $c_{0.05} = 2.518$
  - Si α baja, se requieren valores de $\mathrm{t}_{\widehat{\beta}_j}$ más grandes para rechazar $\mathrm{H_0}$
    - Si $\mathrm{H_0}$ se rechaza al 5%, automáticamente se rechaza al 10%
  - Si grados de libertad aumentan, distribución t converge a una normal estándar
    - Si $n - k - 1 = 120$ y $\alpha = 0.05$,
      $c_{0.05} = 1.658$ de una distribución t [display invttail(120,0.05)]
      $c_{0.05} = 1.645$ de una distribución $\mathrm{N}(0, 1)$ [display invnormal(0.95)]
    - Cuando $n - k - 1 > 120$, ambas distribuciones son similares

> Ejemplo. Análisis del salario (wage1.dta)

$$\log(\widehat{salario}) = \underset{(0.104)}{0.284} + \underset{(0.007)}{0.092}\, educ + \underset{(0.0017)}{0.0041}\, exper + \underset{(0.003)}{0.022}\, antig$$

$$n = 526,\ \mathrm{R}^2 = 0.316$$

- Errores estándar en paréntesis
- Interesados en probar $\mathrm{H}_0 : \beta_{exper} = 0$ vs $\mathrm{H}_\mathrm{a} : \beta_{exper} > 0$
- ¿Grados de libertad?
- Valores críticos de $\phi(z)$: $c_{0.1} = 1.282$, $c_{0.05} = 1.645$, $c_{0.01} = 2.326$
- Estadístico para $\widehat{\beta}_{exper}$:

$$\mathrm{t}_{exper} = \frac{0.0041}{0.0017} \approx 2.41$$

- Rechazamos H_0

  exper es estadísticamente significativa (mayor a cero) a un nivel del 1%
- Rendimiento esperado de 3 años más de experiencia (fijando antigüedad y educación) es 1.2% [3(0.0041) = 0.0123] (bajo, pero positivo)

Cuando $\mathrm{H}_\mathrm{a} : \beta_j < 0$, la regla de rechazo (prueba de una cola) es

$$\mathrm{t}_{\widehat{\beta}_j} < -c$$

- $c > 0$ en tablas (distribución t es simétrica, c viene de $\mathrm{H}_\mathrm{a} : \beta_j > 0$)

Ejemplo. Si $\alpha = 0.05$ y $g.l. = 18$, $c = 1.734$ y H_0 se rechaza al 5% si $\mathrm{t}_{\widehat{\beta}_j} < -1.734$

[Gráfica]

- Dado que $\mathrm{H}_\mathrm{a} : \beta_j < 0$, $\mathrm{t}_{\widehat{\beta}_j} < 0$ para rechazar H_0
  - Si $\mathrm{t}_{\widehat{\beta}_j} > 0$, no da evidencia en favor de H_a

Ejemplo. Análisis del desempeño de estudiantes en un examen
Utilizamos *compns* para calidad del profesor, *prsnl* para atención a estudiantes e *inscr* para tamaño de la escuela

$$\widehat{examen} = \underset{(6.113)}{2.274} + \underset{(0.00010)}{0.00046}\, compns + \underset{(0.04)}{0.048}\, prsnl - \underset{(0.00022)}{0.00020}\, inscr$$

$$n = 408,\ \mathrm{R}^2 = 0.0541$$

- Errores estándar en paréntesis
- Interesados en probar si escuelas grandes tienen bajo desempeño

* $H_0 : \beta_{inscr} = 0$ vs $H_a : \beta_{inscr} < 0$
- Observamos $\widehat{\beta}_{inscr} < 0$, ¿es por error muestral?
- $n - k - 1 = 408 - 4 = 404$
- Valor crítico de $\phi(z)$ (¿por qué?) al 5%: $-c_{0.05} = -1.65$
- Estadístico para $\widehat{\beta}_{inscr}$:

$$\mathrm{t}_{inscr} = \frac{-0.00020}{0.00022} \approx -0.91$$

- No rechazamos H_0

  inscr no es estadísticamente significativa a un nivel del 5%
- $\mathrm{t}_{comp} = 4.6 > 2.33 = c_{0.01}$, se rechaza $H_0 : \beta_{comp} = 0$ a un nivel del 1%
- $\mathrm{t}_{prsnl} = 1.2 < 1.28 = c_{0.1}$, no se rechaza $H_0 : \beta_{prsnl} = 0$ incluso a un nivel del 10%

4.2.2 Hipótesis alternativas de dos lados

En la práctica, es común probar $H_0 : \beta_j = 0$ contra $H_a : \beta_j \neq 0$

- Para ver si x_j tiene un efecto ceteris paribus en y
  - Sin especificar si el efecto es positivo o negativo
  - Previene que escojamos H_a después de ver los resultados de la regresión

Regla de rechazo se basa en el valor absoluto del estadístico t:

$$|\mathrm{t}_{\widehat{\beta}_j}| > c$$

- El valor crítico c se determina al escoger α de forma que el área en cada cola de la distribución sea $\alpha/2$
  - Si $\alpha = 0.05$, c es el percentil 97.5 de una distribución t_{n-k-1}

Ejemplo. Si $\alpha = 0.05$ y $n - k - 1 = 25$, entonces $c_{0.05} = 2.060$
En Stata: display invttail(25,0.025)

[Gráfica]

- Si H_0 se rechaza en favor de H_a, decimos que "x_j es estadísticamente significativa o diferente de cero a un nivel de $(100 - \alpha)$%"

Ejemplo. Calificaciones universitarias en una escala de 0 a 4

$$\begin{array}{lllllllll} \widehat{califuni} = & 1.39 & + & 0.412\, califprep & + & 0.015\, examen & - & 0.083\, faltas \\ & (0.33) & & (0.094) & & (0.011) & & (0.026) \end{array}$$

$$n = 141,\ \mathrm{R}^2 = 0.234$$

– Errores estándar en paréntesis

– Grados de libertad: $141 - 4 = 137$

– Valores críticos de $\phi(z)$: $c_{\frac{0.1}{2}} = 1.65$, $c_{\frac{0.05}{2}} = 1.96$, $c_{\frac{0.01}{2}} = 2.58$

  * En Stata: invnormal(0.95), invnormal(0.975), invnormal(0.995)

– $\mathrm{t}_{califprep} = 4.38 > 2.58$, estadísticamente significativa al 1%

– $\mathrm{t}_{examen} = 1.36$, no es estadísticamente significativa al 10%

– $\mathrm{t}_{faltas} = -3.19 < -2.58$, estadísticamente significativa al 1%

  * Diferencia promedio predicha en $califuni$ entre estudiantes con misma $califprep$ y $examen$ si uno de ellos falta toda la semana es 0.42 (0.083×5 porque $\Delta faltas = 5$)

4.2.3 Pruebas de hipótesis para otros valores de β_j

$\mathrm{H}_0 : \beta_j = 0$ es la hipótesis nula más común pero podemos hacer pruebas para cualquier valor hipotético a_j

- $\mathrm{H}_0 : \beta_j = a_j$
- $\beta_j = 1$ o $\beta_j = -1$ son comunes

El estadístico t apropiado es

$$\mathrm{t}_j = \frac{\widehat{\beta}_j - a_j}{\mathrm{errest}\left(\widehat{\beta}_j\right)}$$

- En general, el estadístico se puede expresar como

$$t = \frac{\text{estimado} - \text{valor hipotético}}{\text{error estándar}}$$

- Mide cuántas desviaciones estándar está $\widehat{\beta}_j$ del valor hipotético a_j

Bajo $\mathrm{H}_0 : \beta_j = a_j$, el estadístico $\mathrm{t}_j \sim \mathrm{t}_{n-k-1}$ por el teorema

- El estadístico t usual se obtiene cuando $a_j = 0$

El estadístico t general se puede usar para hacer una PH contra $\mathrm{H_a}$ de uno y dos lados

- El valor crítico c lo encontramos como antes, la diferencia está en cómo se calcula el estadístico t

Si $\mathrm{H}_0 : \beta_j = 1$ y $\mathrm{H_a} : \beta_j > 1$, rechazamos H_0 si $\mathrm{t} > c$ y decimos que "$\widehat{\beta}_j$ es estadísticamente mayor a 1" a un nivel de $100 \cdot \alpha\%$

$$\mathrm{t} = \frac{\widehat{\beta}_j - 1}{\mathrm{errest}\left(\widehat{\beta}_j\right)}$$

Preguntas. Análisis del crimen en universidades con relación al tamaño del campus

- Modelo poblacional:

$$\log(crimen) = \beta_0 + \beta_1 \log(inscr) + u$$

- Modelo de elasticidad constante
  - No captura efecto ceteris paribus (¿por qué?)
  - β_1 es la elasticidad del *crimen* con respecto a *inscr*
  - Equivale a $crimen = \exp(\beta_0) inscr^{\beta_1} \exp(u)$, ¿qué pasa si $\beta_0 = u = 0$?

[Gráfica]

- $\mathrm{H}_0 : \beta_1 = 0$ es obvio; más interesante $\mathrm{H}_0 : \beta_1 = 1$ vs $\mathrm{H}_0 : \beta_1 > 1$
  - Si se rechaza H_0, el crimen es un problema grave en campus grandes

$$\begin{array}{rccc} \widehat{\log(crimen)} = & -6.63 & + & 1.27 \log(inscr) \\ & (1.03) & & (0.11) \end{array}$$

$$n = 97,\ \mathrm{R}^2 = 0.585$$

- Observamos que $\widehat{\beta}_1 > 1$, pero ¿es suficiente para concluir que $\beta_1 > 1$?
- ¿Estadístico t?
  - En este caso, el t que reporta Stata no nos sirve (¿por qué?)
- ¿Grados de libertad?
- Valores críticos: $c_{0.05} = 1.66, c_{0.01} = 2.37$
- ¿Decisión? ¿Se rechaza o no se rechaza H_0?

Si $\mathrm{H}_0 : \beta_j = -1$ y $\mathrm{H_a} : \beta_j \neq -1$, rechazamos H_0 si $|\mathrm{t}| > c_{\alpha/2}$ y decimos que "$\widehat{\beta}_j$ es estadísticamente diferente de -1" al nivel apropiado

$$\mathrm{t} = \frac{\widehat{\beta}_j - (-1)}{\mathrm{errest}\left(\widehat{\beta}_j\right)}$$

Preguntas. Análisis de precios de casas y características de la comunidad

- Modelo poblacional:

$$\log(precio) = \beta_0 + \beta_1 \log(oxn) + \beta_2 \log(dist) + \beta_3 habit + \beta_4 alumprof + u$$

- Modelo estimado:

$$\widehat{\log(prc)} = \begin{array}{lllll} 11.08 & -0.954\log(oxn) & -0.134\log(dst) & +0.255h & -0.052apr \\ (0.32) & (0.117) & (0.043) & (0.019) & (0.006) \end{array}$$

$$n = 506,\ \mathrm{R}^2 = 0.581$$

- Todos los coeficientes son estadísticamente diferentes de cero con bajos niveles de significancia pero $\mathrm{H}_0 : \beta_j = -1$ (no $\mathrm{H}_0 : \beta_j = 0$)
- ¿Cuál es el valor del estadístico t?
- Conclusión: Poca evidencia de que la elasticidad sea diferente de -1

4.2.4 Cálculo de valores-p para pruebas t

Enfoque clásico: Definir $\mathrm{H_a}$, escoger α que determina c, comparar t contra c, conclusión (rechazar H_0 o no rechazar H_0)

- Hay arbitrariedad al escoger α (cada quien puede escoger diferente α)
- Ej. H_0 puede no rechazarse al 5% pero si al 10%
- En lugar de hacer la prueba para diferentes niveles de α, usamos el valor $-$ p

El valor $-$ p de la prueba

- Responde la siguiente pregunta: Dado el valor observado del estadístico t, ¿cuál es el valor más pequeño de α para el que se rechazaría H_0?
- Es el valor de α si usamos el valor de t como el valor de c

Ejemplo. $\mathrm{H}_0 : \beta_j = 0$, $\mathrm{H_a} : \beta_j \neq 0$, $g.l = 40$, $\mathrm{t} = 1.85$

- $c_{0.05} = 2.021$, no se rechaza H_0
- $c_{0.1} = 1.684$, se rechaza H_0
- $0.05 <$ valor $-$ p < 0.1
- valor $-$ p $= \mathbb{P}\left(|\mathrm{t}_{40}| > 1.85\right) = 2\cdot \mathbb{P}\left(\mathrm{t}_{40} > 1.85\right) = 2(0.0359) = 0.0718$
- En Stata: display ttail(40, 1.85)*2

[Gráfica]

Siempre se reporta en decimal porque es una probabilidad: $0 <$ valor $-$ p < 1

- Ej. Si $\mathrm{H}_0 : \beta_j = 0$ y $\mathrm{H_a} : \beta_j \neq 0$,

$$\text{valor} - \text{p} = \mathbb{P}\left(|T| > |\mathrm{t}|\right)$$

  donde $T = \mathrm{t}_{n-k-1}$ y t es el valor del estadístico de prueba

Interpretación del valor $-$ p:

- Probabilidad de observar un estadístico t tan extremo al observado si H_0 fuera cierta
  - Un $\text{valor} - \text{p}$ pequeño es evidencia contra H_0
  - Un $\text{valor} - \text{p}$ grande da poca evidencia contra H_0
- Ej. $\text{valor} - \text{p} = 0.5$ significa que observaremos valores de t tan extremos al obtenido 50% de las veces si H_0 fuera cierta
- Sintetiza la fuerza o debilidad de la evidencia empírica contra H_0

Una vez calculado el $\text{valor} - \text{p}$, se puede hacer una prueba clásica para cualquier α

- Ej. H_0 se rechaza si $\text{valor} - \text{p} < \alpha$, sino H_0 no se rechaza al nivel $(100 \cdot \alpha)\%$

Para una H_a de un lado

- Si $\mathrm{H}_0 : \beta_j = 0$
  - Para $\mathrm{H}_\mathrm{a} : \beta_j > 0$, solo calcular si $\widehat{\beta}_j > 0$ $(\mathrm{t} > 0)$ y $\text{valor} - \text{p} = \mathbb{P}(T > \mathrm{t})$
  - Para $\mathrm{H}_\mathrm{a} : \beta_j < 0$, solo calcular si $\widehat{\beta}_j < 0$ $(\mathrm{t} > 0)$ y $\text{valor} - \text{p} = \mathbb{P}(T < \mathrm{t}) = \mathbb{P}(T > |\mathrm{t}|)$
- $\text{valor} - \text{p de un lado} = \frac{\text{valor}-\text{p de dos lados}}{2}$

No es crucial reportar el $\text{valor} - \text{p}$ para el estadístico t pero sí para el estadístico F porque los valores críticos varían

4.2.5 Lenguaje de pruebas de hipótesis clásicas

Si, H_0 no se rechaza, ¿cuál de los siguientes enunciados es correcto?

- H_0 no se rechaza al nivel $(100 - \alpha)\%$
- H_0 se acepta al nivel $(100 - \alpha)\%$

Ejemplo.

$$\mathrm{H}_0 : \beta_j = 1,\ \mathrm{t} = 0.393$$
$$\mathrm{H}_0 : \beta_j = 0.99,\ \mathrm{t} = 0.308$$

- Ambas no pueden ser ciertas
- Decimos: Los datos no permiten rechazarlos a un nivel $(100 - \alpha)\%$

4.2.6 Significancia económica (práctica) vs significancia estadística

Al correr una regresión, debemos observar:

- Significancia estadística: Tamaño del estadístico $\mathrm{t}_{\widehat{\beta}_j}$
- Significancia económica: Magnitud y signo del coeficiente estimado $\widehat{\beta}_j$

Ej. Cuando probamos $\mathrm{H}_0 : \beta_j = 0$, $\mathrm{t}_{\widehat{\beta}_j} = \frac{\widehat{\beta}_j}{\mathrm{errest}\left(\widehat{\beta}_j\right)}$ puede ser significativo porque

- $\widehat{\beta}_j$ es grande, o
- errest $\left(\widehat{\beta}_j\right)$ es pequeño

Solo enfocarnos en la significancia estadística puede llevarnos a la conclusión falsa de que una variable independiente es 'importante' para explicar y aunque su efecto estimado sea pequeño

Ejemplo. Tasa de participación en planes de retiro

$$\begin{array}{lllllllll} \widehat{tpartic} = & 80.29 & + & 5.44\,tcomplem & + & 0.269\,ansplan & - & 0.00013\,emplead \\ & (0.78) & & (0.52) & & (0.045) & & (0.00004) \end{array}$$

$$n = 1,534,\ \mathrm{R}^2 = 0.1$$

- $\mathrm{t}_{emplead} = \frac{-0.00013}{0.00004} = -3.25$
- Estadísticamente significativo (valor $-$ p de dos lados = 0.001)
- ¿Significancia económica?
  - Fijando *tcomplem* y *ansplan*, si la empresa crece en 10 mil empleados,
  $$\widehat{tpartic} = 10,000(-0.00013) = -1.3\%$$
  - Gran crecimiento en empleados con efecto moderado en la participación
    - Sí afecta, pero impacto es pequeño en términos prácticos

Cuando el tamaño de muestra crece ($n \to \infty$), los errores estándar disminuyen

- Parámetros estimados con precisión
- Efectos estadísticamente significativos (porque $n \to \infty$)
- Hay que interpretar la magnitud

Relación entre el tamaño de muestra (n) y el nivel de significancia (α)

- Si $n \to \infty$, usar una α baja porque los errores estándar disminuyen
- Si $n \to 0$, usar una α alta porque es difícil encontrar significancia

Ejemplo. Tasa de productos defectuosos (por cada 100)

$$\begin{array}{lllllllll} \widehat{\log(tdef)} = & 12.46 & - & 0.029\,hrcpct & - & 0.962\,\log(vents) & + & 0.761\,\log(emple) \\ & (5.69) & & (0.023) & & (0.453) & & (0.407) \end{array}$$

$$n = 29,\ \mathrm{R}^2 = 0.262$$

- La variable de interés son las horas de capacitación ($hrcpct$)
  - $\mathrm{H}_0 : \beta_{hrcpct} = 0$ y $\mathrm{H}_\mathrm{a} : \beta_{hrcpct} < 0$
  - $\beta_{hrcpct} = -0.029$, ¿cuánto baja la tasa con 5 horas más de capacitación?
  - Efecto no trivial
- $\mathrm{t}_{hrcpct} = -0.029/0.023 = -1.26, g.l. = 29 - 4 = 25$ (n pequeña, α alta)
  - $c_{0.05} = -1.71$, $hrcpct$ no es estadísticamente significativa al 5%
  - $c_{0.1} = -1.32$, $hrcpct$ casi estadísticamente significativa al 10%
- $\mathrm{valor} - \mathrm{p} = \mathbb{P}\left(T_{25} < -1.26\right) = 0.11$
  - Bajo para deberse a error muestral pero sujeto a interpretación

Los errores estándar pueden ser altos por multicolinealidad aún cuando $n \to \infty$

- Difícil estimar con precisión los efectos parciales cuando algunas variables independientes están altamente correlacionadas
  - Como cuando el tamaño de muestra es pequeño
- Soluciones
  - Recolectar más datos
  - Cambiar enfoque del análisis (quitar o cambiar variables independientes)

Guía para discutir la significancia en RLM

1. Si la variable es estadísticamente significativa a niveles usuales (10, 5, 1%), discutir la magnitud del coeficiente para entender la importancia económica
   - Cuidado: unidades de medición, forma de la variable (ej. logaritmo)
2. Si la variable no es estadísticamente significativa,
   - ¿Efecto esperado (signo)?
   - ¿Efecto grande (magnitud)? Si efecto es grande, calcular $\mathrm{valor} - \mathrm{p}$ para el estadístico t.
     - Si n pequeña, $\mathrm{valor} - \mathrm{p} < 0.2$
     - Si $\mathrm{valor} - \mathrm{p} > 0.2$, coeficiente grande por error muestral
3. Si estadístico t es pequeño y el signo es 'incorrecto', ignorar

Si la variable es estadísticamente significativa con signo "incorrecto" y efecto grande, difícil de resolver

- Pensar en el modelo o en los datos (ej. puede ser por omitir una variable relevante)

4.3 Intervalos de confianza

Podemos construir IC para un parámetro poblacional β_j bajo los supuestos del MLC

- Dan un *rango* de valores para β_j
- Diferentes a estimados puntuales
- También conocidos como estimados por intervalos

Sabemos que $\frac{\widehat{\beta}_j - \beta_j}{\text{errest}\left(\widehat{\beta}_j\right)} \sim \mathrm{t}_{n-k-1}$, entonces un **intervalo de confianza** al 95% está dado por

$$\widehat{\beta}_j \pm c_{\alpha/2} \cdot \text{errest}\left(\widehat{\beta}_j\right)$$

- $c_{\alpha/2}$ es el percentil 97.5 de una distribución t_{n-k-1}
- Cota inferior del IC: $\underline{\beta_j} = \widehat{\beta}_j - c_{\alpha/2} \cdot \text{errest}\left(\widehat{\beta}_j\right)$
- Cota superior del IC: $\overline{\beta_j} = \widehat{\beta}_j + c_{\alpha/2} \cdot \text{errest}\left(\widehat{\beta}_j\right)$

Significado: Si calculamos $\underline{\beta_j}$ y $\overline{\beta_j}$ para muchas muestras, el parámetro poblacional (no conocido) β_j estaría en el intervalo $\left(\underline{\beta_j}, \overline{\beta_j}\right)$ para 95% de las muestras

$$\mathbb{P}\left(\underline{\beta_j} < \beta_j < \overline{\beta_j}\right) = 1 - \alpha$$

- Para la muestra particular que usamos para construir el IC, no sabemos si contendrá a β_j
- Esperamos tener una de las muestras para las que el IC contiene a β_j, pero no hay garantía

Elementos para construir un IC:

- $\widehat{\beta}_j$
- errest $\left(\widehat{\beta}_j\right)$
- c de la distribución t_{n-k-1}, para lo que necesitamos saber
  - Grados de libertad $(n - k - 1)$
  - Nivel confianza (95%)

> Ejemplo. Si los grados de libertad son $n - k - 1 = 25$ y $\alpha = 0.05$, $c = 2.06$, entonces el IC es
>
> $$\left[\widehat{\beta}_j - 2.06 \cdot \text{errest}\left(\widehat{\beta}_j\right), \widehat{\beta}_j + 2.06 \cdot \text{errest}\left(\widehat{\beta}_j\right)\right]$$

Grados de libertad $(n - k - 1)$ y valor crítico (c) para un IC al 95%:

- Si $n-k-1>120$, $\mathrm{t}_{n-k-1} \approx \mathrm{N}$, entonces usamos el percentil 97.5 de Φ

$$\widehat{\beta}_j \pm 1.96 \cdot \text{errest}\left(\widehat{\beta}_j\right)$$

- Si $n-k-1>50$, $c \approx 2$, entonces la regla de aproximación es

$$\widehat{\beta}_j \pm 2 \cdot \text{errest}\left(\widehat{\beta}_j\right)$$

- Si $n-k-1<50$, usar tablas o Stata

Si cambia α de 0.05 a 0.1 o 0.01

- Para un IC al 90%, c es el percentil 95 de t_{n-k-1} o Φ
  - Ej. g.l.$= n-k-1=25$, $c=1.71$, IC al 90% es (más angosto)

$$\widehat{\beta}_j \pm 1.71 \cdot \text{errest}\left(\widehat{\beta}_j\right)$$

- Para un IC al 99%, c es el percentil 99.5 de t_{n-k-1} o Φ
  - Ej. g.l.$= n-k-1=25$, $c=2.79$, IC al 99% es (más amplio)

$$\widehat{\beta}_j \pm 2.79 \cdot \text{errest}\left(\widehat{\beta}_j\right)$$

Con IC, podemos hacer PH de 2 lados:

- Si $\mathrm{H}_0 : \beta_j = a_j$ y $\mathrm{H}_\mathrm{a} : \beta_j \neq a_j$, rechazamos H_0 al 5% si $a_j \notin$ IC del 95%

Ejemplo. Modelo de elasticidad constante para estudiar el efecto ceteris paribus del márgen de ganancias (ganancia/ventas) en el gasto I&D

$$\begin{array}{lllll} \log(\widehat{invdes}) = & -4.38 & + \; 1.084 \log(ventas) & + & 0.0217\, margengan \\ & (0.47) & \quad (0.060) & & (0.0128) \end{array}$$

$$n=32, \; \mathrm{R}^2 = 0.918$$

- Elasticidad de I&D respecto a ventas:
  - 1% más de ventas asociado con 1.084% más en gasto I&D
- Grados de libertad: $32-2-1=29$
  - Usamos percentil 97.5 de la distribución t_{29}: $c=2.045$
- IC al 95% para $\beta_{\log(ventas)}$: $1.084 \pm 2.045 \times 0.06$ o $(0.961, 1.21)$
  - $0 \notin$ IC porque se espera que gasto I&D crezca con el tamaño de la empresa
  - $1 \in$ IC, no podemos rechazar $\mathrm{H}_0 : \beta_{\log(ventas)} = 1$ vs $\mathrm{H}_\mathrm{a} : \beta_{\log(ventas)} \neq 1$ al 5%
    - La elasticidad estimada no es estadísticamente diferente de 1 al 5%

- IC al 95% para $\beta_{margengan}$: $0.0217 \pm 2.045 \times 0.0128$ o $(-0.0045, 0.0479)$
  - No rechazamos $\text{H}_0 : \beta_{margengan} = 0$ vs $\text{H}_\text{a} : \beta_{margengan} \neq 0$ al 5%

Los intervalos de confianza dependen de los supuestos hechos para construirlos

- Si la variable omitida se correlaciona con otras variables independientes, el IC no es confiable porque estimados MCO sesgados
- Si hay heterocedasticidad, errest $\left(\widehat{\beta}_j\right)$ no es válido para estimar desvest $\left(\widehat{\beta}_j\right)$
  - IC no será al 95%
- Si normalidad del error no se cumple, el IC no es válido
  - Aunque esto se puede relajar cuando $n \to \infty$

4.4 PH para una combinacion lineal de parámetros

Hasta ahora PH o IC para una sola β_j, aquí cómo probar una sola hipótesis con varias β_j's
Para ilustrar, modelo poblacional para gente trabajadora con título de preparatoria,:

$$\log(salar) = \beta_0 + \beta_1 carrtec + \beta_2 licen + \beta_3 exper + u$$

- Hipótesis de interés: ¿Un año de carrera técnica vale lo mismo que un año de licenciatura?
  - $\text{H}_0 : \beta_1 = \beta_2$, ambas escuelas generan mismo aumento porcentual en salario (ceteris paribus)
  - $\text{H}_\text{a} : \beta_1 < \beta_2$, un año de carrera técnica vale menos que un año de licenciatura
- No podemos usar estadísticos t individuales para $\widehat{\beta}_1$ y $\widehat{\beta}_2$
- Reexpresamos hipótesis como
  - $\text{H}_0 : \beta_1 - \beta_2 = 0$
  - $\text{H}_\text{a} : \beta_1 - \beta_2 < 0$
- El estadístico quedaría

$$\frac{\widehat{\beta}_1 - \widehat{\beta}_2}{\text{errest}\left(\widehat{\beta}_1 - \widehat{\beta}_2\right)}$$

  y mismo proceso que antes (α, g.l., c, regla de rechazo $\text{t} < -c$ o $\text{valor} - \text{p} < \alpha$)

- ¿errest $\left(\widehat{\beta}_1 - \widehat{\beta}_2\right)$? errest $\left(\widehat{\beta}_1 - \widehat{\beta}_2\right) \neq$ errest $\left(\widehat{\beta}_1\right)$ − errest $\left(\widehat{\beta}_2\right)$
  - Por probabilidad, Var $\left(\widehat{\beta}_1 - \widehat{\beta}_2\right)$ y desvest $\left(\widehat{\beta}_1 - \widehat{\beta}_2\right) = \sqrt{\text{Var}\left(\widehat{\beta}_1 - \widehat{\beta}_2\right)}$

- Entonces, $\text{errest}\left(\widehat{\beta}_1 - \widehat{\beta}_2\right) = \sqrt{\text{errest}^2\left(\widehat{\beta}_1\right) + \text{errest}^2\left(\widehat{\beta}_2\right) - 2s_{1,2}}$

- Alternativas:
  - En Stata, usar el comando test
  - Reescribir el modelo para obtener directamente el error estándar de interés

- Reescribir modelo siempre funciona
  - Si $\theta_1 = \beta_1 - \beta_2$, entonces $\mathrm{H_0} : \theta_1 = 0$ vs $\mathrm{H_a} : \theta_1 < 0$, y $\mathrm{t} = \frac{\widehat{\theta}_1}{\text{errest}\left(\widehat{\theta}_1\right)}$
  - Definimos $\beta_1 = \theta_1 + \beta_2$, entonces

$$\begin{aligned} \log(salario) &= \beta_0 + (\theta_1 + \beta_2)\, carrtec + \beta_2 licen + \beta_3 exper + u \\ &= \beta_0 + \theta_1 carrtec + \beta_2 univ + \beta_3 exper + u \end{aligned}$$

  - Se crea una nueva variable $univ = carrtec + licen$, no se estima β_1 y se obtiene $\text{errest}\left(\widehat{\theta}_1\right)$ directo

Ejemplo. Salario y estudios universitarios

$$\begin{array}{rcccccccc} \widehat{\log(salar)} = & 1.472 & + & 0.0667\, carrtec & + & 0.0769\, licen & + & 0.0049\, exper \\ & (0.021) & & (0.0068) & & (0.0023) & & (0.0002) \end{array}$$

$$n = 6{,}763,\ \ \mathrm{R}^2 = 0.222$$

- ¿$\widehat{\beta}_1$ y $\widehat{\beta}_2$ son estadísticamente y económicamente significativas?
- ¿$\widehat{\beta}_1 - \widehat{\beta}_2$?
- ¿$\text{errest}\left(\widehat{\beta}_1 - \widehat{\beta}_2\right)$?

$$\begin{array}{rcccccccc} \widehat{\log(salar)} = & 1.472 & - & 0.0102\, carrtec & + & 0.0769\, univ & + & 0.0049\, exper \\ & (0.021) & & (0.0069) & & (0.0023) & & (0.0002) \end{array}$$

$$n = 6{,}763,\ \ \mathrm{R}^2 = 0.222$$

- $\mathrm{t}_{\widehat{\theta}_1} = \frac{-0.0102}{0.0069} = -1.48$
- $\text{valor} - \mathrm{p} = 0.07$, algo de evidencia pero no fuerte contra $\mathrm{H_0}$
- IC al 95% con aproximación Φ:

$$-0.0102 \pm 1.96 \cdot 0.0069 \quad \text{o} \quad (-0.0237, 0.033)$$

4.5 Pruebas de varias restricciones lineales: Prueba F

Hasta ahora hipótesis de una sola restricción, aquí varias restricciones

- Antes: Una β_j o una combinación lineal de β_j's
- Uso mas común: Probar si un conjunto de variables independientes no tiene efecto parcial sobre la variable dependiente

4.5.1 Probar restricciones de exclusión (PH múltiple o conjunta)

H_0 : Un conjunto de variables no tiene efecto en y, una vez que controlamos por otras

- Estadísticos t individuales no son apropiados
  - Hipótesis no pone restricciones en los otros parámetros
  - ¿Cuántas t significativas para rechazar?
  - No es claro cómo se rechazaría H_0
- Necesitamos probar las restricciones de exclusión de forma conjunta

Comparamos 2 modelos: con y sin restricciones

- Modelo sin restricciones o no restringido incluye todos los parámetros
- Modelo restringido siempre tiene menos parámetros

SCR es útil para hacer pruebas de hipótesis conjuntas

- SCR siempre aumenta cuando se quitan variables del modelo
- Relativo a SCR del modelo con todas las variables, ¿cuánto aumenta SCR cuando quitamos variables del modelo?

Ejemplo. ¿La productividad en el beisbol tiene efecto en el salario?

Modelo no restringido:

$\log(salar) = \beta_0 + \beta_1 ansliga + \beta_2 promjueg + \beta_3 prombat + \beta_4 hruns + \beta_5 carrbat + u$

- Productividad: Promedio de bateo, home runs, carreras al bat
- H_0 : Productividad en el beisbol no tiene efecto en el salario
  - $\mathrm{H}_0 : \beta_3 = 0,\ \beta_4 = 0,\ \beta_5 = 0$ (una vez que controlamos por los años en la liga y el promedio de juegos)
  - 3 restricciones de exclusión
- $\mathrm{H_a}$: H_0 no es cierta
  - Al menos una de las tres (β_3, β_4 o β_5) es diferente de cero
- Para ilustrar, estimamos el modelo

$$\widehat{\log(salar)} = \underset{(0.27)}{11.19} + \underset{(0.0121)}{0.0687}\,ansliga + \underset{(0.0026)}{0.0126}\,promjueg$$

$$+ \underset{(0.0011)}{0.00098}\,prombat + \underset{(0.0161)}{0.0144}\,hruns + \underset{(0.0072)}{0.0108}\,carrbat$$

$$n = 353, \qquad \mathrm{SCR} = 183.186 \qquad \mathrm{R}^2 = 0.6278$$

- ¿t individuales de variables de productividad son significativas?

Modelo restringido (sin variables en H_0):

$$\log(salar) = \beta_0 + \beta_1 ansliga + \beta_2 promjueg + u$$

$$\widehat{\log(salar)} = \underset{(0.11)}{11.22} + \underset{(0.0125)}{0.0713}\,ansliga + \underset{(0.0013)}{0.0202}\,promjueg$$

$$n = 353, \qquad \mathrm{SCR} = 198.311 \qquad \mathrm{R}^2 = 0.5971$$

- Incremento en SCR de 183.186 a 198.311 ¿es suficientemente grande para rechazar H_0?
  - Necesitamos un estadístico con distribución conocida bajo H_0 para obtener un valor crítico c a partir de α

Caso general (ambos modelos con intercepto):

- Modelo sin restricciones con k variables independientes ($k+1$ parámetros):

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

- Suponemos que hay q restricciones de exclusión (tienen coeficientes cero)
  - Últimas q variables se quitan del modelo
- $\mathrm{H}_0 : \beta_{k-q+1} = 0, \ldots, \beta_k = 0$
- $\mathrm{H}_\mathrm{a} : \mathrm{H}_0$ es falsa
  - Al menos uno de los parámetros $(\beta_{k-q+1}, \ldots, \beta_k)$ es distinto de cero
- Modelo con restricciones (impuestas):

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_{k-q} x_{k-q} + u$$

Estadístico F o razón F:

$$\mathrm{F}_d = \frac{(\mathrm{SCR}_r - \mathrm{SCR}_{sr})/q}{\mathrm{SCR}_{sr}/(n-k-1)}$$

- SCR_r del modelo restringido
- SCR_{sr} del modelo sin restricciones

- ≥ 0 porque $\mathrm{SCR}_r \geq \mathrm{SCR}_{sr}$ (por lo que si $\mathrm{F} < 0$, los SCR están invertidos)
- q son los grados de libertad del numerador: $g.l._r - g.l._{sr} = (n-k-1+q)-(n-k-1)$
- $n-k-1$ son los grados de libertad del denominador: $g.l._{sr}$
- Ambos modelos se estiman con mismo n (cuidado cuando hay faltantes)
- Denominador es el estimador insesgado de $\sigma^2 = \mathrm{Var}(u)$ en modelo sin restricciones
- Compara el aumento relativo en SCR al ir del modelo sin restricciones al restringido

Para calcular el estadístico F, necesitamos: $g.l._{sr}$, q, SCR_{sr} y SCR_r

- Ej. $n = 353$, $k+1 = 6$, $n-k-1 = g.l._{sr} = 353-6 = 347$, $q = 3$
- Para seguir, necesitamos la distribución muestral del estadístico F bajo H_0
  - Conociendo esa distribución, podemos escoger c y definir la regla de rechazo

Teorema. Distribución F
Bajo supuestos MLC, RLM.1 a RLM.6 y H_0,

$$\mathrm{F}_d = \frac{(\mathrm{SCR}_r - \mathrm{SCR}_{sr})/q}{\mathrm{SCR}_{sr}/(n-k-1)} \sim \mathrm{F}_{q,\,n-k-1}$$

- Intuición: F es una división de 2 variables aleatorias χ^2 independientes divididas por sus grados de libertad respectivos
  - Esa es la definición de una variable aleatoria F
- Rechazamos H_0 si F es grande
  - Si $\alpha = 0.05$, c_α es el percentil 95 de la distribución $\mathrm{F}_{q,\,n-k-1}$
    - c depende entonces de α, q y $n-k-1$
  - Valores críticos al 10%, 5%, 1% de la distribución F de tablas o Stata
    - display invFtail (3,60,0.05)
- Regla de rechazo: Rechazamos H_0 en favor de H_a al $100\alpha\%$ si

$$\mathrm{F} > c_\alpha$$

  - Ej. Si $q = 3$, $n-k-1 = 60$,
    - $\alpha = 0.05$, $c_\alpha = 2.76$, rechazamos H_0 al 5% si $\mathrm{F} > 2.76$
    - $\alpha = 0.01$, $c_\alpha = 4.13$, rechazamos H_0 al 1% si $\mathrm{F} > 4.13$

[Gráfica]

- En aplicaciones, $q << n-k-1$
  - Si $n-k-1$ pequeño, los parámetros en el modelo sin restricciones se estiman con poca precisión
  - Si $n-k-1 > 120$, $g.l. = \infty$ en tablas F

- Si rechazamos H_0, decimos que $x_{k-q+1}, \ldots, x_k$ son estadísticamente significativas de forma conjunta (o conjuntamente significativas) al $100\alpha\%$
  - Prueba no nos dice qué variable es significativa
  - Pueden ser todas, puede ser solo una
- Si H_0 no se rechaza, las variables son conjuntamente no significativas
  - Justifica quitarlas del modelo

> Ejemplo. ¿La productividad en el beisbol tiene efecto en el salario?
>
> $q = 3$, $n - k - 1 = 347$, $c_{0.05} = 2.6$, $c_{0.01} = 3.78$
>
> $$\mathrm{F}_d = \frac{(\mathrm{SCR}_r - \mathrm{SCR}_{sr})}{\mathrm{SCR}_{sr}} \cdot \frac{n-k-1}{q} = \frac{(198.311 - 183.186)}{183.186} \cdot \frac{347}{3} \approx 9.55$$
>
> ¿Conclusión?
>
> - H_0 : Productividad no tiene efecto en el salario
> - ¿Rechazamos H_0 o no?
> - ¿Significado?

- ¿Por que estadísticos t individuales no son significativos pero el estadístico F sí?
  - MC dificulta asignar el efecto parcial a cada variable (por eso t's no significativas)
  - MC es menos importante para probar si variables de productividad son conjuntamente significativas
- Estadístico F es útil para probar la exclusión de un grupo de variables cuando están altamente correlacionadas
  - Ej. Probar si variables que miden el desempeño de una empresa afectan los salarios de CEOs

4.5.2 Relación entre estadísticos F y t

Si $\mathrm{H}_0 : \beta_k = 0$ y $q = 1$, ¿tenemos 2 formas de probar la hipótesis sobre el coeficiente?

$$\mathrm{t}^2_{n-k-1} \sim \mathrm{F}_{1,\,n-k-1}$$

- Se obtiene el mismo resultado si H_a es de 2 lados, pero
  - t es más flexible porque se puede usar directamente para H_a de un lado
  - t es más fácil de obtener para 1 restricción
- Casos posibles:
  - t's no significativas pero F es significativa

- t significativa pero grupo de variables (F) no es significativo
    * Lógicamente inconsistente (se rechaza $\mathrm{H}_0 : \beta_1 = 0$, pero no se rechaza $\mathrm{H}_0 : \beta_1 = 0, \beta_2 = 0, \beta_3 = 0$)
    * Puede abusarse si no se reporta adecuadamente
    * Ej. Aceptación de créditos: minoría (efecto marginal), edad (no efecto) y otros (ingreso, riqueza, buró), con prueba conjunta entre minoría y edad no significativa para concluir que minoría no es importante

- Comúnmente, si cada variable es estadísticamente significativa, el conjunto de variables también será estadísticamente significativo
  - No hay inconsistencia lógica en rechazar ambas hipótesis nulas

4.5.3 Forma R^2 del estadístico F

El estadístico F se puede obtener usando las R^2's de los modelos

- Restringido: $\mathrm{SCR}_r = \mathrm{SCT}(1 - \mathrm{R}_r^2)$
- No restringido: $\mathrm{SCR}_{sr} = \mathrm{SCT}(1 - \mathrm{R}_{sr}^2)$

Forma R^2 del estadístico F:

$$\mathrm{F}_d = \frac{\left(\mathrm{R}_{sr}^2 - \mathrm{R}_r^2\right)/q}{(1 - \mathrm{R}_{sr}^2)/(n - k - 1)}$$

- No aplica para probar todas las restricciones lineales
  - Solo cuando $\mathrm{SCT}_{sr} = \mathrm{SCT}_r$ y misma variable dependiente
- ≥ 0 siempre porque $\mathrm{R}_{sr}^2 > \mathrm{R}_r^2$ (R_{sr}^2 primero vs SCR_r primero)
- Los valores de R^2's no se tienen que elevar al cuadrado
- Versión útil porque $0 < \mathrm{R}^2 < 1$, mientras que SCR depende de unidades de medición

Ejemplo. ¿La productividad en el beisbol tiene efecto en el salario?

$$\mathrm{F}_d = \frac{(0.6278 - 0.5971)}{(1 - 0.6278)} \cdot \frac{347}{3} \approx 9.54$$

Misma conclusión que antes (diferencia por redondeo)

Ejemplo. Datos faltantes
Tenemos: $Obs = 1,388$, faltantes $= 197$, $k+1 = 6$, $\mathrm{R}_{sr}^2 = 0.0387$, $\mathrm{R}_r^2 = 0.0364$

- $n = 1,191$
- g.l. numerador: $q = 2$
- g.l. denominador: $g.l._{sr} = 1,191 - 6 = 1,185$
- $\alpha = 0.05$, $c_\alpha = 3.0$

∗ display invFtail (2,1185,0.05)

$$F_d = \frac{(0.0387 - 0.0364)}{(1 - 0.0387)} \cdot \frac{1,185}{2} \approx 1.42$$

¿Conclusión?

4.5.4 Cálculo de valores − p para pruebas F

Para pruebas F:

$$\text{valor} - \text{p} = \mathbb{P}\left(\mathcal{F} > F_d\right)$$

- $\mathcal{F}$ es la variable aleatoria $F_{q,\,n-k-1}$
- F_d es el valor del estadístico F

Interpretación:

- Probabilidad de observar un valor de F tan grande como el obtenido dado que H_0 es cierta

Utilidad:

- Ayuda tener una mejor idea de la fuerza de la evidencia en contra de H_0
- Un valor − p pequeño es evidencia contra H_0 (ej. valor − p = 0.016)
- Al tener el valor − p, podemos hacer una prueba F para cualquier α
  - Ej. Si valor − p = 0.024, H_0 se rechaza al 5% pero no se rechaza al 1%

Stata:

- Permite hacer pruebas de varias restricciones (comando test)
- Reporta valores − p en automático
- Aborda el tema de datos faltantes
- Es menos probable cometer un error que al hacerlo a mano

4.5.5 Estadístico F para la significancia total de una regresión

Siempre podemos probar si ninguna variable independiente tiene un efecto en y (en modelo de k variables independientes)

- $H_0 : \beta_1 = \beta_2 = \ldots = \beta_k = 0$
- H_a : Al menos una $\beta_j \neq 0$
- En esencia prueba si $\mathbb{E}\left(y|x_1, x_2, \ldots, x_k\right) = \mathbb{E}\left(y\right)$
- Conocer los valores de las variables independientes no afecta el valor esperado de y

- Se usa para determinar significancia total de la regresión

Modelo restringido (k restricciones impuestas):

$$y = \beta_0 + u$$

- $\mathrm{R}_r^2 = 0$ (no hay variables que expliquen y)

Estadístico

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

- R^2 de regresión de y sobre $x_1, x_2, \ldots, x_k$
- Solo válido para probar exclusión conjunta de todas las variables independientes (ej. EMH)

Si no rechazamos H_0, no hay evidencia de que las variables independientes ayuden a explicar y

- Buscar otras variables para explicar y

Si $\mathrm{valor} - \mathrm{p} \approx 0$, rechazamos H_0

- Variables independientes explican algo de la variación en y (aún si R^2 baja)

R^2 baja puede resultar en estadístico F muy significativo

- Por eso calcular F para probar significancia conjunta y no solo R^2

4.5.6 Pruebas de restricciones lineales generales

Los estadísticos F se utilizan principalmente para probar restricciones de exclusión

- También se pueden usar restricciones que no solo excluyen variables independientes

Para ilustrar:

$$\log(precio) = \beta_0 + \beta_1 \log(avaluo) + \beta_2 \log(lote) + \beta_3 \log(pies2) + \beta_4 cuartos + u$$

- Si el avalúo es racional: $\beta_1 = 1$ y las otras variables no deberían ayudar a explicar el precio al controlar por el avalúo
- $\mathrm{H}_0 : \beta_1 = 1,\ \beta_2 = 0,\ \beta_3 = 0,\ \beta_4 = 0$
- 4 restricciones (3 de ellas de exclusión)
- Como antes:
  - Estimamos modelo sin restricciones
  - Imponemos las restricciones para obtener modelo restringido (cuidado)
- Modelo sin restricciones: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + u$
  - $g.l._{sr} = 5$ y obtenemos SCR_{sr}
- Modelo restringido: $\quad y = \beta_0 + x_1 + u$

  - Para imponer restricción $\beta_1 = 1$, estimamos $y - x_1 = \beta_0 + u$
  - Obtenemos SCR_r

- Estadístico F:

$$\text{F}_d = \frac{(\text{SCR}_r - \text{SCR}_{sr})/4}{\text{SCR}_{sr}/5}$$

  - No podemos usar forma R^2, variables dependientes son diferentes ($\text{SCT}_{sr} \neq \text{SCT}_r$)
  - Usamos forma SCR si variable dependiente en modelo restringido es diferente

Ejemplo. Precios de vivienda

$$\widehat{\log(precio)} = \underset{(0.57)}{0.264} + \underset{(0.151)}{1.043}\log(avaluo) + \underset{(0.0386)}{0.0074}\log(lote)$$

$$- \underset{(0.1384)}{0.1032}\log(pies2) + \underset{(0.0221)}{0.0338}\, cuartos$$

$$n = 88, \qquad \text{SCR} = 1.822 \qquad \text{R}^2 = 0.773 \qquad \text{SCR}_r = 1.880$$

- Si usamos t's individuales para cada hipótesis, no rechazamos ninguna
- Pero hipótesis de racionalidad es una hipótesis conjunta

$$\text{F}_d = \frac{(1.88 - 1.822)}{1.822} \cdot \frac{83}{4} = 0.661$$

- $c_{0.05} = 2.5$ para $\text{F}_{4,\,83}$
  - En Stata: test (lavaluo = 1) (llote lpies2 cuartos)
- ¿Hay evidencia contra hipótesis de que avalúos son racionales?

4.6 Reportar resultados de regresiones

Al reportar los resultados de un análisis de RLM, debemos incluir:

- Variable dependiente claramente indicada
- Variables independientes listadas en la primera columna
- Coeficientes:
  - Interpretar variables clave (saber unidades de medición)
  - Discutir importancia económica
- Errores estándar (en lugar de t's) en paréntesis debajo de los coeficientes
  - Nos forza a pensar en H_0 a probar (no siempre es $\text{H}_0 : \beta_j = 0$)

- Facilita calcular intervalos de confianza

- R^2:
  - Medida de bondad de ajuste
  - Facilita el cálculo de la F para restricciones de exclusión [SCR, σ no esenciales]

- Número de observaciones

- (Opcional) Estrellas para significancia respecto a $H_0 : \beta_j = 0$ (*,**,***)

[Gráfica]