

# Notas de Econometría

Pavel Solís  
2025

## 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: 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

- Valor esperado
- Varianza (para precisión de los estimadores)

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

- 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  $x_1, x_2, \dots, x_k$  y se distribuye normal con media cero y varianza  $\sigma^2$ :

$$u \sim N(0, \sigma^2)$$

Observaciones sobre el supuesto RLM.6:

- RLM.6 es más fuerte que supuestos anteriores
  - Si  $u$  es independiente de  $x_j$ 
$$\mathbb{E}(u \mid x_1, x_2, \dots, x_k) = \mathbb{E}(u) = 0$$
$$\text{Var}(u \mid x_1, x_2, \dots, x_k) = \text{Var}(u) = \sigma^2$$
- 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 MCO  $\hat{\beta}_j$  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

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

- $y$  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 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.  $\log(\text{precio})$  vs  $\text{precio}$
  - A veces, RLM.6 es claramente falsa como cuando  $y$  solo toma algunos valores
    - \* 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
- 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

$$\hat{\beta}_j \sim N[\beta_j, \text{Var}(\hat{\beta}_j)]$$

donde  $\text{Var}(\hat{\beta}_j) = \frac{\sigma^2}{\text{SCT}_j(1-R_j^2)}$ ,  $j = 1, 2, \dots, k$ , entonces

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

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

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

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

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

Normalidad de los estimadores MCO es aproximadamente cierta para muestras grandes

- Aún sin normalidad de los errores

Problema:  $\text{desvest}(\hat{\beta}_j)$  depende de  $\sigma$

## 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

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

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

Teorema. Distribución t para Estimadores Estandarizados

Bajo supuestos MLC, RLM.1 a RLM.6,

$$\frac{\hat{\beta}_j - \beta_j}{\text{errest}(\hat{\beta}_j)} \sim t_{n-k-1}$$

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

Observaciones sobre el teorema:

- Dos últimos teoremas son diferentes
  - Antes,  $\text{desvest}(\hat{\beta}_j)$  y distribución N
  - Ahora,  $\text{errest}(\hat{\beta}_j)$  y distribución t
- Distribución t
  - Se obtiene porque la constante  $\sigma$  en  $\text{desvest}(\hat{\beta}_j)$  se reemplaza con la variable aleatoria  $\hat{\sigma}$  en  $\text{errest}(\hat{\beta}_j)$
  - Demostración difícil y poco informativa, pero en esencia muestra que es la división de una normal estándar  $N(0, 1)$  sobre la raíz cuadrada de una  $\chi^2_{n-k-1}$
  - 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  $\beta_j$

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

$$H_0 : \beta_j = 0, \quad j = 1, 2, \dots, k$$

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

Ejemplo.

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

$H_0 : \beta_2 = 0$  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  $H_0 : \beta_j = 0$  y se define

$$t_{\hat{\beta}_j} = \frac{\hat{\beta}_j}{\text{errest}(\hat{\beta}_j)}$$

- 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  $\hat{\beta}_{educ}$
- Reportado por Stata
- Propiedades:
  - $t_{\hat{\beta}_j}$  tiene el mismo signo que  $\hat{\beta}_j$  porque  $\text{errest}(\hat{\beta}_j) > 0$
  - Fijando  $\text{errest}(\hat{\beta}_j)$ , mayores valores de  $\hat{\beta}_j$  implican mayores valores de  $t_{\hat{\beta}_j}$

¿Qué tan lejos está  $\hat{\beta}_j$  de cero?

- $|\hat{\beta}_j| \gg 0$  puede ser evidencia en contra de  $H_0 : \beta_j = 0$
- Pero hay error de muestreo en estimado  $\hat{\beta}_j$ 
  - Por eso comparamos el tamaño de  $\hat{\beta}_j$  contra su error muestral
  - $\text{errest}(\hat{\beta}_j)$  es el estimado de  $\text{desvest}(\hat{\beta}_j)$
- Entonces,  $t_{\hat{\beta}_j}$  mide cuántas desviaciones estándar  $\hat{\beta}_j$  se aleja de cero
  - Si  $|t_{\hat{\beta}_j}| \gg 0$ , rechazamos  $H_0$

**Regla de rechazo** depende de

- Hipótesis alternativa  $H_a$  (un lado, dos lados)
- Nivel de significancia de la prueba  $\alpha$

$$\alpha = \mathbb{P}(\text{Rechazar } H_0 \mid H_0)$$

- Valores comunes:  $\alpha = 0.1, 0.05, 0.01$
- Requiere conocer la distribución muestral de  $t_{\hat{\beta}_j}$
- Por teorema,  $t_{\hat{\beta}_j} \sim t_{n-k-1}$  cuando  $H_0$  es cierta

Notación:

Pregunta. ¿Cuál de las opciones siguientes es correcta?

- $H_0 : \beta_j = 0$
- $H_0 : \hat{\beta}_j = 0$

- 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  $H_a : \beta_j > 0$

- Para escoger una regla de rechazo, decidimos el nivel de significancia  $\alpha$  o  $\mathbb{P}(\text{Rechazar } H_0 \mid H_0)$ 
  - Si  $\alpha = 0.05$ , estamos dispuestos a rechazar  $H_0$  cuando es cierta 5% de las veces
- Bajo  $H_0$ ,  $t_{\hat{\beta}_j} \sim t_{n-k-1}$  y  $\mathbb{E}(t_{\hat{\beta}_j}) = 0$  (por teorema) pero bajo  $H_a$ ,  $\mathbb{E}(t_{\hat{\beta}_j}) > 0$ 
  - Buscamos un valor positivo ‘suficientemente grande’ de  $t_{\hat{\beta}_j}$  para rechazar  $H_0 : \beta_j = 0$  en favor de  $H_a : \beta_j > 0$
  - Valores negativos de  $t_{\hat{\beta}_j}$  no dan evidencia a favor de  $H_a$
- ‘Suficientemente grande’ es el **valor crítico**  $c$  dado el nivel de significancia  $\alpha$ :  $c_\alpha$ 
  - Si  $\alpha = 0.05$ ,  $c$  es el percentil 95 de una distribución  $t_{n-k-1}$
  - Para obtener  $c$ , se necesita  $\alpha$  y los grados de libertad
  - Los valores de  $c$  se obtienen de tablas o en Stata [`display invttail(gl, \alpha)`]
- Regla de rechazo (prueba de una cola): Se rechaza  $H_0$  en favor de  $H_a$  con un nivel de significancia  $100\alpha$  si

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

Ejemplo. Si  $\alpha = 0.05$  y  $n - k - 1 = 28$ , entonces  $c_{0.05} = 1.701$ . Si  $t_{\hat{\beta}_j} \leq 1.701$ , no rechazamos  $H_0$  al 5%

[Gráfica]

- Patrón: Si  $\alpha$  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  $\alpha$  baja, se requieren valores de  $t_{\hat{\beta}_j}$  más grandes para rechazar  $H_0$ 
  - \* Si  $H_0$  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  $N(0, 1)$  [`display invnormal(0.95)`]
  - \* Cuando  $n - k - 1 > 120$ , ambas distribuciones son similares

Ejemplo. Del análisis del salario (wage1.dta)

$$\log(\widehat{\text{salario}}) = \begin{matrix} 0.284 \\ (0.104) \end{matrix} + \begin{matrix} 0.092 \text{educ} \\ (0.007) \end{matrix} + \begin{matrix} 0.0041 \text{exper} \\ (0.0017) \end{matrix} + \begin{matrix} 0.022 \text{antig} \\ (0.003) \end{matrix}$$

$$n = 526, R^2 = 0.316$$

- Errores estándar en paréntesis

- Interesados en probar  $H_0 : \beta_{exper} = 0$  vs  $H_a : \beta_{exper} > 0$
- Grados de libertad:  $526 - 4 = 522$
- Valores críticos de  $\phi(z)$ :  $c_{0.1} = 1.282$ ,  $c_{0.05} = 1.645$ ,  $c_{0.01} = 2.326$
- Estadístico para  $\hat{\beta}_{exper}$ :

$$t_{exper} = \frac{0.0041}{0.0017} \approx 2.41$$

- 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  $H_a : \beta_j < 0$ , la regla de rechazo (prueba de una cola) es

$$t_{\hat{\beta}_j} < -c$$

- $c > 0$  en tablas (distribución t es simétrica,  $c$  viene de  $H_a : \beta_j > 0$ )

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

[Gráfica]

- Dado que  $H_a : \beta_j < 0$ ,  $t_{\hat{\beta}_j} < 0$  para rechazar  $H_0$ 
  - Si  $t_{\hat{\beta}_j} > 0$ , no da evidencia en favor de  $H_a$

Ejemplo.

$$\widehat{examen} = \begin{matrix} 2.274 \\ (6.113) \end{matrix} + \begin{matrix} 0.00046 \textit{comp} \\ (0.00010) \end{matrix} + \begin{matrix} 0.048 \textit{prsnl} \\ (0.04) \end{matrix} - \begin{matrix} 0.00020 \textit{inscr} \\ (0.00022) \end{matrix}$$

$$n = 408, R^2 = 0.0541$$

- Errores estándar en paréntesis
- *comp* para calidad del profesor, *prsnl* para atención a estudiantes, *inscr* para tamaño de la escuela
- Interesados en probar si escuelas grandes tienen bajo desempeño
  - \*  $H_0 : \beta_{inscr} = 0$  vs  $H_a : \beta_{inscr} < 0$
- ¿ $t_{\hat{\beta}_{inscr}} < 0$  observado es por error muestral?
- $n - k - 1 = 408 - 4 = 404$

– Valor crítico de  $\phi(z)$  al 5%:  $-c_{0.05} = -1.65$

– Estadístico para  $\hat{\beta}_{inscr}$ :

$$t_{inscr} = \frac{-0.00020}{0.00022} \approx -0.91$$

– No rechazamos  $H_0$

\* *inscr* no es estadísticamente significativa a un nivel del 5%

–  $t_{comp} = 4.6 > 2.33 = c_{0.01}$ , se rechaza  $H_0 : \beta_{comp} = 0$  a un nivel del 1%

–  $t_{prsnl} = 1.2 < 1.28 = c_{0.1}$ , 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:

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

- El valor crítico  $c$  se determina al escoger  $\alpha$  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.

$$\widehat{califuni} = 1.39 + 0.412 califprep + 0.015 examen - 0.083 faltas$$

(0.33)      (0.094)                      (0.011)                      (0.026)

$$n = 141, R^2 = 0.234$$

- 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)`
- $t_{califprep} = 4.38 > 2.58$ , estadísticamente significativa al 1%
- $t_{examen} = 1.36$ , no es estadísticamente significativa al 10%
- $t_{faltas} = -3.19 < -2.58$ , 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 ( $0.083 \times 5$  porque  $\Delta faltas = 5$ )

### 4.2.3 Pruebas de hipótesis para otros valores de $\beta_j$

$H_0 : \beta_j = 0$  es la hipótesis nula más común pero podemos hacer pruebas para cualquier valor hipotético  $a_j$

- $H_0 : \beta_j = a_j$
- $\beta_j = 1$  o  $\beta_j = -1$  son comunes

El estadístico t apropiado es

$$t_j = \frac{\hat{\beta}_j - a_j}{\text{errest}(\hat{\beta}_j)}$$

- En general, el estadístico se puede expresar

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

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

Bajo  $H_0 : \beta_j = a_j$ , el estadístico  $t_j \sim t_{n-k-1}$  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  $H_a$  de uno y dos lados

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

Si  $H_0 : \beta_j = 1$  y  $H_a : \beta_j > 1$ , rechazamos  $H_0$  si  $t > c$  y decimos que “ $\hat{\beta}_j$  es estadísticamente mayor a 1” a un nivel de  $100\alpha\%$

$$t = \frac{\hat{\beta}_j - 1}{\text{errest}(\hat{\beta}_j)}$$

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

- Modelo poblacional:

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

- Modelo de elasticidad constante

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

[Gráfica]

- $H_0 : \beta_1 = 0$  es obvio; más interesante  $H_0 : \beta_1 = 1$  vs  $H_0 : \beta_1 > 1$

- Si se rechaza  $H_0$ , el crimen es un problema grave en campus grandes

$$\widehat{\log(\text{crimen})} = \begin{matrix} -6.63 \\ (1.03) \end{matrix} + \begin{matrix} 1.27 \log(\text{inscr}) \\ (0.11) \end{matrix}$$

$$n = 97, R^2 = 0.585$$

- $\hat{\beta}_1 > 1$ , 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  $H_0 : \beta_j = -1$  y  $H_a : \beta_j \neq -1$ , rechazamos  $H_0$  si  $|t| > c_{\alpha/2}$  y decimos que “ $\hat{\beta}_j$  es estadísticamente diferente de  $-1$ ” al nivel apropiado

$$t = \frac{\hat{\beta}_j - (-1)}{\text{errest}(\hat{\beta}_j)}$$

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

- Modelo poblacional:

$$\log(\text{precio}) = \beta_0 + \beta_1 \log(\text{oxn}) + \beta_2 \log(\text{dist}) + \beta_3 \text{habit} + \beta_4 \text{alprof} + u$$

- Modelo estimado:

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

$$n = 506, R^2 = 0.581$$

- Todos los coeficientes son estadísticamente diferentes de cero con bajos niveles de significancia pero  $H_0 : \beta_j = -1$  (no  $H_0 : \beta_j = 0$ )
- ¿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  $H_a$ , escoger  $\alpha$  que determina  $c$ , comparar  $t$  contra  $c$ , conclusión (rechazar  $H_0$  o no rechazar  $H_0$ )

- Hay arbitrariedad al escoger  $\alpha$  (cada quien puede escoger diferente  $\alpha$ )
- Ej.  $H_0$  puede no rechazarse al 5% pero si al 10%
- En lugar de hacer la prueba para diferentes niveles de  $\alpha$ , 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  $\alpha$  para el que se rechazaría  $H_0$ ?
- Es el valor de  $\alpha$  si usamos el valor de  $t$  como el valor de  $c$

Ejemplo.  $H_0 : \beta_j = 0$ ,  $H_a : \beta_j \neq 0$ ,  $g.l = 40$ ,  $t = 1.85$

- $c_{0.05} = 2.021$ , no se rechaza  $H_0$
- $c_{0.1} = 1.684$ , se rechaza  $H_0$
- $0.05 < \text{valor } -p < 0.1$
- $\text{valor } -p = \mathbb{P}(|t_{40}| > 1.85) = 2\mathbb{P}(t_{40} > 1.85) = 2(0.0359) = 0.0718$
- En Stata, display `ttail(40, 1.85)*2`

[Gráfica]

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

- Ej. Si  $H_0 : \beta_j = 0$  y  $H_a : \beta_j \neq 0$ ,

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

donde  $T = t_{n-k-1}$  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 valor  $-p$  pequeño es evidencia contra  $H_0$

- Un valor  $-p$  grande da poca evidencia contra  $H_0$
- Ej. valor  $-p = 0.5$  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 valor  $-p$ , se puede hacer una prueba clásica para cualquier  $\alpha$

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

Para una  $H_a$  de un lado

- Si  $H_0 : \beta_j = 0$ 
  - $H_a : \beta_j > 0$ , solo calcular si  $\hat{\beta}_j > 0$  ( $t > 0$ ) y valor  $-p = \mathbb{P}(T > t)$
  - $H_a : \beta_j < 0$ , solo calcular si  $\hat{\beta}_j < 0$  ( $t < 0$ ) y valor  $-p = \mathbb{P}(T < t) = \mathbb{P}(T > |t|)$
- valor  $-p$  de un lado =  $\frac{\text{valor}-p \text{ de dos lados}}{2}$

No es crucial reportar el valor  $-p$  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.

$$H_0 : \beta_j = 1, t = 0.393$$

$$H_0 : \beta_j = 0.99, t = 0.308$$

- 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  $t_{\hat{\beta}_j}$
- Significancia económica: Magnitud y signo del coeficiente estimado  $\hat{\beta}_j$

Ej. Cuando probamos  $H_0 : \beta_j = 0$ ,  $t_{\hat{\beta}_j} = \frac{\hat{\beta}_j}{\text{errest}(\hat{\beta}_j)}$  puede ser significativo porque

- $\hat{\beta}_j$  es grande, o

- $\text{errest}(\hat{\beta}_j)$  es pequeño

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

Ejemplo. Tasa de participación en planes de retiro

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

$$n = 1,534, R^2 = 0.1$$

- $t_{emplead} = \frac{-0.00013}{0.00004} = -3.25$
- 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 es poco en términos prácticos

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

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

Relación entre el tamaño de muestra ( $n$ ) y el nivel de significancia ( $\alpha$ )

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

Ejemplo. Tasa de productos defectuosos (por cada 100)

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

$$n = 29, R^2 = 0.262$$

- La variable de interés son las horas de capacitación ( $hrcpct$ )
  - $H_0 : \beta_{hrcpct} = 0$  y  $H_a : \beta_{hrcpct} < 0$
  - $\beta_{hrcpct} = -0.029$ , ¿cuánto baja la tasa con 5 horas más de capacitación?

- Efecto no trivial
- $t_{hrcpct} = -0.029/0.023 = -1.26, g.l. = 29 - 4 = 25$  ( $n$  pequeña,  $\alpha$  alta)
  - $c_{0.05} = -1.71$ ,  $hrcpct$  no es estadísticamente significativa al 5%
  - $c_{0.1} = -1.32$ ,  $hrcpct$  casi estadísticamente significativa al 10%
- valor  $-p = \mathbb{P}(T_{25} < -1.26) = 0.11$ 
  - Bajo para deberse a error muestral pero sujeto a interpretación

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

- 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 valor  $-p$  para el estadístico  $t$ .
    - Si  $n$  pequeña, valor  $-p < 0.2$
    - Si valor  $-p > 0.2$ , coeficiente grande por error muestral
3. Si estadístico  $t$  pequeño y signo “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  $\beta_j$  bajo los supuestos del MLC

- Dan un *rango* de valores para  $\beta_j$
- Diferentes a estimados puntuales

- También conocidos como estimados por intervalos

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

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

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

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

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

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

Elementos para construir un IC:

- $\hat{\beta}_j$
- $\text{errest}(\hat{\beta}_j)$
- $c$  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

$$\left[ \hat{\beta}_j - 2.06 \cdot \text{errest}(\hat{\beta}_j), \hat{\beta}_j + 2.06 \cdot \text{errest}(\hat{\beta}_j) \right]$$

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

- Si  $n - k - 1 > 120$ ,  $t_{n-k-1} \approx N$ , entonces usamos el percentil 97.5 de  $\Phi$

$$\hat{\beta}_j \pm 1.96 \cdot \text{errest}(\hat{\beta}_j)$$

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

$$\hat{\beta}_j \pm 2 \cdot \text{errest}(\hat{\beta}_j)$$

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

Si cambia  $\alpha$  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  $\Phi$ 
  - Ej. g.l. =  $n - k - 1 = 25$ ,  $c = 1.71$ , IC al 90% es (más angosto)

$$\hat{\beta}_j \pm 1.71 \cdot \text{errest}(\hat{\beta}_j)$$

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

$$\hat{\beta}_j \pm 2.79 \cdot \text{errest}(\hat{\beta}_j)$$

Con IC, podemos hacer PH de 2 lados:

- Si  $H_0 : \beta_j = a_j$  y  $H_a : \beta_j \neq a_j$ , rechazamos  $H_0$  al 5% si  $a_j \notin \text{IC del 95\%}$

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

$$\log(\widehat{invdes}) = \begin{matrix} -4.38 \\ (0.47) \end{matrix} + \begin{matrix} 1.084 \log(ventas) \\ (0.060) \end{matrix} + \begin{matrix} 0.0217 margengan \\ (0.0128) \end{matrix}$$

$$n = 32, R^2 = 0.918$$

- 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  $(0.961, 1.21)$ 
  - $0 \notin \text{IC}$  porque se espera que gasto I&D crezca con el tamaño de la empresa
  - $1 \in \text{IC}$ , no podemos rechazar  $H_0 : \beta_{\log(ventas)} = 1$  vs  $H_a : \beta_{\log(ventas)} \neq 1$  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  $H_0 : \beta_{margengan} = 0$  vs  $H_a : \beta_{margengan} \neq 0$  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,  $\text{errest}(\hat{\beta}_j)$  no es válido para estimar  $\text{desvest}(\hat{\beta}_j)$ 
  - IC no será al 95%
- Si normalidad del error no se cumple, el IC no es válido
  - Aunque esto se puede relajar cuando  $n \rightarrow \infty$

## 4.4 PH para una combinacion lineal de parámetros

Hasta ahora PH o IC para una sola  $\beta_j$ , aquí cómo probar una sola hipótesis con varias  $\beta_j$ 's

Para ilustrar, modelo poblacional para gente trabajadora con título de preparatoria,:

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

- Hipótesis de interés: ¿Un año de carrera técnica vale lo mismo que un año de licenciatura?
  - $H_0 : \beta_1 = \beta_2$ , ambas escuelas generan mismo aumento porcentual en salario (ceteris paribus)
  - $H_a : \beta_1 < \beta_2$ , un año de carrera técnica vale menos que un año de licenciatura
- No podemos usar estadísticos t individuales para  $\hat{\beta}_1$  y  $\hat{\beta}_2$
- Reexpresamos hipótesis como
  - $H_0 : \beta_1 - \beta_2 = 0$
  - $H_a : \beta_1 - \beta_2 < 0$
- El estadístico quedaría

$$\frac{\hat{\beta}_1 - \hat{\beta}_2}{\text{errest}(\hat{\beta}_1 - \hat{\beta}_2)}$$

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

- ¿ $\text{errest}(\hat{\beta}_1 - \hat{\beta}_2)$ ?  $\text{errest}(\hat{\beta}_1 - \hat{\beta}_2) \neq \text{errest}(\hat{\beta}_1) - \text{errest}(\hat{\beta}_2)$ 
  - Por probabilidad,  $\text{Var}(\hat{\beta}_1 - \hat{\beta}_2)$  y  $\text{desvest}(\hat{\beta}_1 - \hat{\beta}_2) = \sqrt{\text{Var}(\hat{\beta}_1 - \hat{\beta}_2)}$
  - Entonces,  $\text{errest}(\hat{\beta}_1 - \hat{\beta}_2) = \sqrt{\text{errest}^2(\hat{\beta}_1) + \text{errest}^2(\hat{\beta}_2) - 2s_{1,2}}$
- 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  $H_0 : \theta_1 = 0$  vs  $H_a : \theta_1 < 0$ , y  $t = \frac{\hat{\theta}_1}{\text{errest}(\hat{\theta}_1)}$

- Definimos  $\beta_1 = \theta_1 + \beta_2$ , entonces

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

- Se crea una nueva variable  $\text{univ} = \text{carrtec} + \text{licen}$ , no se estima  $\beta_1$  y se obtiene  $\text{errest}(\hat{\theta}_1)$  directo

Ejemplo. Salario y estudios universitarios

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

$$n = 6,763, R^2 = 0.222$$

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

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

$$n = 6,763, R^2 = 0.222$$

- $t_{\hat{\theta}_1} = \frac{-0.0102}{0.0069} = -1.48$
- valor  $-p = 0.07$ , algo de evidencia pero no fuerte contra  $H_0$
- IC al 95% con aproximación  $\Phi$ :

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

## 4.5 Pruebas de varias restricciones lineales: Prueba F

Hasta ahora hipótesis de una sola restricción, aquí varias restricciones

- Antes: Una  $\beta_j$  o una combinación lineal de  $\beta_j$ '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:

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

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

$$\begin{aligned} \widehat{\log(\text{salar})} = & \quad 11.19 \quad + \quad 0.0687 \text{ansliga} \quad + \quad 0.0126 \text{promjueg} \\ & (0.27) \quad (0.0121) \quad (0.0026) \\ & + 0.00098 \text{prombat} \quad + \quad 0.0144 \text{hruns} \quad + \quad 0.0108 \text{carrbat} \\ & (0.0011) \quad (0.0161) \quad (0.0072) \end{aligned}$$

$$n = 353, \quad \text{SCR} = 183.186 \quad R^2 = 0.6278$$

- ¿t individuales de variables de productividad son significativas?

Modelo restringido (sin variables en  $H_0$ ):

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

$$\begin{aligned} \widehat{\log(\text{salar})} = & \quad 11.22 + \quad 0.0713 \text{ansliga} \quad + \quad 0.0202 \text{promjueg} \\ & (0.11) \quad (0.0125) \quad (0.0013) \end{aligned}$$

$$n = 353, \quad \text{SCR} = 198.311 \quad R^2 = 0.5971$$

- 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  $\alpha$

Caso general (ambos modelos con intercepto):

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

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

- Suponemos que hay  $q$  restricciones de exclusión (tienen coeficientes cero)
  - Últimas  $q$  variables se quitan del modelo
- $H_0 : \beta_{k-q+1} = 0, \dots, \beta_k = 0$
- $H_a : H_0$  es falsa
  - Al menos uno de los parámetros  $(\beta_{k-q+1}, \dots, \beta_k)$  es distinto de cero
- Modelo con restricciones (impuestas):

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

Estadístico F o razón F:

$$F_d = \frac{(\text{SCR}_r - \text{SCR}_{sr}) / q}{\text{SCR}_{sr} / (n - k - 1)}$$

- $\text{SCR}_r$  del modelo restringido
- $\text{SCR}_{sr}$  del modelo sin restricciones
- $\geq 0$  porque  $\text{SCR}_r \geq \text{SCR}_{sr}$  (por lo que si  $F < 0$ , 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  $\sigma^2 = \text{Var}(u)$  en modelo sin restricciones
- Compara el aumento relativo en SCR al ir del modelo sin restricciones al restringido

Para calcular el estadístico F, necesitamos:  $g.l._{sr}$ ,  $q$ ,  $\text{SCR}_{sr}$  y  $\text{SCR}_r$

- 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  $H_0$ ,

$$F_d = \frac{(\text{SCR}_r - \text{SCR}_{sr})/q}{\text{SCR}_{sr}/(n - k - 1)} \sim F_{q, n-k-1}$$

- Intuición: F es una división de 2 variables aleatorias  $\chi^2$  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  $\alpha = 0.05$ ,  $c_\alpha$  es el percentil 95 de la distribución  $F_{q, n-k-1}$

- \*  $c$  depende entonces de  $\alpha$ ,  $q$  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

$$F > c_\alpha$$

- Ej. Si  $q = 3$ ,  $n - k - 1 = 60$ ,

- \*  $\alpha = 0.05$ ,  $c_\alpha = 2.76$ , rechazamos  $H_0$  al 5% si  $F > 2.76$

- \*  $\alpha = 0.01$ ,  $c_\alpha = 4.13$ , rechazamos  $H_0$  al 1% si  $F > 4.13$

[Gráfica]

- En aplicaciones,  $q \ll n - k - 1$

- 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  $x_{k-q+1}, \dots, x_k$  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?

$$q = 3, n - k - 1 = 347, c_{0.05} = 2.6, c_{0.01} = 3.78$$

$$F_d = \frac{(\text{SCR}_r - \text{SCR}_{sr})}{\text{SCR}_{sr}} \cdot \frac{n - k - 1}{q} = \frac{(198.311 - 183.186)}{183.186} \cdot \frac{347}{3} \approx 9.55$$

¿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  $H_0 : \beta_k = 0$  y  $q = 1$ , ¿tenemos 2 formas de probar la hipótesis sobre el coeficiente?

$$t_{n-k-1}^2 \sim F_{1, n-k-1}$$

- 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  $H_0 : \beta_1 = 0$ , pero no se rechaza  $H_0 : \beta_1 = 0, \beta_2 = 0, \beta_3 = 0$ )
    - \* 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:  $SCR_r = SCT(1 - R_r^2)$
- No restringido:  $SCR_{sr} = SCT(1 - R_{sr}^2)$

Forma  $R^2$  del estadístico F:

$$F_d = \frac{(R_{sr}^2 - R_r^2) / q}{(1 - R_{sr}^2) / (n - k - 1)}$$

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

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

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

Misma conclusión que antes (diferencia por redondeo)

Ejemplo. Datos faltantes

Tenemos:  $Obs = 1,388$ , faltantes = 197,  $k+1 = 6$ ,  $R_{sr}^2 = 0.0387$ ,  $R_r^2 = 0.0364$

- $n = 1,191$
- 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:

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

- $\mathcal{F}$  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  $\alpha$ 
  - 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)

- $H_0 : \beta_1 = \beta_2 = \dots = \beta_k = 0$
- $H_a : \text{Al menos una } \beta_j \neq 0$
- En esencia prueba si  $\mathbb{E}(y|x_1, x_2, \dots, x_k) = \mathbb{E}(y)$
- 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):

$$y = \beta_0 + u$$

- $R_r^2 = 0$  (no hay variables que expliquen  $y$ )

Estadístico

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

- $R^2$  de regresión de  $y$  sobre  $x_1, x_2, \dots, x_k$



- 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 valor  $-p \approx 0$ , 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:

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

- Si el avalúo es racional:  $\beta_1 = 1$  y las otras variables no deberían ayudar a explicar el precio al controlar por el avalúo
- $H_0 : \beta_1 = 1, \beta_2 = 0, \beta_3 = 0, \beta_4 = 0$
- 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:  $y = \beta_0 + x_1 + u$ 
  - Para imponer restricción  $\beta_1 = 1$ , estimamos  $y - x_1 = \beta_0 + u$
  - Obtenemos  $SCR_r$
- Estadístico F:

$$F_d = \frac{(SCR_r - SCR_{sr})/4}{SCR_{sr}/5}$$

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

Ejemplo. Precios de vivienda

$$\widehat{\log(\text{precio})} = \begin{array}{ccc} 0.264 & + & 1.043 \log(\text{avaluo}) + 0.0074 \log(\text{lote}) \\ (0.57) & & (0.151) \quad (0.0386) \end{array}$$
$$- 0.1032 \log(\text{pies2}) + 0.0338 \text{cuartos}$$
$$(0.1384) \quad (0.0221)$$

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

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

$$F_d = \frac{(1.88 - 1.822)}{1.822} \cdot \frac{83}{4} = 0.661$$

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

## 4.6 Reportar resultados de regresión

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  $H_0 : \beta_j = 0$ )
  - Facilita calcular intervalos de confianza
- $R^2$ :
  - Medida de bondad de ajuste
  - Facilita el cálculo de la F para restricciones de exclusión [SCR,  $\sigma$  no esenciales]
- Número de observaciones
- (Opcional) Estrellas para significancia respecto a  $H_0 : \beta_j = 0$  (\*, \*\*, \*\*\*)

[Gráfica]