TEMA 4
INFERENCIA Y PREDICCIÓN

S. Álvarez, A. Beyaert, M. Camacho, M. González, A. Quesada
Departamento de Métodos Cuantitativos para la Economía y la Empresa
UNIVERSIDAD DE MURCIA

Lo que estudiaremos en este tema:

UNIVERSIDAD DE **MURCIA**

Bibliografía básica: Wooldridge, 2008, cap.4 y 5

1. Distribuciones muestrales de los estimadores MCO

UNIVERSIDAD DE MURCIA

- Para hacer inferencia estadística es necesario conocer la distribución muestral de los estimadores MCO.
- Esta distribución depende de la distribución subyacente de los términos de error.
- Añadimos:<u>RLM.6 Normalidad de los términos de error e independencia de las explicativas:</u>

$$\rightarrow \ \varepsilon \sim \mathrm{N}(0,\sigma^2)$$

$$\rightarrow \ \varepsilon \text{ es independiente de las variables explicativas}$$

- RLM.1-RLM.6: **Supuestos del modelo lineal clásico (MLC)**.
- RLM.1-RLM.6 garantizan que el estimador MCO es el más preciso de entre todos los estimadores insesgados (lineales y no lineales). Es el estimador **eficiente**

¿Qué relación tiene RLM.6 con los supuestos RLM.1-RLM.5?

- Distribución de la variable dependiente condicionada a las variables explicativas:

$$E(y_i | x_{11}, x_{21}, \cdots, x_{(k-1)N}, x_{kN}) = \beta_0 + \beta_1 x_{1i} + \ldots + \beta_k x_{ki}$$

$$V(y_i | x_{11}, x_{21}, \cdots, x_{(k-1)N}, x_{kN}) = \sigma^2$$

$$\Rightarrow y_i | x_{11}, x_{21}, \cdots, x_{(k-1)N}, x_{kN} \sim N(\beta_0 + \beta_1 x_{1i} + \ldots + \beta_k x_{ki}, \sigma^2)$$

- Distribución de los estimadores MCO condicionada a las variables explicativas:

$$\hat{\beta}_j | x_{11}, x_{21}, \cdots, x_{(k-1)N}, x_{kN} \sim N\left(\beta_j, V\left(\hat{\beta}_j\right)\right) \Rightarrow \quad \frac{\hat{\beta}_j - \beta_j}{\sqrt{V(\hat{\beta}_j)}} \sim N\left(0,1\right)$$

- <u>Nota:</u> Aunque todas las distribuciones son condicionadas a los valores de las variables explicativas, para simplificar la notación en las ecuaciones del resto del tema no aparece dicha condición.

$$y_i = \beta_0 + \beta_1 x_{1i} + \ldots + \beta_k x_{ki} + \varepsilon_i$$

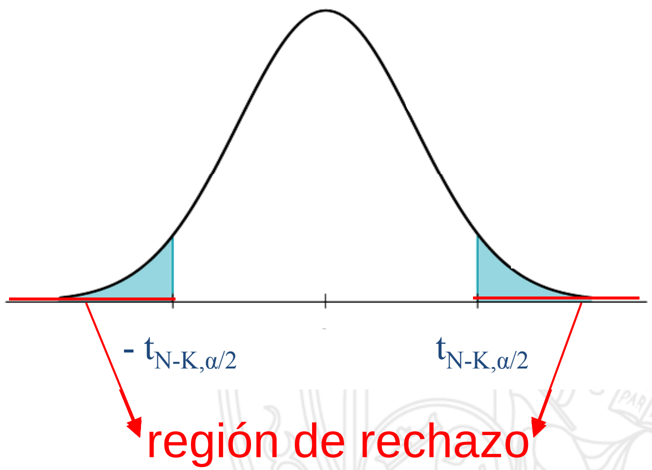
- Queremos contrastar una hipótesis sobre el valor de un β_j, por lo que necesitamos un estadístico con función de distribución conocida para realizar el contraste.
- Bajo RLM.1-RLM.6: $$\frac{\hat{\beta}_j - \beta_j}{\sqrt{V(\hat{\beta}_j)}} \sim N(0,1)$$
- Como $V(\hat{\beta}_j)$ es desconocida, la estimamos por $\hat{V}(\hat{\beta}_j)$ y obtenemos el **estadístico** para contrastar hipótesis sobre β_j:

$$\frac{\hat{\beta}_j - \beta_j}{\sqrt{\hat{V}(\hat{\beta}_j)}} \sim t_{N-K}$$ **estadístico t**

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

- **Estadístico de contraste:**

$$t = \frac{\hat{\beta}_j - \beta_j^0}{\sqrt{\hat{V}(\hat{\beta}_j)}} \underset{H_0}{\sim} t_{N-K}$$



- **Regla de rechazo:** $|t| > t_{N-K;\alpha/2} \Rightarrow RH_0$

- $t_{N-K;\alpha/2}$: valor crítico (cuantil $(100-\alpha/2)\%$ a la izquierda de una t-Student con N-K grados de libertad).
- α : nivel de significatividad (o tamaño) del contraste.

UNIVERSIDAD DE MURCIA

- **Caso particular:** Contraste de significatividad individual:

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

- **Estadístico de contraste:**

$$t = \frac{\hat{\beta}_j}{\sqrt{\hat{V}(\hat{\beta}_j)}} \underset{H_0}{\sim} t_{N-K}$$

- **Regla de rechazo:** $|t| > t_{N-K;\alpha/2} \Rightarrow RH_0$

- $RH_0 \Rightarrow \hat{\beta}_j$ es significativo. Significa que, según la información muestral, $\beta_j \neq 0$, y por tanto x_j tiene un efecto parcial no nulo sobre y.

- no $RH_0 \Rightarrow \hat{\beta}_j$ no es significativo. Significa que, según la información muestral, $\beta_j = 0$ y por tanto x_j no tiene efecto sobre y.

UNIVERSIDAD DE MURCIA

(a) $H_0: \beta_j = \beta_j^0$

$H_1: \beta_j > \beta_j^0$

(b) $H_0: \beta_j = \beta_j^0$

$H_1: \beta_j < \beta_j^0$

- **Estadístico de contraste**:

$$t = \frac{\hat{\beta}_j - \beta_j^0}{\sqrt{\hat{V}(\hat{\beta}_j)}} \underset{H_0}{\sim} t_{N-K}$$

- **Regla de rechazo**:

(a) $t > t_{N-K;\alpha} \Rightarrow RH_0$

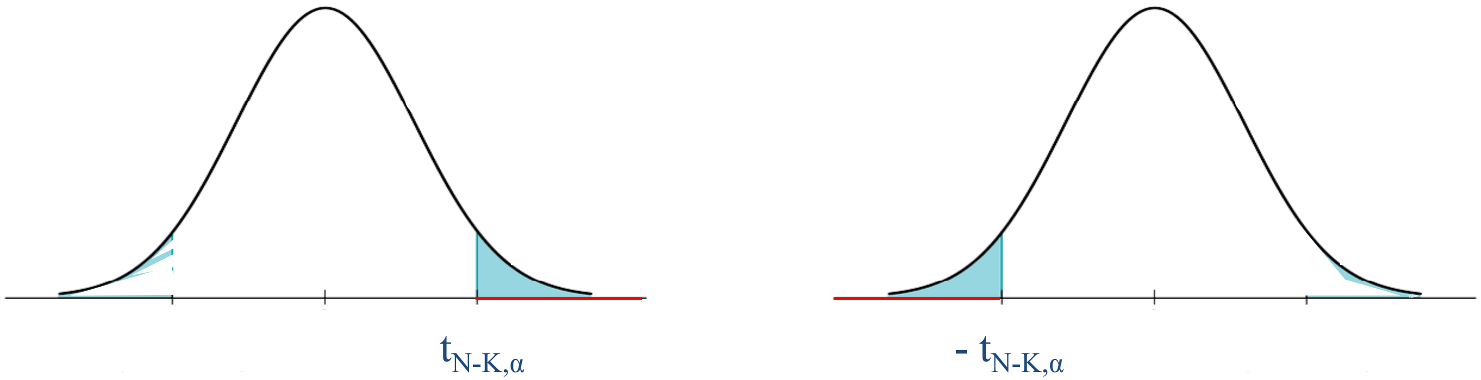
(b) $t < -t_{N-K;\alpha} \Rightarrow RH_0$

- Nota: En el caso (a) la hipótesis nula es en realidad $H_0: \beta_j \leq \beta_j^0$, en el caso (b) la hipótesis nula es en realidad $H_0: \beta_j \geq \beta_j^0$.

UNIVERSIDAD DE MURCIA

- La **región crítica** está en una sola cola y depende de la hipótesis alternativa:

(a) $H_0{:}\beta_j = \beta_j^0 \quad H_1{:}\beta_j > \beta_j^0$ (b) $H_0{:}\beta_j = \beta_j^0 \quad H_1{:}\beta_j < \beta_j^0$



(a) $t_{N\text{-}K;\alpha}$: cuantil $(100\text{-}\alpha)\%$ a la izquierda de una *t*-Student con N-K grados de libertad.

(b) $-t_{N\text{-}K;\alpha}$: cuantil $\alpha\%$ a la izquierda de una una *t*-Student con N-K grados de libertad.

- Un intervalo de confianza (IC) se construye pivotando (o despejando) β_j en las desigualdades de la siguiente expresión:

$$P\left(-t_{N-K,\alpha/2} \leq \frac{\hat{\beta}_j - \beta_j}{\sqrt{\hat{V}(\hat{\beta}_j)}} \leq t_{N-K,\alpha/2} \right) = (100-\alpha)\%$$

- **Intervalo de confianza** al $(100-\alpha)\%$ para β_j: $\hat{\beta}_j \pm t_{N-K;\alpha/2} \sqrt{\hat{V}(\hat{\beta}_j)}$

- **Significado:** Proporciona un intervalo de valores posibles para el parámetro poblacional, con un nivel de confianza de $(100-\alpha)\%$.

- **Significado de nivel de confianza**: Si se calcularan IC con ese método, para un número infinito de muestras, un $(100\text{-}\alpha)\%$ de esos intervalos contendrían el verdadero valor del parámetro.

- **Relación entre IC y contraste de hipótesis**: los IC al $(100\text{-}\alpha)\%$ sirven para realizar contrastes de hipótesis bilaterales de un único parámetro a un nivel de significación $\alpha\%$:

$$H_0: \beta_j = \beta_j^0$$

$$H_1: \beta_j \neq \beta_j^0$$

Regla de rechazo: si β_j^0 no está en el IC $\Rightarrow \mathrm{RH}_0$.

Ejemplo:

$$H_0 : \beta_1 = \beta_2$$
$$H_1 : \beta_1 \neq \beta_2$$

- Equivalentemente:

$$H_0 : \beta_1 - \beta_2 = 0$$
$$H_1 : \beta_1 - \beta_2 \neq 0$$

- Si $\delta = \beta_1 - \beta_2$, entonces:

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

4. Contraste de hipótesis de una combinación lineal de parámetros

UNIVERSIDAD DE MURCIA

- **Estadístico de contraste**: $t=\dfrac{\hat{\delta}}{\sqrt{\hat{V}(\hat{\delta})}} \underset{H_0}{\sim} t_{N-K}$

- **Regla de rechazo**: $|t|>t_{N-K;\alpha/2} \Rightarrow RH_0$

- Una forma útil de calcular $\hat{\delta}$ y $\hat{V}(\hat{\delta})$ es por **reparametrización**:

  Sustituimos $\beta_1=\delta+\beta_2$ en el modelo

$$y_i=\beta_0+\delta x_{1i}+\beta_2\left(x_{1i}+x_{2i}\right)+\ldots+\beta_k x_{ki}+\varepsilon_i$$

  y estimamos para obtener $\hat{\delta}$ y $\sqrt{\hat{V}(\hat{\delta})}$

UNIVERSIDAD DE MURCIA

Ejemplo:

$$H_0: \beta_1 = \beta_2$$
$$\beta_3 = 0$$
$$H_1: \text{no } H_0$$

- Modelo no restringido (MNR): $y_i = \beta_0 + \beta_1 x_{1i} + \ldots + \beta_k x_{ki} + \varepsilon_i$
- Modelo restringido (MR): $y_i = \beta_0 + \beta_1 (x_{1i} + x_{2i}) + \beta_4 x_{4i} + \ldots + \beta_k x_{ki} + \varepsilon_i$

- **Estadístico de contraste**: $F = \dfrac{(SCE_R - SCE_{NR})/q}{SCE_{NR}/(N-K)}$

q : número de restricciones bajo la hipótesis nula.

SCE_{NR} : suma de los cuadrados de los residuos del MNR.

SCE_R : suma de los cuadrados de los residuos del MR.

$(N-K)$: grados de libertad del MNR.

$$F=\frac{\left(SCE_R - SCE_{NR}\right)/q}{SCE_{NR}/\left(N-K\right)}$$

- (SCE_R-SCE_{NR})>0 porque SCE siempre aumenta cuando se imponen restricciones en el modelo $\Rightarrow$ estadístico ≥ 0 , y mide el incremento relativo en la SCE cuando se pasa del MNR al MR.

- **Distribución del estadístico**: $F \underset{H_0}{\sim} F_{q,N-K}$

- **Regla de rechazo**: $F > F_{q,N-K;\alpha} \Rightarrow RH_0$

El **estadístico *F*** puede utilizarse para realizar **contrastes bilaterales de una restricción** (q=1).

- Si q=1 $\Rightarrow$ $t^2_{N\text{-}K} = F_{1,N\text{-}K}$
- Los dos métodos llevan al mismo resultado.
- El **estadístico t es más flexible** porque se puede utilizar también si la alternativa es unilateral.

Ejemplo: $H_0: \beta_1 = \beta_2 = \beta_3 = 0$

$H_1: \text{no } H_0$

- **Estadístico de contraste**: $F = \dfrac{(SCE_R - SCE_{NR})/q}{SCE_{NR}/(N-K)} \underset{H_0}{\sim} F_{q,N-K}$

MNR: $y_i = \beta_0 + \beta_1 x_{1i} + \ldots + \beta_k x_{ki} + \varepsilon_i$

MR: $y_i = \beta_0 + \beta_4 x_{4i} + \ldots + \beta_k x_{ki} + \varepsilon_i$

- **Regla de rechazo:** $F > F_{q,N-K;\alpha} \Rightarrow RH_0$
- $RH_0 \Rightarrow \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3$ son conjuntamente significativos. Significa que, según la información muestral, **al menos una** de las variables x_1, x_2 ó x_3 tiene un efecto parcial no nulo sobre y.
- $\text{no } RH_0 \Rightarrow \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3$ son conjuntamente no significativos. Significa que, según la información muestral, **ninguna** de las variables x_1, x_2 ó x_3 tiene un efecto parcial no nulo sobre y.

Caso particular: **Contraste de significatividad conjunta**.

$$H_0: \beta_1 = \beta_2 = \ldots = \beta_k = 0$$
$$H_1: \text{no } H_0$$

- MR: $y_i = \beta_0 + \varepsilon_i$
- **Estadístico de contraste**: $F = \dfrac{\left(SCE_R - SCE_{NR}\right)/k}{SCE_{NR}/\left(N-K\right)} \underset{H_0}{\sim} F_{k,N-K}$
- **Regla de rechazo**: $F > F_{k,N-K;\alpha} \Rightarrow RH_0$

UNIVERSIDAD DE MURCIA

- Si hay multicolinealidad fuerte $\Rightarrow \mathrm{Var}(\hat{\beta})$ grande

$\Rightarrow$ estadísticos t pequeños $\Rightarrow$ mayor probabilidad de no RH_0

$\Rightarrow$ contrastes de significatividad individual RH_0 con menor frecuencia

$\Rightarrow$ pueden aparecer variables no relevantes cuando en realidad sí lo son.

- Si hay multicolinealidad fuerte entre un grupo de variables, los contrastes de exclusión de la F son más apropiados para contrastar la significatividad conjunta de ese grupo de variables.

- **Indicio muy claro de multicolinealidad** entre un grupo de variables: el estadístico F rechaza la hipótesis nula de significatividad conjunta, pero los estadísticos t no rechazan la hipótesis nula de significatividad individual para ninguna variable $\Rightarrow$ **contradicción**

$$\left.\begin{array}{l} H_0: \beta_1 = \beta_2 = 0 \\ H_1: \text{no } H_0 \end{array}\right\} F > F_{q,N-K;\alpha} \Rightarrow RH_0$$

$$\left.\begin{array}{l} H_0: \beta_1 = 0 \\ H_1: \beta_1 \neq 0 \end{array}\right\} |t| < t_{N-K;\alpha/2} \Rightarrow \text{no } RH_0 \quad \text{y} \quad \left.\begin{array}{l} H_0: \beta_2 = 0 \\ H_1: \beta_2 \neq 0 \end{array}\right\} |t| < t_{N-K;\alpha/2} \Rightarrow \text{no } RH_0$$

- **Regla práctica** para evitar la multicolinealidad:
  - eliminar una o más variables independientes $\Rightarrow$ ¡ojo! sesgo por quitar relevante
  - Si las variables que se eliminan tienen estadístico $|t| < 1$, el estimador resultante puede ser preferible en términos de precisión: la menor varianza compensa el sesgo generado por eliminación de relevante

- Bajo RLM.1-RLM.5 los estimadores MCO se distribuyen aproximadamente como una normal para muestras suficientemente grandes:

$$\frac{\hat{\beta}_j - \beta_j}{\sqrt{\hat{V}(\hat{\beta}_j)}} \approx N(0,1)$$

- Esto implica que los contrastes basados en la *t* y la *F* siguen aproximadamente distribuciones *t* y *F* en muestras grandes, aunque no se cumpla RLM.6.

UNIVERSIDAD DE MURCIA

- **Objetivo**: Predecir el valor de y asociado a determinados valores de las variables explicativas no muy alejados de los valores muestrales.

- $x_1^0, x_2^0, \ldots, x_k^0$ son los valores de las variables explicativas para los que hay que predecir el valor de y.

- El modelo dice: $y^0 = \beta_0 + \beta_1 x_1^0 + \ldots + \beta_k x_k^0 + \varepsilon^0$

  $\Rightarrow E[y^0 | x_1^0, \ldots, x_k^0] = \beta_0 + \beta_1 x_1^0 + \ldots + \beta_k x_k^0$ es el valor esperado de y^0

- No conocemos $\beta_0, \beta_1, \ldots, \beta_k$ sino sus estimaciones

  $\Rightarrow \hat{y}^0 = \hat{\beta}_0 + \hat{\beta}_1 x_1^0 + \ldots + \hat{\beta}_k x_k^0$ es el valor predicho de y^0

- **Bajo RLM.1-RLM.5** $\hat{y}^0$ es el predictor óptimo de y^0: es la predicción más precisa que podemos obtener usando la información disponible.

UNIVERSIDAD DE MURCIA

- El intervalo de predicción (IP) proporciona una medida de la incertidumbre asociada a la predicción.

- El IP permite hacer inferencia.

- **Error de predicción:** $e^0 = y^0 - \hat{y}^0 = (\beta_0 + \beta_1 x_1^0 + \ldots + \beta_k x_k^0) + \varepsilon^0 - \hat{y}^0$

$$E\left(e^0/X^P\right) \overset{\text{RLM.1-RLM.3}}{=} 0$$

- $$V\left(e^0/X^P\right) = V(\varepsilon^0/X^P) + V(\hat{y}^0/X^P) \overset{\text{RLM.4}}{=} \sigma^2 + V(\hat{y}^0/X^P)$$

  donde X^P contiene tanto los valores de las variables explicativas en la muestra, como los valores de las variables explicativas para los que hay que predecir el valor de y.

- ¿Cómo estimamos $V(e^0/X^P)$?

UNIVERSIDAD DE MURCIA

- ¿Cómo estimamos $V(e^0/X^P)$?

$$\hat{V}\left(e^0/X^P\right) = s^2 + \hat{V}(\hat{y}^0/X^P)$$

- Estimación de $\hat{V}(\hat{y}^0/X^P)$:

$$E[y^0 \mid x_1^0,\dots,x_k^0] = \theta = \beta_0 + \beta_1 x_1^0 + \dots + \beta_k x_k^0$$

$$\Rightarrow \beta_0 = \theta - \beta_1 x_1^0 - \dots - \beta_k x_k^0$$

- Sustituyendo esta expresión en el modelo, obtenemos el modelo reparametrizado:

$$y = \theta + \beta_1 (x_1 - x_1^0) + \dots + \beta_k (x_k - x_k^0) + \varepsilon$$

- Estimamos el modelo reparametrizado y obtenemos $\hat{V}(\hat{\theta}/X^P)$

$$\Rightarrow \quad \hat{V}(\hat{y}^0/X^P) = \hat{V}(\hat{\theta}/X^P)$$

UNIVERSIDAD DE MURCIA

- Distribución de e^0

$$e^0/X^P \sim N(0,V(e^0/X^P)) \Rightarrow \frac{\overbrace{y^0-\hat{y}^0}^{e^0}}{\sqrt{V(e^0/X^P)}} \sim N(0,1) \Rightarrow \frac{\overbrace{y^0-\hat{y}^0}^{e^0}}{\sqrt{\hat{V}(e^0/X^P)}} \sim t_{N-K}$$

por ser función lineal de $\hat{\beta}_j$ y ε^0 que son Normales

- **Intervalo de predicción:** IC $(100-\alpha)\%$ para y^0

$$\hat{y}^0 \pm t_{N-K;\alpha/2}\sqrt{\hat{V}(e^0/X^P)}$$

UNIVERSIDAD DE MURCIA

Lo que hemos aprendido:

- Bajo RLM1-RLM6, los estimadores MCO siguen una distribución normal.
- Bajo RLM1-RLM6, los estadísticos *t* y *F* tienen distribuciones *t*-Student y *F* de Snedecor bajo la hipótesis nula.
- El estadístico *t* se usa para contrastar hipótesis de una única restricción frente a alternativas unilaterales o bilaterales.
- El estadístico *F* se usa para contrastar hipótesis de una o más restricciones frente a alternativas bilaterales.
- Bajo RLM1-RLM6, se construyen IC para cada β_j. Los IC se pueden usar para contrastar hipótesis sobre β_j frente a una alternativa bilateral.

Lo que hemos aprendido:

UNIVERSIDAD DE MURCIA

- El estadístico F para la significatividad conjunta contrasta la hipótesis nula de que todos los parámetros, excepto el término constante, son 0.
- Los métodos para realizar contrastes y construir IC son aproximadamente válidos sin el supuesto RLM6.
- Con la predicción puntual se obtiene un único valor de la predicción pero no se está dando información sobre la incertidumbre que comporta esta predicción.
- La predicción por intervalos proporciona una medida de la probabilidad de que dicho intervalo contenga el valor que se desea predecir.