Econometría

Diplomado Banco Central de Honduras

Instituto de Economía

Pontificia Universidad Católica de Chile

Juan Ignacio Urquiza – Junio 2022

Repaso de Estadística

- Propiedades de los Estimadores:
  - Insesgamiento, eficiencia, consistencia
  - Teorema Central del Límite
- Inferencia:
  - Test de hipótesis
  - Intervalo de confianza

Apéndice C – Wooldridge

Estimadores como VA

- Un estimador *t* es una variable aleatoria que es función de los datos muestrales.
- Para evaluar a un estimador debemos analizar distintas propiedades de su distribución de probabilidades.
- Comenzaremos con el estudio de propiedades de muestra pequeña o finita, y luego analizaremos propiedades asintóticas que tienen que ver con el comportamiento de los estimadores a medida que aumenta el tamaño de la muestra.

Propiedades de Estimadores

- Muestras Repetidas de Igual Tamaño
  - <u>Insesgamiento</u>:
    - Sea $t(x)$ estimador de un parámetro θ, y sea $f(t,\theta)$ la función de densidad de probabilidad de t.
    - Entonces, t es insesgado si $E(t) = \theta$.
    - En palabras, el valor esperado del estimador coincide con el del parámetro a estimar.
    - <u>*Importante*</u>: el insesgamiento es una propiedad para todo tamaño de muestra; no sólo para las grandes.

Media Muestral

- Sea Y_1, Y_2, …, Y_n una muestra aleatoria de una población con media μ_Y.
- Entonces,

$$\bar{Y} = \frac{1}{n}(Y_1 + Y_2 + \cdots + Y_n) = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

- <u>Valor Esperado:</u>

$$E(\bar{Y}) = E\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(Y_i)$$

$$= \frac{1}{n}\sum_{i=1}^{n} \mu_Y = \mu_Y$$

Insesgamiento

Propiedades de Estimadores

- Muestras Repetidas de Igual Tamaño

  - <u>Eficiencia Relativa</u>:

    - Sean t_1 y t_2 estimadores *insesgados*.
    - Entonces, t_1 es más eficiente que t_2 si Var(t1)<Var(t2).
    - En palabras, el estimador t_1 es más eficiente si su varianza es más pequeña.

Propiedades de Estimadores

- Muestras Repetidas de Igual Tamaño
  - <u>Eficiencia Relativa</u>:
    - Sean t_1 y t_2 estimadores *sesgados*.
    - Entonces, t_1 es más eficiente que t_2 si $ECM_1 < ECM_2$.
    - $ECM_i = E[(t_i - \theta)^2]$: Error Cuadrático Medio.

$$\boxed{\mathbf{ECM = Var(}t\mathbf{) + sesgo^2}}$$

Muestras Grandes

- Se denomina teoría asintótica al estudio de los estimadores para muestras grandes ($n \longrightarrow \infty$).

- Se revisarán dos temas:
  - Convergencia en probabilidad.
    - Consistencia
  - Convergencia en distribución.
    - Teorema Central del Límite

Convergencia en Probabilidad

- **Definición:**
  - Una secuencia de variables aleatorias $\{x_n\}$ converge en probabilidad a una constante c si

$$\lim_{n\to\infty} \Pr(|x_n - c| > \varepsilon) = 0$$

para cualquier $\varepsilon > 0.$

  - Alternativamente,

$$\lim_{n\to\infty} \Pr(|x_n - c| < \varepsilon) = 1$$

  - *Intuición*: a medida que aumenta el tamaño de la muestra, la distribución de x_n se concentra cada vez más alrededor de la constante c.

Convergencia en Probabilidad

- En general, si x_n converge en probabilidad a una constante c, entonces escribimos:

$$\text{plim } x_n = c$$

- <u>*Propiedades del operador plim*</u>:
  - $\text{plim } (x + y) = \text{plim } x + \text{plim } y$
  - $\text{plim } (x * y) = \text{plim } x * \text{plim } y$
  - $\text{plim } (x / y) = \text{plim } x / \text{plim } y$, si $\text{plim } y \neq 0$

Propiedades de Estimadores

- <u>Consistencia</u>:
  - El estimador $t(x)$ es consistente para θ si:

$$\lim_{n\to\infty} \Pr(|t-\theta| > \varepsilon) = 0$$

  - Es decir, *t* es consistente si converge en probabilidad al parámetro θ.

  - <u>*Nota*</u>: si $E(t_n) = \theta$ y $V(t_n) \to 0$ cuando $n \to \infty$, entonces t_n es consistente.

Media Muestral

- **Varianza:**

$$\mathrm{var}(\bar{Y}) = E[\bar{Y} - E(\bar{Y})]^2$$

$$= E[\bar{Y} - \mu_Y]^2$$

$$= E\left[\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right) - \mu_Y\right]^2 \quad = E\left[\frac{1}{n}\sum_{i=1}^{n}(Y_i - \mu_Y)\right]^2$$

$$= E\left\{\left[\frac{1}{n}\sum_{i=1}^{n}(Y_i - \mu_Y)\right] \times \left[\frac{1}{n}\sum_{j=1}^{n}(Y_j - \mu_Y)\right]\right\}$$

Media Muestral

- <u>Varianza</u>:

$$\operatorname{var}(\bar{Y}) = E\left\{\left[\frac{1}{n}\sum_{i=1}^{n}(Y_i - \mu_Y)\right]\times\left[\frac{1}{n}\sum_{j=1}^{n}(Y_j - \mu_Y)\right]\right\}$$

$$= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}E\left[(Y_i - \mu_Y)(Y_j - \mu_Y)\right]$$

$$= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\operatorname{cov}(Y_i, Y_j)$$

$$= \frac{1}{n^2}\sum_{i=1}^{n}\sigma_Y^2 \quad = \frac{\sigma_Y^2}{n}$$

Consistencia

Convergencia en Distribución

- Teorema Central del Límite:
  - Definición: Sea Y_1, Y_2, …, Y_n una muestra aleatoria de una población con media μ y varianza σ^2. Entonces,

$$\sqrt{n}(\bar{Y}_n - \mu) \xrightarrow{d} N(0, \sigma^2)$$

    donde *d* significa que la expresión converge en distribución a una Normal cuando *n* tiende a infinito.
  - Alternativamente,

$$Z_n = \frac{\bar{Y}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} N(0,1)$$

Test de Hipótesis

- Esta técnica de inferencia estadística nos permite evaluar si la información contenida en los datos avala o no una conjetura sobre la población que se está estudiando.
- Nociones básicas:
  - ***Hipótesis nula e hipótesis alternativa***
  - ***Estadístico de contraste***
  - ***Nivel de significancia***
  - ***Valor crítico***
  - ***Región de aceptación/rechazo***

Planteamiento de Hipótesis

- El punto de partida consiste en establecer la hipótesis a evaluar; es decir, la hipótesis nula (H_0), cuya validez queremos contrastar.
- Frente a esta hipótesis, se plantea una hipótesis alternativa (H_1) que se cumple cuando la nula es falsa.
- Por ejemplo:
  - H_0: $E(Y) = \mu_{Y,0}$
  - H_1: $E(Y) \neq \mu_{Y,0}$ (*test bilateral o "de dos colas"*)
  - $H_1 : E(Y) > \mu_{Y,0}$ o $\mathrm{H}_1 : E(Y) < \mu_{Y,0}$

    (*test unilateral o "de una cola"*)

Estadístico de Contraste

- Una vez definidas H_0 y H_1, hay que determinar un estadístico de contraste que mida la discrepancia entre la información muestral y la hipótesis nula.
  - Idea: si la discrepancia es grande, el valor del estadístico estará dentro de valores poco probables cuando H_0 sea verdadera, lo que indica evidencia contraria a H_0, y viceversa.
- El estadístico de contraste es una variable aleatoria puesto que depende de los datos muestrales.
- Deberá tener una distribución conocida (exacta o aproximada) bajo la hipótesis nula.

Error Tipo I / Error Tipo II

- Cuando realizamos un contraste, podemos cometer dos tipos de errores:
  - Podemos rechazar H_0 cuando es cierta (*error tipo I*)
  - Podemos no rechazar H_0 cuando es falsa (*error tipo II*)
- En la práctica, no es posible determinar si hemos cometido o no un error.
- Sin embargo, es posible especificar la probabilidad α de cometer un error del tipo I.

$$\alpha = \Pr(\text{error tipo I}) = \Pr(\text{rechazar } H_0 \mid H_0 \text{ verdadero})$$

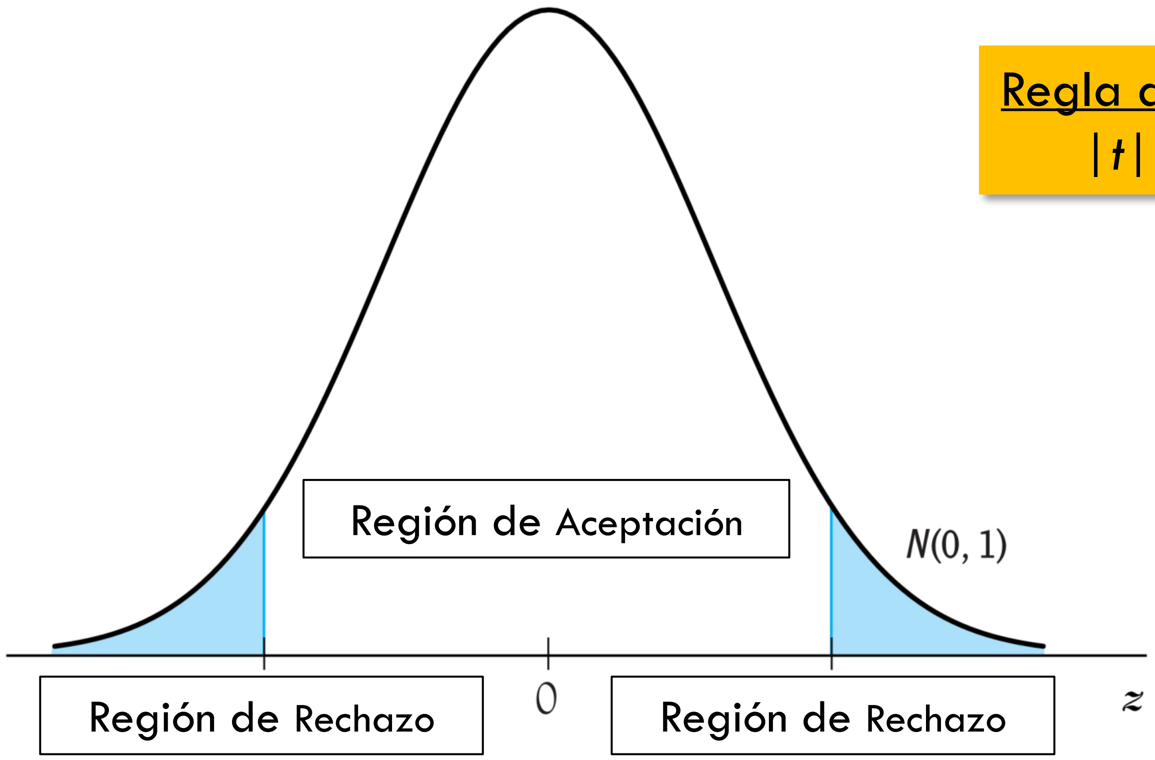
Nivel de Significancia o Tamaño

Valor Crítico / Región de Rechazo

- Entonces, se fija un valor para α tan pequeño como uno desee – en general: $\alpha = 0,1;\ 0,05;\ 0,01$.
  - Interpretación:
    - $\alpha = 0,05$ implica que el investigador estaría dispuesto a rechazar H_0 incorrectamente un 5% de las veces – *tolerancia*.
- Una vez establecido el nivel de significancia, es posible hallar un valor crítico contra el cual contrastar el valor del estadístico.
- Los valores del estadístico T que resulten en el rechazo de H_0 conforman la región de rechazo.

Región de Aceptación/Rechazo

- **Test Bilateral:**

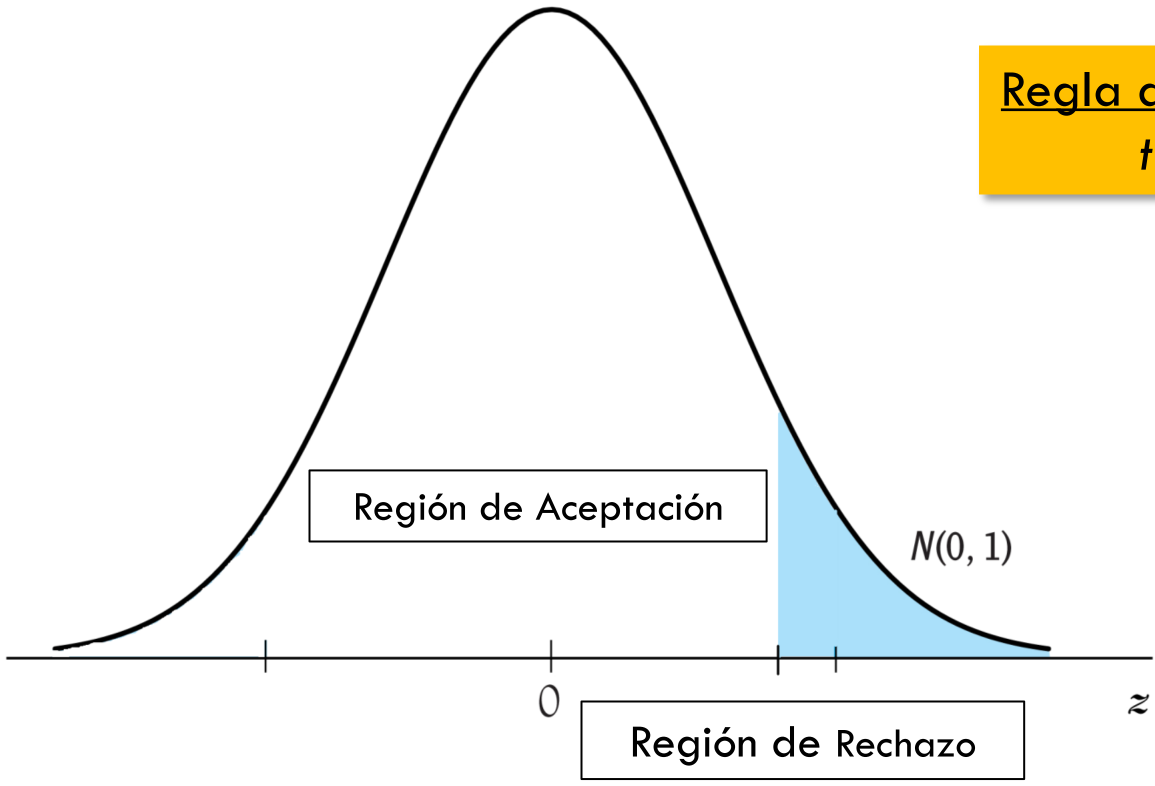


Regla de Decisión:

$|t| > |c|$

Región de Aceptación/Rechazo

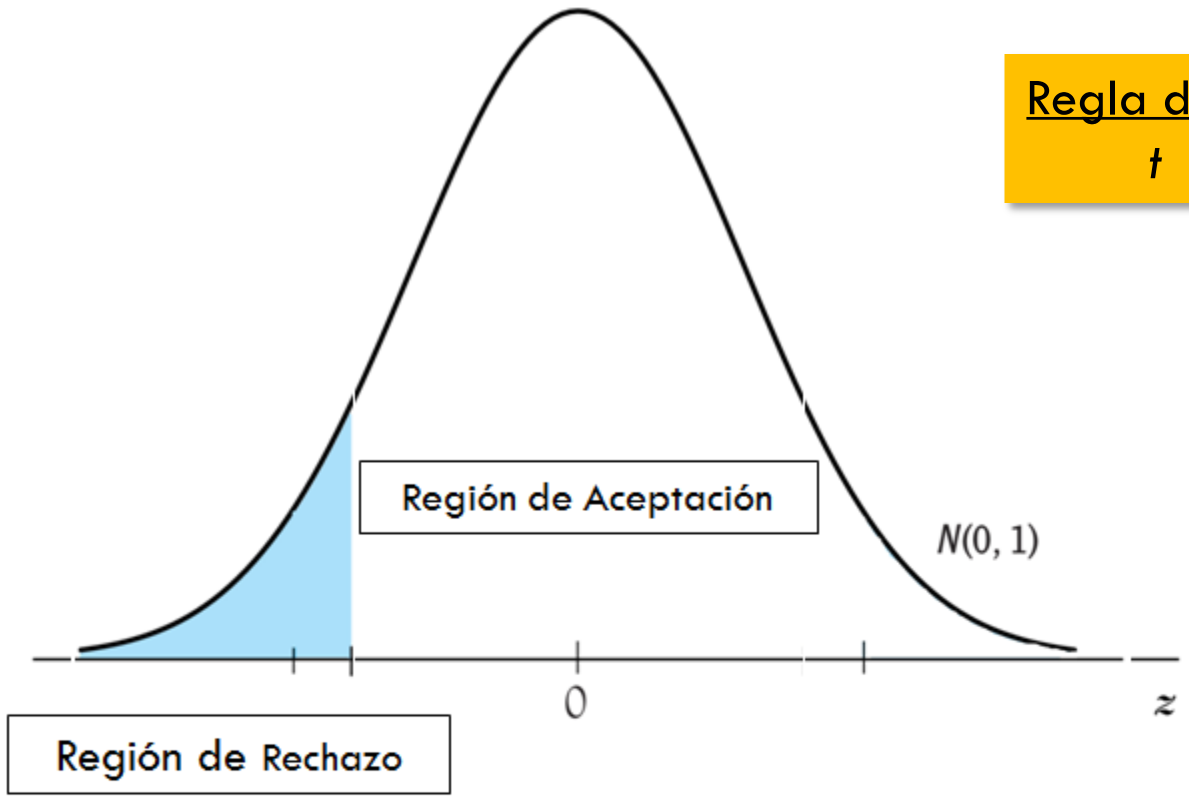
- **Test Unilateral:**



Regla de Decisión:
t > c

Región de Aceptación/Rechazo

- **Test Unilateral:**



Regla de Decisión:

$t < -c$

Test de Hipótesis

- $Y \sim N(\mu_Y, \sigma_Y^2)$ – σ_Y^2 conocido.

Caso 1

- Entonces, bajo H_0:

$$\bar{Y} \sim N\left(\mu_0, \frac{\sigma_Y^2}{n}\right)$$

- Por lo tanto,

$$T = \frac{\bar{Y} - \mu_0}{\sigma_Y / \sqrt{n}} \sim N(0,1)$$

Test de Hipótesis

- $Y \sim N(\mu_Y, \sigma_Y^2)$ – σ_Y^2 desconocido.

Caso 2

- Hay que estimarlo:

$$s_Y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$$

$$\sqrt{s_Y^2/n} \equiv SE(\bar{Y}) = error\ estandar\ de\ \bar{Y}$$

- Por lo tanto,

$$T = \frac{\bar{Y} - \mu_0}{s_Y/\sqrt{n}} \sim t_{n-1}$$

Test de Hipótesis

- ¿Si desconocemos tanto la distribución de *Y* como su varianza σ_Y^2?
- Teorema Central del Límite!
- Por lo tanto,

Caso 3

$$T = \frac{\bar{Y} - \mu_0}{s_Y/\sqrt{n}} \xrightarrow{a} N(0,1)$$

P-value

- Probabilidad de obtener un estadístico tanto o más adverso para la hipótesis nula como el computado, asumiendo que H_0 es verdadera.
  - $P\text{-}value = \Pr(T > t \mid H_0)$
  - $P\text{-}value = \Pr(T < t \mid H_0)$
  - $P\text{-}value = \Pr(\,|T| > |t| \mid H_0) = 2^*\Pr(T > |t| \mid H_0)$
- Cuanto menor sea el *p*-value, menos probable es que la distribución obtenida para el estadístico bajo la nula sea correcta y, por lo tanto, es menos probable que estemos bajo la nula.
- Es decir, un *p*-value chico es evidencia en contra de H_0.

Intervalo de Confianza

- Debido al error de muestreo, es imposible conocer el valor exacto de la media poblacional a partir de la información contenida en los datos muestrales.
- Sin embargo, es posible construir un set de valores que contenga al verdadero valor poblacional para un nivel de probabilidad dado.
- Intervalo de Confianza (test bilateral al 95%):

$$\{\bar{Y} \pm 1.96 * (\sigma_Y/\sqrt{n})\}$$ Caso 1

$$\{\bar{Y} \pm t_{\propto/2}^{n-1} * SE(\bar{Y})\}$$ Caso 2

$$\{\bar{Y} \pm 1.96 * SE(\bar{Y})\}$$ Caso 3