Clase: Tests de Hipótesis

Profesor: Ricardo Pasquini

Facultad de Ciencias Empresariales - Universidad Austral

8 de Abril de 2024

Qué sabemos sobre β poblacional?

- Nuestro interés es poder decir algo sobre β (i.e., el valor del coeficiente poblacional) usando los valores de $\hat{\beta}$ (valores de los coeficientes estimados).
- Propiedades de $\hat{\beta}$:
  - $E[\hat{\beta}] = \beta$ (es *insesgado*)
  - $Var[\hat{\beta}] = \frac{\sigma^2}{\sum(x_i - \bar{x})^2}$, donde $\sigma^2 = Var(\epsilon)$

Qué sabemos sobre β poblacional? (cont.)

- Estimación de la $Var[\hat{\beta}]$:
  - $\hat{\sigma}^2 = \frac{\sum e_i^2}{n-1}$
  - El error estándar: $\sqrt{\hat{Var}[\hat{\beta}]}$

Podemos decir algo más, por ejemplo, sobre un valor puntual de β en la población?

- Test de Hipótesis:
  - Entender la distribución de $\hat{\beta}$
  - Lógica del test e implementación

La distribución de $\hat{\beta}$

- Métodos para obtener la distribución de $\hat{\beta}$:
  1. Supuesto sobre la distribución del error
  2. Teoría asintótica (muestra grande)
  3. Métodos computacionales (Bootstrap)

El supuesto de normalidad

- Supuesto de normalidad del error.
- Implicaciones para la distribución de $\hat{\beta}$:
  - $\hat{\beta} \sim N(\beta, Var(\beta))$
  - Estándarización de $\hat{\beta}$: $Z = \frac{\hat{\beta}-\beta}{\sqrt{Var(\beta)}} \sim N(0,1)$

La lógica del Test de Hipótesis

- Lógica del Test de Hipótesis:
  1. Supuestos y distribución de $\hat{\beta}$
  2. Obtención del estimador y verificación de supuestos
  3. Rechazo del supuesto si se observa un valor improbable

El supuesto de normalidad

- El primer método consiste en hacer un supuesto sobre la distribución del error. Supondremos que se distribuye normalmente.
- Si el error se distribuye normalmente entonces $\hat{\beta}$ se distribuye normalmente con centro en el valor esperado de beta, y varianza dada por la varianza de beta.

$$\hat{\beta} \sim N(\beta, Var(\beta))$$

- Por las propiedades de la Normal, sabemos que el valor estandarizado del coeficiente, sigue la distribución *Normal Estándar*

$$Z = \frac{\hat{\beta} - \beta}{\sqrt{Var(\beta)}} \sim N(0, 1)$$

- Es decir, podríamos usar la distribución normal estándar como base para entender el comportamiento del estimador.. pero hay un problema... hay dos cosas que necesitamos para estandarizar el coeficiente que no conocemos:
  - 1) el valor poblacional de beta, y
  - 2). σ^2 (la varianza del error), que como vimos más arriba es

El supuesto de normalidad (cont.)

- Afortunadamente podemos solucionar los dos problemas:
  - 1) El valor poblacional de beta va a ser un supuesto que realizaremos usando la misma lógica del test de hipótesis (ver más adelante).
  - 2) Más arriba ya vimos que podíamos estimar σ^2 en base a los datos.
    - Este último procedimiento, sin embargo, no es inocuo. Se puede demostrar (no lo hacemos) que cuando usamos el valor estimado, el coeficiente de beta ahora se distribuye T-student con n-1 grados de libertad:

$$T = \frac{\hat{\beta} - \beta}{\sqrt{\hat{Var}(\beta)}} \sim T_{n-1}$$

La lógica del Test de Hipótesis

- La lógica del Test de Hipótesis es la siguiente:
  1. Voy a realizar una serie de supuestos que me permitirán arribar a una distribución para el estimador. Es decir, me permitirán decir con qué probabilidad espero observar cada valor del estimador.
  2. Voy a ir a los datos, obtener el estimador, y ver si esa evidencia me permite rechazar los supuestos.
     - Todo lo que voy a poder hacer con el test es rechazar (o no rechazar) un supuesto, en particular, un supuesto que me sea de interés de investigación. Es decir, nunca voy a concluir que el supuesto es válido, sino, a lo sumo, que no tengo evidencia para rechazarlo.
  3. Pero cuando rechazamos los supuestos? La idea es rechazar si surge un valor del estimador que (de acuerdo con los supuestos) surge con muy baja probabilidad.

La lógica del Test de Hipótesis (cont.)

- En este contexto, el supuesto que voy a hacer es un valor determinado para β (en la población).
  - Por ejemplo, típicamente de interés es probar el supuesto de que $\beta = 0$.
    - La razón por la que este test es típicamente de interés es porque, si fuera válido, implicaría que, de acuerdo a nuestro modelo, X no tiene efecto sobre Y. Y en muchas circunstancias de investigación queremos saber si hay o no hay efecto!
    - Denotamos este test como $H_0 : \beta = 0$ versus $H_a : \beta \neq 0$.

La lógica del Test de Hipótesis (cont.)

- Pero también me podrían interesar otros valores ($\beta = 1$ en el caso de que el modelo esté probando un efecto tipo elasticidad - ver ejemplos práctica-).
- Una vez que hice ese supuesto, casi que cuento con una distribución para el estimador $\hat{\beta}$. Estrictamente cuento con una distribución para el valor de $\hat{\beta}$ transformado, T, que me servirá para lo mismo. Veamos:
  - Para simplificar la exposición supongamos que queremos testear $H_0 : \beta = 0$. Notemos que:
    - Por lo que dijimos anteriormente sabemos que T tiene distribución T-student:

$$\hat{T} = \frac{\hat{\beta} - 0}{\sqrt{\hat{Var}(\beta)}} \sim T_{n-1}$$

La lógica del Test de Hipótesis (cont.)

- ¿Cuándo rechazar el supuesto? La idea que propone el Test es la siguiente: Vamos a rechazar el supuesto, si observamos un valor de $\hat{T}$ que ocurriría con muy poca probabilidad.
  - Veamos un ejemplo:
    - Supongamos que $\beta = 0$. Si $\beta = 0$ entonces lo que espero encontrar $\hat{\beta}$ muy cercano a 0 con alta probabilidad. Siguiendo la definición de $\hat{T}$ esto también implica que espero encontrar T cercano a 0.

La lógica del Test de Hipótesis (cont.)

- Supongamos adicionalmente que encontramos un valor de $\hat{\beta} = 10$ y además medimos que $\sqrt{Va\hat{r}(\beta)} = 2$, con $n = 100$. Es decir $\hat{T} = 5$. ¿Cuán probable es encontrar $\hat{T} = 5$ en una distribución T-student con 99 grados de libertad? La respuesta rápida (ustedes pueden buscar ese valor) es que es muy improbable (ocurre con mucho menos del 1

La lógica del Test de Hipótesis (cont.)

- En general vamos a utilizar dos criterios para decidir cuándo un resultado es improbable:
  1. Estableciendo una región de rechazo. Vamos a identificar en la distribución de T cuales son los valores a partir de los cuales ocurren resultados con menos de un *nivel de significancia* (por ejemplo, un valor usual es rechazar si el valor ocurre con menos de un 5
  2. Midiendo la probabilidad de ocurrencia del valor que obtuvimos para el valor que efectivamente medimos ($\hat{T}$) en la distribución. A esto se lo conoce como **P-valor** o **P-value**.

La lógica del Test de Hipótesis (cont.)

- Como la distribución de T es contínua lo que hacemos es medir la probabilidad de obtener un valor mayor a T (y si es un test a dos-colas miramos la probabilidad de obtener un valor mayor a $\hat{T}$ y menor a -$\hat{T}$).
  - Por esta razón a veces encontrarán el P-valor denotado como:
    - $P(\hat{T} > T)$ o $P(|\hat{T} > T|)$.

La lógica del Test de Hipótesis (cont.)

- Si el P-valor es muy bajo (por ejemplo menor al 5
- Para ver ejemplos gráficos de la distribución T, de las regiones de rechazo y de las mediciones de P-valor, no dejen de leer el capítulo 4 del libro de Wooldridge.