ˇˇ

Econometría I - EAE- 250-A

Prediccion

Ezequiel Garcia-Lembergman

Instituto de Economía - Pontificia Universidad Católica de Chile

Introducción

- Hoy vamos a discutir como hacer predicciones con intervalo de confianza.
  - ¿Con qué probabilidad puedo descartar que una acción caerá a un precio menor a 5 dado sus características?
  - ¿Con qué probabilidad puedo estar seguro que una persona con determinadas caracteristicas re-pagara su deuda?

Intervalos de Confianza para la predicción

- Suponga que se obtiene los estimadores MCO, $\hat{\beta}$ y que se cumplen los supuestos clasicos para una muestra de n observaciones.
- Dado los $\hat{\beta}$ y valores para las variables x (x_τ), es facil construir la prediccion $\hat{y_\tau}$.
- Queremos obtener intervalos de confianza para una predicción a partir de la línea de regresión de MCO
- Esto me va a permitir decir no solo la prediccion para un individuo en particular, sino tambien entre que rangos esta la verdadera variable de ese individuo con cierto nivel de confianza.
- Por ejemplo, si usted tiene que determinar si darle un credito o no a Sofia, estas herramientas van a permitirle decir: dadas las caracteristicas de Sofia, la prediccion es que va a repagar el credito con un 70% de probabilidad. Ademas, con 95% de confianza la probabilidad va a estar entre 50% y 80%.

Ejemplo: Prediccion de Notas en la universidad

- Suponga que el director de una universidad esta decidiendo que alumnos va a admitir el siguiente anio. Para ello quiere predecir el promedio que los aplicantes obtendran en la universidad (entre 0 y 4). En base a datos de ex alumnos de la universidad, el director estima el siguiente modelo:

$$NotaUniversidad_i = \beta_0 + \beta_1 SAT_i + 0.40 NotaHighSchool_i + u$$

  Obtiene por MCO:
- donde SAT es el puntaje en la SAT que va de 0 a 1500).
- Al estimar por MCO obtiene:

$$NotaUni\hat{v}ersidad = 1 + 0.00149 SAT + 0.40 NotaHighSchool$$

- Una aplicante, Sofia, tiene $SAT = 900$ y su promedio de notas en el high school era de 3. Su vector de x's es: $x_\tau = (1, 900, 3)$. Reemplazando en la prediccion, predecimos que va a obtener nota 3 en la universidad.

$$NotaUni\hat{v}ersidad_\tau = 3$$

Prediccion individual y prediccion promedio

- Hoy vamos a construir intervalos de confianza para dos tipos de predicciones:

  1. **Prediccion individual:** busca predecir el valor de y para un individuo en particular. Para el individuo τ, Predecir: $y_\tau = x_\tau^\top \beta + u_\tau$.
     $\rightarrow$ Ejemplo: el y predicho para un individuo con caracteristicas $x_\tau^\top$.

  2. **Prediccion media o promedio:** busca predecir el valor promedio de y: Predecir: $E[y_\tau | X] = x_\tau^\top \beta$.
     $\rightarrow$ Ejemplo, el promedio de y para los individuos con caracteristicas $x_\tau^\top$

- Ambas alternativas dan lugar a la misma predicción ($\hat{y}_\tau$), pero diferentes intervalos de confianza, ya que difieren en la varianza del error de predicción.

Predicción individual: intervalos de confianza

- Definiciones:
  - y_τ el valor para el cual se desea construir un intervalo de confianza (e.g: τ es un individuo que no esta en la muestra).
  - sea $x_{\tau,1}, \ldots, x_{\tau,k}$ los nuevos valores de las variables independientes
  - u_τ el error no observado para el individuo τ.
- Usando los $\hat{\beta}$ de nuestra regresion MCO original, el mejor estimador (MELI) de y_τ es

$$\hat{y}_\tau = \hat{\beta}_0 + \hat{\beta}_1 x_{\tau,1} + \ldots + \hat{\beta}_k x_{\tau,k} = x_\tau^\top \hat{\beta}$$

- En la practica esto es bien sencillo. ‘’MCO me dio los estimadores $\hat{\beta}$, asi que decime los valores para la variable independiente, reemplazo y te doy mi prediccion”.
- Pero, queremos construir los intervalos de confianza para esa prediccion.

Predicción individual

- Definiendo el error de prediccion como $\hat{e}_\tau = y - y_\tau$.
- El intervalo para y_τ es

$$\hat{y}_\tau - t_{\alpha/2} s.e.(\hat{e}_\tau) \leq y_\tau \leq \hat{y}_\tau + t_{\alpha/2} s.e.(\hat{e}_\tau)$$

donde

$$s.e.(\hat{e}_\tau) = \sqrt{\hat{\sigma}^2 (x_\tau^\top (X^\top X)^{-1} x_\tau + 1)}$$

- Noten que los dos limites son calculables con los datos. Es decir, me permite decir: para un individuo con caracteristicas x_τ, predigo el valor $\hat{y}_\tau$. Ademas, con $1 - \alpha\%$ de confianza, el valor verdadero y_τ se encuentra entre los limites.

Ejemplo: Prediccion de Notas en la universidad

- Suponga que el director de una universidad esta decidiendo que alumnos va a admitir el siguiente anio. Para ello quiere predecir el GPA que los aplicantes obtendran en la universidad. En base a datos de ex alumnos de la universidad, el director hace la siguiente regresion MCO.

$$\widehat{NotaUniversidad}_i = 1 + 0.00149 SAT + 0.40 NotaHighSchool_i$$

- donde SAT es el puntaje en la SAT (de 0 a 1500).
- Una aplicante, Sofia, tiene $SAT = 900$ y su GPA en el high school era de 3. Eso les da una prediccion de la nota en la universidad de $\hat{y}_{sofia} = 3$.
- Suponga que, ademas, cuenta con los siguientes datos: $\sqrt{\hat{\sigma}^2(x_\tau^\top (X^\top X)^{-1} x_\tau) + 1)} = 0.204$
- Construya el intervalo de confianza para $\alpha = 5\%$.
- Nota de sofia en la universidad: $2.6 < NotaUniversidad_{sofia} < 3.4$.
- Entonces, puede concluir que la prediccion para Sofia es 3. Ademas, con 95% de confianza Sofia obtendra un GPA en la universidad entre 2.6 y 3.4.

Predicción individual en Modelo Simple

- Se puede demostrar que en el modelo simple de una variable explicativa:

$$\hat{y}_{\tau} - t_{\alpha/2} s.e.(\hat{e}_{\tau}) \leq y_{\tau} \leq \hat{y}_{\tau} + t_{\alpha/2} s.e.(\hat{e}_{\tau})$$

con $s.e.(\hat{e}_{\tau}) = \sqrt{\hat{\sigma}^2 \left(\frac{(x_{\tau 1} - \overline{x})^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2} + \frac{1}{n} + 1 \right)}$.

- Esto implica que la incertidumbre de la predicción es menor cuando:
  - $x_{\tau 1}$ está cerca de $\overline{x}$
  - El tamaño de la muestra es más grande
  - El valor de σ^2 es menor

Predicción promedio

- Ahora se busca encontrar un intervalo de confianza para la persona promedio de la poblacion.
- Como antes, el mejor estimador (MELI) de $E[y_\tau|X]$ también es

$$\hat{y}_\tau = x_\tau^\top \hat{\beta}$$

- Sea ϵ_τ el error de predicción

$$\epsilon_\tau = E[y_\tau|X] - \hat{y}_\tau = x_\tau^\top(\beta - \hat{\beta})$$

- Se puede demostar que:

$$Var[\epsilon_\tau|X] = \sigma^2 x_\tau^\top (X^\top X)^{-1} x_\tau$$

$\rightarrow$ Noten que ahora la varianza sera menor porque no esta el error no observable u.

- El intervalo de confianza para $E[y_\tau|X]$ al $1-\alpha$ es

$$\hat{y}_\tau - t_{\alpha/2} s.e.(\hat{\epsilon}_\tau) \leq E[y_\tau|X] \leq \hat{y}_\tau + t_{\alpha/2} s.e.(\hat{\epsilon}_\tau)$$

donde $s.e.(\hat{\epsilon}_\tau) = \sqrt{\hat{\sigma}^2 x_\tau^\top (X^\top X)^{-1} x_\tau}$.

Conclusiones

- Podemos construir predicciones y sus intervalos de confianza (que difieren dependiendo de si son individuales o promedios)
- Mientrás más poder explicativo tiene nuestro modelo, más precisas serán las predicciones.