Tema 2 – Regresión lineal : estimación

- $Y = \underbrace{\beta_0 + \beta_1 X_1 + \dots + \beta_k X_k}_{\text{función de regresión poblacional (FRP)}} + \varepsilon$

$$\text{función de regresión poblacional (FRP)} = \mathbb{E}[Y \mid X_1, \dots, X_k]$$

Y = var. dependiente / endógena
X_i = var. indep. / exógena

- Residuo de la regresión : $\hat{\varepsilon}_i = Y_i - \hat{Y}_i = Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_{1i})$

- método de los mínimos cuadrados ordinarios (MCO)

$$= \min_{\beta_0, \beta_1} \sum_{i=1}^{n} \hat{\varepsilon}_i^2 .$$

$\Leftrightarrow$ Ecuaciones normales $\begin{cases} \sum \hat{\varepsilon}_i = 0 \\ \sum X_{1i} \hat{\varepsilon}_i = 0 \end{cases}$

$$\Leftrightarrow \quad \hat{\beta}_1 = \frac{\widehat{\mathrm{cov}}(X_1, Y)}{\widehat{\mathrm{var}}(X_1)} \quad , \quad \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}_1 .$$

- Coeficiente de determinación $R^2 = \dfrac{\mathrm{var}(\hat{Y}_i)}{\mathrm{var}(Y_i)} = 1 - \dfrac{\mathrm{var}(\hat{\varepsilon}_i)}{\mathrm{var}(Y_i)}$

ya que $Y_i = \hat{Y}_i + \hat{\varepsilon}_i$ y $\hat{Y}_i \perp \hat{\varepsilon}_i$. ...

– Nótese que $R^2 = r^2$ = coef. de correlación al cuadrado $= \mathrm{corr}(Y_i, \hat{Y}_i)^2$

- SCT = SCE + SCR SC = suma cuadrados

$$\Leftrightarrow \sum (Y_i - \bar{Y})^2 = \sum (\hat{Y}_i - \bar{Y})^2 + \sum \hat{\varepsilon}_i^2$$

$$\therefore \; R^2 = \frac{SCE}{SCT} = 1 - \frac{SCR}{SCT} .$$

- Regresión múltiple:

$$Y = X\beta + \varepsilon \quad \Rightarrow \quad \hat{\beta}_{MCO} = (X^T X)^{-1} X^T Y .$$

Otra vez, $R^2 = \dfrac{\mathrm{var}(\hat{Y})}{\mathrm{var}(Y)} = \dfrac{SCE/n}{SCT/n} = 1 - \dfrac{SCR}{SCT}$, $SCR = \sum \hat{\varepsilon}_i^2$, $SCT = \sum (Y_i - \bar{Y})^2$.

- R^2 corregido $= 1 - \dfrac{SCR/n-k-1}{SCT/n-1} = 1 - \dfrac{\hat{\sigma}^2}{S_Y^2}$.

- Modelo de coeficientes beta $\begin{cases} Z_Y := Y_i - \bar{Y} / S_Y \\ Z_j := \dfrac{X_{ij} - \bar{X}_j}{S_{X_j}} \end{cases}$

Tema 4 – Regresión lineal: Inferencia

- Supuesto 1 : – $Y_i = \beta_0 + \cdots + \beta_K X_{Ki} + \varepsilon_i$
  - Matriz $\mathbb{E}[X_i X_i^T] > 0$ the definite
- Supuesto 2 : – Exogeneidad, ie:
  $\mathbb{E}[\varepsilon_i | x_i] = 0$.
- Supuesto 3 : $(X_{1i}, \dots, X_{Ki}, Y_i)$ son iid.

- Bajo estos supuestos tenemos:
  ① Insesgadez : $\mathbb{E}[\hat{\beta}_j] = \beta_j$ y $\mathbb{E}[\hat{\beta}_j | \underline{X}] = \beta_j$.

- Si asumimos que $\mathbb{E}[X_i^4] < \infty$, $\mathbb{E}[Y_i^4] < \infty$ entonces
  $\hat{\beta}_1 \xrightarrow{d} N(\beta_1, \sigma^2_{\hat{\beta}_1})$ donde $\sigma^2_{\hat{\beta}_1} = \frac{1}{n} \frac{var((X_{1i} - \mu_X)\varepsilon_i)}{var(X_{1i})^2}$.
  ↳ Simplemente una aplicación del CLT.

- Def error estándar $= ee(\hat{\beta})$ es un estimador de la desviación típica de la distribución muestral de $\hat{\beta}$.
  eg. $ee(\hat{\beta}_1) = \sqrt{\hat{\sigma}^2_{\hat{\beta}_1}} \approx \sqrt{\widehat{var}(\hat{\beta}_1)}$

- Supuesto 4 : Homocedasticidad $var(\varepsilon_i | X) = \sigma^2 \ \forall i$.
  ↳ Si incluimos este supuesto entonces
  $var(\hat{\beta} | X) = \sigma^2 (X^T X)^{-1}$
  $\Leftrightarrow var(\hat{\beta}_j) = \frac{\sigma^2}{SCT_j (1 - R_j^2)}$
  donde – $SCT_j = \sum_{t=1}^n (X_{jt} - \bar{X}_j)^2$
  – R_j = coef. si regresamos X_j con todos los X_i $i \neq j$.
  ⚠ Nótese que si hay multicolinealidad perfecta, ie. $R_j = 1$, el estimador MCO no se puede estimar.

  ↳ Bajo este supuesto $\hat{\sigma}^2 = \frac{\sum_{i=1}^n \hat{\varepsilon}_i^2}{n-K-1}$ es un estimador insesgado.

  ↳ Teorema Gauss-Markov : bajo este supuesto, los MCO son ELIO (insesgados y óptimos).

- Normalidad : si $\varepsilon \sim N(0, \sigma^2)$ entonces
  $\hat{\beta}_j \sim N(\beta_j, var(\hat{\beta}_j))$.

  ↓ coef. asimet. kurtosis

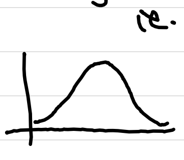
  Test : estadístico Jarque-Bera $JB = n\left[\frac{S^2}{6} + \frac{(K-3)^2}{24}\right]$
  $\sim \chi^2_2$ bajo $\mathcal{H}_0 : \varepsilon \sim N$

4.3 Inferencia

- Error tipo I = $\mathbb{P}(\text{Rechazo } \mathcal{H}_0 \,|\, \mathcal{H}_0)$, tipo II = $\mathbb{P}(\text{Acepto } \mathcal{H}_0 \,|\, \mathcal{H}_1)$.

① $\boxed{\text{Inferencia de } \beta_j \text{ – homocedasticidad + normalidad}}$

- $\mathcal{H}_0 : \beta_j = 0$, $\mathcal{H}_1 : \beta_j \neq 0$
- $t = \frac{\hat{\beta}_j}{ee(\hat{\beta}_j)} \sim t_{n-(K+1)}$ bajo homoce. + normalidad, ie. $\varepsilon_i \sim N(0, \sigma^2) \ \forall i$.



Si hacemos $\mathcal{H}_1 : \beta_j \neq 0$ utilizamos $|t| > t_{n-K-1, \alpha/2}$

Si hacemos $\mathcal{H}_1 : \beta_j > 0$ utilizamos $t > t_{n-K-1, \alpha}$

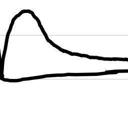
- p-valor = $\mathbb{P}$(observar nuestros datos | $\mathcal{H}_0$).

② $\boxed{\text{Inferencia de } \beta_j \text{ – hetero. + no normalidad}}$

- $\frac{\hat{\beta}_j - \beta_j}{ee(\hat{\beta}_j)} \xrightarrow{d} N(0,1)$ así que hacer contrastes asintóticos.

③ $\boxed{\text{contraste de hipótesis de conjunto : F}}$

- $\mathcal{H}_0 : \beta_{K-q+1} = \beta_{K-q+2} = \cdots = \beta_K = 0$
  $\mathcal{H}_1$: al menos una $\neq 0$.
- Ecuac. irrestrictiva $Y_i = \beta_0 + \cdots + \beta_K X_{Ki} + \varepsilon_{iNR}$
  Ecuac. restringida $Y_i = \beta_0 + \cdots + \beta_{K-q} X_{(K-q)i} + \varepsilon_{iR}$
- Homocedasticidad y normalidad
  – $\frac{(SCR_R - SCR_{NR})/q}{SCR_{NR}/n-K-1} \sim F_{q, n-K-1}$ $\quad SCR = \sum_{i=1}^n \hat{\varepsilon}_i^2$



  Rechazamos si $> F_{q, n-K-1, \alpha}$.
- Heterocedasticidad
  – $F_q^R \xrightarrow{d} \chi^2_q$ o $\frac{1}{q} F_q^R \longrightarrow F_{q, \infty}$
  cuando $n \to \infty$.

4.5 Predicción

$\hat{Y}^0 = \hat{\beta}_0 + \hat{\beta}_1 X_1^0 + \cdots + \hat{\beta}_K X_K^0$ y $\hat{\varepsilon}^0 = Y^0 - \hat{Y}^0$

por lo que $var(\hat{\varepsilon}^0 | X) = \underline{var(\varepsilon^0)} + var(\hat{Y}^0)$
$= \sigma^2 \simeq \hat{\sigma}^2$ estimador insesgado.

Podemos hacer un intervalo de confianza aprox del 95%.

$\hat{Y}^0 \pm 1.96\, ee(\hat{\varepsilon}^0) \simeq \hat{Y}^0 \pm 2 \left\{ ee(\hat{Y}^0)^2 + \hat{\sigma}^2 \right\}^{1/2}$.

Tema 6 - Regresión con Heterocedasticidad y Autocorr

6.1 Heterocedasticidad

- Recuerda que incluso en este supuesto $\hat{\beta}^{MCO}$ es insesgado y consistente, pero ineficiente en general.
- $\Sigma_{\varepsilon\varepsilon'} := \mathbb{E}[\varepsilon\varepsilon'|X] = \begin{pmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_n^2 \end{pmatrix}$

  $\Rightarrow \hat{\beta}_{MCP} = (X^T \Sigma_{\varepsilon\varepsilon'}^{-1} X)^T X^T \Sigma_{\varepsilon\varepsilon'}^{-1} y$

  ↳ Mínimos cuadrados ponderados (ie. simplemente multiplicamos por $1/\sigma_i$ para hacer todo homocedástico).

- Normalmente no conocemos σ_i así que utilizamos $\hat{\beta}^{MCO}$.
- Contrastes de heterocedasticidad.

  ↳ Supón que $\varepsilon_i^2 = \alpha_0 + \alpha_1 X_{1i} + \dots + \alpha_k X_{ki} + e_i$ entonces $\mathcal{H}_0 : var(\varepsilon_i) = \sigma^2 \Leftrightarrow \mathcal{H}_0 : \alpha_1 = \dots = \alpha_k = 0$.

  En la práctica regresamos $\hat{\varepsilon}_i^2 = \alpha_0 + \dots + \alpha_k X_{ki} + e_i$ y calculamos el estadístico $F = \dfrac{R^2/k}{(1-R^2)/n-k-1} \sim F_{k,n-k-1}$.

  ① Contraste de Breush-Pagan (BP)

  $LM = n \cdot R^2 \sim \chi^2_k$.

  ② contraste de White = calcula el LM pero en la regresión de $\hat{\varepsilon}_i^2$ se incluyen cuadrados (ie. X_{ji}^2) y efectos cruzados (ie $X_{ji}X_{li}$).

6.2 Autocorrelación

↳ $\mathbb{E}[\varepsilon_s \varepsilon_j | \underline{X}] \neq \mathbb{E}[\varepsilon_s|X]\,\mathbb{E}[\varepsilon_j|X]$.

① contraste de Durbin-Watson (DW)

$$DW = \frac{\sum_{t=1}^T (\hat{\varepsilon}_t - \hat{\varepsilon}_{t-1})^2}{\sum_{t=1}^T \hat{\varepsilon}_t^2} \simeq 2(1-\hat{\rho})$$

$\hat{\varepsilon}_t = \hat{\rho}\hat{\varepsilon}_{t-1} + e_t$.

Autocorrel. de primer orden.

Si no hay autocorr. $DW \simeq 2$ ya que $\hat{\rho} \simeq 0$.

En las tablas ofrecen valores inferiores y superiores d_i, d_s. Hay dos casos:

DW<2 ① $\mathcal{H}_0 : \hat{\rho} = 0$, $\mathcal{H}_1 : \hat{\rho} > 0$

$\underbrace{\quad}_{\mathcal{H}_1}\ d_i\ \underbrace{\quad}_{\text{no concl.}}\ d_s\ \underbrace{\quad}_{\mathcal{H}_0}\ 2$

DW>2 ② $\mathcal{H}_0 : \hat{\rho} = 0$, $\mathcal{H}_1 : \hat{\rho} < 0$

$2\ \underbrace{\quad}_{\mathcal{H}_0}\ 4-d_s\ \underbrace{\quad}_{\text{no concl.}}\ 4-d_i\ \underbrace{\quad}_{\mathcal{H}_1}\ 4$.

ie. en general

$\underbrace{\quad}_{\mathcal{H}_1}\ d_i\ \underbrace{\quad}_{\text{no conc.}}\ d_s\ \quad 2\quad \underbrace{\quad}_{\mathcal{H}_0}\ 4-d_s\ \underbrace{\quad}_{\text{no conc.}}\ 4-d_i\ \underbrace{\quad}_{\mathcal{H}_1}$

Tema 7 - Variables Explicativas Dicotómicas

7.1 Modelos ANOVA - análisis de varianza.

- $Y_i = \beta_0 + \alpha_1 D_{1i} + \dots + \alpha_m D_{mi} + \varepsilon_i$, hay m variables Dummy, i.e. $D_{ri} \in \{0,1\}$.

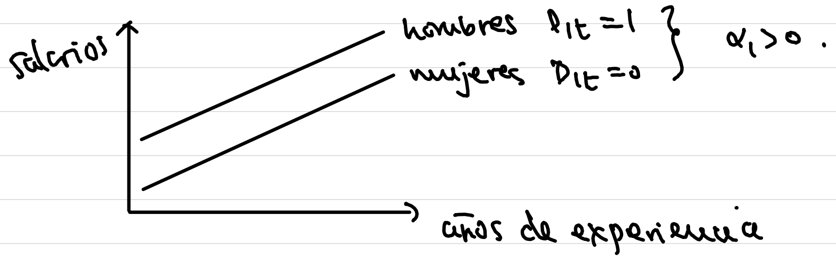
  ej: si tenemos Y_i = salarios y queremos incluir el nivel de estudio $\in \{1,2,\dots,8\}$, utilizamos 7 variables dummy (no 8 porque sino habrá multicolinealidad: $\sum D_k = 1$ ⇒ "trampa de la var. dummy").

- Si $\alpha_i \approx \alpha_j$ podemos hacer el contraste $H_0: \alpha_i - \alpha_j = 0$ utilizando el estadístico $\left| \frac{\hat{\alpha}_i - \hat{\alpha}_j}{ee(\hat{\varepsilon}_i - \hat{\varepsilon}_j)} \right| > t_{n-k-1, \alpha/2}$.

7.2 Modelo ANCOVA - incluimos dummy y cuantitativas

- Ejemplo $Y_t = \beta_0 + \beta_1 X_{1t} + \alpha_1 D_{1t} + \varepsilon_t$.

ej:



7.3 Interacciones

- Se pueden incluir cosas como $\alpha_1 D_{1i} X_{1i}$ lo cual hace que la pendiente cambie.

7.4 Estacionalidad - movimiento oscilante regular y repetitivo.

- Podemos desestacionalizar la serie haciendo una regresión tipo ANOVA, ej. $Y_t = \alpha_1 D_{1t} + \dots + \alpha_{12} D_{12t} + \varepsilon_t$ y luego analizar $Y_t - \hat{Y}_t$, la cual no debería tener estacionalidad.

7.5 Regresión por tramos

- $Y_i = \beta_0 + \alpha_0 D_{1i} + \beta_1 X_{1i} + \alpha_1 D_{1i} X_{1i} + \varepsilon_i$ y supón que $D_{1i} = \begin{cases} 0 & \text{si } X_{1i} \le j^* \\ 1 & \text{si } X_{1i} > j^* \end{cases}$

$$\Leftrightarrow \text{Dos regresiones} \begin{cases} Y_i = \beta_0 + \beta_1 X_{1i} + \varepsilon_i & X_{1i} \le j^* \\ Y_i = \beta_0 + \alpha_0 + (\beta_1 + \alpha_1) X_{1i} + \varepsilon_i & X_{1i} > j^* \end{cases}$$

Pero como tienen que coincidir en $X_{1i} = j^*$ vemos que $\alpha_0 = -\alpha_1 j^*$ por lo que obtenemos una regresión

$Y_i = \beta_0 + \beta_1 X_{1i} + \alpha_1 D_{1i} (X_{1i} - j^*) + \varepsilon_i$

Tema 8 - Análisis de especificación

8.1 Selección de variables

- Inclusión de variables irrelevantes
  - AIC $= \ln\left(\frac{\sum_{i=1}^{n} \hat{\varepsilon}_t^2}{n}\right) + \frac{2(k+1)}{n}$ ← se intenta penalizar más la adición de regresores.
  - Schwarz $= \frac{k+1}{n}\ln(n) + \ln\left(\frac{\sum \hat{\varepsilon}_t^2}{n}\right)$ ←
  - Efectos: $\hat{\beta}_j$ sigue siendo insesgado y consistentes pero $ee(\hat{\beta}_j)$ deja de ser eficiente.
  - Utilizar contrastes tipo t sobre β_j o F sobre conjuntos para evitar la sobreespecificación.

- Omisión de variables relevantes
  - Los problemas de subespecificación son más graves: si la variable omitida es un factor determinante de Y y está correl. con X_i entonces $\hat{\beta}_j^{MCO}$ se vuelve insesgado
    ↳ sesgo de variable omitida.

$$\hat{\beta}_1 \xrightarrow{P} \beta_1 + \rho_{X\varepsilon}\frac{\sigma_\varepsilon \sigma_X}{\sigma_X^2}$$

por lo que si $cov(X,\varepsilon) \neq 0$ problema!

  - Podemos incluir variables de control para mantener la insesgadez del resto de estimadores.

8.2 Mala especificación funcional

- Contraste RESET

$$Y_i = \beta_0 + \dots + \beta_k X_{ki} + \delta_1 \hat{Y}_i^2 + \delta_2 \hat{Y}_i^3 + \varepsilon_i$$

$\hat{Y}_i$ valores ajustados a la ecuac. normal.

  - Contrastamos $H_0: \delta_1 = \delta_2 = 0$. Hacemos un contraste tipo F o LM $\sim \chi_2^2$.

- Modelos no anidados: comparamos modelos que no están incluidos uno en otro, ej: comparamos el clásico con $Y_i = \beta_0 + \dots + \beta_k Z_{ki} + \varepsilon_i$ (en vez de X_{ji}).
  1. Hacer $Y_i = \sum \delta_j X_{ji} + \sum \alpha_j Z_{ji} + \varepsilon_i$ y hacer contraste F con $H_0: \alpha_j = 0 \ \forall j$.
  2. "Prueba J"
     - $Y_i = \sum \beta_j X_{ji} + \phi_1 \hat{Y}_i^z + \varepsilon_i$ y contrastar $H_0: \phi_1 = 0$. Si significativo rechazamos el modelo $Y_i = \sum \beta_j X_{ji} + \varepsilon_i$. Hacemos lo mismo con $\hat{Y}_i^x$.

8.3 Errores de medida

- Supón que Y_i^* es la variable sin error de medición y que medimos $Y_i = Y_i^* + \omega_i$. Entonces si regresamos $Y_i = \beta_0 + \dots + \beta_k X_{ki} + \varepsilon_i + \omega_i$ encontramos que

$$\text{plim}\, \hat{\beta}_1 = \frac{cov(X,Y^*) + cov(X,\omega)}{var(X)} = \beta_1 \quad \text{si}$$

$cov(X,\omega) = 0$. El único problema será que la varianza será mayor.

- También hay errores en la variable explicativa ($X_1 = X_1^* + \omega_1$). Existen dos posibles escenarios:
  1. $cov(X_1, \omega_1) = 0$
     $\hat{\beta}_1$ es consistente pero la varianza es mayor.
  2. $cov(X_1^*, \omega_1) = 0 \Rightarrow cov(X_1, \omega_1) = \sigma_\omega^2 \neq 0$.
     $\hat{\beta}_1$ es inconsistente e infraestimada, ie. $|\hat{\beta}_1| < |\beta_1|$
     Además los demás estimadores $\hat{\beta}_i$ se vuelven inconsistentes.

- Variables proxy: no podemos medir la variable exógena, ej. coef. intelectual, y utilizamos una var. que la aprox., ej. nivel de estudios.
  X_i^*: inobservable, X_i: proxy, $X_i^* = \delta_0 + \delta_1 X_i + e_i$
  - sustituimos X_i^* por X_i: "sol^n^ por sust. de var. omitida".
    - utilizar lógica sust. para ver cuando serán insistentes: ver exogeneidad.

8.4 Otras fuentes de invalidez

- Problemas de selección muestral, ie. datos perdidos.
- Causalidad simultánea ⇒ no podemos asumir que $cov(X_i, \varepsilon_i) = 0$.
- Errores estándares inconsistentes (ie. por tratamiento inapropiado de heterocedasticidad).

↳ $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$
$X_i = \gamma_0 + \gamma_1 Y_i + v_i$
$\Rightarrow cov(X_i, \varepsilon_i) = \frac{\gamma_1 \sigma_\varepsilon^2}{1 - \gamma_1 \beta_1}$

Da lugar a un MCO sesgado e inconsistente.

Resumen Importante

- Contraste global
  $\mathcal{H}_0: \beta_0 = \cdots = \beta_k = 0$ $\qquad F_{k,n-k-1} = \dfrac{R^2/k}{(1-R^2)/n-k-1}$

- Contraste algunas β_i
  $\mathcal{H}_0: \beta_{i_1} = \cdots = \beta_{i_q} = 0$ $\qquad F_{q,n-k-1} = \dfrac{(SCR_R - SCR_{NR})/q}{SCR_{NR}/n-k-1}$.