**PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ**
**CIENCIAS SOCIALES**
**CICLO 2022-2**

**Fundamentos de Econometría**
**Práctica Dirigida 4**


| | | |
|---|---|---|
| Profesor: | Juan Palomino | juan.palominoh@pucp.pe |
| Jefes de Práctica: | Tania Paredes | tania.paredes@pucp.edu.pe |


Fecha: 17 – 09 – 2022

## 1. Interpretación de inferencia estadística

a. Explique la relación entre nivel de confianza e intervalos de confianza

a. Definimos el nivel de confianza como $(1-\alpha)\cdot 100\%$, de tal forma que si "$\alpha$" es 1, 5 o 10 obtenemos niveles de confianza de 99%, 95% o 90% respectivamente. Con dicho nivel de confianza $(1-\alpha)\cdot 100\%$, construimos los intervalos de confianza (IC). Si es una distribución de dos colas, como una t-Student, el percentil de la distribución a utilizar sera $t_{1-\frac{\alpha}{2}}$. Si es una distribución de una cola, como una distribución F (de Fisher), el percentil de la distribución a utilizar será $F_{1-\alpha}$.

De esta forma, en el modelo bivariado, la probabilidad de que el $\beta_2$ del Proceso Generador de Datos (PGD) se encuentre dentro del intervalo de confianza es el nivel de confianza.

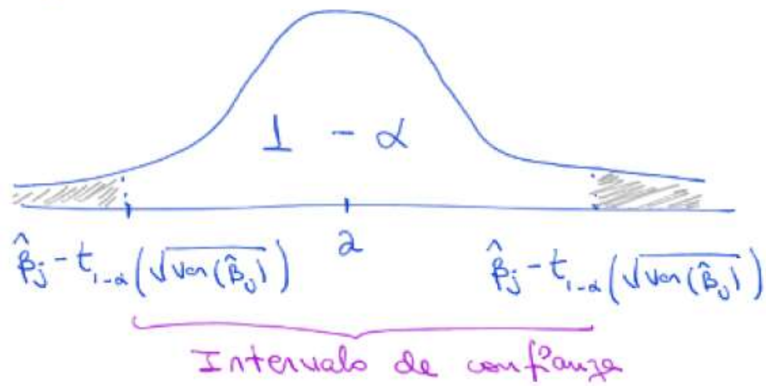
$$Pr\left(\hat{\beta}_2 - t_{1-\frac{\alpha}{2}}\sqrt{Var(\hat{\beta}_2)} < \beta_2 < \hat{\beta}_2 + t_{1-\frac{\alpha}{2}}\sqrt{Var(\hat{\beta}_2)}\right) = 1-\alpha$$

En el modelo multivariado, sería lo siguiente:

$$Pr\left(\hat{\beta}_j - t_{1-\frac{\alpha}{2}}\sqrt{\hat{\sigma}^2 c_{jj}} < \beta_j < \hat{\beta}_j + t_{1-\frac{\alpha}{2}}\sqrt{\hat{\sigma}^2 c_{jj}}\right) = 1-\alpha$$

Donde $\hat{\sigma}^2$ es la varianza muestral y $c_{jj}$ la celda $_{jj}$ de la matriz $(X'X)^{-1}$ (ver PD3 para mayor detalle).

De forma gráfica, asumiendo $H_0: \beta_j = a$




b. Explique la relación entre nivel de confianza y nivel de significancia. Luego, a través del test *t*, analice los casos de rechazo o no rechazo de una hipótesis nula del estimador igual a cero.

b. Si el nivel de confianza es $(1-\alpha) \times 100\%$, dicho $\alpha$ es el nivel de significancia. De tal forma que al sumar ambos valores obtengo $100\%$.

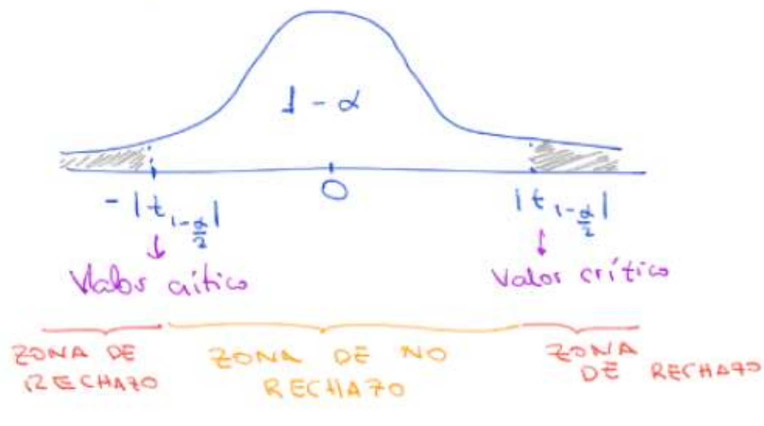
Realizamos el estadístico t y lo comparamos con el valor crítico de la distribución t.

$$\underbrace{\left|\frac{\beta_j - a}{\sqrt{\hat{\sigma}^2 c_{jj}}}\right|}_{\text{estadístico } t} > \underbrace{t_{1-\frac{\alpha}{2}}^{(n-k)}}_{\text{Valor crítico}} \Rightarrow \text{Rechazo de } H_0$$

$$\left|\frac{\beta_j - a}{\sqrt{\hat{\sigma}^2 c_{jj}}}\right| < t_{1-\frac{\alpha}{2}}^{(n-k)} \Rightarrow \text{No rechazo de } H_0$$

Como la hipótesis es $H_0: \beta_j = 0$ ; entonces $a = 0$.

De forma gráfica:




c. ¿Cuáles son los tres niveles de significancia más usados?

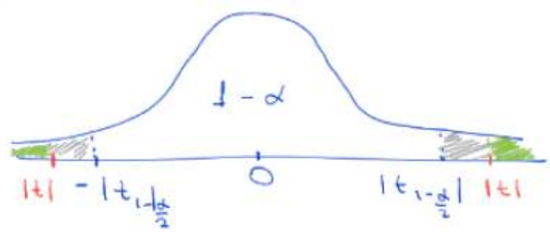
c. Los tres niveles de significancia más usados son 1%, 5% y 10%.

d. Defina qué es el p-value y clasifíquelo según los niveles de significancia de 1c.

d. El p-value es el menor nivel de significancia con el que podemos rechazar una nula. Dado que esta variable es continua, se plantean las siguientes categorías.

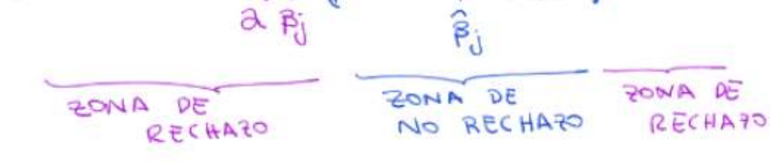
* Si p-value > 0.1, el parámetro no es significativo
* Si 0.05 < p-value ≤ 0.1, el parámetro es significativo al 10%.
* Si 0.01 < p-value ≤ 0.05; " " " " 5%
* Si p-value ≤ 0.01; " " " " 1%.

De forma gráfica, es el área sombreada de color verde del gráfico anterior.

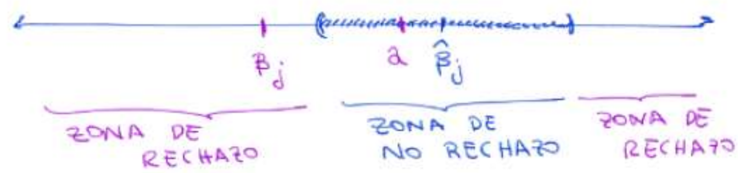



e. Explique el error tipo I y el error tipo II y proponga ejemplos

e. Intuitivamente, el error tipo I es "el error más grave" que se puede realizar, pues estoy rechazando algo que es "verdadero". De forma estricta, el error tipo I implica rechazar la hipótesis nula a pesar que es "verdadera".
Sea el siguiente ejemplo. Se plantea que $H_0: \beta_j = a$.
Nuestras estimaciones indican que $\hat{\beta}_j$ es lejano a $a$, y rechazamos la $H_0$, pero en realidad $\beta_j$ es muy cercano a $a$.



Otro ejemplo está asociado con la significancia del estimador. Si p-value es 4%, implica que tengo 4% de probabilidades de cometer el error tipo I.
El error tipo II es lo contrario, es "aceptar" una hipótesis cuando debió ser rechazada. Si $H_0: \beta_j = a$, de forma gráfica:




La probabilidad de cometer este error es "$\beta$", el cual es la potencia de la prueba (no confundir con parámetros del modelo econométrico $\beta_j$). Para reducirlo, a mayor muestra, mayor potencia.

2. **Hipótesis lineales y Test F**
   a. A partir de la función de producción Cobb-Douglas con progreso tecnológico a la Hicks, elabore un modelo econométrico

② F. Producción Cobb Douglas con progreso tecnológico a la Hicks: $Y = AK^{\beta_2}L^{\beta_3}$. El modelo econométrico implica aplicar logaritmos (ver PD1)

$$\ln Y = \ln A + \beta_2 \ln K + \beta_3 \ln L + u$$ , pero A es desconocido (es la PTF)

$$\ln Y = \beta_1 + \beta_2 \ln K + \beta_3 \ln L + u.$$

   b. Diseñe una hipótesis lineal de tal forma que pueda rechazar la hipótesis de rendimientos constantes a escala. Luego, plantee dicha hipótesis de forma matricial

(b) $H_0: \beta_2 + \beta_3 = 1$ ; rendimientos ctes a escala

De forma matricial: $[0 \;\; 1 \;\; 1]\begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} = [1]$

Aquí $q = 1$ (número de filas del resultado de la hipótesis)

c. A partir de la siguiente base de datos, construya el estadístico F

(c) Dado que la matriz x tiene en sus columnas las variables constante (1) ; $\ln(K)$ y $\ln(L)$. Si la matriz $(x'x)^{-1}$ es la siguiente:

$$(x'x)^{-1} = \begin{bmatrix} 3.009 & 0.034 & -0.559 \\ 0.034 & 0.205 & -0.271 \\ -0.559 & -0.271 & 0.447 \end{bmatrix}$$

y $\hat{\beta} = \begin{bmatrix} 1.171 \\ 0.376 \\ 0.603 \end{bmatrix}$

y el estadístico F es:

$$F = \frac{(R\hat{\beta} - r)'[R(x'x)^{-1}R']^{-1}[R\hat{\beta} - r]/q}{\hat{\varepsilon}'\hat{\varepsilon}/(n-k)}$$

Remplazando:

- $R\hat{\beta} - r = [0 \;\; 1 \;\; 1] \begin{bmatrix} 1.171 \\ 0.376 \\ 0.603 \end{bmatrix} - 1 = \underline{-0.022}$

- $R(X'X)^{-1}R = [0 \;\; 1 \;\; 1] \begin{bmatrix} 3.009 & 0.034 & -0.559 \\ 0.034 & 0.205 & -0.271 \\ -0.559 & -0.271 & 0.447 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$

$$= \underline{0.110}$$

- $\hat{\varepsilon}'\hat{\varepsilon} = \underline{0.852}$

- $q = 1$

- Sea $N = 27$ y evidentemente $K = 3$ $\therefore$ $\underline{N - K = 24}$

$$\therefore \quad \underbrace{F}_{\text{estadístico } F} = \frac{(-0.022)(0.110)(-0.022)/1}{0.852/24} = \underline{0.116}$$

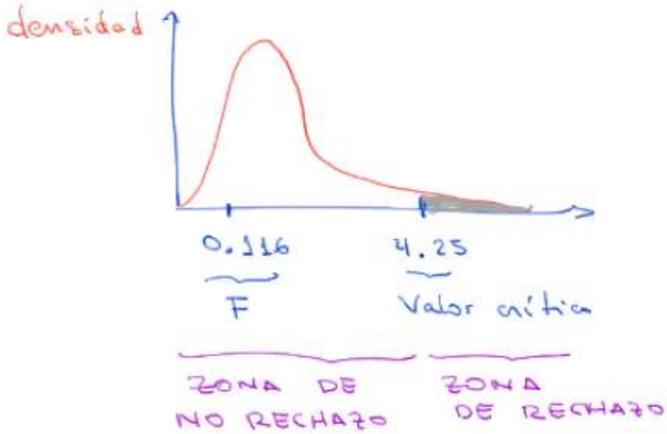
d. Bajo este planteamiento, ¿se rechaza la hipótesis nula? Use el test F y justifique usando el gráfico de la función de densidad F.

(d) El valor crítico $F_{(1,24)} = 4.25$

Si estadístico F es menor al valor crítico, no se rechaza la $H_0$.

Test F: $F < F_{(1,24)}$

Función de densidad




3. **Estimación con restricciones lineales**
   Un investigador quiere explicar la evolución de la tasa de criminalidad (C) medida como el número de delitos reportados a la policía por cada 1000 personas usando la información de la tasa de desempleo (U) y de la tasa de deserción escolar (D) a nivel regional

| Source | SS | df | MS | | |
|---|---|---|---|---|---|
| | | | | Number of obs = | 23 |
| | | | | F( 2, 20) = | 12093.02 |
| Model | 47601.9509 | 2 | 23800.9755 | Prob > F = | 0.0000 |
| Residual | 39.3631692 | 20 | 1.96815846 | R-squared = | 0.9992 |
| | | | | Adj R-squared = | 0.9991 |
| Total | 47641.3141 | 22 | 2165.51428 | Root MSE = | 1.4029 |

| C | Coef. | Std. Err. | t | P>\|t\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| U | 1.521364 | .0427094 | 35.62 | 0.000 | 1.432274 | 1.610454 |
| D | .2547178 | .0926254 | 2.75 | 0.012 | .0615046 | .447931 |
| _cons | -15.36337 | 1.980653 | -7.76 | 0.000 | -19.49494 | -11.2318 |

El investigador quiere estimar este modelo imponiendo la condición de que la suma de los efectos marginales del desempleo y de la tasa de deserción escolar sobre la tasa de criminalidad es igual a dos. Si además usted cuenta con la siguiente información:

$$(X'X)^{-1} = \begin{bmatrix} 1.993227 & & \\ -0.040630 & 0.000927 & \\ 0.078727 & -0.078727 & 0.004359 \end{bmatrix}$$

Donde $X = [i \;\; U \;\; D]$, se le pide:

a. Obtener el vector de parámetros estimados del modelo restringido.

a) Queremos hallar $\tilde{\beta}_1$, $\tilde{\beta}_2$ y $\tilde{\beta}_3$ (los estimadores restringidos).
También nos dan información sobre la hipótesis nula $\tilde{\beta}_2 + \tilde{\beta}_3 = a$

Antes teníamos un modelo irrestricto $\rightarrow \min \hat{\varepsilon}'\hat{\varepsilon}$

$y = x\hat{\beta} + \hat{\varepsilon}$

donde $\hat{\beta} = (x'x)^{-1}x'y$

Ahora tenemos un modelo restringido

$\min \tilde{\varepsilon}'\tilde{\varepsilon} = (y - x\tilde{\beta})'(y - x\tilde{\beta})$

s.a. $R\tilde{\beta} = r$

$\tilde{\beta} = \hat{\beta} + (x'x)^{-1}R'[R(x'x)^{-1}R']^{-1}(r - R\hat{\beta})$

✓ Definimos $\widehat{R\beta} = r \rightarrow r - R\hat{\beta}$

$\underset{1\times 3}{[0 \quad 1 \quad 1]} \underset{3\times 1}{\begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix}} = a$ ; $\quad R = [0 \quad 1 \quad 1]$, $\quad r = a$

✓ Además si: $r = a \rightarrow R\hat{\beta} - a = \hat{\beta}_2 + \hat{\beta}_3 - a$

o $a - \hat{\beta}_2 - \hat{\beta}_3$

✓ Por otra parte : $\boxed{\text{Podemos aproximar así}}$

$$\begin{bmatrix} \tilde{\beta}_1 \\ \tilde{\beta}_2 \\ \tilde{\beta}_3 \end{bmatrix} = \underbrace{\begin{bmatrix} -15,3 \\ 1,52 \\ 0,25 \end{bmatrix}}_{\substack{\hat{\beta} \\ \text{(datos de la tabla)}}} + \underbrace{\begin{bmatrix} 1,9 & -0,04 & 0,07 \\ -0,04 & 0,0009 & -0,001 \\ 0,04 & -0,001 & 0,004 \end{bmatrix}}_{(X'X)^{-1} = A} \underbrace{\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}}_{R'} \underbrace{\left\{ [0 \ 1 \ 1] A \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \right\}^{-1}}_{[R(X'X)^{-1}R']^{-1}} \cdots$$

$$(2 - \underbrace{1,52}_{\hat{\beta}_2} \underbrace{-0,25}_{\hat{\beta}_3})$$

Finalmente se obtiene $\begin{bmatrix} \tilde{\beta}_1 \\ \tilde{\beta}_2 \\ \tilde{\beta}_3 \end{bmatrix} = \begin{bmatrix} -9,6 \\ 1,3 \\ 0,6 \end{bmatrix}$

b. Determine cuál de los dos modelos se ajusta mejor a los datos si se sabe que la SCR del modelo restringido es igual a 75.44174

$$F = \frac{(SCR_R - SCR_I)/q}{SCR_I/(n-k)}$$

; $q = 1 \rightarrow$ # de restricciones
$n = 23$ (de la tabla)
$k = 3 \rightarrow$ # parámetros

✓ $SCR_R = 75{,}44$ (dato)

✓ $SCR_U = 39{,}3631$ (de la tabla)

$$\boxed{F = \frac{\tilde{e}'\tilde{e} - \hat{e}'\hat{e} / q}{\hat{e}'\hat{e} / n-k}}$$

$$F = \frac{(75{,}44 - 39{,}36) / 1}{(39{,}36)/23-3} = \boxed{18{,}33}$$

Comparamos F hallado con F de la tabla

$$\boxed{18{,}33 > F_{0.95}(1,20) = 4{,}3}$$

L Se rechaza Ho: $\beta_2 + \beta_3 = 2$ al 95% de confianza
5% de significancia

## 4. Laboratorio

a. A partir de la siguiente base de datos adjuntada (ver archivo datos.dta adjuntado), realice una regresión del efecto de las demás variables sobre el ingreso.

```
. regress income educ jobexp race

      Source |       SS           df       MS      Number of obs   =        20
-------------+----------------------------------   F(3, 16)        =     29.16
       Model |  1538.92019         3  512.973396   Prob > F        =    0.0000
    Residual |  281.505287        16  17.5940804   R-squared       =    0.8454
-------------+----------------------------------   Adj R-squared   =    0.8164
       Total |  1820.42548        19  95.8118671   Root MSE        =    4.1945

------------------------------------------------------------------------------
      income |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   1.981124   .3231024     6.13   0.000     1.296178     2.66607
      jobexp |   .6419622   .1811106     3.54   0.003     .2580248      1.0259
        race |   .5707931   2.871949     0.20   0.845    -5.517466    6.659052
       _cons |  -7.863763   5.369166    -1.46   0.162    -19.24589    3.518362
------------------------------------------------------------------------------
```

b. Verifique el valor del estadístico *t* de cada estimador.
Recordamos:

$$t - estadistico = \frac{\widehat{\beta_2} - \beta_{2-HO}}{\sqrt{Var(\widehat{\beta_2})}}$$

Ejemplo para el caso de educ:

$$t - estadistico = \frac{1.981124 - 0}{0.3231024} = 6.131566$$

Esto mismo se puede comprobar para los otros estimadores.

c. Interprete cada estimador en función a su significancia y los intervalos de confianza

- **En función del t estadístico**

Hallamos el valor crítico. Tanto en Stata como en la mayoría de literatura econométrica, se suele trabajar con un nivel de significancia de 5%. Por ello, hallamos el valor crítico en cualquier tabla de distribución *t-Student* es:
$t_{1-\frac{0.05}{2}}(N-k) = 2.12.$
Comparamos con el t estadístico y, para el caso de educación:

Si $|t_{estadistico}| > t_{tabla}$, se rechaza la Hipótesis nula de $\beta_{2-HO} = 0$, dado que:

$$|6.131566| > 2.12$$

Lo mismo se repite para los demás estadísticos t.

- **En función del p-value:**

Stata por defecto calcula el p-value. En función de *PD4 parte 1d* definimos la significancia.

- **En función del IC:**

Vemos que el IC no considera en algunos casos cero; por lo tanto, se rechaza la Ho.

d. Aplique el test F explicado en la clase anterior. Para ello, diseñe la hipótesis lineal en función de matrices e interprete el resultado.

- **En función del comando *regress (reg)*:**

$$H0: \beta_2 = 0, \beta_3 = 0, \beta_4 = 0$$
$$H1: \beta_j \neq 0$$

Matricialmente
Ho:

```
 0 1 0 0         B1          0
[0 0 1 0]    [   B2   ] = [  0    ]
 0 0 0 1         B3          0
                 B4
```

Donde en esta matríz *k=4* y *q=3*. Por defecto Stata calcula el test F. El estadístico F es 29.16 (ver imagen de 4a) y el valor crítico es $F(q, N-k) = F(3{,}16)$. El p-value *(Prob > F)* es 0.0000, lo cual implica que se rechaza la H0 donde los 3 parámetros son iguales a cero.

- **En función del comando *test*:**

```
. test educ jobexp race //obtenemos exactamente los mismos resultados

 ( 1)  educ = 0
 ( 2)  jobexp = 0
 ( 3)  race = 0

       F(  3,    16) =   29.16
            Prob > F =    0.0000
```