Econometría Aplicada Intermedia Semana 1

Edinson Tolentino
MSc Economics
email: edinsontolentinor@pacifico.com.pe

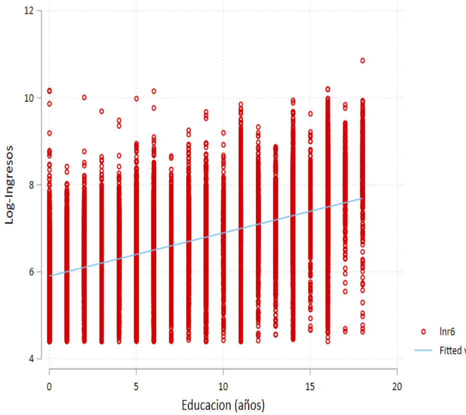
Twitter: @edutoleraymondi

7 de agosto de 2022

Contenido

Introducción

- Deseamos responder una prenguta general: ¿Cuánto puedo recibir de ingresos dado mas años de educación?
- Para ello debemos tratar de responder las siguientes preguntas:
  - Donde conseguir la información: ENAHO
    Se presenta como trabajar bases de datos de ENAHO
  - ¿Relación de los ingresos y años de educacion ?: positiva o negativa
    Se analizará variables como años de educacion y los ingresos laborales



Análisis Exploratorio

La fuente de datos proviene de la ENAHO , se trabaja con los modulos 300 y 500. Ademas se toma en cuenta los siguientes puntos:

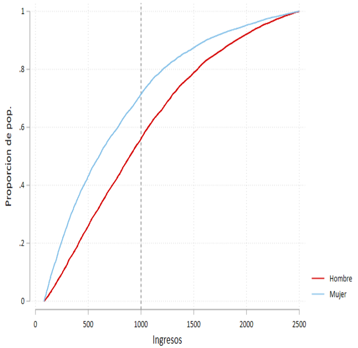
- Solo se toma en cuenta a los jefes de hogares
- PEA Ocupada
- Observaciones superiores al 5 % de distribución de los ingresos.
- No missing values en la variable años educación

La presente sección busca poder tener un análisis exploratorio de las variables como **r6** (ingresos laborales) y **reduca** años de educación en la base de datos del 2021 para responder nuestra pregunta.

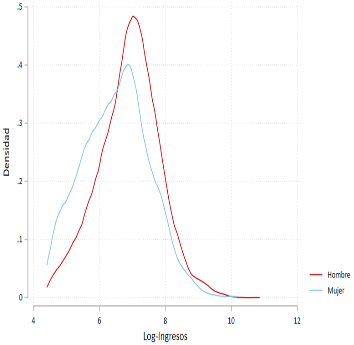
- Son los ingresos laborales diferentes a nivel del percentil 5 y el precentil 75
- Son los años de educación de un trabajador diferentes a nivel del percentil 5 y el precentil 75

Análisis Exploratorio

La proporción de mujeres y honbres que perciben ingresos de 1,000 soles es diferentes



La distribución de los ingresos es heterogenea



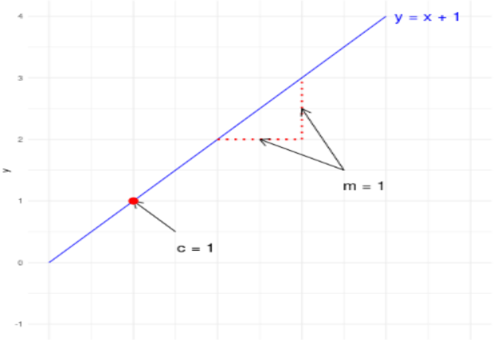
Ecuación de la Recta

- **Regresion Lineal** supone que el promedio de la variable producto y sobre muchas observaciones tiene una relación directa con el **insumo** de variables $x_1, \cdots, x_p$
- Recordando la informacion de la ecuación de una **linea recta** de la formación del colegio:

$$y = mx + c$$

.

Donde m es la pendiente de linea y c es el intercepto y



Regresión Lineal Simple

- **Regresion lineal simple** refiere el caso donde existe solo una variable de insumo y solo una variable de producto. Existe una tecnica facil de visualizar. Dado el conjunto de data, se puede plotear los valores de y y x, ademas se puede encontrar la **mejor linea de predicción**.
- Linea de regresion estimada:

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

- Donde los estimadores son:
  - Pendiente

$$\hat{\beta} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}$$

  - Intercepto

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$$

Regresión Lineal Simple: aplicación I

- Luego del análisis exploratorio, se procede a poder observar regresiones lineales bivariadas siguiendo la metodología de Mincer
- **Los retornos a la educación**, siguiendo la propuesta de Mincer
- Se propone la siguiente ecuación:

$$r6 = \hat{\alpha} + \hat{\beta} reduca + \varepsilon$$

Donde: $r6$ denota los ingresos laborales, *reduca* los años de educación de la trabajador(a).

Cuadro: Ingresos - Dep. Var

	(1)
reduca	145.45***
	(1.98)
Constant	106.24***
	(19.97)
Observaciones	26737
Adj. R^2	0.168

Errores estandar en parentesis ()
Fuente: INEI - 2020.
Elaboracion: Autor

Regresión Lineal Simple: medidas de ajustes

- **Bondad de ajuste**: una primera garantía sobre la calidad de la regresión proviene de verificar si nuestra línea de regresión, basada en los coeficientes estimados, se ajusta bien a los datos.
- Para chequear esto se usa el R^2, este debe estar entre 0 (no predicción) y 1 (perfecta predicción)
- Entonces:

$$R^2 = \frac{ESS}{TSS}$$

- ESS: sumatoria de las desviaciones al cuadrado de la predicción del promedio

$$ESS = \sum_{i=1}^{n} \left(\hat{Y}_i - \bar{Y}\right)^2$$

- TSS: sumatoria de las desviaciones al cuadrado DE Y_i del promedio

$$TSS = \sum_{i=1}^{n} \left(Y_i - \bar{Y}\right)^2$$

Regresión Lineal Simple: significancia

Regresión Lineal Simple: significancia

- El estimador de OLS (MCO) , $\hat{\beta}_0$ y $\hat{\beta}_1$ son calculados de una muestra aletoria , ellos son asimismo variables altorias. Dado que tenemos diferentes muestras , tendremos diferentes $\hat{\beta}_0$ y $\hat{\beta}_1$ todo el tiempo.

Regresión Lineal Simple: significancia

- El estimador de OLS (MCO) , $\hat{\beta}_0$ y $\hat{\beta}_1$ son calculados de una muestra aletoria , ellos son asimismo variables altorias. Dado que tenemos diferentes muestras , tendremos diferentes $\hat{\beta}_0$ y $\hat{\beta}_1$ todo el tiempo.
- Necesidad de realizar una inferencia , dado que no conocemos **el verdadero valor de** β_0 y β_1

Regresión Lineal Simple: significancia

- El estimador de OLS (MCO) , $\hat{\beta}_0$ y $\hat{\beta}_1$ son calculados de una muestra aletoria , ellos son asimismo variables altorias. Dado que tenemos diferentes muestras , tendremos diferentes $\hat{\beta}_0$ y $\hat{\beta}_1$ todo el tiempo.
- Necesidad de realizar una inferencia , dado que no conocemos **el verdadero valor de** β_0 y β_1
- Se utiliza la prueba t (dado que no se conoce la población)

$$t = \frac{\hat{\beta} - \beta}{SE(\hat{\beta})}$$

Regresión Lineal Simple: significancia

- El estimador de OLS (MCO) , $\hat{\beta}_0$ y $\hat{\beta}_1$ son calculados de una muestra aletoria , ellos son asimismo variables altorias. Dado que tenemos diferentes muestras , tendremos diferentes $\hat{\beta}_0$ y $\hat{\beta}_1$ todo el tiempo.
- Necesidad de realizar una inferencia , dado que no conocemos **el verdadero valor de** β_0 y β_1
- Se utiliza la prueba t (dado que no se conoce la población)

$$t = \frac{\hat{\beta} - \beta}{SE(\hat{\beta})}$$

- El error estandar o desviación estandar del estimador : $SE(\hat{\beta})$

Regresión Lineal Simple: significancia

- El estimador de OLS (MCO) , $\hat{\beta}_0$ y $\hat{\beta}_1$ son calculados de una muestra aletoria , ellos son asimismo variables altorias. Dado que tenemos diferentes muestras , tendremos diferentes $\hat{\beta}_0$ y $\hat{\beta}_1$ todo el tiempo.
- Necesidad de realizar una inferencia , dado que no conocemos **el verdadero valor de** β_0 y β_1
- Se utiliza la prueba t (dado que no se conoce la población)

$$t = \frac{\hat{\beta} - \beta}{SE(\hat{\beta})}$$

- El error estandar o desviación estandar del estimador : $SE(\hat{\beta})$
- Para evaluar la significancia (se usara STATA a través de una prueba bilateral t-test) con la siguientes hipotesis nula:

$$H_o : \beta = 0$$

Regresión Lineal Simple: significancia

- El estimador de OLS (MCO) , $\hat{\beta}_0$ y $\hat{\beta}_1$ son calculados de una muestra aletoria , ellos son asimismo variables altorias. Dado que tenemos diferentes muestras , tendremos diferentes $\hat{\beta}_0$ y $\hat{\beta}_1$ todo el tiempo.
- Necesidad de realizar una inferencia , dado que no conocemos **el verdadero valor de** β_0 y β_1
- Se utiliza la prueba t (dado que no se conoce la población)

$$t = \frac{\hat{\beta} - \beta}{SE(\hat{\beta})}$$

- El error estandar o desviación estandar del estimador : $SE(\hat{\beta})$
- Para evaluar la significancia (se usara STATA a través de una prueba bilateral t-test) con la siguientes hipotesis nula:

$$H_o : \beta = 0$$

  - 99 %, 95 % , 90 % niveles de confianza

Regresión Lineal Simple: significancia

- El estimador de OLS (MCO) , $\hat{\beta}_0$ y $\hat{\beta}_1$ son calculados de una muestra aletoria , ellos son asimismo variables altorias. Dado que tenemos diferentes muestras , tendremos diferentes $\hat{\beta}_0$ y $\hat{\beta}_1$ todo el tiempo.
- Necesidad de realizar una inferencia , dado que no conocemos **el verdadero valor de** β_0 y β_1
- Se utiliza la prueba t (dado que no se conoce la población)

$$t = \frac{\hat{\beta} - \beta}{SE(\hat{\beta})}$$

- El error estandar o desviación estandar del estimador : $SE(\hat{\beta})$
- Para evaluar la significancia (se usara STATA a través de una prueba bilateral t-test) con la siguientes hipotesis nula:

$$H_o : \beta = 0$$

  - 99 %, 95 % , 90 % niveles de confianza
  - **P-value** en los documentos ***, **, *, donde 0.01,0.05, 0.1

Regresión Lineal Simple: aplicación II

Stata implementa el modelo de regresión se implementa mediante los comandos `reg` cuya sintaxis es:

Syntax

reg depvar [indepvars] [if] [in] [weight], [options]

```
eststo: reg r6 reduca

      Source |       SS           df       MS      Number of obs   =    26,737
-------------+----------------------------------   F(1, 26735)     =   5399.91
       Model |  1.3054e+10         1  1.3054e+10   Prob > F        =    0.0000
    Residual |  6.4632e+10    26,735  2417521.97   R-squared       =    0.1680
-------------+----------------------------------   Adj R-squared   =    0.1680
       Total |  7.7687e+10    26,736  2905702.05   Root MSE        =    1554.8

------------------------------------------------------------------------------
          r6 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      reduca |   145.4499   1.979339    73.48   0.000     141.5703    149.3295
       _cons |   106.2389   19.97219     5.32   0.000      67.0924    145.3855
------------------------------------------------------------------------------
```

Regresión Lineal Simple: formas funcionales

- La regresión anterior puede ser estimada de forma diferente:

$$r6 = \hat{\alpha} + \hat{\beta} reduca + \varepsilon$$

$$r6 = \hat{\alpha} + \hat{\beta} ln(reduca) + \varepsilon$$

$$ln(r6) = \hat{\alpha} + \hat{\beta} reduca + \varepsilon$$

$$ln(r6) = \hat{\alpha} + \hat{\beta} ln(reduca) + \varepsilon$$

- Donde las interpretaciones alrededor de la pendiente cambian $\hat{\beta}$:

Model	Dependent Variable	Independent Variable	Interpretation of β
Level-level	y	x	$dy = \beta dx$
Level-log	y	log(x)	$dy = (\beta/100)\% dx$
Log-level	log(y)	x	$\% dy = (100\beta) dx$
Log-log	log(y)	log(x)	$\% dy = \beta \% dx$

Regresión Lineal Simple: aplicacion III

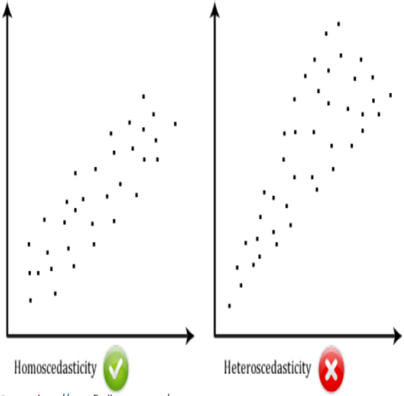
Modelo	Interpretación de coeficientes
Semielasticidad	un cambio porcentual en y dado el incremento de una unidad de x $(100.\beta)$
Elasticidad	un cambio porcentual en y dado el incremento porcentual de una unidad de x (β)

Cuadro: Ingresos - Dep. Var

	(r6)	ln(r6)	(r6)	ln(r6)
reduca	145.45***	0.10***		
	(1.98)	(0.00)		
lnreduca			858.18***	0.59***
			(15.68)	(0.01)
Constant	106.24***	5.90***	-337.30***	5.61***
	(19.97)	(0.01)	(34.14)	(0.02)
Observaciones	26737	26737	25238	25238
Adj. R^2	0.168	0.242	0.106	0.164

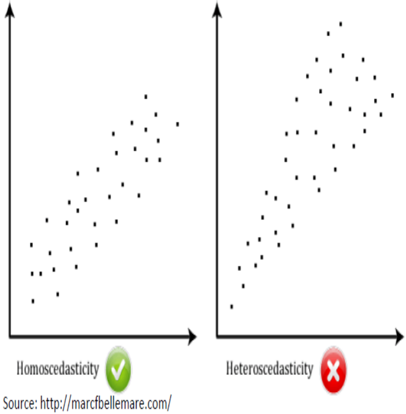
Errores estandar en parentesis ()
Fuente: INEI - 2020.
Elaboracion: Autor

Regresión Lineal Simple: heterocedasticidad



Source: http://marcfbellemare.com/

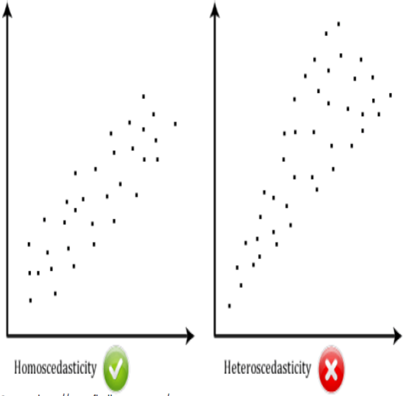
Regresión Lineal Simple: heterocedasticidad



- Dado:

$$E(\varepsilon_i \mid X_i)$$

Regresión Lineal Simple: heterocedasticidad



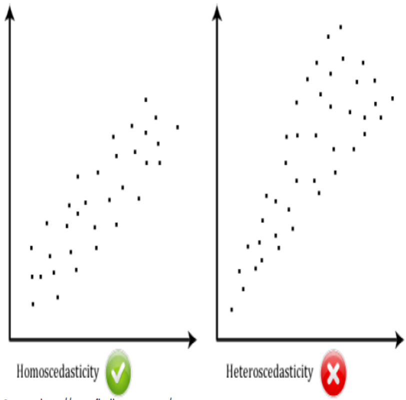
Source: http://marcfbellemare.com/

- Dado:

$$E(\varepsilon_i \mid X_i)$$

- Significa que el valor esperado del termino de error condicional sobre X_i es cero.

Regresión Lineal Simple: heterocedasticidad



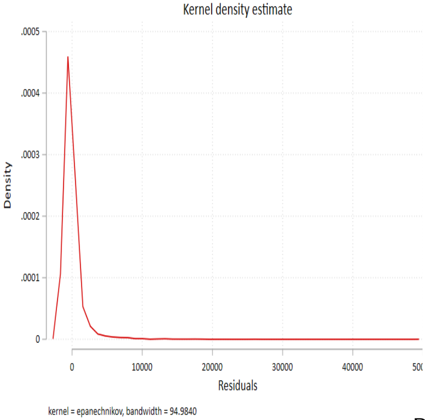
Source: http://marcfbellemare.com/

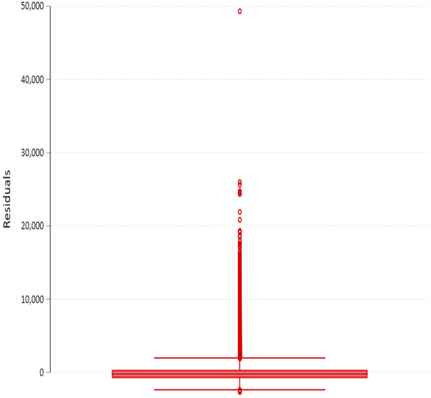
- Dado:

$$E(\varepsilon_i \mid X_i)$$

- Significa que el valor esperado del termino de error condicional sobre X_i es cero.
- Si la varianza de ε_i no depende sobre el valor de X_i, por ejemplo , la varianza ε_i es constante dado X_i, entonces el error es llamada homocedastico

Regresión Lineal Simple: aplicacion IV





Definitivamente los errores del modelo bivariado no son normales

Regresión Lineal Simple: conclusiones

- Analizar la inclusión de mas de una variable independiente (Xs)
- Revisión de la presencia de heterocedasticidad a través de pruebas robustas
    - Deteccion de pruebas de errores i.i.d
    - Ajuste del modelo a través de los errores robustos de White (comando robust)
- Siempre poder comprender la inclusión de contexto de literatura y marco teorico en las estimaciones

Regresión lineal Multivariada

Regresión lineal Multivariada

- La ecuación general de los salarios puede ser formulada como sigue:

$$ln\left(w_i\right) = \alpha + \gamma_1 Educ_i + \gamma_2 Exp_i + \gamma_3 Exp_i^2 + \pi Male_i + X_i' \beta + \mu_i \qquad (1)$$

Regresión lineal Multivariada

- La ecuación general de los salarios puede ser formulada como sigue:

$$In\left(w_i\right) = \alpha + \gamma_1 Educ_i + \gamma_2 Exp_i + \gamma_3 Exp_i^2 + \pi Male_i + X_i'\beta + \mu_i \qquad (1)$$

- Donde: $\mu_i \sim N\left(0, \sigma^2\right)$
- $In\left(w_i\right)$: logaritmo de ingresos laborales
- $Educ_i$: años de educación
- Exp_i: experiencia laboral en años
- $Male_i$: variable dummy $== 1$, si la persona es hombre

Regresión lineal Multivariada

- La ecuación general de los salarios puede ser formulada como sigue:

$$ln\left(w_i\right) = \alpha + \gamma_1 Educ_i + \gamma_2 Exp_i + \gamma_3 Exp_i^2 + \pi Male_i + X_i'\beta + \mu_i \qquad (1)$$

- Donde: $\mu_i \sim N\left(0, \sigma^2\right)$
- $ln\left(w_i\right)$: logaritmo de ingresos laborales
- $Educ_i$: años de educación
- Exp_i: experiencia laboral en años
- $Male_i$: variable dummy $== 1$, si la persona es hombre
- Analizaremos los efectos marginales de la variable **Exp** (por ejemplo, experiencia de la fuerza laboral) sobre los ingresos laborales.

Regresión lineal Multivariada: Método Delta

Regresión lineal Multivariada: Método Delta

- El efecto marginal para Exp esta dado por:

$$\frac{\partial ln(w)}{\partial exp} = \hat{\gamma}_2 + 2\hat{\gamma}_3 Exp$$

Regresión lineal Multivariada: Método Delta

- El efecto marginal para Exp esta dado por:

$$\frac{\partial ln(w)}{\partial exp} = \hat{\gamma}_2 + 2\hat{\gamma}_3 Exp$$

- En orden de encontrar el punto de inflexión, la derivada de la ecuación mostrada lineas arriba debe ser igualada a cero, la cual resolvera el estado estacionario

Regresión lineal Multivariada: Método Delta

- El efecto marginal para Exp esta dado por:

$$\frac{\partial ln(w)}{\partial exp} = \hat{\gamma}_2 + 2\hat{\gamma}_3 Exp$$

- En orden de encontrar el punto de inflexión, la derivada de la ecuación mostrada lineas arriba debe ser igualada a cero, la cual resolvera el estado estacionario

$$Exp_{estacionario} = \frac{\hat{\gamma}_2}{2\hat{\gamma}_3}$$

Regresión lineal Multivariada: Método Delta

- El efecto marginal para Exp esta dado por:

$$\frac{\partial ln(w)}{\partial exp} = \hat{\gamma}_2 + 2\hat{\gamma}_3 Exp$$

- En orden de encontrar el punto de inflexión, la derivada de la ecuación mostrada lineas arriba debe ser igualada a cero, la cual resolvera el estado estacionario

$$Exp_{estacionario} = \frac{\hat{\gamma}_2}{2\hat{\gamma}_3}$$

$$\hat{\triangle} = -\frac{\hat{\gamma}_2}{2\hat{\gamma}_3}$$

Regresión lineal Multivariada: Método Delta

- El efecto marginal para Exp esta dado por:

$$\frac{\partial ln(w)}{\partial exp} = \hat{\gamma}_2 + 2\hat{\gamma}_3 Exp$$

- En orden de encontrar el punto de inflexión, la derivada de la ecuación mostrada lineas arriba debe ser igualada a cero, la cual resolvera el estado estacionario

$$Exp_{estacionario} = \frac{\hat{\gamma}_2}{2\hat{\gamma}_3}$$

$$\hat{\triangle} = -\frac{\hat{\gamma}_2}{2\hat{\gamma}_3}$$

- Nota: esta última expresión es una función no lineal de los dos estimadores de OLS

Regresión lineal Multivariada: Método Delta

- Método utilizado para poder analizar parámetros no-lineales, donde se cálcula varianzas no lineal.
- Supuesto, nosotros deseamos testear la proporción que el punto de cambio en el logaritmo de salario y experiencia es 20 años
- Esto es formalmente expresado a través de la hipótesis:

$$H_o : \triangle = 20 \text{ vs } H_a : \triangle \neq 20$$

- La varianza muestral para el estimado es derivado usando el método delta y esta dado por:

$$Var\left(\hat{\triangle}\right) = \left(\frac{\partial\hat{\triangle}}{\partial\hat{\gamma}_2}\right)^2 Var\left(\hat{\gamma}_2\right) + \left(\frac{\partial\hat{\triangle}}{\partial\hat{\gamma}_3}\right)^2 Var\left(\hat{\gamma}_3\right)\cdots$$

$$= \cdots + 2\left(\frac{\partial\hat{\triangle}}{\partial\hat{\gamma}_2}\right)\left(\frac{\partial\hat{\triangle}}{\partial\hat{\gamma}_3}\right) Cov\left(\hat{\gamma}_2, \hat{\gamma}_3\right)$$

- Esta expresión provee una aproximación asintótica de la varianza muestral para el punto en estimación

Test de Heterocedasticidad

Test de Heterocedasticidad

- La prueba de **White Koenker** es util para poder detectar la presencia de heterocedasticidad
  1. El test parte de estimar los residuos de MCO (OLS) de la regresión original
  2. Luego de obtener el residuo podemos elevar al cuadrado (definido como: $\hat{\varepsilon}_i$)
  3. Se realiza un regresión auxiliar que se especifica con los rergesores de la ecuación original:

$$\hat{\varepsilon}_i = \phi + \delta_1 Educ_i + \delta_2 Exp + \delta_3 Exp_i^2 + \delta_4 Male_i + \kappa' X_i + \xi_i \tag{2}$$

- Por ultimo, se calcula , n x $R^2 \sim \chi^2_{k-1}$

Test de Heterocedasticidad

- La prueba de **White Koenker** es util para poder detectar la presencia de heterocedasticidad
  1. El test parte de estimar los residuos de MCO (OLS) de la regresión original
  2. Luego de obtener el residuo podemos elevar al cuadrado (definido como: $\hat{\varepsilon}_i$)
  3. Se realiza un regresión auxiliar que se especifica con los rergesores de la ecuación original:

$$\hat{\varepsilon}_i = \phi + \delta_1 Educ_i + \delta_2 Exp + \delta_3 Exp_i^2 + \delta_4 Male_i + \kappa' X_i + \xi_i \tag{2}$$

- Por ultimo, se calcula , n x $R^2 \sim \chi^2_{k-1}$
- Esta última expresión es el Multiplicador de lagrange (LM) para heterocedasticidad

Test de Heterocedasticidad

- La prueba de **White Koenker** es util para poder detectar la presencia de heterocedasticidad
  1. El test parte de estimar los residuos de MCO (OLS) de la regresión original
  2. Luego de obtener el residuo podemos elevar al cuadrado (definido como: $\hat{\varepsilon}_i$)
  3. Se realiza un regresión auxiliar que se especifica con los rergesores de la ecuación original:

$$\hat{\varepsilon}_i = \phi + \delta_1 Educ_i + \delta_2 Exp + \delta_3 Exp_i^2 + \delta_4 Male_i + \kappa' X_i + \xi_i \tag{2}$$

- Por ultimo, se calcula , n x $R^2 \sim \chi^2_{k-1}$
- Esta última expresión es el Multiplicador de lagrange (LM) para heterocedasticidad
- Nota: **El test de LM asume que los errores del logaritmo de salarios son i.i.d y no distribuido normalmente**

Test para Heterocedasticidad

Test para Heterocedasticidad

- La expresión convencional para la matrix de varianza y covarianza de OLS esta dado por:

$$var\left(\hat{\beta}\right) = \sigma^2 \left(X'X\right)^{-1} \tag{3}$$

Test para Heterocedasticidad

- La expresión convencional para la matrix de varianza y covarianza de OLS esta dado por:

$$var\left(\hat{\beta}\right) = \sigma^2 \left(X'X\right)^{-1} \tag{3}$$

- Sin embargo, ante la presencia de heterocedasticidad, se tiene:

$$var\left(\hat{\beta}\right) = \left(X'X\right)^{-1} X'E\left(uu'\right)\left(X'X\right)^{-1} \tag{4}$$

Test para Heterocedasticidad

- La expresión convencional para la matrix de varianza y covarianza de OLS esta dado por:

$$var\left(\hat{\beta}\right) = \sigma^2 \left(X'X\right)^{-1} \tag{3}$$

- Sin embargo, ante la presencia de heterocedasticidad, se tiene:

$$var\left(\hat{\beta}\right) = \left(X'X\right)^{-1} X'E\left(uu'\right)\left(X'X\right)^{-1} \tag{4}$$

- Donde, $E\left(uu'\right) = \Omega$, y Ω es una matriz diagonal de orden n x n

Test para Heterocedasticidad

- La expresión convencional para la matrix de varianza y covarianza de OLS esta dado por:

$$var\left(\hat{\beta}\right) = \sigma^2 \left(X'X\right)^{-1} \tag{3}$$

- Sin embargo, ante la presencia de heterocedasticidad, se tiene:

$$var\left(\hat{\beta}\right) = \left(X'X\right)^{-1} X'E\left(uu'\right)\left(X'X\right)^{-1} \tag{4}$$

- Donde, $E\left(uu'\right) = \Omega$, y Ω es una matriz diagonal de orden n x n
- La expresión puede ser re-expresada como:

$$var\left(\hat{\beta}\right) = \left(X'X\right)^{-1} X'\Omega\left(X'X\right)^{-1} \tag{5}$$

Test para Heterocedasticidad

Test para Heterocedasticidad

- En el caso de 5 x 5 (por ejemplo, donde solo existe 5 observaciones) , la matriz Ω puede ser expresada como:

Test para Heterocedasticidad

- En el caso de 5 x 5 (por ejemplo, donde solo existe 5 observaciones) , la matriz Ω puede ser expresada como:

$$\Omega = \begin{bmatrix} \sigma_1^2 & 0 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & 0 & 0 & 0 \\ 0 & 0 & \sigma_3^2 & 0 & 0 \\ 0 & 0 & 0 & \sigma_4^2 & 0 \\ 0 & 0 & 0 & 0 & \sigma_5^2 \end{bmatrix}$$

Test para Heterocedasticidad

- En el caso de 5 x 5 (por ejemplo, donde solo existe 5 observaciones) , la matriz Ω puede ser expresada como:

$$\Omega = \begin{bmatrix} \sigma_1^2 & 0 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & 0 & 0 & 0 \\ 0 & 0 & \sigma_3^2 & 0 & 0 \\ 0 & 0 & 0 & \sigma_4^2 & 0 \\ 0 & 0 & 0 & 0 & \sigma_5^2 \end{bmatrix}$$

- el termino σ^2 determinado en la diagonal son desconocidos

Test para Heterocedasticidad

- En el caso de 5 x 5 (por ejemplo, donde solo existe 5 observaciones) , la matriz Ω puede ser expresada como:

$$\Omega = \begin{bmatrix} \sigma_1^2 & 0 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & 0 & 0 & 0 \\ 0 & 0 & \sigma_3^2 & 0 & 0 \\ 0 & 0 & 0 & \sigma_4^2 & 0 \\ 0 & 0 & 0 & 0 & \sigma_5^2 \end{bmatrix}$$

- el termino σ^2 determinado en la diagonal son desconocidos
- Por tanto, como se puede estimar los elementos en la diagonal principal en la matriz Ω?

Test para Heterocedasticidad

Test para Heterocedasticidad

- Eicker-White-Huber (EWH) sugieren reemplazar los elementos de la diagonal por los residuos al cuadrado de la regresión de OLS.

Test para Heterocedasticidad

- Eicker-White-Huber (EWH) sugieren reemplazar los elementos de la diagonal por los residuos al cuadrado de la regresión de OLS.
- Los residuos al cuadrado (pro ejemplo, $sq_{residual_i}$) actual como proxy empiricas para la varianza del error, el cual es asumida que tiene variación a través de las n observaciones dado la heterocedasticidad.

Test para Heterocedasticidad

- Eicker-White-Huber (EWH) sugieren reemplazar los elementos de la diagonal por los residuos al cuadrado de la regresión de OLS.
- Los residuos al cuadrado (pro ejemplo, $sq_{residual_i}$) actual como proxy empiricas para la varianza del error, el cual es asumida que tiene variación a través de las n observaciones dado la heterocedasticidad.
- La matriz Ω podria ser escrita empiricamente como:

$$\hat{\Omega} = \begin{bmatrix} sq_{residual_1} & 0 & 0 & 0 & 0 \\ 0 & sq_{residual_2} & 0 & 0 & 0 \\ 0 & 0 & sq_{residual_3} & 0 & 0 \\ 0 & 0 & 0 & sq_{residual_4} & 0 \\ 0 & 0 & 0 & 0 & sq_{residual_5} \end{bmatrix}$$

Test para Heterocedasticidad

Test para Heterocedasticidad

- La expresión de la formula puede ser re-escrita como:

Test para Heterocedasticidad

- La expresión de la formula puede ser re-escrita como:

$$var\left(\hat{\beta}\right) = \left(X'X\right)^{-1} X'\hat{\Omega}\left(X'X\right)^{-1} \tag{6}$$

Test para Heterocedasticidad

- La expresión de la formula puede ser re-escrita como:

$$var\left(\hat{\beta}\right) = \left(X'X\right)^{-1} X'\hat{\Omega}\left(X'X\right)^{-1} \tag{6}$$

- Esta expresión puede ahora ser estimada usando los residuos al cuadrado de la forma original de la regresión propuesta.

Test para Heterocedasticidad

- La expresión de la formula puede ser re-escrita como:

$$var\left(\hat{\beta}\right) = \left(X'X\right)^{-1} X'\hat{\Omega}\left(X'X\right)^{-1} \tag{6}$$

- Esta expresión puede ahora ser estimada usando los residuos al cuadrado de la forma original de la regresión propuesta.
- Por tanto, nosotros conocemos tener una matriz estimable de matriz de varianzas-covarianzas.

Test para Heterocedasticidad

Test para Heterocedasticidad

- Esta matriz de varianza-covarianza proporciona un estimador consistente de la matriz independientemente de la estructura de heterocedasticidad.

Test para Heterocedasticidad

- Esta matriz de varianza-covarianza proporciona un estimador consistente de la matriz independientemente de la estructura de heterocedasticidad.
- A veces se la denomina la matriz de varianza-covarianza consistente con heterocedasdicidad (HCVC, heteroscedasdicity-consistent variance-covariance, siglas en ingles).

Test para Heterocedasticidad

- Esta matriz de varianza-covarianza proporciona un estimador consistente de la matriz independientemente de la estructura de heterocedasticidad.
- A veces se la denomina la matriz de varianza-covarianza consistente con heterocedasdicidad (HCVC, heteroscedasdicity-consistent variance-covariance, siglas en ingles).
- Por tanto, La corrección solo ajusta la matriz de varianza-covarianza

Test para Heterocedasticidad

- Esta matriz de varianza-covarianza proporciona un estimador consistente de la matriz independientemente de la estructura de heterocedasticidad.
- A veces se la denomina la matriz de varianza-covarianza consistente con heterocedasdicidad (HCVC, heteroscedasdicity-consistent variance-covariance, siglas en ingles).
- Por tanto, La corrección solo ajusta la matriz de varianza-covarianza
- Esta matriz de varianza-covarianza consistente de heterocedasticidad solo es válida asintóticamente $\cdots$ por lo que su uso en muestras pequeñas es cuestionable.

Test para Heterocedasticidad

- Esta matriz de varianza-covarianza proporciona un estimador consistente de la matriz independientemente de la estructura de heterocedasticidad.
- A veces se la denomina la matriz de varianza-covarianza consistente con heterocedasdicidad (HCVC, heteroscedasdicity-consistent variance-covariance, siglas en ingles).
- Por tanto, La corrección solo ajusta la matriz de varianza-covarianza
- Esta matriz de varianza-covarianza consistente de heterocedasticidad solo es válida asintóticamente $\cdots$ por lo que su uso en muestras pequeñas es cuestionable.
- La HCVC se conoce como un estimador robusto de la matriz de varianza-covarianza de MCO.

Hipotesis de Test usando HCVC

Hipotesis de Test usando HCVC

- En presencia de heterocedasticidad, la matrix de varianca-covarianza robusta puede ser usado para testear cualquier hipotesis que considere un solo parametro, recordando la prueba: t-test.

Hipotesis de Test usando HCVC

- En presencia de heterocedasticidad, la matrix de varianca-covarianza robusta puede ser usado para testear cualquier hipotesis que considere un solo parametro, recordando la prueba: t-test.
- El t-test puede ser un t-test asintotico en este caso.

Hipotesis de Test usando HCVC

- En presencia de heterocedasticidad, la matrix de varianca-covarianza robusta puede ser usado para testear cualquier hipotesis que considere un solo parametro, recordando la prueba: t-test.
- El t-test puede ser un t-test asintotico en este caso.
- Por tanto, Que pasa con las pruebas conjuntas cuando se utiliza el HCVC

Hipotesis de Test usando HCVC

- En presencia de heterocedasticidad, la matrix de varianca-covarianza robusta puede ser usado para testear cualquier hipotesis que considere un solo parametro, recordando la prueba: t-test.
- El t-test puede ser un t-test asintotico en este caso.
- Por tanto, Que pasa con las pruebas conjuntas cuando se utiliza el HCVC
- El uso de la prueba F para probar proposiciones que incorporan restricciones conjuntas no es válido en presencia de heterocedasticidad.

Hipotesis de Test usando HCVC

- En presencia de heterocedasticidad, la matrix de varianca-covarianza robusta puede ser usado para testear cualquier hipotesis que considere un solo parametro, recordando la prueba: t-test.
- El t-test puede ser un t-test asintotico en este caso.
- Por tanto, Que pasa con las pruebas conjuntas cuando se utiliza el HCVC
- El uso de la prueba F para probar proposiciones que incorporan restricciones conjuntas no es válido en presencia de heterocedasticidad.
- En consecuencia, una prueba alternativa es proporcionada en el test de Wald, que se puede calcular utilizando la matriz de covarianza de varianza robusta.

Hipotesis de Test usando HCVC

Hipotesis de Test usando HCVC

- Por ejemplo, se desea testear la significancia conjunta de dos efectos estimados, el cual esta descrito con los sub-indices **m** y **s**.

Hipotesis de Test usando HCVC

- Por ejemplo, se desea testear la significancia conjunta de dos efectos estimados, el cual esta descrito con los sub-indices **m** y **s**.
- La hipotesis nula esta dada por:

$$H_o : \gamma_m = \gamma_s = 0 \text{ vs } H_a : H_o \text{ es falso}$$

Hipotesis de Test usando HCVC

- Por ejemplo, se desea testear la significancia conjunta de dos efectos estimados, el cual esta descrito con los sub-indices **m** y **s**.
- La hipotesis nula esta dada por:

$$H_o : \gamma_m = \gamma_s = 0 \text{ vs } H_a : H_o \text{ es falso}$$

- Se define entonces, un vector fila de dimensión 1 x 2 a través de sus estimadores $\begin{bmatrix} \hat{\gamma}_m & \hat{\gamma}_s \end{bmatrix}$

Hipotesis de Test usando HCVC

- Por ejemplo, se desea testear la significancia conjunta de dos efectos estimados, el cual esta descrito con los sub-indices **m** y **s**.
- La hipotesis nula esta dada por:

$$H_o : \gamma_m = \gamma_s = 0 \text{ vs } H_a : H_o \text{ es falso}$$

- Se define entonces, un vector fila de dimensión 1 x 2 a través de sus estimadores $\begin{bmatrix} \hat{\gamma}_m & \hat{\gamma}_s \end{bmatrix}$
- La matrix de HCVC es una sub-matriz definida como:

$$\hat{V}_{robust} = \begin{bmatrix} Var(\hat{\gamma}_m) & Cov(\hat{\gamma}_m, \hat{\gamma}_s) \\ Cov(\hat{\gamma}_m, \hat{\gamma}_s) & Var(\hat{\gamma}_s) \end{bmatrix}$$

Hipotesis de Test usando HCVC

Hipotesis de Test usando HCVC

- La forma de la matriz de test de Wald esta formulada como:

$$Wald = \left[(\hat{\gamma}_m - \gamma_m) \quad (\hat{\gamma}_s - \gamma_s) \right] \hat{V}_{robust}^{-1} \left[(\hat{\gamma}_m - \gamma_m) \quad (\hat{\gamma}_s - \gamma_s) \right]'$$

Hipotesis de Test usando HCVC

- La forma de la matriz de test de Wald esta formulada como:

$$Wald = \left[(\hat{\gamma}_m - \gamma_m) \quad (\hat{\gamma}_s - \gamma_s) \right] \hat{V}_{robust}^{-1} \left[(\hat{\gamma}_m - \gamma_m) \quad (\hat{\gamma}_s - \gamma_s) \right]'$$

- Bajo la hipotesisi nula planteada, $H_o : \gamma_m = \gamma_s = 0$, entonces

$$Wald = \left[\hat{\gamma}_m \quad \hat{\gamma}_s \right] \hat{V}_{robust}^{-1} \left[\hat{\gamma}_m \quad \hat{\gamma}_s \right]' \sim \chi_2^2$$

Hipotesis de Test usando HCVC

- La forma de la matriz de test de Wald esta formulada como:

$$Wald = \left[(\hat{\gamma}_m - \gamma_m) \quad (\hat{\gamma}_s - \gamma_s)\right] \hat{V}_{robust}^{-1} \left[(\hat{\gamma}_m - \gamma_m) \quad (\hat{\gamma}_s - \gamma_s)\right]'$$

- Bajo la hipotesisi nula planteada, $H_o : \gamma_m = \gamma_s = 0$, entonces

$$Wald = \left[\hat{\gamma}_m \quad \hat{\gamma}_s\right] \hat{V}_{robust}^{-1} \left[\hat{\gamma}_m \quad \hat{\gamma}_s\right]' \sim \chi_2^2$$

- Este test estadistico esta distribuido bajo una chi-cuadrado con dos grados de libertad.

Parte Practica

Los archivos de la parte practica lo puede descargar del siguiente QR