Econometría Aplicada
Modelos de Regresion Lineal
Semana 1
Aplicación

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

Twitter: @edutoleraymondi

Educate Perú

7 de agosto de 2022

Contenido

Introducción

EDÚCATE PERÚ CONSULTORES

Introducción

- La consultora **Marilyn Loden** , quien acuñó el termino **techo de cristal** en

Introducción

Introducción

- En 1991, la Comisión de **Techos de Vidrio (glass ceiling)** del Departamento de Trabajo de EE. UU. (1991-1996) definió el **techo de vidrio** como:

Introducción

- En 1991, la Comisión de **Techos de Vidrio (glass ceiling)** del Departamento de Trabajo de EE. UU. (1991-1996) definió el **techo de vidrio** como:
- ··· esas barreras artificiales basadas en prejuicios organizacionales o de actitud que impiden que las personas calificadas asciendan en su organización a puestos de nivel gerencial.

Introducción

- En 1991, la Comisión de **Techos de Vidrio (glass ceiling)** del Departamento de Trabajo de EE. UU. (1991-1996) definió el **techo de vidrio** como:
- $\cdots$ esas barreras artificiales basadas en prejuicios organizacionales o de actitud que impiden que las personas calificadas asciendan en su organización a puestos de nivel gerencial.
- El concepto de **techo de cristal** se convirtió en el objeto de un estudio de investigación sistemático desde finales de los años ochenta.

Introducción

EDÚCATE PERÚ CONSULTORES

- En 1991, la Comisión de **Techos de Vidrio (glass ceiling)** del Departamento de Trabajo de EE. UU. (1991-1996) definió el **techo de vidrio** como:
- $\cdots$ esas barreras artificiales basadas en prejuicios organizacionales o de actitud que impiden que las personas calificadas asciendan en su organización a puestos de nivel gerencial.
- El concepto de **techo de cristal** se convirtió en el objeto de un estudio de investigación sistemático desde finales de los años ochenta.
- Los sociólogos estuvieron a la vanguardia de las primeras investigaciones y (algunos) economistas se interesaron a fines de la década de 1990.

Introducción

- En 1991, la Comisión de **Techos de Vidrio (glass ceiling)** del Departamento de Trabajo de EE. UU. (1991-1996) definió el **techo de vidrio** como:
- $\cdots$ esas barreras artificiales basadas en prejuicios organizacionales o de actitud que impiden que las personas calificadas asciendan en su organización a puestos de nivel gerencial.
- El concepto de **techo de cristal** se convirtió en el objeto de un estudio de investigación sistemático desde finales de los años ochenta.
- Los sociólogos estuvieron a la vanguardia de las primeras investigaciones y (algunos) economistas se interesaron a fines de la década de 1990.
- El enfoque original de los economistas en esta área era la promoción, no el pago.

Introducción

- En 1991, la Comisión de **Techos de Vidrio (glass ceiling)** del Departamento de Trabajo de EE. UU. (1991-1996) definió el **techo de vidrio** como:
- $\cdots$ esas barreras artificiales basadas en prejuicios organizacionales o de actitud que impiden que las personas calificadas asciendan en su organización a puestos de nivel gerencial.
- El concepto de **techo de cristal** se convirtió en el objeto de un estudio de investigación sistemático desde finales de los años ochenta.
- Los sociólogos estuvieron a la vanguardia de las primeras investigaciones y (algunos) economistas se interesaron a fines de la década de 1990.
- El enfoque original de los economistas en esta área era la promoción, no el pago.
- A principios de la década de 2000, las regresiones cuantílicas se utilizaban de forma rutinaria en esta investigación para centrarse en la remuneración.

Introducción

Introducción

- La socióloga Catherine Berheide, quien acuñó la frase **piso pegajoso** (sticky floor) en 1992

Introducción

EDÚCATE PERÚ CONSULTORES

- La socióloga Catherine Berheide, quien acuñó la frase **piso pegajoso** (sticky floor) en 1992
- El **piso pegajoso** se refiere a los casos en los que las mujeres ocupan puestos de baja categoría, bajos salarios y de baja movilidad en el mercado laboral.

Introducción

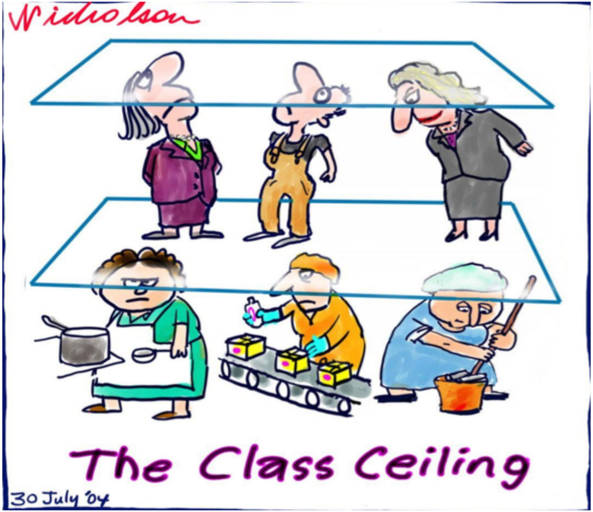
- La socióloga Catherine Berheide, quien acuñó la frase **piso pegajoso** (sticky floor) en 1992
- El **piso pegajoso** se refiere a los casos en los que las mujeres ocupan puestos de baja categoría, bajos salarios y de baja movilidad en el mercado laboral.
- Donde se dice:

Introducción

- La socióloga Catherine Berheide, quien acuñó la frase **piso pegajoso** (sticky floor) en 1992
- El **piso pegajoso** se refiere a los casos en los que las mujeres ocupan puestos de baja categoría, bajos salarios y de baja movilidad en el mercado laboral.
- Donde se dice:
- $\cdots$ la mayoría de las mujeres deberían tener la suerte de tener el techo de cristal como su problema $\cdots$ .muchas están atascadas en el piso pegajoso.

Introducción

EDÚCATE PERÚ CONSULTORES



Descripción de Información

- La información que se utilizará es proveniente de la base de datos de la Encuesta Nacional de Hogares (ENAHO). Se procesa la base de datos del modulo 300 y 500 donde se analizará los ingresos mensuales de los trabajadores.

Cuadro: Descripción de variables

Variables	Descripción
lnr6	Logaritmo ingreso mensual (Soles)
r6	Ingreso mensual (Soles)
reduca	Años de educación
rmujer	==1, mujer
rexper	Años de experiencia
rexpersq	Años de experiencia cuadrado
rpareja	== 1, casado
rsoltero	== 1, soltero

Pregunta 1

Cuadro: Estadisticas descriptivas

	Trabajadores	Promedio	Mediana	Min.	Max.	Std
r6	26737	1,396.86	944.75	81.50	52,063.25	1,705
reduca	26737	8.87	11.00	0.00	18.00	5
rmujer	26737	0.29	0.00	0.00	1.00	0
rexper	26737	33.47	33.00	2.00	84.00	15
rexpersq	26737	1,337.00	1,089.00	4.00	7,056.00	1,078
rpareja	26737	0.64	1.00	0.00	1.00	0
rsoltero	26737	0.09	0.00	0.00	1.00	0

Fuente: INEI - 2020.
Elaboracion: Autor

Pregunta 1

Cuadro: Estadisticas descriptivas

	Trabajadores	Promedio	Mediana	Min.	Max.	Std
r6	26737	1,396.86	944.75	81.50	52,063.25	1,705
reduca	26737	8.87	11.00	0.00	18.00	5
rmujer	26737	0.29	0.00	0.00	1.00	0
rexper	26737	33.47	33.00	2.00	84.00	15
rexpersq	26737	1,337.00	1,089.00	4.00	7,056.00	1,078
rpareja	26737	0.64	1.00	0.00	1.00	0
rsoltero	26737	0.09	0.00	0.00	1.00	0

Fuente: INEI - 2020.
Elaboracion: Autor

- Alrededor de 29 % de los trabajadores jefes de hogares son mujeres.
- En promedio, la experiencia de los trabajadores es de 33 años y sus años de educación son de 9 años.
- El 64 % de los trabajadores tiene una condición civil de pareja (casado(a) o conviviente).

Pregunta 2

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

Pregunta 3

Pregunta 3

- Se realizará la estimación de la ecuación de salarios de los trabajadores peruanos utilizando la información 2021 (ENAHO)

Pregunta 3

- Se realizará la estimación de la ecuación de salarios de los trabajadores peruanos utilizando la información 2021 (ENAHO)
- Para realizar el análisis se propone dos ecuaciones:

$$lnr6_i = \beta_0 + \beta_1 reduca_i + \mu_i \quad (1)$$

$$lnr6_i = \beta_0 + \beta_1 reduca_i + \beta_2 rexper_i + \beta_3 rexpersq_i$$
$$+\beta_4 rpareja_i + \beta_5 rsoltero_i + \phi rmujer_i + \mu_i (2)$$

Pregunta 3

1 Estime la regresión de la ecuación 1 y 2 usando OLS. Programación en STATA
   - Use el nivel de significancia de 0.05 para testear la presencia de heterocedasticidad.
   - ¿Qué es lo que usted concluye?

Pregunta 3

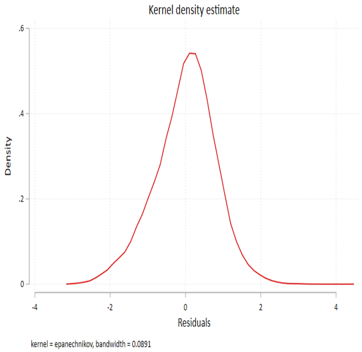
Cuadro: Log (Ingresos) -Mincer

	(1)	(2)
reduca	0.10***	0.09***
	(0.00)	(0.00)
rmujer		-0.29***
		(0.01)
rexper		0.02***
		(0.00)
rexpersq		-0.00***
		(0.00)
rpareja		0.06***
		(0.01)
rsoltero		-0.07***
		(0.02)
Constant	5.90***	5.87***
	(0.01)	(0.03)
Observaciones	26737	26737
Adj. R^2	0.242	0.287
Region FE	66728.5	65099.2

Errores estandar en parentesis ()
Fuente: INEI - 2021.
Elaboracion: Autor

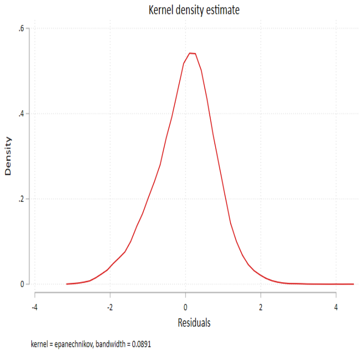
Pregunta 3

Figura 1:



Pregunta 3

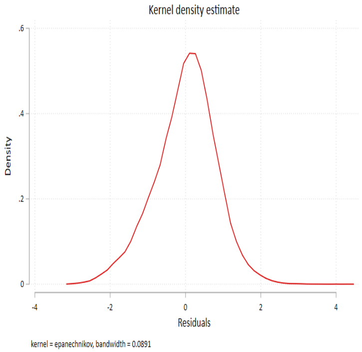
Figura 1:



- La distribución es muy puntiaguda con colas más alargadas en comparación con una distribución normal.

Pregunta 3

Figura 1:



- La distribución es muy puntiaguda con colas más alargadas en comparación con una distribución normal.
- La prueba confirma que las dos proposiciones clave que gobiernan la normalidad (es decir, simetría y mesokurtosis) son ambas decididamente rechazadas por los datos en este caso.

Pregunta 3

El modelo de regresion lineal, se estimo sus errores y se realizo un grafico de distribición (lado izquierdo) y caja (lado derecho)

Figura 1:

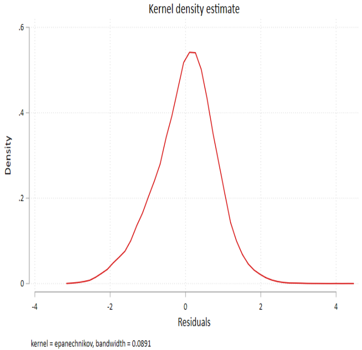
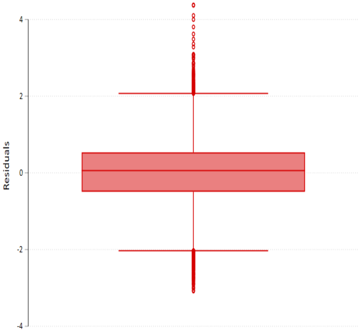


Figura 2:



Pregunta 3

- ¿Cuál es la diferencia entre los test Breusch-Pagan/Cook-Weisberg y el test de White/Koenker ?
- El Breusch-Pagan/Cook-Weisberg test asume que los errores de la ecuación original se distribuyen de manera normal.
- Por otro lado, el test de White/Koenker solo asume que los errores de la ecuación original son identica e independientemente distribuidos (i.i.d).
- Por lo tanto, es útil probar los residuos del modelo de regresión de la ecuación original para determinar la **normalidad** para decidir cuál de estas pruebas de heterocedasticidad usar.
- El comando relevante en STATA, dada la normalidad es violada, es:

hettest rmujer reduca rexper rexpersq rpareja rsoltero, *iid*

Pregunta 3

- El test de Koenker/White heteroscedasticity requiere la estimación del siguiente modelo de regresión auxiliar :

$$\hat{\varepsilon}_i^2 = \delta_0 + \delta_1 reduca_i + \delta_2 rexper_i + \delta_3 rexper_i^2 +$$
$$+ \delta_4 rpareja_i + \delta_5 rsoltero_i + \phi rmujer_i + \xi_i (3)$$

- Hipotesis

$$H_o\colon \delta_0 = \delta_1 = \delta_2 = \delta_3 = \delta_4 = \delta_5 = \phi = 0$$
$$\text{vs } H_a\colon H_o \text{ no es verdad}$$

- Bajo la hipotesisi Nula , el test de LM es definido como

$$n \times R^2 \sim \chi^2_{k-1}$$

```
Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: reduca rmujer rexper rexpersq rpareja rsoltero

         chi2(6)      =   224.30
         Prob > chi2  =   0.0000
```

Pregunta 3

- El test de Koenker/White heteroscedasticity requiere la estimación del siguiente modelo de regresión auxiliar :

$$\hat{\varepsilon}_i^2 = \delta_0 + \delta_1 reduca_i + \delta_2 rexper_i + \delta_3 rexper_i^2 +$$
$$+ \delta_4 rpareja_i + \delta_5 rsoltero_i + \phi rmujer_i + \xi_i (3)$$

- Hipotesis

$$H_o\colon \delta_0 = \delta_1 = \delta_2 = \delta_3 = \delta_4 = \delta_5 = \phi = 0$$
$$\text{vs } H_a\colon H_o \text{ no es verdad}$$

- Bajo la hipotesisi Nula , el test de LM es definido como

$$n \times R^2 \sim \chi^2_{k-1}$$

```
Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: reduca rmujer rexper rexpersq rpareja rsoltero

         chi2(6)      =   224.30
         Prob > chi2  =   0.0000
```

- La hipotesis nula de homocedasticidad tendra que se analizada para el modelo de log(wage)

Pregunta 3

- El test de Koenker/White heteroscedasticity requiere la estimación del siguiente modelo de regresión auxiliar :

$$\hat{\varepsilon}_i^2 = \delta_0 + \delta_1 reduca_i + \delta_2 rexper_i + \delta_3 rexper_i^2 +$$
$$+ \delta_4 rpareja_i + \delta_5 rsoltero_i + \phi rmujer_i + \xi_i (3)$$

- Hipotesis

$$H_o\colon \delta_0 = \delta_1 = \delta_2 = \delta_3 = \delta_4 = \delta_5 = \phi = 0$$
$$\text{vs } H_a\colon H_o \text{ no es verdad}$$

- Bajo la hipotesisi Nula , el test de LM es definido como

$$n \times R^2 \sim \chi^2_{k-1}$$

```
Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: reduca rmujer rexper rexpersq rpareja rsoltero

         chi2(6)      =   224.30
         Prob > chi2  =   0.0000
```

- La hipotesis nula de homocedasticidad tendra que se analizada para el modelo de log(wage)

Pregunta 3

Cuadro: Ecuación de μ^2

	(4)	
reduca	0.09***	(0.00)
rmujer	-0.29***	(0.01)
rexper	0.02***	(0.00)
rexpersq	-0.00***	(0.00)
rpareja	0.06***	(0.01)
rsoltero	-0.07***	(0.02)
Constant	5.87***	(0.03)
Observaciones	26737	
R^2	0.287	
Adj. R^2	0.287	
R^2 x n		
Region FE	√	

Errores estandar en parentesis ()
Fuente: INEI - 2021.
Elaboracion: Autor

Pregunta 3

Cuadro: Ecuación de μ^2

	(4)	
reduca	0.09***	(0.00)
rmujer	-0.29***	(0.01)
rexper	0.02***	(0.00)
rexpersq	-0.00***	(0.00)
rpareja	0.06***	(0.01)
rsoltero	-0.07***	(0.02)
Constant	5.87***	(0.03)
Observaciones	26737	
R^2	0.287	
Adj. R^2	0.287	
R^2 x n		
Region FE	$\surd$	

Errores estandar en parentesis ()
Fuente: INEI - 2021.
Elaboracion: Autor

- Por tanto, $n \times R^2$ es igual a 224.30

Pregunta 3

Cuadro: Ecuación de μ^2

	(4)	
reduca	0.09***	(0.00)
rmujer	-0.29***	(0.01)
rexper	0.02***	(0.00)
rexpersq	-0.00***	(0.00)
rpareja	0.06***	(0.01)
rsoltero	-0.07***	(0.02)
Constant	5.87***	(0.03)
Observaciones	26737	
R^2	0.287	
Adj. R^2	0.287	
R^2 x n		
Region FE	$\surd$	

Errores estandar en parentesis ()
Fuente: INEI - 2021.
Elaboracion: Autor

- Por tanto, $n \times R^2$ es igual a 224.30
- Dado un valor de significancia de 0.05, la distribución χ^2_{k-1} arroja el valor de 12.59 (χ^2_6)

Pregunta 3

EDÚCATE PERÚ CONSULTORES

Cuadro: Ecuación de μ^2

	(4)	
reduca	0.09***	(0.00)
rmujer	-0.29***	(0.01)
rexper	0.02***	(0.00)
rexpersq	-0.00***	(0.00)
rpareja	0.06***	(0.01)
rsoltero	-0.07***	(0.02)
Constant	5.87***	(0.03)
Observaciones	26737	
R^2	0.287	
Adj. R^2	0.287	
R^2 x n		
Region FE	$\surd$	

Errores estandar en parentesis ()
Fuente: INEI - 2021.
Elaboracion: Autor

- Por tanto, $n \times R^2$ es igual a 224.30
- Dado un valor de significancia de 0.05, la distribución χ^2_{k-1} arroja el valor de 12.59 (χ^2_6)
- Evidenciandoque se rehaza la H_o(homocedasticidad)

Pregunta 3

Cuadro: Ecuación de μ^2

	(4)	
reduca	0.09***	(0.00)
rmujer	-0.29***	(0.01)
rexper	0.02***	(0.00)
rexpersq	-0.00***	(0.00)
rpareja	0.06***	(0.01)
rsoltero	-0.07***	(0.02)
Constant	5.87***	(0.03)
Observaciones	26737	
R^2	0.287	
Adj. R^2	0.287	
R^2 x n		
Region FE	$\surd$	

Errores estandar en parentesis ()
Fuente: INEI - 2021.
Elaboracion: Autor

- Por tanto, n x R^2 es igual a 224.30
- Dado un valor de significancia de 0.05, la distribución χ^2_{k-1} arroja el valor de 12.59 (χ^2_6)
- Evidenciandoque se rehaza la H_o(homocedasticidad)
- Concluyendo que existe presencia de **heterocedasticidad**

Pregunta 3

Cuadro: Ecuación de μ^2

	(4)	
reduca	0.09***	(0.00)
rmujer	-0.29***	(0.01)
rexper	0.02***	(0.00)
rexpersq	-0.00***	(0.00)
rpareja	0.06***	(0.01)
rsoltero	-0.07***	(0.02)
Constant	5.87***	(0.03)
Observaciones	26737	
R^2	0.287	
Adj. R^2	0.287	
R^2 x n		
Region FE	$\surd$	

Errores estandar en parentesis ()
Fuente: INEI - 2021.
Elaboracion: Autor

- Por tanto, $n \times R^2$ es igual a 224.30
- Dado un valor de significancia de 0.05, la distribución χ^2_{k-1} arroja el valor de 12.59 (χ^2_6)
- Evidenciandoque se rehaza la H_o(homocedasticidad)
- Concluyendo que existe presencia de **heterocedasticidad**
- Se necesita corregir la matriz de varianza-covarianza usando la opción robust antes del análisis

Pregunta 4

EDÚCATE PERÚ CONSULTORES

2 Usando el t-test asintótico y al nivel de significancia de 0.05 se le pide testear la proposición: el log de salario alcanza un máximo valor alrededor de los 50 años. Reporte todos los cálculos relevantes. ¿Qué es lo que usted concluye?

Pregunta 4

EDÚCATE PERÚ CONSULTORES

Cuadro: Log (Ingresos) -Mincer Robustos

	(1)		(2)		(3)	
reduca	0.10***	(0.00)	0.10***	(0.00)	0.09***	(0.00)
rmujer			-0.34***	(0.01)	-0.29***	(0.01)
rexper					0.02***	(0.00)
rexpersq					-0.00***	(0.00)
rpareja					0.06***	(0.01)
rsoltero					-0.07***	(0.02)
Constant	5.90***	(0.01)	6.02***	(0.01)	5.87***	(0.03)
Observaciones	26737		26737		26737	
Adj. R^2	0.242		0.267		0.287	
Region FE					$\surd$	

Errores estandar en parentesis ()
Fuente: INEI - 2021.
Elaboracion: Autor

Pregunta 4

- La derivada obtenida , esta determinada como:

$$\frac{\partial ln(w)}{\partial edad} = \hat{\beta}_2 + 2\hat{\beta}_3 edad$$

- El punto de estado estacionario es cálculado al igualar en cero la derivada obtenida y despejar el valor de la variable *edad*
- Para la presente aplicación, se tiene:

$$age_{estacionario} = \frac{\hat{\beta}_2}{-2\hat{\beta}_3} = \frac{0.0582714}{2x(-0.000578)} = 49.7287 = \hat{\triangle}$$

- Por tanto, deseamos testear la proposión planetada a través de la hipoteisis

$$H_o : \triangle = 50 \text{ vs } H_a : \triangle \neq 50$$

Pregunta 4

- Usando el **Metodo Delta**, la varianza muestral esta dado por:

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

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

- Recordando:

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

- Donde, para el presente caso:

Pregunta 4

- Usando el **Metodo Delta**, la varianza muestral esta dado por:

$$Var\left(\hat{\triangle}\right) = \left(\frac{\partial\hat{\triangle}}{\partial\hat{\gamma}_{2}}\right)^{2} Var\left(\hat{\gamma}_2\right) + \left(\frac{\partial\hat{\triangle}}{\partial\hat{\gamma}_{3}}\right)^{2} Var\left(\hat{\gamma}_3\right)\cdots$$

$$= \cdots + 2\left(\frac{\partial\hat{\triangle}}{\partial\hat{\gamma}_{2}}\right)\left(\frac{\partial\hat{\triangle}}{\partial\hat{\gamma}_{3}}\right) Cov\left(\hat{\gamma}_2, \hat{\gamma}_3\right)$$

- Recordando:

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

- Donde, para el presente caso:

$$\frac{\partial\hat{\triangle}}{\partial\hat{\gamma}_2} = -\frac{1}{2\hat{\gamma}_3}$$

Pregunta 4

- Usando el **Metodo Delta**, la varianza muestral esta dado por:

$$Var\left(\hat{\triangle}\right) = \left(\frac{\partial\hat{\triangle}}{\partial\hat{\gamma}_{\ 2}}\right)^2 Var\left(\hat{\gamma}_2\right) + \left(\frac{\partial\hat{\triangle}}{\partial\hat{\gamma}_{\ 3}}\right)^2 Var\left(\hat{\gamma}_3\right)\cdots$$

$$= \cdots + 2\left(\frac{\partial\hat{\triangle}}{\partial\hat{\gamma}_{\ 2}}\right)\left(\frac{\partial\hat{\triangle}}{\partial\hat{\gamma}_{\ 3}}\right) Cov\left(\hat{\gamma}_2, \hat{\gamma}_3\right)$$

- Recordando:

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

- Donde, para el presente caso:

$$\frac{\partial\hat{\triangle}}{\partial\hat{\gamma}_2} = -\frac{1}{2\hat{\gamma}_3}$$

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

Pregunta 4

- El valor de las derivadas encontradas son calculadas:

Pregunta 4

- El valor de las derivadas encontradas son calculadas:

$$\frac{\partial \hat{\triangle}}{\partial \hat{\beta}_2} = -\frac{1}{2\hat{\beta}_3} = -22.5468$$

Pregunta 4

- El valor de las derivadas encontradas son calculadas:

$$\frac{\partial\hat{\triangle}}{\partial\hat{\beta}_2} = -\frac{1}{2\hat{\beta}_3} = -22.5468$$

$$\frac{\partial\hat{\triangle}}{\partial\hat{\beta}_3} = -\frac{\hat{\beta}_2}{2\hat{\beta}_3^2} = 69051.769$$

- Los elementos obtenidos a través de la **robust** de la matriz de varianza-covarianza son:

```
. matrix list vage

symmetric vage[2,2]
              rexper     rexpersq
  rexper   1.948e-06
rexpersq  -2.550e-08   3.635e-10
```

Pregunta 4

EDÚCATE PERÚ CONSULTORES

- Los elementos obtenidos a través de la **robust** de la matriz de varianza-covarianza son:

```
. matrix list vage

symmetric vage[2,2]
              rexper    rexpersq
  rexper   1.948e-06
rexpersq  -2.550e-08   3.635e-10
```

- Donde:

$$Var(\hat{\beta}_2) = 0.0000001948$$

$$Var(\hat{\beta}_3) = 0.00000000003635$$

$$Cov(\hat{\beta}_2, \hat{\beta}_3) = -0.000000002550$$

Pregunta 4

- Por tanto, la varianza muestral es calculada reemplazando en el método delta:

$$Var\left(\hat{\triangle}\right) = (-22.5468)^2 x 0.0000001948 + (69051.769)^2 x 0.00000000003635 + \cdots$$

$$= \cdots + 2x\left(-22.5468 x 69051.769\right)\left(-0.00000000255\right) = 0.18116328$$

- Nosotros deseamos testear la hipotesis:

$$H_o : \triangle = 50 \text{ vs } H_a : \triangle \neq 50$$

- El test-t es expresado en el presente caso como:

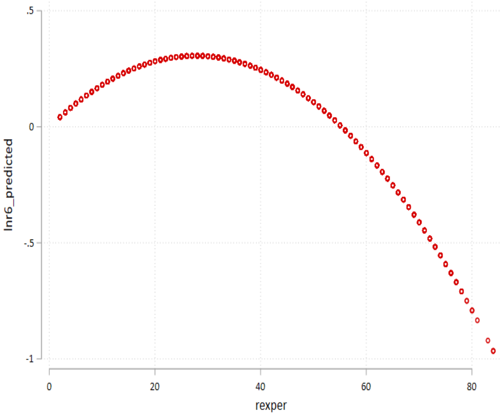
$$t = \frac{\hat{\triangle} - \triangle}{\sqrt{var(\triangle)}} = \frac{27.670369 - 50}{\sqrt{0.18116}} = -52.462195 \sim t_{\infty}$$

- **Dado el valor critico de ± 1.96, por tanto se puede rechazar la hipotesis nula** :

$$\mid -52.462195 \mid > \mid -1.96 \mid \Rightarrow \mid t_{value} \mid > \mid t_{critico} \mid$$

Pregunta 4

Figura : Perfil de salarios-edad para empleados de Perú



Pregunta 5

3 Usando el nivel de significancia de 0.05 determine si las ecuaciones de regresión logaritmica salarial estimada por separada entre hombres y mujeres son estadisticas preferibles al modelo de regresion logaritmica salarial estimado en la ecuacion 1.

Pregunta 5

- **Lo solución pasa por implementar el test de Chow**
- Estime la regresión usando la información para una sub-muestra solo para hombres y extraiga la suma al cuadrado de residuos (sum of squared residuals)
- Estime la regresión usando la información para una sub-muestra solo para mujeres y extraiga la suma al cuadrado de residuos (sum of squared residuals)
- La suma de residuos al cuadrado (**RSS**, siglas en ingles) para los dos regresiones especificas de genero, puede entonces ser resumido para el retorno de no restringidos, la cual será definida como RSS^{μ}
- Este es entonces comparado dado la suma de residuos al cuadrado de la forma del modelo restringido (tambien conocido como pooled) , la cual se define como RSS^{c}

Pregunta 5

Cuadro: Ecuación (1) - Log (Ingresos)

	Todos		Mujer		Hombre	
reduca	0.09***	(0.00)	0.08***	(0.00)	0.09***	(0.00)
rmujer	-0.29***	(0.01)				
rexper	0.02***	(0.00)	0.03***	(0.00)	0.02***	(0.00)
rexpersq	-0.00***	(0.00)	-0.00***	(0.00)	-0.00***	(0.00)
rpareja	0.06***	(0.01)	0.01	(0.03)	0.07***	(0.02)
rsoltero	-0.07***	(0.02)	-0.01	(0.03)	-0.11***	(0.03)
Constant	5.87***	(0.03)	5.46***	(0.05)	5.91***	(0.04)
Observaciones	26737		7727		19010	
Adj. R^2	0.287		0.271		0.259	
RSS	17857.9		5713.6		12117.3	

Errores estandar en parentesis
Fuente: INEI - 2021.
Elaboracion: Autor

Pregunta 3

Pregunta 3

- La cantidades de RSS de cada ecuación estimada permite construir el test-F para implementar el test de Chow:

Pregunta 3

- La cantidades de RSS de cada ecuación estimada permite construir el test-F para implementar el test de Chow:

$$F = \frac{\frac{RSS^c - RSS^u}{DF^c - DF^u}}{\frac{RSS^u}{DF^u}}$$

Pregunta 3

- La cantidades de RSS de cada ecuación estimada permite construir el test-F para implementar el test de Chow:

$$F = \frac{\frac{RSS^c - RSS^u}{DF^c - DF^u}}{\frac{RSS^u}{DF^u}}$$

- Donde, *DF* representa los grados de libertad. Por tanto, dado los resultados de las regresiones , tenemos

Pregunta 3

- La cantidades de RSS de cada ecuación estimada permite construir el test-F para implementar el test de Chow:

$$F = \frac{\frac{RSS^c - RSS^u}{DF^c - DF^u}}{\frac{RSS^u}{DF^u}}$$

- Donde, *DF* representa los grados de libertad. Por tanto, dado los resultados de las regresiones , tenemos
- Modelo restringido(Constraint)

$$RSS^u = RSS^h + RSS^m = 12,117.3 + 5,713.6 = 17830.9$$

Pregunta 3

- La cantidades de RSS de cada ecuación estimada permite construir el test-F para implementar el test de Chow:

$$F = \frac{\frac{RSS^c - RSS^u}{DF^c - DF^u}}{\frac{RSS^u}{DF^u}}$$

- Donde, DF representa los grados de libertad. Por tanto, dado los resultados de las regresiones , tenemos
- Modelo restringido(Constraint)

$$RSS^u = RSS^h + RSS^m = 12,117.3 + 5,713.6 = 17830.9$$

- Modelo no restringido (No constraint)

$$RSS^c = 17857.9$$

Pregunta 3

- La cantidades de RSS de cada ecuación estimada permite construir el test-F para implementar el test de Chow:

$$F = \frac{\frac{RSS^c - RSS^u}{DF^c - DF^u}}{\frac{RSS^u}{DF^u}}$$

- Donde, DF representa los grados de libertad. Por tanto, dado los resultados de las regresiones , tenemos
- Modelo restringido(Constraint)

$$RSS^u = RSS^h + RSS^m = 12,117.3 + 5,713.6 = 17830.9$$

- Modelo no restringido (No constraint)

$$RSS^c = 17857.9$$

- Los grados de libertad para el modelo restringido esta expresado como:

$$DF^c = 26,737 - 7 = 26,730$$

Pregunta 3

- La cantidades de RSS de cada ecuación estimada permite construir el test-F para implementar el test de Chow:

$$F = \frac{\frac{RSS^c - RSS^u}{DF^c - DF^u}}{\frac{RSS^u}{DF^u}}$$

- Donde, DF representa los grados de libertad. Por tanto, dado los resultados de las regresiones , tenemos
- Modelo restringido(Constraint)

$$RSS^u = RSS^h + RSS^m = 12,117.3 + 5,713.6 = 17830.9$$

- Modelo no restringido (No constraint)

$$RSS^c = 17857.9$$

- Los grados de libertad para el modelo restringido esta expresado como:

$$DF^c = 26,737 - 7 = 26,730$$

- Los grados de libertad para el modelo no restringido esta esperesado como:

Pregunta 3

- La cantidades de RSS de cada ecuación estimada permite construir el test-F para implementar el test de Chow:

$$F = \frac{\frac{RSS^c - RSS^u}{DF^c - DF^u}}{\frac{RSS^u}{DF^u}}$$

- Donde, DF representa los grados de libertad. Por tanto, dado los resultados de las regresiones , tenemos
- Modelo restringido(Constraint)

$$RSS^u = RSS^h + RSS^m = 12,117.3 + 5,713.6 = 17830.9$$

- Modelo no restringido (No constraint)

$$RSS^c = 17857.9$$

- Los grados de libertad para el modelo restringido esta expresado como:

$$DF^c = 26,737 - 7 = 26,730$$

- Los grados de libertad para el modelo no restringido esta esperesado como:

$$DF^u = (19010 - 6) + (7,727 - 6) = 26,725$$

Pregunta 3

- La cantidades de RSS de cada ecuación estimada permite construir el test-F para implementar el test de Chow:

$$F = \frac{\frac{RSS^c - RSS^u}{DF^c - DF^u}}{\frac{RSS^u}{DF^u}}$$

- Donde, DF representa los grados de libertad. Por tanto, dado los resultados de las regresiones , tenemos
- Modelo restringido(Constraint)

$$RSS^u = RSS^h + RSS^m = 12,117.3 + 5,713.6 = 17830.9$$

- Modelo no restringido (No constraint)

$$RSS^c = 17857.9$$

- Los grados de libertad para el modelo restringido esta expresado como:

$$DF^c = 26,737 - 7 = 26,730$$

- Los grados de libertad para el modelo no restringido esta esperesado como:

$$DF^u = (19010 - 6) + (7,727 - 6) = 26,725$$

$$DF^c - DF^u = 26,730 - 26,725 = 5$$

Pregunta 3

EDÚCATE PERÚ CONSULTORES

- La cantidades de RSS de cada ecuación estimada permite construir el test-F para implementar el test de Chow:

$$F = \frac{\frac{RSS^c - RSS^u}{DF^c - DF^u}}{\frac{RSS^u}{DF^u}}$$

- Donde, DF representa los grados de libertad. Por tanto, dado los resultados de las regresiones , tenemos
- Modelo restringido(Constraint)

$$RSS^u = RSS^h + RSS^m = 12,117.3 + 5,713.6 = 17830.9$$

- Modelo no restringido (No constraint)

$$RSS^c = 17857.9$$

- Los grados de libertad para el modelo restringido esta expresado como:

$$DF^c = 26,737 - 7 = 26,730$$

- Los grados de libertad para el modelo no restringido esta esperesado como:

$$DF^u = (19010 - 6) + (7,727 - 6) = 26,725$$

$$DF^c - DF^u = 26,730 - 26,725 = 5$$

- A que se refiere este ultimo termino encontrado 5?

Pregunta 5

Pregunta 5

- ¿Cuáles son las 5 restricciones sobre este caso?

Pregunta 5

- ¿Cuáles son las 5 restricciones sobre este caso?
- $H_o : [\beta_1]^h = [\beta_1]^m$, diferencias de genero en educación
- $[\beta_2]^h = [\beta_2]^m$, diferencias de genero en experiencial (lineal)
- $[\beta_3]^h = [\beta_3]^m$, diferencias de genero en experiencia (cuadrático)
- $[\beta_4]^h = [\beta_4]^m$, diferencias de genero en pareja
- $[\beta_5]^h = [\beta_5]^m$, diferencias de genero en soltero
- $H_a : H_o$ no es verdad

Pregunta 5

EDÚCATE PERÚ CONSULTORES

- ¿Cuáles son las 5 restricciones sobre este caso?
- $H_o : [\beta_1]^h = [\beta_1]^m$, diferencias de genero en educación
- $[\beta_2]^h = [\beta_2]^m$, diferencias de genero en experiencial (lineal)
- $[\beta_3]^h = [\beta_3]^m$, diferencias de genero en experiencia (cuadrático)
- $[\beta_4]^h = [\beta_4]^m$, diferencias de genero en pareja
- $[\beta_5]^h = [\beta_5]^m$, diferencias de genero en soltero
- $H_a : H_o$ no es verdad
- Nota: Estamos probando las diferencias de genero en las interseciones de otras variables, ¿Porqué?

Pregunta 5

- El valor critico para el F-stadistico al nivel de significancia de 0.05 es de : $F_{6,\infty} = 2.099$, ¿Qué concluye?

$$F = \frac{\frac{17857.9 - 17830.9}{5}}{\frac{17830.9}{26,725}} = 8.0935343 \sim F(5, 26, 737)$$

- Sin embargo, aunque parece decisivo los resultados econtrados, el test de Chow es invalido en este presente caso.
- Nosotros conocemos que existe la presencia de heterocedasticidad, ello tambien se evidencia cuando se realiza las estimaciones por separado (hombre y mujer)

Pregunta 5

Pregunta 5

- Se requiere un enfoque alternativo para probar la hipótesis de conjunta, que sea menos engorrosa.

Pregunta 5

- Se requiere un enfoque alternativo para probar la hipótesis de conjunta, que sea menos engorrosa.
- Esto podría lograrse mediante la inclusión de **términos de interacción** entre la variable dummy femenina y cada variable en la ecuación logarítmica del salario.

Pregunta 5

- Se requiere un enfoque alternativo para probar la hipótesis de conjunta, que sea menos engorrosa.
- Esto podría lograrse mediante la inclusión de **términos de interacción** entre la variable dummy femenina y cada variable en la ecuación logarítmica del salario.
- Entonces podríamos probar la hipótesis de que **las estimaciones para estas interacciones son conjuntamente significativamente diferentes de cero** utilizando una **corrección robusta de heterocedasticidad para la matriz de varianza-covarianza**.

Pregunta 5

EDÚCATE PERÚ CONSULTORES

- Se requiere un enfoque alternativo para probar la hipótesis de conjunta, que sea menos engorrosa.
- Esto podría lograrse mediante la inclusión de **términos de interacción** entre la variable dummy femenina y cada variable en la ecuación logarítmica del salario.
- Entonces podríamos probar la hipótesis de que **las estimaciones para estas interacciones son conjuntamente significativamente diferentes de cero** utilizando una **corrección robusta de heterocedasticidad para la matriz de varianza-covarianza**.
- Si esta hipótesis conjunta no puede rechazarse, los puntos de datos se pueden agrupar por género.

Pregunta 5

- La intuición detras del test puede ser explicado a través del siguiente modelo:

$$lnwage_i = \alpha_0 + \alpha_1 educ_i + \alpha_2 edad_i + \alpha_3 edad_i^2$$

$$+\alpha_4 casado_i + \alpha_5 soltero_i + \pi mujer_i +$$

$$+\gamma_1 (mujer_i) x (educ_i) + \gamma_2 (mujer_i) x (edad_i) + \gamma_3 (mujer_i) x (edad_i^2)$$

$$+\gamma_4 (mujer_i) x (casado_i) + \gamma_5 (mujer_i) x (soltero_i)$$

- Hipotesis

$$H_o : \gamma_1 = \gamma_2 = \gamma_3 = \gamma_4 = \gamma_5 = 0$$

Pregunta 5

- Vamos a centrarnos sobre la variable educación y su parametro γ_1
- Si mujer (variable dummy) $= 0$, entonces (en promedio) la tasa marginal de retorno de la educación es α_1
- Esto representa los retornos a la educación de los hombres
- Si mujer (variable dummy) $= 1$, entonces (en promedio) la tasa marginal de retorno de la educación es $\alpha_1 + \gamma_1$
- Se define, (en promedio) que la tasa marginal de retorno para mujer esta dado por $\alpha_1 + \gamma_1 = \alpha^*$
- El efecto de la educación sobre el logaritmo de salario es α_1 si el individuo es hombre (por ejemplo, *female* $= 0$) y α^* si es mujer (por ejemplo, *female* $= 1$)
- Ahora la de los parametros de interacción es $\gamma_1 = \alpha^* - \alpha_1$
- Por tanto, la interaccion del parametro representa las **diferencias de genero** en el retorno a la educación.

Pregunta 5

Cuadro: Ecuación (1) - Log (Salarios)

	Pooled	Mujer	Hombre	Interactivo
reduca	0.0864***	0.0838***	0.0883***	0.0883***
	(0.0012)	(0.0023)	(0.0014)	(0.0015)
educa x mujer				-0.0045*
				(0.0026)
Observaciones	26737	7727	19010	26737
Adj. R^2	0.287	0.271	0.259	0.288
RSS	17857.9	5713.6	12117.3	17830.9

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