Programa de Especialización en Econometría Aplicada
Centro de Formación Continua -UNI
Practica
Clase 1

Edinson Tolentino
Docente
email: edinson.tolentino@gmail.com

Twitter: @edutoleraymondi

Universidad Nacional de Ingeneria

11 de julio de 2025

Contenido

Logro de la sesión

- Aplicar los conocimientos de pruebas de inferencias para parametros no lineales
- Aplicar el método delta
- Aplicación de pruebas de inferencia sobre base de datos de ingresos a nivel de personas

Introducción

Introducción

- Jacob Mincer publico su libro "Schooling, Experience and Earnings" (1974)

- Jacob Mincer publico su libro “Schooling, Experience and Earnings” (1974)
- Mincer modelo el logaritmo de ingresos como una funcion de años de educacion y años potenciales de experiencia en el mercado laboral
- Por lo tanto, la ecuación de Mincer:

$$\log w_x = X\beta + rSch + \beta_1 Exp - \beta_2 Exp^2$$

Introducción

- La consultora **Marilyn Loden** , quien acuñó el termino **techo de cristal** en 1978 , es también la autora del libro Liderazgo femenino o ¿cómo triunfar en los negocios sin ser uno de los chicos?.

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:
- · · · 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.

- 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.
- 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.

- 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

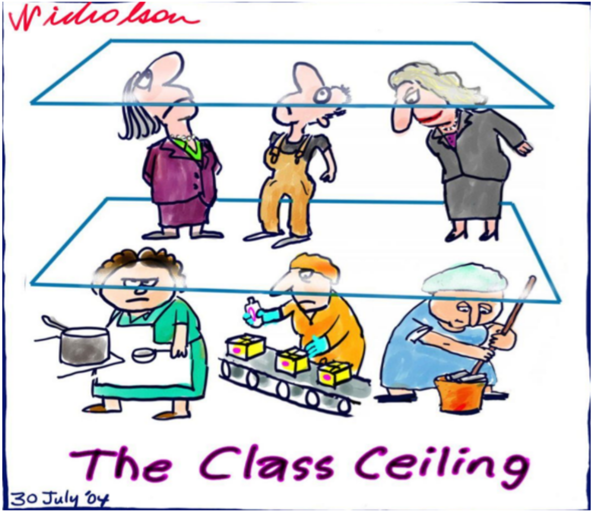
- La socióloga Catherine Berheide, quien acuñó la frase **piso pegajoso** (sticky floor) en 1992

- 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.

- 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:

- 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



- 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
ln *wage*	Logaritmo ingreso mensual (Soles)
reduca	Años de educación
rmujer	==1, mujer
redad	Años de edad
redadsq	Años de edad al cuadrado
rpareja	== 1, con pareja
rDpto	Según departamentos(dummies)

Cuadro: Estadisticas descriptivas

	Personas	Promedio	Mediana	Min.	Max.	Std
r6	24774	1,443.89	996.04	0.00	52,063.25	1,745
reduca	24774	8.70	11.00	0.00	18.00	5
rmujer	24774	0.28	0.00	0.00	1.00	0
redad	24774	46.80	47.00	18.00	69.00	12
redadsq	24774	2,330.71	2,209.00	324.00	4,761.00	1,109
rpareja	24774	0.66	1.00	0.00	1.00	0

Fuente: ENAHO - 2021.
Elaboracion: Autor

Introducción

Cuadro: Estadisticas descriptivas

	Personas	Promedio	Mediana	Min.	Max.	Std
r6	24774	1,443.89	996.04	0.00	52,063.25	1,745
reduca	24774	8.70	11.00	0.00	18.00	5
rmujer	24774	0.28	0.00	0.00	1.00	0
redad	24774	46.80	47.00	18.00	69.00	12
redadsq	24774	2,330.71	2,209.00	324.00	4,761.00	1,109
rpareja	24774	0.66	1.00	0.00	1.00	0

Fuente: ENAHO - 2021.
Elaboracion: Autor

- Alrededor de 28 % de los trabajadores son mujeres.
- En promedio, la experiencia de los trabajadores es de 47 años y sus años de educación son de 9 años.
- El 66 % de los trabajadores tiene una condición civil : con pareja.

Introducción

- Se realizará la estimación de la ecuación de salarios de los trabajadores peruanos utilizando la información 2020 (ENAHO)

- Se realizará la estimación de la ecuación de salarios de los trabajadores peruanos utilizando la información 2020 (ENAHO)
- Para realizar el análisis se propone tres ecuaciones:

$$\ln wage_i = \alpha_0 + \alpha_1 reduca_i + \mu_i \tag{1}$$

$$\begin{aligned}\ln wage_i &= \alpha_0 + \alpha_1 reduca_i + \pi mujer_i \\ &+ \alpha_2 redad_i + \alpha_3 redad_i^2 + \alpha_4 rpareja_i + \mu_i \end{aligned} \tag{2}$$

$$\begin{aligned}\ln wage_i &= \alpha_0 + \alpha_1 reduca_i + \pi mujer_i \\ &+ \alpha_2 redad_i + \alpha_3 redad_i^2 + \alpha_4 rpareja_i + \sum_{i=2}^{25} \lambda_i dpto_i + \mu_i \end{aligned} \tag{3}$$

Pregunta 1

1. Estime la regresión de la ecuación 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 1

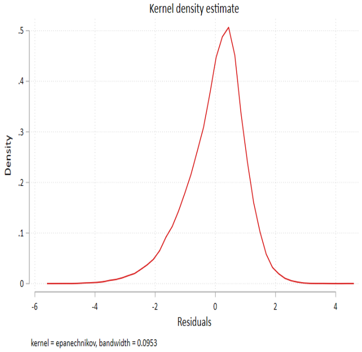
Cuadro: Modelo Lineal

	(1)	(2)
reduca	0.072***	0.071***
	(0.00)	(0.00)
rmujer		-0.560***
		(0.02)
redad		0.095***
		(0.00)
redadsq		-0.001***
		(0.00)
rpareja		-0.134***
		(0.02)
Constant	6.085***	4.450***
	(0.02)	(0.11)
Observations	24774	24774

Errores estandar en parentesis.
Fuente: EnAHO 2021.
Elaboracion: Autor
***, **, * denote statistical significance at the 1 %, 5 % and 10 % levels respectively for zero.

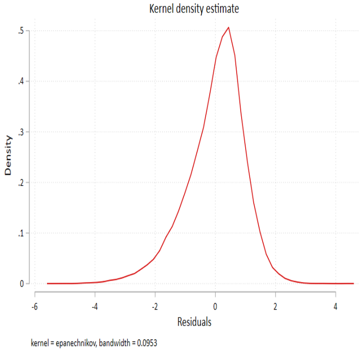
Pregunta 1

Figura 1:



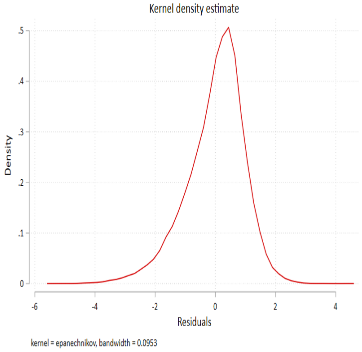
Pregunta 1

Figura 1:



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

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 1

Figura 1:

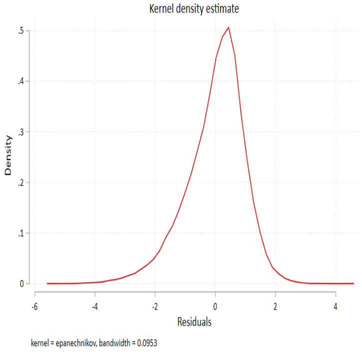
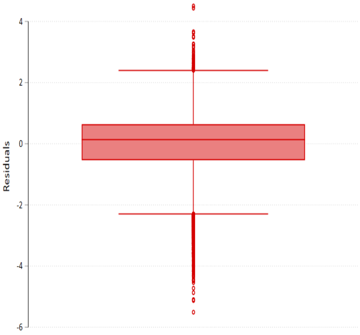


Figura 2:



Pregunta 1

- ¿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 mujer neduca exper expersq civil, iid

Pregunta 1

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

$$\varepsilon_i^2 = \delta_0 + \delta_1 reduca_i + \phi rmujer_i + \delta_2 redad_i + \delta_3 redad_i^2 + \delta_4 rpareja_i + \xi_i \qquad (4)$$

- Hipotesis

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

- 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: neduca mujer exper expersq civil

         chi2(5)      =   407.51
         Prob > chi2  =   0.0000
```

Pregunta 1

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

$$\varepsilon_i^2 = \delta_0 + \delta_1 reduca_i + \phi rmujer_i + \delta_2 redad_i + \delta_3 redad_i^2 + \delta_4 rpareja_i + \xi_i \qquad (4)$$

- Hipotesis

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

- 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: neduca mujer exper expersq civil

         chi2(5)      =   407.51
         Prob > chi2  =   0.0000
```

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

Pregunta 1

Cuadro: Modelo Lineal

	ε^2_i
reduca	0.016**
	(0.01)
rmujer	1.260***
	(0.10)
redad	-0.090***
	(0.02)
redadsq	0.001***
	(0.00)
rpareja	1.263***
	(0.09)
Constant	1.674***
	(0.53)
Observations	24774

Errores estandar en parentesis.
Fuente: EnAHO 2021.
Elaboracion: Autor
***, **, * denote statistical significance at the 1 %, 5 % and 10 % levels respectively for zero.

Pregunta 1

Cuadro: Modelo Lineal

	ε^2_i
reduca	0.016**
	(0.01)
rmujer	1.260***
	(0.10)
redad	-0.090***
	(0.02)
redadsq	0.001***
	(0.00)
rpareja	1.263***
	(0.09)
Constant	1.674***
	(0.53)
Observations	24774

Errores estandar en parentesis.
Fuente: EnAHO 2021.
Elaboracion: Autor
***, **, * denote statistical significance at the 1 %, 5 % and 10 % levels respectively for zero.

- Por tanto, $n \times R^2$ es igual a $24,774$

Cuadro: Modelo Lineal

	ε^2_i
reduca	0.016**
	(0.01)
rmujer	1.260***
	(0.10)
redad	-0.090***
	(0.02)
redadsq	0.001***
	(0.00)
rpareja	1.263***
	(0.09)
Constant	1.674***
	(0.53)
Observations	24774

Errores estandar en parentesis.
Fuente: EnAHO 2021.
Elaboracion: Autor
***, **, * denote statistical significance at the 1 %, 5 % and 10 % levels respectively for zero.

- Por tanto, $n \times R^2$ es igual a $24,774$
- Dado un valor de significancia de 0.05, la distribución de la tabla χ^2_{k-1} arroja el valor de 11.07 (χ^2_5)

Pregunta 1

Cuadro: Modelo Lineal

	ε^2_i
reduca	0.016**
	(0.01)
rmujer	1.260***
	(0.10)
redad	-0.090***
	(0.02)
redadsq	0.001***
	(0.00)
rpareja	1.263***
	(0.09)
Constant	1.674***
	(0.53)
Observations	24774

Errores estandar en parentesis.
Fuente: EnAHO 2021.
Elaboracion: Autor
***, **, * denote statistical significance at the 1 %, 5 % and 10 % levels respectively for zero.

- Por tanto, $n \times R^2$ es igual a $24,774$
- Dado un valor de significancia de 0.05, la distribución de la tabla χ^2_{k-1} arroja el valor de 11.07 (χ^2_5)
- Evidenciando que se rehaza la H_o(homocedasticidad)

Pregunta 1

Cuadro: Modelo Lineal

	ε^2_i
reduca	0.016**
	(0.01)
rmujer	1.260***
	(0.10)
redad	-0.090***
	(0.02)
redadsq	0.001***
	(0.00)
rpareja	1.263***
	(0.09)
Constant	1.674***
	(0.53)
Observations	24774

Errores estandar en parentesis.
Fuente: EnAHO 2021.
Elaboracion: Autor
***, **, * denote statistical significance at the 1 %, 5 % and 10 % levels respectively for zero.

- Por tanto, $n \times R^2$ es igual a $24,774$
- Dado un valor de significancia de 0.05, la distribución de la tabla χ^2_{k-1} arroja el valor de 11.07 (χ^2_5)
- Evidenciando que se rehaza la H_o(homocedasticidad)
- Concluyendo que existe presencia de **heterocedasticidad**

Pregunta 1

Cuadro: Modelo Lineal

	ε^2_i
reduca	0.016**
	(0.01)
rmujer	1.260***
	(0.10)
redad	-0.090***
	(0.02)
redadsq	0.001***
	(0.00)
rpareja	1.263***
	(0.09)
Constant	1.674***
	(0.53)
Observations	24774

Errores estandar en parentesis.
Fuente: EnAHO 2021.
Elaboracion: Autor
***, **, * denote statistical significance at the 1 %, 5 % and 10 % levels respectively for zero.

- Por tanto, $n \times R^2$ es igual a $24,774$
- Dado un valor de significancia de 0.05, la distribución de la tabla χ^2_{k-1} arroja el valor de 11.07 (χ^2_5)
- Evidenciando que 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 2

2 Use el comando robust para corregir la presencia de heterocedasticidad y re-estime el modelo (2). Luego responda las siguientes preguntas

1. Use la estructura del Test de Wald, realice usando un nivel de significancia de 0.05 para testear:

$$H_o : \alpha_4 = \pi = 0$$

2. Utilizando el Test de Wald en STATA, Python (test), ¿Qué es lo que usted concluye?

- Según el modelo :

$$\log r6_i = \alpha_0 + \alpha_1 reduca_i + \pi rmujer_i + \alpha_2 redad_i + \alpha_3 redadsq_i + \alpha_4 rpareja_i + \mu_i$$

Cuadro: Modelo Lineal robusto

	ln *wage*		ln *wage*		ln *wage*	
reduca	0.072***	(0.00)	0.072***	(0.00)	0.071***	(0.00)
rmujer			-0.476***	(0.02)	-0.560***	(0.03)
redad					0.095***	(0.01)
redadsq					-0.001***	(0.00)
rpareja					-0.134***	(0.02)
Constant	6.085***	(0.02)	6.218***	(0.01)	4.450***	(0.11)
Observations	24774		24774		24774	

Errores estandar en parentesis.
Fuente: EnAHO 2021.
Elaboracion: Autor
***, **, * denote statistical significance at the 1 %, 5 % and 10 % levels respectively for zero.

Pregunta 2

- La hipotesis nula es expresada como:

$$H_o : \alpha_4 = \pi = 0$$

$$H_a : H_o - no - verdad$$

- La matriz de la forma del test de Wald es especificado como:

$$Wald = \begin{bmatrix}(\hat{\alpha}_4 - \alpha_4) & (\hat{\pi} - \pi)\end{bmatrix} \hat{V}_{robust}^{-1} \begin{bmatrix}(\hat{\alpha}_4 - \alpha_4) & (\hat{\pi} - \pi)\end{bmatrix}'$$

- El test de Wald implementa las restricciones de la H_o

$$Wald = \begin{bmatrix}\hat{\alpha}_4 & \hat{\pi}\end{bmatrix} \hat{V}_{robust}^{-1} \begin{bmatrix}\hat{\alpha}_4 & \hat{\pi}\end{bmatrix}'$$

- El valor del test de Wald para una significancia conjunta de los dos coeficientes es **934.1019** dado una distribución de chi-cuadrado con **2** grados de libertad de **5.99** (ver tabla de distribucion de una χ^2)
  - STATA : 646.08286
  - Python : 934.1019

Pregunta 3

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 maximo alrededor de 50 años. Reporte todos los cálculos relevantes. ¿Qué es lo que usted concluye?

Pregunta 3

Cuadro: Modelos Lineales Robustos

	(1)	(2)	(3)
reduca	0.07***	0.07***	0.07***
	(0.00)	(0.00)	(0.00)
rmujer		-0.48***	-0.56***
		(0.02)	(0.03)
Observaciones	24774	24774	24774
R^2	0.0858	0.111	0.130
Controls			$\surd$

Fuente: ENAHO - 2021.
Elaboracion: Autor
***, **, * denote statistical significance at the 1 %, 5 % and 10 % levels respectively for zero.

Pregunta 3

- La derivada obtenida , esta determinada como:

$$\frac{\partial \log(wage)}{\partial redad} = \hat{\beta}_2 + 2\hat{\beta}_3 redad$$

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

$$redad_{estacionario} = \frac{\hat{\beta}_2}{-2\hat{\beta}_3} = \frac{0.095}{2x(-0.001)} = 43 = \hat{\triangle}$$

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

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

Pregunta 3

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

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

- Recordando:

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

- Donde, para el presente caso:

Pregunta 3

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

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

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

- 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 3

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

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

- 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 3

- El valor de las derivadas encontradas son calculadas:

Pregunta 3

- El valor de las derivadas encontradas son calculadas:

$$\frac{\partial \hat{\triangle}}{\partial \hat{\beta}_2} = -\frac{1}{2\hat{\beta}_3} = 452.16$$

Pregunta 3

- El valor de las derivadas encontradas son calculadas:

$$\frac{\partial \hat{\triangle}}{\partial \hat{\beta}_2} = -\frac{1}{2\hat{\beta}_3} = 452.16$$

$$\frac{\partial \hat{\triangle}}{\partial \hat{\beta}_3} = -\frac{\hat{\beta}_2}{2\hat{\beta}_3^2} = 39031.71$$

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

	const	number__reduca	number__rmujer	number__redad	number__redadsq	number__rpareja
const	0.013004	-2.401348e-05	-1.583186e-04	-5.641416e-04	5.928826e-06	-1.540953e-04
number__reduca	-0.000024	2.053916e-06	-2.325714e-06	1.616597e-07	-1.187166e-10	2.660956e-07
number__rmujer	-0.000158	-2.325714e-06	6.486965e-04	-9.613457e-06	1.026090e-07	3.902510e-04
number__redad	-0.000564	1.616597e-07	-9.613457e-06	2.622671e-05	-2.823949e-07	-1.115742e-05
number__redadsq	0.000006	-1.187166e-10	1.026090e-07	-2.823949e-07	3.099567e-09	1.258704e-07
number__rpareja	-0.000154	2.660956e-07	3.902510e-04	-1.115742e-05	1.258704e-07	4.884660e-04

Pregunta 3

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

	const	number_reduca	number_rmujer	number_redad	number_redadsq	number_rpareja
const	0.013004	-2.401348e-05	-1.583186e-04	-5.641416e-04	5.928826e-06	-1.540953e-04
number_reduca	-0.000024	2.053916e-06	-2.325714e-06	1.616597e-07	-1.187166e-10	2.660956e-07
number_rmujer	-0.000158	-2.325714e-06	6.486965e-04	-9.613457e-06	1.026090e-07	3.902510e-04
number_redad	-0.000564	1.616597e-07	-9.613457e-06	2.622671e-05	-2.823949e-07	-1.115742e-05
number_redadsq	0.000006	-1.187166e-10	1.026090e-07	-2.823949e-07	3.099567e-09	1.258704e-07
number_rpareja	-0.000154	2.660956e-07	3.902510e-04	-1.115742e-05	1.258704e-07	4.884660e-04

- Donde:

$$Var(\hat{\beta}_2) = 2.622671455450803 8e-05 = 0.000026226$$

$$Var(\hat{\beta}_3) = 3.099567058003456e-09 = 0.000000003099$$

$$Cov(\hat{\beta}_2, \hat{\beta}_3) = -2.8239486279027537e-07 = -0.0000002823$$

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

$$Var\left(\hat{\triangle}\right) = (452.16)^2 x0.000026226 + (39031.71)^2 x0.000000003099 + \cdots$$

$$= \cdots + 2x\left(452.16x39031.71\right)\left(-0.0000002823\right) \equiv 0.1164$$

- 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:

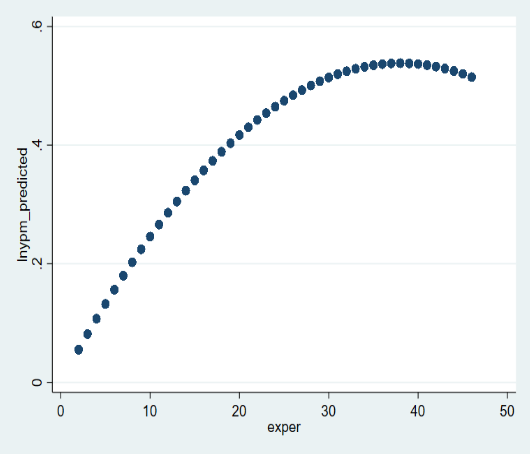
$$t = \frac{\hat{\triangle} - \triangle}{\sqrt{var(\triangle)}} = \frac{43 - 50}{\sqrt{0.1164}} = -20.51 \sim t_{\infty}$$

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

$$|-20.57| > |-1.96| \Rightarrow |t_{value}| > |t_{critico}|$$

- Qué concluye?

Figura : Perfil de salarios-edad para empleados de Perú



Pregunta 4

- 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 (2).

- **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 4

Cuadro: Modelos Lineales Robustos

	ln *wage*	ln *wage*	ln *wage*
reduca	0.07***	0.07***	0.07***
	(0.00)	(0.00)	(0.00)
Observaciones	24774	24774	24774
R^2	0.0858	0.111	0.130
Controls			√

Fuente: ENAHO - 2021.
Elaboracion: Autor
***, **, * denote statistical significance at the 1 %, 5 % and 10 % levels respectively for zero.

Pregunta 4

Pregunta 4

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

Pregunta 4

- 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 4

- 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 4

- 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

$$RSS^u = RSS^h + RSS^m = 25,565.242 + 13,372.516 = 38,937.758$$

Pregunta 4

- 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

$$RSS^u = RSS^h + RSS^m = 25,565.242 + 13,372.516 = 38,937.758$$

- Modelo no restringido

$$RSS^c = 39,186.371$$

Pregunta 4

- 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

$$RSS^u = RSS^h + RSS^m = 25,565.242 + 13,372.516 = 38,937.758$$

- Modelo no restringido

$$RSS^c = 39,186.371$$

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

$$DF^c = 24,774 - 6 = 24,768$$

Pregunta 4

- 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

$$RSS^u = RSS^h + RSS^m = 25,565.242 + 13,372.516 = 38,937.758$$

- Modelo no restringido

$$RSS^c = 39,186.371$$

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

$$DF^c = 24,774 - 6 = 24,768$$

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

Pregunta 4

- 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

$$RSS^u = RSS^h + RSS^m = 25,565.242 + 13,372.516 = 38,937.758$$

- Modelo no restringido
$$RSS^c = 39,186.371$$
- Los grados de libertad para el modelo restringido esta expresado como:

$$DF^c = 24,774 - 6 = 24,768$$

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

$$DF^u = (7,014 - 5) + (17,760 - 5) = 24,764$$

Pregunta 4

- 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

$$RSS^u = RSS^h + RSS^m = 25,565.242 + 13,372.516 = 38,937.758$$

- Modelo no restringido

$$RSS^c = 39,186.371$$

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

$$DF^c = 24,774 - 6 = 24,768$$

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

$$DF^u = (7,014 - 5) + (17,760 - 5) = 24,764$$

$$DF^c - DF^u = 24,768 - 24,764 = 4$$

Pregunta 4

- 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

$$RSS^u = RSS^h + RSS^m = 25,565.242 + 13,372.516 = 38,937.758$$

- Modelo no restringido

$$RSS^c = 39,186.371$$

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

$$DF^c = 24,774 - 6 = 24,768$$

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

$$DF^u = (7,014 - 5) + (17,760 - 5) = 24,764$$

$$DF^c - DF^u = 24,768 - 24,764 = 4$$

- A que se refiere este ultimo termino encontrado 4?

Pregunta 4

Pregunta 4

- ¿Cuáles son las 4 restricciones sobre este caso?

Pregunta 4

- ¿Cuáles son las 4 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 edad (lineal)
- $[\beta_3]^h = [\beta_3]^m$, diferencias de genero en edad (cuadrático)
- $[\beta_4]^h = [\beta_4]^m$, diferencias de genero en civil
- $H_a : H_o$ no es verdad

Pregunta 4

- ¿Cuáles son las 4 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 edad (lineal)
- $[\beta_3]^h = [\beta_3]^m$, diferencias de genero en edad (cuadrático)
- $[\beta_4]^h = [\beta_4]^m$, diferencias de genero en civil
- $H_a : H_o$ no es verdad
- Nota: Estamos probando las diferencias de genero en las interseciones de otras variables, ¿Porqué?

Pregunta 4

Pregunta 4

- El valor critico para el F-stadistico al nivel de significancia de 0.05 es de : $F_{5,\infty} = 2.099$, ¿Qué concluye?

Pregunta 4

- El valor critico para el F-stadistico al nivel de significancia de 0.05 es de : $F_{5,\infty} = 2.099$, ¿Qué concluye?

$$F = \frac{\frac{39{,}186.371 - 38{,}937.758}{4}}{\frac{38{,}937.758}{24{,}764}} = 39.52 \sim F(4, 38, 937.758)$$

- El valor critico para el F-stadistico al nivel de significancia de 0.05 es de : $F_{5,\infty} = 2.099$, ¿Qué concluye?

$$F = \frac{\frac{39{,}186.371 - 38{,}937.758}{4}}{\frac{38{,}937.758}{24{,}764}} = 39.52 \sim F(4, 38, 937.758)$$

- Sin embargo, aunque parece decisivo los resultados econtrados, el test de Chow es invalido en este presente caso.

Pregunta 4

- El valor critico para el F-stadistico al nivel de significancia de 0.05 es de : $F_{5,\infty} = 2.099$, ¿Qué concluye?

$$F = \frac{\frac{39{,}186.371 - 38{,}937.758}{4}}{\frac{38{,}937.758}{24{,}764}} = 39.52 \sim F(4, 38, 937.758)$$

- 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 4

Pregunta 4

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

Pregunta 4

- 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 4

- 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 4

- 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 4

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

$$\begin{aligned} \ln wage_i &= \alpha_0 + \alpha_1 educ_i + \pi mujer_i + \alpha_2 redad_i + \alpha_3 redad_i^2 + \alpha_4 rpareja_i \\ &= +\gamma_1 (mujer_i) x (educ_i) + \gamma_2 (mujer_i) x (exper_i) \\ &= +\gamma_3 (mujer_i) x (exper_i^2) + \gamma_4 (mujer_i) x (civil_i) \end{aligned}$$

- Hipotesis

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

Pregunta 4

- Dado la ecuacion con interaccion :

$$
\begin{aligned}
\ln wage_i \quad &= \alpha_0 + \alpha_1 educ_i + \pi \underbrace{mujer_i}_{=1} + \alpha_2 redad_i + \alpha_3 redad_i^2 + \alpha_4 rpareja_i \\
&= +\gamma_1 (\underbrace{mujer_i}_{==1}) x (educ_i) + \gamma_2 (mujer_i) x (exper_i) \\
&= +\gamma_3 (mujer_i) x (exper_i^2) + \gamma_4 (mujer_i) x (civil_i)
\end{aligned}
$$

- Si $rmujer_i == 1$, entonces la tasa marginal de retorno de la educación es α_1
- Si $rmujer_i == 1$, entonces la tasa marginal de retorno de la educación es $\alpha_1 + \gamma_1$
- Por tanto, la tasa marginal de retorno para mujer esta dado por $\alpha_1 + \gamma_1 = \alpha^*$
- Entonces, los parametros de interacción seran: $\gamma_1 = \alpha^* - \alpha_1$

Pregunta 4

Cuadro: Modelos Lineales Robustos

	Interaccion	(1)	mujer	hombre
reduca	0.07***	0.07***	0.08***	0.07***
	(0.00)	(0.00)	(0.00)	(0.00)
educa x mujer	0.01***			
	(0.00)			
Observaciones	24774	24774	7014	17760
R^2	0.136	0.130	0.132	0.104
Controls	$\surd$	$\surd$	$\surd$	$\surd$

Fuente: ENAHO - 2021.
Elaboracion: Autor
***, **, * denote statistical significance at the 1 %, 5 % and 10 % levels respectively for zero.