

Econometria Aplicada

Clase 1

Modelos Lineales

Edinson Tolentino

email: edinson.tolentino@gmail.com

Twitter: [@edutoleraymondi](https://twitter.com/edutoleraymondi)

Educate Peru

6 de abril de 2024

Introducción

Data y Variables

Pregunta 1

Pregunta 2

Pregunta 3

Pregunta 4

Introducción

Data y Variables

Pregunta 1

Pregunta 2

Pregunta 3

Pregunta 4

- ▶ 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

- ▶ 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
- ▶ Dado la siguiente ecuacion

$$\frac{\Delta Y_t}{h_t Y_t} = r$$

- ▶ 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
- ▶ Dado la siguiente ecuacion

$$\frac{\Delta Y_t}{h_t Y_t} = r$$

- ▶ Y_t representa los potenciales ingresos
- ▶ h_t representa la proporcion de periodo t utilizado para adquirir **capitla humano** (costo de oportunidad)

- ▶ 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
- ▶ Dado la siguiente ecuacion

$$\frac{\Delta Y_t}{h_t Y_t} = r$$

- ▶ Y_t representa los potenciales ingresos
- ▶ h_t representa la proporcion de periodo t utilizado para adquirir **capitla humano** (costo de oportunidad)
- ▶ Si el individuo se dedica full-time en educación

$$h_t = 1 \Rightarrow Y_s = Y_0 e^{rs}$$

- ▶ 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
- ▶ Dado la siguiente ecuacion

$$\frac{\Delta Y_t}{h_t Y_t} = r$$

- ▶ Y_t representa los potenciales ingresos
- ▶ h_t representa la proporcion de periodo t utilizado para adquirir **capitla humano** (costo de oportunidad)
- ▶ Si el individuo se dedica full-time en educación

$$h_t = 1 \Rightarrow Y_s = Y_0 e^{rs}$$

- ▶ Para años post-school, Mincer asume que h_t se reduce linealmente con la experiencia

$$h_t = h_0 - \left(\frac{h_0}{T} \right) t$$

- Entonces, para x años de post-school de experiencia laboral

$$Y_x = Y_s \exp \left(r \int_0^x h_t \partial t \right) \quad (1)$$

Donde :

$$\int_0^x h_t \partial t = h_0 x - \frac{1}{2} \frac{h_0}{T} x^2$$

Adicionalmente :

$$Y_s = Y_0 e^{rs} \equiv X\beta e^{rs}$$

Retomando(1)

$$Y_x = Y_0 e^{rs} \exp \left[r \left(h_0 x - \frac{1}{2} \frac{h_0}{T} x^2 \right) \right]$$

Entonces,

$$\log Y_x = \log Y_0 + rs + rh_0 x - \left(\frac{rh_0}{2T} \right) x^2$$

Dado :

$$w_x = (1 - h_x) Y_x$$

$$\log w_x = X\beta + rs + rh_0 x - r \left(\frac{h_0}{2T} \right) x^2 + \log(1 - h_x) \quad (2)$$

- Entonces, para x años de post-school de experiencia laboral

$$Y_x = Y_s \exp \left(r \int_0^x h_t \partial t \right) \quad (1)$$

Donde :

$$\int_0^x h_t \partial t = h_0 x - \frac{1}{2} \frac{h_0}{T} x^2$$

Adicionalmente :

$$Y_s = Y_0 e^{rs} \equiv X\beta e^{rs}$$

Retomando(1)

$$Y_x = Y_0 e^{rs} \exp \left[r \left(h_0 x - \frac{1}{2} \frac{h_0}{T} x^2 \right) \right]$$

Entonces,

$$\log Y_x = \log Y_0 + rs + rh_0 x - \left(\frac{rh_0}{2T} \right) x^2$$

Dado :

$$w_x = (1 - h_x) Y_x$$

$$\log w_x = X\beta + rs + rh_0 x - r \left(\frac{h_0}{2T} \right) x^2 + \log(1 - h_x) \quad (2)$$

- Por lo tanto, la ecuacion (2) equivala a la ecuación de Mincer:

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





- ▶ 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?.

- ▶ 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:

- ▶ 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.

- ▶ 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:
- ▶ ... 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.

- ▶ 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.
- ▶ 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.





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



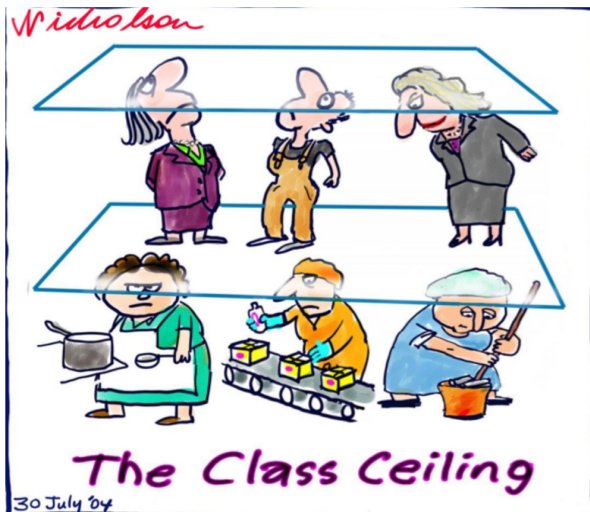
- ▶ La socióloga Catherine Berheide, quien acuñó la frase **piso pegajoso** (sticky floor) en 1992
- ▶ El **piso pegajoso** se refiere a los casos en los que las mujeres ocupan puestos de baja categoría, bajos salarios y de baja movilidad en el mercado laboral.



- ▶ 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:
- ▶ ... la mayoría de las mujeres deberían tener la suerte de tener el techo de cristal como su problema ... muchas están atascadas en el piso pegajoso.



Introducción

Data y Variables

Pregunta 1

Pregunta 2

Pregunta 3

Pregunta 4

- La información que se utilizará es proveniente de la base de datos de la Encuesta Nacional de Hogares (ENAHOG). 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: Estadísticas 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

Cuadro: Estadísticas 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.

- ▶ Se realizará la estimación de la ecuación de salarios de los trabajadores peruanos utilizando la información 2020 (ENAH0)

- ▶ Se realizará la estimación de la ecuación de salarios de los trabajadores peruanos utilizando la información 2020 (ENAH0)
- ▶ Para realizar el análisis se propone tres ecuaciones:

$$\ln wage_i = \alpha_0 + \alpha_1 reduca_i + \mu_i \quad (1)$$

$$\ln wage_i = \alpha_0 + \alpha_1 reduca_i + \pi mujer_i + \alpha_2 redad_i + \alpha_3 redad_i^2 + \alpha_4 rpareja_i + \mu_i \quad (2)$$

$$\ln wage_i = \alpha_0 + \alpha_1 reduca_i + \pi mujer_i + \alpha_2 redad_i + \alpha_3 redad_i^2 + \alpha_4 rpareja_i + \sum_{i=2}^{25} \lambda_i dpto_i + \mu_i \quad (3)$$

Introducción

Data y Variables

Pregunta 1

Pregunta 2

Pregunta 3

Pregunta 4

- 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?

Cuadro: Modelo Lineal

	(1)	(2)
reduca	0.072*** (0.00)	0.071*** (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*** (0.02)	4.450*** (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.

Figura 1:

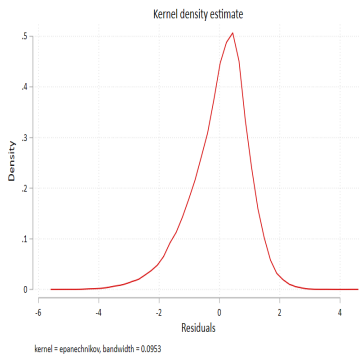
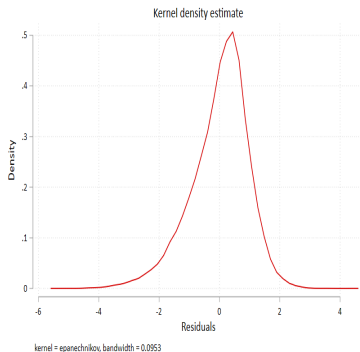
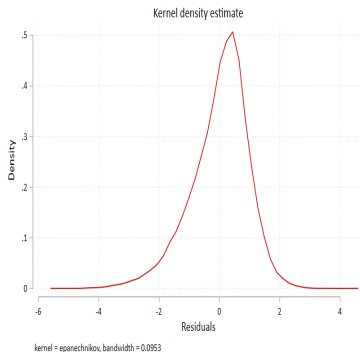


Figura 1:



- 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.

Figura 1:

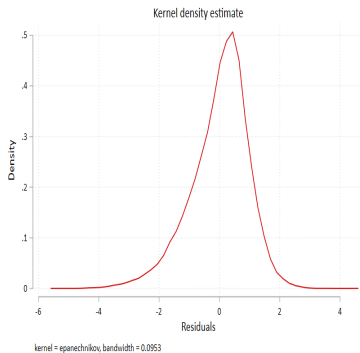
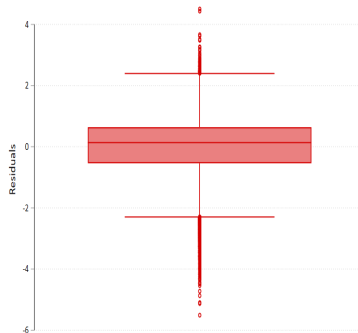


Figura 2:



- ▶ ¿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 idéntica 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

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

$$\varepsilon_i^2 = \delta_0 + \delta_1 \text{reduca}_i + \phi \text{rmujer}_i + \delta_2 \text{redad}_i + \delta_3 \text{redad}_i^2 + \delta_4 \text{rpareja}_i + \xi_i \quad (4)$$

- Hipotesis

$$H_o: \delta_0 = \delta_1 = \phi = \delta_2 = \delta_3 = \delta_4 = 0$$

vs $H_a: H_o$ no es verdad

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

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

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity

Ho: Constant variance

Variables: neduca mujer exper expersq civil

chi2(5) = 407.51

Prob > chi2 = 0.0000

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

$$\varepsilon_i^2 = \delta_0 + \delta_1 \text{reduca}_i + \phi \text{rmujer}_i + \delta_2 \text{redad}_i + \delta_3 \text{redad}_i^2 + \delta_4 \text{rpareja}_i + \xi_i \quad (4)$$

- Hipotesis

$$H_o: \delta_0 = \delta_1 = \phi = \delta_2 = \delta_3 = \delta_4 = 0$$

vs $H_a: H_o$ 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)

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 paréntesis.

Fuente: EnAHO 2021.

Elaboracion: Autor

***, **, * denote statistical significance at the 1 %, 5 % and 10 % levels respectively for zero.

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 paréntesis.

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 paréntesis.

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)

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 paréntesis.

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_0 (homocedasticidad)

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 paréntesis.

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_0 (homocedasticidad)
- ▶ Concluyendo que existe presencia de **heterocedasticidad**

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 paréntesis.

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_0 (homocedasticidad)
- ▶ Concluyendo que existe presencia de **heterocedasticidad**
- ▶ Se necesita corregir la matriz de varianza-covarianza usando la opción robust antes del análisis

Introducción

Data y Variables

Pregunta 1

Pregunta 2

Pregunta 3

Pregunta 4

- 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 :

$$\log r6_i = \alpha_0 + \alpha_1 \text{reduca}_i + \pi \text{rmujer}_i + \alpha_2 \text{redad}_i + \alpha_3 \text{redadsq}_i + \alpha_4 \text{rpareja}_i + \mu_i$$

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.

- ▶ La hipótesis 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:

$$Wald = [(\hat{\alpha}_4 - \alpha_4) \quad (\hat{\pi} - \pi)] \hat{V}_{robust}^{-1} [(\hat{\alpha}_4 - \alpha_4) \quad (\hat{\pi} - \pi)]'$$

- ▶ El test de Wald implementa las restricciones de la H_o

$$Wald = [\hat{\alpha}_4 \quad \hat{\pi}] \hat{V}_{robust}^{-1} [\hat{\alpha}_4 \quad \hat{\pi}]'$$

- ▶ 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 distribución de una χ^2)
 - ▶ STATA : 646.08286
 - ▶ Python : 934.1019

Introducción

Data y Variables

Pregunta 1

Pregunta 2

Pregunta 3

Pregunta 4

- 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?

Cuadro: Modelos Lineales Robustos

	(1)	(2)	(3)
reduca	0.07*** (0.00)	0.07*** (0.00)	0.07*** (0.00)
rmujer		-0.48*** (0.02)	-0.56*** (0.03)
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.

- ▶ 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 calculado al igualar en cero la derivada obtenida y despejar el valor de la variable *exper*
- ▶ Para la presente aplicación, se tiene:

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

- ▶ Por tanto, deseamos testear la proposición planetada a través de la hipotesis

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

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

$$\begin{aligned} Var(\hat{\Delta}) &= \left(\frac{\partial \hat{\Delta}}{\partial \hat{\gamma}_2} \right)^2 Var(\hat{\gamma}_2) + \left(\frac{\partial \hat{\Delta}}{\partial \hat{\gamma}_3} \right)^2 Var(\hat{\gamma}_3) \\ &+ 2 \left(\frac{\partial \hat{\Delta}}{\partial \hat{\gamma}_2} \right) \left(\frac{\partial \hat{\Delta}}{\partial \hat{\gamma}_3} \right) Cov(\hat{\gamma}_2, \hat{\gamma}_3) \end{aligned}$$

- Recordando:

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

- Donde, para el presente caso:

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

$$\begin{aligned} Var(\hat{\Delta}) &= \left(\frac{\partial \hat{\Delta}}{\partial \hat{\gamma}_2} \right)^2 Var(\hat{\gamma}_2) + \left(\frac{\partial \hat{\Delta}}{\partial \hat{\gamma}_3} \right)^2 Var(\hat{\gamma}_3) \\ &+ 2 \left(\frac{\partial \hat{\Delta}}{\partial \hat{\gamma}_2} \right) \left(\frac{\partial \hat{\Delta}}{\partial \hat{\gamma}_3} \right) Cov(\hat{\gamma}_2, \hat{\gamma}_3) \end{aligned}$$

- Recordando:

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

- Donde, para el presente caso:

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

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

$$\begin{aligned} Var(\hat{\Delta}) &= \left(\frac{\partial \hat{\Delta}}{\partial \hat{\gamma}_2} \right)^2 Var(\hat{\gamma}_2) + \left(\frac{\partial \hat{\Delta}}{\partial \hat{\gamma}_3} \right)^2 Var(\hat{\gamma}_3) \\ &+ 2 \left(\frac{\partial \hat{\Delta}}{\partial \hat{\gamma}_2} \right) \left(\frac{\partial \hat{\Delta}}{\partial \hat{\gamma}_3} \right) Cov(\hat{\gamma}_2, \hat{\gamma}_3) \end{aligned}$$

- Recordando:

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

- Donde, para el presente caso:

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

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

Pregunta 3

- ▶ El valor de las derivadas encontradas son calculadas:

Pregunta 3

- El valor de las derivadas encontradas son calculadas:

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

- El valor de las derivadas encontradas son calculadas:

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

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

- 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

- 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:

$$Var(\hat{\beta}_2) = 2.6226714554508038e - 05 = 0.000026226$$

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

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

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

$$\begin{aligned} \text{Var}(\hat{\Delta}) &= (452.16)^2 \times 0.000026226 + (39031.71)^2 \times 0.000000003099 + \dots \\ &= \dots + 2 \times (452.16 \times 39031.71) \times (-0.0000002823) \equiv 0.1164 \end{aligned}$$

- ▶ Nosotros deseamos testear la hipótesis:

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

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

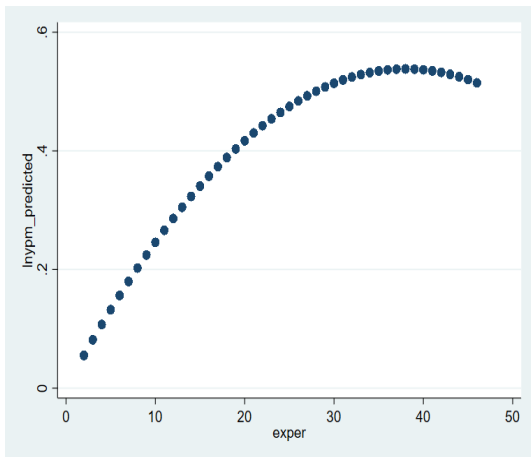
$$t = \frac{\hat{\Delta} - \Delta}{\sqrt{\text{var}(\hat{\Delta})}} = \frac{43 - 50}{\sqrt{0.1164}} = -20.51 \sim t_{\infty}$$

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

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

- ▶ Qué concluye?

Figura : Perfil de salarios-edad para empleados de Perú



Introducción

Data y Variables

Pregunta 1

Pregunta 2

Pregunta 3

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 estadísticas 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 específicas de género, puede entonces ser resumido para el retorno de no restringidos, la cual será definida como RSS^u
- ▶ Este es entonces comparado dado la suma de residuos al cuadrado de la forma del modelo restringido (también conocido como pooled), la cual se define como RSS^c

Cuadro: Modelos Lineales Robustos

	<i>ln wage</i>	<i>ln wage</i>	<i>ln wage</i>
reduca	0.07*** (0.00)	0.07*** (0.00)	0.07*** (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 cantidad 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 cantidad 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 está expresado como:

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

- ▶ Los grados de libertad para el modelo no restringido está expresado 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

- ▶ ¿Cuáles son las 4 restricciones sobre este caso?

- ▶ ¿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

- ▶ ¿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 intersecciones de otras variables, ¿Porqué?**

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?

- 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.

- 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

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

- ▶ 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.

- ▶ 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**.

- ▶ 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.

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

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

- Hipotesis

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

```
-----  
      name: <unnamed>  
      log: D:_Econometricas1/Tablas/resultados_2.log  
log type: text  
opened on: 12 Aug 2021, 22:59:06  
  
. test i1 i2 i3 i4  
  
( 1) i1 = 0  
( 2) i2 = 0  
( 3) i3 = 0  
( 4) i4 = 0  
  
      F( 4, 37486) = 33.71  
      Prob > F = 0.0000  
  
. log close  
      name: <unnamed>  
      log: D:_Econometricas1/Tablas/resultados_2.log  
log type: text  
closed on: 12 Aug 2021, 22:59:06  
-----
```

- Dado la ecuacion con interaccion :

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

- 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: $\gamma_1 = \alpha^* - \alpha_1$

Cuadro: Modelos Lineales Robustos

	Interaccion	(1)	mujer	hombre
reduca	0.07*** (0.00)	0.07*** (0.00)	0.08*** (0.00)	0.07*** (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	✓	✓	✓	✓

Fuente: ENAHO - 2021.

Elaboracion: Autor

***, **, * denote statistical significance at the 1 %, 5 % and 10 % levels respectively for zero.