

Programa de Especialización en Econometría Aplicada
Centro de Formación Continua -UNI
Modelos Variable Dependiente Limitada
Clase 2
Práctica

Edinson Tolentino

MSc Economics

email: edinson.tolentino@gmail.com

Twitter: @edutoleraymondi

Universidad Nacional de Ingeniería

30 de noviembre de 2024



1 Introducción

2 Data

3 Pregunta 1

4 Pregunta 2

5 Pregunta 3

6 Pregunta 4

7 Pregunta 5





- Un problema a la hora de estimar la ecuación de salarios en las mujeres (por ejemplo, logaritmo de salarios) evidencia que la muestra en empleo puede no ser una extracción aleatoria de la población.



- Un problema a la hora de estimar la ecuación de salarios en las mujeres (por ejemplo, logaritmo de salarios) evidencia que la muestra en empleo puede no ser una extracción aleatoria de la población.
- En otras palabras, las mujeres observadas en el empleo pueden ser un grupo selectivo en términos de sus **no observables** (por ejemplo, capacidad, motivación).



- Un problema a la hora de estimar la ecuación de salarios en las mujeres (por ejemplo, logaritmo de salarios) evidencia que la muestra en empleo puede no ser una extracción aleatoria de la población.
- En otras palabras, las mujeres observadas en el empleo pueden ser un grupo selectivo en términos de sus **no observables** (por ejemplo, capacidad, motivación).
- La fuente del problema es que los **no observables** que determinan la participación en el empleo y los que determinan los logaritmos de los salarios están potencialmente correlacionados.



- Un problema a la hora de estimar la ecuación de salarios en las mujeres (por ejemplo, logaritmo de salarios) evidencia que la muestra en empleo puede no ser una extracción aleatoria de la población.
- En otras palabras, las mujeres observadas en el empleo pueden ser un grupo selectivo en términos de sus **no observables** (por ejemplo, capacidad, motivación).
- La fuente del problema es que los **no observables** que determinan la participación en el empleo y los que determinan los logaritmos de los salarios están potencialmente correlacionados.
- En tales circunstancias, la estimación convencional de ecuaciones de logaritmos salariales por MCO puede producir estimaciones de coeficientes sesgadas dada la presencia de **sesgo de selección**.





- **Heckman** (1976) propone una metodología el cual eventualmente proviene de un procedimiento estandar para la corrección de **sesgo selección**



- **Heckman** (1976) propone una metodología el cual eventualmente proviene de un procedimiento estandar para la corrección de **sesgo selección**
- Esto llega a ser conocido como el **procedimiento de corrección Heckman**



- **Heckman** (1976) propone una metodología el cual eventualmente proviene de un procedimiento estandar para la corrección de **sesgo selección**
- Esto llega hacer conocido como el **procedimiento de corrección Heckman**
- El método podría implementarse utilizando un enfoque en dos pasos o un enfoque de ecuaciones simultáneas utilizando la Máxima Verosimilitud de información completa (FIML).



- Dada la información de la Encuesta Nacional de Hogares **ENAHOG**, la cual contiene información de los jefes de hogar mujeres en Perú durante el 2021.

Cuadro: Descripción de variables

Variables	Descripción
<i>rlfp_i</i>	== 1 , mujer que pertenece a la fuerza laboral
<i>rpereja_i</i>	== 1 , mujer con pareja
<i>redad_i</i>	edad en años
<i>redadsq_i</i>	edad en años cuadrado
<i>reduca_i</i>	años de educación
<i>lnr6prin</i>	Logaritmo ingreso mensual (Soles)
<i>rn6_i</i>	Numero de hijos de 0 a 6 años
<i>rn12_i</i>	Numero de hijos de 7 a 12 años
<i>rexper_i</i>	experiencia laboral (años)
<i>rexpersq_i</i>	experiencia laboral cuadrado (años)



Cuadro: Estadísticas descriptivas

	Personas	Promedio	Mediana	Min.	Max.	Std
log-salarios (Ocup. princ)	7778	6.37	6.42	4.22	10.17	1
Ingreso laboral mensual (ocup. princ.)	7778	993.42	615.73	68.33	26,120.17	1,349
años educacion	10932	8.14	9.00	0.00	18.00	5
Edad	10932	52.93	52.00	18.00	98.00	16
Edad cuadrado	10932	3,069.14	2,704.00	324.00	9,604.00	1,813
Persona con pareja	10932	0.87	1.00	0.00	1.00	0
Numero de hijos	10932	1.29	1.00	0.00	9.00	1
Numero de hijos 6 años	10932	0.28	0.00	0.00	5.00	0
Numero de hijos 6-12 años	10932	0.32	0.00	0.00	5.00	0
Experiencia laboral	7778	33.76	33.00	1.00	84.00	15
Experiencia laboral cuadrado	7778	1,376.44	1,089.00	1.00	7,056.00	1,148
=1 pertenece a la PEA	10932	0.71	1.00	0.00	1.00	0

Fuente: ENAHO - 2021.

Elaboracion: Autor



Cuadro: Estadísticas descriptivas

	Personas	Promedio	Mediana	Min.	Max.	Std
log-salarios (Ocup. princ.)	7778	6.37	6.42	4.22	10.17	1
Ingreso laboral mensual (ocup. princ.)	7778	993.42	615.73	68.33	26,120.17	1,349
años educacion	10932	8.14	9.00	0.00	18.00	5
Edad	10932	52.93	52.00	18.00	98.00	16
Edad cuadrado	10932	3,069.14	2,704.00	324.00	9,604.00	1,813
Persona con pareja	10932	0.87	1.00	0.00	1.00	0
Numero de hijos	10932	1.29	1.00	0.00	9.00	1
Numero de hijos 6 años	10932	0.28	0.00	0.00	5.00	0
Numero de hijos 6-12 años	10932	0.32	0.00	0.00	5.00	0
Experiencia laboral	7778	33.76	33.00	1.00	84.00	15
Experiencia laboral cuadrado	7778	1,376.44	1,089.00	1.00	7,056.00	1,148
=1 pertenece a la PEA	10932	0.71	1.00	0.00	1.00	0

Fuente: ENAHO - 2021.

Elaboracion: Autor

- Se posee un total de 10,932 observaciones a lo largo de las variables (demográficas y de capital humano) para las personas.



Cuadro: Estadísticas descriptivas

	Personas	Promedio	Mediana	Min.	Max.	Std
log-salarios (Ocup. princ)	7778	6.37	6.42	4.22	10.17	1
Ingreso laboral mensual (ocup. princ.)	7778	993.42	615.73	68.33	26,120.17	1,349
años educacion	10932	8.14	9.00	0.00	18.00	5
Edad	10932	52.93	52.00	18.00	98.00	16
Edad cuadrado	10932	3,069.14	2,704.00	324.00	9,604.00	1,813
Persona con pareja	10932	0.87	1.00	0.00	1.00	0
Numero de hijos	10932	1.29	1.00	0.00	9.00	1
Numero de hijos 6 años	10932	0.28	0.00	0.00	5.00	0
Numero de hijos 6-12 años	10932	0.32	0.00	0.00	5.00	0
Experiencia laboral	7778	33.76	33.00	1.00	84.00	15
Experiencia laboral cuadrado	7778	1,376.44	1,089.00	1.00	7,056.00	1,148
=1 pertenece a la PEA	10932	0.71	1.00	0.00	1.00	0

Fuente: ENAHO - 2021.

Elaboracion: Autor

- Se posee un total de 10,932 observaciones a lo largo de las variables (demográficas y de capital humano) para las personas.
- Cuatro de las variables solo reportan observaciones para 7,778 mujeres quienes perciben un ingreso laboral

Pregunta 1





- La ecuación de logaritmo de salarios puede ser estimado bajo un modelo de regresión **Censurado** o **Truncado**

Pregunta 1





- **Modelo de regresión censurada:**

- ▶ Se posee observaciones para todos los valores de X pero no sobre la variable producto Y , cuando el evento no ocurre (por ejemplo, cuando $y = 0$)



- **Modelo de regresión censurada:**

- ▶ Se poseen observaciones para todos los valores de X pero no sobre la variable producto Y , cuando el evento no ocurre (por ejemplo, cuando $y = 0$)

- **Modelo de regresión truncada:**

- ▶ No se poseen observaciones para todos los valores de X o los valores de la variable producto Y , cuando el evento no ocurre (por ejemplo, cuando $y = 0$)



- **Modelo de regresión censurada:**

- ▶ Se posee observaciones para todos los valores de X pero no sobre la variable producto Y , cuando el evento no ocurre (por ejemplo, cuando $y = 0$)

- **Modelo de regresión truncada:**

- ▶ No se posee observaciones para todos los valores de X o los valores de la variable producto Y , cuando el evento no ocurre (por ejemplo, cuando $y = 0$)

- **Respuesta:** modelo de regresión truncado

Pregunta 2





- Estime el modelo probit para la fuerza laboral femenina (empleo de las mujeres)

$$\text{prob} [lfp_i = 1] = \Phi(\alpha_0 + \alpha_1 educ_i + \alpha_2 edad_i + \alpha_3 edad_i^2 \\ + \alpha_4 pareja_i + \alpha_5 rnh6_i + \alpha_6 rnh12_i)$$

- 1 A que edad las mujeres maximizan su participación en el mercado laboral
- 2 Use el nivel de significancia de 0.05 para determinar si la información sobre la población de la fuerza laboral alcanza su máximo a los 34 años
- 3 Interprete precisamente los efectos marginales para las variables **rnh6** y **rnh12**

Pregunta 2.1



```
name: <unnamed>
log: C:396L2/Tablas/resultados_2.log
log type: text
opened on: 22 Jun 2024, 04:00:47
```

```
.      eststo: probit rflp $Zs
```

```
Iteration 0: Log likelihood = -6568.0864
Iteration 1: Log likelihood = -5793.1131
Iteration 2: Log likelihood = -5792.0513
Iteration 3: Log likelihood = -5792.0513
```

Probit regression

Number of obs = 10,932
LR chi2(6) = 1552.07
Prob > chi2 = 0.0000
Pseudo R2 = 0.1182

Log likelihood = -5792.0513

rflp	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
reduca	-.0313924	.0028709	-10.93	0.000	-.0370192	-.0257656
rpajeja	-.165457	.0435874	-3.80	0.000	-.2508866	-.0800273
redad	.0527756	.0050695	10.41	0.000	.0428395	.0627116
redadsq	-.0007804	.0000454	-17.20	0.000	-.0008693	-.0006914
rn6	-.2420865	.0248881	-9.73	0.000	-.2908661	-.1933068
rn12	-.0610507	.0248356	-2.46	0.014	-.1097275	-.0123739
_cons	.6937773	.1438517	4.82	0.000	.4118332	.9757214

(est1 stored)

```
.      log close
name: <unnamed>
log: C:396L2/Tablas/resultados_2.log
log type: text
closed on: 22 Jun 2024, 04:00:50
```


Pregunta 2.1





Cuadro: Modelo Probit (1)

==1 pertenece a la PEA		
años educacion	-0.03***	(0.00)
Persona con pareja	-0.17***	(0.04)
Edad	0.05***	(0.01)
Edad cuadrado	-0.00***	(0.00)
Numero de hijos 6 años	-0.24***	(0.02)
Numero de hijos 6-12 años	-0.06**	(0.02)
Constant	0.69***	(0.14)
Observaciones	10932	
Pseudo. R ²	0.118	
Log-L	-5792.1	
Grados de Libertad (k)	7	

Fuente: ENAHO - 2021.

Elaboracion: Autor

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



Cuadro: Modelo Probit (1)

==1 pertenece a la PEA		
años educacion	-0.03***	(0.00)
Persona con pareja	-0.17***	(0.04)
Edad	0.05***	(0.01)
Edad cuadrado	-0.00***	(0.00)
Numero de hijos 6 años	-0.24***	(0.02)
Numero de hijos 6-12 años	-0.06**	(0.02)
Constant	0.69***	(0.14)
Observaciones	10932	
Pseudo. R ²	0.118	
Log-L	-5792.1	
Grados de Libertad (k)	7	

Fuente: ENAHO - 2021.

Elaboracion: Autor

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

- La edad a la cual las mujeres maximizan su edad en la participación del salarios sera:

$$z_i = \dots + \hat{\alpha}_2 edad_i + \hat{\alpha}_3 edad_i^2 + \dots$$

$$\frac{\partial z}{\partial edad} = -\hat{\alpha}_2 + 2\hat{\alpha}_3 edad_i$$

$$edad_{ss} = -\frac{\hat{\alpha}_2}{2\hat{\alpha}_3}$$

- Reemplazando los datos de la estimación realizada

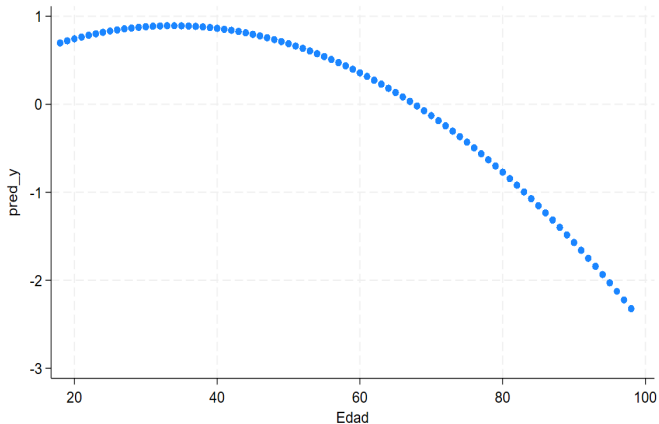
$$edad_{ss} = -\frac{0.05277}{2(-0.0007804)}$$

$$edad_{ss} = \hat{\Delta} = 33.81$$

Pregunta 2.1



- Entonces, la participación laboral en Perú alcanza su máximo a la edad de 34 años, ¿Será plausible dicho resultado?, ¿que crees?



Pregunta 2.2





- Use el nivel de significancia de 0.05 para determinar si la información sobre la población de la fuerza laboral alcanza su máximo a los 34 años
- Entonces

$$H_0 : \Delta = 34$$

$$H_a : \Delta \neq 34$$

Pregunta 2.2



```
-----
name: <unnamed>
log: C:396L2/Tablas/resultados_3.log
log type: text
opened on: 22 Jun 2024, 04:00:50
```

```
. nlcom - _b[redad]/(2*_b[redadsq]) -34
      _nl_1: - _b[redad]/(2*_b[redadsq]) -34
```

```
-----
      rflp | Coefficient Std. err.      z    P>|z|    [95% conf. interval]
-----+-----
      _nl_1 |  -.1853663    1.385356   -0.13   0.894   -2.900614    2.529881
-----
```

```
. gen pred_y=_b[redad]*redad + _b[redadsq]*redadsq
. scatter pred_y redad
```

```
. graph export "$Imagen/t1.png", replace
file C:396L2/Imagen/t1.png saved as PNG format
```

```
. log close
name: <unnamed>
log: C:396L2/Tablas/resultados_3.log
log type: text
closed on: 22 Jun 2024, 04:00:51
-----
```

Pregunta 2.3





Cuadro: Efectos Marginales probit

	Efectos Marginales	
años educacion	-0.009***	(0.00)
Persona con pareja	-0.049***	(0.01)
Edad	0.016***	(0.00)
Edad cuadrado	-0.000***	(0.00)
Numero de hijos 6 años	-0.072***	(0.01)
Numero de hijos 6-12 años	-0.018**	(0.01)
Observations	10932	

Errores estandar en parentesis.

Fuente: INEI -2021.

Elaboracion: Autor

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



Cuadro: Efectos Marginales probit

	Efectos Marginales	
años educacion	-0.009***	(0.00)
Persona con pareja	-0.049***	(0.01)
Edad	0.016***	(0.00)
Edad cuadrado	-0.000***	(0.00)
Numero de hijos 6 años	-0.072***	(0.01)
Numero de hijos 6-12 años	-0.018**	(0.01)
Observations	10932	

Errores estandar en parentesis.

Fuente: INEI -2021.

Elaboracion: Autor

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

- Intretando los efectos marginales de manera precisa **rn6** y **rn12**
- Un niño adicional menor o igual a 6 años de edad **reduce** la participación femenina en el mercado laboral en **7.2 puntos porcentuales**, en promedio y *ceteris paribus*
- Un niño adicional menor de edad es 7-12 años de edad **reduce** la participación femenina en el mercado laboral en **1.8 puntos porcentuales**, en promedio y *ceteris paribus*

Pregunta 2.3



```
name: <unnamed>
log: C:396L2/Tablas/resultados_4.log
log type: text
opened on: 22 Jun 2024, 04:00:51
```

```
. quietly probit rflp $Zs
. margins , dydx(*) post
```

Average marginal effects
Model VCE: OIM

Number of obs = 10,932

Expression: Pr(rflp), predict()
dy/dx wrt: reduca rpareja redad redadsq rnh6 rnh12

	Delta-method						
	dy/dx	std. err.	z	P> z	[95% conf. interval]		
reduca	-.0093765	.0008467	-11.07	0.000	-.0110359	-.0077717	
rpareja	-.0494196	.0129987	-3.80	0.000	-.0748965	-.0239428	
redad	.0157633	.0014931	10.56	0.000	.0128368	.0186898	
redadsq	-.0002331	.000013	-17.87	0.000	-.0002586	-.0002075	
rnh6	-.0723078	.0073521	-9.83	0.000	-.0867176	-.057898	
rnh12	-.018235	.007413	-2.46	0.014	-.0327642	-.0037058	

```
. log close
name: <unnamed>
log: C:396L2/Tablas/resultados_4.log
log type: text
closed on: 22 Jun 2024, 04:00:51
```

Pregunta 2.3



```
-----  
name: <unnamed>  
log: C:\396L2\Tablas\resultados_5.log  
log type: text  
opened on: 22 Jun 2024, 04:00:51
```

```
. quietly probit rflp $Zs  
.  
mfx
```

```
Marginal effects after probit  
y = Pr(rflp) (predict)  
= .72730172
```

variable	dy/dx	Std. err.	z	P> z	[95% C.I.]	X
reduca	-.0104313	.00095	-10.93	0.000	-.012301 -.008561	8.1352
rpareja*	-.0528448	.01333	-3.96	0.000	-.078969 -.026721	.870655
redad	.0175367	.00169	10.36	0.000	.014217 .020856	52.9257
redadsq	-.0002593	.00002	-17.04	0.000	-.000289 -.000229	3069.14
rnhd6	-.0804424	.00825	-9.75	0.000	-.096615 -.06427	.282473
rnhd12	-.0202864	.00825	-2.46	0.014	-.036465 -.004108	.318057

```
-----  
(*) dy/dx is for discrete change of dummy variable from 0 to 1
```

```
. log close  
name: <unnamed>  
log: C:\396L2\Tablas\resultados_5.log  
log type: text  
closed on: 22 Jun 2024, 04:00:52  
-----
```



- Use los coeficientes de la ecuación del probit (1) para calcular los **seudo-residuos** del modelo probit y resuma los valores obtenidos
- Cuánto es la **inversa del Ratio de Mills** relacionada a estos seudo residuos?
- Por qué el termino provee es util ara la aplicación?
- La inversa del Ratio de Mills se mide como:

$$= \frac{\phi(z_i)}{\Phi(z_i)}$$

Pregunta 3



```
-----
name: <unnamed>
log: C:396L2/Tablas/resultados_6_1.log
log type: text
opened on: 22 Jun 2024, 04:00:57

.      /*
>      probit rflp reduca rpareja redad redadsq rnh6 rnh12
>
>      predict phat          /* prediccion de la regresion probit*/
>      gen z=invnorm(phat)    /*valores del indice probit estandarizado*/
>
>      ////The Inverse Mills
>      gen den_z=normalden(z) /*valor de densidad de probabilidad */
>      gen sct=den_z/phat     /*Inversa de ratio de Mills*/
>
>      */
.      log close
name: <unnamed>
log: C:396L2/Tablas/resultados_6_1.log
log type: text
closed on: 22 Jun 2024, 04:00:57
-----
```



Los pseudo residuos estan dados por:

$$\hat{\mu} = \frac{y_i - \Phi(\hat{\theta}_i)}{\Phi(\hat{\theta}_i) [1 - \Phi(\hat{\theta}_i)]} \phi(\hat{\theta}_i)$$

Si $y_i = 1$, entonces:

$$\hat{\mu} = \frac{\phi(\hat{\theta}_i)}{\Phi(\hat{\theta}_i)}$$

La inversa del ratio de Mills

Si $y_i = 0$, entonces:

$$\hat{\mu} = -\frac{\phi(\hat{\theta}_i)}{1 - \Phi(\hat{\theta}_i)}$$

El complemento de la inversa del ratio de Mills



```
-----
name: <unnamed>
log: C:396L2/Tablas/resultados_6_2.log
log type: text
opened on: 22 Jun 2024, 04:00:57

.
/*
> quietly probit rflp reduca rpareja redad redadsq rnh6 rnh12
> predict phat /* prediccion de la regresion probit*/
> gen resid=((lfp-phat)/(phat*(1-phat)))*normalden(z)
> sum resid
> sum resid if rflp==1
> */
. log close
name: <unnamed>
log: C:396L2/Tablas/resultados_6_2.log
log type: text
closed on: 22 Jun 2024, 04:00:57
-----
```




- El resumen estadístico para los **pseudo-residuos** son reportados

```
-----  
      name: <unnamed>  
      log: C:396L2/Tablas/resultados_7.log  
log type: text  
opened on: 22 Jun 2024, 04:00:58  
  
.      *Resumen residuos  
.      su residu  
  
Variable |      Obs      Mean  Std. dev.      Min      Max  
-----+-----  
  residu |    10,932    2.20e-10   .714811  -1.943027   2.247264  
  
.      log close  
      name: <unnamed>  
      log: C:396L2/Tablas/resultados_7.log  
log type: text  
closed on: 22 Jun 2024, 04:00:58  
-----
```

- El valor promedio para las 10,932 observaciones es de cero (-0.0000000000220)
- El rango de los pseudo-residuo es -1.9 hacia 2.24



- Los **speudo-residuos** para el modelo probit cuando el evento de interes ocurre (participación femenina, $lfp_i = 1$ son definidos como:

$$\frac{\phi(z_i)}{\Phi(z_i)}$$

- Donde z_i es el **indice estandarizado probit** para la participación de empleo de la i -esima mujer
- Específicamente, el **indice estandarizado probit** en este caso sera:

$$z_i = \alpha_0 + \alpha_1 edad_i + \alpha_3 edad_i^2 + \alpha_4 rpareja_i + \alpha_5 rnh6_i + \alpha_6 rnh12_i$$

Pregunta 3





- La **Inversa del ratio de Mills** (o pseudo-residuos) una medida empirica de los **no observables** en la ecuación de determinación en el empleo (participacion de empleo).



- La **Inversa del ratio de Mills** (o pseudo-residuos) una medida empirica de los **no observables** en la ecuación de determinación en el empleo (participacion de empleo).
- Ello es diseñado para corregir la ecuación de logaritmo de salarios dada cualquier **sesgo de seleccion (bias selection)** dentro del empleo.



- La **Inversa del ratio de Mills** (o pseudo-residuos) una medida empirica de los **no observables** en la ecuación de determinación en el empleo (participacion de empleo).
- Ello es diseñado para corregir la ecuación de logaritmo de salarios dada cualquier **sesgo de seleccion (bias selection)** dentro del empleo.
- Este enfoque trata el efecto de selección (truncamiento) como un problema de variable omitida donde la **variable omitida** es la **Inversa del ratio de Mills**



- La **Inversa del ratio de Mills** (o pseudo-residuos) una medida empirica de los **no observables** en la ecuación de determinación en el empleo (participacion de empleo).
- Ello es diseñado para corregir la ecuación de logaritmo de salarios dada cualquier **sesgo de seleccion (bias selection)** dentro del empleo.
- Este enfoque trata el efecto de selección (truncamiento) como un problema de variable omitida donde la **variable omitida** es la **Inversa del ratio de Mills**
- Por construcción, los pseudo-residuos toman valores **positivos** cuanndo el **evento ocurre**



- La **Inversa del ratio de Mills** (o pseudo-residuos) una medida empirica de los **no observables** en la ecuación de determinación en el empleo (participacion de empleo).
- Ello es diseñado para corregir la ecuación de logaritmo de salarios dada cualquier **sesgo de seleccion (bias selection)** dentro del empleo.
- Este enfoque trata el efecto de selección (truncamiento) como un problema de variable omitida donde la **variable omitida** es la **Inversa del ratio de Mills**
- Por construcción, los pseudo-residuos toman valores **positivos** cuanndo el **evento ocurre**
- Por otro lado, los pseudo-residuos toman valores **negativos** cuanndo el **evento no ocurre**



- Estime la regresión de MCO:

$$\ln r6prin_i = \beta_0 + \beta_1 educ_i + \beta_2 rpareja_i + \beta_3 exper_i + \beta_4 exper_i^2 + v_i$$

- Donde: $i = 1, \dots, 5,877$, dado el término $v_i \sim N(0, \sigma^2)$

Pregunta 4



```
-----  
name: <unnamed>  
log: C:396L2/Tablas/resultados_8.log  
log type: text  
opened on: 22 Jun 2024, 04:00:58
```

```
. reg lnr6prin $Vs sct if rflp==1 ,r
```

```
Linear regression
```

Number of obs	=	7,778
F(5, 7772)	=	674.66
Prob > F	=	0.0000
R-squared	=	0.2739
Root MSE	=	.88455

lnr6prin	Coefficient	Robust std. err.	t	P> t	[95% conf. interval]	
reduca	.0833971	.0032629	25.56	0.000	.0770009	.0897933
rexper	.0318502	.0033979	9.37	0.000	.0251893	.038511
rexpersq	-.0005541	.0000637	-8.70	0.000	-.0006789	-.0004293
rpreja	-.0202588	.0330277	-0.61	0.540	-.085002	.0444843
sct	.2604463	.1308616	1.99	0.047	.0039224	.5169703
_cons	5.266789	.0628651	83.78	0.000	5.143557	5.390022

```
-----  
log close  
name: <unnamed>  
log: C:396L2/Tablas/resultados_8.log  
log type: text  
closed on: 22 Jun 2024, 04:00:58  
-----
```

Pregunta 4



Resultados del modelo MCO

Cuadro: Modelo seleccion

	OLS		(1)		(2)	
años educacion	0.08779***	(0.00)	0.08340***	(0.00)	0.08340***	(0.00)
Experiencia laboral	0.02716***	(0.00)	0.03185***	(0.00)	0.03185***	(0.00)
Experiencia laboral cuadrado	-0.00045***	(0.00)	-0.00055***	(0.00)	-0.00055***	(0.00)
Persona con pareja	0.00744	(0.03)	-0.02026	(0.03)	-0.02026	(0.03)
sct			0.26045**	(0.13)		
resids					0.26045**	(0.13)
Constant	5.32388***	(0.06)	5.26679***	(0.06)	5.26679***	(0.06)
Observations	7778		7778		7778	

Errores estandar en parentesis.

Fuente: EnAHO 2021.

Elaboracion: Autor

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

Los retornos por un año mas de experiencia, incrementan los ingresos del jefe de hogar mujer en 2.7 % en promedio y manteniendo todo lo demas constante.

¿Cuál es el problema de dicha estimación MCO?

Respuesta: La submuestra de trabajadoras (jefe de hogar) puede representar un grupo selectivo en términos de sus no observables y por lo tanto, las estimaciones de la ecuación del logarítmico de salario pueden estar sujetas a un sesgo de selectividad y OLS produce coeficientes sesgados.



Cuadro: Modelo de MCO solo participacion laboral

	OLS	
años educacion	0.08340***	(0.00)
Experiencia laboral	0.03185***	(0.00)
Experiencia laboral cuadrado	-0.00055***	(0.00)
Persona con pareja	-0.02026	(0.03)
sct	0.26045**	(0.13)
Constant	5.26679***	(0.06)
Observations	7778	

Errores estandar en parentesis.

Fuente: ENAHO 2021.

Elaboracion: Autor

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

El parametro $\rho\sigma_v$ Este parametro es el producto de la correlación entre el coeficiente y el error estandar del logaritmo del salario

$$E[\ln wage | lfp = 1, X] = X_i\beta + \rho\sigma_v \frac{\phi(w_i\gamma)}{\Phi(w_i\gamma)}$$

Asimismo la **inversa del ratio de Mills** esta representado por el coeficiente sct



- El **sct** es la variable $\frac{\phi(z_i)}{\Phi(z_i)}$
- El valor estimado para la **inversa del ratio de Mills** es de 0.26044 , indicando una correlación positiva entre el componente no observable que determina la participación del empleo y el log de salarios.
- Esta relación positiva es considerada como una buena percepción para las mujeres en el mercado laboral (mas motivadas)
- La mayor motivación de las mujeres se debe a los altos salarios que pueden percibir al entrar al mercado laboral versus las mujeres que no ingresan al mismo.
- Una mujer que elige pertenecer al mercado laboral (tener un empleo), en promedio, gana un mayor retribucion salarial que una mujer con el mismo conjunto de características observables y dada una asignación aleatoria en la población.

Pregunta 4



- Dado la ecuación del logaritmo de salarios , se puede calcular el efecto de seleccion:

$$E [\ln wage | lfp = 1, X] = 4.66813 + \dots + 0.26 \frac{\phi(w_i \gamma)}{\Phi(w_i \gamma)}$$

efecto de seleccion:

$$0.26 \frac{\phi(w_i \gamma)}{\Phi(w_i \gamma)}$$

```
-----
name: <unnamed>
log: C:396L2/Tablas/resultados_9.log
log type: text
opened on: 22 Jun 2024, 04:00:58

.      sum sct if rflp==1

Variable |      Obs      Mean   Std. dev.   Min      Max
-----+-----
sct      |      7,778   .4195113   .210578   .1204085   2.247264

.      log close
name: <unnamed>
log: C:396L2/Tablas/resultados_9.log
log type: text
closed on: 22 Jun 2024, 04:00:58
-----
```

- Se puede calcular el promedio del efecto de seleccion, como se observa posee un valor de

$$\frac{\phi(w_i \gamma)}{\Phi(w_i \gamma)} = 0.41951$$

Pregunta 4



Pregunta 4



- Entonces el efecto de seleccion sobre el logaritmo de salarios sera $+0.26044 \times 0.419511 = +0.1092$

Pregunta 4



- Entonces el efecto de seleccion sobre el logaritmo de salarios sera $+0.26044 \times 0.419511 = +0.1092$
- El salario de una mujer seleccionada dentro del mercado laboral, en promedio y manteniendo todos los demas factores constantes, es de 10.92 % mas comparada con otra misma mujer pero no seleccionada de forma aleatoria de la población dada el mismo conjunto de características observables.



- Entonces el efecto de seleccion sobre el logaritmo de salarios sera $+0.26044 \times 0.419511 = +0.1092$
- El salario de una mujer seleccionada dentro del mercado laboral, en promedio y manteniendo todos los demas factores constantes, es de 10.92 % mas comparada con otra misma mujer pero no seleccionada de forma aleatoria de la población dada el mismo conjunto de características observables.
- Asimismo, debemos determinar si el coeficiente es estadísticamente significativo y diferente de cero (hipotesis significancia individual)
- Podemos observar que el **t-test** es calculado como 1.19, dado la corrección de la presencia de heterocedasticidad bajo la MVC robustas
- Dado los valores criticos al 95 % , 1.96, la hipotesis nula de **no sesgo de selección sobre la muestra** no es rechazada, sin embargo, al 90 % se confirma la presencia de sesgo de seleccion en la aplicación



- Cual es el rol de las variables rnh6 y rnh12 en la aplicación?
- ¿Que es lo que usted puede observar en la regresión MCO al implementarlos?



- Cual es el rol de las variables rn_{h6} y rn_{h12} en la aplicación?
- Dado estas dos variables son incluidas en la participación laboral y no en la ecuación de log-salarios, dado que ellos son designados como unos **instrumentos**
- Estos instrumentos identificaran los parametros en la ecuación de logaritmo de salarios (modelo de selección)

Pregunta 5





Cuadro: Modelo de MCO

OLS		
años educacion	0.02***	(0.00)
Persona con pareja	0.06*	(0.03)
Experiencia laboral	-0.10***	(0.01)
Experiencia laboral cuadrado	-0.00***	(0.00)
Edad	0.11***	(0.02)
Edad cuadrado	0.00*	(0.00)
Numero de hijos 6 años	-0.01	(0.02)
Numero de hijos 6-12 años	-0.02	(0.02)
Constant	4.02***	(0.29)
Observations	7778	

Errores estandar en parentesis.

Fuente: ENAHO 2021.

Elaboracion: Autor

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

- Entonces, se puede observar que la variable rnh6 no es estadística significativa, la cual indicaría ser un buen instrumento de identificación en el modelo de seleccion
- Asimismo, la variable rnh12 no es un buen instrumento al ser significativa.

Pregunta 5





Cuadro: Modelo Heckman

Heckman		
log-salarios (Ocup. princ)		
años educacion	0.08***	(0.00)
Experiencia laboral	0.03***	(0.00)
Experiencia laboral cuadrado	-0.00***	(0.00)
Persona con pareja	-0.02	(0.03)
Constant	5.27***	(0.07)
<hr/>		
select		
años educacion	-0.03***	(0.00)
Persona con pareja	-0.17***	(0.04)
Edad	0.05***	(0.01)
Edad cuadrado	-0.00***	(0.00)
Numero de hijos 6 años	-0.24***	(0.02)
Numero de hijos 6-12 años	-0.06**	(0.02)
Constant	0.69***	(0.14)
<hr/>		
/mills		
lambda	0.26*	(0.14)
<hr/>		
Observaciones	10932	
Pseudo. R ²		
λ	0.26	
σ	0.90	
ρ	0.29	

Fuente: ENAHO - 2021.

Elaboracion: Autor

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

- Se puede observar

$$\lambda = \rho \cdot \sigma$$

- Donde:

- ρ : el coeficiente de correlacion entre las **no observables** en el empleo y el logaritmo de salario (0.28904)
- σ : el error estandar del logaritmo del salario es (0.90106)

- Por tanto,

$$\lambda = (0.28904) \times (0.90106) = +0.26004$$