

# Modelos No Lineales

## Clase Modelo

Edinson Tolentino

email: [edinson.tolentino@gmail.com](mailto:edinson.tolentino@gmail.com)

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

Educate Peru

2 de febrero de 2024

- 1 Introducción
- 2 Data y Variables
- 3 Pregunta 1
- 4 Pregunta 2
- 5 Pregunta 3
- 6 Pregunta 4
- 7 Pregunta 5
- 8 Pregunta 6

- La propuesta del presente ejercicio es para examinar el impacto de variables **demografico** y **economica** sobre **laborar en una administracion publica**
- Las variables demograficas relacionadas al genero y edad
- Las variables economicas (o oferta laboral) relacionada hacia las características individuales **empleo principal** (por ejemplo, salarios del empleo y principios horas de trabajo)
- En otras palabras, examinaremos los factores que se encuentran asociados (correlaciones) para poder laborar en una entidad publica.

- La información que se utilizará es proveniente de la base de datos de la Encuesta Nacional de Hogares (ENAH). Se procesa la base de datos del modulo 500 donde se analizará los determinantes de una persona jefe de hogar labore en la administrtracion publica.

**Cuadro:** Descripción de variables

Variables	Descripción
rpublica	== 1 , si el jefe de hogar labora en la Adm. Publica
lnr6	Logaritmo ingreso mensual (Soles)
redad	años de edad
rmujer	==1, mujer

**Cuadro:** Estadísticas descriptivas

	Personas	Promedio	Mediana	Min.	Max.	Std
==1, Adm. Publica	22624	0.11	0.00	0.00	1.00	0
rmujer	22624	0.28	0.00	0.00	1.00	0
==1, Situacion de empleo informal	22624	0.74	1.00	0.00	1.00	0
Nivel de ingresos (log)	22624	6.93	6.96	4.79	10.86	1
edad	22624	45.69	46.00	19.00	65.00	11
reduca_niv==Elemental	22624	0.31	0.00	0.00	1.00	0
reduca_niv==Secundaria	22624	0.42	0.00	0.00	1.00	0
reduca_niv==Superior no universitaria	22624	0.13	0.00	0.00	1.00	0
reduca_niv==Superior universitaria	22624	0.14	0.00	0.00	1.00	0

Fuente: ENAHO - 2021.

Elaboracion: Autor

- Se realizará la estimación de la ecuación de salarios de los trabajadores peruanos utilizando la información 2021 (ENAH0)
- Se propone la siguiente especificación :

$$prob[second_i = 1] = \Phi(\alpha_0 + \alpha_1 rmujer_i + \alpha_2 rinfo_i + \alpha_3 lnr6_i + \alpha_4 redad + \gamma X) \quad (1)$$

- Donde  $i = 1, 2, \dots, n$  y  $\Phi(\bullet)$  denota la función de distribución acumulada para el operador de una normal estandar
- $X$  determina variables de control como nivel educativo.
- Ello implica una un modelo de regresión probit dado al operador CDF especifico

- Comparamos el modelo probit dado la inclusion y no inclusion de variables control

**Cuadro:** Modelo No Lineal: Probit

	(1)		(2)	
<hr/>				
=1, Adm. Publica				
rmujer	0.42***	(0.03)	0.36***	(0.03)
=1, Situacion de empleo informal	-1.03***	(0.03)	-0.81***	(0.03)
Nivel de ingresos (log)	0.49***	(0.02)	0.35***	(0.02)
edad	0.01***	(0.00)	0.02***	(0.00)
Constant	-4.96***	(0.17)	-4.60***	(0.18)
<hr/>				
Observaciones	22624		22624	
Pseudo. R <sup>2</sup>	0.288		0.320	
Log-L	-5509.6		-5265.4	
Grados de Libertad (k)	5		9	
Controls			✓	

Fuente: ENAHO - 2021.

Elaboracion: Autor

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

Cuadro: Modelos Lineal vs No Lineal

	OLS		Probit	
main				
rmujer	0.05***	(0.00)	0.36***	(0.03)
=1, Situacion de empleo informal	-0.18***	(0.01)	-0.81***	(0.03)
Nivel de ingresos (log)	0.04***	(0.00)	0.35***	(0.02)
edad	0.00***	(0.00)	0.02***	(0.00)
Elemental	0.00	(.)	0.00	(.)
Secundaria	0.00	(0.00)	0.14***	(0.04)
Superior no universitaria	0.09***	(0.01)	0.64***	(0.05)
Superior universitaria	0.19***	(0.01)	0.87***	(0.05)
Constant	-0.22***	(0.02)	-4.60***	(0.18)
Observaciones	22624		22624	
R <sup>2</sup>	0.251			
Pseudo. R <sup>2</sup>			0.320	
Log-L	-2352.1		-5265.4	
Grados de Libertad (k)			9	

Fuente: ENAHO - 2021.

Elaboracion: Autor

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

- Cuales son las diferencias entre el Modelo MCO vs Probit
- Existe una diferencia entre la lectura de los resultados del coeficientes, justifique su respuesta



- 1 Estime el modelo de la ecuación (1) . Interprete precisamente los estimadores de máxima verosimilitud (maximum likelihood) para  $\alpha_1$  y  $\alpha_3$  para este caso.

**Cuadro:** Modelos No Lineal: Probit

	Probit	
==1, Adm. Publica		
rmujer	0.36***	(0.03)
==1, Situacion de empleo informal	-0.81***	(0.03)
Nivel de ingresos (log)	0.35***	(0.02)
edad	0.02***	(0.00)
Elemental	0.00	(.)
Secundaria	0.14***	(0.04)
Superior no universitaria	0.64***	(0.05)
Superior universitaria	0.87***	(0.05)
Constant	-4.60***	(0.18)
Observaciones	22624	
R <sup>2</sup>	0.320	
Pseudo. R <sup>2</sup>	-5265.4	
Log-L	9	

Fuente: ENAHO - 2021.

Elaboracion: Autor

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

- Provea una interpretación para los estimadores  $\alpha_1$  (variable mujer) y  $\alpha_3$  (correspondiente logaritmo de salarios)?.
- Si el jefe de hogar es una mujer se observa un **disminuye** en el **indice estadarizado de probit por laborar en una Adm. Publica en 0.104 desviaciones estandar** respecto a su par hombre , en promedio y manteniendo todo lo demas constante
- Un incremento de 10 % en los salarios mensuales , **aumenta el indice estandarizado probit para laborar en una Adm. Publica en 0.035 desviaciones estandar**, en promedio y manteniendo todo lo demas constante

- El índice estandarizado de probit ( $z$ ) es expresado como:

$$z_i = \alpha_0 + \alpha_1 rmujeri + \alpha_2 rinfo_i + \alpha_3 \ln r6_i + \alpha_4 redad_i + \gamma X$$

- Donde:

$$\frac{\partial z}{\partial \ln r6} = \hat{\alpha}_3 = 0.35$$

$$\partial \ln r6 = \frac{\partial r6}{r6}$$

- En otras palabras,  $\partial \ln r6$  es el cambio proporcional en el salario mensual
- Un incremento de **10 %** en el salario mensual es expresado como una proporción de **0.10**
- Entonces,  $\partial \ln r6 = 0.10$

$$\partial z = 0.35 \times \partial \ln r6$$

$$\partial z = 0.35 \times 0.10 = 0.035$$

- Usando el comando **margins** calcule los efectos marginales (impacto) sobre las covariables y analice las probabilidades para laborar en una Adm. Publica.

## Cuadro: Efectos Marginales probit

	Efectos Marginales
rmujer	0.046*** (0.00)
==1, Situacion de empleo informal	-0.104*** (0.00)
Nivel de ingresos (log)	0.045*** (0.00)
edad	0.002*** (0.00)
Elemental	0.000 (.)
Secundaria	0.014*** (0.00)
Superior no universitaria	0.085*** (0.01)
Superior universitaria	0.132*** (0.01)
Observations	22624

Errores estandar en parentesis.

Fuente: INEI -2021.

Elaboracion: Autor

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

- Las variables **rmujer** y **rinfo** son variables demográficas en la ecuación de oferta laboral. Ambas son variables dummies.
- El efecto estimado sobre estas variables dummies son conocidos **efecto impacto**

**Cuadro:** Efectos Marginales probit

	Efectos Marginales
rmujer	0.046*** (0.00)
==1, Situacion de empleo informal	-0.104*** (0.00)
Nivel de ingresos (log)	0.045*** (0.00)
edad	0.002*** (0.00)
Elemental	0.000 (.)
Secundaria	0.014*** (0.00)
Superior no universitaria	0.085*** (0.01)
Superior universitaria	0.132*** (0.01)
Observations	22624

Errores estandar en parentesis.

Fuente: INEI -2021.

Elaboracion: Autor

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

- Las variables **rmujer** y **rinfo** son variables demográficas en la ecuación de oferta laboral. Ambas son variables dummies.
- El efecto estimado sobre estas variables dummies son conocidos **efecto impacto**
- El efecto genero estimado sugiere que si el trabajador es mujer , en promedio y ceteris paribus , **aumenta** la probabilidad de laborar en el sector publico en **4.6** puntos porcentuales
- El efecto genero estimado sugiere que si el trabajador es informal , en promedio y ceteris paribus , **reduce** la probabilidad de laborar en el sector publico en **10.4** puntos porcentuales

- El índice estandarizado de probit calculado presenta un valor de 1.51 ( $z = 1.51$ ) considerando las características promedio. ¿Cómo interpretar dicho valor?

- Utilizaremos la información del modelo probit (coeficiente) y los resultados del efectos marginal para corroborar.
- Calcule el **efecto marginal** para **lnr6** sobre la probabilidad del segundo empleo
- En general , la formula para el efecto marginal de una  $k^{th}$  variable sobre la probabilidad de **laboar el la Adm. Publica** para el modelo probit esta dado por:

$$\frac{\partial Prob(AdmPublica = 1)}{\partial X_k} = \phi(z) \times \beta_k$$

- Donde  $\beta_k$  es el coeficiente del modelo probit correspondiente para la  $X_k$  variable.



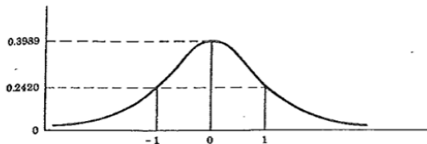
- Se conoce el valor de  $\beta_k$  en este caso
- Según el modelo del probit estimado para  $\ln r6 = 0.35$
- Nosotros conocemos el valor de  $z$  usando la muestra de la data es 1.51
- Nosotros evaluamos la curva normal estandar para este valor en  $z$
- En otras palabras, nosotros calculamos  $\phi(1.51)$

¿Cómo procesamos y analizamos esto?

$$\phi(-1.51) = \phi(1.51)$$

- Calculamos la *pdf* de la table conteniendo las coordenadas de la distribución normal estandar

### Ordinates of the Normal Curve



#### Example

$$z = \frac{x - \mu}{\sigma}$$

$$z = 0 : \text{ordinate} = 0.3989$$

$$z = 1 : \text{ordinate} = 0.2420$$

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.3989	.3989	.3989	.3988	.3986	.3984	.3982	.3980	.3977	.3973
.1	.3970	.3965	.3961	.3956	.3951	.3945	.3939	.3932	.3925	.3918
.2	.3910	.3902	.3894	.3885	.3876	.3867	.3857	.3847	.3836	.3825
.3	.3814	.3802	.3790	.3778	.3765	.3752	.3739	.3725	.3712	.3697
.4	.3683	.3668	.3653	.3637	.3621	.3605	.3589	.3572	.3555	.3538
.5	.3521	.3503	.3485	.3467	.3448	.3429	.3410	.3391	.3372	.3352
.6	.3332	.3312	.3292	.3271	.3251	.3230	.3209	.3187	.3166	.3144
.7	.3123	.3101	.3079	.3056	.3034	.3011	.2989	.2966	.2943	.2920
.8	.2897	.2874	.2850	.2827	.2803	.2780	.2756	.2732	.2709	.2685
.9	.2661	.2637	.2613	.2589	.2565	.2541	.2516	.2492	.2468	.2444
1.0	.2420	.2396	.2371	.2347	.2323	.2299	.2275	.2251	.2227	.2203
1.1	.2179	.2155	.2131	.2107	.2083	.2059	.2036	.2012	.1989	.1965
1.2	.1942	.1919	.1895	.1872	.1849	.1826	.1804	.1781	.1758	.1736
1.3	.1714	.1691	.1669	.1647	.1626	.1604	.1582	.1561	.1539	.1518
1.4	.1497	.1476	.1456	.1435	.1415	.1394	.1374	.1354	.1334	.1315
1.5	.1295	.1276	.1257	.1238	.1219	.1200	.1182	.1163	.1145	.1127
1.6	.1109	.1092	.1074	.1057	.1040	.1023	.1006	.0989	.0973	.0957
1.7	.0940	.0925	.0909	.0893	.0878	.0863	.0848	.0833	.0818	.0804
1.8	.0790	.0775	.0761	.0748	.0734	.0721	.0707	.0694	.0681	.0669
1.9	.0656	.0644	.0632	.0620	.0608	.0596	.0584	.0573	.0562	.0551

- Por tanto, calculamos  $\phi(1.51) = 0.1276$  usando la tabla

$$\frac{\partial \text{Prob}(\text{AdmPublica} = 1)}{\partial \ln r6} = \phi(z) \times \beta_{\ln r6}$$

$$\frac{\partial \text{Prob}(\text{AdmPublica} = 1)}{\partial \ln r6} = \phi(1.51) \times \beta_{\ln r6}$$

$$\frac{\partial \text{Prob}(\text{AdmPublica} = 1)}{\partial \ln r6} = \phi(1.51) \times \beta_{\ln r6}$$

$$\frac{\partial \text{Prob}(\text{AdmPublica} = 1)}{\partial \ln r6} = \phi(1.51) \times (0.35)$$

$$\frac{\partial \text{Prob}(\text{AdmPublica} = 1)}{\partial \ln r6} = 0.1276 \times (0.35)$$

$$\frac{\partial \text{Prob}(\text{AdmPublica} = 1)}{\partial \ln r6} = 0.0445$$

- Esto es (aproximadamente) el estimado para el efecto marginal reportado por STATA usando **margins** del cual es  $0.0445 \equiv 0.0445$

- Por lo tanto:

$$\frac{\partial \text{Prob}(\text{AdmPublica} = 1)}{\partial \ln r6} = 0.0445$$

- Ahora, interpretando el efecto de 5 % en un incremento en lo salarios
- El efecto estimado para este cambio es calculado como:

$$0.0445 \times 0.05 = 0.0022$$

- Por tanto, un incremento en los salarios de 5 % reduce el pluriempleo o segunda ocupación en 0.22 puntos porcentual , en promedio & *ceteris paribus*.

- La varianza de una variable aleatoria distribuida logistica es  $\frac{\pi^2}{3}$
- Use esta información para proveer un estimado aproximado del coeficiente del modelo logistico correspondiente al modelo estimado probit obtenido por  $\alpha_3$
- Interprete este logit estimado

- El coeficiente probit es un coeficiente estandarizado definido como:

$$\beta_{probit} \div \sigma$$

- Sin embargo, dado  $\sigma$  es el igual a 1 en el probit (dado el supuesto de la normal estandar) es usualmente expresada como  $\beta_{probit}$
- En contraposición , el coeficiente es no estandarizado
- Si nosotros estandarizamos el coeficiente logit por dividir esto a través de esto , por su derivación estandar esto podria ser comparado por el coeficiente probit

- La **varianza** para el logístico esta dado por:

$$\frac{\pi^2}{3}$$

- La **desviación estandar** para el logit esta dado por

$$\frac{\pi}{\sqrt{3}} = \sigma_{logit}$$

- Por tanto:

$$\frac{\beta_{logit}}{\sigma_{logit}} = \beta_{probit}$$

$$\beta_{logit} = \beta_{probit} \times \sigma_{logit}$$

- Entonces

$$\frac{\pi}{\sqrt{3}} = 1.8138$$

- Por tanto,  $\beta_{logit} \approx 0.35 \times 1.8138 \equiv .6320012$  (diferente computacionalmente a lo encontrado en stata)
- Esto ahora tiene un **log odds ratio** interpretación
- Por tanto, un incremento de **10%** en los salarios ,en promedio y ceteris paribus , aumenta el odds ratio de laborar en la Adm. Publica en 0.63 (aproximado)



► Comparacion de los efectos marginales de los modelos no lineales

**Cuadro:** Efectos Marginales Modelos No Lineales

	MPL	Probit	Logit
rmujer	0.050*** (0.00)	0.046*** (0.00)	0.046*** (0.00)
==1, Situacion de empleo informal	-0.184*** (0.01)	-0.104*** (0.00)	-0.110*** (0.00)
Nivel de ingresos (log)	0.042*** (0.00)	0.045*** (0.00)	0.040*** (0.00)
redad	0.003*** (0.00)	0.002*** (0.00)	0.002*** (0.00)
Observaciones	22624	22624	22624
Pseudo. R <sup>2</sup>			
Log-L			
Grados de Libertad (k)			
Controls			

Errores estandar en parentesis.

Fuente: INEI -2021.

Elaboracion: Autor

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