

Laboratorio 1 - Microeconometría

Fernando Cotrina Lejabo

Abril 2025

1 Regresión Lineal

1. Importe la base de datos a utilizar y familiarícese con las observaciones y las variables disponibles, i.e., observe el tamaño de la muestra, los nombres y etiquetas de las variables disponibles.

Tenemos la siguiente tabla:

Table 1: Descripción de la base de datos

Información	Valor	Detalle
Observaciones	91,668	Número total de individuos en la muestra
Variables	31	Número total de variables disponibles
Fecha	29 Ago 2020	Fecha de creación/actualización

Variable	Tipo	Descripción
conglome	str6	Número de conglomerado
vivienda	str3	Número de selección de vivienda
hogar	str2	Número secuencial del hogar
codperso	str2	Número de orden de la persona
ubigeo	str6	Ubicación geográfica
estrato	byte	Estrato geográfico
sexo	byte	Sexo del individuo (1=Hombre, 2=Mujer)
edad	byte	Edad en años cumplidos
niv_edu	byte	Nivel educativo alcanzado
ocu500	byte	Indicador de PEA (Población Económicamente Activa)
fac500a	float	Factor de expansión para empleo/ingresos
ing_lab	float	Ingreso laboral mensual
ing_total	float	Ingreso total del hogar
ln_ing_lab	float	Logaritmo natural del ingreso laboral
pared	float	Material de paredes (1=Ladrillo)
pisos	float	Material de pisos (1=Madera/parquet)
serv_agua	float	Acceso a agua potable (1=Sí)
sshh	float	Baño conectado a red pública (1=Sí)
serv_luz	float	Servicio eléctrico (1=Sí)
area	byte	Zona geográfica (1=Urbano, 2=Rural)

2. Considere que la población en edad de trabajar para el caso peruano comprende a las personas entre 14 y 64 años: haga los cambios pertinentes. (**Ver archivo .do**)

3. Observe las características estadísticas de la variable de ingreso mensual. Haga lo mismo con ayuda visual: Muestre y guarde en formato Stata Graph (gph) un histograma de esta variable. ¿Observa algún potencial problema? ¿Qué es un outlier?

Podemos visualizar la siguiente distribución:

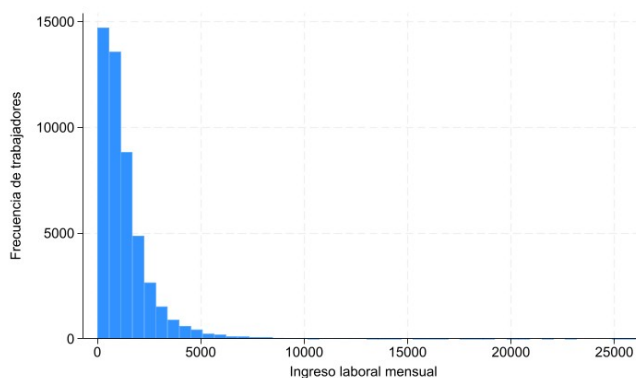


Figure 1: Distribución de los Ingresos Laborales Mensuales.

¿Qué es un outlier? Un outlier (valor atípico) es una observación que se desvía significativamente del resto de los datos. Vemos varios problemas:

- (a) La Curtosis es de 30.61, muy superior a 3 (valor de referencia para una normal), lo que sugiere una distribución con colas pesadas y picos pronunciados.
- (b) La Asimetría es de 3.77 lo que indica una fuerte asimetría positiva (cola derecha larga).
- (c) El valor mínimo es 1.75, probablemente un error de medición. Además, el valor máximo es de 26,018.33 extremadamente alto comparado con el percentil 99% (6,718) y la media (1,313.76). Esto sugiere valores atípicos que distorsionan la media.

Table 2: Estadísticos descriptivos del ingreso laboral

Estadístico	Valor
Observaciones	49,136
Media	1,313.76
Desviación estándar	1,371.84
Varianza	1,881,933.92
Mínimo	1.75
Máximo	26,018.33
Asimetría	3.77
Curtosis	30.61
Suma	64,552,821.89

Table 3: Distribución percentil del ingreso laboral

Percentil	Valor	Percentil	Valor
1%	35.92	75%	1669.42
5%	110.00	90%	2704.42
10%	199.50	95%	3651.08
25%	471.17	99%	6718.00
50% (Mediana)	975.67		

4. Trimming/Acotando: Acote la variable excluyendo a individuos con ingresos mensuales muy altos o bajos. Muestre las características estadísticas de la nueva versión. ¿Los ingresos se distribuyen normalmente? Muestre un histograma de esta nueva versión, pero ahora mejore su gráfico y guárdelo en Portable Document Format (PDF) y en formato Joint Photographic Experts Group (JPG). Concepto nuevo: winsorize (E2TEE9-2025 I).

Tenemos lo siguiente.

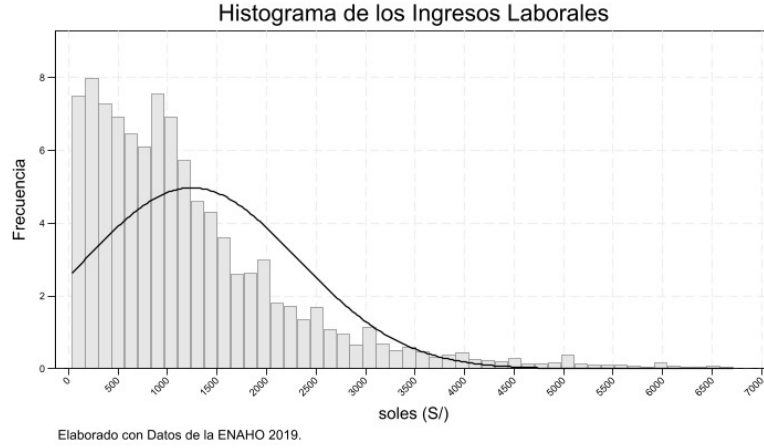


Figure 2: Distribución de los Ingresos Laborales Mensuales Mejorado.

Podemos observar ahora que la distribución de los ingresos tiene una forma muy parecida a una normal. Además, observando las estadísticas descriptivas tenemos que su asimetría es mucho menor que antes y la curtosis también disminuyó.

Table 4: Estadísticos descriptivos del ingreso laboral (actualizado)

Estadístico	Valor
Observaciones	48,151
Media	1,244.69
Desviación estándar	1,070.97
Varianza	1,146,978
Mínimo	36.08
Máximo	6,711.83
Asimetría (Skewness)	1.74
Curtosis	6.80
Suma total	48,151

Table 5: Distribución percentil del ingreso laboral (actualizado)

Percentil	Valor	Percentil	Valor
1%	56.00	75%	1640.92
5%	127.33	90%	2595.08
10%	211.17	95%	3428.67
25%	482.58	99%	5198.33
50% (Mediana)	975.75		

5. Plantee una regresión lineal en la que muestre la relación entre el ingreso laboral de la persona y su edad. Luego, plantee la misma relación, pero usando como variable dependiente el logaritmo del salario. ¿Cómo se interpretan los resultados?.

Estos fueron los resultados:

- (a) Modelo 1: Un año más de edad incrementa el ingreso mensual en aproximadamente 6.25 soles.
(b) Modelo 2: Un año más de edad incrementa el ingreso mensual en aproximadamente 0.35

Table 6: Resultados de regresión lineal (modelos simplificados)

Variable	Modelo 1: <code>ing_lab</code>	Modelo 2: <code>ln_ing_lab</code>
Edad	6.252*** (0.350)	0.0035*** (0.00035)
Constante	991.243*** (13.999)	6.594*** (0.0148)
Observaciones	48,151	
R-cuadrado	0.0057	0.0021

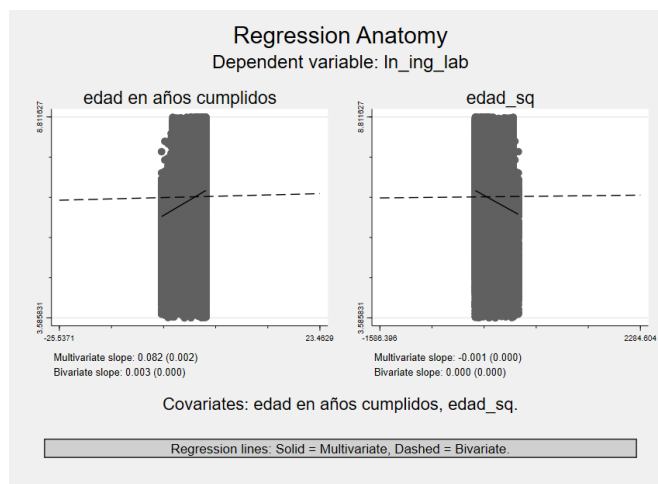
- Significancia: *** $p < 0.01$.
- Modelo 2: Valor = $e^{0.0034877} - 1 = 0.0035$

6. Plantee el modelo anterior, pero ahora incluya como variable de control la edad al cuadrado de la persona (edad²). ¿Qué puede decir de la nueva especificación? ¿Cómo se interpretan los resultados?

Table 7: Regresión del logaritmo del ingreso laboral con efecto cuadrático de edad

Variable	Coefficiente	Error Est.	t	$P > t $
Edad	0.081545	0.002172	37.55	0.000
Edad ²	-0.000969	0.000027	-36.39	0.000
Constante	5.185451	0.041303	125.55	0.000
Observaciones	48,151			
R ²	0.0289			
R ² ajustado	0.0288			
F-stat	715.33 ($p = 0.000$)			

7. Muestre gráficamente lo que halló en el apartado anterior utilizando el comando `reganat`.



2 Variables Factoriales

8. Estime el mismo modelo reemplazando edad y edad² con su contraparte factorial y muestre que los resultados son los mismos. Muestre la información guardada luego de correr la regresión.

Estos fueron los resultados:

Table 8: Regresión cuadrática del log-ingreso laboral

Variable	Coefficiente	(Error Est.)
Edad	0.081 545***	(0.002191)
Edad ²	−0.000 969***	(0.000027)
Constante	5.185 451***	(0.041243)
Observaciones	48,151	
R ²	0.0289	

Errores estándar robustos entre paréntesis.

*** p < 0.001

9. Pruebe que la edad como polinomio tiene significancia sobre los ingresos..

La tabla 9 muestra que para ambos casos de significancia, individual y grupal, el pvalue es menor a 0.05, por lo que nuestro coeficiente es estadísticamente significativo.

Table 9: Pruebas de significancia para el modelo cuadrático

Prueba	Hipótesis Nula (H_0)	Estadístico F	Valor p
Individual	$\beta_{\text{Edad}^2} = 0$	1279.35	0.0000
Conjunta	$\beta_{\text{Edad}} = \beta_{\text{Edad}^2} = 0$	713.12	0.0000

Grados de libertad: 1 y 48,148 (individual); 2 y 48,148 (conjunta)

3 Estandarización

10. Estandarice las variables material de los pisos, las paredes, el acceso a servicios básicos y el equipamiento del hogar. **(Ver archivo .do)**
11. Cree el indicador de riqueza como la agregación (suma) de estas variables estandarizadas. **(Ver archivo .do)**
12. Estandarice el indicador.

Table 10: Comparación de indicadores de riqueza

Indicador	Obs	Media	Desv. Est.	Mínimo	Máximo
IR (Original)	48.151	2.988	1.869	−1.749	8.358
s_IR (Estandarizado)	48.151	-5.74×10^{-9}	1	−2.535	2.874

13. Corra tres regresiones: el modelo univariado, el bivariado y uno en el que controle por riqueza, todos contra los ingresos estandarizados. Muestre una tabla en STATA comparando los betas estimados obtenidos en cada regresión. Muestre significancia con asteriscos. **(Ver archivo .do)**
14. Haga lo mismo que en el apartado anterior, pero esta vez muestre los resultados como tablas en una cuadrícula de STATA comparando los betas estimados obtenidos en cada regresión. Esto ayuda a copiar y pegar como Excel. Muestre significancia con asteriscos. **(Ver archivo .do)**

Table 11: Resultados de regresión múltiple

Variable	(1)	(2)	(3)
Edad	0.00584*** (17.85)	0.0762*** (39.13)	0.0659*** (35.90)
Edad ²		-0.000873*** (-34.82)	-0.000751*** (-32.16)
IR Est.			0.394*** (89.53)
Constante	-0.237*** (-18.10)	-1.506*** (-44.53)	-1.310*** (-39.91)
Observaciones	48,151	48,151	48,151

Nota: Estadísticos t entre paréntesis.

*p<0.05, **p<0.01, ***p<0.001

15. Haga lo mismo que en el apartado anterior, pero esta vez mejore la tabla y expórtela en formato Excel (xls). Muestre significancia con asteriscos, muestre las etiquetas de las variables en vez del nombre, muestre solo tres decimales, agregue un título y una nota al pie. (**Ver archivo .do**)