

# Uso de Stata en el análisis de regresión

Maestría en Economía Aplicada  
Año 2022



UNIVERSIDAD  
**AUSTRAL**

Valores que inspiran

# Regresión lineal: pasos previos

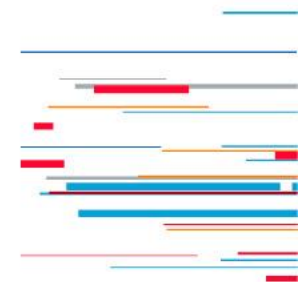
Antes de estimar un modelo de regresión, es útil, como vimos en la primera clase de este curso, hacer una descripción estadística de los datos, así como también construir gráficos que nos permitan entender mejor cómo se comportan las variables. Por otra parte, también puede resultar valioso calcular algunas correlaciones entre la variable dependiente y algunos regresores claves, para lo cual nos valdremos del comando **correlate** y **pwcorr**.

# Comando *regress*

El comando **regress** estima modelos lineales mediante Mínimos Cuadrados Ordinarios (MCO). Luego de su ejecución, Stata muestra la tabla ANOVA, estadísticos sobre la bondad del ajuste, estimación de los coeficientes, errores estándares, estadísticos  $t$ ,  $p$  – *values* e intervalos de confianza. La sintaxis del comando es:

***regress depvar [indepvars] [if] [in] [weight] [, options]***

A continuación, veremos un ejemplo de regresión lineal simple y luego pasaremos a analizar en detalle cómo estimar modelos de regresión lineal múltiple con sus diferentes especificidades.



# Regresión lineal simple: estimación

Como una primera aproximación, veamos un ejemplo de regresión lineal simple utilizando la base *lifeexp.dta* (precargada en Stata). Consideraremos como dependiente a la variable *lexp* (esperanza de vida al nacer), y como independiente a la variable *ln\_gnppc* (logaritmo natural del GNP per cápita). Para estimar este sencillo modelo tipearemos:

```
. regress lexp ln_gnppc
```

Source	SS	df	MS	Number of obs	=	63
Model	873.264865	1	873.264865	F(1, 61)	=	97.09
Residual	548.671643	61	8.99461709	Prob > F	=	0.0000
Total	1421.93651	62	22.9344598	R-squared	=	0.6141
				Adj R-squared	=	0.6078
				Root MSE	=	2.9991

lexp	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
ln_gnppc	2.768349	.2809566	9.85	0.000	2.206542 3.330157
_cons	49.41502	2.348494	21.04	0.000	44.71892 54.11113

# Regresión lineal simple: cálculo de valores ajustados

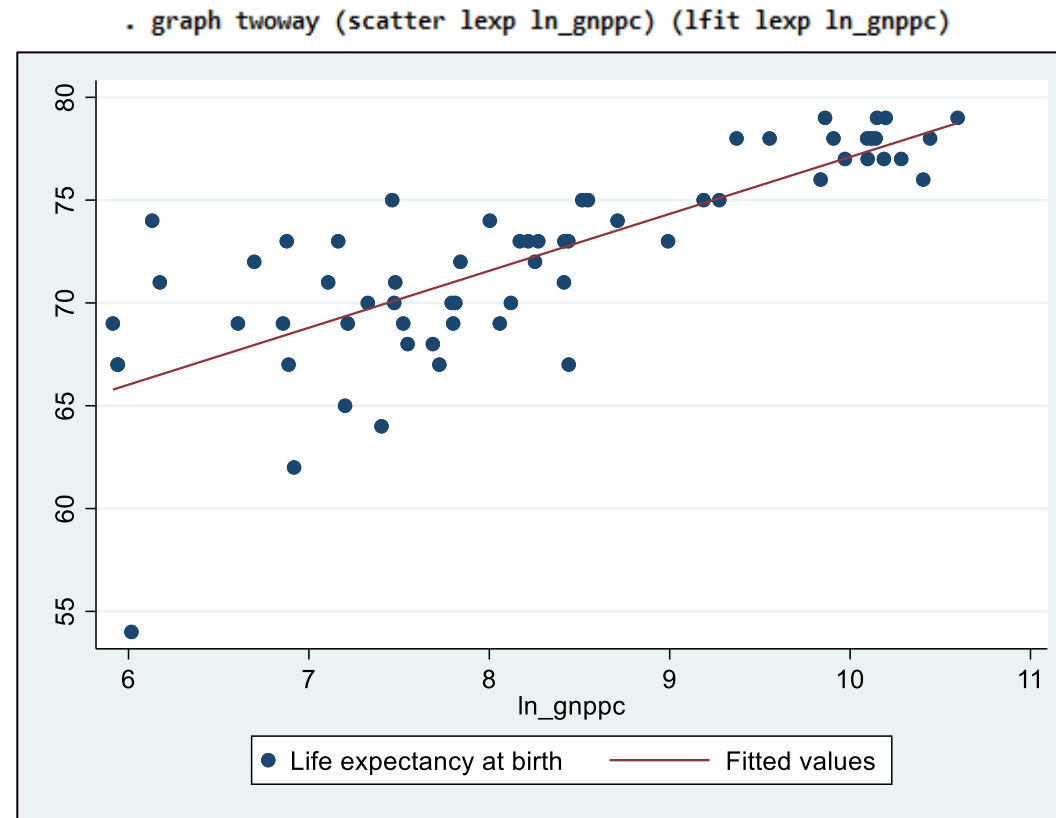
A partir del modelo estimado en el paso anterior, podemos obtener el valor ajustado (o predicho) de la esperanza de vida al nacer, para cada observación que no tenga *missing values* en la variable independiente. Esto lo podemos hacer mediante el comando **predict**, el cual genera una nueva variable con los valores predichos para la variable dependiente del modelo estimado previamente.

```
. predict p_lexp  
(option xb assumed; fitted values)  
(5 missing values generated)  
  
. list ln_gnppc lexp p_lexp in 1/10
```

	ln_gnppc	lexp	p_lexp
1.	6.697034	72	67.95475
2.	6.131227	74	66.3884
3.	10.19728	79	77.64465
4.	6.173786	71	66.50622
5.	7.68708	68	70.69555
6.	10.14172	78	77.49084
7.	.	73	.
8.	7.106606	71	69.08859
9.	8.43815	73	72.77477
10.	8.546752	75	73.07542

# Regresión lineal simple: gráfico de la recta de regresión estimada

Podemos visualizar el ajuste del modelo superponiendo la recta de regresión estimada previamente con un diagrama de dispersión para las variables:



# Regresión lineal múltiple: estimación

Ahora pasaremos a estimar diferentes modelos de regresión lineal múltiple. Para ello, siguiendo a Cameron y Trivedi (2009) trabajaremos con la base *mus03data.dta*, la cual contiene datos de gastos médicos de personas de 65 años o más que califican para atención médica bajo el programa Medicare de EE. UU.

Las variables con las que trabajaremos son:

```
. describe totexp suppins phylim actlim totchr age female income
```

variable name	storage type	display format	value label	variable label
<b>totexp</b>	double	%12.0g		<b>Total medical expenditure</b>
<b>suppins</b>	float	%9.0g		<b>=1 if has supp priv insurance</b>
<b>phylim</b>	double	%12.0g		<b>=1 if has functional limitation</b>
<b>actlim</b>	double	%12.0g		<b>=1 if has activity limitation</b>
<b>totchr</b>	double	%12.0g		<b># of chronic problems</b>
<b>age</b>	double	%12.0g		<b>Age</b>
<b>female</b>	double	%12.0g		<b>=1 if female</b>
<b>income</b>	double	%12.0g		<b>annual household income/1000</b>

# Regresión lineal múltiple: estimación

En primer lugar, estimaremos un modelo para explicar el gasto médico total de los individuos:

```
. regress totexp suppins phylim actlim totchr age female income
```

Source	SS	df	MS	Number of obs	=	3,064
Model	5.3081e+10	7	7.5829e+09	F(7, 3056)	=	61.43
Residual	3.7723e+11	3,056	123440213	Prob > F	=	0.0000
Total	4.3031e+11	3,063	140487727	R-squared	=	0.1234
				Adj R-squared	=	0.1213
				Root MSE	=	11110

totexp	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
suppins	829.3004	418.5569	1.98	0.048	8.618926	1649.982
phylim	2426.572	521.3988	4.65	0.000	1404.244	3448.899
actlim	3695.046	566.0882	6.53	0.000	2585.094	4804.998
totchr	1939.73	166.2386	11.67	0.000	1613.779	2265.68
age	-77.37197	33.20847	-2.33	0.020	-142.4852	-12.25877
female	-1257.174	412.6097	-3.05	0.002	-2066.194	-448.1532
income	7.223121	9.284613	0.78	0.437	-10.9816	25.42784
_cons	7370.567	2501.867	2.95	0.003	2465.054	12276.08



# Regresión lineal múltiple: estimación

En segundo lugar, estimaremos el mismo modelo que en la diapositiva anterior, pero utilizando el logaritmo natural del gasto médico como variable dependiente ( $\ln\_totexp$ ). ¿Cómo deben interpretarse los coeficientes?

```
. generate ln_totexp=ln(totexp)
(109 missing values generated)

. regress ln_totexp suppins phylim actlim totchr age female income
```

Source	SS	df	MS	Number of obs	=	2,955
Model	1264.72124	7	180.674463	F(7, 2947)	=	124.98
Residual	4260.16814	2,947	1.44559489	Prob > F	=	0.0000
Total	5524.88938	2,954	1.87030785	R-squared	=	0.2289
				Adj R-squared	=	0.2271
				Root MSE	=	1.2023

ln_totexp	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
suppins	.2556428	.0462264	5.53	0.000	.1650034 .3462821
phylim	.3020598	.0569709	5.30	0.000	.190353 .4137666
actlim	.3560054	.0621118	5.73	0.000	.2342185 .4777923
totchr	.3758201	.0184227	20.40	0.000	.3396974 .4119429
age	.0038016	.0036561	1.04	0.299	-.0033672 .0109705
female	-.0843275	.0455442	-1.85	0.064	-.1736292 .0049741
income	.0025498	.0010194	2.50	0.012	.000551 .0045486
_cons	6.703737	.27676	24.22	0.000	6.161075 7.2464

# Regresión lineal múltiple: estimación

Ahora reemplacemos la variable de ingreso por su logaritmo natural ( $\ln\_income$ ). ¿Cómo debe interpretarse el coeficiente de esta última variable?

```
. regress ln_totexp suppins phylim actlim totchr age female ln_income
```

Source	SS	df	MS	Number of obs	=	2,873
Model	1227.73543	7	175.390776	F(7, 2865)	=	123.25
Residual	4076.96358	2,865	1.42302394	Prob > F	=	0.0000
Total	5304.69902	2,872	1.84704005	R-squared	=	0.2314
				Adj R-squared	=	0.2296
				Root MSE	=	1.1929

ln_totexp	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
suppins	.2299794	.0467824	4.92	0.000	.1382488 .32171
phylim	.2990087	.0573565	5.21	0.000	.1865446 .4114728
actlim	.3741294	.0627597	5.96	0.000	.2510706 .4971882
totchr	.3695771	.0185086	19.97	0.000	.3332856 .4058687
age	.0049231	.0036811	1.34	0.181	-.0022947 .012141
female	-.0700955	.0456081	-1.54	0.124	-.1595234 .0193324
ln_income	.080909	.0264125	3.06	0.002	.0291196 .1326984
_cons	6.474008	.2910909	22.24	0.000	5.90324 7.044777

# Regresión lineal múltiple: cálculo de valores ajustados

Tomando como referencia el primero de los modelos de regresión múltiple que vimos anteriormente, podemos emplear el comando **predict** para crear una variable que contenga los valores de *totexp* predichos por el modelo. Veamos algunas estadísticas descriptivas de ambas variables:

```
. quietly regress totexp suppins phylim actlim totchr age female income
. predict p_totexp
(option xb assumed; fitted values)
```

. summarize p_totexp, detail					. summarize totexp, detail				
Fitted values					Total medical expenditure				
Percentiles		Smallest			Percentiles		Smallest		
1%	426.8721	-463.2244			1%	0	0		
5%	1363.101	-414.3383			5%	112	0		
10%	2020.08	-401.8279	Obs	3,064	10%	393	0	Obs	3,064
25%	3829.5	-372.9354	Sum of Wgt.	3,064	25%	1271	0	Sum of Wgt.	3,064
50%	6247.188		Mean	7030.889	50%	3134.5		Mean	7030.889
			Std. Dev.	4162.886				Std. Dev.	11852.75
		Largest					Largest		
75%	9962.329	20253.95			75%	7151	104823		
90%	12935.71	20355.93	Variance	1.73e+07	90%	17050	108256	Variance	1.40e+08
95%	14716.31	20853.95	Skewness	.5761701	95%	27367	123611	Skewness	4.165058
99%	17334.43	22727.29	Kurtosis	2.685262	99%	62346	125610	Kurtosis	26.26796

# Regresión lineal múltiple: cálculo de valores ajustados

Cuando la variable dependiente del modelo está en logaritmos (por ejemplo, el segundo y tercer modelo de regresión múltiple de los que vinimos analizando), la predicción de la variable en su escala original es un poco más compleja. Dado que  $\exp[E(\ln y)] \neq E(y)$ , no basta con aplicar la función inversa del logaritmo a los pronósticos para obtener los valores predichos de la variable original, sino que deben implementarse algunos ajustes adicionales. Para más detalles, ver Cameron y Trivedi (2009) y Wooldridge (2010).

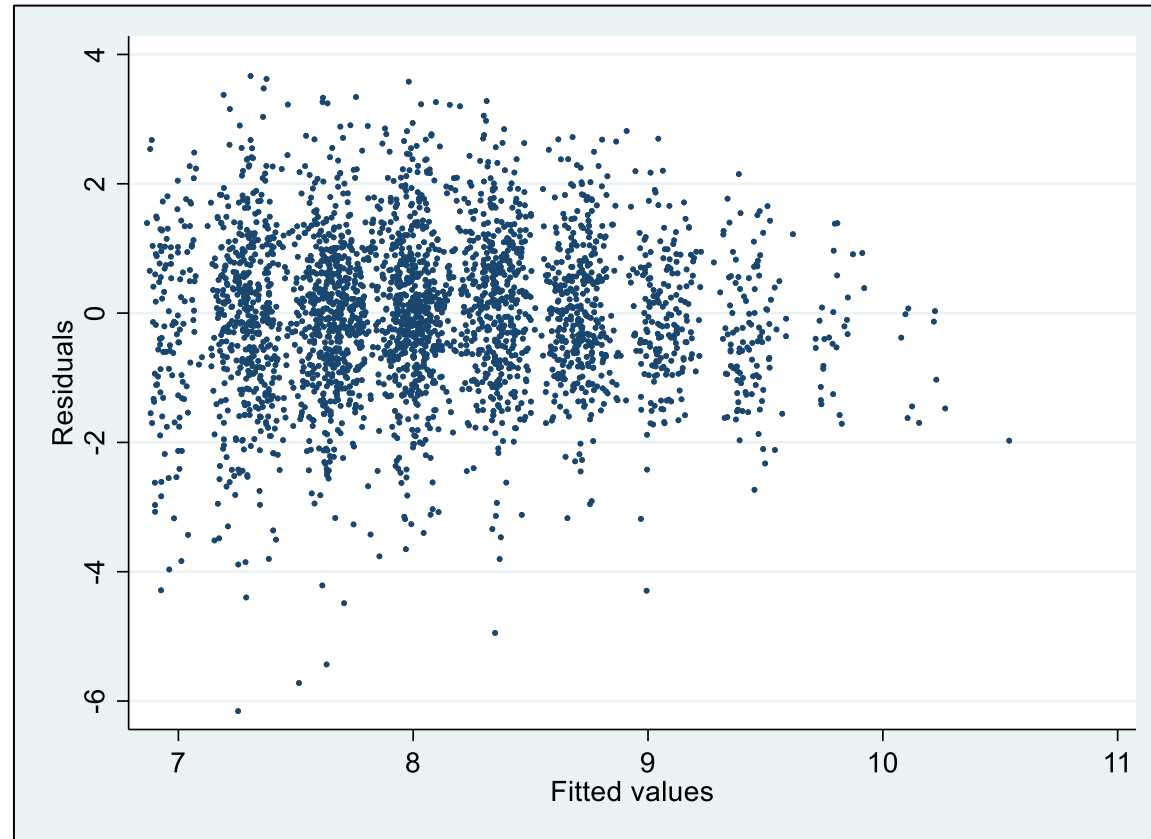
# Regresión lineal múltiple: inspección gráfica de heterocedasticidad

Uno de los supuestos clásicos en el análisis de regresión es el de homocedasticidad del término de error  $u$ , lo cual implica que su varianza, condicional al vector de variables explicativas, es constante. Cuando este supuesto no se cumple, una de las consecuencias es que las versiones por default de los errores estándares de los estimadores MCO no son válidas, lo cual afecta directamente la inferencia acerca de los parámetros del modelo.

En una primera aproximación, veremos cómo podemos inspeccionar gráficamente la presencia de heterocedasticidad, mediante los comandos **rvfplot** y **rvpplot**. El primero grafica un diagrama de dispersión de los residuos contra los valores ajustados de la variable dependiente, mientras que el segundo realiza el mismo tipo de gráfico reemplazando los valores ajustados por los valores de algunas de las variables explicativas.

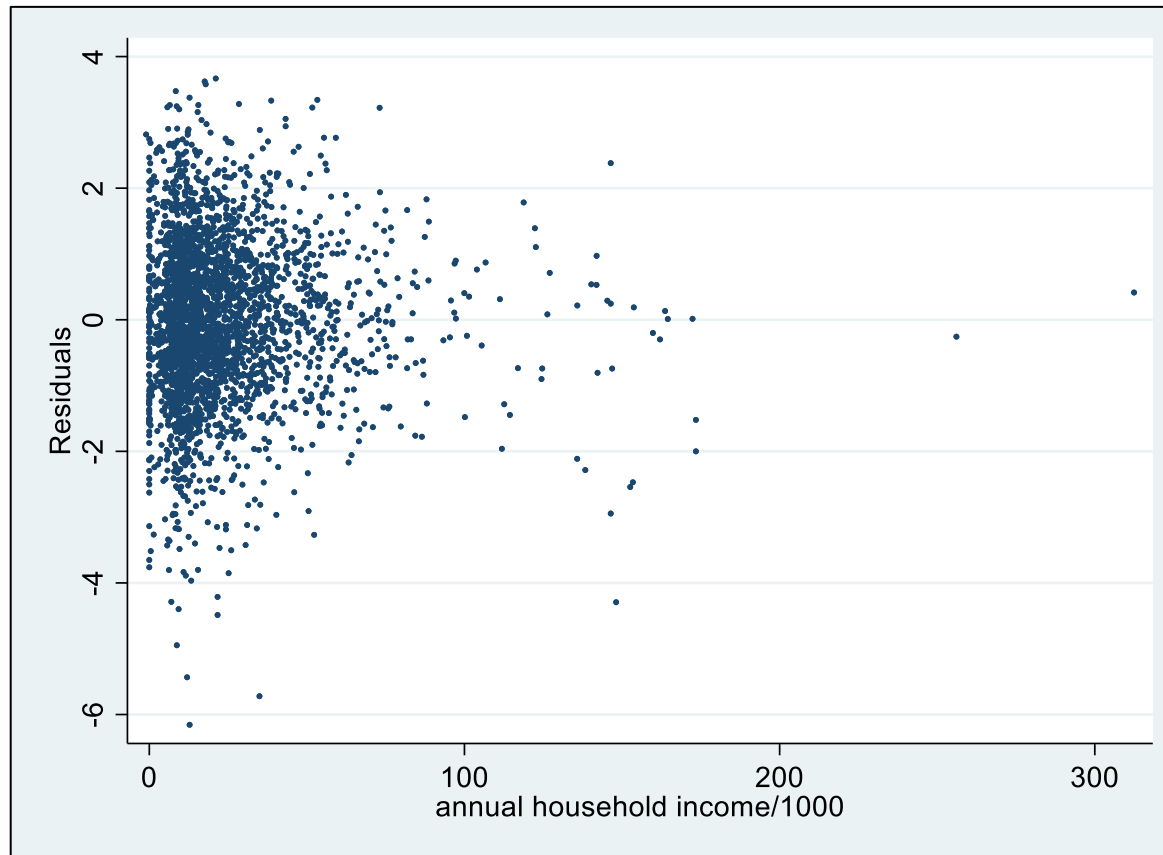
# Regresión lineal múltiple: inspección gráfica de heterocedasticidad

```
. quietly regress ln_totexp suppins phylim actlim totchr age female income  
. rvfplot
```



# Regresión lineal múltiple: inspección gráfica de heterocedasticidad

```
. quietly regress ln_totexp suppins phylim actlim totchr age female income  
. rvpplot income
```



# Regresión lineal múltiple: tests para heterocedasticidad

Un procedimiento más preciso para evaluar la presencia de heterocedasticidad en el modelo consiste en la aplicación de algún test de hipótesis. Para ello, una vez estimado el modelo, podemos ejecutar diferentes comandos en Stata:

- **estat hettest varlist, options:** ejecuta la prueba ML de Breusch-Pagan para heterocedasticidad, en la que la hipótesis nula es que el término de error es homocedástico. En *varlist* pueden incluirse todos los regresores o un subconjunto de éstos (por ejemplo, si se sospecha que sólo ciertas variables independientes están relacionadas con la heterocedasticidad). La opción **iid** relaja el supuesto del test de que los errores se distribuyen normalmente.
- **whitetst:** ejecuta la prueba de White para heterocedasticidad, la cual es una “versión modificada” de la prueba de Breusch-Pagan, puesto que incorpora los términos cuadráticos y los productos de los regresores (en la regresión auxiliar del test). La hipótesis nula es que el término de error es homocedástico.

Por otra parte, no existe un comando en Stata que ejecute directamente la prueba de Goldfeld-Quandt para heterocedasticidad. Sin embargo, mediante algunos pasos, se puede implementar este test manualmente.



# Regresión lineal múltiple: tests para heterocedasticidad

```
. quietly regress ln_totexp suppins phylim actlim totchr age female income
. estat hettest suppins phylim actlim totchr age female income, iid

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
Ho: Constant variance
Variables: suppins phylim actlim totchr age female income

      chi2(7)      =    93.13
      Prob > chi2   =    0.0000

. whitetst

White's general test statistic : 139.9029  Chi-sq(31)  P-value = 8.8e-16
```

# Regresión lineal múltiple: estimadores de varianza alternativos

Una forma de lidiar con el problema de heterocedasticidad en el término de error es utilizar estimadores de varianza (VCE, por sus siglas en inglés) robustos. Así, puede corregirse el cálculo de los errores estándares de los estimadores de los parámetros, permitiendo realizar inferencia estadística “válida”. Para ello, debemos utilizar la opción **vce(robust)** en el comando **regress**.

Por otra parte, los modelos estimados utilizando datos de corte transversal (*cross-section*) pueden presentar problemas de autocorrelación en los errores de observaciones que pertenecen a un mismo grupo o *cluster*. En consecuencia, los VCE por default y robustos son inválidos, afectando la inferencia. Este problema puede subsanarse utilizando la opción **vce(cluster clustervar)**.

# Regresión lineal múltiple: estimadores de varianza alternativos

Pongamos el foco en el uso de la opción `vce(robust)`:

```
. regress ln_totexp suppins phylim actlim totchr age female income
```

Source	SS	df	MS	Number of obs	=	2,955
Model	1264.72124	7	180.674463	F(7, 2947)	=	124.98
Residual	4260.16814	2,947	1.44559489	Prob > F	=	0.0000
				R-squared	=	0.2289
				Adj R-squared	=	0.2271
Total	5524.88938	2,954	1.87030785	Root MSE	=	1.2023

ln_totexp	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
suppins	.2556428	.0462264	5.53	0.000	.1650034 .3462821
phylim	.3020598	.0569709	5.30	0.000	.190353 .4137666
actlim	.3560054	.0621118	5.73	0.000	.2342185 .4777923
totchr	.3758201	.0184227	20.40	0.000	.3396974 .4119429
age	.0038016	.0036561	1.04	0.299	-.0033672 .0109705
female	-.0843275	.0455442	-1.85	0.064	-.1736292 .0049741
income	.0025498	.0010194	2.50	0.012	.000551 .0045486
_cons	6.703737	.27676	24.22	0.000	6.161075 7.2464

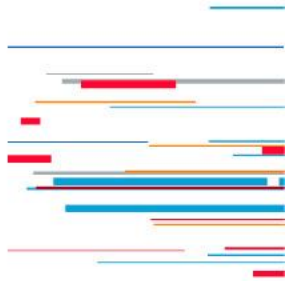
```
. regress ln_totexp suppins phylim actlim totchr age female income, vce(robust)
```

Linear regression

Number of obs	=	2,955
F(7, 2947)	=	126.97
Prob > F	=	0.0000
R-squared	=	0.2289
Root MSE	=	1.2023

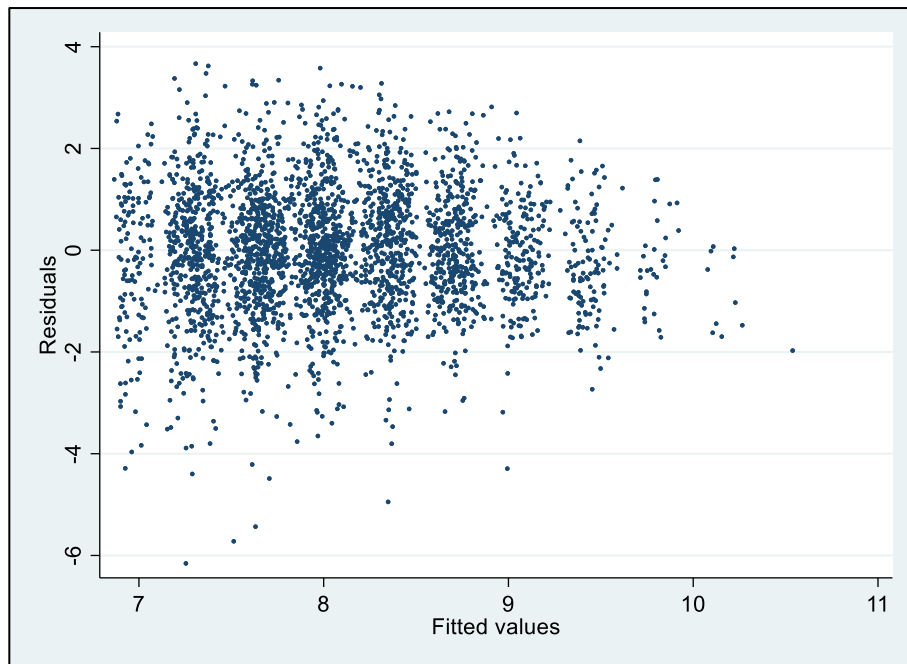
ln_totexp	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]
suppins	.2556428	.0465982	5.49	0.000	.1642744 .3470112
phylim	.3020598	.057705	5.23	0.000	.1889136 .415206
actlim	.3560054	.0634066	5.61	0.000	.2316797 .4803311
totchr	.3758201	.0187185	20.08	0.000	.3391175 .4125228
age	.0038016	.0037028	1.03	0.305	-.0034587 .011062
female	-.0843275	.045654	-1.85	0.065	-.1738444 .0051894
income	.0025498	.0010468	2.44	0.015	.0004973 .0046023
_cons	6.703737	.2825751	23.72	0.000	6.149673 7.257802



# Regresión lineal múltiple: inspección gráfica de outliers

Al igual que en el caso de la heterocedasticidad, para inspeccionar visualmente la presencia de *outliers* (valores atípicos) podemos utilizar un diagrama de dispersión de los residuos contra los valores ajustados de la variable dependiente. Por otra parte, con el comando **predict** y la opción **residual**, podemos crear una variable que contenga los residuos.

```
. quietly regress ln_totexp suppins phylim actlim totchr age female income, vce(robust)
```



```
. predict p_u, residual  
(109 missing values generated)  
  
. predict p_ln_totexp  
(option xb assumed; fitted values)  
  
. list totexp ln_totexp p_ln_totexp p_u if p_u < -5
```

	totexp	ln_tot~p	p_ln_t~p	p_u
1.	3	1.098612	7.254341	-6.155728
2.	6	1.791759	7.513358	-5.721598
3.	9	2.197225	7.631211	-5.433987

# Regresión lineal múltiple: colinealidad

La colinealidad implica la existencia de relaciones lineales entre las variables independientes del modelo. Salvo que alguna de estas relaciones sea exacta (que, en ese caso, impediría la estimación de los parámetros), no se rompe ningún supuesto clásico de la regresión lineal. No obstante, la presencia de cierto grado de correlación entre los regresores puede incrementar la varianza de los estimadores de los parámetros, haciéndolos más imprecisos. Un estadístico que, con sus limitaciones, permite evaluar “cuánto se infla” la varianza es el **factor inflacionario de la varianza** (VIF, por sus siglas en inglés):

```
. quietly regress ln_totexp suppins phylim actlim totchr age female income, vce(robust)
. vif
```

Variable	VIF	1/VIF
phylim	1.63	0.612872
actlim	1.62	0.618416
totchr	1.16	0.860297
age	1.11	0.900535
income	1.09	0.921201
suppins	1.06	0.947491
female	1.03	0.970835
Mean VIF	1.24	

# Regresión lineal múltiple: test sobre coeficientes

El comando **test** ejecuta tests de hipótesis en base al procedimiento de Wald, el cual hace uso de los coeficientes estimados y de los VCE.

Un ejercicio común es testear si los coeficientes de dos variables del modelo son estadísticamente diferentes entre sí. Para ello, una vez estimado el modelo, debemos escribir:

***test variable1 = variable2***

```
. quietly regress ln_totexp suppins phylim actlim totchr age female income, vce (robust)

. test phylim = actlim

( 1)  phylim - actlim = 0

      F( 1, 2947) =    0.27
      Prob > F =    0.6054
```

# Regresión lineal múltiple: test sobre coeficientes

Otro procedimiento común es testear la significancia conjunta de un subconjunto de regresores del modelo. Para ello, una vez estimado el modelo, debemos escribir:

*test varlist*

```
. test phylim actlim totchr

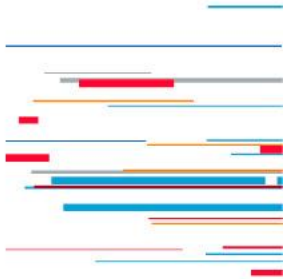
( 1)  phylim = 0
( 2)  actlim = 0
( 3)  totchr = 0

      F( 3, 2947) = 272.36
      Prob > F = 0.0000
```

# Regresión lineal múltiple: comando margins

El comando **margins** permite calcular diferentes tipos de respuesta de la variable dependiente (obtenidas a partir de las predicciones de un modelo previamente estimado) y evaluarlas en valores fijos de algunos regresores.

Por ejemplo, supongamos que tenemos un modelo en que ninguna de las variables ha sido transformada mediante la aplicación del logaritmo, pero queremos estimar las elasticidades o semielasticidades para diferentes regresores. Dado que éstas son función de las variables del modelo (es decir, no son constantes), es útil evaluarlas en ciertos valores de los regresores para obtener una medida resumen. Usualmente, se eligen las medias de los regresores para evaluar estas funciones. Veamos algunos ejemplos.





# Regresión lineal múltiple: comando margins

```
. quietly regress totexp suppins phylim actlim totchr age female income, vce (robust)

. margins, eyex (income) atmeans

Conditional marginal effects           Number of obs   =    3,064
Model VCE       : Robust

Expression   : Linear prediction, predict()
ey/ex w.r.t. : income
at           : suppins       =   .5812663 (mean)
              phylim        =   .4255875 (mean)
              actlim        =   .2836162 (mean)
              totchr        =   1.754243 (mean)
              age           =   74.17167 (mean)
              female        =   .5796345 (mean)
              income        =   22.47472 (mean)

-----
```

	Delta-method				
	ey/ex	Std. Err.	t	P> t	[95% Conf. Interval]
income	.0230892	.0267642	0.86	0.388	-.0293885 .0755669

```
. margins, eydx (income age) atmeans

Conditional marginal effects           Number of obs   =    3,064
Model VCE       : Robust

Expression   : Linear prediction, predict()
ey/dx w.r.t. : age income
at           : suppins       =   .5812663 (mean)
              phylim        =   .4255875 (mean)
              actlim        =   .2836162 (mean)
              totchr        =   1.754243 (mean)
              age           =   74.17167 (mean)
              female        =   .5796345 (mean)
              income        =   22.47472 (mean)

-----
```

	Delta-method				
	ey/dx	Std. Err.	t	P> t	[95% Conf. Interval]
age	-.0110046	.0051293	-2.15	0.032	-.0210619 -.0009473
income	.0010273	.0011909	0.86	0.388	-.0013076 .0033623

# Regresión lineal múltiple: efectos de interacción

En ocasiones, se suelen construir modelos que permitan que el efecto de un regresor sobre la variable dependiente dependa la magnitud de otra variable explicativa. Por ejemplo, podemos pensar que el efecto que tiene *female* sobre *ln\_totexp* es función lineal de *age*. Para permitir esto, debemos incorporar el producto entre *female* y *age* como regresor del modelo, para lo cual Stata tiene su propia sintaxis:

```
. regress ln_totexp suppins phylim actlim totchr c.age i.female c.age#i.female income, vce (robust)
```

In_totexp	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
suppins	.2535287	.0465601	5.45	0.000	.1622351	.3448224
phylim	.3038538	.0576991	5.27	0.000	.1907191	.4169885
actlim	.3628639	.0633256	5.73	0.000	.2386969	.4870308
totchr	.3754977	.0187006	20.08	0.000	.3388302	.4121652
age	.0160656	.0059273	2.71	0.007	.0044436	.0276876
1.female	1.415688	.5391991	2.63	0.009	.358443	2.472933
female#c.age						
1	-.0202516	.0072342	-2.80	0.005	-.0344361	-.006067
income	.0026963	.0010352	2.60	0.009	.0006666	.004726
_cons	5.795954	.4446481	13.03	0.000	4.924101	6.667806

Nótese que debemos utilizar los prefijos **c.** y **i.** para indicarle a Stata que la variable *age* es continua y que la variable *female* es discreta.



# Regresión lineal múltiple: exportación a Excel

Mediante el uso del comando **outreg2** podemos exportar los resultados de una regresión a diferentes tipos de archivo (documentos, hojas de cálculo, etc.) Dado que **outreg2** es un comando creado por usuarios, debemos instalarlo escribiendo:

**ssc install outreg2**

Evidentemente, antes de exportar los resultados de una regresión es necesario estimar el modelo. Además, es recomendable guardar temporalmente las estimaciones de un modelo, dado que esto nos será de utilidad cuando debamos trabajar con múltiples regresiones. Para ello, una vez estimado el modelo, debemos escribir:

**estimates store *name***

donde ***name*** es el nombre del modelo establecido por el usuario.

# Regresión lineal múltiple: exportación a Excel

Una posible sintaxis del comando **outreg2** es la siguiente:

**outreg2 [estlist] using filename [, options]**

donde **estlist** es una lista de modelos. Veamos un ejemplo:

```
. quietly regress totexp suppins phylim actlim totchr age female income, vce (robust)
. estimates store modelo_1
. outreg2 [modelo_1] using modelos_ols.xls, replace
modelos_ols.xls
dir : seeout
```

	A	B	C
1			
2		(1)	
3	VARIABLES	totexp	
4			
5	suppins	829.3**	
6		(411.5)	
7	phylim	2,427***	
8		(536.9)	
9	actlim	3,695***	
10		(687.6)	
11	totchr	1,940***	
12		(180.9)	
13	age	-77.37**	
14		(36.57)	
15	female	-1,257***	
16		(416.6)	
17	income	7.223	
18		(8.369)	
19	Constant	7,371***	
20		(2,734)	
21			
22	Observations	3,064	
23	R-squared	0.123	
24	Robust standard errors in parentheses		
25	*** p<0.01, ** p<0.05, * p<0.1		

# Regresión lineal múltiple: exportación a Excel

Si estimamos otros modelos y queremos guardarlo en la misma hoja de cálculo que el anterior, debemos hacer lo siguiente:

```
. gen ln_totexp = ln(totexp)
(109 missing values generated)

. gen ln_income = ln(income)
(88 missing values generated)

. quietly regress ln_totexp suppins phylim actlim totchr age female income, vce (robust)

. estimates store modelo_2

. quietly regress ln_totexp suppins phylim actlim totchr age female ln_income, vce (robust)

. estimates store modelo_3

. outreg2 [modelo_2 modelo_3] using modelos_ols.xls, append
modelos_ols.xls
dir : seeout
```

	A	B	C	D
1				
2		(1)	(2)	(3)
3		modelo_1	modelo_2	modelo_3
4	VARIABLES	totexp	ln_totexp	ln_totexp
5				
6	suppins	829.3**	0.256***	0.230***
7		(411.5)	(0.0466)	(0.0474)
8	phylim	2,427***	0.302***	0.299***
9		(536.9)	(0.0577)	(0.0583)
10	actlim	3,695***	0.356***	0.374***
11		(687.6)	(0.0634)	(0.0634)
12	totchr	1,940***	0.376***	0.370***
13		(180.9)	(0.0187)	(0.0188)
14	age	-77.37**	0.00380	0.00492
15		(36.57)	(0.00370)	(0.00373)
16	female	-1,257***	-0.0843*	-0.0701
17		(416.6)	(0.0457)	(0.0459)
18	income	7.223	0.00255**	
19		(8.369)	(0.00105)	
20	ln_income			0.0809***
21				(0.0273)
22	Constant	7,371***	6.704***	6.474***
23		(2,734)	(0.283)	(0.298)
24				
25	Observations	3,064	2,955	2,873
26	R-squared	0.123	0.229	0.231
27	Robust standard errors in parentheses			
28	*** p<0.01, ** p<0.05, * p<0.1			

# Bibliografía

- Cameron, A. C., & Trivedi, P. K. (2009). *Microeconometrics using Stata*. Texas: Stata Press.
- Rodríguez, G. (2019). *Stata Tutorial: Updated for Version 16*. Princeton University.
- StataCorp. (2019). *Getting Started with Stata for Windows: Release 16*. Texas: Stata Press.
- Wooldridge, J. M. (2010). *Introducción a la econometría. Un enfoque moderno*. México: Cengage Learning.



UNIVERSIDAD  
**AUSTRAL**

Valores que inspiran

