

Estimaciones y comandos Stata

Econometría básica

Rodrigo Taborda

Facultad de Economía

Universidad del Rosario

Primera versión 18 de Enero de 2011

Ésta versión 3 de marzo de 2011

1. Introducción

Este documento presenta conjuntamente resultados de estimación de regresión para el curso Econometría Básica y comandos de Stata para su obtención. Usa información obtenida mediante una encuesta hecha a los estudiantes de cursos previos de Econometría Básica. Información relacionada con esta encuesta, resultados y propósito se encuentra [acá](#).

Todos los comandos de Stata se presentan de manera que serán usados en un archivo .do o en el espacio de comandos de Stata. Las opciones presentadas en este documento son de mi preferencia y no implican una solución única para uso de Stata y presentación de resultados econométricos. Usted puede encontrar opciones que son perfectamente válidas. Mi selección de opciones dependen exclusivamente de dos cosas: primero, haber intentado otras posibilidades y llegar a la conclusión que la opción presentada es la menos complicada para mi, y, segundo, mi pereza de no buscar nuevas opciones potencialmente mejores.

Un objetivo subyacente de usar Stata de la manera presentada en este documento es no perder el tiempo en tareas que el computador puede hacer por usted. Objetivos secundarios son mantener un control y registro de lo que usted hace y asegurar la posibilidad de reproducir los resultados de una manera rápida y confiable.

El documento ofrece los comandos de Stata en recuadro con fondo gris. Usted debe construir un archivo (.do) incluyendo todos los comandos, y luego ejecutarlo en Stata. Los comandos y opciones presentados no son exhaustivos y usted debe recurrir a las opciones de ayuda de Stata, así como otros manuales y libros para ampliar su conocimiento y uso del programa.

2. Preliminares de Stata

2.1. Preliminar del preliminar

Stata considera una línea como un comando. Esto funciona bien para comandos cortos, en el caso de comandos largos es una fuente de confusión. Este problema se soluciona indicándole a Stata cuando termina un comando, sin importar si tiene varias líneas.

```
#delimit ;
```

Este comando le dice a Stata que un interprete como un comando todo lo que haya hasta que encuentre un punto y coma.

2.2. Cargar datos de fuente externa

Mi recomendación para llevar datos a Stata es que usted tenga los datos en el formato menos complicado posible. Para mi se trata de datos guardados separados por coma (.csv) o por tabulaciones (.tab). Asegurese que sus datos numéricos no tienen caracteres alfabéticos, de lo contrario Stata tomará su variable como texto y no un número. Comprobar el formato de archivo y su contenido debe hacerse antes de llevar los datos a Stata, preferiblemente.

```
insheet using "D:\el\lugar\donde\tenga\su\archivo.csv", comma clear;
```

Este comando le dice a Stata que busque su archivo en la ubicación indicada en el computador. Las comillas no son necesarias si sus carpetas ya archivos no tienen caracteres raros o espacios, de lo contrario tiene que usar comillas. Las opciones del comando van despues de la coma: se trata de un archivo en formato .csv, por eso la opcion (comma), y se pide que limpie los datos que existan en memoria para ser reemplazados por los que se van a cargar, por eso la opción (clear).

2.3. Guardar datos

Una vez los datos estén en Stata, los pude guardar en formato Stata para uso posterior. El archivo quedará guardado con extensión (.dta), no es necesario que lo especifique, recomiendo que lo haga para evitar confusiones. El archivo será compatible para Stata en la versión que

esté siendo usada y posteriores. Si necesita que su archivo sea legible para versiones previas (8 y 9) use el comando (saveold) en lugar de (save).¹

```
save "D:\el\lugar\donde\tenga\su\archivo.dta", replace;
```

2.4. Cargar datos de Stata

Una vez usted tiene sus datos básicos en formato Stata los puede usar repetidamente para su trabajo. El comando para invocar datos que ya existen en formato Stata es:

```
use "D:\el\lugar\donde\tenga\su\archivo.dta", clear;
```

El comando tiene solo la opción (clear), que pide limpiar cualquier base de datos existente en Stata y usar la solicitada.

3. Organícese

Organice sus datos de manera que sepa que esta haciendo y no pierda tiempo despues intentando averiguarlo.

3.1. Asigne etiquetas a las variables

Asigne etiquetas a las variables de manera que no dependa del nombre o código de la variable, sino de un nombre que tenga sentido y pueda usar automaticamente en el análisis gráfico o de regresión.

Fácilmente las variables que continene la información tienen un nombre como: var1, x1, y1, pregunta1. Además Stata no permite que su variable tenga nombre extenso con caracteres particulares, como: nombre del nieto, tiene niños, código. La posibilidad de asignar etiquetas a las variables le permitirá asignar a una variable un alias, usted seguirá usando el nombre en los comandos, pero el alias aparecerá en los resultados.

¹Si todavía usa Stata 8 o 9, actualícese.

```
label var id120 "Edad";  
label var id130 "Estatura (cm)";  
label var id140 "Peso (Kg)";
```

El comando indica que se asigne una etiqueta a una variable (`label var`), el nombre de la variable es (`id120`) y entre comillas se asigna el alias (`Edad`).

3.2. Comente, comente, comente

Agregue un comentario a todo lo que haga, para que después no se le olvide lo que hizo. Además así puede compartir su trabajo en el archivo (`.do`), y quien lo lea debería entender. Si no lo entiende, sus comentarios son malos. Comentarios en Stata comienzan con un asterisco (*). Otra opción es encerrar el comentario entre los siguientes símbolos (`/* */`).

```
*Etiqueta para la variable id120 y la  
*voy a nombrar Edad;  
    label var id120 "Edad";  
/*Etiqueta para la variable id130 y la voy a nombrar Estatura (cm)*/;  
    label var id130 "Estatura (cm)";
```

4. Inspeccione sus datos

Haga una inspección de los datos de manera visual y cuantitativa, explorando estadísticas descriptivas como: Media, Mínimo, Máximo, Desviación estándar.

4.1. Gráfica

Una manera rápida de inspeccionar los datos de corte transversal es haciendo un histograma. Si tiene datos de series de tiempo, una gráfica de la serie contra el tiempo es apropiada.

```

hist id120, width(1) percent
    title("Distribucion edad")
    subtitle("Años")
    scheme(s1mono)
;
graph export "D:\el\lugar\donde\tenga\su\grafica\hist_edad", as(eps) replace;

```

El comando (`hist`) pide hacer un histograma sobre la variable (`id120`). La opción (`percent`) pide que grafique los porcentajes. La opción (`width(1)`) pide que el ancho de las barras del histograma sea igual a 1, especifiqué esta opción porque los datos de edad se encuentran por años completos y facilita la visualización de la información. La opción (`scheme(s1mono)`) pide a Stata que haga la gráfica con un esquema de colores monocromático. Las opciones (`title`) y (`subtitle`) asignan título y subtítulo.

El comando (`graph export`) pide a Stata que exporte la gráfica a formatos disponibles para su procesador de texto, una opción apropiada para Word puede ser (`.emf`) Windows Enhanced Metafile o (`.wmf`) Windows Metafile, si usa Latex como editor de texto, una opción es (`.eps`) encapsulated PostScript.

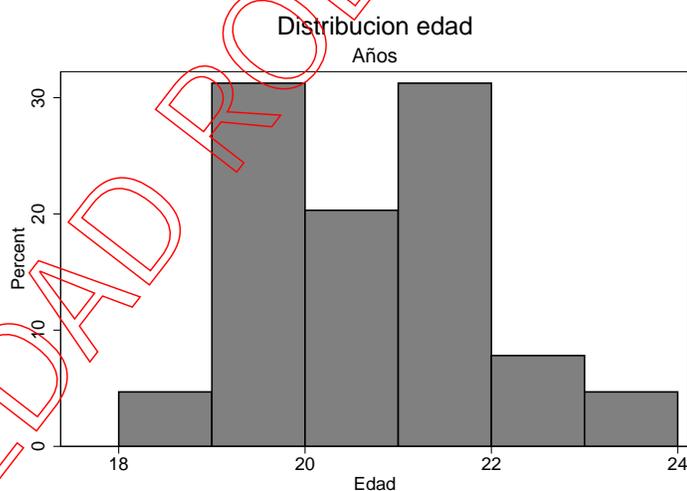


Figura 1. Distribución de la edad de los estudiantes.

Note cómo todas las opciones definidas en el comando de gráfica aparecen correctamente, incluyendo la etiqueta de la variable, en el eje horizontal inferior, (`Edad`).

4.2. Tabla estadísticas descriptivas

Una tabla con estadísticas descriptivas siempre es útil.

Cuadro 1. Estadísticas descriptivas edad y estatura

	Media	Desv. Est.	Min	Max
Edad	20.2	1.3	18.0	24.0
Estatura	166.3	8.8	150.0	185.0

```
sum id120 id130;
```

El comando (`sum`) genera las estadísticas descriptivas de las variables (`id120`) y (`id130`). Si agrega la opción (`detail`) puede obtener muchas otras estadísticas descriptivas de la variable.

5. Regresión

A partir de una conjetura basada en la teoría o lógica económica, se lleva a cabo una regresión. En este ejemplo se sugiere que el desempeño de los estudiantes en la universidad depende de su puntaje ICFES, ecuación 1.

$$\text{Notas en la universidad} = f(\text{ICFES}) \quad (1)$$

La ecuación a estimar es la ecuación 2.

$$\text{Notas en la universidad}_i = \beta_0 + \beta_1 \text{ICFES}_i + u_i \quad (2)$$

El comando para regresión lineal simple entre el promedio del primer semestre (variable `aca110`) y el puntaje ICFES (variable `aca105`) es:

```
reg aca110 aca105;
```

El comando toma como variable dependiente la primer variable de la lista, y como independiente o explicativa todas las que sigan. En este caso no tiene opciones, pero existen múltiples posibilidades que se exploran más adelante.

Los resultados se presentan en la tabla 2. La lectura de la tabla se hace por columnas. Cada columna presenta los resultados de una regresión. La segunda columna presenta la regresión cuya variable dependiente es la nota promedio del primer semestre, y cuya variable explicativa es el puntaje del ICFES. Las columnas siguientes usan como variable dependiente la nota del promedio del último semestre antes de responder la encuesta, y las notas de los cursos de estadística, matemáticas, el último curso de macroeconomía, y el último curso de microeconomía.

Esta manera de reportar resultados de regresión es estandard en economía. Note que en la parte inferior de la tabla se encuentra el número de observaciones usadas en la regresión, el estadístico R^2 y el F . En la parte superior de la tabla se encuentran los coeficientes correspondientes a la variable explicativa, ICFES en este caso y la constante. Los debajo de cada coeficiente se encuentra entre paréntesis el error estandard, y como superíndice de cada coeficiente una letra a , b o c que indica el nivel de significancia del coeficiente. a para menor a 1%, b para menor a 5% y c para menor a 10%.²

²Una nomenclatura alternativa es entre uno y 3 asteriscos para los mismos niveles de significancia.

Cuadro 2. Resultados de regresión, ICFES explica desempeño académico

VARIABLES	(1) 1er sem.	(2) Ultimo sem.	(3) Estadis.	(4) Matemat.	(5) Macro.	(6) Micro.	(7) Econometria.
ICFES	0.0157 ^b (0.00737)	0.00847 (0.00611)	0.00187 (0.0102)	0.00961 (0.0101)	0.00383 (0.0104)	0.0162 (0.00989)	-0.00287 (0.0130)
Constant	2.995 ^a (0.446)	3.343 ^a (0.370)	3.829 ^a (0.615)	3.286 ^a (0.613)	3.371 ^a (0.626)	2.708 ^a (0.598)	4.015 ^a (0.784)
Observations	64	64	64	64	64	64	59
R ²	0.068	0.030	0.001	0.014	0.002	0.042	0.001
F	4.518	1.921	0.0340	0.900	0.137	2.701	0.0490

Errores estandard en parentesis. a p<0.01. b p<0.05. c p<0.1.

Una opción de análisis ante una regresión con una variable explicativa es hacer un gráfico de dispersión de datos que incluya la ecuación estimada de regresión. Se presentan dos graficas correspondientes a aquellas donde las variables dependientes son promedio primer semestre y promedio último semestre.

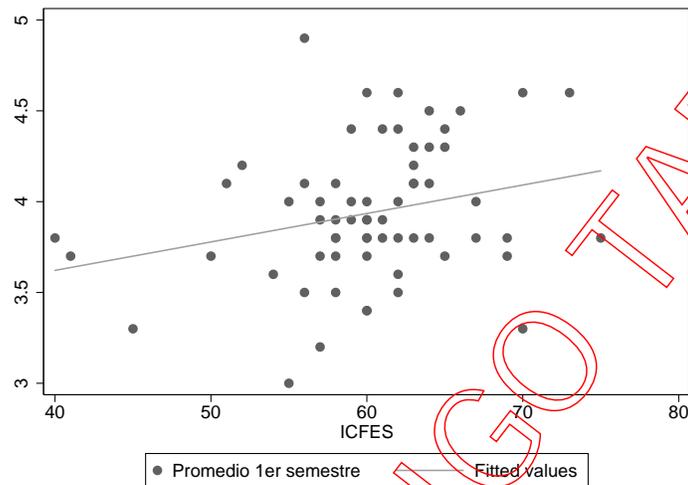


Figura 2. Dispersión de datos y ajuste de regresión.

Nota: Gráfica correspondiente a regresión donde variable dependiente es el promedio del primer semestre y variable explicativa el puntaje ICFES.

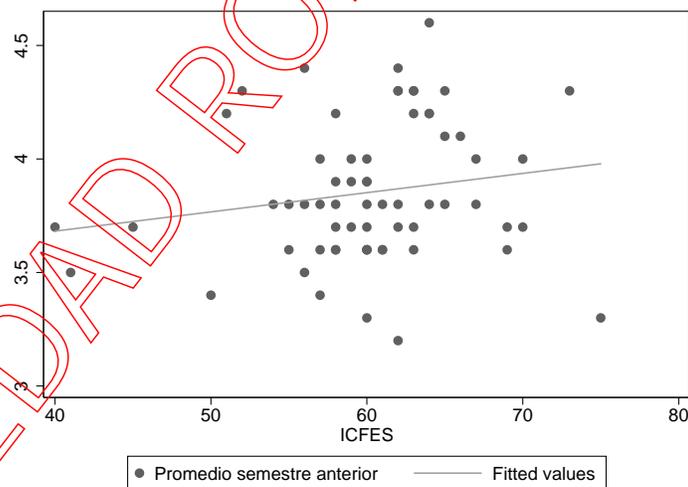


Figura 3. Dispersión de datos y ajuste de regresión.

Nota: Gráfica correspondiente a regresión donde variable dependiente es el promedio del último semestre y variable explicativa el puntaje ICFES.

El comando para obtener esta gráfica entre el promedio del primer semestre (variable aca110) y el puntaje ICFES (variable aca105) es:

```

twoway (scatter aca110 aca105) (lfit aca110 aca105),
    title("")
    subtitle("")
    scheme(s1mono)
    ;
graph export "D:\el\lugar\donde\tenga\su\grafica\hist_edad", as(eps) replace;

```

El comando (`twoway`) indica que vamos a hacer una gráfica que incluye información en dos ejes. La capacidad de este comando es que puede superponer dos tipos de gráficos con la misma característica de tener dos ejes. De modo que se superponen la gráfica de dispersión de datos (`scatter`), y la gráfica de ajuste lineal de regresión (`lfit`). El comando tiene las mismas opciones que ya fueron discutidas anteriormente. Pero note que las opciones están asignadas al comando (`twoway`) y no a los comandos intermedios (`scatter`) y (`lfit`), de ser necesario podrían agregarse opciones a estos sub-comandos.

5.1. Interpretación de resultados

Es difícil distinguir entre las gráficas 9 y 3. Sin embargo los resultados de regresión reportados en la tabla 2 sugieren que estadísticamente la regresión entre nota promedio del primer semestre e ICFES es estadísticamente significativa a niveles altos, mientras que pasa lo contrario con la regresión de nota del último semestre. Una posible conclusión de este ejercicio es que el examen ICFES solo explica el desempeño del estudiante en el primer semestre, pero no en los siguientes.

5.2. Condicione los resultados de su regresión

En un gran número de comandos de Stata se pueden condicionar los resultados a una submuestra de datos. Continuando con el ejemplo anterior es posible obtener los resultados de regresión discriminando por género. El comando es el siguiente:

```

reg aca110 aca105 if id110num == 0;
reg aca110 aca105 if id110num == 1;

```

Cuadro 3. Resultados de regresión, ICFES explica desempeño académico primer semestre. Hombres y Mujeres.

VARIABLES	(1) 1er sem. Hombres	(2) 1er sem. Mujeres
ICFES	0.0264 ^c (0.0130)	0.0133 (0.00952)
Constant	2.304 ^a (0.807)	3.151 ^a (0.566)
Observations	24	40
R-squared	0.159	0.049
F	4.146	1.942

Errores estandard en parentesis. a p<0.01. b p<0.05. c p<0.1.

El comando es similar al de regresión pero agrega un condicional (if) que solicita hacer la regresión solo para las observaciones que cumplen con la condición según la variable (id110num). La variable (id110num) es dicotómica y contiene el género 0 (hombres) y 1 (mujeres). Note que para escribir el condicional se incluyen dos signos igual (==).

Los resultados de la regresión se presentan en la tabla 3.

6. Manipulación de variables

Stata ofrece dos opciones para generar nuevas variables a partir de una base de datos existente. El comando (gen) genera nuevas variables usando muchas funciones matemáticas. El comando (egen) es denominado extensiones al comando (gen) y ofrece posibilidades particulares que ahorran mucho tiempo y esfuerzo.

```
gen variable_inservible = aca105 * aca110;
egen promedio_estadistica = mean(aca121);
by acano100: egen promedio_estadistica_transporte = mean(aca121);
```

El primer comando genera la variable (variable_inservible) que es igual al producto de la variable que contiene el puntaje ICFES y el promedio del primer semestre. El segundo comando genera una nueva variable que contiene el promedio de la nota del curso de estadística (aca121) de todos los estudiantes. El tercer comando genera una nueva variable que contiene el promedio de la nota del curso de estadística (aca121) de todos los estudiantes según la variable (acano100). (acano100) es una variable categórica que contiene el medio de transporte

de los estudiantes. 1 (Automovil propio), 2 (Automovil no propio), 3 (Transmilenio), 4 (Bus), 5 (Taxi), 6 (Camina). De manera que al solicitar el promedio del curso de estadística, se genera una nueva variable que ya no es la misma para todos los estudiantes, sino que cambia para los grupos definidos por el medio de transporte.

7. Inferencia

7.1. Los estimadores tienen propiedades muestrales

y son estimadores muestrales, si tuviésemos una muestra diferente obtendríamos un valor diferente, si tuviésemos mil muestras diferentes obtendríamos mil estimadores diferentes. Es por ello que a pesar de que la estimación de regresión lineal ofrece un solo estimador, se pueden definir propiedades muestrales sobre el estimador. Esta sección comienza mostrando la obtención de muchos estimadores y su distribución.

En primer lugar se van a generar histogramas para el coeficiente y constante de una regresión a partir de la estimación de 10, 50, 100 y 500 regresiones de

$$\text{Notas en la universidad} = f(\text{ICFES})$$

Note como al incrementar el número de repeticiones el promedio de los estimadores de cada una de las repeticiones de la estimación con el 75 % de los datos iniciales se acerca cada vez más al estimador de mínimos cuadrados ordinarios presentado en el cuadro 2.

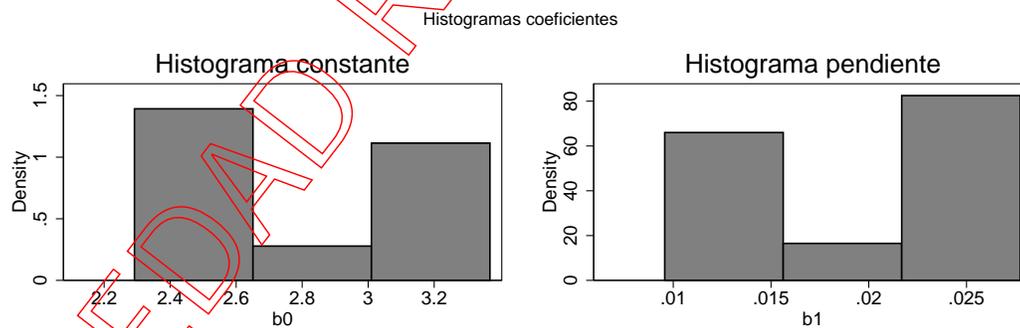


Figura 4. Histograma de coeficientes constante y pendiente de la regresión.

Nota: Gráfica muestra el histograma de 10 repeticiones de la regresión.

Los comandos para desarrollar estos resultados se presentan a continuación, solo se presenta la simulación para 500 repeticiones. Estos comandos tienen un poco más de dificultad respecto a los presentados anteriormente.

Cuadro 4. Estadísticas de coeficientes constante y pendiente de 10 repeticiones de la regresión

Variable	Mean	Std. Dev.	Min.	Max.
b0	2.808	0.437	2.292	3.369
b1	0.019	0.007	0.01	0.028
n	48	0	48	48
N		10		

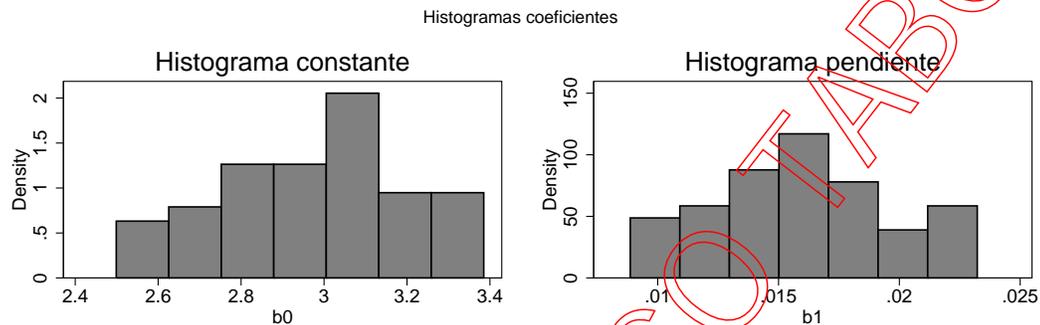


Figura 5. Histograma de coeficientes constante y pendiente de la regresión.

Nota: Gráfica muestra el histograma de 50 repeticiones de la regresión.

Cuadro 5. Estadísticas de coeficientes constante y pendiente de 50 repeticiones de la regresión

Variable	Mean	Std. Dev.	Min.	Max.
b0	2.985	0.224	2.499	3.385
b1	0.016	0.004	0.009	0.023
n	48	0	48	48
N		50		

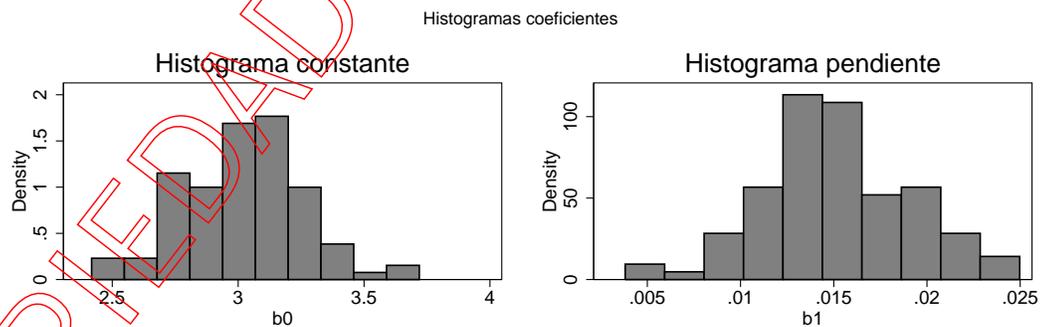


Figura 6. Histograma de coeficientes constante y pendiente de la regresión.

Nota: Gráfica muestra el histograma de 100 repeticiones de la regresión.

Cuadro 6. Estadísticas de coeficientes constante y pendiente de 100 repeticiones de la regresión

Variable	Mean	Std. Dev.	Min.	Max.
b0	3.025	0.244	2.418	3.719
b1	0.015	0.004	0.004	0.025
n	48	0	48	48
N	100			

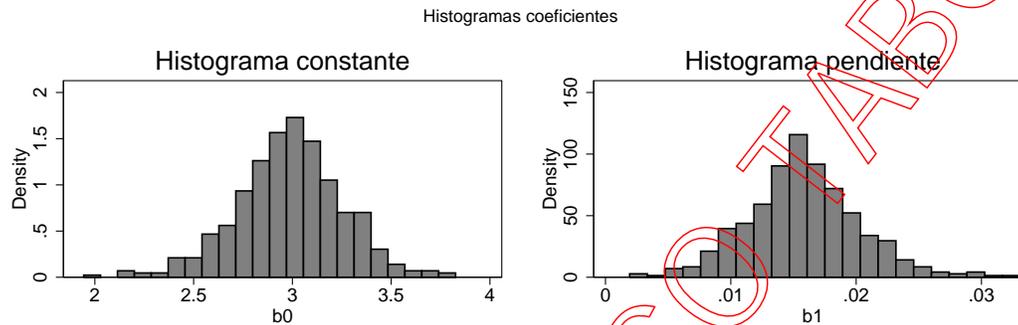


Figura 7. Histograma de coeficientes constante y pendiente de la regresión.

Nota: Gráfica muestra el histograma de 500 repeticiones de la regresión.

Cuadro 7. Estadísticas de coeficientes constante y pendiente de 500 repeticiones de la regresión

Variable	Mean	Std. Dev.	Min.	Max.
b0	2.987	0.27	1.943	3.826
b1	0.016	0.005	0.002	0.033
n	48	0	48	48
N	500			

```

/*COMANDOS PARA SIMULACION. 500 REPETICIONES*/
**COMANDO PARA ABRIR PROCESO DE RECOLECCION DE DATOS DE LAS 500 REGRESIONES;
postfile postfile500rep ronda b0 b1 n using
"D:/el/lugar/donde/tenga/su/archivo/reg_icfes_500rep", replace;
**COMANDO PARA REPETIR PROCEDIMIENTOS;
foreach ronda of numlist 1/500 {;
**COMANDO PARA OBTENER UNA MUESTRA DE LOS DATOS EXISTENTES
Y HACER LOS MISMOS CALCULOS ANTERIORES;
**COMANDO PARA ABRIR BASE DE DATOS;
use D:/el/lugar/donde/tenga/su/archivo/results-survey, clear;

```

```

/*CON EL 75% DE LOS DATOS*/
sample 75;
**COMANDO PARA EJECUTAR LA REGRESION;
/*EJECUTA LA REGRESION*/;
quietly regress aca110 aca105;
**COMANDO PARA GUARDAR EL COEFICIENTE DE LA REGRESION;
/*SE GUARDA LA RONDA, EL COEFICIENTE DE LA CONSTANTE Y LA PENDIENTE,*/
EL NUMERO DE OBSERVACIONES*/
post postfile500rep () (_b[_cons]) (_b[aca105]) (_N);
};
**COMANDO PARA CERRAR PROCESO DE RECOLECCION DE DATOS DE LAS 500 REGRESIONES;
postclose postfile500rep;
**ANALISIS DE LOS COEFICIENTES DE PENDIENTE Y CONSTANTE;
**CARGAR DATOS;
use D:/el/lugar/donde/tenga/su/archivo/reg_icfes_500rep, clear;
**ESTADISTICAS DESCRIPTIVAS;
summarize b0 b1 n;
/*Exportar tabla de estadísticas a latex*/
sutex b0 b1 n,
    file(D:/el/lugar/donde/tenga/su/archivo/reg_icfes_500rep_stats.tex)
minmax
title("Estadísticas de coeficientes constante y pendiente de 500
repeticiones de la regresión")
placement(h!)
key(fig:reg_icfes_500rep)
replace;
**HISTOGRAMA COEFICIENTES;
histogram b0,
    title("Histograma constante")
scheme(simono)
saving(D:/el/lugar/donde/tenga/su/archivo/reg_icfes_500rep_constante,
replace);
histogram b1,
    title("Histograma pendiente")

```

```

scheme(simono)
saving(D:/el/lugar/donde/tenga/su/archivo/reg_icfes_500rep_pendiente,
replace);
graph combine
D:/el/lugar/donde/tenga/su/archivo/reg_icfes_500rep_constante.gph
D:/el/lugar/donde/tenga/su/archivo/reg_icfes_500rep_pendiente.gph,
iscale(1.5)
title("Histogramas coeficientes")
xsize(12)
scheme(simono);
graph export D:/el/lugar/donde/tenga/su/archivo/reg_icfes_500rep,
as(eps) replace;

```

7.2. Intervalos de confianza

Conociendo las propiedades muestrales de un estimador, es posible definir los intervalos de confianza y conjuntamente con el supuesto de normalidad sobre los errores, calificar su significancia y probabilidad de que el coeficiente sea estadísticamente significativo o cero.

Cuadro 8. Resultados de regresión, ICFES explica desempeño académico

VARIABLES	(1) 1er sem.	(2) Ultimo sem.
ICFES	0.0157 ^b (0.00737) 2.126 (0.0375) 0.000933 - 0.0304	0.00847 (0.00611) 1.386 (0.171) -0.00375 - 0.0207
Constant	2.995 ^a (0.446) 6.717 (6.62e-09) 2.104 - 3.886	3.343 ^a (0.370) 9.038 (0) 2.604 - 4.083
Observations	64	64
R-squared	0.068	0.030
F	4.518	1.921
Errores estandard en parentesis. a p<0.01. b p<0.05. c p<0.1.		

En la tabla 8 se presentan los mismos resultados de regresión de las dos primeras regresiones de la tabla 2 pero bajo el coeficiente se presenta el error estándar, estadístico t, valor P y el intervalo de confianza. En las siguientes gráficas se presenta el intervalo de confianza.

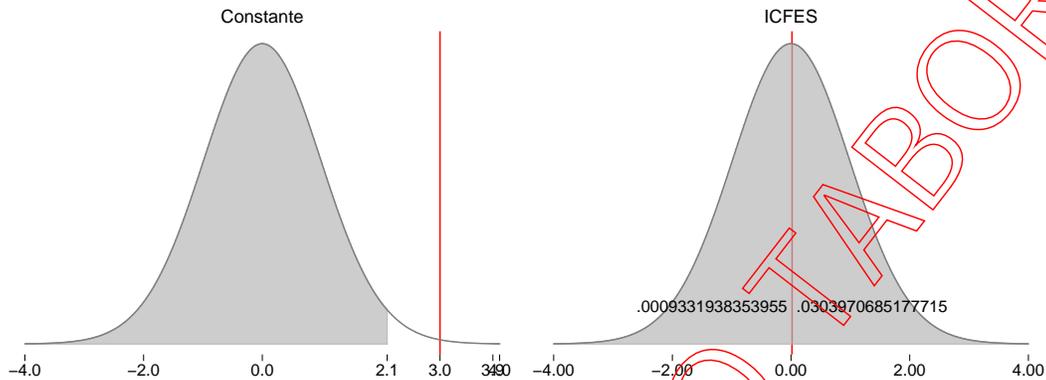


Figura 8. Intervalos de confianza para reg 1 tabla 8.
Nota: Gráfica muestra coeficiente e intervalo de confianza.

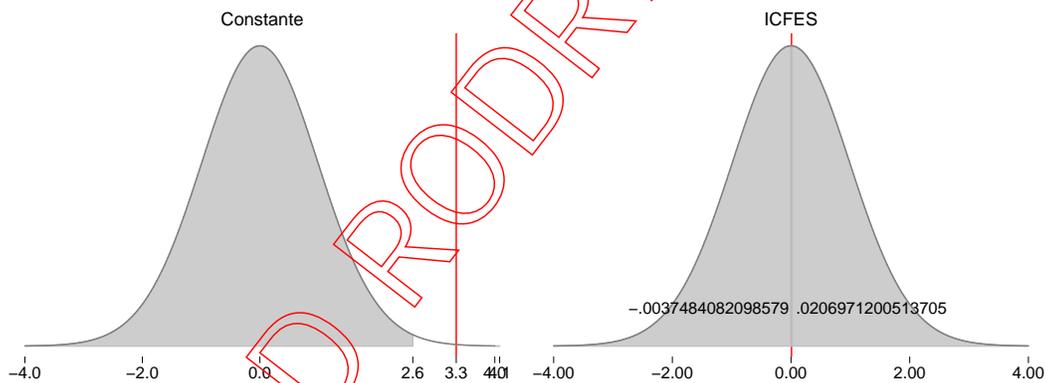


Figura 9. Intervalos de confianza para reg 2 tabla 8.
Nota: Gráfica muestra coeficiente e intervalo de confianza.

8. Variable dependiente limitada

Los modelos de variable dependiente limitada son fáciles de estimar en Stata. Una ventaja particular que tiene Stata es el cálculo de la magnitud de los efectos y graficas asociados a los mismos. En particular existe el grupo de comandos diseñados por Scott Long y Jeremy Freese llamado Spost que facilitan su obtención. Para la utilización de estos comandos usted debe haber instalado el paquete de programas incluidos en Spost en Stata.

Con los mismos datos usados en los resultados anteriores, se estimará un modelo de variable dependiente limitada. Se construye una variable binaria a partir del promedio del primer semestre. La variable toma un valor de 1 si el promedio del primer semestre es superior a 4 y 0 si es inferior.³

```
gen aca110_binaria = 0;
    label var aca110_binaria "Promedio 1er semestre - binaria";
replace aca110_binaria = 1 if aca110 > 4;
```

En el recuadro anterior se presenta una manera de definir una nueva variable con las características definidas anteriormente. Esta nueva variable es `aca110_binaria`. Con esta variable se va a estimar un modelo de variable dependiente limitada a partir de los resultados del examen ICFES.

```
probit aca110_binaria aca105;
probit aca110_binaria aca105 id110num;
logit aca110_binaria aca105;
logit aca110_binaria aca105 id110num;
```

El comando de estimación de modelos probit y logit se ejecuta de manera similar al comando de regresión. La primer variable después del comando `probit` o `logit` corresponde a la variable dependiente binaria. En este comando también se puede condicionar la muestra con el operador `if`.

Los mismos resultados de estimación presentados en la tabla 9 con la estimación de efecto marginal se presentan en la tabla 10.

Los comandos para obtener el rango de probabilidad para todos los valores observados de la variable ICFES son los siguientes.

```
qui probit aca110_binaria aca105;
prgen aca105, gen(icfes) rest(mean) ci;
```

³Ningún estudiante reportó promedio inferior a 3.

Cuadro 9. Resultados de regresión variable dependiente limitada, ICFES explica desempeño académico

VARIABLES	LABELS	(1) Probit 1	(2) Probit 2	(3) Logit 1	(4) Logit 2
aca105	ICFES	0.0434 (0.0284)	0.0464 (0.0290)	0.0706 (0.0477)	0.0750 (0.0485)
id110num	Genero		0.173 (0.347)		0.270 (0.566)
Constant	Constant	-3.031 ^c (1.736)	-3.323 ^c (1.832)	-4.929 ^c (2.927)	-5.359 ^c (3.058)
Observations		64	64	64	64

Errores estandar en parentesis. a $p < 0.01$. b $p < 0.05$. c $p < 0.1$.

Cuadro 10. Efecto marginal de regresión variable dependiente limitada, ICFES explica desempeño académico

VARIABLES	(1) Probit 1	(2) Probit 2	(3) Logit 1	(4) Logit 2
ICFES	0.0159 (0.0103)	0.0169 (0.0105)	0.0158 (0.0105)	0.0167 (0.0107)
Genero		0.0625 (0.124)		0.0595 (0.123)
Observations	64	64	64	64

Errores estandar en parentesis. a $p < 0.01$. b $p < 0.05$. c $p < 0.1$.

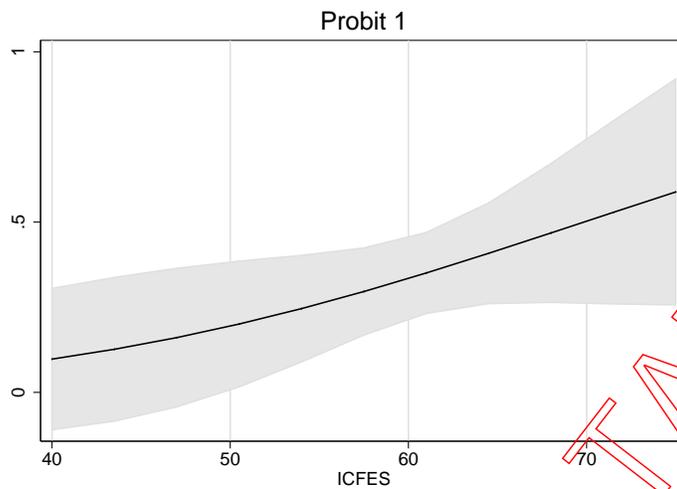


Figura 10. Probabilidad de obtener nota superior a 4 para diferentes valores del ICFES.

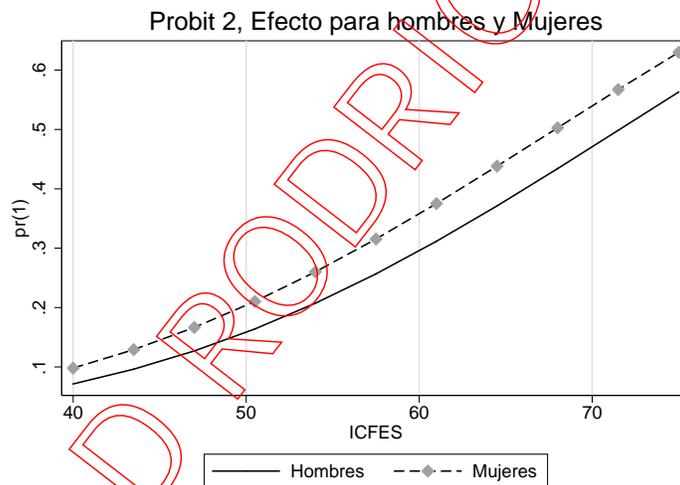


Figura 11. Probabilidad de obtener nota superior a 4 para diferentes valores del ICFES.

```
graph twoway (rarea icfesplub icfespllb icfesx, color(gs14))
  (connected icfespl icfesx,
    lpattern(solid) msize(zero) color(black))
  ,xlabel(40(10)75, grid) scheme(simono) legend(off)
  title("Probit 1");
```

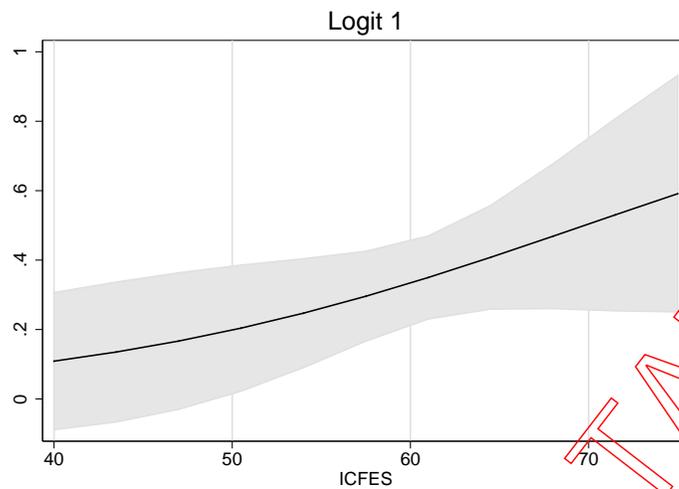


Figura 12. Probabilidad de obtener nota superior a 4 para diferentes valores del ICFES.

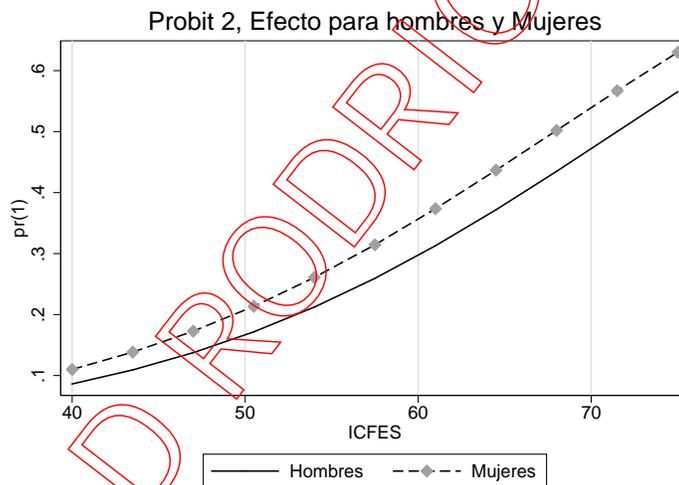


Figura 13. Probabilidad de obtener nota superior a 4 para diferentes valores del ICFES.

En primer lugar debe estimarse el modelo probit o logit. Posteriormente se usa el comando `prgen` que hace parte de los comandos del paquete `spost` de Long and Freese (2006) para obtener la probabilidad asociada a los valores de una de las variables explicativas. En este caso se trata de la variable ICFES. El resto de variables explicativas se definen en la media (`rest(mean)`) y se solicita que se generen variables con los intervalos de confianza (`ci`). Finalmente se grafican las variables. El comando `prgen` genera variables que contienen la información necesarias con los sufijos `p0` y `p1`, correspondientes a que no ocurra u ocurra el evento.

ub y lb para límite superior e inferior.

Existen comandos adicionales asociados a la estimación de modelos con variable dependiente limitada. Es posible obtener el efecto marginal con el comando `mfx`, siempre se debe ejecutar este comando inmediatamente después de la estimación de este tipo de modelos.

```
probit aca110_binaria aca105;  
mfx, at(icfes=60)
```

El comando anterior genera la probabilidad del evento codificado como 1 para la estimación previa del modelo probit o logit evaluada cuando el puntaje del examen icfes es igual a 60.

Recuerde cómo para obtener el efecto marginal se evalúa la densidad de la distribución usada en la estimación en el valor deseado (60 en el ejemplo anterior) y se multiplica por el coeficiente obtenido en la estimación probit o logit. Este mismo resultado puede calcularse con los siguientes comandos.

```
probit aca110binaria aca105;  
nlcom (normalden(_b[cons] + _b[aca105] * 60) * _b[dttime]);
```

9. Modelos de datos de panel

La estimación de modelos de datos de panel en Stata es sencilla. En primer lugar debe establecerse que los datos tienen estructura de panel. Debe existir una variable que identifique los individuos y una variable que identifique el tiempo. La base de datos que se usa como ejemplo (Baum, 2006) en este manual corresponde a información de 48 estados de Estados Unidos que contiene tasa de mortalidad en accidentes de tráfico en carreteras (`fatal`), impuesto a la cerveza (`beertax`), consumo de bebidas alcohólicas (`spircons`), tasa de desempleo (`unrate`) y el ingreso percapita (`perinck`). El comando para definir esta característica de los datos es (`xtset individuo fecha`).

```
use http://www.stata-press.com/data/imeus/traffic.dta, clear;
xtset state year;
```

La estimación de modelos con efectos fijos y aleatorios se reduce a incluir la opción en el comando (xtreg). Seguido de la estimación se guardan los resultados para la prueba posterior de especificación de Hausman.

```
xtreg fatal beertax spircons unrate perinck, fe;
estimates store fijos;

xtreg fatal beertax spircons unrate perinck, re;
estimates store aleatorios;

hausman fijos aleatorios;
```

Referencias

Baum, C. (2006). *An introduction to modern econometrics using Stata* (2nd ed.). Stata Press books. StataCorp LP.

Long, J. S. and J. Freese (2006). *Regression Models for Categorical Dependent Variables using Stata* (2nd ed.). Stata Press books. StataCorp LP.

PROPIEDAD RODRIGO TABORDA