

Tema 1

Gretl y la Econometría

Contenido

1.1. Introducción	2
1.2. ¿Qué es la Econometría?	2
1.2.1. ¿Para qué sirve la Econometría?	4
1.3. Un estudio econométrico	6
1.4. Los datos y su manejo	7
1.4.1. Fuentes de datos	9
1.4.2. El software econométrico	9
1.5. Introducción a Gretl	11
1.5.1. Análisis descriptivo de una variable	14
1.5.2. Relaciones entre variables	19

1.1. Introducción

Este curso se dirige a aquellas personas interesadas en aprender a interpretar información estadística sobre la realidad económica. La herramienta básica es un modelo econométrico que conjuga los esquemas teóricos sobre el funcionamiento de la Economía con las técnicas estadísticas de análisis de datos. Un modelo puede tener una estructura muy compleja, pero en este curso nos centramos en el modelo más sencillo, y que da nombre a la asignatura, el **modelo de regresión lineal general**. Este modelo explica el comportamiento de una única variable económica o de otra índole más general.

Por otro lado, este curso tiene un carácter totalmente aplicado, en el que los ejemplos prácticos sirven para introducir los conceptos estadístico-econométricos. Así, una parte importante del curso se dedica a estudiar casos prácticos, en los que el estudiante aprenderá a manejar un software econométrico y a interpretar adecuadamente los resultados obtenidos. El paquete econométrico a utilizar es Gretl; se trata de software de libre uso, fácil de manejar y que tiene acceso a las bases de datos que se estudian en muchos libros de introducción al análisis econométrico.

Este primer tema se organiza de la siguiente forma: la sección 2 presenta la disciplina que nos ocupa en este curso, la Econometría. La sección 3 describe un ejemplo de estudio econométrico, destacando cuáles son los elementos que integran un modelo econométrico. La sección 4 se ocupa de los datos económicos, sus características, las principales fuentes de obtención de datos y los programas informáticos que sirven para almacenar y procesar los datos. El software Gretl se introduce en el apartado 5, en el que se incluye el esquema de una primera sesión práctica de uso de Gretl. Los dos últimos apartados son un repaso a los conceptos de probabilidad e inferencia estadística que se aplicarán posteriormente, y que se acompaña de una sesión de práctica en Gretl.

1.2. ¿Qué es la Econometría?

En la toma de decisiones de carácter económico suele ser muy útil disponer de información en forma de datos cuantitativos. Por ejemplo, a la hora de elegir unos estudios universitarios podemos guiarnos por nuestras preferencias personales, pero también por factores como las expectativas de salario en la rama elegida o la facilidad con la que esperamos conseguir un empleo. Si se trata de la compra-venta de un piso, nos interesa conocer la situación del mercado inmobiliario. Para ello podemos recopilar datos de precios y de algunas características de los pisos que puedan influir en el precio como, por ejemplo, su tamaño o si es una vivienda usada que necesita reforma. Supongamos que en la sección de anuncios de un periódico local aparecen los siguientes datos sobre 50 pisos en venta en el centro de una ciudad:

- Precio del piso, en miles de euros.
- Tamaño del piso, en metros cuadrados hábiles.
- Estado del piso: si necesita reforma o está para entrar a vivir.

Indicador	Tamaño	Precio	A reformar	Indicador	Tamaño	Precio	A reformar
1	55	210,354	no	26	110	476,600	no
2	59	309,520	no	27	110	456,769	no
3	60	366,617	no	28	115	500,643	no
4	60	299,304	si	29	125	619,000	no
5	60	369,650	no	30	135	645,253	no
6	65	273,460	si	31	135	625,000	no
7	65	155,000	si	32	140	522,800	si
8	70	228,384	no	33	150	390,660	no
9	70	246,415	no	34	150	504,850	si
10	70	255,000	si	35	150	715,204	no
11	75	150,253	si	36	150	570,000	si
12	77	352,800	no	37	160	751,265	no
13	80	366,000	si	38	180	583,000	si
14	80	298,000	si	39	180	738,000	no
15	80	312,530	no	40	180	552,931	si
16	83	240,400	no	41	190	691,200	no
17	85	278,569	si	42	195	811,400	no
18	91	390,658	no	43	200	691,000	si
19	92	216,364	si	44	200	1110,000	no
20	100	402,600	no	45	230	961,620	no
21	100	272,300	si	46	230	661,000	no
22	100	360,607	no	47	240	841,417	no
23	100	570,000	no	48	240	588,992	si
24	100	480,809	no	49	245	841,400	si
25	100	186,314	si	50	250	1051,000	no

Tabla 1.1: Datos sobre precio de vivienda ocupada

Estos datos aparecen en la Tabla 1.1. En base a esta información, si nos ofrecen un piso de $100 m^2$ reformado a un precio de $525000€$, diríamos que el piso parece caro ya que su precio supera el promedio de precios de los pisos de estas características incluidos en la muestra:

$$\frac{402,6 + 360,607 + 570 + 480,809}{4} = 453,504 \text{ miles de euros}$$

Sin embargo, ¿qué podemos decir si se tratara de un piso de $90 m^2$ a reformar? ¿O de un piso de $50 m^2$ reformado? No tenemos datos para replicar el procedimiento anterior. Un econométra podría ayudar a dar respuesta a estas cuestiones. En el Gráfico 1.1, que representa conjuntamente el precio y el tamaño de cada piso, se ve un patrón o *relación estable* entre tamaño de un piso y su precio. Esta relación se puede trasladar a un *modelo* útil para responder a las preguntas que planteamos. Las técnicas econométricas nos permiten cuantificar, a partir del modelo y los datos, la influencia que tiene el tamaño del piso o su estado en el precio del mismo. La respuesta podría ser, por ejemplo: *La estimación del precio medio de un piso a reformar de $90 m^2$ es de 297350 euros, aunque el precio puede oscilar entre 152711 y 441989 euros a un nivel de confianza del 90%. Además, si se trata de un piso reformado, la estimación del precio medio se incrementa en más de 100000 euros, siendo factibles precios entre 210521 y 556639 euros.*

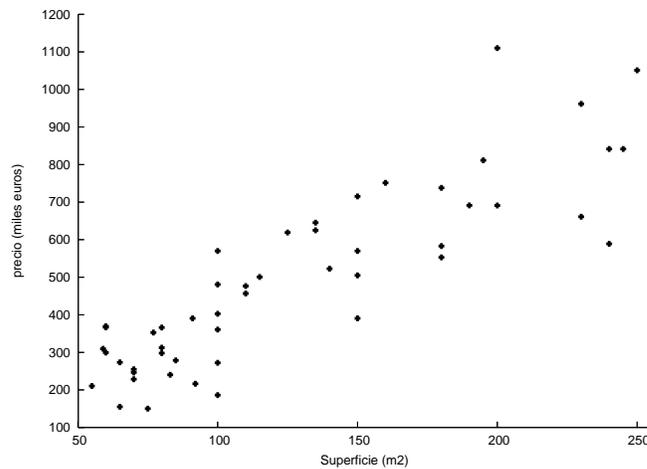


Gráfico 1.1: Diagrama de dispersión superficie-precio de pisos

La Econometría es una rama de la Economía que utiliza la estadística para medir o cuantificar las relaciones existentes entre variables económicas. Es una materia interdisciplinar que utiliza la teoría económica, la matemática, la estadística y los métodos computacionales. En palabras de Ramanathan (2002):

*En términos sencillos, la **econometría** se ocupa de la aplicación de métodos estadísticos a la economía. A diferencia de la estadística económica, que es principalmente datos estadísticos, la econometría se distingue por la unificación de teoría económica, instrumentos matemáticos y metodología estadística. En términos más generales, la econometría se ocupa de (1) estimar relaciones económicas, (2) confrontar la teoría económica con los datos y contrastar hipótesis relativas al comportamiento económico, y (3) predecir el comportamiento de variables económicas.*

1.2.1. ¿Para qué sirve la Econometría?

El objetivo de un estudio econométrico es comprender mejor un fenómeno económico y, como resultado, poder realizar predicciones de la evolución futura del fenómeno de interés. El instrumento básico es el **modelo**, que ayuda a entender las relaciones entre variables económicas y sirve para evaluar los efectos de distintas medidas o políticas económicas. Algunos ejemplos en los que la Econometría es de utilidad son:

- Un analista del mercado de activos puede estar interesado en analizar y cuantificar la relación entre el precio de un activo y distintas características de la empresa que ofrece ese activo así como del estado general de la economía.
- Los directivos de Iberdrola pueden estar interesados en analizar los factores que afectan a la demanda de electricidad.
- El grupo Eroski puede estar interesado en cuantificar el efecto de distintos niveles de publicidad sobre sus ventas y sus beneficios.

- El servicio de estudios del Ministerio de Economía y del Banco de España o del Banco Central Europeo quiere analizar el impacto de las políticas monetarias y fiscales sobre el desempleo, la inflación, las exportaciones e importaciones, los tipos de interés, etc.
- Si un organismo quiere implementar políticas para corregir, por ejemplo, la discriminación salarial por sexo, en primer lugar debe conocer cuáles son los principales factores determinantes del problema y, en segundo lugar, analizar las posibles medidas a tomar, estudiando cuáles pueden ser los efectos de dichas medidas.
- Un gobierno regional puede necesitar previsiones sobre la evolución de la población para planificar la necesidad de servicios sociales y las necesidades de financiación que conllevan. También debe tener información precisa sobre su capacidad de financiación, por lo que le interesa disponer de predicciones relativas a la recaudación impositiva.
- Si una persona quiere contratar un préstamo, le interesa conocer cuál va a ser la evolución de los tipos de interés.

En los últimos años hemos asistido a una mayor difusión y utilización de los métodos econométricos gracias, entre otras razones, a la mayor disponibilidad y calidad de los datos y al desarrollo de los métodos de computación. Además, la aplicación de la Econometría no se restringe al ámbito estrictamente económico, sino que proporciona procedimientos de estudio de datos que pueden aplicarse al campo de las Ciencias Sociales. Por ejemplo, para:

- Analizar si el endurecimiento de las penas, como la introducción de la pena de muerte, tiene como consecuencia la disminución de la tasa de criminalidad.
- Analizar la efectividad de las medidas de seguridad vial, como el carnet por puntos, en la reducción del número de muertes en accidentes de tráfico.
- Predecir los resultados de una competición deportiva como, por ejemplo, el número de goles que marcará la selección de Inglaterra en un mundial de fútbol.
- Analizar cuál puede ser el efecto sobre los votantes en las próximas elecciones de una determinada medida, por ejemplo, prohibir fumar en lugares públicos, legalizar los matrimonios entre personas del mismo sexo, etc.
- Estudiar si hay diferencias en el voto dependiendo de si se trata de elecciones locales, regionales o europeas.
- Analizar si las medidas restrictivas sobre la publicidad de tabaco y alcohol reducen el consumo de estos productos.

Los comienzos de la Econometría pueden situarse en la década de los treinta del siglo pasado. Su coincidencia en el tiempo con la Gran Depresión no es casual: como consecuencia de ésta, los economistas de la época estaban interesados en poder predecir los ciclos económicos que observaban. Entre ellos destaca Keynes, que defendía la intervención del gobierno en la actividad económica para mitigar estas crisis. Así, los primeros econométricos se ocuparon de dar respuesta a problemas macroeconómicos con objeto de asesorar a los gobiernos en la implantación de políticas económicas.

En un comienzo, se aplicaron a los datos económicos métodos estadísticos que ya habían sido utilizados en ciencias naturales. Sin embargo, estos métodos no podían reproducirse miméticamente en el ámbito económico, sino que había que adaptarlos o desarrollar nuevos métodos de acuerdo a las características propias que poseen las variables socioeconómicas. Así, en la econometría se han desarrollado dos grandes áreas: la *econometría teórica*, cuyo objetivo es desarrollar métodos de estudio y análisis de datos y determinar sus propiedades, y la *econometría aplicada*, que se ocupa de utilizar estos métodos para responder a los problemas de interés en la práctica. En este curso ponemos mayor énfasis en la parte aplicada. Se trata de proporcionar al alumno las herramientas necesarias para que sea capaz de llevar a cabo un proyecto aplicado. Para ello, es indispensable dedicar tiempo al conocimiento de los métodos e instrumentos básicos del análisis econométrico, ya que son el requisito previo para una buena aplicación práctica.

1.3. Un estudio econométrico

Uno de nuestros objetivos específicos es que, al final del curso, el estudiante debe ser capaz de estructurar y desarrollar un trabajo de investigación. Hoy día, una persona que disponga de un ordenador en su casa puede llevar a cabo un pequeño proyecto econométrico. Así, un estudio econométrico consta de las siguientes etapas, Heij, de Boer, Franses, Kloek & Dijk (2004):

- *Formulación del problema.* Se trata de determinar la cuestión de interés. Debemos plantear de forma precisa las preguntas que nos interesa responder. Por ejemplo, si se trata de conocer la situación del mercado inmobiliario en una ciudad, podemos plantearnos la siguiente pregunta: ¿cuál es el precio de los pisos en esa ciudad y qué factores lo determinan? La teoría económica puede ayudarnos a enfocar el problema, a determinar qué variables están involucradas y cuál puede ser la relación entre ellas.
- *Recolección de datos* estadísticos relevantes para el análisis. En el ejemplo anterior, es fácil recolectar datos sobre el precio de pisos, su tamaño y otras características que pueden influir en su precio (ver Tabla 1.1). Los resultados del análisis van a depender en gran medida de la calidad de los datos. Sin embargo, no siempre es sencillo obtener los datos relevantes para el análisis. Podemos encontrar problemas como la ausencia de algún dato, cambios en la definición de una variable, fallos en el método de recogida, tener una cantidad insuficiente de datos o no disponer de información relativa a una variable.
- *Formulación y estimación del modelo.* De la unión de las teorías y cuestiones planteadas en la primera etapa con los datos se llega a un **modelo econométrico**. Por ejemplo, podemos plantear que, en media, el precio de un piso, Y , depende de su tamaño, X . Un posible modelo econométrico que recoge esta teoría es:

$$Y|X \sim N(\alpha + \beta X, \sigma^2)$$

Es decir, el precio de los pisos dado un tamaño, por ejemplo 100 m^2 , se distribuye alrededor de su media $\alpha + \beta 100$ según una normal de varianza σ^2 . Al formular el

modelo hemos elegido la forma funcional de la relación entre las variables y la naturaleza estocástica de la variable de interés o endógena, Y . El objetivo es obtener un modelo relevante y útil para dar respuesta a nuestros objetivos.

El siguiente paso es la estimación de los parámetros desconocidos de la distribución y que son de interés para el análisis. En el ejemplo del precio de los pisos, interesan los parámetros de su media, α y β . La estimación consiste en utilizar los datos y toda la información relevante para aprender algo sobre los parámetros desconocidos. En la interpretación de los resultados de estimación es importante tener en cuenta que *no conocemos* el valor de los parámetros, por lo que únicamente vamos a hacer afirmaciones del tipo “*con un 95 % de confianza, el aumento del impuesto sobre carburantes no afecta al consumo de gasolina*”.

Existen muchos métodos de estimación. La elección entre uno u otro depende de las propiedades del modelo econométrico seleccionado. Es decir, una mala selección del modelo también influye en la validez de las estimaciones. Un curso introductorio de Econometría, como este, se suele centrar en el estudio del modelo de regresión lineal y su estimación mediante *mínimos cuadrados ordinarios*, que son instrumentos sencillos y muy útiles en la práctica.

- *Análisis del modelo*. Se trata de estudiar si el modelo elegido es adecuado para recoger el comportamiento de los datos. Por ejemplo, si es correcto asumir que el tamaño del piso influye en su precio, si la relación lineal entre ambas variables es correcta, etc. Consiste en una serie de contrastes diagnósticos que valoran si el modelo está correctamente especificado, es decir, si los supuestos realizados son válidos. Si es necesario, se modifica el modelo en base a los resultados obtenidos en los contrastes.
- *Aplicación del modelo*. Una vez obtenido un modelo *correcto*, se utiliza para responder a las cuestiones de interés.

Dado que para la realización de un proyecto econométrico es necesario conocer dónde obtener los datos y manejar un software específico de análisis econométrico, vamos a extendernos un poco en estos dos puntos.

1.4. Los datos y su manejo

¿Cómo se obtienen datos económicos? No proceden de experimentos controlados sino que los economistas, al igual que otros investigadores del campo de las Ciencias Sociales, obtienen los datos de la observación de la realidad. En un experimento controlado, como los realizados en laboratorios, el investigador tiene control sobre las condiciones del estudio. Por ejemplo, para analizar el efecto de un fertilizante, podemos aplicar distintas dosis de fertilizante sobre un conjunto de sembrados, controlando también el grado de humedad o la luz que recibe cada planta. Además, se puede repetir el experimento, manteniendo las mismas condiciones o alterando algunas como las dosis o el grado de humedad. Obviamente, aunque las cantidades elegidas sean exactamente las mismas, no esperamos que el resultado, por ejemplo, el crecimiento de las plantas, sea idéntico entre experimentos porque las semillas utilizadas

son distintas o porque hay pequeños errores de medida. Estas diferencias naturales en los resultados de los experimentos se conocen como *variaciones muestrales*.

Los datos obtenidos de experimentos controlados son típicos de las Ciencias Naturales y se conocen como *datos experimentales*. Los datos que son resultado de un proceso que tiene lugar en la sociedad, y que no es controlable por una o varias personas, se conocen como *datos no experimentales*. Esta característica ha sido un factor importante en el desarrollo de las técnicas econométricas y debemos tenerlo en cuenta en la interpretación de los resultados.

Clasificación de los datos económicos. Los datos económicos pueden ser de diferentes tipos, lo que va a determinar el análisis que realicemos. Una primera clasificación distingue entre datos *cuantitativos*, aquéllos que toman valores numéricos dentro de un rango de valores, como precio o tamaño de un piso, y datos *cualitativos*, que aparecen como categorías o atributos, como por ejemplo el sexo, la profesión o el estado de un piso. Los seis primeros temas de este curso se centran en el análisis de datos cuantitativos. El tema siete considera situaciones en las que algún factor explicativo es cualitativo.

Una segunda clasificación distingue entre *datos de series temporales* y *datos de sección cruzada*. Los primeros se refieren a observaciones recogidas en sucesivos momentos de tiempo, normalmente regulares, como años, trimestres o meses. Ejemplos de datos temporales son el Producto Interior Bruto (PIB) de la Contabilidad Nacional trimestral, el número mensual de afiliaciones a la Seguridad Social o el valor diario del IBEX35. Los segundos se refieren a valores que toman diferentes agentes en un momento del tiempo, por ejemplo, la población desempleada en el año 2005 en cada uno de los países de la Unión Europea (UE), el salario medio en cada sector industrial en el 2006 o el gasto realizado en libros de texto por un conjunto de familias en septiembre pasado. También es posible tener una combinación de datos de sección cruzada y series temporales, por ejemplo, las puntuaciones obtenidas por los estudiantes de Econometría en los cursos 2004-05, 2005-06 y 2006-07. Cuando se encuesta a los mismos individuos a lo largo del tiempo, como la tasa de paro y el crecimiento del PIB desde 1990 hasta 2006 para los 25 países de la UE, se conocen con el nombre de *datos de panel* o *datos longitudinales*. En este curso nos centraremos en el análisis de datos de sección cruzada. Las técnicas que utilizemos también se pueden aplicar en series temporales, aunque en ocasiones su estudio es más complejo.

Una tercera clasificación se establece en función del nivel de agregación. Se conocen como *datos microeconómicos* o *microdatos* los referidos al comportamiento de agentes económicos como individuos, familias o empresas. Un ejemplo es la Encuesta de Población Activa, elaborada por el INE y publicada en http://www.ine.es/prodyser/micro_epa.htm. Los *datos macroeconómicos* o *macrodatos* son los datos referidos a ciudades, regiones o naciones que son resultantes de la agregación sobre agentes individuales, como son los resultados de la Contabilidad Nacional. Por ejemplo, la Contabilidad Nacional Trimestral de España, elaborada también por el INE y publicada en http://www.ine.es/inebmenu/mnu_cuentas.htm.

1.4.1. Fuentes de datos

Encontrar y recopilar datos no es siempre sencillo. En ocasiones es muy costoso coleccionar los datos adecuados a la situación y manejarlos. Sin embargo, esta tarea se ha visto favorecida en los últimos años por la mejora en la recogida de datos y el hecho de que muchos organismos permiten acceder a sus bases de datos en la *World Wide Web*. Algunos organismos que publican datos macroeconómicos son:

- Instituto Vasco de Estadística (EUSTAT): <http://www.eustat.es>.
- Banco de España: <http://www.bde.es> → Estadísticas. También publica el **Boletín estadístico mensual** y el Boletín de coyuntura mensual.
- Instituto Nacional de Estadística (INE): <http://www.ine.es> → Inebase o Banco tempus. Están disponibles, por ejemplo, los resultados de la encuesta de población activa, la Contabilidad Nacional o el **boletín estadístico mensual**. Además, en *enlaces* se encuentran otras páginas web de servicios estadísticos.
- EUROSTAT: Es la Oficina Estadística de la Unión Europea, se encarga de verificar y analizar los datos nacionales recogidos por los Estados Miembros. El papel de Eurostat es consolidar los datos y asegurarse de que son comparables utilizando una metodología homogénea. La información en términos de tablas estadísticas, boletines estadísticos e informativos, incluso working papers se puede encontrar en la dirección: <http://europa.eu.int/comm/eurostat>.
- Organización para la Cooperación y Desarrollo Económico (OCDE): <http://www.oecd.org>, Statistical portal, statistics. Están disponibles algunas series de las publicaciones **Main Economic Indicators** (mensual) o Comercio internacional.
- Fondo Monetario Internacional (FMI): <http://www.imf.org>. Para obtener datos sobre un amplio conjunto de países también se puede consultar su publicación **Estadísticas Financieras Internacionales** (mensual y anual).

Muchos manuales de Econometría incluyen una base de datos que se analizan en el texto como ilustración a la materia. En este curso utilizaremos principalmente los datos incluidos en Ramanathan (2002), que están accesibles como archivos de muestra en Gretl.

1.4.2. El software econométrico

El desarrollo de los ordenadores ha permitido almacenar una gran cantidad de datos, a la vez que ha facilitado su manejo. Existen en la actualidad un amplio conjunto de paquetes para el análisis econométrico que realizan complejas operaciones mediante unas instrucciones muy sencillas. Si los datos están disponibles en papel, las hojas de cálculo, como EXCEL, son un instrumento sencillo para introducir y preparar los datos y realizar operaciones sencillas. Sin embargo, en general es conveniente utilizar programas econométricos específicos. Algunos de los más populares en los cursos de Econometría son:

- **EViews**, desarrollado por Quantitative Micro Software, contiene una amplia gama de técnicas de análisis econométrico. Muchos manuales de Econometría contienen un CD con ejemplos prácticos en Eviews. Su página web con la información del programa es <http://www.eviews.com>.
- **SHAZAM**, elaborado en la Universidad British of Columbia (Canadá), incluye técnicas para estimar muchos tipos de modelos econométricos. Más información se puede obtener en <http://shazam.econ.ubc.ca>, donde se puede ejecutar el programa remotamente.
- **Gretl**, acrónimo de *Gnu Regression, Econometric and Time Series* (Biblioteca Gnu de Regresión Econometría y Series Temporales), elaborado por Allin Cottrell (Universidad Wake Forest). Es software libre, muy fácil de utilizar. También da acceso a bases de datos muy amplias, tanto de organismos públicos, como el Banco de España, como de ejemplos recogidos en textos de Econometría.
- **RATS**, acrónimo de *Regression Analysis of Time Series*. Contiene una amplia gama de técnicas de análisis econométrico con especial dedicación al Análisis de Series Temporales. Su web es: <http://www.estima.com>
- **R**, software libre para cómputo estadístico y gráficos. Consiste en un lenguaje, un entorno de ejecución, un debugger y la habilidad de correr programas guardados en archivos de tipo script. Su diseño fue influenciado por dos lenguajes existentes: S y Scheme. Página web: <http://www.r-project.org>

Un objetivo de este curso es que el estudiante se familiarice con el uso de programas econométricos. Por su sencillez y accesibilidad, en este curso introductorio se utiliza el programa Gretl para estudiar casos prácticos. En la página

http://gretl.sourceforge.net/gretl_espanol.html

se encuentra toda la información en castellano relativa a la instalación y manejo del programa. El manual, en inglés, se encuentra en la carpeta *en/*.

Junto con el programa se pueden cargar los datos utilizados como ejemplos de aplicaciones econométricas en los siguientes libros de texto Davidson & Mackinnon (2004), Greene (2003), Gujarati (1997), Ramanathan (2002), Stock & Watson (2003), Verbeek (2004), Wooldridge (2003).

Al instalar Gretl automáticamente se cargan los datos utilizados en Ramanathan (2002) y Greene (2003). El resto se pueden descargar de la página:

http://gretl.sourceforge.net/gretl_data.html

en la opción *textbook datasets*. Este curso se estructura sobre casos prácticos presentados en Ramanathan (2002) y en Wooldridge (2003) y ejercicios a resolver con ayuda de Gretl. La unión de teoría y práctica permiten al alumno un autoaprendizaje tanto de los contenidos básicos del curso de Análisis de Regresión como de la utilización del software Gretl.

1.5. Introducción a Gretl

La primera sesión con el programa Gretl consiste en una práctica guiada en la que se aprenderá a crear un fichero, introducir los datos de la Tabla 1.1 y realizar un análisis descriptivo.

Preparación del fichero. Al ejecutar Gretl, aparece la siguiente ventana principal:

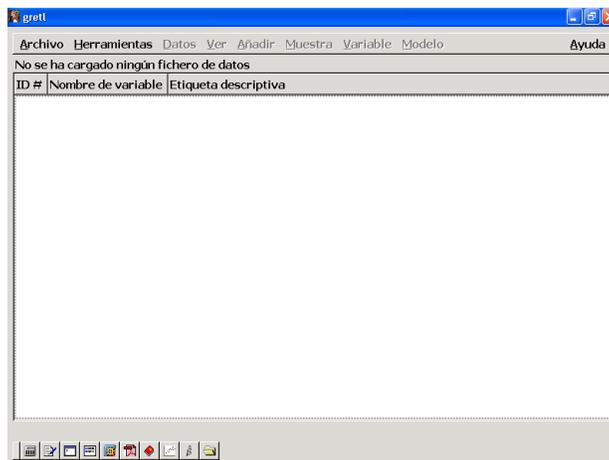


Gráfico 1.2: Pantalla inicial de Gretl

Como todavía no se ha cargado ningún fichero, varias opciones del menú principal, en gris claro, no están disponibles. Los datos a analizar no están incluidos en la base de Gretl, por lo que vamos a la opción *Archivo* → *Nuevo conjunto de datos Control+N*. Completamos la información que va solicitando el programa:

- *número de observaciones*, en la Tabla 1.1 se incluyen 50 pisos. Pinchar en *Aceptar*.
- El tipo de datos que utilizamos. En este caso, marcamos *de sección cruzada y Adelante*.
- Si el paso anterior se ha realizado correctamente, confirmamos la estructura del conjunto de datos pinchando en *Aceptar*. Al pinchar en *Atrás* se recupera sólo la ventana de tipo de datos, por lo que esta opción no permite corregir un error en el número de observaciones.
- En la última ventana marcaremos *Sí* queremos empezar a introducir los datos.
- En la siguiente ventana escribimos el *Nombre de la primera variable*, por ejemplo *m2*. No se pueden utilizar la letra *ñ*, acentos ni más de 15 caracteres para nombrar a las variables. Tras *Aceptar*, se abre una hoja de cálculo, de modo que en la pantalla aparece:

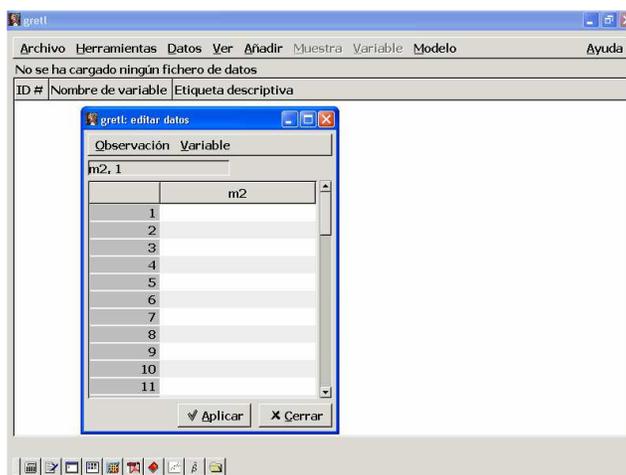


Gráfico 1.3: Añadir datos: hoja de cálculo de Gretl

Para incluir los datos de la variable $m2$, vamos a la celda correspondiente, por ejemplo la primera, y pinchamos sobre ella con la tecla izquierda del ratón; tras teclear la cifra, 55, damos a la tecla *Entrar*. Si por error no tecleamos algún dato, por ejemplo, la segunda observación de 59 m^2 , nos situaremos en la fila posterior, en este caso en el primer dato de 60 m^2 , y vamos a *observación* \rightarrow *insertar obs.* Se crea una nueva fila en blanco por encima de la anterior. Para guardar las modificaciones en la sesión de trabajo hay que pinchar en *Aplicar*.

Podemos añadir más variables con la opción *Variable* \rightarrow *Añadir* del menú de la hoja de cálculo. Por ejemplo, creamos una nueva variable que denominamos *Reforma*. Esta variable es cualitativa, por lo que asociamos a la situación *a reformar = sí* el valor 0 y a la otra opción, *a reformar = no* el valor 1. Una vez que se han incluido todos los datos, vamos a *Aplicar* y *Cerrar* la hoja de cálculo. Si no habíamos guardado los últimos cambios realizados, al cerrar la hoja de cálculo aparece un cuadro que nos pide confirmar los cambios. Las series creadas deben aparecer así en la pantalla:

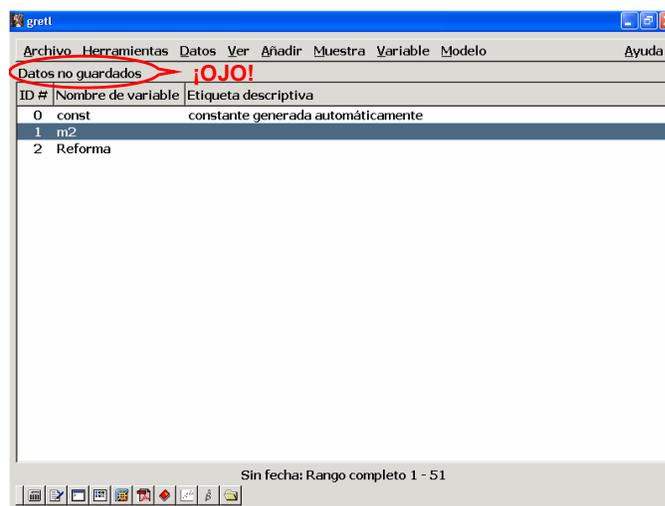


Gráfico 1.4: Fin de carga de datos con hoja de cálculo

Es recomendable guardar los datos ya incorporados en un fichero de datos Gretl mediante la opción del menú principal *Archivo* \rightarrow *Guardar datos*. En el siguiente cuadro añadimos el directorio y el nombre del fichero de datos, por ejemplo, *pisos*. Por defecto, grabará los datos con la extensión *gdt*. Para usar estos datos en una sesión posterior, sólo hay que pinchar dos veces sobre el fichero.

Con frecuencia, los datos están almacenados en otra hoja de cálculo, como EXCEL. Por ejemplo, en el fichero EXCEL *pisos.xls* se encuentran las variables $m2$ y *precio* de la Tabla 1.1. Añadir los datos de *precio* al fichero de Gretl es muy sencillo. Una vez abierto el fichero *pisos.gdt*, hay que:

- Utilizar la opción del menú principal *Archivo* \rightarrow *Añadir datos* \rightarrow *EXCEL*
- Dar el nombre y ubicación del fichero EXCEL, *pisos.xls*.
- Dar la celda a partir de la cual hay que empezar a importar los datos. En este caso la variable *precio* empieza en la celda B1, donde está su nombre, e importaremos los datos desde *columna 2, fila 1*. Para añadir las dos variables, $m2$ y *precio*, comenzaremos a importar datos en *columna 1, fila 1*. Finalmente, hay que pinchar en *Aceptar*.

Para comprobar si no hay errores en los datos vamos a *Datos* → *seleccionar todos* y luego activamos la hoja de cálculo mediante *Datos* → *Editar valores* o bien mostramos los datos en pantalla con *Datos* → *Mostrar valores* → *Todas las variables*. Debe aparecer la siguiente ventana:

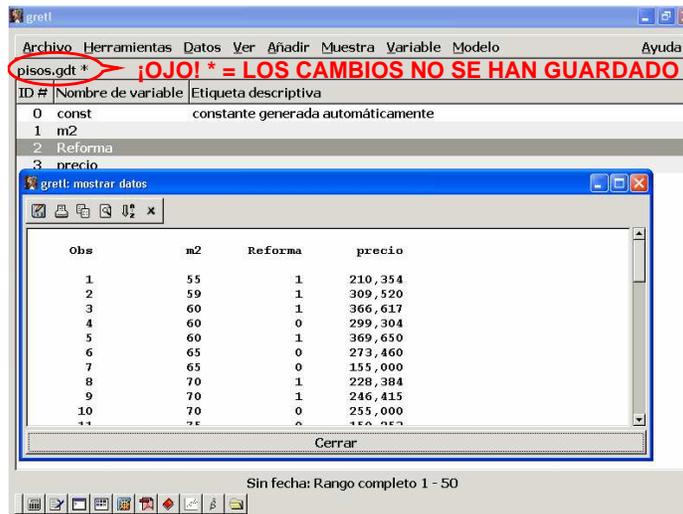


Gráfico 1.5: Fichero con datos de tres variables

Una vez que los datos se han cargado correctamente, los almacenamos en el mismo fichero *pisos.gdt* pinchando en *Archivo* → *Guardar datos*. Una vez guardadas las modificaciones, en la pantalla de Gretl aparece el nombre del fichero sin el asterisco *.

Notas explicativas. Al crear un fichero, nos interesa incluir notas explicativas del trabajo ya realizado. En Gretl es posible añadir esta información en dos apartados, uno general y otro específico de cada variable. Es posible añadir una breve descripción de cada variable y que aparezca como *etiqueta descriptiva* junto con el nombre de la variable. Por ejemplo, añadiremos la nota informativa sobre la interpretación de la variable *Reforma*:

Valor 0 si el piso está para reformar, valor 1 si está reformado

Marcamos con el ratón la variable y vamos a *Variable* → *editar atributos*. El cuadro siguiente en el apartado *descripción* escribimos el texto y pinchamos en *Aceptar* (ver Gráfico 1.6).

Las etiquetas descriptivas son útiles para saber la fuente de datos o las unidades de medida. Por ejemplo, para la variable *precio* y *m2* añadiremos las siguientes etiquetas descriptivas:

Variable	Etiqueta descriptiva	Nombre a mostrar en gráficos
<i>precio</i>	Precio de pisos en miles de euros	Precio (miles euros)
<i>m2</i>	Tamaño de pisos en metros cuadrados	Superficie (m2)

La opción *Datos* → *Editar información* da lugar a un cuadro que permite añadir texto informativo, por ejemplo,

Datos utilizados en el tema 1 de Análisis de regresión con Gretl

Finalmente, la opción *Datos* → *Ver descripción* permite visualizar la información de la estructura del conjunto de datos junto con las notas explicativas añadidas. Si todo el proceso se ha realizado correctamente, en pantalla debe aparecer el siguiente cuadro:

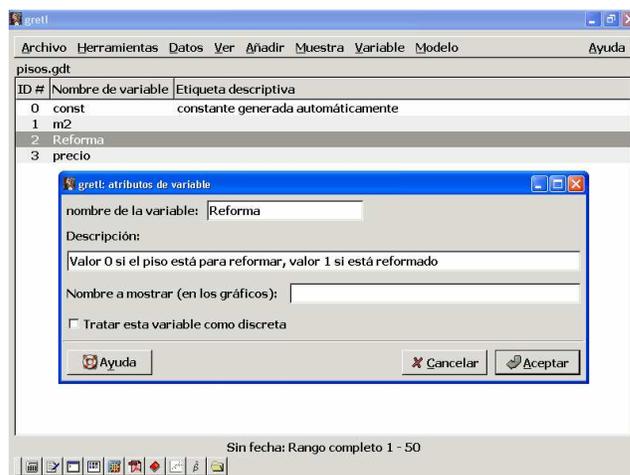


Gráfico 1.6: Cuadro de descripción de variables



Gráfico 1.7: Fichero con descripción de variables

1.5.1. Análisis descriptivo de una variable

Una vez incorporados los datos, vamos a obtener una visión general de los mismos. El objetivo del análisis descriptivo es resumir un conjunto de datos, extrayendo las características e información más relevante para el estudio. En primer lugar, sintetizaremos la información de cada una de las variables y en una segunda etapa, obtendremos una primera idea de las relaciones existentes entre las variables. Para ello se utilizan *gráficos* y números-resumen conocidos como *estadísticos descriptivos*¹. El análisis descriptivo de una única variable que proporciona Gretl se encuentra en la opción *variable* del menú principal; un resumen de este análisis se obtiene en el *menú auxiliar* que aparece al pinchar con la tecla derecha del ratón sobre la variable.

El gráfico más utilizado para resumir datos de sección cruzada de una única variable económica es el **histograma**, que aparece con la opción del menú auxiliar *Gráfico de frecuencias*. Se trata de un diagrama de barras que en el eje horizontal o abscisa representa los va-

¹Este apartado es un resumen de los conceptos mínimos relevantes. Explicaciones más detalladas se encuentran en manuales como Peña & Romo (1997).

lores de la variable divididos en intervalos. Sobre cada intervalo se dibuja una barra, cuya superficie refleja el número de observaciones que pertenecen a dicho intervalo. Si, por ejemplo, pinchamos con la tecla derecha del ratón sobre la variable *precios* y vamos a *Gráfico de frecuencias*, aparece el cuadro de opciones del histograma en la que fijamos:

- *Número de intervalos*: Por defecto aparecen 7 intervalos, que es un número entero próximo a \sqrt{N} , siendo N el número de observaciones, en este caso 50.
- *Valor mínimo intervalo izquierdo y grosor del intervalo*: todos los intervalos deben tener la misma amplitud. Por defecto, se eligen de manera que el punto central o marca de clase de los intervalos primero y último sean, respectivamente, los valores mínimo y máximo que toma la variable en el conjunto de datos.

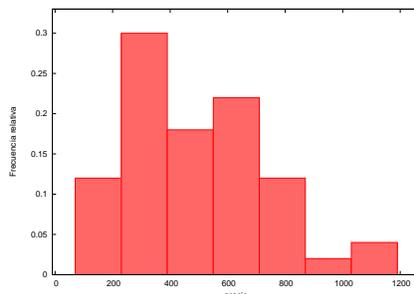


Gráfico 1.8: Histograma de frecuencias relativas

Usando las opciones estándar de Gretl obtenemos el Gráfico 1.8. Si pinchamos sobre el gráfico, se despliega un menú auxiliar que permite hacer cambios en el gráfico (*editar*) o guardarlo en diversos formatos (portapapeles, postscript, etc). La opción *guardar a sesión como icono* guarda el gráfico a lo largo de la sesión de Gretl. Es decir, una vez cerrada la ventana del gráfico, se recupera pinchando en el cuarto símbolo de la *barra de herramientas* situada en parte inferior derecha de la ventana principal (*vista iconos de sesión*) y, a continuación, pinchando dos veces en el icono *gráfico 1*.

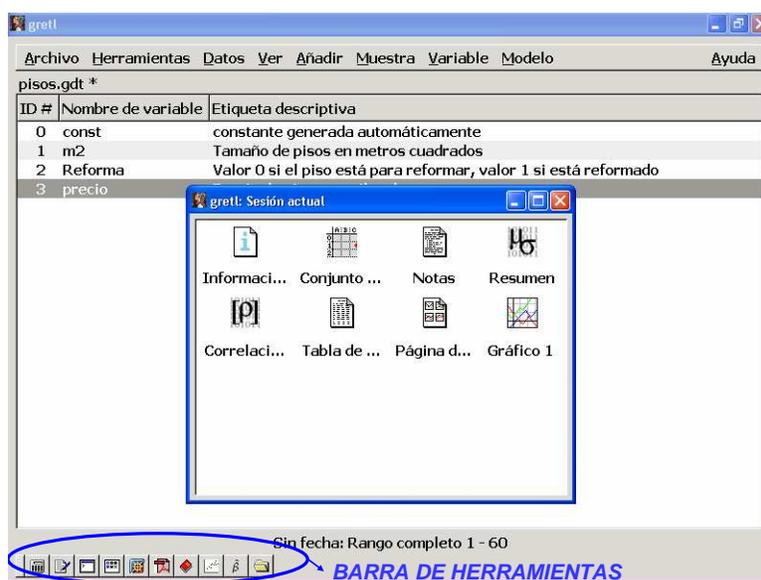


Gráfico 1.9: Iconos de la sesión

Para ver la tabla con la distribución de frecuencias representada en el histograma, hay que marcar la variable correspondiente e ir a la opción *Variable* → *Distribución de frecuencias*. Por ejemplo, la tabla de distribución de frecuencias de la variable *precio* es:

Distribución de frecuencias para *precio*, observaciones 1-50 número de cajas = 7, media = 489,858, desv.típ.=237,416

intervalo	punto medio	frecuencia	rel	acum.
< 230,23	150,25	6	12,00%	12,00% ****
230,23 - 390,19	310,21	15	30,00%	42,00% *****
390,19 - 550,15	470,17	9	18,00%	60,00% *****
550,15 - 710,11	630,13	11	22,00%	82,00% *****
710,11 - 870,06	790,08	6	12,00%	94,00% ****
870,06 - 1030,0	950,04	1	2,00%	96,00%
>= 1030,0	1110,0	2	4,00%	100,00% *

Tabla 1.2: Distribución de frecuencias del precio de 50 pisos

En la primera columna aparecen los intervalos en que se han dividido los valores que toma la variable *precio* y la segunda incluye el punto medio o **marca de clase** del intervalo. La columna *frecuencia* es lo que se conoce como **frecuencia absoluta** de un intervalo, es decir, el número de pisos con precio en ese intervalo. Por ejemplo, en la Tabla 1.1 hay 15 pisos cuyo precio se encuentra entre 230232€ y 390190€. La columna, *rel*, contiene la **frecuencia relativa** de cada intervalo, es decir, la fracción de observaciones que hay en cada tramo. Con estas frecuencias se ha construido el histograma anterior. Por ejemplo, los 15 pisos con precio en el intervalo [230,232; 390,190) constituyen el 30 % del total de los 50 pisos. Y, como todos los intervalos son de igual amplitud, la altura de la segunda barra del histograma es la frecuencia relativa asociada en tanto por uno, es decir, 0,3. Si a la frecuencia relativa de un intervalo se le suman las frecuencias relativas de los anteriores se obtiene la **frecuencia relativa acumulada** hasta cada intervalo, que aparece en la columna *acum*. Por ejemplo, en el conjunto de pisos que estudiamos, un 42 % de ellos tiene un precio inferior a 390190€.

La descripción numérica de una variable se encuentra en la opción del mismo menú auxiliar *Estadísticos descriptivos* o en el menú principal, *Variable* → *Estadísticos principales*. El resultado para la variable *precio* es la Tabla 1.3:

Estadísticos principales, usando las observaciones 1 - 50
para la variable 'precio' (50 observaciones válidas)

Media	489,86	Desviación típica	237,42
Mediana	466,68	C.V.	0,48466
Mínimo	150,25	Asimetría	0,68052
Máximo	1110,0	Exc. de curtosis	-0,19251

Tabla 1.3: Estadísticos descriptivos del precio de 50 pisos

Esta ventana tiene un nuevo menú. La opción *Copiar* permite importar la tabla a un fichero MS Word, Latex o simplemente, como aparece en pantalla (*Texto plano*). Estos estadísticos

descriptivos reflejan algunas características de la distribución recogidas en el histograma. La media y la mediana son medidas de posición, la desviación típica y el coeficiente de variación son medidas de dispersión, mientras que la asimetría y exceso de curtosis son medidas de forma de la distribución.

Las **medidas de posición** dan una idea de la situación o centro del conjunto de puntos. La *media* es el valor promedio. Si disponemos de N datos de una variable x_1, x_2, \dots, x_N , la media, o también momento muestral de primer orden, se define como:

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_N}{N} = \frac{1}{N} \sum_{i=1}^N x_i$$

La media es un estadístico poco robusto frente a la presencia de valores extremos: observaciones anómalas van a tener una gran influencia en el valor que tome. Por ejemplo, si el piso número 50 tuviera un precio *muy alto*, por ejemplo, 1350 miles de euros en lugar de 1051, entonces el precio medio aumentaría en casi 6000 euros, situándose en 495,84 miles de euros.

En general, interesan estadísticos cuyo valor no varíe mucho ante cambios en los valores de unas pocas observaciones, por muy grandes que sean esas variaciones. La *mediana*, que es el valor central de la distribución, posee esta propiedad. Así, la mediana del *precio* es 466,68 miles de euros.

Las medidas de posición proporcionan un valor representativo del conjunto de datos que debe complementarse con una medida del error asociado. Para valorar la representatividad de este único valor se utilizan las **medidas de dispersión**, que informan de si las observaciones están poco concentradas (o muy dispersas) alrededor de su centro. Una medida sencilla es la diferencia entre los valores máximo y mínimo que toman los datos en la muestra, lo que se conoce como *recorrido*. Es decir,

$$\text{Recorrido} = \text{Máximo} - \text{Mínimo}$$

En el ejemplo, tenemos que el recorrido de los precios es $1110 - 150,25 = 959,75$ miles de euros. Esta medida sólo tiene en cuenta dos valores, los extremos. Otras medidas se elaboran con todos los datos, por ejemplo, la desviación típica, que es la raíz cuadrada positiva de la varianza. La varianza de un conjunto de datos se define como un promedio de los cuadrados de las desviaciones de los datos a la media. Gretl calcula la varianza, S^{*2} o S_x^{*2} , como:

$$S_x^{*2} = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_N - \bar{x})^2}{N - 1} = \frac{1}{N - 1} \sum_{i=1}^N (x_i - \bar{x})^2$$

Por tanto, la *desviación típica*, S_x^* , se calcula según:

$$S_x^* = + \sqrt{\frac{1}{N - 1} \sum_{i=1}^N (x_i - \bar{x})^2}$$

Varianza y desviación típica son medidas de la dispersión de los datos alrededor de la media. Tiene el valor mínimo cero cuando todos los datos de la variable toman el mismo valor. La ventaja de la desviación típica es que tiene las mismas unidades de medida que la variable original. En general, cuanto más próxima a cero esté S_x^* , más concentrados estarán los datos

alrededor de la media y ésta será más representativa del conjunto de observaciones. Sin embargo, al depender S_x^* de las unidades de medida, no es fácil comparar su representatividad en dos conjuntos de datos. Para solucionar este problema se utiliza el *coeficiente de variación*, $C.V.$, que es una medida adimensional de la dispersión, y se define como:

$$C.V. = \frac{S_x^*}{|\bar{x}|} \quad \text{si } \bar{x} \neq 0$$

En el ejemplo de precios tenemos que $C.V. = 0,485 < 1$, la dispersión de los datos es pequeña en relación a su nivel, por lo que consideramos que la media sí es bastante representativa del conjunto de datos.

Media y desviación típica son los estadísticos-resumen más conocidos. Se acompañan de las **medidas de forma**, que reflejan otras características del histograma. La asimetría de una distribución se refiere a si los datos se distribuyen de forma simétrica alrededor de la media o no. El *coeficiente de asimetría* se define como:

$$\text{Coeficiente de asimetría} = \frac{1}{N} \sum_{i=1}^N \left(\frac{x_i - \bar{x}}{S_x} \right)^3 = \frac{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^3}{S_x^3}$$

con $S_x = \sqrt{(N-1)/N} \times S_x^* = \sqrt{\sum_i (x_i - \bar{x})^2 / N}$. El coeficiente de asimetría es cero cuando los datos se distribuyen simétricamente alrededor de la media, es positivo cuando la cola derecha (asociada a valores por encima de la media) es más larga que la izquierda siendo negativa en caso contrario. En el ejemplo de los precios de los pisos, observamos que la asimetría es positiva, lo que se corresponde con una media mayor que la mediana, es decir, $\bar{x} > \text{Mediana}(X)$.

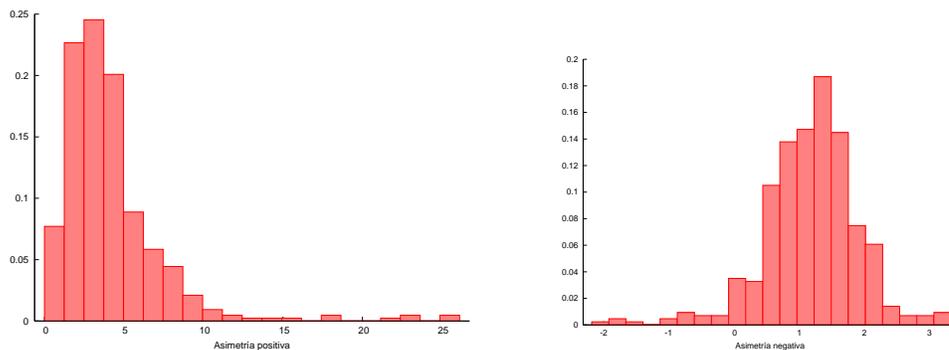


Gráfico 1.10: Tipos de asimetría

El coeficiente de curtosis es una medida del apuntamiento de la distribución y se define:

$$\text{Curtosis} = \frac{1}{N} \sum_{i=1}^N \left(\frac{x_i - \bar{x}}{S_x} \right)^4 = \frac{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^4}{S_x^4}$$

Este coeficiente mide la cantidad de observaciones que se encuentran en las colas en relación con las situadas alrededor de la media. El nivel de referencia es tres, que es el valor de la

curtosis de la distribución *normal*. Así, se define el *exceso de curtosis* como:

$$\text{Exc. de curtosis} = \frac{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^4}{S_x^4} - 3 \quad (1.1)$$

Un exceso de curtosis positivo indica mayor peso de observaciones en la cola y mayor apuntamiento que la distribución normal, mientras que si es negativo indica menor número de observaciones en la cola y menor apuntamiento.

Cuando tenemos un conjunto de variables, Gretl permite recoger en una única tabla los estadísticos descriptivos de todas las variables. El proceso es el siguiente:

1. Seleccionar las variables de interés pinchando simultáneamente la tecla izquierda del ratón y la tecla *Control*.
2. Ir a *Ver* → *Estadísticos principales* o utilizar *Estadísticos descriptivos* en el menú auxiliar que aparece al pinchar la tecla derecha del ratón sobre las variables seleccionadas.

Así, con los datos de la Tabla 1.1 se obtiene la siguiente tabla de estadísticos descriptivos:

Estadísticos principales, usando las observaciones 1 - 50

Variable	MEDIA	MEDIANA	MIN	MAX
m2	127,34	105,00	55,000	250,00
Reforma	0,62000	1,0000	0,00000	1,0000
precio	489,86	466,68	150,25	1110,0

Variable	D.T.	C.V.	ASIMETRÍA	EXC.CURTOSIS
m2	59,048	0,46370	0,67091	-0,77954
Reforma	0,49031	0,79083	-0,49445	-1,7555
precio	237,42	0,48466	0,68052	-0,19251

Tabla 1.4: Estadísticos descriptivos del conjunto de datos

donde D.T. indica desviación típica, MIN es mínimo y MAX denota el máximo. Al interpretar estos resultados, hay que tener en cuenta que la variable *Reforma* no es una variable cuantitativa continua, sino una variable cualitativa discreta, que sólo toma valores 1 ó 0.

1.5.2. Relaciones entre variables

Cuando el conjunto de datos contiene, por ejemplo, dos variables cuantitativas nos interesa estudiar la relación o asociación que existe entre ellas. En general, al analizar dos (o más) variables, podemos establecer una relación de causalidad entre ellas. Por ejemplo, podemos pensar que el precio de un piso puede ser consecuencia del tamaño de la vivienda, pero no al revés. Se llama variable independiente o exógena, x , a la que causa el efecto y variable dependiente o endógena, y , a la que lo recibe. La relación entre estas variables puede estudiarse con gráficos o expresarse numéricamente mediante, por ejemplo, el coeficiente de correlación. Todos estos elementos del análisis descriptivo de un conjunto de variables se realiza con el menú que se despliega en la opción *Ver* de Gretl.

Representación gráfica. El diagrama de dispersión o *scatterplot* da una primera idea de la relación entre dos variables. Es el gráfico que representa cada punto (x_i, y_i) , $i = 1, \dots, N$ en el plano: la variable x aparece en el eje de abscisas y la variable y en el eje de ordenadas. Por ejemplo, para obtener con Gretl el Gráfico 1.11, precio sobre superficie, podemos seguir uno de los siguientes pasos:

- Ver \rightarrow Gráficos \rightarrow Gráfico X-Y (*scatter*) y en el cuadro *Definir el gráfico* marcar:
Variable de eje X Elegir \rightarrow $m2$
Variables de eje Y Añadir \rightarrow *precio*
- O bien seleccionar las variables *precio* y $m2$ pinchando simultáneamente la tecla izquierda del ratón y la tecla *Control* e ir al menú auxiliar, *Gráfico de dos variables XY*. En el siguiente cuadro, se selecciona la variable de la abscisa, $m2$.

Al pinchar en *Aceptar* aparece el Gráfico 1.11 que, además de la nube de puntos, incluye una recta-síntesis de la relación, la recta de regresión mínimo cuadrática que veremos más adelante.

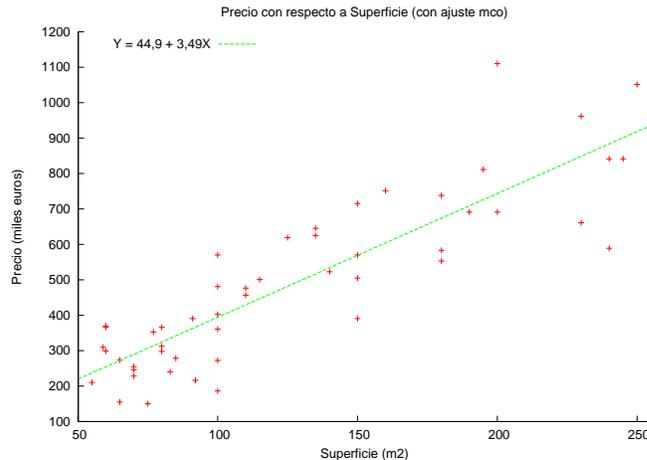


Gráfico 1.11: Diagrama de dispersión superficie-precios (2)

Al pinchar sobre el gráfico aparece un menú auxiliar que sirve para:

- Exportar el gráfico a ficheros en diferentes formatos en *Guardar como Windows metafile (EMF)...*, *PNG...*, *postscript (EPS)...*, *PDF...*
- Copiar/exportar el gráfico a otros ficheros con *Copiar al portapapeles*.
- Guardar el fichero en la sesión de Gretl en *Guardar la sesión como icono*.
- Realizar cambios en el fichero con *Editar*. En la pestaña *Principal* se controla el título del gráfico, el tamaño y tipo de letra, el color de las líneas/puntos, el dibujo del marco completo, la situación de texto explicativo de las variables representadas (*posición de la clave*) o la eliminación de la recta-resumen. La escala y la explicación de los ejes se modifica en *Eje X* y *Eje Y*. En *líneas* se controla la representación de los datos, tipo de línea o punto, y el texto explicativo de las variables. *Etiquetas* permite añadir texto en el gráfico y *salida a fichero* incluye varios formatos para guardar el gráfico.

El gráfico de dispersión permite distinguir la posible relación, lineal o no, que existe entre las variables. Se dice que hay una **relación lineal positiva** entre ambas variables cuando al aumentar x , aumenta en promedio el valor de y (figura b en el Gráfico 1.12). Diremos que hay una **relación lineal negativa** entre ambas variables cuando observamos que al aumentar x , disminuye en promedio el valor de y (figura c). En el ejemplo, se observa una clara relación lineal positiva entre precio y tamaño del piso.

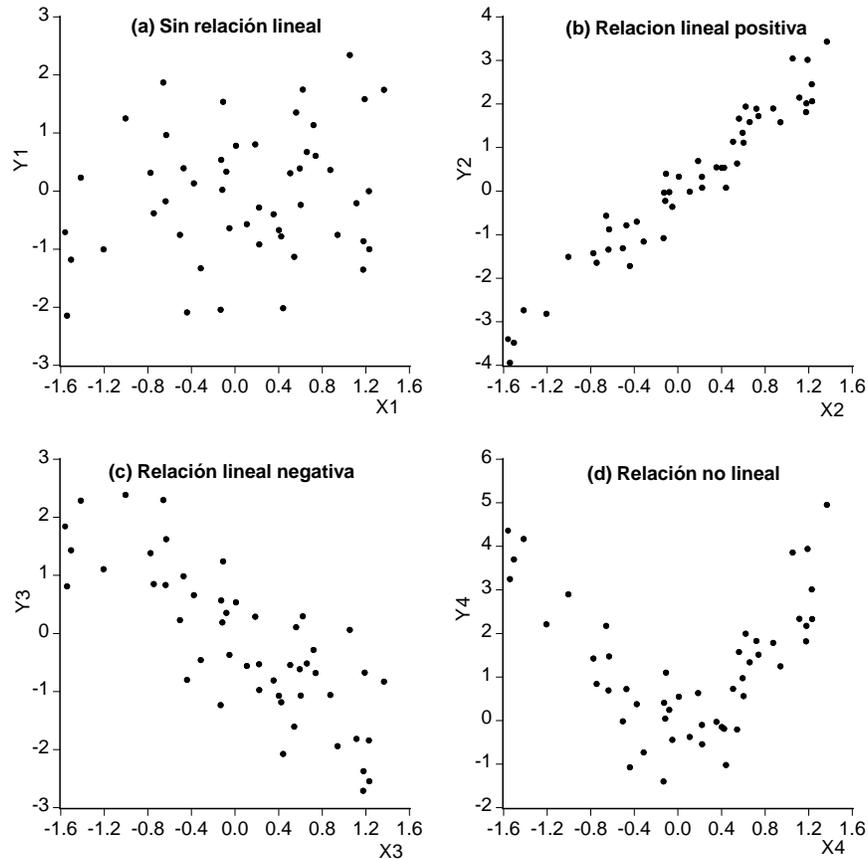


Gráfico 1.12: Diagramas de dispersión

Covarianza y correlación. La **covarianza** es una medida del grado de asociación lineal entre dos variables. Si se tienen N pares de datos de dos variables, $(x_1, y_1) \dots (x_N, y_N)$, la covarianza se denota por S_{xy} y se define:

$$S_{xy} = cov(x, y) = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})$$

siendo \bar{x} e \bar{y} las medias aritméticas de las variables. La covarianza depende de las unidades de medida de las variables, lo que no permite comparar la relación entre distintos pares de variables medidas en unidades diferentes. En estos casos se utiliza el **coeficiente de correlación lineal** entre x e y , que se define:

$$r_{xy} = corr(x, y) = \frac{S_{xy}}{S_x S_y} = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^N (y_i - \bar{y})^2}}$$

El coeficiente de correlación lineal y la covarianza tienen el mismo signo: son positivos si existe relación lineal directa o positiva (figura b en el Gráfico 1.12), son negativos si existe relación lineal inversa o negativa (figura c) y toma valor cero si x e y son independientes (figura a) o cuando la relación, si existe, es no lineal (figura d). Además, su valor no depende del orden en que se consideren las variables, es decir, $S_{xy} = S_{yx}$ y $r_{xy} = r_{yx}$. A diferencia de la covarianza, el coeficiente de correlación es una medida adimensional de la relación que toma valores entre -1 y 1, $-1 \leq r_{xy} \leq 1$: un coeficiente de correlación igual a uno en valor absoluto indica que las variables están relacionadas linealmente de forma exacta y los datos se sitúan sobre una línea.

En Gretl, si se marcan las variables que interesan y se va a *Ver* \rightarrow *Matriz de correlación* se obtiene una tabla (matriz) con los coeficientes de correlación para cada par de variables consideradas. El resultado para los datos de precios, tamaño y reforma de los pisos es:

Coeficientes de correlación, usando las observaciones 1 - 50
valor crítico al 5% (a dos colas) = 0,2787 para n = 50

m2	Reforma	precio	
1,0000	0,0440	0,8690	m2
	1,0000	0,2983	Reforma
		1,0000	precio

Tabla 1.5: Matriz de coeficientes de correlación

Por ejemplo, el coeficiente de correlación entre el precio y el tamaño de los pisos se encuentra en la primera fila, columna tercera, (precio-m2). Es decir, $r_{precio,m2} = 0,869$, lo que indica que hay una fuerte relación lineal positiva entre estas variables. Hay que tener en cuenta que este coeficiente se define para variables cuantitativas, por lo que no lo aplicamos a la variable *Reforma*.

Bibliografía

Davidson, D. y J. Mackinnon (2004), *Econometric Theory and Methods*, Oxford University Press.

Greene, W. (2003), *Econometric Analysis*, 5ª edn., Prentice-Hall.

Gujarati, D. (1997), *Econometría básica*, 4ª edn., McGraw-Hill.

Heij, C., de Boer, P., Frances, P., Kloek, T. y H. Van Dijk (2004), *Econometric Methods with Applications in Business and Economics*, Oxford University Press.

Peña, D. y J. Romo (1997), *Introducción a la Estadística para las Ciencias Sociales*, McGraw-Hill.

Ramanathan, R. (2002), *Introductory Econometrics with Applications*, 5ª edn., South-Western.

Stock, J. y M. Watson (2003), *Introduction to Econometrics*, Addison-Wesley.

Verbeek, M. (2004), *A Guide to Modern Econometrics*, 2ª edn., John Wiley.

Wooldridge, J. M. (2003), *Introductory Econometrics. A Modern Approach*, 2ª edn., South-Western.

