



Actividad A3

Gestión de los datos en Gretl

Pilar González y Susan Orbe

Dpto. Economía Aplicada III (Econometría y Estadística)

Contenido

- 1 Actividad A3.1. Generar ficheros de datos en Gretl.
- 2 Actividad A3.2. Modificar ficheros de datos en Gretl.
- 3 Actividad A3.3. Análisis descriptivo de datos en Gretl.
- 4 Actividad A3.4. Análisis gráfico en Gretl.
- 5 Actividad A3.5. Análisis de una sola variable.

Contenido

- 1 Actividad A3.1. Generar ficheros de datos en Gretl.
- 2 Actividad A3.2. Modificar ficheros de datos en Gretl.
- 3 Actividad A3.3. Análisis descriptivo de datos en Gretl.
- 4 Actividad A3.4. Análisis gráfico en Gretl.
- 5 Actividad A3.5. Análisis de una sola variable.

A3.1. Alquiler de sombrillas.

Enunciado.

Una empresa familiar desea conocer la relación existente entre el número de sombrillas alquiladas en la playa (S , en unidades por semana) y el precio de alquiler por día que se mantiene constante a lo largo de la semana (P , en euros). La empresa ha guardado los datos semanales de la temporada anterior desde la primera semana de mayo hasta la penúltima de septiembre:

t	1	2	3	4	5	6	7	8	9	10	11
S	347	251	339	310	305	317	364	403	323	322	401
P	4,5	4,5	4,5	5	5	6	6	7	7	8	9
t	12	13	14	15	16	17	18	19	20	21	
S	394	424	442	462	491	437	422	433	393	429	
P	9	10	10	12	15	15	12	10	10	9	

A3.1. Alquiler de sombrillas.

Cuestiones.

- a. Introduce manualmente los datos en Gretl.
- b. Introduce los atributos para cada una de las variables.
- c. Introduce información sobre el conjunto de datos.
- d. Guarda estos datos en un fichero de formato Gretl denominado `sombrilla.gdt`.

A3.1. Matriculación de turismos.

Enunciado.

Se desea analizar la evolución de la venta de coches en la Comunidad Autónoma Vasca en los últimos años, para lo que se cuenta con datos sobre las siguientes variables:

MATCAV: número total de turismos matriculados en la CAV.

MATSP: número total de turismos matriculados en el estado español.

price_p: precio del barril de petróleo Brent en dólares.

Estos datos se encuentran en el fichero `Matriculaciones.xls`, en la hoja denominada `Series matriculacion`.

A3.1. Matriculación de turismos.

Cuestiones.

- a. Importa estos datos a un fichero Gretl.
- b. Dota de atributos a las variables.
- c. Guarda estos datos en un fichero de formato Gretl denominado `matriculaciones.gdt`.
- d. Posteriormente te informan de que los datos son de series temporales (enero de 2004 a diciembre de 2012). Da a los datos del fichero `matriculaciones.gdt` la estructura de datos de series temporales.

A3.1. Salarios.

Enunciado.

Abre el fichero `wage1.gdt` que se encuentra en la carpeta denominada `Wooldridge`, correspondiente al libro `Wooldridge (2001)`.

- Guarda este fichero en tu propio directorio.
- Sólo interesan las variables `wage`, `educ`, `exper`, `tenure`, `nonwhite`, `female`, `married`, `numdep` y `smsa`. Borra el resto de las variables del fichero.
- Cambia los nombres de las variables y las descripciones para que aparezcan en castellano.

`wage` - salario: salario medio por hora en dólares

`exper` - `exper`: años de experiencia

`female` - `mujer`: = 1 si es mujer

`numdep` - `numdep`: número dependientes

`tenure` - `antigüedad`: años trabajando en la misma empresa

`educ` - `educ`: años de educación

`nonwhite` - `raza`: = 1 si no es blanco

`married` - `casado`: = 1 si está casado

`smsa` - `ciudad`: = 1 si vive en una metrópoli

- Guarda el fichero con el nombre `salarios.gdt`.

A3.1. Casas rurales en Bizkaia.

Enunciado.

La página web nekatur.net dispone de una gran cantidad de información sobre agroturismos y casas rurales del País Vasco. El fichero de datos NBizkaia.txt contiene información sobre las siguientes variables:

- precio medio de la habitación (PR, en euros)
- precio del desayuno por persona (PD, en euros)
- número de habitaciones (HB)
- localización (LOC)
- cercanía a algún parque natural (PN)
- cercanía a algún lago o embalse (LG)
- cercanía a alguna playa (PL)
- tener acceso a wifi (WIFI)
- poseer la Q de calidad (Q)
- tener el compromiso de calidad turística (KT).

A3.1. Casas rurales en Bizkaia.

Codificación en el fichero de texto *NBizkaia.txt*.

Las variables LOC, PN, LG, PL, WIFI, Q y KT que se encuentran en el fichero NBizkaia.txt están codificadas de la siguiente forma:

LOC: A si está alejado del centro urbano
 C si se encuentra en el centro urbano

PN, LG y PL: V si están a más de 20 km
 R si están a menos de 1km
 N si no tienen este servicio cerca
 P si están en el propio servicio

WIFI: G si es la conexión es gratuita
 N si no se dispone de conexión
 S si la conexión conlleva pago

Q y KT: N si no la tienen
 S si la tienen

A3.1. Casas rurales en Bizkaia.

Cuestiones.

- a. Importa el fichero de datos NBizkaia.txt. Ten en cuenta que el formato del fichero *NBizkaia.txt* es de texto, los datos son de sección cruzada, las variables están separadas por tabulador y en los números el decimal viene indicado por una coma (,).
- b. Observa la recodificación que hace Gretl de las variables cualitativas.
- c. Dota de atributos a todas las variables y verifica que las variables cualitativas han sido consideradas como variables discretas.
- d. Guarda el fichero de datos en el archivo NekaturBizkaia.gdt.

Contenido

- 1 Actividad A3.1. Generar ficheros de datos en Gretl.
- 2 Actividad A3.2. Modificar ficheros de datos en Gretl.**
- 3 Actividad A3.3. Análisis descriptivo de datos en Gretl.
- 4 Actividad A3.4. Análisis gráfico en Gretl.
- 5 Actividad A3.5. Análisis de una sola variable.

A3.2. Alquiler de sombrillas.

Añadir observaciones.

Buscando en sus archivos, la empresa familiar de sombrillas encuentra observaciones para la última semana de septiembre de la temporada anterior:

<i>t</i>		22
<i>S</i>		379
<i>P</i>		9

Introduce manualmente esta nueva observación en el fichero `sombrillas.gdt`.

A3.2. Alquiler de sombrillas.

Añadir variables.

Un asesor de esta empresa familiar considera que el alquiler de sombrillas puede depender también de las condiciones climatológicas por lo que sugiere incluir las siguientes variables en este estudio:

T : temperatura media de la semana (en grados centígrados).

V : si es una semana de viento o no.

Introduce las siguientes variables en el fichero `sombrillas.gdt` con los datos de la siguiente tabla. En lo que se refiere a la variable V , dále el valor 1 para las semanas sin viento y el valor 2 para las semanas con viento.

t	1	2	3	4	5	6	7	8	9	10	11
T	26	23	25	23	23	25	28	33	28	30	33
V	sí	sí	no	sí	sí	no	no	no	no	sí	sí
t	12	13	14	15	16	17	18	19	20	21	22
T	32	34	35	36	40	36	34	36	33	35	32
V	sí	no	no	no	no	no	sí	no	no	no	sí

A3.2. Matriculación de turismos.

Importar nuevas variables de Excel.

Se considera que la evolución de la situación económica puede haber afectado también la evolución de la venta de coches en la Comunidad Autónoma Vasca en los últimos años. Para medir esta variable se utiliza como proxy el Índice de Producción Industrial (IPI).

En el fichero `matriculaciones.xls`, en la hoja denominada IPI, podemos encontrar la siguiente información:

IPISP: Índice de Producción Industrial del estado español (datos brutos).

IPISPVA: tasa de variación anual del IPI del estado español.

IPICAV: Índice de Producción Industrial de la CAV (desestacionalizado).

IPICAVVA: tasa de variación anual del IPI de la CAV.

A3.2. Matriculación de turismos.

Cuestiones.

- a. Importa estos datos a Gretl.
- b. Dota de atributos a las variables.
- c. Guarda estos datos en un fichero de formato Gretl, denominado
`matriculaciones.gdt`.

A3.2. Simulación de datos.

Enunciado.

Abre un *Nuevo conjunto de datos* en Gretl con las siguientes características:

- datos de sección cruzada.
- tamaño muestral: 1000.

a. Simula las siguientes variables:

u , extraída de una distribución normal de media cero y varianza 9.

X_2 , extraída de una distribución normal de media seis y varianza veinticinco.

X_3 , extraída de una distribución uniforme comprendida entre 16 y 65.

b. Simula la siguiente variable

$$Y = 3 + 0,5 X_2 - 2,4 X_3 + u$$

c. Guarda los datos en el fichero con el nombre `simulados.gdt`

Contenido

- 1 Actividad A3.1. Generar ficheros de datos en Gretl.
- 2 Actividad A3.2. Modificar ficheros de datos en Gretl.
- 3 Actividad A3.3. Análisis descriptivo de datos en Gretl.**
- 4 Actividad A3.4. Análisis gráfico en Gretl.
- 5 Actividad A3.5. Análisis de una sola variable.

A3.3. Análisis descriptivo de datos en Gretl.

Casas rurales en Bizkaia.

- a. Obtén los estadísticos descriptivos de las variables cuantitativas del fichero `NekaturBizkaia.gdt`.

¿Cuál es el rango del precio medio por habitación?, ¿y el del número de habitaciones?, ¿cuáles son el precio mínimo y máximo por persona que hay que pagar por desayunar?
- b. Obtén la matriz de correlaciones entre las variables cuantitativas. Comenta los resultados.

Salarios.

- a. Obtén los estadísticos descriptivos de las variables salario, educación, experiencia y antigüedad del fichero `salarios.gdt`. Comenta los resultados.
- b. Obtén la matriz de correlaciones de las variables anteriores. Comenta los resultados.

Contenido

- 1 Actividad A3.1. Generar ficheros de datos en Gretl.
- 2 Actividad A3.2. Modificar ficheros de datos en Gretl.
- 3 Actividad A3.3. Análisis descriptivo de datos en Gretl.
- 4 Actividad A3.4. Análisis gráfico en Gretl.**
- 5 Actividad A3.5. Análisis de una sola variable.

A3.4. Gráficos de series temporales.

Alquiler de sombrillas.

Abre el fichero de datos `sombrillas.gdt` y dibuja los gráficos de series temporales para las variables `sombrillas` alquiladas, `precio` y `temperatura`.

- a. Representa las variables en un gráfico conjunto.
- b. Interpreta los resultados:
 - ¿Observas un comportamiento tendencial en alguna de las series?
 - ¿Observas un comportamiento estacional en alguna de las series?

A3.4. Gráficos de series temporales.

Matriculación de turismos.

Abre el fichero de datos `matriculaciones.gdt`:

- a. Representa los gráficos de series temporales de las variables matriculaciones en la CAV, matriculaciones en España y precio del petróleo:
 - en el mismo gráfico.
 - en gráficos separados.
- b. Representa en el mismo gráfico las series del IPI de España, IPI de la CAV, tasa de variación anual del IPI de España, tasa de variación anual IPI de la CAV.
- c. Interpreta los resultados:
 - ¿Observas un comportamiento tendencial en alguna de las series?
 - ¿Observas un comportamiento estacional en alguna de las series?

A3.4. Gráficos bivariantes.

Alquiler de sombrillas.

Abre el fichero de datos `sombrillas.gdt` y representa gráficamente la relación entre las siguientes variables:

- Sombrillas alquiladas frente al precio.
- Sombrillas alquiladas frente a la temperatura.
- Sombrillas alquiladas frente al viento.
- Interpreta los resultados.

A3.4. Gráficos bivariantes.

Matriculación de turismos.

Abre el fichero de datos `matriculaciones.gdt` y representa gráficamente la relación entre las siguientes variables:

- Matriculaciones en la CAV frente al precio del petróleo.
- Matriculaciones en la CAV frente al IPI de la CAV.
- Matriculaciones en la CAV frente al IPI de la CAV (tasa variación anual).
- Interpreta los resultados.

A3.4. Gráficos bivariantes.

Salarios.

Abre el fichero de datos `salarios.gdt` y representa gráficamente la relación entre las siguientes variables:

- Salarios frente a la educación.
- Salarios frente a la experiencia.
- Salarios frente a la antigüedad.
- Experiencia frente a la antigüedad.
- Experiencia frente a la educación.
- Educación frente a la antigüedad.
- Interpreta los resultados.

A3.4. Gráficos trivariantes.

Alquiler de sombrillas.

Abre el fichero de datos `sombrillas.gdt` y representa gráficamente la relación entre las variables `sombrillas` alquiladas, `precio` y `temperatura`.

Matriculación de turismos.

Abre el fichero de datos `matriculaciones.gdt` y representa gráficamente la relación entre las siguientes variables:

- Matriculaciones en España frente al precio del petróleo y al IPI de España.
- Matriculaciones en España frente al precio del petróleo y al IPI de España (tasa variación anual).
- Interpreta los resultados.

A3.4. Gráficos trivariantes.

Salarios.

Abre el fichero de datos `salarios.gdt` y representa gráficamente la relación entre las siguientes variables:

- Salarios frente a la educación y a la experiencia.
- Salarios frente a la experiencia y la antigüedad.
- Salarios frente a la antigüedad y la educación.
- Experiencia frente a la educación y la antigüedad.
- Interpreta los resultados.

Contenido

- 1 Actividad A3.1. Generar ficheros de datos en Gretl.
- 2 Actividad A3.2. Modificar ficheros de datos en Gretl.
- 3 Actividad A3.3. Análisis descriptivo de datos en Gretl.
- 4 Actividad A3.4. Análisis gráfico en Gretl.
- 5 Actividad A3.5. Análisis de una sola variable.

A3.5. Análisis de una sola variable. Distribución.

Alquiler de sombrillas.

Abre el fichero de datos `sombrillas.gdt` y analiza la distribución de las variables `sombrillas` alquiladas, `precio` y `temperatura`.

- Realiza los contrastes de normalidad de las tres variables.
- Obtén la distribución de frecuencias de las variables `sombrillas` alquiladas y `temperatura` junto a los contrastes de normalidad y gamma.
- Interpreta los resultados.

¿Alguna de las variables proviene de una distribución normal?

¿Alguna de las variables proviene de una distribución gamma?

¿Qué características puedes observar de las distribuciones de frecuencias que has obtenido?

A3.5. Análisis de una sola variable. Distribución.

Matriculación de turismos.

Abre el fichero de datos `matriculaciones.gdt`.

- Estima las funciones de densidad de las variables matriculaciones en la CAV y matriculaciones en España utilizando el kernel gaussiano.
- Obtén las distribuciones de frecuencias de las dos variables mencionadas junto con la información del contraste de normalidad.
- Obtén los gráficos Q-Q de dichas variables.
- Interpreta los resultados.
 - ¿Son las características de las funciones de densidad estimadas similares?, ¿en qué se diferencian?
 - ¿Qué características puedes observar de las distribuciones de frecuencias que has obtenido?
 - ¿Coincide la información proporcionada por las distribuciones de frecuencias y los gráficos Q-Q?

A3.5. Análisis de una sola variable. Distribución.

Casas rurales en Bizkaia.

Abre el fichero de datos `NekaturBizkaia.gdt`.

- a. Analiza la distribución de la variable número de habitaciones.
- b. Obtén la distribución de frecuencias de la variable precio medio con la información de los contrastes.
- c. Obtén la distribución de frecuencias para cada una de las variables cualitativas.
 - c.1. ¿Cuántas casas rurales están a menos de un kilómetro de la playa?
 - c.2. ¿Cuántas casas rurales están a más de 20 kilómetros de un lago?
 - c.3. ¿Cuántas casas rurales están en un parque natural?
 - c.4. ¿Cuántas casas rurales están a menos de 1 kilómetro de la playa y de un parque natural?, ¿cuántas están a más de 20 kilómetros?
 - c.5. ¿Qué porcentaje de casas rurales están alejadas del centro urbano?
 - c.6. ¿Cuántas casas rurales disponen de conexión a internet?, ¿cuántas de ellas lo ofrecen de forma gratuita?
 - c.7. ¿Qué porcentaje de casas rurales no disponen de la Q de calidad? ¿Cuántas no tienen adquirido el compromiso de calidad turística?

A3.5. Análisis de una sola variable. Gráficos de caja.

Casas rurales en Bizkaia.

- d. Obtén los gráficos de caja de la variable precio medio empleando como variable de separación las variables discretas existentes.
- e. Interpreta los resultados.
 - ¿Qué dirías sobre la distribución de la variable precio medio?
 - ¿Existe evidencia muestral de la distribución de la que proviene la variable número de habitaciones?
 - ¿Cuál es la información que realmente aporta las distribuciones de frecuencias de las variables LOC y WIFI?
 - ¿Cuáles son las variables discretas que más diferencias logran en la variable precio medio?

A3.5. Análisis de una sola variable. Gráficos de caja.

Consumo de pollo.

Abre el fichero de datos `pollo.gdt` y analiza la información que te proporcionan los distintos gráficos de cajas.

- Gráfico de caja del consumo de pollo simple y empleando la variable $X6$ como variable de separación.
- Gráfico de caja de la variable renta.
- Gráfico de cajas conjunto de todas las variables precio.
- Gráfico de cajas individual de las variables precio empleando como variable de separación $X6$.
- Interpreta y relaciona todos los resultados obtenidos.