

Task T3

Data management in Gretl

Pilar González and Susan Orbe

Dpt. Applied Economics III (Econometrics and Statistics)

Contents

- 1 T3.1. Generating Gretl data files.
- 2 T3.2. Modifying Gretl data files.
- 3 T3.3. Data analysis using Gretl.
- 4 T3.4. Graphical analysis using Gretl.
- 5 T3.5. Univariate analysis.

Contents

- 1 T3.1. Generating Gretl data files.
- 2 T3.2. Modifying Gretl data files.
- 3 T3.3. Data analysis using Gretl.
- 4 T3.4. Graphical analysis using Gretl.
- 5 T3.5. Univariate analysis.

T3.1. Generating Gretl data files.

Beach umbrella rental.

A family firm wants to analyse the relationship between the volume of beach umbrellas rented (U , in units per week), the daily rental price that is constant throughout the week (P , en euros). The table below shows weekly data from the first week of May 2013 until the next to last week of September 2013:

t	1	2	3	4	5	6	7	8	9	10	11
U	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
U	394	424	442	462	491	437	422	433	393	429
P	9	10	10	12	15	15	12	10	10	9

T3.1. Generating Gretl data files.

Beach umbrella rental.

- a. Enter these data into Gretl manually.
- b. Add a description for each variable.
- c. Write down some information about the data set.
- d. Save these data in a Gretl file called `umbrellas.gdt`.

T3.1. Generating Gretl data files.

Registered vehicles.

The automobile industry wants to analyse the evolution of sales in the Basque Country in the last years. There is information available on these variables:

RV: number of registered vehicles in the Basque Country.

RVSP: number of registered vehicles in Spain.

BOP: Brent oil price (in dollars per barrel).

These data may be found in the Excel file `vehicles.xls`, in the sheet called `Registered vehicles`.

T3.1. Generating Gretl data files.

Registered vehicles.

- a. Import the data from the Excel data file into Gretl.
- b. Add a description for each variable.
- c. Save these data in a Gretl data file called `vehicles.gdt`.
- d. The data set consists of monthly time series from January 2004 to December 2012. Change the structure of the data in the file `vehicles.gdt` from cross-section to time series data.

T3.1. Generating Gretl data files.

Wages.

Load the data file `wage1.gdt` you may find in the sample folder Wooldridge (Wooldridge (2001)).

- a. Save this data file in your own folder.
- b. We are only interested in the variables `wage`, `educ`, `exper`, `tenure`, `nonwhite`, `female`, `married`, `numdep` and `smsa`. Delete the rest of the variables.
- c. Change the name and the attributes of the variables as follows:

<i>wage</i> - W	<i>educ</i> - ED	<i>exper</i> - EX
<i>nonwhite</i> - NW	<i>female</i> - F	<i>married</i> - M
<i>numdep</i> - ND	<i>tenure</i> - T	<i>smsa</i> - C: = 1 if live in a big city
- d. Save this file as `wages.gdt`.

T3.1. Generating Gretl data files.

Holiday cottages in Biscay.

The website nekatur.net provides a lot of information about holiday cottages in Biscay. The data file `HCBiscay.txt` contains data on the variables:

- average price of a room (RP, in euros)
- number of rooms (HB)
- breakfast price (BP, in euros)
- WiFi access (WIFI)
- location (LOC)
- proximity to a natural park (NP)
- proximity to a lake or a reservoir (LK)
- proximity to a beach (BE)
- Tourism Quality Q Certificate (Q)
- Commitment to quality tourism award (C)

T3.1. Generating Gretl data files.

Holiday cottages in Biscay.

The **codes** used in the file *HCBiscay.txt* for the variables WIFI, LOC, NP, LK, BE, Q and C are:

WIFI	<i>F</i> : free WiFi access <i>N</i> : no WiFi access <i>P</i> : WiFi access for an additional fee
LOC	<i>A</i> : far from the town center <i>C</i> : in the town center
NP, LK and BE	<i>V</i> : holiday cottage more than 20 km from the service (park, lake or beach) <i>R</i> : holiday cottage less than 1 km from the service (park, lake or beach) <i>N</i> : there is no service near the holiday cottage <i>P</i> : holiday cottage located in the service
Q and C	<i>N</i> : if the holiday cottage does not have the certificate or award <i>Y</i> : if the holiday cottage has got the certificate or award.

T3.1. Generating Gretl data files.

Holiday cottages in Biscay.

- a. Import the file `HCBiscay.txt` into Gretl, taking into account that it is a text file containing cross-section data, tabulator separated. The decimal is written as a point (.).
- b. Look at the coding Gretl applies to the qualitative variables.
- c. Add a description for each variable and verify that the qualitative variables have been defined as discrete variables.
- d. Save the data in the Gretl file `cottages.gdt`.

Contents

- 1 T3.1. Generating Gretl data files.
- 2 T3.2. Modifying Gretl data files.
- 3 T3.3. Data analysis using Gretl.
- 4 T3.4. Graphical analysis using Gretl.
- 5 T3.5. Univariate analysis.

T3.2. Modifying Gretl data files.

Beach umbrella rental. Add observations.

Searching in their old files, the manager of the family firm finds more observations on beach umbrella rental, in particular, some data for the last week of September 2013.

t	22
U	379
P	9

Enter these observations into the data file `umbrellas.gdt` manually.

T3.2. Modifying Gretl data files.

Beach umbrella rental. Add variables.

The family firm manager thinks that the volume of rentals depend as well on some **meteorological** conditions such as:

T : weekly average temperature (in degrees Celsius).

W : windy week.

Enter these variables into the data file `umbrellas.gdt` manually. The data are shown in the table below. With respect to the variable W , give the value 1 to windy weeks and the value 2 to non windy weeks.

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

T3.2. Modifying Gretl data files.

Registered vehicles.

The experts of the automobile industry consider that the evolution of sales may depend as well on the general economic situation in the Basque Country. The variable Industrial Production Index (IPI) will be used as a proxy for the economic climate.

The Excel file `vehicles.xls` (sheet IPI) contains time series data on:

IPI: Industrial Production Index in Spain (raw data).

IPIR: annual variation rate of the Spanish IPI.

IPIBC: Industrial Production Index in the Basque Country (seasonally adjusted data).

IPIBCR: annual variation rate of the Basque IPI.

T3.2. Modifying Gretl data files.

Registered vehicles.

- a. Import these data from the Excel file into Gretl.
- b. Add a description for each new variable.
- c. Save all the changes in the data file `vehicles.gdt`.

T3.2. Modifying Gretl data files.

Simulation.

Create a new data set in Gretl with these characteristics:

- cross-section data.
- sample size: 1000.

a. Simulate the variables:

u , from a normal distribution with zero mean and variance 9.

X_2 , from a normal distribution with mean 6 and variance 25.

X_3 , from a uniform distribution [minimum 16, maximum 65].

b. Simulate the variable $Y = 3 + 0.5 X_2 - 2.4 X_3 + u$.

c. Save the data in a Gretl file called `simulate.gdt`.

Contents

- 1 T3.1. Generating Gretl data files.
- 2 T3.2. Modifying Gretl data files.
- 3 T3.3. Data analysis using Gretl.**
- 4 T3.4. Graphical analysis using Gretl.
- 5 T3.5. Univariate analysis.

T3.3. Data analysis using Gretl.

Holiday cottages in Biscay.

- a. Compute the descriptive statistics for the quantitative variables included in the file `cottages.gdt`.

What is the range of the room price? What is the range of the number of rooms? What is the minimum and maximum price of breakfast?

- b. Compute the correlation matrix for these quantitative variables. Comment on the results obtained.

Wages.

- a. Compute the descriptive statistics for the variables `wage`, `education`, `experience` and `tenure` included in the file `wages.gdt`. Comment on the results.
- b. Compute the correlation matrix for the same variables. Comment on the results.

Contents

- 1 T3.1. Generating Gretl data files.
- 2 T3.2. Modifying Gretl data files.
- 3 T3.3. Data analysis using Gretl.
- 4 T3.4. Graphical analysis using Gretl.
- 5 T3.5. Univariate analysis.

T3.4. Time series graphs.

Beach umbrella rental.

Open the data file `umbrellas.gdt` and draw the time series plots of variables price and temperature.

- a. Draw both plots on a single graph.
- b. Interpret the results:

Do you observe an increasing or decreasing trend in the variables?

Do you observe a seasonal behaviour in the variables?

T3.4. Time series graphs.

Registered vehicles.

Open the data file `vehicles.gdt`:

- a. Graph the variables number of registered vehicles in the Basque Country, number of registered vehicles in Spain and Brent oil price:
 - on the same graph.
 - in multiple graphs.
- b. Graph the variables Spanish IPI, Basque IPI, annual variation rate of the Spanish IPI and annual variation rate of the Basque IPI on the same graph.
- c. Interpret the results:

Do you observe a trend in the variables?

Do you observe a seasonal behaviour in the variables?

T3.4. Scatterplots.

Beach umbrella rental.

Open the data file `umbrellas.gdt` and graph:

- Number of rented umbrellas against price.
- Number of rented umbrellas against temperature.
- Number of rented umbrellas against W .
- Interpret the results.

T3.4. Scatterplots.

Registered vehicles.

Open the data file `vehicles.gdt` and graph:

- Number of registered vehicles in the Basque Country against Brent oil price.
- Number of registered vehicles in the Basque Country against the Basque IPI.
- Number of registered vehicles in the Basque Country against the annual variation rate of the Basque IPI.
- Comment on the results obtained.

T3.4. Scatterplots.

Wages.

Open the data file `wages.gdt` and graph:

- a. Wages against education.
- b. Wages against experience.
- c. Wages against tenure.
- d. Experience against tenure.
- d. Experience against education.
- e. Education against tenure.
- f. Interpret the results.

T3.4. 3D Graphs.

Beach umbrella rental.

Open the data file `umbrellas.gdt` and draw a plot of the relationship among the variables number of rented umbrellas, price and temperature.

Registered vehicles.

Open the data file `vehicles.gdt` and draw a plot of the relationship among the variables:

- Number of registered vehicles in Spain against the price of Brent and the Spanish IPI.
- Number of registered vehicles in Spain against the price of Brent and the annual variation rate of the Spanish IPI.
- Comment on the results.

T3.4. 3D Graphs.

Wages.

Open the data file `wages.gdt` and graph:

- Wages against education and experience.
- Wages against experience and tenure.
- Wages against tenure and education.
- Experience against education and tenure.
- Comment on the results.

Contents

- 1 T3.1. Generating Gretl data files.
- 2 T3.2. Modifying Gretl data files.
- 3 T3.3. Data analysis using Gretl.
- 4 T3.4. Graphical analysis using Gretl.
- 5 T3.5. Univariate analysis.

T3.5. Univariate analysis. Distribution.

Beach umbrella rate.

Open the data file `umbrellas.gdt` and analyse the distribution of the variables number of rented umbrellas, price and temperature.

- Run normality tests on the three variables considered.
- Compute the frequency distribution of the variables number of rented umbrellas and temperature and run the tests against the normal and gamma distributions.
- Comment on the results.

Could you conclude that these variables come from a normal distribution?

Could you conclude that these variables come from a gamma distribution?

What characteristics do these frequency distributions show?

T3.5. Univariate analysis. Distribution.

Registered vehicles.

Open the data file `vehicles.gdt`. Consider the variables number of registered vehicles in the Basque Country and the number of registered vehicles in Spain.

- Estimate their density functions.
- Obtain their frequency distributions and run the test against the normal distribution.
- Draw the Q-Q plots.
- Comment on the results.

Are the characteristics of the two estimated density functions similar? What are the differences?

What characteristics do you observe in the frequency distributions obtained?

Do the information in the frequency distributions coincide with the one in the Q-Q plots?

T3.5. Univariate analysis. Distribution.

Holiday cottages in Biscay.

Open the data file `cottages.gdt`.

- a. Analyse the distribution of the variable number of rooms.
- b. Obtain the frequency distribution of the variable room price and run the tests against the normal and gamma distributions.
- c. Obtain the frequency distribution for the qualitative variables.
 - c.1. How many holiday cottages are less than 1 km from the beach?
 - c.2. How many holiday cottages are more than 20 km from a lake?
 - c.3. How many holiday cottages are located in a natural park?
 - c.4. How many holiday cottages are less than 1 km from a beach?, and from a natural park?
How many holiday cottages are more than 20 km from a beach?, and from a natural park?
 - c.5. What percentage of holiday cottages are far from the urban center?
 - c.6. How many holiday cottages offer WiFi access? How many holiday cottages offer free WiFi access?
 - c.7. What percentage of holiday cottages do not have the Q-quality certificate? How many holiday cottages have not received the Commitment to quality tourism award?

T3.5. Univariate analysis. Boxplot.

Holiday cottages in Biscay.

- d. Draw the factorized boxplots for the variable room price using all the discrete variables included in the file as factors.
- e. Comment on the results.

What would you say about the distribution of the variable room price?

Is there any evidence in the sample about what distribution the variable number of rooms comes from?

What is the information provided by the frequency distributions of the variables LOC and WIFI?

What are the discrete variables that generate more differences in the variable price of a room?

T3.5. Univariate analysis. Boxplot.

Chicken consumption.

Open the data file `chicken.gdt` and analyse the information provided by:

- Simple boxplot of the variable chicken consumption.
- Factorized boxplot of the variable chicken consumption using the variable `X6` as the factor.
- Boxplot of the variable income.
- Joint boxplot of the price variables.
- Individual factorized boxplot of the price variables using `X6` as the factor.
- Interpret and compare all the results obtained.