

Example 3.3

Analysis of data in Gretl: descriptive statistics

Pilar González and Susan Orbe

Dpt. Applied Economics III (Econometrics and Statistics)

Contents

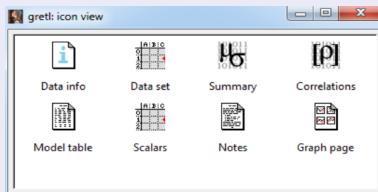
- 1 3.3.1. Gretl session.
- 2 3.3.2. Visitors to Bilbao.
- 3 3.3.3. Pizza consumption.

- 1 3.3.1. Gretl session.
- 2 3.3.2. Visitors to Bilbao.
- 3 3.3.3. Pizza consumption.

Example 3.3.1. Gretl session.

Descriptive statistics in a Gretl session.

Clicking on the icon *session icon view* in the main window Toolbar, you may see the objects included by default in a Gretl session.



Summary contains by default the table of descriptive statistics for all the variables in the data file.

Correlations contains by default the correlation matrix for all the variables in the data file.

Contents

1 3.3.1. Gretl session.

2 3.3.2. Visitors to Bilbao.

3 3.3.3. Pizza consumption.

Example 3.3.2. Visitors to Bilbao.

Questions.

Open the file `tourism.gdt`.

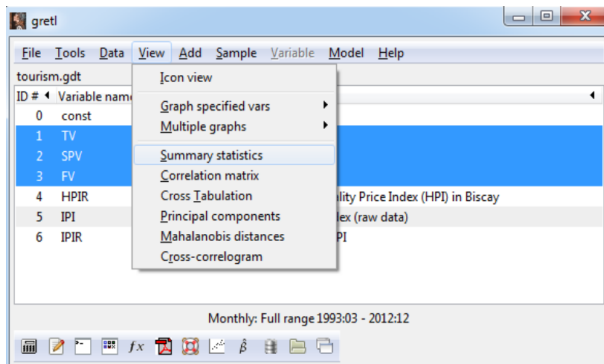
- Compute the descriptive statistics of the variables total visitors, visitors coming from Spain and foreign visitors.
- Save the output in MS format (Word).
- Compute the correlation matrix among total visitors, visitors coming from Spain and foreign visitors.
- Comment on the results.

Example 3.3.2. Visitors to Bilbao.

Summary statistics.

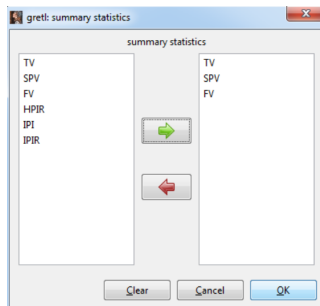
To obtain the **descriptive statistics** of one or more variables, go up to the Menu Bar and click

View -> Summary statistics



Example 3.3.2. Visitors to Bilbao.

Summary statistics.



Select the variables of interest

gretl: summary statistics

	Mean	Median	Minimum	Maximum
TV	39601.	38339.	12623.	82149.
SPV	26870.	26624.	10125.	43984.
FV	12731.	11790.	1518.0	38193.

	Std. Dev.	C.V.	Skewness	Ex. kurtosis
TV	15291.	0.38612	0.38006	-0.57204
SPV	8232.6	0.30638	0.089240	-0.93659
FV	7740.7	0.60802	0.83820	0.23915

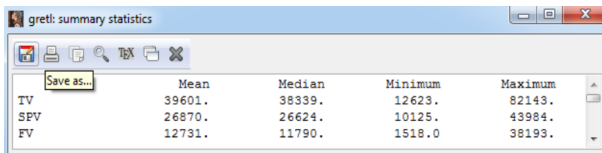
	5% perc.	95% perc.	IQ range	Missing obs.
TV	17807.	66488.	23889.	0
SPV	13682.	39977.	13760.	0
FV	3038.1	27628.	11125.	0

Output

Example 3.3.2. Visitors to Bilbao.

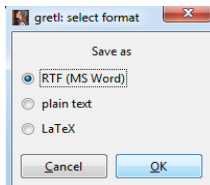
Summary statistics.

To **save** the output in Word format, go up to the menu bar in the **summary statistics** window and click on the left icon (*Save as ...*).



	Mean	Median	Minimum	Maximum
TV	39601.	38339.	12623.	82143.
SPV	26870.	26624.	10125.	43984.
FV	12731.	11790.	1518.0	38193.

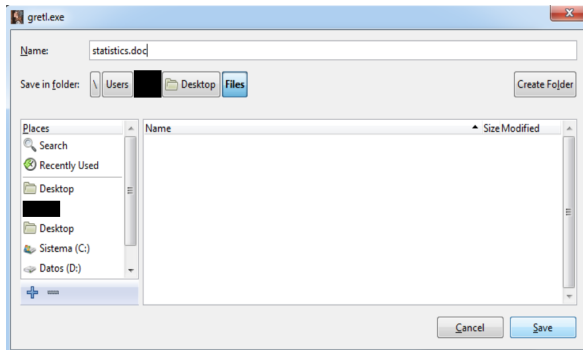
Select the format RTF (MS Word) in the dialog box.



Example 3.3.2. Visitors to Bilbao.

Summary statistics.

Save this file in your own folder (in this case, in the Desktop).

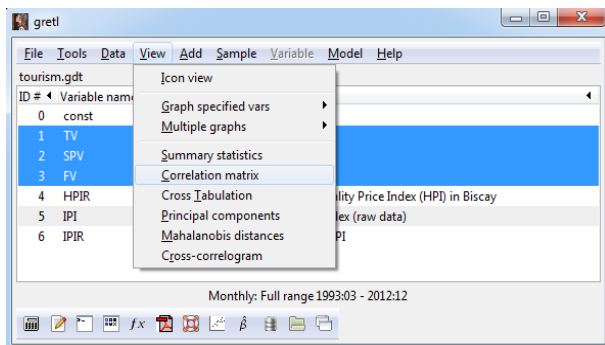


Example 3.3.2. Visitors to Bilbao.

Correlation matrix.

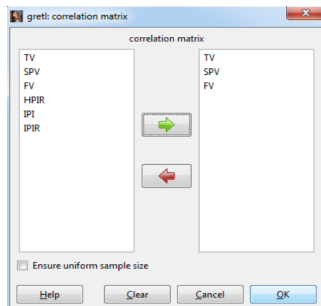
To obtain the **correlation matrix** for two or more variables, go up to the Menu Bar and click

View -> Correlation matrix

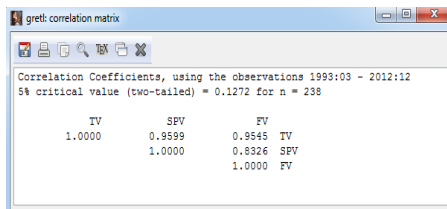


Example 3.3.2. Visitors to Bilbao.

Correlation matrix.



Select the variables of interest



Output

Example 3.3.2. Visitors to Bilbao.

Results.

- The sample mean of total visitors is 39601, while the sample mean of Spanish visitors is 26870 and the sample mean of foreign visitors 12731. Therefore, on average, the flow of foreign visitors to Bilbao is less than half of the flow of Spanish visitors.
- Given the values of the coefficients of variation and the IQ rank, it may be concluded that the sample variability of the variable foreign visitors is much higher (almost double) than the variability of Spanish visitors.
- As expected, the correlation coefficients are quite high, close to 1.

Contents

- 1 3.3.1. Gretl session.
- 2 3.3.2. Visitors to Bilbao.
- 3 3.3.3. Pizza consumption.

Example 3.3.3. Pizza consumption.

Questions.

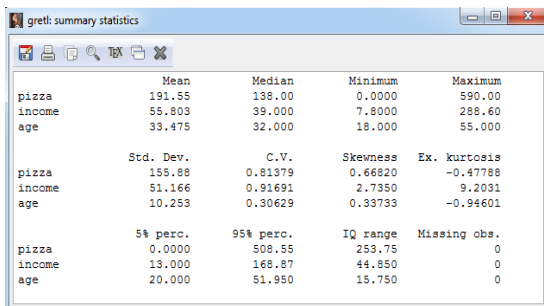
Open the file `pizza.gdt` that contains information on pizza consumption and some characteristics of the individuals in the sample.

- Compute the descriptive statistics of the variables consumption, income and age.
- Compute the correlation matrix among consumption, income and age.
- Save the correlation matrix in a Latex file.
- Comment on the results.

Example 3.3.3. Pizza consumption.

Summary statistics.

Following the steps explained above, you may obtain the **descriptive statistics** of the variables consumption, income and age.



The screenshot shows a window titled 'gretl: summary statistics'. It contains two tables of descriptive statistics for the variables 'pizza', 'income', and 'age'.

	Mean	Median	Minimum	Maximum
pizza	191.55	138.00	0.0000	590.00
income	55.803	39.000	7.8000	288.60
age	33.475	32.000	18.000	55.000

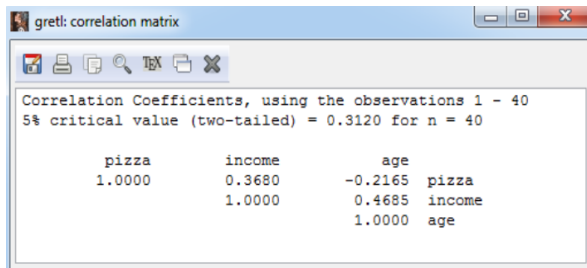
	Std. Dev.	C.V.	Skewness	Ex. kurtosis
pizza	155.88	0.81379	0.66820	-0.47788
income	51.166	0.91691	2.7350	9.2031
age	10.253	0.30629	0.33733	-0.94601

	5% perc.	95% perc.	IQ range	Missing obs.
pizza	0.0000	508.55	253.75	0
income	13.000	168.87	44.850	0
age	20.000	51.950	15.750	0

Example 3.3.3. Pizza consumption.

Correlation matrix.

Following the steps explained above, you may obtain the **correlation matrix** among consumption, income and age.



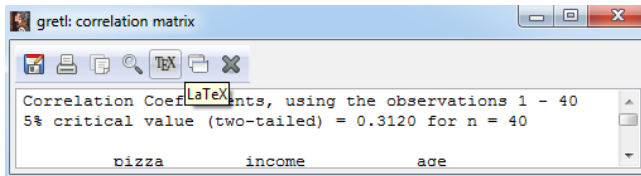
Example 3.3.3. Pizza consumption.

The menus at the top of the windows **gretl: summary statistics** and **gretl: correlation matrix** enable us to:

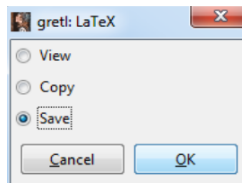
- **Save** the results in several formats: RTF(MS Word), plain text and LaTeX.
- **Print** the results.
- **Copy** the results in several formats (RTF(MS Word), plain text and LaTeX).
- **Work with the output in LaTeX format.** Thus, you can *View* the output in LaTeX format, *Copy* the output including the LaTeX commands to paste it to another file, *Save* the output including the LaTeX commands in a .tex file.

Example 3.3.3. Pizza consumption.

To save the correlation matrix in [LaTeX format](#), click on the icon *TeX* in the **correlation matrix** window.

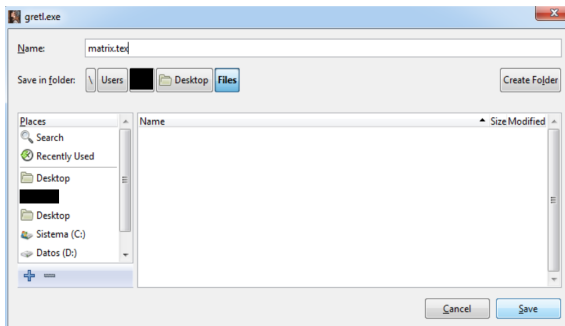


Mark Save in the dialog box.



Example 3.3.3. Pizza consumption.

Save this file in your folder (in this case, in the Desktop).



Example 3.3.3. Pizza consumption.

Results.

- The sample mean of pizza consumption is \$191.55, the sample mean of income is 55.803 thousands of dollars and the sample mean of age is 33.475 years.
- Given the results obtained for the standard deviation, the coefficient of variation and the IQ rank, it may be concluded that the sample variability of the variables consumption and income is quite high.
- The correlation between consumption and income is positive, while the correlation between consumption and age is negative. That is, the larger the income, the larger the level of consumption but the older the client, the lower the level of consumption.
- The correlation between income and age is positive, that is, the older the richer.