Solution to Task T3.

Data management in Gretl.

Task T3.1. Generating Gretl data files

Beach umbrella rental

a. Enter data into Gretl manually.

`File --> New data set`

Number of observations: 21

Structure of data set: time series

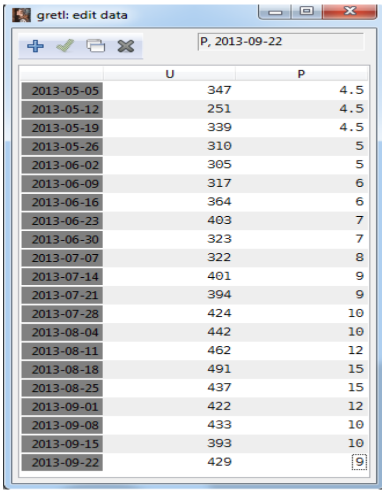
Daily date to represent week: Sunday

Time series frequency: weekly

Starting observation: 2013-05-05 (first week of may 2013)

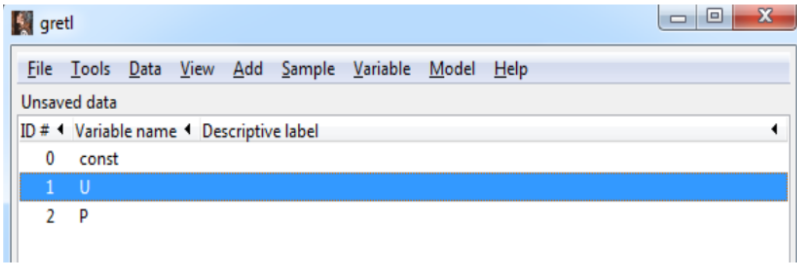
Select: start entering data values.

Give a name to each variable and start entering data values manually.



	U	P
2013-05-05	347	4.5
2013-05-12	251	4.5
2013-05-19	339	4.5
2013-05-26	310	5
2013-06-02	305	5
2013-06-09	317	6
2013-06-16	364	6
2013-06-23	403	7
2013-06-30	323	7
2013-07-07	322	8
2013-07-14	401	9
2013-07-21	394	9
2013-07-28	424	10
2013-08-04	442	10
2013-08-11	462	12
2013-08-18	491	15
2013-08-25	437	15
2013-09-01	422	12
2013-09-08	433	10
2013-09-15	393	10
2013-09-22	429	9

The main window shows variables U and P.

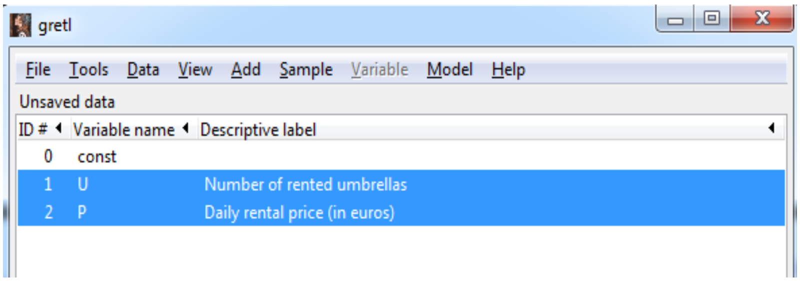


Unsaved data

ID #	Variable name	Descriptive label
0	const	
1	U	
2	P	

b. Add a description for each variable.

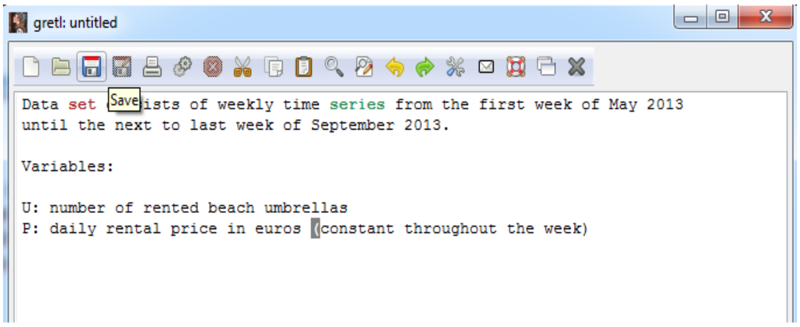
Highlight the variables in the main window, right-click and select the *Edit attributes* option from the pulldown menu. Write down the description for each variable.



ID #	Variable name	Descriptive label
0	const	
1	U	Number of rented umbrellas
2	P	Daily rental price (in euros)

c. Write down some notes on the data set.

Click on the *Session icon view* icon of the Toolbar. Then, click on the *Data info* icon and write down the text shown in the window below.



```
Data set consists of weekly time series from the first week of May 2013
until the next to last week of September 2013.

Variables:

U: number of rented beach umbrellas
P: daily rental price in euros (constant throughout the week)
```

Before quitting click on the *Save* icon.

d. Save the data file.

`File --> Save data as ...`

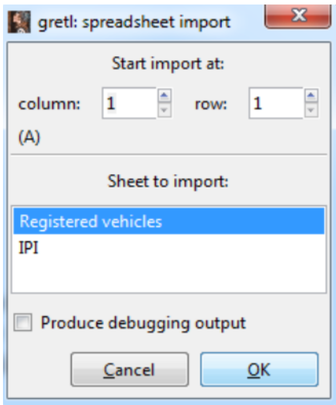
Select the folder where you want to save the data and use the name `umbrellas.gdt`.

Registered vehicles

a. Read data from an Excel file.

`File --> Open data --> User file`

Go to the folder where the Excel data file (`vehicles.xls`) is. Load this file into Gretl. In the dialog box, select the sheet, and the row and column where the data start.



gretl: spreadsheet import

Start import at:

column: 1 row: 1

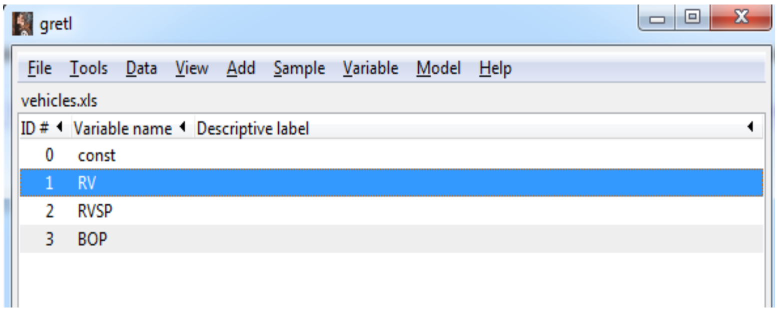
(A)

Sheet to import:

Registered vehicles
IPI

□ Produce debugging output

The three imported variables appear in the main window.

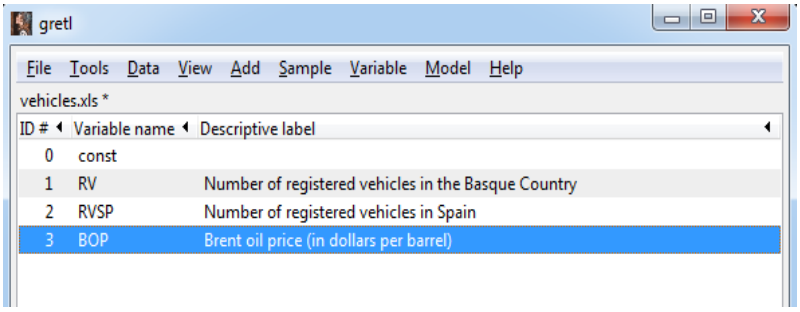


vehicles.xls

ID #	Variable name	Descriptive label
0	const	
1	RV	
2	RVSP	
3	BOP	

b. Add a description for each variable.

Highlight the variables in the main window, right-click and select the *Edit attributes* option. Write down the description for each variable.



vehicles.xls *

ID #	Variable name	Descriptive label
0	const	
1	RV	Number of registered vehicles in the Basque Country
2	RVSP	Number of registered vehicles in Spain
3	BOP	Brent oil price (in dollars per barrel)

c. Save the data file.

```
File --> Save data as ...
```

Go to the folder where you want to save the data and use the name `vehicles.gdt`.

d. Change the data set structure.

```
Data --> Dataset structure
```

Then, mark the Time series option and give the time series frequency and the starting observation.

DON'T forget to save the changes to this data file

Wages

a. Save the data file in your folder.

```
File --> Open data  --> Sample file ...
```

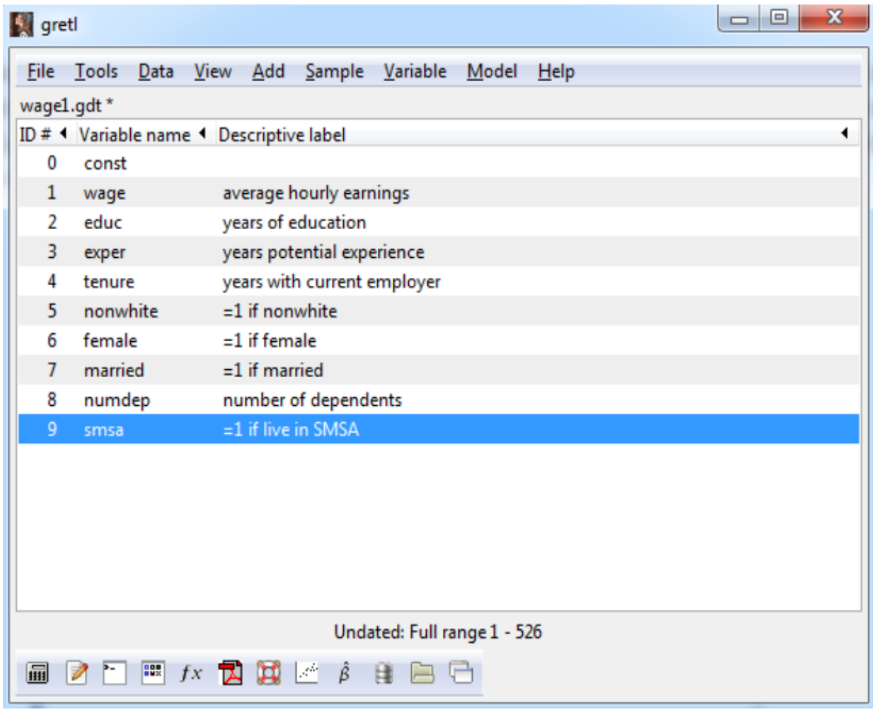
Choose the *Wooldridge* tab and load the file called `wage1`.

Save this data file in your own folder.

```
File --> Save data as ...
```

b. Delete the variables you are not interested in.

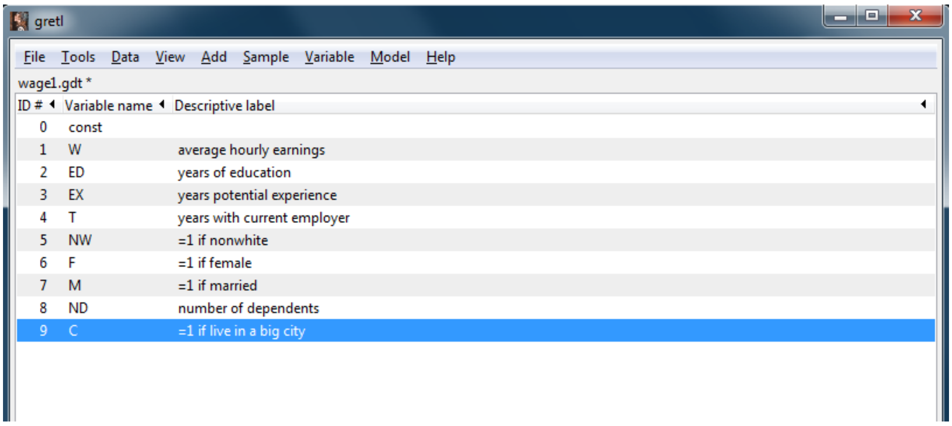
Highlight all the variables you are NOT interested in using the cursor. Right-click and select the *Delete* option from the pulldown menu. After having done this, the main window looks as follows:


gretl
File Tools Data View Add Sample Variable Model Help
wage1.gdt *
Undated: Full range 1 - 526

ID #	Variable name	Descriptive label
0	const	
1	wage	average hourly earnings
2	educ	years of education
3	exper	years potential experience
4	tenure	years with current employer
5	nonwhite	=1 if nonwhite
6	female	=1 if female
7	married	=1 if married
8	numdep	number of dependents
9	smsa	=1 if live in SMSA

c. Change the name and the attributes of the variables.

Highlight each variable, right-click and select the option *Edit attributes*. Then, write down the new names and attributes.



ID #	Variable name	Descriptive label
0	const	
1	W	average hourly earnings
2	ED	years of education
3	EX	years potential experience
4	T	years with current employer
5	NW	=1 if nonwhite
6	F	=1 if female
7	M	=1 if married
8	ND	number of dependents
9	C	=1 if live in a big city

d. Do NOT forget to save all the changes in a data file called `wages.gdt`.

```
File --> Save data as ...
```

Holiday cottages in Biscay

a and b. Load the data file `HCBiscay.txt` and check the codes that Gretl assigns to the qualitative variables.

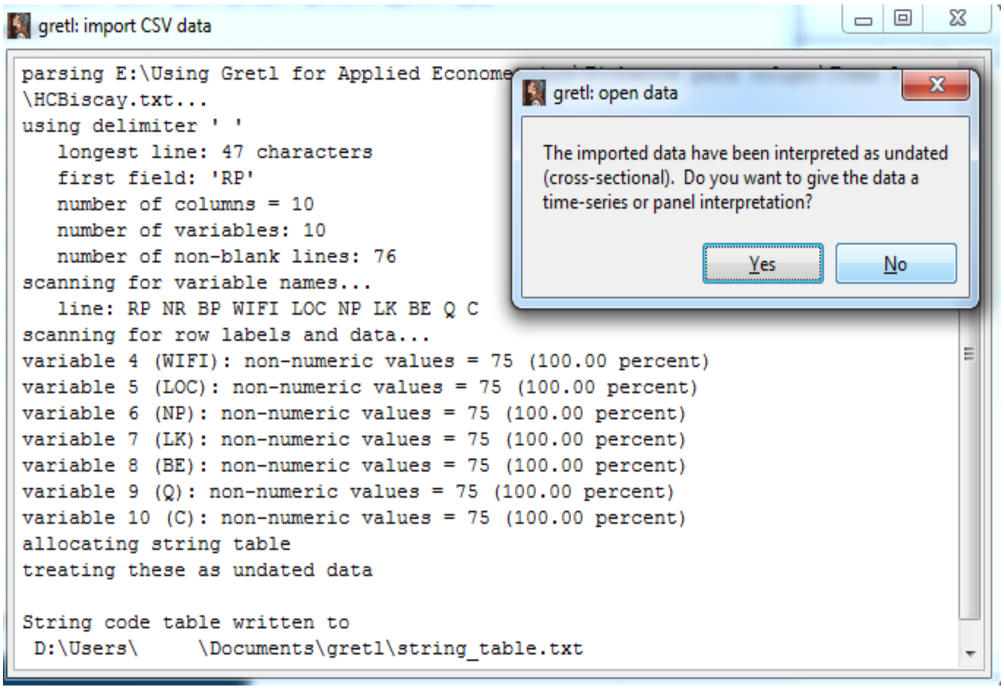
```
File --> Open data    --> User file
```

Look for the folder where the data file `HCBiscay.txt` is and load the data file.

When a data file in text format is loaded, Gretl generates two windows that provide all the data information.

The window **gretl: import CSV data** shows:

- The name of the data file imported: HCBiscay.txt.
- The number of observations: 76
- The qualitative variables with non-numeric values: WIFI, LOC, NP, LK, BE, Q and C.
- At the bottom, you can find the name of the file where the codes for the qualitative variables, that is, the numeric values assigned by Gretl to the categories of the qualitative variables are: string_table.txt.

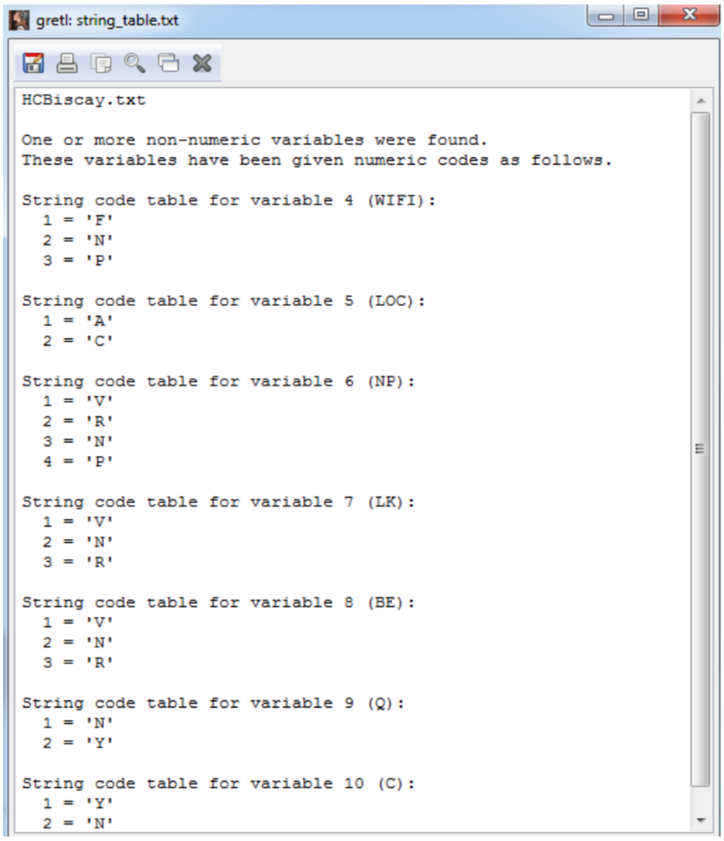


```
parsing E:\Using Gretl for Applied Econome
\HCBiscay.txt...
using delimiter ' '
   longest line: 47 characters
   first field: 'RP'
   number of columns = 10
   number of variables: 10
   number of non-blank lines: 76
scanning for variable names...
   line: RP NR BP WIFI LOC NP LK BE Q C
scanning for row labels and data...
variable 4 (WIFI): non-numeric values = 75 (100.00 percent)
variable 5 (LOC): non-numeric values = 75 (100.00 percent)
variable 6 (NP): non-numeric values = 75 (100.00 percent)
variable 7 (LK): non-numeric values = 75 (100.00 percent)
variable 8 (BE): non-numeric values = 75 (100.00 percent)
variable 9 (Q): non-numeric values = 75 (100.00 percent)
variable 10 (C): non-numeric values = 75 (100.00 percent)
allocating string table
treating these as undated data

String code table written to
 D:\Users\     \Documents\gretl\string_table.txt
```

The imported data have been interpreted as undated (cross-sectional). Do you want to give the data a time-series or panel interpretation?

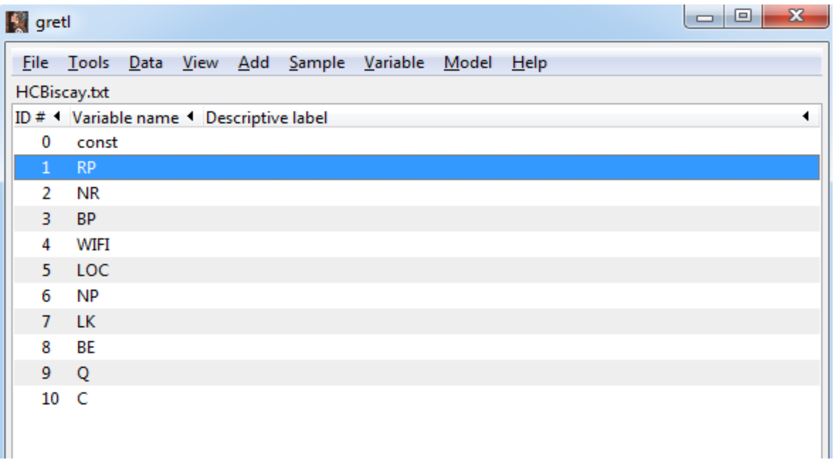
The window **gretl: string_table.txt** shows the codes assigned to the categories of the qualitative variables. For instance, the label A assigned in the text file `HCBiscay.txt` to a holiday cottage far from the town center is transformed into the number 1 in the Gretl format file.

Note that the same numerical value is not always assigned to the same label. For instance, the label R becomes the numerical value 2 for the qualitative variable NP, and the numerical value 3 for LK. This is due to the fact that Gretl always assigns the numerical value 1 to the first label that detects for each variable.



```
HCBiscay.txt

One or more non-numeric variables were found.
These variables have been given numeric codes as follows.

String code table for variable 4 (WIFI):
   1 = 'F'
   2 = 'N'
   3 = 'P'

String code table for variable 5 (LOC):
   1 = 'A'
   2 = 'C'

String code table for variable 6 (NP):
   1 = 'V'
   2 = 'R'
   3 = 'N'
   4 = 'P'

String code table for variable 7 (LK):
   1 = 'V'
   2 = 'N'
   3 = 'R'

String code table for variable 8 (BE):
   1 = 'V'
   2 = 'N'
   3 = 'R'

String code table for variable 9 (Q):
   1 = 'N'
   2 = 'Y'

String code table for variable 10 (C):
   1 = 'Y'
   2 = 'N'
```

All the imported variables, quantitative and qualitative, are listed in the main window of Gretl.



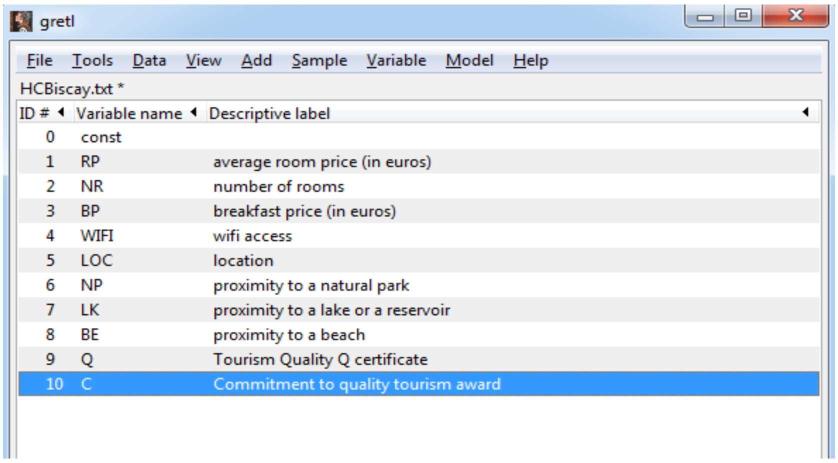
ID #	Variable name	Descriptive label
0	const	
1	RP	
2	NR	
3	BP	
4	WIFI	
5	LOC	
6	NP	
7	LK	
8	BE	
9	Q	
10	C	

Highlight ONE of the variables, right-click and select the option *Display values* from the pulldown menu to see the codes used in the initial file `HCBiscay.txt`.

Highlight ALL the variables, right-click and select the option *Display values* to see the numerical values assigned by Gretl.

c. Add a description for each variable and verify that all the qualitative variables have been defined as discrete variables.

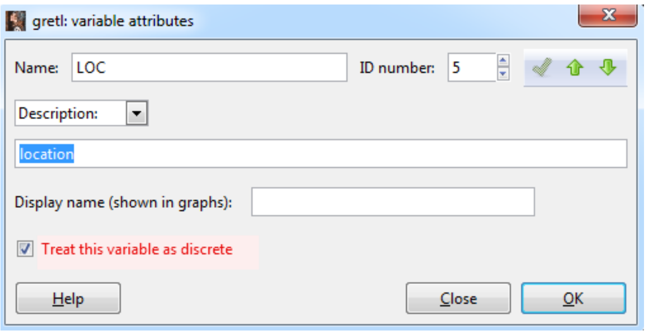
Highlight each variable in the main window, right-click and select the *Edit attributes* option. Write down the description of the variable.



HCBiscay.txt *

ID #	Variable name	Descriptive label
0	const	
1	RP	average room price (in euros)
2	NR	number of rooms
3	BP	breakfast price (in euros)
4	WIFI	wifi access
5	LOC	location
6	NP	proximity to a natural park
7	LK	proximity to a lake or a reservoir
8	BE	proximity to a beach
9	Q	Tourism Quality Q certificate
10	C	Commitment to quality tourism award

Let's verify whether the qualitative variable *LOC* has been defined as a discrete variable. Highlight the variable *LOC* using the cursor, right-click and select the option *Edit attributes*. As you may see in the window below, the option Treat this variable as discrete is on.



Name: LOC ID number: 5

Description:

location

Display name (shown in graphs):

☑ Treat this variable as discrete

Repeat this commands for the rest of the qualitative variables.

d. Do NOT forget to save all the changes in a data file called `cottages.gdt`.

Task T3.2. Modifying Gretl data files

Beach umbrella rental

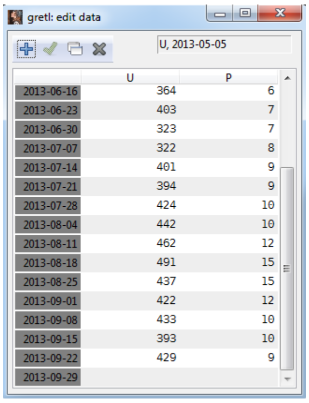
I. Add new observations manually. Click

```
Data --> Add observations ...
```

and write down the number of observations to add. Then, click

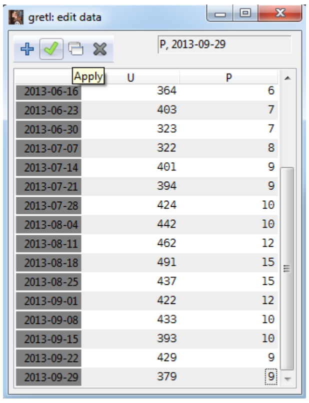
```
Data --> Edit values
```

and write down the new values in the empty row.



U, 2013-05-05

	U	P
2013-06-16	364	6
2013-06-23	403	7
2013-06-30	323	7
2013-07-07	322	8
2013-07-14	401	9
2013-07-21	394	9
2013-07-28	424	10
2013-08-04	442	10
2013-08-11	462	12
2013-08-18	491	15
2013-08-25	437	15
2013-09-01	422	12
2013-09-08	433	10
2013-09-15	393	10
2013-09-22	429	9
2013-09-29		



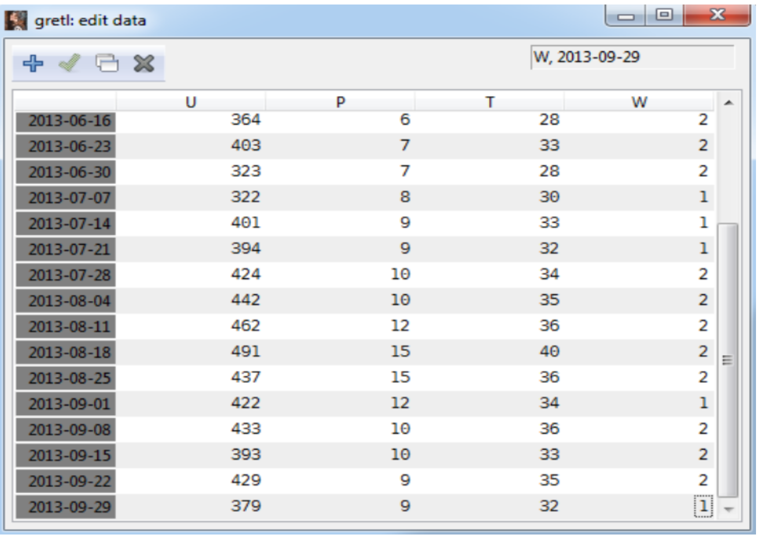
P, 2013-09-29

	U	P
2013-06-16	364	6
2013-06-23	403	7
2013-06-30	323	7
2013-07-07	322	8
2013-07-14	401	9
2013-07-21	394	9
2013-07-28	424	10
2013-08-04	442	10
2013-08-11	462	12
2013-08-18	491	15
2013-08-25	437	15
2013-09-01	422	12
2013-09-08	433	10
2013-09-15	393	10
2013-09-22	429	9
2013-09-29	379	9

II. Add variables manually.

```
Data --> Edit values
```

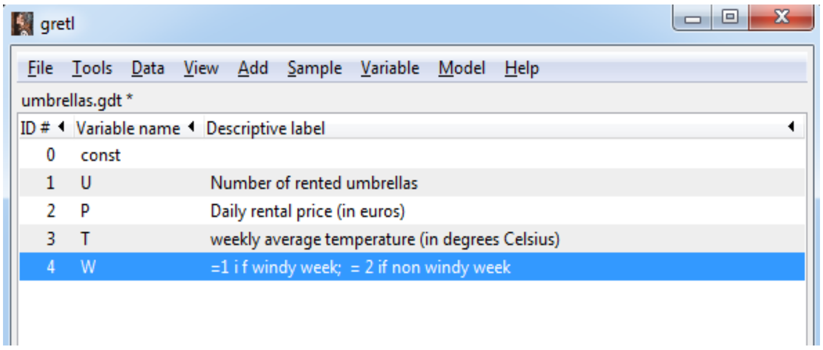
Click on the symbol + , select *Add variable* and write down the sample values for the two variables.



W, 2013-09-29

	U	P	T	W
2013-06-16	364	6	28	2
2013-06-23	403	7	33	2
2013-06-30	323	7	28	2
2013-07-07	322	8	30	1
2013-07-14	401	9	33	1
2013-07-21	394	9	32	1
2013-07-28	424	10	34	2
2013-08-04	442	10	35	2
2013-08-11	462	12	36	2
2013-08-18	491	15	40	2
2013-08-25	437	15	36	2
2013-09-01	422	12	34	1
2013-09-08	433	10	36	2
2013-09-15	393	10	33	2
2013-09-22	429	9	35	2
2013-09-29	379	9	32	1

Follow the procedure explained in Task T3.1 to edit the attributes for the two new varaibles.



umbrellas.gdt *

ID #	Variable name	Descriptive label
0	const	
1	U	Number of rented umbrellas
2	P	Daily rental price (in euros)
3	T	weekly average temperature (in degrees Celsius)
4	W	=1 i f windy week; = 2 if non windy week

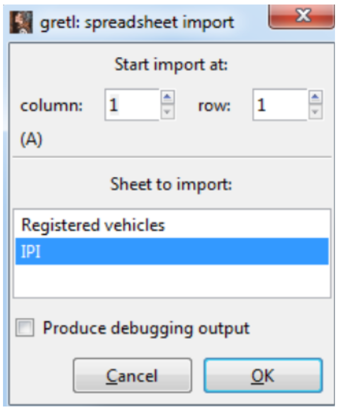
Do NOT forget to save all the changes into your data file.

Registered vehicles

a. Import new data.

```
File --> Append data   --> Excel
```

Open the folder where the Excel file is and load it. In the dialog box, select the sheet where the new data are and the row and column where the data start.



Start import at:

column: 1 row: 1

(A)

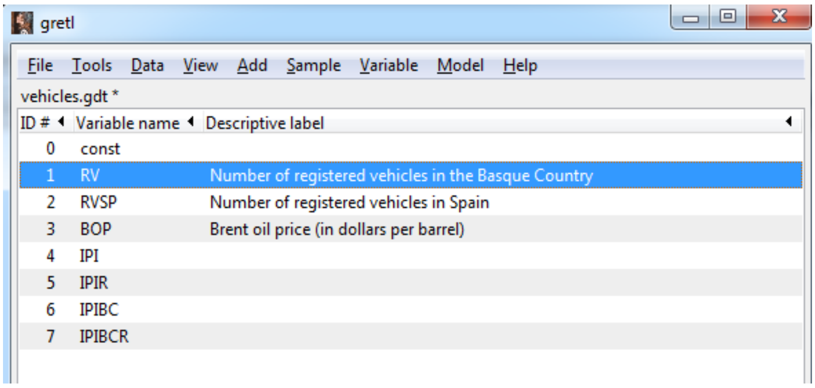
Sheet to import:

Registered vehicles

IPI

□ Produce debugging output

The four imported variables are listed in the main window.

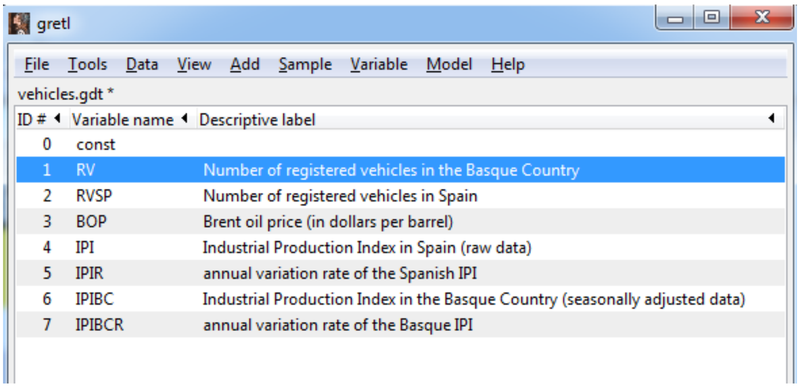


vehicles.gdt *

ID #	Variable name	Descriptive label
0	const	
1	RV	Number of registered vehicles in the Basque Country
2	RVSP	Number of registered vehicles in Spain
3	BOP	Brent oil price (in dollars per barrel)
4	IPI	
5	IPIR	
6	IPIBC	
7	IPIBCR	

b. Add a description for each new variable.

Highlight each variable using the cursor, right-click and select the option *Edit attributes*. Write down the description of each variable.



vehicles.gdt *

ID #	Variable name	Descriptive label
0	const	
1	RV	Number of registered vehicles in the Basque Country
2	RVSP	Number of registered vehicles in Spain
3	BOP	Brent oil price (in dollars per barrel)
4	IPI	Industrial Production Index in Spain (raw data)
5	IPIR	annual variation rate of the Spanish IPI
6	IPIBC	Industrial Production Index in the Basque Country (seasonally adjusted data)
7	IPIBCR	annual variation rate of the Basque IPI

c. Do NOT forget to save all the changes into the data file `vehicles.gdt`.

Simulation

To create a new data set, click

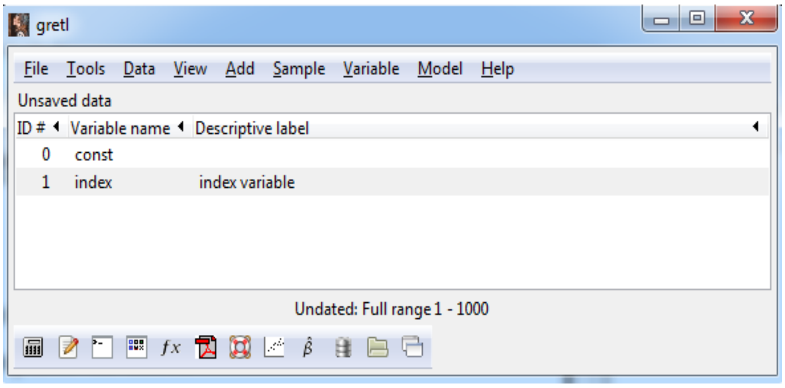
`File --> New data set`

Number of observations: 1000

Structure of data set: cross-sectional

DO NOT mark start entering data values, because the data are going to be simulated.

Only the variable index is shown in the main window.



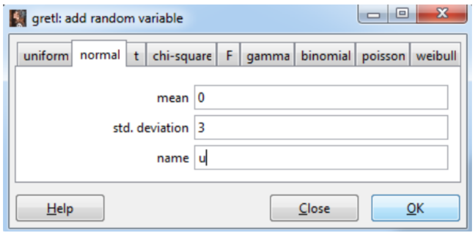
Unsaved data

ID #	Variable name	Descriptive label
0	const	
1	index	index variable

Undated: Full range 1 - 1000

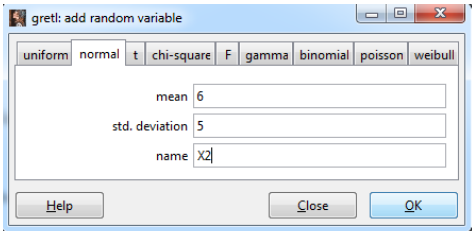
a y b. Simulate the variables:

`Add -- > Random variable ...`



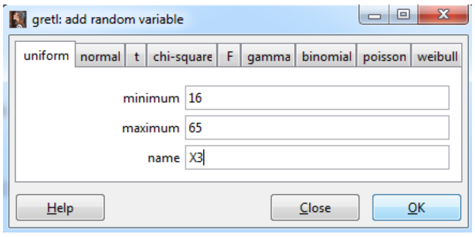
mean 0
std. deviation 3
name u

N(0, 9)



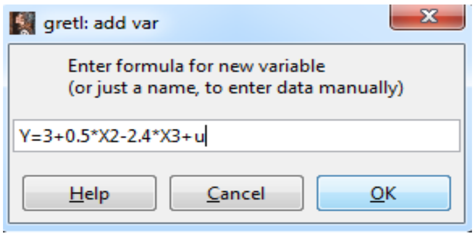
mean 6
std. deviation 5
name X2

N(6, 25)



minimum 16
maximum 65
name X3

U(16, 65)

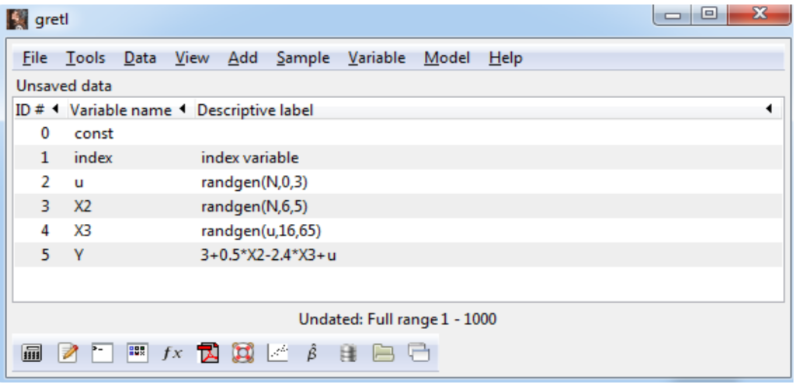


Enter formula for new variable
(or just a name, to enter data manually)

```
Y=3+0.5*X2-2.4*X3+u
```

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

Now, all the generated variables are listed in the main window.



Unsaved data

ID #	Variable name	Descriptive label
0	const	
1	index	index variable
2	u	randgen(N,0,3)
3	X2	randgen(N,6,5)
4	X3	randgen(u,16,65)
5	Y	3+0.5*X2-2.4*X3+u

Undated: Full range 1 - 1000

d. Do NOT forget to save all the changes in a data file called `simulation.gdt`.

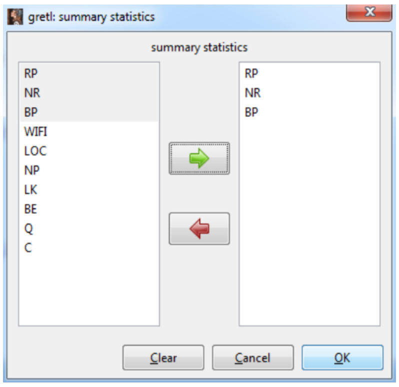
Task T3.3. Data analysis using Gretl

Holiday cottages in Biscay

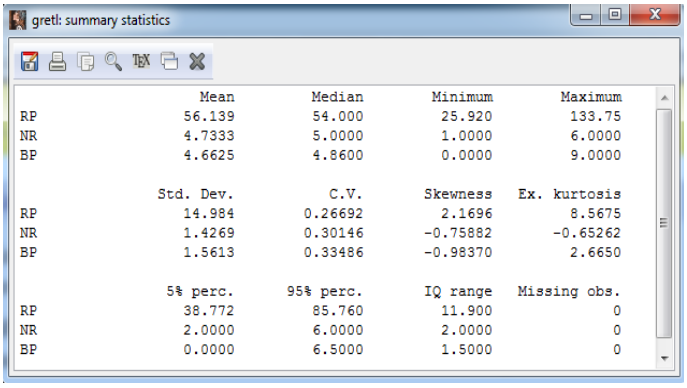
a. Descriptive statistics.

`View --> Summary statistics`

Select the quantitative variables: price of a room, number of rooms and price of breakfast.



The results are shown in the window below.



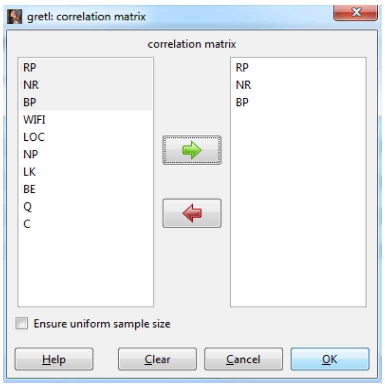
```
                 Mean         Median        Minimum        Maximum
RP             56.139         54.000         25.920         133.75
NR             4.7333         5.0000         1.0000         6.0000
BP             4.6625         4.8600         0.0000         9.0000

            Std. Dev.           C.V.       Skewness   Ex. kurtosis
RP             14.984        0.26692         2.1696         8.5675
NR             1.4269        0.30146       -0.75882       -0.65262
BP             1.5613        0.33486       -0.98370         2.6650

             5% perc.      95% perc.       IQ range   Missing obs.
RP             38.772         85.760         11.900              0
NR             2.0000         6.0000         2.0000              0
BP             0.0000         6.5000         1.5000              0
```

- Range of the average price of a room: 133.75 - 25.92 = 107.83.
- Range of the number of rooms: 6 - 1 = 5.
- Minimum price of breakfast: €0.
- Maximum price of breakfast: €9.

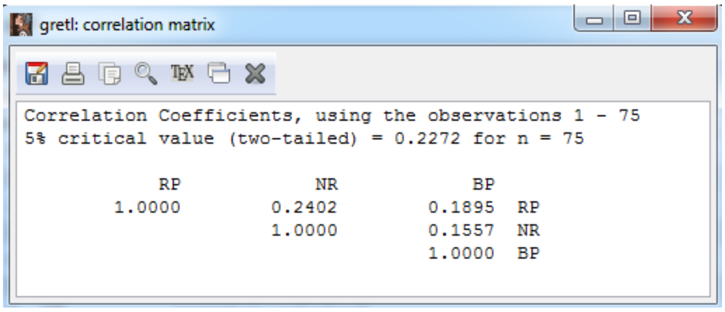
b. Correlation matrix.

`View --> Correlation matrix`

Select the quantitative variables: price of a room, number of rooms and price of breakfast.



The output window shows that the pairwise correlation between the variables is positive but quite small.



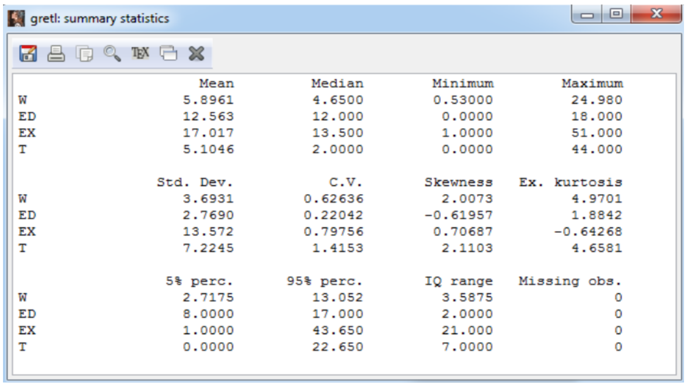
```
Correlation Coefficients, using the observations 1 - 75
5% critical value (two-tailed) = 0.2272 for n = 75

             RP            NR            BP
         1.0000        0.2402        0.1895  RP
                       1.0000        0.1557  NR
                                     1.0000  BP
```

Wages

a. Descriptive statistics.

```
View --> Summary statistics
```

Select the variables wage, education, experience and tenure. The results appear in the window below.



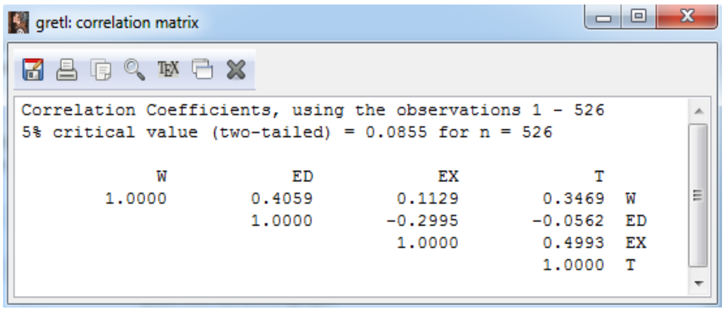
```
                   Mean         Median        Minimum        Maximum
W                5.8961         4.6500        0.53000         24.980
ED               12.563         12.000         0.0000         18.000
EX               17.017         13.500         1.0000         51.000
T                5.1046         2.0000         0.0000         44.000

              Std. Dev.           C.V.       Skewness   Ex. kurtosis
W                3.6931        0.62636         2.0073         4.9701
ED               2.7690        0.22042       -0.61957         1.8842
EX               13.572        0.79756        0.70687       -0.64268
T                7.2245         1.4153         2.1103         4.6581

               5% perc.      95% perc.       IQ range   Missing obs.
W                2.7175         13.052         3.5875              0
ED               8.0000         17.000         2.0000              0
EX               1.0000         43.650         21.000              0
T                0.0000         22.650         7.0000              0
```

This table shows the main descriptive statistics for the selected variables. Note that the variables that show more variability are experience and tenure.

b. Correlation matrix.

`View --> Correlation matrix`

Select the variables wage, education, experience and tenure. The results are shown in the window below.



```
Correlation Coefficients, using the observations 1 - 526
5% critical value (two-tailed) = 0.0855 for n = 526

             W            ED            EX             T
        1.0000        0.4059        0.1129        0.3469  W
                      1.0000       -0.2995       -0.0562  ED
                                    1.0000        0.4993  EX
                                                  1.0000  T
```

The variable wage is positively correlated with the variables education, experience and tenure. The correlation coefficient between wage and experience is quite small.

Furthermore, education is negatively correlated with experience, the correlation coefficient between education and tenure is close to zero, and the correlation between experience and tenure is strong and positive.

Task T3.4. Graphical analysis using Gretl.

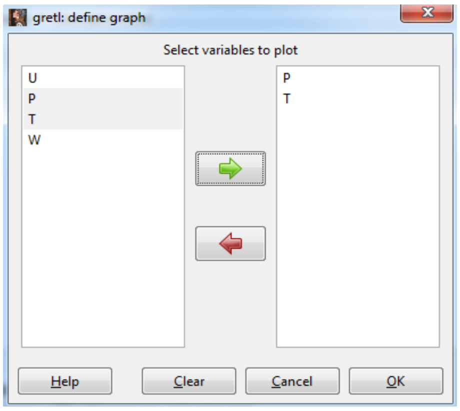
Time series graphs.

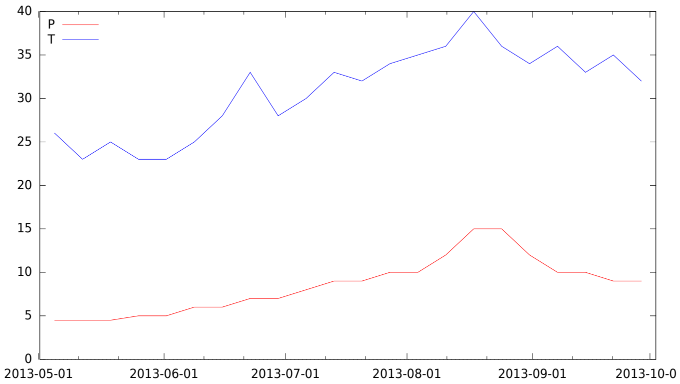
Beach umbrella rental

To draw several time series on a single graph, click

```
View --> Graph specified vars    --> Time series plot ...
```

Select the variables price and temperature.





It may be observed that:

- Both series show an increasing trend until the middle of August when they start decreasing until the end of the sample.
- The series do not present seasonal behaviour.

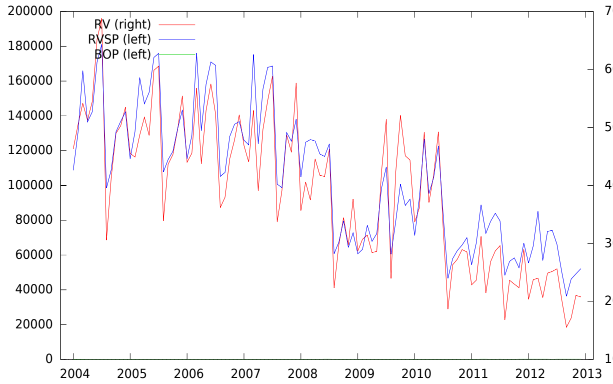
Registered vehicles

a. Vehicles registered and price of Brent.

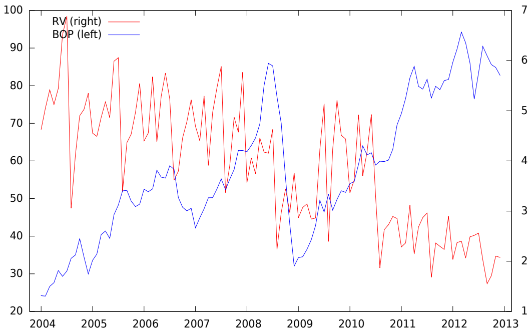
To draw several time series on a single graph, click

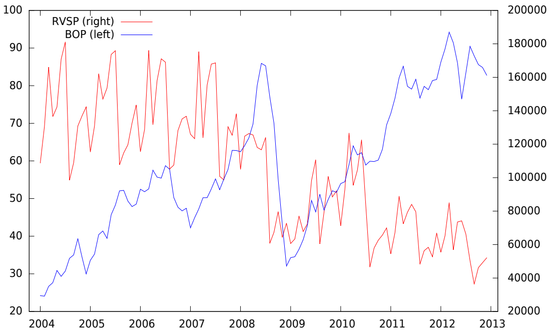
```
View --> Graph specified vars    --> Time series plot ...
```

Select the variables of interest.



It is not possible to see the series BOP in this graph because the scale of BOP is much smaller than the scale of the number of vehicles registered in Spain and in the Basque Country. Therefore, in this case, it is better to draw two plots: one of RV and BOP and another one of $RVSP$ and BOP

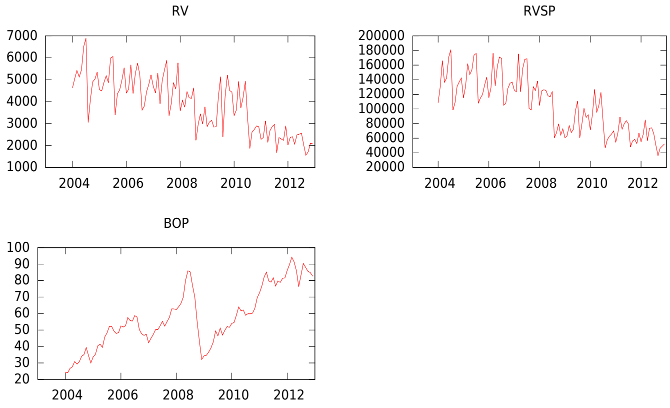




To draw several time series in separate small graphs, click

```
View --> Multiple graphs    --> Time series ...
```

and select the variables of interest.

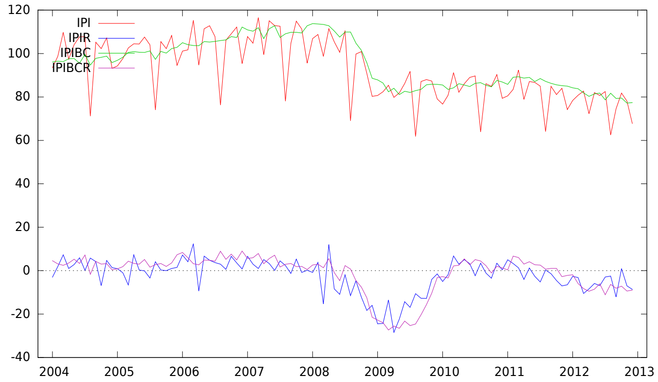


b. Industrial Production Index.

To draw the series related to IPI on a single graph, click

```
View --> Graph specified vars    --> Time series plot ...
```

and select the variables of interest.



c. Results.

- Trend. The number of registered vehicles series show both a decreasing trend while the trend of the price of Brent series is positive. The IPI series show a slightly increasing trend up to 2008 when they start decreasing reaching negative values. The recession suffered from 2008 to 2010 may be clearly seen in the graph. The annual variation rates turn to be positive at the beginning of 2010, although they became negative again in 2012.

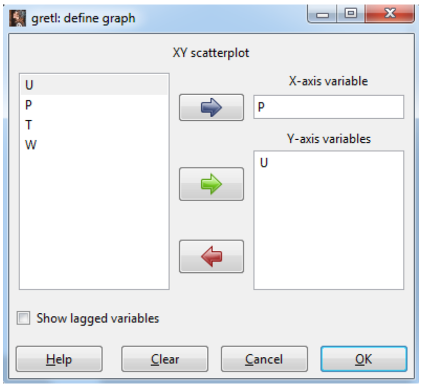
- Seasonality. The series of registered vehicles (both in Spain and in the Basque Country) and the Spanish IPI show seasonal behaviour. It was not expected to observe a seasonal component neither in the IPI of the Basque Country because they are seasonally adjusted data nor in the variation rates because they are annual variation rates.

Scatterplots.

Beach umbrella rental

`View -->Graph specified vars --> X-Y scatter ...`

Select the variable in the Y-axis (umbrella, U) and in the X-axis (price, P).



gretl: define graph

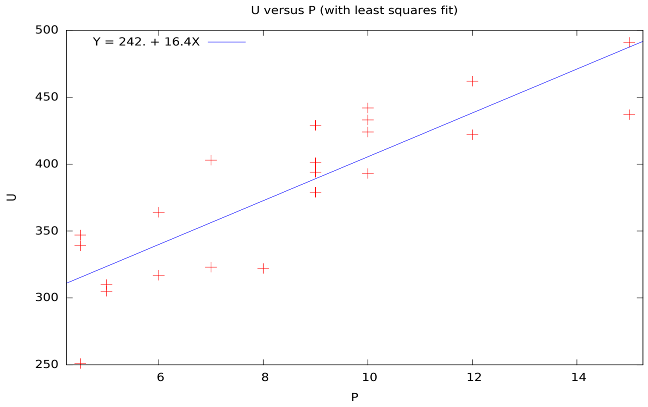
XY scatterplot

U
P
T
W

X-axis variable: P

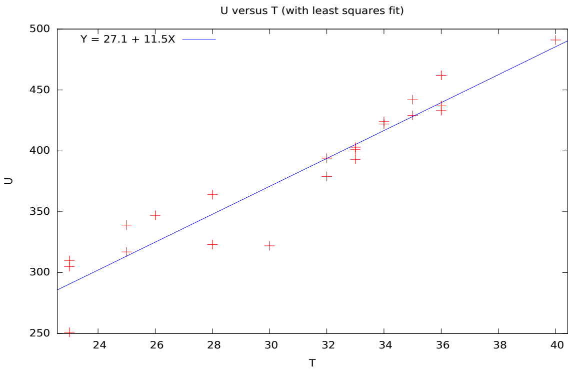
Y-axis variables: U

□ Show lagged variables

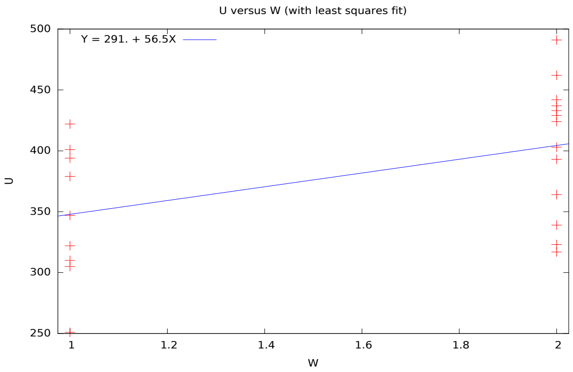


The graph shows a positive relationship between the number of rented umbrellas and the price.

Umbrellas against temperature



Umbrellas against W

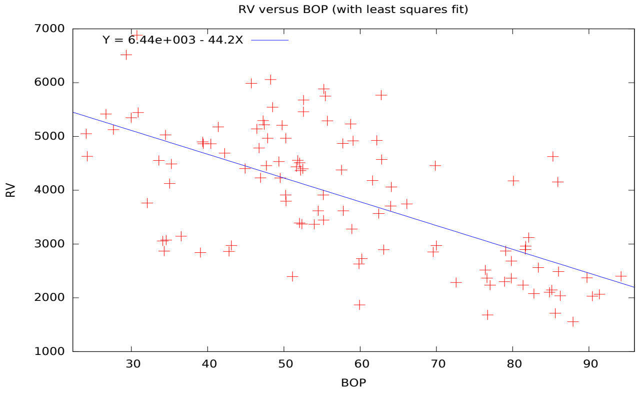


It may be observed in these graphs that there is a positive relationship between the number of rented umbrellas and the temperature, and the number of umbrellas rented and W. That is, the higher the temperature is (the less wind there is), the more umbrellas are rented.

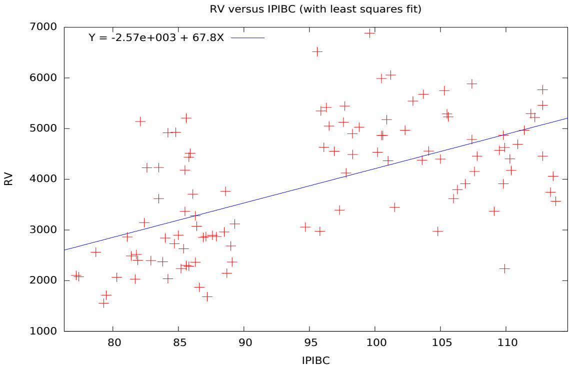
Registered vehicles

```
View -->Graph specified vars   --> X-Y scatter ...
```

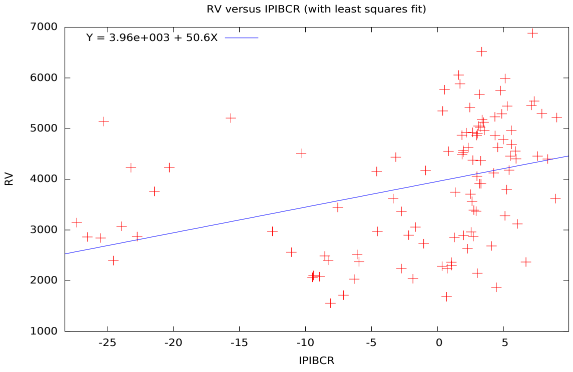
Select the variables in the Y-axis (vehicles, RV) and in the X-axis (price of oil, BOP).



RV against IPIBC



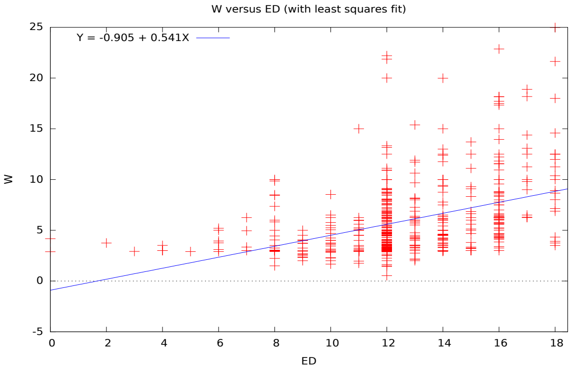
RV against IPIBCR



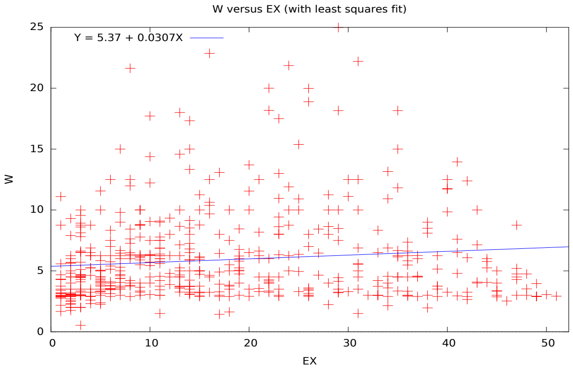
These graphs show that there is negative relationship between the number of vehicles registered and the price of Brent, while the relationship between the number of vehicles registered and the IPI (both in raw data and in annual rates) is positive. The IPI indicator is usually used as a proxy to measure the level of economic activity. The higher the economic activity is, the more vehicles are registered.

Wages

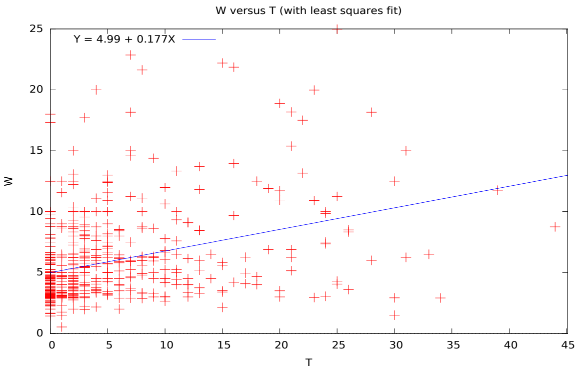
wages against *education*



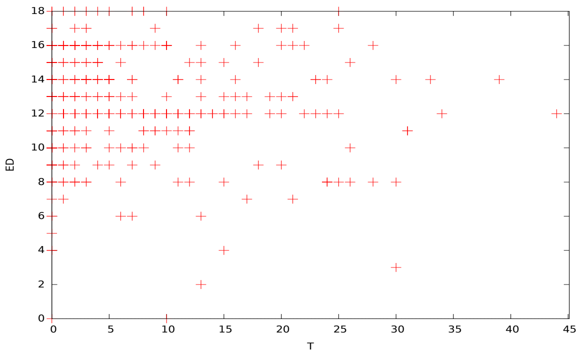
wages against *experience*



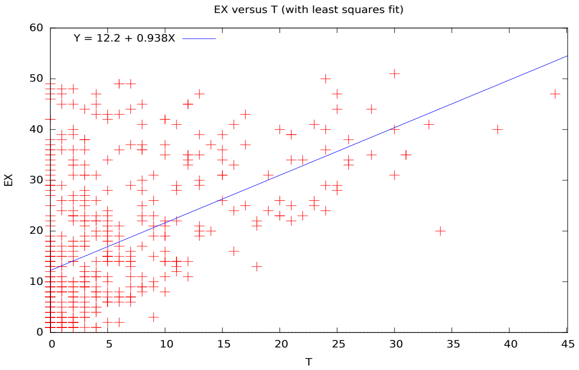
wages against *tenure*



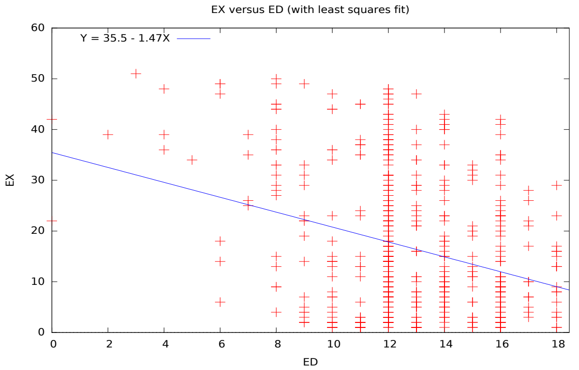
education against *tenure*



experience against *tenure*



experience against *education*



Analysing the graphs, it may be concluded that there is a positive relationship between wages and education and between wages and tenure. The larger the education/tenure is, the higher the wages are. However, the relationship between wages and experience does not seem to be very strong.

There is a strong and positive relationship between experience and tenure, whereas the relationship between education and experience or tenure is quite weak.

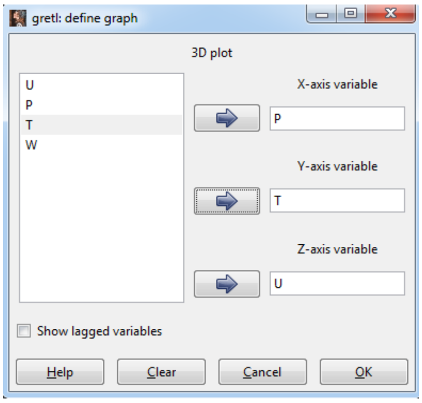
3D graphs.

Beach umbrella rental

To obtain a 3D graph, click:

```
View --> Graph specified vars   --> 3D plot...
```

Select the variables for the Z-axis (number of umbrellas, U), for the X-axis (price, P) and for the Y-axis (temperature, T).



gretl: define graph

3D plot

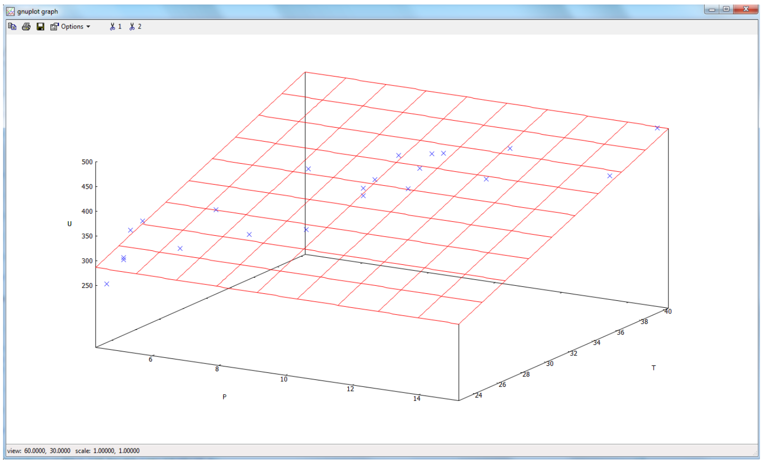
U
P
T
W

X-axis variable: P

Y-axis variable: T

Z-axis variable: U

□ Show lagged variables



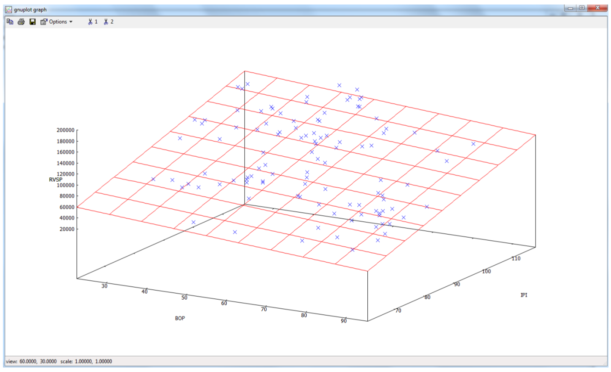
gnuplot graph

view: 60.0000, 30.0000 scale: 1.00000, 1.00000

It may be observed that the relationship between number of umbrellas and temperature is strong and positive and the relationship between number of umbrellas and price is not so strong and negative.

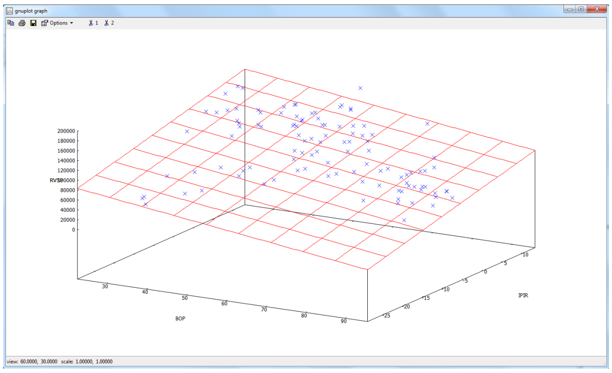
Registered vehicles

$RVSP$ against BOP and IPI



view: 60.0000, 30.0000 scale: 1.00000, 1.00000

$RVSP$ against BOP and $IPIR$

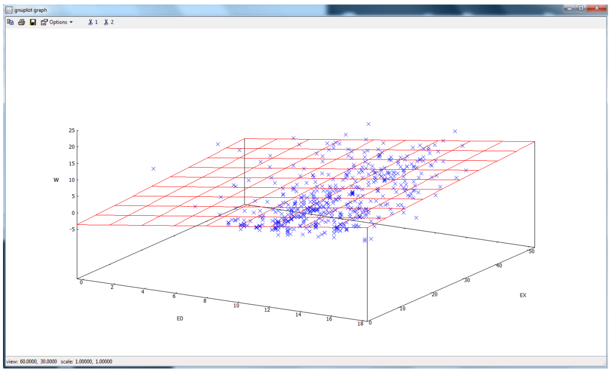


view: 60.0000, 30.0000 scale: 1.00000, 1.00000

It may observed in the graph that if the price of Brent increases the number of registered vehicles falls, while if the IPI (or its annual variation rate) increases, the number of registered vehicles goes up.

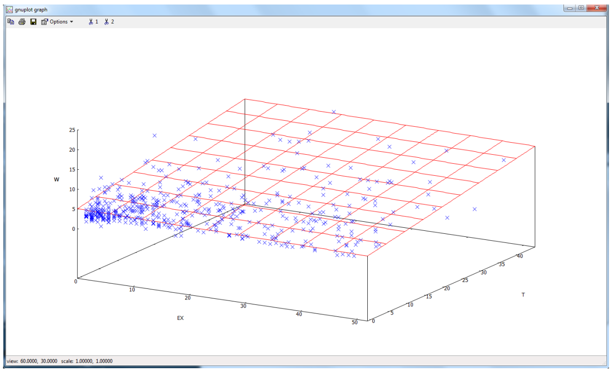
Wages

$wages$ against $education$ and $experience$



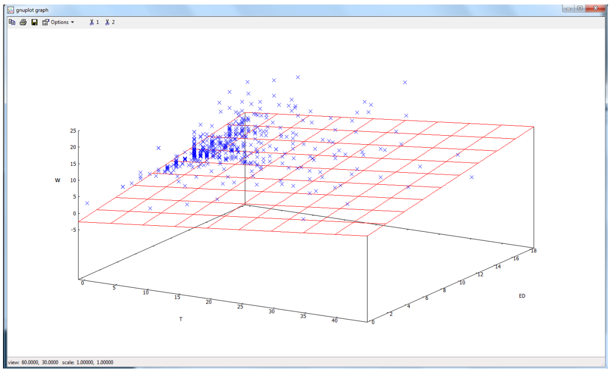
view: 60.0000, 30.0000 scale: 1.00000, 1.00000

$wages$ against $experience$ and $tenure$



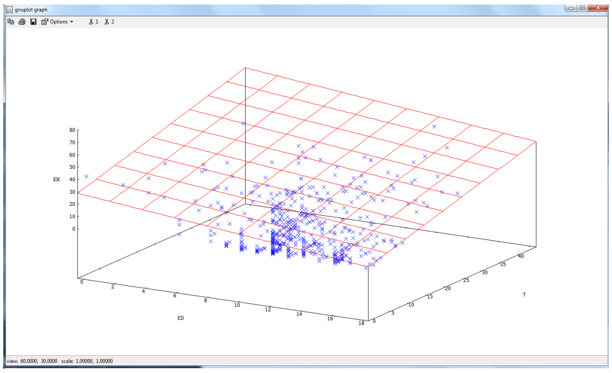
view: 60.0000, 30.0000 scale: 1.00000, 1.00000

$wages$ against $tenure$ and $education$



view: 60.0000, 30.0000 scale: 1.00000, 1.00000

$experience$ against $education$ and $tenure$



view: 60.0000, 30.0000 scale: 1.00000, 1.00000

It may be concluded that the relationship between wages and education or tenure is positive, while the relationship between wages and experience is quite weak.

Task T3.5. Univariate analysis.

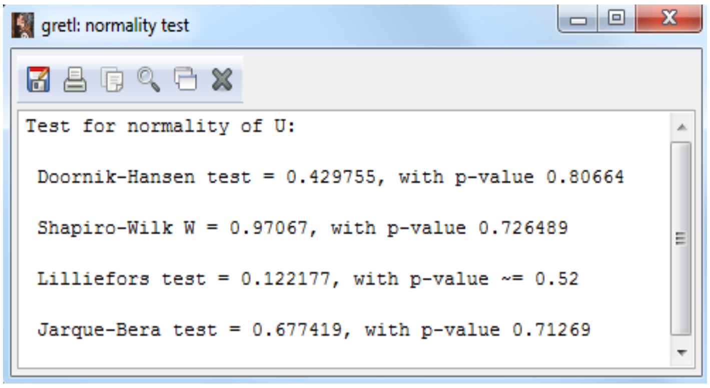
Beach umbrella rentals

a. Normality tests.

Select the variable and click

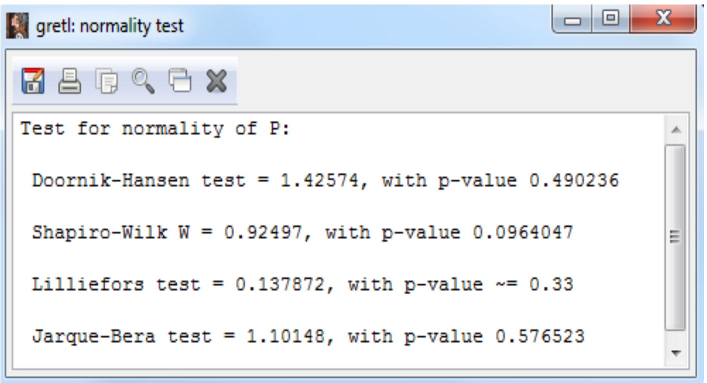
```
Variable --> Normality test
```

Rented umbrellas



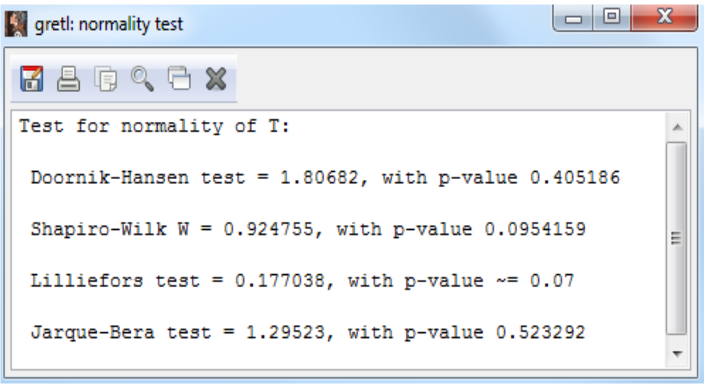
```
Test for normality of U:

 Doornik-Hansen test = 0.429755, with p-value 0.80664

 Shapiro-Wilk W = 0.97067, with p-value 0.726489

 Lilliefors test = 0.122177, with p-value ~= 0.52

 Jarque-Bera test = 0.677419, with p-value 0.71269
```

Price



```
Test for normality of P:

 Doornik-Hansen test = 1.42574, with p-value 0.490236

 Shapiro-Wilk W = 0.92497, with p-value 0.0964047

 Lilliefors test = 0.137872, with p-value ~= 0.33

 Jarque-Bera test = 1.10148, with p-value 0.576523
```

Temperature



```
Test for normality of T:

 Doornik-Hansen test = 1.80682, with p-value 0.405186

 Shapiro-Wilk W = 0.924755, with p-value 0.0954159

 Lilliefors test = 0.177038, with p-value ~= 0.07

 Jarque-Bera test = 1.29523, with p-value 0.523292
```

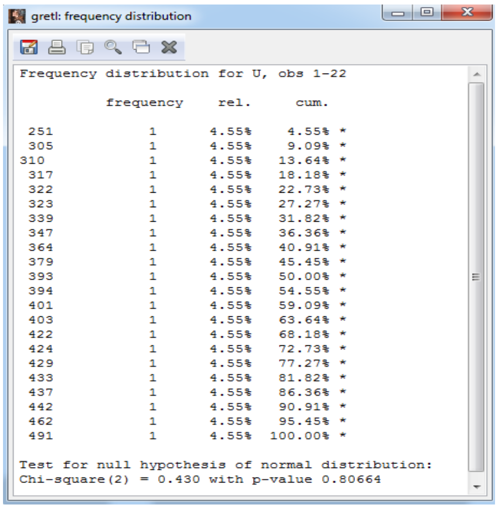
b. Frequency distribution.

To obtain the frequency distribution and the result of the normality test, click

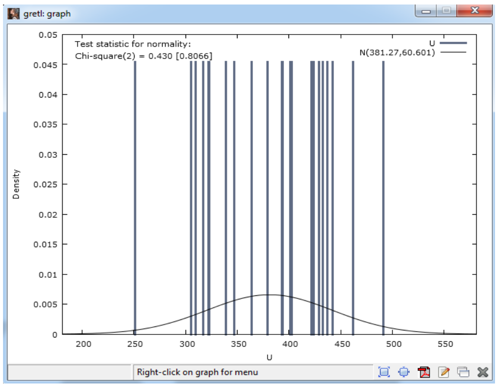
```
Variable --> Frequency distribution ...
```

and mark Test against normal distribution and show plot.

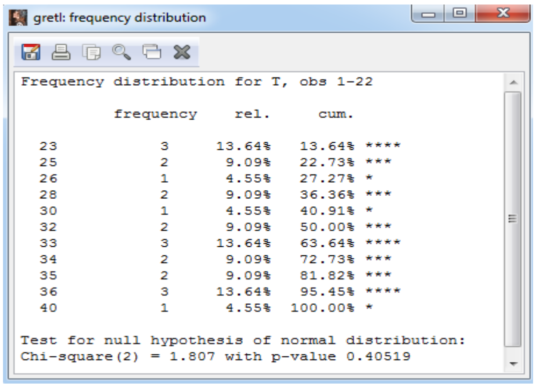
Rented umbrellas



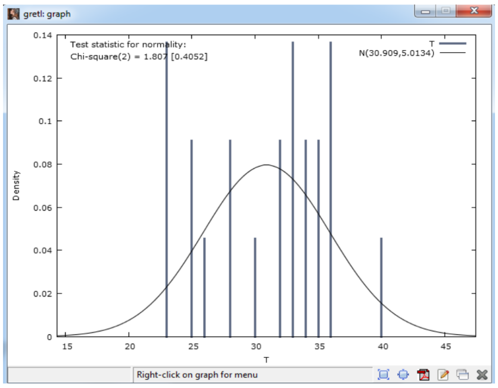
```
Frequency distribution for U, obs 1-22

          frequency    rel.     cum.

 251          1       4.55%    4.55% *
 305          1       4.55%    9.09% *
310           1       4.55%   13.64% *
 317          1       4.55%   18.18% *
 322          1       4.55%   22.73% *
 323          1       4.55%   27.27% *
 339          1       4.55%   31.82% *
 347          1       4.55%   36.36% *
 364          1       4.55%   40.91% *
 379          1       4.55%   45.45% *
 393          1       4.55%   50.00% *
 394          1       4.55%   54.55% *
 401          1       4.55%   59.09% *
 403          1       4.55%   63.64% *
 422          1       4.55%   68.18% *
 424          1       4.55%   72.73% *
 429          1       4.55%   77.27% *
 433          1       4.55%   81.82% *
 437          1       4.55%   86.36% *
 442          1       4.55%   90.91% *
 462          1       4.55%   95.45% *
 491          1       4.55%  100.00% *

Test for null hypothesis of normal distribution:
Chi-square(2) = 0.430 with p-value 0.80664
```



Temperature



```
Frequency distribution for T, obs 1-22

          frequency    rel.     cum.

 23           3      13.64%   13.64% ****
 25           2       9.09%   22.73% ***
 26           1       4.55%   27.27% *
 28           2       9.09%   36.36% ***
 30           1       4.55%   40.91% *
 32           2       9.09%   50.00% ***
 33           3      13.64%   63.64% ****
 34           2       9.09%   72.73% ***
 35           2       9.09%   81.82% ***
 36           3      13.64%   95.45% ****
 40           1       4.55%  100.00% *

Test for null hypothesis of normal distribution:
Chi-square(2) = 1.807 with p-value 0.40519
```

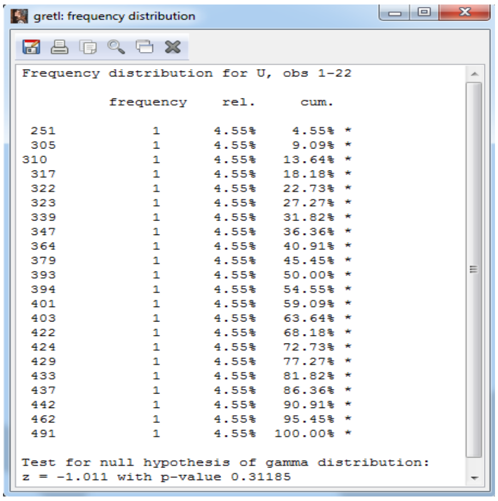


To obtain the frequency distribution and the result of the gamma distribution test, highlight the variable of interest using the cursor and click

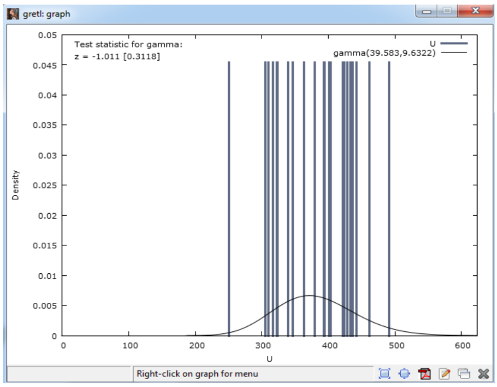
```
Variable --> Frequency distribution ...
```

and mark Test against gamma distribution and show plot.

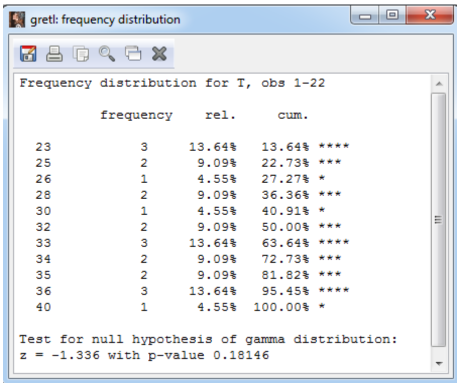
Rented umbrellas



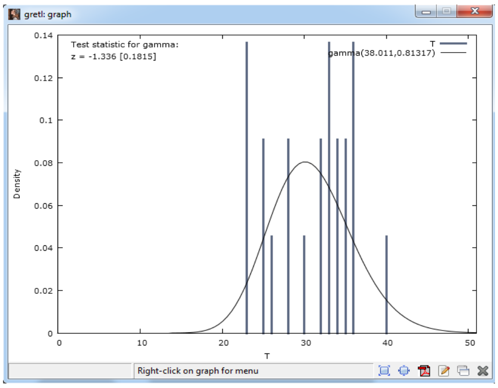
```
Frequency distribution for U, obs 1-22

          frequency    rel.     cum.

 251          1       4.55%    4.55% *
 305          1       4.55%    9.09% *
310           1       4.55%   13.64% *
 317          1       4.55%   18.18% *
 322          1       4.55%   22.73% *
 323          1       4.55%   27.27% *
 339          1       4.55%   31.82% *
 347          1       4.55%   36.36% *
 364          1       4.55%   40.91% *
 379          1       4.55%   45.45% *
 393          1       4.55%   50.00% *
 394          1       4.55%   54.55% *
 401          1       4.55%   59.09% *
 403          1       4.55%   63.64% *
 422          1       4.55%   68.18% *
 424          1       4.55%   72.73% *
 429          1       4.55%   77.27% *
 433          1       4.55%   81.82% *
 437          1       4.55%   86.36% *
 442          1       4.55%   90.91% *
 462          1       4.55%   95.45% *
 491          1       4.55%  100.00% *

Test for null hypothesis of gamma distribution:
z = -1.011 with p-value 0.31185
```



Temperature



```
Frequency distribution for T, obs 1-22

            frequency     rel.      cum.

  23            3     13.64%   13.64% ****
  25            2      9.09%   22.73% ***
  26            1      4.55%   27.27% *
  28            2      9.09%   36.36% ***
  30            1      4.55%   40.91% *
  32            2      9.09%   50.00% ***
  33            3     13.64%   63.64% ****
  34            2      9.09%   72.73% ***
  35            2      9.09%   81.82% ***
  36            3     13.64%   95.45% ****
  40            1      4.55%  100.00% *

Test for null hypothesis of gamma distribution:
z = -1.336 with p-value 0.18146
```



c. Results.

- At a 5 % significance level, the null hypothesis of a normal distribution is not rejected neither for the variable U nor for the variable T.
- At a 5 % significance level, the null hypothesis of a gamma distribution is not rejected neither for the variable U nor for the variable T.

Registered vehicles

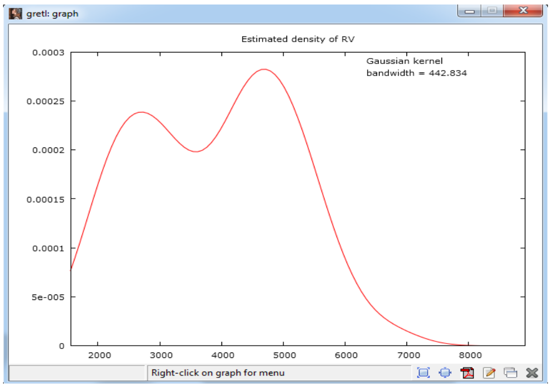
a. Estimate the density functions.

To estimate the density function, highlight the variable of interest using the cursor and click

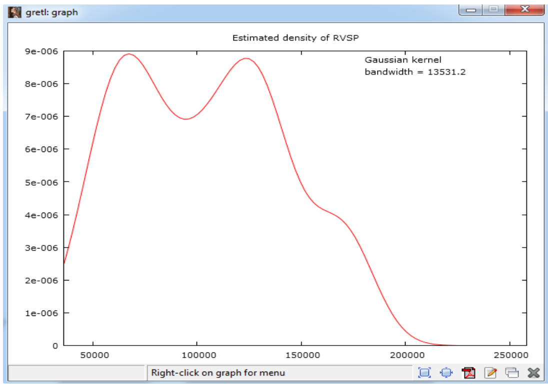
```
Variable --> Estimated density plot ...
```

and select the option *Gaussian kernel*.

RV

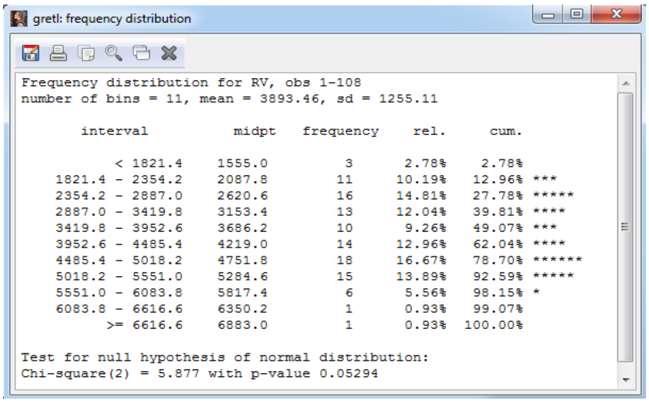


RVSP

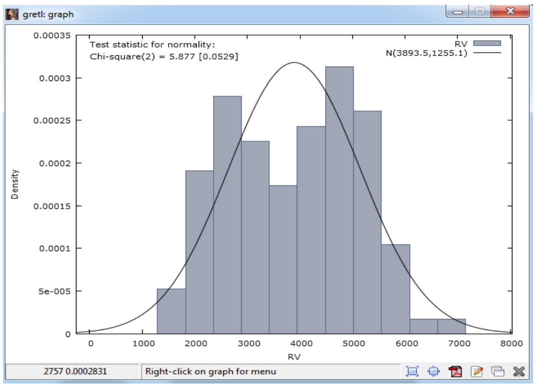


b. Frequency distribution and normality test.

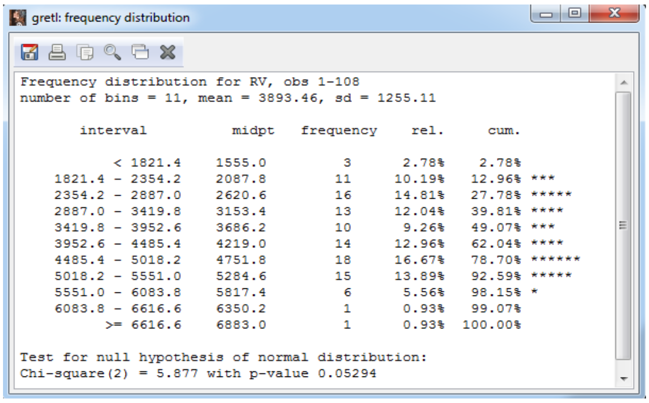
RV



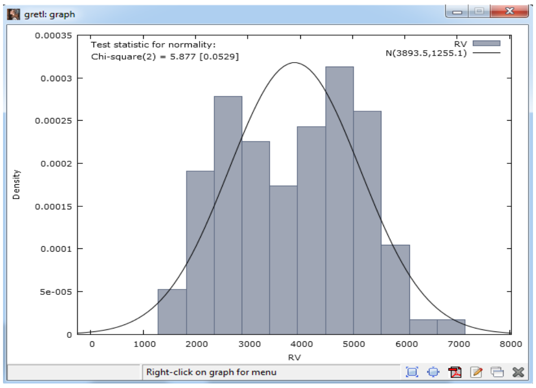
```
Frequency distribution for RV, obs 1-108
number of bins = 11, mean = 3893.46, sd = 1255.11

       interval          midpt   frequency    rel.     cum.

           < 1821.4     1555.0        3      2.78%    2.78%
    1821.4 - 2354.2     2087.8       11     10.19%   12.96% ***
    2354.2 - 2887.0     2620.6       16     14.81%   27.78% *****
    2887.0 - 3419.8     3153.4       13     12.04%   39.81% ****
    3419.8 - 3952.6     3686.2       10      9.26%   49.07% ***
    3952.6 - 4485.4     4219.0       14     12.96%   62.04% ****
    4485.4 - 5018.2     4751.8       18     16.67%   78.70% ******
    5018.2 - 5551.0     5284.6       15     13.89%   92.59% *****
    5551.0 - 6083.8     5817.4        6      5.56%   98.15% *
    6083.8 - 6616.6     6350.2        1      0.93%   99.07%
          >= 6616.6     6883.0        1      0.93%  100.00%

Test for null hypothesis of normal distribution:
Chi-square(2) = 5.877 with p-value 0.05294
```



RVSP



```
Frequency distribution for RV, obs 1-108
number of bins = 11, mean = 3893.46, sd = 1255.11

       interval          midpt   frequency    rel.     cum.

           < 1821.4     1555.0        3      2.78%    2.78%
    1821.4 - 2354.2     2087.8       11     10.19%   12.96% ***
    2354.2 - 2887.0     2620.6       16     14.81%   27.78% *****
    2887.0 - 3419.8     3153.4       13     12.04%   39.81% ****
    3419.8 - 3952.6     3686.2       10      9.26%   49.07% ***
    3952.6 - 4485.4     4219.0       14     12.96%   62.04% ****
    4485.4 - 5018.2     4751.8       18     16.67%   78.70% ******
    5018.2 - 5551.0     5284.6       15     13.89%   92.59% *****
    5551.0 - 6083.8     5817.4        6      5.56%   98.15% *
    6083.8 - 6616.6     6350.2        1      0.93%   99.07%
          >= 6616.6     6883.0        1      0.93%  100.00%

Test for null hypothesis of normal distribution:
Chi-square(2) = 5.877 with p-value 0.05294
```

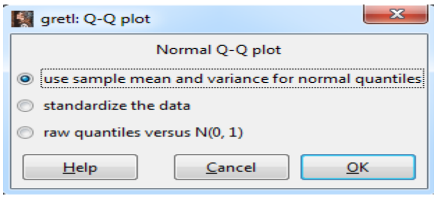


c. Q-Q plots.

To obtain the Q-Q plot, highlight the variable of interest using the cursor, click

`Variable --> Normal Q-Q plot ...`

and mark use sample mean and variance for normal quantiles.



gretl: Q-Q plot

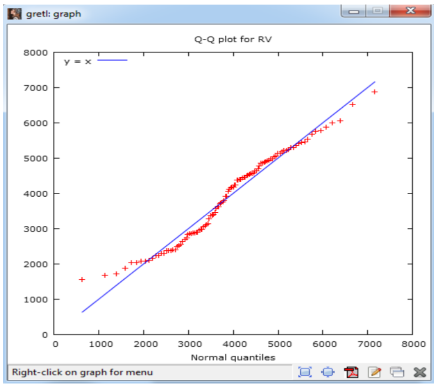
Normal Q-Q plot

● use sample mean and variance for normal quantiles

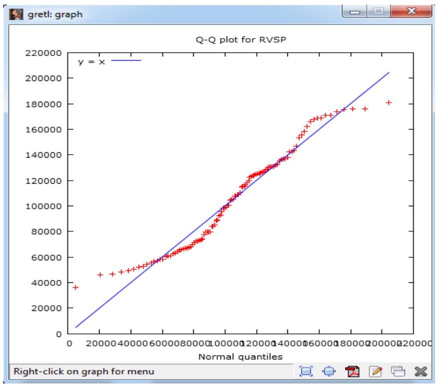
○ standardize the data

○ raw quantiles versus N(0, 1)

RV



RVSP



d. Results:

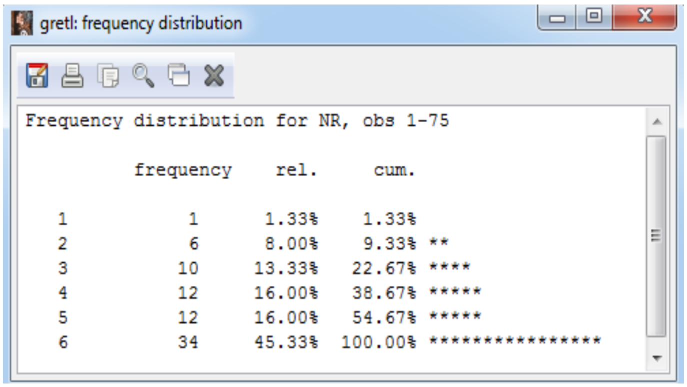
- Both distributions have two modes.
- The Q-Q plots suggest that the distribution of the variable RV is closer to the normal than the distribution of variable $RVSP$.
- The normality tests conclude that, at the 5 % significance level, the null hypothesis of normality is rejected for the variable $RVSP$ but it is not rejected for the variable RV.

Holiday cottages in Biscay

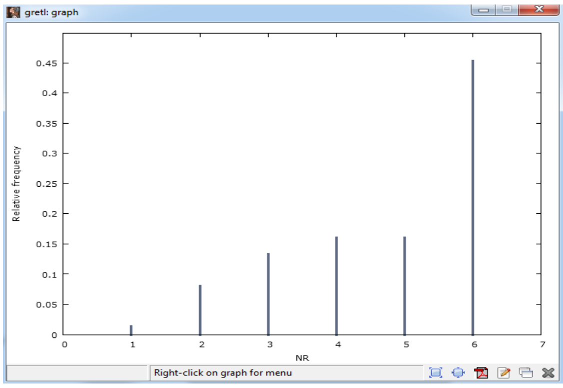
a. To analyse the distribution of the variable number of rooms, click

```
Variable --> Frequency distribution ...
```

and mark Show data only.

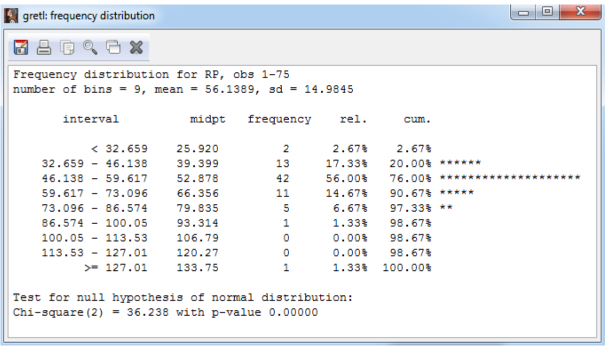


```
Frequency distribution for NR, obs 1-75

           frequency    rel.     cum.

  1             1      1.33%    1.33%
  2             6      8.00%    9.33% **
  3            10     13.33%   22.67% ****
  4            12     16.00%   38.67% *****
  5            12     16.00%   54.67% *****
  6            34     45.33%  100.00% ****************
```



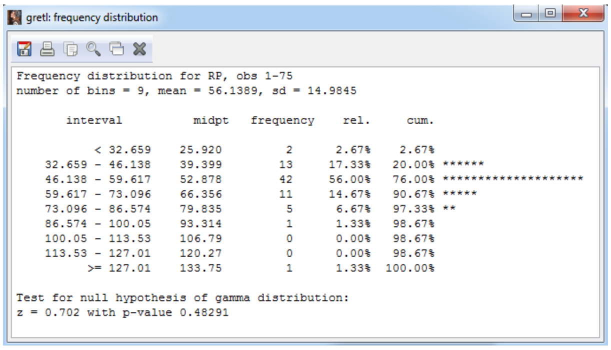
b. Frequency distribution of variable room price and normality tests.

Normal distribution



```
Frequency distribution for RP, obs 1-75
number of bins = 9, mean = 56.1389, sd = 14.9845

       interval          midpt   frequency    rel.     cum.

            < 32.659    25.920         2      2.67%    2.67%
    32.659 - 46.138    39.399        13     17.33%   20.00% ******
    46.138 - 59.617    52.878        42     56.00%   76.00% ********************
    59.617 - 73.096    66.356        11     14.67%   90.67% *****
    73.096 - 86.574    79.835         5      6.67%   97.33% **
    86.574 - 100.05    93.314         1      1.33%   98.67%
    100.05 - 113.53    106.79         0      0.00%   98.67%
    113.53 - 127.01    120.27         0      0.00%   98.67%
           >= 127.01    133.75         1      1.33%  100.00%

Test for null hypothesis of normal distribution:
Chi-square(2) = 36.238 with p-value 0.00000
```

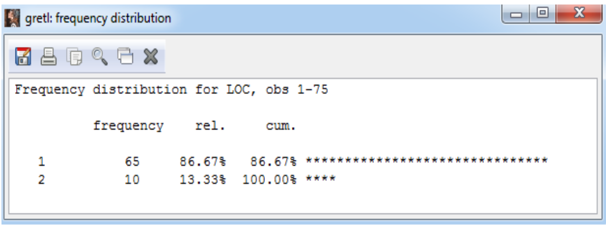
Gamma distribution



```
Frequency distribution for RP, obs 1-75
number of bins = 9, mean = 56.1389, sd = 14.9845

       interval          midpt   frequency    rel.     cum.

            < 32.659    25.920         2      2.67%    2.67%
    32.659 - 46.138    39.399        13     17.33%   20.00% ******
    46.138 - 59.617    52.878        42     56.00%   76.00% ********************
    59.617 - 73.096    66.356        11     14.67%   90.67% *****
    73.096 - 86.574    79.835         5      6.67%   97.33% **
    86.574 - 100.05    93.314         1      1.33%   98.67%
    100.05 - 113.53    106.79         0      0.00%   98.67%
    113.53 - 127.01    120.27         0      0.00%   98.67%
           >= 127.01    133.75         1      1.33%  100.00%

Test for null hypothesis of gamma distribution:
z = 0.702 with p-value 0.48291
```

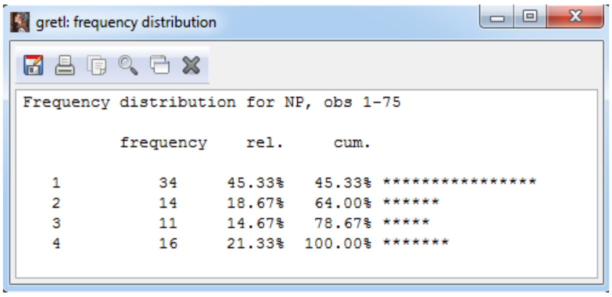
c. Frequency distribution of the qualitative variables.

LOC



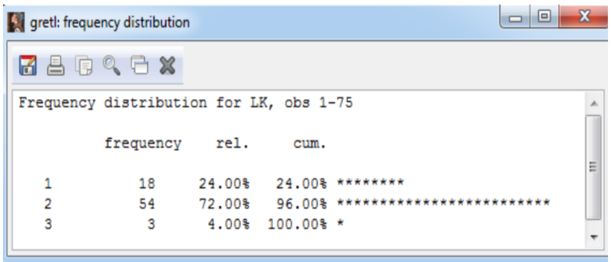
```
Frequency distribution for LOC, obs 1-75

          frequency    rel.     cum.

  1          65     86.67%   86.67% *******************************
  2          10     13.33%  100.00% ****
```

NP



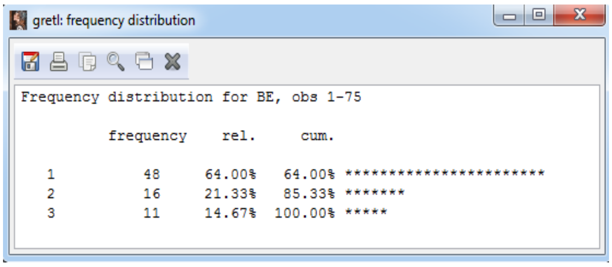
```
Frequency distribution for NP, obs 1-75

          frequency    rel.     cum.

  1          34     45.33%   45.33% ****************
  2          14     18.67%   64.00% ******
  3          11     14.67%   78.67% *****
  4          16     21.33%  100.00% *******
```

LK



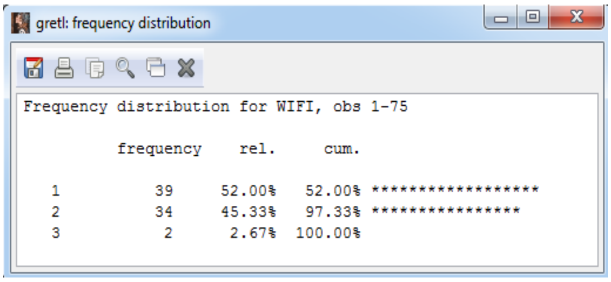
```
Frequency distribution for LK, obs 1-75

          frequency    rel.     cum.

  1          18     24.00%   24.00% ********
  2          54     72.00%   96.00% *************************
  3           3      4.00%  100.00% *
```

BE



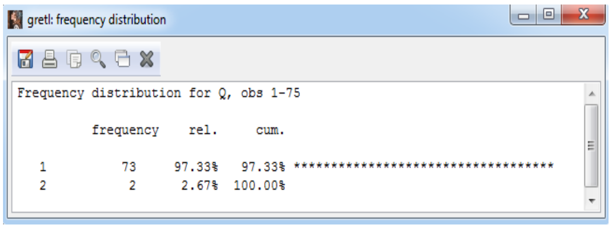
```
Frequency distribution for BE, obs 1-75

          frequency    rel.     cum.

  1          48     64.00%   64.00% ***********************
  2          16     21.33%   85.33% *******
  3          11     14.67%  100.00% *****
```

WiFi



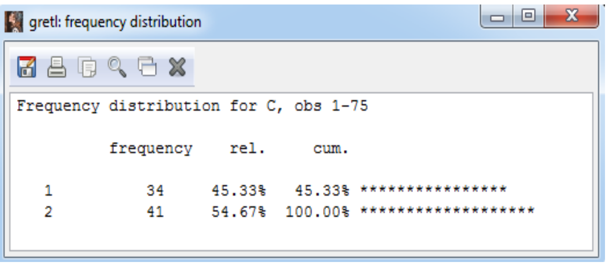
```
Frequency distribution for WIFI, obs 1-75

          frequency    rel.     cum.

  1          39     52.00%   52.00% ******************
  2          34     45.33%   97.33% ****************
  3           2      2.67%  100.00%
```

Q



```
Frequency distribution for Q, obs 1-75

          frequency    rel.     cum.

  1          73     97.33%   97.33% ***********************************
  2           2      2.67%  100.00%
```

C



```
Frequency distribution for C, obs 1-75

          frequency    rel.     cum.

  1          34     45.33%   45.33% ****************
  2          41     54.67%  100.00% *******************
```

c.1. It refers to the category R = 3 of the variable BE. Therefore, 14.67 % of the holiday cottages are less than 1 km from a beach.

c.2. It refers to the category V = 1 of the variable LK. Therefore, 24 % of the holiday cottages are more than 20 km from a lake or reservoir.

c.3. It refers to the category P = 4 of the variable NP. Therefore, 21.33 % of the holiday cottages are inside a natural park.

c.4. It refers to the category R = 3 of the variable BE. Therefore, 14.67 % of the holiday cottages are less than 1 km from a beach.

It refers to the category R = 2 of the variable NP. Therefore, 18.67 % of the holiday cottages are less than 1 km from a natural park.

It refers to the category V = 1 of the variable BE. Therefore, 64 % of the holiday cottages are more than 20 km from a beach.

It refers to the category V = 1 of the variable NP. Therefore, 45.33 % of the holiday cottages are more than 20 km from a natural park.

c.5. It refers to the category A = 1 of the variable LOC. Therefore, 86.67 % of the holiday cottages are not in the the town center.

c.6. It refers to the categories G = 1 and S = 3 of the variable WiFi. Therefore, 54.67 % of the holiday cottages offer WiFi.

It refers to the category G = 1 of the variable WiFi. Therefore, 52 % of the holiday cottages offer free WiFi.

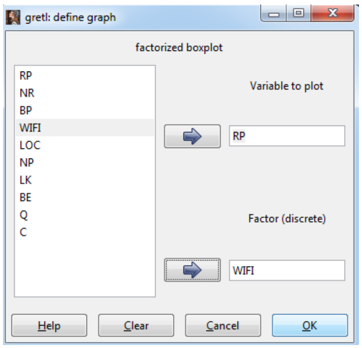
c.7. It refers to the category N = 1 of the variable Q. Therefore, 97.33 % of the holiday cottages do not have the Tourism Quality Q certificate.

It refers to the category N = 2 of the variable C. Therefore, 54.67 % of the holiday cottages do not have the Commitment to Quality Tourism Award.

d. Boxplot of the variable average price of a room.

```
Variable --> Boxplot
```

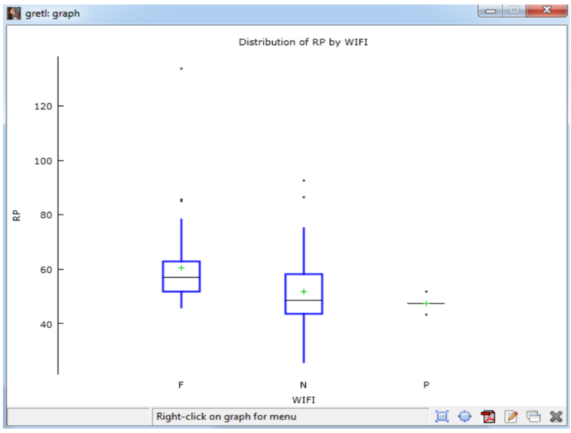
and mark Factorized. In the dialog box select the variable price of a room and the factor WiFi.


gretl: define graph

factorized boxplot

Variable to plot: RP

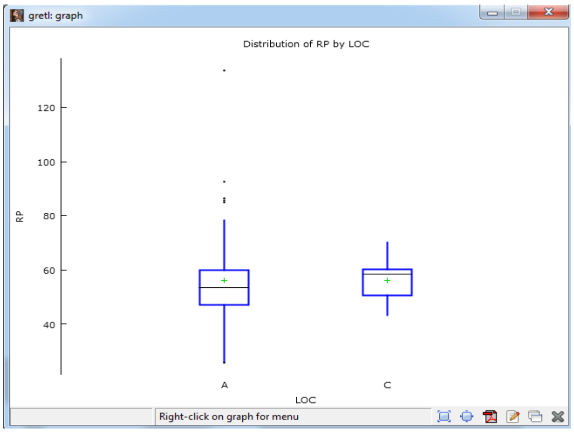
Factor (discrete): WIFI

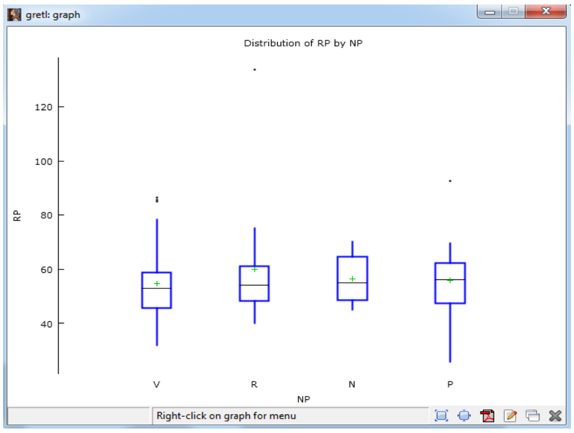


Using the rest of the qualitative variables as factors:

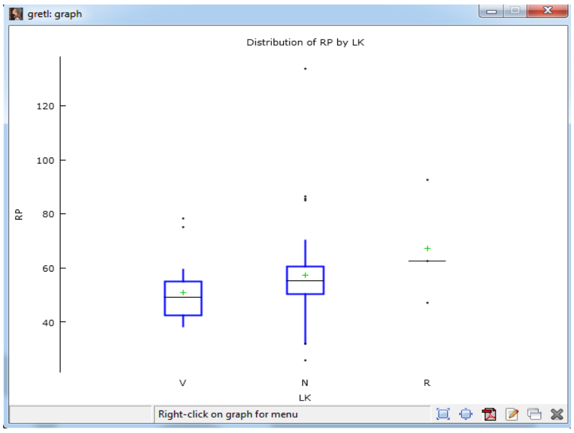
LOC

NP

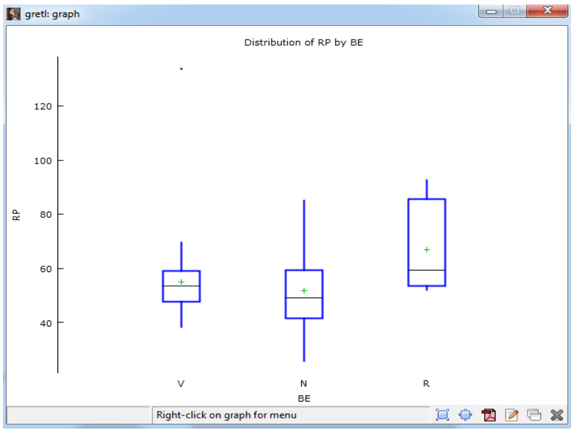




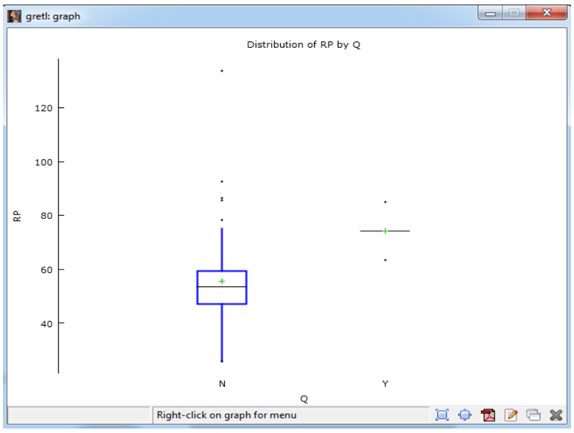
LK



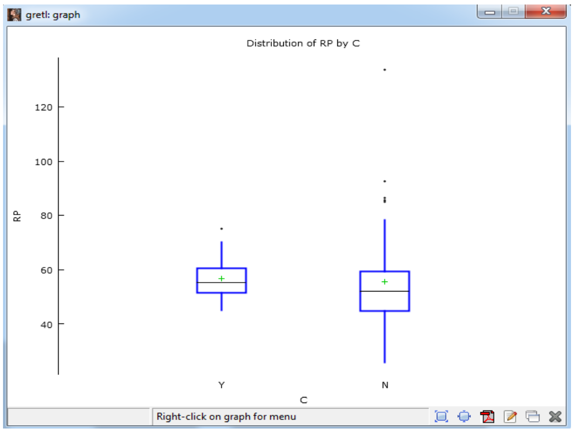
BE



Q



C



e. Results.

- Given the results of the tests, the null hypothesis that the variable price follows a normal distribution is rejected at a 5 % significance level, but it is not rejected the null hypothesis that the variable price follows a gamma distribution.

- No.

- The frequency distribution of the variable LOC is concentrated in the category "far from the town center".

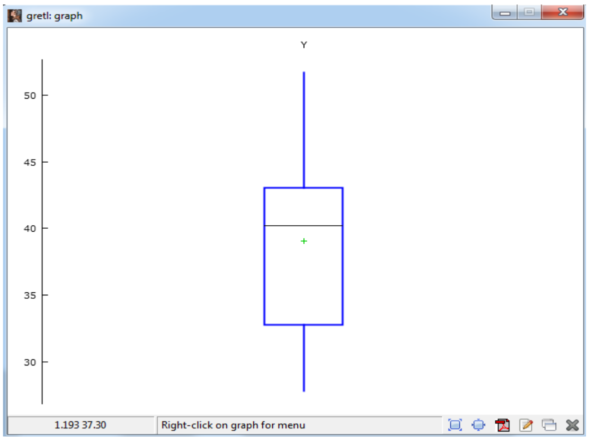
  The frequency distribution of the variable WiFi is concentrated in the categories "free WiFi access" and "No WiFi access".

- First, having the Tourism Quality Q certificate, followed by WiFi access and proximity to a lake or a beach.

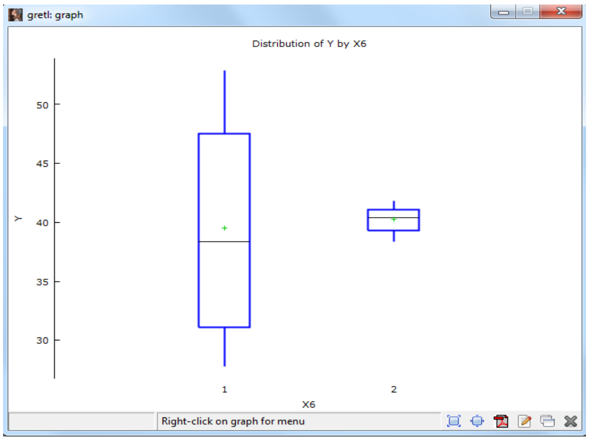
Chicken consumption

a. and b. Boxplots.

Simple boxplot



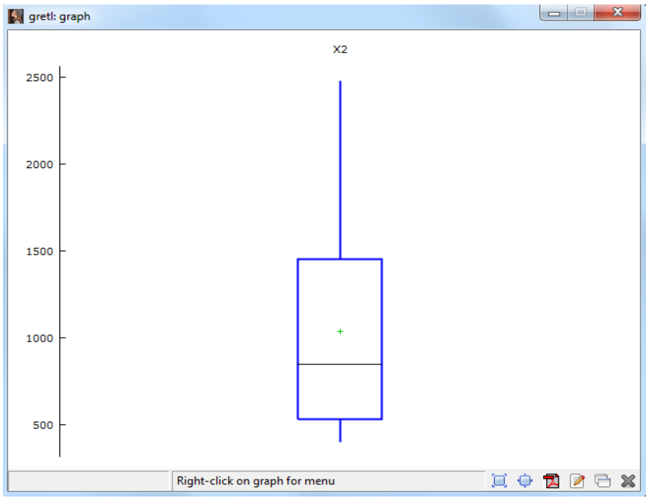
Factorized boxplot (X6)



c. Boxplot of variable income.

`Variable --> Boxplot`

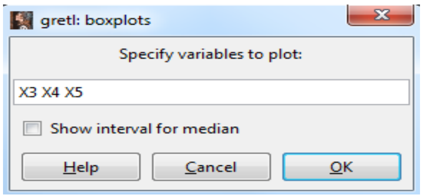
and mark Simple boxplot.



d. Joint boxplot of all the price variables.

`View --> Graphs specified vars --> Boxplots ...`

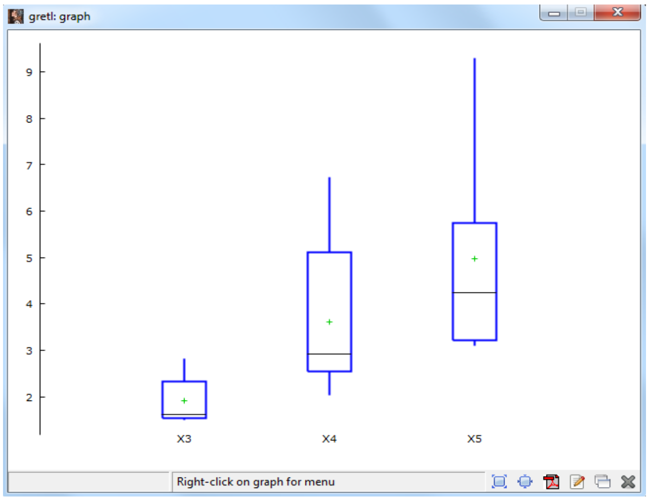
Then, select the price variables.


gretl: boxplots

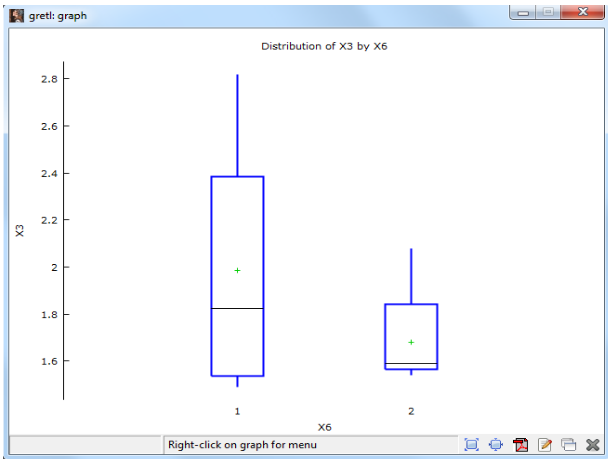
Specify variables to plot:

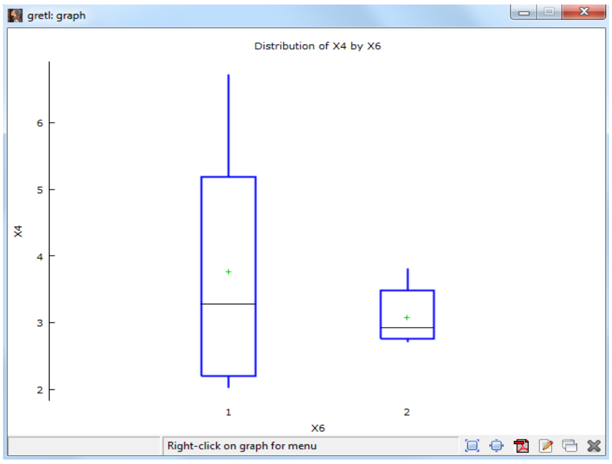
X3 X4 X5

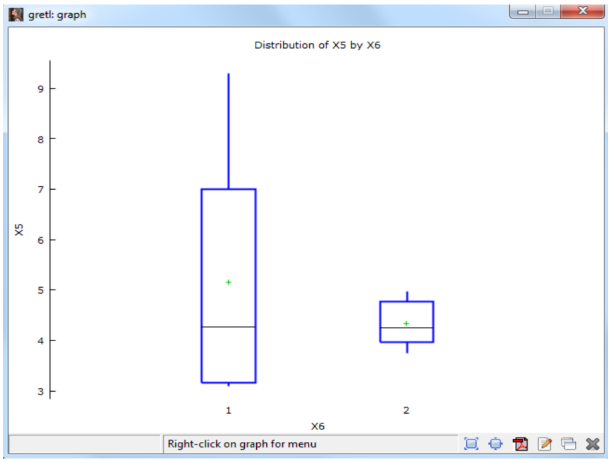
□ Show interval for median



e. Factorized boxplot for each price variable.







f. Results.

- The variability observed in the variable consumption of chicken comes mostly from the period of avian flu epidemic. Note that the variability observed in the years without epidemic is very small.
- The dispersion in the distribution of the variable chicken price is much smaller than the dispersion in the distributions of the beef price and of the pork price.
- The biggest difference in the price of a product due to the avian flue epidemic is observed in the price of chicken.