eman ta zabal zazu

Universidad del País Vasco | Euskal Herriko Unibertsitatea

CAMPUS OF INTERNATIONAL EXCELLENCE

Example 5.1

OLS estimation: cross-section data

Pilar González and Susan Orbe

Dpt. Applied Economics III (Econometrics and Statistics)

Contents

Contents

Example 5.1.1. Ordinary Least Squares.

Pizza consumption.

Estimate a regression model to determine the consumption of pizza (data file `pizza.gdt`).

a. Regress pizza consumption on age and income. Assume that the relationship between the variables is linear. Save the results to the session as an icon.

b. Restrict the sample to the clients over 35 years old. Estimate the model using this restricted sample and save the results to the session as an icon. Do you obtain the same results? Restore the original sample.

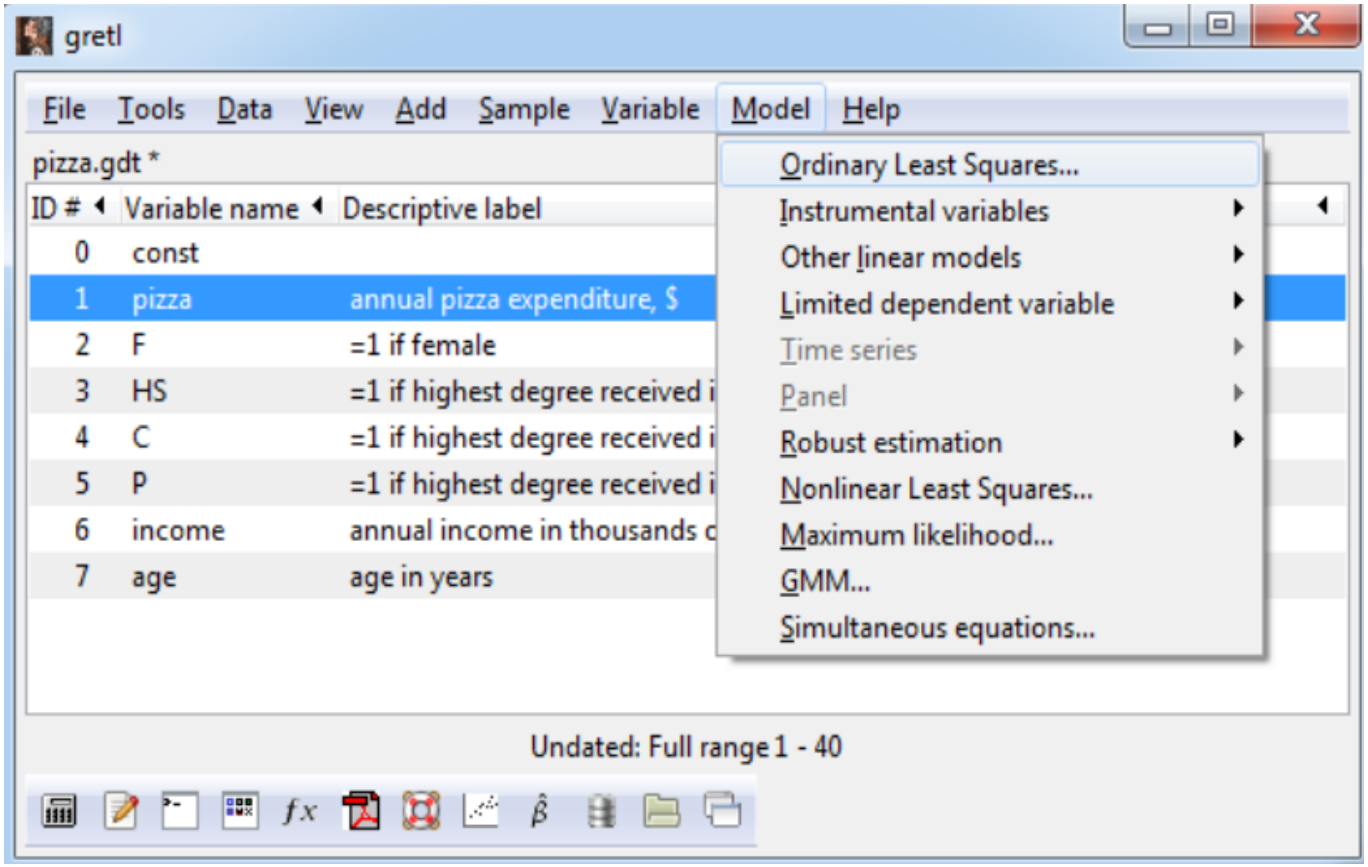
c. Restrict the sample to the clients with only high school education. Estimate the model using this restricted sample and save the results to the session as an icon. Do you obtain the same results? Restore the original sample.

d. Interpret the results. Save the session as pizza5.1.1.

Example 5.1.1. Ordinary Least Squares.

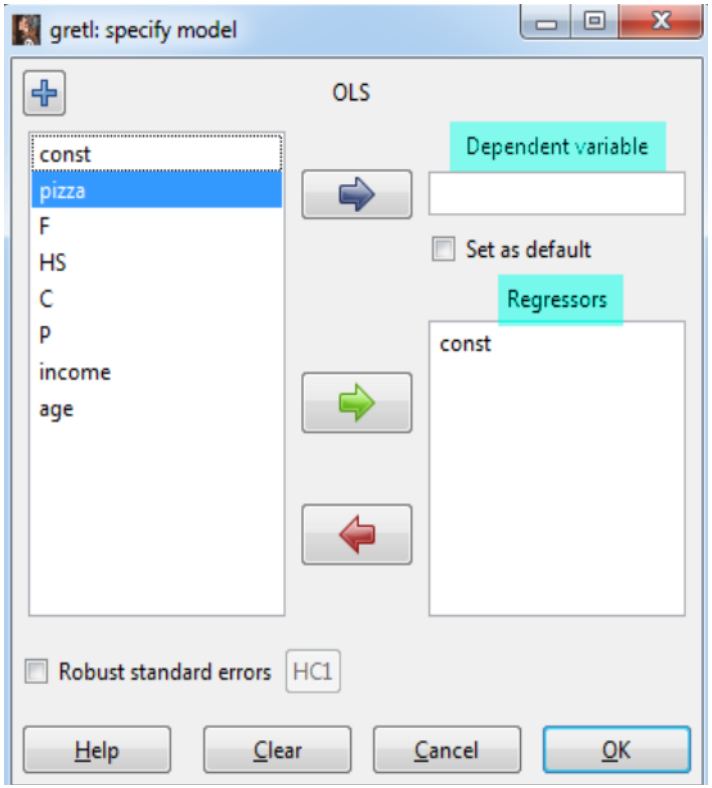
To estimate a regression model by OLS, select the option *Ordinary Least Squares* from the **Model** pulldown menu.

```
Model -> Ordinary Least Squares ...
```



Example 5.1.1. Ordinary Least Squares.

The options in the dialog box allow you to specify the model you want to estimate. You have to select the dependent variable, and the regressors among the variables included in the left box.



Example 5.1.1. Ordinary Least Squares.

Notice that by default, Gretl assumes that you want to estimate an intercept and includes this in the regressors list by default ("const").

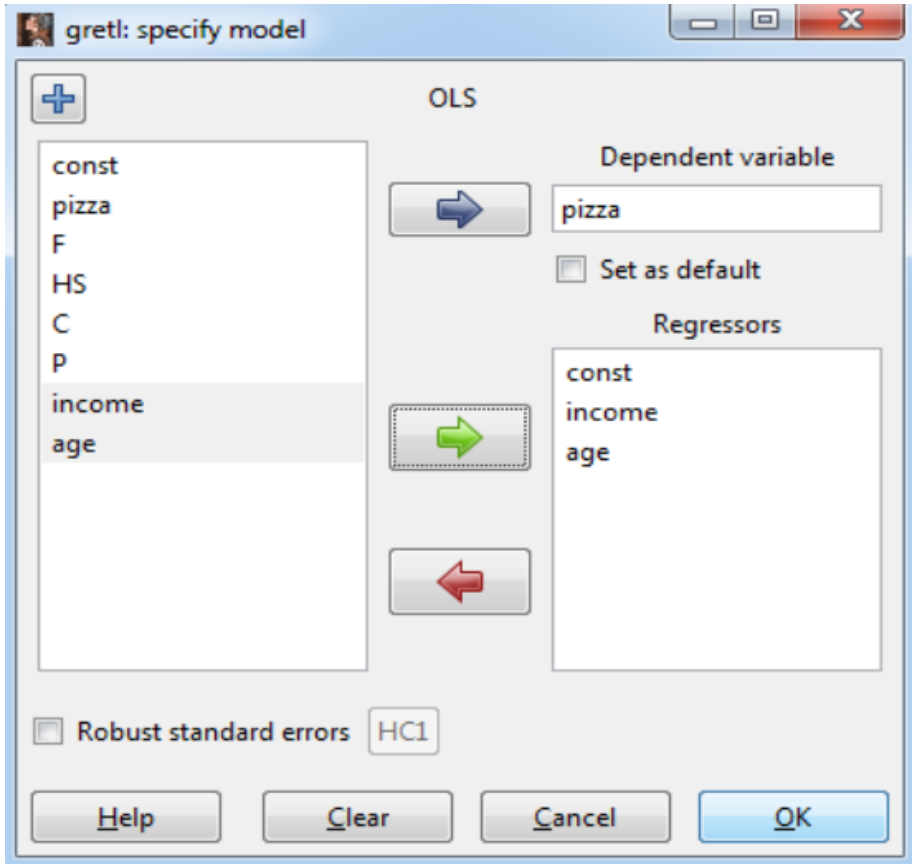
To include income as a regressor, highlight the variable in the left box using the cursor and click on the green arrow.

To include pizza as the dependent variable, highlight the variable in the left box using the cursor and click on the blue arrow.

If you want the variable pizza to be the dependent variable in all the regression models, mark Set as default.

Example 5.1.1. Ordinary Least Squares.

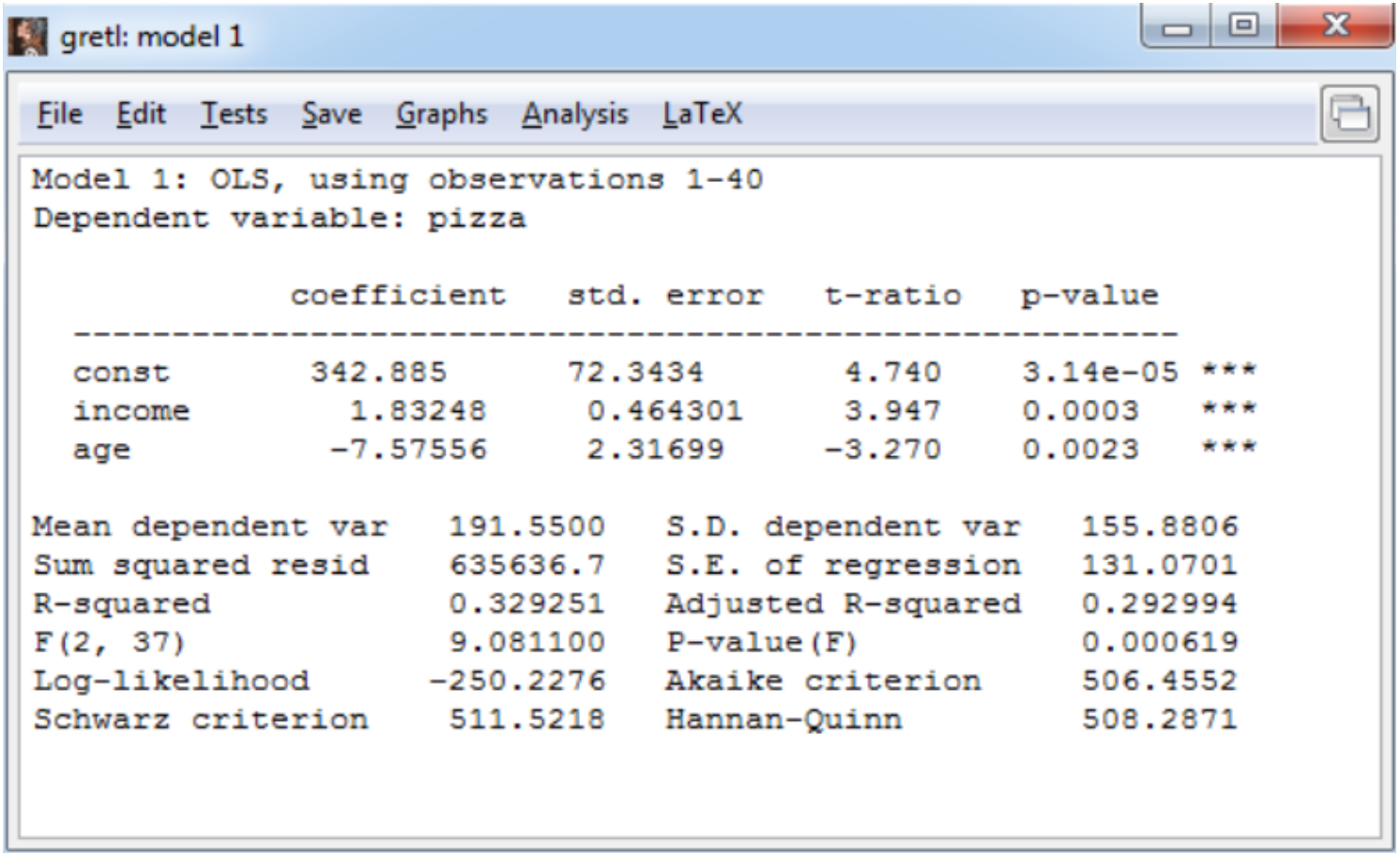
Model: $pizza_i = \beta_1 + \beta_2\, income_i + \beta_3\, age_i + u_i \quad i = 1, ..., N$



There are several options at the bottom of the window: you may ask for **Help**, **Clear** the selected variables or **Cancel** the estimation. To estimate the specified model click **OK**.

Example 5.1.1. Ordinary Least Squares.

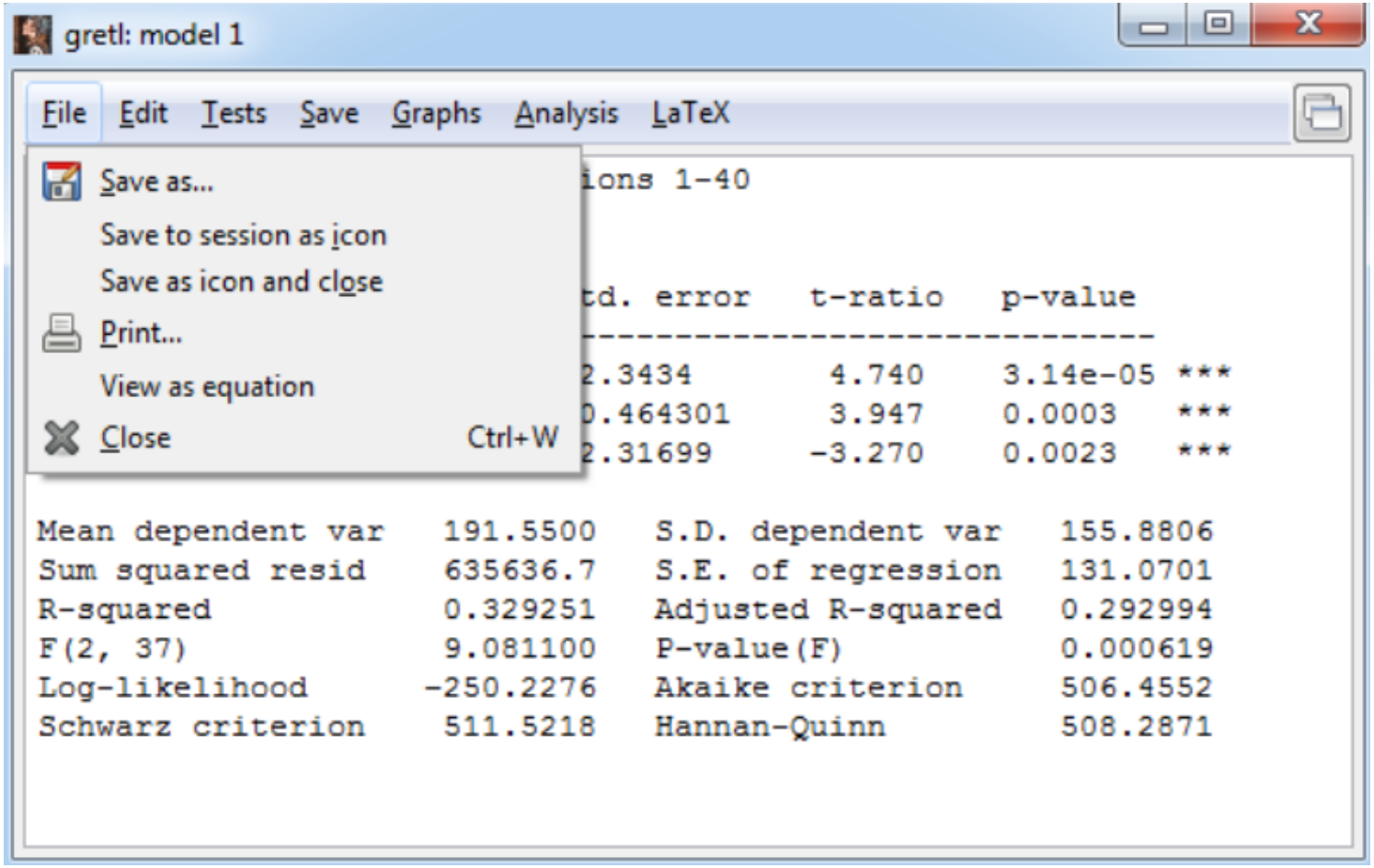
Estimation results.



There is a menu with several options at the top of the output window.

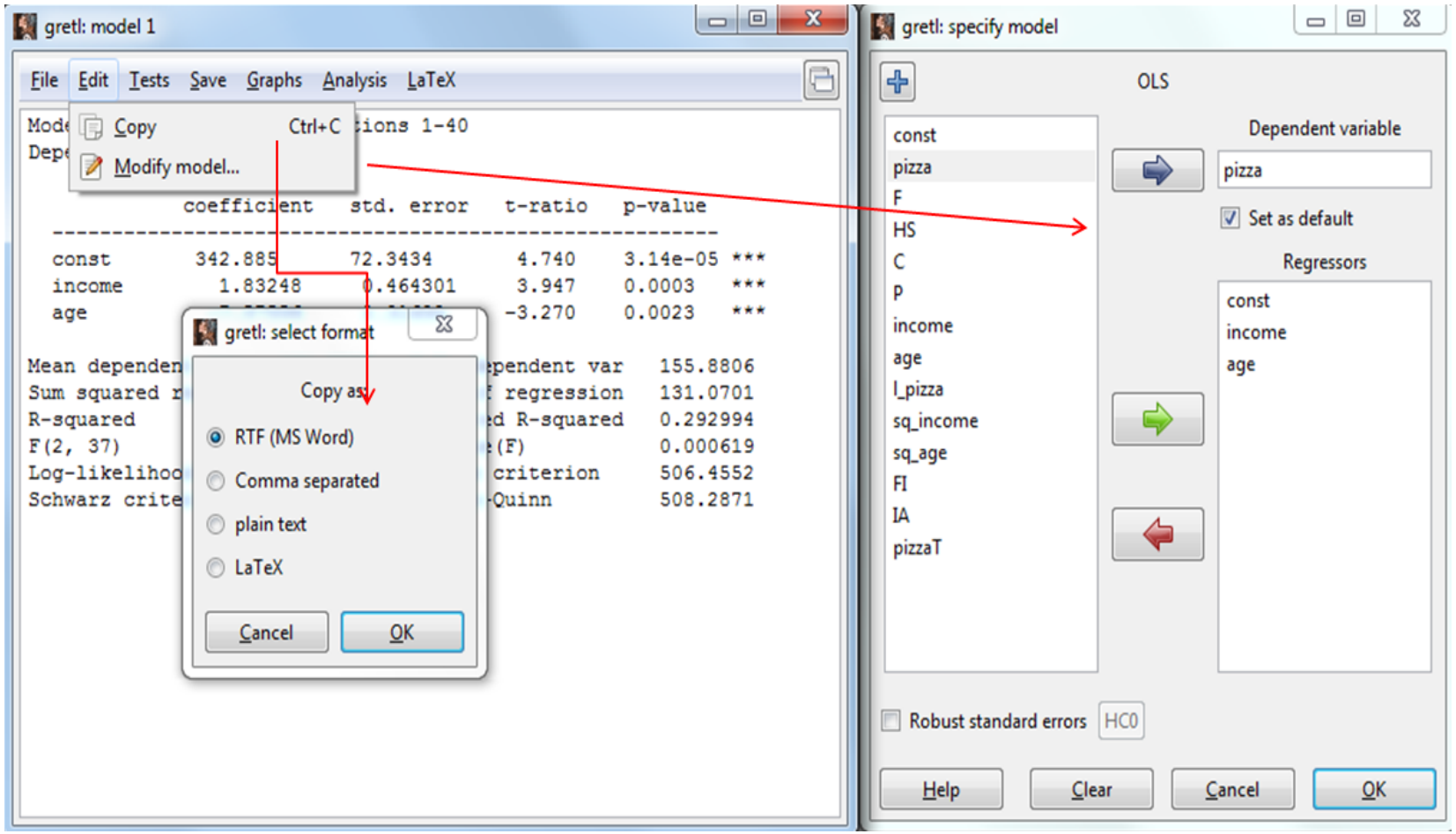
Example 5.1.1. Ordinary Least Squares.

File. To save the results in several formats (Word, comma separated, plain text or LaTeX) or as an icon to the session, to print them or to view them in equation form.



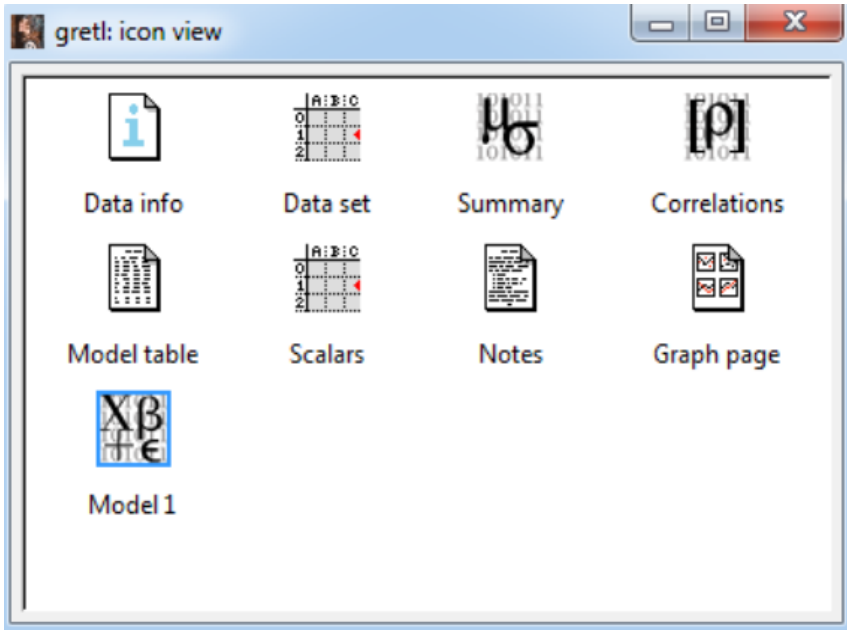
Example 5.1.1. Ordinary Least Squares.

Edit. To copy the results in several formats (Word, comma separated, plain text or LaTeX), or to modify the estimated model changing either the dependent variable or the regressors.



Example 5.1.1. Ordinary Least Squares.

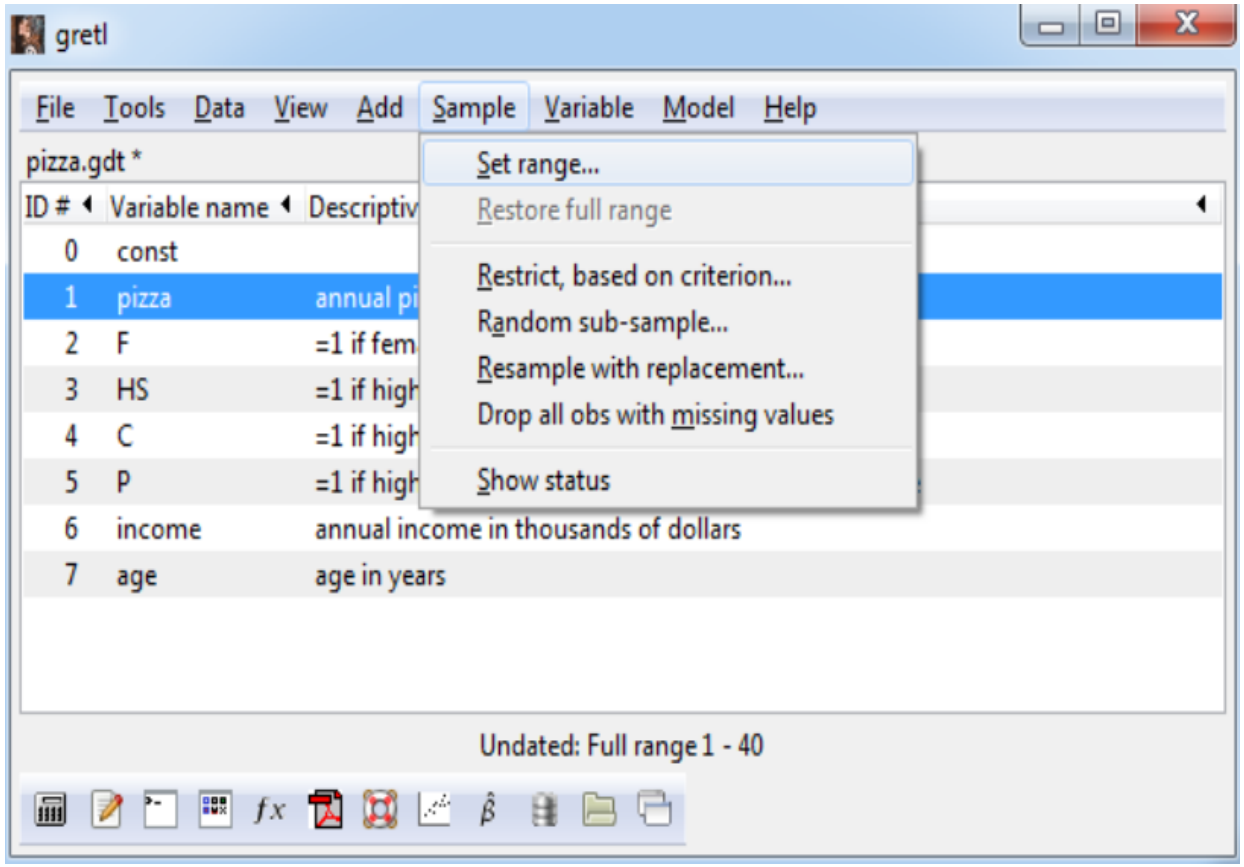
If you save the model to the session as an icon, a new icon appears in the session.



By default, this icon is called *Model 1* because it is the first model estimated in the present session. Right-clicking on this icon it is possible to change the name of this icon (`Rename`) or to retrieve the estimation results (`Display`).

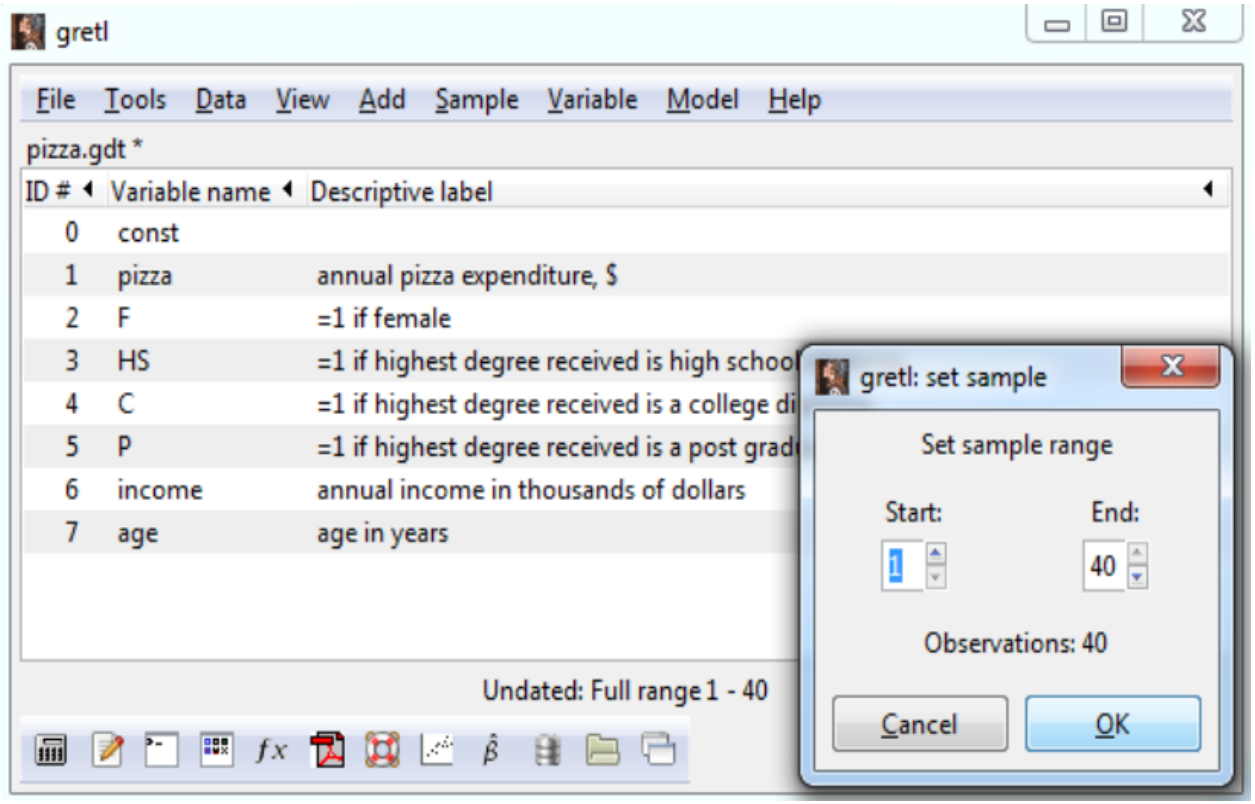
Example 5.1.1. Ordinary Least Squares.

To restrict a sample, go up to the **Sample** pulldown menu where several criteria are offered to restrict the sample.



Example 5.1.1. Ordinary Least Squares.

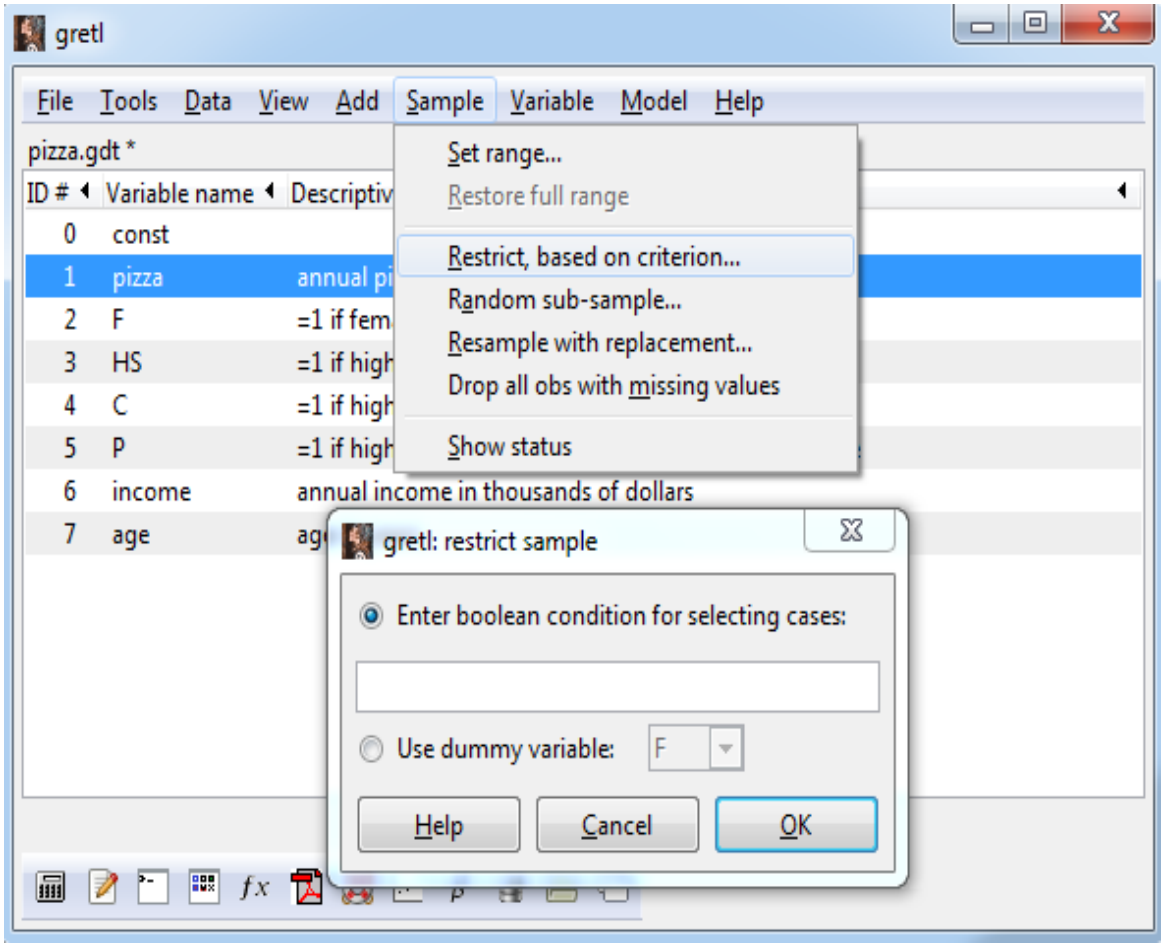
For instance, the option *Set range* allows to select a subsample of consecutive observations. It is necessary to select the first observation and the last one. This is not the option of interest in this example.



Example 5.1.1. Ordinary Least Squares.

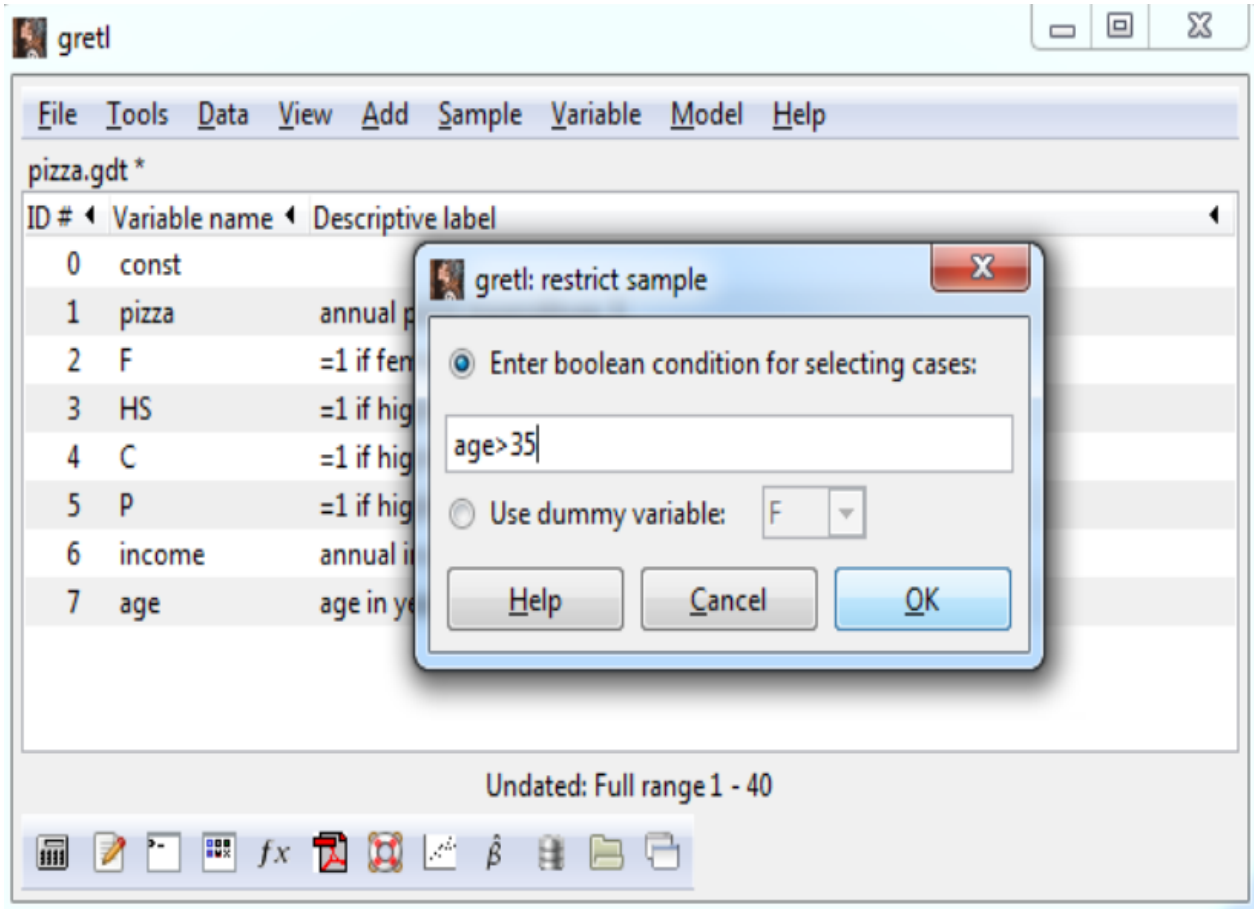
The option *Restrict, based on criterion ...* offers two criteria:

1) Enter boolean condition for selecting cases.
2) Use dummy variable.



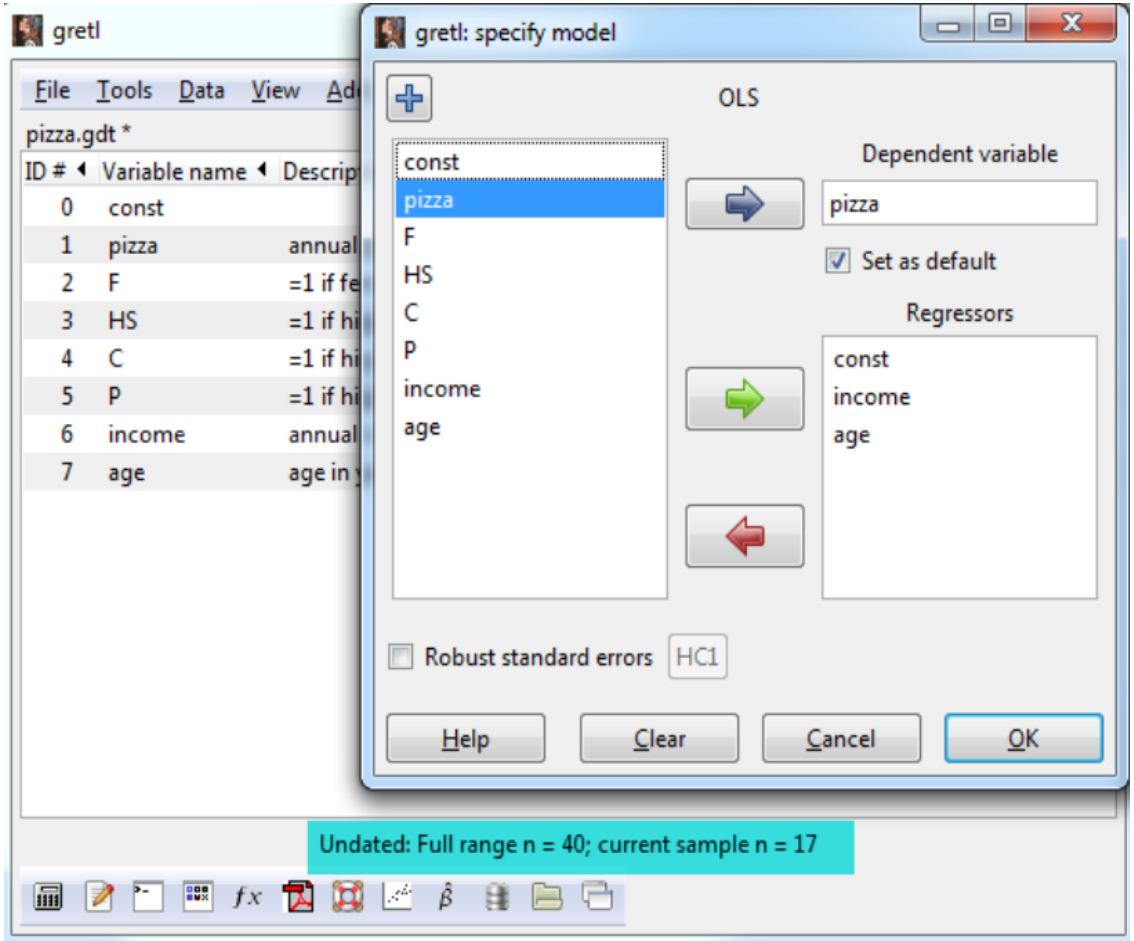
Example 5.1.1. Ordinary Least Squares.

Using the first option, it is possible to establish a criterion such as clients older than 35 years old.



Example 5.1.1. Ordinary Least Squares.

Estimate the model using the restricted sample of size 17.



Example 5.1.1. Ordinary Least Squares.

Estimation results using the restricted sample.

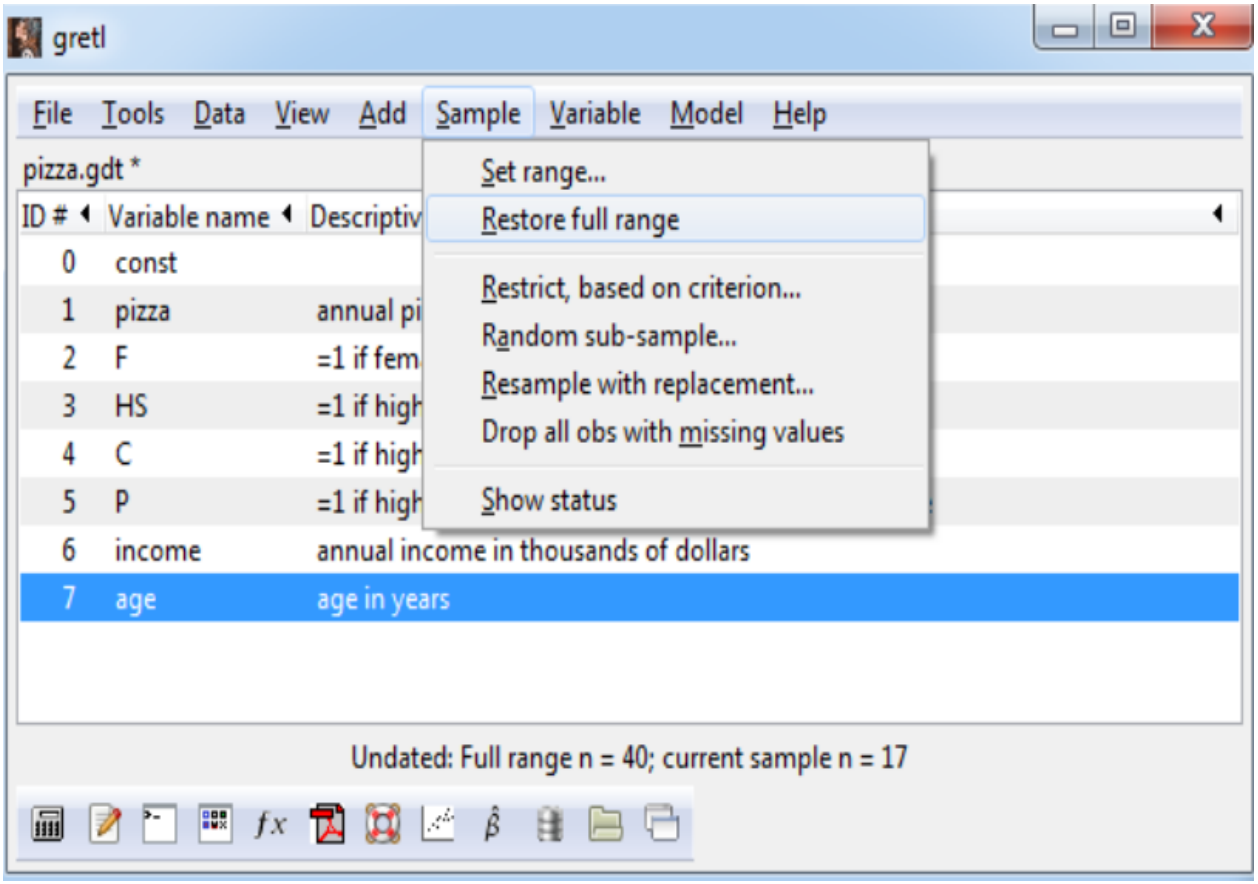
gretl: model 2

File Edit Tests Save Graphs Analysis LaTeX

```
Model 2: OLS, using observations 1-17
Dependent variable: pizza

              coefficient   std. error   t-ratio   p-value
  ---------------------------------------------------------
  const        299.969      247.180       1.214    0.2450
  income         1.59059      0.499671    3.183    0.0066   ***
  age           -6.22178      5.57067    -1.117    0.2828

Mean dependent var   161.3529   S.D. dependent var   162.9540
Sum squared resid    236193.4   S.E. of regression   129.8883
R-squared            0.444073   Adjusted R-squared   0.364655
F(2, 14)             5.591574   P-value(F)           0.016411
Log-likelihood      -105.2051   Akaike criterion     216.4102
Schwarz criterion    218.9098   Hannan-Quinn         216.6587
```

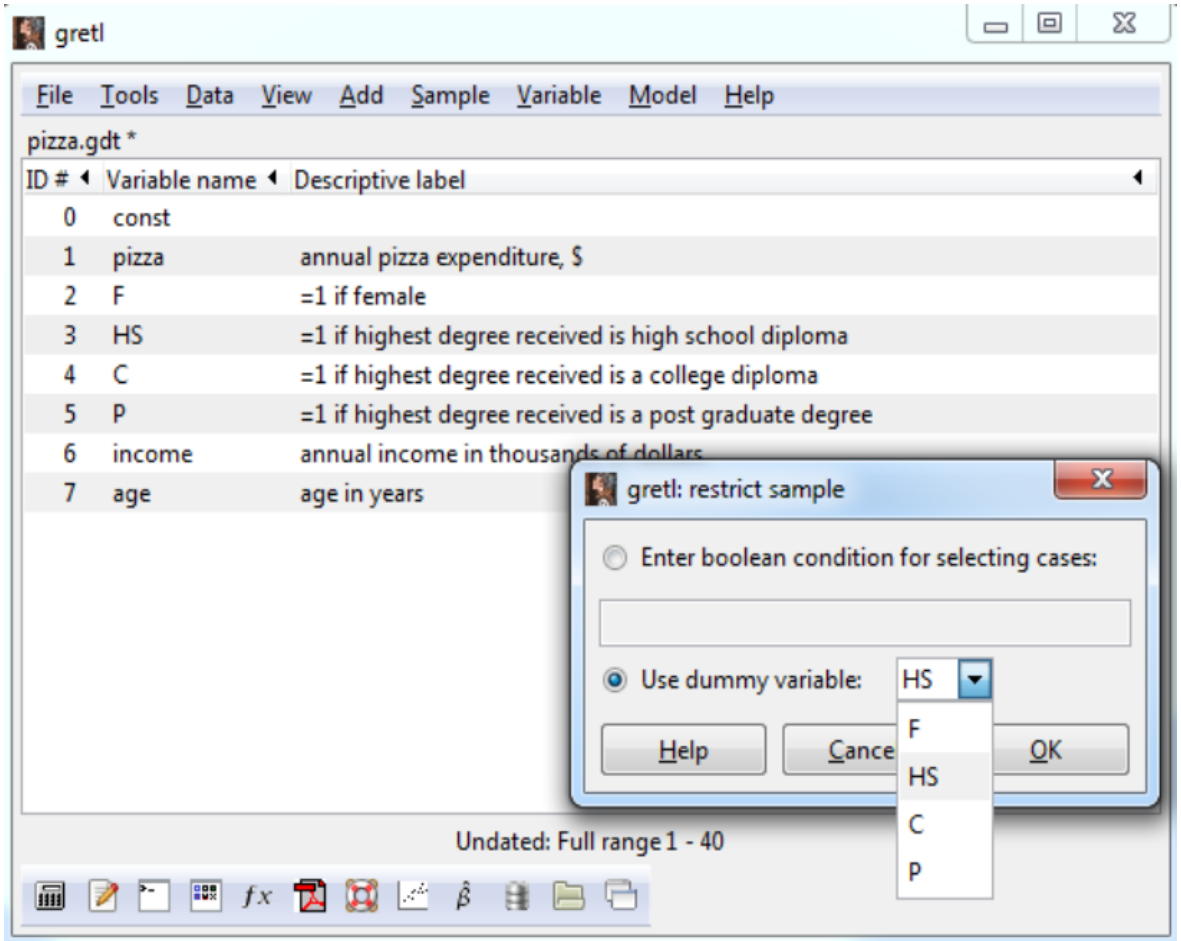
Example 5.1.1. Ordinary Least Squares.

To return to the full sample, select the option *Restore full range* in the **Sample** pulldown menu.



Example 5.1.1. Ordinary Least Squares.

The second option of *Restrict based on criterion ...* allows us to restrict the sample using a criterion based on a dummy variable, in this case, HS (maximum level of studies is high school).



Example 5.1.1. Ordinary Least Squares.

Estimation results using the restricted sample of size 15.

gretl: model 3

File Edit Tests Save Graphs Analysis LaTeX

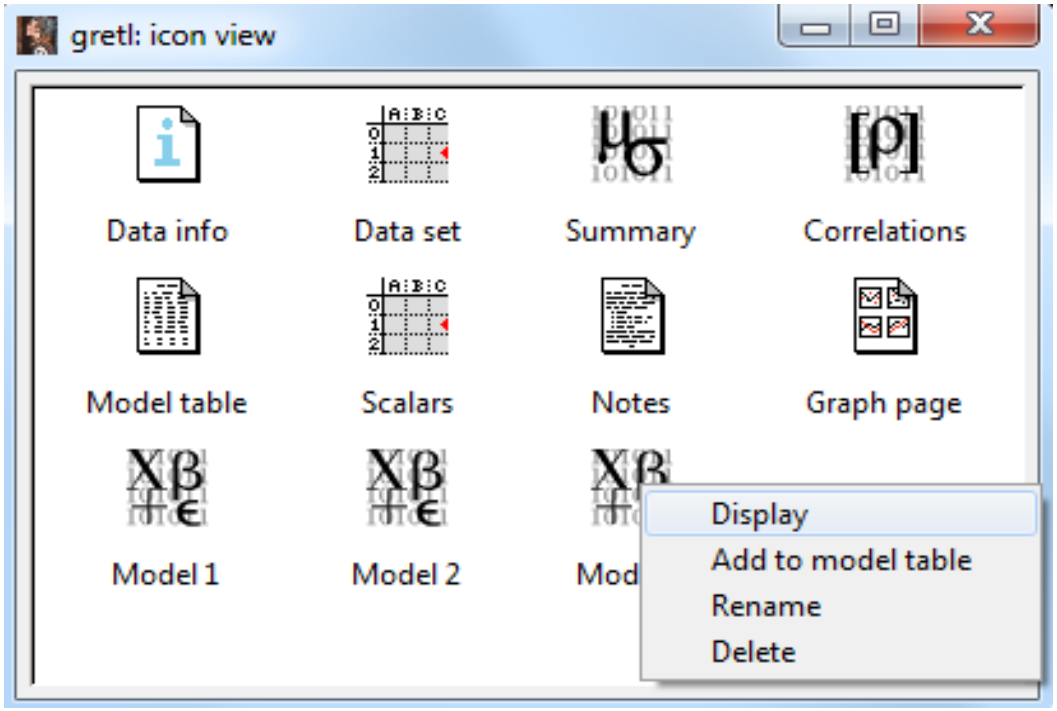
Model 3: OLS, using observations 1-15
Dependent variable: pizza

	coefficient	std. error	t-ratio	p-value	
const	352.069	115.456	3.049	0.0101	**
income	3.90353	1.44769	2.696	0.0194	**
age	-8.56052	3.46725	-2.469	0.0295	**

Mean dependent var	212.4667	S.D. dependent var	155.7245
Sum squared resid	186169.9	S.E. of regression	124.5558
R-squared	0.451638	Adjusted R-squared	0.360244
F(2, 12)	4.941677	P-value(F)	0.027190
Log-likelihood	-91.98181	Akaike criterion	189.9636
Schwarz criterion	192.0878	Hannan-Quinn	189.9410

Example 5.1.1. Ordinary Least Squares.

If the three estimated models have been saved as icons, there is one icon for each model in the session. Right-clicking on these icons, you may retrieve the estimation results.



Do not forget to save the session with the name pizza5.1.1!

Example 5.1.1. Ordinary Least Squares.

Results I. Full sample.

SRF $$\widehat{pizza}_i = 342.885 + 1.8324\ income_i - 7.57556\ age_i \quad i = 1, \ldots, 40$$

- Interpretation of the estimated coefficients:

$\hat{\beta}_1$: The estimated pizza consumption for a newborn baby with no income amounts to \$342.885.

$\hat{\beta}_2$: It is estimated that pizza consumption increases by \$1.8324 when annual income increases by \$1000 holding age constant.

$\hat{\beta}_3$: It is estimated that pizza consumption decreases by \$7.57556 when age increases by one year holding annual income constant.

- Coefficient of determination:

R^2: 32.9251 % of the total sample variation in pizza consumption is explained by the variations in annual income and age.

Example 5.1.1. Ordinary Least Squares.

Results II. Clients over 35 years old.

SRF $$\widehat{pizza}_i = 299.969 + 1.59059\ income_i - 6.22178\ age_i \quad i = 1, \ldots, 17$$

- Interpretation of the estimated coefficients:

$\hat{\beta}_2$: It is estimated that the amount of pizza consumed by the clients over 35 years old increases by \$1.59059 when income increases by \$1000, holding age fixed.

$\hat{\beta}_3$: It is estimated that the amount of pizza consumed by the clients over 35 years old decreases by \$6.22178 when age increases by 1 year, holding income fixed.

- Coefficient of determination:

R^2: 44.4073 % of the sample variation in pizza consumption for the clients over 35 years old is explained by the variations in annual income and age.

Example 5.1.1. Ordinary Least Squares.

Results III. Maximum level of studies: high school.

SRF $\quad \widehat{pizza}_i = 352.069 + 3.90353\ income_i - 8.56052\ age_i \quad i = 1, \ldots, 15$

- Interpretation of the estimated coefficients:

$\hat{\beta}_2$: It is estimated that the amount of pizza consumed by the clients with only secondary studies increases by \$3.90353 when income increases by \$1000, holding age fixed.

$\hat{\beta}_3$: It is estimated that the amount of pizza consumed by the clients with only secondary studies decreases by \$8.56052 when age increases by one year, holding income fixed.

- Coefficient of determination:

R^2: 45.1638 % of the sample variation in pizza consumption for the clients with only high school education is explained by the variation in annual income and age

Contents

Example 5.1.2. Exploring the estimation results.

Pizza consumption.

Open the session pizza5.1.1.

a. Estimate a regression model to determine pizza consumption as a function of income and age that enables us to measure the following effect:

   The expected change in pizza consumption due to a change in income depends on the age of the client

b. Save all the results to the session as icons.

c. Explore the **Save** command in the menu bar of the estimation results window.

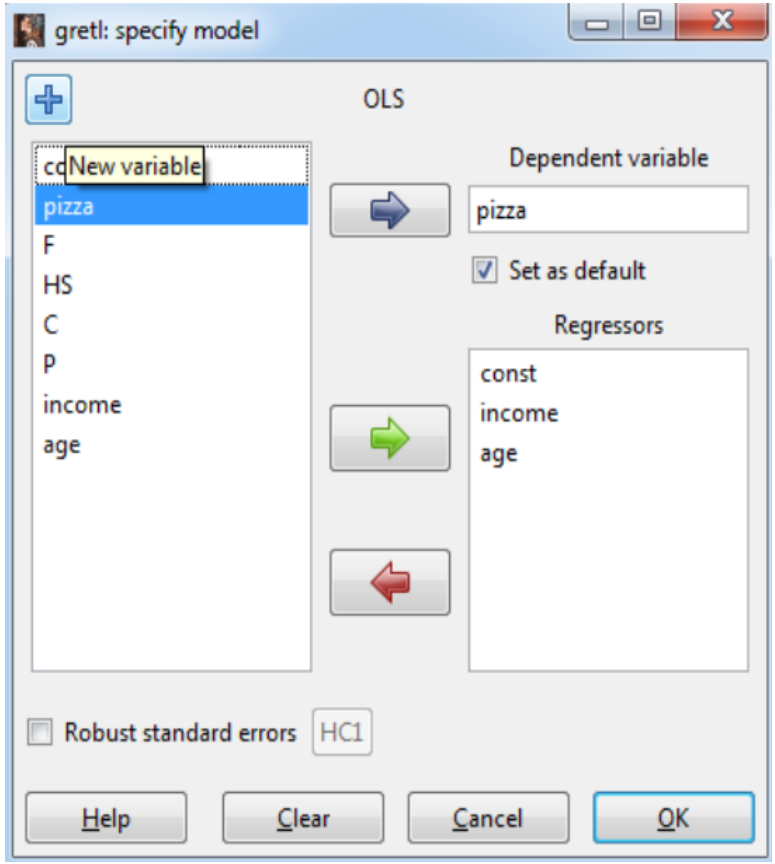
d. Obtain the covariance matrix of the OLS estimators of the regression coefficients.

e. Interpret the results and save the session under the name pizza5.1.2.

Example 5.1.2. Exploring the estimation results.

`Model -> Ordinary Least Squares ...`

This command yields the dialog box shown below.

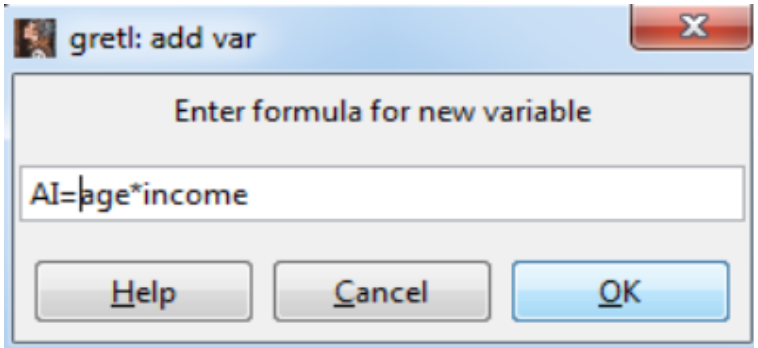


Example 5.1.2. Exploring the estimation results.

$$pizza_i = \beta_1 + \beta_2\, income_i + \beta_3\, age + \beta_4\, (age_i \times income_i) + u_i$$

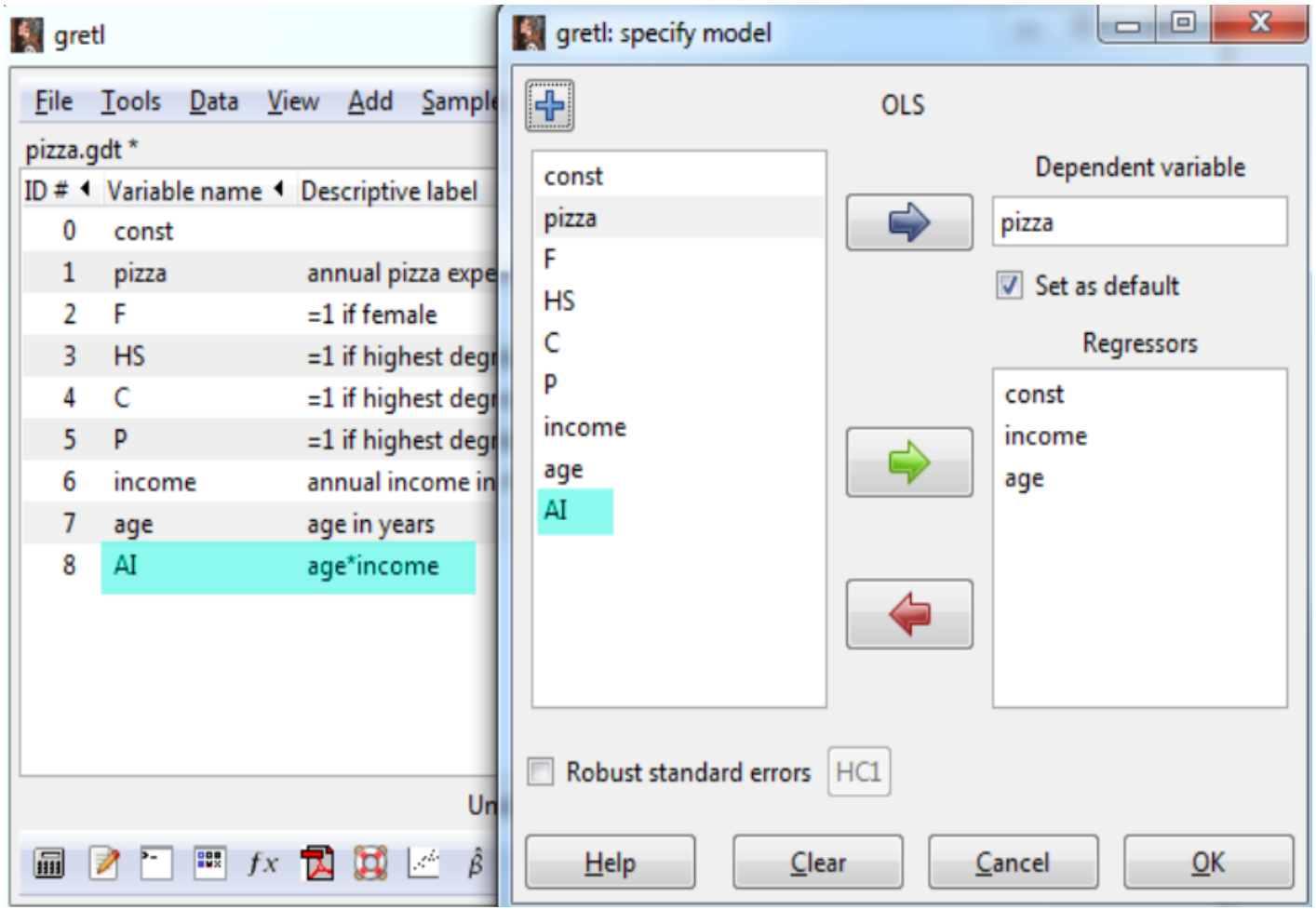
The effect "the expected change in pizza consumption due to a change in income depends on the age of the client" is included in the specification of the model by means of a new term, which is an interaction between income and age.

Therefore, prior to the estimation of the model, it is necessary to generate this interaction term $(age_i \times income_i)$ clicking on the icon $+$ located in the upper left side of the `gretl:specify model` window. Then you define the new variable in the dialog box.



Example 5.1.2. Exploring the estimation results.

The new variable appears both in the main window and in the `gretl:specify model` window.



Example 5.1.2. Exploring the estimation results.

Estimation results.

gretl: model 4

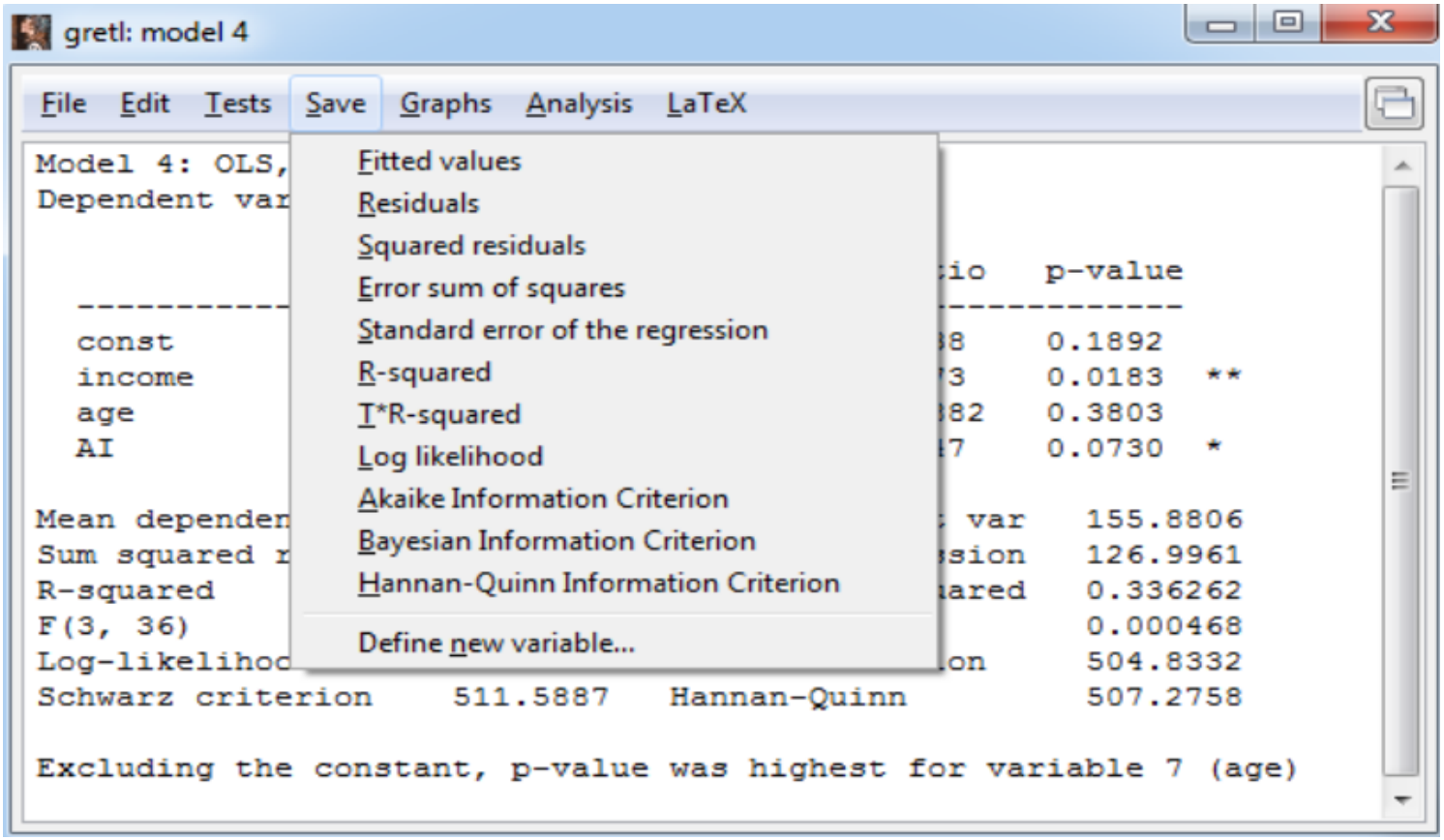
File Edit Tests Save Graphs Analysis LaTeX

```
Model 4: OLS, using observations 1-40
Dependent variable: pizza

             coefficient   std. error    t-ratio   p-value
  ------------------------------------------------------------
  const      161.465       120.663        1.338    0.1892
  income       6.97991       2.82277      2.473    0.0183   **
  age         -2.97742       3.35210     -0.8882   0.3803
  AI          -0.123239      0.0667187   -1.847    0.0730   *

Mean dependent var   191.5500   S.D. dependent var   155.8806
Sum squared resid    580608.7   S.E. of regression   126.9961
R-squared            0.387319   Adjusted R-squared   0.336262
F(3, 36)             7.586038   P-value(F)           0.000468
Log-likelihood      -248.4166   Akaike criterion     504.8332
Schwarz criterion    511.5887   Hannan-Quinn         507.2758

Excluding the constant, p-value was highest for variable 7 (age)
```

Save the results to the session as an icon.

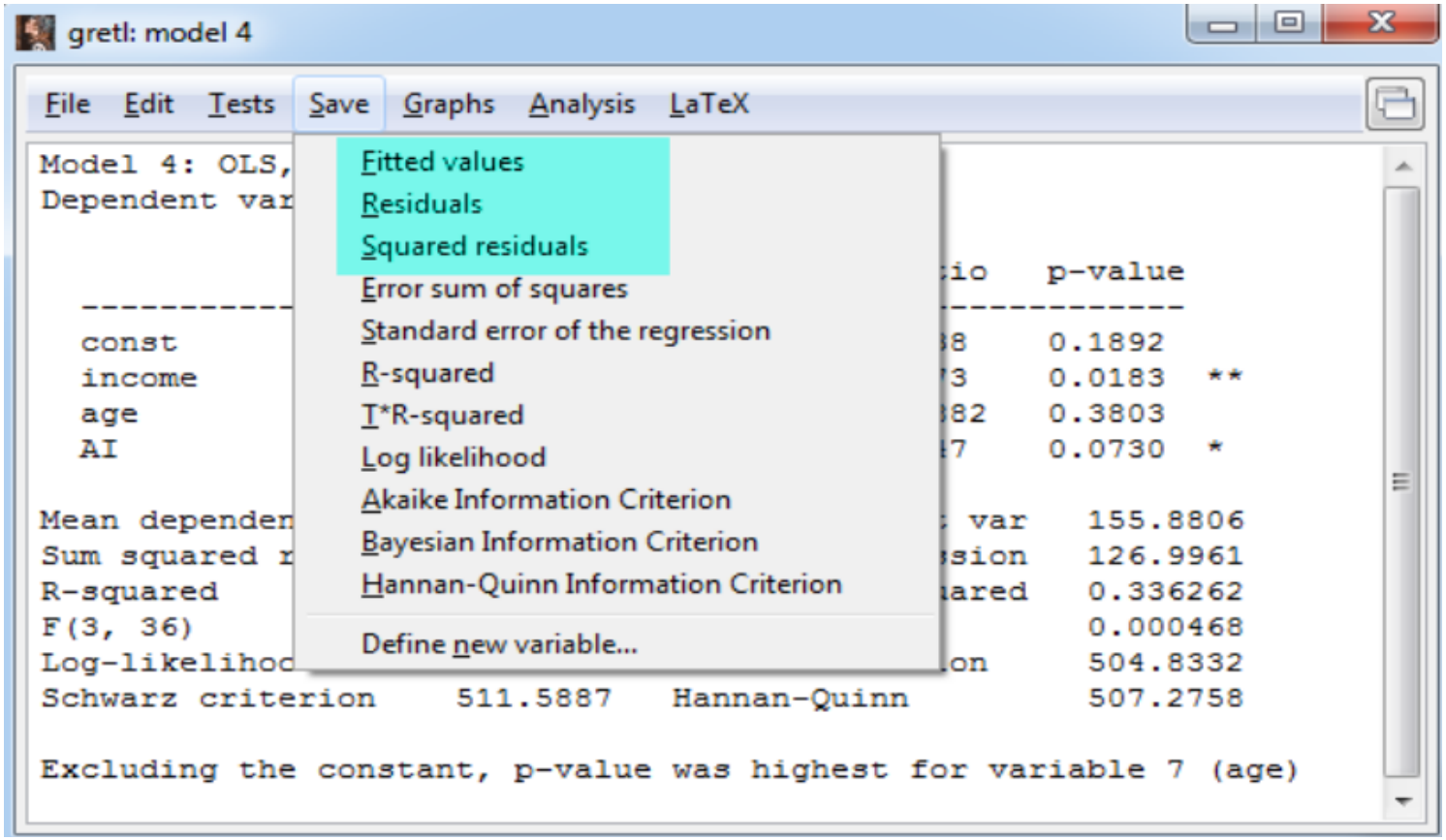
Example 5.1.2. Exploring the estimation results.

To save the estimation results use the command **Save** in the menu bar of the estimation results window.



Example 5.1.2. Exploring the estimation results.

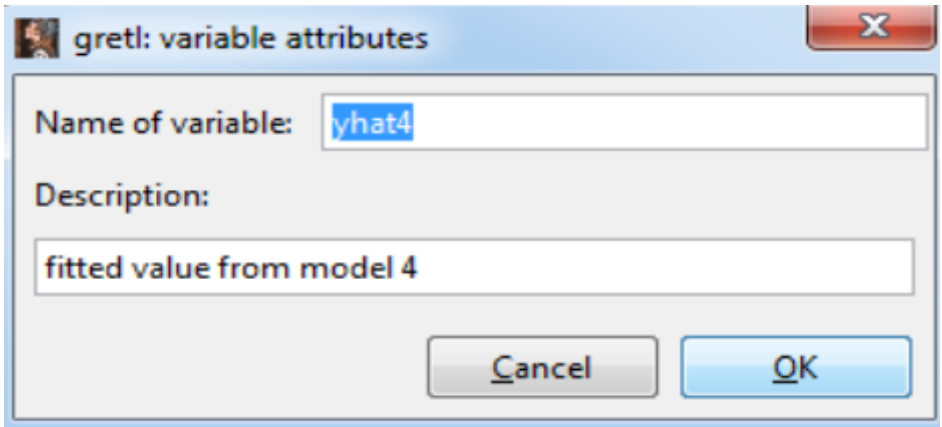
Let's look at the first three option of the **Save** pulldown menu.



Example 5.1.2. Exploring the estimation results.

- Fitted values $\left\{\widehat{pizza}_i\right\}_{i=1}^{N=40}$

`Save -> Fitted values` yields the dialog box below.

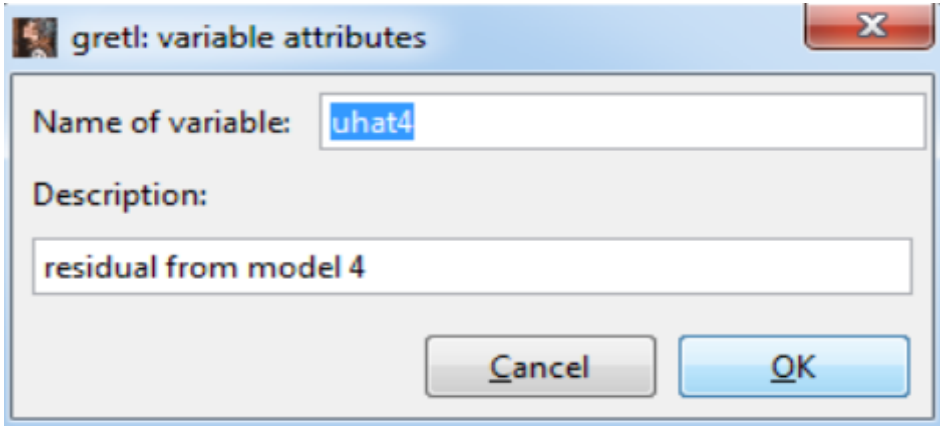


The fitted series will be named by default *yhat4* where the cardinal "4" indicates the specific estimated model, in this case the fourth one. The name and the description of this variable may be changed.

Example 5.1.2. Exploring the estimation results.

- Residuals $\{\hat{u}_i\}_{i=1}^{N=40}$

`Save -> Residuals` yields the dialog box below.

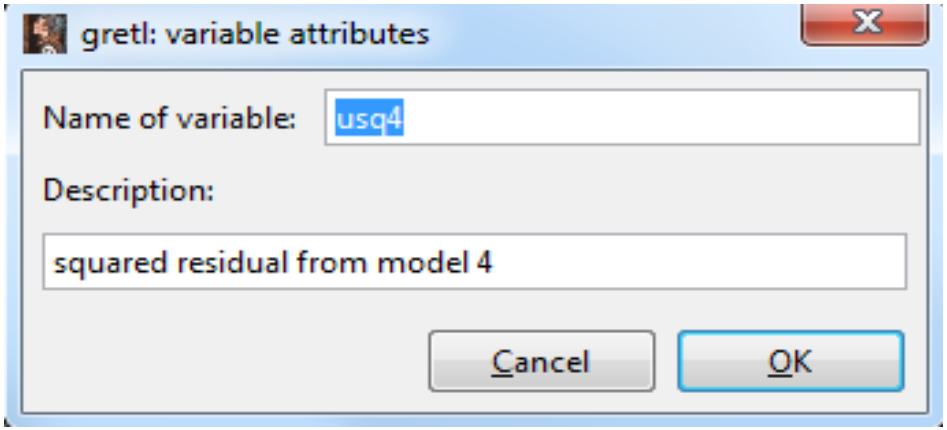


The residuals will be named by default *uhat4* where the cardinal "4" indicates the specific estimated model, in this case the fourth one. The name and the description of this variable may be changed.

Example 5.1.2. Exploring the estimation results.

- Squared residuals $\left\{\hat{u}_i^2\right\}_{i=1}^{N=40}$

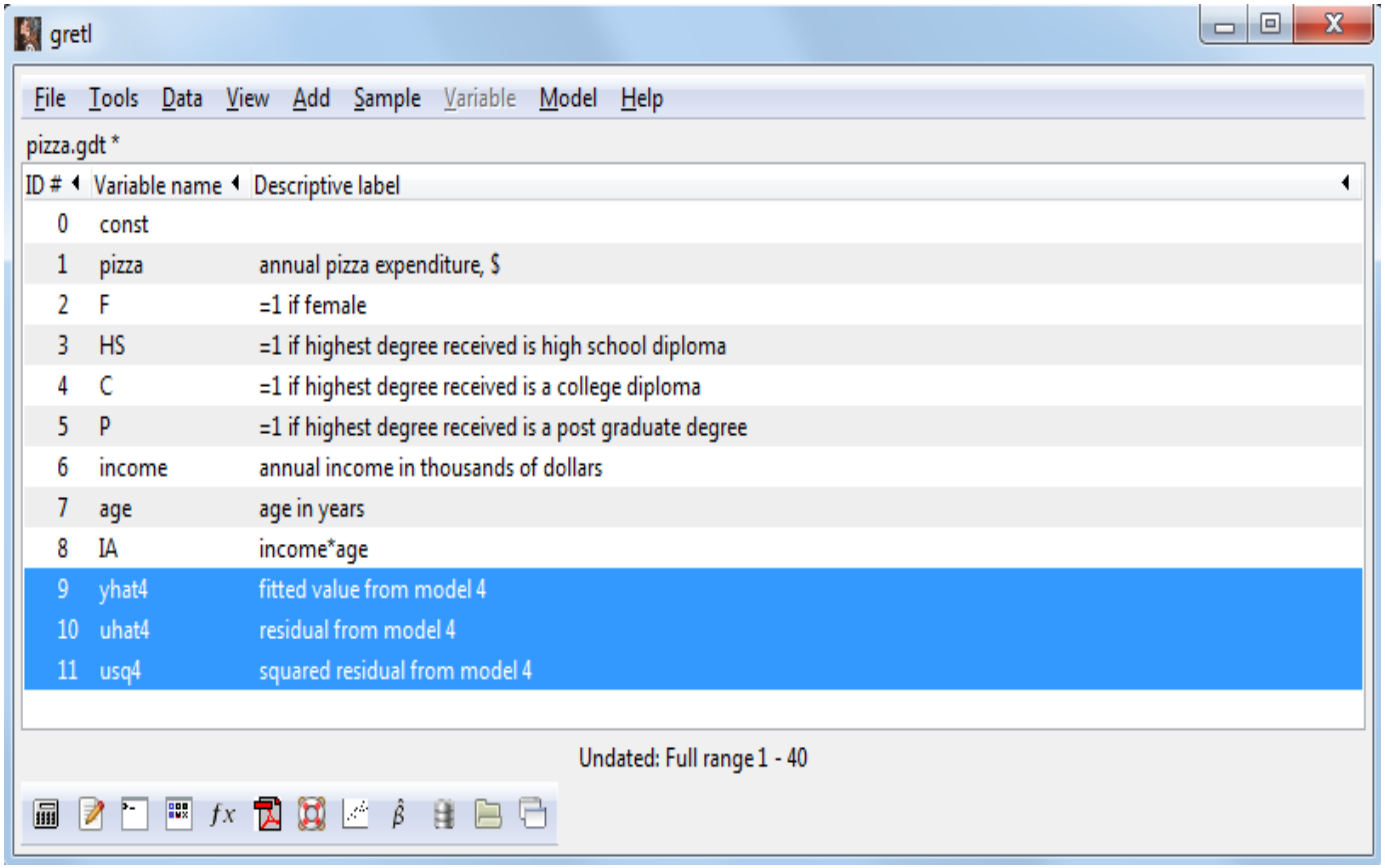
`Save -> Squared residuals` yields the dialog box below.



The squared residuals will be named by default *usq4* where the cardinal "4" indicates the specific estimated model, in this case the fourth one. The name and the description of this variable may be changed.

Example 5.1.2. Exploring the estimation results.

The saved variables *yhat4*, *uhat4* and *usq4* appear in the main window of Gretl in the same order they have been generated. Furthermore they are included as well in the icon *Data set* of the session.



Example 5.1.2. Exploring the estimation results.

You may save the rest of the statistics that appear in the **Save** pulldown menu in a similar way.

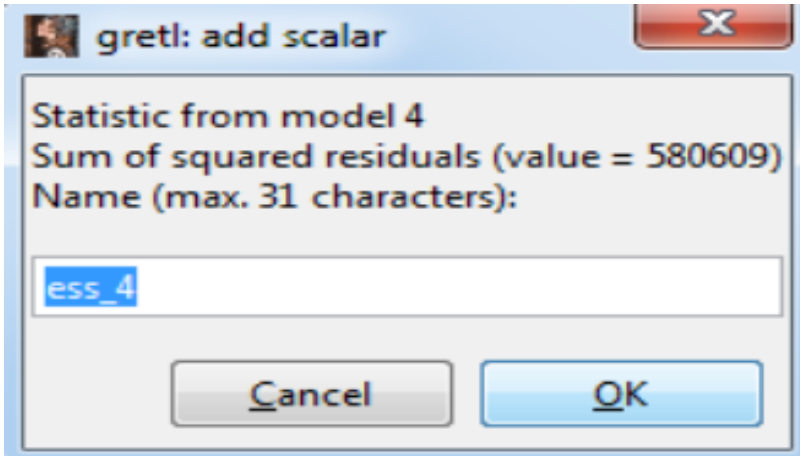
These are the names assigned by default:

- Error sum of squares: ess_#
- Standard error of the regression: sigma_#
- Coefficient of determination, R-squared: rsq_#
- T*R-squared: trsq_#
- Log likelihood: lnl_#
- Akaike Information Criterion: aic_#
- Bayesian Information Criterion: bic_#
- Hannan-Quinn Information Criterion: hqc_#

where # stands for the number of the estimated model.

Example 5.1.2. Exploring the estimation results.

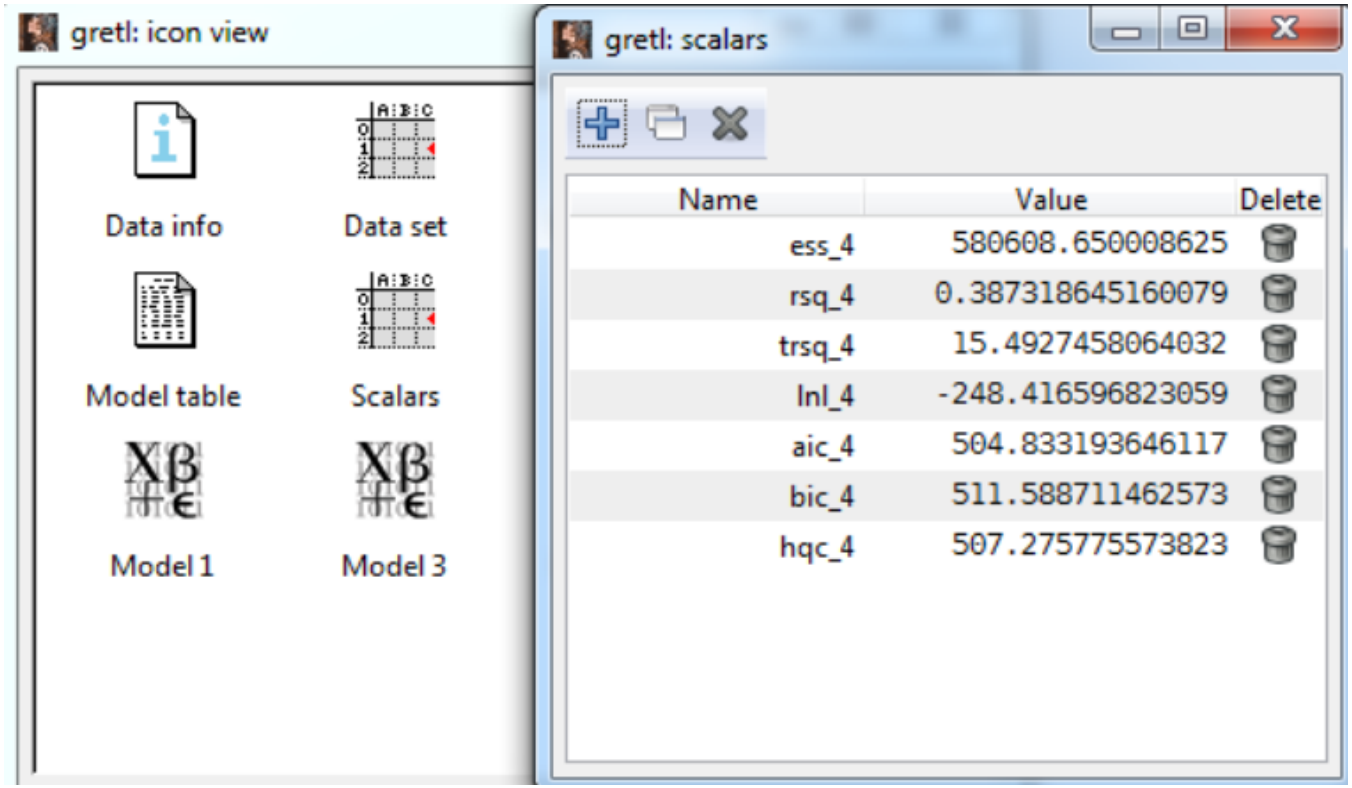
The **Save** pulldown menu includes some goodness-of-fit measures and the information criteria. For instance, you may save the sum of squared residuals. The assigned name by default can be changed.



The rest of the values are saved in a similar way.

Example 5.1.2. Exploring the estimation results.

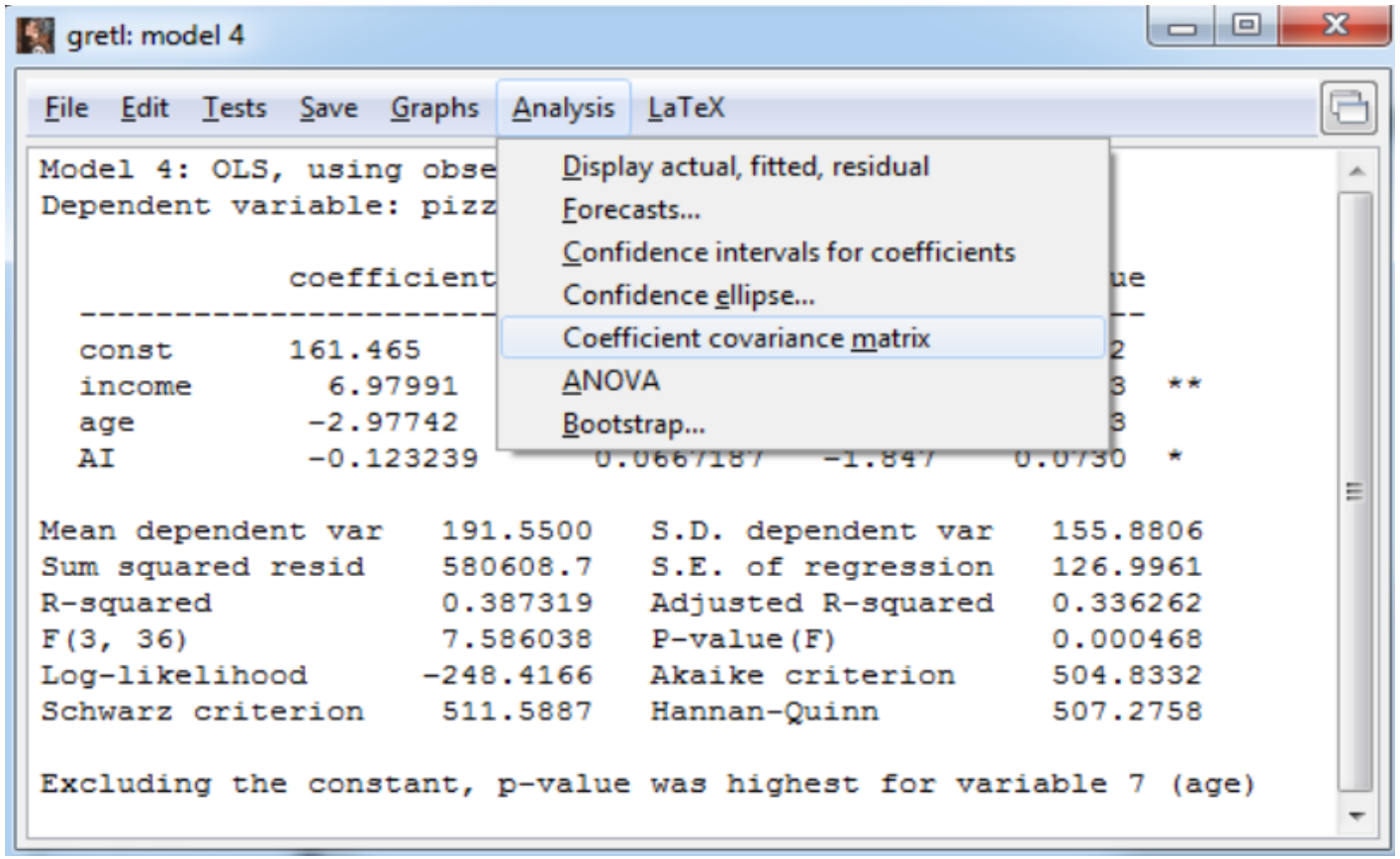
All the values saved may be retrieved from the session, double-clicking on the icon *Scalars*.



Example 5.1.2. Exploring the estimation results.

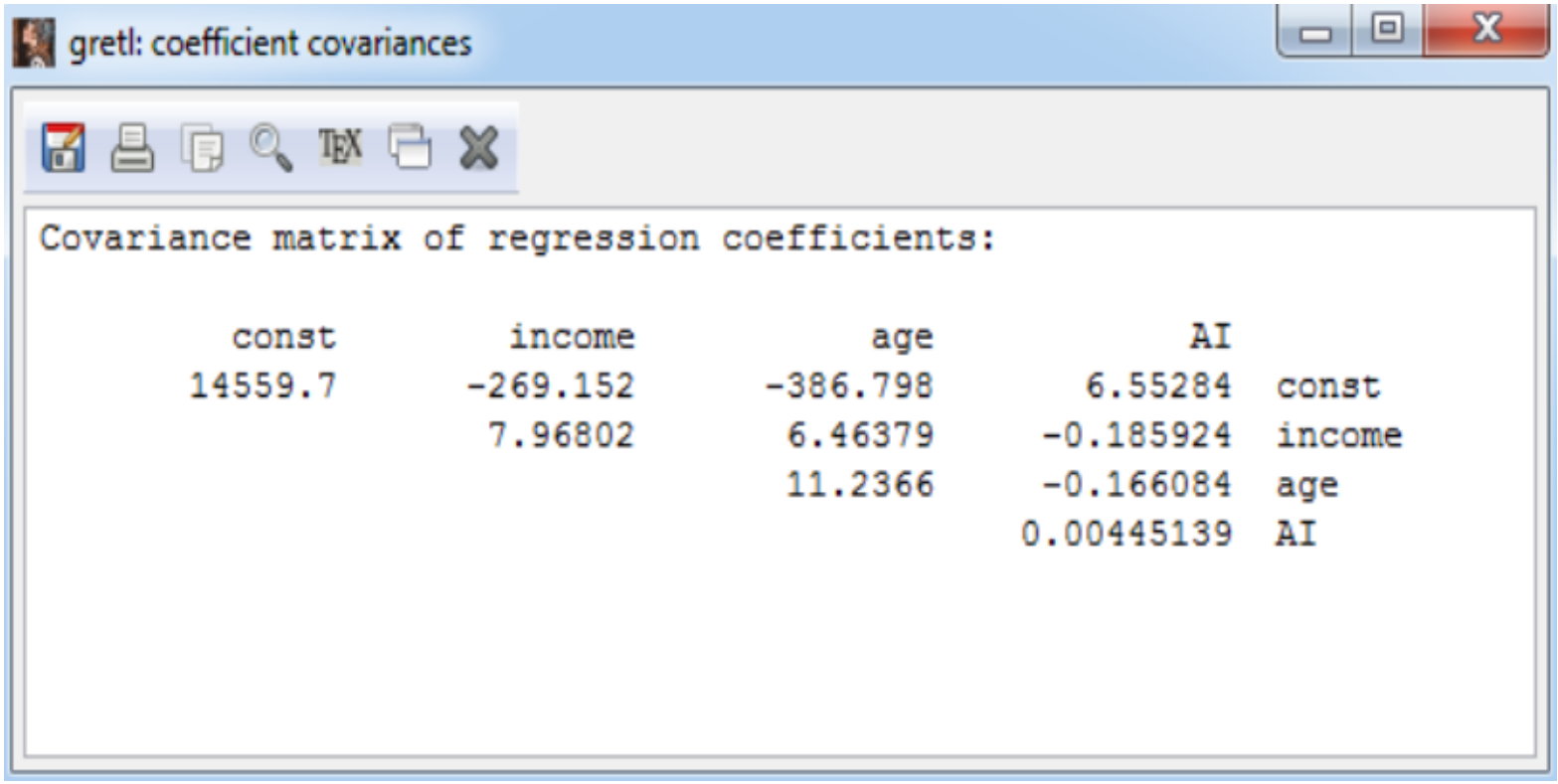
To estimate the covariance matrix of the OLS estimator of the coefficients, select the option *Coefficient covariance matrix* in the **Analysis** pulldown menu of the estimation results window.

`Analysis -> Coefficient covariance matrix`



Example 5.1.2. Exploring the estimation results.

The results are shown in the figure below. Since the covariance matrix is symmetric only the upper half of the matrix appears.



Example 5.1.2. Exploring the estimation results.

Results (I).

$$\widehat{pizza_i} = 161.465 + 6.979991\, income_i - 2.97742\, age - 0.1232239\, (age_i \times income_i)$$

- The estimated pizza consumption amounts to \$161.465 when the variables annual income and age take the value zero.

- Marginal effects:

- Income marginal effect. It is estimated that the pizza consumption increases by $(6.979991 - 0.1232239 \times age_i)$ dollars when income increases by \$1000 holding age constant. This increment is not constant throughout the sample because it depends on the client's age: the older the client, the smaller the income marginal effect.
- Age marginal effect. It is estimated that the pizza consumption decreases by $(-2.97742 - 0.1232239\, income_i)$ dollars when the age increases by one year holding income constant. This increment is not constant throughout the sample because it depends on the income of the individual: the higher the income of the client, the more negative the age marginal effect.

Example 5.1.2. Exploring the estimation results.

Results (II).

- Coefficient of determination:

  R^2: 38.7319 % of the sample variation in pizza consumption is explained by the variations in annual income and age, according to this estimated regression model.

- Estimated covariance matrix of the OLS estimators:

$$\widehat{V}(\hat{\beta}) = \hat{\sigma}^2 (X'X)^{-1} = \begin{pmatrix} 14559.7 & -269.152 & -386.798 & 6.55284 \\ & 7.96802 & 6.46379 & -0.185924 \\ & & 11.2366 & -0.166084 \\ & & & 0.0045139 \end{pmatrix}$$

Contents

Example 5.1.3. Graphs.

Pizza consumption.

Open the session pizza5.1.2.

a. Estimate a regression model where pizza consumption depends on age and gender. In this model the effect of age on consumption should depend on gender.

b. Obtain all the residuals plots and save them as icons to the session and in pdf format.

c. Obtain all the plots of the fitted values and save them as icons to the session and in pdf format.

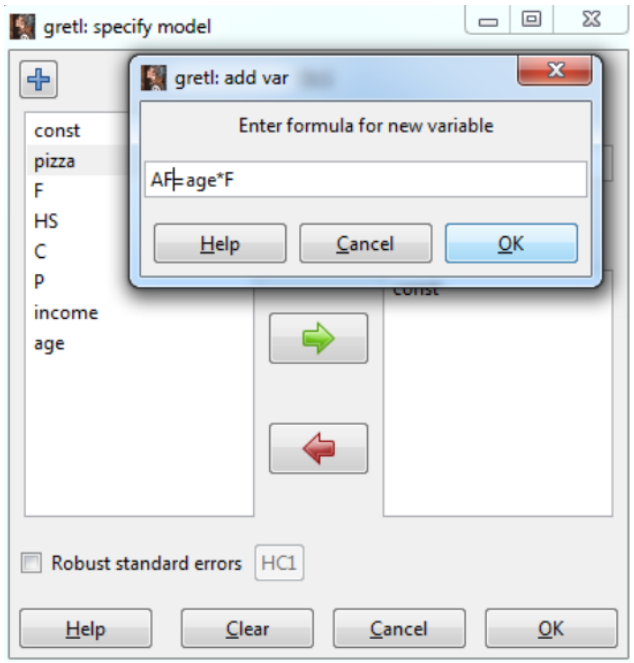
d. Interpret the results and save the session as pizza5.1.3.

Example 5.1.3. Graphs.

$$pizza_i = \beta_1 + \beta_2\, age_i + \beta_3\, F_i + \beta_4\, (age_i \times F_i) + u_i$$

`Model -> Ordinary Least Squares ...`

Then click on the icon [+] (upper left side of the specificiation window) to add the new regressor included in this model.



Example 5.1.3. Graphs.

Estimation results

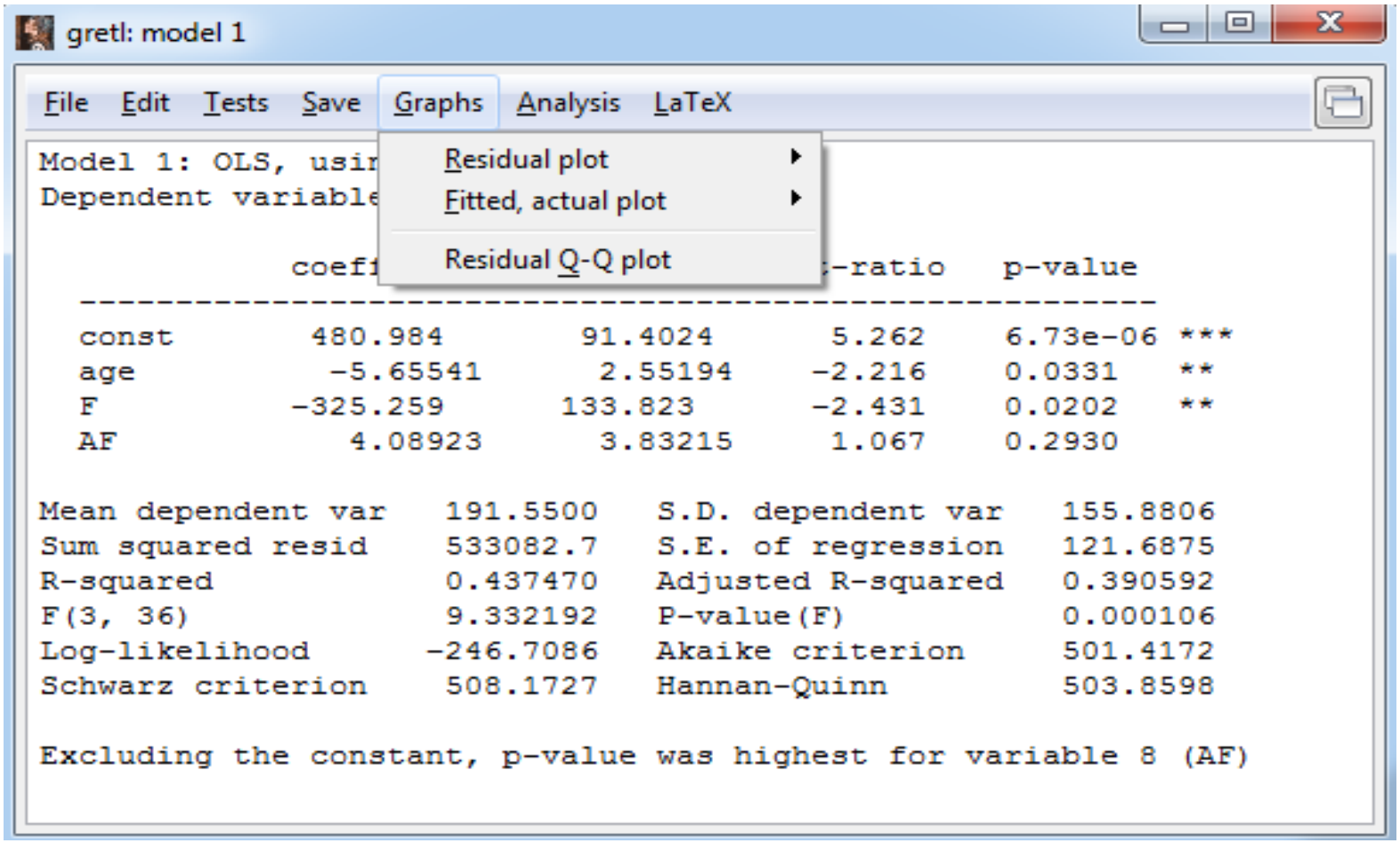
gretl: model 1

File Edit Tests Save Graphs Analysis LaTeX

```
Model 1: OLS, using observations 1-40
Dependent variable: pizza

              coefficient   std. error   t-ratio   p-value
  ---------------------------------------------------------
  const        480.984       91.4024      5.262    6.73e-06 ***
  age           -5.65541      2.55194    -2.216    0.0331   **
  F           -325.259      133.823      -2.431    0.0202   **
  AF             4.08923      3.83215      1.067    0.2930

Mean dependent var   191.5500   S.D. dependent var   155.8806
Sum squared resid    533082.7   S.E. of regression   121.6875
R-squared            0.437470   Adjusted R-squared   0.390592
F(3, 36)             9.332192   P-value(F)           0.000106
Log-likelihood      -246.7086   Akaike criterion     501.4172
Schwarz criterion    508.1727   Hannan-Quinn         503.8598

Excluding the constant, p-value was highest for variable 8 (AF)
```

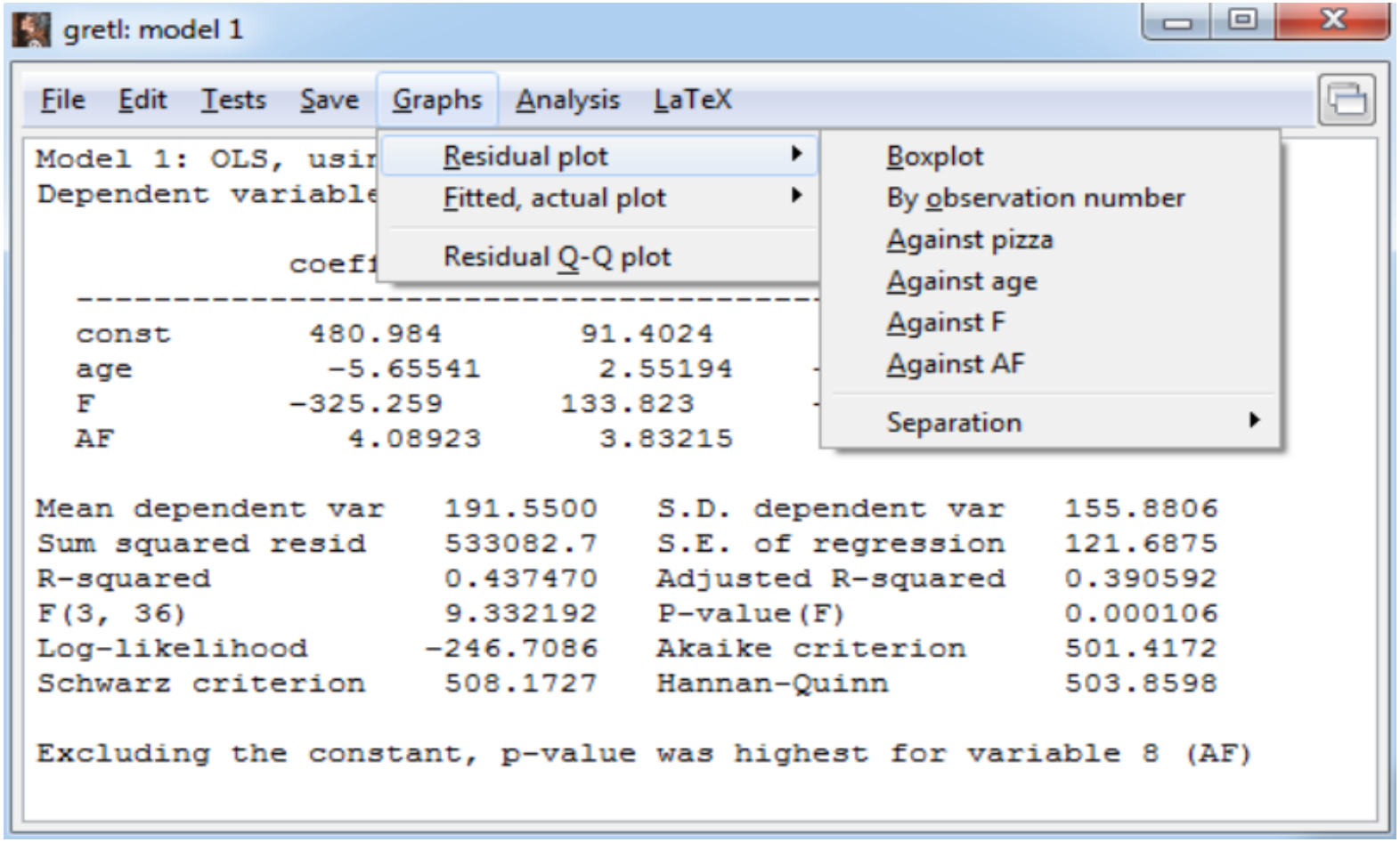
Example 5.1.3. Graphs.

Go to the **Graphs** option in the menu bar of the estimation output window to see the plots that can be obtained: the residuals plot, the fitted/actual values plot and the residual Q-Q plot.



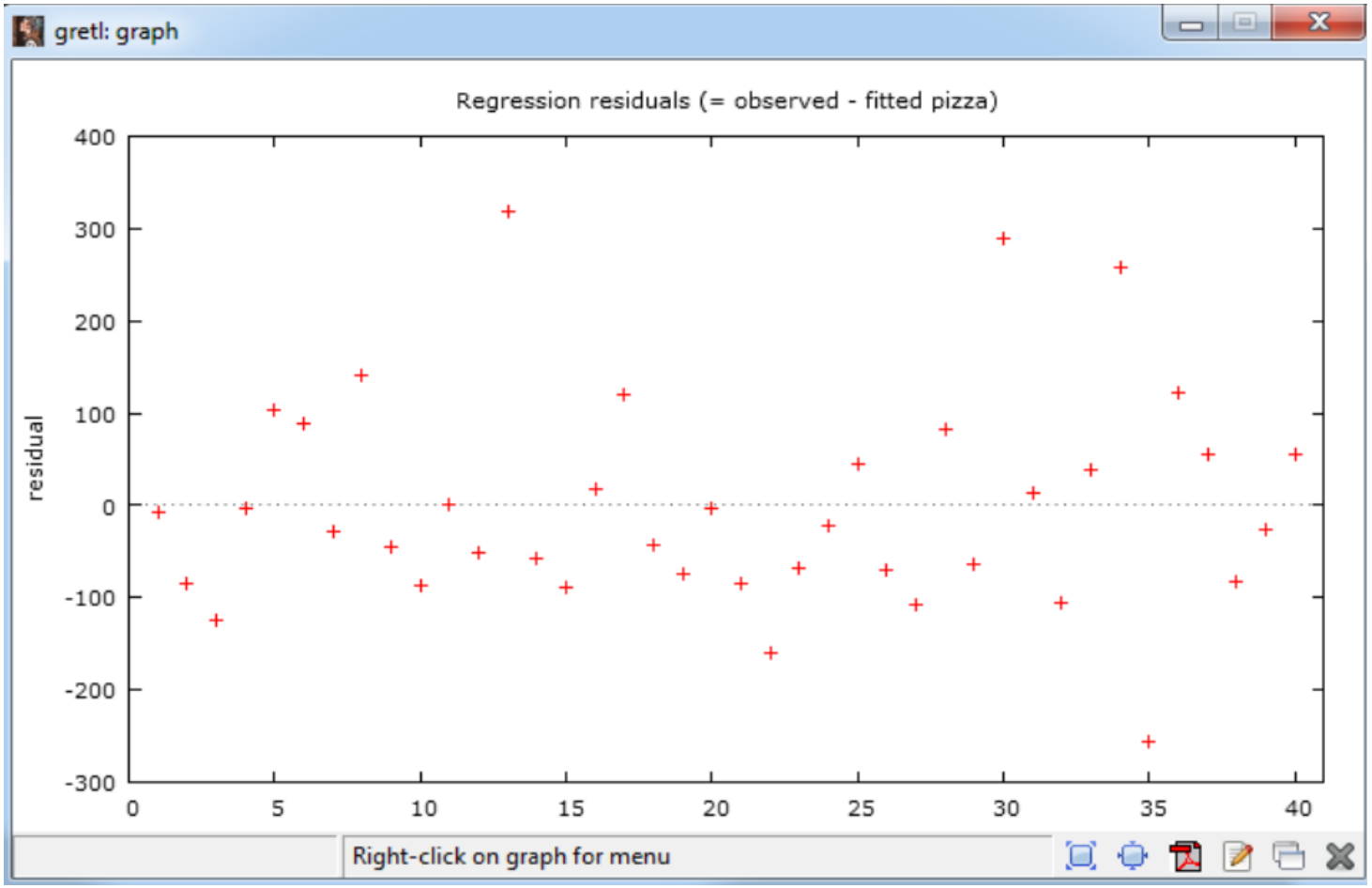
Example 5.1.3. Graphs.

With respect to the residuals plots there are several options.



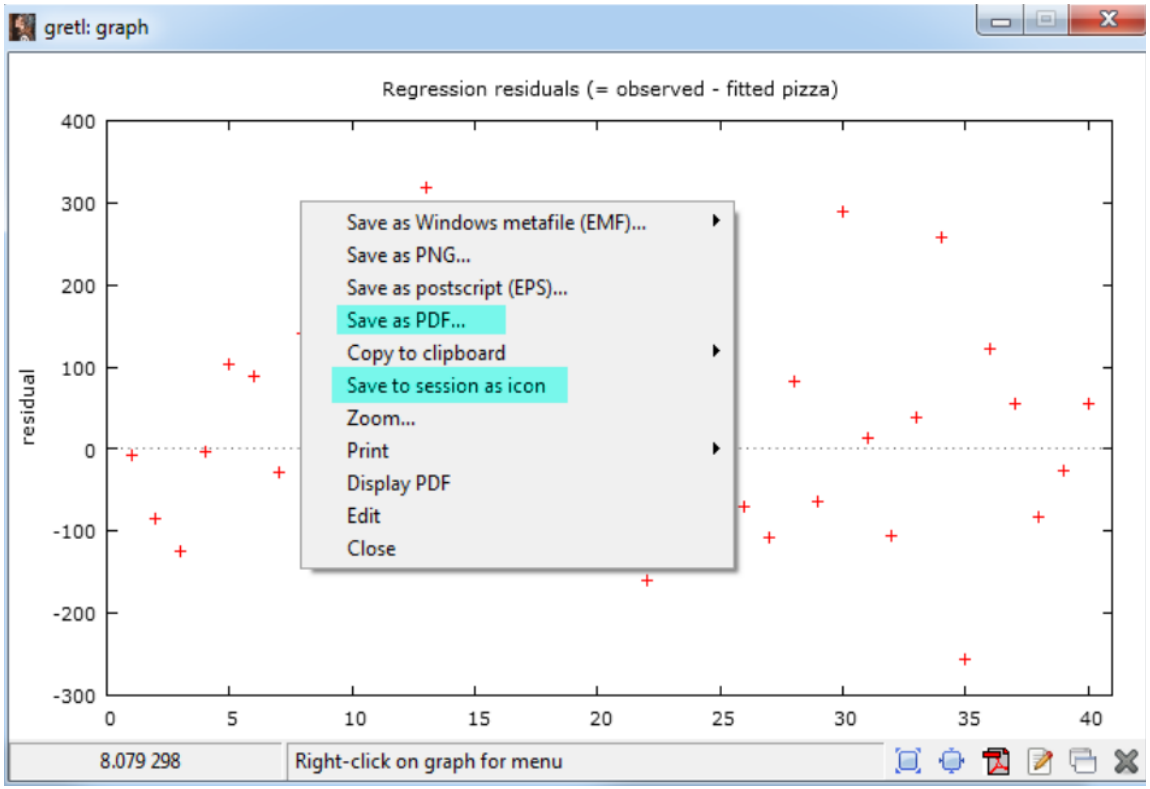
Example 5.1.3. Graphs.

Residual plot by observation number (RG1).



Example 5.1.3. Graphs.

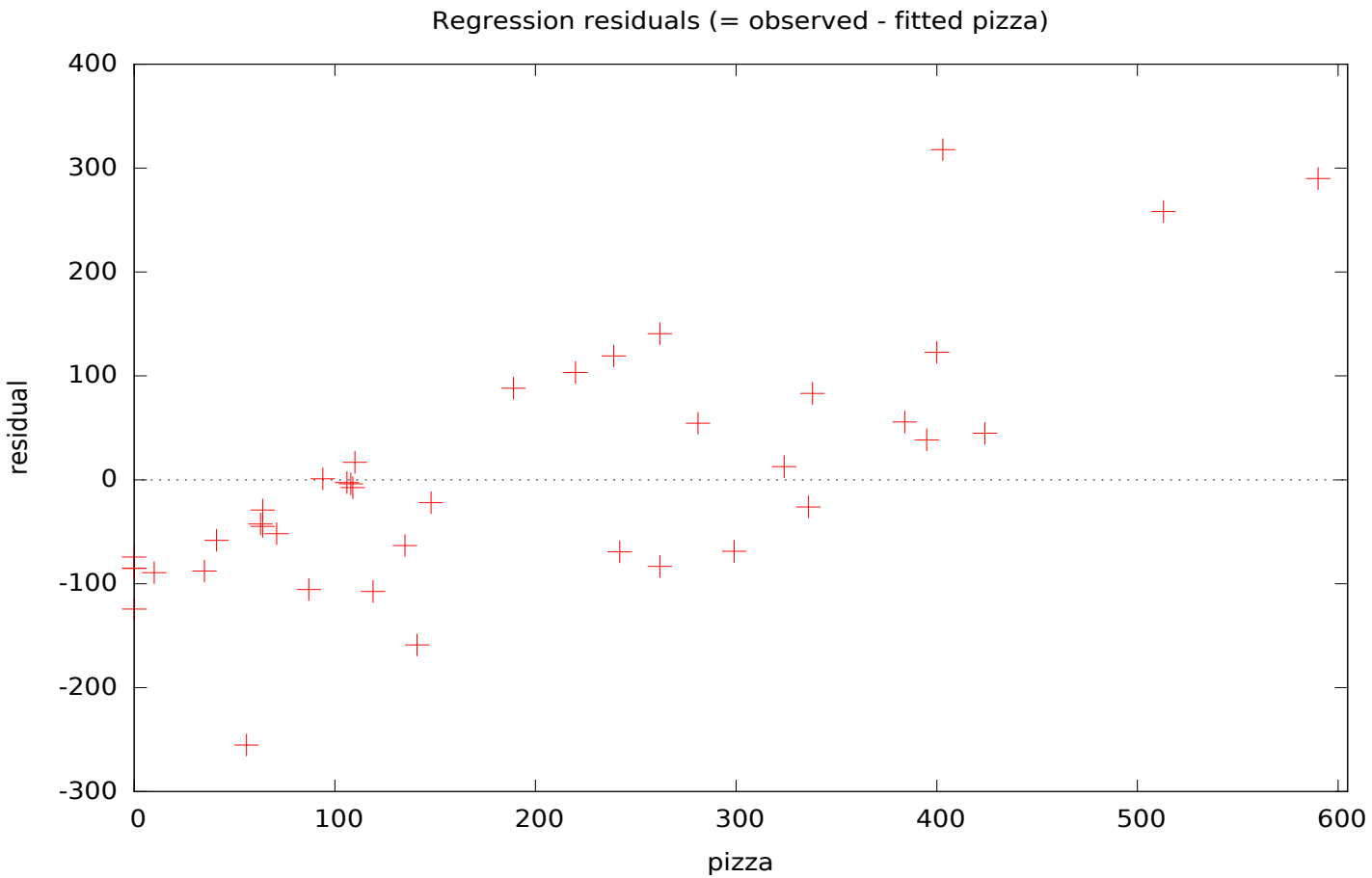
To save this plot as an icon, right-click on the plot and select the option *Save to session as icon*.



To save this plot in pdf format, right-click on the plot and select the option *Save as PDF ....*

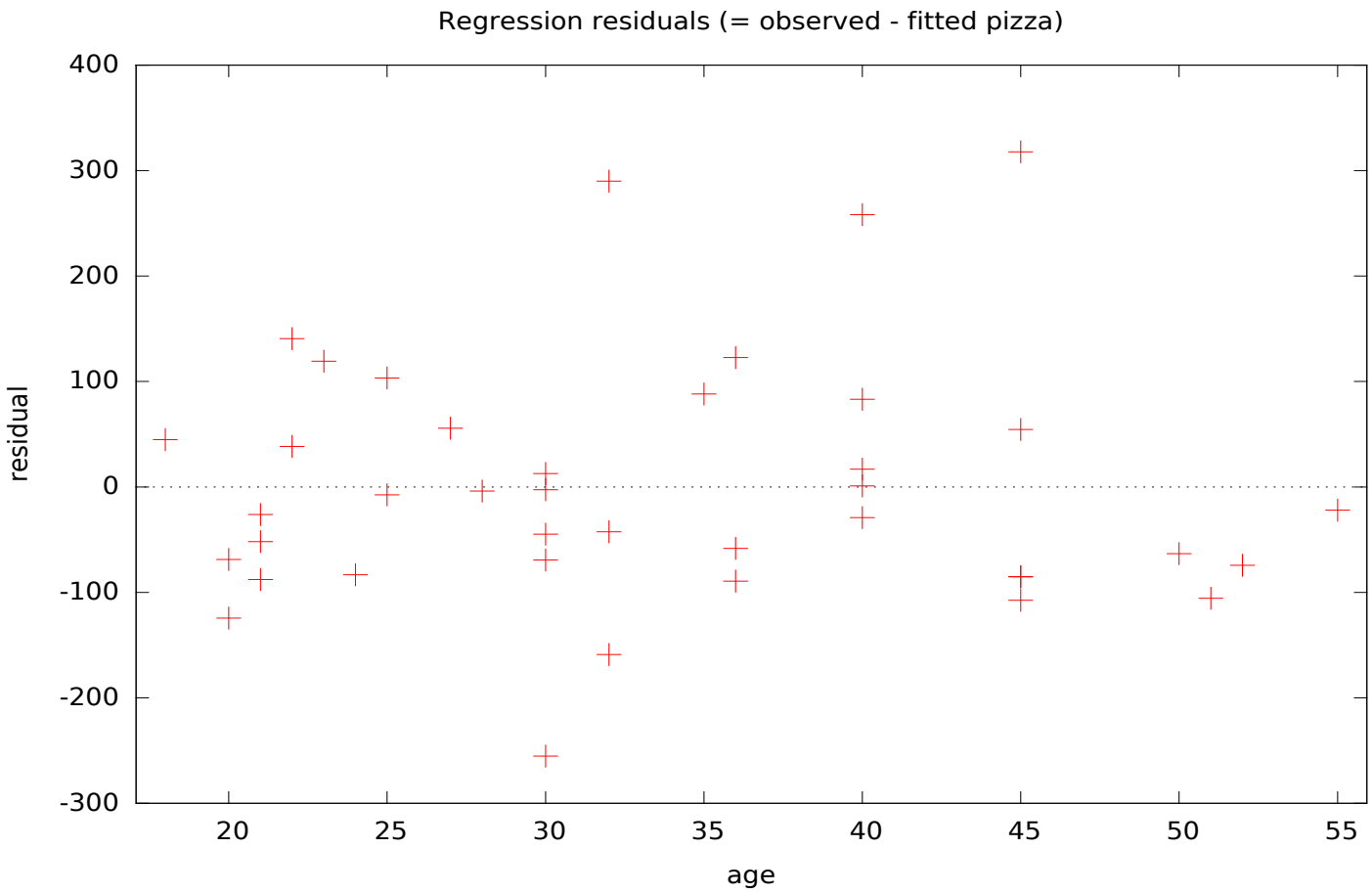
Example 5.1.3. Graphs.

Residual plot against the dependent variable (RG2).



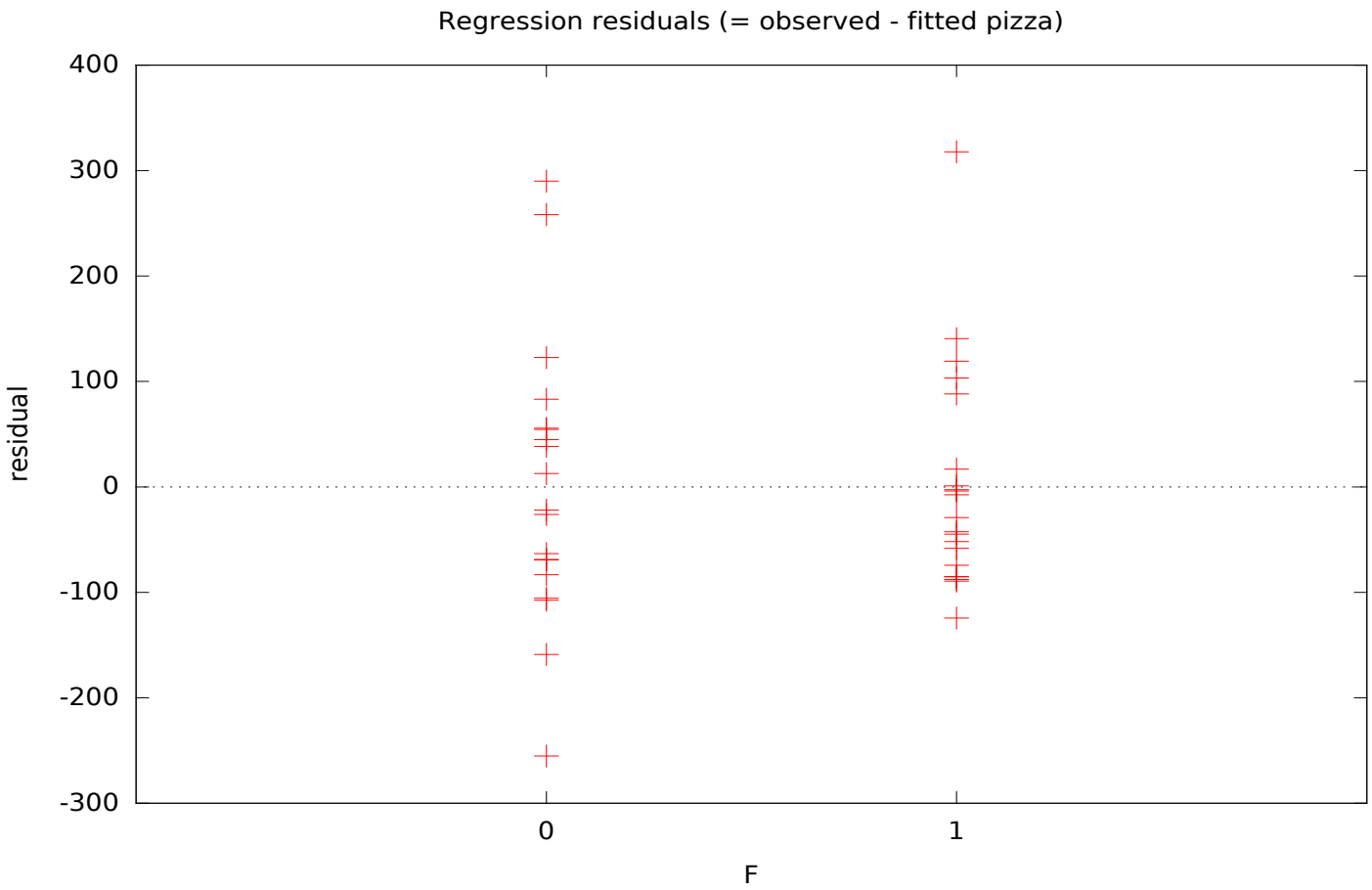
Example 5.1.3. Graphs.

Residual plot against age (RG3).



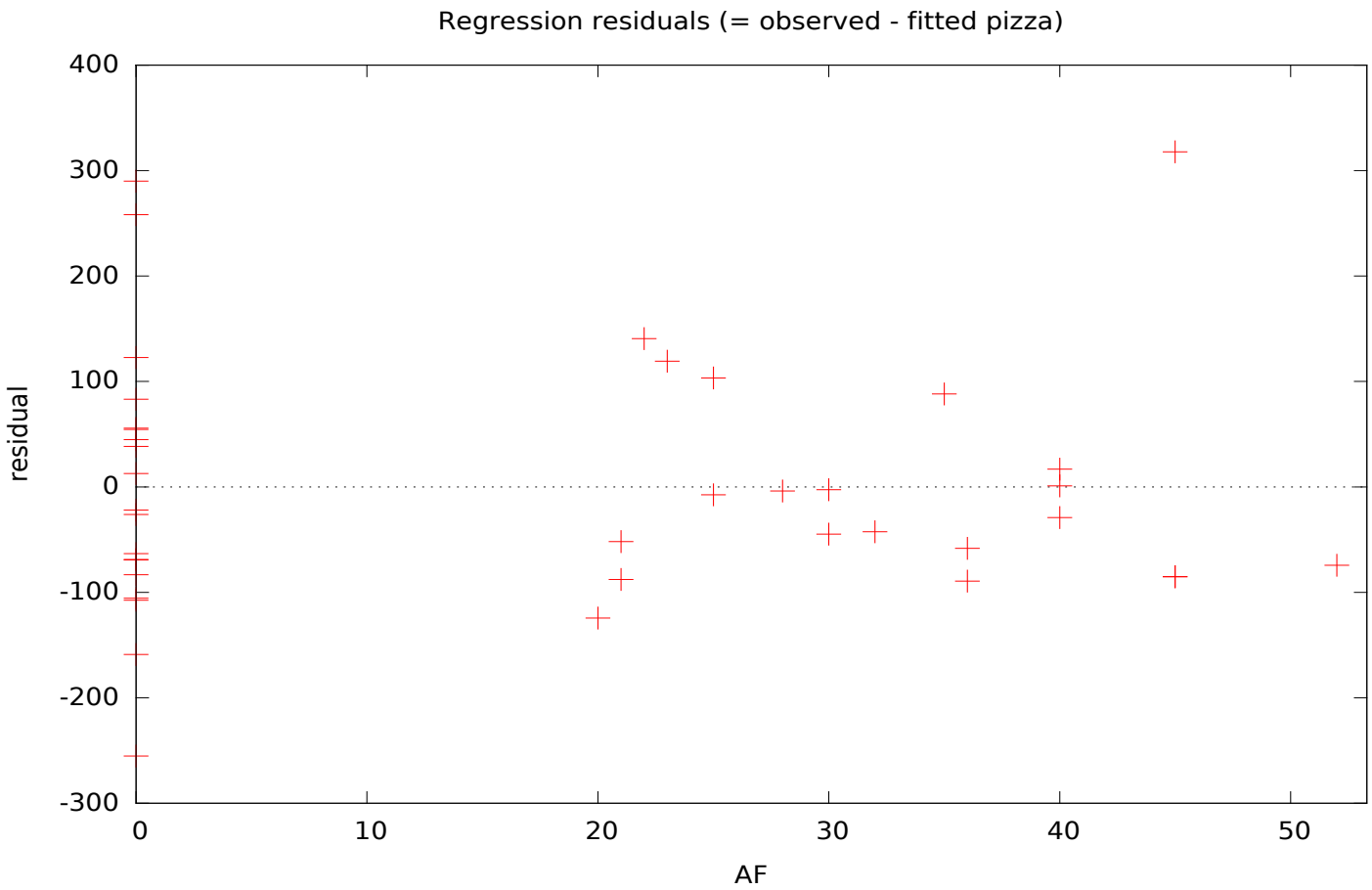
Example 5.1.3. Graphs.

Residual plot against the dummy variable F (RG4).



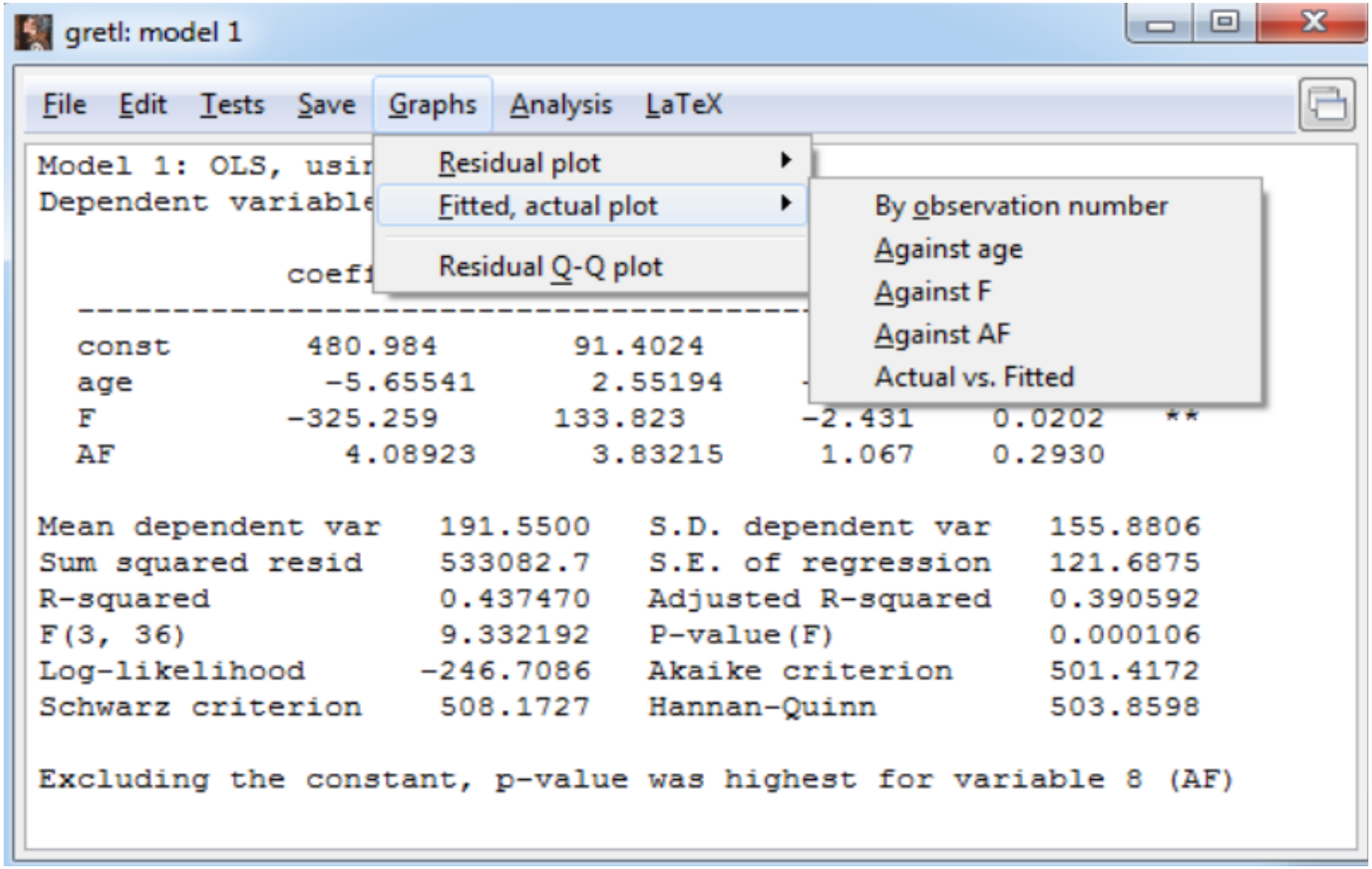
Example 5.1.3. Graphs.

Residual plot against the new variable AF (RG5).



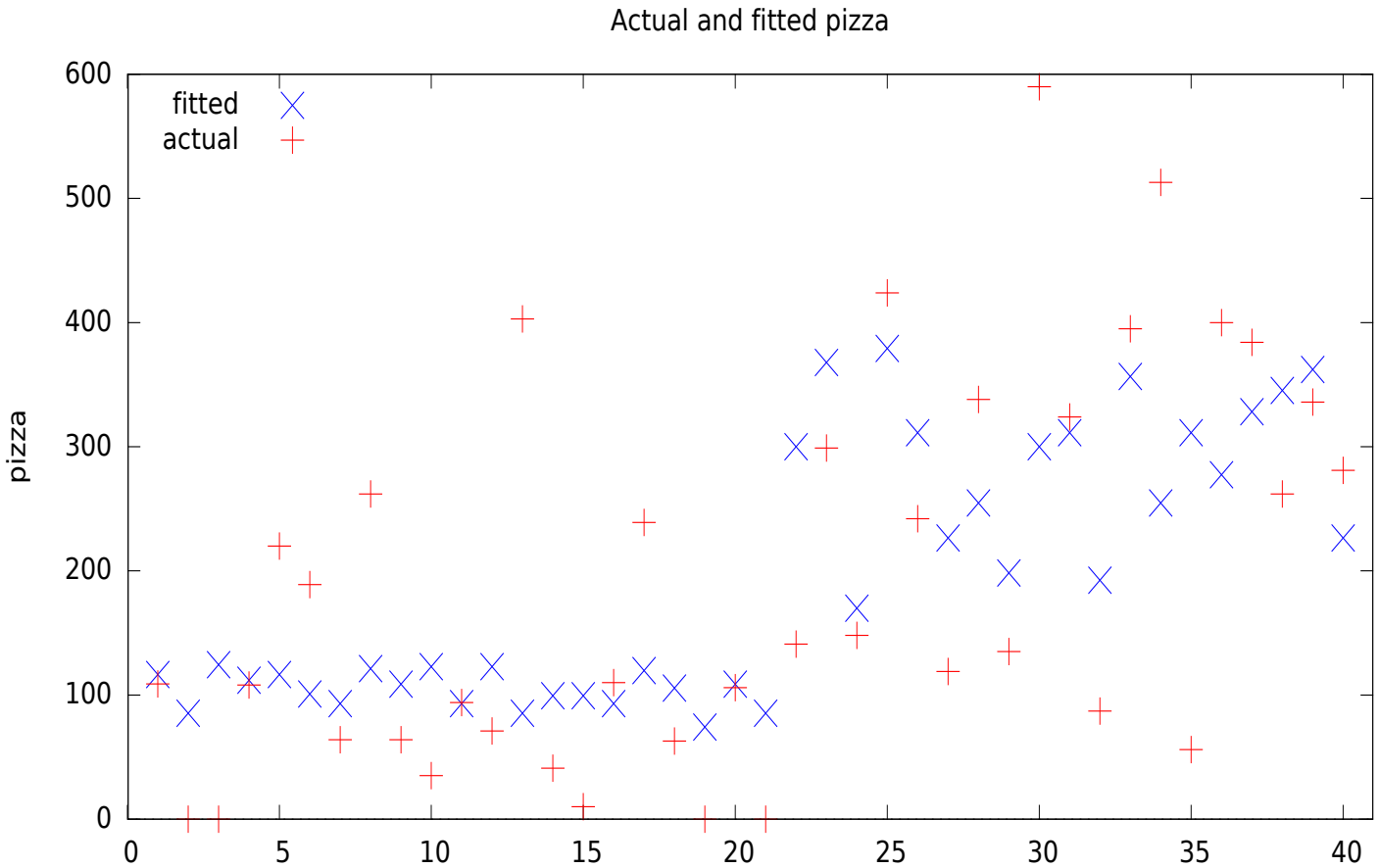
Example 5.1.3. Graphs.

With respect to the fitted and actual values plots several options are available.



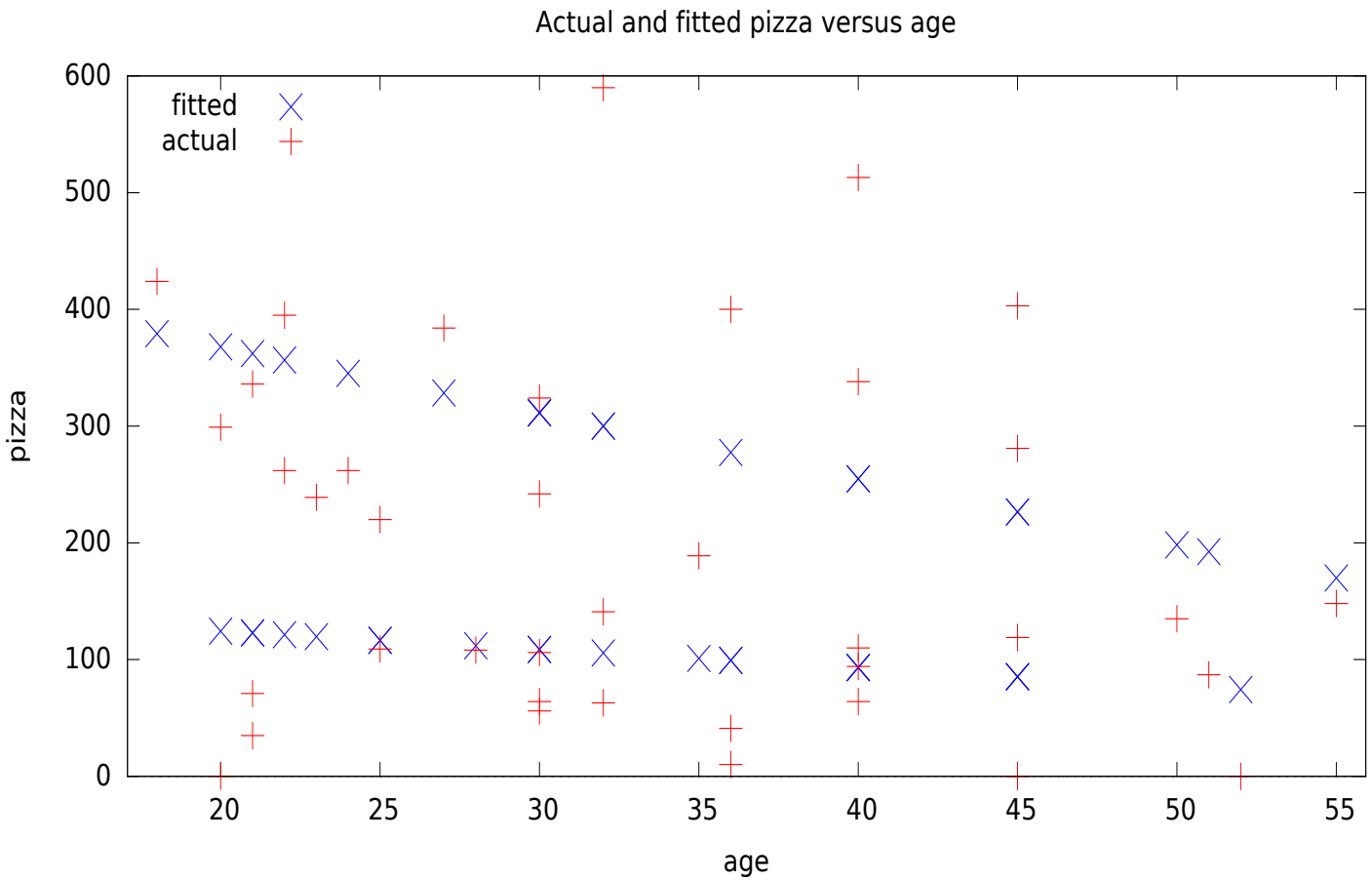
Example 5.1.3. Graphs.

Fitted and actual values plot by number of observation (FAG1).



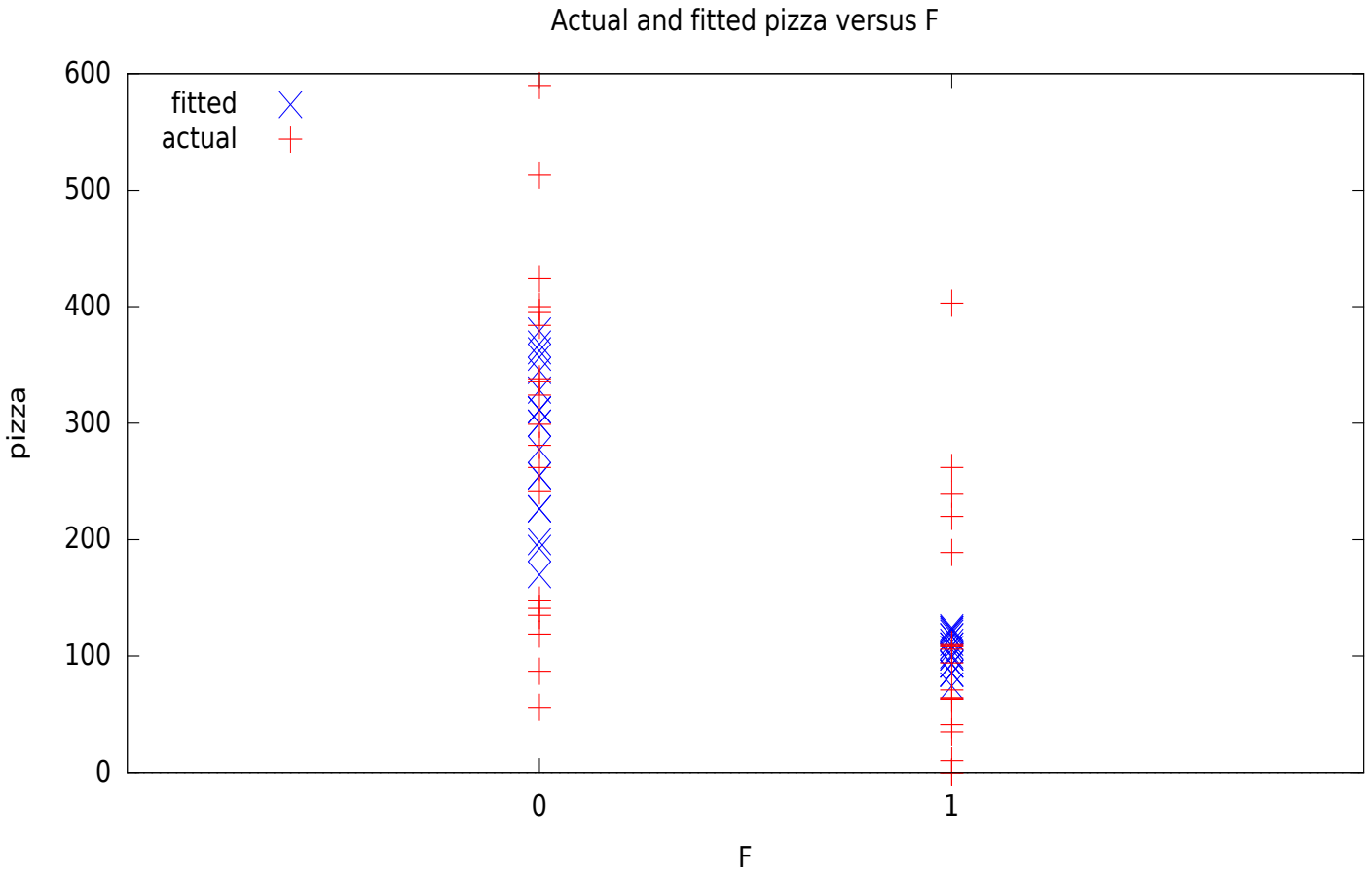
Example 5.1.3. Graphs.

Fitted and actual values against age (FAG2).



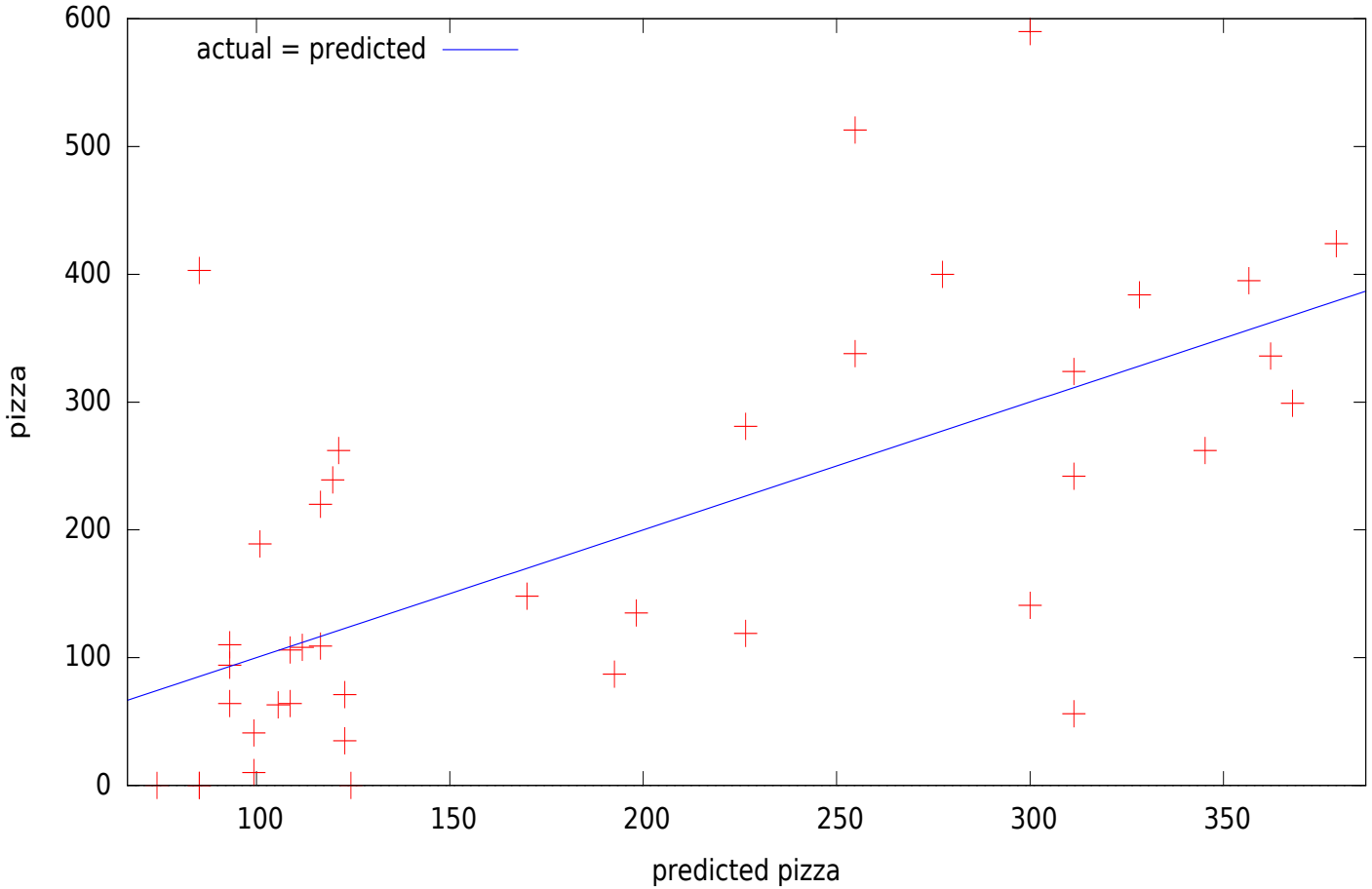
Example 5.1.3. Graphs.

Fitted and actual values against the dummy variable F (FAG3).



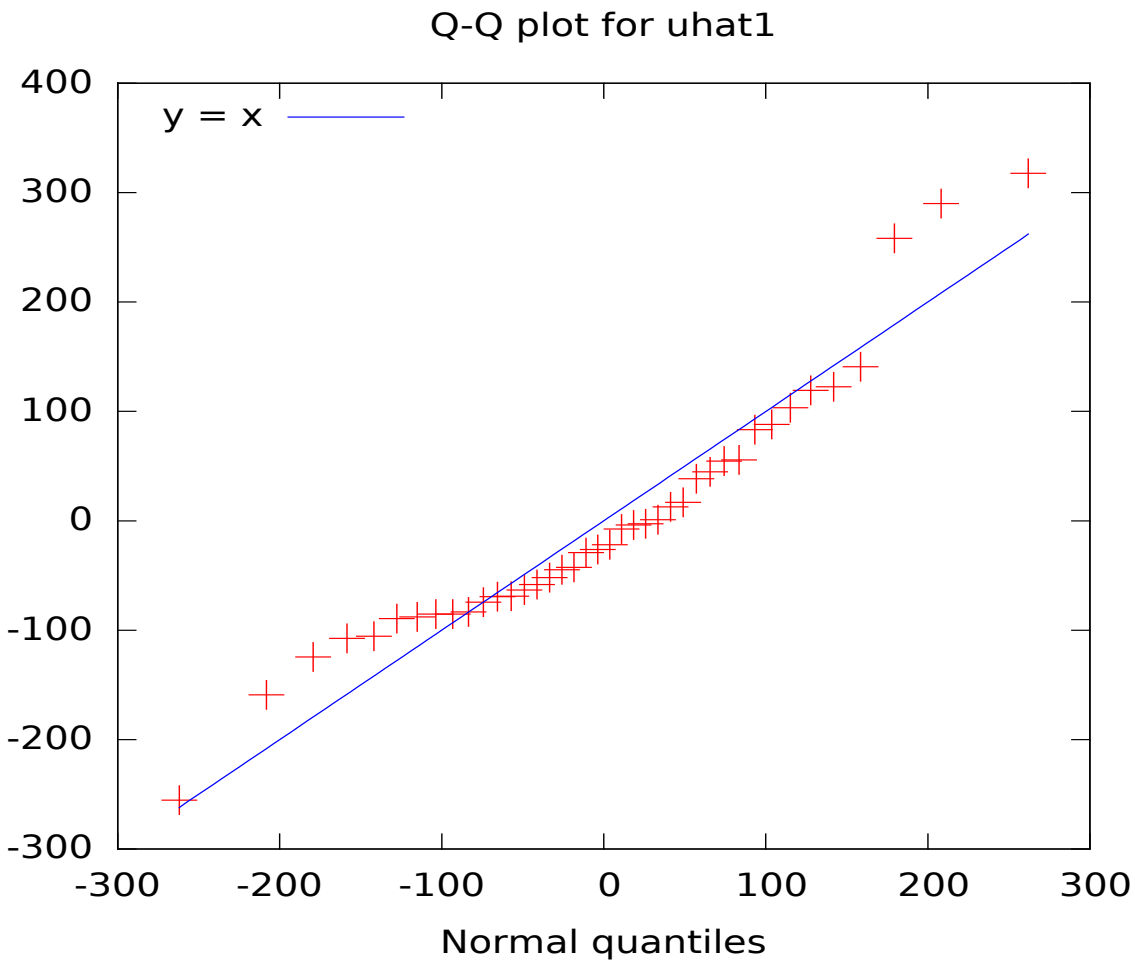
Example 5.1.3. Graphs.

Actual values versus fitted values graph (FAG4).



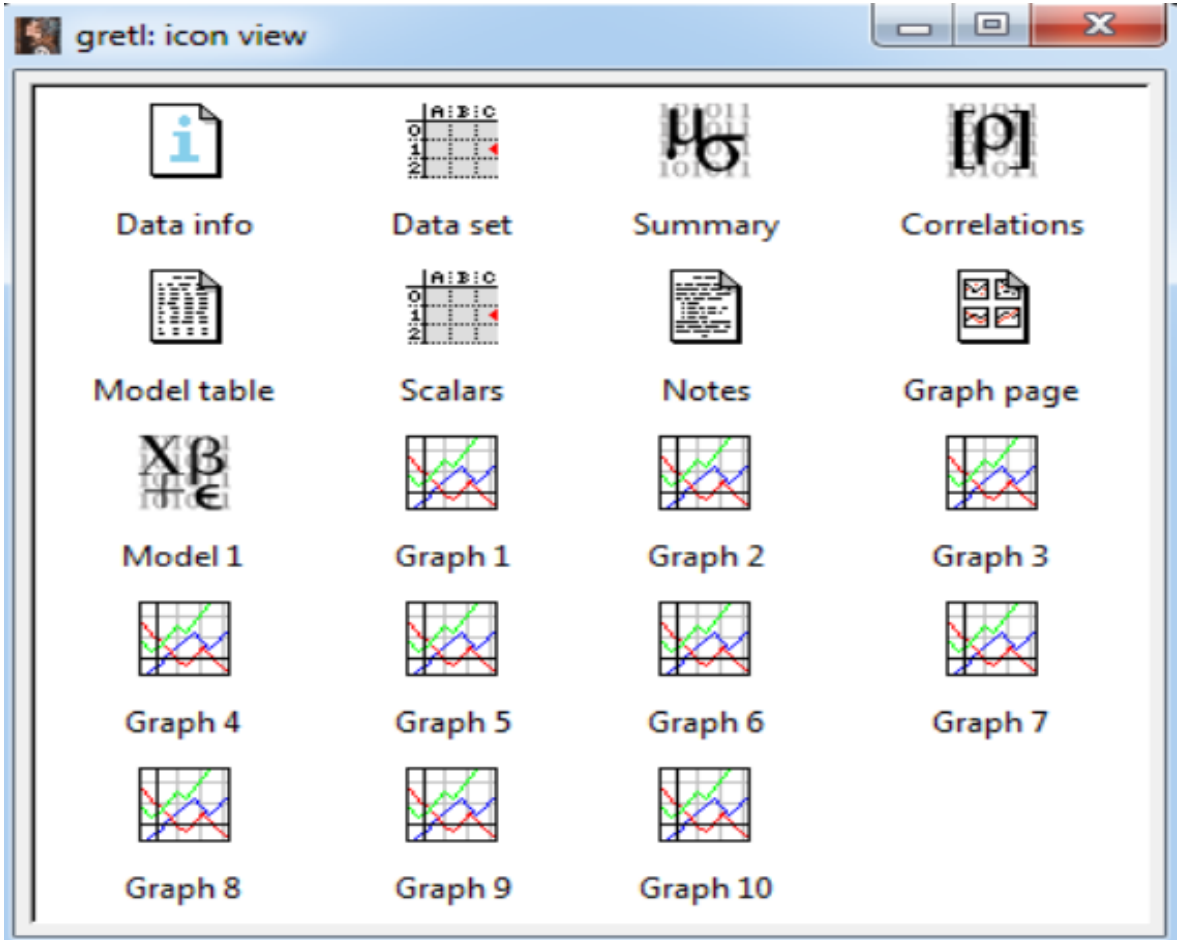
Example 5.1.3. Graphs.

Residual Q-Q plot.



Example 5.1.3. Graphs.

The graphs can be retrieved from the session.



Example 5.1.3. Graphs.

Results (I).

SRF: $\widehat{pizza}_i = 480.984 - 5.65541\, age_i - 325.259\, F_i + 4.08923\, (age_i \times F_i)$

- The estimated pizza consumption for newborn men amounts to \$480.984.
- Marginal effects:
  - Age marginal effect: The estimated change in the consumption of pizza amounts to $(-5.65541 + 4.08923 F_i)$ dollars when age increases by one year, holding gender constant. This change is not constant throughout the sample because it depends on the gender of the client. Thus, this change is estimated to be -5.65541 dollars for a man and -0.56618 dollars for a woman.
  - Gender effect: The estimated difference in the consumption of pizza between a woman and a man of the same age amounts to $(-325.259 + 4.08923\, age_i)$ dollars. This change is not constant throughout the sample because it depends on the age of the client. Thus, it is estimated that women consume less pizza than men but this difference decreases as the clients get older.
- Coefficient of determination:

  R^2: 43.7470 % of the sample variation in pizza consumption is explained by the variations in age and gender, according to this estimated model.

Example 5.1.3. Graphs.

Results (II).

- Interpretation of the residuals graphs:

  RG1: The residuals oscillate randomly around their mean (zero).

  RG2: The residuals are not randomly distributed: they are an increasing function of pizza consumption.

  RG3: The variability of the residuals seems to increase with age.

  RG4: The variability of the residuals related to the male category seems to be higher than the variability of the residuals for females.

  RG5: You can only observe the variability of the residuals against age for the female category. It does not seem to be constant. The residuals related to the male category are on the ordinate axis.

- Interpretation of the fitted and actual values graphs:

  FAG1 : The fit is very poor for the first 20 observations.

  FAG2: It seems that the effect of age is not well reflected.

  FAG3: It seems that the fit is worse for women than for men.

  FAG4: The global fit is quite poor since the observations are not close to the blue line.

- Residual Q-Q plot:

  There is no evidence that the residuals come from a normal distribution because the observations in the tails are far away from the main diagonal.