BETA.DI

USER MANUAL

BETA.DI

USER MANUAL

Kaloyan Haralampiev
Andrey Dyukmedzhiev

The Beta.Di software and the User manual are sponsored by "St. Kliment Ohridski" University of Sofia
Grant 064/2008

TABLE OF CONTENTS

OVERVIEW 5

1. HOW TO INSTALL BETA.DI 7

2. CREATION OF A DATA COLLECTION 10

3. CREATIN OF A CONTINGENCY TABLE 14
- **3.1. Creation of a contingency table from a Data Collection** 14
- **3.2. Creation of a contingency table from an external source** 17

4. THE RESULTS 20
- **4.1. Result Table** 22
- **4.2. Graphics** 25

5. GETTING HELP 28

OVERVIEW

Beta.Di is a Windows Application enabling analyzing categorical data from statistical social surveys, build on standard Multiple Documents Interface (MDI). User interface is clear and intuitive. The available actions are implemented as menu commands and the most of them are duplicated with tool buttons on a tool bar.

Beta.Di is applied for an inference of the relative frequencies in the rows of a contingency table. Inference is based on Bayesian Categorical Data Analysis in the following way:

A common situation in the social surveys is the following one – the representative sample with size n is randomly drawn from a population with size N. The joint sample distribution of the absolute frequencies over values of two categorical variables forms contingency tables:

	y_1	y_2	…	y_c	Total
x_1	f_{11}	f_{12}	…	f_{1c}	$f_{1\bullet}$
x_2	f_{21}	f_{22}	…	f_{2c}	$f_{2\bullet}$
⋮	⋮	⋮		⋮	⋮
x_r	f_{r1}	f_{r2}	…	f_{rc}	$f_{r\bullet}$
Total	$f_{\bullet 1}$	$f_{\bullet 2}$	…	$f_{\bullet c}$	n

where:

x_i are the values (categories) of the independent variable;

y_j are the values (categories) of the dependent variable;

f_{ij} is the absolute frequency in row i and column j;

$f_{i\bullet}$ is the sum of all absolute frequencies in row i;

$f_{\bullet j}$ is the sum of all absolute frequencies in column j;

r is the number of the rows;

c is the number of columns;

$$n = \sum_{i=1}^{r}\sum_{j=1}^{c} f_{ij}$$

It can be proved that the unknown relative frequency in a particular row of a contingency table have a posterior probability which is Beta distributed with:

$$\alpha = f_{ij} + 1$$

$$\beta = f_{i\bullet} - f_{ij} + c - 1$$

i.e.:

$$P(\pi_{j|i} = x \mid DI) = \frac{(f_{i\bullet} + c - 1)!}{f_{ij}!(f_{i\bullet} - f_{ij} + c - 2)!} x^{f_{ij}} (1-x)^{f_{i\bullet} - f_{ij} + c - 2}$$

where:

$\pi_{j|i} = \dfrac{\hat{f}_{ij}}{\hat{f}_{i\bullet}}$ is the unknown relative frequency in column j given the row i;

$\hat{f}_{ij}$ is the unknown absolute frequencies in row i and column j in the entire population;

$\hat{f}_{i\bullet}$ is the sum of all unknown absolute frequencies in row i in the entire population;

D are the available data consisting of sample absolute frequencies f_{ij};

I represents the prior information. (The posterior probability is obtained under assumption that there is no prior information, i.e. the prior probability is uniformly distributed.)

The Beta.Di software allows to calculate the posterior probability, to form Beta distribution and to calculate its mean and standard deviation.

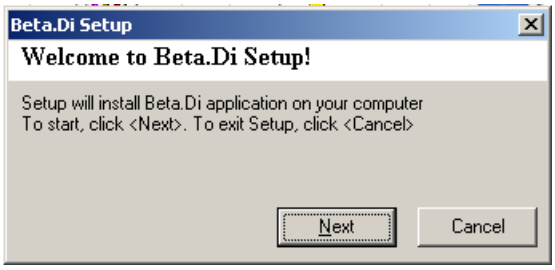
The user can:

- Create a new data collection;
- Use an existing data collection;
- Create a new contingency table;
- Use an existing contingency table;
- Create a new table with means and standard deviations of the Beta distributions;
- Create graphs of the Beta distributions.

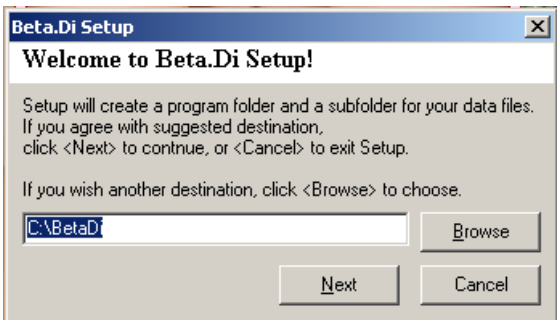
1. HOW TO INSTALL BETA.DI

Run the Beta.Di setup file:

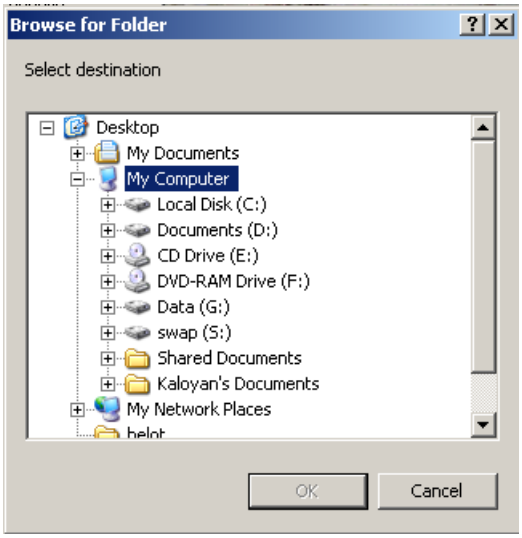
The following window appears:



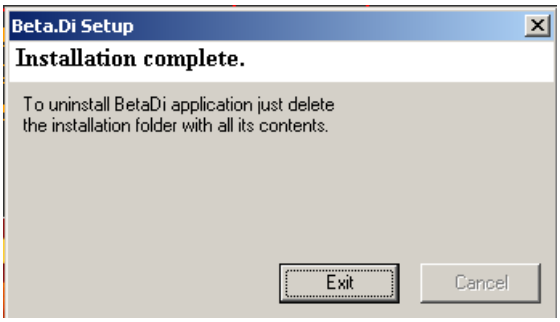
Click **Next**.
Consecutively the following two windows appear:



If the suggested folder is acceptable click **Next**.
Warning: Setup will create a new folder *Beta.Di* and a subfolder *DATA*.
Warning: The installation requires only 2 MB disc space and does not write entries in the Windows Registry.
If the suggested folder is not acceptable click **Browse**.
The following window appears:



Choose a folder. Click **OK** and **Next**.
The following window appears:



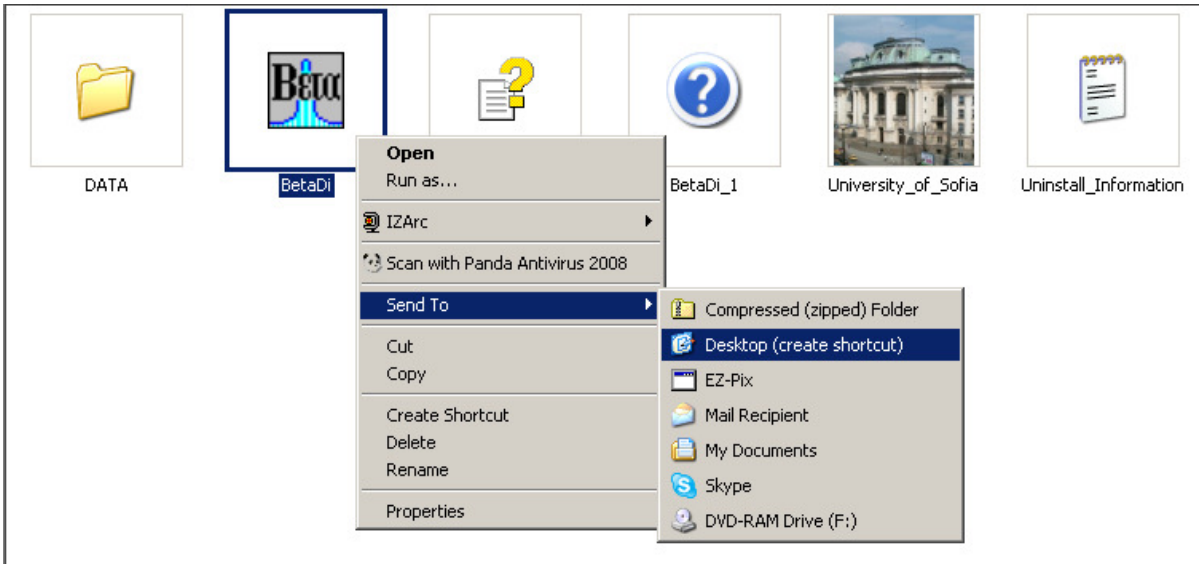
Click **Exit**.

Warning: Beta.Di Setup neither creates a Desktop icon, a Quick Lunch icon, a Start Menu icon, nor file associations between BetaDi.exe and data files it handles – users can create these manually at their own discretion.

To create Desktop icon right-click on Beta.Di icon. From the menus choose:

Send To

Desktop (create shortcut)



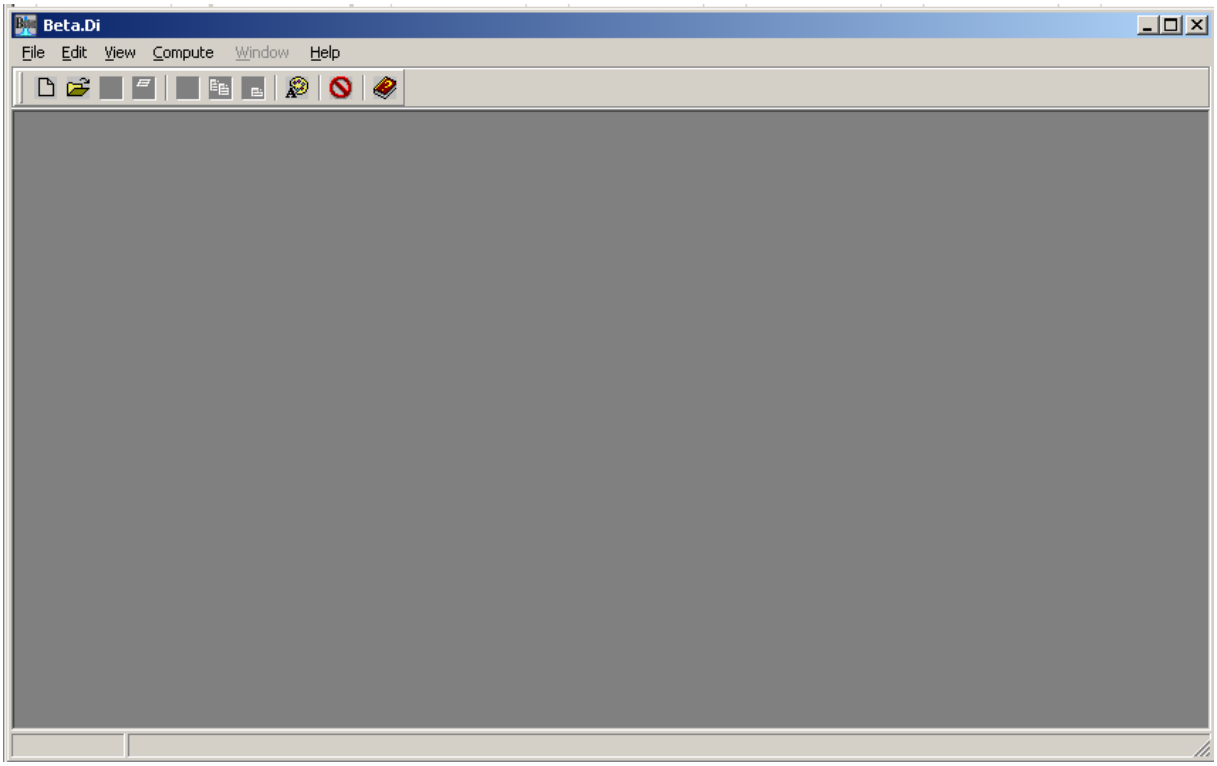
To create Quick Lunch icon first create Desktop icon and then drag it to the Quick Lunch taskbar.

To create Start Menu icon just copy Beta.Di to the folder *Document and Settings\<User>\Start Menu\Programs*.

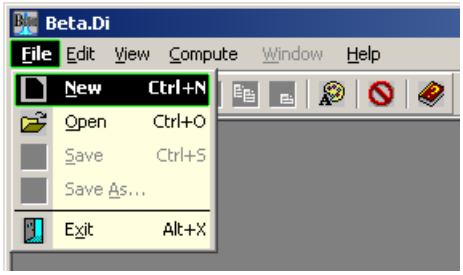
Warning: To uninstall BetaDi just delete the folder Beta.Di with all its contents.

2. CREATION OF A DATA COLLECTION

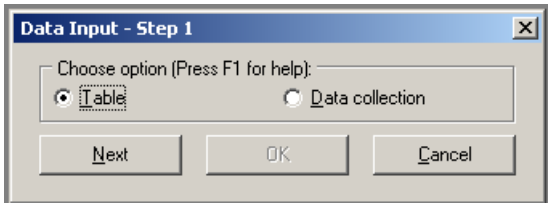
When Beta.Di is started the following window appears:



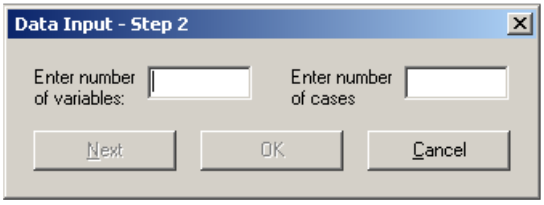
Form the menus choose:
File
New



The following window appears:



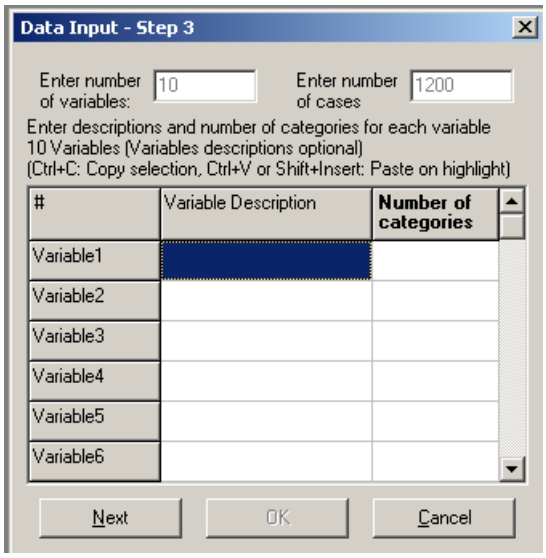
Check *Data collection* and click **Next**.
The following window appears:



Enter the number of the variables and the number of the cases. Click **Next**.

Warning: The maximum possible number of the variables is 999. The maximum possible number of the cases is 9999.

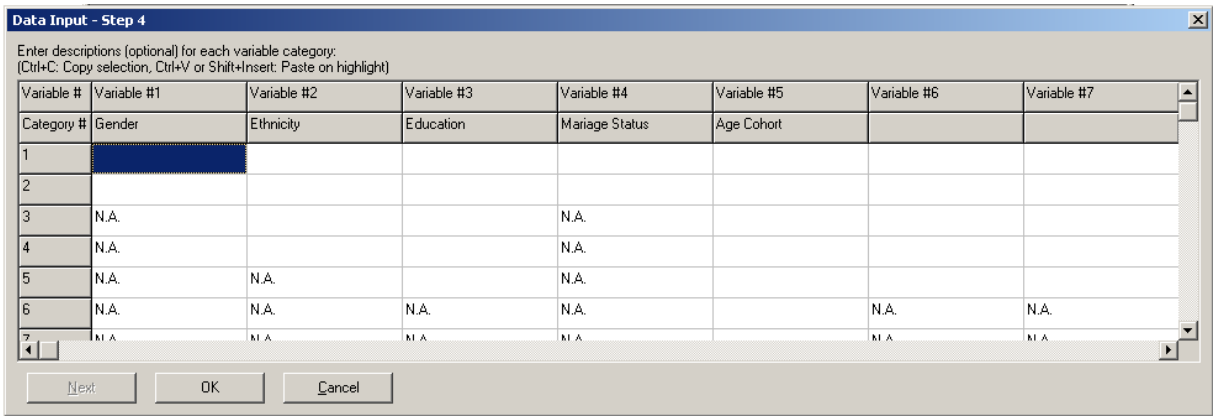
The following window appears:



Enter the description of the variables (or paste them from Clipboard). Enter the number of categories. Click **Next**.

Warning: The descriptions of the variables are optional. However, the number of categories **must** be entered.

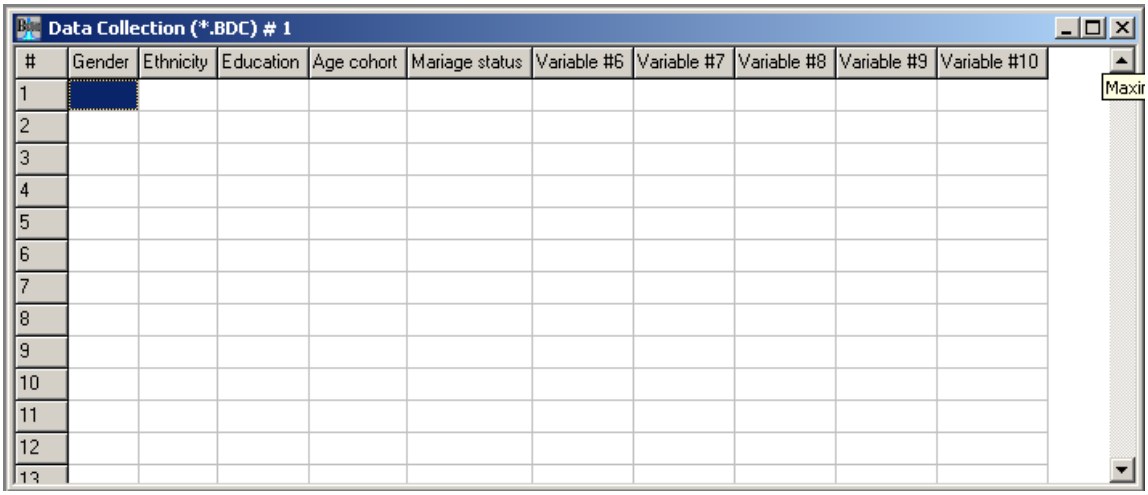
The following window appears:



Enter the descriptions of the variable categories (or paste them from Clipboard) and click **OK**.

Warning: The descriptions of the variable categories are optional.

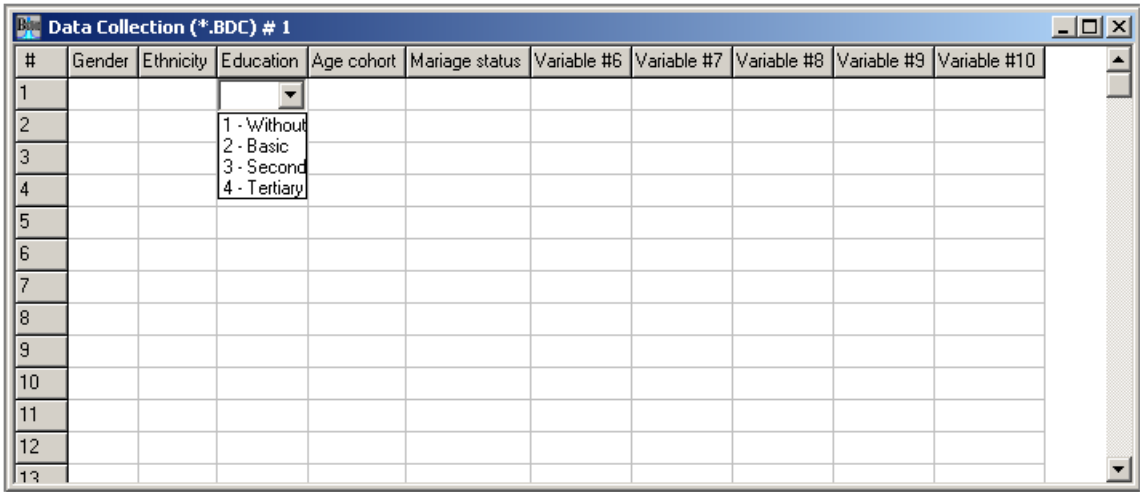
The following window appears:



The data entry could be done in one of the three following ways:

- Enter the codes of the categories (Column *Category #* in the **Data Input – Step 4** window) for each case;

- Press **F11** and select from drop-down list. Press **Enter** to confirm the choice:

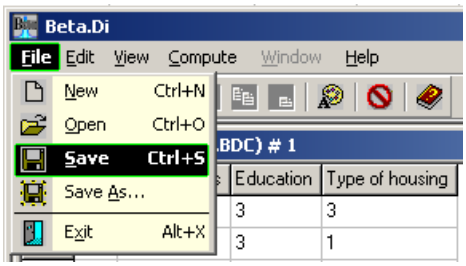


- Paste from Clipboard.

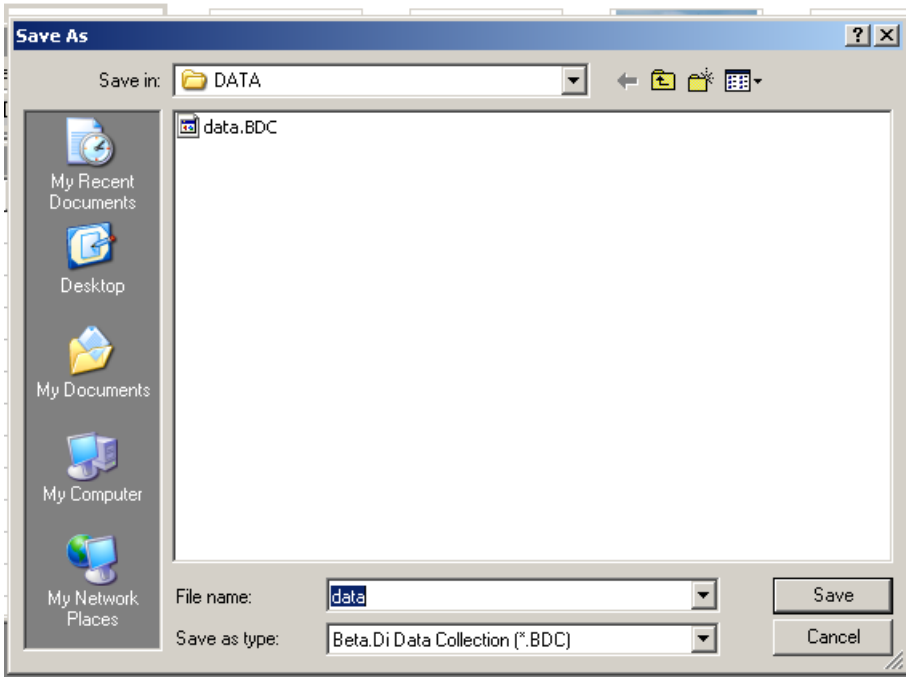
After the data are entered from the menus choose:

File

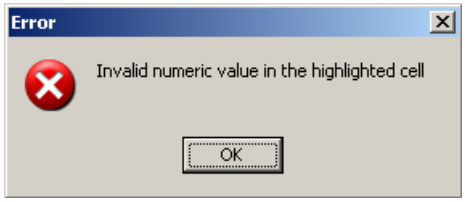
Save



The standard Windows File Save dialog box appears:



Type the file name and click **Save**.
Warning: If an inadmissible value is entered the following window appears:



Click **OK** and edit the value in the highlighted cell.

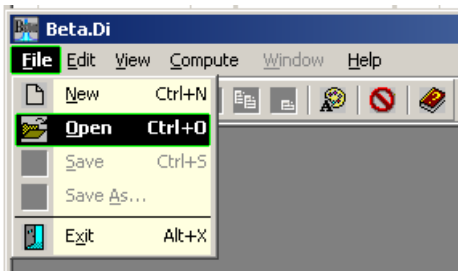
3. CREATION OF A CONTINGENCY TABLE

3.1. Creation of a contingency table from a Data Collection

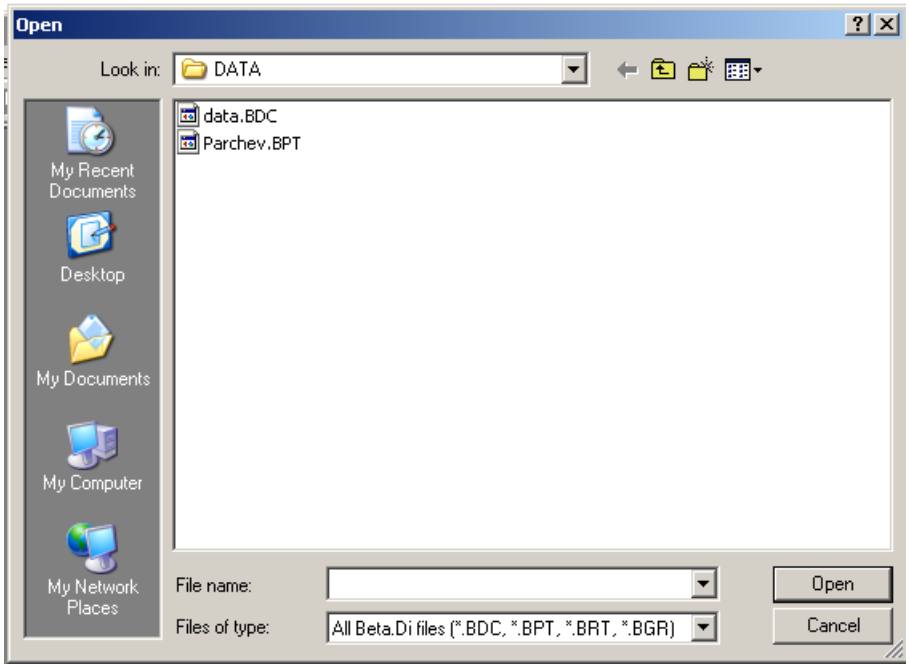
If there is no open Data Collection from the menus choose:

File

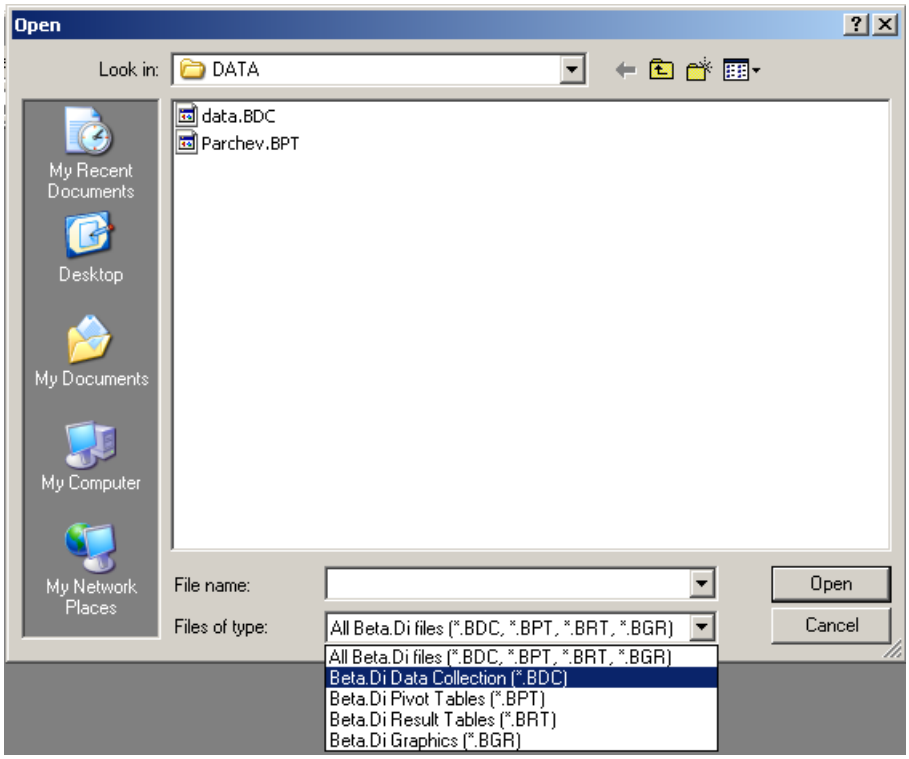
Open



The standard Windows File Open dialog box appears:



Select *Files of type*: *Beta.Di Data Collection* (**.BDC*):

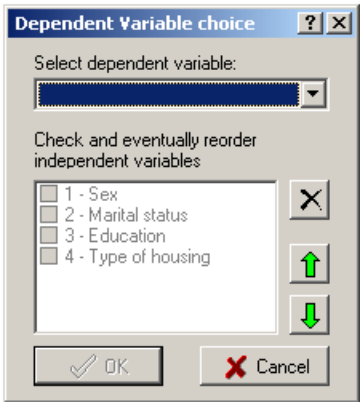


Select the name of the Data Collection file from the list and click **Open**.
From the menus choose:

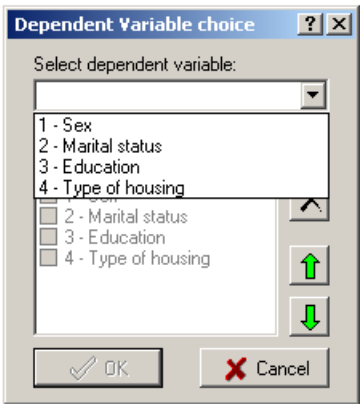
Compute

Create Pivot Table

The following window appears:



Select the dependent variable from the drop-down combo-box:



Warning: The dependent variable automatically is placed in columns.

Check the independent variables in the list box.

Reorder the independent variables in hierarchical order by the up and down arrows or dragging list items with the mouse.

Warning: The independent variables automatically are placed in rows. Besides, the second variable is nested within the first variable, the third variable is nested within the first two variables and so on.

Click **OK** and a window with contingency table appears:

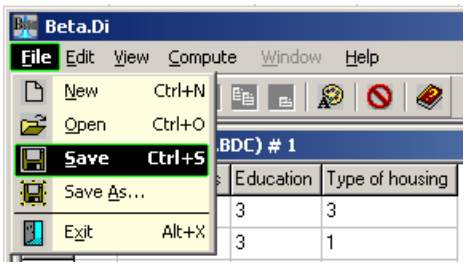
Pivot Table (*.BPT) #2

Independent Variables		Dependent Variable: Marital status				
Education	Sex	single	married	cohabitation	divorced	widowed
up to 4-th grade	male	4	4	1	1	0
	female	2	3	1	0	1
primary school (up to 7-th grade)	male	5	9	1	2	0
	female	6	15	0	2	5
secondary school up to college (2 years)	male	178	234	22	62	5
	female	102	277	19	98	34
university - all degrees	male	27	28	3	10	0
	female	19	58	2	24	3

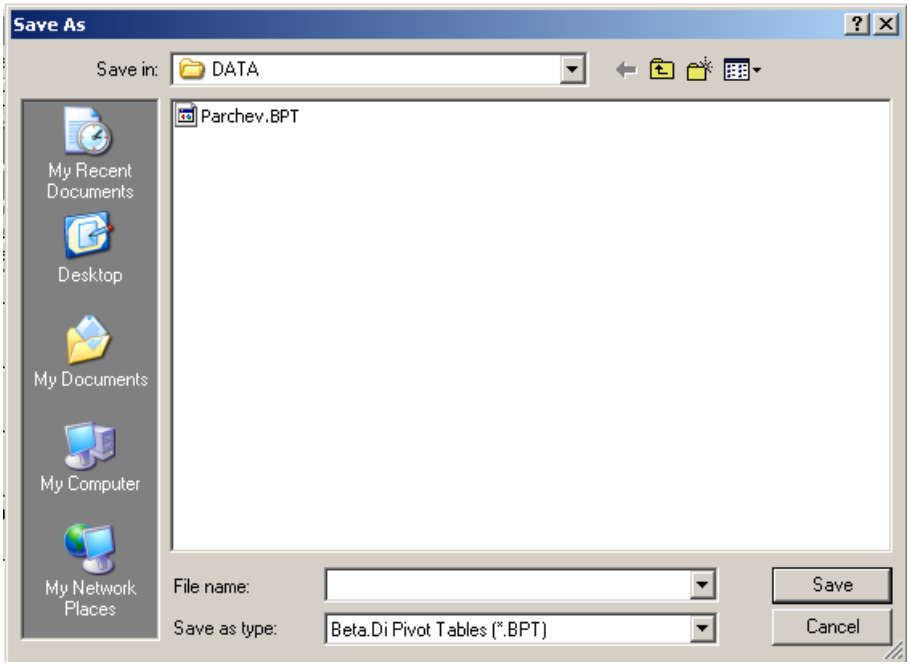
The contingency table could be saved as a separate file. From the menus choose:

File

Save



The standard Windows File Save dialog box appears:



Type the file name and click **Save**.

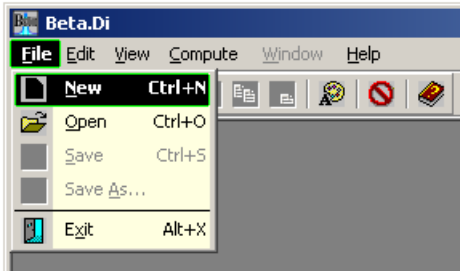
3.2. Creation of a contingency table from an external source

If there is already prepared contingency table from some external source (for instance an output of an another software or a publication) but the raw data are not accessible then the blank contingency tables could be created and it can be filled up manually.

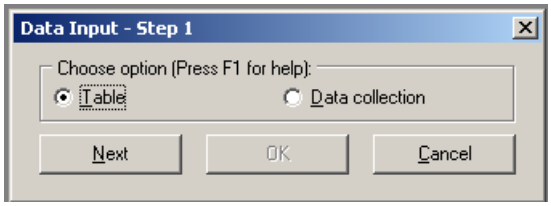
From the menus choose:

File

New

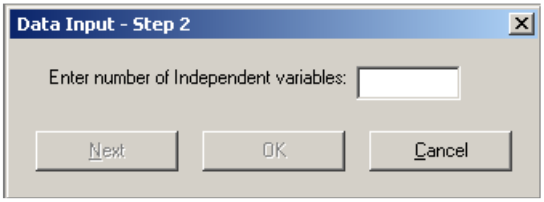


The following window appears:



Check *Table* and click **Next**.

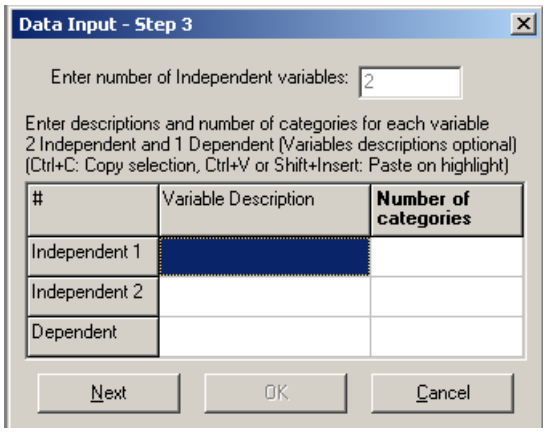
The following window appears:



Enter the number of the independent variables and click **Next**.

Warning: The maximum possible number of the independent variables is 9.

The following window appears:



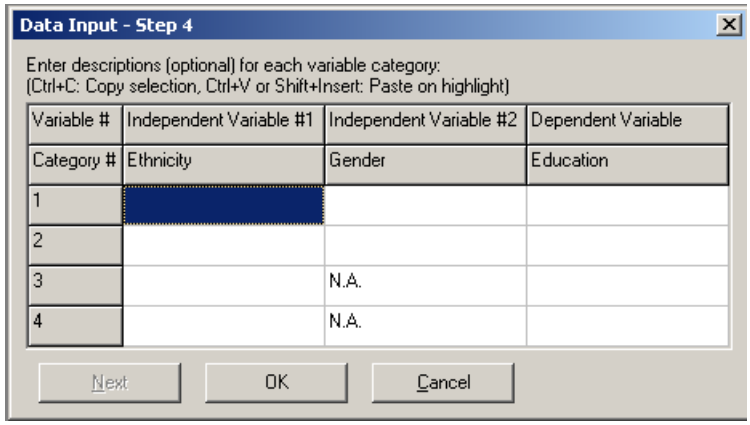
Enter the description of the variables (or paste them from Clipboard). Enter the number of categories. Click **Next**.

Warning: The dependent variable is placed automatically in the last row.

Warning: The descriptions of the variables are optional. However, the number of categories **must** be entered.

Warning: The independent variables automatically are placed in rows. Besides, the second variable is nested within the first variable, the third variable is nested within the first two variables and so on.

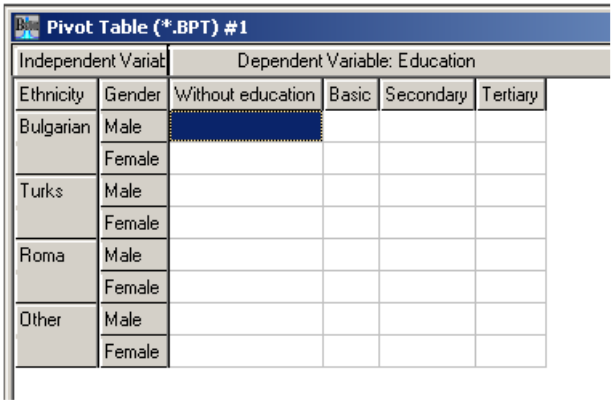
The following window appears:



Enter the descriptions of the variable categories (or paste them from Clipboard) and click **OK**.

Warning: The descriptions of the variable categories are optional.

The following window appears:

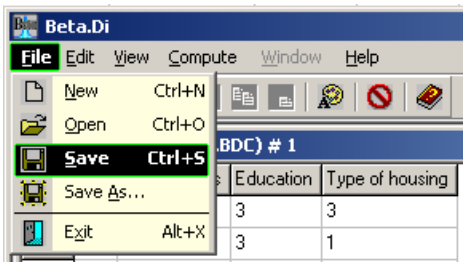


The absolute frequencies could be entered in one of the two following ways:
- Enter manually;
- Paste from Clipboard.

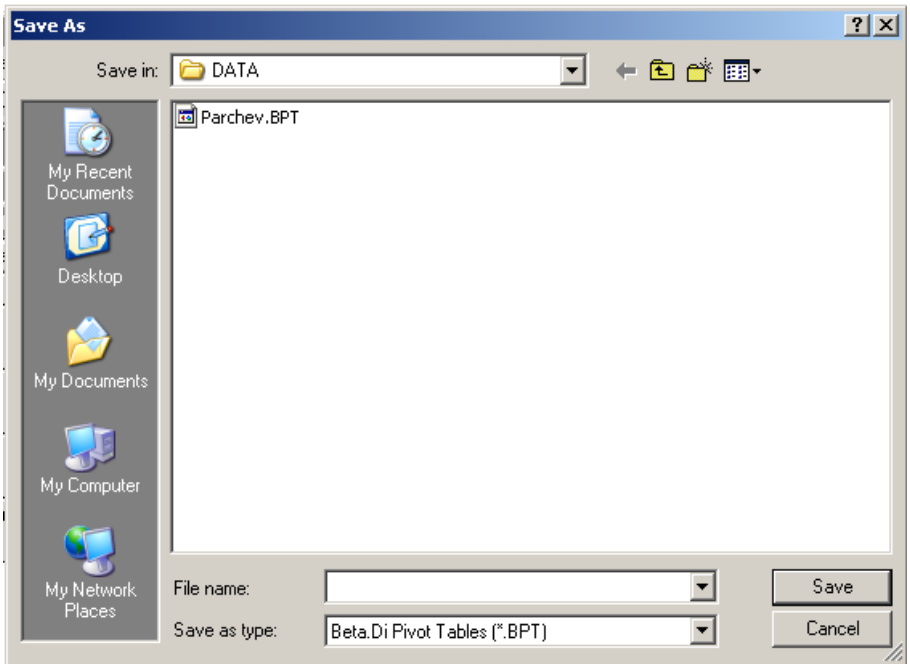
The contingency table could be saved as a separate file. From the menus choose:

File

Save



The standard Windows File Save dialog box appears:



Type the file name and click **Save**.

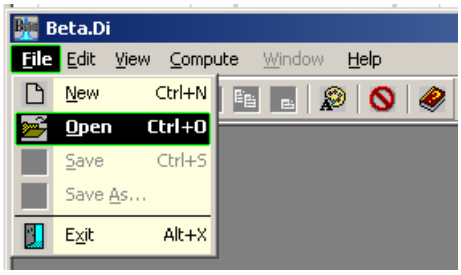
4. THE RESULTS

If there is no open contingency table the possible actions are:
- Create a contingency table from a Data Collection (Chapter 3.1.);
- Create a contingency table from an external source (Chapter 3.2.);
- Open an existing Pivot Table.

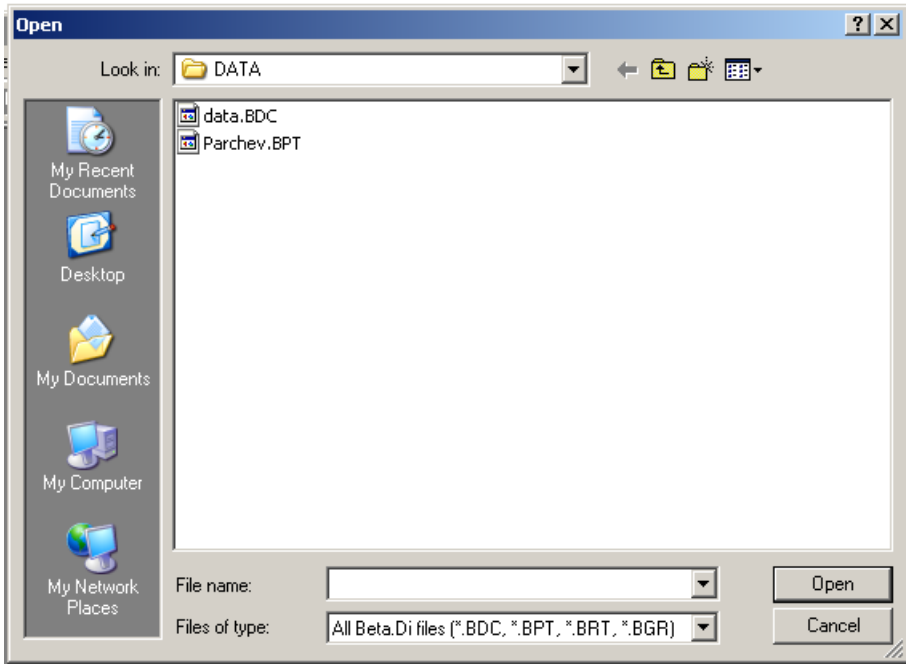
To open an existing Pivot Table from the menus choose:

File

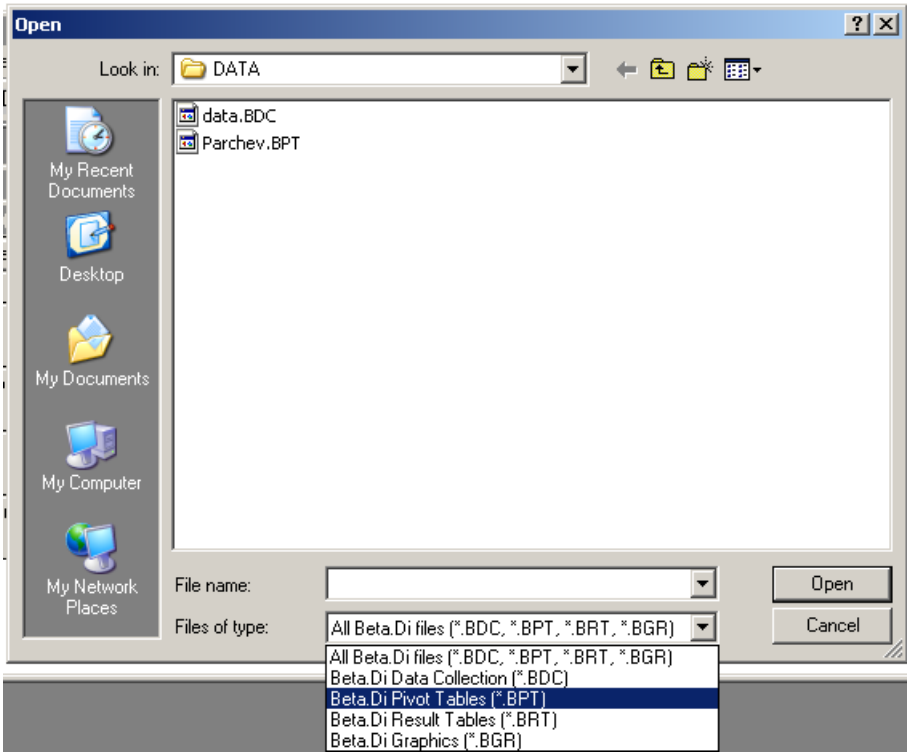
Open



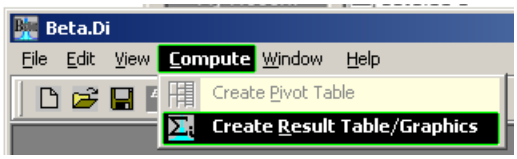
The standard Windows File Open dialog box appears:



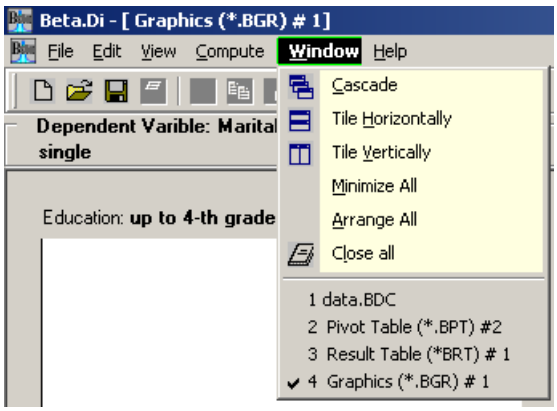
Select *Files of type*: *Beta.Di Pivot Tables* (*.*BPT*):



Select the name of the Pivot Table file from the list and click **Open**.
From the menus choose:
Compute
Create Result Table/Graphics



Two output windows appear – Result Table and Graphics.
Warning: The Graphics output window is on top. To navigate between windows from the menus choose **Window** and select file name from the list:



4.1. Result Table

The Result Table output window looks like as following:

Independent Variables		Dependent Variable: Marital status									
Education	Sex	single		married		cohabitation		divorced		widowed	
		Mean	Std. deviation	Mean	Std. deviation	Mean	Std. deviation	Mean	Std. deviation	Mean	Std. deviation
up to 4-th grade	male	0.33	0.12	0.33	0.12	0.13	0.08	0.13	0.08	0.07	0.06
	female	0.25	0.12	0.33	0.13	0.17	0.10	0.08	0.08	0.17	0.10
primary school (up to 7-th grade)	male	0.27	0.09	0.45	0.10	0.09	0.06	0.14	0.07	0.05	0.04
	female	0.21	0.07	0.48	0.09	0.03	0.03	0.09	0.05	0.18	0.07
secondary school up to college (2 years)	male	0.35	0.02	0.46	0.02	0.05	0.01	0.12	0.01	0.01	0.00
	female	0.19	0.02	0.52	0.02	0.04	0.01	0.19	0.02	0.07	0.01
university - all degrees	male	0.38	0.06	0.40	0.06	0.05	0.03	0.15	0.04	0.01	0.01
	female	0.18	0.04	0.53	0.05	0.03	0.02	0.23	0.04	0.04	0.02
Eta Squared		0.03		0.01		0.01		0.01		0.03	

The Result Table output window consists of three kinds of results – means and standard deviations of the Beta distributions and a measure of differences between row means – Eta-squared:

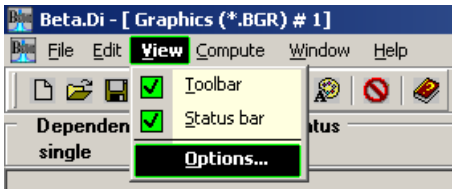
$$mean_{ij} = \frac{f_{ij} + 1}{f_{i\bullet} + c}$$

$$standard\ deviation_{ij} = \sqrt{\frac{mean_{ij}(1 - mean_{ij})}{f_{i\bullet} + c + 1}}$$

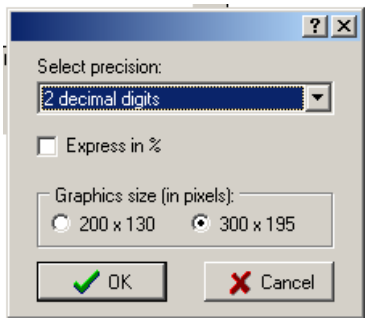
All results in the Result Table output window are presented in form of ratios rounded to the second decimal digit. To change the form and the precision from the menus choose:

View

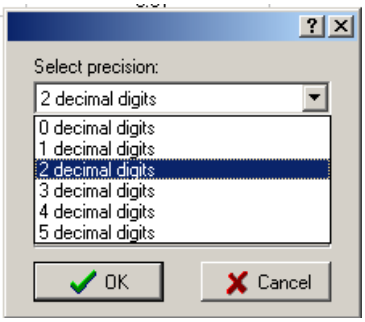
Options…



The following window appears:



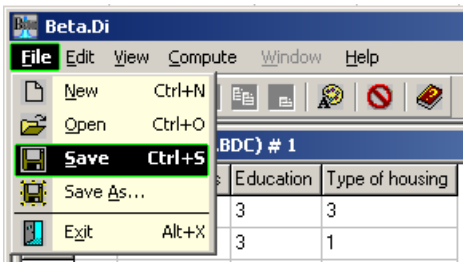
Select the precision from the drop-down combo-box:



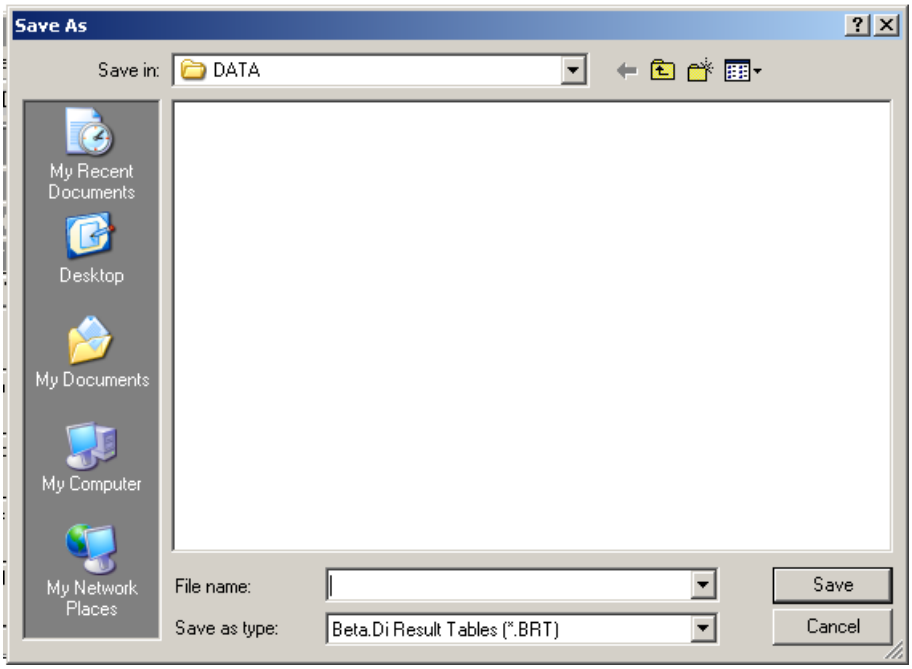
If necessary check *Express in %*. Click **OK**.
Warning: This operation affects all active Result Table output windows.

Independent Variables		Dependent Variable: Marital status									
Education	Sex	single		married		cohabitation		divorced		widowed	
		Mean	Std. deviation	Mean	Std. deviation	Mean	Std. deviation	Mean	Std. deviation	Mean	Std. deviation
up to 4-th grade	male	33.3	11.8	33.3	11.8	13.3	8.5	13.3	8.5	6.7	6.2
	female	25.0	12.0	33.3	13.1	16.7	10.3	8.3	7.7	16.7	10.3
primary school (up to 7-th grade)	male	27.3	9.3	45.5	10.4	9.1	6.0	13.6	7.2	4.5	4.3
	female	21.2	7.0	48.5	8.6	3.0	2.9	9.1	4.9	18.2	6.6
secondary school up to college (2 years)	male	35.4	2.1	46.4	2.2	4.5	0.9	12.5	1.5	1.2	0.5
	female	19.3	1.7	52.0	2.2	3.7	0.8	18.5	1.7	6.5	1.1
university - all degrees	male	38.4	5.7	39.7	5.7	5.5	2.6	15.1	4.2	1.4	1.4
	female	18.0	3.6	53.2	4.7	2.7	1.5	22.5	3.9	3.6	1.8
Eta Squared		3.4		0.7		0.8		1.0		3.1	

The Result Table could be saved as a separate file. From the menus choose:
File
Save



The standard Windows File Save dialog box appears:

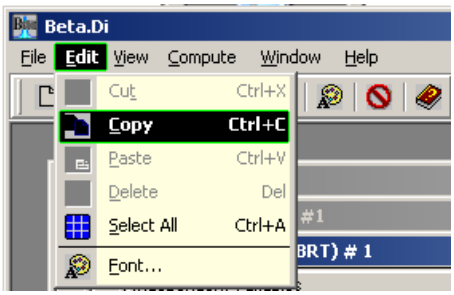


Type the file name and click **Save**.

Warning: The Result Tables could be copied to Clipboard as a whole table. From the menus choose:

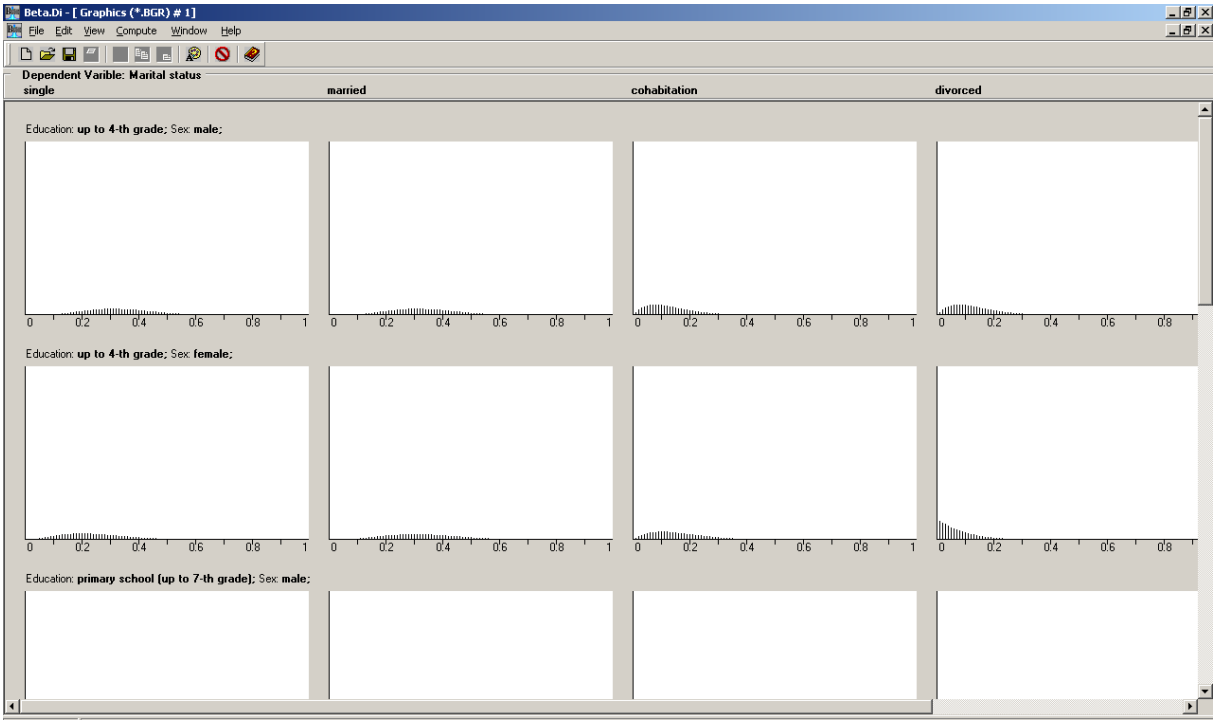
Edit

Copy



4.2. Graphics

The Graphics output window looks like as following:

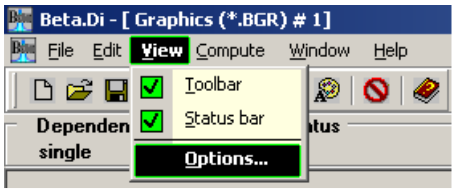


The Graphics output window consists of bar charts of the probability density function of the all Beta distributions.

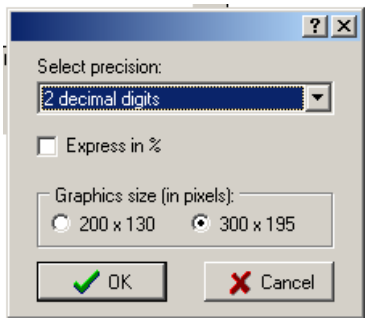
All charts in the Graphics output window are presented in size 300 x 195 pixels. To change the size from the main menus choose:

View

Options…

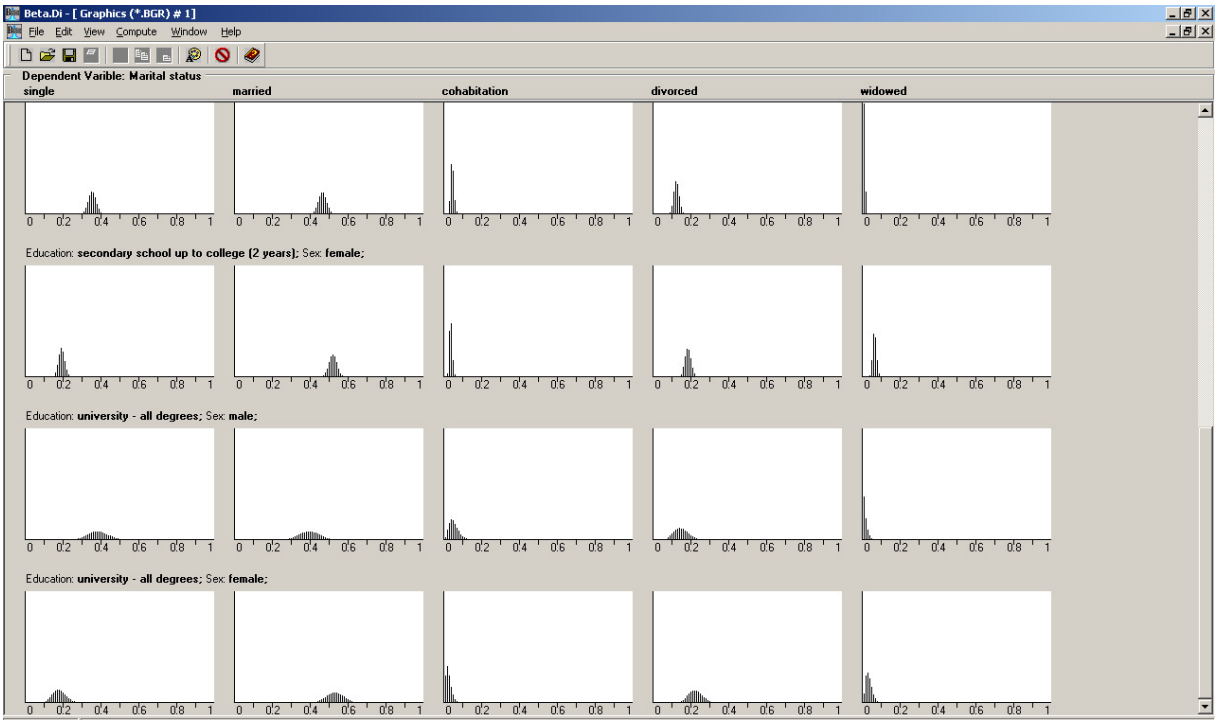


The following window appears:



Check *200* x *130*. Click **OK**.

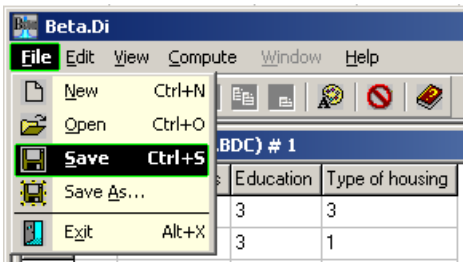
Warning: This operation affects all active Graphics output windows.



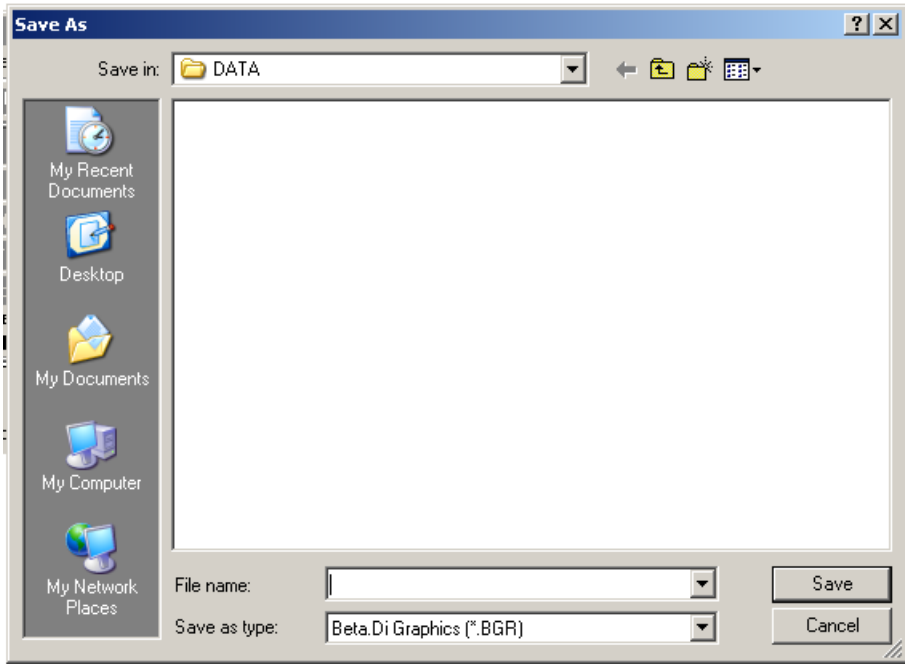
The Graphics could be saved as a separate file. From the menus choose:

File

Save



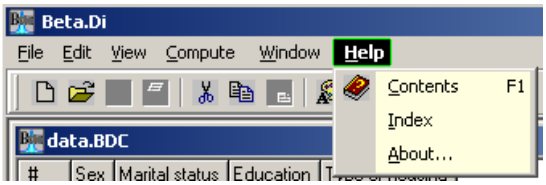
The standard Windows File Save dialog box appears:



Type the file name and click **Save**.

5. GETTING HELP

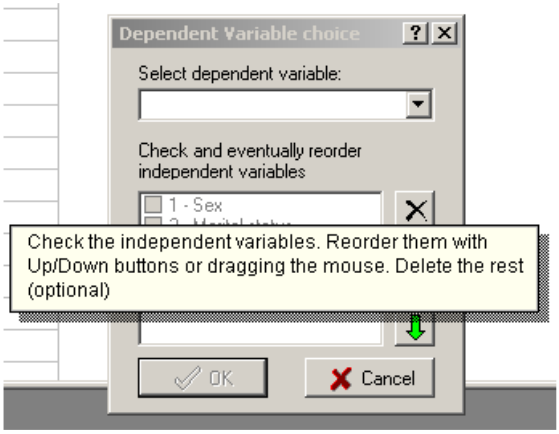
Beta.Di provides all standard Windows help topics – **Help Contents**, **Help Index** and **Help About**:



Besides Beta.Di provides context-sensitive help in one of the two following ways:
- pop-up windows;
- direct link to the Help window.

The pop-up help window is available if there is a question button in the upper-right corner of the dialog window title bar:

Click the question-mark, move the mouse on the desired topic and click again.
The pop-up help window appears:



If there is no question button in the upper-right corner of the dialog window title bar then pressing **F1** will open separate Help window.