



[HTTP://WWW.ICSSRDATA SERVICE.IN/](http://www.icssrdataservice.in/)

# ICSSR Data Service

Indian Social Science Data Repository

## SPSS: User Guide



Indian Council of Social Science Research

**SPSS 22: User Guide**

**Contents**

<b>1.</b>	<b>Introduction</b>	<b>1</b>
<b>2.</b>	<b>Opening Data File</b>	<b>1</b>
<b>3.</b>	<b>Importing Text Data</b>	<b>4</b>
<b>4.</b>	<b>Value Labels</b>	<b>7</b>
<b>5.</b>	<b>Reports</b>	<b>9</b>
<b>6.</b>	<b>Descriptive Statistics</b>	<b>11</b>
	<b>6.1 Frequency Calculation</b>	
	<b>6.2 Comparing Two Variables</b>	
<b>7.</b>	<b>Sort Cases</b>	<b>17</b>
<b>8.</b>	<b>Select Cases</b>	<b>18</b>
<b>9.</b>	<b>Split Data</b>	<b>19</b>
<b>10.</b>	<b>Custom Tables</b>	<b>20</b>
	<b>10.1 Two Way Table</b>	
	<b>10.2 Three Way Table</b>	
<b>11.</b>	<b>Compute Variable</b>	<b>23</b>
<b>12.</b>	<b>Recoding of Data</b>	<b>25</b>
<b>13.</b>	<b>Rank Cases</b>	<b>27</b>
<b>14.</b>	<b>Weight Cases</b>	<b>29</b>
<b>15.</b>	<b>Aggregate Data</b>	<b>29</b>
<b>16.</b>	<b>Merging Data</b>	<b>31</b>
	<b>16.1 Adding New Variables</b>	
	<b>16.1.1 Adding Variables when Numbers of Cases in Two Files are Uneven</b>	
	<b>16.1.1.1 Restructuring Data</b>	
	<b>16.2 Adding New Cases</b>	
<b>17.</b>	<b>Output Files</b>	<b>37</b>
	<b>17.1 Run Script</b>	
<b>18.</b>	<b>Syntax</b>	<b>40</b>
<b>19.</b>	<b>Graphs</b>	<b>41</b>

**1. Introduction**

SPSS (Statistical Package for Social Science) is a comprehensive package for data analysis developed by IBM Corporation. The software facilitates its users to collect, handle, clean and present the data” in the forms of tables, graphs, descriptive statistics, reports, trend lines, etc. SPSS can read almost any type of data e.g. numeric, alphanumeric, binary, date, currency and so on. The data in SPSS can be viewed and edited from dropdown menus and dialogue boxes.

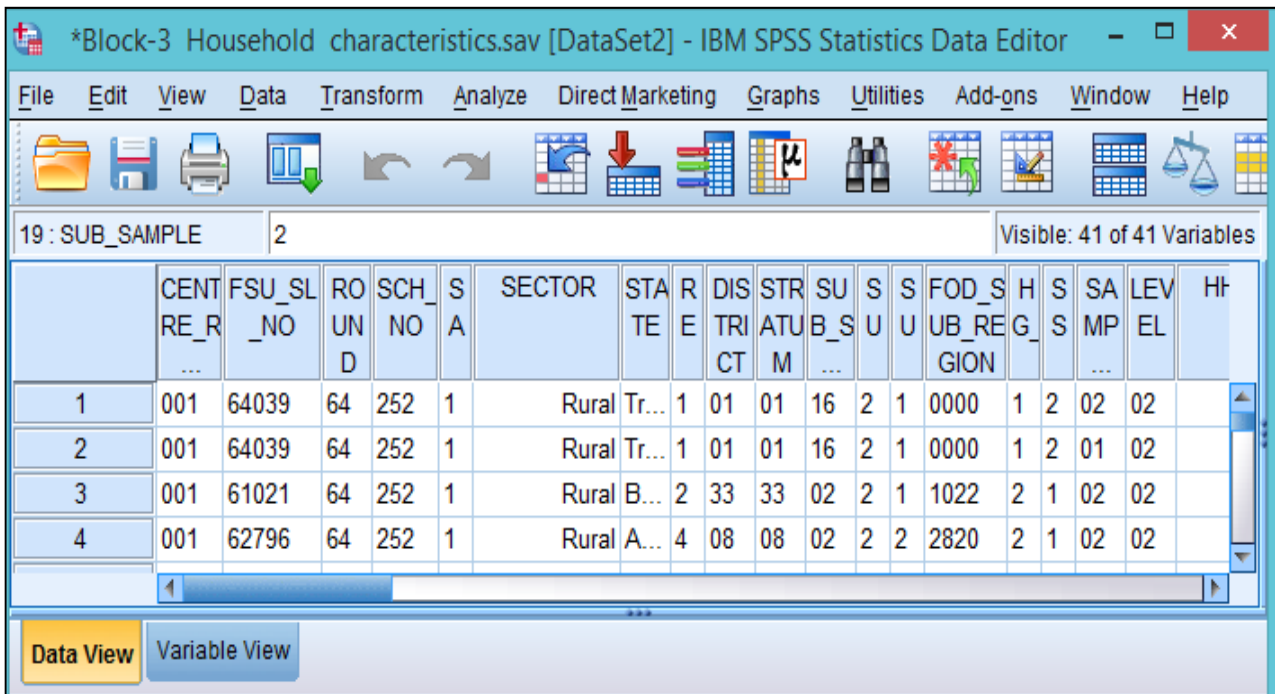
In this manual, the example of dataset for NSSO round “Schedule 25.2: Participation and Expenditure in Education, 64th Round” is used. This survey was conducted during the period of July 2007 to June 2008.

**2. Opening a Data File in SPSS**

To open a dataset or file, select the following functions from the dropdown menus:

**File >Open >Data**

Mainly, two types of “views” are available in SPSS to view a data file, i.e. "Data View" and "Variable View".



**Fig. 1: Data View**

Fig. 1 is the screenshot of "Data View" in SPSS, where rows represent the numbers of households, and columns represent the variables for which information has been collected from the households.

The "Variable View" is shown in Fig. 2, where each row represents the number of variable and each column represent the nature of variable.

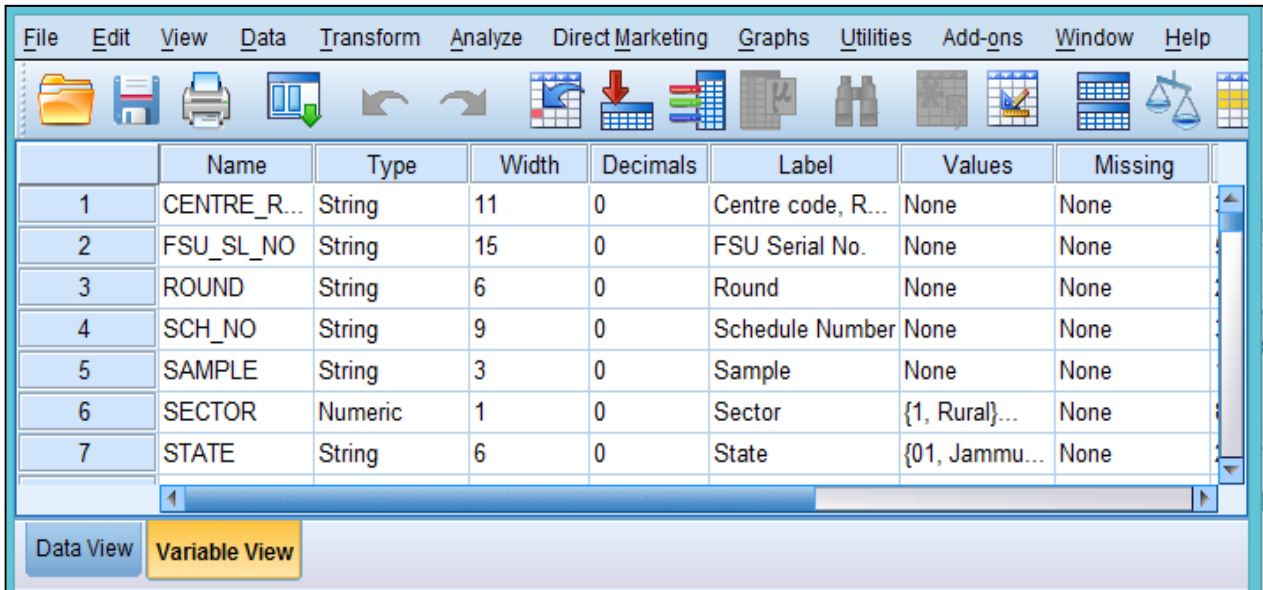


Fig. 2: Variable View

In Fig. 2, the first column provides "name" of variable which is generally written in short form using underscores. Second column provides "type" of variable, i.e. whether it is numeric, alphanumeric, string, date, etc.

To check, click at "String" option. On clicking at "String", Option Box shown in Fig. 3 will appear. Select the appropriate type in the available option box and click on "OK".

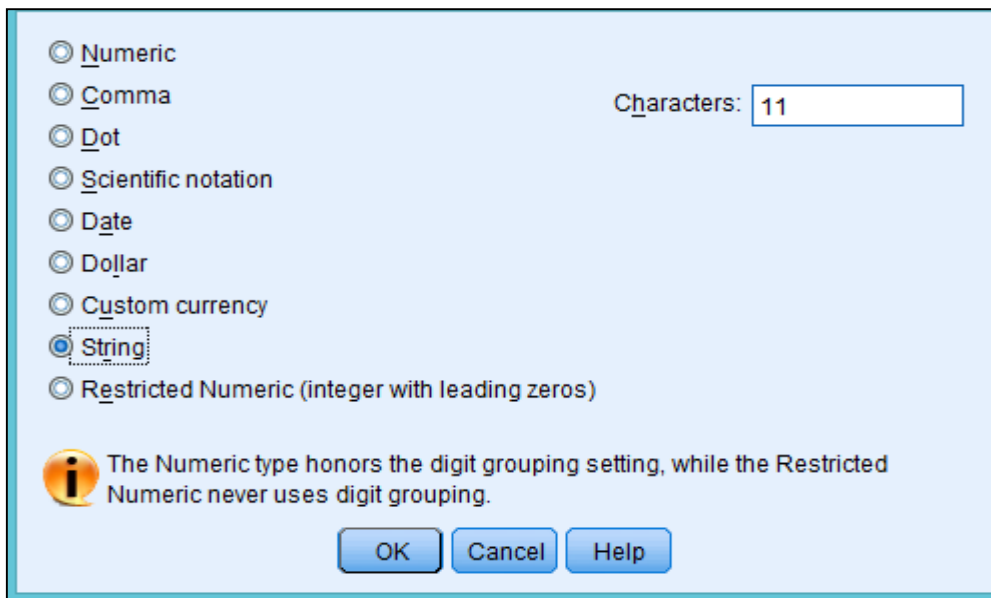
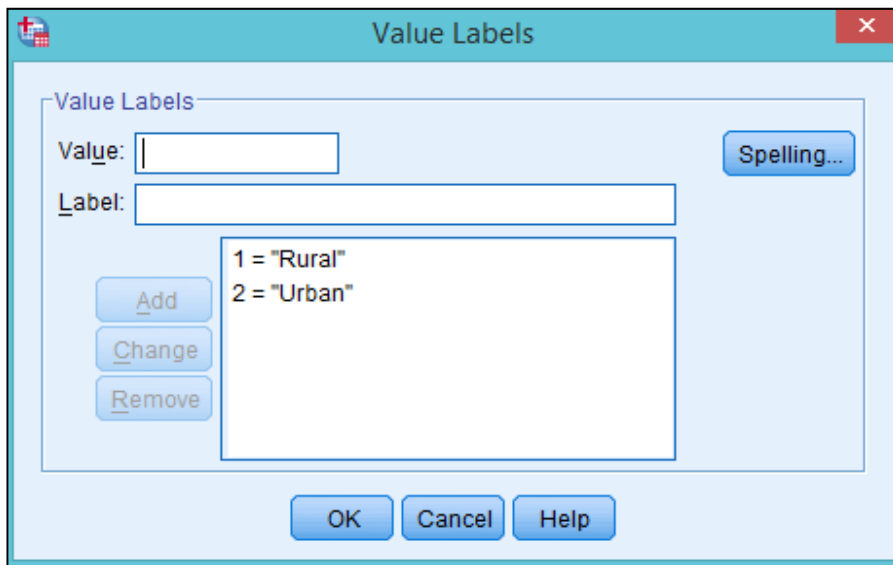


Fig. 3: Option Box to Define Type of Variable

The third column provides expected length of a given variable; whereas the fourth column provides number of decimal places for a given variable. The fifth column provides "Label", where users can provide full name of the variable whose short form is given in Column one.

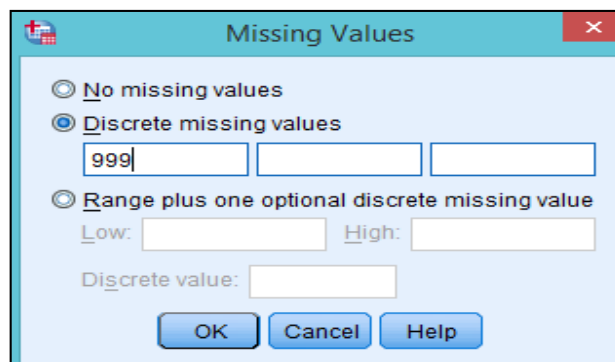
The sixth column in Fig. 2, provides "Values" assigned to a variable, e.g. whether the information provided in the dataset is for rural and urban population, here 1 denotes rural and 2 denotes urban population.

On clicking at the "Values" button, the Option Box that appears is shown in Fig. 4, type "1" in the "Value" box and name of variable in the "Label" box; Click at "OK".



**Fig. 4: Option Box for Value Labels**

Seventh column in Fig. 2 provides information on "Missing" values in the dataset. In case, there are missing values in the dataset, one can assign values to them which will appear in the dataset. For assigning value to missing values, click at "missing", a new Option Box will appear as shown in Fig. 5. Select "Discrete missing values" from the Option Box and write the value you want to assign, click at "OK". For example, 999 is assigned for the missing values here.



**Fig. 5: Assigning Value to Missing Values**

### 3. Importing Text Data

Generally, most of the social sciences data are in text format providing information on each household, individual and firm, etc. In this section, the functions to import text data in SPSS file is described.

To import text data, select: **File > Read Text Data** from the dropdown menus in SPSS. The first (of six step) “Text Import Wizard” shown in Fig. 6.1 will appear to facilitate import of text data.

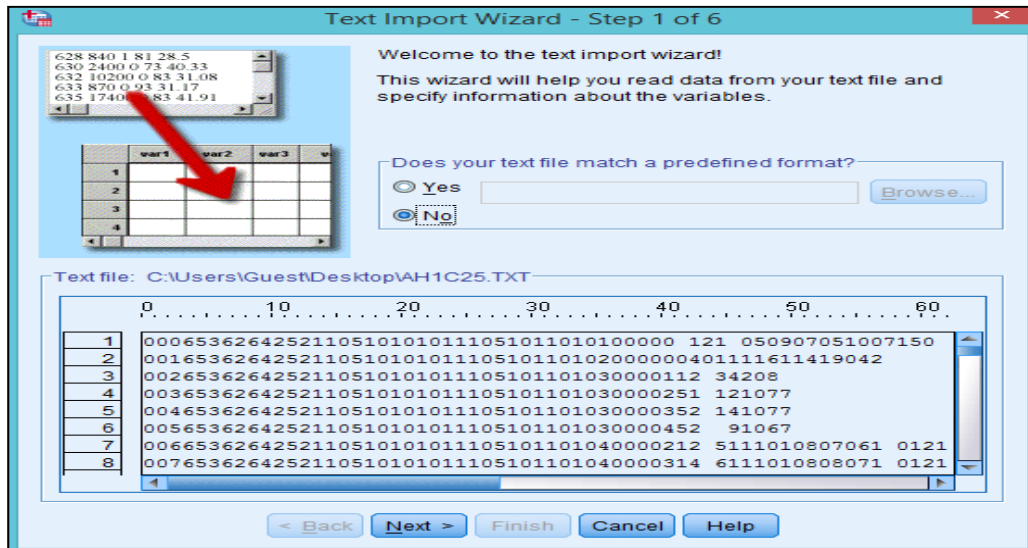


Fig. 6.1: Text Import Wizard (Step-1)

This wizard seeks information on pre-defined format of text file. In this case, since the file is not pre-defined, leave it unfilled and click at “Next” button. The second Text Import Wizard will appear as shown in Fig. 6.2.

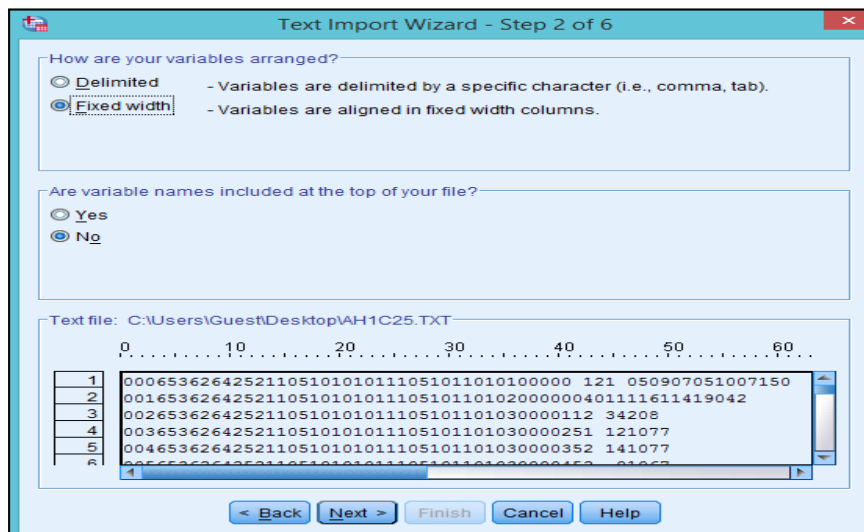


Fig. 6.2: Text Import Wizard (Step 2)

Select appropriate format from which variables are be imported viz. **delimited and fixed** width. In most of files, variables are arranged in fixed width, and in this case also, we have the fixed width data.

So, here  **Fixed width** - Variables are aligned in fixed width columns. option is selected. Click at the appropriate option. Choose "Next" button to proceed as shown in Fig. 6.2. The third Text Import Wizard will appear as displayed in Fig. 6.3.

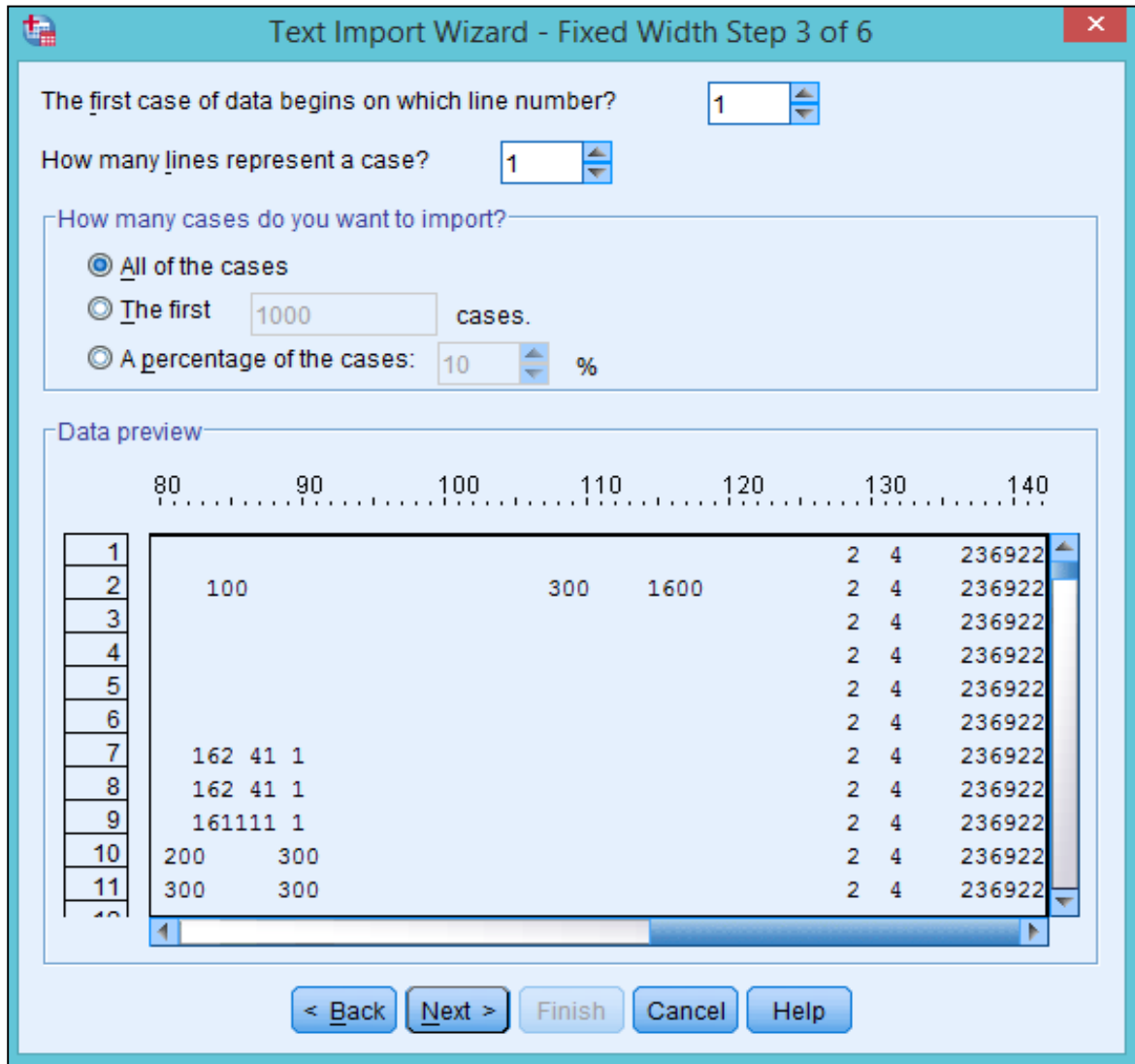


Fig. 6.3: Text Import Wizard (Step 3)

In the third Text Import Wizard reproduced in Fig. 6.3, each line represent a case. Select "1" in the first box as shown in the above in Fig. 6.3. Similarly, here the data begins with first line, so "1" is selected again in the second box. Next, choose how many case do you want to import? Since, all the cases are to be imported, select "all the cases" option. Click at "Next" button to get the fourth Text Import Wizard as shown in Fig. 6.4.

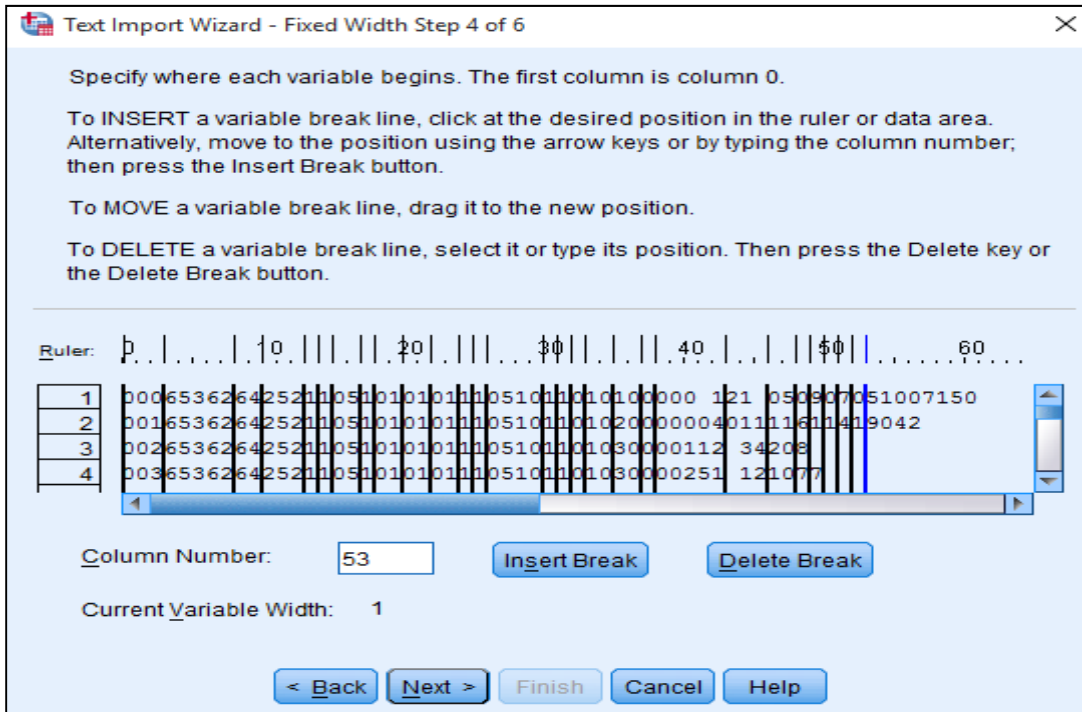


Fig. 6.4: Text Import Wizard (Step 4)

In order to extract variables from the dataset, the data is to broken in columns appropriately. In this example, variables are extracted by clicking on 3<sup>rd</sup> column, 8<sup>th</sup> column and so on which are represented in black lines as shown in Fig. 6.4. These breaks are created in accordance with the layout file provided by NSSO. After breaking the variables in appropriate columns, click at "Next" to get the fifth of six Text Import Wizard as shown in Fig. 6.5.

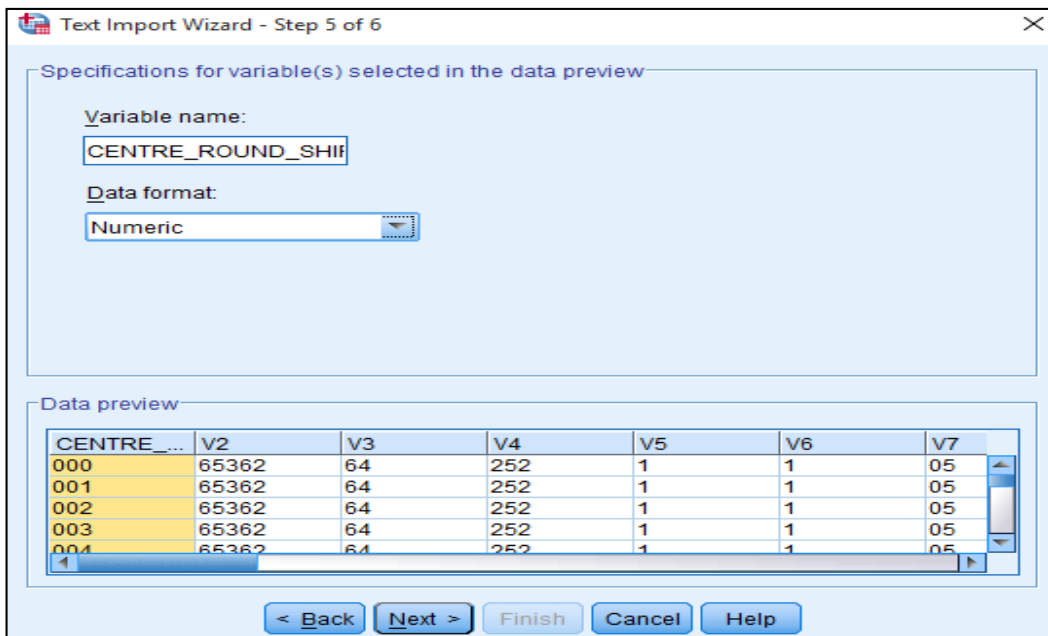


Fig. 6.5: Text Import Wizard (Step 5)



In fifth Text Import Wizard(Fig. 6.5), name and nature to each variable is to be assigned. In this example, the first variable is selected and assigned the name "CENTER\_ROUND\_SHIFT" and "numeric" as its nature. Likewise, assign names and nature to each variables. Click at "Next" button to get to the sixth and last Text Import Wizard as shown in Fig.6.6.

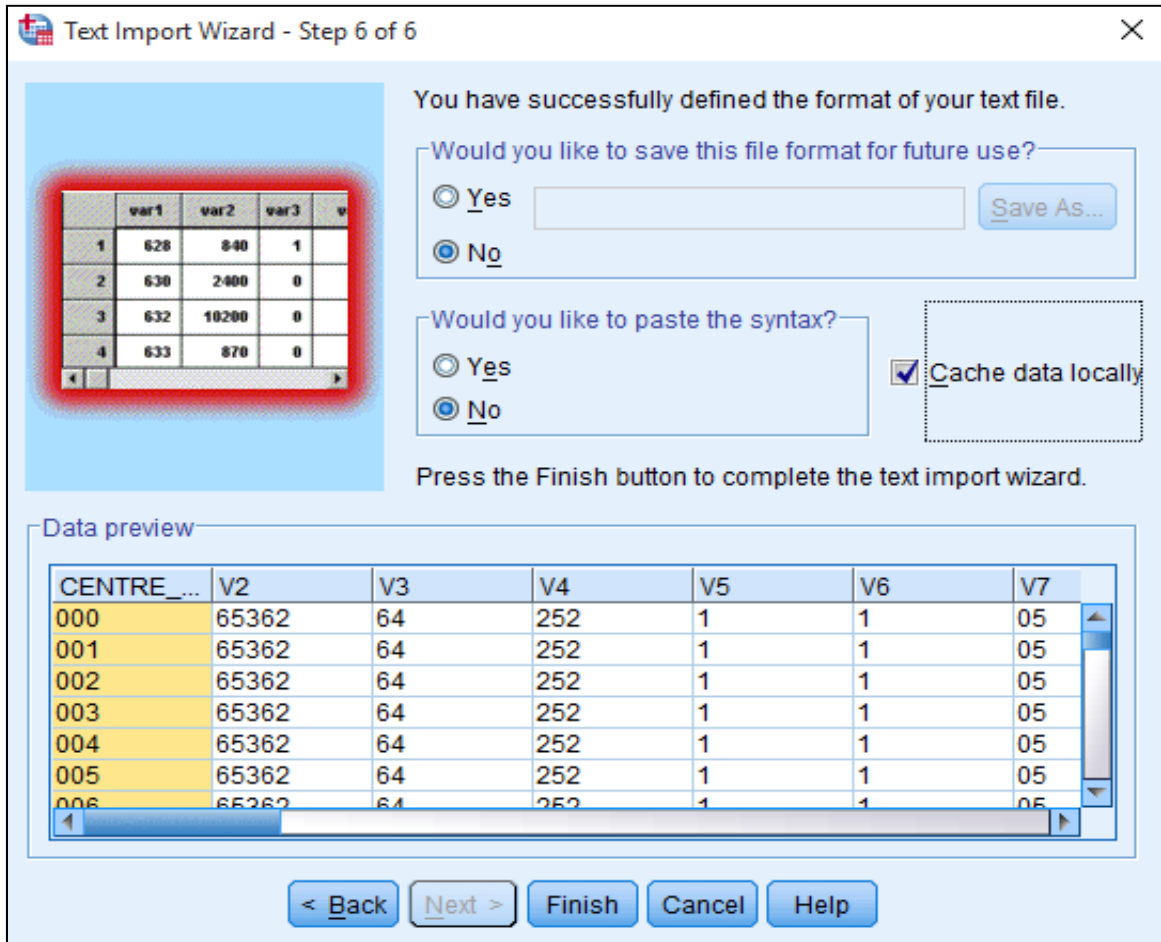


Fig. 6.6: Text Import Wizard (Step 6)

In the six and last step, select "Yes" or "No" options in the check boxes as shown in Fig. 6.6. Click at the "Finish" button to imported the data in the SPSS format as shown in Fig. 1. Cross check the number of observations in the SPSS file with the information provided in the supporting documents. In NSSO data, the supporting documents refers to the "README" file.

#### 4. Value Labels

The "Value Labels" function is available under the "View" function. This function is to be used to display the ordinal variable (i.e. rural=1 and urban=2), either in values or in labels in the window. To use value labels function, select the following options from the dropdown menu:


Select, View >  Value labels

As shown in Fig. 7, values of the variable are displayed as 1 in the "Sector" column.

The screenshot shows the SPSS Data View window. The 'SECTOR' column is highlighted in yellow, and all values in this column are '1'. The window title is '1 : SECTOR' and it shows 'Visible: 41 of 41 Variables'. The status bar at the bottom indicates 'IBM SPSS Statistics Processor is ready', 'Unicode:ON', and 'Weight On'.

	CENTRE_R	FSU_SL_NO	ROUND	SCH_NO	S_A	SECTOR	STATE	RE	DISRICT	STRATUM	SUB_S	S_U	S_U	F_U
1	001	64039	64	252	1	1	16	1	01	01	16	2	1	0
2	001	64039	64	252	1	1	16	1	01	01	16	2	1	0
3	001	61021	64	252	1	1	10	2	33	33	02	2	1	1
4	001	62796	64	252	1	1	28	4	08	08	02	2	2	2
5	001	62059	64	252	1	1	28	1	14	14	01	4	1	2
6	001	69148	64	252	1	1	18	2	07	07	01	2	2	1
7	001	63230	64	252	1	1	27	4	27	27	02	1	1	2
8	001	64159	64	252	1	1	12	1	02	02	01	1	2	0
9	001	61635	64	252	1	1	09	2	42	42	01	2	1	0
10	001	61150	64	252	1	1	09	4	41	41	01	3	2	0

Fig. 7: Display of Value Labels of Variable

In order to change the value labels of variable, click at  Value Labels. On clicking at "Value Label", screen shot of output that is displayed is shown in Fig. 8. The label of the variable is shown as "Rural" in the Sector variable.

The screenshot shows the SPSS Data View window after the value labels have been changed. The 'SECTOR' column now displays the word 'Rural' for all 10 rows. The window title is '1 : SECTOR' and it shows 'Visible: 41 of 41 Variables'. The status bar at the bottom indicates 'IBM SPSS Statistics Processor is ready', 'Unicode:ON', and 'Weight On'.

	CENTRE_R	FSU_SL_NO	ROUND	SCH_NO	S_A	SECTOR	STATE	RE	DISRICT	STRATUM	SUB_S	S_U	S_U	F_U
1	001	64039	64	252	1	Rural	Tr...	1	01	01	16	2	1	0
2	001	64039	64	252	1	Rural	Tr...	1	01	01	16	2	1	0
3	001	61021	64	252	1	Rural	B...	2	33	33	02	2	1	1
4	001	62796	64	252	1	Rural	A...	4	08	08	02	2	2	2
5	001	62059	64	252	1	Rural	A...	1	14	14	01	4	1	2
6	001	69148	64	252	1	Rural	A...	2	07	07	01	2	2	1
7	001	63230	64	252	1	Rural	M...	4	27	27	02	1	1	2
8	001	64159	64	252	1	Rural	A...	1	02	02	01	1	2	0
9	001	61635	64	252	1	Rural	U...	2	42	42	01	2	1	0
10	001	61150	64	252	1	Rural	U...	4	41	41	01	3	2	0

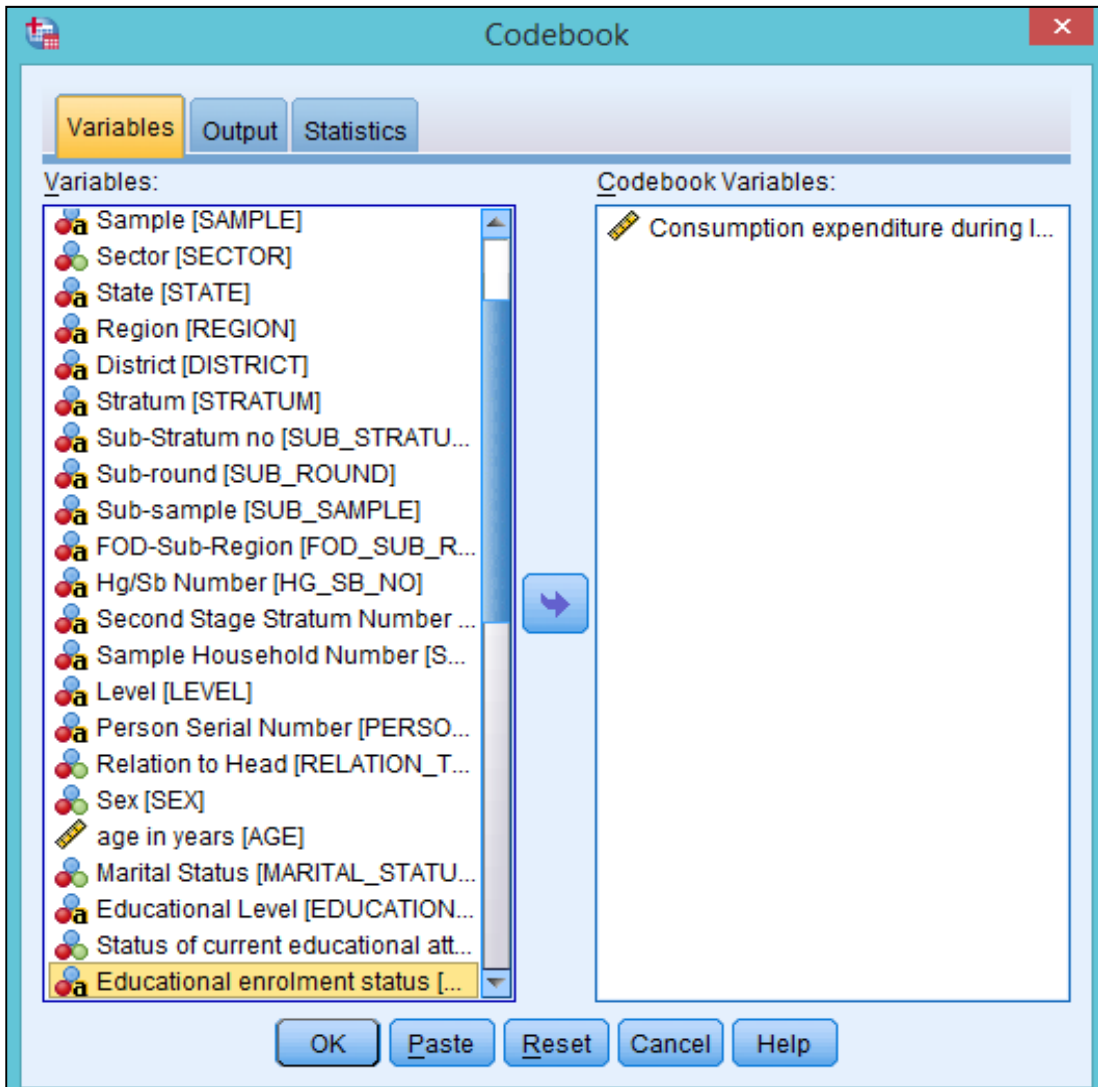
Fig. 8: Display of Changed Value Labels of Variable

## 5. Reports

The reports menu of SPSS provides basic information or summary of various variables in the dataset. The basic information includes valid cases, missing values, mean, median and mode, etc. To generate reports, select the following options from the dropdown menus:

**Analyze > Reports > Codebook or OLAP Cues or Case Summaries**

In this example, "Codebook" is selected first. On selection of "Codebook", Option Box that is displayed, is shown in Fig.9.



**Fig. 9: Codebook**

In the example given in Fig. 9, "consumption expenditure" is selected from the variable list in order to view information about it. Click at "OK" button, as a result Table 1 given below is displayed.

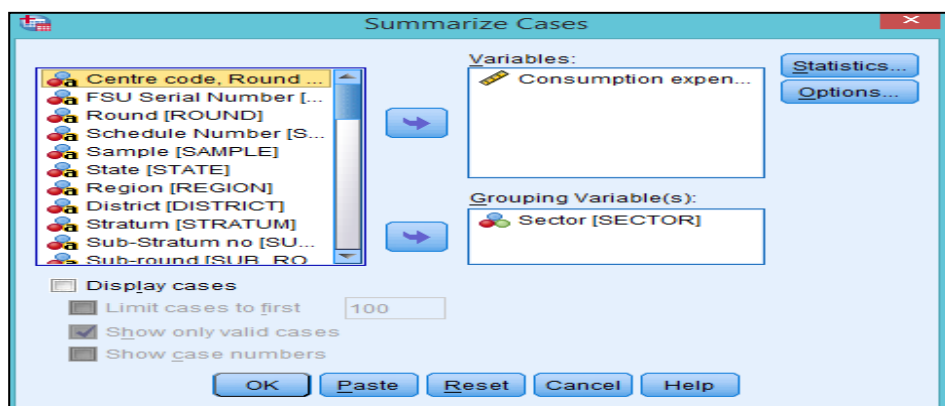
		Value
Standard Attributes	Position	49
	Label	Consumption expenditure during last 30 days on Free collections(Rs.)
	Type	Numeric
	Format	F4
	Measurement	Scale
	Role	Input
	Valid	240077
Central Tendency and Dispersion	Missing	205883
	Mean	248.32
	Standard Deviation	229.858
	Percentile 25	100.00
	Percentile 50	200.00
	Percentile 75	300.00

**Table 1: Display of Information Related to a Variable**

Table 1 above provides information on position, label, type, format, measurement, role, valid, missing, mean, standard deviation, and percentile of consumption expenditure, etc.

To summarize the cases available in the variable, select the options given below from the dropdown menus:

**Analyze > Reports > Case Summarize**



**Fig. 10: Summarize Cases of a Variable**

Select required "Variables" and "Grouping Variables" from left-end option window to right-hand selection box using arrow buttons. In this example, "Consumption expenditure" in variables and "Sector" in grouping variable is selected. After selecting both the variables, click at "OK" button. Resultant output is shown in Table 2 and Table 3.

	Cases					
	Included		Excluded		Total	
	N	Percent	N	Percent	N	Percent
Consumption expenditure during last 30 days on Total (Rs.) * Sector	445944	100.0%	16	0.0%	445960	100.0%

Table 2.1: Case Summarize of a Variable

Sector	Consumption expenditure during last 30 days on Total (Rs.)
Rural	290155
Urban	155789
Total	445944

Table 2.2: Case Summarize of a Variable

Table 2.1 and Table 2.2 provides number of cases involved in the consumption expenditure variable. Here, almost 100 percent cases are valid, involving 290155 rural and 155789 urban cases.

## 6. Descriptive Statistics

Descriptive statistics in SPSS is used to find out certain basic statistics of variables, e.g. frequencies of various variables, their summary, relationship between two variables, etc.

### 6.1 Frequency Calculation

To calculate the frequencies of rural and urban household in sample survey, select the following options from dropdown menu:

**Analyze > Descriptive Statistics > Frequencies**

On selection of the above mentioned menu items, "Frequencies" calculation box will appear as shown in Fig. 11. Select "sector" from **variable(s)** and checkbox for "**Display frequency tables**". Click at "OK". Table 4 will be displayed with results for frequencies calculation of specified variable(s).

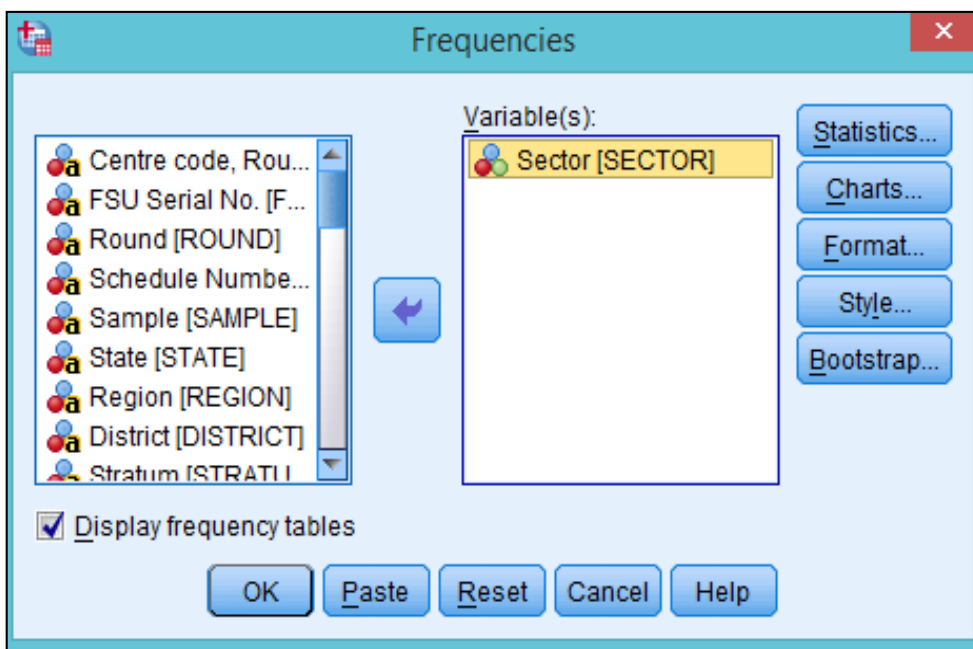


Fig. 11: Frequencies Calculation

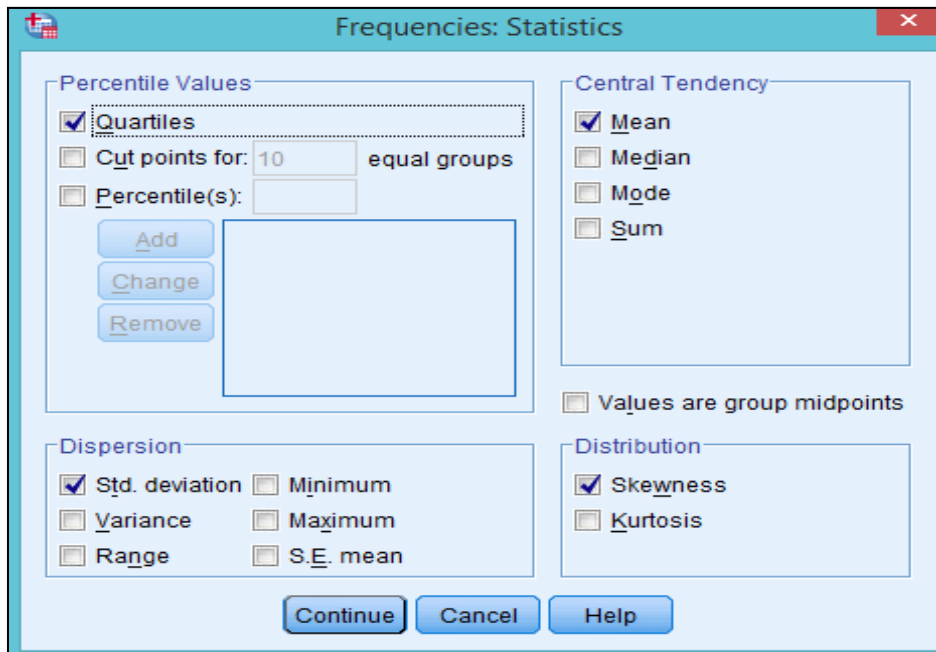
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Rural	63318	63.0	63.0	63.0
	Urban	37263	37.0	37.0	100.0
	Total	100581	100.0	100.0	

Table 3: Display of Frequencies Calculation Result

Table 3 reveals that there are 63,318 rural and 37,263 urban households selected for the survey. Results are also shown in terms of percentage, where 63 percent of households are rural and 37 percent are urban.

Further, click at **Statistics...** option in the "Frequencies" calculation box, to calculate several other statistics as shown in the Option Box for "Frequency : Statistics" given in Fig. 12.

You can calculate the percentile values, e.g. deciles, quartiles, etc which is shown in the upper left column of the Option Box in Fig. 12. You can also calculate the measures of central tendency, like mean, median, mode, etc. using "Statistics" function. Similarly, the measures of dispersion (like standard deviation, variance, and range) and distribution (skewness, kurtosis) can also be calculated with selection of appropriate options given in the left-side and right-side panel in the lower half of the Option Box, respectively (Fig. 12).



**Fig. 12: Frequencies: Statistics calculation**

For calculating mean, standard deviation, quartiles and skewness of the total consumption of households; check the appropriate boxes, e.g. **Quartiles, Std. Deviation, Mean, and Skewness**, as shown in the Fig. 12. Click at "**Continue**". Table 4 will be displayed as output of these selections.

N	Valid	100578
	Missing	3
Mean		3798.02
Std. Deviation		2931.657
Skewness		3.884
Std. Error of Skewness		.008
Percentiles	25	2000.00
	50	3000.00
	75	4700.00

**Table 4: Calculation of Different Statistics**

The Table 4 displays the mean, standard deviation, skewness and quartile class of total consumption expenditure of households. While the above two rows namely, "valid" and "missing" reveals that there are three missing values in the dataset and the remaining 100578 values are valid.

## 6.2 Comparing Two Variables (Cross Tabulation)

Relationship between two variables can be found using "Cross Tabulation" option available in SPSS. Select the following options from the dropdown menus for cross tabulation:

**Analyze > Descriptive Statistics > Crosstabs**

On selection of the above mentioned options from dropdown menu, the Option Box given in Fig. 13 will appear. In the example given here, relationship between "Social Group" of population and "Land Possessed" by them is calculated. Select "social group" variable in the **row(s)** and "land possessed" in the **column(s)** from the Option Box given in Fig. 13.

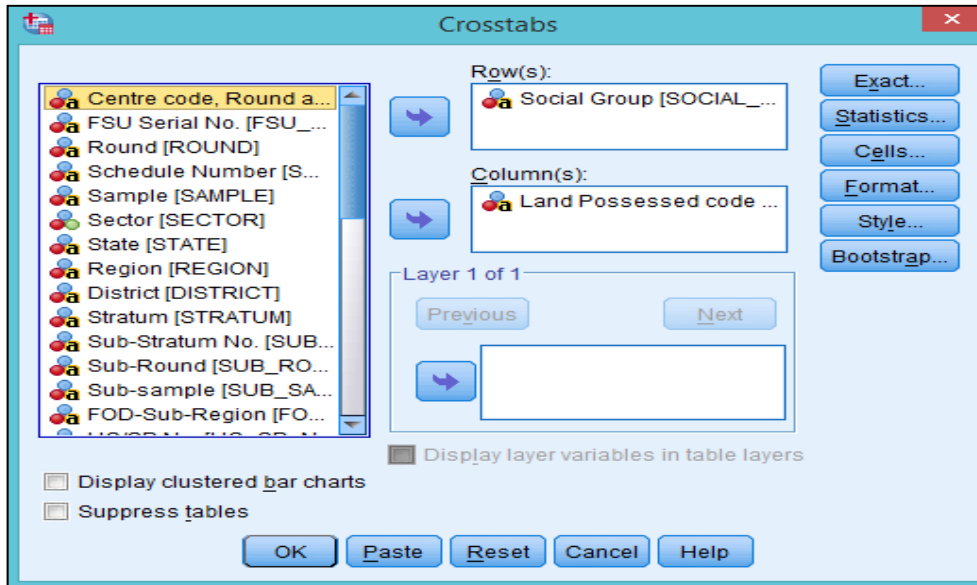


Fig. 13: Cross Tabulation

Click on "Cells" option given in the top right-hand panel (Fig. 13), to calculate the row-wise and column wise percentage. In this example, row wise percentage is calculated and observed counts, which is shown in the Fig. 14.

Further, check the "Counts" option and select "Percentages" option, to calculate the cross tabulation of two variables. In this example, cross tabulation of two variables, i.e. **Social Group \* Land Possessed code** is calculated as shown in Fig. 14.

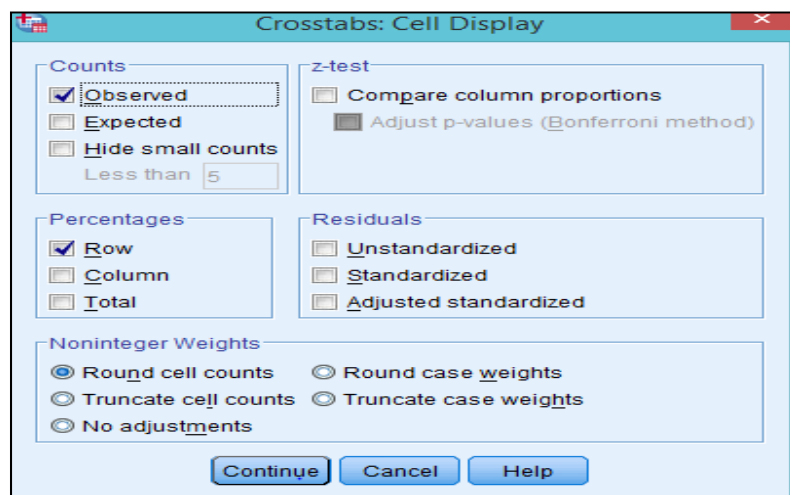


Fig. 14: Crosstabs: Cell Display



The resultant output has been displayed in Table 5, where "land possessed" variable is divided into two parts viz. 3.01 to 4.00 acres and greater than 8.00 acres. And, "social group" is classified into schedule tribe, schedule caste, OBCs and Others. This output reveals that most of the households have 3.01 to 4.00 acres of land.

			Land Possessed code		Total
			3.01 - 4.00 .....	greater than 8.00	
Social Group	scheduled tribe	Count	144	1	145
		% within Social Group	99.3%	0.7%	100.0%
	scheduled caste	Count	86	0	86
		% within Social Group	100.0%	0.0%	100.0%
	other backward class	Count	126	3	129
		% within Social Group	97.7%	2.3%	100.0%
	Others	Count	53	3	56
		% within Social Group	94.6%	5.4%	100.0%
Total		Count	409	7	416
		% within Social Group	98.3%	1.7%	100.0%

Table 5: Display of Result for Cross Tabulation between Variables

Further, one can also cross tab these two variables with one or more categorical variables e.g. for rural and urban sector, by using the layer option in "Crosstabs" as shown in Fig. 15. Select "Sector" variable in the "Layer" option and click at "OK". This layer option is not confined to only one categorical variable. As such, you can select more than one layer options for cross tabulation.

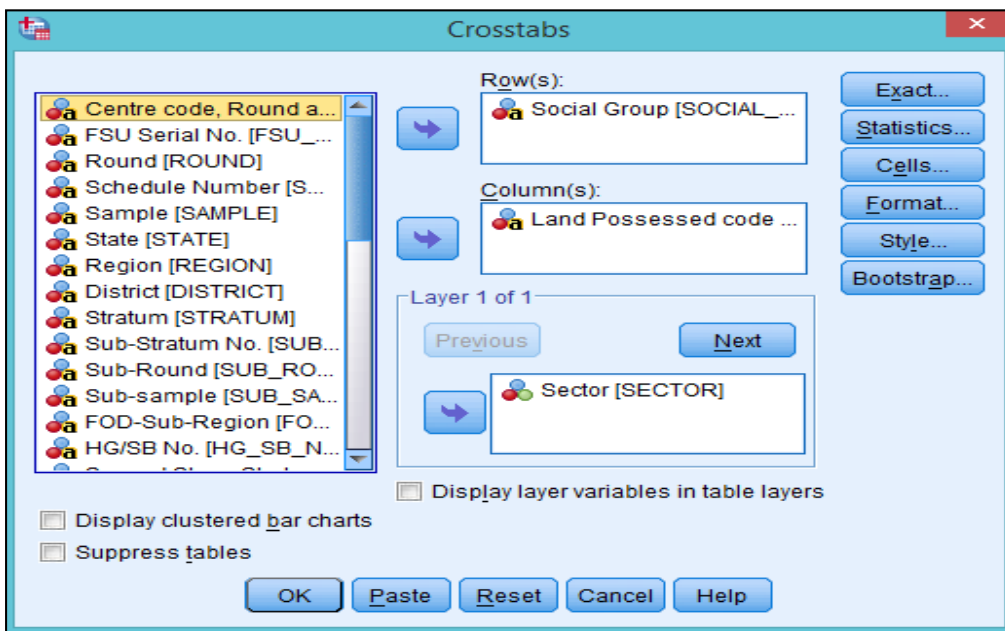


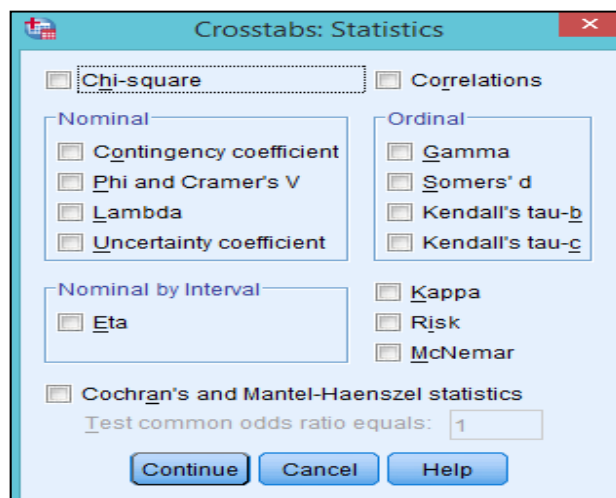
Fig. 15: Crosstabs Calculation Window

Resultant cross tabulation results are given in Table 6, where cross tabulation is calculated between **Social Group, Land Possessed code and Sector**.

Sector	Land Possessed code			Total		
		3.01 - 4.00 .....	greater than 8.00			
<b>Rural</b>	Social Group	scheduled tribe	22492	17436343	167831	17626666
		scheduled caste	971	34602163	123130	34726264
		other backward class	22311	67290135	1417870	68730316
		others	11003	36765957	1404156	38181116
		<b>Total</b>		<b>56777</b>	<b>156094598</b>	<b>3112987</b>
<b>Urban</b>	Social Group	scheduled tribe	3547	1946002	3070	1952619
		scheduled caste	5471	9471957	8848	9486276
		other backward class	13913	22913112	70716	22997741
		others	20811	28619421	169369	28809601
		<b>Total</b>		<b>43742</b>	<b>62950492</b>	<b>252003</b>
<b>Total</b>	Social Group	scheduled tribe	26039	19382345	170901	19579285
		scheduled caste	6442	44074120	131978	44212540
		other backward class	36224	90203247	1488586	91728057
		others	31814	65385378	1573525	66990717
		<b>Total</b>		<b>100519</b>	<b>219045090</b>	<b>3364990</b>

**Table 6: Crosstabs Calculation for Social Group, Land Possessed code and Sector**

Further, as shown in Crosstabs Calculation Window (Fig. 16), various statistics between the two variables can also be calculated, e.g. chi-square, correlations, etc. For that, click at **Statistics...** option in the "Crosstab" window and select the required statistics to be calculated from the Option Box, then click at "Continue" to get the results.



**Fig. 16: Crosstabs Statistics Calculation**

7. Sort Cases

The "Sort Cases" data in SPSS is used to sort the data in ascending or descending order. This task can be performed by selecting following options from the dropdown menus:

**Data > Sort Cases... Select Ascending / Select Descending**

Select "ascending order" to sort your data in ascending order, else select "descending order". In the example given here, "State codes" selected for sorting in "ascending order" using "Sort Order" option as shown in Fig. 17.

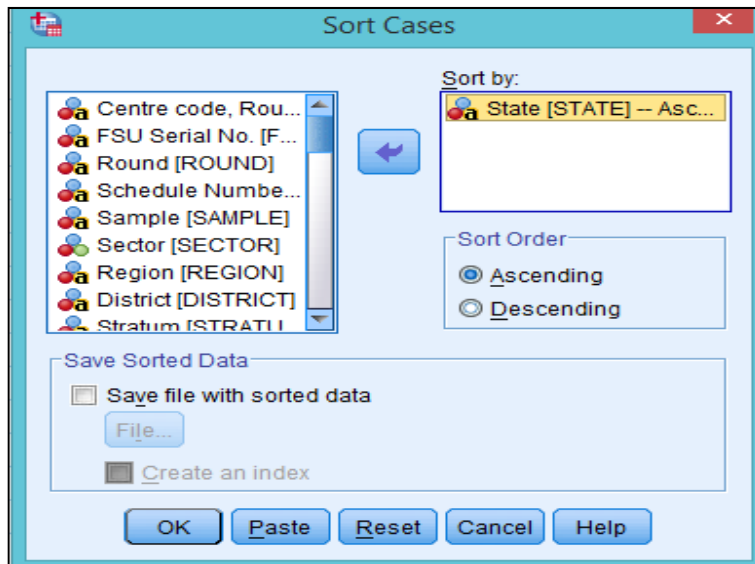


Fig. 17: Sort Cases

Table 7 below displays data arranged in ascending order by state codes.

	CENTRE_R...	FSU_SL_NO	ROUND	SCH_NO	SA	SECTOR	STATE	RE	DIS	STR	SU	S	S	FOD_S	H	S	SA	LEV	HH_SIZE	NIC_200	NCO_2004
1	001	65528	64	252	1	1	01	3	04	04	02	1	1	0120	1	1	03	02	14	01111	920
2	001	65665	64	252	1	1	01	3	02	02	01	3	2	0122	1	1	03	02	6	01134	920
3	001	65569	64	252	1	1	01	1	13	13	04	1	1	0110	1	1	02	02	7	45201	931
4	001	65691	64	252	1	1	01	3	06	06	02	4	2	0121	1	1	01	02	6	01111	611
5	001	65528	64	252	1	1	01	3	04	04	02	1	1	0120	1	1	01	02	9	01111	920
6	001	65545	64	252	1	1	01	1	13	13	03	2	1	0110	1	1	02	02	7	45201	931
7	001	65691	64	252	1	1	01	3	06	06	02	4	2	0121	1	2	02	02	3	01111	611
8	001	65583	64	252	1	1	01	3	03	03	02	2	2	0120	1	1	02	02	8	45202	931
9	001	65691	64	252	1	1	01	3	06	06	02	4	2	0121	1	2	01	02	4	01111	611
10	001	65547	64	252	1	1	01	2	09	09	01	3	2	0111	1	1	02	02	11	45203	931
11	001	65503	64	252	1	1	01	3	06	06	03	1	2	0121	2	2	02	02	7	01111	611

Table 7: Display of Result in Sorting Cases Order

8. Select Cases

The "Select Cases" option in SPSS works as the filter function. Select options given below from the dropdown menu to filter the data:

**Data > Select Cases**

On selection of above-mentioned options, Option Box that is displayed is given in Fig. 18. Select the appropriate variable, e.g. "Sector" to bring it to the "Use filter variable" box. Click at "OK" as shown in Fig. 18. The data is filtered out and is displayed as shown in Table 8.

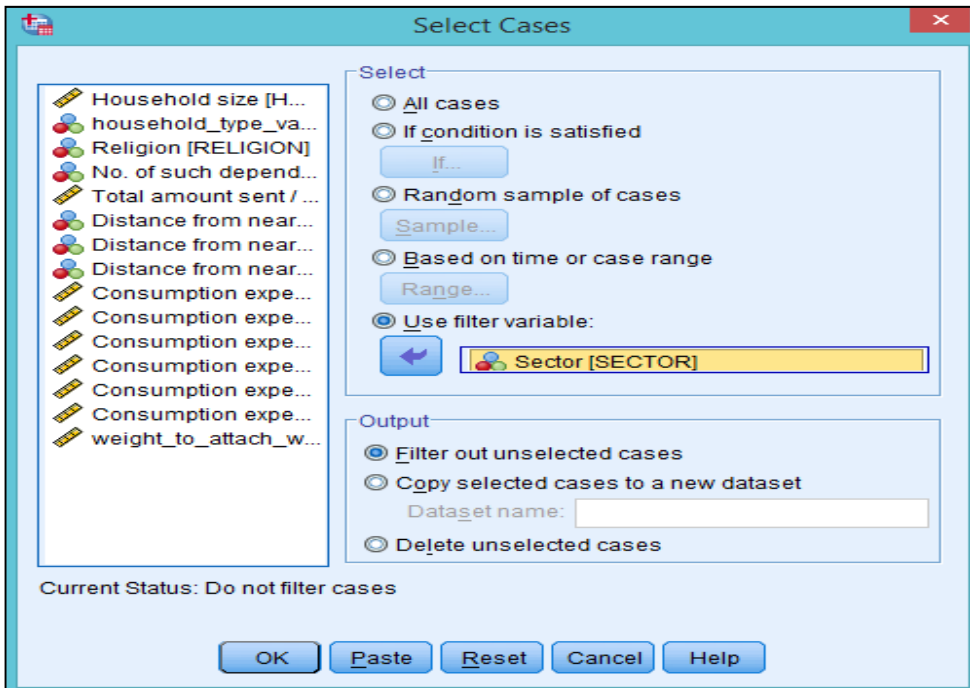


Fig. 18: Selecting and Filtering Cases

The image shows the SPSS Data View window for a dataset named '\*Block-3 Household char'. The window displays a table with 11 rows and 7 columns. The columns are labeled 'CENT RE\_R', 'FSU\_SL\_NO', 'RO UN D', 'SCH\_NO', 'S A', and 'SECTOR'. The 'SECTOR' column contains the value 'Rural' for all 11 rows. The 'Data View' tab is selected at the bottom.

	CENT RE_R	FSU_SL_NO	RO UN D	SCH_NO	S A	SECTOR
1	001	64039	64	252	1	Rural
2	001	64039	64	252	1	Rural
3	001	61021	64	252	1	Rural
4	001	62796	64	252	1	Rural
5	001	62059	64	252	1	Rural
6	001	69148	64	252	1	Rural
7	001	63230	64	252	1	Rural
8	001	64159	64	252	1	Rural
9	001	61635	64	252	1	Rural
10	001	61150	64	252	1	Rural
11	001	63155	64	252	1	Rural

Table 8: Filtering Cases on Sector

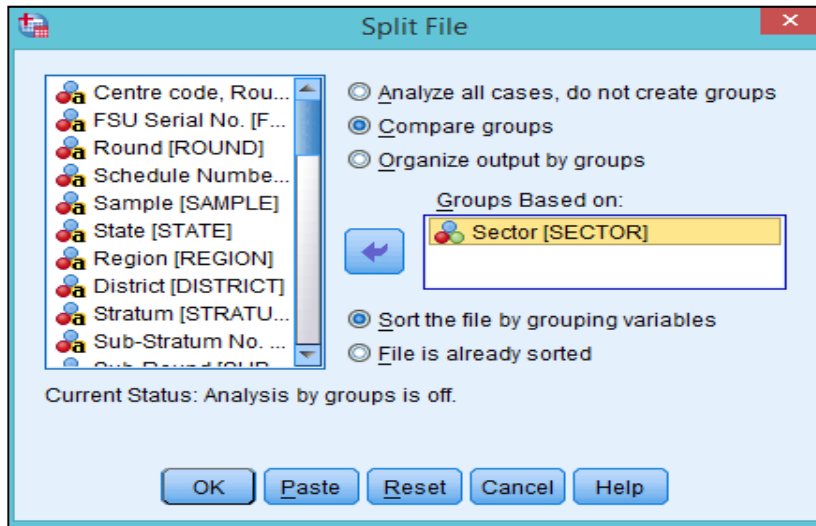
9. Split Data

Split data command is very useful to compare the information of a variable in terms of categorical variables. In the example given here, "consumption expenditure "is compared for rural and urban sectors.

Select the following options from dropdown menu to do so:

**Data > Split File...**

On selection of options given above, the Option Box that appears is shown below in Fig. 19. Select **Compare groups** and take "Sector" variable in the Selection Box. Click at "OK". Select the option **Split by SECTOR** that appears at the bottom right-end of the Option Box.



**Fig. 19: Option Box for Split File**

The output file is displayed as shown in Table 9. Here, descriptive statistics of the variable "Consumption Expenditure" is calculated and the result is separated for rural and urban sectors.

Sector		N	Min.	Max.	Sum	Mean	Std. Deviation
<b>Rural</b>	Consumption expenditure during last 30 days on Total (Rs.)	159260258	100	53500	468634075955	2942.57	1865.105
	Valid N (list wise)	159260258					
<b>Urban</b>	Consumption expenditure during last 30 days on Total (Rs.)	63246236	1	80000	333380137424	5271.15	3916.585
	Valid N (list wise)	63246236					

**Table 9: Split File based on Sector**

## 10. Custom Tables

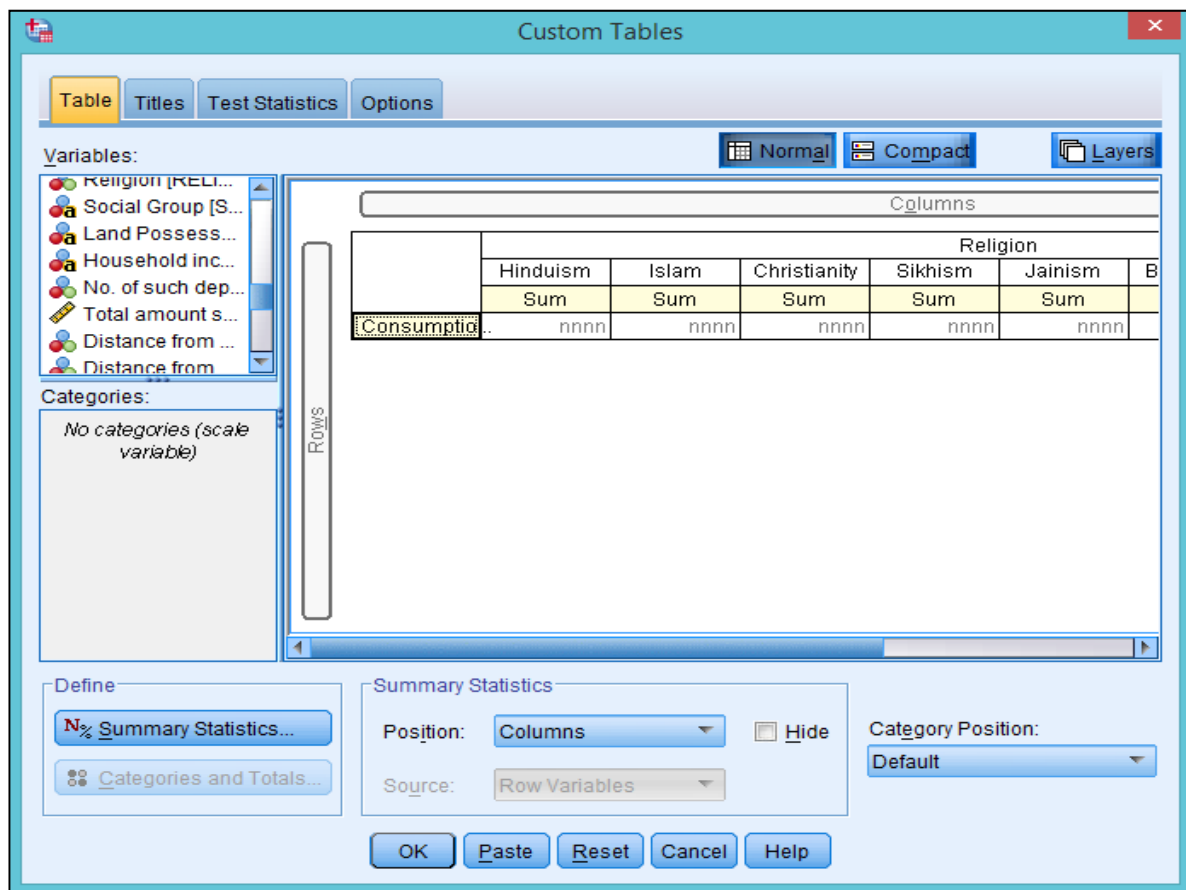
Custom table function in SPSS is useful to prepare a wide range of tables, e.g. two way tables, three way tables, four way tables, etc.

### 10.1 Two Way Tables

Two way tables calculate the relationship between two variables. For example, total consumption expenditure of households hailing from different religious groups can be calculated by selecting options given below from the dropdown menus:

**Analyze > Tables > Custom Tables.**

On selection of options given above, Option Box given in Fig. 20 will appear.



**Fig. 20: Calculation of Custom Tables**

Select "**Consumption Expenditure**" variable from the variables list to Rows and "**Religion**" to Columns as shown in Fig. 20. Select any statistics, like sum, count, mean, percentage, std. deviation and so on,

by clicking at **Summary Statistics...** from the **Define** menu. Select the desired statistics and click at **Apply to Selection** as shown in Fig. 21. Click at "**OK**".

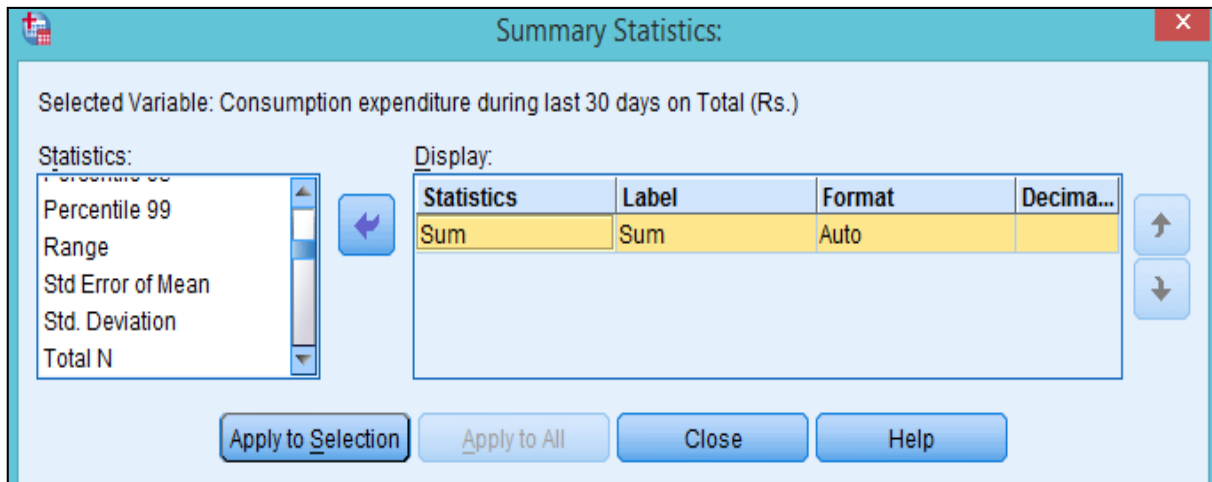


Fig. 21: Calculation of Two-way Custom Tables

Table 10 will appear as output with "religion" shown in "columns" and "consumption" in "row".

	Religion							
	Hinduism	Islam	Christianity	Sikhism	Jainism	Buddhism	Zoroastrianism	No Response
	Sum	Sum	Sum	Sum	Sum	Sum	Sum	Sum
Consumption	655313872428	92345683331	20680331076	1762758933	23693233082	3788032118	4260282502	170019908

Table 10: Two-way Custom Table: Religion Vs Consumption Expenditure

### 10.2 Three Way Tables

Three way tables provide information on two variables with one categorical variable. In continuation of previous example, we can calculate "total expenditure of households" for different religion with a categorical variable such as rural and urban, i.e. total expenditure of different religious groups in rural and urban sector. To do so, select the options given below from the dropdown menus:

**Analyze > Tables > Custom Tables**

On selection of above-mentioned options, screenshot of Option Box that is displayed is given in Fig. 22. Select "**Consumption Expenditure**" from rows and "**Religion**" from columns. Select "**Sector**" variable in rows (before consumption) for calculating consumption expenditure by rural and urban population and choose the required statistics. Click at "OK" as shown in Fig. 22. The resultant output is given in Table 11.

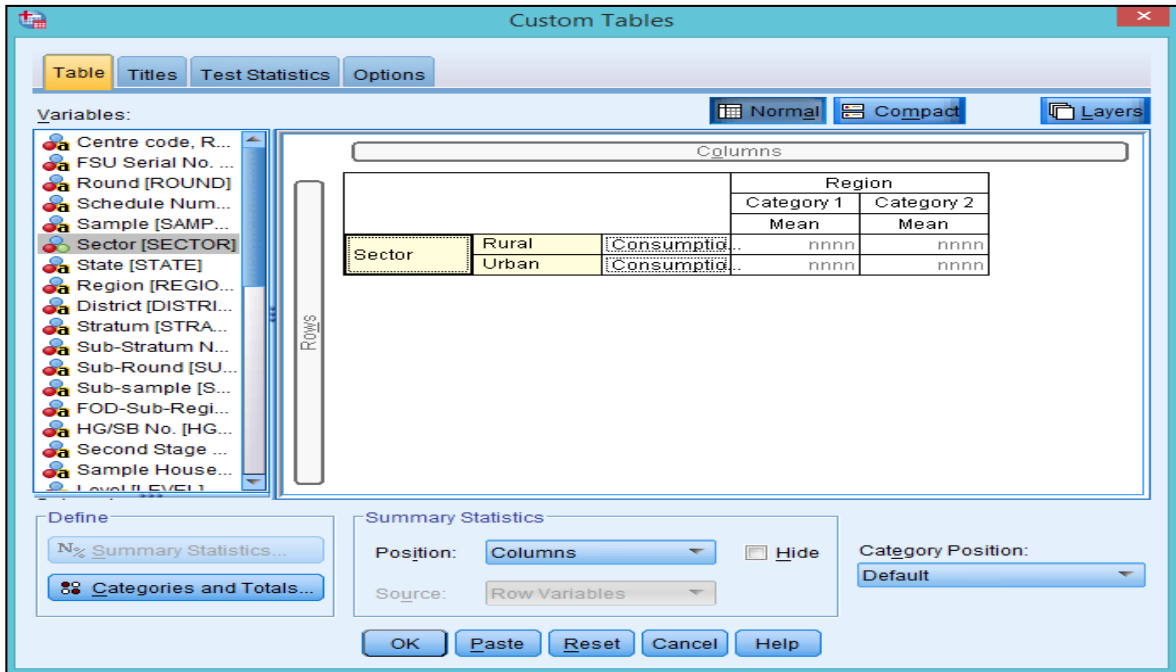


Fig. 22: Calculation of Three-way Custom Tables

Likewise, you may calculate four ways, five ways tables by selecting categorical variables in rows and columns.

		Region					
		1	2	3	4	5	6
		Sum	Sum	Sum	Sum	Sum	Sum
SectorRural	Consumption expenditure during last 30 days on Purchase (Rs.)	104669299104	101575051718	61525523744	36782466440	39967622178	4314904812
Urban	Consumption expenditure during last 30 days on Purchase (Rs.)	123867218545	74612338570	64969819244	29540417949	26766574845	2854422799

Table 11: Three-way Custom Table

Table 11 given above displays total consumption of different religion belonging to rural and urban areas.



## 11. Compute Variable

New variables can be calculated by adding, subtracting, multiplying (etc.) two or more numerical variables using "compute variable" option. In this example, sum of expenditure of two variables, namely, gift and free collection is being calculated. Select the following options from the dropdown menus:

**Transform > Compute Variable.**

On selection of above mentioned options, an Option Box will appear on the screen wherein write the name of new variable in the "Target Variables" box and select the two variables in "Numeric Expression". In this example, "Gift\_free" is selected in "Target Variables" box, "gifts and loans" and "free collection" variables are selected in "Numeric Expression". To add values of these two variables, "addition" function is required as shown in Fig. 23. Click at "OK" button. As a result, "Gift\_free", i.e. a new variable will be computed in the dataset and result will be displayed as shown in Table 12. Likewise additional statistical values can be computed by combining two, three or more numbers of variables.

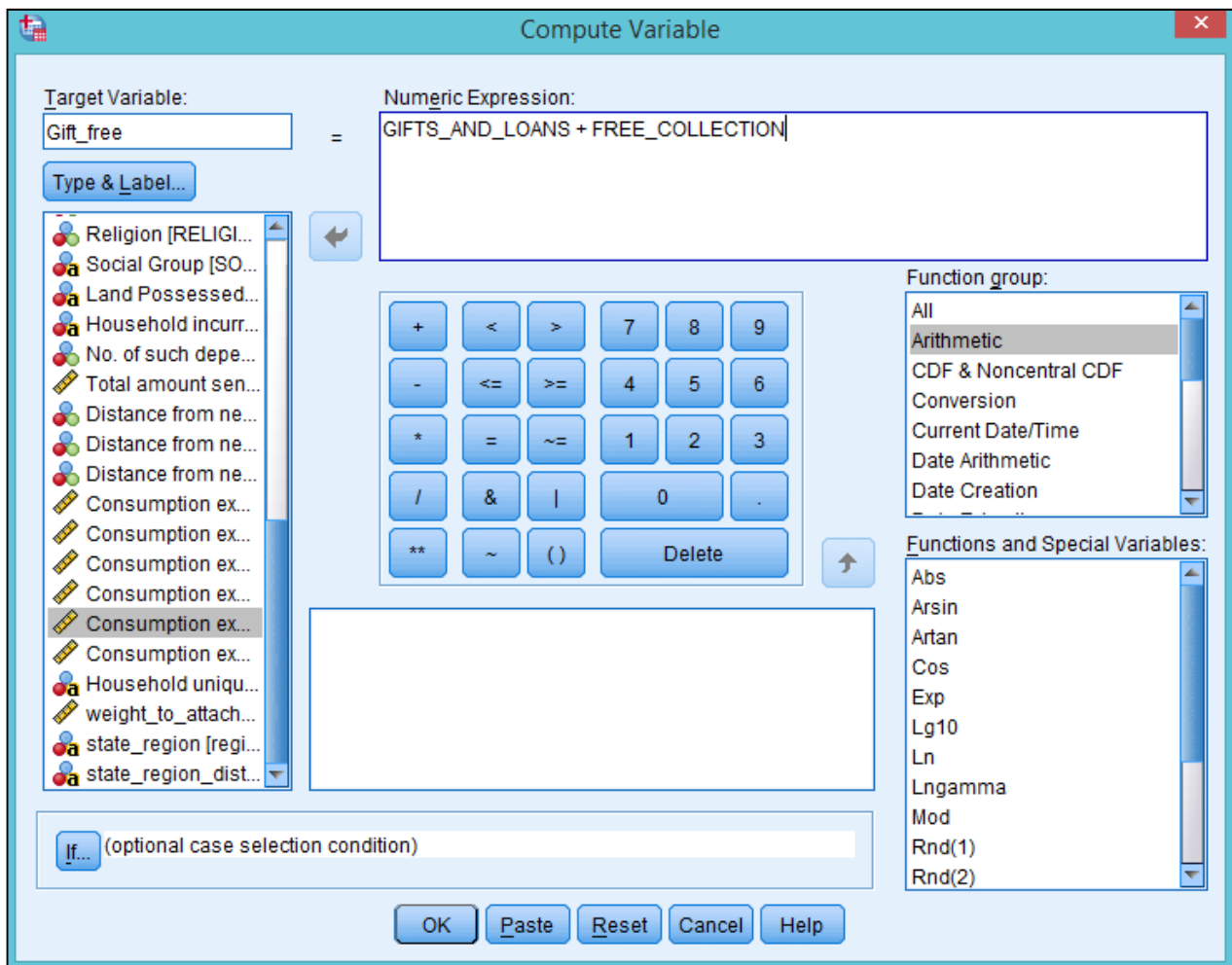


Fig. 23: Computing Additional Variable

The screenshot shows the IBM SPSS Statistics Data Editor window. The title bar reads '\*Block-3 Household characteristics.sav [DataSet1] - IBM SPSS Statistics Data Editor'. The menu bar includes File, Edit, View, Data, Transform, Analyze, Direct Marketing, Graphs, Utilities, Add-ons, Window, and Help. The toolbar contains various icons for file operations and analysis. The main window displays a data table with 11 rows and 12 columns. The first column is labeled 't\_c' and contains values 1 through 11. The second column is labeled 'gift\_free' and contains values 150.00 and 245.00 for rows 9 and 10 respectively. The other columns are labeled 'var'. The status bar at the bottom indicates 'IBM SPSS Statistics Processor is ready', 'Unicode:ON', and 'Weight On'.

t_c	gift_free	var	var	var	var	var	var	var	var	var	var
1											
2											
3											
4											
5											
6											
7											
8											
9	150.00										
10	245.00										
11											

Table 12: Computed Variable- "gift\_free"

Further, the option **If...** (optional case selection condition) given at the bottom can also be used for calculating the new variable involving categorical variable. Continuing with our example, to add "gifts and loans" and "free collection "only for "rural" sector, select "If". The Option Box given in Fig. 24 will appear. Click on at **Include if case satisfies condition:** and select "Sector" variable. Use "=" sign from the symbol box given in Fig. 24, select type 1 for rural sector, and click at "Continue" as shown in Fig. 24. A new variable will be created only for rural areas. Similarly, this function can be performed for urban areas by selecting type 2 in the "include if case satisfies condition" function.

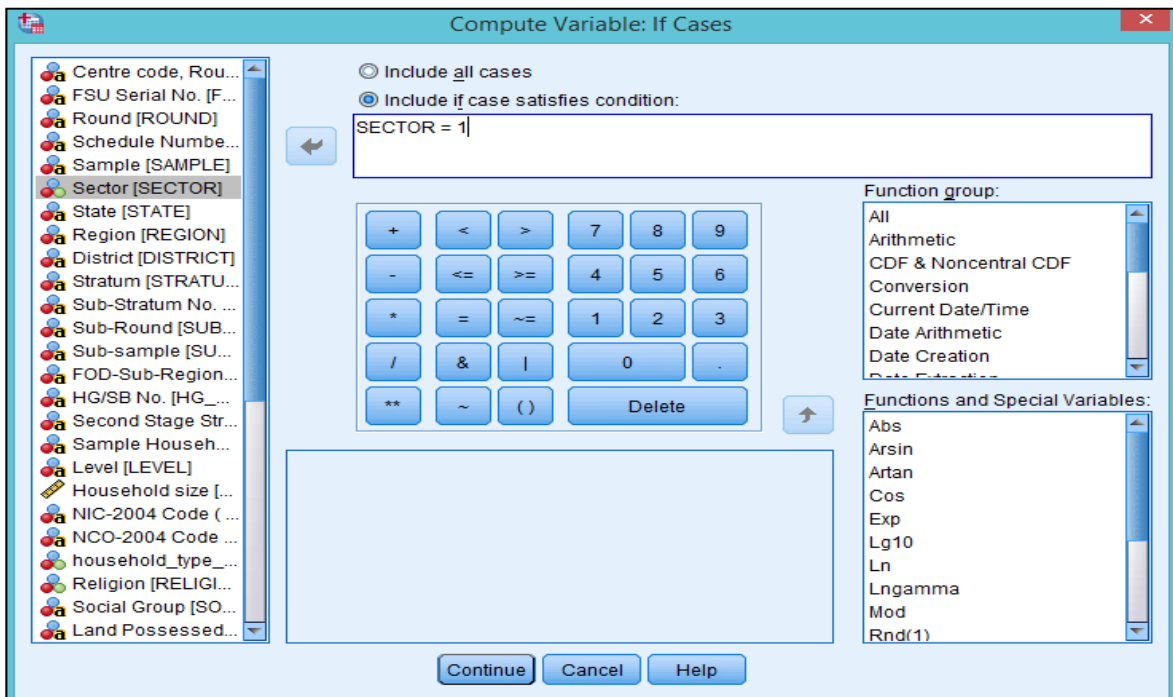


Fig. 24: Compute Variable: If Cases

1: DIST_FROM_SEC...	4	GIFTS_AND LOANS	filter_\$	var	var	var	var	var	var	var	var	var	var	var	var	var
1	in: reg... gr... 2	60000	Not Selected													
2	in: reg... gr... 2	60000	Not Selected													
3	in: reg... gr... 2	60000	Not Selected													
4	in: reg... gr... 2	60000	Not Selected													
5	in: reg... gr... 2	60000	Not Selected													
6	in: reg... gr... 2	60000	Not Selected													
7	al: others... 0... 2	26000	Selected													
8	al: others... 0... 2	26000	Selected													
9	al: others... 0... 2	26000	Selected													
10	al: others... 0... 2	26000	Selected													
11	in: reg... 3... 2	20000	Not Selected													
12	in: reg... 3... 2	20000	Not Selected													
13	in: reg... 3... 2	20000	Not Selected													
14	in: reg... 3... 2	20000	Not Selected													
15	in: reg... 3... 2	20000	Not Selected													
16	in: reg... 3... 2	20000	Not Selected													
17	in: reg... 3... 2	20000	Not Selected													
18	al: self... gr... 1 1	19600	Selected													
19	al: self... gr... 1 1	19600	Selected													
20	al: self... gr... 1 1	19600	Selected													
21	al: self... gr... 1 1	19600	Selected													
22	al: self... gr... 1 1	19600	Selected													
23	al: self... gr... 1 1	19600	Selected													

Table 13: Computed Variable: If Cases

## 12. Recoding of Data

The "recoding of data" option is used when the two or more variables are to be combined. In the example given above, there is a variable called, "distance from home to primary school", with value varying from 1 to 5 kilometres . For combining values 1 and 2 into a single value e.g. 10, options to be selected are as follows:

**Transform > Recode Into different Variables.**

On selection of options given above, Option Box that is displayed is given in Fig. 25.1 with header "Recode Into different Variables".

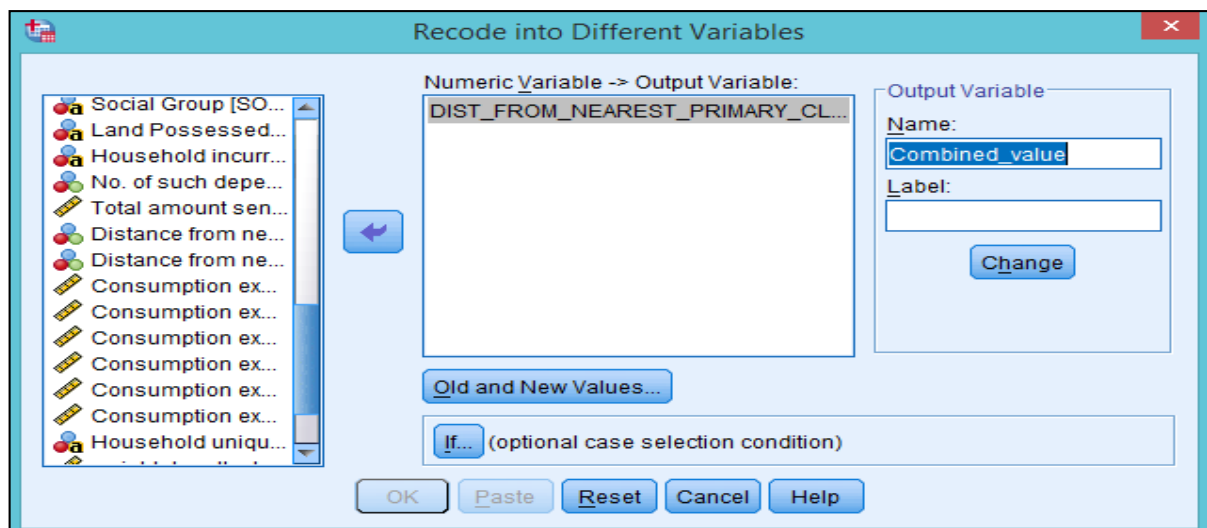
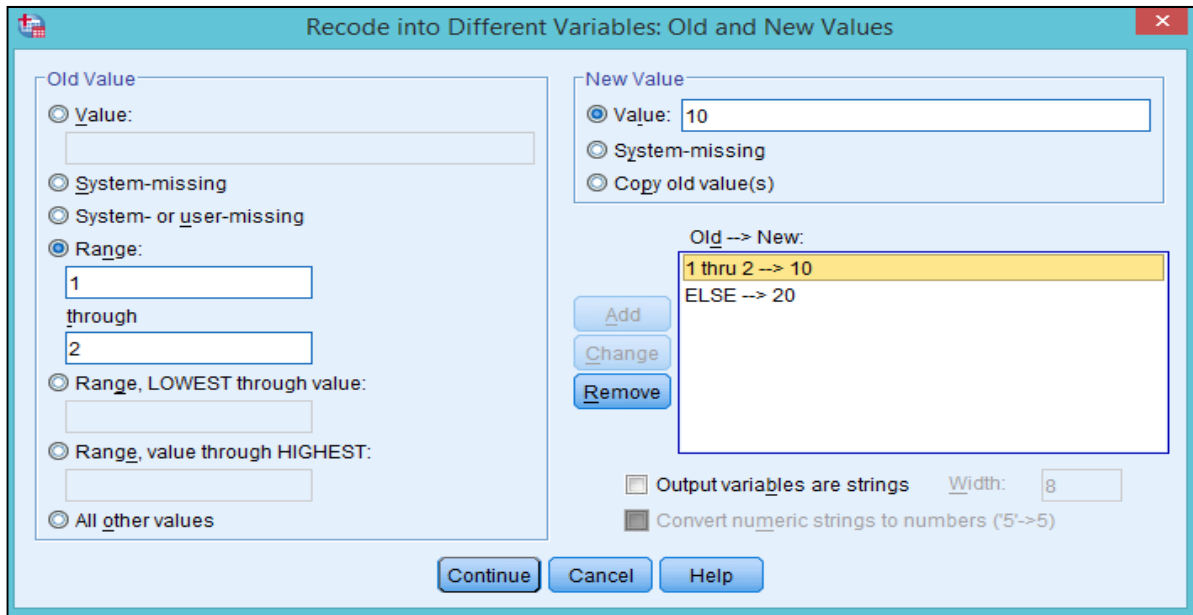


Fig. 25.1: Recoding of Different Variables

Select "distance variable" from list of variables into Selection Box, type the name you want to assign to the new variable. Click at **Old and New Values...**. A new Option Box will appear as shown in Fig. 25.2.



**Fig.25.2:Recoding of Different Variables: Old and New Values**

Select the "range" and write "1" and "2" in the boxes under "Range" of Old Value menu, and select the "Value" under the New Value menu and click at "Add" button. Select all other values, e.g. "20" and click at "Continue". As a result, a new variable with combined values will be created in the dataset as shown in Table 14.

	combined_value	var	var	var	var	var	var	var	var
1	10.00								
2	10.00								
3	10.00								
4	10.00								
5	10.00								
6	10.00								
7	10.00								
8	10.00								
9	10.00								
10	10.00								
11	10.00								

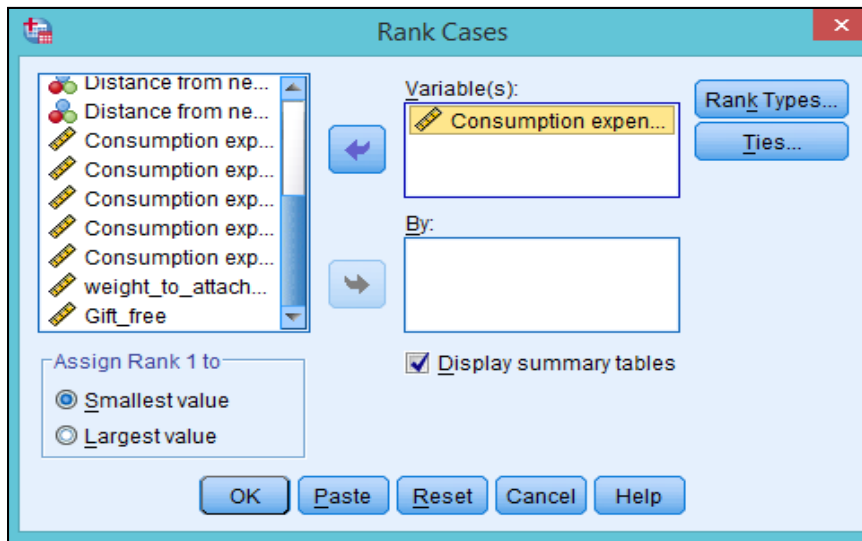
**Table 14: Combined Values for Recoding of Different Variables**

### 13. Rank Cases

The "rank cases" option can be used to distribute the whole dataset into various equal parts i.e. from 1 to 100 or 10 to 100 etc. Here, we will calculate "decile" class of "total consumption expenditure" variable i.e. distributing the consumption expenditure into 10 equal parts. The options for calculation of "decile" class is as follows:

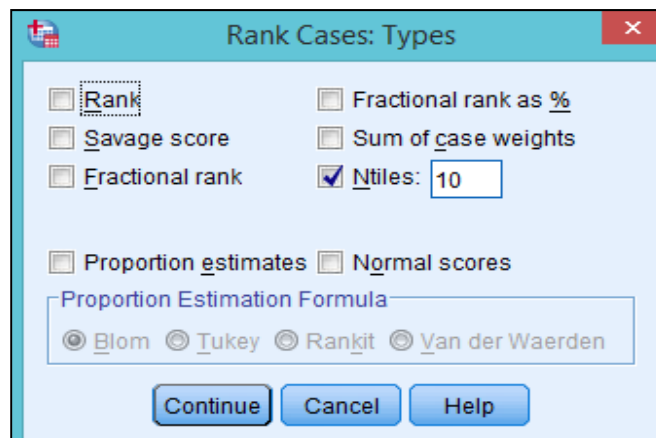
**Transform > Rank Cases.**

A new Option Box is displayed as shown in Fig. 26. Select the "consumption expenditure" variable in the "variable(s)" box and Click on **Rank Types...** as shown in Fig. 26.



**Fig. 26: Calculation of Rank Cases**

The new Option Box that is displayed is given in Fig. 26.1. Check the "Ntiles" box and type 10 into it. Click on "Continue", a new variable with the name of **NTOTAL** will be created and a new Option Box(Fig. 26.3) prompting users to assign value to the "Tie", i.e. low, high or low.



**Fig. 26.1: Calculation of Rank Cases: Types**

In this example, "mean" value is assigned to the "Tie" as shown in Fig. 26.3. Click at "Mean" and "Continue". Resultantly, Table 15 is displayed that shows newly created variable based on "decile" values.

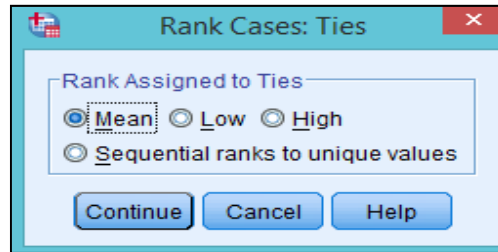


Fig. 26.3: Calculation of Rank Cases: Ties

	NTOTAL	var	var	var	var	var	var	var	var	var	var
1	.										
2	.										
3	.										
4	1										
5	1										
6	1										
7	1										
8	1										
9	1										
10	1										
11	1										

Table 15: Calculation of Rank Cases

To distribute the dataset into 4 equal parts i.e. **quartile**, click at "Ntiles" and type "4" in the box, click at "Continue" and "OK" as shown in Fig. 26.4. A new variable with the quartiles will be created in the dataset as similar to Table 15.

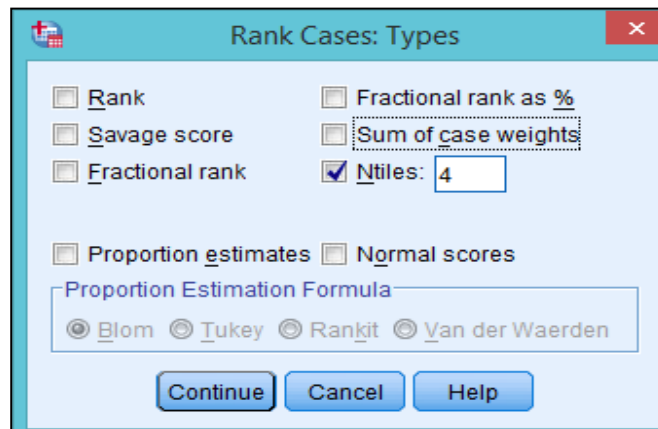


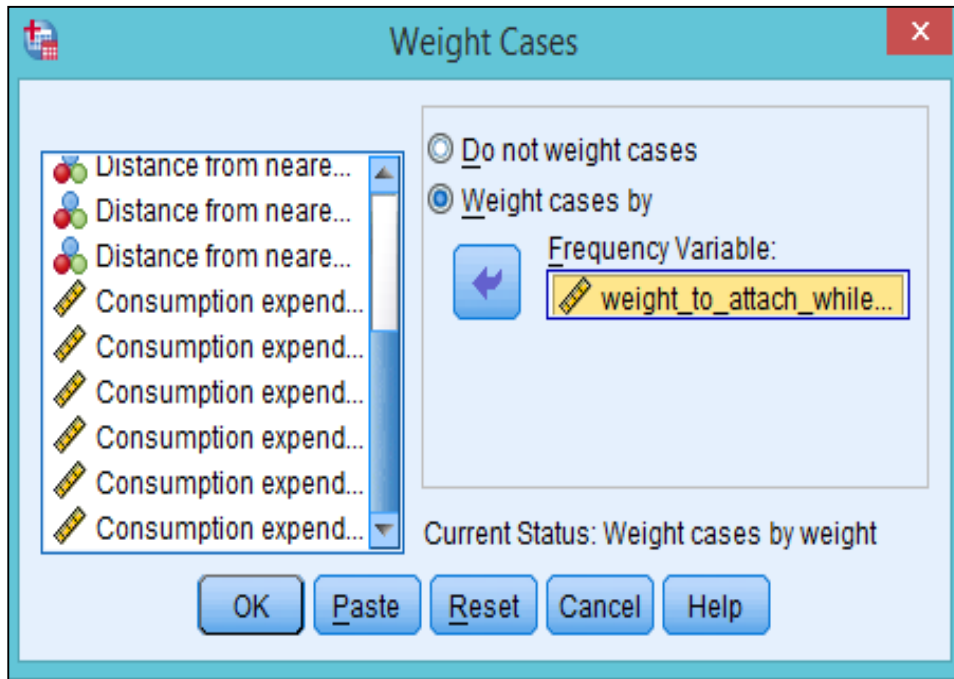
Fig. 26.4: Calculation of Rank Cases: Types

## 14. Weight Cases

The "weight cases" command in SPSS is very useful for taking an estimate of the data from sample data. The NSSO data set provide the "**Multiplier**" which is to be used while estimating the data. In SPSS, cases are weighted by a variable with the help of following function:

**Data > Weight Cases.**

On selection of above mentioned options, the Option Box is that is displayed is shown in Fig. 27.



**Fig. 27: Calculation of Weight Cases**

Select the "**Weight cases by**" and select the "**weight to attach variable**" in the "Frequency Variable" box and click at "OK". The **Weight On** result will appear in the bottom right of the window.

## 15. Aggregate Data

The "aggregate data" command of SPSS convert the data files from the disaggregated level to aggregated level. For example, the dataset that we are dealing with provides information on consumption expenditure for each household for different states of India. To create a dataset on state-wise consumption expenditure information, select the following options from the dropdown menus:

**Data > Aggregate**

On selection of above mentioned options, the Option Window that is displayed is given in Fig. 28.

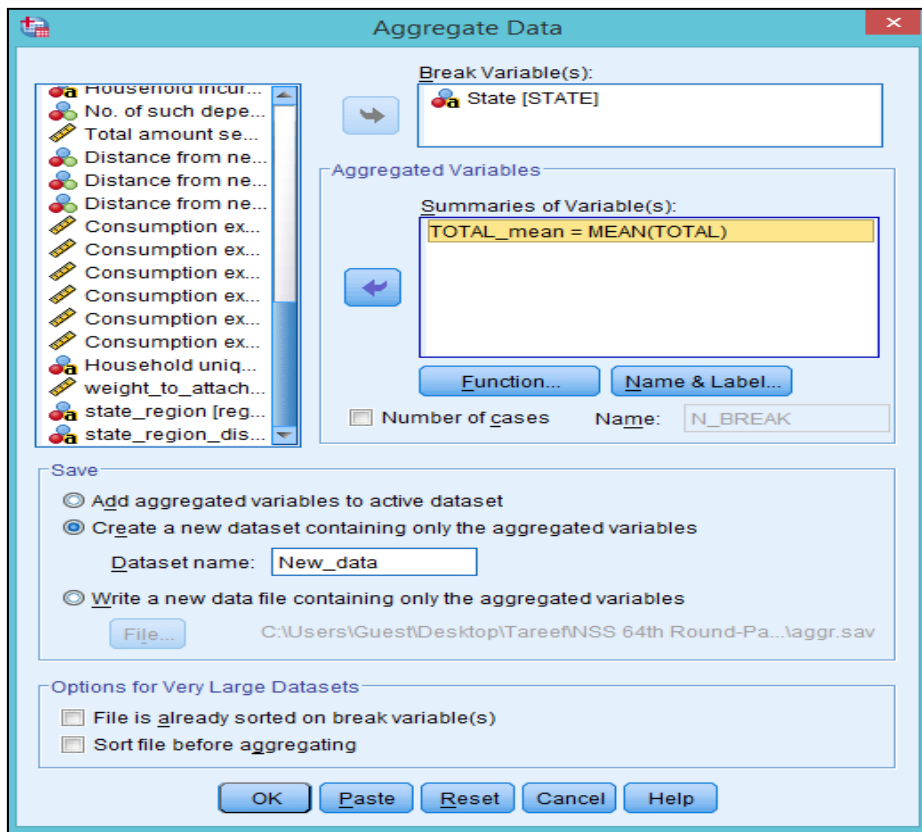


Fig. 28: Calculation of Aggregate Data

From the Option Box, select state variable in **Break Variable(s):** and desired variable to be aggregated in the **Summaries of Variable(s):**. In this example, "total consumption expenditure" is selected. For selecting desired statistics, click at **Function...** below the **Summaries of Variable(s):**. The new Option Box that appear is shown in Fig. 28.1.

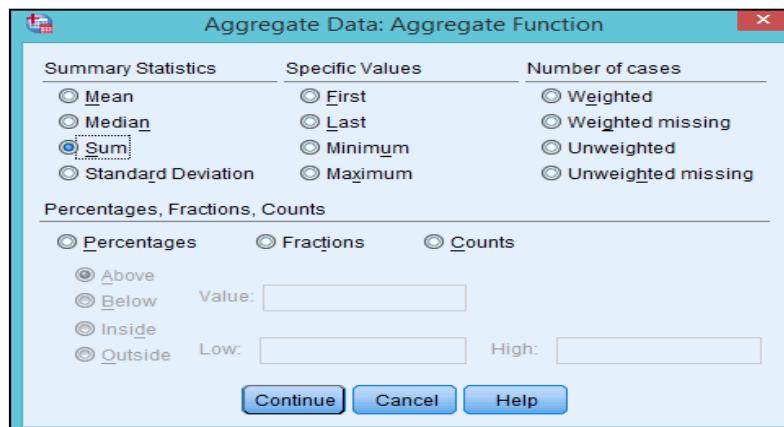


Fig. 28.1: Calculation of Aggregate Data: Aggregate Function



Select desired statistics from the Option Box 28.1. In this example, "Sum" is selected. Click at "Continue". Now go back to the "Aggregate Data" Option Box and from the "Save" menu given in the Middle of Option Box, select  Create a new dataset containing only the aggregated variables as shown in Fig. 28.2 and type the name of new data set in the "dataset name:" In this example, we have typed the "New data" in the box. Click at ; a new data set will be created with the assigned name as shown in Table 16.

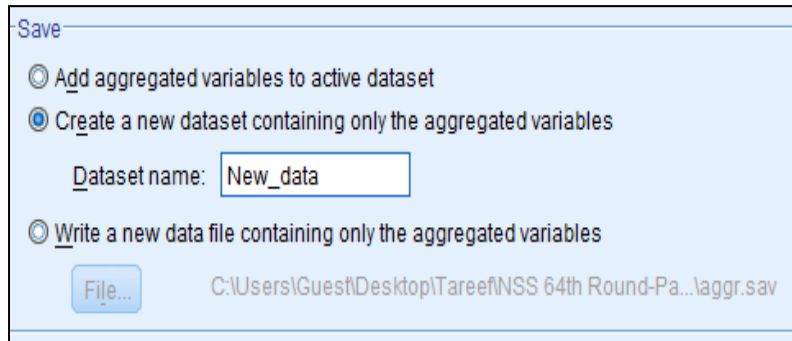


Fig. 28.2: Calculation of Aggregate Data: Aggregate Function (Step 2)

	STATE	TOTAL_sum	var	var	var	var	var	var	var
1	Jammu & Kashmir	4782.52							
2	Himachal Pradesh	3906.32							
3	Punjab	5146.32							
4	Chandigarh	7050.98							
5	Uttaranchal	3899.94							
6	Haryana	4843.15							
7	Delhi	6107.65							
8	Rajasthan	3886.26							
9	Uttar Pradesh	3477.04							
10	Bihar	2830.09							
11	Sikkim	3254.82							
12	Arunachal Pradesh	3674.43							
13	Nagaland	5710.88							

Table 16: Aggregated Data

## 16. Merging Data

“Merging Data” is used to merge two variables from two different files. In other words, “Merging Data” is used to add a variable in the dataset. Examples of use of “Merging Data” option is given below:

### 16.1 Adding New Variables

Data > Merge Files > Add Variables.

On selection of above-mentioned functions from the dropdown menus, "Add Variable Wizard "given Fig. 29 is displayed. Select a file to be merged and click at "Continue". However, before using this option both the files must be sorted by the identifier variable. In this example, it is HHID (Household ID).

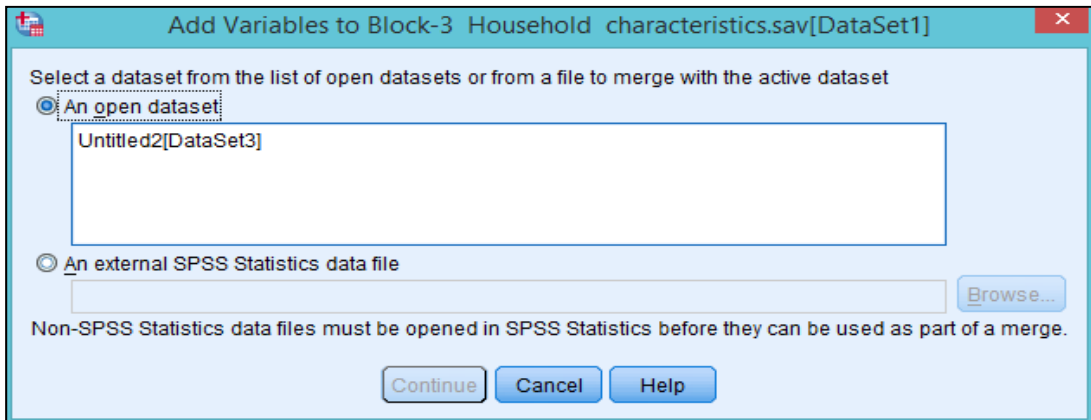


Fig. 29: Adding New Variables

The next stage "Add Variable Wizard" that appears is shown in Fig. 29.1. The common variable here is "HHID" which is to be selected in **Key Variables:** after selecting the  **Match cases on key variables**. Select the  **Active dataset is keyed table**, and click at "OK". As a result, a new variable **42** **Tution\_fee** will be added in the dataset as Var00001 as shown in Table 17.

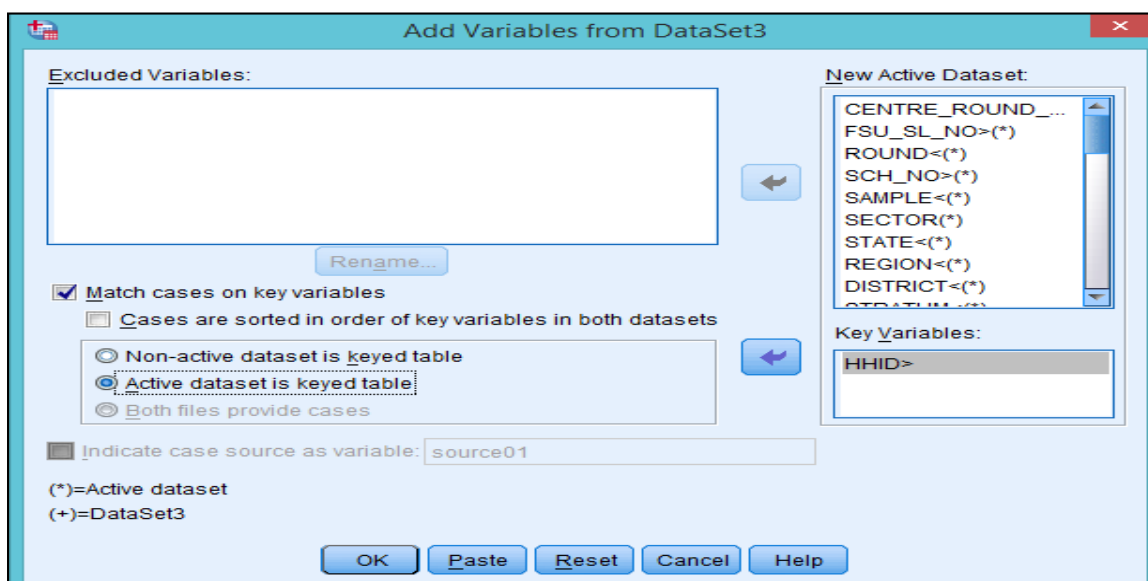


Fig. 29.1: Adding Variables to Dataset

\*Block-3 Household characteristics.sav [DataSet1] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help

6 : VAR00001 45.00 Visible: 46 of 46 Variables

	L	NTOTAL	VAR00001	var	var	var	var	var	var	var	var
1	4.090	5	4521.00								
2	6.173	2	1.00								
3	4.688	4	1.00								
4	8.298	3	485.00								
5	1.715	6	85.00								
6	0.815	4	45.00								
7	0.750	1	25.00								
8	5.818	5	78.00								
9	9.025	2	69.00								
10	0.045	8	36.00								
11	8.495	5	75.00								

Data View Variable View

IBM SPSS Statistics Processor is ready Unicode:ON Filter On Weight On

Table 17: Added Variables

### 16.1.1 Adding Variables When Numbers of Cases in Two Files are Uneven

While merging two datasets, it is necessary that the number of cases in two files must be the same. If the number of cases are not equal, there will be greater number of missing values and that would be difficult for any calculations. To handle it, the dataset having higher number of cases need to be restructured.

#### 16.1.1.1 Restructuring Data

In the example taken here, there are two files, one provides information on households where numbers of cases are lesser than the other file which provides information on individuals. Now to merge these two files, restructuring of data is required. Follow the steps given below for restructuring of files:

**Data > Restructure...**

The Restructuring Wizard (Step 1) that appears is shown in Fig. 30.1. Select the second option, i.e.

Restructure selected cases into variables for transposing the information of individuals from cases to variables, so that the number of cases in this file becomes equal to the number of cases in other file.

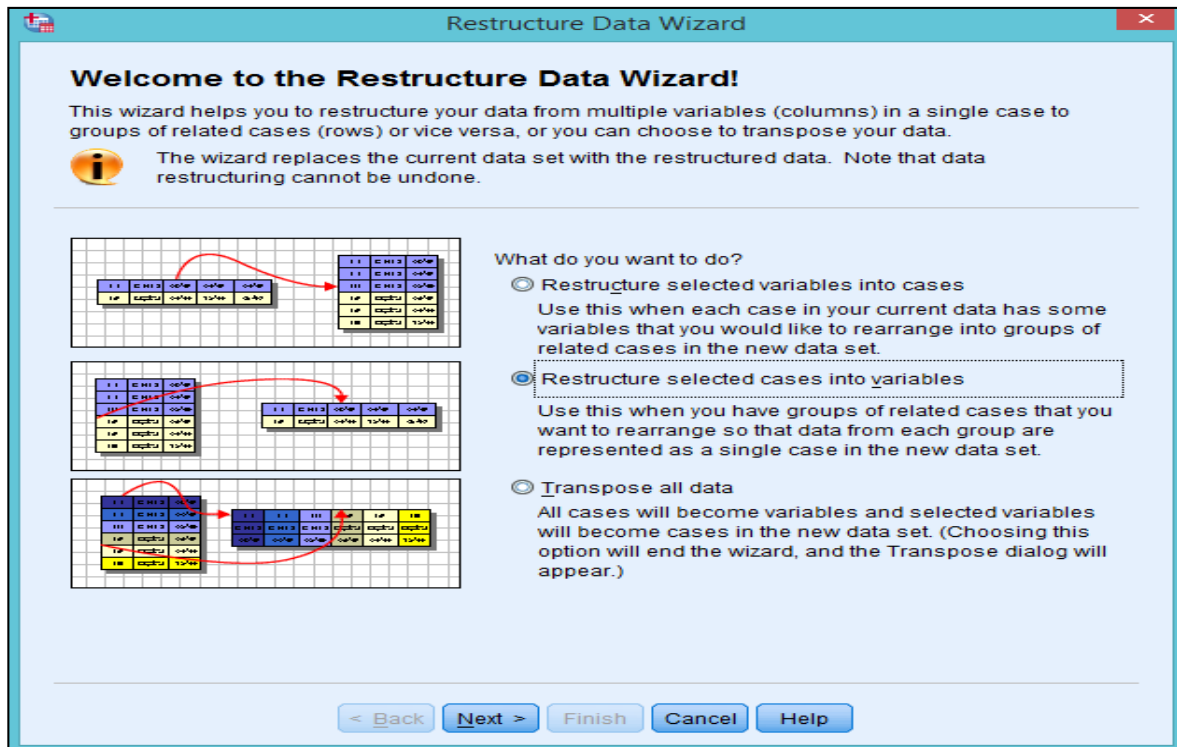


Fig. 30.1: Restructure Data Wizard (Step 1)

After selecting "restructure selected cases into variables", Click at "Next". The second step of the Wizard that appears is shown in Fig. 30.2.

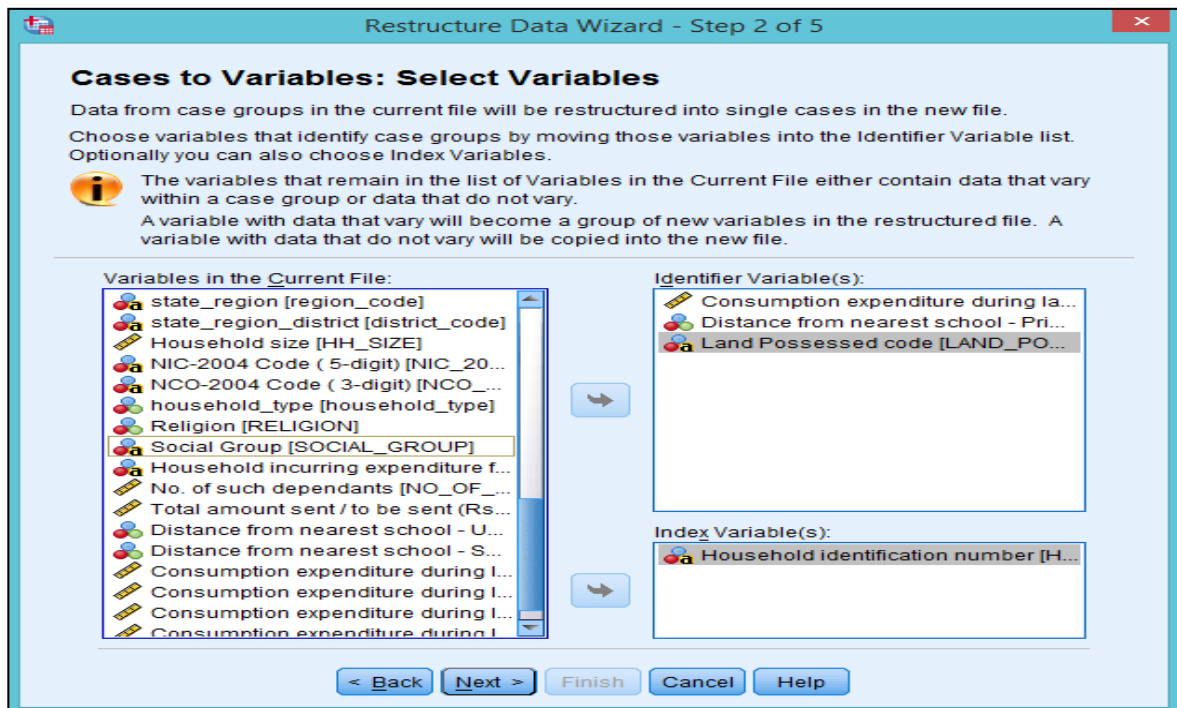


Fig. 30.2: Restructure Data Wizard (Step 2)

Now, select the "HHID" (i.e. the same values for the cases) to the "Index variables", and the variables which need to be restructured should be selected to the "Identifier variables" and Click "Next" as shown in Fig. 30.2.

Second step of Wizard that appears is shown in Fig. 30.3. Select the **Yes - data will be sorted by the Identifier and Index variables** and again click on "Next".

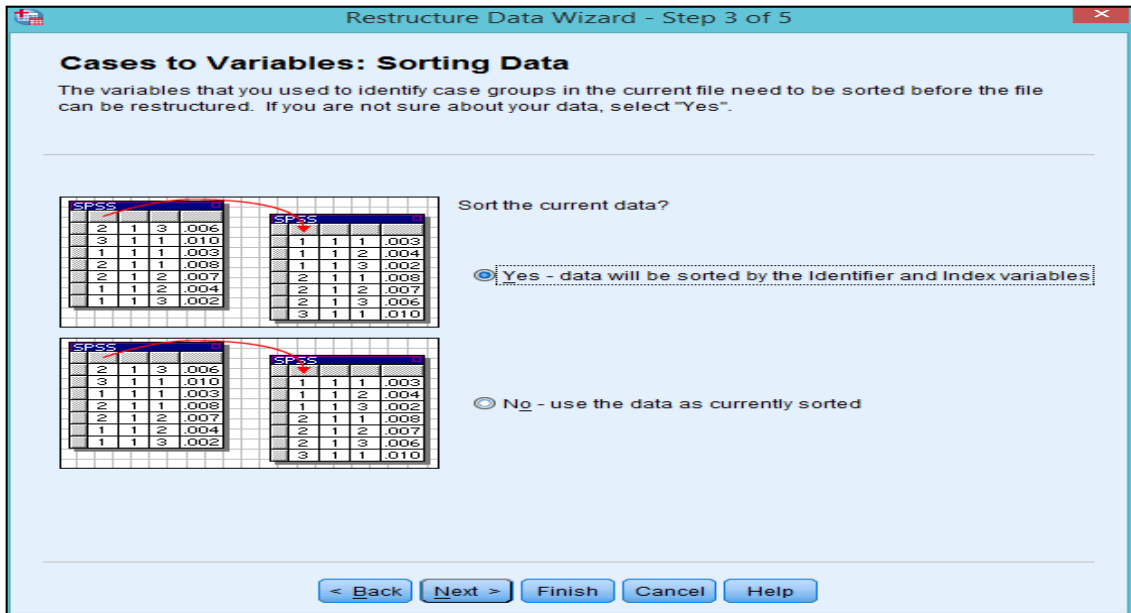


Fig. 30.3: Restructure Data Wizard (Step 3)

The fourth step wizard that appears is shown in Fig. 30.4. Click on "Next" button.

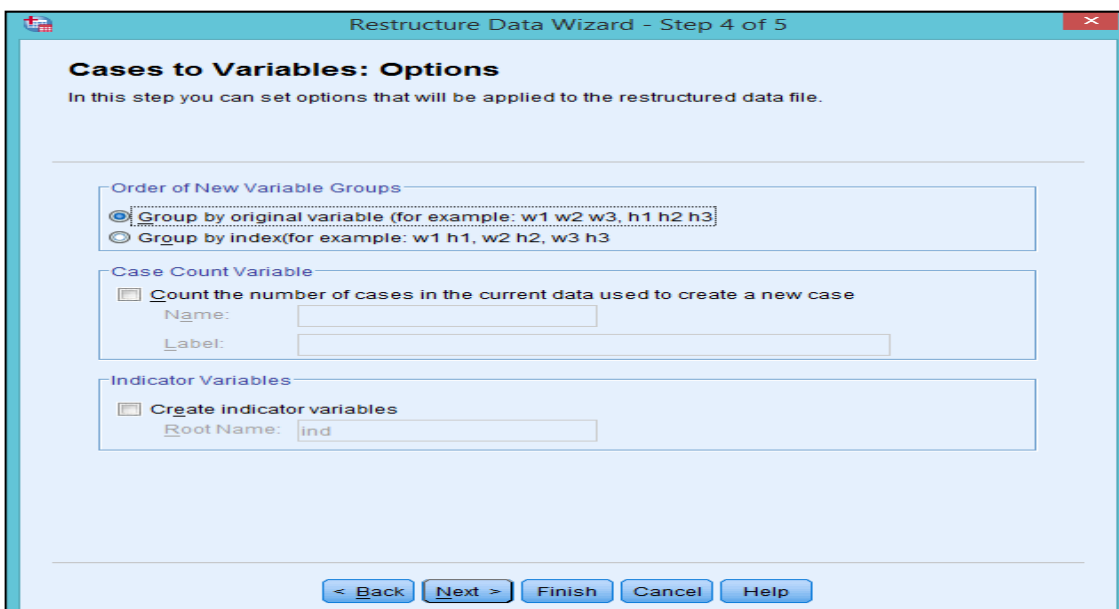


Fig. 30.4: Restructure Data Wizard (Step 4)

Click at "Next" to get the final and fifth step of Wizard as shown in Fig. 30.5. Click at "Finish" button in that window. The data will be restructured in terms of number of households. Now, two datasets can be merged.

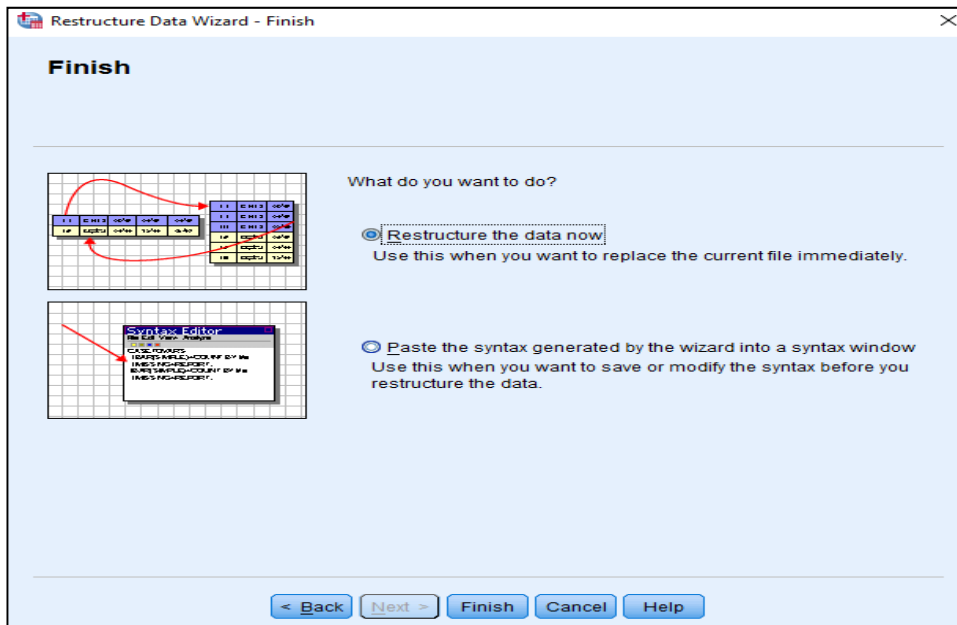


Fig. 30.5: Restructure Data Wizard (Final Step)

## 16.2 Adding New Cases

Similar to "adding variable", you can add new cases in the dataset using menu option from the dropdown menu given below:

**Data > Merge File > Add Cases.**

On selection of the options given above, the Option Box that appears is shown in Fig. 31. Select "An external SPSS Statistics data file" and click at "Browse" button.

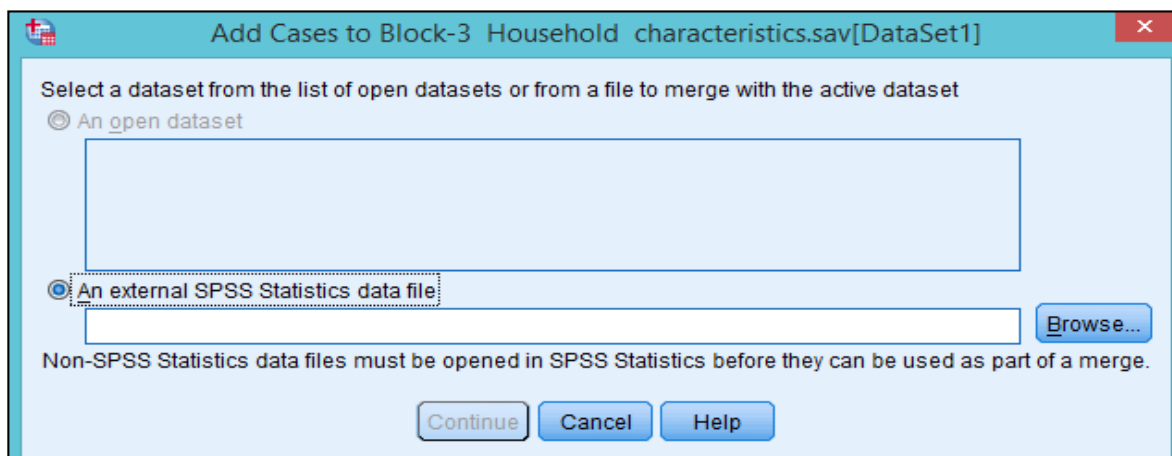
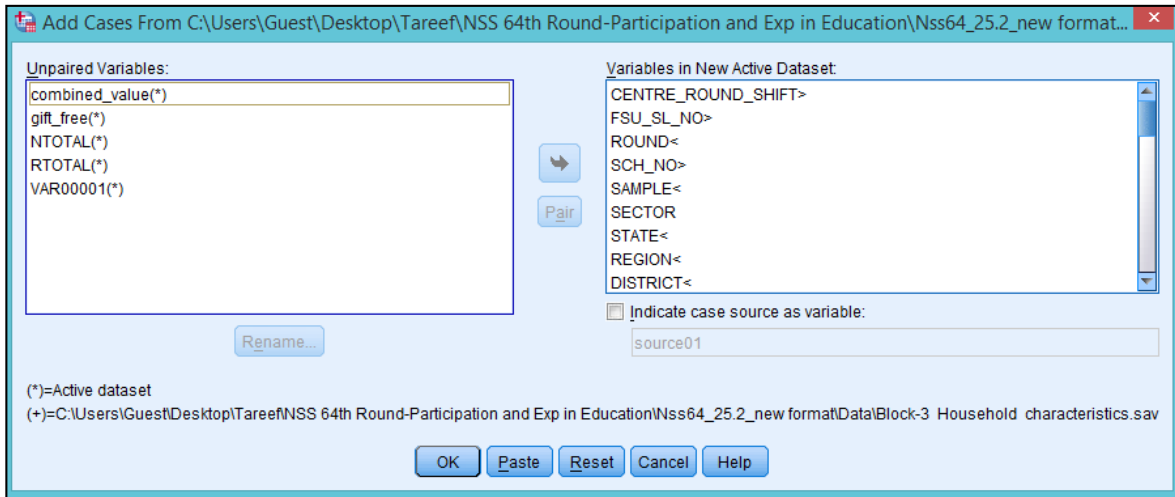


Fig. 31: Adding New Cases: Browse and Selecting File

Now, select the file from where the cases are to be added, and click "OK". Subsequently, the Option Box that appears is shown in Fig. 32.



**Fig. 32: Adding New Cases: Renaming or Deleting Unpaired Variables**

Fig. 32 shows the unpaired variables which are not available in one dataset, are shown in the unpaired variables box. You can handle these unpaired variables either through renaming or deleting them from the file, then click "OK". The new cases will be added to the file as shown in Table 18.

14596	16549	64	252	1		2	08		4	23	23	01	4	2	0823	1	2	04	02		8	26921	932		23			1	2	0	2
14597	16552	64	252	1		2	08		2	22	22	01	3	2	0820	1	1	01	02		11	74940	313		21			1	3	0	2
14598	16546	64	252	1		2	08		5	13	13	01	4	2	0820	1	1	02	02		12	36911	121		21			1	3	0	2
14599	15820	64	252	1		2	08		2	12	12	01	2	1	0820	1	1	04	02		5	60231	241		21			1	3	0	2
14600	15897	64	252	1		2	08		5	05	05	01	1	1	0820	1	2	04	02		7	45201	931		23			2	3	0	2
14601	15818	64	252	1		2	08		1	20	20	01	4	2	0811	1	2	02	02		3	01211	612		21			1	3	0	2
14602	16512	64	252	1		2	08		2	09	09	01	4	2	0823	2	2	02	02		2				29			1	3	0	2
14603	15878	64	252	1		2	08		4	31	31	01	1	1	0823	1	1	02	02		5	52393	241		21			1	9	0	2
14604	15831	64	252	1		2	08		2	12	33	02	1	2	0820	1	1	01	02		5	74999	734		21			2	9	0	2
14605	16561	64	252	1		2	08		4	32	32	01	3	2	0823	1	2	01	02		8	52201	241		21			1	9	0	2
14606	15028	64	252	1		2	08		3	25	25	01	1	1	0812	1	2	03	02		6	01111	611		21			1	3	0	2
14607	15006	64	252	1		2	08		3	26	26	02	1	1	0812	1	1	02	02		6	75112	214		22			1	9	0	2
14608	15816	64	252	1		2	08		5	14	14	01	4	2	0810	1	1	03	02		6	45201	712		21			1	2	0	2
14609	16556	64	252	1		2	08		2	09	09	01	3	1	0823	1	2	01	02		6	51498	241		21			1	1	0	2
14610	15877	64	252	1		2	08		2	22	22	01	1	2	0820	1	1	03	02		4	45201	931		23			1	2	0	2

**Table 18: Result Display of Adding New Cases: Renaming or Deleting Unpaired Variables**

## 17. Output Files

The "output" files are the files, where the final calculation appears. For example, calculating the frequencies of households from rural and urban sector, shown in Fig. 33. The tables in output files can also be edited, by double clicking it. The table will be activated. Now, select options from the dropdown menus to edit the table. In this example, "Table Looks" is used by selecting following option from dropdown menu:

**Format > Table Looks**

Select "classic" look of the table, as shown in Table 19. There are many other options that can use, e.g. sorting the tables, transposing the tables, cell formats, etc.

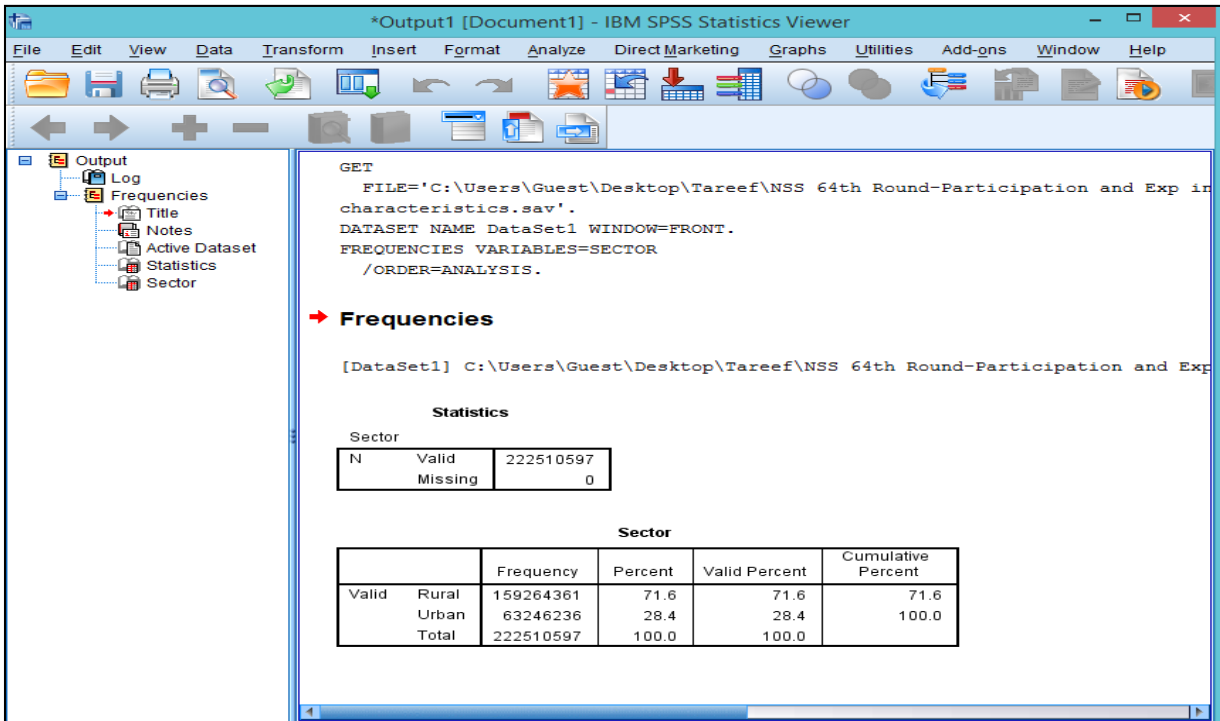


Fig. 33: Creation of Output File

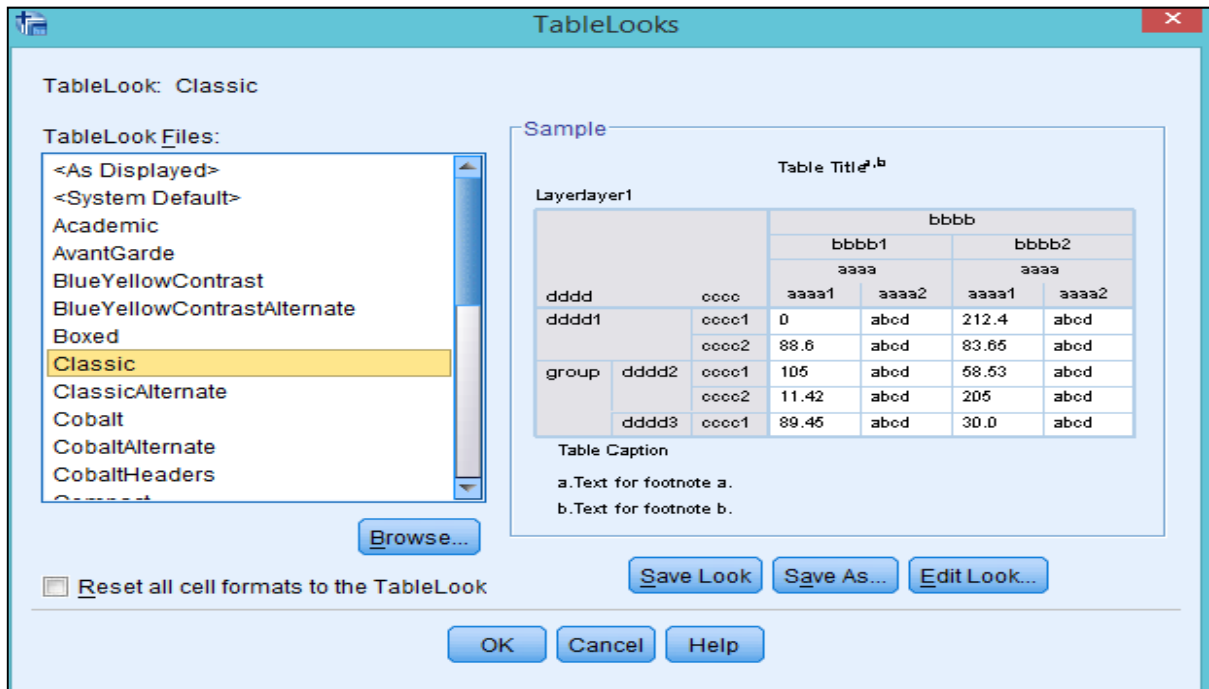


Table 19: TableLooks of Output File

Further, click at "cell properties", the following wizard will appear as shown in Fig. 34. You can change the font and background, format value, alignment and margins. The change of font and background is shown in Fig. 34.



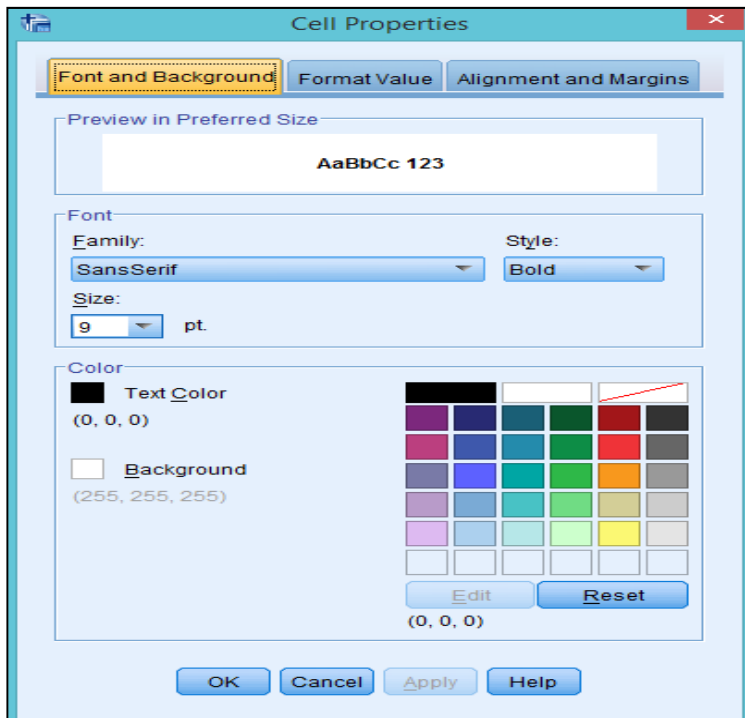


Fig. 34: Cell Properties- Font and Background

Click at "format value", wizard that appears is shown in Fig. 35 where the edition in terms of value format can be edited. Similarly, alignment and margins can also be defined.

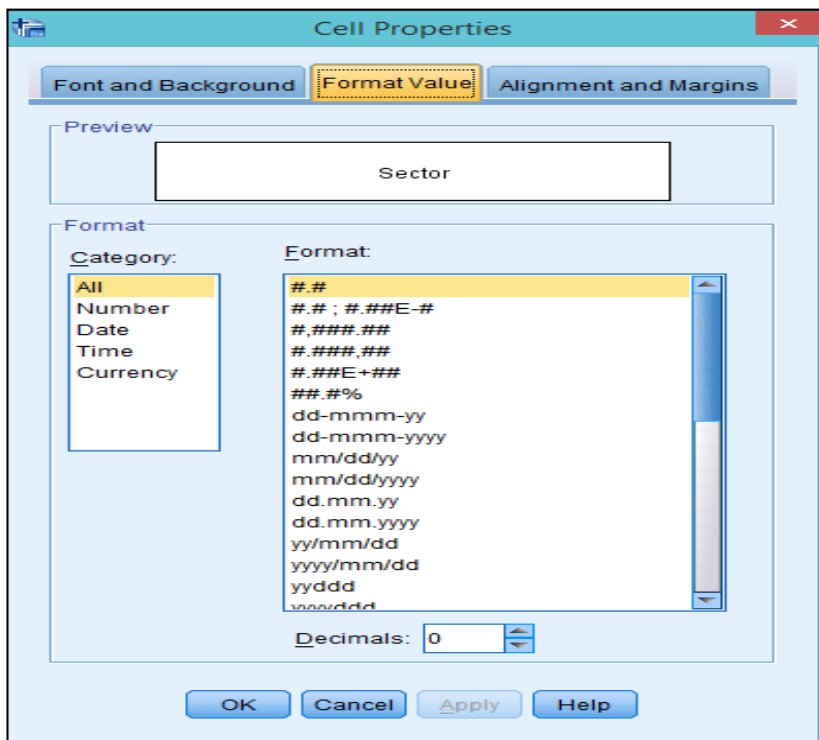


Fig. 35: Cell Properties- Format Value

## 17.1 Run Script

You may also use the "run script" option from the **Toolbars** to do the same calculation for other files. On clicking at "run script", the wizard that opens is shown in Fig. 36.

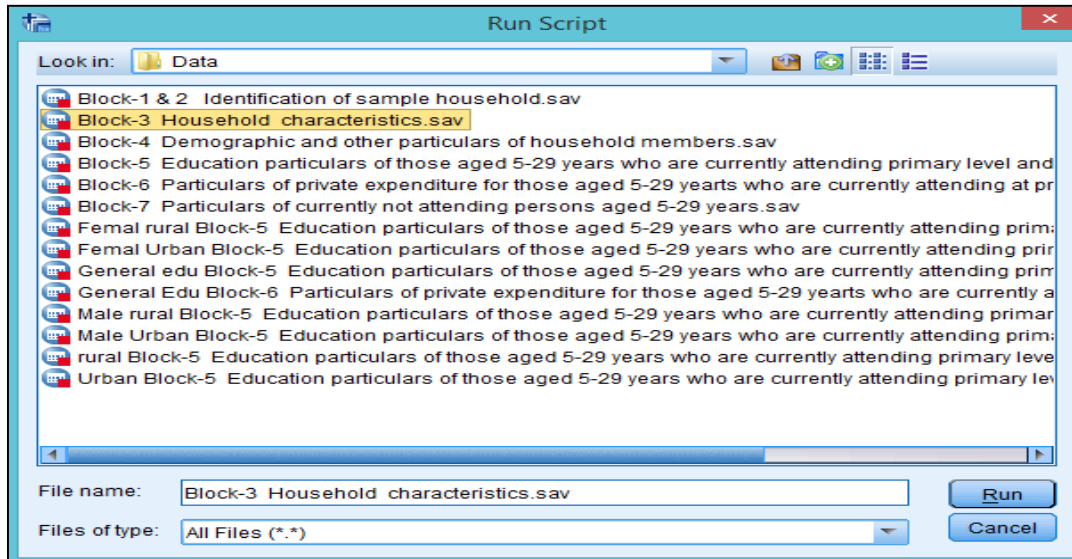


Fig. 36: Run Script

Select any file from your system and click on "Run". The same calculation will be done for different file.

## 18. Syntax

Syntax in SPSS is the third window, where you can run the earlier commands and the results will be produced on a single click. To open a new syntax file, go to:

**File > New > Syntax.**

The following Syntax Editor will appear as shown in Fig. 37.

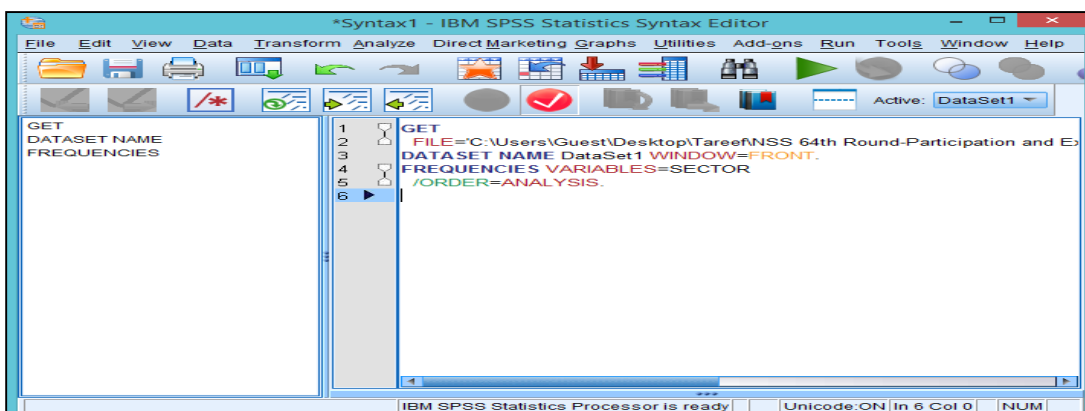


Fig. 37: Syntax Editor

Now, copy the commands from the output file, and paste it in the syntax file. The name of file can also be changed. Click at "Run" from the toolbars after selecting required command (s) The same calculation will be done for another file.

## 19. Graphs

In SPSS, one can draw a wide range of graphs, e.g. histogram, bar, line, area, scatter, plot, etc. To prepare graph, select the "Graphs" from the dropdown menu. The Option Box for "Chart Builder" that appears is as shown in Fig. 38. Select desired type of graph from the list which are given in the "Gallery". "Bar diagram" is selected in this example. The resultant bar diagram is shown in Fig. 38, where "Sector" has been taken on the X-axis and "Expenditure" on Y-axis.

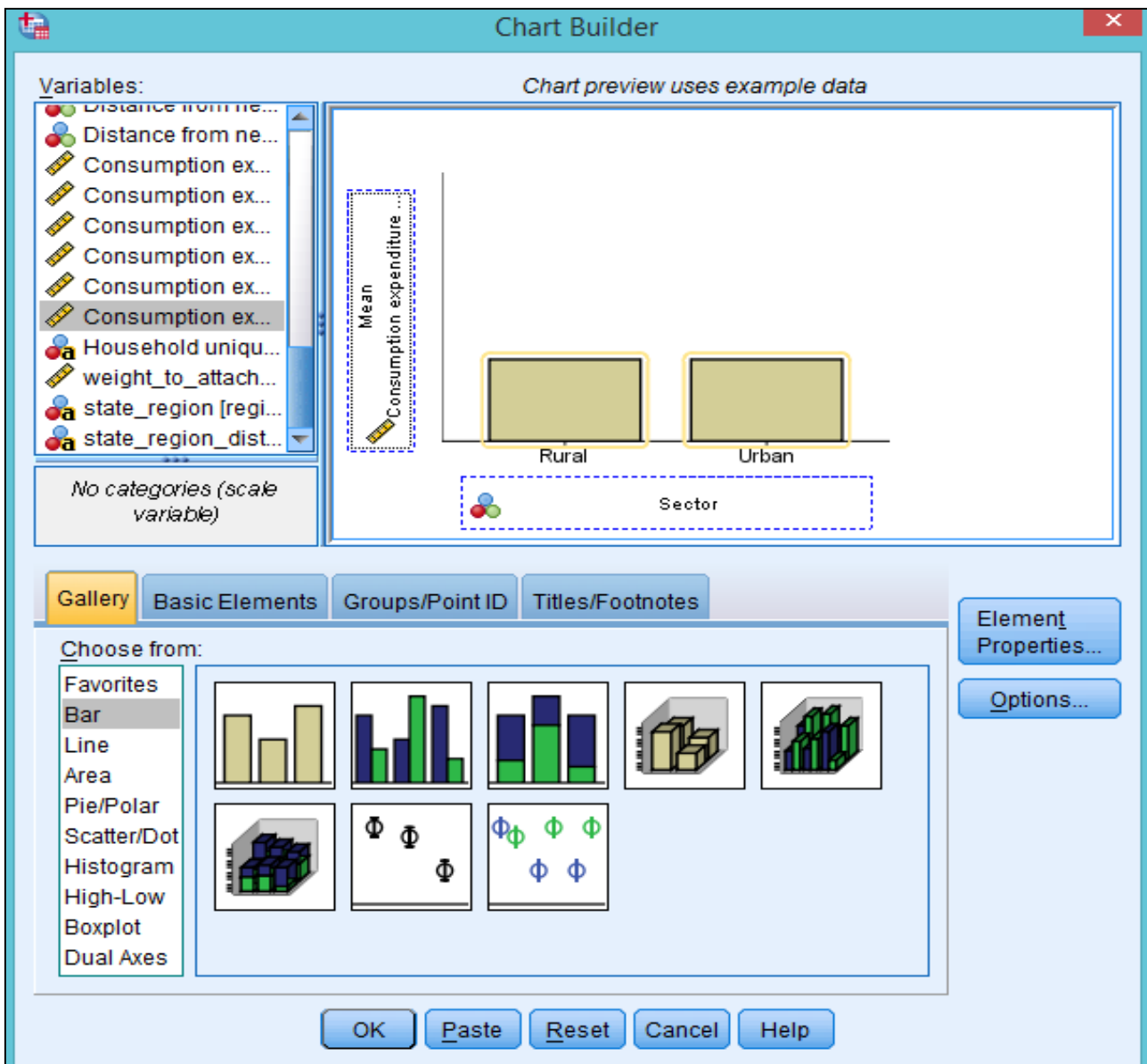


Fig. 38: Chart Builder Window and Creation of Graph

To plot average expenditure of households in rural and urban areas, select "Mean" from the "Statistics" menu in the "Element Properties" as shown in Fig. 39.

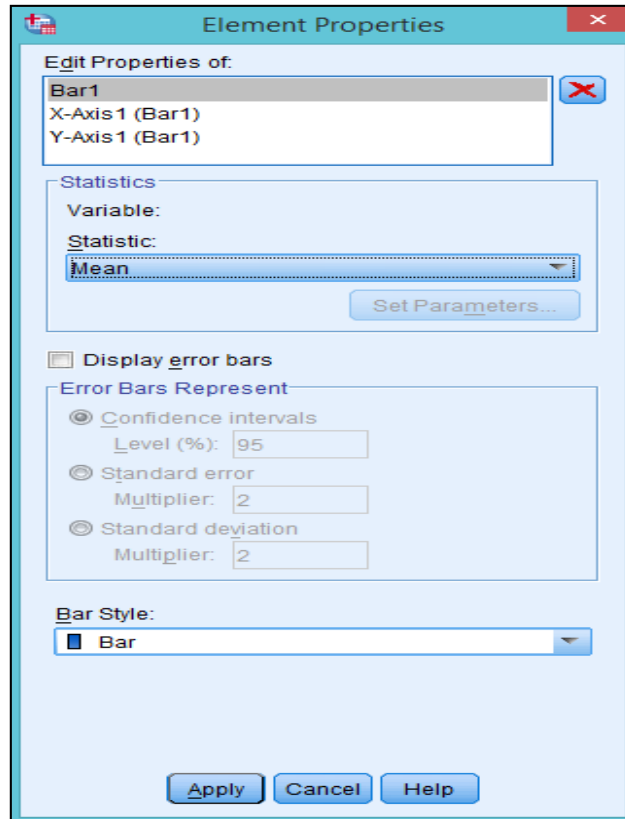


Fig. 39: Chart Builder- Element Properties

On taking the steps mentioned above, the graph produced as output is shown in Fig. 40.

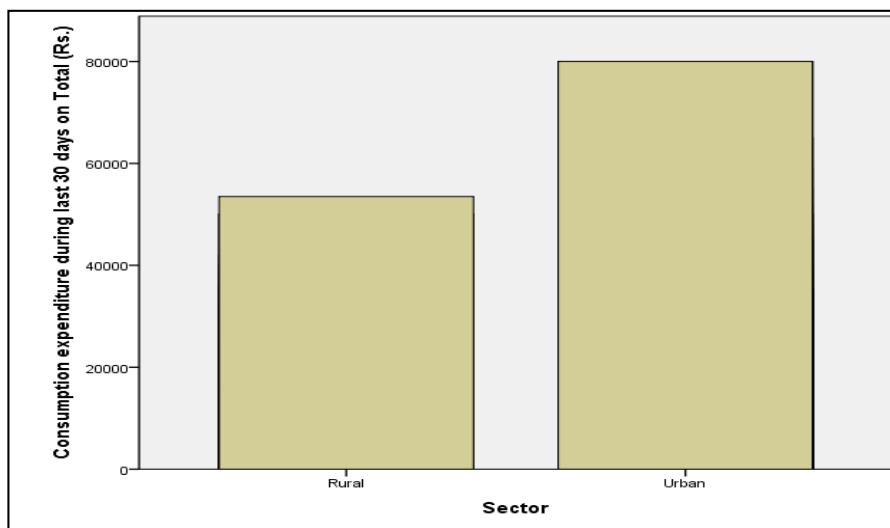


Fig. 40: Bar Chart

For further information please contact

**Dr. Upendra Choudhury**

Member-Secretary  
Indian Council of Social Science Research  
Aruna Asaf Ali Marg  
New Delhi - 110067  
E-mail: [ms@icssr.org](mailto:ms@icssr.org)

**Dr. Jagdish Arora**

Director  
Information and Library Network Centre  
An IUC of UGC  
Infocity, Gandhinagar -382007, Gujarat  
E-mail: [director@inflibnet.ac.in](mailto:director@inflibnet.ac.in)



**Indian Council of Social Science Research**

JNU Institutional Area, Aruna Asaf Ali Marg  
New Delhi - 110067 (INDIA)

For more information : <http://www.icssrdataservice.in/>.