

ICSSR Data Service

Indian Social Science Data Repository

SPSS: User Guide



Indian Council of Social Science Research

SPSS 22: User Guide

Contents

1.	Introd	uction	1									
2.	Openi	ng Data File	1									
3.	Impor	ting Text Data	4									
4.	Value	Labels	7									
5.	Report	ts	9									
6.	Descri	Descriptive Statistics										
	6.1	Frequency Calculation										
	6.2	Comparing Two Variables										
7.	Sort Ca	ases	17									
8.	Select	Cases	18									
9.	Split D	ata	19									
10.	Custor	n Tables	20									
	10.1	Two Way Table										
	10.2	Three Way Table										
11.	Compu	ute Variable	23									
12.	Recod	ing of Data	25									
13.	Rank C	Cases	27									
14.	Weigh	t Cases	29									
15.	Aggre	gate Data	29									
16.	Mergi	ng Data	31									
	16.1	Adding New Variables										
		16.1.1 Adding Variables when Numbers of Cases in Two Files are Uneven										
		16.1.1.1 Restructuring Data										
	16.2	Adding New Cases										
17.	Outpu	t Files	37									
	17.1	Run Script										
18.	Syntax	(40									
19.	Graph	S	41									

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".

ta *Bloo	:k-3 Ho	usehold	cha	racter	isti	cs.sav [Data:	Set2] -	IBM	SPSS	5 Sta	tist	tics	Data E	dit	tor	-		×	:
<u>F</u> ile <u>E</u> dit	<u>V</u> iew	<u>D</u> ata	<u>T</u> ran:	sform	Ar	nalyze Direct	<u>M</u> ark	etin	g <u>(</u>	<u>G</u> raph	IS I	<u>J</u> tili	ties	Add-	ons	1	<u>W</u> ind	ow	<u>H</u> elp	
)		C			ŀ			μ				*5	V			4	2	
19 : SUB_S	AMPLE	2												١	Visi	ble:	41 of	41 V	ariabl	es
	CENT RE_R	FSU_SL	RO UN D	SCH_ NO	S A	SECTOR	STA TE	R E	DIS TRI CT	STR ATU M	SU B_S	S U	S U	FOD_S UB_RE GION	H G_	S S	SA MP	LEV EL	Hŀ	
1	001	64039	64	252	1	Rural	Tr	1	01	01	16	2	1	0000	1	2	02	02		
2	001	64039	64	252	1	Rural	Tr	1	01	01	16	2	1	0000	1	2	01	02		
3	001	61021	64	252	1	Rural	В	2	33	33	02	2	1	1022	2	1	02	02		
4	001	62796	64	252	1	Rural	A	4	80	08	02	2	2	2820	2	1	02	02		÷
	4																			
Data Viev	/ Variab	le 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.

<u>F</u> ile	<u>E</u> dit	<u>V</u> iew <u>D</u> ata	Transform A	nalyze Dire	ect <u>M</u> arketing	<u>G</u> raphs <u>U</u> tilitie	s Add- <u>o</u> ns	<u>W</u> indow <u>H</u> elp
				¥ 🎬				
		Name	Туре	Width	Decimals	Label	Values	Missing
1		CENTRE_R	String	11	0	Centre code, R	None	None 🖆
2	2	FSU_SL_NO	String	15	0	FSU Serial No.	None	None !
3	}	ROUND	String	6	0	Round	None	None :
4	Ļ	SCH_NO	String	9	0	Schedule Number	None	None :
5	5	SAMPLE	String	3	0	Sample	None	None
6	;	SECTOR	Numeric	1	0	Sector	{1, Rural}	None I
7	'	STATE	String	6	0	State	{01, Jammu	None :
	1	1						
Data	View	Variable View						

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".

Value Labels	×
Value Labels Val <u>u</u> e: Label:	Spelling
Add Change Remove	
OK Cancel Help	

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.

Missing Val	ues ×
© <u>N</u> o missing values (© <u>D</u> iscrete missing values 999	
© Range plus one optional di Low: ∐i	iscrete missing value
Discrete value:	
OK Cancel	Help

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. 6.1will appear to facilitate import of text data.

Те	ext Import Wizard - Step 1 of 6
628 840 1 81 28.5 630 2400 0 78 40 33 630 870 0 93 31 17 635 1740 9 83 41.91	Welcome to the text import wizard! This wizard will help you read data from your text file and specify information about the variables.
eart ear2 ear3 e	Does your text file match a predefined format? • Yes • No
Text file: C:\Users\Guest\Des	ktop\AH1C25.TXT
1 0006536264252110	5101010111051011010100000 121 050907051007150 📥
2 0016536264252110	51010101110510110102000000401111611419042
3 0026536264252110	510101011105101101030000112 34208
<u>4</u> 0036536264252110	510101011105101101030000251 121077
6 0056536264252110	510101011105101101030000452 91067
7 0066536264252110	510101011105101101040000212 5111010807061 0121
8 0076536264252110	510101011105101101040000314 6111010808071 0121
4	
< <u>B</u> ack	Next > Finish Cancel Help

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 predefined, leave it unfilled and click at "Next" button. The second Text Import Wizard will appear as shown in Fig. 6.2.

Text Import Wizard - Step 2 of 6
How are your variables arranged? Delimited - Variables are delimited by a specific character (i.e., comma, tab). Fixed width - Variables are aligned in fixed width columns.
Are variable names included at the top of your file? ^O Yes [●] No
Image: City Sers (Guest) Desktop (AH1C25.TXT 0
6 Back Next > Finish Cancel Help

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.

ta Text Import Wizar	d - Fixed W	/idth Step 3 of	6	×
The first case of data begins on which line	number?	1 🚔		
How many lines represent a case?	*			
How many cases do you want to import?-				
All of the cases				
© <u>T</u> he first 1000 cases	S.			
\bigcirc A percentage of the cases: 10	\$			
Data preview				
	110	100	100	140
80		120	130	
1		2	4	236922 📥
2 100	300 1	.600 2	4	236922
		2	4	236922
5		2	4	236922
6		2	4	236922
7 162 41 1		2	4	236922
8 162 41 1		2	4	236922
9 161111 1		2	4	236922
10 200 300		2	4	236922
11 300 300		2	4	236922
1				
< <u>B</u> ack (<u>N</u> ext >)	Finish	Cancel Help		

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.

Text Import Wizard - Fixed Width Step 4 of 6	<
Specify where each variable begins. The first column is column 0.	
To INSERT a variable break line, click at the desired position in the ruler or data area. Alternatively, move to the position using the arrow keys or by typing the column number; then press the Insert Break button.	
To MOVE a variable break line, drag it to the new position.	
To DELETE a variable break line, select it or type its position. Then press the Delete key or the Delete Break button.	
<u>Buler:</u> ⊅	_
1 00065362642520105001010101005100101000000 121 050907051007150 2 0016536264252010500101010100510010000004011111601409042 3 0026536264252010500101010100510010010300000112 34208 4 0036536264252010500101010101000510010010300000251 121 077	
<u>C</u> olumn Number: 53 In <u>s</u> ert Break <u>D</u> elete Break	
Current Variable Width: 1	
< Back Next > Finish Cancel Help	

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 3rd column, 8th 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.

text Import	Wizard - Ste	p 5 of 6				×
Specification	s for variable	e(s) selected	in the data pr	eview		
	name:	2 LUI				
CENTRE		SHIF				
Data form	nat					
Numeric	:]				
□Data preview						
	[[
CENTRE	. V2	V3	V4	V5	V6	V7
000	65362	64	252	1	1	05
001	65262	64	252	1	1	05
002	65262	64	252	1	1	05
004	65362	64	252	1	1	05
4						
	<	Back Next	> Finish	Cancel	Help	
		Fig. 6.5: Te	xt Import V	Vizard (Ste	p 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.

🔒 Text	Impor	t Wiza	rd - S	tep 6	of 6						Х
						You have	success	fully defined	the format of yo	our text file.	
						-Would ye	ou like to	save this file	e format for futu	re use?	
	var1	var2	yar3			© <u>Y</u> es				Save As	
1	628	8-40	1			No					1
2	630	2400	0			_					
3	632	10200	0			-Would ye	ou like to	paste the sy	ntax?		
4	633	870	0							acha data la a	
	0000000000		and the second	•		© 1 <u>6</u> 3				ache data loca	апу
						<u>N</u> 0					
Data p	reviev	v				i iess uie	,	, .	piete tre text in	, inport wizard.	
CEN	TRE_	V2			V3		V4	V5	V6	V7	
000		653	362		64		252	1	1	05 🔺	
001		653	362		64		252	1	1	05	
002		653	362		64		252	1	1	05	
003		653	362		64		252	1	1	05	
004		053	362		64 64		252	1	1	05	
005		653	262		64 64		252	1	1	05	
4										4	
				: <u>B</u> ac	k (Next >	Finish	Cancel	Help		

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 > 14 Value labels

<u>F</u> ile <u>E</u> dit <u>V</u> ie	w <u>D</u> at	a <u>T</u> ransfo	orn <u>A</u> r	nalyze I	Dire	ct <u>M</u> arketi⊨ <u>G</u> raj	ohs <u>U</u>	tiliti	ies A	dd- <u>o</u> n	s <u>W</u> i	ndo	wĿ	lelp
) 🛄			-	× 🎬	ł			ų	3	H		*
1: SECTOR 1 Visible: 41 of 41 Variables														
	CENT RE_R	FSU_SL _NO	RO UN D	SCH_ NO	S A	SECTOR	STA TE	R E	DIS TRI CT	STR ATU M	SU B_S	S U	S U	F
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 1	61150	64	252	1	1	09	4	41	41	01	3	2	0
Data View	Variab	le View												
		BM SPSS	Stati	stics P	roce	essor is ready	l	Jnio	code:	ON	W	eigh	nt O	n

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

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.

<u>F</u> ile <u>E</u> dit <u>V</u> iew <u>D</u> ata <u>T</u> ransforn <u>A</u> nalyze Direct <u>M</u> arketi⊨ <u>G</u> raphs <u>U</u> tilities Add- <u>o</u> ns <u>W</u> indow <u>H</u> elp														
) 🛄			1	🖬 🔡	Ļ			4		9-6		*
1:SECTOR		1							Visit	ole: 41	1 of 4	1 Va	aria	bles
	CENT RE_R	FSU_SL _NO	RO UN D	SCH_ NO	S A	SECTOR	STA TE	R E	DIS TRI CT	STR ATU M	SU B_S	S U	S U	F
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	В	2	33	33	02	2	1	1
4	001	62796	64	252	1	Rural	A	4	80	80	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 -
Data View Variable View														
	IBM SPSS Statistics Processor is ready Unicode:ON Weight On													

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.

	Codebook		
Variables Output Statistics			
Variables:	<u>C</u> odebook Variables:		
 Sample [SAMPLE] Sector [SECTOR] State [STATE] Region [REGION] District [DISTRICT] Sub-Stratum no [SUB_STRATU Sub-Stratum no [SUB_STRATU Sub-sample [SUB_SAMPLE] FOD-Sub-Region [FOD_SUB_R Hg/Sb Number [HG_SB_NO] Second Stage Stratum Number Sample Household Number [S Level [LEVEL] Person Serial Number [PERSO Relation to Head [RELATION_T Sex [SEX] Marital Status [MARITAL_STATU Status of current educational att Educational enrolment status [Consumption expenditure during I		
OK Paste Reset Cancel Help			

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.

SPSS 22: User Guide

		Value
Standard Attributes	Position	49
	Label	Consumption expenditure during last 30 days on Free collections(Rs.)
	Туре	Numeric
	Format	F4
	Measurement	Scale
	Role	Input
Ν	Valid	240077
	Missing	205883
Central Tendency and Dispersion	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

Summarize Cases					
Centre code, Round FSU Serial Number [Round [ROUND] Schedule Number [S Sample [SAMPLE] State [STATE] Region [REGION] District [DISTRICT] Stratum [STRATUM] Sub-Stratum no [SU Sub-round [SUB RO Display cases Limit cases to first Show only valid cases Show case numbers OK Paste	Variables: Consumption expen Options Grouping Variable(s): Sector [SECTOR] Reset Cancel Help				

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	included		Excluded		
	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).

Frequencies ×					
Variable(s): Centre code, Rou Charts Round [ROUND] Schedule Numbe Schedule Numbe Schedule Numbe Schedule Numbe Statistics Charts Format Style Bootstrap Monostrap Monostrap					

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 are 63,318 rural and 37,263 urban households selected for the survey. Result in 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).

😘 Frequencies: Sta	Frequencies: Statistics				
Percentile Values Quartiles Cut points for: 10 equal groups Percentile(s): Add Change Remove	Central Tendency Mean Median Mode Sum				
Dispersion Std. deviation Minimum Variance Maximum Range S.E. mean Continue Cancel	■ Values are group midpoints Distribution ▼ Skewness ■ Kurtosis				

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.

Ν	Valid	100578
	Missing	3
Mean		3798.02
Std. Deviatio	on	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.

D	Crosstabs	×
Centre code, Round a FSU Serial No. [FSU Round [ROUND] Schedule Number [S Sample [SAMPLE] Sector [SECTOR] State [STATE] Region [REGION] District [DISTRICT] Stratum [STRATUM] Sub-Stratum No. [SUB Sub-Stratum No. [SUB Sub-Sample [SUB_SA FOD-Sub-Region [FO Display clustered bar charts Suppress tables	Row(s): Column(s): Column(s): Calumn(s): Column(s): Column(s): Reset Cancel Help	Exact Statistics C <u>e</u> lls Eormat Style Bootstr <u>a</u> p

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.

Crosstabs: Cell Display					
Counts Counts Expected Hide small counts Less than 5	Z-test				
Percentages Row Column Total	Residuals Unstandardized Standardized Adjusted standardized				
Noninteger Weights					
Continue Cancel Help					

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	
			3.01 - 4.00	greater than	
				8.00	Total
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.

t	Crosstabs	×
Centre code, Round a FSU Serial No. [FSU Round [ROUND] Schedule Number [S Sample [SAMPLE] State [STATE] Region [REGION] District [DISTRICT] Stratum [STRATUM] Sub-Stratum No. [SUB Sub-Stratum No. [SUB Sub-St	Row(s): Social Group [SOCIAL Column(s): Column(Exact Statistics Cells Eormat Style Bootstr <u>a</u> p

Fig. 15: Crosstabs Calculation Window

			Land Posse			
				3.01 - 4.00	greater than	
Sector					8.00	Total
Rural	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
	Total		56777	156094598	3112987	159264362
Urban	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
	Total		43742	62950492	252003	63246237
Total	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
	Total		100519	219045090	3364990	222510599

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

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 <u>Statistics...</u> 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.

Crosstabs: St	tatistics ×							
C <u>h</u> i-square	Correlations							
Nominal	Ordinal							
Contingency coefficient	🔲 <u>G</u> amma							
Phi and Cramer's V	Somers' d							
🔲 Lambda	🔲 Kendall's tau- <u>b</u>							
Uncertainty coefficient	🔲 Kendall's tau- <u>c</u>							
Nominal by Interval	🗐 <u>K</u> appa							
🕅 <u>E</u> ta	🔲 R <u>i</u> sk							
	McNemar							
Cochr <u>a</u> n's and Mantel-Ha	enszel statistics							
Test common odds ratio	equals: 1							
Continue Cancel Help								

Fig. 16: Crosstabs Statistics Calculation

7. Sort Cases

The "**Sort Cases**" data in SPPS is used to sort the data in ascending or descending order. This task can be perform 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.

ta s	ort Cases ×
Centre code, Rou FSU Serial No. [F Round [ROUND] Schedule Numbe Sample [SAMPLE] Sector [SECTOR] Region [REGION] District [DISTRICT]	Sort by:
Save Sorted Data Save file with sorted d File Create an index OK Paste	Reset Cancel Help

Fig. 17: Sort Cases

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

<u>F</u> ile	<u>E</u> dit	View	Data								Date	iset	1] - 1	DIV	1 31	255 Sta	tist	ICS.	Data		01		_	
			Data	Tra	nsform		Analyze Dire	ect <u>I</u>	Marketing	G	aphs	U	tilitie	в	Ac	dd- <u>o</u> ns	V	/ind	ow	<u>H</u> elp				
		e)			2		ŀ		μ	A	Å	×	5		4		4			A 14		A	B
1:ST/	1: STATE 01 Visible: 45 of 45 Variables											les												
		CENT RE_R	FSU_SL _NO	RO UN D	SCH_ NO	S A	SECTOR		STATE	R E	DIS TRI CT	STR ATU M	SU B_S	S U	S U	FOD_S UB_RE GION	H G_	S S	SA MP	LEV EL	HH_SIZE	NIC_200 4_CODE	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	
-	В	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	
1	0	001	65547	64	252	1	1	01		2	09	09	01	3	2	0111	1	1	02	02	11	45203	931	
1	1	001	65503	64	252	1	1	01		3	06	06	03	1	2	0121	2	2	02	02	7	01111	611	
Data	Image: Constraint of the second se																							

 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.

t a	Select Cases	×						
 Household size [H household_type_va Religion [RELIGION] No. of such depend Total amount sent / Distance from near Distance from near Distance from near Consumption expe weight_to_attach_w 	Select All cases If condition is satisfied If Random sample of cases Sample Based on time or case range Range If sector [SECTOR]							
Current Status: Do not filter cases								

Fig. 18: Selecting and Filtering Cases

			3	Blo	ck-3 H	Hou	usehold	char
<u>F</u> ile	<u>E</u> dit	⊻iew	<u>D</u> ata	Tra	nsform		<u>A</u> nalyze	Dire
		CENT RE_R	FSU_SL _NO	RO	SCH_ NO	S A	SECT	OR
1		001	64039	64	252	1		Rural
2	!	001	64039	64	252	1	1	Rural
3	:	001	61021	64	252	1	1	Rural
- 4		001	62796	64	252	1		Rural
5		001	62059	64	252	1	1	Rural
6	;	001	69148	64	252	1	1	Rural
7	•	001	63230	64	252	1		Rural
8	•	001	64159	64	252	1	1	Rural
9		001	61635	64	252	1		Rural
10	D	001	61150	64	252	1		Rural
1.	1	001	63155	64	252	1	I	Rural
Data	View	Variab	le View					_
Jata								

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 <u>© Compare groups</u> and take "Sector" variable in the Selection Box. Click at "OK". Select the option <u>Split by SECTOR</u> that appears at the bottom right-end of the Option Box.

ta	Split File ×
Centre code, Rou FSU Serial No. [F Round [ROUND] Schedule Numbe Sample [SAMPLE] State [STATE] Region [REGION] District [DISTRICT] Stratum [STRATU Sub-Stratum No	 Analyze all cases, do not create groups Compare groups Organize output by groups Groups Based on: Sector [SECTOR] Sort the file by grouping variables File is already sorted groups is off.

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.

							Std.
Sector		Ν	Min.	Max.	Sum	Mean	Deviation
Rural	Consumption expenditure during last 30 days on Total (Rs.)	159260258	100	53500	468634075955	2942.57	1865.105
	Valid N (list wise)	159260258					
Urban	Consumption expenditure during last 30 days on Total (Rs.)	63246236	1	80000	333380137424	5271.15	3916.585
	Valid N (list wise)	63246236					
L	Table 9: 5	Split File ba	sed or	ו Secto	r		

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.

ta		Custom ⁻	Tables				×		
Table Titles Test Statistics	Options								
<u>V</u> ariables:			Ē	Norm <u>a</u> l	Co <u>m</u> pact	Lay	ers		
Social Croup IS	Columns								
Land Possess									
💑 Household inc		Hinduism	Islam	Christianity	Sikhism	gion Jainism	в		
No. of such dep		Sum	Sum	Sum	Sum	Sum			
Piotal amount s Distance from	Consumptio	nnnn	nnnn	nnnn	nnnn	nnnn			
Categories: No categories (scale variable)									
Deline	-Summary S	taustics							
№% Summary Statistics	Pos <u>i</u> tion:	Columns	-	🔲 <u>H</u> ide	Cat <u>e</u> gory Posi	tion:	_		
Seategories and Totals	So <u>u</u> rce:	Source: Row Variables							
	ОК	Paste Res	et Cancel	Help					

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 shown in Fig. 21. Click at "**OK**".

ţ			Summary	Statistics:			X
	Selected Variable: (Statistics:	Consumption exper	nditure during last 30 <u>D</u> isplay:	days on Total (Rs.)			
	Percentile 99		Statistics	Label	Format	Decima	
	Range		Sum	Sum	Auto		7
	Std Error of Mean						+
	Std. Deviation						
	Total N	-					
	(Apply to <u>S</u> election	<u>A</u> pply to All	Close	Help		

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	Zoroastriani sm	No Respons e
	Sum	Sum	Sum	Sum	Sum	Sum	Sum	Sum
Consumpti on	655313872 428	923456833 31	206803310 76	17627589 33	236932330 82	37880321 18	426028250 2	1700199 08

 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.

1	Custom Tables	×
Table Titles Test Statistics	Options	
<u>V</u> ariables:	III Normal 🗄 Compact	Layers
Centre code, R	C <u>o</u> lumns	
Round [ROUND]	Region	
a Sample [SAMP	Mean Mean	
Sector [SECTOR]	Sector Urban Consumption nnnn nnnn	
a Region [REGIO District [DISTRI		
Stratum [STRA Since and stratum N		
Sub-Round [SU Sub-sample [S		
FOD-Sub-Regi		
Sampla Llavas		
Define	Summary Statistics	
N _% <u>S</u> ummary Statistics	Position: Columns T Hide Category Position: Default	-
S <u>Categories and Totals</u>	Source: Row Variables	
	OK Paste Reset Cancel Help	

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
UrbanConsumption 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.





ţ.			*Bloc	k-3 Househo	old charact	teristics.sav	[DataSet1]	- IBM SPSS S	Statistics D	ata Editor			×
<u>F</u> ile	<u>E</u> dit	View	<u>D</u> ata <u>T</u> ran	sform <u>A</u> naly	ze Direct <u>N</u>	Aarketing (aphs <u>U</u> til	ities Add- <u>o</u> n	s <u>W</u> indow	<u>H</u> elp			
								*,	4	☆ 🎹			AB
											Visible: 42	of 42 Variab	les
		t_c ,	gift_free	var	var	var	var	var	var	var	var	var	
	1		-										
	2												
	3		-										
4	4												
	5												-
	7	1											
	, 8												
	9	ur	150.00										
1	0		245.00										
1	1	ri	-										Ŧ
		4										••••••	
Data	View	Variable	e View										
							IBM SPSS	Statistics Proce	essor is read	y Unic	ode:ON	Weight On	

Table 12: Computed Variable- "gift_free"

Further, the option for (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 loans at the 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.

t	Compute Variable: If Cases	×
Centre code, Rou FSU Serial No. [F Round [ROUND] Schedule Numbe Sample [SAMPLE] Sector [SECTOR] State [STATE] Region [REGION] Stratum [STRATU Sub-Stratum No Sub-Stratum No Sub	Compute Variable: If Cases Include all cases Include if case satisfies condition: SECTOR = 1 Image: satisfies condition: Image: satisfies condition: SECTOR = 1 Image: satisfies condition: Image: satisfies condition:	ables:
 household_type Religion [RELIGI Social Group [SO Land Possessed 	Ln Lngamma Mod Rnd(1)	T
	Continue Cancel Help	

Fig. 24: Compute Variable: If Cases

🍓 *Block-4 🛙	Demogra	phic an	d other	r partic	culars o	of hous	ehol	d me	mbers	s.sav [[DataSe	1] - IBM	SPSS Stat	istics Dat	a Editor								-		×
<u>F</u> ile <u>E</u> dit	View	<u>D</u> ata	Trans	form	Anal	yze (Direc	t <u>M</u> ar	keting	<u>G</u> r	raphs	Utilities	Add- <u>o</u>	ons <u>W</u>	ndow <u>H</u> elp										
😂 H				2	2	l	đ		. 3		å	*	4		4	14 (ABG							
1: DIST_FRO	M_SEC	4																					Visible: 51	l of 51 Varial	bles
	ehold		. L						G	IFTS	AND	LOANS	filte	er S	var	var	var	var	var	var	var	var	var	var	
1	in: reg		. gr	2.								60000	Not S	Selected											
2	in: reg		. gr	2.								60000	Not S	Selected											
3	in: reg		. gr	2.								60000	Not S	Selected											
	in: reg		. gr	2.								60000	Not S	Selected											
	in: reg		. gr	2.								60000	Not \$	Selected											
	in: reg		. gr	2.								60000	Not S	Selected											
7	al: othe	rs	. 0	2.				0.				26000	\$	Selected											
8	al: othe	rs	. 0	2.				0.				26000	\$	Selected											
9	al: othe	rs	. 0	2.				0.				26000	\$	Selected											
10	al: othe	rs	. 0	2.				0.				26000	\$	Selected											
	in: reg		. 3	2.								20000	Not 3	Selected											
12	in: reg		. 3	2.								20000	Not S	Selected											
-13-	in: reg		. 3	2.								20000	Not S	Selected											
	in: reg		. 3	2.								20000	Not S	Selected											
	in: reg		. 3	2.								20000	Not \$	Selected											
	an: reg		. 3	2.								20000	Not \$	Selected											
17	an: 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											-
	1	_		_	_	_	_	_		_	_	_				***									
Data View	Variable	e View																							
																				IBM SPSS S	Statistics Proc	essor is ready	Filte	r On	

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".

t a	Recode into Different Variables	×
 Social Group [SO Land Possessed Household incurr No. of such depe Total amount sen Distance from ne Distance from ne Consumption ex Household uniqu Total and the set of the	Numeric Variable -> Output Variable: Output Variable DIST_FROM_NEAREST_PRIMARY_CL Name: Combined_value Label: Label: Change Old and New Values Change V Paste Reset Cancel Help	
	Fig. 25.1: Recoding of Different Variables	

Recode into Different V	/ariables: Old and New Values
Old Value Value: System-missing System- or user-missing Range: 1 through 2 Range, LOWEST through value: Range, value through HIGHEST: All other values Continue	New Value Image: Value: 10 System-missing Copy old value(s) Old> New: 1 thru 2> 10 ELSE> 20 Add Change Remove Output variables are strings Width: Convert numeric strings to numbers ('5'->5) Cancel

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.

t a			×	Block-3 H	lousehold	characterist	tics.sav [Dat	aSet1] - IBN	1 SPSS Stat	istics Data E	ditor		×
<u>F</u> ile	<u>E</u> dit	View	<u>D</u> ata	<u>T</u> ransform	<u>A</u> nalyze	Direct <u>M</u> arke	ting <u>G</u> raphs	s <u>U</u> tilities	Add- <u>o</u> ns	<u>W</u> indow <u>H</u>	lelp		
					∼ 📱	i 📥 =	۲ ۲	#					A.
											Visi	ble: 43 of 43	Variables
		с	ombined	l_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									~
		1											
Data \	View	Variable	View										
							IBN	I SPSS Statist	ics Processo	r is ready	Unicode:O	N Weig	ht On

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.

ta 🛛	Rank Cases	
 Distance from ne Distance from ne Consumption exp Consumption exp Consumption exp Consumption exp Consumption exp Consumption exp Gonsumption exp Gift_free 	Variable(s): Rank Types ✓ Consumption expen <u>T</u> ies By:	
Assign Rank 1 to	Display summary tables	
Smallest value		
Largest value		
ОК	Paste Reset Cancel Help	

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.

ta Rank	Cases: Types	۲.
Rank Savage score <u>F</u> ractional rank	 Fractional rank as <u>%</u> Sum of <u>case weights</u> <u>N</u>tiles: 10 	
Proportion <u>e</u> stimate Proportion Estimation <u>Blom</u> <u>Tukey</u>	es 🔄 N <u>o</u> rmal scores n Formula 🔊 Ran <u>k</u> it 🎯 <u>V</u> an der Waerden	
Continue	Cancel Help	

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

e l			*Block-3 H	lousehold	characterist	ics.sav [Dat	aSet1] - IBN	A SPSS Stat	istics Data E	ditor		×
<u>F</u> ile <u>E</u> d	lit <u>V</u> iew	<u>D</u> ata	Transform	<u>A</u> nalyze	Direct <u>M</u> arket	ing <u>G</u> raphs	s <u>U</u> tilities	Add- <u>o</u> ns	Window H	elp		
🔁 I				∼ 📱	1 📥 🗐	ч	*					A.
										Visi	ble: 45 of 45 \	/ariables
	N	OTAL	var	var	var	var	var	var	var	var	var	,
1												
2												
3												
4	_	1										
5		1										
5	_	1										
- 1		1										
9	_	1										
10	_	. 1										
11		1										-
	4											
Data Vie	w Varial	le View										
						IBI	M SPSS Statis	tics Process	or is ready	Unicode:O	N Weigh	t On

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.

ta w	/eight Cases	x
 Distance from neare Distance from neare Distance from neare Distance from neare Consumption expend 	 Do not weight cases Weight cases by Frequency Variable: weight_to_attach_while Current Status: Weight cases by weight Reset Cancel Help	

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 Eunction... below the Summaries of Variable(s):. The new Option Box that appear is shown in Fig. 28.1.

ta Aggre	egate Data: Aggreg	ate Function
Summary Statistics	Specific Values	Number of cases
© <u>M</u> ean	© <u>F</u> irst	© Weighted
© Media <u>n</u>	© <u>L</u> ast	Weighted missing
© <u>S</u> um	Minim <u>u</u> m	O Unweighted
Standard Deviation	© Ma <u>x</u> imum	O Unweighted missing
Percentages, Fractions, C	Counts	
○ Percentages	Fractions OC	ounts
Outside Low:		High:
6	continue Cancel	Help

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 OK ;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)

t)		*Untitled	2 [New_Data] - IBM SPS	S Statistic	s Data Edito	r			×
<u>F</u> ile <u>E</u> dit	<u>V</u> iew <u>D</u> ata <u>T</u> ran	sform <u>A</u> nalyze C)irect <u>M</u> arketing	<u>G</u> raphs	<u>U</u> tilities	Add- <u>o</u> ns <u>W</u>	<u>(</u> indow <u>H</u> elp			
	🖨 🛄 🖬	r 🛥 📕		ĸ			s 🔛	 		ABC
								Visit	le: 2 of 2 Va	riables
	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								3
8	Rajasthan	3886.26								
9	Uttar Pradesh	3477.04								
10	Bihar	2830.09								
11	Sikkim	3254.82								
12	Arunachal Pradesh	3674.43								
12	Magaland	6740.88								
Data View	Variable View									
Open data o	locument				IBM SPS	S Statistics Pro	cessor is ready	Unicod	ie:ON	

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).

ţ,	Add Variables to Block-3 Household characteristics.sav[DataSet1]	×
Sel	ect a dataset from the list of open datasets or from a file to merge with the active dataset An <u>o</u> pen dataset	
	Untitled2[DataSet3]	
0	An external SPSS Statistics data file	
		Browse
Nor	n-SPSS Statistics data files must be opened in SPSS Statistics before they can be used as pai	rt of a merge.
	Continue Cancel Help	

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 $\boxed{\text{Key Variables:}}$ after selecting the $\boxed{\text{Match cases on key variables}}$. Select the $\boxed{\text{Active dataset is keyed table}}$, and click at "OK". As a result, a new variable $\boxed{42 \quad \text{Tution_fee}}$ will be added in the dataset as Var00001 as shown in Table 17.

Add Variables from Data	Set3	×
Excluded Variables:		New Active Dataset:
Rename Match cases on key variables Cases are sorted in order of key variables in both datasets Non-active dataset is keyed table Active dataset is keyed table Eoth files provide cases	•	CENTRE_ROUND FSU_SL_NO>(*) ROUND<(*) SCH_NO>(*) SAMPLE<(*) SECTOR(*) STATE<(*) REGION<(*) DISTRICT<(*) CTDATLUL.(*) Key Variables: HHID>
Indicate case source as variable: source01 (*)=Active dataset (+)=DataSet3		
OK Paste Reset Cance	Help	0

Fig. 29.1: Adding Variables to Dataset

SPSS 22: User Guide

t)			*Blo	ck-3 Househ	old charac	teristics.sa	v [DataSet1] - IBM SPSS	Statistics D	ata Editor			x
<u>F</u> ile	<u>E</u> dit	<u>V</u> iew	<u>D</u> ata <u>T</u> ra	nsform <u>A</u> nal	yze Direct <u>I</u>	<u>M</u> arketing	<u>G</u> raphs <u>U</u>	tilities Add- <u>o</u>	ns <u>W</u> indov	v <u>H</u> elp			
					E			*;	4	3			ABC
6 : VA	R00001		45.00								Visible: 46	of 46 Varia	ibles
		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									
1	0	0.045	8	36.00									
1	1	8.495	5	75.00									-
		•											
Data	View	Variable	e View										
							IBM SPSS St	atistics Proces:	sor 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.

Re	structure Data Wizard ×
Welcome to the Restruction This wizard helps you to restructure your groups of related cases (rows) or vice ve Image: The wizard replaces the current restructuring cannot be undone.	ure Data Wizard! data from multiple variables (columns) in a single case to orsa, or you can choose to transpose your data. data set with the restructured data. Note that data
	 What do you want to do? Restructure selected variables into cases Use this when each case in your current data has some variables that you would like to rearrange into groups of related cases in the new data set. Restructure selected cases into variables Use this when you have groups of related cases that you want to rearrange so that data from each group are represented as a single case in the new data set. Transpose all data All cases will become variables and selected variables will become cases in the new data set. (Choosing this option will end the wizard, and the Transpose dialog will appear.)
< Back	ext > Finish Cancel Help

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.

ta Restructure Da	ta Wizard - Step 2 of 5
Cases to Variables: Select Va	ariables
Data from case groups in the current file will be	e restructured into single cases in the new file.
Choose variables that identify case groups by r Optionally you can also choose Index Variables	noving those variables into the Identifier Variable list.
The variables that remain in the list of V within a case group or data that do not v	ariables in the Current File either contain data that vary ary.
A variable with data that vary will becom variable with data that do not vary will be	e a group of new variables in the restructured file. A copied into the new file.
Variables in the <u>C</u> urrent File:	I <u>d</u> entifier Variable(s):
🚜 state_region [region_code] 🖉 🖆	Consumption expenditure during la
state_region_district [district_code]	Distance from nearest school - Pri
Household size [HH_SIZE]	a Land Possessed code [LAND_PO
NIC-2004 Code (5-digit) [NIC_20	
household type [household type]	
Religion [RELICION]	
Social Group (SOCIAL_GROUP)	
B Household incurring expenditure f	
No. of such dependants INO OF	
Total amount sent / to be sent (Rs	
Distance from nearest school - U	Index Variable(s):
💫 Distance from nearest school - S	Household identification number [H
Consumption expenditure during I	
< <u>B</u> ack <u>N</u> ext >	Finish Cancel Help

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 so	ted by t	he Identifier an	d Index va	iables and a	gain c	lick on "N	ext".				

Cases to Variables: Sort	ing Data
The variables that you used to identify c can be restructured. If you are not sure	ase groups in the current file need to be sorted before the file about your data, select "Yes".
SP55	Sort the current data?
2 1 3 006 3 1 1 1 1 003 1 1 1 1 1 003 2 1 2 008 1 1 3 002 2 1 2 009 1 1 3 002 2 1 2 004 2 1 3 002 1 1 2 004 2 1 3 002 1 1 2 004 2 1 2 006 1 1 2 004 2 1 2 006 1 1 2 004 2 1 2 006 3 1 1 1 010 3 1 1 010	
2 1 3 006 3 1 1 003 1 1 2 004 2 1 1 003 1 1 2 004 2 1 1 003 1 1 2 004 2 1 1 0007 2 1 1 002 1 1 2 007 2 1 1 006 1 1 2 007 2 1 1 008 2 1 1 002 2 1 3 006 3 1 1 3 002 2 1 3 006	\bigcirc No - use the data as currently sorted

Fig. 30.3: Restructure Data Wizard (Step 3)

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

Group by original variable (for example: w1 w2 w3, h1 h2 h3) Group by index(for example: w1 h1, w2 h2, w3 h3 se Count Variable Count the number of cases in the current data used to create a new case Name: Label:	o create a new case
se Count Variable <u>Count the number of cases in the current data used to create a new case Name: Label: </u>	o create a new case
Se Count Variable <u>Count the number of cases in the current data used to create a new case Name: Label: </u>	o create a new case
Count the number of cases in the current data used to create a new case Name: Label:	o create a new case
Name: Label:	
Label:	
icator Variables	
Create indicator variables	
Roothame. Ind	

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.

t)	Add Cases to Block-3 Household characteristics.sav[DataSet1]	×
Sele	ect a dataset from the list of open datasets or from a file to merge with the active dataset	
© /	An <u>o</u> pen dataset	
0.7	An external SPSS Statistics data file	
		Browse
Non	-SPSS Statistics data files must be opened in SPSS Statistics before they can be used as par	t of a merge.
	Continue Cancel Help	

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.

Add Cases From C:\Users\Guest\Desktop\Tareef\NSS 6	4th Round	-Participation and Exp in Education\Nss64_25.2_new format 🗙						
Unpaired Variables:		Variables in New Active Dataset:						
combined_value(*)		CENTRE_ROUND_SHIFT>						
gift_free(*)		FSU_SL_NO>						
NTOTAL(*)		ROUND<						
RTOTAL(*)	-	SCH_NO>						
VAR00001(*)		SAMPLE<						
	P <u>a</u> ir	SECTOR						
		STATE<						
		REGION<						
		DISTRICT<						
		Indicate case source as variable:						
R <u>e</u> name		source01						
(*)=Active dataset (+)=C:\Users\Guest\Desktop\Tareef\NSS 64th Round-Participation and Exp in Education\Nss64_25.2_new format\Data\Block-3 Household characteristics.sav								
ОК	aste <u>R</u> eso	Cancel Help						

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	80	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	80	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	80	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	80	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	80	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	80	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	80	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.

ta 👘	Cell Propertie	es	×	<								
Font and Background	Format Value	Alignment	and Margins									
Preview in Preferred S	Size											
	AaBbCc 123											
Font												
Eamily:		Style:										
SansSerif		Bold	-									
<u>S</u> ize: 9 ▼ pt.												
Text Color												
(0, 0, 0)												
Background												
(255, 255, 255)												
		Edit	<u>R</u> eset									
	(0, 0	0, 0)										
OK Cancel Apply Help												

Fig. 34: Cell Properties- Font and Background

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

de l			Cell Propertie	es		×
Font an	d Back <u>o</u>	round	Format Value	Alignment and	Margins	
Preview	~				1	1
			Sector			
Format	t	E	armat:			
Categ	ory:		innac.			
All		#	.#		-	
Numt	per	#	.#;#.##E-#			
Date		#	, ### .##			
Time		#	.###,##			
Curre	ncy	#	.##E+##			
		#	#.#%			
		d	d-mmm-yy			
		d	d-mmm-yyyy			
		n	nm/dd/yy			
		n	nm/dd/yyyy			
		d	d.mm.yy			
		d	d.mm.yyyy			
		Y1	//mm/dd			
		Y.	/yy/mm/dd			
		X,	/ddd		-	
		100	aavada			
		<u>[</u>	ecimals: 0			
		ок	Cancel Appl	y Help		

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.



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.



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.

ta	Element Properties	×
E <u>d</u> it Propert	ies of:	
Bar1		$\mathbf{\times}$
X-Axis1 (Ba	ir1)	
Y-AXIS1 (Ba	iri)	
_Statistics-		
Variable:		
Statistic:		
Mean		-
	Set Parameters	
🔲 Display	error bars	
Error Bars	Represent	
Onfid	lence intervals	
Level	(%): 95	
© S <u>t</u> anda	ard error	
Multip	plier: 2	
Standa	ard deviation	
Multi	plier. 2	
Bar Style:		
📕 Bar		-
	Apply Cancel Help	

Fig. 39: Chart Builder- Element Properties

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



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

Dr. Jagdish Arora

Director Information and Library Network Centre An IUC of UGC Infocity, Gandhinagar -382007, Gujarat E-mail: 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/.