

PSPP Users Guide

GNU PSPP Statistical Analysis Software
Release 0.6.2



This manual is for GNU PSPP version 0.6.2, software for statistical analysis.

Copyright © 1997, 1998, 2004, 2005 Free Software Foundation, Inc.

Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.1 or any later version published by the Free Software Foundation; with no Invariant Sections, with the Front-Cover Texts being “A GNU Manual,” and with the Back-Cover Texts as in (a) below. A copy of the license is included in the section entitled “GNU Free Documentation License.”

(a) The FSF’s Back-Cover Text is: “You have the freedom to copy and modify this GNU manual.”

Table of Contents

1	Graphical User Interface	1
	DummyNode	1
1.1	Primary Windows	1
1.2	Data Editor	2
1.2.1	Data View	3
1.2.2	Variable View	4
1.2.3	Data Entry and Editing	4
1.2.3.1	Edit Variable Attributes	5
1.2.3.2	New Variable	9
1.2.3.3	New Case	9
1.2.3.4	Read Variable in Data View	10
1.2.3.5	Change Data Value	10
1.2.3.6	Select Entire Row or Column	10
1.2.3.7	Clear Variables	10
1.2.3.8	Clear Cases	11
1.2.3.9	Sort Cases	12
1.2.4	Status Bar	13
1.3	Output Viewer	13
	Examples	13
1.3.2	Save Output	14
1.4	Syntax Editor	14
1.4.1	Enter Commands	15
	Example	15
1.4.2	Run	16
1.4.3	Save	17
1.5	Main Menu	17
1.5.1	Overview	17
1.5.2	File	17
1.5.2.1	New	17
1.5.2.2	Open	17
1.5.2.3	Import Delimited Text Data	17
1.5.2.4	Save	20
1.5.2.5	Save As	20
1.5.2.6	Display Data File Information	20
1.5.2.7	Recently Used Data	21
1.5.2.8	Recently Used Files	21
1.5.2.9	Quit	21
1.5.3	Edit	21
1.5.3.1	Insert Variable	21
	Example	21
1.5.3.3	Insert Cases	21
1.5.3.4	Go To Case	21
1.5.3.5	Cut	21

1.5.3.6	Copy	21
1.5.3.7	Paste	21
1.5.3.8	Find	22
1.5.4	View	22
1.5.4.1	Status Bar	22
1.5.4.2	Font	22
1.5.4.3	Grid Lines	22
1.5.4.4	Value Labels	22
	Value Labels	22
1.5.4.6	Variables	22
1.5.5	Data	22
1.5.5.1	Sort Cases	22
1.5.5.2	Transpose	23
1.5.5.3	Split File	23
	Compare Groups	24
	Organize Output by Groups	24
1.5.5.6	Weight Cases	24
1.5.6	Transform	24
1.5.6.1	Compute	24
	Example	25
1.5.6.3	Rank Cases	28
	Rank Types	29
1.5.6.5	Recode into Same Variables	30
	Example	30
	Example	34
1.5.6.8	Recode into Different Variables	35
1.5.7	Analyze	36
1.5.7.1	Descriptive Statistics Definitions	36
1.5.7.2	Frequencies	38
	Example	39
1.5.7.4	Descriptives	42
	Example	42
1.5.7.6	Explore	44
	Example	44
1.5.7.8	Crosstabs	48
	Statistics and Measures of Association	48
1.5.7.10	Compare Means	52
1.5.7.11	One-Sample T-Test	52
1.5.7.12	Independent Samples T Test	53
	Example	54
1.5.7.14	Paired Samples T Test	56
1.5.7.15	One-Way Anova	58
	Example	58
	Example	60
1.5.7.18	Bivariate Correlation	61
	Example	62
1.5.7.20	Factor Analysis	63
1.5.7.21	Extraction	64

1.5.7.22	Rotation Dialogue Box	65
	Example	66
1.5.7.24	Reliability	69
1.5.7.25	Split-half Reliability	69
1.5.7.26	Chronbach's Alpha	69
	Example	69
1.5.7.28	Non-Parametric Statistics	71
1.5.7.29	Binomial	71
	Example	71
1.5.7.31	Chi-Square	73
	Example	74
1.5.7.33	ROC Curve	74
1.5.7.34	Test Variable	75
1.5.7.35	State Variable	75
1.5.7.36	Display	75
	Example	75
1.5.8	Utilities	78
1.5.8.1	Variables	78
	Example	79
1.5.8.3	Data File Comments	79
1.5.8.4	Display Comments in Output	79
1.5.8.5	Reset	79
1.5.9	Windows	79
1.5.9.1	Minimize all Windows	79
1.5.9.2	Split	79
1.5.10	Toolbar	80

1 Graphical User Interface

This chapter describes the graphical user interface.

DummyNode

use as a temp dummy node for all the ref commands until they get their proper names

1.1 Primary Windows

The PSPP graphical user interface provides a set of windows through which you store, edit and analyze data. You can choose to perform tasks and operations through interactive drop down menus in the data editor window or through command line statements in the syntax editor. The three primary windows include the Data Editor, Syntax Editor, and the Output Viewer.

Data Editor Central to PSPP, launches when you start a session. Similar in organization to a spreadsheet with rows and columns of data cells, this is where the data is stored, operated on, and edited.

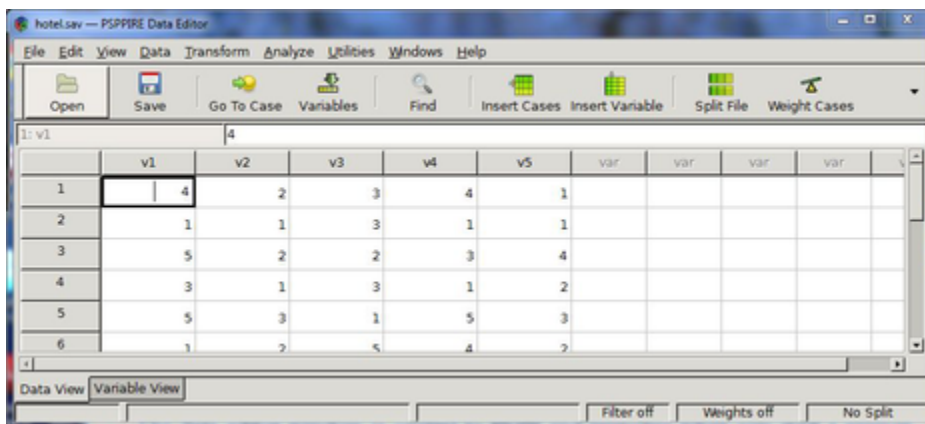
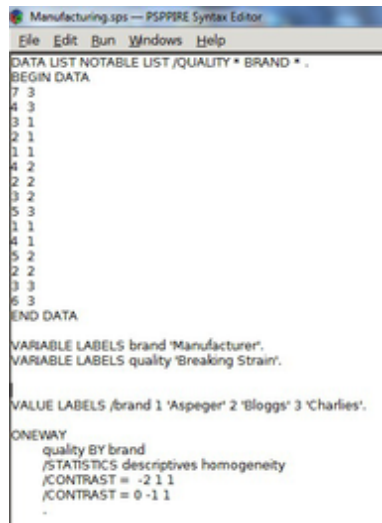


Figure 1.1: Data editor window

Syntax Editor Enter command line instructions to edit, import and analyze data. This has the advantage of enabling batch processing and code reuse. Launch the syntax editor from within the data editor by selecting File->Open->Syntax.



```

Manufacturing.sps — PSPPIRE Syntax Editor
File Edit Run Windows Help
DATA LIST NOTABLE LIST /QUALITY * BRAND * .
BEGIN DATA
7 3
4 3
3 1
2 1
1 1
4 2
2 2
3 2
5 3
1 1
4 1
5 2
2 2
3 3
6 3
END DATA
VARIABLE LABELS brand 'Manufacturer'.
VARIABLE LABELS quality 'Breaking Strain'.
VALUE LABELS /brand 1 'Aspeger' 2 'Bloggs' 3 'Charlies'.
ONEWAY
  quality BY brand
  /STATISTICS descriptives homogeneity
  /CONTRAST = -2 1 1
  /CONTRAST = 0 -1 1
  .

```

Figure 1.2: Syntax editor window

Output Viewer Displays executed instructions as well as results of statistical analyses including, text, tables, and charts.

Note: this window is not visible upon start up and only pops open after the first instruction set is executed.

1.2 Data Editor

There are two alternate views: *data view* and *variable view*. Data view shows the numerical value held by each datum while the other displays the metadictionary data of each variable. Toggle between views by clicking on the *Data View* or *Variable View* tab at the bottom left hand corner of the window.

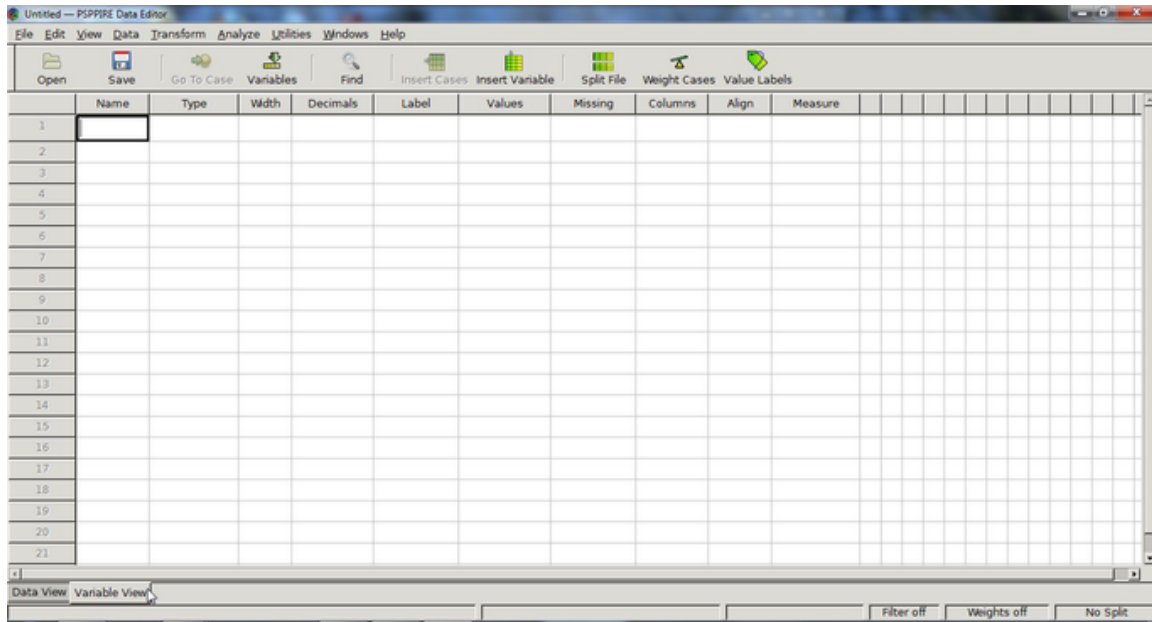


Figure 1.3: Data editor window

1.2.1 Data View

SPSS lists variables in columns and cases in rows, e.g., in the hotel customer satisfaction survey (see figure x.) each row represents the set of answers given by each respondent and each column a question.

	v1	v2	v3	v4	v5
1	4	2	3	4	1
2	1	1	3	1	1
3	5	2	2	3	4
4	3	1	3	1	2
5	5	3	1	5	3
6	1	2	5	4	2
7	3	2	4	3	1
8	1	4	5	2	1
9	3	2	3	1	2
10	2	5	4	2	1
11	4	2	2	3	5
12	2	1	4	1	1
13	1	2	5	5	2
14	2	3	3	3	1
15	4	1	1	1	3
16	1	1	5	1	2

Figure 1.4: Data view

1.2.2 Variable View

Display variable attributes associated with each case, where variable attributes (see [\[DummyNode\], page 1](#)) include name, type, width, decimal, variable label, value label, missing values, columns, align, and measurement.

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	v1	Numeric	8	0	I am satisfied with the lev	{1,"Strongly Disagree"}_	None	8	Right	Ordinal
2	v2	Numeric	8	0	The value for money was	{1,"Strongly Disagree"}_	None	8	Right	Ordinal
3	v3	Numeric	8	0	The staff were slow in res	{1,"Strongly Disagree"}_	None	8	Right	Ordinal
4	v4	Numeric	8	0	My concerns were dealt w	{1,"Strongly Disagree"}_	None	8	Right	Ordinal
5	v5	Numeric	8	0	There was too much nois	{1,"Strongly Disagree"}_	None	8	Right	Ordinal

Figure 1.5: Variable view

1.2.3 Data Entry and Editing

There are several functions available in the data editor for entering, editing and sorting data.

1.2.3.1 Edit Variable Attributes

Edit variable attributes such as Name, Type, Width, Decimals, Label, and Values, Missing, Columns, Align, and Measure move in variable view.

Name

Click a cell in the *Name* column.

Type a new name in the cell.

Press the enter key.

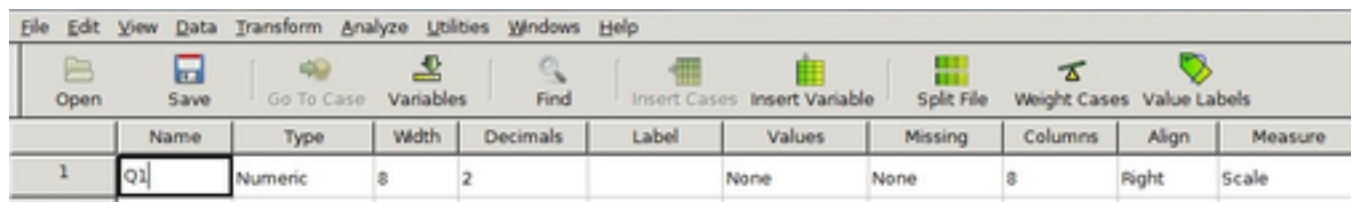


Figure 1.6: Edit *name* in variable view.

Type

Click a cell in the *Type* column.

Click on the grey box that appears to bring up a *Variable Type* dialogue box.

Choose a variable type (see [DummyNode], page 1) from the list: Numeric, Comma, Dot, Scientific Notation, Date, Dollar, Custom Currency, and String.

Enter the width and/or number of decimal places, if relevant.

Press OK.

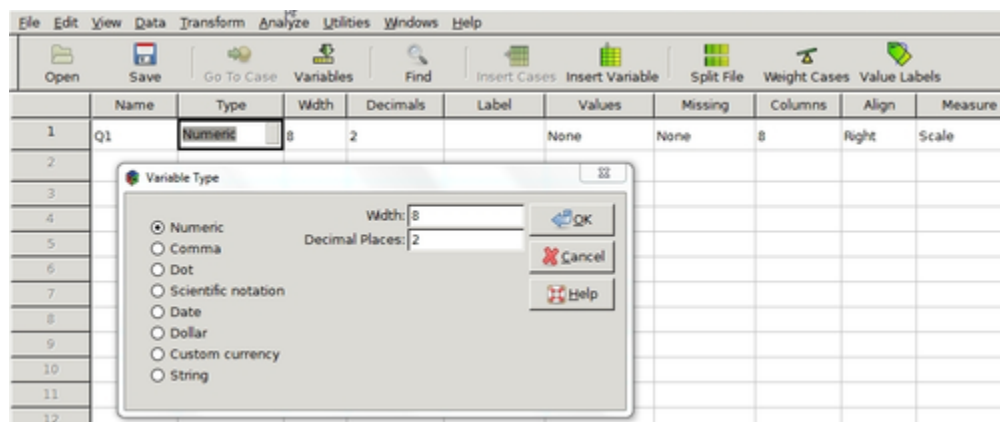


Figure 1.7: Edit *type* in variable view.

Width

Click a cell in the *Width*(see [DummyNode], page 1) column.

Select the up-down arrow that appears in the cell to alter the value.

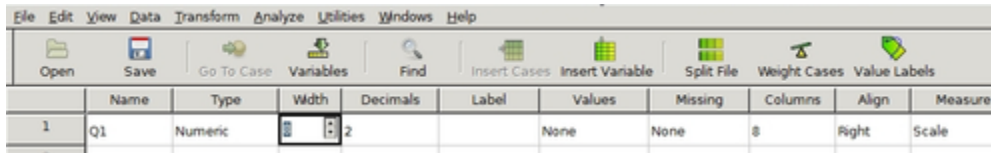


Figure 1.8: Edit *width* in variable view.

Decimals

Click a cell in the *Decimals*(see [DummyNode], page 1) column.

Select the up-down arrow that appears inside the cell to alter the value.

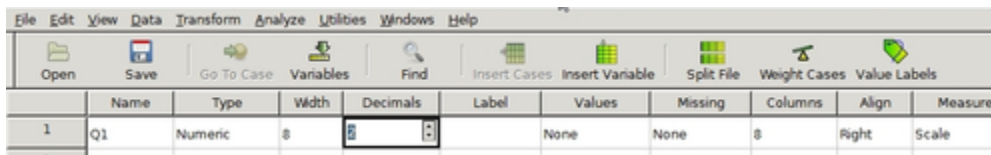


Figure 1.9: Edit *decimals* in variable view.

Label

Click a cell in the *Label* column.

Type a variable label in the cell.

Press enter.

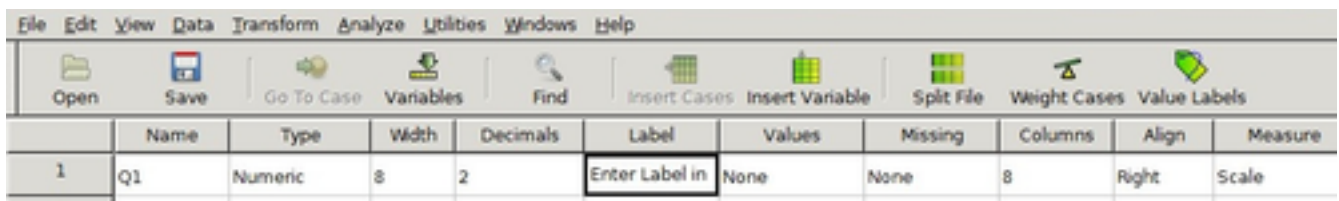


Figure 1.10: Edit *label* in variable view.

Values

Click a cell in the *Values*(see [DummyNode], page 1) column.

Click the gray box that appears inside the cell to bring up the *Value Labels dialogue* box.

Enter the *Value label*.

Click the *Add* button.

Repeat to add other labels, if necessary.

Press OK.

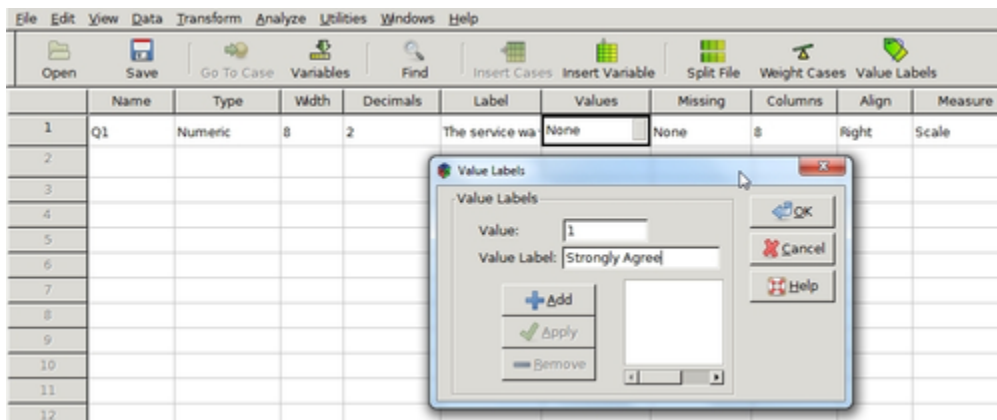


Figure 1.11: Edit *values* in variable view.

Missing

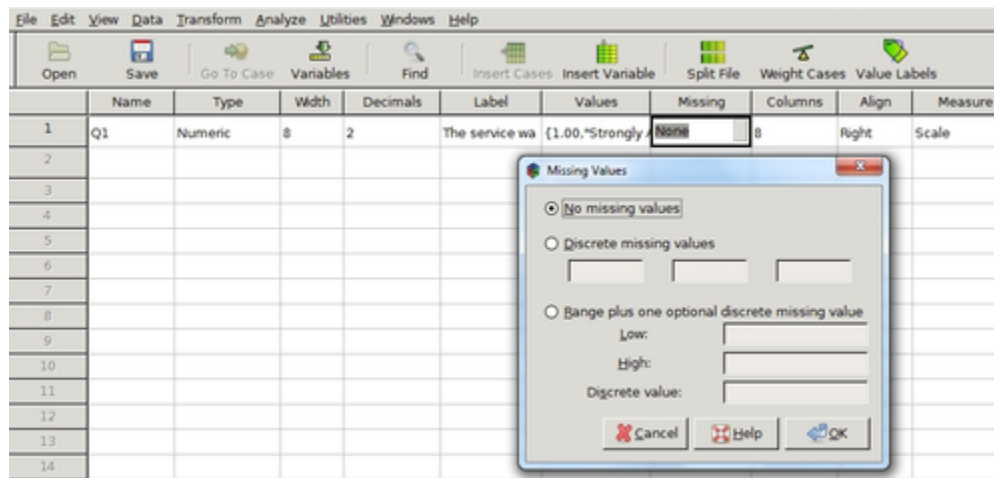
Select a cell in the *Missing* column.

Click the gray box that appears inside the cell to bring up the *Missing Values* dialogue box.

Enter either up to three discrete missing values or specify a numerical range plus one discrete value.

See [DummyNode], page 1 for further information.

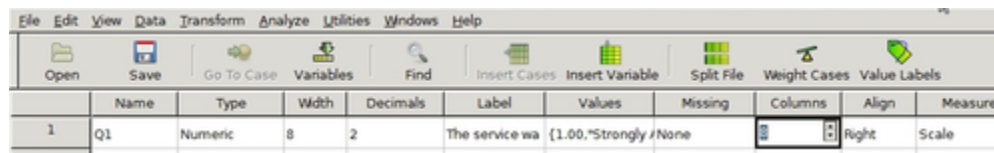
Press OK.

Figure 1.12: Edit *missing* in variable view.

Columns

Select a cell in the *Columns* (see also [DummyNode], page 1) column.

Click the up-down arrows that appears inside the cell to alter the value.

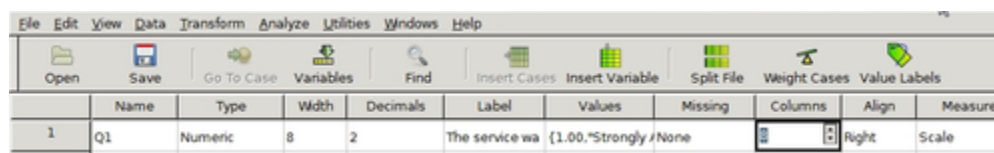
Figure 1.13: Edit *columns* in variable view.

Align

Align values left, right or center in a cell

Click a cell in the *Align* column.

Select *left*, *right*, or *center* from the drop down menu that appears.

Figure 1.14: Edit *align* in variable view.

Measure

Select a cell in the column *Measure* (see [DummyNode], page 1).

Select either *nominal*, *ordinal*, or *scale* from the drop down menu.

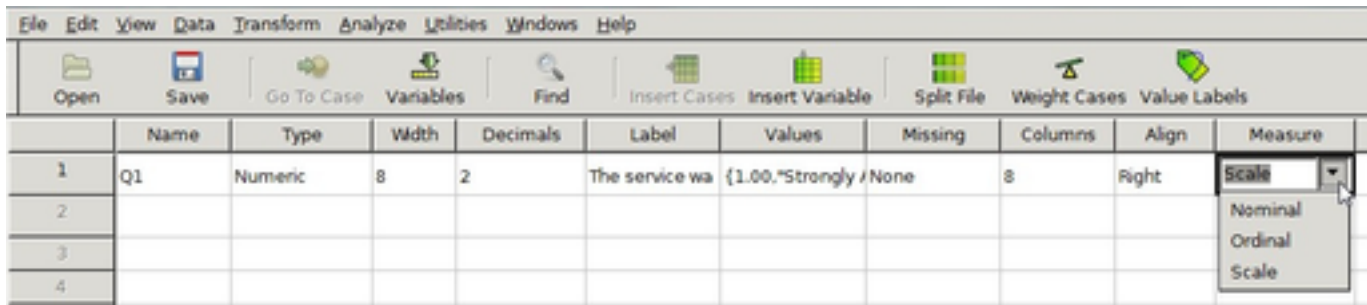


Figure 1.15: Edit *measure* in variable view.

1.2.3.2 New Variable

Enter a new variable by selecting *insert variable* from the *edit* drop down menu or tool bar. This function inserts a new variable with a generic name. In addition, if you already have at least one variable defined, then you can add a new variable to the left of the current column by right clicking on the variables title and selecting insert variable from the pop-up menu.

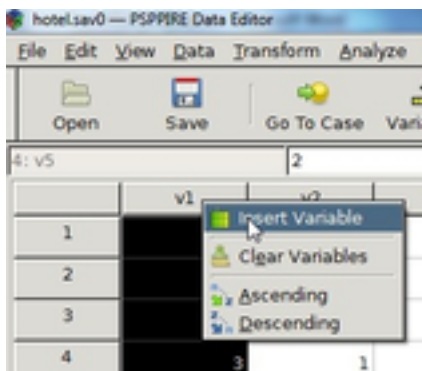


Figure 1.16: Insert new variable

1.2.3.3 New Case

There are a few ways to add a new case to an existing variable:

1. Select Edit->Insert Case from the drop down menu.
2. Select *Insert Cases* from the toolbar.
3. Right click on the row number above the insertion point to bring up a pop menu, and select *Insert Cases*.

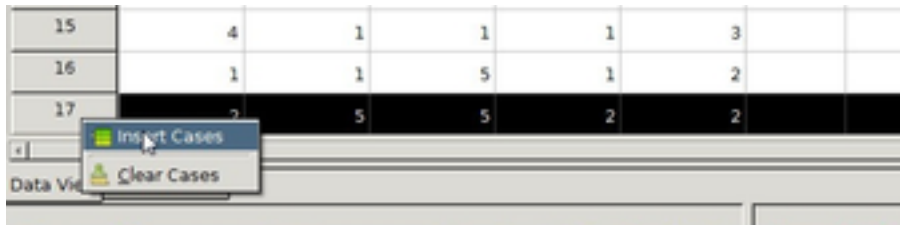


Figure 1.17: Insert new case

1.2.3.4 Read Variable in Data View

Hover the mouse over the variables name at the top of the column to view its label.

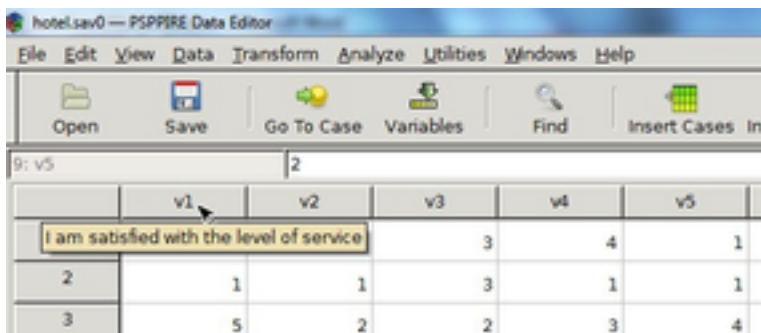


Figure 1.18: Hover mouse over variable title to view its label.

1.2.3.5 Change Data Value

Switch to *data view*, if you are not there already.

Select the cell you want to change.

Type the new value.

Press enter.

1.2.3.6 Select Entire Row or Column

Right click on the column title or row number to select the entire column or row, respectively.

1.2.3.7 Clear Variables

There are a couple of ways to clear columns:

In data view:

1. Right click on the variable name at the top of the column.
Select *Clear Variables* from the pop-up menu.
2. Right click on the column to select it.
Select Edit->Clear Variables from the drop down menu.

In variable view:

Right click on the number title to the left of the row.
Select *Clear Variables* from the pop-up menu.

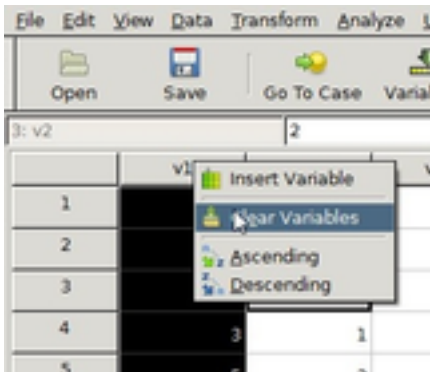


Figure 1.19: Delete variables

1.2.3.8 Clear Cases

There are a couple of ways to clear cases:

1. Right click on the row number to the left of the row.
Select *Clear Cases* from the pop-up menu.
2. Right click on the row number to the left of the row.
Select Edit->Clear Cases from the drop down menu.

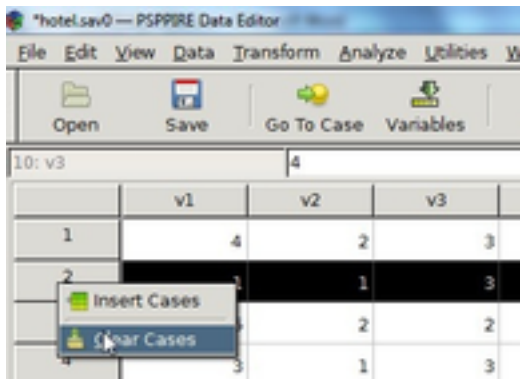


Figure 1.20: Delete cases

1.2.3.9 Sort Cases

There are two alternative ways to sort cases:

1. Right click on the variable title and select *Ascending* or *Descending* from the pop-up menu.
2. Right click on the variable title and select Data->Sort Cases (see [DummyNode], page 1).

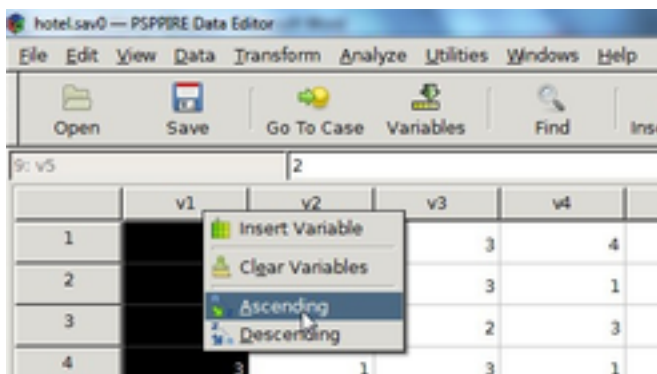


Figure 1.21: Sort cases

1.2.4 Status Bar

Located at the bottom of the data editor window, provides information about the data under process.

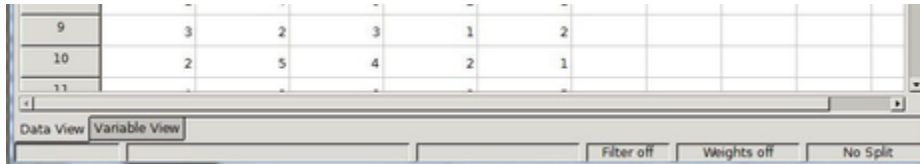


Figure 1.22: Status bar

Filter Status turns on when you filter a subset of cases in the current file for analysis.

Weight Status states the variable you are weighting the data against in the analysis.

Split File Status indicates there are separate data into groups.

1.3 Output Viewer

The output viewer displays results, charts, and the instruction set executed, whether invoked from a menu command or the syntax window.

Note: The output viewer is not visible when PSPP loads, and appears upon execution of the first set of instructions.

Examples

1.

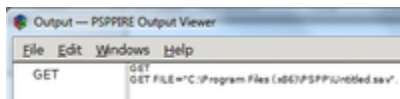


Figure 1.23: Output viewer displays open file instructions.

2.

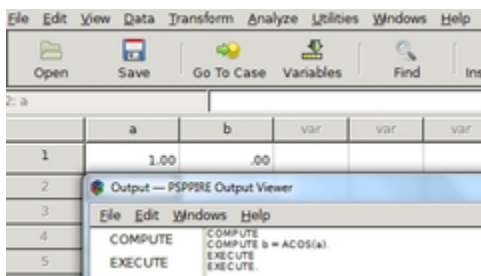


Figure 1.24: Output viewer displays executed statistics functions.

The right pane of the output viewer displays results such as tables, text, and charts. The left pane gives the corresponding outline.

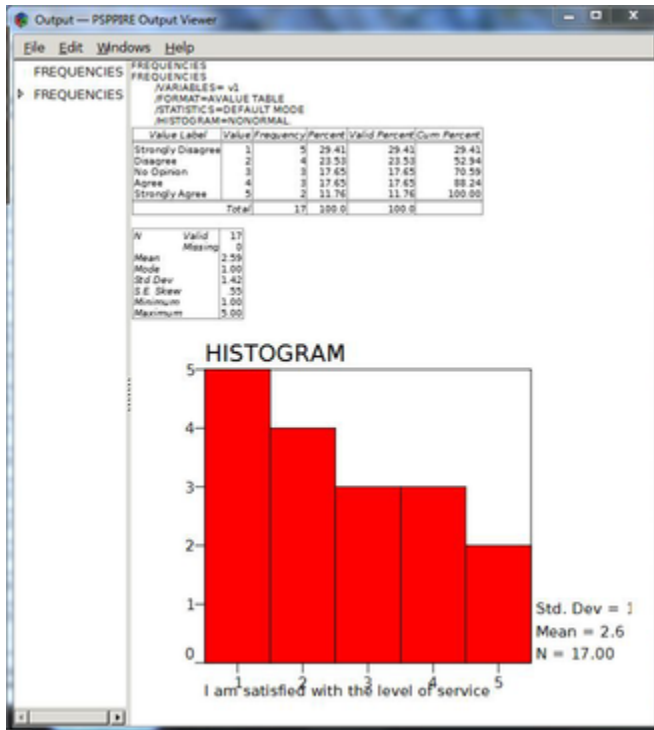


Figure 1.25: Output viewer panes

1.3.2 Save Output

To save the results in the output viewer select File->Export. File types include PDF, HTML, Open Document, Text, and Postscript.

1.4 Syntax Editor

Use the Syntax Editor window to enter command line instructions. For syntax and rules see ([DummyNode], page 1).

Enter commands into the syntax window and then execute them by selecting run from the menu. The instruction set can optionally be saved for later reuse. For example, the following set of commands in figure x populates the data editor with a list of random numbers.

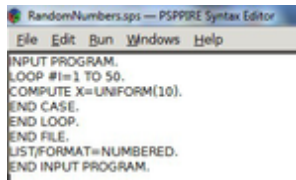


Figure 1.26: Enter command line instructions in the syntax editor.

1.4.1 Enter Commands

Most data-editor-menu functions enable you to paste the command line equivalent to that function directly into a syntax window from within the functions dialog box, by selecting the paste button.

Example

Place a command line instruction that generates descriptive statistics in the syntax editor window.

Open file xxx.sav.

Select Analyze -> Descriptive Statistics -> Descriptives (refer to section x)

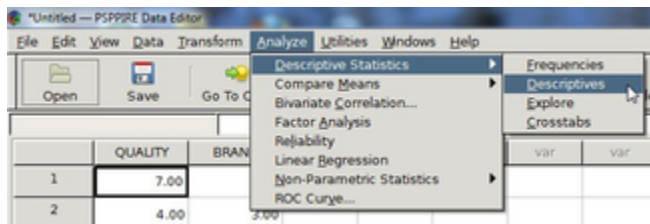


Figure 1.27: *Descriptives* function in the data editor drop down menu.

Select *variables and statistics* in the *Descriptives* dialogue box

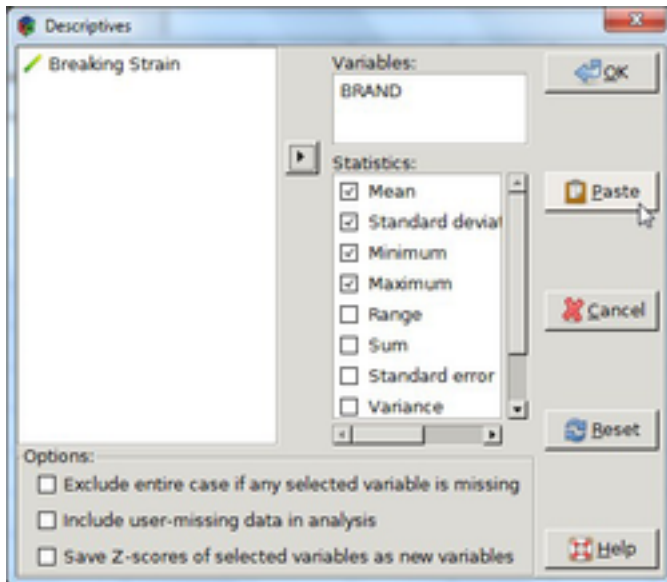


Figure 1.28: *Paste* button in the *Descriptives* dialogue box.

Click the *Paste* button to insert the command line equivalent of the function into the syntax editor.

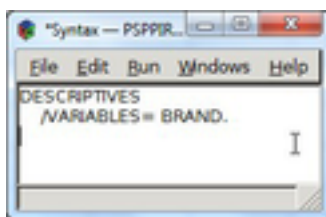


Figure 1.29: Command line equivalent of *Descriptives* function in the drop-down menu.

1.4.2 Run

Select run to execute code in the syntax window. Options include:

- | | |
|--|-------------------|
| Execute all commands | Run->All |
| Execute highlighted commands | Run->Selection |
| Execute a single line at current cursor position | Run->Current Line |
| Execute at cursor position onwards | Run->To End |

1.4.3 Save

Save a set of commands by selecting *Save* in the *File* drop down menu.

1.5 Main Menu

1.5.1 Overview

Perform tasks and operations on data with functions available through the drop down menus and toolbar. See also [\[DummyNode\]](#), page 1.

File Perform all file handling operations such opening, closing and saving files.

Edit Move data and text around with cut, copy, and paste operations. In addition, can use to insert data and cases and perform searches.

View Use to change fonts, display extra information such as a status bar and value labels, and toggle between *data* and *variable* view.

Data Perform operations on data sets, such as sort, transpose, apply weights, extract subsets of cases, and split files.

Transform Replace select data with the result of applying mathematical or other operations on it.

Analyze Perform statistical analyses on the data.

Utilities Use to insert comments and view summary information of each variable.

Windows Use to toggle between open windows, split and minimize them.

Help Provide help contents.

1.5.2 File

1.5.2.1 New

Open a new data or syntax file.

1.5.2.2 Open

Open data, syntax, system, or portable files.

1.5.2.3 Import Delimited Text Data

Import all or part of a delimited text data file into the data editor. A series of dialogue boxes guides you through this process as follows:

Importing Textual Data Dialogue Box Import all or part of the file.

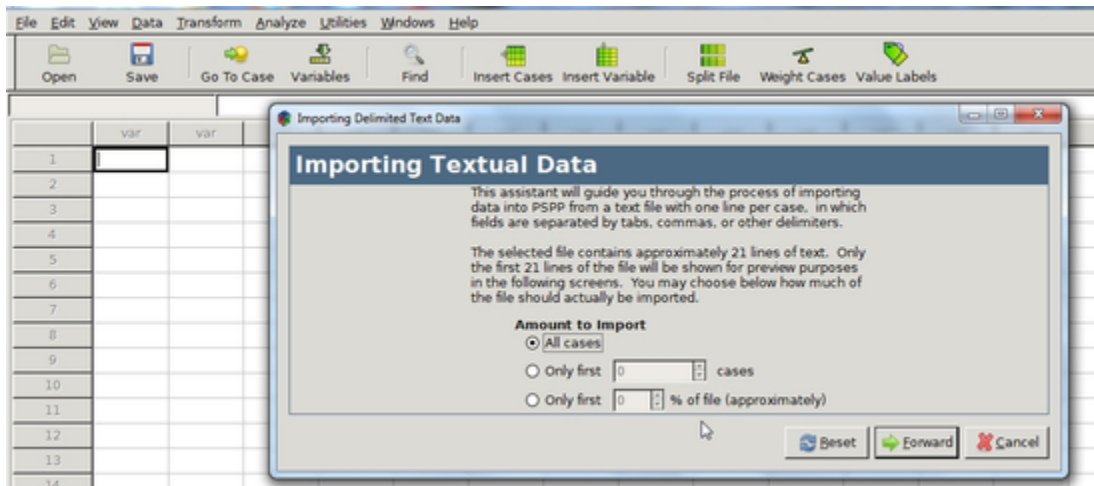


Figure 1.30: Import textual data dialogue box

First Line of Data Dialogue Select the first line of data.

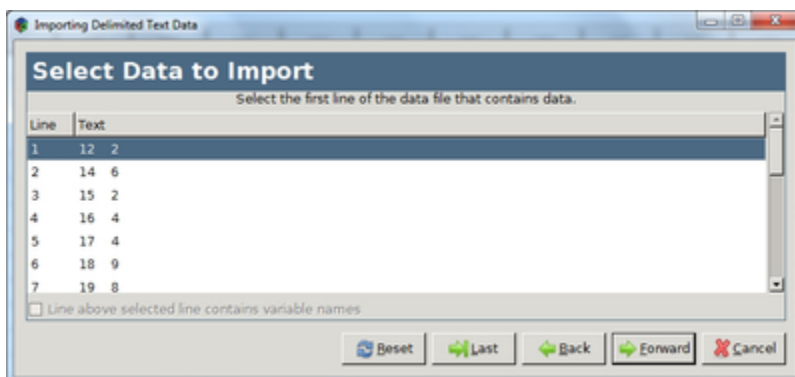


Figure 1.31: Select first line of data.

Choose Separators Dialogue Choose a delimiter. Options include:

Space

Colon (:)

Pipe (|)

Tab

Comma (,)

Semicolon (;)

Bang (!)

Hyphen (-)

Slash (/)

Custom Separator Choose how to quote separator characters as demonstrated below in figure x.

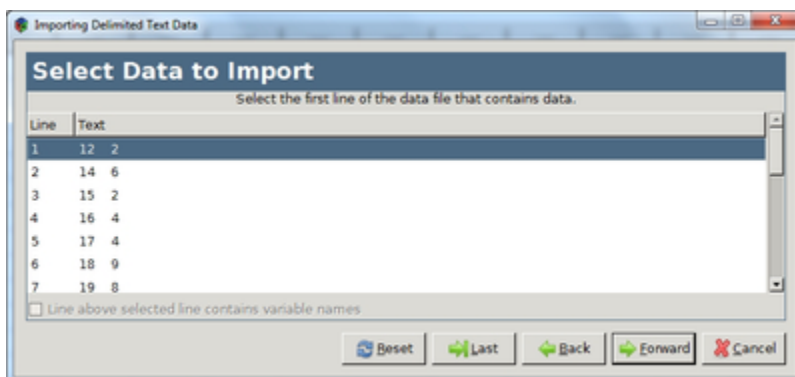


Figure 1.32: Choose data separators dialogue box

Adjust Variable Formats Verify and edit variable attributes (see [DummyNode], page 1).

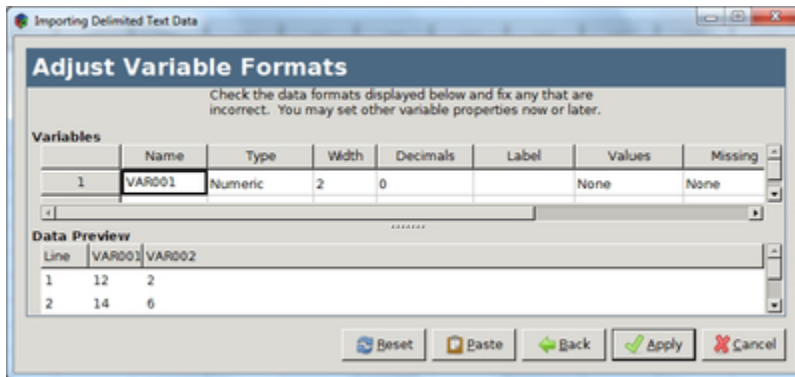


Figure 1.33: Check and fix data formats.

Select the *Apply* button to import the data.

1.5.2.4 Save

Save the current file.

1.5.2.5 Save As

Save the current file with a name and location of your choosing.

1.5.2.6 Display Data File Information

Working File Display the dictionary of the currently opened file.

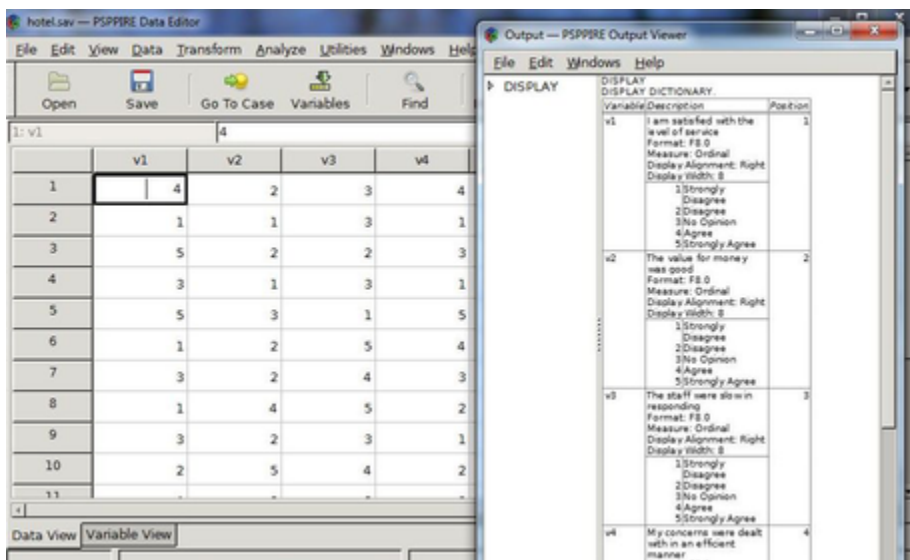


Figure 1.34: Dictionary of current file.

External File Display the dictionary of a data file that is not yet open.

1.5.2.7 Recently Used Data

Links to recently used data files.

1.5.2.8 Recently Used Files

Links to recently used syntax files.

1.5.2.9 Quit

Close the current data file and shutdown PSPP.

1.5.3 Edit

1.5.3.1 Insert Variable

Insert a new variable in data view to the left of the current column. The editor fills in system-missing values for all cases as a placeholder. By default, PSPP inserts a numerical variable type. See also [DummyNode], page 1.

Example

Select Edit->Insert Variable to enter a new variable.

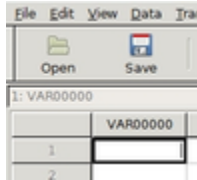


Figure 1.35: Insert variable in data editor.

1.5.3.3 Insert Cases

Insert a new case in the selected row if it is empty, or, above otherwise.

1.5.3.4 Go To Case

Jump to the selected case number.

1.5.3.5 Cut

Remove the selected text and place it elsewhere by selecting paste at the desired insertion point.

1.5.3.6 Copy

Copies text onto the clipboard. Highlight the text and then select Edit->Copy.

1.5.3.7 Paste

Places the contents stored on the clipboard at the insertion point.

1.5.3.8 Find

Look for a case.

1.5.4 View

1.5.4.1 Status Bar

Turn the status bar on or off. See [DummyNode], page 1 for further information.

1.5.4.2 Font

View and edit the current font settings.

1.5.4.3 Grid Lines

Make the grid lines in the data editor window visible or invisible.

1.5.4.4 Value Labels

Show the value label associated with each result in data view.

Value Labels

The data editor window on the right of figure x displays the value labels of the numerical survey data displayed in the window to the left.

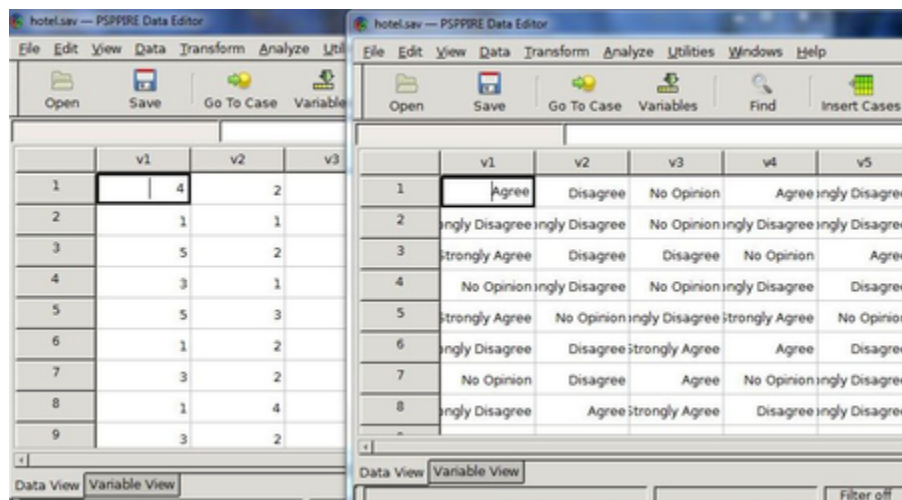


Figure 1.36: Value labels of the data in the left window.

1.5.4.6 Variables

Switch the data editor between data and variable view.

1.5.5 Data

1.5.5.1 Sort Cases

Sort cases in ascending or descending order based on one or more variables.

1.5.5.2 Transpose

Exchange rows and columns on data that is in the reverse order to the PSPP file format. Standard file format lists cases in rows and variables in columns.

1.5.5.3 Split File

Split data into groups using grouping variables. If you select more than one grouping variable then you need to nest each successive category within the previous one. sp 1

Note: Cases should be sorted with the *Sort Cases* command first such that they are in the same order as the grouping variables listed under the *Groups based on* textbox.

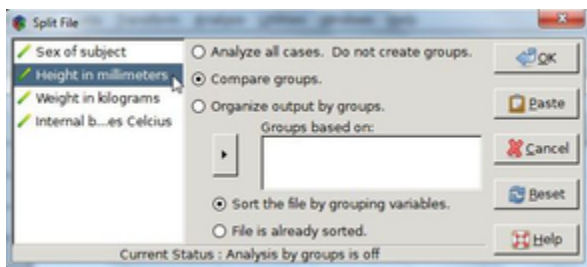


Figure 1.37: *Split File* Dialogue Box

Group a file with physiology data according to height.

	sex	height	weight	temperature	var	var	var	var
32	.00	1800	88.9	37.03				
33	.00	1801	90.5	34.42				
34	.00	1801	88.9	37.12				
35	.00	1801	90.9	39.03				
36	.00	1801	92.1	35.98				
37	.00	1801	91.7	38.98				
38	.00	1802	87.7	35.03				
39	.00	1884	88.6	39.97				
40	.00	1903	91.5	37.52				

Figure 1.38: Physiology data file

Open the file xxx.sav.
 Select Data->Split File.
 Select the *Compare groups* radio button in the pop up window.
 Click *Height in millimeters* to highlight the variable.

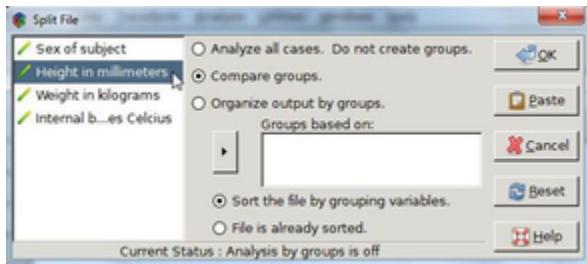


Figure 1.39: *Split-File* dialogue box

Select the arrow to the left of the *Groups based on* textbox to move the *height* variable inside that textbox.

Click OK to see the results in the output viewer.

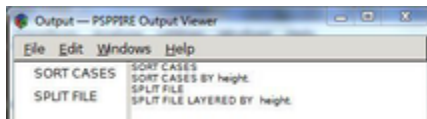


Figure 1.40: Split-file results in output viewer

Compare Groups

Show split-file group output along side each other. For example, the output viewer displays a chart for each group in a single window.

Organize Output by Groups

Show results for each split-file group separately.

1.5.5.6 Weight Cases

Assign weights to variables. Note: Cases with a zero, negative, or missing value for the weighting variable are not included in the analysis.

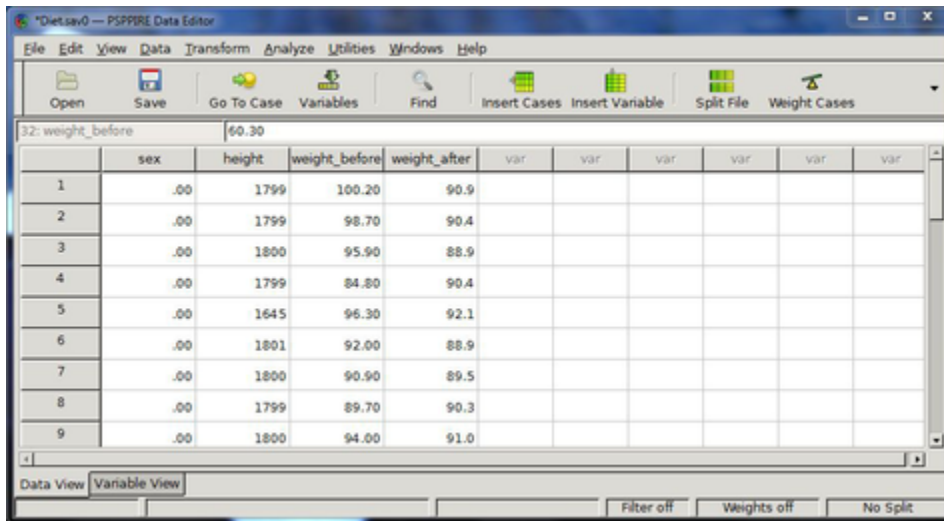
1.5.6 Transform

1.5.6.1 Compute

Perform operations on numeric and string variables. Compute new variables or replace existing ones.

Example

Compute the absolute difference in weight before and after participants follow a new diet plan (see figure x).



	sex	height	weight_before	weight_after	var	var	var	var	var	var
1	.00	1799	100.20	90.9						
2	.00	1799	98.70	90.4						
3	.00	1800	95.90	88.9						
4	.00	1799	84.80	90.4						
5	.00	1645	96.30	92.1						
6	.00	1801	92.00	88.9						
7	.00	1800	90.90	89.5						
8	.00	1799	89.70	90.3						
9	.00	1800	94.00	91.0						

Figure 1.41: Volunteers weights before and after dieting.

Open the file xxx.sav.

Select Transform->Compute from the drop-down menu to bring up a Compute Variable dialog box.

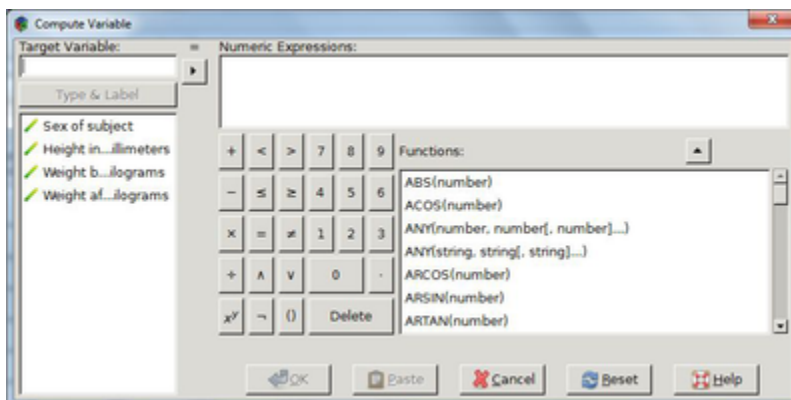


Figure 1.42: *Compute* dialog box

Enter *weightChange* in the Target Variable textbox and click the *emphType* and *Label* button below it; This variable holds the result of the calculation.

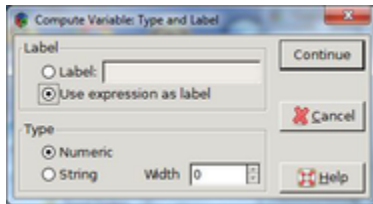


Figure 1.43: Compute solution *Type* and *Label*

Select the *Label* radio button in the *Type and Label* dialogue box, and enter *weightChange*.

Press *Continue*.

Select the *ABS(number)* function from the list of functions and press the up arrow to place the function up in the *Numeric Expressions* textbox.

Select *weightAfter* from the list of variables located on the left of the dialog box.

Note: when you hover the mouse over the list of variables this brings up their names.

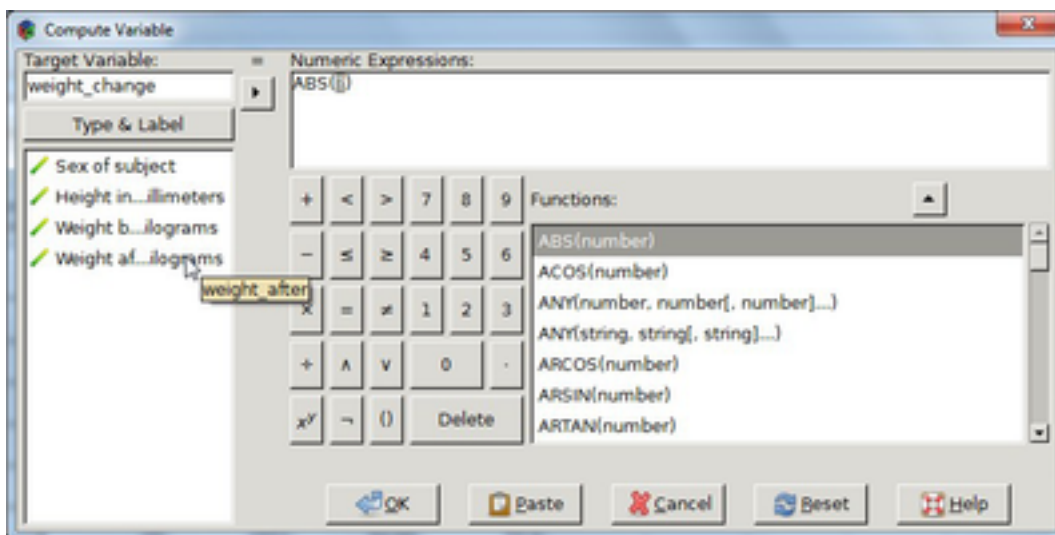


Figure 1.44: Enter numeric expression into (compute) dialogue box.

Click in between the parenthesis of the function in the *Numeric Expressions* textbox to move the cursor position.

Click the right arrow to the left of the Numerical Expressions textbox to move ABS(weightAfter) variable to the cursor position in the box.

Delete the question mark if it is still present in the ABS arguments.

Enter a space followed by a minus sign after *weightAfter*, as follows: ABS(weightAfter -)

Select the *weightBefore* variable from the list on the left side of the Compute Variable dialog box.

Place the cursor after the minus sign of the mathematical expression in the *Numeric Expressions* textbox .

Click the right arrow.

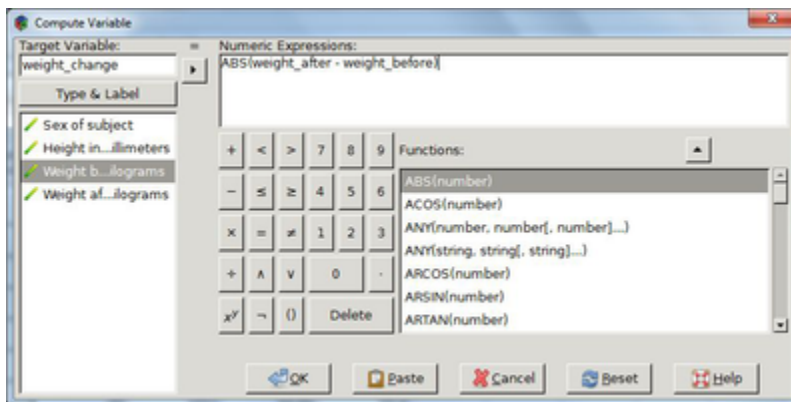


Figure 1.45: Enter formula into *compute* variable dialogue box.

Click OK to add the result to the data file.

	sex	height	weight_before	weight_after	weight_change	var	var	var	var	var
1	.00	1799	100.20	90.9	9.31					
2	.00	1799	98.70	90.4	8.31					
3	.00	1800	95.90	88.9	7.04					
4	.00	1799	84.80	90.4	5.56					
5	.00	1645	96.30	92.1	4.23					
6	.00	1801	92.00	88.9	3.12					
7	.00	1800	90.90	89.5	1.39					
8	.00	1799	89.70	90.3	.61					
9	.00	1800	94.00	91.0	2.97					

Figure 1.46: New variable *weight change* from calculation in *compute* dialogue box.

1.5.6.3 Rank Cases

Create new variables that hold ranks, Savage scores, Normal scores, and percentile values for numeric variables in the current data file. PSPPP automatically names and labels these such as to indicate original variable name and selected measures.

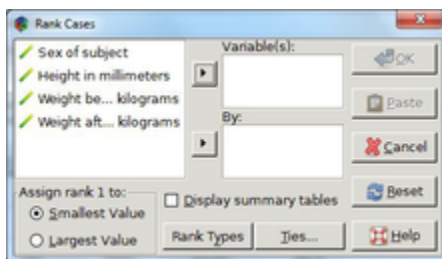


Figure 1.47: Rank cases by one or more variables.

Select the *Smallest Value* option from the *Assign rank 1 to* box for ascending order and *Largest Value* for descending.

Rank numerical variables by moving them into the *Variables* textbox (highlight the variable then click the right arrow).

Organize rankings into subgroups by moving the variables upon which the others will depend into the *By* textbox, e.g., move *gender* and *height* to calculate ranks for each combination of gender and height.

Rank Types

Bring up the *Rank Cases Types* dialogue box by clicking on the *Rank Types* Button towards the bottom of the *Rank Cases* dialogue. This dialogue box includes the following options: Rank, Savage score, Fractional rank, Fractional rank as a percent, Sum of case weights, Ntiles, Proportion Estimates, and Normal Scores.

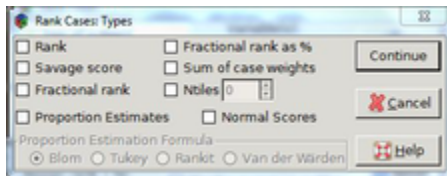


Figure 1.48: *Rank case types* dialogue box.

Rank The value of the new variable is its rank.

Savage Scores calculated using an exponential distribution

Fractional Rank Return rank divided by the sum of the weights.

Fractional Rank as a Percent Return rank divided by the number of cases as a percent.

Sum of Case Weights is returned.

Ntiles Return a rank according to the percentile group. The size of each percentile group is given by $100/Ntile$. For example, if $Ntile=4$, then rank 1 corresponds to the 1st 25th

percentile, rank 2 corresponds to cases between the 25th and 50th percentile, rank 3 to cases between the 50th and 75th percentile and rank 4 to the last percentile group.

Proportion Estimates Estimate of the cumulative proportion of the distribution corresponding to a specific rank.

Normal Scores Return z scores corresponding to the estimated cumulative proportion.

Choose from a range of proportion estimate formulae when calculating proportion estimates and normal scores; options include Blom, Tukey, Rankit, or Van der Waerden.

Blom Ranking variable based on proportion estimates which employ the formula $(r - 3/8)/(w + 1/8)$, where w is the sum of the case weights and r is the rank.

Tukey Apply the formula $(r - 1/3)/(w + 1/3)$, where r is the rank and w is the sum of the case weights.

Rankit Estimate expected values using $(r - 1/2)/w$, where w is the number of observations and r is the rank which ranges from unity to w .

Van der Waerden Return Van der Waerdens transformation, given by the formula $r/(w + 1)$, where w is the sum of the case weights and r is the rank, ranging from 1 to w .

Ties Use to select the method for assigning rankings to cases with the same value on the original variable.

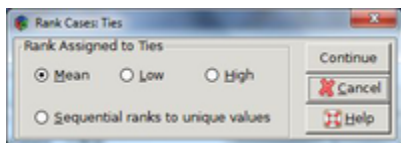


Figure 1.49: *Rank case ties* dialogue box

1.5.6.5 Recode into Same Variables

Recode different values of the same type (numeric or string) into a variable. There are two options available: 1) change a single value to a new one or 2) change a range of values to a category.

Example

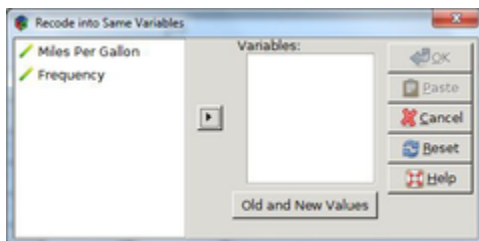
Increase the value 12 Miles per Gallon (MPG) to 13 in a data set listing the mileage per gallon of a specific brand of car.

	MPG	Frequency	var	var	var	var	var	var	var
1	12	2							
2	14	6							
3	15	2							
4	16	4							
5	17	4							
6	18	9							
7	19	8							
8	20	3							
9	21	5							
10	22	5							
11	23	3							
12	24	4							
13	25	5							
14	26	3							
15	28	3							

Figure 1.50: Mileage-per-gallon data file

Open file xxx.sav

Select Transform->codeRecode into Same Variables to bring up a Recode dialogue box.

Figure 1.51: *Recode into same variables* pop-up box

Select *Miles Per Gallon* and click the right arrow to move the variable into the *Variables* textbox.

Click the *Old and New Values* button to bring up its dialogue box.

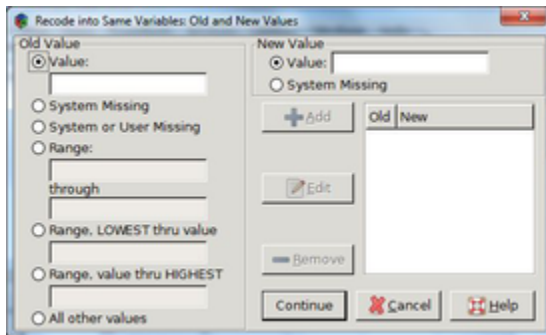


Figure 1.52: *Recode into same variables: Old and new values* pop-up box

Type the number 12 in the *Value* text box under *Old Value* on the left side.

Type the number 13 in the *Value* text box under *New Value* on the right side.

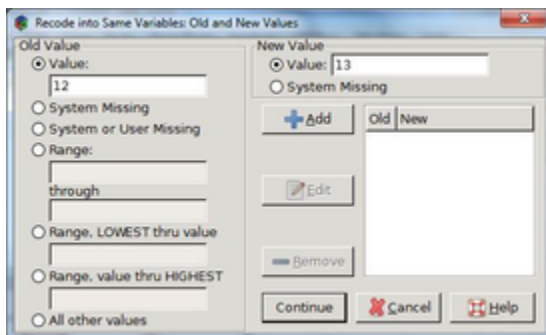


Figure 1.53: Replace cases with a specified value with a new one.

Click the *Add* button. The number 12 now appears under *old* and 13 under *new*.

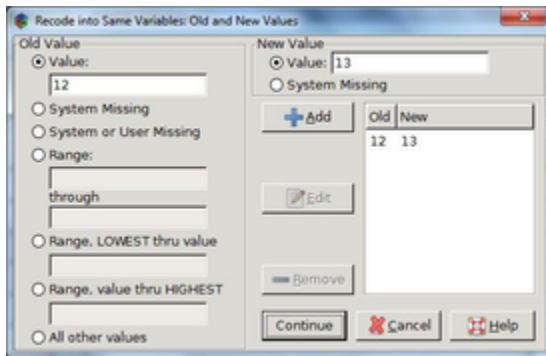


Figure 1.54: Display old and new case values in *Recode into same variables* pop-up box.

Example

Convert the mileage per gallon values in the xxx.sav automobile data set into three categories, as follows:

If *Miles Per Gallon* is less than 19 MPG then give it a value of one.

If *Miles Per Gallon* is greater than or equal to 19 MPG and less than 24 then give it a value of 2.

If *Miles Per Gallon* is greater or equal to 24 then give it a value of three.

Select Transform->Recode into Same Variables.

Select the *range LOWEST thru value* radio button, and enter 18.

Enter the number 1 in the value field under *New Value* on the right side of the dialogue box.

Click the *Add button*, to get the *LOWEST-18* listed under *old* and 1 under *new*.

Repeat with the other value ranges:

Select the *emphRange-through* radio button and then enter 19 in the first text box and 23 in the second.

Enter the number 2 in the value text box under *New Value*, and click the *Add* button.

Select the *range thru HIGHEST* radio button and enter 24.

Enter the number 3 in the value field under *New Value*, and click the *Add* button.

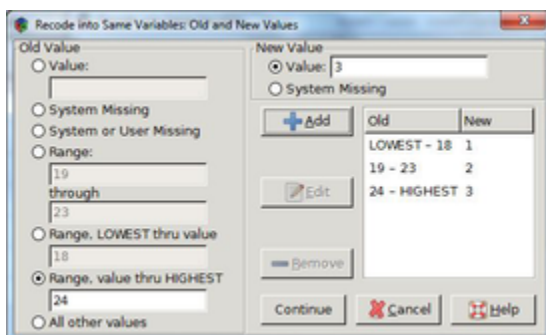


Figure 1.55: Collapse a range of values into categories with *recode into same variables*.

The screenshot shows the SPSS Data Editor window for a file named 'MPG.sav'. The window title bar includes 'File Edit View Data Transform Analyze' and 'Open Save Go To Case Var'. The main area displays a table with two columns: 'MPG' and 'Frequency'. The data is as follows:

	MPG	Frequency
1	1	2
2	1	6
3	1	2
4	1	4
5	1	4
6	1	9
7	2	8
8	2	3
9	2	5
10	2	5
11	2	3
12	3	4
13	3	5
14	3	3
15	3	3

! and New Values dialogue box disappears.

into New Variables dialogue box to put the new values in the

Figure 1.56: Calculated *frequency* data inserted into data editor.

1.5.6.8 Recode into Different Variables

Change variable values and save into a new one. There are two options available; change a single value to a new one or change a range of values to a category. Note, that the new variable type must be the same as the original; numeric or string. (see also [Section 1.5.6.5 \[Recode into Same Variables\]](#), page 30).

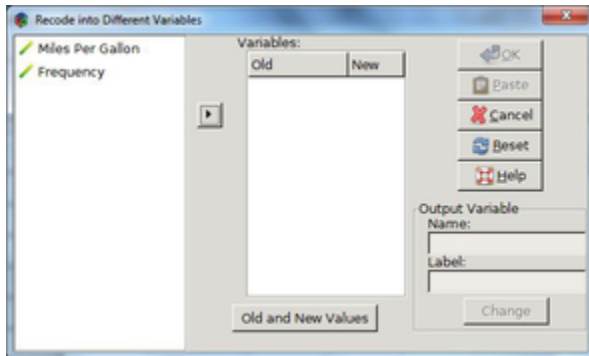


Figure 1.57: Recode case values into a new variable.

1.5.7 Analyze

1.5.7.1 Descriptive Statistics Definitions

Central Tendency Measure of the central position of a set of data.

Mean Average value, or first moment (see definition below), defined as sum of all data within a set divided by the number of cases in the set. Note: the mean is susceptible to

outlying values, e.g., the average temperature over a month in a region that has had a few unusually hot days will be higher than usual.

Note: **Moments** are the sums of integer powers of the values, where a higher integer power corresponds to a higher moment e.g. the variance is the second moment, skewness the third, and kurtosis the fourth.

Median The central value of an ordered set of numbers. In other words, 50% of the values fall above or below it. When the number of data values is even the median is the average of the middle two. The median is not as susceptible to outlying values as the mean.

Mode Most frequently occurring value. Note that there can be more than one mode in a data set.

Sum The addition of all values across the different cases excluding missing ones.

Dispersion A measure of the spread or variation in data values about the mean.

Standard Deviation A measure of the spread in values about the mean equal to the square root of the sum of the squares of the difference between each value and the mean, for all values, divided by one less than the number of cases.

Variance The standard deviation squared, or the second moment.

Range The difference between the largest and smallest numerical values in the set.

Minimum The smallest numerical value in the set.

Maximum The largest numerical value in the set.

S.E. Mean The standard error of the mean is a measure of by how much the value of the mean varies with different samples from the same distribution.

Distribution Statistics that describe the shape and symmetry of a distribution.

Skewness A measure of the asymmetry of a distribution. Estimate this by taking the third moment (proportional to the sum of the cube of the difference of each data value and the mean). Unlike the mean and variance the definition is such as to make it a dimensionless quantity, and this pure number characterizes the shape of the distribution. Positive values indicate the distribution has a more heavily pronounced right tail, while negative ones indicate a more prominent left one. When both tails are equal the skewness is zero.

Kurtosis A measure of the relative peakedness or flatness compared to a normal distribution. Estimate this by taking the fourth moment (proportional to the sum of the fourth power of the difference of each data value and the mean). Like the skewness, this is also a dimensionless quantity. A distribution with negative kurtosis is termed leptokurtic,

while a distribution with a positive value as platykurtic, and an in-between one is termed mesokurtic.

1.5.7.2 Frequencies

Calculate statistics including central tendency (mean, median, mode, and sum), dispersion (standard deviation, variances, range, minimum, maximum, standard error of the mean) and distribution (skewness and kurtosis). Optionally, view frequency tables and charts.

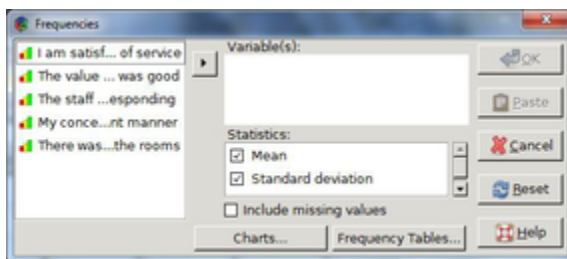


Figure 1.58: *Frequencies* dialogue box

The frequencies dialogue box lists all variables along the left side. In order to perform statistical analyses or plot charts move the variables into the *Variables* dialogue box. To do this, first, click on the variable until it is highlighted and then select the right arrow.

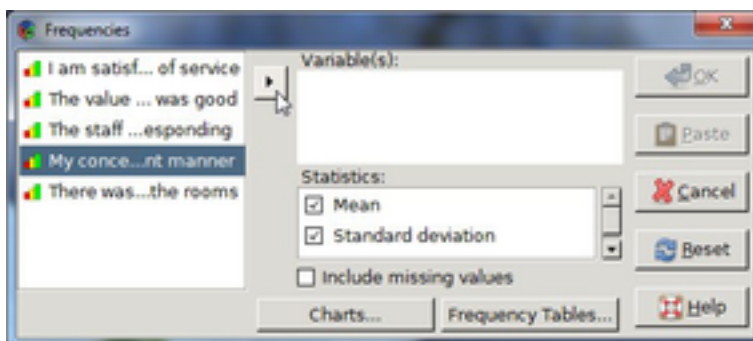


Figure 1.59: Select variables in the *frequencies* dialogue box.

Choose whether to display frequency tables, and in what way to order the tables; ascending or descending value or ascending or descending frequency.

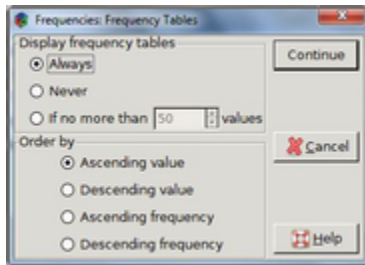


Figure 1.60: *Frequency tables* pop-up box

Plot histograms or pie charts with the Chart Dialogue box.

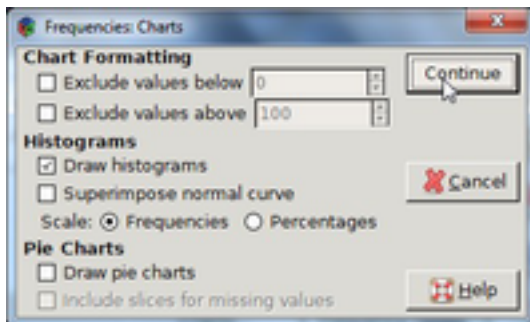


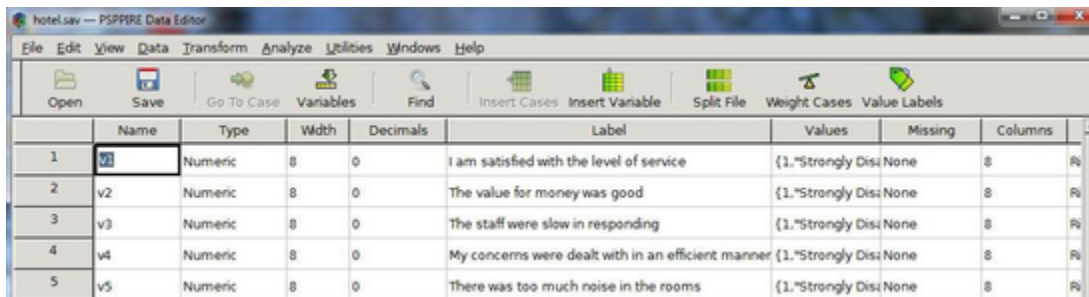
Figure 1.61: *Frequency charts* pop-up box.

Histogram options include scaling with either frequencies or percentages and superimposing a normal curve.

Pie charts come with the option to include slices for missing values.

Example

Analyze the extent to which hotel customers were satisfied with the level of service in the following survey.

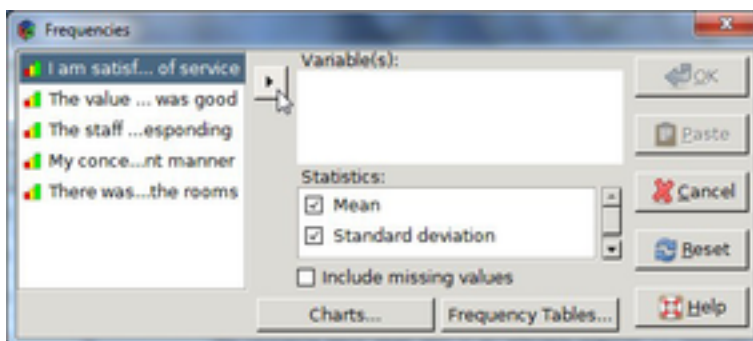


	Name	Type	Width	Decimals	Label	Values	Missing	Columns	
1	v1	Numeric	8	0	I am satisfied with the level of service	{1."Strongly Dis; None		8	R
2	v2	Numeric	8	0	The value for money was good	{1."Strongly Dis; None		8	R
3	v3	Numeric	8	0	The staff were slow in responding	{1."Strongly Dis; None		8	R
4	v4	Numeric	8	0	My concerns were dealt with in an efficient manner	{1."Strongly Dis; None		8	R
5	v5	Numeric	8	0	There was too much noise in the rooms	{1."Strongly Dis; None		8	R

Figure 1.62: Customer satisfaction survey

Select Analyze-> Descriptive Statistics-> Frequencies to bring up the *Frequencies* dialogue box.

Select the first variable from the list on the left side then click the right arrow to move it inside the *Variables* dialogue box.

Figure 1.63: Pick variables in the *frequencies* dialogue box.

Select the *charts* button to bring up the *Charts* dialogue box.

Select *Draw Histograms* and press *Continue* to close this dialogue box.

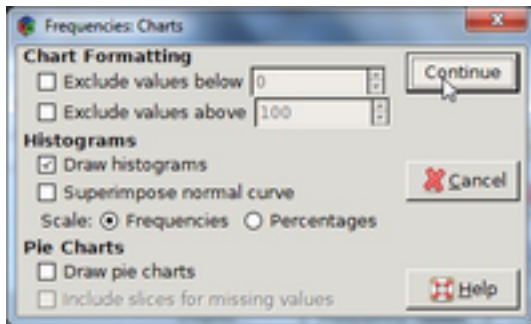


Figure 1.64: Plot histograms of selected data.

Select *Mode* from the *Statistics* options in the *Frequencies* dialogue box.

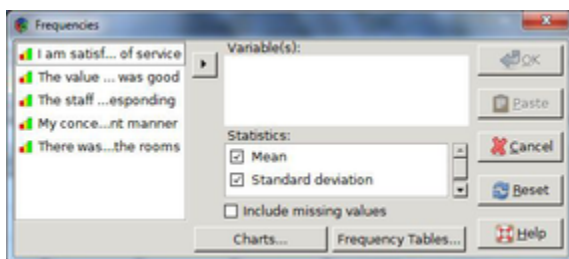


Figure 1.65: Select statistical analyses in the *frequencies* pop-up.

Press OK to display the results in the output viewer.

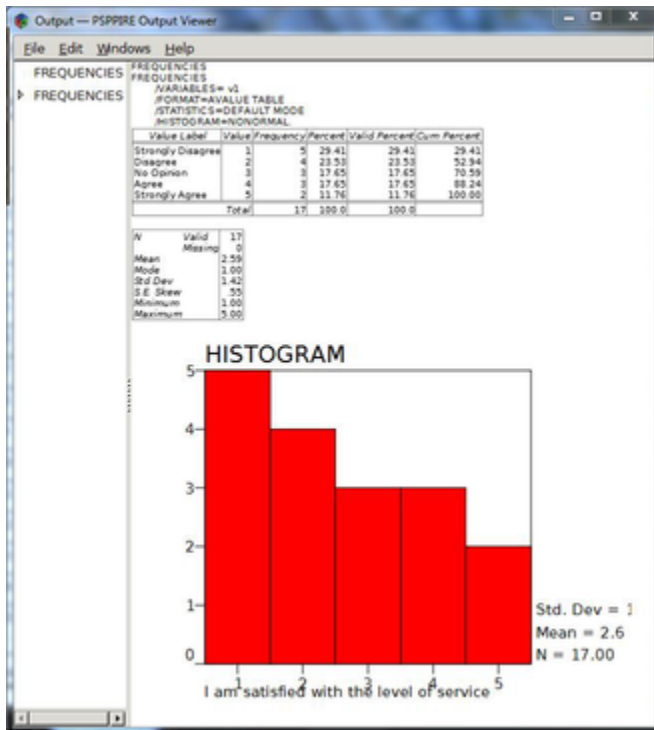


Figure 1.66: Results of *frequencies* function displayed in output viewer.

1.5.7.4 Descriptives

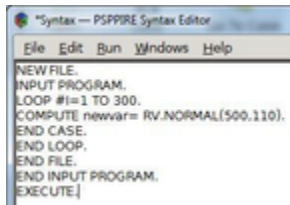
Display univariate summary statistics and z scores. Statistics include central tendency (mean), dispersion (standard deviation, variance, range, minimum, maximum, and standard error of the mean), and distribution (kurtosis and skewness).

Z-score is a dimensionless variable that classifies how many standard deviations a value is from the mean. Also known as standardizing or normalizing. Based on normal theory, these are best for quantitative variables with symmetrical distributions. Note, if the distribution is skewed the z scores will be as well.

Example

Begin by making a data set in the syntax editor. Write a small program to generate normally distributed numbers that represent SAT test scores with a mean of 500 and standard deviation of 110 as follows:

Select File->New->Syntax to open a syntax window.

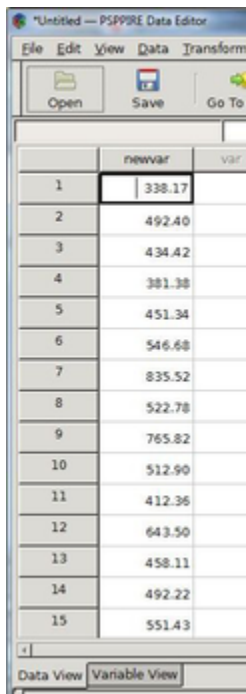


```
*Syntax — PSPPIRE Syntax Editor
File Edit Run Windows Help
NEW FILE.
INPUT PROGRAM.
LOOP #i=1 TO 300.
COMPUTE newvar= RV.NORMAL(500,110).
END CASE.
END LOOP.
END FILE.
END INPUT PROGRAM.
EXECUTE.]
```

Command lines shown in figure x. into the syntax window.

Figure 1.67: Generate data (random numbers) in the syntax editor.

Select Run->All to execute the program and populate the data editor with random numbers drawn from a normal distribution.



	newvar	var
1	338.17	
2	492.40	
3	434.42	
4	381.38	
5	451.34	
6	546.68	
7	835.52	
8	522.78	
9	765.82	
10	512.90	
11	412.36	
12	643.50	
13	458.11	
14	492.22	
15	551.43	

Figure 1.68: New data generated with command line instructions in the syntax editor.

Next, calculate the z scores using the *Descriptives* menu command.

Select Analyze->Descriptive Statistics->Descriptives to bring up the *Descriptives* dialogue box.

Select the *newvar* variable by clicking on it in the variables list to the left of the dialogue box, and then by clicking the right arrow to move it over into the *Variables* textbox.

Select the *save Z scores of selected variables as new variables* check box.

Click OK to generate a list of z scores under the new variable in the data editor window.



	newvar	Znewvar
1	338.17	-1.48
2	492.40	-.08
3	434.42	-.61
4	381.38	-1.09
5	451.34	-.45
6	546.68	.41
7	835.52	3.03
8	522.78	.19
9	765.82	2.39
10	512.90	-.10
11	412.36	-.81
12	643.50	1.29
13	458.11	-.39
14	492.22	-.08
15	551.43	.45

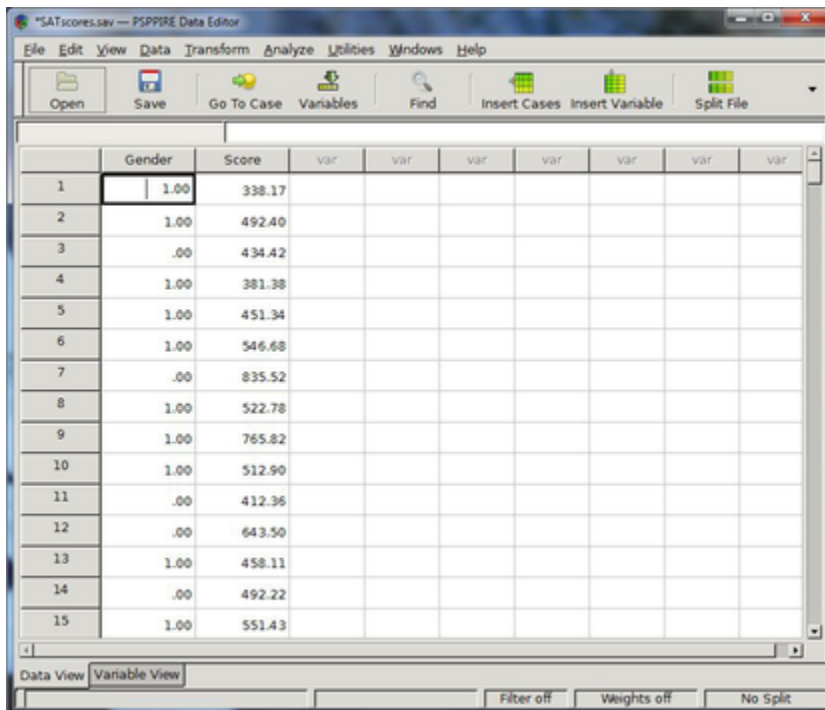
Figure 1.69: *Z-scores* generated with the *descriptives* function.

1.5.7.6 Explore

Provides descriptive statistics including percentiles and extremes, and in addition, enables you to examine subgroups of data. Descriptive statistics include the mean, median, variance, standard deviation, minimum, maximum, range, skewness, and kurtosis.

Example

Generate descriptive statistics of the following data set of SAT test scores versus gender for everyone, independent of gender, or break it down into subgroups and do a separate analysis for each gender type.



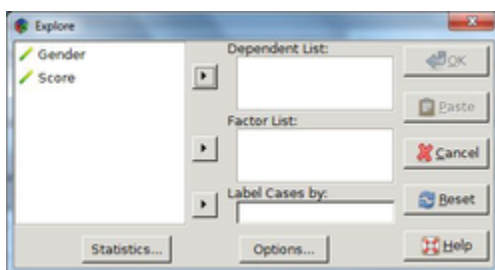
The screenshot shows the SPSS Data Editor window titled '*SATscores.sav - PSPSRE Data Editor'. The window contains a data grid with 15 rows and 10 columns. The first two columns are labeled 'Gender' and 'Score'. The 'Gender' column contains values 1.00, 1.00, .00, 1.00, 1.00, 1.00, .00, 1.00, 1.00, 1.00, .00, .00, 1.00, .00, and 1.00. The 'Score' column contains values 338.17, 492.40, 434.42, 381.38, 451.34, 546.68, 835.52, 522.78, 765.82, 512.90, 412.36, 643.50, 458.11, 492.22, and 551.43. The remaining columns are labeled 'var' and are empty. The window also shows a menu bar (File, Edit, View, Data, Transform, Analyze, Utilities, Windows, Help) and a toolbar with icons for Open, Save, Go To Case, Variables, Find, Insert Cases, Insert Variable, and Split File. At the bottom, there are tabs for 'Data View' and 'Variable View', and buttons for 'Filter off', 'Weights off', and 'No Split'.

	Gender	Score	var	var	var	var	var	var	var
1	1.00	338.17							
2	1.00	492.40							
3	.00	434.42							
4	1.00	381.38							
5	1.00	451.34							
6	1.00	546.68							
7	.00	835.52							
8	1.00	522.78							
9	1.00	765.82							
10	1.00	512.90							
11	.00	412.36							
12	.00	643.50							
13	1.00	458.11							
14	.00	492.22							
15	1.00	551.43							

Figure 1.70: Data sheet of SAT scores by gender.

Start by looking at all test scores independent of gender.

Select Analyze-> Descriptive Statistics-> Explore to open the *Explore* dialogue box.

Figure 1.71: *Explore* dialogue box

Select *Score* and press the right arrow button next to the *Dependent List* box.

Click the *Statistics* button at the bottom left side of the dialogue box to bring up the *Statistics* pop-up box.

Select the *Descriptives* and *Extremes* check boxes, and then press continue.

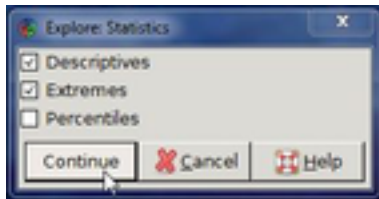


Figure 1.72: *Explore* statistic functions

Press OK on the *Explore* dialogue box to display the results in the output editor.

Cases			
	Valid	Missing	Total
	N	Percent/N	Percent
Score	300	100%/0	0%/300

Case Number	Value
ScoreHighest1	71835.52
2	282805.10
3	114773.54
4	5765.82
5	60761.71
ScoreLowest1	121179.03
2	56182.98
3	158234.39
4	300246.40
5	214258.52

	Statistic	Std. Error
ScoreMean	501.53	6.37
95% Confidence Interval for Mean: Lower Bound	488.97	
Upper Bound	514.05	
5% Trimmed Mean	499.22	
Median	501.25	
Variance	121179.95	
Std. Deviation	110.36	
Minimum	179.03	
Maximum	835.52	
Range	656.49	
Interquartile Range	137.38	
Skewness	.00	.14
Kurtosis	.24	.28

Figure 1.73: *Explore* statistical analysis results displayed in output viewer.

Repeat the analysis, examining each gender group separately, where 0 represents male.

Select Analyze->Descriptive Statistics->Explore to open the Explore dialogue box.

Select *Score* and press the right arrow button beside the *Dependent List* textbox.

Select *Gender* and press the right arrow button beside the *Factor List* textbox.

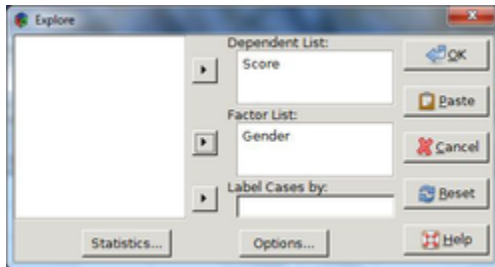


Figure 1.74: Compare SAT scores and gender with the *Explore* function.

Press the Statistics button at the bottom of the Explore dialogue box.

Select the *Descriptives* and *Extremes* check boxes and then press the continue button.

Press OK on the Explore dialogue box to display the results.

The output viewer lists different statistics for each group. For example, the lowest and highest test scores are 182.98 and 835.52, respectively, for the males, and, 179.03 and 805.10 for the female group.

		Cases			
		Valid	Missing	Total	
Gender	N	Percent/N	Percent	N	Percent
Score .00	144	100%	0%	144	100%
Score 1.00	156	100%	0%	156	100%

Gender	Case Number	Value
Score .00 Highest	1	7835.52
	2	68761.71
	3	253710.48
	4	92703.49
	5	283693.34
Score .00 Lowest	1	56182.98
	2	158234.39
	3	163265.45
	4	296271.90
	5	242285.87
1.00 Highest	1	282805.10
	2	114773.54
	3	9763.82
	4	290729.39
	5	291725.04
1.00 Lowest	1	121179.03
	2	305246.40
	3	214258.52
	4	60279.05
	5	24285.90

Gender	Statistic	Std. Error	
Score .00	Mean	502.53	9.30
	95% Confidence Interval for Mean	Lower Bound	484.14
		Upper Bound	520.92
	5% Trimmed Mean	503.49	
	Median	501.62	
	Variance	12462.02	
	Std. Deviation	111.63	
	Minimum	182.98	
	Maximum	835.52	
	Range	652.54	
	Interquartile Range	152.08	
Skewness	-.08	.20	
Kurtosis	1.8	.46	
1.00	Mean	500.57	8.77
	95% Confidence Interval for Mean	Lower Bound	483.25
		Upper Bound	517.90
	5% Trimmed Mean	500.18	
	Median	498.50	
	Variance	11996.46	
	Std. Deviation	109.53	
	Minimum	179.03	
	Maximum	805.10	
	Range	626.07	
	Interquartile Range	122.77	
Skewness	.08	.19	
Kurtosis	.36	.39	

Figure 1.75: Compare SAT scores and gender with the *Explore* function.

1.5.7.8 Crosstabs

Examine bivariate relationships through cross tabulation. This tool is useful for determining if one variable is dependent upon another, e.g., in a customer satisfaction survey you might discover a dependency of customer satisfaction levels on price.

Statistics and Measures of Association

Chi-Square Use Pearson's Chi-Square to test if the frequency in each cell is at least five; when the frequency is less than five and the contingency table is 2 x 2 in size then Fisher's

exact test is performed. When both variables are quantitative this function also returns a linear-by-linear association test.

Correlations Return Spearmans correlation coefficient and rho for numeric data when the values in the rows and columns are ordered. When both variables are quantitative returns the Pearsons R.

Nominal Data Measure the degree of association with several coefficients.

Phi Measure of association that returns the square root of the chi-square coefficient divided by the sample size. Typically, apply this to 2 x 2 tables although it can handle larger sizes. For 2 x 2 tables phi is identical to the correlation coefficient, but for larger ones the interpretation can be more complex.

Contingency Coefficient An adjustment to phi in order to make it suitable for tables larger than 2 x 2. The coefficient is equal to the square root of chi-square divided by chi-square plus the sample size. Note: For tables smaller than 5-by-5 this underestimates the level of association.

Cramer's V This measure of association gives accurate results when the number of rows and columns are equal, independent of size. V is equal to the square root of chi-square divided by the product of the sample size and the length of the smallest row or column.

Lamda Returns the percent by which you reduce the error when estimating the value of the dependent variable based upon on the value of the independent one. A lambda of zero means that there is no advantage in using the distribution of the independent variable to estimate the value of the dependent variable.

Uncertainty Coefficient Otherwise known as the entropy coefficient, is the percent reduction in error accounting for the variance in the dependent variable, where variance is defined in terms of entropy. When the uncertainty coefficient is zero, the independent variable is not helpful in predicting the dependent variable.

Ordinal Data

Gamma Returns the strength of association for ordinal data, ignoring tied (cases have equal values on both variables) pairs. Gamma is defined as the number of concordant pairs minus the number discordant ones over the sum of each type of pair. Cases form a concordant pair when the ranking agrees; meaning one case has a higher value than another case on both variables. When one case has a higher value on one variable, but a lower value on another the ranking disagrees and the cases form a discordant pair. Values of gamma range from -1 to 1. A value of 1 indicates perfect agreement between

the rankings, while a value of -1 means there is perfect disagreement. When gamma is equal to zero then there is no association between cases.

Note: Since gamma ignores ties, the strength of association between two variables might be over-exaggerated.

Somers'd Modification to Gamma that accounts for pairs tied on the dependent variable; values range from -1 for negative relationships to +1 for positive ones.

Kendall's Tau B A modification to gamma that takes into account tied pairs tied on both the dependent and independent variable. Kendall's Tau-B is symmetrical in that the same value is produced regardless of which variable is considered independent. Tau-b can only reach its maximum or minimum values of 1 and -1 for square tables.

Kendall's Tau C This is a variant of Kendall's Tau-B that can reach the maximum and minimum values of +1 and -1 for non-square tables.

Nominal by Interval

Eta Gauge of association when the dependent variable is a quantitative measure and the independent variable is categorical. Eta ranges from 0 to 1, with increasing values indicating a higher degree of association.

Cohen's Kappa Kappa is a measure of the degree to which two raters agree over their sorting of items into mutually exclusive categories, and takes into account agreement that occurs by chance. Values range from -1 to +1 and represent the proportion of agreements between raters after adjusting for those that take place by chance. A value of +1 indicates perfect agreement and -1 perfect disagreement. When kappa is zero, then there is no relationship between the raters and any agreements or disagreements are due to chance only.

Estimated Relative Risk The observed risk of having the outcome of interest be greater due to a specific factor. Use the odds ratio as an estimate or measure of risk if the occurrence of the factor is rare. A confidence interval greater than one indicates there is a greater chance of the event occurring in the presence of the factor. Note: this is only suitable for 2 x 2 tables.

Example

Consider a medical study examining the relative risk of patients having cerebella toxicity who have taken a generic chemotherapeutic agent rather than a non-generic one.

	Toxicity+	Toxicity-
Generic+	11	14
Generic-	3	31

Toxicity Versus Chemotherapeutic Agent

Enter the data from table x into a PSPP data sheet as follows in figure x. below , where +1 represents positive (+) and 0 negative (-):

	Freq	Toxicity	Generic
1	11.00	1	1
2	3.00	1	0
3	14.00	0	1
4	31.00	0	0
5			

Figure 1.76: Toxicity data in the data editor

Weight cases by selecting Data->Weight Cases, and then choose the *Weight Cases By:* option (see figure x).

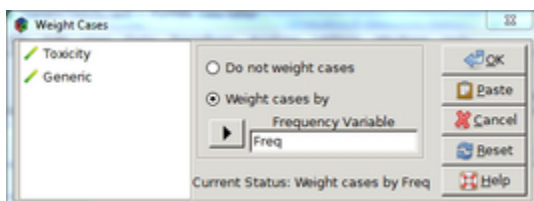


Figure 1.77: *Weight cases* function

Select Analyze-> Descriptive Statistics -> Crosstabs to open the *Crosstabs* dialogue box.

Move the variable *toxicity* from the list on the left side of the dialogue box into *Rows* and *Generic* into *Columns*.

Click *Statistics* to bring up the *Statistics* dialogue box.

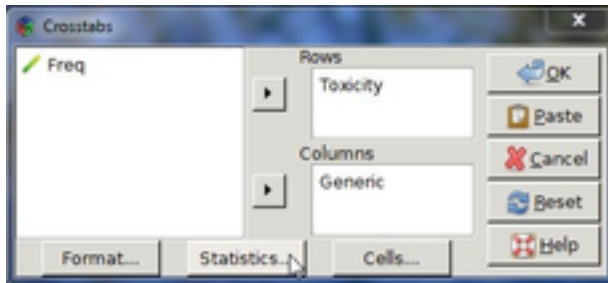


Figure 1.78: *Crosstabs* dialogue box.

Select the *Risk* check box and then click *Continue* to close this dialogue box.

Click OK on the *Crosstabs* dialogue box to display the results in the output viewer.

Output — SPSS Output Viewer

WEIGHT BY Freq
CROSSTABS
CROSSTABS
CROSSTABS
/TABLES=Toxicity BY Generic
/FORMAT=AVALUE LABELS TABLES PIVOT
/STATISTICS=CHISQ RISK
/CELLS=COUNT ROW COLUMN TOTAL

		Cases					
		Valid	Missing	Total			
		N/Percent	N/Percent	N/Percent	N/Percent		
Toxicity * Generic		59.00	100.0%	.00	0.0%	59.00	100.0%

Toxicity	Generic		Total
	0	1	
0	31.0 68.9%	14.0 31.1%	45.0 100.0%
1	9.2% 56.0%	76.3% 52.5%	76.3%
1	3.0 21.4%	11.0 78.6%	14.0 100.0%
	8.8% 5.3%	44.0% 18.6%	23.7% 23.7%
Total	34.0 100.0%	25.0 100.0%	59.0 100.0%
	57.6% 37.6%	42.4% 42.4%	100.0% 100.0%

Statistic	Value	d.f.	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	9.851	1.00	.00		
Likelihood Ratio	10.071	1.00	.00		
Fisher's Exact Test				1.00	.00
Continuity Correction	8.001	1.00	.00		
Linear-by-Linear Association	9.681	1.00	.00		
N of Valid Cases	59.00				

Statistic	Value	95% Confidence Interval	
		Lower	Upper
Odds Ratio for Generic (0 / 1)	8.12	1.95	33.73
For cohort Toxicity = 0	3.21	1.16	8.93
For cohort Toxicity = 1	.40	.24	.66
N of Valid Cases	59.00		

Figure 1.79: *Crosstab* statistics results in output viewer.

1.5.7.10 Compare Means

1.5.7.11 One-Sample T-Test

Tests if the mean of a single variable differs from a specified constant.

Compare the mean of the measured lifetime of 100 soft white light bulbs to that claimed by the manufacturer (700 hours).

Open the file LightBulbHours.sav.

Select Analyze-> Compare Means -> One Sample T Test.

Select the *hour* variable from the list on the left side of the dialogue box and then click the right arrow to move it into the *Test Variables* textbox.

Enter the value 700 into the *Test Value* textbox.

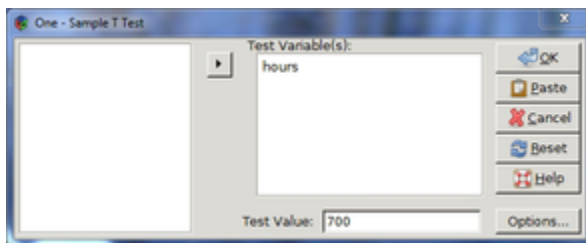


Figure 1.80: *One-Sample T Test* dialogue box

Click OK to view the results in the output window.

	N	Mean	Std. Deviation	S.E. Mean
hours	100	699.94	35.82	3.58

Test Value = 700.000000						
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
hours	-0.293	99	.771	-.06	-7.17	7.05

Figure 1.81: *One-Sample T Test* output results

1.5.7.12 Independent Samples T Test

Compare means of the same variable taken from two independent samples. Although the test requires normally distributed data, minor departures from this assumption are tolerable, and there is no requirement that the samples are the same size.

Example

Compare mean heights, in centimeters, of ten twelve-year-old boys to ten twelve-year-old girls.

Enter the following *height* and *gender* data into PSPP, where 0 represents female and 1 male.

152.71	0
142.57	0
155.48	0
148.79	0
167.74	0
149.51	0
142.00	0
150.21	0
171.19	0
141.99	0
152.67	1
139.65	1
152.19	1
156.94	1
151.86	1
157.79	1
152.49	1
157.52	1
156.95	1
141.61	1

Name the first variable *Height* and the second *Gender*.

Select Analyze-> Compare Means -> Independent Samples T Test.

Select *Height* from the list on the left side of the dialogue box and then click the upper right arrow to move it into the *Test Variables* box.

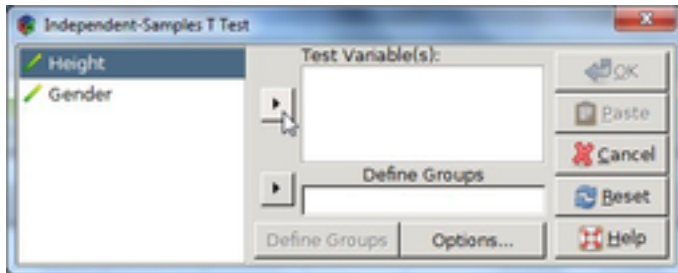


Figure 1.82: *Independent-Samples T Test* dialogue box

Select *Gender* and click the lower right arrow to move the variable into the *Define Groups* area.

Click the *Define Groups* button at the bottom of the dialogue box.

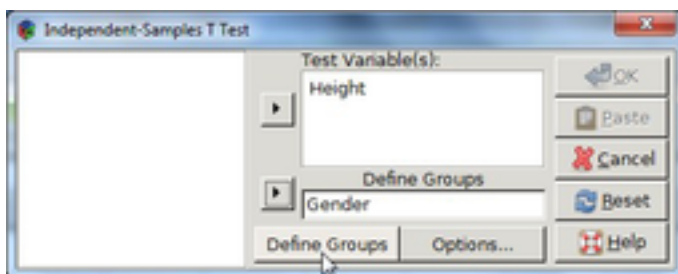


Figure 1.83: Define variables and groups in the *Independent-Samples T Test* dialogue.

Enter 0 into the *Group1 Value* textbox of the *Define Groups* dialogue box.

Enter 1 into the *Group2 Value* textbox of the *Define Groups* dialogue box.

Select *Continue*.

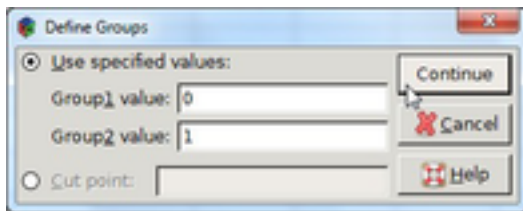


Figure 1.84: Organize data into groups.

Click OK on the *Independent-Samples T Test* dialogue box to examine the results in the output viewer.

Gender	N	Mean	Std. Deviation	S.E. Mean
Height0	10	152.22	18.20	5.73
1	10	151.97	6.45	2.04

	Levene's Test for Equality of Variances		t-test for Equality of Means						
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
								Lower	Upper
Height	1.66	.21	0718.90	95	.25	3.82	-7.77	8.27	
			0713.20	95	.25	3.82	-7.87	8.38	

Figure 1.85: Results of the *Independent-Samples T Test* in the output viewer.

1.5.7.14 Paired Samples T Test

Compare means of two sets of quantitative data, where data sets are organized into pairs with a definite relationship, e.g., one set of data might represent a measurement taken before some event and the other a measurement after. Each pair is then the measurement before coupled with the measurement after.

Compare the means of the following measurements (figure x) from nine asthmatics of their peak expiratory flow rate before and after a walk on a cold winters day.

	After	Before
1	302.00	314.00
2	199.00	240.00
3	230.00	338.00
4	314.00	390.00
5	220.00	296.00
6	258.00	256.00
7	328.00	391.00
8	332.00	404.00
9	231.00	290.00
10		

Figure 1.86: Before and after data

Select Analyze->Compare Means-> Paired Samples T Test.

Highlight the *Before* variable with the mouse and then click the right arrow to place it in the *Test Variable(s)* dialogue box under *Var1*.

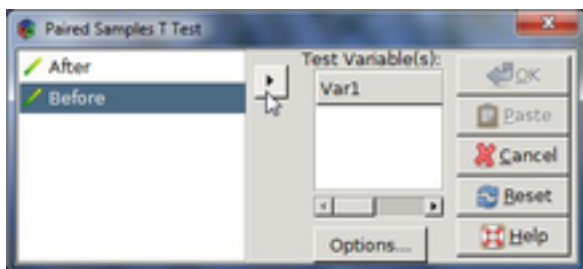


Figure 1.87: *Paired Samples T Test* dialogue box.

Repeat the same steps with the *After* variable to place it under *Var2*.

sp 1

Click OK to display the results in the Output window.

	Mean	N	Std. Deviation	Std. Error	Mean
Pair 0 Before	324.333	9	60.37	20.12	
After	268.222	9	51.17	17.06	

	N	Correlation	Sig.
Pair 0 Before & After	9	.82	.01

	Paired Differences							
	Mean	Std. Deviation	Std. Error	Mean	95% Confidence Interval of the Difference	t	Sig.	Detailed
Pair 0 Before - After	56.11	34.17	11.39	29.84	82.38	4.93	.00	

Figure 1.88: *Paired Samples T Test* output results.

1.5.7.15 One-Way Anova

Indicates if there are significant differences between means of three or more independent groups, where the independent variable consists of two or more categorical independent groups and the dependent one is either interval or ratio. Note: this only tells you if the means of the groups are significantly different and not which ones.

Optional Statistics include:

Descriptives (mean, standard deviation, standard error of the mean, minimum, maximum, and 95% confidence intervals for each dependent variable)

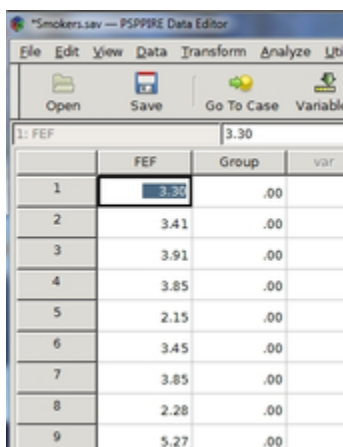
Homogeneity of variance test that calculates the Levene statistic to look for the equality of variances

Specify contrasts the you want tested by the t statistic with the *Contrasts* button in the dialogue box.

Example

Determine if the mean of each smoker group's lung capacity varies, where the different groups include non-smoker, light-smoker, medium-smoker, and heavy-smoker.

Open the data file XXX.sav. This lists forced expiratory flow measurements under the variable *FEF* and the corresponding group type (non-smoker=0, light-smoker=1, medium-smoker=2, heavy-smoker=3) under *Groups*.



	FEF	Group	var
1	3.30	.00	
2	3.41	.00	
3	3.91	.00	
4	3.85	.00	
5	2.15	.00	
6	3.45	.00	
7	3.85	.00	
8	2.28	.00	
9	5.27	.00	

Figure 1.89: Data describing the lung capacity of several groups of smokers.

Select Analyze-> Compare Means-> One-Way ANOVA

Click on *FEF* from the list of variables on the left side of the dialogue box then select the upper right arrow to move it into the *Dependent Variables* textbox.

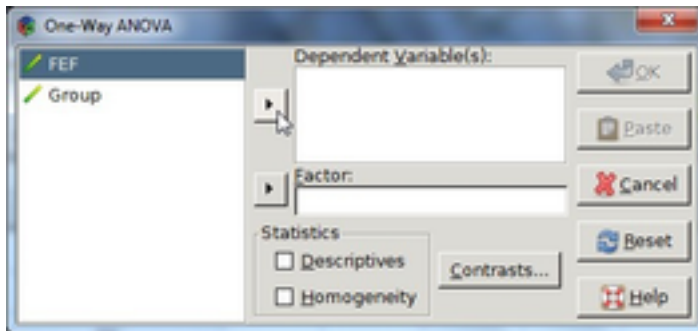


Figure 1.90: *One-Way ANOVA* dialogue box

Select the *Group* variable and click the lower right arrow button to place it inside the *Factor* textbox.

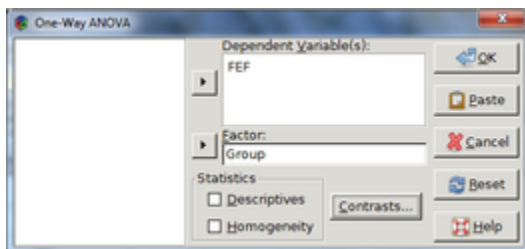


Figure 1.91: Select dependent and independent variables in the *One-Way ANOVA* dialogue.

Click OK to displays the results in the output viewer.

ONEWAY					
ONEWAY VARIABLES= FEF BY Group					
	Sum of Squares	df	Mean Square	F	Significance
FEF(Between Groups)	173.25	3	57.75	57.769416	.00
Within Groups	488.49799	41	11.914585		
Total	661.84799	44			

Figure 1.92: *One-Way ANOVA* output

Contrast

Contrast is a weighted sum of the means and enables you to test your research hypothesis as opposed to the statistical one.

Example

Check if the mean lung capacity of non-smokers varies from that of smokers, independent of how many they smoke. There are four different groups, of which three are smokers of different intensities. This means the non-smoking group should have a mean with a weighting of unity and the others a weighting of one-third.

Select *Contrasts* on the *One-Way ANOVA* dialogue box.

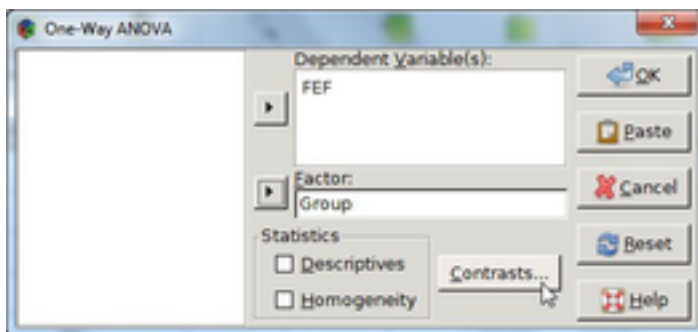


Figure 1.93: Use *contrasts* button to calculate weighted sum of the means.

Enter 1.0 in the *Coefficients* textbox and then click the *Add* button.

Add the values -0.333, -0.333, and -0.334 to the *Coefficients* list.

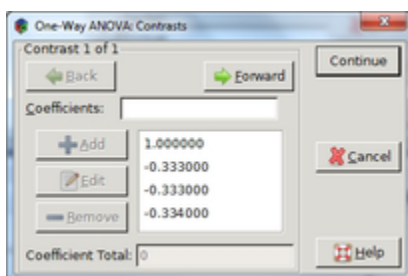


Figure 1.94: *Contrasts* dialogue box

Press *Continue* to close the *Contrasts* dialogue box.

Press OK on the *One-Way ANOVA* dialogue box to display the results in the output viewer.

Output — PSPSS Output Viewer

File Edit Windows Help

ONEWAY

ONEWAY VARIABLES= FEF BY Group
/CONTRAST= 1 -0.333 -0.333 -0.334

	Sum of Squares	df	Mean Square	F	Significance
FEF (Between Groups)	373.35	2	186.675	57.78	.00
Within Groups	488.49794	61	8.00816		
Total	661.84799	63			

Group		1	2	3
Contrast 1		1	-0.333	-0.333

	Contrast	Value of Contrast	Std. Error	t	df	Sig. (2-tailed)
FEF (Assume equal variances)	1	94	06.1465	7.94	61	.00
Does not assume equal	1	94	06.146534936	7.94	61	.00

Figure 1.95: *One-Way ANOVA* output

1.5.7.18 Bivariate Correlation

Computes Pearson's correlation coefficient, defined as a value between -1 and +1 that indicates the degree of association between variables. A positive value means large values of one variable are associated with large values of the other and, likewise, small values of one variable are associated with small values of the other. A negative value indicates that large values of one variable are associated with small values of the other and vice-versa.

Example

Use this feature to investigate the correlation, if any, between ice cream sales and outside temperature.

Open a new data file.

Enter the following pairs of values listed in table x of the number of ice cream cones sold each day one summer at a specified store along with the corresponding daytime high temperature.

Number of Cones Sold	Temperature (Fahrenheit)
171.00	76.00
161.00	71.00
156.00	66.00
151.00	61.00
154.00	73.00
143.00	61.00
146.00	64.00
157.00	69.00
171.00	78.00
171.00	78.00
166.00	76.00
168.00	76.00
176.00	81.00
181.00	83.00
181.00	86.00
144.00	60.00
174.00	87.00
177.00	89.00
178.00	86.00
186.00	91.00
164.00	75.00
161.00	74.00
163.00	72.00
149.00	69.00
161.00	71.00
155.00	74.00
170.00	86.00
179.00	91.00
176.00	71.00
165.00	82.00

Ice Cream Cones Sold Versus Outside Temperature

Select Analyze-> Bivariate Correlation, to bring up the *Bivariate Correlation* dialogue box.

Select *Cones* and click the right arrow button to move it into the right-hand-side textbox.

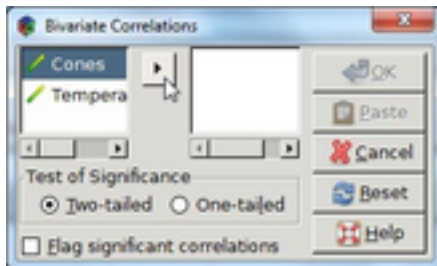


Figure 1.96: *Bivariate Correlations* dialogue box

Repeat the same steps with the *Temperature* variable.

Click OK to display the results in the output viewer.

Output — PSPPIRE Output Viewer

File Edit Windows Help

CORRELATION: CORRELATIONS
/VARIABLES = Cones Temperature
/PRINT = TWOTAIL, SIG.

		Cones	Temperature
Cones	Pearson Correlation	1.00	.89
	Sig. (2-tailed)		.00
	N	30	30
Temperature	Pearson Correlation	.89	1.00
	Sig. (2-tailed)	.00	
	N	30	30

CORRELATION: CORRELATIONS
/VARIABLES = Cones Temperature
/PRINT = TWOTAIL, SIG.

		Cones	Temperature
Cones	Pearson Correlation	1.00	.89
	Sig. (2-tailed)		.00
	N	30	30
Temperature	Pearson Correlation	.89	1.00
	Sig. (2-tailed)	.00	
	N	30	30

Figure 1.97: emphBivariate Correlations output

1.5.7.20 Factor Analysis

Helps identify the underlying variables, or factors, to explain correlations. Variables should be either interval or ratio, and not categorical.

Note: PSPP assumes the data has a nearly bivariate normal distribution for each pair of variables and that observations are independent.

Two main statistic functions are available: *Extraction* and *Rotation* (see fig x).

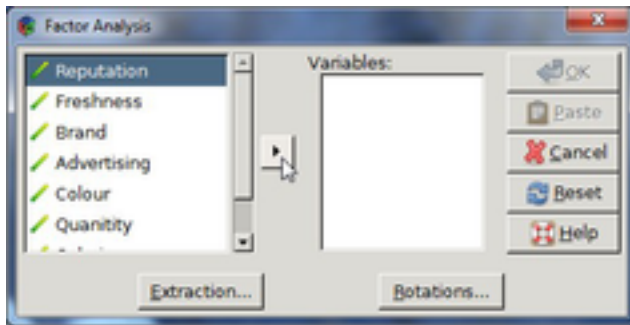


Figure 1.98: *Factor Analysis* dialogue box

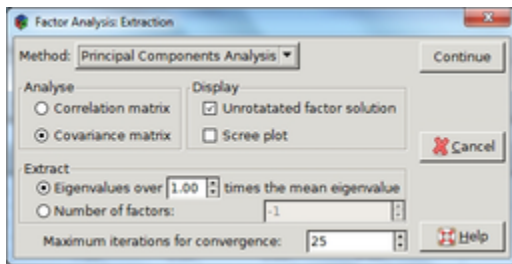


Figure 1.99: *Factor Analysis Extraction* dialogue box

Method The `emphMethod` button on the `emphExtraction` dialogue box enables you to choose the method of factor extraction. Options include `emphPrincipal Components Analysis` and `emphPrincipal Axis Factoring`.

Principal Components Analysis Use when the correlation matrix is singular. This function assembles uncorrelated linear combinations of the observed

variables such that each successive component accounts for a progressively smaller portion of the variance.

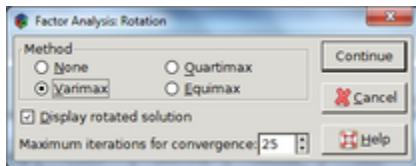
Principal Axis Factoring An iterative procedure that extracts and replaces factors from the current correlation matrix with new estimates repeatedly, until the convergence criterion for extraction is satisfied.

Analyze Choose either a correlation or covariance matrix.

Display Display the unrotated factor solution and Scree plot.

Unrotated Factor Solution Display the unrotated factor loadings, communalities, and the solution eigenvalues.

Scree Plot A plot of the variance for each factor.



whose eigenvalues exceed a specified factor above the mean number of factors to keep.

Convergence Specify the maximum number of steps the algorithm can take to estimate the solution.

Figure 1.100: *Factor Analysis Rotation* dialogue box

1.5.7.22 Rotation Dialogue Box

Method Choose the type of factor rotation method. Options include none, varimax, quartimax, and equimax.

Varimax An orthogonal rotation of the factor axes which has the effect of differentiating the original variables by the extracted factor. This solution

provides results that make it easier to identify each variable with a single factor.

Quartimax An orthogonal rotation that minimizes the number of factors needed to explain each variable.

Equimax A combination of the varimax and quartimax rotations.

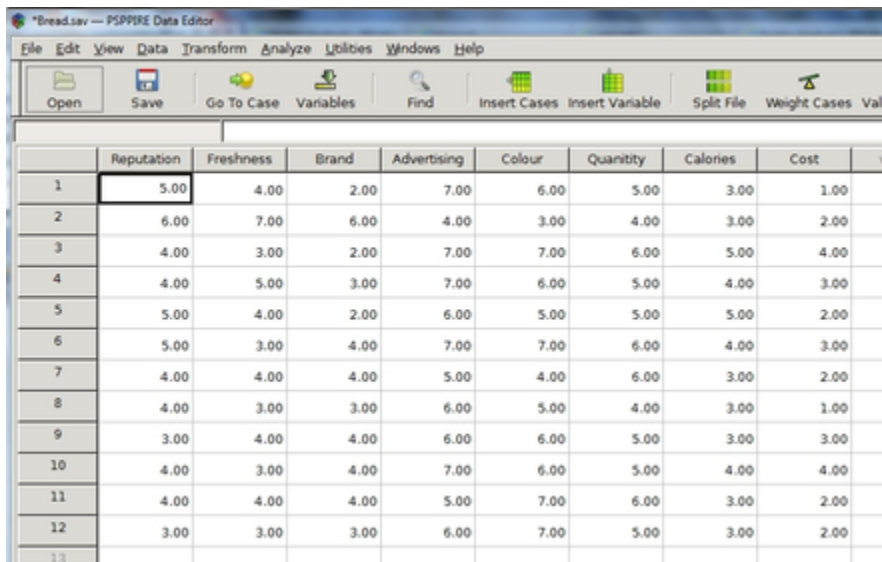
Display Rotated Solution Display the solution in this way if you have selected a rotation method.

Maximum Iterations for Convergence Choose the maximum number of steps the algorithm can take to reach convergence. The default value is twenty-five.

Example

Perform factor analysis to determine the leading factors that drive people to purchase specific brands of bread.

Enter the data in figure x, or open file xxx. sav.



	Reputation	Freshness	Brand	Advertising	Colour	Quantity	Calories	Cost	Value
1	5.00	4.00	2.00	7.00	6.00	5.00	3.00	1.00	
2	6.00	7.00	6.00	4.00	3.00	4.00	3.00	2.00	
3	4.00	3.00	2.00	7.00	7.00	6.00	5.00	4.00	
4	4.00	5.00	3.00	7.00	6.00	5.00	4.00	3.00	
5	5.00	4.00	2.00	6.00	5.00	5.00	5.00	2.00	
6	5.00	3.00	4.00	7.00	7.00	6.00	4.00	3.00	
7	4.00	4.00	4.00	5.00	4.00	6.00	3.00	2.00	
8	4.00	3.00	3.00	6.00	5.00	4.00	3.00	1.00	
9	3.00	4.00	4.00	6.00	6.00	5.00	3.00	3.00	
10	4.00	3.00	4.00	7.00	6.00	5.00	4.00	4.00	
11	4.00	4.00	4.00	5.00	7.00	6.00	3.00	2.00	
12	3.00	3.00	3.00	6.00	7.00	5.00	3.00	2.00	
13									

Figure 1.101: Survey data describing influential factors in purchasing decisions.

Select Analyze-> Factor Analysis to bring up *Factor Analysis* dialogue box.

Click the first variable, *Reputation*, and then press the right arrow button to move it into the *Variables* dialogue box.

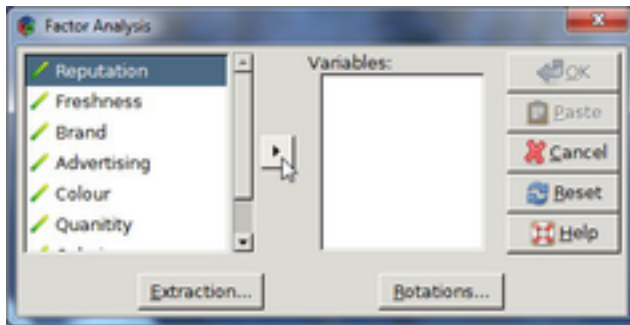


Figure 1.102: Pick variables for *Factor Analysis*.

Repeat with all the other variables until no more remain on the left side.

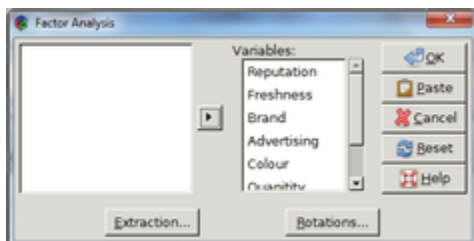


Figure 1.103: All variables are selected for *Factor Analysis*.

Click the *Extraction* button.

Select the *Principal Components Analysis* method.

Select the *Correlation matrix* option under *Analyse*.

Select *Unrotated factor solution* and *Scree plot* under *Display*.

Click *Continue* to close this dialogue box.

Select the *Rotations* button in the *Factor Analysis* dialogue box.

Select the *Varimax* option under *Method*.

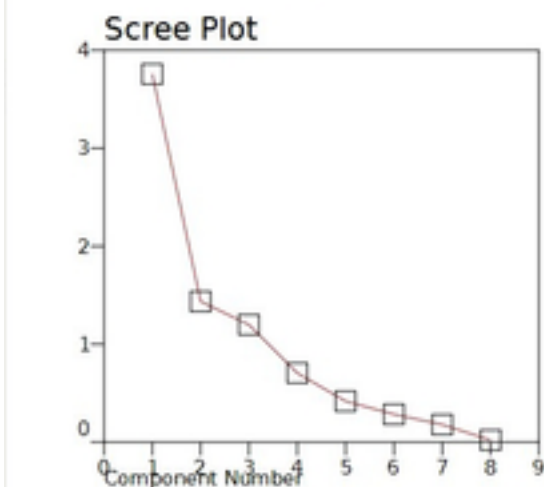
Select *Display rotated solution*.

Click *Continue* to close the dialogue box.

Click the OK button to display the results in the output viewer.

	Initial	Extraction
Reputation	1.00	.78
Freshness	1.00	.82
Brand	1.00	.90
Advertising	1.00	.79
Colour	1.00	.79
Quantity	1.00	.56
Calories	1.00	.91
Cost	1.00	.87

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.76	46.96	46.96	3.76	46.96	46.96	2.09	26.16	26.16
2	1.44	17.96	64.92	1.44	17.96	64.92	1.90	23.79	49.95
3	1.20	14.97	79.89	1.20	14.97	79.89	2.40	29.95	79.89
4	.71	8.82	88.71						
5	.42	5.23	93.94						
6	.28	3.56	97.49						
7	.18	2.24	99.73						
8	.02	.27	100.00						



	Component		
	1	2	3
Reputation	.52	-.64	.32
Freshness	.82	-.35	-.07
Brand	.68	-.08	-.65
Advertising	-.83	.18	.29
Colour	-.85	-.22	-.14
Quantity	-.59	-.09	-.46
Calories	.56	.75	.18
Cost	-.53	-.52	-.57

	Component		
	1	2	3
Reputation	.52	-.64	.32
Freshness	.82	-.35	-.07
Brand	.68	-.08	-.65
Advertising	-.83	.18	.29
Colour	-.85	-.22	-.14
Quantity	-.59	-.09	-.46
Calories	.56	.75	.18
Cost	-.53	-.52	-.57

*****two output figs in one pic above

1.5.7.24 Reliability

Estimate the internal consistency reliability. Options include Split-half reliability and Chronbach's Alpha.

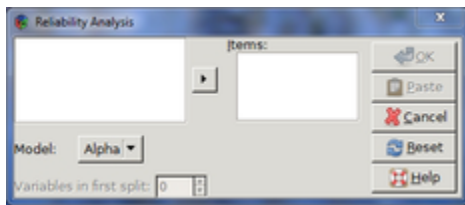


Figure 1.105: *Reliability Analysis* dialogue box

1.5.7.25 Split-half Reliability

An estimation based on the correlation of two equivalent forms of the scale. This is a three-step process: 1) Randomly divide all items that measure the same construct into two sets, 2) administer the whole instrument and divide the total score for each item in half, and 3) estimate the split-half reliability by calculating the correlation of these two sets.

1.5.7.26 Chronbach's Alpha

This estimation is equivalent to the average of all possible split-half estimates, and is an indication of the correlation among the variables.

Example

Examine the internal reliability of the following pass-fail test data, where zero represents fail and one pass.

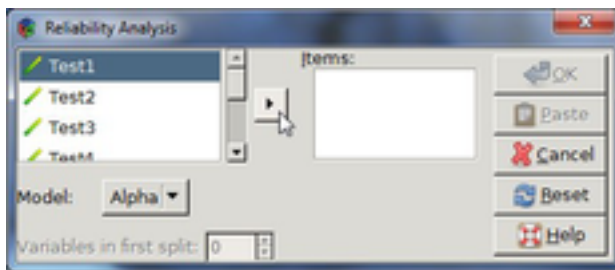
	Test1	Test2	Test3	Test4	Test5	Test6	Test7	Test8
1	.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
3	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
4	1.00	.00	1.00	1.00	.00	1.00	1.00	1.00
5	1.00	.00	.00	1.00	1.00	1.00	1.00	1.00
6	.00	1.00	1.00	.00	.00	1.00	1.00	1.00
7	.00	1.00	.00	.00	1.00	1.00	1.00	1.00
8	1.00	.00	.00	.00	1.00	.00	1.00	1.00
9	.00	.00	.00	1.00	.00	1.00	1.00	1.00
10	.00	.00	.00	1.00	1.00	.00	.00	.00
11								

Figure 1.106: Fail-pass test data

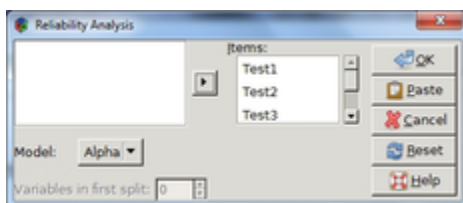
Enter the test data into a new data file, or open the file xxx.sav.

Select Analyze-> Reliability.

Click the first variable and select the right arrow to move it into the *epphItems* textbox.

Figure 1.107: Pick variables for *Reliability Analysis*

Repeat with the other variables until they are all listed in the *Items* text box.

Figure 1.108: All variables are chosen for *Reliability Analysis*.

Click OK to display the results. In this instance, *Cronbachs Alpha* is 0.61 indicating the internal consistency is questionable.

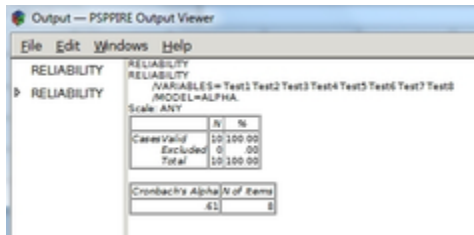


Figure 1.109: *Reliability Analysis* output

1.5.7.28 Non-Parametric Statistics

Provides Binomial statistics and Chi-Square for an arbitrary distribution of data. Tests the differences between the observed and expected distribution of a sample, and can also test the difference between two or more samples.

1.5.7.29 Binomial

Compare the observed frequencies of dichotomous data to the expected values from the binomial distribution. The primary purpose of this test is to observe if the number of outcomes for each variable differs, e.g., you can verify a fair coin by comparing the frequency of heads and tails in several coin tosses.

Example

The following list of *HotCold* and *Range* data indicates the temperatures at which customers ordered their gourmet drinks on a midsummers day; available menu options include ice-cold, luke-warm, warm, or steaming-hot. The first column, *emphHotCold* is a dichotomous variable with a value of zero for ice-cold drinks and one otherwise. The next

column, `emphRange` varies from 0 to 3 depending upon temperature range of the drink, where 0=cold, 1=luke-warm, 2=warm, and 3= hot.

HotCold	Range
0	0
0	0
1	2
1	2
0	0
1	1
1	1
1	1
1	2
1	1
1	3
1	3
1	2
0	0
0	0
0	0
1	1
0	0
0	0
1	2
1	2
0	0
1	3
0	0

Gourmet Drink Temperatures

Start by examining the data in the first column; *HotCold*.

Copy table x.x into a data file, or open xxx.sav.

Select Analyze->Non-Parametric Statistics->Binomial.

Select *HotCold* and click the right arrow button to place it in the *Test Variable List* textbox.

Verify *Test proportion* has the value 0.5.

Verify *Get from data* is selected under *Define Dichotomy*.

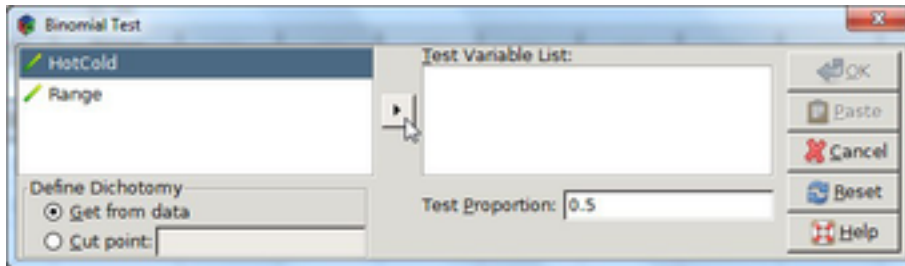


Figure 1.110: *Binomial* dialogue box

Click OK to display the results in the output viewer.

```

NPAR TESTS
  /BINOMIAL(0.5) = HotCold

```

	Category	N	Observed Prop.	Test Prop.	Exact Sig. (2-tailed)
HotCold Group2	00	11	.45	.50	.69
HotCold Group2	100	14	.56	.50	
Total		25	1.00		

Figure 1.111: *Binomial Test* output

1.5.7.31 Chi-Square

Used to check if distributions of categorical variables differ, where categorical variables are used to label categorical data, e.g. assign a value of 1 for values in range a to c, value of 2 for values in range d to f, and so on.

Example

Enter the data in table x from the previous section(binomial) into a new data file.

Select Analyze->Non-Parametric Statistics->Chi-Square.

Select the *Range* variable, and then click the right arrow to place it into the *Test Variable List* textbox.

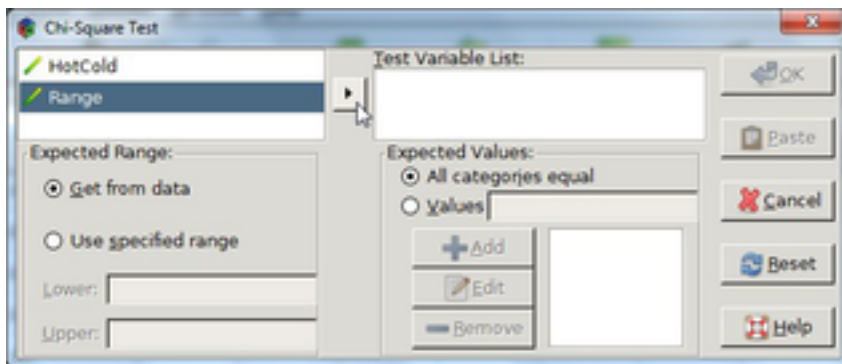


Figure 1.112: *Chi-Square Test* dialogue box

Verify that *All categories equal* is selected under *Expected Values*.

Verify that the *Get from data* option under *Expected Range* is selected.

Click OK to display the results in the output viewer.

NPAR TEST				
NPAR TESTS				
/CHISQUARE = Range				
	Observed N	Expected N	Residual	
.00	11	6.25	4.75	
1.00	5	6.25	-1.25	
2.00	6	6.25	-.25	
3.00	3	6.25	-3.25	
Total	25			

	Range
Chi-Square	3.56
df	3
Asymp. Sig.	.14

Figure 1.113: *Chi-Square Test* output

1.5.7.33 ROC Curve

Use to analyze binary classification systems. Examples of binary classification schemes include passing or failing newly made items during an inspection, and labeling patients to be either with or without a medical problem. The Roc Curve function returns an ROC curve with the option to display a diagonal reference line, standard error and confidence interval, as well as all coordinate points.

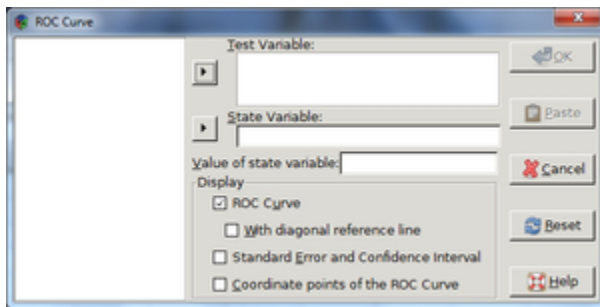


Figure 1.114: *ROC Curve* dialogue box

1.5.7.34 Test Variable

Continuous numerical data holding the test results.

1.5.7.35 State Variable

Indicates which category is positive.

1.5.7.36 Display

Gives options for the output viewer including:

ROC curve with or without a diagonal reference line

Display standard error and confidence interval

Display coordinate points of the ROC curve

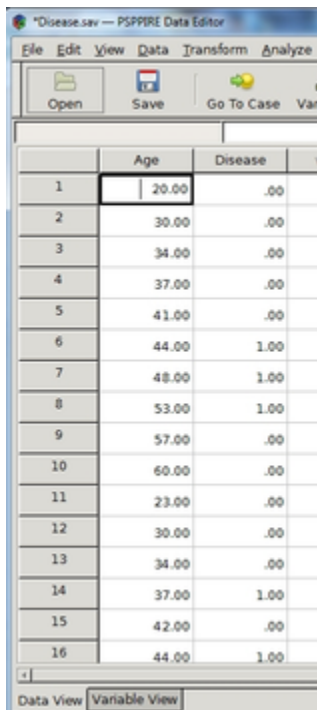
Display ROC Curve Produce a plot of one classification rate from a binary classification scheme versus the other as the criterion changes, e.g., you can plot the true positive rate versus the false positive rate for a disease as the threshold value between the two classifications varies.

Display Diagonal Reference Line Include a diagonal reference line in the plot. An ROC curve that coincides with this line indicates the distribution of each classification group is similar.

Display Standard Error and Confidence Interval Show the standard error, confidence interval, and area in a separate table in the output viewer. The area ranges in value from 0 to 1, where 1 indicates the distribution of each classification group does not overlap and 0.5 means both distributions are the same.

Example

Make an ROC curve of the following medical test data describing the dependence of a disease with age. The *age* variable holds the patients ages and the corresponding *disease* column is either 0 indicating a negative diagnosis or 1 otherwise.



The screenshot shows the SPSS Data Editor window titled "Disease.sav - PSPPRE Data Editor". The menu bar includes File, Edit, View, Data, Transform, and Analyze. The toolbar contains icons for Open, Save, Go To Case, and Variable View. The main data grid displays 16 rows of data with columns for Case Number, Age, and Disease. The Disease column contains binary values (.00 or 1.00). At the bottom, there are tabs for "Data View" and "Variable View".

	Age	Disease
1	20.00	.00
2	30.00	.00
3	34.00	.00
4	37.00	.00
5	41.00	.00
6	44.00	1.00
7	48.00	1.00
8	53.00	1.00
9	57.00	.00
10	60.00	.00
11	23.00	.00
12	30.00	.00
13	34.00	.00
14	37.00	1.00
15	42.00	.00
16	44.00	1.00

Figure 1.115: Medical test data

Open the file xxx.sav.

Select Analyze-> ROC Curve.

Select *Age* from the list of variables on the left side of the dialogue box, and click the upper right arrow button to move it under *Test Variable*.

Select *Disease* and click the lower right arrow button to move it into the *State Variable* box.

Select all the check boxes under *Display*.

Enter the number 1 in the *Value of state variable* box.

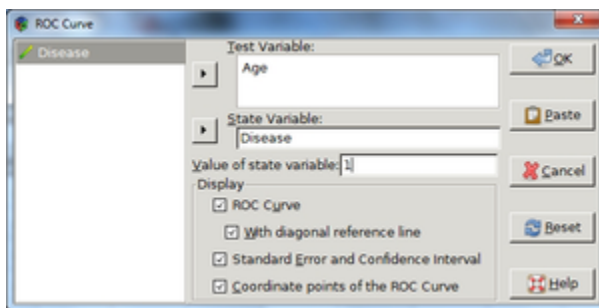


Figure 1.116: *ROC Curve* dialogue box

Click OK to display the ROC curve and accompanying tables in the output viewer.

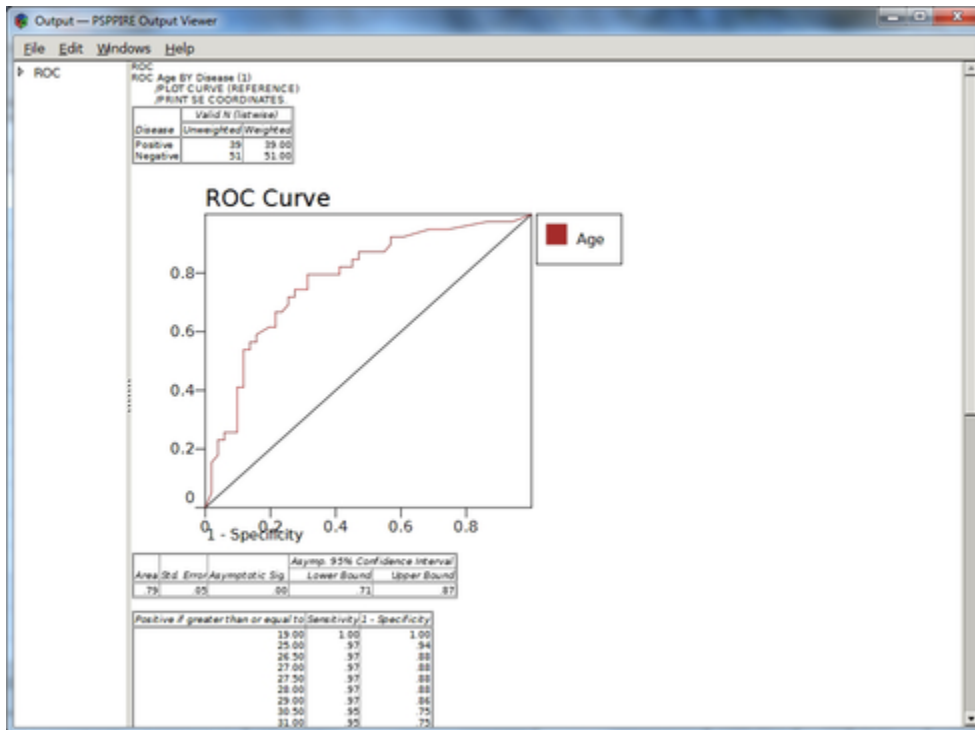


Figure 1.117: *ROC Curve* output

1.5.8 Utilities

1.5.8.1 Variables

View variable properties and/or go-to it in the data editor window. Properties include variable name and label, Data format, user-missing values, measurement level, and value labels. For a definition of each variable property, refer to [DummyNode], page 1.

Example

Select Utilities-> Variables to bring up the *Variables* dialogue box.

Select a variable from the list in the left column to view its properties on the right side of the dialogue box.

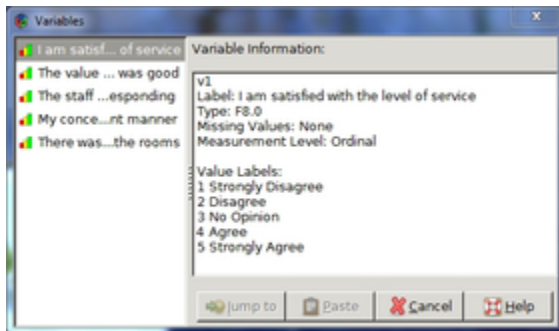


Figure 1.118: View variable properties in *Variables* dialogue box.

1.5.8.3 Data File Comments

This feature enables you to add comments to a data file.

1.5.8.4 Display Comments in Output

Display comments in the output viewer.

1.5.8.5 Reset

Clears the current comment.

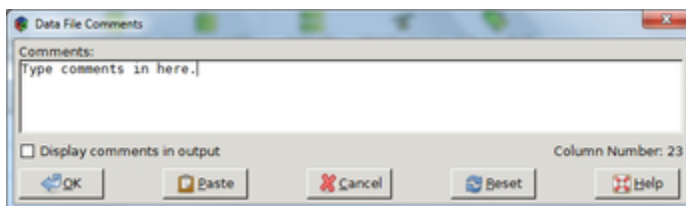


Figure 1.119: *Comments* dialogue box

1.5.9 Windows

1.5.9.1 Minimize all Windows

Collapse all windows. Click the PSPP icon on the tool bar to open them again.

1.5.9.2 Split

Split the current data editor window (see figure x).

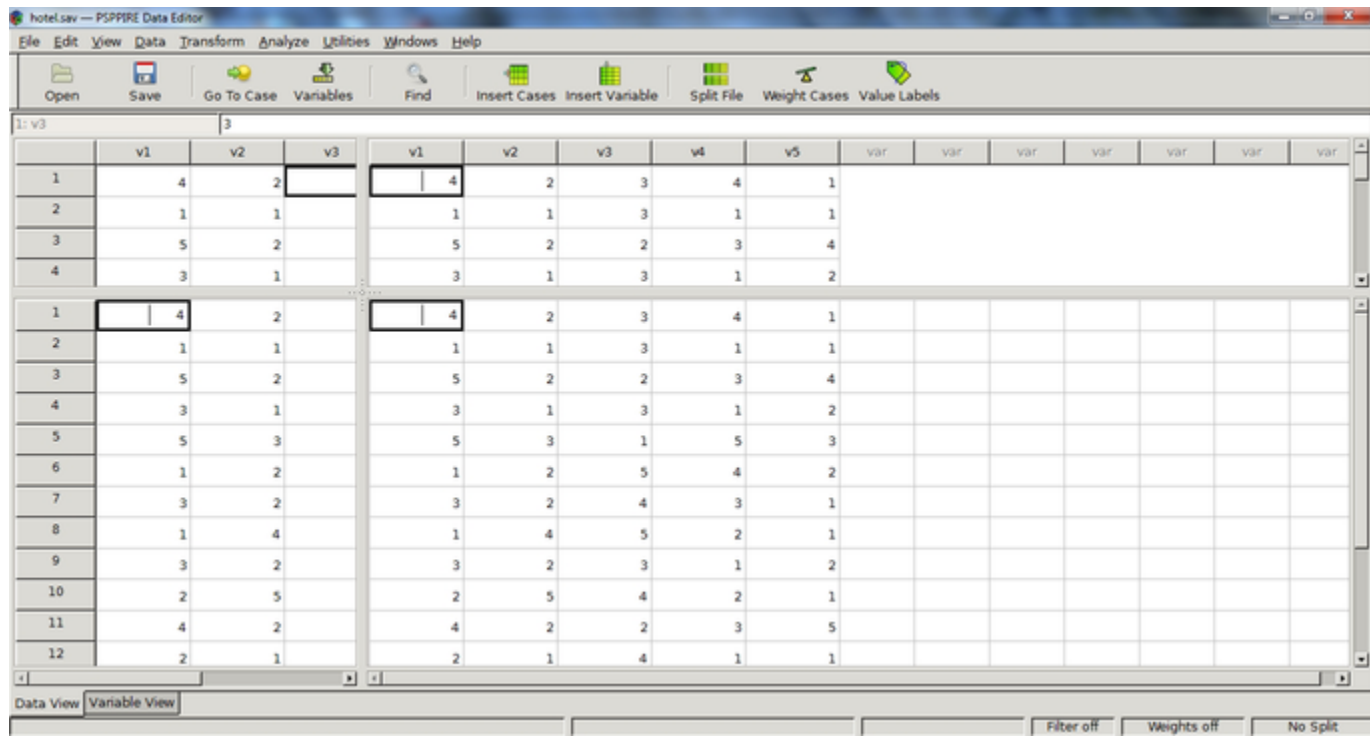


Figure 1.120: Split data-editor window

1.5.10 Toolbar

SPSS provides a toolbar across the top with short cuts to more commonly used menu commands.

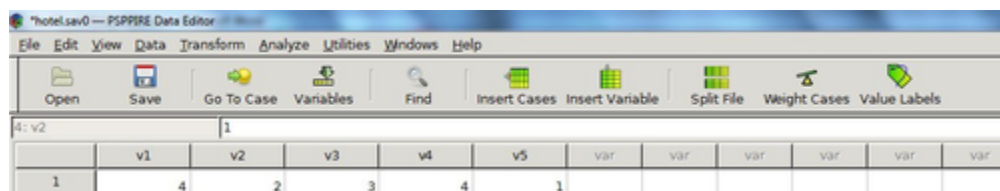


Figure 1.121: Data editor toolbar