

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	Your rights and obligations	1
2	Graphical User Interface	2
	DummyNode	2
2.1	Primary Windows	2
2.2	Data Editor	2
2.2.1	Data View	2
2.2.2	Variable View	3
2.2.3	Data Entry and Editing	3
2.2.3.1	New Variable	3
2.2.3.2	Edit Variable Attributes	3
2.2.3.3	New Case	5
2.2.3.4	Read Variable in Data View	5
2.2.3.5	Change Data Value	5
2.2.3.6	Select Entire Row or Column	5
2.2.3.7	Clear Variables	5
2.2.3.8	Clear Cases	6
2.2.3.9	Sort Cases	6
2.2.4	Status Bar	6
2.3	Output Viewer	7
	Examples	7
	Example	7
2.3.3	Save Output Contents	7
2.4	Syntax Editor	7
2.4.1	Enter Commands	7
	Example	7
2.4.2	Run	8
2.4.3	Save	8
2.5	Main Menu	8
2.5.1	Overview	8
2.5.2	File	9
2.5.2.1	New	9
2.5.2.2	Open	9
2.5.2.3	Import Delimited Text Data	9
2.5.2.4	Save	10
2.5.2.5	Save As	10
2.5.2.6	Display Data File Information	10
2.5.2.7	Recently Used Data	10
2.5.2.8	Recently Used Files	11
2.5.2.9	Quit	11
2.5.3	Edit	11
2.5.3.1	Insert Variable	11
	Example	11

2.5.3.3	Insert Cases	11
2.5.3.4	Go To Case	11
2.5.3.5	Cut	11
2.5.3.6	Copy	11
2.5.3.7	Paste	11
2.5.3.8	Find	11
2.5.4	View	11
2.5.4.1	Status Bar	11
2.5.4.2	Font	11
2.5.4.3	Grid Lines	11
2.5.4.4	Value Labels	12
	Value Labels	12
2.5.4.6	Variables	12
2.5.5	Data	12
2.5.5.1	Sort Cases	12
2.5.5.2	Transpose	12
2.5.5.3	Split File	12
	Compare Groups	13
	Organize Output by Groups	13
	Example	13
2.5.5.7	Weight Cases	13
2.5.6	Transform	13
2.5.6.1	Compute	13
	Example	13
2.5.6.3	Rank Cases	15
	Rank Types	15
2.5.6.5	Recode into Same Variables	16
	Example	16
	Example	18
2.5.6.8	Recode into Different Variables	19
2.5.7	Analyze	19
2.5.7.1	Descriptive Statistics Definitions	19
2.5.7.2	Frequencies	20
	Example	21
2.5.7.4	Descriptives	21
	Example	22
2.5.7.6	Explore	22
	Example	22
2.5.7.8	Crosstabs	23
	Statistics and Measures of Association	24
2.5.7.10	Compare Means	26
2.5.7.11	One-Sample T-Test	26
2.5.7.12	Independent Samples T Test	27
	Example	28
2.5.7.14	Paired Samples T Test	29
2.5.7.15	One-Way Anova	29
	Example	30
	Example	30

2.5.7.18	Bivariate Correlation	31
Example		32
2.5.7.20	Factor Analysis	33
2.5.7.21	Extraction	33
2.5.7.22	Rotation Dialogue Box	34
Example		35
2.5.7.24	Reliability	36
2.5.7.25	Split-half Reliability	36
2.5.7.26	Chronbach's Alpha	36
Example		36
2.5.7.28	Non-Parametric Statistics	37
2.5.7.29	Binomial	37
Example		37
2.5.7.31	Chi-Square	39
Example		39
2.5.7.33	ROC Curve	39
2.5.7.34	Test Variable	39
2.5.7.35	State Variable	39
2.5.7.36	Display	39
Example		40
2.5.8	Utilities	40
2.5.8.1	Variables	41
Example		41
2.5.8.3	Data File Comments	41
2.5.8.4	Display Comments in Output	41
2.5.8.5	Reset	41
2.5.9	Windows	41
2.5.9.1	Minimize all Windows	41
2.5.9.2	Split	41
2.5.10	Toolbar	41

Appendix A	GNU Free Documentation License	42
.....		
A.1	ADDENDUM: How to use this License for your documents....	48

1 Your rights and obligations

PSPP is not in the public domain. It is copyrighted and there are restrictions on its distribution, but these restrictions are designed to permit everything that a good cooperating citizen would want to do. What is not allowed is to try to prevent others from further sharing any version of this program that they might get from you.

Specifically, we want to make sure that you have the right to give away copies of PSPP, that you receive source code or else can get it if you want it, that you can change these programs or use pieces of them in new free programs, and that you know you can do these things.

To make sure that everyone has such rights, we have to forbid you to deprive anyone else of these rights. For example, if you distribute copies of PSPP, you must give the recipients all the rights that you have. You must make sure that they, too, receive or can get the source code. And you must tell them their rights.

Also, for our own protection, we must make certain that everyone finds out that there is no warranty for PSPP. If these programs are modified by someone else and passed on, we want their recipients to know that what they have is not what we distributed, so that any problems introduced by others will not reflect on our reputation.

Finally, any free program is threatened constantly by software patents. We wish to avoid the danger that redistributors of a free program will individually obtain patent licenses, in effect making the program proprietary. To prevent this, we have made it clear that any patent must be licensed for everyone's free use or not licensed at all.

The precise conditions of the license for PSPP are found in the GNU General Public License. You should have received a copy of the GNU General Public License along with this program; if not, write to the Free Software Foundation, Inc., 51 Franklin Street, Fifth Floor, Boston, MA 02110-1301 USA. This manual specifically is covered by the GNU Free Documentation License (see [Appendix A \[GNU Free Documentation License\]](#), page 42).

2 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

2.1 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.

-INSERT FIGURE

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.

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.

2.2 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.

-INSERT IMAGE

2.2.1 Data View

PSPP 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.

-INSERT FIGURE

2.2.2 Variable View

Display variable attributes associated with each case, where variable attributes (see [DummyNode], page 2) include name, type, width, decimal, variable label, value label, missing values, columns, align, and measurement.

-INSERT FIGURE

2.2.3 Data Entry and Editing

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

2.2.3.1 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.

-INSERT FIGURE

2.2.3.2 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.

-INSERT IMAGE

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 2) 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.

-INSERT IMAGE

Width

Click a cell in the *Width*(see [DummyNode], page 2) column.

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

-INSERT FIGURE

Decimals

Click a cell in the *Decimals* (see [DummyNode], page 2) column.

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

-INSERT FIGURE

Label

Click a cell in the *Label* column.

Type a variable label in the cell.

Press enter.

-INSERT FIGURE

Values

Click a cell in the *Values* (see [DummyNode], page 2) 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.

-INSERT FIGURE

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 2 for further information.

Press OK.

-INSERT FIGURE

Columns

Select a cell in the *Columns* (see also [DummyNode], page 2) column.

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

-INSERT FIGURE

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.

-INSERT FIGURE

Measure

Select a cell in the column *Measure* (see [DummyNode], page 2).

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

-INSERT FIGURE

2.2.3.3 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*.

-INSERT FIGURE

2.2.3.4 Read Variable in Data View

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

-INSERT FIGURE

2.2.3.5 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.

2.2.3.6 Select Entire Row or Column

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

2.2.3.7 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.

-INSERT FIGURE

2.2.3.8 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.

-INSERT IMAGE

2.2.3.9 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 2).

-INSERT FIGURE

2.2.4 Status Bar

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

-INSERT FIGURE

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

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

Split File Status indicates there are separate data into groups.

ihere

2.3 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

Output viewer displays the last set of instructions executed.

1.
-INSERT FIGURE
2.
-INSERT FIGURE

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

Example

-INSERT FIGURE

2.3.3 Save Output Contents

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

2.4 Syntax Editor

Use the Syntax Editor window to enter command line instructions. For syntax and rules see ([DummyNode], page 2).

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.

-INSERT FIGURE

2.4.1 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)

-INSERT FIGURE

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

-INSERT FIGURE

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

-INSERT IMAGE

2.4.2 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

2.4.3 Save

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

ihere

2.5 Main Menu

2.5.1 Overview

Perform tasks and operations on data with functions available through the drop down menus and toolbar. See also [\[DummyNode\]](#), page 2.

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.

2.5.2 File

2.5.2.1 New

Open a new data or syntax file.

2.5.2.2 Open

Open data, syntax, system, or portable files.

2.5.2.3 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.

-INSERT FIGURE

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

-INSERT FIGURE

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.

-INSERT FIGURE
ihere

Adjust Variable Formats Verify and edit variable attributes (see [DummyNode], page 2).

-INSERT FIGURE

Select the *Apply* button to import the data.

2.5.2.4 Save

Save the current file.

2.5.2.5 Save As

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

2.5.2.6 Display Data File Information

Working File Display the dictionary of the currently opened file.

-INSERT FIGURE

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

2.5.2.7 Recently Used Data

Links to recently used data files.

2.5.2.8 Recently Used Files

Links to recently used syntax files.

2.5.2.9 Quit

Close the current data file and shutdown PSPP.

2.5.3 Edit

2.5.3.1 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 2.

Example

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

-INSERT FIGURE

2.5.3.3 Insert Cases

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

2.5.3.4 Go To Case

Jump to the selected case number.

2.5.3.5 Cut

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

2.5.3.6 Copy

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

2.5.3.7 Paste

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

2.5.3.8 Find

Look for a case.

2.5.4 View

2.5.4.1 Status Bar

Turn the status bar on or off. See [\[DummyNode\]](#), page 2 for further information.

2.5.4.2 Font

View and edit the current font settings.

2.5.4.3 Grid Lines

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

2.5.4.4 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.

-INSERT FIGURE

2.5.4.6 Variables

Switch the data editor between data and variable view.

2.5.5 Data

2.5.5.1 Sort Cases

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

2.5.5.2 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.

2.5.5.3 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.

-INSERT FIGURE

Group a file with physiology data according to height.

-INSERT IMAGE

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.

-INSERT IMAGE

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.

-INSERT IMAGE

ihere

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.

Example

Using a data file containing physiology data, the file is grouped based on *height* using the Split File menu command.

-INSERT FIGURE

Select Data->Split File.

Select the Compare Groups radio button in the dialogue box.

Click *height* in millimeters to highlight the variable.

-INSERT FIGURE

Press the right arrow.

Hit OK.

2.5.5.7 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.

2.5.6 Transform

2.5.6.1 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).

-INSERT FIGURE

Open the file xxx.sav.

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

-INSERT FIGURE

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

-INSERT FIGURE

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.

-INSERT FIGURE

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.

-INSERT FIGURE

Click OK to add the result to the data file.

-INSERT FIGURE

ihere

2.5.6.3 Rank Cases

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

-INSERT FIGURE

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.

-INSERT FIGURE

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.

ihere use quotation to indent

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.

-INSERT FIGURE

ihere last

2.5.6.5 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.

-INSERT FIGURE

Open file xxx.sav

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

-INSERT FIGURE

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.

-INSERT FIGURE

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.

-INSERT FIGURE

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

-INSERT FIGURE

Press *Continue* to close the *New Values* dialogue box

Click OK on the *Recode into Same Variable* dialogue box.

The value 12 MPG is now 13 MPG.

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*.

-INSERT FIGURE(title will describe)

Click *Continue*. The *Old and New Values* dialogue box disappears.

Click *OK* on the *Recode into New Variables* dialogue box to put the new values in the *MPG* variable column.

-INSERT FIGURE

ihere lastlast

2.5.6.8 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 2.5.6.5 \[Recode into Same Variables\]](#), page 16).

-INSERT FIGURE

2.5.7 Analyze

2.5.7.1 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.

ihereastlastlast

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.

ihere last last last

2.5.7.2 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.

-INSERT FIGURE

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.

-INSERT FIGURE

-INSERT FIGURE

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

-INSERT FIGURE

Plot histograms or pie charts with the Chart Dialogue 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.

-INSERT FIGURE

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.

-INSERT FIGURE

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

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

-INSERT FIGURE

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

-INSERT FIGURE

Press OK to display the results in the output viewer.

-INSERT FIGURE

2.5.7.4 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.

Enter the following command lines shown in figure x. into the syntax window.

-INSERT FIGURE

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

-INSERT FIGURE

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.

-INSERT FIGURE

ihere last last last last

2.5.7.6 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.

-INSERT FIGURE

Start by looking at all test scores independent of gender.

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

-INSERT FIGURE

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.

-INSERT FIGURE

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

-INSERT FIGURE

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

ihereastlastlastlast

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.

-INSERT FIGURE

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.

-INSERT FIGURE

2.5.7.8 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 Pearsons 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 Fishers 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.

Lambda 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 (-):

-INSERT IMAGE

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

-INSERT IMAGE

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.

-INSERT IMAGE

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.

-INSERT IMAGE

ihrelastlastlastlast

2.5.7.10 Compare Means

2.5.7.11 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.

-INSERT IMAGE

Click OK to view the results in the output window.

-INSERT IMAGE

2.5.7.12 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.

-INSERT IMAGE

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.

-INSERT IMAGE

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*.

-INSERT IMAGE

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

-INSERT IMAGE

2.5.7.14 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.

-INSERT IMAGE

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

-INSERT IMAGE

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 Variables* dialogue box under *Var1*.

-INSERT IMAGE

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.

-INSERT IMAGE

iherefifthlast

2.5.7.15 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.

-INSERT IMAGE

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*.

-INSERT IMAGE

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.

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

-INSERT IMAGE

Click OK to displays the results in the output viewer.

-INSERT IMAGE

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.

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.

-INSERT IMAGE

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.

-INSERT IMAGE

2.5.7.18 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.

-INSERT IMAGE

Repeat the same steps with the *Temperature* variable.

Click OK to display the results in the output viewer.

-INSERT IMAGE

2.5.7.20 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).

-INSERT IMAGE

2.5.7.21 Extraction

- INSERT IMAGE

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.

Extract Specify to keep factors whose eigenvalues exceed a specified factor above the mean eigenvalue, or choose a specific number of factors to keep.

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

2.5.7.22 Rotation Dialogue Box

-INSERT IMAGE

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.

-INSERT IMAGE

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.

-INSERT IMAGE

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

-INSERT IMAGE

Click the @emphExtraction 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.

-INSERT IMAGE

2.5.7.24 Reliability

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

- INSERT IMAGE

2.5.7.25 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.

2.5.7.26 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.

-INSERT IMAGE

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.

-INSERT IMAGE

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

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

-INSERT IMAGE

2.5.7.28 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.

2.5.7.29 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*.

-INSERT IMAGE

Click OK to display the results in the output viewer.

-INSERT IMAGE

2.5.7.31 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.

-INSERT IMAGE

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.

-INSERT IMAGE

2.5.7.33 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.

-INSERT IMAGE

2.5.7.34 Test Variable

Continuous numerical data holding the test results.

2.5.7.35 State Variable

Indicates which category is positive.

2.5.7.36 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.

-INSERT IMAGE

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.

-INSERT IMAGE

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

-INSERT IMAGE

2.5.8 Utilities

2.5.8.1 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 2.

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.

-INSERT IMAGE

2.5.8.3 Data File Comments

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

2.5.8.4 Display Comments in Output

Display comments in the output viewer.

2.5.8.5 Reset

Clears the current comment.

-INSERT IMAGE

2.5.9 Windows

2.5.9.1 Minimize all Windows

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

2.5.9.2 Split

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

-INSERT IMAGE

2.5.10 Toolbar

PSPP provides a tool bar across the top with short cuts to more commonly used menu commands.

-INSERT IMAGE

Appendix A GNU Free Documentation License

Version 1.2, November 2002

Copyright © 2000,2001,2002 Free Software Foundation, Inc.
51 Franklin St, Fifth Floor, Boston, MA 02110-1301, USA

Everyone is permitted to copy and distribute verbatim copies of this license document, but changing it is not allowed.

0. PREAMBLE

The purpose of this License is to make a manual, textbook, or other functional and useful document *free* in the sense of freedom: to assure everyone the effective freedom to copy and redistribute it, with or without modifying it, either commercially or non-commercially. Secondly, this License preserves for the author and publisher a way to get credit for their work, while not being considered responsible for modifications made by others.

This License is a kind of “copyleft”, which means that derivative works of the document must themselves be free in the same sense. It complements the GNU General Public License, which is a copyleft license designed for free software.

We have designed this License in order to use it for manuals for free software, because free software needs free documentation: a free program should come with manuals providing the same freedoms that the software does. But this License is not limited to software manuals; it can be used for any textual work, regardless of subject matter or whether it is published as a printed book. We recommend this License principally for works whose purpose is instruction or reference.

1. APPLICABILITY AND DEFINITIONS

This License applies to any manual or other work, in any medium, that contains a notice placed by the copyright holder saying it can be distributed under the terms of this License. Such a notice grants a world-wide, royalty-free license, unlimited in duration, to use that work under the conditions stated herein. The “Document”, below, refers to any such manual or work. Any member of the public is a licensee, and is addressed as “you”. You accept the license if you copy, modify or distribute the work in a way requiring permission under copyright law.

A “Modified Version” of the Document means any work containing the Document or a portion of it, either copied verbatim, or with modifications and/or translated into another language.

A “Secondary Section” is a named appendix or a front-matter section of the Document that deals exclusively with the relationship of the publishers or authors of the Document to the Document’s overall subject (or to related matters) and contains nothing that could fall directly within that overall subject. (Thus, if the Document is in part a textbook of mathematics, a Secondary Section may not explain any mathematics.) The relationship could be a matter of historical connection with the subject or with related matters, or of legal, commercial, philosophical, ethical or political position regarding them.

The “Invariant Sections” are certain Secondary Sections whose titles are designated, as being those of Invariant Sections, in the notice that says that the Document is released

under this License. If a section does not fit the above definition of Secondary then it is not allowed to be designated as Invariant. The Document may contain zero Invariant Sections. If the Document does not identify any Invariant Sections then there are none.

The “Cover Texts” are certain short passages of text that are listed, as Front-Cover Texts or Back-Cover Texts, in the notice that says that the Document is released under this License. A Front-Cover Text may be at most 5 words, and a Back-Cover Text may be at most 25 words.

A “Transparent” copy of the Document means a machine-readable copy, represented in a format whose specification is available to the general public, that is suitable for revising the document straightforwardly with generic text editors or (for images composed of pixels) generic paint programs or (for drawings) some widely available drawing editor, and that is suitable for input to text formatters or for automatic translation to a variety of formats suitable for input to text formatters. A copy made in an otherwise Transparent file format whose markup, or absence of markup, has been arranged to thwart or discourage subsequent modification by readers is not Transparent. An image format is not Transparent if used for any substantial amount of text. A copy that is not “Transparent” is called “Opaque”.

Examples of suitable formats for Transparent copies include plain ASCII without markup, Texinfo input format, LaTeX input format, SGML or XML using a publicly available DTD, and standard-conforming simple HTML, PostScript or PDF designed for human modification. Examples of transparent image formats include PNG, XCF and JPG. Opaque formats include proprietary formats that can be read and edited only by proprietary word processors, SGML or XML for which the DTD and/or processing tools are not generally available, and the machine-generated HTML, PostScript or PDF produced by some word processors for output purposes only.

The “Title Page” means, for a printed book, the title page itself, plus such following pages as are needed to hold, legibly, the material this License requires to appear in the title page. For works in formats which do not have any title page as such, “Title Page” means the text near the most prominent appearance of the work’s title, preceding the beginning of the body of the text.

A section “Entitled XYZ” means a named subunit of the Document whose title either is precisely XYZ or contains XYZ in parentheses following text that translates XYZ in another language. (Here XYZ stands for a specific section name mentioned below, such as “Acknowledgements”, “Dedications”, “Endorsements”, or “History”.) To “Preserve the Title” of such a section when you modify the Document means that it remains a section “Entitled XYZ” according to this definition.

The Document may include Warranty Disclaimers next to the notice which states that this License applies to the Document. These Warranty Disclaimers are considered to be included by reference in this License, but only as regards disclaiming warranties: any other implication that these Warranty Disclaimers may have is void and has no effect on the meaning of this License.

2. VERBATIM COPYING

You may copy and distribute the Document in any medium, either commercially or noncommercially, provided that this License, the copyright notices, and the license notice saying this License applies to the Document are reproduced in all copies, and

that you add no other conditions whatsoever to those of this License. You may not use technical measures to obstruct or control the reading or further copying of the copies you make or distribute. However, you may accept compensation in exchange for copies. If you distribute a large enough number of copies you must also follow the conditions in section 3.

You may also lend copies, under the same conditions stated above, and you may publicly display copies.

3. COPYING IN QUANTITY

If you publish printed copies (or copies in media that commonly have printed covers) of the Document, numbering more than 100, and the Document's license notice requires Cover Texts, you must enclose the copies in covers that carry, clearly and legibly, all these Cover Texts: Front-Cover Texts on the front cover, and Back-Cover Texts on the back cover. Both covers must also clearly and legibly identify you as the publisher of these copies. The front cover must present the full title with all words of the title equally prominent and visible. You may add other material on the covers in addition. Copying with changes limited to the covers, as long as they preserve the title of the Document and satisfy these conditions, can be treated as verbatim copying in other respects.

If the required texts for either cover are too voluminous to fit legibly, you should put the first ones listed (as many as fit reasonably) on the actual cover, and continue the rest onto adjacent pages.

If you publish or distribute Opaque copies of the Document numbering more than 100, you must either include a machine-readable Transparent copy along with each Opaque copy, or state in or with each Opaque copy a computer-network location from which the general network-using public has access to download using public-standard network protocols a complete Transparent copy of the Document, free of added material. If you use the latter option, you must take reasonably prudent steps, when you begin distribution of Opaque copies in quantity, to ensure that this Transparent copy will remain thus accessible at the stated location until at least one year after the last time you distribute an Opaque copy (directly or through your agents or retailers) of that edition to the public.

It is requested, but not required, that you contact the authors of the Document well before redistributing any large number of copies, to give them a chance to provide you with an updated version of the Document.

4. MODIFICATIONS

You may copy and distribute a Modified Version of the Document under the conditions of sections 2 and 3 above, provided that you release the Modified Version under precisely this License, with the Modified Version filling the role of the Document, thus licensing distribution and modification of the Modified Version to whoever possesses a copy of it. In addition, you must do these things in the Modified Version:

- A. Use in the Title Page (and on the covers, if any) a title distinct from that of the Document, and from those of previous versions (which should, if there were any, be listed in the History section of the Document). You may use the same title as a previous version if the original publisher of that version gives permission.

- B. List on the Title Page, as authors, one or more persons or entities responsible for authorship of the modifications in the Modified Version, together with at least five of the principal authors of the Document (all of its principal authors, if it has fewer than five), unless they release you from this requirement.
- C. State on the Title page the name of the publisher of the Modified Version, as the publisher.
- D. Preserve all the copyright notices of the Document.
- E. Add an appropriate copyright notice for your modifications adjacent to the other copyright notices.
- F. Include, immediately after the copyright notices, a license notice giving the public permission to use the Modified Version under the terms of this License, in the form shown in the Addendum below.
- G. Preserve in that license notice the full lists of Invariant Sections and required Cover Texts given in the Document's license notice.
- H. Include an unaltered copy of this License.
- I. Preserve the section Entitled "History", Preserve its Title, and add to it an item stating at least the title, year, new authors, and publisher of the Modified Version as given on the Title Page. If there is no section Entitled "History" in the Document, create one stating the title, year, authors, and publisher of the Document as given on its Title Page, then add an item describing the Modified Version as stated in the previous sentence.
- J. Preserve the network location, if any, given in the Document for public access to a Transparent copy of the Document, and likewise the network locations given in the Document for previous versions it was based on. These may be placed in the "History" section. You may omit a network location for a work that was published at least four years before the Document itself, or if the original publisher of the version it refers to gives permission.
- K. For any section Entitled "Acknowledgements" or "Dedications", Preserve the Title of the section, and preserve in the section all the substance and tone of each of the contributor acknowledgements and/or dedications given therein.
- L. Preserve all the Invariant Sections of the Document, unaltered in their text and in their titles. Section numbers or the equivalent are not considered part of the section titles.
- M. Delete any section Entitled "Endorsements". Such a section may not be included in the Modified Version.
- N. Do not retitle any existing section to be Entitled "Endorsements" or to conflict in title with any Invariant Section.
- O. Preserve any Warranty Disclaimers.

If the Modified Version includes new front-matter sections or appendices that qualify as Secondary Sections and contain no material copied from the Document, you may at your option designate some or all of these sections as invariant. To do this, add their titles to the list of Invariant Sections in the Modified Version's license notice. These titles must be distinct from any other section titles.

You may add a section Entitled “Endorsements”, provided it contains nothing but endorsements of your Modified Version by various parties—for example, statements of peer review or that the text has been approved by an organization as the authoritative definition of a standard.

You may add a passage of up to five words as a Front-Cover Text, and a passage of up to 25 words as a Back-Cover Text, to the end of the list of Cover Texts in the Modified Version. Only one passage of Front-Cover Text and one of Back-Cover Text may be added by (or through arrangements made by) any one entity. If the Document already includes a cover text for the same cover, previously added by you or by arrangement made by the same entity you are acting on behalf of, you may not add another; but you may replace the old one, on explicit permission from the previous publisher that added the old one.

The author(s) and publisher(s) of the Document do not by this License give permission to use their names for publicity for or to assert or imply endorsement of any Modified Version.

5. COMBINING DOCUMENTS

You may combine the Document with other documents released under this License, under the terms defined in section 4 above for modified versions, provided that you include in the combination all of the Invariant Sections of all of the original documents, unmodified, and list them all as Invariant Sections of your combined work in its license notice, and that you preserve all their Warranty Disclaimers.

The combined work need only contain one copy of this License, and multiple identical Invariant Sections may be replaced with a single copy. If there are multiple Invariant Sections with the same name but different contents, make the title of each such section unique by adding at the end of it, in parentheses, the name of the original author or publisher of that section if known, or else a unique number. Make the same adjustment to the section titles in the list of Invariant Sections in the license notice of the combined work.

In the combination, you must combine any sections Entitled “History” in the various original documents, forming one section Entitled “History”; likewise combine any sections Entitled “Acknowledgements”, and any sections Entitled “Dedications”. You must delete all sections Entitled “Endorsements.”

6. COLLECTIONS OF DOCUMENTS

You may make a collection consisting of the Document and other documents released under this License, and replace the individual copies of this License in the various documents with a single copy that is included in the collection, provided that you follow the rules of this License for verbatim copying of each of the documents in all other respects.

You may extract a single document from such a collection, and distribute it individually under this License, provided you insert a copy of this License into the extracted document, and follow this License in all other respects regarding verbatim copying of that document.

7. AGGREGATION WITH INDEPENDENT WORKS

A compilation of the Document or its derivatives with other separate and independent documents or works, in or on a volume of a storage or distribution medium, is called

an “aggregate” if the copyright resulting from the compilation is not used to limit the legal rights of the compilation’s users beyond what the individual works permit. When the Document is included in an aggregate, this License does not apply to the other works in the aggregate which are not themselves derivative works of the Document.

If the Cover Text requirement of section 3 is applicable to these copies of the Document, then if the Document is less than one half of the entire aggregate, the Document’s Cover Texts may be placed on covers that bracket the Document within the aggregate, or the electronic equivalent of covers if the Document is in electronic form. Otherwise they must appear on printed covers that bracket the whole aggregate.

8. TRANSLATION

Translation is considered a kind of modification, so you may distribute translations of the Document under the terms of section 4. Replacing Invariant Sections with translations requires special permission from their copyright holders, but you may include translations of some or all Invariant Sections in addition to the original versions of these Invariant Sections. You may include a translation of this License, and all the license notices in the Document, and any Warranty Disclaimers, provided that you also include the original English version of this License and the original versions of those notices and disclaimers. In case of a disagreement between the translation and the original version of this License or a notice or disclaimer, the original version will prevail.

If a section in the Document is Entitled “Acknowledgements”, “Dedications”, or “History”, the requirement (section 4) to Preserve its Title (section 1) will typically require changing the actual title.

9. TERMINATION

You may not copy, modify, sublicense, or distribute the Document except as expressly provided for under this License. Any other attempt to copy, modify, sublicense or distribute the Document is void, and will automatically terminate your rights under this License. However, parties who have received copies, or rights, from you under this License will not have their licenses terminated so long as such parties remain in full compliance.

10. FUTURE REVISIONS OF THIS LICENSE

The Free Software Foundation may publish new, revised versions of the GNU Free Documentation License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns. See <http://www.gnu.org/copyleft/>.

Each version of the License is given a distinguishing version number. If the Document specifies that a particular numbered version of this License “or any later version” applies to it, you have the option of following the terms and conditions either of that specified version or of any later version that has been published (not as a draft) by the Free Software Foundation. If the Document does not specify a version number of this License, you may choose any version ever published (not as a draft) by the Free Software Foundation.

A.1 ADDENDUM: How to use this License for your documents

To use this License in a document you have written, include a copy of the License in the document and put the following copyright and license notices just after the title page:

```
Copyright (C) year your name.  
Permission is granted to copy, distribute and/or modify this document  
under the terms of the GNU Free Documentation License, Version 1.2  
or any later version published by the Free Software Foundation;  
with no Invariant Sections, no Front-Cover Texts, and no Back-Cover  
Texts. A copy of the license is included in the section entitled ‘‘GNU  
Free Documentation License’’.
```

If you have Invariant Sections, Front-Cover Texts and Back-Cover Texts, replace the “with...Texts.” line with this:

```
with the Invariant Sections being list their titles, with  
the Front-Cover Texts being list, and with the Back-Cover Texts  
being list.
```

If you have Invariant Sections without Cover Texts, or some other combination of the three, merge those two alternatives to suit the situation.

If your document contains nontrivial examples of program code, we recommend releasing these examples in parallel under your choice of free software license, such as the GNU General Public License, to permit their use in free software.