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

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
      1.5.3.2 Insert Cases ............................................... 21
      1.5.3.3 Go To Case ............................................... 21
      1.5.3.5 Cut ......................................................... 21
1.5.7 Analyze
1.5.6 Transform
1.5.5 Data
1.5.4 View

1.5.7.21 Extraction
1.5.7.20 Factor Analysis
Example

1.5.7.15 One-Way Anova
1.5.7.14 Paired Samples T Test
Example
1.5.7.12 Independent Samples T Test
1.5.7.11 One-Sample T-Test
1.5.7.10 Compare Means
Statistics and Measures of Association

1.5.7.6 Explore
Example

1.5.7.2 Frequencies
1.5.6.8 Recode into Different Variables
Example

1.5.6.3 Rank Cases
Rank Types
1.5.6.5 Recode into Same Variables
Example
1.5.6.7 Recode into Different Variables
Example

1.5.6.1 Compute
Example

1.5.5.6 Weight Cases
Organize Output by Groups
1.5.5.6 Weight Cases

1.5.5 Data
1.5.5.1 Sort Cases
1.5.5.2 Transpose
1.5.5.3 Split File
Compare Groups

1.5.4 View
1.5.4.1 Status Bar
1.5.4.2 Font
1.5.4.3 Grid Lines
1.5.4.4 Value Labels
Value Labels

1.5.3.7 Paste
1.5.3.8 Find

1.5.2 Compute

1.5.1 Copy

1.5.3.6 Copy
1.5.3.7 Paste
1.5.3.8 Find

1.5.4 View
1.5.4.1 Status Bar
1.5.4.2 Font
1.5.4.3 Grid Lines
1.5.4.4 Value Labels
Value Labels

1.5.5 Data
1.5.5.1 Sort Cases
1.5.5.2 Transpose
1.5.5.3 Split File
Compare Groups
Organize Output by Groups
1.5.5.6 Weight Cases

1.5.6 Transform
1.5.6.1 Compute
Example
1.5.6.3 Rank Cases
Rank Types
1.5.6.5 Recode into Same Variables
Example
1.5.6.8 Recode into Different Variables

1.5.7 Analyze
1.5.7.1 Descriptive Statistics Definitions
1.5.7.2 Frequencies
Example
1.5.7.4 Descriptives
Example
1.5.7.6 Explore
Example
1.5.7.8 Crosstabs
Statistics and Measures of Association
1.5.7.10 Compare Means
1.5.7.11 One-Sample T-Test
1.5.7.12 Independent Samples T Test
Example
1.5.7.14 Paired Samples T Test
1.5.7.15 One-Way Anova
Example
1.5.7.18 Bivariate Correlation
Example
1.5.7.20 Factor Analysis
1.5.7.21 Extraction
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.

![Syntax editor window](image)

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.
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.
Chapter 1: Graphical User Interface

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.

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.](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 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.](image)

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.](image)

### 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.](image)

### 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.](image)

### Values
Chapter 1: Graphical User Interface

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.](image)

### 1.2.3.2 New Variable

Enter a new variable by selecting insert variable from the edit drop down menu or toolbar. 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](image)

### 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.
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.

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](image)

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-\(\rightarrow\)Clear Cases from the drop down menu.

![Image](image1.png)

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-\(\rightarrow\)Sort Cases (see [DummyNode], page 1).

![Image](image2.png)

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.

![Status bar](image)

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.

![Output viewer](image)

Figure 1.23: Output viewer displays open file instructions.

2.

![Output viewer](image)

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.
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 dialogue 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.

![Descriptives Dialogue Box](image)

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.

![Command Line Equivalent](image)

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 as 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.

![Importing Textual Data Dialogue Box](image)

Figure 1.30: Import textual data dialogue box

First Line of Data Dialogue

Select the first line of data.

![Select Data to Import](image)

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).

![Adjust Variable Formats](image)

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.

![Dictionary of current file](image)

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.](image)

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.

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.

![Split File Dialogue Box](image)

Figure 1.37: Split File Dialogue Box

Group a file with physiology data according to height.

![Physiology data file](image)

Figure 1.38: Physiology data file
Open the file xxx.sav.
Select Data->Split File.
Select the \textit{Compare groups} radio button in the pop up window.
Click \textit{Height in millimeters} to highlight the variable.

![Split-File dialogue box](image)

\textbf{Figure 1.39: Split-File dialogue box}

Select the arrow to the left of the \textit{Groups based on} textbox to move the \textit{height} variable inside that textbox.

Click OK to see the results in the output viewer.

![Split-file results in output viewer](image)

\textbf{Figure 1.40: Split-file results in output viewer}

\textbf{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.

\textbf{Organize Output by Groups}
Show results for each split-file group separately.

\textbf{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.

\textbf{1.5.6 Transform}

\textbf{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).

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 dialogue box.

Figure 1.42: Compute dialogue box
Enter *weightChange* in the Target Variable textbox and click the *Type and Label* button below it; This variable holds the result of the calculation.

![Figure 1.43: Compute solution Type and Label](image)

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.](image)
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.](image-url)
Click OK to add the result to the data file.

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. PSPP automatically names and labels these such as to indicate original variable name and selected measures.

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.](image)

**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 \( \frac{r - 3/8}{w + \frac{1}{8}} \), where \( w \) is the sum of the case weights and \( r \) is the rank.

- **Tukey** Apply the formula \( \frac{r - 1/3}{w + 1/3} \), where \( r \) is the rank and \( w \) is the sum of the case weights.

- **Rankit** Estimate expected values using \( \frac{r - 1/2}{w} \), 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 \( \frac{r}{w + 1} \), 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](image)

### 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.
Open file `xxx.sav`

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

Figure 1.50: Mileage-per-gallon data file

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](image)

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.](image)
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 *range 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*. 
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.
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
mesorkurtic.

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.
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 1.62: Customer satisfaction survey](image)

![Figure 1.63: Pick variables in the frequencies dialogue box.](image)
Select the *charts* button to bring up the *Charts* dialogue box.

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

![Image of Charts dialogue box](image1.png)

Figure 1.64: Plot histograms of selected data.

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

![Image of Frequencies dialogue box](image2.png)

Figure 1.65: Select statistical analyses in the *frequencies* pop-up.
Chapter 1: Graphical User Interface

Press OK to display the results in the output viewer.

![Output Viewer](image)

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.

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

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

```
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.
```

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

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.

![Data Editor with Z-scores](image)

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.
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](image)

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

![Figure 1.73: Explore statistical analysis results displayed in output viewer.](image)
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.

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 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
Correlations

Return Spearman's correlation coefficient and rho for numeric data when the values in the rows and columns are ordered. When both variables are quantitative returns the Pearson's 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 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**

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 cerebellar toxicity who have taken a generic chemotherapeutic agent rather than a non-generic one.

<table>
<thead>
<tr>
<th>Generic</th>
<th>Toxicity+</th>
<th>Toxicity-</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generic+</td>
<td>11</td>
<td>1</td>
</tr>
<tr>
<td>Generic-</td>
<td>3</td>
<td>31</td>
</tr>
</tbody>
</table>

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 (-):

![Figure 1.76: Toxicity data in the data editor](image)

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

![Figure 1.77: Weight cases function](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.

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.

![One-Sample T Test dialogue box](image)

Figure 1.80: One-Sample T Test dialogue box

Click OK to view the results in the output window.

![One-Sample T Test output results](image)

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.

<table>
<thead>
<tr>
<th>Height</th>
<th>Gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>152.71</td>
<td>0</td>
</tr>
<tr>
<td>142.57</td>
<td>0</td>
</tr>
<tr>
<td>155.48</td>
<td>0</td>
</tr>
<tr>
<td>148.79</td>
<td>0</td>
</tr>
<tr>
<td>167.74</td>
<td>0</td>
</tr>
<tr>
<td>149.51</td>
<td>0</td>
</tr>
<tr>
<td>142.00</td>
<td>0</td>
</tr>
<tr>
<td>150.21</td>
<td>0</td>
</tr>
<tr>
<td>171.19</td>
<td>0</td>
</tr>
<tr>
<td>141.99</td>
<td>0</td>
</tr>
<tr>
<td>152.67</td>
<td>1</td>
</tr>
<tr>
<td>139.65</td>
<td>1</td>
</tr>
<tr>
<td>152.19</td>
<td>1</td>
</tr>
<tr>
<td>156.94</td>
<td>1</td>
</tr>
<tr>
<td>151.86</td>
<td>1</td>
</tr>
<tr>
<td>157.79</td>
<td>1</td>
</tr>
<tr>
<td>152.49</td>
<td>1</td>
</tr>
<tr>
<td>157.52</td>
<td>1</td>
</tr>
<tr>
<td>156.95</td>
<td>1</td>
</tr>
<tr>
<td>141.61</td>
<td>1</td>
</tr>
</tbody>
</table>
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](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.

![Figure 1.83: Define variables and groups in the Independent-Samples T Test dialogue.](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.

Click OK on the Independent-Samples T Test dialogue box to examine the results 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.

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 Variables 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.

Click OK to display the results in the Output window.

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.

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](image)

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.](image)

Click OK to displays the results in the output viewer.

![Figure 1.92: One-Way ANOVA output](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.

![One-Way ANOVA](image)

Figure 1.93: Use *contrast* 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.

![One-Way ANOVA: Contrasts](image)

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.

![One-Way ANOVA output](image)

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.

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

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.

![Bivariate Correlations dialogue box](image)

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.

![Bivariate Correlations output](image)

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).

**Method** The Method button on the Extraction dialogue box enables you to choose the method of factor extraction. Options include Principal Components Analysis and Principal Axis Factoring.

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

![Factor Analysis Rotation dialogue box](image)

- **Extract** Specify to keep factors whose eigenvalues exceed a specified factor above the mean number of factors to keep.
- **Maximum Iterations for 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.

![Survey data](image.png)

**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.
1.5.7.24 Reliability

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

![Reliability Analysis dialogue box](image)

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.
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 ephhItems textbox.

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.

![Reliability Analysis output](image)

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, *HotCold* 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.

<table>
<thead>
<tr>
<th>HotCold</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

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*.

![Binomial dialogue box](image)

Figure 1.110: *Binomial* dialogue box

Click OK to display the results in the output viewer.

![Binomial Test output](image)

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.

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.

**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.
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.
Figure 1.115: Medical test data

<table>
<thead>
<tr>
<th>Age</th>
<th>Disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>20.00</td>
</tr>
<tr>
<td>2</td>
<td>30.00</td>
</tr>
<tr>
<td>3</td>
<td>34.00</td>
</tr>
<tr>
<td>4</td>
<td>37.00</td>
</tr>
<tr>
<td>5</td>
<td>41.00</td>
</tr>
<tr>
<td>6</td>
<td>44.00</td>
</tr>
<tr>
<td>7</td>
<td>48.00</td>
</tr>
<tr>
<td>8</td>
<td>53.00</td>
</tr>
<tr>
<td>9</td>
<td>57.00</td>
</tr>
<tr>
<td>10</td>
<td>60.00</td>
</tr>
<tr>
<td>11</td>
<td>33.00</td>
</tr>
<tr>
<td>12</td>
<td>35.00</td>
</tr>
<tr>
<td>13</td>
<td>38.00</td>
</tr>
<tr>
<td>14</td>
<td>37.00</td>
</tr>
<tr>
<td>15</td>
<td>42.00</td>
</tr>
<tr>
<td>16</td>
<td>44.00</td>
</tr>
</tbody>
</table>
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.

![ROC Curve](image)

**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.

![Variables dialogue box](image)

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.

![Comments dialogue box](image)

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

Figure 1.121: Data editor toolbar