An introduction to IBM SPSS Statistics

Contents

1 Introduction.................................................................................................................................................1
2 Entering your data.........................................................................................................................................2
3 Preparing your data for analysis ..................................................................................................................10
4 Exploring your data: univariate analysis .....................................................................................................14
5 Generating descriptive statistics for metric variables ...............................................................................17
6 Exploring your data: analysing more than one variable .............................................................................21
7 Getting started with inferential statistics..................................................................................................27

1 Introduction

The aim of this short guide is to provide an introduction to IBM SPSS Statistics (hereafter: SPSS), a specialist statistical analysis software that is widely used in the social sciences, including business and management research. It covers the following topics:

1. Entering your data
2. Preparing your data for analysis
3. Exploring your data: univariate analysis
4. Exploring your data: analysing more than one variable
5. Getting started with inferential statistics

The routines described are those introduced in Chapter 13. The guide assumes that you are already familiar with the contents of that chapter and therefore concentrates on how to run the analysis techniques referred to in that chapter rather than when and why you would use them or how they should be interpreted.

The guide is written around SPSS Version 21. Most of the functionality referred to in the guide is also available in earlier versions although the user interface has changed somewhat.

This is not intended to be a comprehensive guide to using SPSS; such a document would be a (long) book in itself. If you wish to go further, there are a large number of books available
(see further reading for Chapter 13 for suggestions). There is a useful tutorial and a statistics coach available through the SPSS Help menu. You may also find that SPSS training or support material is available through your institution.

1.1 When to use SPSS

Many basic analysis projects involving data exploration, descriptive statistics and simple inferential statistics can be successfully completed using a spreadsheet package such as Microsoft Excel. SPSS comes into its own for more advanced projects, especially those requiring statistical routines not available in standard spreadsheet packages and those involving multivariate analysis. If your project involves either of the latter, consider using SPSS (or a similar package) for your data analysis.

2 Entering your data

We will demonstrate data entry using a small dataset on customer satisfaction that is available for download (customer satisfaction.sav) on the companion website (along with other datasets).

On opening SPSS you will see a start-up window that allows you to choose various options (Figure 1). These include: opening an existing SPSS data file, opening another file type and typing in your own data. Select the Type in data and click OK. This will take you to the SPSS Data Editor.

Figure 1 – SPSS Start-up window
2.1 The SPSS Data Editor

The Data Editor includes a menu bar along the top of the screen and tabs in the bottom left corner which allow you to choose one of two views:

- The Data View (Figure 2), which is where you enter data. It is laid out as a case-by-variable matrix. Each numbered row represents a case (e.g. a respondent) in your data and each column represents a variable. When you name your variables, the name will appear in the column header.

Figure 2 – SPSS Data View

- The Variable View (Figure 3), which is where you can define the details of each variable by giving them labels, defining values, etc.

Figure 3 – SPSS Variable View
Switch between the views by clicking on the appropriate tab.

2.2 Adding variables

To add a variable to your dataset, select the Variable View tab. We will create a new variable called Gender that records the gender (male or female) of respondents. Type a suitable name for the variable in the column headed Name in the first row (Figure 4). This will create a new variable and populate other cells in the same row.

Figure 4 – Adding a new variable in the Variable View

<table>
<thead>
<tr>
<th></th>
<th>Name</th>
<th>Type</th>
<th>Width</th>
<th>Decimals</th>
<th>Label</th>
<th>Values</th>
<th>Missing</th>
<th>Columns</th>
<th>Align</th>
<th>Measure</th>
<th>Role</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Gender</td>
<td>Numeric</td>
<td>0</td>
<td>2</td>
<td>None</td>
<td>None</td>
<td>0</td>
<td>Right</td>
<td>Unknown</td>
<td>Input</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

It will also allocate the variable name as a column header in the Data View (Figure 5).

Figure 5 – New variable in the Data View

Keep variable names short but meaningful. Avoid, for example, using Q1, Q2, Q3, etc. to indicate responses to questions but instead use a simple naming convention, using abbreviations if required such as CSat1 for ‘Customer satisfaction question 1’. Note that SPSS will not accept spaces or some characters (such as *, ? or /) as names.

2.3 Defining variables

Next you need to define the properties of your new variable. This is done through the Variable View by editing the values in the different columns for each variable as required. Taking each column in turn and working from left to right:
2.3.1 Type

This defines the type of variable. Click on the right-hand side of the cell to open up the Variable Type dialogue box (Figure 6). The default is Numeric which is suitable for most purposes but you can choose others if required. String should only be used if the data is in the form of text. (Note: the Variable Type dialogue box also allows you to set the Width and Decimal Places values for the variable.)

Figure 6 – Variable Type dialogue box

2.3.2 Width

This sets the maximum number of characters in the data for that column. Only change this if you need more than the default 8 characters.

2.3.3 Decimals

This sets the number of decimal places that will be shown for the variable. The default is 2; you can change this by clicking on the cell and using the up/down arrows or by typing in the desired value. You can set the value to 0 for variables containing only whole numbers.

2.3.4 Label

Here you can enter a longer and more descriptive name for your variables. This is the variable name that will appear in your output. This is a very useful option.

(Hint: SPSS default in the dialogue boxes is to show variables by their labels. Whilst labels are good for output they are not always convenient when shown in dialogue boxes. To change
the default setting to show names rather than variable labels select File > Options > General and select Display Names in the Variable Lists area.)

2.3.5 Values

In the Values column you can define the numerical codes that you enter for nominal variables (such as gender) by allocating them labels. The labels will appear in your output. It is another very useful option. To assign a label:

1. Click on the right-hand side of the cell in the Values column for that variable to open up the Value Labels dialogue box (Figure 7).
2. To define a value, enter the number in the Value box and the definition in the Label box and click Add.
3. This creates the definition in the box.
4. Repeat the process to add further definitions, clicking Add each time (you can edit them using the Change or Remove buttons).
5. Once complete, press OK and the values will be defined for that variable.

Figure 7 – Value Labels dialogue box

2.3.6 Missing values

If there is no value entered in a particular cell for a particular case, SPSS will automatically treat and report it as a system-missing value. However, SPSS also allows you to define specific values in the dataset as representing missing values. This can be helpful if, for example, you want to distinguish between values that are missing because the respondent did not reply and those that are missing because of a technical data collection problem. Click on the cell in the Missing Values column if you wish to define specific missing values.
2.3.7 Columns

This cell determines the width of the column in the Data View. Unless you have changed the Width value it can be left as the default.

2.3.8 Align

This allows you to specify the alignment for the column; it can be left as the default for most purposes.

2.3.9 Measure

This is an important option as it allows you to identify the level of measurement for each variable. Click on the cell to open up a drop-down box (Figure 8). It gives three options: scale, ordinal and nominal. Scale is used for both interval and ratio scale data. Since Gender is nominal, select this measurement level.

Figure 8 – Measure drop-down box

![Measure drop-down box]

2.3.10 Role

The final column is labelled Role. The options in this column support particular types of analysis routine. We will not be using these options.

2.4 Adding more variables

You can add more variables by repeating the procedure described above. As you do so, SPSS will allocate the variable name to consecutive columns in the Data View. Figure 9 shows the complete set of variables for the customer satisfaction dataset.
2.5 Entering data into the Data View

Once the variables have been created you can enter data directly into the Data View. Figure 10 shows the complete customer satisfaction dataset in the Data View.

If you have assigned value labels to any of your variables you can show the text rather than the numbers in the Data View by clicking on the Value Labels icon in the ribbon (Figure 11).
2.6 Importing data

If your dataset already exists in another format (such as Excel), you can usually import that data into SPSS rather than retyping it manually. There are various options for doing this. At start up you can select the Open another type of file option in the start-up window and then choose the appropriate file. Alternatively, once SPSS is running, select File > Open > Data and choose the file type in the Files of Type drop down. (Note: if you are importing data from an Excel worksheet that has column headers in the first row, SPSS will ask whether you wish to use those as the variable names.)

Once the data are imported, you can amend details of each variable (such as the value labels) using the procedures outlined earlier.

2.7 Saving and managing your data

Ensure that you save your work regularly (Select: File > Save). Once you have created your dataset, ensure that you back it up in a secure place, not on your PC or laptop. If you make any changes to your master dataset, record those changes and create a duplicate back-up.

Give files a meaningful name. It is also helpful to date them as this makes it easier to track back if you need to do so.

2.8 SPSS Viewer

As you work, SPSS opens a second window, the SPSS Viewer which displays output such as graphs and test results (Figure 12). It also records other events such as errors. As well as the menu bar and the output area, the Viewer contains a tree diagram of the current output.

Figure 12 – SPSS Viewer
Output files need to be saved as separate files (File > Save) and can be printed as required. They are given the file extension .spv.

Individual items (such as tables or graphs) can be copied and pasted into word processing packages if required. Right click on the item and select Copy. Alternatively, if you wish to paste the output as an image (e.g. a JPEG), right click and select Copy Special and select the desired type.

3 Preparing your data for analysis

Once your data are entered you can follow the steps in Chapter 13 to prepare your data for analysis. SPSS has a number of routines that can help you with this process, including:

3.1 Calculating summated scores for multi-item scales

Suppose you have a scale that used three items to measure customers’ intention to recommend your product (see the Intend to recommend.sav dataset). In your dataset you have three variables (named Rec_1 to Rec_3) that represent responses to the three questions. You now need to calculate a new variable which is the mean of the three items. You can use the Compute Variable command to do this.

1. Select Transform > Compute Variable to open up the Compute Variable dialogue box (Figure 13).
2. In the Target Variable box enter the name you want to give to the new variable (here: Rec_Sum).
3. Next you have to tell SPSS how to calculate the new variable. You do this in the Numeric Expression box. You can copy the variable names from the left-hand box by selecting them and clicking on the arrow button and adding the mathematical operators (+, -, <, etc.) in the dialogue box or typing them in directly. Alternatively, SPSS offers a range of functions in the Function group that will enter the numeric expression for you. Select Function group > statistical to open up a list of statistical functions in the Functions and Special Variables box. Double-click on Mean (or use the adjacent up arrow) to insert the appropriate function expression (MEAN?,?) into the Numeric Expression box. The question marks (?) indicate that you need to specify relevant variable names in the expression. This you can now do as shown in Figure 13.
4. Click OK to create the new variable.

Figure 13 – Compute Variable dialogue box

3.2 Reverse coding variables

You can also use the Compute Variable command for reverse coding of Likert-type scales by using the following equation:

\[ \text{new value} = (\text{highest value in scale} + 1) - \text{original score} \]

So for a 7-point scale with an original score of 2, the reverse coded value would be:

\[ \text{new value} = (7 + 1) - 2 = 8 - 2 = 6 \]

Figure 14 shows this being done using the Compute Variable command to reverse code the variable Rec_1 which is measured on a 7-point scale into a new variable Rec_1_Rev.

Figure 14 – Compute Variable dialogue box for reverse coding
3.3 Recoding variables

SPSS can be used to recode a variable, for example to group a nominal or metric variable into a smaller number of categories. Two command options are available for recoding. We recommend the Recode into Different Variables as this ensures that you keep your original data. We will illustrate how it works by using it as an alternative way of reverse coding the Rec_1 variable in the earlier example.

1. Select Transform > Recode into Different Variables to open up the dialogue box (Figure 15).
2. In the left-hand box, select the Rec_1 variable and click on the right arrow to send it to the Numeric Variable → Output Variable box.
3. In the Output Variable area, give the new variable a name and label.

Figure 15 – Recode into Different Variable dialogue box
4. Click on Change to enter the new name into the Numeric Variable → Output Variable box.

5. Now click on the Old and New Values… button to open up a new dialogue box where you specify the details of the recoding (Figure 16).

6. Select the Value option in the Old Value area. Enter 1 in the box. Select the Value option in the New Value area and enter 7 in the box. Click Add. The Old → New box shows that 1 will be recoded as 7.

7. Repeat this procedure until you have specified all the required recoding. Click Continue to return to the Recode into Different Variables dialogue box.

8. Click OK to create the new variable.

Figure 16 – Old and New Values dialogue box

3.4 Checking scale reliability (Cronbach’s alpha)

SPSS can be used to check the reliability of multi-item scales by calculating Cronbach’s alpha. To do so:

1. Select Analyze > Scale > Reliability Analysis to open up the Reliability Analysis dialogue box (Figure 17).

2. Select the required variables in the left hand box and send them to the Items: box using the right arrow.

3. Check that the Model drop-down box is set to Alpha.

4. If desired give the scale to be tested a label. You can also click on the Statistics button to obtain additional statistics about the items and the scale.

Figure 17 – Reliability Analysis dialogue box
5. Click OK. The basic output is shown in Figure 18. It reports an alpha of 0.208, indicating low scale reliability: the summated scale that we calculated earlier would be an unreliable measure of customer’s Intention to recommend.

Figure 18 – Reliability Analysis output (Cronbach’s alpha)

4 Exploring your data: univariate analysis

This section looks at using SPSS for univariate data exploration with the techniques discussed in Chapter 13. It covers:

1. Creating a frequency table
2. Generating descriptive statistics
3. Creating graphs in SPSS

We will use the customer satisfaction.sav data introduced earlier.
4.1 Creating a frequency table

To create a frequency table for a categorical variable:

1. Select Analyze > Descriptives > Frequencies to open up the Frequencies dialogue box (Figure 19).
2. Select the desired variable(s) in the left-box and click the right arrow to send them to the Variable(s) box.
3. Tick the Display frequency tables box.
4. If you wish to add a graph (chart) to the output, click on the Charts button to open up the Frequencies: Charts dialogue box (Figure 20).

5. In the Charts dialogue box, select the desired Chart Type and Chart Value (frequencies or percentages).

Figure 20 – Frequencies: Charts dialogue box

6. Click on Continue to return to the Frequencies dialogue box.

7. Click on OK to create the output as shown in Figure 21 (the pie chart has been resized).
Generating descriptive statistics for metric variables

There are several ways of generating descriptive statistics for metric variables in SPSS. The method we show here has the advantage of allowing you to create histograms at the same time.

1. Select Analyze > Descriptives > Frequencies to open up the Frequencies dialogue box.
2. Select the desired variable(s) in the left-box and click the right arrow to send them to the Variable(s) box.
3. Deselect the Display frequency tables box if it has been ticked.
4. Click on the Statistics button to open up the Frequencies: Statistics dialogue box (Figure 22).
5. In the Statistics dialogue box, select the desired statistics.
6. Click on Continue to return to the Frequencies dialogue box.
7. If you wish to add a histogram to the output, click on the Charts button to open up the Frequencies: Charts dialogue box (Figure 20).

8. In the Charts dialogue box, select the desired Chart Type Histograms: and select the Show normal curve on histogram option.

9. Click on Continue to return to the Frequencies dialogue box.

10. Click on OK to create the output as shown in Figure 23 (the pie chart has been resized).

Figure 23 – Descriptive statistics and histogram
5.1 Creating graphs in SPSS

As we have already seen, a number of SPSS routines offer the option of producing chart outputs. It is also possible to produce graphs directly from your data using the Chart Builder. We will demonstrate its use by creating a box plot.

1. Select Graphs > Chart Builder to open up the Chart Builder dialogue box (Figure 24). (Note: when you first open Chart Builder you will be presented with a dialogue box reminding you to define your variables. As you will have done this at data entry, click OK to proceed to the main dialogue box.)
2. The Chart Builder is divided into three main areas: a list of available variables, a chart preview area and a series of tabs, with Gallery as the default which shows a range of chart types.

Figure 24 – Chart Builder

3. To create a box plot, select Boxplot in the Choose from list under the Gallery tab. A choice of box plot icons will appear under the Gallery tab. Drag and drop the 1-D box plot (Hint: hold your mouse over the chart icons in the Gallery to see what they represent) from there into the chart preview area (Figure 25). Doing this also opens

---

1 The Chart Builder is a relatively new feature in SPSS. If you have used earlier versions, you may prefer to use the Legacy Dialogs option which presents each graph type separately.
up an Element Properties dialogue box which you can use to change how you want elements to be displayed.

**Figure 25 – Chart Builder and Element Properties area for 1-D box plot**

4. In the chart preview area you will see a rectangle marked ‘X-Axis?’. Drag and drop the variable from the Variables list that you want to display on the chart into that rectangle (in this case Customer Commitment). The box plot now takes shape (Figure 26).

**Figure 26 – Chart preview for 1-D box plot, variable added**

5. Click OK to create the output (Figure 27).
Other chart types can be created by selecting different options from the Gallery.

Once a graph has been output in the Viewer you can double-click on it to open up a Chart Editor. This allows you to change features of the chart such as colour and add or remove data labels. The SPSS Help tutorial provides more details on how to use the Chart Editor.

### 6 Exploring your data: analysing more than one variable

This section looks at using SPSS for exploring more than one variable at a time (bivariate and multivariate analysis) to apply the techniques discussed in Chapter 13. It covers:

1. Comparing means and other descriptive statistics for different groups
2. Exploring the association between categorical variables
3. Exploring the association between two metric variables

We will use the customer satisfaction.sav dataset introduced earlier.

#### 6.1 Comparing different groups or different variables

A number of techniques can be applied using SPSS to compare different groups or different metric variables.

##### 6.1.1 Comparing means and other descriptive statistics for different variables

To compare descriptive statistics for more than one metric variable, you can use the same procedures as described above for generating a single variable using Analyze > Descriptive Statistics > Frequencies. Simply add all the variables you wish to inspect to the Variables area to create a summary table. An example output is shown in Figure 28.
6.1.2 Compare means and other descriptive statistics for groups

To generate a simple table comparing means and other descriptive statistics for different categories groups of a nominal variable:

- Select Analyze > Compare Means > Means to open the Means dialogue box (Figure 29).
- Send the grouping (categorical) variable to the Independent List box. Send the metric variable to be used for comparison to the Dependent List box.

Figure 29 – Means dialogue box

- Click on Options to open up the Means: Options dialogue box. The default statistics are mean, number of cases and standard deviation. Add any additional statistics you require and click Continue (Figure 30) to return to the Means dialogue box.
6.1.3 Visualising differences in group means

You can use the Chart Builder to create charts comparing groups. We will illustrate the basic principles by creating a bar chart comparing the mean satisfaction scores of male and female customers.

1. Select Graphs > Chart Builder to open up the Chart Builder dialogue box. Select Bar from the Choose from list under the Gallery tab and then drag the Simple Bar icon into the chart preview area (Figure 32).
2. Drag and drop Gender the Variables list into the X-Axis box in the chart preview area and Customer satisfaction into the Y-Axis box. The bar chart now takes shape (Figure 33).

Figure 33 – Chart preview for Simple Bar chart, variables added

3. Click OK. The output is shown in Figure 34.
The procedure just outlined can be used to create other chart types such as box plots by selecting different options from the Gallery.

6.2 Exploring the association between categorical variables

Contingency tables can be generated using the SPSS Crosstabs command to explore the associations between categorical variables. We will create a simple cross tabulation of gender and store location to illustrate the process.

1. Select Analyze > Descriptive Statistics > Crosstabs to open the Crosstabs dialogue box (Figure 35).
2. Decide on the variables to be included in the table. In this case Gender will be the row headers and Store location the column headers so send them to the Row(s) and Column(s) boxes respectively.
3. Tick Display clustered bar charts.
4. Click on Cells to open up the Crosstabs: Cell Display dialogue box (Figure 36). Specify what you want to appear in the table. The default is observed counts. Select Row in the percentages box. Click Continue.

Figure 36 – Crosstabs: Cell Display dialogue box

![Crosstabs: Cell Display dialogue box](image)

5. In the Crosstabs dialogue click OK. The output is shown in Figure 37; the bar chart has been resized.

Figure 37 – Contingency table and clustered bar chart showing store location by gender

![Contingency table and clustered bar chart](image)
In addition to creating clustered bar charts through the Crosstabs function, graphs including stacked bar charts can also be created using the Chart Builder.

### 6.3 Creating scatter plots to explore the association between metric variables

Use the Chart Builder to create scatter plots as follows:

1. Select Graphs > Chart Builder.
2. In the Gallery, select Scatter/Dot.
3. Drag and drop the Simple Scatter icon into the chart preview area.
4. Drag and drop the relevant variables into the X- and Y-Axis boxes.
5. Click OK to create the scatter plot.

### 7 Getting started with inferential statistics

In Chapter 13 we introduced the use of inferential statistics to help us draw conclusions about a larger population on the basis of a sample set of data. In this section of the guide we give brief details of how to carry them out in SPSS. You should refer to Chapter 13 for details of when and why to select particular techniques and how to interpret their output.

#### 7.1 Calculating and charting confidence intervals

To calculate confidence intervals for the mean:

- Select Analyze > Descriptive Statistics > Explore to open up the Explore dialogue box (Figure 38).
- Send the variable(s) for which you want to calculate the confidence interval to the Dependent List.
- Click on the Statistics and check that the Descriptives option has been ticked in the Explore: Statistics dialogue box. You can also change the desired confidence interval (the default is 95%). Click on Continue.
- In the Display area, select the Statistics radar button.
• Click on OK to produce the output (Figure 39). The confidence interval for the mean is included in the table.

**Figure 39 – Explore output including 95% confidence interval for the mean**

<table>
<thead>
<tr>
<th>Case Processing Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cases</strong></td>
</tr>
<tr>
<td>Valid</td>
</tr>
<tr>
<td>N</td>
</tr>
<tr>
<td>Customer satisfaction</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Descriptives</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Customer satisfaction</strong></td>
</tr>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>95% Confidence Interval for Mean</td>
</tr>
<tr>
<td>5% Trimmed Mean</td>
</tr>
<tr>
<td>Median</td>
</tr>
<tr>
<td>Variance</td>
</tr>
<tr>
<td>Std. Deviation</td>
</tr>
<tr>
<td>Minimum</td>
</tr>
<tr>
<td>Maximum</td>
</tr>
<tr>
<td>Range</td>
</tr>
<tr>
<td>Interquartile Range</td>
</tr>
<tr>
<td>Skewness</td>
</tr>
<tr>
<td>Kurtosis</td>
</tr>
</tbody>
</table>

If you want to compare confidence intervals for different groups of a nominal variable, follow the steps above but include the grouping variable in the Factor List in the Explore dialogue box. This will generate the confidence intervals for the means of each group.

You can also use the Explore routine to generate box plots and other graphical output (select Both or Plots in the Display box) and click on Plots to choose the type of chart you want.
7.1.1 Charting confidence intervals

To produce a chart of confidence intervals of the mean satisfaction level of male and female customers:

1. Select Graphs > Chart Builder to open up the Chart Builder dialogue box. Select Bar from the Choose from list under the Gallery tab and then drag the Simple Error Bar icon into the chart preview area (Figure 40).
2. Drag and drop Gender from the Variables list into the X-Axis box in the chart preview area and Customer satisfaction into the Y-Axis box. The error bar chart now takes shape; the default is to display a 95% confidence interval for the mean.

Figure 40 – Chart preview for Error Bar chart, variables added

3. Click OK. The output is shown in Figure 42.

Figure 41 – Graph of 95% confidence intervals for the mean of customer satisfaction by gender
7.2 Carrying out tests of difference

This section outlines how to run tests of difference referred to in Chapter 13.

7.2.1 Running t-tests in SPSS

SPSS offers the following suite of *t*-tests for comparing means:

- One-sample *t*-test
- Paired two-sample *t*-test
- Independent two-sample *t*-test

These can be found by selecting Analyze > Compare Means. For details of when to use them, relevant hypotheses and test assumptions see Chapter 13.

We will illustrate their application using the independent two-sample *t*-test to test whether there is a statistically significant difference in satisfaction levels between male and female customers in the customer satisfaction.sav dataset.

1. Select Analyze > Compare Means > Independent Samples T-Test.
2. In the dialogue box send Customer satisfaction to the Test Variables box.
3. Send Gender to the Grouping Variable box and Click on Define Groups. A dialogue box opens. Select Use specified values and enter 0 as Group 1 and 1 as Group 2. (Note: the variable is coded 0 = male and 1 = female.) Click Continue. The groups are now shown in the Grouping Variable box (Figure 42).

Figure 42 – Independent Sample T-Test dialogue box
4. Click OK. The output is shown in Figure 43. It includes a test for equality of variance and reports the test results both with and without the assumption that variances are equal. In this case Levene’s test for equality of variance is not significant at the 5% level: the \( p \)-value (Sig. in the output) = 0.631 which is greater than 0.05 so you would assume the variances are equal. The \( p \)-value for the \( t \)-test for the means is given in the column Sig. (2-tailed) and is 0.038. This is less than 0.05 so you would conclude that the difference is statistically significant at the 5% level. See Chapter 13 for further details on the output.

As noted in Chapter 13, you should also comment on the practical significance of the difference. The results can be reported in a simplified version of the table or if referred to in text using the format \( t(\text{df}) = \text{test statistic} \) (t) to 2 decimal places, \( p = \text{p-value} \) to 2 decimal places, so: \( t(18) = 2.24, p = 0.04 \).

**Figure 43 – Independent two-sample \( t \)-test output**

### Group Statistics

<table>
<thead>
<tr>
<th>Gender</th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Std. Error Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>10</td>
<td>4.000</td>
<td>1.02200</td>
<td>0.32669</td>
</tr>
<tr>
<td>Female</td>
<td>10</td>
<td>3.700</td>
<td>1.15900</td>
<td>0.36667</td>
</tr>
</tbody>
</table>

### Independent Samples Test

<table>
<thead>
<tr>
<th>Customer satisfaction</th>
<th>Levene's Test for Equality of Variances</th>
<th>( t )-test for Equality of Means</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( F )</td>
<td>Sig</td>
</tr>
<tr>
<td>Equal variances assumed</td>
<td>2.39</td>
<td>.031</td>
</tr>
<tr>
<td>Equal variances not assumed</td>
<td>2.24</td>
<td>17.764</td>
</tr>
</tbody>
</table>

### 7.2.2 One-way ANOVA

We will illustrate the use of One-Way ANOVA to test whether there is a statistically significant difference in the engagement levels of employees of different ages (Employee engagement.sav).
1. Select Analyze > Compare Means > One-Way ANOVA.
2. In the dialogue box send Age group to the Factor box.
3. Send Engagement to the Dependent List box.
4. Click on Options and select Descriptives and Homogeneity of variance. Click Continue.
5. In the One-Way ANOVA dialogue box, click OK. The output is shown in Figure 44.
The first table in Figure 44 (Descriptives) gives the descriptive statistics for the sample. The second table, labelled Test of Homogeneity of Variances, reports the results of Levene’s test for equality of variance. The null hypothesis for this test is that the variances of the groups are equal. If the result of this test is significant (i.e. $p < 0.05$) we would conclude that variances are significantly different. In this case it is not significant ($0.724 > 0.05$) so we would accept the null hypothesis and conclude that the variances of the groups are equal.

The third table in Figure 44 (ANOVA) contains the results of the ANOVA test of whether or not the population means are different. The test statistic is in the column marked $F$ (i.e. 10.690) and its $p$-value is in the column marked Sig. In this case $p = 0.000$ which is less than 0.05 (in fact it is less than 0.01) so that you would conclude that there is a statistically significant difference between the mean satisfaction levels of different age groups.

The results can be reported in a simplified version of the table or be referred to in text using the format $F$ (between groups df, within groups df) = $F$ statistic to 2 decimal places, $p = p$-value to 2 decimal places, so: $F(4, 170) = 10.690, p < 0.001$. Ensure you comment on the practical significance of the results.
It is important to note that ANOVA tells you whether or not there is a difference in the group means but does not pinpoint where that difference lies. A range of what are known as post hoc tests can be used to test for such differences. These are beyond the scope of this short guide but see further reading for this Chapter for where to find more details. Post hoc tests can be run in SPSS.

### 7.2.3 Mann-Whitney test

The Mann-Whitney test is a nonparametric test that tests for differences in the ranked scores of two independent groups. The null hypothesis is one of no difference between the groups.

The example used here relates to the waiting time in seconds experienced by a random sample of customers from two different branches of a café franchise (waiting times.sav). Examination of the waiting times suggested that they were not normally distributed. It was therefore decided to use the nonparametric Mann-Whitney test in preference to an independent 2-sample \( t \)-test. The test is whether or not there is a difference in the distribution of the waiting times between the two groups. The test is run as follows:

1. Select Analyze > Nonparametric Tests > Legacy Dialogs > 2 Independent Samples\(^2\).
2. Send the dependent metric variable (Waiting times) to the Test Variable List box.
3. Send the independent categorical variable (Group) to the Grouping Variable Box.
4. Click Define Groups and enter the values for Group 1 (enter 1 = Branch A) and Group 2 (enter 2 = Branch B) as defined for the grouping variable. Click Continue.
5. Check that Mann-Whitney U is ticked in the Test Type section of the dialogue box.
6. Click OK. The output is shown in Figure 45.

**Figure 45 – Mann-Whitney test output**

---

\(^2\) This test and other non-parametric tests can also be run using the Analyze > Nonparametric Tests > Independent Samples routine
The first table (Ranks) shows the mean rank of each group, indicating that that group had the highest waiting times. The second table (Test Statistics) reports the test statistic (U) as 3.500 and the 2-tailed p-value (Asymp.Sig.) as 0.000. As the p-value is below 0.05 (in fact it is below 0.01), we would conclude that the difference in waiting times between the customer groups is statistically significant. Remember to comment on the practical significance of the results. (Hint: to help understand the practical significance it is useful to look at the median of each group. You can generate these using the Analyze > Compare Means > Means command and adding Median in the Options dialogue box.)

### 7.3 Carrying out tests of association

Here we will look at how to run the nonparametric and parametric tests discussed in Chapter 13.

#### 7.3.1 Chi-squared test of association, Phi and Cramer’s V

We will demonstrate how to use these tests by running the chi-squared test reported in Chapter 13, testing whether or not driving frequency and household income are associated. The data are contained in the dataset Driving frequency.sav. In this test we will include Phi and Cramer’s V. To run the test:

1. Select Analyze > Descriptive Statistics > Crosstabs.
2. In the dialogue box, send Annual household income to the Row(s) box and Driving frequency to the Column(s) box.
3. Click Statistics and tick the Chi-square and Phi and Cramer’s V boxes. Click Continue.
4. In the Crosstabs dialogue box, click Cells and tick the Row box in the Percentages area (you can also select other boxes if required). Click Continue.

5. Click OK to generate the output (Figure 46):
The first table in Figure 46 (Case Processing Summary) provides a summary of the cases included in the analysis. The second table (Annual household income * Driving frequency Crosstabulation) is the contingency table of the data, showing counts and per cent of row totals. The third table (Chi-Square Tests) contains the test result. The relevant one is marked Pearson Chi-Square and gives a chi-squared test statistic (labelled Value) of 43.685 with 6 degrees of freedom (df) and a p-value (Assymp. Sig.) of 0.000. This is below 0.05 so we would therefore conclude that the association between annual household income and driving frequency is statistically significant.

The table in Figure 46 marked Symmetric Measures contains the output of the Phi and Cramer’s V tests. Phi is relevant for 2x2 tables so as this is a 3x4 table, we should ignore the values forPhi and use Cramer’s V instead. The following effect size descriptors are suggested when reporting the magnitude of the association based on the value of Phi or Cramer’s V:

- 0.1 small effect size
- 0.3 medium effect size
• 0.5 large effect size

The reported value for Cramer’s V is 0.134 which suggests a small effect size for the association between the two variables, even though the relationship is highly statistically significant. As always, be sure to comment on the practical significance of your findings.

Note that below the table SPSS advises you that no expected cell values are below 5. This allows you to check that the assumption regarding the number of cells with an expected frequency below 5 is satisfied by your data.

7.3.2 Fisher’s exact test

When running a chi-squared analysis as described above on a 2x2 contingency table, SPSS includes Fisher’s Exact Test. An example is shown in Figure 47 which tests the association between customer gender and preferred store format from the customer satisfaction.sav dataset.

Figure 47 – Chi-squared test output including Fisher’s exact test
The results are in the Chi-Square Tests table in the row labelled Fisher’s Exact Test. The appropriate p-value is the one marked Exact Sig. (2-sided). This is reported as 0.023, which is below the significance level of 0.05 so we would therefore conclude that the association between restaurant usage and store purchase is statistically significant at the 5% level. Phi is shown as 0.600 in the table marked Symmetric Measures which suggests a large effect size.

7.3.3 Correlation analysis in SPSS

SPSS can be used to calculate Pearson’s r, Spearman’s rho and Kendall’s tau using the Bivariate correlation command. SPSS will also report tests of statistical significance. We will show this applied to measure and test the correlation between satisfaction and commitment in the customer satisfaction.sav dataset. The routine is similar for the other tests.

1. Select Analyze > Correlate > Bivariate to open the Bivariate Correlations dialogue box (Figure 48).
2. Send the variables to be analysed to the Variables: box. (Note: SPSS output is in the form of a matrix that reports all possible paired combinations of the variables included in the analysis so that you can add more than two variables in the Variables: box.)
3. Check that Pearson is selected in the Correlation Coefficients area.
4. Select desired Test of Significance (here: two-tailed).
5. Check Flag Significant Correlations is selected.

Figure 48 – Bivariate Correlations dialogue box

6. Click OK. The output is shown in Figure 49.
Figure 49 – Pearson’s $r$ output

<table>
<thead>
<tr>
<th></th>
<th>Customer satisfaction</th>
<th>Customer commitment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Customer satisfaction</td>
<td>Pearson Correlation</td>
<td>$.799**</td>
</tr>
<tr>
<td></td>
<td>Sig. (2-tailed)</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>20</td>
</tr>
<tr>
<td>Customer commitment</td>
<td>Pearson Correlation</td>
<td>.799*</td>
</tr>
<tr>
<td></td>
<td>Sig. (2-tailed)</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>20</td>
</tr>
</tbody>
</table>

*Correlation is significant at the 0.01 level (2-tailed).

The output reports the correlation coefficient and flags those that are (statistically) significant. Refer to Chapter 13 for guidance on interpreting the strength of the correlation coefficient.

7.3.4 Linear regression

SPSS offers a comprehensive range of tools for carrying out different types of regression including bivariate and multivariate linear regression. We will illustrate its use for bivariate regression by analysing the relationship between satisfaction and commitment in the customer satisfaction database. The researcher has developed the conceptual model shown in Figure 50:

Figure 50 – Conceptual model of the relationship between customer satisfaction and commitment

We will test the following hypotheses, derived from Figure 50, at a 5% significance level:

$H_1$: Customer satisfaction has an impact on customer commitment.

$H_0$: Customer satisfaction has no impact on customer commitment.

To run the test:
1. Select Analyze > Regression > Linear to open up the Linear Regression dialogue box.
2. Send Commitment to the Dependent: box and Satisfaction to the Independent(s): box.
   (Note: if you are carrying out multiple regression you would enter the additional independent variables here as well.)
3. In the Method drop-down box, ensure Enter is selected.
4. Click OK. The output is shown in Figure 51.

Figure 51 – Bivariate regression output

Next step is to interpret the output which SPSS presents in four tables. The first indicates which variables have been entered and removed into the model. In the other tables, pay particular attention to the following:

- In the Model Summary table:
  - $R^2$ (R Square = 0.639).
- In the ANOVA table which tests the significance of the $R^2$: 
  -
o Regression and residual degrees of freedom \((df = 1 \text{ and } 18 \text{ respectively})\).

o The test statistic; for ANOVA this is an \(F\)-statistic \((F = 31.849)\).

o The \(p\)-value \(= 0.000\). Compare the \(p\)-value to your desired significance level (Alpha), in this case 0.05. Since \(0.000 < 0.05\), you would conclude that the \(R^2\) is statistically significant.

- In the Coefficients table:
  o The \(b\) coefficient (Customer satisfaction = 0.892).
  o The test statistic \((t = 5.643)\).
  o The \(p\)-value \(= 0.000\). Compare the \(p\)-value to your desired significance level (Alpha), in this case 0.05. Since \(0.000 < 0.05\), you would conclude that the \(b\) coefficient for Customer satisfaction is statistically significant.
  o The 95% confidence interval \((0.560, 1.224)\) which indicates a range of plausible values for the \(b\) coefficient.

On the basis of this analysis you would reject \(H_0\) at the 5\% level and conclude that customer satisfaction has a statistically significant impact on commitment. As with other tests, you should also comment on the practical significance of your findings.

The results of regression analysis are most easily reported in tabular form.

### 7.3.4.1 Testing the assumptions of regression analysis

As noted in Chapter 13, linear regression analysis requires a number of assumptions to be met. One of those is that the relationship being analysed should be linear. You can use a scatterplot to check this assumption (see above for how to create one). In addition SPSS can be used to test other assumptions. Two of these are:

- Normality of errors. The difference between the observed values and the values predicted by the regression model are known as errors or residuals. The test assumes that these errors are normally distributed for each value of the independent variable. You can test this assumption using a normal probability plot of the residuals. To create this, Click Plots in the Linear Regression dialogue box and tick the Normal Probability Plot box before you run the regression analysis. The output is shown in Figure 52. For the assumption to hold, the points should lie in a reasonably straight line.
• Constant variance of errors. The assumption of homogeneity of variance (homoscedasticity) requires that the variance of the errors is constant across all values of the independent variable. You can check this visually using a plot of standardised predicted values against standardised residuals. To create such a plot Click Plots in the Linear Regression dialogue box. Send *ZRESID to the Y box and *ZPRED to the X box before you run the analysis. The output is shown in Figure 53. For the assumption to hold, the plot should show no observed pattern, for example if the error value changes considerably as the value of the independent variable changes. That does not seem to be the case in this example, so the assumption of constant variance looks reasonable.

For further tests of assumptions, see the recommended further reading for Chapter 13.
7.3.5 Multiple linear regression

The same routine can also be used to carry out multiple linear regression as shown in the Research in practice example in Chapter 13 by entering all of the independent variables into the variables box. This will run the regression with all of the independent variables included. Other options are possible but they are outside the scope of this guide.

Multiple linear regression also requires that your data meet similar assumptions to those of bivariate regression. In addition, it also requires that the independent variables are not too strongly correlated with one another. The term multicolinearity is used to refer to the correlation between independent variables in regression analysis. If there is too much multicolinearity it causes problems with identifying the separate effects of the independent variables. It is possible, for example, to find a multiple regression model where the $R^2$ is statistically significant but none of the individual regression coefficients are. In our experience, multicolinearity is often a problem in student research projects. Partly this is due to a failure to think about the extent to which the independent variables are really conceptually distinct when formulating the conceptual model and partly it is due to poor choice of measurement scales. It is therefore important to check for multicolinearity as part of the analysis. A simple approach is to calculate the correlation coefficient for each pair of independent variables. Correlations above 0.9 suggest high multicolinearity.

More sophisticated diagnostic tests can be run in SPSS that will help detect problems due to the combined effects of two or more independent variables. If your independent variables exhibit high multicolinearity you may have to consider dropping one or more variables from the model or even reverting to bivariate regression and reporting each result separately, making a note of the resulting limitations. You can find more details in the references in the further reading area of Chapter 13.