Graphs and Statistics With SPSS 16.
Contents

1. INTRODUCTION .................................................................................................................. 3
  1.1 VARIABLES ...................................................................................................................... 3

2. STARTING SPSS .................................................................................................................. 5
  2.1 SAVING AND LOADING SPSS DATA ............................................................................. 6

3. GRAPHS ............................................................................................................................... 7
  3.1 THE CHART BUILDER ...................................................................................................... 7
  3.2 VARIABLES AND TYPES OF CHART ........................................................................... 9
  3.3 READING A BOXPLOT ..................................................................................................... 10
  3.4 CHART EDITOR .............................................................................................................. 11
  3.5 EXERCISES FOR GRAPHS ............................................................................................ 12

4. STATISTICAL TECHNIQUES ............................................................................................. 13
  4.1 SPSS STATISTICS OVERVIEW ..................................................................................... 13
  4.2 FREQUENCIES ............................................................................................................... 14
  4.3 CROSSTABS ................................................................................................................... 14
  4.4 INDEPENDENT T-TEST ................................................................................................. 16
  4.5 MANN-WHITNEY U TEST .............................................................................................. 17
  4.6 CORRELATIONS ............................................................................................................. 17
  4.7 ONE WAY ANOVA .......................................................................................................... 18
  4.8 TESTING FOR EQUALITY OF VARIANCE ..................................................................... 19
  4.9 EXERCISES FOR SIMPLE STATISTICS ....................................................................... 20

5. FURTHER READING ............................................................................................................ 24
1. Introduction

This document introduces you to simple statistics and graphs using the software SPSS, version 16, which is available on the Managed XP Service, and can be installed on personal Windows Vista and XP computers and Macs with older versions of OSX. PASW 18 is available for machines with Windows 7 and OSX Snow Leopard.

The document outlines the procedures required to perform statistical analyses on datasets, but it does not teach statistical techniques. It is assumed that the reader knows sufficient statistical theory to decide which analysis is most appropriate for their study. The document then helps the user use SPSS to carry out the chosen analyses.

That said, as the processes outlined in this document depend on variable measurement it is worth revising that.

1.1 Variables

Variables are the characteristics of units, for example the weight of a person, the colour of a person's eyes, the drug that a person is given, or the amount of glass fibre in a pipe. There are various categories of variables as seen below:

<table>
<thead>
<tr>
<th>Variable Type</th>
<th>Typical Values</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Binary/Yes/No</td>
<td>Yes/No Variable</td>
<td>Gender</td>
</tr>
<tr>
<td>Dichotomous</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nominal</td>
<td>More than two categories</td>
<td>Colour, Drink</td>
</tr>
<tr>
<td>Ordinal</td>
<td>Has an order to it</td>
<td>Military Rank</td>
</tr>
<tr>
<td>Counted</td>
<td>Positive integers</td>
<td>Number of children</td>
</tr>
<tr>
<td>Scaled</td>
<td>Wider range of values</td>
<td>Hads</td>
</tr>
<tr>
<td>Continuous</td>
<td>Zero has significance</td>
<td>Temperature</td>
</tr>
</tbody>
</table>

It is helpful to distinguish between three types of variable that you can get in statistical analysis. The first type is **Nominal**. This is a value where the options are qualified by a name only e.g. blood group, favourite pop-groups, colour, group allocated to. When using these types of variables you should not be able to say that one order of values is correct and another order of values is wrong. A special case of this, statistically, is when there are only two possible answers e.g. gender, drinks coffee. A good way to check is to see if you can re-phrase the question so that a simple yes or no response would contain all the information you are interested in. If you find that you want to ask more questions, then your variable is NOT binary.

The next type is **Ordinal**. In this you still have distinct options, but these options have an order. The most common example is social class but it also
covers military rank (or rank in any organisation). Most questions where people are asked how much they like something are ordinal, as are most Likert scales.

Finally there are Scale variables. These not only have order like ordinal variables but also the difference between values has meaning. It is fairly evident that the relationship between an archbishop and a bishop is not the same as the relationship between a curate and a vicar yet these would both be next to each other on a scale. However, the difference between having 2 children instead of 1 child is exactly the same as having 5 children instead of 4. The scale variables are what most people think of as variables. In SPSS these are called continuous.

This may seem trivial, but all analyses require an understanding of the nature of the variables.
2. Starting SPSS

After installation on a PC the software will be available from the Start menu, under Programs. On the Managed XP Service, SPSS is also available from the Programs menu.

When SPSS starts you will see the following dialogue box.

There are five options and you are most likely to either type data into SPSS or open an existing data file. These are covered in the following two sections. You can choose not to load this initial dialogue box, in which case starting SPSS will automatically place you in the Type in data mode.
2.1 Saving and Loading SPSS Data

Once you have entered all your data into SPSS you should save it into a data file. Ideally you should save your data many times before this, say after defining your variables, then after inputting data for several units, and so on until your data is complete.

To save your data file either click the File menu and select Save, or click the usual Save button. You will see a typical Windows Save dialogue. From here you can specify the name for your data file, and use the Save in field to specify the drive and folder in which to save the file. You can even use the New Folder button to create a new folder in which to save your data file.

SPSS data files are usually saved with the extension .sav, but in the Save as type field you can see some other formats for SPSS data. Data can be saved in text files with the extension .txt or .dat, or as Excel data with the extension .xls. You can also save data in other, less common, formats.

When you return to SPSS you can load existing SPSS data files using Open from the File menu, or by using the usual Open button from the Toolbar. All the data files ending in .sav will be available, and you can load any data file by selecting it, then clicking the Open button. In addition you can load data from other applications, as described in the following sections.
3. Graphs

This section is new as SPSS has redesigned the graph interface. This isn’t the first time since SPSS was created, but it is the first time that the old interface has been retired since the Windows version of SPSS.

Before the Windows version, the “plot” command was used to create fairly simple symbol plots (if you want, using caseplot you can still see what these looked like). Next, there was “Graph” in which you selected a graph type and then filled in a dialogue box, this is now being retired and is only available through syntax. Then they added Xgraph which is the “interactive graphics” but it seems that this is also being retired, as it is found under legacy dialogue boxes.

Finally, they have introduced the new graphics interface, ggraph. The interface is surprisingly different as you can see from the shortness of the Graphs menu.

3.1 The Chart Builder

The new interface is called the chart builder and the old graphs at present are in Legacy dialogue boxes. The chart builder, for the first time, uses information about the variables used. When you select the chart builder, you get the following screen:

If you have defined the variable properties click the OK button, you have not or are unsure, you can click the Define Variable Properties button, but there are ways to alter this later if necessary. This dialogue box always appears unless you tick the Don’t show... checkbox.
You will see the Chart Builder dialogue box. Initially, there will be no graph displayed in the Chart preview area.

The **Chart Builder** works by drag and drop. Drag a chart type from the **Gallery** into the **Chart preview** area. If your preferred chart type is not available, you can add new chart type to the **Gallery** using the **Basic Elements** tab.

When you drag a chart type into the **Chart preview** area, the **Element Properties** dialogue box will appear. For the time being, you can ignore this.
The **Chart preview** area in the original dialogue box will display your selected chart type. It will contain blank fields in which you can specify variables.

You can now build your chart by dragging variables from the **Variables** box into the element fields indicated by blue dotted lines in the **Chart preview** area. The element fields vary according to which chart you are drawing. In the pie chart they are called **Angle Variable** and **Slice by**.

You can change the properties of these elements using the **Element Properties** dialogue box, which appeared when you selected a chart type. Once you have the graph in the form you require, click the **OK** button to create the graph in your output.

### 3.2 Variables and types of Chart

**Single variable charts**

<table>
<thead>
<tr>
<th>Measurement</th>
<th>Chart type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nominal</td>
<td>Bar chart</td>
</tr>
<tr>
<td>Ordinal</td>
<td>Bar chart</td>
</tr>
<tr>
<td>Continuous</td>
<td>Histogram</td>
</tr>
</tbody>
</table>
### Bi-variate charts

<table>
<thead>
<tr>
<th>Measurement level</th>
<th>Measurement level</th>
<th>Chart type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nominal</td>
<td>Nominal</td>
<td>Clustered barchart</td>
</tr>
<tr>
<td>Nominal</td>
<td>Ordinal</td>
<td>Clustered barchart</td>
</tr>
<tr>
<td>Binary</td>
<td>Continuous</td>
<td>Population Pyramid</td>
</tr>
<tr>
<td>Nominal</td>
<td>Continuous</td>
<td>Boxplot</td>
</tr>
<tr>
<td>Ordinal</td>
<td>Ordinal</td>
<td>Clustered barchart</td>
</tr>
<tr>
<td>Ordinal</td>
<td>Continuous</td>
<td>Boxplot</td>
</tr>
<tr>
<td>Continuous</td>
<td>Continuous</td>
<td>Scatter plot</td>
</tr>
</tbody>
</table>

#### 3.3 Reading a boxplot

Of all the charts available, the one that causes most difficulty in interpreting is the Boxplot as this is a specialist descriptive chart. I am therefore going to take a single Boxplot and add labels indicating what the various components mean.
3.4 Chart Editor

Having created a graph, it will almost certainly not be perfect, so you will need to adjust it. To do that, double-click the graph to view it in the Chart Editor window.

Use the menus to change or add detail to the graph. Most options will use Element Properties dialogue boxes similar to the one used in creating the graph. You should experiment with the system and find out what it can do. It is a fairly basic graphing package, however, so if you need to produce a very complex or precise graph, you will need to use different software.
3.5 Exercises for Graphs

Demonstration
1. From the course disk, open ‘Favourite Food.sav’ in SPSS.
2. Create a simple barchart for the season of birthday [season] which we have respondents.
3. Create a histogram for the times eaten in a year [Times].
4. Create a clustered bar chart for season of year by salty.
5. Create a boxplot for cost of serving [cost] by gender.
6. Create a scatterplot for cost of serving [cost] by times eaten in the last year (Times).

Exercise
1. From the course disk, open ‘Breast Cancer.sav’ in SPSS.
2. Create a simple barchart for Pathological Tumour size (categories) [Pathcat].
3. Create a histogram for the Pathological Tumour size (cm) [Pathsize].
4. Create a clustered bar chart for Pathological Tumour size (categories) [Pathcat] by Status[status].
5. Create a boxplot for Pathological Tumour size (cm) [Pathsize] by Status[Status].
6. Create a scatterplot of Pathological Tumour size (cm) [Pathsize] by Age (years) [Age]
7. Edit this graph and put a regression line on it.
8. Create a pie-chart for Progesterone Receptor Status[pr]
9. Try to recreate any of the above graphs using the interactive graphs.
10. Try editing the interactive graphs.
4. Statistical Techniques

If you are using SPSS you will want to carry out some statistics. The **Analyze** menu contains many categories. Each of these leads to sub-menus, which have several options. In addition, some of these options can carry out a variety of tests.

4.1 SPSS Statistics Overview

Performing analyses in SPSS is similar to creating graphs; any selection will present you with a complex dialogue box that offers you more control over the analysis performed. Using linear regression as an example you would see:

The variables are listed on the left-hand side, and you use the small arrow buttons to place a selected variable into an appropriate field. Again there are buttons along the bottom of the dialogue box that let you define precisely the type of results you wish to produce.

Once you have defined the analysis click the **OK** button to carry out the tests and send the results to the Output window. A single analysis dialogue may produce many tables of statistics. Again, the more you understand your data the more you can produce meaningful and useful results. The following sections outline some of the basic analyses available.
4.2 Frequencies

**Use for:** Straightforward description of a variable. The options chosen differ whether one is dealing with continuous or categorical variables.

**Limitations:** Tables need turning off for variables with a wide range of responses.

**See:** Field p 70-71

Frequencies is the technique used to get a basic description of the data. It will not only produce a frequency count, but will also calculate a wide range of statistics, and produce bar charts and histograms. To calculate frequencies, click the **Analyze** menu and select **Descriptive Statistics**, then select **Frequencies** from the sub-menu. This will bring up the following dialogue box:

Select a variable to test then use the little arrow button to move this variable into the **Variables** field. Use the **Statistics**, **Charts**, or **Format** buttons to define the required outcome, then click the **OK** button to perform the test.

4.3 Crosstabs

**Use for:** Exploring the relationship of two categorical variables.

**Limitations:** To use the normal chi-squared value you need to have **expected values** of greater than five. This caution is conservative and can be got around by using the exact statistics. Tables with more than 25 cells are clumsy and difficult to interpret.

**See:** Field p 681-694

Crosstabs is the technique used in SPSS to produce cross tabulation of two variables. As this technique treats the values as categories, it is only sensible to use this with categorical data. To perform Crosstabs, click the **Analyze** menu, choose **Descriptive Statistics**, then select **Crosstabs**. This will produce the following dialogue box.
You can choose various tests to apply by clicking the **Statistics** button. Here you will find the Two-way Pearson **Chi-square** test. This is the usual Chi-square test; the one-way Chi-square test, which appears under **Nonparametric tests**, is rarely used.

To avoid getting only cell counts in your table, click the **Cells** button to get the following dialogue box. Here you can specify the type of results that you wish to have displayed in the cells.

Click the **Continue** button to return to the original dialog, then click the **OK** button to perform the Crosstabs analysis.
4.4 Independent T-test

Use for: Comparing the difference in mean value between two groups of one continuous variable when groups are specified by a binary variable. Indirectly this is testing the difference in size between the two groups.

Limitations: Each group is assumed to be normally (Gaussian) distributed!

See: Field p 269-303

If you are comparing the same measurement carried out on two groups of individuals, it is usual to carry out a T-test on the data. To do this, click the Analyze menu and select Compare Means, then from the sub-menu select Independent-Samples T Test. This will produce the following dialogue box

Use the first arrow to select test variables and the lower arrow to select the grouping variable. In order to perform the T-test SPSS needs to know how the two groups are defined by the grouping variable, so you should click the Define Groups button to access the following dialogue box.

Specify the values then click the Continue button to return to the original dialogue box, then click the OK button to perform the T-test analysis.
4.5 Mann-Whitney U Test

**Use for:** Comparing the difference in size between two groups of one continuous variable when the groups are defined by a binary variable.

**Limitations:** This does not test the difference between median values! Each group is not assumed to be normally (Gaussian) distributed.

**See:** Field p 523-534

If the data is thought not to be normal, then you must use a nonparametric test. Click the **Analyze** menu and select **Nonparametric Tests**, then from the sub-menu select **2 Independent Samples**.

![Image of Two-Independent-Samples Tests dialog box]

The test type is set to the Mann-Whitney U test. Again select test variables and grouping variables, then define the groups. Click the **OK** button to perform the test.

4.6 Correlations

**Use for:** Exploring the relationship between two continuous variables. **Pearson’s** assumes normality except for the shared variation; **Spearman’s** when this cannot be assumed for both variables.

**Limitations:** Neither of these will work with nominal variables although you can use the Spearman's with ordinal variables. Binary variables are ordinal.

**See:** Field p 125-130

To see if two variables are related, click the **Analyze** menu and select **Correlate**, then from the sub-menu select **Bivariate**. You will see the following dialogue box.
From here you can also perform a non-parametric test by selecting **Spearman** as the correlation coefficient.

### 4.7 One way ANOVA

**Use for:** Detecting the difference in mean values of different groups where there is more than two distinct groups.

**Limitations:** It is assumed that the populations are normally distributed and have equal variance. It also assumes that the samples are independent of each other, which means that each sample is from a completely separate set of units.

**See:** Field Chapter 8 pages 309 to 362

This is an extension of the T-test for when there are more than two independent groups. It is possible to compare all pairs separately afterwards with multiple comparison tests that are less conservative than doing a Bonferroni correction after doing multiple T-tests.

From the **Analyze** menu select **Compare Means** and then **One-way ANOVA**. You will see the following dialogue box:
In the **Dependent List** section, specify the variable that you are interested in the means of. You may have more than one continuous variable in the dependent list; SPSS will do a separate ANOVA for each continuous variable you put in the dependent list. In the **Factor** box specify the categorical variable that defines the groups that you want to compare.

To compare pairs of groups, click the **Post Hoc** button to produce the following dialogue box.

![One-Way ANOVA: Post Hoc Multiple Comparisons dialog](image)

Normally you should use either the **Tukey** or the **Scheffe**, according to which is popular in your subject. The one exception being if you want to test each of the other treatment against a control, in which case the **Dunnett** seems to be a good choice and if you have unequal number of cases then use the **Games-Howell**.

The others options ticked here may be used if they are accepted by your referees and you have reason to want to use them, on the whole they tend to be mildly conservative. The options that are not ticked should not be used in any circumstance, since this procedure has been developed. They either do not control the error rate, are liberal (being liberal in statistics is a fault worse than being conservative) or are strongly conservative.

### 4.8 Testing for Equality of Variance

*(and what to do if equality is not satisfied in One-way ANOVA)*

ANOVA, with equal number of cases in each group, is fairly robust if the variance within each group is not similar to that in other groups. So with planned experiments you normally do not need to check for the equality of variance.
However, if you are not in the situation to carry out a balanced experiment then you need to first check the equality of variance. If this is shown not to be equal, you should move to an adaptation of the F-Test. To do this from the main dialogue box in One-way ANOVA, click the Options button to get the following dialogue box:

**Homogeneity of variance** just means that each group units have similar variance. This will carry out a Levene’s test for you on the variances. For large samples it may well be worth going to the examine procedure to do a more nuanced analysis as even quite small difference in variance may result in significant results. **Welch** is an alternative test for the F-test which is adjusted for difference in variance.

## 4.9 Exercises for simple statistics

**Demonstration**

1. From the course disk, open the file ‘Favourite Food.sav’
2. Using the **Analyze** menu, subsection **Descriptive Statistics**, obtain the frequencies for season born in (season) and gender (gender). What percentage of the respondents were male?

   ![Image](image.png)

What percentage of respondents were born in the spring and summer?

3. Using the **Analyze** menu, subsection **Descriptive Statistics**, obtain the frequencies for the cost of an average serving (costs) and the times eaten in the last year (times). Make sure at the same time that you also obtain from the
Statistics button the mean, median, standard deviation, minimum and maximum. What are the maximum values of cost and times?

<table>
<thead>
<tr>
<th>Times</th>
<th>Cost</th>
</tr>
</thead>
</table>

4. Using the Analyze menu, subsection Descriptive Statistics, obtain the crosstabulation of season of birthday (season) and whether an individual favourite food is salty (salty). In the Cells have the count and the row and column percentage. Under the Statistics button obtain the chi-squared test. What percentage of people born in the winter liked salty food?

What is the p-value for the chi-squared test?

Is it significant?

5. Using the Analyze menu, subsection Compare Means, carry out an independent t-test for cost of an average serving (cost) between the genders (gender). Which gender on average has more expensive taste in food

What is the t-value

Is this value significant?

6. Using the Analyze menu, subsection Nonparametric Tests, compare to independent samples using the Mann-Whitney U-Test. Are the genders significantly different if we do not assume a normal distribution?

Which has the higher values?

7. Using the Analyze menu, subsection Compare Means to One-Way ANOVA and compare cost by season. What are the mean values and standard error for each season?

<table>
<thead>
<tr>
<th>Spring</th>
<th>Summer</th>
<th>Autumn</th>
<th>Winter</th>
</tr>
</thead>
</table>

Is this a significant result?

Does Levene’s tell us to use Welch-test?

Yes ☐ No ☐
8. Using the **Analyze** menu, subsection **Correlate**, carry out a bivarariate correlation cost of an average serving (cost) by the no of times eaten in the last year (times). What are the pearson’s correlation and the Spearman’s correlation and are they significant?

<table>
<thead>
<tr>
<th>Correlation-type</th>
<th>value</th>
<th>P-value</th>
<th>Significant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spearman</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Is cost and times correlated?

**Exercise**

1. From the course disk, open ‘breast cancer.sav’ in SPSS.
2. Use **Frequencies** to obtain the percentage in the data set of the following:
   a) Positive estrogen receptor status?  
   [ ]
   b) Positive progesterone receptor status?  
   [ ]
   c) Had died?  
   [ ]
   d) A pathological tumour size of less than two cms?  
   [ ]
   e) Had affected the lymph nodes?  
   [ ]

3. Use **Frequencies** to obtain the mean, standard deviation, minimum and maximum for the following variables: Age (in years)[Age], Pathological Tumour Size(cm)[Pathsize] and Time (months)[Time]

<table>
<thead>
<tr>
<th></th>
<th>mean</th>
<th>s.d.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tumour size</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4. Crosstabulate Tumour Size (cat)[Pathcat] by the status[status] obtaining the chi-squared value. Is the result significant?  
   Yes  [ ]  No  [ ]

5. What else from histological Grad [histgrad], estrogen receptor status[er] and progesterone receptor status[pr] might have an effect on the status[status] of the women from this data set?

<table>
<thead>
<tr>
<th></th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Histological grad</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Estrogen receptor status</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Progesterone receptor status</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
6. Obtain a t-test for the age(in years)[age] against status[status]. What is the average age difference between those who died and those still alive(censored) including the confidence interval for this

<table>
<thead>
<tr>
<th>mean</th>
<th>Lower bound</th>
<th>Upper bound</th>
</tr>
</thead>
</table>

Is this result significant?
Yes [ ] No [ ]

7. Obtain a Mann-Whitney U-Test for the age(in years)[age] against which status[status]. Does this give any different result to the T-Test?
Yes [ ] No [ ]

8. Carry out a One-Way ANOVA of time (months) against Pathological Tumour size and fill in the means in the following data?

<table>
<thead>
<tr>
<th>Tumour size</th>
<th>No of cases</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt; 5 cm</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-5 cm</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;= 2 cm</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

What is problematic about this analysis?

9. Obtain all the Pearson and Spearmans correlation for age(in years)[age], Pathologic Tumor Size (cm)[pathsize] and Time (months)[Time]. Are the correlations positive, negative or not significant?

<table>
<thead>
<tr>
<th>Variable 1</th>
<th>Variable 2</th>
<th>positive</th>
<th>negative</th>
<th>Not-sig</th>
</tr>
</thead>
<tbody>
<tr>
<td>age</td>
<td>Tumor Size</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>age</td>
<td>Time</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tumor Size</td>
<td>Time</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

10. Save this as an output file called “Breast Cancer.spo”

11. Go to File and select Export and save to a word file. Try opening the file in Word.

**Homework or for when you are bored**

1. Revise the topics covered in this session and read further either by reading the relevant sections in Field (see course notes) or by working through the SPSS tutorial or by using your own choice of SPSS book and statistic books
2. Take a copy of the data set Clean Air.sav and have a go at doing a good descriptive and two way analysis of the data. This is a data set on clean air in Boston. You may recode or calculate new variables in order to make sense of the data.

3. Produce a report of your results either by using the export (see above) or by cutting and pasting the relevant bits to Word (see course notes).

5. Further Reading

SPSS Books


This is a book that is both a statistics text and an SPSS primer. It covers a large number of techniques (all introduced in this course and more) along with the background theory of how they work. For those who want to go further, I have referenced the relevant pages in Field for each statistical technique covered in this userguide.

**SPSS Made Simple**, Kinnear & Grey, ISBN 0-8637-7350-8, £12.43 for v 15

A good introduction to SPSS, costing about half the price of a single manual. It is written by academics in Aberdeen.


More than a manual, it details why and how to use SPSS for analysing your data. It has become a classic, but costs about twice as much as Kinnear & Grey. If you want a book from SPSS, this is preferable to a manual.

General Statistics Books


A basic introduction to statistical thinking.


This book is a good read, even if the closest you get to statistics is reading what someone else has done. It goes through the basic ways that research may be reported to mislead.