

Good Tables for Excel Users

Contents

1	Introduction	2
2	Examples of tables	3
3	Producing tables	5
4	Using tables	7
4.1	Exploring the data	7
4.2	Aggregating data	7
4.3	Presenting the summary table	8
5	Further topics	11
5.1	Other summaries	11
5.2	Weighting	11
5.3	Multiple responses	12
5.4	Modelling	14
5.5	Adding a measure of precision	15
6	In conclusion	15

October 2002

© Statistical Services Centre

The University of Reading, UK

<http://www.reading.ac.uk/ssc>

1 Introduction

This guide is one of a series on effective statistics for Excel users. We have an overall guide called “Excel for statistics: what you can and cannot do”. There we explain that Excel is good for data entry, organization, and descriptive statistics, i.e. graphs and tables. We suggest you add a statistics package if you need more complicated statistical analyses. Other “good-practice” guides for Excel users are on data entry and graphics.

Excel’s extensive facilities for tabulating data are called “pivot tables”. Everyone who is considering using Excel for statistical work should know about them. Those introduced are surprised how simple they are, and amazed at their power.

To construct a pivot table, your data must be in the standard row-column form that Excel calls a “list”. This is also the starting point for those who wish to do simple statistics with Excel, using our add-in, called SSC-Stat. Our tutorial on SSC-Stat emphasizes the importance of “lists” for the data layout if you are doing any statistical work. The key points in that tutorial form a useful introduction to this guide for those who are beginners in using Excel for data analysis.

In Section 2 we give an example of data in list format and also of typical tables. Excel’s facilities for tabulation consist of a wizard to help you produce the initial table, as described in Section 3. They are followed by facilities to edit the table, so it is in the desired form for your use.

These tabulation facilities parallel those for producing a graph. There you again start with a wizard, and then edit the resulting axes, legends and so on. Graphs are used for two purposes. There are exploratory graphs, to help you, in evaluating your data. Then there are presentation graphs that form part of your report to others.

Tables are the same. We show in Section 4 that Excel tabulation facilities can help with data exploration. Presentation is also important and we see how Excel is used to produce tables for reports.

With large data sets the analysis often proceeds in stages. In this case summaries of the raw data become the information for the next part in the analysis. Producing these summaries is a third use of tabulation. This is sometimes called “aggregating” the data and we give an example in Section 4.

In Section 5 we describe some analyses that need more than Excel’s pivot tables can provide easily. The simplest procedure is then to add a statistics package. There are many, including EpiInfo, SPSS, SAS, or Stata.

2 Examples of tables

Fig. 2a shows some data from a simple survey. There were 36 farmers from 4 villages. The information in column G is the yield of rice and other columns indicate how much fertilizer they applied, and which variety of rice was used.

Fig. 2a An example of survey data

	A	B	C	D	E	F	G	H	I
1	Village	Field	Size	Fertiliser	Fertlevel	Variety	Yield	infected	insects
2	Sabey	3	2	2.5	3: >2	Old	53.6	yes	144
3	Sabey	13	5	1.5	2: 0.5-2	Old	44.6	yes	340
4	Sabey	4	5	3	3: >2	Old	50.7	no	0
5	Sabey	20	1.5	1	2: 0.5-2	Trad	33.6	yes	489
6	Sabey	19	5	2.5	3: >2	New	62.1	no	0
7	Sabey	10	4	1.5	2: 0.5-2	Trad	30.6	no	0
8	Sabey	8	4.5	2	3: >2	Old	37.7	no	0
9	Sabey	11	3.5	0	1: none	Trad	24.3	no	0
10	Sabey	14	5	2	3: >2	New	56.8	no	0
11	Sabey	12	4.5	2.5	3: >2	Old	59.3	yes	35
12	Kesen	9	1.5	2	3: >2	Old	40.4	no	0
13	Kesen	8	7	0	1: none	Trad	25.8	no	0
14	Kesen	1	2.5	1.5	2: 0.5-2	Old	40.7	no	0
15	Kesen	4	8	0	1: none	Trad	27.6	yes	107
16	Kesen	6	4.5	2.5	3: >2	Old	48.7	no	0
17	Kesen	3	2	0.5	2: 0.5-2	Trad	27	no	0

Figs 2b to 2d show three simple pivot tables. The first is a **one-way table** showing how many of the 36 farmers used each of the 3 varieties. This is sometimes called a “tally” because it just gives a frequency count.

Fig. 2b A one-way table of frequencies

	A	B
1		
2		
3	Count of Variety	
4	Variety	Total
5	New	4
6	Old	17
7	Trad	15
8	Grand Total	36
9		

The second and third are both **two-way tables**. *Fig. 2c* gives the count of farmers using the 3 varieties in each of the 4 villages. We will see in a later section how these data can be displayed as percentages

Fig. 2c A two-way table of frequencies

	A	B	C	D	E	F
1						
2						
3	Count of Variety	Village				
4	Variety	Kesen	Nanda	Niko	Sabey	Grand Total
5	New		2		2	4
6	Old	3	7	2	5	17
7	Trad	4	5	3	3	15
8	Grand Total	7	14	5	10	36
9						

Fig. 2d gives the mean yield of the farmers in each group. Using both figures together, we see that the 2 farmers who were from village **Nanda** and used the **New** variety, had a mean yield of 59.8 q/ha¹ (almost 6 tons per hectare).

Fig. 2d A two-way table of means

	A	B	C	D	E	F
1						
2						
3	Average of Yield	Village				
4	Variety	Kesen	Nanda	Niko	Sabey	Grand Total
5	New		59.8		59.5	59.6
6	Old	43.3	46.4	36.1	49.2	45.4
7	Trad	24.9	36.3	26.9	29.5	30.0
8	Grand Total	32.8	44.7	30.6	45.3	40.6
9						

So the general idea is that some columns of data may be “**categorical**”. These are sometimes called “**class variables**” or “**factors**”. Examples of categorical data are

- Yes/No
- Excellent/Good/Mediocre/Bad/Awful
- Young/Middling/Old
- England/Scotland/Wales/NI

In the survey data shown in **Fig. 2a** the two columns of categorical data we used were those that indicated the village and the variety. Often, as in the examples above, these columns will be given in words, but this is not essential. In **Fig. 2e** we have a table with a single categorical column, showing the amount of fertilizer applied. Here we have three “statistics” that give the number and percentage of farmers applying each level of fertilizer, and also their mean yield.

¹ A quintile is 1/10 of a ton.

Fig. 2e A table showing three summaries

	A	B	C	D
1				
2				
3		Data ▾		
4	Fertiliser ▾	Count	Percentage	Mean Yield
5	0	9	25%	27.4
6	0.5	2	6%	33.7
7	1	2	6%	39.9
8	1.5	6	17%	38.3
9	2	8	22%	44.4
10	2.5	6	17%	56.5
11	3	3	8%	47.6
12	Total	36	100%	40.6
13				

We sometimes distinguish between “ordered categorical data” and “nominal data”. An example above of ordered data is the amount of fertilizer, or “Young/Middling/Old” where we would like the table to keep this order and not have the categories in alphabetical order. With nominal data, there is no particular order, so it may not matter what order Excel uses.

In the two-way table in **Fig. 2c**, the frequencies are the same as in the last column of **Fig. 2b**. This is the “Total” or “Margin” or “All” column of the table in **Fig. 2b**. We usually use the word “Total” in this guide. We see that the totals of a two-way table are one-way tables and so on.

The word “Total” is not entirely satisfactory, as is seen in **Fig. 2d**, where it gives the means² (rather than the totals) for each village or variety. This is why some software uses other terms, such as the “Margin”.

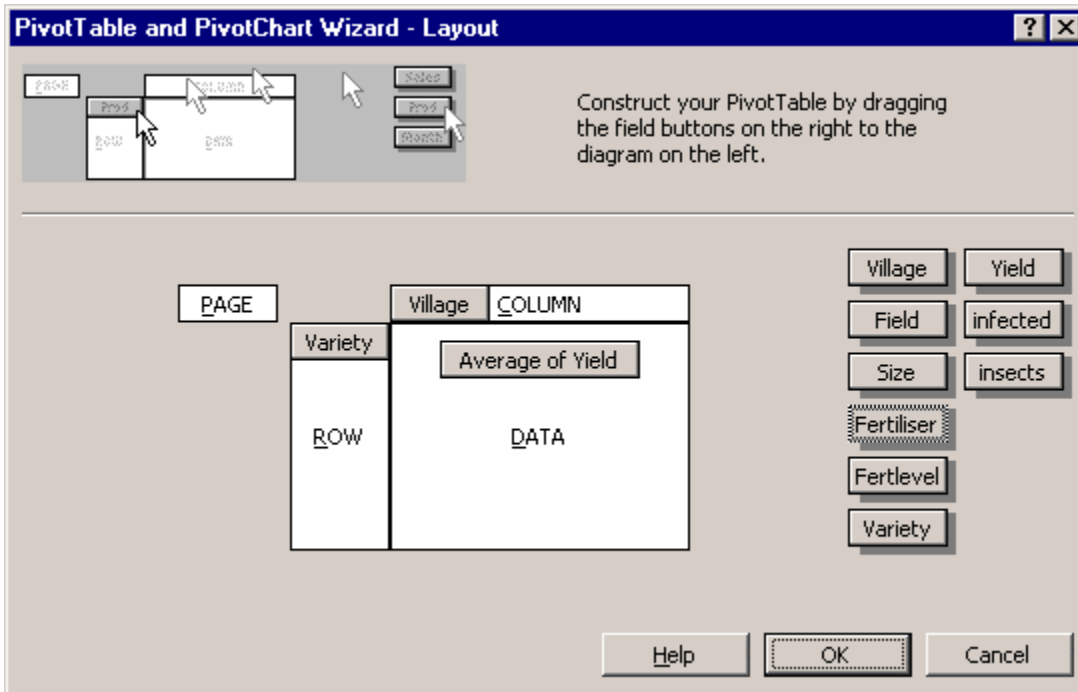
3 Producing tables

To produce a pivot table, start from the sheet containing the data and use **Data => Pivot Table** and **Pivot Chart Report**. This launches the Table Wizard, from where the key step is to define the initial layout³ of your proposed table, as shown in **Fig. 3a**. Here you specify which (categorical) columns will give the rows and columns of the table. Then you show what is to be tabulated. In this example, we have set up the screen for the table in **Fig. 2d** above.

² The name can be edited, for example to “Mean”. Once changed anywhere, then all occurrences within the table change automatically.

³ The pivot table wizard has three steps. The layout button is on the third step.

Fig. 3a One screen from Excel's pivot table wizard



When analyzing a survey you may wish to produce multiple tables on the same sheet. One way is simply to make copies of the original table. They can then each be modified to present the data you would like. **Fig. 3b** shows the same information as in **Fig. 2c**, but modified, to give the row and column percentages. From **Fig. 2c** there were 7 farmers in Village Nanda, who used the Old variety of rice. From the first table in **Fig. 3b** we see that they were 50% of the Nanda farmers, while the second table informs us that they were 41% of the farmers using this variety.

Fig. 3b Tables of row and column percentages

	A	B	C	D	E	F	G	H	I	J	K
1											
2											
3	Village %	Varie ▾					Variety %	Varie ▾			
4	Village ▾	New	Old	Trad	Total		Village ▾	New	Old	Trad	Total
5	Kesen	0%	43%	57%	100%		Kesen	0%	18%	27%	19%
6	Nanda	14%	50%	36%	100%		Nanda	50%	41%	33%	39%
7	Niko	0%	40%	60%	100%		Niko	0%	12%	20%	14%
8	Sabey	20%	50%	30%	100%		Sabey	50%	29%	20%	28%
9	Total	11%	47%	42%	100%		Total	100%	100%	100%	100%

To produce the percentages in a table is one example of modifying the information. One way to do this is to right-click anywhere in the table and then choose Field Settings and then Options. Alternatively double-click in the cell A3 in **Fig. 3b**, to go straight to the options menu.

4 Using tables

In this section we indicate how tables are used for data exploration, aggregation and for presentation.

4.1 Exploring the data

It is useful to look at individual observations in more detail. This may be because certain categories are particularly interesting, or perhaps because they are surprising, and may indicate oddities in the data.

Double clicking on any cell of the table shows all the data for just the observations that make up the cell. They are produced on a new sheet. This is sometimes called “drill-down”. An alternative way is to right-click on the cell and then choose “**Group and Outline, => Show Detail**”

As an example, we show the data for the 7 farmers from Nanda that used the Old variety in *Fig. 4a*

Fig. 4a Using the “drill-down” facility to see individual rows of data

	A	B	C	D	E	F	G	H	I
1	Village Field Size Fertiliser Fertlevel Variety Yield infected insects								
2	Nanda	25	4	2	3: >2	Old	42.4	yes	348
3	Nanda	14	2.5	1.5	2: 0.5-2	Old	45.8	yes	148
4	Nanda	26	4	2	3: >2	Old	41.3	no	0
5	Nanda	24	6	1.5	2: 0.5-2	Old	42.2	no	0
6	Nanda	12	3	1	2: 0.5-2	Old	46.2	no	0
7	Nanda	21	4	3	3: >2	Old	49.3	no	0
8	Nanda	2	3.5	2.5	3: >2	Old	57.4	yes	403

4.2 Aggregating data

Data are available at different “levels” in a study. For the rice survey the information in *Fig. 2a* is all at the “farmer level”. There may also be information at the “village level”, for example the location, distance to a main road, whether there is a school or bank in the village and so on.

Similarly an educational survey may have some information at the local authority level, the school level, the classroom level and the pupil level. A medical study may have information about individual patients and also about the visits that each patient made to a hospital.

Where information is available at multiple levels, this usually corresponds to multiple tables in a database package. In a spreadsheet the information from each level is often in its own rectangle and each level may be on a separate sheet.

Tabulating the data then provides a powerful tool for moving information from a lower level⁴. For example, in *Fig. 4b* we show some summaries that are at “village level”. For example the mean yields in the two villages, Kesen and Niko, are lower than in the other two villages. We see also that the mean fertilizer applied is lower.

⁴ If the SSC-Stat add-in is installed an alternative is the **Analysis => Summarise** dialogue. This permits a wide range of statistics including percentiles and presents the results as one-way tables, similar to Fig. 4b.

Fig. 4b Aggregating data from the farmer to the village level

Village	Count	Mean Fertiliser	Min Yield	Mean Yield	Max Yield
Kesen	7	0.9	19.1	32.8	48.7
Nanda	14	1.9	25.8	44.7	61.4
Niko	5	0.1	24.7	30.6	40.4
Sabey	10	1.9	24.3	45.3	62.1

These summaries could now be merged with other data that was collected at the village level.

4.3 Presenting the summary table

There are many ways in which the appearance of the information in the table can be changed. If all or part of the table is selected, then the Format menu may be used to format either the data or the cells. This was done in many of the tables given in earlier sections, to specify the width of the columns and the number of decimals.

Fig. 4c Examples of tables after using autoformat

Village	Variety	Mean Yield	Mean Yield	Village
Kesen				
	Trad	24.9	Trad	Kesen
	Old	43.3	Old	Niko
Niko			New	Nanda
	Trad	26.9		Sabey
	Old	36.1		
Nanda				
	Trad	36.3		
	Old	46.4		
	New	59.8		
Sabey				
	Trad	29.5		
	Old	49.2		
	New	59.5		

An auto-format or a conditional format can also be used, and two examples are shown in **Fig. 4c**. Once the table is as you would like it, then it can be copied to Word or Powerpoint as part of a report or presentation⁵.

The name “pivot-table” signifies the idea that the rows can be dragged to become the columns, etc. If you were to decide that a variable was not needed in the table, it can be removed. And other category columns can be added.

Sometimes a table becomes much clearer if the categories making up the rows or the columns are reordered. An example is in **Fig. 4d** and shows a **three-way table** ordered so the villages using the new variety are at the top. The data are also to some extent “triangular” and this

⁵ Take care, however in copying from Excel. The default is to copy the whole Excel workbook. This has the advantage that the table can then be edited further from within Word or Powerpoint. However, this may be a heavy overhead if the data set is large. We often do a “paste special” to copy just the values.

corresponds to the fact that farmers growing the new varieties used more fertilizer. The re-ordering of the categories in a table can either be done manually⁶, or using **Excel's Data => Sort** facility⁷.

Fig. 4d A three-way table, with villages in a user-defined order

1						
2						
3	Mean Yield		Fert			
4	Village	Variety	1: none	2: 0.5-2	3: >2	Grand Mean
5	Sabey	New			59.5	59.5
6		Old		44.6	50.3	49.2
7		Trad	24.3	32.1		29.5
8	Sabey Mean		24.3	36.3	53.4	45.3
9	Nanda	New			59.8	59.8
10		Old		44.7	47.6	46.4
11		Trad	37.6	25.8	39.3	36.3
12	Nanda Mean		37.6	40.0	47.5	44.7
13	Kesen	Old		40.7	44.6	43.3
14		Trad	24.2	27.0		24.9
15	Kesen Mean		24.2	33.9	44.6	32.8
16	Niko	Old	31.8	40.4		36.1
17		Trad	26.9			26.9
18	Niko Mean		28.1	40.4		30.6
19	Grand Mean		27.4	37.7	49.2	40.6

Categories can also be omitted or grouped⁸. There is also flexibility on whether to show all, some or none of the totals. As an example **Fig. 4e** shows the same as **Fig. 4d**, but with the villages grouped.

Fig. 4e Data grouped to simplify the presentation

1						
2						
3	Mean Yield		Fert			
4	Village2	Variety	1: none	2: 0.5-2	3: >2	Mean
5	Sabey/Nanda	New			59.6	59.6
6		Old		44.7	49.0	47.5
7		Trad	31.0	30.0	39.3	33.7
8	Kesen/Niko	Old	31.8	40.6	44.6	40.4
9		Trad	25.5	27.0		25.7
10	Mean		27.4	37.7	49.2	40.6
11						

The parallel between the graphs and the tables is made even clearer by the ease with which a chart can be produced. In **Fig. 4f** we show a table of the mean yields for each level of fertilizer and each variety. The corresponding Pivot chart is in **Fig. 4g**.

⁶ Click on the cell to move and drag the yellow section.

⁷ You can sort within a pivot table. Mark the cell, or the part of the table that will be used and then use **Data => Sort**.

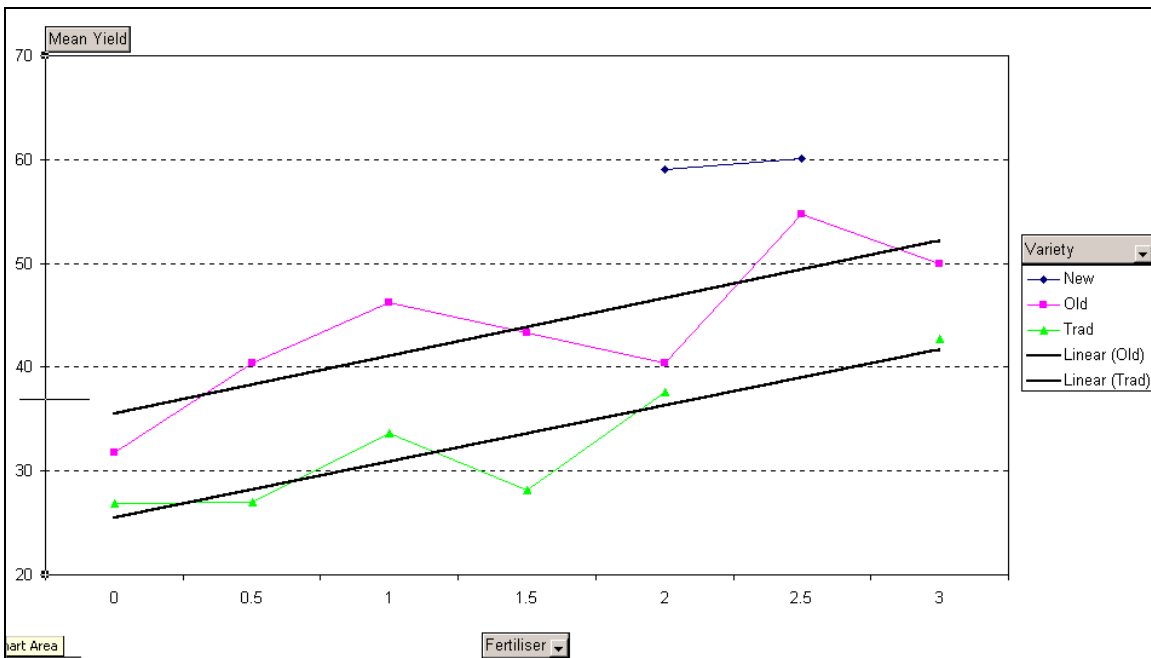
⁸ To group categories, select those to group and right-click. Then choose **Groups** and **Outline => Group**.

Fig. 4f Mean yields by fertilizer and variety

1					
2					
3	Mean Yield	Variety			
4	Fertiliser	New	Old	Trad	Mean
5	0		31.8	26.9	27.4
6	0.5		40.4	27.0	33.7
7	1		46.2	33.6	39.9
8	1.5		43.3	28.2	38.3
9	2	59.1	40.5	37.7	44.4
10	2.5	60.1	54.8		56.5
11	3		50.0	42.7	47.6
12	Mean	59.6	45.4	30.0	40.6

These two methods of presentation complement each other. The table shows the actual values, and also has extra columns for the margins. The graph is easier to assimilate and also has the added lines that indicate the trend with increasing fertilizer.

Fig. 4g Data as in Fig. 4f, given as a pivot chart



5 Further topics

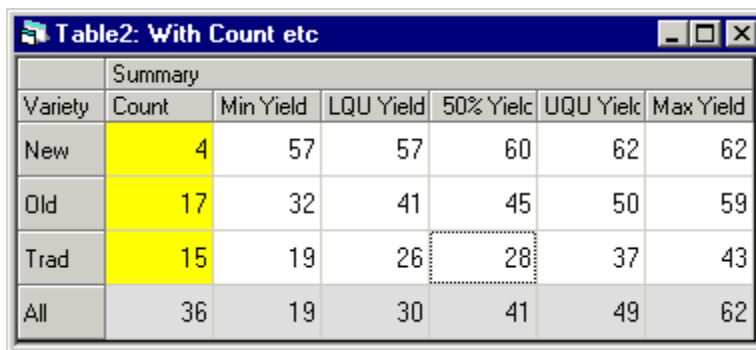
In this section we describe some topics concerned with tables that exceed Excel's capabilities, at least those that can be done easily. The user can then either struggle with the software, with which they are familiar, or move to other software.

5.1 Other summaries

Excel permits tables of frequencies, including percentages, or summary statistics. It includes a limited set of summary statistics however. For example in **Fig. 2d** we showed a table of the mean yield for each variety and village. There is however no function to produce a pivot table with medians.

An example for the data used in this guide is in **Fig. 5a**⁹. There we see that overall the median yield was 41 q/ha, with a quarter of the farmers having a yield in excess of 49 q/ha.

Fig. 5a A table showing medians and other percentage points



Summary						
Variety	Count	Min Yield	LQU Yield	50% Yield	UQU Yield	Max Yield
New	4	57	57	60	62	62
Old	17	32	41	45	50	59
Trad	15	19	26	28	37	43
All	36	19	30	41	49	62

5.2 Weighting

We emphasise the importance of weighting data when constructing tables in the SSC guide titled "Approaches to the Analysis of Survey Data". Weights are important if the way that units were sampled does not reflect their presence in the population. Usually the weights will be provided as an additional column of data. Statistics packages, such as SPSS permit these weights to be used, when calculating summaries in tables. In **Fig. 5b** we show the same table as in figure **Fig. 2d**, but where the yields have been weighted by the size of each farmer's field¹⁰. This was also recorded as the third column, called **Size**, in the data shown in **Fig. 2a**.

⁹ We used Instat+ to produce this figure, and some others in this section. Other statistics packages, such as SPSS could equally have been used.

¹⁰ The way the data were collected is also important if we wish to calculate the means in a way that is "fair". If farmers were equally likely to be included in the sample, then the weights, used in **Fig. 5b** should be used if we later wish to be fair to each hectare. If we sampled farmers with probability proportional to their field size, then we should not weight again.

Fig. 5b Mean yields weighted by the size of the field

Variet	Village				
	Kesen	Nanda	Niko	Sabey	All
New		59.3		59.5	59.4
Old	44.9	45.9	36.1	48.6	44.9
Trad	25.3	39.0	27.8	28.6	32.0
All	31.0	43.1	31.5	46.8	39.9

5.3 Multiple responses

In some surveys there are multiple response questions. For example a health survey might include the question:

List the reasons you were away from work during the past month?

This can alternatively be phrased as a series of “multiple dichotomy” questions, each requiring a Yes/No answer, such as:

Were you away from work during the past month because of:

- Illness
- Holiday
- Children
- Other reasons

A simple example on forms of exercise is shown in **Fig. 5c**. Here the first columns of data correspond to some general questions, the sex and age group of the respondent. Then the multiple response question concerns whether the person did exercise of any of three different types in the last week.

Fig. 5c A simple example of multiple response data

	X1	X2 - F	X3 - F	X4 - F	X5 - F	X6 - F
	Person	Sex	Age	First	Second	Third
1	1	Male	51+	Walking	Cycling	Jogging
2	2	Male	<21	Cycling	*	*
3	3	Male	<21	*	*	*
4	4	Male	21-50	Jogging	Cycling	*
5	5	Male	<21	Cycling	Walking	*
6	6	Female	51+	Walking	Jogging	*
7	7	Female	21-50	Walking	*	*
8	8	Female	21-50	*	*	*
9						

Alternatively, in **Fig. 5d** the information on exercise is coded in a way that corresponds to a sequence of simple questions, like “Did you walk?”

Fig. 5d An alternative input of multiple response data

	X1	X2 - F	X3 - F	X4 - F	X5 - F	X6 - F
	Person	Sex	Age	Walking	Jogging	Cycling
1	1	Male	51+	Yes	Yes	Yes
2	2	Male	<21	No	No	Yes
3	3	Male	<21	No	No	No
4	4	Male	21-50	No	Yes	Yes
5	5	Male	<21	Yes	No	Yes
6	6	Female	51+	Yes	Yes	No
7	7	Female	21-50	Yes	No	No
8	8	Female	21-50	No	No	No

In **Fig. 5e** we show the sort of table that might be required. The new feature of this table is that the three columns in **Fig. 5d** have been recorded as a single dimension of the table. We see in **Fig. 5e** here that the 8 respondents recorded 11 activities. The calculation of percentages is also more “interesting”, because they may be given either in relation to the number of respondents, or the number of responses.

Fig. 5e Number of respondents

	Sex		
	Male	Female	All
Walking	2	2	4
Jogging	2	1	3
Cycling	4	0	4
Responding	4	2	6
Respondents	5	3	8
All	8	3	11

Fig. 5f Percentage of respondents

	Sex		
	Male	Female	All
Walking	40	67	50
Jogging	40	33	38
Cycling	80	0	50
Respondents	5	3	8
Responses	8	3	11
All	160	100	138

For example, in **Fig. 5f** we see for the males that the percentages for walking jogging and cycling are 40%, 40% and 80%. This corresponds to the fact that 2 out of the 5 males walked or jogged, while 4 of them cycled, see **Fig. 5e**. In all therefore there were 8 responses for the 5 males, and this corresponds to the 160%¹¹.

¹¹ We do not suggest that one would usually present such small numbers in percentage terms. The example is used merely to illustrate the concepts.

5.4 Modelling

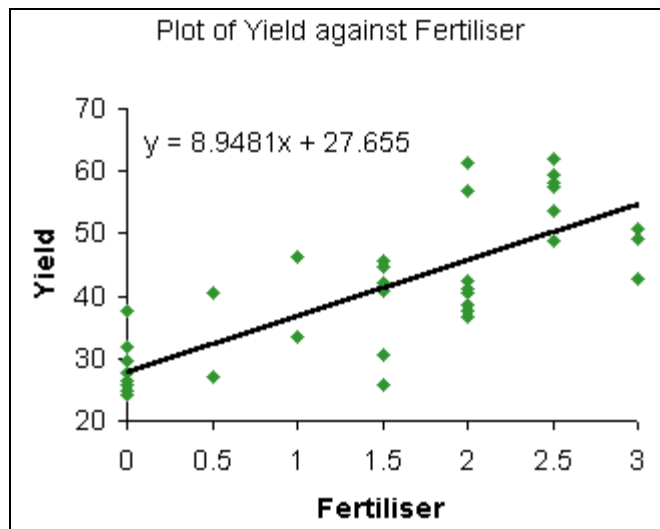
The final themes that we explore are crucial for some studies. In this guide we have so far considered only descriptive statistics. Often however the data for analysis have been sampled from some larger population, and we would like our conclusions to apply to the population and not merely be a summary for the particular observations in the sample.

One aspect of this generalization concerns exactly which tables, or graphs need to be produced in a report. We should try to summarise the data as simply as possible. However tables that are too simplistic may mislead the reader.

For example we could use a chi-square test with a 2-way table of frequencies, to check whether the two categorical variables are related, e.g. **Fig. 2c**. If not, then the one-way table in **Fig. 2b** would provide a simpler summary of the data. Now the more modern approach is called log-linear modelling. This is roughly the same as the chi-square test, but is much more general, because it is not limited to just 2-way tables. All statistics packages include facilities for log-linear modelling.

An second example is the two-way table shown in **Fig. 4f**, and in graphical form in **Fig. 4g**. In **Fig. 4g** the two main lines look close to being parallel. If a parallel-line model is sufficient then the table in **Fig. 4f**, and the three lines in **Fig. 4g** are more complicated than is needed¹². In that case we challenge the reader to assess whether the graph in **Fig. 5g** would be sufficient to indicate that a unit of fertilizer would increase the yields by an estimated 8.9 units¹³. The effect of variety would be reported as a small separate table, or perhaps included within the text of the report¹⁴.

Fig. 5g A simpler graph than Fig. 6d



Simple examples, such as the above, are within the framework of what is called “linear models”. This includes regression modelling and analysis of variance that are also in all statistics

¹² The “comparison of regressions” analysis does indicate the adequacy of the parallel lines model. There is no evidence to support the more complicated model where each variety has a different slope.

¹³ It is not sufficient. Using a statistics package to fit the parallel line model gives a much lower slope of 5.26 units. (The reason for the difference is that the improved varieties also tend to have more fertilizer added, and so bias the slope, if they are not included in the model.)

¹⁴ The means for zero fertilizer are estimated to be 30.0, 35.7 and 47.8 for the Trad, Old and New.

packages. More recently, many statistics packages also include facilities for “generalized linear models”, and these allow the same types of decision to be made for situations where the distribution of the variable of interest is other than normal

5.5 Adding a measure of precision

Once the data are to be considered as a sample, then it is also useful to give a measure of precision¹⁵ of the summary values that are presented in tables and graphs. For example, in addition to knowing the effect of fertiliser in the parallel-line model is a slope of 5.26, it is sometimes useful to know what degree of confidence can be attached to this value.

If the data may be assumed to be a simple random sample then the calculation of measures of precision is covered in statistics courses and is part of the routine output when using any statistics package¹⁶. However many sample surveys are designed in a more complicated way. It is still possible for standard errors to be calculated, but only a few packages, such as Stata or SAS include such facilities.

6 In conclusion

Apart from the few paragraphs at the end of Section 7, this guide is concerned with descriptive statistics. We suggest that many training courses could cover the topics described here in more detail. This would help trainees to understand basic concepts better¹⁷. The materials covered here are also a useful component of data summary and presentation in many real studies.

The presentation of results from any study is usually a mixture of tables, graphs and results quoted in the text. We give some information on the preparation of good tables for reports in our guide titled “Informative presentation of tables graphs and statistics”. Users are often prepared to spend considerable time on making graphs as presentable as possible. We suggest that the preparation of presentable tables is equally important.

Finally the preparation of tables can be used to motivate the teaching of inference and modelling methods as we indicate at the end of Section 5.

¹⁵ This is given either in the form of the standard error, or as confidence limits.

¹⁶ The standard error of the slope for the parallel line model is then calculated to be 0.96 so the 95% confidence limits for the slope are about 4.30 to 6.22.

¹⁷ We find many who have formal training in statistics but remain confused by percentages, and percentiles. They are natural topics to cover when teaching about the tabulation of data.