

## Introduction

The analysis and reporting stage of your survey is extremely important. All too often, piles of questionnaires are left sitting in boxes while people wonder about the best way of dealing with them. The longer the questionnaires sit there, the more out of date the information they contain can become. It is therefore essential to begin work on your analysis as soon as possible.

## Forward planning

It is important to think about the analysis stage of your research before starting to collect the data as this will help you to identify what type of information you need to gather. For example, do you want to report results as percentages or calculate average scores? (See also the section on summarising data later in this guideline). It is not essential to have your entire analysis planned in advance as the data can sometimes present ideas you might not have previously thought about. However, you might need to consider the following:

- **Question scales** (see also the guideline on 'Writing questions'). For example, if you want to calculate a mean score then you will need to use a numerical scale.
- **Comparisons with previous surveys.** If you want to compare your results with a previous year then you need to make sure that the question wording and scales are the same.
- **The sub-groups you are interested in.** If you want to examine differences in the results by geographical area you need to ensure that your sample is made up of a sufficient number of people from each area and that you include a question (e.g. postcode) to enable you to carry out this analysis.
- **The number of 'open' questions** (see also the guideline on 'Writing questions'). Having to read through and 'code' too many of these can dramatically increase your analysis time.

## Coding

All survey responses need to be 'coded', but the coding of 'closed' and 'open' questions is a different process. In the case of 'closed' questions, coding simply involves assigning a numerical value to each response (e.g. yes = 1, no = 2) since the range of available response options are already known. This allows for more effective data entry and analysis. Coding of 'open' questions involves drawing up a list of categories (a 'coding frame') into which answers can be allocated or 'coded'. This process is time-consuming and requires careful attention.

## Data entry

Coded data can either be scanned or entered by hand. Scanning can be expensive, specialist software is required and the questionnaire would also need to be set up in 'scannable' format. On the other hand, scanning is usually quicker, with less room for human error. Manual data entry is cheaper in the short term, but more time-consuming and can be subject to input error. However, specialist packages such as Snap, Pinpoint or SPSS Data Entry reduce the scope for this error considerably.

## Data cleaning

It is recommended that you 'clean' your data (i.e. look for and correct any errors in the data set) before carrying out any work on it. This involves running a few basic checks:

- Looking for duplicates in any variables which should be unique (e.g. an ID code).
- Looking at the minimum and maximum values for each variable (e.g. a response of '6' when 1 to 4 are the only possible options suggests an error).

- Looking at the 'likeliness' of responses (e.g. if an 'age' value is listed as 555 this is likely to be a keying error).
- Looking for responses to questions that should have been skipped via routing (e.g. 'if Q6 = no, go to Q10' but Q7-Q9 have been answered).

① *Scanning and specialist data entry packages will reduce error since they can be set up to reject invalid responses.*

### Software for data analysis

A range of specialist software is available for undertaking quantitative data analysis, although Microsoft Excel is capable of running basic descriptive statistics as well as a range of more complex statistical analyses. Specialist packages are not necessary, but can be more user-friendly. We use SPSS for the analysis of quantitative data.

### Summarising data

A range of tools (descriptive statistics) can be used to describe/summarise data, including:

- **Frequency** – the 'count' or number of responses in each category.
- **Percentage (%)** – the proportion of people who give a particular response or set of responses.
- **Mean** – the arithmetic average (used with standard deviation to illustrate variance).
- **Standard deviation** – the average spread of scores around the mean.
- **Median** – the 'middle value'.
- **Mode** – the most common response to a question.
- **Range** – the difference between the highest and lowest response value (e.g. range of ages).

It is good practice to report the base (i.e. the number of respondents who have answered a particular question) and the number of non-responses when using these tools to describe your data.

### Presenting data

The way in which you present your data should make the results clear and easy to see at a glance. Tables and charts are two of the main ways of presenting a summary of your data, and a few tips on using these in research reports are outlined below:

- Try to avoid too much detail on percentages – sample data is not accurate to this level and rounded figures have more of an impact (e.g. use 80% as opposed to 79.85%).
- Use bold type to make important figures or text in the table stand out.
- Align figures to the right as this makes them easier to read.
- Don't try to put too much information on one chart.
- Use a sensible scale on charts – don't try to mask or emphasise certain results.
- Only use pie charts when the data add up to 100%.
- Remember that some people will only have a black and white copy of the report, so choose colours for charts carefully.

The following examples illustrate how the presentation of the same data can be improved through different use of tables and charts:

## Example 1 – improving tables

The figures in the following table are difficult to read due to the centre alignment, unnecessary decimal points and % signs in each cell. Single spacing makes the table appear cramped and the number of respondents involved is not shown:

Very satisfied	25.58%
Quite satisfied	51.94%
Neither satisfied nor dissatisfied	14.73%
Quite dissatisfied	4.26%
Very dissatisfied	3.49%

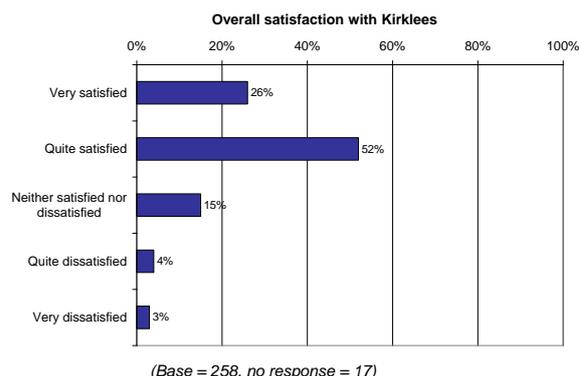
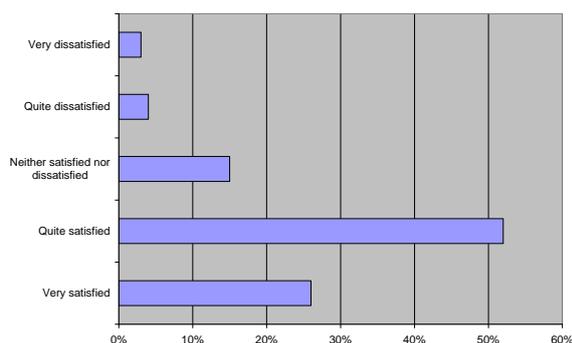
Adding the base, frequency of each response and number of non-responses helps interpretation. Rounded, right-aligned figures and increased spacing make the results easier to see at a glance. Finally, a title makes the content of the table clear:

Overall satisfaction with Kirklees	Freq	%
Very satisfied	67	26
Quite satisfied	134	52
Neither satisfied nor dissatisfied	39	15
Quite dissatisfied	10	4
Very dissatisfied	8	3
<b>Base</b>	<b>258</b>	

(No response = 17)

## Example 2 – improving charts

The first chart (a default Excel bar chart) can be improved by removing the grey background, adding % values to the bars, altering the scale to run from 0-100%, adding a title and reporting the base and number of non-responses at the bottom of the chart:



## Calculating sampling error

Survey results from a sample are an approximation of the results for the whole population and are unlikely to be exactly the same if the survey were to be repeated. If you have used a probability sampling technique (see guideline on 'Sampling') it is useful to calculate the sampling error for your results. Sampling error is reported as a 'confidence interval' - the range within which the 'true' value is expected to lie (for example +/- 3%). Confidence intervals are based on the size of the whole population, the size of your sample and the confidence level (usually 95%), i.e. the number of times out of 100 the 'true' value would fall within the stated range. For example, if 75% of your sample gave

a particular response and the confidence interval is +/- 3% then you can be 95% confident that the result for the whole population would lie between 72% and 78%.

More information can be found under 'confidence intervals' and 'categorical data' at [www.stats.gla.ac.uk/steps/glossary](http://www.stats.gla.ac.uk/steps/glossary), and a simple 'calculator' for working out the confidence intervals for your results is available at <http://www.surveysystem.com/sscalc.htm>.

### Looking at differences

Cross-tabulations can be used to examine any differences in your results between groups of interest, e.g. males and females. However, you should be careful not to make too much of small differences in results as these are likely to be due to sampling error.

A range of statistical tests (for example Chi-square) are available to help you when looking at the significance of any differences between results. Most software packages will perform these tests for you, but it is important that the results are understood and interpreted correctly.

Further information on statistical tests, including a 'statistical advisor' and an extensive statistics glossary can be found at <http://www.statsoftinc.com/textbook/stathome.html>.

### Structuring your report

Your final report needs to be clear, well organised and logical. The content may vary according to particular projects, but as a guide it is useful to include the following sections in a quantitative research report:

- Executive Summary
- Introduction/background
- Methods
- Findings
- Conclusions and recommendations
- Appendices (including full tables of results for reference and a copy of the questionnaire)

### What next?

If you would like to know more about this topic or any other aspect of research or consultation, please contact us:

<p><b>The Council's Corporate Research and Consultation Team (CRAC)</b></p>		<p>01484 221748 (internal 860-1748)  <a href="mailto:consultation@kirklees.gov.uk">consultation@kirklees.gov.uk</a></p>
-----------------------------------------------------------------------------	-------------------------------------------------------------------------------------	-----------------------------------------------------------------------------------------------------------------------------