

# Statistical Report Writing

*School of Mathematics , The University of Manchester.*

# Contents

<b>1</b>	<b>Summary</b>	<b>2</b>
<b>2</b>	<b>Introduction</b>	<b>2</b>
<b>3</b>	<b>Structure</b>	<b>2</b>
3.1	Title . . . . .	3
3.2	Summary/Abstract . . . . .	3
3.3	Introduction . . . . .	3
3.4	Methods . . . . .	3
3.5	Results, conclusions and recommendations . . . . .	4
3.6	General discussion . . . . .	4
3.7	References . . . . .	5
3.8	Appendix . . . . .	5
<b>4</b>	<b>Content</b>	<b>5</b>
<b>5</b>	<b>Presentation</b>	<b>5</b>
5.1	Figures, Plots, diagrams, <i>etc.</i> . . . . .	5
5.2	Tables . . . . .	6
5.3	Computer output . . . . .	6
<b>6</b>	<b>Use of English</b>	<b>6</b>
<b>7</b>	<b>Style</b>	<b>7</b>
<b>8</b>	<b>Detailed Preparation</b>	<b>8</b>
<b>9</b>	<b>Afterwards</b>	<b>8</b>
<b>10</b>	<b>References</b>	<b>9</b>
<b>11</b>	<b>Appendix: Statistics project reports - some interesting quotes</b>	<b>10</b>

# 1 Summary

As the work environment grows in complexity, the importance of skillful communication becomes essential. In addition to the need to develop adequate statistical skills you will find it necessary to effectively communicate to others the results of statistical studies. It is of little use performing complex and sophisticated statistical analyses if the conclusions of such analyses are not *meaningfully* transmitted to the rest of the community.

This document gives general guidance about the writing of reports, with particular reference to statistical reports following a statistical analysis. While this is immediately relevant for some of the courses you are taking, it may also be of help to you in your future job(s), some of which will hopefully use statistics.

# 2 Introduction

An essential part of a statistician's work is concerned with presenting his results to "customers" who are not usually too conversant with statistics. These notes are aimed at helping you to present your statistical analyses more clearly.

We shall consider the following aspects of report writing:

1. Structure.
2. Content.
3. Presentation.
4. Use of English
5. Style.
6. Detailed preparation.

Your main aim in report writing is to get your message across. Anything which diminishes your report's clarity or tends to irritate the reader will detract from this goal. Higham (1993) is an excellent book which deals with the more general challenge of mathematical writing. It is available in the John Rylands University Library.

# 3 Structure

There are various possibilities; the following one has many advantages. The report is made up of several components, listed below in the order in which they should appear. Some reports may well be better split up into sections, while shorter reports will not require as full a structure as suggested below. At all

costs, avoid the narrative approach, which says “First I did this, then I did that, etc.” Such approaches make the report difficult to follow.

### **3.1 Title**

Every report needs a title, preferably on a separate title page, which should also include the author’s address and department or other affiliation, and the date. The title should be short and informative. Titles such as “Project 1” or “Coursework 1” are non-informative and as such totally unacceptable; so is a title such as “The statistical analysis of two data sets”.

### **3.2 Summary/Abstract**

A brief summary – the shorter the better, and certainly no more than half a page – should follow, to tell the reader what the report is about, in general terms, and why it may be worth reading (!). A summary/abstract should contain a skeleton outline of

- the problem,
- what you have done,
- your conclusions

and should not, if at all possible, use symbols or numbers, and should contain technical terms only if they are unavoidable and widely understood.

### **3.3 Introduction**

This should describe the general context and background, including a general description of what data are available, the manner in which and the purpose for which they were collected. It should also provide the aims of the investigation, together with some indication of the methods used. It may well be an expanded version of the summary, and in general should not exceed two pages in length.

### **3.4 Methods**

These should now be described in a fair amount of detail, including any theory that is necessary. How much detail is always difficult to decide. Aim to write as though for someone of comparable standing, e.g. a student in the same year as yourself, who does not know anything in detail about the subject of the report. The reader should be able to repeat the study on the basis of your report. Don’t be afraid of putting very detailed and long descriptions into one or more appendices.

This component may well have to be split into sections, e.g.

1. a preliminary analysis, using graphical and simple descriptive methods,
2. a full-scale analysis.

The second section may further have to be split into two subsections e.g.

- (a) a description, explanation or development of the methodology to be used,
- (b) the actual analysis.

Subsection (a) will probably need to be quite technical, certainly more technical than the rest of the report. It should describe in detail the precise mathematical/statistical tools and techniques to be used and reveal the manner in which they lead to the desired results. Do not shy away from using mathematical notation (get well familiar with the capabilities of your word-processor to produce mathematical expressions) but be careful to define every bit of the notation you use. A mathematical expression is totally meaningless if any of the symbols used in it is undefined. Similarly pages of mathematical development and explanations are rendered useless and total waste of space if any of the symbols involved in these explanations are not defined. I

In the subsection containing the actual analysis it is important that you describe your analysis sufficiently clearly and carefully to enable it to be reproduced by the reader.

### **3.5 Results, conclusions and recommendations**

Give the main results and conclusions. Subsidiary results and deductions can be left in this section if they do not detract from the flow of the presentation of the main results and conclusion; otherwise they may be gathered into a separate component, be combined with the description of methods, or be put into an appendix. Your conclusions should be expressed in a way that can be understood by a non-statistician and should make sense even if the reader of your report had omitted to read the middle section on detailed statistical analysis. Try to report your conclusion in the context of the experiment from which the data came.

### **3.6 General discussion**

It may be appropriate to give some account of previous investigation of the same or related problems, or to relate the present conclusion to others in a connected area. It is worth discussing how far the original aim was successfully achieved, if it was not, why not, and how you might have done things differently. Reservations about the data also belong here. (In real life, data sets may well contain errors or have been obtained from a badly designed experiment). This general discussion can in fact appear not as a section in itself but can be subsumed in the Introduction.

### 3.7 References

These should always be included if you refer to books or papers from journals. The simplest, and widely-used style, is to refer to authors in the text in the format “Surname (Year of Publication)”, and to give the full references in the References section as shown at the end of this article i.e. for a journal article reference give Author’s surname, author’s initials, title of paper including subtitle if there is one, journal name, year of publication, volume, first and last page numbers of article, in that order; for a book reference give, in the following order, the author’s surname, the author’s initials, (year of publication) Title of book (in italics), publishing company, city of publication.

### 3.8 Appendix

Whatever does not fit naturally into the main body of the report and cannot reasonably be omitted should go into an appendix. However, you don’t want the report to suffer from appendicitis! A case of the appendix or appendices tail wagging the main text dog makes the report hard to read.

## 4 Content

Here are some guidelines.

1. The report must be self-contained: everything should either be stated explicitly, be justified by a reference to some other source, or be common knowledge. In particular, the source of any data should be clearly indicated.
2. It is not usually sensible to describe everything that you did in detail. Indeed, if you modify your approach, there is no need to describe the earlier approach more than briefly.

## 5 Presentation

Reports should be attractive and easy to find one’s way through. Hence, they should be neat, clean and tidy, and must be *legible*. It is advisable that you try to word-process the report– not a difficult thing to do on the School of Mathematics or other University PCs available to you using Word or Latex.

### 5.1 Figures, Plots, diagrams, *etc.*

Tables, figures, graphs, and diagrams should be titled and numbered (and preferably given a caption). Sections (which should have headings) and pages should

also be numbered. Headings should stand out (by being underlined or on a separate line, for example) and should **never** appear as the last line of a page. Think of a report as needing signposts, provided by the section numbers. Similarly, graphs need to be neatly drawn, *carefully labelled* and titled: scales should be chosen sensibly and the units indicated. Freehand graphs, carefully drawn, may be appropriate; a sketch graph with various amendments, and axes carelessly drawn, will never be. As far as possible, tables and diagrams should be placed in the text near to the passage referring to them.

## 5.2 Tables

Tables need titles, numbering and captions, should be boxed in, and figures in columns should be aligned correctly, usually so that the decimal points fall in the same column. Rows and columns usually need labels or headings, and units must be stated in, or perhaps close to, the table.

## 5.3 Computer output

Computer output needs some thought. For a student project, it may not be necessary to copy the results from the output so that they have the same appearance as the rest of the report; it should suffice if the output is **edited appropriately** first and then pasted onto the body of the report or otherwise attach it.

A possibly helpful reference is Chapman & Mahon (1986).

## 6 Use of English

It is very important that you think carefully about your use of English. The object of writing, as of speaking, is to communicate. Writing, in particular, requires very careful use of language, especially when abstract and difficult concepts (such as those involved in statistical modelling and analysis) are to be communicated. It is easy to read and understand English which is well written and follows the rules of good usage. However, as the quality of writing deteriorates, it becomes progressively more difficult for the reader to work out its intended meaning. Finally it becomes impossible. While many project reports are written to a very high standard, many others contain much that is unintelligible. An obvious, and easily corrected, problem occurs when the author fails to use properly constructed sentences. (A good first test check here is to identify the verb in each sentence!) A much more insidious (and very frequent) problem occurs when the author unconsciously assumes that the reader is somehow following his/her quite unwritten train of thought. Thus the reader is presumed to know exactly what the author is talking about—for example, what variables are currently being considered, what is being held fixed, and what is being allowed to vary - when in fact none of this

has ever been stated. Attached to this document are some illustrative quotations, with commentary where appropriate, which have been taken from recent project reports. In most cases you will find that the sentences look just fine - until you try to decode the meaning. How should you attempt to improve the quality of your writing? It is important to understand that, like learning to drive, this is something you have to work at. No doubt an ability to write well comes more easily to some than to others, but everyone needs to put some effort into acquiring this skill, and anyone can improve if they do. Of course there are plenty of books on how to write well, but perhaps the most useful thing you can do is to read widely (books, articles in *quality* newspapers, etc ) and learn to observe how others write.

## 7 Style

1. Be brief and to the point and use shorter rather than longer words. Thus, ‘schools and colleges’ is to be preferred to ‘educational establishments’ if the two phrases are used with the same meaning. If you find that you are writing about something not relevant to your heading then either the material belongs elsewhere, possibly even to a new separate section, or you are rambling and the material should be omitted.
2. In general, try to avoid the use of the personal pronoun “I”, as it can get very irritating to the reader. Impersonal verbs and the passive form are usually to be preferred, e.g. “A regression analysis was carried out” instead of “I carried out a regression analysis”.
3. *Use of tenses*: A scientific report should be written in the present tense e.g. ”The statistical analysis shows that.....”, not ”The statistical analysis showed that.....”. You use the past tense only when describing events that occurred in the past e.g. ”The data were collected in a study conducted by Royal Statistical Society.....”.
4. Try to get your spelling correct! The main problem is knowing which words’ spelling you are shaky on. If in doubt, look up the word in a good dictionary (Oxford, Cassells, for example) or use a spell-checker.
5. Avoid using words inappropriately. For example, the similar-sounding words “tendency” and “trend” are not the same and cannot be used interchangeably. Do not use the word “significant” other than in a statistical sense.
6. If it is not too late, try to write grammatically: whole sentences, each with its main verb; no telegraphic style. Thus, the following is not acceptable.

“A standard analysis is possible. By regression.”

Mathematical/Statistical writing follows the same grammar and syntax as ordinary English.

7. A common error: the word “data” is treated as if it were a singular noun instead of as a plural noun, which it is! (Data=plural of the singular noun *datum*.) The correct use of the word is “Data *are* ...”.
8. A good style, which is a pleasure to read, is not something which comes naturally to mathematicians on the whole. Nevertheless, it is worth trying to aim for it. The only way to develop a good style is to read good books, practice writing and read critically what you have written.

Gowers (1973) and Fowler (1983) (see references at the end) are useful books on grammar, punctuation and style.

## 8 Detailed Preparation

Two main points:-

1. Plan the report *before* you start writing it.
2. Be prepared to revise it, possibly extensively, after you have written the first version.

If the report covers work over an extended period of time, say several weeks, it is usually sensible to write the first draft piecemeal as the results come in from the analysis. This will help to avoid a much larger task at the end of the exercise if you leave everything to the last moment. You may also find that the process of writing up may itself suggest new ideas and so may influence your plans for the later parts of the analysis.

One final, perhaps obvious, point: you don't need to write the sections of the report in the order in which they will appear in the completed report: it may often be easier to write the Introduction and Conclusions *after* drafting the other sections of the report.

## 9 Afterwards

After you have finished your report READ IT AGAIN. Check for

- Incomplete/incorrect punctuation.
- Incorrect line or paragraph breaks.

- Parentheses not matching.
- Wrong font.
- Missing words or phrases.
- Incorrect formatting of equations.
- Wrong numbers in tables.

Even if there are no technical errors you will often find better ways to express what you are trying to say.

Finally put your document aside for a day or two and then read it again. If at this stage parts of it do not make sense to you, they will not make sense to other readers of your report either. You will need to rethink the way you expressed your thoughts and the way you presented your material.

## 10 References

- Chapman, M. & Mahon, B. (1986), *Plain Figures*, HMSO, London.
- Fowler, H. W. (1983), *Dictionary of Modern English Usage* (2nd edition revised by Sir Ernest Gowers), Oxford University Press, Oxford and New York.
- Gowers, Sir Ernest (1973), *The Complete Plain Words* (2nd edition revised by Sir Bruce Fraser), Penguin Books, Harmondsworth, Middlesex.
- Higham, Nicholas J. (1993), *Handbook of Writing for the Mathematical Sciences*, Society for Industrial and Applied Mathematics, Philadelphia.

## 11 Appendix: Statistics project reports - some interesting quotes

"Note the sizable difference in minimum and maximum ages."

*This actually meant (I think): "Note the sizable difference between the minimum age of those in group A and the minimum age of those in group B" (with a similar observation for the maximum ages).*

"From this data, there appears to be a general increase in haemoglobin levels"

*This sentence, which appeared in isolation, suggests that haemoglobin levels are increasing. With what? The author didn't say. With the passage of time (not a consideration in the project), or as some other variable increases? If so, which? The reader can only guess.*

"The grouping of the data"

*This probably meant the "distribution of the data".*

"...however, if one compares each of the groups to the sample it is the group with PSR who bear the most similarity"

"Mcv against age however does show any distinct influence on PSR."

"The reasoning behind my method of analysis is that figure 3 did not show any influence on PSR when two explanatory variables were taken simultaneously."

"The interquartile range and all the data cover a larger set of volumes."

"It is clear from the plot that haemoglobin is proportional to age in absence and presence of proliferative Sickle Retinopathy after the trim."

"A t-test concluded that one could be 95% confident that, on average, the mean of age1 is higher than the mean of age0."

"However, it should be noted that this level lies directly on the line of the haemoglobin concentrations of those men who do not have PSR."

"As you move in a direction of 45 degrees, individuals tend to show PSR present."

"Working along the horizontal axis of figure 16..."

"A 95% confidence interval gives (-18.25, -8.85) with a p-value of less than 0.05 so there are clearly differences between the two data sets."

*Confidence interval for what? Test for what?*

"Although the third plot indicates that Hb increases with MCV, the difference between those with PSR and those who do not have the disorder is likely to be due to the difference in age, i.e., nothing can be taken from the differing values for the two groups."

"Noticeably, there are no males below 10 who have PSR, while there are almost as many males over 40 with PSR as there are males under 20, with most of the values falling between 20 and 40."

"This model is unsuitable as it assumes that month has a constant affect on temperature for each pressure level"

"...by the red ink which indicates the differences not to be equal"

”...as the difference in any two pressure levels is always proportional to the difference between the same two pressure levels for a different month to the same extent as any other two pressure levels for the same two months.”

This could have been expressed by saying that the value of a certain function was independent of one of its arguments - a little mathematical notation would have made everything clear!

”The assumptions of this model are that pressure is independent of month and that differences in temperature at the same pressure levels occur over different months.”

”...and a similar trend of similarity between the two was noted.”

”By using a weight  $w(i)$  it could be possible to obtain better fits of residuals for some of the pressure levels which were not as effective.”

*What is a "fit of residuals" and what is an ineffective pressure level?*