

Communicating Statistical Results

Jennifer A. Hoeting and Geof H. Givens*
Colorado State University

Spring, 1997

Writing understandable, interpretable prose describing statistical results is a challenging endeavor. Knowledge of complicated statistical techniques are of little help unless you can explain the results of your analyses. In this class you will be asked to write several papers summarizing data analyses you have conducted. This document describes some key features of good papers.

1 Your Responsibility to the Reader

The most important single thing to remember is that you are telling a story. From your point of view, the story may begin when we give you data to analyze. However the story really began much earlier, when the scientist first identified some questions which s/he found fascinating.

It is your job to help the reader share this fascination by telling the story from the beginning: what are the questions of interest and why are they interesting? Without this motivating information, your methods and results lack context.

The story you tell should be a cohesive narrative of ideas that leads to your conclusions. It should not be a diary describing everything you did during the analysis. Do not describe failed approaches and dead ends, your tribulations with the latest statistical software, or any other distracting material. Do not include computer code or output.

2 Format for Papers that Convey Statistical Results

There are many reasonable scientific paper formats. We prefer to read papers which are written in traditional order (see below) because this order facilitates story-telling and emphasizes the scientific context of the statistical analysis. Ehrenberg (1982) recommends nearly a reverse order, which can work quite well, especially for more technical papers and journal articles. We believe the reverse ordering is more difficult to do well, however.

*This document may not be reproduced in part or in full without the explicit consent of the authors.

A paper written in traditional order might have main sections as follows:

Abstract: This is a brief summary of (i) the questions of interest, (ii) the study design and methods of analysis, and (iii) the results.

Introduction: The paper should begin with a clear statement of the scientific questions to be addressed and why they are interesting. Describe the objectives of the study, related research, and other background information. You may prefer to discuss the data collection methods and to describe the features of the data (figures, tables, summary statistics, etc.) in this section or in a subsequent section.

Methods: Describe the study design and analysis methodology. Describe the data if you have not already done so. Emphasize the relationships between the methods/design and the questions of interest. Why did you choose a particular approach? How do your methods address the questions of interest?

Results: Describe the main results of the statistical analysis. Interpret these results in light of the study objectives and questions of interest. **Provide some measure of uncertainty with each result included in your paper.** This section should be organized in a logical manner, which is typically not the order in which you performed your analyses. (Some authors prefer to combine the methods and results sections in a single section, *Analysis and Results*.)

Conclusions: This section usually includes a summary of the main findings of the paper, detailed discussion of how these findings relate to the questions of interest, comparisons of the results with those of other studies, and suggestions for future research or possible improvements in your approach. Avoid repeating material from the results section.

References: In order to provide sufficient context, it will probably be necessary to cite related research in relevant scientific fields. You do not need to provide references for standard statistical techniques in your papers for this class.

Technical appendices: For complicated statistical analyses, a technical appendix is appropriate. The appendix might include details of nonstandard statistical methodology or other material which would have interrupted the flow of the body of the paper. For this class, do not list the data, and never include a printout of computer code used for analyses.

3 Keep it Simple

The best analysis is the simplest one that completely answers the questions of interest. The fanciest analysis is not the best analysis. Needless introduction of sophisticated techniques dilutes your report, increases abstraction, and decreases comprehension. Carry out a deft, parsimonious, and elegant analysis.

Your class studies sophisticated statistical methods, and you are given free rein to apply the techniques you learn to data. Naturally, you may feel a desire to demonstrate your mastery of the most complex statistical methods in your report. Resist this urge.

4 Describing Data

Many authors jump straight into the statistical analysis without mentioning what the data look like. Early in the paper, you should describe the data collection methods, and provide graphical, tabular, or other summary information about the data.

Data description is not a rote chore. Do not simply report univariate marginal means, standard deviations, and so forth for each variable. Choose your descriptions in light of the questions of interest. **Early descriptive figures and tables are the reader's guiding map for the entire analysis.** Choose scatterplots, paired boxplots, or cross-tabulations which illuminate points which are later confirmed or refined with your analysis. Create these displays to enable the reader to anticipate the rest of your paper. Your displays should allow the reader to gauge the reasonableness and plausibility of your methods and results. If you can show the results of your statistical analysis in a picture, the statistical results that you present will be much more convincing.

For example, Figures 1 and 2 show two descriptive plots for a dataset which was collected to investigate the relationship between gas mileage and automobile characteristics. Figure 1 is not useful. It is a histogram of the numbers small, medium, and large cars. This information tells us nothing about the question of interest. In contrast, Figure 2 shows that small cars tend to be substantially more fuel efficient than other cars.

Figures 3 and 4 show two more descriptive figures from the same dataset. Figure 3 shows that there may be a relationship between mileage and car weight. However, Figure 4 also shows this relationship, along with other features whose presence was used in the methods section to motivate why different models for manual and automatic cars were reasonable.

5 Figures

Each figure should be chosen to convey a message which is relevant and important to your story. Create the figure to convey the message and eliminate all extraneous features. Discuss and describe your figure in the body of the paper. Otherwise, omit it.

Making clear and informative figures is as difficult as it is necessary. Statisticians, computer scientists, cognitive psychologists, and other researchers have a lot of useful advice. Books by Cleveland (1985) and Tufte (1983) are good references.

Here are a few basic points to keep in mind:

1. Label all axes. Provide meaningful axis limits and tick marks.
2. Avoid deceptive or distracting aspect ratios (i.e., the ratio of figure width to height).
3. Arrange plots on the page in a manner which facilitates comparisons between plots. Choose the most sensible and informative horizontal or vertical alignment, axis scale and position, and plot types.
4. If you ask the reader to visually compare several plots, it is usually best to use the same axis scales in those plots.

Figure 1

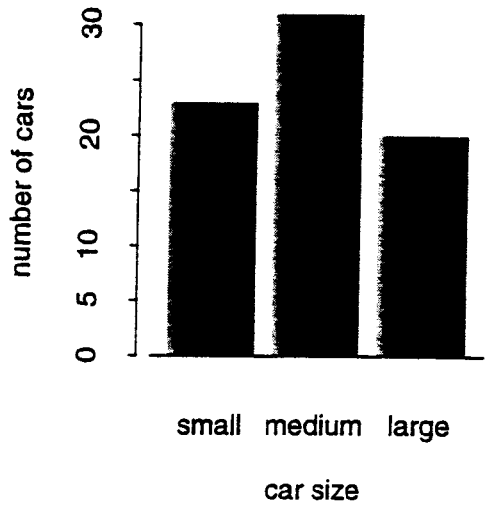


Figure 2

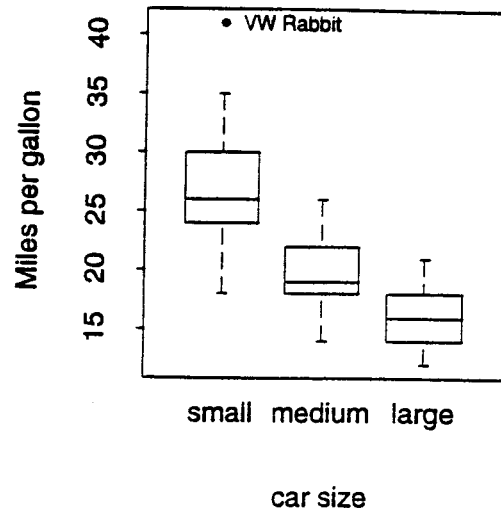


Figure 3

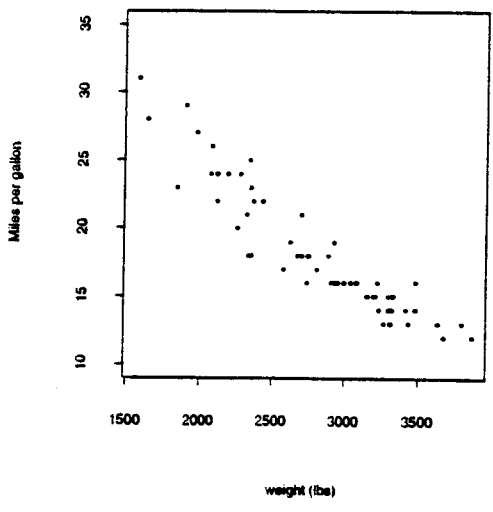


Figure 4

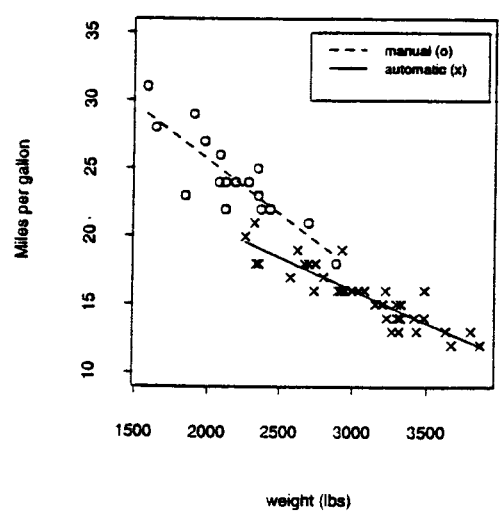


Table 1: A bad table. Table 2 shows an improvement.

Method	Model	Log score	Predictive Coverage %
Adjusted R ²	1,2,3,4,5,8,11,12,13,15	43.384321	67.521199
C _p	1,2,3,4,11,13	42.39846832	63.98365
MC ³	model averaging	29.3498233	80.365151
Occam's Window	model averaging	31.4536	80.65231651
Stepwise (15%)	3,4,8,9,13,15	39.51319	63.546632

Table 2: An improved version of Table 1.

Method	Predictors in model	Log score	Predictive Coverage %
MC ³	model averaging	29	80
Occam's Window	model averaging	31	81
Stepwise (15%)	3 4 8 9 13 15	40	64
C _p	1 2 3 4 11 13	42	64
Adjusted R ²	1 2 3 4 5 8 11 12 13 15	43	68

5. Keep it simple. Too many colors, lines, plotting characters, dimensions, or labels are distracting.
6. Make it legible. Consider the scale and clarity of the plot and its components.
7. Provide a clear and complete caption.
8. Remember that figure creation is an iterative process.

6 Tables

As much care must be taken with tables as with figures. Ehrenberg (1981) believes that writers, not readers, are usually to blame when the message of a paper is lost in a sea of numbers.

Some guidelines for good tables are below.

1. Round severely. Include only those significant digits that you need to convey the message. This makes tables easier to read and understand, and avoids implying more accuracy than intended. Do *not* use all of the digits that SAS prints out.
2. Order entries in your tables to convey information. In Table 1, the methods are in alphabetical order, which is irrelevant. In Table 2, the methods are sorted according to ranking of log score so it is easy to see which method has the smallest log score.
3. Give clear and complete descriptive captions.

Table 3: This table is easier to read than Table 4 because comparisons are made within a column.

Method	Estimate	SD
MLE	3.7	1.1
MOM	3.5	0.8
LS	4.1	1.9

Table 4: This table is harder to read than Table 3.

	MLE	MOM	LS
Estimate	3.7	3.5	4.1
SD	1.1	0.8	1.9

4. Organize tables so that comparisons can be made columnwise. The human eye is much better at comparing a column of numbers than a row of numbers. For example, Table 3 is easier to read than Table 4.

7 Miscellanea

1. If you include a figure or table in a paper, you must discuss it in the body of the paper. If you do not have anything to say about a table or a figure, then it serves no purpose in the paper.
2. Scientific writing requires special skills in addition to those required for drafting clear and correct prose. Gopen and Swan (1990) and Higham (1993) address these skills. Improve your style, syntax, and grammar with reference to such books as *The Elements of Style* (Strunk and White, 1979). Keep a dictionary and a thesaurus in your office. Use computer spelling check programs.
3. Allow time for careful revisions. We have revised this paper nine times, and it is far from perfect.
4. The words *data*, *strata*, and *criteria* are plural.
5. Avoid the word *prove* in your conclusions. Statistical results almost never prove anything, they support conclusions.
6. Avoid using the word *significant* in any context except the formal statistical one. Instead, use *important*, *substantial*, or another choice depending on your meaning.

7. Include proper punctuation when a sentence includes a formula which is set off from the text. Frequently, a period or comma should follow the equation.
8. Be concise. You will be amazed how much we delete from some students' papers.
9. Avoid footnotes whenever possible.
10. Adopt a serious, scientific tone. Use simple, telegraphic prose, and minimize jargon.

8 Grading

In this class, papers will be evaluated according to the following criteria:

Statistical appropriateness. Appropriateness of the methods and models chosen to analyze the data. Technical execution of the analysis. Diagnostics. Quality of graphs and tables.

Scientific appropriateness. Accuracy of the statement of the questions of interest. Thoughtfulness and simplicity of your statistical analyses. Interpretation of the results of the analysis. Does the analysis answer the scientific questions of interest?

Writing. Organization, logic, clarity, and accuracy of report.

These criteria will be weighted equally.

9 References

- Cleveland, W.S. (1985) *The Elements of Graphing Data*, second edition, Wadsworth, Pacific Grove, California.
- Day, R.A. (1988), *How to Write and Publish a Scientific Paper*, third edition, Cambridge University Press: Cambridge.
- Ehrenberg, A. (1981), "The Problem of Numeracy", *The American Statistician*, **35**, 67-71.
- Ehrenberg, A. (1982), "Writing Technical Papers or Reports", *The American Statistician*, **36**, 326-329.
- Gopen, G. and Swan, J. (1990), "The Science of Scientific Writing", *American Scientist*, **78**, 550-558.
- Higham, N.J. (1993), *Handbook of Writing for the Mathematical Sciences*, Society for Industrial and Applied Mathematics, Philadelphia.
- Strunk Jr., W. and White, E.B. (1979) *The Elements of Style*, third edition, Macmillan, New York.
- Tufte, E.R. (1983), *The Visual Display of Quantitative Information*, Graphics Press, Cheshire, Connecticut.