

Communicating Statistical Results

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

August 26, 2002

Being a statistician is no excuse for being a poor writer. Scientists rely on us to provide understandable, interpretable conclusions, yet, sadly, their expectations of us are astonishingly low. Excellent communication skills will win you enormous amounts of respect and gratitude, not to mention business opportunities.

We recently dined with two statisticians who consult (and hire consultants) for major telecommunications and pharmaceutical companies. We asked our friends what was the most important skill a statistician should know when entering the job market. Without hesitation, they both responded that outstanding communication skills were paramount. They gave little priority to knowledge of overdispersed log-linear models, quadratic discriminant analysis, or other complicated statistical techniques.

In this class you will be asked to write several statistical reports summarizing data analyses you have conducted. This paper describes some key features of good reports. The *Handbook of Writing for the Mathematical Sciences* (Higham, 1993) is an outstanding reference for further advice.

1 Your Responsibility to the Reader

You are telling a story. Frequently, a statistician's involvement in the story begins when s/he receives a dataset from a scientist and is asked to analyze it. From your point of view, the story may begin when you received the data. However the story really began much earlier, when the scientist first identified some questions that 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 Report Format

There are many reasonable report formats. We prefer to read reports that are written in chronological order, 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

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

well, especially for more technical papers and journal articles. We believe the reverse ordering is more difficult to do well, however.

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

1. *Abstract.* This is a brief summary of (i) the questions of interest, (ii) the analysis methods, and (iii) the results.
2. *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 (graphs, tables, summary statistics, etc.) in this section or in a subsequent section.
3. *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?
4. *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 writers prefer to combine the methods and results sections in a single section, *Analysis and Results*.)
5. *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.
6. *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 reports for this class, but references for advanced techniques should be given.
7. *Technical appendices.* For complicated statistical analyses, a technical appendix is appropriate. The appendix might include details of nonstandard statistical methodology or other material that 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 Simplicity

The fanciest analysis is not the best analysis. The best analysis is the simplest one that completely answers the questions of interest. 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 reports jump straight into the analysis without mentioning what the data look like. Early in the report, 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 graphs and tables are the reader's guiding map for the entire analysis.** Choose scatterplots, paired boxplots, or cross-tabulations that illuminate points that are later confirmed or refined with your analysis. Create these displays to enable the reader to anticipate the rest of your report. Show the reader how your simple plots give rough information about the questions of interest. Your displays should allow the reader to gauge the reasonableness and plausibility of your methods and results.

For example, Figures 1 and 2 show two descriptive plots for a dataset that was collected to investigate the relationship between city gas mileage and automobile characteristics for a set of 57 models of cars. Figure 1 is pointless. It is a histogram of the numbers models of 3-, 4-, and 6 and 8 cylinder cars. This information tells us nothing about the question of interest. In contrast, Figure 2 shows that cars with fewer cylinders are more fuel efficient.

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 Graphs

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

Making clear and informative graphs is as difficult as it is necessary. Books by Cleveland (1985) and Tufte (1983) are good references.

In the meantime, remember a few basic points.

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 that facilitates comparisons between plots. Choose the most sensible and informative horizontal or vertical alignment, axis scale and position, and plot types.
4. Keep it simple. Too many colors, lines, plotting characters, dimensions, or labels are distracting.
5. Make it legible. Consider the scale and clarity of the plot and its components.
6. Provide a clear and complete caption.

Figure 1: # of cylinders

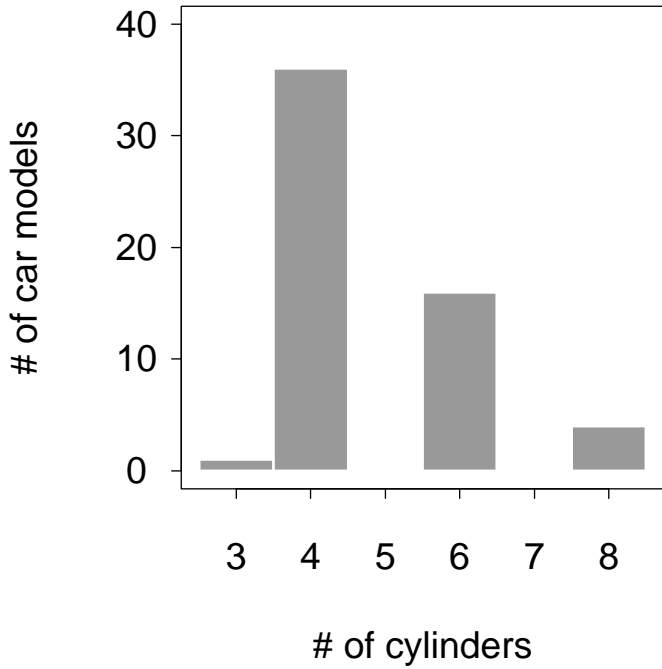


Figure 2: # of cylinders by city MPG

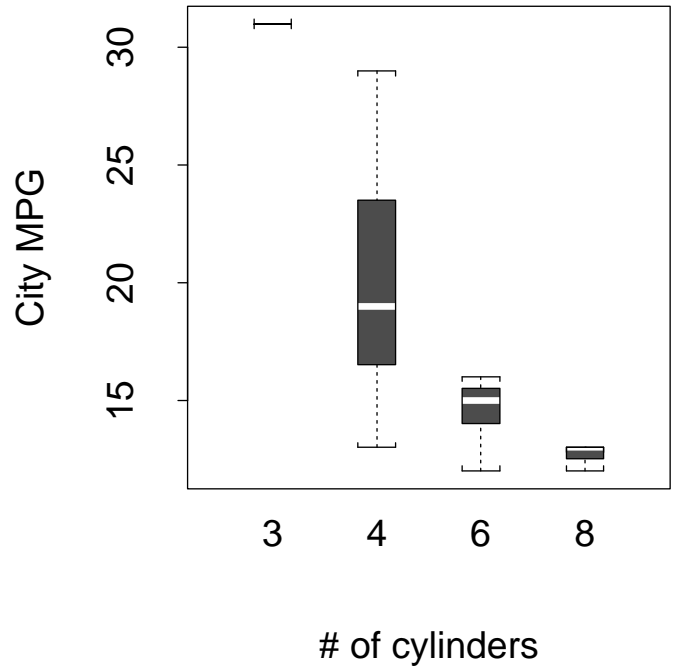


Figure 3: city MPG versus car weight

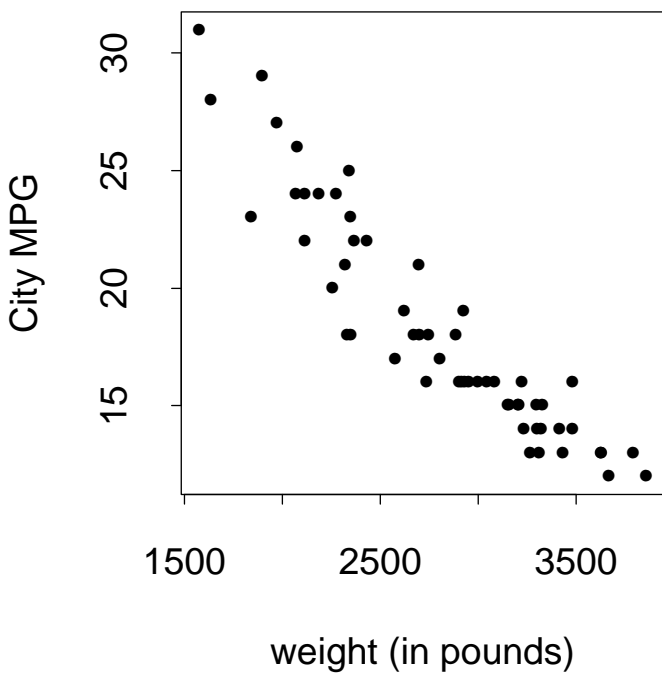


Figure 4: city MPG versus car weight

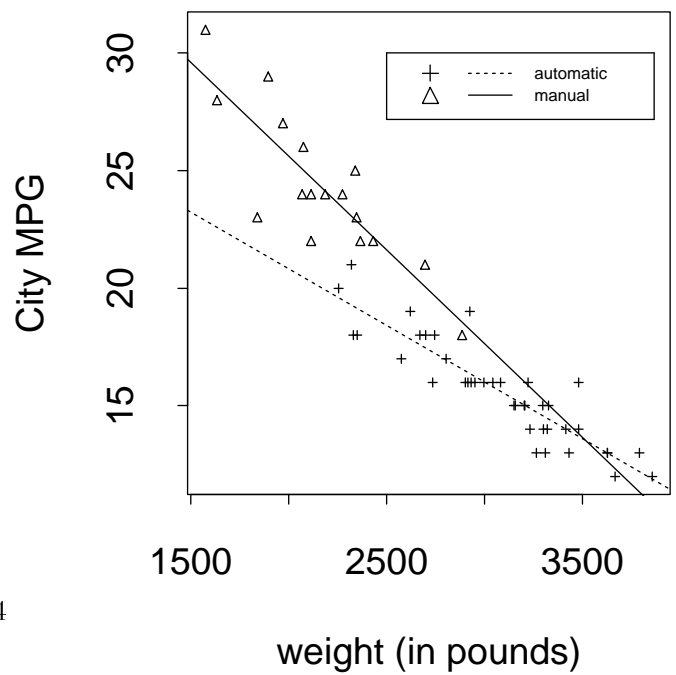


Table 1: A bad table. Table 2 shows an improvement.

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

6 Tables

As much care must be taken with tables as with graphs. 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.
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.
4. Organize 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 harder to read than Table 4.

7 Miscellanea

1. Improve your style, syntax, and grammar with reference to such books as *The Elements of Style* (Strunk and White, 1979). Keep a thesaurus and dictionary in your office. Use computer spelling check programs. Day's (1988) amusing "commandments" include the following. A preposition is a poor word to end a sentence with. Verbs has to agree with their subject. Remember to never split an infinitive. When dangling, don't use participles. Join clauses

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

	MLE	MOM	LS
Estimate	3.7	3.5	4.1
SD	1.1	0.8	1.9

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

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

good, like a conjunction should. Don't write a run-on sentence it is difficult when you got to punctuate it so it makes sense when the reader reads what you wrote. About sentence fragments.

2. Scientific writing requires special skills in addition to those required for drafting clear and correct prose. Higham (1993) and Gopen and Swan (1990) address these skills.
3. Allow time for careful revisions. We have revised this paper nine times, and it is far from perfect.
4. 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.
5. The words *data*, *strata*, and *criteria* are plural.
6. Avoid the word *prove* in your conclusions. Statistical results rarely prove anything, they support conclusions.
7. Avoid using the word *significant* in any context except the formal statistical one. Instead, use *important*, *substantial*, or another choice depending on your meaning.
8. Include proper punctuation when a sentence includes a formula that is set off from the text. Frequently, a period or comma should follow the equation.
9. Be concise. You will be amazed how much we delete from some students' reports.
10. Avoid footnotes whenever possible.
11. Adopt a serious, scientific tone. Use simple, telegraphic prose, and minimize jargon.

8 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.

9 Appendix: Splus code for the figures

```
postscript(file="cars.ps",print.it=F)
par(mfrow=c(2,2))
hist(cars$cylinders,xlab="# of cylinders",ylim=c(0,40),
ylab="# of car models",axes=F)
axis(1,at=seq(2.5,7.5,by=1),labels=c(3:8))
axis(2,at=seq(0,40,by=10))
box(n=1)
title("Figure 1: # of cylinders")

boxplot(split(cars$cityMPG,cars$cylinders),xlab="# of cylinders",
ylab="City MPG")
title("Figure 2: # of cylinders by city MPG")

plot(cars$weight,cars$cityMPG,ylab="City MPG",xlab="weight (in pounds)")
title("Figure 3: city MPG versus car weight")

plot(cars$weight,cars$cityMPG,ylab="City MPG",xlab="weight (in pounds)",type="n")
title("Figure 4: city MPG versus car weight")
a<-as.logical(cars$transmission)
points(cars$weight[a],cars$cityMPG[a],pch=3)
abline(lsfite(cars$weight[a],cars$cityMPG[a]),lty=2)
points(cars$weight[!a],cars$cityMPG[!a],pch=2)
abline(lsfite(cars$weight[!a],cars$cityMPG[!a]),lty=1)
legend(3000,30,legend=c("automatic","manual"),lty=c(2:1),marks=c(3,2),cex=.4)
title("Figure 4: city MPG versus car weight")
dev.off()
```