

## STUMBLING BLOCKS ON THE ROAD TOWARDS STATISTICAL LITERACY

Herman Callaert, Limburgs Universitair Centrum, Belgium

*If statistical literacy for all (in whatever sense) is a goal worthwhile to be pursued, the educational task of the statistical community is enormous. A particularly difficult area is the wealth of statistical information apparently out of the control of professional statisticians. If new pedagogy and new content is needed in the statistics courses we teach (and which deserves a lot of fundamental research and international co-operation), what about the pedagogy and content of the statistics “on the street”? Is there any action statisticians and statistical societies should take here?*

### INTRODUCTION

Recent history shows a substantial increase in the concern for professional research on and guidance of the statistical education process. Among a variety of publications in this field, the paper by Moore (1997) together with the accompanying discussion papers, quite convincingly shows that a reform movement in statistical education is (might be) on its way, and that several colleagues in the statistical community are taking it seriously. The discussion paper by Hawkins (1997) focuses on statistical literacy for all, stating fundamental questions about the concept of statistical literacy itself and about a change of emphasis in our teaching objectives. Challenges for bringing about adequate changes in statistics education are enormous and resistance to it has many faces (see, for example, the discussion paper by Garfield (1997)). A major part of the effort for bringing statistical literacy to everyone concentrates on improving the education offered by the statistical community, and rightly so. However, people's (il)literacy might also be influenced by “daily life” exposure, and this “forced learning” may well overrule whatever is left from a single introductory course taken years ago. In this paper we illustrate different kinds of “everyday exposure” indicating possibly disastrous side-effects. We briefly sketch a couple of pitfalls in the area of visualization, software and standard textbooks. This is just the top of the iceberg, and every area can unfortunately be illustrated by numerous examples. The question is raised whether not only individuals but also statistical societies should play a visible role here.

### GRAPHS

Instances where data are displayed through graphs are abundant and nobody escapes from seeing them. From elementary to elaborate, graphs show up in newspapers and on TV, in popular magazines and professional journals, on web-pages and so on. Hence being surrounded in daily life by a multitude of eye-catching multicolored displays of data, learning, at least at the descriptive level, how to understand the message from the picture should be easy.

But what do people see when exposed to a graph ? This is a double question. The first one, and the most natural in the context of an ICOTS conference, is about perception and learning, about how information is presented and about what people make of it. Progress of research into these topics is amply reported on in other publications. The second question can be rephrased as an exclamation: unbelievable, what people are seeing when exposed to graphs ! I assume that, some time ago, when graphs had to be drawn by hand, some (not all) people considered it worthwhile to first think hard about what data to display and how. A nice example is the “Carte Figurative” de Charles Joseph Minard drawn in 1869 about Napoleon’s 1812 campaign in Russia (see Tufte 1983). Nowadays, anyone with whatever dataset and a PC can pick from a multitude of fancy displays. And it doesn’t take a week’s reading of your favorite newspaper or magazine to find out that one really does it. It is amazing how many recent graphical displays still comply with the rules given in Wainer’s 1984 paper. Since many people only remember the global impression they get from a quick glance at a graph (not everyone looks carefully at labels on the axis nor at accompanying text) a graph should be clear, instructive and honest. Creating this type of graphs for use in statistics courses is not simple, even for the professional statistician. So what with the numerous graphs all around and their influence on the statistical literacy of people ?

## SOFTWARE

Statistical packages are becoming ever more powerful and diversified. Many are aimed at specific research questions encountered in the design and analysis of experiments. Some are explicitly being developed with statistical education purposes in mind. One example in this field is MEDASS, and interesting ideas related to the development of this type of software can be found in Biehler (1997). Not all statistical software packages comply to the same high standards, let it be minimal user friendliness, acceptable information about procedures and results, or even mathematical correctness.

But a lot of those packages are good, and even excellent, for doing what they are made for. They also are improving steadily, under the closely watching eye of the statistical community, who uses them, tests them, helps in their development and writes critical reviews. A drawback is that most of this software is specialized, expensive and confined to being used by specialists, with at least a bit and often a lot of statistical background. For a variety of reasons, many people have found their own route of escape. Some use statistical software so infrequently that any learning curve pertaining to a new package is considered too steep, even without trying. Another problem is the wealth of existing packages for a variety of purposes and the resulting confusion on choosing the right software for a vaguely defined “statistical need”. Last but not least, price plays a role, and an important one, as well for schools as for individual students, teachers and researchers, who have much too small grants or personal money as their only resources. The result is that many people are clever enough to maximally get their money’s worth, and use what they already paid for anyway. Whatever formula and statistical analysis tool they can get out of their favorite word processor or spreadsheet will be used first, and often will satisfactorily answer the basic statistical questions they have. If this really is the way many people perform statistical analysis (a simple one, and the only one they ever need), it might be wise for the statistical community to take some interest in this type of popular packages too. Some developers might appreciate helpful hints on statistical statements in their displayed results and in their on-line manuals, since for many of their customers it might be a first contact with the world of statistics or it can be the only way to profit in some form from continuing education in this field.

#### STANDARD TEXTBOOKS

Apart from daily life encounter, quite a few people come across statistical concepts in somewhat more specialized settings. This certainly is the case for many students and researchers from academia, government and industry, who are working on a variety of subjects. Taking specialized courses and attending colloquia is one thing, but “find it out for yourself and meet the deadline” is not uncommon too. Nowadays there is a wealth of information relatively easily accessible, and virtual libraries are combining classical books and journals with fancy information-and-communication technology. Being overwhelmed with information is one thing, and not completely new. I still remember my first visit to a huge university library... that was overwhelming too. But

given that people go through the effort of searching for articles and books so that they can learn statistical concepts and methods needed in their particular situation, it still is unclear that they get very helpful information. Of course, variety in level and quality is enormous here, and fortunately many publications in statistics apply to very high standards, appropriate to the purpose they are intended to meet. This does not mean that the statistical community should not continue to critically appraise what is being published and support improvement. A typical example is the ongoing effort for improving the presentation of statistical aspects of scientific research in biomedical journals. Much of what is advised here (e.g. Bailar and Mosteller, 1992) can be transposed to a variety of other fields, and, put into its proper context, can help as a guideline for writing textbooks, certainly those starting with "Introduction to ..." in their title. How many research assistants (and others) from fields outside statistics run into a library and grab an "Introduction to ..." book to look up the statistical tool they need, is a number which is hard to guess. I think it is not low. Hence, carefully writing books, and certainly also basic textbooks, may play an important role in improving the "statistical literacy" of a vague, but possibly not unimportant, number of students, assistants and researchers. That also here some improvement is warranted is illustrated in Callaert (1997).

## REFERENCES

- Bailar, J.C. and Mosteller, F. (1992). Guidelines for Statistical Reporting in Articles for Medical Journals. In: Medical Uses of Statistics, 2nd Ed., J.C. Bailar and F. Mosteller (ed), Boston, Ma : NEJM Books.
- Biehler, R. (1997). Software for Learning and for Doing Statistics, *International Statistical Review*, 65, 167-189.
- Callaert, H. (1997). Nonparametric Hypotheses for the Two-Sample Location Problem, Manuscript.
- Garfield, J. (1997). Discussion of the paper by D.S. Moore, *International Statistical Review*, 65, 137-141.
- Hawkins, A. (1997). Discussion of the paper by D.S. Moore, *International Statistical Review*, 65, 141-146.
- Moore, D .S. (1997). New Pedagogy and New Content: The case of Statistics, *International Statistical Review*, 65, 123-137.
- Tufte, E. R. (1983). The Visual Display of Quantitative Information, Cheshire, Comm.: Graphics Press.
- Wainer, H. (1984). How to Display Data Badly, *The American Statistician*, 38, 137-147.