

Some Principles for Developing Software for Teaching Statistics

Peter Schwanenberg - St Augustin, Germany

1. A survey of the development of statistical applications

The application of statistical methods has changed with the development of data processing. The major periods are:

Before electronical data processing: Computations had to be done via pencil and paper with additional access to tables and calculating machines. For large data sets calculations were tedious and not all methods were practicable. On the other hand, before starting computing one checked very carefully whether there would be a useful result.

Mainframes: Most commonly used statistics were known by the 1930s, but had to wait for computers to become practical. Statisticians were eager users and statistical programs were among the first general purpose programs. There were many difficulties to overcome when accessing a computer and handling the commands and the control languages. The output produced by such programs was voluminous and the interesting results were buried among many additional data.

Personal computers: With the appearance of personal computers, interactive computing and menu-controlled programs became available. The easy use of statistical programs which are as powerful as the earlier mainframe versions led to widespread use by statisticians and even by office staff.

Following the enormous progress in hardware development, high-quality graphics is an important feature of an increasing number of programs.

2. Some deficiencies in statistical programs for teaching

From the teaching point of view, statistical packages have the deficiency that they only produce final results and do not show the intermediate steps. So a student

who fails to find the correct result by a paper and pencil computation will find little help in locating the error from such packages. If a package is to be useful in lecturing, it is essential that it have as an option the possibility of displaying and documenting intermediate steps.

By monitoring individual steps, a discussion and exploration of the data will be possible and the impact of individual decisions in the sequence of computations can be controlled. The graphical display of intermediate results, in particular, offers great advantages, as one can then visualise more complex features than can be indicated by statistical measures which only represent a special aspect. Testing using a given measure always means testing a predetermined aspect, whereas a visual analysis makes it possible to explore the data without predetermined hypotheses.

Another important requirement is the independence of descriptive statistics. For hypothesis testing this means a stricter separation of the description of the sample from the values of the test statistics. A deterrent example is the standard output for multiple regression where the descriptive results for the sample have to be picked out from among a huge set of numbers representing the values of different test statistics. The separation also makes it possible to bring to bear individual decisions on the conduct of the problem-solving process.

Doing descriptive statistics intensively is not always welcomed in the academic area where acceptance of statistical methods often depends more on the complexity of the mathematics than the efficiency of the practical process of decision making. According to such a criterion, methods of hypothesis testing have more standing because they involve probability calculus.

A further important point in teaching is to reveal steps within an algorithm where you have an operational choice from the problem background. The possible choices are not obvious because in this context so-called rules of thumb are used. As the choice of such rules has an influence on the results, teaching requirements demand that:

- (i) Rules of thumb must be documented in the program in such a way that the user always knows what options are available. It should be made obvious when one of several possible rules is taken.
- (ii) If several rules are applicable, the student should be given an option which would allow him/her to test the consequences of each rule.
- (iii) The program should be open so that one can add his/her own rules.

When testing an existing program, one should, for example, look at the process of constructing histograms, especially at the construction of the intervals on the variable axis. Many programs generate intervals of equal length. This will lead to useless results for many problems arising in practice. Some authors recommend data transformations for these cases so that equal-length intervals can be applied to the transformed data. However, transforming the data is an additional difficulty for many students as one often does not know how to choose the transformation and to interpret the transformed data. In this case, the program should give better support for retransforming the data so that one can see the results on the original data. Considering that intervals found by formal rules cannot seriously compete with intervals derived from the real problem, it is an inevitable demand for a statistical program to give the user an option for constructing

intervals of his/her own choice.

3. Example demonstrating the power of graphical analysis

When doing hypothesis testing, the sample is described by some standard measures. This is considered to give a satisfactory impression of the data. The following data sets given by Anscombe (1973) demonstrate that one can have an identical description by the most commonly used measures even where the scatterplots vary greatly.

The data sets $\{(x_i, y_i); i = 1, 3\}$ and (x_4, y_4) shown below have the same summary statistics but differ substantially in general character.

Data						
	1	2	3	4	5	6
	x	y ₁	y ₂	y ₃	x ₄	y ₄
1	10	9.14	7.46	8.04	8	6.58
2	8	8.14	6.77	6.95	8	5.76
3	13	8.74	12.74	7.58	8	7.71
4	9	8.77	7.11	8.81	8	8.84
5	11	9.26	7.81	8.33	8	8.47
6	14	8.10	8.84	9.96	8	7.04
7	6	6.13	6.08	7.24	8	5.25
8	4	3.10	5.39	4.26	19	12.50
9	12	9.13	8.15	10.84	8	5.56
10	7	7.26	6.42	4.82	8	7.91
11	5	4.74	5.73	5.68	8	6.89

Summary Statistics	
x-mean:	9
x-standard deviation:	3.16
y-mean:	7.5
y-standard deviation:	1.94
simple regression:	$y = 3 + 0.5x$
R-squared:	0.67
correlation:	0.82

Examining these measures one might suppose strong analogies in the scatterplots. A graphical analysis using CRICKET GRAPH by Cricket Software Inc. displays significant differences in structure - see Figure 1 below.

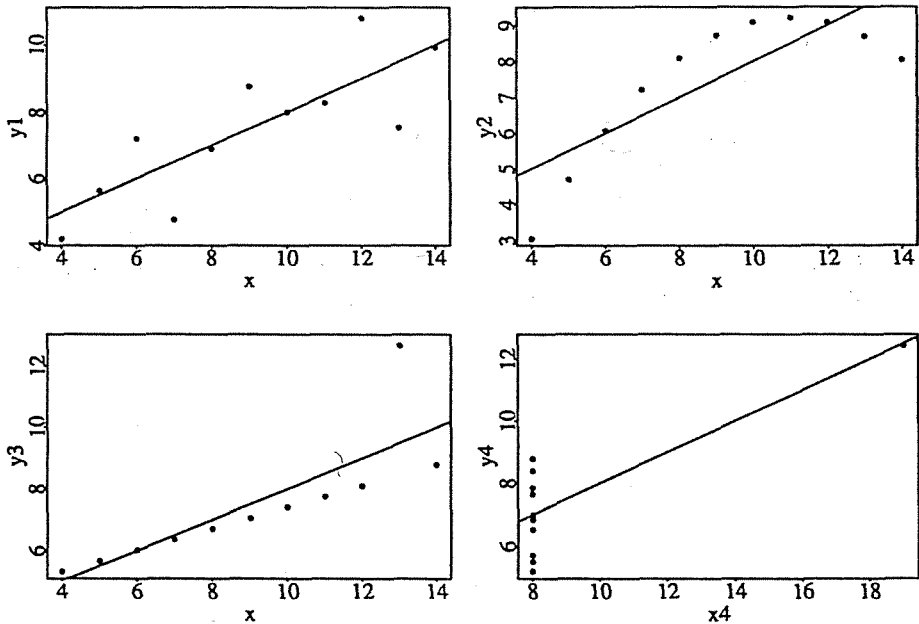


FIGURE 1
Scatterplots of the four data sets

4. Conclusions

Finally, I list some important points for producing statistical software for teaching:

- (i) A direct and elementary access to the data by graphical analysis should be possible.
- (ii) Rules of thumb used in a program should be documented, as well as possible alternatives.
- (iii) Intermediate results should be available as an option, preferably with additional graphical displays.
- (iv) More emphasis on descriptive methods, and a stricter separation of descriptive output and the values of test statistics are recommended.
- (v) The methods available should be offered in a well-organised manner.
- (vi) The program should know about possible meaningful sequences of analysis, and offer methods in a context-sensitive way.
- (vii) The decision process within a session should be open to user input so that users can bring in their personal experience and knowledge.

These points can be helpful in judging whether software can meaningfully be used in teaching statistics. Some aspects are already realised in existing programs, e.g. DATA DESK and LINDO. DATA DESK by Odesta Corporation is a desktop data analysis and display environment for statistics on the Macintosh. LINDO by Lindo Systems Inc. is not a statistical program but intended for linear integer and quadratic programming. It is given here because of its possibilities of getting intermediate results.

Reference

Anscombe, F J (1973) Graphs in statistical analysis. *American Statistician* 27, 17-21.