Streamlined Statistics : An Innovative Polytechnic Approach Reviewed

Mike Camden - Wellington, New Zealand

1. Introduction

In their jobs, many people need to deal with: large sets of data; a feeling that they don't have the mathematical skills they need to deal with this data; and some software - probably a spreadsheet package or graphics package.

These people also have lots of common sense.

This paper describes an approach to statistics which aims to build numeracy and graphic skills onto people's common sense, and to give skills in using software on statistical data. It examines the subject, 3150 "Statistics" for New Zealand Science Technicians, then evaluates our experience at Wellington Polytechnic with this and similar subjects that we have presented to a wide range of students.

2. Statistics for Science Technicians (Statistics 3150)

This introductory statistics subject for the New Zealand Certificate in Science was developed by a group of statisticians and teachers, and launched in 1986. Second and third year subjects followed.

The aims of the introductory course are as follows: First, to introduce students to the most used concepts and techniques of statistics. Second, to enable students to grow in numerate judgement. Third, to achieve skills in collecting and exploring data, representing them graphically, and reporting conclusions. Its contents are set out below. Each topic is allocated a given number of class hours and computer hours. Very deliberately, the subject starts off with Exploratory Data Analysis (Topics 1, 2 and 3). The second and third subjects start off with further work on EDA.

TABLE 1
Topics for Statistics 3150

	Topics	Time (hours) Class	Comp	Total
0	Statistical Literacy	0	0	0
1	Exploring Single Variables	8	6	14
2	Exploring Pairs of Variables	7	6	13
3	Exploring Series Variables	4	5	9
4	Probability Concepts	8	0	8
5	Probability Models	14	2	16
6	Sampling and Data Quality	8	2	10
7	Estimation	6	1	7
8	Statistical Quality Assurance	10	4	14
9	Significance	3	0	3
10	Least Squares Regression	2	4	6
	Total	70	30	10

We had some difficulty persuading the examining authority to accept a topic numbered zero, and with no hours allocated to it! We felt that "Statistical Literacy" should permeate the whole subject, and defined it like this:

Topic 0 : Statistical Literacy: The student should be able to ...

(i) communicate clearly:
  (a) record data and transcribe the results of calculation accurately, and to a suitable precision;
  (b) construct tables and graphs which are clear conveyors of information;
  (c) equip tables and graphs with titles and labels;
  (d) name variables correctly and state their units;
  (e) state the source of data;
  (f) where possible, check that numerical results are reasonable.

(ii) use statistical computer packages:
  (a) use a statistical package for every appropriate objective of this prescription;
  (b) understand the role of computing in data exploration and statistical inference;
  (c) approach new statistical software with confidence.

3. Principles and prejudices

The beliefs, principles, and possibly prejudices, which the group used in developing these subjects are similar to those which guide our teaching. This section states our philosophy for the learning of statistics by our client student groups.

This philosophy is designed for a client group which excludes future statisticians, but includes future and current laboratory technicians, research workers, policy developers, nurses, computer professionals, environmental health officers, electronics and engineering technicians, and various others.

Exploration versus inference: Skills in exploration (or EDA) are a much better starting point, and are much more useful to our target group, than skills in inference. This principle is based on the beliefs that:

(i) inference, especially hypothesis testing, is often done inappropriately;
(ii) if hypothesis testing is to be done at all, it should be left to the professionals;
(iii) even where formal inference is appropriate, an exploration of the data for non-standard features is necessary;
(iv) exploration often yields more information from the data than formal inference, as stated by John Tukey at the NZ Statistical Association Conference in 1977.

An alternative path to useful numeracy: The exploration of data with graphs is an alternative path to skills and confidence with numbers. These exploratory skills can be based on experience, rather than on the mathematical skills which eluded many people at school, and they are among the skills most often needed in today's workplace. We believe that this approach bypasses the blockages of "Maths Anxiety".

Exploring data with graphs involves two sorts of skills: mechanical skills in making graphs, and the much more challenging skills of reading the graphs. Many statistics texts, even some recent ones, spend large efforts on making graphs, and hardly any on interpreting them.

Humans communicated with pictures for thousands of years before they used written words or numbers. However, it took humans until 1786 (Tufte, 1982) to begin publishing statistical data graphics. Our teaching experience makes it very clear that the making and reading of graphs is a skill which must be learnt, but which feels natural once it is learnt.

A healthy diet of raw data: Students flourish if they're fed on a diet of data-sets which are raw, fresh, high fibre, locally-grown, and which contain some bugs and blemishes. The data-sets need to have bulk, which means they are multivariate. Our reasons? First, real data is full of interest for learners (and full of surprises for teachers). Second, our client group will have to struggle in their jobs with data that has missing values, clerical errors, and unexpected features. Third, our client group will need to read research literature, and will benefit from a level of cynicism about the data on which the research is based. And finally, real data-sets are always multivariate, and relationships among variables are much more interesting than the properties of a single variable.

Time-series data-sets are readily available to teachers, though they are often "cooked" rather than raw. Case data-sets are harder for teachers to come by. Hence, the NZ Statistical Association published the *Data Bundle* (Camden, 1989), a collection of 25 data-sets, with suggestions on how they can be explored.

The place of software tools: A statistical package with graphical features is just as essential to the learner as a pencil. Students should use a package from the start of the course onwards, to store and edit data-sets, to do calculations, and to draw graphs.

Software is developing very fast, but not necessarily in the ideal direction for our purposes. Perhaps non-statistical software developers may soon realise what the statistical uses of graphs really are!

We are faced with two key questions.