
Review of Seeing Statistics

Dr Sandra Bonellie
Napier University

s.bonellie@napier.ac.uk

Seeing Statistics is an introductory statistics course in the form of web based interactive material. The author states that 'Statistics is inherently a geometric discipline' and aims to introduce various statistical concepts through the use of interactive graphics. The version reviewed here is release 1.0 which includes material on displaying data, summary statistics, binomial and normal distributions, inference and confidence intervals and correlation and regression. Release 2.0 will also look at multiple comparisons, categorical data and non-parametric statistics.

The material is accessed through the Web using a serial number which is supplied with the book which accompanies the material. The book gives instructions on how to use the material on the Web and is not a textbook in itself. It has a chapter on navigation which runs through a sample tour of the Web site using as an illustration the chapter on Describing the Centre. This is almost not needed since the Web site is well set out and easy to navigate.

I reviewed Seeing Statistics using Netscape 4.6 running on Windows 95 and experienced no serious problems with loading or running the material. The only minor problem on the technical side which I encountered involved a simulation of tossing two dice. The dice should have fallen independently but this did not work for the browser I was using. However a warning message appeared alerting me to the problem.

Structure

On entering the web site a Table Of Contents can be obtained. There are 14 chapters listed of which all but 3 are available in Release 1.00. Typically a chapter begins with an overview of the topics covered in that section. A section entitled 'Why am I learning this?' then summarises what could be described as the learning outcomes for that chapter.

The author states three guiding principles used in designing the material. These are that the material should be visual, active and engaging. At the core of the material therefore are a set of applets which provide an opportunity for the student to interact with the material. These fall into 3 broad categories namely applets which involve:

- simulation, eg rolling a die or sampling from a specified distribution
- interactive graphics, eg moving a line on a scatter plot and observing the effect
- calculation, eg obtaining binomial or normal probabilities

The applets involving simulation worked well and show the opportunities which are available using the web which would not be possible using a standard text book.

In general I found the material to be well set out. Each screen was clear and easy to read and was suited to the medium. The temptation to overload the screens with text was avoided by providing links to glossaries or to additional help as required.

By clicking on a Discovery icon at the side of the screen the student is given a set of questions to guide them though the current section. The following is an example of the discovery material from the section on the median.

Seeing Statistics
by Gary McClelland
ISBN 053437090X

published by
Duxbury, Thomson Learning

'What would happen to our estimate of the centre or typical value if, due to a typo, we had entered 140 instead of 14 for the highest value? Would the estimate of the centre change? Why not?'

These questions are clearly designed to prompt understanding by encouraging the students to think for themselves. In general the answers to such questions are to be found by working through the material which follows. There is not however a direct link and I did feel that it would have been helpful to supply guides to the answers either directly or by providing a link to part of the text which answers the question.

Each chapter ends with a set of exercises for the student to work through. Again a disappointing feature of this was the lack of solutions to these examples. On a few occasions a Check Up icon gives access to a question where the student is required to enter an answer such as a mean or a median. Information on whether the answer is correct or not is then given together with an explanation. More use could have been made of this facility.

Applications

Most of the topics covered in the material are illustrated using data from a true/false quiz on statistics. Students are invited to take try the quiz for themselves and a theme throughout is what score someone with no statistical knowledge might expect to attain given he or she guesses the answer to each question. An example of the type of questions included in the quiz is given in Fig 1. It is made clear that they are not expected to know the answers at this stage. However, while the idea of asking

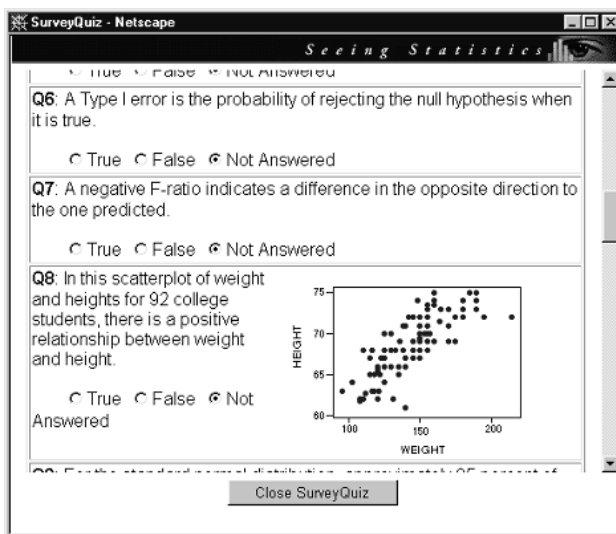


Fig 1 Questions from the statistical knowledge quiz

the students to participate in the data collection is commendable, I do wonder if this quiz would not be off putting for most people.

In addition to the examples given in the main body of the web material, each chapter includes an Applications icon which allows the student to access data from psychology, business, engineering or biology. There is also a facility for the student to enter their own data and to work with this.

Content

While finding some useful simulations in the material on the whole I found the statistical content of Seeing Statistics was disappointing and in some cases misleading.

Simple plots

The first graphical technique which is suggested is called a simple plot and is used to illustrate a univariate set of data by plotting the observations against the observation number. A plot of this type is not appropriate unless there is some ordering inherent to the observations. The data used to illustrate this plot is a set of scores from the statistical knowledge quiz, plotted against observation number. Although the author does not suggest that there is any inherent ordering to the scores, the use of this plot is potentially misleading.

Describing the centre

The main reason for the use of the simple plot would seem to be to develop the plot to illustrate measures of location and spread. The median is introduced as the measure of centre which minimises the sum of the absolute errors and the mean as the measure minimising the sum of the squares of the errors. These concepts are illustrated using an applet of a simple plot where the student can move the line representing the centre of the data up or down and observe the effect on either the sum of the absolute errors or the sum of squares.

Using this approach the fact that the median is the middle value of the data set is secondary to the fact that the median minimises the sum of the absolute errors. This seems to complicate unnecessarily a straightforward concept which most students have no problem dealing with.

The reason for taking this approach seems to be to illustrate the concepts of mean and median geometrically. The applet to illustrate that the mean minimises the sum of squares is shown in Fig 2. The

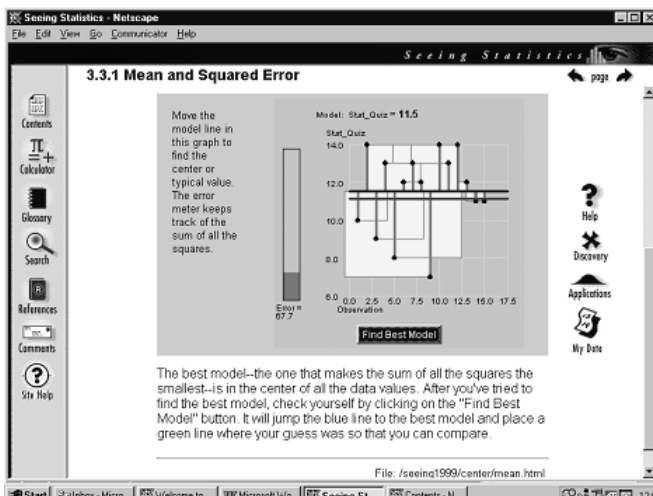


Fig 2 Describing the Centre: Introduction to the mean

'squares' for the error at each data point are drawn and change size as the line representing the centre of the data is moved up or down. Personally I did not feel this would encourage any greater understanding of the mean in students. In fact the reverse might be the case.

Normal Distribution

The normal distribution is introduced by using the scores from the statistical knowledge quiz again. There is mention that this is the essentially the normal approximation to the binomial distribution at the end of the chapter but no reference is made to the distinction between discrete and continuous distributions which must give the student a limited understanding of the topic.

On a more positive note, the author provides applets to calculate normal probabilities which also show the corresponding area under the normal curve. In addition explanation of how to calculate the probabilities using tables is given, again with each probability illustrated graphically.

Correlation and Regression

These topics lend themselves to the interactive, geometric approach taken by this material. Applets which allow the student to move a line about a scatter plot and observe the effect on the gradient and the intercept work well. Also useful are applets which

allow the student to see typical graphs for different correlation coefficients. These are illustrated in Fig 3.

Unfortunately the accompanying text states, incorrectly, that: 'The higher the correlation the more often the points will fall on a line. Also the line itself will be steeper'

Conclusions

This material makes good use of the facilities available on the web. The underlying philosophy of helping students gain understanding through actively engaging with the interactive graphics is a sound one. The technical side of the material is of a high standard and the material is easy to use and to navigate.

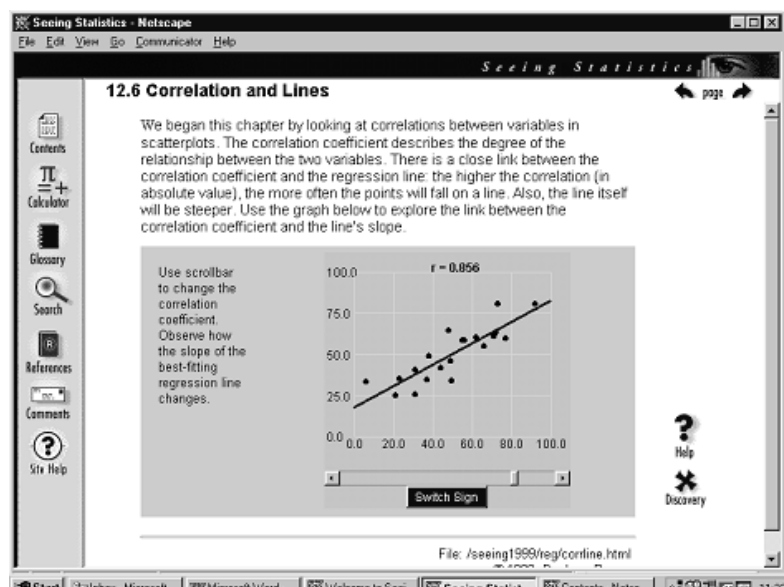


Fig 3 Correlation and Regression

However there are some topics where the desire for a good interactive graphic seems to lead to unnecessary complications such as the approach taken to introduce the median and the mean. There are also occasions when a full or adequate explanation of the topic is not given as in the treatment of the normal distribution. It is also worrying to see misleading statements occurring in the text.

Overall some of the individual applets would be useful to illustrate particular points either as a demonstration to a class or for use by individual students. The shortcomings of the statistical content however are such that I could not recommend Seeing Statistics as suitable material for an introductory statistics course.