

CAST of Thousands

Mark Brewer
BioSS, The Macaulay
Institute

m.brewer@bioSS.ac.uk



CAST stands for Computer-Assisted Statistics Teaching, and is an on-line learning tool, viewed from within a Web browser, developed by Doug Stirling at Massey University, New Zealand. Doug has a good number of years' experience in developing computer-based learning materials, and all this experience is in evidence in the current version of CAST, v2.1.

Contents of CAST

CAST consists of a complete introductory course in statistics, augmented by two advanced modules. There are two completed alternative versions of CAST: a "general" version; and a "biometrics" version, which contains examples slanted towards biology, agriculture and the health sciences. A third version ("business") is under development, containing examples aimed at business and economics students. Further, each of these different flavours of CAST have both "student" and "lecturer" editions, the latter containing little text but larger versions of the interactive diagrams—ideal for demonstration using a projector of some sort in lectures.

The topics covered by CAST are:

1. Introduction
2. Displaying Variation
3. Bivariate Numerical Data
4. Time Series
5. Categorical Data
6. Data Collection
7. Sampling and Variability
8. Inference
9. Regression Models
10. Comparing Groups
11. Contingency Tables

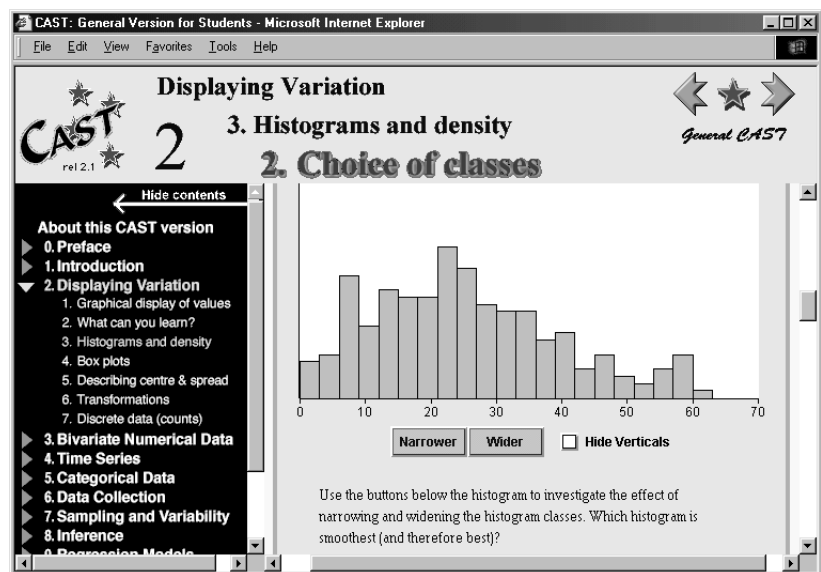


Fig 1 Different histograms from same data possible

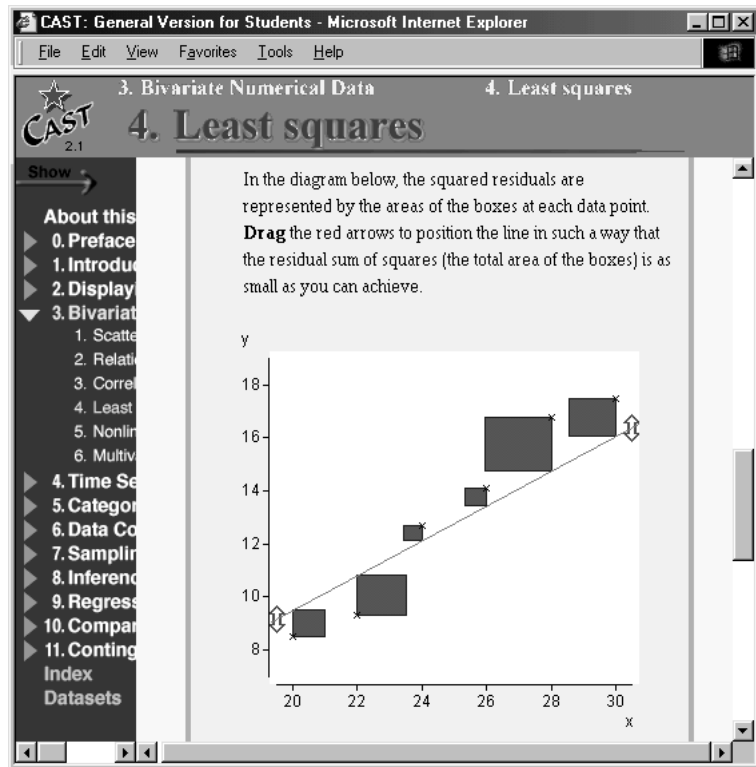
Fig 2 Least Squares illustrated

To a large extent, the topics above should cover the material in any standard University first-level course. Topic 4 is the only one to look slightly out of place, perhaps (at least in a UK context). That Topic, “Time Series” is present only in skeleton form in the biometrics version.

After a gentle introduction to the ideas of data collection, storing and interpretation, Topic 2 describes methods for displaying and summarising data. Figure 1 illustrates the look of a CAST window in this Topic, when using Internet Explorer on Windows 98. When quite deep into the material, a frame at the top of the browser screen provides navigation and context information—we can see that we are in Section 2.3.2, and that we can move forwards or backwards one step via the arrows in the top right. On the left-hand side of the window, we see a contents listing, enabling us to jump rapidly from one Section to another.

This listing of contents can be shrunk to enable more of the actual learning material to be viewed (“shrunk” appears more accurate than “hidden”). The main body of the window in Figure 1 shows part of an applet where the user can obtain different histograms from the same set of data, increasing or decreasing the bin widths incrementally. Thus, the user gets to see and understand an important detail of data display by direct experimentation, rather than simply being instructed by the text. This is a clear illustration of the ethos behind CAST; the user is never very far from having something to do—and something genuinely interesting and informative, at that.

A chapter on Bivariate data follows in Topic 3, covering scatterplots, correlation, basic linear regression and multivariate data display. The issue of least-squares is explained with the aid of the applet shown in Figure 2; dragging the small arrows on the plot moves the line, and the aim is stated of making the total area of the squares as small as possible. This is a clear and intuitive way of demonstrating what can appear a complicated concept when studied mathematically. Figure 2 also illustrates the appearance of the contents listing after “Hide contents” (see Figure 1) has been selected; usefully, the user is still able to navigate through CAST in this mode. Topic 4 on Time Series deals with running mean-type smoothers, exponential smoothing, structural time series and control charts.



Topic 5 covers categorical data: after discussing issues of data display, a brief discussion on logistic regression is given, where the user has the chance to try to fit a logistic curve by eye (this section is not overly technical, given the notional level for the material). Topic 6 discusses data collection, including sections on the interpretation of relationships (i.e. cause and effect, and so on), survey design, and experimental design. Following logically, Topic 7 is concerned with sampling theory and measuring variability. Here, the Normal distribution is introduced formally, and the user has several opportunities to play with sliding scales, as in Figure 3. Here, the sliding scales change the mean and variance of the Normal distribution; the X scale adapts to represent the different location and scales, but the Z scale stays fixed—showing the usefulness of the “standard” Normal.

$$\chi^2$$

The final topics deal with the more complex material. Topic 8 covers estimation and inference, including calculation and interpretation of confidence intervals and hypothesis testing. The estimation section goes as far as kernel density estimation—see Figure 4. Here, the sliding scale varies the kernel function bandwidth, and the resulting density plot illustrates clearly the effect on the smoothness of the density, highlighting the bias-variance trade-off. Lastly, Topics 9, 10 and 11 cover regression, analysis of variance, issues of marginal and conditional probabilities, and the distribution.

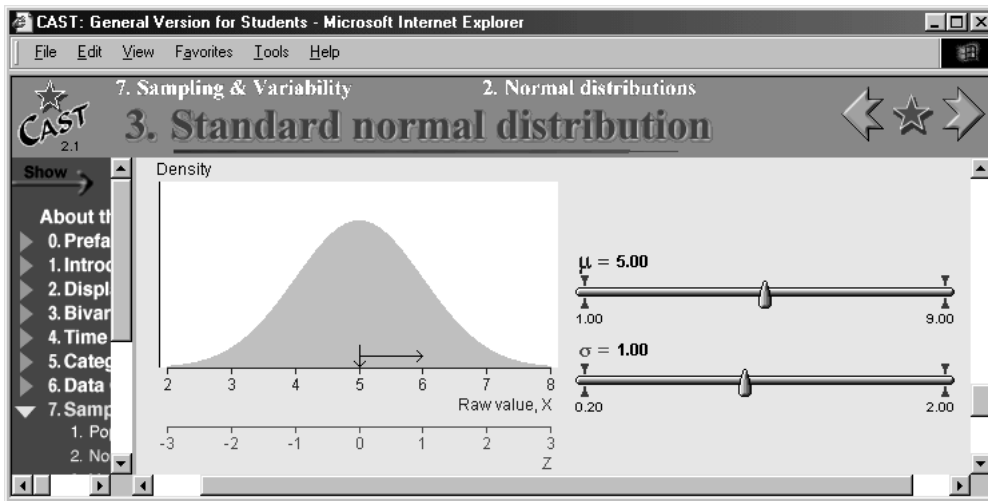


Fig 3 Normal distribution and sliding scales

There are also three further advanced modules available as case-study illustrations and extensions of the concepts learnt in the main CAST program: one on nonlinear regression, one on simulation and sport, and one on multiple regression (this latter is currently under development). To illustrate, the simulation and sport module begins with a simulation of a tennis match; sliding scales are used to vary the probabilities of each player winning a point given serve, and the effect of these probabilities can be seen by generating an entire match worth of points. Following this, there are further simulation-based studies of football, golf handicaps and athletics.

There is a further option here, where access to CAST is regarded as crucial—a “pure” stand-alone version not requiring access to Massey University can be bought for upwards of US\$150. However, having used CAST on numerous occasions, I have not encountered a problem with access.

Accessing CAST

Applets Aplenty

CAST can be accessed entirely over the Internet, at <http://cast.massey.ac.nz/>

The style of CAST involves the use of very many interactive java applets. The applets are, in general, extremely well thought-out and implemented. It is these applets which provide *genuine interactive elements* to CAST—this interactivity is not merely a selection of hyperlinks or

Alternatively a “stand-alone” version of CAST can be downloaded and installed for use either on individual PCs or over an Intranet. Both of these options involve entering CAST through a log-in screen—for educational use, there is currently no charge for the PIN code required. The checking of the PIN code requires connection to a server at Massey University, and although this should ordinarily take no more than a few seconds (even over a modem), there is always the possibility of network/machine problems disrupting either a computer workshop or live demonstration.

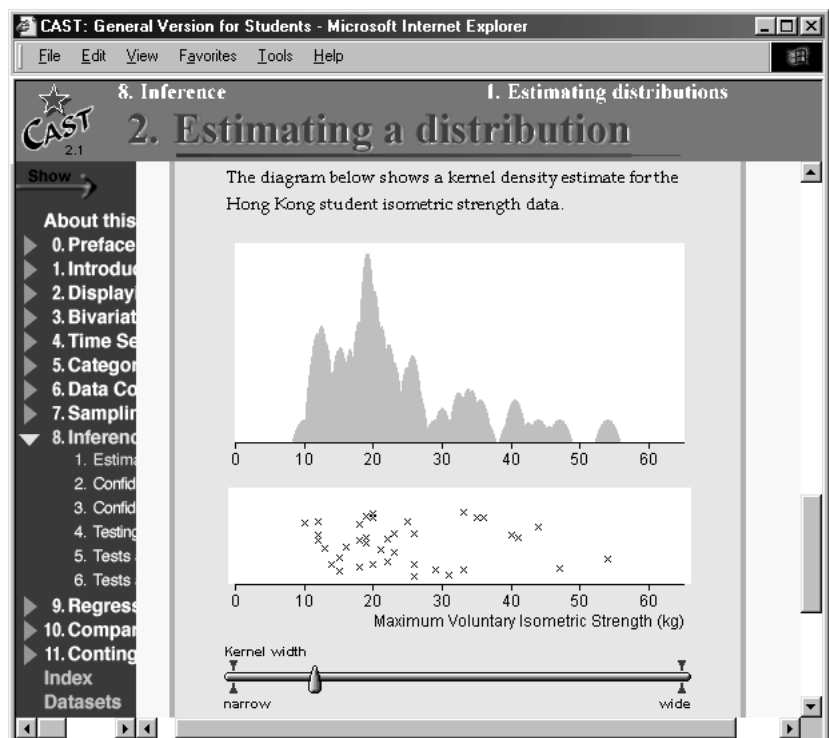


Fig 4 Kernel density estimation

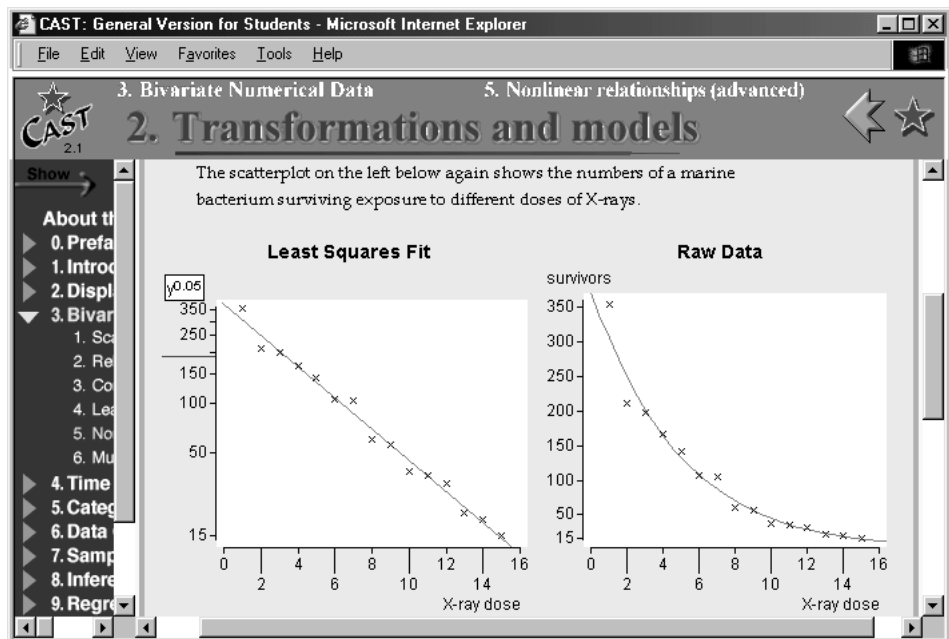
Fig 5 Making it clear that log transformation is appropriate

pop-up windows, it is a sequence of explanations of concepts which utilise fully the tools at hand—in this case, the Java capability of most modern browsers. The current version of CAST has browser requirements at log-in which mean that the system will not work with Netscape browsers higher than 6.0, nor with Mozilla. All recent versions of Internet Explorer and Netscape 4.5 to 4.7 are known to be compatible with CAST's log-in procedure. The aforementioned "stand-alone" version of CAST, however, is compatible with any (recent) browser, since in that case the log-in procedure is not required.

Examples of several applets within CAST have been provided already, but Figure 5 shows yet another. The left-hand plot shows a transformed response variable plotted against an explanatory variable. The form of the response transformation can be varied by moving the short horizontal line on the y-axis. This changes the scale of the y-axis too, so that one might observe the effects of the transformation—indeed, further to this, the effects on the original scale are shown in the right-hand plot. Via this interactive mechanism, the workings of the Box-Cox family of transformations are easily understood. In the example of Figure 5, it is clear that a log transformation is appropriate.

The success of the response transformation applet is indicative of the applets as a whole; issues dealt with by other applets not yet mentioned include: the dangers of extrapolation; rotating 3D scatterplots; the meaning of "95% confidence"; between- and within-group variation in ANOVA; and many, many more. Hopefully the examples described will have provided enough of the flavour of CAST.

Of course, this article (on a static medium) cannot hope to do justice to the interactive elements of CAST—the only solution is to *try it for yourself!*



Summary

CAST is a perfect illustration of how on-line learning *should* be. CAST is not a hastily-converted set of static lecture notes, with a handful of dynamic features thrown in as an afterthought. Rather, it is a complete system making full use of the fact that the user is sat at a computer. The interactive features are not merely pointless "games", nor are they simply there to provide a break from reading text from the screen. Instead, they are genuinely useful learning tools in their own right, demonstrating complex concepts, analyses and mathematics.

While CAST may not be best suited for use as the sole learning mechanism for any single University-level course, it contains many elements which lecturers could either direct students towards (perhaps via direct links in their own Web resources) or use in lectures/tutorials as a means of exposition.

In conclusion, CAST ought to be seriously investigated by (a) anyone teaching introductory statistics, no matter what the level, and even (b) anyone considering the development of a (scientific) on-line teaching resource.