

SOFTWAREING 'INSIGHT'

Peter J. Smith, Department of Statistics and Operations Research,
RMIT University, Australia

How do we gain insight into statistical processes using statistical software? The key word here in the context of tertiary education is 'insight'. At RMIT University, all our undergraduate degree statistics subjects (both theoretical and applied) have weekly laboratory software classes using selections of MINITAB, EXCEL, SPLUS, STATISTICA, SAS and SPSS. Helpings from this menu are designed to prepare students for real-world problems with a diverse range of software choices, not in the context of blackbox training, but rather a considered utilisation which doesn't die when the software dates.

We attempt to give students insights into the impact and structure of say, boxplots, as a beneficial data-analytic tool, rather than giving them a specific training in interpreting MINITAB boxplots or STATISTICA boxplots which fails when outliers are drawn differently or axes swapped. This encourages teaching which is software rich, but software independent. Clearly, we all must use software, and use it well. This is a stimulating and challenging reality. Much software is more advanced than the students' ability to understand its statistical workings. Our challenge is to take the lead from software---it is often clever, rapid, lucid and short-cutting---and use this sophistication to teach with insight.

For example, in teaching elementary statistical theory, probability mass functions are often algebraically defined and manipulated. A poisson mass function may be found as a calculus limit from binomial mass functions with increasing n and shrinking p . This may no longer be a satisfactory explanation for the poisson approximation to the binomial for statistics students weak in calculus. Their insight may be triggered by self-made graphics of binomial mass functions with increasing n and shrinking p , all tiled comparatively across a screen in a student-centred live laboratory session. Other notable examples of significant insights include: software simulations of cumulative winnings at a roulette wheel; the on-screen construction of a likelihood as a curve with an obvious maximum.

Our final example shows how software may be used to cut the inference. often, we put off exposing students to statistical inference until ‘the second half of the book’ ---after probability, random variables and data gathering. We can take a lead here from software: calculations for the ‘standard error of the mean’ are simply (routinely) done on a single column of numbers. The software doesn’t wait for probability theory before supplying the answer. The standard error concept is thus moved up front, to ‘the front of the book’, to the very first time that a mean is mentioned. The inference aspect can be teased out using live sampling distribution simulations held in stem-and-leaf plot form. At least then there is reinforcement for the basic reporting of means using standard errors rather than standard deviations. Here, not only can software help give us insight into the behaviour of means, by its indifference to textbookish theory, it can help us get our reporting right.