

TEACHING BASIC STATISTICAL CONCEPTS USING COMPUTER DISPLAY TOOLS

Siva Ganesh and S Ganesalingam, Institute of Information Sciences and Technology,
Massey University, New Zealand

This talk provides a brief demonstration of some of the computer intensive tools available for teaching important basic statistical concepts to introductory level tertiary students. The programs/packages demonstrated are ActivStats (a PC or Mac based tool that includes video clips, simulation, animation, narration, text and interactive experiments; also links with student edition of DataDesk), DISCUS (a PC based interactive, dynamic and adaptable suite of EXCEL based worksheets/macros), QUERCUS (a PC based statistical tutoring package suitable for students majoring in biological sciences; also links with MINITAB as a background tool), and Models'n'Data (a highly interactive Mac based program for teaching statistical concepts).

INTRODUCTION

Teachers of 'Statistics' have been discussing extensively the difficulties encountered by students in introductory level statistics courses. The major complaint by the students is that the fundamental concepts of statistics are abstract! Introduction of computer technology and multimedia to illustrate fundamental concepts can be regarded as a significant and innovative step forward. The hope is that the students will be able to digest these concepts by means of computer simulations in an interactive nature.

One major aspect of computer based teaching of statistics is that the basic statistical concepts can be illustrated easily using graphical and other display tools. This conference presentation attempts to provide an overall introduction to four of the commonly used software packages for teaching/learning basic statistics. Each of these packages is given a basic introduction followed by a description of what aspects of statistical concepts they cover and an example session, in separate sections of this paper. Information given are precise, but the intention is to give an impression about the tools and to direct the readers to (web)sites where more information can be acquired.

ACTIVSTATS

This software package presents an introductory statistics course in a multimedia format. ActivStats [1] can be used on both Windows and Macintosh platforms. The structure includes interactive experiments, simulation, animation, narration, text and video clips.

Statistical concepts covered are the material contained in any standard basic course. These include ideas of variation and distribution; regression and correlation; sample surveys and designed experiments; simple ideas of probability; and the basic concepts of hypothesis testing and confidence intervals, comparing two means, inference for proportions and inference for regressions. The main feature that make ActivStats unusual is the extensive use of video and animated illustrations (a multimedia concept), which describe the background to data gathering. The student edition of DataDesk [2] (a statistical and graphics package) is linked to the use of ActivStats. Reviewers describe ActivStats as a 'rich learning environment', see Currall et.al. (1997), for a detailed review.

An example

ActivStats presents its lessons in a 'lesson book', which has three general presentation areas. Each area responds to a single mouse click and operates in a new window. The following shows a lesson book from section 5.1 Density Curves, of Normal distribution. This falls under the general topic of Chapter 1 'Understanding Data'.

The screenshot shows a window titled "5-1 Density Curves" with a blue title bar. The main content area is titled "Normal Distributions" and contains the text: "and, in later lessons, a basis for going beyond the data to draw broader conclusions." Below this is a section titled "Density Curves" with the following text: "Histograms show the relative frequency of each interval along the range of data values by the size of the bar over that interval. This lesson discusses a way to make that idea more general." This is followed by an icon of a histogram and a red "Activity" icon with the text: "How should we generalize histograms? Histograms use the size of each bar to indicate the relative frequency of values in the corresponding interval. Density curves generalize this idea." Next is the text: "Density curves relate relative frequencies to intervals of data values. Thus, they help us to know which values are common and which are rare." This is followed by an icon of a hand pointing to a graph and a red "Activity" icon with the text: "Practice relating relative frequencies and values. Density curves relate intervals of data values to their relative frequencies. This Activity shows how this relationship works." Below this is the text: "Density Curves satisfy only a few properties." This is followed by an icon of a person with a pencil and a red "Activity" icon with the text: "Review the properties of Density Curves. This Activity offers a quick review of the properties of density curves." At the bottom, the text reads: "Density curves are used throughout statistics to relate intervals of data values to the relative frequency of observations at those values." The window has a vertical sidebar on the right labeled "Index Glossary Marks" and a footer in the bottom right corner that says "5-1 Density Curves".

Clicking on the first icon “How should we generalize histograms?”, plays a video clip explaining the basics and gives a chance (or more) to explore further by means of ‘practice questions’. Similar activities are carried out with the remaining icons, in particular, the last one reviews the lesson by means of simple ‘fill in the blanks’ questions.

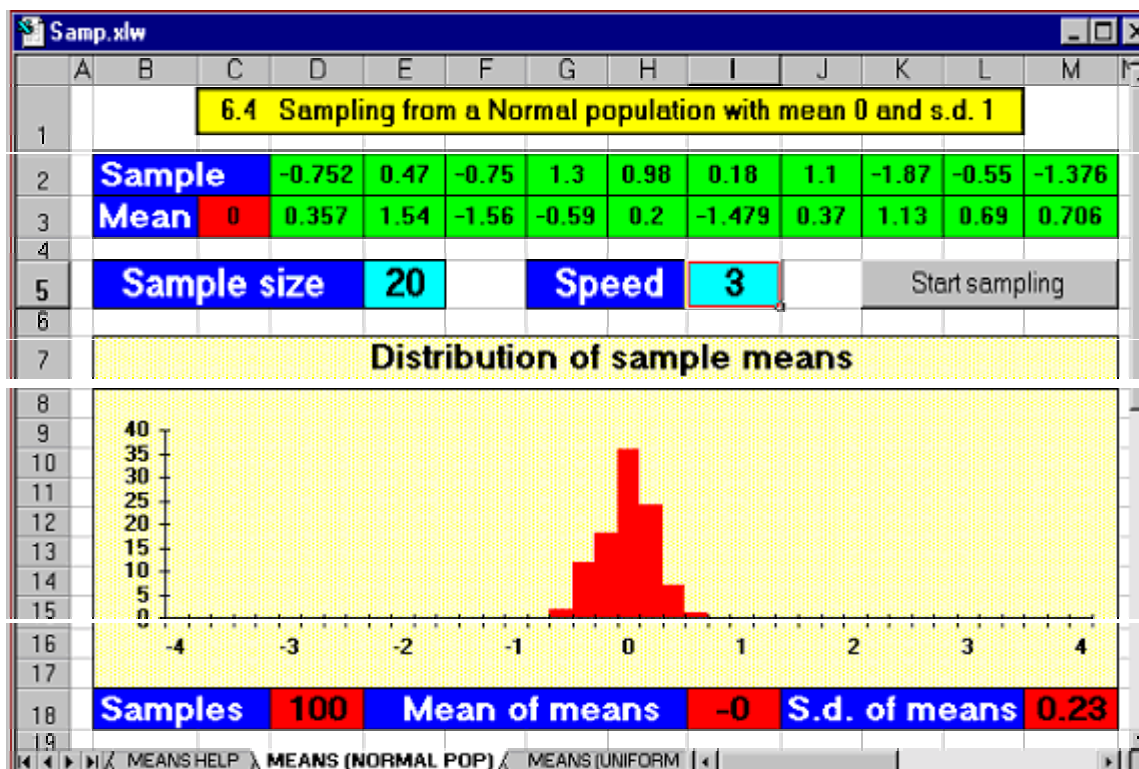
DISCUS

DISCUS [3], Discovering Important Statistical Concepts Using Spreadsheets, a suite of EXCEL workbooks, is claimed to be an interactive, dynamic and adaptable package for teaching statistics at introductory level. Gray (1996), claims that DISCUS contains a selection of data sets, charts, tables, and graphs of various kinds including simulation exercises, and the materials are designed to be used “off the shelf by students working on their own”. The workbooks cover the following topics: Descriptive Statistics, Probability, Binomial and Poisson Distributions, Continuous Distributions, Sampling, Regression and Hypothesis Testing. The package includes printed work-materials describing the aim of various sections followed by information on what appears on the screen and what the students need to know. The printed material also includes, a detailed list of ‘what to do’ exercises, which involve interactive tasks and ‘what if?’ and other questions. As EXCEL workbooks are portable, DISCUS can be used on both the Windows and the Macintosh platforms.

An Example

A brief look at the workbook on ‘Sampling’ is considered here. This workbook (like the others) contains a ‘Welcome’ worksheet followed by ‘Introduction’ and ‘Help’ worksheets describing the contents. The other worksheets are set-aside for examples as well as interactive learning. The following figure shows the worksheet for investigating the sampling distribution of the mean when sampling from Normal populations.

The only cells in the worksheet that can be altered to examine the behaviour of the distribution are the ones next to ‘Sample size’ and ‘Speed’. These are shaded in ‘light blue’. The window shows the results of simulating 100 samples, each of size 20, including the histogram of these 100 sample means.



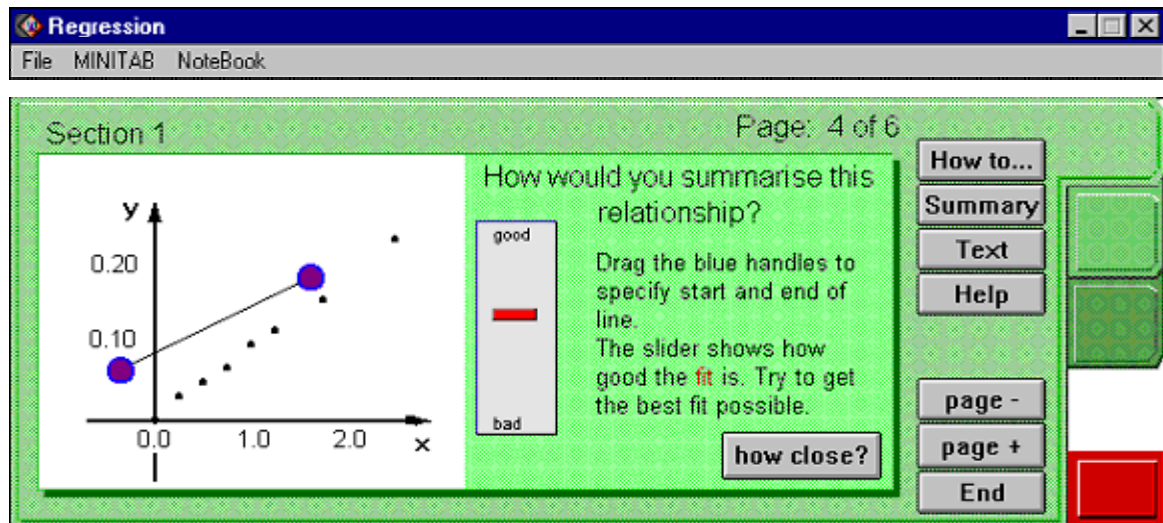
QUERCUS

QUERCUS [4] is an interactive statistical tutoring package, aimed at helping bioscience students learn basic statistics. It uses MINITAB [5] as a background tool for carrying out various analyses. QUERCUS consists of two types of module, workbooks and exercise books, plus a variety of ancillary and resource materials including an introduction to MINITAB. There are twelve workbooks; each is intended to be run in a one-hour supervised computer laboratory session, and covers topics such as Visualising Data, Measuring location and variability, Sampling distributions, Confidence intervals and Significance testing, Regression and Correlation, Residuals and transformations, and ANOVA. There is an exercise book for each workbook, providing extra practice with the material covered, and a short self-assessment test for students and teachers to monitor progress and problems. A student guidebook is available to take the learner through the various modules and exercise books. For a detailed review of QUERCUS, see Gray (1997).

An Example

The module on 'Exploring Relationships I: Regression' is considered here. It consists of sections on 1. Exploring Relationships, 2. The linear model, and 3. Making predictions. For

example, the following figure shows how, by interactive simulation, the students could move a line around the screen to find the one that best models data points.



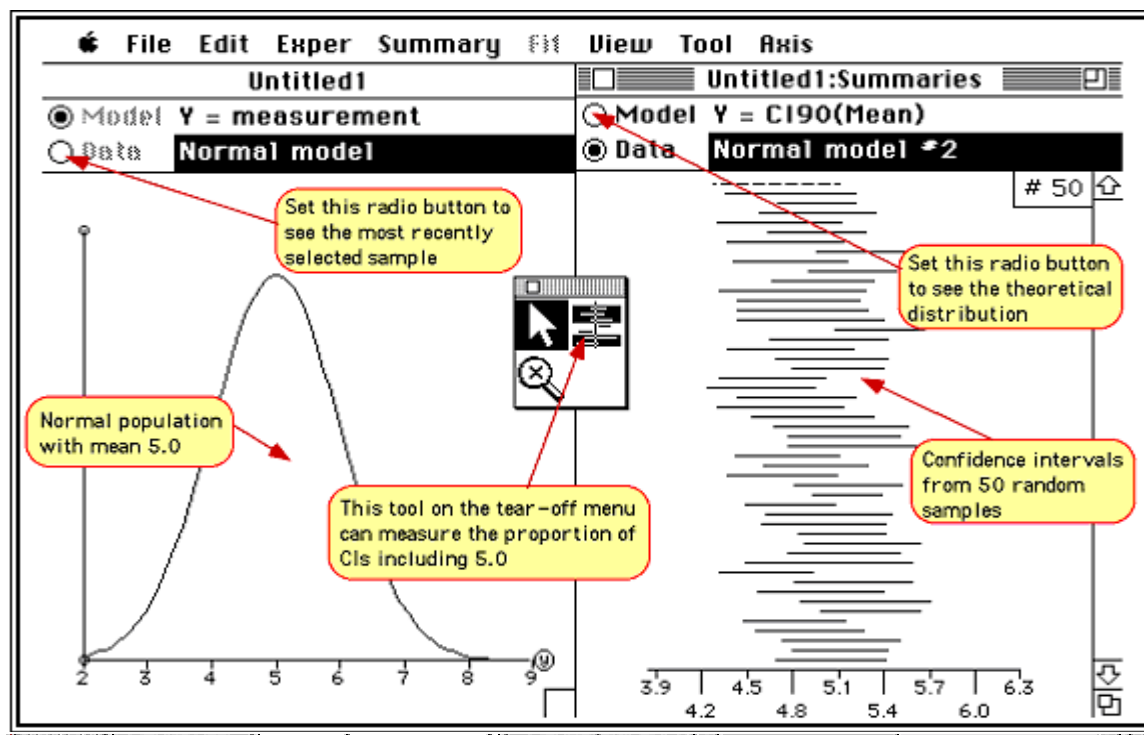
Note that MINITAB (and a Note Book) can be started from the menu; the 'How to ...' button gives help on using QUERCUS; 'Help' button gives hints relating to the task on each page; and 'page +' and 'page -' buttons help move on to next/previous tasks.

MODELS'N'DATA

Models'n'Data [6] has been designed to help teach statistics in a flexible manner in which a wide variety of exercises and demonstrations can be performed. Not only does it display and analyse real data sets, but it can also create, manipulate and display a wide variety of statistical models. With Models'n'Data you can demonstrate most statistical concepts in introductory courses and many from more advanced courses. By repeatedly generating data from models and then examining these repeated random samples, you can build up sampling distributions of many types of summary. You can fit models to data, graphically, and also graphically display data in a wide range of ways (most of the displays can be interactively modified with tools and controls). The display and analysis features cover everything required in introductory statistics courses and exceed the features in many 'major' statistics programs. Sets of exercises have been written (see for example Stirling (1995) to help teach the concepts. Models'n'Data is available only on the Macintosh platform.

An Example

A simple example on showing how simulations can demonstrate properties of confidence intervals is considered here. The demonstration is based on a normal model (mean 5 and standard deviation 1) with sample size 20. The annotated figure below shows the frame where the demonstration is displayed.



The command 'Take Sample' from the 'Exper' menu generates and displays a random sample. Repeated use of the command displays different random samples to show their variability. The command 'Confidence Intervals' from the 'Summary' menu presents a dialog box where a 90% confidence interval for the mean can be specified. The student indicates that the sampling distribution is to be built up, and a second window appears on the right of the screen displaying the confidence interval graphically. Subsequent use of the command 'Take Sample' automatically calculates the confidence interval from each sample selected and these are displayed in a horizontal stack, indicating the sample-to-sample variability of the confidence intervals.

SUMMARY

In conclusion, each of the four software packages outlined here may be regarded as a resource which students can consult in their own time to learn or reinforce the ideas taught through an introductory statistics course. ActivStats is very impressive since it makes good use of multimedia and integration with DataDesk. The package DISCUS could well be regarded as an excellent addition to the study material available for students, in particular, the animated look at sampling distributions and the other appealing simulations make this product a worthy acquisition. QUERCUS is a sophisticated and an impressive product, and should be a valuable additional resource for those learning (and teaching) statistics at elementary level in the context of bio-science. Models'n'Data is a program for teaching statistical concepts to a general audience, whether at introductory or advanced level. It not only displays and analyses real data sets, but it can also create, manipulate and display a wide variety of statistical models. As with any teaching, the benefits of computer-based student learning of basic statistics needs to be assessed effectively!

REFERENCES

- Currall, J., Young, S. and Bowman, A. (1997). Software Reviews, CTI Statistics[#].
 Gray, R. (1996). Software Reviews, CTI Statistics[#].
 Gray, R. (1997). Software Reviews, CTI Statistics[#].
 Stirling, D.W. (1995). *Statistical Exercises Using Models'n'Data*. Wiley (1995), Brisbane.*

[#]<http://www.stats.gla.ac.uk/cti/>

- [1] ActivStats, Data Description Inc., New York, <http://www.datadesk.com>
 [2] DataDesk, Data Description Inc., New York, <http://www.datadesk.com>
 [3] DISCUS, Hunt, N. and Tyrrell, S., email: DNHunt@coventry.ac.uk
 [4] QUERCUS, STAMS, Glasgow, <http://www.stams.strath.ac.uk/external/quercus/>
 [5] MINITAB Inc., State College, PA, <http://www.minitab.com>
 [6] Models'n'Data, Stirling, D.W., <http://fims-www.massey.ac.nz/~stats/ModelsAndData>