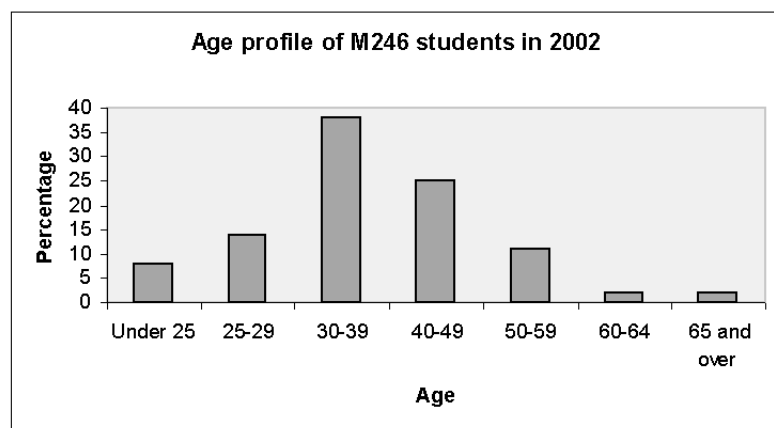

Distance Teaching of Statistics to Undergraduates at the Open University

Jane Williams
The Open University

j.e.williams@open.ac.uk

At the beginning of this academic year, February 2002, there were 27,817 registrations on courses run by the Faculty of Mathematics and Computing. This figure includes post-graduate students and students registered in Singapore. Of these, 7,055 registrations were for level one entry courses and 1,863 registrations were for level two and level three statistics courses.

The ages of our students range from 18 to those who are over 65. As an example, the following chart shows the age profile of students registered on M246 *Elements of Statistics* for the February 2002 presentation of the course.



There are no entry qualifications and students are advised as to the best place to begin their studies by Advisors and Academic Staff based in the thirteen Regional Centres throughout the UK. To continue with the M246 example, 21% of the students were completely new to the University and 19% had qualifications below those required for enrolment at a conventional university.

What courses do they study?

MU120 *Open Mathematics* is a 30-point entry course for students with only basic (literacy and) numeracy skills. The course aims to build confidence in mathematics and to incorporate mathematical thinking into everyday activities. At the beginning of the course, students are introduced to statistical ideas including summary statistics, index numbers, and graphical representations. The use of a graphics calculator is integral and ideas of regression are introduced towards the end of the course.

MST121 *Using Mathematics* is a 30-point entry course for students who are more confident in their mathematical abilities. It is arranged in four blocks, the final one being devoted entirely to statistics and covers basic rules of probability, probability distributions including geometric and normal distributions, sampling distributions, Central Limit Theorem, confidence intervals, hypothesis testing and fitting lines to data. Use of computer software (Mathcad and OUStat) is part of the course.

Also presented in Singapore.

MDST242 *Statistics in Society* is a 30-point course at second level, suitable for students from across the University who need to include some statistics in their degree profile. Only basic mathematical skills are required. The course uses

statistical techniques to investigate everyday situations and to give an understanding of statistical ideas - exploratory data analysis, basic statistical inference and relevance of statistics to decision making. It does this through investigating questions such as the following:

- Are we getting better off?
- Does class size affect pupils' performance?
- How much do pupils truant?
- Is my child developing normally?

M246 Elements of Statistics is a 30-point course at second level. Ideally students should have completed MST121 before registering for this course. It is an introductory statistics course that emphasises the practical nature of the subject. SSC (Student Statistical Calculator - non-standard software) is used to analyse data and to develop understanding of statistical concepts such as exploratory data analysis, estimating confidence intervals, hypothesis testing and regression. Also presented in Singapore.

M343, Applications of Probability is a 30-point course at level 3. For this course students need to be competent in algebra and calculus. The course introduces models to describe patterns of events that occur in time and space, eg earthquakes and occurrence of plant species, and situations that occur at discrete time points, such as gambler's ruin. Probability models are developed for events which may occur at any time, eg the spread of an epidemic. Other situations involving probability include genetics and fluctuation in the stock market. Also presented in Singapore.

M346 Linear Statistical Modelling is a 30-point course at level 3. Students need a good understanding of the basic ideas of statistics before starting this course, together with an ability to apply ideas and interpret results. It covers statistical modelling where a response variable depends on one or several exploratory variables and uses problems and data to explore questions, such as:

- How well does a patient respond to treatment, given age and severity of disease?
- How do different strains of wheat compare when grown in various conditions?

The statistical tool used is GENSTAT.
Also presented in Singapore.

What awards do these courses contribute to?

The awards listed below are those that students studying a statistics course may be aiming towards. Not all students are interested in an award. Students may study MU120, MST121 and not do any further statistics. These students may go on to take awards not listed here.

Diploma in Statistics (120 points)

M246, M343, M346 + 30 points from a list that includes MDST242

To obtain an honours degree, a student must study 360 points, including 120 points at level three.

BA/BSc, BA/BSc (Hons)

This degree can be tailored to a student's own requirements - a student can take courses from any area of the curriculum.

BA/BSc (Hons) Mathematical Sciences

BA/BSc (Hons) Computing and Mathematical Sciences
M246, M343 and M346 are optional courses for above two degrees.

BA/BSc (Hons) Economics and Mathematical Sciences

MST121, M246 and M346 are compulsory courses for this degree. M343 is optional.

BA/BSc (Hons) Health Studies

BSc (Hons) Natural Sciences

BA/BSc (Hons) Social Sciences

MDST242 is an optional course for above three degrees.

How do we support our students?

- Academic staff based in regional centres
- Advisors and Student Support Staff based in regional centres
- Local tutors
- Face to face tutorial sessions and/or dayschools (optional)
- Phone and email contact with tutors and regional centres
- Correspondence tuition - this is the major part of a tutor's job; quick turnaround and positive encouragement is vital.
- Telephone helpline - operated by academic staff from the Faculty
- Computer conferences - run by OUSA (Open University Student Association)

What is next for the Statistics department?

BM240 Quantitative Methods in Business is a 30-point course at second level, written specifically for the undergraduate business studies degree and is due for first presentation starting in February 2003. The aim of the course is to help students make sense of the large amounts of quantitative data produced by and for business. Students will learn to perform simple quantitative analyses using EXCEL and to interpret results of analyses performed by others. Students require basic

mathematical skills.

M248 *Analysing Data* is a 30-point course at second level, which will replace M246, using MINITAB to replace SSC. It is due for first presentation in February 2003.

U247, *Statistics in Society* is a 15-point service course to be presented in place of MDST242. First presentation is scheduled for February 2005. The course will use SPSS which is particularly suitable for students from the Social Science Faculty and from the School of Health and Social Welfare.

Java Applications for Teaching Statistics

Andrew Bertie
The Open University

a.bertie@open.ac.uk

M²⁴⁸ *Analysing Data* is a new introductory statistics course presented by the Open University from 2003. As the course title suggests, the emphasis is on learning statistics through analysing data rather than mathematics. MINITAB will be provided for data analysis, but it was decided to develop further software to support the teaching of more difficult statistical concepts. In the past the OU has developed such software in the C++ programming language. For this course, however, it was decided to develop the software in Java.

Advantages of Java

You may have already come across Java programs in the form of *applets* embedded in web pages. There are already hundreds of statistical applets, and a few complete courses in statistics that use applets, available on the web. But Java can be used to write stand-alone *applications* just as in C++. The M248 Java applications and Sun's Java Runtime Environment (JRE) will be delivered to students on the same CD-ROM as MINITAB and the course data files, saving students the costs of downloading the JRE and accessing the applets over the Web – problems highlighted by Hunt and Tyrrell (2000). So why choose Java and not C++? There are a number of general advantages in choosing Java as a development language that flow from it being designed for the Web.

The most important of these is its platform independence summed up by Sun's slogan Write Once, Run Anywhere (WORA). A program written in Java should require no modification to run on PCs (with their different versions of Windows), Macs, Unix or LINUX systems. This is an important advantage in distance teaching, where students can own a wide range of PCs with different specifications, ages, and operating systems.

There are some features of Java that make it particularly attractive to developers of statistical software. Its platform independence extends to the storage of numerical data types, ensuring statistical algorithm produces exactly the same numerical accuracy and behaviour on any hardware or system.

A striking feature of Java is the platform independence of its graphics. The modern PC provides exciting opportunities for interactive, dynamic and animated statistical graphics – particularly important in teaching statistics. But, even more than programming languages, computer graphics has suffered from lack of portability and standardisation. Traditional programming languages, such as Fortran and C, lacked standard graphical libraries that would allow the statistician to build plots from graphical primitives. The Java standard libraries provide superb platform independent support for graphics. It has all the graphical primitives you would expect, plus sophisticated handling of colour, stroke, text, geometry, transformations, animation and rendering. These libraries make it possible to construct any kind of statistical display, from basic exploratory plots to stunning professional-looking presentation graphics. They

also make it possible to develop animations for educational use. Unlike other languages, any graphical code you write using these libraries will run on any system capable of running Java - in keeping with Java's WORA philosophy.

The Java standard libraries also provide excellent platform independent support for building graphical user interfaces (GUIs). Using these libraries you can design a modern graphical interface to your statistical application involving multiple windows, dialogue boxes, menus, buttons, icons, fields etc. Remarkably, Java's WORA principle applies here too: the code you write will work on different platforms and operating systems without modification. Java allows you to choose a *look-and-feel* for your interface. You can choose a look-and-feel to suit Windows, Unix or Macintosh computer systems, or you can specify a standard Java look-and-feel that will look the same on all these platforms. No other programming language or system offers this power to design platform independent graphical user interfaces.

The World Wide Web is by definition an international network, so Java is designed to cope with differences between countries and languages. It provides support for *internationalization* - the process of enabling a program to run correctly in any country. With increasing globalization of distance education, this is an important advantage. A minimal level of internationalization of data display is possible with little effort. It is easy to write a Java program that automatically detects the geographical location it is running in, and formats numbers, percentages, currency, dates and times appropriately.

There is increasing pressure on universities to make their software accessible to disabled people. Java's *accessibility* package supports *assistive technologies* - such as audible text readers and screen magnification.

Java is virtually free. You can download everything you need to develop, compile and run Java programs for the cost of an Internet connection. The Java Development Kit (JDK), various utilities, tools, documentation and tutorials are all free from Sun's website, and Sun permit the free redistribution of the JRE. All you need to write Java code is a basic text editor, although specialized Java editors are also available for little or no cost.

What is Java?

Java is a structured language descended from Algol (through C), but the most important thing about it is that it's *object-orientated* (O-O). Essentially this means that the language supports designing a program that closely models the entities/objects that you are interested in, and the interactions between them. These entities can be abstract concepts, such as a statistical distribution. This approach is quite different to a traditional FORTRAN

program say, that is essentially a collection of functions/ algorithms. O-O ways of thinking are essential in the case of GUIs.

As an O-O language, Java is designed to overcome the problems of traditional languages such as FORTRAN and C, and encourage the development of correct, readable, easily modified and reusable programming code. Java's O-O approach can be applied to statistical concepts, methods and data in powerful and expressive ways that would be difficult in traditional languages. All this could also be said of C++, but Java is simpler, easier to learn and less prone to frustrating run-time errors than C++.

The Java language itself is quite small and quickly learnt - the syntax and basic data types are very similar to C. Beyond that is the class that represents an object, similar to C++. All data (variables and constants) and methods (functions) must belong to a class (unlike C++). Classes can inherit from other classes, so you can have an hierarchy of classes. The real skill in Java programming is choosing classes that produce an elegant and general solution to a problem, which can be easily reused and extended.

Sun provides a vast number of ready-made classes known as the Java API, which are fully documented on Sun's website:
<http://java.sun.com/products/jdk/1.4/docs/api>

There are also class libraries provided by other vendors, some free, some for a price. For example, a good free library for matrix computations is documented at <http://math.nist.gov/javanumerics/jama/doc>. The Sun API classes include the usual mathematical functions, random number generators (including one for the standard normal). There are classes for displaying and editing tables, and the 2D Graphics API provides graphical primitives. Beyond that there are no specifically statistical classes - which led me to develop my own class library called *Java Statistical Classes (JSC)*.

For a more detailed introduction to Java, see Murdoch (1999). Chambers(2000) describes Java as "...a more powerful engine for statistical software than, for example, C or C++."

The advantages of implementing statistical systems in Java are discussed at length by Lang (2000).

SUStats

SUStats (Software for Understanding Statistics) is a suite of 21 Java programs (applets/applications) designed to support the teaching of some statistical concepts in *M248*. They support the teaching of:

- The relative frequency "settling down" concept of

probability.

- Sampling distributions.
- Modelling variation and data distributions.
- Central limit theorem
- Confidence intervals
- Hypothesis testing, power
- Parameter estimation
- Least squares regression

Figure 1 shows one of the SUStats applets that demonstrate the relative frequency concept of probability. The *Rolling a die* applet simulates an experiment involving rolling a die in which two faces of the die are defined as “success” and shows how the relative frequency of successes settle-down to values close to 1/3. The die shown below the graph is rapidly spinning, but not randomly; it is cycling through its faces from one to six in sequence. The student clicks on the **Stop** button to stop it spinning and select the face value shown on the die, but the die is spinning so rapidly that their selection will in effect be random - like rolling the die. This follows a suggestion of Velleman and Moore (1996) that software should allow students to make random selections, since they may not regard computer-generated random numbers as genuinely random. As selections are made, the relative number of successes are plotted in the graph and the current proportion is displayed numerically.

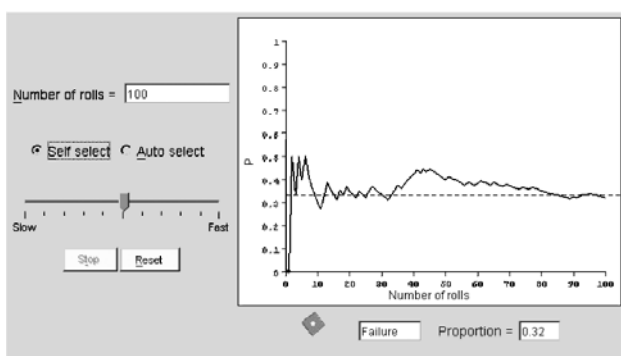


Figure 1: The *Rolling a die* applet

Figure 2 shows the *Hypothesis testing* applet that simulates fixed-level and significance testing of the means of six standard distributions and an arbitrary distribution called ‘Your distribution’ that the student defines by manipulating its pdf curve. There is a tabbed panel for each of the distributions. Figure 2 shows the applet with the **Your distribution** tab selected. Each panel has the same layout. The population is represented in the top-left of each panel – by a plot of the distribution and fields for its parameters, corresponding mean μ and standard deviation σ of the population. In the bottom-left area of each panel are controls for the simulated testing, consisting of fields for specifying the test mean, alternative hypothesis, significance level, sample size and number of samples, a check box for switching to single-step mode, and buttons for starting the simulation and resetting the applet.

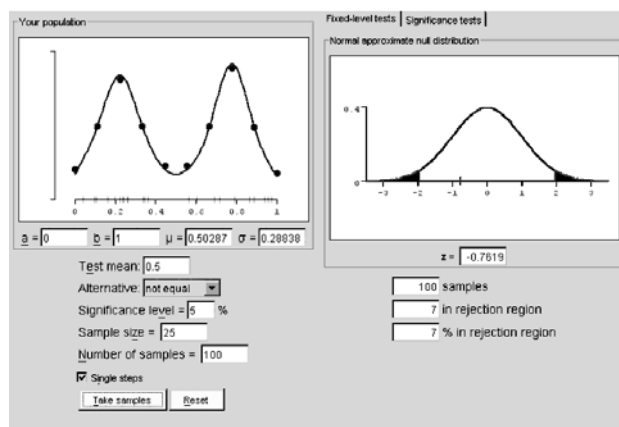


Figure 2: The *Hypothesis testing* applet

The right half of the applet, consisting of two tabbed panels labelled **Fixed-level tests** and **Significance tests**, displays the simulation results. When the **Take samples** button is clicked, the specified number of samples of the specified size are randomly generated from the population. For each sample, the test statistic and significance probability are calculated. In the **Fixed-level tests** panel the value of the test statistic is counted if it lies in the rejection region and summary results about the simulation appear below the plot of the null distribution plot. In the **Significance tests** panel the values of the significance probabilities are plotted in a histogram below the null distribution plot.

Figure 3 shows the *Principle of least squares* applet. Students can enter data points by clicking on the plot, typing or pasting data into the table, selecting a real data example or generating a random sample. The scatterplot in Figure 3 shows a fitted least squares line, confidence intervals of the mean and prediction intervals. The overlapping squares represent the squared residuals. The influence of individual data points can be explored by dragging their markers with the mouse or editing their values in the table – the fitted line, intervals and squares changing immediately. An option is provided to fit the least squares line through the origin. Clicking on the **Yours** button allows the student to fit a line by hand manipulating its position using three beads.

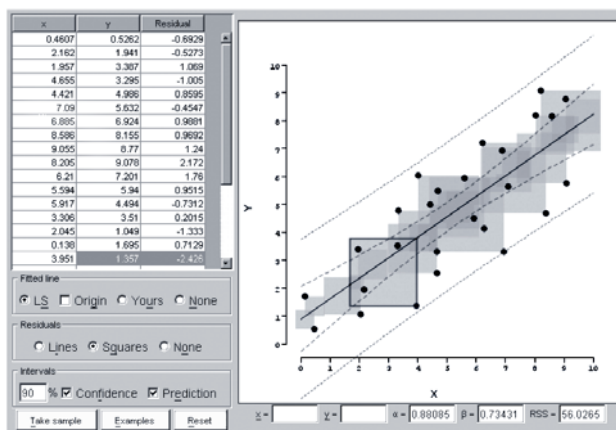


Figure 3: The *Principle of least squares* applet

Open University students will use the stand-alone application versions of these applets with printed computing activities, but basic instructions and suggested activities can be found in the web pages containing the applets at <http://www2.open.ac.uk/CES/projects/SUStats/SUStatsApplets.html>

The JSC library of reusable classes that supported these developments is being extended to cover a wide range of statistical methods and graphics and will be made available to the statistical computing community when completed.

Acknowledgements

The development of *SUStats* was a collaborative effort of the *M248* course team consisting of members of the Open University Statistics Department. Gillian Iossif (course chair), Paul Garthwaite, Paddy Farrington, Pia Larsen, and Kevin McConway provided ideas for applets and contributed to their design.

References

- [1] J M Chambers (2000) Users, programmers, and statistical software, *Journal of Computational and Graphical Statistics*, **9**, 404-422.
- [2] Hunt, N and Tyrrell, S (2000). Learning statistics on the web – DISCUSS. *Teaching Statistics*, **22**(3), 85-90.
- [3] Lang, D T (2000). The Omegahat environment: new possibilities for statistical computing. *Journal of Computational and Graphical Statistics*, **9**, 423-451.
- [4] Murdoch, D J (1999). What's Java? *Chance*, **12**, 44-47.
- [5] Velleman, P and Moore, D S (1996) Multimedia for teaching statistics: promises and pitfalls, *American Statistician*, **50**, 217-225.

Your advert could be here!

This newsletter reaches teachers and course developers at all levels and in all areas of mathematics, statistics and OR in all higher education institutions in the UK. It also has an overseas mailing list.

The total circulation at present exceeds 2300.

Workshops, conferences, meetings etc will be given free publicity.

Advertisement rates till Dec 2002 VAT exclusive

	colour	mono	Quarter page (landscape or portrait)	75
Full page back cover	475		one page insertion	350
Full page inside cover	450		extra pages (each)	100
Full page in newsletter	400	300		
Half page	260	150		

Reductions for regular or bulk advertisers. Copy date for the next issue is 10 October 2002.

Further details from the editor and at
<http://ltsn.mathstore.ac.uk/newsletter/advert/adinfo2002.htm>