

From Brunelli, Lina & Cicchitelli, Giuseppe (editors). **Proceedings of the First Scientific Meeting (of the IASE)**. Università di Perugia (Italy), 1994. Pages 255-263. Copyright holder: University of Perugia. Permission granted by Dipartimento di Scienze Statistiche to the IASE to make this book freely available on the Internet. This pdf file is from the IASE website at <http://www.stat.auckland.nz/~iase/publications/proc1993>. Copies of the complete Proceedings are available for 10 Euros from the ISI (International Statistical Institute). See <http://isi.cbs.nl/sale-iase.htm> for details.

255

A ROLE FOR COMPUTER INTENSIVE METHODS IN INTRODUCING STATISTICAL INFERENCE

Peter Jones, Kay Lipson and Brian Phillips
Swinburne University of Technology
John Street, Hawthorn 3122, Australia

I. Introduction

Introductory statistics courses have become increasingly prevalent in a wide variety of undergraduate and graduate programs over the past few years. This has resulted in the study of inferential statistics becoming the norm rather than the exception. Whilst statistics was once a course chosen by the more mathematically able student, many current students of statistics have little mathematical aptitude or expertise (Tanis, 1992 for example). As a result, many introductory statistics courses have moved away from including much of the statistical theory that underpins inference to become basically technique oriented "recipe book" courses. The inherent danger of producing students capable of performing various complex statistical tests without really knowing what they are doing is obvious, particularly if the course includes the use of a sophisticated statistical computer package.

To overcome the problem of having to teach inferential statistics to mathematically unsophisticated students it has been suggested by several statistics educators that the appropriate use of the computer in the classroom may help. In particular, computer based simulation is seen as a valuable tool, for instance:

Simulation can help convey both the hard idea that random variation has a pattern in the long run and specific facts such as the central limit theorem (Moore, 1992, p. 23).

and

Teachers can use simulation to illustrate ideas that are not otherwise accessible to beginning students. For example, there are many proofs of the central limit theorem, and many of them are short. But none are particularly intuitive or accessible to students who know nothing of moment generating functions. On the other hand, it is easy to demonstrate the central limit theorem, making the underlying definitions clearer and illustrating methods of simulation in the process (Thisted and Velleman, 1992, p. 49).

There is, however, a lack of substantive research into the effectiveness of simulation exercises and opinion is divided amongst statistics educators as to the advantages of such activities. As noted by Hawkins:

ICOTS 2 delegates were treated to "101 ways of prettying up the Central Limit Theorem on screen", but if the students are not helped to see the purpose of the CLT, and if the software does not take them beyond what is still, for them, an abstract representation, then the software fails (Hawkins, 1990, p. 28).

Although untested, it would seem that exposing students to large numbers of computer simulations does not necessarily clarify for them some of the more fundamental issues in statistics. It may well be that this apparent lack of effectiveness is due to our lack of knowledge of how to teach effectively with computer technology and this is not surprising, given our relative lack of experience and theory to guide us. One area in which some theoretical work has been done is in determining the features of the software that may in theory prove to be more effective, such as the use of multiple representations, and dynamic interactive displays (Hawkins, 1990; Kaput, 1992; Rubin *et al.*, 1988). It is not the purpose of this paper to delve into the reasons why the particular software being used may or may not be effective, although this is in itself an extremely important issue. Our purpose is to suggest a rationale as to why computer based simulations are not as helpful as we might suppose, and to propose an alternative path leading to statistical inference, which potentially avoids this problem.

2. Statistical inference and the sampling distribution

A critical step in developing the theory of statistical inference is the idea of a sampling distribution – the recognition that the estimates of a population parameter will vary and that this variation will conform to a predictable pattern. Yet, for all its importance, experience and research have shown that the idea is generally poorly understood (Moore, 1992; Rubin, 1990, for example). One reason for this might be the way in which the idea has been traditionally introduced in statistics courses, using a deductive approach based on probability theory (Johnson and Bhattacharyya, 1987; Mendenhall *et al.*, 1990 for example). Such explanations are usually expressed in highly mathematical language, which tends to make the argument largely inaccessible to all but the more mathematically able. But perhaps more importantly, it is a theoretical development that is difficult to relate to the physical process of drawing a sample from a population. Statistics educators have come to recognise

that there are deficiencies with a purely theory based explanation, and now often accompany or replace this with an empirical argument. The alternative interpretation uses the long run relative frequency approach to probability, where the sampling distribution is viewed as the result of taking repeated samples of a fixed size from a population and calculating the value of the sample statistic for each (Devore and Peck, 1986; Ott and Mendenhall, 1990, for example). The empirical approach has the advantages of being more readily related to the actual physical process of sampling and of requiring minimal use of formal mathematical language.

Because the computer has an obvious role in this empirical development of the idea of a sampling distribution, by enabling the repeated sampling to be carried out and summarising the results, a number of instructional sequences have been developed built around these capabilities (Norusis, 1988, for example). These approaches have been widely promoted and are now commonplace activities in introductory statistics courses. In our experience, however, they are less successful than statistics educators might have hoped. One reason for this may be the problem that students have when endeavouring to integrate their empirical experience of the sampling distribution with the theoretical model of the sampling distribution that is used in classical statistical inference. Eventually, in the classical approach, students must accept that the sampling distribution of the statistic of interest may be modelled by a known probability distribution, and that knowledge of this probability distribution is the basis for further estimation and hypothesis testing.

3. Using computer intensive methods

An alternative path to inference is provided by the developing area of Computer Intensive Methods. Basically, these methods are called "computer intensive" because they involve the computation of the statistic of interest for many data sets. According to Diaconis and Efron:

The payoff for such intensive computation is freedom from two limiting statistical factors that have dominated statistical theory since its beginnings: the assumption that the data conform to a bell-shaped curve and the need to focus on statistical measures whose theoretical properties can be analysed mathematically (Diaconis and Efron, 1983, p. 96).

For the practising statistician, there are often definite advantages of using a computer intensive method rather than a classical method when restrictive assumptions about the nature of the data do not hold, or traditionally used test statistics are not appropriate. From the pedagogic

point of view, there may also be advantages to the student of statistics.

Essentially, applying a computer intensive method means that the entire inference problem is dealt with from an empirical perspective. Students are not required to make the conceptual connection between the empirical and theoretical sampling distributions, but actually use an empirical sampling distribution as the basis for the inference. Consider the following example, taken from a typical introductory statistics text by Weiss and Hassett:

A highway official wants to compare two brands of paint used for striping roads. Ten stripes of each paint are run across the highway. The number of months that each stripe lasts is given below.

Brand A		Brand B	
35.6	36.1	37.2	36.4
37.0	35.8	39.2	37.5
34.9	34.9	37.2	40.5
36.0	38.8	38.8	38.2
36.6	36.5	37.7	36.6

Based on the sample data, does there appear to be a *difference* in mean lasting time between the two paints? Use $\alpha = 0.05$ (Weiss and Hassett, 1987, p. 427).

Classical method of solution

From the question, the students must recognise that they are required to perform a hypothesis test for the equality of means with a two-sided alternative thus:

$$H_0: \mu_1 = \mu_2$$

$$H_1: \mu_1 \neq \mu_2$$

If we assume that the two samples are taken independently from two normally distributed populations with means μ_1 and μ_2 respectively, and the respective standard deviations are unknown, then a hypothesis test for comparing the means can be carried out by consideration of the test statistic

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_1^2 / n_1 + s_2^2 / n_2}}$$

which has a distribution which can be approximated by a t -distribution with degrees of freedom given by

$$df = \frac{(s_1^2 / n_1 + s_2^2 / n_2)^2}{\frac{(s_1^2 / n_1)^2}{n_1 - 1} + \frac{(s_2^2 / n_2)^2}{n_2 - 1}}$$

where n_1 and n_2 are the sizes of the samples, \bar{x}_1 and \bar{x}_2 are the means of the samples, and s_1^2 and s_2^2 are the sample variances. Substituting in these formulae gives a value of the test statistic that is then compared with a value from the tables¹.

Rather than using a calculator to perform these calculations, such procedures are generally carried out using a standard statistical package such as Minitab. This is most likely how the student would be expected to proceed:

```
MTB > TwoSample 95.0 'BrandA' 'BrandB';
SUBC> Alternative 0.

TWO SAMPLE T FOR BrandA VS BrandB
      N   MEAN   STDEV   SE MEAN
BrandA 10   36.22   1.14    0.36
BrandB 10   37.98   1.33    0.42
95 PCT CI FOR MU BrandA - MU BrandB: (-2.93, -0.59)
TTEST MU BrandA = MU BrandB (VS NE): T = -3.18 P = 0.0055 DF = 17
```

Of course the student must now relate the computer output to the original question, which would be, on the basis of the P -value, to reject the null hypothesis and to conclude that there is a difference in mean lasting time between the two brands of paint.

Computer intensive method of solution

The observed mean difference in the time that the paint lasts for the two brands is 1.76 months. This difference may be due to a real difference

¹ The formulae given are those that are used when we cannot assume the variances of the two populations are equal. The assumption of equal variances was often made in the past because the distribution of the test statistic had otherwise only an approximate (t -distribution). The common use of computers has meant that this is no longer an issue. Whether or not the assumption of equal variances is made the classical methods of solution are very similar.

In this example, an empirical distribution of 500 observations was generated, and displayed as a stem and leaf plot. From the plot it may be seen that there was only one value of the mean difference as far or further from zero than 1.76, giving our estimated P -value as 1 in 500 or 0.002. Thus we see that such a difference in means is very unlikely to have been caused by chance alone, and conclude that there is a real difference in the mean lasting times of the two brands of paint.

Comparison of classical inference and computer intensive methods

The essential components of a hypothesis test have been well described by Noreen:

Three ingredients are usually required for a hypothesis test: a hypothesis, a test statistic, and some means of generating the probability distribution of the test statistic under the assumption that the hypothesis is true (Noreen, 1989, p. 2).

It is in this process of defining a test statistic and determining its distribution that the essential difference between the classical and computer intensive methods lies. The test statistic and its distribution are often complex, and their origins obscure to students of classical statistical methodologies. The essential purpose of the process, to determine the likelihood of such a set of outcomes under certain assumptions, becomes easily lost. In classical hypothesis testing, the fundamental logic of the hypothesis test becomes entangled with the theory underlying the sampling distribution of the test statistic. Using a computer intensive method, an intuitively obvious test statistic is selected for which an empirical distribution is generated. The students remain linked to the overriding purpose of the exercise – the inference. Using the technology in this way amounts to forming a partnership between the student and the computer in which the computer takes on the lower level tasks of performing the numerous of calculations, whilst the student undertakes the higher order tasks of actually applying the logic of inference to the particular situation. In fact, the student must keep in mind the overall purpose of the exercise in order to carry it out successfully. Whilst the student of classical inference needs to recognise certain situations, and to make the appropriate assumptions, the students of computer intensive methods have been equipped with valuable problem solving skills that may have application in circumstances far removed from the one at hand.

The question arises as to the long term effect of the technology on the students' understanding of statistical inference when computer intensive methods are used. It is reasonable to contend that the act of carrying out

hypothesis testing using computer intensive methods may actually enhance students' understanding of inference, a long term effect of the technology. It may even be that this deeper understanding could then be applied to classical hypothesis testing when students are exposed to this procedure, although one might ask why, in this computer age, this would be necessary.

Thus, it seems possible that using computer intensive method may enable beginning students of statistics, especially those without strong mathematical backgrounds, to appreciate the important principles of statistical inference. Whilst at this stage little research exists to support this theory, what does exist seems encouraging (Simon *et al.*, 1976, for example). A challenge for statistical educators in the 1990's is to determine if this is indeed the direction we should be following.

Bibliography

- Devore J. and Peck R. (1986), *Statistics*, West Publishing Company, St. Paul.
- Diaconis P. and Efron B. (1983), Computer-Intensive Methods in Statistics, *Scientific American* (May), pp. 96-108.
- Hawkins A. (1990), Success and Failure in Statistical Education - A UK Perspective, in D. Vere-Jones (ed.), *Proceedings of the Third International Conference on Teaching Statistics*, International Statistical Institute, Voorburg, Vol. 1, pp. 24-32.
- Johnson R. and Bhattacharyya G. (1987), *Statistics Principles and Methods*, John Wiley & Sons, New York.
- Kaput, J.J. (1992), Technology and Mathematics Education, in D. A. Grouws (eds.), *Handbook of Research on Mathematics Teaching and Learning*, MacMillan, pp. 515-555.
- Mendenhall W., Wackerly D.D. and Scheaffer R.L. (1990), *Mathematical Statistics with Applications*, 4th edition, PWSP-Kent Publishing Company, Boston.
- Moore D.S. (1992), Teaching Statistics as a Respectable Subject, in F. Gordon and S. Gordon (eds.), *Statistics for the Twenty-First Century*, MAA Notes No. 26, Mathematical Association of America, pp. 14-25.
- Noreen E. W. (1989), *Computer Intensive Methods for Testing Hypotheses: An Introduction*, John Wiley & Sons, New York.
- Norusis M.J. (1988), *SPSS/PC+ Studentware*, SPSS Inc., Chicago.
- Ott L. and Mendenhall W. (1990), *Understanding Statistics*, 5th edition, PWSP-KENT Publishing Company, Boston.

- Rubin A., Bruce B., Roseberry A. and Du Mouchel W. (1988), Getting an Early Start: Using Interactive Graphics to Teach Statistical Concepts in High School, in *Proceedings of the Section on Statistical Education*, American Statistical Association, pp. 6-15.
- Rubin A., Bruce B. and Tenney Y. (1990), Learning about sampling: Trouble at the core of statistics, in D. Vere-Jones (ed.), *Proceedings of the Third International Conference on Teaching Statistics*, International Statistical Institute, Voorburg, Vol. 2, pp. 314-319.
- Simon J.L., Atkinson D.T. and Shevokas C. (1976), Probability and Statistics: Experimental Results of a Radically Different Teaching Method, *Mathematical Education*, November, pp. 733-739.
- Stirling D. (1991), *Models 'n' Data*, Intellimation, Santa Barbara.
- Tanis E.A. (1992), Computer Simulations to Motivate Understanding, in F. Gordon and S. Gordon (eds.), *Statistics for the Twenty-First Century*, MAA Notes No. 26, Mathematical Association of America, pp. 217-225.
- Thisted R.A. and Velleman P.F. (1992), Computers and Modern Statistics, in D.C. Hoaglin and D.S. Moore (eds.), *Perspectives on Contemporary Statistics*, Mathematical Association of America, pp. 41-53.
- Weiss N. and Hassett M. (1987), *Introductory Statistics*, 2nd edition, Addison-Wesley, Reading, Massachusetts.