

Inferential statistics and the graphics calculator*

Barry Kissane
The Institute of Education
Murdoch University, Australia
kissane@murdoch.edu.au

ABSTRACT: Recent teaching and learning of elementary statistics have been influenced by the use of statistical packages on microcomputers, which have permitted data storage and flexible data analysis. Scientific calculators have routinely provided statistical capabilities for some time, but generally they have been too limited to be used as alternatives to computer statistics packages, at least at the undergraduate level. In this paper, graphics calculators are regarded as devices that combine some of the advantages of each of these two kinds of technology for early work in statistics. The first generation of graphics calculators, while providing significant data analysis opportunities, were insufficient for the needs of students of early undergraduate statistics, since the important aspects of inferential statistics were not accessible. Later models include capabilities dealing with hypothesis testing, the construction of confidence intervals and the tabulation of probability distributions. It is suggested that these meet most of the statistical needs of introductory courses. The small size and cost of graphics calculators increase the prospect that individual students will have ready access to them at all times, with significant curriculum implications identified. Programming capabilities of graphics calculators permit student explorations dealing with important concepts in statistical inference to be conducted, some examples of which are described in the paper.

Introduction

In the space of one generation, the practice of statistics has changed considerably as a consequence of developments in technology. The computer has had three significant effects on statistical practice. Firstly, data storage and statistical computations are now routinely handled by computers rather than more tedious and error-prone human equivalents. Secondly, the processing capabilities of the computer have given rise to new data analysis opportunities, most notably exploratory data analysis, first popularised by John Tukey in the 1970's. The third change wrought by computers, which allow for visual displays of data to be readily

made, is to provide graphical as well as numerical representations of data. Whereas computers with suitable software for statistical purposes were at first only available on mainframes, located remotely from users, it is now quite common for statistical analyses to be conducted on desktop microcomputers equipped with powerful statistical software.

Changes of such magnitude can reasonably be expected to exert considerable influence on the undergraduate statistics curriculum. It is clearly important that students learn statistics in realistic ways, so that they become adept at making use of suitable technology, such as computer software packages. Employers of graduates these days naturally expect that students will have acquired some competence with suitable combinations of hardware and software by the time they enter the job market. To bring this goal about, students need to be taught to use particular combinations of software and hardware efficiently and some space for this new need must be created in the curriculum.

As well as curriculum space, physical equipment must also be provided in order for students to have reasonable access to appropriate computers and software; this is often much more difficult for institutions to provide adequately, since much elementary statistics teaching involves large undergraduate classes. One solution to this problem is to insist that students purchase their own computer and suitable statistical software, but there are many students for whom this is not yet a reasonable expectation, particularly students in developing countries or from less affluent families in developed countries. Questions of access to technology are clearly of critical importance since, without reasonable access *for all students*, it is unreasonable and unrealistic to demand student competence. In many circumstances, access to technology is much easier to provide with graphics calculators than with microcomputers, as argued elsewhere (e.g., Kissane, Bradley & Kemp 1994; Kissane, 1995). Even in relatively affluent countries such as Australia and the United States, student access to computers is often rather meagre; the situation is clearly much worse in most parts of most developing countries and in many parts of Asia. The case for graphics calculators as an alternative to computers seems particularly strong in economic situations where resources for teaching and learning mathematics are limited.

In addition, it seems unwise to consider appropriate hardware and software for statistics without acknowledging the need to consider technological needs of other aspects of the mathematics curriculum, as Biehler (1991) noted:

A major problem with mathematical tools is that each one requires specific knowledge, for both students and teachers. As time in school is very limited, more complicated tools for probability education cannot be chosen without taking into account which tools should be used in the entire mathematics curriculum. This is why the usefulness outside probability is important. (p.181)

With reasonable access to technology, we might expect that teaching elementary inferential statistics might focus on statistical concepts and thinking more than was previously possible, to take advantage of the potential of technology to take care of most of the relevant routine computation. As well as appropriate content concerned with using the software efficiently, the modern statistics curriculum must also develop the underlying statistical ideas so that students understand them well enough to be able to critically interpret the results of data analyses.

The evolution of the statistical calculator

Early models of scientific calculators included some statistical capabilities dealing with univariate data or, for more expensive models, bivariate data. These calculators typically contained commands for calculating means, standard deviations (both biased and unbiased estimates) as well as summary data for using raw score formulas for computing statistics of these kinds. In the case of calculators equipped for handling bivariate data, least squares linear regression coefficients and Pearson product-moment correlation coefficients were also calculated automatically. Some calculators included commands for accessing cumulative normal distribution tables, obviating the need for students to refer to tables of values.

The major advantage of scientific calculators was associated with arithmetic: students permitted to use such a calculator did not have to engage in a great deal of tedious arithmetic computation in order to calculate statistics of interest. The computational easing also meant that data could be analysed directly, rather than having to be rounded to produce more convenient numbers. Actual measurements could be used, rather than integer approximations for computational convenience, and it was no longer necessary to group data into convenient intervals for computational ease. In theory, although not necessarily in practice, students could analyse their data directly, given access to such a calculator. For the purpose of conducting inferential tests, the calculator provided most of the necessary raw materials for computing statistics (such as a t-statistic), and required of users only that they manipulate the calculator results appropriately.

The main limitation of these calculators concerned data storage. Because of memory limitations, scientific calculators do not store individual data points as they are entered. Rather, appropriate raw score summary statistics are obtained as data are entered, and these are used to compute statistics on request. A consequence of this is that students do not know for sure what data have been entered, and most key punching errors are not detectable. Similarly, although mechanisms for deleting data are available, it is difficult to undertake data editing and not possible to undertake data transformations. As well as storage limitations, scientific calculators did not allow flexibility of data analysis; users were restricted to a small range of numerical descriptors of a data set.

These limitations were all overcome with the development of graphics calculators, mainly because data were stored in the calculators. Early models, in the mid and late 80's, provided for the storage of data, thus allowing easy checking, editing and transformations of data to be carried out. The second generation of graphics calculators, in the early 1990's, incorporated space for several variables (usually six) and organised them into data lists for ease of access and handling. As the adjective 'graphics' suggests, an additional advantage of these was their capacity to create graphical displays of data as well as numerical summaries. To an extent, this generation of graphics calculators allowed students to take advantage of technology in the three ways noted above: automatic storage and computation, exploratory data analysis and visual displays. Compared with a computer statistics package, graphics calculator limitations were considerable: data sets cannot be too large, screens are small and (generally) monochromatic and a limited range of analyses (only concerned with descriptive statistics) were available. Despite these limitations, graphics calculators seemed promising for student work in statistics, as their small size and relative cost made it easier to provide technological support for large numbers of students than is the case for computers.

Recently, a third generation of graphics calculators has appeared, offering a little more support for students of introductory statistics. The substantial differences from the second generation for statistics include the provision of hypothesis testing capabilities, automatic computation of confidence intervals and the routine availability of the main probability distributions likely to be encountered by students in the early years of their study of statistics. Three of the four manufacturers of graphics calculators aimed at the upper secondary and lower undergraduate levels (Casio, Sharp and Texas Instruments) have produced third generation calculators with these capabilities, with, naturally enough, small differences in styles of operation and capabilities. For each calculator, data stored (by key entry, electronic transfer or calculator simulation) can be subjected to inferential statistical analysis in flexible and quite powerful ways. In most cases, an intuitive local syntax is used to effect suitable analyses.

Examples of inferential capabilities

To illustrate some of these capabilities, the screens below are taken from a Casio cfx-9850G Plus graphics calculator. Figure 1 shows the calculator instructions needed to perform a one-way analysis of variance when there are three levels of a particular treatment involved. The raw data have been entered into Lists 1, 2 and 3 as indicated.



Figure 1: Setting up a one-way analysis of variance on a Casio cfx-9850G Plus

The resulting calculator output is shown in Figure 2. In this case, the calculator screen size does not permit all of the output to be displayed on a single screen, and it is necessary for the user to scroll down to obtain the last three rows (as shown on the second screen of Figure 2). Other information of relevance to such an analysis, such as cell means and standard deviations, is available in the calculator in the usual way, but is not reported at the same time as the ANOVA summary table. This information can be used to conduct ad hoc post comparisons tests or to interpret better the results of the ANOVA.

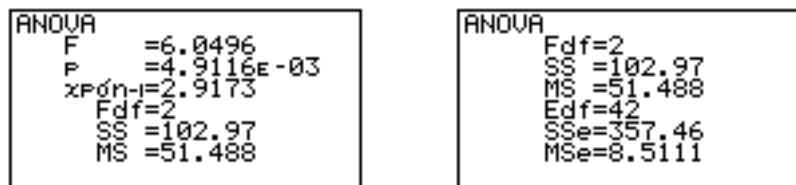


Figure 2: ANOVA summary table

Some kinds of hypothesis testing (such as t-tests and z-tests for means) are aided by graphical displays of results as well as numerical tests of significance. Another example is Chi-squared hypothesis testing associated with contingency tables, such as tests of independence and tests of fit to a model. Unlike tests of means, raw data for contingency tables are stored in data matrices, using the matrix area of

calculators. Figure 3 shows the output for a particular analysis in graphical rather than numerical form, to illustrate this kind of possibility. As for other analyses, further information is available from the calculator than what is routinely provided by the significance test. In this case, a matrix of expected values is automatically constructed and can be recalled by the user if desired.

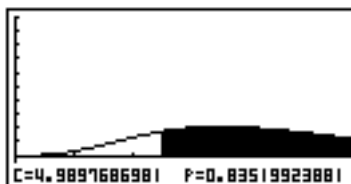


Figure 3: Graphical display of a Chi-squared test on a Casio cfx-9850G Plus

Figure 4 shows an example of the construction of a confidence interval, once data have been entered into the calculator. Third generation calculators typically include both one-sample and two-sample confidence intervals, using either z or t distributions. The choices of which confidence interval to construct are, of course, made by the user.

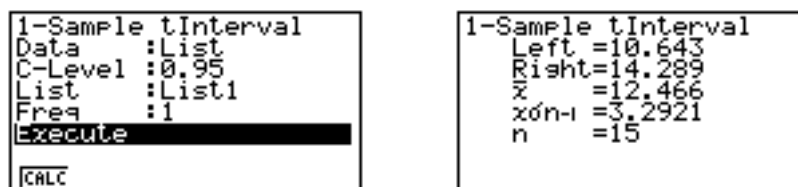


Figure 4: Calculator screens showing the construction of a confidence interval about a single mean on a Casio cfx-9850G Plus

Figure 5 shows an example of the third embellishment of graphics calculators, involving values for the main probability distributions of relevance to introductory study of inferential statistics. Although some graphics calculators provided normal distribution values (as did some scientific calculators in fact), third generation models routinely provide information about the distributions typically encountered by students when first studying inferential statistics: normal, Student's- t , binomial, Poisson and Chi-squared. The screen on the left of Figure 5 displays the details of the probability distribution sought (in this case, the cumulative binomial, with $p = 0.25$ and $n = 20$), while the screen on the right shows the actual value, in this case $\text{Prob}(x \leq 5) = 0.61717$.

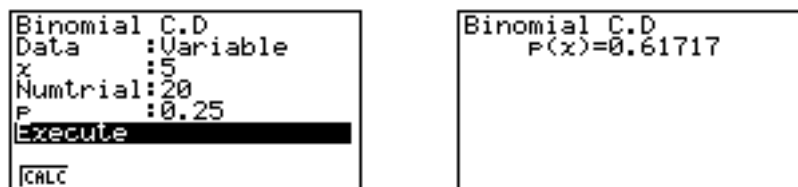


Figure 5: Cumulative binomial distribution values from a Casio cfx-9850G Plus

For all three kinds of inferential analysis (hypothesis testing, confidence interval construction and probability determination), communication between the user and the calculator is fairly intuitive, so that students familiar with the operations of their

calculator for basic descriptive statistics will not need a lot of extra time to master the commands needed.

Learning with a calculator

One of the reasons for delaying the study of inferential statistics until the end of secondary school or the early undergraduate years is that it is notoriously difficult for students to learn, on account of the probabilistic ideas involved. This point is emphasised in a recent major treatise on learning probability:

"... probability is a difficult subject to learn and teach. ... In probability, paradoxes or counterintuitive ideas occur at the very heart of the subject, in the definition, and subsequently in relatively simple applications. This is borne out by the vast psychological research on people's misconceptions as well as in the difficulty experienced by pupils in applying probabilistic notions in problems. People do not seem to have probabilistic intuition in the same way they have geometric or visual intuition." (Kapadia & Borovcnik, 1991, p.2)

The first and most obvious implication of the use of a modern graphics calculator for statistical inference is that less time may be needed for many computational aspects, such as constructing confidence intervals and conducting hypothesis tests. While clearly a graphics calculator will be quicker and more convenient than performing relevant computations by hand, it is likely for most students that it will also be more affordable, accessible and available than a computer statistics package. The convenience of the calculator, assuming that it is either owned by or on long-term loan to a student, far exceeds that of a computer for very many students (especially those without personal access to appropriate hardware and software at home). This is particularly so at the introductory levels, often the first undergraduate year, where class enrolments frequently exceed the number of available computers on a campus, hours of use are somewhat restricted, security issues make computer access inconvenient and, even when the computer is readily available, statistical software is often quite sophisticated and a little daunting for students at the introductory levels. As noted above, it is frequently necessary for students to develop eventually some expertise with computer software for statistics, however inconvenient it may be to access computers on which to use it; this does not necessarily mean however that it is a wise educational strategy to *start* with such software.

Mathematics teachers (at all levels and in all countries) seem to agree that there is insufficient time available for teaching and learning. So, activities that seem likely to demand less student time than the present activities are probably advantageous. The time saved by avoiding hand computations, waiting for access to a computer or navigating the menu structures or data requirements of a large statistics package can be used to engage in learning activities to develop better intuitions about inferential statistics. Concepts such as those of sampling distribution, hypothesis test, confidence interval, statistic, error rates and statistical significance are critically important to understanding statistical inference and very difficult to acquire, as suggested in the quote above. Providing more time to develop the concepts, activities to illuminate them and opportunities to explore them will not resolve the difficulties for students, but it is not unreasonable to expect that it may lessen them. The following sections briefly describe some of the learning possibilities available with a modern graphics calculator.

Sampling distributions

The notion of a sampling distribution is critical to classical statistical inference. One way to help students' develop intuitions about this concept is to provide them with the means of exploring what happens when successive random samples are taken from a population. A short calculator program is needed to do this, using the pseudo-random number generator built into all graphics calculators. An example of such a program for the Casio cfx-9850G is shown in Figure 6; the program assumes that a (finite) population of M values is stored in List 1 and has the effect of selecting a simple random sample of size N from this population, storing it in List 6 and displaying the mean of the sample. (The program includes only eight lines of code, but two calculator screens are necessary to display them all.)

```

=====SRANSAMP=====
"SAMPLE SIZE"?→N#
N→Dim List 6#
Dim List 1→M#
For 1→K To N#
List 1[[Int (MRan#+1)]]
→List 6[K]#
Next#
Mean(List 6)#
  
```

Figure 6: A program to select simple random samples

Students can either be given program steps and asked to enter them into their calculator, or electronic transfer of programs between calculators or between a computer and calculators is possible. When executed, the program selects a sample of the desired size, stores it and then displays the mean of the sample. For example, output from three successive runs of the program is shown in Figure 7.

SAMPLE SIZE?	11.80
10	
SAMPLE SIZE?	11.10
10	
	12.00

Figure 7 : Means of three samples from a population obtained from a program

There are at least two ways of using such a program to help develop a notion of a sampling distribution. One possibility is for each of the students in a class to select a sample, find its mean and report the result. The set of results can then be analysed as an empirical sampling distribution. It is not a very large jump from this idea to that of the general notion of a sampling distribution, and even a class of thirty students will readily generate a distribution in this way that shows some of the important features. A second way of using such a program is for students to generate sets of samples by themselves and construct empirical sampling distributions. In either case, explorations of the effects of varying the sample size are relatively easy to carry out. Ideas of this kind are not new. (They were suggested in Kissane (1981), for example.) What is new is the prospect that they might be accessible easily to all students in a class.

Programs of this kind can also be written to automatically select a number of samples of a certain size and then to generate an empirical distribution of the resulting means, effectively automating the process of repetitive sampling. A similar sort of approach can be taken to explore resampling methods for inferential statistics, such as those described by Borovcnik & Peard (1996, p.260).

Confidence intervals

The idea of a confidence interval is not an easy one to grasp and the construction of confidence intervals can be a tedious operation. With a modern graphics calculator, the construction of a confidence interval is easy, freeing up some space to think about the ideas involved. Thus, students can use a random sample chosen as described above to readily construct a confidence interval for the mean of a population, and compare their interval with the actual population mean. As with sampling distributions, each member of the class can do this empirically and the results of all compared, or the same student can repeat the operation several times and record the outcome each time. Roughly speaking (since the results of course only hold in general rather than in particular cases), a class will find that about 90% of their 90% confidence intervals will actually contain a population mean, while almost all of their 99% confidence intervals will do so.

Another way in which a graphics calculator may be helpful in developing students' concepts of confidence intervals arises from its capacity to quickly explore the relationships between confidence intervals and levels of confidence or between confidence intervals and hypothesis testing. To illustrate some of these possibilities, Figure 8 shows a pair of confidence intervals constructed about the difference between the means of two independent samples, using a t -distribution. Students can use the graphics calculator commands to generate quickly a succession of confidence intervals to different levels of confidence, in order to see that the width of the intervals increases with increasing confidence. Discussions about why that should be so will help students understand the ideas involved.



Figure 8: 95% and 99% confidence intervals for the difference of two means

Similarly, comparisons of confidence intervals with the associated t -test can lead to fruitful discussions exploring the relationship between the construction of a confidence interval and the notion of statistical significance, as realised with a hypothesis test. In this case, a graphical representation of the relevant two-tailed test is shown in Figure 9.



Figure 9: Graphical representation of t -test of difference between two means

Students can use the calculator to find the smallest probability level that would lead to a confidence interval that included zero; such a question will demand that they explore the connections between the relevant ideas.

Strictly speaking, of course, these kinds of activities can only *suggest* what the

relationships between important concepts are; the theoretical results concern infinite populations and rest in part on assumptions of normality of population distributions. More sophisticated students might explore informally the effects of violating such assumptions, by comparing results with populations (approximately) normally distributed with those from populations which are clearly not normally distributed. For such explorations, it is most convenient to work with simulated data, using the calculator's random number generation capabilities. While some graphics calculators provide a special command for simulating random normal deviates, useful for these sorts of explorations, for others a transformation of uniform random numbers is needed. One possibility is $z = \cos(2\pi a) \sqrt{-2 \ln b}$, which generates a random variable z approximately distributed as $N(0,1)$ from a pair of independent random variables a and b , each of which is uniformly distributed on the interval $(0,1)$.

Hypothesis testing

There are several ways in which informal calculator activities may help to develop student intuitions about hypothesis testing. As noted earlier, some of these may involve exploring the links between hypothesis testing and the construction of confidence intervals. In a similar way, students can use a random sampling program like the one above to select a sample and to then conduct an hypothesis test. If each of the students in a class goes through the same exercise, it is possible to see that the type 1 error rate describes how often a null hypothesis is rejected when it is true. As before results are only approximations, of course, but will help students to appreciate the meanings of the relevant concepts.

Another kind of possibility is to explore the relationships between significance tests and data, especially for small data sets, to develop an intuitive feel for the fragility of results from small samples. For example, for the contingency table shown in Figure 10, the null hypothesis of no association will be rejected at the 5% level of significance.

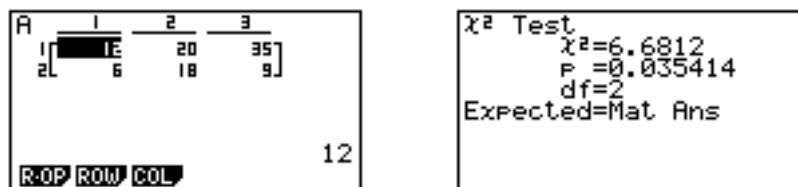


Figure 10: A Chi-squared test of association on a 2 x 3 table

To help interpret the meaning of such a significance test, students can readily change the raw data and conduct further tests of the same kind. For example, Figure 11 shows that the null hypothesis of no association would not have been rejected at the 5% level, had the frequencies of 18 and 9 instead been 17 and 10.

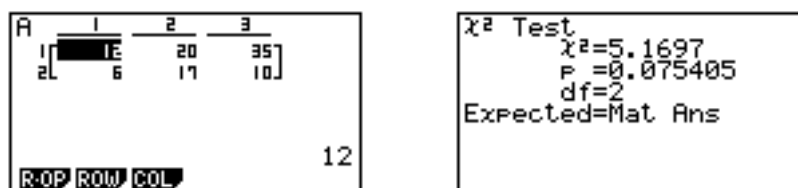


Figure 11: Examining the effect on the test of a slight change in the original data

Similar kinds of explorations can be conducted for other hypothesis tests such as t -tests and z -tests. Such activities may help students to appreciate some of the limitations of classical hypothesis testing, as well as providing some powerful encouragement to students for making sure that data entries are checked carefully.

Conclusion

Graphics calculators appear to be powerful and flexible tools for introductory teaching and learning of inferential statistics, and may be more accessible to students than are computers. The limited requirements of first courses involving inferential statistics are probably met by most modern graphics calculators, which are also useful for other mathematical purposes as well. This paper has suggested a number of ways in which the calculators might be used to enhance students' intuitions about important inferential ideas such as sampling distributions, confidence intervals and hypothesis testing.

References

- Biehler, R. 1991, Computers in probability education, In R. Kapadia & M. Borovcnik (Eds), *Chance Encounters: Probability in Education*, Dordrecht, Kluwer Academic, pp 169-211.
- Borovcnik, M. & Peard, R. 1996, Probability, In Bishop, A.J., Clements, K., Keitel, C., Kilpatrick, J. & Laborde, C. (Eds.), *International Handbook of Mathematics Education*, Dordrecht, Kluwer Academic, pp 239-287.
- Kapadia, R. & Borovcnik, M. 1991, The educational perspective, In R. Kapadia & M. Borovcnik (Eds), *Chance Encounters: Probability in Education*, Dordrecht, Kluwer Academic, pp 1-26.
- Kissane, B. 1981, Activities in inferential statistics, *National Council of Teachers of Mathematics Yearbook*, Reston, VA.: NCTM, 182-193.
- Kissane, B. 1995, The importance of being accessible: The graphics calculator in mathematics education, *Proceedings of the First Asian Technology Conference in Mathematics*, Association of Mathematics Educators: Singapore, 161-170.
- Kissane, B. 1997, Chance and data: New opportunities provided by the graphics calculator, in W.C. Yang & Y.A. Hasan (Eds), *Computer Technology in Mathematical Research and Teaching*, Penang, Malaysia, School of Mathematical Sciences, pp 80-88.
- Kissane, B., Bradley, J., & Kemp, M. 1994, Graphics calculators, equity and assessment, *Australian Senior Mathematics Journal*, 8(2), 31-43.

* This paper has been published as follows:

- Kissane, Barry 1998, 'Inferential statistics and the graphics calculator', in Wei-Chi Yang, Kiyoshi Shirayanagi, Sung-Chi Chu & Gary Fitz-Gerald (eds), *Proceedings of the Third Asian Technology Conference in Mathematics, Tsukuba, Japan*, Singapore, Springer (pp 111-121).