Inferential statistics: Using a graphics calculator to enhance understanding of how changes in sample values affect the significance of outcomes

Marian Kemp
M.Kemp@murdoch.edu.au
Teaching and Learning Centre
Murdoch University
Australia

Jen Bradley
J.Bradley@murdoch.edu.au
School of Chemical & Mathematical Sciences
Murdoch University
Australia

Abstract*: Students often have difficulty with understanding statistical inference*. *In this paper we demonstrate how technology can be used to aid students in developing an understanding of some concepts in hypothesis testing. The graphics calculator provides an ideal tool for students to investigate how changes in sample values affect the significance of the outcomes. Two scenarios will be presented: the concept of the proportion for a single population, specifically addressing the idea of a majority and secondly the concept of the mean difference from two related populations, specifically addressing the idea of a significant difference. In these scenarios the effects of both changes in the sample values (i.e., sample proportion or sample mean difference) and the sample sizes will be explored. The changes that are readily observable through the use of the calculator will be related back to underlying concepts of variation in the standard errors (i.e., the standard deviations of the distribution of the sample values). Whilst the explorations could also be carried out using a spreadsheet such as Excel, graphics calculators offer more accessible learning experiences which involve students in active participation and critical thinking. These activities would be suitable at either upper secondary school or first year undergraduate level, whenever students are first introduced to inferential statistics.*

1. Introduction

Chance and data courses at the secondary school level usually focus on descriptive statistics: calculating ranges, means, modes, medians and representing data in graphical form. In Western Australia at the first year undergraduate level, introductory statistics courses start with descriptive statistics and progress to inferential statistics. Studies of probability distributions lead to calculations for confidence intervals and hypothesis tests for one and two samples, including z-tests for proportions and t-tests for means. Given that "many published research reports, as well as popular media accounts, utilize ideas of statistical confidence and significance" [1], it is important that students have a good understanding of what statistical significance means. Indeed an understanding of "sampling and of making inferences about populations based on samples, are fundamental to prediction and decision making in many aspects of life" [2]. However, when students are introduced to inferential statistics their teachers often find that "although many students can complete calculations they are often unable to understand underlying processes or properly interpret the results of the calculations" [1]. The investigations detailed in this paper are designed to enable students to grasp those underlying processes so they can properly interpret the results of their calculations.

2. Background

The literature reveals a number of possible reasons why inferential statistics is commonly difficult for students to understand as well as some procedures that teachers have put in place to try to improve the students' understanding of this branch of statistics. Many of these include the use of technology in a variety of ways. This paper highlights some of these issues and actions. The

authors then offer ways of investigating and exploring hypothesis testing which students can undertake using a graphics calculator.

It is important to bear in mind two important points when discussing the use of technology in developing understanding: technology is a tool for learning and must be used thoughtfully by teachers [3]; and students need to develop their own skills in using and choosing the technology that is most suitable for the purpose [4].

2.1 Using technology in developing statistical understanding

There has been widespread development of computer technology over the last 30 years. In 1990 Moore [5] predicted that that the use of technology would greatly change the way that people teach statistics. Subsequently in 1997 he argues that " technology strongly influences what we teach and how we teach. The case for substantial change in statistics instruction is built on strong synergies between content, pedagogy and technology" [6]. There are many computer programs, Java applets and sophisticated graphics calculators, all of which can be used to help develop students' understanding of descriptive and inferential statistics [7]. Reports of educators' use of technology, including computer software, graphics calculators and the Internet, show variation in effectiveness [8], including a report of research where there is no conclusive evidence of improvement [9]. Therefore, it is important to note the advantages and disadvantages of the various technological offerings and make careful decisions about when and where to use technology in the classroom [10].

In 2007 Garfield and Ben-Zvi [11] produced a review of how students learn statistics and note that "even a well-designed simulation is unlikely to be an effective teaching tool unless students' interaction with it is carefully structured". They also point out that although teachers would like some information on how to select a particular tool, the research literature does not give specific evidence as such. However, they highlight the fact that some recent studies have determined that:

> developing a deep understanding of statistics concepts is quite challenging and should not be underestimated. Research suggests that it takes time, a well-thought out learning trajectory, activities and discussion questions to develop deep understanding. Good reasoning about important concepts can be developed very carefully using enough time and revisiting of these ideas [12].

2.2 Understanding variation

Watson [13] draws attention to some of the misconceptions that secondary students have when considering sample size, randomness and variation, which together hinder their understanding of inferential statistics. She suggests that informal work with samples and inference should start in the primary school and points out that "Developing reasoning related to sampling may be associated with developing literacy and social skills rather than developing numeracy skills". This is endorsed by Ben-Zvi [7]. Watson and Kelly [14] indicate the need for students to not only consider averages, as is commonly done, but also "measures of spread and variation such as the standard deviation" at the school level in order to achieve the understanding of variation which is later needed for inferential statistics. This is in accord with the research reported by Shaughnessy [15] which highlights that "in the United States far more time is spent on notions of average, the 'signal' in data, than on the noise. The noise in data is variability, and often variability is just as important in data analysis as averages are, if not more so".

Unlike mathematics, statistics is contextually bound and requires data. Contexts relevant to the students are more likely to motivate and interest them. Students should have opportunities to choose their own samples and make comparisons [16]. However, they need to understand that if they are to make inferences then the samples must be representative of the whole population. The

Internet provides access to many possible data sets (e.g., Australian Bureau of Statistics, www.abs.gov.au) or students can collect data for a particular project. The second author regularly collects data from students in large first year undergraduate statistics units and uses this data throughout the unit (in lectures, in tutorials and in assessment tasks). Simple data displays very readily display the variation in the data, which can be further explored through analysis.

2.3 Sampling distributions

An understanding of statistical inference requires a real knowledge of the nature of sampling distributions because it involves "drawing conclusions about the characteristics of a population based on empirical observations of samples taken from the population" [17]. This involves an understanding of the idea of random sampling and how it is representative of the population, comprehending the difference between a sample statistic and a population parameter. It requires students to know how the size of the sample, and thus the characteristics of the sampling distribution, as well as the actual values of the data collected, determines the results and conclusions drawn from significance tests.

Technology can be used to aid students' learning about sampling distributions. Modern graphics calculators include a range of capabilities relevant to work in this area, including numerical and graphical data analysis, the construction of confidence intervals and a range of elementary hypothesis tests. In addition to these, programs have been written for graphics calculators [18] to allow students to take random samples and to explore the resulting sampling distributions. These can be used to develop an understanding of the Central Limit Theorem, if used appropriately. Java applets for computers and computer packages designed for teaching and learning statistics can be used for these sorts of purposes as well.

Research into the use of simulations to help understand the characteristics of sampling distributions was undertaken by Chance, delMas and Garfield [19]. A Sampling Distribution program SIM was developed and an activity was designed to develop students' understanding. Five research studies investigated how students could build up a sampling distribution of their own and were then given a set of questions designed to assess their understanding. The results suggested that there was some benefit. However, when the activities were redesigned to help students recognise their own misconceptions there was significant improvement.

A combination of practical and technological activities has been used at Murdoch University by the second author for more than a decade to aid students' understanding of the distribution of sample values and the Central Limit Theorem. During the first lecture of a large first year statistics unit, a questionnaire is completed by the students. This data set (each record having about 20 fields) is put into an *Excel* spread sheet and used as a data source in both tutorials and assessment tasks with students either working with the graphics calculator or *SPSS* or *Excel*. One of the questions asked of the students is their height. As there are generally more females than males, the students are given a "*population*" of female heights and asked to use random number tables to select two samples – one of size 9 and the other of size 25. They are then asked to calculate the mean and standard deviation for each of their samples using a calculator and to give the values to their tutor. These are then collated and a computer display created showing the three distributions (population heights, sample means for sample size 9 and sample means for sample size 25) one under the other on the same scale. During a lecture the displays and values (e.g. mean and standard deviation of all the sample means for each sample size) are discussed and related to the Central Limit Theorem. Whilst it is possible to simulate this activity on either a computer or a graphics calculator, the fact

of the students *owning* the data, both original and sample values, has been demonstrated to aid the learning processes.

2.4 Confidence intervals and hypothesis tests

Frequently introduced before formal hypothesis testing, confidence intervals can form a link between "sampling variability and hypothesis testing" [20]. There are statistical software for computers, programs (many inbuilt) for graphics calculators and Java applets on the Web which can generate many confidence intervals, some graphically, so that students can evaluate the proportion of intervals that do not contain the population parameter. They then can see that "the probability connected to a confidence interval is not the probability of being correct; rather, it is the relative frequency of correct statements that will be made in the long run by using this method" [21]. The concept of values being inside or outside the interval, i.e., at the tail ends of the distributions, then naturally leads to the idea of *usual* or *unusual* values and concepts for hypothesis testing. There are then several forms of technology which will perform the underlying calculations for the various tests. The more complex the data and hypothesis tests required, the more specialised the form of the technology – an important concept for the students to understand.

2.5 Graphics calculators: accessibility and suitability

Increasingly sophisticated graphics calculators provide a large range of tools for developing statistical understanding [22]. Both computer packages and graphics calculators provide opportunities to enhance student learning. The authors would agree that, because of the size of computer screens and computing power, ideally students should have access to statistical packages such as they are likely to be required to use when in the workforce. An important task in teaching students to use technology is to make them understand the role of each kind of technology, from its use as part of the learning process to its use in industry or research. However, there are some immediate and some not so obvious advantages to using a graphics calculator. These hand-held computers are relatively cheap and accessible, especially where computer laboratories are not always available and software licences expensive [23]. Many models have suitable software built in. Less obviously these graphics calculators give very easy access to the exploration of the magnitude of changes in sample characteristics and values which lead to changes in statistical inferences are made from the data [24].

The large screen and variety of statistical displays makes the use of a statistical package on a computer *the* most desirable tool for some activities. The accessibility (cost, size, power source, 24/7 availability) and certain other features make the graphics calculator *the* most desirable tool for other activities. The following **investigations** detailed in this paper cannot be carried out easily on a computer as, in general, hypothesis tests when only summary sample variables are known are not possible – raw data is required. They are designed to enable students to explore for themselves, in the context of hypothesis testing, **what** happens **if** certain values change.

3 Investigations

3.1 Investigation 1: Is 51% a majority?

A common use of proportions is to show that a majority of individuals fall into a particular category – be it the proportion of cars that are red or the proportion of Australian adults in favour of the development of more nuclear power stations. Ask any member of the general public, "*Is 51% a majority*?" and they would probably answer, "*Yes*!". Of course they would be quite correct if the 51% was obtained using the entire target population. However, when we ask a class of students, "*Is

51% a majority?", we are really asking if 51% is **significantly** larger than 50% given the size of the sample that gave us the 51%. The answer is often, "*I am not convinced it is a majority.*"

Using a graphics calculator it is very quick and easy to work through several examples seeing **what** happens **if** this value changes. The screens below in Figure 1 are from a Casio fx-9860G using a 1-sample z-test for proportions from the *STAT – TEST* menu using a sample proportion of 0.51 (51%) from a sample of size 100.
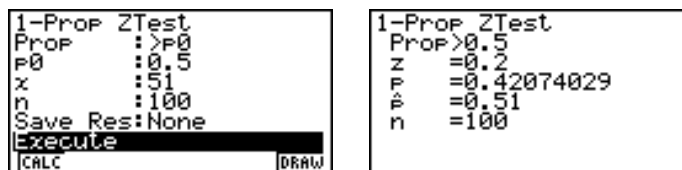

Figure 1: Testing a proportion with a graphics calculator.

Thus, for testing the null hypothesis that the population proportion is equal to 0.5 against the alternative hypothesis that the population proportion is greater than 0.5, using the P-value of 0.4207 and a 5% level of significance we would not reject the null hypothesis and would conclude that there is really not sufficient evidence to claim a majority. The P-value in this case suggests that 42.07% of the time when the population proportion is actually 0.5, a result of 0.51 or greater will occur; in other words, such a value is *not at all unusual* and a good deal more likely than the 5% figure chosen to define statistically *significant*.

So what sample proportion is needed for this particular sample to convince us there is a majority? Very quickly, through repeated use of the 3 steps: *EXIT* – change 51 – *EXE*, students can explore for themselves and see that a sample proportion of 59% is required before we are convinced there is a majority, as represented by a P-value less than 5%. A useful exercise is to make a note of the values such as those shown in the following Table 1.

Table 1: P-values associated with various sample proportions for sample size 100

sample proportion	0.51	0.52	0.53	0.54	0.55	0.56	0.57	0.58	0.59
P-value	0.4207	0.3446	0.2742	0.2118	0.1586	0.1151	0.0808	0.0548	0.0359

Most students will find this 59% value a very surprising result and are likely to ask, "*Why do we need such a large sample proportion to be convinced of a majority?*". The answer is because in this case the standard error for proportions (standard deviation for distribution of sample proportions) is a relatively large 0.05. Thus to be at least 1.645 standard errors above the hypothesised population value of 0.5, the sample proportion needs to be $1.645 \times 0.05 = 0.08225$ larger i.e. have a value of 0.58225 or more.

"*So when can we conclude that there is a majority with a sample proportion of 51%?*" In the above exercise the sample proportion was changed. The following Table 2 gives the results of an exercise where the sample proportion is kept the same but the sample size is changed. Again it only takes a few seconds on the graphics calculator to make the changes and perform the tests. The most difficult part is working out the *x* value having chosen the sample size, *n*. However, this can be easily overcome by entering $0.51 \times n$ (e.g. 0.51×200) for *x*.

Table 2: P-values associated with various sample sizes when sample proportion is 0.51

sample size	100	200	500	1000	2000	5000	6000	7000
P-value	0.4207	0.3886	0.3274	0.2635	0.1855	0.0786	0.0607	0.0471

A little more trial and adjustment yields a P-value of 0.0495 for a sample size of 6800. Students find this another surprising result. "*Why is such a large sample size required before a sample proportion of 51% leads to a conclusion of a majority using a 5% level of significance?*". The answer, again, is due to the size of the standard error. In this case 1.645 standard errors have to be less than 0.01 (i.e. 0.51 – 0.50). Thus the standard error has to be less than 0.006079.

These exercises lead very easily into a discussion about the effect of the sample size on the standard error. Once more it is very quick and easy to generate a set of standard errors for different sample sizes – which could be added to the previous table as shown in Table 3 below.

Table 3: Standard errors and P-values associated with various sample sizes when sample proportion is 0.51

sample size	100	200	500	1000	2000	5000	6000	7000
standard error	0.05000	0.03536	0.02236	0.01581	0.01118	0.00707	0.00645	0.00598
P-value	0.4207	0.3886	0.3274	0.2635	0.1855	0.0786	0.0607	0.0471

Use of the formula for the standard error gives $n = 0.5 \times 0.5 \div 0.006079^2 = 6765.1$ i.e., giving a minimum sample size of 6766. However, since $x = 0.51n = 0.51 \times 6766 = 3450.66$ is not a whole number, this value of n will not work. Working on the calculator it is possible to quickly show that the smallest value greater than 6766 for which a sample proportion of 0.51 is possible is 6800.

If time permits, a final exercise, which combines the first two exercises, can be used to construct a table such as that shown in Table 4 below. Here the **P-values** are displayed for different sample proportions, $\hat{p}$, and sample sizes, n, with standard errors, SE, also shown.

Table 4: P-values for different sample proportions and sample sizes, with standard errors

n SE $\hat{p}$	0.51	0.52	0.53	0.54	0.55	0.56	0.57	0.58	0.59
100 0.05	0.4207	0.3446	0.2742	0.2118	0.1586	0.1151	0.0808	0.0548	0.0359
200 0.03536	0.3886	0.2858	0.1981	0.1289	0.0786	0.0448			
500 0.02236	0.3274	0.1855	0.0899	0.0368					
1000 0.01581	0.2635	0.1029	0.0289						
2000 0.01118	0.1855	0.0368							
5000 0.00707	0.0786	0.0023							
6000 0.00645	0.0607	0.0010							
7000 0.00598	0.0471								

3.2 Investigation 2: Is there a difference in means?

Consider data from two dependent samples. These can come from an experimental design of either repeated measures (e.g. body temperatures of patients with a fever before and two hours after taking a certain medication) or paired samples (e.g. fever patients paired according to disease and initial body temperature, given either a placebo or the medication and two hours later having their temperatures recorded). The question being asked is, "*Does the medication reduce the patient's temperature?*" (repeated measure) or, "*Does the medication work better than the placebo in reducing the patient's temperature?*", (paired samples). More explicitly: is there a **significant difference** (specifically reduction) in temperatures, on average, or is the average temperature after

taking the medication ***significantly different*** from (specifically less than) the average temperature after taking the placebo?

Again, using a graphics calculator it is very quick and easy to work through several examples seeing ***what*** happens ***if*** this value changes. The screens in Figure 2 below are from a Casio fx-9860G using a 1-sample t-test for means from the STAT – TEST menu with a sample mean difference of 1 and a sample standard deviation of 1 for a sample of size 5. (For the temperature scenarios assume the repeated measures design is temperature difference *before minus after* and the paired design is temperature difference *placebo minus medication*.)
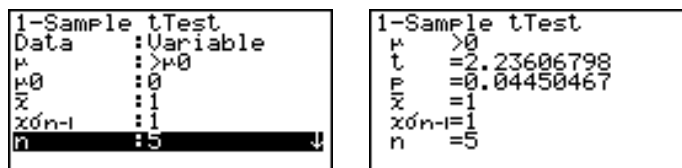

Figure 2: Dependent samples t-test via a graphics calculator

Thus, for testing the null hypothesis that the population mean difference is equal to 0 against the alternative hypothesis that the population mean difference is greater than 0, we note that the P-value of 0.0445 suggests that a temperature reduction this large or larger is expected only 4.45% of the time if the null hypothesis were true. This is an *unusually small* probability, so we would reject the null hypothesis at a 5% level of significance and would conclude that there is some support for a reduction in temperature after medication.

The exercises for proportions essentially had two variables: the sample proportion and the sample size. Equivalent exercises for means need to consider three variables: the sample mean, $\bar{x}$, the sample standard deviation, s, and the sample size, n.

An initial exploration for the students is to hold both the sample size and sample standard deviation constant and look at the effect of varying the sample mean. As with tests for proportions, it literally takes only a couple of seconds to make changes on the calculator to the initially performed test and a couple more to note the result in a table such as that shown in Table 5 below for $n = 10$ and $s = 3$.

Table 5: P-values associated with various sample mean differences when $n = 10$ and $s = 3$

sample mean	1	1.2	1.4	1.6	1.7	1.75	1.8
P-value	0.1596	0.1188	0.0871	0.0630	0.0534	0.0491	0.0451

A parallel exercise repeats the one above but using a different sample standard deviation. This could lead to constructing a two way table such as that shown in Table 6 below giving the P-values for the indicated sample mean and standard deviation values but keeping the sample size as 10.

Table 6: P-values for different sample mean differences and standard deviations when $n = 10$

s \ $\bar{x}$	1	1.2	1.4	1.6	1.7	1.75	1.8
3.0	0.1596	0.1188	0.0871	0.0630	0.0534	0.0490	0.0451
2.5	0.1188	0.0817	0.0551	0.0368			
2.0	0.0742	0.0451					
1.5	0.0321						

In this scenario the sample mean needs to be at least 1.8332 (t-distribution with 9 degrees of freedom) standard errors above the hypothesised population mean of zero. Therefore, the sample mean needs to be at least 1.8332 standard errors. Thus $\bar{x} > 1.8332 \times s \div \sqrt{10}$ $\therefore$ $\bar{x} > 0.5797s$.

One of the key discussion points with students when performing the proportions exercises should be how the sample size affects the standard error (proportions). A similar discussion for means needs to consider how both the sample size and the sample standard deviation affect the standard error (means).

Table 7 (below), displays the results of an exercise designed to look at the effect of change in sample size while holding the sample mean difference at 1 and the sample standard deviation at 2.

Table 7: P-values associated with various sample sizes when $\bar{x} = 1$ and $s = 2$

sample size	5	6	7	9	11	12	13	14
P-value	0.1631	0.1376	0.1170	0.0860	0.0641	0.0556	0.0483	0.0420

The next exercise looks at the effect of changing the sample standard deviation, holding the sample mean difference at 1 and the sample size at 10. Results are shown below in Table 8.

Table 8: P-values associated with various sample standard deviations when $\bar{x} = 1$ and n = 10

sample standard deviation	5	4	3	2	1.8	1.7	1.6
P-value	0.2714	0.2248	0.1596	0.0741	0.0564	0.0479	0.0398

A suitable follow-up explores the size of the standard error as both the sample size and the sample standard deviation change. Some examples are given in the partially completed Table 9 below.

Table 9: Standard errors for various sample sizes, *n*, and standard deviations, *s*.

n \ s	1.0	1.5	2.0	2.5	3.0	3.5	4.0
5	0.4472	0.6708	0.8944	1.1180	1.3416	1.5652	1.7889
8	0.3536	0.5303					1.4142
10	0.3162	0.4743					1.2649
15	0.2582	0.3873					1.0328
20	0.2236	0.3354					0.8944
25	0.2000	0.3000	0.4000	0.5000	0.6000	0.7000	0.8000

Suitable discussions with students might relate to the direct effect of the sample standard deviation (i.e. a doubling of *s* produces a doubling of the standard error) versus the inverse square root effect of the sample size (i.e. a quadrupling of *n* produces a halving of the standard error).

Unlike the investigation for proportions, the choice of starting values for the sample mean difference, sample standard deviation and sample size displayed in the above tables are somewhat arbitrary. It would be much better to use a specific example, arising as an example in a lecture or in a tutorial exercise, as the starting point for this investigation for means.

4. Conclusion

Over the last 30 plus years, more and more statistical computer packages have been developed. Some have been designed to aid students develop an understanding of probability and statistics, while others have been designed for professional statisticians. Sophisticated graphics calculators

increasingly provide a large range of tools for enhancing student learning and developing statistical understanding. The authors agree that students should ideally have access to statistical packages, particularly to prepare them for the workforce, but the statistical facilities on many models of graphics calculators provide excellent, attractive learning tools for students. The second author hires out graphics calculators to those students who have no prior experience of using one in the school system. Just over half way through every semester there are requests from many students who had previously decided to make do with their own calculator, to hire a graphics calculator, as they have been impressed by how the learning of their peers has been enhanced. These graphics calculators give very easy access to the exploration of the magnitude of changes in sample characteristics and values which lead to changes in the statistical inferences that are being made from the data. Much of this exploration is not possible using a computer package. The investigations detailed in this paper allow students to better understand the mechanisms involved in statistical inference rather than just use (any) technology as a *black box* means to an end.

References

[1] Chance, B., delMas, R., & Garfield, J. (2004). Reasoning about sampling distributions. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 295-318). AA Dortrecht, Netherlands: Kluwer Academic, p.295.

[2] Australian Education Council. (1991). *A national statement on mathematics for Australian schools. Carlton,* Victoria.

[3] Schuyten, G., & Thas, O. (2007). Statistical thinking in computer-based learning environments. *International Statistical Review*, 75(3), 365-371.

[4] Bradley, J., & Kemp, M. (2002). *The "plus" provided by graphics calculators in teaching undergraduate statistics.* Paper presented at the 2nd International Conference on the Teaching of Mathematics, Crete.

[5] Moore, D. S. (1990). Uncertainty. In L.A.Steen (Ed), *On the shoulders of giants: New approaches to numeracy* (pp. 95-137). Washington, DC: National Academic Press.

[6] Moore, D. S. (1997). New pedagogy and new content: The case of statistics. *International Statistical Review,* 65(2), 123-141, p.123.

[7] Ben-Zvi, D. (2000). Towards understanding the role of technological tools in statistical learning. *Mathematical thinking and learning,* 2(1&2), 127-155.

[8] Garfield, J., & Ben-Zvi, D. (2007). How students learn statistics revisited: A current review of research on teaching and learning statistics. *International Statistical Review*, 75(3), 372-396

[9] Collins, L. B., & Mittag, K. C. (2005). Effect of calculator technology on student achievement in an introductory statistics course. *Statistics Education Research Journal*, 4(1), 7-15. http://www.stat.auckland.ac.nz/serj, p.376.

[10] Kemp, M., Kissane, B., & Bradley, J. (1998). *Learning undergraduate statistics: The role of the graphics calculator.* Paper presented at the International Conference on the Teaching of Mathematics, Samos, Greece.

[11] Garfield, J., & Ben-Zvi, D. (2007). How students learn statistics revisited: A current review of research on teaching and learning statistics. *International Statistical Review*, 75(3), 372-396, p.376.

[12] Garfield, J., & Ben-Zvi, D. (2007). How students learn statistics revisited: A current review of research on teaching and learning statistics. *International Statistical Review*, 75(3), 372-396, p.379.

[13] Watson, J. M. (2004). Developing reasoning about samples. In D. Ben-Zvi & J. Garfield (Eds.), *The Challenge of Developing Statistical Literacy, Reasoning and Thinking* (pp. 277-294). AA Dortrecht, Netherlands: Kluwer Academic, p.278.

[14] Watson, J. M, & Kelly, B., A. (2007). Assessment of students' understanding of variation. *Teaching Statistics*, 29(3), 80-88, p.80.

[15] Shaughnessy, J. M. (2006). Research on students' understanding of some big concepts in statistics. In G. Burrill, F & P. Elliott, C (Eds.), *Thinking and reasoning with data and chance: Sixty-eigth yearbook* (pp. 77-98). Reston, VA: National Council of Teachers of Mathematics, p.86.

[16] Watson, J. M. (2004). Developing reasoning about samples. In D. Ben-Zvi & J. Garfield (Eds.), *The Challenge of Developing Statistical Literacy, Reasoning and Thinking* (pp. 277-294). AA Dortrecht, Netherlands: Kluwer Academic.

[17] delMas, R. C. (2004). A comparison of mathematical and statistical reasoning. In D. Ben-Zvi & J. Garfield (Eds.), *The Challenge of Developing Statistical Literacy, Reasoning and Thinking* (pp. 7-95). AA Dortrecht, Netherlands: Kluwer Academic, p.88.

[18] Kissane, B., & Kemp, M. (2006). *Mathematics with a graphics calculator: Casio fx-9860G AU*. Mirrabooka. WA: The Mathematical Association of Western Australia (Inc).

[19] Chance, B., delMas, R., & Garfield, J. (2004). Reasoning about sampling distributions. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 295-318). AA Dortrecht, Netherlands: Kluwer Academic.

[20] Bertie, A., & Farrington, P. (2003). Teaching confidence intervals with Java applets. *Teaching Statistics*, 25(3), 70-74, p.70.

[21] Koehler, M. H. (2006). Using graphing calculator simulations in teaching statistics. In G. Burrill, F & P. Elliott, C (Eds.), *Thinking and reasoning with data and chance: Sixty-eigth yearbook* (pp. 257-272). Reston, VA: National Council of Teachers of Mathematics, p.267.

[22] Kissane, B. (1998). *Inferential statistics and the graphics calculator*. Paper presented at the Third Asian Technology Conference on Mathematics, Tsukuba, Japan.

[23] Bradley, J., Kemp, M., & Kissane, B. (1994). Graphics calculators: A (brief) case of technology. *Australian Senior Mathematics Journal*, 8(2), 23-30.

[24] Koehler, M. H. (2006). Using graphing calculator simulations in teaching statistics. In G. Burrill, F & P. Elliott, C (Eds.), *Thinking and reasoning with data and chance: Sixty-eigth yearbook* (pp. 257-272). Reston, VA: National Council of Teachers of Mathematics.