ENHANCING STUDENT UNDERSTANDING IN STATISTICAL INFERENCE – ASSESSING THE EFFECTIVENESS OF A COMPUTER INTERACTION

FRANCIS, Glenda, KOKONIS, Sue & LIPSON, Kay
Swinburne University of Technology
Australia

A new simulation has been developed to facilitate developmental learning of statistical inference. This simulation has been designed in the light of current multimedia design principals and cognitive theory. While many simulations have been developed to help students understand a variety of statistical concepts, evaluations of what these simulations actually achieve have been relatively scarce. This paper presents a model for the evaluation of simulations. In particular, the paper discusses the way in which the development of the students' conceptual understanding has been assessed. Some preliminary results from the evaluation of this specific project are presented.

INTRODUCTION

Traditionally students find it very difficult to grasp the logic behind inference, but such understanding is important if significance testing is to be used intelligently. While many simulations have been produced as aids to student understanding, very little evaluation of such simulations has been reported (delMas, Garfield & Chance, 1999).

One such evaluation conducted by the authors, involved talking with the students as they worked their way through a computer based simulation (Lipson, Kokonis, & Francis, 2003). This study revealed some of the shortcomings of the interactive, and also provided an insight into the cognitive stages students progressed through in developing an understanding of the logic behind hypothesis testing. The research suggested that students passed through distinct stages in the construction of a schema for hypothesis testing and that students need to fully grasp each stage before being able to progress to the next stage. Briefly, these stages are:

- Recognition: Forming an understanding of the empirical sampling distribution
- Integration: The assimilation of an observed sample result with the sampling distribution. This involves locating the sample proportion on the sampling distribution and making a judgement about whether such a sample proportion is 'likely' to have come from the population used to generate the sampling distribution.
- Contradiction: Recognising that there is an inconsistency between the sample and the hypothesised population.
- Explanation: Considering possible statistical explanations for the contradiction between the observed sample and the hypothesised population.

The explanation stage is where many students run into problems. If the sample is inconsistent with the population, they instinctively look for problems with the sample, and don't tend to consider the possibility that it is the hypothesis which may be incorrect – that the sample may come from a different population.

The authors have used the information from the earlier study to develop a new simulation; one which can be used independently by students. This simulation was designed according to principals of multi-media design, and with the aim of leading students through the conceptual stages articulated in the previous study (Lipson, Francis, & Kokonis, 2006).

This new simulation was trialled with students in a first year introductory course in statistics, and the development of the students' conceptual understanding was tracked across several weeks of study. As some studies suggest that simulations are most effective when introduced before theoretical discussion of the topic (Zhou et al, 2005), the simulation was introduced at different stages for different tutorial groups, with the aim of comparing the progress in understanding across the different timings. For some students the simulation was introduced before a lecture on the theory behind inference and for other students it was introduced after the theoretical lecture.

THE SIMULATION

The simulation was written in three parts, and was designed to specifically address critical junctures in the students' conceptual development. Based on the theories of multi media design (Mayer, 2002) the context was kept extremely simple, so that there was nothing in the scenario to distract students from the concepts being expounded. The simulation is based around taking samples of jelly beans from a (large) jar, which contains both red and black jelly beans. Part I aims to develop an understanding of sampling distributions, and looks at whether individual samples are consistent or inconsistent with this sampling distribution. This aims to address the 'recognition' and 'integration' phases identified in the previous study.

Bob's jar of jelly beans is introduced. This is a population with a known proportion of black jelly beans (40%). Student take repeated samples from Bob's jar and calculate the proportion of black jelly beans in each sample. The sample proportions are recorded one by one and a histogram built up as the sample proportions are added. A normal curve is added to the histogram, and likely/unlikely regions displayed on the histogram. Throughout the simulation, questions prompt the students to build their understanding of what this histogram represents and to reflect on what sorts of sample proportions we could expect in samples taken from a population where 40% of the jelly beans are black. There are explicit questions asking whether specific sample proportions are likely or unlikely in samples taken from this population. After exploring samples of size 20, the process is repeated with samples of size 50 and the students are prompted to reflect on the effect of sample size on the sampling distribution.

Parts II and III of the simulation focus on addressing the 'contradiction' and 'explanation' phases of conceptual understanding. Part II of the simulation begins again with building a sampling distribution, recapping what was done in Part I. At this point a new jar (population) is introduced. The proportion of black jelly beans in the new jar (Kay's jar) is unknown. A single sample is taken from Kay's jar, and students are asked to speculate if this Kay's jar has the same proportion of black jelly beans as Bob's. This is typically where students have had most difficulty in the past – making the conceptual leap from exploring the sorts of sample proportions that could be expected from a population, to taking a single sample and using it to draw a conclusion about the population it was drawn from. Part III extends the concept further to introduce confidence intervals. Students use the simulation to explore the potential populations the sample might have come from.

TESTING CONCEPTUAL UNDERSTANDING

A series of both formal and informal assessments were used throughout the semester to monitor the development of conceptual understanding. Some of the assessments used open ended questions and others used multiple choice questions. The tests were aimed at measuring the following skills/concepts:

- To understand what is meant by a population and a sample.
- To appreciate what an empirical sampling distribution represents. (*Recognition*)
- To identify if a particular sample is consistent with a given sampling distribution. (*Integration*)
- When a sample is inconsistent with the (null) hypothesis sampling distribution, to recognise that that this represents a contradiction. (*contradiction*)
- To recognise that this apparent contradiction could be explained by postulating that the sample came from a different population. (*explanation*)
- To understand how the Binomial Test relates to this process.

The set of questions presented in Figures 1 & 2 are typical of one type of questioning used throughout the semester. They relate specifically to the theory behind hypothesis testing.

In the 1990's Australia Post published a report in which they claimed that at least 96% of letters are delivered on time. A journalist decided to test that claim by posting several letters. Of the 59 letters that he posted, he found that 52 were delivered on time, and thus he wrote an article in the newspaper with the headline "Doubt on Letters Promise".

1. What is the population in this study?
   - o All letters sent via Australia Post
   - o All letters sent by the journalist
   - o All Australian adults
   - o All Australians who send letters

2. What proportion of the journalist's letters were delivered on time?
   Sample proportion = _____________

3. Suppose Australia Post's claim is correct, and 96% of all letters were delivered on time. We produced a sampling distribution of 100 samples of size 59 taken from a population where 96% of letters are delivered on time – this sampling distribution is displayed in Figure 1.

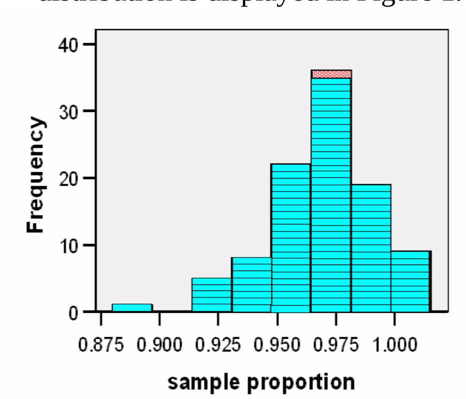

Figure 1: Sampling distribution for 100 samples of size 59.

One of the blocks in the sampling distribution is highlighted. What does this block represent?

- o The median number of letters delivered on time
- o One letter delivered on time by Australia post
- o One sample of 59 letters in which 97.5% were delivered on time
- o One sample of 59 letters which were all delivered on time
- o The mean number of letters delivered on time

4. Figure 2 shows the sampling distribution for samples of size 59, with unlikely sample proportions shaded and a normal curve fitted. Click on the part of the distribution where the journalist's sample falls.

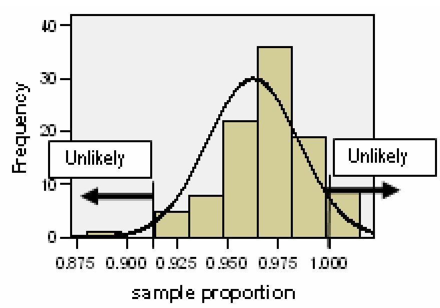

Figure 1: Multiple choice version of sampling theory test questions (Part 1)

5. Does the journalist's sample look like it belongs to this distribution?
   - Yes
   - No
6. What conclusion can we reach from this?
   - We can't conclude anything
   - Australia Post's claim that 96% of letters are delivered on time is probably true
   - Australia Post's claim that 96% of letters are delivered on time is probably false
   - Australia post delivers more than 96% of letters on time

7. What SPSS procedure would we use to estimate the probability of getting a sample with proportion 0.88 from a population where the proportion was 0.96?
   - The 'binomial test' procedure
   - The 'frequencies' procedure
   - The 'one-sample t-test' procedure
   - The 'explore' procedure

Figure 2: Multiple choice version of sampling theory test questions (Part 2).

Sets of questions similar to this one were given to the students several times throughout the semester, to keep track of their progress. The particular set of questions given in Figures 1 & 2 was used twice; on the first occasion as open ended questions and then a week later in multiple choice format. Each of these formats has its own advantages and disadvantages. Some responses to the open ended questions were ambiguous and it wasn't possible to tell whether the student understood the theory or not. For example, while some responses to the first question 'what is the population in this study' were clear and unambiguous: "*All letters sent through Australia Post*", "*All letters sent by the journalist*", other responses were less precise: "*all letters*", or "*all of the letters*". Perhaps "*all letters*" is more likely to refer to all letters sent via Australia Post, while "*all of* the *letters*' is more likely to refer to all of the letters sent by the journalist. However it's not entirely clear that this was what the students intended. Giving multiple choice options avoids this problem, but certainly doesn't offer the students as much scope for revealing misconceptions as the open ended form! In order to make the responses from the multiple choice version as comparable as possible to the responses to open ended questions, the options given in the multiple choice should be modelled on the types of responses students actually give to the open ended questions.

TIMING OF PRESENTATION AND TESTING

Half of the tutorial groups completed part I of the simulation (the introduction to the sampling distribution) before the first lecture on the theory behind inference. The remainder of the tutorial groups completed part I of the simulation after the lecture. Only the 76 students who attended the lecture on inference are included in this study. The sequence of activities and tests for both groups of students is displayed in Table 1.

Table 1
Sequence of Activities and Tests for Each Group

Week	*Group 1*	*Group 2*
3	Simulation Part I	
4	Lecture	Lecture
4	Test 1	Test 1
4	Simulation Part II	Simulation Part I
5		Simulation Part II
5	Test 2	Test 2

Further testing was conducted over the following three weeks, and some of these tests included open ended questions of the same style as those given here, but based on a different scenario. Questions on inference will also be included in the end of semester examination. As yet, only the first two of these tests have been administered and analysed.

RESULTS

The responses to two of the test questions (3 & 6) are considered here. Question 3 was designed to assess whether students understand what a sampling distribution represents. The percentage of students in each group who answered this question correctly is displayed in Table 2, for both tests 1 and 2.

Table 2
Performance of Each Group on Selected Test Questions

	Percentage of correct answers				
	Q3 Test 1	*Q3 Test 2*	*Q6 Test 1*	*Q6 Test 3*	*n*
Group 1(Sim I before lecture)	42%	77%	5%	37%	43
Group 2(Sim I after lecture)	6%	39%	6%	27%	33

At the time of the first test, students in group 1 had completed part I of the simulation, which focuses on constructing sampling distributions. The group who had attended the lecture, but had not yet done the simulation performed very poorly. This is possibly because there was very little time between the lecture and the test in which to absorb and integrate the information from the lecture.

At the time of the second test, all students had completed all three parts of the simulation. Students who had been exposed to the part I of the simulation before the lecture substantially out performed those who did not use the simulation until after the lecture. This suggests that exposure to the simulation before the lecture helped students to integrate the information given in the lecture.

Question 6 which is aimed at assessing the ‘explanation’ phase of theoretical understanding. This is the phase which students find most difficult. The percentage of correct responses to this question are also presented in Table 1, for both tests 1 and 2. At the time of test 1, the information from the lecture had clearly not had much impact on the students’ understanding of this explanation phase. Only 5% of the students gave responses indicating that Australia Post might be overstating the percentage of letters delivered on time. As part I of the simulation stops after the construction of the sampling distribution, it offers no help to the students with this question, and both groups had a similar poor performance.

Performance was substantially better on test 2 than test 1, but the group who used the simulation before the lecture once again seem to have gained some additional benefit. However, for both groups, it is still only a very low proportion of students who are able to see that a possible explanation for the contradiction could be that the population proportion is less than 96%. Following these results, more time was spent on this concept in the lectures and it will be interesting to see whether results improve further throughout the semester.

From the responses to the open ended question on test 1, it is evident that students, having recognised an inconsistency between the sample and the sampling distribution, are more likely to search for an explanation related to the sample than an explanation related to the population. Many students gave answers here that indicated mistrust of the journalist, for example “*the journalists sample may have been biased* ”. The other most common type of response was to simply report that this was an unlikely sample proportion without then questioning the population from which the sample may have been drawn. The multiple choice alternatives given for this item on test 2 did not offer enough scope for students to repeat this mistake, since almost all of the alternatives refer to the population, and none specifically refer to the sample. The improvement in performance between tests 1 and 2 might in part reflect this shortcoming in the possible responses offered. The next test, to be administered in week 8 uses a similar style of questions to those reported here, and like test 1, are open ended in format. If the apparent improvement in performance between weeks 4 and 5, is solely due to the multiple choice format, this will be evident in the analysis of these further test responses.

LIMITATIONS AND FUTURE WORK

The results presented here need to be viewed with some caution. As existing groups were used, there may be other factors which differ across the groups apart from just the order in which the learning materials were administered. Work is currently underway investigating possible differences in the backgrounds of the students in the different groups, to try to identify any potential confounding factors.

There were also several problems with the set of questions reported here. The 'block' which was highlighted in the empirical sampling distribution happened to be at the top of the highest column of the histogram. This distracted students into focussing on the 'typical' values in the distribution and they gave responses relating to the mean, the median or the mode, rather than looking at what each individual block in the distribution represents. Some of the improvement in correct response rate across the semester just refects that the students have learnt what we are looking for in this type of question, rather than an improvement in the number of students who understand what a sampling distribution represents. In later versions of these questions a different block of the histogram was chosen. The second issue was some confusion created for students by introducing a specific sample other than the journalist's sample. Several students answered the remainder of the questions as if the journalist's sample had a proportion of 0.975 (the proportion in the highlighted block) instead of a proportion of 0.88. In later versions of the test the order of the questions 2 and 3 were reversed, so that the proportion in the journalists sample was mentioned immediately before it was displayed on the sampling distribution.

Changes to the multiple choice options for question 6 will also be made next semester to more accurately reflect the sort of responses students give to the open ended questions.

REFERENCES

delMas, R., Garfield, J., & Chance, B. (1999). A model of classroom research in action: Developing simulation activities to improve student's statistical reasoning. *Journal of Statistics Education, 7*(7).

Lipson, K., Francis, G., & Kokonis, S. (2006, July 2-9). *Developing a computer interaction to enhance student understanding in statistical inference.* Paper presented at the ICOTS7 conference, Salvador.

Lipson, K., Kokonis, S., & Francis, G. (2003, August 11-12). *Investigation of students' experiences with a web-based computer simulation.* Paper presented at the IASE Satellite conference on Statistics and the Internet, Berlin.

Zhou, G., Brouwer, W., Nocente, N., & Martin, B. (2005). Enhancing conceptual learning through computer based applets: the effectiveness and implications. *Journal of Interactive Learning Research*, *16*(1), 31-50.

Mayer, R. E. (2002). Cognitive theory and the design of multimedia instruction: An example of the two-way street between cognition and instruction. *New directions for teaching and learning, 89*(Spring), 55-71.