ASSESSING CONCEPTUAL CHANGE IN LEARNING STATISTICS

Sue Finch and Geoff Cumming, La Trobe University, Australia

StatPlay, our multimedia for introductory statistics, aims to promote understanding of fundamental concepts. We describe assessment of students' use of StatPlay's Sampling Playground for learning about sampling variability, standard error and sampling distributions. Students worked with StatPlay's interactive simulations, and undertook a range of prediction, estimation, labelling and explanation tasks. Results suggest that StatPlay's striking visual representations can promote good understanding of sampling variability, and that confrontation with prior misconceptions and experience with the results of repeated sampling can be valuable, in particular for understanding standard error.

Many intuitive statistical ideas are flawed and appear resistant to our best attempts at teaching (Shaughnessy, 1992). A view of learning as conceptual change requires a shift from these 'naïve statistics' ideas, based on everyday beliefs about probability and statistics, to accurate conceptions (Thomason, Cumming and Zangari, 1994). The potential of a computer environment for learning statistical concepts is widely recognised, but simulations and demonstrations do not always meet their promise, and careful evaluation is not routinely undertaken (Cohen and Chechite, 1997).

StatPlay is a multimedia learning environment for introductory statistics students (Thomason et al., 1994). It aims to promote conceptual change by paying close attention to the deep-seated and pervasive misconceptions in statistics; by providing striking, multiple, dynamically-linked representations of statistical concepts; and by allowing the user to explore relations between representations.

USING STATPLAY'S SAMPLING PLAYGROUND

Understanding variability is one of the 'big ideas' in statistics (Gal and Garfield, 1997). Learners first meet the idea of quantifying variability in a set of observations, followed by the difficult notion that there is variability also in sample statistics. In StatPlay's Sampling PlayGround (PG) users can explore variability in sample means and can sample repeatedly from a population of any shape, taking samples of any size. You can sample at various speeds; the slowest speed gives an animation of sample points and the sample mean. Sample results can be viewed in two modes: Figure 1 shows the means for a dozen successive samples; Figure 2 shows as ‘heaps’ the distribution of sampled points and the distribution of sample means.

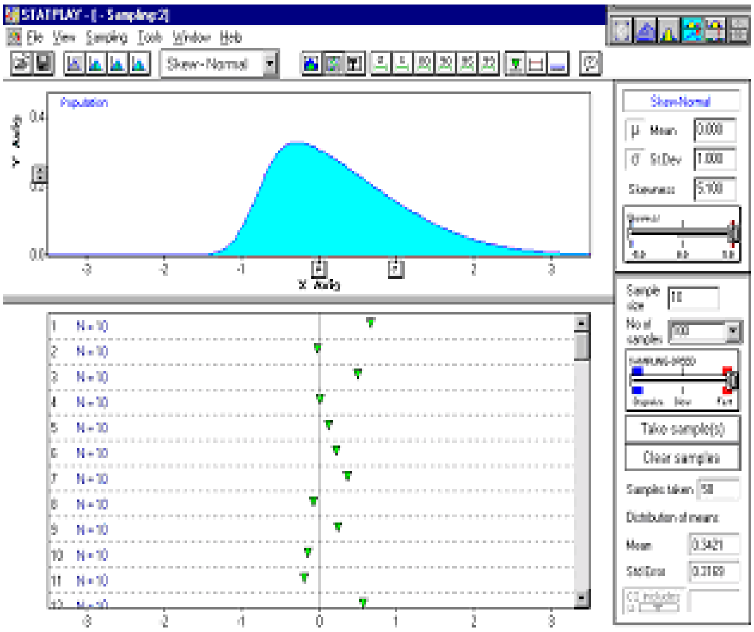

Figure 1. StatPlay’s Sampling PlayGround showing in the lower panel sample means of samples of size 10 from a skewed distribution.

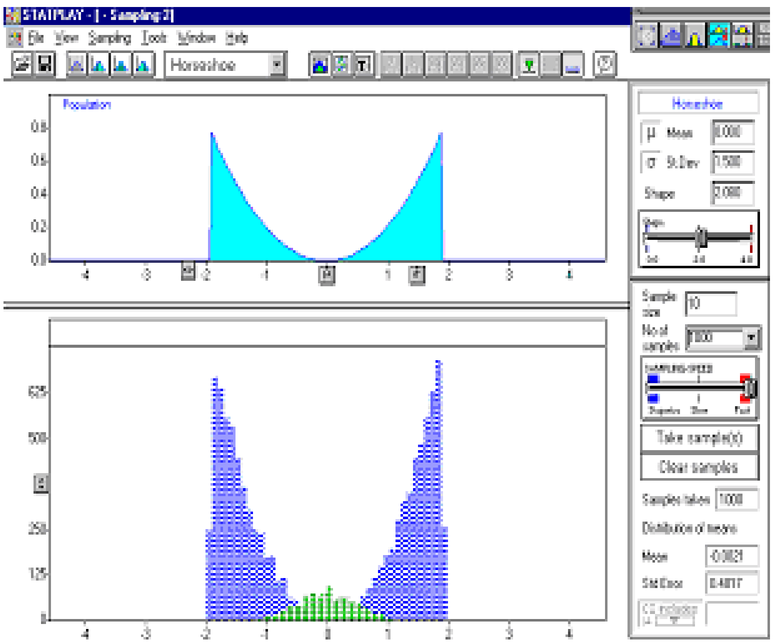

Figure 2. The ‘heaps’ view in StatPlay’s Sampling PlayGround showing in the lower panel means of samples of size 10 (small heap) and all sample points (large heap) from a horseshoe distribution.

Using the Sampling PG teachers can first introduce the idea that there is variability in sample means; scrolling through successive samples shows the ‘dance of the means’

(Figure 1). Sample means can be cumulated to show quantification of variability in a ‘heap’ of sample means. Using (a simulation of) an infinite number of samples, the sampling distribution of the mean is displayed and standard error can be introduced more formally. Students can explore the effects of changing sample size on the distribution of means, and the effects of sampling from different populations.

StatPlay’s Sampling PG has been used extensively in teaching first year psychology students at LaTrobe University over the past four years. Students work in small groups guided by worksheets and pre-recorded voice ‘demos’ (Les and Maillardet, 1998). This has proved a very effective way to guide efficient use of the software and to present applied problems. Students make predictions, run sampling experiments, identify multiple representations and, with guidance, systematically examine factors affecting sampling variability.

EVALUATING UNDERSTANDING

Our ongoing empirical evaluation of StatPlay considers a range of learning activities including, for example, different ways of thinking about variability in means (including standard error) and comparing students’ own estimates with StatPlay’s simulations. Concepts relating to sampling, sampling distributions and the Law of Large Numbers are assessed. Evaluation uses a range of methods for assessing conceptual understanding, including qualitative methods. Some examples follow.

PREDICTING VARIABILITY IN MEANS

Two teaching goals have been for students to develop (a) knowledge that means vary from sample to sample, and (b) sensitivity to the changes in this variation with changes in sample size. Assessment activities have asked students to mark on a scale plausible positions of a number of means for samples of a particular size from a specified population.

Two studies to date suggest the potential of StatPlay in promoting conceptual change. In a single-group, pretest-posttest quasi-experiment psychology students saw a lecture demonstration then undertook tutorial activities with StatPlay. They estimated variability in means for different sample sizes. While almost no sensitivity was evident prior to the StatPlay activities (mean sensitivity score = 1.1 where 1 indicates no sensitivity), students could accurately estimate relative changes in variability with sample size after (mean = 1.6 where 1.8 indicates appropriate sensitivity). Subsequently, an experimental study compared students undertaking StatPlay activities with control students (taught by the same tutor) working on similar activities but with pencil, paper and pocket calculators. A substantial increase in sensitivity to sample size was observed in the StatPlay group (from 1.4 to 1.9 on the scale above), with virtually no change from before to after the activities in the control group (remaining at 1.4). Together these results suggest that students can develop good intuitions about variability in sampling means by working with appropriate visual representations; the next pedagogical challenge is to ensure these intuitions carry over into applied practice.

CONFRONTING MISCONCEPTIONS

It has been suggested that students need to directly confront their misconceptions of statistical concepts to achieve better understanding (Shaughnessy, 1992). In two studies we compared students asked to make predictions based on judgments of their original ideas about sampling variability (confront group) with those undertaking similar learning activities without such judgements (control group).

In one experimental study, 18 students made predictions of variability in sample means and gave estimates of standard error (SE), then ran StatPlay simulations and evaluated their predictions relative to StatPlay’s results. In a similar amount of time (total 2 hours), 14 control subjects ran and examined the results of simulations. Subjects were learning introductory statistics but had not yet met in class the material covered. A post-test revealed few differences in estimates of variability in means and explanations of sampling variablity. However confronted subjects were better able to estimate SE, sketch a relevant sampling distribution, explain SE and provide a visual representation of SE. Few controls were able to provide any information about SE.

In a larger two-group class experiment, the same experimental manipulations were used. Students answered problems in class; there was a post-test 3 weeks later. Activities were concurrent with lectures. Most problems were presented as ‘real’ research problems with fixed response choices; in each case, students were asked to write down an explanation of their responses. Explanations were generally classed as appropriate (statistical), partly appropriate (ambiguous) or inappropriate (non-statistical).

Responses to a pre-test question about making inferences based on small samples suggest that the majority of students’ first intuitive response is a non-statistical one. On the final worksheet question, more confronted subjects gave correct explanations of SE. On the post-test, there was little group difference on the task of estimating variability in means for different sample sizes. Groups differed however on the nature of explanations given for SE. For example, half the confronted subjects were able to define SE (cf 20% of controls). Together the two studies suggest that confrontation was effective for learning the complex concept of SE. The constructivist nature of the confrontation activity may have contributed.

RECOGNISING VISUAL REPRESENTATIONS

StatPlay's visual representations aim to provide strong 'take home' images of important concepts, and to facilitate understanding by promoting links with other representations, such as formulae and words.

After laboratory sessions covering two PGs, students were given a label-to-pictures matching task with 6 concept labels and 4 pictures of StatPlay PGs—two of the Sampling PG (where the labels 'sampling variability', 'standard errror' and 'CLT' were appropriate) and two of a more complex PG (where 'Type I error', 'Type II error' and 'Power' were also appropriate). The picture best described as *sampling variability* was labelled as such by 84% of students; another 7% used some other appropriate label. There was only one instance (in > 100 students) of an incorrect label. The picture best described as *CLT* was labelled as such by 25%; 35% used one or more other valid labels (16% used both *CLT* and other valid labels). Incorrect labels were applied by 23% of subjects (although one-third also applied appropriate labels). The labelling of the more complex PG was not done so well. Results suggest that well-designed visual representations can indeed help in understanding sampling variability. Visual representations can also be confusing; we have found however that visual tasks can be a powerful way for identifying such confusion. Misconceptions that the student may otherwise find hard to articulate, and an expert teacher may not anticipate can be uncovered. Some students, for example, wrongly apply aspects of a sampling distribution to the population.

EXPLAINING SAMPLING VARIABILITY

Developing skills in communicating statistical ideas is an important curricular goal (Gal and Garfield, 1997). First year students at the end of a year long introductory course (with StatPlay) were given an open-ended pencil and paper task of explaining *sampling variability* to another student. Two aspects of the explanations were examined—the characteristic said to vary (sample observations, samples, means or some other statistic), and the way in which the characteristic was said to vary (in larger samples: less variation, more variation, closer to a normal distribution).

Nearly 90% of the 104 students mentioned one or more of the four characteristics; 66% referred to means, 40% referred to samples, and only 9% mentioned sample observations alone. Students explained in a variety of ways—with words, with pictures (some from StatPlay), with formulae, and with metaphor. Although not explicitly asked to do so, around 18% of subjects described a way in which the characteristic varied; 70% of these referred to less variation in larger samples. The quality of explanations varied but it is clear that the vast majority of students were aware of a central concern of inferential statistics—that every sample, and every sample mean can be different.

CONCLUSION

Building intuitions about statistical uncertainty and how it can be quantified depends in part on knowing what kinds of variations to expect from sample to sample. Common learning experiences give little of this, as students labour to sketch a sample distribution, calculate a mean or generate some output from a computing package. StatPlay, in contrast, allows students to dynamically explore the characteristics of many equally plausible samples from a given population in a single session—laying the groundwork for coming to terms with variability. Our studies to date suggests that the vivid visual representations of StatPlay can help students in understanding and estimating sampling variability, and, with appropriate activities, in understanding SE.

ACKNOWLEDGEMENT

This research has been supported by the Australian Research Council.

REFERENCES

Cohen, S. and Chechite, R. (1997). Probability distributions, Assessment and Instructional Software. In Gal, I. and Garfield, J. (Eds.) *The Assessment Challenge in Statistics Education*, Amsterdam: IOS Press.

Gal, I. and Garfield, J. (1997). *The Assessment Challenge in Statistics Education*, Amsterdam: IOS Press.

Les, L. and Maillardet, R. (1998). Online explanations for learners: The ‘Play it again Sam’ facility. Proceedings of ICOTS-5, The Fifth International Conference on Teaching Statistics, Singapore, June.

Shaughnessy, J. M. (1992). Research in probability and statistics: Reflections and directions. In Grouws, D. (Ed.), *Handbook of research on mathematics teaching and learning*, (pp.465-494), New York: Macmillan.

Thomason, N., Cumming, G. and Zangari, M. (1994). Understanding central concepts of statistics and experimental design in the social sciences. In Beattie, K., McNaught, C. and Wills, S. (Ed.s), *Interactive multimedia in University education: Designing for change in teaching and learning*, Amsterdam: North-Holland.