

CONSTRUCTING INFERENTIAL CONCEPTS THROUGH BOOTSTRAP AND RANDOMIZATION-TEST SIMULATIONS: A CASE STUDY

Maxine Pfannkuch and Stephanie Budgett

Department of Statistics, The University of Auckland, New Zealand

m.pfannkuch@auckland.ac.nz

Statisticians such as Cobb (2007) have promoted the use of computer intensive methods such as bootstrapping and the randomization test in introductory statistics courses. One of their arguments for using these simulations is that the logic of inference is conceptually more accessible to students than the traditional approach. In this paper we test indirectly the claim that simulations assist in the construction of inferential concepts using an analytical tool that is based on the versatile thinking framework for conceptual development. Using the tool, which identifies nine possible modes of student interaction with representations, we analyse two introductory statistics students' interactions with the Visual Inference Tools (VIT) bootstrap confidence interval construction and randomization test modules. Our findings suggest that, for these students, the VIT simulations were facilitating the development of statistical inferential concepts.

INTRODUCTION

Researchers such as Hesterberg (2006), Cobb (2007), and Engel (2010) have promoted the use of bootstrap and randomization-test simulations to improve students' understanding of inferential concepts and the logic of inference. They believe that the classical approach based on the sampling distribution with its mathematical procedures is inaccessible to many students. The challenge for such claims is to provide evidence that such simulations do assist in the construction of inferential concepts. It is conceivable that certain types of software, such as the use of spreadsheets, could continue to obscure the conceptual infrastructure in a similar way to the classical mathematical approach, whereas purpose-built software for learning such as the *Visual Inference Tools* (VIT) developed by Chris Wild (see: <http://www.stat.auckland.ac.nz/~wild/VIT>) may make the process more transparent. Hence the research question, which will form the focus of this paper, is "Do students develop concepts of statistical inference through experiencing and thinking with dynamic visualization simulations?" Since it is difficult to obtain such evidence we will gain an indication indirectly through an analysis of two students' interactions, one with the bootstrap VIT module and the other with the randomization-test VIT module. Before we embark on this analysis, we will refer to some literature and propose an analysis tool. We will then use the tool to analyze the students' interactions with the VIT modules, and draw some conclusions.

BACKGROUND

The introductory statistics curriculum was shaped by what was computable in the first half of the twentieth century, and there are now many reasons why the approach to statistical inference needs to change in education. First, the increasing power of technology means that the preferred method of inference in statistical practice is becoming empirical simulation methods (Hesterberg, 2006) resulting in the gap between statistical practice and education rapidly widening. Second, the logic of inference needs to be center-stage in introductory statistics courses, rather than the normal distribution, which Cobb (2007, p. 7-8) characterizes as "an intellectual albatross" and "a fraud." In particular, there is a disconnection between the randomization method and the type of inference it supports. We commit a fraud when we pretend "that choosing at random from a large normal population is a good model for randomly assigning treatments." Third, the simulation methods can be used for a variety of quantities of interest such as medians, quartiles, and measures of spread rather than being restricted to the mean. Fourth, in simulations the same type of thinking is used across a range of different situations and can be generalized "to other designs, other test statistics, and other data structures" (Cobb, 2007, p. 13). Fifth, many students find the mathematical approach to inference inaccessible resulting in the underpinning concepts remaining obscure. Part of the problem is that the sampling distribution is conceptually difficult because it can be viewed as both a process and an object, a procept (Gray & Tall, 1994). The sampling distribution can be perceived as an *object*, a probability distribution with its own properties and as being formed from the

process of drawing multiple random samples and recording from each sample a statistic such as the mean. “It is hard for students to make the transition from a one-time process like taking a limit at a point, or taking a single sample and computing the mean, to an abstract entity like a derivative or sampling distribution, created by applying the process repeatedly” (Cobb, 2007, p. 7).

Using and thinking with representations, consequently, play an important part in developing statistical thinking. Thomas (2008, p. 10) believes that the flexible use of representations assists students to develop rich schemas and conceptual understanding, which he summarizes in a *versatile thinking* framework comprising three elements:

- *process/object versatility*—the ability to switch at will in any given representational system between a perception of a mathematical entity as a process or an object;
- *visuo/analytic versatility*—the ability to exploit the power of visual schemas by linking them to relevant logico/analytic schemas;
- *representational versatility*—the ability to work seamlessly within and between representations, and to engage in procedural and conceptual interactions with representations.

Graham, Pfannkuch and Thomas (2009, p. 687) proposed an analytical tool (Table 1) that can be employed to differentiate qualitatively between representational activity that results in procedural thinking and the flexible use of representations that may lead to the construction of conceptual thinking. This tool identifies three types of activity that a student may engage in when interacting with a given representation – surface observation of the representation, deep or property observation of the representation, actions on the representation – in order to gain further understanding or information. For example, when engaging with a box plot a student: may notice the line in the box part is not in the middle, a surface observation; or may conjecture the distribution is skewed, a deep or property observation; or may take an action on the representation such as producing the dot plot of the data.

Table 1 Possible Modes of Student Interaction with Representations

Interaction	Procedural View		Conceptual View	
			Process	Structural/Object
Surface Observation	Procedural Surface Observation (PRSO)	Process Surface Observation (PSO)	Object Surface Observation (OSO)	
Deep or Property Observation	Procedural Property Observation (PRPO)	Process Property Observation (PPO)	Object Property Observation (OPO)	
Action on the Representation	Procedural Tool (PRT)	Conceptual Process Tool (CPT)	Conceptual Object Tool (COT)	

Each of these three modes is influenced by the thinking that the representation of the statistical idea evokes for the student – a procedural view, conceptual process view, or conceptual object view – and this depends on the schema evoked and used to understand it. By a procedural view we mean paying attention to, or using, a specific method or algorithm for carrying out a process. For example, a method for observing patterns in data is to plot the data, a procedural action, whereas *finding patterns in data* involves recognizing a process, which is a conceptual view. The other type of conceptual view in the tool is where, for example, a distribution (a collection of individual case values) is viewed as a conceptual entity or object, which has its own properties and which can be acted upon. Hence Table 1 shows there are nine different possible modes of student interaction with representations divided into those underpinned by a procedural perspective on representation and those where a conceptual view prevails. We now use the framework as an analytical tool on transcription data from two student interviews to provide indirect evidence that the VIT modules, designed for learning concepts underpinning bootstrap and randomization-test inference, promote the formation of concepts.

METHOD

This study is part of a large collaborative research project involving 33 team members and over 2000 first-year university students. Over 200 students volunteered to be interviewed and

therefore two students from each of the six classes were randomly selected. Four of the 12 interviewees were given a task to complete, two of which are reported in this paper. The students whose interviews are analyzed would have been introduced to confidence intervals using a classical mathematical approach within a three-week teaching module the year before at school. They had no previous experience of the bootstrap method or experimental design and the randomization test. Their prior experience is typical of about 60% of the cohort. In four 50-minute lectures in a class of about 450 students they were introduced to the bootstrap method and the randomization test. The lectures included hands-on activities, which occurred before lecturer demonstrations of the VIT modules. Students used the VIT modules in an assignment. They were interviewed about one week after they had handed in their assignments. The interviews were video-recorded and later transcribed for analysis. The analysis was conducted by the authors and then validated by Mike Thomas, the creator of the versatile thinking framework and Table 1.

The VIT modules have three dynamically linked screens (see Figures 1 and 2), each of which can be viewed one at a time, sequentially, in order that the ideas behind bootstrap and randomization-test inference are gradually introduced and reinforced (see Budgett, Pfannkuch, Wild, and Regan, (2013) for further detail). The dynamic visualizations use a myriad of representations such as dot plots, numeric values, a variation-band for re-sampled medians, and a distribution of re-sampled medians/re-randomized differences in the means. This should encourage students to work and think seamlessly within and between representations, to switch between a perception of a statistical entity as a process or an object, and to exploit the power of visual schemas by linking them to relevant logico/analytic schemas. In other words, the VIT modules should encourage versatile thinking and hence conceptual understanding.

One student, S1, was given raw data on survival times in days for 20 randomly sampled guinea pigs who were given a drug, and was asked to find the typical survival time (Figure 1). The other student, S2, was given raw data on decrease in blood pressure for 21 black male volunteers with high blood pressure who were randomly assigned to either the calcium tablet group or the placebo group, and was asked whether the added calcium intake reduced blood pressure (Figure 2). Both tasks were adapted from Hesterberg, Moore, Monaghan, Clipson, and Epstein (2009). Using the appropriate VIT module on a provided computer the students were asked to teach the interviewer how to carry out such analyses.

RESULTS

Using Table 1 as a tool we will first analyze S1's interactions as he constructs a bootstrap confidence interval and then S2's interactions as he constructs a tail proportion. The coding from Table 1 is included in order to make transparent how the data were mapped against this framework.

The Bootstrap Task

S1 entered the guinea pig survival time in days into a spreadsheet and then chose the bootstrap construction module from a suite of modules. That is, he recognized the problem scenario and chose the correct method for analysis, carrying out a procedural action on the spreadsheet representation (PRT): "basically what we want to do is bootstrap confidence interval construction." After some time thinking and exploring he decided that the median would be the appropriate quantity of interest.

Once he obtained the plot of the sample on which the median was plotted (see top screen of Figure 1) he continued:

So you've got a median here and that's our sample. So we've got the 20 survival times of guinea pigs and what we're trying to do now is consider this sample as the population and now we're going to sample with replacement from that population and so basically iNZight [VIT] does that for us so we don't have to spend years doing this ... you can see it's sampling on the population with replacement, establishing the plot diagram down here.

He recognized the sample plot as an object or entity and attributed to it a property (OPO) by reconceiving the sample plot as though it were the population distribution, a deep property conceptualization. With this perspective on the image he is able to use the software to act on the representation as a conceptual object (COT), taking multiple samples with replacement in order to plot a band of re-sample medians (see middle screen Figure 1).

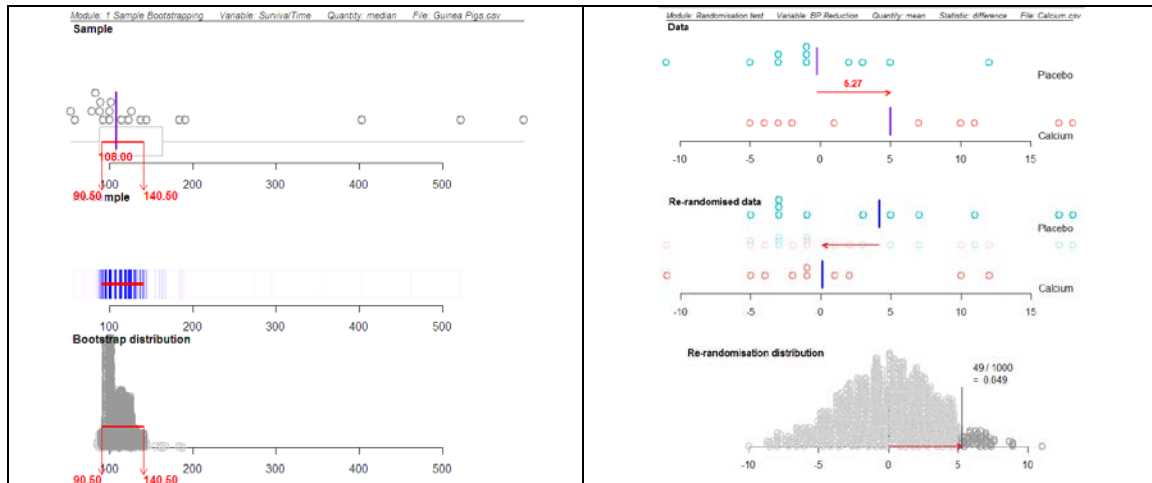


Figure 1. VIT module showing bootstrap confidence interval for the population median survival time (days) of guinea pigs

Figure 2. VIT randomization-test module showing tail proportion for difference between blood pressure mean reductions

S1 then put the third screen in operation so the three screens were dynamically linked (see Figure 1). Before doing so he remarked, “if we want a good distribution we want to go to a thousand” indicating a conceptual overview of the process of the distribution of re-sample medians (PPO) that would enable him to act on it in a conceptual manner (CPT). He was able to work seamlessly between several representations such as the vertical bar and the dot representing the median as he said, “below that we just see the plotted version of that [re-sample median] instead ... yeah, differently plotted.” He was able to switch between a perception of a statistical entity as an object and a process, as seen in his comment: “And so now you can see the bootstrap distribution [OPO] being constructed as the population is re-sampled [PPO], the population is sampled a thousand times, right?”

Once the bootstrap distribution was built up he not only recognized it as an object in its own right but also realized that he would need to operate on this conceptual object (COT) to “show the confidence interval” (see bottom screen Figure 1). He continued: “So this confidence interval contains what I was saying before, the majority of the re-samples ... the majority of the medians, re-sampled medians lie between 90.5 and 140.5.” Here he demonstrates the ability to extract a property from this object, using an OPO conceptual view. He then guessed that the majority meant 95%. On the display the lines representing the confidence interval then dynamically move up and are placed on the middle and top screens (see Figure 1). Using a PPO he noted this process:

So basically it shows the confidence interval on this plot instead and as we move up we see the confidence interval here. You can see that the actual sample, the sample of the actual population, that the median lies in between that [the confidence interval representation].

Finally he is able to interpret the confidence interval representation as an object in its own right and infer a property from it, which is an OPO:

Once we have this confidence interval established we can go back and do some statistical inference and say that it is highly likely that the population median of survival time after the guinea pigs undergo treatment will lie between 90.5 days and 140.5.

Using the terminology *highly likely* is not statistically correct, however, we note that S1 recognized that he was making an inference about the population from a sample.

The Randomization-test Task

S2’s approach to the calcium/placebo task was to use the bootstrap confidence interval module and if that did not work, “I’d have to go back and do randomization.” His mode of operation with the raw data representation was very much a procedural surface observation (PRSO) with his action on the choice of module being a procedure (PRT). This initial procedural view was further reinforced when he entered the data into the spreadsheet. The data for the task were presented to S2 in two columns. Consequently, he used the same data structure for the spreadsheet,

which resulted in the module not working. It should be noted that the students had not experienced entering such data in a spreadsheet as in the assignment the spreadsheet of data was given to them. Before S2 sorted out what he should be doing he stated that the randomization test:

Gives you a mean, so if you have a mean from the original, like say you have a mean difference in those on the original data, you could then generate a tail proportion which tells you how likely that mean is to eventuate when chance is acting alone. So it [the test] resamples but it takes away any label on it [the data], so it takes away the calcium or the placebo group [label] and then resamples again.

His statement using language such as *generate* and *resamples again* indicate he has a conceptual overview of the process that would enable him to act on the VIT module in a conceptual manner (CPT). He also seems to have the ability to link his visual schema to the relevant analytical schema. He incorrectly uses *resamples again*, language for the bootstrap module, rather than re-randomizes or re-assigns to the two groups.

When S2 put the top screen of Figure 2 into action he was able to interpret the data representation, “it gives a mean of both the placebo and calcium groups and shows the calcium group has a greater mean decrease,” an object property observation (OPO). Similarly he was able to interpret the changing red arrows in the middle screen of Figure 2: “It’s giving them no labels and then re-randomizing them. So for each mean it is generating, it’s showing whether they [the group means] are larger or lower.” At this stage he was asked how this procedure helped to answer the research question to which he responded:

It’s taken a sample but you do not know whether it is a chance thing looking at the original data. Then it re-randomizes the data giving you a set of data and in that way you’ll be able to see if chance is acting alone. So in the re-randomized data if you had that sort of margin [difference in means in original data] quite frequently you’d say that chance could be acting alone and the tail proportion would be high.

Again S2 demonstrates he has a conceptual overview of the process enabling him to act on the representations in a conceptual way (CPT). S2 now put the third screen (Figure 2) into operation and the following conversation took place with the interviewer (I) as the re-randomization distribution built up and when he operated on this distribution conceptual object (COT) to show the tail proportion. Note how he works seamlessly within and between the representations.

S2: *So it [the middle screen means] will go down and show the full set of re-randomized data means.*

I: *What are those things building up here in the bottom?*

S2: *The difference in means between the re-randomized data so it [the distribution] is going higher (PPO), a bell curve (OPO).*

I: *Why are you looking at the difference in means?*

S2: *To see if the difference in the original data is by chance or not by chance. So that’s showing the likelihood of getting the original difference in means, so 49 chances out of 1000 you generate a difference in means equal to or larger than the original (OPO).*

I: *Why are you interested in that?*

S2: *To see if calcium is effective as a way of treating blood pressure and whether it actually causes any difference. ... It’s very likely that the calcium is the cause of the difference (OPO).*

I: *What would you say if the tail proportion was 34%?*

S2: *It’s plausible that it could not be due to calcium. It could be due to a number of other factors, just chance acting alone (OPO).*

I: *So can you say what it is due to?*

S2: *You cannot (OPO) because it’s not below 10% (PRPO).*

It is noteworthy that at this point S2 now correctly identifies the statistic being generated and in the original data when he specifically refers to the *difference in means*. As the re-randomization distribution is built up he observes that a property of the process (PPO) will result in similarities to the normal distribution (OPO) and switches between a perception of a statistical entity as a process or an object. Using an OPO conceptual view he extracts properties from the re-randomization distribution object by comparing the original difference in the means or greater to the difference in means generated under chance acting alone. He then reveals deep property observations from the tail proportion object: causation when the tail proportion is small, and that he is unable to make a conclusion when the tail proportion is large, at which stage he recalls the

procedural property observation (PRPO) that a tail proportion of more than 10% (a guideline provided in teaching) does not allow any causes including chance to be offered as an explanation.

The analysis of these students' interactions with the VIT representations indicates that they were often employing *a conceptual view of the processes or objects* represented on the screen (a PPO or OPO), that is, making deep property observations. They were also able to act on some of the objects represented (COT) during their knowledge construction. Therefore, indirectly, the evidence is that such thinking might be leading towards the formation of conceptual understanding.

CONCLUSION

Versatile thinking is characteristic of a person who has developed rich mathematical or statistical schemas (Thomas, 2008). Our claim is that these students, through their explanations of how a bootstrap confidence interval/tail proportion is constructed, seemed to demonstrate the three elements of versatile thinking and hence indirectly we conjecture that they were developing rich statistical schema, including conceptual understanding of inference. Furthermore, our theoretical analysis of the students' explanations using Table 1, a representational versatility perspective, showed how the students interacted conceptually with the dynamic representations. It should be noted that our analysis only demonstrates the nature of the students' interactions with the dynamic external representations and not the effect on their cognitive structures. Such interactions, however, seemed to suggest the VIT modules assist in promoting in the students the development of rich statistical inference schema. The simulations seem to allow students to visualize and experience sampling variability/random re-allocation to groups and how the bootstrap/re-randomization distribution evolves. We conjecture that the students did have a visual schema for these concepts, as they seem to be able to visualize what would occur before they acted on the representations. Hesterberg (2006), Cobb (2007) and Engel's (2010) claims that empirical simulations promote students' conceptual understanding of the logic of inference seem to be indirectly borne out by our analysis of these students' interactions with the dynamic visualizations of the VIT modules. Further research, however, is needed to validate these findings.

ACKNOWLEDGEMENTS

Work in this paper is supported in part by a grant from the Teaching and Learning Research Initiative (http://www.tlri.org.nz/sites/default/files/projects/9295_summary%20report.pdf).

REFERENCES

- Budgett, S., Pfannkuch, M., Regan, M., & Wild, C. J. (2013). Dynamic visualizations and the randomization test. *Technology Innovations in Statistics Education*, 7(2), <http://escholarship.org/uc/item/9dg6h7wb>
- Cobb, G. (2007). The introductory statistics course: A Ptolemaic curriculum? *Technology Innovations in Statistics Education*, 1(1), 1-15.
- Engel, J. (2010). On teaching bootstrap confidence intervals. In C. Reading (Ed.), *Proceedings of the Eighth International Conference on Teaching Statistics*. Voorburg, The Netherlands: International Statistical Institute.
- Graham, A., Pfannkuch, M., & Thomas, M. O. J. (2009). Versatile thinking and the learning of statistical concepts. *ZDM – International Journal on Mathematics Education*, 41(5), 681-695.
- Gray, E. & Tall, D. O. (1994). Duality, ambiguity & flexibility: A proceptual view of simple arithmetic. *Journal for Research in Mathematics Education*, 26, 115-141.
- Hesterberg, T. (2006). Bootstrapping students' understanding of statistical concepts. In G. Burrill (Ed.), *Thinking and reasoning with data and chance. Sixty-eighth National Council of Teachers of Mathematics Yearbook* (pp. 391-416). Reston, VA: NCTM
- Hesterberg, T., Moore, D., Monaghan, S., Clipson, A., & Epstein, R. (2009). Bootstrap methods and permutation tests. In D. Moore, G. McCabe & B. Craig (Eds.), *Introduction to the practice of statistics* (6th edition) (pp. 16-1-16-60). New York: Freeman.
- Thomas, M. O. J. (2008). Conceptual representations and versatile mathematical thinking. *Proceedings of ICME-10* (CD version of proceedings). Copenhagen, Denmark, 1-18. http://www.icme10.dk/proceedings/pages/regular_pdf/RL_Mike_Thomas.pdf