MODEL-BASED INFORMAL INFERENCE

Richard Lehrer[1], Ryan Seth Jones[1], and Min-Joung Kim[2]
[1]Vanderbilt University's Peabody College USA
[2]Louisiana State University USA
rich.lehrer@vanderbilt.edu

Informal inference refers to making decisions informed by conceptions and representations of sampling variability without recourse to theorems governing formal approaches. To support informal inference, grade 6 middle-school students invented and revised models of chance processes they experienced first-hand by generating repeated measures. These processes provided an accessible interpretation of variability as composed of fixed (signal) and random (error) components. Conducting an analysis of variance and then creating random device analogs of these sources informed model building. Students then generated empirical sampling distributions of model parameters to assist model test and revision. Students used these sampling distributions to guide informal inference. We report on the intelligibility of this form of model-based inference as indicated by results of a flexible interview.

INTRODUCTION

Statistical practice is aimed at making inference. This truism is nonetheless difficult to secure for learners, partly because ideas like probability density and the logic of hypothesis testing are notoriously challenging. An alternative approach, informal inference, eschews these formalisms in order to provide entrée to learners in ways that are consistent with statistical practices of inference (Makar & Rubin, 2009; Pratt & Ainley, 2008). Yet the intention to support learning about the grounds of inference informally raises uncertainty about productive means to do so (Pfannkuch, 2011).

Our approach to informal inference engages students in one of the foundations of formal inference, modeling chance, in ways that we conjecture are accessible to comparatively young students in the early middle school years. Rather than engaging students in probability distributions as models of chance, or in re-sampling as more accessible versions of these distributions, students invent and revise models of chance processes they experience first-hand by generating repeated measures or by producing a product. The longer progression of learning that situates this form of model-based inference is summarized in Lehrer, Kim, Ayers, and Wilson (2014) and is founded in broader approach, data modeling, that aims to re-integrate data and chance throughout K-8 schooling (English, 2012; Lehrer & Romberg, 1996). Here we sketch the trajectory of learning that culminates in model-based inference:

- Students measure the same attribute of an object or produce something. The intention is to foster an image of repeated process (Saldanha & Thompson, 2003) and to coordinate this image with a collection. Collectively, these are seeds of distribution.
- Students invent and compare visualizations of the resulting data. The intention is to foster an image of shape of the data as a result of designers' choices about how to structure the data (e.g., order cases, create intervals).
- Students invent and compare measures that estimate true measure (measurement process) or target value (production process). These inventions and comparisons develop statistics as measures of signal and noise (Konold & Pollatsek, 2002; Lehrer, Kim, & Jones, 2011; Petrosino, Lehrer, & Schauble, 2003).
- Students measure uncertainty in the behavior of visible, culturally available random devices, such as spinners or dice that are familiar to them from games of chance. These informal seeds are developed to encompass sampling distribution, and statistics are re-purposed as descriptions of a sampling distribution.
- Students re-visit the processes of measure and production by analyzing sources of error. Different sources of error are modeled by the random devices previously investigated and observed measures or products are represented as a combination of signal (e.g. sample median) and noise.

In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9, July, 2014), Flagstaff, Arizona, USA.* Voorburg, The Netherlands: International Statistical Institute. iase-web.org

- Students investigate model fit, using student-invented criteria such as approximations to shape of the data, the center of the data, and the spread of the data. Sampling distributions of model parameters inform model fit.
- Students make informal inference about new sample statistics in light of their belief about the adequacy of their model. Ideally, this involves use of the sampling distribution of model parameters as a guide to probabilistic inference.

The trajectory of learning that culminates in model-based inference requires coordination among concepts and practices that often develop over a considerable period of time (Pfannkuch & Wild, 2000). The extent to which these concepts and practices are intelligible to students remain an open question in our research. To address this question, we conducted a design experiment (Cobb, Confrey, diSessa, Lehrer, & Schauble, 2003) involving four classes of sixth-grade students and their teacher. The design experiment instantiated the learning trajectory described above. Before and after completion of instruction, we administered and scored student responses to a paper-and-pencil assessment instrument. The latter indicated clear gains in student knowledge along six dimensions of learning (Lehrer et al., in press). To provide a finer-grained portrait of student understandings of model-based inference, we conducted individual flexible interviews that serve as the primary source for the results and conclusions about model-based inference developed here.

METHOD

Participants

We selected a convenience sample of 13 grade-6 students. Interviews were conducted primarily by Kim and ranged from 30 minutes to 58 minutes.

Structure and Intentions of the Flexible Interview

The flexible interview was composed of four scenarios intended to elicit students' conceptions of (a) the nature of chance and probability, (b) characteristics and effects of changes in sample size on sampling distributions, (c) criteria for evaluating the goodness of fit for a model of variability composed as fixed signal and random noise, (d) revising models to change parameters to produce different distributions, and (e) criteria guiding informal inference in a framework of models of a novel production process for batteries.

The first scenario asked students to interpret the meaning of "there is a 40% chance of rain tomorrow. How might the forecaster have come up with 40%?" The scenario indicated the incidence of outcome-based understandings of probability in which 40% could be interpreted as certainty about tomorrow's weather (Konold, 1989). In contrast, images of probability as statements about long-term processes would be indicated by students' reference imagined accumulated records of past events or forecasting models based on a similar interpretation. The second scenario presented a hidden mystery 2-color (red-blue) spinner in TinkerPlots (Konold & Miller, 2005). We asked the students to speculate about the structure's spinner and if uncertain, to suggest the number of times that it should be run, and to explain why. The intention was to probe students' understanding of the value of a larger sample for inferring structure of the device. We did not allow the student to run the spinner a large number of times, but instead ran it once 10 times, again eliciting students' estimates of the structure of the spinner. Students suggested the number of samples of size 10 that would help make prediction more stable. The purpose was to probe students' anticipation of sample-to-sample variation. We elicited how one would use TinkerPlots to obtain a sampling distribution. Students either spontaneously generated an empirical sampling distribution or were assisted to do so. They then responded to questions about the meaning of a case and the meaning of statistics of center and variability of a sampling distribution. Students predicted the effects of change in sample size on the center and variability of the sampling distribution.

The third scenario addressed students' conceptions of model fit. Students viewed a signal-noise model of the measurement of the circumference of one person's head and were asked to judge its adequacy in light of a single run of the model. If students suggested running the model repeatedly, they saw empirical sampling distributions of the median and IQR, and were then probed about their assessment of model fit. The intention of this item was to elicit student criteria for model fit and the extent to which sampling variability was coordinated with model fitting. Students were challenged to revise the model presented to change the shape of the data from Gaussian to uniform or to U-shape with same median and range. The intention was to probe students' understandings of the coordination between model structure and resulting distribution of outcomes.

The fourth scenario addressed students' conceptions of model-based inference. Students first interpreted the meaning of a display and statistics of center and variability in a test of battery life. The intention was to probe students' conceptions of statistics as measures of central tendency and consistency of battery life. Students viewed a model of the production process with sources of variability of battery life, and the model-based sampling distributions of median and IQR. In light of these sources of information, students inferred whether or not a new sample of batteries made with a new production method made a difference. We recorded student justifications with an eye toward their use of characteristics of the sampling distributions to guide inference.

Coding of Student Responses

Two of us (Kim and Lehrer) used a method of constant comparative analysis to analyze a few cases of student responses informed by the expectations that guided our construction of the interview and generated a set of codes. We worked collaboratively to try out the codes on the larger sample, and as we discovered exceptions or the need for further differentiation, we revised and refined the coding scheme. In this analysis, we collapse codes to develop a more general portrait of student conceptions.

RESULTS

Our interpretation of student responses is summarized according the core disciplinary concepts elicited across the four scenarios.

Interpretations of Probability

All students (100%) interpreted the statement about the likelihood of rain as 40% to refer not to an outcome but instead as an indicator of uncertainty. Most (77%, 10 students) made explicit reference to past records of rainfall under similar conditions. The remaining students cited factors contributing to the likelihood of rain in weather forecasting models, such as the percent cloud cover. In the context of inferring the structure of a chance device, the mystery spinner, three students ventured a guess about its structure (e.g., 50-50 "cause its like typical"), but most students (77%, 10 students) indicated that they could not know. All students seemed to anticipate sample-to-sample variability, and most (85%, 11 students) suggested that inference about the structure of the spinner would require a large number of trials (e.g., 100 or more). Two students believed that 20 or fewer spins would suffice. Students' rationales generally drew upon their experiences with random devices. For example, "If I just run it like ten times, anything could happen in those ten times, but if I run it a hundred times, then you have more precise of a data…probably see what it's gonna do." Another student referred to an iterative process: "I like to start low because then it's easier to look at, um, the data you have and then I like to get larger from there."

Empirical Sampling Distribution

Most students identified each case of a sampling distribution with a sample (77%), but a few (n =3) were uncertain, slipping between a sample size (10) and the number of samples (100) in the second scenario. For example, one student gestured toward a case and mentioned that it represented 10 spins and then informed the interviewer that the statistic of percent blue represented "20% out of 100." A few turns later, the same student mentioned that 20% represented 2 outcomes of one color. Seven students (54%) associated the median statistic of the sampling distribution with the structure of the spinner, with the remaining students using the center clump to infer a small range of potential structures. All students forecast that a change in sample size (from 10 to 50 repetitions) would not affect the median. However, only four students (31%) forecast a decrease in sampling variability with an increase in sample size, reasoning that more extreme percentages would be "a lot harder" to obtain. Four other students predicted an increase in sampling variability due to "more chances." The remaining five students (38%) forecast no substantive change beyond what one might expect from sample-to-sample variability (e.g., “except there might be a few more.”).

Models, Model Fit and Informal Inference

In the scenario involving production of batteries, most students (92%) interpreted median battery life as an average battery life and the IQR as a measure of consistency. Many offered additional contextualization to make sense of the situation: “I guess they don’t come out the same energy in all the batteries.” Nearly all students (92%) modified model parameters appropriately to generate uniform or U-shaped distributions, thus coordinating the structure of the model to the resulting distribution of outcomes. Most (77%) judged model fit by comparing the output of the model statistics to the real-world sample statistics. Only three explicitly mentioned shape as a criterion for model fit. Students further judged the model as an adequate description of the process in light of the sampling distribution, especially the sampling distribution of the IQR (92%). However, most did not spontaneously generate this empirical approximation, but instead, interpreted it when we provided it. Most students then went on (85%) to judge claims about longer-lasting or more consistent batteries by contrasting sample statistics with the model-based sampling distribution.

CONCLUSIONS

The findings of this study, coupled with the results of a larger-scale assessment, collectively suggest the utility of grounding informal inference in the process of building and refining models of chance processes. We conjecture that processes involving signal and noise are apt points of entrée for several reasons. First, these processes are often tangible and hence afford students opportunities to coordinate process with a distribution of outcomes. Second, the accessibility of the process provides opportunities to conduct an analysis of sources of variability and to model the uncertainty of variability of each source. This renders the modeling inherent in formal inference accessible and visible to students, albeit in a more direct form. Coupled with sampling distributions of model parameters, students can engage in statistical inference. Of course, students in this study had not yet mastered the intricate coordination among chance, models, and inference. In this light, we consider these students as having made significant, albeit incomplete, progress in understanding the foundations of statistical, as contrasted to everyday inference. Because this research represents a comparatively early iteration in a design cycle, we anticipate that further re-design of instruction may support students to coordinate more firmly relations among models of processes and informal inferences about these processes.

REFERENCES

Cobb, P., Confrey, J., diSessa, A., Lehrer, R., & Schauble, L. (2003). Design experiments in educational research. *Educational Researcher*, *32*(1), 9-13.

English, L. D. (2012). Data modeling with first-grade students. *Educational Studies in Mathematics*, *81*, 15-30.

Konold, C., & Miller, C. D. (2005). *TinkerPlots: Dynamic data exploration* (v.2.0). Emeryville, CA: Key Curriculum.

Konold, C. (1989). Informal conceptions of probability. *Cognition & Instruction*, *6*(1), 59-98.

Konold, C., & Pollatsek, A. (2002). Data analysis as the search for signals in noisy processes. *Journal of Research in Mathematics Education*, *33*(4), 259-289.

Lehrer, R., Kim, M-J., Ayers, E., & Wilson, M. (2014). Toward establishing a learning progression to support the development of statistical reasoning. In J. Confrey and A. Maloney (Eds.), *Learning over time: Learning trajectories in mathematics education*. Charlotte, NC: Information Age Publishers.

Lehrer, R., Kim, M-J., & Jones, R. (2011). Developing conceptions of statistics by designing measures of distribution. *International Journal of Mathematics Education*, *43*(5), 723-736.

Lehrer, R., & Romberg, T. (1996). Exploring children’s data modeling. *Cognition and Instruction*, *14*(1), 69–108.

Makar, K., & Rubin, A. (2009). A framework for thinking about informal statistical inference. *Statistics Education Research Journal*, *8*(1), 82-105.

Petrosino, A., Lehrer, R., & Schauble, L. (2003). Structuring error and experimental variation as distribution in the fourth grade. *Mathematical Thinking and Learning*, *5*(2&3), 131-156.

Pfannkuch, M. (2011). The role of context in developing informal statistical inferential reasoning: A classroom study. *Mathematical Thinking and Learning*, *13*(1 &2), 27-46.

Pfannkuch, M., & Wild, C.J. (2000). Statistical thinking and statistical practice: Themes gleaned from professional statisticians. *Statistical Science*, *15*(2), 132-152

Pratt, D., & Ainley, J. (2008). Introducing the special issue of informal inferential reasoning. *Statistics Education Research Journal*, *7*(2), 3-4.

Saldanha, L., & Thompson, P. (2003). Conceptions of sample and their relationship to statistical inference. *Educational Studies in Mathematics*, *51*, 257-270.