MODEL COMPARISONS AS A MEANS OF PROVIDING AN INFORMAL QUANTITATIVE ESTIMATION FOR STATISTICAL UNCERTAINTY

Michal Dvir and Dani Ben-Zvi
LINKS I-CORE, Faculty of Education, The University of Haifa, Israel
dvirmich@gmail.com

Abstract. *A statistical inquiry is at its core the process of correlating between two worlds: the real world in which the actual phenomenon is observed and studied, and a theoretical one, probabilistic in nature, affording different types of possible explanations – statistical models that can represent the phenomenon. The objective of this study is to investigate young learners' informal statistical reasoning and the role model comparisons play as part of it. Specifically, we focus on model comparisons that occur as young learners investigate probabilistic questions that emerge while conducting a real world statistical investigation. We offer two illustrative examples showcasing how a specific type of model comparison was associated with young learners' informal attempts to quantify their levels of uncertainty. Theoretical and practical implications are suggested in conclusion.*

THEORETICAL BACKGROUND

A *model* is fundamentally a representation, usually a simplified version, constructed or being used in order to provide a description or explanation for a phenomenon (Hesse, 1962; Lehrer & Schauble, 2010). A model could be a physical object, or a theoretical entity analogous to the original phenomenon (Garfield & Ben-Zvi, 2008; Lehrer & Schauble, 2010; Wild & Pfannkuch, 1999). However, not just any representation should be considered as a model: a specific explanatory purpose for its construction is required. Furthermore, as models are not exact copies of the original represented situation, a model may prove ill-suited for the purpose of its construction (Hesse, 1962). Thus, an ongoing process of evaluation and refinement – a *modeling* process - is usually required (Lesh, Carmona, & Post, 2002).

As a model is defined by its purpose, it would be considered a *statistical model* (and its accompanying modeling process would be considered a *statistical modeling* process) if the purpose of its construction is statistical in nature. Two main characteristics are agreed upon in defining what could be considered a statistical purpose: 1) variability associated with the investigated phenomenon; and 2) probability associated with the investigated explanation (Brown & Kass, 2009; Budgett & Pfannkuch, 2015; Garfield & Ben-Zvi, 2008).

As our focus is on young learners' statistical modeling processes, there is a need for some adjustments regarding the probabilistic requirement. This is due to the fact that young learners have not yet learned and mastered its formal terminology and tools. Considering the role probability serves in a formal statistical setting, we suggest that an informal quantitative estimation of the uncertainty accompanying the statistical inquiry is a reasonable informal alternative. Thus, by "*informal quantitative estimation of the uncertainty*" we mean an attempt to measure aspects contributing to the uncertainty not by means of formal procedures and calculations.

In particular, two types of quantifiable uncertainties should be accounted for: 1) *contextual uncertainty* – stemming from the unknown behavior of the real world phenomenon that is being explored; and 2) *statistical uncertainty* - stemming from the non-deterministic nature of the statistical process and tools (Manor & Ben-Zvi, 2015). Thus, a statistical model should offer an informal quantitative estimation of both in conjunction with the explanation or description it provides.

Dvir and Ben-Zvi (forthcoming) offer an illustrative example for a manifestation of the informal alternative probabilistic requirement suggested above, associated with a specific type of *model comparison*: a comparison between two rivalling explanations for the investigated phenomenon, distinguished by their source. One, originating from a preconceived conjecture regarding the phenomenon - represented by a *Conjecture Model*, while the other originating from real data that was collected - represented by a *Data Model*. In their study, the researchers suggest that comparing the two models can promote young learners' awareness to aspects of uncertainty, leading them to offer more accurate descriptions of the phenomenon. However, as these previous findings focused on the real world aspect of young learners' investigation, only contextual uncertainty was meaningfully addressed.

To expand our understandings, we focus in this study on similar model comparisons, between Data and Conjecture Models, but ones that were conducted in a complimentary context: exploring specific probabilistic questions that arose during a real world data exploration. It is within this complimentary context that we ask: *What role can comparing between a data model and a conjecture model play in promoting students' informal quantitative estimation of statistical uncertainty?*

SETTING AND METHOD

To respond to this question, a case study was conducted as part of the Connections project - a longitudinal design and research project (beginning in 2005) aiming at promoting young learners' statistical reasoning in a technology-enhanced and inquiry-based learning environment (Gil & Ben-Zvi, 2011; Makar, Bakker, & Ben-Zvi, 2011). This case study was part of the 2016 iteration, in which a team of six researchers accompanied the participation of a whole class of sixth graders (ages 11-12), as they participated in a 19-lesson long learning trajectory.

The participants

The class of sixth graders consisted of 26 students from an elementary public school in northern Israel. This school was chosen as it has a relatively open minded approach and great willingness to cooperate with the research team and the planned learning trajectory. The same class had also participated in the previous 2015 iteration of the Connections project. After reviewing the 2016 data, two pairs were chosen as their work provided relatively clear illustrations of general commonalties found across most participating pairs. We focus here on these two pairs.

The first pair – Ami and Gil: Ami is diagnosed as a gifted child, with very high academic achievements and language skills. So much so that he is at times not understood by his peers, leading to some social difficulties. He has vast everyday knowledge but has some trouble with writing. Gil – great athlete, hard worker with relatively high achievements. He is very well loved but tends to be a bit bashful around adults and could appear a bit closed off.

The second pair – Orr and Erez: Orr has relatively high articulation skills, however is at times insecure and impulsive. He has high achievements in mathematics but has some difficulties with reading comprehension and writing. Erez is diagnosed as a gifted boy, as well as having a learning disability (dyslexia). He has highly developed reasoning skills and high achievement in mathematics, but also has some difficulties with reading and writing.

The learning trajectory

The 2016 learning trajectory was specifically designed to include both exploratory data analysis (EDA) activities (Tukey, 1977) in which young learners collected and analyzed real world data, as well as activities in which the subject of inquiry was probabilistic questions arising from the real world data exploration, according to the Integrated Modeling Approach (IMA, Manor & Ben-Zvi, in press). The probabilistic questions explored, addressed issues such as the advantages of random sampling and the relation between sample size and its representativeness.

During the learning trajectory, three full investigation cycles were conducted by the participants, working in pairs. Each cycle began in the real world: posing a research question, formulating a conjecture, planning and executing relevant data collection, and analyzing the data obtained. To provide more opportunities for statistical reasoning and modeling, the participants were then asked to informally infer onto a larger population (Makar, Bakker, & Ben-Zvi, 2011). Sample size, as well as the targeted population were gradually increased, in accordance to the growing samples heuristic (Bakker, 2004; Ben-Zvi, 2006). As inferring onto a larger population promoted the emergences of relevant probabilistic questions regarding sampling methods or sample representativeness, a new world was introduced: the probability world, or as the participants named it later on: the "tool's world."

Within the probability world, new tools were offered in order to investigate the probabilistic questions that emerged. In particular - the TinkerPlots (Konold & Miller, 2015) Sampler[1] was introduced, as a means to explore such questions by modeling a hypothetical population and drawing multiple random samples from it. Thus showcasing the distinction between the two worlds: in the probability world, contrary to the real world, the population is known and generating a sample from it is much easier. Drawing multiple samples, comparing between them and eventually constructing a sampling distribution, provided ample opportunities for exploring both the instigating probabilistic issues, as well as new ones emerging from their explorations. It is the actions and articulations of the participants during these types of explorations that this paper focuses on. Specifically, we focus on evidence for students' developing methods of estimating how probabilistic features of the statistical inquiry may affect the level of confidence in their informal inference. Thus, in the context of students' investigations in the probability world, we ask: *What role does comparing between a Data Model and a Conjecture Model, within a probability world inquiry, play in promoting students' informal quantitative estimation of statistical uncertainty?*

Data collection and Analysis

All of the students' actions and articulations were videotaped throughout the learning trajectory. Whole class discussions were captured using three video cameras, one of which was mobile. Working in pairs, Camtasia was used to simultaneously capture each pairs' articulations and gestures, and their computer screens. All artefacts created, such as drawings of conjectures, were collected as well. Field notes were also taken by a researcher during each lesson.

For both pairs, the entirety of the video data collected of their work was viewed. Any scene including the creation or employment of a visual representation was thoroughly transcribed and qualitatively analyzed according to the interpretive microanalysis approach (Siegler, 2006). Key interpretations have been triangulated by a research team that included the co-authors and another senior Connections researcher. Commonalities regarding the role model comparisons played in the modeling processes were noticed and generally described. For each recurrent role, the pairs' transcripts were then revisited with the specific purpose of exploring these commonalities. Several other pairs are currently being explored to further investigate our focus role for model comparisons conducted within the probabilistic world, and further triangulation sessions are planned to re-inspect the preliminary results.

PRELIMINARY RESULTS

The preliminary analysis of data highlighted a shared role that model comparisons held in the probabilistic investigations the students had conducted. Two examples were chosen as they best illustrate the potential connections between the type of model comparison that this article focuses on (Data vs. Conjecture Models) and emergent expressions of informal quantitative estimations of statistical uncertainty. All of the examples we provide were taken from the pairs' probabilistic inquiries, in which – using the TinkerPlots Sampler – they modeled their conjectured population and drew and explored multiple random simulated samples to investigate whether a sample size 60 could represent a population of 1300 students. The first example shows how comparing a Data Model and a Conjecture Model accompanied a process of exploring sampling variability leading to an informal quantitative estimation for its limitations. The second example shows how a comparison between a Data Model and a Conjecture Model preceded and promoted the emergence of an informal quantitative estimation for sample representativeness.

The first example, taken from the work of Ami and Gil, showcases how an initial perception of sample variability as immense and immeasurable was accompanied by articulations of a general sense of relatively low confidence in sample representativeness. Ami expresses his own sense of uncertainty, and relates it to uncertainty derived from the statistical tools: "I believe a tiny bit that it [the sample] does [represent the modeled population], almost not at all, I can't be sure". The conjecture model at that point was that "everything changes" – each sample will be completely different than the other. The visual expression of this would have been, had this conjecture been true, that the shape of the distribution of data of each sample will be completely different from the shape of any other sample. By drawing multiple samples and exploring the actual variability between them via a Data Model consisting of multiple circumventing lines (adding another circumventing line for each new sample drawn, Fig. 1), and comparing the measure of sampling variability portrayed in the data model to their conjecture of immeasurability, a new understanding was implied by the construction of a new model. Not "everything changes", but there is rather a top limit represented by dots Ami chose to add and explained: "look at the dots here [at the top of each column of the girls' distribution] – if it [data from the next sample] surpasses them, I would be surprised" (Fig. 2).

Therefore, the new model (Fig. 2) provides an estimation for the limitations of the sampling variability. The measure of variability implied by the new model could be considered an informal quantitative estimation of the contribution of sampling variability to the statistical uncertainty in this informal inferential context. As Ami began this segment articulating a connection between his perception that the sampling variability is immeasurable and his low confidence in the statistical tools, a quantitative (however informal) estimation of the first, implies a similar understanding regarding the second.

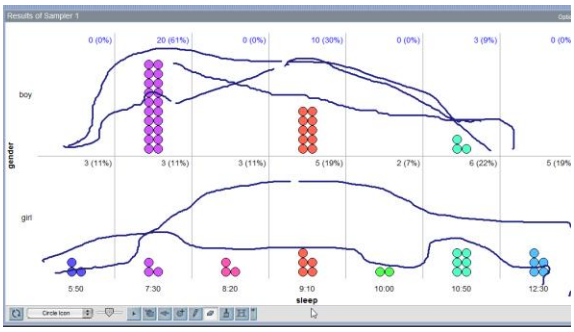

Figure 1: Ami and Gil's Data Model consisting of multiple circumventing lines for each former sample distributions.

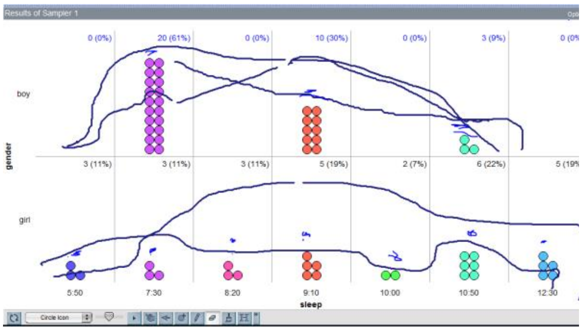

Figure 2: Ami and Gil's inferred limitations of the sampling variability represented by dots added to the top of each non empty column.

In the second example, taken from the work of Orr and Erez, the construction of the pair's model for an informal confidence interval indicating samples whose means were considered "okay" (Fig. 3), although somewhat guided by the accompanying researcher, was a natural progression of the pair's investigation. This model (Fig. 3) served as an informal inference from the pair's prior inquiry. In particular, while investigating a sampling distribution, the pair had articulated a distinction between two types of simulated samples: samples they had initially considered "trustworthy", indicated by a circle Erez hand drew circumventing the middle column (Fig. 4) indicating the pairs' conjecture, and those they had considered too "spread out" (as can be seen in the "pyramid" shape, as Orr named it, Erez had hand drew to indicate what they see in the data, Fig 4.). This distinction seemed to be the prime motivation to further explore the notion of trustworthiness, as Erez explained: "I think that because it [the range of the sampling distribution] is so spread out so I can trust it [a sample] less". This created a climate in which the researcher's following suggestion – compare the mean of the sampling distribution with the mean of the real sample and identify which simulated samples are "reasonable" - was accepted and meaningfully understood. The result of which was the pairs' new hand drawn line circumventing the three middle columns in the sampling distribution to indicate the "reasonable" samples (those with a mean in the range 40-55, Fig. 3), a new model they constructed with a similar purpose of a (informal) confidence interval. Thus the model comparison from which this distinction was drawn – the Data Model (the pyramid) and the Conjecture Model (the circle) (Fig. 4), seemed to serve here as a catalyst for the guided emergence of a new tool, an informal confidence interval. The new tool provides an informal quantitative estimation, in this example even more clearly quantitative (40-55) of a key aspect of the statistical uncertainty associated with inferring from a sample: reasonable deviation of a sample's mean (in the case – the mean of a sub group of a sample). After its construction the researcher asked the pair regarding the meaning of the new drawing (Fig. 3), to which Erez replied:

Erez	It [the circumventing line, Fig 3.] means that we can trust 71%.
Researcher	So there is no need to have a bigger sample, then?
Erez	Yes. It [the mean in 71% of the samples] is relatively close to reality [the mean of the real sample]

Note that the sum of percentages of data in the three middle columns is 17%+ 31% +23%=71%, showing how Erez directly connects the new model to his "trust", or confidence in the ability of sample size 60 to represent the population and provides a quantitative estimation: "71%".

Both of the examples described above, share a key action: comparing a Conjecture Model representing the pair's expectation regarding an aspect of the simulated samples' behavior (variability of their distribution and variability of their means) with a Data Model representing the actual behavior observed while drawing multiple simulated samples. In both, this comparison resulted in the construction of a new model, which seems to have a slightly different purpose: describe a quantitative, albeit informal, estimation of the explored variability. More explicitly in the second example, this informal quantitative estimation of variability, was also extended to describe a similar estimation of the level of statistical uncertainty each pair had associated with the form of variability they were investigating.

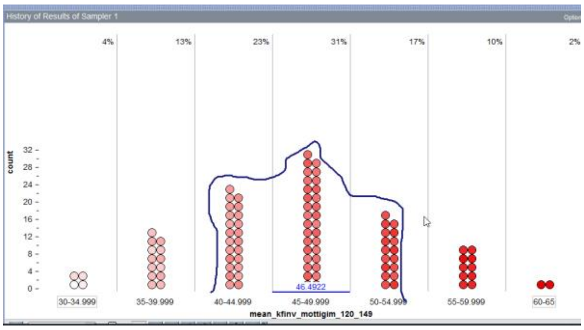

Figure 3: Orr and Erez' informal confidence interval.

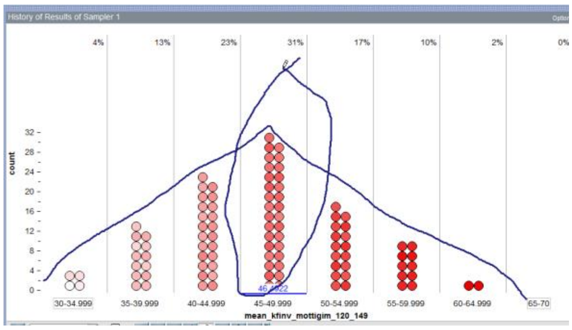

Figure 4: The Conjecture Model (the circle) placed on top of the Data Model (the pyramid).

SUGGESTED IMPLICATIONS

The role of comparisons between Data and Conjecture Models has been previously described as a catalyst for the informal statistical modeling process (Dvir & Ben-Zvi, forthcoming). However, illustrations provided for this claim were limited to real world investigations. Therefore, certain aspects distinguishing statistical modeling and the unique reasoning it requires from general modeling endeavours had not yet been thoroughly discussed. Specifically, while within the compound of a real world investigation *contextual* uncertainty can be explored, and informally quantitatively estimated, via model comparison, fewer opportunities for deep explorations of issues concerning *statistical* uncertainty are provided in the current literature.

Thus, our current interest is to go beyond such limitations, stepping out of the realm of a real world investigation, alongside our young students' further explorations of probabilistic questions that their real world inquiry had provoked. In doing so, it is our claim that the role of model comparison as a driving force can be expanded: we now consider it as a catalyst of the statistical modeling process during *both* a real world investigation as well as during a probabilistic investigation. The details may vary: the conjecture may be regarding a real world phenomenon or a probabilistic one, and the data being explored may be a real sample or multiple simulated samples. However, in both worlds– comparing a preconceived conjecture (or one that emerges as the investigation progresses) about the phenomenon that is explored, with information received by suitable data, could promote the emergence of informal quantitative estimations of different aspects of the uncertainty accompanying the statistical inferential process. Even more so when both models (Data, Conjecture) are portrayed visually one on top of the other (as in Fig. 4).

The expanded role of model comparisons, and further insight regarding the reasoning processes accompanying such comparisons, lead us to suggest the following pedagogical implications.

1. Incorporating probabilistic investigations with data exploration activities can provide more opportunities for young learners to explore and develop means of informally estimating aspects contributing to statistical uncertainty.
2. Specifically directing young learners to articulate and visually represent their underlying conjectures as well as explicitly promoting comparing these with the relevant data – whether during a real world investigation, or a probabilistic one – could further their current inquiry.

3. Through the course of deep exploration of both worlds' phenomenon – real and probabilistic – complex statistical concepts can be introduced and explored, such as sampling variability and sampling distribution

Our findings suggest that inclusion of probabilistic inquiries is beneficial in elucidating complex but vital aspects of statistical reasoning. Further research on the role of model comparisons, in the context of informal inference and in other statistical contexts should provide more applicable insights both in gaining a deeper understanding of young learners reasoning with statistical models and modelling, as well as for future designs of learning environments to promote it.

NOTES

[1] With the Sampler engine, students can design and run probability simulations, then plot the results to give a visual representation of the outcomes over many samples.

REFERENCES

Bakker, A. (2004). *Design research in statistics education: On symbolizing and computer tools*. Utrecht, the Netherlands: CD Beta Press.

Ben-Zvi, D. (2006). Scaffolding students' informal inference and argumentation. In A. Rossman & B. Chance (Eds.), *Proceedings of the Seventh International Conference on Teaching of Statistics*. Voorburg, The Netherlands: International Statistical Institute.

Brown, E. N., & Kass, R. E. (2009). What is statistics? *The American Statistician*, *63*(2), 105-123.

Budgett, S., & Pfannkuch, M. (2015). Building conditional probability concepts through reasoning from an eikosogram model: A pilot study. In *Proceedings of the Ninth International Research Forum on Statistical Reasoning, Thinking and Literacy* (pp. 10-23). Paderborn, Germany: University of Paderborn.

Dvir, M., & Ben-Zvi, D. (2018). *The role of model comparison in young learners' reasoning with statistical models and modeling*. Manuscript submitted for publication.

Hesse, M. B. (1962). *Forces and fields. The concept of action at a distance in the history of physics*. Mineola, NY: Dover.

Garfield, J., & Ben-Zvi, D. (2008). *Developing Students' Statistical Reasoning: Connecting Research and Teaching Practice*. New York City, New York: Springer.

Gil, E., & Ben-Zvi, D. (2011). Explanations and context in the emergence of students' Informal Inferential Reasoning. *Mathematical Thinking and Learning, 13*(1&2), 87-108.

Konold, C., & Miller, C. (2015). *TinkerPlots* (Version 2.3.1) [Computer software]. University at Massachusetts. Online: http://www.srri.umass.edu/tinkerplots.

Lehrer, R., & Schauble, L. (2010). What kind of explanation is a model? In M.K. Stein (Ed.), *Instructional Explanations in the Disciplines* (pp. 9-22). New York: Springer.

Lesh, R., Carmona, G., & Post, T. (2002). Models and modeling: Representational fluency. In D. Mewborn, P. Sztajn, D. White, H. Wiegel, L. Bryant, & K. Nooney (Ed.), *Proceedings of the 24th Annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (Vol. 1, pp. 89-98). Columbus, OH: ERIC Clearinghouse for Science, Mathematics, and Environmental Education.

Makar, K., Bakker, A., & Ben-Zvi, D. (2011). The reasoning behind informal statistical inference. *Mathematical Thinking and Learning, 13*(1-2), 152-173.

Manor, H., & Ben-Zvi, D. (2015). Students' emergent articulations of models and modeling in making informal statistical inferences. In *Proceedings of the Ninth International Research Forum on Statistical Reasoning, Thinking and Literacy* (pp. 107-117). Paderborn, Germany: University of Paderborn.

Manor, H., & Ben-Zvi, D. (in press). Students' emergent articulations of statistical models and modeling in making informal statistical inferences. *Statistics Education Research Journal.*

Siegler, R. S. (2006). Microgenetic analyses of learning. In W. Damon & R.M. Lerner (Series Eds.) & D. Kuhn & R.S. Siegler (Vol. Eds.), *Handbook of child psychology: Volume 2: Cognition, perception, and language* (6th ed., pp. 464–510). Hoboken, NJ: Wiley.

Tukey, J. (1977). *Exploratory data analysis.* Reading, MA: Addison-Wesley.

Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry (with discussion papers). *International Statistical Review, 67*(3), 223-265.