STRUCTURING CONTEXTS FOR STATISTICAL TREATMENT: INITIALIZING STATISTICAL REASONING

Erna Lampen
School of Education, University of the Witwatersrand, South Africa
Christine.lampen@wits.ac.za

Teaching from everyday contexts drew my attention to the importance of learning to structure situations for statistical treatment. In this theoretical paper I argue that Mallows' notion of solving a zeroth problem as part of the data-handling process should receive priority in introductory statistics courses for teachers who learn statistics through solving authentic problems. I discuss an example of such a structuring process and reflect on the demands it posed for the students and for me as lecturer.

EVERYDAY CONTEXTS IN THE STATISTICS CLASSROOM

Context is defined as "the real-world phenomena, settings, or conditions from which data are drawn or about which data pertain"(Langrall, Nisbet & Mooney, 2006). As such, contexts are unstructured, since they "have vaguely defined or unclear goals and unstated constraints, require learners to make judgments about the problem and defend them often by expressing personal opinions or beliefs about the problem interpretation and represent uncertainty about which concepts, rules, and principles are necessary for the solution or how they are organized" (Shin, Jonassen & McGee, 2003) An example of an unstructured problem in my statistics course centered on the question "What is a reasonable price to pay for a used car?" Instead of giving the students a data set to work with, I decided that as future teachers of statistics they have to learn to make decisions about the data they will need and where to source such data. When I posed this problem the first time, I thought I was positioning the class at the start of a data-handling cycle: *Step 1: Ask the question* and I was ready to continue with *Step 2: Generate data*. In stead I found us drawn into the complexity of what it means in context for individuals to answer such a seemingly simple question. The context is structured differently by students when the focus is on buying a car for themselves, than when considering advising someone or disinterestedly describing what is available in the market. We were drawn back into context to a more fundamental *Step 0*.

THE ZEROTH PROBLEM IN THE DATA-HANDLING PROCESS

Mallows (1998) draws our attention to the context when he motivates attention to the Zeroth problem:

> …namely deciding what the relevant population is, what the relevant data are, and just how these relate to the purpose of the study (There is an even earlier problem: choosing what problem to study!)

He continues to frame the Zeroth Problem as follows:

> Problem (0): Considering the relevance of the observed data, and other data that might be observed to the substantive problem.

Mallows (1998) uses the term "data" in the wider sense of "information about the context". He further explains the relevance of context to the statistical process and indicates the importance of taking account of everyday reasoning in context.

> Before we see any data at all, if we understand even partially what the problem is about, we must have some idea as to what the data would be like. We must have knowledge of concomitant variables, these are indeed what define the problem! So we are always in a position of having some "knowledge" (which we should recognize might be in error). But we should always leave ourselves open to what the next stage of data-gathering and data-analysis will teach us…The formulation of clean questions is often an important part of an inquiry. (Mallows, 1998, p. 3).

It transpired that I didn't pose a "clean question" and that my understanding of "the concomitant variables" was not the same as that of the students. Nor did my students and I have a shared focus - while my viewpoint was already informed by the norms, processes and tools of statistics, my students' reasoning was firmly grounded in everyday concerns. Mallow's discussion of the relationship between context and statistics considering the Zeroth Problem also problematizes the

In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics (ICOTS8, July, 2010), Ljubljana, Slovenia*. Voorburg, The Netherlands: International Statistical Institute. www.stat.auckland.ac.nz/~iase/publications.php

relationship between descriptive reasoning, guided by intuition and context knowledge and normative reasoning guided by the knowledge field of statistics. The notion of statistical cognition relates reasoning in everyday contexts and reasoning statistically with a view on teaching and learning.

STATISTICAL COGNITION

Beyth-Marom, Fidler and Cumming (2008, p. 20) created the construct "statistical cognition" to integrate research about statistical reasoning in context, research about statistical theory and its practice, and research about statistics education in an evidence based practice. I will interpret statistical cognition at classroom level to argue for its relevance as a meta-view on teaching and learning statistics from unstructured contexts. But first I will summarize Beyth-Marom et al.'s view. They call on a taxonomy of reasoning from judgment and decision-making literature in their description of statistical cognition (p. 21):

Descriptive:	(1) Decisions people make; (2) How people decide
Normative:	(1) Logically consistent decision procedures; (2) How people should decide
Prescriptive:	(1) How to help people to make good decisions; (2) How to train people to make better decisions.

The construct provides a view of three dialectically related knowledge fields: everyday life, statistics and the knowledge field of pedagogy and didactics. However, the dialectic is problematic as the facets comprise discourses that are fundamentally different in nature and often lead to conflicting problem-solving processes and irreconcilable decisions. Everyday reasoning discourses are segmentally organized (Bernstein, 1999) so that the car salesman, the prospective buyer and the journalist will produce different structurings to answer the question of a reasonable price for a used car. Pervasive reasoning fallacies related to data-based situations (Kahneman, Slovic & Tversky, 1982) are also descriptive models of everyday reasoning. An example of such a reasoning fallacy is people's use of *the law of small numbers* when they base their decisions on limited personal experience, despite overwhelming but less personal evidence to the contrary.

On the other hand normative models describe "good" reasoning based on the norms of a knowledge field such as Statistics (Baron, 1988). Good statistical thinking is based on normative models of reasoning expressed as statistical theorems and laws. *The law of large numbers* is an example of a normative model. Being scientific, these models are strongly bounded (Bernstein, 1999) and access to reasoning with them requires re-contextualization of everyday knowledge as well as knowledge of specific statistical language and processes.

Finally, prescriptive models describe how we ought to think, measured against a normative ideal. Prescriptive models are used in education and training, to try to get students to conform to "better" ways of thinking and reasoning. Theories of learning and teaching statistics belong to the prescriptive category, since teachers need to know what good statistical reasoning is, in order to "prescribe" through teaching interventions how learners ought to reason. That *statistical reasoning is based on awareness of variation and distribution of measurements* is an example of a prescriptive model, which prescribes teaching for understanding of variation as a pre-requisite for understanding statistical tools and their use.

STATISTICAL COGNITION AT CLASSROOM LEVEL

During teaching and learning of statistics, the three aspects of statistical cognition are realized in a complex process: teachers *prescribe* by making decisions about tasks and the value of student contributions based on their knowledge of the descriptive facet as well as their understanding of the normative aspect–how they understand students ought to reason about data and data-based situations. At classroom level, the descriptive facet comprises everyday contexts–where people reason about data-based situations without or with limited statistical training–and therefore students' pre-knowledge about the problem and context at hand. The statistical content that must be learned is part of the discipline of statistics and belongs to the normative facet.

The relationship between descriptive reasoning and reasoning according to the norms of statistics then constitutes decisions about structuring the everyday context to make it treatable with statistics. I will introduce a sense of direction by using D→N to represent the reasoning processes that enable access tot the knowledge field of statistics. As such, D→N constitutes aspects of “doing statistics”, such as asking meaningful questions and deciding what data to collect (Gal & Garfield, 1997, p. 5). The reverse process N→D represents the reasoning processes involved in the interpretation of data and statistical measures in context. This reverse arrow constitutes aspects of “using statistics”. In evidence based decision-making the normative facet determines the evidence (Beyth-Marom et al., p. 21), for example a correlation between two variables, but it helps little if the two variables compared are not reasonably related in context. I argue that the reasoning processes D→N, which is “stepping over” into the knowledge field of statistics, must be made explicit in teaching. Learning to reason statistically depend on careful understanding and analysis of the context into assumptions, variables with related units of measurement and conjecturing relationships between such variables. The richer the context, the more potential it has to yield a strong relationship to statistical reasoning demanded by the normative facet.

SOLVING THE ZEROTH PROBLEM TO STRUCTURE A CONTEXT

In an introductory statistics course for secondary mathematics teachers which formed part of my PhD study, I looked for answers to the pedagogical question of how to get students to consider which aspects are salient to a data-based problem when they don’t yet have an imagination of the norms of statistics. I introduced the course with an unstructured context to be structured for statistical treatment. My teaching goal was to make the structuring process explicit so that the students become aware of the choices and assumptions they make during the structuring. I will describe aspects of structuring of a context early in the course, namely the context of prices of used cars. I then will reflect on the pedagogical implications.

Owning a car is a dream of many teachers who yearn for a new car, but struggle to afford even a used car. In the first session of the course, I asked the question: What is a reasonable price to pay for a used car? This is surely a reasonable question to ask if one is looking to buy a used car, but looking at the distribution of prices for a particular kind of used car is not something one is likely to do. Yet, one could reasonably be expected to “look around” and so get a feel for asking prices. I reasoned that this was a situation with potential to use everyday reasoning in context to elicit a need for and appreciation for descriptive statistical treatment. The class then continued to structure the context as a group. The structuring proceeded in non-linear fashion, with the following processes and aspects of statistical reasoning emerging.

Orientation in context

Students scanned their personal experiences and gave pointed examples to populate the descriptive reasoning space. These examples were around naïve opinions of what “reasonable” would mean, for example, “I saw a Mercedes in the newspaper for R150 000. That’s what I think I can afford to pay for a car.” As the discussion developed the orientation became markedly more impersonal and focused on used cars as a class of objects.

Emergence of variables

Students engaged with the examples, noticing and marking alternatives which introduced variables, for example, “How old was that Mercedes?”; “You can get a brand new car for R150 000, but a small car.” A wide range of concomitant variables were introduced (make, colour of car, model, kilometre reading, radio, CD player, leather upholstery, etc.), leading to a sense of unwieldiness (too much information to deal with).

Search for systemic structure

Students searched for external solutions like the possibility of existing structures to determine prices (price fixing, book prices of new and used cars), almost as if they would like that structure to solve the problem. The notion emerged that there is variation even within systemic structures.

Search for structure among variables

Students argue about what would constitute the main variable: price or age or kilometer reading or other factors like what an individual could afford. They gradually narrowed down variables through reasoning about relationships between them. For example: “A car has a lifespan, it can do only so much kilometers, even if it goes for service regularly.” “It depends on how much the car has done, I’ll pay more for a car that is ten years old but has few kilometres.”

Making assumptions explicit

Hidden assumptions hindered the search for structure among variables, for example the notion that the lifespan of a car is more related to its age than to its kilometer reading based on the assumption that cars generally do an unstated “average” amount of kilometers in a year.

Search for data as “more examples”

Students expressed the need to match their reasoning to a larger set of examples: “Well, let’s look in the JunkMail.”

Awareness of issues around the validity of sources and sampling

As they populated the everyday reasoning space with examples of possible sources of data, students raised issues about representation and trustworthiness, for example: “Who advertises in JunkMail?”, “If you look on the internet you get the same cars on more than one website.”, “You can’t compare a Toyota to a Mercedes.”

IMPLICATIONS FOR TEACHING

According to Sfard (2000), the struggle for shared focus in mathematical (statistical) discussion is characterized by tension between the need for intuitive acceptability of arguments and operative rigour. As teacher, my participation in the discussion was from a position of statistical norms or operative rigour, while my students maintained strong focus on intuitive acceptability in terms of their everyday/ descriptive positions. It was my task to let them experience the need for a different way to reason, and to guide the discussion to “bring into being” statistical objects through the need to communicate (Sfard, 2008). It is important to note that none of these discussions were “statistical” in the normative sense. No data was available and informal reasoning in the context reigned. Yet, their everyday reasoning shows awareness of important aspects of statistical thinking: awareness of variables, the need for more information, the need to validate sources of information, the need to sample fairly and the need to control some variables while allowing others to vary randomly. A sense of the need for tools to investigate the situation was evident from the reluctance of the class to accept any specific amount as *the* reasonable price. Also evident was the need for extended general knowledge, pertaining to the context in an endeavour to search for cause-effect relationships between price and other variables. This process of structuring the context is omitted if teaching and learning starts with a prepared dataset. I had a sense that the students didn’t want to yield to data before they had done a fair amount of structuring as described above. They didn’t need any data in order to state hypotheses like “We think the price will come down if the kilometre reading increases.” Allowing the students opportunity to structure the context provided me as lecturer opportunities to mark aspects of statistical reasoning in their talk and to motivate the use of statistics to validate and quantify their informal but reasonable hypotheses. I argue that the students have started to develop a gaze (Bernstein, 1999) which can be described as statistical, through the process of structuring an everyday context for statistical treatment.

REFERENCES

Baron, J. (1988). *Thinking and Deciding*. Cambridge: Cambridge University Press.

Bernstein, B. (1999). Vertical and Horizontal Discourse:an essay. *British Journal of Sociology of Education, 20*(2), 157 - 173.

Beyth-Marom, R., Fidler, F., & Cumming, G. (2008). Statistical Cognition: Towards evidence-based practice in statistics and statistics education [Electronic Version]. *Statistics Education Research Journal*, *7*, 20-39. Online: www.stat.auckland.ac.nz/serj.

Gal, I., & Garfield, J. B. (1997). Curricular Goals and Assessment Challenges in Statistics Education. In I. Gal & J. B. Garfield (Eds.), *The Assessment Challenge in Statistics Education* (pp. 1-16). Amsterdam: IOS Press.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgement Under Uncertainty: Heuristics and Biases*. Cambridge, UK: Cambridge University Press.

Langrall, C., Nisbet, S., & Mooney, E. (2006). *The interplay between students' statistical knowledge and context knowledge in analyzing data*. Paper presented at the Seventh International Conference of Teachers of Statistics.

Mallows, C. (1998). The Zeroth Problem. *The American Statistician*, *52*(1), 1-9.

Sfard, A. (2000). Steering (Dis)Course between Methaphors and Rigor:Using focal analysis to investigate an emergence of mathematical objects. *Journal for Research in Mathematics Education 31*(3), 296 - 327.

Sfard, A. (2008). *Thinking as Communication*. New York: Cambridge University Press

Shin, N., Jonassen, D. H., & McGee, S. (2003). Predictors of Well-Structured and Ill-Structured Problem Solving in an Astronomy Simulation. *Journal of Research in Science Teaching, 40*(1), 6–33.