Journal of Mathematical Behavior 35 (2014) 1–30

Contents lists available at ScienceDirect

The Journal of Mathematical Behavior

journal homepage: www.elsevier.com/locate/jmathb

Conceptual issues in understanding the inner logic of statistical inference: Insights from two teaching experiments

Luis A. Saldanha*, Patrick W. Thompson

Arizona State University, USA

ARTICLE INFO

Article history:
Available online 7 April 2014

Keywords:
Statistical inference
Stochastic event
Sampling distribution
Margin of error

ABSTRACT

We report on a sequence of two classroom teaching experiments that investigated high school students' understandings as they explored connections among the ideas comprising the *inner logic* of statistical inference—ideas involving a core image of sampling as a repeatable process, and the organization of its outcomes into a distribution of sample statistics as a basis for making inferences. Students' responses to post-instruction test questions indicate that despite understanding various individual components of inference—a sample, a population, and a distribution of a sample statistic—their abilities to coordinate and compose these into a coherent and well-connected scheme of ideas were usually tenuous. We argue that the coordination and composition required to assemble these component ideas into a coherent scheme is a major source of difficulty in developing a deep understanding of inference.

Published by Elsevier Inc.

1. Introduction

Statistical inference is one of the most sophisticated and important schemes of ideas in introductory statistics. While data analysis techniques focus on the information content and structure of specific collections of data, it is statistical inference that situates data with respect to a population from which it might be drawn, allowing us to examine the extent to which information extracted from the data in hand can be generalized to the population. The importance to everyday citizenship of understanding statistical inference is clear. Citizens are frequently confronted with published reports of opinion surveys, justifications and implications of policy, and reports of drug trials and experimental medical treatments—many of which speak about margins of error or confidence. Citizens are also confronted with conflicting reports: "Vitamin B-6 can help protect against heart attacks" versus "vitamin B-6 has little effect on the likelihood of heart attacks" or "reform mathematics curricula are effective" versus "reform mathematics curricula are ineffective or downright damaging". Ideas of sampling and statistical inference are important for understanding the degree to which data-based claims are warranted, and that conflicting claims are not necessarily a sign of confusion or duplicity.

With regard to the teaching of statistical inference, the National Council of Teachers of Mathematics (NCTM, 2000) states that in grades 9–12, students should:

* Corresponding author at: School of Mathematical and Statistical Sciences, Arizona State University, P.O. Box 871804, Tempe, AZ 85287-1804, USA. Tel.: +1 480 965 3951.

E-mail address: Luis.Saldanha@asu.edu (L.A. Saldanha).

http://dx.doi.org/10.1016/j.jmathb.2014.03.001
0732-3123/Published by Elsevier Inc.

- use simulations to explore the variability of sample statistics from a known population, and to construct sampling distributions;
- understand how sample statistics reflect the values of population parameters, and use sampling distributions as the basis for informal inference;
- evaluate published reports that are based on data by examining the design of the study, the appropriateness of the data analysis, and the validity of conclusions; and
- understand how basic statistical techniques are used to monitor process characteristics in the workplace (NCTM, 2000, p. 324).

Though the NCTM's recommendations seem straightforward, the presence of "understand" in two of them makes them problematic. It is unclear what it means to understand how sample statistics reflect the values of population parameters. Nor is it clear what it means to understand the use of sampling distributions as the basis for informal inference. It is not even clear what it means to understand sampling distributions.

Statistics teachers share the common experience that students find it very difficult to understand, for example, confidence intervals. Students often understand a 95% confidence interval inappropriately, as indicating that we are 95% confident that the population parameter being estimated will be within that specific interval. Their meaning of "95% confident" is usually tacit, unwittingly suggesting to themselves that the statistical machinery of confidence intervals is activated only to produce that one specific interval, rather than understanding it as a claim about the *method* by which such intervals are produced. Statistics teachers, also, are not immune to thinking that "95% confidence" refers to a population parameter being in a specific interval (Thompson, Liu & Saldanha, 2007).[1] These difficulties underscore a distinction we will highlight in this article as foundational for a coherent understanding of statistical inference. The distinction is between thinking of sampling as a one-time event versus thinking of it as a repeatable event whose resulting distribution of a statistic provides a basis for making confident inferences about a population.

2. Background literature and framing

There is ample evidence that students tend to focus on individual samples and statistical summaries of them instead of on how collections of a statistic's values are distributed. Kahneman and Tversky (1972) concluded that people tend to make judgments about the likelihood of a sample based on how closely the sample represents the population from which it is drawn. They later refined their hypothesis to distinguish between two types of inferential reasoning: what they called *singular* and *distributional* reasoning (Kahneman & Tversky, 1982). When reasoning singularly, a person focuses on a particular outcome and what might have caused it. When reasoning distributionally, a person considers any particular outcome as one instance of a class of similar outcomes, and sees probability statements about individual outcomes as being actually about relative frequencies of similar outcomes in the class (Kahneman & Tversky, 1982, p. 518).

Konold and his colleagues supported Kahneman and Tversky's distinction between singular and distributional reasoning, finding clear evidence that people frequently and persistently interpret probabilistic questions about an event as a request to assess the likelihood that it will, in fact, be the next outcome (Konold, 1989, 1991; Konold, Pollatsek, Well, Lohmeier, & Lipson, 1993). Gigerenzer and colleagues (Gigerenzer, 1994; Hertwig & Gigerenzer, 1999; Sedlmeier & Gigerenzer, 1997) likewise have pointed out that students tend to focus on events as singular phenomena, and make judgments about them that are (from a frequentist perspective) non-probabilistic. Sedlmeier and Gigerenzer's (1997) review of research on the relationship between sample size and the confidence with which people make sample-based predictions is especially salient. They argued that people often reason about tasks that involve a distribution of sample statistics as if the question at hand is about individual samples.

The relationship of the above literature on probabilistic reasoning to issues of understanding statistical inference is that gathering a sample and calculating a statistic from it can be viewed as a *stochastic event*. In our usage, one conceives of an event stochastically when one understands that event as but one instance of a variety of outcomes potentially generated by an underlying repeatable random process. A person conceives of a sample's mean stochastically by seeing this one mean as but one instance of the repeatable process "collect a sample and calculate its mean", together with the anticipation that, if repeated, one would get a variety of sample means.

Combining the idea of stochastic conception of sample with the idea that any sample somehow reflects the underlying population, one can anticipate that the values of a sample statistic will vary somewhat under repeated sampling, and that aggregates of such values will naturally be internally "diverse". That people often conceive (what we see as) stochastic events non-stochastically has important implications for how they draw inferences and how they understand instruction aimed at developing a normative understanding of statistical inference. It is not straightforward for students to view, say, a class'

[1] Velleman (1997) addressed these issues nicely when he said "... the confidence interval is the random quantity whereas the population parameter is fixed and unchanging. Interpretations of confidence intervals should reflect this distinction. When we say, 'with 90% confidence, $63.5 \le \mu \le 65.5$,' we do *not* mean that '90% of the time μ will be between 63.5 and 65.5,' but rather that in the long run, 90% of the intervals we compute from independently drawn samples will include the true mean" (Velleman, 1997, p. 18/5).