Journal of Mathematical Behavior 30 (2011) 235–254

Contents lists available at ScienceDirect

The Journal of Mathematical Behavior

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

Elementary and middle grade students' constructions of typicality

Aisling M. Leavy[a,*,1], James A. Middleton[b]

[a] Mary Immaculate College – University of Limerick, South Circular Road, Limerick, Ireland
[b] Center for Research on Education in Science, Mathematics, Engineering, and Technology, Arizona State University, Tempe, AZ, USA

ARTICLE INFO

Available online 5 May 2011

Keywords:
Statistical reasoning
Mathematical thinking
Elementary students
Middle grade students
Typicality
Data and statistics

ABSTRACT

This study addresses the measures chosen by students when selecting or constructing indices to properties of distributions of data. A series of individual teaching experiments were conducted to provide insight into the development of five 4th to 8th grade students' conceptualizations of distribution over the course of 8 weeks of instruction. During the course of the teaching experiment (emergent) statistical tasks and analogous teacher activities were created and refined in an effort to support the development of understanding. In the process of development, attempts were made by students to coordinate center and variability when constructing measures to index properties of distributions. The results indicate that consideration of representativeness was a major factor that motivated modification of approaches to constructing indices of distributions, and subsequent coordination of indices of variation and center. In particular, the defining features of student's self-constructed "typical" values and notions of spread were examined, resulting in a model of development constituting eight "categories" ranging from the construction of values that did not reflect properties of the data (Category 1) to measures employing conceptual use of the mean in combination with other indices of center and spread (Category 8).

1. Introduction

1.1. Concept of distribution

This study is concerned primarily with describing the development of descriptors that represent distributions of data. Distribution is a statistical term for the arrangement of the observations along a scale of measurement (Hardyck & Petrinovich, 1969). In practical terms, if individual data points (e.g. *frequency*, *magnitude*) from an assortment of observations are arranged (i.e. plotted, tabulated) using a common scale of measurement, the result is a picture of the set of data that embodies its structural properties as a whole. Repeated sets of observations will result in sets of differently shaped plots—each of which varies in its location (e.g. center) and its density (e.g. variation). Here is where our primary interest lies: How do students conceptualize these sets of data? What are the features of students' depictions of distribution specifically related to variation and center that hold promise for coherent development of statistical reasoning for complex data sets?

Briefly, in formal statistics, distributions are indexed by *parameters*. Parameters are numerical, descriptive summaries of a population. Theoretical distributions are often described on the basis of the center and the variability of the distribution. In an applied situation (such as a group of students examining the distribution of a particular variable), once the variable and scale

* Corresponding author. Tel.: +353 87 4177551/61 204978; fax: +353 61 345598.
E-mail addresses: Aisling.Leavy@mic.ul.ie (A.M. Leavy), James.Middleton@asu.edu (J.A. Middleton).
[1] Tel.: +1 480 965 9644.

0732-3123/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmathb.2011.03.001

of measurement have been selected, a group of scores may be obtained and the distribution of scores can be represented using tabular, graphical, or algebraic means. A mathematical treatment of distribution entails the use of such summary characteristics of the distribution that identify algebraically center and variation. The mean (μ) is the most commonly used measure of *center*. It is the weighted average of the possible values of a variable and is also referred to as the expected value of the variable, $E(X)$. Interpreted in terms of the density of a distribution, the mean can be considered as its center of mass. *Variation* represents the spread of the distribution around a measure of location and is typically denoted by some difference function between individual values in the distribution and its collective center. The variance is the most commonly used index of variation. These measures have been described as emerging "as ways of describing how specific data sets are distributed within the space of values" (Cobb, 1999, p. 10). More commonly, these characteristics describe different geometric properties of a distribution including its center, its variation, its symmetry, and its peakedness. Each of these properties contributes to the statistician's understanding of the peculiarities of a set of data.

Both the Principles and Standards for School Mathematics (NCTM, 2000) and the Guidelines for Assessment and Instruction in Statistics Education (GAISE) Report (Franklin et al., 2007) emphasize the importance of statistics and the centrality of the notion of typicality (i.e. representativeness). Throughout the grades there is a visible emphasis on distributions of data with specific attention to descriptions of shape (grades 3–5), measures of center and spread (grades 6–8), and the selection and calculation of summary statistics (grades 9–12). Researchers and practitioners are only beginning to identify and understand the ways in which elementary and middle school students come to index distributions and the extent to which students are able to provide rich and comprehensive descriptions of distributions of data in the ways envisioned by the principles and standards.

2. Theoretical perspective

2.1. Understandings arising from research on the teaching and learning of statistics

2.1.1. Central tendency

Studies of descriptive statistics have focused on children's conceptions of central tendency and representativeness (Cai, 1998, 2000; Hancock, Kaput & Goldsmith, 1992; Mevarech, 1983; Mokros & Russell, 1995; Pollatsek, Lima, & Well, 1981). Findings suggest that the concept of the mean, in particular, is difficult to understand. One plausible explanation is that many students find it difficult to understand that the mean can be a number not actually represented in the data set. Another factor may be the use of the 'fair share' analogy to describe the mean, a comparison which makes little sense when the calculated mean represents a value that cannot possibly be represented in reality (e.g. a mean of 3.2 persons per household). Moreover, the simplicity of the typical 'sum/n' formula may lead to the misconception that the mean is elementary and straightforward (Mokros & Russell, 1995; Strauss & Bichler, 1988).

Beyond computation one of the overriding features of the mean that constitutes difficulty for children concerns the concept of representative value. Hancock et al. (1992) in a study of fifth through eighth graders found that students had difficulties reasoning about aggregate (e.g. mean). Students did not recognize instances in which the mean could be used to typify a data set, as indicated by the lack of instances where the mean was used to compare two groups of unequal size. This difficulty understanding the representative nature of some descriptive statistics extends to conceptualizations of the median. Research has indicated that procedural fluency in computing the median does not indicate the development of associated conceptual knowledge; students able to calculate medians may not necessarily recognize medians as measures of center or as group descriptors of data (Bakker, 2004; Konold & Higgins, 2003). In fact, many students see the median as a feature associated with a particular data value in the middle of the group rather than as a characterization of the entire group (Bakker, Biehler, & Konold, 2005). The ability, then, of students to compute representative values when specifically instructed (Mokros & Russell, 1995) compared to their inability to construct representative values in other situations (Hancock et al., 1992) suggests that students may not understand the role that representative values play in data analysis.

While we are aware of the inherent difficulties associated with the mean, and to a lesser extent the median, as representative values, we have less insight into other measures that students may use to represent data sets, and the ways in which these measures relate to and index landmarks and trends in the data. Regardless of what measure is used, these summary statistics can have little meaning for students unless the data set can be thought of as more than a series of numbers (Mokros & Russell, 1995) – an understanding of the data set as a unit must exist. When this understanding is not present, and the data set is only seen as a set of disjoint numbers, representative values such as the mean and median have little significance. The importance of having a distributional view of data is illustrated in Bakker's (2004) study of 7th graders. Bakker found that students did not take the distribution into account when considering mean and median; this was particularly evident in symmetric distributions where students did not realize that the mean and median would be similar in value. Perhaps then, one reason why efforts to support students construct understandings of the mean as a representative value have not been as successful as anticipated may be because a distributional perspective on data has not been used.

2.1.2. Variation

Variation has received scant attention both instructionally and in research terms, compared to central tendency (Shaughnessy, 1992, 1997). This is somewhat surprising as variation is a key concept in statistics and fundamental to all aspects of data analysis – it is the attempt to account for and model variation in data that defines statistics as a discipline.