Use of the Arithmetic Mean : An Investigation of Four Properties Issues and Preliminary Results

Marjorie Roth Leon and Judith S Zawojewski - Evanston, Illinois, USA

1. Introduction and overview

Teaching statistics has become increasingly important in recent years, as evidenced by recommendations for curriculum reform found in the *Standards for Curriculum and Evaluation* (1989) and *Everybody Counts* (1989), and by the commitment to curriculum development by the American Statistical Association-National Council of Teachers of Mathematics Joint Committee on the Curriculum in Statistics and Probability (Gnanadesikan et al., 1987; Landwehr et al., 1987; Landwehr and Watkins, 1986) and the Used Numbers Project (Russell and Friel, 1989).

The arithmetic mean is probably the most commonly taught and encountered statistic today, appearing in numerous everyday contexts. It is also the one statistic that is consistently included in the elementary school curriculum in the United States. Further, the mean is a fundamental concept in statistics courses because of its predominant use in most inferential statistics and formulae.

Although the mean commands a predominant place in standard curricula and coursework, superficial understanding of the mean by both children and adults is widespread (Carpenter et al., 1981; Goodchild, 1988; Mevarech, 1983; Pollatsek et al., 1980; Zawojewski, 1986). People are able to compute the mean for use in simple situations, but are unable to use it in complex contexts. Pollatsek et al (1980, p.191) refer to students' use of the mean as a "computational rather than a conceptual act. Knowledge of the mean seems to begin and end with an impoverished formula".

Only recently have attempts been made to identify the component properties of the mean. Goodchild (1988) identified three types of meaning for the mean, i.e. the mean as: (1) a representative number; (2) a measure of location; and (3) an expected value. Goodchild found that 13 and 14-year-olds commonly interpret the mean as a location, but need more help with the representativeness and expectation interpretations.

Strauss and Bichler (1988) identified seven properties of the mean and conducted interviews with 8, 10, 12, and 14-year-old Israeli students to assess their ability to apply each of the seven properties to simulated real-world problems. Different age groups demonstrated different courses of reasoning to support their understanding of specific properties. Results indicated that three of the seven properties were mastered by virtually all of the 12 and 14-year-olds.

Goodchild's work focussed on rich uses and interpretations of the mean in context, while Strauss and Bichler examined properties that are within the concept of the mean. Both aspects are important, but represent different types of questions.

Our investigation utilised Strauss and Bichler's conceptual organisation of the properties of the mean, and had three purposes: (1) to investigate what happens developmentally to those statistical properties that were not mastered by age 12-14; (2) to examine the qualities of the statistical properties, including their relative conceptual difficulty and their relative ability to evoke the concept of the mean; and (3) to determine the relative effectiveness of different testing formats in assessing subjects' knowledge about the component properties of the mean.

The four statistical properties selected for investigation were as follows: Property A: The mean is located between the extreme values; Property B: The sum of the deviations is zero; Property F: When the mean is calculated, a value of zero, if present, must be taken into account; Property G: The mean value is representative of the values that were averaged.

The two testing formats employed were story vs numerical presentation. Both are commonly used in standard mathematics curricula in the United States.

2. Methods and results

One hundred and forty-five subjects (42 fourth grade students, 61 eighth grade students, and 42 college students) served as experimental subjects.

A 16-item test was devised, and administered in four different versions. Items on the four versions were identical, but were randomly permuted to minimise order effects, with the stipulation that each of the sixteen items had to appear in one of the first four item positions on one of the four versions of the test. Space was provided on the test sheet for subjects to write justifications for their answers to the first four test items. Thus, each of the sixteen items was accompanied by a written justification one-fourth of the time. This format reduced subject fatigue, maintained subject motivation, and made the study feasible from a time perspective for classroom teachers whose classes participated in the study. There were four items per statistical property, with two of these four items being presented in story format, and two being presented in numerical format. Tests were administered by the experimenters or the subjects' classroom teacher.

A 3 (year in school) $\times$ 4 (statistical property) $\times$ 2 (testing format) repeated measures analysis of variance was conducted, the number of items answered with a value corresponding to the arithmetic mean serving as the dependent variable. Statistically significant main effects were obtained for year in school, statistical property, and testing format. Post hoc comparisons (Scheffe's test) were conducted.

For year in school, post hoc analyses indicated that the mean score for college students was significantly higher than that of the fourth graders (1.29 vs .67 respectively), but that eighth graders' and fourth graders' mean scores did not differ significantly from each other.

Regarding statistical property, post hoc comparisons indicated that the mean score for Property A (1.33) differed significantly from the mean score for Property F (.66) and Property G (.70), but did not differ significantly from Property B (1.20). The mean score for Property B differed significantly from Properties F and G, while the mean score for Properties F and G did not differ significantly from each other.

With testing format, story format was found to be superior to numerical format (i.e. mean scores of 1.07 vs .78 respectively).

Additionally, two 2×2 interactions attained statistical significance. These were year in school × statistical property and statistical property × testing format. These interactions will not be analysed or discussed in detail in this manuscript due to space constraints.

3. Conclusions

Age effects: It was not surprising that performance improved with age. The concept of the mean is introduced formally into the mathematics curriculum in the fourth grade in the United States, reviewed yearly through eighth grade, and reintroduced in high school algebra, and college statistics and mathematics. It is also encountered in academic disciplines such as life science, geography, language, arts, psychology, and business. Additionally, the concept of the mean is encountered in its applied version in many everyday life contexts. These varied exposures to the concept of the mean form a continually growing knowledge base for using and understanding the mean as one ages.

Statistical property effects: Results of this study indicate that subjects found Properties F and G to be almost twice as difficult to understand as they did Properties A and B. Four possible explanations exist for this phenomenon, with these explanations not necessarily being mutually exclusive.

First, Properties A and B deal with the idea of the data distribution and its relationship to the mean, which is very similar to Goodchild's interpretation of the mean as a measure of location. Alternatively, in Goodchild's terminology, Properties F and G address a representative interpretation of the mean. Goodchild's research indicates that the mean-as-location was a commonly-understood interpretation of the mean, while a mean-as-a-representative-measure interpretation was more difficult for persons to grasp.

Second, Properties A and B are computationally based, since they can be solved by applying knowledge about number and operations, while Properties F and G, as previously stated, index the mean as a representative value. As Pollatsek et al. (1980), and Mevarech (1983) argue, most people find it relatively easy to understand the mean as a computational construct, and relatively difficult to understand the mean as a representative value.

Third, Strauss and Bichler state that Properties A and B reflect what they term the statistical aspect of the mean, Property F reflects what they refer to as the abstract aspect of the mean, and Property G reflects what they call the representative aspect of the mean. Our data suggest that the statistical aspect of the mean is the most easily understood by the majority of subjects, while the abstract and representative aspects prove more difficult to master.

Fourth, differences in the difficulty of understanding the various statistical properties could reflect a construction artifact of the problems. Properties A and B contained two possible response choices (yes or no), while Properties F and G required open-ended responses that ranged hypothetically from negative to positive infinity. Therefore, subjects could have correctly answered Propety A or B items at a 50% accuracy level by chance alone, while the corresponding chance accuracy level for Property F or G items would be many times lower.

Testing format: Items presented in story format were significantly easier to solve than were items presented in numerical format. A possible explanation for this may be that story format provides a concrete context, while numerical format provides an abstract context. Piaget states that abstract reasoning is the province of formal operational thought. Recent research, however, indicates that while concrete reasoning (i.e. as described in Piaget's concrete operational stage of thinking) is mastered by most Western children between the ages of 6 and 11, formal operational thought was activated by only two-thirds of a college and middle-aged adult sample that was presented with unfamiliar problems to solve (Keating, 1980), and that only 50%-60% of college students and older adults applied formal operational reasoning to problems in general (Neimark, 1975).

4. Future directions

The following issues remain unresolved, either because data analysis has not yet been completed, or because additional research questions need to be posed in future studies. These issues may be outlined as follows:

(i) Assessing the extent to which students answered items representing Properties F and G with "median-like" responses, which in fact represent appropriate alternative responses for these properties. These data have already been obtained, but not yet analysed. Analysis of this data will shed increasing light on the extent to which the phrase "Choose one number that best describes the typical number in the set" represents an everyday, commonsense, meaning of the word "average".

(ii) Identifying the type of explanations that subjects provided in their written justifications of their responses. These data are also awaiting analysis. A point of interest is the degree of correspondence between our justification categories and those of Strauss and Bichler, which will help assess the similarity between the in-depth, individual interview methodology employed by the latter researchers, and the group testing, paper-and-pencil methodology employed in the present research.

(iii) The ramifications of changing the word "typical" to "average", "mean", or "representing the whole set of values". This concern represents an additional question for future investigations involving the properties of the arithmetic mean.

(iv) The effect of changing the answers to Property F and G items to non-whole number results. Again, this concern represents an additional question for future studies.

References

Board on Mathematical Sciences [and] Mathematical Sciences Education Board (1989) *Everybody Counts : A Report to the Nation of the Future of Mathematics Education.* National Academy Press, Washington DC.

Carpenter, T P, Corbitt, M K, Kepner, H S, Lindquist, M M and Reys, R E (1981) *Results from the Second Mathematics Assessment of the National Assessment of Educational Progress.* National Council of Teachers of Mathematics, Reston, VA.

Gnanadesikan, M, Scheaffer, R L and Swift, J (1987) *The Art and Techniques of Simulation.* Dale Seymour, Palo Alto, CA.

Goodchild, S (1988) School pupils' understanding of average. *Teaching Statistics* **10**, 77-81.

Keating, D (1980) Thinking processes in adolescence. In: J Adelson (ed) *Handbook of Adolescent Psychology.* Wiley, New York.

Landwehr, J M, Swift, J and Watkins, A E (1987) *Exploring Surveys and Information from Samples.* Dale Seymour, Palo Alto, CA.

Landwehr, J M and Watkins, A E (1986) *Exploring Data.* Dale Seymour, Palo Alto, CA.

Mevarech, A R (1983) A deep structure model of students' statistical misconceptions. *Educational Studies in Mathematics* **14**, 415-429.

National Council of Teachers of Mathematics, Committee on Standards for School Mathematics (1989) *Curriculum and Evaluation Standards for School Mathematics.* The Council, Reston, VA.

Neimark, E D (1975) Intellectual development during adolescence. In: F D Horowitz (ed) *Review of Child Development Research* (Vol 1). University of Chicago Press, Chicago.

Newman, C M, Obremski, T E and Scheaffer, R L (1987) *Exploring Probability.* Dale Seymour, Palo Alto, CA.

Pollatsek, A, Lima, S and Well, A D (1980) Concept of computation : students' understanding of the mean. *Educational Studies in Mathematics* **12**, 191-204.

Russell, S J and Friel, S N (1989) Collecting and analysing real data in the elementary school classroom. In: P R Trafton and A P Shulte (eds) *New Directions for Elementary School Mathematics, 1989 Yearbook.* National Council of Teachers of Mathematics, Reston, VA, 134-148.

Strauss, S S and Bichler, E (1988) The development of children's concepts of the arithmetic average. *Journal for Research in Mathematics Education* **19(1)**, 74-80.

Zawojewski, J S (1986) *The Teaching and Learning Processes of Junior High Students Under Different Modes of Instruction in Measures of Central Tendency.* Unpublished doctoral dissertation, Northwestern University, Evanston, Illinois.