# Coordinating Student Learning and Teacher Activity – The Case of Savannah: Motivating an Understanding of Representativeness through Examination of Distributions of Data

Dr. Aisling Leavy, Mary Immaculate College, Limerick

This research study investigated the measures chosen by five 4th to 8th grade students when selecting statistical measures to describe distributions of data. Over the course of eight weeks of instruction, individual teaching experiments were conducted to investigate the development of understanding of distribution. The results indicate that consideration of representativeness was a major factor that motivated modification of approaches to constructing indices of distributions. This paper outlines the case of a 4th grade student who participated in the study and narrates her journey as she grappled with increasingly complex statistical ideas relating to representativeness.

## Theoretical Perspective

Recent approaches to curriculum design and research in statistics education engage students in examining *distributions* of data rather than working with isolated data values devoid of context. A distribution refers to the arrangement of data values along a scale of measurement (Hardyck & Petrinovich, 1969) and distributions may be depicted using a variety of graphical representations. Distributions are commonly described using *parameters*: measures of centre, variability, skew (symmetry) and kurtosis ("peakedness"). The parameters that are the focus of study at primary school level are center (mean, median, and mode) and variability (range).

An important skill when examining a distribution of data is the ability to find trends and patterns in the data and thus infer properties. This is in essence what the parameters of a distribution describe – underlying signals (Konold & Pollatsek, 2002) or messages communicated by the data. In primary school mathematics one way to focus students on the signals emitted by a distribution is to ask them to elect a value that best represents the distribution – often referred to as a *typical value*. This notion of typicality is referenced in many curricula and by national bodies such as in the Principles and Standards for School Mathematics (NCTM, 2000).

Recent research in statistics education examines primary students (Cobb, 1999; Konold, Robinson, Khalil, Pollatsek, Well, Wing & Mayr, 2002; Leavy & Middleton, 2001; Lehrer & Schauble, 2002; Mokros & Russell, 1995) and preservice teachers (Leavy, 2004; Leavy & O'Loughlin, 2006) understandings of and efforts to represent (or typify) distribution. Research examining the measures children use to represent and summarize distributions of data indicates that a variety of measures are used. The most frequently identified approaches used to determine representative values are the measures of center and spread such as modal clumps, ranges, and intervals. Student understandings of appropriate measures to represent data vary and include conceptions of average as modal (values occurring with the greatest frequency) and middle (akin to the median).

## Purpose of the Study

The research from which this paper was motivated examines student's decision-making processes when selecting statistical measures to represent a distribution of data. In this line of inquiry, effort was made to ascertain what (formal or informal) measures students utilize in representing distributions, and how these measures develop and contribute to an understanding of distribution shape and structure. For the remainder of the paper, the measures students select to represent distributions are referred to as *representative* measures.

## Design and Methodology

### *Participants*

Five participants from a variety of urban schools in the Southwestern United States participated in the study. Students had not received any specialized instruction or experiences with data prior to participation in the study. The 4th grade student, who is the focus of this paper, had few experiences with data, limited mainly to pictogram and pie chart interpretation. All other students had received curricular instruction relating to graphical representations in addition to calculating measures of central tendency.

### *Procedure*

Teaching experiment methodology was used to uncover student's conceptual understanding of distribution. Teaching experiment methodology consisted of a clinical interview phase, a teaching phase and an analysis phase. Clinical interviews lasted 60 minutes and were video taped. Two clinical interviews were conducted with each student, one at the outset of the study to determine initial understanding of distribution and another at the end of the study to assess changes in understanding of distribution.

Teaching episodes took place following the initial clinical interview and consisted of each student working individually on a variety of model-eliciting activities. Each teaching episode was characterized by the presentation of specific tasks designed to elicit models of students understanding of certain statistical phenomena. The tasks used were derived from a variety of sources, the majority being developed by the instructor and informed by a previous pilot study (Leavy, 2003), in addition to original tasks used by permission of the Cobb & McClain group at Vanderbilt (McGatha, Cobb, McClain, 2002). Problems were presented in several forms (e.g. graphical, tabular etc.) and students were asked to describe distributions and determine representative values for the distributions. In other cases students collected their own data and constructed typical values for the data. All data sets were presented predominantly on line plots or on stem and leaf plots chosen due to their structure facilitating the display of individual data values, the visibility of which were considered important for the construction of representative values. Eight teaching episodes, lasting approximately 50 minutes each, were carried out with Savannah; each episode was audio taped and several sessions were videotaped. These episodes were used to construct a model of her mathematical thinking and guide her to develop more sophisticated ways of reasoning about data.

### *Data Analysis*

The analysis phase involved examining the data resulting from the clinical interviews and the teaching episodes. Data from the clinical interview were used to construct a hypothetical learning trajectory (Simon, 1995) for Savannah. Ongoing analysis occurred between teaching episodes, and a retrospective analysis focused on the cumulative episodes. Results from the ongoing analysis were used to inform the direction of the study and the construction of tasks for subsequent teaching episodes. The main sources of data were the researcher's field notes, audio and video records of the interactions, and samples of the student's written work. Savannah's verbalizations, inscriptions, and models were analyzed as to how they revealed her schemes for structuring and depicting distributions. Hence, the teaching episodes provided the means to generate in-depth knowledge of Savannah's conceptual understanding of data and how it was fostered in a coherent instructional sequence.

**Results: A Case Study of Savannah**

In the reporting of the results emphasis is placed on understandings of the statistical concepts examined in the initial clinical interview and the subsequent development of student's conceptual schemes over the course of the teaching experiment. Each of the five participants in the teaching experiment constitutes a separate developmental case. For the purpose of this paper, I will focus on the case of Savannah the 4th grade student. This case casts light on particular statistical understandings and misconceptions the roots of which have important implications for the design of statistical instruction particularly within the Irish primary education system.

*Background on Savannah*

Savannah was 9 years old and stated that mathematics was her favourite subject. She was in a 4th grade non-differentiated mathematics class in public school. Her in-class performance in mathematics placed her in the upper 10% of the students. As a result, Savannah would be placed in a group for more mathematically able students in 5th grade.

*Initial approaches to constructing representative measures*

During the initial clinical interview and first teaching episodes Savannah constructed representative measures based on her *own experience of the phenomenon* under investigation. In these situations Savannah seemed not to comprehend that data referred to an actual event, this was indicated by her inability to attend to the data presented on the graphs. The following task (Task 1, Appendix A) is an example of one such episode in which Savannah demonstrated the ability to read values from the graph but the inability to use the data to answer the question.

> Teacher: How many gummi bears would you say are in a packet of candy?
> Savannah: Ah.. like … maybe 25 or no .. probably 26.
> Teacher: Why do you say 26 gummi bears? …… (long pause)
> Teacher: And did those students (pointing to the graph) get as many as 26?
> Savannah: No, but … I think you'd get more. It really depends on … just … I always get more.

As can be seen from the example above, Savannah used her own experience of the size of a packet of Gummi Bears to answer the question. The value Savannah elected actually fell outside the range of values presented on the line plot and resulted in the presentation of an *idiosyncratic* and non-representative data value.

Initial analysis of Savannah's response suggested that the situation presented in the task had particular relevance for Savannah and influenced her response. To test this hypothesis, Savannah was presented with a number of tasks and asked to interpret the data presented on them. As with the 'Gummi Bear' task, she continued to use her own experiences of the problem situations to construct representative values. It seemed that Savannah (erroneously) perceived that graphs were merely display tools used to present a 'picture' of data. In her view, graphs were objects the purpose of which was to read off discrete values rather than presenting a means to examine distributional structure – hence explaining her difficulties with questions that demanded an examination of structure. The following transcript (Task 2, Appendix A) arising from the first teaching episode provides another example of her tendency to revert to her own experiences when determining representative values.

> Teacher: Based on this graph, generally, how many medals do you think a country won in the Olympics? …
> Savannah: Yeah .. em .. (sighs) okay 8.
> Teacher: Why 8?
> Savannah: Because .em .well I watched the Olympics and they said that it had been on for a while and they said she has already won 3 medals. And actually I would probably say about 6.
> Teacher: 6 okay. Why 6?
> Savannah: Because I was watching the Olympics and a girl had .. this swimmer had only had .. em .. only 3 medals .. and it was almost over so she couldn't get that much more.
> Teacher: Okay.

*Teacher Activities to Promote the Development of Understanding*

Due to the difficulties in getting Savannah to attend to the data presented in tasks, several teaching episodes were devoted to data collection activities. The intent was that by providing experiences in collecting data and constructing graphs, Savannah would begin to understand that the data presented in graphs represented real situations. These data could then be used as information to determine representativeness. Examples of such activities were counting the numbers of raisins in boxes of raisins, the number of M&Ms in packets, recording resting and active heart rates etc. The data were collected and graphed on line plots from which Savannah was required to describe and interpret her data. As a result of engaging in such data modeling (Lehrer & Romberg, 1996) activities Savannah began to attend closely to the structure of the data, and midway through the teaching experiment paid attention to the structure of data sets that she had not constructed. It cannot be overemphasized that, in Savannah's case, the activity of collecting and graphing data was a necessary precondition to understanding data. However, the transition was not effortless and as can be seen from the following transcript (Task 3, Appendix A) she did at times refer back to her own data collection experiences and neglected the data on display.

> Teacher: Based on this data, if you were the general manager how many raisins would you say are generally in a box of raisins?
> Savannah: I'd probably say 17.
> Teacher: Tell me why.
> Savannah: Because last week I counted up and measured and most boxes I had got were 17.
> Teacher: Now imagine that you didn't count them up last week .. that you didn't know there were 17. So imagine you were Anne Browne and you only had those 30 boxes …
> Savannah: Then I'd probably say 26.
> Teacher: Why?
> Savannah: Most people got the .. 7 that's what it was. There is more of that [26] than that [17]. And it would probably be closest to that [26].

As can be seen from the example above, the results of the data collection activities in which Savannah had been involved were more salient than the data presented in the task. Hence her experiences took on greater importance and influenced her responses. Savannah made the transition from collecting her own data and making inferences from these data to being presented with data sets collected by others. Great care was taken in explaining how these data were collected and Savannah was often presented with descriptive scenarios related to the problems. The provision of context to problems facilitated Savannah in understanding and attending to the data. Her representative values eventually became less related to her own experiences and she began to construct representative values using one particular strategy.

*Emerging Understandings of Representativeness*

The nature of the representative values provided by Savannah remained steadfast throughout the teaching experiment. All her typical values shared the same characteristic: they were modal values. Regardless of the shape of the distribution presented to Savannah, she always chose the mode as the value that represented the data set. Savannah never adjusted the representative value from the mode to allow for outliers, skew or gaps in the data. Even when her attention was drawn to the shape of the distribution and to the inability of a modal value to represent all the data, she remained unwavering in her decision that the mode represented the majority of the data.

When asked to construct a representative value from a multimodal data set, Savannah's initial strategy was to name all the modes that appeared in the data set. This pattern occurred across several tasks (Task 4, Appendix A), and Savannah was reluctant to examine the data so as to choose one specific typical value from the several possibilities.

> Teacher: If somebody came into Ms. Murphy's class and said "Generally how tall are the students in your class. What would Ms. Murphy say?"
> Savannah: em .. 48
> Teacher: Okay. Why would you say 48?
> Savannah: 48 actually 48 or 43 because they have the most amount of people .. heights.

In subsequent teaching episodes Savannah was presented with scenario's that required her to choose one of the several possible modal values as the representative value. Savannah examined the data sets and chose modal values that had the largest number of data values clustered around it. This strategy indicated that she had some notion of the most frequent value being a representative measure (Task 5, Appendix A).

> Teacher: So it says that Jodie told the principal that the typical height of a 6th grade student was 43 or 48 inches. Why do you think she said that?
> Savannah: Because 43 and 48 em have the highest .. em .. number inches .. eh .. of a person in 6th grade.
> Teacher: However, the principal said: "... You must choose either 43 or 48 inches. Look at the data so that you can make your decision ..." Which value do you think Jodie should choose .. – 43 or 48 inches? Why?
> Savannah: 43 because it comes up first.
> Savannah laughs
> Savannah: Oh I have a better idea. 48 because there are more numbers that are closer to 48 than 43.
> Teacher: Great. For example?
> Savannah: Em next to 48 there's 49, 50, 47.
> Savannah: For 43 there's only 44 and 42.

*Teacher Activities Highlighting Attention to the Limitations of the Mode as a Representative Measure*

While the mode is an appropriate representative measure for many distributions of symmetric shape, it becomes less representative once the distribution becomes skewed. Savannah used the mode as a blanket measure regardless of distributional shape resulting in non-appropriate representative measures on several occasions. As a result, it was considered important that Savannah's attention be drawn to the limitations of the mode and that she be encouraged to consider other measures that represent a data set.

Many attempts were made to draw Savannah's attention to the importance of a typical value being representative. Savannah understood that the mode was not accurate in describing all the values in a distribution and expressed this understanding on a number of occasions. On one occasion she was presented with a scenario of another student deciding not to choose a modal value and adjusting the typical value from the mode to reflect the data more accurately. Savannah admitted the value of adjusting scores to increase representativeness but voiced concerned regarding the *truth* of such approaches (Task 6, Appendix A).

> Teacher: What do you think is the typical value? How high do you think a 4th grader is?
> Savannah: 30 .. 31
> Teacher: Why?
> Savannah: Because most kids are.
> Teacher: Okay now if you look at the next part of the question.
> Teacher reads the explanation
> Teacher: Do you understand that even though 31 was the most it was also the smallest height?
> Savannah: Yeah.
> Teacher: She says there were lots of values higher than 31 so she moved the typical value up higher to reflect more of the students heights…. Do you think that is a good strategy?
> Savannah: No. Because it just wouldn't be .. it wouldn't be true.
> Teacher: Okay?
> Savannah: The .. the .. if most students are 31 … but well it's more fair.

It was thought that following introduction to the median that Savannah may have found it appropriate to use the median as a representative value. She chose, however, to select modal values even in the cases of distributions in which the median would have been the appropriate measure and the mode inappropriate. Savannah was presented with the situation of a student her age choosing a median value as the typical value and asked what she thought of the strategy. While she admitted that there were merits associated with using the median she stated that her first choice would remain the mode.

Data from the clinical interview indicated that Savannah had no conception of the mean. In the teaching episodes, the mean presented difficulties for Savannah and it took several episodes for her to construct a meaningful understanding of the mean. She was initially introduced to the mean through a fair share model. She was encouraged to carry out fair share problems by physically distributing candy fairly amongst a number of agents. Savannah succeeded in solving each of the problems by distributing the candy. However, she often found it difficult to relate the answer back to the question and understand why the answer made sense. Savannah was also introduced to a balance approach to the mean. While she was not encouraged to balance values at equal distances on either side of the mean, she was exposed to the notion that the mean is the point on which the distribution balances. Initially it seemed as if Savannah had grasped the notion, as she was quite accurate at eyeballing the balance point of graphs, however, subsequent probing identified that she was in fact gauging the median. She was identifying a balance point of the distribution by finding the point on the graph that divided all the data values in half. She was not conceptualizing balance as a weighted quantity. Savannah never grasped the concept of the mean as a representative measure and did not suggest that the mean could be used to summarize a distribution of data.

*Final Understandings of Representativeness*

In the final clinical interview Savannah was presented with the *Gummi Bear* task. Her response to this task indicated that she was attending to the presented data and not making references to her own experiences. The actual representative value provided by Savannah represented the mode of the distribution. Her choice of the mode reflects her strong belief that the mode is representative of the data because it is the most frequently occurring data value. Savannah's justification for the mode being the typical value is more sophisticated than her previous arguments in that she incorporates the notion of the importance of the typical value being surrounded by other values. This indicates a transition to the notion of variability in that she has incorporated a 'cluster' idea, and may signal an attempt at coordinating the notions of location and variability.

Savannah did, however, realize that the mode was not representative of all the values in the distribution. The terms she used to express this notion were the concepts of "fair" and "true." She considered that the modal values told the truth about the data, the truth being that most values occurred at the mode. Hence in some ways the term typical meant the most frequently occurring value. The mode, however, according to Savannah was not always 'fair' because it did not represent everyone in the data set. Truth, however in Savannah's mind, was primary to fairness. As the mode told the truth, even to the detriment of being fair, she felt compelled to use the mode in determining representative values (Task 7, Appendix A).

> Teacher: How long can a 3rd grade student hold their breath?
> Savannah: 14 or 7 seconds.
> Teacher: Okay. Why?
> Savannah: Because they both .. have the most amount of people .. 3 people have 14 and 3 people have 7 and one person has 5 .. 3 people held their breath for 14 which is the most. But still 3 people held their breath for 7 seconds so I think that 14 and 7 are the only ones that matter and 24 doesn't matter 'cause it is all the way down there and only one person got it.
> Teacher: Great thank you.
> Savannah: They already got they [person who got 24 seconds] got a lot of stuff and then .. it's not really fair.
> Teacher: Oh okay. On who?
> Savannah: Okay the people who held the most [the value at 24]. They beat that persons scores [14].. but still most people got it and ..
> Savannah: It is not really fair but it's true.

**Summary of Savannah's Schemes**

When asked to construct a representative value, Savannah displayed the tendency to ignore the data presented to her and rely on her own experiences of the presented situations. This resulted in the construction of *idiosyncratic* and thus non-representative measures. This predisposition to focus on her own experiences was overcome by engaging Savannah in data modeling activities. As a result of modeling data Savannah began to understand the meaning of data values presented on graphs, and with this comprehension came the keenness to examine graphically represented data.

Subsequently, Savannah's construction of representative values was exclusively focused on identifying *modal values* in a data set. Savannah considered the modal value as constituting an exact and accurate typical value, whilst not always being a fair value, she did consider the mode as embodying an authentic and truthful representation of the data. It seemed that the mode was truthful because when used as a typical value it represented the 'most' occurrences of data in contrast to the median (or mean) which did not necessarily represent the most frequently occurring data values. In a sense, Savannah considered the mode as accurately portraying or reflecting the majority of the data values. This frequentist view of data is not atypical and has been found in other studies. When presented with multimodal data Savannah used the distribution of data around each mode as a determining factor indicating an understanding of a typical value being representative.

Savannah also exhibited understanding of the *median* as constituting the middle of a data set and could identify the median when presented with a variety of forms of data. She did not, however, utilize the median as a representative measure. Thus she demonstrated procedural understanding of the median without the corresponding conceptual understandings (Hiebert and Lefevre, 1986). In other studies, this development of skills rather than concepts (Resnick & Ford, 1981) has been shown to result in students being able to calculate medians but not necessarily recognizing medians as measures of center or as group descriptors of data (Bakker, 2004; Konold & Higgins, 2003). In fact, other studies indicate that 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).

Attempts were made to introduce Savannah to the concept of mean. However, she demonstrated difficulties in constructing a rich understanding of the mean and did not conceive of the mean as a representative measure. Following completion of the teaching experiment, Savannah's understanding of the mean was limited to a fair share approach to the mean, and calculation of the mean could be carried out only through the use of concrete materials. Similarly, other studies have indicated that an overriding feature of the mean that presents difficulty for children is representativeness. A study of fifth through eighth graders found that 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 (Hancock, Kaput & Goldsmith, 1992). The ability, 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 do not typically understand the role that representative values play in data analysis.

In summary, many experiences in collecting and graphing data were required before Savannah understood the meaning of data. The use of activities that required data collection, organization, and depiction resulted in Savannah demonstrating the ability to construct and interpret data using a variety of graphical representations. Her limited experiences in school mathematics coupled with her age and grade level resulted in Savannah never moving beyond her use of modal values in representing a data set. She understood the limitations of such measures but was reluctant to relinquish use of the mode. The concept of variability was not explored in any detail with Savannah due to her reluctance to consider the importance of any features of a data set apart from the mode. She did examine the distribution of data around the mode indicating her awareness of variability, however this was only in the context of the mode. However, it does indicate the incorporation of a 'cluster' idea, which might be thought of as a precursor or first look at variability.

## References

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

Bakker, A., Biehler, R., and Konold, C. (2005). Should young students learn about boxplots? In G. Burrill (Ed.), Curricular Development in Statistics Education. International Association for Statistical Education (IASE) Roundtable, Lund, Sweden, 28 June-3 July, (pp. 163-173), http://www.stat.auckland.ac.nz/~iase/publications.(??)

Cobb, P. (1999). Individual and collective mathematical development: The case of statistical data analysis. Mathematical Thinking and Learning, 1(1), 5-43.

Hancock, C., Kaput, J.J., & Goldsmith, L.T. (1992). Authentic inquiry with data: critical barriers to classroom implementation. Educational Psychologist, 27(3), 337-364.

Hardyck, C.D., & Petrinovich, L.F. (1969). Introduction to Statistics for the Behavioral Sciences. Philadelphia, Saunders.

Hiebert, J., & Lefevre, P. (1986). Conceptual and procedural knowledge in mathematics: An introductory analysis. In J. Hiebert (Ed.), Conceptual and procedural knowledge: The case of mathematics (pp. 1-27). Hillsdale, NJ: Lawrence Erlbaum Associates.

Konold, C., & Higgins, T. (2003). Reasoning about data. In J. Kilpatrick, G. Martin, & D. Schifter (Eds), A research companion to NCTM's Standards. Reston, VA: NCTM.

Konold, C., & Pollatsek, A. (2002). Data analysis as the search for signals in noisy processes. Journal for Research in Mathematics Education, 33, 259-289.

Konold, C., Robinson, A., Khalil, K., Pollatsek, A., Well, A., Wing R., & Mayr, S. (2002). *Students' use of modal clumps to summarize data*. Paper presented at The Sixth International Conference on Teaching Statistics, Cape Town, South Africa.

Leavy, A.M. (2003). Gifted Students' Understanding of Statistics: Analysis of Data Arising from a Small Group Teaching Experiment. Gifted and Talented International, 18(1), 17-26.

Leavy, A.M. (2004). Indexing Distributions of Data: Preservice Teachers' Notions of Representativeness. School Science and Mathematics, 104(3), 119-134.

Leavy, A.M. & Middleton, J.A. (2001, April). Middle Grade Students Understanding of the Statistical Concept of Distribution. American Educational Research Association Annual Conference in Seattle, Washington, USA.

Leavy, A.M. & O'Loughlin, N. (2006) Moving Beyond the Arithmetic Average: Preservice Teachers Understanding of the Mean. *Journal of Mathematics Teacher Education, 9*(1), 53-90.

Lehrer, R., & Romberg, T. (1996). Exploring children's data modeling. *Cognition and Instruction, 14*(1), 69-108.

Lehrer, R., & Schauble, L. (2002). Inventing data structures for representational purposes: Elementary grade students' classification models. Mathematical Thinking and Learning, 2, 51-74.

McGatha, M., Cobb, P., & McClain, K. (2002). An analysis of students' initial statistical understandings: Developing a conjectured learning trajectory. *Journal of Mathematical Behavior*, *16*(3), 339-355.

Mokros, J., & Russell, S.J. (1995). Children's concept of average and representativeness. *Journal for Research in Mathematics Education, 26*(1), 20-39.

National Council of Teachers of Mathematics (2000). *Principles and Standards for School Mathematics*. Reston, VA: Author.

Simon, M.A. (1995). Reconstructing mathematics pedagogy from a constructivist perspective. *Journal for Research in Mathematics Education, 26,* 114-145.

**Appendix**

<u>Task 1</u> *Number of 'Gummi Bears' in a packet*

A class of students were interested in examining the number of Gummi Bears in a packet. They each counted the number of bears in a packet and put their results on a line plot.

What does the graph tell you about the number of Gummi bears students counted?

From examining the graph, generally how many Gummi Bears would you expect to find in a packet of Gummi Bears? Why? Here are their results:

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | X | | | | | | | |
| | X | | X | | | | | | | |
| | X | | X | | | | | | | |
| X | X | X | X | X | X | | X | | | |
| X | X | X | X | X | X | X | X | X | | |
| 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |

<u>Task 2</u> *The Winter Olympics Problem*

The 1984 Winter Olympics were held in Sarajevo, Yugoslavia. The data were reorganized into a line plot. Describe the data. What is the typical number of medals won by a country?

| | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| X | | | | | | | | | | | | | | | | | | | | | | | | |
| X | | | | | | | | | | | | | | | | | | | | | | | | |
| X | X | | X | | | | X | | | | | | | | | | | | | | | | | |
| X | X | X | X | X | X | | X | X | | | | X | | | | | | | | | | | X | X |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 |

<u>Task 3</u> *The Raisin Data Controversy*

The general manager of sun maid raisins, Ann Brown, has received an inquiry regarding the average number of raisins that are in a box of sun maid raisins. She went down to the factory floor and picked up 30 boxes of raisins, counted the number of raisins in each box and the put the results on a line plot. Here is the number of raisins that he found in the 30 boxes:

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | X | |
| | | | | | | | | | X | |
| | | | | | | | X | | X | |
| | | | | | | | X | | X | |
| | | X | X | | | | X | X | X | |
| | X | X | X | X | X | X | X | X | X | |
| X | X | X | X | X | X | X | X | X | X | |
| 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 |

Based on this data, if you were the general manager how many raisins would you say are generally in a box of raisins?

Task 4 *Class Height Data*

On school sports day, Miss Murphy's 5th grade basketball team played a game of basketball against Mr. Cody's basketball team. Miss Murphy's team won the game. Jamie, a student in Mr. Cody's class, believes that the game was not fair. She believes that students in Miss Murphy's class are taller than students in Mr. Cody's class. So, when choosing a basketball team, Miss Murphy has a lot of tall students from which to choose. Jamie decided to measure the height of each student in both classes and construct a line plot for each class. Examine the graphs.

What is the typical height of a student in Miss Murphy's class?

Height of Miss Murphy's students

| | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | X | | | | | X | | | | | | | | | | |
| | | | X | | | | X | X | X | X | | | | | | | | |
| | | X | X | | | | X | X | X | X | | | | | | | | |
| X | | X | X | X | | | X | X | X | X | | | X | X | | | X | X |
| 40 | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 49 | 50 | 51 | 52 | 53 | 54 | 55 | 56 | 57 | 58 |

Task 5

A student in the 6th grade class, Jodie, was also asked to calculate the height of the students in the 6th grade class. So, Jodie measured each member of the class using a tape measure and recorded it on a line plot. This is the line plot of the student's heights.

Height of 6th grade students

| | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | X | | | | | X | | | | | | | | | | |
| | | | X | | | | X | X | X | X | | | | | | | | |
| | | X | X | | | | X | X | X | X | | | | | | | | |
| X | | X | X | X | | | X | X | X | X | | | X | X | | | X | X |
| 40 | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 49 | 50 | 51 | 52 | 53 | 54 | 55 | 56 | 57 | 58 |

Jodie told the principal that the typical height of a 6th grade student was 43 or 48 inches. However, the principal said:

"Tell me one number. You must choose either 43 or 48 inches. Look at the data so that you can make your decision between the two."

Which value do you think Jodie should choose as the typical height – 43 or 48 inches? Why?

Task 6 *What do you think?*

| | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| X | | | | | | | | | | | | | | | | | | | | | | | | |
| X | | | | | | | | | | | | | | | | | | | | | | | | |
| X | X | | X | | | | X | | | | | | | | | | | | | | | | | |
| X | X | X | X | X | X | | X | X | | | | X | | | | | | | | | | | X | X |
| 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 4 | 4 | 4 | 4 | 4 | 4 |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 0 | 1 | 2 | 3 | 4 | 5 |

Another student who did this task during the summer chose 34 inches as the typical height even though 31 inches was the height that occurred most. Maria said that she chose 34 because:

"I originally chose 31 as a typical value because it occurred the most. But then I looked at the data and saw that there were lots of people who were taller than 31 inches. In fact, even though 31 was the one that occurred most it was also the smallest value. So, I didn't think that I should give the smallest height as the height of most students. There were lots of values higher than 31 so I moved the typical value up higher to reflect more of the students heights. However, I made sure that I kept my new value close to the 31 inches"

What do you think? Is this a good strategy? Why?

Task 7 *How long can you hold your breath?*

The 3rd grade classroom in the same school decided to carry out a similar study. They wanted to see how long they could hold their breaths. They also made a line graph of their results. What is the typical length of time a 3rd grader can hold their breath?

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | X | | | | | | | X | | | | | | | | | | |
| | X | X | X | | | | | X | X | X | | | | | | | | | |
| X | X | X | X | X | | | X | X | X | X | X | | | | | | | | X |
| 5 | 6 | 7 | 8 | 9 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 |
| | | | | | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 0 | 1 | 2 | 3 | 4 |

Task 6 *What do you think?*

| | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| X | X | | X | | | | X | | | | | | | | | | | | | | | | | |
| X | X | X | X | X | X | | X | X | | | | X | | | | | | | | | | | X | X |
| X | | | | | | | | | | | | | | | | | | | | | | | | |
| 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 4 | 4 | 4 | 4 | 4 | 4 |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 0 | 1 | 2 | 3 | 4 | 5 |

Another student who did this task during the summer chose 34 inches as the typical height even though 31 inches was the height that occurred most. Maria said that she chose 34 because:

"I originally chose 31 as a typical value because it occurred the most. But then I looked at the data and saw that there were lots of people who were taller than 31 inches. In fact, even though 31 was the one that occurred most it was also the smallest value. So, I didn't think that I should give the smallest height as the height of most students. There were lots of values higher than 31 so I moved the typical value up higher to reflect more of the students heights. However, I made sure that I kept my new value close to the 31 inches"

What do you think? Is this a good strategy? Why?

*Task 7* *How long can you hold your breath?*

The 3rd grade classroom in the same school decided to carry out a similar study. They wanted to see how long they could hold their breaths. They also made a line graph of their results. What is the typical length of time a 3rd grader can hold their breath?

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | X | | | | | | | X | | | | | | | | | | |
| | X | X | X | | | | | X | X | X | | | | | | | | | |
| X | X | X | X | X | | | X | X | X | X | X | | | | | | | | X |
| 5 | 6 | 7 | 8 | 9 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 |
| | | | | | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 0 | 1 | 2 | 3 | 4 |