Statistics Unique Place in Math

Paul Dale

Abstract

This unit explains statistics unique place in the mathematics curriculum. It introduces students to statistical thinking and questioning by implementing the first two components - Formulate Questions and Collect Data - of the statistical problem solving process as set forth by the *Guidelines for Assessment and Instruction in Statistics Education (GAISE) Report.* Students will evaluate questions to determine if they require variability (and therefore a statistical investigation) or if they only require a single deterministic answer, as is the case for the majority of the mathematics curriculum. Students will then design and implement a plan to collect data to answer a statistical question about their class, "Who has a bigger arm span in the class - boys or girls?"

Rationale

Statistics is important to teach to young students and teachers must introduce statistics and probability to their students per Common Core and Pennsylvania State Standards in the sixth grade. The National Council of Teachers of Mathematics (NCTM) included "Data Analysis and Probability" as one of the five content strands in their influential document, *Curriculum and Evaluation Standards for School Mathematics* (NCTM, 1989). This change was in response to the "advances in technology and modern methods of data analysis in the 1980s, coupled with the data richness of society in the information age". The document continued to provide "the basis for reform of mathematics curricula in many states [and] the acceptance of and interest in statistics as part of mathematics education gained strength". Other educational groups joined the NCTM in emphasizing the need for students to learn statistics.

Why has data and statistics gained much attention by these groups? Students must develop statistical literacy to thrive in the modern economy. Individuals who acquire this skill can amount more success than those without. Consider Lynn Steen's words in *Mathematics and Democracy: The Case for Quantitative Literacy* (Steen 2011):

Quantitative literacy, also called numeracy, is the natural tool for comprehending information in the computer age. The expectation that ordinary citizens be quantitatively literate is primarily a phenomenon of the late twentieth century… Unfortunately, despite years of study and life

experience in an environment immersed in data, many educated adults remain functionally illiterate… Quantitative literacy empowers people by giving them tools to think for themselves [sic], to ask intelligent questions of experts, and to confront authority confidently. These are the skills required to thrive in the modern world.

Teachers can therefore serve their students well by providing the tools they need to develop their statistical literacy.

Since statistics is a relatively new subject area for many teachers, many teachers on the front lines fail to comprehend its unique and important place in the curriculum. Despite the subject gaining well-founded attention by specialists, some teachers may see it as simply another topic to check-off among the list of curricula.

This curriculum unit intends to expose teachers and their students to the statistical process, taking them from point A to point B while developing the reasoning and thinking that is essential to the subject.

Background

Statistics is a methodological tool useful in many fields of study and facets of life with real-world implications. A statistically literate person views data (and the world) with a discerning eye, questioning collection methods, inferring potential misguidance or veracity, and deciding the likelihood of answers. These cognitive processes are the essence of statistics and its important value, but most times it is not what comes across to students.

Most times, teachers approach statistics from a mathematical standpoint conflating the two subjects. Students are not exposed to the critical reasoning portion of statistics as described above. Instead of asking students to engage in the usefulness of the statistics' process, they are asked to answer step-by-step mathematical procedures (e.g. determine the mean, median, and mode of this data set; look up the corresponding p-value). The problem is that rote calculation as it is currently taught "does little to develop statistical thinking" (p.1 SET)

It is helpful to look at the reasons why this happens. One explanation is how teachers were first exposed to the subject in their own schooling. Often times, statistics is taught as a part of a mathematics class at the pre-college level, appearing as just another piece of the mathematical puzzle. Their classes also emphasized procedural mathematics where finding a deterministic answer to a question was the primary focus, rather than learning a coherent set of statistical tools (p.1 SET).

Another reason for this misguided focus comes from the types of statistical questions on large-scale assessments. Although there exists an acknowledgement by the statistics' education community to assess statistical reasoning, most items on these tests still assess procedural competency. Because of this, many teachers and curriculum-developers use these types of questions to guide their instruction. See Chapter 7 in the *Statistical Education of Teachers* for helpful information that contrasts procedural question types with sound conceptual statistics questions. Teachers can also find informative and appropriate statistical assessments online through the Levels of Conceptual Understanding in Statistics (LOCUS) project. Both of these resources help teachers gain an understanding of the statistical skills teachers should aim to teach to their students.

The question remains of how exactly to teach statistics to students. Two important documents will help teachers answer this question: *SET: Statistical Education of Teachers* and the *GAISE: Guidelines for Assessment and Instruction in Statistics Education: A PreK-12 Curriculum Framework*. Both of these were developed by *Statistics Teacher*, an online journal published by the American Statistical Association – National Council of Teachers of Mathematics Joint Committee on Curriculum in Statistics and Probability for Grades K-12.

Statistics Teacher concludes teachers should "emphasize the statistical problem-solving process… through projects that allow teachers to carry out the process from beginning to end" (p. 43 SET). This means teachers can best serve their students by developing their own statistical questions in the classroom, collecting and analyzing the data, and interpreting the results. This allows students to construct their own statistical foundational knowledge at a young age. With this in place, students can further develop these concepts and apply them more generally in various contexts that introduced in school and their lives outside of it.

This process is also supported by the fact that statistics is a highly context dependent subject. In order for students to know and apply any statistical concepts, they must be familiar with whatever context is given to them. For example, if a district curriculum has students comparing the data of a town's 5K and Half-Marathon times (as the case would be on an assessment item in *SET*) , students would first need to understand the context of the details in order to move past and apply statistical reasoning and conceptual knowledge. Therefore, it only makes sense that students first learn the concepts through first-hand experiences with their own data and questions.

Teachers can use the following investigative process and framework, as set forth in the *GAISE,* to create their own statistical activities with their students:

I. Formulate Questions - this unit will have students evaluate questions to understand what qualifies statistical questions and what does not. *The key distinction is the anticipation of variability. "The formulation of a statistics*

question requires an understanding of the difference between a question that anticipates a deterministic answer and a question that anticipates an answer based on data that vary. The question, "How tall am I?'" will be answered with a single height. It is not a statistics question. The question "How tall are adult men in the USA?" would not be a statistics question if all these men were exactly the same height! The fact that there are differing heights, however, implies that we anticipate an answer based on measurements of height that vary. This is a statistics question (p. 11 *GAISE*). As per Common Core and PA State Standards, students will need to recognize statistical questions based on this key distinction.

II. Collect Data - this unit involves students accounting for the variability of arm span lengths in their class. This acknowledgement of variability is a key component of statistical thinking.
III. Analyze Data
IV. Interpret Results

The *GAISE* document can be found at http://www.amstat.org/asa/education/Guidelines-for-Assessment-and-Instruction-in-Statistics-Education-Reports.aspx.

Objectives

- Students will engage in the problem-solving process of statistics by first formulating a statistical question (given by teacher if students need support doing so)
- Recognize a statistical question as one that anticipates variability in the data related to the question and accounts for it in the answers. *For example, "How old am I?" is not a statistical question, but "How old are the students in my school?" is a statistical question because one anticipates variability in students' ages*.
- Distinguish between questions that require a statistical investigation and those that do not.

Students need to understand how statistics works as a measurement of a collection of variable data. Students gravitate to overt measures such as the tallest in the class or the longest arm span? Students need to understand that statistical questions will not only take the longest or tallest, but will also include variable data in the answers. They need to know that they will take all variable data into account to provide an answer to the question. Therefore, to show proof of learning, students will need to explain how they will answer a statistical question with variable data as evidence. This is done in two ways: graphing and evaluating the appearance of it, and/or mathematical approaches such as comparison of means or medians.

Moving beyond simply identifying statistical questions as ones that include variable data, students will evaluate statistical questions to see which is most clear and understandable to answer the question raised. They will decide which question is best and explain why. For example, students will be asked to identify which group has the longest arm span - boys or girls.

If given the following, "Who has the longest arm span in the class?" or "What is the shortest arm span in the class?", students will be prompted to express those questions will only provide one answer and will not ultimately answer the statistical question posed. If given a comprehensive and accurate statistical question, students will have to explain these questions are both clear, aligned to the statistical question, and will include variable data as evidence in the answer.

- Students will engage in the statistical problem-solving process by collecting data

To develop statistical thinking, students need to explore data and decide how to approach it. For example, students will need identify appropriate variables for addressing a statistical question and distinguish between categorical and quantitative variables.

- Students will engage in the statistical problem-solving process by analyzing and displaying
- Display numerical data in plots on a number line, including dot plots, histograms, and box plots.

To effectively analyze data and effectively answer statistical questions, students must be able to sketch graphs that accurately represent the data in comprehensible ways. Students must be able to use labels and scales appropriately, in addition to contrasting mediums when comparative data is considered.

Students (and teachers too) must also understand what graphs are appropriate, or not appropriate, for answering a statistical question. For example, if teachers and students need to compare data to answer a statistical question (e.g. boys vs. girls arm span length), effective graphical displays are back-to-back stem and leaf plots, histograms, dot plots, or box plots. An example of an inappropriate graph would be the bar graph which is helpful only when exploring categorical data. Students need to understand what graph will help them answer their statistical questions, and they need to understand how to effectively create comprehensible versions of them.

Students will therefore create stem and leaf plots, histograms, dot plots, and box plots and use to it to explain how it reflects the data in the context of the statistical question. In order to increase the effectiveness of the graph, students will include contrast (where appropriate), labels, and scale.

Students will use these graphical or numerical summaries to explain the structures of the data. Students will use reasoning strategies to compare groups based on quantitative data (i.e. compare shapes, center, and variability.

- Students will engage in the statistical problem-solving process by interpreting data

After students have collected and analyzed data, they must be prompted to consider how their results might apply to other populations. They must also think about the implications of data collection methods.

Strategies

- **5e Lesson Plan** - http://enhancinged.wgbh.org/research/eeeee.html
- **Think-Write-Pair-Share** - https://www.nasa.gov/sites/default/files/files/4-TWPS_Template.pdf

Classroom Activities, Resources, and Appendices

Unit Overview:

Prerequisite: Students should be familiar with measuring in both centimeters and inches although this is easy enough to introduce along the way with this lesson; students should also have familiarity with graphs and using data to create graph although this will not be necessary if the teacher demonstrates this process while students copy or observe the teacher

Objective: Use the statistical problem-solving process (formulate questions, collect data, analyze data, and interpret results) to answer the statistical question: Who has a bigger arm span in our class - boys or girls?

Specific Outcomes:
SWBAT evaluate various questions based on a scenario to decide which one is the best statistical question that will determine who typically has the bigger arm span in the class (statistical questions must include variability, unlike mathematical questions which have a deterministic answer).
SWBAT design (or select from a set of choices depending on their current understanding) a plan that will allow them to answer the statistical question.
SWBAT implement the plan, organize, and analyze the data by describing what groups graphically have more or less and what are the differences if any
SWBAT express the data in a back-to-back stem and leaf plot or histogram (if students are unfamiliar, the teacher should be prepared to create these while students observe or reproduce the teacher's example)
SWBAT use the graph and the data to answer the question by stating specific comparative observations from the graph

Materials: Scenario posted on overhead or on individual cards for students to read; Sets of four questions on individual cards labeled QUESTION 1, 2, 3, and 4 (see below for question information); Measuring tape - you will need to measure all students so remember to consider

the approximate longest arm span in the class and account for when selecting/creating your tape measure. If not available, you can use a piece of masking tape on the wall; then, use a ruler to copy markings onto the tape.

Lesson 1 - Recognizing Statistical Questions

Common Core Standard included in this lesson: Grade 6: Statistics and Probability: Develop understanding of statistical variability.

CCSS.MATH.CONTENT.6.SP.A.1 Recognize a statistical question as one that anticipates variability in the data related to the question and accounts for it in the answers. *For example, "How old am I?" is not a statistical question, but "How old are the students in my school?" is a statistical question because one anticipates variability in students' ages.*
http://www.corestandards.org/Math/Content/6/SP/A/1/

Post the following scenario on overhead projector or handwritten on board:
Scenario: *Kids in 6th grade experience growth spurts - a period of time when your body (legs, arms, bones, etc.) grows quickly. It is known that boys and girls begin this process at different times with girls often beginning around age ten whereas boys begin at approximately age 12. One way to measure growth is to look at a person's arm span (i.e. the distance between your outstretched arms from fingertip to fingertip). If our class wanted to test this established idea, what statistical question could we ask about arm span?*

Engage students: Give sets of question cards (three incorrect, one correct) to groups or pairs. Students are instructed to read each and evaluate which statistical question would best explore the research topic. Students should be prepared to explain why they choose the statistical question. Explanations must highlight a contrast in the first two questions being deterministic (i.e. only providing one answer) and reason the fourth question will compare the two groups with variable data.

1. What is the longest arm span measure in the class?
2. What is the shortest arm span measure in the class?
3. What is the average arm span of both boys and girls combined?
4. Who has a bigger arm span in the class - boys or girls? (CORRECT CHOICE: most useful for exploring this statistical topic)

Exemplary response:
We want to compare boys and girls arm spans in our class. QUESTION 1 and 2 would only give you two numbers and it does not include any information about comparing boys and girls. QUESTION 3 would include everyone's arm span measures so you could not tell if the girls had bigger arm span or if the boys do. QUESTION 4 is the best question because it would allow you

to compare the groups' typical arm span (accounting for all students in both groups) to see who is bigger right now - girls or boys.

Share out responses. Transition students attention to the lesson objective.

Explore concept: Tell students they will now look at these questions and determine which are considered STATISTICAL and NOT STATISTICAL. Students can use the following flowchart questions to guide them:

- Will the answer include only one piece of information: For example, "How old am I?" If so, then the question is NOT STATISTICAL.
- Will the answer include different pieces of information (variability)? For example, "How old are the students in my school?" If so, then the question is a STATISTICAL QUESTION.

Students are instructed to make these determinations with a partner and then post their responses on a T-Chart under STATISTICAL or NOT STATISTICAL for all students to see and observe. Give students time to complete this task, and then review their responses.

Explain concept: Tell students they will further explore STATISTICAL AND NON-STATISTICAL QUESTIONS by looking at list of questions, noting if it is statistical or not, and then explaining why they gave their answer. Students can work with their partner or independently.

Questions to evaluate:

1. How old are you? (NOT STATISTICAL - would only include one answer)
2. What is the most common way kids in our class travel to school each day - walk, bus, car, bicycle, or other? (STATISTICAL - answer would include the different ways students travel to school, and then determine which as the most of the given groups)
3. Are you male or female? (NOT STATISTICAL - only includes one response)
4. Among the sixth graders in the state of Pennsylvania, which group has more - boys or girls? (STATISTICAL - answer would account for the number of boys and girls in the state and then compare them to see which is more)
5. What is the average arm span among sixth grade boys in the state of Pennsylvania? (Statistical - considers all boys' arm spans and gives mathematical average)
6. About how long does it take you to travel to school in minutes? (NOT STATISTICAL - answer would include one response)
7. What is the average travel time to school for students in our class? (STATISTICAL - considers the variability of all students' travel times)

8. What is the most popular sport among students in our class - soccer, baseball/softball, running, basketball, or other? (STATISTICAL - considers all of the different responses of students in the class)
9. In which month were you born? (NOT STATISTICAL - only includes one answer)
10. What is the most popular birthday month in our class? (STATISTICAL - includes all variable student responses)

Elaborate with follow-up questions:

Have students complete the following sentence frame:

A. Statistical questions are answered with _____ (variability/many different pieces of information that answer the question/variable data that answer the question).
B. A non-statistical question, like "How old am I?" is not a statistical question because the it _____ (will only include one response/answer).
C. The main difference between a statical question like "How old are the students in my school?" and a non-statistical question like "How old am I?" is _____ (the first statistical question will include all of the different ages or variable ages in providing the answer, whereas the second non-statistical question will only give one answer).

Evaluate: During this time, students will evaluate their knowledge level of the lesson concept. Students will give a quantitive rating using the following criteria:

1. I need more help. I still do not understand what makes a question a statistical question.
2. I need a little bit more help. I understand today's material, but I cannot explain what makes a question statistical.
3. I know what a statistical question is and I can explain it to someone who might need help understanding it.

Lesson 2 - Designing and Implementing a plan to collect data

Common Core Standard included in this lesson: Grade 6: Statistics and Probability:

Develop understanding of statistical variability.

CCSS.MATH.CONTENT.6.SP.A.2 Understand that a set of data collected to answer a statistical question has a distribution which can be described by its center, spread, and overall shape.
http://www.corestandards.org/Math/Content/6/SP/A/2/

Summarize and describe distributions.
CCSS.MATH.CONTENT.6.SP.B.4
Display numerical data in plots on a number line, including dot plots, histograms, and box plots.
http://www.corestandards.org/Math/Content/6/SP/B/4/

Engage: Post the following on the board: *THINK-WRITE-PAIR-SHARE: Yesterday, we learned about statistical questions. Today, we will design and implement a plan to answer a specific statistical question about our class: Who has a bigger arm span in the class - boys or girls? What can we do to answer this question?*

Students work in groups to write down their plan. Teachers can provide scaffolds by giving the following prompt, omitting the text in parentheses:

To compare boys and girls arm span, first we should (measure and record arm span lengths of both groups). Then, we should (use the data to create a graph OR find the average/mean for both groups to observe who has typically bigger arm span).

Students share their responses as teacher takes note. The students and the teacher should ultimately agree they should measure all students arm spans, record and organize the data, and then graph or compute measures of center (mean or median) to compare the groups.

Implement the plan:

Conduct measurements and record data: The teacher can review how to measure arm span with students, and also record appropriate data in a whole class table labeled boys and girls (this can be entered into a spreadsheet shown on a projector or on chart paper). Students copy the aggregated data for their own purposes.

Graph data: The teacher and the students can decide which graphing approach is best to help them answer the statistical question. If students do not offer specific ways, the teacher can provide the outline for creating a back-to-back stem and leaf plot and a histogram.

The teacher can guide students in using the data to create both the stem and leaf plot and the histogram.

Interpret results:

Students make observations using *I notice* statements. Students make observations by referring to the graph by describing who has more or less.

Appendices

Standards

CCSS.Math.Content.6.SP.A.1: Recognize a statistical question as one that anticipates variability in the data related to the question and accounts for it in the answers. *For example, "How old am I?" is not a statistical question, but "How old are the students in my school?" is a statistical question because one anticipates variability in students' ages*.

CCSS.Math.Content.6.SP.B.4: Display numerical data in plots on a number line, including dot plots, histograms, and box plots.

CCSS.Math.Content.6.SP.B.5: Summarize numerical data sets in relation to their context, such as by:

CCSS.Math.Content.6.SP.B.5.a: Reporting the number of observations.

CCSS.Math.Content.6.SP.B.5.b: Describing the nature of the attribute under investigation, including how it was measured and its units of measurement.

Annotated Bibliography

Bargagliotti, A. E. (2017). Statistical Education of Teachers. *Statistics Teacher*. Retrieved from http://www.statisticsteacher.org/files/2017/03/EDU-SET.pdf

Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., Scheaffer, R. (2005). *Guidelines for Assessment and Instruction in Stastics Education (GAISE) Report*. Retrieved from http://www.statisticsteacher.org/files/2017/02/GAISEPreK-12_Full.pdf.