Asking Statistical Questions: A Teachers' Guide

1. Where did the term "statistical question" come from?
To the best of our knowledge, it was coined for the Common Core math standards. Other terms for the same (or similar) concept are "Investigative question", "Research question", "Big Question".

2. What is a statistical question?

A statistical question is a question about a real-world context that

a) addresses variability and

b) can be answered with data.

Non-statistical: What is LeeAnn's favorite color? While this can be answered with data (by asking LeeAnn), it does not address variability, since LeeAnn presumably has only one favorite color.

Statistical: What is the most common favorite color in our classroom? This can be answered by collecting data (asking each student his or her favorite color) and recognizes that different students may have different answers. (We couldn't just say "What is the class's favorite color?" unless we defined how to address the issue of variability ---will we choose the color with the most 'votes'?)

Non-statistical: If a tree falls in a forest, and no one can hear or see it, and there are no sensors human or otherwise around, does it make a sound? This cannot be answered with data, and so is not statistical.

Non-statistical: What is the airspeed velocity of an unladen swallow? This is non-statistical because it assumes all unladen swallows have the same velocity. A follow-up question, such as "African or European?" might point out that there is variation between species, but were we to collect data, we would see variation between individuals. This question could be rephrased into a statistical question:

Statistical: What is the *typical* (or most common, or most extreme, or....) of an unladen European swallow.

3. What else indicates a statistical question?

Suppose the question *can* be answered with data? How much data is required? If it can be answered with a single observation, then it is not statistical. If it requires looking at all of the data available for a variable (or a substantial portion), then it is statistical.

Non-statistical: Did anyone select Red as a favorite color? We can answer this with a single observation.

Statistical: What proportion of people selected Red as their favorite color? Now we need to examine everyone.

Exercise: Choose a data set and generate as many statistical questions as you can in 5 minutes.

4. Are there grey areas?

Yes indeed! But remember: the purpose is not to classify questions as "statistical" or "non", but to develop a healthy sense of curiosity that will help us analyze and understand our data. In other words, it is more important that students learn to *ask* statistical questions, than to know whether or not what they're doing is statistical questioning.

5. Are there other types of questions?

Within the broad category of Statistical Questions, we can make narrower distinctions. These are not so important for students to learn, but can be helpful in guiding class conversations.

Summary vs. Relationship questions. A summary question is one that involves a single variable. (The favorite color question is an example of that). A relationship question considers two or more variables. Is red more popular among boys than girls? is a relationship question because it involves the relationship between gender and favorite color.

Relationship questions include a sub-class of comparative questions, in which students compare themselves (a single observation) to the group. "I wonder whether I throw away more recyclables than the rest of the class?" "I wonder if the sugar content of my snacks is about the same as the rest of the class?" At first glance, these may not seem statistical, since they can be answered with a single observation. However, because they compare to the entire class, they also address the variability in the classroom.

A Matter of Scale. There are big questions and then there are BIG questions. Is human activity responsible for global climate change? How has the geographical distribution of native California lizards changed over the last 5 years is a statistical question, but is BIG in that it probably has to be broken into many smaller questions. ("How do we define a 'geographical distribution'? What sorts of lizards do we include or exclude? How will we measure change in geographical distribution?") Despite this, it is compelling, and maybe we can bite off a chunk and answer a smaller piece to address this .For instance: "How has the number of reported alligator lizard sightings changed over the last five years in Los Angeles County?"

Questions can also vary in the level of detail. Consider this pair:(a) Is there a relationship between the sugar content of food and its perceived healthy-level? (b) Do foods with high sugar content tend to receive a lower healthy-level rating? The first question is somewhat vague and general: it simply asks whether a relationship between two variables (sugar content and health-level rating) exists. The second one goes further and proposes a possible direction for the relationship: high sugar means low health rating.

At some level, questions might be so vague that they are worthless: are any of the variables in our data set related to any others? is an example. Whenever possible, get students to ask questions about *these* data—something that couldn't be applied to any generic data set.

Questions about our data vs. inferential questions.

Inferential questions are questions about the wider world. These are questions that are sometimes prompted by the data. For instance, compare these two: (a) "In our class, was Red a more popular color among the girls than it was among the boys?" (b) "Do girls prefer red more than boys prefer red?"

The first question is about the data at hand. The next is about all girls and all boys and is much broader in scope.

Both types of questions are good. But it is important for students to know which type they are asking and answering, and to consider whether they data at-hand are appropriate for answering.

6. Are there any prompts or stems I can use to get my students started?

BEFORE students see data: Get them to ask questions that begin with "I wonder if...." These questions should be limited to the context. I wonder if there is more littering on Friday than on Monday on campus? I wonder if snacks are healthier when eaten at home?

These questions can be used to motivate data collection (whet their appetite for learning) or can be used after data collection to get them interested in analyzing the data. If after data collection,

a good follow-up is "What data did we collect that addresses that question?" If the answer is "none", then it may be a good question, and perhaps might inspire them to do their own data-collection campaign!

AFTER students see data: The best place to start is with "I see that...." statements. These are facts about the data. "I see that in for our class, food eaten at home costs less than food eaten while at school." "I see that I threw away more recyclables than the rest of the class." "I see that my sugar content of my snacks is typical for this class (about average)." It is important that students can explain how they reach these conclusions. They should have a graphical or numerical summary of the data that supports this statement. You should make sure that they are NOT making wider inferences about the larger world. You can press them on this by asking them for specifics: "You said that food eaten at home costs less than food eaten while at school. Do you think this is always true? Do you think it will be true next month? for the class next door? At all schools?" [Answers: to all questions, the best answer is "we don't know because we don't have data that tells us. But I can make a guess or a prediction about what we would see if we did get data."] Get them to be specific about which group they and sets of circumstances they are talking about, and help them become aware that their statements must apply ONLY to the data collected and not beyond.

AFTER they have described what they see, they can make predictions (state hypotheses) about the larger world. These statements begin with "I expect that..." or "I wonder if..." "I wonder if this is true for all classes?" "I expect that if I were to collect data tomorrow, my graph would look very much like it does today."