Lesson 16: Estimating population proportions

Goals

- Compare (orally) proportions for the same category from different samples of a population.
- Comprehend that the term "proportion" refers to a number between 0 and 1 that represents the fraction of the data within a certain category.
- Use the proportion of a random sample that is within a certain category to make inferences about the population, and explain (orally and in writing) the reasoning.

Learning Targets

- I can estimate the proportion of population data that are in a certain category based on a sample.

Lesson Narrative

In the previous lesson, students used samples to estimate measures of centre of a population. In this lesson, students estimate population proportions. The term **proportion** is used in statistics to refer to a number from 0 to 1 that represents the fraction of the data belonging to a given category.

Students see that if a sample is representative of the population, then we can use proportional reasoning to make predictions about the population. However, students need to understand that, due to sampling variability, these predictions are estimates, not exact answers like they get when working with actual proportional relationships.

The activity about examining a distribution of proportions from many different samples is included as an optional opportunity to deepen students' understanding of sampling variability.

Addressing

- Convert a rational number to a decimal using long division; know that the decimal form of a rational number terminates in 0s or eventually repeats.
- Analyse proportional relationships and use them to solve real-world and mathematical problems.
- Use random sampling to draw inferences about a population.
- Use data from a random sample to draw inferences about a population with an unknown characteristic of interest. Generate multiple samples (or simulated samples) of the same size to gauge the variation in estimates or predictions. For example, estimate the mean word length in a book by randomly sampling words from the book; predict the winner of a school election based on randomly sampled survey data. Gauge how far off the estimate or prediction might be.

- Use measures of centre and measures of variability for numerical data from random samples to draw informal comparative inferences about two populations. For example, decide whether the words in a chapter of a year 8 science book are generally longer than the words in a chapter of a year 5 science book.

Building Towards

- Understand that statistics can be used to gain information about a population by examining a sample of the population; generalisations about a population from a sample are valid only if the sample is representative of that population. Understand that random sampling tends to produce representative samples and support valid inferences.
- Use data from a random sample to draw inferences about a population with an unknown characteristic of interest. Generate multiple samples (or simulated samples) of the same size to gauge the variation in estimates or predictions. For example, estimate the mean word length in a book by randomly sampling words from the book; predict the winner of a school election based on randomly sampled survey data. Gauge how far off the estimate or prediction might be.

Instructional Routines

- Stronger and Clearer Each Time
- Compare and Connect
- Discussion Supports
- Notice and Wonder

Required Materials

Paper bags
Pre-printed slips, cut from copies of the blackline master

Required Preparation

Print and cut up slips from the Reaction Times blackline master. Prepare one set of slips in a paper bag for every 2 students.

Student Learning Goals

Let's estimate population proportions using samples.

16.1 Getting to School

Warm Up: 5 minutes

The purpose of this warm-up is for students to compute the fraction of individuals whose responses fall in a specified category. This activity gives students time to think about how to compute these fractions from qualitative data.

For the second and third questions, students may debate whether to include the 10 minute times or not. According to the wording of the question asked, it does ask for *more than* 10 minute times, so maybe exactly 10 minutes should not count (since $10 \not> 10$). On the other hand, all of the values are listed as whole numbers, so a student who takes 10 minutes and 1 second to get to school may have rounded down to 10, but should have been counted. Noticing the large difference in answers for the third question, it may be worth clarifying the data in this instance, even for an estimate.

Monitor for students who include the 10 minute times for the second and third questions as well as those who do not.

Launch

Give students 2 minutes of quiet work time followed by a whole-class discussion.

Student Task Statement

A teacher asked all the students in one class how many minutes it takes them to get to school. Here is a list of their responses:

20	10	15	8	5	15	10	5
20	5	15	10	3	10	18	5
25	5	5	12	10	30	5	10

What fraction of the students in this class say:

a. it takes them 5 minutes to get to school?

b. it takes them more than 10 minutes to get to school?

1. If the whole school has 720 students, can you use this data to estimate how many of them would say that it takes them more than 10 minutes to get to school?

   Be prepared to explain your reasoning.

Student Response

a. $\frac{7}{24}$

b. $\frac{9}{24}$ or equivalent (or $\frac{15}{24}$ or equivalent if 10 minutes is included)

1. About 270 students (or 450 students if 10 minutes is included) since $\frac{9}{24}$ of 720 is 270 (or $\frac{15}{24}$ of 720 is 450)

Activity Synthesis

Select students to share their methods for calculating the solutions. Include previously identified students who did or did not include the 10 minute values in their calculations.

If it does not arise during the discussion, explain that answering the third question with the data at hand is only accurate if the sample data is representative of the school. It is possible that the class happens to contain only students who get a ride to school, but much of the school rides the bus.

16.2 Reaction Times

15 minutes

In previous lessons, students examined the estimation of the mean and median for populations using data from a sample. In this activity, students apply similar reasoning to estimating the proportion of a population that matches certain characteristics. Students collect a sample of 20 reaction times and compute the fraction of responses in their sample that are in a given range. Then, in the discussion, students compare their estimations to the known population proportion and use the class's proportions to gauge the accuracy of their estimate.

Instructional Routines

- Compare and Connect

Launch

Arrange students in groups of 2. Distribute bags of slips cut from the blackline master.

Tell students that, in statistics, a **proportion** is a number between 0 and 1 that represents the fraction of the data that fits into the desired category. For example, with the data: {Yes, Yes, Yes, No, Maybe} the proportion of "Yes" answers is $\frac{3}{5}$ or 0.6.

Introduce the context: All 120 sixth formers at a secondary school were asked to click a button as soon as they noticed a box change colour and the response time was recorded in seconds. These 120 response times represent the population for this activity. Their responses are written on the slips of paper in the bag.

When selecting a sample of 20, each value does not need to be replaced before taking the next one.

Allow students 10 minutes of partner work time followed by a whole-class discussion.

Action and Expression: Internalise Executive Functions. Chunk this task into more manageable parts. After students have solved the first 2-3 problems, check-in with either select groups of students or the whole class. Invite students to share the strategies they have used so far as well as any questions they have before continuing.
Supports accessibility for: Organisation; Attention

Anticipated Misconceptions

The reason a proportion is between 0 and 1 is because we are specifically looking for the proportion of a desired category out of the whole group. This is a "part out of whole" fraction, which will be a number between 0 and 1.

Student Task Statement

The track coach at a secondary school needs a student whose reaction time is less than 0.4 seconds to help out at track meetings. All the year 13s in the school measured their reaction times. Your teacher will give you a bag of papers that list their results.

1. Work with your partner to select a random sample of 20 reaction times, and record them in the table.

2. What **proportion** of your sample is less than 0.4 seconds?
3. Estimate the proportion of all year 13s at this school who have a reaction time of less than 0.4 seconds. Explain your reasoning.
4. There are 120 year 13s at this school. Estimate how many of them have a reaction time of less than 0.4 seconds.
5. Suppose another group in your class comes up with a different estimate than yours for the previous question.
    a. What is another estimate that would be *reasonable*?
    b. What is an estimate you would consider *unreasonable*?

Student Response

Answers vary. Sample response:

1. 0.5, 0.44, 0.51, 0.38, 0.42, 0.79, 0.39, 0.46, 0.34, 0.3, 0.36, 0.41, 0.31, 0.82, 0.35, 0.36, 0.48, 0.72, 0.74, 0.45
2. $\frac{8}{20} = 0.4$
3. 0.4 since the sample was chosen at random and is likely representative.
4. 48 since $0.4 \times 120 = 48$.
5.
    a. 50

b. 100

Activity Synthesis

The purpose of the discussion is for students to see how multiple sample proportions can help revise their estimates and give an idea of how accurate the individual estimates from samples might be.

Ask the groups to share the proportion from their sample that had fast reaction times and display the results for all to see.

Some questions for discussion:

- "Using the class's data, how accurate do you think your group's estimate is? Explain or show your reasoning." (Students should mention the variability of the proportions from the samples influencing the accuracy of the estimate.)
- "The actual proportion for this population is 0.5. How close was your estimate? Explain why your estimate was not exactly the same." (Each sample might be slightly different since they do not include all of the values, but they should be close.)
- "If each group had 40 reaction times in their samples instead of 20, do you think the estimate would be more or less accurate?" (The estimate should be more accurate since there is more information available.)

Representing, Speaking: Compare and Connect. Invite students to create a visual display of their work to share with other students. Displays should include a representation of the sample reaction times, the proportion of fast reaction times, and an estimate for the number of twelfth-graders with fast reaction times. As students investigate each other's work, ask students to compare their samples and proportion of fast reaction times. Listen for and amplify the language students use to justify the accuracy of their estimate for the proportion of fast reaction times using the definition of a random sample. This will help students compare their sample proportions in order to see how accurate their population estimates might be.
Design Principle(s): Optimise output (for comparison); Cultivate conversation

16.3 A New Comic Book Hero

15 minutes

In the previous activity, students collected their own sample and computed an estimate for the population proportion based on the sample. In this activity, students use a different context to practise exploring proportions from samples and their extension to populations. The optional activity following this one continues with this context exploring sampling variability.

Instructional Routines

- Stronger and Clearer Each Time

- Notice and Wonder

Launch

Keep students in groups of 2.

Tell students that three comic books, *The Adventures of Super Sam*, *Beyond Human,* and *Mysterious Planets,* are all planning to add a new superhero to their stories. A survey was sent to dedicated readers of each series to ask what type of ability the new hero should have: fly, freeze, or another power.

Display the tables from the Task Statement for all to see. Ask students, "What do you notice? What do you wonder?"

Students may notice:

- There are 4 different responses: fly, freeze, super strength, and invisibility.
- The number of responses for each of the 4 different responses.
- There are 20 responses.

Students may wonder:

- Will the decision for the new hero's power be based only on this survey?
- What proportion chose each of the different powers?
- Is this sample of 20 responses representative of the population?

Give students 5–7 minutes of partner work time followed by a whole-class discussion.

Representation: Internalise Comprehension. Demonstrate and encourage students to use colour coding and annotations to highlight connections between representations in a problem. For example, use colour to highlight connections between the same qualitative data in a survey of new hero special ability.
Supports accessibility for: Visual-spatial processing

Student Task Statement

Here are the results of a survey of 20 people who read *The Adventures of Super Sam* regarding what special ability they think the new hero should have.

response	what new ability?
1	fly
2	freeze
3	freeze
4	fly
5	fly
6	freeze
7	fly
8	super strength
9	freeze
10	fly
11	freeze
12	freeze
13	fly
14	invisibility
15	freeze
16	fly
17	freeze
18	fly

19	super strength
20	freeze

1. What proportion of this sample want the new hero to have the ability to fly?

2. If there are 2 024 dedicated readers of *The Adventures of Super Sam*, estimate the number of readers who want the new hero to fly.

   Two other comic books did a similar survey of their readers.

   - In a survey of people who read *Beyond Human*, 42 out of 60 people want a new hero to be able to fly.
   - In a survey of people who read *Mysterious Planets*, 14 out of 40 people want a new hero to be able to fly.

3. Do you think the proportion of all readers who want a new hero that can fly are nearly the same for the three different comic books? Explain your reasoning.

4. If you were in charge of these three comics, would you give the ability to fly to any of the new heroes? Explain your reasoning using the proportions you calculated.

Student Response

1. $\frac{8}{20} = 0.4$

2. About 810 since $0.4 \times 2\,024 \approx 810$

3. The proportion for *Mysterious Planets* seems close to the proportion for *The Adventures of Super Sam* but not for *Beyond Human*, since the proportion for *Mysterious Planets* is 0.35 which is close to the 0.4 for *The Adventures of Super Sam* while the proportion for *Beyond Human* is 0.7.

4. Answers vary. Sample responses:

   - I would give the new heroes of all of the comics the ability to fly. Based on these results, about 40% of the readers of *Mysterious Planets* and *The Adventures of Super Sam* want the hero to have that new ability, so I think it would satisfy many of their readers. About 70% of the readers of *Beyond Human* want this new ability, so they will most likely be happy with this choice.
   - Only for the new hero of *Beyond Human*. In the data for *The Adventures of Super Sam*, more people voted for freeze than fly, so I don't think it would make sense to choose flight for that comic rather than freeze. It is hard to tell for *Mysterious Planets* without the actual data. The votes might have been split among the other choices more evenly and flight might have been the most popular choice.

Activity Synthesis

The purpose of the discussion is for students to recognise the value of a proportion and why a sample may be necessary to estimate the proportion for the population.

Some questions for discussion:

- "Explain why you think a sample was used instead of the population for this situation." (The population is too large to ask all of the readers about their preference. Also, the authors may want the power of the new hero to be a surprise to some readers, so they want to get some information without telling everyone about what is coming.)
- "Based on the proportions computed in each of these samples, would you suggest to the authors of any of the comic books to give their new hero the ability to fly?"
- "Although the proportion of responses from the *Mysterious Planets* sample who chose flight was only 0.35, it was the most popular choice (freeze, super strength, invisibility, and other powers split the remaining votes). Does this information change your answer to the previous question?"
- "Of the three estimates, which do you think is most accurate? Explain your reasoning." (*Beyond Human* has the largest sample, so it may be the most accurate.)

Writing, Speaking, Listening: Stronger and Clearer Each Time. Use this routine to help students improve their writing by providing them with multiple opportunities to clarify their explanations through conversation. Give students time to meet with 2–3 partners to share their response to: "Would you give the ability to fly to any of the new heroes?" Provide prompts for feedback that will help students strengthen their ideas and clarify their language. For example, "Which proportions did you consider?", "How did you calculate each proportion?", "Can you say more about…?", etc. Give students 1–2 minutes to revise their writing based on the feedback they received.
Design Principle(s): Optimise output (for explanation)

16.4 Flying to the Shelves

Optional: 10 minutes

This optional activity goes beyond year-level expectations to deepen students' understanding of sampling variability.

This activity continues the comic book context introduced in the previous activity. There is not a measure of variability such as range or IQR for proportions since the data are qualitative rather than quantitative, so other methods must be employed to determine the accuracy of an estimate. Students look at dot plots showing the results from multiple samples to gauge the accuracy of the estimates for population proportions. This activity will provide a foundation for work in later years.

Instructional Routines

- Discussion Supports

Launch

Arrange students in groups of 2.

Help students make sense of the dot plot. Each dot in the dot plot represents the proportion from a random sample of 20 readers. 50 random samples were taken and the 50 proportions are plotted on the dot plot. For example, the 0.4 proportion from the previous activity would be represented by one of the 4 dots at 0.4 on the first dot plot.

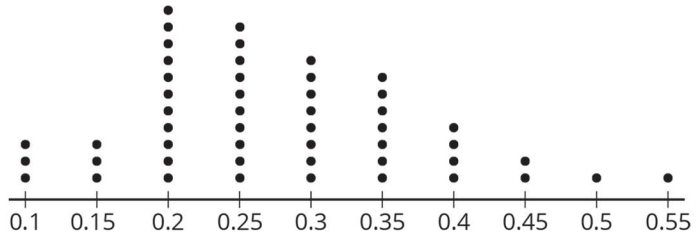
Ask, "Are any of the sample proportions greater than or equal to 0.5? What does this mean?" (Yes, 2 dots are greater than or equal to 0.5. This means that, in those samples, at least half of the people prefer the new hero to have the ability to fly.)

Give students 5–7 minutes of partner work time followed by a whole-class discussion.

Action and Expression: Internalise Executive Functions. Chunk this task into more manageable parts to support students who benefit from support with organisational skills in problem solving. For example, pause to check for understanding after 3-5 minutes of work time.
Supports accessibility for: Organisation; Attention

Student Task Statement

The authors of *The Adventures of Super Sam* chose 50 different random samples of readers. Each sample was of size 20. They looked at the sample proportions who prefer the new hero to fly.



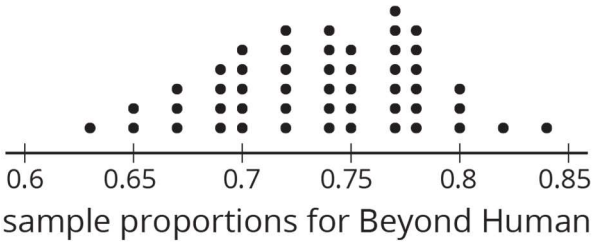
sample proportions for The Adventures of Super Sam

1. What is a good estimate of the proportion of *all* readers who want the new hero to be able to fly?

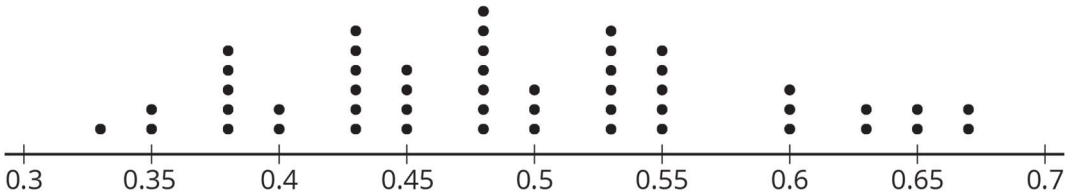
2. Are most of the sample proportions within 0.1 of your estimate for the population proportion?

3. If the authors of *The Adventures of Super Sam* give the new hero the ability to fly, will that please most of the readers? Explain your reasoning.

   The authors of the other comic book series created similar dot plots.



sample proportions for Beyond Human



sample proportions for Mysterious Planets

4. For each of these series, estimate the proportion of all readers who want the new hero to fly.
   - *Beyond Human*:
   - *Mysterious Planets*:
5. Should the authors of either of these series give their new hero the ability to fly?
6. Why might it be more difficult for the authors of *Mysterious Planets* to make the decision than the authors of the other series?

Student Response

1. Answers vary. Sample response: 0.3 since the centre of the distribution is near there.
2. Yes
3. No. Only about 30% of readers seem to want the new hero to fly, so there may be a more preferred super power.
4. Answers vary. Sample response:
   - *Beyond Human*: about 0.73
   - *Mysterious Planets*: about 0.5
5. *Beyond Human* should give the power to fly to its new hero, since almost $\frac{3}{4}$ voted for that option.
6. Not only do we estimate the population proportion to be about 0.5, but the values are quite varied among all the sample proportions.

Are You Ready for More?

Draw an example of a dot plot with at least 20 dots that represent the sample proportions for different random samples that would indicate that the population proportion is above 0.6, but there is a lot of uncertainty about that estimate.

Student Response

Answers vary. Responses should include a majority of the dots greater than or near 0.6, but there should be a lot of variability in the sample proportions.

Activity Synthesis

The purpose of the discussion is for students to talk about how the variability in sample proportions affects their trust in the estimates of the population proportion.

Consider these questions for discussion:

- "When estimating a population mean or median from a random sample, measures of variability from a sample can be used to help gauge the accuracy of the estimate. With proportions there is not a measure of variability in the same way. How did the information in this activity guide your thoughts about the accuracy of the population estimate?" (Many samples were taken and their proportions computed. The variability of these sample proportions showed how much to trust the population estimate.)
- "How does the distribution of values in the dot plots of sample proportions affect your trust in an estimate of population proportion?" (The more variability, the less certainty in the estimate.)
- "How would the distributions change if the number of responses in each sample were increased?" (The centre should remain about the same, the variability should decrease or, in other words, the dots in the dot plot should get closer together towards the centre.)

Speaking: Discussion Supports. Use this routine to support whole-class discussion. For each response or observation that is shared, ask students to restate what they heard using precise mathematical language. Consider providing students time to restate what they hear to a partner, before selecting one or two students to share with the class. Ask the original speaker if their peer was accurately able to restate their thinking. Call students' attention to any words or phrases that helped clarify the original statement. This will provide more students with an opportunity to produce language as they interpret the reasoning of others.
Design Principle(s): Support sense-making

Lesson Synthesis

Consider asking these discussion questions to clarify the main ideas of the lesson:

- "When using data, what is a proportion? How is it calculated?" (A proportion is the fraction of the data that are in a certain category. It is calculated by counting the number of data values in the category and dividing by the total number of data values in the sample.)
- "In order to say that more than half of the people in a sample responded with a certain answer, what would the proportion for that answer be?" (Any value greater than 0.5)
- "A random sample indicates that a 0.45 proportion of people shopping at a certain store prefer wheat bread to white bread. The store has 3 000 customers. Estimate the number of people shopping at the store who prefer wheat bread." (About 1 350 since $0.45 \times 3\,000 = 1\,350$)

16.5 More than 48 Grams

Cool Down: 5 minutes

This cool-down assesses student's understanding of estimating proportions for a population based on a random sample.

Student Task Statement

A chemical engineer is trying to increase the amount of the useful product in a reaction. She performs the reaction with her new equipment 10 times and gets the following amounts of the useful product in grams:

47.1 48.2 48.3 47.5 48.5 48.1 47.2 48.2 48.4 48.3

1. What proportion of the reactions were above the 48 grams threshold?
2. Other chemists typically get 65% of their reactions to produce more than 48 grams. Should the engineer say that she was able to increase the useful product when compared to the other chemists?

Student Response

1. 0.7 since 7 of the 10 reactions had more than 48 grams of the useful product.
2. Answers vary. Sample response: She could be optimistic, but her proportion does not seem far from what others have done. She should run more reactions to be more sure of the improvement. With only 10 values in her data set, 0.7 (and 0.6) is as close to 0.65 as she could get.

Student Lesson Summary

Sometimes a data set consists of information that fits into specific categories. For example, we could survey students about whether they have a pet cat or dog. The categories for these data would be {neither, dog only, cat only, both}. Suppose we surveyed 10 students. Here is a table showing possible results:

option	number of responses
neither dog nor cat	2
dog only	4
cat only	1
both dog and cat	3

In this sample, 3 of the students said they have both a dog and a cat. We can say that the **proportion** of these students who have a both a dog and a cat is $\frac{3}{10}$ or 0.3. If this sample is representative of all 720 students at the school, we can predict that about $\frac{3}{10}$ of 720, or about 216 students at the school have both a dog and a cat.

In general, a proportion is a number from 0 to 1 that represents the fraction of the data that belongs to a given category.

Glossary

- proportion

Lesson 16 Practice Problems

Problem 1 Statement

Tyler wonders what proportion of students at his school would dye their hair blue, if they were allowed to. He surveyed a random sample of 10 students at his school, and 2 of them said they would. Kiran didn't think Tyler's estimate was very accurate, so he surveyed a random sample of 100 students, and 17 of them said they would.

a. Based on Tyler's sample, estimate what proportion of the students would dye their hair blue.

b. Based on Kiran's sample, estimate what proportion of the students would dye their hair blue.

c. Whose estimate is more accurate? Explain how you know.

Solution

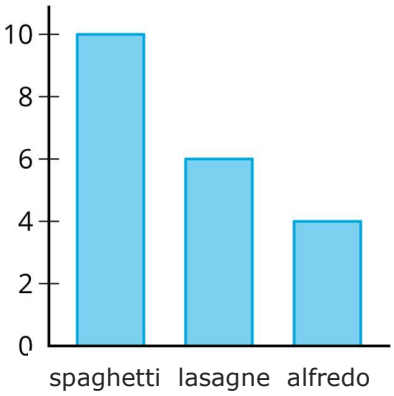
a. 0.20

b. 0.17

c. Kiran's estimate is probably more accurate, because he used a much larger sample than Tyler. Sample proportions from larger samples tend to be more tightly clustered around that value of the population proportion. Note it is still possible for Tyler's estimate to be more accurate, coincidentally.

Problem 2 Statement

Han surveys a random sample of students about their favourite pasta dish served by the cafeteria and makes a bar graph of the results.



Estimate the proportion of the students who like lasagne as their favourite pasta dish.

Solution

0.3 (or $\frac{6}{20}$ or equivalent)

Problem 3 Statement

Elena wants to know what proportion of people have cats as pets. Describe a process she could use to estimate an answer to her question.

Solution

Answers vary. Sample response: Find a random sample of about 50 people and ask them if they have a pet cat. Once the responses are recorded, count the number of "yes" answers and divide that by 50 to get an estimate of the population proportion.

Problem 4 Statement

The science teacher gives daily homework. For a random sample of days throughout the year, the median number of problems is 5 and the IQR is 2. The Spanish teacher also gives daily homework. For a random sample of days throughout the year, the median number of problems is 10 and the IQR is 1. If you estimate the median number of science homework problems to be 5 and the median number of Spanish problems to be 10, which is more likely to be accurate? Explain your reasoning.

Solution

The Spanish estimate is more likely to be accurate. When the measure of variability (the IQR) is larger, it is hard to get a good estimate of the population.

Problem 5 Statement

Diego wants to survey a sample of students at his school to learn about the percentage of students who are satisfied with the food in the cafeteria. He decides to go to the cafeteria on a Monday and ask the first 25 students who purchase a lunch at the cafeteria if they are satisfied with the food.

Do you think this is a good way for Diego to select his sample? Explain your reasoning.

Solution

No, this is not a good way to select a sample. Explanations vary. Sample explanation: Students who are buying lunch at the cafeteria may be choosing to buy their lunch because they like the cafeteria food. Students who bring lunch from home won't be included in the sample.