Chief Reader Report on Student Responses:

2021 AP® Statistics Free-Response Questions

• Number of Students Scored	184,111		
• Number of Readers	1,080		
• Score Distribution	Exam Score	N	%At
	5	29,790	16.2
	4	36,649	19.9
	3	40,153	21.8
	2	31,693	17.2
	1	45,826	24.9
• Global Mean	2.85		

The following comments on the 2021 free-response questions for AP® Statistics were written by the Chief Reader, Dr. Ken Koehler, PhD. They give an overview of each free-response question and of how students performed on the question, including typical student errors. General comments regarding the skills and content that students frequently have the most problems with are included. Some suggestions for improving student preparation in these areas are also provided. Teachers are encouraged to attend a College Board workshop to learn strategies for improving student performance in specific areas.

© 2021 College Board.
Visit College Board on the web: collegeboard.org.

Question #1 **Task:** Exploring Data

Max. Points: 4 **Mean Score:** 1.28

What were the responses to this question expected to demonstrate?

The primary goals of this question were to assess a student's ability to (1) determine values for the five-number summary of data provided in a table and in a dotplot; (2) identify potential outliers using a method based on the five-number summary; (3) identify potential outliers using a method based on the sample mean and standard deviation; and (4) explain why the method based on the five-number summary would tend to identify more potential outliers than the method based on the sample mean and standard deviation for a data sampled from a distribution strongly skewed to the right.

This question primarily assesses skills in skill category 2: Data Analysis. Skills required for responding to this question include (2.C) Calculate summary statistics, relative positions of points within a distribution, correlation, and predicted response, and (4.B) Interpret statistical calculations and findings to assign meaning or assess a claim.

This question covers content from Unit 1: Exploring One-Variable Data of the course framework in the AP Statistics Course and Exam Description. Refer to topic 1.7, and learning objectives UNC-1.I, and UNC-1.K.

How well did the responses address the course content related to this question? How well did the responses integrate the skills required on this question?

- In part (a) most responses identified some components of a five-number summary, but many responses omitted some components. Some responses included components that are not part of the five-number summary, such as the mean, standard deviation, range or interquartile range.

- In part (b-i) most responses correctly identified the two potential outliers and provided justification by calculating the upper and lower outlier criteria. However, some responses incorrectly calculated the outlier criteria by adding/subtracting $1.5 \times \text{IQR}$ to the median, rather than to the appropriate quartile values. Additionally, some responses omitted the calculation of the lower outlier criteria.

- In part (b-ii) most responses correctly identified the one potential outlier and provided justification by calculating the upper and lower outlier criteria. However, some responses incorrectly calculated the outlier criteria by adding/subtracting 1 standard deviation or $3 \times$ standard deviation, rather than $2 \times$ standard deviation, to the mean. Additionally, some responses omitted the calculation of the lower outlier criteria.

- In part (c) many responses correctly indicated that in samples from a more severely right-skewed distribution, the sample mean is pulled more toward the extreme values in the right tail of the distribution and the standard deviation gets larger while the sample quartiles (or median) and IQR are not impacted as much. Some responses mentioned that the sample mean and standard deviation are not resistant to outliers but did not explicitly state that the mean and standard deviation tend to increase as skewness becomes more severe. Many responses did not provide an explanation that linked the impact of a right-skewed distribution on the relevant summary statistics to the impact on the outlier criteria.

© 2021 College Board.
Visit College Board on the web: collegeboard.org.

What common student misconceptions or gaps in knowledge were seen in the responses to this question?

Common Misconceptions/Knowledge Gaps	*Responses that Demonstrate Understanding*
• Failing to identify all components of a five-number summary. • Including summary statistics that are not part of a five-number summary.	• Identification of the five-number summary for the distribution of length of stay requires: minimum is 5 days, Q1 is 6 days, median is 7 days, Q3 is 8 days, and maximum is 21 days.
• Using an incorrect formula to calculate the outlier criteria for the $1.5 \times \text{IQR}$ rule. Failing to calculate the lower outlier criteria.	• $\text{IQR} = 8 - 6 = 2$ days. • Lower fence $= 6 - 1.5 \times 2 = 3$ days. There are no data values less than 3 days. • Upper fence $= 8 + 1.5 \times 2 = 11$ days. The data values of 12 days and 21 days are potential outliers because they are greater than 11 days.
• Failing to communicate how the impact of a right-skewed distribution on the relevant summary statistics had an impact on the outlier criteria	• In a strongly right-skewed distribution, the mean is pulled towards the right, and the standard deviation is inflated. This shifts the interval of non-outliers in Method B towards the right, identifying fewer points as potential outliers. Method A doesn't shift as much because Q3 and the IQR are resistant to outliers.

Based on your experience at the AP® Reading with student responses, what advice would you offer teachers to help them improve the student performance on the exam?

Some teaching tips:

- Encourage students to use correct labels for the values in the five-number summary. Acceptable labels are minimum or min; first quartile or Q1; Median, Med or Q2; third quartile or Q3; maximum or max. Provide opportunities for your students to practice finding these values from a graphical display like a dotplot or stemplot.
- Remind your students to identify the lower boundary as well as the upper boundary for procedures for identifying outliers. The students must identify the value(s) of the outlier(s), not just identify how many outliers there are.
- Discuss the reasoning behind the outlier identification criteria. Both the $1.5 \times \text{IQR}$ rule and the 2 standard deviation rule are dependent on a measure of location and a measure of variability.
    - For the $1.5 \times \text{IQR}$ rule, the boundaries for outliers are dependent upon Q1 and Q3 (location) as well as IQR (variability).
    - For the 2 standard deviation rule the boundaries for outliers are dependent upon $\bar{x}$ (location) and standard deviation (variability).
- Explore how skewness affects the statistics that make up these two outlier rules. For distributions that are more severely skewed, two things are true:
    - Location**:** The mean is "pulled" towards the long tail, and the location of Q1 and Q3 remain relatively unchanged.
    - Variability**:** The standard deviation increases due to the presence of the extreme values in the long tail, and the IQR remains relatively unchanged.
- Discuss how the effects of skewness on location and variability affect the outlier criteria
    - $1.5 \times \text{IQR}$ rule**:** This method creates boundaries for outliers that are based upon values that are relatively unaffected by the skew of a distribution (Q1, Q3, and IQR).

  - 2 standard deviation rule**:** This method creates boundaries for outliers that are based upon values that are affected by the skew of a distribution (mean and standard deviation).
  - The result is that the boundaries for outliers for the $1.5 \times \text{IQR}$ rule tend to define a narrower interval that is not pulled as much toward the long tail than the boundaries for outliers obtained from the 2 standard deviation rule, which will tend to define a wider interval that is pulled more toward the long tail. Therefore, in a skewed distribution, the $1.5 \times \text{IQR}$ rule might identify more data points as potential outliers than the 2 standard deviation rule.
- Provide opportunities for students to explain statistical concepts.
  - When comparing two methods, require a direct comparison.
  - Start with providing students with concrete examples and follow with exercises that help them progress towards generalized conceptual understanding.

What resources would you recommend to teachers to better prepare their students for the content and skill(s) required on this question?

- The *AP Statistics Course and Exam Description* (CED), effective Fall 2020, includes instructional resources for AP Statistics teachers to develop students' broader skills. Please see page 227 of the CED for examples of key questions and instructional strategies designed to develop skill 2.A, describe data presented numerically or graphically. A table of representative instructional strategies, including definitions and explanations of each, is included on pages 213-223 of the CED. The strategy "Quickwrite," for example, may be helpful in developing students' abilities to explain why one method might detect more possible outliers than another in a right-skewed distribution.
- AP Classroom provides two videos for topic 1.7, both focused on the relevant content and skills for this question. The first focuses on skill 2.C, calculating summary statistics ..., and discusses summary statistics that can be used to describe the center and variability of a distribution of quantitative data. The second focuses on skill 4.B, interpreting statistical calculations ..., and discusses outliers, resistant and nonresistant summary statistics, and which measures of center and variability are best for describing a distribution. Both videos are framed in the relevant context of the safety of drinking water in Flint, MI.
- AP Classroom also provides topic questions for formative assessment of topic 1.7 and access to the question bank, which is a searchable database of past AP Questions on this topic.
- The Online Teacher Community features many resources shared by other AP Statistics teachers. For example, to locate resources to give your students practice determining outliers, try entering the keyword "outlier" in the search bar, then selecting the drop-down menu for "Resource Library." When you filter for "Classroom-Ready Materials," you may find worksheets, data sets, practice questions, and guided notes, among other resources.

© 2021 College Board.
Visit College Board on the web: collegeboard.org.

Question #2 **Task:** Collecting Data

Max. Points: 4 **Mean Score:** 0.92

What were the responses to this question expected to demonstrate?

The primary goals of this question were to assess a student's ability to (1) describe bias that could be introduced by allowing subjects to self-report results instead of recording results by fitting each subject with a monitor; (2) explain the statistical benefit of using random sampling to obtain a representative sample of subjects from a target population; and (3) provide an explanation of whether a statistically significant outcome from a particular type of study may be used to justify a conclusion about a cause-and-effect relationship.

This question primarily assesses skills in skill category 1: Selecting Statistical Methods. Skills required for responding to this question include (1.C) Describe an appropriate method for gathering and representing data, and (4.A) Make an appropriate claim or draw an appropriate conclusion.

This question covers content from Unit 3: Collecting Data of the course framework in the AP Statistics Course and Exam Description. Refer to topics 3.2, and 3.4, and learning objectives DAT-2.B, and DAT-2.E.

How well did the responses address the course content related to this question? How well did the responses integrate the skills required on this question?

- In part (a) most responses were able to indicate a bias that self-reporting data would introduce and provide a reason for it. However, not many responses linked potential for bias to systematic underreporting or systematic overreporting of miles walked across most subjects. Furthermore, very few responses linked the bias to using a sample statistic (e.g., sample mean miles walked) to estimate a relevant population parameter (e.g., population mean miles walked).

- In part (b) many responses indicated that a representative sample allowed for results of the study to be generalized to the target population. However, quite a few responses discussed the ease and efficiency of taking a representative sample as opposed to a census, which would be true of a non-representative sample as well. Very few responses were written in the context of the problem by indicating that a representative sample allowed for estimation or inference about cholesterol levels, or inference about the relationship between cholesterol levels and miles walked, in the target population.

- In part (c) many responses correctly indicated a causal inference cannot be made from an observational study. However, many responses argued that confounding variables were not controlled for but failed to establish that a confounding variable must be associated with cholesterol level AND also associated with amount of walking. Furthermore, some responses stated that a claim of a cause-and-effect relationship would be valid based on the result of the significance test.

What common student misconceptions or gaps in knowledge were seen in the responses to this question?

Common Misconceptions/Knowledge Gaps	*Responses that Demonstrate Understanding*
• Many responses did not establish a systematic overreporting or systematic underreporting of miles walked that results in a biased estimate.	• Many more subjects would report a higher number of miles walked than they actually walked, while relatively few subjects would report a lower number of miles walked than they actually walked.

© 2021 College Board.
Visit College Board on the web: collegeboard.org.

• Very few responses linked the bias to an estimate of a relevant population parameter.	• This would result in the sample mean miles walked to be an overestimate of the true mean miles walked for all adults in the target population.
• Many responses discussed a benefit of using a representative sample in general terms and not with respect to this study.	• A representative sample allows results of the study to be generalized to the target population. This allows us to use the results of the study to draw conclusions about the difference in cholesterol levels for those who walk fewer miles per day and those who walk more miles per day in the target population.
• Some responses indicated that it was valid to make a claim about a cause-and-effect relationship simply because the hypothesis test showed a statistically significant result.	• No, this would not be a valid claim because the researchers did not randomly assign the amount of walking to the subjects.
• Some responses did not make it clear whether it would be valid to claim that increased walking causes a decrease in average cholesterol levels.	• No, it would not be valid to make the claim that increased walking causes a decrease in average cholesterol levels. This was an observational study, and a cause-and-effect relationship cannot be established from an observational study.
• Many responses that used a confounding argument did not clearly convey the idea of confounding.	• No, there are potential confounding variables that were not controlled for in this study. It is possible that those with a healthy diet tend to walk more than those with an unhealthy diet. It is reasonable to think that those with a healthy diet tend to have lower cholesterol levels and those with an unhealthy diet tend to have higher cholesterol levels. If this is the case, researchers won't be able to determine if the reduced levels of cholesterol were due to the increased amount of walking or to the healthier diet.

Based on your experience at the AP® Reading with student responses, what advice would you offer teachers to help them improve the student performance on the exam?

- When asked to describe bias that could result from the method of collecting data, students should be encouraged to do three things: 1. Identify a source of the bias (e.g., volunteers were used, subjects self-reported results, …), 2. Describe why responses for members of the sample would differ in some systematic way from members of the population of interest (e.g., … therefore, subjects in the sample are more likely to … than those in the general population), and 3. Explain what the result will be when using the sample data to estimate a population parameter (e.g., This will result in a sample mean that will overestimate the true population mean, or this will result in a sample correlation that will tend to be larger than the population correlation).
    - **TIP:** It is important for teachers to help their students see the big picture and understand how units covered earlier in the course are interrelated to units covered later in the course. Return to concepts such as representative samples and sources of bias from the Collecting Data unit, for example, in student exercises developed for later units of the course, such as Sampling Distributions and Statistical Inference.
- Answers should always be given in the context of the problem, so when students are referring to "the study," students should use language to indicate an understanding of what the purpose of the study was.

© 2021 College Board.
Visit College Board on the web: collegeboard.org.

- When asked about the benefit of a specific statistical procedure, students need to be sure that their response is not something that is also true of procedures that lack the key feature(s) of the named procedure. For example, if asked about the benefit of using a simple random sample, responses should not discuss something that is also true of sampling methods that do not use random selection.
- When asked a ‘yes’ or ‘no’ question, responses should explicitly say ‘yes’ or ‘no’ without ambiguity.
- When students use statistical terminology (e.g., confounding variables), it is important they use the terminology correctly and, if necessary, provide an explanation, or illustration, that demonstrates a clear understanding of what that terminology means.

What resources would you recommend to teachers to better prepare their students for the content and skill(s) required on this question?

- The *AP Statistics Course and Exam Description* (CED), effective Fall 2020, includes instructional resources for AP Statistics teachers to develop students’ broader skills. Please see page 225 of the CED for examples of key questions and instructional strategies designed to develop skill 1.C, describe an appropriate method for gathering and representing data. A table of representative instructional strategies, including definitions and explanations of each, is included on pages 213-223 of the CED. The strategy “Graphic Organizer,” for example, may help students to organize ideas and information related to study design.
- AP Classroom provides two videos focused on the content and skills needed to answer this question.
    - The video for topic 3.4 discusses sampling methods that lead to bias and ways in which a sampling method might systematically lead to over/under estimates (see DAT-2.E.1), all within the context of college “success” data. Key takeaways of this video were especially relevant to this question: “Bias arises when certain responses are systematically favored over others,” and “When describing bias, explain how the sample may systematically differ from the population and the resulting direction of bias.”
    - The video for topic 3.2 develops DAT-2.B, identify appropriate generalizations and determinations based on observational studies*,* which was also relevant to this question.
- AP Classroom also provides topic questions for formative assessment of topics 3.2 and 3.4, as well as access to the question bank, which is a searchable database of past AP Questions on this topic.
- The Online Teacher Community features many resources shared by other AP Statistics teachers. For example, to locate resources to give your students practice discussing causation, try entering the keyword “causation” in the search bar, then selecting the drop-down menu for “Resource Library.” When you filter for “Classroom-Ready Materials,” you may find worksheets, data sets, practice questions, and guided notes, among other resources.

© 2021 College Board.
Visit College Board on the web: collegeboard.org.

Question #3 **Task:** Probability and Sampling Distributions

Max. Points: 4 **Mean Score:** 0.73

What were the responses to this question expected to demonstrate?

The primary goals of this question were to assess a student's ability to (1) define a random variable and identify its distribution; (2) identify the value of a binomial probability; (3) identify and interpret the expected value of a binomial random variable; and (4) use the expected value of a random variable or the probability of a specific event to counter a claim.

This question primarily assesses skills in skill category 3: Using Probability and Simulation. Skills required for responding to this question include (3.A) Determine relative frequencies, proportions, or probabilities using simulation or calculations, (3.B) Determine parameters for probability distributions, and (4.B) Interpret statistical calculations and findings to assign meaning or assess a claim.

This question covers content from Unit 4: Probability, Random Variables, and Probability Distributions of the course framework in the AP Statistics Course and Exam Description. Refer to topics 4.10, and 4.11, and learning objectives UNC-3.B, UNC-3.C, and UNC-3.D.

How well did the responses address the course content related to this question? How well did the responses integrate the skills required on this question?

- In part (a) many responses were unable to correctly identify the random variable as the number of gift cards received by a particular employee in a 52-week year, but responses generally were able to indicate that the binomial distribution should be used, identify the values of the parameters of the binomial distribution, define the event of interest, and calculate the correct probability.

- In part (b), most responses correctly calculated the expected value, but many responses had difficulty interpreting the expected value as an average over a large number of 52-week years.

- Most responses to part (c) were able to determine that Agatha did not have a strong argument. Many responses clearly based that decision on a relevant probability or expected value, linking it to the likelihood of Agatha not receiving a gift card in a 52-week year.

What common student misconceptions or gaps in knowledge were seen in the responses to this question?

Common Misconceptions/Knowledge Gaps	*Responses that Demonstrate Understanding*
• Misunderstanding or misusing vocabulary associated with random variables and probability o Random variable o Distribution o Expected value	• Let the random variable of interest X represent the number of gift cards that a particular employee receives in a 52-week year. • X has a binomial distribution. • If the random process of selecting one employee each week was repeated for a very large number of years, each employee can expect to receive about 0.26 gift cards per year, on average.

© 2021 College Board.
Visit College Board on the web: collegeboard.org.

• Many responses had difficulty defining a random variable.
  ○ It is not whether someone gets a gift card.
  ○ It is not the employees.
  ○ It is not the probability of getting a gift card.
	• Let the random variable of interest X represent the number of gift cards that a particular employee receives in a 52-week year.

• Confusion about how a variable is distributed. Many responses said normal or uniform or confused the distribution with the physical distribution of the cards (e.g., the employer handed the gift cards out randomly).	• X has a binomial distribution. ○ Because each employee has probability $\frac{1}{200}$ *or* 0.005 of being selected each week to receive a gift card and each week's selection is independent from every other week, X has a binomial distribution with $n = 52$ repeated independent trials and probability of success $p = 0.005$ for each trial.

• In part (a-ii) an error made in calculating the probability often resulted from failure to specify the correct event:
  ○ Calculating the probability of an employee receiving exactly one gift card in a 52-week year
  ○ Giving parallel solutions by computing probabilities for more than one event
  ○ Misinterpreting the complement of a discrete event. Many students said:
$P(X \geq 1) = 1 - P(X \leq 1)$
$1 - P(X \leq 0) = 1 - binomcdf(n = 52,$
$p = 0.005, x = 1)$
	• The event of interest may be correctly specified in many ways:
	  ○ $P(\text{at least one gift card})$
	  ○ $P(X \geq 1)$
	  ○ $1 - P(X = 0)$
	  ○ $1 - P(\text{none})$
	  ○ $\sum_{k=1}^{52} \binom{52}{k} (0.005)^k (0.995)^{52-k}$
	• The following calculator syntax is acceptable, but it is better to avoid such syntax. If calculator syntax is used, parameters and events must be clearly identified.
	  ○ $\text{binomcdf}(n = 52, p = 0.005,$ $\text{lower bound} = 1, \text{upper bound} = 52)$
	  ○ $1 - \text{binompdf}(n = 52, p = 0.005,$ $x = 0)$
	  ○ $1 - \text{binomcdf}(n = 52, p = 0.005,$ $x \text{ or upper bound} = 0)$

© 2021 College Board.
Visit College Board on the web: collegeboard.org.

• Common errors made in responses to part (a-ii) were calculating the probability for only one week and not for a 52-week year, i.e. , $p = \frac{1}{200} = 0.005$ or multiplying the probability for one week by 52, i.e., $(52)(0.005) = 0.26$	• $1 - P(X = 0) = 1 - (0.995)^{52} = 0.2295$
• Misunderstanding that expected values are averarges over many trials.	• If the random process of selecting one employee each week was repeated for a very large number of years, each employee can expect to receive about 0.26 gift cards per year, on average.
• Not interpreting the expected value correctly (it is not a probability) o 0.26 chance of getting a gift card o 26% chance of getting a gift card o 26% of employees will receive a gift card	• If the random process of selecting one employee each week was repeated for a very large number of years, each employee can expect to receive about 0.26 gift cards per year, on average.
• Misconception that the average or expected value must be a value the random variable could take. o 0.26 gift cards, so the average is 0 gift cards, or the average is 1 gift card.	• The average is 0.26 gift cards.
• In part (c) failing to bring the probability or expected value into the justification of the decision.	• The probability of an employee not receiving a gift card in a 52-week year is 0.77. • The probability of an employee getting at least one gift card in a 52-week year is 0.23. • The average number of gift cards that an employee would expect to receive in a 52-week year is 0.26.
• In part (c) not linking the probability or expected value to the decision. Not stating why the probabiity of 0.23 or 0.77 would justify the decision made.	• It is quite likely that a particular employee will fail to receive a gift card for an entire 52-week year because the probability of an employee getting at least one gift card in a 52-week year is only 0.23. • An employee receiving 0 gift cards is not unusual because the average number of gift cards that an employee would expect to receive in a 52-week year is 0.26.

© 2021 College Board.
Visit College Board on the web: collegeboard.org.

• In part (c) some responses wanted to link the probability to an alpha level or discuss statistical significance. Such responses confused interpreting a probability with a significance test. o The probability an employee will never receive a gift card in a 52-week year is 0.77. This is greater than 0.05 so this is not statistically significant.	• The probability an employee will never receive a gift card in a 52-week year is 0.77. This is high, so it is not unusual that Agatha did not receive a gift card. • The probability an employee will receive at least one gift card is 0.23 in a 52-week year. This is pretty low, so it is not unlikely that Agatha did not receive a gift during that time.
• Responses attempted to justify a decision based on a non-relevant probability. o The chance an employee gets a gift card is $\frac{1}{200}$ so Agatha does not have a strong argument.	• The probability an employee will never receive a gift card in a 52-week year is 0.77. This is high, so it is not unusual that Agatha did not receive a gift card.

• Poor communucation of reasoning, e.g., not stating if a probability should be considered large or small.
  o The probability an employee will never receive a gift card in a 52-week year is 0.77.
  o The probability an employee will receive at least one gift card is 0.23 in a 52-week year.
	• The probability an employee will never receive a gift card in a 52-week year is 0.77. This is high, so it is not unusual that Agatha did not receive a gift card.
	• The probability an employee will receive at least one gift card is 0.23 in a 52-week year. This is pretty low, so it is not unlikely that Agatha did not receive a gift during that time.

Based on your experience at the AP® Reading with student responses, what advice would you offer teachers to help them improve the student performance on the exam?

Stress correct use of vocabulary! Play games, make flashcards, give vocabulary quizzes, or create a word wall. Students need to practice using vocabulary words, especially statistical terminology, correctly.

When introducing a new distribution, spend time connecting the distribution to a specific type of random variable.

When solving probability problems involving a particular distribution, require students to identify the random variable, state the distribution and specify values for the parameters. Teachers should focus on probability notation rather than calculator syntax. When using calculator syntax, everything must be clearly labeled.

Ask students to interpret the values they get. The more they interpret the values the more the students will understand about what they are finding and why.

Teachers should frequently ask. “why?” Why is Agatha wrong? Why is that expected value a decimal?

© 2021 College Board.
Visit College Board on the web: collegeboard.org.

Teach students to "close the loop" when providing a rationale. Finish making the argument connecting the probability to the valid argument; provide a statement about the rarity/likeliness of the event taking place. Why does the probability you computed support, or provide evidence against, the claim? Every decision should have an explanation or a justification.

Tell students not to assume the person reading their response knows what they are thinking. If the student provides a number as part of their justification of a decision, they need to say how that number helps support their decision. Teach them to explicitly say what they mean and finish their thoughts. Don't use "it;" be clear what "it" is referring to.

Teachers should give students practice with making predications and decisions based on probability alone. Do some problems of this sort after inference, so students learn that not everything needs to be a hypothesis test. When teaching probability, add parts to questions that require students use probability to support an argument or make a prediction. For example, ask students to determine if an event is likely or not.

Have students practice answering a question using words in the stem of the problem. For example, "The probability that an employee receives at least one gift card in a 52-week year is 0.2295" or "Agatha does not have a strong argument that the selection process was not truly random, because …"

What resources would you recommend to teachers to better prepare their students for the content and skill(s) required on this question?

- The *AP Statistics Course and Exam Description* (CED), effective Fall 2020, includes instructional resources for AP Statistics teachers to develop students' broader skills. Please see pages 229-230 of the CED for examples of key questions and instructional strategies designed to develop skills 3.A and 3.B and page 232 for questions and instructional strategies designed to develop skill 4.B, interpret statistical calculations and findings to assign meaning or assess a claim. A table of representative instructional strategies, including definitions and explanations of each, is included on pages 213-223 of the CED. The strategy "Sentence Starters," for example, may help students to practice communication skills: "The probability an employee will *never receive a gift card* in *a 52-week year* is *0.70*. This is *high*, so it is *not unusual* that Agatha did not receive a gift card."
- AP Classroom videos for Topic 4.10 and 4.11 are especially helpful for developing the content and skills needed to answer this question.
    - The video for topic 4.10 discusses defining a random value, identifying the distribution and values of interest, determining probabilities using the binomial probability formula, and answering a question in context.
    - The video for topic 4.11 develops skill 3.B, determine parameters for probability distributions, applied to the binomial distribution, which was especially relevant to this question.
- AP Classroom also provides topic questions for formative assessment of topics 4.10 and 4.11, as well as access to the question bank, which is a searchable database of past AP Questions on these topics.
- The Online Teacher Community features many resources shared by other AP Statistics teachers. For example, to locate resources to give your students practice using a binomial distribution, try entering the keywords "binomial distribution" in the search bar, then selecting the drop-down menu for "Resource Library." When you filter for "Classroom-Ready Materials," you may find worksheets, data sets, practice questions, and guided notes, among other resources.

© 2021 College Board.
Visit College Board on the web: collegeboard.org.

Question #4 **Task:** Inference

Max. Points: 4 **Mean Score:** 1.54

What were the responses to this question expected to demonstrate?

The primary goals of this question were to assess a student's ability to (1) identify an appropriate inference procedure to test a claim about a population proportion; (2) identify the appropriate null hypothesis and the appropriate alternative hypothesis; (3) check conditions required for accurate application of the identified inference procedure; (4) compute the value of a test statistic and the corresponding *p*-value; (5) state and justify a conclusion about the claim; and (6) determine whether a Type I or Type II error could have been made and describe a consequence of the identified type of error.

This question primarily assesses skills associated with inference, including skills in skill category 1: Selecting Statistical Methods; skill category 3: Using Probability and Simulation; and skill category 4: Statistical Argumentation. Skills required for responding to this question include (1.B) Identify key and relevant information to answer a question or solve a problem, (1.E) Identify an appropriate inference method for significance tests, (1.F) Identify null and alternative hypotheses, (3.E) Calculate a test statistic and find a *p*-value, provided conditions for inference are met, (4.A) Make an appropriate claim or draw an appropriate conclusion, (4.C) Verify that inference procedures apply in a given situation, and (4.E) Justify a claim using a decision based on significance tests.

This question covers content from Unit 6: Inference for Categorical Data: Proportions of the course framework in the AP Statistics Course and Exam Description. Refer to topics 6.4, 6.5, 6.6, and 6.7, and learning objectives DAT-3.B, UNC-5.A, VAR-6.D, VAR-6.E, VAR-6.F, and VAR-6.G.

How well did the responses address the course content related to this question? How well did the responses integrate the skills required on this question?

This question adapted the standard three-section rubric for inference questions and added another section to assess the response to the type of error question in part (b). The first section includes the statement of the null and alternative hypotheses, in the context of the study, and specification of the test statistic using words or formula from part (a). The second section includes verifying conditions for applying the test and computation of the values of the test statistic and *p*-value from part (a). The third section includes the statement of the conclusion, in the context of the study, with justification based on the results reported in part (a). The fourth section includes reporting the appropriate type of error and stating a consequence that follows from that type of error from part (b).

Section 1:

- A substantial minority of responses failed to implement an inference procedure and instead relied upon a plea just using the sample statistic.
- Most responses recognized that a one-proportion *z*-test was appropriate in this context. Several responses achieved this by reporting the formula for the *z*-test.
- Most responses recognized that the alternative hypothesis was right-tailed. However, some responses failed to properly convey the concept of population proportion by using nonstandard notation in the statement of hypotheses.
- A substantial minority of responses did not include sufficient context by excluding mention of the response variable.

Section 2:

- Most responses recognized that conditions must be checked before conducting a hypothesis test; however, a substantial minority of responses failed to properly check those conditions.
- The check of the independence condition was frequently incomplete. Most responses cited the condition of random sampling, but some failed to check the 10% condition. Further, some responses simply stated "it was random" without indicating that data were obtained from a random sample.
- Most responses reported correct values of the test statistic and *p*-value as found from their calculator, with few attempting to directly calculate the value of test statistic from its formula. For those that showed a test statistic

© 2021 College Board.
Visit College Board on the web: collegeboard.org.

formula, a large number incorrectly used the sample statistic $(\hat{p})$ in the standard error formula instead of the hypothesized value (p_0).
- For the few responses that used a critical value approach, some failed to properly identify the correct critical value from the table of z-scores.
- Some responses reported a confidence interval and a hypothesis test. Great care had to be taken on behalf of the reader to score each approach and report the weaker of the two scores.

Section 3:
- Most responses made a correct decision with justification based on the relationship between p-value and α but many failed to state their conclusion in terms of the alternative hypothesis. Very few responses stated a conclusion that was in opposition to their decision.
- Many responses were considered minimal when the conclusion did not provide context.
- Some responses included an interpretation of a p-value; however, most p-value interpretations were incorrect.

Section 4:
- Most responses correctly identified the type of error associated with their decision from part (a). However, some responses incorrectly described the identified error.
- Many responses failed to provide a consequence of the identified type of error and, instead, simply defined the type of error in context.

What common student misconceptions or gaps in knowledge were seen in the responses to this question?

Most errors occurred in responses that either poorly organized their work or poorly communicated their ideas. More specific mistakes are noted in the table below.

Common Misconceptions/Knowledge Gaps	*Responses that Demonstrate Understanding*
• Does not refer to a population proportion in stating the hypotheses and/or uses nonstandard notation. (e.g., "$H_0 : \hat{p} = 0.4,\ H_a : \hat{p} > 0.4$")	• Let p be the population proportion of customers that would place an order if offered a $10 coupon. $H_0 : p = 0.4,\ H_a : p > 0.4$
• Correctly states the name of the appropriate test but uses an inappropriate formula. (e.g., $z = \frac{0.4 - \frac{38}{90}}{\sqrt{\frac{\left(\frac{38}{90}\right)\left(\frac{42}{90}\right)}{90}}} = -0.475$)	• The test conducted is a one-proportion z-test and the value of the test statistic is $z = \frac{\frac{38}{90} - 0.4}{\sqrt{\frac{(0.4)(0.6)}{90}}} = 0.430$.
• Fails to include sufficient context in identifying the population parameter or stating the hypotheses.	• The parameter, p, is the population proportion of all customers of the pet supply company who would place an order within 30 days after receiving an email with a coupon for $10 off the next purchase.
• Fails to acknowledge the independence condition is checked by BOTH the random selection AND the 10% rule.	• The independent observations condition for performing a one-sample z-test is satisfied because 1) this is a random sample of 90 customers, and 2) it is reasonable to assume the company has at least 900 customers.

© 2021 College Board.
Visit College Board on the web: collegeboard.org.

• Fails to appropriately check the values for the expected number of successes/failures is sufficiently large (at least 5 or 10). For example, $n\hat{p} > 30$ and $n(1-\hat{p}) > 30$.	• The sample size is large enough to support a condition of normality of the sampling distribution because $90(0.4) = 36 > 10$ and $90(0.6) = 54 > 10$.
• Conclusion is stated in terms of the null hypothesis instead of the alternative hypothesis. (e.g., "We have enough evidence to suggest that 40% of customers would place an order if offered a \$10 coupon.")	• Because $p\text{-value} = 0.33 > \alpha = 0.05$, we do not have convincing statistical evidence to suggest that more than 40% of customers would place an order if offered a \$10 coupon.
• Conclusion is stated in terms that suggest the alternative hypothesis has been "proven" untrue. (e.g., "There is no evidence that the manager's belief is correct.")	• Because $p\text{-value} = 0.33 > \alpha = 0.05$, we do not have sufficient evidence to suggest that the manager's belief is correct.
• States a type II error in a context that implies the hypothesis test was done incorrectly. (e.g., "We did not find evidence that more than 40% of customers would place an order if offered a \$10 coupon but there actually was evidence that more than 40% would place an offer if offered a \$10 coupon.")	• Although an interpretation of the error in context was not necessary, a correct interpretation would be, "We did not find convincing evidence that more than 40% of customers would place an order if offered a \$10 coupon when, actually, more than 40% of customers would place an order if offered a \$10 coupon."
• Provides a definition of a type II error in context without a consequence of that error. (e.g., "Type II error. The manager would conclude that there is no evidence for their claim when their claim is actually true.")	• A consequence of incorrectly concluding that it is *not* true that more than 40% of customers would place an order if offered a \$10 coupon would be that the manager ends the coupon promotion and sales decline.

Based on your experience at the AP® Reading with student responses, what advice would you offer teachers to help them improve the student performance on the exam?

- Provide opportunities for students to practice writing skills from the beginning of the course.
    - Assign previously released AP problems as assessments.
    - Teach students organizational strategies (e.g., state/plan/do/conclude).
- Encourage students to define the population parameter of interest in context.
- Emphasize the importance of checking conditions before conducting inference procedures.
- Assess student's ability to quickly decide on appropriate inference procedures.
- To help minimize errors in the use of notation, encourage students to practice writing hypotheses in context using complete sentences.
- Encourage students to name inference procedures in words instead of by formula.
- Have students write the proper check for a large enough sample size for a one-proportion inference procedure as $np_0 > 10$ and $n(1-p_0) > 10$, with the specific values for n and p_0 from the prompt.
- Emphasize that the independence condition is checked by 1) random sampling AND 2) the sample size is less than 10% of the population.
- Encourage students to use a hypothesis test instead of a confidence interval to provide justification about a statistical claim.
- Remind students to provide justification for a hypothesis test conclusion using the *p*-value.

© 2021 College Board.
Visit College Board on the web: collegeboard.org.

- Have students practice writing conclusions for hypothesis tests in terms of having enough statistical evidence to support (or not having enough statistical evidence to support) the alternative hypothesis in context.
- Teach students to organize error concepts in a HOT box (H_0 is *T*rue) to help with understanding. E.g.,

		Truth: H_0 is True	Truth: H_0 is False
Decision	Reject H_0	Type I Error	Power
	Fail to Reject H_0		Type II Error

- Remind students that when they are asked to make a choice, they should pick just one choice and explain their reasoning.
- Emphasize that consequences have tangible impacts and do not involve "thinking" or "feeling" or "concluding." Provide real-world examples and allow students to practice consequences that follow from different decisions.

What resources would you recommend to teachers to better prepare their students for the content and skill(s) required on this question?

- The AP Statistics *Course and Exam Description* (CED), effective Fall 2020, includes instructional resources for AP Statistics teachers to develop students' broader skills.
    - Section 1: Please see page 226 for examples of key questions and instructional strategies designed to develop skills 1.E, identify an appropriate method for significance tests, and 1.F, identify null and alternative hypotheses.
    - Section 2: Please see page 232 for examples of key questions and instructional strategies designed to develop skill 4.C, verify that inference procedures apply in a given situation, and page 230 for skill 3.E, calculate a test statistic and find a *p*-value, provided conditions for inference are met.
    - Section 3: Please see pages 231, 232 for examples of key questions and instructional strategies designed to develop skills 4.A, make an appropriate claim or draw an appropriate conclusion and 4.E, justify a claim using a decision based on a significance test.
    - Section 4 pulls together several of the skills developed above: 1.B, 3.A, 4.A, and 4.B.
- A table of representative instructional strategies, including definitions and explanations of each, is included on pages 213-223 of the CED. The strategy "Error analysis," for example, may help students to recognize how to avoid errors, such as implicitly accepting the null hypotheses.
- AP Classroom videos for topics 6.4 through 6.7 are a rich resource for helping students to develop understanding of the content and mastery of the skills featured in these topics and this question. Each topic features two videos, each focused on a different skill and presented in context.
- AP Classroom also provides topic questions for formative assessment of topics 6.4 through 6.7, as well as access to the question bank, which is a searchable database of past AP Questions on these topics.
- The Online Teacher Community features many resources shared by other AP Statistics teachers. For example, to locate resources to give your students practice identifying and giving a consequence of a Type I or Type II error, try entering the keyword "Error" in the search bar, then selecting the drop-down menu for "Resource Library." When you filter for "Classroom-Ready Materials," you may find worksheets, data sets, practice questions, and guided notes, among other resources.

Question #5 **Task:** Multi-Focus

Max. Points: 4 **Mean Score:** 1.45

What were the responses to this question expected to demonstrate?

The primary goals of this question were to assess a student's ability to (1) recognize whether comparisons between samples should be based on proportions instead of counts when sample sizes are different; (2) identify appropriate proportions to compute from a table of counts; (3) construct and label a segmented bar chart; (4) use a segmented bar chart to make a comparison; (5) identify an appropriate inference procedure for investigating whether the distribution of a categorical random variable differs across populations; and (5) identify the null and alternative hypotheses for a chi-square test of homogeneity.

This question assesses skills in multiple skill categories, including skill category 1: Selecting Statistical Methods; skill category 2: Data Analysis; and skill category 4: Statistical Argumentation. Skills required for responding to this question include (1.E) Identify an appropriate inference method for significance tests, (1.F) Identify null and alternative hypotheses, (2.B) Construct numerical or graphical representations of distributions, (2.D) Compare distributions or relative positions of points within a distribution, and (4.B) Interpret statistical calculations and findings to assign meaning or assess a claim.

This question covers content from multiple units, including Unit 1: Exploring One-Variable Data, Unit 2: Exploring Two-Variable Data, and Unit 8: Inference for Categorical Data: Chi-Square of the course framework in the AP Statistics Course and Exam Description. Refer to topics 1.4, 2.2, 2.3, and 8.5, and learning objectives UNC-1.C, UNC-1.P, UNC-1.R, VAR-8.I, and VAR-8.J.

How well did the responses address the course content related to this question? How well did the responses integrate the skills required on this question?

- In part (a) most responses recognized the need to compare sample proportions instead of counts when comparing the results for the different cities. If a response recognized that a teen was selected from each city, it most often correctly computed the sample proportion for each city and explicitly compared the three values. Many incorrect responses computed a proportion for the combined Detroit and San Diego samples, computed proportions based on the overall total of teens, or computed proportions based on the total number of "Yes" responses. Responses that did not include a specific answer to the question about the correctness of the claim or did not provide a directional comparison (e.g., higher, lowest) of computed values were scored no higher than partial (P).

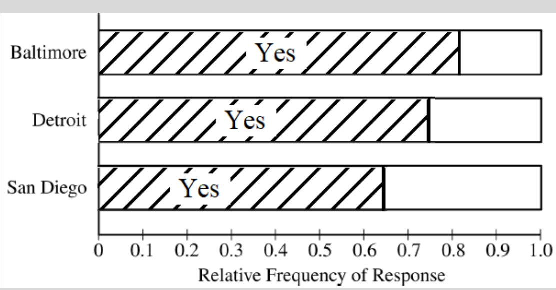
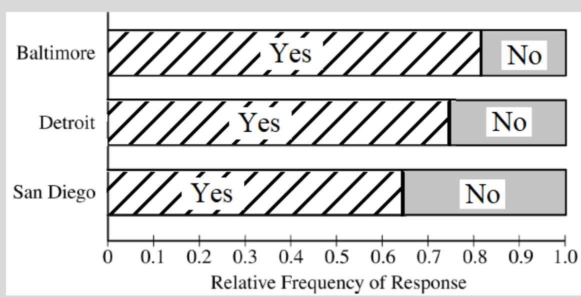
- In part (b-i) most responses correctly segmented the bar chart; the majority of these responses also labeled the segments or provided a key. Unfortunately, some responses attempted to overlay the two proportions or construct side-by-side bar graphs for each city which does not demonstrate an understanding of a segmented bar graph.

- In part (b-ii) the majority of responses correctly identified San Diego as having the smallest proportion and provided the correct value.

- In part (c) most responses identified a chi-square test but did not correctly specify a chi-square test <u>of homogeneity</u>. In stating hypotheses, many responses did not include the context of the proportions of teens in the three cities who consumed a soft drink. Some responses provided an incorrect alternative hypothesis using wording that indicated that the proportions for the three cities must be three different values rather than only at least two values must be different.

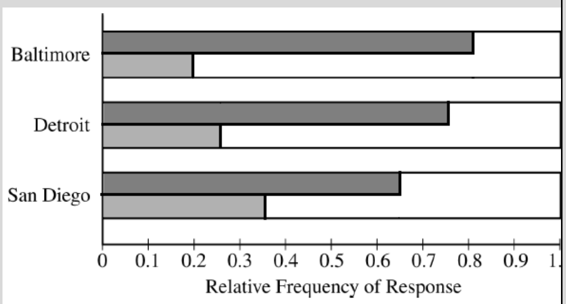
© 2021 College Board.
Visit College Board on the web: collegeboard.org.

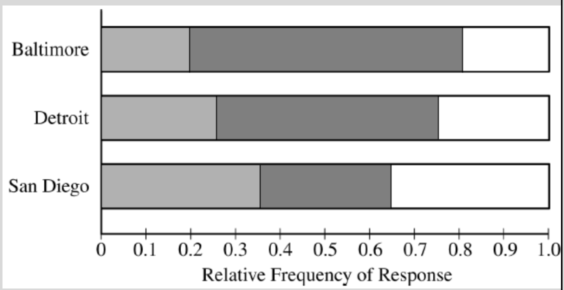
What common student misconceptions or gaps in knowledge were seen in the responses to this question?

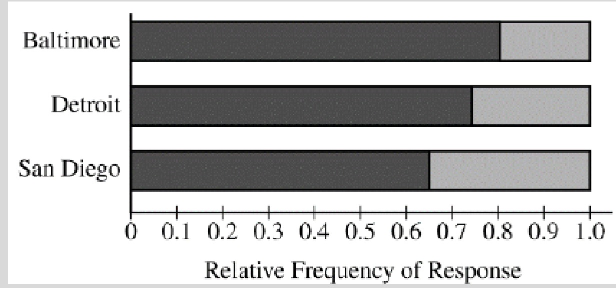
Common Misconceptions/Knowledge Gaps	*Responses that Demonstrate Understanding*
• Using the term "population size" when referring to the "sample size."	• Because the sample sizes are different for the three cities, the researcher should not use counts to compare the likelihood of a selecting a teen who consumed a soft drink from each city.
• Recognizing that there is a problem with the researcher's claim due to different sample sizes is not a complete answer. Students need to include their reasoning and, if appropriate, provide a correct approach. Answers which only stated that the samples sizes were different without addressing how the sample size contributes to "likelihood" did not receive full credit.	• The researcher's claim is incorrect. Although Baltimore had the fewest "Yes" responses, it also had the smallest sample size. Because the sample sizes for the three cities were different, the researcher should compare the proportion of "Yes" responses and not the counts.
• Computing incorrect proportions, e.g., $\frac{727}{3,441} \approx 0.211$, $\frac{1232}{3,441} \approx 0.358$, $\frac{1482}{3,441} \approx 0.431$. Proportions should be computed for each individual city because a single teen was randomly selected from each city.	• For the sample from Baltimore, the proportion of teens who consumed a soft drink in the past week is $\frac{727}{904} \approx 0.804$; for the sample from Detroit, the proportion is $\frac{1,232}{1,663} \approx 0.741$; and for the sample from San Diego, the proportion is $\frac{1,482}{2,280} = 0.65$.
• Labels are required on segmented bar graphs, and all segments must be labeled to indicate an understanding that the bar represents the whole sample and total 100%. An example of insufficient labels: 	• All segments are clearly labeled for each bar: 

© 2021 College Board.
Visit College Board on the web: collegeboard.org.

- Some responses provided side-by-side or overlapping bars. For example:





- 

- Using statistics in the hypotheses rather than parameters. For example: $H_0 : X^2 = 0$ versus $H_a : X^2 \neq 0$ or $H_0 : \hat{p}_B = \hat{p}_D = \hat{p}_{SD}$ versus H_a: at least one $\hat{p}$ is different

- $H_0 : p_B = p_D = p_{SD}$ versus H_a : at least one p_i is different, where p_i is the proportion of all teens from city *i* who consumed a soft drink in the past week

- Some reponses stated the alternative hypothesis incorrectly by indicating that all three proportions must be different. Examples of incorrect wording are "proportions are different for each city," "the proportion is different in Baltimore, Detroit, and San Diego," "the proportion of all teens who consumed a soft drink in the past week is different for all three cities."

- H_0 : There is no difference in the proportions of all teens who consumed a soft drink in the past week across the three cities.

  H_a : The proportions of all teens who consumed a soft drink in the past week are different for at least two of the three cities.

- Hypotheses that test for an association between variables rather than a comparison of population distributions are incorrect. For example: H_0 : There is no association between soda consumption and city.

- H_0 : The distribution of soda consumption by all teens is the same for the three cities.

  H_a : At least one of the distributions of soda consumption by all teens is different for the three cities.

© 2021 College Board.
Visit College Board on the web: collegeboard.org.

Based on your experience at the AP® Reading with student responses, what advice would you offer teachers to help them improve the student performance on the exam?

Teaching tips:

- Develop exercises to help students use vocabulary correctly. Population refers to the entire group from which the sample was chosen; sample refers to the units selected that provide the data for analysis. Population size and sample size should be discussed whenever sampling occurs in a problem.
- Students need to read the question carefully and understand the information given before answering the question. In particular, conditional probability can be difficult due to the same words being used in multiple ways. Understanding the difference between $P(A \cap B)$ and $P(A \mid B)$ when written in words is a difficult concept for students. Students should be presented with multiple versions of this type of wording.
- Emphasize that a segmented bar represents a whole group; multiple segmented bars allow for the comparison of multiple groups. A single segmented bar totals 100%, and each segment represents the relative frequency of a particular response within a group. Collectively, the relative frequencies are the distribution for the group among the categories (commonly referred to as simply the "distribution").
- Chi-square tests need to be identified completely by specific name: chi-square goodness-of-fit test, chi-square test of independence, or chi-square test of homogeneity.
- Emphasize that a chi-square test of homogeneity is a test to investigate if the groups (genus) are the same (homo) with respect to their distributions (collection of relative frequencies) among categories. The test considers the distribution within each group and tests if those distributions are the same for all groups. This concept can be related to the segmented bar graphs. If the bars for the different groups have similar patterns, then the groups have similar distributions, and the null hypothesis of same distributions is not likely to be rejected. Students should be encouraged to make the connection between the visual display of segmented bar graphs and the written hypotheses of a chi-square test of homogeneity, especially when there are more than two categories for each group.
- When presenting count data collected from several different populations in a table, consider leaving off the "Total" column containing the counts for the combined samples. The "Total" column can be misleading to students when interpreting the data. Encourage students to think about the individual samples as being distinct. Note that the "Total" column is relevant under the assumption (usually of the null hypothesis) that the distribution among categories is the same for all the populations and is used when calculating the test statistic (similar to the pooled proportion of a two-sample z-test for proportions).
- It is recommended that chi-square test of independence and chi-square test of homogeneity be taught separately. Although the computations are the same when performing these tests, the concepts are very different. Students need to clearly understand how the data was collected and the question of interest; they should practice writing hypotheses for different situations before doing test calculations.

What resources would you recommend to teachers to better prepare their students for the content and skill(s) required on this question?

- The *AP Statistics Course and Exam Description* (CED), effective Fall 2020, includes instructional resources for AP Statistics teachers to develop students' broader skills. Please see page 226 of the CED for examples of key questions and instructional strategies designed to develop skills 1.E and 1.F, pages 227, 228 for skills 2.B and 2.D, and page 232 for questions and instructional strategies designed to develop skill 4.B. A table of representative instructional strategies, including definitions and explanations of each, is included on pages 213-223 of the CED. The strategy "Sketch and Switch," for example, may be modified to help students to practice constructing well-labeled segmented bar graphs, as required for this question.
- AP Classroom videos for topics 1.4, 2.2, 2.3, and 8.5 are especially helpful for developing the content and skills needed to answer this question.
  - The videos for topic 1.4 introduce constructing displays for categorical data. The video for topic 2.2 develops how to construct segmented bar graphs.
  - The video for topic 2.3 develops content and skills related to conditional relative frequencies.
  - The videos for topic 8.5, especially the first of the two videos, develop understanding of when to use a chi-square test for homogeneity vs. a chi-square test for independence.
- AP Classroom also provides topic questions for formative assessment of topics 1.4, 2.2, 2.3, and 8.5, as well as access to the question bank, which is a searchable database of past AP Questions on these topics.

© 2021 College Board.
Visit College Board on the web: collegeboard.org.

- The Online Teacher Community features many resources shared by other AP Statistics teachers. For example, to locate resources to give your students practice naming a specific χ^2 test, try entering the keyword "homogeneity" in the search bar, then selecting the drop-down menu for "Resource Library." When you filter for "Classroom-Ready Materials," you may find worksheets, data sets, practice questions, and guided notes, among other resources.

© 2021 College Board.
Visit College Board on the web: collegeboard.org.

Question #6 **Task:** Investigative Task

Max. Points: 4 **Mean Score:** 1.56

What were the responses to this question expected to demonstrate?

The primary goals of this question were to assess a student's ability to (1) compare two distributions using information provided by side-by-side boxplots; (2) use information in a scatterplot to compare trends across time for two sets of data; (3) use information in a scatterplot to describe the relationship between two variables; (4) compare the rates of change for one variable as another variable changes for two data sets displayed in a scatterplot; (5) use information presented in several graphs to explain how a variable could be a confounding variable with respect to the relationship between two other variables.

This question assesses skills in multiple categories, including skill category 1: Selecting Statistical Methods; skill category 2: Data Analysis; and skill category 4: Statistical Argumentation. Skills required for responding to this question include (1.C) Describe an appropriate method for gathering and representing data, (2.A) Describe data presented numerically or graphically, (2.D) Compare distributions or relative positions of points within a distribution, (4.A) Make an appropriate claim or draw an appropriate conclusion, and (4.B) Interpret statistical calculations and findings to assign meaning or assess a claim.

This question covers content from multiple units, including Unit 1: Exploring One-Variable Data, Unit 2: Exploring Two-Variable Data, and Unit 3: Collecting Data of the course framework in the AP Statistics Course and Exam Description. Refer to topics 1.9, 2.4, 2.8, and 3.5, and learning objectives DAT-1.A, DAT-1.H, UNC-1.N, and VAR-3.A.

How well did the responses address the course content related to this question? How well did the responses integrate the skills required on this question?

- In part (a) most responses correctly addressed at least one of the characteristics of a distribution (shape, center, variability, outliers), and many did so using explicit comparison phrases (e.g., "is greater than"). However, some responses didn't address all of the characteristics or didn't use comparative language. Also, some responses did not include sufficient context because the response variable was not mentioned.

- In part (b) most responses correctly described the trend as positive (increasing) for the new stadium. Fewer responses, however, correctly described the trend in the old stadium as neither increasing or decreasing (or increasing at a slower rate than the new stadium). Also, some responses did not include sufficient context by identifying the stadiums and including the names of the explanatory and response variables.

- In part (c-i) most responses correctly described at least one of the characteristics of an association between two quantitative variables (direction, form, strength). In part (c-ii), most responses correctly stated that the rates were about the same in the two stadiums, but many responses simply repeated words from the statement of the question without providing an explanation.

- In part (d) some responses provided a correct explanation of confounding in the context of this problem, but most responses were unable to explain the concept of confounding or use information provided in the graphs to explain how one variable could be a confounding variable in the relationship between average attendance and another variable, or both.

What common student misconceptions or gaps in knowledge were seen in the responses to this question?

Common Misconceptions/Knowledge Gaps	*Responses that Demonstrate Understanding*
• Some responses did not address all of the relevant characteristics of the boxplots (center, spread, and existence of unusual values). (Note: Because complete shape information cannot be obtained from a boxplot, comments about shape were not required.)	• The median average attendance is greater at the new stadium than the old. Likewise, the range of average attendance across years is greater at the new stadium than the old. Neither distribution has outliers.
• Many responses did not use comparative language when attempting to compare the distributions. For example, "The median attendance in the new stadium is about 25,000 while the median attendance at the old stadium is about 16,000." The word "while" (or equivalent) does not count as a comparison.	• The median average attendance is greater at the new stadium than the old. Likewise, the range of average attendance across years is greater at the new stadium than the old. Neither distribution has outliers.
• Some responses did not include sufficient context in part (a) because they did not refer to the response variable (average attendance).	• The median average attendance is greater at the new stadium than the old. Likewise, the range of average attendance across years is greater at the new stadium than the old. Neither distribution has outliers.
• Many responses gave descriptions of shape that were not supported by the boxplots (e.g., "the old distribution is approximately normal" or "the new distribution is skewed to the right").	• The distribution of average attendance for the old stadium is roughly symmetric, and the distribution of average attendance for the new stadium is skewed to the left.
• Some responses referred to the mean rather than the median average attendance.	• The median average attendance is greater at the new stadium than the old.
• Some responses referred to the IQR (or range) as a region, not a measure of spread. For example, "the whole IQR of the new stadium is above the whole IQR of the old stadium."	• The middle 50% of the values for the new stadium are greater than the middle 50% of the values for the old stadium.
• Some responses described the trends in both stadiums as positive, without distinguishing that the rate of increase is greater in the new stadium.	• The trend in average attendance over time at both stadiums is positive, but average attendance is increasing faster at the new stadium.
• Some responses did not include sufficient context in part (b) because they did not refer to one or more of the following: stadiums, explanatory variable (time or years), response	• The trend in average attendance over time in the old stadium is neither positive nor negative, while the trend in average attendance over time in the new stadium is positive.

© 2021 College Board.
Visit College Board on the web: collegeboard.org.

variable (average attendance). Forgetting the explanatory variable was most common.	
• Some responses did not address all three characteristics of an association (direction, form, strength). Form and strength were missing much more often than direction.	• There is a strong, positive, linear association between number of games won and average attendance.
• Some responses suggested that changes in number of games won causes attendance to increase. While this may be true, association between two variables does not necessarily imply causation.	• There is a strong, positive, linear association between number of games won and average attendance.
• Some responses did not provide an explanation in part (c-ii) that went beyond repeating words in the statement of the question.	• A line drawn through the points for the old stadium has about the same slope as a line drawn through the points for the new stadium.
• Many responses had a vague idea of confounding but were not able to convey the meaning of confounding in this context.	• Number of games won and stadium are confounded because we cannot know which of these variables is causing the increase in attendance.
• Many responses were unable to justify the existance of a counfounding variable by showing how it is related to both the explanatory variable and the response variable (or by showing how two explanatory variables are related to each other and to the response variable).	• In part (a) the boxplots show that attendance was higher in the new stadium. In part (c) Graph I shows that attendance was higher when the team had more wins. Finally, Graph II in part (c) shows that the team had more wins in the new stadium.
• Some responses tried to explain the idea of confounding, but made statements that were too strong. For example, "It was the number of games won that caused the increase in attendance, not the new stadium."	• It could be that the number of games won caused the increase in attendance, not the new stadium.

Based on your experience at the AP® Reading with student responses, what advice would you offer teachers to help them improve the student performance on the exam?

- Insist that students use comparative phrases (e.g., "is greater than," "is approximately the same as") when comparing distributions, trends, or associations.
- Make sure students always include context in any description, comparison, and explanation by using the variable names (e.g., "average attendance," "time," "number of games won"), not just the group names (e.g., "old stadium," "new stadium"). Using the labels on the graph axes is usually sufficient for context.
- Develop exercises to help students practice the correct use of statistical vocabulary, including terms like IQR, range, and correlation.
- Develop exercises to help students remember to address all four primary characteristics when describing/comparing distributions of a quantitative variable: shape, center, variability, and outliers/gaps. Some teachers use acronyms like SOCS to help students remember these characteristics.

© 2021 College Board.
Visit College Board on the web: collegeboard.org.

- Make sure students are aware of the limitations of boxplots. Boxplots don't reveal peaks (e.g., normality) or gaps.
- Make sure students know that the line segment in the box of a boxplot represents the median, not the mean.
- Help students understand the difference between the strength of an association and the direction of an association. Describing the association in the old stadium as "weaker" does not necessarily imply that the growth in attendance over time is smaller.
- Develop exercises to help students remember to address all four primary characteristics when describing the association between two quantitative variables: direction, form, strength, and unusual features.
- Insist that students provide complete explanations and stress that simply repeating some words in the statement of the question is not sufficient.
- Make sure that students understand the difference between explanatory and response variables and that associations do not imply causation.
- Make sure students understand that there must be three associations for confounding to exist: a relationship between one explanatory variable and the response variable, a relationship between a second explanatory variable (the potential confounding variable) and the response variable, and a relationship between the two explanatory variables.
- Make sure students understand the consequence of confounding: If two explanatory variables are confounded, it is impossible to determine which explanatory variable may be causing a change in the response variable. The idea of confounding is very important for students to understand, especially in the context of observational studies.

What resources would you recommend to teachers to better prepare their students for the content and skill(s) required on this question?

- The *AP Statistics Course and Exam Description* (CED), effective Fall 2020, includes instructional resources for AP Statistics teachers to develop students' broader skills. Please see page 225 of the CED for examples of key questions and instructional strategies designed to develop skills 1.C, pages 227, 228 for skills 2.A and 2.D, and pages 231, 232 for questions and instructional strategies designed to develop skills 4.A and 4.B. A table of representative instructional strategies, including definitions and explanations of each, is included on pages 213-223 of the CED. The strategy "Two wrongs make a right," for example, may help students to improve comparison skills, as required for this question.
- AP Classroom videos for topics 1.9, 2.4, 2.8, and 3.5 are especially helpful for developing the content and skills needed to answer this question.
- AP Classroom also provides topic questions for formative assessment of topics 1.9, 2.4, 2.8, and 3.5, as well as access to the question bank, which is a searchable database of past AP Questions on these topics.
- The Online Teacher Community features many resources shared by other AP Statistics teachers. For example, to locate resources to give your students practice describing confounding variables, try entering the keyword "confounding" in the search bar, then selecting the drop-down menu for "Resource Library." When you filter for "Classroom-Ready Materials," you may find worksheets, data sets, practice questions, and guided notes, among other resources.

© 2021 College Board.
Visit College Board on the web: collegeboard.org.