Chief Reader Report on Student Responses:

2019 AP® Statistics Free-Response Questions

- Number of Students Scored 219,392
- Number of Readers 849
- Score Distribution

Exam Score	N	%At
5	32,332	14.7
4	40,379	18.4
3	58,321	26.6
2	42,366	19.3
1	45,994	21.0

- Global Mean 2.87

The following comments on the 2019 free-response questions for AP® Statistics were written by the Chief Reader, Kenneth Koehler from Iowa State University. 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.

Question #1 **Max. Points:** 4 **Mean Score:** 1.68

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 features of a distribution of sample data using information provided by a histogram; (2) identify potential outliers; (3) sketch a boxplot; and (4) comment on an advantage of displaying data as a histogram rather than as a boxplot.

This question primarily assesses skills in skill category 2: Data Analysis. Skills required for responding to this question include (2.A) Describe data presented numerically or graphically, (2.B) Construct numerical or graphical representations of distributions, (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 topics 1.6, 1.7, and 1.8, and learning objectives UNC-1.H, UNC-1.K, UNC-1.L, and UNC-1.M.

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 addressed the shape of the distribution, with many correctly identifying the bimodal feature by talking about two peaks, or two clusters, of data. Center was addressed by most responses, with many describing the center of the entire distribution and others describing the centers of the two clusters (locations of the two peaks). Students had more difficulty describing variation, and almost all responses that addressed variation did so in reference to the entire distribution; very few responses addressed variation separately within each cluster. Most responses that addressed variation did so either by recognizing bounds for (or an approximation of) the value of the range, or by recognizing bounds which contained all data values.

- In part (b) most responses that tried to justify a claim of no outliers tried to use the $1.5 \times \text{IQR}$ beyond the first or third quartile criterion; very few responses tried to use the 2 or 3 standard deviations from the mean criterion. Unfortunately, some students correctly computed an appropriate criterion but failed to indicate why the computation led to a conclusion of no outliers. Most students were able to draw an appropriate boxplot.

- Many students correctly indicated that the bimodal feature of the distribution was evident in the histogram but not in the boxplot. Many other students stated that the boxplot revealed a greater degree of skewness than the histogram; unfortunately, these responses do not address a feature that is evident in the histogram but not the boxplot.

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

Common Misconceptions/Knowledge Gaps	*Responses that Demonstrate Understanding*
• Describing the shape of the distribution as approximately normal. This implies a bell-shaped or unimodal mound shape for the distribution that is not apparent in the histogram	• The distribution of the sample of room sizes is bimodal and roughly symmetric with most room sizes falling into two clusters: 100 to 200 square feet and 250 to 350 square feet.
• Failure to address variability when describing a distribution from information in a histogram.	• The range of the distribution is somewhere between 150 and 250 square feet.

• Incorrect use of the statistical term "range," e.g., "the range of the data is from 100 to 350 square feet." In statistics the range is a single number representing the distance between the maximum and minimum values.	• The range of the distribution is somewhere between 150 and 250 square feet. • The room sizes range from 100 to 350 square feet.
• Using definitive language for summary statistics, such as mean, median, range, when exact values cannot be determined from the histogram. The histogram shown in the problem indicates that the minimum is between 100 and 150 square feet and the maximum is between 300 and 350 square feet. Consequently the range is somewhere between 150 and 250 square feet. The range cannot be determined exactly from the histogram.	• The range of the distribution is somewhere between 150 and 250 square feet. • The sample median is about 250 square feet.
• Computing the ends of the inner fences in part (b) without either stating a conclusion about outliers or without linking the conclusion to the values of the observations relative to the ends of the inner fences. A justified conclusion is needed to complete the response.	• The interquartile range is $\text{IQR} = 292 - 174 = 118$ square feet. The ends of the inner fences are $Q_1 - 1.5(\text{IQR}) = -3,$ and $Q_3 + 1.5(\text{IQR}) = 469.$ No outliers are present because the minimum room size of 134 square feet is larger than -3 and the maximum room size of 315 square feet is smaller than 469.
• Some students used incorrect formulas for the inner fences, such as median $\pm 1.5(\text{ IQR})$.	• The interquartile range is $\text{IQR} = 292 - 174 = 118$ square feet. The ends of the inner fences are $Q_1 - 1.5(\text{IQR}) = -3,$ and $Q_3 + 1.5(\text{IQR}) = 469.$
• Some responses to part (c) presented statements about skewness being more clearly revealed by the boxplot than the histogram. The question asked about a characteristic of the shape of the room size distribution. Bimodality, not skewness, is the obvious characteristic of the distribution of room sizes that is apparent from the histogram.	• The histogram clearly shows the bimodal nature of the distribution of room sizes, but this is not apparent in the boxplot.

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:

- Make sure students know that describing the shape of a distribution as approximately normal implies a bell-shape or unimodal mound-shape. Emphasize that “normal” or “approximately normal” should be used thoughtfully as opposed to being used as an automatic response.
- When describing distributions, make sure students know to address, in context, the shape, center, and variability, and comment on any unusual features.
- Emphasize that in a statistical setting, the word “range” is a noun referring to a single number that is the difference between the maximum and minimum data values. Stating “the range of the data is from 100 to 350” is not the same as stating “all of the data are in the interval from 100 to 350.”
- When describing the center or spread of a distribution based only on a histogram, have students practice with language that conveys uncertainty arising from not having exact data values. Statements such as “the median is approximately 250,” or “the range is between 150 and 250,” are examples that convey such uncertainty. Definitive statements such as “the median is 250” or “the range is 150,” cannot be made from a histogram alone.
- Communication is important. After computing values of relevant statistics, have students practice completing the response by stating a conclusion and using the values of the statistics to justify the conclusion.
- Give students practice computing fences from the correct formulae. Also, if a student remembers a formula incorrectly during the exam, emphasize that they should follow through with their computation as subsequent components may still earn credit.
- Encourage students to read the question carefully. In part (c) the question asks the student to recognize a characteristic of the shape of a particular distribution (room sizes) that is apparent in the histogram and *not* apparent in the boxplot. Recognizing a shape feature of the boxplot that is *more* apparent than in the histogram not only reverses the direction of the requested comparison, but also recognizes a shape characteristic present in both (instead of a shape characteristic that is present in the histogram but absent from the boxplot).
- Boxplots should not be described as “normal” or “approximately normal” because modality cannot be seen from a boxplot. Remind students that boxplots do give some indication of symmetry or skewness.

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

The updated AP Statistics *Course and Exam Description* (CED), effective Fall 2019, includes instructional resources for AP Statistics teachers to develop students’ broader skills. Please see page 227 of the CED for examples of questioning and instructional strategies designed to develop the skill of *describing data presented graphically*, which was important for this question. 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’ ability to describe their observations in written responses.

In general, review of previous exam questions and chief reader reports will give teachers insight into what constitutes strong statistical reasoning, as well as common student errors and how to address them in the classroom. 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. Beginning in August 2019, you’ll have access to the full range of questions from past exams in AP Classroom. You’ll have access to:

- an online library of AP questions relevant to your course
- personal progress checks with new formative questions
- a dashboard to display results from progress checks and provide real-time insights

Question #2 **Max. Points:** 4 **Mean Score:** 1.77

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 components of an experiment; (2) determine if an experiment has a control group; and (3) describe how experimental units can be randomly assigned to treatments.

This question primarily assesses skills in skill category 1: Selecting Statistical Methods. Skills required for responding to this question include (1.B) Identify key and relevant information to answer a question or solve a problem and (1.C) Describe an appropriate method for gathering and representing data.

This question covers content from Unit 3: Collecting Data of the course framework in the AP Statistics Course and Exam Description. Refer to topic 3.5 and learning objectives VAR-3.A, VAR-3.B, and VAR-3.C.

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 identified the treatments as the sprays with the four different concentrations of the fungus (0 ml/L, 1.25ml/L, 2.5ml/L, or 3.75ml/L). Responses indicated more confusion about the identification of experimental units. Many students correctly indicated the twenty containers, or the communities of insects within the containers, but many other students incorrectly identified individual insects as the experimental units. The key is to realize that treatments (sprays) are applied to entire containers, not individual insects. Furthermore, the response is recorded for containers, the number of insects alive in a container one week after spraying. Many responses correctly identified the response variable as the number of insects alive in each container.

- Many responses to part (b) correctly stated that the experiment has a control group by identifying the containers that were sprayed with the solution containing no fungus as the control group or stating that some containers were given a treatment with no fungus.

- The vast majority of students made a good attempt to describe a method for randomly assigning the sprays to the containers. Some descriptions were incomplete, or the communication was not sufficiently clear.

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 indicated that the insects are the experimental units, rather than the containers. In this case, the decision about which treatment to apply was made container by container, not insect by insect, making the containers the experimental units.	• Because the treatments were applied to containers and the response was measured on containers, the experimental units are the 20 containers, each containing the same number of insects.

• Some responses provided ambiguous response variables, such as “number of insects alive,” which could refer to the total number of insects alive in all containers or the number alive in each container.	• The response variable is the number of insects alive in each container one week after spraying.
• Some responses to part (b) state that there is no control group because all containers were sprayed. There is a lack of understanding that a control group may be treated with an inactive substance.	• Because the 0 ml/L concentration contains no fungus, the containers that are sprayed with the 0 ml/L concentration form the control group.
• Many responses to part (b) identified the control group as the containers that receive the 0 ml/L mixture, but also said these containers did not receive a treatment. This is an incorrect statement, and it often contradicted the response in part (a) that included 0 ml/L as a treatment.	• Because the 0 ml/L concentration contains no fungus, the containers that are sprayed with the 0 ml/L concentration form the control group.
• Some responses to part (c) that used slips of paper to do the random assignment forgot to mix/shuffle the slips before using them to assign treatments to containers (or containers to treatments). Likewise, many responses were ambiguous about whether to select slips without replacement.	• Using 20 equally sized slips of paper, label 5 slips with 0 ml/L, 5 slips with 1.25 ml/L, 5 slips with 2.5 ml/L, and 5 slips with 3.75 ml/L. Mix the slips of paper in a hat. For each container, select a slip of paper from the hat (without replacement) and spray that container with the treatment selected.
• Some responses that used a random number generator (or table of random digits) forgot to indicate that numbers were to be used without replacement. Some responses did not define the interval of values from which they were selecting (and the interval of values they were ignoring) when using a table of random digits.	• Label each container with a unique integer from 1 to 20. Then use a random number generator to choose 15 integers from 1 to 20 without replacement. Use the first 5 of these numbers to identify the 5 containers that will receive the 0 ml/L treatment. Use the second 5 of these numbers to identify the 5 containers that will receive the 1.25 ml/L treatment. Use the third 5 of these numbers to identify the 5 containers that will receive the 2.5 ml/L treatment. The remaining 5 containers will receive the 3.75 ml/L treatment.
• Some responses didn’t explicitly describe what random device (e.g., slips of paper, random number generator, table of random digits) they were using. For example, “After	• Label each container with a unique integer from 1 to 20. Then use a random number generator to choose 15 integers from 1 to 20 without replacement. Use the first 5 of these numbers to identify the 5 containers that will receive the 0 ml/L treatment. Use the second 5 of these numbers to

numbering the containers from 1 to 20, select 5 random numbers from 1 to 20."	identify the 5 containers that will receive the 1.25 ml/L treatment. Use the third 5 of these numbers to identify the 5 containers that will receive the 2.5 ml/L treatment. The remaining 5 containers will receive the 3.75 ml/L treatment.
• Some responses "pre-grouped" the experimental units into 4 groups of 5 containers and then randomly assigned sprays to the groups. This was not scored essentially correct because it does not allow for all possible random assignments of sprays to containers.	• Label each container with a unique integer from 1 to 20. Then use a random number generator to choose 15 integers from 1 to 20 without replacement. Use the first 5 of these numbers to identify the 5 containers that will receive the 0 ml/L treatment. Use the second 5 of these numbers to identify the 5 containers that will receive the 1.25 ml/L treatment. Use the third 5 of these numbers to identify the 5 containers that will receive the 2.5 ml/L treatment. The remaining 5 containers will receive the 3.75 ml/L treatment.

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:

- To practice identifying experimental units, have students identify the smallest collection of things to which a single treatment is applied.
- Make sure students understand that there should be one value of the response variable for each experimental unit.
- Make sure students understand that there are at least two types of control groups in general: groups of experimental units that receive a treatment with an inactive ingredient (as in this experiment), and groups of experimental units that receive no treatment at all (not in this experiment).
- Remind students that using slips of paper isn't random unless they impose the randomness by mixing/shuffling and that they should be selecting without replacement.
- In describing how to use a random number generator (or table of random digits) for random assignment or random sampling, insist that students include details about whether numbers are used with or without replacement and clearly identify intervals of values from which numbers are selected and intervals of values they are not using.
- When randomly assigning treatments to units (or units to treatments), it is best to have one label/slip per experimental unit, selected without replacement.
- Make sure students explicitly describe the random device (e.g., slips of paper, random number generator, table of random digits) they are using for random assignment. It is not sufficient to indicate, for example, that after numbering the containers from 1 to 20, select 5 random numbers from 1 to 20.
- The only time experimental units should be "pre-grouped" is in a randomized block design, where a relevant blocking variable is used to create the groups.

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

The updated AP Statistics *Course and Exam Description* (CED), effective Fall 2019, includes instructional resources for AP Statistics teachers to develop students' broader skills. Please see page 73 of the CED for sample strategies that focus on topics related to experimental design. 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 determine whether a study involves random sampling versus random assignment.

In general, review of previous exam questions and chief reader reports will give teachers insight into what constitutes strong statistical reasoning, as well as common student errors and how to address them in the classroom.

The Online Teacher Community features many resources shared by other AP Statistics teachers. For example, try entering the keywords "experimental units" in the search bar, then selecting the drop-down menu for "Resource Library." When you filter for "Classroom-Ready Materials," you may find resources such as handouts, data sets, practice questions, and guided notes that can help your students develop a better understanding of experimental design.

Beginning in August 2019, you'll have access to the full range of questions from past exams in AP Classroom. You'll have access to:

- an online library of AP questions relevant to your course
- personal progress checks with new formative questions
- a dashboard to display results from progress checks and provide real-time insights

Question #3 **Max. Points:** 4 **Mean Score:** 1.39

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) use information in a two-way table of relative frequencies to compute joint, marginal, and conditional probabilities; (2) recognize if events are independent; and (3) compute a probability for a binomial distribution.

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 and (3.B) determine parameters for probability distributions.

This question covers content from Unit 4: Probability Rules, Random Variables, and Probability Distributions of the course framework in the AP Statistics Course and Exam Description. Refer to topics 4.3, 4.5, 4.6, and 4.10, and learning objectives VAR-4.A, VAR-4.D, VAR-4.E, UNC-3.B, and UNC-3.C.

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 obtained the joint probability from the table of relative frequencies provided in the stem of the problem, although some responses assumed independence of the two events, without justification, and multiplied marginal probabilities to obtain a joint probability. Many responses correctly computed the conditional probability as $P(\text{never} \mid \text{woman}) = \dfrac{P(\text{never and woman})}{P(\text{woman})} = \dfrac{0.0636}{0.53} = 0.12.$ Many responses correctly computed the probability of the union of two events, responding never or being a woman, as $P(\text{never or woman}) = P(\text{never}) + P(\text{woman}) - P(\text{never and woman}) = 0.12 + 0.53 - 0.0636 = 0.5864.$ Forgetting to include $P(\text{never and woman})$ in the calculation was a common error.

- Independence of the events of being a person who responds never and being a woman could be demonstrated by indicating that the conditional probability computed in section (iii) of part (a), $P(\text{never} \mid \text{woman}) = 0.12$, is equal to the value of $P(\text{never})$ given in the frequency table. Alternatively, the response could indicate that $P(\text{never and woman}) = 0.0636$ reported in section (i) of part (a) is the product of the values for $P(\text{never})$ and $P(\text{woman})$ given in the frequency table. Many responses were unable to complete the justification for independence, and responses were not as strong for part (b) as for the other two parts of the question.

- Many responses to part (c) were able to indicate that the binomial distribution should be used, identify the values of the parameters of the appropriate binomial distribution, and calculate the correct probability. Some common errors are discussed below.

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

Common Misconceptions/Knowledge Gaps	*Responses that Demonstrate Understanding*
• Without justification, assuming the events are independent in part (a), subpart (i), and computing the joint probability as $P(\text{woman}) \times P(\text{never}) = (0.53)(0.12) = 0.0636$ instead of reading the joint probability directly from the frequency table.	• $P(\text{never and woman}) = 0.0636.$

• In the response to part (a), subpart (ii), not subtracting the joint probability, $P(\text{never and woman})$, in the calculation for $P(\text{A or B})$.	• $P(\text{never or woman})$ $= P(\text{never}) + P(\text{woman}) - P(\text{never and woman})$ $= 0.12 + 0.53 - 0.0636$ $= 0.5864$
• Confusion in the use of the symbols $\cap$ (intersection) and | (given)	• $P(\text{never} \cup \text{woman})$ $= P(\text{never}) + P(\text{woman}) - P(\text{never} \cap \text{woman})$ $= 0.12 + 0.53 - 0.0636$ $= 0.5864$
• Confusing $P(\text{never} \mid \text{woman})$ with $P(\text{woman} \mid \text{never})$	$P(\text{never} \mid \text{woman}) = \dfrac{P(\text{never and woman})}{P(\text{woman})} = \dfrac{0.0636}{0.53} = 0.12$
• Misunderstanding independence, saying the following are valid explanations of why events are NOT independent ○ $P(\text{woman}) \neq P(\text{never})$ ○ $P(\text{man and never}) \neq P(\text{woman and never})$ ○ $P(\text{never} \mid \text{woman}) \neq P(\text{woman})$ ○ $P(\text{woman}) \neq P(\text{man})$ ○ $P(\text{woman}) \neq P(\text{woman} \mid \text{never})$	• Statements that can be used to justify independence are ○ $P(\text{woman and never}) = 0.0636$ is the same as $P(\text{woman}) \times P(\text{never}) = (0.53)(0.12) = 0.0636$ ○ $P(\text{never} \mid \text{woman}) = \dfrac{0.0636}{0.53} = 0.12$ is the same as $P(\text{never}) = 0.12$ ○ $P(\text{woman} \mid \text{never}) = \dfrac{0.0636}{0.12} = 0.53$ is the same as $P(\text{woman}) = 0.53$
• Misunderstanding independence, saying the following are valid explanations of why events are independent ○ $P(\text{woman and never})$ is close to $P(\text{man and never})$ ○ $P(\text{woman})$ is close to $P(\text{man})$	• Statements that can be used to justify independence are ○ $P(\text{woman and never}) = 0.0636$ is the same as $P(\text{woman}) \times P(\text{never}) = (0.53)(0.12) = 0.0636$ ○ $P(\text{never} \mid \text{woman}) = \dfrac{0.0636}{0.53} = 0.12$ is the same as $P(\text{never}) = 0.12$ ○ $P(\text{woman} \mid \text{never}) = \dfrac{0.0636}{0.12} = 0.53$ is the same as $P(\text{woman}) = 0.53$
• Confusing independence with correlation. • Using a chi-square test to assess independence for a two-way contingency table. • Confusing independent observations with independent events. • Confusing independence with mutually exclusive events. • Confusing independence with lack of cause and effect. For example, saying "Just because	• Statements that can be used to justify independence are ○ $P(\text{woman and never}) = 0.0636$ is the same as $P(\text{woman}) \times P(\text{never}) = (0.53)(0.12) = 0.0636$ ○ $P(\text{never} \mid \text{woman}) = \dfrac{0.0636}{0.53} = 0.12$ is the same as $P(\text{never}) = 0.12$

you are a woman does not mean that you have to say no."	○ $P(\text{woman} \mid \text{never}) = \frac{0.0636}{0.12} = 0.53$ is the same as $P(\text{woman}) = 0.53$
• Not justifying a valid explanation of independence using the probabilities found in the table. For example, simply stating $P(\text{woman and never}) = P(\text{woman}) \times P(\text{never})$ without showing that this relationship is satisfied by the values in the table of relative frequencies.	• Statements that can be used to justify independence that reference values in the table of relative frequencies are ○ $P(\text{woman and never}) = 0.0636$ is the same as $P(\text{woman}) \times P(\text{never}) = (0.53)(0.12) = 0.0636$ ○ $P(\text{never} \mid \text{woman}) = \frac{0.0636}{0.53} = 0.12$ is the same as $P(\text{never}) = 0.12$ ○ $P(\text{woman} \mid \text{never}) = \frac{0.0636}{0.12} = 0.53$ is the same as $P(\text{woman}) = 0.53$
• Failure to make decision. For example, reporting $P(\text{never} \mid \text{woman}) = \frac{0.0636}{0.53} = 0.12$ is the same as $P(\text{never}) = 0.12,$ without including a statement that this shows that the events are independent.	• The event of randomly selecting a woman is independent of the event of randomly selecting a person who says never because $P(\text{never} \mid \text{woman}) = \frac{0.0636}{0.53} = 0.12$ is the same as $P(\text{never}) = 0.12.$
• In part (c), not recognizing that a binomial distribution should be used. Using a normal or geometric distribution instead.	• Define X as the number of people in a random sample of 5 people who always take their medicine as prescribed. Then X has a binomial distribution with $n = 5$ and $p = 0.54,$ and $P(X \geq 4) = 0.2415.$
• In part (c), not using correct upper bound in the binomial cumulative distribution function. For example, correctly stating $P(\text{at least four always take medicine})$ and then calculating a probability equal to $1 - P(X \leq 4).$	• Define X as the number of people in a random sample of 5 people who always take their medicine as prescribed. Then X has a binomial distribution with $n = 5$ and $p = 0.54,$ and $P(X \geq 4) = 1 - P(X \leq 3) = 1 - 0.7585 = 0.2415.$
• In part (c), forgetting to include binomial coefficients in the probability formula. For example, reporting $P(X \geq 4) = (0.54)^4(0.46)^1 + (0.54)^5(0.46)^0.$	• Define X as the number of people in a random sample of 5 people who always take their medicine as prescribed. Then X has a binomial distribution with $n = 5$ and $p = 0.54,$ and $P(X \geq 4) = \binom{5}{4}(0.54)^4(0.46)^1 + \binom{5}{5}(0.54)^5(0.46)^0 = 0.2415.$

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 communication and application skills for using any probability formula. When communication was weak in how the probability formula applied to the events of the problem, the responses were also weak.

- When using a formula given on the formula sheet, clearly define the labels in the context of the problem. For example, using W and N to indicate the events "woman" and "never" is better than using the generic A and B labels shown on the formula sheet, but using "woman" and "never" provide additional clarity.
- Have students practice answering a question using words in the stem of the problem. For example, in part (b) "Are the events ... independent?" should have an answer of "Yes, the events ... are independent." Completing the response by clearly stating a decision is important. A response that did not state a decision about whether the events are independent, including a check mark without words, was scored as an incorrect response in part (b), even if the response displayed numerical values and formulas that could be used for a valid justification.
- Students should be encouraged to write responses using words and standard statistical notation rather than using notation for calculator functions. In part (c), for example, it is best to define a random variable X as the number of people in a random sample of 5 people who always take their medicine as prescribed, and indicate that X has a binomial distribution with $n = 5$ and $p = 0.54$. Having clearly indicated the use of a binomial distribution with $n = 5$ and $p = 0.54$, the subsequent probability statement only needs to define the event, in this case $X \geq 4$, and report the correct value of the probability, i.e., $P(X \geq 4) = 0.24149$. Displaying calculator functions used in a calculation is not necessary to score full credit.

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

The updated AP Statistics *Course and Exam Description* (CED), effective Fall 2019, includes instructional resources for AP Statistics teachers to develop students' broader skills. See page 90 of the CED for sample strategies that focus on topics related to probability. 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, allows students to analyze existing solutions to determine where errors have occurred, and may help prevent them from making similar types of errors themselves.

In general, review of previous exam questions and chief reader reports will give teachers insight into what constitutes strong statistical reasoning, as well as common student errors and how to address them in the classroom.

The Online Teacher Community features many resources shared by other AP Statistics teachers. For example, to locate resources to help your students with conditional and marginal probabilities, try entering the keywords "two-way table" in the search bar, then selecting the drop-down menu for "Resource Library." When you filter for "Classroom-Ready Materials," you may find resources such as handouts, data sets, practice questions, and guided notes.

Beginning in August 2019, you'll have access to the full range of questions from past exams in AP Classroom. You'll have access to:

- an online library of AP questions relevant to your course
- personal progress checks with new formative questions
- a dashboard to display results from progress checks and provide real-time insights

Question #4 Max. Points: 4 Mean Score: 1.65

What were the responses to this question expected to demonstrate?

The primary goals of this question were to assess a student's ability to perform an appropriate hypothesis test to address a particular question. More specific goals are to assess students' ability to state appropriate hypotheses; identify the name of an appropriate statistical test, check appropriate assumptions/conditions for performing the named test; calculate a correct test statistic and *p*-value; and draw a correct conclusion, with justification, in the context of the study.

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.E) Identify an appropriate inference method for significance tests, (1.F) Identify null and alternative hypotheses, (4.C) Verify that inference procedures apply in a given situation, (3.E) Calculate a test statistic and find a *p*-value, provided conditions for inference are met, and (4.E) Justify a claim using a decision based on a significance test.

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 6.10 and 6.11, and learning objectives VAR-6.H, VAR-6.I, VAR-6.J, VAR-6.K, and DAT-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?

Unlike the past when questions involving tests of hypotheses were scored with respect to a four-section scoring guideline, this question was scored with a three-section guideline. The first section includes the statement of the null and alternative hypotheses and the specification of the test statistic using words or formula. The second section includes verifying conditions for applying the test and computation of the values of the test statistic and *p*-value. The third section included the statement of the conclusion, in the context of the study, with justification based on the results reported in the second section. This will generally be the structure of scoring guidelines for tests of hypotheses in the future.

Section 1:

- Most responses recognized that a hypothesis test should be performed and the test involved proportions for two populations. Most responses indicated that the null hypothesis involved equality of two quantities, even if the responses did not clearly refer to population proportions. Specification of the correct one-sided alternative hypothesis was less well accomplished. Common errors involved non-existent or incomplete descriptions of parameters.

- Most responses indicated that a two-sample *z*-test should be used, although a substantial number of responses did not complete the identification of the test by including that it is a test for the difference in two population proportions.

Section 2:

- Most responses made attempts to verify conditions. Common errors are discussed below.

- Responses showed that most students knew how to use their calculator to correctly calculate the value of a test statistic and the associated *p*-value. A common error was that the *p*-value did not match the alternative hypothesis stated in the response.

- Some responses used a confidence interval to conduct the test. Because the alternative hypothesis is one-sided, it took special care to apply this approach correctly, especially when a two-sided confidence interval was used.

Section 3:

- Essentially every response that computed the value of a test statistic, or computed a confidence interval, stated a conclusion in the context of a comparison of the proportions of kochia plants resistance to glyphosate in 2014 and 2017. Some responses that rejected the null hypothesis of equal proportions were incomplete because they failed to indicate if the data supported the alternative hypothesis of a higher proportion of resistant plants in 2017. Other common errors are addressed below.

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

Common Misconceptions/Knowledge Gaps	*Responses that Demonstrate Understanding*
• Does not refer to population proportions in the statement of hypotheses (e.g., refers to means, or states hypotheses in words and never refers to population proportions or uses non-standard notation without defining the notation).	• Let p_{14} represent the proportion of the population of kochia plants that were resistant to glyphosate in 2014. Let p_{17} represent the proportion of the population of kochia plants that were resistant to glyphosate in 2017. The null hypothesis is $\mathrm{H}_0 : p_{17} - p_{14} = 0$ and the alternative hypothesis is $\mathrm{H}_a : p_{17} - p_{14} > 0$.
• States a two-sided alternative hypothesis or reverses the direction of the one-sided alternative hypothesis.	• Let p_{14} represent the proportion of the population of kochia plants that were resistant to glyphosate in 2014. Let p_{17} represent the proportion of the population of kochia plants that were resistant to glyphosate in 2017. The null hypothesis is $\mathrm{H}_0 : p_{17} - p_{14} = 0$ and the alternative hypothesis is $\mathrm{H}_a : p_{17} - p_{14} > 0$.
• Uses subscripts that do not clearly distinguish the populations, e.g. $\mathrm{H}_0 : p_1 - p_2 = 0$ and $\mathrm{H}_a : p_1 - p_2 > 0$.	• Let p_{14} represent the proportion of the population of kochia plants that were resistant to glyphosate in 2014. Let p_{17} represent the proportion of the population of kochia plants that were resistant to glyphosate in 2017. The null hypothesis is $\mathrm{H}_0 : p_{17} - p_{14} = 0$ and the alternative hypothesis is $\mathrm{H}_a : p_{17} - p_{14} > 0$.
• Provides incomplete or no definition of parameters used in the response.	• Let p_{14} represent the proportion of the population of kochia plants that were resistant to glyphosate in 2014. Let p_{17} represent the proportion of the population of kochia plants that were resistant to glyphosate in 2017. The null hypothesis is $\mathrm{H}_0 : p_{17} - p_{14} = 0$ and the alternative hypothesis is $\mathrm{H}_a : p_{17} - p_{14} > 0$.
• Refers to "2 sample *z*-test" without reference to a difference in proportions.	• An appropriate test procedure is a two-sample *z*-test for a difference in population proportions.

• Correctly indicates that a two-sample test for the difference in proportions should be used but contradicts that statement by applying a different test.	• An appropriate test procedure is a two-sample z-test for a difference in population proportions.
• Verifying conditions: ○ Does not indicate that there are <u>two independent</u> random sample<u>s.</u> ○ Does not verify that all four of the expected (or observed) counts are greater than 10 to check the approximate normality of the sampling distribution of $\hat{p}_{17} - \hat{p}_{14}$. ○ Does not check that it is reasonable to assume that the individuals within <u>each</u> sample can be viewed as independent even though the sampling is done without replacement (10% condition).	• Verifying conditions ○ The first condition for applying the test is that the data are gathered from independent random samples from the populations of kochia plants in the Western United States in 2014 and 2017. We are told that a random sample of 61 kochia plants was selected in 2014 and a second random sample of 52 kochia plants was selected in 2017. It is reasonable to assume that the 2017 sample of plants was in no way influenced by the 2014 sample of plants<u>.</u> ○ The second condition is that the sampling distribution of the test statistic is approximately normal. This condition is satisfied because the expected counts under the null hypothesis are all greater than 10. The pooled estimate of the proportion of resistant plants is $\hat{p}_c = \frac{(61)(0.197)+(52)(0.385)}{61+52} = 0.2835.$ The estimates of the expected counts are $61(0.2835) = 17.29,$ $61(1-0.2835) = 43.71,$ $52(0.2835) = 14.74,$ and $51(1-0.2835) = 37.25,$ all of which are greater than 10. ○ Because the sampling must have been done without replacement, the independence condition for each sample should be checked, but it is reasonable to assume that each population has millions of plants, so it is reasonable to assume that sample sizes are less than 10% of the respective population sizes.
• Computes an incorrect value for the test statistic or *p*-value.	• Using the pooled estimate of the proportion of resistant plants, $\hat{p}_c = 0.2835,$ the value of the test statistic is $$z = \frac{0.385 - 0.197}{\sqrt{\left(\frac{(0.2835)(0.7165)}{61} + \frac{(0.2835)(0.7165)}{52}\right)}} = 2.21.$$ The *p*-value is 0.0135.
• Provides a two-sided confidence interval for the difference in two population proportions instead of a test statistic, but does not adjust for the one-sided alternative hypothesis.	• A 90% confidence interval for $p_{17} - p_{14}$ is $(0.049, 0.327)$. Because the lower end of this confidence interval is above zero, we have evidence at the $\alpha = 0.05$ level of significance to reject the null hypothesis of equal proportions and conclude that the population proportion of resistant plants is higher in 2017 than in 2014.

• States a conclusion that implies that the alternative hypothesis has been proven rather than having convincing evidence to support the alternative hypothesis.	• Because the *p*-value is less than $\alpha = 0.05$, we have convincing statistical evidence to conclude that the proportion of resistant plants in the 2017 population of kochia plants is greater than the proportion of resistant plants in the 2014 population of kochia plants.
• States a conclusion in terms of only the null hypothesis without stating that there is convincing evidence to support the alternative hypothesis that the population proportion of resistant plants is higher in 2017 than in 2014.	• Because the *p*-value is less than $\alpha = 0.05$, we have convincing statistical evidence to conclude that the proportion of resistant plants in the 2017 population of kochia plants is greater than the proportion of resistant plants in the 2014 population of kochia plants.
• States a conclusion that is inconsistent with the stated alternative hypothesis. In this case, with $\mathrm{H_a} : p_{17} - p_{14} > 0$, states the null hypothesis may be rejected and that there is evidence to support the claim that the proportion of resistant plants is greater in 2014.	• Because the *p*-value is less than $\alpha = 0.05$, we have convincing statistical evidence to conclude that the proportion of resistant plants in the 2017 population of kochia plants is greater than the proportion of resistant plants in the 2014 population of kochia plants.
• Reports a *p*-value but neglects to justify the conclusion by stating that the *p*-value is less than the alpha value of 0.05 given in the stem of the question.	• Because the *p*-value is less than $\alpha = 0.05$, we have convincing statistical evidence to conclude that the proportion of resistant plants in the 2017 population of kochia plants is greater than the proportion of resistant plants in the 2014 population of kochia plants.

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?

- Encourage students to clearly define all parameters in the context of the problem.
- When labeling parameters, encourage students to use labels that will help them keep track of the direction of the alternative hypothesis, e.g., p_{17}, p_{14} instead of p_1, p_2.
- Encourage students to clearly communicate the correct procedure. (This is usually more correctly done when they use words, not formulas.)
- Provide students with problems where they are given sample proportions instead of observed counts.
- Encourage students to check all three conditions for both samples.
- Students should not conduct both a significance test and a confidence interval, unless specifically asked to do so. A confidence interval and a hypothesis test are often considered parallel responses, so each are read separately and the score for the weaker solution is awarded.
- Applying a confidence interval approach for a one-sided alternative hypothesis is difficult because students need to adjust the confidence level to align with the level of significance. Encourage students to conduct a significance test when they are asked to determine if the data provide convincing evidence for a claim, rather than a confidence interval.
- Encourage students not to show calculator syntax when calculating a *p*-value.
- Encourage students to make a direct comparison of the *p*-value to alpha level and use the result of that comparison to support their conclusion.
- Encourage students to clearly communicate that the data provide convincing evidence to support the alternative hypothesis (and not just state that there is no evidence for the null hypothesis).

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

The updated AP Statistics *Course and Exam Description* (CED), effective Fall 2019, includes instructional resources for AP Statistics teachers to develop students' broader skills. Please see page 128 of the CED for sample strategies that focus on topics related to inference for categorical data. 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, can help students practice communicating their conclusions from a hypothesis test by providing them with a sentence structure to model their responses after.

In general, review of previous exam questions and chief reader reports will give teachers insight into what constitutes strong statistical reasoning, as well as common student errors and how to address them in the classroom.

The Online Teacher Community features many resources shared by other AP Statistics teachers. For example, to locate resources to help your students with testing for the difference in population proportions, try entering the keywords "two-sample z-test" in the search bar, then selecting the drop-down menu for "Resource Library." When you filter for "Classroom-Ready Materials," you may find resources such as handouts, data sets, practice questions, and guided notes.

Beginning in August 2019, you'll have access to the full range of questions from past exams in AP Classroom. You'll have access to:

- an online library of AP questions relevant to your course
- personal progress checks with new formative questions
- a dashboard to display results from progress checks and provide real-time insights

Question #5 **Max. Points:** 4 **Mean Score:** 1.62

What were the responses to this question expected to demonstrate?

The primary goals of this question are to assess a student's ability to (1) evaluate a percentile of a normal distribution; (2) evaluate a probability for a normal distribution; and (3) compute an expected value for a random variable with two possible outcomes.

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 and (3.B) Determine parameters for probability distributions.

This question covers content from multiple units, including Unit 1: Exploring One-Variable Data, Unit 4: Probability Rules, Random Variables, and Probability Distributions and Unit 5: Sampling Distributions of the course framework in the AP Statistics Course and Exam Description. Refer to topics 1.10, 4.8, and 5.2, and learning objectives VAR-2.B, VAR-5.C, and VAR-6.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?

- Most responses made a serious attempt to compute the 25^{th} percentile of the smartphone battery lifetime distribution in part (a). To receive complete credit, responses needed to clearly indicate that the objective was to find the 25^{th} percentile of a normal distribution with mean 30 months and standard deviation 8 months. This was most effectively done using words and/or graphs of normal density functions. As indicated below in the discussion of common errors, responses that relied solely on calculator function syntax did not receive full credit unless arguments corresponding to normal distribution parameters were carefully labeled.

- In responses to part (b), the vast majority of students were able to use a calculator to find the probability that a randomly selected battery would need to be replaced within 24 months of the date of purchase. To receive complete credit, the response needed to indicate that the probability that the lifetime of a randomly selected battery is less than 24 months is computed for a normal distribution with mean 30 months and standard deviation 8 months.

- Part (c) was the most difficult part of this question. A substantial number of responses did not display an appropriate method for computing an expectation. Students were expected to incorporate the probability computed in part (b) into the computation of the expectation in part (c). Some responses that displayed a relevant formula for the computation of the expected gain incorrectly assigned the probability from part (b) to \$50 instead of $-\$150$.

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 failed to convey that battery life span has a normal distribution with mean 30 months and standard deviation 8 months, vital information for developing complete responses for both part (a) and part (b).	• The 25^{th} percentile of the standard normal distribution is $-0.6745.$ Consequently, the 25^{th} percentile of a normal distribution with mean 30 months and standard deviation 8 months is $30 + 8(-0.6745) = 24.6$ months.

• Many responses tried to convey information about the distribution through calculator function syntax with unlabeled inputs. For example: in part (a) invnorm(0.25, 30, 8) rather than invnorm(0.25, μ=30, σ=8); or in part (b), normalcdf(-9999, 24, 30, 8) rather than normalcdf(-9999, 24, μ=30, σ=8). Unlabeled calculator function syntax did not receive complete credit.	• The 25th percentile of the standard normal distribution is -0.6745. Consequently, the 25th percentile of a normal distribution with mean 30 months and standard deviation 8 months is $30 + 8(-0.6745) = 24.6$ months. • The probability that a randomly selected customer will need to request a replacement because the battery fails within 24 months from the date of purchase is $P(\text{life span} \le 24 \text{ months}) = P\left(Z \le \frac{24-30}{8}\right) = 0.2266$.
• Calculator commands were incorrectly applied in some responses. For example, using normalcdf to find a normal distribution percentile in part (a), or using normalpdf to calculate the probability in part (b).	• The 25th percentile of the standard normal distribution is -0.6745. Consequently, the 25th percentile of a normal distribution with mean 30 months and standard deviation 8 months is $30 + 8(-0.6745) = 24.6$ months. • The probability that a randomly selected customer will need to request a replacement because the battery fails within 24 months from the date of purchase is $P(\text{life span} \le 24 \text{ months}) = P\left(Z \le \frac{24-30}{8}\right) = 0.2266$.
• In part (b), some responses failed to provide sufficient information about the boundary and direction for the outcome of interest that the lifetime of a randomly selected battery is less than 24 months.	• The probability that a randomly selected customer will need to request a replacement because the battery fails within 24 months from the date of purchase is $P(\text{life span} \le 24 \text{ months}) = P\left(Z \le \frac{24-30}{8}\right) = 0.2266$.
• Some responses to part (c) used probabilities that did not sum to one in the expectation calculation.	• The company's expected gain for each warranty purchased is $(\$50) \times P(\text{life span} > 24 \text{ months}) + (-\$150) \times P(\text{life span} \le 24 \text{ months})$ $= (\$50) \times (0.7734) + (-\$150) \times (0.2266) = \$4.68$.
• Some responses to part (c) switched the two probabilities: matching the probability of a \$50 gain with the $-\$150$ outcome, and the probability of a $-\$150$ gain with the \$50 outcome.	• The company's expected gain for each warranty purchased is $(\$50) \times P(\text{life span} > 24 \text{ months}) + (-\$150) \times P(\text{life span} \le 24 \text{ months})$ $= (\$50) \times (0.7734) + (-\$150) \times (0.2266) = \$4.68$.
• Some responses to part (c) incorrectly used percentages in place of probabilities in calculating the expectation.	• The company's expected gain for each warranty purchased is $(\$50) \times P(\text{life span} > 24 \text{ months}) + (-\$150) \times P(\text{life span} \le 24 \text{ months})$ $= (\$50) \times (0.7734) + (-\$150) \times (0.2266) = \$4.68$.

• Some responses to part (c) provided clearly unreasonable values for the company's expected gain for each warranty purchased because they were either less than $-\$150$ or greater than $\$50$.	• The company's expected gain for each warranty purchased is $(\$50) \times P(\text{life span} > 24 \text{ months}) + (-\$150) \times P(\text{life span} \le 24 \text{ months})$ $= (\$50) \times (0.7734) + (-\$150) \times (0.2266) = \$4.68.$

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?

- To identify the statistical distribution being used, students should be encouraged to use words and standard statistical notation, or a sketch of a normal curve with tick marks on the horizontal axis locating the mean, and the mean ± 1 (or more) standard deviations, rather than using calculator function syntax.
- If calculator function syntax is used to identify a normal distribution, students should be reminded that they need to label the mean and standard deviation parameters in the arguments of the calculator function.
- Remind students to use standard notation for parameters of a probability distribution. If non-standard symbols are used, students should define what the symbols mean.
- Encourage students to identify the boundary and direction of regions of interest by using correct statistical notation, e.g., $P(X < 24)$, or providing a sketch of the normal distribution with the boundaries of the region of interest marked and the corresponding area under the curve shaded.
- In computing the expectation of a discrete random variable, remind students
    - to check that the probabilities sum to one.
    - to check that each outcome is associated with the correct probability.
    - that percentages cannot be substituted for probabilities.
    - that an expected value of a random variable cannot be outside the range of values that the random variable takes on.
- Some student responses did not provide sufficient details of their calculation. Insist that students show their work.

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

The updated AP Statistics *Course and Exam Description* (CED), effective Fall 2019, includes instructional resources for AP Statistics teachers to develop students' broader skills. Please see pages 229-230 of the CED for examples of questioning and instructional strategies designed to develop the skill of *using probability and simulation*. A table of representative instructional strategies, including definitions and explanations of each, is included on pages 213-223 of the CED. The strategy "Debriefing," for example, can help students clarify their misconceptions by discussing their understanding of percentiles and probability distributions.

In general, review of previous exam questions and chief reader reports will give teachers insight into what constitutes strong statistical reasoning, as well as common student errors and how to address them in the classroom.

The Online Teacher Community features many resources shared by other AP Statistics teachers. For example, to locate resources to help your students with calculations related to random variables and probability distributions, try entering the keywords "expected value" in the search bar, then selecting the drop-down menu for "Resource Library." When you filter for "Classroom-Ready Materials," you may find resources such as handouts, data sets, practice questions, and guided notes.

Beginning in August 2019, you'll have access to the full range of questions from past exams in AP Classroom. You'll have access to:

- an online library of AP questions relevant to your course

- personal progress checks with new formative questions
- a dashboard to display results from progress checks and provide real-time insights

Question #6 **Max. Points:** 4 **Mean Score:** 0.75

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 the population to which results from a random sample may be generalized; (2) describe a disadvantage of using a sample mean rather than a sample median to indicate typical values when the sample distribution is skewed; (3) describe how the theoretical sampling distribution of the sample median could be constructed; (4) construct an approximate confidence interval for a population median using results from a bootstrap procedure; and (5) interpret a confidence interval.

This question represents the investigative task of the free-response section. The investigative task is intended to assess understanding of several content areas contained in the course framework and also to assess the ability to extend statistical reasoning by integrating statistical ideas and applying them in a new context or in a non-routine way.

This question primarily assesses skills in multiple skill categories, including skill category 2: Data Analysis, skill category 3: Use Probability and Simulation, and skill category 4: Statistical Argumentation. Skills required for responding to this question include (2.C) Calculate summary statistics, relative positions of points within a distribution, correlation, and predicted response, (3.A) Determine relative frequencies, proportions, or probabilities using simulation or calculations, (3.C) Describe probability distributions, (3.B) Determine parameters for probability distributions, (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 3: Collecting Data, Unit 4: Probability Rules, Random Variables, and Probability Distributions and Unit 5: Sampling Distributions of the course framework in the AP Statistics Course and Exam Description. Refer to topics 1.7, 1.8, 3.2, 4.2, and 5.3, and learning objectives UNC-1.K, UNC-1.M, DAT-2.A, DAT-2.B, UNC-3.H, and UNC-2.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 were able to identify a population of one-bedroom apartments, but some responses failed to limit the description to one-bedroom apartments listed on the web site.

- In part (b) many responses recognized that the sample mean would be affected by the right skewness of the distribution of rental prices causing it to tend to be larger than typical rental prices, while the sample median tends to be closer to the typical rental prices. Some responses only stated that the sample mean would tend to be larger than the sample median and did not complete the argument by indicating which would tend to be closer to typical rental prices.

- Many responses talked about taking samples of 50 one-bedroom apartments from the population that was described in part (b) and computing the median of each sample. The medians for all possible samples of 50 one-bedroom apartments would need to be computed to obtain the theoretical sampling distribution of the sample median. However, many responses indicated that a sample of all possible samples would be selected, which would only yield an approximation to the theoretical sampling distribution of the sample median.

- Many students were able to compute percentiles from the results in the table presented in part (d). Most students who completed part (d) were able to count the number of medians in the table that fell between the percentiles computed in part (d) to provide a response to part (e).

- Interpretation of the confidence interval was less well done in responses to part (f) than the construction of the confidence interval. A surprising number of responses ignored the work done in parts (d) and (e) and constructed the confidence interval as

$$\text{(Average median)} \pm \text{(some multiplier)(standard deviation of sample median)},$$

where the average and standard deviation were computed from the information in the table of bootstrap results preceding part (d). While this response may have some merit, it did not follow the instruction in part (f) to "use your values from parts (d) and (e)."

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

Common Misconceptions/Knowledge Gaps	*Responses that Demonstrate Understanding*
• In part (a), not including "one-bedroom apartments" or "web site" in describing the population.	• Because random sampling was used, the results of the sample may be generalized to the population of rental prices for one-bedroom apartments in the city that are listed on this particular web site at the time the sample was taken.
• In part (a) specifying one-bedroom apartments "similar" to those on the website. This could be something similar to the population, but it is not the population.	• Because random sampling was used, the results of the sample may be generalized to the population of rental prices for one-bedroom apartments in the city that are listed on this particular web site at the time the sample was taken.
• In part (b), recognizing that the sample mean tends to be larger than the median for right skewed distributions, but not indicating that this tends to result in a sample mean that is larger than typical rental prices.	• Because the distribution of rental prices is skewed to the right, the sample median provides a better indicator of typical rental prices than the sample mean. Some very large rental prices may result in a sample mean that is substantially larger than the more typical rental prices. As a result, the sample mean would overestimate the typical rental price, whereas the sample median would be a more accurate representation of typical rental prices
• In part (d), not indicating that medians must be computed for all possible samples of 50 one-bedroom apartments that could be selected from the population. For example, this includes responses that indicate "take many samples of 50 one-bedroom apartments." This would not provide the **theoretical** sampling distribution for the sample median, only an approximation.	• To determine the sampling distribution of median rental prices for random samples of 50 one-bedroom apartments from this population, Emma would need to obtain every possible sample of 50 one-bedroom apartments from this web site and compute the median of each sample. The collection of all possible sample medians is the theoretical sampling distribution for sample median.
• In part (d) computing a mean and standard deviation from the information in the table and using a normal distribution based on those values to compute percentiles. Inspection of the table suggests that a normal distribution would not provide a good approximation to the bootstrap distribution of sample medians. Use	• Because $n = 15{,}000,$ compute $(0.05)(n) = (0.05)(15{,}000) = 750$ and compute $(0.95)(15{,}000) = 14{,}250.$ The 5th percentile is a value, say $x_{0.05},$ such that at least 750 values in the table are less than or equal to $x_{0.05}$ and at least 14,250 are greater than or equal to $x_{0.05}.$ Cumulate frequencies starting with the smallest sample median listed in the table and going toward the largest (going down columns) until

the actual distribution provided by the table.	you first reach 750 values, to obtain $x_{0.05} = \$2,500$. Similarly, $x_{0.95} = \$2,950$.
• In part (e) not following directions. For example, calculating $95\% - 5\%$ to get 90% rather than adding frequencies and dividing the sum by 15,000.	• The percentage of bootstrap medians between (and including) the values found in part (d) for the 5th and 95th percentiles is $\frac{14,404}{15,000} \times 100\% = 96.03\%$.
• For part (f), not following instructions by failing to use the values computed in parts (d) and (e) to construct and interpret the confidence interval. • Not providing a correct interpretation of whatever confidence interval presented in the response.	• From the results in part (d) and part (e), an approximate 96% confidence interval for the median rental price of all one-bedroom apartments listed on this Web site for this city is $(\$2,500, \$2,900)$. We are approximately 96 percent confident that the median rental price of all one-bedroom apartments listed on this Web site is between \$2,500 and \$2,950.

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?

- As you go through examples of random samples in the course, have students clearly describe the population each random sample was selected from.
- Encourage students to carefully read every question and make sure their response fully answers the question. Provide practice for students to explain "why" instead of just listing traits.
- It is good practice to use simulations to help students understand sampling distributions for summary statistics (using applets or activities) but be sure to explain that simulated sampling distributions are approximations to theoretical sampling distributions, with the approximations becoming more accurate as the number of samples used in the simulation is increased. Stress that the theoretical sampling distribution is obtained by evaluating the statistic of interest for all possible samples of a particular size that could be selected from the population, which may be impractical to compute.
- Questions may arise about performing a simulation to examine the sampling distribution of a statistic of interest (mean, proportion, median, standard deviation) when the population is unknown. This would be an opportunity to talk about taking a large random sample from the population, and subsequently taking samples from the original sample (sampling with replacement) to mimic taking repeated samples from the unknown population. This is the idea of the bootstrap.
- Use examples to illustrate when it is appropriate to use a normal distribution and that sometimes it is NOT appropriate.
- Stress to students to follow the directions of the question. The investigative task, Question 6, often has directions to follow in order to scaffold through the question. Practice throughout the course with examples from investigative tasks from exams published on the College Board website. Do not save work on investigative tasks until the end of the course.
- To help students better understand how to interpret a confidence interval, present examples of good and bad interpretations throughout the latter part of the course.

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

The updated AP Statistics *Course and Exam Description* (CED), effective Fall 2019, includes instructional resources for AP Statistics teachers to develop students' broader skills. Please see pages 231-232 of the CED for examples of questioning and instructional strategies designed to develop the skill of *statistical argumentation*. 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, allows students to analyze both correct and incorrect interpretations of a confidence interval and help prevent them from making errors in their own interpretations.

In general, review of previous exam questions and chief reader reports will give teachers insight into what constitutes strong statistical reasoning, as well as common student errors and how to address them in the classroom.

The Online Teacher Community features many resources shared by other AP Statistics teachers. For example, to locate resources to help your students construct and interpret confidence intervals, try entering the keywords "confidence interval" in the search bar, then selecting the drop-down menu for "Resource Library." When you filter for "Classroom-Ready Materials," you may find resources such as handouts, data sets, practice questions, and guided notes.

Beginning in August 2019, you'll have access to the full range of questions from past exams in AP Classroom. You'll have access to:

- an online library of AP questions relevant to your course
- personal progress checks with new formative questions (including formative Investigative Tasks)
- a dashboard to display results from progress checks and provide real-time insights