
AP[®] Statistics

Sample Student Responses and Scoring Commentary

Inside:

Free Response Question 6

- Scoring Guideline**
- Student Samples**
- Scoring Commentary**

AP[®] STATISTICS

2019 SCORING GUIDELINES

Question 6

Intent of Question

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.

Solution

Part (a):

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 website at the time the sample was taken.

Part (b):

Because the distribution of the 50 rental prices in the sample 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 results 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.

Part (c):

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 website and compute the median of each sample. The collection of all possible sample medians is the theoretical sampling distribution for sample median.

Part (d):

- (i) $(0.05)(15,000) = 750$ and $(0.95)(15,000) = 14,250$. 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 you first reach 750 values, to obtain $x_{0.05} = \$2,500$.
- (ii) Similarly, $x_{0.95} = \$2,950$.

Part (e):

The percentage of bootstrap medians between (and including) the values found in part (d) for the 5th and 95th percentiles is

$$\frac{14,404}{15,000} \times 100\% \approx 96.03\%$$

AP[®] STATISTICS

2019 SCORING GUIDELINES

Question 6 (continued)

Part (f):

From the results in part (d) and part (e), an approximate 96 percent confidence interval for the median rental price of all one-bedroom apartments listed on this website for this city is (\$2,500, \$2,950). We are approximately 96 percent confident that the median rental price of all one-bedroom apartments listed on this website for this city is between \$2,500 and \$2,950.

Scoring

This question is scored in four sections. Section 1 consists of parts (a) and (b), section 2 consists of part (c), section 3 consists of parts (d) and (e), and section 4 consists of part (f). Sections 1, 2, 3, and 4 are each scored as essentially correct (E), partially correct (P), or incorrect (I).

Section 1 is scored as follows:

Essentially correct (E) if the response satisfies the following three components:

1. The correct population (listings of one-bedroom apartments on the website) is identified in part (a).
2. In part (b), identifying that using the sample mean instead of the sample median overestimates the typical rental price.
3. The disadvantage of using the sample mean that is reported in part (b) is correctly linked to some feature of the distribution (e.g. skewness) that is evident in the histogram.

Partially correct (P) if the response satisfies only two of the three components.

Incorrect (I) if the response does not meet the criteria for E or P.

Note: Responses that refer to the mean being larger than the median in a skewed right distribution alone is not sufficient to satisfy component 2.

Section 2 is scored as follows:

Essentially correct (E) if the response satisfies the following two components:

1. Indicates that Emma would need to obtain every possible sample of 50 one-bedroom apartments.
2. Indicates that Emma would need to compute the median rental price for each sample.

Partially correct (P) if the response satisfies only one of the two components.

Incorrect (I) if the response does not satisfy the criteria for E or P.

Section 3 is scored as follows:

Essentially correct (E) if the response satisfies the following two components:

1. Correct values for the 5th percentile and the 95th percentile are reported in part (d).
2. The correct percentage of bootstrap samples that produced sample medians at or between the two values, if they are plausible, reported in part (d) is reported in part (e).

Partially correct (P) if the response satisfies only one of the two components.

Incorrect (I) if the response does not satisfy the criteria for E or P.

Note: Plausible values for part (d) will be considered values between 2,345 and 3,062.5.

AP[®] STATISTICS
2019 SCORING GUIDELINES

Question 6 (continued)

Section 4 is scored as follows:

Essentially correct if the response in part (f) satisfies the following three components:

1. Uses \$2500 and \$2950 or the values of the percentiles reported in part (d) as the endpoints of the confidence interval.
2. Indicates an approximate 90 or 96 percent level of confidence or a level of confidence consistent with part (e).
3. Makes a correct statement in context indicating that the confidence interval is for the median.

Partially correct if the response satisfies only two of the three components.

Incorrect if the response does not satisfy the criteria for E or P.

Note: Since rental prices from the population are discrete values, the true confidence level of the interval from part (d) is unknown. A correctly calculated part (e) is a way to estimate the confidence level; from Emma's sample the confidence level is estimated to be approximately 96 percent. The process described in part (d) for calculating the interval will result in a confidence level of at least 90 percent. For these reasons, confidence levels of either 90 or 96 percent satisfy component 2.

Each essentially correct (E) section counts as 1 point, and a partially correct (P) section counts as $\frac{1}{2}$ point.

- | | |
|----------|-----------------------------|
| 4 | Complete Response |
| 3 | Substantial Response |
| 2 | Developing Response |
| 1 | Minimal Response |

If a response is between two scores (for example, $2\frac{1}{2}$ points), use a holistic approach to decide whether to score up or down depending on the strength of the response and communication.

6A1

6A1

STATISTICS
SECTION II
Part B
Question 6

Spend about 25 minutes on this part of the exam.
Percent of Section II score—25

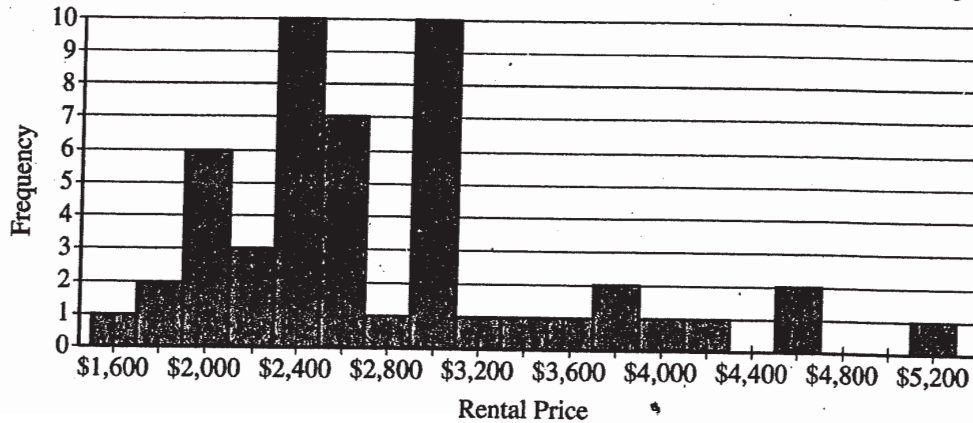
Directions: Show all your work. Indicate clearly the methods you use, because you will be scored on the correctness of your methods as well as on the accuracy and completeness of your results and explanations.

6. Emma is moving to a large city and is investigating typical monthly rental prices of available one-bedroom apartments. She obtained a random sample of rental prices for 50 one-bedroom apartments taken from a Web site where people voluntarily list available apartments.

(a) Describe the population for which it is appropriate for Emma to generalize the results from her sample.

The population of interest is all ^{Web-available} listed one-bedroom apartments in this large city. (voluntarily)

The distribution of the 50 rental prices of the available apartments is shown in the following histogram.



Unauthorized copying or reuse of any part of this page is illegal.

GO ON TO THE NEXT PAGE.

(b) Emma wants to estimate the typical rental price of a one-bedroom apartment in the city. Based on the distribution shown, what is a disadvantage of using the mean rather than the median as an estimate of the typical rental price?

The distribution for the rental price is strongly skewed to the right. The mean is sensitive/not resistant to outliers, and thus is abnormally high, not representative of the typical ^{monthly} rental price of all ^{web-}available listed one-bedroom apartment in this city.

(c) Instead of using the sample median as the point estimate for the population median, Emma wants to use an interval estimate. However, computing an interval estimate requires knowing the sampling distribution of the sample median for samples of size 50. Emma has one point, her sample median, in that sampling distribution. Using information about rental prices that are available on the Web site, describe how someone could develop a theoretical sampling distribution of the sample median for samples of size 50.

~~Determine the sample median for any possible sample of 50 web-listed~~

~~Find the median of all web-listed one-bedroom apartments - this ^{is this large city} will be the center of the sampling distribution. Then, determine the median of all possible samples of 50 web-listed one-bedroom apartments in this large city. Centered at the population median, this forms the theoretical sampling distribution.~~

Find the median rental price for all possible combinations of 50 ^{available} web-listed one-bedroom apartments in this city. Record all medians in a table with the corresponding frequency. Find the 50th percentile median; this will be the center of the sampling distribution. These medians, together, represent the sampling distribution for median ^{monthly rental prices} of samples of ^{available} web-listed one-bedroom apartments in this large city of size 50.

Unauthorized copying or reuse of any part of this page is illegal.

GO ON TO THE NEXT PAGE.

Because Emma does not have the resources to develop the theoretical sampling distribution, she estimates the sampling distribution of the sample median using a process called bootstrapping. In the bootstrapping process, a computer program performs the following steps.

- Take a random sample, with replacement, of size 50 from the original sample.
- Calculate and record the median of the sample.
- Repeat the process to obtain a total of 15,000 medians.

Emma ran the bootstrap process, and the following frequency table is the bootstrap distribution showing her results of generating 15,000 medians.

Bootstrap Distribution of Medians					
Median	Frequency	Median	Frequency	Median	Frequency
2,345	1	2,585	1	2,825	247
2,390	13	2,587.5	171	2,837.5	7
2,395	18	2,600	22	2,847.5	1
2,400	56	2,612.5	1,190	2,872.5	317
2,445	4	2,625	174	2,885	10
2,447.5	56	2,672.5	5	2,950	700
2,450	55	2,675	1,924	2,962.5	93
2,475	3	2,687.5	1,341	2,972.5	6
2,495	66	2,700	2,825	2,975	65
2,497.5	136	2,735	35	2,985	12
2,500	1,899	2,747.5	619	2,987.5	1
2,522.5	2	2,750	2	2,995	6
2,525	945	2,795	278	3,000	2
2,550	1,673	2,812.5	16	3,062.5	3

The bootstrap distribution provides an approximation of the sampling distribution of the sample median. A confidence interval for the median can be constructed using a percentage of the values in the middle of the bootstrap distribution.

(d) Use the frequency table to find the following.

- (i) Value of the 5th percentile:
 $.05(15000) = 750$
 750th smallest median is \$2500.

- (ii) Value of the 95th percentile:
 750th largest median is \$2950

According to the frequency table, the 5th percentile for the median rental price generated with the "bootstrapping" is \$2500.

According to the frequency table, the 95th percentile for the median rental price generated with the bootstrapping method is \$2950.

6A4

6A4

- (e) Find the percentage of bootstrap medians in the table that are equal to or between the values found in part (d).

Sum of frequencies 2500-2950 inclusive

$$= 700 + 10 + \dots + 1899 = 14404$$

$$\frac{14404}{15000} = .9603 = \text{proportion of medians between 2500-2950 inclusive}$$

The percentage of bootstrap medians in the table between or equal to the \$2500 and \$2950 found in part (d-i) and (d-ii) respectively is 96.03%.

- (f) Use your values from parts (d) and (e) to construct and interpret a confidence interval for the median rental price.

I am 96.03% confident that the median monthly rental price for all ^{available} web-listed one-bedroom apartments in this large city is between \$2500 and \$2950.

6B1

6B1

STATISTICS
SECTION II

Part B

Question 6

Spend about 25 minutes on this part of the exam.

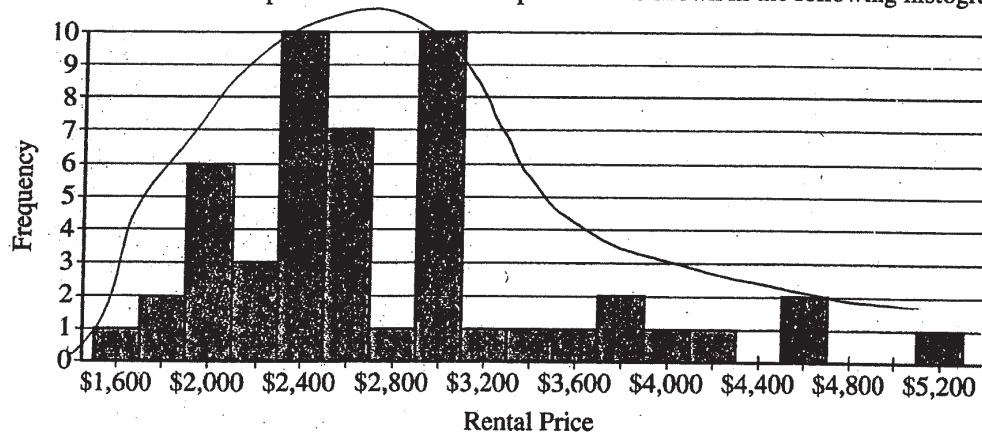
Percent of Section II score—25

Directions: Show all your work. Indicate clearly the methods you use, because you will be scored on the correctness of your methods as well as on the accuracy and completeness of your results and explanations.

6. Emma is moving to a large city and is investigating typical monthly rental prices of available one-bedroom apartments. She obtained a random sample of rental prices for 50 one-bedroom apartments taken from a Web site where people voluntarily list available apartments.
- (a) Describe the population for which it is appropriate for Emma to generalize the results from her sample.

The population of one-bedroom apartments on the website she went on.

The distribution of the 50 rental prices of the available apartments is shown in the following histogram.



Unauthorized copying or reuse of
any part of this page is illegal.

GO ON TO THE NEXT PAGE.

- (b) Emma wants to estimate the typical rental price of a one-bedroom apartment in the city. Based on the distribution shown, what is a disadvantage of using the mean rather than the median as an estimate of the typical rental price?

Since the distribution of prices is skewed to the right, that means the mean is greater than the median of the distribution. So the disadvantage would be getting a higher estimate of the typical rental price so Emma would be misinformed on what the typical price should be (thinking it's higher) and spending more than she needs to.

- (c) Instead of using the sample median as the point estimate for the population median, Emma wants to use an interval estimate. However, computing an interval estimate requires knowing the sampling distribution of the sample median for samples of size 50. Emma has one point, her sample median, in that sampling distribution. Using information about rental prices that are available on the Web site, describe how someone could develop a theoretical sampling distribution of the sample median for samples of size 50.

By using the website, someone could randomly pick \hat{n} one bedroom apartments from the website 50 and use those to develop a distribution of the sample median through calculations, and you will repeat by obtaining multiple samples (around 1000) of size 50 and calculating the median for each.

Because Emma does not have the resources to develop the theoretical sampling distribution, she estimates the sampling distribution of the sample median using a process called bootstrapping. In the bootstrapping process, a computer program performs the following steps.

- Take a random sample, with replacement, of size 50 from the original sample.
- Calculate and record the median of the sample.
- Repeat the process to obtain a total of 15,000 medians.

Emma ran the bootstrap process, and the following frequency table is the bootstrap distribution showing her results of generating 15,000 medians.

Median	Frequency	Median	Frequency	Median	Frequency
2,345	1	2,585	1	2,825	247
2,390	13	2,587.5	171	2,837.5	7
2,395	18	2,600	22	2,847.5	1
2,400	56	2,612.5	1,190	2,872.5	317
2,445	4	2,625	174	2,885	10
2,447.5	56	2,672.5	5	2,950	700
2,450	55	2,675	1,924	2,962.5	93
2,475	3	2,687.5	1,341	2,972.5	6
2,495	66	2,700	2,825	2,975	65
2,497.5	136	2,735	35	2,985	12
2,500	1,899	2,747.5	619	2,987.5	1
2,522.5	2	2,750	2	2,995	6
2,525	945	2,795	278	3,000	2
2,550	1,673	2,812.5	16	3,062.5	3

The bootstrap distribution provides an approximation of the sampling distribution of the sample median. A confidence interval for the median can be constructed using a percentage of the values in the middle of the bootstrap distribution.

(d) Use the frequency table to find the following.

(i) Value of the 5th percentile: $(.05)(15000) = 750$ medians

2500

(ii) Value of the 95th percentile: $(.95)(15000) = 14250$ medians

2950

6B4

6B4

- (e) Find the percentage of bootstrap medians in the table that are equal to or between the values found in part (d).

b/w 2500 and 2950

there are 14404 medians so

$$\frac{14404}{15000} = \boxed{96.027\%}$$

- (f) Use your values from parts (d) and (e) to construct and interpret a confidence interval for the median rental price.

~~2500, 2950~~ (2500, 2950)

96.027% of intervals constructed
have a median that lies b/w
2500 and 2950.

6C1

6C1

STATISTICS

SECTION II

Part B

Question 6

Spend about 25 minutes on this part of the exam.

Percent of Section II score—25

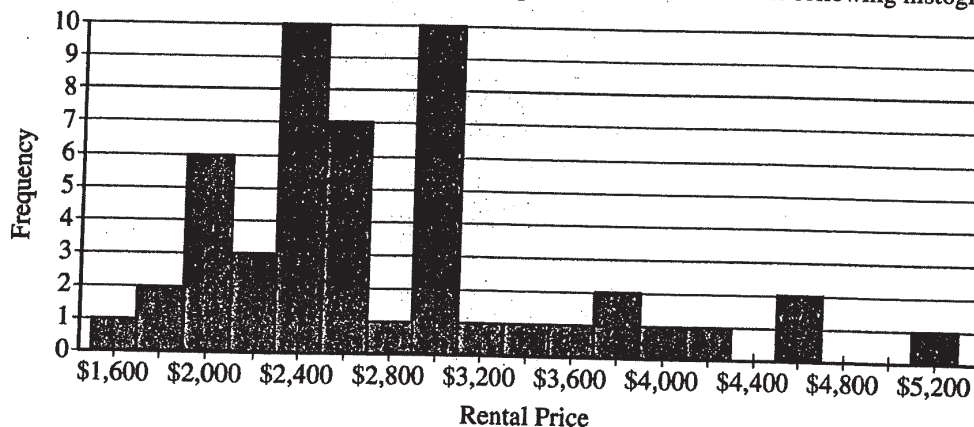
Directions: Show all your work. Indicate clearly the methods you use, because you will be scored on the correctness of your methods as well as on the accuracy and completeness of your results and explanations.

6. Emma is moving to a large city and is investigating typical monthly rental prices of available one-bedroom apartments. She obtained a random sample of rental prices for 50 one-bedroom apartments taken from a Web site where people voluntarily list available apartments.

- (a) Describe the population for which it is appropriate for Emma to generalize the results from her sample.

Emma can generalize her results to one-bedroom apartments listed on the website she used.

The distribution of the 50 rental prices of the available apartments is shown in the following histogram.



Unauthorized copying or reuse of
any part of this page is illegal.

GO ON TO THE NEXT PAGE.

- (b) Emma wants to estimate the typical rental price of a one-bedroom apartment in the city. Based on the distribution shown, what is a disadvantage of using the mean rather than the median as an estimate of the typical rental price?

Emma's data ~~cont~~ is positively skewed and contains outliers. Therefore, the mean will not be as representative as the median. This is because the median is not affected by outliers.

- (c) Instead of using the sample median as the point estimate for the population median, Emma wants to use an interval estimate. However, computing an interval estimate requires knowing the sampling distribution of the sample median for samples of size 50. Emma has one point, her sample median, in that sampling distribution. Using information about rental prices that are available on the Web site, describe how someone could develop a theoretical sampling distribution of the sample median for samples of size 50.

Emma can use a random number generator to get random samples of 50 from the website. Then, she calculates the median of ~~the~~ each sample. She can repeat this in order to calculate an interval estimate.

Because Emma does not have the resources to develop the theoretical sampling distribution, she estimates the sampling distribution of the sample median using a process called bootstrapping. In the bootstrapping process, a computer program performs the following steps.

- Take a random sample, with replacement, of size 50 from the original sample.
- Calculate and record the median of the sample.
- Repeat the process to obtain a total of 15,000 medians.

Emma ran the bootstrap process, and the following frequency table is the bootstrap distribution showing her results of generating 15,000 medians.

Bootstrap Distribution of Medians					
Median	Frequency	Median	Frequency	Median	Frequency
2,345	1	2,585	1	2,825	247
2,390	13	2,587.5	171	2,837.5	7
2,395	18	2,600	22	2,847.5	1
2,400	56	2,612.5	1,190	2,872.5	317
2,445	4	2,625	174	2,885	10
2,447.5	56	2,672.5	5	2,950	700
2,450	55	2,675	1,924	2,962.5	93
2,475	3	2,687.5	1,341	2,972.5	6
2,495	66	2,700	2,825	2,975	65
2,497.5	136	2,735	35	2,985	12
2,500	1,899	2,747.5	619	2,987.5	1
2,522.5	2	2,750	2	2,995	6
2,525	945	2,795	278	3,000	2
2,550	1,673	2,812.5	16	3,062.5	3

The bootstrap distribution provides an approximation of the sampling distribution of the sample median. A confidence interval for the median can be constructed using a percentage of the values in the middle of the bootstrap distribution.

(d) Use the frequency table to find the following.

(i) Value of the 5th percentile:

$$2626.303 \text{ invNorm}(0.05, 2584.16, 119.583) = 2387.4635$$

(ii) Value of the 95th percentile:

$$3019.696 \text{ invNorm}(0.95, 2584.16, 119.583) = 2780.8565$$

Unauthorized copying or reuse of
any part of this page is illegal.

GO ON TO THE NEXT PAGE.

6C4

6C4

- (e) Find the percentage of bootstrap medians in the table that are equal to or between the values found in part (d).

90% of bootstrap medians are equal to or between the values from part (d)

- (f) Use your values from parts (d) and (e) to construct and interpret a confidence interval for the median rental price.

μ = median rental price

assumptions:

- i) SRS of apartments
- ii) $n > 30$ (CLT)
- iii) $n < 10\%$ of population

(2582.6, 2585.8)

I am 90% confident that the median of the bootstrap distribution is between 2582.6 and 2585.8

AP[®] STATISTICS

2019 SCORING COMMENTARY

Question 6

Note: Student samples are quoted verbatim and may contain spelling and grammatical errors.

Overview

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, (3A) 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.

AP[®] STATISTICS
2019 SCORING COMMENTARY

Question 6 (continued)

Sample: 6A

Score: 4

In part (a) the response indicates the population of interest is all voluntarily listed web available one-bedroom apartments in this large city, which satisfies component 1 of section 1. In part (b) the response indicates that the mean would be abnormally high and not representative of the typical rental price, which satisfies component 2 of section 1. Also in part (b) the response indicates the distribution of rental price is skewed to the right, which satisfies component 3 of section 1. Because all three components are satisfied, section 1 was scored as essentially correct.

In part (c) the response indicates obtaining all possible samples of 50 one-bedroom apartments, which satisfies component 1 of section 2. Also, in part (c), the response indicates that the median will be found for each sample, which satisfies component 2 of section 2. Because both components are satisfied, section 2 was scored as essentially correct.

In part (d) the response correctly calculates the 5th and 95th percentile, which satisfies component 1 of section 3. In part (e) the response correctly calculates the percentage at or between the values that were reported in part (d), which satisfies component 2 of section 3. Because both components are satisfied, section 3 was scored as essentially correct.

In part (f) the response uses the values reported in part (d) for the confidence interval for median, which satisfies component 1 of section 4. The response uses the value reported in part (e) as the confidence level, which satisfies component 2 of section 4. The response correctly interprets the confidence interval for the median in context, which satisfies component 3 of section 4. Because all three components are satisfied, section 4 was scored as essentially correct.

Because four sections are scored as essentially correct, the response earned a score of 4.

AP[®] STATISTICS
2019 SCORING COMMENTARY

Question 6 (continued)

Sample: 6B

Score: 3

In part (a) the response indicates the population as one-bedroom apartments on the website, which satisfies component 1 of section 1. In part (b) the response indicates the mean would be a higher estimate of the typical rental price, which satisfies component 2 of section 1. The response indicates the distribution of prices is skewed to the right, which satisfies component 3 of section 1. Because all three components are satisfied, section 1 was scored as essentially correct.

In part (c) the response indicates multiple samples should be obtained, rather than all possible samples, which does not satisfy component 1 of section 2. The response does indicate that sample medians will be calculated, which satisfies component 2 of section 2. Because one of the two components is satisfied, section 2 was scored as partially correct.

In part (d) the response correctly calculates the 5th and 95th percentile, which satisfies component 1 of section 3. In part (e) the response correctly calculates the percentage of values at or between the values that are reported in part (d), which satisfies component 2 of section 3. Because both components are satisfied, section 3 was scored as essentially correct.

In part (f) the response used the values reported in part (d) for the confidence interval for median, which satisfies component 1 of section 4. The response uses the value reported in part (e) as the confidence level, which satisfies component 2 of section 4. The response incorrectly interprets the confidence interval for the median and does not include context, which does not satisfy component 3 of section 4. Because two of the three components are satisfied, section 4 was scored as partially correct.

Because two sections were scored as essentially correct and two sections were scored as partially correct, the response earned a score of 3.

AP[®] STATISTICS
2019 SCORING COMMENTARY

Question 6 (continued)

Sample: 6C

Score: 1

In part (a) the response indicates the population as one-bedroom apartments listed on the website, which satisfies component 1 of section 1. In part (b) the response indicates that the mean is not as representative as the median but does not clearly communicate what the mean is not representative of, or that the mean overestimates the typical value, which does not satisfy component 2 of section 1. The response indicates that the distribution is positively skewed, which satisfies component 3 of section 1. Because two of three components are satisfied, section 1 was scored as partially correct.

In part (c) the response does not indicate that all possible samples are obtained for a sampling distribution, which does not satisfy component 1 of section 2. The response does indicate the calculation of medians for samples, which does satisfy component 2 of section 2. Because one of the two components is satisfied, section 2 was scored as partially correct.

In part (d) incorrect values are given for the 5th and 95th percentiles, which does not satisfy component 1 of section 3. In part (e) an incorrect value for the percentage of bootstrap medians that are between and including the values in part (d) is given, which does not satisfy component 2 of section 3. Because none of the components are satisfied, section 3 was scored as incorrect.

In part (f) the values indicated as the endpoints of the confidence interval are not the values indicated in part (d) or the correct values, which does not satisfy component 1 of section 4. The confidence level given is the level indicated in part (e), which satisfies component 2 of section 4. The response provides an incorrect interpretation of the confidence interval for medians, which does not satisfy component 3 of section 4. Because one of the three components are satisfied, section 4 was scored as incorrect.

Because two sections were scored as partially correct and two sections were scored as incorrect, the response earned a score of 1.