
AP[®] Statistics

Sample Student Responses and Scoring Commentary

Inside:

Free Response Question 1

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

AP[®] STATISTICS

2019 SCORING GUIDELINES

Question 1

Intent of Question

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.

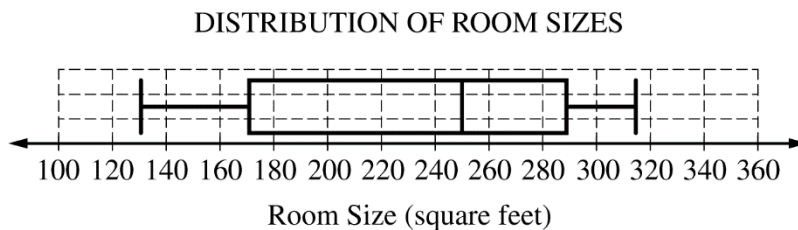
Solution

Part (a):

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. The center of the distribution is between 200 and 300 square feet. The range of the distribution is between 150 and 250 square feet. There are no apparent outliers.

Part (b):

The interquartile range is $IQR = 292 - 174 = 118$ square feet. There are no potential outliers because the minimum room size of 134 square feet does not fall below $Q_1 - 1.5(IQR) = -3$ square feet, and the maximum room size of 315 square feet does not exceed $Q_3 + 1.5(IQR) = 469$ square feet.



Part (c):

The histogram clearly shows the bimodal nature of the distribution of room sizes, but this is not apparent in the boxplot.

AP[®] STATISTICS

2019 SCORING GUIDELINES

Question 1 (continued)

Scoring

This question is scored in three sections. Section 1 consists of part (a); Section 2 consists of the outlier determination in part (b); Section 3 consists of the boxplot sketch in part (b) and part (c). Each section is scored as essentially correct (E), partially correct (P), or incorrect (I).

Section 1 is scored as follows:

Essentially correct (E) if the description of the distribution of room sizes satisfies the following four components:

1. The shape is bimodal OR there are two peaks OR there are two clusters.
2. The center is between 200 and 300 square feet.
3. The spread is addressed by stating the range is a value between 150 and 250 square feet OR the interquartile range is a value between 50 and 150 square feet OR all room sizes are between 100 and 350 square feet.
4. The response includes context.

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

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

Notes:

- Shape: Component 1 cannot be satisfied if a response describes the histogram as unimodal or describes the entire histogram as normal or approximately normal.
- Shape: A response that addresses symmetry, while appropriate, does not impact the scoring of section 1.
- Center: A response that states one cluster of the distribution is centered between 150 and 200 square feet and the other cluster is centered between 250 and 300 square feet satisfies both components 1 and 2.
- Center:
 - Responses that address center using interval language such as “the mean of the distribution is *between* 200 and 300” must, for any single measure of center, provide an interval with lower endpoint not below 200 square feet, and with upper endpoint not above 300 square feet to satisfy component 2.
 - Responses that address center using approximate language such as “the median of the distribution is *approximately* 225” must, for any single measure of center, specify a numeric value that is not less than 200 square feet, and that is not greater than 300 square feet to satisfy component 2.
 - Responses that use definitive language such as “the mean of the distribution *is* 231.4” must identify the corresponding numeric value correctly to satisfy component 2. Specifically, the median of the distribution can be correctly identified as any value between 250 and 253.5 square feet, inclusive; the mean of the distribution is 231.4 square feet; and the center (or average) of the distribution can be any value that is a correct median or mean.

AP[®] STATISTICS

2019 SCORING GUIDELINES

Question 1 (continued)

- Spread: A response recognizing all values in the sample fall between 100 and 350 square feet (or between 134 and 315 square feet) satisfies component 3 *only* for these exact endpoints and need not appeal to a specific measure of spread such as range or interquartile range (IQR).
- Spread:
 - Responses that appeal to a specific measure of spread using interval language, such as “the IQR is *between* 50 and 150,” must provide bounds appropriate to the corresponding measure of spread. For range, the lower endpoint must not be below 150 square feet and the upper endpoint cannot exceed 250 square feet; for IQR, the lower endpoint must not be below 50 square feet, with upper endpoint not to exceed 150 square feet; for standard deviation, the lower endpoint must not be below 25 square feet, with upper endpoint not to exceed 100 square feet.
 - Responses that appeal to a specific measure of spread using approximate language, such as “the range is *approximately* 250,” must specify a numeric value within the bounds appropriate to that measure of spread. For range, the value must be between 150 and 250 square feet (inclusive); for IQR, the value must be between 50 and 150 square feet (inclusive); for standard deviation, the value must be between 25 and 100 square feet (inclusive). Responses that appeal to a specific measure of spread using definitive language, such as “the range of the distribution *is* 181,” must identify the corresponding numeric value correctly to satisfy component 3. Specifically, the range of the distribution is 181 square feet; the IQR of the distribution is 118 square feet; and the standard deviation of the distribution is 68.12 square feet.

Section 2 is scored as follows:

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

1. Computation of both upper and lower outlier boundary fences that also shows the fences formulas either in words, symbols $Q_1 - 1.5(\text{IQR})$ and $Q_3 + 1.5(\text{IQR})$, or with values substituted from the table $174 - 1.5(118)$ and $292 + 1.5(118)$, or $(174 - 177)$ and $(292 + 177)$.
2. A correct decision regarding the presence of outliers.
3. Correct justification that compares the data with the fences.

Partially correct (P) if the response satisfies only two of the three components OR if the response omits exactly one of the fences but otherwise satisfies all three components.

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

Notes:

- A response that identifies both fence formulas using symbols, but does not substitute values for all symbols, must also include the correct fence values of -3 and 469 to satisfy component 1.
- In place of an appeal to fences, a response may compute outlier bounds representing k standard deviations from the sample mean, where k is a number from 2 to 3 (inclusive), and must include formulas for both endpoints either in words, symbols $\bar{x} \pm k(\text{standard deviation})$, or with values substituted from the table. When $k = 2$ the outlier bounds are $(95.16, 367.64)$; when $k = 3$ the bounds are $(27.04, 435.76)$.

AP[®] STATISTICS
2019 SCORING GUIDELINES

Question 1 (continued)

- A response that identifies the standard deviation bounds using symbols, but that does not substitute values for all symbols, does not satisfy component 1 unless the correct numeric bounds are provided.
- Component 3 is satisfied if the response states the outlier decision criterion: any data values falling outside of the interval from -3 to 469 are potential outliers.

Section 3 is scored as follows:

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

1. A correct sketch of the boxplot.
2. A response for part (c) that indicates the bimodal shape of the room size distribution is apparent in the histogram but not in the boxplot.

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

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

Notes:

- The boxplot must be completely correct to satisfy component 1. Specifically:
 - The minimum is positioned between grid lines at 120 and 140 square feet.
 - Q_1 is positioned between grid lines at 160 and 180 square feet.
 - The median is positioned between grid lines at 240 and 260 square feet.
 - Q_3 is positioned between grid lines at 280 and 300 square feet.
 - The maximum is positioned between grid lines at 300 and 320 square feet.
- If a *mean* is included as a part of the boxplot, component 1 cannot be satisfied.
- A response based on skewness or symmetry does not satisfy component 2.
- A response stating the unimodal OR normal shape of the histogram of room sizes is apparent in the histogram but not in the boxplot will satisfy component 2 *only* if the shape description in section 1 component 1 was also unimodal OR normal, respectively.

AP[®] STATISTICS
2019 SCORING GUIDELINES

Question 1 (continued)

4 Complete Response

Three sections essentially correct

3 Substantial Response

Two sections essentially correct and one section partially correct

2 Developing Response

Two sections essentially correct and no sections partially correct

OR

One section essentially correct and one or two sections partially correct

OR

Three sections partially correct

1 Minimal Response

One section essentially correct

OR

No sections essentially correct and two sections partially correct

1A1

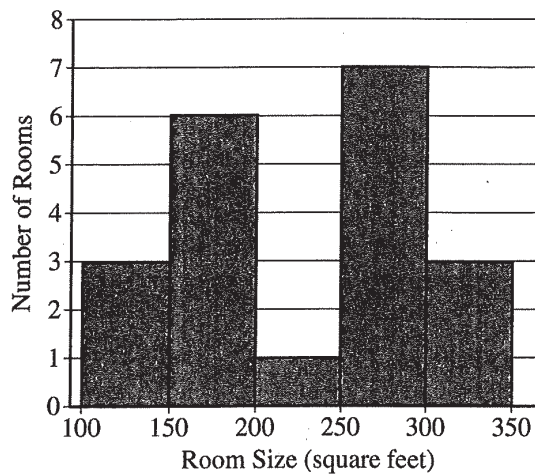
1A2

STATISTICS
SECTION II
Part A
Questions 1-5

Spend about 1 hour and 5 minutes on this part of the exam.
Percent of Section II score—75

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.

1. The sizes, in square feet, of the 20 rooms in a student residence hall at a certain university are summarized in the following histogram.



- (a) Based on the histogram, write a few sentences describing the distribution of room size in the residence hall.

The distribution of room size in the residence hall is bimodal and roughly symmetric. The spread of the distribution can be represented by its range of approximately 250. This range is not exact since the minimum room size in square feet is between 100 and 150 and the maximum room size is between 300 and 350. The center of the distribution can be shown by its median of approximately 250 square feet.

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

GO ON TO THE NEXT PAGE.

1A2

1A2

(b) Summary statistics for the sizes are given in the following table.

| Mean | Standard Deviation | Min | Q1 | Median | Q3 | Max |
|-------|--------------------|-----|-----|--------|-----|-----|
| 231.4 | 68.12 | 134 | 174 | 253.5 | 292 | 315 |

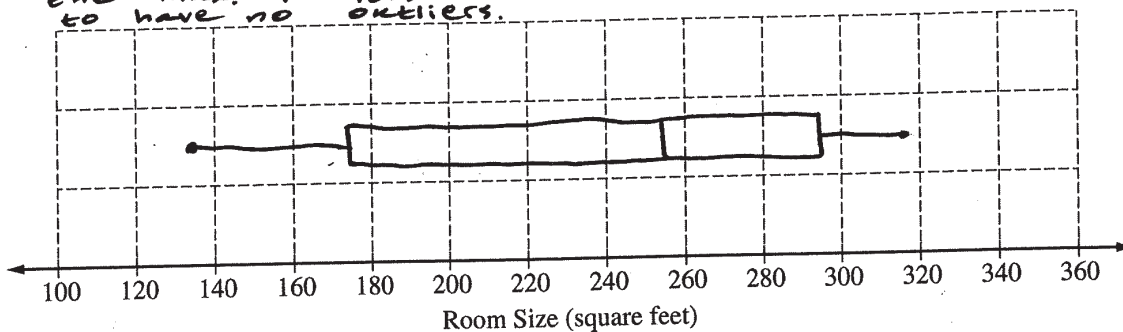
Determine whether there are potential outliers in the data. Then use the following grid to sketch a boxplot of room size.

$$IQR = 292 - 174 = 118.$$

$$\text{upper bound} = Q_3 + 1.5IQR = 292 + 1.5(118) = 469.$$

$$\text{lower bound} = Q_1 - 1.5IQR = 174 - 1.5(118) = -3.$$

Since the min. is greater than the lower bound and the max. is less than the upper bound, the data seem to have no outliers.



(c) What characteristic of the shape of the distribution of room size is apparent from the histogram but not from the boxplot?

The histogram shows us that the distribution of room size is bimodal while the boxplot does not.

1B1

1B1

STATISTICS

SECTION II

Part A

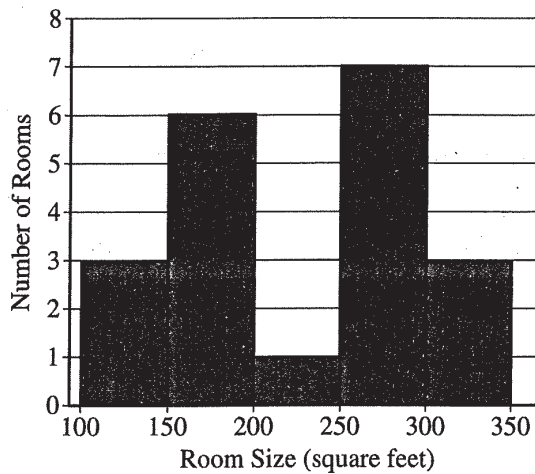
Questions 1-5

Spend about 1 hour and 5 minutes on this part of the exam.

Percent of Section II score—75

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.

- The sizes, in square feet, of the 20 rooms in a student residence hall at a certain university are summarized in the following histogram.



- Based on the histogram, write a few sentences describing the distribution of room size in the residence hall.

The distribution room sizes in the residence hall seems to be centered around 200 to 300 sqft since the median room size is between 200 and 300 sqft. There do not seem to be any outliers in the distribution. The distribution seems to be bimodal and fairly symmetric. The IQR of the distribution is between 50 and 150.

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

GO ON TO THE NEXT PAGE.

1B2

1B2

(b) Summary statistics for the sizes are given in the following table.

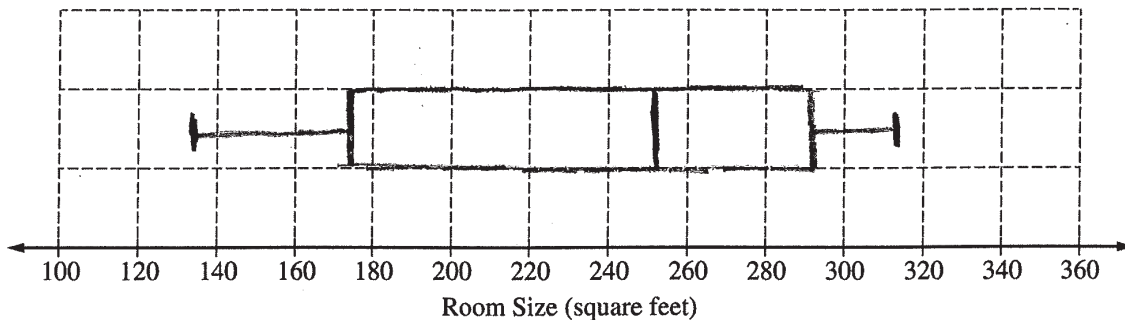
| Mean | Standard Deviation | Min | Q1 | Median | Q3 | Max |
|-------|--------------------|-----|-----|--------|-----|-----|
| 231.4 | 68.12 | 134 | 174 | 253.5 | 292 | 315 |

Determine whether there are potential outliers in the data. Then use the following grid to sketch a boxplot of room size.

$$UF = 292 + IQR(1.5) = 292 + 177 = 469$$

$$LF = 174 - 177 = -3$$

There are no potential outliers in the data since all data values fall within the upper and lower fences.



(c) What characteristic of the shape of the distribution of room size is apparent from the histogram but not from the boxplot?

The boxplot shows that the distribution of room sizes is skewed to the left. This is not shown in the histogram.

101

STATISTICS
SECTION II
Part A
Questions 1-5

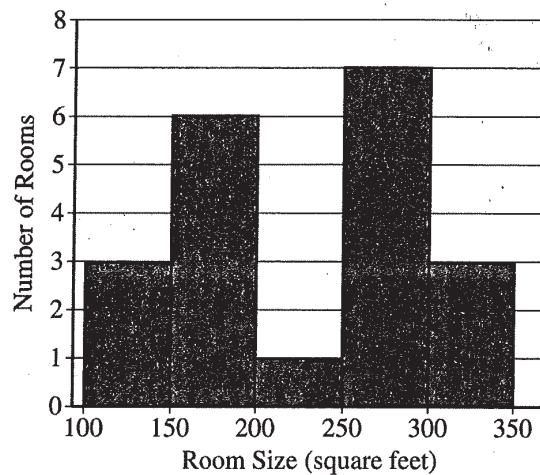
101

Spend about 1 hour and 5 minutes on this part of the exam.

Percent of Section II score—75

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.

1. The sizes, in square feet, of the 20 rooms in a student residence hall at a certain university are summarized in the following histogram.



500V

- (a) Based on the histogram, write a few sentences describing the distribution of room size in the residence hall.

The shape of the histogram is bimodal with two peaks. One peak is at 150-200 ft² and the other peak is at 250-300 ft². There seem to be no plausible outliers. The center is at 250 square feet which is the median. The range is 250 square feet and the variability isn't too large.

1C2

1C2

1.5x

(b) Summary statistics for the sizes are given in the following table.

| Mean | Standard Deviation | Min | Q1 | Median | Q3 | Max |
|-------|--------------------|-----|-----|--------|-----|-----|
| 231.4 | 68.12 | 134 | 174 | 253.5 | 292 | 315 |

Determine whether there are potential outliers in the data. Then use the following grid to sketch a boxplot of room size.

$$IQR = Q_3 - Q_1 = 292 - 174 = 118$$

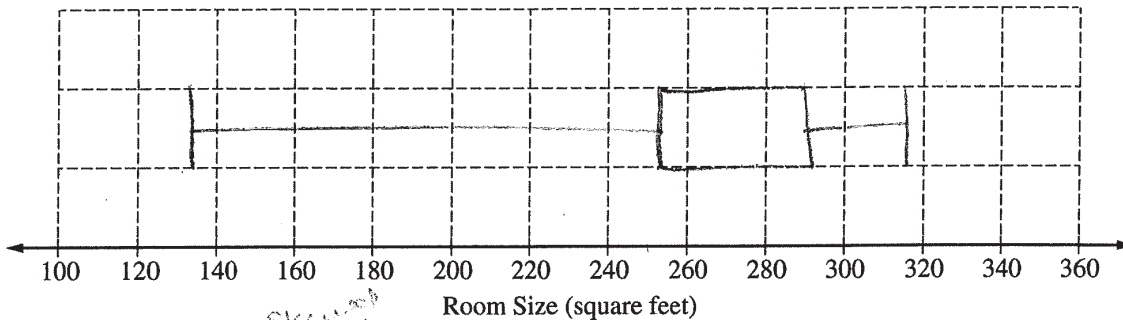
$$1.5(IQR) = 177$$

$$Q_3 + 1.5IQR = 292 + 177 = 469$$

$$Q_1 - 1.5IQR = 174 - 177 = -3$$

$$134 > -3 \checkmark$$

There are no outliers in the data as all the values fall within 1.5 x interquartile range below Q_1 and above Q_3 .



(c) What characteristic of the shape of the distribution of room size is apparent from the histogram but not from the boxplot?

The boxplot shows that the distribution of room size is skewed left which is not apparent from the histogram.

AP[®] STATISTICS

2019 SCORING COMMENTARY

Question 1

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) 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.

Sample: 1A

Score: 4

In part (a) the response recognizes that there is a bimodal nature of the histogram through the statement “[t]he distribution of room size in the residence hall is bimodal and roughly symmetric,” which satisfies component 1 of section 1. Addressing symmetry for this distribution is appropriate but not required for this question. The response recognizes a valid approximate value for the median by stating that the histogram has a “median of approximately 250 square feet,” which satisfies component 2 of section 1. The response recognizes a valid approximate value for the range by stating the histogram has a “range of approximately 250.” which satisfies component 3 of section 1. The response further emphasizes the uncertainty in the range estimate by providing an interval that contains the minimum data value and an interval that contains the maximum data value. Such emphasis is not required to satisfy component 3. Both room size and square feet are included in the response; either one of these would provide context; therefore, component 4 is satisfied. This response satisfies all four components because it correctly addresses the shape, center, and spread of the distribution in context. Therefore, section 1 was scored as essentially correct.

In part (b) the response includes the computation of the lower fence, $Q_1 - 1.5(IQR) = -3$, and the computation of the upper fence, $Q_3 + 1.5(IQR) = 469$, which satisfies component 1 of section 2. The response communicates the decision that “the data seem to have no outliers,” which satisfies component 2 of section 2. The response justifies the decision of no outliers by recognizing that all data fall between the lower and upper fence through the statement “[s]ince the min. is greater than the lower bound and the max. is less than the upper bound,” which satisfies component 3 of section 2. Because the computation, decision, and justification components are all satisfied, section 2 was scored as essentially correct.

In part (b) the boxplot sketch is completely correct, which satisfies component 1 of section 3. In part (c) the bimodal shape of the histogram is recognized as the shape characteristic that is not apparent in the boxplot, which satisfies component 2 of section 3. Because the response satisfies components 1 and 2, section 3 was scored as essentially correct.

Because three sections were scored as essentially correct, the response earned a score of 4.

AP[®] STATISTICS
2019 SCORING COMMENTARY

Question 1 (continued)

Sample: 1B

Score: 3

In part (a) the response recognizes the bimodal nature of the histogram through the statement “[t]he distribution seems to be bimodal and fairly symmetric,” which satisfies component 1 of section 1. Addressing symmetry for this distribution is appropriate but not required for this question. The response recognizes a valid interval for the median of the histogram by stating that “the median room size is between 200 and 300 sqft,” which satisfies component 2 of section 1. The response recognizes a valid interval for the interquartile range by stating, “[t]he IQR of the distribution is between 50 and 150,” which satisfies component 3 of section 1. Both room size and square feet are included in the response; either one of these would provide context; therefore, component 4 is satisfied. This response satisfies all four components because it correctly addresses the shape, center, and spread of the distribution in context. Therefore section 1 was scored as essentially correct.

In part (b) the response includes the computation of the lower fence, $174 - 177 = -3$, and the computation of the upper fence, $292 + 177 = 469$, which satisfies component 1 of section 2. A correct decision with correct justification is provided by stating “[t]here are no potential outliers in the data since all data fall within the upper and lower fences.” Therefore, components 2 and 3 are also satisfied. Because the computation, decision, and justification components are all satisfied, section 2 was scored as essentially correct.

In part (b) the boxplot is drawn correctly, which satisfies component 1 of section 3. Because the response recognizes a shape characteristic of the boxplot (left skewness) that is not apparent in the histogram, component 2 of section 3 is not satisfied. Symmetry or skewness can be argued from either the histogram or boxplot. However, the bimodal shape of the histogram is not at all apparent from the boxplot. Because the response satisfies only one of the two components, section 3 was scored as partially correct.

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

AP[®] STATISTICS
2019 SCORING COMMENTARY

Question 1 (continued)

Sample: 1C

Score: 2

In part (a) the response describes the centers of both modes as “[t]he shape of the histogram is bimodal with two peaks. One peak is at 150-200 ft² and the other peak is at 250-300 ft²,” which satisfies both components 1 and 2 of section 1. The response goes on to address the center of the entire distribution by stating, “[t]he center is at 250 square feet which is the median.” Because this response provides a correct exact value for the median of the entire distribution, it also would satisfy component 2 of section 1. However, the response states that “the range is 250 square feet,” which is an incorrect exact value of the range, component 3 of section 1 is not satisfied. Because square feet are included in the response, component 4 of section 1 is satisfied. This response satisfies three of the four components because it correctly addresses the shape and center of the distribution in context, but does not correctly address the spread of the distribution; therefore, section 1 was scored as partially correct.

In part (b) the response included the computation of the lower fence, $174 - 177 = -3$, and the computation of the upper fence, $292 + 177 = 469$, which satisfies component 1 of section 2. A correct decision with correct justification is provided by stating “[t]here are no outliers in the data as all the values fall within $1.5 \times$ interquartile range below Q_1 and above Q_3 ”; therefore, components 2 and 3 are also satisfied. Because the computation, decision, and justification components are all satisfied, section 2 was scored as essentially correct.

In part (b) the boxplot is not sketched correctly. All five numbers of the five-number summary must be correctly positioned, and the sketch in this response does not correctly position the first quartile and omits the median; therefore, component 1 of section 3 is not satisfied. Because the response recognizes a shape characteristic of the boxplot (skewed left) that is not apparent in the histogram, component 2 of section 3 is not satisfied. Approximate symmetry or mild skewness can be argued from either the histogram or boxplot. However, the bimodal shape of the histogram is not at all apparent from the boxplot. Because the response does not satisfy either component, section 3 was scored as incorrect.

Because one section was scored as partially correct, one section was scored as essentially correct, and one section was scored as incorrect, the response earned a score of 2.