

# 2019 AP<sup>®</sup> STATISTICS FREE-RESPONSE QUESTIONS

## STATISTICS

### SECTION II

#### Part A

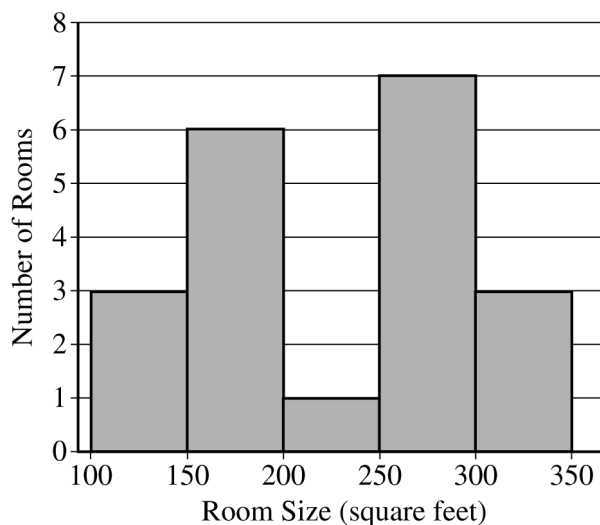
#### Questions 1-5

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.

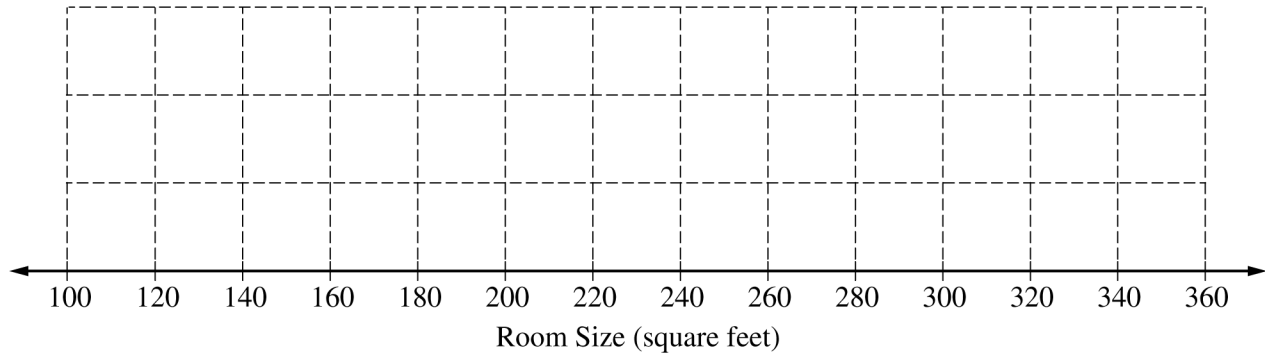


- (a) Based on the histogram, write a few sentences describing the distribution of room size in the residence hall.
- (b) Summary statistics for the sizes are given in the following table.

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

## 2019 AP<sup>®</sup> STATISTICS FREE-RESPONSE QUESTIONS

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



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

## 2019 AP<sup>®</sup> STATISTICS FREE-RESPONSE QUESTIONS

2. Researchers are investigating the effectiveness of using a fungus to control the spread of an insect that destroys trees. The researchers will create four different concentrations of fungus mixtures: 0 milliliters per liter (ml/L), 1.25 ml/L, 2.5 ml/L, and 3.75 ml/L. An equal number of the insects will be placed into 20 individual containers. The group of insects in each container will be sprayed with one of the four mixtures, and the researchers will record the number of insects that are still alive in each container one week after spraying.

- (a) Identify the treatments, experimental units, and response variable of the experiment.

Treatments:

Experimental units:

Response variable:

- (b) Does the experiment have a control group? Explain your answer.
- (c) Describe how the treatments can be randomly assigned to the experimental units so that each treatment has the same number of units.

**AP<sup>®</sup> 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.

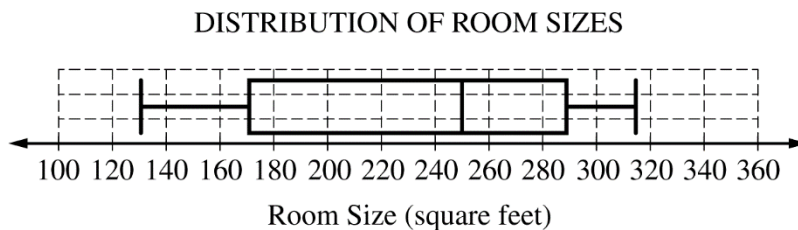
**Solution**

**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<sup>®</sup> 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<sup>®</sup> 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  $Q_1 - 1.5(IQR)$  and  $Q_3 + 1.5(IQR)$ , 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<sup>®</sup> 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.