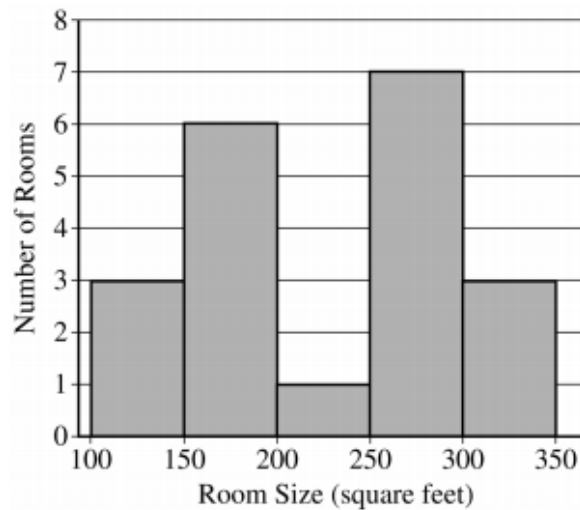


**Spend about 1 hour and 5 minutes on this part of the exam.**

**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.

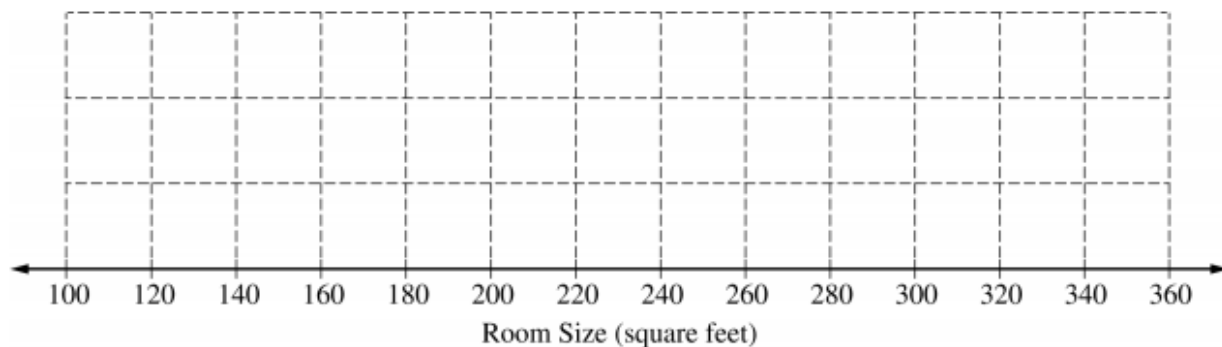


- (a) Based on the histogram, write a few sentences describing the distribution of room size in the residence hall.

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

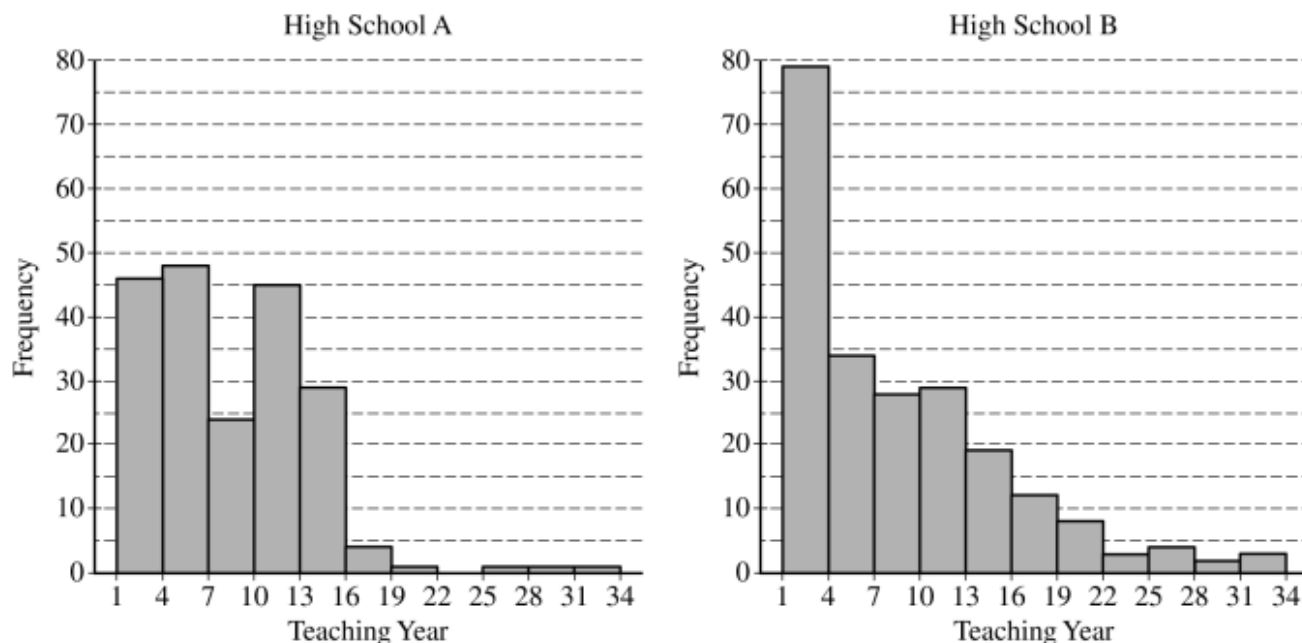
Mean	Standard Deviation	Min	Q1	Median	Q3	Max
231.4	68.12	134	174	253.5	292	315

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?

5. The following histograms summarize the teaching year for the teachers at two high schools, A and B.

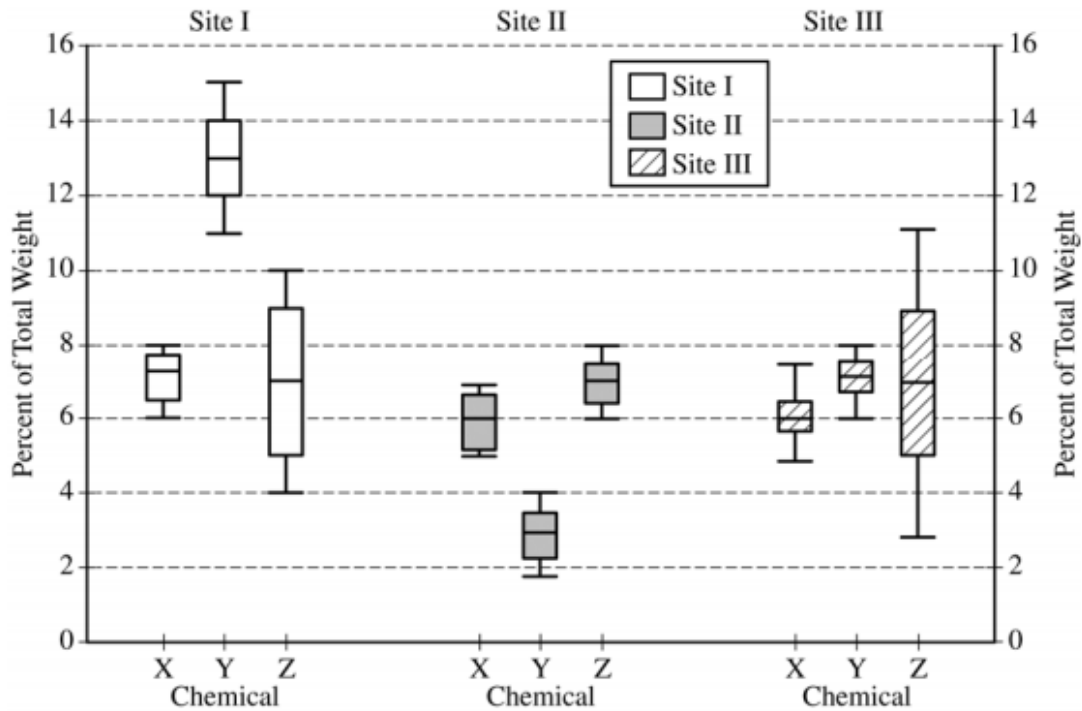


Teaching year is recorded as an integer, with first-year teachers recorded as 1, second-year teachers recorded as 2, and so on. Both sets of data have a mean teaching year of 8.2, with data recorded from 200 teachers at High School A and 221 teachers at High School B. On the histograms, each interval represents possible integer values from the left endpoint up to but not including the right endpoint.

- (a) The median teaching year for one high school is 6, and the median teaching year for the other high school is 7. Identify which high school has each median and justify your answer.

- (b) An additional 18 teachers were not included with the data recorded from the 200 teachers at High School A. The mean teaching year of the 18 teachers is 2.5. What is the mean teaching year for all 218 teachers at High School A?
- (c) The standard deviation of the teaching year for the 221 teachers at High School B is 7.2. If one teacher is selected at random from High School B, what is the probability that the teaching year for the selected teacher will be within 1 standard deviation of the mean of 8.2 ? Justify your answer.

4. The chemicals in clay used to make pottery can differ depending on the geographical region where the clay originated. Sometimes, archaeologists use a chemical analysis of clay to help identify where a piece of pottery originated. Such an analysis measures the amount of a chemical in the clay as a percent of the total weight of the piece of pottery. The boxplots below summarize analyses done for three chemicals—X, Y, and Z—on pieces of pottery that originated at one of three sites: I, II, or III.



- (a) For chemical Z, describe how the percents found in the pieces of pottery are similar and how they differ among the three sites.

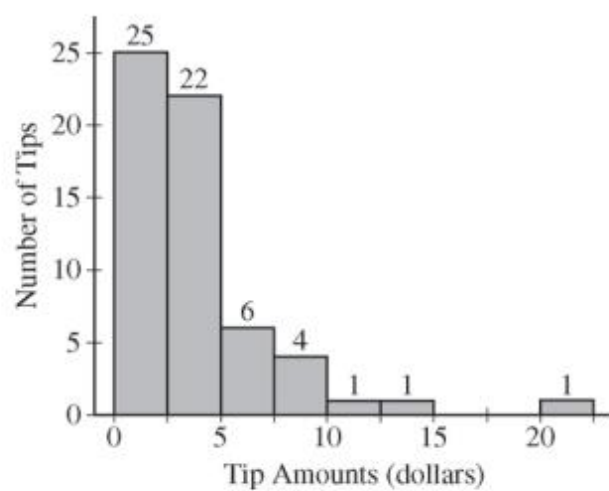
(b) Consider a piece of pottery known to have originated at one of the three sites, but the actual site is not known.

- (i) Suppose an analysis of the clay reveals that the sum of the percents of the three chemicals X, Y, and Z is 20.5%. Based on the boxplots, which site—I, II, or III—is the most likely site where the piece of pottery originated? Justify your choice.

- (ii) Suppose only one chemical could be analyzed in the piece of pottery. Which chemical—X, Y, or Z—would be the most useful in identifying the site where the piece of pottery originated? Justify your choice.

2016 Question #1

1. Robin works as a server in a small restaurant, where she can earn a tip (extra money) from each customer she serves. The histogram below shows the distribution of her 60 tip amounts for one day of work.



- (a) Write a few sentences to describe the distribution of tip amounts for the day shown.

- (b) One of the tip amounts was \$8. If the \$8 tip had been \$18, what effect would the increase have had on the following statistics? Justify your answers.

The mean:

The median:

**Spend about 65 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. Two large corporations, A and B, hire many new college graduates as accountants at entry-level positions. In 2009 the starting salary for an entry-level accountant position was \$36,000 a year at both corporations. At each corporation, data were collected from 30 employees who were hired in 2009 as entry-level accountants and were still employed at the corporation five years later. The yearly salaries of the 60 employees in 2014 are summarized in the boxplots below.



- (a) Write a few sentences comparing the distributions of the yearly salaries at the two corporations.

- (b) Suppose both corporations offered you a job for \$36,000 a year as an entry-level accountant.
- (i) Based on the boxplots, give one reason why you might choose to accept the job at corporation A.
- (ii) Based on the boxplots, give one reason why you might choose to accept the job at corporation B.

1. An administrator at a large university is interested in determining whether the residential status of a student is associated with level of participation in extracurricular activities. Residential status is categorized as on campus for students living in university housing and off campus otherwise. A simple random sample of 100 students in the university was taken, and each student was asked the following two questions.

- Are you an on campus student or an off campus student?
- In how many extracurricular activities do you participate?

The responses of the 100 students are summarized in the frequency table shown.

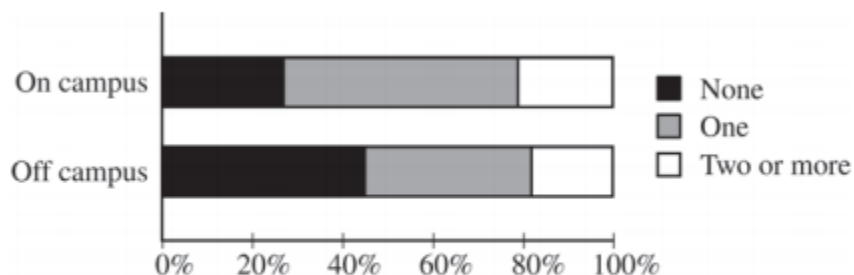
Level of Participation in Extracurricular Activities	Residential Status		Total
	On campus	Off campus	
No activities	9	30	39
One activity	17	25	42
Two or more activities	7	12	19
Total	33	67	100

- (a) Calculate the proportion of on campus students in the sample who participate in at least one extracurricular activity and the proportion of off campus students in the sample who participate in at least one extracurricular activity.

On campus proportion:

Off campus proportion:

The responses of the 100 students are summarized in the segmented bar graph shown.



- (b) Write a few sentences summarizing what the graph reveals about the association between residential status and level of participation in extracurricular activities among the 100 students in the sample.

- (c) After verifying that the conditions for inference were satisfied, the administrator performed a chi-square test of the following hypotheses.

$H_0$  : There is no association between residential status and level of participation in extracurricular activities among the students at the university.

$H_a$  : There is an association between residential status and level of participation in extracurricular activities among the students at the university.

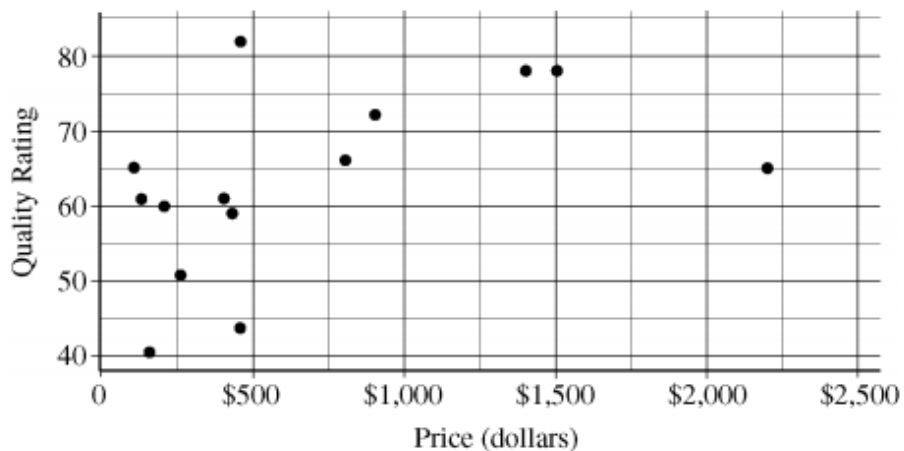
The test resulted in a  $p$ -value of 0.23. Based on the  $p$ -value, what conclusion should the administrator make?

---

**Part A****Questions 1-5****Spend about 65 minutes on this part of the exam.****Percent of Section II score—75**

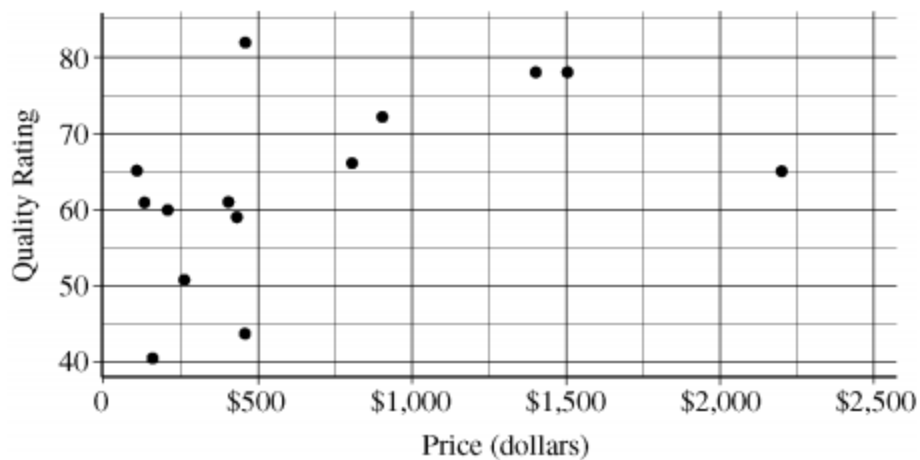
**Directions:** Show all your work. Indicate clearly the methods you use, because you will be scored on the correctness of your methods as well as on the accuracy and completeness of your results and explanations.

1. The scatterplot below displays the price in dollars and quality rating for 14 different sewing machines.



- (a) Describe the nature of the association between price and quality rating for the sewing machines.
- (b) One of the 14 sewing machines substantially affects the appropriateness of using a linear regression model to predict quality rating based on price. Report the approximate price and quality rating of that machine and explain your choice.

- (c) Chris is interested in buying one of the 14 sewing machines. He will consider buying only those machines for which there is no other machine that has both higher quality and lower price. On the scatterplot reproduced below, circle all data points corresponding to machines that Chris will consider buying.



**Part A****Questions 1-5****Spend about 65 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. A professional sports team evaluates potential players for a certain position based on two main characteristics, speed and strength.
  - (a) Speed is measured by the time required to run a distance of 40 yards, with smaller times indicating more desirable (faster) speeds. From previous speed data for all players in this position, the times to run 40 yards have a mean of 4.60 seconds and a standard deviation of 0.15 seconds, with a minimum time of 4.40 seconds, as shown in the table below.

	Mean	Standard Deviation	Minimum
Time to run 40 yards	4.60 seconds	0.15 seconds	4.40 seconds

Based on the relationship between the mean, standard deviation, and minimum time, is it reasonable to believe that the distribution of 40-yard running times is approximately normal? Explain.

- (b) Strength is measured by the amount of weight lifted, with more weight indicating more desirable (greater) strength. From previous strength data for all players in this position, the amount of weight lifted has a mean of 310 pounds and a standard deviation of 25 pounds, as shown in the table below.

	Mean	Standard Deviation
Amount of weight lifted	310 pounds	25 pounds

Calculate and interpret the  $z$ -score for a player in this position who can lift a weight of 370 pounds.

- (c) The characteristics of speed and strength are considered to be of equal importance to the team in selecting a player for the position. Based on the information about the means and standard deviations of the speed and strength data for all players and the measurements listed in the table below for Players A and B, which player should the team select if the team can only select one of the two players? Justify your answer.

	Player A	Player B
Time to run 40 yards	4.42 seconds	4.57 seconds
Amount of weight lifted	370 pounds	375 pounds