

Name Key

Date \_\_\_\_\_

Unit I Review

AP Statistics

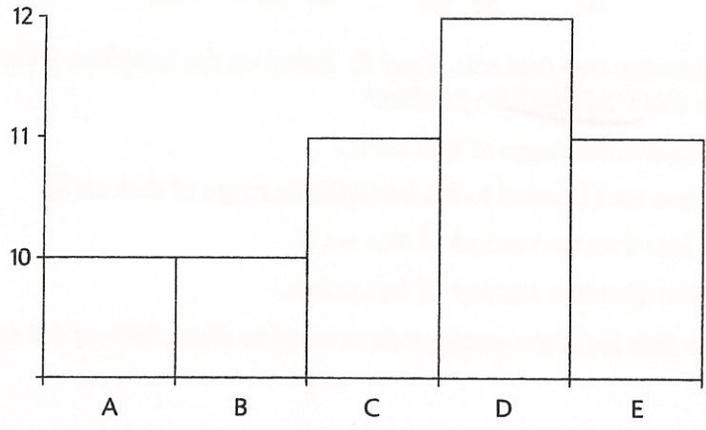
Part I

1) Which of the following is/are suitable to determine whether a distribution is skewed?

- I. Stemplot
- II. Histogram
- III. Cumulative frequency charts

- A. III only
- B. I and II
- C. I and III
- D. II and III
- E. I, II, and III**

2) Which statement is best represented by this histogram?



- A. The histogram is skewed right.
- B. Division D sold more than A and B combined.
- C. Division D made 1/3 of all sales.
- D. Division C sold more than division B.**
- E. Division C sold twice as much as division B.

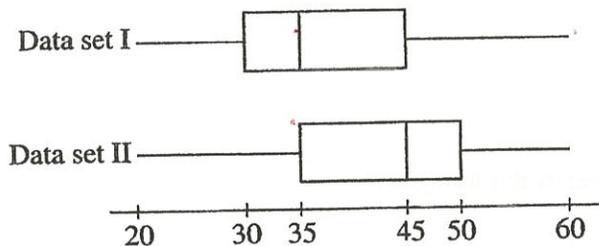
3) Which of the following statements is false?

- A. A symmetric distribution can have outliers.
- B. A skewed left distribution cannot have outliers on the right.**
- C. A skewed right distribution may have outliers on the right.
- D. A symmetric distribution may have more outliers than a skewed distribution.
- E. Stemplots are suitable for showing outliers.

4) Which of the following would be greatly affected by outliers?

- I. IQR
  - II. range
  - III. median
- A. I only
  - B. II only
  - C. I and II
  - D. I and III
  - E. II and III

5)



The boxplots shown above summarize two data sets, I and II. Based on the boxplots, which of the following statements about these two data sets CANNOT be justified?

- ~~(A)~~ The range of data set I is equal to the range of data set II.
- ~~(B)~~ The interquartile range of data set I is equal to the interquartile range of data set II.
- ~~(C)~~ The median of data set I is less than the median of data set II.
- (D)** Data set I and data set II have the same number of data points.
- ~~(E)~~ About 75% of the values in data set II are greater than or equal to about 50% of the values in data set I.

6) The following list is ordered from smallest to largest: 25, 26, 26, 30,  $y, y, y$ , 33, 150. Which of the following statements is (are) true?

*Right skewed*

- ~~I.~~ The mean is greater than the median
  - ~~II.~~ The mode is 26
  - ~~III.~~ There are no outliers in the data
- a. I only
  - b. I and II only
  - c. III only
  - d. I and III only
  - e. II and III only

*150*

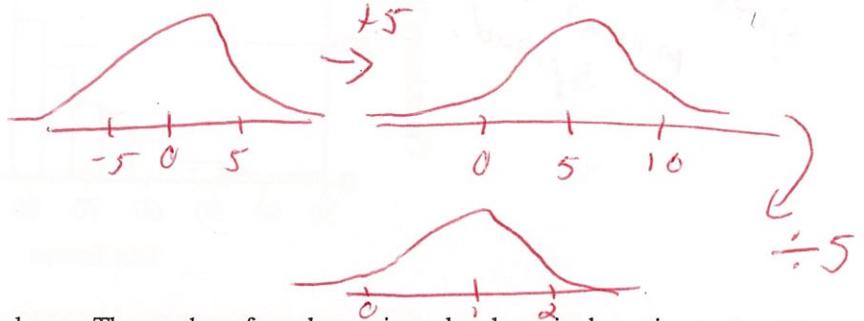
*↑  
all the same*

7) Which of the following is not true about histograms?

- A. The bars must touch each other.
- B. The width of each class must be the same.
- C. A histogram for a given set of data may be symmetric or skewed depending on the selection of class width and boundaries.
- D. Histograms may have gaps and clusters.
- E.** Histograms may represent categorical data.

8) The mean of a bell-shaped, symmetric distribution is 0, and the standard deviation of the distribution is 5. If 5 is added to each value in the distribution and then the resulting values are each divided by 5, what would be the new mean and standard deviation?

- A. mean = 0, standard deviation = 1
- B. mean = 1, standard deviation = 1
- C. mean = 0, standard deviation = 2
- D. mean = 1, standard deviation = 2
- E. mean = 5, standard deviation = 1



9) The salary schedule of a school district has six columns. The number of employees in each column is shown in the chart that follows. The salary increases as you move from column 1 to column 2 to column 3, and so on.

Column	1	2	3	4	5	6
# of Employees	32	58	110	230	650	90

Which of the following statements is accurate?

- A. The distribution is skewed right with a mean salary that is greater than the median salary.
- B. The distribution is skewed right with a median salary that is greater than the mean salary.
- C. The distribution is skewed left with a mean salary that is greater than the median salary.
- D. The distribution is skewed left with a median salary that is greater than the mean salary.
- E. Without knowing the salary for each column, the relationship between mean and median salaries cannot be determined.

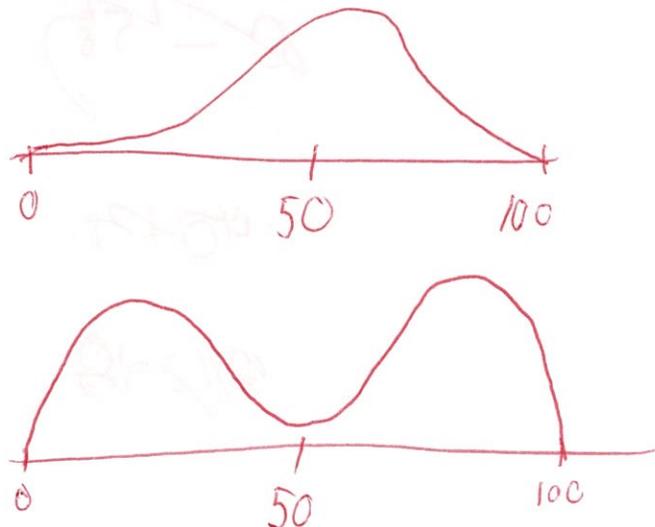
10) The following is the 5-number summary of a set of animal weights:  $a$ , 12,  $b$ , 16, and  $c$ . If a boxplot is drawn to show outliers, what is maximum length of either boxplot whisker?

- A. the larger of  $(12 - a)$  and  $(c - 16)$
- B. the larger of  $(b - a)$  and  $(c - b)$
- C. 6 *mod.*
- D. 8
- E. cannot be determined from the given data

11) If two distributions are symmetric and have the same range and the same mean, then which of the following statements are true?

- I. They have the same standard deviation.
- II. They have the same median.
- III. They have the same IQR.

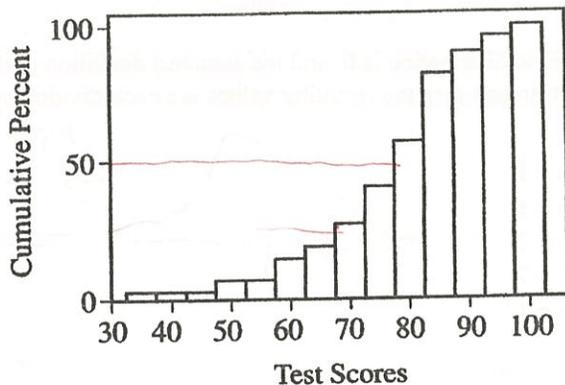
- A. I only
- B. II only
- C. III only
- D. II and III
- E. I, II, and III



12)

AP STATISTICS  
TEST SCORES

Greater Variability  
means wider  
spread.



Left Skew

The figure above shows a cumulative relative frequency histogram of 40 scores on a test given in an AP Statistics class. Which of the following conclusions can be made from the graph?

- (A) There is greater variability in the lower 20 test scores than in the higher 20 test scores. 30-75      75-100
- ~~(B) The median test score is less than 50.~~
- ~~(C) Sixty percent of the students had test scores above 80.~~ pass
- ~~(D) If the passing score is 70, most students did not pass the test.~~ 25%
- ~~(E) The horizontal nature of the graph for test scores of 60 and below indicates that those scores occurred most frequently.~~ most

13)

The five-number summary for a one-variable dataset is {5, 18, 20, 40, 75}. If you wanted to construct a modified boxplot for the dataset (that is, one that would show outliers if there are any), what would be the maximum possible length of the right side "whisker"?

- a. 35
- b. 33
- c. 5
- d. 55
- e. 53

$IQR \times 1.5$

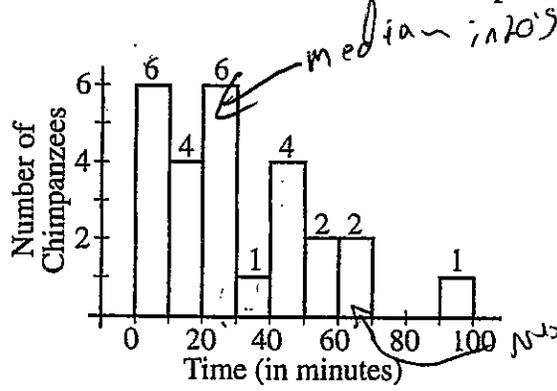
$24 \times 1.5 = 36$

~~$40 - 58$~~

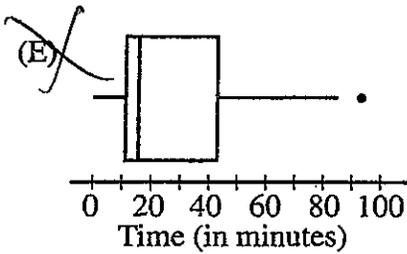
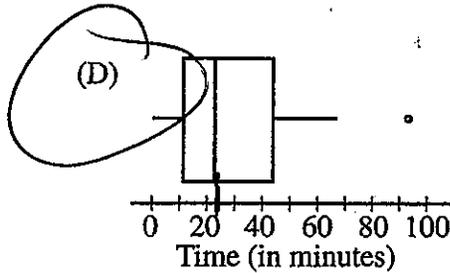
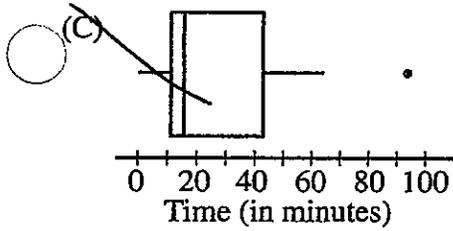
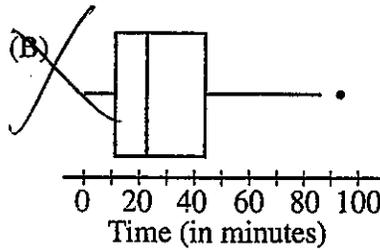
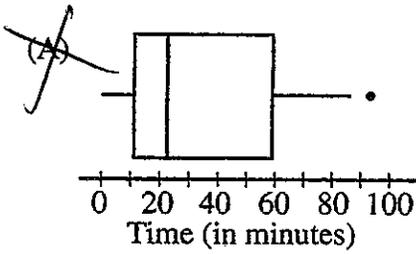
~~$40 + 75$~~

~~$10 + 18$~~

14) The histogram below displays the times, in minutes, needed for each chimpanzee in a sample of 26 to complete a simple navigational task.



It was determined that the largest observation, 93, is an outlier since  $Q_3 + 1.5(Q_3 - Q_1) = 87.125$ . Which of the following boxplots could represent the information in the histogram?



**AP<sup>®</sup> STATISTICS**  
**2011 SCORING GUIDELINES (Form B)**

**Question 1**

**Intent of Question**

The primary goals of this question were to assess students' ability to (1) describe and use a procedure for estimating medians from histograms; (2) use graphical displays to compare two different distributions; (3) use graphical and numerical information to compare the means for two groups.

**Solution**

**Part (a):**

The median is the value with half of the P-T ratios at or below it and half of the values at or above it.

For  $n$  observations in a group, use  $\frac{n+1}{2}$  to find the position of the median in the ordered list of observations.

For states west of the Mississippi ( $n = 24$ ) the median falls between the 12th and 13th value in the ordered list, and both the 12th and 13th values fall in the interval 15–16. For states east of the Mississippi ( $n = 26$ ) the median falls between the 13th and 14th value in the ordered list, and both of these values also fall in the interval 15–16.

From the histogram, cumulative frequencies for the two groups are shown in the table below.

Interval	West	East
12–13	1	2
13–14	$1+4=5$	$2+4=6$
14–15	$1+4+6=11$	$2+4+4=10$
15–16	$1+4+6+3=14$	$2+4+4+11=21$

Thus, the median P-T ratio for both groups is at least 15 students per teacher and at most 16 students per teacher.

**Part (b):**

The shapes of the two histograms are different. The histogram for states that are west of the Mississippi River is unimodal and skewed to the right, whereas the histogram for states that are east of the Mississippi River is unimodal and nearly symmetric.

As noted in part (a), the medians of the two distributions are about the same, between 15 and 16 for both distributions.

The histograms also show that there is more variability in the P-T ratios for states that are west of the Mississippi River. Although the greatest and least values for each group are not known, the range can be approximated. The range for the west is at most  $22 - 12 = 10$ , and the range for the east is at most  $19 - 12 = 7$ .

**AP<sup>®</sup> STATISTICS**  
**2011 SCORING GUIDELINES (Form B)**

**Question 1 (continued)**

**Part (c):**

The medians of the two distributions are about the same, as determined in part (a). The distribution of P-T ratios for states that are west of the Mississippi River is skewed to the right, indicating that the mean will probably be higher than the median. The rough symmetry for the east group indicates that the mean will be close to the median. Thus, the mean for the west group will probably be greater than the mean for the east group.

**Scoring**

Parts (a), (b), and (c) are scored as essentially correct (E), partially correct (P), or incorrect (I).

**Part (a)** is scored as follows:

Essentially correct (E) if a correct estimation method is described and appropriate estimates (values between 15 and 16, inclusive) are provided.

Partially correct (P) for any of the following:

- The response describes a correct estimation method, but the estimates are not provided.
- The response describes a method that conveys the idea of median as the middle value but is not entirely correct (for example, it describes the 12th value rather than the average of the 12th and 13th values), *AND* it provides reasonable estimates.
- The response gives an incomplete description of the method *AND* provides reasonable estimates.
- The response shows work only on the histograms *AND* correct estimates are provided, *BUT* no verbal explanation of the method is given.

Incorrect (I) if the response fails to meet the criteria for E or P.

**Part (b)** is scored as follows:

Essentially correct (E) if appropriate comparative statements are made for the centers, the shapes, and the spreads of the two groups.

*Note:* The shape of the east histogram can be described as skewed, *approximately* symmetric, or *approximately* normal. However, if the shape is described as symmetric or normal, part (b) cannot be scored an E.

Partially correct (P) if all three comparative statements are not made, but correct information regarding all three characteristics (center, shape, and spread) is provided for both groups, *OR* if only two of the three comparative statements are made.

*Note:* If a comparative statement about medians is made in part (a) or in part (c), this can count for a comparison of center in part (b).

Incorrect (I) if at most one comparative statement is made *AND* the response does not include correct information about all three characteristics (center, shape, and spread) for both groups.

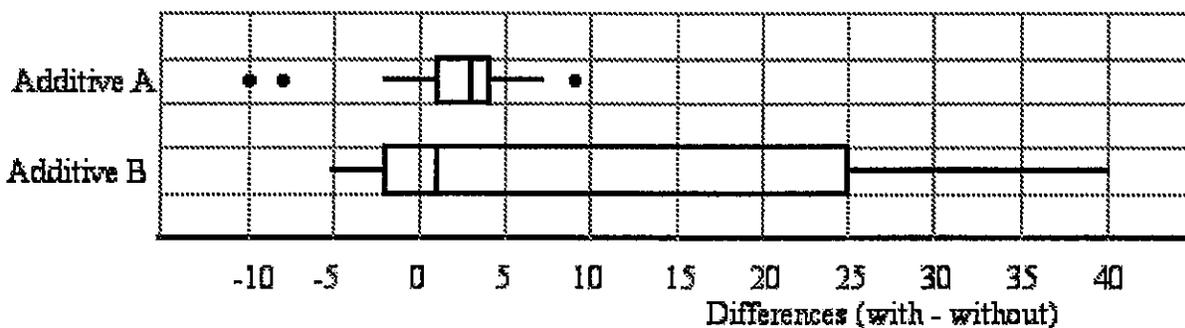
**AP<sup>®</sup> STATISTICS  
2004 SCORING GUIDELINES**

**Question 1**

**Solution**

**Part (a):**

	Additive A	Additive B
IQR	$4 - 1 = 3$	$25 - (-2) = 27$
$1.5 \times \text{IQR}$	4.5	40.5
$Q_1 - 1.5 \times \text{IQR}$	$1 - 4.5 = -3.5$	$-2 - 40.5 = -42.5$
$Q_3 + 1.5 \times \text{IQR}$	$4 + 4.5 = 8.5$	$25 + 40.5 = 65.5$
	3 outliers	no outliers



**Part (b i):**

Additive A is better at increasing the mileage in the greatest number of cars. The mileage increased for at least seventy-five percent of the cars with additive A, whereas the mileage decreased for more than twenty-five percent of the cars with additive B.

**Part (b ii):**

Additive B appears to produce a higher mean mileage gain than additive A. The boxplot for additive B clearly shows that the distribution of differences is skewed to the right, which will pull the average towards the larger values. The mean difference for additive B will be substantially greater than the median of 1. On the other hand, the distribution of differences for additive A has much less variability, as seen by comparing the lengths of the two boxes, and appears to be skewed to the left. The mean difference for additive A will be less than the median of 3.