

AP[®] STATISTICS

2009 SCORING GUIDELINES

Question 6

Intent of Question

The primary goals of this investigative task were to assess a student's ability to (1) define a parameter and state a correct pair of hypotheses; (2) explain how a particular statistic measures skewness; (3) use the observed value of the statistic and a simulated sampling distribution to make a conclusion about the shape of the population; and (4) create a new statistic and explain how it measures skewness.

Solution

Part (a):

The parameter of interest is $\mu =$ population mean miles per gallon (mpg) of a particular car model.

The null and alternative hypotheses are as follows:

$$H_0: \mu = 27$$

$$H_a: \mu < 27$$

Part (b):

If the distribution is right-skewed, one would expect the mean to be greater than the median.

Therefore the ratio $\frac{\text{sample mean}}{\text{sample median}}$ should be large (greater than 1).

Part (c):

Because we are testing for right-skewness, the estimated p -value will be the proportion of the simulated statistics that are greater than or equal to the observed value of 1.03. The dotplot shows that 14 of the 100 values are more than 1.03. Because this simulated p -value (0.14) is larger than any reasonable significance level, we do not have convincing evidence that the original population is skewed to the right and conclude that it is plausible that the original sample came from a normal population.

Part (d):

One possible statistic is $\frac{\text{maximum} - \text{median}}{\text{median} - \text{minimum}}$

If the distribution is right-skewed, one would expect the distance from the median to the maximum to be larger than the distance from the median to the minimum; thus the ratio should be greater than 1.

Scoring

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

AP[®] STATISTICS

2009 SCORING GUIDELINES

Question 6 (continued)

Part (a) is scored as follows:

Essentially correct (E) if the student correctly states the hypotheses with a lower-tailed alternative hypothesis *AND* correctly defines the parameter of interest by referring to the mean of the population in context (e.g., population mean mpg, true mean fuel efficiency, mean mpg of all cars of this model, mean mpg of this car model).

Partially correct (P) if only one component is correct (hypotheses or definition of parameter).

Incorrect (I) otherwise.

Notes:

- If a student attempts to define the parameter more than once (e.g., saying “the parameter is . . .” and then later saying $\mu = \dots$ ”), then these are treated as parallel solutions, and the worst attempt is scored.
- A symbol other than μ in the hypotheses must be explicitly and correctly defined to get credit for the parameter component.
- If words are used in the hypotheses to describe the parameter,
 - the words do not count as the definition of the parameter, and
 - the words must be consistent with the definition of the parameter, or no credit is given for the parameter component.

Part (b) is scored as follows:

Essentially correct (E) if the student states that large values (or values greater than 1) of the statistic indicate right-skewness *AND* justifies the answer with a correct statement of how right-skewness affects the relationship between the mean and median.

Partially correct (P) if the student:

- states that large values of the statistic indicate right-skewness *BUT* only argues that in a normal (or symmetric) distribution the ratio should be close to 1 (i.e., does not discuss the right-skewness)
OR
- makes a correct statement of how right-skewness affects the relationship between the mean and median *BUT* does not state that large values of the statistic indicate right-skewness
OR
- has reversed the relationship between mean and median in right-skewed distributions or reversed left- and right-skewness *AND* states that small values of the statistic indicate right-skewness.

Incorrect (I) if the student says large values indicate right-skewness but gives no explanation or an incorrect explanation.

Part (c) is scored as follows:

Essentially correct (E) if the student states that it is plausible that the sample came from a normal population *AND* justifies the choice with specific numerical evidence from the dotplot describing the relative location of the value 1.03 (e.g., 14 percent of the values are above 1.03).

AP[®] STATISTICS

2009 SCORING GUIDELINES

Question 6 (continued)

Partially correct (P) if the student:

- states that it is plausible that the sample came from a normal population *AND* justifies the choice by describing the relative location of the value 1.03 in the dotplot without specific numerical evidence (e.g., 1.03 is toward the middle of the distribution, 1.03 is within two standard deviations of the mean)
OR
- states that the sample came from a population that is skewed to the right *AND* justifies the choice by describing the relative location of the value 1.03 in the dotplot (e.g., *only* 14 percent of the values are above 1.03, 1.03 is in the tail of the distribution).

Incorrect (I) if the student does not refer to the relative location of 1.03 in the dotplot.

Note:

Common incorrect responses include the following:

- Simply describing the shape of the dotplot to justify normality.
- Saying that the dots are centered around 1, so the sample came from a normal population.
- Arguing that 1.03 is close to 1 (without describing its relative position in the dotplot).
- Stating that the sample size (or number of samples) is large, so the distribution is normal, or that the sample size is too small to make a conclusion.
- Stating an answer with no explanation.

Part (d) is scored as follows:

Essentially correct (E) if the student defines a reasonable statistic *AND* identifies the values that indicate right-skewness with a correct explanation of how the right-skewness affects the components of the statistic.

Partially correct (P) if the student:

- defines a reasonable statistic but fails to adequately justify the values that indicate right-skewness
OR
- does not define a reasonable statistic but uses values from the table (min, Q1, med, Q3, max) to describe a reasonable method for identifying right-skewness (e.g., if the distance from the median to maximum is greater than the distance from the minimum to the median, the distribution is skewed to the right).

Incorrect (I) if the method does not include a comparison (e.g., simply checking for outliers on one side only).

Notes:

- The statistic must be a formula that produces a single numerical value. For example, (maximum - median) > (median - minimum) is not a statistic.
- Any statistic that uses summary values not in the table (such as the mean) is incorrect.

AP[®] STATISTICS

2009 SCORING GUIDELINES

Question 6 (continued)

Some other acceptable variations of the statistic $\frac{\text{maximum} - \text{median}}{\text{median} - \text{minimum}}$ include the reciprocal (values < 1 indicate right-skewness), reversing the order of subtraction in the numerator or denominator (values < -1 indicate right-skewness), and the difference of the numerator and denominator (values > 0 indicate right-skewness).

Some other acceptable statistics are listed below. For each statistic, values > 1 indicate right-skewness. Other variations, including reciprocals and statistics based on differences are also acceptable.

1. $\frac{\text{maximum} - Q3}{Q1 - \text{minimum}}$
2. $\frac{Q3 - \text{median}}{\text{median} - Q1}$
3. $\frac{\left(\frac{Q1 + Q3}{2}\right)}{\text{median}}$
4. $\frac{\left(\frac{\text{minimum} + \text{maximum}}{2}\right)}{\text{median}}$

Each essentially correct (E) response counts as 1 point, and each partially correct (P) response 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 determine whether to score up or down, depending on the strength of the response and communication.

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 graded on the correctness of your methods as well as on the accuracy and completeness of your results and explanations.

6. A consumer organization was concerned that an automobile manufacturer was misleading customers by overstating the average fuel efficiency (measured in miles per gallon, or mpg) of a particular car model. The model was advertised to get 27 mpg. To investigate, researchers selected a random sample of 10 cars of that model. Each car was then randomly assigned a different driver. Each car was driven for 5,000 miles, and the total fuel consumption was used to compute mpg for that car.

- (a) Define the parameter of interest and state the null and alternative hypotheses the consumer organization is interested in testing.

$\mu =$ ~~average~~^{true} population mean mpg for this model

$$H_0: \mu = 27$$

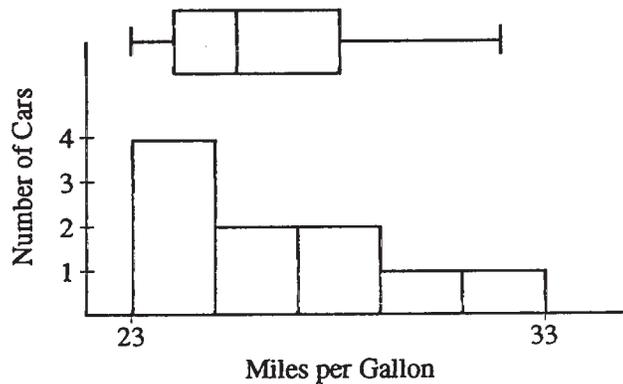
$$H_a: \mu < 27$$

GO ON TO THE NEXT PAGE.

A2

6A2

One condition for conducting a one-sample t -test in this situation is that the mpg measurements for the population of cars of this model should be normally distributed. However, the boxplot and histogram shown below indicate that the distribution of the 10 sample values is skewed to the right.



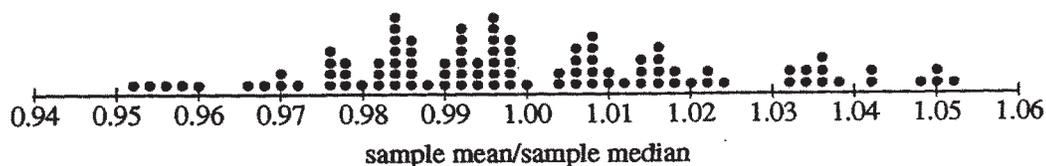
- (b) One possible statistic that measures skewness is the ratio $\frac{\text{sample mean}}{\text{sample median}}$. What values of that statistic (small, large, close to one) might indicate that the population distribution of mpg values is skewed to the right? Explain.

If the statistic is large, then the population distribution of mpg values is skewed to the right (assuming that the sample is a simple random sample). Since the mean is more sensitive/less resistant to extreme values than the median, it will be higher than the median when the skew is to the right, and thus the ratio would be large.

GO ON TO THE NEXT PAGE.

AS

- (c) Even though the mpg values in the sample were skewed to the right, it is still possible that the population distribution of mpg values is normally distributed and that the skewness was due to sampling variability. To investigate, 100 samples, each of size 10, were taken from a normal distribution with the same mean and standard deviation as the original sample. For each of those 100 samples, the statistic $\frac{\text{sample mean}}{\text{sample median}}$ was calculated. A dotplot of the 100 simulated statistics is shown below. 6A3



In the original sample, the value of the statistic $\frac{\text{sample mean}}{\text{sample median}}$ was 1.03. Based on the value of 1.03 and the dotplot above, is it plausible that the original sample of 10 cars came from a normal population, or do the simulated results suggest the original population is really skewed to the right? Explain.

The simulated results ~~suggest~~ ^{fail to} suggest that the ~~the~~ ^{original} sample came from a skewed population. There are 14 points out of 100 that have a statistic $\frac{\text{sample mean}}{\text{sample median}}$ greater than 1.03, which means that a value of 1.03 or greater could occur around 14% of the time by chance in a normally distributed population. Generally, a ~~value~~ ^{probability} of less than 10% is needed to reject a null hypothesis, so this simulation does not provide enough evidence to say that it isn't plausible that the original sample came from a normal population.

GO ON TO THE NEXT PAGE.

(d) The table below shows summary statistics for mpg measurements for the original sample of 10 cars.

Minimum	Q1	Median	Q3	Maximum
23	24	25.5	28	32

Choosing only from the summary statistics in the table, define a formula for a different statistic that measures skewness.

~~$(Q_3 - \text{Median}) / (\text{Median} - Q_1)$~~

$$\frac{Q_3 - \text{Median}}{\text{Median} - Q_1}$$

What values of that statistic might indicate that the distribution is skewed to the right? Explain.

Values ~~are~~ substantially greater than 1 would indicate a distribution skewed to the right, since it would indicate that the distance between Q_3 and the median is much greater than the distance between the median and Q_1 , and thus that the right tail is longer than the left, which indicates skewness.

GO ON TO THE NEXT PAGE.

B1

STATISTICS
SECTION II
Part B
Question 6

6B1

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 graded on the correctness of your methods as well as on the accuracy and completeness of your results and explanations.

6. A consumer organization was concerned that an automobile manufacturer was misleading customers by overstating the average fuel efficiency (measured in miles per gallon, or mpg) of a particular car model. The model was advertised to get 27 mpg. To investigate, researchers selected a random sample of 10 cars of that model. Each car was then randomly assigned a different driver. Each car was driven for 5,000 miles, and the total fuel consumption was used to compute mpg for that car.

- (a) Define the parameter of interest and state the null and alternative hypotheses the consumer organization is interested in testing.

μ is the mean fuel efficiency
in miles per gallon of all cars of
the particular car model.

$$H_0: \mu = 27$$

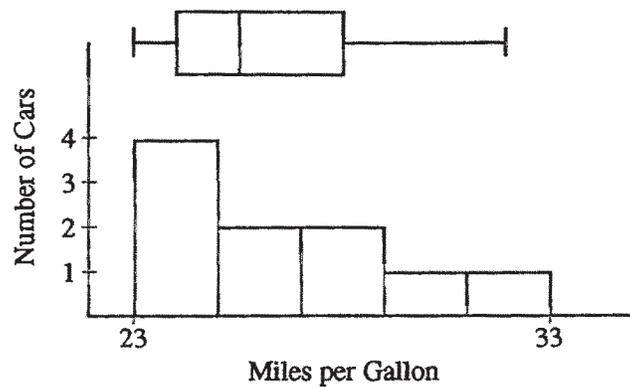
$$H_a: \mu < 27$$

GO ON TO THE NEXT PAGE.

B2

One condition for conducting a one-sample t -test in this situation is that the mpg measurements for the population of cars of this model should be normally distributed. However, the boxplot and histogram shown below indicate that the distribution of the 10 sample values is skewed to the right.

6B2



- (b) One possible statistic that measures skewness is the ratio $\frac{\text{sample mean}}{\text{sample median}}$. What values of that statistic (small, large, close to one) might indicate that the population distribution of mpg values is skewed to the right? Explain.

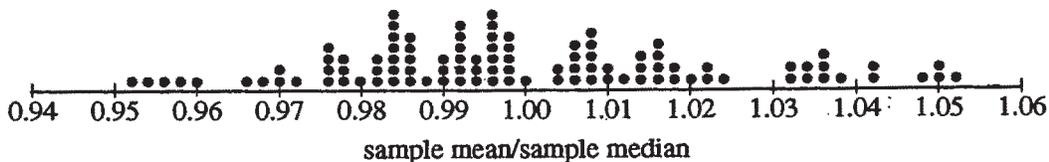
A large value would indicate skewness to the right because if there is skew to the right, then the sample mean would be pulled higher than the sample median, making the ratio larger than 1.

GO ON TO THE NEXT PAGE.

B3

6B3

- (c) Even though the mpg values in the sample were skewed to the right, it is still possible that the population distribution of mpg values is normally distributed and that the skewness was due to sampling variability. To investigate, 100 samples, each of size 10, were taken from a normal distribution with the same mean and standard deviation as the original sample. For each of those 100 samples, the statistic $\frac{\text{sample mean}}{\text{sample median}}$ was calculated. A dotplot of the 100 simulated statistics is shown below.



In the original sample, the value of the statistic $\frac{\text{sample mean}}{\text{sample median}}$ was 1.03. Based on the value of 1.03 and the dotplot above, is it plausible that the original sample of 10 cars came from a normal population, or do the simulated results suggest the original population is really skewed to the right? Explain.

It is plausible that the original sample came from a normal population because the distribution is centered at 1 which is no skew, and the distribution is mound shape and roughly symmetrical, meaning that it is equally likely to get skewness to the left or to the right.

GO ON TO THE NEXT PAGE.

(d) The table below shows summary statistics for mpg measurements for the original sample of 10 cars.

6 B4

B4

Minimum	Q1	Median	Q3	Maximum
23	24	25.5	28	32

Choosing only from the summary statistics in the table, define a formula for a different statistic that measures skewness.

$$\frac{Q_3 - \text{median}}{\text{median} - Q_1}$$

What values of that statistic might indicate that the distribution is skewed to the right? Explain.

Large values would indicate skewness to the right because if there is skew to the right, Q_3 and the median will be further apart than the median and Q_1 , making the value greater than 1.

GO ON TO THE NEXT PAGE.

C1

STATISTICS

SECTION II

~~Part B~~

Question 6

6C1

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 graded on the correctness of your methods as well as on the accuracy and completeness of your results and explanations.

6. A consumer organization was concerned that an automobile manufacturer was misleading customers by overstating the average fuel efficiency (measured in miles per gallon, or mpg) of a particular car model. The model was advertised to get 27 mpg. To investigate, researchers selected a random sample of 10 cars of that model. Each car was then randomly assigned a different driver. Each car was driven for 5,000 miles, and the total fuel consumption was used to compute mpg for that car.

- (a) Define the parameter of interest and state the null and alternative hypotheses the consumer organization is interested in testing.

parameter of interest $\rightarrow \mu = \text{true } \overset{\text{mean}}{\text{fuel efficiency}}$ of the car

$$H_0: \mu = 27 \text{ mpg}$$

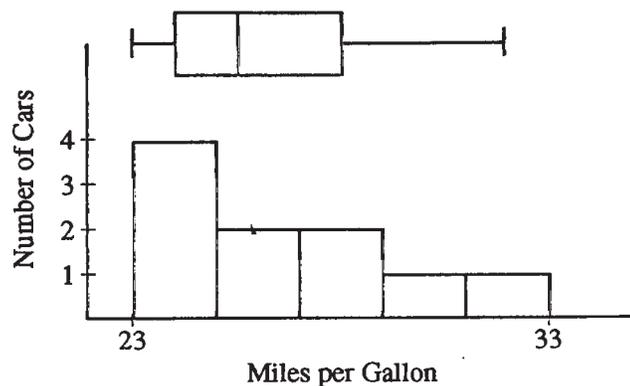
$$H_a: \mu \neq 27 \text{ mpg}$$

GO ON TO THE NEXT PAGE.

C2

6C2

One condition for conducting a one-sample t -test in this situation is that the mpg measurements for the population of cars of this model should be normally distributed. However, the boxplot and histogram shown below indicate that the distribution of the 10 sample values is skewed to the right.

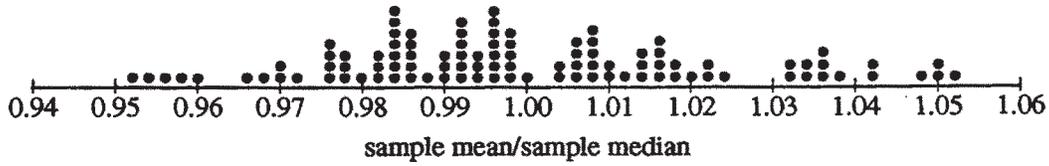


- (b) One possible statistic that measures skewness is the ratio $\frac{\text{sample mean}}{\text{sample median}}$. What values of that statistic (small, large, close to one) might indicate that the population distribution of mpg values is skewed to the right? Explain.

If a population is skewed to the right, the mean will be greater than the median, resulting in a large $\frac{\text{sample mean}}{\text{sample median}}$ ratio.

GO ON TO THE NEXT PAGE.

- 6C3 (c) Even though the mpg values in the sample were skewed to the right, it is still possible that the population distribution of mpg values is normally distributed and that the skewness was due to sampling variability. 6C3
 To investigate, 100 samples, each of size 10, were taken from a normal distribution with the same mean and standard deviation as the original sample. For each of those 100 samples, the statistic $\frac{\text{sample mean}}{\text{sample median}}$ was calculated. A dotplot of the 100 simulated statistics is shown below.



In the original sample, the value of the statistic $\frac{\text{sample mean}}{\text{sample median}}$ was 1.03. Based on the value of 1.03 and the dotplot above, is it plausible that the original sample of 10 cars came from a normal population, or do the simulated results suggest the original population is really skewed to the right? Explain.

The data suggest the original population could have been normal. If normal, the $\frac{\text{sample mean}}{\text{sample median}}$ ratio should be 1 and the dot plot above appears to be centered around 1 with a fairly normal shape.

GO ON TO THE NEXT PAGE.

C4

(d) The table below shows summary statistics for mpg measurements for the original sample of 10 cars. 6C4

Minimum	Q1	Median	Q3	Maximum
23	24	25.5	28	32

Choosing only from the summary statistics in the table, define a formula for a different statistic that measures skewness.

if,

$$|\text{median} - \text{minimum}| \neq |\text{median} - \text{maximum}|$$

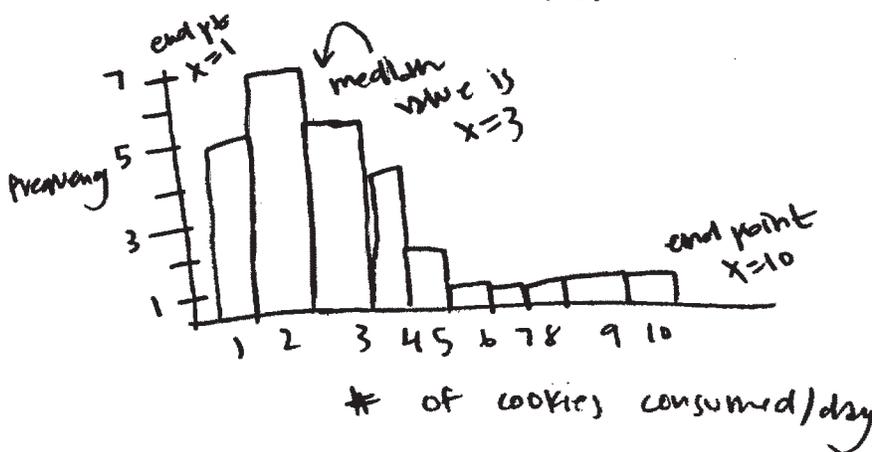
skewness is present

What values of that statistic might indicate that the distribution is skewed to the right? Explain.

if

$$|\text{median} - \text{minimum}| < |\text{median} - \text{maximum}|$$

then the distribution is likely skewed to the right because median has moved closer to the left side of the distribution where more points are concentrated



$$|3-1| < |3-10|$$

$$2 < 7 \quad \checkmark$$

GO ON TO THE NEXT PAGE.

AP[®] STATISTICS

2009 SCORING COMMENTARY

Question 6

Overview

The primary goals of this investigative task were to assess a student's ability to (1) define a parameter and state a correct pair of hypotheses; (2) explain how a particular statistic measures skewness; (3) use the observed value of the statistic and a simulated sampling distribution to make a conclusion about the shape of the population; and (4) create a new statistic and explain how it measures skewness.

Sample: 6A

Score: 4

In part (a) the student correctly defines the parameter of interest, including the concepts of mean and population in context, and correctly states the null and alternative hypotheses using standard notation. Part (a) was scored as essentially correct. In part (b) the student correctly states that large values of the statistic indicate that the distribution is skewed to the right and gives a justification that describes the relationship between the mean and median when the distribution is skewed right. Part (b) was scored as essentially correct. In part (c) the student correctly states that there is not convincing evidence that the original sample came from a skewed population and gives specific numerical evidence from the dotplot ("14 points out of 100"). Although not necessary, the student goes on to correctly explain when there would be convincing evidence that the distribution was skewed right. Part (c) was scored as essentially correct. In part (d) the student provides a reasonable statistic to measure skewness. The student then correctly identifies the values of the statistic that indicate right-skewness and justifies the response by discussing how the components of the statistic are affected by right-skewness. Part (d) was scored as essentially correct. The entire answer, based on all four parts, was judged a complete response and earned a score of 4 points.

Sample: 6B

Score: 3

In part (a) the student correctly defines the parameter of interest, including the concepts of mean and population in context, and correctly states the null and alternative hypotheses using standard notation. Part (a) was scored as essentially correct. In part (b) the student correctly states that large values of the statistic indicate that the distribution is skewed to the right and gives a justification that describes the relationship between the mean and median when the distribution is skewed right. Part (b) was scored as essentially correct. In part (c) the student correctly says, "It is plausible that the original sample came from a normal population." However, the justification provided is not based on the relative location of the value 1.03 in the dotplot. Instead, the student simply describes the shape and center of the dotplot and concludes that it is equally likely to get skewness to the left or to the right when sampling from a normal population. This is a true statement, but it does not address the question about the shape of the *original* population. Part (c) was scored as incorrect. In part (d) the student provides a reasonable statistic to measure skewness. The student then correctly identifies the values of the statistic that indicate right-skewness and justifies the response by discussing how the components of the statistic are affected by right-skewness. Part (d) was scored as essentially correct. With three parts essentially correct and one part incorrect, the entire answer was judged a substantial response and earned a score of 3 points.

AP[®] STATISTICS
2009 SCORING COMMENTARY

Question 6 (continued)

Sample: 6C

Score: 2

In part (a) the student correctly defines the parameter of interest but incorrectly uses a two-sided alternative hypothesis. Part (a) was scored as partially correct. In part (b) the student correctly states the relationship between the mean and median when the distribution is skewed to the right and then concludes that large values of the statistic indicate that the distribution is skewed to the right. Part (b) was scored as essentially correct. In part (c) the student does not use the relative location of 1.03 to make a decision about the shape of the original population. The student simply states that this distribution is centered at 1, which is expected when sampling from a normal distribution. Although this is a true statement, it does not give any information about the shape of the *original* population. Part (c) was scored as incorrect. In part (d), the student does not provide a formula for a single statistic that measures skewness. Instead, the student provides an inequality. However, the student then correctly explains how the inequality could be used to evaluate (right) skewness. Part (d) was scored as partially correct. With one part essentially correct, two parts partially correct, and one part incorrect, the entire answer was judged a developing response and earned a score of 2 points.