



## AP<sup>®</sup> Statistics 2009 Scoring Guidelines

### The College Board

The College Board is a not-for-profit membership association whose mission is to connect students to college success and opportunity. Founded in 1900, the association is composed of more than 5,600 schools, colleges, universities and other educational organizations. Each year, the College Board serves seven million students and their parents, 23,000 high schools and 3,800 colleges through major programs and services in college readiness, college admissions, guidance, assessment, financial aid, enrollment, and teaching and learning. Among its best-known programs are the SAT<sup>®</sup>, the PSAT/NMSQT<sup>®</sup> and the Advanced Placement Program<sup>®</sup> (AP<sup>®</sup>). The College Board is committed to the principles of excellence and equity, and that commitment is embodied in all of its programs, services, activities and concerns.

© 2009 The College Board. College Board, Advanced Placement Program, AP, AP Central, SAT, and the acorn logo are registered trademarks of the College Board. PSAT/NMSQT is a registered trademark of the College Board and National Merit Scholarship Corporation.

Permission to use copyrighted College Board materials may be requested online at:  
[www.collegeboard.com/inquiry/cbpermit.html](http://www.collegeboard.com/inquiry/cbpermit.html).

Visit the College Board on the Web: [www.collegeboard.com](http://www.collegeboard.com).  
AP Central<sup>®</sup> is the official online home for AP teachers: [apcentral.collegeboard.com](http://apcentral.collegeboard.com).

# AP<sup>®</sup> STATISTICS 2009 SCORING GUIDELINES

## Question 1

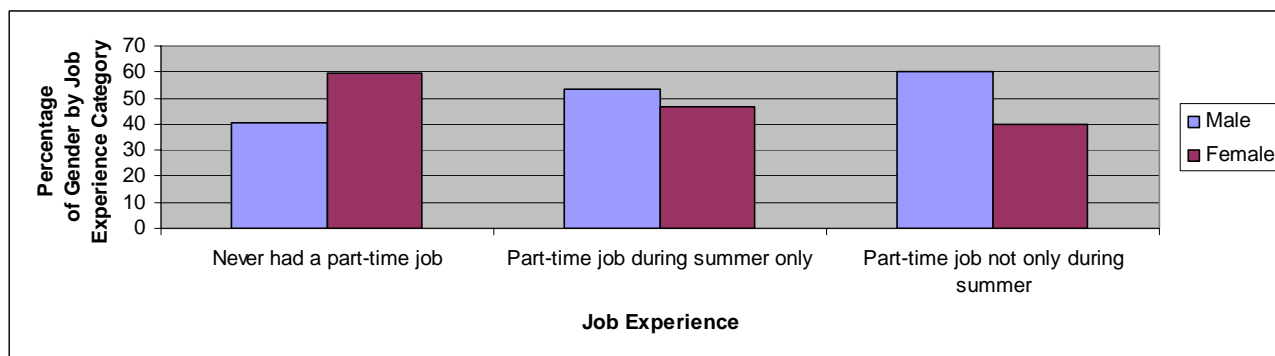
### Intent of Question

The primary goals of this question were to assess a student’s ability to (1) construct an appropriate graphical display for comparing the distributions of two categorical variables; (2) summarize from this graph the relationship of the two categorical variables; and (3) identify the appropriate statistical procedure to test if an association exists between two categorical variables and state appropriate hypotheses for the test.

### Solution

#### **Part (a):**

A side-by-side bar graph of the data is shown below.



#### **Part (b):**

The females in the sample were more likely to have never had a part-time job than were the males. Females were less likely than males to have had a part-time job during summer only or to have had a part-time job not only during summer. If there is no association between the two variables, one would expect the heights of the bars for gender within each job-experience category to be about the same. Based on the sample data summarized in the graphical display, there appears to be an association between job experience and gender.

#### **Part (c):**

The appropriate test of significance is the chi-square test of association or independence.

The appropriate hypotheses are:

$H_0$ : There is no association between gender and job experience.

$H_a$ : There is an association between gender and job experience.

OR

# AP<sup>®</sup> STATISTICS

## 2009 SCORING GUIDELINES

### Question 1 (continued)

$H_0$ : Gender and job experience are independent.

$H_a$ : Gender and job experience are not independent.

#### **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 an appropriate graph is drawn correctly and all labels are provided.

Partially correct (P) if the graphical display is drawn correctly but some labels are missing.

Incorrect (I) if a partial graph is drawn (such as for one job-experience category only) *OR* if the graphical display is incorrectly drawn or inappropriate.

*Note:* A segmented bar graph, a bar graph conditioning on gender, and pie charts are acceptable graphical displays.

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

Essentially correct (E) if the student correctly discusses percentages for both genders within each category of job experience.

Partially correct (P) if the student correctly discusses percentages for both genders within only one or two of the job-experience categories.

Incorrect (I) otherwise.

*Notes:*

- Stating that there is no obvious gender difference for “Part-time job during summer only” is acceptable.
- If an incomplete graph is drawn in part (a), the student may receive an (E) for part (b) if a discussion of the percentages for both genders in all three job-experience categories is provided.

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

Essentially correct (E) if the student recommends using the chi-square test of association or independence and states correct hypotheses in context.

Partially correct (P) if the student recommends using an incorrect test but states correct hypotheses in context for a chi-square test of association or independence *OR* if the student names the correct test but does not give correct hypotheses in context.

Incorrect (I) if the student recommends any incorrect statistical test and does not state correct hypotheses in context for a chi-square test of association or independence.

# AP<sup>®</sup> STATISTICS

## 2009 SCORING GUIDELINES

### Question 1 (continued)

*Notes:*

- If a student names merely the “chi-square test,” this will be accepted if the hypotheses for a chi-square test of association or independence are provided.
- If a student recommends the chi-square test of homogeneity but states correct hypotheses for that test in context, score part (c) as partially correct.

**4 Complete Response**

All three parts essentially correct

**3 Substantial Response**

Two parts essentially correct and one part partially correct

**2 Developing Response**

Two parts essentially correct and no part partially correct

*OR*

One part essentially correct and one or two parts partially correct

*OR*

Three parts partially correct

**1 Minimal Response**

One part essentially correct and no parts partially correct

*OR*

No parts essentially correct and two parts partially correct

**AP<sup>®</sup> STATISTICS  
2009 SCORING GUIDELINES**

**Question 2**

**Intent of Question**

The primary goals of this question were to assess a student's ability to (1) calculate a percentile value from a normal probability distribution; (2) recognize a binomial scenario and calculate an appropriate probability; and (3) use the sampling distribution of the sample mean to find a probability for the mean of five observations.

**Solution**

**Part (a):**

Let  $X$  denote the stopping distance of a car with new tread tires where  $X$  is normally distributed with a mean of 125 feet and a standard deviation of 6.5 feet. The  $z$ -score corresponding to a cumulative probability of 70 percent is  $z = 0.52$ . Thus, the 70th percentile value can be computed as:

$$x = \mu_x + z\sigma_x = 125 + 0.52(6.5) = 128.4 \text{ feet.}$$

**Part (b):**

From part (a), it was found that a stopping distance of 128.4 feet has a cumulative probability of 0.70. Thus the probability of a stopping distance greater than 128.4 is  $1 - 0.70 = 0.30$ .

Let  $Y$  denote the number of cars with the new tread pattern out of five cars that stop in a distance greater than 128.4 feet.  $Y$  is a binomial random variable with  $n = 5$  and  $p = 0.30$ .

$$\begin{aligned} P(Y \geq 2) &= 1 - P(Y \leq 1) = 1 - \left[ \binom{5}{0}(0.3)^0(0.7)^5 + \binom{5}{1}(0.3)^1(0.7)^4 \right] \\ &= 1 - 0.5282 = 0.4718. \end{aligned}$$

**Part (c):**

Let  $\bar{X}$  denote the mean of the stopping distances of five randomly selected cars. All tires have the new tread pattern. Because the stopping distance for each of the five cars has a normal distribution, the distribution of  $\bar{X}$  is normal with a mean of 125 feet and a standard deviation of  $\frac{6.5}{\sqrt{5}} = 2.91$  feet. Thus,

$$P(\bar{X} > 130) = P\left(Z > \frac{130 - 125}{6.5/\sqrt{5}}\right) \approx P(Z > 1.72) = 0.0427.$$

**Scoring**

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

# AP<sup>®</sup> STATISTICS

## 2009 SCORING GUIDELINES

### Question 2 (continued)

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

Essentially correct (E) if the student clearly indicates which distribution is being used, along with the parameters ( $\mu_x$  and  $\sigma_x$ ), and correctly calculates the percentile value with appropriate justification (except for minor arithmetic or transcription errors). There are three components: distribution, parameters, and calculation of distance.

*Notes:*

- The standard notation  $N(125, 6.5)$  defines distribution and parameters. Also, the z-score formula setup implies distribution and parameters. This applies only in part (a), because approximate normality is given in the stem of the problem.
- If the calculator command  $\text{invNorm}(0.70, 125, 6.5)$  is provided along with 128.4 feet *AND* an appropriately labeled sketch of a normal distribution is supplied, then the response should be scored as essentially correct (E). An appropriately labeled sketch must include correct labels for center and spread.
- If the calculator command  $\text{invNorm}(0.70, \mu = 125, \sigma = 6.5)$  is provided along with 128.4 feet, then the response should be scored as essentially correct (E).

Partially correct (P) if the student correctly supplies only two out of the three components.

*Note:* If the calculator command  $\text{invNorm}(0.70, 125, 6.5)$  is provided along with 128.4 feet, then the response should be scored as partially correct (P).

Incorrect (I) if the student correctly supplies at most one of the components.

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

Essentially correct (E) if the student recognizes this probability as an application of the binomial distribution and sets up the problem correctly by first finding the probability for  $p$ , the probability of a success, and then using this  $p$  to find the correct binomial probability. There are three components: distribution, parameters, and calculation.

*Note:* If the calculator command  $1-\text{binomcdf}(5, 0.3, 1)$  is provided along with 0.4718 and an identification of the distribution and its parameters—e.g., by the standard notation  $B(5, 0.3)$  or  $\text{Bin}(5, 0.3)$ —then the response should be scored as essentially correct (E).

Partially correct (P) if the student correctly supplies only two out of the three components.

*Notes:*

- As long as the student identifies the distribution and parameters—e.g., by the standard notation  $B(5, 0.3)$  or  $\text{Bin}(5, 0.3)$ —the binomial formula does not need to be set up to receive full credit. However, the binomial formula setup can suffice for identifying two of the three components: distribution and parameters.
- If the calculator command  $1-\text{binomcdf}(5, 0.3, 1)$  is provided along with 0.4718, then the response should be scored as partially correct (P).

Incorrect (I) if the student correctly supplies at most one of the components.

# AP<sup>®</sup> STATISTICS

## 2009 SCORING GUIDELINES

### Question 2 (continued)

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

Essentially correct (E) if the student recognizes that the distribution of the sample mean will be approximately normal with the appropriate mean and standard deviation and calculates the probability correctly. There are three components: sampling distribution, parameters, and calculation.

*Notes:*

- The z-score formula setup suffices only for parameters in part (c).
- If the calculator command  $\text{Normalcdf}(130, \infty, 125, 2.91)$  AND an appropriately labeled sketch of a normal distribution are provided along with the value obtained using the calculator, 0.0428, then the response should be scored as essentially correct (E). An appropriately labeled sketch must include correct labels for center and spread.
- If the calculator command  $\text{Normalcdf}(130, \infty, \mu = 125, \sigma = 2.91)$  is provided along with 0.0428, then the response should be scored as essentially correct (E).

Partially correct (P) if the student correctly supplies only two out of the three components.

*Note:* If the calculator command  $\text{Normalcdf}(130, \infty, 125, 2.91)$  is provided along with 0.0428, then the response should be scored as partially correct (P).

Incorrect (I) if the student correctly supplies at most one of the components.

*Notes:*

- The calculator solution is 0.0428. If this is the only information provided, the response is scored as incorrect (I).
- If a  $t$  distribution is used, then the response should be scored as incorrect (I).

A student should be penalized only once for using calculator syntax—that is, look at parts (a), (b), and (c) together.

#### **4 Complete Response**

All three parts essentially correct

#### **3 Substantial Response**

Two parts essentially correct and one part partially correct

#### **2 Developing Response**

Two parts essentially correct and no part partially correct

OR

One part essentially correct and one or two parts partially correct

OR

Three parts partially correct

**AP<sup>®</sup> STATISTICS  
2009 SCORING GUIDELINES**

**Question 2 (continued)**

**1 Minimal Response**

One part essentially correct and no parts partially correct

*OR*

No parts essentially correct and two parts partially correct

# AP<sup>®</sup> STATISTICS

## 2009 SCORING GUIDELINES

### Question 3

#### **Intent of Question**

The primary goals of this question were to assess a student's ability to (1) describe a randomization process required for comparing two groups in a randomized experiment; and (2) describe a potential consequence of using self-selection instead of randomization.

#### **Solution**

##### **Part (a) (completely randomized design):**

Each student will be assigned a unique random number using a random number generator on a calculator, statistical software, or a random number table. The assigned numbers will be listed in ascending order. The students with the lowest 12 numbers in the ordered list will receive the instructional program that requires physically dissecting frogs. The students with the highest 12 numbers will receive the instructional program that uses computer software to simulate the dissection of a frog.

##### **Part (a) *alternative* (randomized block design):**

Students will be paired or placed into blocks of size two, based on having similar pretest scores. So, the first block will contain the two students with the two lowest pretest scores, the second block will contain the two students with the third- and fourth-lowest pretest scores, and so on, with the last block containing the two students with the two highest pretest scores. In each block, the students will be assigned a unique random number using a random number generator on a calculator, statistical software, or a random number table. The student in each block with the lower random number will receive the instructional program that requires physically dissecting frogs, and the student with the higher random number will receive the instructional program that uses computer software to simulate the dissection of a frog.

##### **Part (b):**

By not randomizing and allowing the students to self-select, there is a potential for changes to occur in the differences between pretest and posttest scores for a particular group because of the characteristics of students who choose a particular instructional method, not because of the instructional method itself. For example, suppose frog-loving students already know a lot about frog anatomy; one would therefore expect these students to be less likely to show a large change between the pretest and posttest scores. Suppose the frog-loving students tend to select the computer simulation method (perhaps because they do not like the notion of dissecting the frogs they love). The possible low change between pretest and posttest scores for the computer simulation group might then be attributed to the students' already knowing a lot about frog anatomy beforehand, not to the instructional method itself. The frog dissection group might see a larger change in scores because the students entering this group are those with the lower pretest scores (less prior knowledge) and who are thus more likely to show greater improvement between pretest and posttest scores.

#### **Scoring**

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

# AP<sup>®</sup> STATISTICS

## 2009 SCORING GUIDELINES

### Question 3 (continued)

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

Essentially correct (E) if a proper method of randomization is described that (1) creates two groups of equal size; *AND* (2) assigns the named treatments to the groups in a manner that knowledgeable statistics users would employ to assign the students to the two instructional groups.

Partially correct (P) if only one of the two criteria above is met.

Incorrect (I) if neither criterion is met.

*Notes:*

- Coin tossing (or equivalent method) using a stopping rule to obtain equal sample sizes requires placing the students in the class in a random order. If this method does not include a random order, at best, part (a) is scored as partially correct.
- In using a random number table, if numbers are specified, the student must work with two-digit numbers. For example, if using the first 24 integers, the student must use 01–24, not 1–24. If the student uses numbers such as 1–24, a solution that would otherwise be essentially correct becomes partially correct, and a partially correct response becomes incorrect.

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

Essentially correct (E) if (1) blocks are formed based on students' having similar pretest scores; *AND* (2) the two students in each block are assigned to different treatments; *AND* (3) the method of randomization used to assign the students in each block to the treatments is correct and can be implemented after reading the student's response (in a manner that knowledgeable statistics users would employ to assign the students to the two instructional groups).

Partially correct (P) if two of the three components above are presented correctly.

Incorrect (I) if no more than one of the three components is presented correctly.

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

Essentially correct (E) if (1) the example gives a reasonable characteristic of the self-selected students in the study; *AND* (2) explains how this characteristic could be associated with changes in the differences between the pretest and posttest scores.

Partially correct (P) if (1) the example gives a reasonable characteristic of the self-selected students in the study; *AND* (2) a weak explanation is provided of how this characteristic could be associated with changes in the differences between pretest and posttest scores.

*Note:* A weak explanation of how a characteristic could be associated with changes in the differences between pretest and posttest scores must at least mention test scores or state that one group will perform better than the other. (Simply mentioning a behavioral difference is not sufficient.)

**AP<sup>®</sup> STATISTICS**  
**2009 SCORING GUIDELINES**

**Question 3 (continued)**

Incorrect (I) if an incorrect or no explanation is provided of how a characteristic could be associated with changes in the differences between pretest and posttest scores

*OR*

the example does not give a reasonable characteristic of the self-selected students in the study

*OR*

a student says that there must be an equal number of students in the class assigned to each treatment.

**4 Complete Response**

Both parts essentially correct

**3 Substantial Response**

One part essentially correct and the other part partially correct

**2 Developing Response**

One part essentially correct and the other part incorrect

*OR*

Both parts partially correct

**1 Minimal Response**

No part essentially correct and only one part partially correct

# AP<sup>®</sup> STATISTICS 2009 SCORING GUIDELINES

## Question 4

### Intent of Question

The primary goals of this question were to evaluate a student's ability to (1) identify and compute an appropriate confidence interval after checking the necessary conditions; (2) interpret the interval in the context of the question; and (3) use the confidence interval to make an inference about whether or not a council member's belief is supported.

### Solution

#### **Part (a):**

Step 1: Identify the appropriate confidence interval by name or formula and check for appropriate conditions.

The two-sample  $t$  interval for  $\mu_N - \mu_S$ , the difference in population mean response times, is

$$(\bar{x}_N - \bar{x}_S) \pm t^* \sqrt{\frac{s_N^2}{n_N} + \frac{s_S^2}{n_S}}$$

where  $\mu_N$  denotes the mean response for calls from the northern fire station and  $\mu_S$  denotes the mean response for calls from the southern fire station.

Conditions:    1. Independent random samples  
                  2. Large samples or normal population distributions

A random sample of 50 calls was selected from the northern fire station, independent of the random sample of 50 calls selected from the southern fire station.

The use of the two-sample  $t$  interval is reasonable because both sample sizes are large ( $n_N = 50 > 30$  and  $n_S = 50 > 30$ ), and by the central limit theorem, the sampling distributions for the two sample means are approximately normal. Therefore the sampling distribution of the difference of the sample means  $\bar{x}_N - \bar{x}_S$  is approximately normal.

Step 2: Correct Mechanics

Unequal variances: Degrees of freedom = 96.

$$\begin{aligned} &(4.3 - 5.3) \pm 1.985 \sqrt{\frac{3.7^2}{50} + \frac{3.2^2}{50}} \\ &-1.0 \pm 1.985 \times .6918 \\ &(-2.37, 0.37) \end{aligned}$$

Step 3: Interpretation

Based on these samples, one can be 95 percent confident that the difference in the population mean response times (northern - southern) is between -2.37 minutes and 0.37 minutes.

# AP<sup>®</sup> STATISTICS 2009 SCORING GUIDELINES

## Question 4 (continued)

### Part (b):

Zero is within the 95 percent confidence interval of plausible values for the difference in population means. Therefore this confidence interval does not support the council member’s belief that there is a difference in mean response times for the two fire stations.

### Scoring

This problem is scored in four sections. Section 1 consists of part (a), step 1; section 2 consists of part (a), step 2; section 3 consists of part (a), step 3; and section 4 consists of part (b). Sections 1, 2, and 3 are scored as essentially correct (E) or incorrect (I), and section 4 is scored as essentially correct (E), partially correct (P), or incorrect (I).

**Section 1:** Identify the appropriate confidence interval by name or formula and check for appropriate conditions.

Section 1 is scored as follows:

Essentially correct (E) if the student does both of the following:

- Indicates that the appropriate procedure is a  $t$  confidence interval for difference in means, either by name or formula.
- Verifies that the sample sizes are large enough to use this procedure by referencing a number (such as 25 or 30) *OR* by stating that 50 is large enough that the central limit theorem applies.

Incorrect (I) if the student omits at least one of the two elements above *OR* names or gives a formula for a two-sample  $z$  confidence interval instead of a two-sample  $t$  confidence interval.

### Section 2: Correct Mechanics

Section 2 is scored as follows:

Essentially correct (E) if the student shows a correct confidence interval, either by displaying numbers in the formula or by writing the numerical interval.

*Note:* The following are acceptable solutions for section 2:

- The following degrees of freedom,  $t^*$  values, and confidence intervals are all acceptable.

<b>Solutions to Step 2</b>			
<b>Procedure</b>	<b>Degrees of freedom</b>	<b><math>t^*</math></b>	<b>Confidence interval for <math>\mu_N - \mu_S</math></b>
Unequal variances	96	1.985	(-2.37, 0.37)
Conservative df	40 (using table)	2.021	(-2.40, 0.40)
Conservative df	49	2.010	(-2.39, 0.39)
Conservative df	50 (using table)	2.009	(-2.39, 0.39)
Pooled variance	98	1.984	(-2.37, 0.37)

# AP<sup>®</sup> STATISTICS

## 2009 SCORING GUIDELINES

### Question 4 (continued)

- Students do not need to state explicitly the degrees of freedom used.
- It is acceptable if a student reports a confidence interval for  $\mu_S - \mu_N$ , in which case the signs are reversed.
- An identifiable minor arithmetic error will not necessarily change a score on section 2 from essentially correct to incorrect.

Incorrect (I) if the numerical interval given is incorrect because of a substantive statistical error, such as failing to square the standard deviations *OR* if the interval is missing or completely wrong.

### Section 3: Interpretation

Section 3 is scored as follows:

Essentially correct (E) if the student correctly interprets the confidence interval (not the confidence level) in context. The interpretation must indicate that the interval is for a difference in mean response times. If the student gives an interpretation of *both* the confidence level and the confidence interval, both must be correct to be scored as essentially correct (E).

Incorrect (I) if a student gives an interpretation of the confidence level *instead of* the confidence interval *OR* if the interpretation of the confidence interval is wrong.

*Note:* The correct interpretation of the confidence interval for step 3 of part (a) may be found in part (b).

### Section 4

Section 4 is scored as follows:

Essentially correct (E) if the student makes a correct conclusion in context, supported by the fact that zero is contained within the 95 percent confidence interval.

Partially correct (P) if the student makes a correct conclusion supported by the fact that zero is contained in the confidence interval, but one or both of the following occurs:

- The student omits the context.
- The student makes a statistically incorrect statement in the explanation, such as stating that the council member's belief is wrong. Such a statement is equivalent to accepting the null hypothesis of no difference in population mean response times.

Incorrect (I) if the student gives the incorrect conclusion *OR* makes no reference to a confidence interval.

*Notes:*

- The answer in part (b) needs to be consistent with the confidence interval given in part (a). If the interval does not cover zero, then part (b) must state that the interval does support the council member's belief because the interval does not contain zero.
- The response for part (b) may be found in the space allocated to part (a).

# AP<sup>®</sup> STATISTICS

## 2009 SCORING GUIDELINES

### Question 4 (continued)

Each essentially correct (E) response counts as 1 point, and a partially correct (P) response in part (b) 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,  $1\frac{1}{2}$  points), use a holistic approach to determine whether to score up or down, depending on the strength of the response and communication.

# AP<sup>®</sup> STATISTICS

## 2009 SCORING GUIDELINES

### Question 5

#### **Intent of Question**

The primary goals of this statistical inference question were to assess a student's ability to (1) interpret a  $p$ -value in context; (2) make an appropriate conclusion about the study based on the  $p$ -value; and (3) based on the conclusion, identify the type of error that could have occurred and a possible consequence of this error in context.

#### **Solution**

##### **Part (a):**

The  $p$ -value of 0.0761 measures the chance of observing a difference between the two sample proportions ( $\hat{p}_{CC} - \hat{p}_{CC+MMR}$ ) as large as or larger than the one observed, if the survival rates for the two treatments (CC alone and CC + MMR) are in fact the same.

##### **Part (b):**

Because the  $p$ -value of 0.0761 is greater than 0.05, the null hypothesis should not be rejected. That is, there is not sufficient evidence to conclude that the treatment "CC alone" produces a higher survival rate than the standard treatment "CC + MMR."

##### **Part (c):**

Because the null hypothesis was not rejected, a Type II error could have occurred. A possible consequence is that CC + MMR would continue as the accepted practice when, in fact, CC alone would result in a higher survival rate.

#### **Scoring**

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

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

A correct interpretation must include the following three components:

- Correct probability phrase (e.g., "The  $p$ -value of 0.0761 measures the chance of . . .") that includes the words "as large as" (or something similar).
- Correct conditional phrase (e.g., "if the survival rates for the two treatments are in fact the same").
- Correct context.

Essentially correct (E) if the response includes all three components.

Partially correct (P) if the response includes the first component and one of the other two components

OR

if the probability phrase is complete except for the omission of the words "as large as" (or something similar) and the other two components are included.

Incorrect (I) if the response includes no more than one component.

# AP<sup>®</sup> STATISTICS

## 2009 SCORING GUIDELINES

### Question 5 (continued)

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

Essentially correct (E) if a correct conclusion (failure to reject  $H_0$ ) is provided in context with appropriate linkage to the  $p$ -value.

Partially correct (P) if a correct conclusion is provided but either the context or linkage is missing  
*OR*  
the student “accepts  $H_0$ ” (or something similar) and provides both context and linkage.

Incorrect (I) if the student rejects  $H_0$   
*OR*  
the student provides neither context nor linkage.

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

A correct response must include the following two components:

- The type of error named is consistent with the conclusion in part (b).
- A consequence is provided (in context) that is consistent with the conclusion in part (b) and is specific with regard to the treatments. The consequence must address how heart attack patients will be treated as a result of the error.

Essentially correct (E) if the response includes both components.

Partially correct (P) if the response includes only the consequence component  
*OR*

the type of error named is consistent with the conclusion in part (b) *AND* a correct definition (either generic or in context) of that error is given, but the consequence component is either missing or incorrect.

Incorrect (I) if the response does not include the consequence component, apart from the exception given above as the second type of a partially correct response.

#### **4 Complete Response**

All three parts essentially correct

#### **3 Substantial Response**

Two parts essentially correct and one part partially correct

#### **2 Developing Response**

*OR* Two parts essentially correct and no part partially correct

One part essentially correct and one or two parts partially correct

*OR* Three parts partially correct

**AP<sup>®</sup> STATISTICS**  
**2009 SCORING GUIDELINES**

**Question 5 (continued)**

**1 Minimal Response**

One part essentially correct and no parts partially correct

*OR*

No parts essentially correct and two parts partially correct

# AP<sup>®</sup> 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<sup>®</sup> 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  $\mu$  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<sup>®</sup> 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<sup>®</sup> 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. $\frac{\text{maximum} - Q3}{Q1 - \text{minimum}}$ | 3. $\frac{\left(\frac{Q1 + Q3}{2}\right)}{\text{median}}$                         |
| 2. $\frac{Q3 - \text{median}}{\text{median} - Q1}$   | 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.