



## **AP<sup>®</sup> Statistics 2012 Scoring Guidelines**

### **The College Board**

The College Board is a mission-driven not-for-profit organization that connects students to college success and opportunity. Founded in 1900, the College Board was created to expand access to higher education. Today, the membership association is made up of more than 5,900 of the world's leading educational institutions and is dedicated to promoting excellence and equity in education. Each year, the College Board helps more than seven million students prepare for a successful transition to college through programs and services in college readiness and college success — including the SAT<sup>®</sup> and the Advanced Placement Program<sup>®</sup>. The organization also serves the education community through research and advocacy on behalf of students, educators, and schools. 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.

© 2012 The College Board. College Board, Advanced Placement Program, AP, SAT and the acorn logo are registered trademarks of the College Board. All other products and services may be trademarks of their respective owners. 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.org](http://www.collegeboard.org).

AP Central is the official online home for the AP Program: [apcentral.collegeboard.org](http://apcentral.collegeboard.org).



# AP<sup>®</sup> STATISTICS

## 2012 SCORING GUIDELINES

### Question 1

#### **Intent of Question**

The primary goals of this question were to assess students' ability to (1) describe a nonlinear association based on a scatterplot; (2) describe how an unusual observation may affect the appropriateness of using a linear model for bivariate numeric data; (3) implement a decision-making criterion on data presented in a scatterplot.

#### **Solution**

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

The data show a weak but positive association between price and quality rating for these sewing machines. The form of the association does not appear to be linear. Among machines that cost less than \$500, there appears to be very little association between price and quality rating. But the machines that cost more than \$500 do generally have better quality ratings than those that cost less than \$500, which causes the overall association to be positive.

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

The sewing machine that most affects the appropriateness of using a linear regression model is the one that costs about \$2,200 and has a quality rating of about 65. Although the other four sewing machines costing more than \$500 generally have higher quality ratings than those costing under \$500, their prices and quality ratings follow a trend that suggests that quality ratings may not continue to increase with higher prices, but instead may approach a maximum possible quality rating. The \$2,200 sewing machine is the most expensive of all but has a relatively low quality rating, which is consistent with a nonlinear model that approaches a maximum possible quality rating and then perhaps decreases. If a linear model were fit to all of the data, this one machine would substantially pull the regression line toward it, resulting in a poor overall fit of the line to the data.

# AP<sup>®</sup> STATISTICS 2012 SCORING GUIDELINES

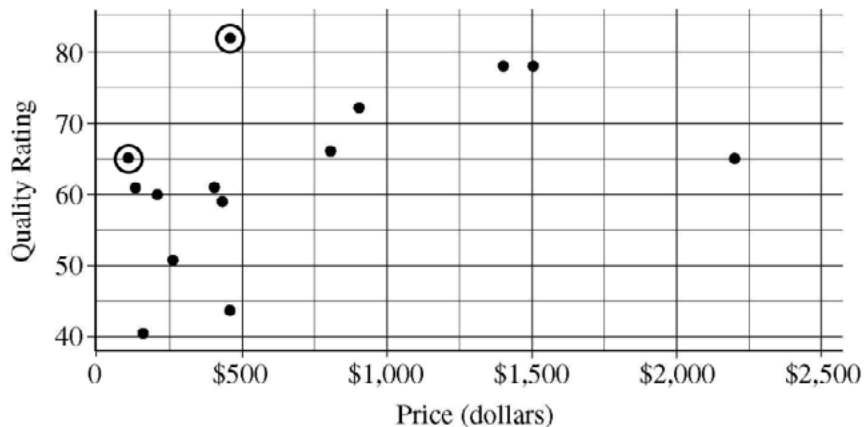
## Question 1 (continued)

### Part (c):

According to Chris's criterion, there are two sewing machine models that he will consider buying:

1. The model that costs a bit more than \$100 and has a quality rating of 65.
2. The model that costs a bit below \$500 and has a quality rating of 81 or 82.

The data points corresponding to these two machines have been circled on the scatterplot below.



### 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 the response correctly describes three aspects of association: direction (positive), strength (weak or moderate), and form (curved or nonlinear), *AND* describes the association in context.

Partially correct (P) if the response correctly describes two aspects of association in context  
*OR*

if the response describes all three aspects of association without context.

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

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

Essentially correct (E) if the response identifies the correct point with reasonable approximations to the price and quality values *AND* gives either of the following two explanations:

1. The point in conjunction with the entire collection of points appears to have a curved (or nonlinear) form.
2. A linear model that includes all the points would result in a poor overall fit to the data, largely owing to the presence and influence of the identified point.

# AP<sup>®</sup> STATISTICS

## 2012 SCORING GUIDELINES

### Question 1 (continued)

Partially correct (P) if the response identifies the correct point with reasonable approximations to the price and quality values *AND* gives a weak explanation of why the point affects the reasonableness of a linear model. The following are examples of weak explanations.

1. The point is an outlier.
2. Removal of the point makes the pattern more linear.
3. The point does not follow the linear pattern of the others.
4. A sewing machine this expensive should have a higher quality rating.
5. There is a much cheaper sewing machine with the same quality rating as this one.
6. The point has considerable influence on the parameters of the least squares regression line.

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

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

Essentially correct (E) if the correct two points are circled *AND* no other points are circled.

Partially correct (P) if the correct two points are circled *AND* one or two other points are circled.

*OR*

if only one of the two correct points is circled *AND* at most one other point is circled.

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

#### **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 one part incorrect

*OR*

One part essentially correct and two parts partially correct

*OR*

One part essentially correct and one part partially correct  
(*BUT* see the exception noted with an asterisk below)

*OR*

All three parts partially correct

#### **1 Minimal Response**

One part essentially correct and two parts incorrect

*OR*

\*Part (c) essentially correct, part (b) partially correct, and part (a) incorrect

*OR*

Two parts partially correct and one part incorrect

**AP<sup>®</sup> STATISTICS**  
**2012 SCORING GUIDELINES**

**Question 2**

**Intent of Question**

The primary goals of this question were to assess students' ability to (1) perform calculations and compute expected values related to a discrete probability distribution; (2) implement a normal approximation based on the central limit theorem.

**Solution**

**Part (a):**

By counting the number of sectors for each value and dividing by 10, the probability distribution is calculated to be:

$x$	\$2	\$1	-\$8
$P(x)$	0.6	0.3	0.1

**Part (b):**

The expected value of the net contribution for one play of the game is:

$$E(x) = \$2(0.6) + \$1(0.3) + (-\$8)(0.1) = \$0.70 \text{ (or 70 cents).}$$

**Part (c):**

The expected contribution after  $n$  plays is  $\$0.70n$ . Setting this to be at least \$500 and solving for  $n$  gives:

$$0.70n \geq 500, \text{ so } n \geq \frac{500}{0.70} \approx 714.286,$$

so 715 plays are needed for the expected contribution to be at least \$500.

**Part (d):**

The normal approximation is appropriate because the very large sample size ( $n = 1,000$ ) ensures that the central limit theorem holds. Therefore, the sample mean of the contributions from 1,000 plays has an approximately normal distribution, and so the sum of the contributions from 1,000 plays also has an approximately normal distribution.

$$\text{The z-score is } \frac{500 - 700}{92.79} \approx -2.155.$$

The probability that a standard normal random variable exceeds this z-score of  $-2.155$  is 0.9844. Therefore, the charity can be very confident about gaining a net contribution of at least \$500 from 1,000 plays of the game.

# AP<sup>®</sup> STATISTICS

## 2012 SCORING GUIDELINES

### Question 2 (continued)

#### **Scoring**

This question is scored in three sections. Section 1 consists of parts (a) and (b); section 2 consists of part (c); and section 3 consists of part (d). Sections 1, 2, and 3 are scored as essentially correct (E), partially correct (P), or incorrect (I).

**Section 1** is scored as follows:

Essentially correct (E) if all three probabilities are filled in correctly in the table in part (a) *AND* the expected value is calculated correctly in part (b), with work shown.

Partially correct (P) if all three probabilities are filled in correctly in the table in part (a) *AND* the expected value is not calculated correctly in part (b),

*OR*

the probabilities in part (a) are not all correct *AND* the expected value in part (b) is calculated appropriately from the probabilities given in part (a) or from the correct probabilities.

Incorrect (I) if the response does not meet the criteria for E or P.

**Section 2** is scored as follows:

Essentially correct (E) if the response addresses the following two components:

1. Provides a solution based on a reasonable calculation, equation, or inequality from the answer given in part (b).
2. Clearly selects the next higher integer as the answer.

Partially correct (P) if the response correctly completes component (1) listed above but not component (2).

Incorrect (I) if the response does not meet the criteria for E or P.

**Section 3** is scored as follows:

Essentially correct (E) if the response correctly addresses the following three components:

1. Indicates the use of a normal distribution with the correct mean and standard deviation.
2. Uses the correct boundary and indicates the correct direction.
3. Has the correct normal probability consistent with components (1) and (2).

Partially correct (P) if the response correctly addresses exactly two of the three components listed above.

Incorrect (I) if the response does not meet the criteria for E or P.

# AP<sup>®</sup> STATISTICS

## 2012 SCORING GUIDELINES

### Question 2 (continued)

#### Notes

- Because the question asks students to use a normal distribution and specifies the parameter values, the response does not have to justify the normal approximation or show how to calculate the parameter values.
- If the response earns credit for component (1) but no direction has been provided for component (2), then the response earns credit for component (3) if the correct probability of 0.9844 is reported.
- If the response does not earn credit for component (1) owing to incorrect identification of the mean and/or standard deviation, then the response can still earn credit for component (2) if the boundary is calculated correctly from the mean and standard deviation indicated in component (1).

#### **4 Complete Response**

All three sections essentially correct

#### **3 Substantial Response**

Two sections essentially correct and one section partially correct

#### **2 Developing Response**

Two sections essentially correct and one section incorrect

*OR*

One section essentially correct and one or two sections partially correct

*OR*

Three sections partially correct

#### **1 Minimal Response**

One section essentially correct and two sections incorrect

*OR*

Two sections partially correct and one section incorrect

# AP<sup>®</sup> STATISTICS

## 2012 SCORING GUIDELINES

### Question 3

#### **Intent of Question**

The primary goals of this question were to assess students' ability to (1) compare two distributions presented with histograms; (2) comment on the appropriateness of using a two-sample  $t$ -procedure in a given setting.

#### **Solution**

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

Household size tended to be larger in 1950 than in 2000. The histograms reveal a much larger proportion of small (1-, 2-, and 3-person) households in 2000 than in 1950. Similarly, the histograms reveal a much smaller proportion of large (5-person and larger) households in 2000 than in 1950. Also, the median household sizes can be calculated to be 5 people per household in 1950 compared with 3 or 4 people per household in 2000. The year 1950 displayed slightly more variability in household sizes than the year 2000. Although the interquartile ranges for both years are the same (3 people), the standard deviation (1950: about 2.6 people; 2000: about 2.1 people) and the range (1950: 13 people; 2000: 11 people) are larger for 1950 than for 2000. Both distributions of household size are skewed to the right. In both years, there are a few households with very large families, as large as 14 people in 1950 and 12 people in 2000.

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

The conditions for applying a two-sample  $t$ -procedure are:

1. The data come from independent random samples or from random assignment to two groups;
2. The populations are normally distributed, or both sample sizes are large;
3. The population sizes are at least 10 (or 20) times the sample sizes.

The first condition is satisfied because independent random samples were selected for the years 1950 and 2000. The second condition is satisfied because the sample sizes (500 in each group) are quite large, despite the right skewness of the distributions of household sizes in the sample data. The third condition is satisfied because the number of households in the large metropolitan area in both 1950 and 2000 would easily exceed  $10 \times 500 = 5,000$ .

#### **Scoring**

This question is scored in four sections. Part (a) has three components: (1) comparing the centers of the two distributions; (2) comparing variability for the two distributions; (3) identifying the shapes of both distributions and including context related to the variable of interest. Section 1 consists of part (a), component 1; section 2 consists of part (a), component 2; section 3 consists of part (a), component 3. Section 4 consists of part (b). Sections 1 and 2 are scored as essentially correct (E) or incorrect (I). Sections 3 and 4 are scored as essentially correct (E), partially correct (P), or incorrect (I).

**Section 1** is scored as follows:

Essentially correct (E) if the response correctly compares center (or location) for both distributions.

Incorrect (I) otherwise.



# AP<sup>®</sup> STATISTICS 2012 SCORING GUIDELINES

## Question 3 (continued)

**Section 2** is scored as follows:

Essentially correct (E) if the response correctly compares variability for both distributions.

Incorrect (I) otherwise.

**Section 3** is scored as follows:

Essentially correct (E) if the response includes context related to the variable of interest (household size) *AND* the response correctly identifies the shapes of both distributions.

Partially correct (P) if the response correctly identifies the shapes of both distributions *BUT* does *NOT* include context related to the variable of interest (household size),

*OR*

if the response correctly identifies the shape of only one distribution *AND* includes context related to the variable of interest (household size).

Incorrect (I) otherwise.

**Section 4** is scored as follows:

Essentially correct (E) if the response correctly states and checks the following two conditions.

1. The data come from independent random samples
2. Normality/sample size conditions.

Partially correct (P) if the response correctly states and checks only one of the two conditions listed above,

*OR*

if the response correctly refers to random samples and large sample size, *BUT* does *NOT* state and check either condition correctly.

Incorrect (I) otherwise.

*Note:* The population size condition does not need to be checked to earn E or P.

Each essentially correct (E) section counts as 1 point. Each partially correct (P) section counts as  $\frac{1}{2}$  point.

<b>4</b>	<b>Complete Response</b>
<b>3</b>	<b>Substantial Response</b>
<b>2</b>	<b>Developing Response</b>
<b>1</b>	<b>Minimal Response</b>

If a response is between two scores (for example,  $2\frac{1}{2}$  points), use a holistic approach to decide whether to score up or down, depending on the overall strength of the response and communication.

# AP<sup>®</sup> STATISTICS 2012 SCORING GUIDELINES

## Question 4

### Intent of Question

The primary goal of this question was to assess students' ability to identify, set up, perform, and interpret the results of an appropriate hypothesis test to address a particular question. More specific goals were to assess students' ability to (1) state appropriate hypotheses; (2) identify the name of an appropriate statistical test and check appropriate assumptions/conditions; (3) calculate the appropriate test statistic and  $p$ -value; (4) draw an appropriate conclusion, with justification, in the context of the study.

### Solution

Step 1: States a correct pair of hypotheses.

Let  $p_{07}$  represent the population proportion of adults in the United States who would have answered "yes" about the effectiveness of television commercials in December 2007. Let  $p_{08}$  represent the analogous population proportion in December 2008.

The hypotheses to be tested are  $H_0: p_{07} = p_{08}$  versus  $H_a: p_{07} \neq p_{08}$ .

Step 2: Identifies a correct test procedure (by name or by formula) and checks appropriate conditions.

The appropriate procedure is a two-sample  $z$ -test for comparing proportions.

Because these are sample surveys, the first condition is that the data were gathered from independent random samples from the two populations. This condition is met because we are told that the subjects were randomly selected in the two different years. Although we are not told whether the samples were selected independently, this is a reasonable assumption given that they are samples of different sizes selected in different years.

The second condition is that the sample sizes are large, relative to the proportions involved. This condition is satisfied because all sample counts (622 "yes" in 2007;  $1,020 - 622 = 398$  "no" in 2007; 676 "yes" in 2008;  $1,009 - 676 = 333$  "no" in 2008) are all at least 10 (or, are all at least 5).

An additional condition may be checked: The population sizes (more than 200 million adults in the United States) are much larger than 10 (or, 20) times the sample sizes.

Step 3: Correct mechanics, including the value of the test statistic and  $p$ -value (or rejection region).

The sample proportions who answered "yes" are:

$$\hat{p}_{07} = \frac{622}{1,020} \approx 0.6098 \quad \text{and} \quad \hat{p}_{08} = \frac{676}{1,009} \approx 0.6700.$$

The combined proportion,  $\hat{p}_c$ , who answered "yes" in these two years is:

$$\hat{p}_c = \frac{622 + 676}{1,020 + 1,009} = \frac{1,298}{2,029} \approx 0.6397.$$

**AP<sup>®</sup> STATISTICS**  
**2012 SCORING GUIDELINES**

**Question 4 (continued)**

The test statistic is:

$$z = \frac{\hat{p}_{07} - \hat{p}_{08}}{\sqrt{\hat{p}_c(1 - \hat{p}_c)\left(\frac{1}{n_{07}} + \frac{1}{n_{08}}\right)}} = \frac{0.6098 - 0.6700}{\sqrt{(0.6397)(1 - 0.6397)\left(\frac{1}{1,020} + \frac{1}{1,009}\right)}} \approx -2.82.$$

The  $p$ -value is  $2P(Z \leq -2.82) \approx 0.0048$ .

Step 4: State a correct conclusion in the context of the study, using the result of the statistical test.

Because this  $p$ -value is smaller than any common significance level such as  $\alpha = 0.05$  or  $\alpha = 0.01$  (or, because this  $p$ -value is so small), we reject  $H_0$  and conclude that the data provide convincing (or, statistically significant) evidence that the proportion of all adults in the United States who would answer “yes” to the question about the effectiveness of television commercials changed from December 2007 to December 2008.

**Scoring**

Each of steps 1, 2, 3, and 4 are scored as essentially correct (E), partially correct (P), or incorrect (I).

**Step 1** is scored as follows:

Essentially correct (E) if the response identifies correct parameters *AND* both hypotheses are labeled and state the correct relationship between the parameters.

Partially correct (P) if the response identifies correct parameters *OR* states correct relationships, but not both.

Incorrect (I) if the response does not meet the criteria for E or P.

*Note:* Either defining the parameter symbols in context, or simply using common parameter notation, such as  $p_{07}$  and  $p_{08}$ , with subscripts clearly relevant to the context, is sufficient.

**Step 2** is scored as follows:

Essentially correct (E) if the response correctly includes the following three components:

1. Identifies the correct test procedure (by name or by formula).
2. Checks for randomness.
3. Checks for normality.

Partially correct (P) if the response correctly includes two of the three components listed above.

Incorrect (I) if the response correctly includes one or none of the three components listed above.

# AP<sup>®</sup> STATISTICS

## 2012 SCORING GUIDELINES

### Question 4 (continued)

**Step 3** is scored as follows:

Essentially correct (E) if the response correctly calculates both the test statistic and a  $p$ -value that is consistent with the stated alternative hypothesis.

Partially correct (P) if the response correctly calculates the test statistic but not the  $p$ -value,

*OR*

if the response calculates the test statistic incorrectly but then calculates the correct  $p$ -value for the computed test statistic.

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

**Step 4** is scored as follows:

Essentially correct (E) if the response provides a correct decision in context, also providing justification based on linkage between the  $p$ -value and conclusion.

Partially correct (P) if the response provides a correct decision, with linkage to the  $p$ -value, but not in context,

*OR*

if the response provides a correct decision in context, but without justification based on linkage to the  $p$ -value.

Incorrect (I) if the response does not meet the criteria for E or P.

*Note:* If the decision is consistent with an incorrect  $p$ -value from step 3, and also in context with justification based on linkage to the  $p$ -value, then step 4 is scored as E.

Each essentially correct (E) step counts as 1 point. Each partially correct (P) step counts as  $\frac{1}{2}$  point.

<b>4</b>	<b>Complete Response</b>
<b>3</b>	<b>Substantial Response</b>
<b>2</b>	<b>Developing Response</b>
<b>1</b>	<b>Minimal Response</b>

If a response is between two scores (for example,  $2\frac{1}{2}$  points), use a holistic approach to decide whether to score up or down, depending on the overall strength of the response and communication.

# AP<sup>®</sup> STATISTICS

## 2012 SCORING GUIDELINES

### Question 5

#### **Intent of Question**

The primary goals of this question were to assess students' ability to (1) describe a Type II error and its consequence in a particular study; (2) draw an appropriate conclusion from a  $p$ -value; (3) describe a flaw in a study and its effect on inference from a sample to a population.

#### **Solution**

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

In the context of the study, a Type II error means failing to reject the null hypothesis that 35 percent of adult residents in the city are able to pass the test when, in reality, less than 35 percent are able to pass the test. The consequence of this error is that the council would not fund the program, and the city would continue to have a smaller proportion of physically fit residents than the council would like.

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

Because the  $p$ -value of 0.97 is larger than  $\alpha = 0.05$ , we fail to reject the null hypothesis. There is not convincing evidence that the proportion of adult residents in the city who are able to pass the physical fitness test is less than 0.35. After all, the sample proportion of  $\hat{p} = 0.416$  is actually higher than 0.35, which is in the opposite direction of the alternative hypothesis.

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

This is not a randomly selected sample because the sample was selected by recruiting volunteers. It seems reasonable to think that volunteers would be more physically fit than the population of city adults as a whole. Therefore, the sample proportion will likely overestimate the population proportion of adult residents in the city who are able to pass the physical fitness test.

#### **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 the response correctly completes the following two components:

1. Describes the error in context by referring to the proportion of adult residents in the city who are able to pass the physical fitness test.
2. Describes the consequence as not funding the program and/or continuing poor physical fitness of the adult residents in the city.

# AP<sup>®</sup> STATISTICS

## 2012 SCORING GUIDELINES

### Question 5 (continued)

#### Notes

- If a response provides more than one description of a Type II error, score the weakest attempt.
- Referring to the symbolic hypotheses is *not sufficient* for context.
- Referring to funding and/or the city council is *not sufficient* for context.
- If a response describes a Type II error incorrectly, the response can get the consequence component correct if it is consistent with the incorrectly described error.
- If a response provides more than one description of a Type II error, the response can get the consequence component correct if the consequence is clearly linked to one of the error descriptions and is consistent with the error to which it is linked.
- If a response gives an incomplete description of a Type II error (for example, “we fail to reject the null hypothesis that the proportion of adult residents who are able to pass is 0.35”), the response can get the consequence component correct if the consequence is consistent with the partial description of the error.
- If a response provides no description of a Type II error, the response cannot get the consequence component correct.

Partially correct (P) if the response correctly completes only one of the two components listed above.

Incorrect (I) if the response correctly completes neither of the two components listed above.

*Note:* Describing the Type II error only in terms of the consequence (for example, “They don’t fund the program when they should”) should get credit for the consequence but should not get credit for the error, because there is no reference to the proportion of adult residents in the city who are able to pass the test.

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

Essentially correct (E) if the response correctly completes the following three components:

1. Links the  $p$ -value to the conclusion by stating that the  $p$ -value is greater than  $\alpha = 0.05$ ,  
*OR*  
by stating that the  $p$ -value is large,  
*OR*  
by correctly interpreting the  $p$ -value.
2. Uses context by referring to the proportion of adult residents who are able to pass the test,  
*OR*  
by referring to the funding of the program.
3. Makes a correct conclusion that describes the lack of evidence for the alternative hypothesis ( $H_a : p < 0.35$ ).

#### Notes

- If a response includes an incorrect interpretation of the  $p$ -value, then the response cannot earn credit for the linkage component, even if the response explicitly compares the  $p$ -value to  $\alpha$  or describes the  $p$ -value as large.
- Referring to the symbolic hypotheses is not sufficient for context.
- Accepting the null hypothesis or some equivalent statement such as “the population proportion is (or is likely to be, or is about) 0.35” cannot receive credit for the conclusion component, even if the student makes additional correct statements about the alternative hypothesis.

# AP<sup>®</sup> STATISTICS

## 2012 SCORING GUIDELINES

### Question 5 (continued)

- Stating that the null hypothesis should not be rejected is not sufficient for the conclusion, because it does not address the direction of  $H_a$ .
- Correctly addressing the consequence (“They don’t fund the program”) is sufficient if the response also indicates that the null hypothesis is not being rejected.
- Drawing a conclusion about the sample proportion (for example, “proportion who passed the test”) is not sufficient for the conclusion, because it does not properly address the parameter in  $H_a$ .

Partially correct (P) if the response correctly completes two of the three components listed above.

Incorrect (I) if the response correctly completes one or none of the three components listed above.

#### Notes

- A response that says the  $p$ -value is very large, recognizes that the sample proportion ( $\hat{p} = 0.416$ ) is greater than  $p = 0.35$ , and consequently concludes there is no evidence to support  $H_a$  (in context) is scored as essentially correct (E).
- A response that *rejects*  $H_0$  is scored as incorrect (I).

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

Essentially correct (E) if the response correctly completes the following three components:

1. States that the sample is not random and/or says that volunteers were used.
2. Describes *how* the sample is “different” with regard to physical fitness or another variable related to the ability to pass the physical fitness test.
3. Addresses the idea of making an inference from the sample to the population by stating that the sample statistic will overestimate the population parameter or that the sample will not be representative of the population.

#### Notes

- If for the first component a student provides additional proposed flaws (for example, “the sample size is too small”), score the weakest attempt.
- Saying only that the sample is different or not representative does not address *how* the sample is different.
- Saying “physically fit people will be overrepresented” or “the results cannot be generalized” or “the results will be inaccurate” lack a specific reference to the population and is not sufficient for the third component.
- Referring to “bias” is not sufficient for the first component unless the concept of bias is clearly explained (for example, saying that the sample proportion will tend to overestimate the population proportion).
- Incorrect application of statistical concepts (for example, saying that the statistic is “skewed,” discussing cause and effect) results in a loss of credit for the third component.

Partially correct (P) if the response correctly addresses two of the three components listed above.

Incorrect (I) if the response correctly addresses one or none of the three components listed above.

**AP<sup>®</sup> STATISTICS**  
**2012 SCORING GUIDELINES**

**Question 5 (continued)**

**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 one part incorrect

*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 two parts incorrect

*OR*

Two parts partially correct and one part incorrect



**AP<sup>®</sup> STATISTICS**  
**2012 SCORING GUIDELINES**

**Question 6**

**Intent of Question**

The primary goals of this question were to assess students' ability to (1) implement simple random sampling; (2) calculate an estimated standard deviation for a sample mean; (3) use properties of variances to determine the estimated standard deviation for an estimator; (4) explain why stratification reduces a standard error in a particular study.

**Solution**

**Part (a):**

Peter can number the students from 1 to 2,000 and then use a calculator or computer to generate 100 unique random numbers between 1 and 2,000 without replacement. If non-unique numbers are generated, the repeated numbers are ignored until 100 unique numbers are obtained. The students whose numbers correspond to the randomly generated numbers are then selected for the sample.

**Part (b):**

The estimated standard deviation of the sampling distribution of the sample mean is:

$$\frac{s}{\sqrt{n}}, \text{ or } \frac{4.13}{\sqrt{100}} = 0.413.$$

**Part (c):**

The variance of Rania's estimator is  $(0.6)^2 \text{Var}(\bar{X}_f) + (0.4)^2 \text{Var}(\bar{X}_m)$ , where  $\text{Var}(\bar{X}_f) = \frac{\sigma_f^2}{n_f}$  represents

the variance of the point estimator for females and  $\text{Var}(\bar{X}_m) = \frac{\sigma_m^2}{n_m}$  represents the variance of the point estimator for males.

The estimated standard deviation is the square root of the variance. Using the respective sample standard deviations  $s_f$  and  $s_m$  for the population parameters, Rania's estimate is calculated as:

$$\sqrt{(0.6)^2 \frac{s_f^2}{n_f} + (0.4)^2 \frac{s_m^2}{n_m}} = \sqrt{(0.6)^2 \frac{(1.80)^2}{60} + (0.4)^2 \frac{(2.22)^2}{40}} \approx \sqrt{0.01944 + 0.01972} \approx 0.198.$$

**Part (d):**

The comparative dotplots from Rania's data reveal that the distribution of the number of soft drinks for females appears to be quite different from that of males. In particular, the centers of the distributions appear to be significantly different. Additionally, the variability of values around the center *within* gender in each of Rania's dotplots appears to be considerably less than the variability displayed in the dotplot of Peter's data. Rania's estimator takes advantage of the decreased variability within gender because her data were obtained by sampling the two genders separately. Peter's estimator has more variability because his data were obtained from a simple random sample of all the high school students.

# AP<sup>®</sup> STATISTICS

## 2012 SCORING GUIDELINES

### Question 6 (continued)

#### **Scoring**

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

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

Essentially correct (E) if the response provides a correct sampling procedure to obtain a simple random sample that includes sufficient information such that two knowledgeable statistics users would implement equivalent methods.

Partially correct (P) if the response provides a plausible sampling procedure for obtaining a simple random sample but does *NOT* include sufficient information as described for E.

Incorrect (I) if the response does not meet the criteria for E or P.

#### *Notes*

- If computer-generated random numbers are used, an explicit statement about ignoring repeats is needed for sufficient information. (Sampling until 100 students are obtained conveys the idea of ignoring repeats.)
- If a table of random numbers is used without adequate specification of the numbers to be used (for example, 0001 to 2,000 or a different set of 2,000 four-digit numbers), the response is scored as incorrect (I).
- If a table of random numbers is used with adequate specification of the numbers to be used, as described above, a statement about ignoring repeats *AND* a statement about ignoring numbers not in the specified range are both needed to earn an E. If either or both are missing, the response is scored as partially correct (P). (The statement about ignoring repeats or the statement about ignoring numbers not in the range of specification may be implicit. For example, "... continue this procedure until 100 students are selected.")
- The procedure of using names or numbers on slips of paper in a hat must indicate some randomization in selection (for example, mixing the slips of papers or randomly choosing from the hat).
- If the sampling procedure does not produce a simple random sample, then the response is scored as incorrect (I).

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

Essentially correct (E) if the correct calculation is performed *AND* sufficient work is shown.

Partially correct (P) if the correct answer is provided but no work is shown.

Incorrect (I) if the response does not meet the criteria for E or P.

#### *Notes*

- The correct formula (either in symbols or with correct numerical substitutions) is sufficient for shown work in either part (b) or part (c).
- A response with only the answer 0.413 with no work shown is scored as P (because dividing by the square root of 100 can be done mentally).

**AP<sup>®</sup> STATISTICS**  
**2012 SCORING GUIDELINES**

**Question 6 (continued)**

- Sufficient shown work with no calculation is scored as partially correct (P). For example, showing  $\frac{4.13}{\sqrt{100}}$  with no calculation.
- Minor calculation errors can be ignored in part (b) or part (c) and considered when determining between two scores.
- The incorrect use of the notation  $\sigma$  instead of  $s$  is considered a minor error in the investigative task and will not reduce the score in either part (b) or part (c).

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

Essentially correct (E) if the response includes the following three components:

1. Combining variances.
2. Using weights (0.6 for females and 0.4 for males).
3. Recognizing variability of sample means (variance divided by sample size),  
*AND*

the response correctly combines these three components to produce the estimated standard deviation for Rania's point estimator.

Partially correct (P) if the response has at least two of the three components *AND* a reasonable attempt to combine these components to produce the (overall) estimated standard deviation for Rania's point estimator.

Incorrect (I) if the response does not meet the criteria for E or P.

*Notes*

- When part (c) is scored partially correct (P), the reasonableness of the numerical result can be considered in determining the overall score.
- The following are algebraically equal and provide the correct estimated standard deviation of Rania's point estimator of 0.198:

$$\sqrt{(0.6)^2 \frac{s_f^2}{n_f} + (0.4)^2 \frac{s_m^2}{n_m}} = \sqrt{(0.6)^2 \frac{s_f^2}{60} + (0.4)^2 \frac{s_m^2}{40}} = \sqrt{(0.6) \frac{s_f^2}{100} + (0.4) \frac{s_m^2}{100}} = \sqrt{\frac{(0.6)s_f^2 + (0.4)s_m^2}{100}}$$

- A response that combines  $s_f^2$  and  $s_m^2$  to obtain a pooled standard deviation for Rania's data, such as

$$s_p = \sqrt{\frac{(n_f - 1)s_f^2 + (n_m - 1)s_m^2}{(n_f - 1) + (n_m - 1)}} = \sqrt{\frac{(59)(1.8)^2 + (39)(2.22)^2}{59 + 39}} \approx 1.97786,$$

contains two of the three components: combining variances and using weights

$\left(\frac{59}{98} \approx 0.6 \text{ and } \frac{39}{98} \approx 0.4\right)$  for part (c). Because pooling is a reasonable attempt to combine the components, the calculation of  $s_p \approx 1.97786$  is scored as partially correct (P). If the response

includes the correct third component,  $\left(\frac{s_p}{\sqrt{n}} \approx \frac{1.97786}{\sqrt{100}} = 0.197786\right)$ , the response is still scored

partially correct, because the pooling is a reasonable, but not the correct, combination of the three components to compute the sample standard deviation of the stratified sample mean.

**AP<sup>®</sup> STATISTICS**  
**2012 SCORING GUIDELINES**

**Question 6 (continued)**

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

Essentially correct (E) if a reasonable justification is provided based on *BOTH* of the following two components:

1. Smaller variability in responses for each gender in comparison with the variability in responses in Peter's data, and
2. Linkage (either explicit or implicit) between the smaller variability in responses for each gender and producing a smaller estimated standard deviation for Rania's combined point estimator than Peter's point estimator.

Partially correct (P) if the benefit of the smaller variability in responses for each gender with stratification (homogeneity within each gender) for these samples is identified, but there is no reference to how it affects the estimated standard error for Rania's point estimator,

*OR*

if the benefit of different centers with stratification (heterogeneity between genders) for the samples is identified, without reference to the resulting smaller variability in responses for each gender and how it affects the estimated standard error for Rania's point estimator,

*OR*

if the response identifies that Rania's combined data for males and females has smaller variability than Peter's data.

Incorrect (I) if the response does not meet the criteria for E or P.

*Notes*

- Comments about shapes of the distributions are extraneous and can be ignored.
- General statements about the benefits of stratification without using the sample data in the dotplots are not sufficient.

Each essentially correct (E) part counts as 1 point. Each partially correct (P) part 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 decide whether to score up or down, depending on the overall strength of the response and communication, especially in the investigative parts—part (c) and part (d)—of the response.