Question 5: Multi-Focus — 4 points

General Scoring Notes

- Each part of the question (indicated by a letter) is initially scored by determining if it meets the criteria for essentially correct (E), partially correct (P), or incorrect (I). The response is then categorized based on the scores assigned to each letter part and awarded an integer score between 0 and 4 (see the table at the end of the question).
- The model solution represents an ideal response to each section of the question, and the scoring criteria identify the specific components of the model solution that are used to determine the score.

	Model Solution	Scoring
(a)	$P(11+ \text{ months} \cap \text{majority regular cards})$ $= \frac{71 + 76 + 112}{500}$ $= \frac{259}{500}$ $= 0.518$	**Essentially correct (E)** if the response satisfies the following two components: 1. Provides the correct probability 2. Shows work for the correct probability **Partially correct (P)** if the response satisfies only one of the two components required for E. **Incorrect (I)** if the response does not meet the criteria for E or P.

Additional Notes:

- An arithmetic or transcription error in a response can be ignored if correct work is shown.
- A correct fraction, decimal, or percentage may satisfy component 1.
- To satisfy component 2 work must demonstrate how the numerator of 259 was computed. This may be accomplished by showing the addition or clearly indicating the appropriate summands on the table.
- A response that satisfies components 1 and 2 for the probability that the collector has been collecting baseball cards for 11 or more months and has a majority of *rare* baseball cards may be scored P.
- A specific probability statement is not required, but if correctly given should be considered a positive in holistic scoring.

Model Solution	Scoring
(b) $P(\text{majority regular cards} \mid \text{fewer than 6 months})$ $= \dfrac{P(\text{majority regular cards} \cap \text{fewer than 6 months})}{P(\text{fewer than 6 months})}$ $= \dfrac{\frac{80}{500}}{\frac{91}{500}}$ $= \dfrac{80}{91}$ $= 0.879$	**Essentially correct (E)** if the response satisfies the following two components: 1. Provides the correct probability 2. Shows work for the correct probability **Partially correct (P)** if the response satisfies only one of the two components required for E. **Incorrect (I)** if the response does not meet the criteria for E or P.

Additional Notes:

- An arithmetic or transcription error in a response can be ignored if correct work is shown.
- A response of $\frac{80}{91}$ satisfies both components 1 and 2.
- A response that satisfies components 1 and 2 for the probability that given a randomly selected collector has been collecting baseball cards for fewer than six months, they have a majority of *rare* baseball cards may be scored P.
- A specific probability statement is not required, but if correctly given should be considered a positive in holistic scoring.

Model Solution	Scoring
(c) (i) Michelle should conduct a chi-square test for independence between months collecting baseball cards and majority card status for all baseball card collectors at the convention. (ii) H_0: There is not an association between months collecting baseball cards and majority card status for all baseball card collectors at the convention. H_a: There is an association between months collecting baseball cards and majority card status for all baseball card collectors at the convention. *OR* H_0: Months collecting cards and majority card status are independent for all baseball card collectors at the convention. H_a: Months collecting cards and majority card status are not independent for all baseball card collectors at the convention.	**Essentially correct (E)** if the response satisfies the following three components: 1. In part (c-i) the response identifies a chi-square test for independence by name 2. In part (c-ii) the response states the correct null hypothesis to imply there is not an association (relationship) *AND* states the correct alternative hypothesis to imply there is an association (relationship) 3. In part (c-ii) the response provides sufficient context for at least one of the hypotheses by including reference to both "months collecting" and "majority card status" **Partially correct (P)** if the response satisfies only two of the three components required for E *OR* if the response identifies a "chi-square test for homogeneity" in part (c-i) by name or formula *AND* components 2 and 3 are consistent with the chi-square test for homogeneity. **Incorrect (I)** if the response does not meet the criteria for E or P.

Additional Notes

- A response that states merely "chi-square" or "chi-square test," without specifying independence, does not satisfy component 1. However, a response that states "chi-square test of association" does satisfy component 1.
- Any discussion of the degrees of freedom for the test should be ignored in scoring.
- If the hypotheses do not explicitly state the population of interest, it should be assumed that the population is all baseball card collectors at the convention. However, if the response clearly references the sample, component 2 is not satisfied.

Model Solution	Scoring
(d) Because the p-value of 0.0075 is less than any reasonable α level such as 0.05 or 0.10, the null hypothesis should be rejected. The data provide convincing statistical evidence that there is an association between months collecting baseball cards and majority card status for all baseball card collectors at the convention.	**Essentially correct (E)** if the response satisfies the following two components: 1. Provides correct comparison of the p-value to alpha (p-value is less than alpha) *AND* provides a correct decision about the null and/or alternative hypothesis 2. States a conclusion, in context, consistent with and in terms of the stated alternative hypothesis in part (c) using nondefinitive language **Partially correct (P)** if the response satisfies only one of the two components required for E. **Incorrect (I)** if the response does not meet the criteria for E or P.

Additional Notes

- Because no alpha value is explicitly given, a response may satisfy component 1 by stating that the p-value is small *AND* providing a correct decision about the null and/or alternative hypothesis.
- An explicit decision about the null hypothesis is not required to satisfy component 1.
- If an explicit decision is stated and the conclusion is inconsistent with the decision, component 1 is not satisfied.
- The decision part of component 1 may be satisfied by implying the decision within the conclusion statement (sufficient evidence/insufficient evidence for the alternative hypothesis).
- To satisfy the context in component 2 the response must include discussion of both "months spent collecting baseball cards" and "which type of card is the majority in the collection."
- If the response omits hypotheses in part (c), assume the correct alternative hypothesis is provided when scoring component 2.
- If the response states incorrect hypotheses in part (c), components 1 and 2 may be satisfied by either referring to the stated alternative hypothesis or referring to the inference question.
- Examples of nondefinitive language in component 2 include "evidence to accept the alternative," "there is evidence for the alternative," "there is not sufficient evidence for the alternative," and "Michelle should conclude that her belief is correct."
- Examples of definitive language in component 2 include "proves the null," "proves the alternative," "accepts the alternative," "there is not evidence for the alternative," "no evidence for the alternative," "therefore the alternative is true," and "there is an association."
- If components 1 and/or 2 are satisfied and the response provides an incorrect interpretation of the p-value, the score is lowered from E to P or P to I.

Scoring for Question 5

Each essentially correct (E) part counts as 1 point, and each partially correct (P) part counts as ½ point.

	Score
Complete Response	**4**
Substantial Response	**3**
Developing Response	**2**
Minimal Response	**1**

If a response is between two scores (for example, 2 ½ points), use a holistic approach to decide whether to score up or down, depending on the strength of the response and quality of the communication.

Question 5: Multi-Focus — 4 points

General Scoring Notes

- Each part of the question (indicated by a letter) is initially scored by determining if it meets the criteria for essentially correct (E), partially correct (P), or incorrect (I). The response is then categorized based on the scores assigned to each letter part and awarded an integer score between 0 and 4 (see the table at the end of the question).
- The model solution represents an ideal response to each part of the question, and the scoring criteria identify the specific components of the model solution that are used to determine the score.

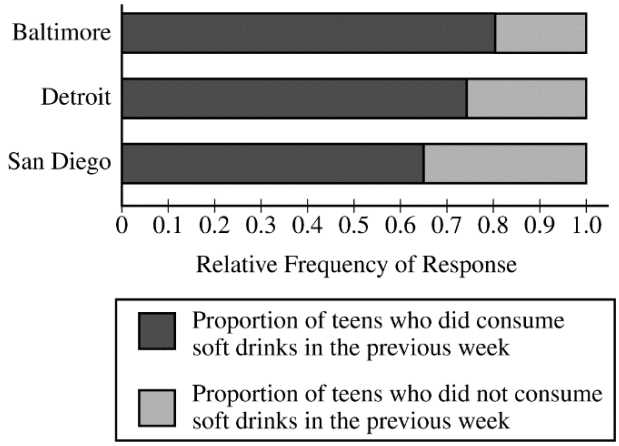
	Model Solution	Scoring
(a)	No, the researcher's claim is not correct. Although the Baltimore survey has the least number of teens who consumed a soft drink in the past week, it also has the least number of teens surveyed among the three cities' samples. Comparing the numbers of teens who consumed a soft drink in the past week is meaningless without considering the sample sizes. The comparison should be based on proportions rather than counts. In fact, the proportion of Baltimore teens who consumed a soft drink in the past week, $\frac{727}{904} \approx 0.804$, is larger than the proportions for either of the other two cities, $\frac{1,232}{1,663} \approx 0.741$ for Detroit and $\frac{1,482}{2,280} = 0.65$ for San Diego.	**Essentially correct (E)** if the response satisfies the following two components: 1. Indicates that the researcher's claim is not correct (or "may not be correct", if proportions are not reported) 2. Provides an explanation that is based on at least one of the following: • The proportions (or relative frequencies), not counts, should be compared because the sample sizes are not equal *OR* • The proportion of Baltimore teens who consumed a soft drink in the past week, $\frac{727}{904} \approx 0.804$, is larger than the proportion for at least one of the two other cities **Partially correct (P)** if the response satisfies component 1 AND states that the sample sizes are not equal *OR* if the response satisfies component 2 only *OR* if the response provides a correct proportion (or relative frequency) for Baltimore and at least one other city. **Incorrect (I)** if the response does not meet the criteria for E or P.

Additional Notes:

- A response that compares the proportion of Baltimore teens who consumed a soft drink in the past week to the combined proportion of the other two cities, $\frac{1,323+1,482}{1,663+2,280}=\frac{2,714}{3,943}\approx 0.688$ is not equivalent to comparing counts and may earn a P.
- If the "yes" count for each city is divided by the same number (e.g., the total number of respondents who consumed a soft drink in the past week, 3,441; or the total sample size, 4,847), then the response is equivalent to comparing counts and should be scored I.
- A response may satisfy component 2 by providing correct numerical values for the proportion of Baltimore teens who did *not* consume a soft drink in the past week, $\frac{177}{904}\approx 0.196,$ AND the proportion of teens who did *not* consume a soft drink in the past week for at least one of the two other cities, either $\frac{431}{1,663}\approx 0.259$ for Detroit and/or $\frac{798}{2,280}=0.35$ for San Diego.
- If work is shown, calculation or transcription errors should be ignored in scoring.
- Statistical notation should be ignored in scoring.

Model Solution	Scoring

(b) **(i)** A segmented bar graph of the relative frequencies based on the information in the table is shown below:



(ii) The proportion of teens who consumed a soft drink in the previous week are shown below:

- Baltimore: $\frac{727}{904} \approx 0.804$
- Detroit: $\frac{1,232}{1,663} \approx 0.741$
- San Diego: $\frac{1,482}{2,280} = 0.65$

San Diego has the smallest proportion of teens (0.65) who consumed a soft drink in the previous week.

Essentially correct (E) if the response satisfies the following four components:

1. Constructs a segmented bar graph in part (b-i), with the bars correctly segmented
2. Includes clear labeling of the proportions of teens who consumed a soft drink in the previous week and the proportions of teens who did not consume a soft drink in the previous week for the segmented bar graph provided in part (b-i)
3. Identifies San Diego as the city with the smallest proportion of teens who consumed a soft drink in the previous week in part (b-ii)
4. Reports the correct numerical value of the proportion of teens who consumed a soft drink in the previous week for the city identified in part (b-ii)

Partially correct (P) if the response satisfies only two or three of the four components.

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

Additional Notes:

- A response that constructs a segmented bar graph with the lengths of the segments representing the relative frequencies of teens who consumed a soft drink in the previous week between 0.75 and 0.85 for Baltimore, between 0.7 and 0.8 for Detroit, and between 0.6 and 0.7 for San Diego satisfies component 1.
- A response that constructs a segmented bar graph with the lengths of the segments representing the relative frequencies of teens who did not consume a soft drink in the previous week between 0.15 and 0.25 for Baltimore, between 0.2 and 0.3 for Detroit, and between 0.3 and 0.4 for San Diego satisfies component 1.
- Incorrect proportions imported from part (a) may be used to satisfy component 1.
- Segmented bar graphs with more than two segments cannot satisfy either component 1 or component 2.
- Labels of "Yes" and "No" may satisfy component 2.
- A response to part (b-ii) that is consistent with an incorrect graph in (b-i) may satisfy components 3 and 4.

Model Solution	Scoring
(c) **(i)** Since the data were collected from independent random samples from the three cities, a chi-square test for homogeneity should be conducted.	**Essentially correct (E)** if the response satisfies the following three components: 1. Identifies a chi-square test for homogeneity by name in part (c-i) 2. States the null hypothesis to imply homogeneous (or equal) proportions AND states the alternative hypothesis to imply that at least two proportions are not the same in part (c-ii) 3. Provides sufficient context for at least one of the hypotheses in part (c-ii) by including the parameters of interest (proportion of teens who consumed a soft drink) AND the populations (cities)
(ii) The appropriate hypotheses are: H_0: There is no difference in the proportion of all teens who consumed a soft drink in the past week across the three cities. H_a: There is at least one difference in the proportion of all teens who consumed a soft drink in the past week across the three cities. *OR* H_0: The proportion of all teens who consumed a soft drink in the past week is the same across the three cities. H_a: The proportion of all teens who consumed a soft drink in the past week differs for at least two of the three cities.	**Partially correct (P)** if the response satisfies component 1 and only one of the other two components *OR* if the response identifies a "chi-square test" in part (c-i) by name or formula AND satisfies component 2. **Incorrect (I)** if the response does not meet the criteria for E or P.

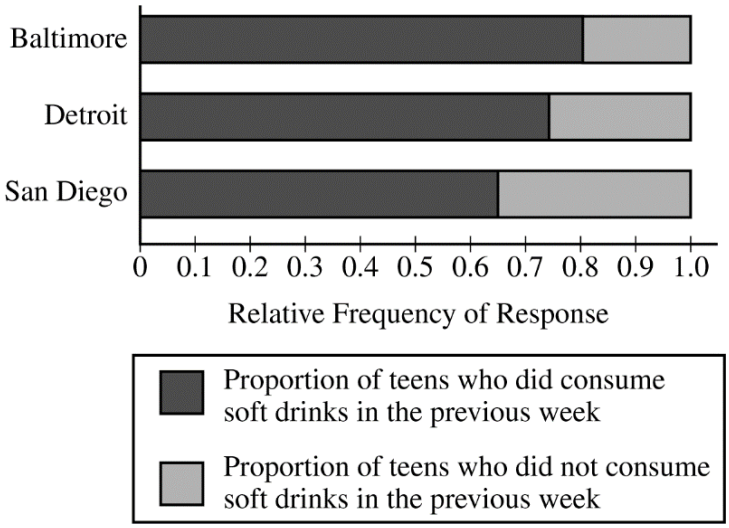
Additional Notes:

- A response that identifies two different tests is considered parallel solutions and the weaker solution is used when scoring component 1.
- Component 1 is not satisfied by the test statistic formula for a chi-square test unless the response includes "Homogeneity."
- Component 1 is not satisfied if the response presents a test statistic formula that is inconsistent with a chi-square test of homogeneity, even if the response identifies a chi-square test of homogeneity by name.
- A response that states the hypotheses in terms of distributions rather than proportions (e.g., H_0: There is no difference in distributions of teens who consumed or did not consume a soft drink in the past week across the three cities) satisfies component 2.
- A response that states either the null hypothesis or the alternative hypothesis by referring to sample proportions does not satisfy component 2.
- A response that uses symbols to describe the hypotheses must clearly identify the parameters in context (proportion of teens who consumed a soft drink) AND reference the populations (cities) in order to satisfy component 3.
- Any attempt to check test conditions should be ignored in scoring.
- Any discussion of the degrees of freedom for the test should be ignored in scoring.

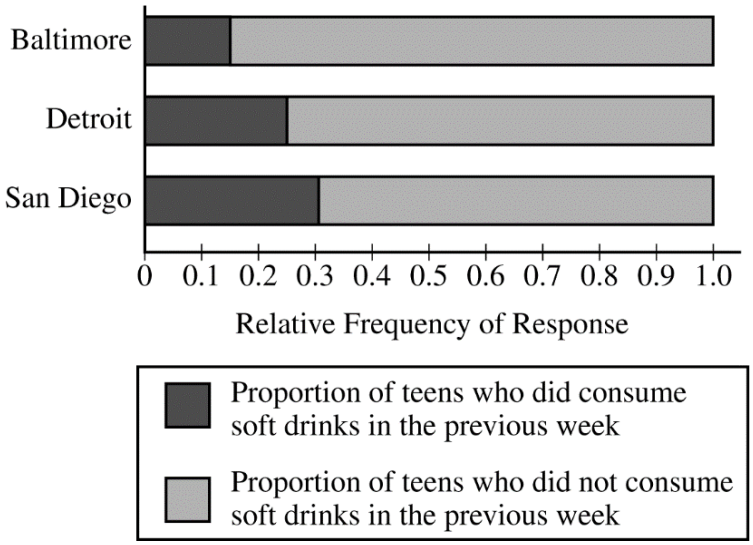
Scoring for Question 5	Score
Complete Response Three parts essentially correct	**4**
Substantial Response Two parts essentially correct and one part partially correct	**3**
Developing Response Two parts essentially correct and no part partially correct *OR* One part essentially correct and one or two parts partially correct *OR* Three parts partially correct	**2**
Minimal Response One part essentially correct and no part partially correct *OR* No part essentially correct and two parts partially correct	**1**

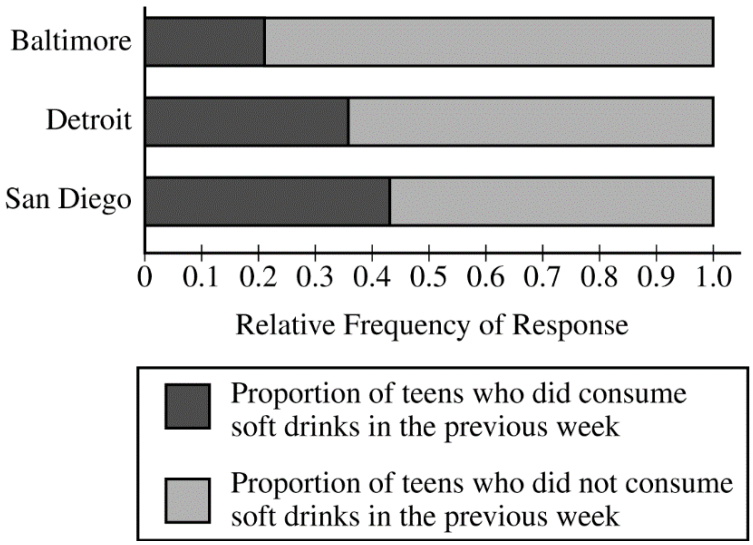
Common acceptable and unacceptable graphs for part (b-i)

Acceptable Graph

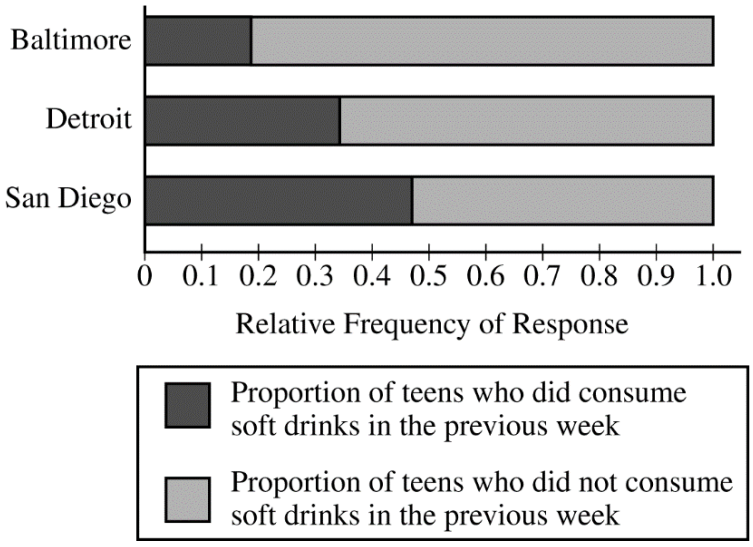


Common Unacceptable Graphs





Common Unacceptable Graphs (continued)



2017

AP CollegeBoard

AP Statistics

Scoring Guidelines

2017 SCORING GUIDELINES

Question 1

Intent of Question

The primary goals of this question were to assess a student's ability to (1) explain statistical terms used when describing the relationship between two variables; (2) interpret the slope of a linear regression equation; and (3) calculate a value of y when given a regression equation, a value of x, and a residual.

Solution

Part (a):

In the context of a scatterplot in which y represents weight and x represents length, the following are defined.

A positive relationship means that wolves with higher values of length also tend to have higher weights.

A linear relationship means that as length increases by one meter, weight tends to change by a constant amount, on average.

A strong relationship means that the data points fall close to a line (or curve).

Part (b):

The slope of 35.02 indicates that two wolves that differ by one meter in length are predicted to differ by 35.02 kilograms in weight, with the longer wolf having the greater weight.

Part (c):

In general, a residual is equal to actual weight minus predicted weight, or equivalently,

$$\text{actual weight} = \text{predicted weight} + \text{residual}.$$

For the wolf with length 1.4 meters and residual of -9.67, the predicted weight is

$$-16.46 + 35.02(1.4) = 32.568 \text{ kilograms}.$$

Therefore, the actual weight of the wolf is $32.568 + (-9.67) = 22.898$ kilograms.

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 includes the following four components:

1. A reasonable definition of positive
2. A reasonable definition of linear
3. A reasonable definition of strong
4. At least one definition in context

2017 SCORING GUIDELINES

Question 1 (continued)

Partially correct (P) if the response includes only three of the four components.

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

Notes:

- The description of a positive relationship should clearly indicate that relatively low values of one variable tend to appear with relatively low values of the other variable, and relatively high values of the first variable tend to appear with relatively high values of the other variable.
    - Examples of acceptable responses:
        - As length increases, so does weight.
        - Longer wolves weigh more.
        - The points on the graph go up as you move from left to right.
    - Examples of unacceptable responses:
        - As length goes up, weight changes.
        - Both length and weight get bigger.
        - The correlation is greater than 0.
- The description of a linear relationship can take one of two approaches: the data pattern (data points exhibit the pattern of a line in the graph) or the constant rate of change (as the explanatory variable changes, the response variable exhibits a constant rate of change).
    - Examples of acceptable responses:
        - The points generally follow a straight line.
        - The relationship between x and y is straight.
        - Length and width have a constant slope.
    - Examples of unacceptable responses:
        - The points all line up.
        - You can draw a straight line through the points.
        - There is a positive correlation.
        - Every increase in x yields a 35.02 increase in y.
- The description of strong should indicate how close points are to a line.
    - Examples of acceptable responses:
        - Observed values are close to predicted values.
        - Deviations from the least-squares regression line are small.
        - The correlation coefficient is close to 1.
    - Examples of unacceptable responses:
        - All the points are close together.
        - The scatterplots are clustered together.
        - There is a high positive correlation.
- Context can be shown by referring to length and weight or by using meters and kilograms.
- Sketches and graphs can be used to help clarify definitions, but a sketch alone cannot satisfy a definition component.

Part (b) is scored as follows:

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

1. The correct value of 35.02 for the slope.
2. An interpretation that includes an increase of a specified amount of weight for each unit increase in length.
3. An indication that the relationship is not exact by using words such as "on average" or "predicted weight."

2017 SCORING GUIDELINES

Question 1 (continued)

Partially correct (P) if the response includes only two of the three components.

Note: If the response identifies the slope as -16.46 (the intercept value), the second component is satisfied only if the response states that for each one-meter increase in length there is a *decrease* in predicted or average weight of 16.46 kilograms.

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

'art (c) is scored as follows:

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

1. A correct computation for the predicted value 32.568 kilograms.
2. A correct computation for the actual weight 22.9 kilograms using the given residual and the predicted value.

Partially correct (P) if the response provides a correct computation for the predicted value but is not able to complete the correct calculation of the actual weight, including if the residual is defined in the wrong direction as (predicted weight) – (actual weight) to give an answer of 42.24 kilograms;

OR

if the response provides an incorrect value for the predicted weight, but then uses that value correctly to determine the actual weight as (predicted weight) + residual = (predicted weight) $+(-9.67)$;

OR

if the response provides a correct answer for the actual weight but does not give sufficient information to determine how it was calculated.

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

Notes:

- The expression $-16.46 \quad 35.02(1.4)$ is enough to satisfy the first component.
- The equation $-16.46 + 35.02(1.4) - 9.67 = 22.9$ satisfies both components.
- Arithmetic mistakes are overlooked if they do not lead to an unreasonable answer (such as a negative value). For example, $32.568 + (-9.67) = 21.9$ satisfies the second component.

2017 SCORING GUIDELINES

Question 1 (continued)

4 Complete Response

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 parts 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

OR

No parts essentially correct and two parts partially correct

2017 SCORING GUIDELINES

Question 2

Intent of Question

The primary goals of this question were to assess a student's ability to (1) construct and interpret a confidence interval for a population proportion and (2) use a confidence interval for a proportion to find a confidence interval for a dollar amount that can be calculated using that proportion.

Solution

Part (a):

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

The appropriate procedure is a one-sample z-interval for a population proportion p. In this case, the population is all customers of the restaurant who ask for a water cup, and p is the proportion of that population who will fill the cup with a soft drink. The appropriate formula is $\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$.

Conditions:

1. Random sample
2. Large sample (number of successes $n\hat{p} \geq 10$ and number of failures $n(1-\hat{p}) \geq 10$)

For condition 1 the stem of the problem states that a random sample of customers who asked for a water cup was used.

For condition 2 the number of successes (filled cup with soft drink) is 23 and the number of failures is 57, both of which are greater than 10.

Step 2: Correct mechanics

The sample proportion is $\hat{p} = \frac{23}{80} = 0.2875$. The confidence interval is

$$0.2875 \pm 1.96\sqrt{\frac{0.2875(1 \quad 0.2875)}{80}}$$

$$= 0.2875 \pm 1.96(0.0506) \qquad \text{or } 0.1883 \text{ to } 0.3867.$$

$$= 0.2875 \pm 0.0992$$

Step 3: Interpretation

We can be 95 percent confident that in the population of all customers of the restaurant who ask for a water cup, the proportion who will fill it with a soft drink is between 0.1883 and 0.3867.

Part (b):

Using the confidence interval in part (a), a 95 percent interval estimate for the number of customers in June who asked for a water cup but then filled it with a soft drink is $3{,}000 \times 0.1883$ to $3{,}000 \times 0.3857$, or 565 to 1,160. At a cost of \$0.25 per customer, a 95 percent interval estimate for the cost to the restaurant in June is \$141.25 to \$290.00.

2017 SCORING GUIDELINES

Question 2 (continued)

Scoring

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

Section 1 is scored as follows:

Essentially correct (E) if the one-sample z-interval for a proportion is identified (either by name or formula) *AND* both conditions (random sampling and large sample) are adequately addressed.

Partially correct (P) if the response identifies the correct procedure *BUT* adequately addresses only one of the two required conditions;
OR
if the response does not identify the correct procedure *BUT* adequately addresses both required conditions.

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

Notes:

- Stating the large sample condition without verifying it is not sufficient. The response must use specific numerical values to adequately address the condition.
- If the response includes additional inappropriate conditions, such as $n \geq 30$ or requiring a normal population, then the response earns at most P for section 1.
- Stating and checking a condition about the size of the sample relative to the size of the population is appropriate but not required.
- Any statement of hypotheses, description of the population, or definition of the parameter should be considered extraneous. However, if such statements are included and incorrect, they are considered as poor communication in terms of holistic scoring.

Section 2 is scored as follows:

Essentially correct (E) if the response gives the correct 95 percent confidence interval. Supporting work is not required, but if included, it must be correct.

Partially correct (P) if the response gives a correct confidence interval with incorrect (but appropriate) supporting work shown;
OR
if the response gives an incorrect but reasonable confidence interval with appropriate supporting work shown — for instance, if a value other than 1.96 is used for the critical value or a value other than $\frac{23}{80}$ is used for $\hat{p}$.

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

Notes:

- Appropriate supporting working must have the form:
  proportion ± (critical value)(SE/SD of proportion).

2017 SCORING GUIDELINES

Question 2 (continued)

- A confidence interval that has both endpoints outside the interval from 0 to 1 is considered unreasonable.

Section 3 is scored as follows:

Essentially correct (E) if the response provides an appropriate interpretation of the interval that includes the following three components:

1. Conveys inference about a population proportion
2. Demonstrates a clear understanding that the parameter is the proportion of the water cup population that fills the cup with a soft drink
3. Mentions 95 percent confidence and interprets it correctly using words such as "We can be 95 percent confident" or "With 95 percent confidence"

Partially correct (P) if the response provides an appropriate interpretation of the interval that includes the first component *AND* only one of the other two components;
OR
if the response provides a correct interpretation of the confidence *level* in context without interpreting the specific interval.

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

Notes:

- Clear indication of an inference to the random sample of 80 customers rather than to the population does not satisfy the first component and is scored I.
- Stating values that are unrealistic as proportions or percentages (including blanks) in the interpretation lowers the score one level (from E to P, or from P to I).
- When both the interpretation and the level of the interval are given, only the interpretation is scored. If the interpretation of the confidence level is incorrect, it is considered as poor communication in terms of holistic scoring.
- Any interpretation that implies the interval has a 95 percent chance (or possibility or probability) of capturing the population proportion is scored I.

Section 4 is scored as follows:

Essentially correct (E) if the response gives a correct interval estimate for the cost to the restaurant *AND* shows enough work to indicate how the interval was found.

Partially correct (P) if the response gives the correct interval estimate for the *number* of customers (565 to 1,160) who would fill the water cup with soda, including showing work, but does not multiply by \$0.25 to find the interval for the cost;
OR
if the response gives the correct interval estimate for the *expected cost* to the restaurant for an individual who asked for a water cup (\$0.05 to \$0.10), including showing work, but does not multiply by 3,000 to find the interval for the total cost;
OR
if the response makes a reasonable attempt to find the endpoints of the interval as (0.1883)(3,000)(0.25) and (0.3867)(3,000)(0.25), but makes an error such as using the sample size of 80 instead of the population size of 3,000;

Question 2 (continued)

OR

if the response gives the correct interval without showing how it was found.

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

Notes:

- The response in section 4 earns an E if it follows correctly from an incorrect interval found in part (a), even if the interval is unreasonable.
- Units ($) are not required to earn an E.

Each essentially correct (E) section counts as 1 point, and a partially correct (P) section counts as ½ point.

4 Complete Response

3 Substantial Response

2 Developing Response

1 Minimal Response

If a response is between two scores (for example, 2½ points), use a holistic approach to decide whether to score up or down, depending on the strength of the response and communication.

2017 SCORING GUIDELINES

Question 3

Intent of Question

The primary goals of this question were to assess a student's ability to (1) calculate a probability from a normal distribution; (2) calculate a weighted probability from two individual probabilities; and (3) calculate a conditional probability for dependent events when individual and joint probabilities are provided.

Solution

Part (a):

Let X denote the diameter of a randomly selected melon from Distributor J. X has an approximately normal distribution with mean 133 mm and standard deviation 5 mm.

The z-score for a diameter of 137 mm is $z = \dfrac{137 - 133}{5} = \dfrac{4}{5} = 0.8$.

Therefore, $P(X > 137) = P(Z > 0.8) = 1 - 0.7881 = 0.2119$.

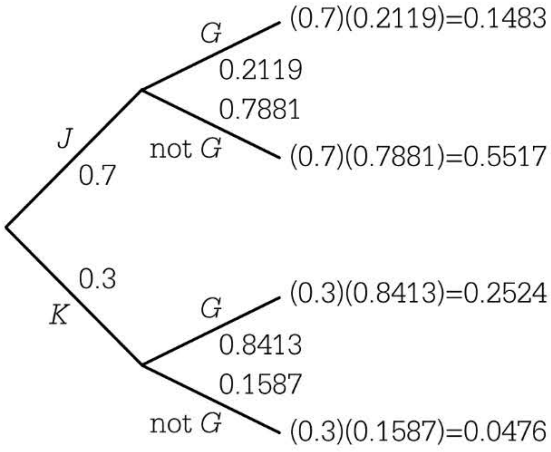
Part (b):

Define events:
J: melon is from Distributor J
K: melon is from Distributor K
G: melon diameter is greater than 137 mm

For a randomly selected melon from the grocery store,

$$\begin{aligned} P(G) &= P(G \mid J) \times P(J) + P(G \mid K) \times P(K) \\ &= (0.2119)(0.7) + (0.8413)(0.3) \\ &= 0.1483 + 0.2524 \\ &= 0.4007 \end{aligned}$$

OR



From the tree diagram, $P(G) = P(G \text{ and } J) + P(G \text{ and } K) = 0.1483 + 0.2524 = 0.4007$.

2017 SCORING GUIDELINES

Question 3 (continued)

Part (c):

Using the events defined in part (b), the requested probability is

$$P(J \mid G) = \frac{P(J \text{ and } G)}{P(G)} = \frac{P(G \mid J)P(J)}{P(G)} = \frac{(0.2119)(0.7)}{0.4007} = \frac{0.1483}{0.4007} = 0.3701.$$

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 provides the following three components:

1. Normality and parameters: Indicates use of a normal (or approximately normal) distribution and clearly identifies the correct parameter values. (Showing correct components in a z-score, labeling the mean and standard deviation in a normalcdf calculator statement, or drawing a normal curve with sufficient values marked to indicate the mean and standard deviation are all sufficient methods for identifying the correct parameter values.)
2. Boundary and direction: Uses the correct boundary value of $x = 137$ or $z = 0.8$ and the correct direction. (Showing correct boundary and direction in a probability statement using "greater than," labeling the lower limit and upper limit in a normalcdf calculator statement, drawing a normal curve with 137 labeled and an indication that the area of interest is to the right of 137, or a conclusion in words using "greater than.")
3. Probability: Reports the correct normal probability consistent with the response's setup described in components 1 and 2.

Partially correct (P) if the response correctly provides only two of the three components.

Incorrect (I) if the response gives the answer with no work shown or otherwise does not satisfy the criteria for E or P.

Notes:

- An inconsistency in calculations lowers the score for part (a) by one level (from E to P, or from P to I). For instance, if the response states the requested probability as $P(X > 137)$, but actually computes $P(X < 137)$.
- An error in statistical notation, such as using s instead of σ for the standard deviation or $\overline{x}$ instead of μ for the mean, does not satisfy component 1.
- A correct direction can be obtained by using a left direction and an upper bound of 137 and then subtracting the value of the cumulative probability from 1. However, simply showing the calculation $1 - 0.7881 = 0.2119$ does NOT give a correct direction.
- Consider any steps of a hypothesis test as extraneous work.
- Standard notations such as N(133,5) or N(133,25) satisfy the first component.
- A sketch of the normal curve with 133 and 138 in the appropriate positions satisfies the first component.
- Use of a value other than 137 does not satisfy the second component.
- If the only error in part (a) is the reversal of the numerator for the z-score $(133 - 137)$, the response is scored as P.

2017 SCORING GUIDELINES

Question 3 (continued)

- For calculator notation:
  - ncdf is sufficient to indicate use of the normal distribution.
  - Stating the lower and upper bounds is sufficient for identifying boundary and direction.
  - Minimum notation, such as "u" for upper, is sufficient for identifying boundary.
  - The minimum value for an upper bound must be at least 4 standard deviations above the mean.

Part (b) is scored as follows:

Essentially correct (E) if the probability is computed correctly, *AND* work is shown that includes correct numerical values using a formula, end results from a tree diagram, or some other appropriate method.

Partially correct (P) if the response provides a correct strategy for finding the probability, such as a formula or tree diagram, but uses one or more incorrect values or does not use the end result from the tree;
OR
if the response gives the correct probability but not enough work is shown to determine how it was found.

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

Notes:

- A response that uses 0.3 for J and 0.7 for K earns a P if no other mistakes are made.
- Tree diagrams:
  - If used, the multiplication is implied, but the addition must be indicated.
  - If not used, both the multiplication and the addition must be indicated.
- The shown work might be found in part (c).
- Probabilities can be shown in a table. Addition is implied if "Total" is labeled, as shown.

Intersection	G	Not G	Total
J	0.1483	0.5517	0.7
K	0.2524	0.0576	0.3
Total	0.4007	0.5993	

Part (c) is scored as follows:

Essentially correct (E) if the probability is computed correctly *AND* work is shown that illustrates how the probability was found.

Partially correct (P) if the response provides a strategy for finding the probability for dependent variables, such as the appropriate formula or end results from the correct tree diagram, but uses one or more incorrect values or does not use the end result from the tree;
OR
if the response gives the correct probability but not enough work is shown to determine how it was found.

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

2017 SCORING GUIDELINES

Question 3 (continued)

Notes:

- Part (c) earns an E if a correct answer follows from incorrect values used from part (a) and/or part (b), provided all values are between 0 and 1, inclusive.
- A transcription error in a response can be ignored if correct work is shown.
- For any part, if the resulting probability or part of the calculation of the probability uses a value that is not between 0 and 1, inclusive, the score is lowered by one level (from E to P, or from P to I).
- In any part, unsupported answers are scored as I.

4 Complete Response

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 parts 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

OR

No parts essentially correct and two parts partially correct

2017 SCORING GUIDELINES

Question 4

Intent of Question

The primary goals of this question were to assess a student's ability to use boxplots to (1) compare multiple sets of data; (2) identify which set of data is most likely to have produced a particular summary value; and (3) determine which variable is most useful for classifying a new observation.

Solution

Part (a):

The median value for the percent of chemical Z in the pottery pieces is similar for all three sites, at about 7 percent. The ranges for the percent of chemical Z are much different for the three sites, with the smallest range of about 2 percent (from 6 percent to 8 percent) at Site II, a range of about 6 percent (from about 4 percent to 10 percent) at Site I, and the largest range of about 8 percent (from about 3 percent to 11 percent) at Site III.

Part (b):

(i) The piece most likely originated at Site III. Although values outside of the range of data observed in the samples would be possible, using the available data results in approximate minimum and maximum sums of the percents for the three chemicals as shown in the table below. Site III is the only site in which 20.5 falls between the sums of the minimum and maximum values.

	Site I		Site II		Site III	
Chemical	Min	Max	Min	Max	Min	Max
X	6	8	5	7	5	7.5
Y	11	15	1.9	4	6	8
Z	4	10	6	8	3	11
Sum	21	33	12.9	19	14	26.5

(ii) Chemical Y would be most useful, because the distribution of the percentages of total weights at the three sites do not overlap. The distributions of chemicals X and Z have substantial overlap.

Scoring

This question is scored in three sections. Section 1 consists of part (a), section 2 consists of part (b-i), and section 3 consists of part (b-ii). Each section is scored as essentially correct (E), partially correct (P), or incorrect (I).

Section 1 is scored as follows:

Essentially correct (E) if the response includes the following three components:
1. Recognition that the medians or centers are almost the same for the three sites
2. Recognition that the variability (ranges, IQRs, spread) is different across the three sites
3. Context is included

Partially correct (P) if the response includes only two of the three components.

Incorrect (I) if the response includes at most one of the three components.

2017 SCORING GUIDELINES

Question 4 (continued)

Notes:

- In all sections, comments about shape should be ignored because complete shape information is not obtainable from boxplots.
- Responses are not required to give numerical values. If responses provide numerical values, any reasonable approximation from the boxplots is acceptable.
- Because the boxplots are all symmetric, it is acceptable if the response discusses means instead of medians.
- Any discussion of chemical X and chemical Y is considered extraneous.
- Context is satisfied by any of the following references: site, chemical, weight, total weight, X, Y, or Z.

Section 2 is scored as follows:

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

1. Site III is chosen.
2. Sums of the minimum and maximum are computed for the three chemicals at each site.
3. A reasonable numerical justification is given involving sums of a statistical measure across the three chemicals to choose Site III.

Partially correct (P) if the response includes only two of the three components.

Incorrect (I) if the response includes at most one of the three components.

Notes:

- If the response computes only the sum of the minimums for Site I and the sum of the maximums for Site II and recognizes that this is sufficient, the response is scored E.
- If an alternative measure is used that involves sums of the three chemicals, such as the sum of the medians or the sums of the first quartiles and sums of the third quartiles, instead of the minimum and maximum sums, the second component is not satisfied, but the third component might be satisfied.
    - If the response explicitly or implicitly compares the alternate sum to the other two sites (for example, by indicating that the sum is the closest to 20.5 percent or by listing the sums for all three sites) the response is scored P.
    - If the response does not have an implicit or explicit comparison, the response is scored I.
- If either Site I or Site II is identified as the correct choice, no matter how that choice is justified, the response is scored I.
- The approximate sums of the medians are 27.5 for Site I, 16 for Site II, and 20 for Site III.

Section 3 is scored as follows:

Essentially correct (E) if the response chooses chemical Y *AND* gives a reasonable justification based on the fact that the distributions of chemical Y are distinctive across sites.

Partially correct (P) if the response chooses chemical Y *AND* provides justification based on the boxplots, but does not clearly explain that the distributions of chemical Y are distinctive across sites;
OR
if the response correctly discusses that the distributions of chemical Y are distinctive across sites, but never explicitly chooses chemical Y as the best choice, for instance, by stating only that there is substantial overlap across sites for chemicals X and Z but no overlap for chemical Y.

2017 SCORING GUIDELINES

Question 4 (continued)

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

Notes:

- To justify that the distributions of chemical Y are distinctive across sites, the justification must address both location and variability of the boxplots; for example, by stating that the boxplots do not overlap for chemical Y.
- If the response chooses chemical X or Z *OR* chooses chemical Y with no reasonable justification, the response is scored I.
- The justification that the distributions of chemical Y are distinctive across sites:
    - The following are acceptable because both location and variability are addressed. Such responses are scored E.
        - The boxplots for chemical Y do not overlap, or the boxplots for chemicals X and Z overlap.
        - All values of Site I are high, all values of Site II are low, and all values of Site III are in the middle.
        - The ranges never intersect.
        - The boxplots share no data.
        - Has completely different percentages at each site.
    - The following are incomplete justifications and are scored P.
        - The boxplots vary.
        - Chemical Y varies the most.
        - Chemical Y has the greatest variation.
        - The variation between/among sites is the largest.
        - The boxplots are different.
        - The medians/means differ.
        - The medians/means are most variable.
        - There is a difference in the percentages of chemical Y for each site.
        - The distribution of percents differs the most among the sites.

2017 SCORING GUIDELINES

Question 4 (continued)

4 Complete Response

Three sections essentially correct

3 Substantial Response

Two sections essentially correct and one section partially correct

2 Developing Response

Two sections essentially correct and no sections partially correct

OR

One section essentially correct and one or two sections partially correct

OR

Three sections partially correct

1 Minimal Response

One section essentially correct

OR

No sections essentially correct and two sections partially correct

2017 SCORING GUIDELINES

Question 5

Intent of Question

The primary goal of this question was to assess a student's 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 a student's ability to (1) state appropriate hypotheses; (2) identify the appropriate statistical test procedure and check appropriate conditions for inference; (3) calculate the appropriate test statistic and *p*-value; and (4) draw an appropriate conclusion, with justification, in the context of the study.

Solution

Step 1: State a correct pair of hypotheses.

The null hypothesis is that age group at diagnosis and gender are independent (that is, they are not associated) for the population of people currently being treated for schizophrenia.

The alternative hypothesis is that age group at diagnosis and gender are not independent for the population of people currently being treated for schizophrenia.

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

The appropriate test is a chi-square test of independence.

The conditions for this test are satisfied because:

1. The question states that the sample was randomly selected.
2. The expected counts for the eight cells of the table are at least 5, as seen in the following table, with expected counts shown below observed counts.

	Age at Diagnosis				
	20 to 29	30 to 39	40 to 49	50 to 59	Total
Women	46	40	21	12	119
	56.91	36.22	17.25	8.62	
Men	53	23	9	3	88
	42.09	26.78	12.75	6.38	

Step 3: Find the value of the test statistic and the *p*-value.

The test statistic is calculated as $\chi^2 = \sum \frac{(O-E)^2}{E}$, or

$$\begin{aligned}\chi^2 &= 2.093 + 0.395 + 0.817 + 1.322 \\ &\quad + 2.830 + 0.534 + 1.105 + 1.788 \\ &= 10.884.\end{aligned}$$

The *p*-value is $P(\chi^2 \geq 10.884) = 0.012$, based on $(4-1) \times (2-1) = 3$ degrees of freedom.

2017 SCORING GUIDELINES

Question 5 (continued)

Step 4: State the conclusion in context, with linkage to the p-value.

Because the p-value is very small (for instance much smaller than $\alpha = 0.05$), we reject the null hypothesis and conclude that the sample data provide strong evidence that there is an association between age group at diagnosis and gender for the population currently being treated for schizophrenia.

Scoring

This question is scored in three sections. Section 1 consists of steps 1 and 2 (stating the correct hypotheses, identifying the appropriate test procedure, and checking the technical conditions); section 2 consists of step 3 (performing the correct mechanics); and section 3 consists of step 4 (stating a correct conclusion with justification). 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 the response correctly includes the following three components:
1. States BOTH hypotheses correctly with context included in at least one of them
2. Identifies a chi-square test of independence by name or formula
3. Verifies appropriate conditions that minimally include the condition for the expected counts and do not include any incorrect conditions (such as normality)

Partially correct (P) if the response includes only two of the three components.

Incorrect (I) if the response includes at most one of the three components.

Notes:
- Stating the expected count condition is not sufficient; the condition must be checked by reporting the expected counts, or minimally by showing that the smallest expected count is at least 5.
- The random sample condition was stated in the stem so need not be explicitly checked.
- If the null and alternative hypothesis are correctly stated in terms of population proportions, component 1 is satisfied. For example:
  $H_0 : p_1 = p_2 = p_3 = p_4$, where p_i is the population proportion of women at each indicated age group, 1, 2, 3, or 4, who are currently being treated for schizophrenia.
  H_a : At least one of the population proportions, p_1, p_2, p_3, p_4, differs from the other three.
  OR
  H_a : The population proportions for the four age groups are not all the same.

Section 2 is scored as follows.

Essentially correct (E) if the response correctly calculates the following two values:
1. The value of the chi-square test statistic
2. The p-value, critical value, or p-value range from chi-square table

Partially correct (P) if the response correctly calculates only one of the two values.

2017 SCORING GUIDELINES

Question 5 (continued)

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

Notes:

- If the response makes an error in one calculation, subsequent calculations are considered correct if they follow correctly from the initial miscalculation.
- With 3 degrees of freedom, the correct critical value is 7.81 for a significance level of 0.05 and 11.34 for a significance level of 0.01.
- Work does not have to be shown for calculations of test statistic or p-value. However, if incorrect work (other than minor arithmetic/transcription errors) is shown it is considered to be an incorrect calculation of the respective component, even if the correct value is given.
- If a response provides a test statistic that is not a chi-square test statistic, section 2 is scored I.

Section 3 is scored as follows.

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

1. A correct conclusion about the alternative hypothesis.
2. Justification of the conclusion based on linkage between the p-value and a reasonable alpha (or linkage between test statistic and critical value).
3. The conclusion is stated in context.

Partially correct (P) if the response includes only two of the three components.

Incorrect (I) if the response includes at most one of the three components.

Notes:

- If the response provides a correct decision, in context, with linkage to the p-value, but the decision is stated in terms of the null hypothesis with no conclusion about the alternative hypothesis, component 1 is not satisfied.
- Incorrect statistical statements are considered incorrect conclusions for the hypothesis test and do not satisfy component 1.
- If the conclusion is consistent with the p-value from section 2, and also in context with justification based on linkage to the p-value, section 3 is scored E.
- If no alpha level is given, the solution must be explicit about the linkage by giving a correct interpretation of the p-value or explaining how the conclusion follows from the p-value. For example, stating that because the p-value is small, we reject the null hypothesis or stating that because the p-value is large, we do not reject the null hypothesis.
- A decision about the null hypothesis (reject H_0 or fail to reject H_0) is not required, but if such a statement is given the scoring of the decision is considered in component 2.

2017 SCORING GUIDELINES

Question 5 (continued)

4 Complete Response

Three sections essentially correct

3 Substantial Response

Two sections essentially correct and one section partially correct

2 Developing Response

Two sections essentially correct and no sections partially correct

OR

One section essentially correct and one or two sections partially correct

OR

Three sections partially correct

1 Minimal Response

One section essentially correct

OR

No sections essentially correct and two sections partially correct

OR

Section 1 partially correct and the other two sections incorrect

2017 SCORING GUIDELINES

Question 6

Intent of Question

The primary goals of this question were to assess a student's ability to (1) calculate probabilities associated with treatment and control group memberships for two different methods of random assignment and (2) justify which method of random assignment is more appropriate in a given situation.

Solution

Part (a):

(i) Let T (tail) represent being assigned to the treatment group and H (head) represent being assigned to the control group for each coin flip. The process stops when either the treatment group or the control group has two members. The outcomes and their probabilities are as follows.

Arrangement	A	B	C	D	E	F
Chip outcomes	TT	THT	THH	HH	HTH	HTT
Calculation	$\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)$	$\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)$	$\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)$	$\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)$	$\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)$	$\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)$
Probability	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8}$

(ii) Man 1 and Man 2 are assigned to the same group for arrangements A and D, so the probability is

$$P(\text{A}) + P(\text{D}) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}.$$

Part (b):

(i) Let T represent being assigned to the treatment group and C represent being assigned to the control group for each chip drawn. The process stops when either the treatment group or the control group has two members. The probabilities differ from the coin flip method because chips are drawn *without* replacement. The outcomes and their probabilities are as follows.

Arrangement	A	B	C	D	E	F
Chip outcomes	TT	TCT	TCC	CC	CTC	CTT
Calculation	$\left(\frac{2}{4}\right)\left(\frac{1}{3}\right)$	$\left(\frac{2}{4}\right)\left(\frac{2}{3}\right)\left(\frac{1}{2}\right)$	$\left(\frac{2}{4}\right)\left(\frac{2}{3}\right)\left(\frac{1}{2}\right)$	$\left(\frac{2}{4}\right)\left(\frac{1}{3}\right)$	$\left(\frac{2}{4}\right)\left(\frac{2}{3}\right)\left(\frac{1}{2}\right)$	$\left(\frac{2}{4}\right)\left(\frac{2}{3}\right)\left(\frac{1}{2}\right)$
Probability	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

(ii) Man 1 and Man 2 are assigned to the same group for arrangements A and D, so the probability is

$$P(\text{A}) + P(\text{D}) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}.$$

2017 SCORING GUIDELINES

Question 6 (continued)

Part (c):

Use the chip method. The chip method gives equal probability to all possible arrangements, but the coin method does not, as shown in the tables from parts (a-i) and (b-i). Furthermore, the coin method is more likely to result in imbalanced treatment groups with regard to students and teachers, based on the probabilities in parts (a-ii) and (b-ii). If food preferences for teachers are different than for students, the imbalance is a problem. For example, if one treatment group consists entirely of students, it would be impossible to know if a difference in the response variable is due to the treatment (type of meal) or the role of the person at the school (teacher or student).

Scoring

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

Part (a) is scored as follows:

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

1. The correct probability for each arrangement in the table of (i).
2. Appropriate justification is shown in (i).
3. The correct probability is given in (ii).

Partially correct (P) if the response includes only two of the three components.

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

Notes:

- For component 2, examples of appropriate justification include the following:
  - Shows calculation of a correct probability for at least one of arrangements A and D *AND* shows calculation of a correct probability for at least one of arrangements B, C, E, and F. Calculations must be linked to a specific arrangement with a label, by proximity to an arrangement in the table, or by following the structure of the table or display of arrangements.
  - Lists correct coin outcomes for at least one of arrangements A and D *AND* lists correct coin outcomes for at least one of arrangements B, C, E, and F (for example, A: TT and B: THT). Lists of outcomes must be linked to a specific arrangement with a label, by proximity to an arrangement in the table, or by following the structure of the table or display of arrangements.
  - Shows a tree diagram that ends with six branches corresponding to the six arrangements. Probabilities or coin outcomes must be listed on the branches, but arrangements do not need to be identified.
  - Shows a tree diagram that ends with eight (or 16) branches where the branches leading to at least one of arrangements A and D are identified, and the branches leading to at least one of arrangements B, C, E, and F are identified. Probabilities or coin outcomes must be listed on the branches.
- If an incorrect probability is given in part (ii), component 3 is satisfied if the probability is between 0 and 1 and:
  - is the result of a minor arithmetic error; *OR*
  - is the result of adding the probabilities of arrangements A and D from the table in (i).

2017 SCORING GUIDELINES

Question 6 (continued)

Part (b) is scored as follows:

Essentially correct (E) if the response includes the following two components:
1. The correct probability for each arrangement in the table of (i).
2. The correct probability is given in (ii).

Partially correct (P) if the response includes only one of the three components.

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

Note: If an incorrect probability is given in (ii), component 2 is satisfied if the probability is between 0 and 1 and

- is the result of a minor arithmetic error; *OR*
- is the result of adding the probabilities of arrangements A and D from the table in (i); *OR*
- work is shown that demonstrates the same incorrect approach is used in both parts (a-ii) and (b-ii). For example, component 2 is satisfied if the response forgets to include $P(\text{D})$ and says $P(\text{A}) = \frac{1}{4}$ in (a-ii) and $P(\text{A}) = \frac{1}{6}$ in part (b-ii). However, unlabeled answers of $\frac{1}{4}$ in (a-ii) and $\frac{1}{6}$ in (b-ii) would not satisfy component 2 because it is unclear if the same approach is being used.

Part (c) is scored as follows:

Essentially correct (E) if the response chooses the chip method AND includes the following three components:
1. Provides a statistical benefit to the chip method. For example, stating that all arrangements are equally likely.
2. Provides a statistical drawback to the coin method. For example, stating that the coin method is more likely to result in imbalanced treatment groups.
3. States that the responses (opinions, food preferences) of teachers and students might be different.

Partially correct if the response chooses the chip method and includes only one or two of the three components.

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

Notes:
- A correct comparative statement can be used to satisfy both components 1 and 2. For example, stating that the coin method is *more likely* to result in a disproportionate number of students in one treatment group, or that the chip method has a *greater chance* of providing roughly equivalent treatment groups.
- Benefits and drawbacks must be about the probabilities of the arrangements (from the tables in (i)), or the probabilities of imbalanced (balanced) groups (from the questions in (ii)), or both.
- If the response chooses the coin method, look back to see if the coin method should be the preferred method based on incorrect work from parts (a) and (b). If so, score part (c) using the corresponding three components.

2017 SCORING GUIDELINES

Question 6 (continued)

- If the response says that it does not matter which method to use, look back to see if the tables are the same in parts (a-i) and (b-i) or if the probabilities are the same in parts (a-ii) and (b-ii). If so, part (c) is partially correct if the response justifies the decision by stating that the two tables or the two probabilities are the same. To be essentially correct, the response also needs to satisfy component 3.
- If the response does not choose a method, deduct one component from the number of correct components.

4 Complete Response

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 parts 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

OR

No parts essentially correct and two parts partially correct

AP® Statistics
2016 Scoring Guidelines

SCORING GUIDELINES

Question

Intent of Question

The primary goals of this question were to assess a student's ability to (1) describe the distribution of a quantitative variable based on a histogram and (2) determine the effect of changing one data value on mean and the median.

Solution

Part (a):

The distribution of Robin's tip amounts is skewed to the right. There is a gap between the largest amount (in the \$20 to \$22.50 interval) and the second largest tip amount (in the \$12.50 to \$15 interval), and the largest tip amount appears to be an outlier. The median tip amount is between \$2.50 and \$5.00. Robin's tip amounts vary from a minimum of between \$0 and \$2.50 to a maximum of between \$20.00 and \$22.50. About 78 percent of the tip amounts are between \$0 and \$5.

Part (b):

ɘ mean: If the \$8 tip had been \$18, the mean would increase by \$10 divided by 60, or $\$\frac{1}{6}$, or about 17 cents.

ɘ median: If the \$8 tip had been \$18, the median would not change because the current median is between \$2.50 and \$5.00, and both \$8 and \$18 are greater than that.

Scoring

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

Part (a) is scored as follows:

Essentially correct (E) if the response includes reasonable comments on the following five components:

1. Shape (skewed right)
2. Outlier (at least one) *OR* gap (one tip amount greater than \$20, next highest at most \$15)
3. Center between \$2.50 and \$5.00 (median) or between \$2.62 and \$5.13 (mean)
4. Variability, by noting that the tip amounts vary from about \$0 to at most \$22.50, or that a majority of tip amounts are between \$0 and a value greater than or equal to \$5, or by providing a correct numerical approximation of a measure of variability
5. Context (tip amounts)

Partially correct (P) if the response includes only three or four of the five components.

Incorrect (I) if the response includes at most two of the five components.

SCORING GUIDELINES

Question 1 (continued)

Part (b) is scored as follows:

Essentially correct (E) if the response includes the following four components:
1. Comments that the mean will increase
2. Correctly justifies why the mean will increase
3. Comments that the median will not change
4. Correctly justifies why the median will not change

Partially correct (P) if the response includes only two or three of the four components.

Incorrect (I) if the response includes at most one of the four components.

4 Complete Response

Both parts essentially correct

3 Substantial Response

One part essentially correct and one part partially correct

2 Developing Response

One part essentially correct and one part incorrect

OR

Both parts partially correct

1 Minimal Response

One part partially correct and one part incorrect

SCORING GUIDELINES

Question

Intent of Question

The primary goals of this question were to assess a student's ability to (1) identify, set up, perform, and interpret the results of an appropriate hypothesis test to address a particular question and (2) assess the effectiveness of treatments in a controlled experiment.

Solution

Part (a):

Step 1: States a correct pair of hypotheses.

H_0 : The proportion of children who would choose each snack is the same regardless of which type of ad is viewed.
H_a : The proportion of children who would choose each snack differs based on which type of ad is viewed.

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

The appropriate procedure is a chi-square test of homogeneity.

The conditions for this test are satisfied because (1) the question states that the children were randomly assigned to groups, and (2) expected counts for the six cells of the table are all at least 5, as seen in the following table that lists expected counts beside observed counts.

Group	Choco-Zuties	Apple-Zuties	Total
A	21 (18.67)	4 (6.33)	25
B	13 (18.67)	12 (6.33)	25
C	22 (18.67)	3 (6.33)	25
Total	56	19	75

Step 3: Calculates the appropriate test statistic and p-value.

The test statistic is calculated as $\chi^2 = \sum \frac{(O-E)^2}{E}$, which is

$$\chi^2 \approx 0.292 + 0.860 + 1.720 + 5.070 + 0.595 + 1.754 \approx 10.291.$$

The p-value is $P(\chi^2_{df2} \geq 10.291) \approx 0.006$.

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

Because the value is very small (for instance, much smaller than $\alpha = 0.05$ we reject the null hypothesis at the 05 level (and at the 01 level). The data provide convincing statistical evidence that the proportions who would choose each snack differ based on which ad is viewed.

SCORING GUIDELINES

Question 2 (continued)

Part (b):

When neither ad was viewed, $\frac{22}{25}$ or 88 percent of the children chose Choco-Zuties whereas only percent chose Apple-Zuties.

When the Choco-Zuties ad was viewed, 84 percent of the children chose Choco-Zuties, which was very similar to the percentage that chose them without viewing any ad. So watching the Choco Zuties ad did not affect the snack choice very much.

n the Apple-Zuties ad was viewed, only $\frac{13}{25}$ or 52 percent of the children chose Choco-Zuties, and percent chose Apple-Zuties. Watching the Apple-Zuties ad seemed to increase the proportion of children choosing Apple-Zuties.

Scoring

This question is scored in four sections. Section 1 consists of steps 1 and 2 in part (a); section 2 consists of step 3 in part (a); section 3 consists of step 4 in part (a); and section 4 consists of part (b). Sections 1, 2, 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 includes the following three components:
1. Both hypotheses are stated correctly with at least one in context.
2. Identifies the correct test procedure (by name or by formula).
3. The technical conditions are checked (all expected counts are greater than or equal to 5).

Partially correct (P) if the response includes only two of the three components.

Incorrect (I) if the response correctly includes at most one of the three components.

Notes for component 1:
- It is acceptable if hypotheses are stated in terms of association/independence instead of proportions.
- Context must be included in at least one of the hypotheses, but is not required to be in both.
- If the hypotheses contain language that suggests that the response refers to the sample data, the component is not satisfied.

Note for component 2: It is acceptable if the test is identified as a test of independence instead of a test of homogeneity.

Notes for component 3:
- The random assignment condition was stated so need not be explicitly mentioned.
- Stating the expected count condition is met is not sufficient for component 3. The condition must be checked by reporting the expected counts and either:
    - noting that all are greater than or equal to 5;

SCORING GUIDELINES

Question 2 (continued)

- noting that all are greater than or equal to 1 and at most 20 percent are less than 5.

- Noting that the smallest expected count is 6.33 and that it is greater than or equal to 5 is sufficient to satisfy this component.
- Component 3 is not satisfied if the expected counts are reported as integers.
- This component is not satisfied if the response includes any of the following inappropriate conditions:
  - The response implies that a random sample was taken, e.g., "SRS – check."
  - The response refers to independence of groups or independence of ads as a required condition.
  - The response indicates that a sample size greater than 30 ensures normality or the response implies normality as a condition.
- A response stating that children are independent can be ignored in the scoring of this component.

Section 2 is scored as follows:

Essentially correct (E) if the response includes the following three components:
1. Correct chi-square test-statistic
2. Correct degrees of freedom
3. Correct *p*-value

Partially correct (P) if the response includes only two of the three components.

Incorrect (I) if the response correctly includes at most one of the three components.

Notes:
- If the response makes an error in one calculation, future calculations are considered correct if they follow correctly from the initial miscalculation.
- A chi-square critical value approach is acceptable: The critical value for $\alpha = 0.05$ is $\chi^2 = 5.99$.

Section 3 is scored as follows:

Essentially correct (E) if the response provides a correct conclusion about the alternative hypothesis context, *AND* provides justification based on linkage between the *p*-value and conclusion.

Partially correct (P) if the response provides a correct conclusion in context, but without justification based on linkage to the *p*-value;
OR
if the response provides a correct conclusion, with linkage to the *p*-value, but not in context;
OR
if the response provides a correct decision stated in terms of the null hypothesis in context, with linkage to the *p*-value, but no conclusion is made about the alternative hypothesis.

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

SCORING GUIDELINES

Question 2 (continued)

Notes:

- If the conclusion is consistent with the *p*-value from section 2, and also in context with justification based on linkage to the *p*-value, section 3 is scored as E.
- If no alpha level is given, the solution must be explicit about the linkage by giving a correct interpretation of the *p*-value or by explaining how the conclusion follows from the *p*-value such as saying: "Because the *p*-value is small, we reject the null hypothesis." or "Because the alue is large, we do not reject the null hypothesis."
- If a conclusion contains language that suggests that the response refers to the sample data, conclusion component is not correct, unless the same error occurred in the statement of hypotheses in section 1.

Section 4 is scored as follows:

Essentially correct (E) if the response, in context, concludes that ad type A had little effect id type B had an effect, both supported by the observed proportions (or counts) from the table.

Partially correct (P) if the response correctly concludes that ad type A had little effect, and ad type B had an effect but does not provide correct numerical justification;
OR
if the response compares all of the proportions (or counts) as required, but without correctly concluding ad effectiveness;
OR
if the response correctly describes the effect of one of the ad types A or B (in context, with correct numerical justification) but not the other;
OR
if the response compares all of the proportions (or counts) as required, but not in context.

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

Each essentially correct (E) section counts as 1 point, and a partially correct (P) section counts as ½ point.

4 Complete Response

3 Substantial Response

2 Developing Response

1 Minimal Response

If a response is between two scores (for example, 2½ points), use a holistic approach to decide whether score up or down, depending on the strength of the response and communication.

SCORING GUIDELINES

Question

Intent of Question

The primary goals of this question were to assess a student's ability to (1) identify explanatory and response variables from the description of a research study; (2) indicate and justify whether a study is observational or experimental; and (3) explain what confounding means in the context of a particular study with a specific confounding variable.

Solution

Part (a):

The explanatory variable is the person's degree of cigarette smoking. The response variable is whether the person develops Alzheimer's disease during the course of the study.

Part (b):

This is an observational study because the people in the study were not assigned to a certain degree of cigarette smoking. Rather, the degree of cigarette smoking for each person was passively observed and recorded, not manipulated by the researchers.

Part (c):

nfounding variable is one that is related to the explanatory variable and possibly influences response variable. In this case it seems plausible that people who exercise more regularly might be more health conscious, therefore, less likely to smoke cigarettes than people who do not exercise regularly. Similarly, it's possible that people who exercise more regularly are less likely to develop Alzheimer's disease than people who do not exercise regularly. If both of these relationships turn out to be true, then smoking cigarettes would be associated with developing Alzheimer's disease due to the association of both variables with exercise, even if there were no cause-and-effect relationship between smoking cigarettes and developing Alzheimer's disease.

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 both variables are described correctly. A correct description includes some degree of status of the variables, such as smoking versus not smoking and developing Alzeimer' versus not developing Alzheimer's.

Partially correct (P) if one variable is described correctly and one is not described correctly;
OR
ıeither variable is described correctly but smoking is mentioned as the explanatory variable, and Alzheimer's is mentioned as the response variable;
OR
the explanatory and response variables are correctly described but are interchanged.

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

SCORING GUIDELINES

Question 3 (continued)

Note: Describing the variables only as smoking and Alzheimer's without a reference to levels/status not sufficient for E. But making the connection between smoking and explanatory and between Alzheimer's and response is sufficient for P.

Part (b) is scored as follows:

Essentially correct (E) if the response selects the correct type of study (observational) and provides the correct explanation that smoking status was not assigned *OR* that smoking status was only observed.

Partially correct (P) if the response selects the correct type of study (observational) and provides correct explanation but does not refer to smoking status as the variable that would have been assigned had it been an experiment; for instance, by simply stating that treatment was not assigned;
OR
if the response selects the correct type of study (observational) and provides an explanation that says that smoking is observed (or that smoking status and Alzheimer's are observed) without a modifier for observed (such as, only, just, merely, simply) *AND* without indicating that treatments were not assigned.

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

Notes:

- A response that states the explanatory variable was not assigned without naming smoking status is sufficient for E if the explanatory variable is correctly defined in part (a).
- A response that states that smoking status is ONLY observed (or that smoking status and Alzheimer's are ONLY observed) is sufficient for E.
- A response that says that Alzheimer's is observed without mentioning smoking status is scored I.
- A response that provides an incorrect statistical explanation (such as, "the study is observational because an experiment must have a control group") lowers the score in part (b) by one level (from E to P or from P to I).
- If Alzheimer's is given as the explanatory variable and smoking is given as the response variable in part (a), then part (b) should be scored accordingly with the two variables interchanged.
- Any phrase that refers to the "effect of smoking on Alzheimer's" or "the association between smoking and Alzheimer's" (rather than smoking status) should be ignored.

Part (c) is scored as follows:

Essentially correct (E) if the response includes the following two components:
1. Provides a reasonable explanation that exercise status is related to smoking status.
2. States that exercise status might influence whether the person develops Alzheimer's disease.

Partially correct (P) if the response describes only one of the two components;
OR
if the response only describes how smoking and exercise jointly influence whether the person develops Alzheimer's;

SCORING GUIDELINES

Question 3 (continued)

if the response mentions both of the associations in each component, but also includes an incorrect statistical statement. For instance, part (a) is correct, but in part (c), the explanatory and response variables are interchanged.

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

4 Complete Response

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 parts 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

OR

No parts essentially correct and two parts partially correct

SCORING GUIDELINES

Question 4

Intent of Question

The primary goals of this question were to assess a student's ability to (1) calculate a probability using basic probability rules or the geometric distribution; (2) recognize that a probability calculation for independent events does not depend on the previous outcomes of those events; and (3) assess whether a claim about the probability of a single event is reasonable based on a calculated probability of a series of those events.

Solution

Part (a):

If the failure rate for the super igniters is 15 percent, then the probability that each igniter fails is 0.15, and the probability that it does not fail is 0.85. Therefore the probability that the first 30 igniters tested do not fail is $(0.85)^{30} \approx 0.0076$. The solution can also be written as $(1 - 0.15)^{30} \approx 0.0076$.

Part (b):

Given that there are no failures in the first 30 trials, the probability that the first failure occurs on the trial is 0.15, and the probability that it does not occur on the 31st but occurs on the 32nd trial $(0.85)(0.15) = 0.1275$. Therefore the probability that the first failure occurs on the 31st or 32nd super igniter tested is $0.15 + 0.1275 = 0.2775$.

Note that this is equivalent to asking for the probability that the first failure occurs on the first or second trial, which is $0.15 + (0.85)(0.15) = 0.2775$.

Part (c):

The result of the probability calculation in part (a) provides a reason to believe that the failure rate of the super igniters is less than 15 percent. The calculated probability of 0.0076 shows that there is less than a 1 percent chance that 30 or more igniters in a row would not fail if the failure rate was ercent. This probability is smaller than conventional significance levels such as $\alpha = 0.05$ $\alpha = 0.01$, and thus is small enough to make it reasonable to believe that the failure rate of the super igniters is less than 15 percent.

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 gives the correct probability *AND* correct justification.

Partially correct (P) if the response correctly notes that the answer is the probability that there will be 30 successes in 30 attempts, but does not carry out a correct probability calculation;
OR

SCORING GUIDELINES

Question 4 (continued)

if the response defines the random variable X as the trial with the first failure, identifies X as having a geometric distribution with $p = 0.15$, and writes the desired probability as $P(X > 30)$, but does not carry out a correct probability calculation;

OR

if the response defines the random variable X as the number of failures in the first 30 attempts, identifies X as a binomial random variable with $p = 0.15$ and $n = 30$, and writes the desired probability as $P(X = 0)$, but does not carry out a correct probability calculation;

OR

if the response gives the correct probability but, in specifying a geometric or binomial distribution, has an incorrect or incomplete definition of parameters or value(s) of the random variable.

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

Note: Justification can be given using the multiplication rule; *OR* by defining X to be the trial with the first failure, recognizing that X has a geometric distribution, and using that information to find $P(X > 30)$; *OR* by defining X to be the number of failures in the first 30 attempts, and then finding $P(X = 0)$ using either probability rules or the binomial distribution with $n = 30$ and $p = 0.15$.

Part (b) is scored as follows:

Essentially correct (E) if the response gives the correct probability AND correct justification.

Partially correct (P) if the response makes a reasonable attempt to calculate a geometric, binomial, or conditional probability, but does not successfully carry out the calculation;

OR

if the response gives the correct probability but, in specifying a geometric or binomial distribution, has an incorrect or incomplete definition of parameters or value(s) of the random variable.

Incorrect (I) if the response finds an incorrect probability resulting from an unreasonable attempt to calculate a geometric, binomial, or conditional probability or otherwise does not meet the criteria for E or P.

Note: Similar to part (a) justification can be given using probability rules; *OR* by stating that X geometric where X is the trial with the first failure, then finding $P(X = 1 \text{ or } X = 2)$; *OR* by stating that s the number of failures in two trials and finding $1 - P(X = 0)$ or $P(X = 1 \text{ or } X = 2)$ using the binomial distribution.

Part (c) is scored as follows:

Essentially correct (E) if the response states that it is reasonable to believe that the failure rate is less than 15 percent *AND* bases this decision on the fact that the probability of 30 consecutive successful launches with a failure rate of 15 percent (that is, answer from part (a)) is small *AND* does so in the context of the situation.

Partially correct (P) if the response otherwise satisfies the criteria for an (E) but does so without any context;

SCORING GUIDELINES

Question 4 (continued)

if the response states a significance level and makes a decision in a context that is appropriate to the given probability in part (a) and the stated significance level but does not explicitly compare the probability and the significance level (no linkage).

Incorrect (I) if the response does not explicitly make a decision about whether it is reasonable to conclude that the failure rate is less than 15 percent (For example: "As seen in Part (a), if the failure rate is 15 percent then the probability of 30 successful launches in a row is very small.");
OR
if the response otherwise does not meet the criteria for E or P.

Notes:

- Justification based on the probability can come by stating a significance level and noting that the probability is smaller than the significance level *OR* by simply stating that the probability of 0.0076 is small *OR* by referring to the expected number of failures (4.5) as being very unlikely because zero failures is more than two standard deviations below 4.5.
- If the response bases the decision on the expected number of failures (4.5) for $n = 30$ and $p = 0.15$ without referencing why zero failures would be considered to be too far below 4.5 to give reason to doubt the stated 15 percent failure rate, the response is scored P.
- If the calculation in part (a) is incorrect, the answer in part (c) needs to be consistent with the answer in part (a), unless the value is recalculated in part (c).

4 Complete Response

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 parts 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
OR
No parts essentially correct and two parts partially correct

SCORING GUIDELINES

Question 5

Intent of Question

The primary goals of this question were to assess a student's ability to (1) construct and interpret a confidence interval for a population proportion; (2) explain why one of the conditions for inference necessary; and (3) explain why a suggested procedure for constructing a confidence interval is incorrect.

Solution

Part (a):

The appropriate procedure is a one-sample z-interval for a population proportion. The problem stated the conditions for inference have been met, so they do not need to be checked. A 95 percent confidence interval for the population proportion is given as $\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$, which is

$$0.37 \pm 1.96\sqrt{\frac{(0.37)(0.63)}{1{,}048}} \approx 0.37 \pm 0.03 = (0.34, 0.40).$$

We are 95 percent confident that the population proportion of all adults in the U.S. who would have chosen the economy statement is between 0.34 and 0.40.

Part (b):

The condition is necessary because the formula for the confidence interval relies on the fact that the binomial distribution can be approximated by a normal distribution which then results in the sampling distribution of $\hat{p}$ being approximately normal. The approximation does not work well unless both $n\hat{p}$ and $n(1-\hat{p})$ are at least 10.

Part (c):

The suggested procedure is not appropriate because one of the requirements for using a two-sample interval for a difference between proportions is that the two proportions are based on two independent samples. In the situation described the two proportions come from a single sample and thus are not independent.

Scoring

This question is scored in three sections. Section 1 consists of the mechanics of the confidence interval in part (a), section 2 consists of interpreting the confidence interval in part (a), and section 3 consists of parts (b) and (c).

Section 1 is scored as follows:

Essentially correct (E) if the response includes the following two components:
1. States the correct procedure by name or formula
2. Calculates the confidence interval

Partially correct (P) if the response includes only one of the two components.

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

SCORING GUIDELINES

Question 5 (continued)

Notes:

- A formula with correct values is sufficient for component 2.
- Component 2 can never be satisfied if the response contains an incorrect formula.

Section 2 is scored as follows:

Essentially correct (E) if the response includes the following four components for the interval interpretation:

1. Estimates a proportion
2. Infers about the population
3. States 95 percent confidence
4. Includes context

Partially correct (P) if the response satisfies components 1 and 2 *AND* satisfies only one of components 3 or 4;
OR
if the response gives a correct interpretation of the confidence *level* in context without interpreting the specific interval.

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

Notes:

- Any indication of an inference to the adults sampled rather than the population does not satisfy component 2 and is scored as I.
- Stating values that are unrealistic as proportions (including blanks) in the interpretation lower the score by one level (from E to P or from P to I).
- When both the interpretation of the interval and the level are given, only the interpretation of the interval is scored.

Section 3 is scored as follows:

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

1. Part (b) states that the condition implies the sampling distribution of $\hat{p}$ is approximately normal *OR* the normal approximation to the binomial distribution is appropriate.
2. Part (c) indicates the procedure is not appropriate because the two proportions come from single sample (dependent) rather than two (independent) samples.

Partially correct (P) if the response includes only one of the two components.

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

Notes:

- Referring to the sampling distribution needing to be normal without explicitly stating $\hat{p}$ satisfies component 1.
- Referring to normal approximation must specify "to binomial" to satisfy component 1.
- Discussing the two ample nterval versus the two roportion nterval is extraneous information.

SCORING GUIDELINES

Question 5 (continued)

- In part (c) component 2 can never be satisfied if a procedure that is not possible for one sample of categorical data is suggested.

4 Complete Response

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 parts 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

OR

No parts essentially correct and two parts partially correct

SCORING GUIDELINES

Question

Intent of Question

The primary goals of this question were to assess a student's ability to (1) use a scatterplot to comment on eport about the relationship between two variables and interpret the slope for the least-squares regression line summarizing this relationship; (2) describe the relationship between two variables scatterplot when a categorical variable is introduced and compare a characteristic of the distribution variable for different categories of individuals in a scatterplot; and (3) describe how the associations between two variables for each category of individuals in a scatterplot differ from the overall association i the same scatterplot.

Solution

Part (a):

The scatterplot supports the newspaper report about number of semesters needed to complete an academic program and starting salary because it shows a positive association between these two variables.

Part

The slope is 1.1594. For each additional semester needed to complete an academic program, predicted starting salary increases by €1,159.40.

Part (c):

For the business majors alone, there is a strong, negative, linear association between number of semesters and starting salary. Business majors who need a greater number of semesters to complete an academic program tend to have lower starting salaries.

Part (d):

Business majors have the lowest median starting salary at around €38,000, followed by physics majors at around €48,000, and then chemistry majors with the highest median starting salary at around €55,000.

Part (e):

The newspaper report should be modified to account for major. Overall, majors that take longer to complete tend to have higher starting salaries, with chemistry the highest, physics the next highest, and business the lowest. However, within a major, students who take a greater number of semesters tend to have lower starting salaries.

Scoring

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

SCORING GUIDELINES

Question 6 (continued)

Section 1 is scored as follows:

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

1. In part (a) the response addresses the positive association.
2. In part (a) the response uses the positive association to justify that the scatterplot support newspaper report.
3. In part (b) the response correctly identifies the numerical value of the slope from the computer output.
4. In part (b) the response interprets the slope as the change in starting salary for each additional semester, in context.
5. In part (b) the interpretation of slope includes nondeterministic language (e.g., "predicted starting salary" or equivalent) when interpreting the slope.

Partially correct (P) if the response includes three or four of the five components.

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

Notes:

- In part (a) the response can use phrases such as "positive association (correlation, relationship)," "increasing relationship," or describe a positive association (e.g., "the starting salaries are higher when there is a greater number of semesters") to satisfy component 1 However, describing the relationship between only two points does not satisfy this component.
- In part (a) comments about linearity and strength should be ignored, even if the response implies these are required (e.g., "Yes, because the relationship is strong, positive, and linear.").
- In part (a) responses that answer "no," "maybe," "kind of," "somewhat," or equivalent do not satisfy component 2.
- In part (a) a response that says "no, because the three clusters of points each have a negative association" does not satisfy component 1 or component 2.
- In part (a) no context is required, but variable names are required in part (b) to satisfy component 4.
- In part (b) a response that states the equation $\hat{y} = 34.018 + 1.1594x$ does not satisfy component 3 unless the slope is specifically identified or used in the interpretation.
- In part (b) a response that incorrectly identifies the numerical value of the slope can still satisfy components 4 and 5 using the incorrect value.
- In part (b) the 1-unit increase in number of semesters must be stated or implied to satisfy component 4 (e.g., for each semester, for every semester). A response that states or implies an unspecified number of semesters does not satisfy this component (for example, as semesters increase).
- In part (b) examples of nondeterministic language include "predicted starting salary," "expected starting salary," "estimated starting salary," "typical starting salary," "average starting salary," "starting salary, on average," "our model predicts," and so on. However, "about," "approximately," and "according to the model" do not satisfy component 5.
- In part (b) no units are required for the change in predicted starting salary (which means OK to say 1.1594 euros, 1159.40, 1.1594, 1,159.40 dollars, and so on).
- In all parts it is acceptable if a response refers to salary rather than starting salary.

SCORING GUIDELINES

Question 6 (continued)

Section 2 is scored as follows:

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

1. In part (c) the response states that the association is negative.
2. In part (c) the response states that the association is strong *OR* linear *OR* both.
3. In part (c) the response refers to both variables (semesters, salary) in context.
4. In part (d) the response correctly compares the three majors.
5. In part (d) the response provides reasonable values for the median salaries or refers to "median starting salaries" when describing a characteristic of the graph.

Partially correct (P) if the response includes three or four of the five components.

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

Notes:

- In part (c) the response can use phrases such as "negative association (correlation, relationship)," "decreasing relationship," "inversely related," or describe a negative association (for example, "the starting salaries are smaller when there is a greater number of semesters") satisfy component 1. However, describing the relationship between only two points does not satisfy this component.
- In part (c) responses that describe the relationship incorrectly as weak or nonlinear do satisfy component 2, even if the other characteristic is described correctly.
- Only drawing a line on the scatterplot does not satisfy component 2.
- In part (d) a response that says "chemistry has the highest median starting salary and business has the lowest median starting salary" (or the equivalent) implies that physics is in the middle, and satisfies component 4.
- In part (d) if no values are provided for the medians, the response must use the phrase "median starting salary" at least once to satisfy component 5.
- In all parts it is acceptable if a response refers to salary rather than starting salary.

Section 3 is scored as follows:

Essentially correct (E) if the response states that there is a negative association for each of the majors *AND* the response notes the overall positive association.

Partially correct (P) if the response states that there is a negative association for each of the majors BUT does not note the overall positive association;
OR
if the response states that there is a negative association for one or two specific majors (for example, for business majors) *AND* the response notes the overall positive association.

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

Notes:

- Additional ways to note the overall positive association include stating that the original report was correct, stating that majors that take longer to complete tend to have higher starting salaries, or the equivalent.

SCORING GUIDELINES

Question 6 (continued)

- A response that does not explicitly name the majors or refer to the three majors but makes a general statement such as "However, for an academic major, there is a negative association" satisfies the requirement to state the negative association for each of the majors.
- A response that states the original report was wrong, or incorrect, or should be retracted, cannot satisfy the requirement to note the overall positive association. However, a response that says the original report might be misleading (or the equivalent) but still notes the overall positive association satisfies the requirement to note the overall positive association.
- In all parts it is acceptable if a response refers to salary rather than starting salary.

4 Complete Response

All three sections essentially correct

3 Substantial Response

Two sections essentially correct and one section partially correct

OR

Sections 1 and 2 partially correct and section 3 essentially correct

OR

Sections 1 and 3 essentially correct and section 2 incorrect

OR

Sections 2 and 3 essentially correct and section 1 incorrect

2 Developing Response

Sections 1 and 2 essentially correct and section 3 incorrect

OR

Section 1 essentially correct and sections 2 and 3 partially correct

OR

Sections 1 and 3 partially correct and section 2 essentially correct

OR

One section essentially correct and one section partially correct

OR

All three sections partially correct

1 Minimal Response

One section essentially correct and two sections incorrect

OR

Two sections partially correct and one section incorrect

OR

One section partially correct and two sections incorrect