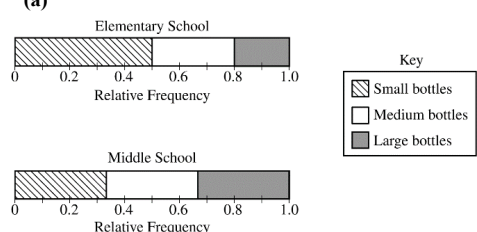


4 points

Question 2: Focus on Exploring Data

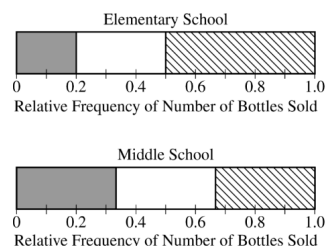
General Scoring Notes

- 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) 	Essentially correct (E) if the response satisfies the following two components: - The elementary school's segmented bar graph is correctly partitioned according to the given proportions - The middle school's segmented bar graph is correctly partitioned to create three equal areas Partially correct (P) if the response only satisfies one of the two components required for E. Incorrect (I) if the response does not meet the criteria for E or P.

Additional Notes:

- Responses that do not use the key given may still satisfy both components as long as a new key is provided that defines small, medium, and large bottles.
- Responses that do not use the key given and do not define a new key may receive credit for component 2 as long as the bar is vertically partitioned to create three equal areas, but do not satisfy component 1.
- Responses need to be within ± 0.05 of the actual cutoff to be acceptable for demonstrating equal areas.
- Responses do not need to be drawn in the order from smallest bottles to largest bottles as shown in the model solution. For example, the graph could be drawn from largest bottles to smallest bottles.



Model Solution	Scoring
(b) No, the segment for small bottles for the elementary school is wider than the segment for small bottles for the middle school; however, the middle school students sold three times as many bottles as the elementary school students. So, if the elementary school sold x number of bottles, the middle school sold $3x$ number of bottles. For example, if the elementary students sold 100 bottles total, then they sold $0.5(100)$ or 50 small bottles. However, because the middle school students sold three times the total number of bottles as the elementary students, they would have sold 300 bottles total and $(0.\bar{3})(300) = 100$ small bottles. Because $100 > 50$, the middle school sold more small bottles, and the elementary school's administrator is not correct.	Essentially correct (E) if the response satisfies the following three components: - States that the elementary school's administrator is incorrect - Provides correct mathematical support verifying the middle school sold more small bottles than the elementary school, consistent with the response to component 1 - Includes context Partially correct (P) if the response satisfies only two of the three components required for E. Incorrect (I) if the response does not meet the criteria for E or P.

Additional Notes:

- A response that shows mathematical support by showing the inequalities $0.5x < 0.\bar{3}(3x)$, $0.5x < x$, $0.5(1) < 0.\bar{3}(3)$, or $(0.5)\left(\frac{1}{4}\right) < (0.\bar{3})\left(\frac{3}{4}\right)$ may satisfy component 2.
- In order to satisfy component 3, a response must include "elementary," "middle," and "bottles."
- Part (b) should be scored consistent with the response to part (a).

Model Solution	Scoring
(c) (i) The mosaic plot shows that the proportion of large bottles sold by High School A was 0.7 and the proportion of large bottles sold by High School B was 0.6. High School A sold a greater proportion of large bottles because $0.7 > 0.6$. (ii) The number of bottles sold is represented by the area of the shaded region. The area of the rectangle representing large bottles sold by High School B is clearly larger than the area of the rectangle representing large bottles sold by High School A. Therefore, High School B sold more large bottles than High School A.	Essentially correct (E) if the response satisfies the following four components: 1. In part (c-i) the response indicates that High School A sold a greater proportion of large bottles 2. In part (c-i) the response bases reasoning on the height of the rectangles or the relative frequencies representing large bottles 3. In part (c-ii) the response indicates that High School B sold a greater number of large bottles 4. In part (c-ii) the response indicates that the area of the rectangle representing large bottles sold by High School B is larger than the area of the rectangle representing large bottles sold by High School A Partially correct (P) if the response satisfies two or three of the four components required for E. Incorrect (I) if the response does not meet the criteria for E or P.

Additional Notes:

- If a response refers to “proportions” rather than relative frequencies in part (c-i), the values 0.7 and 0.6 must be provided to satisfy component 2.
- If the response bases reasoning on the heights of rectangles in part (c-i), the response must clearly state that the rectangles representing large bottles are referenced.
- If the response bases reasoning on the areas of rectangles in part (c-ii), the response must clearly state that the areas representing large bottles are referenced.
- A response that provides reasonable estimates of the large bottle areas may satisfy component 4.

Scoring for Question 2	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 <i>OR</i> One part essentially correct and one or two parts partially correct <i>OR</i> Three parts partially correct	2
Minimal Response One part essentially correct and no part partially correct <i>OR</i> No part essentially correct and two parts partially correct	1

4 points

Question 5: Multi-Focus

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: - Provides the correct probability - 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) $P(\text{majority regular cards} \mid \text{fewer than 6 months})$

$$= \frac{P(\text{majority regular cards} \cap \text{fewer than 6 months})}{P(\text{fewer than 6 months})}$$

$$= \frac{\frac{80}{500}}{\frac{91}{500}}$$

$$= \frac{80}{91}$$

$$= 0.879$$

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

- Provides the correct probability
- 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) <i>AND</i> 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 <i>AND</i> 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) <i>AND</i> 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 $\frac{1}{2}$ point.

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

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

Question 6: Investigative Task**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)	(i) The relative frequencies of successful and unsuccessful treatments in each clinic are presented in the following table: <table> <tr> <th></th><th>Clinic A</th><th>Clinic B</th></tr> <tr> <td>Unsuccessful Treatment</td><td>$\frac{51}{139} \approx 0.3669$</td><td>$\frac{33}{68} \approx 0.4853$</td></tr> <tr> <td>Successful Treatment</td><td>$\frac{88}{139} \approx 0.6331$</td><td>$\frac{35}{68} \approx 0.5147$</td></tr> </table> (ii) Clinic A appears to be more successful in treating allergy sufferers than Clinic B. Clinic A was successful for 63.3% of the allergy sufferers it treated, while Clinic B was successful for only 51.5% of the allergy sufferers it treated.		Clinic A	Clinic B	Unsuccessful Treatment	$\frac{51}{139} \approx 0.3669$	$\frac{33}{68} \approx 0.4853$	Successful Treatment	$\frac{88}{139} \approx 0.6331$	$\frac{35}{68} \approx 0.5147$	Essentially correct (E) if the response satisfies the following two components: - The response to part (a-i) provides correct numerical values for the relative frequencies in all four cells of the table - The response to part (a-ii) identifies the clinic that is more successful in treating allergy sufferers and includes a justification based on the relative frequencies reported in part (a-i) Partially correct (P) if the response satisfies only one of the two components OR if the response provides correct numerical values for the relative frequencies in only two or three of the cells of the table, and correctly identifies the clinic that has greater success, with or without providing adequate justification based on the values of the reported relative frequencies. Incorrect (I) if the response does not meet the criteria for E or P.
	Clinic A	Clinic B									
Unsuccessful Treatment	$\frac{51}{139} \approx 0.3669$	$\frac{33}{68} \approx 0.4853$									
Successful Treatment	$\frac{88}{139} \approx 0.6331$	$\frac{35}{68} \approx 0.5147$									

Additional Notes:

- Responses may be given as proportions or percentages.
- To satisfy component 1, numerical values for relative frequencies must be accurately reported to at least two decimal places, e.g., 0.36 or 0.37, 0.48 or 0.49, 0.63 or 0.64, 0.51 or 0.52.
- If a response in part (a-i) reports the relative proportions of the total (dividing the count in each of the four cells by 207) instead of the conditional relative frequencies, then component 2 may be satisfied in either of the following ways:
 - Concluding that Clinic A is more successful in treating allergies based on the larger proportion of successful cases (0.43 vs. 0.17)
 - Concluding that Clinic B is more successful in treating allergies based on the smaller proportion of unsuccessful cases (0.16 vs. 0.25).

- If a response reports incorrect values in part (a-i), then the justification in component 2 may use either the relative frequencies reported in the table in part (a-i) OR the correct conditional relative frequencies given in the model solution.
- To satisfy component 2, the response must make it clear that Clinic A is being compared to Clinic B, either by explicitly mentioning both clinics OR by reporting both relative frequencies that are being compared.

Model Solution	Scoring
(b) No. This is an observational study; it is not a randomized experiment. Cause and effect can only be established with a well-designed, randomized experiment. There may be other variables, besides where a patient was treated, that affect the success rates for treating allergies. For example, Clinic A may mostly treat mild allergy cases that are easy to treat successfully, while Clinic B may mostly treat severe allergy cases that are more difficult to treat successfully.	Essentially correct (E) if the response satisfies the following two components: 1. Indicates “no,” a causal inference is not justified 2. Provides at least one of the following explanations: • Notes that this is an observational study • Indicates that this is not a randomized experiment • Identifies a plausible confounding variable that could affect the success rates for treating allergies at the two clinics Partially correct (P) if the response satisfies component 2 but fails to indicate that researchers are not justified in making a causal inference <i>OR</i> if the response satisfies component 1 AND includes one of the following as justification: • A statement that association does not imply causation • A statement that there may be a confounding variable without specifying a reasonable confounding variable. Incorrect (I) if the response does not satisfy the criteria for E or P.

Additional Notes:

- A response that states that treatments (Clinic A, Clinic B) are not assigned (randomly assigned, imposed) to the patients satisfies component 2.
- A response that identifies a reasonable confounding variable does not need to explain why it is a confounding variable in order to satisfy component 2.
- The strength of the response in part (b) should be considered if holistic scoring is necessary.

Model Solution	Scoring
(c) (i) Clinic A: More successful in treating mild allergies. Clinic A successfully treated 75.0% of mild allergy sufferers, while Clinic A successfully treated only 28.6% of severe allergy sufferers. Clinic B: More successful in treating mild allergies. Clinic B successfully treated 91.7% of mild allergy sufferers, while Clinic B successfully treated only 42.9% of severe allergy sufferers. (ii) Clinic A: More likely to treat mild allergy sufferers than severe allergy sufferers. Of the 139 allergy sufferers treated at Clinic A, 104 (74.8%) suffered from mild allergies, while only 35 (25.2%) suffered from severe allergies. Clinic B: More likely to treat severe allergy sufferers than mild allergy sufferers. Of the 68 allergy sufferers treated at Clinic B, 56 (82.4%) suffered from severe allergies, while only 12 (17.6%) suffered from mild allergies.	Essentially correct (E) if the response satisfies the following four components: 1. The response to part (c-i) indicates that both Clinic A and Clinic B are more successful in treating mild allergies than treating severe allergies 2. The response to part (c-i) provides a justification based on correctly reported proportions (or percentages) of successfully treated allergy sufferers of each severity for each clinic 3. The response to part (c-ii) indicates that Clinic A is more likely to treat mild allergy sufferers, while Clinic B is more likely to treat severe allergy sufferers 4. The response to part (c-ii) provides a justification based on correctly reported frequencies or proportions (or percentages) of severe or mild allergy sufferers treated by each clinic Note: See the Q6 Correct Relative Frequencies tables at the end of this scoring guideline for the relative frequencies that should be used in each designated part. Partially correct (P) if the response satisfies 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 to part (c-i) may satisfy component 2 in either of the following ways:
 - For each clinic, reports the correct proportion (or percentage) of severe allergy sufferers successfully treated AND the correct proportion (or percentage) of mild allergy sufferers successfully treated.
 - Makes a correct reference to the mosaic plot, comparing the ratio of shaded area to open area within the “Mild” rectangle to the ratio of shaded area to open area within the “Severe” rectangle for each clinic. Numerical values are not required.
- A response to part (c-ii) may satisfy component 4 in any of the following ways:
 - For each clinic, reports the correct proportion (or percentage) of treated allergy sufferers with severe allergies. (Clinic A: 0.252, Clinic B: 0.824)
 - For each clinic, reports the correct proportion (or percentage) of treated allergy sufferers with mild allergies. (Clinic A: 0.748, Clinic B: 0.176)

- Reports the correct proportion (or percentage) of treated allergy sufferers with severe allergies for one clinic AND reports the correct proportion (or percentage) of treated allergy sufferers with mild allergies for the other clinic.
- For each clinic, reports the correct frequency of treated allergy sufferers with mild allergies and the correct frequency of treated allergy sufferers with severe allergies. (Clinic A: 104 mild and 35 severe; Clinic B: 12 mild and 56 severe)
- Makes a correct reference to the mosaic plot, comparing the area or width of the “Mild” rectangle to the area or width of the “Severe” rectangle within each clinic. Numerical values are not required.

Model Solution	Scoring
(d) The more successful clinic identified in part (a-ii) is Clinic A, which is different from the physician’s conclusion that Clinic B is better when taking allergy severity into account. This happens because for both clinics the success rate is much higher for mild allergy sufferers (75.0 percent versus 28.6 percent for Clinic A, and 91.7 percent versus 42.9 percent for Clinic B). Clinic A treats mostly mild allergy sufferers (74.8 percent of its patients) while Clinic B treats mostly severe allergy sufferers (82.4 percent of its patients). Therefore, when combining the results across allergy severity categories to obtain the table from part (a-i), the facts that Clinic A treats a larger proportion of mild allergy sufferers and mild allergy sufferers have a higher success rate make it appear as if Clinic A is better overall.	Essentially correct (E) if the response satisfies the following two components: 1. The response indicates that within each clinic, the proportion of successfully treated mild allergy sufferers is greater than the proportion of successfully treated severe allergy sufferers (0.75 vs. 0.286 for Clinic A and 0.917 vs. 0.429 for Clinic B) <i>OR</i> the response indicates that the overall proportion of successfully treated mild allergy sufferers (0.767) is much larger than the overall proportion of successfully treated severe allergy sufferers (0.374) 2. The response indicates that the allergy sufferers treated by Clinic A include a much greater proportion of mild cases than the allergy sufferers treated by Clinic B (sample proportions: 0.748 vs. 0.176) <i>OR</i> the response indicates that the allergy sufferers treated by Clinic B includes a much greater proportion of severe cases than the allergy patients treated by Clinic A (sample proportions: 0.824 vs 0.252) <i>OR</i> the response indicates that of the allergy sufferers treated at Clinic A, a higher proportion are mild (0.748) than severe (0.252), and of the allergy sufferers treated by Clinic B, a higher proportion are severe (0.824) than mild (0.176) Partially correct (P) if the response satisfies only one of the two components. Incorrect (I) if the response does not meet the criteria for E or P.

Additional Notes:

- A response that does not use the answer from (c-i) cannot satisfy component 1.
- A response that does not use the answer from (c-ii) cannot satisfy component 2.
- A response that compares successful treatments between clinics (0.75 vs. 0.917 or 0.286 vs. 0.429), does not satisfy component 1 or component 2.

- A response does not need to repeat the calculated values from part (c), but if it does, the numbers must be the same as those that were reported in part (c).
- A response that only compares the proportion of successful (mild and severe) treatments at Clinic A to the proportion of successful (mild and severe) treatments at Clinic B should be scored I.
- If the response in part (a-ii) concludes that Clinic B is more successful in treating allergies, then part (d) should be scored as follows:
 - **Essentially correct (E)** if the response includes the following four components:
 - (i) At least one of the two components listed for E above.
 - (ii) The response states that overall, Clinic B has a higher proportion of successfully treated allergy sufferers.
 - (iii) The response states that Clinic B has a higher proportion of successfully treated mild allergy sufferers.
 - (iv) The response states that Clinic B has a higher proportion of successfully treated severe allergy sufferers.
 - **Partially correct (P)** if the response satisfies component (i) AND two of the three remaining components (ii)–(iv) listed for E in this note.
- If the response in part (a-ii) concludes that Clinic B is more successful, the response can be scored no higher than 3.
- The strength of the response in part (d) should be considered if holistic scoring is necessary.

Q6 Correct Relative Frequencies

(a-i)

	Clinic A	Clinic B
Unsuccessful treatment	$\frac{51}{139} = 0.367$	$\frac{33}{68} = 0.485$
Successful treatment	$\frac{88}{139} = 0.633$	$\frac{35}{68} = 0.515$

(c-i)

	Clinic A	Clinic B
Successful treatment - Mild	$\frac{78}{104} = 0.75$	$\frac{11}{12} = 0.917$
Successful treatment - Severe	$\frac{10}{35} = 0.286$	$\frac{24}{56} = 0.429$

(c-ii)

	Clinic A	Clinic B
Mild cases treated	$\frac{104}{139} = 0.748$	$\frac{12}{68} = 0.176$
Severe cases treated	$\frac{35}{139} = 0.252$	$\frac{56}{68} = 0.824$

NOTE: To satisfy the components in part (c), the comparison needs to be made by severity WITHIN each clinic (a vertical comparison), not Clinic A to Clinic B.

Scoring for Question 6

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: - Indicates that the researcher's claim is not correct (or "may not be correct", if proportions are not reported) - 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 <i>OR</i> - 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 <i>OR</i> if the response satisfies component 2 only <i>OR</i> 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: Relative Frequency of Response Proportion of teens who did consume soft drinks in the previous week Proportion of teens who did not consume soft drinks in the previous week (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: - Constructs a segmented bar graph in part (b-i), with the bars correctly segmented - 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) - 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) - 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. (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.	Essentially correct (E) if the response satisfies the following three components: - Identifies a chi-square test for homogeneity by name in part (c-i) - 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) - 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) 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.

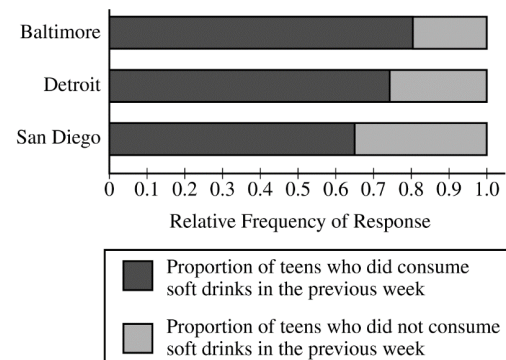
Additional Notes:

- 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.

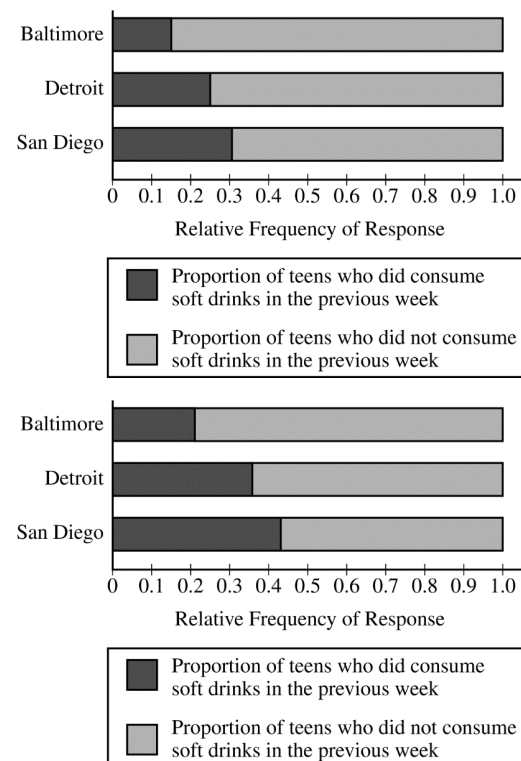
Scoring for Question 5	Score
Complete Response Three parts essentially correct	4
Substantial Response Two parts essentially correct and one part partially correct	3
Developing Response Two parts essentially correct and no part partially correct <i>OR</i> One part essentially correct and one or two parts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

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

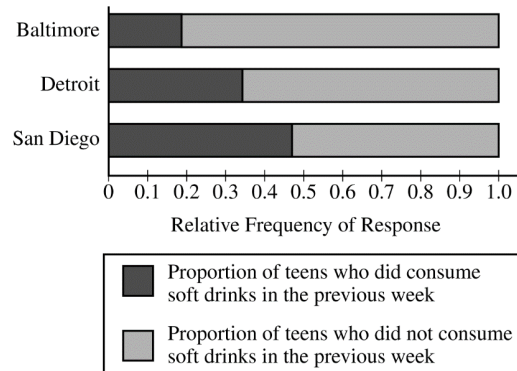
Acceptable Graph



Common Unacceptable Graphs



Common Unacceptable Graphs (continued)



AP[®] Statistics

2016 Scoring Guidelines

© 2016 The College Board. College Board, Advanced Placement Program, AP, AP Central, and the acorn logo are registered trademarks of the College Board.

Visit the College Board on the Web: www.collegeboard.org.

AP Central is the official online home for the AP Program: apcentral.collegeboard.org.

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):

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.

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

$$\begin{aligned}\chi^2 &\approx \\ &0.292 + 0.860 + \\ &1.720 + 5.070 + \\ &0.595 + 1.754 \approx 10.291.\end{aligned}$$

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 p -value is very small (for instance, much smaller than $\alpha = 0.05$) we reject the null hypothesis at the 5 level (and at the 1 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.

When 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:

- Both hypotheses are stated correctly with at least one in context.
- Identifies the correct test procedure (by name or by formula).
- 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)

- o 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:
 - o The response implies that a random sample was taken, e.g., “SRS – check.”
 - o The response refers to independence of groups or independence of ads as a required condition.
 - o 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 p -value 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 and 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):

Confounding 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 Alzheimer's versus not developing Alzheimer's.

Partially correct (P) if one variable is described correctly and one is not described correctly;

OR

either 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 percent. This probability is smaller than conventional significance levels such as $\alpha = 0.05$ or $\alpha = 0.01$, 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 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 sampling 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 sample interval versus the two proportion interval 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 a report 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 in 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”) to 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 not 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

AP[®] Statistics

2014 Scoring Guidelines

© 2014 The College Board. College Board, Advanced Placement Program, AP, AP Central, and the acorn logo are registered trademarks of the College Board.
Visit the College Board on the Web: www.collegeboard.org.
AP Central is the official online home for the AP Program: apcentral.collegeboard.org.

2014 SCORING GUIDELINES

Question 1

Intent of Question

The primary goals of this question were to assess a student's ability to (1) calculate conditional proportions from a two-way table; (2) comment on association between two categorical variables as displayed in a graph; and (3) draw an appropriate conclusion from the p -value of a chi-square test.

Solution**Part (a):**

The proportion of on campus residents who participate in at least one extracurricular activity is $\frac{17+7}{33} = \frac{24}{33} \approx 0.727$. The proportion of off campus residents who participate in at least one extracurricular activity is $\frac{25+12}{67} = \frac{37}{67} \approx 0.552$.

Part (b):

The graph reveals that on campus residents in this sample are more likely to participate in extra-curricular activities than off campus residents. The proportions who participate in two or more extra-curricular activities are similar between the two groups but slightly greater for on campus residents (on campus: 0.212, off campus: 0.179). On campus residents have a greater proportion who participate in one activity (on campus: 0.515, off campus: 0.373) and a smaller proportion who participate in no extracurricular activities (on campus: 0.273, off campus: 0.448) than off campus residents.

Part (c):

The p -value of 0.23 is greater than conventional significance levels such as $\alpha = 0.10$ or $\alpha = 0.05$ or $\alpha = 0.01$. Therefore, the p -value indicates that the sample data do not provide strong enough evidence to conclude that participation in extracurricular activities differs between on and off campus residents in the population of all students at the university.

2014 SCORING GUIDELINES

Question 1 (continued)

Scoring

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

Part (a) is scored as follows:

Essentially correct (E) if the response correctly performs *both* calculations with work shown.

Partially correct (P) if the response correctly performs one of the two calculations with work shown;
OR

if the response provides both correct answers with no work shown;
OR

if the response calculates the proportion of students involved in exactly one extracurricular activity rather than the proportion of students involved in at least one extracurricular activity for both groups, with work shown.

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

Note: Answers reported as fractions rather than decimals are acceptable.

Part (b) is scored as follows:

Essentially correct (E) if the response correctly compares proportions between the two groups of students for at least two of the three categories.

Partially correct (P) if the response correctly lists proportions for at least two categories for the two groups but does not make an explicit comparison between the two groups;
OR

if the response correctly compares the relative values of the proportions between the two groups of students for only one of the three categories.

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

Notes:

- A response without any reference to percentages or proportions is scored as at most P, (for example, a response that attempts to compare counts).
- A response that treats bar graphs as distributions of a quantitative variable lowers the score one level (that is, from E to P, or from P to I) in part (b).

2014 SCORING GUIDELINES

Question 1 (continued)

Part (c) is scored as follows:

Essentially correct (E) if the response states a correct conclusion in the context of the study *AND* provides correct justification/decision of that conclusion based on linkage to the p -value.

Partially correct (P) if the response provides no conclusion in context but does provide correct justification/decision based on linkage to the p -value;

OR

if the response provides a correct conclusion in context but with incorrect or missing justification/decision based on linkage to the p -value;

OR

if the response provides the conclusion in context and correct justification/decision based on linkage to the p -value but states a conclusion equivalent to accepting the null hypothesis.

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

Notes:

- Justification based on the p -value can be given by stating a significance level and noting that the p -value is larger than the significance level OR by simply stating that the p -value is large.
- A conclusion that is equivalent to “accept the null hypothesis”, either as a stated decision or as a conclusion in context, cannot be scored as E. Such a response is scored as P if it includes both content and correct justification based on linkage to p -value. If such a response lacks either context or linkage, it is scored as I.

4 Complete Response

All three parts essentially correct

3 Substantial Response

Two parts essentially correct and one part partially correct

2 Developing Response

Two parts essentially correct and one part incorrect

OR

One part essentially correct and one or two parts partially correct

OR

Three parts partially correct

1 Minimal Response

One part essentially correct and two parts incorrect

OR

Two parts partially correct and one part incorrect

2014 SCORING GUIDELINES

Question 2

Intent of Question

The primary goals of this question were to assess a student's ability to (1) calculate a probability; (2) assess whether a claim about randomness is questionable in light of a calculated probability; and (3) determine whether a description of a simulation method achieves a correct simulation of a random process.

Solution

Part (a):

The probability that all 3 people selected are women can be calculated using the multiplication rule, as follows:

$P(\text{all three selected are women})$

$= P(\text{first is a woman}) \times P(\text{second is a woman} | \text{first is a woman}) \times P(\text{third is a woman} | \text{first two are women})$

$= \frac{3}{9} \times \frac{2}{8} \times \frac{1}{7} \approx 0.012$

Part (b):

The probability calculated in part (a) does provide a reason to doubt the manager's claim that the selections were made at random. The calculation shows that there is only about a 1.2% chance that random selection would have resulted in three women being selected. The probability is small enough that it may cast doubt on the manager's claim that the selections were made at random.

Part (c):

No, the process does not correctly simulate the random selection of three women from a group of nine people of whom six are men and three are women. The random selection of three people among nine is done *without* replacement. However, in the simulation with the dice, the three dice rolls in any given trial are independent of one another, indicating a selection process that is done *with* replacement.

2014 SCORING GUIDELINES

Question 2 (continued)

Scoring

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

Part (a) is scored as follows:

Essentially correct (E) if the response correctly computes the probability of selecting the three women, and shows how the probability was computed.

Partially correct (P) if the response shows only one of the following:

Gives the correct probability of $\frac{1}{84}$ (0.012 or 0.011 is acceptable) but does not show how it was

computed;

OR

Correctly shows how the probability should be computed, but does not carry the computation through correctly;

OR

Correctly computes (showing work) only the numerator, or only the denominator of the correct

answer. (For example, $\frac{1}{9} \times \frac{1}{8} \times \frac{1}{7} \approx 0.002$, or $\frac{3}{9} \times \frac{2}{9} \times \frac{1}{9} \approx 0.008$, or $\frac{3}{9} \times \frac{3}{8} \times \frac{3}{7} \approx 0.054$);

OR

Mistakenly assumes independence and calculates (showing work) the binomial probability

$$\frac{1}{3} \times \frac{1}{3} \times \frac{1}{3} = \frac{1}{27} \approx 0.037.$$

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

Part (b) is scored as follows:

Essentially correct (E) if the response states that the probability from part (a) is small (or insufficiently small), makes an appropriate decision consistent with the probability being small (or insufficiently small), and does so in the context of this situation.

Partially correct (P) if the response shows only one of the following:

Otherwise satisfies the criteria for an E but does so without any context;

OR

States a significance level and makes a decision in 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;

OR

Otherwise satisfies the criteria for an E but does not explicitly make a decision about whether there is reason to doubt the manager's claim. (For example: "The probability of selecting the three women from among the nine employees is very small so it is unlikely to occur by chance.")

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

2014 SCORING GUIDELINES

Question 2 (continued)

Notes:

- Each of the following situations is one in which a response that otherwise would be scored as E should be scored as P, and a response that otherwise would be scored as P should be scored as I:
 - The response includes a statement that the small probability *proves* that the manager did not make the selection at random (or any equivalent wording).
 - The response includes a statement that clearly interprets the probability from part (a) to be the probability that the manager selected the three people at random.
- Each of the following situations is one in which the response is scored as I:
 - The decision is inconsistent with the justification (e.g., "The probability is very small, so there is no reason to doubt the manager's claim").
 - The response states or implies that because the selection of three women was *not impossible*, there is *no reason to doubt* the manager's claim.

Part (c) is scored as follows:

Essentially correct (E) if the response answers no *AND* states that the dice outcomes in the proposed simulation are independent *AND* states that the genders of the selected convention attendees are dependent. The table below shows statements that should be considered equivalent to the required statements of independence and dependence.

<u>Independence</u> of dice outcomes	<u>Dependence</u> of genders
- The three dice outcomes are independent. - The probability of rolling a 5 or a 6 is the same on all three dice. - The dice simulation actually simulates sampling with replacement.	- The genders of the three people are dependent (or not independent). - The probability of selecting a woman changes after each selection. - The people are sampled without replacement.

OR

Essentially correct if the response answers no *AND* computes the correct probability that a trial of the simulation will indicate the selection of three women $\left(\frac{1}{3} \times \frac{1}{3} \times \frac{1}{3} \approx 0.037\right)$ *AND* states that the probability is different from the probability found in part (a).

Partially correct (P) if the response correctly answers no and either:

States only that the dice outcomes are independent or states only that the genders of the selected convention attendees are dependent, but not both;

OR

Otherwise meets the criteria for E but has poor communication. An example of poor communication is: "No, because it selects with replacement. It isn't possible for the same person to be selected twice." (There is an apparent shift between the two sentences from describing the simulation to describing the actual selection of people, but that is not made clear.)

2014 SCORING GUIDELINES

Question 3

Intent of Question

The primary goals of this question were to assess a student's ability to (1) perform a probability calculation from a normal distribution; (2) explain an implication of examining the distribution of a sample mean rather than the distribution of a single measurement; and (3) perform a probability calculation involving independent events using the multiplication rule.

Solution**Part (a):**

Because the distribution of the daily number of absences is approximately normal with mean 120 students and standard deviation 10.5 students, the z-score for an absence total of 140 students is

$z = \frac{140 - 120}{10.5} \approx 1.90$. The table of standard normal probabilities or a calculator reveals that the probability that 140 or fewer students are absent is 0.9713. So the probability that more than 140 students are absent (and that the school will lose some state funding) is $1 - 0.9713 = 0.0287$.

Part (b):

High School A would be *less* likely to lose state funding. With a random sample of 3 days, the distribution of the sample mean number of students absent would have less variability than that of a single day. With less variability, the distribution of the sample mean would concentrate more narrowly around the mean of 120 students, resulting in a smaller probability that the mean number of students absent would exceed 140.

In particular, the standard deviation of the sample mean number of absences, $\bar{x}$, is

$\frac{\sigma}{\sqrt{n}} = \frac{10.5}{\sqrt{3}} = 6.062$. So the z-score for a sample mean of 140 is $\frac{140 - 120}{6.062} \approx 3.30$. The probability that

High School A loses funding using the suggested plan would be $1 - 0.9995 = 0.0005$, as determined from the table of standard normal probabilities or from a calculator, which is less than a probability of 0.0287 obtained for the plan described in part (a).

Part (c):

For any one typical school week, the probability is $\frac{2}{5} = 0.4$ that the day selected is not Tuesday, not Wednesday, or not Thursday. Therefore, because the days are selected independently across the three weeks, the probability that none of the three days selected would be a Tuesday or Wednesday or Thursday is $(0.4)^3 = 0.064$.

2014 SCORING GUIDELINES

Question 3 (continued)

Scoring

Parts (a), (b), and (c) were 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. Indicates use of a normal distribution and clearly identifies the correct parameter values (showing correct components of a z-score calculation is sufficient).
2. Uses the correct boundary value (140, 140.5, or 141 is acceptable).
3. Reports the correct normal probability consistent with components 1 and 2

OR

if the response reports a probability of 0.025 with justification based on the empirical rule for an acceptable boundary value (140, 140.5, or 141 is acceptable).

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

OR

if the response provides an incorrect probability of 0.05 with justification based on the empirical rule for an acceptable boundary value (140, 140.5, or 141 is acceptable).

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

Note: An inconsistency in calculations lowers the score for part (a) by one level (that is, from E to P or from P to I).

Part (b) is scored as follows:

Essentially correct (E) if the response provides the correct answer of less likely *AND* the following three components:

1. Clearly references the distribution of the sample mean.
2. Indicates that the variability of the distribution is smaller.
3. Indicates that the distribution is centered at 120.

OR

if the response provides the correct answer of less likely *AND* the following two components:

1. Correctly calculates the probability that the sample mean would exceed 140 (arithmetic errors are not penalized).
2. Correctly compares this probability to the probability in part (a).

Partially correct (P) if the response provides the correct answer of less likely *AND* only two of the following three components:

1. Clearly references the distribution of the sample mean.
2. Indicates that the variability of the distribution is smaller.
3. Indicates that the distribution is centered at 120.

OR

if the response provides the correct answer of less likely *AND* correctly calculates the probability that the sample mean would exceed 140 (arithmetic errors are not penalized) *BUT* does not correctly compare this probability with the probability in part (a).

2014 SCORING GUIDELINES

Question 3 (continued)

Incorrect (I) if the response does not meet the criteria for E or P, including if the response provides the incorrect answer or provides the correct answer of less likely with no explanation or an incorrect explanation.

Note: An equivalent approach is to use the total number of absences for 3 days. The sampling distribution of the total number of absences for the 3 days is approximately normal, with mean

$3(120) = 360$ absences and standard deviation $3(6.026) \approx 18.187$ absences. The z-score for a total of

$3(140) = 420$ absences is: $\frac{420 - 360}{18.187} \approx 3.30$. Such a response is scored E if the response provides the

correct answer of less likely and references the distribution of the sample total, and includes the correct mean and standard deviation.

Part (c) is scored as follows:

Essentially correct (E) if the response correctly calculates the probability *AND* shows sufficient work.

Partially correct (P) if the response reports the correct probability but shows no work or does not show sufficient work;

OR

if the response uses the multiplication rule involving three events but does so incorrectly and/or with an incorrect probability of not selecting a Tuesday, Wednesday, or Thursday.

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

4 Complete Response

All three parts essentially correct

3 Substantial Response

Two parts essentially correct and one part partially correct

2 Developing Response

Two parts essentially correct and one part incorrect

OR

One part essentially correct and one or two parts partially correct

OR

Three parts partially correct

1 Minimal Response

One part essentially correct and two parts incorrect

OR

Two parts partially correct and one part incorrect

2014 SCORING GUIDELINES

Question 4

Intent of Question

The primary goals of this question were to assess a student's ability to (1) describe why the median might be preferred to the mean in a particular context; (2) compare the relative merits of two sampling plans; and (3) describe a consequence of nonresponse in a particular study.

Solution

Part (a):

The median is less affected by skewness and outliers than the mean. With a variable such as income, a small number of very large incomes could dramatically increase the mean but not the median.

Therefore, the median would provide a better estimate of a typical income value.

Part (b):

Method 2 is better than Method 1. A sample obtained from Method 1 could be biased because of the voluntary nature of the response. It is plausible that class members with larger incomes might be more likely to return the form than class members with smaller incomes. The mean income for such a sample would overestimate the mean income of all class members. With Method 2, despite the smaller sample size, the random selection is likely to result in a sample that is more representative of the entire class and produce an unbiased estimate of mean yearly income of all class members.

2014 SCORING GUIDELINES

Question 4 (continued)

Scoring

This question is scored in three sections. Part (b) has two components: (1) identifying a relevant characteristic for each sampling method; (2) indicating the effect of the biased method on the estimate of the mean income. Section 1 consists of Part (a); section 2 consists of part (b), component 1; and section 3 consists of part (b), component 2. 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 includes the following two components:

1. Describes how skewness or outliers affect the mean or do not affect the median.
2. Makes a conjecture about a relevant characteristic of the distribution of incomes, such as skewness or an outlier.

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

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

Notes:

- For Component 1, examples of responses that are acceptable include:
 - The mean is affected by skewness (outliers).
 - The median is not affected by skewness (outliers).
 - The mean is greater (less) than the median when there is right (left) skewness or outliers.
- For Component 1, examples of responses that are not acceptable include:
 - Don't use the mean for skewed distributions or distributions with outliers.
 - Use the median for skewed distributions or distributions with outliers.
 - Responses that include an incorrect statement about means and/or medians, such as for right skewed distributions, the median will be higher than the mean.
- It is possible to satisfy *both* components with a single sentence, such as, "If there was a billionaire in the sample, the mean would be higher than the median."
- If a response argues that using the *mean* is a more appropriate way to estimate the typical income, then reduce the score in section 1 by one level (that is, from E to P or from P to I).

Section 2 is scored as follows:

Essentially correct (E) if the response chooses Method 2 *AND* includes the following two components:

1. Identifies a relevant characteristic of Method 1.
2. Identifies a relevant characteristic of Method 2.

Partially correct (P) if the response chooses Method 2 *AND* includes only one of the two components listed above

OR

if the response includes both components but does not choose a method.

Incorrect (I) if the response chooses Method 1 *OR* otherwise does not meet the criteria for E or P.

2014 SCORING GUIDELINES

Question 4 (continued)

Notes:

- Responses that do not explicitly choose Method 2 can still earn an E for section 2 if the choice is clearly implied. The choice of Method 2 is clearly implied if the response only discusses negative characteristics of Method 1 and only discusses positive characteristics of Method 2, such as, Method 1 is biased but Method 2 uses a random sample.
- Responses that *compare* the two methods can satisfy *both* components, such as, saying that Method 1 is more biased or that nonresponse will be less of a problem with Method 2.
- Responses that refer to the nonresponse bias as *voluntary response bias*, *response bias*, *undercoverage* can still earn an E.
- Discussions of conditions for inference should be considered extraneous and ignored.

Section 3 is scored as follows:

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

1. Indicates the incomes of responders may be different from the incomes of nonresponders.
2. Indicates the biased sampling method may produce a misleading estimate/conclusion about the mean income, including direction, for example, "The sample mean is likely to be higher than the mean of the population."

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

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

Notes:

- A single sentence can satisfy the first component of section 2 and the first component of section 3. (For example, "In method 1, rich people are more likely to respond.")
- For component 2, either direction is acceptable but the direction must be consistent with the identified bias. Saying only that Method 2 will be more accurate or more representative does not satisfy component 2.
- If a response addresses possible nonresponse bias in Method 2, the response can still satisfy both components of section 3.
- Responses that focus on the larger sample size in Method 1 can satisfy component 2 if such responses describe the effect as reducing the variability of the estimate. (For example, "I would use Method 1 since the larger sample size would give less variability of the mean.")
- Responses that focus on untruthful survey answers can satisfy component 2 if the effect on the estimate is appropriate. (For example, "People contacted in Method 2 might say they make more money than they actually do. This would make the estimated mean too high.")

2014 SCORING GUIDELINES

Question 4 (continued)

4 Complete Response

All three sections essentially correct

3 Substantial Response

Two sections essentially correct and one section partially correct

2 Developing Response

Two sections essentially correct and one section incorrect

OR

One section essentially correct and one or two sections partially correct

OR

Three sections partially correct

1 Minimal Response

One section essentially correct and two sections incorrect

OR

Two sections partially correct and one section incorrect

2014 SCORING GUIDELINES

Question 5

Intent of Question

The primary goal of this question was to assess students' ability to identify, set up, perform, and interpret the results of an appropriate hypothesis test to address a particular question. More specific goals were to assess students' ability to (1) state appropriate hypotheses; (2) identify the 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: States a correct pair of hypotheses.

Let μ_{diff} represent the population mean difference in purchase price (woman – man) for identically equipped cars of the same model, sold to both men and women by the same dealer, in the county.

The hypotheses to be tested are $H_0 : \mu_{\text{diff}} = 0$ versus $H_a : \mu_{\text{diff}} > 0$.

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

The appropriate procedure is a paired t -test.

The conditions for the paired t -test are:

1. The sample is randomly selected from the population.
2. The population of price differences (woman – man) is normally distributed, or the sample size is large.

The first condition is met because the car models and the individuals were randomly selected. The sample size ($n = 8$) is not large, so we need to investigate whether it is reasonable to assume that the population of price differences is normally distributed. The dotplot of sample price differences reveals a fairly symmetric distribution, so we will consider the second condition to be met.

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

The test statistic is $t = \frac{585 - 0}{\frac{530.71}{\sqrt{8}}} \approx 3.12$.

The p -value, based on a t -distribution with $8 - 1 = 7$ degrees of freedom, is 0.008.

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

Because the p -value is very small (for instance, smaller than $\alpha = 0.05$), we reject the null hypothesis. The data provide convincing evidence that, on average, women pay more than men in the county for the same car model.

2014 SCORING GUIDELINES

Question 5 (continued)

Scoring

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

Step 1 is scored as follows:

Essentially correct (E) if the response identifies the correct parameter *AND* states correct hypotheses.

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

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

Note: Defining the parameter symbol in context or simply using common parameter notation is sufficient.

Step 2 is scored as follows:

Essentially correct (E) if the response identifies the correct test procedure (by name or by formula) *AND* checks *both* conditions correctly.

Partially correct (P) if the response correctly completes two of the three components (identification of procedure, check of randomness condition, check of normality condition).

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

Note: The random sampling condition can be verified by referring to the random selection of car models or to the random selection of male and female car buyers.

Step 3 is scored as follows:

Essentially correct (E) if the response correctly calculates both the test statistic and the p -value.

Partially correct (P) if the response correctly calculates the test statistic but not the p -value;
OR
 if the response calculates the test statistic incorrectly but then calculates the correct p -value for the computed test statistic.

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

Note: If the response identifies a z -test for a mean as the correct procedure in step 2, then the response can earn a P in step 3 if both the test statistic and the p -value are calculated correctly.

2014 SCORING GUIDELINES

Question 5 (continued)

Step 4 is scored as follows:

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

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

OR

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

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

Notes:

- If the conclusion is consistent with an incorrect p -value from step 3 and also in context with justification based on linkage to the p -value, step 4 is scored as E.
- A response that performs a two-sample t -test with correct calculations should fail to reject H_0 . A conclusion that is equivalent to “accept H_0 ” (such as “we conclude that women pay the same amount as men, on average”), either as a stated decision or as a conclusion in context, cannot be scored as E. Such a response will be scored as P provided that the conclusion is in context with linkage. Such a response will be scored as I if it lacks either context or linkage.

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

4 Complete Response

3 Substantial Response

2 Developing Response

1 Minimal Response

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

2014 SCORING GUIDELINES

Question 6

Intent of Question

The primary goals of this question were to assess a student's ability to (1) calculate and interpret a residual value; (2) answer questions about residual plots; (3) compare associations between two scatterplots; and (4) identify an appropriate explanatory variable to include in a regression model based on residuals from simpler regression models.

Solution**Part (a):**

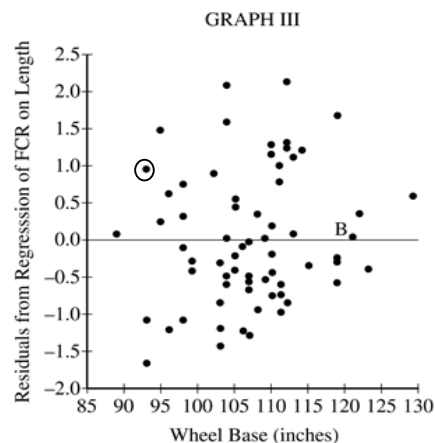
For a car with length 175 inches, the predicted value for the car's FCR, based on the least squares regression line, is

$$\text{predicted FCR} = -1.595789 + 0.0372614(175) \approx 4.92 \text{ gallons per 100 miles.}$$

The actual FCR for the car is 5.88, so the residual is $5.88 - 4.92 = 0.96$. The residual value means that the car's FCR is 0.96 gallons per 100 miles greater than would be predicted for a car of its length.

Part (b):

- (i) The point with a wheel base of 93 inches and a residual of 0.96 gallons per 100 miles is circled in graph III below.



- (ii) Point B corresponds to a car with an actual FCR that is very close to the FCR that would be predicted for a car with its length by the regression model which predicts FCR using the explanatory variable length.

2014 SCORING GUIDELINES

Question 6 (continued)

Part (c):

Graph II reveals a moderate association that is positive and linear. In contrast, there is a weak association that is positive and linear in graph III. The association between engine size and residual (from predicting FCR based on length) is stronger than the association between wheel base and residual (from predicting FCR based on length).

Part (d):

Engine size is a better choice than wheel base for including with length in a regression model for predicting FCR. The stronger association between engine size and residual (from predicting FCR based on length) indicates that engine size is more useful than wheel base for reducing the variability in FCR values that remains unexplained (as indicated by residuals) after predicting FCR based on length.

Scoring

Parts (a), (b), (c), and (d) were 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 two components:

1. A correct residual value with supporting calculation.
2. A correct interpretation of the residual value, in context.

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

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

Notes:

- If the residual value is incorrect, the interpretation should be considered correct if it follows from the incorrect residual value.
- Correct interpretation of the residual must include the correct direction and magnitude of the FCR value away from the predicted FCR value.
- A calculated residual value which is slightly different from 0.96 due to the number of significant digits is acceptable.

Part (b) is scored as follows:

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

1. Circles the correct point in graph III.
2. Provides a reasonable interpretation of the car associated with point B having a residual near 0 that refers to predicting FCR based on length.

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

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

2014 SCORING GUIDELINES

Question 6 (continued)

Note: A correct response for the second component must include reference to the observed FCR value of the car represented by point B, not the point B itself.

Part (c) is scored as follows:

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

1. A description of form *AND* direction for both graphs.
2. A description of the strength of association for both graphs.
3. A comparison between the two graphs.

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

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

Notes:

- Part (c) is focused on the comparison of graph II and graph III. Inferences drawn from patterns in these graphs are considered in part (d).
- Linear is needed for form in graph II.
- Graph III may be described as having no association between wheel base and the residuals of FCR based on length, which is sufficient for describing the form, direction and strength of association of graph III.

Part (d) is scored as follows:

Essentially correct (E) if the response indicates the correct choice with a sound justification based on the following two components:

1. The strong(er) association.
2. Reducing the variability that remains unexplained in the model which predicts FCR based on length.

Partially correct (P) if the response indicates the correct choice and provides a justification based on only one of the two components which are listed above.

Incorrect (I) if the response indicates the incorrect choice;

OR

if the response indicates the correct choice but does not mention either of the two components which are listed above.

Note: Describing the variables in graph II and graph III as residuals is not required but can be used positively in holistic scoring. Incorrect descriptions of graph II or graph III or the variables in graphs are not acceptable.

2014 SCORING GUIDELINES

Question 6 (continued)

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

4 Complete Response

3 Substantial Response

2 Developing Response

1 Minimal Response

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