

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) Let X represent the amount of gold applied to a necklace randomly selected from necklaces produced with this machine. The random variable X has an approximately normal distribution with mean 300 mg and standard deviation 5 mg. Then, $P(296 < X < 304)$ $= P(X < 304) - P(X \leq 296)$ $= P\left(Z < \frac{304 - 300}{5}\right) - P\left(Z \leq \frac{296 - 300}{5}\right)$ $= P(Z < 0.8) - P(Z \leq -0.8)$ $\approx 0.7881 - 0.2119 \approx 0.5763.$	Essentially correct (E) if the response satisfies the following three components: - Indicates the use of a normal distribution with mean 300 and standard deviation 5 - Specifies the correct event (boundary value and direction) or an event consistent with values reported in component 1 - Provides the correct probability of 0.5763 or a probability consistent with components 1 and 2 Partially correct (P) if the response satisfies only two of the three components. Incorrect (I) if the response does not satisfy the criteria for E or P.

Additional Notes:

Component 1

- A response may satisfy component 1 by any of the following or a combination of the following:
 - Graphical:** Displaying a graph of a normal density function with the appropriate scale on the horizontal axis showing the mean and the standard deviation for the distribution of the amount of gold applied to the necklaces.
 - Calculator function syntax:** Labeling correct values of the mean and standard deviation in a “normalcdf” statement, such as:
 $\text{normalcdf}(\text{lower} = 296, \text{upper} = 304, \text{mean} = 300, \text{standard deviation} = 5).$
 Correct specification of the lower and upper bounds is not required to satisfy component 1.
 - Words:** Using a statement such as “normal distribution with mean 300 and standard deviation 5.”
 - Standard Notation:** Using standard notation such as $N(300, 5)$ or $N(300, (5)^2)$.
 - Z-score:** Displaying the correct mean and standard deviation in a z-score calculation that includes “z,”
 such as $z = \frac{304 - 300}{5}.$

Component 2

- A response may satisfy component 2 by any of the following or a combination of the following:
 - Graphical: Displaying a graph of a normal density function with the region of interest ($296 < \text{amount of gold} < 304$ or $-0.8 < Z < 0.8$) clearly identified. The shaded area does not need to be proportional, but the boundaries should be on the proper side of the mean and the shading should be in the proper direction.
 - Calculator function syntax: Labeling the lower and upper bounds of the region of interest in a “normalcdf” statement, such as:
 - `normalcdf(lower = 296, upper = 304, mean = 300, standard deviation = 5)`.
 - `normalcdf(lower = -0.8, upper = 0.8,  $\mu = 0$ ,  $\sigma = 1$ )`.
 - Words: Specifying the correct event in words with correct numerical value for the boundary value and correct direction, such as “the probability that the amount of gold applied to a randomly selected necklace is between 296 and 304 mg” or $P(296 < \text{amount of gold} < 304)$.
 - Standard Notation: Using standard notation such as $P(296 < X < 304)$ or $P\left(\frac{296 - 300}{5} < Z < \frac{304 - 300}{5}\right)$ or $P(-0.8 < Z < 0.8)$.

General

- Minor errors in statistical notation may be ignored. However, this is considered poor communication if holistic scoring is required.
- If the only error in the response to part (a) is the reversal of the numerator for the z-score, $(300 - 296)$ or $(300 - 304)$, the response is scored P.
- An arithmetic or transcription error in a response can be ignored if correct work is shown.
- A response that satisfies components 1 and 2 by including more than calculator syntax (e.g., standard random variable notation and probability notation) is considered good communication and may be considered if holistic scoring is required.

Model Solution	Scoring
(b) (i) If the machine is working properly, and the sample mean amount of gold, $\bar{X}$, has a sampling distribution that follows a normal distribution with mean 300 mg and standard deviation $\frac{5}{\sqrt{2}} \approx 3.5355$ mg, then $P(\bar{X} > 303)$ $= P\left(Z > \frac{303 - 300}{\frac{5}{\sqrt{2}}}\right)$ $\approx P(Z > 0.8485)$ $\approx 0.198.$ (ii) Observing a sample mean amount of 303 mg would not provide convincing evidence that the population mean amount of gold being applied by the machine is something other than 300 mg because the probability of observing a sample mean that differs from 300 mg by 3 mg or more is large, around $0.198(2) = 0.396$.	Essentially correct (E) if the response satisfies components 1, 4, and 5 <i>AND</i> at least one of components 2 or 3: 1. In part (b-i), indicates the use of a normal (or an approximately normal) distribution and identifies the correct parameter values (mean 300 mg and standard deviation $\frac{5}{\sqrt{2}} \approx 3.5355$ mg) with work shown or a correct formula 2. In part (b-i), specifies the correct event (boundary value and direction) 3. In part (b-i), provides the correct probability of 0.198 or the probability consistent with components 1 and 2 4. In part (b-ii), indicates whether observing a sample mean of 303 mg would provide convincing evidence that the population mean is not 300 mg, consistent with the response to part (b-i) 5. In part (b-ii), provides justification based on the probability obtained in part (b-i) or a correctly computed probability in part (b-ii) Partially correct (P) if the response does not meet the requirements for E and satisfies at least three of the five components <i>OR</i> the response satisfies only components 4 and 5, consistent with the response to part (b-i) <i>OR</i> the response satisfies only component 1 and one of the other four components. Incorrect (I) if the response does not meet the criteria for E or P.

Additional Notes:**Component 1**

- A response may satisfy component 1 by any of the following or a combination of the following:
 - Graphical: Displaying a graph of a normal density function with the appropriate scale on the horizontal axis showing mean and standard deviation for the distribution of mean amount of gold.
 - Calculator function syntax: Labeling correct values of the mean and standard deviation in a “normalcdf” statement, such as “normalcdf” or “ncdf”, e.g.,
 - $\text{normalcdf}\left(\text{lower} = 303, \text{upper} = \infty, \text{mean} = 300, \text{standard deviation} = \frac{5}{\sqrt{2}}\right)$.
 Correct specification of the upper and lower bounds is not required to satisfy component 1.
 - Words: Using a statement such as “approximately normal distribution with mean 300 and standard deviation $\frac{5}{\sqrt{2}}$.”
 - Standard Notation: Using standard notation such as $N\left(300, \frac{5}{\sqrt{2}}\right)$ or $N\left(300, \left(\frac{5}{\sqrt{2}}\right)^2\right)$.
 - Z-score: Displaying the correct mean and standard deviation in a z-score calculation, which includes “z,” such as $z = \frac{303 - 300}{\frac{5}{\sqrt{2}}}$.
 - To satisfy the supporting work criterion of component 1 the response must clearly indicate that the standard deviation for $\bar{X}$ is computed by dividing 5 by the square root of 2. This may be indicated with $\frac{5}{\sqrt{2}}$, words, or standard notation such as $\frac{\sigma}{\sqrt{n}}$.

Component 2

- A response may satisfy component 2 by any of the following or a combination of the following:
 - Graphical: Displaying a graph of a normal density function with the region of interest ($\bar{X} > 303$ or $Z > 0.8485$) clearly identified. The shaded area does not need to be proportional, but the boundary should be on the proper side of the mean and the shading should be in the proper direction.
 - Calculator function syntax: Labeling the lower and upper bounds of the region of interest in a “normalcdf” statement, such as:
 - $\text{normalcdf}\left(\text{lower} = 303, \text{upper} = \infty, \text{mean} = 300, \text{standard deviation} = \frac{5}{\sqrt{2}}\right)$.
 - $\text{normalcdf}(\text{lower} = 0.8485, \text{upper} = \infty, \mu = 0, \sigma = 1)$.
 - Words: Specifying the correct event in words with correct numerical value for the boundary and correct direction, such as the “probability that the mean amount of gold applied to two randomly selected necklaces is greater than 303 mg” or $P(\text{mean amount of gold} > 303)$.
 - Standard Notation: Using standard notation such as or $P(\bar{X} > 303)$ or $P\left(Z > \frac{303 - 300}{3.5355}\right)$ or $P(Z > 0.8485)$.

Component 3

- A response may satisfy component 3 if the reported probability is consistent with components 1 and 2. If a normal distribution with a mean of 300 mg and standard deviation 5 mg is used, the probability reported should be 0.274.

General

- Minor errors in statistical notation may be ignored. However, this is considered poor communication and may be considered if holistic scoring is required.
- If the only error in part (b) is the reversal of the numerator for the z -score ($300 - 303$), the response is scored P.
- An arithmetic or transcription error in a response can be ignored if correct work is shown.

Alternate Solution:

- A response that calculates the probability that the total amount of gold on the two randomly selected necklaces is greater than 606 mg (2×303 mg) can be scored essentially correct if the response demonstrates:
 - The use of a normal (or approximately normal) distribution and the correct mean and standard deviation for the distribution of the total amount of gold applied to two randomly selected necklaces, (e.g., $N(600, 7.071)$, or mean = $2(300) = 600$ mg and standard deviation = $\sqrt{2(5^2)} \approx 7.071$ mg, or $z = \frac{606 - 600}{7.071} \approx 0.8485$), satisfying component 1.
 - Specifies the correct event, including the correct boundary value of $x_{total} = 2(303) = 606$ or $z \approx 0.8485$, and the correct direction, satisfying component 2.
 - The probability is correctly computed using the mean and the standard deviation of the total amount of gold applied to two randomly selected necklaces. An arithmetic error can be ignored if correct work is shown and the result is between 0 and 0.5.
- While the probability of a difference in part (b-ii) is twice the probability found in part (b-i), it is not necessary for a response to double the probability in part (b-i). However, if the probability in part (b-i) is doubled, that should be considered if holistic scoring is required.

Model Solution	Scoring
(c) (i) The sampling distribution of the sample range for random samples of size $n = 2$ from a normal distribution with standard deviation $\sigma = 5$ is skewed to the right. Almost all values of the simulated ranges are between 0 mg and about 25 mg and the center of the distribution is about 6 mg. (ii) As the value of the population standard deviation increases, the variation (spread) in the distribution of the sample range increases and the mean of the distribution of the sample range also increases.	Essentially correct (E) if the response satisfies four or five of the following five components: 1. In part (c-i), describes the shape as positively skewed (or skewed to the right) 2. In part (c-i), describes the center of the distribution as about 6 mg 3. In part (c-i), describes the spread of the simulated sample ranges as having most values between 0 mg and 30 mg 4. In part (c-ii), indicates that the mean (or center) of the distribution of the sample range increases as the population standard deviation increases 5. In part (c-ii), indicates that the variation (or spread) of the distribution of the sample range increases as the population standard deviation increases Partially correct (P) if the response satisfies three of the five components OR the response satisfies only components 4 and 5 OR the response only satisfies components 2 and 4 OR the response satisfies only components 3 and 5. Incorrect (I) if the response does not meet the criteria for E or P.

Additional Notes:

- The centrality component of the response to part (c-i) may be satisfied with any reasonable description of center, such as a comment on the approximate value of mean or median of the distribution or indication that the center is somewhere between 4 and 10 mg.
- The spread component of the response to part (c-i) may be satisfied with a reasonable statement about the inter-quartile range. Any value between 5 and 7, inclusive, is considered reasonable.
- Although the actual range of the sample distribution of the sample range is infinite, a comment that the range is any value between 20 and 30 may be scored as satisfying the spread component of the response to part (c-i).
- The response to part (c-ii) need not comment on the skewness of the distributions. Any comment on the skewness of the distributions should be considered as extraneous in scoring the response.

Model Solution	Scoring
(d) (i) No, a sample range of 10 mg is not unusual if the machine is working properly with a standard deviation of 5 mg. Although observing a sample range around 10 mg or greater is much more likely if the population standard deviation is 8 mg or 12 mg than when the population standard deviation is 5 mg, the graph of the sampling distribution of the sample range for samples of size 2 from a normal distribution with $\sigma = 5$ mg indicates that 10 mg is not an unusual value for the range when $\sigma = 5$ mg. There is about a 20% chance that a random sample of two necklaces would yield a range of 10 mg or more when the machine is working properly. (ii) No, Cleo's sample mean of 303 mg and range of 10 mg do not indicate that the machine is not working properly. As noted in part (b-i), the probability that the sample mean would be equal to or greater than 303 mg when the machine is working properly is almost 20% so having a sample mean of 303 mg is not unusual. Furthermore, it is less than one standard deviation, $\frac{5}{\sqrt{2}} = 3.5355$ mg, away from 300 mg. As indicated in part (d-i), the probability of a range of 10 mg or greater when the population standard deviation is 5 mg is also about 20%, so not unusual. There is not statistically significant evidence to show the machine is not working properly.	Essentially correct (E) if the response satisfies four or five of the following five components: 1. In part (d-i), indicates a sample range of 10 mg is not unusual 2. In part (d-i), justifies the response to component 1 by indicating that the graph of the sampling distribution of the sample range when $\sigma = 5$ mg shows that values of 10 mg or greater occur often 3. In part (d-ii), indicates there is not convincing evidence that the machine is not working properly 4. In part (d-ii), justifies the response by indicating that a sample mean of 303 mg would not be unusual when the machine is working properly with one of the following • 303 mg is less than one standard deviation away from 300 mg • The probability that the mean is greater than or equal to 303 mg was shown in (b) to have probability of 0.198, which is not unusual • Some other reasonable justification 5. In part (d-ii), justifies the response by indicating that a sample range of 10 mg is not unusual when the machine is working properly as discussed in part (d-i) Partially correct (P) if the response does not meet the requirements for E and satisfies at least three of the five components OR the response satisfies components 1 and 2 OR the response satisfies component 3 and either component 4 or 5. Incorrect (I) if the response does not meet the criteria for E or P.

Additional Notes:

- An argument that a sample range of 10 mg or more is more likely to occur (or closer to the center of the sampling distribution of the range) when $\sigma = 8$ mg or $\sigma = 12$ mg is not required for component 2, but by itself does not satisfy component 2. Inclusion of this argument should be considered if holistic scoring is required

- Other reasonable justifications to satisfy component 4 include arguments based on a z -value close to zero, a one-sample z -test or one-sample t -test for a population mean, or a one sample z -interval or t -interval for a population mean.

Alternate Solutions:

- A response that uses the intersection (or union) of the event ‘the sample mean is greater than or equal to 303 mg’ and the event ‘the sample range is greater than or equal to 10 mg’ can satisfy components 3, 4, and 5.
 - Components 3, 4, and 5 are satisfied if the response shows, for example:
 - Let $P(M) = P(\text{mean} \geq 303) = 0.198$ (from part b-ii).
 - Let $P(R) = P(\text{range} \geq 10 \text{ if } \sigma = 5) \approx 0.2$ (from part d-i).
 - For Intersection: $P(M \text{ and } R) = P(M)P(R) = (0.198)(0.2) = 0.0396$, because this probability is small, this is evidence that the machine is not working properly.
 - For Union: $P(M \text{ or } R) = P(M) + P(R) - P(M)P(R) = 0.198 + 0.2 - (0.198)(0.2) = 0.3584$, because this probability is large, this is not evidence that the machine is not working properly.

General Notes:

- Context (amount of gold in necklaces) is not required in parts (a), (b), (c), or (d) of the response. However, including context is considered good communication and may be considered if holistic scoring is required.

Scoring for Question 6

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.	

2018**AP[®]** **CollegeBoard**

AP Statistics

Scoring Guidelines

2018 SCORING GUIDELINES**Question 1****Intent of Question**

The primary goals of this question were to assess a student's ability to (1) identify various values in regression computer output; (2) interpret the intercept of a regression line in context; (3) interpret the coefficient of determination (r^2) in context; and (4) identify an outlier from a scatterplot.

Solution**Part (a):**

The estimate of the intercept is 72.95. It is estimated that the average time to finish checkout if there are no other customers in line is 72.95 seconds.

Part (b):

The coefficient of determination is $r^2 = 73.33\%$. This value indicates that 73.33% of the variability in the times it takes customers to finish checkout, including time waiting in line, can be explained by knowing how many customers are in line in front of the selected customer.

Part (c):

The outlier is the point with $x = 3$ and y close to 0. This point is considered an outlier because the combination of x and y values differs from the pattern of the rest of the data. Specifically, the value of y (time to finish checkout) is much lower than would be expected when there are $x = 3$ customers in line in front of the selected customer, given the remaining data.

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 satisfies the following three components:

1. Correctly identifies 72.95 as the intercept.
2. Communicates the concept of a y -intercept in a context that includes both time and zero customers.
3. Indicates that the value of the intercept is a prediction by using language such as “predicted,” “estimated,” or “average” value of y .

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

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

2018 SCORING GUIDELINES

Question 1 (continued)

Notes:

- Regression equations (such as $\hat{y} = 72.95 + 174.40x$) cannot be used to satisfy identification of the intercept in component 1, unless the intercept is explicitly labeled as such.
- A regression equation cannot be used to satisfy component 3.
- Incorrect regression equations are treated as extraneous and do not affect the scoring of any component.
- A response that interprets 72.95 as a slope does not satisfy components 1 or 2.

Part (b) is scored as follows:

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

1. Correctly identifies 73.33% as the coefficient of determination.
2. Provides a correct (possibly generic) interpretation of r^2 .
3. Interpretation includes context.

Partially correct (P) if the response satisfies only two of the three components;

OR

if the response satisfies the three components, but reverses the roles of number of customers in line and time to finish checkout in the interpretation.

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

Notes:

- In component 2 the correct interpretation of the coefficient of determination can take any of several equivalent forms, such as:
 - The percent variability in y that is attributed to the linear relationship between y and x or between x and y .
 - The proportion of the total variability in the dependent variable y that is explained by the independent variable x .
 - The proportion of variation in y that is accounted for by the linear model.
 - The proportionate reduction of total variation of the y values that is associated with the use of the independent variable x .
 - The proportionate reduction in the sum of the squares of vertical deviations obtained by using the least-squares line instead of the naïve prediction of $\bar{y}$.
- In component 2 common *incorrect* interpretations of the coefficient of determination include:
 - The percent variability in the *predicted* y values that is explained by the linear relationship between y and x .
 - The percent variability in the *data* that is explained by the linear relationship between y and x .
 - The percent variability that is explained by the linear relationship between y and x .
 - The percent variability in y that is *on average* explained by the linear relationship between y and x .
- For component 3 context must include mention of time or customers.

2018 SCORING GUIDELINES**Question 1 (continued)**

Part (c) is scored as follows:

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

1. Correctly identifies the outlier.
2. Describes an unusual feature of the identified scatter plot point, relative to the remaining data points, that is sufficient to identify it as the outlier. Examples include:
 - The combination of x and y values is unusual compared to the other points.
 - The value of y is much lower than would be expected (or predicted), given the remaining data.
 - The residual for the point is unusually large relative to the other residuals.

Partially correct (P) if the response satisfies component 1 but does not satisfy component 2.

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

Notes:

- In the absence of any point being circled on the graph, component 1 can still be satisfied by explicitly referring to the coordinates of the outlier. Valid coordinates for outlier identification must specify an x value of 3 and a y value that is strictly between 0 and 250.
- A response that does not make a comparison to the remaining data points, such as stating the outlier has a large residual or is nowhere near the regression line, does not satisfy component 2.
- A response that makes a comparison to the remaining data points based upon an unusual feature that is *insufficient* for outlier identification, such as stating the point is the only point with that particular y value, does not satisfy component 2.
- In the absence of explicit numerical calculation, a response that appeals to the influence that the outlier has on the regression coefficient estimates or on the sample correlation coefficient does not satisfy component 2.

2018 SCORING GUIDELINES**Question 1 (continued)****4 Complete Response**

Three parts essentially correct

3 Substantial Response

Two parts essentially correct and one part partially correct

2 Developing Response

Two parts essentially correct and no parts partially correct

OR

One part essentially correct and one or two parts partially correct

OR

Three parts partially correct

1 Minimal Response

One part essentially correct

OR

No parts essentially correct and two parts partially correct

2018 SCORING GUIDELINES

Question 2

Intent of Question

The primary goals of this question were to assess a student's ability to (1) calculate the sample size when given the endpoints of a confidence interval for a proportion; (2) explain how bias could be present in a particular survey method; and (3) estimate a proportion from sample data collected using a method designed to decrease bias.

Solution**Part (a):**

Using the standard formula for a confidence interval for one proportion, the interval (0.584 to 0.816) is found as follows. $\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ where $\hat{p} = \frac{0.584 + 0.816}{2} = 0.7$, the margin of error is $0.816 - 0.7 = 0.116$, and $z^* = 1.96$.

Solving $1.96 \sqrt{\frac{0.7(1-0.7)}{n}} = 0.116$ yields $n = \frac{(1.96)^2(0.7)(1-0.7)}{(0.116)^2} \approx 59.95$. The sample size was 60.

Part (b):

Bias might have been introduced because students responded directly to the environmental science teacher. Because the students would know that an environmental science teacher cares about the environment, they might say yes when they actually don't recycle. This would result in a point estimate that is greater than the proportion of all students who would respond yes to the question.

Part (c):

(i) The expected number is $(300)\left(\frac{1}{2}\right) = 150$.

(ii) The point estimate is based on expecting 150 students to be required to say no and 150 students to truthfully answer the question. Of the 213 answers of no, we expect that $213 - 150 = 63$ were from students who truthfully answered the question. That means we expect that the remaining $150 - 63 = 87$ students truthfully answered the question and responded yes. So the point estimate for the proportion of all students at the high school who would respond yes to the question is $\frac{87}{150} = 0.58$.

2018 SCORING GUIDELINES

Question 2 (continued)

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 satisfies the following five components:

1. Uses a standard error in the form $\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ where $\hat{p}$ is between 0 and 1.
2. Shows evidence that $\hat{p} = 0.7$ was correctly used in the standard error.
3. Shows evidence that 0.116 was correctly used as the margin of error in the calculation.
4. Shows evidence that $z^* = 1.96$ was correctly used as the critical value in the calculation.
5. Includes a single, positive whole-number answer.

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

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

Notes:

- Using an equation in the form $n = \frac{z^2 \hat{p}(1-\hat{p})}{\text{MOE}^2}$ satisfies component 1.
- A value of 0.21 in the numerator of the standard error implies that $\hat{p} = 0.7$ was correctly used in the standard error and satisfies component 2.
- An equation such as $0.816 = 0.7 + \text{MOE}$ implies that 0.116 was correctly used for the margin of error and satisfies component 3.
- Statements that suggest a whole-number answer is approximate (such as, “about 60” or “ ≈ 60 ”) satisfy component 5.
- Algebraic work between the set-up and final answer does not need to be shown to satisfy component 5.
- When calculating the values 0.7, 0.116, or 1.96, ignore minor arithmetic errors or transcription errors if they can be identified by the work shown.

2018 SCORING GUIDELINES**Question 2 (continued)**

Part (b) is scored as follows:

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

1. Explains why the responses to the survey might differ from the truth about student recycling in this context (for example, the survey was not anonymous, the question was asked by an authority figure).
2. Explains how the responses to the survey might differ from the truth about student recycling (for example, “students might say yes when they actually don’t recycle,” “students lie and say yes,” “students don’t recycle but lie to the teacher”).
3. Describes the effect of the bias on the point estimate (or the proportion, percentage, number of yes responses in the sample) and doesn’t contradict the bias described.

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

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

Notes:

- To satisfy component 1 the response must provide a reason that is based on a bias created by the teacher asking students in person. For example, a response that addresses the wording of the question, voluntary response, or sampling variability does not satisfy component 1.
- To satisfy component 2 the response needs to explicitly contrast what the students say with what they do.
- Evidence used to address component 3 cannot also be used to address component 2. For example, a response that says “Students might lie, producing an estimate that is too high” addresses the effect of the bias on the point estimate but should not be combined with the statement about students lying to infer that students do not actually recycle. However, a response that says “Students may lie and say yes, producing an estimate that is too high” satisfies both components 2 and 3.
- If the response is clearly about the population proportion and not about the point estimate, component 3 cannot be satisfied.
- Statements such as “the interval will be too high” do not satisfy component 3 because they don’t specifically address the point estimate.

2018 SCORING GUIDELINES**Question 2 (continued)**

Part (c) is scored as follows:

Essentially correct (E) if the response gives an answer of 150 in (c-i) and gives an answer of 0.58 (or equivalent) in (c-ii).

Partially correct (P) if the response gives an answer of 150 in (c-i) and gives an answer of 0.42 (or equivalent) in (c-ii);

OR

if the response does not give an answer of 150 in (c-i) but gives an answer of 0.58 (or equivalent) with supporting work in (c-ii).

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

Notes:

- In part (c-i) the answer must be a single number. Responses such as “at least 150” or “147–153” are incorrect. However, responses such as “about 150” or “ ≈ 150 ” are acceptable.
- In part (c-ii) the proportion can be described verbally (e.g., “87 out of 150”).
- In part (c-ii) if the response clearly indicates that 0.58 (or 0.42) is the population proportion, lower the overall score in part (c) by one level (that is, from E to P, or from P to I). Using probability notation such as $P(\text{yes})$ does not clearly indicate a population proportion.
- In part (c-ii) if the response includes a point estimate of 0.58 or 0.42 but uses a confidence interval as the final answer, lower the overall score in part (c) by one level (that is, from E to P, or from P to I).
- If the answer is incorrect in part (c-i) and the answer in part (c-ii) uses numerator = 87 and denominator = 300 – answer to (c-i), the response should be scored P.

2018 SCORING GUIDELINES**Question 2 (continued)****4 Complete Response**

Three parts essentially correct

3 Substantial Response

Two parts essentially correct and one part partially correct

2 Developing Response

Two parts essentially correct and no parts partially correct

OR

One part essentially correct and one or two parts partially correct

OR

Three parts partially correct

1 Minimal Response

One part essentially correct

OR

No parts essentially correct and one or two parts partially correct

2018 SCORING GUIDELINES

Question 3

Intent of Question

The primary goals of this question were to assess a student's ability to (1) compute a probability based on a weighted mixture of two populations; (2) compute a conditional probability; and (3) recognize a binomial random variable and compute the probability associated with it.

Solution

Part (a):

Let L denote left-handed, M denote multiple birth, and S denote single birth.
The probability that a randomly selected child born in the region is left-handed is:

$$P(L) = P(M)P(L | M) + P(S)P(L | S) = (0.035)(0.22) + (0.965)(0.11) = 0.0077 + 0.10615 = 0.11385.$$

Part (b):

From part (a), $P(L) = 0.11385$. Therefore,

$$P(M | L) = \frac{P(L \text{ and } M)}{P(L)} = \frac{(0.035)(0.22)}{0.11385} = \frac{0.0077}{0.11385} \approx 0.0676.$$

Part (c):

Let X represent the number of children who are left-handed in a random sample of 20 children from the region. X has a binomial distribution with $n = 20$ and $p = 0.11385$ (found in part (a)). Using the binomial distribution,

$$\begin{aligned} P(X \geq 3) &= 1 - P(X \leq 2) \\ &= 1 - \binom{20}{0}(0.11385)^0(0.88615)^{20} - \binom{20}{1}(0.11385)^1(0.88615)^{19} - \binom{20}{2}(0.11385)^2(0.88615)^{18} \\ &\approx 1 - 0.598 \approx 0.402. \end{aligned}$$

2018 SCORING GUIDELINES**Question 3 (continued)****Scoring**

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

Part (a) is scored as follows:

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

Partially correct (P) if the response provides a reasonable strategy for finding the probability, such as a formula or tree diagram, but uses one or more inappropriate values;

OR

if the response gives the correct probability but not enough work is shown to determine how that probability was found.

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

Note: A reasonable strategy needs to include summing the results of two multiplications.

Part (b) is scored as follows:

Essentially correct (E) if the probability is computed correctly, with work shown that includes appropriate numerical values for both the numerator and denominator.

Partially correct (P) if the response includes a numerator and denominator in calculating the conditional probability, with one appropriate term (numerator or denominator) and the other inappropriate.

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

Note: Appropriate values include incorrectly calculated values from part (a).

2018 SCORING GUIDELINES

Question 3 (continued)

Part (c) is scored as follows:

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

1. Uses a calculation based on the binomial distribution to find the probability of the number of children in the sample who are left-handed.
2. Specifies appropriate values for n and p .
3. Uses correct endpoint value for the probability.
4. Uses correct direction to calculate the probability of at least three left-handed children.
5. Correctly calculates a binomial probability consistent with the previous work.

Partially correct (P) if the response satisfies component 1 and only two or three of the other four components;
OR

if components 2, 3, 4, and 5 are met, and the response does not explicitly indicate the binomial distribution is used by name or formula.

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

Notes:

- “Appropriate” values include incorrectly calculated values from part (a) or a recalculated probability from part (b).
- An unlabeled numerical value in a calculator statement cannot be used to satisfy a component.
- A response which calculates $P(X \leq 3)$ satisfies component 3 but does not satisfy component 4.
- A normal approximation to the binomial is not appropriate because $np = 20 \times 0.11385 = 2.277 < 5$.
A response using the normal approximation can score at most P. To earn a score of P, the response must include all of the following:
 - a correct mean and standard deviation based on the binomial parameters
 - clear indication of boundary and direction with a z-score or diagram
 - the probability computed correctly

Notes for all parts:

- If the resulting probability or part of the calculation of the probability uses a value that is not between 0 and 1, inclusive, the score for that part is lowered by one level (that is, from E to P, or from P to I).
- An arithmetic or transcription error in a response can be ignored if correct work is shown. For example, $0.0077 + 0.10615 = 0.1385$.

2018 SCORING GUIDELINES**Question 3 (continued)****4 Complete Response**

Three parts essentially correct

3 Substantial Response

Two parts essentially correct and one part partially correct

2 Developing Response

Two parts essentially correct and no parts partially correct

OR

One part essentially correct and one or two parts partially correct

OR

Three parts partially correct

1 Minimal Response

One part essentially correct

OR

No parts essentially correct and two parts partially correct

2018 SCORING GUIDELINES**Question 4****Intent of Question**

The primary goals of this question were to assess a student's ability to (1) determine whether a cause-and-effect conclusion can be made based on how a study was conducted and (2) set up, perform, and interpret the results of a hypothesis test, in the context of the problem.

Solution**Part (a):**

Yes, it would be reasonable to conclude that the new procedure causes a reduction in recovery time, for patients similar to those in the study. The patients in the study were randomly assigned to the two procedures, which reduces the chance that confounding variables will affect the results. Therefore the statistically significant reduction in mean recovery time can be attributed to the new procedure being superior to the standard procedure.

Part (b):

Step 1: State a correct pair of hypotheses.

Let μ_S represent the mean recovery time among all patients similar to those in the study if they were to receive the standard treatment.

Let μ_N represent the mean recovery time among all patients similar to those in the study if they were to receive the new treatment.

The hypotheses to be tested are $H_0 : \mu_S = \mu_N$ versus $H_a : \mu_S > \mu_N$.

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

The appropriate procedure is a two-sample t -test for a difference between means.

Because this is an experiment, the first condition is that subjects were randomly assigned to one treatment group or the other. In this case the condition is satisfied because we were told that the subjects were randomly assigned to either the standard or new procedure.

The second condition is that the recovery times of the two populations are normally distributed or the sample sizes are sufficiently large to presume that the distribution of the difference in the sample means is approximately normal. In this case the condition is met because the sample sizes of 110 and 100 are both sufficiently large.

2018 SCORING GUIDELINES

Question 4 (continued)

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

$$\text{The test statistic is } t = \frac{\bar{x}_S - \bar{x}_N}{\sqrt{\frac{s_S^2}{n_S} + \frac{s_N^2}{n_N}}} = \frac{217 - 186}{\sqrt{\frac{34^2}{110} + \frac{29^2}{100}}} \approx 7.13.$$

The p -value is the area greater than 7.13 for a t -distribution with $df = 207.18$, which is essentially 0 (8.36×10^{-12}).

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

Because the p -value is very small, we have sufficient evidence to conclude that for patients similar to the ones in the study, those receiving the new procedure would have less recovery time, on average, than those receiving the standard procedure.

Scoring

This question is scored in three sections. Section 1 consists of part (a); section 2 consists of step 1, step 2, and the test statistic in step 3 in part (b); and section 3 consists of the p -value in step 3 and step 4 in part (b). Sections 1, 2, and 3 are each scored essentially correct (E), partially correct (P), or incorrect (I).

Section 1 is scored as follows:

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

1. Correctly states that it is reasonable to make a causal conclusion.
2. Justifies the causal conclusion based on random assignment of patients to procedures (or procedures to patients);
OR
justifies the causal conclusion by stating that a randomized experiment was conducted.
3. Includes the context of the situation.

Partially correct (P) if the response satisfies component 1 *AND* provides WEAK justification of the causal conclusion by stating that there was random assignment or a randomized experiment was conducted, but with no context;

OR

by stating that an experiment was conducted or there was assignment (without the word “randomized”) *AND* the response includes context of the situation;

OR

by stating that the study design reduces the chance of confounding variables or balances the effects of uncontrolled variables across both groups in context without explicitly referring to the random assignment.

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

2018 SCORING GUIDELINES

Question 4 (continued)

Notes:

- If the response states that it is *not* reasonable to make a causal conclusion because the result could have been due to random chance *AND* explains that there is evidence for a causal conclusion based on random assignment of patients to procedures or by stating that a randomized experiment was conducted, then the response is scored E.
- If the response discusses aspects of an experiment other than random assignment (such as, control, replication, or large samples), then those aspects are considered extraneous and the response can be scored E unless those aspects are incorrect for this study (such as, blocking is a requirement, or the study used blocking, or the study used a placebo) in which case the score should be lowered one level (that is, from E to P, or from P to I).
- If the response correctly states in context that it is reasonable to make a causal conclusion but includes incorrect or contradictory justification (such as, random selection of patients), then the response is scored I.

Section 2 is scored as follows:

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

1. Parameters are defined correctly.
2. Hypotheses imply equality in the null and correct direction in the alternative.
3. Correct test is identified by name or formula.
4. Correct test statistic for a difference in means is calculated.

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

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

Notes:

- If standard symbols are used for the parameters with appropriate group labels (such as, μ_S, μ_N), component 1 is satisfied.
- If the correct test is identified, but the response states an incorrect formula or uses incorrect notation in the formula, component 3 is not satisfied.
- A pooled two-sample t -test is acceptable for component 3, but the student must also state and comment on the plausibility of the equal population variances assumption.
- If the response identifies a z -test for equal means as the correct test identification, component 3 is not satisfied but component 4 could be satisfied.

Confidence Interval approach:

- If a single two-sample t -interval for the difference in means is used, components 3 and 4 can be satisfied. Component 3 is satisfied if the t -interval is correctly identified by name or formula. Component 4 is satisfied if the correct interval is calculated. If an alpha level is stated, then an appropriate adjustment to the confidence level must be made because the appropriate test is one-sided.
- If two one-sample t -intervals are used, while not a recommended approach, component 3 is not satisfied but component 4 could be satisfied. Component 4 is satisfied if both intervals are calculated correctly.

2018 SCORING GUIDELINES

Question 4 (continued)

Section 3 is scored as follows:

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

1. Makes reference to an approximately correct p -value that is consistent with the test statistic and alternative hypothesis for a difference in means.
2. Correctly justifies the conclusion based on the size of the p -value or the test statistic.
3. Correctly states the conclusion in context.

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

Incorrect (I) if the response does not meet the criteria for E or P or includes a justification not based on the inferential results.

Notes:

Component 1:

- Is satisfied if the response makes reference to a large test statistic without referring to a p -value.

Component 2:

- No alpha level is needed to provide justification of the conclusion based on the size of the p -value.
- Is satisfied if the response states the p -value without reference to size, but it is contiguous to the conclusion and clearly indicates a continuous train of thought.
- A correct interpretation of the p -value with a complete explanation that obtaining a test statistic at least this extreme is unlikely due to chance alone is considered justification based on the size of the p -value.
- If an incorrect interpretation of the p -value is given, the score is lowered one level (that is, from E to P, or from P to I).
- A decision about the null hypothesis (reject H_0 or fail to reject H_0) is not required, but if an incorrect decision is stated based on the given p -value then component 2 is not satisfied.
- If a rejection region approach is used, a reasonable critical value replaces the p -value.

Component 3:

- A correct conclusion must be related to the alternative hypothesis in order to satisfy component 3.
- The following responses do not satisfy component 3:
 - States or implies that the null hypothesis is *accepted*
 - States or implies that the alternative hypothesis has been *proven*
 - States the conclusion in past tense (unless the response did not satisfy a component of section 2 for the use of past tense)

2018 SCORING GUIDELINES

Question 4 (continued)

Confidence Interval approach:

- If a single two-sample t -interval for the difference in means is used:
 - Component 1 is satisfied if the response indicates that zero is either included or not included in the calculated interval.
 - Component 2 is satisfied if the response indicates that the bounds are either both above or both below zero (consistent with alternative hypothesis) and uses that as justification for the conclusion.
 - Component 3 is satisfied if the conclusion is stated in context.
- If two one-sample t -intervals are used (which is not recommended) the response is scored at most P if all three components are satisfied, otherwise scored I:
 - Component 1 is satisfied if the response states that the intervals do not overlap.
 - Component 2 is satisfied if the conclusion indicates that the confidence interval for the new procedure lies below the confidence interval for the standard procedure.
 - Component 3 is satisfied if the conclusion is stated in context.

Note: If the three sections of the response are scored as E, to earn a score of 4 as a complete response, both conditions in step 2 must be correctly stated and justified. Additional condition(s) inappropriate for a two-sample t -test must not be stated. Otherwise, the response earns a score of 3 a substantial response.

2018 SCORING GUIDELINES

Question 4 (continued)

4 Complete Response

Three sections essentially correct with conditions for inference

3 Substantial Response

Three sections essentially correct without conditions for inference

OR

Two sections essentially correct and one section partially correct

2 Developing Response

Two sections essentially correct and no sections partially correct

OR

One section essentially correct and one or two sections partially correct

OR

Three sections partially correct

1 Minimal Response

One section essentially correct

OR

No sections essentially correct and one or two sections partially correct

2018 SCORING GUIDELINES

Question 5

Intent of Question

The primary goals of this question were to assess a student's ability to (1) determine which of two histograms represents data with a larger median; (2) calculate the mean of a combined data set when the separate means and sample sizes are known; and (3) calculate the probability that an individual randomly chosen from a finite population will have a value within one standard deviation of the mean, when provided with values for the mean, standard deviation, and all members of the population.

Solution**Part (a):**

The median teaching year for High School A is any value with 100 data values at or below it and 100 data values at or above it. The median teaching year for High School B is the 111th value in the ordered list of values. For High School A the median is in the interval that starts at 7 and ends just before 10, because there are only 94 data values below 7 and 106 data values of at least 7. Therefore the median cannot be less than 7. For High School B the median is in the interval that starts at 4 and ends just before 7 because there are more than half (113) of the data values less than 7. Therefore the median must be less than 7. So High School A must be the one with a median of 7, and High School B must be the one with a median of 6.

Another way to determine which school has the median of 7 is to notice that the distribution for High School B is highly skewed to the right, whereas the distribution for High School A is bimodal with a few possible outliers on the right. A distribution that is highly right-skewed is likely to have a substantially larger mean than median. The mean of both distributions is given as 8.2 years, so it makes sense that the highly right-skewed distribution (High School B) is the one with the bigger gap between the mean and median and, therefore, the one with the lower median of 6.

Part (b):

The mean for the original 200 teachers was given as 8.2 years, and the mean for the additional 18 teachers is 2.5 years. Therefore the mean for the combined data set is:

$$\frac{(200)(8.2) + (18)(2.5)}{200 + 18} = \frac{1,640 + 45}{218} \approx 7.73 \text{ years.}$$

Part (c):

The interval mean plus or minus 1 standard deviation on either side of the mean is 8.2 ± 7.2 , or from 1.0 year to 15.4 years. Because teaching year is recorded as an integer, the interval includes teaching years 1 to 15. The number of teachers in that interval can be found by adding the heights of the five bars in the histogram for the intervals from 1 to 16, which includes $79 + 34 + 28 + 29 + 19 = 189$. Therefore the probability is $\frac{189}{221} \approx 0.8552$.

2018 SCORING GUIDELINES**Question 5 (continued)****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 satisfies the following three components:

1. States that the median is 6 for High School B and the median is 7 for High School A.
2. Provides a reasonable explanation of how the decision was made.
3. Provides the definition of the median or explicitly applies the definition of a median as a criterion in reaching their decision.

OR

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

1. States that the median is 6 for High School B and the median is 7 for High School A.
2. States that High School B shows a skewed distribution (or High School A shows a less skewed distribution).
3. Provides a reasonable explanation of how the more skewed distribution (High School B) would be the one with a larger separation between the mean and median.

Partially correct (P) if the response satisfies the first component and only one of the other two components required for E.

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

Note: An incorrect statistical statement in the response will result in E being lowered to P, but not P being lowered to I. For example,

- If either distribution is described as left skewed, normal, or approximately normal;
- If the discussion would indicate a median different than 7 for High School A or a median different than 6 for High School B.

Part (b) is scored as follows:

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

1. The correct answer that the mean is 7.73.
2. Enough work to show that the answer was obtained as a weighted average of the two individual means.

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

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

2018 SCORING GUIDELINES**Question 5 (continued)**

Part (c) is scored as follows:

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

1. Calculates that the appropriate interval is 1 to 15.4 or 1 to 15 teaching years.
2. Correctly sums the counts of data values in the numerator based on the intervals provided.
3. Computes the probability using 221 as the denominator.

Partially correct (P) if the response satisfies only two of the three components;

OR

if the response reports the correct probability (0.8552) without supporting work.

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

Notes:

- If the response attempts to use the Empirical Rule or normal distribution to provide the desired probability, the response is scored I.
- If an incorrect count is shown in component 2, for instance by including the interval from 16 to 19, then component 3 is satisfied if that incorrect count is divided by 221 to find the reported probability.
- It is acceptable if the count is slightly off because of difficulty reading the exact heights of the bars in the histogram.
- If only one of component 2 or component 3 is missing, but the correct probability (0.8552) is reported, the response can be scored E.
- If the response recognizes that all values in the histogram bins up to 16 fall within one standard deviation of the mean and reports the interval as 1 to 16, component 1 is satisfied.

2018 SCORING GUIDELINES**Question 5 (continued)****4 Complete Response**

Three parts essentially correct

3 Substantial Response

Two parts essentially correct and one part partially correct

OR

Part (a) essentially correct and two parts partially correct

2 Developing Response

Two parts essentially correct and no parts partially correct

OR

Part (b) or part (c) 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 one or two parts partially correct

2018 SCORING GUIDELINES

Question 6

Intent of Question

The primary goals of this question were to assess a student's ability to (1) describe what constitutes a Type II error for a specific hypothesis test; (2) specify a rejection region in terms of values of the sample mean; (3) compute the power of a test for a specific value in the alternative hypothesis; (4) recognize the definition of power; and (5) understand the impact of increasing the sample size on the power of a test.

Solution**Part (a):**

A Type II error occurs when the alternative hypothesis is true, but the null hypothesis is not rejected. In this situation a Type II error would happen if the mean systolic blood pressure of the population of employees is greater than 122 mmHg, but the null hypothesis that it is 122 mmHg is not rejected. In other words a Type II error would happen if the mean blood pressure for the population of employees is higher than the national average, but the test does not conclude that it is higher.

Part (b):

The test is one-sided and the standard deviation is known, so the null hypothesis will be rejected if the test statistic $z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} > 1.645$. With $\mu = 122$, $\sigma = 15$, and $n = 100$, we get $\frac{\bar{x} - 122}{1.5} > 1.645$. Therefore, $\bar{x} > 124.4675$.

Part (c):

If the actual population mean is 125, with $\sigma = 15$ and $n = 100$, then the sampling distribution of $\bar{x}$ is approximately normal with mean of 125 and standard deviation $\frac{15}{\sqrt{100}} = 1.5$. Therefore,

$$P(\bar{x} > 124.4675) = P\left(\frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} > \frac{124.4675 - 125}{1.5}\right) = P(z > -0.355) = 0.64.$$

Part (d):

The probability found in part (c) is called the power of the test.

Part (e):

If the sample size is increased from 100 to something larger, the probability of rejecting the null hypothesis when the population mean is 125 will be higher than it is for a sample of size 100. Intuitively, more data provide a higher probability of a correct conclusion. The technical explanation is that the rejection region will still be $z > 1.645$, but the sampling distributions of the sample mean will have a smaller standard deviation; therefore, the minimum value of $\bar{x}$ for which we would reject the null hypothesis would be lower and, in return, the probability the null hypothesis is rejected will increase.

2018 SCORING GUIDELINES

Question 6 (continued)

Scoring

This question is scored in three sections. Section 1 consists of part (a) and part (b), section 2 consists of part (c), and section 3 consists of part (d) and part (e). 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 satisfies the following four components:

1. Part (a) includes in the description of a Type II error the fact that the alternative hypothesis is true, either generically or in context of the situation.
2. Part (a) includes in the description of a Type II error the fact that the null hypothesis is not rejected, either generically or in context of the situation.
3. Part (b) includes a correct z-score for the upper 5 percent tail and indicates the correct direction for the rejection region.
4. Part (b) includes $\mu_{\bar{x}} = 122$, $\sigma_{\bar{x}} = 1.5$, and the resulting $\bar{x}$ value.

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.

Notes:

- If the response in part (a) does not include context, the number of components satisfied is reduced by one (that is, from four to three, or from three to two, and so on). Context includes a reference to units, to blood pressure, to employees, etc.
- If a response in part (a) is clearly referring to an individual's blood pressure as opposed to the mean blood pressure of all employees, neither components 1 nor 2 are satisfied.

Section 2 is scored as follows:

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

1. Recognizes that the null hypothesis will be rejected when $\bar{x} \geq 124.4675$, as found in part (b).
2. Provides the correct sampling distribution for the sample mean when the true mean is 125, including correct values for the mean and standard deviation, either explicitly or by plugging them into the test statistic formula.
3. Provides evidence of using the normal curve and finds the correct probability value.

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

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

Notes:

- Components 1 and 3 can still be satisfied if errors made in finding the rejection region in part (b) are carried into part (c).
- A calculator statement that does not include labels for input values does not satisfy component 2 but may still satisfy components 1 and 3.

2018 SCORING GUIDELINES

Question 6 (continued)

Section 3 is scored as follows:

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

1. Part (d) specifies power as the name of the probability.
2. Part (e) correctly states that the probability would be greater.
3. Part (e) correctly implies that the standard deviation of the sampling distribution decreases, either explicitly or by substituting values into a formula.
4. Part (e) indicates the minimum value of $\bar{x}$ for which the null hypothesis is rejected decreases, either explicitly or by substituting values into a formula.

Note: Component 4 can still be satisfied if a response indicates that a maximum value of $\bar{x}$ for which the null hypothesis is rejected increases if this direction is consistent with answers in parts (b) and (c).

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.

2018 SCORING GUIDELINES

Question 6 (continued)

4 Complete Response

Three sections essentially correct

3 Substantial Response

Two sections essentially correct and one section partially correct

2 Developing Response

Two sections essentially correct and no sections partially correct

OR

One section essentially correct and one or two sections partially correct

OR

Three sections partially correct

1 Minimal Response

One section essentially correct

OR

No sections essentially correct and two sections partially correct

OR

Sections 1 and 2 incorrect, and section 3 partially correct with exactly three of the four components satisfied

OR

Section 1 partially correct with exactly three of the four components satisfied, and sections 2 and 3 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

$$\frac{17 + 7}{33} = \frac{24}{33} \approx 0.727. \text{ The proportion of off campus residents who participate in at least one}$$

$$\text{extracurricular activity is } \frac{25 + 12}{67} = \frac{37}{67} \approx 0.552.$$

Part (b):

The graph reveals that on campus residents in this sample are more likely to participate in extracurricular activities than off campus residents. The proportions who participate in two or more extracurricular 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

$$z = \frac{140 - 120}{10.5} \approx 1.90. \text{ The table of standard normal probabilities or a calculator reveals that the}$$

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

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

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):

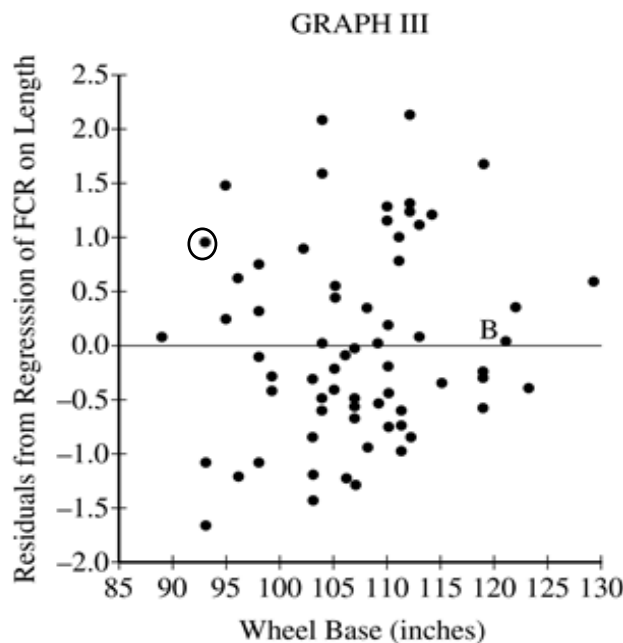
For a car with length 175 inches, the predicted value for the car's FCR, based on the least squares regression line, is

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

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.