

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.	

AP[®] Statistics 2015 Scoring Guidelines

© 2015 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.

2015 SCORING GUIDELINES**Question 1****Intent of Question**

The primary goals of this question were to assess a student's ability to (1) compare features of two distributions of data displayed in boxplots and (2) identify statistical measures that are important in making decisions based on data sets.

Solution**Part (a):**

The median salary is approximately the same for both corporations. The range and interquartile range of the salaries are greater for Corporation A than for Corporation B. The two highest salaries at Corporation A are outliers while Corporation B has no outliers.

Part (b):

- (i) Five years after starting, at least 3 out of 30 (10%) of the salaries at Corporation A are greater than the maximum salary at Corporation B. If I accept the offer from Corporation A, I might be able to make a higher salary at Corporation A than at Corporation B.
- (ii) Five years after starting, the minimum salary at Corporation B is greater than at Corporation A. In fact, at Corporation A it looks like some people are still making the starting salary of \$36,000 and never received a raise in the five years since they were hired. So if I work at Corporation A, I might never receive a raise in salary.

Scoring

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

Part (a) is scored as follows:

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

1. A correct comparison of center.
2. A correct comparison of spread.
3. A discussion of the outliers for Corporation A.
4. The response is in context.

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

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

Note: Any mention of shape should be ignored because complete shape information cannot be determined from a boxplot.

2015 SCORING GUIDELINES

Question 1 (continued)

Part (b) is scored as follows:

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

1. In part (b-i) a relevant statistical measure is identified (or described) or a relevant statistical comparison is provided that supports the choice of Corporation A.
2. In part (b-i) an explanation is provided for why the measure or comparison is relevant.
3. In part (b-ii) a relevant statistical measure is identified (or described) or a relevant statistical comparison is provided that supports the choice of Corporation B.
4. In part (b-ii) an explanation is provided for why the measure or comparison is relevant.

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

Incorrect (I) if the response includes none or one of the four components.

Note: If a response does not provide a statistical measure or comparison in part (b-i) or (b-ii), the second and fourth components can still be satisfied if an acceptable explanation is provided that would follow from a relevant statistical measure or comparison. For example, if the response in part (b-i) only states “At Corporation A, I have the potential to earn a higher salary,” the second component is satisfied.

4 Complete Response

Both parts essentially correct

3 Substantial Response

One part essentially correct and one part partially correct

2 Developing Response

One part essentially correct and one part incorrect

OR

Both parts partially correct

1 Minimal Response

One part partially correct and one part incorrect

2015 SCORING GUIDELINES

Question 2

Intent of Question

The primary goals of this question were to assess a student's ability to (1) use confidence intervals to test a question about a proportion and (2) understand the relationship between sample size and margin of error in a confidence interval for a proportion.

Solution**Part (a):**

- (i) No. The confidence interval is $(0.09, 0.21)$, which includes the value of 0.20. Therefore, it is plausible that the computer program is generating discounts with a probability of 0.20, and the confidence interval does not provide convincing statistical evidence that the program is not working as intended.
- (ii) No. The confidence interval includes values from 0.09 to 0.21, so any value in that interval is a plausible value for the probability that the computer is using to generate discounts.

Part (b):

The formula for computing the margin of error for a proportion includes the square root of the sample size in the denominator. For a random sample that is four times the size of the original sample, the margin of error can be determined by dividing the margin of error of the original sample by two. Therefore, the new margin of error is 0.03.

Part (c):

Using the margin of error of 0.03 obtained from the second sample, the confidence interval for p is 0.15 ± 0.03 or $(0.12, 0.18)$. The interval does not include 0.20, and therefore, there is convincing evidence that the computer program is not working as intended and is not generating discounts with a probability of 0.20.

Scoring

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

Section 1 is scored as follows:

Essentially correct (E) if the response states that because the interval contains 0.20, it does not provide convincing statistical evidence that the computer program is not working as intended.

Partially correct (P) if the response indicates that it is necessary to check whether the value of 0.20 is in the computed interval, but there are errors in implementation. Examples of errors include:

- The response notes that 0.20 is within the interval but does not draw a conclusion.
- The response has an arithmetic error in the computation of the endpoints of the interval but provides a correct conclusion with justification that is consistent with the computed interval.

2015 SCORING GUIDELINES**Question 2 (continued)**

Incorrect (I) if the response does not recognize how to use the confidence interval to check whether the computer is working correctly;

OR

if the response states that the interval shows that the proportion is equal to 0.20;

OR

if the response notes that 0.20 is within the interval and concludes that the program is not working as intended;

OR

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

Section 2 is scored as follows:

Essentially correct (E) if the response concludes that there is not convincing statistical evidence that the computer program generates the discount with a probability of 0.20 *AND* justifies the conclusion by noting that there are values other than 0.20 in the interval.

Partially correct (P) if the response correctly concludes that there is not convincing evidence that the computer program generates the discount with a probability of 0.20, but provides incomplete reasoning to justify the conclusion.

Examples of incomplete reasoning include:

- stating that 0.20 is a plausible value for the proportion of discounts without giving further explanation;
- indicating that having 0.20 in the interval does not prove that 0.20 is the true proportion; and
- providing a generic statistical argument, such as reference to the fact that the null hypothesis should never be accepted, or stating that not rejecting the null hypothesis is not proof that 0.20 is the true proportion of bills discounted.

Note: If an incorrect interval is computed in part (a-i) that does not contain 0.20, and in part (a-ii) the response concludes that the program is not generating discounts with a probability of 0.20 because 0.20 is not in the interval, section 2 is scored as P.

Incorrect (I) if the response states that there is evidence that the computer program generates the discount with a probability of 0.20;

OR

if the response correctly concludes that there is not convincing evidence that the computer program generates the discount with a probability of 0.20 *AND* provides incorrect or no justification;

OR

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

Section 3 is scored as follows:

Essentially correct (E) if the response gives the correct value of 0.03 as the new margin of error by using the correct formula or by providing a correct explanation that recognizes that quadrupling the sample size divides the margin of error from the original sample by two.

Note: If the response relies on the margin of error formula and calculates a value that would round to 0.030, the response is scored as E.

2015 SCORING GUIDELINES

Question 2 (continued)

Partially correct (P) if the response uses the correct margin of error formula or provides a correct explanation that recognizes that quadrupling the sample size divides the margin of error from the original sample by two, but calculates an incorrect margin of error that is less than 0.06;

OR

if the response does not calculate a margin of error but provides a correct explanation that recognizes that quadrupling the sample size divides the margin of error from the original sample by two;

OR

if the response gives the correct margin of error without correct justification.

Incorrect (I) if the response does not recognize in any way that the new margin of error depends on the square root of the sample size;

OR

if the response calculates a margin of error that is greater than or equal to 0.06;

OR

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

Section 4 is scored as follows:

Essentially correct (E) if the conclusion states that there is convincing evidence that the computer program is not working as intended because, based on the new margin of error, the interval does not contain 0.20;

OR

if the conclusion states the intended value of 0.20 is greater than the upper boundary of 0.18;

OR

if the conclusion states the margin of error is smaller than the difference between the sample proportion and the intended long-run proportion of 0.20.

Notes:

- If the margin of error was computed incorrectly in part (b), but a correct answer to part (c) is consistent with this incorrect margin of error, section 4 is scored as E.
- If no specific margin of error or confidence interval was given in parts (b) or (c), but the response provides a complete and correct conclusion with reference to a smaller margin of error (or narrower confidence interval), section 4 is scored as E.

Partially correct (P) if an interval is incorrectly constructed using the margin of error from part (b), but a correct conclusion is given and justified for the computed interval;

OR

if an interval is computed correctly using the margin of error from part (b) and an argument is made based on whether or not 0.20 is in the interval, but the conclusion is incorrect;

OR

if no specific margin of error or confidence interval has been given in parts (b) or (c), but the response concludes that because the margin of error has decreased (or the confidence interval is narrower), 0.20 is not in the interval.

Incorrect (I) if an interval is computed correctly using the margin of error from part (b) and a correct conclusion is given, but no argument is made based on whether or not 0.20 is in the interval;

OR

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

2015 SCORING GUIDELINES**Question 2 (continued)**

Notes:

- If a response includes a confidence level, the level can be ignored because no confidence level was provided.
- If the response bases a conclusion on the relative location of 0.20 within the interval (for example, 0.20 is near the edge of the interval), the response is scored as I.
- A response that provides additional incorrect explanation lowers the score in section 4 by one level (that is, from E to P, or from P to I).

Each essentially correct (E) section counts as 1 point. Each partially correct (P) section 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.

SCORING GUIDELINES**Question****Intent of Question**

The primary goals of this question were to assess a student's ability to (1) perform a probability calculation from a discrete random variable; (2) calculate the expected value of a discrete random variable; (3) perform a conditional probability calculation from a discrete random variable; and (4) use probabilistic thinking to make a prediction about how an expected value will change given a condition about the random variable.

Solution**Part (a):**

The probability that at least one ATM is working when the mall opens is:

$$P(X \geq 1) = 0.21 + 0.40 + 0.24 = 0.85.$$

Part (b):

The expected value of the number of ATMs that are working when the mall opens is:

$$E(X) = 0(0.15) + 1(0.21) + 2(0.40) + 3(0.24) = 1.73 \text{ machines.}$$

Part (c):

probability that all three ATMs are working when the mall opens, given that at least one ATM is working is:

$$P(X = 3 \mid X \geq 1) = \frac{P(X = 3 \text{ and } X \geq 1)}{P(X \geq 1)} = \frac{P(X = 3)}{P(X \geq 1)} = \frac{0.24}{0.85} \approx 0.282$$

Part (d):

Given that at least one ATM is working when the mall opens, the expected value of the number of working ATMs would be greater than the expected value calculated in part (b). By eliminating the possibility of 0 working ATMs, the probabilities for 1, 2, and 3 working ATMs all increase proportionally, so the expected value must increase.

Scoring

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

Part (a) is scored as follows:

Essentially correct (E) if the probability is computed correctly with work shown.

Partially correct (P) if the correct answer is given, but no work is shown;

OR

appropriate work is shown but the answer is incorrect or missing;

OR

SCORING GUIDELINES

Question 3 (continued)

one of the incorrect cumulative probabilities $P(X < 1)$, $P(X \leq 1)$, or $P(X > 1)$ is stated AND computed correctly.

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

Note: The probability can be calculated as $1 - P(X = 0) = 1 - 0.15 = 0.85$.

Part (b) is scored as follows:

Essentially correct (E) if the expected value is computed correctly with work shown.

Partially correct (P) if the correct answer is given, but no work is shown;

OR

if appropriate work is shown but the answer is incorrect or missing.

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

Part (c) is scored as follows:

Essentially correct (E) if the probability is computed correctly, with work shown that includes correct 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 (numerator or denominator) correct in numerical value and the other incorrect;

OR

if the correct answer is given, but no work is shown.

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

Part (d) is scored as follows:

Essentially correct (E) if the response provides the correct answer (greater than) with a reasonable explanation based on the fact that with $X = 0$ eliminated, the probabilities for $X = 1$, $X = 2$, $X = 3$ all increase;

OR

if the response provides the correct answer (greater than) with a reasonable explanation based on fact that with $X = 0$ eliminated, the balance point of the distribution increases;

OR

if the response provides the correct answer (greater than) and the conditional expected value is computed correctly with work shown.

Partially correct (P) if the response provides the correct answer (greater than) with a weak explanation, such as "Yes, because 0 is eliminated";

OR

SCORING GUIDELINES**Question 3 (continued)**

if the response provides the correct answer (greater than) with a reasonable but incorrect attempt to calculate the conditional expected value using a revised probability distribution with only the values $X = 1$, $X = 2$, and $X = 3$, and corresponding probabilities greater than or equal to those in the original probability distribution.

Incorrect (I) if the response does not provide the correct answer (greater than);

OR

if the response provides the correct answer (greater than) with an incorrect explanation or no explanation.

Note: The conditional expected value is:

$$1\left(\frac{0.21}{0.85}\right) + 2\left(\frac{0.40}{0.85}\right) + 3\left(\frac{0.24}{0.85}\right) \approx 1(0.247) + 2(0.471) + 3(0.282) = 2.04 \text{ machines.}$$

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 score up or down, depending on the overall strength of the response and communication.

2015 SCORING GUIDELINES

Question 4

Intent of Question

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

Solution

Step 1: States a correct pair of hypotheses.

Let p_{asp} represent the population proportion of adults similar to those in the study who would have developed colon cancer within the six years of the study if they had taken a low-dose aspirin each day. Similarly, let p_{plac} represent the population proportion of adults similar to those in the study who would have developed colon cancer within the six years of the study if they had taken a placebo each day.

The hypotheses to be tested are $H_0 : p_{\text{asp}} = p_{\text{plac}}$ versus $H_a : p_{\text{asp}} < p_{\text{plac}}$ or equivalently, $H_0 : p_{\text{asp}} - p_{\text{plac}} = 0$ versus $H_a : p_{\text{asp}} - p_{\text{plac}} < 0$.

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

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

Because this is a randomized experiment, the first condition is that the volunteers were randomly assigned to one treatment group or the other. The condition is satisfied because we are told that the volunteers were randomly assigned to take a low-dose aspirin or a placebo.

The second condition is that the sample sizes are large, relative to the proportions involved. The condition is satisfied because all sample counts are large enough; that is, 15 with colon cancer in aspirin group, 26 with colon cancer in placebo group, $500 - 15 = 485$ cancer-free in aspirin group, and $500 - 26 = 474$ cancer-free in placebo group.

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

The sample proportions who developed colon cancer are $\hat{p}_{\text{asp}} = \frac{15}{500} = 0.030$ and $\hat{p}_{\text{plac}} = \frac{26}{500} = 0.052$.

The combined sample proportion who developed colon cancer is $\hat{p}_{\text{combined}} = \frac{15 + 26}{500 + 500} = 0.041$.

The test statistic is $z = \frac{0.030 - 0.052}{\sqrt{0.041(1 - 0.041)\left(\frac{1}{500} + \frac{1}{500}\right)}} \approx -1.75$ (-1.7542 from calculator).

The p -value is $P(Z \leq -1.75) = 0.0401$ (0.0397 from calculator), where Z has a standard normal distribution.

2015 SCORING GUIDELINES

Question 4 (continued)

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

Because the p -value is less than the given significance level of $\alpha = 0.05$, we reject the null hypothesis. The data provide convincing statistical evidence that the proportion of all adults similar to the volunteers who would develop colon cancer if they had taken a low-dose aspirin every day is less than the proportion of all adults similar to the volunteers who would develop colon cancer if they had not taken a low-dose aspirin every day.

Scoring

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

Step 1 is scored as follows:

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

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

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

Note: Either defining the parameters in context, or simply using common parameter notation with subscripts clearly relevant to the context, such as p_{asp} and p_{plac} , is sufficient.

Step 2 is scored as follows:

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

1. Identifies the correct test procedure (by name or by formula).
2. Notes that the use of random assignment satisfies the randomness condition.
3. Checks for approximate normality of the test statistic by citing that all four counts are larger than some standard criterion such as 5 or 10.

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

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

Notes:

- For the randomness component, it is (minimally) acceptable to say “random assignment — check” but not acceptable to say “random — check” or “SRS — check.” The important concept here is that it is random assignment, and not random sampling, that is required. If the response implies that the study used a random sample, the randomness component is not satisfied, regardless of whether random assignment is correctly addressed.
- The normality check may use the expected counts under the null hypothesis in place of observed counts.

2015 SCORING GUIDELINES

Question 4 (continued)

Step 3 is scored as follows:

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

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

OR

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

OR

if the response reports the correct p -value but no calculations or test statistic are shown.

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

Note: The p -value is considered correct if it is consistent with the alternative stated in the response and the calculated test statistic, even if those are incorrect.

Step 4 is scored as follows:

Essentially correct (E) if the response provides a correct conclusion in context, with justification based on linkage between the p -value and the given $\alpha = 0.05$.

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:

- The conclusion must be related to the alternative hypothesis.
- If the p -value is incorrect, then step 4 is scored as E if the response includes proper linkage and a conclusion in context consistent with that p -value.
- If the p -value is less than 0.05, wording that states or implies that the alternative hypothesis is *proven* lowers the score one level (that is, from E to P or P to I) in step 4.
- If the p -value is incorrect and greater than 0.05, wording that states or implies that the null hypothesis is *accepted* lowers the score one level (that is, from E to P or P to I) in step 4.

2015 SCORING GUIDELINES**Question 4 (continued)**

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.

SCORING GUIDELINES**Question****Intent of Question**

The primary goals of this question were to assess a student's ability to (1) use the information provided by a scatterplot to describe the relationship between two quantitative variables; (2) interpret and use the information given by lines displayed on a scatterplot; and (3) use a regression equation to estimate a predicted value of y for a given x value.

Solution**Part**

There is a moderately strong, positive, linear relationship between height and arm span so that taller students tend to have longer arm spans.

Part (b):

The line in Graph 2 is the one that is helpful. For each student, the graph illustrates whether arm is equal to height (points on the line), arm span is greater than height (points above the line), or arm span is less than height (points below the line).

Classification	Square	Tall Rectangle	Short Rectangle
Frequency	3	4	5

Part (c):

The predicted arm span is $\hat{y} = 11.74 + 0.8247x = 11.74 + 0.8247(61) = 62.05$ inches.

Scoring

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

Part (a) is scored as follows:

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

1. The relationship is approximately linear.
2. The relationship is positive.
3. There is a moderate to strong association (or relationship).
4. The response is given in context.

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

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

SCORING GUIDELINES

Question 5 (continued)

Notes:

- Correlation alone is not sufficient to indicate a linear relationship. However, linear correlation sufficient to indicate a linear relationship.
- Reporting that the correlation coefficient is 0.81, the correlation coefficient is close to 1.0, or some other value or range of values for the correlation coefficient is not sufficient to satisfy component 3.

Part (b) is scored as follows:

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

- The line in Graph 2 is selected.
- A reasonable explanation for selecting the line in Graph 2 that links the body shapes to the regions defined by the line is given. The explanation should indicate that square shapes (arm span is equal to height) are represented by points on the line, tall rectangle shapes (arm span less than height) are represented by points below the line, and short rectangle shapes (arm span is greater than height) are represented by points above the line.
- Correct counting of body types demonstrated by reporting correct frequencies (3, 4, and 5) reporting correct proportions $\left(\frac{1}{4} = \frac{3}{12}, \frac{1}{3} = \frac{4}{12}, \text{ and } \frac{5}{12}\right)$ for the square, tall rectangle, and short rectangle body types, respectively.
- The correct frequencies (3, 4, and 5) are reported in the table.

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

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

Notes:

- Selecting the regression line on Graph 1 cannot satisfy either component 1 or component 2.
- To satisfy component 1, it is sufficient to refer to the $y = x$ line without explicitly mentioning Graph 2.
- Use of incorrect labels, such as regression line or least-squares regression line, in referring to the $y = x$ line in Graph 2 is an incorrect use of terminology that should be strongly discouraged, but it is ignored in this scoring rubric.
- The explanation for selecting Graph 2 is acceptable if it explicitly includes at least two of the following: (i) square body shapes (arm span equal to height) are represented by points on the line, (ii) tall rectangle body shapes (arm span less than height) are represented by points *below* the line, or (iii) short rectangle body shapes (arm span greater than height) are represented by points *above* the line.
- Incorrectly counting the points on the $y = x$ line in Graph 2 is considered a minor error. Frequencies (2, 4, and 5) are accepted for both component 3 and component 4. Reporting corresponding proportions $\left(\frac{2}{11}, \frac{4}{11}, \frac{5}{11}\right)$ or $\left(\frac{2}{12}, \frac{4}{12}, \frac{5}{12}\right)$ in the table satisfies component 3 but does not satisfy component 4.
- Frequencies reported in the table as (2, 5, 4) or (3, 5, 4), interchanging the counts for tall and short rectangle body shapes, satisfies component 3, but does not satisfy component 4. Reporting corresponding proportions, $\left(\frac{2}{11}, \frac{5}{11}, \frac{4}{11}\right)$ or $\left(\frac{3}{12}, \frac{5}{12}, \frac{4}{12}\right)$, in the table does not satisfy either component 3 or component 4.

SCORING GUIDELINES**Question 5 (continued)**

Part (c) is scored as follows:

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

1. Correct formula for predicting arm span with 61 inserted for x .
2. Correct value for the predicted arm span.
3. Units for the predicted arm span given as inches.

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

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

Notes:

- Any value for the predicted arm span between 61.5 and 62.5 is acceptable; values outside this interval do not satisfy component 2. This allows for inaccuracy in obtaining a value for the predicted arm span directly from the line displayed on Graph 1 and for rounding in applying the prediction formula; but it excludes predictions based on the $y = x$ line and other unreasonable predictions.
- Reporting a prediction equation that is similar to the equation given in the stem, with 61 inserted for x , satisfies components 1 and 2 if the value of the predicted arm span is between 61.5 and 62.5 inches. Otherwise, neither of those two components is satisfied. This could occur if a student enters the data from the graph into a calculator to compute the least squares regression line, instead of using the formula provided in the stem.

4 Complete Response

All three parts essentially correct

3 Substantial Response

Two parts essentially correct and one part partially correct

2 Developing Response

Two parts essentially correct and one part incorrect

OR

One part essentially correct and two parts partially correct

OR

One part essentially correct, one part partially correct, and one part incorrect

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

2015 SCORING GUIDELINES

Question 6

Intent of Question

The primary goals of this question were to assess a student's ability to (1) describe how sample data would differ using two different sampling methods; (2) describe the sampling distribution of the sample mean for two different sampling methods; and (3) choose the sampling method that will result in the best estimate of the population mean.

Solution**Part (a):**

No, a sample obtained using Method 2 will not be representative of all tortillas made that day. The sample obtained using Method 2 will only represent the tortillas from one production line, not from the entire population because the distributions of diameters for the two production lines are different.

Part (b):

Method 1 was most likely used to select this sample. The bimodal shape in the histogram of sample data indicates that tortillas were selected from both production lines, which is what would happen using Method 1. Method 2 would be likely to produce a unimodal distribution of diameters centered at either 5.9 inches or 6.1 inches.

Part (c):

Method 2 would result in less variability in the sample of 200 tortillas on a given day because the sample comes from only one production line. Because the distributions of diameters are not the same for the two production lines, selecting tortillas from both lines as in Method 1 would result in more variable sample data.

Part (d):

The sampling distribution of the sample mean diameter for samples obtained using Method 1 would be approximately normal with mean 6 inches and standard deviation $\frac{0.11}{\sqrt{200}} \approx 0.0078$ inch.

Part (e):

Method 1 would result in less variability in the sample means over the 365 days, because with Method 2, roughly half of the sample means will be around 5.9 inches and the other half will be around 6.1 inches. With Method 1, however, the sample means will all be very close to 6 inches, as indicated by the standard deviation in part (d).

Part (f):

Method 1 is more likely to produce a sample mean close to 6 inches. Because the sample mean is an unbiased estimator for both methods, the manager should pick the method that would result in less variability in the distribution of the sample mean. Based on the answer to part (e), Method 1 results in less variability in the distribution of the sample mean.

2015 SCORING GUIDELINES

Question 6 (continued)

Scoring

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

1. In part (a) the response says no *AND* either argues that the sample will only be selected from one production line and not the entire population (that is, the sample will only represent one production line) *OR* argues that the tortillas from the two productions lines are different.
2. In part (b) the response chooses Method 1 *AND* refers to a relevant characteristic of the histogram (shape, center, or variability) that matches what would be expected when using Method 1 (or that does not match what would be expected when using Method 2).
3. In part (c) the response chooses Method 2 *AND* either justifies by stating that the sample comes from only one production line (does not come from both production lines) *OR* justifies by comparing the possible range of diameters for the two methods.

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.

Note: If a response includes more than one justification in an individual part, score the weaker of the two justifications. For example, a response for part (a) that says “No, because only one line was selected and because the sample size was too small” does not satisfy the first component because the sample size argument is incorrect.

Section 2 is scored as follows:

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

1. The shape is approximately normal.
2. The mean is 6 inches.
3. The standard deviation is $\frac{0.11}{\sqrt{200}}$ inch.

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.

Note: It is not necessary to include units (inches) for the mean or standard deviation.

Section 3 is scored as follows:

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

1. In part (e) the response chooses Method 1 *AND* describes the sampling distribution of the sample mean for Method 2 as having some *sample means* close to the mean of production line A (5.9 inches) and the other *sample means* close to the mean of production line B (6.1 inches).

2015 SCORING GUIDELINES

Question 6 (continued)

2. In part (f) the response chooses Method 1 *AND*
- refers to a correct answer in part (e);
- OR
- describes the sampling distribution of the sample mean for Method 2 as having some *sample means* close to the mean of production line A (5.9, less than 6) and the other *sample means* close to the mean of production line B (6.1, greater than 6);
- OR
- argues that on a single day, it would be preferable to get a sample with a mean around 6 rather than getting a sample that would have a mean around 5.9 (less than 6) or a mean around 6.1 (greater than 6).

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

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

Notes:

- In part (e), the response must be clear that there is *more than one* mean being described (365 sample means).
 - Correct: The sample means will be around 5.9 or 6.1.
 - Not correct: The sample mean will be around 5.9 or 6.1.
- Parts (e) and (f) are not satisfied if the response does not imply the sample means *vary* from the population means in both methods (or does not imply that the sample mean varies from the population mean when making the single day argument in part (f)). However, if the mistake is made in part (e), do not penalize the response in part (f) for the same mistake.
 - Correct: The sample means will be around 5.9 or 6.1 instead of close to 6.
 - Not correct: The sample means will be 5.9 or 6.1 instead of close to 6.
 - Not correct: The sample means will be 6 instead of around 5.9 or 6.1.

4 Complete Response

All three sections essentially correct

3 Substantial Response

Two sections essentially correct and one section partially correct

2 Developing Response

Two sections essentially correct and one section incorrect

OR

One section essentially correct and either one or two sections partially correct

OR

All three sections partially correct

1 Minimal Response

One section essentially correct and two sections incorrect

OR

Two sections partially correct and one section incorrect

AP[®] Statistics 2014 Scoring Guidelines

© 2014 The College Board. College Board, Advanced Placement Program, AP, AP Central, and the acorn logo are registered trademarks of the College Board.

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

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

2014 SCORING GUIDELINES**Question 1****Intent of Question**

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

Solution**Part (a):**

The proportion of on campus residents who participate in at least one extracurricular activity is

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

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

Part (b):

The graph reveals that on campus residents in this sample are more likely to participate in 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):

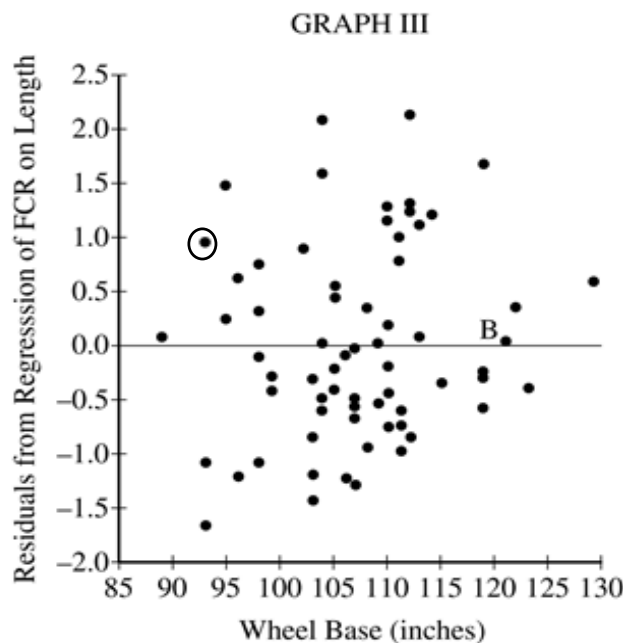
For a car with length 175 inches, the predicted value for the car's FCR, based on the least squares regression line, is

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

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.