

Question 5: Multi-Focus

4 points

General Scoring Notes

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

Model Solution	Scoring
A i. Let random variable X represent the number of bedrooms in a randomly selected newly built house in the 2024 sample from Country B. The probability that a randomly selected house from the 2024 sample had fewer than 3 bedrooms is the probability that the house had either 1 or 2 bedrooms, which is $P(X < 3) = P(X = 1) + P(X = 2)$ $P(X < 3) = 0.12 + 0.22$ $P(X < 3) = 0.34.$ ii. The average number of bedrooms per house for the sample of newly built houses in 2024 is $E(X) = 1(0.12) + 2(0.22) + 3(0.28) + 4(0.22) + 5(0.14) + 6(0.02)$ $E(X) = 3.10 \text{ bedrooms.}$	Essentially correct (E) if the response satisfies at least three of the following four components: - In part A (i) the response correctly calculates the probability of 0.34. - In part A (i) the response provides supporting work or justification for component 1. - In part A (ii) the response correctly calculates the mean of 3.10. - In part A (ii) the response provides supporting work or justification for component 3. Partially correct (P) if the response satisfies only two of the components required for E. Incorrect (I) if the response does not meet the criteria for E or P.

Scoring Notes:

- An arithmetic or transcription error in a response can be ignored if correct work is shown.
- Supporting work for finding the expected value must include at least two of the terms in the equation to show the pattern, such as $1(0.12) + 2(0.22) + \dots$.
- A response to part A (ii) that indicates use of the appropriate formula for the mean of a discrete random variable in words (e.g., “the sum of x times p of x for all values of x ”) or describes an appropriate method (e.g., using the given distribution to create a hypothetical sample of 100 and then calculating the mean) may satisfy component 4.

Model Solution	Scoring
B i. Let μ = the population mean number of bedrooms in newly built houses in 2024 from Country B. The null hypothesis is $H_0: \mu = 2.9$ and the alternative hypothesis is $H_a: \mu \neq 2.9$. ii. A Type I error would be determining that the population mean number of bedrooms in newly built houses in 2024 from Country B is not equal to 2.9 when it is in fact 2.9.	Essentially correct (E) if the response satisfies the following four components: 1. In part B (i) the response states the correct equality for the null hypothesis with the value 2.9. 2. In part B (i) the response states the correct two-sided alternative hypothesis consistent with the null hypothesis. 3. In part B (i) the response provides sufficient context for the parameter by including a reference to the population mean AND the response variable (bedrooms) AND the sampling units (newly built houses). 4. In part B (ii) the response correctly defines a Type I error in context. Partially correct (P) if the response satisfies only three of the four components required for E. Incorrect (I) if the response does not meet the criteria for E or P.

Scoring Notes:

- Components 1 and 2 may be satisfied without regard to the symbol (or lack of symbol) used to represent the population parameter.
- Neither context nor the concept of the population is required to satisfy components 1 or 2.
- A response that states the hypotheses in words (e.g., “the null hypothesis is that the mean is 2.9, and the alternative hypothesis is that the mean is not equal to 2.9”) may satisfy components 1 and 2.
- A response that states the hypotheses in words AND refers to the population in context (e.g., “the null hypothesis is that the population mean number of bedrooms in newly built houses in 2024 from Country B is equal to 2.9, and the alternative hypothesis is that the population mean number of bedrooms in newly built houses in 2024 from Country B is not equal to 2.9”) may satisfy components 1, 2, and 3.
- If the response clearly refers to the **sample** mean instead of the **population** mean using words or a symbol (e.g., $\bar{x}$ or $\hat{\mu}$), then component 3 is not satisfied unless the symbol used is defined as the **population** mean.
- The phrase “mean number of bedrooms” or “mean number of bedrooms for houses” is not sufficient for the population aspect of component 3.
- A response may satisfy the population aspect of component 3 by the following:
 - Referring to the population by using words such as “population,” “all,” or “true” when defining the parameter.
 - Using notation such as μ when defining the hypothesis statements.
- A response that refers to “bedrooms” OR “rooms” may satisfy context for component 4.

Model Solution	Scoring
C Because the value 2.9 is not contained within the 97% confidence interval, the null hypothesis should be rejected. Therefore, there is convincing statistical evidence, at the $\alpha = 0.03$ level of significance, that the population mean number of bedrooms in newly built houses in 2024 from Country B is not equal to 2.9 (or is different than that in 2017).	Essentially correct (E) if the response satisfies the following two components: 1. States a conclusion consistent with and in terms of the alternative hypothesis using nondefinitive language 2. Provides a justification for the conclusion by indicating that 2.9 is not contained within the confidence interval Partially correct (P) if the response satisfies only one of the two components required for E. Incorrect (I) if the response does not meet the criteria for E or P.

Scoring Notes:

- A response that only provides an interpretation of the given confidence interval is scored I.
- A response may satisfy the conclusion aspect of component 1 by using words such as “there is evidence to support the alternative,” “there is statistical evidence that H_a is true,” “I am 97% confident there is a difference in means,” or “because 2.9 is not included in our interval, it is not a plausible value for the mean.”
- If an explicit decision is stated and the conclusion is inconsistent with the decision, component 1 is not satisfied. A response that incorrectly indicates that 2.9 is contained in the confidence interval and then provides an otherwise correct conclusion based on 2.9 being contained in the interval satisfies component 1 but not component 2.
- A response that provides a conclusion that is consistent with an incorrect alternative hypothesis identified in part B satisfies component 1.
- A definitive response that states that the average number of bedrooms is 3.10 (the value of the sample mean or any other number) does not satisfy component 1, even if the response makes additional correct statements about the alternative hypothesis.
- A response that draws a conclusion that is clearly about the sample mean (e.g., a statement about “average number of bedrooms for newly built houses in the study”) does not satisfy component 1.
- If the conclusion includes a definitive statement (e.g., “this proves that we have enough evidence that the mean number of bedrooms in newly built homes in 2024 is 2.9” or “Rodney is correct; the mean number of bedrooms in newly built homes in 2004 is 2.9”), then component 1 is not satisfied.
- A response that reverse engineers the given confidence interval to obtain a standard error using a reasonable critical value for a 97% confidence interval and then uses the standard error to compute a t -statistic and p -value to reach a correct conclusion satisfies component 1 but does not satisfy component 2.

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

Question 3: Focus on Probability and Sampling Distributions**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) The random variable X is the dollar value of the cash prize in a bath fizzy. (i) The proportion of bath fuzzies containing \$1 is equal to the $P(X = \\$1)$ and $P(X = \$1)$ $= 1 - (0.2 + 0.05 + 0.05 + 0.01 + 0.01)$ $= 0.68.$ (ii) The proportion of bath fuzzies that contain at least \$10 is equal to the $P(X \geq \\$10)$ and $P(X \geq \$10)$ $= 0.05 + 0.05 + 0.01 + 0.01$ $= 0.12.$	Essentially correct (E) if the response satisfies at least three of the following four components: 1. Correctly calculates the proportion of bath fuzzies containing \$1 2. Provides correct supporting work for the value calculated in component 1 3. Correctly calculates the proportion of bath fuzzies containing at least \$10 4. Provides correct supporting work for the value calculated in component 3 Partially correct (P) if the response satisfies only two of the four components. Incorrect (I) if the response does not meet the criteria for E or P.

Additional Notes:

- A response that provides a correct percentage, instead of a proportion, may satisfy components 1 and 3, such as 68 percent or 12 percent.
- An arithmetic or transcription error in a response can be ignored if correct work is shown.

Model Solution	Scoring
(b) Given a bath fizzy contains at least \$10, then the probability that it contains \$100 is $P(X = \$100 \mid X \geq \$10)$ $= \frac{0.01}{0.12}$ $\approx 0.0833.$	Essentially correct (E) if the response satisfies the following two components: 1. Correctly calculates the requested probability 2. Shows work consistent with their response to part (a-ii) Partially correct (P) if the response satisfies only one of the two components. Incorrect (I) if the response does not meet the criteria for E or P.

Additional Notes:

- A specific conditional probability statement is not required, but if correctly given should be considered a positive in holistic scoring.
- An arithmetic or transcription error in a response can be ignored if correct work is shown.

Model Solution	Scoring
(c) The expected value of the distribution of X is $E(X) = 1(0.68) + 5(0.2) + 10(0.05)$ $+ 20(0.05) + 50(0.01) + 100(0.01)$ $= \$4.68.$ The expected value is the mean of the cash prizes that result from the long run of many, many trials of randomly selecting bath fuzzies and determining the amount each contains.	Essentially correct (E) if the response satisfies both components 1 and 2 <i>AND</i> at least two of components 3–5: 1. States the correct expected value 2. Shows appropriate work to calculate the expected value 3. Interpretation includes the concept of repeating the selection process over a long period of time 4. Interpretation includes the concept of an average or mean 5. Interpretation includes the context of receiving a cash prize Partially correct (P) if the response does not meet the criteria for E but satisfies two or three of components 1–4. Incorrect (I) if the response does not meet the criteria for E or P.

Additional Notes:

- Supporting work for finding the expected value must include at least two of the terms in the equation to show the pattern, such as $1(0.68) + 5(0.2) + \dots$
- Calculator notation does not satisfy component 2, such as $1 - \text{VAR STATS}(L1, L2)$.
- Responses that satisfy only components 1 and 5 or only components 2 and 5 receive a score of I but should be considered a positive in holistic scoring.
- An arithmetic or transcription error in a response can be ignored if correct work is shown.

Model Solution	Scoring
(d) The expected value of the distribution of X in euros is $4.68(0.89) \approx 4.17$ euros.	Essentially correct (E) if the response correctly calculates the number of euros for the expected value showing work and including units. Partially correct (P) if the response correctly calculates the number of euros for the expected value but is missing either work or units. Incorrect (I) if the response does not meet the criteria for E or P.

Additional Notes:

- A response that does not have at least one decimal place in the final response (e.g., rounded to an integer) should be scored no more than P.
- The response can either use the expected value calculated in part (c) or first convert all the values in the probability distribution to euros to find the new expected value.
- An arithmetic or transcription error in a response can be ignored if correct work is shown.

Scoring for Question 3

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