

AP[®] Statistics

Scoring Guidelines

2019 SCORING GUIDELINES

Question 1

Intent of Question

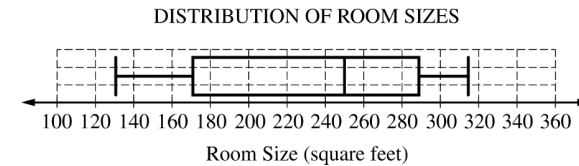
The primary goals of this question were to assess a student's ability to (1) describe features of a distribution of sample data using information provided by a histogram; (2) identify potential outliers; (3) sketch a boxplot; and (4) comment on an advantage of displaying data as a histogram rather than as a boxplot.

Solution**Part (a):**

The distribution of the sample of room sizes is bimodal and roughly symmetric with most room sizes falling into two clusters: 100 to 200 square feet and 250 to 350 square feet. The center of the distribution is between 200 and 300 square feet. The range of the distribution is between 150 and 250 square feet. There are no apparent outliers.

Part (b):

The interquartile range is $IQR = 292 - 174 = 118$ square feet. There are no potential outliers because the minimum room size of 134 square feet does not fall below $Q_1 - 1.5(IQR) = -3$ square feet, and the maximum room size of 315 square feet does not exceed $Q_3 + 1.5(IQR) = 469$ square feet.

**Part (c):**

The histogram clearly shows the bimodal nature of the distribution of room sizes, but this is not apparent in the boxplot.

2019 SCORING GUIDELINES

Question 1 (continued)

Scoring

This question is scored in three sections. Section 1 consists of part (a); Section 2 consists of the outlier determination in part (b); Section 3 consists of the boxplot sketch in part (b) and part (c). Each section is scored as essentially correct (E), partially correct (P), or incorrect (I).

Section 1 is scored as follows:

Essentially correct (E) if the description of the distribution of room sizes satisfies the following four components:

1. The shape is bimodal OR there are two peaks OR there are two clusters.
2. The center is between 200 and 300 square feet.
3. The spread is addressed by stating the range is a value between 150 and 250 square feet OR the interquartile range is a value between 50 and 150 square feet OR all room sizes are between 100 and 350 square feet.
4. The response includes context.

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

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

Notes:

- Shape: Component 1 cannot be satisfied if a response describes the histogram as unimodal or describes the entire histogram as normal or approximately normal.
- Shape: A response that addresses symmetry, while appropriate, does not impact the scoring of section 1.
- Center: A response that states one cluster of the distribution is centered between 150 and 200 square feet and the other cluster is centered between 250 and 300 square feet satisfies both components 1 and 2.
- Center:
 - Responses that address center using interval language such as “the mean of the distribution is *between* 200 and 300” must, for any single measure of center, provide an interval with lower endpoint not below 200 square feet, and with upper endpoint not above 300 square feet to satisfy component 2.
 - Responses that address center using approximate language such as “the median of the distribution is *approximately* 225” must, for any single measure of center, specify a numeric value that is not less than 200 square feet, and that is not greater than 300 square feet to satisfy component 2.
 - Responses that use definitive language such as “the mean of the distribution is 231.4” must identify the corresponding numeric value correctly to satisfy component 2. Specifically, the median of the distribution can be correctly identified as any value between 250 and 253.5 square feet, inclusive; the mean of the distribution is 231.4 square feet; and the center (or average) of the distribution can be any value that is a correct median or mean.

2019 SCORING GUIDELINES

Question 1 (continued)

- Spread: A response recognizing all values in the sample fall between 100 and 350 square feet (or between 134 and 315 square feet) satisfies component 3 *only* for these exact endpoints and need not appeal to a specific measure of spread such as range or interquartile range (IQR).
- Spread:
 - Responses that appeal to a specific measure of spread using interval language, such as “the IQR is *between* 50 and 150,” must provide bounds appropriate to the corresponding measure of spread. For range, the lower endpoint must not be below 150 square feet and the upper endpoint cannot exceed 250 square feet; for IQR, the lower endpoint must not be below 50 square feet, with upper endpoint not to exceed 150 square feet; for standard deviation, the lower endpoint must not be below 25 square feet, with upper endpoint not to exceed 100 square feet.
 - Responses that appeal to a specific measure of spread using approximate language, such as “the range is *approximately* 250,” must specify a numeric value within the bounds appropriate to that measure of spread. For range, the value must be between 150 and 250 square feet (inclusive); for IQR, the value must be between 50 and 150 square feet (inclusive); for standard deviation, the value must be between 25 and 100 square feet (inclusive). Responses that appeal to a specific measure of spread using definitive language, such as “the range of the distribution is 181,” must identify the corresponding numeric value correctly to satisfy component 3. Specifically, the range of the distribution is 181 square feet; the IQR of the distribution is 118 square feet; and the standard deviation of the distribution is 68.12 square feet.

Section 2 is scored as follows:

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

1. Computation of both upper and lower outlier boundary fences that also shows the fences formulas either in words, symbols $Q_1 - 1.5(IQR)$ and $Q_3 + 1.5(IQR)$, or with values substituted from the table $174 - 1.5(118)$ and $292 + 1.5(118)$, or $(174 - 177)$ and $(292 + 177)$.
2. A correct decision regarding the presence of outliers.
3. Correct justification that compares the data with the fences.

Partially correct (P) if the response satisfies only two of the three components OR if the response omits exactly one of the fences but otherwise satisfies all three components.

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

Notes:

- A response that identifies both fence formulas using symbols, but does not substitute values for all symbols, must also include the correct fence values of -3 and 469 to satisfy component 1.
- In place of an appeal to fences, a response may compute outlier bounds representing k standard deviations from the sample mean, where k is a number from 2 to 3 (inclusive), and must include formulas for both endpoints either in words, symbols $\bar{x} \pm k(\text{standard deviation})$, or with values substituted from the table. When $k = 2$ the outlier bounds are $(95.16, 367.64)$; when $k = 3$ the bounds are $(27.04, 435.76)$.

2019 SCORING GUIDELINES

Question 1 (continued)

- A response that identifies the standard deviation bounds using symbols, but that does not substitute values for all symbols, does not satisfy component 1 unless the correct numeric bounds are provided.
- Component 3 is satisfied if the response states the outlier decision criterion: any data values falling outside of the interval from -3 to 469 are potential outliers.

Section 3 is scored as follows:

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

1. A correct sketch of the boxplot.
2. A response for part (c) that indicates the bimodal shape of the room size distribution is apparent in the histogram but not in the boxplot.

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

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

Notes:

- The boxplot must be completely correct to satisfy component 1. Specifically:
 - The minimum is positioned between grid lines at 120 and 140 square feet.
 - Q_1 is positioned between grid lines at 160 and 180 square feet.
 - The median is positioned between grid lines at 240 and 260 square feet.
 - Q_3 is positioned between grid lines at 280 and 300 square feet.
 - The maximum is positioned between grid lines at 300 and 320 square feet.
- If a *mean* is included as a part of the boxplot, component 1 cannot be satisfied.
- A response based on skewness or symmetry does not satisfy component 2.
- A response stating the unimodal OR normal shape of the histogram of room sizes is apparent in the histogram but not in the boxplot will satisfy component 2 *only* if the shape description in section 1 component 1 was also unimodal OR normal, respectively.

2019 SCORING GUIDELINES

Question 1 (continued)

4 Complete Response

Three sections essentially correct

3 Substantial Response

Two sections essentially correct and one section partially correct

2 Developing Response

Two sections essentially correct and no sections partially correct

OR

One section essentially correct and one or two sections partially correct

OR

Three sections partially correct

1 Minimal Response

One section essentially correct

OR

No sections essentially correct and two sections partially correct

2019 SCORING GUIDELINES

Question 2

Intent of Question

The primary goals of this question were to assess a student's ability to (1) identify components of an experiment; (2) determine if an experiment has a control group; and (3) describe how experimental units can be randomly assigned to treatments.

Solution**Part (a):**

Treatments: Sprays with four different concentrations of the fungus (0 ml/L, 1.25 ml/L, 2.5 ml/L, and 3.75 ml/L)

Experimental units: 20 containers, each containing the same number of insects

Response variable: Number of insects that are still alive in each container one week after spraying

Part (b):

Yes. Because the 0 ml/L concentration contains no fungus, the containers that are sprayed with the 0 ml/L concentration form the control group.

Part (c):

Label each container with a unique integer from 1 to 20. Then use a random number generator to choose 15 integers from 1 to 20 without replacement. Use the first five of these numbers to identify the five containers that will receive the 0 ml/L treatment. Use the second five of these numbers to identify the five containers that will receive the 1.25 ml/L treatment. Use the third five of these numbers to identify the five containers that will receive the 2.5 ml/L treatment. The remaining five containers will receive the 3.75 ml/L treatment.

(Alternative solution) Using 20 equally sized slips of paper, label five slips with 0 ml/L, five slips with 1.25 ml/L, five slips with 2.5 ml/L, and five slips with 3.75 ml/L. Mix the slips of paper in a hat. For each container, select a slip of paper from the hat (without replacement) and spray that container with the treatment selected.

2019 SCORING GUIDELINES

Question 2 (continued)

Scoring

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

Part (a) is scored as follows:

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

1. Identifies the 4 concentrations (or mixtures or sprays) as the treatments
2. Identifies the 20 containers as the experimental units
3. Identifies the number of insects that are still alive in each container as the response variable

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

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

Notes:

- Listing the four treatments satisfies component 1 (including ml/L is not required). However, if the list does not include all four treatments, component 1 is not satisfied.
- To satisfy component 1, the response must refer to plural concentrations/mixtures/sprays (e.g., the mixtures, the levels of the concentration). Referring only to the explanatory variable (concentration) does not satisfy component 1.
- The following responses satisfy component 2: “the 20 containers”; “the containers”; “the 20 groups of insects”; or “the groups of insects in each container.” References to only “groups of insects” do not satisfy component 2 because it is unclear if these groups are formed by treatment or by container.
- To satisfy component 3, it must be clear that the response variable is being measured separately for each experimental unit. A response that says only “number of insects alive” does not satisfy component 3 because it could be referring to the total number of insects alive.
- To satisfy component 3, the response must be stated as a variable by using “number of” or equivalent. For example, “insects alive in each container” is not a variable and would not satisfy component 3.
- If the response states that the insects are the experimental units, then component 3 can still be satisfied by providing a binary response variable for each insect (e.g., whether the insect lived or died, survival status).

2019 SCORING GUIDELINES

Question 2 (continued)

Part (b) is scored as follows:

Essentially correct (E) if the response indicates that there is a control group and justifies this claim by identifying the control group or by explaining that there is a treatment which contains no fungus.

Partially correct (P) if the response indicates that there is no control group because every container is sprayed with some mixture

OR

if the response states that there is a control group but implies that 0 ml/L is not a treatment (e.g., “the containers with 0 ml/L form a control group because they don’t receive a treatment”; “yes, there is a group that got no treatment”).

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

Notes:

- The response does not need to explain the purpose of a control group.
- The response does not need to explicitly say “yes”—it can be implied by stating that there is a control group or saying “the control group is”

Part (c) is scored as follows:

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

1. Creates appropriate labels for the units/treatments (e.g., label the containers from 1 through 20, label 20 slips of paper with five for each treatment)
2. Describes how to correctly implement the random assignment process
3. The random assignment process results in an equal number of experimental units assigned to each treatment

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

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

Notes:

- If the response states that insects are the experimental units in part (a), the response in part (c) can be in terms of insects or containers. In either case, the same three components are used to determine the score.
- If the response states that the containers are the experimental units in part (a), but only describes how to assign insects to treatments in part (c), component 1 is not satisfied.
- For responses that use slips of paper:
 - If the number of slips of paper is not equal to the number of experimental units, then component 1 is not satisfied. The slips of paper do not need to be specifically identified as equally-sized.
 - If the slips of paper are not mixed/shuffled or the slips are not “selected at random,” component 2 is not satisfied. Sampling without replacement is implied when using slips of paper, unless the response specifies sampling with replacement.

2019 SCORING GUIDELINES

Question 2 (continued)

- For responses that use random number generators (or a 20-sided die):
 - If the initial assignment of numbers to units does not give each unit the same probability of being assigned to each treatment (e.g., units are represented by different numbers of integers), then component 1 is not satisfied.
 - If the response does not indicate that the numbers are selected without replacement or that different numbers must be used, the response does not satisfy component 2. The response does not need to specify the interval of numbers from which they are selecting (e.g., randomly generate a number from 1 to 20).
- For responses that use a table of random digits:
 - If the initial assignment of numbers to units does not give each unit the same probability of being assigned to each treatment, component 1 is not satisfied. For example, responses that use the labels 1 to 20 (not 01 to 20) do not satisfy component 1 because label 1 has a $\frac{1}{10}$ probability of being selected but label 20 has a $\frac{1}{100}$ probability of being selected.
 - If the response does not indicate that the numbers are selected without replacement or that different numbers must be used, the response does not satisfy component 2. The response does not need to specify the interval of numbers from which they are selecting or state that the numbers corresponding to unused labels will be skipped (e.g., skip numbers 00 and 21 to 99).
- For responses that use a 4-sided die (or random integers from 1 to 4):
 - If the die is rolled for each experimental unit, then component 3 is not satisfied because an equal number of units per treatment is not guaranteed.
 - If the die is rolled for each experimental unit until treatments are “full,” then component 1 is not satisfied because this setup doesn’t allow for all possible random assignments to be equally likely (unless the order of the units is randomized initially).
- If a response groups the experimental units before any random assignment (e.g., forms five groups of four containers or four groups of five containers), and then randomly assigns treatments to the groups or randomly assigns treatments within each group, component 1 is not satisfied. However, if a response forms groups in the context of a randomized block design with a reasonable blocking variable, component 1 can be satisfied.
- If a response describes two different random assignment processes in detail (e.g., how to randomly assign insects to containers and how to assign containers to treatments), both descriptions are scored according to the three components and the lower score is used.
- Responses that assign experimental units only to groups and not to treatments (e.g., randomly select five containers and put them in group 1) do not satisfy component 3.
- If the response randomly assigns insects to containers, the containers must be assigned to a treatment to satisfy component 3. In this case, the assignment of treatment to container does not need to be at random to satisfy component 3.

2019 SCORING GUIDELINES

Question 2 (continued)

4 Complete Response

Three parts essentially correct

3 Substantial Response

Two parts essentially correct and one part partially correct

2 Developing Response

Two parts essentially correct and no parts partially correct

OR

One part essentially correct and one or two parts partially correct

OR

Three parts partially correct

1 Minimal Response

One part essentially correct

OR

No parts essentially correct and two parts partially correct

2019 SCORING GUIDELINES

Question 3

Intent of Question

The primary goals of this question were to assess a student's ability to (1) use information from a two-way table of relative frequencies to compute joint, marginal, and conditional probabilities; (2) recognize whether two events are independent; and (3) compute a probability for a binomial distribution.

Solution

Part (a):

$$(i) \quad P(\text{never and woman}) = 0.0636$$

(ii)

$$\begin{aligned} P(\text{never or woman}) &= P(\text{never}) + P(\text{woman}) - P(\text{never and woman}) \\ &= 0.12 + 0.53 - 0.0636 \\ &= 0.5864 \end{aligned}$$

$$(iii) \quad P(\text{never} \mid \text{woman}) = \frac{P(\text{never and woman})}{P(\text{woman})} = \frac{0.0636}{0.53} = 0.12$$

Part (b):

Yes, the event of being a person who responds never is independent of the event of being a woman because

$$P(\text{never} \mid \text{woman}) = P(\text{never}) = 0.12.$$

Part (c):

Define X as the number of people in a random sample of five people who always take their medicine as prescribed. Then X has a binomial distribution with $n = 5$ and $p = 0.54$, and

$$P(X \geq 4) = \binom{5}{4}(0.54)^4(0.46)^1 + \binom{5}{5}(0.54)^5(0.46)^0 \approx 0.19557 + 0.04592 \approx 0.24149.$$

Scoring

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

Part (a) is scored as follows:

Essentially correct (E) if the response reports correct values of the probabilities for (i), (ii), and (iii).

Partially correct (P) if only one or two of the probabilities are correct.

Incorrect (I) if none of the probabilities are correct.

Notes:

- Assuming independence for events never and woman in (i) without referencing the result in part (b) does not satisfy (i).
- Alternative solutions for (ii) include $0.0564 + 0.0636 + 0.1384 + 0.3280 = 0.5864$ and $0.0564 + 0.53 = 0.5864$.

2019 SCORING GUIDELINES

Question 3 (continued)

Part (b) is scored as follows:

Essentially correct (E) if the response indicates that the events are independent, gives an explanation of independence using the events in the problem, *AND* provides appropriate justification using numbers from the table.

Note: Examples of valid explanations with appropriate justifications include:

- $P(\text{woman and never}) = 0.0636$ is the same as $P(\text{woman}) \times P(\text{never}) = (0.53)(0.12) = 0.0636$.
- $P(\text{never} | \text{woman}) = \frac{0.0636}{0.53} = 0.12$ is the same as $P(\text{never}) = 0.12$.
- $P(\text{woman} | \text{never}) = \frac{0.0636}{0.12} = 0.53$ is the same as $P(\text{woman}) = 0.53$.

Partially correct (P) if the response indicates that the events are independent *AND* gives an explanation of independence using the events in the problem but does not provide justification using numbers from the table

OR

if the response uses a correct method of illustrating that events are independent but makes an arithmetic mistake or a transcription mistake that results in concluding that these two events are not independent.

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

Part (c) is scored as follows:

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

1. Clearly indicates a binomial distribution with $n = 5$ and $p = 0.54$.
2. Indicates the correct boundary value and direction of the event.
3. Reports the correct probability.

Partially correct (P) if the response satisfies component 1 but it does not satisfy one or both of the other two components

OR

if the response does not satisfy component 1 but both of the other two components are satisfied.

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

2019 SCORING GUIDELINES

Question 3 (continued)

Notes:

- The response $B(5, 0.54)$ satisfies component 1.
- Components 1 and 2 are satisfied by displaying the correct formula for computing the binomial probability using the correct values for n and p , e.g.,

$$\binom{5}{4}(0.54)^4(0.46)^1 + \binom{5}{5}(0.54)^5(0.46)^0$$

- Only component 1 is satisfied if the correct binomial distribution is used in an incorrect probability formula, e.g.,

$$\binom{5}{4}(0.54)^4(0.46)^1$$

- For component 2, the boundary value and direction may be described in words, e.g., $P(\text{at least four people})$.
- Component 2 may be satisfied by displaying a bar graph of a binomial distribution with the appropriate bars shaded.
- The response of $1 - \text{binomcdf}(n = 5, p = 0.54, \text{upper bound} = 3) = 0.24$ is scored E since n , p , and the boundary value are clearly identified.

The response of $1 - \text{binomcdf}(n = 5, p = 0.54, 3) = 0.24$ is scored P since n and p are clearly identified and the boundary value is not identified.

The response of $1 - \text{binomcdf}(5, 0.54, 3) = 0.24$ is scored I.

- A normal approximation to the binomial is not appropriate since $np = 5 \times 0.54 = 2.7$ and $2.7 < 5$. A response using the normal approximation can score at most P. To score P, the response must include all of the following:

- an indication that the probability calculated is a normal approximation for the binomial probability
- a correct mean and standard deviation based on the binomial parameters
- clear indication of boundary and direction with a z-score or diagram
- the probability computed correctly.

An example of a response which meets these four criteria is

$$P\left(Z \geq \frac{4 - np}{\sqrt{np(1-p)}}\right) = P\left(Z \geq \frac{4 - (5)(0.54)}{\sqrt{(5)(0.54)(0.46)}}\right) \approx 0.1217, \text{ and the binomial distribution is mentioned.}$$

2019 SCORING GUIDELINES

Question 3 (continued)

4 Complete Response

Three parts essentially correct

3 Substantial Response

Two parts essentially correct and one part partially correct

2 Developing Response

Two parts essentially correct and no parts partially correct

OR

One part essentially correct and one or two parts partially correct

OR

Three parts partially correct

1 Minimal Response

One part essentially correct

OR

No parts essentially correct and two parts partially correct

2019 SCORING GUIDELINES

Question 4

Intent of Question:

The primary goals of this question were to assess a student's ability to perform an appropriate hypothesis test to address a particular question. More specific goals were to assess students' ability to state appropriate hypotheses, identify the appropriate statistical test procedure, check appropriate assumptions/conditions for inference; calculate a correct test statistic and p -value; and draw a correct conclusion, with justification, in the context of the study.

Solution

Section 1:

Let p_{14} represent the proportion of the population of kochia plants in the western United States that were resistant to glyphosate in 2014. Let p_{17} represent the proportion of the population of kochia plants in the western United States that were resistant to glyphosate in 2017.

The null hypothesis $H_0 : p_{17} - p_{14} = 0$ is to be tested against the alternative hypothesis $H_a : p_{17} - p_{14} > 0$.

An appropriate inference procedure is a two-sample z -test for a difference in proportions. The formula for the test statistic is:

$$z = \frac{\hat{p}_{17} - \hat{p}_{14}}{\sqrt{\left(\frac{\hat{p}_c(1 - \hat{p}_c)}{n_{17}} + \frac{\hat{p}_c(1 - \hat{p}_c)}{n_{14}} \right)}}$$

where $\hat{p}_c = \frac{n_{14}\hat{p}_{14} + n_{17}\hat{p}_{17}}{n_{14} + n_{17}}$ is a pooled estimate of the proportion of resistant plants for 2014 and 2017 combined.

Section 2:

The first condition for applying the test is that the data are gathered from independent random samples from the populations of kochia plants in the western United States in 2014 and 2017. The question indicates that a random sample of 61 kochia plants was taken in 2014 and a second random sample of 52 kochia plants was taken in 2017. It is reasonable to assume that the 2017 sample of plants was in no way influenced by the 2014 sample of plants.

The second condition is that the sampling distribution of the test statistic is approximately normal. This condition is satisfied because the expected counts under the null hypothesis are all greater than 10. The pooled estimate of the proportion of resistant plants is $\hat{p}_c = \frac{(61)(0.197) + (52)(0.385)}{61 + 52} \approx 0.2835$. The estimates of the expected counts are $61(0.2835) \approx 17.29$, $61(1 - 0.2835) \approx 43.71$, $52(0.2835) \approx 14.74$, $52(1 - 0.2835) \approx 37.26$, all of which are greater than 10.

Because sampling must have been done without replacement, the independence condition for each sample should be checked. Information on the population sizes of kochia plants is not given for either 2014 or 2017, but it is reasonable to assume that each population has millions of plants. Therefore it is reasonable to assume that the sample sizes are less than 10 percent of the respective population sizes.

2019 SCORING GUIDELINES

Question 4 (continued)

Using the pooled estimate of the proportion of resistant plants, $\hat{p}_c \approx 0.2835$, the value of the test statistic is:

$$z = \frac{0.385 - 0.197}{\sqrt{\frac{(0.2835)(0.7165)}{61} + \frac{(0.2835)(0.7165)}{52}}} \approx 2.21$$

The p -value is 0.0135.

Section 3:

Because the p -value is less than $\alpha = 0.05$, there is convincing statistical evidence to conclude that the proportion of resistant plants in the 2017 population of kochia plants is greater than the proportion of resistant plants in the 2014 population of kochia plants.

Scoring

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

Section 1 is scored as follows:

Essentially correct (E) if the response satisfies components 1 and 4 *AND* at least one of the remaining components:

1. Hypotheses imply equality of proportions in the null hypothesis and correct direction in the alternative hypothesis, which utilize an appropriate population parameter in words or symbols.
2. Identifies parameters that are population proportions.
3. Both parameters are correctly defined as proportions of resistant plants in 2014 and 2017.
4. The two-sample z -test for proportions is identified by name or formula

Partially correct (P) if the response does not meet the requirement for E, but at least two of the components are satisfied.

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

Notes:

- Correct ways to state the null hypothesis that satisfy component 1:

$$H_0 : p_{17} = p_{14} \text{ or } H_0 : p_{17} - p_{14} = 0$$

$$H_0 : p_{17} \leq p_{14} \text{ or } H_0 : p_{17} - p_{14} \leq 0$$

$$H_0 : p_{14} \geq p_{17} \text{ or } H_0 : p_{14} - p_{17} \geq 0$$

Correct ways to state the alternative hypothesis that satisfy component 1:

$$H_a : p_{17} > p_{14} \text{ or } H_a : p_{17} - p_{14} > 0$$

$$H_a : p_{14} < p_{17} \text{ or } H_a : p_{14} - p_{17} < 0$$

Incorrect ways to state the null hypothesis that do not satisfy component 1:

$$H_0 : p_{17} < p_{14} \text{ or } H_0 : p_{17} - p_{14} < 0$$

$$H_0 : p_{14} > p_{17} \text{ or } H_0 : p_{14} - p_{17} > 0$$

Incorrect ways to state the alternative hypothesis that do not satisfy component 1:

$$H_a : p_{17} \neq p_{14} \text{ or } H_a : p_{17} - p_{14} \neq 0$$

$$H_a : p_{17} < p_{14} \text{ or } H_a : p_{17} - p_{14} < 0$$

2019 SCORING GUIDELINES

Question 4 (continued)

- Examples for components 2 and 3:
 - Satisfies both components 2 and 3:
 - p_{17} is the proportion of resistant plants
 - p_{14} is the proportion of resistant plants
 - Satisfies component 2 but not component 3:
 - p_1 is the proportion of resistant plants
 - p_2 is the proportion of resistant plants
 - p_{17} is the proportion of plants
 - p_{14} is the proportion of plants
 - p_{17}, p_{14}
 - p_1, p_2
- If the test is correctly identified by name, but then an incorrect formula is stated, this is considered to be a parallel response and component 4 is not satisfied.
- If the test identifies an unpooled two sample z -test for a difference in proportions as the correct test or formula, component 4 is satisfied.

Section 2 is scored as follows:

Essentially correct (E) if the response satisfies components 1 and 2 *AND* at least two of the remaining components:

1. Notes that the use of random samples in 2014 and 2017 satisfies the randomness condition.
2. Checks for approximate normality of the test statistic by showing that the expected numbers of resistant and non-resistant kochia plants are both larger than some commonly accepted criterion (e.g. 5 or 10) for both samples.
3. Notes that the populations of kochia plants must be extremely large in both years, thus satisfies the independence (10%) conditions.
4. Reports a correct value of the z -test statistic.
5. Reports a p -value that is consistent with the stated alternative hypothesis and reported test statistic.

Partially correct (P) if the response does not meet the criteria for E, but at least two of the five components are satisfied.

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

Notes:

- For the randomness component it is minimally acceptable to say “random samples—check” or “SRSs—check.” The important concept is that the study used two independent random samples. Although it is not known if a SRS was taken versus another type of random sample, it is minimally acceptable to indicate SRSs since the sampling method is unknown. If the response implies that random assignment was used, the randomness component is not satisfied.

2019 SCORING GUIDELINES

Question 4 (continued)

- To satisfy component 2, the response must include actual numbers, or a formula with numbers plugged in, as well as a clear indication of comparison of the four quantities to some standard criterion, such as 5 or 10, or the statement that each such quantity is large enough. If a formula with numbers is used, simplification is NOT required.

Examples of acceptable quantities (comparison still must be made):

- 12, 49, 20, 32
- 12.017, 48.983, 20.02, 31.98
- $61(0.197)$, $61(1 - 0.197)$, $52(0.385)$, $52(1 - 0.385)$

Examples of unacceptable quantities:

- $n_{17}\hat{p}_{17}$, $n_{17}(1 - \hat{p}_{17})$, $n_{14}\hat{p}_{14}$, $n_{14}(1 - \hat{p}_{14})$
- $n_{17}p_{17}$, $n_{17}(1 - p_{17})$, $n_{14}p_{14}$, $n_{14}(1 - p_{14})$
- $n_{17}\hat{p}_c$, $n_{17}(1 - \hat{p}_c)$, $n_{14}\hat{p}_c$, $n_{14}(1 - \hat{p}_c)$
- $61\hat{p}_c$, $61(1 - \hat{p}_c)$, $52\hat{p}_c$, $52(1 - \hat{p}_c)$
- The test statistics for the pooled and unpooled z -tests are 2.21 and 2.22 respectively, thus they are close to the same value. If the response provides the unpooled formula but then states a pooled test statistic, component 4 is satisfied. If the response provides the pooled formula but then states an unpooled test statistic, component 4 is satisfied.
- If the response uses a critical value approach rather than a p -value approach, then the correct critical value of -1.645 or 1.645 , that is consistent with the alternative hypothesis, satisfies component 5.
- If the response did not satisfy component 1 in section 1 because a two-tailed alternative was stated or the direction of the alternative was incorrect, then the p -value in component 5 should be consistent with the stated alternative. If the response omits hypotheses or other incorrect hypotheses are stated, assume the correct alternative hypothesis is provided when scoring component 5.

Section 3 is scored as follows:

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

- Provides justification of the conclusion based on a correct comparison between a stated p -value and an alpha value of 0.05.
- Provides a correct conclusion consistent with the alternative hypothesis.
- The conclusion is stated in context.

Partially correct (P)

if the response satisfies components 1 and 2

OR

if the response satisfies components 2 and 3

OR

if the response satisfies components 1 and 3 AND, based on the p -value from section 2, either

- the conclusion correctly rejects the null hypothesis but does not state that there is convincing evidence for the alternative hypothesis

OR

- the conclusion correctly fails to reject the null hypothesis but does not state there is not convincing evidence for the alternative hypothesis.

2019 SCORING GUIDELINES

Question 4 (continued)

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

Notes:

- If the conclusion is consistent with a reasonable, but incorrect, p -value from section 2, and is presented in context with justification based on comparison of the p -value to the level of significance, then section 3 is scored E.
- If the response implies that the outcome of the hypothesis test is a “proof” of either a true or false null, the score is lowered one level (that is, from E to P, or from P to I).
- If an incorrect interpretation of the p -value is given, the score is lowered one level (that is, from E to P, or from P to I).
- If the response uses a critical value approach rather than a p -value approach, then the correct critical value of -1.645 or 1.645 replaces the p -value in section 2, and comparison of the test statistic from section 2 to the critical value (e.g. $2.21 > 1.645$) satisfies component 1.
- If the response clearly states a reasonable level of significance that differs from 0.05 and provides a justification and conclusion in context based on that justification, the response is scored E.
- If the response provides the incorrect comparison between the stated p -value and the level of significance, but the conclusion is consistent with the given comparison and the alternative hypothesis, then component 2 is satisfied.
- If the response did not satisfy component 1 in section 1 because a two-tailed alternative was stated or the direction of the alternative was incorrect, then the conclusion component 2 should be consistent with the stated alternative. If the response states other incorrect hypotheses or omits hypotheses, assume the correct alternative hypothesis is provided when scoring component 2.

2019 SCORING GUIDELINES

Question 4 (continued)

Alternative Approach

Two-Sided Confidence Interval for Difference in Two Population Proportions

Section 1 is scored E, P, or I according to the guidelines in section 1 for a 2 sample z -test for proportions.

Notes:

- To satisfy component 4, the two-sample confidence interval for a difference in two proportions should be identified by name or formula by referring to the z -distribution and two proportions.

Section 2 is scored as follows:

Essentially correct (E) if the response satisfies components 1 and 2 *AND* at least one of the remaining components:

- Notes that the use of random samples in 2014 and 2017 satisfies the randomness condition.
- Checks for approximate normality of the test statistic by showing that the observed numbers of resistant and non-resistant kochia plants are both larger than some commonly accepted criterion (say 5 or 10) for both samples.
- Notes that the populations of kochia plants must be extremely large in both years, thus satisfies the independence (10%) conditions.
- Reports the correct confidence interval that is consistent with the stated alternative hypothesis.

Partially correct (P) if the response does not meet the criteria for E, but at least two of the four components are satisfied.

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

Notes:

- Examples of correct 90% confidence intervals to address a one-sided alternative for $\alpha = 0.05$ are:
(-0.327, -0.049)
(0.049, 0.327)
- Examples of correct 95% confidence intervals to address a two-sided alternative for $\alpha = 0.05$ are:
(-0.353, -0.022)
(0.022, 0.353)

Section 3 is scored E, P or I according to the guidelines in section 3 for a 2 sample z -test for proportions.

Notes:

- Component 1 is satisfied if a confidence interval that is consistent with the alternative hypothesis is given and the appropriate interval endpoint(s) are compared to zero.

Overall Notes:

- If the response constructs two separate one-proportion z -intervals for 2014 and 2017, then sections 1 and 2 are scored as above, and section 3 is scored I.

2019 SCORING GUIDELINES

Question 4 (continued)

Alternative Approach

Chi-square test for homogeneity

The value of the Pearson chi-square test statistic (uncorrected) is 4.8821 with 1 degree of freedom and a p -value of 0.02714. This value is the same as the square of the pooled z -statistic, so it is the same test, but the p -value is for a two-sided alternative. This p -value could be divided by 2 to obtain an appropriate p -value for the one-sided alternative, but the sample data needs to be examined to determine the correct direction of the alternative.

Section 1 is scored E, P, or I according to the guidelines in section 1 for a 2 sample z -test for proportions.

Notes:

- Examples of unacceptable hypotheses:
 $H_0 : p_{17}$ and p_{14} are independent or $H_0 : p_{17}$ and p_{14} have no association
 $H_a : p_{17}$ and p_{14} are dependent or $H_a : p_{17}$ and p_{14} have an association
 $H_a : p_{17} \neq p_{14}$ or $H_a : p_{17} \geq p_{14}$ or $H_a : p_{14} \leq p_{17}$
- To satisfy component 4, a chi-square test for homogeneity is identified by name or formula.

Section 2 is scored E, P, or I according to the guidelines in section 2 for a 2 sample z -test for proportions.

Notes:

- To satisfy component 4, the reported value of the chi-square test statistic is 4.8821.
- If a one sided alternative hypothesis is given in section 1, then to satisfy component 5, the reported value of the p -value is 0.01357.
- If a two sided alternative hypothesis is given in section 1, then to satisfy component 5, the reported p -value is 0.02714.

Section 3 is scored E, P or I according to the guidelines in section 3 for a 2 sample z -test for proportions.

Notes:

- If the response clearly indicates that the two sample proportions were used to determine the correct one-sided direction and a p -value of 0.0135 was used as justification, section 3 is scored E.
- If a correct conclusion is reached based on the p -value of 0.02714 and a two-sided alternative, then section 3 is scored at most P.
- If the final response justifies the conclusion based on the p -value of 0.0135 but does not explicitly indicate how the correct direction was determined, section 3 is scored at most P.

2019 SCORING GUIDELINES

Question 4 (continued)

4 Complete Response

Three sections essentially correct

3 Substantial Response

Two sections essentially correct and one section partially correct

2 Developing Response

Two sections essentially correct and no sections partially correct

OR

One section essentially correct and one or two sections partially correct

OR

Three sections partially correct

1 Minimal Response

One section essentially correct

OR

No section essentially correct and two sections partially correct

2019 SCORING GUIDELINES

Question 5

Intent of Question

The primary goals of this question were to assess a student's ability to (1) evaluate a percentile of a normal distribution; (2) evaluate a probability for a normal distribution; and (3) compute an expected value for a random variable with two possible outcomes.

Solution**Part (a):**

The 25th percentile of the standard normal distribution is -0.6745 . Consequently the 25th percentile of a normal distribution with mean 30 months and standard deviation 8 months is $30 + 8(-0.6745) = 24.6$ months.

Part (b):

The probability that a randomly selected customer will need to request a replacement because the battery fails within 24 months from the date of purchase is

$$P(\text{life span} \leq 24 \text{ months}) = P\left(Z \leq \frac{24 - 30}{8}\right) = P(Z \leq -0.75) \approx 0.2266.$$

Part (c):

The company's expected gain for each warranty purchased is

$$\begin{aligned} &(\$50) \times P(\text{life span} > 24 \text{ months}) + (-\$150) \times P(\text{life span} \leq 24 \text{ months}) \\ &= (\$50) \times (0.7734) + (-\$150) \times (0.2266) \approx \$4.68. \end{aligned}$$

Scoring

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

Part (a) is scored as follows:

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

1. Indicates the use of a normal distribution with a mean of 30 and a standard deviation of 8.
2. Sets up a correct approach for finding the 25th percentile of the battery life span distribution.
3. Reports the correct value of the 25th percentile (24.6) or a percentile value that is 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 meet the criteria for E or P.

Notes:

- Component 1 may be satisfied in either part (a) or (b).
- Incorrect statistical notation in specifying the distribution mean or standard deviation (for example, $\bar{x} = 30$ or $s = 8$) results in Component 1 not being met the first time it appears.

2019 SCORING GUIDELINES

Question 5 (continued)

- Component 1 may be satisfied by one of the following:
 - **Graphical**: Displaying a graph of a normal density function with the horizontal axis clearly labeled using the mean and standard deviation for the battery life span distribution.
 - **Calculator syntax**: Labeling the mean and standard deviation in an inverse normal cdf calculator statement, for example, $\text{invNorm}(0.25, \mu = 30, \sigma = 8)$.
 - **z-score**: Showing correct components in a standard z-score calculation, for example, $z = \frac{x-30}{8}$, or $-0.6745 = \frac{x-30}{8}$, or $x = 30 + 8(-0.6745)$.
 - **Notation**: Using standard notation for a normal distribution, for example, $N(30, 8)$ or $\text{Normal}(30, 64)$.
- Component 2 may be satisfied by one of the following:
 - **Graphical**: Identifying the lower-tail area corresponding to the probability of 0.25 in a graph of a normal density function.
 - **Calculator syntax**: Stating the correct percentile in an inverse normal cdf calculator statement, for example, $\text{invNorm}(0.25, \mu = 30, \sigma = 8)$.
 - **z-score**: Equating the z-score of the 25th percentile of the battery life span distribution to -0.6745 , for example, $-0.6745 = \frac{x-30}{8}$ or $x = 30 + 8(-0.6745)$.

Part (b) is scored as follows:

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

1. Indicates the use of a normal distribution with a mean of 30 and a standard deviation of 8.
2. Specifies the correct event, including the correct boundary value and direction.
3. Reports the correct probability (0.2266) or a probability consistent with components 1 and 2.

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

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

Notes:

- Component 1 may be satisfied in either part (a) or (b).
 - Incorrect statistical notation in specifying the distribution mean or standard deviation (for example, $\bar{x} = 30$ or $s = 8$) results in Component 1 not being met the first time it appears.
- Component 1 may be satisfied by one of the following:
 - **Graphical**: Displaying a graph of a normal density function with the horizontal axis clearly labeled using the mean and standard deviation for the battery life span distribution.
 - **Calculator syntax**: Labeling the mean and standard deviation in a normal cdf calculator statement, for example, $\text{normalcdf}(0, \text{upper} = 24, \mu = 30, \sigma = 8)$.
 - **z-score**: Showing correct components in a standard z-score calculation, for example, $z = \frac{x-30}{8}$, or $z = \frac{24-30}{8}$, or $\frac{24-30}{8} = -0.75$.
 - **Notation**: Using standard notation for a normal distribution, for example, $N(30, 8)$ or $\text{Normal}(30, 64)$.

2019 SCORING GUIDELINES

Question 5 (continued)

- Component 2 may be satisfied by one of the following:
 - **Graphical**: A normal density graph labeled with boundary and direction corresponding to the region of interest.
 - **Calculator syntax**: Labeling the upper bound in a normal cdf calculator statement. The lower bound does not need to be labeled but must be less than or equal to 0, for example, $\text{normalcdf}(0, \text{upper} = 24, \mu = 30, \sigma = 8)$.
 - **In words**: Specifying the correct event, boundary value, and direction, in words: for example, $P(\text{battery life span} \leq 24 \text{ months})$, or “probability requires battery replacement within 24 months.”
 - **Random variable**: Specifying the boundary value and direction using a random variable, for example, $P(X \leq 24)$ or $P(Z \leq -0.75)$.

Part (c) is scored as follows:

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

1. Provides an expected value calculation with two appropriate probabilities that add to 1 and are paired with the correct outcomes.
2. Reports the correct expected value (4.68), or a reasonable expected value consistent with work shown.

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

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

Notes:

- For component 1, appropriate probabilities that add to 1 are:
 - probabilities consistent with the value computed in part (b); or
 - the correct probabilities; or
 - if there is no probability provided in part (b), probabilities explicitly defined in context in part (c).
- If the correct probability is used, due to rounding considerations, any number between 4.00 and 4.70 is acceptable as a correct expected value for component 2.
- A reasonable expected value for component 2 can be any number between $-\$150$ and $+\$50$ that is consistent with work shown.

2019 SCORING GUIDELINES

Question 5 (continued)

4 Complete Response

Three parts essentially correct

3 Substantial Response

Two parts essentially correct and one part partially correct

2 Developing Response

Two parts essentially correct and no parts partially correct

OR

One part essentially correct and one or two parts partially correct

OR

Three parts partially correct

1 Minimal Response

One part essentially correct

OR

No parts essentially correct and two parts partially correct

2019 SCORING GUIDELINES

Question 6

Intent of Question

The primary goals of this question were to assess a student's ability to (1) recognize the population to which results from a random sample may be generalized; (2) describe a disadvantage of using a sample mean rather than a sample median to indicate typical values when the sample distribution is skewed; (3) describe how the theoretical sampling distribution of the sample median could be constructed; (4) construct an approximate confidence interval for a population median using results from a bootstrap procedure; and (5) interpret a confidence interval.

Solution**Part (a):**

Because random sampling was used, the results of the sample may be generalized to the population of rental prices for one-bedroom apartments in the city that are listed on this particular website at the time the sample was taken.

Part (b):

Because the distribution of the 50 rental prices in the sample is skewed to the right, the sample median provides a better indicator of typical rental prices than the sample mean. Some very large rental prices result in a sample mean that is substantially larger than the more typical rental prices. As a result the sample mean would overestimate the typical rental price, whereas the sample median would be a more accurate representation of typical rental prices.

Part (c):

To determine the sampling distribution of median rental prices for random samples of 50 one-bedroom apartments from this population, Emma would need to obtain every possible sample of 50 one-bedroom apartments from this website and compute the median of each sample. The collection of all possible sample medians is the theoretical sampling distribution for sample median.

Part (d):

- (i) $(0.05)(15,000) = 750$ and $(0.95)(15,000) = 14,250$. The 5th percentile is a value, say $x_{0.05}$, such that at least 750 values in the table are less than or equal to $x_{0.05}$ and at least 14,250 are greater than or equal to $x_{0.05}$. Cumulate frequencies starting with the smallest sample median listed in the table and going toward the largest (going down columns) until you first reach 750 values, to obtain $x_{0.05} = \$2,500$.
- (ii) Similarly, $x_{0.95} = \$2,950$.

Part (e):

The percentage of bootstrap medians between (and including) the values found in part (d) for the 5th and 95th percentiles is

$$\frac{14,404}{15,000} \times 100\% \approx 96.03\%$$

2019 SCORING GUIDELINES

Question 6 (continued)

Part (f):

From the results in part (d) and part (e), an approximate 96 percent confidence interval for the median rental price of all one-bedroom apartments listed on this website for this city is (\$2,500, \$2,950). We are approximately 96 percent confident that the median rental price of all one-bedroom apartments listed on this website for this city is between \$2,500 and \$2,950.

Scoring

This question is scored in four sections. Section 1 consists of parts (a) and (b), section 2 consists of part (c), section 3 consists of parts (d) and (e), and section 4 consists of part (f). Sections 1, 2, 3, and 4 are each scored as essentially correct (E), partially correct (P), or incorrect (I).

Section 1 is scored as follows:

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

1. The correct population (listings of one-bedroom apartments on the website) is identified in part (a).
2. In part (b), identifying that using the sample mean instead of the sample median overestimates the typical rental price.
3. The disadvantage of using the sample mean that is reported in part (b) is correctly linked to some feature of the distribution (e.g. skewness) that is evident in the histogram.

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

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

Note: Responses that refer to the mean being larger than the median in a skewed right distribution alone is not sufficient to satisfy component 2.

Section 2 is scored as follows:

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

1. Indicates that Emma would need to obtain every possible sample of 50 one-bedroom apartments.
2. Indicates that Emma would need to compute the median rental price for each sample.

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

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

Section 3 is scored as follows:

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

1. Correct values for the 5th percentile and the 95th percentile are reported in part (d).
2. The correct percentage of bootstrap samples that produced sample medians at or between the two values, if they are plausible, reported in part (d) is reported in part (e).

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

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

Note: Plausible values for part (d) will be considered values between 2,345 and 3,062.5.

2019 SCORING GUIDELINES

Question 6 (continued)

Section 4 is scored as follows:

Essentially correct if the response in part (f) satisfies the following three components:

1. Uses \$2500 and \$2950 or the values of the percentiles reported in part (d) as the endpoints of the confidence interval.
2. Indicates an approximate 90 or 96 percent level of confidence or a level of confidence consistent with part (e).
3. Makes a correct statement in context indicating that the confidence interval is for the median.

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

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

Note: Since rental prices from the population are discrete values, the true confidence level of the interval from part (d) is unknown. A correctly calculated part (e) is a way to estimate the confidence level; from Emma's sample the confidence level is estimated to be approximately 96 percent. The process described in part (d) for calculating the interval will result in a confidence level of at least 90 percent. For these reasons, confidence levels of either 90 or 96 percent satisfy component 2.

Each essentially correct (E) section counts as 1 point, and a 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 strength of the response and communication.

2018

AP[®] CollegeBoard

AP Statistics

Scoring Guidelines

2018 SCORING GUIDELINES

Question 1

Intent of Question

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

Solution

Part (a):

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

Part (b):

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

Part (c):

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

Scoring

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

Part (a) is scored as follows:

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

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

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

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

2018 SCORING GUIDELINES

Question 1 (continued)

Notes:

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

Part (b) is scored as follows:

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

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

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

OR

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

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

Notes:

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

2018 SCORING GUIDELINES

Question 1 (continued)

Part (c) is scored as follows:

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

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

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

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

Notes:

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

2018 SCORING GUIDELINES

Question 1 (continued)

4 Complete Response

Three parts essentially correct

3 Substantial Response

Two parts essentially correct and one part partially correct

2 Developing Response

Two parts essentially correct and no parts partially correct

OR

One part essentially correct and one or two parts partially correct

OR

Three parts partially correct

1 Minimal Response

One part essentially correct

OR

No parts essentially correct and two parts partially correct

2018 SCORING GUIDELINES

Question 2

Intent of Question

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

Solution

Part (a):

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

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

Part (b):

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

Part (c):

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

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

2018 SCORING GUIDELINES

Question 2 (continued)

Scoring

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

Part (a) is scored as follows:

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

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

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

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

Notes:

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

AP STATISTICS
2018 SCORING GUIDELINES

Question 2 (continued)

Part (b) is scored as follows:

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

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

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

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

Notes:

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

AP STATISTICS
2018 SCORING GUIDELINES

Question 2 (continued)

Part (c) is scored as follows:

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

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

OR

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

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

Notes:

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

2018 SCORING GUIDELINES

Question 2 (continued)

4 Complete Response

Three parts essentially correct

3 Substantial Response

Two parts essentially correct and one part partially correct

2 Developing Response

Two parts essentially correct and no parts partially correct

OR

One part essentially correct and one or two parts partially correct

OR

Three parts partially correct

1 Minimal Response

One part essentially correct

OR

No parts essentially correct and one or two parts partially correct

2018 SCORING GUIDELINES

Question 3

Intent of Question

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

Solution

Part (a):

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

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

Part (b):

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

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

Part (c):

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

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

2018 SCORING GUIDELINES

Question 3 (continued)

Scoring

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

Part (a) is scored as follows:

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

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

OR

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

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

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

Part (b) is scored as follows:

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

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

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

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

2018 SCORING GUIDELINES

Question 3 (continued)

Part (c) is scored as follows:

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

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

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

OR

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

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

Notes:

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

Notes for all parts:

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

2018 SCORING GUIDELINES

Question 3 (continued)

4 Complete Response

Three parts essentially correct

3 Substantial Response

Two parts essentially correct and one part partially correct

2 Developing Response

Two parts essentially correct and no parts partially correct

OR

One part essentially correct and one or two parts partially correct

OR

Three parts partially correct

1 Minimal Response

One part essentially correct

OR

No parts essentially correct and two parts partially correct

2018 SCORING GUIDELINES

Question 4

Intent of Question

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

Solution

Part (a):

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

Part (b):

Step 1: State a correct pair of hypotheses.

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

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

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

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

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

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

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

2018 SCORING GUIDELINES

Question 4 (continued)

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

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

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

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

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

Scoring

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

Section 1 is scored as follows:

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

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

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

OR

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

OR

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

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

2018 SCORING GUIDELINES

Question 4 (continued)

Notes:

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

Section 2 is scored as follows:

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

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

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

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

Notes:

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

Confidence Interval approach:

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

2018 SCORING GUIDELINES

Question 4 (continued)

Section 3 is scored as follows:

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

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

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

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

Notes:

Component 1:

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

Component 2:

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

Component 3:

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

2018 SCORING GUIDELINES

Question 4 (continued)

Confidence Interval approach:

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

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

2018 SCORING GUIDELINES

Question 4 (continued)

4 Complete Response

Three sections essentially correct with conditions for inference

3 Substantial Response

Three sections essentially correct without conditions for inference

OR

Two sections essentially correct and one section partially correct

2 Developing Response

Two sections essentially correct and no sections partially correct

OR

One section essentially correct and one or two sections partially correct

OR

Three sections partially correct

1 Minimal Response

One section essentially correct

OR

No sections essentially correct and one or two sections partially correct

2018 SCORING GUIDELINES

Question 5

Intent of Question

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

Solution**Part (a):**

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

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

Part (b):

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

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

Part (c):

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

2018 SCORING GUIDELINES

Question 5 (continued)

Scoring

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

Part (a) is scored as follows:

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

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

OR

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

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

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

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

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

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

Part (b) is scored as follows:

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

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

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

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

2018 SCORING GUIDELINES

Question 5 (continued)

Part (c) is scored as follows:

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

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

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

OR

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

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

Notes:

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

2018 SCORING GUIDELINES

Question 5 (continued)

4 Complete Response

Three parts essentially correct

3 Substantial Response

Two parts essentially correct and one part partially correct

OR

Part (a) essentially correct and two parts partially correct

2 Developing Response

Two parts essentially correct and no parts partially correct

OR

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

OR

Three parts partially correct

1 Minimal Response

One part essentially correct

OR

No parts essentially correct and one or two parts partially correct

2018 SCORING GUIDELINES

Question 6

Intent of Question

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

Solution**Part (a):**

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

Part (b):

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

$$\bar{x} > 124.4675.$$

Part (c):

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

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

Part (d):

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

Part (e):

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

2018 SCORING GUIDELINES

Question 6 (continued)

Scoring

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

Section 1 is scored as follows:

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

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

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

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

Notes:

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

Section 2 is scored as follows:

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

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

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

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

Notes:

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

2018 SCORING GUIDELINES

Question 6 (continued)

Section 3 is scored as follows:

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

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

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

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

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

2018 SCORING GUIDELINES

Question 6 (continued)

4 Complete Response

Three sections essentially correct

3 Substantial Response

Two sections essentially correct and one section partially correct

2 Developing Response

Two sections essentially correct and no sections partially correct

OR

One section essentially correct and one or two sections partially correct

OR

Three sections partially correct

1 Minimal Response

One section essentially correct

OR

No sections essentially correct and two sections partially correct

OR

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

OR

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

AP[®] Statistics 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,

$$H_0 : p_{\text{asp}} - p_{\text{plac}} = 0 \text{ versus } H_a : p_{\text{asp}} - p_{\text{plac}} < 0.$$

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.

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

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

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

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 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 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:

- Correct formula for predicting arm span with 61 inserted for x .
- Correct value for the predicted arm span.
- 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

OR 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

OR 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