2022

AP® Statistics

Scoring Guidelines

Question 1: Focus on Exploring Data — 4 points

General Scoring Notes

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

	Model Solution	Scoring
(a)	The scatterplot reveals a strong, positive, roughly linear association between the mass and length of bullfrogs. There are no points that seriously deviate from the straight-line pattern of the points in the plot.	**Essentially correct (E)** if the response provides a description that includes at least three of components 1-4 and component 5: 1. Direction of association (positive or increasing) 2. Strength of association (strong) 3. Form of association (linear or approximately linear) 4. Unusual features (no points with large discrepancies from the pattern (straight line) exhibited by most of the points on the plot) 5. Context (association between length and mass of bullfrogs) **Partially correct (P)** if the response satisfies only one or two components out of components 1-4 and component 5 *OR* if the response satisfies at least three out of components 1-4 but does not satisfy component 5. **Incorrect (I)** if the response does not meet the criteria for E or P.

Additional Notes:

- To satisfy component 4, it is sufficient to simply indicate that there are no unusual features.
- To satisfy component 5, it is minimally sufficient for the response to refer to the association or relationship between mass and length without explicitly mentioning bullfrogs.
- The strength of the response in part (a) may be considered if holistic scoring is needed.

	Model Solution	Scoring
(b)	The value of the slope of the least-squares regression line is 6.086. This value indicates that the predicted mass of a bullfrog increases by 6.086 grams for each additional millimeter of length.	**Essentially correct (E)** if the response satisfies the following three components: 1. Identifies the value of the slope as 6.086 2. Provides an interpretation that references an increase of a number of grams of mass for each one-millimeter increase in length 3. Indicates that the slope represents a change in a prediction using non-deterministic language such as "predicted," "estimated," "expected," or "average" **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.

Additional Notes:

- The value of the slope, 6.086, may be rounded to 6.09 or 6.1, but not to 6, to satisfy the numerical requirement in component 1.
- A response that only contains 6.086 in the interpretation satisfies component 1.
- A calculation of slope may satisfy component 1, provided that two points from the line are used in the calculation.
- Units of measurements must be correctly specified for both mass and length to satisfy component 2.
- It is not required to refer specifically to the "least-squares regression line."

	Model Solution	Scoring
(c)	The coefficient of determination is $r^2 \approx 0.819$. This value indicates that 81.9% of the variation in bullfrog mass can be explained by variation in bullfrog length as described by the least-squares line.	**Essentially correct (E)** if the response provides a correct interpretation of r^2 in context. **Partially correct (P)** if the response provides a generic interpretation (no context) *OR* if the response provides a reasonable but incorrect interpretation of r^2 in context. **Incorrect (I)** if the response does not satisfy the criteria for E or P.

Additional Notes:

- Correct interpretations of r^2 include the concept that part of the variation in the response (dependent or *y*) variable is explained by the linear relationship with the explanatory (independent or *x*) variable. The response can take any of several equivalent forms, such as:
  - The proportion of the total variability in the dependent (response) variable *y* that is explained by the independent (explanatory) variable *x*.
  - The proportion of variation in *y* that is accounted for by the linear model.
  - The proportionate reduction of the 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 sample mean to predict values of *y*.
- Correct interpretation of r^2 must explicitly relate to the dependent variable. Mention of the data, predicted values, or no mention of the dependent variable are incorrect interpretations. Common incorrect interpretations include:
  - The percent (or proportion or part of the total) variability in the predicted *y*-values that is explained by the linear relationship between *y* and *x*.
  - The percent (or proportion or part of the total) variability in the data that is explained by the linear relationship between *y* and *x*.
  - The percent (or proportion or part of the total) variability that is explained by the linear relationship between *y* and *x*.
  - The percent (or proportion or part of the total) variability in *y* that is on average explained by the linear relationship between *y* and *x*.
- A reasonable but incorrect interpretation of r^2 with context might include the following responses:
  - 81.9% of the variation in mass and length can be accounted for by the least-squares regression line.
  - 81.9% of the variability in predicted mass is accounted for by the length.
- For context, the response variable (*y*) must be identified as mass, and the explanatory variable (*x*) must be identified as length.
- An interpretation of the correlation between mass and length, $r = \sqrt{0.819} = 0.905$, is not considered a reasonable interpretation of r^2.
- The value of the percentage (81.9%) or proportion (0.819) of variation does not need to be specified, but if an incorrect value is specified, the score is lowered by one level, from E to P or from P to I.
- The strength of the response in part (c) may be considered if holistic scoring is needed.

Model Solution	Scoring
(d) (i) The largest residual in absolute value belongs to the bullfrog with length 162 millimeters and mass 356 grams. (ii) The least-squares regression line overestimates the mass of the bullfrog with length 162 millimeters. Plot 2 shows that the point for the bullfrog with length 162 millimeters is below the least-squares regression line.	**Essentially correct (E)** if the response satisfies the following two components: 1. The response to part (d-i) identifies the correct bullfrog (length between 160 and 165 millimeters, mass between 350 and 375 grams) 2. The response to part (d-ii) explicitly indicates whether the linear model overestimates or underestimates mass for the bullfrog identified in part (d-i) and provides a correct justification based on a comparison of the identified observation to the least-squares regression line **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.

Additional Notes:

- The comparison of the observation to the regression line in the response to part (d-ii) is satisfied if the response does one of the following:
  - Correctly indicates if the observation is below (above) the least-squares regression line in Plot 2.
  - Notes that observed mass is smaller (larger) than the mass predicted by the least-squares regression line.
  - Marks the observation selected in part (d-i) on Plot 2 with an indication of the vertical distance from the least-squares regression line.
  - Notes the correct sign of the residual.
- Numerical values are not required in the response to part (d-ii). If a numerical value is given for the predicted mass, however, it must be reasonable. A numerical value for the predicted mass could be computed with the formula given in the stem, e.g., $-546 + (6.086)(162) = 439.9$ grams, for a bullfrog of length 162 millimeters, or a value can be read from the line shown in Plot 2. Any value between 425 and 450 should be considered a reasonable value. Showing work is not required.
- The word overestimate with the calculated predicted value of mass is enough to satisfy component 2.
- If the wrong observation is identified in part (d-i), the response to part (d) may be scored P if the response to part (d-ii) correctly compares that observation to the least-squares regression line and states the correct conclusion about overestimating or underestimating mass with justification.
- It is not required to refer specifically to the "least-squares regression line."

Scoring for Question 1
Each essentially correct (E) part counts as 1 point, and each partially correct (P) part counts as ½ point.

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

If a response is between two scores (for example, 2 ½ points), use a holistic approach to decide whether to score up or down, depending on the strength of the response and quality of the communication.

Question 2: Focus on Collecting Data — 4 points

General Scoring Notes

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

Model Solution	Scoring
(a) Treatments: New drug, placebo. Experimental units: The 72 people who receive the new drug or placebo. Response variable: Improvement in acne severity	**Essentially correct (E)** if the response satisfies the following three components: 1. Identifies the treatments as new drug and placebo 2. Identifies the experimental units as the 72 people (subjects, participants, twins) in the experiment 3. Identifies the response variable as the improvement in acne severity **Partially correct (P)** if the response satisfies only two of the three components. **Incorrect (I)** if the response does not satisfy the criteria for E or P.

Additional Notes:

- To satisfy component 1, identification of the treatments must include both the placebo and the new drug.
- To satisfy component 2, the response must indicate that the experimental units are individual people. The response could refer to participants, subjects, twins, or members of the pairs of twins without explicitly mentioning the number 72. However, a response that states or implies that there are 36 experimental units (e.g., "the pairs of twins") does not satisfy component 2.
- To satisfy component 3, the response must include the context of "acne" and "improvement" (e.g., "improvement in acne severity," "acne improvement score"), but it does not need to include a reference to the scale, the dermatologist, two-week time periods, or treatments. Reasonable synonyms for improvement can be used, such as using "reduction" or "change" or by including the verbal descriptions of the scale ("no improvement" to "complete cure"). However, a description of a binary outcome (e.g., "whether or not the acne improves") does not satisfy component 3.
- For responses that indicate the 36 pairs of twins are the experimental units, component 3 may be satisfied by indicating that the response variable is the improvement in acne severity or by indicating that the response variable is the difference in improvement in acne severity.
- If the response provides parallel solutions (i.e., two or more complete solutions without choosing or indicating which is to be scored), the response is scored based on the weaker of the two solutions. For example, if a response says that the experimental units are "the 72 participants and the scores from 0 to 100," component 2 is not satisfied.

	Model Solution	Scoring
(b)	Improvement scores will vary due to many factors, including initial acne severity, what treatment is received, and other variables such as diet and genetics. Because the pairs of twins are similar in initial acne severity, pairing allows for the variation in improvement scores due to the treatment received to be distinguished from variation due to initial acne severity, unlike in a completely randomized design. Consequently, using the matched-pairs design will provide a more precise estimate of the mean difference in improvement in acne severity for the new drug compared to the placebo and make it easier to find convincing evidence that the new drug is better, if it really is better.	**Essentially correct (E)** if the response describes a statistical advantage of a matched-pairs design AND satisfies the following three components: 1. The advantage pertains to an inference made after collecting the data (e.g., the ability to distinguish between the effects of the treatments or the precision of the estimate of the drug effect) 2. Indicates that the matched-pairs design is better by using a comparative word (e.g., easier, clearer, greater) or by making an explicit comparison to a completely randomized design 3. Includes context (e.g., "drug," "improvement," "acne," or "twins") **Partially correct (P)** if the response describes a statistical advantage of a matched-pairs design AND satisfies one or two of the three components. **Incorrect (I)** if the response does not satisfy the criteria for E or P.

Additional Notes:

- To be considered an advantage of a matched-pairs design, the advantage described must be true for a matched-pairs design and not be true for a completely randomized design. For example, saying that "random assignment allows us to conclude cause-and-effect" is true of both designs. Similarly, "this allows the dermatologist to make conclusions about people with differing acne severity" is true of both designs. Also, "reduces bias" and "reduces variability in the estimates of the individual treatment means" is true of neither design.
- Responses that describe only the set-up of a matched-pairs experiment do not satisfy the requirement to describe an advantage of a matched-pairs design. For example, the response "in a matched-pairs design, the members of each pair will be similar in terms of acne severity" does not describe an advantage. However, "in a matched-pairs design, we can compare two people with similar acne severity" does describe an advantage.
- Advantages of a matched-pairs design that satisfy component 1 include "makes it easier to determine if the drug is effective," "gives a better estimate of the effect of the new drug," "reduces variability in the estimate of the drug effect," "makes the difference between the drug and the placebo more easily distinguishable," and "gives a clearer picture of how well the drug works."
- Advantages of a matched-pairs design that don't satisfy component 1 include "accounts for a source of variability," "controls for potentially confounding variables," "allows you to distinguish variation due to severity from variation due to treatment," "each person can be compared to someone similar," "reduces variability," "more balanced treatment groups," and "more accurate results."

- It is acceptable to provide a disadvantage of a completely randomized design rather than an advantage of the matched-pairs design (e.g., "The completely randomized design will make it harder to find convincing evidence that the new drug is better").
- It is acceptable to use the term "blocking" as a synonym for "pairing."
- A response that states that a matched-pairs design requires a smaller sample size to get power or precision equal to that in a completely randomized design and describes this advantage in context should be scored E.

	Model Solution	Scoring
(c)	For each pair of twins, label one person as twin A and label the other person as twin B. For each pair of twins, toss a coin. If the coin lands on heads, twin A gets the placebo and twin B gets the active drug. If the coin lands on tails, twin A gets the active drug and twin B gets the placebo. *OR* Label the members of each pair of twins as "Twin 1" and "Twin 2." Using a random number generator, generate an integer from 1 to 2. Give the drug to the twin whose number is selected and the placebo to the twin whose number is not selected. Repeat for all pairs of twins. *OR* Label 1 notecard "A" and another notecard "B." For each pair of twins, shuffle the cards and give one card to each twin. The twin who gets "A" receives the drug and the twin who gets "B" receives the placebo.	**Essentially correct (E)** if the response randomly assigns the two treatments within pairs of twins AND satisfies the following three components: 1. Uses a random process (e.g., flipping a coin, using a random number generator, shuffling cards) that gives each twin in a pair a 50% probability of getting the drug and a 50% probability of getting the placebo 2. Describes how to use the random process to assign one specific twin in each pair to the drug and the other twin to the placebo 3. Indicates that the random assignment process will be completed for each pair of twins **Partially correct (P)** if the response randomly assigns the two treatments within pairs of twins AND satisfies only two of the three components for E. **Incorrect (I)** if the response does not satisfy the criteria for E or P.

Additional Notes:

- A response that does not randomly assign both treatments within pairs of twins should be scored incorrect (I). Examples include a response that describes a completely randomized design, describes a crossover design where each person receives both treatments, uses pairs other than twins, does not use random assignment, or indicates that both twins in a pair receive the same treatment.
- For responses that use slips of paper or selecting items from a hat, the slips must be shuffled (or blindly drawn) or the hat mixed or shaken to have a random process and satisfy component 1.
- To satisfy component 2, the response must describe what to do for each possible outcome of the random process and specify which treatment each twin receives. For example, none of the following descriptions satisfy component 2:
  - "Roll a die. If it is 1–3, give the first twin the drug and the second twin the placebo." (Response doesn't describe what to do if the die is 4–6.)
  - "Have one member of each pair flip a coin. If it is heads, that twin gets the drug. If it is tails, that twin gets the placebo." (Response doesn't indicate what treatment the other twin will receive.)
  - "Flip a coin. If it is heads, give one twin the drug and the other twin the placebo. If it is tails, do the reverse." (Response doesn't specify which twin is getting the drug.)
  - "Label one slip of paper "A" and a second slip of paper "B." Mix them in a hat and have each member of the pair choose one slip." (Response doesn't specify if A represents the new drug or the placebo.)

- Ignore any discussion about randomly selecting 36 pairs of twins to obtain subjects for the experiment. Likewise, ignore any discussion about how to perform the analysis for a paired design (e.g., "subtract the improvement scores for each pair of twins").
- It is acceptable to refer to each pair of twins as a block.

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

Question 3: Focus on Probability and Sampling Distributions — 4 points

General Scoring Notes

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

	Model Solution	Scoring
(a)	Random variable A, which represents the amount of shampoo in a randomly selected bottle, follows a normal distribution with mean 0.6 liter and standard deviation 0.04 liter. Then, the probability that a randomly selected bottle is underfilled is $P(A < 0.5) = P\left(Z < \frac{0.5 - 0.6}{0.04} = -2.5\right) \approx 0.0062.$	**Essentially correct (E)** if the response includes the following three components: 1. Indicates the use of a normal (or approximately normal) distribution and identifies the correct parameter values (mean 0.6 and standard deviation 0.04) 2. Specifies the correct event (boundary value and direction), or an event consistent with values reported in component 1 3. Provides the correct probability of 0.0062 or probability consistent with components 1 and 2 **Partially correct (P)** if the response satisfies only two of the three components *OR* if the response fails to satisfy component 1 and 2, but shows the correct z-score formula, z-score value, and correct probability (e.g., $\frac{0.5 - 0.6}{0.04} = -2.5,$ resulting in a probability of 0.0062). **Incorrect (I)** if the response does not satisfy the criteria for E or P.

Additional Notes:

Component 1

- A response may satisfy component 1 by any of the following or a combination of the following:
  - Graphical: Displaying a graph of a normal density function with the appropriate scale on the horizontal axis showing the mean and standard deviation for the distribution of shampoo amount.

- Calculator function syntax: Labeling correct values of the mean and standard deviation in a "normalcdf" statement, such as
  $\text{normalcdf}(\text{lower} = -\infty, \text{upper} = 0.5, \text{mean} = 0.6, \text{standard deviation} = 0.04)$.
  Correct specification of the upper and lower bounds is not required to satisfy component 1.
- Words: Using a statement such as "normal distribution with mean 0.6 and standard deviation 0.04."
- Standard Notation: Using standard notation such as $N(0.6, 0.04)$ or $N\left(0.6, (0.04)^2\right)$.
- *Z*-score: Displaying the correct mean and standard deviation in a *z*-score calculation that includes "*z*," such as $z = \frac{0.5 - 0.6}{0.04}$.

Component 2

- A response may satisfy component 2 by any of the following or a combination of the following:
  - Graphical: Displaying a graph of a normal density function with the region of interest ($A < 0.5$ or $Z < -2.5$) clearly identified. The shaded area does not need to be proportional, but the boundary should be on the proper side of the mean, and the shading should be in the proper direction.
  - Calculator function syntax: Identifying the lower and upper bounds of the region of interest in a "normalcdf" statement, such as:
    - $\text{normalcdf}(\text{lower} = -\infty, \text{upper} = 0.5, \text{mean} = 0.6, \text{standard deviation} = 0.04)$
    - $\text{normalcdf}(\text{lower} = -\infty, \text{upper} = -2.5, \mu = 0, \sigma = 1)$

    Correct specification of the mean and standard deviation is not required to satisfy component 2.
  - Words: Specifying the correct event in words with correct numerical values for the boundary value and correct direction, such as "the probability that the amount of shampoo is less than 0.5 liter" or $P(\text{amount of shampoo} < 0.5)$.
  - Standard Notation: Using standard notation such as: $P(A < 0.5)$ or $P\left(z < \frac{0.5 - 0.6}{0.04}\right)$ or $P(Z < -2.5)$.

General

- It is not necessary to define the random variable *A* because it is defined in the stem. It is not necessary to define the random variable *Z* because it is standard notation. Any other random variable must be defined correctly.
- An error in statistical notation, such as using *s* instead of σ for the population standard deviation or using $\bar{x}$ instead of μ for the population mean, does not satisfy component 1.
- If the only error in the response to part (a) is the reversal of the numerator for the *z*-score $(0.6 - 0.5)$, the response is scored P.
- An arithmetic or transcription error in a response can be ignored if correct work is shown.

Model Solution	Scoring
(b) (i) The random variable of interest, X, is the number of underfilled bottles in a box of 10 bottles. The distribution of X is binomial with parameters $n = 10$ and $p = 0.0062$. (ii) The crate will be rejected by the warehouse if two or more underfilled bottles are found in the box. The probability of that is $P(X \geq 2) = 1 - P(X \leq 1)$ $= 1 - \binom{10}{1}(0.0062)^1(0.9938)^9$ $- \binom{10}{0}(0.0062)^0(0.9938)^{10}$ $\approx 0.0017.$	**Essentially correct (E)** if the response satisfies the following four components: 1. Defines a random variable as the number of underfilled bottles in a box of 10 bottles in the response to part (b-i) 2. Indicates that the random variable has a binomial distribution with parameters $n = 10$ and $p = 0.0062$ (or the probability from part (a)). The parameters may be located in the response to either part (b-i) or part (b-ii) 3. Provides supporting work for the calculation of the probability in part (b-ii) that identifies the event of interest 4. Calculates the correct probability of approximately 0.0017, or a probability consistent with the response to part (a) or part (b-i) **Partially correct (P)** if the response satisfies only two or three of the four components. **Incorrect (I)** if the response does not satisfy the criteria for E or P.

Additional Notes:

Component 1

- A response may satisfy component 1 if the response indicates that the random variable is the number of underfilled bottles and $n = 10$ is used in the description of its distribution.

Component 2

- A response may satisfy component 2 by any of the following:
  - Binomial formula: Using the binomial formula with correct *n* and *p* values. For example: $1 - \binom{10}{1}(0.0062)^1(0.9938)^9 - \binom{10}{0}(0.0062)^0(0.9938)^{10}$.
  - Words or standard notation: Using a statement such as "binomial distribution with $n = 10$ and $p = 0.0062$," or using standard notation such as X B(10, 0.0062).
  - Calculator function syntax: Labeling correct parameter values in a "binomcdf" or "binompdf" statement such as:
    - 1 – binomcdf ($n = 10$, $p = 0.0062$, upper bound = 1)
    - 1 – binompdf ($n = 10$, $p = 0.0062$, $x = 0$) – binompdf ($n = 10$, $p = 0.0062$, $x = 1$)
  - Referring to a "box" does not satisfy the requirement for parameter $n = 10$.

Component 3

- A response may satisfy component 3 by any of the following:
  - Graphical display: Displaying a bar graph of binomial probabilities with appropriate bars shaded.

- Words or standard notation: Specifying the correct event in words with identification of the correct numerical boundary and correct direction, such as "probability that X is at least two" or "probability that at least two bottles are underfilled" or P(at least two bottles are underfilled). Identification of the distribution and parameters may be obtained from the response to part (b-i).
- Random variable: $P(X \geq 2)$ or $1 - P(X \leq 1)$. Identification of the distribution and parameters may be obtained from the response to part (b-i).
- Probability formula: e.g., $1 - \binom{10}{1}(0.0062)^1(0.9938)^9 - \binom{10}{0}(0.0062)^0(0.9938)^{10}$.
- Calculator function notation: Using calculator function notation with clearly defined arguments. For example:
  - "1 – binomcdf$(n = 10,\ p = 0.0062,$ upper bound $= 1)$" satisfies component 3 because the binomial parameters and the boundary value are clearly labeled.
  - "1 – binomcdf$(n = 10,\ p = 0.0062, 1)$" does not satisfy component 3 because the boundary value is not labeled.
  - "1 – binomcdf$(10,\ 0.0062,$ upper bound $= 1)$" does not satisfy component 3 because the binomial parameters are not labeled.

- Because $np = (10)(0.0062) = 0.062$ is less than 5, the normal approximation to the binomial distribution is not an appropriate method to calculate the probability, and a response that uses this method does not satisfy component 3. However, a response that uses the normal approximation to the binomial distribution may satisfy component 4 if it displays the correct mean and standard deviation of the binomial distribution AND provides a clear indication of the appropriate collection of possible outcomes included in the event using a diagram or a z-score, e.g., $P\left(Z \geq \frac{2 - (10)(0.0062)}{\sqrt{(10)(0.0062)(0.9938)}}\right)$ or $1 - P\left(Z \leq \frac{1 - (10)(0.0062)}{\sqrt{(10)(0.0062)(0.9938)}}\right)$. (Note that $\sqrt{(10)(0.0062)(0.9938)} \approx 0.248$.)
- An arithmetic or transcription error in a response can be ignored if correct work is shown.

Model Solution	Scoring
(c) The company should use the original programming for the filling machine. For the original programming of the filling machine, the probability of an underfilled bottle is $P(A < 0.5) = P\left(Z < \frac{0.5 - 0.60}{0.04}\right)$ $= P(Z < -2.5) \approx 0.0062.$ For the adjusted programming of the filling machine, the probability of an underfilled bottle is $P(A < 0.5) = P\left(Z < \frac{0.5 - 0.56}{0.03}\right)$ $= P(Z < -2.0) \approx 0.02275.$ Because the probability of an underfilled bottle is greater for the adjusted programming, this would result in more rejected shipments. The company should continue with the original machine programming.	**Essentially correct (E)** if the response satisfies the following two components by comparing either probabilities or z-scores: Comparing probabilities: 1. Correctly calculates the probability of underfilling a bottle as 0.023 for the adjusted programming of the filling machine 2. Provides a correct conclusion about which programming (adjusted or original) should be recommended based on a comparison of the probabilities calculated for the original and adjusted programming *OR* Comparing z-scores: 1. Correctly calculates the z-score for the adjusted programming 2. Provides a correct conclusion about which programming (adjusted or original) should be recommended based on a comparison of the z-scores (e.g., a higher z-score results in more bottles being underfilled) calculated for the original and adjusted programming **Partially correct (P)** if the response satisfies only one of the two components required for an E. **Incorrect (I)** if the response does not satisfy the criteria for E or P.

Additional Notes:

- A response that correctly uses the binomial distribution to find the probability that a crate will be rejected with correct values and justification should be scored E. For the original programming, this probability is 0.0017, and for the adjusted programming, this probability is 0.0206.
  Adjusted programming:
  Let Y represent the number of underfilled shampoo bottles in a box of 10 using the adjusted programming.
  $P(Y \geq 2) = 1 - P(Y \leq 1)$
  $= 1 - \binom{10}{1}(0.02275)^1(0.97725)^9$
  $-\binom{10}{0}(0.02275)^0(0.97725)^{10}$
  ≈ 0.0206
- A response that incorrectly computes the probability that a crate will be rejected, with or without justification, should be scored P if it provides a correct conclusion based on comparing that probability to the probability computed in part (b-ii).

- Component 2 is not satisfied if no recommendation is made for choice of programming. A response stating "yes" or "no" is not sufficient for indicating a choice of programming.
- An arithmetic or transcription error in a response can be ignored if correct work is shown.

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

Question 4: Focus on Inference — 4 points

General Scoring Notes

- This two-part question is scored in four sections. Each section is initially scored by determining if it meets the criteria for essentially correct (E), partially correct (P), or incorrect (I). Part (a) includes three sections that may appear in any order in the response. The first section includes identification of the appropriate confidence interval in part (a). The second section includes verifying the conditions for inference in part (a) and calculating the values of the endpoints of the confidence interval. The third section includes the interpretation of the confidence interval in part (a). The fourth section includes the response to part (b). The response is then categorized based on the scores assigned to each section and awarded an integer score between 0 and 4 (see the table at the end of the question).
- The model solution represents an ideal response to each section of the question, and the scoring criteria identify the specific components of the model solution that are used to determine the score.

	Model Solution	Scoring
(a) Section 1	The appropriate procedure is a one-sample z-interval for the proportion of all teenagers in the United States who would respond that they use a video streaming service every day.	**Essentially correct (E)** if the response satisfies the following two components: 1. Identifies the appropriate procedure as a one-sample z-interval by name or formula or by the calculations of the correct confidence interval endpoint values 2. States that the parameter of interest is the population proportion **Partially correct (P)** if the response satisfies only one of the two components. **Incorrect (I)** if the response does not meet the criteria for E or P.

Additional Notes:

- The response to component 2 concerning the statement of "population proportion" can be found in any of the three sections of part (a).
- Any notation used to represent sample proportion or population proportion should remain consistent throughout part (a).

	Model Solution	Scoring
(a) **Section 2**	This survey selected a random sample of 920 teenagers in the United States, which enables the interval to be generalized to the population of interest. This sample of 920 teenagers is less than 10% of the total number of teenagers in the United States, which is required as sampling was conducted without replacement from a finite population. In addition, there were more than 10 successes and 10 failures as $(920)(0.59) = 542.8$ (or 543) responded that they use a streaming service daily and $(920)(0.41) = 377.2$ (or 377) responded that they did not. Thus, the sample size is large enough to support the assumption that the sampling distribution of $\hat{p}$ is approximately normal. Therefore, a 95% confidence interval for the population proportion is given by $\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.59 \pm 1.96\sqrt{\frac{(0.59)(0.41)}{920}}$, which is 0.59 ± 0.032, and the interval is $(0.558, 0.622)$.	**Essentially correct (E)** if the response satisfies the following four components: 1. States that a random sample was selected 2. Indicates 920 is less than ten percent of all teenagers in the United States 3. Verifies that there are at least 10 successes and failures by calculating the following values $n\hat{p} = (920)(0.59) \approx 542.8$ and $n(1-\hat{p}) = (920)(1-0.59) \approx 377.2$ 4. Reports the values for a correct interval consistent with the procedure stated in Section 1 **Partially correct (P)** if the response satisfies either component 3 or component 4 and at least one of the other three components. **Incorrect (I)** if the response does not meet the criteria for E or P.

Additional Notes:

- Stating the large sample condition without verification is not sufficient for component 3.
- If the response includes an inappropriate check of conditions, such as $n > 30$, then component 3 is not satisfied.
- Supporting work, showing formulas or calculations, is not required for component 4.
- If the interval values are correct, the use of a one-sample *z* procedure for proportion can be used to satisfy component 1 of Section 1.
- A response that uses the value of $x = 543$ will result in an interval of $(0.5584, 0.6219)$, and a response that uses the value of $x = 542$ will result in an interval of $(0.5573, 0.6209)$. These interval endpoint values may be used to satisfy component 4.
- If the response includes supporting work for calculating the confidence interval that displays a correct formula with correct values inserted for $\hat{p}$, *n*, and *z*, then component 4 is satisfied even if values for the endpoints of the confidence interval are not displayed or calculated incorrectly.
- A response that computes an interval in percentages rather than proportions may satisfy component 4 if the response correctly indicates the use of percentages, $(55.8\%, 62.2\%)$.
- Minor errors or omissions when checking assumptions may be considered if holistic scoring is required.

	Model Solution	Scoring
(a) **Section 3**	We can be 95% confident that the proportion of all teenagers in the United States who would respond that they use a streaming service every day is between 0.558 and 0.622.	**Essentially correct (E)** if the response satisfies the following two components: 1. Indicates 95% confidence and interprets the interval using words such as "we are 95% confident" or "with 95% confidence" and provides interval endpoint values consistent with calculations in Section 2 2. Conveys inference about a population proportion in the proper context, i.e., "the true proportion of U.S. teenagers," "the population proportion of U.S. teenagers," or "the proportion of all U.S. teenagers" **Partially correct (P)** if the response satisfies only one of the two components. **Incorrect (I)** if the response does not meet the criteria for E or P.

Additional Notes:

- For component 1, responses that use the same incorrect proportion interval endpoint values calculated in Section 2 satisfies this portion of component 1.
- Section 3 is scored I if the response includes unreasonable values for a proportion or percentage.
- Responses must be consistent with the use of the terms proportion and percentage in component 2. A response that refers to the population percentage must use percentage values to satisfy component 2, and a response that refers to the population proportion must use proportion values.
- Any interpretation of the level of the interval, whether correct or incorrect, is considered extraneous and does not satisfy component 1.
- Clear indication of an inference to the sample of 920 teenagers, rather than the population of teenagers, does not satisfy component 2.
- Section 3 is scored I if it implies that the population proportion is a random variable. For example, it indicates that there is a 95 percent chance that the population proportion is between 0.558 and 0.622.
- A response that implies that we have found the actual percentage of U.S. teenagers who respond that they use a streaming service (deterministic) does not satisfy component 1.

	Model Solution	Scoring
(b) **Section 4**	The 95% confidence interval of $(0.558, 0.622)$ indicates that any value between 0.558 and 0.622 is a plausible value for the proportion of all teenagers in the United States who use video streaming every day. Because the value 0.5 is not contained in the interval, the sample data provide convincing statistical evidence that the proportion of all teenagers in the United States who would use a streaming service every day is not 0.5.	**Essentially correct (E)** if the response satisfies the following two components: 1. Provides a correct conclusion (there is convincing evidence or there is not convincing evidence that 0.50 is a plausible value), consistent with the interval calculated in part (a) 2. Provides correct justification based on whether the value of 0.5 is contained in the interval calculated in part (a) **Partially correct (P)** if the response satisfies only one of the two components. **Incorrect (I)** if the response does not meet the criteria for E or P.

Additional Notes:

- Any use of a hypothesis test to answer Section 4 cannot be used to satisfy component 2. However, the statement of a conclusion that 0.50 is not a plausible value based on the results of the hypothesis test can be used to satisfy component 1.
- A response that correctly uses incorrectly calculated proportion interval endpoint values can satisfy components 1 and 2 if the response is consistent with interval values.
- A response that includes a correct interpretation of the confidence interval in Section 4 can receive credit for the interpretation in Section 3 if no interpretation was provided in Section 3.
- A response that presents a correct conclusion in the context of the population percentage may satisfy both component 1 and component 2.
- A response that indicates that there is evidence that the proportion of teenagers in the United States who would use a streaming service every day is greater than 0.5 based on the fact that the entire confidence interval is above 0.5 may satisfy both component 1 and component 2.
- A response that does not provide the correct conclusion and justification but does recognize that the value is not in the interval may be considered a positive if holistic scoring is required.

Scoring for Question 4
Each essentially correct (E) part counts as 1 point, and each partially correct (P) part counts as ½ point.

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

If a response is between two scores (for example, 2 ½ points), use a holistic approach to decide whether to score up or down, depending on the strength of the response and quality of the communication.

Question 5: Multi-Focus — 4 points

General Scoring Notes

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

	Model Solution	Scoring
(a)	The sample median reduction in blood pressure for those eating dark chocolate, 7 mmHg, is greater than the sample median reduction in blood pressure for those eating white chocolate, about 0 mmHg.	**Essentially correct (E)** if the response satisfies the following three components: 1. The medians are correctly determined with the median for dark chocolate equal to 7 mmHg and the median for white chocolate between −2 mmHg and 2 mmHg, inclusive 2. The computed medians are compared correctly 3. The context of the response variable (reduction in blood pressure) and the treatment groups (dark chocolate and white chocolate) is included **Partially correct (P)** if the response satisfies two of the three components and, if stated, the reported values of the medians are reasonable *OR* if the response satisfies one of components 2 or 3 and the reported values of the medians are reasonable. **Incorrect (I)** if the response does not meet the criteria for E or P.

Additional Notes:

- To be scored as reasonable, the reported value of the median for the dark chocolate group must be between 5 mmHg and 8 mmHg, inclusive, and the reported value for the median of the white chocolate group must be between −2 mmHg and 3 mmHg, inclusive.

	Model Solution	Scoring
(b)	The researcher's conclusion may not necessarily be true because looking at the difference in sample means alone does not consider the variability in the sampling distribution of the differences in sample means. Another random assignment of subjects to dark and white chocolate may result in a sample mean reduction in blood pressure for subjects assigned to dark chocolate that is smaller than the mean reduction for subjects assigned to white chocolate. The variability in the sampling distribution of potential differences in sample means must be considered in making a conclusion about convincing statistical evidence. Further, an inference procedure can assess the likelihood that the observed difference in sample means occurred by random chance if the population means are equal.	**Essentially correct (E)** if the response satisfies at least two of the following three components: 1. Indicates that the researchers failed to consider the variability in the sampling distribution of differences in sample means (or that this should be considered) OR states that a different random assignment could produce different sample means 2. Includes an explanation that a difference of 5.66 mmHg could have occurred by random chance, even if the population means are equal 3. Indicates that an inference procedure is needed **Partially correct (P)** if the response satisfies only one of the three components. **Incorrect (I)** if the response does not meet the criteria for E or P.

Additional Notes:

- A response may be scored E if it states that no conclusion can be made about the relationship between the population mean blood pressure reduction for those who consume dark chocolate and those who consume white chocolate because the sample was drawn only from the subpopulation of *healthy* adults.

Model Solution	Scoring
(c) The observed value of the sample statistic $\bar{x}_{dark} - \bar{x}_{white}$ is 5.66 mmHg. The graph of simulation results reveals that a difference of 5.66 mmHg or larger occurred in only 3 of the 120 trials. Thus, the p-value is approximately equal to $\frac{3}{120} = 0.025$. Thus, assuming that there is no difference in mean blood pressure reduction, there is an approximate probability of 0.025 of getting a difference of 5.66 mmHg or larger by chance alone. Because this approximate p-value is less than 5 percent, there is convincing evidence that adding dark chocolate to the diet will result in a greater mean reduction in blood pressure than adding white chocolate to the diet for people similar to those who participated in the study.	**Essentially correct (E)** if the response satisfies the following four components: 1. Calculates the correct p-value of 0.025 2. Provides supporting work from the simulation for the calculation of the p-value 3. Provides a correct conclusion in the context of whether there is convincing evidence that adding dark chocolate to the diet will result in greater average reduction in blood pressure than adding white chocolate to the diet 4. Justifies the conclusion by directly comparing the calculated p-value to 0.05 **Partially correct (P)** if the response satisfies only two or three of the four components *OR* if the conclusion (component 3) is correct but is not in context and at least one additional component is satisfied *OR* if the conclusion (component 3) is in context and correct, but the justification is based on an analysis of the data shown in the dotplots (e.g., a two-sample t-confidence interval, a two-sample t-test, an appropriate nonparametric test, or an "exact" randomization test). **Incorrect (I)** if the response does not meet the criteria for E or P.

Additional Notes:

- Component 2 is satisfied by simply noting that only three of the simulated trials resulted in a difference of at least 5.66 mmHg, the observed difference.
- A response that incorrectly evaluates the p-value may still satisfy components 3 and 4 if the response is consistent with the use of the incorrect p-value.
- The hypotheses are not required because they are given in the stem.

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

Question 6: Investigative Task — 4 points

General Scoring Notes

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

Model Solution

(a) **(i)** The relative frequencies of successful and unsuccessful treatments in each clinic are presented in the following table:

	Clinic A	Clinic B
Unsuccessful Treatment	$\frac{51}{139} \approx 0.3669$	$\frac{33}{68} \approx 0.4853$
Successful Treatment	$\frac{88}{139} \approx 0.6331$	$\frac{35}{68} \approx 0.5147$

(ii) Clinic A appears to be more successful in treating allergy sufferers than Clinic B. Clinic A was successful for 63.3% of the allergy sufferers it treated, while Clinic B was successful for only 51.5% of the allergy sufferers it treated.

Scoring

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

1. The response to part (a-i) provides correct numerical values for the relative frequencies in all four cells of the table
2. The response to part (a-ii) identifies the clinic that is more successful in treating allergy sufferers and includes a justification based on the relative frequencies reported in part (a-i)

Partially correct (P) if the response satisfies only one of the two components
OR
if the response provides correct numerical values for the relative frequencies in only two or three of the cells of the table, and correctly identifies the clinic that has greater success, with or without providing adequate justification based on the values of the reported relative frequencies.

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

Additional Notes:

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

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

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

Additional Notes:

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

	Model Solution	Scoring
(c)	**(i)** **Clinic A:** More successful in treating mild allergies. Clinic A successfully treated 75.0% of mild allergy sufferers, while Clinic A successfully treated only 28.6% of severe allergy sufferers. **Clinic B:** More successful in treating mild allergies. Clinic B successfully treated 91.7% of mild allergy sufferers, while Clinic B successfully treated only 42.9% of severe allergy sufferers. **(ii)** **Clinic A:** More likely to treat mild allergy sufferers than severe allergy sufferers. Of the 139 allergy sufferers treated at Clinic A, 104 (74.8%) suffered from mild allergies, while only 35 (25.2%) suffered from severe allergies. **Clinic B:** More likely to treat severe allergy sufferers than mild allergy sufferers. Of the 68 allergy sufferers treated at Clinic B, 56 (82.4%) suffered from severe allergies, while only 12 (17.6%) suffered from mild allergies.	**Essentially correct (E)** if the response satisfies the following four components: 1. The response to part (c-i) indicates that both Clinic A and Clinic B are more successful in treating mild allergies than treating severe allergies 2. The response to part (c-i) provides a justification based on correctly reported proportions (or percentages) of successfully treated allergy sufferers of each severity for each clinic 3. The response to part (c-ii) indicates that Clinic A is more likely to treat mild allergy sufferers, while Clinic B is more likely to treat severe allergy sufferers 4. The response to part (c-ii) provides a justification based on correctly reported frequencies or proportions (or percentages) of severe or mild allergy sufferers treated by each clinic **Note:** See the Q6 Correct Relative Frequencies tables at the end of this scoring guideline for the relative frequencies that should be used in each designated part. **Partially correct (P)** if the response satisfies two or three of the four components. **Incorrect (I)** if the response does not meet the criteria for E or P.

Additional Notes:

- A response to part (c-i) may satisfy component 2 in either of the following ways:
  - For each clinic, reports the correct proportion (or percentage) of severe allergy sufferers successfully treated AND the correct proportion (or percentage) of mild allergy sufferers successfully treated.
  - Makes a correct reference to the mosaic plot, comparing the ratio of shaded area to open area within the "Mild" rectangle to the ratio of shaded area to open area within the "Severe" rectangle for each clinic. Numerical values are not required.
- A response to part (c-ii) may satisfy component 4 in any of the following ways:
  - For each clinic, reports the correct proportion (or percentage) of treated allergy sufferers with severe allergies. (Clinic A: 0.252, Clinic B: 0.824)
  - For each clinic, reports the correct proportion (or percentage) of treated allergy sufferers with mild allergies. (Clinic A: 0.748, Clinic B: 0.176)

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

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

Additional Notes:

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

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

Q6 Correct Relative Frequencies

(a-i)

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

(c-i)

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

(c-ii)

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

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

Scoring for Question 6 Each essentially correct (E) part counts as 1 point, and each partially correct (P) part counts as ½ point.	**Score**
Complete Response	**4**
Substantial Response	**3**
Developing Response	**2**
Minimal Response	**1**
If a response is between two scores (for example, 2 ½ points), use a holistic approach to decide whether to score up or down, depending on the strength of the response and quality of the communication.	