2026

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 five-number summary for the Breed H goats is as follows: Minimum: 48 pounds Q1: 56 pounds Median: 64 pounds Q3: 72 pounds Maximum: 80 pounds When there are an even number of observations, the median is typically calculated by taking the mean of the two middle values or $\frac{62 + 66}{2} = 64$ pounds.	**Essentially correct (E)** if the response satisfies the following four components: 1. The response correctly identifies the minimum and maximum values. 2. The response correctly identifies the lower quartile. 3. The response correctly identifies the upper quartile. 4. The response correctly identifies the median. **Partially correct (P)** if the response satisfies only two or three of the four components required for E. **Incorrect (I)** if the response does not meet the criteria for E or P.

Additional Notes:

- The median is the middle value when data are ordered. Because there are 14 data points, the median can be any value between 62 pounds and 66 pounds. Therefore, any value between 62 pounds and 66 pounds, inclusive, may satisfy component 4 except a calculation of the mean (i.e., 64.07).
- A response that correctly states a median other than 64 may satisfy components 2 and 3 by correctly calculating the appropriate quartiles (e.g., “a median of 62 may result in a lower quartile of 55.5, and a median of 66 may result in an upper quartile of 72.5”).
- Any discussion of the mean, standard deviation, or IQR should be disregarded in scoring.
- If the response uses an incorrect procedure to determine the median and the same incorrect procedure is used for both quartiles, components 2 and 3 may still be satisfied.
- Inclusion or omission of units of measurement should be disregarded in scoring.
- If the response includes exactly five unlabeled numbers expressed together as a vertical or horizontal list, interpret the numbers as being labeled as the minimum, lower quartile, median, upper quartile, and maximum, in this order.
- A response that includes only the five correct values for the five-number summary without providing a complete set of labels or without putting them in an ordered list may earn a score of P.

	Model Solution	Scoring
B	The medians for the distribution of weight for Breed H goats and the distribution of weight for Breed J goats are approximately the same. (The median for each distribution is approximately 64 pounds.) The range for the distribution of weight for Breed H goats is smaller (32 pounds) than the range for the distribution of weight for Breed J goats (approximately 40 pounds). The IQR for the distribution of weight for Breed H goats is smaller (16 pounds) than the IQR for the distribution of weight for Breed J goats (approximately 24 pounds).	**Essentially correct (E)** if the response satisfies the following three components: 1. The response correctly indicates that the median for the distribution of weight for Breed H goats is approximately the same as or the same as the median for the distribution of weight for Breed J goats or provides a correct comparison of the median for each group based on the response in part A. 2. The response correctly indicates that the range or IQR for the distribution of weight for Breed H goats is less than the range or IQR for the distribution of weight for Breed J goats or provides a correct comparison of the range or IQR for each group based on the response in part A. 3. The response includes context when describing the distribution of the weight of goat breeds. **Partially correct (P)** if the response satisfies only two of the three components required for E. **Incorrect (I)** if the response does not meet the criteria for E or P.

Additional Notes:

- A response need not include specific numerical values to satisfy any given component or to score E on part B.
- A response that indicates the medians are not the same and justifies the decision with reasonable values may satisfy component 1 (e.g., “the Breed H median of 64 is greater than the Breed J median of 63”).
- A response that discusses comparing means without a comparison of medians does not satisfy component 1.
- A response that only lists values for center and/or spread and does not compare them does not satisfy components 1 and/or 2.
- Context for component 3 must include the variable (weight or pounds), the two groups (H and J), and goats or breeds.
- Discussions of shape or outliers should be disregarded in scoring.

	Model Solution	Scoring
C	i. From the stem-and-leaf plot, it is apparent that the distribution of weight for the Breed H goats is bimodal. From a boxplot, it would not be apparent that the distribution of weight for Breed H goats is bimodal. ii. From the boxplot alone, it can be determined that the distribution of weight for Breed H goats is approximately symmetric. However, a boxplot may appear symmetric without revealing other characteristics of the distribution, such as whether it is bimodal or uniform.	**Essentially correct (E)** if the response satisfies the following two components: 1. In part C (i) the response indicates that the stem-and-leaf plot makes it apparent that the distribution of the weight for the Breed H goats is bimodal. 2. In part C (ii) the response indicates that a boxplot does not give specific values for the data, only uses summary statistics, or does not indicate peaks. **Partially correct (P)** if the response satisfies only one of the two components required for E. **Incorrect (I)** if the response does not meet the criteria for E or P.

Additional Notes:

- Component 1 may be satisfied by using a description of bimodal such as “two peaks” or “two clusters.”
- A response indicating that a boxplot does not show the frequency of data values may satisfy component 2.
- A response that indicates the boxplot is normal or approximately normal cannot be scored an E. It may be scored P if components 1, 2, or both are satisfied.

Scoring for Question 1	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 2: Focus on Sampling and Experimental Design — 4 points

General Scoring Notes

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

Model Solution	Scoring
A i. There are two treatments: • one-half cup of coffee grounds added to the soil of the rosebush weekly • no coffee grounds added to the soil of the rosebush ii. The experimental units are the rosebushes. iii. The response variable is the number of roses on each rosebush.	**Essentially correct (E)** if the response satisfies the following three components: 1. In part A (i) the response correctly identifies the treatments. 2. In part A (ii) the response correctly identifies the experimental units. 3. In part A (iii) the response correctly identifies the response variable. **Partially correct (P)** if the response satisfies only two of the three components required for E. **Incorrect (I)** if the response does not meet the criteria for E or P.

Additional Notes:

- A response that identifies the treatments as "coffee grounds" and "no coffee grounds" may satisfy component 1.
- A response that identifies the experimental units as "soil" satisfies component 2.
- Identification of the response variable as a mean, or average, does not satisfy component 3.

	Model Solution	Scoring
B	Holly will give each rosebush a unique number ranging from 1 to 30. She will use a random number generator to generate 15 unique numbers from 1 to 30, inclusive. The 15 rosebushes corresponding to these 15 generated numbers will go into Group 1 and will have one-half cup of coffee grounds added to their soil weekly. The rosebushes corresponding to the remaining 15 numbers will go into Group 2 and will not have coffee grounds added to their soil. *Alternative Solution:* Using 30 equally sized slips of paper, Holly will label 15 slips "one-half cup of coffee grounds added to the soil weekly" and the other 15 slips "no coffee grounds added to the soil." She will then mix the slips of paper in a bowl. For each rosebush, Holly will select a slip of paper (without replacement) to determine the treatment the rosebush receives.	**Essentially correct (E)** if the response satisfies the following three components: 1. The response creates the appropriate labels for the units and treatments. 2. The response describes how to correctly implement the random process so that every possible random assignment is equally likely. 3. The response indicates a random process that results in 15 rosebushes being assigned to each treatment (coffee grounds weekly or no coffee grounds). **Partially correct (P)** if the response satisfies only two of the three components required for E. **Incorrect (I)** if the response does not meet the criteria for E or P.

Additional Notes:

- For responses that use slips of paper or marbles (or equivalent) to represent treatments:
  - To satisfy component 1, some slips must be labeled or assigned to represent "one-half cup of coffee grounds added to the soil weekly" and some labeled or assigned to represent "no coffee grounds added to the soil" (e.g., "blue marbles represent 'one-half cup of coffee grounds weekly' and yellow marbles represent 'no coffee grounds'").
  - To satisfy component 2, the slips of paper must be mixed or shuffled, and the response must clearly link the treatment selected to a rosebush (e.g., "for each rosebush a slip is selected to determine the treatment").
  - To satisfy component 3, the response must indicate that there are 15 slips (or marbles of a specific color) for each treatment, and the response must indicate that slips of paper (or marbles) are selected without replacement.
- For responses that use numbered slips of paper:
  - To satisfy component 1, the rosebushes must be labeled 1 to 30 (or equivalent unique identifiers).
  - To satisfy component 2, the slips of paper must be mixed or shuffled, and the response must clearly link the number selected to a rosebush that receives one-half cup of coffee grounds added to its soil weekly or a rosebush that receives no coffee grounds added to its soil.
  - To satisfy component 3, the response must indicate that slips of paper are selected without replacement and that 15 rosebushes are assigned to each treatment.

- For responses that use a random number generator with rosebushes labeled from 1 to 30 (or an equivalent interval):
  - To satisfy component 1, the rosebushes must be labeled 1 to 30.
  - To satisfy component 2, the response indicates that the random number generator selects numbers 1 to 30, and the response clearly links the number selected to a rosebush that receives the one-half cup of coffee grounds added to its soil weekly or a rosebush that receives no coffee grounds added to its soil.
  - To satisfy component 3, the response must indicate that numbers are selected without repeats and that 15 rosebushes are assigned to each treatment.
- For responses that flip a coin for each rosebush (or roll a die and note odd/even, generate a random number 1 or 2, or equivalent):
  - To satisfy component 1, each outcome must be linked to a treatment (e.g., "heads = one-half cup of coffee grounds added to its soil weekly and tails = receives no coffee grounds added to its soil").
  - To satisfy component 2, the response must clearly link the outcome of the coin flip to a rosebush (e.g., "for each rosebush, Holly will flip a coin to determine the treatment"). Note: If the response includes a stopping rule (e.g., when 15 rosebushes are assigned one treatment, the remaining rosebushes are assigned the other treatment), component 2 is not satisfied because this plan increases the probability that the last rosebushes have the same treatment, which does not meet the equally likely random assignment requirement.
  - To satisfy component 3, the response must indicate that 15 rosebushes are assigned to each treatment using a stopping rule. If there is no stopping rule, component 3 is not satisfied.
- If a response uses a random number generator or slips of paper with the numbers 1 to 30 and does not initially number the rosebushes from 1 to 30, component 1 may be satisfied if the response indicates a link between the number selected and the rosebush (e.g., "if 3 is selected, the third rosebush receives the one-half cup of coffee grounds weekly").
- A response that uses a nonrandom process should be scored I (e.g., "number the rosebushes from 1 to 30 and assign one-half cup of coffee grounds weekly to odd-numbered rosebushes").
- A response that uses blocking does not satisfy component 2 because all possible random assignments are not equally likely.
- If the response describes two ways to perform the random assignment, assign the score for the weaker assignment process.

	Model Solution	Scoring
C	Assuming that coffee grounds have no effect on the number of roses, it is unlikely $(\text{probability} < 0.05)$ that there would be a difference in the sample mean number of roses as big or bigger than the observed difference in sample means by random chance alone.	**Essentially correct (E)** if the response satisfies the following three components: 1. The response includes a reference to a difference in sample results. 2. The response indicates that the results of the experiment are unlikely to happen by random chance. 3. The response includes context. **Partially correct (P)** if the response satisfies only two of the three components required for E. **Incorrect (I)** if the response does not meet the criteria for E or P.

Additional Notes:

- To satisfy component 1, the response must include a reference to a difference in results (e.g., difference in means) or a comparison of results (e.g., "the results from the coffee group are greater than the results from the control group").
- To satisfy component 2, a response may:
  - state that the difference in results is unlikely to happen by random chance;
  - state that the difference in results is larger than expected by random chance; or
  - state that the p-value is less than 0.05.
- To satisfy component 3, the response must minimally include a reference to the number of roses.

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	For a normal distribution with mean $\mu = 109$ seconds and standard deviation $\sigma = 16$ seconds, the probability that a randomly selected performance of the team song will last longer than 120 seconds is $P\left(Z > \frac{120 - 109}{16}\right)$ $= P(Z > 0.6875)$ $= 0.2459$.	**Essentially correct (E)** if the response satisfies the following three components: 1. The response indicates the use of a normal distribution and identifies the correct parameter values (mean 109 and standard deviation 16). 2. The response specifies the correct event, including the correct boundary value (performance $= 120$ or $z = 0.6875$) and the correct direction. 3. The response provides the correct probability of 0.2459 or probability consistent with component 1 and component 2. **Partially correct (P)** if the response satisfies only two of the three components required for E. **Incorrect (I)** if the response does not meet the criteria for E or P.

Additional Notes:

Component 1

- A response that indicates a *t*-distribution does not satisfy 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 (109) and standard deviation (16).
  - **Calculator function syntax:** Labeling correct values of the mean and standard deviation in an appropriate calculator statement, such as
    $\text{normalcdf}(\text{lower} = 120, \text{upper} = \infty, \text{mean} = 109, \text{standard deviation} = 16)$,
    $\text{normalcdf}(\text{lower} = 0.6875, \text{upper} = \infty, \text{mean} = 0, \text{standard deviation} = 1)$, or
    $\text{normaldist}(\text{mean} = 109, \text{standard deviation} = 16)$.
    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 109 and standard deviation 16.”

- **Standard notation:** Using standard notation such as $N(109,16)$ or $N\left(109,16^2\right)$.
- ***Z*-score:** Displaying the correct mean and standard deviation in a *z*-score calculation that includes "*z*," such as $z = \frac{120 - 109}{16}$.

- 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.

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 $(\text{performance} > 120 \text{ or } Z > 0.6875)$ 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.
  - **Words:** Specifying the correct event in words with correct numerical values for the boundary value and correct direction, such as "the probability that the length of the performance is more than 120 seconds" or $P(\text{performance} > 120)$.
  - **Standard notation:** Using standard notation, such as $P(\text{performance} > 120)$, or $P\left(Z > \frac{120 - 109}{16}\right)$, or $P(Z > 0.6875)$.
  - **Calculator function syntax:** Labeling the lower and upper bounds of the region of interest in an appropriate calculator statement such as
    $\text{normalcdf}(\text{lower} = 120, \text{upper} = \infty, \text{mean} = 109, \text{standard deviation} = 16)$ or
    $\text{normalcdf}(\text{lower} = 0.6875, \text{upper} = \infty, \text{mean} = 0, \text{standard deviation} = 1)$.
    An upper bound need not be specified. The minimum value for a numerical specification of an upper bound must be at least four standard deviations above the mean (i.e., at least 173).
- A response that includes a lower boundary of 121 instead of 120 does not satisfy component 2. Note that the probability associated with a boundary of 121 is 0.2266.

General

- If the response only includes a *z*-score formula with a reversal of the numerator for the *z*-score $(109 - 120)$ and the resulting probability, the response is scored as P, even though components 1 and 2 are not satisfied.
- An arithmetic or transcription error in a response can be ignored if correct work is shown.
- For a response that is scored I, the number of components satisfied should be considered if holistic scoring is required.

	Model Solution	Scoring
B	Let the random variable of interest X represent the number of performances of the team song lasting more than 120 seconds. Because the probability that a performance of the team song lasting more than 120 seconds is 0.2459 and each performance is independent of other performances of the team song, X has a binomial distribution with $n = 10,$ repeated independent trials, and a probability of success that is the same $(p \approx 0.2459)$ for each trial. The probability that at least three performances of the team song last more than 120 seconds is $P(X \geq 3) = 1 - P(X \leq 2)$ $P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2)$ $P(X \leq 2) = \binom{10}{0}(0.2459)^0(0.7541)^{10} +$ $\binom{10}{1}(0.2459)^1(0.7541)^9 +$ $\binom{10}{2}(0.2459)^2(0.7541)^8$ $P(X \leq 2) = 0.0595 + 0.1939 + 0.2846$ $P(X \leq 2) = 0.5380,$ so $P(X \geq 3) = 1 - P(X \leq 2)$ $P(X \geq 3) = 1 - 0.5380$ $P(X \geq 3) = 0.4620$.	**Essentially correct (E)** if the response satisfies the following three components: 1. The response describes the distribution as binomial with $n = 10$ and $p \approx 0.2459$ or a probability consistent with the response in part A. 2. The response provides supporting work for the calculation of the probability that identifies the event of interest (i.e., identifies the correct boundary AND direction for the event). 3. The response calculates the correct probability of 0.4620 or a probability consistent with components 1 and 2. **Partially correct (P)** if the response satisfies only two of the three components required for E. **Incorrect (I)** if the response does not meet the criteria for E or P.

Additional Notes:

Component 1

- A response may satisfy component 1 by any of the following:
  - **Words and standard notation:** Using a statement such as "binomial distribution with $n = 10$ and $p = 0.2459$" or using standard notation such as $X \sim B(10, 0.2459)$.
  - **Binomial formula:** Using the binomial formula with correct n and p values. For example, $= 1 - \binom{10}{0}(0.2459)^0(0.7541)^{10} - \binom{10}{1}(0.2459)^1(0.7541)^9 - \binom{10}{2}(0.2459)^2(0.7541)^8$.
  - **Calculator function syntax:** Labeling correct parameter values with an appropriate calculator statement, such as:
    - $1 - \text{binomcdf}(n = 10, p = 0.2459, \text{upper bound} = 2)$
    - $1 - \text{binomcdf}(n = 10, p = 0.2459, \text{x} = 2)$
    - $\text{binomcdf}(n = 10, p = 0.2459)$
    - $\text{binomialdist}(n = 10, p = 0.2459)$

- $1 - \begin{pmatrix} \text{binompdf}(n = 10, p = 0.2459, \text{upper bound} = 0) \\ +\text{binompdf}(n = 10, p = 0.2459, \text{upper bound} = 1) \\ +\text{binompdf}(n = 10, p = 0.2459, \text{upper bound} = 2) \end{pmatrix}$

- $1 - \text{binomcdf}(10, 0.2459, \text{upper bound} = 2)$ does not satisfy component 1 because the binomial parameters are not labeled.
- $\text{binomialdist}(10, 0.2459)$ does not satisfy component 1 because the binomial parameters are not labeled.

Component 2

- A response may satisfy component 2 by any of the following:
  - **Graphical:** 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 three" or "probability that at least three team songs are more than 120 seconds long" or "P (at least three songs are more than 120 seconds long)."
  - **Random variable:** $P(X \geq 3)$ or $1 - P(X \leq 2)$.
  - **Probability formula:** For example,
    $1 - \binom{10}{0}(0.2459)^0(0.7541)^{10} - \binom{10}{1}(0.2459)^1(0.7541)^9 - \binom{10}{2}(0.2459)^2(0.7541)^8$.
  - **Calculator function notation:** Using the calculator function notation with clearly defined arguments. For example:
    - $1 - \text{binomcdf}(n = 10, p = 0.2459, \text{upper bound} = 2)$
    - $1 - \text{binomcdf}(n = 10, p = 0.2459, \text{x} = 2)$
    - $1 - \begin{pmatrix} \text{binompdf}(n = 10, p = 0.2459, \text{upper bound} = 0) \\ +\text{binompdf}(n = 10, p = 0.2459, \text{upper bound} = 1) \\ +\text{binompdf}(n = 10, p = 0.2459, \text{upper bound} = 2) \end{pmatrix}$

- $1 - \text{binomcdf}(n = 10, p = 0.2459, 2)$ does not satisfy component 2 because the boundary value is not labeled.
- Because $np = (10)(0.2459) = 2.459$ 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 1. However, a response that uses the normal approximation to the binomial distribution may satisfy component 2 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, such as $P\left(Z \geq \frac{3 - (10)(0.2459)}{\sqrt{(10)(0.2459)(0.7541)}}\right)$. Note that $\sqrt{(10)(0.2459)(0.7541)} = 1.362$.

Component 3

- A response that calculated the correct probability in part A will result in a range between 0.4441 and 0.4744 and may satisfy component 3.
- An arithmetic or transcription error in a response can be ignored if correct work is shown.

General

- For a response that is scored I, the number of components satisfied should be considered if holistic scoring is required.

Model Solution	Scoring
C i. The random variable Y is the number of the games at which the first performance of the team song lasts longer than 120 seconds. The distribution of Y is geometric with parameter $p \approx 0.2459$. The mean of Y is $\mu_Y = \frac{1}{p} = \frac{1}{0.2459} \approx 4.07$ games. ii. The standard deviation of Y is $\sigma_Y = \frac{\sqrt{1-p}}{p} = \frac{\sqrt{1-0.2459}}{0.2459}$ $= \frac{0.8684}{0.2459} \approx 3.53$ games.	**Essentially correct (E)** if the response satisfies the following four components: 1. In part C (i) the response indicates the use of a geometric distribution with $p \approx 0.2459$ OR a probability that is consistent with part A. 2. In part C (i) the response correctly calculates the mean of a geometric distribution OR a mean that is consistent with component 1. 3. In part C (ii) the response correctly calculates the standard deviation of a geometric distribution OR a standard deviation consistent with component 1. 4. In part C (i) or C (ii) the response provides supporting work for at least one of the correct calculations in component 2 or 3. **Partially correct (P)** if the response satisfies only two or three of the four components required for E. **Incorrect (I)** if the response does not meet the criteria for E or P.

Additional Notes:

Component 1

- A response may satisfy component 1 by any of the following:
  - **Words:** Using a statement such as "geometric distribution with probability 0.2459."
  - **Standard notation:** Using standard notation such as $G(0.2459)$.
  - **Geometric formula:** Displaying the correct probability in the formula for a mean or standard deviation of a geometric distribution.
  - **Calculator function syntax:** Labeling the correct value of the probability in an appropriate calculator statement, such as $\text{geometpdf}(p = 0.2459, x)$ or $\text{geometcdf}(p = 0.2459, x)$. Specification of x is not required to satisfy component 1.

General

- An arithmetic or transcription error in a response can be ignored if correct work is shown.
- A response with the mean rounded to 4 or 5 does not satisfy component 2 but may satisfy component 4.
- A response with a standard deviation rounded to 3 or 4 does not satisfy component 3 but may satisfy component 4.
- If the responses are not labeled, assume they are in the correct order for scoring purposes.
- For a response that is scored P or I, the number of components satisfied should be considered if holistic scoring is required.

	Model Solution	Scoring
D	Over repeated trials, the number of games a person attends until a performance of the team song is longer than 120 seconds deviates, on average, from the mean of 4.07 by 3.53 games.	**Essentially correct (E)** if the response satisfies the following three components: 1. The response indicates that the standard deviation is the average or typical variation from the mean. 2. The response indicates repeated trials or in the long run. 3. The response interprets the value of the standard deviation within the context of the problem. **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:

- To satisfy component 3, a response must clearly define the variable including the number of events in context (i.e., "number of games," "number of songs," or "number of performances") and a success outcome (e.g., "song longer than 120 seconds" or "long performance").
- Numerical values are not required in the interpretation, but if numerical values are included, those values must be consistent with the response in part C. If they are inconsistent, lower the score from E to P or P to I.
- A response with an incorrect measurement unit, such as "3.53 seconds" instead of "3.53 games," does not satisfy component 3.
- The quality of communication in part D should be considered if holistic scoring is required.

Scoring for Question 3 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 4: Focus on Inference — 4 points

General Scoring Notes

- This question is scored in three sections. Each section is initially scored by determining if it meets the criteria for essentially correct (E), partially correct (P), or incorrect (I). The first section includes statements of the null and alternative hypotheses and identification of the appropriate hypothesis test. The second section includes verifying the conditions for the test identified in the first section and calculating the value of the test statistic and the corresponding *p*-value. The third section includes the conclusion for the test identified in the first section. 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
Section 1	Let μ_C represent the population mean number of oranges on trees when fertilized with Brand C for all trees similar to those in the study. Let μ_N represent the population mean number of oranges on trees when fertilized with Brand N for all trees similar to those in the study. The null and alternative hypotheses are: $H_0 : \mu_C = \mu_N$ $H_a : \mu_C \neq \mu_N$ An appropriate inference procedure is a two-sample *t*-test for a difference of population means.	**Essentially correct (E)** if the response satisfies the following three components: 1. The response identifies a two-sample *t*-test for a difference of two population means by name (e.g., "two-sample *t*-test for means" or "two-sample *t*-test") or by formula. 2. The response states a correct null hypothesis of equal means AND a correct two-sided alternative hypothesis of unequal means. 3. The response provides sufficient context by referencing the two groups (Brand C fertilizer and Brand N fertilizer) AND by referencing the populations (all orange trees treated with Brand C and all orange trees treated with Brand N). **Partially correct (P)** if the response satisfies only two of the three components required for E. **Incorrect (I)** if the response does not meet the criteria for E or P.

Additional Notes:

- If the response identifies the correct test by name but also states an incorrect formula, then component 1 is not satisfied.
- If the response identifies the test using the correct formula but equating it with a *z* instead of a *t*, then component 1 is not satisfied.
- Neither group names nor the concept of the population is required to satisfy component 2.

- A response that states the hypotheses in words (e.g., "the null hypothesis is that the means are equal and the alternative hypothesis is that the means are not equal") may satisfy component 2. If the hypotheses also include the group names (Brand C and Brand N) and reference to the populations, then both components 2 and 3 may be satisfied.
- Group or population aspects of component 3 may be satisfied with the statement in the hypotheses, the definition of the parameters, or the statement of the conclusion, provided that there are no incorrect references to groups or population elsewhere in the response (e.g., using $\bar{x}$ in hypotheses or referring to samples in the conclusion).
- If the response clearly refers to sample means instead of population means using words or symbols (e.g., $\bar{x}_C$ and $\bar{x}_N$), then component 2 may be satisfied, but component 3 is not satisfied unless the symbols used are defined as population means.
- A response may satisfy the population aspect of component 3 by:
  - Referring to the population using words such as "population," "all," or "true" when defining the parameters.
  - Using notation such as μ_C, μ_N, μ_1, or μ_2 when defining the hypothesis statements. Note that subscripts C and N would be sufficient to also satisfy the group aspect of component 3, whereas subscripts of 1 and 2 are not sufficient to satisfy the group aspect of component 3 without further clarification.

Matched Pairs Approach

- Because a response presenting a matched pairs *t*-test approach cannot satisfy component 1, such a response cannot meet the criteria for a score of E. If the response satisfies components 2 and 3, it may be scored P; otherwise, it is scored I. (Note that null and alternative hypotheses about μ_d satisfy component 2 ONLY if matched pairs is explicitly named as the approach.)
- A response that attempts a matched pairs approach using a confidence interval may be scored P if component 3 is satisfied and the response indicates that the confidence interval is for the mean difference. Otherwise, it is scored I.

Confidence Interval Approach

- If a two-sample *t*-interval for a difference in population means is identified correctly by name (e.g., "two-sample *t*-interval for the difference in population means" or "two-sample *t*-interval") or by formula, then component 1 may be satisfied.
- If the response uses individual one-sample *t*-intervals for the two means, which is not a correct approach, then component 1 is not satisfied.
- For a response using a confidence interval approach, component 2 is satisfied if the response clearly indicates that the confidence interval is used to assess the correct two-sided alternative to the null hypothesis of equal means.

	Model Solution	Scoring
Section 2	The randomization condition for performing the two-sample *t*-test for a difference of two population means is satisfied because the farmer randomly assigned the brand of fertilizer to the trees. Because the sample size is greater than 30 for each brand, the sampling distribution for $\bar{x}_C - \bar{x}_N$ is approximately normal. The value of the test statistic is $t = \frac{\bar{x}_C - \bar{x}_N}{\sqrt{\frac{s_C^2}{n_C} + \frac{s_N^2}{n_N}}} = \frac{141 - 148}{\sqrt{\frac{(15)^2}{58} + \frac{(19)^2}{58}}}$ $\approx \frac{-7}{3.179} \approx -2.202.$ The corresponding *p*-value with $df \approx 108.174$ is $2 * P(t < -2.202) \approx 0.0298.$	**Essentially correct (E)** if the response satisfies the following four components: 1. The response checks the randomization condition by referring to the random assignment of treatments to experimental units. 2. The response indicates that each of the distributions had a sample size greater than 30, so the sampling distribution of the difference in population means is approximately normal. 3. The response provides the correct *t*-statistic. 4. The response provides the correct *p*-value, consistent with the reported test statistic and the stated alternative hypothesis. **Partially correct (P)** if the response satisfies only two or three of the four components required for E. **Incorrect (I)** if the response does not meet the criteria for E or P.

Additional Notes:

- Component 1 is not satisfied if the response indicates that the randomization condition is met because the sample was selected at random from all orange trees or all of the farmer's orange trees. This would include a response that states, "random samples–check" or "SRSs–check."
- It is minimally acceptable to state "randomly assigned" to satisfy component 1.
- It is minimally acceptable to state "sample greater than 30 in BOTH types of fertilizer" to satisfy component 2.
- Values for the test statistic and *p*-value depend on how the degrees of freedom were determined. The following combinations of *t*-values and *p*-values all satisfy components 3 and 4, respectively:
  - The degrees of freedom for $t \approx -2.202$ from the model solution were computed with the Satterthwaite approximation used by many calculators:

$$df = \frac{\left(\frac{s_C^2}{n_C} + \frac{s_N^2}{n_N}\right)^2}{\frac{1}{n_C - 1}\left(\frac{s_C^2}{n_C}\right)^2 + \frac{1}{n_N - 1}\left(\frac{s_N^2}{n_N}\right)^2} = \frac{\left(\frac{(15)^2}{58} + \frac{(19)^2}{58}\right)^2}{\frac{1}{58 - 1}\left(\frac{(15)^2}{58}\right)^2 + \frac{1}{58 - 1}\left(\frac{(19)^2}{58}\right)^2} \approx 108.174.$$

  - Using $t \approx -2.202$ and $(58 - 1) = 57$ degrees of freedom, the resulting *p*-value is 0.032.
  - Using a pooled *t*-test with $t \approx -2.202$ and 114 degrees of freedom (which would be appropriate because the standard errors for the two samples are similar), the pooled estimate of the standard error is 17.117 and the resulting *p*-value is 0.0297.
- Discussion of degrees of freedom will be treated as extraneous in scoring section 2.

- If the response incorrectly identifies the test as a *z*-test in section 1, the *z*-statistic of -2.202 satisfies component 3 and a *p*-value of 0.0276 satisfies component 4.
- If the response does not identify the *z*-test in section 1 but reports the test statistic as $z \approx -2.202$ (instead of *t*) then component 3 is not satisfied; however, component 4 may still be satisfied if the correct *p*-value of 0.0276 is reported with a *z*-statistic of -2.202.
- A response that reports the correct value for the *t*-statistic but contains errors in supporting work may still satisfy component 3.
- If the response compares the value of the test statistic to a critical value instead of computing a *p*-value, then a comparison consistent with the stated alternative hypothesis satisfies component 4.
- If the response omits the hypotheses, the correct two-sided alternative hypothesis is assumed when scoring component 4.
- If an incorrect alternative hypothesis is stated, then the *p*-value must be consistent with the stated alternative hypothesis to satisfy component 4.
- If the response satisfies component 4, any supporting work for the *p*-value may be treated as extraneous.

Matched Pairs Approach

- Section 2 of a response using a matched pairs *t*-test or a matched pairs confidence interval must be scored I.

Confidence Interval Approach

- If the stated alternative hypothesis is correct or no alternative hypothesis is provided:
  - If a two-sided 95% confidence interval for $\mu_C - \mu_N$ is correctly calculated as $(-13.3, -0.7)$ then component 3 is satisfied.
- If the two-sided confidence interval is correctly interpreted based on whether 0 is in the interval, then component 4 is satisfied.
- If the stated alternative hypothesis is incorrect (1-sided), the confidence interval approach must be consistent with the stated alternative to satisfy components 3 and 4:
  - An interval consistent with the stated alternative will satisfy component 3. A lower one-sided 95% confidence interval for $\mu_C - \mu_N$ is $(-\infty, -1.727)$ or an upper one-sided 95% confidence interval for $\mu_C - \mu_N$ is $(-12.274, \infty)$.
  - If only the upper end of the lower one-sided confidence interval is used to reach a conclusion or only the lower end of the upper one-sided confidence interval is used to reach a conclusion, then component 4 is satisfied.
- A response that provides an interval for $\mu_N - \mu_C$ should be scored the same way as the interval for $\mu_C - \mu_N$.

	Model Solution	Scoring
Section 3	Because the *p*-value of 0.0298 is less than the significance level of 0.05, the null hypothesis should be rejected. There is convincing statistical evidence that the population mean number of oranges on trees when fertilized with Brand C is different from the population mean number of oranges on trees when fertilized with Brand N for all trees similar to those in the study.	**Essentially correct (E)** if the response satisfies the following two components: 1. The response provides a correct comparison of the *p*-value to alpha (*p*-value is less than/greater than alpha) AND provides a correct decision about the null and/or alternative hypothesis. 2. The response states a conclusion in context, consistent with and in terms of the stated alternative hypothesis using nondefinitive language. **Partially correct (P)** if the response satisfies only one of the two components required for E. **Incorrect (I)** if the response does not meet the criteria for E or P.

Additional Notes:

- To satisfy component 1, the response must clearly identify the number that is compared to alpha as a *p*-value (which can be identified anywhere in the response).
- To satisfy the *p*-value comparison in component 1, the response can compare the value of the test statistic to an appropriate critical value (e.g., $|t| > 1.98$).
- An explicit decision about the null hypothesis is not required to satisfy component 1.
- If an explicit decision is stated and the conclusion is inconsistent with the decision, component 1 is not satisfied.
- The decision part of component 1 may be satisfied by implying the decision within the conclusion statement (sufficient evidence/insufficient evidence for the alternative hypothesis).
- If the comparison and decision are consistent with an incorrect *p*-value (or an incorrect value of the test statistic or an incorrect confidence interval), the response may satisfy component 1.
- To satisfy the context in component 2, the response must include reference to means, groups (Brand C fertilizer and Brand N fertilizer), the sampling units (trees), and the variable of interest (number of oranges on the trees).
- A response that clearly refers to samples in the conclusion cannot satisfy component 3 in section 1 but may satisfy components 1 and 2 in section 3.
- If the response omits hypotheses, assume the correct alternative hypothesis is provided when scoring component 2.
- If the response states incorrect hypotheses, component 2 may be satisfied by either stating a conclusion in terms of the stated alternative hypothesis or by answering the inference question.
- Examples of nondefinitive language in component 2 include “evidence to accept the alternative,” “there is evidence for the alternative,” or “there is not sufficient evidence for the alternative.”

- Examples of definitive language in component 2 include "accepts the null," "proves the null," "proves the alternative," "accepts the alternative," "there is not evidence for the alternative," or "no evidence for the alternative."
- If components 1 and/or 2 are satisfied and the response also provides an incorrect interpretation of the *p*-value, the score is lowered from E to P or P to I.

Confidence Interval Approach

- Component 1 is satisfied if the justification is based on whether 0 is included in the confidence interval.
- If only the upper end of the lower one-sided confidence interval is used to reach a conclusion or only the lower end of the upper one-sided confidence interval is used to reach a conclusion, then component 1 is satisfied. A lower one-sided 95% confidence interval for $\mu_C - \mu_N$ is $(-\infty, -1.727)$ or an upper one-sided 95% confidence interval for $\mu_C - \mu_N$ is $(-12.274, \infty)$. These values are consistent with a pooled standard deviation as well.
- If no alternative hypothesis is specified in the response, then assume the correct alternative hypothesis is provided when scoring component 2.
- If the response includes an incorrect interpretation of the confidence interval, then the score for section 3 is lowered from E to P or from P to I.

Matched Pairs Approach

- Components should be scored according to the scoring guidelines using the reported *p*-value or by comparing the *t*-statistic to an appropriate critical value or comparing 0 to the reported confidence interval.

Rejection Region Approach

- A response that uses a comparison of the value of the test statistic to a critical value in lieu of reporting and comparing the *p*-value to a significance level may satisfy component 1. The one-sided test critical values for a significance level of 0.05 for commonly used degrees of freedom (*df*) are shown in the following table. The critical values for a two-sided test are also shown.

df	Critical Value (one-sided test)	Critical Value (two-sided test)
114	1.658	1.981
108.174	1.659	1.982
57	1.672	2.002

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

Question 5: Multi-Focus — 4 points

General Scoring Notes

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

	Model Solution	Scoring
A	i. $P(\text{Football}) = \frac{2,461}{4,193} \approx 0.587$ ii. $P(25 \text{ to} < 30 \mid \text{Football}) = \frac{1,326}{2,461}$ $P(25 \text{ to} < 30 \mid \text{Football}) \approx 0.539$	**Essentially correct (E)** if the response satisfies at least three of the following four components: 1. In part A (i) the response indicates the correct probability. 2. In part A (i) the response provides supporting work for the correct probability calculated in component 1. 3. In part A (ii) the response indicates the correct conditional probability. 4. In part A (ii) the response provides supporting work for the correct conditional probability calculated in component 3. **Partially correct (P)** if the response satisfies only two of the four components required for E. **Incorrect (I)** if the response does not meet the criteria for E or P.

Additional Notes:

- An arithmetic or transcription error in a response can be disregarded if correct work is shown.
- If the probabilities are presented in a different order than asked for by the question, then they must have appropriate labels. In this case, sufficient labels include probability notation, context, or identification of the subparts (i and ii).
- Decimal or percentage answers may satisfy components 1 and 3.
- A response of $\frac{2461}{4193}$ satisfies components 1 and 2.
- A response of $\frac{1326}{2461}$, $\frac{\left(\frac{1326}{4193}\right)}{\left(\frac{2461}{4193}\right)}$, or $\frac{\left(\frac{1326}{4193}\right)}{0.587}$ may satisfy components 3 and 4.

	Model Solution	Scoring
B	i. *b* represents the probability of all professional athletes that play football, so it corresponds to the calculation from part A (i). ii. *x* represents the joint probability of a professional athlete being a football player and in the "$25 \le \text{Age} < 30$" age-group.	**Essentially correct (E)** if the response satisfies the following two components: 1. In part B (i) the response correctly identifies *b* as the value corresponding to part A (i). 2. In part B (ii) the response correctly identifies *x* in context. **Partially correct (P)** if the response satisfies only one of the two components required for E. **Incorrect (I)** if the response does not meet the criteria for E or P.

Additional Notes:

- A response may satisfy component 1 by stating the proportion represents all professional athletes who play football.
- A response may satisfy component 1 by providing the value given in part A (i).
- In part B (i) if a response correctly provides an answer of A (i) but uses words that imply a conditional probability, the response has parallel solutions and does not satisfy component 1.
- In part B (ii) a response that indicates that ***x*** represents the probability of the intersection of playing football and being in the age-group that is between 25 and 30 may satisfy component 2.
- A response that clearly indicates the entire group of athletes is used (i.e., 4193), can satisfy component 2 without the word "and" (e.g., "randomly selected athlete is a football player in the age range $25 \le \text{Age} < 30$").
- The following interpretations do not satisfy component 2 because they describe conditional probabilities or conditional frequencies:
  - Randomly selected football player in the age-range $25 \le \text{Age} < 30$
  - Probability of an age-range $25 \le \text{Age} < 30$ who plays football
  - All athletes aged $25 \le \text{Age} < 30$ who play football
  - Footballers aged 25 to 30
  - Football players from the 25 to 30 group
  - Probability that a randomly selected athlete plays football given they are in the 25 to 30 age-group
- In part B (ii) if the response clearly indicates that *x* is a count and not a probability, component 2 is not satisfied.

Model Solution	Scoring
C i. No, the events "Baseball" and "35 ≤ Age" are not mutually exclusive because $P(\text{Baseball} \cap 35 \le \text{Age}) = \frac{61}{4{,}193}$ $= 0.0145,$ which is not 0. ii. No, the events are not independent because $P(35 \le \text{Age} \mid \text{Baseball}) \neq P(35 \le \text{Age})$ $\frac{61}{1{,}216} \neq \frac{121}{4{,}193}$ $0.0501 \neq 0.0289$.	**Essentially correct (E)** if the response satisfies component 4 and at least two of the other three components: 1. In part C (i) the response draws a conclusion that the events are not mutually exclusive. 2. In part C (i) the response provides the correct justification based on the joint probability or joint frequency. 3. In part C (ii) the response draws a conclusion that the events are not independent. 4. In part C (ii) the response supports the conclusion that events are not independent with a reasonable statistical rationale that includes an appropriate comparison of probabilities. **Partially correct (P)** if the response does not satisfy the criteria for E and satisfies at least two of the four components. **Incorrect (I)** if the response does not meet the criteria for E or P.

Additional Notes:

- A response that references 61 baseball players who are $35 \le \text{Age}$ may satisfy component 2.
- A response that includes any appropriate comparison may satisfy component 4. For example:
  - $P(\text{Baseball} \mid 35 \le \text{Age}) = \frac{61}{121} \approx 0.504$ is not equal to $P(\text{Baseball}) = \frac{1{,}216}{4{,}193} \approx 0.290.$
  - $P(35 \le \text{Age} \mid \text{Not baseball}) = \frac{(19 + 41)}{(516 + 2{,}461)} \approx 0.0202$ is not equal to $P(35 \le \text{Age}) = \frac{121}{4{,}193} \approx 0.029.$
  - $P(\text{Baseball}) \times P(35 \le \text{Age}) = \left(\frac{1{,}216}{4{,}193}\right) \times \left(\frac{121}{4{,}193}\right) \approx 0.008$ is not equal to $P(\text{Baseball} \cap 35 \le \text{Age}) = \frac{61}{4{,}193} \approx 0.015.$
  - A response may satisfy component 4, if it compares the observed counts of 61 with the expected count of 35.09 with either work shown for the expected count or labels for both the observed and expected counts.
- Notation is not required to satisfy component 4 if appropriate fractions or probabilities are shown.
- The number of components satisfied should be considered if holistic scoring is required.

	Model Solution	Scoring
D	No, because we have data for the entire population of players for these three professional sports leagues. No random sampling took place. Inference is not needed when you have data for the entire population.	**Essentially correct (E)** if the response satisfies the following two components: 1. The response indicates that it is not appropriate to conduct a chi-square test for independence. 2. The response provides a valid reason why a chi-square test for independence is not appropriate. **Partially correct (P)** if the response satisfies component 1 but gives a weak explanation about why it is inappropriate to use a chi-square test for independence. *OR* If the response does not explicitly indicate that a chi-square test for independence is not appropriate, but the reasoning for not using a chi-square test for independence is appropriate. **Incorrect (I)** if the response does not meet the criteria for E or P.

Additional Notes:

- Appropriate justification for component 2 may be:
  - The data does not represent a random sample.
  - The data represents a census.
  - The sample is not less than 10% of the population.
- Weak justification for component 2 may be:
  - An incorrect attempt to check a valid condition
  - Referencing a different chi-square test
  - Referencing that multiple samples were taken
  - Referencing that in part C (ii), it was shown that the variables are not independent
  - Indicating that conditions are not met without specifying which condition
- A response that states “proceed with caution” and satisfies component 2 may be scored a P.

Scoring for Question 5

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 6: Investigative Task — 4 points

General Scoring Notes

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

Model Solution	Scoring
A i. There is a moderately strong, positive, linear relationship between the number of hits and number of runs. ii. $\hat{y} = -372.2 + 0.823(1,250) = 656.55$, where $\hat{y}$ represents the predicted number of runs.	**Essentially correct (E)** if the response provides a description that includes at least two of the first three components as well as components 4 and 5: 1. In part A (i) the response states the correct direction (positive). 2. In part A (i) the response states the correct strength (strong or moderately strong). 3. In part A (i) the response states the correct form of association (linear). 4. In part A (i) the response includes context (number of hits and number of runs). 5. In part A (ii) the response indicates the correct prediction for number of runs with supporting work. **Partially correct (P)** if the response does not satisfy the criteria for E and satisfies three or more of the five components. **Incorrect (I)** if the response does not satisfy the criteria for E or P.

Additional Notes:

- An arithmetic or transcription error in a response can be ignored if correct work is shown.
- A response that indicates "as the number of hits increases, the number of runs tends to increase" may satisfy components 1 and 4.
- A response that provides 656 or 657 with supporting work for part A (ii) may satisfy component 5.

	Model Solution	Scoring
B	i. The team represented by point A is classified as having a salary that is less than the median but has a higher number of hits and higher number of runs compared to other teams with the same salary classification. ii. The strength of the linear relationship seems to be stronger for teams with total salaries greater than the median than for teams with total salaries less than the median. The points for individual teams with a total salary greater than the median have a tighter cluster around a linear pattern than the wider cluster of the points for individual teams with a total salary less than the median.	**Essentially correct (E)** if the response satisfies the following three components: 1. In part B (i) the response describes point A as having a higher number of hits and a higher number of runs when compared to teams with the same classification. 2. In part B (ii) the response indicates the relationship is stronger for teams with total salaries greater than the median than for teams with total salaries less than the median. 3. In part B (ii) the response gives a reasonable explanation for the answer in component 2. **Partially correct (P)** if the response satisfies only two of the three components required for E. **Incorrect (I)** if the response does not meet the criteria for E or P.

Additional Notes:

- Examples of responses that may satisfy component 3 include:
  - Teams with a total salary greater than the median have points that more tightly cluster about the line than those with a total salary less than the median.
  - Teams with a total salary greater than the median follow a clearer linear pattern than those with a total salary less than the median.
  - Teams with a total salary less than the median have points that are more widely scattered from the linear pattern than teams with a total salary greater than the median.
  - Teams with a total salary less than the median have points that are more of a cloud shape than teams with a total salary greater than the median.
- A response that says the strength is about the same for both groups does not satisfy component 2 but may satisfy component 3.

Model Solution	Scoring
C i. $t^* = 2.05$, $df = 30 - 2 = 28$ ii. The 95% confidence interval for the mean number of runs for all teams with 1,250 hits using the critical value from part C (i) and the given standard error of 17.48 is $656.55 \pm 2.05(17.48) = (620.72, 692.38)$. iii. The 95% prediction interval for the number of runs for a single team with 1,250 hits using the critical value from part C (i) and the given standard error of 56.78 is $656.55 \pm 2.05(56.78) = (540.15, 772.95)$.	**Essentially correct (E)** if the response satisfies at least four of the following five components: 1. In part C (i) the response indicates the correct critical value to at least two decimal places. 2. In part C (ii) the response correctly calculates the lower and upper bounds of the 95% confidence interval of the mean response, consistent with C (i). 3. In part C (ii) the response shows correct work for the upper and lower bounds calculated in component 2. 4. In part C (iii) the response correctly calculates the lower and upper bounds of the 95% prediction interval for a team with 1,250 hits, consistent with C (i). 5. In part C (iii) the response shows correct work for the upper and lower bounds calculated in component 4. **Partially correct (P)** if the response satisfies two or three of the five components required for E. **Incorrect (I)** if the response does not meet the criteria for E or P.

Additional Notes:

- An arithmetic or transcription error in a response can be ignored if correct work is shown.
- If a response uses an incorrect point estimate found in part A (ii) but uses a correct critical value and standard error and correct work is shown in at least one of parts C (ii) and C (iii), then the response may earn a score of E.
- A response that does not give the critical value to two decimal places but instead uses the entire critical value from the calculator will find an interval of $(620.74, 692.36)$ for part C (ii) and an interval of $(540.24, 772.86)$ in part C (iii) and may earn a score of E.
- A response that uses an incorrect critical value may still earn a score of E. Two common incorrect critical values and their resulting intervals are $t^* = 1.70$ (626.83, 686.27) and (560.02, 753.08) and $t^* = 1.96$ (622,29, 690.81) and (545.26, 767.84).

	Model Solution	Scoring
D	i. A distribution of sample means should have less variability than a distribution of individual observations. We would expect individuals to fall further from the population mean than the average of a sample of individuals. When taking the mean of a sample of individuals the individual extremes are averaged, causing less variability. ii. The prediction interval is wider than the confidence interval because the standard error for a distribution of individuals would be higher than the standard error for a distribution of means. Based on the standard error formulas, the variability for the prediction interval is higher than that for the confidence interval because the prediction interval formula includes an additional s^2 under the square root. The critical value is the same; therefore, the margin of error will be larger for prediction intervals. This makes sense because the distribution of means does not vary as much as the distribution of individuals.	**Essentially correct (E)** if the response satisfies the following three components: 1. In part D (i) the response indicates that there would be less variability for a distribution of means than for a distribution of individual observations. 2. In part D (i) the response gives a reasonable explanation for the response in component 1. 3. In part D (ii) the response indicates that the prediction interval was wider than the confidence interval because the standard error is higher for a distribution of individuals than the standard error for a distribution of means. **Partially correct (P)** if the response satisfies only two of the three components required for E *OR* the response satisfies only component 3. **Incorrect (I)** if the response does not meet the criteria for E or P.

Additional Notes:

- For component 2, responses need to focus on how means and individuals vary and not the values of the standard errors provided in part C. The following are examples of reasonable responses:
  - The distribution of sample means has less variability because averaging several values reduces the effect of extreme observations.
  - Extreme values tend to balance each other out when calculating a mean, so the variability of means is smaller.
  - Because individual high and low values are averaged together, the sample means vary less than individual data points.
  - Individual observations can vary widely, but averages of groups tend to stay closer to the population mean.
  - A single observation can be far from the population mean, but a sample mean is less likely to be that far away from the population mean.
  - The standard deviation for sample means is divided by the square root of n, resulting in less variability than individual observations.

- For component 3, a response needs to explain why the prediction interval is wider, based on the answer in part D (i) or the given formulas. The following are examples of responses that satisfy component 3:
  - The standard error is higher for the prediction interval because there is an additional s^2 under the square root.
  - There will be more variability for the prediction of individual values than for the prediction of means.

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.	