

Question 5: Multi-Focus

4 points

General Scoring Notes

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

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

Scoring Notes:

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

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

Scoring Notes:

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

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

Scoring Notes:

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

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

Question 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) The inference procedure that should be used to estimate the mean price, in dollars (\$), of this type of whistle at all stores that sell the whistle is a one-sample t-interval for a population mean. (ii) The parameter of interest is the mean whistle price, in dollars (\$), of this type of whistle at all stores that sell the whistle.	Essentially correct (E) if the response satisfies the following three components: - In part (a-i) the response identifies a one-sample t-interval for a population mean by name (e.g., “one sample t-interval”) - In part (a-ii) the response identifies the correct parameter by including reference to the mean - In part (a-ii) the response provides sufficient context for the parameter by including reference to the population (all stores that sell the whistle) and the variable of interest (the price of this type of whistle) Partially correct (P) if the response only satisfies one or two of the three components required for E. Incorrect (I) if the response does not meet the criteria for E or P.

Additional Notes:

- In part (a-i) a response that refers to a “test” does not satisfy component 1.
- In part (a-i) a response that refers to “a mean,” singular, satisfies the one-sample aspect of component 1. If the response states “one-sample,” it is not required to also include “for a mean” to satisfy component 1.
- A response that addresses “mean” in part (a-i) does not satisfy component 2 unless it is also addressed in part (a-ii).
- In part (a-ii) a response may satisfy the variable of interest aspect of component 3 with a minimum of “price” and “whistle.”
- In part (a-ii) a response may satisfy the population aspect of component 3 by using words such as: “population,” “all,” or “true,” or by using the symbol μ .
- In part (a-ii) if the response clearly refers to the sample mean instead of the population mean using words or symbols (e.g., $\bar{x}$), then component 3 is not satisfied unless the symbol used is defined as the population mean.
- For responses that are scored P, the number of components satisfied should be considered if holistic scoring is required.

Model Solution	Scoring
(b) (i) The distribution of the sample of whistle prices appears slightly skewed to the right, because the mean is slightly higher than the median. (ii) Based on the $1.5 \times \text{IQR}$ rule, there are not whistle prices in this sample that would be considered outliers. A whistle price is an outlier using this method if it is more than $1.5 \times \text{IQR}$ below the first quartile (Q_1) or more than $1.5 \times \text{IQR}$ above the third quartile (Q_3). Because $Q_1 - 1.5 \times \text{IQR} = 4.51 - 1.5(5.475 - 4.51) = 3.0625,$ and the minimum value (4.25) is greater than 3.0625, there are no outliers to the left. Because $Q_3 + 1.5 \times \text{IQR} = 5.475 + 1.5(5.475 - 4.51) = 6.9225,$ and the maximum value (6.58) is less than 6.9225, there are no outliers to the right.	Essentially correct (E) if the response satisfies the following four components: - In part (b-i) the response indicates the distribution is skewed to the right or approximately symmetric - In part (b-i) the response justifies the response to component 1 with appropriate reasoning based on the summary statistics - In part (b-ii) the response provides a justification for stating there are no outliers by correctly calculating the lower and upper outlier criteria with work shown - In part (b-ii) the response states whether there are outliers based on the calculated lower and upper outlier criteria Partially correct (P) if the response satisfies three of the four components required for E. <i>OR</i> satisfies components 1 and 2 <i>OR</i> components 3 and 4 of components 1–4 required for E. Incorrect (I) if the response does not meet the criteria for E or P.

Additional Notes:

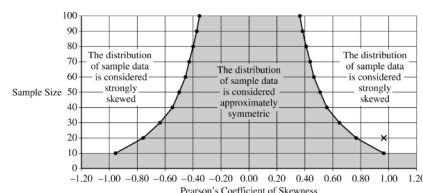
- If the response states skewed right, to satisfy component 2 the response should refer to the median being less than the mean. If the response states approximately symmetric, the response should refer to the median and mean being close together.
- A response that indicates skewed left does not satisfy components 1 or 2.
- If the response says the distribution is skewed to the right, examples of responses that satisfy component 2 include:
 - the mean is higher than the median,
 - the ratio between the mean and median is greater than 1,
 - the difference between the maximum and median is greater than the difference between the median and minimum, and
 - the difference between the maximum and third quartile is greater than the difference between the first quartile and minimum.
- For responses that score P, the number of components satisfied should be considered if holistic scoring is required.

Model Solution

Scoring

- (c) (i) Pearson's coefficient of skewness is
- $$\frac{3(5.12 - 4.885)}{0.743} \approx 0.949.$$

(ii)



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

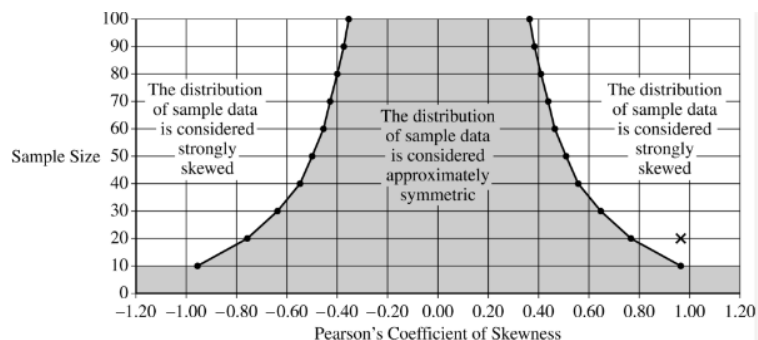
1. In part (c-i) the response calculates Pearson's coefficient of skewness with work shown
2. In part (c-ii) the response includes an x at approximately (0.949, 20) or an x consistent with the calculation in part (c-i)

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:

- An arithmetic or transcription error in a response can be ignored if correct work is shown.
- Responses must include an x that is located on the line for sample size of 20 and located between 0.90 and 1.00 on the horizontal axis or be consistent with the value given in part (c-i) to satisfy component 2.



Model Solution

Scoring

- (d) (i) Looking at the graph in part (c), for a sample size of 20, and a skewness coefficient of 0.949, this point falls in "the distribution of sample data is considered strongly skewed" region. Therefore, we would consider the shape of the distribution of the sample of whistle prices to be strongly skewed.
- (ii) No, based on the response to part (d-i), Julio's data would not satisfy the normality condition because neither of the criteria listed are met. Julio only has a sample size of 20, which is less than 30, and Pearson's coefficient of skewness indicates the distribution of sample data is strongly skewed.

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

1. In part (d-i) the response indicates the sample is strongly skewed or makes a statement consistent with the graph in part (c)
2. In part (d-i) the response uses or refers to the graph or the Pearson coefficient of skewness to make the correct determination of skewness
3. In part (d-ii) the response states that the condition of normality is not met or makes a statement consistent with the explanation in part (d-i)
4. In part (d-ii) the response explains the sample size of 20 is less than 30
5. In part (d-ii) the response explains that the distribution of sample data is strongly skewed or makes a statement consistent with the explanation in part (d-i)

Partially correct (P) if the response does not meet the criteria for E but satisfies three or four of components 1–5 required for E.

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

Additional Notes:

- A response that determines the sample is strongly skewed, but says the normality condition is satisfied, should be scored I.
- In part (d-i) a response that indicates the distribution is strongly skewed to the left does not satisfy component 1.
- If a response in part (c-ii) has an x on the curve, the response in part (d-i) must indicate the sample is skewed (not strongly skewed or approximately symmetric), or indicate it is not possible to choose whether it is strongly skewed or approximately symmetric to satisfy component 1.
- In part (d-i) a response that indicates the distribution is skewed, but not strongly skewed, because the Pearson coefficient of skewness is not that far from the curve on the graph satisfies component 1 and component 2.
- In part (d-i) a response that indicates the distribution is skewed, but not strongly skewed, may satisfy components 3–5 if in part (d-ii) the response indicates the conditions are met because the distribution is not very skewed and there are no outliers. The response does not need to indicate the sample size is less than 30.
- In part (d-ii) a response that indicates the sample size is small without explicitly referring to the values 20 and 30 may satisfy component 4.
- The quality of communication, or the number of components satisfied for responses with score P, should be considered if holistic scoring is required.

Scoring for Question 6

Each essentially correct (E) part counts as 1 point, and each partially correct (P) part counts as 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_d represent the true mean difference (placebo minus omega-3) of irritability scores for all people with this medical condition. The null hypothesis is $H_0: \mu_d = 0$ and the alternative hypothesis is $H_a: \mu_d > 0$. The appropriate inference procedure is a matched pairs t-test for a mean difference.	Essentially correct (E) if the response satisfies the following four components: - Identifies a paired t-test for a population mean difference by name or by formula - States the hypotheses using a single mean (e.g., μ_d, μ) - States the correct equality for the null hypothesis (e.g., $H_0: \mu_d = 0$) AND states the correct direction for the alternative hypothesis (e.g., $H_a: \mu_d > 0$) - Provides sufficient context for the parameter, by including reference to the <i>population</i> mean difference AND the sampling units (people with the medical condition) AND the response variable (irritability score) Partially correct (P) if the response satisfies three of the four components. Incorrect (I) if the response does not meet the criteria for E or P.

Additional Notes:

- A response that states the null hypothesis as $H_0: \mu_d \leq 0$ may satisfy component 3.
- A response that states the name of the procedure as t -test, or one-sample t -test may satisfy component 1.
- To satisfy component 2, the hypotheses must be stated in terms of a mean. If a symbol other than μ or $\bar{X}$ is used to denote the mean, it must be clearly defined as a mean (but does not need to reflect the context of irritability score). It is acceptable to use μ_n to denote the mean.

- A response that states the hypotheses in words (e.g., “the null hypothesis is that the mean difference is 0, and the alternative hypothesis is that the mean difference is greater than 0”) may satisfy components 2 and 3. Neither context nor the concept of the *population* is required to satisfy component 2 or 3.
- A response that states the hypotheses in words [e.g., “the null hypothesis is that the mean difference in irritability score (placebo minus omega-3) for all people with this medical condition is equal to 0 and the alternative hypothesis is that the mean is greater than 0”) may satisfy components 2, 3, and 4.
- The elements of component 4 do not have to be satisfied with the statement of the hypotheses. They may be satisfied by work presented anywhere in the response, most likely by the statement of the conclusion.
- If the statement of the hypotheses refers to population mean and the conclusion refers to sample mean (or vice versa), then the population aspect of component 4 is not satisfied.
- If the response clearly refers to the *sample* mean instead of the *population* mean using words or a symbol (e.g., $\bar{x}$), then component 4 is not satisfied unless the symbol used is defined as the *population* mean.
- A response may satisfy the population aspect of component 4 by doing the following:
 - referring to population in the statement of the conclusion of the inferential procedure.
 - using notation such as μ when defining the hypothesis statements.
- A response may satisfy the sampling units aspect of component 4 by referring to “patients” or “people with this medical condition” or a similar statement.
- 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 by formula using a z -percentile instead of a t -percentile, then component 1 is not satisfied.

Confidence Interval Approach:

- If a one-sample t -interval for a population mean is identified correctly by name (e.g., “one-mean t -interval” or “one-sample t -interval” or “ t -interval”) or by formula, then component 1 is satisfied.
- If a response uses a one-sample t -interval for a population mean, then components 2 and 3 are satisfied if the response indicates that it is a confidence interval for the mean difference in irritability score (placebo minus omega-3).

Two-Sample Approach:

- A response that clearly defines the *population* difference in means with clearly defined parameters for the means of placebo and omega-3 AND the sampling units (people with the medical condition or people similar to those in the study) AND the response variable (irritability score) may earn credit for component 4.

	Model Solution	Scoring
Section 2	The independence condition for performing a paired t-test for a mean difference is satisfied because the data were obtained from a randomized experiment where the week in which the patient received the treatment was randomly assigned. The sampling distribution of the mean difference must be approximately normal. Although the sample size is less than 30 ($n = 19$), this is satisfied because the boxplot for the sample differences shows an approximately symmetric distribution with no outliers. The value of the test statistic is: $t = \frac{\bar{x}_d - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{1.789 - 0}{\frac{2.485}{\sqrt{19}}} \approx 3.138$ Using 18 degrees of freedom, the corresponding p-value is $P(t > 3.138) \approx 0.0028.$	Essentially correct (E) if the response satisfies the following four components: 1. Checks the independence condition by referring to the random assignment 2. Indicates that the distribution of irritability scores for the differences (placebo minus omega-3) is not badly skewed supports the assumption that the sampling distribution of $\bar{x}_d$ is approximately normal 3. Correctly reports the value of the t-statistic consistent with the named test 4. Correctly reports the p-value, consistent with the stated alternative hypothesis and reported test statistic Partially correct (P) if the response satisfies only two or three of the four components. Incorrect (I) if the response does not meet the criteria for E or P.

Additional Notes:

- Component 1 is not satisfied if the response indicates that the independence condition is met because the sample was selected at random from all patients with the medical condition.
- Component 1 is satisfied if the response indicates that the treatment was randomly assigned even if the terms “independent” or “independence” are not used.
- Discussion of degrees of freedom will be treated as extraneous in scoring for section 2.
- If the response incorrectly identifies the test as a z -test in section 1, the correct z -statistic of 3.138 satisfies component 3 and a p -value of 0.00085 satisfies component 4. However, if the test statistic is not reported, component 3 is not satisfied, but if the correct p -value (consistent with the alternative hypothesis) is reported then component 4 is satisfied.
- If the response does not identify a z -test in section 1 but reports the test statistic as $z = 3.138$ (instead of t) then component 3 is not satisfied; however, component 4 may still be satisfied if the correct p -value of 0.00085 is reported with a z -statistic of 3.138.
- 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 a two-tailed alternative hypothesis is stated, then the p -value must be consistent with the stated alternative hypothesis to satisfy component 4.
- 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 satisfies component 4, any supporting work for the p -value may be treated as extraneous.

- If no test statistic is reported, and the p -value is consistent with the stated alternative hypothesis or equal to 0.0028 component 4 is satisfied.

Confidence Interval Approach:

- If the stated alternative hypothesis is correct or no alternative hypothesis is provided:
 - If either a one-sided 95 percent confidence interval for μ_d is correctly calculated as $(0.8004, \infty)$ or a two-sided 90 percent confidence interval for μ_d is correctly calculated as $(0.8004, 2.7776)$ then component 3 is satisfied.
 - If only the lower end of the confidence interval for μ_d is used to reach a conclusion, then component 4 is satisfied.
- If the stated alternative hypothesis is incorrect (two-sided or reversed direction), the confidence interval approach must be consistent with the stated alternative to satisfy components 3 and 4:
 - An interval consistent with the stated two-sided alternative will satisfy component 3. A two-sided 95 percent confidence interval for μ_d is $(0.591, 2.987)$. If the two-sided confidence interval is correctly interpreted based on whether zero is in the interval, then component 4 is satisfied.
 - An interval consistent with the incorrect reversed alternative will satisfy component 3. If either a two-sided 90 percent confidence interval for μ_d is correctly calculated as $(0.8004, 2.7776)$, or a one-sided 95 percent confidence interval for μ_d is correctly calculated as $(-\infty, 2.7776)$, then component 3 is satisfied. If the two-sided confidence interval is correctly interpreted based on whether zero is in the interval, then component 4 is satisfied; if only the upper end of the lower one-sided confidence interval is used to reach a conclusion, then component 4 is satisfied.
- If the difference calculated is μ_d , omega-3 minus placebo, the intervals will be the negative of those provided above.

Two-Sample Approach:

- A response that indicates that the distribution of irritability scores for each of the two samples, placebo AND omega-3, are not badly skewed may satisfy component 2.
- If the response correctly reports the value of the t -statistic as 2.256, then the response may satisfy component 3.
- Values for the p -value depend on how the degrees of freedom were determined. The following p -values all assume the correct alternative hypothesis and satisfy component 4:
 - The degrees of freedom for $t \approx 2.256$ on a two-sample solution computed with the Satterthwaite approximation used by many calculators is:

$$df = \frac{\left(\frac{s_{\text{placebo}}^2}{n_{\text{placebo}}} + \frac{s_{\text{omega-3}}^2}{n_{\text{omega-3}}} \right)^2}{\frac{1}{n_{\text{placebo}} - 1} \left(\frac{s_{\text{placebo}}^2}{n_{\text{placebo}}} \right)^2 + \frac{1}{n_{\text{omega-3}} - 1} \left(\frac{s_{\text{omega-3}}^2}{n_{\text{omega-3}}} \right)^2} = \frac{\left(\frac{(2.987)^2}{19} + \frac{(1.739)^2}{19} \right)^2}{\frac{1}{19-1} \left(\frac{(2.987)^2}{19} \right)^2 + \frac{1}{19-1} \left(\frac{(1.739)^2}{19} \right)^2}$$

$$\approx 28.94,$$

resulting in a p -value of 0.0159 or a p -value of 0.0318 if a two-sided alternative hypothesis was stated.

- Using $t \approx 2.256$ and $(19 + 19 - 2) = 36$ degrees of freedom, the resulting p -value is 0.0151 or a p -value of 0.0302 if a two-sided alternative hypothesis was stated.

- Using $t \approx 2.256$ and $(19 - 1) = 18$ degrees of freedom, the resulting p -value is 0.0184 or a p -value of 0.0368 if a two-sided alternative hypothesis was stated.
- If no test statistic is reported, and the p -value is consistent with the stated alternative hypothesis or equal to 0.0159 or 0.0151 or 0.0184, component 4 is satisfied.

	Model Solution	Scoring
Section 3	Because the p -value ≈ 0.0028 is less than the significance level, $\alpha = 0.05$, the null hypothesis should be rejected. The data provide convincing statistical evidence that for patients similar to those in the study, the true mean difference (placebo minus omega-3) in irritability scores for people with this medical condition is greater than zero. This suggests the omega-3 fatty acids are helpful in reducing irritability scores in people with this medical condition.	Essentially correct (E) if the response satisfies the following two components: - Provides correct comparison of the p-value to alpha (p-value is less than/greater than alpha) <i>AND</i> provides a correct decision about the null and/or alternative hypothesis - States a conclusion in context, consistent with, and in terms of, the alternative hypothesis using non-deterministic language 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:

- In order to satisfy component 1, the response must clearly identify the number that is compared to alpha as a p -value.
- If the response provides an unreasonable p -value (that is clearly identified as the p -value) and correctly compares it to alpha, component 1 may be satisfied.
- 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., $t_\alpha = 1.734$, and $3.13 > 1.734$.
- An explicit decision 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).
- To satisfy the context in component 2, the response must include the population parameter, sampling/experimental units, and the response variable.
- 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 non-deterministic language in component 2 include “evidence to accept the alternative,” “there is evidence for the alternative,” “there is not sufficient evidence for the alternative.”
- Examples of deterministic language in component 2 include “proves the null,” “proves the alternative,” “accepts the alternative,” “there is not evidence for the alternative,” and “no evidence for the alternative.”
- 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.
- If components 1 and/or 2 are satisfied and the response provides an incorrect interpretation of the p -value, the score is lowered from E to P or P to I.
- In section 1, if component 4 is not satisfied only because of the omission of sampling units (patients), in section 3 component 2 may be satisfied by an appropriate conclusion that is only missing sampling units.

Confidence Interval Approach:

- Component 2 should be scored according to the rubric and component 1 should be scored with regard to a comparison of zero to the appropriate end of the reported confidence interval.
- If no alternative hypothesis is specified in the response, then assume the correct alternative hypothesis is provided when scoring component 2.
- If an incorrect two-sided alternative hypothesis is specified, then component 2 is satisfied if the justification is based on whether zero is included in the confidence interval.
- 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.

Two-Sample Approach:

- Components should be scored according to the rubric using the reported p -value, or by comparing the t -statistic to an appropriate critical value.

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 <i>OR</i> One section essentially correct and one or two sections partially correct <i>OR</i> Three sections partially correct	2
Minimal Response One section essentially correct and no section partially correct <i>OR</i> No section essentially correct and two sections partially correct	1