

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	An appropriate inference procedure is a one-sample z-test for a population proportion. The null hypothesis is $H_0: p = 0.22$, and the alternative hypothesis is $H_a: p > 0.22$, where p = the true proportion of students at Karen's high school that use the application at least once per week.	Essentially correct (E) if the response satisfies the following four components: - Identifies a one-sample z-test for a population proportion by name (e.g., "one-proportion z-test" but not merely "one-sample z-test") or by formula - States the correct equality for the null hypothesis with the value 0.22 - States the correct direction for the one-sided alternative hypothesis consistent with the null hypothesis - Provides sufficient context for the parameter by including reference to the population proportion (true proportion of students at Karen's high school) AND the sampling units (students) AND the response variable (using the application) Partially correct (P) if the response satisfies three of the four components required for E. Incorrect (I) if the response does not meet the criteria for E or P.

Scoring Notes:

- If the response identifies the correct test by name but also states an unreasonable formula, then component 1 is not satisfied.
- If the response identifies the test using the correct formula but equating it with a t instead of a z , then component 1 is not satisfied.
- A response that states the null hypothesis as $H_0: p \leq 0.22$ satisfies component 2.

- Components 2 and 3 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 2 or 3.
- A response that states the hypotheses in words (e.g., "the null hypothesis is that the proportion is 0.22, and the alternative hypothesis is that the proportion is greater than 0.22") may satisfy components 2 and 3.
- A response that states the hypotheses in words and refers to the population in context (e.g., "the null hypothesis is that the population proportion of students at Karen's high school that use the application at least once per week is equal to 0.22 and the alternative hypothesis is that the population proportion is greater than 0.22") may satisfy components 2, 3, and 4.
- If the null hypothesis is incorrect, the response can satisfy component 3 with a correct alternative hypothesis.
- The elements of component 4 do not have to be satisfied with the statement of the hypotheses. They may be satisfied with the statement in the hypotheses, definition of the parameter, or the statement of the conclusion.
- If the statement of the hypotheses refers to population proportion and the conclusion refers to sample proportion (or vice versa), then the population aspect of component 4 is not satisfied.
- If the response clearly refers to the **sample** proportion instead of the **population** proportion using a $\hat{p}$, then component 4 is not satisfied unless the symbol used is defined as the **population** proportion.
- A response may satisfy the population aspect of component 4 by the following:
 - Referring to the population by using words such as "population," "all," or "true" when defining the parameter or in the statement of the conclusion of the inferential procedure.
 - Using notation such as p , p_0 , or π when defining the hypothesis statements.

Confidence Interval Approach:

- If a one-sample z -interval for a population proportion is identified correctly by name (e.g., "one-proportion z -interval" but not merely "one-sample z -interval") or by formula, then component 1 is satisfied.
- If a response uses a one-sample z -interval for a population proportion, then component 4 is satisfied if the response indicates that it is a confidence interval for the true proportion of students at Karen's high school that use the application at least once per week.

	Model Solution	Scoring
Section 2	The independent observation condition for performing the one-sample z-test for a population proportion is satisfied. This is because the data were obtained from a random sample of 130 high school students from Karen's high school. Also, the sample of 130 students is less than 10% of the total number of students at this large high school, because $130 < 0.10(2,000)$ and the total number of students in Karen's high school is greater than 2,000. The 10% condition is required as sampling was conducted without replacement from a finite population. The number of expected successes and expected failures were both more than 10 because $130(0.22) = 28.6$ and $130(0.78) = 101.4$. Thus, the sample size is large enough to support the assumption that the sampling distribution of $\hat{p}$ is approximately normal. $\hat{p} = \frac{38}{130} \approx 0.2923$ Test statistic: $z = \frac{0.2923 - 0.22}{\sqrt{\frac{0.22(1 - 0.22)}{130}}} \approx 1.99$ $P(z > 1.99) \approx 0.023$	Essentially correct (E) if the response satisfies the following four components: - Checks the independence condition by referring to the random sample of 130 students AND indicates that 130 is less than or equal to 10% of 2,000 (i.e., $(130) \leq (0.10)N$) - Checks that the sample size is large enough to support the assumption that the sampling distribution of $\hat{p}$ is approximately normal by verifying that the number of expected successes and expected failures were at least 10 by calculating the following values: $np_0 = 130(0.22) = 28.6$ and $n(1 - p_0) = 130(0.78) = 101.4$ - Correctly reports the value of the z-statistic - Reports the value for the correct p-value consistent with the test statistic and the procedure stated in Section 1 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.

Scoring Notes:

- To satisfy the reference to the random selection of 130 students in component 1, it is minimally acceptable to state "random sample – check" or "SRS – check." However, component 1 is not satisfied if the response implies that random **assignment** was used or only states "random – check."
- To satisfy component 2, a direct comparison must be made against a standard criterion (5 or 10) using either actual values of the observed successes and failures (38 and 92) OR values for the expected successes and failures (28.6 and 101.4) OR formulas for the expected number of successes and failures with values inserted (or defined elsewhere), such as $130(0.22)$ and $130(1 - 0.22)$.
- If the response includes an inappropriate check of conditions, such as $n > 30$, then component 2 is not satisfied.
- A response that reports the correct value for the z -statistic but contains errors in supporting work satisfies component 3.
- A response that inputs correct values into the z -statistic formula but computes an incorrect value for the z -statistic, satisfies component 3.

- If the response incorrectly uses the sample proportion in calculating the standard error, the response does not satisfy component 3 but may satisfy component 4 ($z = 1.81$ and $p = 0.0351$).
 - The following combinations of z -statistics and p -values may satisfy component 4 but not component 3: $z = 1.76$ and $p = 0.039$ OR $z = 1.915$ and $p = 0.0276$.
- If the response satisfies component 4, any supporting work for the p -value may be treated as extraneous.
- To satisfy component 4, the p -value must be consistent with the alternative hypothesis and either the reported test statistic OR the correct test statistic (1.99).
- If the response compares the value of the test statistic to a critical value instead of reporting a p -value, then the critical value (1.645), or a critical value consistent with the stated alternative hypothesis, satisfies component 4.
- If the response omits identifying the hypotheses, the correct one-sided alternative hypothesis is assumed when scoring component 4.
- If the response omits a test statistic, the correct test statistic 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.

Confidence Interval Approach:

- If either a one-sided 95% confidence interval is correctly calculated as $(0.227, \infty)$ or a two-sided 90% confidence interval is correctly calculated as $(0.227, 0.358)$, then component 3 is satisfied.
- If the alternative hypothesis corresponds to $H_a: p > 0.22$ and only the lower end of a confidence interval is used to reach a conclusion, then component 4 is satisfied.
- Application of a confidence interval approach must be consistent with the stated alternative to satisfy component 4.
- A two-sided 95% confidence interval is $(0.214, 0.370)$ and does not satisfy component 3.

	Model Solution	Scoring
Section 3	Because this p -value is less than the $\alpha = 0.05$ significance level, the null hypothesis should be rejected. There is convincing statistical evidence that the population proportion of students at Karen's high school that use the application at least once per week is greater than Country W's proportion of 0.22.	Essentially correct (E) if the response satisfies the following two components: 1. Provides 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. States a conclusion in context, consistent with, and in terms of, the 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.

Scoring 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).
- 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. This also applies to incorrect p -values that are unreasonable (e.g., p -value = 3.2).
- 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., $z > 1.645$).
- If the response includes any statement equivalent to "accept the null hypothesis," component 1 is not satisfied.
- An explicit decision (Fail to Reject/Reject) is not required to satisfy component 1. The decision part of component 1 may be satisfied by implying the decision within the conclusion statement (insufficient evidence/sufficient evidence for the alternative hypothesis).
- If an explicit decision is stated and the conclusion is inconsistent with the decision, component 1 is not satisfied.
- To satisfy the context in component 2, the response must reference proportion, students (may be implied), and using the application OR answers the inference question.
- 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 incorrect hypothesis or by stating the conclusion by answering the inference question given in the stem (e.g., "the results from this study do not provide convincing statistical evidence that the proportion of students at Karen's high school that use the application at least once per week is greater than 0.22").
- Definitive language, such as the following, does not satisfy component 2: "proves the null," "proves the alternative," "accepts the alternative," and "there is no evidence for the alternative."
- Nondefinitive language, such as the following, is required to satisfy component 2: "evidence to accept the alternative," "there is evidence for the alternative," and "there is not sufficient evidence for the alternative."

- 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.

Confidence Interval Approach:

- If the alternative hypothesis is specified correctly as $H_a: p > 0.22$, then component 1 is satisfied if the justification is based on whether 0.22 is below the lower end of the confidence interval. If the alternative hypothesis is stated in the wrong direction, then component 1 is satisfied if the justification is based on whether 0.22 is above the upper end of the confidence interval.
- If an incorrect two-sided alternative hypothesis is specified, then component 1 is satisfied if the justification is based on whether 0.22 is included in the confidence interval.
- 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.

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 4: Focus on Inference

4 points

General Scoring Notes

- This question is scored in four sections. Each section is initially scored by determining if it meets the criteria for essentially correct (E), partially correct (P), or incorrect (I). The first section includes statements of the null and alternative hypotheses and identification of the appropriate hypothesis test in part (a). The second section includes verifying the conditions for the test identified in part (a) and calculating the value of the test statistic and the corresponding p -value. The third section includes the conclusion for the test identified in part (a). The fourth section includes the response to part (b). The response is then categorized based on the scores assigned to each section and awarded an integer score between 0 and 4 (see the table at the end of the question).
- The model solution represents an ideal response to each section of the question, and the scoring criteria identify the specific components of the model solution that are used to determine the score.

	Model Solution	Scoring
(a) Section 1	Let p represent the proportion of all customers of the pet supply company who would place an order within 30 days after receiving an e-mail with a coupon for \$10 off the next purchase. The null hypothesis is $H_0: p = 0.40$, and the alternative hypothesis is $H_a: p > 0.40$. An appropriate test is a one-sample z-test for a population proportion.	Essentially correct (E) if the response satisfies the following three components: - States the correct equality for the null hypothesis for a proportion (e.g., $p = 0.40$) AND the correct direction of the one-sided alternative hypothesis for a proportion (e.g., $p > 0.40$) - Provides sufficient context for the parameter by including reference to the <i>population</i> proportion AND the sampling units (customers) AND the response variable (placing an order after receiving a coupon) - Identifies a one-sample z-test for a population proportion by name (e.g., “one-proportion z-test”) but <u>not</u> merely “one-sample z-test”) or by formula Partially correct (P) if the response does not meet the criteria for E but satisfies either component 1 and/or component 3. Incorrect (I) if the response does not meet the criteria for E or P.

Additional Notes:

- The elements of component 2 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 proportion and the conclusion refers to sample proportion (or vice versa), then the population aspect of component 2 is not satisfied.

- A response that states the null hypothesis as $H_0: p \leq 0.40$ may satisfy component 1.
- To satisfy component 1, the hypotheses must be stated in terms of a proportion. If a symbol other than p or π is used to denote the proportion, it must be clearly defined as a proportion (but does not need to reflect the context of customers who would place an order within 30 days after receiving a coupon) in order for the response to satisfy component 1. It is acceptable to use “ p_0 ” to denote the proportion.
- A response that states the hypotheses in words (e.g., “the null hypothesis is that the proportion is 0.40, and the alternative hypothesis is that the proportion is greater than 0.40”) may satisfy component 1. Neither context nor the concept of the *population* is required to satisfy component 1.
- A response that states the hypotheses in words (e.g., “the null hypothesis is that the proportion of all customers who would place an order within 30 days after receiving a coupon is equal to 0.40, and the alternative hypothesis is that the proportion is greater than 0.40”) may satisfy component 1 and component 2.
- If the response clearly refers to the *sample* proportion instead of the *population* proportion using words or a symbol (e.g., $\hat{p}$), then component 2 is not satisfied unless the symbol used is defined as the *population* proportion.
- A response may satisfy the population aspect of component 2 by doing the following:
 - referring to population in the statement of the conclusion of the inferential procedure
 - using notation such as p , p_0 , or π when defining the hypothesis statements.
- A response may satisfy the sampling units aspect of component 2 by referring to “people who place an order” or similar statement.
- If the response identifies the correct test by name, but also states an incorrect formula, then component 3 is not satisfied.
- If the response identifies the test by formula using a t -percentile instead of a z -percentile, then component 3 is not satisfied.

Confidence Interval Approach:

- If a one-sample z -interval for a population proportion is identified correctly by name (e.g., “one-proportion z -interval”) but not merely “one-sample z -interval”) or by formula, then component 3 is satisfied.
- If a response uses a one-sample z -interval for a population proportion, then component 2 is satisfied if the response indicates that it is a confidence interval for the proportion of all customers who would place an order within 30 days after receiving a coupon, even if the hypotheses are not stated.

	Model Solution	Scoring
(a) Section 2	The independent observations condition for performing the one-sample z-test for a population proportion is satisfied because the data were obtained from a random sample of 90 customers who placed an order in the past year and, because sampling of customers is done without replacement, it is assumed that this large online company has more than $10(90) = 900$ customers. The sample size is large enough to support an assumption that the sampling distribution of $\hat{p}$ is approximately normal because $(90)(0.4) = 36$ and $(90)(1 - 0.4) = 54$ are both at least 10. The value of the sample proportion is $\hat{p} = \frac{38}{90} \approx 0.422$ and the value of the test statistic is $z = \frac{\frac{38}{90} - 0.40}{\sqrt{\frac{(0.40)(0.60)}{90}}} \approx 0.430$. The corresponding p-value is $P(z > 0.430) \approx 0.333$.	Essentially correct (E) if the response satisfies the following four components: 1. Checks the independence condition by referring to the random selection of 90 customers AND indicating that the company is assumed to have at least 900 customers (i.e., $90 \leq 0.10N$) 2. Checks that the sample size is large enough to support the assumption that the sampling distribution of $\hat{p}$ is approximately normal by verifying that $(90)(0.4)$ and $(90)(1 - 0.4)$ are both at least 10 (or 5) 3. Correctly reports the value of the z-statistic 4. Correctly reports the p-value, consistent with the reported test statistic and stated alternative hypothesis 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:

- In order to satisfy the reference to the random selection of 90 customers in component 1 it is minimally acceptable to state “random sample – check” or “SRS – check.” However, component 1 is not satisfied if the response implies that random *assignment* was used or only states “random - check.”
- In order to satisfy component 2, the response must include actual values of the observed successes and failures, or values for the expected successes and failures, or formulas for the expected number of successes and failures with values inserted AND the response must make a comparison of the two values with some standard criterion, such as 5 or 10. If expressions such as $(90)(0.4)$ and $(90)(1 - 0.4)$ are used, simplification is not required.
 - Examples of acceptable quantities (comparisons must still be made):
 - 38 and 52 (observed counts)
 - 36 and 54 (expected counts under the null hypothesis)
 - $(90)(0.4)$ and $(90)(1 - 0.4)$
 - $(90)(0.4222)$ and $(90)(1 - 0.422)$
 - Unless values of all parameters are explicitly defined in the response, the following quantities are unacceptable:
 - $90p$, $90(1 - p)$, np , $n(1 - p)$
 - $90\hat{p}$, $90(1 - \hat{p})$, $n\hat{p}$, $n(1 - \hat{p})$
 - $n(0.4)$, $n(0.6)$, $n(0.4222)$, $n(1 - 0.4222)$

- When computing the test statistic, using a $\hat{p}$ of 0.42 in the numerator results in a test statistic equal to 0.39 with a p -value of 0.36. These values satisfy components 3 and 4.
- If the response uses $\hat{p}$ in the null standard error formula to calculate the z -statistic, component 3 may be satisfied.
- A response that reports the correct value for the z -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 the response compares the value of the test statistic to a critical value instead of reporting a p -value, then the critical value (1.645), or a critical value consistent with the stated alternative hypothesis, satisfies component 4.
- If a two-tailed alternative hypothesis is stated or the direction of the stated one-tailed alternative hypothesis is incorrect, then the p -value must be consistent with the stated alternative hypothesis to satisfy component 4.
- If the response omits identifying the hypotheses, the correct one-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.

Confidence Interval Approach:

- If either a one-sided 95 percent confidence interval is correctly calculated as $(0.337, 1)$, or a two-sided 90 percent confidence interval is correctly calculated as $(0.337, 0.508)$, then component 3 is satisfied.
- If only the lower end of a confidence interval is used to reach a conclusion, then component 4 is satisfied.
- Application of a confidence interval approach must be consistent with the stated alternative to satisfy component 4. A two-sided 95 percent confidence interval is $(0.320, 0.524)$, and a lower one-sided 95 percent confidence interval is $(0, 0.508)$.

	Model Solution	Scoring
(a) Section 3	Because the p -value is greater than $\alpha = 0.05$, the null hypothesis should not be rejected. The results from this study do not provide convincing statistical evidence that the manager's belief is correct. That is, there is not convincing statistical evidence that more than 40 percent of all customers of the pet supply company would place an order within 30 days after receiving an e-mail with a coupon for \$10 off the next purchase.	Essentially correct (E) if the response satisfies the following two components: 1. Provides a correct justification of the conclusion based on whether the p-value is less than $\alpha = 0.05$ (or a comparison of the value of the test statistic to an appropriate critical value, e.g., $z < -1.645$) 2. States a correct conclusion consistent with the stated alternative hypothesis OR states a conclusion that answers the inference question (e.g., states the conclusion in terms of the manager's belief) 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:

- Although including proper context (the concept of population proportion and referencing the response variable) is important in stating the conclusion, context displayed in stating the conclusion is considered in scoring component 2 of Section 1.
- The response need not make an explicit decision about the null hypothesis (reject H_0 or fail to reject H_0) in order to satisfy component 1. However, if an incorrect decision is stated, then component 1 is not satisfied.
- If the conclusion and justification 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 and component 2.
- If the response omits hypotheses, assume the correct alternative hypothesis, $H_a: p > 0.40$, was provided when scoring component 1 and component 2.
- If the conclusion includes a definitive statement (e.g., "this proves that we do not have enough evidence to claim..." or "there is no evidence..."), then component 2 is not satisfied.
- If the response includes a statement that is equivalent to accepting the null hypothesis (e.g., "we conclude that the proportion of customers who will place an order is 0.40"), then component 2 is not satisfied.
- If the response includes an incorrect interpretation of the p -value, then the score for Section 3 is lowered one level (that is, from E to P or from P to I).
- The clarity and quality of the statement of the conclusion and the statement of the justification may be used in a holistic approach to decide whether to score up or down (e.g., raising a score of 2.5 to 3 or reducing a score of 2.5 to 2).

Confidence Interval Approach:

- If the alternative hypothesis is specified correctly as $H_a: p > 0.40$, then component 1 is satisfied if the justification is based on whether 0.40 is below the lower end of the confidence interval. If the alternative hypothesis is stated in the wrong direction, then component 1 is satisfied if the justification is based on whether 0.40 is above the upper end of the 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 0.40 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 one level (that is, from E to P or from P to I).

	Model Solution	Scoring
(b) Section 4	Because the null hypothesis was not rejected in part (a), a Type II error could have been made. A Type II error occurs when the null hypothesis is false and is not rejected. In this case, a Type II error is made by failing to reject the null hypothesis that 40 percent (or less) of all customers of the pet supply company would place an order within 30 days after receiving an e-mail with a coupon for \$10 off the next purchase, when in fact, more than 40 percent would do so. Consequently, the manager may decide not to use the coupon promotion when it actually would result in more than 40 percent of their customers making a purchase within 30 days.	Essentially correct (E) if the response satisfies the following two components: 1. States that a Type II error could have been made 2. Provides a reasonable interpretation of the consequence of the stated error AND uses sufficient context, by including at minimum "those who would place an order" or "coupon" or "sales" (e.g., indicating the manager may not use the coupon promotion when it would actually lead to more than 40% of customers placing an order) 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:

- If the response to part (a) rejects the null hypothesis, then
 - Component 1 is satisfied if the response states that a Type I error could have been made.
 - Component 2 is satisfied if the response provides a reasonable interpretation of the consequence of a Type I error AND uses sufficient context by including at minimum "those who would place an order" or "coupon" or "sales."
- If the response states that a Type II error could have been made, followed by an incorrect description of a Type II error (e.g., "did not find convincing evidence that more than 40% of customers will place an order when there actually was evidence of more than 40%"), component 1 is not satisfied.
- The clarity and quality of the statement of the consequence may be used in a holistic approach to decide whether to score up or down (e.g., raising a score of 2.5 to 3 or reducing a score of 2.5 to 2).

Scoring for Question 4

Each essentially correct (E) part counts as 1 point, and each partially correct (P) part counts as $\frac{1}{2}$ point.

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

If a response is between two scores (for example, 2 $\frac{1}{2}$ points), use a holistic approach to decide whether to score up or down, depending on the strength of the response and quality of the communication.