2025

AP® Statistics

Sample Student Responses and Scoring Commentary

Inside:

Free-Response Question 4

- ☑ **Scoring Guidelines**
- ☑ **Student Samples**
- ☑ **Scoring Commentary**

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: 1. 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 2. States the correct equality for the null hypothesis with the value 0.22 3. States the correct direction for the one-sided alternative hypothesis consistent with the null hypothesis 4. 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: 1. 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., $(130) \le (0.10)N$) 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 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 $n(1 - p_0) = 130(0.78) = 101.4$ 3. Correctly reports the value of the z-statistic 4. 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

Begin your response to **QUESTION 4** on this page.

1 sample z test for p

p = The true proportion of high school students at Karens school that use the app to help them with their homework at least once a week.

$H_0: p = 0.22$ $H_A: p > 0.22$ $\alpha: 0.05$

Conditions:

Random: Karen choose a random sample of 130 students at her school

10% Condition: $130 \leq 0.1(2{,}000)$ 2,000 = All students at Karens school

Large Counts Condition: $0.22(130) = 28.6 \geq 10$

$0.78(130) = 101.4 \geq 10$

Since all of the conditions are met we can perform a 1 sample z test for p

$$\frac{0.2923 - 0.22}{\sqrt{\frac{0.22 \cdot 0.78}{130}}} = 1.9902 \qquad P(Z \geq 1.9902) = 0.02328$$

Since the p value of $0.02328 < 0.05$ we reject H_0, and have convincing evidence for H_A, that the true proportion of students at Karens school that use the app to help them with homework at least once a week is greater than 0.22

GO ON TO THE NEXT PAGE.

Question 4

Begin your response to **QUESTION 4** on this page.

State: 1 propZtest

$H_0: p = 0.22$

$H_a: p > 0.22$

$\alpha = 0.05$

$p = 38/130 = 0.292$

p = the true proportion of all students at Karen's school who use the app to help them with their homework at least once per week

Plan: random sample of 130 students from Karens school ✓

normal distribution ✓

$130 > 30$

$(130)(0.292) > 10$

$37.96 > 10$ ✓

$(130)(1-0.292) > 10$

$92.04 > 10$ ✓

DO:

$$\frac{0.292 - 0.22}{\sqrt{\frac{0.292(1-0.292)}{130}}} = 1.9902$$

$= 0.9767$

p-value $1 - 0.9767 = 0.0233$

Conclude: The p-value of 0.0233 is lower than the alpha of 0.05 so we reject the null hypothesis, there is convincing evidence that the proportion of students at Karen's school who use the app to help them with their homework at least once per week is greater than the proportion of her country.

GO ON TO THE NEXT PAGE.

Question 4

Begin your response to **QUESTION 4** on this page.

4.) 1-PropZ test

H_0 : $\hat{P}_W = \hat{P}_H$

H_a : $\hat{P}_W < \hat{P}_H$

$\hat{P}_W$: Proportion of Country W's Students who use app to help with homework

$\hat{P}_H$: Proportion of Karen's highschool that uses app to help with homework

$n > 30$	$130 > 30$	✓
Random	Random	✓
$n(p) > 10$	$130(.292) > 10$	✓
$n(1-p) > 10$	$130(1-.292) > 10$	✓

$\hat{P}_H = \frac{38}{130} = 0.292$

$\alpha = .05$

Do: 1-PropZ test

$P_W = 0.22$

$X = 38$

$n = 130$

Prop: $> P_W$

$Z = 1.99$

P-Value $= 0.023$

GO ON TO THE NEXT PAGE.

Question 4

Continue your response to **QUESTION 4** on this page.

There is convincing statistical evidence at the .05 = α level to reject the null hypothesis that the porportion of Karen's highschool who use the app is equal to the porportion of country wide students who use the app. Because, .023 < .05 = α

GO ON TO THE NEXT PAGE.

Use a pencil or a pen with black or dark blue ink. Do NOT write your name. Do NOT write outside the box.

Question 4

Note: Student samples are quoted verbatim and may contain spelling and grammatical errors.

Overview

NEW for 2025: The question overviews can be found in the *Chief Reader Report on Student Responses* on AP Central.

Sample: 4A
Score: 4

The response earned the following: Section 1 – E, Section 2 – E, and Section 3 – E.

In Section 1 the response correctly identifies the one sample *z*-test for a proportion, satisfying component 1; correctly states the null hypothesis, satisfying component 2; correctly states the alternative hypothesis, satisfying component 3; and refers to the population parameter by using *p* as well as "true" in the definition of the parameter and uses "students" and "app," satisfying component 4. The response for this section was scored essentially correct (E).

In Section 2 the response correctly checks the independence condition with "choose a random sample" and indicating that "$130 \leq 0.1(2{,}000)$," satisfying component 1; correctly checks the large counts condition by verifying the number of expected successes and failures are at least 10, satisfying component 2; reports a correct *z*-statistic, satisfying component 3; and gives a correct *p*-value, satisfying component 4. The response for this section was scored essentially correct (E).

In Section 3 the response provides a correct comparison of the *p*-value to alpha "$0.02328 < 0.05$," and makes a correct decision of "reject H_0," satisfying component 1. The response states a correct conclusion in context and in terms of the alternative hypothesis using nondefinitive language ("convincing evidence for H_A, that the true proportion of students at Karens school that use the app to help them with homework at least once a week is greater than 0.22"), satisfying component 2. The response for this section was scored essentially correct (E).

Sample: 4B
Score: 3

The response earned the following: Section 1 – E, Section 2 – P, and Section 3 – E.

In Section 1 the response correctly identifies the "1 prop *z* test" satisfying component 1. (Although the response uses an incorrect formula for the *z*-statistic, it is not an unreasonable formula, therefore component 1 is still satisfied). The response correctly states the null hypothesis, satisfying component 2, and it correctly states the alternative hypothesis, satisfying component 3. The response refers to the population parameter by using *p* as well as "true" in the definition of the parameter and uses "students" and "app," satisfying component 4. The response for this section was scored essentially correct (E).

In Section 2 the response includes "random sample ✓"; however, it does not indicate that the sample is less than 10% of the population, so component 1 is not satisfied. The response correctly checks the large counts condition by verifying the number of observed successes and failures are at least 10; however, the response also includes an inappropriate check of conditions "$130 > 30$," so it does not satisfy component 2. The response reports a correct *z*-statistic (even though the numbers in the formula are not correct in the denominator), satisfying component 3, and it and gives a correct *p*-value, satisfying component 4. For satisfying only two of the four components, the response for this section was scored partially correct (P).

Question 4 (continued)

In Section 3 the response provides a correct comparison of the *p*-value to alpha, "0.0233 is lower than the alpha of 0.05," and makes a correct decision of "reject the null," satisfying component 1. The response states a correct conclusion in context and in terms of the alternative hypothesis using nondefinitive language ("convincing evidence that the proportion of students at Karen's school who use the app to help them with their homework at least once per week is greater than the proportion of her country"), satisfying component 2. The response for this section was scored essentially correct (E).

Sample: 4C
Score: 1

The response earned the following: Section 1 – I, Section 2 – P, and Section 3 – P.

In Section 1 the response correctly identifies the "1 - Prop *Z* test," satisfying component 1. The response does not include the value of .22 in the null hypothesis, so component 2 is not satisfied. The response does state the correct direction for the alternative, consistent with the null, so component 3 is satisfied. (The response does show a less-than symbol, but it uses that symbol correctly). The response uses a *p*-hat that is not defined as the population in words (true, all, etc.), so component 4 is not satisfied. Because only one of the four components is satisfied, the response for this section was scored incorrect (I).

In Section 2 the response does not check the independence condition as all it states is "random ✓" (which is insufficient), and it omits the 10% rule, so component 1 is not satisfied. The response correctly checks the large counts condition by verifying the number of observed successes and failures are at least 10; however, it includes an inappropriate check of conditions "130 > 30 ," so component 2 is not satisfied. The response reports a correct *z*-statistic, satisfying component 3, and it gives a correct *p*-value, satisfying component 4. For satisfying only two of the four components, the response for this section was scored partially correct (P).

In Section 3 the response provides a correct comparison of the *p*-value to alpha ".023 < 0.5" and makes a correct decision of "reject the null hypothesis," satisfying component 1. The response incorrectly states a conclusion in terms of the null hypothesis and omits the word "students" in connection to Karen's school, so it does not satisfy component 2. Because only one of the two components is satisfied, the response for this section was scored partially correct (P).