2024

AP® Statistics

Sample Student Responses and Scoring Commentary

Inside:

Free-Response Question 1

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

Question 1: 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 p_{younger} represent the proportion of all exercise center members from 18 to 55 years of age who would be interested in taking online fitness classes, and p_{older} represent the proportion of all exercise center members 56 years or older who would be interested in taking online fitness classes. The null hypothesis is $H_0: p_{\text{younger}} = p_{\text{older}}$ and the alternative hypothesis is $H_a: p_{\text{younger}} \neq p_{\text{older}}$. An appropriate inference procedure is a two-sample *z*-test for a difference of population proportions.	**Essentially correct (E)** if the response satisfies the following three components: 1. Identifies a two-sample *z*-test for a difference of population proportions by name (e.g., "two-proportion *z*-test" or "two-sample *z*-test") or by formula 2. States a correct null hypothesis of equal proportions *AND* a correct two-sided alternative hypothesis of unequal proportions 3. Provides sufficient context by referencing the two groups (18 to 55 years of age and 56 years or older) *AND* the populations (all exercise center members) **Partially correct (P)** if the response satisfies 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 equates it with a *t* instead of a *z*, then component 1 is not satisfied.
- Neither the names of the groups nor the concept of *population* is required to satisfy component 2.

- A response that states the hypotheses in words (e.g., the null hypothesis is that the proportions are equal, and the alternative hypothesis is that the proportions are not equal) may satisfy component 2. If the hypotheses also include the group names (younger and older members) and reference to the populations, then both components 2 and 3 may be satisfied.
- Group or population aspects of component 3 may be satisfied anywhere in the response, provided there are no incorrect references to groups or population elsewhere in the response (e.g., using $\hat{p}$ in hypotheses, referring to samples in conclusion, or describing groups as "those who are interested in online exercise classes" and "those who are not interested in online exercise classes").
- If the response clearly refers to *sample* proportions instead of *population* proportions using words or symbols (e.g., $\hat{p}_o$ and $\hat{p}_y$), then component 2 may be satisfied, but component 3 is not satisfied unless the symbols used are defined as *population* proportions.
- A response may satisfy the population aspect of component 3 by
  - Referring to the population by using words such as: "population," "all," or "true" when defining the parameter.
  - Using notation such as p_1, p_2, p_y, p_o, π_y, π_o, pi_y, pi_o, when defining the hypothesis statements.

    Note that subscripts y and o 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.

Confidence Interval Approach:

- If a two-sample *z*-interval for a difference in population proportions is identified correctly by name (e.g., "two-proportion *z*-interval" or "two-sample *z*-interval") or by formula, then component 1 may be satisfied.
- If the response uses individual one-sample *z*-intervals for the two proportions, 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 proportions.
- If a response uses a two-sample *z*-interval for a population proportion, it must include references to group names and populations to satisfy component 3. For example, component 3 is satisfied if the response indicates that it is a confidence interval for the difference in population proportions of younger and older exercise center members.

Chi-Square Test for Homogeneity Approach:

- If the response uses a chi-square test approach, identifying the procedure name as a "chi-square test for homogeneity," component 1 may be satisfied.
- If the response identifies the procedure as a "chi-square test for independence" or just a "chi-square test," then component 1 is not satisfied.

	Model Solution	Scoring
Section 2	The independent observations condition for performing the two-sample z-test for a difference in population proportions is satisfied because the data were obtained from a random sample of 170 exercise center members ages 18 to 55 years and a second random sample of 230 exercise center members ages 56 years and older. The 10% condition must be met by both samples because sampling of exercise center members is done without replacement. There are more than $10(170) = 1{,}700$ adults from 18 to 55 years of age who are members of the exercise center and more than $10(230) = 2{,}300$ adults ages 56 years and older who are members of the exercise center. The value of the sample proportions are $\hat{p}_{\text{younger}} = \frac{51}{170} = 0.3$ and $\hat{p}_{older} = \frac{79}{230} \approx 0.3435.$ The combined proportion is $\hat{p}_{\text{c}} = \frac{170(0.3) + 230(0.3435)}{170 + 230} \approx 0.325.$ The sample size is large enough to support an assumption that the sampling distribution of $\hat{p}_{\text{younger}} - \hat{p}_{\text{older}}$ is approximately normal because $170(0.325) = 55.25$, $(170)(1 - 0.325) = 114.75$, $230(0.325) = 74.75$, and $(230)(1 - 0.325) = 155.25$ are all at least 10. The value of the test statistic is $z = \frac{0.3 - 0.3435}{\sqrt{0.3250(1 - 0.3250)}\sqrt{\frac{1}{170} + \frac{1}{230}}}$ $z \approx -0.918.$ The corresponding p-value is $2 * P(z < -0.918) \approx 0.359.$	**Essentially correct (E)** if the response satisfies the following four components: 1. Checks the independence condition by referring to the two random samples *AND* indicates that there are more than $10(170) = 1{,}700$ exercise center members from 18 to 55 years of age and more than $10(230) = 2{,}300$ exercise center members ages 56 years and older 2. Checks that the sample sizes are large enough by verifying that $170(0.325) = 55.25$, $(170)(1 - 0.325) = 114.75$, $230(0.325) = 74.75$, and $(230)(1 - 0.325) = 155.25$ are all at least 5 (or 10) 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 required for E. **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 exercise center members in component 1, it is minimally acceptable to state "random samples—check" or "SRSs—check." However, component 1 is not satisfied if the response implies that random *assignment* was used or only states "random—check."

- If the response states that the random sample of members in each age group is less than 10 percent of total membership in that group (with sample sizes provided), component 1 may be satisfied. For example, "$170 < 10\%$ of all members ages 18 to 55, and $230 < 10\%$ of all members ages 56 and older."
- In order to satisfy component 2, the response must include 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 four values with some standard criterion, such as 5 or 10. If expressions such as $170(0.325)$ and $(170)(1 - 0.325)$ are used, simplification is not required.
- Examples of acceptable quantities (comparisons must still be made):
  - $55.25,\ 114.75,\ 74.75,$ and 155.25
  - $(170)(0.325),\ (170)(1 - 0.325),\ 230(0.325),$ and $(230)(1 - 0.325)$
  - 51, 119, 79, and 151 (observed counts)
  - $(170)(0.3),\ (170)(0.7),\ (230)(0.3435),$ and $(230)(0.6565)$
- Unless values of relevant parameters are explicitly identified in the response, the following quantities are unacceptable:
  - $170p_1,\ 170(1 - p_1),\ n_1p_1,\ n_1(1 - p_1),\ 230p_2,\ 230(1 - p_2),\ n_2p_2,\ n_2(1 - p_2)$
  - $170\hat{p}_1,\ 170(1 - \hat{p}_1),\ n_1\hat{p}_1,\ n_1(1 - \hat{p}_1),\ 230\hat{p}_2,\ 230(1 - \hat{p}_2),\ n_2\hat{p}_2,\ n_2(1 - \hat{p}_2)$
- A response that reports the correct value for the *z*-statistic but contains errors in supporting work may still satisfy component 3.
- A response that reports a value for the *z*-statistic using the incorrect estimate of the standard error of the difference $\hat{p}_{younger} - \hat{p}_{older}$ based on individual estimates for each group may still satisfy components 3 and 4, in which case $z = -0.9237$ and p-value $= 0.3557$.
- A response that uses Table A to determine the *p*-value from the rounded $z = -0.92$ should report p-value $= 0.3576$.
- A response that inputs correct values into the *z*-statistic formula but computes an incorrect value for the *z*-statistic, may 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.

Confidence Interval Approach:

- If the stated alternative hypothesis is correct or no alternative hypothesis is provided:
  - If a two-sided 95 percent confidence interval for $p_{\text{younger}} - p_{\text{older}}$ is correctly calculated as $(-0.1357,\ 0.0488)$, 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 the stated alternative hypothesis is incorrect (one-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 percent confidence interval for $p_{\text{younger}} - p_{\text{older}}$ is $(-1, 0.3395)$ or an upper one-sided 95 percent confidence interval for $p_{\text{younger}} - p_{\text{older}}$ is $(-0.1209, 1)$.
  - 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 $p_{\text{younger}} - p_{\text{older}}$ should be scored the same way as the interval for $p_{\text{older}} - p_{\text{younger}}$.

Chi-Square Test for Homogeneity Approach:

- Component 2 is satisfied if the response verifies that all four expected counts are at least 5 (or 10), so that the test statistic has an approximate chi-square distribution.
- Component 3 is satisfied if the response correctly reports the chi-square statistic (no work is required).

	Model Solution	Scoring
Section 3	Because the *p*-value of approximately 0.359 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 population proportion of exercise center members from 18 to 55 years of age who would be interested in taking online fitness classes is different from the population proportion of adults ages 56 years and older who would be interested in taking online fitness classes.	**Essentially correct (E)** if the response satisfies the following two components: 1. Provides correct comparison of the *p*-value to alpha (*p*-value is 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 non-definitive 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:

- In order 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 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., $|z| > 1.96$.
- 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 references to proportions, the groups (ages 18 to 55, 56 or older), the sampling units (members), and the variable of interest (interest in online fitness classes).
- 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 non-definitive language in component 2 include “evidence to accept the alternative,” “there is evidence for the alternative,” and “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," and "no 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:

- Component 1 is satisfied if the justification is based on whether zero 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 percent confidence interval for $p_{\text{younger}} - p_{\text{older}}$ is $(-1, 0.3395)$ or an upper one-sided 95 percent confidence interval for $p_{\text{younger}} - p_{\text{older}}$ is $(-0.1209, 1)$.
- 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 1	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 1

Begin your response to **QUESTION 1** on this page.

STATISTICS

SECTION II

Total Time—1 hour and 30 minutes

6 Questions

SECTION II, Part A

Suggested Time—1 hour and 5 minutes

5 Questions

Directions: Show all your work. Indicate clearly the methods you use, because you will be scored on the correctness of your methods as well as on the accuracy and completeness of your results and explanations.

1. A large exercise center has several thousand members from age 18 to 55 years and several thousand members age 56 and older. The manager of the center is considering offering online fitness classes. The manager is investigating whether members' opinions of taking online fitness classes differ by age. The manager selected a random sample of 170 exercise center members ages 18 to 55 years and a second random sample of 230 exercise center members ages 56 years and older. Each sampled member was asked whether they would be interested in taking online fitness classes.

51 | 170 79 | 230

The manager found that 51 of the 170 sampled members ages 18 to 55 years and that 79 of the 230 sampled members ages 56 years and older said they would be interested in taking online fitness classes.

At a significance level of $\alpha = 0.05$, do the data provide convincing statistical evidence of a difference in the proportion of all exercise center members ages 18 to 55 years who would be interested in taking online fitness classes and the proportion of all exercise center members ages 56 years and older who would be interested in taking online fitness classes? Complete the appropriate inference procedure to justify your response.

2 sample z test for a difference in proportions

$H_0: p_1 = p_2$

$H_a: p_1 \neq p_2$

where p_1 represents the proportion of all exercise center members ages 18 to 55 years who would be interested in taking online fitness classes and p_2 represents the proportion of all exercise center members ages 56 years and older who would be interested in taking online fitness classes.

Question 1

Continue your response to **QUESTION 1** on this page.

conditions:

1) random - question states that a random sample of 170 exercise center members ages 18 to 55 years was selected and a random sample of 230 exercise center members age 56 years and older was selected ✓

2) independent - $n \leq 10\%\ N$ ✓

$n_1 = 170 \leq 10\%\ N$ reasonable to assume over 1700 people aged 18-55 are member of the exercise center, making this true

$n_2 = 230 \leq 10\%\ N$ reasonable to assume over 2300 people aged 56+ are members of the exercise center, making this true

states both N = several thousand

3) normal -

$\hat{p}_c = \frac{x_1 + x_2}{n_1 + n_2} = \frac{51 + 79}{170 + 230} = .325$

$n_1\hat{p}_c \geq 10 = 170(.325) = 55.25$ ✓

$n_1(1-\hat{p}_c) \geq 10 = 170(.675) = 114.75$ ✓

$n_2\hat{p}_c \geq 10 = 230(.325) = 74.75$ ✓

$n_2(1-\hat{p}_c) \geq 10 = 230(.675) = 155.25$ ✓

calculations:

stat → tests →

6: 2-prop z test

$x_1 = 51$ $x_2 = 79$

$n_1 = 170$ $n_2 = 230$

$p_1 \neq p_2$

$z = .9177765091$

$p = .3587357818$

Because the p value of .36 > $\alpha = 0.05$ we fail to reject H_0. There is not convincing statistical evidence that the proportion of all exercise center members ages 18-55 who would take an online fitness class differs from the proportion of all center members 56+ who would take an online fitness class.

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.

0143170

Q5261/3

Question 1

Begin your response to **QUESTION 1** on this page.

STATISTICS

SECTION II

Total Time—1 hour and 30 minutes

6 Questions

SECTION II, Part A

Suggested Time—1 hour and 5 minutes

5 Questions

Directions: Show all your work. Indicate clearly the methods you use, because you will be scored on the correctness of your methods as well as on the accuracy and completeness of your results and explanations.

1. A large exercise center has several thousand members from age 18 to 55 years and several thousand members age 56 and older. The manager of the center is considering offering online fitness classes. The manager is investigating whether members' opinions of taking online fitness classes differ by age. The manager selected a random sample of 170 exercise center members ages 18 to 55 years and a second random sample of 230 exercise center members ages 56 years and older. Each sampled member was asked whether they would be interested in taking online fitness classes.

   $51/170 = .3$ $79/230 = .34$

   The manager found that 51 of the 170 sampled members ages 18 to 55 years and that 79 of the 230 sampled members ages 56 years and older said they would be interested in taking online fitness classes.

   At a significance level of $\alpha = 0.05$, do the data provide convincing statistical evidence of a difference in the proportion of all exercise center members ages 18 to 55 years who would be interested in taking online fitness classes and the proportion of all exercise center members ages 56 years and older who would be interested in taking online fitness classes? Complete the appropriate inference procedure to justify your response.

P: P_1 = proportion of exercise center members ages 18 to 55 years interested in online classes

P_2 = proportion of exercise center members ages 56+ year interested in online classes

H: $H_0: P_1 = P_2$ $\alpha = .05$

$H_a: P_1 \neq P_2$

A: Random? - Stated a random sample of 170 18-55 year olds and 230 56+ year olds was taken

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 1

Continue your response to **QUESTION 1** on this page.

Independent?- safe to assume there are at least 1,700 exercise center members between 18 and 55 and at least 2,300 members age 56+ : Independence

Large?- $.3(170) = 51$ $.7(170) = 119$ $.34(230)$ $.66(230)$

$51 \geq 30$ $119 \geq 30$ $78 \geq 30$ $152 \geq 30$

It is safe to use normal approximation

N: 2 porportion Z-test

T: $Z = -.918$

O: $p = .359$

M: Since the p-value $(.359) > \alpha (.05)$ we fail to reject the null

S: There is not a difference in the porportion of all exercise members ages 18 to 55 years old who would be intrested in taking online fitness and the porportion of all exercise members ages 56+ year old intrested in the same.

Question 1

Begin your response to **QUESTION 1** on this page.

STATISTICS

SECTION II

Total Time—1 hour and 30 minutes

6 Questions

SECTION II, Part A

Suggested Time—1 hour and 5 minutes

5 Questions

Directions: Show all your work. Indicate clearly the methods you use, because you will be scored on the correctness of your methods as well as on the accuracy and completeness of your results and explanations.

1. A large exercise center has several thousand members from age 18 to 55 years and several thousand members age 56 and older. The manager of the center is considering offering online fitness classes. The manager is investigating whether members' opinions of taking online fitness classes differ by age. The manager selected a random sample of 170 exercise center members ages 18 to 55 years and a second random sample of 230 exercise center members ages 56 years and older. Each sampled member was asked whether they would be interested in taking online fitness classes.

   The manager found that 51 of the 170 sampled members ages 18 to 55 years and that 79 of the 230 sampled members ages 56 years and older said they would be interested in taking online fitness classes.

   At a significance level of $\alpha = 0.05$, do the data provide convincing statistical evidence of a difference in the proportion of all exercise center members ages 18 to 55 years who would be interested in taking online fitness classes and the proportion of all exercise center members ages 56 years and older who would be interested in taking online fitness classes? Complete the appropriate inference procedure to justify your response.

2 Sample Z

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 1

Continue your response to **QUESTION 1** on this page.

State: $H_0 = p_1 - p_2$

$Ha: p_1 - p_2$

$\alpha = .05$

2 Sample z test

Let p_1 represent the true proportion of exercise center members ages 18 to 55 years who would be interested in taking online fitness classes.

Let p_2 represent the true proportion of exercise center members ages 56 and older who would be interested in taking online fitness classes.

Plan:

Random ✓

Both age groups were randomly sampled.

*All conditions met!

Independent ✓

It is safe to assume that both 170 is less than 10% of all 18-55 year olds and 230 is less than 10% of all people 56 years old or older.

Large ✓

~~$np(1-p) \geq 10$~~ CLT would apply for both

$n_1(p_1) \geq 10$

$n_1(1-p_1) \geq 10$

$n_2(p_2) \geq 10$

$n_2(1-p_2) \geq 10$

Do:

Conclude:

Since the p-value .02 is less than α at .05 we reject the H_0 and we do have convincing evidence that there is a difference in age groups that would participate in online fitness classes.

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.

0002410

Q5261/3

Question 1

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

Overview

The primary goals of the question were to assess a student's ability to (1) identify an appropriate procedure for conducting a hypothesis test for a two-sample *z*-test for a difference of population proportions; (2) identify the correct hypotheses for conducting a two-sample *z*-test for a difference of population proportions; (3) check the conditions for the hypothesis test for a two-sample *z*-test for a difference of population proportions; (4) calculate the test statistic and *p*-value for a two-sample *z*-test for a difference of population proportions; (5) compare the *p*-value to a significance level to make a decision regarding the hypotheses; and (6) determine an appropriate conclusion for a hypothesis test for a two-sample *z*-test for a difference of population proportions.

This question primarily assesses skills in skill category 1: Selecting Statistical Methods, skill category 3: Using Probability and Simulation, and skill category 4: Statistical Argumentation. Skills required for responding to this question include (1.E) Identify an appropriate inference method for significance tests, (1.F) Identify null and alternative hypotheses, (3.E) Calculate a test statistic and find a *p*-value, provided conditions for inference are met, (4.C) Verify that inference procedures apply in a given situation, and (4.E) Justify a claim using a decision based on significance tests.

This question covers content from Unit 6: Inference for Categorical Data: Proportions of the course framework in the AP Statistics Course and Exam Description. Refer to topics 6.10 and 6.11 and learning objectives VAR-6.H, VAR-6.I, VAR-6.J, VAR-6.K, and DAT-3.D.

Sample: 1A
Score: 4

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

In section 1 the response correctly identifies the procedure as a "2 sample z test for a difference in proportions," satisfying component 1. The response includes correct null and two-sided alternative hypotheses in terms of p_1 and p_2, satisfying component 2. The two proportions are correctly defined, referencing the two groups being compared, and the response includes an appropriate reference to population ("all"), satisfying component 3. Note that the use of a standard symbol for the parameter (p) also satisfies the population aspect of component 3. Section 1 was scored essentially correct (E).

In section 2 the response states that two random samples were taken and that the sample sizes of 170 and 230 are each less than 10% of the corresponding population size, together satisfying component 1. The response correctly calculates the expected number of successes and failures for each sample based on the pooled estimate of the common proportion under the null hypothesis and notes that all four exceed 10, satisfying component 2. The reported positive value of the *z* statistic is correct for a two-sided alternative, satisfying component 3. A correct *p*-value is given, satisfying component 4. Section 2 was scored essentially correct (E).

In section 3 the response makes a correct comparison of the *p*-value to $\alpha = 0.05$ and makes a correct decision, satisfying component 1. A conclusion in context, in terms of the alternative hypothesis, is provided using non-definitive language ("not convincing statistical evidence"), satisfying component 2. Section 3 was scored essentially correct (E).

Question 1 (continued)

Sample: 1B
Score: 2

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

In section 1 the response correctly identifies the procedure as a "2 porportion Z-test" satisfying component 1. The response includes correct null and two-sided alternative hypotheses using two proportions, satisfying component 2. The response includes definitions of the two proportions that satisfy the "group" aspect of component 3 but do not satisfy the population aspect. However, the use of standard population notation for proportions used in the hypotheses satisfies the population aspect. Also, the response included "all" in the conclusion, which satisfies component 3 of section 1. Section 1 was scored essentially correct (E).

In section 2 the response notes that separate random samples were taken from each age group, and the 10% condition is checked, satisfying component 1. The response shows the calculation of the observed number of successes and failures for each age group but compares them to 30. Comparing counts to 30 is not standard and suggests that a large sample size condition for means is being checked, not a large count condition, hence fails to satisfy component 2. Correct values of both the *z* statistic and the *p*-value are given, satisfying components 3 and 4. Section 2 was scored partially correct (P).

In section 3 the response makes a correct comparison of the *p*-value to $\alpha = 0.05,$ and makes a correct decision, satisfying component 1. The conclusion "[t]here is not a difference …" is too definitive, failing to satisfy component 2. Section 3 was scored partially correct (P).

Sample: 1C
Score:1

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

In section 1 the response identifies the procedure in two different places, first as "2 Sample Z," and elaborating this description later as "2 Sample Z test." The first statement is insufficient to satisfy component 1, but the second statement can be considered a clarification and does satisfy component 1. Incorrect notation is used in defining both the null and alternative hypotheses $\left(\text{“}H_0 = p_1 - p_2\text{”}\right.$ and $\left.\text{“}H_a \neq p_1 - p_2\text{”}\right),$ failing to satisfy component 2. The response includes definitions of the two proportions, addressing both the group and population ("true") aspects, satisfying component 3. Section 1 was scored partially correct (P).

In section 2 the response correctly states, "Both age groups were randomly sampled." The check of the 10% conditions states that "less than 10% of all 18-55 year olds" and similarly for the 56+ age group. The word "all" in this context was interpreted as shorthand for all exercise club members. Because the population aspect is assessed in section 1, this was considered a minor error in section 2, so the response satisfies component 1. The response does not provide an appropriate check that sample sizes are large enough for either sample. The statement "CLT would apply for both" does not provide a numerical check and fails to satisfy component 2. The response does not include a value for the *z* statistic, failing to satisfy component 3. The response provides an incorrect *p*-value $(0.02),$ failing to satisfy component 4. Section 2 was scored incorrect (I).

In section 3 the hypotheses are incorrect or ambiguous; therefore, this section was evaluated as if the correct hypotheses were included. The response makes a correct comparison of the stated *p*-value to $\alpha = 0.05$ and makes a consistent decision to reject the null hypothesis, satisfying component 1. The response does not mention proportions, failing to satisfy component 2. Section 3 was scored partially correct (P).