2024

AP® Statistics

Scoring Guidelines

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:

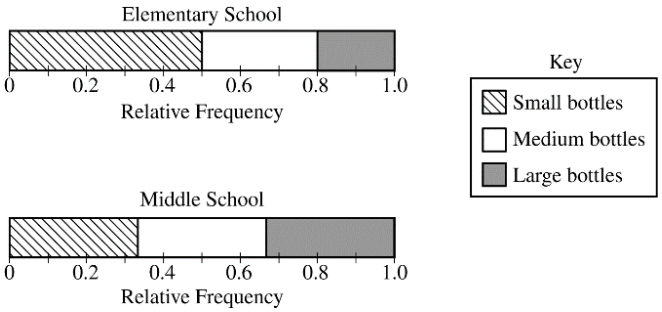
- Component 1 is satisfied if the justification is based on whether zero is included in the confidence interval.
  - If only the upper end of the lower one-sided confidence interval is used to reach a conclusion or only the lower end of the upper one-sided confidence interval is used to reach a conclusion, 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 2: Focus on Exploring Data — 4 points

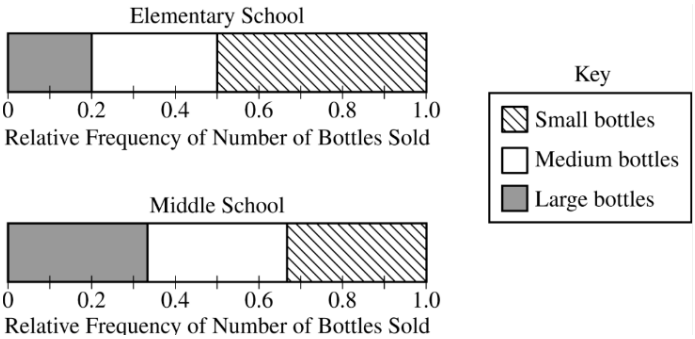
General Scoring Notes

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

Model Solution	Scoring
(a) 	**Essentially correct (E)** if the response satisfies the following two components: 1. The elementary school's segmented bar graph is correctly partitioned according to the given proportions 2. The middle school's segmented bar graph is correctly partitioned to create three equal areas **Partially correct (P)** if the response only satisfies one of the two components required for E. **Incorrect (I)** if the response does not meet the criteria for E or P.

Additional Notes:

- Responses that do not use the key given may still satisfy both components as long as a new key is provided that defines small, medium, and large bottles.
- Responses that do not use the key given and do not define a new key may receive credit for component 2 as long as the bar is vertically partitioned to create three equal areas, but do not satisfy component 1.
- Responses need to be within ±0.05 of the actual cutoff to be acceptable for demonstrating equal areas.
- Responses do not need to be drawn in the order from smallest bottles to largest bottles as shown in the model solution. For example, the graph could be drawn from largest bottles to smallest bottles.



	Model Solution	Scoring
(b)	No, the segment for small bottles for the elementary school is wider than the segment for small bottles for the middle school; however, the middle school students sold three times as many bottles as the elementary school students. So, if the elementary school sold x number of bottles, the middle school sold $3x$ number of bottles. For example, if the elementary students sold 100 bottles total, then they sold $0.5(100)$ or 50 small bottles. However, because the middle school students sold three times the total number of bottles as the elementary students, they would have sold 300 bottles total and $\left(0.\overline{3}\right)(300) = 100$ small bottles. Because $100 > 50,$ the middle school sold more small bottles, and the elementary school's administrator is not correct.	**Essentially correct (E)** if the response satisfies the following three components: 1. States that the elementary school's administrator is incorrect 2. Provides correct mathematical support verifying the middle school sold more small bottles than the elementary school, consistent with the response to component 1 3. Includes context **Partially correct (P)** if the response satisfies only two of the three components required for E. **Incorrect (I)** if the response does not meet the criteria for E or P.

Additional Notes:

- A response that shows mathematical support by showing the inequalities $0.5x < 0.\overline{3}(3x),\ 0.5x < x,$ $0.5(1) < 0.\overline{3}(3),$ or $(0.5)\left(\frac{1}{4}\right) < \left(0.\overline{3}\right)\left(\frac{3}{4}\right)$ may satisfy component 2.
- In order to satisfy component 3, a response must include "elementary," "middle," and "bottles."
- Part (b) should be scored consistent with the response to part (a).

Model Solution	Scoring
(c) (i) The mosaic plot shows that the proportion of large bottles sold by High School A was 0.7 and the proportion of large bottles sold by High School B was 0.6. High School A sold a greater proportion of large bottles because $0.7 > 0.6$. (ii) The number of bottles sold is represented by the area of the shaded region. The area of the rectangle representing large bottles sold by High School B is clearly larger than the area of the rectangle representing large bottles sold by High School A. Therefore, High School B sold more large bottles than High School A.	**Essentially correct (E)** if the response satisfies the following four components: 1. In part (c-i) the response indicates that High School A sold a greater proportion of large bottles 2. In part (c-i) the response bases reasoning on the height of the rectangles or the relative frequencies representing large bottles 3. In part (c-ii) the response indicates that High School B sold a greater number of large bottles 4. In part (c-ii) the response indicates that the area of the rectangle representing large bottles sold by High School B is larger than the area of the rectangle representing large bottles sold by High School A **Partially correct (P)** if the response satisfies 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:

- If a response refers to "proportions" rather than relative frequencies in part (c-i), the values 0.7 and 0.6 must be provided to satisfy component 2.
- If the response bases reasoning on the heights of rectangles in part (c-i), the response must clearly state that the rectangles representing large bottles are referenced.
- If the response bases reasoning on the areas of rectangles in part (c-ii), the response must clearly state that the areas representing large bottles are referenced.
- A response that provides reasonable estimates of the large bottle areas may satisfy component 4.

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

Question 3: Focus on Sampling and Experimental Design — 4 points

General Scoring Notes

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

	Model Solution	Scoring
(a)	This is an observational study. The researchers had the car owners estimate their mileage. The car owners were not randomly assigned a car model, so no treatment was imposed.	**Essentially correct (E)** if the response satisfies the following three components: 1. Identifies an observational study 2. Provides a justification based on no treatment being imposed 3. Includes context **Partially correct (P)** if the response satisfies only two of the three components required for E. **Incorrect (I)** if the response does not meet the criteria for E or P.

Additional Notes:

- A response that states the study is an experiment receives a score of I for part (a).
- A response that states "No random assignment of treatment" may satisfy component 2.
- A response may satisfy component 3 with car models, drivers, cars, owners, or members. Mileage alone does not satisfy component 3.

	Model Solution	Scoring
(b)	Number the days in the experiment from 1 to 70. Using a random number generator, generate 35 unique integers from 1 to 70, inclusive. Assign the days with those 35 unique integers for James to drive the car with autopilot and assign the remaining 35 days for James to drive the car without autopilot. (Alternative solution) Using 70 equally sized slips of paper, label 35 "with autopilot" and 35 "without autopilot." Mix the slips of paper in a bag. Each day for the 70 days, select a slip of paper (without replacement) to determine the driving method for that day.	**Essentially correct (E)** if the response satisfies the following three components: 1. Creates appropriate labels for the units/treatments 2. Describes how to correctly implement the random process so that every possible random assignment is equally likely 3. The response indicates a random process that results in 35 days assigned to using autopilot and 35 days assigned to not using autopilot **Partially correct (P)** if the response satisfies only two of the three components required for E. **Incorrect (I)** if the response does not meet the criteria for E or P.

Additional Notes:

- For responses that use slips of paper (or marbles or equivalent) to represent treatments
  - To satisfy component 1 some slips must be labeled or assigned to represent autopilot and some labeled or assigned to represent no autopilot (e.g., blue marbles represent autopilot and yellow marbles represent no autopilot).
  - To satisfy component 2 the slips of paper must be mixed/shuffled, and the response must clearly link the treatment selected to a day (e.g., each day James selects a slip to determine the driving method).
  - To satisfy component 3 the response must indicate that there are 35 slips (or marbles of a specific color) for each treatment and the response must indicate that slips of paper (marbles) are selected without replacement.
- For responses that use slips of paper labeled from 1 to 70 (or an equivalent interval)
  - To satisfy component 1 the days must be labeled 1 to 70.
  - To satisfy component 2 the slips of paper must be mixed/shuffled, and the response must clearly link the day number selected to autopilot or no autopilot.
  - To satisfy component 3 the response must indicate that slips of paper are selected without replacement and that 35 days are assigned to each treatment.
- For responses that use a random number generator with days labeled from 1 to 70 (or an equivalent interval)
  - To satisfy component 1 the days must be labeled 1 to 70.
  - To satisfy component 2 the response indicates that the random number generator selects numbers 1 to 70 inclusive, and the response clearly links the day number selected to autopilot or no autopilot.
  - To satisfy component 3 the response must indicate that numbers are selected without repeats and that 35 days are assigned to each treatment.

- For responses that flip a coin for each day (or roll a die and note odd/even, generate a random number 1 or 2, or equivalent)
  - To satisfy component 1 each outcome must be linked to a treatment (e.g., heads equals autopilot, tails equals no autopilot).
  - To satisfy component 2 the response must clearly link the outcome of the coin flip to a day (e.g., each day James flips a coin to determine the driving method). Note: If the response includes a stopping rule (e.g., when 35 days are assigned one treatment, the remaining days are assigned the other treatment), component 2 is not satisfied because this plan increases the probability that the last days will have the same treatment, which does not meet the equally likely random assignment requirement.
  - To satisfy component 3 the response must indicate that 35 days are assigned to each treatment using a stopping rule. If there is no stopping rule, component 3 is not satisfied.
- If a response uses a random number generator or slips of paper with the numbers 1 to 70 and does not initially number the days from 1 to 70, component 1 may be satisfied if the response indicates a link between the number selected and the day (e.g., if 3 is selected, James uses autopilot on the third day).
- Responses that do not use any random process should be scored I. For example, "number the days from 1 to 70 and use autopilot on odd-numbered days."
- Responses that use blocking do not satisfy component 2 because all possible random assignments are not equally likely.
- If the response describes two ways to perform the random assignment, assign the score for the weaker assignment process.

	Model Solution	Scoring
(c)	In order to generalize his findings to all Model D cars in his club, he would need to randomly select Model D cars from the club. He would then need to carry out a similar study using the Model D cars that were randomly sampled from the club.	**Essentially correct (E)** if the response satisfies the following three components: 1. The response indicates that more cars must be sampled 2. The response indicates that random sampling is required for generalization 3. The response is in context, which would include sampling from the population of interest (Model D cars from his club) **Partially correct (P)** if the response satisfies only two of the three components required for E. **Incorrect (I)** if the response does not meet the criteria for E or P.

Additional Notes:

- Component 3 may be satisfied by using "members" of the club or "owners" in the club rather than cars.
- Any discussion of experimental design beyond the sample selection (such as conditions or two-sample design) should be ignored.

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

Question 4: Focus on Probability and Sampling Distributions — 4 points

General Scoring Notes

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

Model Solution	Scoring
(a) Let G represent the number of geodes a player opens until finding a red crystal. G follows a geometric distribution with $p = 0.08$. (i) $\mu = E(G) = \frac{1}{0.08} \approx 12.5$ geodes (ii) $\sigma_G = \frac{\sqrt{1 - 0.08}}{0.08} \approx 11.99$ geodes	**Essentially correct (E)** if the response satisfies at least three of the following four components: 1. In part (a-i) the response correctly calculates the mean 2. In part (a-i) the response provides supporting work for the calculation of the correct mean 3. In part (a-ii) the response correctly calculates the standard deviation 4. In part (a-ii) the response provides supporting work for the calculation of the correct standard deviation **Partially correct (P)** if the response satisfies only two of the four components required for E. **Incorrect (I)** if the response does not meet the criteria for E or P.

Additional Notes:

- An arithmetic or transcription error in a response can be ignored if correct work is shown.
- A response with the mean rounded to 12 or 13 does not satisfy component 1.
- A response with a standard deviation of 12 geodes satisfies component 3.

	Model Solution	Scoring
(b)	(i) $P(Y = 3) = (0.92)^2(0.08) \approx 0.067712$ (ii) $P(Y = 4) = 1 - P(Y = 1 \text{ or } 2 \text{ or } 3)$ $\approx 1 - (0.08 + 0.0736 + 0.067712)$ ≈ 0.778688 *OR* If Conrad opens 4 geodes, then he either finds no red geodes or he finds a red geode on the fourth one he opens; therefore, $P(Y = 4) = (0.92)^4 + (0.92)^3(0.08)$ $\approx 0.778688.$	**Essentially correct (E)** if the response satisfies at least three of the following four components: 1. In part (b-i) the response correctly calculates the probability of opening 3 geodes 2. In part (b-i) the response provides supporting work for the calculation of the correct probability 3. In part (b-ii) the response correctly calculates the probability of opening 4 geodes, consistent with the response to part (b-i) 4. In part (b-ii) the response provides supporting work for the calculation of the correct probability **Partially correct (P)** if the response satisfies only two of the four components required for E. **Incorrect (I)** if the response does not meet the criteria for E or P.

Additional Notes:

- A response may satisfy component 2 or component 4 by the following or a combination of the following:
  - Probability formula: Displaying a correct formula for computing the geometric probability, such as:
    - $(1 - 0.08)^2(0.08)$ or $(0.92)^2(0.08)$ for part (b-i)
  - Calculator function syntax: Using calculator function notation with the correct value of the parameter identified, such as:
    - geompdf$(p = 0.08, x = 3)$ for part (b-i)
    - 1 – geomcdf$(p = 0.08, x = 3)$ for part (b-ii)
    - binompdf$(n = 4, p = 0.08, x = 0)$ + geompdf$(p = 0.08, x = 4)$ for part (b-ii)
    - binompdf$(n = 4, p = 0.92, x = 4)$ + geompdf$(p = 0.08, x = 4)$ for part (b-ii)
- An arithmetic or transcription error in a response can be ignored if correct work is shown.

Model Solution	Scoring
(c) (i) $\mu = E(Y)$ $\approx (1)(0.08) + (2)(0.0736)$ $+(3)(0.0677) + (4)(0.778688)$ $\approx 0.08 + 0.1472 + 0.2031 + 3.1148$ ≈ 3.545 geodes. (ii) The mean of 3.545 geodes is the average number of geodes that result from a long run of many, many trials of opening randomly selected geodes and counting the number opened until either a red geode is found or the fourth geode is opened.	**Essentially correct (E)** if the response satisfies both components 1 and 2 and at least two of components 3–5: 1. In part (c-i) the response states the correct mean of the distribution consistent with values calculated in part (b) 2. In part (c-i) the response shows appropriate work to calculate the mean using the values calculated in part (b) 3. In part (c-ii) the interpretation includes the concept of repeating the selection process over a long period of time 4. In part (c-ii) the interpretation includes the concept of an average or mean 5. In part (c-ii) the interpretation includes the context of number of geodes opened **Partially correct (P)** if the response does not meet the criteria for E but satisfies two or three of components 1–4. **Incorrect (I)** if the response does not meet the criteria for E or P.

Additional Notes:

- Supporting work for finding the expected value must include at least two of the terms in the equation to show the pattern, such as $1(0.08) + 2(0.0736) + \ldots$.
- Calculator notation does not satisfy component 2, such as 1-VAR STATS(L1, L2).
- An arithmetic or transcription error in a response can be ignored if correct work is shown.
- The response $\frac{1}{(.08 + .0736 + .0677 + .0623)} \approx 3.526$ does not satisfy component 1 or 2.
- In part (c-i) if the response has incorrect values in part (b), but uses the values from the correct probability distribution, the response may satisfy components 1 and 2.
- The numerical value of the mean is not required to satisfy components 3–5.

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

Question 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 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)	$P(11+ \text{ months} \cap \text{majority regular cards})$ $= \frac{71 + 76 + 112}{500}$ $= \frac{259}{500}$ $= 0.518$	**Essentially correct (E)** if the response satisfies the following two components: 1. Provides the correct probability 2. Shows work for the correct probability **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.
- A correct fraction, decimal, or percentage may satisfy component 1.
- To satisfy component 2 work must demonstrate how the numerator of 259 was computed. This may be accomplished by showing the addition or clearly indicating the appropriate summands on the table.
- A response that satisfies components 1 and 2 for the probability that the collector has been collecting baseball cards for 11 or more months and has a majority of *rare* baseball cards may be scored P.
- A specific probability statement is not required, but if correctly given should be considered a positive in holistic scoring.

Model Solution	Scoring
(b) $P(\text{majority regular cards} \mid \text{fewer than 6 months})$ $= \frac{P(\text{majority regular cards} \cap \text{fewer than 6 months})}{P(\text{fewer than 6 months})}$ $= \frac{\frac{80}{500}}{\frac{91}{500}}$ $= \frac{80}{91}$ $= 0.879$	**Essentially correct (E)** if the response satisfies the following two components: 1. Provides the correct probability 2. Shows work for the correct probability **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.
- A response of $\frac{80}{91}$ satisfies both components 1 and 2.
- A response that satisfies components 1 and 2 for the probability that given a randomly selected collector has been collecting baseball cards for fewer than six months, they have a majority of *rare* baseball cards may be scored P.
- A specific probability statement is not required, but if correctly given should be considered a positive in holistic scoring.

Model Solution	Scoring
(c) (i) Michelle should conduct a chi-square test for independence between months collecting baseball cards and majority card status for all baseball card collectors at the convention. (ii) H_0: There is not an association between months collecting baseball cards and majority card status for all baseball card collectors at the convention. H_a: There is an association between months collecting baseball cards and majority card status for all baseball card collectors at the convention. *OR* H_0: Months collecting cards and majority card status are independent for all baseball card collectors at the convention. H_a: Months collecting cards and majority card status are not independent for all baseball card collectors at the convention.	**Essentially correct (E)** if the response satisfies the following three components: 1. In part (c-i) the response identifies a chi-square test for independence by name 2. In part (c-ii) the response states the correct null hypothesis to imply there is not an association (relationship) *AND* states the correct alternative hypothesis to imply there is an association (relationship) 3. In part (c-ii) the response provides sufficient context for at least one of the hypotheses by including reference to both "months collecting" and "majority card status" **Partially correct (P)** if the response satisfies only two of the three components required for E *OR* if the response identifies a "chi-square test for homogeneity" in part (c-i) by name or formula *AND* components 2 and 3 are consistent with the chi-square test for homogeneity. **Incorrect (I)** if the response does not meet the criteria for E or P.

Additional Notes

- A response that states merely "chi-square" or "chi-square test," without specifying independence, does not satisfy component 1. However, a response that states "chi-square test of association" does satisfy component 1.
- Any discussion of the degrees of freedom for the test should be ignored in scoring.
- If the hypotheses do not explicitly state the population of interest, it should be assumed that the population is all baseball card collectors at the convention. However, if the response clearly references the sample, component 2 is not satisfied.

Model Solution	Scoring
(d) Because the p-value of 0.0075 is less than any reasonable α level such as 0.05 or 0.10, the null hypothesis should be rejected. The data provide convincing statistical evidence that there is an association between months collecting baseball cards and majority card status for all baseball card collectors at the convention.	**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 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 stated alternative hypothesis in part (c) using nondefinitive language **Partially correct (P)** if the response satisfies only one of the two components required for E. **Incorrect (I)** if the response does not meet the criteria for E or P.

Additional Notes

- Because no alpha value is explicitly given, a response may satisfy component 1 by stating that the p-value is small *AND* providing a correct decision about the null and/or alternative hypothesis.
- 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).
- To satisfy the context in component 2 the response must include discussion of both "months spent collecting baseball cards" and "which type of card is the majority in the collection."
- If the response omits hypotheses in part (c), assume the correct alternative hypothesis is provided when scoring component 2.
- If the response states incorrect hypotheses in part (c), components 1 and 2 may be satisfied by either referring to the stated alternative hypothesis or referring to the inference question.
- Examples of nondefinitive language in component 2 include "evidence to accept the alternative," "there is evidence for the alternative," "there is not sufficient evidence for the alternative," and "Michelle should conclude that her belief is correct."
- Examples of definitive language in component 2 include "proves the null," "proves the alternative," "accepts the alternative," "there is not evidence for the alternative," "no evidence for the alternative," "therefore the alternative is true," and "there is an association."
- 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.

Scoring for Question 5 Each essentially correct (E) part counts as 1 point, and each partially correct (P) part counts as ½ point.	**Score**
Complete Response	**4**
Substantial Response	**3**
Developing Response	**2**
Minimal Response	**1**
If a response is between two scores (for example, 2 ½ points), use a holistic approach to decide whether to score up or down, depending on the strength of the response and quality of the communication.	

Question 6: Investigative Task — 4 points

General Scoring Notes

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

Model Solution	Scoring
(a) (i) 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: 1. In part (a-i) the response identifies a one-sample *t*-interval for a population mean by name (e.g., "one sample *t*-interval") 2. In part (a-ii) the response identifies the correct parameter by including reference to the mean 3. 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:

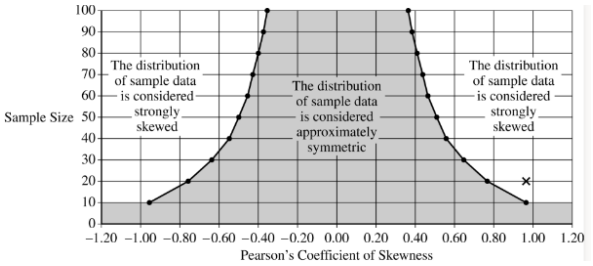
- In part (a-i) a response that refers to a "test" does not satisfy component 1.
- In part (a-i) a response that refers to "a mean," singular, satisfies the one-sample aspect of component 1. If the response states "one-sample," it is not required to also include "for a mean" to satisfy component 1.
- 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: 1. In part (b-i) the response indicates the distribution is skewed to the right or approximately symmetric 2. In part (b-i) the response justifies the response to component 1 with appropriate reasoning based on the summary statistics 3. 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 4. 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. *OR* satisfies components 1 and 2 *OR* 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.

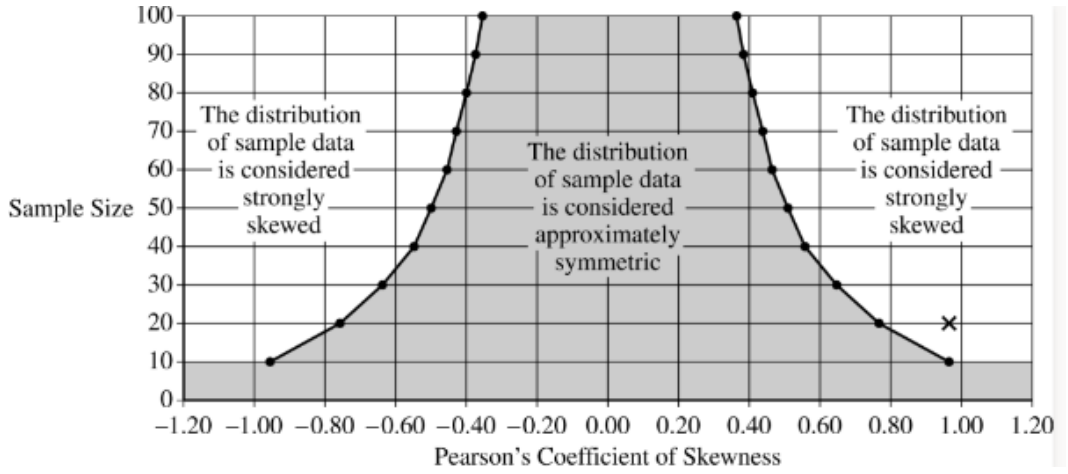
Additional Notes:

- 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.