2021

AP® Statistics

Scoring Guidelines

Question 1: 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)	The five-number summary of the distribution of length of stay is: Minimum = 5 days Lower quartile (Q_1) = 6 days Median = 7 days Upper quartile (Q_3) = 8 days Maximum = 21 days	**Essentially correct (E)** if the response provides correct values for ALL FIVE of the summary statistics with labels (minimum, lower quartile, median, upper quartile, and maximum). **Partially correct (P)** if the response provides correct values for only THREE or FOUR of the summary statistics with labels. **Incorrect (I)** if the response does not meet the criteria for E or P.

Additional Notes:

- Any discussion of the mean, IQR, or the standard deviation of length of stay should be ignored in scoring.
- Inclusion or omission of units of measurement (days) has no bearing on scoring.
- If the response includes exactly 5 unlabeled numbers expressed together as a vertical or horizontal list, interpret the numbers as being labeled as the minimum, lower quartile, median, upper quartile, and maximum, respectively.
- A response that includes only five numbers that are correct values for the five-number summary without providing a complete set of labels or not putting them in an ordered list may be scored P.

Model Solution	Scoring
(b) **(i)** The patients who stayed for 12 days and 21 days are considered outliers using method A. An outlier using method A is a value greater than $1.5 \times \text{IQR}$ above the third quartile (Q_3) or more than $1.5 \times \text{IQR}$ below the first quartile (Q_1). Because $Q_1 - 1.5 \times \text{IQR} = 6 - 1.5(8-6) = 3,$ then any values below 3 are considered outliers. There are no such values. Because $Q_3 + 1.5 \times \text{IQR} = 8 + 1.5(8-6) = 11,$ then any values above 11 are considered outliers. **(ii)** The patient who stayed for 21 days is the only outlier using method B. An outlier using method B is a value located 2 or more standard deviations above, or below, the mean. Because $\text{Mean} \pm 2 \times \text{SD} = 7.42 \pm 2(2.37),$ then any value that is outside of the interval $(2.68, 12.16)$ is considered an outlier.	**Essentially correct (E)** if the response satisfies the following four components: 1. Correctly identifies the two outliers in part (b-i) as the patients who stayed for 12 days and 21 days 2. Provides a justification for part (b-i) by calculating the lower and upper outlier criteria for the $1.5 \times \text{IQR}$ rule (e.g., "using method A, an outlier is any value below 3 days or above 11 days") 3. Correctly identifies the one outlier in part (b-ii) as the patient who stayed for 21 days 4. Provides a justification for part (b-ii) by calculating the lower and upper outlier criteria for the 2 standard deviations rule (e.g., "using method B, an outlier is any value below 2.68 days or above 12.16 days") **Partially correct (P)** if the response satisfies only two or three of the four components. **Incorrect (I)** if the response does not meet the criteria for E or P.

Additional Notes:

- A response for part (b-ii) that manually computes the standard deviation as 2.374 and then uses it to construct an interval of $(2.672, 12.168)$ satisfies component 4.
- Component 1 and component 2 are satisfied if the response to part (b-i) uses correct calculations with incorrect values of summary statistics reported in the response to part (a).

	Model Solution	Scoring
(c)	Quartiles and the IQR are less sensitive to extreme values in strongly skewed distributions than the mean and standard deviation. Relative to the quartiles, the mean is pulled more toward the extreme values in the longer tail of a strongly skewed distribution. For a distribution that is strongly skewed to the right, the sample mean will be pulled more toward the extreme values in the longer right tail of the distribution than the sample median, and the ratio of the standard deviation to the IQR will tend to be larger than that for more nearly symmetric distributions. As a result, this pulls the value of the outlier criterion for method B, $\text{Mean} + 2 \times \text{SD}$, more toward the extreme values in the right tail of the distribution than the outlier criterion for method A, $Q_3 + 1.5 \times \text{IQR}$. This decreases the ability of method B to identify outliers relative to method A, which means that method A may identify more outliers than method B for a distribution that is strongly skewed to the right.	**Essentially correct (E)** if the response satisfies the following two components: 1. Indicates that the mean is pulled more toward the extreme values in the longer right tail for a strongly right-skewed distribution than the quartiles (or median) OR indicates that the ratio of the standard deviation to the IQR tends to be larger for strongly skewed distributions than for more nearly symmetric distributions 2. Provides an explanation that links effects of skewness on an increased ability of method A to detect outliers relative to method B (e.g., "the larger shift in the mean relative to the shift in the median (or quartiles) has a greater effect on decreasing the ability of method B to detect outliers compared to method A" OR "the larger increase in the standard deviation, relative to the IQR, results in a greater increase in the range of non-outlier values for method B compared to method A") **Partially correct (P)** if the response satisfies only one of the two components. **Incorrect (I)** if the response does not meet the criteria for E or P.

Scoring for Question 1	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 2: Focus on Collecting 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) Keeping daily journals could introduce response bias due to the self-reporting by subjects who may have a poor or incomplete memory of the amount of walking that was done. If most subjects who keep daily journals underreport the number of miles walked per day because they cannot remember all of their walking at the end of the day, then the estimate of mean daily miles walked for the target population will be biased too low. Wearing activity trackers would likely provide a more accurate record of daily miles walked by each subject in the study.	**Essentially correct (E)** if the response satisfies the following two components: 1. Indicates that keeping a daily journal could result in a bias that would be avoided by using activity trackers AND provides a reasonable explanation 2. Provides a description of a bias that refers to at least one of the following: • The use of a daily journal may result in a systematic/consistent underreporting, or systematic/consistent overreporting of daily miles walked • The use of a daily journal may result in a biased estimation (underestimation or overestimation) of a population parameter (e.g., mean daily miles walked for the members of the target population) **Partially correct (P)** if the response satisfies only one of the two components. **Incorrect (I)** if the response does not meet the criteria for E or P.

Additional Notes:

- A response does not need to specifically name a type of bias (e.g., response bias).
- The response may refer to the explanatory variable as “activity level.”
- The direction of the bias need not be specified in order to satisfy component 1.
- Examples of reasonable explanations for indicating that keeping a daily journal may result in a bias include:
  - “Because the subjects are self-reporting their daily miles walked.”
  - “Because the subjects may not accurately recall their daily miles walked.”
  - “Because the subjects may forget to complete an entry in their journal.”
- The direction of the bias must be specified in order to satisfy component 2.

- The response must indicate the underreporting or overreporting is systematic across the subjects (or there is a tendency to underreport or overreport) in order to satisfy component 2. Examples of responses that satisfy component 2 include:
    - "The subjects in the study may consistently underreport their daily miles walked."
    - "Subjects are likely to underreport their daily miles walked."
    - "Most subjects may overreport their daily miles walked."
    - "The bias may result in an estimate of the mean daily miles walked by members of the target population that is lower than the target population mean."
- A response that indicates the underreporting or overreporting for only some people does not satisfy component 2 (e.g., "Some people might record higher miles than they actually walk.").

	Model Solution	Scoring
(b)	It is necessary to have a representative sample of subjects from the population in order to make an unbiased inference about the difference between the mean cholesterol levels for all adult members of the target population who walk fewer miles per day and the mean cholesterol levels for all adult members of the target population who walk more miles per day.	**Essentially correct (E)** if the response satisfies the following two components: 1. Provides an explanation that the use of a representative sample is necessary in order to make a valid generalization about the target population 2. Refers to estimation, or inference, for cholesterol levels in the target population OR an association between cholesterol level and amount of walking in the target population **Partially correct (P)** if the response satisfies only one of the two components. **Incorrect (I)** if the response does not meet the criteria for E or P.

Additional Notes:

- A response that discusses the accuracy or validity of a significance test does not satisfy component 1 unless the response makes it clear that the inference is being generalized to the target population**.**
- In order to satisfy component 2, the response need not state a specific population parameter(s).
- If a parameter is specified, it must be relevant to cholesterol level or the association between cholesterol level and amount of walking. Some examples include:
  - Individual population mean cholesterol level
  - One or more differences between population mean cholesterol levels
  - Individual population median cholesterol level
  - One or more differences between population median cholesterol levels
  - A population correlation between cholesterol level and amount of walking
  - A population regression model for cholesterol level and amount of walking

Model Solution	Scoring
(c) No, since the treatments (amounts of walking) were not randomly assigned to the subjects in the study, it would not be valid to claim that increased walking causes a decrease in average cholesterol levels for adults in the target population. The researchers would only be able to conclude that cholesterol level has a negative association with daily miles walked for adults in the target population. There may be one or more confounding variables that are the actual cause of the relationship. For example, people who walk more may be more concerned about maintaining a healthy diet and eat more foods that are low in cholesterol, while people who walk less may eat more foods that are high in cholesterol. Consequently, the association between cholesterol and daily miles walked could actually be caused by differences in diets and not differences in amount of walking.	**Essentially correct (E)** if the response satisfies the following two components: 1. Indicates that a causal inference cannot be made 2. Provides a valid explanation that is based on one of the following: • the lack of (random) assignment of treatments to subjects • being an observational study/not an experiment • the existence of a possible confounding variable that is associated with amount of walking and associated with cholesterol level **Partially correct (P)** if the response satisfies only component 1 AND provides a weak explanation. **Incorrect (I)** if the response does not meet the criteria for E or P.

Additional Notes:

- A response that provides an explanation that is based on the existence of a possible confounding variable may or may not identify a specific confounding variable. In either case, the response must indicate that the confounding variable has an association with amount of walking AND also indicate that the confounding variable has an association with cholesterol level in order to satisfy component 2. Examples of responses that satisfy component 2:
  - A response that identifies a reasonable confounding variable: "Diet could be a confounding variable. People who walk more may tend to eat more foods that are low in cholesterol, while people who walk less may tend to eat more foods high in cholesterol."
  - A response that does not identify a confounding variable: "There could be a confounding variable that has an association with cholesterol level and also has an association with amount of walking."
- If a response identifies a specific confounding variable, then any variable that is reasonable (e.g., diet, weight, body mass index, etc.) should be accepted in scoring component 2.
- In component 2, the following are examples of weak explanations:
  - The response indicates the existence of a confounding variable but does not indicate that the confounding variable has an association with amount of walking AND an association with cholesterol level.
  - The response communicates that an association between cholesterol level and amount of walking does not imply that there is a causal relationship between cholesterol level and amount of walking. However, a general statement, without context, that association does not imply causation should be scored incorrect (I).

- A response that only references specific elements of an experiment (e.g., placebo, control group, replication) aside from assignment of treatments to subjects should be scored incorrect (I).
- A response that states that a causal relationship can be concluded due to the statistically significant result and goes on to say that there may be a confounding variable that is associated with amount of walking and cholesterol level (e.g., diet) should be read as parallel solutions and scored incorrect (I).
- Responses in parts (a) or (b) cannot be carried down to part (c) to satisfy component 2 unless the response in part (c) refers to specific statements in part (a) or (b).

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 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) (i) Let the random variable of interest X represent the number of gift cards that a particular employee receives in a 52-week year. Because each employee has probability $\frac{1}{200} = 0.005$ of being selected each week to receive a gift card and each week's selection is independent from every other week, X has a binomial distribution with $n = 52$ repeated independent trials and probability of success $p = 0.005$ for each trial. **(ii)** The probability that a particular employee receives at least one gift card in a 52-week year is: $P(X \geq 1) = 1 - P(X = 0)$ $= 1 - \binom{52}{0}(0.005)^0(0.995)^{52}$ $= 1 - 0.7705$ $= 0.2295$	**Essentially correct (E)** if the response satisfies the following four components: 1. Defines the random variable as the number of gift cards that a particular employee receives in a 52-week year 2. Describes the distribution as binomial with $n = 52$ and $p = 0.005$ 3. Identifies the event of interest (i.e., identify the correct boundary AND direction for the event) in the calculation of the probability in part (a-ii) 4. Provides supporting work to identify the correct probability of 0.2295 (or 0.230, if rounded) OR a probability consistent with components 2 and 3 **Partially correct (P)** if the response satisfies only two or three of the four components. **Incorrect (I)** if the response does not meet the criteria for E or P.

Additional Notes:

- A response that states $X \sim B(52, 0.005)$ satisfies component 2.
- A response that states the random variable is distributed by a distribution that is not binomial (e.g., normal or uniform) and then uses the binomial calculation does not satisfy component 2.
- Stating that gift cards are distributed randomly is not a distribution and does not, in itself, satisfy component 2. Component 2 can still be satisfied if the response goes on to use the binomial distribution.
- In order to satisfy component 2 using calculator function notation, the sample size and probability parameter must be clearly identified.
  - The following satisfy component 2:
    - binomcdf(n or trials = 52, $p = 0.005$, 1, 52)

    - 1 – binomcdf(n or trials = 52, $p = 0.005$, 0)
    - 1 – binompdf(n or trials = 52, $p = 0.005$, 0)
  - The following do not satisfy component 2 because the parameter or sample size is not clearly labeled:
    - binomcdf(52, 0.005, lower bound = 1, upper bound = 52)
    - 1 – binomcdf(52, $p = 0.005$, upper bound or $x = 0$)
    - 1 – binompdf(n or trials = 52, 0.005, $x = 0$)
- In order to satisfy component 3, the supporting work must identify the event of interest, i.e., $X \geq 1$, the boundary is 1, and the direction is greater than or equal to, or at least.
  - Possible ways to do this include:
    - Probability notation, e.g. $P(X \geq 1)$, $1 - P(X = 0)$
    - Summing probabilities, e.g. $\sum_{k=1}^{52} \binom{52}{k} (0.005)^k (0.995)^{52-k}$
    - Description in words $P(\text{employee receives at least one gift card})$, $1 - P(\text{employee receives no gift cards})$
    - Graphical, a bar graph of binomial probabilities with appropriate bars shaded
    - Using calculator function syntax with clearly labeled parameters (e.g. $p = 0.005$, $n = 52$) and clearly labeled event boundaries (e.g., lower bound = 1, upper bound = 52)
  - The following satisfy component 3:
    - binomcdf(n or trials = 52, $p = 0.005$, lower bound=1, upper bound = 52)
    - 1 – binomcdf(n or trials = 52, $p = 0.005$, upper bound or $x = 0$)
    - 1 – binompdf(n or trials = 52, $p = 0.005$, $x = 0$)
  - The following do not satisfy component 3 because the event boundaries are not clearly labeled.
    - binomcdf(n or trials = 52, $p = 0.005$, 1, 52)
    - 1 – binomcdf(n or trials = 52, $p = 0.005$, 0)
    - 1 – binompdf(n or trials = 52, $p = 0.005$, 0)
- A response will satisfy component 3 if the probability is computed for a geometric distribution with the first success within the first 52 weeks ($X \leq 52$) (e.g. response that states geometcdf(0.005, x or upper bound = 52)).
- Because $np = (52)(0.005) = 0.26$ is less than 5, the normal approximation to the binomial distribution is not an appropriate method to calculate the probability, and a response that uses this method does not satisfy component 4. However, a response that uses the normal approximation to the binomial distribution may satisfy component 3 if it displays the correct mean and standard deviation of the binomial distribution AND provides a clear indication of the appropriate collection of possible outcomes included in the event using a diagram or a ***z***-score, e.g., $1 - P\left(Z \leq \frac{0 - (52)(0.005)}{\sqrt{(52)(0.005)(0.995)}}\right)$, $P\left(Z \geq \frac{1 - (52)(0.005)}{\sqrt{(52)(0.005)(0.995)}}\right)$, or $P\left(Z \geq \frac{0.5 - (52)(0.005)}{\sqrt{(52)(0.005)(0.995)}}\right)$.

Model Solution	Scoring
(b) The expected value for the number of gift cards a particular employee will receive in a 52-week year is $np = 52(0.005) = 0.26$. If the random process of selecting one employee each week to receive a gift card is repeated for a very large number of years, each employee can expect to receive about 0.26 gift cards per year, on average, or about one gift card every four years.	**Essentially correct (E)** if the response satisfies the following two components: 1. Correctly calculates the expected value AND provides supporting work for the calculation of the expected value 2. Provides a reasonable interpretation of the expected value that includes *at least two* of the following three aspects: • The concept of repeating the selection process over a long period of time • The concept of an average or mean • The context of receiving gift cards **Partially correct (P)** if the response satisfies only one of the two components. **Incorrect (I)** if the response does not meet the criteria for E or P.

Additional Notes:

- A response may satisfy component 1 if the reported expected value is consistent with the distribution of the random variable identified in the response to part (a-i), AND supporting work for the calculation of the expected value in part (b) is shown.
- Examples of supporting work that satisfies component 1 include:
  - $np = 52(0.005) = 0.26$ or $np = 52(0.005)$
  - $np = \frac{52}{200}$
  - $52(0.005) = 0.26$
  - $np = 0.26$, if the values of n and p are reported in the response to part (a)
- A response that incorrectly calculates the expected value may still satisfy component 2 using the incorrect expected value in the interpretation.

	Model Solution	Scoring
(c)	No, Agatha's experience does not constitute strong evidence that the selection process was not truly random. In fact, it is quite likely $(\text{probability} = (0.995)^{52} \approx 0.7705)$ that a particular employee will fail to receive a gift card for an entire 52-week year.	**Essentially correct (E)** if the response satisfies the following three components: 1. Indicates that Agatha does not have a strong argument that the selection process was not truly random 2. Provides a relevant probability or expected value 3. Provides an explanation that correctly links the probability or expected value to the decision **Partially correct (P)** if the response satisfies only two of the three components. **Incorrect (I)** if the response does not meet the criteria for E or P.

Additional Notes:

- Examples that satisfy component 2:
    - The probability that Agatha will receive at least one gift card in a 52-week year is 0.2295, or the value computed in part (a-ii).
    - The probability that Agatha will fail to receive a gift card for an entire 52-week year is 0.7705, or the complement of the value computed in part (a-ii).
    - The expected value computed in part (b).
    - Stating AT MOST 52 out of 200 employees will win a gift card (or AT LEAST 148 will not win).
- A response that indicates that Agatha does have a strong argument that the selection process was not truly random (or responds "yes") that is adequately supported by an explanation based on an incorrectly calculated probability in part (a-ii) OR an incorrectly calculated expected value in part (b) is scored E.
- If a response gives two arguments, treat them as parallel solutions and score the weaker solution.

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 Inference — 4 points

General Scoring Notes

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

	Model Solution	Scoring
(a) **Section 1**	Let p represent the proportion of all customers of the pet supply company who would place an order within 30 days after receiving an e-mail with a coupon for $10 off the next purchase. The null hypothesis is $H_0: p = 0.40$, and the alternative hypothesis is $H_a: p > 0.40$. An appropriate test is a one-sample *z*-test for a population proportion.	**Essentially correct (E)** if the response satisfies the following three components: 1. States the correct equality for the null hypothesis for a proportion (e.g., $p = 0.40$) AND the correct direction of the one-sided alternative hypothesis for a proportion (e.g., $p > 0.40$) 2. Provides sufficient context for the parameter by including reference to the *population* proportion AND the sampling units (customers) AND the response variable (placing an order after receiving a coupon) 3. Identifies a one-sample *z*-test for a population proportion by name (e.g., "one-proportion *z*-test" but <u>not</u> merely "one-sample *z*-test") or by formula **Partially correct (P)** if the response does not meet the criteria for E but satisfies either component 1 and/or component 3. **Incorrect (I)** if the response does not meet the criteria for E or P.

Additional Notes:

- The elements of component 2 do not have to be satisfied with the statement of the hypotheses. They may be satisfied by work presented anywhere in the response, most likely by the statement of the conclusion.
- If the statement of the hypotheses refers to population proportion and the conclusion refers to sample proportion (or vice versa), then the population aspect of component 2 is not satisfied.

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

Confidence Interval Approach:

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

Model Solution		Scoring
(a) **Section 2**	The independent observations condition for performing the one-sample z-test for a population proportion is satisfied because the data were obtained from a random sample of 90 customers who placed an order in the past year and, because sampling of customers is done without replacement, it is assumed that this large online company has more than $10(90) = 900$ customers. The sample size is large enough to support an assumption that the sampling distribution of $\hat{p}$ is approximately normal because $(90)(0.4) = 36$ and $(90)(1-0.4) = 54$ are both at least 10. The value of the sample proportion is $\hat{p} = \frac{38}{90} \approx 0.422$ and the value of the test statistic is $z = \frac{\frac{38}{90} - 0.40}{\sqrt{\frac{(0.40)(0.60)}{90}}} \approx 0.430$. The corresponding p-value is $P(z > 0.430) \approx 0.333$.	**Essentially correct (E)** if the response satisfies the following four components: 1. Checks the independence condition by referring to the random selection of 90 customers AND indicating that the company is assumed to have at least 900 customers (i.e., $90 \le 0.10N$) 2. Checks that the sample size is large enough to support the assumption that the sampling distribution of $\hat{p}$ is approximately normal by verifying that $(90)(0.4)$ and $(90)(1-0.4)$ are both at least 10 (or 5) 3. Correctly reports the value of the z-statistic 4. Correctly reports the p-value, consistent with the reported test statistic and stated alternative hypothesis **Partially correct (P)** if the response satisfies only two or three of the four components. **Incorrect (I)** if the response does not meet the criteria for E or P.

Additional Notes:

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

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

Confidence Interval Approach:

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

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

Additional Notes:

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

Confidence Interval Approach:

- If the alternative hypothesis is specified correctly as $H_a: p > 0.40$, then component 1 is satisfied if the justification is based on whether 0.40 is below the lower end of the confidence interval. If the alternative hypothesis is stated in the wrong direction, then component 1 is satisfied if the justification is based on whether 0.40 is above the upper end of the confidence interval.

- If no alternative hypothesis is specified in the response, then assume the correct alternative hypothesis is provided when scoring component 2.
- If an incorrect two-sided alternative hypothesis is specified, then component 2 is satisfied if the justification is based on whether 0.40 is included in the confidence interval.
- If the response includes an incorrect interpretation of the confidence interval, then the score for Section 3 is lowered one level (that is, from E to P or from P to I).

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

Additional Notes:

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

Scoring for Question 4 Each essentially correct (E) part counts as 1 point, and each partially correct (P) part counts as 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 5: Multi-Focus — 4 points

General Scoring Notes

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

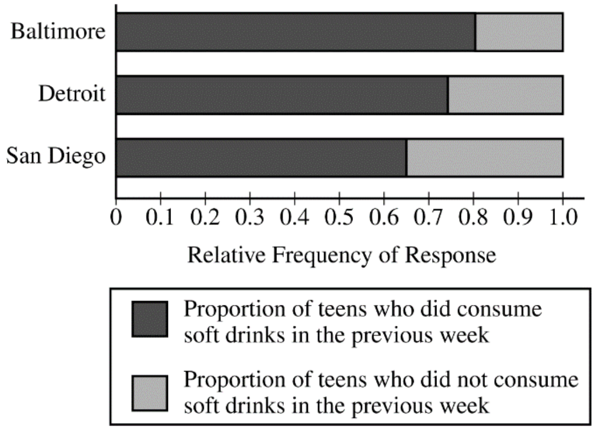
Model Solution	Scoring
(a) No, the researcher's claim is not correct. Although the Baltimore survey has the least number of teens who consumed a soft drink in the past week, it also has the least number of teens surveyed among the three cities' samples. Comparing the numbers of teens who consumed a soft drink in the past week is meaningless without considering the sample sizes. The comparison should be based on proportions rather than counts. In fact, the proportion of Baltimore teens who consumed a soft drink in the past week, $\frac{727}{904} \approx 0.804$, is larger than the proportions for either of the other two cities, $\frac{1,232}{1,663} \approx 0.741$ for Detroit and $\frac{1,482}{2,280} = 0.65$ for San Diego.	**Essentially correct (E)** if the response satisfies the following two components: 1. Indicates that the researcher's claim is not correct (or "may not be correct", if proportions are not reported) 2. Provides an explanation that is based on at least one of the following: • The proportions (or relative frequencies), not counts, should be compared because the sample sizes are not equal *OR* • The proportion of Baltimore teens who consumed a soft drink in the past week, $\frac{727}{904} \approx 0.804$, is larger than the proportion for at least one of the two other cities **Partially correct (P)** if the response satisfies component 1 AND states that the sample sizes are not equal *OR* if the response satisfies component 2 only *OR* if the response provides a correct proportion (or relative frequency) for Baltimore and at least one other city. **Incorrect (I)** if the response does not meet the criteria for E or P.

Additional Notes:

- A response that compares the proportion of Baltimore teens who consumed a soft drink in the past week to the combined proportion of the other two cities, $\frac{1,323 + 1,482}{1,663 + 2,280} = \frac{2,714}{3,943} \approx 0.688$ is not equivalent to comparing counts and may earn a P.
- If the "yes" count for each city is divided by the same number (e.g., the total number of respondents who consumed a soft drink in the past week, 3,441; or the total sample size, 4,847), then the response is equivalent to comparing counts and should be scored I.
- A response may satisfy component 2 by providing correct numerical values for the proportion of Baltimore teens who did *not* consume a soft drink in the past week, $\frac{177}{904} \approx 0.196,$ AND the proportion of teens who did *not* consume a soft drink in the past week for at least one of the two other cities, either $\frac{431}{1,663} \approx 0.259$ for Detroit and/or $\frac{798}{2,280} = 0.35$ for San Diego.
- If work is shown, calculation or transcription errors should be ignored in scoring.
- Statistical notation should be ignored in scoring.

Model Solution

(b) **(i)** A segmented bar graph of the relative frequencies based on the information in the table is shown below:



(ii) The proportion of teens who consumed a soft drink in the previous week are shown below:

- Baltimore: $\frac{727}{904} \approx 0.804$
- Detroit: $\frac{1,232}{1,663} \approx 0.741$
- San Diego: $\frac{1,482}{2,280} = 0.65$

San Diego has the smallest proportion of teens (0.65) who consumed a soft drink in the previous week.

Scoring

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

1. Constructs a segmented bar graph in part (b-i), with the bars correctly segmented
2. Includes clear labeling of the proportions of teens who consumed a soft drink in the previous week and the proportions of teens who did not consume a soft drink in the previous week for the segmented bar graph provided in part (b-i)
3. Identifies San Diego as the city with the smallest proportion of teens who consumed a soft drink in the previous week in part (b-ii)
4. Reports the correct numerical value of the proportion of teens who consumed a soft drink in the previous week for the city identified in part (b-ii)

Partially correct (P) if the response satisfies only two or three of the four components.

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

Additional Notes:

- A response that constructs a segmented bar graph with the lengths of the segments representing the relative frequencies of teens who consumed a soft drink in the previous week between 0.75 and 0.85 for Baltimore, between 0.7 and 0.8 for Detroit, and between 0.6 and 0.7 for San Diego satisfies component 1.
- A response that constructs a segmented bar graph with the lengths of the segments representing the relative frequencies of teens who did not consume a soft drink in the previous week between 0.15 and 0.25 for Baltimore, between 0.2 and 0.3 for Detroit, and between 0.3 and 0.4 for San Diego satisfies component 1.
- Incorrect proportions imported from part (a) may be used to satisfy component 1.
- Segmented bar graphs with more than two segments cannot satisfy either component 1 or component 2.
- Labels of "Yes" and "No" may satisfy component 2.
- A response to part (b-ii) that is consistent with an incorrect graph in (b-i) may satisfy components 3 and 4.

Model Solution	Scoring
(c) **(i)** Since the data were collected from independent random samples from the three cities, a chi-square test for homogeneity should be conducted. **(ii)** The appropriate hypotheses are: H_0: There is no difference in the proportion of all teens who consumed a soft drink in the past week across the three cities. H_a: There is at least one difference in the proportion of all teens who consumed a soft drink in the past week across the three cities. *OR* H_0: The proportion of all teens who consumed a soft drink in the past week is the same across the three cities. H_a: The proportion of all teens who consumed a soft drink in the past week differs for at least two of the three cities.	**Essentially correct (E)** if the response satisfies the following three components: 1. Identifies a chi-square test for homogeneity by name in part (c-i) 2. States the null hypothesis to imply homogeneous (or equal) proportions AND states the alternative hypothesis to imply that at least two proportions are not the same in part (c-ii) 3. Provides sufficient context for at least one of the hypotheses in part (c-ii) by including the parameters of interest (proportion of teens who consumed a soft drink) AND the populations (cities) **Partially correct (P)** if the response satisfies component 1 and only one of the other two components *OR* if the response identifies a "chi-square test" in part (c-i) by name or formula AND satisfies component 2. **Incorrect (I)** if the response does not meet the criteria for E or P.

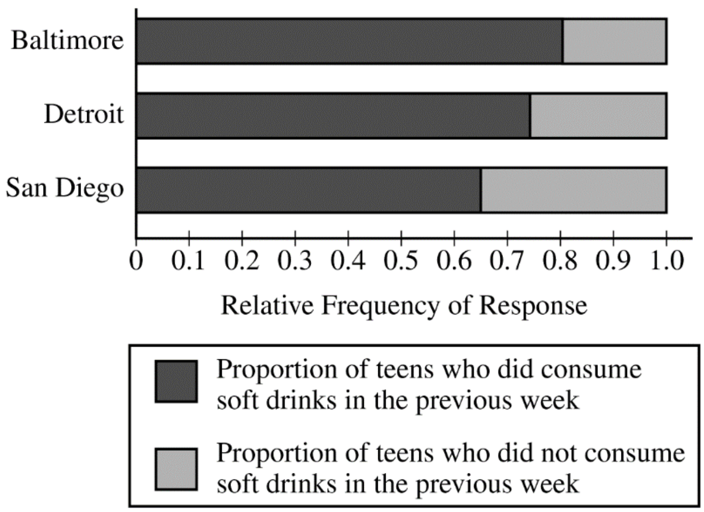
Additional Notes:

- A response that identifies two different tests is considered parallel solutions and the weaker solution is used when scoring component 1.
- Component 1 is not satisfied by the test statistic formula for a chi-square test unless the response includes "Homogeneity."
- Component 1 is not satisfied if the response presents a test statistic formula that is inconsistent with a chi-square test of homogeneity, even if the response identifies a chi-square test of homogeneity by name.
- A response that states the hypotheses in terms of distributions rather than proportions (e.g., H_0: There is no difference in distributions of teens who consumed or did not consume a soft drink in the past week across the three cities) satisfies component 2.
- A response that states either the null hypothesis or the alternative hypothesis by referring to sample proportions does not satisfy component 2.
- A response that uses symbols to describe the hypotheses must clearly identify the parameters in context (proportion of teens who consumed a soft drink) AND reference the populations (cities) in order to satisfy component 3.
- Any attempt to check test conditions should be ignored in scoring.
- Any discussion of the degrees of freedom for the test should be ignored in scoring.

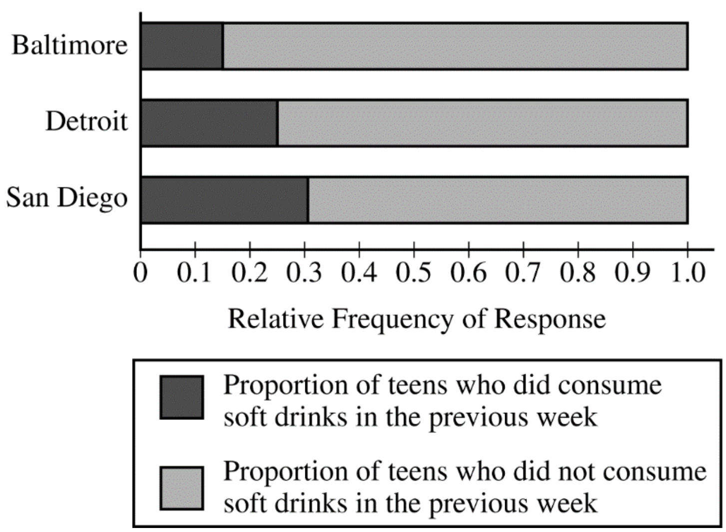
Scoring for Question 5	Score
Complete Response Three parts essentially correct	**4**
Substantial Response Two parts essentially correct and one part partially correct	**3**
Developing Response Two parts essentially correct and no part partially correct *OR* One part essentially correct and one or two 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**

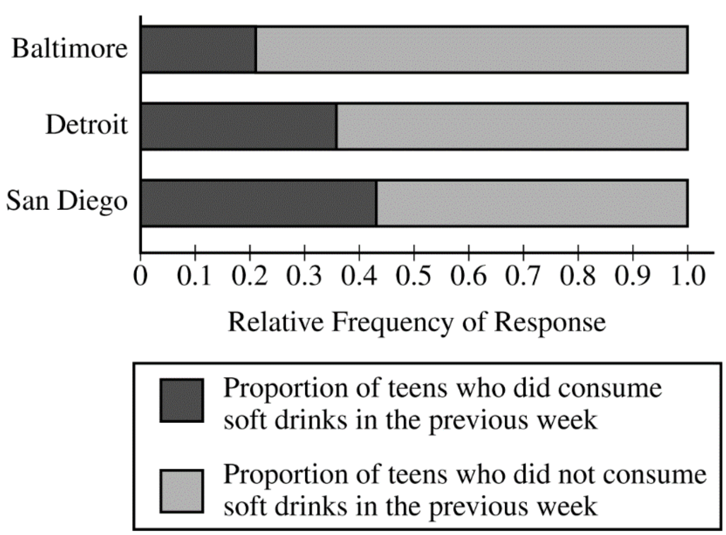
Common acceptable and unacceptable graphs for part (b-i)

Acceptable Graph

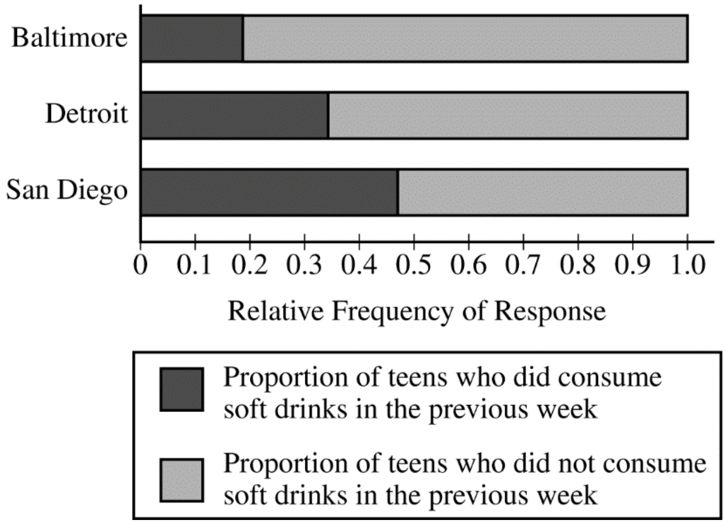


Common Unacceptable Graphs





Common Unacceptable Graphs (continued)



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) The boxplots reveal that the team tended to have a higher average per-game attendance during years in the new stadium than during years in the old stadium because the median value of about 25,000 attendees per game during years in the new stadium is greater than the median value of about 16,000 attendees per game during years in the old stadium. The interquartile ranges (IQRs) are similar, which indicates that variability in average per-game attendance is similar during the years in the two stadiums, but the range of average per-game attendance is slightly larger during the years in the new stadium. There are no years with average per-game attendance identified as an outlier for either stadium.	**Essentially correct (E)** if the response satisfies the following three components: 1. Indicates that the median average per-game attendance is greater in the new stadium *OR* Indicates that the average per-game attendance is usually (typically) greater in the new stadium 2. Indicates that the variability in average per-game attendance is greater in the new stadium *OR* Indicates that the variability in the two stadiums is roughly the same according to the IQRs 3. Provides sufficient context by including the response variable (average attendance or attendance) or the units of the response variable (thousands of people or people) **Partially correct (P)** if the response satisfies only two of the three components *OR* if the response satisfies only component 1 *OR* if the response satisfies only component 2. **Incorrect (I)** if the response does not meet the criteria for E or P.

Additional Notes:

- To satisfy component 1, the response may refer only to “center” and does not need to specify which measure of center (e.g., median, mean) is being used.

- To satisfy component 2, a response may refer only to "variability" or "spread" and does not need to specify which measure of variability (e.g., range or IQR) is being used. However, if the response states that the variability is about the same, the response must explicitly refer to the IQRs.
- To satisfy components 1 and 2, the stadiums must be identified (e.g., "old," "new"), and an explicit comparison phrase (e.g., "greater than," "about the same as") must be used. Separate lists of characteristics alone or summary statistics alone do not count as a comparison.
- To satisfy components 1 and 2, numerical values are not required. However, if they are included, they should be reasonably correct. Numerical values can be reported in units of people (e.g., median = 16,000) or in thousands of people (e.g., median = 16).
- Any mention of shape is ignored in the scoring of part (a) because complete shape information cannot be obtained from a boxplot. Statements about shape not clearly supported by the boxplots (e.g., "the old stadium distribution is approximately normal") should be considered a negative in terms of holistic scoring. However, statements about shape that are supported by the boxplots (e.g., "the old stadium distribution is roughly symmetric," "the new stadium distribution is skewed to the left") should be considered a positive.

Model Solution	Scoring
(b) During the years in the new stadium, the average per-game attendance increases linearly, from about 16,000 people in 2000 to about 27,000 people in 2016. However, during the years in the old stadium, there is no obvious increasing or decreasing trend over time for the average per-game attendance. The average attendance appears to vary about an average of approximately 16,000 attendees per game from 1970 to 1999.	**Essentially correct (E)** if the response satisfies the following three components: 1. Describes the direction of the trend in average per-game attendance in the new stadium as increasing (positive) 2. Describes the direction of the trend in average per-game attendance in the old stadium as relatively constant (e.g., “no association,” “flat”) *OR* Describes the direction of the trend in the old stadium as positive but less steep (e.g., “less positive,” “flatter”) than the trend in the new stadium 3. Provides sufficient context by including the two groups (old stadium, new stadium) AND the explanatory variable (time or year) AND the response variable (average attendance or attendance) or the units of the response variable (thousands of people or people) **Partially correct (P)** if the response satisfies only two of the three components **Incorrect (I)** if the response does not meet the criteria for E or P.

Additional Notes:

- Only describing an association or correlation as “strong” or “weak” addresses strength and not the direction of the trend and does not satisfy components 1 or 2.
- Only describing an association as “linear” or “non-linear” addresses form and not the direction of the trend and does not satisfy components 1 or 2.
- Numerical values, including years (e.g., “from 2000 to 2016”), are not required for any component. However, a response that includes years in numerical form (e.g., “1970”) satisfies the context requirement for the explanatory variable in component 3.
- Component 1 can also be satisfied if the response provides an estimated value for the correlation for the new stadium that is positive.
- Component 2 can also be satisfied if the response provides an estimated value for the correlation for the old stadium of 0 (or approximately 0). Providing positive correlations for both stadiums does not satisfy component 2 because it is impossible to compare the steepness of the trends using their correlations.

Model Solution	Scoring
(c) **(i)** Graph I indicates a strong, positive, linear relationship between average per-game attendance and the number of games won during the 47 years of the team's existence. Average per-game attendance increases linearly, with an average increase of about 500 attendees per game for each additional game won. Variation about the linear trend in attendance is relatively small and about the same for any number of games won. **(ii)** No. Graph II suggests that the rates at which average per-game attendance increases as the number of games won increases are about the same for the two stadiums. A line drawn through the points for the old stadium has about the same slope as (or may have a slightly larger slope than) a line drawn through the points for the new stadium.	**Essentially correct (E)** if the response satisfies four or five of the following five components: 1. In part (c-i) describes the direction of the relationship as positive 2. In part (c-i) describes the form of the relationship as linear or nearly linear 3. In part (c-i) describes the strength of the relationship as very strong, strong, or moderately strong 4. In part (c-ii) indicates that the rates are about the same for the two stadiums (or slightly larger for the old stadium) 5. In part (c-ii) provides an explanation that indicates that if a line were drawn through the points for the old stadium, the slope would be roughly the same (or slightly greater than) the slope of a line through the points for the new stadium **Partially correct (P)** if the response satisfies only three of the five components. **Incorrect (I)** if the response does not meet the criteria for E or P.

Additional Notes:

- A response that provides an estimated value of the correlation satisfies component 1 if the estimated correlation is positive, but an estimated correlation cannot satisfy components 2 or 3.
- A response need not include the word "positive" to satisfy component 1. For example, "the average attendance is higher when the team has more wins" satisfies component 1. Likewise, a response need not include the word "strong" to satisfy component 3. For example, "variation about the linear trend in attendance is relatively small" satisfies component 3.
- Correct comments on homogeneous variation in part (c-i) (i.e., the variation about the linear trend in attendance is about the same for any number of games won) should be considered a positive in terms of holistic scoring.
- Responses that satisfy all 5 components should be considered a positive in terms of holistic scoring.
- Context is not required in part (c) because it has already been assessed in parts (a) and (b).

Model Solution	Scoring
(d) The number of games won could be a confounding variable for assessing the potential effect of opening the new stadium on average per- game attendance. The boxplots in part (a) show that average per-game attendance tended to be higher for games in the new stadium than for games in the old stadium, but the cause of the increase may actually be that attendees were more excited about attending games for teams that were better at winning. The scatterplots in part (c) show that average per-game attendance has a strong positive correlation with games won, and the team tended to win more games playing in the new stadium than in the old stadium.	**Essentially correct (E)** if the response provides an explanation that satisfies the following four components: 1. States that there is an association between attendance and one of the explanatory variables (stadium, year, wins) 2. States that there is an association between attendance and a different one of the explanatory variables (stadium, year, wins) 3. States that there is an association between the two explanatory variables (stadium, year, wins) identified in components 1 and 2 4. Explains the idea of confounding by describing that the variable identified as a potential confounding variable could be the cause of the association between attendance and the other explanatory variable identified in components 1 and 2 *OR* Explains the idea of confounding by stating that it is impossible to know which of the two explanatory variables identified in components 1 and 2 may be the cause of the increase in attendance **Partially correct (P)** if the response satisfies only three of the four components *OR* if the response satisfies only two of the four components and justifies at least one of the statements in components 1 through 3 by referring to the appropriate graph from parts (a) through (c) (e.g., "based on the boxplots," "in part (b)"). **Incorrect (I)** if the response does not otherwise meet the criteria for E or P.

Additional Notes:

- The response can use any combination of 2 of the 3 explanatory variables (stadium and wins, stadium and year, year and wins). The response cannot introduce a new variable (e.g., weather, having popular players) to satisfy any component.
- An incorrectly described association cannot be used to satisfy components 1 through 3.
- To satisfy component 4 the response must discuss all three variables: the response variable (attendance) and the two explanatory variables from components 1 and 2.

- The strength of the response to part (d), especially communication in component 4, can be considered in holistic scoring. A well-communicated response to part (d) should be considered a positive. A poorly-communicated response to part (d), should be considered a negative.

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.