Question 2: 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	Sampling method I is not an appropriate sampling method for the farmer to use to estimate the proportion of cabbage plants in the field that are affected by aphids. Sampling method I is a convenience sample where region 3 is not selected randomly. If the farmer's belief is correct, there may be fewer cabbage plants that are affected by aphids in region 3 than in most other regions of the cabbage field because region 3 is in the row farthest from the river. This may lead to an underestimate of the proportion of cabbage plants in the field that are damaged by aphids.	**Essentially correct (E)** if the response satisfies the following three components: 1. Indicates sampling method I is not an appropriate sampling method to obtain the estimate 2. Provides an explanation that refers to at least one of the following: • The region is likely not representative of the entire cabbage field. • The sampling process does not include random selection. • The sample is a convenience sample. 3. Provides sufficient context by including any two of the following: • Sampling unit (region) • Population (cabbage field or the 25 regions in the field) • Statistic or parameter (proportion of cabbage plants that are damaged by aphids) OR count (number of cabbage plants that are damaged by aphids) **Partially correct (P)** if the response satisfies component 1 AND either component 2 or component 3 *OR* if the response does not indicate whether or not sampling method I is an appropriate sampling method AND satisfies both components 2 and 3. **Incorrect (I)** if the response does not meet the criteria for E or P.

Scoring Notes:

References only to "a field" are sufficient to satisfy the population description in component 3.

	Model Solution	Scoring
B	The selection of row E is likely to provide an overestimate of the proportion of all cabbage plants in the field that are damaged by aphids. If the farmer's belief that the extent of aphid damage is greater for the regions in the cabbage field closer to the river is correct, then row E, which is the row of regions located closest to the river, is likely to have a greater proportion of cabbage plants damaged by aphids than regions farther from the river.	**Essentially correct (E)** if the response satisfies the following three components: 1. Indicates that the selection of row E is likely to produce an overestimate 2. Provides a justification that is based on the location of row E as the row located closest to the river 3. Links the location to why row E is likely to produce an overestimate (e.g., by referring to the farmer's belief that the extent of aphid damage is greater for the regions in the cabbage field closer to the river) **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.

Scoring Notes:

A response that indicates the selection of row E is likely to provide an underestimate may be scored P if the response provides a justification that ignores the farmer's belief (either by omission or by specific mention) but includes a valid reason for why the regions further from the river may have more aphid damage.

Model Solution	Scoring
C The farmer should write the region numbers from row A, 1 through 5, onto same-size slips of paper, then put the numbers into a hat, mix well, and select one of the numbers. The farmer should repeat this process for the region numbers of each of the other rows (i.e., row B, 6 through 10; row C, 11 through 15; row D, 16 through 20; row E, 21 through 25) and select one number from each row. This process will result in the selection of one region from each row. The farmer will examine every cabbage plant in each of the selected regions for aphid damage to determine the proportion of cabbage plants in the selected regions that are damaged by aphids. *Alternative Solution:* The farmer should use a random number generator to generate one two-digit integer from 01 to 05, one two-digit integer from 06 to 10, one two-digit integer from 11 to 15, one two-digit integer from 16 to 20, and one two-digit integer from 21 to 25. For each integer selected, the farmer should select the corresponding numbered region and examine every cabbage plant in each of the selected regions for aphid damage to determine the proportion of cabbage plants in the selected regions that are damaged by aphids.	**Essentially correct (E)** if the response satisfies the following four components: 1. Describes a random selection process that indicates the groupings used for the strata (by rows) 2. Describes how to correctly implement a random selection process for which the selection of regions within each stratum is equally likely 3. Describes a random selection process for which the selections across strata are independent 4. Describes a random selection process that results in the selection of one region from each stratum **Partially correct (P)** if the response satisfies only two or three of the four components required for E. **Incorrect (I)** if the response does not meet the criteria for E or P.

Scoring Notes:

- A response may satisfy component 1 by referring to the five regions in the rows instead of using row letters.
- For responses that use cards or slips of paper:
  - If the number of slips of paper (number of cards) does not equal 5 for each random selection, then component 1 is not satisfied. Slips of paper (cards) do not need to be specifically identified as equally sized.
  - If the response does not describe a thorough mixing (shuffling) of the slips of paper (cards), then component 2 is not satisfied.
- For responses that use a random number generator or table of random digits:
  - If it is not clear that a random selection process allows the selection of regions within each stratum to be equally likely, then component 2 is not satisfied.
  - If the response does not clearly indicate that a random number is generated from the region numbers within each of the five rows (e.g., only describes the generation of five random numbers from the two-digit integers from 01 to 25), then component 4 is not satisfied.

- For responses that use a fair die:
  - If a five-sided fair die is rolled for each of the five rows and the response clearly indicates the region numbers assigned to the values on the die for each roll, then component 4 is satisfied.
  - If a six-sided die is rolled for each of the five rows and the response clearly indicates which number is excluded (e.g., "if a 6 is rolled, roll again until a non-6 number is achieved") AND the response clearly indicates the region numbers (or columns) assigned to the values on the die for each roll, then component 4 is satisfied.
  - If a 25-sided fair die is rolled for each of the five selections, then component 4 is not satisfied without sufficient further justification because the selection of one region from each stratum is not guaranteed.
- If a response describes two separate random selection processes in detail (e.g., describes how to use a random number generator and slips of paper in a hat), score both descriptions according to the four components and use the lower score.
- If a response indicates that separate samples were taken from each of the strata, then component 3 is satisfied.
- A response that selects columns instead of regions within strata cannot satisfy component 3.

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 parts 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 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)

- 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.