2025

AP® Statistics

Sample Student Responses and Scoring Commentary

Inside:

Free-Response Question 2

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

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 2

Begin your response to **QUESTION 2** on this page.

A) Sampling method I is not an appropriate method for the farmer to use to estimate the proportion of cabbage plants in the field that are damaged by aphids because one region that is not randomly selected may not be representative of every region, especially since it is expected to be in the row with the least aphid impact.

B) The selection of row E is likely to provide an overestimate of the propotion of cabbage plants in the field that are damaged by aphids because row E is closest to the river, and the farmer believes that regions closer to the river are more affected by aphid damage, so the damage in row E is likely higher than the damage in other rows.

C) Use a computer to generate a random number between 1 and 5 inclusive and choose the corresponding row A region. Generate a random number between 6 and 10 inclusive and choose the corresponding row B region. Generate a random number between 11 and 15 inclusive and choose the corresponding row C region. Generate a random number between 16 and 20 inclusive and choose the corresponding row D region. Generate a random number between 21 and 25 inclusive and choose the corresponding row E region.

GO ON TO THE NEXT PAGE.

Question 2

Begin your response to **QUESTION 2** on this page.

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 damaged by aphids. This is not an appropriate method because the sample was not randomly selected. Additionally, our sample size is not large enough for us to come up with an appropriate conclusion due to a lack of a sample, and no randomness.

B) By using sampling method number 2, and if the farmers belief is correct, the selection of only using Row E is very likely to provide an overestimate of the proportion of cabbage plants in a field that are damaged by aphids. The farmer had previously stated that he believes that the extent of aphid damage is greater for the regions in the cabbage field closer to the river. In this case, Row E is the closest to the river and will most likely have the most aphid damage because of this and if the farmers belief is correct.

C) To implement sample 3 which requires a random selection of one region from each rows A, B, C, D, E, we must conduct a selection process that guarantees us a random selection. For each row, I would put 5 of the region numbers from the rows into a hat, and pick 1. For example, I will write the numbers 1-5 on slips of paper and blindly choose one. Then, I would repeat this process with each row, ending with 5 number slips chosen. By doing this, we are randomly selecting a certain plot from each row to get the best and most random sample for our experiment/ study.

GO ON TO THE NEXT PAGE.

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

Question 2

Begin your response to **QUESTION 2** on this page.

A. Sampling method 1 is not an appropriate sampling method for the farmer to use because it will only take data from region 3, which is one of the farthest regions from the river.

B. The selection of row E is likely to provide an overestimate of the proportion of cabbage plants in the field that are damaged by aphids. The farmer believes the extent of aphid damage is greater for the regions in the field closer to the river, and row E is the closest row to the river.

C. Assign each row a number 1-5. Using a RNG, select one number (ignore repeats). The row whose corresponding number was chosen will be examined to determine the aphid damage of every cabbage plant in that row.

GO ON TO THE NEXT PAGE.

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

Question 2

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

Overview

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

Sample: 2A
Score: 4

This response earned the following: part A – E, part B – E, and part C – E.

In part A the response indicates that sampling method I is not appropriate, satisfying component 1; indicates that the method does not use random sampling, satisfying component 2; and includes "proportion of cabbage plants" (parameter), "in the field" (population), and "region" (sampling unit), satisfying component 3. This part of the response was scored essentially correct (E).

In part B the response indicates that "row E is likely to provide an overestimate," satisfying component 1; states that "row E is closest to the river," satisfying component 2; and links the higher damage closer to the river to the farmer's belief, satisfying component 3. This part of the response was scored essentially correct (E).

In part C the response describes a random selection process that satisfies component 1 by using the rows as the groupings, satisfies component 2 by using a random number generator to pick a number at random from an appropriate set of numbers for each row, satisfies component 3 by picking a random number from each row separately, and satisfies component 4 by picking one number from each row. This part of the response was scored essentially correct (E).

Sample: 2B
Score: 3

This response earned the following: part A – E, part B – E, and part C – P.

In part A the response indicates that sampling method I is not appropriate, satisfying component 1; indicates that the sample was not randomly selected, satisfying component 2; and includes "proportion of cabbage plants" (parameter) and "in the field" (population), satisfying component 3. Note that the response did not mention a sampling unit, but only two of the three context components are required to satisfy component 3. This part of the response was scored essentially correct (E).

In part B the response indicates that "using row E is very likely to provide an overestimate," satisfying component 1; states that "Row E is the closest to the river," satisfying component 2; and links the greater damage in row E to the farmer's belief, satisfying component 3. This part of the response was scored essentially correct (E).

In part C the response describes a random selection process that satisfies component 1 by using the rows as the groupings. The response does not satisfy component 2 because it does not describe thorough mixing of the slips of paper. The response satisfies component 3 by picking a random number from each row separately and satisfies component 4 by picking one number from each row. For satisfying only three of the four components, this part of the response was scored partially correct (P).

Question 2 (continued)

Sample: 2C
Score: 1

This response earned the following: part A – I, part B – E, and part C – I.

In part A the response indicates that sampling method I is not appropriate, satisfying component 1. The response indicates that the region is farthest from the river, but that does not provide a reason why the region is likely not representative of the entire cabbage field, so it does not satisfy component 2. The response includes “region” (sampling unit) but does not include the population or a parameter, so it does not satisfy component 3. Because this part of the response only satisfies component 1, it was scored incorrect (I).

In part B the response indicates that ”The selection of row E is likely to provide an overestimate,” satisfying component 1; states that “row E is the closest row to the river,” satisfying component 2; and links the greater damage in row E to the farmer’s belief, satisfying component 3. This part of the response was scored essentially correct (E).

In part C the response describes a random selection process that selects only one row. Components 1, 2, 3, and 4 are not satisfied because the response is using rows as clusters rather than strata. Because none of the components are satisfied, this part of the response was scored incorrect (I).