2025

AP® Statistics

Free-Response Questions

STATISTICS

SECTION II

TIME – 1 HOUR AND 30 MINUTES

Directions:

Section II has 6 free-response questions and lasts 1 hour and 30 minutes.

You may use the available paper for scratch work and planning, but you must write your answers in the free-response booklet. Label parts (e.g., A, B, C) and sub-parts (e.g., i, ii, iii) as needed. Use a pencil or a pen with black or dark blue ink to write your responses.

Show all your work. Indicate clearly the methods you use, because you will be scored on the correctness of your methods as well as on the accuracy and completeness of your results and explanations. Correct answers without supporting work may not receive credit.

You may use a handheld calculator in this section. Use of a handheld graphing calculator is expected.

Reference information, including lists of formulas and tables, is available in this application and can be accessed throughout the exam.

You may pace yourself as you answer the questions in this section, or you may use these optional timing recommendations:

It is suggested that you spend about 1 hour and 5 minutes on questions 1 through 5 and about 25 minutes on question 6.

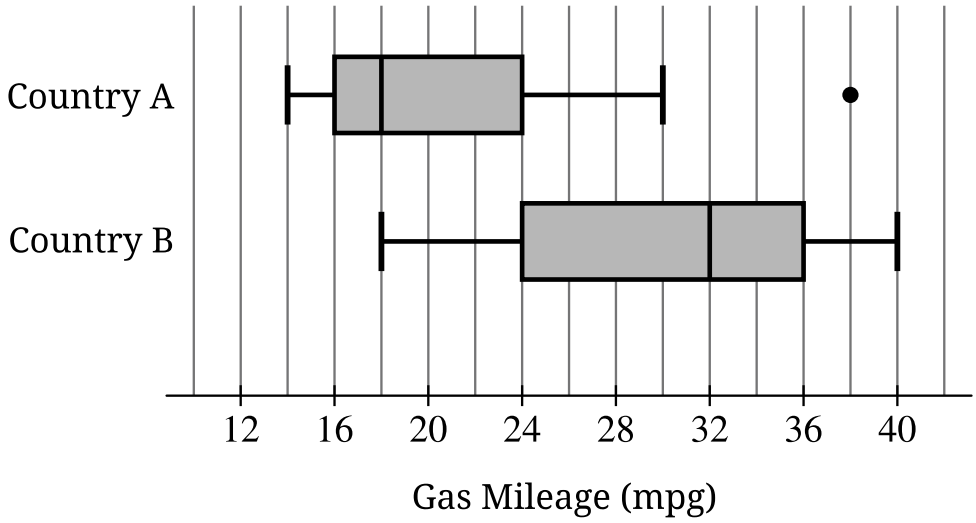
You can go back and forth between questions in this section until time expires. The clock will turn red when 5 minutes remain—**the proctor will not give you any time updates or warnings.**

Note: This exam was originally administered digitally. It is presented here in a format optimized for teacher and student use in the classroom.

During the AP Exam administration, students have access to reference information. To see the reference information for this course, please visit AP Central: https://apcentral.collegeboard.org/courses/ap-statistics/exam

1. The manager of an automotive company is interested in comparing the gas mileages for cars manufactured in Country A and cars manufactured in Country B. The manager selected a random sample of 100 cars manufactured in Country A and a random sample of 100 cars manufactured in Country B. The gas mileages for each sample, in miles per gallon (mpg), are summarized in the boxplots.

Boxplots of Gas Mileage for Each Country



A. Compare the distributions of gas mileage for the sample of cars manufactured in Country A and the sample of cars manufactured in Country B.

B. For the distribution of gas mileage for the sample of cars manufactured in Country A, would you expect the mean to be greater than 18 mpg, less than 18 mpg, or equal to 18 mpg? Justify your answer.

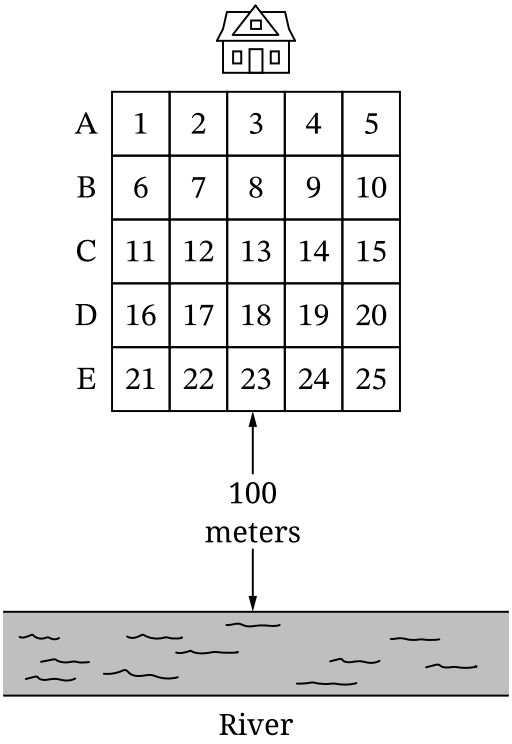
C. The manager will create a new boxplot with the combined data from the sample of cars manufactured in Country A and the sample of cars manufactured in Country B.

i. What is the range of the combined data set? Justify your answer.

ii. What is a possible value of the median of the combined data set? Justify your answer by referencing the boxplots shown.

2. Aphids are tiny insects that feed on plants such as cabbage plants. A farmer wants to reduce the number of aphids in a cabbage field. A river is located 100 meters south of the cabbage field. The farmer divides the field into 25 regions of equal size, as shown in the diagram. Each region has approximately the same number of cabbage plants.

Farmer's House and Cabbage Field



The farmer would like to estimate the proportion of cabbage plants in the field that are affected by aphids and believes that the extent of aphid damage is greater for the regions in the cabbage field closer to the river. To obtain the estimate, the farmer is considering three sampling methods.

- Sampling method I: Select region 3, which is closest to the farmer's house and farthest from the river. Examine every cabbage plant in the region for aphid damage.
- Sampling method II: Randomly select one row (A, B, C, D, or E). For every region in the selected row, examine every cabbage plant for aphid damage.
- Sampling method III: Randomly select one region from each of rows A, B, C, D, and E. For each selected region, examine every cabbage plant for aphid damage.

A. Explain whether sampling method I is an appropriate sampling method for the farmer to use to estimate the proportion of cabbage plants in the field that are damaged by aphids.

B. Using sampling method II, the farmer randomly selected row E and examined every cabbage plant in row E. If the farmer's belief is correct, determine whether the selection of row E is likely to provide an overestimate or an underestimate of the proportion of cabbage plants in the field that are damaged by aphids. Justify your answer.

C. Using the information provided in the diagram of the cabbage field, describe how to implement sampling method III, which requires a random selection of one region from each of rows A, B, C, D, and E.

3. Ms. Fey is a manager at a restaurant. To improve the dining experience for her customers, she uses a digital music service to create a playlist of songs that will be played in the restaurant. The playlist contains 1,000 songs and consists of four different types of music in the following quantities: 200 country songs, 400 pop songs, 100 rock songs, and 300 jazz songs. The digital music service will select songs at random from the playlist to be played in the restaurant. Any song can be replayed at any time.

A.

i. Suppose one song is selected at random to be played. What is the probability that the song is a rock song? Show your work.

ii. Suppose two songs are selected at random to be played. What is the probability that both songs are rock songs? Show your work.

B. In every one-hour period, 20 songs will be played at random and any song can be replayed at any time. Ms. Fey is interested in how many rock songs will be played in a typical one-hour period.

i. Define the random variable of interest to Ms. Fey, and state how the random variable is distributed.

ii. What is the expected value for the random variable in part B (i)? Show your work.

C. Recall that in every one-hour period, 20 songs will be played at random and any song can be replayed at any time.

i. Determine the probability that 4 or more rock songs in a particular one-hour period will be played. Show your work.

ii. Suppose 4 rock songs are played during a particular one-hour period. Does this provide strong evidence that the song selection process was not truly random? Justify your answer without performing an inference procedure.

4. A software application (app) lets users enter questions to receive answers in the form of images, texts, or videos. Research indicates that 22 percent of high school students in Country W use the app to help them with their homework at least once per week. Karen is an AP Statistics student in Country W at a high school that has more than 2,000 students. She believes the proportion of all students at her school who use the app to help them with their homework at least once per week is greater than the proportion for her country. To investigate her belief, she took a simple random sample of 130 students from her school and found that 38 of the sampled students use the app to help them with their homework at least once per week.

Is there convincing statistical evidence, at a 0.05 significance level, to support Karen's belief? Justify your answer with the appropriate inference procedure.

5. According to a 2017 national survey in Country B, the mean number of bedrooms in newly built houses was 2.9. Rodney, a researcher, believes the mean number of bedrooms in newly built houses in the country was different in 2024 than it was in 2017. To investigate his belief, he took a large random sample of newly built houses in Country B in 2024 and recorded the number of bedrooms in each house. The distribution of the number of bedrooms for the sampled houses is summarized in the table.

Distribution of the Number of Bedrooms for the Houses Sampled in 2024

Number of Bedrooms	1	2	3	4	5	6
Proportion of Houses	0.12	0.22	0.28	0.22	0.14	0.02

A.

i. A house from the sample will be selected at random. What is the probability that the house had fewer than 3 bedrooms? Show your work.

ii. What is the mean number of bedrooms for the sample of newly built houses in 2024? Show your work.

B. Rodney will use a one-sample *t*-test for a population mean to test his belief.

i. In the context of Rodney's investigation, state the hypotheses for the test.

ii. Explain, in context, what a Type I error would be for Rodney's hypothesis test.

C. A different researcher, Keisha, suggests using a confidence interval to investigate whether the mean number of bedrooms in newly built houses in 2024 in Country B was different from 2.9.

Assume the conditions for inference have been met. Using Rodney's data, Keisha calculated a one-sample 97 percent confidence interval to estimate the population mean as $(3.01, 3.19)$. Based on the confidence interval, what conclusion can be made for Rodney's hypothesis test in part B at $\alpha = 0.03$? Justify your answer.

6. Stefan, a psychologist, conducted a study to investigate the effect of time of day on reading comprehension in children. One hundred children volunteered, with their parents' consent, to participate in the study. Fifty of the children were randomly assigned to read a story at 9 a.m. and then answer 25 questions about it. The remaining 50 children were assigned to read the same story at 3 p.m. and answer the same 25 questions. The reading comprehension for each child was measured by a reading score, which was determined by the number of questions that were answered correctly about the story. Stefan is interested in comparing the mean reading scores for the two times of day. Table 1 shows the results of Stefan's study.

Table 1: Summary Statistics of Reading Scores

	n	Mean	Standard Deviation
9 a.m.	50	15.2	4.12
3 p.m.	50	17.9	4.43

Stefan found the conditions for inference were met and conducted a two-sample t-test for the difference in two population means. Let μ_{AM} represent the mean reading score for all children, similar to those in the study, who would read the story at 9 a.m. Let μ_{PM} represent the mean reading score for all children, similar to those in the study, who would read the story at 3 p.m. Stefan's hypotheses are as shown.

$$H_0: \mu_{\text{AM}} = \mu_{\text{PM}}$$
$$H_a: \mu_{\text{AM}} \neq \mu_{\text{PM}}$$

A. The p-value for Stefan's hypothesis test was 0.002. State an appropriate conclusion, at the 5 percent significance level, for Stefan's test in the context of the investigation. Justify your answer.

B. Explain why it was appropriate for Stefan to conduct a two-sample t-test for the difference in two population means instead of a paired t-test for the population mean difference.

C. Researchers are usually interested in the practical importance of their results as well as the statistical significance of the hypothesis test. The practical importance of the results indicates whether the observed results are meaningful in real life. For example, in an investigation of the heights of two groups of students, a difference in the two group means of 3.8 inches is much more meaningful, or has more practical importance, than a difference in the two group means of only 0.2 inches.

One indicator of practical importance is effect size. A common method for measuring effect size for the difference in two group means is Cohen's *d* coefficient. Cohen's *d* coefficient compares the absolute value of the difference in the means of the two groups to the pooled variability of the observed data values from the two groups.

Cohen's *d* coefficient can be calculated using $d = \frac{|\bar{x}_1 - \bar{x}_2|}{s_p}$, where s_p represents the pooled standard deviation, $\bar{x}_1$ represents the sample mean for the first group, and $\bar{x}_2$ represents the sample mean for the second group. When the sizes of the groups are equal, s_p is calculated as $s_p = \sqrt{\frac{s_1^2 + s_2^2}{2}}$, where s_1 represents the sample standard deviation for the first group and s_2 represents the sample standard deviation for the second group.

Consider the summary statistics from Stefan's study in Table 1.

i. Calculate Cohen's *d* coefficient for Stefan's study. Show your work.

ii. Higher values of Cohen's *d* indicate greater practical importance and lower values of Cohen's *d* indicate less practical importance. Typically, we use the intervals listed in Table 2 to help interpret practical importance.

Table 2: Guidelines for Interpreting Cohen's *d* Coefficient

Cohen's *d* Coefficient	Practical Importance
$0 \le d \le 0.20$	Not very meaningful in real life
$0.20 < d < 0.80$	Somewhat meaningful in real life
$d \ge 0.80$	Very meaningful in real life

Based on your answer to part C (i) and the information in Tables 1 and 2, describe the practical importance of Stefan's results, in context.

D. Suppose the results of Stefan's study, summarized in Table 1, instead had a standard deviation for the 9 a.m. reading scores, s_1, and a standard deviation for the 3 p.m. reading scores, s_2, that were both greater than 4.43. Assume the group sample sizes and the means are not changed.

i. Would the Cohen's *d* coefficient in this new situation be smaller than, larger than, or the same as the Cohen's *d* coefficient calculated in part C (i)? Explain your answer.

ii. Does the Cohen's *d* coefficient described in part D (i) indicate that Stefan's observed difference in the means in the new situation would have more practical importance than, less practical importance than, or the same practical importance as what was originally determined in part C (ii)? Explain your answer.

STOP

END OF EXAM