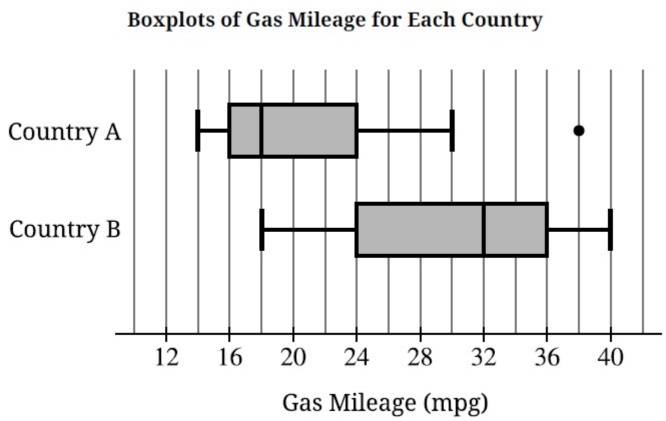
1. (Exploring Data). The manager of an automotive company is interested in comparing the gas mileages for cars manufactured in Country A and cars manufactured in Country B. The manager selected a random sample of 100 cars manufactured in Country A and a random sample of 100 cars manufactured in Country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



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.

Solution:
A. The distribution of gas mileage for the sample of cars manufactured in Country A has a lower center than the distribution of gas mileage for the sample of cars manufactured in Country B. The median gas mileage for the sample of cars manufactured in Country A (18 mpg) is less than the median gas mileage for the sample of cars manufactured in Country B (32 mpg).

The range of the gas mileages for the sample of cars manufactured in Country A (24 mpg) is slightly greater than the range of the gas mileages for the sample of cars manufactured in Country B (22 mpg). However, the IQR of the gas mileages for the sample of cars manufactured in Country A (8 mpg) is less than the IQR of the gas mileages for the sample of the cars manufactured in Country B (12 mpg).

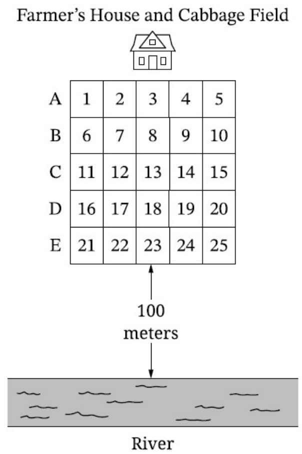
The car manufactured in Country A with 38 mpg (the maximum of the sample of cars manufactured in Country A) is an outlier, while the distribution of gas mileage for the sample of cars manufactured in Country B has no outliers

B. The mean of the distribution of gas mileage for the sample of cars manufactured in Country A is expected to be greater than 18 mpg, the median of the distribution. Because the distribution of gas mileage for the sample of cars manufactured in Country A has an outlier to the right (or is skewed to the right), the mean of the distribution (which is not resistant) is expected to be pulled above the median (which is resistant) toward the higher values of gas mileage.

C.
i. The maximum value in the combined data is 40 mpg because 40 mpg is the maximum gas mileage for the sample of cars manufactured in Country B, and as shown in the boxplot, all the gas mileages for the sample of cars manufactured in Country A are less than 40 mpg. The minimum value in the combined data is 14 mpg, because 14 mpg is the minimum mpg for the sample of cars manufactured in Country A, and as shown in the boxplot, all the gas mileages for the sample of cars manufactured in Country B are greater than 14. Thus, the range of the combined data set is 40-14=26 mpg.

ii. In the combined data, there are 200 gas mileages. The median is a value where at least half, or 100, of the gas mileages in the combined data are less than or equal to the median value and at least half, or 100, of the gas mileages in the combined data are greater than or equal to the median value. From the boxplot for the sample of cars manufactured in Country A, the third quartile, Q3, is 24 mpg indicating there are at least 75 gas mileages less than or equal to 24 mpg and at least 25 gas mileages greater than or equal to 24 mpg. From the boxplot for the sample of cars manufactured in Country B, the first quartile, Q1, is 24 mpg indicating there are at least 25 gas mileages less than or equal to 24 mpg and at least 75 gas mileages greater than or equal to 24 mpg. Thus, in the combined data set, there are at least 100 gas mileages less than or equal to 24 mpg and at least 100 gas mileages greater than or equal to 24 mpg, which implies 24 is the value of the median of the combined data set.

2. (Sampling and Experimental Design). 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.



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.

Solution:

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

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.

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.

3 (Probability and Sampling Distributions). 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.

Solution:

A

i. $P(\text{Rock Song}) = \frac{100}{1{,}000} = 0.10$

ii. $P(\text{Both Rock Songs}) = (0.10)(0.10) = 0.01$

B

i. Let the random variable of interest, X, represent the number of the 20 songs played in one hour that are rock songs. It is stated that any song can be replayed at any time, which establishes that each rock song has probability $\frac{100}{1{,}000} = 0.10$ of being selected each hour and each song is independent from every other song. Therefore, X has a binomial distribution with $n = 20$ independent trials and probability of success $p = 0.10$ for each trial.

ii. The expected value for the number of rock songs played in one hour is $np = 20(0.10) = 2$ songs.

C

i. The probability that in a particular hour 4 or more rock songs will be played is $P(X \geq 4) = 1 - P(X \leq 3)$

$$P(X \geq 4) = 1 - \binom{20}{0}(0.10)^0(0.90)^{20} + \binom{20}{1}(0.10)^1(0.90)^{19}$$

$$+\binom{20}{2}(0.10)^2(0.90)^{18} + \binom{20}{3}(0.10)^3(0.90)^{17}$$

$P(X \geq 4) = 1 - 0.867 = 0.133.$

ii. No, the probability that 4 or more rock songs would be played in an hour is 0.133, which is high enough to be reasonably attributed to chance alone. This probability is not small enough to provide evidence that the selection process was not truly random.

4 (Inference). 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.

Solution:

Section 1 An appropriate inference procedure is a one-sample z-test for a population proportion. The null hypothesis is $\mathrm{H}_0 : p = 0.22$, and the alternative hypothesis is $\mathrm{H}_\mathrm{a} : p > 0.22$, where $p =$ the true proportion of students at Karen's high school that use the application at least once per week.

Section 2 The independent observation condition for performing the one-sample z-test for a population proportion is satisfied. This is because the data were obtained from a random sample of 130 high school students from Karen's high school. Also, the sample of 130 students is less than 10% of the total number of students at this large high school, because $130 < 0.10(2,000)$ and the total number of students in Karen's high school is greater than 2,000. The 10% condition is required as sampling was conducted without replacement from a finite population.

The number of expected successes and expected failures were both more than 10 because $130(0.22) = 28.6$ and $130(0.78) = 101.4$. Thus, the sample size is large enough to support the assumption that the sampling distribution of $\hat{p}$ is approximately normal.

$$\hat{p} = \frac{38}{130} \approx 0.2923$$

Test statistic:

$$z = \frac{0.2923 - 0.22}{\sqrt{\frac{0.22(1-0.22)}{130}}} \approx 1.99$$

$P(z > 1.99) \approx 0.023$

Section 3 Because this p-value is less than the $\alpha = 0.05$ significance level, the null hypothesis should be rejected. There is convincing statistical evidence that the population proportion of students at Karen's high school that use the application at least once per week is greater than Country W's proportion of 0.22.

5 (Multi-Focus). 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.

Solution:

A

i. Let random variable X represent the number of bedrooms in a randomly selected newly built house in the 2024 sample from Country B. The probability that a randomly selected house from the 2024 sample had fewer than 3 bedrooms is the probability that the house had either 1 or 2 bedrooms, which is

$$\begin{aligned} P(X < 3) &= P(X = 1) + P(X = 2) \\ P(X < 3) &= 0.12 + 0.22 \\ P(X < 3) &= 0.34. \end{aligned}$$

ii. The average number of bedrooms per house for the sample of newly built houses in 2024 is

$$\begin{aligned} E(X) &= 1(0.12) + 2(0.22) + 3(0.28) + \\ &\quad 4(0.22) + 5(0.14) + 6(0.02) \\ E(X) &= 3.10 \text{ bedrooms.} \end{aligned}$$

B

i. Let $\mu =$ the population mean number of bedrooms in newly built houses in 2024 from Country B. The null hypothesis is $\mathrm{H_0} : \mu = 2.9$ and the alternative hypothesis is $\mathrm{H_a} : \mu \neq 2.9$.

ii. A Type I error would be determining that the population mean number of bedrooms in newly built houses in 2024 from Country B is not equal to 2.9 when it is in fact 2.9.

C Because the value 2.9 is not contained within the 97% confidence interval, the null hypothesis should be rejected. Therefore, there is convincing statistical evidence, at the $\alpha = 0.03$ level of significance, that the population mean number of bedrooms in newly built houses in 2024 from Country B is not equal to 2.9 (or is different than that in 2017).

6 (Investigative Task). 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.

$$\text{H}_0 : \mu_{\text{AM}} = \mu_{\text{PM}}$$
$$\text{H}_\text{a} : \mu_{\text{AM}} \neq \mu_{\text{PM}}$$

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 $d = \dfrac{|\bar{x}_1 - \bar{x}_2|}{s_p}$, 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 $s_p = \sqrt{\dfrac{s_1^2 + s_2^2}{2}}$, 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**
$0 \leq d \leq 0.20$	Not very meaningful in real life
$0.20 < d < 0.80$	Somewhat meaningful in real life
$d \geq 0.80$	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.

Solution:

A Because the p-value of 0.002 is less than the level of significance of 0.05, the null hypothesis should be rejected. There is convincing statistical evidence of a difference between the mean reading score for all children, similar to those who participated in the study, who would read the story at 9 a.m. and the mean reading score for all children, similar to those who participated in the study, who would read the story at 3 p.m.

B It was appropriate for Stefan to conduct a two-sample t-test instead of a paired t-test because the two groups are independent. Stefan used random assignment to place the 100 volunteer children into two groups, and there is no indication that the two groups of 50 children are paired in any meaningful way (e.g., age, reading comprehension level)

C

i. The value of the pooled standard deviation is

$$s_p = \sqrt{\frac{(4.12)^2 + (4.43)^2}{2}} \approx 4.28.$$

Therefore, the value of Cohen's d is

$$d = \frac{|15.2 - 17.9|}{4.28} \approx 0.63.$$

ii. Based on Table 2, a Cohen's d value of 0.63 would indicate that Stefan's results were somewhat practical or meaningful in real life.

D

i. The value of Cohen's d would decrease. If the standard deviation for the a.m. group and p.m. group were both greater than 4.43, the pooled standard deviation would be greater than 4.28. With a larger value in the denominator and the same value in the numerator, the value of Cohen's d would be smaller than 0.63.

ii. The lower Cohen's d value would indicate less practical importance than that of the original results of Stefan's study.

Problems adapted from the College Board released tests.