• 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

$$P(X < 3) = P(X = 1) + P(X = 2)$$
$$P(X < 3) = 0.12 + 0.22$$
$$P(X < 3) = 0.34.$$

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 $\text{H}_0 : \mu = 2.9$ and the alternative hypothesis is $\text{H}_\text{a} : \mu \neq 2.9$.

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).

• 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 $\mu_{\text{AM}}$ represent the mean reading score for all children, similar to those in the study, who would read the story at 9 a.m. Let $\mu_{\text{PM}}$ 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 $s_p = \sqrt{\frac{s_1^2 + s_2^2}{2}}$, 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** |
|---|---|
| $0 \leq d \leq 0.20$ | 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)

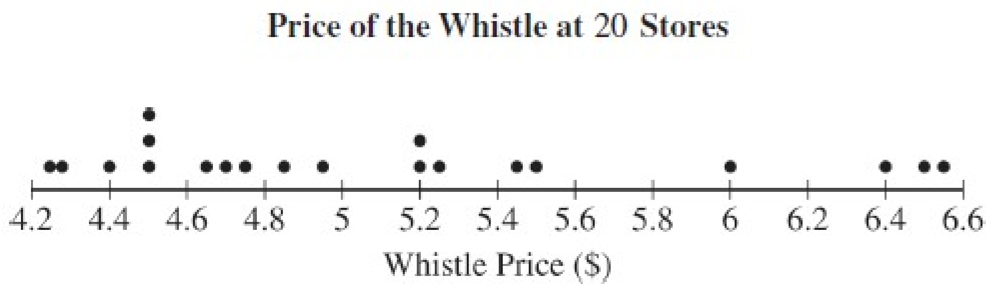
i. The value of Cohen's $d$ would decrease. If the standard deviation for the a.m. group and p.m. group were both greater than 4.43, the pooled standard deviation would be greater than 4.28. With a larger value in the denominator and the same value in the numerator, the value of Cohen's $d$ would be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

• A company sells a certain type of whistle. The price of the whistle varies from store to store. Julio, a statistician at the company, wants to estimate the mean price, in dollars ($), of this type of whistle at all stores that sell the whistle.

(A) (**i**) Identify the appropriate inference procedure for Julio to use.
(**ii**) Describe the parameter for the inference procedure you identified in part (a-i) in context.

Julio called the managers of 20 randomly selected stores that sell the whistle and recorded the price of the whistle at each store. Following is a dotplot of Julio's data.




The summary statistics for Julio's data are shown in the following table.

**Summary Statistics for Julio's Data**

| Sample Size Deviation | Mean Minimum | Standard $Q_1$ | Median | $Q_3$ | Maximum | | |
|---|---|---|---|---|---|---|---|
| 20 | 5.12 | 0.743 | 4.25 | 4.51 | 4.885 | 5.475 | 6.58 |

(B) Julio wants to examine some characteristics of the distribution of the sample of whistle prices.

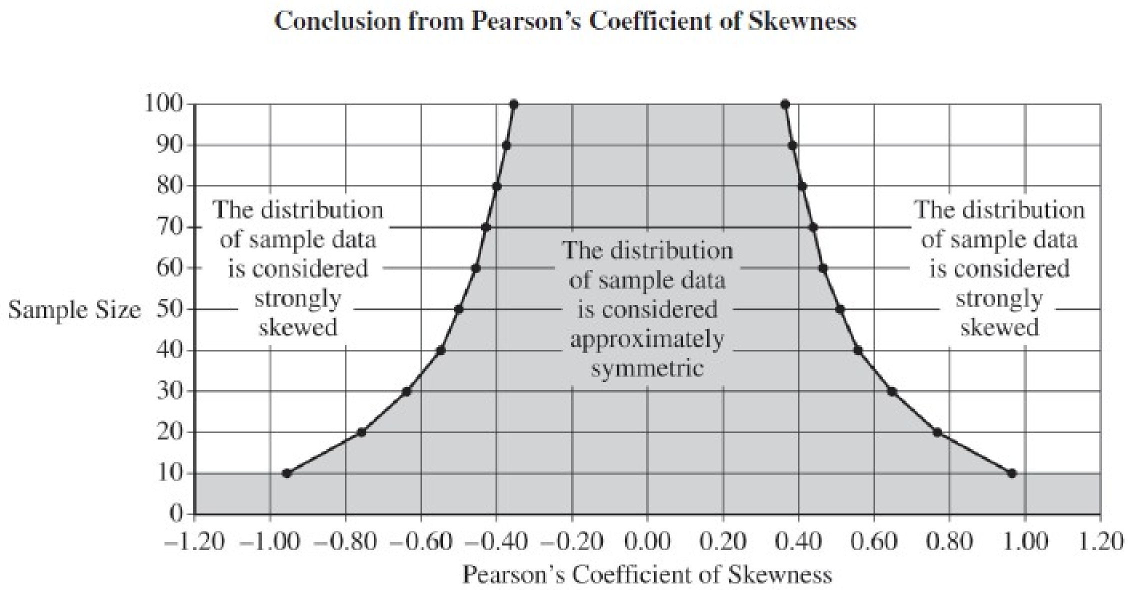
i. Describe the shape of the distribution of the sample of whistle prices. Justify your response using appropriate values from the summary statistics table.

ii. Using the $1.5 \times$ IQR rule, determine whether there are any outliers in the sample of whistle prices. Justify your response.

It can often be difficult to determine whether the distribution of sample data is skewed by looking at a graph of the data and the summary statistics, particularly when the sample size is small. Thus, statisticians sometimes measure how skewed a data set is. One such measure is Pearson's coefficient of skewness, which is calculated using the following formula.

$$\text{Pearson's Coefficient of Skewness} = \frac{3(\bar{x} - m)}{s}$$

In the formula, $\bar{x}$ is the sample mean, $m$ is the sample median, and $s$ is the sample standard deviation.
(C) (**i**) Calculate Pearson's coefficient of skewness for Julio's sample of 20 whistle prices. Show your work.
The following graph shows conclusions that can be made about the shape of the distribution of sample data based on Pearson's coefficient of skewness and sample size.




(**ii**) Indicate the value of the Pearson's coefficient of skewness you calculated in part (c-i) for the appropriate sample size by marking it with an "X" on the preceding graph.

(D) Consider your work in part (c).

(**i**) What should you conclude about the shape of the distribution of the sample of whistle prices? Justify your response.

Julio's inference procedure in part (a-i) needs one of the following requirements to be satisfied to verify the normality condition.

- The sample size is greater than or equal to 30.
- If the sample size is less than 30, the distribution of the sample data is not strongly skewed and does not have outliers.

(ii) Using your response to (d-i) and the preceding requirements, is the normality condition satisfied for Julio's data? Explain your response.

*Solutions:*

(A) (i) The inference procedure that should be used to estimate the mean price, in dollars ($), of this type of whistle at all stores that sell the whistle is a one-sample $t$-interval for a population mean.

(ii) The parameter of interest is the mean whistle price, in dollars ($), of this type of whistle at all stores that sell the whistle.

(b) (i) The distribution of the sample of whistle prices appears slightly skewed to the right, because the mean is slightly higher than the median.

(ii) Based on the $1.5 \times \text{IQR}$ rule, there are not whistle prices in this sample that would be considered outliers. A whistle price is an outlier using this method if it is more than $1.5 \times \text{IQR}$ below the first quartile ($Q_1$) or more than $1.5 \times \text{IQR}$ above the third quartile ($Q_3$). Because

$$\begin{aligned} Q_1 - 1.5 \times \text{IQR} &= 4.51 - 1.5(5.475 - 4.51) \\ &= 3.0625, \end{aligned}$$

and the minimum value (4.25) is greater than 3.0625, there are no outliers to the left. Because

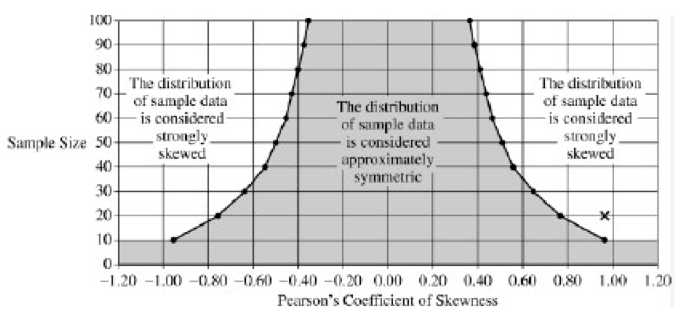
$$\begin{aligned} Q_3 + 1.5 \times \text{IQR} &= 5.475 + 1.5(5.475 - 4.51) \\ &= 6.9225, \end{aligned}$$

and the maximum value (6.58) is less than 6.9225, there are no outliers to the right.

(C) (i) Pearson's coefficient of skewness is

$$\frac{3(5.12 - 4.885)}{0.743} \approx 0.949.$$

(ii)




(D) (i) Looking at the graph in part (c), for a sample size of 20, and a skewness coefficient of 0.949, this point falls in "the distribution of sample data is considered strongly skewed" region. Therefore, we would consider the shape of the distribution of the sample of whistle prices to be strongly skewed.

(ii) No, based on the response to part (d-i). Julio's data would not satisfy the normality condition because neither of the criteria listed are met. Julio only has a sample size of 20, which is less than 30, and Pearson's coefficient of skewness indicates the distribution of sample data is strongly skewed.

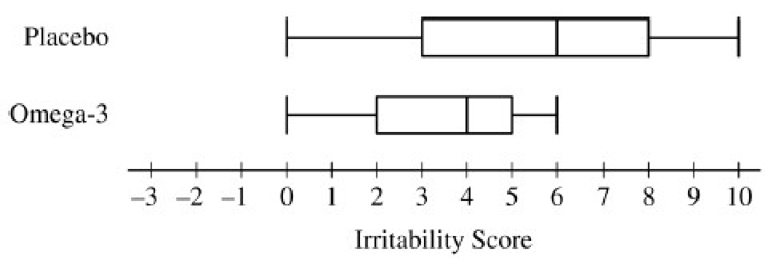
• A medical researcher completed a study comparing an omega-3 fatty acids supplement to a placebo in the treatment of irritability in patients with a certain medical condition. Nineteen patients with the medical condition volunteered to participate in the study. The study was conducted using the following weekly schedule.

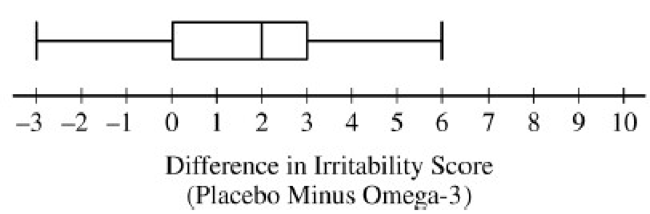
- Week 1: Each patient took a randomly assigned treatment, omega-3 supplement or placebo.
- Week 2: The patients did not take either the omega-3 supplement or the placebo. This was necessary to reduce the possibility of any carryover effect from the assigned treatment taken during week 1.
- Week 3: Each patient took the treatment, omega-3 supplement or placebo, that they did not take during week 1.

At the end of week 1 and week 3, each patient's irritability was given a score on a scale of 0 to 10, with 0 representing no irritability and 10 representing the highest level of irritability.

For each patient, the two irritability scores and the difference in their scores (placebo minus omega-3) were recorded. The results are summarized in the table and boxplots.

| | $n$ | Mean | Standard Deviation |
|---|---|---|---|
| Placebo | 19 | 5.421 | 2.987 |
| Omega-3 | 19 | 3.632 | 1.739 |
| Difference (placebo minus omega-3) | 19 | 1.789 | 2.485 |




The researcher claims the omega-3 supplement will decrease the mean irritability score of all patients with the medical condition similar to the volunteers who participated in the study. Is there convincing statistical evidence to support the researcher's claim at a significance level of $\alpha = 0.05$? Complete the appropriate inference procedure to support your answer.

*Solution:*

**Section 1**

Let $\mu_d$ represent the true mean difference (placebo minus omega-3) of irritability scores for all people with this medical condition.

The null hypothesis is $H_0$: $\mu_d = 0$ and
The alternative hypothesis is $H_a$: $\mu_d > 0$.

The appropriate inference procedure is a matched pairs $t$-test for a mean difference.

**Section 2**

The independence condition for performing a paired $t$-test for a mean difference is satisfied because the data were obtained from a randomized experiment where the week in which the patient received the treatment was randomly assigned.

The sampling distribution of the mean difference must be approximately normal. Although the sample size is less than 30 ($n = 19$), this is satisfied because the boxplot for the sample differences shows an approximately symmetric distribution with no outliers.

The value of the test statistic is:

$$t = \frac{\overline{x}_d - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{1.789 - 0}{\frac{2.485}{\sqrt{19}}} \approx 3.138$$

Using 18 degrees of freedom, the corresponding p-value is $P(t > 3.138) \approx 0.0028$.

**Section 3**

Because the $p$-value $\approx 0.0028$ is less than the significance level, $\alpha = 0.05$, the null hypothesis should be rejected. The data provide convincing statistical evidence that for patients similar to those in the study, the true mean difference (placebo minus omega-3) in irritability scores for people with this medical condition is greater than zero. This suggests the omega-3 fatty acids are helpful in reducing irritability scores in people with this medical condition.

---

*Problems adapted from the College Board released tests.*