Exercises

Teaching reading An educator believes that new reading activities in the classroom will help elementary school pupils improve their reading ability. She recruits 44 third-grade students and randomly assigns them into two groups. One group of 21 students does these new activities for an 8-week period. A control group of 23 third-graders follows the same curriculum without the activities. At the end of the 8 weeks, all students are given the Degree of Reading Power (DRP) test, which measures the aspects of reading ability that the treatment is designed to improve. Comparative boxplots and summary statistics for the data from Fathom are shown below.

(a) Based on the graph and numerical summaries, write a few sentences comparing the DRP scores for the two groups.

(b) Is the mean DRP score significantly higher for the students who did the reading activities? Carry out an appropriate test to support your answer.

(c) Can we conclude that the new reading activities caused an increase in the mean DRP score? Explain.

(d) Construct and interpret a 95% confidence interval for the difference in mean DRP scores. Explain how this interval provides more information than the significance test in part (b).

Who talks more—men or women? Researchers equipped random samples of 56 male and 56 female students from a large university with a small device that secretly records sound for a random 30 seconds during each 12.5-minute period over two days. Then they counted the number of words spoken by each subject during each recording period and, from this, estimated how many words per day each subject speaks. The female estimates had a mean of 16,177 words per day with a standard deviation of 7520 words per day. For the male estimates, the mean was 16,569 and the standard deviation was 9108.

(a) Do these data provide convincing evidence of a difference in the average number of words spoken in a day by male and female students at this university? Carry out an appropriate test to support your answer.

(b) Interpret the P-value from part (a) in the context of this study.

Paired or unpaired? In each of the following settings, decide whether you should use paired t procedures or two-sample t procedures to perform inference. Explain your choice.

(a) To test the wear characteristics of two tire brands, A and B, each brand of tire is randomly assigned to 50 cars of the same make and model.

(b) To test the effect of background music on productivity, factory workers are observed. For one month, each subject works without music. For another month, the subject works while listening to music on an MP3 player. The month in which each subject listens to music is determined by a coin toss.

(c) A study was designed to compare the effectiveness of two weight-reducing diets. Fifty obese women who volunteered to participate were randomly assigned into two equal-sized groups. One group used Diet A and the other used Diet B. The weight of each woman was measured before the assigned diet and again after 10 weeks on the diet.

A better drug? In a pilot study, a company's new cholesterol-reducing drug outperforms the currently available drug. If the data provide convincing evidence that the mean cholesterol reduction with the new drug is more than 10 milligrams per deciliter (mg/dl) greater than with the current drug, the company will begin the expensive process of mass-producing the new drug. For the 14 subjects who were assigned at random to the current drug, the mean cholesterol reduction was 54.1 mg/dl with a standard deviation of 11.93 mg/dl. For the 15 subjects who were randomly assigned to the new drug, the mean cholesterol reduction was 68.7 mg/dl with a standard deviation of 13.3 mg/dl. Graphs of the data reveal no outliers or strong skewness.

(a) Carry out an appropriate significance test. What conclusion would you draw? (Note that the null hypothesis is $H_0: \mu_1 - \mu_2 = 0$.)

(b) Based on your conclusion in part (a), could you have made a Type I error or a Type II error? Justify your answer.
Dr. Teresa Amabile conducted a study involving 47 college students, who were randomly assigned to two treatment groups. The 23 students in one group were given a list of statements about external reasons (E) for writing, such as public recognition, making money, or pleasing their parents. The 24 students in the other group were given a list of statements about internal reasons (I) for writing, such as expressing yourself and enjoying playing with words. Both groups were then instructed to write a poem about laughter. Each student's poem was rated separately by 12 different poets using a creativity scale. The 12 poets' ratings of each student's poem were averaged to obtain an overall creativity score.

We used Fathom software to randomly reassign the 47 subjects to the two groups 1000 times, assuming the treatment received doesn't affect each individual's average creativity rating. The dotplot shows the approximate randomization distribution of $\bar{x}_I - \bar{x}_E$.

(a) Why did researchers randomly assign the subjects to the two treatment groups?

(b) In the actual experiment, $\bar{x}_I - \bar{x}_E = 4.15$. What conclusion would you draw? Justify your answer with appropriate evidence.

(c) Based on your conclusion in part (b), could you have made a Type I error or a Type II error? Justify your answer.
**A. Remember Cuss and BS**

The activity group's center (51.5) is higher than the control (41.5). The control group has a larger spread (17.1) compared to the activity (11.0) spread. The shape of the activity group appears to be slightly skewed to the left while the shape of the control group appears more symmetric. Neither group appears to have any outliers.

**B. Parameters:**

- $\mu_1 = \text{actual mean DEP for 3rd graders doing the activity}$
- $\mu_2 = \text{actual mean DEP for 3rd graders not doing activity}$

$H_0: \mu_1 = \mu_2 \quad (\text{the same})$

$H_A: \mu_1 > \mu_2 \quad (\text{activity better})$

- $\sigma$ is unknown (t-inference)
- Random - This was a randomized comparative study
- Normal - The box plot shows neither outliers nor strong skewness
- Independent - Due to random assignment these two groups can be viewed as independent
TEST - 2 SAMPLE T TEST FOR \( \mu_1 - \mu_2 \)

\[ t^* : \quad \text{df}_1 = 21 - 1 = 20 \quad \text{Select the Constructive} \quad \text{df} = 20 \]
\[ \text{df}_2 = 23 - 1 = 22 \quad \alpha = 0.05 \]

Calc

\[ t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{51.5 - 41.5}{\sqrt{\frac{11.2^2}{21} + \frac{17.1^2}{23}}} \]

\[ t = 2.33 \]

\[ \text{df} = 37.9 \]

\[ \text{P-value} : \quad P (t_{20} > 2.33) = t_{cdf}(2.33, 37.9, 20) = 0.0152 \]

Conclude: Since the p-value \(0.0152\) < 0.05, we reject \( H_0 \).
We have sufficient evidence to conclude there is a difference in the actual mean DRP scores for 3rd graders doing the activity and those not doing the activity.

Since this was a randomized controlled experiment, we can conclude that the activities caused an increase in the mean DRP Score.

95% 2 SAMPLE T INTERVAL. FOR \( \mu_1 - \mu_2 \)

Calc

\[ \bar{X}_1 - \bar{X}_2 + t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \]

\[ 51.5 - 41.5 \pm 2.086 \sqrt{\frac{11.2^2}{21} + \frac{17.1^2}{23}} \]

\[ t_{20} = t_{inv}(0.05, 20) = -2.086 \]

Conclude: We are 95% confident the interval 10.3 to 18.97 captures the difference in actual mean DRP scores for 3rd graders that do the activity and those that do not.
(H1) Women: $\mu_1 = 56$  \[ \bar{x}_1 = 16.177 \]  $s_1 = 7.529$

(\mu_2) Men: $\mu_2 = 56$  \[ \bar{x}_2 = 16.569 \]  $s_2 = 9.108$

$\mu_1 =$ actual mean number of words spoken by women per day
$\mu_2 =$ actual mean number of words spoken by men per day

$H_0: \mu_1 = \mu_2$

$H_A: \mu_1 \neq \mu_2$

**Conditions:**
Random: both samples randomly selected
Normal: both samples are over 30 ($n > 30$)
Independent: the groups (men/women) are independent.
Reasonable to assume there are more than 560 men and 560 women at the school
$\sigma$ is unknown (t-inference)

**Test:** 2 sample $t$-test for $\mu_1 - \mu_2$

$$t_{df} = \frac{16.177 - 16.569}{\sqrt{\frac{7.529^2}{56} + \frac{9.108^2}{56}}} = -2.48$$

$P(\text{value} < -2.48) = 0.020$

**Conclusion:** Since the $p$-value ($0.020 > 0.05$), we fail to reject $H_0$. We do not have enough evidence to conclude that male and female students speak a different number of words per day on average.
Interpretation of p-value in context:

If males and females at this university speak the same number of words per day on average, then we have about an 80% chance of selecting a sample where the difference between the average number of words spoken per day by males and females is as large as or larger than the difference we actually saw.
PAIRED T OR 2 SAMPLE T

a) Test 2 different tires - 2 sample T-test

b) Test effect of background music. The same subjects worked a month without music then a month with music - paired T-test

c) Subject were randomly assigned to 2 groups. One group used diet A and the other group diet B - 2 sample T-test.
65. \( M_1 = \) actual mean cholesterol reduction using new drug
\( M_2 = \) actual mean cholesterol reduction using current drug

\[ H_0: \ M_1 - M_2 = 10 \]
\[ H_A: \ M_1 - M_2 > 10 \]

Mean cholesterol reduction with new drug is more than 10 mg/dl than the current drug.

Conditions: \( \mu \) is 

Random: This was a randomized controlled experiment.

Normal: Both samples has fewer than 30 observations, but we were told that no strong skewness or outliers were detected.

Independent: Due to random assignment, these 2 groups are independent. Also knowing one patient's reduction in cholesterol gives no information about another patient's reduction in cholesterol.

TEST: 2-Sample T-TEST for \( M_1 - M_2 \)

\( M_1: \) new drug \( - n_1 = 15 \)
\( \bar{x}_1 = 68.7 \text{ mg/dl} \)
\( s_1 = 13.3 \text{ mg/dl} \)

\( M_2: \) current drug \( - n_2 = 14 \)
\( \bar{x}_2 = 54.1 \text{ mg/dl} \)
\( s_2 = 11.93 \text{ mg/dl} \)

\[ d.f. = 14 \]
\[ d.f. = 13 \text{ (Conservative)} \]

\[ \bar{x}_1 - \bar{x}_2 = 68.7 - 54.1 - (10) = 4.6 \]

\[ t = \frac{4.6}{\sqrt{13.3^2 (1/15) + 11.93^2 (1/14)}} = 4.6/1.686 = 2.71 \]

\[ \mu = 0.982 \]
165 continued  Note—Cannot use Co1 for this ques.

\[ t = 1.982 \quad \text{df} = 13 \]

\[ P \text{ value} = P(t > 1.982) = 2 \cdot \text{cdf}(1.982, 99, 13) = 0.1726 \]

**Conclude:** Since the \( P \text{ value} (0.1726) > 0.05 \), we fail to reject \( H_0 \). We do not have enough evidence to conclude that the mean cholesterol reduction is more than 10 mg/dl more for the new drug than for the current drug.

165B Since we Fail to Reject \( H_0 \) based on \( \alpha = 0.05 \) which means we did not find a Type I error; we could only have the possibility of a Type II error.
57A Researchers randomly assign the subjects to the 2 treatment groups to help balance out the effects of external variables.

57B The actual experiment $X_I - X_E = 4.15$
The graph of the sampling distribution shows about 5 of the 1,000 differences were 4.15 or bigger.

\[
\begin{align*}
\text{Ho: } & \mu_I - \mu_E = 0 \\
\text{Ha: } & \mu_I - \mu_E > 0 \\
\end{align*}
\]

\[
\Pr \left( \bar{X}_I - \bar{X}_E \geq 4.15 \right) = 0.005
\]
Since $P_{\text{value}} = 0.005 < 0.05$
then we reject Ho

Conclude: We would conclude that the mean rating for those with interlunar reasons is significantly higher than those with external reasons.

57C This is a type I error, since we rejected $\text{Ho: } \mu_I - \mu_E = 0$