# SECTION 11.2B Exercises

51. **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.[36]
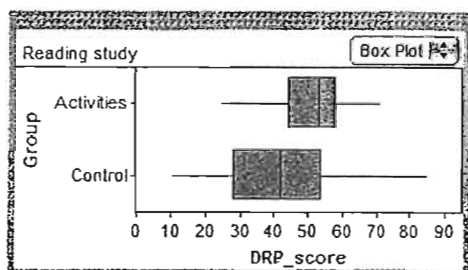
Reading study

	Group	
	Activities	Control
	51.4762	41.5217
DRP_score	21	23
	11.0074	17.1487

S1 = mean ( )
S2 = count ( )
S3 = stdDev ( )

(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).

53. **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.

59. **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.[42]

(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.

65. **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 of blood (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 not $H_0: \mu_1 - \mu_2 = 0$.)

(b) Based on your conclusion in part (a), could you have made a Type I error or a Type II error? Justify your answer.

57. **Rewards and creativity** 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.[40] 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 $\bar{x}_I - \bar{x}_E$.

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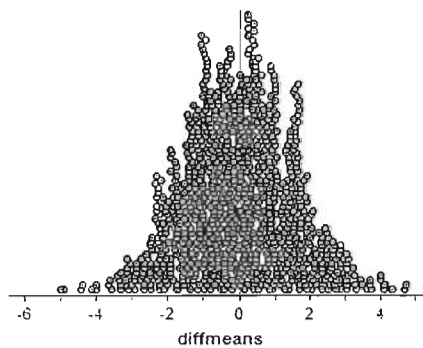
(a) Why did researchers randomly assign the subjects to the two treatment groups?

(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.

|5| GROUPS

WITH ACTIVITIES $n_1 = 21$ $\bar{x}_1 = 51.5$ $S_1 = 11.0$
CONTROL $n_2 = 23$ $\bar{x}_2 = 41.5$ $S_2 = 17.1$

(A) Remember Cuss and BS

The activity groups 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:

μ_1 = actual mean DRP for 3rd graders doing the activity
μ_2 = actual mean DRP for 3rd graders Not doing activity

$H_0: \mu_1 = \mu_2$ (the same) $H_0: \mu_1 - \mu_2 = 0$
$H_A: \mu_1 > \mu_2$ (activity better)   OR $H_A: \mu_1 - \mu_2 > 0$

CONDITIONS

σ IS UNKNOWN (t inference)
Random - this was a randomized comparitive study
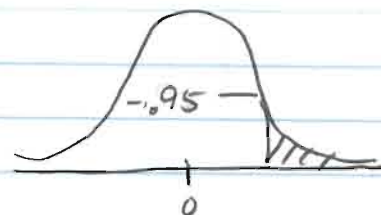Normal - The box plot shows neither outliers or strong skewness
INDEPENDENT - Due to random assignment these 2 Groups can be viewed as independent

51B CONT

TEST - 2 SAMPLE T TEST FOR $\mu_1 - \mu_2$

t^* : $df_1 = 21-1 = 20$ ⟶ Select the conservative $\boxed{df = 20}$

$df_2 = 23 - 1 = 22$ ⟶ $\alpha = .05$

-.95

10.0

Calc

2-Samp TTest

*POOLED-NO

↓

$t = 2.33$ ✓

$p = .0127$ ✓

$df = 37.9$

$$t_{20} = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)^0}{\sqrt{S_1^2/n_1 + S_2^2/n_2}} = \frac{51.5 - 41.5}{\sqrt{\frac{11.0^2}{21} + \frac{17.1^2}{23}}}$$

$$\boxed{t_{20} = 2.33}$$

P value : $P(t_{20} > 2.33) = tcdf(2.33, E99, 20) = \underline{.0152}$

Conclude: Since p-value (.0152) < .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

51C Since this was a randomized controlled experiment, we can conclude that the activities caused an increase in the mean DRP Score.

51D 95% 2 SAMPLE T INTERVAL FOR $\mu_1 - \mu_2$

CALC

2 SAMP TINT

*Pooled: NO

USING FULL Decimals

↓

(1.233, 18.68)

df = 37.9

$$\bar{x}_1 - \bar{x}_2 \pm t^* \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

$$51.5 - 41.5 \pm 2.086 \sqrt{\frac{11^2}{21} + \frac{17.1^2}{23}}$$

10 ± 8.966

$(1.034, 18.966)$

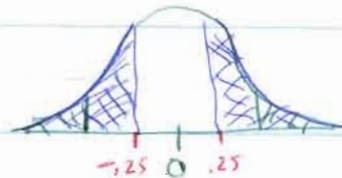$t_{20}^* = INVT(.025, 20) = -2.086$

CONCLUDE: WE ARE 95% Confident the interval 1.03 to 18.97 captures the difference in actual mean DRP Scores for 3RD Graders that do the activity and those that do not.

53A

(μ_1) Women: $n_1 = 56$ $\bar{X}_1 = 16,177$ $S_1 = 7520$
(μ_2) Men: $n_2 = 56$ $\bar{X}_2 = 16569$ $S_2 = 9108$

μ_1 = actual mean number of words spoken by women per day
μ_2 =   "   "   "   "   "   "   by men per day

$H_0: \mu_1 = \mu_2$
$H_A: \mu_1 \neq \mu_2$

-.25   0   .25

Conservative
df = 55

CONDITIONS:
Random: both samples randomly selected
Normal: both samples are over 30 (n=56)
Independent: the groups (men + women) are independent.
   Reasonable to assume there are more than
   560 men and 560 women at the school
σ is UNKNOWN (t inference)

TEST: 2 SAMPLE T-test for $\mu_1 - \mu_2$

$$t_{55} = \frac{16177 - 16569}{\sqrt{\dfrac{7520^2}{56} + \dfrac{9108^2}{56}}} = \frac{-392}{} = -.248 \checkmark$$

(CALC)
2SAMP TTEST
Pooled (NO)
↓
$\mu_1 \neq \mu_2$
$t = -.248 \checkmark$

$P = .80$
$df \approx 106.2$

PVALUE: $P(t \leq -.248)$ or $P(t \geq .248) = \boxed{.8050}$
   tcdf $(-E99, -.248, 55) = .4025 \times 2$ ↗

CONCLUDE: Since the p value (.8050) > .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.

53B

Interpretation of P-value in context:

IF MALES AND FEMALES AT THIS UNIVERSITY SPEAK THE <u>SAME</u> 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.

59   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
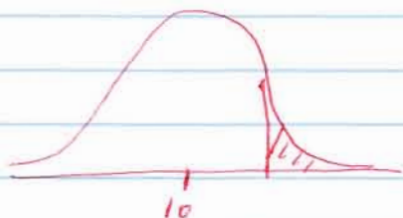MONTH WITH MUSIC — PAIRED T TEST

(c) SUBJECT WERE RANDOMLY ASSIGNED TO 2
GROUPS. ONE GROUP USED DIET A AND
THE OTHER GROUP DIET B — 2 SAMPLE
                         T test.

[65] μ_1 = actual mean cholesterol reduction using new drug
μ_2 = actual " " " using current drug

$H_0: \mu_1 - \mu_2 = 10$
$H_A: \mu_1 - \mu_2 > 10$ mean cholesterol reduction with
                         new drug is more than 10 mg/dl
                         then the current drug

Conditions: μ is unknown (t inference)
    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 viewed as independent. Also, knowing one patients
       reduction in cholesterol gives no information about
       another patients reduction in cholesterol.

TEST: 2 SAMPLE T TEST FOR $\mu_1 - \mu_2$

μ_1: new drug — $n_1 = 15$ $\bar{X}_1 = 68.7 \text{ mg/dl}$ $S_1 = 13.3 \text{ mg/dl}$
μ_2: current drug — $n_2 = 14$ $\bar{X}_2 = 54.1$ $S_2 = 11.93$

$df_1 = 14$
$df_2 = 13$ ← Conservative df   (df = 13)

$$t = \frac{68.7 - 54.1 - (10)}{\sqrt{\frac{(13.3)^2}{15} + \frac{(11.93)^2}{14}}} = \frac{4.6}{4.686} = \frac{4.6}{4.686} = 0.982$$

65 CONTINUED   NOTE — Cannot use Calc for this Ques.

$$t = .982 \qquad df = 13$$

$$P value = P(t > .982) = tcdf(.982, E99, 13) = .1720$$

Conclude: Since the pvalue (.1720) >.05, We fail
to reject Ho. 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

65B   Since we Fail to Reject Ho based on $\alpha = .05$
which means we did not find a Type I
error; we could only have the
possibility of a Type 2 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 $\bar{X}_I - \bar{X}_E = 4.15$
     THE GRAPH OF THE SAMPLING DISTRIBUTION
     SHOWS ABOUT 5 OF THE 1,000 differences
     were 4.15 OR BIGGER

$H_0: \mu_I - \mu_E = 0$
$H_A: \mu_I - \mu_E > 0$

0
diff means
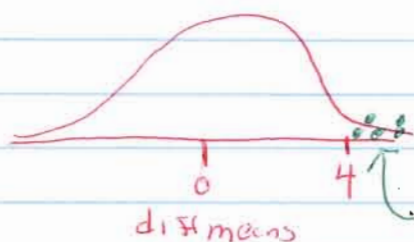
$5/1000 = .005 \longrightarrow P(\bar{X}_I - \bar{X}_E \geq 4.15) = .005$

Since P value $= .005 < .05$
then we Reject H_0

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

57C This is a type I error, since we
     rejected $H_0: \mu_I - \mu_E = 0$