

## 9.E: Two-Sample Problems (Exercises)

These are homework exercises to accompany the Textmap created for "Introductory Statistics" by Shafer and Zhang.

### 9.1: Comparison of Two Population Means: Large, Independent Samples

#### Basic

##### Q9.1.1

Construct the confidence interval for  $\mu_1 - \mu_2$  for the level of confidence and the data from independent samples given.

a. 90% confidence,

$$n_1 = 45, \bar{x}_1 = 27, s_1 = 2 \quad (9.E.1)$$

$$n_2 = 60, \bar{x}_2 = 22, s_2 = 3$$

b. 99% confidence,

$$n_1 = 30, \bar{x}_1 = -112, s_1 = 9 \quad (9.E.2)$$

$$n_2 = 40, \bar{x}_2 = -98, s_2 = 4$$

##### Q9.1.2

Construct the confidence interval for  $\mu_1 - \mu_2$  for the level of confidence and the data from independent samples given.

a. 95% confidence,

$$n_1 = 110, \bar{x}_1 = 77, s_1 = 15 \quad (9.E.3)$$

$$n_2 = 85, \bar{x}_2 = 79, s_2 = 21$$

b. 90% confidence,

$$n_1 = 65, \bar{x}_1 = -83, s_1 = 12 \quad (9.E.4)$$

$$n_2 = 65, \bar{x}_2 = -74, s_2 = 8$$

##### Q9.1.3

Construct the confidence interval for  $\mu_1 - \mu_2$  for the level of confidence and the data from independent samples given.

a. 99.5% confidence,

$$n_1 = 130, \bar{x}_1 = 27.2, s_1 = 2.5 \quad (9.E.5)$$

$$n_2 = 155, \bar{x}_2 = 38.8, s_2 = 4.6$$

b. 95% confidence,

$$n_1 = 68, \bar{x}_1 = 215.5, s_1 = 12.3 \quad (9.E.6)$$

$$n_2 = 84, \bar{x}_2 = 287.8, s_2 = 14.1$$

##### Q9.1.4

Construct the confidence interval for  $\mu_1 - \mu_2$  for the level of confidence and the data from independent samples given.

a. 99.9% confidence,

$$n_1 = 275, \bar{x}_1 = 70.2, s_1 = 1.5 \quad (9.E.7)$$

$$n_2 = 325, \bar{x}_2 = 63.4, s_2 = 1.1$$

b. 90% confidence,

$$n_1 = 120, \bar{x}_1 = 35.5, s_1 = 0.75 \quad (9.E.8)$$

$$n_2 = 146, \bar{x}_2 = 29.6, s_2 = 0.80$$

##### Q9.1.5

Perform the test of hypotheses indicated, using the data from independent samples given. Use the critical value approach. Compute the  $p$ -value of the test as well.

- a. Test  $H_0 : \mu_1 - \mu_2 = 3$  vs  $H_a : \mu_1 - \mu_2 \neq 3$  @  $\alpha = 0.05$

$$\begin{aligned}n_1 &= 35, \bar{x}_1 = 25, s_1 = 1 \\n_2 &= 45, \bar{x}_2 = 19, s_2 = 2\end{aligned}\tag{9.E.9}$$

- b. Test  $H_0 : \mu_1 - \mu_2 = -25$  vs  $H_a : \mu_1 - \mu_2 < -25$  @  $\alpha = 0.10$

$$\begin{aligned}n_1 &= 85, \bar{x}_1 = 188, s_1 = 15 \\n_2 &= 62, \bar{x}_2 = 215, s_2 = 19\end{aligned}\tag{9.E.10}$$

#### Q9.1.6

Perform the test of hypotheses indicated, using the data from independent samples given. Use the critical value approach. Compute the  $p$ -value of the test as well.

- a. Test  $H_0 : \mu_1 - \mu_2 = 45$  vs  $H_a : \mu_1 - \mu_2 > 45$  @  $\alpha = 0.001$

$$\begin{aligned}n_1 &= 200, \bar{x}_1 = 1312, s_1 = 35 \\n_2 &= 225, \bar{x}_2 = 1256, s_2 = 28\end{aligned}\tag{9.E.11}$$

- b. Test  $H_0 : \mu_1 - \mu_2 = -12$  vs  $H_a : \mu_1 - \mu_2 \neq -12$  @  $\alpha = 0.10$

$$\begin{aligned}n_1 &= 35, \bar{x}_1 = 121, s_1 = 6 \\n_2 &= 40, \bar{x}_2 = 135, s_2 = 7\end{aligned}\tag{9.E.12}$$

#### Q9.1.7

Perform the test of hypotheses indicated, using the data from independent samples given. Use the critical value approach. Compute the  $p$ -value of the test as well.

- a. Test  $H_0 : \mu_1 - \mu_2 = 0$  vs  $H_a : \mu_1 - \mu_2 \neq 0$  @  $\alpha = 0.01$

$$\begin{aligned}n_1 &= 125, \bar{x}_1 = -46, s_1 = 10 \\n_2 &= 90, \bar{x}_2 = -50, s_2 = 13\end{aligned}\tag{9.E.13}$$

- b. Test  $H_0 : \mu_1 - \mu_2 = 20$  vs  $H_a : \mu_1 - \mu_2 > 20$  @  $\alpha = 0.05$

$$\begin{aligned}n_1 &= 40, \bar{x}_1 = 142, s_1 = 11 \\n_2 &= 40, \bar{x}_2 = 118, s_2 = 10\end{aligned}\tag{9.E.14}$$

#### Q9.1.8

Perform the test of hypotheses indicated, using the data from independent samples given. Use the critical value approach. Compute the  $p$ -value of the test as well.

- a. Test  $H_0 : \mu_1 - \mu_2 = 13$  vs  $H_a : \mu_1 - \mu_2 < 13$  @  $\alpha = 0.01$

$$\begin{aligned}n_1 &= 35, \bar{x}_1 = 100, s_1 = 2 \\n_2 &= 35, \bar{x}_2 = 88, s_2 = 2\end{aligned}\tag{9.E.15}$$

- b. Test  $H_0 : \mu_1 - \mu_2 = -10$  vs  $H_a : \mu_1 - \mu_2 \neq -10$  @  $\alpha = 0.10$

$$\begin{aligned}n_1 &= 146, \bar{x}_1 = 62, s_1 = 4 \\n_2 &= 120, \bar{x}_2 = 73, s_2 = 7\end{aligned}\tag{9.E.16}$$

#### Q9.1.9

Perform the test of hypotheses indicated, using the data from independent samples given. Use the  $p$ -value approach.

- a. Test  $H_0 : \mu_1 - \mu_2 = 57$  vs  $H_a : \mu_1 - \mu_2 < 57$  @  $\alpha = 0.10$

$$\begin{aligned}n_1 &= 117, \bar{x}_1 = 1309, s_1 = 42 \\n_2 &= 133, \bar{x}_2 = 1258, s_2 = 37\end{aligned}\tag{9.E.17}$$

- b. Test  $H_0 : \mu_1 - \mu_2 = -1.5$  vs  $H_a : \mu_1 - \mu_2 \neq -1.5$  @  $\alpha = 0.20$

$$\begin{aligned}n_1 &= 65, \bar{x}_1 = 16.9, s_1 = 1.3 \\n_2 &= 57, \bar{x}_2 = 18.6, s_2 = 1.1\end{aligned}\tag{9.E.18}$$

### Q9.1.10

Perform the test of hypotheses indicated, using the data from independent samples given. Use the  $p$ -value approach.

- a. Test  $H_0 : \mu_1 - \mu_2 = -10.5$  vs  $H_a : \mu_1 - \mu_2 > -10.5$  @  $\alpha = 0.01$

$$\begin{aligned} n_1 &= 64, \bar{x}_1 = 85.6, s_1 = 2.4 \\ n_2 &= 50, \bar{x}_2 = 95.3, s_2 = 3.1 \end{aligned} \quad (9.E.19)$$

- b. Test  $H_0 : \mu_1 - \mu_2 = 110$  vs  $H_a : \mu_1 - \mu_2 \neq 110$  @  $\alpha = 0.02$

$$\begin{aligned} n_1 &= 176, \bar{x}_1 = 1918, s_1 = 68 \\ n_2 &= 241, \bar{x}_2 = 1782, s_2 = 146 \end{aligned} \quad (9.E.20)$$

### Q9.1.11

Perform the test of hypotheses indicated, using the data from independent samples given. Use the  $p$ -value approach.

- a. Test  $H_0 : \mu_1 - \mu_2 = 50$  vs  $H_a : \mu_1 - \mu_2 > 50$  @  $\alpha = 0.005$

$$\begin{aligned} n_1 &= 72, \bar{x}_1 = 272, s_1 = 26 \\ n_2 &= 103, \bar{x}_2 = 213, s_2 = 14 \end{aligned} \quad (9.E.21)$$

- b. Test  $H_0 : \mu_1 - \mu_2 = 7.5$  vs  $H_a : \mu_1 - \mu_2 \neq 7.5$  @  $\alpha = 0.10$

$$\begin{aligned} n_1 &= 52, \bar{x}_1 = 94.3, s_1 = 2.6 \\ n_2 &= 38, \bar{x}_2 = 88.6, s_2 = 8.0 \end{aligned} \quad (9.E.22)$$

### Q9.1.12

Perform the test of hypotheses indicated, using the data from independent samples given. Use the  $p$ -value approach.

- a. Test  $H_0 : \mu_1 - \mu_2 = 23$  vs  $H_a : \mu_1 - \mu_2 < 23$  @  $\alpha = 0.20$

$$\begin{aligned} n_1 &= 314, \bar{x}_1 = 198, s_1 = 12.2 \\ n_2 &= 220, \bar{x}_2 = 176, s_2 = 11.5 \end{aligned} \quad (9.E.23)$$

- b. Test  $H_0 : \mu_1 - \mu_2 = 4.4$  vs  $H_a : \mu_1 - \mu_2 \neq 4.4$  @  $\alpha = 0.05$

$$\begin{aligned} n_1 &= 32, \bar{x}_1 = 40.3, s_1 = 0.5 \\ n_2 &= 30, \bar{x}_2 = 35.5, s_2 = 0.7 \end{aligned} \quad (9.E.24)$$

## Applications

### Q9.1.13

In order to investigate the relationship between mean job tenure in years among workers who have a bachelor's degree or higher and those who do not, random samples of each type of worker were taken, with the following results.

	n	$\bar{x}$	s
Bachelor's degree or higher	155	5.2	1.3
No degree	210	5.0	1.5

- Construct the 99% confidence interval for the difference in the population means based on these data.
- Test, at the 1% level of significance, the claim that mean job tenure among those with higher education is greater than among those without, against the default that there is no difference in the means.
- Compute the observed significance of the test.

### Q9.1.14

Records of 40 used passenger cars and 40 used pickup trucks (none used commercially) were randomly selected to investigate whether there was any difference in the mean time in years that they were kept by the original owner before being sold. For cars the mean was 5.3 years with standard deviation 2.2 years. For pickup trucks the mean was 7.1 years with standard deviation 3.0 years.

- Construct the 95% confidence interval for the difference in the means based on these data.

- b. Test the hypothesis that there is a difference in the means against the null hypothesis that there is no difference. Use the 1% level of significance.
- c. Compute the observed significance of the test in part (b).

#### Q9.1.15

In previous years the average number of patients per hour at a hospital emergency room on weekends exceeded the average on weekdays by 6.3 visits per hour. A hospital administrator believes that the current weekend mean exceeds the weekday mean by fewer than 6.3 hours.

- a. Construct the 99% confidence interval for the difference in the population means based on the following data, derived from a study in which 30 weekend and 30 weekday one-hour periods were randomly selected and the number of new patients in each recorded.

	n	$\bar{x}$	s
Weekends	30	13.8	3.1
Weekdays	30	8.6	2.7

- a. Test at the 5% level of significance whether the current weekend mean exceeds the weekday mean by fewer than 6.3 patients per hour.
- b. Compute the observed significance of the test.

#### Q9.1.16

A sociologist surveys 50 randomly selected citizens in each of two countries to compare the mean number of hours of volunteer work done by adults in each. Among the 50 inhabitants of Lilliput, the mean hours of volunteer work per year was 52, with standard deviation 11.8. Among the 50 inhabitants of Blefuscu, the mean number of hours of volunteer work per year was 37, with standard deviation 7.2.

- a. Construct the 99% confidence interval for the difference in mean number of hours volunteered by all residents of Lilliput and the mean number of hours volunteered by all residents of Blefuscu.
- b. Test, at the 1% level of significance, the claim that the mean number of hours volunteered by all residents of Lilliput is more than ten hours greater than the mean number of hours volunteered by all residents of Blefuscu.
- c. Compute the observed significance of the test in part (b).

#### Q9.1.17

A university administrator asserted that upperclassmen spend more time studying than underclassmen.

- a. Test this claim against the default that the average number of hours of study per week by the two groups is the same, using the following information based on random samples from each group of students. Test at the 1% level of significance.

	n	$\bar{x}$	s
Upperclassmen	35	15.6	2.9
Underclassmen	35	12.3	4.1

- b. Compute the observed significance of the test.

#### Q9.1.18

An kinesologist claims that the resting heart rate of men aged 18 to 25 who exercise regularly is more than five beats per minute less than that of men who do not exercise regularly. Men in each category were selected at random and their resting heart rates were measured, with the results shown.

	n	$\bar{x}$	s
Regular exercise	40	63	1.0

	$n$	$\bar{x}$	$s$
No regular exercise	30	71	1.2

- Perform the relevant test of hypotheses at the 1% level of significance.
- Compute the observed significance of the test.

#### Q9.1.19

Children in two elementary school classrooms were given two versions of the same test, but with the order of questions arranged from easier to more difficult in Version *A* and in reverse order in Version *B*. Randomly selected students from each class were given Version *A* and the rest Version *B*. The results are shown in the table.

	$n$	$\bar{x}$	$s$
Version A	31	83	4.6
Version B	32	78	4.3

- Construct the 90% confidence interval for the difference in the means of the populations of all children taking Version *A* of such a test and of all children taking Version *B* of such a test.
- Test at the 1% level of significance the hypothesis that the *A* version of the test is easier than the *B* version (even though the questions are the same).
- Compute the observed significance of the test.

#### Q9.1.20

The Municipal Transit Authority wants to know if, on weekdays, more passengers ride the northbound blue line train towards the city center that departs at 8 : 15 *a. m.* or the one that departs at 8 : 30 *a. m.* The following sample statistics are assembled by the Transit Authority.

	$n$	$\bar{x}$	$s$
8:15 a.m. train	30	323	41
8:30 a.m. train	45	356	45

- Construct the 90% confidence interval for the difference in the mean number of daily travelers on the 8 : 15 *a. m.* train and the mean number of daily travelers on the 8 : 30 *a. m.* train.
- Test at the 5% level of significance whether the data provide sufficient evidence to conclude that more passengers ride the 8 : 30 *a. m.* train.
- Compute the observed significance of the test.

#### Q9.1.21

In comparing the academic performance of college students who are affiliated with fraternities and those male students who are unaffiliated, a random sample of students was drawn from each of the two populations on a university campus. Summary statistics on the student GPAs are given below.

	$n$	$\bar{x}$	$s$
Fraternity	645	2.90	0.47
Unaffiliated	450	2.88	0.42

Test, at the 5% level of significance, whether the data provide sufficient evidence to conclude that there is a difference in average GPA between the population of fraternity students and the population of unaffiliated male students on this university campus.

### Q9.1.22

In comparing the academic performance of college students who are affiliated with sororities and those female students who are unaffiliated, a random sample of students was drawn from each of the two populations on a university campus. Summary statistics on the student GPAs are given below.

	n	$\bar{x}$	s
Sorority	330	3.18	0.37
Unaffiliated	550	3.12	0.41

Test, at the 5% level of significance, whether the data provide sufficient evidence to conclude that there is a difference in average GPA between the population of sorority students and the population of unaffiliated female students on this university campus.

### Q9.1.23

The owner of a professional football team believes that the league has become more offense oriented since five years ago. To check his belief, 32 randomly selected games from one year's schedule were compared to 32 randomly selected games from the schedule five years later. Since more offense produces more points per game, the owner analyzed the following information on points per game (ppg).

	n	$\bar{x}$	s
ppg previously	32	20.62	4.17
ppg recently	32	22.05	4.01

Test, at the 10% level of significance, whether the data on points per game provide sufficient evidence to conclude that the game has become more offense oriented.

### Q9.1.24

The owner of a professional football team believes that the league has become more offense oriented since five years ago. To check his belief, 32 randomly selected games from one year's schedule were compared to 32 randomly selected games from the schedule five years later. Since more offense produces more offensive yards per game, the owner analyzed the following information on offensive yards per game (oypg).

	n	$\bar{x}$	s
oypg previously	32	316	40
oypg recently	32	336	35

Test, at the 10% level of significance, whether the data on offensive yards per game provide sufficient evidence to conclude that the game has become more offense oriented.

## Large Data Set Exercises

### Large Data Sets are absent

25. Large Data Sets 1A and 1B list the SAT scores for 1,000 randomly selected students. Denote the population of all male students as Population 1 and the population of all female students as Population 2.
  - a. Restricting attention to just the males, find  $n_1$ ,  $\bar{x}_1$  and  $s_1$ . Restricting attention to just the females, find  $n_2$ ,  $\bar{x}_2$  and  $s_2$ .
  - b. Let  $\mu_1$  denote the mean SAT score for all males and  $\mu_2$  the mean SAT score for all females. Use the results of part (a) to construct a 90% confidence interval for the difference  $\mu_1 - \mu_2$ .
  - c. Test, at the 5% level of significance, the hypothesis that the mean SAT scores among males exceeds that of females.
26. Large Data Sets 1A and 1B list the SAT scores for 1,000 randomly selected students. Denote the population of all male students as Population 1 and the population of all female students as Population 2.
  - a. Restricting attention to just the males, find  $n_1$ ,  $\bar{x}_1$  and  $s_1$ . Restricting attention to just the females, find  $n_2$ ,  $\bar{x}_2$  and  $s_2$ .

- b. Let  $\mu_1$  denote the mean SAT score for all males and  $\mu_2$  the mean SAT score for all females. Use the results of part (a) to construct a 95% confidence interval for the difference  $\mu_1 - \mu_2$ .
  - c. Test, at the 10% level of significance, the hypothesis that the mean SAT scores among males exceeds that of females.
27. Large Data Sets 7A and 7B list the survival times for 65 male and 75 female laboratory mice with thymic leukemia. Denote the population of all such male mice as Population 1 and the population of all such female mice as Population 2.
- a. Restricting attention to just the males, find  $n_1$ ,  $\bar{x}_1$  and  $s_1$ . Restricting attention to just the females, find  $n_2$ ,  $\bar{x}_2$  and  $s_2$ .
  - b. Let  $\mu_1$  denote the mean survival for all males and  $\mu_2$  the mean survival time for all females. Use the results of part (a) to construct a 99% confidence interval for the difference  $\mu_1 - \mu_2$ .
  - c. Test, at the 1% level of significance, the hypothesis that the mean survival time for males exceeds that for females by more than 182 days (half a year).
  - d. Compute the observed significance of the test in part (c).

### Answers

1. a. (4.20, 5.80)  
b. (-18.54, -9.46)
- 2.
3. a. (-12.81, -10.39)  
b. (-76.50, -68.10)
- 4.
5. a.  $Z = 8.753$ ,  $\pm z_{0.025} = \pm 1.960$ , reject  $H_0$ ,  $p\text{-value} = 0.0000$   
b.  $Z = -0.687$ ,  $-z_{0.10} = -1.282$ , do not reject  $H_0$ ,  $p\text{-value} = 0.2451$
- 6.
7. a.  $Z = 2.444$ ,  $\pm z_{0.005} = \pm 2.576$ , do not reject  $H_0$ ,  $p\text{-value} = 0.0146$   
b.  $Z = 1.702$ ,  $z_{0.05} = -1.645$ , reject  $H_0$ ,  $p\text{-value} = 0.0446$
- 8.
9. a.  $Z = -1.19$ ,  $p\text{-value} = 0.1170$  do not reject  $H_0$   
b.  $Z = -0.92$ ,  $p\text{-value} = 0.3576$  do not reject  $H_0$
- 10.
11. a.  $Z = 2.68$ ,  $p\text{-value} = 0.0037$ , reject  $H_0$   
b.  $Z = -1.34$ ,  $p\text{-value} = 0.1802$  do not reject  $H_0$
- 12.
13. a.  $0.2 \pm 0.4$   
b.  $Z = -1.466$ ,  $-z_{0.050} = -1.645$ , do not reject  $H_0$  (exceeds by 6.3 or more)  
c.  $p\text{-value} = 0.0869$
- 14.
15. a.  $5.2 \pm 1.9$   
b.  $Z = -1.466$ ,  $-z_{0.050} = -1.645$ , do not reject  $H_0$  (exceeds by 6.3 or more)  
c.  $p\text{-value} = 0.0708$
- 16.
17. a.  $Z = 3.888$ ,  $z_{0.01} = 2.326$ , reject  $H_0$  (upperclassmen study more)  
b.  $p\text{-value} = 0.0001$
- 18.
19. a.  $5 \pm 1.8$   
b.  $Z = 4.454$ ,  $z_{0.01} = 2.326$ , reject  $H_0$  (Test A is easier)  
c.  $p\text{-value} = 0.0000$
- 20.
21.  $Z = 0.738$ ,  $\pm z_{0.025} = \pm 1.960$ , do not reject  $H_0$  (no difference)
- 22.
23.  $Z = -1.398$ ,  $-z_{0.10} = -1.282$ , reject  $H_0$  (more offense oriented)

- 24.
25. a.  $n_1 = 419$ ,  $\bar{x}_1 = 1540.33$ ,  $s_1 = 205.40$ ,  $n_2 = 581$ ,  $\bar{x}_2 = 1520.38$ ,  $s_2 = 217.34$   
 b.  $(-2.24, 42.15)$   
 c.  $H_0 : \mu_1 - \mu_2 = 0$  vs  $H_a : \mu_1 - \mu_2 > 0$  . Test Statistic:  $Z = 1.48$ . Rejection Region:  $[1.645, \infty)$  Decision: Fail to reject  $H_0$ .
- 26.
27. a.  $n_1 = 65$ ,  $\bar{x}_1 = 665.97$ ,  $s_1 = 41.60$ ,  $n_2 = 75$ ,  $\bar{x}_2 = 455.89$ ,  $s_2 = 63.22$   
 b.  $(187.06, 233.09)$   
 c.  $H_0 : \mu_1 - \mu_2 = 182$  vs  $H_a : \mu_1 - \mu_2 > 182$  . Test Statistic:  $Z = 3.14$ . Rejection Region:  $[2.33, \infty)$ . Decision: Reject  $H_0$ .  
 d.  $p\text{-value} = 0.0008$

## 9.2: Comparison of Two Population Means: Small, Independent Samples

### Basic

In all exercises for this section assume that the populations are normal and have equal standard deviations.

#### Q9.2.1

Construct the confidence interval for  $\mu_1 - \mu_2$  for the level of confidence and the data from independent samples given.

- a. 95% confidence,

$$\begin{aligned} n_1 &= 10, \bar{x}_1 = 120, s_1 = 2 \\ n_2 &= 15, \bar{x}_2 = 101, s_1 = 4 \end{aligned} \quad (9.E.25)$$

- b. 99% confidence,

$$\begin{aligned} n_1 &= 6, \bar{x}_1 = 25, s_1 = 1 \\ n_2 &= 12, \bar{x}_2 = 17, s_1 = 3 \end{aligned} \quad (9.E.26)$$

#### Q9.2.2

Construct the confidence interval for  $\mu_1 - \mu_2$  for the level of confidence and the data from independent samples given.

- a. 90% confidence,

$$\begin{aligned} n_1 &= 28, \bar{x}_1 = 212, s_1 = 6 \\ n_2 &= 23, \bar{x}_2 = 198, s_1 = 5 \end{aligned} \quad (9.E.27)$$

- b. 99% confidence,

$$\begin{aligned} n_1 &= 14, \bar{x}_1 = 68, s_1 = 8 \\ n_2 &= 20, \bar{x}_2 = 43, s_1 = 3 \end{aligned} \quad (9.E.28)$$

#### Q9.2.3

Construct the confidence interval for  $\mu_1 - \mu_2$  for the level of confidence and the data from independent samples given.

- a. 99.9% confidence,

$$\begin{aligned} n_1 &= 35, \bar{x}_1 = 6.5, s_1 = 0.2 \\ n_2 &= 20, \bar{x}_2 = 6.2, s_1 = 0.1 \end{aligned} \quad (9.E.29)$$

- b. 99% confidence,

$$\begin{aligned} n_1 &= 18, \bar{x}_1 = 77.3, s_1 = 1.2 \\ n_2 &= 32, \bar{x}_2 = 75.0, s_1 = 1.6 \end{aligned} \quad (9.E.30)$$

#### Q9.2.4

Construct the confidence interval for  $\mu_1 - \mu_2$  for the level of confidence and the data from independent samples given.

- a. 99.5% confidence,

$$\begin{aligned} n_1 &= 40, \bar{x}_1 = 85.6, s_1 = 2.8 \\ n_2 &= 20, \bar{x}_2 = 73.1, s_1 = 2.1 \end{aligned} \quad (9.E.31)$$

b. 99.9% confidence,

$$\begin{aligned} n_1 &= 25, \bar{x}_1 = 215, s_1 = 7 \\ n_2 &= 35, \bar{x}_2 = 185, s_1 = 12 \end{aligned} \quad (9.E.32)$$

### Q9.2.5

Perform the test of hypotheses indicated, using the data from independent samples given. Use the critical value approach.

a. Test  $H_0 : \mu_1 - \mu_2 = 11$  vs  $H_a : \mu_1 - \mu_2 > 11$  @  $\alpha = 0.025$

$$\begin{aligned} n_1 &= 6, \bar{x}_1 = 32, s_1 = 2 \\ n_2 &= 11, \bar{x}_2 = 19, s_1 = 1 \end{aligned} \quad (9.E.33)$$

b. Test  $H_0 : \mu_1 - \mu_2 = 26$  vs  $H_a : \mu_1 - \mu_2 \neq 26$  @  $\alpha = 0.05$

$$\begin{aligned} n_1 &= 17, \bar{x}_1 = 166, s_1 = 4 \\ n_2 &= 24, \bar{x}_2 = 138, s_1 = 3 \end{aligned} \quad (9.E.34)$$

### Q9.2.6

Perform the test of hypotheses indicated, using the data from independent samples given. Use the critical value approach.

a. Test  $H_0 : \mu_1 - \mu_2 = 40$  vs  $H_a : \mu_1 - \mu_2 < 40$  @  $\alpha = 0.10$

$$\begin{aligned} n_1 &= 14, \bar{x}_1 = 289, s_1 = 11 \\ n_2 &= 12, \bar{x}_2 = 254, s_1 = 9 \end{aligned} \quad (9.E.35)$$

b. Test  $H_0 : \mu_1 - \mu_2 = 21$  vs  $H_a : \mu_1 - \mu_2 \neq 21$  @  $\alpha = 0.05$

$$\begin{aligned} n_1 &= 23, \bar{x}_1 = 130, s_1 = 6 \\ n_2 &= 27, \bar{x}_2 = 113, s_1 = 8 \end{aligned} \quad (9.E.36)$$

### Q9.2.7

Perform the test of hypotheses indicated, using the data from independent samples given. Use the critical value approach.

a. Test  $H_0 : \mu_1 - \mu_2 = -15$  vs  $H_a : \mu_1 - \mu_2 < -15$  @  $\alpha = 0.10$

$$\begin{aligned} n_1 &= 30, \bar{x}_1 = 42, s_1 = 7 \\ n_2 &= 12, \bar{x}_2 = 60, s_1 = 5 \end{aligned} \quad (9.E.37)$$

b. Test  $H_0 : \mu_1 - \mu_2 = 103$  vs  $H_a : \mu_1 - \mu_2 \neq 103$  @  $\alpha = 0.10$

$$\begin{aligned} n_1 &= 17, \bar{x}_1 = 711, s_1 = 28 \\ n_2 &= 32, \bar{x}_2 = 598, s_1 = 21 \end{aligned} \quad (9.E.38)$$

### Q9.2.8

Perform the test of hypotheses indicated, using the data from independent samples given. Use the critical value approach.

a. Test  $H_0 : \mu_1 - \mu_2 = 75$  vs  $H_a : \mu_1 - \mu_2 > 75$  @  $\alpha = 0.025$

$$\begin{aligned} n_1 &= 45, \bar{x}_1 = 674, s_1 = 18 \\ n_2 &= 29, \bar{x}_2 = 591, s_1 = 13 \end{aligned} \quad (9.E.39)$$

b. Test  $H_0 : \mu_1 - \mu_2 = -20$  vs  $H_a : \mu_1 - \mu_2 \neq -20$  @  $\alpha = 0.005$

$$\begin{aligned} n_1 &= 30, \bar{x}_1 = 137, s_1 = 8 \\ n_2 &= 19, \bar{x}_2 = 166, s_1 = 11 \end{aligned} \quad (9.E.40)$$

### Q9.2.9

Perform the test of hypotheses indicated, using the data from independent samples given. Use the  $p$ -value approach. (The  $p$ -value can be only approximated.)

a. Test  $H_0 : \mu_1 - \mu_2 = 12$  vs  $H_a : \mu_1 - \mu_2 > 12$  @  $\alpha = 0.01$

$$\begin{aligned} n_1 &= 20, \bar{x}_1 = 133, s_1 = 7 \\ n_2 &= 10, \bar{x}_2 = 115, s_1 = 5 \end{aligned} \quad (9.E.41)$$

b. Test  $H_0 : \mu_1 - \mu_2 = 46$  vs  $H_a : \mu_1 - \mu_2 \neq 46$  @  $\alpha = 0.10$

$$\begin{aligned} n_1 &= 24, \bar{x}_1 = 586, s_1 = 11 \\ n_2 &= 27, \bar{x}_2 = 535, s_1 = 13 \end{aligned} \quad (9.E.42)$$

### Q9.2.10

Perform the test of hypotheses indicated, using the data from independent samples given. Use the  $p$ -value approach. (The  $p$ -value can be only approximated.)

a. Test  $H_0 : \mu_1 - \mu_2 = 38$  vs  $H_a : \mu_1 - \mu_2 < 38$  @  $\alpha = 0.01$

$$\begin{aligned} n_1 &= 12, \bar{x}_1 = 464, s_1 = 5 \\ n_2 &= 10, \bar{x}_2 = 432, s_1 = 6 \end{aligned} \quad (9.E.43)$$

b. Test  $H_0 : \mu_1 - \mu_2 = 4$  vs  $H_a : \mu_1 - \mu_2 \neq 4$  @  $\alpha = 0.005$

$$\begin{aligned} n_1 &= 14, \bar{x}_1 = 68, s_1 = 2 \\ n_2 &= 17, \bar{x}_2 = 67, s_1 = 3 \end{aligned} \quad (9.E.44)$$

### Q9.2.11

Perform the test of hypotheses indicated, using the data from independent samples given. Use the  $p$ -value approach. (The  $p$ -value can be only approximated.)

a. Test  $H_0 : \mu_1 - \mu_2 = 50$  vs  $H_a : \mu_1 - \mu_2 > 50$  @  $\alpha = 0.01$

$$\begin{aligned} n_1 &= 30, \bar{x}_1 = 681, s_1 = 8 \\ n_2 &= 27, \bar{x}_2 = 625, s_1 = 8 \end{aligned} \quad (9.E.45)$$

b. Test  $H_0 : \mu_1 - \mu_2 = 35$  vs  $H_a : \mu_1 - \mu_2 \neq 35$  @  $\alpha = 0.10$

$$\begin{aligned} n_1 &= 36, \bar{x}_1 = 325, s_1 = 11 \\ n_2 &= 29, \bar{x}_2 = 286, s_1 = 7 \end{aligned} \quad (9.E.46)$$

### Q9.2.12

Perform the test of hypotheses indicated, using the data from independent samples given. Use the  $p$ -value approach. (The  $p$ -value can be only approximated.)

a. Test  $H_0 : \mu_1 - \mu_2 = -4$  vs  $H_a : \mu_1 - \mu_2 < -4$  @  $\alpha = 0.05$

$$\begin{aligned} n_1 &= 40, \bar{x}_1 = 80, s_1 = 5 \\ n_2 &= 25, \bar{x}_2 = 87, s_1 = 5 \end{aligned} \quad (9.E.47)$$

b. Test  $H_0 : \mu_1 - \mu_2 = 21$  vs  $H_a : \mu_1 - \mu_2 \neq 21$  @  $\alpha = 0.01$

$$\begin{aligned} n_1 &= 15, \bar{x}_1 = 192, s_1 = 12 \\ n_2 &= 34, \bar{x}_2 = 180, s_1 = 8 \end{aligned} \quad (9.E.48)$$

## Applications

### Q9.2.13

A county environmental agency suspects that the fish in a particular polluted lake have elevated mercury level. To confirm that suspicion, five striped bass in that lake were caught and their tissues were tested for mercury. For the purpose of comparison, four striped bass in an unpolluted lake were also caught and tested. The fish tissue mercury levels in mg/kg are given below.

Sample 1 (from polluted lake)	Sample 2 (from unpolluted lake)
0.580	0.382

Sample 1 (from polluted lake)	Sample 2(from unpolluted lake)
0.711	0.276
0.571	0.570
0.666	0.366
0.598	

- Construct the 95% confidence interval for the difference in the population means based on these data.
- Test, at the 5% level of significance, whether the data provide sufficient evidence to conclude that fish in the polluted lake have elevated levels of mercury in their tissue.

#### Q9.2.14

A genetic engineering company claims that it has developed a genetically modified tomato plant that yields on average more tomatoes than other varieties. A farmer wants to test the claim on a small scale before committing to a full-scale planting. Ten genetically modified tomato plants are grown from seeds along with ten other tomato plants. At the season's end, the resulting yields in pound are recorded as below.

Sample 1(genetically modified)	Sample 2(regular)
20	21
23	21
27	22
25	18
25	20
25	20
27	18
23	25
24	23
22	20

- Construct the 99% confidence interval for the difference in the population means based on these data.
- Test, at the 1% level of significance, whether the data provide sufficient evidence to conclude that the mean yield of the genetically modified variety is greater than that for the standard variety.

#### Q9.2.15

The coaching staff of a professional football team believes that the rushing offense has become increasingly potent in recent years. To investigate this belief, 20 randomly selected games from one year's schedule were compared to 11 randomly selected games from the schedule five years later. The sample information on rushing yards per game (rypg) is summarized below.

	n	$\bar{x}$	s
rypg previously	20	112	24
rypg recently	11	114	21

- Construct the 95% confidence interval for the difference in the population means based on these data.
- Test, at the 5% level of significance, whether the data on rushing yards per game provide sufficient evidence to conclude that the rushing offense has become more potent in recent years.

### Q9.2.16

The coaching staff of professional football team believes that the rushing offense has become increasingly potent in recent years. To investigate this belief, 20 randomly selected games from one year's schedule were compared to 11 randomly selected games from the schedule five years later. The sample information on passing yards per game (pypg) is summarized below.

	n	$\bar{x}$	s
pypg previously	20	203	38
pypg recently	11	232	33

- Construct the 95% confidence interval for the difference in the population means based on these data.
- Test, at the 5% level of significance, whether the data on passing yards per game provide sufficient evidence to conclude that the passing offense has become more potent in recent years.

### Q9.2.17

A university administrator wishes to know if there is a difference in average starting salary for graduates with master's degrees in engineering and those with master's degrees in business. Fifteen recent graduates with master's degree in engineering and 11 with master's degrees in business are surveyed and the results are summarized below.

	n	$\bar{x}$	s
Engineering	15	68,535	1627
Business	11	63,230	2033

- Construct the 90% confidence interval for the difference in the population means based on these data.
- Test, at the 10% level of significance, whether the data provide sufficient evidence to conclude that the average starting salaries are different.

### Q9.2.18

A gardener sets up a flower stand in a busy business district and sells bouquets of assorted fresh flowers on weekdays. To find a more profitable pricing, she sells bouquets for 15 dollars each for ten days, then for 10 dollars each for five days. Her average daily profit for the two different prices are given below.

	n	$\bar{x}$	s
\$15	10	171	26
\$10	5	198	29

- Construct the 90% confidence interval for the difference in the population means based on these data.
- Test, at the 10% level of significance, whether the data provide sufficient evidence to conclude the gardener's average daily profit will be higher if the bouquets are sold at \$10 each.

### Answers

- (16.16, 21.84)
  - (4.28, 11.72)
- 
- (0.13, 0.47)
  -
- 
- $T = 2.787$ ,  $t_{0.025} = 2.131$ , reject  $H_0$
  - $T = 1.831$ ,  $\pm t_{0.025} = \pm 2.023$ , do not reject  $H_0$
-

7. a.  $T = -1.349$ ,  $-t_{0.10} = -1.303$ , reject  $H_0$   
b.  $T = 1.411$ ,  $\pm t_{0.05} = \pm 1.678$ , do not reject  $H_0$
- 8.
9. a.  $T = 2.411$ ,  $df = 28$ , p-value  $> 0.01$ , do not reject  $H_0$   
b.  $T = 1.473$ ,  $df = 49$ , p-value  $< 0.10$ , reject  $H_0$
- 10.
11. a.  $T = 2.827$ ,  $df = 55$ , p-value  $< 0.01$ , reject  $H_0$   
b.  $T = 1.699$ ,  $df = 63$ , p-value  $< 0.10$ , reject  $H_0$
- 12.
13. a.  $0.2267 \pm 0.2182$   
b.  $T = 1.699$ ,  $df = 63$ ,  $t_{0.05} = 1.895$ , reject  $H_0$  (elevated levels)
- 14.
15. a.  $-2 \pm 17.7$   
b.  $T = -0.232$ ,  $df = 29$ ,  $-t_{0.05} = -1.699$ , do not reject  $H_0$  (not more potent)
- 16.
17. a.  $5305 \pm 1227$   
b.  $T = 7.395$ ,  $df = 24$ ,  $\pm t_{0.05} = \pm 1.711$ , reject  $H_0$  (different)

### 9.3 Comparison of Two Population Means: Paired Samples

#### Basic

In all exercises for this section assume that the population of differences is normal.

1. Use the following paired sample data for this exercise.

<i>Population 1</i>	35	32	35	35	36	35	35
<i>Population 2</i>	28	26	27	26	29	27	29

(9.E.49)

- a. Compute  $\bar{d}$  and  $s_d$ .
  - b. Give a point estimate for  $\mu_1 - \mu_2 = \mu_d$ .
  - c. Construct the 95% confidence interval for  $\mu_1 - \mu_2 = \mu_d$  from these data.
  - d. Test, at the 10% level of significance, the hypothesis that  $\mu_1 - \mu_2 > 7$  as an alternative to the null hypothesis that  $\mu_1 - \mu_2 = 7$ .
2. Use the following paired sample data for this exercise.

<i>Population 1</i>	103	127	96	110
<i>Population 2</i>	81	106	73	88
<i>Population 1</i>	90	118	130	106
<i>Population 2</i>	70	95	109	83

(9.E.50)

- a. Compute  $\bar{d}$  and  $s_d$ .
  - b. Give a point estimate for  $\mu_1 - \mu_2 = \mu_d$ .
  - c. Construct the 90% confidence interval for  $\mu_1 - \mu_2 = \mu_d$  from these data.
  - d. Test, at the 1% level of significance, the hypothesis that  $\mu_1 - \mu_2 < 247$  as an alternative to the null hypothesis that  $\mu_1 - \mu_2 = 24$ .
3. Use the following paired sample data for this exercise.

<i>Population 1</i>	40	27	55	34
<i>Population 2</i>	53	42	68	50

(9.E.51)

- a. Compute  $\bar{d}$  and  $s_d$ .
- b. Give a point estimate for  $\mu_1 - \mu_2 = \mu_d$ .
- c. Construct the 99% confidence interval for  $\mu_1 - \mu_2 = \mu_d$  from these data.

- d. Test, at the 10% level of significance, the hypothesis that  $\mu_1 - \mu_2 \neq -12$  as an alternative to the null hypothesis that  $\mu_1 - \mu_2 = -12$ .
4. Use the following paired sample data for this exercise.

Population 1	196	165	181	201	190
Population 2	212	182	199	210	205

(9.E.52)

- Compute  $\bar{d}$  and  $s_d$ .
- Give a point estimate for  $\mu_1 - \mu_2 = \mu_d$ .
- Construct the 98% confidence interval for  $\mu_1 - \mu_2 = \mu_d$  from these data.
- Test, at the 2% level of significance, the hypothesis that  $\mu_1 - \mu_2 \neq -20$  as an alternative to the null hypothesis that  $\mu_1 - \mu_2 = -20$ .

### Applications

5. Each of five laboratory mice was released into a maze twice. The five pairs of times to escape were:

Mouse	1	2	3	4	5
First release	129	89	136	163	118
Second release	113	97	139	85	75

- Compute  $\bar{d}$  and  $s_d$ .
  - Give a point estimate for  $\mu_1 - \mu_2 = \mu_d$ .
  - Construct the 90% confidence interval for  $\mu_1 - \mu_2 = \mu_d$  from these data.
  - Test, at the 10% level of significance, the hypothesis that it takes mice less time to run the maze on the second trial, on average.
6. Eight golfers were asked to submit their latest scores on their favorite golf courses. These golfers were each given a set of newly designed clubs. After playing with the new clubs for a few months, the golfers were again asked to submit their latest scores on the same golf courses. The results are summarized below.

Golfer	1	2	3	4	5	6	7	8
Own clubs	77	80	69	73	73	72	75	77
New clubs	72	81	68	73	75	70	73	75

- Compute  $\bar{d}$  and  $s_d$ .
  - Give a point estimate for  $\mu_1 - \mu_2 = \mu_d$ .
  - Construct the 99% confidence interval for  $\mu_1 - \mu_2 = \mu_d$  from these data.
  - Test, at the 1% level of significance, the hypothesis that on average golf scores are lower with the new clubs.
7. A neighborhood home owners association suspects that the recent appraisal values of the houses in the neighborhood conducted by the county government for taxation purposes is too high. It hired a private company to appraise the values of ten houses in the neighborhood. The results, in thousands of dollars, are

House	County Government	Private Company
1	217	219
2	350	338
3	296	291
4	237	237
5	237	235
6	272	269
7	257	239

House	County Government	Private Company
8	277	275
9	312	320
10	335	335

- Give a point estimate for the difference between the mean private appraisal of all such homes and the government appraisal of all such homes.
  - Construct the 99% confidence interval based on these data for the difference.
  - Test, at the 1% level of significance, the hypothesis that appraised values by the county government of all such houses is greater than the appraised values by the private appraisal company.
8. In order to cut costs a wine producer is considering using duo or 1 + 1 corks in place of full natural wood corks, but is concerned that it could affect buyers's perception of the quality of the wine. The wine producer shipped eight pairs of bottles of its best young wines to eight wine experts. Each pair includes one bottle with a natural wood cork and one with a duo cork. The experts are asked to rate the wines on a one to ten scale, higher numbers corresponding to higher quality. The results are:

Wine Expert	Duo Cork	Wood Cork
1	8.5	8.5
2	8.0	8.5
3	6.5	8.0
4	7.5	8.5
5	8.0	7.5
6	8.0	8.0
7	9.0	9.0
8	7.0	7.5

- Give a point estimate for the difference between the mean ratings of the wine when bottled are sealed with different kinds of corks.
  - Construct the 90% confidence interval based on these data for the difference.
  - Test, at the 10% level of significance, the hypothesis that on the average duo corks decrease the rating of the wine.
9. Engineers at a tire manufacturing corporation wish to test a new tire material for increased durability. To test the tires under realistic road conditions, new front tires are mounted on each of 11 company cars, one tire made with a production material and the other with the experimental material. After a fixed period the 11 pairs were measured for wear. The amount of wear for each tire (in mm) is shown in the table:

Car	Production	Experimental
1	5.1	5.0
2	6.5	6.5
3	3.6	3.1
4	3.5	3.7
5	5.7	4.5
6	5.0	4.1
7	6.4	5.3
8	4.7	2.6

Car	Production	Experimental
9	3.2	3.0
10	3.5	3.5
11	6.4	5.1

- Give a point estimate for the difference in mean wear.
  - Construct the 99% confidence interval for the difference based on these data.
  - Test, at the 1% level of significance, the hypothesis that the mean wear with the experimental material is less than that for the production material.
10. A marriage counselor administered a test designed to measure overall contentment to 30 randomly selected married couples. The scores for each couple are given below. A higher number corresponds to greater contentment or happiness.

Couple	Husband	Wife
1	47	44
2	44	46
3	49	44
4	53	44
5	42	43
6	45	45
7	48	47
8	45	44
9	52	44
10	47	42
11	40	34
12	45	42
13	40	43
14	46	41
15	47	45
16	46	45
17	46	41
18	46	41
19	44	45
20	45	43
21	48	38
22	42	46
23	50	44
24	46	51
25	43	45

Couple	Husband	Wife
26	50	40
27	46	46
28	42	41
29	51	41
30	46	47

- Test, at the 1% level of significance, the hypothesis that on average men and women are not equally happy in marriage.
- Test, at the 1% level of significance, the hypothesis that on average men are happier than women in marriage.

## Large Data Set Exercises

### Large Data Sets are absent

- Large Data Set 5 lists the scores for 25 randomly selected students on practice SAT reading tests before and after taking a two-week SAT preparation course. Denote the population of all students who have taken the course as Population 1 and the population of all students who have not taken the course as Population 2.
  - Compute the 25 differences in the order **after - before**, their mean  $\bar{d}$ , and their sample standard deviation  $s_d$ .
  - Give a point estimate for  $\mu_d = \mu_1 - \mu_2$ , the difference in the mean score of all students who have taken the course and the mean score of all who have not.
  - Construct a 98% confidence interval for  $\mu_d$ .
  - Test, at the 1% level of significance, the hypothesis that the mean SAT score increases by at least ten points by taking the two-week preparation course.
- Large Data Set 12 lists the scores on one round for 75 randomly selected members at a golf course, first using their own original clubs, then two months later after using new clubs with an experimental design. Denote the population of all golfers using their own original clubs as Population 1 and the population of all golfers using the new style clubs as Population 2.
  - Compute the 75 differences in the order **original clubs - new clubs**, their mean  $\bar{d}$ , and their sample standard deviation  $s_d$ .
  - Give a point estimate for  $\mu_d = \mu_1 - \mu_2$ , the difference in the mean score of all students who have taken the course and the mean score of all who have not.
  - Construct a 90% confidence interval for  $\mu_d$ .
  - Test, at the 1% level of significance, the hypothesis that the mean SAT score increases by at least ten points by taking the two-week preparation course.
- Consider the previous problem again. Since the data set is so large, it is reasonable to use the standard normal distribution instead of Student's  $t$ -distribution with 74 degrees of freedom.
  - Construct a 90% confidence interval for  $\mu_d$  using the standard normal distribution, meaning that the formula is  $\bar{d} \pm z_{\alpha/2} \frac{s_d}{\sqrt{n}}$ . (The computations done in part (a) of the previous problem still apply and need not be redone.) How does the result obtained here compare to the result obtained in part (c) of the previous problem?
  - Test, at the 1% level of significance, the hypothesis that the mean golf score decreases by at least one stroke by using the new kind of clubs, using the standard normal distribution. (All the work done in part (d) of the previous problem applies, except the critical value is now  $z_\alpha$  instead of  $t_\alpha$  (or the  $p$ -value can be computed exactly instead of only approximated, if you used the  $p$ -value approach).) How does the result obtained here compare to the result obtained in part (c) of the previous problem?
  - Construct the 99% confidence intervals for  $\mu_d$  using both the  $t$ - and  $z$ -distributions. How much difference is there in the results now?

## Answers

- $\bar{d} = 7.4286$ ,  $s_d = 0.9759$
  - $\bar{d} = 7.4286$
  - (6.53, 8.33)
  - $T = 1.162$ ,  $df = 6$ ,  $t_{0.10} = 1.44$ , do not reject  $H_0$

- 2.
3. a.  $\bar{d} = -14.25$ ,  $s_d = 1.5$   
 b.  $\bar{d} = -14.25$   
 c.  $(-18.63, -9.87)$   
 d.  $T = -3.000$ ,  $df = 3$ ,  $\pm t_{0.05} = \pm 2.353$ , reject  $H_0$
- 4.
5. a.  $\bar{d} = 25.2$ ,  $s_d = 35.6609$   
 b.  $\bar{d} = 25.2$   
 c.  $25.2 \pm 34.0$   
 d.  $T = 1.580$ ,  $df = 4$ ,  $t_{0.10} = 1.533$ , reject  $H_0$  (takes less time)
- 6.
7. a. 3.2  
 b.  $3.2 \pm 7.5$   
 c.  $T = 1.392$ ,  $df = 9$ ,  $t_{0.10} = 2.821$ , do not reject  $H_0$  (government appraisals not higher)
- 8.
9. a. 0.65  
 b.  $0.65 \pm 0.69$   
 c.  $T = 3.014$ ,  $df = 10$ ,  $t_{0.10} = 2.764$ , reject  $H_0$  (experimental material wears less)
- 10.
11. a.  $\bar{d} = 16.68$ ,  $s_d = 10.77$   
 b.  $\bar{d} = 16.68$   
 c.  $(11.31, 22.05)$   
 d.  $H_0 : \mu_1 - \mu_2 = 10$  vs  $H_a : \mu_1 - \mu_2 > 10$ . Test Statistic:  $T = 3.1014$ ,  $df = 11$ . Rejection Region:  $[2.492, \infty)$  Decision: Reject  $H_0$ .
- 12.
13. a. (1.6266, 2.6401) Endpoints change in the third decimal place.  
 b.  $H_0 : \mu_1 - \mu_2 = 1$  vs  $H_a : \mu_1 - \mu_2 > 1$ . Test Statistic:  $Z = 3.6791$ . Rejection Region:  $[2.33, \infty)$  Decision: Reject  $H_0$ . The decision is the same as in the previous problem.  
 c. Using the  $t$ -distribution, (1.3188, 2.9478) Using the  $z$ -distribution, (1.3401, 2.9266) There is a difference.

## 9.4: Comparison of Two Population Proportions

### Basic

1. Construct the confidence interval for  $p_1 - p_2$  for the level of confidence and the data given. (The samples are sufficiently large.)

- a. 90% confidence

$$\begin{aligned} n_1 &= 1670, \hat{p}_1 = 0.42 \\ n_2 &= 900, \hat{p}_2 = 0.38 \end{aligned} \quad (9.E.53)$$

- b. 95% confidence

$$\begin{aligned} n_1 &= 600, \hat{p}_1 = 0.84 \\ n_2 &= 420, \hat{p}_2 = 0.67 \end{aligned} \quad (9.E.54)$$

2. Construct the confidence interval for  $p_1 - p_2$  for the level of confidence and the data given. (The samples are sufficiently large.)

- a. 98% confidence

$$\begin{aligned} n_1 &= 750, \hat{p}_1 = 0.64 \\ n_2 &= 800, \hat{p}_2 = 0.51 \end{aligned} \quad (9.E.55)$$

- b. 99.5% confidence

$$n_1 = 250, \hat{p}_1 = 0.78 \quad (9.E.56)$$

$$n_2 = 250, \hat{p}_2 = 0.51$$

3. Construct the confidence interval for  $p_1 - p_2$  for the level of confidence and the data given. (The samples are sufficiently large.)

a. 80% confidence

$$n_1 = 300, \hat{p}_1 = 0.255 \quad (9.E.57)$$

$$n_2 = 400, \hat{p}_2 = 0.193$$

b. 95% confidence

$$n_1 = 3500, \hat{p}_1 = 0.147 \quad (9.E.58)$$

$$n_2 = 3750, \hat{p}_2 = 0.131$$

4. Construct the confidence interval for  $p_1 - p_2$  for the level of confidence and the data given. (The samples are sufficiently large.)

a. 99% confidence

$$n_1 = 2250, \hat{p}_1 = 0.915 \quad (9.E.59)$$

$$n_2 = 2525, \hat{p}_2 = 0.858$$

b. 95% confidence

$$n_1 = 120, \hat{p}_1 = 0.650 \quad (9.E.60)$$

$$n_2 = 200, \hat{p}_2 = 0.505$$

5. Perform the test of hypotheses indicated, using the data given. Use the critical value approach. Compute the  $p$ -value of the test as well. (The samples are sufficiently large.)

a. Test  $H_0 : p_1 - p_2 = 0$  vs  $H_a : p_1 - p_2 > 0$  @  $\alpha = 0.10$

$$n_1 = 1200, \hat{p}_1 = 0.42 \quad (9.E.61)$$

$$n_2 = 1200, \hat{p}_2 = 0.40$$

b. Test  $H_0 : p_1 - p_2 = 0$  vs  $H_a : p_1 - p_2 \neq 0$  @  $\alpha = 0.05$

$$n_1 = 550, \hat{p}_1 = 0.61 \quad (9.E.62)$$

$$n_2 = 600, \hat{p}_2 = 0.67$$

6. Perform the test of hypotheses indicated, using the data given. Use the critical value approach. Compute the  $p$ -value of the test as well. (The samples are sufficiently large.)

a. Test  $H_0 : p_1 - p_2 = 0.05$  vs  $H_a : p_1 - p_2 > 0.05$  @  $\alpha = 0.05$

$$n_1 = 1100, \hat{p}_1 = 0.57 \quad (9.E.63)$$

$$n_2 = 1100, \hat{p}_2 = 0.48$$

b. Test  $H_0 : p_1 - p_2 = 0$  vs  $H_a : p_1 - p_2 \neq 0$  @  $\alpha = 0.05$

$$n_1 = 800, \hat{p}_1 = 0.39 \quad (9.E.64)$$

$$n_2 = 900, \hat{p}_2 = 0.43$$

7. Perform the test of hypotheses indicated, using the data given. Use the critical value approach. Compute the  $p$ -value of the test as well. (The samples are sufficiently large.)

a. Test  $H_0 : p_1 - p_2 = 0.25$  vs  $H_a : p_1 - p_2 < 0.25$  @  $\alpha = 0.005$

$$n_1 = 1400, \hat{p}_1 = 0.57 \quad (9.E.65)$$

$$n_2 = 1200, \hat{p}_2 = 0.37$$

b. Test  $H_0 : p_1 - p_2 = 0.16$  vs  $H_a : p_1 - p_2 \neq 0.16$  @  $\alpha = 0.02$

$$n_1 = 750, \hat{p}_1 = 0.43 \quad (9.E.66)$$

$$n_2 = 600, \hat{p}_2 = 0.22$$

8. Perform the test of hypotheses indicated, using the data given. Use the critical value approach. Compute the  $p$ -value of the test as well. (The samples are sufficiently large.)

a. Test  $H_0 : p_1 - p_2 = 0.08$  vs  $H_a : p_1 - p_2 > 0.08$  @  $\alpha = 0.025$

$$n_1 = 450, \hat{p}_1 = 0.67 \quad (9.E.67)$$

$$n_2 = 200, \hat{p}_2 = 0.52$$

b. Test  $H_0 : p_1 - p_2 = 0.02$  vs  $H_a : p_1 - p_2 \neq 0.02$  @  $\alpha = 0.001$

$$n_1 = 2700, \hat{p}_1 = 0.837 \quad (9.E.68)$$

$$n_2 = 2900, \hat{p}_2 = 0.854$$

9. Perform the test of hypotheses indicated, using the data given. Use the critical value approach. Compute the  $p$ -value of the test as well. (The samples are sufficiently large.)

a. Test  $H_0 : p_1 - p_2 = 0$  vs  $H_a : p_1 - p_2 < 0$  @  $\alpha = 0.005$

$$n_1 = 1100, \hat{p}_1 = 0.22 \quad (9.E.69)$$

$$n_2 = 1300, \hat{p}_2 = 0.27$$

b. Test  $H_0 : p_1 - p_2 = 0$  vs  $H_a : p_1 - p_2 \neq 0$  @  $\alpha = 0.01$

$$n_1 = 650, \hat{p}_1 = 0.35 \quad (9.E.70)$$

$$n_2 = 650, \hat{p}_2 = 0.41$$

10. Perform the test of hypotheses indicated, using the data given. Use the critical value approach. Compute the  $p$ -value of the test as well. (The samples are sufficiently large.)

a. Test  $H_0 : p_1 - p_2 = 0.15$  vs  $H_a : p_1 - p_2 > 0.15$  @  $\alpha = 0.10$

$$n_1 = 950, \hat{p}_1 = 0.41 \quad (9.E.71)$$

$$n_2 = 500, \hat{p}_2 = 0.23$$

b. Test  $H_0 : p_1 - p_2 = 0.10$  vs  $H_a : p_1 - p_2 \neq 0.10$  @  $\alpha = 0.10$

$$n_1 = 220, \hat{p}_1 = 0.92 \quad (9.E.72)$$

$$n_2 = 160, \hat{p}_2 = 0.78$$

11. Perform the test of hypotheses indicated, using the data given. Use the critical value approach. Compute the  $p$ -value of the test as well. (The samples are sufficiently large.)

a. Test  $H_0 : p_1 - p_2 = 0.22$  vs  $H_a : p_1 - p_2 > 0.22$  @  $\alpha = 0.05$

$$n_1 = 90, \hat{p}_1 = 0.72 \quad (9.E.73)$$

$$n_2 = 75, \hat{p}_2 = 0.40$$

b. Test  $H_0 : p_1 - p_2 = 0.37$  vs  $H_a : p_1 - p_2 \neq 0.37$  @  $\alpha = 0.02$

$$n_1 = 425, \hat{p}_1 = 0.772 \quad n_2 = 425, \hat{p}_2 = 0.331 \quad (9.E.74)$$

12. Perform the test of hypotheses indicated, using the data given. Use the critical value approach. Compute the  $p$ -value of the test as well. (The samples are sufficiently large.)

a. Test  $H_0 : p_1 - p_2 = 0.50$  vs  $H_a : p_1 - p_2 < 0.50$  @  $\alpha = 0.10$

$$n_1 = 40, \hat{p}_1 = 0.65 \quad (9.E.75)$$

$$n_2 = 55, \hat{p}_2 = 0.24$$

b. Test  $H_0 : p_1 - p_2 = 0.30$  vs  $H_a : p_1 - p_2 \neq 0.30$  @  $\alpha = 0.10$

$$n_1 = 7500, \hat{p}_1 = 0.664 \quad (9.E.76)$$

$$n_2 = 1000, \hat{p}_2 = 0.319$$

## Applications

In all the remaining exercises the samples are sufficiently large (so this need not be checked).

13. Voters in a particular city who identify themselves with one or the other of two political parties were randomly selected and asked if they favor a proposal to allow citizens with proper license to carry a concealed handgun in city parks. The results are:

	Party A	Party B
Sample size, $n$	150	200
Number in favor, $x$	90	140

- Give a point estimate for the difference in the proportion of all members of Party A and all members of Party B who favor the proposal.
  - Construct the 95% confidence interval for the difference, based on these data.
  - Test, at the 5% level of significance, the hypothesis that the proportion of all members of Party A who favor the proposal is less than the proportion of all members of Party B who do.
  - Compute the  $p$ -value of the test.
14. To investigate a possible relation between gender and handedness, a random sample of 320 adults was taken, with the following results:

	Men	Women
Sample size, $n$	168	152
Number of left-handed, $x$	24	9

- Give a point estimate for the difference in the proportion of all men who are left-handed and the proportion of all women who are left-handed.
  - Construct the 95% confidence interval for the difference, based on these data.
  - Test, at the 5% level of significance, the hypothesis that the proportion of men who are left-handed is greater than the proportion of women who are.
  - Compute the  $p$ -value of the test.
15. A local school board member randomly sampled private and public high school teachers in his district to compare the proportions of National Board Certified (NBC) teachers in the faculty. The results were:

	Private Schools	Public Schools
Sample size, $n$	80	520
Proportion of NBC teachers,	0.175	0.150

- Give a point estimate for the difference in the proportion of all teachers in area public schools and the proportion of all teachers in private schools who are National Board Certified.
  - Construct the 90% confidence interval for the difference, based on these data.
  - Test, at the 10% level of significance, the hypothesis that the proportion of all public school teachers who are National Board Certified is less than the proportion of private school teachers who are.
  - Compute the  $p$ -value of the test.
16. In professional basketball games, the fans of the home team always try to distract free throw shooters on the visiting team. To investigate whether this tactic is actually effective, the free throw statistics of a professional basketball player with a high free throw percentage were examined. During the entire last season, this player had 656 free throws, 420 in home games and 236 in away games. The results are summarized below.

	Home	Away
Sample size, $n$	420	236
Free throw percent, $\hat{p}$	81.5%	78.8%

- a. Give a point estimate for the difference in the proportion of free throws made at home and away.
  - b. Construct the 90% confidence interval for the difference, based on these data.
  - c. Test, at the 10% level of significance, the hypothesis that there exists a home advantage in free throws.
  - d. Compute the  $p$ -value of the test.
17. Randomly selected middle-aged people in both China and the United States were asked if they believed that adults have an obligation to financially support their aged parents. The results are summarized below.

	China	USA
Sample size, $n$	1300	150
Number of yes, $x$	1170	110

Test, at the 1% level of significance, whether the data provide sufficient evidence to conclude that there exists a cultural difference in attitude regarding this question.

18. A manufacturer of walk-behind push mowers receives refurbished small engines from two new suppliers,  $A$  and  $B$ . It is not uncommon that some of the refurbished engines need to be lightly serviced before they can be fitted into mowers. The mower manufacturer recently received 100 engines from each supplier. In the shipment from  $A$ , 13 needed further service. In the shipment from  $B$ , 10 needed further service. Test, at the 10% level of significance, whether the data provide sufficient evidence to conclude that there exists a difference in the proportions of engines from the two suppliers needing service.

### Large Data Set Exercises

#### Large Data Sets are absent

19. Large Data Sets 6A and 6B record results of a random survey of 200 voters in each of two regions, in which they were asked to express whether they prefer Candidate A for a U.S. Senate seat or prefer some other candidate. Let the population of all voters in region 1 be denoted Population 1 and the population of all voters in region 2 be denoted Population 2. Let  $p_1$  be the proportion of voters in Population 1 who prefer Candidate A, and  $p_2$  the proportion in Population 2 who do.
- a. Find the relevant sample proportions  $\hat{p}_1$  and  $\hat{p}_2$ .
  - b. Construct a point estimate for  $p_1 - p_2$ .
  - c. Construct a 95% confidence interval for  $p_1 - p_2$ .
  - d. Test, at the 5% level of significance, the hypothesis that the same proportion of voters in the two regions favor Candidate A, against the alternative that a larger proportion in Population 2 do.
20. Large Data Set 11 records the results of samples of real estate sales in a certain region in the year 2008 (lines 2 through 536) and in the year 2010 (lines 537 through 1106). Foreclosure sales are identified with a 1 in the second column. Let all real estate sales in the region in 2008 be Population 1 and all real estate sales in the region in 2010 be Population 2.
- a. Use the sample data to construct point estimates  $\hat{p}_1$  and  $\hat{p}_2$  of the proportions  $p_1$  and  $p_2$  of all real estate sales in this region in 2008 and 2010 that were foreclosure sales. Construct a point estimate of  $p_1 - p_2$ .
  - b. Use the sample data to construct a 90% confidence for  $p_1 - p_2$ .
  - c. Test, at the 10% level of significance, the hypothesis that the proportion of real estate sales in the region in 2010 that were foreclosure sales was greater than the proportion of real estate sales in the region in 2008 that were foreclosure sales. (The default is that the proportions were the same.)

### Answers

1. a. (0.0068, 0.0732)  
b. (0.1163, 0.2237)
- 2.
3. a. (0.0210, 0.1030)  
b. (0.0001, 0.0319)
- 4.
5. a.  $Z = 0.996$ ,  $z_{0.10} = 1.282$ ,  $p\text{-value} = 0.1587$ , do not reject  $H_0$   
b.  $Z = -2.120$ ,  $\pm z_{0.025} = \pm 1.960$ ,  $p\text{-value} = 0.0340$ , reject  $H_0$
- 6.

7. a.  $Z = -2.602$ ,  $-z_{0.005} = -2.576$ , p-value = 0.0047, reject  $H_0$   
b.  $Z = 2.020$ ,  $\pm z_{0.01} = \pm 2.326$ , p-value = 0.0434, do not reject  $H_0$
- 8.
9. a.  $Z = -2.85$ , p-value = 0.0022, reject  $H_0$   
b.  $Z = -2.23$ , p-value = 0.0258, do not reject  $H_0$
- 10.
11. a.  $Z = 1.36$ , p-value = 0.0869, do not reject  $H_0$   
b.  $Z = 2.32$ , p-value = 0.0204, do not reject  $H_0$
- 12.
13. a.  $-0.10$   
b.  $-0.10 \pm 0.101$   
c.  $Z = -1.943$ ,  $-z_{0.05} = -1.645$ , reject  $H_0$  (fewer in Party A favor)  
d. p-value = 0.0262
- 14.
15. a. 0.025  
b.  $0.025 \pm 0.0745$   
c.  $Z = 0.552$ ,  $z_{0.10} = 1.282$ , do not reject  $H_0$  (as many public school teachers are certified)  
d. p-value = 0.2912
- 16.
17.  $Z = 4.498$ ,  $\pm z_{0.005} = \pm 2.576$ , reject  $H_0$  (different)
- 18.
19. a.  $\hat{p}_1 = 0.355$  and  $\hat{p}_2 = 0.41$   
b.  $\hat{p}_1 - \hat{p}_2 = -0.055$   
c.  $(-0.1501, 0.0401)$   
d.  $H_0 : p_1 - p_2 = 0$  vs  $H_a : p_1 - p_2 < 0$ . Test Statistic:  $Z = -1.1335$ . Rejection Region:  $(-\infty, -1.645]$ . Decision: Fail to reject  $H_0$ .

## 9.5 Sample Size Considerations

### Basic

1. Estimate the common sample size  $n$  of equally sized independent samples needed to estimate  $\mu_1 - \mu_2$  as specified when the population standard deviations are as shown.
  - a. 90% confidence, to within 3 units,  $\sigma_1 = 10$  and  $\sigma_2 = 7$
  - b. 99% confidence, to within 4 units,  $\sigma_1 = 6.8$  and  $\sigma_2 = 9.3$
  - c. 95% confidence, to within 5 units,  $\sigma_1 = 22.6$  and  $\sigma_2 = 31.8$
2. Estimate the common sample size  $n$  of equally sized independent samples needed to estimate  $\mu_1 - \mu_2$  as specified when the population standard deviations are as shown.
  - a. 80% confidence, to within 2 units,  $\sigma_1 = 14$  and  $\sigma_2 = 23$
  - b. 90% confidence, to within 0.3 units,  $\sigma_1 = 1.3$  and  $\sigma_2 = 0.8$
  - c. 99% confidence, to within 11 units,  $\sigma_1 = 42$  and  $\sigma_2 = 37$
3. Estimate the number  $n$  of pairs that must be sampled in order to estimate  $\mu_d = \mu_1 - \mu_2$  as specified when the standard deviation  $s_d$  of the population of differences is as shown.
  - a. 80% confidence, to within 6 units,  $\sigma_d = 26.5$
  - b. 95% confidence, to within 4 units,  $\sigma_d = 12$
  - c. 90% confidence, to within 5.2 units,  $\sigma_d = 11.3$
4. Estimate the number  $n$  of pairs that must be sampled in order to estimate  $\mu_d = \mu_1 - \mu_2$  as specified when the standard deviation  $s_d$  of the population of differences is as shown.
  - a. 90% confidence, to within 20 units,  $\sigma_d = 75.5$
  - b. 95% confidence, to within 11 units,  $\sigma_d = 31.4$
  - c. 99% confidence, to within 1.8 units,  $\sigma_d = 4$

5. Estimate the minimum equal sample sizes  $n_1 = n_2$  necessary in order to estimate  $p_1 - p_2$  as specified.
  - a. 80% confidence, to within 0.05 (five percentage points)
    - i. when no prior knowledge of  $p_1$  or  $p_2$  is available
    - ii. when prior studies indicate that  $p_1 \approx 0.20$  and  $p_2 \approx 0.65$
  - b. 90% confidence, to within 0.02 (two percentage points)
    - i. when no prior knowledge of  $p_1$  or  $p_2$  is available
    - ii. when prior studies indicate that  $p_1 \approx 0.75$  and  $p_2 \approx 0.63$
  - c. 95% confidence, to within 0.10 (ten percentage points)
    - i. when no prior knowledge of  $p_1$  or  $p_2$  is available
    - ii. when prior studies indicate that  $p_1 \approx 0.11$  and  $p_2 \approx 0.37$
6. Estimate the minimum equal sample sizes  $n_1 = n_2$  necessary in order to estimate  $p_1 - p_2$  as specified.
  - a. 80% confidence, to within 0.02 (two percentage points)
    - i. when no prior knowledge of  $p_1$  or  $p_2$  is available
    - ii. when prior studies indicate that  $p_1 \approx 0.78$  and  $p_2 \approx 0.65$
  - b. 90% confidence, to within 0.05 (five percentage points)
    - i. when no prior knowledge of  $p_1$  or  $p_2$  is available
    - ii. when prior studies indicate that  $p_1 \approx 0.12$  and  $p_2 \approx 0.24$
  - c. 95% confidence, to within 0.10 (ten percentage points)
    - i. when no prior knowledge of  $p_1$  or  $p_2$  is available
    - ii. when prior studies indicate that  $p_1 \approx 0.14$  and  $p_2 \approx 0.21$

### Applications

7. An educational researcher wishes to estimate the difference in average scores of elementary school children on two versions of a 100-point standardized test, at 99% confidence and to within two points. Estimate the minimum equal sample sizes necessary if it is known that the standard deviation of scores on different versions of such tests is 4.9.
8. A university administrator wishes to estimate the difference in mean grade point averages among all men affiliated with fraternities and all unaffiliated men, with 95% confidence and to within 0.15. It is known from prior studies that the standard deviations of grade point averages in the two groups have common value 0.4. Estimate the minimum equal sample sizes necessary to meet these criteria.
9. An automotive tire manufacturer wishes to estimate the difference in mean wear of tires manufactured with an experimental material and ordinary production tire, with 90% confidence and to within 0.5 mm. To eliminate extraneous factors arising from different driving conditions the tires will be tested in pairs on the same vehicles. It is known from prior studies that the standard deviations of the differences of wear of tires constructed with the two kinds of materials is 1.75 mm. Estimate the minimum number of pairs in the sample necessary to meet these criteria.
10. To assess the relative happiness of men and women in their marriages, a marriage counselor plans to administer a test measuring happiness in marriage to  $n$  randomly selected married couples, record their test scores, find the differences, and then draw inferences on the possible difference. Let  $\mu_1$  and  $\mu_2$  be the true average levels of happiness in marriage for men and women respectively as measured by this test. Suppose it is desired to find a 90% confidence interval for estimating  $\mu_d = \mu_1 - \mu_2$  to within two test points. Suppose further that, from prior studies, it is known that the standard deviation of the differences in test scores is  $\sigma_d \approx 10$ . What is the minimum number of married couples that must be included in this study?
11. A journalist plans to interview an equal number of members of two political parties to compare the proportions in each party who favor a proposal to allow citizens with a proper license to carry a concealed handgun in public parks. Let  $p_1$  and  $p_2$  be the true proportions of members of the two parties who are in favor of the proposal. Suppose it is desired to find a 95% confidence interval for estimating  $p_1 - p_2$  to within 0.05. Estimate the minimum equal number of members of each party that must be sampled to meet these criteria.
12. A member of the state board of education wants to compare the proportions of National Board Certified (NBC) teachers in private high schools and in public high schools in the state. His study plan calls for an equal number of private school teachers and public school teachers to be included in the study. Let  $p_1$  and  $p_2$  be these proportions. Suppose it is desired to find a 99% confidence interval that estimates  $p_1 - p_2$  to within 0.05.

- a. Supposing that both proportions are known, from a prior study, to be approximately 0.15, compute the minimum common sample size needed.
- b. Compute the minimum common sample size needed on the supposition that nothing is known about the values of  $p_1$  and  $p_2$ .

### Answers

1.
  - a.  $n_1 = n_2 = 45$
  - b.  $n_1 = n_2 = 56$
  - c.  $n_1 = n_2 = 234$
- 2.
3.
  - a.  $n_1 = n_2 = 33$
  - b.  $n_1 = n_2 = 35$
  - c.  $n_1 = n_2 = 13$
- 4.
5.
  - a.
    - i.  $n_1 = n_2 = 329$
    - ii.  $n_1 = n_2 = 255$
  - b.
    - i.  $n_1 = n_2 = 3383$
    - ii.  $n_1 = n_2 = 2846$
  - c.
    - i.  $n_1 = n_2 = 193$
    - ii.  $n_1 = n_2 = 128$
- 6.
7.  $n_1 = n_2 \approx 80$
- 8.
9.  $n_1 = n_2 \approx 34$
- 10.
11.  $n_1 = n_2 \approx 769$

---

This page titled [9.E: Two-Sample Problems \(Exercises\)](#) is shared under a [CC BY-NC-SA 3.0](#) license and was authored, remixed, and/or curated by [Anonymous](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

- [9.E: Two-Sample Problems \(Exercises\)](#) has no license indicated.