Chapter 7

Comparing Two Populations

7.1 Study Suggestions

Only one new formula appears in Chapter 7, the confidence interval for the difference of two proportions. (It is also stated that the hypothesis test for comparing two proportions reduces to the familiar Fisher's test.) Chapter 7 focuses on four types of studies, either finite populations or Bernoulli trials matched with either controlled or observational studies. It is important to note that a controlled study must have subjects assigned to treatments by randomization.

A colleague of mine who has taught introductory statistics several times from early versions of my text once remarked that Chapter 7 provides perhaps the best illustration of how my approach to introductory statistics differs from the standard approach. A text that follows the standard approach does not emphasize, and in fact may not even mention, the differences between these four types of studies because, after all, the mathematical derivations and formulas are the same for each study type. The four types of studies differ greatly, however, in how they are executed, and how they should be interpreted. If, like me, your teacher wants you to become a critical interpreter of the research and conclusions of others, then these latter issues—execution and interpretation—need to be emphasized in the course.

If in Chapter 5 you were interested in the informal ways to check the assumptions of Bernoulli trials, then you may want to take the time to study the examples in the text that show how the techniques of Chapter 7 can put the earlier informal examinations into the familiar framework of hypothesis testing.

The discussion of the limitations of observational studies is continued in Chapter 9; see especially the material on Simpson's paradox and standardized rates.

As you continue to work through the text, remember that the qualitative differences between controlled and observational studies presented in Chapter 7 for a dichotomous response are equally valid for a multicategory or numerical result, as presented in Chapters 11 and 16, respectively.

Chapter 7 introduces the idea of the *practical importance* of results. I find this to be an easy, yet intellectually honest, way to help students to understand the issue of sample size. (A reviewer of the first version of the text opined that I should drop Clyde Gaines's Three-Point Basket study from the book because, "You don't need to be a statistician to analyze the data!" Chapter 7 illustrates, however, that some statistical training can help one realize that for the goal of estimating a difference, Clyde's sample size was much too small.)

7.2 Solutions to Odd-Numbered Exercises

1. (b) The values $n_1 = 5{,}000$ and $n_2 = 5{,}000$ are given. It follows that $n = 10{,}000$, and

$$a = \hat{p}_1 n_1 = 0.42(5{,}000) = 2{,}100.$$

Similarly, $c = 2{,}000$. Subtraction and addition yield $b = 2{,}900$, $d = 3{,}000$, $m_1 = 4{,}100$, and $m_2 = 5{,}900$. The standardized value of the test statistic for Fisher's test,

$$z = \frac{\sqrt{9{,}999}[2{,}100(3{,}000) - 2{,}900(2{,}000)]}{\sqrt{5{,}000(5{,}000)(4{,}100)(5{,}900)}},$$

is 2.03. The approximate P-value for the third alternative is twice the area under the standard normal curve to the right of 2.03. This area is $2(0.0212) = 0.0424$.

3. (b) $\hat{p}_1 = 0.62$, $\hat{p}_2 = 0.55$, and $\hat{p}_1 - \hat{p}_2 = 0.07$. The 95 percent confidence interval is 0.07 $\pm$

$$1.96\sqrt{\frac{0.62(0.38)}{756} + \frac{0.55(0.45)}{756}} =$$

$$0.07 \pm 0.05 = [0.02, 0.12].$$

(c) From the information given, it follows that $a = 0.62(756) = 469$, $b = 287$, $n_1 = 756$, $c = 0.55(756) = 416$, $d = 340$, $n_2 = 756$, $m_1 = 885$, $m_2 = 627$, and $n = 1512$. The standardized value of the test statistic for Fisher's test,

$$z = \frac{\sqrt{1{,}511}[469(340) - 287(416)]}{\sqrt{756(756)(885)(627)}},$$

is 2.77. The approximate P-value for the third alternative is twice the area under the standard normal curve to the right of 2.77. This area is $2(0.0028) = 0.0056$.

5. (b) $\hat{p}_1 = 0.480$, $\hat{p}_2 = 0.400$, and $\hat{p}_1 - \hat{p}_2 = 0.080$. The 95 percent confidence interval is 0.080 $\pm$

$$1.96\sqrt{\frac{0.480(0.520)}{25} + \frac{0.400(0.600)}{25}} =$$

$$0.080 \pm 0.274 = [-0.194, 0.354].$$

7. (a) The standardized value of the test statistic for Fisher's test,

$$z = \frac{\sqrt{1{,}863}[2(1{,}362) - 251(249)]}{\sqrt{253(1{,}611)(251)(1{,}613)}},$$

is 6.35. The approximate P-value for the third alternative is twice the area under the standard normal curve to the right of 6.35. This area is smaller than 0.0004.

(b) The assumption of Bernoulli trials should be discarded.

9. (a) The standardized value of the test statistic for Fisher's test is

$$z = \frac{\sqrt{98}[21(30) - 24(24)]}{\sqrt{45(54)(45)(54)}} = 0.22.$$

The approximate P-value for the third alternative is twice the area under the standard normal curve to the right of 0.22. This area equals $2(0.4129) = 0.8258$.

(b) $\hat{p}_1 = 0.467$, $\hat{p}_2 = 0.444$, and $\hat{p}_1 - \hat{p}_2 = 0.023$. The 95 percent confidence interval for $p_1 - p_2$ is 0.023 $\pm$

$$1.96\sqrt{\frac{0.467(0.533)}{45} + \frac{0.444(0.556)}{54}} =$$

$$0.023 \pm 0.197 = [-0.174, 0.230].$$

11. Let the first population be 1983 and the second population be 1984. With this identification, $p_1 - p_2$ can be interpreted as the decrease in awareness from 1983 to 1984. $\hat{p}_1 = 0.648$, $\hat{p}_2 = 0.590$, and $\hat{p}_1 - \hat{p}_2 = 0.058$. The 95 percent confidence interval for $p_1 - p_2$ is 0.058 $\pm$

$$1.96\sqrt{\frac{0.648(0.352)}{1{,}072} + \frac{0.590(0.410)}{2{,}632}} =$$

$$0.058 \pm 0.034 = [0.024, 0.092].$$

13. Let the first population be 1984 and the second population be 1983. With this identification, $p_1 - p_2$ can be interpreted as the increase in knowledge from 1983 to 1984. $\hat{p}_1 = 0.664$, $\hat{p}_2 = 0.650$, and $\hat{p}_1 - \hat{p}_2 = 0.014$. The 95 percent confidence interval for $p_1 - p_2$ is 0.014 $\pm$

$$1.96\sqrt{\frac{0.664(0.336)}{1{,}552} + \frac{0.650(0.350)}{695}} =$$

$$0.014 \pm 0.043 = [-0.029, 0.057].$$

At the 95 percent level, the level of knowledge might not have changed since the interval includes 0. The interval is narrow; hence, the change in knowledge, if any, was likely small.

15. Let the first population be 1984 and the second population be 1983. With this identification, $p_1 - p_2$ can be interpreted as the increase in knowledge from 1983 to 1984. $\hat{p}_1 = 0.609$, $\hat{p}_2 = 0.659$, and $\hat{p}_1 - \hat{p}_2 = -0.050$. The 95 percent confidence interval for $p_1 - p_2$ is -0.050 $\pm$

$$1.96\sqrt{\frac{0.609(0.391)}{1{,}552} + \frac{0.659(0.341)}{695}} =$$

$$-0.050 \pm 0.043 = [-0.093, -0.007].$$

7.3 Exam Questions

1. Suppose we are estimating the effectiveness of a drug (population 1) versus a placebo (population 2) and obtain a 99% confidence interval for $p_1 - p_2$ of $[0.02, 0.04]$. A physician points out that unless $p_1 - p_2 \geq 0.05$, the drug is not worth using. This is an example for which ___________ is a more important consideration than statistical significance.
   (a) common sense
   (b) practical importance
   (c) statistical power
   (d) a confounding factor

2. Kramer selects a random sample of size 400 from the population (one) of all white cars registered in California in 1993. He also selects an independent random sample of size 400 from the population (two) of all red cars registered in California in 1993. A car is labeled a failure if during 1993 it was involved in one or more accidents that were reported to the police. Kramer obtains a total of 60 failures, of which only 20 were white cars.

   Which type of study is this?
   (a) Controlled study on finite populations.
   (b) Observational study on finite populations.
   (c) Controlled study on Bernoulli trials.
   (d) Observational study on Bernoulli trials.

3. Refer to the previous question.
   (a) Describe, in words, what the numbers p_1 and p_2 represent in this study.
   (b) Compute the point estimates of p_1 and p_2.

4. Refer to the two previous questions. Obtain a 95% confidence interval for the proportion of white cars that are successes minus the proportion of red cars that are successes.

5. Use the information below to answer the questions that follow.
   - On Monday afternoon a golf pro attempted fifty putts from the distance of five feet, and obtained 32 successes (made putts) and 18 failures (missed putts).
   - On Tuesday afternoon, the same pro repeated the study and obtained 34 successes and 16 failures.
   - Assume that the shots on Monday were Bernoulli trials with probability of success equal to p_2.
   - Assume that the shots on Tuesday were Bernoulli trials with probability of success equal to p_1.
   - Note that p_2 and p_1 might be equal or unequal.

   (a) This study is an example of one of the four types of studies described in Chapter 7 of the text. Which one?
   (b) Of the following statements, which one best describes $p_1 - p_2$?
      i. It equals the amount the pro's probability of success increased from Monday to Tuesday.
      ii. It equals the amount the pro's probability of success decreased from Monday to Tuesday.
      iii. It equals 0.04.
      iv. It equals the pro's probability of success on Monday.
      v. It equals the pro's probability of success on Tuesday.
   (c) Construct the 95 percent confidence interval for $p_1 - p_2$.

6. A researcher uses the following sampling scheme. First, the researcher selects a random sample of 100 persons from a superpopulation of interest. Second, the researcher divides, by randomization, the 100 selected persons into two equal sized treatment groups.

   Which of the four types of studies is the researcher performing?

7. A researcher has two treatments for breast cancer. The response is a success if the woman is still alive five years after the cancer is discovered and a failure otherwise. The researcher selects n subjects at random from the population of all women with breast cancer and assigns, by randomization, an equal number of subjects to each treatment. She obtains an approximate P-value of 0.0124 for the two-sided alternative. Is the P-value still a valid measure of chance if the subjects are not a random sample from the population? Briefly justify your answer.

8. A researcher has two treatments for prostate cancer. The response is a success if the man is still alive five years after the cancer is discovered and a failure otherwise. The researcher selects n subjects at random from the population of all men with prostate cancer and assigns, by randomization, an equal number of subjects to each treatment. The researcher obtains $[0.01, 0.05]$ as a 95 percent confidence interval for $p_1 - p_2$. Which of the following is the correct interpretation of this interval?

   (a) At the 95 percent confidence level, the first treatment has between a 1- and 5-percentage-point-higher probability of success than the second treatment.

   (b) At the 95 percent confidence level, the second treatment has between a 1- and 5-percentage-point-higher probability of success than the first treatment.

   (c) At the 95 percent confidence level, the probability of success on the first treatment may be as much as 5 percentage points higher or 1 percentage point lower than the probability of success on the second treatment.

   (d) At the 95 percent confidence level, the probability of success on the second treatment may be as much as 5 percentage points higher or 1 percentage point lower than the probability of success on the first treatment.

9. Bob plans to collect data to test the null hypothesis that $p_1 = p_2$ versus the third alternative. His simulation study indicates that if $p_1 = 0.60$ and $p_2 = 0.80$, the estimated power is 80 percent. Which of the following statements is the most accurate interpretation of the results of Bob's simulation study?

   (a) Bob can be 80 percent confident that $p_1 = 0.60$ and $p_2 = 0.80$.

   (b) If, in fact, $p_1 = 0.60$ and $p_2 = 0.80$, there is an estimated 80 percent chance that Bob's data will lead to rejecting the null hypothesis.

   (c) If, in fact, $p_1 = 0.60$ and $p_2 = 0.80$, there is an estimated 80 percent chance that Bob's data will lead to not rejecting the null hypothesis.

10. Bob plans to collect data to test the null hypothesis that $p_1 = p_2$ versus the third alternative. Bob states, "My simulation study indicates that the estimated power is 20 percent." Which of the following statements is the best criticism of Bob's statement?

    (a) An estimated power of 20 percent is too low to be of any practical value.

    (b) Without knowledge of the sample sizes, one cannot interpret the meaning of the power.

    (c) Without a definition of the populations being compared, one cannot interpret the meaning of the power.

    (d) It is impossible to interpret the power because Bob has failed to report the values of p_1 and p_2 used in his simulation study.

11. Kramer selects a random sample of size 400 from the population of all red cars registered in

California in 1993. He also selects an independent random sample of size 400 from the population of all white cars registered in California in 1993. A car is labeled a failure if during 1993 it was involved in one or more accidents that were reported to the police. Kramer obtains a total of 60 failures, of which only 20 were white cars.

George hears of Kramer's study and tells Elaine, "Buy a white car; they are safer than red cars."

Which of the following is the best criticism of George's comment?

(a) George should reserve comment until he has investigated the power of Kramer's study.

(b) Because it is difficult to obtain a random sample, Kramer's findings must interpreted cautiously.

(c) The color of a car might be confounded with other factors that influence the frequency of accidents.

(d) The difference Kramer found might not be of practical importance.

12. Elaine performs a completely randomized design to compare her ability to shoot three-point baskets from the left (treatment 1) and right (treatment 2) corners of a basketball court. On the assumption of independent Bernoulli trials, Elaine tests the null hypothesis that $p_1 = p_2$ and obtains $\mathbf{P} = 0.2000$ for the third alternative.

George hears of Elaine's study and states, "Clearly, Elaine is equally skilled from either location."

Which of the following is the best criticism of George's comment?

(a) George should reserve comment until he has investigated the power of Elaine's study.

(b) Elaine's shots might not be Bernoulli trials, so the P-value might be invalid.

(c) The location of the shot might be confounded with other factors that influence the outcome.

(d) The difference Elaine found might not be of practical importance.

13. Elaine performs a completely randomized design to compare her ability to shoot three-point baskets from the left (treatment 1) and right (treatment 2) corners of a basketball court. On the assumption of independent Bernoulli trials, Elaine obtains $[0.001, 0.002]$ as the 95 percent confidence interval for $p_1 - p_2$.

George hears of Elaine's study and states, "Clearly, in a shooting contest Elaine should attempt to shoot from the left corner rather than from the right corner."

Which of the following is the best criticism of George's comment?

(a) George should reserve comment until he has investigated the power of Elaine's study.

(b) Elaine's shots might not be Bernoulli trials, so the confidence level might not be valid.

(c) The location of the shot might be confounded with other factors that influence the outcome.

(d) The difference Elaine found might not be of practical importance.

7.4 Solutions to Exam Questions

1. (b).

2. (b).

3. (a) p_1 equals the proportion (*not number*) of white cars in California that did not have an accident during 1993. p_2 equals the proportion of red cars in California that did not have an accident during 1993.

   (b) The point estimates are

$$\hat{p}_1 = 380/400 = 0.95 \text{ and}$$

$$\hat{p}_2 = 360/400 = 0.90.$$

4. First,

$$\sqrt{\frac{(0.95)(0.05)}{400} + \frac{(0.90)(0.10)}{400}} = 0.01854.$$

The confidence interval is

$$(0.950 - 0.900) \pm 1.96(0.01854) =$$

$$0.050 \pm 0.036 = [0.014, 0.086].$$

5. (a) An observational study with Bernoulli trials. (The first 50 shots—subjects—were observed on Monday and the last 50 on Tuesday; hence, randomization was not possible.)

   (b) i. (Answer iii is incorrect because $\hat{p}_1 - \hat{p}_2 = 0.04$; $p_1 - p_2$ is unknown.)

   (c) The 95 percent confidence interval for $p_1 - p_2$ is

$$(0.68 - 0.64) \pm$$

$$1.96\sqrt{\frac{0.68(0.32)}{50} + \frac{0.64(0.36)}{50}} =$$

$$0.04 \pm 0.19 = [-0.15, 0.23].$$

6. A controlled study of finite populations.

7. Yes. Without the assumption of a random sample, the study becomes a CRD, as studied in Chapters 1 through 3. Fisher's test is appropriate for a CRD or a random sample.

8. (a).

9. (b).

10. (d).

11. (c).

12. (a).

13. (d).