Now suppose that we are planning a new study to attempt to determine the mean hemoglobin level for the population of children under the age of 6 who have been exposed to high levels of lead. We are told that the general population of children in this age group has mean hemoglobin level $\mu = 12.29$ g/100 ml and standard deviation $\sigma = 0.85$ g/100 ml. If the true mean hemoglobin level for exposed children is 0.5 g/100 ml lower than that for unexposed children, we want the test to have 80% power to detect this difference. We plan to use a one-sided test conducted at the 0.05 level of significance. What sample size will we need for this study?

We begin by assembling all the pieces necessary to carry out a sample size calculation. Since a one-sided test will be performed at the $\alpha = 0.05$ level, $z_\alpha = 1.645$. We want a power of 0.80; therefore, $\beta = 0.20$ and $z_\beta = 0.84$. The hypothesized mean of the population is $\mu_0 = 12.29$ g/100 ml, and the alternative mean is 0.5 units less than this, or $\mu_1 = 11.79$ g/100 ml. We do not know the standard deviation of the hemoglobin levels for exposed children; however, we are willing to assume that it is the same as that for unexposed children. Therefore, $\sigma = 0.85$ g/100 ml. Putting these pieces together,

$$\begin{aligned} n &= \left[\frac{(z_\alpha + z_\beta)(\sigma)}{(\mu_1 - \mu_0)}\right]^2 \\ &= \left[\frac{(1.645 + 0.84)(0.85)}{(11.79 - 12.29)}\right]^2 \\ &= 17.8. \end{aligned}$$

Thus, a sample of size 18 would be required.

10.8 Review Exercises

1. What is the purpose of a test of hypothesis?
2. Does a hypothesis test ever prove the null hypothesis? Explain.
3. What is a p-value? What does the p-value mean in words?
4. Briefly explain the relationship between confidence intervals and hypothesis testing.
5. Under what circumstances might you use a one-sided test of hypothesis rather than a two-sided test?
6. Describe the two types of errors that can be made when you conduct a test of hypothesis.
7. Explain the analogy between type I and type II errors in a test of hypothesis and the false positive and false negative results that occur in diagnostic testing.
8. List four factors that affect the power of a test.
9. The distribution of diastolic blood pressures for the population of female diabetics between the ages of 30 and 34 has an unknown mean μ_d and standard deviation $\sigma_d = 9.1$ mm Hg. It may be useful to physicians to know whether the mean of this population is equal to the mean diastolic blood pressure of the general population of females in this age group, 74.4 mm Hg [12].
   (a) What is the null hypothesis of the appropriate test?
   (b) What is the alternative hypothesis?
   (c) A sample of ten diabetic women is selected; their mean diastolic blood pressure is $\bar{x}_d = 84$ mm Hg. Using this information, conduct a two-sided test at the $\alpha = 0.05$ level of significance. What is the p-value of the test?
   (d) What conclusion do you draw from the results of the test?
   (e) Would your conclusion have been different if you had chosen $\alpha = 0.01$ instead of $\alpha = 0.05$?

10. *E. canis* infection is a tick-borne disease of dogs that is sometimes contracted by humans. Among infected humans, the distribution of white blood cell counts has an unknown mean μ and a standard deviation σ. In the general population, the mean white blood cell count is 7250/mm^3 [13]. It is believed that persons infected with *E. canis* must on average have lower white blood cell counts.
    (a) What are the null and alternative hypotheses for a one-sided test?
    (b) For a sample of 15 infected persons, the mean white blood cell count is $\bar{x} = 4767/\text{mm}^3$ and the standard deviation is $s = 3204/\text{mm}^3$ [14]. Conduct the test at the $\alpha = 0.05$ level.
    (c) What do you conclude?
11. Body mass index is calculated by dividing a person's weight by the square of his or her height; it is a measure of the extent to which the individual is overweight. For the population of middle-aged men who later develop diabetes mellitus, the distribution of baseline body mass indices is approximately normal with an unknown mean μ and standard deviation σ. A sample of 58 men selected from this group has mean $\bar{x} = 25.0$ kg/m^2 and standard deviation $s = 2.7$ kg/m^2 [15].
    (a) Construct a 95% confidence interval for the population mean μ.
    (b) At the 0.05 level of significance, test whether the mean baseline body mass index for the population of middle-aged men who do develop diabetes is equal to 24.0 kg/m^2, the mean for the population of men who do not. What is the p-value of the test?
    (c) What do you conclude?
    (d) Based on the 95% confidence interval, would you have expected to reject or not to reject the null hypothesis? Why?
12. The population of male industrial workers in London who have never experienced a major coronary event has mean systolic blood pressure 136 mm Hg and mean diastolic blood pressure 84 mm Hg [16]. You might be interested in determining whether these values are the same as those for the population of industrial workers who **have** suffered a coronary event.
    (a) A sample of 86 workers who have experienced a major coronary event has mean systolic blood pressure $\bar{x}_s = 143$ mm Hg and standard deviation $s_s = 24.4$ mm Hg. Test the null hypothesis that the mean systolic blood pressure for the population of industrial workers who have experienced such an event is identical to the mean for the workers who have not, using a two-sided test at the $\alpha = 0.10$ level.

(b) The same sample of men has mean diastolic blood pressure $\bar{x}_d = 87$ mm Hg and standard deviation $s_d = 16.0$ mm Hg. Test the null hypothesis that the mean diastolic blood pressure for the population of workers who have experienced a major coronary event is identical to the mean for the workers who have not.
(c) How do the two groups of workers compare?

13. Over the years, the Food and Drug Administration of the United States (FDA) has worked very hard to avoid making type II errors. A type II error occurs when the FDA approves a drug that is not both safe and effective. Despite the agency's efforts, however, bad drugs do on occasion get through to the public. For example, Omniflox, an antibiotic, had to be recalled less than six months after its approval due to reports of severe adverse reactions, which included a number of deaths. Similarly, Fenoterol, an inhaled drug intended to relieve asthma attacks, was found to increase the risk of death rather than decrease it [17]. Is there any way for the FDA to completely eliminate the occurrence of type II errors? Explain.

14. Data from the Framingham Study allow us to compare the distributions of initial serum cholesterol levels for two populations of males: those who go on to develop coronary heart disease and those who do not. The mean serum cholesterol level of the population of men who do not develop heart disease is $\mu = 219$ mg/100 ml and the standard deviation is $\sigma = 41$ mg/100 ml [18]. Suppose, however, that you do not know the true population mean; instead, you hypothesize that μ is equal to 244 mg/100 ml. This is the mean initial serum cholesterol level of men who eventually develop the disease. Since it is believed that the mean serum cholesterol level for the men who do not develop heart disease cannot be higher than the mean level for men who do, a one-sided test conducted at the $\alpha = 0.05$ level of significance is appropriate.
(a) What is the probability of making a type I error?
(b) If a sample of size 25 is selected from the population of men who do not go on to develop coronary heart disease, what is the probability of making a type II error?
(c) What is the power of the test?
(d) How could you increase the power?
(e) You wish to test the null hypothesis

$$H_0: \mu \geq 244 \text{ mg/100 ml}$$

against the alternative

$$H_A: \mu < 244 \text{ mg/100 ml}$$

at the $\alpha = 0.05$ level of significance. If the true population mean is as low as 219 mg/100 ml, you want to risk only a 5% chance of failing to reject H_0. How large a sample would be required?
(f) How would the sample size change if you were willing to risk a 10% chance of failing to reject a false null hypothesis?

15. In Norway, the distribution of birth weights for full-term infants whose gestational age is 40 weeks is approximately normal with mean $\mu = 3500$ grams and standard deviation $\sigma = 430$ grams [19]. An investigator plans to conduct a study to determine whether the birth weights of full-term babies whose mothers smoked throughout pregnancy have the same mean. If the true mean birth weight for the infants whose mothers smoked is as low as 3200 grams (or as high as 3800 grams), the investigator wants to risk only a 10% chance of failing to detect this difference. A two-sided test conducted at the 0.05 level of significance will be used. What sample size is needed for this study?

16. The Bayley Scales of Infant Development yield scores on two indices—the Psychomotor Development Index (PDI) and the Mental Development Index (MDI)—which can be used to assess a child's level of functioning in each of these areas at approximately one year of age. Among normal healthy infants, both indices have a mean value of 100. As part of a study assessing the development and neurologic status of children who have undergone reparative heart surgery during the first three months of life, the Bayley Scales were administered to a sample of one-year-old infants born with congenital heart disease. The data are contained in the data set `heart` [20] (Appendix B, Table B.12); PDI scores are saved under the variable name `pdi`, while MDI scores are saved under `mdi`.
(a) At the 0.05 level of significance, test the null hypothesis that the mean PDI score for children born with congenital heart disease who undergo reparative heart surgery during the first three months of life is equal to 100, the mean score for healthy children. Use a two-sided test. What is the p-value? What do you conclude?
(b) Conduct the analogous test of hypothesis for the mean MDI score. What do you conclude?
(c) Construct 95% confidence intervals for the true mean PDI score and the true mean MDI score for this population of children with congenital heart disease. Does either of these intervals contain the value 100? Would you have expected that they would?

Bibliography

[1] National Center for Health Statistics, Fulwood, R., Kalsbeek, W., Rifkind, B., Russell-Briefel, R., Muesing, R., LaRosa, J., and Lippel, K., "Total Serum Cholesterol Levels of Adults 20–74 Years of Age: United States, 1976–1980," *Vital and Health Statistics*, Series 11, Number 236, May 1986.

[2] Gauvreau, K., and Pagano, M., "Why 5%?," *Nutrition*, Volume 10, 1994, 93–94.

[3] Kaplan, N. M., "Strategies to Reduce Risk Factors in Hypertensive Patients Who Smoke," *American Heart Journal*, Volume 115, January 1988, 288–294.

[4] Tsou, V. M., Young, R. M., Hart, M. H., and Vanderhoof, J. A., "Elevated Plasma Aluminum Levels in Normal Infants Receiving Antacids Containing Aluminum," *Pediatrics*, Volume 87, February 1991, 148–151.

[5] National Center for Health Statistics, Fulwood, R., Johnson, C. L., Bryner, J. D., Gunter, E. W., and McGrath, C. R., "Hematological and Nutritional Biochemistry Reference Data for Persons 6 Months–74 Years of Age: United States, 1976–1980," *Vital and Health Statistics*, Series 11, Number 232, December 1982.

0.10. Note that p for the test that assumes equal variances is actually a little less than 0.05. However, since we had no reason to believe that the variances should be equal—and, in fact, the sample standard deviations s_1 and s_2 suggest that they are unlikely to be the same—it is safer to use the modified test. Although this test is less precise than the traditional two-sample t-test if the variances truly are identical, it is more reliable if they are not.

11.4 Review Exercises

1. What is the main difference between paired and independent samples?
2. Explain the purpose of paired data. In certain situations, what might be the advantage of using paired samples rather than independent ones?
3. When should you use the two-sample t-test? When must the modified version of the test be applied?
4. What is the rationale for using a pooled estimate of the variance in the two-sample t-test?
5. A crossover study was conducted to investigate whether oat bran cereal helps to lower serum cholesterol levels in hypercholesterolemic males. Fourteen such individuals were randomly placed on a diet that included either oat bran or corn flakes; after two weeks, their low-density lipoprotein (LDL) cholesterol levels were recorded. Each man was then switched to the alternative diet. After a second two-week period, the LDL cholesterol level of each individual was again recorded. The data from this study are shown below [10].

	LDL (mmol/l)	
Subject	**Corn Flakes**	**Oat Bran**
1	4.61	3.84
2	6.42	5.57
3	5.40	5.85
4	4.54	4.80
5	3.98	3.68
6	3.82	2.96
7	5.01	4.41
8	4.34	3.72
9	3.80	3.49
10	4.56	3.84
11	5.35	5.26
12	3.89	3.73
13	2.25	1.84
14	4.24	4.14

(a) Are the two samples of data paired or independent?
(b) What are the appropriate null and alternative hypotheses for a two-sided test?
(c) Conduct the test at the 0.05 level of significance. What is the p-value?
(d) What do you conclude?

6. Suppose that you are interested in determining whether exposure to the organochlorine DDT, which has been used extensively as an insecticide for many years, is associated with breast cancer in women. As part of a study that investigated this issue, blood was drawn from a sample of women diagnosed with breast cancer over a six-year period and from a sample of healthy control subjects matched to the cancer patients on age, menopausal status, and date of blood donation. Each woman's blood level of DDE—an important byproduct of DDT in the human body—was measured, and the difference in levels for each patient and her matched control calculated. A sample of 171 such differences has mean $\bar{d} = 2.7$ ng/ml and standard deviation $s_d = 15.9$ ng/ml [11].
(a) Test the null hypothesis that the mean blood levels of DDE are identical for women with breast cancer and for healthy control subjects. What do you conclude?
(b) Would you expect a 95% confidence interval for the true difference in population mean DDE levels to contain the value 0? Explain.

7. The following data come from a study that examines the efficacy of saliva cotinine as an indicator for exposure to tobacco smoke. In one part of the study, seven subjects—none of whom were heavy smokers and all of whom had abstained from smoking for at least one week prior to the study—were each required to smoke a single cigarette. Samples of saliva were taken from all individuals 2, 12, 24, and 48 hours after smoking the cigarette. The cotinine levels at 12 hours and at 24 hours are shown below [12].

	Cotinine Levels (nmol/l)	
Subject	**After 12 Hours**	**After 24 Hours**
1	73	24
2	58	27
3	67	49
4	93	59
5	33	0
6	18	11
7	147	43

Let μ_{12} represent the population mean cotinine level 12 hours after smoking the cigarette and μ_{24} the mean cotinine level 24 hours after smoking. It is believed that μ_{24} must be lower than μ_{12}.
(a) Construct a one-sided 95% confidence interval for the true difference in population means $\mu_{12} - \mu_{24}$.
(b) Test the null hypothesis that the population means are identical at the $\alpha = 0.05$ level of significance. What do you conclude?

8. A study was conducted to determine whether an expectant mother's cigarette smoking has any effect on the bone mineral content of her otherwise healthy child. A sample of 77 newborns whose mothers smoked during pregnancy has mean bone mineral content $\bar{x}_1 = 0.098$ g/cm and standard deviation $s_1 = 0.026$ g/cm; a sample of 161 infants whose mothers did not smoke has mean $\bar{x}_2 = 0.095$ g/cm and standard deviation $s_2 = 0.025$ g/cm [13]. Assume that the underlying population variances are equal.
   (a) Are the two samples of data paired or independent?
   (b) State the null and alternative hypotheses of the two-sided test.
   (c) Conduct the test at the 0.05 level of significance. What do you conclude?

9. In an investigation of pregnancy-induced hypertension, one group of women with this disorder was treated with low-dose aspirin, and a second group was given a placebo. A sample consisting of 23 women who received aspirin has mean arterial blood pressure 111 mm Hg and standard deviation 8 mm Hg; a sample of 24 women who were given the placebo has mean blood pressure 109 mm Hg and standard deviation 8 mm Hg [14].
   (a) At the 0.01 level of significance, test the null hypothesis that the two populations of women have the same mean arterial blood pressure.
   (b) Construct a 99% confidence interval for the true difference in population means. Does this interval contain the value 0?

10. As part of the Women's Health Trial, one group of women were encouraged to adopt a low-fat diet while a second group received no dietary counseling. A year later, the women in the intervention group had successfully maintained their diets. At that time, a study was undertaken to determine whether their husbands also had a reduced level of fat intake [15].
   (a) In the intervention group, a sample of 156 husbands has mean daily fat intake $\bar{x}_1 = 54.8$ grams and standard deviation $s_1 = 28.1$ grams. In the control group, a sample of 148 husbands has mean intake $\bar{x}_2 = 69.5$ grams and standard deviation $s_2 = 34.7$ grams. Calculate separate 95% confidence intervals for the true mean fat intakes of men in each group. Use these intervals to construct a graph like Figure 11.3. Does the graph suggest that the two population means are likely to be equal to each other?
   (b) Formally test the null hypothesis that the two groups of men have the same mean dietary fat intake using a two-sided test. What do you conclude?
   (c) Construct a 95% confidence interval for the true difference in population means.
   (d) A researcher might also be interested in knowing whether the men differ with respect to the intake of other types of food, such as protein or carbohydrates. In the intervention group, the husbands have mean daily carbohydrate intake $\bar{x}_1 = 172.5$ grams and standard deviation $s_1 = 68.8$ grams. In the control group, the men have mean carbohydrate intake $\bar{x}_2 = 185.5$ grams and standard deviation $s_2 = 69.0$ grams. Test the null hypothesis that the two populations have the same mean carbohydrate intake. What do you conclude?

11. The table below compares the levels of carboxyhemoglobin for a group of nonsmokers and a group of cigarette smokers. Sample means and standard deviations are shown [16]. It is believed that the mean carboxyhemoglobin level of the smokers must be higher than the mean level of the nonsmokers. There is no reason to assume that the underlying population variances are identical.

Group	n	Carboxyhemoglobin (%)
Nonsmokers	121	$\bar{x} = 1.3, s = 1.3$
Smokers	75	$\bar{x} = 4.1, s = 2.0$

   (a) What are the null and alternative hypotheses of the one-sided test?
   (b) Conduct the test at the 0.05 level of significance. What do you conclude?

12. Suppose that you wish to compare the characteristics of tuberculosis meningitis in patients infected with HIV and those who are not infected. In particular, you would like to determine whether the two populations have the same mean age. A sample of 37 infected patients has mean age $\bar{x}_1 = 27.9$ years and standard deviation $s_1 = 5.6$ years; a sample of 19 patients who are not infected has mean age $\bar{x}_2 = 38.8$ years and standard deviation $s_2 = 21.7$ years [17].
   (a) Test the null hypothesis that the two populations of patients have the same mean age at the 0.05 level of significance.
   (b) Do you expect that a 95% confidence interval for the true difference in population means would contain the value 0? Why or why not?

13. Consider the numbers of community hospital beds per 1000 population that are available in each state in the United States and in the District of Columbia. The data for both 1980 and 1986 are on your disk in a data set called `bed` [18] (Appendix B, Table B.13). The values for 1980 are saved under the variable name `bed80`; those for 1986 are saved under `bed86`. A second data set, called `bed2`, contains the same information in a different format. The numbers of beds per 1000 population for both calendar years are saved under the variable name `bed`, and an indicator of year under the name `year`.
   (a) Generate descriptive statistics for the numbers of hospital beds in each year.
   (b) Since there are two observations for each state—one for 1980 and one for 1986—the data are actually paired. A common error in analyzing this type of data is to ignore the pairing and assume that the samples are independent. Compare the mean number of community hospital beds per 1000 population in 1980 to the mean number of beds in 1986 using the two-sample t-test. What do you conclude?
   (c) Now compare the mean number of beds in 1980 to the mean number in 1986 using the paired t-test.
   (d) Comment on the differences between the two tests. Do you reach the same conclusion in each case?
   (e) Generate a 95% confidence interval for the true difference in the mean number of hospital beds in 1980 and 1986.

14. The data set `lowbwt` contains information for a sample of 100 low birth weight infants born in two teaching hospitals in Boston, Massachusetts [19] (Appendix B,

Table B.7). Measurements of systolic blood pressure are saved under the variable name `sbp` and indicators of gender—with 1 representing a male and 0 a female—under the name `sex`.

(a) Construct a histogram of systolic blood pressure measurements for this sample. Based on the graph, do you believe that blood pressure is approximately normally distributed?

(b) Test the null hypothesis that among low birth weight infants, the mean systolic blood pressure for boys is equal to the mean for girls. Use a two-sided test at the 0.05 level of significance. What do you conclude?

15. The Bayley Scales of Infant Development yield scores on two indices—the Psychomotor Development Index (PDI) and the Mental Development Index (MDI)—that can be used to assess a child's level of functioning at approximately one year of age. As part of a study investigating the development and neurologic status of children who had undergone reparative heart surgery during the first three months of life, the Bayley Scales were administered to a sample of one-year-old infants born with congenital heart disease. The children had been randomized to one of two different treatment groups, known as "circulatory arrest" and "low-flow bypass." The groups differed in the specific way in which the reparative surgery was performed. Unlike circulatory arrest, low-flow bypass maintains continuous circulation through the brain; although it is felt to be preferable by some physicians, it also has its own associated risk of brain injury. The data for this study are saved in the data set `heart` [20] (Appendix B, Table B.12). PDI scores are saved under the variable name `pdi`, MDI scores under `mdi`, and indicators of treatment group under `trtment`. For this variable, 0 represents circulatory arrest and 1 is for low-flow bypass.

(a) At the 0.05 level of significance, test the null hypothesis that the mean PDI score at one year of age for the circulatory arrest treatment group is equal to the mean PDI score for the low-flow group. What is the p-value for this test?

(b) Test the null hypothesis that the mean MDI scores are identical for the two treatment groups. What is the p-value?

(c) What do these tests suggest about the relationship between a child's surgical treatment group during the first three months of life and his or her subsequent developmental status at one year of age?

Bibliography

[1] Packard, F. R., *The Life and Times of Ambroise Paré*, 1510–1590, New York: Paul B. Hoeber, 1921.

[2] Allred, E. N., Bleecker, E. R., Chaitman, B. R., Dahms, T. E., Gottlieb, S. O., Hackney, J. D., Hayes, D., Pagano, M., Selvester, R. H., Walden, S. M., and Warren, J., "Acute Effects of Carbon Monoxide Exposure on Individuals with Coronary Artery Disease," *Health Effects Institute Research Report Number 25*, November 1989.

[3] Markowski, C. A., and Markowski, E. P., "Conditions for the Effectiveness of a Preliminary Test of Variance," *The American Statistician*, Volume 44, November 1990, 322–326.

[4] Moser, B. K., and Stevens, G. R., "Homogeneity of Variance in the Two-Sample Means Test," *The American Statistician*, Volume 46, February 1992, 19–21.

[5] Zempsky, W. T., Rosenstein, B. J., Carroll, J. A., and Oski, F. A., "Effect of Pancreatic Enzyme Supplements on Iron Absorption," *American Journal of Diseases of Children*, Volume 143, August 1989, 966–972.

[6] Satterthwaite, F. W., "An Approximate Distribution of Estimates of Variance Components," *Biometrics Bulletin*, Volume 2, December 1946, 110–114.

[7] SHEP Cooperative Research Group, "Prevention of Stroke by Antihypertensive Drug Treatment in Older Persons with Isolated Systolic Hypertension: Final Results of the Systolic Hypertension in the Elderly Program (SHEP)," *Journal of the American Medical Association*, Volume 265, June 26, 1991, 3255–3264.

[8] Kien, C. L., Liechty, E. A., and Mullett, M. D., "Effects of Lactose Intake on Nutritional Status in Premature Infants," *Journal of Pediatrics*, Volume 116, March 1990, 446–449.

[9] Barrett-Connor, E. L., Cohn, B. A., Wingard, D. L., and Edelstein, S. L., "Why Is Diabetes Mellitus a Stronger Risk Factor for Fatal Ischemic Heart Disease in Women Than in Men?," *Journal of the American Medical Association*, Volume 265, February 6, 1991, 627–631.

[10] Anderson, J. W., Spencer, D. B., Hamilton, C. C., Smith, S. F., Tietyen, J., Bryant, C. A., and Oeltgen, P., "Oat-Bran Cereal Lowers Serum Total and LDL Cholesterol in Hypercholesterolemic Men," *American Journal of Clinical Nutrition*, Volume 52, September 1990, 495–499.

[11] Wolff, M. S., Toniolo, P. G., Lee, E. W., Rivera, M., and Dubin, N., "Blood Levels of Organochlorine Residues and Risk of Breast Cancer," *Journal of the National Cancer Institute*, Volume 85, April 21, 1993, 648–652.

[12] DiGiusto, E., and Eckhard, I., "Some Properties of Saliva Cotinine Measurements in Indicating Exposure to Tobacco Smoking," *American Journal of Public Health*, Volume 76, October 1986, 1245–1246.

[13] Venkataraman, P. S., and Duke, J. C., "Bone Mineral Content of Healthy, Full-term Neonates: Effect of Race, Gender, and Maternal Cigarette Smoking," *American Journal of Diseases of Children*, Volume 145, November 1991, 1310–1312.

[14] Schiff, E., Barkai, G., Ben-Baruch, G., and Mashiach, S., "Low-Dose Aspirin Does Not Influence the Clinical Course of Women with Mild Pregnancy-Induced Hypertension," *Obstetrics and Gynecology*, Volume 76, November 1990, 742–744.

[15] Shattuck, A. L., White, E., and Kristal, A. R., "How Women's Adopted Low-Fat Diets Affect Their Husbands," *American Journal of Public Health*, Volume 82, September 1992, 1244–1250.

[16] Jarvis, M. J., Tunstall-Pedoe, H., Feyerabend, C., Vesey, C., and Saloojee, Y., "Comparison of Tests Used to Distinguish Smokers from Nonsmokers," *American Journal of Public Health*, Volume 77, November 1987, 1435–1438.

[17] Berenguer, J., Moreno, S., Laguna, F., Vicente, T., Adrados, M., Ortega, A., González-LaHoz, J., and Bouza, E., "Tuberculosis Meningitis in Patients Infected with the Human Immunodeficiency Virus," *The New England Journal of Medicine*, Volume 326, March 5, 1992, 668–672.

[18] National Center for Health Statistics, *Health United States 1988*, Public Health Service, Hyattsville, MD: March 1989.

TABLE 12.2
Stata output illustrating the one-way analysis of variance and the Bonferroni multiple comparisons procedure

Analysis of Variance

Source	SS	df	MS	F	Prob > F
Between groups	1292.40	2	646.20	45.38	0.0000
Within groups	1822.72	128	14.24		
Total	3115.12	130	23.96		

Comparison of wtchange by group
(Bonferroni)

Row Mean–Col Mean	1	2
2	-3.20 0.000	
3	-7.80 0.000	-4.60 0.000

12.4 Review Exercises

1. When you are testing the equality of several population means, what problems arise if you attempt to perform all possible pairwise t-tests?
2. What is the idea behind the one-way analysis of variance? What two measures of variation are being compared?
3. What are the properties of the F distribution?
4. Describe the purpose of the Bonferroni correction for multiple comparisons.
5. Consider the F distribution with 8 and 16 degrees of freedom.
   (a) What proportion of the area under the curve lies to the right of $F = 2.09$?
   (b) What value of F cuts off the upper 1% of the distribution?
   (c) What proportion of the area under the curve lies to the left of $F = 4.52$?
6. Consider the F distribution with 3 and 30 degrees of freedom.
   (a) What proportion of the area under the curve lies to the right of $F = 5.24$?
   (b) What proportion of the area under the curve lies to the left of $F = 2.92$?
   (c) What value of F cuts off the upper 2.5% of the distribution?
   (d) What value of F cuts off the upper 0.1%?
7. A study of patients with insulin-dependent diabetes was conducted to investigate the effects of cigarette smoking on renal and retinal complications. Before examining the results of the study, you wish to compare the baseline measures of sys-

tolic blood pressure across four different subgroups: nonsmokers, current smokers, ex-smokers, and tobacco chewers. A sample is selected from each subgroup; the relevant data are shown below [4]. Means and standard deviations are expressed in mm Hg. Assume that systolic blood pressure is normally distributed.

	n	$\bar{x}$	s
Nonsmokers	269	115	13.4
Current Smokers	53	114	10.1
Ex-smokers	28	118	11.6
Tobacco Chewers	9	126	12.2

   (a) Calculate the estimate of the within-groups variance.
   (b) Calculate the estimate of the between-groups variance.
   (c) At the 0.05 level of significance, test the null hypothesis that the mean systolic blood pressures of the four groups are identical. What do you conclude?
   (d) If you find that the population means are not all equal, use the Bonferroni multiple comparisons procedure to determine where the differences lie. What is the significance level of each individual test?
8. One of the goals of the Edinburgh Artery Study was to investigate the risk factors for peripheral arterial disease among persons 55 to 74 years of age. You wish to compare mean LDL cholesterol levels, measured in mmol/liter, among four different populations of subjects: patients with intermittent claudication or interruptions in movement, those with major asymptomatic disease, those with minor asymptomatic disease, and those with no evidence of disease at all. Samples are selected from each population; summary statistics are shown below [5].

	n	$\bar{x}$	s
Intermittent Claudication	73	6.22	1.62
Major Asymptomatic Disease	105	5.81	1.43
Minor Asymptomatic Disease	240	5.77	1.24
No Disease	1080	5.47	1.31

   (a) Test the null hypothesis that the mean LDL cholesterol levels are the same for each of the four populations. What are the degrees of freedom associated with this test?
   (b) What do you conclude?
   (c) What assumptions about the data must be true for you to use the one-way analysis of variance technique?
   (d) Is it necessary to take an additional step in this analysis? If so, what is it? Explain.
9. A study was conducted to assess the performance of outpatient substance abuse treatment centers. Three different types of units were evaluated: private for-profit (FP), private not-for-profit (NFP), and public. Among the performance measures considered were minutes of individual therapy per session and minutes of group

therapy per session. Samples were selected from each type of treatment center; summary statistics are shown in the table [6].

Treatment Centers	Individual Therapy			Group Therapy		
	n	$\bar{x}$	s	n	$\bar{x}$	s
FP	37	49.46	15.47	30	105.83	42.91
NFP	312	54.76	11.41	296	98.68	31.27
Public	169	53.25	11.08	165	94.17	27.12

(a) Given these numbers, how do the different types of treatment centers compare with respect to average minutes of therapy per session?
(b) Construct 95% confidence intervals for the mean minutes of individual therapy per session in each type of treatment center. Do the same for the mean minutes of group therapy. In either case, do you notice anything that suggests that the population means may not be the same?
(c) Test the null hypothesis that the mean minutes of individual therapy per session are identical for each type of center. If necessary, carry out a multiple-comparisons procedure.
(d) Test the null hypothesis that the mean minutes of group therapy per session are the same for each type of center. Again, use a multiple comparisons procedure if necessary.
(e) How do the different treatment centers compare to each other?

10. For the study discussed in the text that investigates the effect of carbon monoxide exposure on patients with coronary artery disease, baseline measures of pulmonary function were examined across medical centers. Another characteristic that you might wish to investigate is age. The relevant measurements are saved on your disk in a data set called `cad` (Appendix B, Table B.14). Values of age are saved under the variable name `age`, and indicators of center are saved under `center`.
(a) Do you feel that it is appropriate to use one-way analysis of variance to evaluate these data? Explain.
(b) For each medical center, what are the sample mean and standard deviation for the values of age?
(c) What is the between-groups estimate of the variance?
(d) What is the within-groups estimate of the variance?
(e) Test the null hypothesis that at the time when the study was conducted, the mean ages for men at all three centers were identical. What do you conclude?

11. The data set `lowbwt` contains information for a sample of 100 low birth weight infants born in two teaching hospitals in Boston, Massachusetts [7] (Appendix B, Table B.7). Systolic blood pressure measurements are saved under the variable name `sbp` and indicators of gender—where 1 represents a male and 0 a female—under the name `sex`.
(a) Assuming equal variances for males and females, use the two-sample t-test to evaluate the null hypothesis that among low birth weight infants, the mean systolic blood pressure for girls is identical to that for boys.

(b) Even though there are only two populations instead of three or more, test the same null hypothesis using the one-way analysis of variance.
(c) In this text, the statement has been made that in the case of two independent samples, the F-test used in the one-way analysis of variance reduces to the two-sample t-test. Do you believe this to be true? Explain briefly.

Bibliography

[1] Allred, E. N., Bleecker, E. R., Chaitman, B. R., Dahms, T. E., Gottlieb, S. O., Hackney, J. D., Hayes, D., Pagano, M., Selvester, R. H., Walden, S. M., and Warren, J., "Acute Effects of Carbon Monoxide Exposure on Individuals with Coronary Artery Disease," *Health Effects Institute Research Report Number 25*, November 1989.

[2] Ott, L., *An Introduction to Statistical Methods and Data Analysis*, North Scituate, MA: Duxbury, 1977.

[3] Wood, P. D., Stefanick, M. L., Dreon, D. M., Frey-Hewitt, B., Garay, S. C., Williams, P. T., Superko, H. R., Fortmann, S. P., Albers, J. J., Vranizan, K. M., Ellsworth, N. M., Terry, R. B., and Haskell, W. L., "Changes in Plasma Lipids and Lipoproteins in Overweight Men During Weight Loss Through Dieting as Compared with Exercise," *The New England Journal of Medicine*, Volume 319, November 3, 1988, 1173–1179.

[4] Chase, H. P., Garg, S. K., Marshall, G., Berg, C. L., Harris, S., Jackson, W. E., and Hamman, R. E., "Cigarette Smoking Increases the Risk of Albuminuria Among Subjects with Type I Diabetes," *Journal of the American Medical Association*, Volume 265, February 6, 1991, 614–617.

[5] Fowkes, F. G. R., Housley, E., Riemersma, R. A., Macintyre, C. C. A., Cawood, E. H. H., Prescott, R. J., and Ruckley, C. V., "Smoking, Lipids, Glucose Intolerance, and Blood Pressure as Risk Factors for Peripheral Atherosclerosis Compared with Ischemic Heart Disease in the Edinburgh Artery Study," *American Journal of Epidemiology*, Volume 135, February 15, 1992, 331–340.

[6] Wheeler, J. R. C., Fadel, H., and D'Aunno, T. A., "Ownership and Performance of Outpatient Substance Abuse Treatment Centers," *American Journal of Public Health*, Volume 82, May 1992, 711–718.

[7] Leviton, A., Fenton, T., Kuban, K. C. K., and Pagano, M., "Labor and Delivery Characteristics and the Risk of Germinal Matrix Hemorrhage in Low Birth Weight Infants," *Journal of Child Neurology*, Volume 6, October 1991, 35–40.