Chapter 8

HYPOTHESIS TESTING: TWO SAMPLE INFERENCE

REVIEW OF KEY CONCEPTS

SECTION 8.1 Paired *t* Test

A study was performed to compare blood-pressure readings obtained from an automated machine versus standard blood-pressure recordings from a mercury sphygmomanometer obtained by a human observer. The study was performed using four different machines, with results given as follows:

		Machine		Human		Difference	
Machine	*n*	Mean	sd	Mean	sd	Mean	sd
A	98	142.5	21.0	142.0	18.1	0.5	11.2
B	84	134.1	22.5	133.6	23.2	0.5	12.1
C	98	147.9	20.3	133.9	18.3	14.0	11.7
D	62	135.4	16.7	128.5	19.0	6.9	13.6

To analyze these data, we let $x_{1i} =$ machine reading for the ith person, $x_{2i} =$ human reading for the ith person, $d_i = x_{1i} - x_{2i} \sim N(\mu_d, \sigma_d^2)$. We wish to test the hypothesis $H_0: \mu_d = 0$ versus $H_1: \mu_d \neq 0$.

We use the *paired t test*. The test statistic t is based on the mean difference $(\bar{d})$ and is given by:

$$t = \frac{\bar{d} - 0}{\frac{s_d}{\sqrt{n}}} \sim t_{n-1} \text{ under } H_0, \text{ where}$$

$$s_d^2 = \sum_{i=1}^{n} \frac{(d_i - \bar{d})^2}{n-1}$$

We reject H_0 if $t > t_{n-1,\, 1-\alpha/2}$ or $t < t_{n-1,\, \alpha/2}$ and accept H_0 otherwise.

Alternatively,

$$p\text{-value} = 2 \times Pr(t_{n-1} > t) \text{ if } t \geq 0$$
$$= 2 \times Pr(t_{n-1} < t) \text{ if } t < 0$$

The results are as follows for each of the four machines:

1. $t = 0.5/\left(11.2/\sqrt{98}\right) = 0.5/1.13 = 0.44 \sim t_{97}$; since $t_{120,\ .75} = 0.677$ and $0.44 < 0.677 < t_{97,\ .75}$, it follows that $p > 2 \times (1 - .75) = .50$
2. $t = 0.5/\left(12.1/\sqrt{84}\right) = 0.5/1.32 = 0.38 \sim t_{83}$; since $t_{83,\ .75} > t_{120,\ .75} = 0.677 > 0.38$, it follows that $p > .50$
3. $t = 14.0/\left(11.7/\sqrt{98}\right) = 14.0/1.18 = 11.85 \sim t_{97}$; since $t > t_{60,\ .9995} = 3.460 > t_{97,\ .9995}$, it follows that $p < 2 \times .0005 = .001$
4. $t = 6.9/\left(13.6/\sqrt{62}\right) = 6.9/1.73 = 3.99$; since $t > 3.460 = t_{60,\ .9995} > t_{61,\ .9995}$, it follows that $p < 0.001$

For Machines A and B, the machine and human readings are comparable; for Machines C and D, the machine reads higher.

In each case, one can develop a 95% CI for μ_d given by

$$\bar{d} \pm t_{n-1,\ .975} \frac{s_d}{\sqrt{n}}$$

For example, for Machine D this is given by

$$6.9 \pm t_{61,\ .975}(1.73)$$
$$6.9 \pm 1.999(1.73) = 6.9 \pm 3.5 = (3.4,\ 10.4)$$

SECTION 8.2 Two-Sample *t* test for Independent Samples—Equal Variances

Infant BP's were collected during the first few days of life in the newborn nursery and again at one week, 1 month, 6 months, 12 months, 18 months, 24 months. One question is, How does the level of consciousness of the infant affect BP? We have the following results for 1 month SBP (systolic blood pressure):

Level of consciousness	Mean	sd	*n*
Quiet sleep	81.9	9.8	64
Awake and quiet	86.1	10.3	175

Notice that in this case, we have two independent samples consisting of different groups of infants, while in Section 8.1, we had two paired samples where each person was used as their own control. Let us assume that σ is the same in both groups. We will check this later. We wish to test the hypothesis $H_0: \mu_1 = \mu_2,\ \sigma_1^2 = \sigma_2^2 = \sigma^2$ versus $H_1: \mu_1 \neq \mu_2,\ \sigma_1^2 = \sigma_2^2 = \sigma^2$. We assume

$$x_i \sim N\left(\mu_1,\ \sigma_1^2\right) \qquad i = 1,\ \ldots,\ n_1$$
$$y_j \sim N\left(\mu_2,\ \sigma_2^2\right) \qquad j = 1,\ \ldots,\ n_2$$

where

x_i = SBP for the ith infant in the quiet sleep group
y_j = SBP for the jth infant in the awake and quiet group
n_1 does not have to be equal to n_2

We base our test on $\bar{x} - \bar{y}$ and use the test statistic

$$t = \frac{\bar{x} - \bar{y}}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1+n_2-2} \text{ under } H_0, \text{ where}$$

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

= pooled variance estimate over the 2 samples

We reject H_0 if $t > t_{n_1+n_2-2,\ 1-\alpha/2}$ or $t < t_{n_1+n_2-2,\ \alpha/2}$ and accept H_0 otherwise.

Alternatively,

$$p\text{-value} = 2 \times Pr\left(t_{n_1+n_2-2} > t\right) \text{ if } t > 0$$

$$p\text{-value} = 2 \times Pr\left(t_{n_1+n_2-2} < t\right) \text{ if } t \le 0$$

This test is called the *two-sample t test for independent samples with equal variances*.

In our case,

$$s^2 = \frac{63(9.8)^2 + 174(10.3)^2}{237} = 103.42$$

$$t = \frac{81.9 - 86.1}{\sqrt{103.42\left(\frac{1}{64} + \frac{1}{175}\right)}} = \frac{-4.2}{1.49}$$

$$= -2.83 \sim t_{239-2} = t_{237}$$

We have $t_{120,\ .995} = 2.617$, $t_{120,\ .9995} = 3.373$. Also, $t_{\infty,\ .995} = 2.576$, $t_{\infty,\ .9995} = 3.291$. Since

$$2.617 < |t| < 3.373,\ 2.576 < |t| < 3.291,$$

and $120 < 237 < \infty$, it follows that $.001 < p < .01$. Thus, there is a significant effect of sleep status on SBP.

A $100\% \times (1 - \alpha)$ CI for $\mu_1 - \mu_2$ is given by

$$\bar{x} - \bar{y} \pm t_{n_1+n_2-2,\ 1-\alpha/2}\ s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

In this example, the 95% CI is given by

$$-4.2 \pm t_{237,\ .975}(1.49)$$

Using Excel, we estimate $t_{237,\ 0.975}$ by 1.97. Thus, the 95% CI is

$$-4.2 \pm 1.97(1.49) = -4.2 \pm 2.9 = (-7.1,\ -1.3)$$

SECTION 8.3 *F* Test for the Equality of Two Variances

How can we verify in the previous example that the underlying sd's are in fact the same in the 2 groups? We wish to test the hypothesis $H_0: \sigma_1^2 = \sigma_2^2,\ \mu_1 \ne \mu_2$ versus $H_1: \sigma_1^2 \ne \sigma_2^2,\ \mu_1 \ne \mu_2$.

We use the variance ratio $F = s_1^2/s_2^2$ as a test statistic. Under H_0, s_1^2/s_2^2 follows an F distribution with $n_1 - 1$ and $n_2 - 1$ *df*. The percentiles of the F distribution are given in Table 9 (Appendix, text). The pth percentile of an F distribution with d_1 and d_2 df is denoted by $F_{d_1,\, d_2,\, p}$.

8.3.1 Characteristics of the *F* Distribution

1. It is a skewed distribution.
2. $F_{n_1-1,\, n_2-1,\, p} = 1/F_{n_2-1,\, n_1-1,\, 1-p}$

This enables us to obtain lower percentiles of an F distribution indirectly from the inverse of the corresponding upper percentile of an F distribution with the *df* reversed. If we use F tables this is important because these tables usually only provide upper percentiles.

For example, suppose we wish to find the upper 2.5% point of an F distribution with 3 and 8 *df*. From Table 9 (Appendix, text),

$$F_{3,\, 8,\, .975} = 5.42$$

Now suppose we wish to find the lower 2.5% point of an F distribution with 3 and 8 *df*. From Table 9,

$$F_{3,\, 8,\, .025} = \frac{1}{F_{8,\, 3,\, .975}} = \frac{1}{14.54} = 0.069$$

We will use the F distribution to test hypotheses concerning the equality of two variances. We reject H_0 if $F = s_1^2/s_2^2$ is either too large or too small. We reject H_0 if $F > F_{n_1-1,\, n_2-1,\, 1-\alpha/2}$ or $F < F_{n_1-1,\, n_2-1,\, \alpha/2}$ and accept H_0 otherwise.

This test is called the *F test for the equality of two variances*.

It is usually easier to put the larger variance in the numerator so that $F > 1$. In our example, from Section 8.2,

$$F = \left(\frac{10.3}{9.8}\right)^2 = 1.105 \sim F_{174,\, 63} \text{ under } H_0$$

From the F table, $F_{\infty,\, 120,\, .975} = 1.31$. Since the percentiles of the F distribution decrease as either the numerator or denominator *df* increases, it follows that $F = 1.105 < 1.31 = F_{\infty,\, 120,\, .975} < F_{174,\, 63,\, .975}$. Also, since $F > 1$, F is also $> F_{174,\, 63,\, .025}$. Therefore, $p > .05$, and there is no significant difference between the variances of the two groups. Another approach is to use the FDIST function of Excel to compute $Pr\left(F_{174,\, 63} > 1.105\right) = .3295$ with p-value $= 2 \times .3295 = .660$ (see Appendix for details).

SECTION 8.4 *t* Test for Independent Samples—Unequal Variances

In the infant BP study referred to in Section 8.2, the following question was asked at 6 months of life: Do you add salt to the baby food? The mean SBPs of the infants were obtained in the subgroups of mothers who answered yes or no to this question with results as follows:

Do you add salt to the baby food?	Mean SBP	sd	*n*
Yes	95.0	13.0	23
No	91.3	9.5	503

We first perform the *F* test to test the hypothesis $H_0: \sigma_1^2 = \sigma_2^2$ versus $H_1: \sigma_1^2 \neq \sigma_2^2$. The test statistic is $F = (13.0/9.5)^2 = 1.873 \sim F_{22,\,502}$ under H_0. The upper critical value, obtained using the FINV function of Excel (see Appendix), is $F_{22,\,502,\,.975} = 1.70$. Since $F = 1.873 > 1.70$, this implies that $p < .05$ and the variances are significantly different. Therefore, we cannot use the equal-variance *t* test as we did in Section 8.2. Instead, we wish to test the hypothesis

$$H_0: \mu_1 = \mu_2,\ \sigma_1^2 \neq \sigma_2^2 \text{ versus } H_1: \mu_1 \neq \mu_2,\ \sigma_1^2 \neq \sigma_2^2$$

We will use the *two-sample t test with unequal variances*. The test statistic is

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim t_d$$

The issue is what degrees of freedom to use for *d*.

There is no exact solution. The most common approach is to use the *Satterthwaite approximation* for *d* given by

$$d' = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}}$$

We reject H_0 if $t > t_{d'',\,1-\alpha/2}$ or $t < t_{d'',\,\alpha/2}$, and accept H_0 otherwise, where d'' is the largest integer $\leq d'$. The *p*-value is given by

$$\begin{aligned} p\text{-value} &= 2 \times Pr(t_{d''} > t) \text{ if } t > 0 \\ &= 2 \times Pr(t_{d''} < t) \text{ if } t \leq 0 \end{aligned}$$

This test is called the *two-sample t test for independent samples with unequal variances*. In this case,

$$\frac{s_1^2}{n_1} = \frac{13^2}{23} = 7.35$$

$$\frac{s_2^2}{n_2} = \frac{9.5^2}{503} = 0.18$$

$$d' = \frac{7.53^2}{\frac{7.35^2}{22} + \frac{0.18^2}{502}} = \frac{56.66}{2.45} = 23.1 \ df$$

We round *down* to the nearest integer $df = 23 = d''$. Therefore,

$$t = \frac{95.0 - 91.3}{\sqrt{\frac{13.0^2}{23} + \frac{9.5^2}{503}}} = \frac{3.7}{2.74} = 1.349 \sim t_{23} \text{ under } H_0.$$

Since $t_{23,\,.90} = 1.319$, $t_{23,\,.95} = 1.714$ and $1.319 < t < 1.714$, it follows that $.05 < p/2 < .10$ or $.10 < p < .20$ (i.e., no significant difference in mean SBP between the two groups of babies). The exact *p*-value obtained from Excel $= 2 \times \Pr(t_{23} > 1.349) = \text{TDIST}(1.349,\ 23,\ 2) = .191$.

The general strategy for comparing means from two independent samples is to first perform the *F* test for the equality of two variances. If the variances are significantly different, then use the two-sample *t* test for independent samples with unequal variances; otherwise, use the two-sample *t* test for independent samples with equal variances.

SECTION 8.5 Sample-Size Determination for Comparing Two Means from Independent Samples

Suppose we wish to test the hypothesis $H_0: \mu_1 = \mu_2,\ \sigma_1^2 \neq \sigma_2^2$ versus $H_1: \mu_1 \neq \mu_2,\ \sigma_1^2 \neq \sigma_2^2$. We intend to use a test with a two-sided significance level $= \alpha$, power $= 1 - \beta$, and projected absolute difference between the two means under $H_1 = |\mu_1 - \mu_2| = \Delta$. How many subjects do we need in each group?

If the sample sizes are expected to be the same in the two groups, then

$$n_1 = n_2 = n = \frac{\left(z_{1-\alpha/2} + z_{1-\beta}\right)^2 (\sigma_1^2 + \sigma_2^2)}{\Delta^2}$$
$$= \text{sample size in each group}$$

If the sample sizes are expected to be different, and $n_2 = kn_1$, where k is the projected ratio of sample sizes in group 2 versus group 1, then

$$n_1 = \frac{\left(z_{1-\alpha/2} + z_{1-\beta}\right)^2 \left(\sigma_1^2 + \sigma_2^2/k\right)}{\Delta^2}$$
$$n_2 = kn_1$$

Example: Blood flow is known to be very reduced in certain parts of the brain in people with stroke. Suppose we plan an observational study to compare cerebral blood flow in "stroke-prone" people (with transient ischemic attacks, or TIAs, where neurologic function is temporarily lost) vs stroke-age normal controls (age 50–79). We will use mean blood flow over the entire brain as an outcome variable. Suppose we expect

	Mean	sd
Stroke prone	45	10
Normal	50	10

How many subjects do we need per group to have a 90% chance of finding a significant difference, if we plan to enroll an equal number of subjects per group? In this case,

$$\alpha = .05,\ \beta = .10,\ \sigma_1 = \sigma_2 = 10,\ \Delta = 5$$
$$z_{1-\alpha/2} = z_{.975} = 1.96$$
$$z_{1-\beta} = z_{.90} = 1.28$$
$$n = \frac{(1.96 + 1.28)^2 (200)}{25} = (10.50)(8) = 83.98 \text{ or } 84 \text{ subjects}$$

We need 84 subjects per group, or 168 subjects in total.

We can also derive power as a function of sample size as follows:

$$\text{Power} = 1 - \beta = \Phi\left(-z_{1-\alpha/2} + \frac{\Delta}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}\right)$$

where $(\mu_1,\ \sigma_1^2),\ (\mu_2,\ \sigma_2^2)$ are the means and variances of the two respective samples,

$$\Delta = |\mu_1 - \mu_2|,\ \alpha = \text{type I error}$$

Suppose 50 people are enrolled per group in the preceding study and a two-sided test will be used with significance level $= .05$. How much power will we have? We have $\alpha = .05$, $n_1 = n_2 = 50$, $\Delta = 5$, $\sigma_1 = \sigma_2 = 10$.

Using the power formula,

$$\text{Power} = \Phi\left(-1.96 + \frac{5}{\sqrt{10^2/50 + 10^2/50}}\right) = \Phi\left(-1.96 + \frac{5}{2}\right) = \Phi(0.540) = .705$$

Thus, we will have 71% power.

PROBLEMS

Ophthalmology
In a study of the natural history of retinitis pigmentosa (RP), 94 RP patients were followed for 3 years [1]. Among 90 patients with complete follow-up, the mean $\pm 1\,se$ of ln (visual-field loss) over 1, 2, and 3 years was 0.02 ± 0.04, 0.08 ± 0.05, and 0.14 ± 0.07, respectively.

8.1 What test procedure can be used to test for changes in ln (visual field) over any given time period?

8.2 Implement the procedure in Problem 8.1 to test for significant changes in visual field over 1 year. Report a *p*-value.

8.3 Answer Problem 8.2 for changes over 2 years.

8.4 Answer Problem 8.2 for changes over 3 years.

8.5 Find the upper 5th percentile of an *F* distribution with 24 and 30 *df*.

8.6 Suppose we have two normally distributed samples of sizes 9 and 15 with sample standard deviations of 13.7 and 7.2, respectively. Test if the variances are significantly different in the two samples.

Pathology
Refer to the heart-weight data in Table 6.1 (Chapter 6, Study Guide).

8.7 Test for a significant difference in mean total heart weight between the diseased and normal groups.

8.8 Test for a significant difference in mean body weight between the diseased and normal groups.

Psychiatry, Renal Disease
Severe anxiety often occurs in patients who must undergo chronic hemodialysis. A set of progressive relaxation exercises was shown on videotape to a group of 38 experimental subjects, while a set of neutral videotapes was shown to a control group of 23 patients who were also on chronic hemodialysis [2]. The results of a psychiatric questionnaire (the State-Trait Anxiety Inventory) are presented in Table 8.1.

Table 8.1 Pretest and Posttest State-Trait Anxiety means and standard deviations for the experimental and control groups of hemodialysis patients

	Pretest			Posttest		
	Mean	sd	*n*	Mean	sd	*n*
Experimental	37.51	10.66	38	33.42	10.18	38
Control	36.42	8.59	23	39.71	9.16	23

Note: A lower score on the test corresponds to less anxiety.

Source: Reprinted with permission of the *Journal of Chronic Diseases*, 1982; **35**(10); 797–802.

8.9 Perform a statistical test to compare the experimental and control groups' pretest scores.

8.10 Perform a statistical test to compare the experimental and control groups' posttest scores.

Hypertension
Blood-pressure measurements taken on the left and right arms of a person are assumed to be comparable. To test this assumption, 10 volunteers are obtained and systolic blood-pressure readings are taken simultaneously on both arms by two different observers, Ms. Jones for the left arm and Mr. Smith for the right arm. The data are given in Table 8.2.

Table 8.2 Effect of arm on level of blood pressure (mm Hg)

Patient	Left arm	Right arm
1	130	126
2	120	124
3	135	127
4	100	95
5	98	102
6	110	109
7	123	124
8	136	132
9	140	137
10	155	156

8.11 Assuming that the two observers are comparable, test whether or not the two arms give comparable readings.

8.12 Suppose we do not assume that the two observers are comparable. Can the experiment, as it is defined, detect differences between the two arms? If not, can you suggest an alternative experimental design so as to achieve this aim?

Ophthalmology
A topic of current interest in ophthalmology is whether or not spherical refraction is different between the left and right eyes. For this purpose refraction is measured in both eyes of 17 people. The data are given in Table 8.3.

Table 8.3 Spherical refraction in the right and left eye

(*i*) Person	(x_i) Spherical refraction OD (right eye) (diopters)	(y_i) Spherical refraction OS (left eye) (diopters)	($d_i = x_i - y_i$) Difference (OD – OS)	(*i*) Person	(x_i) Spherical refraction OD	(y_i) Spherical refraction OS	($d_i = x_i - y_i$) Difference (OD – OS)
1	+1.75	+2.00	–0.25	9	0	0.50	–0.50
2	–4.00	–4.00	0	10	–1.00	–1.25	+0.25
3	–1.25	–1.00	–0.25	11	+0.50	–1.75	+2.25
4	+1.00	+1.00	0	12	–8.50	–5.00	–3.50
5	–1.00	–1.00	0	13	+0.50	+0.50	0
6	–0.75	+0.25	–1.00	14	–5.25	–4.75	–0.50
7	–2.25	–2.25	0	15	–2.25	–2.50	+0.25
8	+0.25	+0.25	0	16	–6.50	–6.25	–0.25
				17	+1.75	+1.75	0

8.13 Is a one-sample or two- sample test needed here?

8.14 Is a one-sided or two-sided test needed here?

8.15 Which of the following test procedures is appropriate to use on these data? (More than one may be necessary.)
a. Paired *t* test
b. Two-sample *t* test for independent samples with equal variances
c. Two-sample *t* test for independent samples with unequal variances
d. One-sample *t* test

8.16 Carry out the hypothesis test(s) in Problem 8.15 and report a *p*-value.

8.17 Estimate a 90% CI for the mean difference in spherical refraction between the two eyes.

Cardiovascular Disease
A study was performed in 1976 to relate the use of oral contraceptives to the levels of various lipid fractions in a group of 163 nonpregnant, premenopausal women ages 21–39. The mean serum cholesterol among 66 current users of oral contraceptives was 201 ± 37 (mg/dL) (mean ± sd), whereas for 97 nonusers it was 193 ± 37 (mg/dL).

8.18 Test for significant differences in mean cholesterol levels between the two groups.

8.19 Report a *p*-value based on your hypothesis test in Problem 8.18.

8.20 Derive a 95% CI for the true mean difference in cholesterol levels between the groups.

8.21 Suppose the two-tailed *p*-value in Problem 8.19 $= .03$ and the two-sided 95% CI in Problem 8.20 $= (-0.6, 7.3)$. Do these two results contradict each other? Why or why not? (*Note*: These values are not necessarily the actual results in Problems 8.19 and 8.20.)

Hypertension
A 1982 study by the Lipid Research CIinics looked at the relationship between alcohol consumption and level of systolic blood pressure in women not using oral contraceptives [3]. Alcohol consumption was categorized as follows: no alcohol use; ≤ 10 oz/week alcohol consumption; > 10 oz/week alcohol consumption. The results for women 30–39 years of age are given in Table 8.4.

Suppose we wish to compare the levels of systolic blood pressure of groups A and B and we have no prior information regarding which group has higher blood pressure.

Table 8.4 Relationship between systolic blood pressure and alcohol consumption in 30–39-year-old women not using oral contraceptives

	Systolic blood pressure (mm Hg)		
	Mean	sd	*n*
A. No alcohol use	110.5	13.3	357
B. ≤ 10 oz per week alcohol consumption	109.1	13.4	440
C. > 10 oz per week alcohol consumption	114.5	14.9	23

Source: Reprinted with permission of the *Journal of Chronic Diseases*, **35**(4), 251–257.

8.22 Should a one-sample or two-sample test be used here?

8.23 Should a one-sided or two-sided test be used here?

8.24 Which test procedure(s) should be used to test the preceding hypotheses?

8.25 Carry out the test in Problem 8.24 and report a *p*-value.

8.26 Compute a 95% CI for the mean difference in blood pressure between the two groups.

8.27 Answer Problem 8.24 for a comparison of groups A and C.

8.28 Answer Problem 8.25 for a comparison of groups A and C.

8.29 Answer Problem 8.26 for a comparison of groups A and C.

Pulmonary Disease

Forced expiratory volume (FEV) is a standard measure of pulmonary function representing the volume of air expelled in 1 second. Suppose we enroll 10 nonsmoking males age 35–39, heights 68–72 inches in a longitudinal study and measure their FEV (L) initially (year 0) and 2 years later (year 2). The data in Table 8.5 are obtained.

Table 8.5 Pulmonary function in nonsmokers at two points in time

Person	FEV year 0 (*L*)	FEV year 2 (*L*)	Person	FEV year 0 (*L*)	FEV year 2 (*L*)
1	3.22	2.95	6	3.25	3.20
2	4.06	3.75	7	4.20	3.90
3	3.85	4.00	8	3.05	2.76
4	3.50	3.42	9	2.86	2.75
5	2.80	2.77	10	3.50	3.32
			Mean	3.43	3.28
			sd	0.485	0.480

8.30 What are the appropriate null and alternative hypotheses in this case to test if mean pulmonary function has decreased over 2 years?

8.31 In words, what is the meaning of a type I and a type II error here?

8.32 Carry out the test in Problem 8.30. What are your conclusions?

Another aspect of the preceding study involves looking at the effect of smoking on baseline pulmonary function and on change in pulmonary function over time. We must be careful, since FEV depends on many factors, particularly age and height. Suppose we have a comparable group of 15 men in the same age and height group as in Table 8.5 who are smokers, and we measure their FEV at year 0 and year 2. The data are given in Table 8.6.

Table 8.6 Pulmonary function in smokers at two points in time

Person	FEV year 0 (*L*)	FEV year 2 (*L*)	Person	FEV year 0 (*L*)	FEV year 2 (*L*)
1	2.85	2.88	9	2.76	3.02
2	3.32	3.40	10	3.00	3.08
3	3.01	3.02	11	3.26	3.00
4	2.95	2.84	12	2.84	3.40
5	2.78	2.75	13	2.50	2.59
6	2.86	3.20	14	3.59	3.29
7	2.78	2.96	15	3.30	3.32
8	2.90	2.74			
			Mean	2.98	3.03
			sd	0.279	0.250

8.33 What are the appropriate null and alternative hypotheses to compare the FEV of smokers and nonsmokers at baseline?

8.34 Carry out the procedure(s) necessary to conduct the test in Problem 8.33.

8.35 Suggest a procedure for testing whether or not the change in pulmonary function over 2 years is the same in the two groups.

Hypertension

A study of the relationship between salt intake and blood pressure of infants is in the planning stages. A pilot study is done, comparing five 1-year-old infants on a high-salt diet with five 1-year-old infants on a low-salt diet. The results are given in Table 8.7.

Table 8.7 Relationship between salt intake and level of systolic blood pressure (SBP)

	High-salt diet	Low-salt diet
Mean SBP	90.8	87.2
sd SBP	10.3	9.2
n	5	5

8.36 If the means and standard deviations in Table 8.7 are considered to be true population parameters, then, using a one-sided test with significance level = .05 , how many infants are needed in each group to have an 80% chance of detecting a significant difference?

8.37 Suppose it is easier to recruit high-salt-diet infants and the investigators decide to enroll twice as many high-salt-diet infants as low-salt-diet infants. How many infants are needed in each group to have a 80% chance of detecting a significant difference using a one-sided test with an α level of .05?

8.38 Suppose the budget will only allow for recruiting 50 high-salt-diet and 50 low-salt-diet infants into the study. How much power would such a study have of detecting a significant difference using a one-sided test with significance level = .05 if the true difference between the groups is 5 mm Hg?

8.39 Answer Problem 8.38 for a true difference of 2 mm Hg.

8.40 Answer Problem 8.38 if a two-sided test is used and the true difference is 5 mm Hg.

8.41 Answer Problem 8.40 if the true difference is 2 mm Hg.

Hypertension

The effect of sodium restriction on blood pressure remains a controversial subject. To test this hypothesis a group of 83 individuals participated in a study of restricted sodium intake (≤ 75 mEq/24 hrs) for a period of 12 weeks. The effect on diastolic blood pressure (DBP) is reported in Table 8.8 [4].

Table 8.8 Effect of sodium restriction on diastolic blood pressure

	Baseline period			Diet period			Change from control (diet-baseline)		
	Mean	sd	*n*	Mean	sd	*n*	Mean	sd	*n*
Age < 40	69.7	8.5	61	69.1	8.5	61	–0.7	6.2	61
Age ≥ 40	77.0	8.0	22	71.9	7.5	22	–5.0	4.7	22
Total							–1.8	6.1	83

8.42 What is the appropriate procedure to test for whether sodium restriction has had an impact on mean DBP?

8.43 Implement the procedure in Problem 8.42 for people age < 40 using the critical-value method.

One of the interesting findings is the difference in response to dietary therapy between people in the two age groups.

8.44 Test the hypothesis that the response to sodium restriction is different in the two groups and report a *p*-value.

8.45 Obtain a 95% CI for the response to dietary therapy in each age group separately.

8.46 Obtain a 95% CI for the difference in response between the 2 age groups.

Suppose the results of this study are to be used to plan a larger study on the effects of sodium restriction on DBP.

8.47 How many subjects need to be enrolled in the larger study to test if sodium restriction results in lower DBP if the mean and sd of decline in DBP over the total group of 83 subjects are used for planning purposes and a 90% chance of detecting a significant difference using a two-sided test with a 5% level of significance is desired?

8.48 Suppose 200 patients are enrolled in the larger study. How much power would the larger study have if a two-sided test with significance level = .05 is used?

Environmental Health

A study was conducted relating lead level in umbilical cord blood and cognitive development [5]. Three groups were identified at birth (high/medium/low) cord blood lead. One issue is the consistency of the differences in blood-lead levels over time between these three groups. The data in Table 8.9 were presented.

Table 8.9 Change in blood-lead level in the first two years of life by cord blood-lead group

Cord blood-lead group		Blood-lead level (mean ± sd)				
		Birth	6 months	12 months	18 months	24 months
Low	Mean ± sd	1.8 ± 0.6	4.6 ± 3.9	5.8 ± 5.1	6.7 ± 5.5	5.4 ± 4.8
	n	85	70	69	65	61
Medium	Mean ± sd	6.5 ± 0.3	7.0 ± 7.8	8.5 ± 7.6	8.3 ± 5.8	7.2 ± 5.0
	n	88	70	70	65	63
High	Mean ± sd	14.6 ± 3.0	7.0 ± 8.7	8.8 ± 6.4	7.6 ± 5.8	7.7 ± 8.5
	n	76	61	60	57	58

8.49 What test procedure can be used to test if there are differences between the observed blood-lead levels at 24 months between the low and high groups (as defined at baseline)?

8.50 Implement the test procedure in Problem 8.49 and report a *p*-value.

8.51 Provide a 95% CI for the difference in mean blood lead levels between the low and high groups at 24 months.

Ophthalmology

Retinitis pigmentosa (RP) is the name given to a family of inherited retinal degenerative diseases that may be transmitted through various modes of inheritance. The most common features include a history of night blindness, loss of visual field, and pigment clumping in the retina. Several reports of lipid abnormalities have been reported in RP patients. However, a consistent trend as regards either excesses or deficiencies in lipid levels has not been apparent. In one study, fatty-acid levels were measured in a group of RP patients and normal controls [6]. The data in Table 8.10 were reported on one particular fatty acid (docosahexaenoic acid) (labeled 22:6w3) in individuals with dominant disease and normal controls.

Table 8.10 Mean levels of plasma 22:6w3 (adjusted for age)(units are nmol/mL plasma)

	Mean	sd	*n*
Dominant affected individuals	34.8	20.8	36
Normal controls	47.8	30.3	68

8.52 What is an appropriate procedure to test if the mean level of 22:6w3 differs between dominant affected individuals and normal controls? State the hypotheses being tested. Is a one-sided or a two-sided test appropriate here?

8.53 Perform the hypothesis test in Problem 8.52 and report a *p*-value.

8.54 Provide a 95% CI for the difference in means between the two groups. What does it mean in words?

Neurology

An article published in 1986 describes physical, social, and psychological problems in patients with multiple sclerosis [7]. Patients were classified as to mild, moderate, and severe disease and were graded on physical health using the McMaster physical health index, and mental health using the Rand mental health index scale. The data are shown in Table 8.11.

Table 8.11 Physical and mental health indices for patients with multiple sclerosis

	Mild			Severe		
	Mean	sd	*n*	Mean	sd	*n*
McMaster physical health index	0.76	0.24	65	0.23	0.12	82
Rand mental health index	157.3	26.2	65	149.1	34.4	82

8.55 What test procedure should be used to test if there are differences in physical health between the two groups?

8.56 Perform the test in Problem 8.55 with the critical-value method using a two-sided test with an a level of .05.

8.57 What test procedure should be used to test if there are differences in mental health between the two groups?

8.58 Perform the test in Problem 8.57 with the critical-value method using an α level of .05.

8.59 What is the most accurate *p*-value corresponding to your answer to Problem 8.58?

Hypertension

As one component of a study of blood pressure (BP) response to nonpharmacologic interventions, subjects with high normal BP (DBP of 80-89) who were moderately overweight were randomized to either an active weight-loss intervention or a control group. Blood-pressure measurements were taken both at baseline and after 18 months of intervention and follow-up. Using change in blood pressure over this period as the outcome, the data in Table 8.12 were obtained for diastolic blood pressure (DBP).

8.60 Assuming that the variances are equal in the two groups, test whether there was a significant difference in mean DBP change between the weight-loss versus the control group.

Table 8.12 Comparison of change in DBP between the weight loss versus the control group

	Change in DBP (mm Hg)		
Treatment Group	*n*	Mean	sd
Weight loss	308	–6.16	5.88
Control	246	–3.91	6.12

Researchers were also interested in the amount of weight loss attributable to the intervention. Mean weight change for each group is given in Table 8.13.

Table 8.13 Comparison of change in weight between the weight-loss group versus the control group

	Change in weight* (kg)		
Treatment group	*n*	Mean	sd
Weight loss	293	–3.83	6.12
Control	235	0.07	4.01

* 18 month weight-baseline weight

8.61 Test whether the variances are the same in the weight-loss and control groups.

8.62 Test whether the intervention was effective in reducing weight in the weight-loss versus the control group.

Cardiology

A study is planned of the effect of drug therapy with cholesterol-lowering drugs on the coronary arteries of patients with severe angina. A pilot study was performed whereby each patient received the placebo treatment and had an angiogram at baseline and again three years later at follow-up. The average diameter of three coronary arteries was measured at baseline and follow-up. The results are given in Table 8.14.

Table 8.14 Change in diameter of coronary arteries in patients with severe angina

Mean change[a]	–0.049
sd	0.237
n	8

[a] Follow-up - baseline.

Suppose we wish to test the hypothesis that there has been significant change over 3 years in the placebo group.

8.63 What test should be used to test this hypothesis?

8.64 Implement the test in Problem 8.63 and report a *p*-value.

8.65 Suppose it is assumed that the mean change over 3 years would be 0 for an active drug group and –0.049 for a placebo group. We intend to randomize subjects to placebo and active drug and compare the mean change in the two groups, How much power would the proposed study have if 500 subjects in each of the drug and placebo groups are available to participate in the study, a two-sided test is used with $\alpha = .05$ and the variances in each group are assumed to be the same?

Cardiovascular Disease, Pediatrics

A family-based behavior-change program was used to modify cardiovascular risk factors among teenage children of patients with ischemic heart disease. The mean baseline level and change in HDL-cholesterol level after a 6-month period in the program is given in Table 8.15.

Table 8.15 Mean baseline level and change in HDL-cholesterol level (mmol/L) among teenage children over a 6-month period

	Mean	sd	*n*
Baseline	1.20	0.32	44
6–months	1.12	0.35	44
Difference (6 months-baseline)	–0.08	0.23	44

8.66 What test procedure can be used to test if there has been a change in mean HDL-cholesterol levels among teenage children who undergo such a program?

8.67 Implement the test procedure mentioned in Problem 8.66 and report a *p*-value (two-sided) as precisely as possible using the appropriate tables.

8.68 Provide a 95% CI for the true mean change over 6 months among teenage children exposed to the program.

8.69 Suppose the results in Problem 8.67 are statistically significant (they may or may not be). Does this necessarily mean that the education program per se is the reason why the HDL-cholesterol levels have changed? If not, is there some way to change the design to allow us to be more confident about the specific effects of the education program?

Endocrinology
A study was performed comparing the rate of bone formation between black and white adults. The data in Table 8.16 were presented [8].

Table 8.16 Comparison of bone-formation rate between black versus white adults

	Blacks ($n = 12$) Mean ± se	Whites ($n = 13$) Mean ± se
Bone-formation rate	0.033 ± 0.007	0.095 ± 0.012

8.70 What method can be used to compare mean bone-formation rate between blacks and whites?

8.71 Implement the method in Problem 8.70 and report a *p*-value.

8.72 Obtain a 95% CI for the mean difference in bone-formation rate between the two groups.

Health Promotion
One goal of an exercise training program is to improve the physical fitness of individuals as measured by their resting heart rate. Suppose it is known that the mean resting heart rate of a group of sedentary individuals is 85 beats/min with a sd of 10 beats/min. A clinical trial is proposed where half the subjects will be randomized to a low intensity exercise program (group A), while the other half will get no exercise training (group B). It is expected that after 6 months, the mean resting heart rate of group A will decline by 3 beats/min while group B will show no mean change in resting heart rate. Assume that the correlation coefficient between resting heart rate measured 6 months apart in the same individual is .8.

8.73 How many subjects are needed in each group to achieve an 80% power if a two-sided test is used with significance level of .05?

8.74 Suppose that 50 subjects are randomized to each group. How much power will the study have if a 2-sided test is used with a significance level of .05?

SOLUTIONS

8.1 Use the paired *t* test because each person is used as his or her own control.

8.2 We have the test statistic

$$t = \frac{\bar{d}}{\frac{s_d}{\sqrt{n}}} = \frac{0.02}{0.04} = 0.50 \sim t_{89} \text{ under } H_0.$$

Since $0.50 < t_{120,\ .975} = 1.980 < t_{89,\ .975}$ it follows that H_0, that there is no mean change in visual-field loss over 1 year, is accepted. Furthermore, since

$$0.50 < t_{120,\ .75} = 0.677 < t_{89,\ .75}$$

it follows that $p > 2 \times (1 - .75) = .50$.

8.3 We have the test statistic

$$t = \frac{0.08}{0.05} = 1.60 \sim t_{89} \text{ under } H_0.$$

Since $1.60 < 1.980$, H_0, that there is no mean change in visual-field loss over 2 years, is accepted. Furthermore, if there were 60 *df*, then since

$$t_{60,\ .9} = 1.296 < 1.60 < t_{60,\ .95} = 1.671,$$

it follows that $2 \times (1 - .95) < p < 2 \times (1 - .90)$, or $.10 < p < .20$. Similarly, if there were 120 *df*, then since $t_{120,\ .9} = 1.289 < 1.60 < t_{120,\ .95} = 1.658$, it would also follow that $.10 < p < .20$. Finally, since there are 89 *df* and we reached the same decision with 60 or 120 *df*, then it follows that $.10 < p < .20$.

8.4 We have the test statistic $t = 0.14/0.07 = 2.000$. The critical value $= t_{89,\ .975} < t_{60,\ .975} = 2.000$. Since $t >$ critical value, H_0 is rejected at the 5% level. Since

$$2.000 < 2.358 = t_{120,\ .99} < t_{89,\ .99},$$

it follows that $p > 2 \times (1 - .99) = .02$. Therefore, $.02 < p < .05$. The exact *p*-value, obtained by computer, is $p = 2 \times Pr(t_{89} > 2.000) = .049$.

8.5 $F_{24,\ 30,\ .95} = 1.89$

8.6 We test the hypothesis H_0: $\sigma_1^2 = \sigma_2^2$ versus H_1: $\sigma_1^2 \neq \sigma_2^2$. We compute the *F* statistic

$$F = \frac{s_1^2}{s_2^2} = \frac{13.7^2}{7.2^2} = 3.62 \sim F_{8,\ 14}$$

under H_0. Since $F_{8,\ 14,\ .975} = 3.29$, $F_{8,\ 14,\ .99} = 4.14$ and $3.29 < 3.62 < 4.14$, it follows that

$$(1-.99)\times 2 < p < (1-.975)\times 2 \text{ or } .02 < p < .05$$

and the variances are significantly different.

8.7 The basic statistics for total heart weight (THW) and body weight (BW) in the two groups are given as follows:

Basic statistics for total heart weight and body weight in diseased and normal males

	Left-heart disease			Normals		
	Mean	sd	n	Mean	sd	n
THW (g)	450.00	139.34	11	317.00	47.09	10
BW (kg)	55.61	11.55	11	56.23	11.54	10

THW Perform the F test for the equality of two variances as follows:

$$F = \frac{s_1^2}{s_2^2} = \left(\frac{139.34}{47.09}\right)^2 = 8.75 \sim F_{10,\,9} \text{ under } H_0$$

We have

$$F_{10,\,9,\,.995} < F_{8,\,9,\,.995} = 6.69 < F$$

Thus, the p-value for F is $< 2(.005) = .01$, and the variances are significantly different. The t test with unequal variances must be used. We have the following test statistic:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{450.0 - 317.0}{\sqrt{\frac{139.34^2}{11} + \frac{47.09^2}{10}}} = \frac{133.0}{44.57} = 2.98$$

The appropriate df (d') must be computed.

$$d' = \frac{\left(\frac{139.34^2}{11} + \frac{47.09^2}{10}\right)^2}{\frac{(139.34^2/11)^2}{10} + \frac{(47.09^2/10)^2}{9}} = \frac{3{,}947{,}286}{316{,}988} = 12.45$$

Thus, there are 12 df. Since $|t| > t_{12,\,.975} = 2.179$, it follows that H_0 is rejected at the 5% level.

8.8 **BW** Again perform the F test for the equality of two variances.

$$F = \frac{s_1^2}{s_2^2} = \left(\frac{11.55}{11.54}\right)^2 = 1.002 \sim F_{10,\,9}$$

which is clearly not statistically significant. The t test with equal variances is therefore used. We have

$$s^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2} = \frac{10(11.55)^2 + 9(11.54)^2}{19} = 133.30$$

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{55.61 - 56.23}{\sqrt{133.30\left(\frac{1}{11} + \frac{1}{10}\right)}} = \frac{-0.62}{5.045} = -0.12 \sim t_{19}$$

which is not statistically significant at the 5% level. Thus, there is no significant difference in mean body weight between the two groups.

8.9 We first compare the variances of the two groups to decide which t test to use. We have the test statistic $F = s_1^2/s_2^2 = 10.66^2/8.59^2 = 1.54 \sim F_{37,\,22}$ under H_0. We note that $F_{37,\,22,\,.975} > F_{\infty,\,30,\,0.975} = 1.79 > F$. Thus, we accept H_0 and conclude that there is no significant difference between the two variances. Therefore, we use a two-sample t test with equal variances. We have the pooled-variance estimate

$$s^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2} = \frac{37(10.66)^2 + 22(8.59)^2}{59} = 98.78$$

We have the test statistic

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{37.51 - 36.42}{\sqrt{98.78\left(\frac{1}{38} + \frac{1}{23}\right)}} = \frac{1.09}{2.626} = 0.42 \sim t_{59} \text{ under } H_0$$

Since $t_{59,\,.975} > t_{60,\,.975} = 2.000 > 0.42$, we accept H_0 at the 5% level.

8.10 We again perform the variance-ratio test. We have $F = s_1^2/s_2^2 = 10.18^2/9.16^2 = 1.24 \sim F_{37,\,22}$ under H_0. Since

$$F_{37,\,22,\,.975} > F_{\infty,\,30,\,.975} = 1.79 > F,$$

we accept H_0 that the variances are the same and use the two-sample t test with equal variances. The pooled-variance estimate is given by

$$s^2 = \frac{37(10.18)^2 + 22(9.16)^2}{59} = 96.28$$

The test statistic for the t test with equal variances is obtained as follows:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{33.42 - 39.71}{\sqrt{96.28\left(\frac{1}{38} + \frac{1}{23}\right)}}$$
$$= \frac{-6.29}{2.592} = -2.43 \sim t_{59} \text{ under } H_0$$

Since $|t| > t_{40,\ .975} = 2.021 > t_{59,\ .975}$, we reject H_0 at the 5% level. Furthermore, since

$$|t| > t_{40,\ .99} = 2.423 > t_{59,\ .99},$$

we have that $p < .02$, while since

$$|t| < t_{60,\ .995} = 2.660 < t_{59,\ .995},$$

we have that $p > .01$. Thus, $.01 < p < .02$.

8.11 This is a classic situation for the use of a paired *t* test where each person is his or her own control. Thus, in this case we test the hypothesis $H_0: \mu_{1i} = \mu_{2i}$ versus

$$H_1: \mu_{1i} - \mu_{2i} = \Delta \neq 0$$

where μ_{1i} = mean blood pressure for the left arm of the *i*th person and μ_{2i} = mean blood pressure for the right arm of the *i*th person. Therefore, we do not assume that each person has the same mean blood pressure for a particular arm. We perform this test by computing the ten differences $d_i = x_{1i} - x_{2i}$ as follows:

i	d_i	i	d_i
1	4	6	1
2	–4	7	–1
3	8	8	4
4	5	9	3
5	–4	10	–1

We have $\bar{d} = 1.50$, $s_d = 3.979$, $s_d/\sqrt{10} = 1.258$. Thus, we compute the paired *t* statistic

$$t = \frac{\bar{d}}{\frac{s_d}{\sqrt{n}}} = \frac{1.50}{1.258} = 1.19$$

This should follow a *t* distribution with 9 *df* under H_0. Since $t_{9,\ .975} = 2.262$, it follows that $p > .05$ and we conclude that mean blood-pressure readings taken on the left and right arms are not significantly different from each other. Indeed, a 95% CI for $\Delta = \mu_{1i} - \mu_{2i}$ is given by

$$\bar{d} \pm 2.262\left(\frac{s_d}{\sqrt{n}}\right) = 1.50 \pm 2.262(1.258)$$
$$= (-1.35,\ 4.35)$$

Thus, we can be reasonably certain that the true mean difference, if any, between the two arms is less than 5 mm Hg.

8.12 If the 2 observers are not comparable, then we cannot separate the differences between the 2 arms from the differences between the 2 observers. We would require that a second measurement be obtained from each person, with Mr. Smith observing the left arm and Ms. Jones observing the right arm. We could then compute

$$d = 0.5 \times [(\text{BP left arm}_{\text{Jones}} - \text{BP right arm}_{\text{Jones}}) + (\text{BP left arm}_{\text{Smith}} - \text{BP right arm}_{\text{Smith}})]$$

to obtain a difference between arms that is not influenced by observers. We could analyze the 10 *d*'s again using the paired *t* test.

8.13 A two-sample test is needed here because we are comparing refractive error between the right and left eyes.

8.14 A two-sided test is needed here because we have no prior idea as to whether the right or left eye has the larger mean refraction.

8.15 A paired *t* test is appropriate here because we are comparing two paired samples consisting of the right and left eyes from the same group of people.

8.16 We first compute the following sample statistics:

$$\bar{d} = \frac{\sum d_i}{17} = \frac{-3.50}{17} = -0.206$$
$$s_d^2 = \frac{\sum (d_i - \bar{d})^2}{16} = \frac{18.404}{16} = 1.150$$

We wish to test the hypothesis $H_0: \mu_d = 0$ versus $H_1: \mu_d \neq 0$, where μ_d = underlying mean difference in spherical refractive error between right and left eyes. We have the test statistic

$$t = \frac{\bar{d}}{\frac{s_d}{\sqrt{n}}} = \frac{-0.206}{\sqrt{\frac{1.150}{17}}}$$
$$= \frac{-0.206}{0.260} = -0.79 \sim t_{16} \text{ under } H_0$$

Since $t_{16,\ .975} = 2.120 > 0.79$, it follows that $p > .05$ and there is no significant difference in mean spherical refractive error between the right and left eyes.

8.17 A two-sided 90% CI for the mean difference in spherical refractive error between right and left eyes is given by

$$\bar{d} \pm t_{n-1,\ .95}\frac{s_d}{\sqrt{n}} = -0.206 \pm t_{16,\ 0.95}(0.260)$$
$$= -0.206 \pm 1.746(0.260)$$
$$= -0.206 \pm 0.454 = (-0.660,\ 0.248)$$

8.18 Clearly, there are no significant differences between the variances, so we use the two-sample t test with equal variances. We use the test statistic

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim t_{n_1+n_2-2} \text{ under } H_0$$

where s^2 is the pooled-variance estimate over the 2 samples. In this case, since $s_1^2 = s_2^2$, we have that

$$s^2 = s_1^2 = s_2^2 = 37^2 .$$

Thus,

$$t = \frac{201 - 193}{\sqrt{37^2\left(\frac{1}{66} + \frac{1}{97}\right)}}$$
$$= \frac{8}{5.904} = 1.36 \sim t_{161} \text{ under } H_0$$

Since $t < t_{\infty,\ 0.975} = 1.960 < t_{161,\ .975}$, we accept H_0 at the 5% level.

8.19 Since $t_{120,\ .9} = 1.289$, $t_{120,\ .95} = 1.658$ and

$$1.289 < 1.36 < 1.658 ,$$

while $t_{\infty,\ .9} = 1.282$, $t_{\infty,\ .95} = 1.645$ and

$$1.282 < 1.36 < 1.645 ,$$

it follows that if we had either 120 or ∞ *df*, then $2 \times (1 - .95) < p < 2 \times (1 - .9)$ or $.1 < p < .2$. Thus, since $120 < 161 < \infty$, it follows that $.1 < p < .2$. The exact p-value obtained by computer is

$$p = 2 \times Pr(t_{161} > 1.36) = .18 .$$

8.20 The 95% CI is given by

$$\bar{x}_1 - \bar{x}_2 \pm t_{161,\ .975}\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} .$$

We use the TINV function of Excel to obtain

$$t_{161,\ .975} = 1.975 .$$

Thus, the 95% CI $= 8 \pm 1.975(5.904) = 8 \pm 11.66$ or $(-3.66,\ 19.66)$.

8.21 These results contradict each other. If the p-value $= .03 < .05$, then the 95% CI must not contain 0, which it does in this case.

8.22 A two-sample test should be used here, because we are comparing 2 finite samples.

8.23 A two-sided test should be used here, because we have no prior conception as to which group has the higher blood pressure.

8.24 The samples being considered are independent and thus we should perform a two-sample t test for independent samples with either equal or unequal variances. We need to perform the F test for the equality of 2 variances in order to decide which of these 2 tests to use. We proceed as follows with the F test. We have the test statistic

$$F = \frac{s_B^2}{s_A^2} = \left(\frac{13.4}{13.3}\right)^2 = 1.015 \sim F_{439,\ 356} .$$

Using a computer program, we find that

$$p = 2 \times Pr\left(F_{439,\ 356} > 1.015\right) = .88 .$$

Thus there is no significant difference between the variances, and we should use the two-sample t test for independent samples with equal variances.

8.25 We have the test statistic

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim t_{n_1+n_2-2} \text{ under } H_0$$

We must compute the pooled-variance estimate as follows:

$$s^2 = \frac{356(13.3)^2 + 439(13.4)^2}{795} = 178.36$$

Therefore,

$$t = \frac{110.5 - 109.1}{\sqrt{178.36\left(\frac{1}{357} + \frac{1}{440}\right)}}$$
$$= \frac{1.4}{0.951} = 1.47 \sim t_{795} \text{ under } H_0$$

Since $t_{120,\ .90} = 1.289$, $t_{120,\ .95} = 1.658$ and

$$1.289 < 1.47 < 1.658 ,$$

it follows that if we had 120 *df*, then $.1 < p < .2$. Similarly, $t_{\infty,\ .90} = 1.282$, $t_{\infty,\ .95} = 1.645$ and

$$1.282 < 1.47 < 1.645 ;$$

thus, if we had ∞ *df*, then $.1 < p < .2$. Therefore, since we obtain the same result for either 120 or ∞ *df* and $120 < 795 < \infty$, it follows that $.1 < p < .2$ and there is no significant difference between the means of the two groups. The exact p-value obtained by computer is

$$p = 2 \times Pr(t_{795} > 1.47) = .14 .$$

8.26 The 95% CI for the true mean difference $(\mu_1 - \mu_2)$ is given by

$$\bar{x}_1 - \bar{x}_2 \pm t_{795,\ .975}\sqrt{s^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)} = 1.4 \pm t_{795,\ .975}(0.951)$$

We use the computer to obtain $t_{795,\ .975} = 1.963$. Therefore, the 95% CI is

$$1.4 \pm 1.963(0.951) = 1.4 \pm 1.9 = (-0.5,\ 3.3).$$

8.27 The variance ratio $= (14.9/13.3)^2 = 1.26 \sim F_{22,\ 356}$ under H_0. Since

$$F < F_{24,\ \infty,\ .975} = 1.64 < F_{22,\ 356,\ .975},$$

it follows that $p > .05$. Therefore, the variances are not significantly different and we must use the two-sample *t* test with equal variances.

8.28 We have the test statistic

$$t = \frac{114.5 - 110.5}{\sqrt{s^2\left(\frac{1}{23}+\frac{1}{357}\right)}} = \frac{4.0}{\sqrt{s^2\left(\frac{1}{23}+\frac{1}{357}\right)}}$$

The pooled-variance estimate is given by

$$s^2 = \frac{356(13.3)^2 + 22(14.9)^2}{378} = 179.52$$

Therefore,

$$t = \frac{4.0}{\sqrt{179.52\left(\frac{1}{23}+\frac{1}{357}\right)}} = \frac{4.0}{2.88}$$
$$= 1.39 \sim t_{378} \text{ under } H_0$$

Since $1.289 = t_{120,\ .90} < t < 1.658 = t_{120,\ .95}$ and

$$1.282 = t_{\infty,\ .90} < t < 1.645 = t_{\infty,\ .95},$$

it follows that if we had either 120 *df* or ∞ *df* then $.10 < p < .20$. Therefore, because $120 < 378 < \infty$ we obtain $.10 < p < .20$.

8.29 The 95% CI is given by $4.0 \pm t_{378,\ .975}(2.88)$. Using the computer, we obtain $t_{378,\ .975} = 1.966$. Therefore, the 95% CI $= 4.0 \pm 1.966(2.88) = 4.0 \pm 5.7 = (-1.7,\ 9.7)$.

8.30 We must use a paired *t* test formulation here because the same people are being compared at two different points in time. Our hypotheses are $H_0:\ \mu_d = 0$ versus $H_1:\ \mu_d > 0$, where $\mu_d =$ underlying mean for (FEV year 0-FEV year 2). We perform a one-sided test because we expect FEV to decline over time.

8.31 A type I error is the probability that we detect a significant decline in FEV level over time given that a true decline does not actually exist. A type II error is the probability that we do not detect a significant decline in FEV level given that a true decline does exist.

8.32 We must compute the 10 paired differences in FEV as follows:

i	d_i	i	d_i
1	0.27	6	0.05
2	0.31	7	0.30
3	–0.15	8	0.29
4	0.08	9	0.11
5	0.03	10	0.18

We have that $\bar{d} = 0.147$, $s_d = 0.150$. The test statistic is given as follows:

$$t = \frac{\bar{d}}{\frac{s_d}{\sqrt{n}}} = \frac{0.147}{\frac{0.150}{\sqrt{10}}} = 3.09 \sim t_9 \text{ under } H_0$$

We find that $t_{9,\ .99} = 2.821$, $t_{9,\ .995} = 3.250$, and thus, $.005 < p < .01$ (one-sided test). Therefore, there is a significant decline in mean FEV level over time.

8.33 We will use a two-sample *t* test for independent samples. The appropriate hypotheses are $H_0:\ \mu_1 = \mu_2$ versus $H_1:\ \mu_1 \neq \mu_2$, where $\mu_1 =$ mean FEV at baseline for smokers and $\mu_2 =$ mean FEV at baseline for nonsmokers.

8.34 We first will test whether the variances in the two groups are significantly different, using the *F* test for the equality of two variances. We have the test statistic

$$F = \frac{s_1^2}{s_2^2} = \frac{0.485^2}{0.279^2} = 3.03 \sim F_{9,\ 14} \text{ under } H_0$$

Since $F < F_{12,\ 14,\ .975} = 3.05 < F_{9,\ 14,\ .975}$ we have that $p > .05$ and thus the variances are not significantly different. Therefore, we use a two-sample *t* test with equal variances. We have the pooled-variance estimate

$$s^2 = \frac{9s_1^2 + 14s_2^2}{23}$$
$$= \frac{9(0.485)^2 + 14(0.279)^2}{23} = 0.139$$

Thus, we have the test statistic

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{3.43 - 2.98}{\sqrt{0.139\left(\frac{1}{10} + \frac{1}{15}\right)}} = \frac{0.45}{0.152}$$
$$= 2.95 \sim t_{23} \text{ under } H_0$$

Since $t_{23,\ .995} = 2.807$, $t_{23,\ .9995} = 3.767$ and

$$2.807 < 2.95 < 3.767,$$

it follows that $.001 < p < .01$. Thus, there is a significant difference in mean FEV level between smokers and nonsmokers at baseline, with smokers having lower FEVs.

8.35 The simplest approach would be to take the difference between year 0 and year 2 FEV for each person in each group and test whether the mean differences are significantly different in the two groups using a two-sample *t* test for independent samples. However, it is well known that adults with high initial levels of pulmonary function tend to decrease more than those with low initial levels (this is known as the "regression to the mean" phenomenon). Thus, if the true rate of decline was the same in smokers and nonsmokers, the nonsmokers might appear to decline faster (as is the case in this data set) because they had higher initial FEVs.We would have to correct the change in FEV for initial level of FEV before looking at differences in change between the two groups.

8.36 We use the sample-size formula in Equation 8.26 (text, Chapter 8) changing $\alpha/2$ to α for a one-sided test. We have

$$n = \frac{(\sigma_1^2 + \sigma_2^2)(z_{1-\alpha} + z_{1-\beta})^2}{\Delta^2} = \text{sample size in each group}$$

In this problem,

$$\sigma_1 = 10.3,\ \sigma_2 = 9.2,\ \Delta = 90.8 - 87.2 = 3.6,$$
$$\alpha = .05,\ \beta = 1 - .80 = .20.$$

Thus, we have

$$n = \frac{(10.3^2 + 9.2^2)(z_{.95} + z_{.80})^2}{3.6^2}$$
$$= \frac{(106.09 + 84.64)(1.645 + 0.84)^2}{12.96} = 90.9$$

Therefore, we need to study 91 infants in each group in order to have an 80% probability of finding a significant difference using a one-sided significance test with $\alpha = .05$.

8.37 We refer to Equation 8.27 (text, Chapter 8) substituting α for $\alpha/2$ and setting $k = 0.5$. We have

$$n_1 = \frac{\left(\sigma_1^2 + \frac{\sigma_2^2}{k}\right)(z_{1-\alpha} + z_{1-\beta})^2}{\Delta^2}$$
$$= \frac{(10.3^2 + \frac{9.2^2}{0.5})(1.645 + 0.84)^2}{3.6^2}$$
$$= \frac{275.37(6.175)}{12.96} = 131.2 \text{ or } 132$$
$$= \text{number of high salt infants}$$
$$n_2 = 0.5 \times n_1 = 66 = \text{number of low salt infants}$$

8.38 We use the power formula given in Equation 8.28 (text, Chapter 8) substituting α for $\alpha/2$ because we are performing a one-sided test. We have

$$\text{Power} = \Phi\left(\frac{|\Delta|}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} - z_{1-\alpha}\right)$$

where

$$n_1 = 50,\ n_2 = 50,\ \Delta = 5,\ \sigma_1 = 10.3,$$
$$\sigma_2 = 9.2,\ \alpha = .05.$$

We have

$$\text{Power} = \Phi\left(\frac{5}{\sqrt{10.3^2/50 + 9.2^2/50}} - z_{.95}\right)$$
$$= \Phi\left(\frac{5}{1.953} - 1.645\right)$$
$$= \Phi(0.915) = .820$$

8.39 If $\Delta = 2$ mm Hg, then

$$\text{Power} = \Phi\left(\frac{2}{\sqrt{10.3^2/50 + 9.2^2/50}} - 1.645\right)$$
$$= \Phi(1.024 - 1.645)$$
$$= \Phi(-0.621) = 1 - \Phi(0.621)$$
$$= 1 - .7327 = .267$$

8.40 If $\Delta = 5$ mm Hg and a two-sided test is used, then from (Equation 8.28) (text, Chapter 8),

$$\text{Power} = \Phi\left(\frac{|\Delta|}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} - z_{1-\alpha/2}\right)$$
$$= \Phi\left(\frac{5}{1.953} - 1.96\right)$$
$$= \Phi(0.600) = .726$$

Note that the two-sided test here has less power (.726) than the one-sided test using the same parameters in Problem 8.38 (.820).

8.41 If $\Delta = 2$ mm Hg, then

$$\begin{aligned}\text{Power} &= \Phi\left(\frac{2}{1.953} - 1.96\right) = \Phi(1.024 - 1.96)\\ &= \Phi(-0.936) = 1 - \Phi(0.936)\\ &= 1 - .825 = .175\end{aligned}$$

8.42 Since each person is being used as his or her own control, we will use the paired *t* test to assess whether sodium restriction has had an impact on diastolic blood pressure.

8.43 We wish to test the hypothesis $H_0: \mu_d = 0$ versus $H_1: \mu_d \neq 0$. We use the test statistic

$$t = \frac{\bar{d}}{\frac{s_d}{\sqrt{n}}} \sim t_{n-1} \text{ under } H_0$$

If we perform the hypothesis test using an α level of .05, then we will reject H_0 if

$$t < t_{n-1,\ .025} \text{ or } t > t_{n-1,\ .975}.$$

In this case,

$$t = \frac{-0.7}{\frac{6.2}{\sqrt{61}}} = -0.882$$

The critical values are $t_{60,\ .025} = -2.000$ and

$$t_{60,\ .975} = 2.000.$$

Since $-2.000 < -0.882 < 2.000$, it follows that we accept H_0 at the 5% level and conclude that sodium restriction does not have a significant impact on mean DBP in the under-40 age group.

8.44 Since we are comparing two independent samples, we must use the two-sample *t* test for independent samples with either equal or unequal variances. To decide between these two tests, we first perform the *F* test for the equality of two variances. We have the test statistic

$$F = \frac{s_1^2}{s_2^2} = \frac{6.2^2}{4.7^2} = 1.74 \sim F_{60,\ 21}$$

under H_0. We refer to the *F* tables and note that

$$F < F_{\infty,\ 30,\ .975} = 1.79 < F_{60,\ 21,\ .975}.$$

It follows that $p > .05$. Thus, there is no significant difference between the variances and we use the *t* test with equal variances. We must first compute the pooled-variance estimate as follows:

$$\begin{aligned}s^2 &= \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = \frac{60(6.2)^2 + 21(4.7)^2}{81}\\ &= \frac{2770.29}{81} = 34.20\end{aligned}$$

We now have the test statistic

$$\begin{aligned}t &= \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{-0.7 + 5.0}{\sqrt{34.20\left(\frac{1}{61} + \frac{1}{22}\right)}}\\ &= \frac{4.3}{1.454} = 2.96 \sim t_{81} \text{ under } H_0\end{aligned}$$

To evaluate the *p*-value, we must compute

$$p = 2 \times Pr(t_{81} > 2.96).$$

Since $t_{60,\ .995} = 2.660$, $t_{60,\ .9995} = 3.460$, and

$$2.660 < 2.96 < 3.460,$$

it follows that if we had 60 *df* that

$$2 \times (1 - .9995) < p < 2 \times (1 - .995)$$

or $.001 < p < .01$. Similarly, since $t_{120,\ .995} = 2.617$, $t_{120,\ .9995} = 3.373$ and $2.617 < 2.96 < 3.373$, it follows that if we had 120 *df* that $.001 < p < .01$. Since $60<81<120$, we conclude that the response to sodium restriction is significantly different in the two age groups ($.001 < p < .01$).

8.45 A 95% CI for true change in mean DBP for the age group < 40 is given by

$$\begin{aligned}-0.7 \pm t_{60,\ .975}\frac{6.2}{\sqrt{61}} &= -0.7 \pm 2.0\frac{6.2}{\sqrt{61}}\\ &= -0.7 \pm 1.59 = (-2.3,\ 0.9).\end{aligned}$$

For the age group ≥ 40, we have the 95% CI

$$\begin{aligned}-5.0 \pm t_{21,\ .975}\frac{4.7}{\sqrt{22}} &= -5.0 \pm 2.080\frac{4.7}{\sqrt{22}}\\ &= -5.0 \pm 2.08 = (-7.1,\ -2.9).\end{aligned}$$

8.46 A 95% CI for the true difference in response between the two groups is given by $4.3 \pm t_{81,\ .975}(1.454)$. We find $t_{81,\ .975}$ using Excel and obtain $t_{81,\ .975} = 1.990$. Thus, the 95% CI =

$$4.3 \pm 1.990(1.454) = 4.3 \pm 2.89 = (1.4,\ 7.2).$$

8.47 We use the sample-size formula

$$n = \frac{\sigma^2\left(z_{1-\beta} + z_{1-\alpha/2}\right)^2}{\Delta^2}$$

In this case, $\Delta = -1.8$, $\sigma = 6.1$, $\alpha = .05$, $\beta = .1$. Therefore, we have that

$$n = \frac{6.1^2(z_{.90} + z_{.975})^2}{(-1.8)^2}$$
$$= \frac{6.1^2(1.28 + 1.96)^2}{(-1.8)^2} = \frac{390.62}{3.24} = 120.6$$

Therefore, we need to study 121 subjects to have a 90% chance of finding a significant difference using a two-sided test with $\alpha = .05$.

8.48 We use the following power formula:

$$\text{Power} = \Phi\left(\frac{\sqrt{n}|\Delta|}{\sigma} - z_{1-\alpha/2}\right)$$
$$= \Phi\left[\frac{\sqrt{200}(1.8)}{6.1} - 1.96\right] = \Phi(4.173 - 1.96)$$
$$= \Phi(2.213) = .987$$

8.49 We wish to test the hypothesis $H_0: \mu_1 = \mu_2$ versus $H_1: \mu_1 \neq \mu_2$. We will use the F test for the equality of two variances ($H_0: \sigma_1^2 = \sigma_2^2$ versus $H_1: \sigma_1^2 \neq \sigma_2^2$) to decide whether to use a two-sample t test with equal or unequal variances. We have the test statistic

$$F = \left(\frac{8.5}{4.8}\right)^2 = 3.14 \sim F_{57,\ 60} \text{ under } H_0.$$

Referring to Table 9 (Appendix, text), we note that

$$F > F_{24,\ 60,\ .975} = 1.88 > F_{57,\ 60,\ .975}.$$

Therefore, $p < .05$ and we must use the two-sample t test with unequal variances.

8.50 We have the test statistic

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{5.4 - 7.7}{\sqrt{\frac{4.8^2}{61} + \frac{8.5^2}{58}}}$$
$$= \frac{-2.3}{1.274} = -1.81 \sim t_{d''} \text{ under } H_0$$

We must interpolate to compute the appropriate *df*. We have that

$$d' = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}} = \frac{(0.378 + 1.246)^2}{\frac{0.378^2}{60} + \frac{1.246^2}{57}}$$
$$= \frac{2.635}{0.0296} = 89.0 \text{ or } d'' = 89 \text{ } df$$

Referring to the t table, we note that $t_{60,\ .95} = 1.671$, $t_{60,\ .975} = 2.000$, $t_{120,\ .95} = 1.658$, $t_{120,\ .975} = 1.980$. Since $1.671 < 1.81 < 2.000$ and $1.658 < 1.81 < 1.980$, we reach the same conclusion with either 60 or 120 *df*. Thus,

$$2 \times (1 - .975) < p < 2 \times (1 - .95)$$

or $.05 < p < .10$. Thus, there is not a significant difference in mean blood-lead level at 24 months between the two groups.

8.51 A 95% CI for the difference in mean blood-lead level is given by

$$\bar{x}_1 - \bar{x}_2 \pm t_{89,\ .975}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = 5.4 - 7.7 \pm t_{89,\ .975}(1.274)$$
$$= -2.3 \pm t_{89,\ .975}(1.274)$$

We find $t_{89,\ .975}$ using Excel and obtain $t_{89,\ .975} = 1.987$. Therefore, the 95% CI is given by $-2.3 \pm 1.987(1.274) = -2.3 \pm 2.53 = (-4.8,\ 0.2)$.

8.52 We first test the hypothesis $H_0: \sigma_1^2 = \sigma_2^2$ versus $H_1: \sigma_1^2 \neq \sigma_2^2$ to decide on the appropriate t test. We have the test statistic

$$F = \frac{s_1^2}{s_2^2} = \left(\frac{30.3}{20.8}\right)^2 = 2.12 \sim F_{67,\ 35} \text{ under } H_0.$$

To evaluate significance, we use the FDIST function of Excel and obtain the p-value

$$2 \times Pr\left(F_{67,\ 35} > 2.12\right) = .008.$$

Thus we must use the two-sample t test with unequal variances to test the hypothesis $H_0: \mu_1 = \mu_2,\ \sigma_1^2 \neq \sigma_2^2$ versus $H_1: \mu_1 \neq \mu_2,\ \sigma_1^2 \neq \sigma_2^2$. We will use a two-sided test, because we have no prior hypotheses as to the direction of any differences found.

8.53 We have the test statistic

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{34.8 - 47.8}{\sqrt{\frac{20.8^2}{36} + \frac{30.3^2}{68}}} = \frac{-13.0}{5.052} = -2.57 \sim t_{d''}$$

The *df* is obtained from the Satterthwaite approximation as follows:

$$d' = \frac{\left(\frac{20.8^2}{36} + \frac{30.3^2}{68}\right)^2}{\frac{(20.8^2/36)^2}{35} + \frac{(30.3^2/68)^2}{67}} = \frac{651.22}{6.85} = 95.1$$

Thus, $d'' = 95$ *df*. Since

$$t_{60,\ .99} = 2.390 < 2.57 < 2.660 = t_{60,\ .995}$$

and

$$t_{120,\ .99} = 2.358 < 2.57 < 2.617 = t_{120,\ .995},$$

it follows that if we had either 60 or 120 *df*, then $2 \times (1 - .995) < p < 2 \times (1 - .99)$ or $.01 < p < .02$. Since $60 < 95 < 120$, it follows that $.01 < p < .02$.

8.54 A 95% CI for the true difference in means is given by $-13.0 \pm t_{95,\ .975}(5.052)$. We use the TINV function of Excel to obtain $t_{95,\ .975} = 1.985$. Thus, the 95% CI is

$$-13.0 \pm 1.985(5.052) = -13.0 \pm 10.0 = (-23.0,\ -3.0).$$

It means that 95% of all possible 95% CIs that could be constructed from repeated random samples of 36 dominants and 68 normals will contain the true difference in means $\mu_1 - \mu_2$.

8.55 We use a two-sample *t* test with either equal or unequal variances. To decide on the appropriate procedure, we first perform the *F* test for the equality of two variances. We have the test statistic

$$F = \frac{s_1^2}{s_2^2} = \left(\frac{0.24}{0.12}\right)^2 = 4.00 \sim F_{64,\ 81} \text{ under } H_0.$$

Since $F_{64,\ 81,\ .975} < F_{24,\ 60,\ .975} = 1.88 < F$, it follows that $p < .05$ and the variances are significantly different. Therefore, we must use the two-sample *t* test with unequal variances.

8.56 We have the test statistic

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{0.76 - 0.23}{\sqrt{\frac{0.24^2}{65} + \frac{0.12^2}{82}}} = \frac{0.53}{0.033} = 16.3$$

To assess significance, we note that $d'' \geq 64$ and

$$t = 16.3 > t_{60,\ .975} = 2.000 > t_{64,\ .975}.$$

Therefore, it follows that $p < .05$ and there are significant differences in mean physical health scope since between the two groups.

8.57 We again perform the *F* test for the equality of two variances. We have the test statistic

$$F = \frac{s_1^2}{s_2^2} = \left(\frac{34.4}{26.2}\right)^2 = 1.72 \sim F_{81,\ 64} \text{ under } H_0.$$

To determine the significance of the results, we use the computer to obtain the *p*-value

$$2 \times Pr\left(F_{81,\ 64} > 1.72\right) = .012.$$

Thus, the variances are significantly different, and we must use the two-sample *t* test with unequal variances.

8.58 We have the test statistic

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{157.3 - 149.1}{\sqrt{\frac{26.2^2}{65} + \frac{34.4^2}{82}}} = \frac{8.2}{5.00} = 1.640$$

Since $t = 1.64 < t_{\infty,\ .975} = 1.96 < t_{d'',\ .975}$ (where d'' is the appropriate *df*), it follows that $p > .05$. Therefore, there is no significant difference in mean mental health score between the two groups.

8.59 To determine the *p*-value, we compute the *df* as follows:

$$d' = \frac{\left(\frac{26.2^2}{65} + \frac{34.4^2}{82}\right)^2}{\frac{(26.2^2/65)^2}{64} + \frac{(34.4^2/82)^2}{81}} = \frac{624.59}{4.31} = 144.8$$

Thus, $d'' = 144$. Since

$$t_{120,\ .9} = 1.289 < 1.64 < t_{120,\ .95} = 1.658$$

and $t_{\infty,\ .9} = 1.282 < 1.64 < t_{\infty,\ .95} = 1.645$, it follows that $2 \times (1 - .95) < p < 2 \times (1 - .9)$ or $.1 < p < .2$. The exact *p*-value obtained by computer is $Pr\left(t_{144} > 1.64\right) \times 2 = .103$.

8.60 We wish to test the hypothesis $H_0:\ \mu_1 = \mu_2,\ \sigma_1^2 = \sigma_2^2$ versus $H_1:\ \mu_1 \neq \mu_2,\ \sigma_1^2 = \sigma_2^2$. We have the test statistic

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim t_{n_1 + n_2 - 2} \text{ under } H_0.$$

The pooled-variance estimate is given by

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = \frac{307(5.88)^2 + 245(6.12)^2}{552} = 35.85$$

Thus,

$$t = \frac{-6.16 + 3.91}{\sqrt{35.85\left(\frac{1}{308} + \frac{1}{246}\right)}} = \frac{-2.25}{0.512}$$
$$= -4.39 \sim t_{552} \text{ under } H_0$$

Since $t_{552,\ .9995} < t_{120,\ .9995} = 3.373 < 4.39$, it follows that $p < .001$. Thus, there is a highly significant difference in mean DBP change between the weight-loss and control groups.

8.61 We wish to test the hypothesis $H_0:\ \sigma_1^2 = \sigma_2^2$ versus $H_1:\ \sigma_1^2 \neq \sigma_2^2$. We have the test statistic

$$F = \left(\frac{6.12}{4.01}\right)^2 = 2.33 \sim F_{292,\ 234} \text{ under } H_0.$$

Since

$$2.33 > F_{24,\ 120,\ .975} = 1.76 > F_{292,\ 234,\ .975}$$

it follows that $p < .05$ and there is a significant difference between the variances.

8.62 We wish to test the hypothesis $H_0:\ \mu_1 = \mu_2,\ \sigma_1^2 \neq \sigma_2^2$ versus $H_1:\ \mu_1 \neq \mu_2,\ \sigma_1^2 \neq \sigma_2^2$. We have the test statistic

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{-3.83 - 0.07}{\sqrt{\frac{6.12^2}{293} + \frac{4.01^2}{235}}}$$
$$= \frac{-3.90}{0.443} = -8.80 \sim t_{d''} \text{ under } H_0$$

The degrees of freedom will be at least as large as 234. Clearly, since $8.80 > t_{120,\ .9995} = 3.373 > t_{234,\ .9995}$, it follows that $p < .001$. Thus, there is a significantly greater mean weight loss in the weight-loss group versus the control group.

8.63 We wish to test the hypothesis $H_0:\ \mu_d = 0$ versus $H_1:\ \mu_d \neq 0$ where μ_d = true mean change in diameter of the average of three coronary arteries. We will use the paired *t* test to test this hypothesis.

8.64 We have the test statistic

$$t = \frac{\bar{d}}{\frac{s_d}{\sqrt{n}}} = \frac{-0.049}{\frac{0.237}{\sqrt{8}}} = \frac{-0.049}{0.084}$$
$$= -0.58 \sim t_7 \text{ under } H_0$$

Since $t_{7,\ .75} = 0.711$, and $|t| < 0.711$, it follows that $p > 2 \times (1 - .75) = .50$. Thus, there is no significant mean change in diameter in the placebo group in this study.

8.65 We use the following power formula for comparing two means:

$$\text{Power} = \Phi\left(-z_{1-\alpha/2} + \frac{|\Delta|}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}\right)$$

where $\alpha = .05$, $n_1 = n_2 = 500$, $\Delta = 0.049$,

$$\sigma_1 = \sigma_2 = 0.237.$$

Thus, the power is given by

$$\text{Power} = \Phi\left(-z_{.975} + \frac{0.049}{\sqrt{0.237^2/500 + 0.237^2/500}}\right)$$
$$= \Phi\left(-1.96 + \frac{0.049}{0.015}\right)$$
$$= \Phi(-1.96 + 3.269)$$
$$= \Phi(1.309) = .90$$

Thus, the study should have 90% power to detect this difference.

8.66 Since each person was used as his or her own control, we use the paired *t* test. The hypotheses to be tested are $H_0:\ \mu_d = 0$ versus $H_1:\ \mu_d \neq 0$, where μ_d = true mean change (6 months-baseline) in HDL cholesterol for teenage children exposed to this behavior-change program.

8.67 We use the test statistic

$$t = \frac{\bar{d}}{\frac{s_d}{\sqrt{n}}} = \frac{-0.08}{\frac{0.23}{\sqrt{44}}}$$
$$= \frac{-0.08}{0.0347} = -2.307 \sim t_{43} \text{ under } H_0$$

We note that $t_{40,\ .975} = 2.021$, $t_{40,\ .99} = 2.423$. Since $2.021 < |t| < 2.423$, it follows that if we had 40 *df*, then $1 - .99 < p/2 < 1 - .975$ or $.01 < p/2 < .025$ or

$$.02 < p < .05.$$

Similarly, since $t_{60,\ .975} = 2.000$, $t_{60,\ .99} = 2.390$ and $2.000 < |t| < 2.390$, it follows that if we had 60 *df*, then $.02 < p < .05$. Therefore, since we have 43 *df* and we reach the same conclusion with either 40 or 60 *df* and $40 < 43 < 60$, it follows that $.02 < p < .05$ and there has been a significant decline in mean HDL cholesterol over the 6-month period.

8.68 A 95% CI for the true mean change over 6 months is given by

$$\bar{d} \pm t_{n-1,\ .975}\frac{s_d}{\sqrt{n}} = -0.08 \pm t_{43,\ .975}(0.0347)$$

We use the computer to obtain $t_{43,\ .975}$, whereby $t_{43,\ .975} = 2.016$. Thus, the 95% CI

$$-0.08 \pm 2.016(0.0347) = -0.08 \pm 0.070 = (-0.15,\ -0.01)$$

8.69 Actually the changes in HDL-cholesterol are in an undesirable direction since low HDL levels are associated with higher coronary risk. We need to establish a control group that is also followed for 6 months and whose change in HDL cholesterol is assessed over a 6-month period. We could compare the mean change in the active treatment group with the mean change in the control group. This would allow us to be more confident of the specific effects of the education program.

8.70 We will assume that the bone-formation rate for blacks is normally distributed with mean μ_1 and variance σ_1^2, and the bone-formation rate for whites is normally distributed with mean μ_2 and variance σ_2^2. We wish to test the hypothesis $H_0: \mu_1 = \mu_2$ versus $H_1: \mu_1 \neq \mu_2$. We will use a two-sample t test with either equal or unequal variances. To decide on the appropriate t test, we will first perform the F test for the equality of 2 variances to test the hypotheses $H_0: \sigma_1^2 = \sigma_2^2$ versus $H_1: \sigma_1^2 \neq \sigma_2^2$. We have the test statistic $F = s_2^2/s_1^2 \sim F_{12,\ 11}$ under H_0. We have

$$s_1 = 0.007\sqrt{12} = 0.024,\ s_2 = 0.012\sqrt{13} = 0.043$$

and $F = (0.043/0.024)^2 = 3.18 \sim F_{12,\ 11}$ under H_0. Since $F < F_{12,\ 12,\ .975} = 3.28 < F_{12,\ 11,\ .975}$, it follows that

$$p > .05$$

and there is no significant difference between the variances. Thus, we will use the two-sample t test with equal variances.

8.71 We must compute the pooled-variance estimate:

$$s^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} = \frac{11(0.024)^2 + 12(0.043)^2}{23} = 0.00126$$

We have the test statistic

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \sim t_{23} \text{ under } H_0$$

$$= \frac{0.033 - 0.095}{\sqrt{0.00126\left(\frac{1}{12}+\frac{1}{13}\right)}} = \frac{-0.062}{0.014} = -4.37$$

Since $|t| = 4.37 > t_{23,\ .9995} = 3.767$, it follows that

$$p < 2(1-.9995) = .001.$$

Thus, there is a highly significant difference in the mean bone-formation rate between blacks and whites.

8.72 A 95% CI for $\mu_1 - \mu_2$ is given by

$$\bar{x}_1 - \bar{x}_2 \pm t_{23,\ .975}\sqrt{s^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)} = -0.062 \pm 2.069(0.014)$$

$$= -0.062 \pm 0.029$$

$$= (-0.091,\ -0.033)$$

8.73 We refer to Equation 8.30 (Chapter 8, text). We have

$$n = \frac{2\sigma_d^2\left(z_{1-\alpha/2} + z_{1-\beta}\right)^2}{\delta^2}$$

We have

$$\sigma_d^2 = \sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2$$

$$= 10^2 + 10^2 - 2(.8)(10)(10)$$

$$= 40$$

$$\alpha = .05,\ 1-\beta = .80,\ \delta = 3$$

Thus,

$$n = \frac{2(40)(z_{.975} + z_{.80})^2}{3^2} = \frac{80(1.96+0.84)^2}{9} = 69.7$$

Thus, we need to enroll 70 subjects per group to have 80% power.

8.74 We refer the the power formula in Equation 8.31 (Chapter 8, text). We have

$$\text{Power} = \Phi\left(-z_{1-\alpha/2} + \frac{\sqrt{n}\delta}{\sigma_d\sqrt{2}}\right)$$

$$= \Phi\left(-z_{.975} + \frac{\sqrt{50}(3)}{\sqrt{40}\sqrt{2}}\right)$$

$$= \Phi\left(-1.96 + \frac{21.21}{8.94}\right)$$

$$= \Phi(0.412) = .66$$

Thus, the study would have 66% power.

REFERENCES

[1] Berson, E. L., Sandberg, M. A., Rosner, B., Birch, D. G., & Hanson, A. H. (1985). Progression of retinitis pigmentosa over a three-year interval. *American Journal of Ophthalmology*, **99**, 246–251.

[2] Alarcon, R. D., Jenkins, C. S., Heestand, D. E., Scott, L. K., & Cantor, L. (1982). The effectiveness of progressive relaxation in chronic hemodialysis patients. *Journal of Chronic Diseases*, **35**(10), 797–802.

[3] Wallace, R. B., Barrett-Connor, E., Criqui, M., Wahl, P., Hoover, J., Hunninghake, D., & Heiss, G. (1982). Alteration in blood pressures associated with combined alcohol and oral contraceptive use—The Lipid Research Clinics prevalence study. *Journal of Chronic Diseases*, **35**(4), 251–257.

[4] Miller, J. Z., Weinberger, M. H., Daugherty, S. A., Fineberg, N. S., Christian, J. C., & Grim, C. E. (1987). Heterogeneity of blood pressure response to dietary sodium restriction in normotensive adults. *Journal of Chronic Disease*, **40**(3), 245–250.

[5] Bellinger, D., Leviton, A., Waternaux, C., Needleman, H., & Rabinowitz, M. (1987). Longitudinal analyses of prenatal and postnatal lead exposure and early cognitive development. *New England Journal of Medicine*, **316**(17), 1037–1043.

[6] Anderson, R. E., Maude, M. B., Lewis, R. A., Newsome, D. A., & Fishman, G. A. (1987). Abnormal plasma levels of polyunsaturated fatty acid in autosomal dominant retinitis pigmentosa. *Experimental Eye Research*, **44**, 155–159.

[7] Harper, A. C., Harper, D. A., Chambers, L. W., Cino, P. M., & Singer, J. (1986). An epidemiological description of physical, social and psychological problems in multiple sclerosis. *Journal of Chronic Disease*, **39**(4), 305–310.

[8] Weinstein, R. S., & Bell, N. H. (1988). Diminished rates of bone formation in normal black adults. *New England Journal of Medicine*, **31**
9(26), 1698–1701.