ANSWER BOOK FOR RASMUSSEN'S AN INTRODUCTION TO STATISTICS WITH DATA ANALYSIS

Lee Panas
Shelley Rasmussen

Brooks/Cole Publishing Company
Pacific Grove, California

AN INTRODUCTION TO STATISTICS

WITH

DATA ANALYSIS

by

SHELLEY RASMUSSEN

Department of Mathematical Sciences
Olney 428T
University of Massachusetts/Lowell
Lowell, MA 01854

Originally published by Brooks/Cole Publishing Company,
Division of Wadsworth, Inc.

ISBN 0-534-13578-1

© 1992 by Wadsworth, Inc.

© 2006 by Shelley Rasmussen

Permission is granted by the author for non-for-profit educational use.

Shelley_Rasmussen@uml.edu

Inferences About Variances

IN THIS CHAPTER

- Parametric inference about a single variance
- Confidence intervals for a variance
- Parametric comparisons of two variances
- Confidence intervals for the ratio of two variances
- Parametric comparisons of more than two variances (Bartlett's test)
- Robust inferences about two or more variances (Levene's test)

Is the variation among drained weights of tomatoes canned in the afternoon the same as the variation among those canned in the morning? Is the variability among plasma estrogen levels the same for monkeys with alcohol in their diets as for monkeys with no alcohol in their diets? Does the variation in rupture times for pieces of stainless steel depend on the level of stress applied to the pieces? Does season affect the variability of earthworm populations in fields?

All of these questions are phrased in terms of variation. Questions about variability are very important in quality control, engineering, and the sciences. In Chapters 10–13 we concentrated on inferences about means (or medians). Now we consider inferences about variances.

In Section 14-1 we consider parametric tests of hypotheses and confidence intervals for a single variance. Then in Section 14-2 we discuss parametric comparisons of two variances, as well as confidence intervals for the ratio of two variances. We cover parametric comparisons of several variances in Section 14-3. Finally, in Section 14-4 we consider inferences about two or more variances that do not require the assumption of Gaussian observations.

14-1 Parametric Inferences About a Variance

Suppose we have a random sample from a Gaussian distribution and we want to make inferences about the variance σ^2 of that distribution. First we will look at an example; then we will outline the significance level approach to a parametric test of hypotheses about a variance, and apply it to the example.

EXAMPLE 14-1

Machines at a factory fill cans with standard-grade tomatoes in puree (based on an example in Duncan, 1974, page 569; from Grant and Leavenworth, 1972, page 41). One responsibility of the quality control manager is to select cans and check the drained weight of the contents. After many checks, the manager has found that for cans filled in the morning, the average drained weight is 21.8 ounces, and the variance is 2.63 ounces^2.

The quality control manager selects a random sample of five cans filled one afternoon. The drained weights (in ounces) are:

22.5 19.5 21.5 20.5 20.0

A plot of these observations is shown in Figure 14-1. We see that the sample values have a fairly symmetrical distribution. The five sample drained weights range from 19.5 ounces to 22.5 ounces.

The manager wants to test the null hypothesis that the variance of drained weights for cans filled during that afternoon equals the morning variance, 2.63 ounces^2. Let's outline the analysis procedure.

The significance level approach to a parametric test of hypotheses about a variance σ^2

1. The hypotheses are H_0: $\sigma^2 = \sigma_0^2$ and H_a: $\sigma^2 \neq \sigma_0^2$, where σ_0^2 is a specified number.

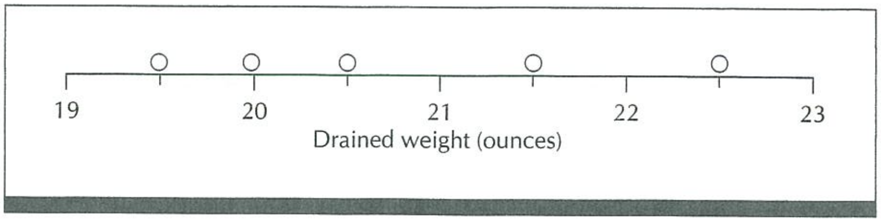

FIGURE 14-1 Drained weights (in ounces) of five cans of tomatoes canned in the afternoon, Example 14-1

2. Let s^2 denote the sample variance and n the sample size. The test statistic is

$$\text{Test statistic} = \frac{(n-1)\,s^2}{\sigma_0^2}$$

3. Assume that we have a random sample from a Gaussian distribution with variance σ^2. Then under the null hypothesis, the test statistic has the chi-square distribution with $n - 1$ degrees of freedom. Very large or very small values of the test statistic are inconsistent with the null hypothesis.
4. Select significance level α.
5. Let X denote a random variable having the chi-square distribution with $n - 1$ degrees of freedom. Find c_1 and c_2 from Table E such that $P(X \leq c_1) = \alpha/2$ and $P(X \geq c_2) = \alpha/2$. The acceptance region is the interval (c_1, c_2). The rejection region includes the intervals $[0, c_1]$ and $[c_2, \infty)$.
6. The decision rule is:

   If $c_1 <$ test statistic $< c_2$, say the results are consistent with the null hypothesis that the population variance equals σ_0^2.

   If test statistic $\leq c_1$ or test statistic $\geq c_2$, say the results are inconsistent with the null hypothesis, suggesting that the population variance does not equal σ_0^2.
7. Collect a sample that satisfies the assumptions in step 3. Calculate the test statistic in step 2. Use the decision rule in step 6 to decide whether the results are consistent with the null hypothesis. Draw conclusions based on the experimental results.

Suppose that, instead of a two-sided alternative, we have the one-sided alternative H_a: $\sigma^2 < \sigma_0^2$. Then in step 5 we find the number c from Table E such that $P(X \leq c) = \alpha$. Values of the test statistic greater than c are consistent with the null hypothesis; values less than or equal to c are inconsistent with the null hypothesis.

If we have the one-sided alternative H_a: $\sigma^2 > \sigma_0^2$, then in step 5 we find the number c from Table E such that $P(X \geq c) = \alpha$. Values of the test statistic less than c are consistent with the null hypothesis; values greater than or equal to c are inconsistent with the null hypothesis.

EXAMPLE 14-1 *(continued)*

Now we can test the hypotheses of interest in Example 14-1. If σ^2 denotes the variance of drained weights among cans filled that afternoon, then we can state the hypotheses as H_0: $\sigma^2 = 2.63$ ounces^2 and H_a: $\sigma^2 \neq 2.63$ ounces^2.

Assume that the five afternoon observations form a random sample from a large production lot. Assume also that these drained weights follow a Gaussian distribution. Figure 14-1 gives us no reason to doubt the Gaussian assumption, although the sample size is very small. We have no way of checking the other assumptions without more information.

Since the sample size is 5, the test statistic equals $4s^2/2.63$, where s^2 is the sample variance for the five afternoon observations. If the assumptions hold, then under the null hypothesis the test statistic has the chi-square distribution with 4 degrees of freedom.

We will use significance level $\alpha = .05$. From Table E we see that if X has the chi-square distribution with 4 degrees of freedom, then $P(X \leq .484) = .025$ and $P(X \geq 11.14) = .025$. Therefore, the acceptance region is (.484, 11.14), the rejection region consists of $[0, .484]$ and $[11.14, \infty)$, and the decision rule is:

If $.484 <$ test statistic < 11.14, say the results are consistent with the null hypothesis that the afternoon variance equals 2.63 ounces².

If test statistic $\leq .484$ or test statistic ≥ 11.14, say the results are inconsistent with the null hypothesis, suggesting that the afternoon variance does not equal 2.63 ounces².

The sample variance of the five observations is $s^2 = 1.45$ ounces², so the test statistic equals $4 \times 1.45/2.63$, or 2.2. Since 2.2 is in the acceptance region, the results are consistent with the null hypothesis. Based on this test of hypotheses, we have no reason to doubt that the variance for drained weights of tomatoes canned that afternoon equals the morning variance (but, of course, the test does not imply that the variance in the afternoon exactly equals the morning variance of 2.63 ounces²).

Confidence Intervals for a Population Variance

Suppose we want to calculate a 100A% *confidence interval for the population variance* σ^2. Let X denote a random variable having the chi-square distribution with $n - 1$ degrees of freedom. Find c_1 and c_2 such that $P(X \leq c_1) = (1 - A)/2$ and $P(X \geq c_2) = (1 - A)/2$. Then our confidence interval for σ^2 has the form

$$\left(\frac{n-1}{c_2} s^2, \frac{n-1}{c_1} s^2\right)$$

EXAMPLE 14-1 *(continued)*

Let's calculate a 95% confidence interval for the variance of drained weights of tomatoes canned that afternoon in Example 14-1. We have $n = 5$ and $s^2 = 1.45$. Since $A = .95$, we have $(1 - A)/2 = .025$. Referring to Table E, we see that $c_1 = .484$ and $c_2 = 11.14$, the same as we used for our test of hypotheses with significance level .05. So a 95% confidence interval for the variance is

$$\left(\frac{4}{11.14} \times 1.45, \frac{4}{.484} \times 1.45\right) = (.52 \text{ ounces}^2, 11.98 \text{ ounces}^2)$$

Note that our null hypothesis variance, 2.63 ounces², is in this confidence interval, in agreement with our test of hypotheses.

To get a confidence interval for the population standard deviation σ, we can take the square root of the upper and lower limits of the confidence interval for the population variance σ^2. For instance, a 95% confidence interval for the standard deviation of drained weights of tomatoes canned that afternoon in Example 14-1 is ($\sqrt{.52}$ ounce, $\sqrt{11.98}$ ounces) or (.72 ounce, 3.46 ounces).

Our analysis gives us no reason to think that the afternoon variance is different from the morning variance. However, as we have mentioned before, there are other important practical considerations in this type of situation. Is the amount of variation acceptable to the company, to consumers, and to the government? If there are tolerance ranges of acceptable values for drained weights, is the production process adequately meeting these tolerances? Generally, no single formal analysis procedure will address all of the questions relevant to a particular experimental situation.

In Section 14-2, we discuss parametric comparisons of two variances.

14-2

Parametric Inferences About Two Variances

Suppose we have two independent random samples from Gaussian distributions. We want to test the null hypothesis that the two population variances are equal. We will use the *variance ratio test* based on an *F* distribution.

The **variance ratio test** is a parametric test for equality of two variances.

We also want to calculate a confidence interval for the ratio of the two population variances. Let's begin with an example.

EXAMPLE 14-2

Researchers designed an experiment to study effects of regular alcohol consumption (Jerome Hojnacki, personal communication, 1986). The participants in the study were 20 adult male squirrel monkeys, of similar age and good health. The researchers randomly divided the monkeys into two equal sized groups. Monkeys in the alcohol group consumed a steady diet of 12% ethyl alcohol for approximately 3 months (ethyl alcohol constituted 12% of their total calories each meal). Monkeys in the control group did not consume alcohol. At the end of the treatment period, the researchers measured plasma estrogen (in nanograms/deciliter) for each monkey. The results are shown below.

Group	Plasma estrogen level (ng/dL)							Sample mean	Sample variance
Alcohol	3.17	2.52	2.59	4.25	3.27	4.92	5.46	3.61	1.33
	2.83	4.80	2.26						
Control	6.57	5.81	5.63	5.75	4.54	5.35	4.16	5.21	.556
	5.12	4.69	4.52						

A plot of the observations is shown in Figure 14-2.

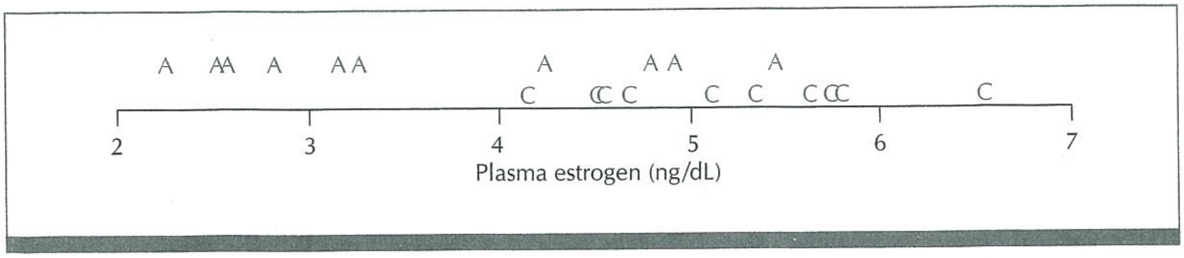

FIGURE 14-2 Plot of plasma estrogen levels in Example 14-2. The symbol A denotes a value for a monkey in the alcohol group; C denotes a value for a monkey in the control group.

The plasma estrogen levels seem to be somewhat lower for the monkeys in the alcohol group (see Exercise 11-5). Is the variation the same for the two groups? Let's outline the significance level approach for comparing the two variances, and then apply it to this example.

The significance level approach to parametric tests of hypotheses about two variances

1. Let σ_1^2 and σ_2^2 denote the variances of the two populations sampled. We want to test the hypotheses $H_0\colon \sigma_1^2 = \sigma_2^2$ and $H_a\colon \sigma_1^2 \neq \sigma_2^2$.
2. The test statistic equals the larger sample variance divided by the smaller sample variance:

$$\text{Test statistic} = \frac{\text{Larger sample variance}}{\text{Smaller sample variance}}$$

3. Assume that we have two independent random samples from Gaussian distributions. The first sample is of size n_1 and comes from a Gaussian distribution with variance σ_1^2. The second sample is of size n_2 and comes from a Gaussian distribution with variance σ_2^2. Let s_1^2 denote the first sample variance and s_2^2 the second sample variance. Then under the null hypothesis, the ratio s_1^2/s_2^2 has the $F(n_1 - 1, n_2 - 1)$ distribution. The ratio s_2^2/s_1^2 has the $F(n_2 - 1, n_1 - 1)$ distribution under the null hypothesis.
4. Select significance level α.
5. Suppose s_1^2 is the larger sample variance, so the test statistic is s_1^2/s_2^2. Let F denote a random variable having the $F(n_1 - 1, n_2 - 1)$ distribution. Find the number c from Table D such that $P(F \leq c) = 1 - \alpha/2$. The acceptance region is the interval $[0, c)$; the rejection region is the interval $[c, \infty)$.
6. The decision rule is:

   If test statistic $< c$, say the results are consistent with the null hypothesis that the two population variances are equal.

   If test statistic $\geq c$, say the results are inconsistent with the null hypothesis, suggesting that the two population variances are not equal.
7. Collect two samples that satisfy the assumptions in step 3. Calculate the test statistic in step 2. Use the decision rule in step 6 to decide whether the two population variances seem to be the same or different. Draw conclusions based on the experimental results.

EXAMPLE 14-2
(continued)

In Example 14-2, we want to test the null hypothesis that the variance of plasma estrogen levels is the same for monkeys fed a steady diet of alcohol as for monkeys not fed alcohol. The alternative hypothesis is that the variances are different for the two groups. Since the alcohol group has the larger sample variance, we let the alcohol group be group 1, and our test statistic is s_1^2/s_2^2.

We assume that we have independent random samples from Gaussian distributions. From Figure 14-2, we see that the Gaussian assumption seems reasonable for the control monkeys, but this is not so clear for the alcohol monkeys. We cannot check the independence assumptions without more information about the experiment. What suggestions about experimental design would you make to these researchers, in order to ensure independence and valid inferences?

We will use significance level $\alpha = .05$. Let F denote a random variable having the $F(9, 9)$ distribution. Since $P(F \leq 4.03) = .975$, the acceptance region is $[0, 4.03)$ and the rejection region is $[4.03, \infty)$. The decision rule is:

If test statistic < 4.03, say the results are consistent with the null hypothesis that the two variances are equal.
If test statistic ≥ 4.03, say the results are inconsistent with the null hypothesis, suggesting that the two variances are not equal.

The sample variance for the alcohol monkeys is $s_1^2 = 1.33$. The sample variance for the control monkeys is $s_2^2 = .556$. Therefore, the test statistic equals 1.33/.556, or 2.4, which is in the acceptance region. Using this test of hypothesis, we have no reason to doubt that the variation in plasma estrogen levels is similar for the alcohol and control monkeys.

Confidence Intervals for the Ratio of Two Population Variances

Suppose we would like to calculate a 100*A*% *confidence interval for the ratio* σ_1^2/σ_2^2 *of the two population variances.* Letting F_1 denote a random variable having an $F(n_1 - 1, n_2 - 1)$ distribution, find the number c_1 such that $P(F_1 \leq c_1) = (1 + A)/2$. Similarly, if F_2 denotes a random variable having the $F(n_2 - 1, n_1 - 1)$ distribution, find the number c_2 such that $P(F_2 \leq c_2) = (1 + A)/2$. A 100*A*% confidence interval for σ_1^2/σ_2^2 has the form

$$\left(\frac{1}{c_1}\frac{s_1^2}{s_2^2},\ c_2\frac{s_1^2}{s_2^2}\right)$$

EXAMPLE 14-2
(continued)

Let's calculate a 95% confidence interval for σ_1^2/σ_2^2 in Example 14-2. Here, σ_1^2 denotes the variance of plasma estrogen values for monkeys consuming alcohol and σ_2^2 denotes the variance for monkeys not consuming alcohol. Since the sample size is 10 for each group, the numerator and denominator degrees of freedom both equal 9. From Table D we see that $c_1 = c_2 = 4.03$. Our 95% confidence interval for σ_1^2/σ_2^2 is

$$\left(\frac{1}{4.03} \times \frac{1.33}{.556},\ 4.03 \times \frac{1.33}{.556}\right) = (.59, 9.6)$$

Note that if the null hypothesis is true, then the ratio of the two population variances is 1. This null hypothesis value is in the confidence interval, agreeing with our test of hypotheses.

Does the ratio σ_1^2/σ_2^2 of the two population variances have units of measurement associated with it? Should the confidence interval (.59, 9.6) for this ratio in Example 14-2 show units? How would you find a confidence interval for the ratio of the two population standard deviations σ_1/σ_2?

The Minitab Appendix for Chapter 14 has an example based on Exercise 14-6, making inferences about the ratio of two variances when the sample sizes are not equal.

Section 14-3 discusses a parametric procedure for comparing more than two variances.

14-3

Parametric Inferences About More Than Two Variances

Suppose we have k independent random samples from Gaussian distributions, where k is greater than or equal to 3. When we test the null hypothesis that the k population variances are equal, we are testing for *homogeneity of variances.* If the null hypothesis of equal variances is true, we say the k population variances are *homogeneous.* If the alternative hypothesis of unequal variances is true, we say the variances are *heterogeneous.*

There are many parametric tests for equality of variances (see, for example, Conover, Johnson, and Johnson, 1981). We will discuss the one most commonly used—*Bartlett's test.*

Bartlett's test is a parametric procedure for testing equality of three or more variances, based on the assumption of independent random samples from Gaussian distributions.

First, we will look at an example.

EXAMPLE 14-3

An engineer subjected uniform pieces of stainless steel to different levels of stress, recording the time to rupture for each piece. He tested six pieces of steel at each of three stress levels. The results are shown below (part of an experiment reported in Schmoyer, 1986; from Garofalo et al., 1961). Stress levels were reported in pounds per square inch (psi).

Stress level (psi)	Rupture time (hours)					
28.84	1,267	1,637	1,658	1,709	1,785	2,437
31.63	170	257	265	570	594	779
34.68	76	87	96	115	122	132

The experimental results are displayed in a scatterplot in Figure 14-3. We see that the time to rupture decreases as stress increases. It also looks like the

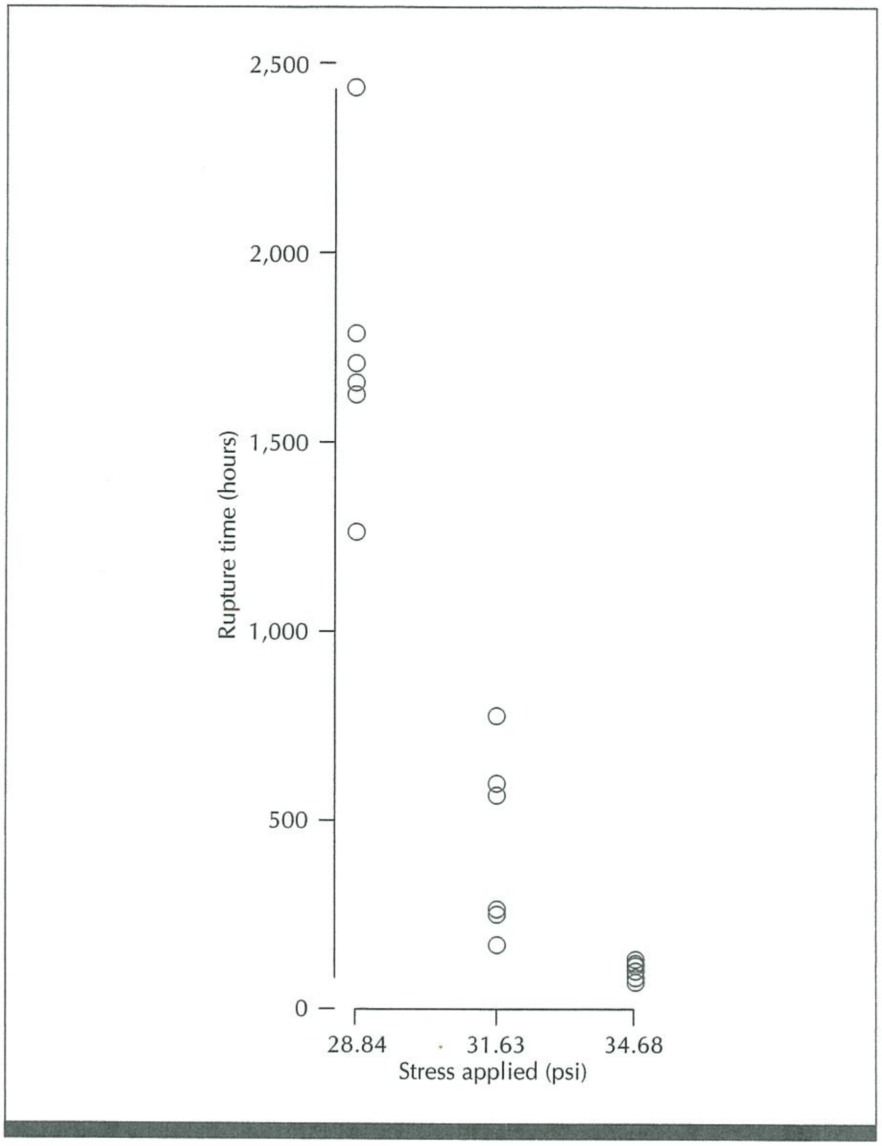

FIGURE 14-3 Scatterplot of rupture time vs. stress applied in Example 14-3

variation in rupture times decreases as stress increases. Let's see how to test the null hypothesis that the three population variances are equal.

The significance level approach to Bartlett's test for equality of several variances

1. Let σ_1^2 through σ_k^2 denote the k population variances. The hypotheses are:

$$H_0: \quad \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2$$
$$H_a: \quad \sigma_1^2 \text{ through } \sigma_k^2 \text{ are not all equal}$$

2. Let n_i denote the sample size and s_i^2 the sample variance of sample i. Let N denote the total sample size, the sum of n_1 through n_k. Let s_r^2 denote the pooled variance estimate (the residual mean square we discussed for one-way analysis of variance in Chapter 12). The numerator of the test statistic is

$$\text{Numerator} = 2.3026\left[(N - k)\log s_r^2 - \sum_{i=1}^{k}(n_i - 1)\log s_i^2\right]$$

where log denotes logarithm base-10. The denominator of the test statistic is

$$\text{Denominator} = 1 + \frac{1}{3(k - 1)}\left(\sum_{i=1}^{k}\frac{1}{n_i - 1} - \frac{1}{N - k}\right)$$

Then the test statistic is

$$\text{Test statistic} = \frac{\text{Numerator}}{\text{Denominator}}$$

3. Assume that we have k independent random samples, one from each of k Gaussian distributions. Then under the null hypothesis, the test statistic has approximately the chi-square distribution with $k - 1$ degrees of freedom. Large values of the test statistic are inconsistent with the null hypothesis.
4. Select significance level α.
5. Find the number c in Table E such that $P(X \le c) = 1 - \alpha$, where X has the chi-square distribution with $k - 1$ degrees of freedom. The acceptance region is $[0, c)$; the rejection region is $[c, \infty)$.
6. The decision rule is:

   If test statistic $< c$, say the results are consistent with the null hypothesis that the k population variances are all equal.

   If test statistic $\ge c$, say the results are inconsistent with the null hypothesis, suggesting that the k population variances are not all equal.
7. Carry out an experiment that satisfies the assumptions in step 3. Calculate the test statistic in step 2. Use the decision rule in step 6 to decide whether the population variances seem to be the same or different. Draw conclusions based on the experimental results.

EXAMPLE 14-3
(continued)

In Example 14-3, we want to test the null hypothesis that the variance in rupture times of uniform pieces of stainless steel is the same for all three stress levels. The alternative hypothesis is that the three variances are not all equal.

Assume that the three samples represent independent random samples from Gaussian distributions. From the plot in Figure 14-3, the Gaussian assumption does not seem unreasonable because all three sample distributions look fairly symmetric. As always, we cannot assess the independence assumption without more information about the experiment. What suggestions would you make regarding experimental design? How would you try to control extraneous sources of variation and ensure independence of observations? Should

Stress level	Sample size	Sample variance
28.84	6	145,877.8
31.63	6	58,509.37
34.68	6	472.6667

$k = 3 \qquad N = 18 \qquad N - k = 15 \qquad s_r^2 = 68{,}286.61$

$$\text{Numerator} = 2.3026[15 \log(68{,}286.61) - 5 \log(145{,}877.8) - 5 \log(58{,}509.37) - 5 \log(472.6667)] = 21.843$$

$$\text{Denominator} = 1 + \frac{1}{3(3-1)}\left(\frac{1}{5} + \frac{1}{5} + \frac{1}{5} - \frac{1}{15}\right) = 1.089$$

$$\text{Test statistic} = \frac{21.843}{1.089} = 20.1 \qquad \text{Degrees of freedom} = 3 - 1 = 2$$

the engineer subject the first six pieces to the first stress level, then reset the equipment and subject the second set of six pieces to the second stress level, then reset the equipment again and subject the third set of pieces to the final stress level? Or would you suggest a different procedure?

If our model assumptions hold, then under the null hypothesis, the test statistic has approximately the chi-square distribution with 2 degrees of freedom. Using significance level $\alpha = .01$, we find $c = 9.21$ from Table E. The acceptance region is $[0, 9.21)$, the rejection region is $[9.21, \infty)$, and the decision rule is:

If test statistic < 9.21, say the results are consistent with the null hypothesis that the variance in rupture times is the same for all three stress levels.
If test statistic ≥ 9.21, say the results are inconsistent with the null hypothesis, suggesting that the variation in rupture times is not the same for all three stress levels.

The calculations we need are outlined in Table 14-1. We see that the test statistic equals 20.1, which is in the rejection region. The results suggest that the variation in rupture times is not the same for all three stress levels, agreeing with what we saw in Figure 14-3.

We can make multiple comparisons to decide which variances seem to be different and which seem to be similar. We use the method of Section 14-2 to calculate confidence intervals for ratios of variances, then use the Bonferroni method (Section 12-1) to get a bound on the overall confidence level for these intervals.

The calculations for our multiple comparisons are outlined in Table 14-2. With three variances, there are three pairwise comparisons. We calculate a 98% confidence interval for each variance ratio. Since each sample size is 6, we use the value $c = 10.97$ from the $F(5, 5)$ distribution for all the intervals. The overall confidence level for the three intervals is at least 94%.

TABLE 14-2 Multiple comparisons of variances in Example 14-3. σ_1^2, σ_2^2, and σ_3^2 denote the variance in rupture times at stress levels 28.84, 31.63, and 34.68 psi, respectively.

Ratio of variances	98% confidence interval
$\frac{\sigma_1^2}{\sigma_2^2}$	$\left(\frac{1}{10.97} \times \frac{145{,}877.8}{58{,}509.37}, 10.97 \times \frac{145{,}877.8}{58{,}509.37}\right) = (.23, 27.35)$
$\frac{\sigma_1^2}{\sigma_3^2}$	$\left(\frac{1}{10.97} \times \frac{145{,}877.8}{472.6667}, 10.97 \times \frac{145{,}877.8}{472.6667}\right) = (28.13, 3{,}385.64)$
$\frac{\sigma_2^2}{\sigma_3^2}$	$\left(\frac{1}{10.97} \times \frac{58{,}509.37}{472.6667}, 10.97 \times \frac{58{,}509.37}{472.6667}\right) = (11.28, 1{,}357.93)$
	Overall confidence level $\geq 1 - (.02 + .02 + .02) = .94$

If 1 is in the confidence interval for a variance ratio, it suggests that the two variances are equal. If 1 is not in the interval, it suggests that the two variances are not equal. From Table 14-2 we see that 1 is in the first confidence interval and not in the other two intervals. These multiple comparisons suggest that the variation in rupture times is similar for stress levels 28.84 and 31.63 psi, while the variation at these two stress levels is much greater than the variation at stress level 34.68 psi. These results agree with our visual evaluation of variation in the three sample distributions illustrated in Figure 14-3.

There is a problem with using the variance ratio test of Section 14-2 for comparing two variances and Bartlett's test of this section for comparing more than two: These procedures are *not robust* to deviations from the Gaussian assumption.

> We say a procedure for testing hypotheses is **robust** if actual significance levels are close to the level we select, even under deviations from assumptions.

When using Bartlett's test and the variance ratio test, if the observations do not exactly follow Gaussian distributions, then the level α we use may be far from the actual significance level of the test. For comparison, t tests and analysis of variance for comparing means *are* quite robust to deviations from the Gaussian assumption, and somewhat robust to small deviations from the equal-variance assumption.

In Section 14-4, we discuss a procedure that does provide a robust test for equality of two or more variances.

14-4 Robust Inferences About Two or More Variances

Suppose we have k independent random samples, one from each of k populations, and we want to test the null hypothesis that the variation in the k

populations is the same. We will discuss a modification of a test proposed by Levene (1960). This modified procedure was recommended by Brown and Forsythe (1974) and shown by Conover, Johnson, and Johnson (1981) to work well in a variety of situations.

> **Levene's (modified) test** is a modified version of a procedure proposed by Levene to test for equality of two or more variances. This test is based on the assumption of independent random samples. The test is robust to deviations from Gaussian observations.

We will outline the p-value approach to Levene's modified test and then apply it to an example.

The p-value approach to Levene's (modified) test for equality of two or more variances

1. The null hypothesis states that the variance is the same in the k populations. The alternative hypothesis is that the variances are not the same in all k populations.
2. Let Y_{ij} denote the jth observation in sample i. Let m_i denote the ith sample median. Define new variables $Z_{ij} = |Y_{ij} - m_i|$. Z_{ij} is the absolute value of the difference between the observation Y_{ij} and the median m_i of sample i. To test our hypotheses, we go through the steps for one-way analysis of variance on the transformed observations Z_{ij}. The test statistic is

$$\text{Test statistic} = \frac{s_B^2}{s_r^2}$$

   where s_B^2 here denotes the between-groups mean square and s_r^2 the residual mean square, based on the Z_{ij}'s.
3. We assume that we have k independent random samples, one from each of k populations. Under the null hypothesis of equality of the k population variances, the test statistic has approximately the $F(k - 1, N - k)$ distribution, where N denotes the total sample size.
4. Carry out an experiment that satisfies the assumptions in step 3. Calculate the test statistic in step 2.
5. Find the p-value $= P(F \geq c_0)$, where c_0 denotes the observed value of the test statistic and F denotes a random variable having the $F(k - 1, N - k)$ distribution.
6. If the p-value is large, say the results are consistent with the null hypothesis that the variances are equal. If the p-value is small, say the results are inconsistent with the null hypothesis, suggesting that the variances are not all equal.

Let's illustrate Levene's modified test for equality of variances with an example.

EXAMPLE 14-4

Does the variation in earthworm populations depend on the time of year? To address this question, researchers divided a field into ten square plots. They watered these plots but did not treat them in any other way. (They added no chemicals, for example.) At three times over a 6-month period, the researchers selected equal sized subplots of the ten plots. (Each subplot was studied just once.) They applied an irritant to the subplots that caused the earthworms to rise to the surface. The researchers recorded total biomass/m² of the earthworms in each subplot. The results are shown below (part of a data set contributed by R. P. Blackshaw and P. J. Diggle to a collection of problems in Andrews and Herzberg, 1985, pages 301–306).

Time	Biomass/m² (values ordered from smallest to largest within the three times)						
1	7.73	8.07	10.61	17.01	17.55	26.98	28.59
	46.42	51.96	81.32				
2	.76	1.82	4.06	4.71	4.73	4.93	5.20
	12.45	37.29	39.57				
3	16.40	17.61	19.34	21.19	24.49	26.63	33.11
	39.12	39.26	53.32				

The observations are plotted in Figure 14-4. The average size of the earthworm populations seems to depend on the time of year. Does the variation

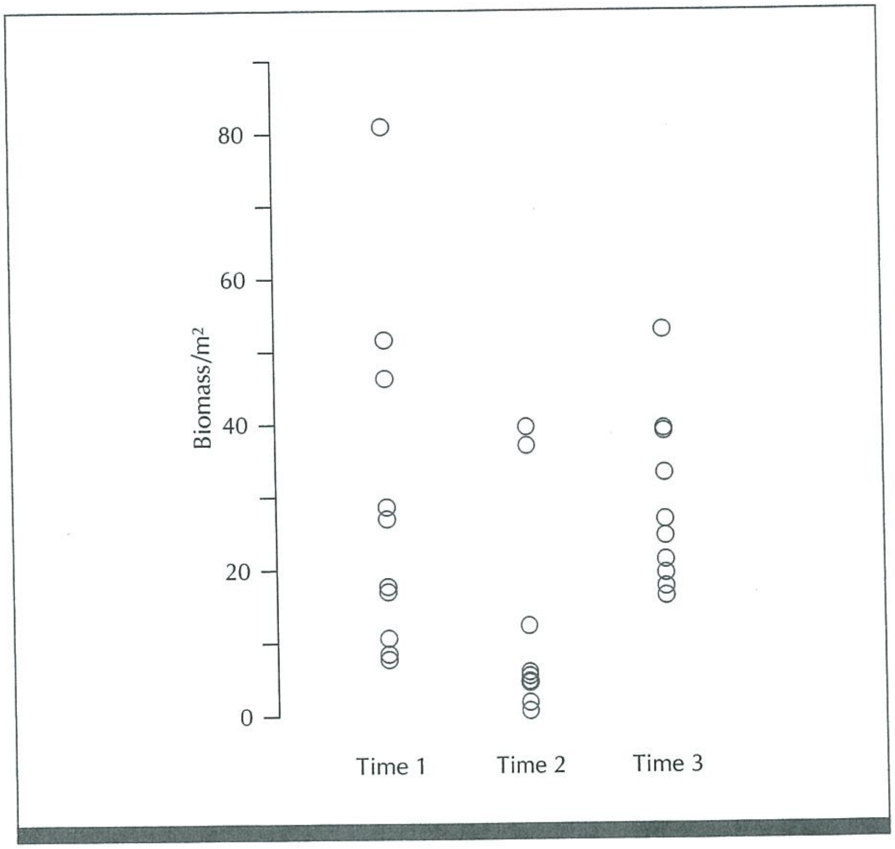

FIGURE 14-4 Plot of the biomass/m² of earthworm populations in Example 14-4

TABLE 14-3 Transformed observations to be used in one-way analysis of variance for Example 14-4

Time	Absolute value of the difference between the observation and the sample median						
1	14.535	14.195	11.655	5.255	4.715	4.715	6.325
	24.155	29.695	59.055				
2	4.07	3.01	.77	.12	.10	.10	.37
	7.62	32.46	34.74				
3	9.16	7.95	6.22	4.37	1.07	1.07	7.55
	13.56	13.70	27.76				

TABLE 14-4 Analysis of variance table for one-way analysis of variance on the values in Table 14-3

Source of variation	Sum of squares	Degrees of freedom	Mean square	Test statistic	*p*-value
Between groups	502	2	251	1.42	.26
Residual	4,778	27	177		
Total	5,280	29			

also depend on the time of year? We wish to test the null hypothesis that the variance of biomass/m^2 of earthworm populations among equal sized subplots of the field is the same at the three times; the alternative hypothesis is that the three variances are not all the same. We assume that we have three independent random samples, but we cannot judge the validity of this independence assumption without more information about the experiment.

The median of the ten observations at the first time is 22.265, the median at the second time is 4.83, and the median at the third time is 25.56. Table 14-3 shows, for each observation, the absolute value of the difference between the observation and its sample median.

Table 14-4 gives the analysis of variance table resulting from one-way analysis of variance on the values in Table 14-3.

The relatively large *p*-value of .26 is consistent with the null hypothesis. [If we had found differences, we could have used multiple comparisons (see Section 12-2) on the transformed observations (Table 14-3) to decide which variances seem to be similar and which different.] Looking at the plot of sample values in Figure 14-4, we see that the variation at time 1 was somewhat larger than the variation at the other two times. However, the conclusion that the variation in earthworm populations appears to be similar for the three times does not seem unreasonable.

Summary of Chapter 14

Parametric procedures for making inferences about one or more variances depend on the assumption that the observations are Gaussian distributed. None of these procedures is robust to deviations from the Gaussian assumption. That is, if the observations are not really from Gaussian distributions, the *p*-values for tests of hypotheses and confidence levels for interval estimates may be very wrong (and therefore meaningless).

Levene's modified procedure for testing equality of variances is robust to deviations from the Gaussian assumption. This means that we can feel comfortable interpreting *p*-values even when the observations do not come from Gaussian distributions.

Exercises for Chapter 14

In each exercise, plot the observations in any ways that seem helpful. Describe the population(s) sampled, whether real or hypothetical. For each procedure, state the assumptions that make the analysis valid. Do these assumptions seem reasonable? What additional information would you like to have about the experiment? Discuss the results of your analysis.

EXERCISE 14-1

An engineer studied the time to rupture for pieces of stainless steel at two levels of stress. He tested six uniform pieces of steel at each of the two stress levels. (This is a separate phase of the experiment discussed in Example 14-3.) The results are shown below (Schmoyer, 1986; from Garofalo et al., 1961). Stress levels are in pounds per square inch (psi).

Stress level (psi)	Rupture time (hours)					
41.69	6.6	9.6	11.2	12.3	19.7	20.4
45.71	1.9	3.9	4.3	4.6	5.7	9.0

a. Plot these observations.

b. Test the null hypothesis that the population variances for rupture times are equal at the two stress levels. Do the assumptions for the analysis seem reasonable?

c. Calculate a confidence interval for the ratio of the two population variances.

EXERCISE 14-2

In Example 10-3, we looked at the average weight in grams of six pairs of twins born to exercised Pygmy goats (Dhindsa, Metcalfe, and Hummels, 1978):

745.5 1,175.0 1,290.0 1,364.5 1,397.5 1,660.0

a. Plot the observations.

b. Test the null hypothesis that the variance in average weights of such pairs of twins is 10,000 grams^2.

c. Calculate a 99% confidence interval for the variance in average weights of pairs of twins born to Pygmy goats treated like those in this experiment.

EXERCISE 14-3

In Example 10-4, we considered the height in inches of five bomb bases sampled in a 15-minute interval (Duncan, 1974, page 43; Hollander and Proschan, 1984, page 42; from Kauffman, 1945):

.826 .829 .831 .836 .840

a. Plot the observations.

b. Test the null hypothesis that the variance for heights of bomb bases pro-

EXERCISE 14-7

In Exercise 11-8, we looked at specific airway resistance 30 minutes after administration of a bronchodilating aerosol by patients using either hand administration or an automatic inhalation device (units not given) (Box, Hunter, and Hunter, 1978, page 158; from a larger study reported by F. J. McIlneath and B. M. Cohen in *J. Med.,* 1970, volume 1, page 229):

Hand:	17.00	22.80	21.60	20.40	11.20	14.00
	52.25	7.50	12.20	18.85	6.05	4.05
Automatic:	11.60	11.60	13.65	17.22	8.25	6.20
	41.50	6.96	8.40	9.00	5.18	3.00

a. Plot the observations.

b. Test the null hypothesis that the variance in specific airway resistance is the same for the two methods of administration.

c. Calculate a 98% confidence interval for the ratio of the two variances.

EXERCISE 14-8

In Exercise 11-7, we looked at sputum histamine levels (μg/g dry weight sputum) for 9 allergic people and 13 nonallergic people, all smokers (Hollander and Wolfe, 1973, page 74; a subset of data in Thomas and Simmons, 1969):

Allergics:	31.0	39.6	64.7	65.9	67.9	100.0
	102.4	1,112.0	1,651.0			
Nonallergics:	4.7	5.2	6.6	18.9	27.3	29.1
	32.4	34.3	35.4	41.7	45.5	48.0
	48.1					

a. Plot the observations.

b. Test the null hypothesis that the variance in sputum histamine levels is the same for allergic and nonallergic smokers.

c. Calculate a 98% confidence interval for the ratio of the two variances.

d. Take the logarithm of each observation. Test the null hypothesis that the variance of the logarithm of sputum histamine level is the same for allergic and nonallergic smokers.

e. Calculate a 98% confidence interval for the ratio of the two variances of the logarithm of sputum histamine level.

f. Discuss your findings.

EXERCISE 14-9

In Exercise 11-6, we considered plasma testosterone levels (nanograms/deciliter) of monkeys on two different diets (Jerome Hojnacki, 1986, personal communication):

duced during that 15-minute interval was .0001 inch². [Recall that the specifications were .830 ± .01 inch and $(.01 \text{ inch})^2 = .0001 \text{ inch}^2$.]

c. Calculate a 98% confidence interval for the variance of heights of bomb bases produced during that 15-minute interval.

EXERCISE 14-4

In Exercise 10-7, we considered determinations of serum iron concentrations (μg/100 ml) using a new method (Hollander and Wolfe, 1973, pages 85–86; a portion of the data in Jung and Parekh, 1970):

96	98	99	100	103	103	104	104	105	105
106	106	107	108	108	108	110	113	114	114

a. Plot the observations.

b. Test the null hypothesis that the variance in serum iron concentration determinations using this method is 100 $(\mu\text{g}/100 \text{ ml})^2$.

c. Calculate a 95% confidence interval for the variance in serum iron concentration determinations using this method.

EXERCISE 14-5

In Exercise 10-1, we considered plasma citrate concentrations (μmol/liter) before breakfast for ten volunteers (from a contribution by E. B. Jensen to a collection of problems in Andrews and Herzberg, 1985, page 237; from Andersen, Jensen, and Schou, 1981).

93	116	125	144	105	109	89	116	151	137

a. Plot the observations.

b. Test the null hypothesis that the variance of before-breakfast plasma citrate concentrations is 1,500 $(\mu\text{mol/liter})^2$.

c. Calculate a 90% confidence interval for the variance of before-breakfast plasma citrate concentrations.

EXERCISE 14-6

In Exercise 11-12, we discussed change in pupil diameter (in millimeters) for volunteers after two different treatments (Box, Hunter, and Hunter, 1978, page 160; from H. W. Elliott, G. Navarro, and N. Nomof, *J. Med.,* 1970, volume 1, page 77).

Morphine:	.08	.8	1.0	1.9	2.0	2.4
Nalbuphine:	−.3	.0	.2	.4	.8	

a. Plot the observations.

b. Test the null hypothesis that the variance of change in pupil diameter is the same after both treatments.

c. Calculate a 90% confidence interval for the ratio of the two variances.

Alcohol:	313.99	152.06	145.64	128.86	262.16	251.29
	505.55	94.79	157.49	171.81		
Control:	632.92	308.56	1239.68	440.38	233.02	142.67
	84.91	342.63	1005.66	735.61		

a. Plot the observations.

b. Test the null hypothesis that the variance in plasma testosterone levels is the same for monkeys on the two diets.

c. Calculate a 90% confidence interval for the ratio of the two variances.

d. Take the logarithm of each observation. Test the null hypothesis that the variance of the logarithm of plasma testosterone level is the same for monkeys on the two diets.

e. Calculate a 90% confidence interval for the ratio of the two variances for the logarithm of plasma testosterone level.

EXERCISE 14-10

In Exercise 11-4, we looked at eight independent determinations (in °C) of the melting point of hydroquinone by each of two analysts (Duncan, 1974, pages 575–576; from Wernimont, 1947, page 8):

Analyst 1:	174.0	173.5	173.0	173.5	171.5	172.5
	173.5	173.5				
Analyst 2:	173.0	173.0	172.0	173.0	171.0	172.0
	171.0	172.0				

a. Plot the observations.

b. Test the null hypothesis that the variance of determinations is the same for the two analysts.

c. Calculate a 95% confidence interval for the ratio of the variances for the two analysts.

EXERCISE 14-11

In Exercise 11-3, we considered the total score of five shots at a target by an experienced shooter using a revolver. There were eight trials with each of two types of ammunition (Snow, 1986):

.22 Magnum	42	43	46	47	46	47	39	47
.22 Long Rifle	41	43	41	41	40	40	45	47

a. Plot the observations.

b. Test the null hypothesis that the variance of total scores by this shooter is the same for the two types of ammunition.

c. Calculate a 95% confidence interval for the ratio of variances for the two types of ammunition.

EXERCISE 14-12 In Exercise 15-24, we discuss a study of the permeability of concrete (inches per hour) with six different levels of asphalt content (Mendenhall and Sincich, 1988, page 495; from Woelfl et al., 1981):

3% asphalt:	1,189	840	1,020	980	**6% asphalt:**	707	927	1,067	822
4% asphalt:	1,440	1,227	1,022	1,293	**7% asphalt:**	853	900	733	585
5% asphalt:	1,227	1,180	980	1,210	**8% asphalt:**	395	270	310	208

a. Plot the observations.

b. Use a parametric analysis to test for equality of variances of permeability at each of the six asphalt contents.

c. Use a nonparametric analysis to test for equality of variances of permeability at each of the six asphalt contents.

d. Compare your answers to parts (b) and (c).

EXERCISE 14-13 In Exercise 15-17, we consider times to failure (in hours) of samples of insulation for electrical motors in accelerated life testing at four temperatures (Nelson, 1986, pages 20–21):

190 °C:	7,228	7,228	7,228	8,448	9,167	9,167
	9,167	9,167	10,511	10,511		
220 °C:	1,764	2,436	2,436	2,436	2,436	2,436
	3,108	3,108	3,108	3,108		
240 °C:	1,175	1,175	1,521	1,569	1,617	1,665
	1,665	1,713	1,761	1,953		
260 °C:	600	744	744	744	912	1,128
	1,320	1,464	1,608	1,896		

a. Plot the observations.

b. Use a parametric analysis to test for equality of variances in failure times at the four temperatures. Use the Bonferroni method to make multiple comparisons.

c. Use a nonparametric analysis to test for equality of variances in failure times at the four temperatures. Use the Bonferroni method to make multiple comparisons.

d. Compare your answers to parts (b) and (c).

e. Take the logarithm of each failure time. Use a parametric analysis to test for equality of variances of the logarithm of failure time at the four temperatures.

f. Use a nonparametric analysis to test for equality of variances of the logarithm of failure time at the four temperatures.

g. Compare your answers to parts (e) and (f).

h. Discuss your findings.

EXERCISE 14-14 In Exercise 15-16, we discuss instrument response at five concentrations of copper in solution for an experiment in atomic absorption spectroscopy (Carroll, Sacks, and Spiegelman, 1988):

Copper in solution (micrograms/milliliter)	Instrument response in absorbance units			
.0	.045	.047	.051	.054
.050	.084	.087		
.100	.115	.116		
.200	.183	.191		
.500	.395	.399		

a. Plot the observations.

b. Use a parametric analysis to test for equality of variances of instrument response across copper concentrations.

c. Use a nonparametric analysis to test for equality of variances of instrument response across copper concentrations.

d. Compare your answers to parts (b) and (c).

EXERCISE 14-15 In Exercise 12-6, we compared the working life (thousands of cycles until failure) of three types of stopwatch (Rice, 1988, page 432; from Natrella, 1963):

Type 1:	1.7	1.9	6.1	12.5	16.5	25.1	30.5	42.1
	82.5							
Type 2:	13.6	19.8	25.2	46.2	46.2	61.1		
Type 3:	13.4	20.9	25.1	29.7	46.9			

a. Plot the observations.

b. Use a parametric analysis to test for equality of variances of working life for the three stopwatch types. Use the Bonferroni method to make multiple comparisons.

c. Use a nonparametric analysis to test for equality of variances of working life for the three stopwatch types. Use the Bonferroni method to make multiple comparisons.

d. Compare your results in parts (b) and (c).

EXERCISE 14-16 In Exercise 12-2, we looked at the stimulation index of men treated with one of three different regimens of a synthetic vaccine for malaria or with a saline regimen (Patarroyo et al., 1988):

Saline control:	1.4	1.0	4.0	**Regimen 2:**	6.6	9.1		
Regimen 1:	1.5	5.6	12.4	**Regimen 3:**	35.1	13.4	.8	3.3

a. Plot the observations.

b. Use a parametric analysis to test for equality of variances of stimulation index under the four regimens. Use the Bonferroni method to make multiple comparisons.

c. Use a nonparametric analysis to test for equality of variances of stimulation index under the four regimens. Use the Bonferroni method to make multiple comparisons.

d. Compare your results in parts (b) and (c).

e. Take the logarithm of each observation. Use a parametric analysis to test for equality of variances of the logarithm of stimulation index under the four regimens.

f. Use a nonparametric analysis to test for equality of variances of the logarithm of stimulation index under the four regimens.

g. Compare your results in parts (e) and (f).

EXERCISE 14-17 In Exercise 13-5, we considered an experiment on iron retention in mice. Researchers measured percentage of iron retained for mice under six sets of conditions (from an example in Rice, 1988, pages 356–357):

Fe^{3+}, 10.2 millimolar:	.71	1.66	2.01	2.16	2.42	2.42	2.56	2.60	3.31
	3.64	3.74	3.74	4.39	4.50	5.07	5.26	8.15	8.24
Fe^{3+}, 1.2 millimolar:	2.20	2.93	3.08	3.49	4.11	4.95	5.16	5.54	5.68
	6.25	7.25	7.90	8.85	11.96	15.54	15.89	18.30	18.59
Fe^{3+}, .3 millimolar:	2.25	3.93	5.08	5.82	5.84	6.89	8.50	8.56	9.44
	10.52	13.46	13.57	14.76	16.41	16.96	17.56	22.82	29.13
Fe^{2+}, 10.2 millimolar:	2.20	2.69	3.54	3.75	3.83	4.08	4.27	4.53	5.32
	6.18	6.22	6.33	6.97	6.97	7.52	8.36	11.65	12.45
Fe^{2+}, 1.2 millimolar:	4.04	4.16	4.42	4.93	5.49	5.77	5.86	6.28	6.97
	7.06	7.78	9.23	9.34	9.91	13.46	18.40	23.89	26.39
Fe^{2+}, .3 millimolar:	2.71	5.43	6.38	6.38	8.32	9.04	9.56	10.01	10.08
	10.62	13.80	15.99	17.90	18.25	19.32	19.87	21.60	22.25

a. Plot the observations.

b. Use a parametric analysis to test for equality of variances of percentage of iron retained under the six sets of conditions. Use the Bonferroni method to make multiple comparisons of variances.

c. Use a nonparametric analysis to test for equality of variances of percentage of iron retained under the six sets of conditions. Compare with the parametric test in part (b).

d. Take the logarithm of each observation. Use a parametric analysis to test for equality of variances of the logarithm of percentage of iron retained under the six sets of conditions.

e. Use a nonparametric analysis to test for equality of variances of the logarithm of percentage of iron retained under the six sets of conditions. Compare with the parametric test in part (d).

f. Discuss your findings.

EXERCISE 14-18 In Exercise 13-2, we considered the effects of pH and temperature on optical density (units not given) of a polymer latex (Gasper, 1988; with permission of ICI Resins US):

pH 9.0, 85 °C:	56.6	38.9
pH 9.0, 95 °C:	39.0	37.5
pH 9.3, 85 °C:	63.0	96.8
pH 9.3, 95 °C:	33.0	33.3

a. Plot the observations.

b. Use a parametric analysis to test for equality of variances under the four sets of conditions.

c. Use a nonparametric analysis to test for equality of variances under the four sets of conditions.

d. Compare your answers to parts (b) and (c).

e. Take the reciprocal of each observation. Use a parametric analysis to test for equality of variances of the reciprocal of optical density under the four sets of conditions.

f. Use a nonparametric analysis to test for equality of variances of the reciprocal of optical density under the four sets of conditions.

g. Compare your answers to parts (e) and (f).

h. Discuss your findings.

EXERCISE 14-19 In Exercise 13-1, we looked at the distances (in feet) a player hit a softball under four sets of conditions (Shaughnessy, 1988):

Dudley Thunder, wood bat:	242	230	250	242
Dudley Thunder, aluminum bat:	270	282	265	277
Worth Red Dot, wood bat:	258	264	265	275
Worth Red Dot, aluminum bat:	290	318	302	310

a. Plot the observations.

b. Use a parametric analysis to test for equality of variances under the four sets of conditions.

c. Use a nonparametric analysis to test for equality of variances under the four sets of conditions.

d. Compare your answers to parts (b) and (c).

EXERCISE 14-20

In Example 13-2, we considered flexural strength (in pounds per square inch) of sheet castings of a polymer under four sets of conditions (Duncan, 1974, page 685; from Gore, 1947):

20 minutes, 100°:	9,500	10,650	9,700	9,950	10,100
20 minutes, 120°:	11,300	11,750	11,600	11,650	11,700
60 minutes, 100°:	11,500	11,650	11,250	11,250	11,900
60 minutes, 120°:	10,900	11,500	11,850	11,700	11,650

a. Plot the observations.

b. Use a parametric analysis to test for equality of variances under the four sets of conditions.

c. Use a nonparametric analysis to test for equality of variances under the four sets of conditions.

d. Compare your results in parts (b) and (c).

EXERCISE 14-21

Consider the data on rupture times of pieces of stainless steel at different stress levels, in Example 14-3.

a. Take the logarithm of each rupture time. Use a parametric analysis to test for equality of variances for the logarithm of rupture time at the three stress levels.

b. Use a nonparametric analysis to test for equality of variances for the logarithm of rupture times at the three stress levels. Compare with your results in part (a).

c. Compare your results in part (a) with what we found in Example 14-3.

EXERCISE 14-22 Consider the data on rupture times of pieces of stainless steel at different stress levels, in Example 14-3.

a. Use a nonparametric analysis to test for equal variances of rupture times at the three stress levels. Use the Bonferroni method to make multiple comparisons.

b. Compare your results in part (a) with what we found in Example 14-3.