Chapter 18

# ANOVA Diagnostics and Remedial Measures

When discussing regression analysis, we emphasized the importance of examining the appropriateness of the regression model under consideration, and noted the effectiveness of residual plots and other diagnostics for spotting major departures from the tentative model. Examination of the appropriateness of analysis of variance models is no less important.

In this chapter, we take up the use of residual plots for diagnosing the appropriateness of analysis of variance models, as well as formal tests for the constancy of the error variance. We also discuss the use of transformations of the response variable as a remedial measure to improve the appropriateness of the analysis of variance model for estimation and test inferences.

For pedagogic reasons, as in regression analysis, we have discussed inference procedures before diagnostics and remedial measures. The actual sequence of developing and using any statistical model is, of course, the reverse:

1. Examine whether the proposed model is appropriate for the set of data at hand.
2. If the proposed model is not appropriate, consider remedial measures, such as transformation of the data or modification of the model.
3. After review of the appropriateness of the model and completion of any necessary remedial measures and an evaluation of their effectiveness, inferences based on the model can be undertaken.

It is not necessary, nor is it usually possible, that an ANOVA model fit the data perfectly. As will be noted later, ANOVA models are reasonably robust against certain types of departures from the model, such as the error terms not being exactly normally distributed. The major purpose of the examination of the appropriateness of the model is therefore to detect serious departures from the conditions assumed by the model.

## 18.1 Residual Analysis

Residual analysis for ANOVA models corresponds closely to that for regression models. We therefore discuss only briefly some key issues in the use of residual analysis for ANOVA models.

## Residuals

The residuals $e_{ij}$ for the ANOVA cell means model (16.2) were defined in (16.20):

$$e_{ij} = Y_{ij} - \hat{Y}_{ij} = Y_{ij} - \bar{Y}_{i\cdot} \tag{18.1}$$

As in regression, semistudentized residuals, studentized residuals, and studentized deleted residuals are often helpful for diagnosing ANOVA model departures. The definitions of these residuals for regression in Chapters 3 and 10 are still applicable for ANOVA models. However, in view of the simple nature of the **X** matrix for ANOVA models, the regression formulas often simplify here. The semistudentized residuals $e_{ij}^*$ in (3.5) for regression remain unchanged:

$$e_{ij}^* = \frac{e_{ij}}{\sqrt{MSE}} \tag{18.2}$$

The studentized residuals $r_{ij}$ in (10.20) become here:

$$r_{ij} = \frac{e_{ij}}{s\{e_{ij}\}} \tag{18.3}$$

where:

$$s\{e_{ij}\} = \sqrt{\frac{MSE(n_i - 1)}{n_i}} \tag{18.3a}$$

Finally, the studentized deleted residuals $t_{ij}$ in (10.26) become here:

$$t_{ij} = e_{ij}\left[\frac{n_T - r - 1}{SSE\left(1 - \frac{1}{n_i}\right) - e_{ij}^2}\right]^{1/2} \tag{18.4}$$

### Comment

For ANOVA model (16.2), it can be shown that the leverage of $Y_{ij}$, defined in (10.18), is given by:

$$h_{ij,ij} = \frac{1}{n_i} \tag{18.5}$$

Hence, the variance of the residual $e_{ij}$ for ANOVA model (16.2) can be obtained by substituting (18.5) into (10.14):

$$\sigma^2\{e_{ij}\} = \frac{\sigma^2(n_i - 1)}{n_i} \tag{18.6}$$

Replacing $\sigma^2$ by the unbiased estimator $MSE$ and taking the square root lead to the estimated standard deviation $s\{e_{ij}\}$ in (18.3a).

When the treatment sample sizes $n_i$ are the same, the leverages of all the observations $Y_{ij}$ are the same. As a result, the estimated standard deviations of the residuals, $s\{e_{ij}\}$, are all the same so that the semistudentized residuals $e_{ij}^*$ and the studentized residuals $r_{ij}$ provide essentially the same information, differing only by a constant factor near 1 unless the treatment sample size is very small. ■

## Residual Plots

Residual plots useful for analysis of variance models include: (1) plots against the fitted values, (2) time or other sequence plots, (3) dot plots, and (4) normal probability plots. All of these plots have been encountered previously. We therefore proceed directly to an example to illustrate the use of residual plots for evaluating the appropriateness of analysis of variance models.

**Example**

Table 18.1 contains a portion of the residuals for the rust inhibitor example of Chapter 17. For ease of presentation, the treatments are shown in the columns of the table. The residuals were obtained from the data in Table 17.2a. For instance, the residual for the first experimental unit treated with brand A rust inhibitor is:

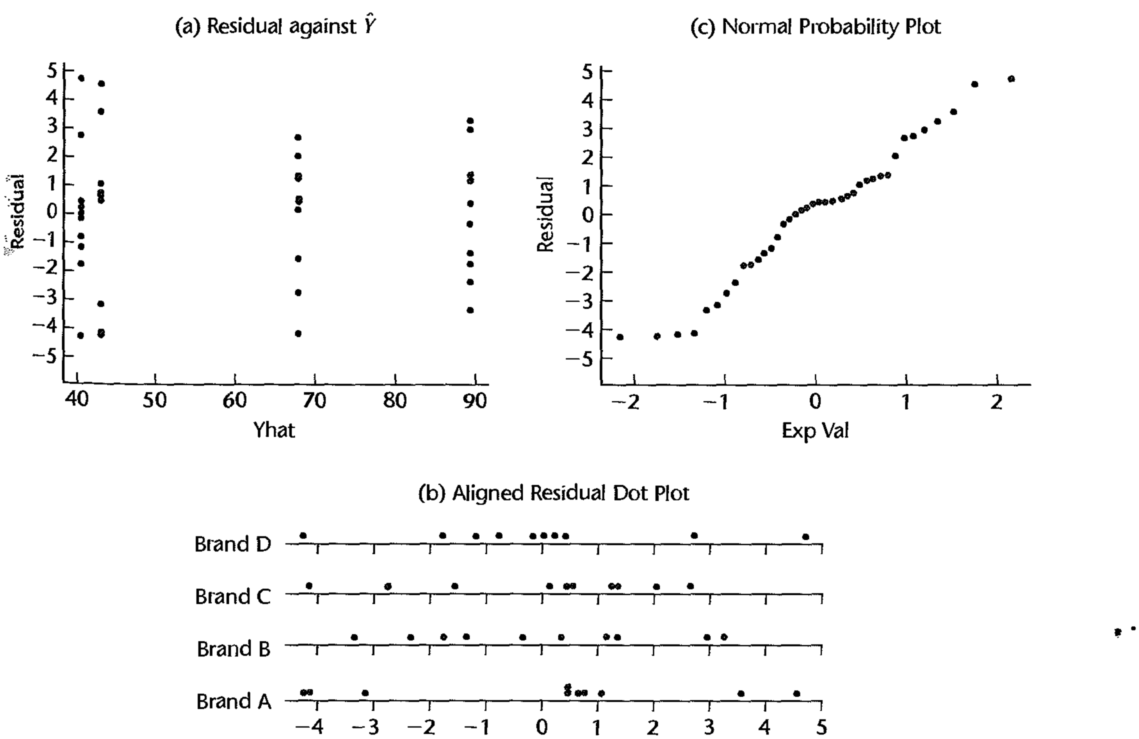
$$e_{11} = Y_{11} - \hat{Y}_{11} = Y_{11} - \bar{Y}_{1\cdot} = 43.9 - 43.14 = .76$$

Figure 18.1 presents three MINITAB diagnostic residual plots. Figure 18.1a contains a *residual plot against the fitted values*. This plot differs in appearance from similar plots for

**TABLE 18.1** Residuals—Rust Inhibitor Example.

| | Brand | | | |
|---|---|---|---|---|
| | A | B | C | D |
| $j$ | $i=1$ | $i=2$ | $i=3$ | $i=4$ |
| 1 | .76 | .36 | .45 | −4.27 |
| 2 | −4.14 | −2.34 | 1.35 | 4.73 |
| 3 | 3.56 | 3.26 | .55 | .23 |
| ... | ... | ... | ... | ... |
| 8 | −4.24 | −1.34 | −2.75 | −1.77 |
| 9 | .46 | 1.36 | −4.15 | .43 |
| 10 | −3.14 | −.34 | 1.25 | −.77 |

**FIGURE 18.1** MINITAB Diagnostic Residual Plots—Rust Inhibitor Example.



regression analysis because the fitted values $\hat{Y}_{ij}$ here are the same for all observations for a given factor level. Recall from (16.17) that $\hat{Y}_{ij} = \bar{Y}_{i.}$.

Figure 18.1b contains *aligned dot plots* of the residuals for each factor level. These plots are similar to the residual plot against the fitted values in Figure 18.1a, except here the residual scale is the horizontal one. An advantage of the plot in Figure 18.1a is that it facilitates an assessment of the relation between the magnitudes of the error variances and the factor level means. A disadvantage is that some of the estimated factor level means may be far apart, making a comparison of the factor levels more difficult. This difficulty is remedied in Figure 18.1b since dot plots can be placed close together to facilitate comparisons between factor levels.

Figure 18.1c contains a *normal probability plot* of the residuals. This plot is exactly the same as for regression models.

No sequence plot of the residuals is presented here because the data for the rust inhibitor example were not ordered according to time or in some other logical sequence.

All of the plots in Figure 18.1, as we shall see, suggest that ANOVA model (16.2) is appropriate for the rust inhibitor data.

## Diagnosis of Departures from ANOVA Model

We consider now how residual plots can be helpful in diagnosing the following departures from ANOVA model (16.2):

1. Nonconstancy of error variance
2. Nonindependence of error terms
3. Outliers
4. Omission of important explanatory variables
5. Nonnormality of error terms

**Nonconstancy of Error Variance.** ANOVA model (16.2) requires that the error terms $\varepsilon_{ij}$ have constant variance for all factor levels. When the sample sizes are not large and do not differ greatly, the appropriateness of this assumption can be studied by using the residuals, semistudentized residuals, or studentized residuals. *Plots of residuals against fitted values* or *dot plots of residuals* are helpful. When the sample sizes differ greatly, studentized residuals should be used in these plots. Constancy of the error variance is shown in these plots by the plots having about the same extent of scatter of the residuals around zero for each factor level. This is the case for the rust inhibitor example in Figures 18.1a and 18.1b.

Figure 18.2 is a prototype residual plot against the fitted values when the error variances are not constant. This plot portrays the case where the error terms for factor level 3 have a larger variance than those for the other two factor levels.

When the sample sizes for the different factor levels are large, *histograms* or boxplots of the residuals for each treatment—arranged vertically and using the same scale, like the dot plots in Figure 18.1b—are an effective means for examining the constancy of the error variance, as well as for assessing whether the error terms are normally distributed.

A number of statistical tests have been developed for formally examining the equality of the $r$ factor level variances; two of these tests will be discussed in Section 18.2.

**Nonindependence of Error Terms.** Whenever data are obtained in a time sequence, a *residual sequence plot* should be prepared to examine if the error terms are serially

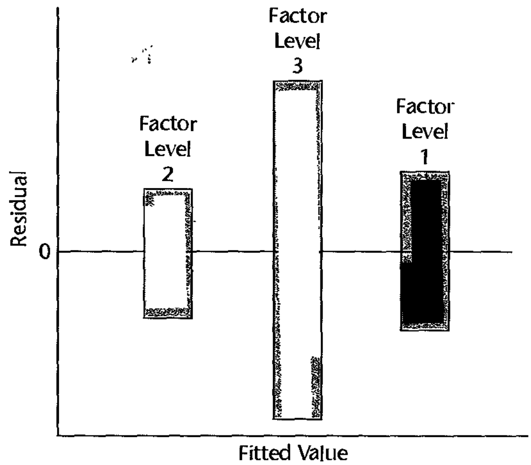



**FIGURE 18.2** Prototype Plot of Residuals against Fitted Values When Error Term Variance Is Not Constant for Factor Levels.

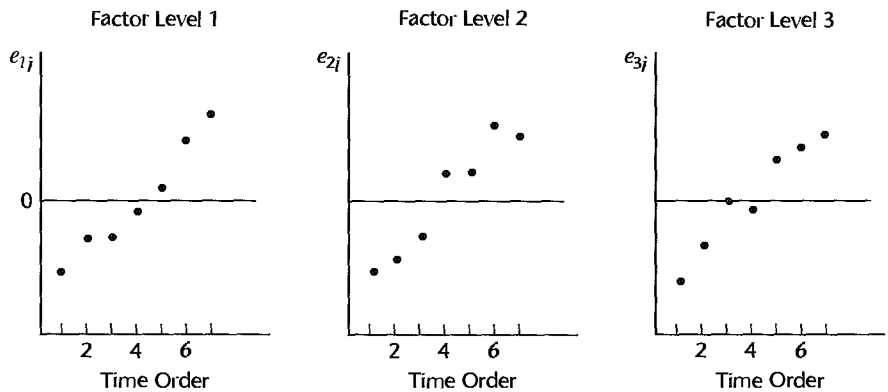



**FIGURE 18.3** Residual Sequence Plots for Group Interaction Study Illustrating Time-Related Effect.

correlated. Figure 18.3 contains the residuals for an experiment on group interactions. Three different treatments were applied, and the group interactions were recorded on videotapes. Seven replications were made for each treatment. Afterward, the experimenter measured the number of interactions by viewing the tapes in randomized order. Figure 18.3 strongly suggests that the experimenter discerned a larger number of interactions as more experience in viewing the tapes was gained. As a result, the residuals in Figure 18.3 appear to be serially correlated. In this instance, an inclusion in the model of a linear term for the time effect might be sufficient to assure independence of the error terms in the revised model.

Time-related effects may also lead to increases or decreases in the error variance over time. For instance, an experimenter may make more precise measurements over time. Figure 18.4 portrays residual sequence plots where the error variance decreases over time.

When the data are ordered in some other logical sequence, such as in a geographic sequence, a plot of the residuals against this ordering is helpful for ascertaining whether the error terms are serially correlated according to this ordering.

**Outliers.** The detection of outliers is facilitated by various plots of the studentized deleted residuals. *Residual plots against fitted values, residual dot plots, box plots,* and *stem-and-*

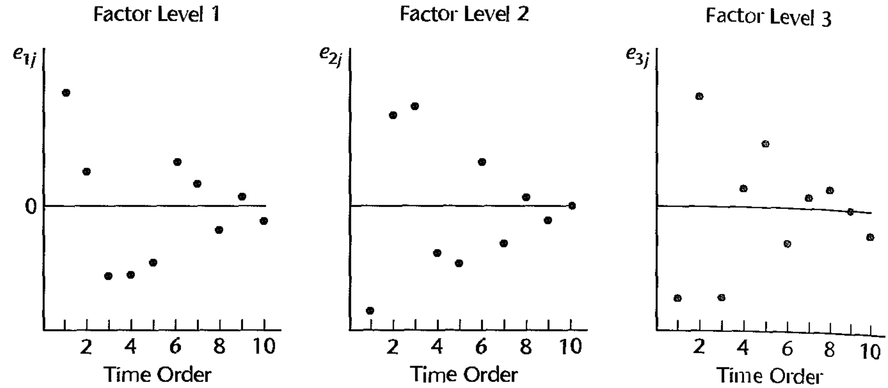



**FIGURE 18.4** **Residual Sequence Plots Illustrating Decreasing Error Variance over Time.**

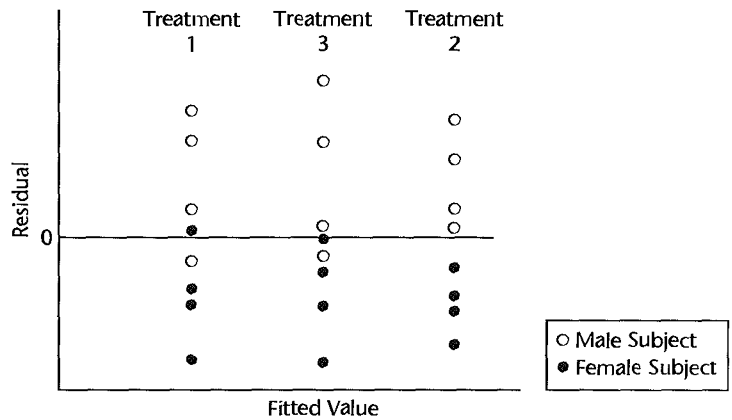



**FIGURE 18.5** **Residual Plot against Fitted Values Illustrating Omission of Important Explanatory Variable.**

*leaf plots* are particularly helpful. These plots easily reveal outlying observations, that is, observations that differ from the fitted value by far more than do other observations. As noted in Chapter 3, it is wise practice to discard outlying observations only if they can be identified as being due to such specific causes as instrumentation malfunctioning, observer measurement blunder, or recording error.

The test for outliers in regression discussed in Chapter 10 is applicable to analysis of variance as well. The appropriate Bonferroni critical value here is $t(1-\alpha/2n_T; n_T-r-1)$. If the largest absolute studentized deleted residual exceeds this critical value, that case should be considered an outlier. Note that the implicit family of tests here consists of the tests on all $n_T$ residuals for the study since we do not know in advance which case will have the largest absolute studentized deleted residual.

Occasionally, a test for an outlier is suggested in advance of the analysis, as when a substitute operator is used for one of the production runs in a manufacturing experiment. Concern about the validity of this response observation might lead to an outlier test. In this case, the Bonferroni critical value would be $t(1-\alpha/2; n_T-r-1)$.

**Omission of Important Explanatory Variables.** Residual analysis may also be used to study whether or not the single-factor ANOVA model is an adequate model. In a learning experiment involving three motivational treatments, the residuals shown in Figure 18.5 were obtained. The residual plot against the fitted values in Figure 18.5 shows no unusual overall pattern. The experimenter wondered, however, whether the treatment effects differ according to the gender of the subject. In Figure 18.5 the residuals for male subjects are shown by open circles, and those for females by dots. The results in Figure 18.5 suggest strongly that for each of the motivational treatments studied, the treatment effects do differ according to gender. Here, an analysis of covariance model, recognizing both motivational treatment and gender of subject as explanatory variables as mentioned in Chapter 15, might be more useful. Analysis of covariance models will be discussed in Chapter 22.

Note that residual analysis here does not invalidate the original single-factor model. Rather, the residual analysis points out that the original model overlooks differences in treatment effects that may be important to recognize. Since there are usually many explanatory variables that have some effect on the response, the analyst needs to identify for residual analysis those explanatory variables that most likely have an important effect on the response.

**Nonnormality of Error Terms.** The normality of the error terms can be studied from *histograms, dot plots, box plots,* and *normal probability plots* of the residuals. In addition, comparisons can be made of observed frequencies with expected frequencies if normality holds, and formal chi-square goodness of fit or related tests can be utilized. The discussion in Chapter 3 about these methods for assessing the normality of the error terms for regression is entirely applicable to ANOVA models.

When the factor level sample sizes are large, the study of normality can be made separately for each treatment. When the factor level sample sizes are small, one can combine the residuals $e_{ij}$ for all treatments into one group, provided that the evidence suggests that there are no major differences in the error variances for the treatments studied. This combining was done in the rust inhibitor example in Figure 18.1c. This figure does not indicate any serious departures from normality. The pattern of the points is reasonably linear except possibly in the tails. The coefficient of correlation between the ordered residuals and their expected values under normality is .987, which also supports the reasonableness of the normality assumption.

When unequal variances of the error terms for the different factor levels are indicated and normality must be examined for the combined data, studentized residuals (18.3) should be used, with *MSE* replaced by the sample variance $s_i^2$ in (16.39) for observations from the $i$th treatment. If ordinary residuals were used, nonnormality might be indicated solely because of the failure of the error terms to have equal variances.

### Comment

As for regression models, the ANOVA residuals $e_{ij}$ are not independent random variables. For ANOVA model (16.2), they are subject to the restrictions in (16.21). Consequently, statistical tests that require independent observations are not exactly appropriate for ANOVA residuals. If, however, the number of residuals for each factor level is not small, the effect of the correlations will only be modest. It has been noted that graphic plots of residuals are less subject to the effects of correlation than are statistical tests because graphic plots contain the individual residuals and not simply functions of them. ■

## 18.2 Tests for Constancy of Error Variance

Several formal tests are available for studying the constancy of the error variance, as required by the ANOVA model. We shall consider two of these, the Hartley test (Ref. 18.1) and the Brown-Forsythe test (Ref. 18.2). Both tests assume that independent random samples are obtained from each population. The Hartley test is simple to carry out, but is applicable only if the sample sizes are equal and if the error terms are normally distributed. The test is designed to be sensitive to substantial differences between the largest and the smallest factor level variances. The Brown-Forsythe test, discussed in Chapter 3, is slightly more difficult to compute but is more generally applicable. The test has been shown to be robust to departures from normality, and sample sizes need not be equal.

Both the Hartley test and the Brown-Forsythe test are often conducted at low $\alpha$ levels when used for testing the constancy of the error variance in the analysis of variance. The reason is that, as we shall note in Section 18.6, the $F$ test for equality of factor level means is robust against nonconstancy of the error variance when the factor level sample sizes are approximately equal, as long as the differences in the variances are not extremely large. Hence, the purpose of using the Hartley or Brown-Forsythe tests in ANOVA is often to determine whether extremely large differences in the error variances exist. For this purpose, a low $\alpha$ level may be employed since only large differences in the error variances need to be detected.

## Hartley Test

We shall describe the Hartley test in general terms. The test considers $r$ normal populations; the variance of the $i$th population is denoted by $\sigma_i^2$. Independent samples of equal size are selected from the $r$ populations; the sample variance for the $i$th population is denoted by $s_i^2$ and the common number of degrees of freedom associated with each sample variance is denoted by $df$. The alternatives to be tested are:

$$
\begin{aligned}
&H_0: \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_r^2 \\
&H_a: \text{not all } \sigma_i^2 \text{ are equal}
\end{aligned}
\tag{18.7}
$$

The Hartley test statistic, denoted by $H^*$, is based solely on the largest sample variance, denoted by $\max(s_i^2)$, and the smallest sample variance, denoted by $\min(s_i^2)$:

$$
H^* = \frac{\max\left(s_i^2\right)}{\min\left(s_i^2\right)}
\tag{18.8}
$$

Values of $H^*$ near 1 support $H_0$, and large values of $H^*$ support $H_a$. The distribution of $H^*$ when $H_0$ holds has been tabulated, and selected percentiles are presented in Table B.10. The distribution of $H^*$ depends on the number of populations $r$ and the common number of degrees of freedom $df$.

The appropriate decision rule for controlling the risk of making a Type I error at $\alpha$ is:

$$
\begin{aligned}
&\text{If } H^* \le H(1-\alpha; r, df), \text{ conclude } H_0 \\
&\text{If } H^* > H(1-\alpha; r, df), \text{ conclude } H_a
\end{aligned}
\tag{18.9}
$$

where $H(1-\alpha; r, df)$ is the $(1-\alpha)100$ percentile of the distribution of $H^*$ when $H_0$ holds, for $r$ populations and $df$ degrees of freedom for each sample variance.

When the Hartley test is used for the single-factor ANOVA model (16.2) with equal sample sizes, $n_i \equiv n$, we have $df = n - 1$. The $r$ normal populations are the normal probability distributions of the $Y$ observations for the $r$ factor levels. The sample variance

$s_i^2$ is the variance of the $n_i$ observations $Y_{ij}$ for the $i$th factor level or equivalently the variance of the $n_i$ residuals $e_{ij}$, defined in (16.39); for $n_i \equiv n$, $s_i^2$ becomes:

$$s_i^2 = \frac{\sum_{j=1}^{n}(Y_{ij} - \bar{Y}_{i\cdot})^2}{n-1} = \frac{\sum_{j=1}^{n} e_{ij}^2}{n-1} \qquad \text{when } n_i \equiv n \tag{18.10}$$

**Example**

The ABT Electronics Corporation performed an experiment to evaluate five types of flux for use in soldering printed circuit boards. A major concern of the firm's reliability engineers was the strength of the soldered joints. To test the five types of flux, 40 printed circuit boards were selected at random. Each of the five flux types was randomly assigned to 8 of the 40 circuit boards and an electronic switch was soldered to each board using the designated flux type. Following a four-week storage period, the 40 circuit boards were tested by an hydraulically operated testing machine which exerted increasing pulling force on each switch. The force (in pounds) required to break a joint, termed the pull strength, is the response of interest. This design is a completely randomized design, with eight replicates of the five treatments corresponding to the five levels of the categorical factor, flux type.

A portion of the observed pull strengths in the experiment is shown in Table 18.2, along with the estimated treatment means $\bar{Y}_{i\cdot}$ and sample variances $s_i^2$. A dot plot of these data is presented in Figure 18.6. Notice that the variability in pull strengths for the third solder type appears to be larger than for the others.

Since approximate normality is required by the Hartley test, normal probability plots of the residuals were first constructed for each treatment (not shown). The approximate normality of the residuals for each treatment was supported by the plots and by the correlation test (the correlations in the five plots are .982, .981, .977, .958, and .939; the critical value for $\alpha = .05$ from Table B.6 is .906).

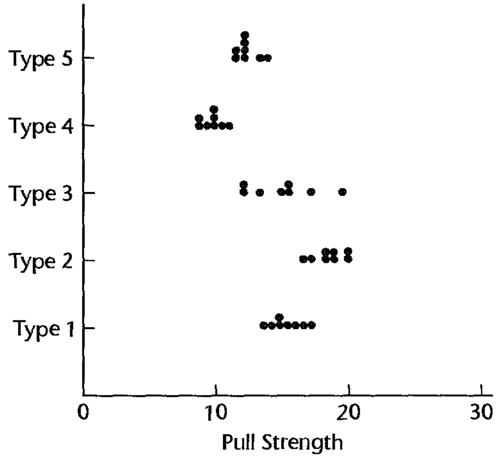
The alternatives for the Hartley test here are:

$$H_0\colon \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_5^2$$
$$H_a\colon \text{not all } \sigma_i^2 \text{ are equal}$$

**TABLE 18.2** **Solder Joint Pull Strengths—ABT Electronics Example.**

| Joint | Flux Type ($i$) | | | | |
|---|---|---|---|---|---|
| $j$ | $i=1$ | $i=2$ | $i=3$ | $i=4$ | $i=5$ |
| 1 | 14.87 | 18.43 | 16.95 | 8.59 | 11.55 |
| 2 | 16.81 | 18.76 | 12.28 | 10.90 | 13.36 |
| ... | ... | ... | ... | ... | ... |
| 7 | 17.40 | 17.16 | 19.35 | 9.41 | 12.05 |
| 8 | 14.62 | 16.40 | 15.52 | 10.04 | 11.95 |
| $\bar{Y}_{i\cdot}$ | 15.420 | 18.528 | 15.004 | 9.741 | 12.340 |
| $\tilde{Y}_i$ | 15.170 | 18.595 | 15.255 | 10.010 | 12.105 |
| $s_i^2$ | 1.531 | 1.570 | 6.183 | .667 | .592 |

**FIGURE 18.6 Dot Plots of Pull Strengths—ABT Electronics Example.**




For level of significance $\alpha = .05$, $r = 5$, and $df = 8 - 1 = 7$, we require $H(.95; 5, 7) = 9.70$. Hence the appropriate decision rule is:

$$\text{If } H^* \leq 9.70, \text{ conclude } H_0$$
$$\text{If } H^* > 9.70, \text{ conclude } H_a$$

From Table 18.2 we see that $\max(s_i^2) = 6.183$ and $\min(s_i^2) = .592$. Hence the test statistic is:

$$H^* = \frac{6.183}{.592} = 10.44$$

Since $H^* = 10.44 > 9.70$, we conclude $H_a$, that the five treatment variances are not equal.

### Comments

1. The Hartley test strictly requires equal sample sizes. If the sample sizes are unequal but do not differ greatly, the Hartley test may still be used as an approximate test. For this purpose, the average number of degrees of freedom would be used for entering Table B.10.

2. The Hartley test is quite sensitive to departures from the assumption of normal populations and should not be used when substantial departures from normality exist. ■

## Brown-Forsythe Test

The Brown-Forsythe test for constancy of the error variance in regression was discussed in Chapter 3. The test was originally developed for use in ANOVA applications and is more general than its use for regression described in Chapter 3. The Brown-Forsythe test, just like the Hartley test, can be used to study the equality of $r$ population variances. Unlike the Hartley test, the Brown-Forsythe test is robust against departures from normality, which often occur together with unequal variances. Also, the Brown-Forsythe test does not require equal sample sizes.

To test the alternatives in (18.7) using the Brown-Forsythe test, we first compute the absolute deviations of the $Y_{ij}$ observations about their respective factor level medians $\tilde{Y}_i$:

$$d_{ij} = |Y_{ij} - \tilde{Y}_i| \tag{18.11}$$

The Brown-Forsythe test then determines whether or not the expected values of the absolute deviations for the $r$ treatments are equal. If the $r$ error variances $\sigma_i^2$ are equal, so will the expected values of the absolute deviations be equal. Unequal error variances imply differing expected values of the absolute deviations. The Brown-Forsythe test statistic is simply the ordinary $F^*$ statistic in (16.55) for testing differences in the treatment means, but now based on the absolute deviations $d_{ij}$ in (18.11):

$$F_{BF}^* = \frac{MSTR}{MSE} \tag{18.12}$$

where:

$$MSTR = \frac{\sum n_i(\bar{d}_{i\cdot} - \bar{d}_{\cdot\cdot})^2}{r-1} \tag{18.12a}$$

$$MSE = \frac{\sum\sum(d_{ij} - \bar{d}_{i\cdot})^2}{n_T - r} \tag{18.12b}$$

$$\bar{d}_{i\cdot} = \frac{\sum_j d_{ij}}{n_i} \tag{18.12c}$$

$$\bar{d}_{\cdot\cdot} = \frac{\sum\sum d_{ij}}{n_T} \tag{18.12d}$$

If the error terms have constant variance and the factor level sample sizes are not extremely small, $F_{BF}^*$ follows approximately an $F$ distribution with $r - 1$ and $n_T - r$ degrees of freedom. Large $F_{BF}^*$ values indicate that the error terms do not have constant variance.

**Example**

Table 18.2 for the ABT Electronics Corporation example provides the sample medians $\tilde{Y}_i$ for the five treatments. The absolute deviations $d_{ij}$ in (18.11) are shown in Table 18.3. We illustrate their calculation for $d_{11}$:

$$d_{11} = |Y_{11} - \tilde{Y}_1| = |14.87 - 15.170| = .300$$

The $F_{BF}^*$ test statistic (18.12) based on the absolute deviations is obtained in the usual manner; it is $F_{BF}^* = 2.94$. For $\alpha = .05$, we require $F(.95; 4, 35) = 2.64$. Since $F_{BF}^* = 2.94 > 2.64$, we conclude $H_a$, that the error terms do not have constant variance. The $P$-value for this test is .034.

**TABLE 18.3** Absolute Deviations of Responses from Treatment Medians—ABT Electronics Example.

| Joint | Flux Type ($i$) | | | | |
|---|---|---|---|---|---|
| $j$ | $i = 1$ | $i = 2$ | $i = 3$ | $i = 4$ | $i = 5$ |
| 1 | .300 | .165 | 1.695 | 1.420 | .555 |
| 2 | 1.640 | .165 | 2.975 | .890 | 1.255 |
| ... | ... | ... | ... | ... | ... |
| 7 | 2.230 | 1.435 | 4.095 | .600 | .055 |
| 8 | .550 | 2.195 | .265 | .030 | .155 |

## 18.3 Overview of Remedial Measures

In the remainder of this chapter, we consider three remedial measures for two common departures from ANOVA model (16.2)—nonconstancy of the error variance and nonnormality of the distribution of the error terms.

1. If the error terms are normally distributed but the variance of the error terms is not constant, a standard remedial measure is to use weighted least squares. We have already considered weighted least squares for nonconstancy of the error variance in regression models. These weighted least squares procedures for regression carry over directly to analysis of variance models.

2. Often, nonconstancy of the error variance is accompanied by nonnormality of the error term distribution. A standard remedial measure here is to transform the response variable $Y$. We shall present two approaches to finding an appropriate transformation to make the error distribution more nearly normal and to help stabilize the variance of the error terms—some simple guides and the Box-Cox procedure. The latter was considered in Chapter 3 for regression models and is directly applicable to analysis of variance models.

3. When there are major departures from ANOVA model (16.2) and transformations are not successful in stabilizing the error variance and bringing the error distribution close to normal, a nonparametric test for the equality of the factor level means may be used instead of the standard $F$ test. We shall consider a nonparametric test that is based on the ranks of the $Y$ observations.

We begin our discussion of remedial measures with weighted least squares.

## 18.4 Weighted Least Squares

When the errors $\varepsilon_{ij}$ are normally distributed but their variances are not the same for the different factor levels, cell means model (16.2) becomes:

$$Y_{ij} = \mu_i + \varepsilon_{ij} \tag{18.13}$$

where $\varepsilon_{ij}$ are independent $N(0, \sigma_i^2)$.

Weighted least squares is a standard remedial measure here, just as for the comparable situation in regression. In fact, we shall use the regression approach to the analysis of variance for implementing weighted least squares. All of the earlier discussion on weighted least squares for regression is applicable to the analysis of variance.

Since the factor level variances $\sigma_i^2$ are usually unknown, they must be estimated. This is ordinarily done by means of the sample variances $s_i^2$ in (16.39), in which case the weight $w_{ij}$ for the $j$th case of the $i$th factor level is:

$$w_{ij} = \frac{1}{s_i^2} \tag{18.14}$$

The test for the equality of the factor level means in (16.54) is now conducted by the general linear test approach described in Chapter 2. The full model is fitted, using the weights in (18.14), and the error sum of squares is obtained, now denoted by $SSE_w(F)$. Next, the reduced model under $H_0$ is fitted and the error sum of squares $SSE_w(R)$ is obtained. Test statistic (2.70) is employed, as usual. We shall see that $df_F = n_T - r$ and $df_R = n_T - 1$. Hence, the general linear test statistic here is:

$$F_w^* = \frac{SSE_w(R) - SSE_w(F)}{r - 1} \div \frac{SSE_w(F)}{n_T - r} \tag{18.15}$$

Since the weights are based on the estimated variances $s_i^2$, the distribution of $F_w^*$ under $H_0$ is only approximately an $F$ distribution with $r - 1$ and $n_T - r$ degrees of freedom. When the factor level sample sizes are reasonably large, the approximation generally is satisfactory. As explained in Chapter 11, bootstrapping can be employed to examine the effect of using estimated weights.

**Example**

Recall in the ABT Electronics example that the normality assumption appears to be reasonably well supported by the data, but the error variance is not constant. Weighted least squares will now be used to test the alternatives:

$$\begin{aligned} &H_0\colon \mu_1 = \mu_2 = \cdots = \mu_5 \\ &H_a\colon \text{not all } \mu_i \text{ are equal} \end{aligned} \tag{18.16}$$

The weights will be based on the sample variances in Table 18.2:

$$w_{1j} = \frac{1}{1.531} = .653 \qquad w_{2j} = \frac{1}{1.570} = .637 \qquad w_{3j} = \frac{1}{6.183} = .162$$

$$w_{4j} = \frac{1}{.667} = 1.499 \qquad w_{5j} = \frac{1}{.592} = 1.689$$

We shall use regression model (16.85) to represent cell means model (18.13):

$$Y_{ij} = \mu_1 X_{ij1} + \mu_2 X_{ij2} + \cdots + \mu_5 X_{ij5} + \varepsilon_{ij} \qquad \text{Full model} \tag{18.17}$$

where:

$$X_1 = \begin{cases} 1 & \text{if case from factor level 1} \\ 0 & \text{otherwise} \end{cases}$$

$$\vdots$$

$$X_5 = \begin{cases} 1 & \text{if case from factor level 5} \\ 0 & \text{otherwise} \end{cases}$$

Note that the factor level means $\mu_i$ play the role of regression coefficients and that the regression model has no intercept.

Table 18.4 repeats from Table 18.2 a portion of the experimental data in column 1 and contains the coding of the indicator variables in columns 2–6 and the weights in column 7. Note, for instance, that the coding for cases from the first treatment is $X_1 = 1$, $X_2 = 0$, $X_3 = 0$, $X_4 = 0$, and $X_5 = 0$, and similarly for cases from the other treatments.

Figure 18.7a contains the MINITAB output when $Y$ in column 1 of Table 18.4 is regressed on $X_1$, $X_2$, $X_3$, $X_4$, and $X_5$ in columns 2–6, using the weights in column 7 and specifying no intercept. We see that $SSE_w(F) = 35.0$.

The reduced model under $H_0$ is given by (16.86):

$$Y_{ij} = \mu_c + \varepsilon_{ij} \qquad \text{Reduced model} \tag{18.18}$$

**TABLE 18.4** Data for Weighted Least Squares Regression—ABT Electronics Examp

| | | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|---|---|---|---|---|---|---|---|---|---|
| | | | Full Model | | | | | Weights | Reduced M |
| $i$ | $j$ | $Y_{ij}$ | $X_{ij1}$ | $X_{ij2}$ | $X_{ij3}$ | $X_{ij4}$ | $X_{ij5}$ | $w_{ij}$ | $X_{ij}$ |
| 1 | 1 | 14.87 | 1 | 0 | 0 | 0 | 0 | .653 | 1 |
| 1 | 2 | 16.81 | 1 | 0 | 0 | 0 | 0 | .653 | 1 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 1 | 7 | 17.40 | 1 | 0 | 0 | 0 | 0 | .653 | 1 |
| 1 | 8 | 14.62 | 1 | 0 | 0 | 0 | 0 | .653 | 1 |
| 2 | 1 | 18.43 | 0 | 1 | 0 | 0 | 0 | .637 | 1 |
| 2 | 2 | 18.76 | 0 | 1 | 0 | 0 | 0 | .637 | 1 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 5 | 7 | 12.05 | 0 | 0 | 0 | 0 | 1 | 1.689 | 1 |
| 5 | 8 | 11.95 | 0 | 0 | 0 | 0 | 1 | 1.689 | 1 |

**FIGURE 18.7** MINITAB Weighted Regression Output for Full and Reduced Models—ABT Electronics Example.

(a) Full Model

```
The regression equation is
Y = 15.4 X1 + 18.5 X2 + 15.0 X3 + 9.74 X4 + 12.3 X5

Predictor        Coef      Stdev    t-ratio       p
Noconstant
X1            15.4200     0.4375      35.24   0.000
X2            18.5275     0.4430      41.82   0.000
X3            15.0037     0.8785      17.08   0.000
X4             9.7413     0.2888      33.73   0.000
X5            12.3400     0.2721      45.36   0.000

Analysis of Variance

SOURCE       DF        SS        MS          F        p
Regression    5    6478.7    1295.7    1295.56    0.000
Error        35      35.0       1.0
Total        40    6513.7
```

(b) Reduced Model

```
The regression equation is
Y = 12.9 X

Predictor        Coef      Stdev    t-ratio       p
Noconstant
X             12.8764     0.4981      25.85   0.000

Analysis of Variance

SOURCE       DF        SS        MS          F        p
Regression    1    6154.5    6154.5     668.28    0.000
Error        39     359.2       9.2
Total        40    6513.7
```

where $\mu_c$ is the common mean response under $H_0$. The corresponding regression model is:

$$Y_{ij} = \mu_c X_{ij} + \varepsilon_{ij} \tag{18.19}$$

where $X_{ij} \equiv 1$. Note that regression model (18.19) has no intercept.

The new $X$ variable is shown in Table 18.4, column 8. Regressing $Y$ in column 1 on $X$ in column 8, using the weights in column 7 and specifying no intercept, leads to the MINITAB output in Figure 18.7b. We see that $SSE_w(R) = 359.2$. We have $n_T - 1 = 40 - 1 = 39$ and $n_T - r = 40 - 5 = 35$. Hence, test statistic (18.15) is:

$$F_w^* = \frac{359.2 - 35.0}{39 - 35} \div \frac{35.0}{35} = 81.05$$

For $\alpha = .01$, we require $F(.99; 4, 35) = 3.908$. Since $F^* = 81.05 > 3.908$, the approximate $F$ test leads to conclusion $H_a$, that the factor level means differ. The approximate $P$-value of the test is 0+.

### Comments

1. The weighted least squares estimates of the factor level means $\mu_i$ are always the estimated factor level means $\bar{Y}_{i\cdot}$, as may be seen by comparing the estimated regression coefficients in Figure 18.7a with the estimated factor level means in Table 18.2. Hence, for ANOVA model (18.13), the weighted and ordinary least squares estimates of the factor level means $\mu_i$ are the same.

2. When the sample variances $s_i^2$ are used as weights, the error sum of squares for the fit of full model (18.17) will always be $SSE_w(F) = n_T - r$. Note that in our example $SSE_w(F) = 35.0$ and $n_T - r = 40 - 5 = 35$.

3. Some analysis of variance computer packages have an option for weighted least squares, with the user specifying the weights. ■

## 18.5 Transformations of Response Variable

When both the model assumptions of constancy of the error variance and normality of the error distributions are violated, a transformation of the response variable is often useful. We describe now two approaches to finding a useful transformation—some simple guides and the Box-Cox procedure.

### Simple Guides to Finding a Transformation

The following are four simple guides to finding a useful transformation. The guides were developed from theoretical considerations to stabilize the error variances, but these transformations often also are helpful in bringing the distribution of the error terms more closely to a normal distribution.

**Variance Proportional to $\mu_i$.** When the variance of the error terms for each factor level (denoted by $\sigma_i^2$) is proportional to the factor level mean $\mu_i$, a square root transformation is helpful:

$$\text{If } \sigma_i^2 \text{ proportional to } \mu_i: \qquad Y' = \sqrt{Y} \qquad \text{or} \qquad Y' = \sqrt{Y} + \sqrt{Y+1} \tag{18.20}$$

This type of situation is often found when the observed variable $Y$ is a count, such as the number of attempts by a subject before the correct solution is found.

**Standard Deviation Proportional to $\mu_i$.** When the standard deviation of the error terms for each factor level is proportional to the factor level mean, a helpful transformation is the logarithmic transformation:

$$\text{If } \sigma_i \text{ proportional to } \mu_i: \qquad Y' = \log Y \tag{18.21}$$

**Standard Deviation Proportional to $\mu_i^2$.** When the error term standard deviation is proportional to the square of the factor level mean for the different factor levels, an appropriate transformation is the reciprocal transformation:

$$\text{If } \sigma_i \text{ proportional to } \mu_i^2: \qquad Y' = \frac{1}{Y} \tag{18.22}$$

**Response Is a Proportion.** At times, the observed variable $Y_{ij}$ is a proportion $p_{ij}$. For instance, the treatments may be different training procedures, the unit of observation is a company training class, and the observed variable $Y_{ij}$ is the proportion of employees in the $j$th class for the $i$th training procedure who benefited substantially by the training. Note that $n_i$ here refers to the number of classes receiving the $i$th training procedure, not to the number of students.

It is well known that for the binomial distribution the variance of the sample proportion depends on the true proportion. When the number of cases on which each sample proportion is based is the same, this variance is:

$$\sigma^2\{p_{ij}\} = \frac{\pi_i(1-\pi_i)}{m} \tag{18.23}$$

Here $\pi_i$ denotes the population proportion for the $i$th treatment and $m$ is the common number of cases on which each sample proportion is based. Since $\sigma^2\{p_{ij}\}$ depends on the treatment proportion $\pi_i$, the variances of the error terms will not be stable if the treatment proportions $\pi_i$ differ. An appropriate transformation for this case is the arcsine transformation:

$$\text{If response is a proportion:} \qquad Y' = 2 \arcsin \sqrt{Y} \tag{18.24}$$

When the proportions $p_{ij}$ are based on different numbers of cases (for instance, in our earlier illustration there may be different numbers of employees in each training class), transformation (18.24) should be employed together with a weighted least squares analysis as described in Section 18.4. The use of the arcsin transformation when the response is a proportion can be an effective, yet simple, remedial measure. A more rigorous approach would involve the use of logistic regression as discussed in Chapter 14.

**Use of Simple Guides.** To examine whether one of the simple transformation guides is applicable, the statistics $s_i^2/\bar{Y}_{i\cdot}$, $s_i/\bar{Y}_{i\cdot}$, and $s_i/\bar{Y}_{i\cdot}^2$ should be calculated for each factor level, where $s_i^2$ is the sample variance of the $Y$ observations for the $i$th factor level, defined in (16.39). Approximate constancy of one of the three statistics over all factor levels would suggest the corresponding transformation as useful for stabilizing the error variance and making the error distributions more nearly normal.

**xample**

Servo-Data, Inc., operates mainframe computers at three different locations. The computers are identical as to make and model, but are subject to different degrees of voltage fluctuation

LE 18.5 e between puter ures at tions (in )— o-Data ple.

| Failure Interval | Location ($i$) | | | | | |
|---|---|---|---|---|---|---|
| | 1 | | 2 | | 3 | |
| $j$ | $Y_{1j}$ | $R_{1j}$ | $Y_{2j}$ | $R_{2j}$ | $Y_{3j}$ | $R_{3j}$ |
| 1 | 4.41 | 2 | 8.24 | 4 | 106.19 | 14 |
| 2 | 100.65 | 13 | 81.16 | 11 | 33.83 | 7 |
| 3 | 14.45 | 6 | 7.35 | 3 | 78.88 | 10 |
| 4 | 47.13 | 9 | 12.29 | 5 | 342.81 | 15 |
| 5 | 85.21 | 12 | 1.61 | 1 | 44.33 | 8 |

| $i$ | $\bar{Y}_{i\cdot}$ | $s_i^2$ | $i$ | $\bar{R}_{i\cdot}$ | $s_i^2$ |
|---|---|---|---|---|---|
| 1 | 50.4 | 1,789 | 1 | 8.4 | 20.3 |
| 2 | 22.1 | 1,103 | 2 | 4.8 | 14.2 |
| 3 | 121.2 | 16,167 | 3 | 10.8 | 12.7 |
| | $\bar{Y}_{\cdot\cdot} = 64.6$ | | | $\bar{R}_{\cdot\cdot} = 8.00$ | |

in the power lines serving the respective installations. Table 18.5 contains the lengths of time between computer failures $Y_{ij}$ for the three locations, for five failure intervals each. The table also contains the ranks $R_{ij}$ (from 1 to 15) for $Y_{ij}$, which we shall use in Section 18.7 for nonparametric analysis. Even though the sample sizes are small, the data suggest highly skewed distributions having nonconstant error variance. This is an observational study because no randomization of treatments to experimental units occurred.

To study whether one of the simple guides is helpful here, we have calculated the following statistics based on the results in Table 18.5.

| $i$ | $\frac{s_i^2}{\bar{Y}_{i\cdot}}$ | $\frac{s_i}{\bar{Y}_{i\cdot}}$ | $\frac{s_i}{\bar{Y}_{i\cdot}^2}$ |
|---|---|---|---|
| 1 | 35.5 | .84 | .017 |
| 2 | 49.9 | 1.50 | .068 |
| 3 | 133.4 | 1.05 | .009 |

The relation $s_i/\bar{Y}_{i\cdot}$ is the most stable, hence the logarithmic transformation (18.21) may be helpful here. We shall continue this example after discussing the use of the Box-Cox procedure for finding an appropriate transformation in the analysis of variance.

## Box-Cox Procedure

The Box-Cox transformation procedure was described in Chapter 3 for regression. As noted there, the Box-Cox procedure identifies a power transformation of the type $Y^\lambda$ to correct for both lack of normality and nonconstancy of the error variance. The procedure is entirely applicable to the analysis of variance. As for regression, the numerical search procedure for ANOVA models considers different values of the parameter $\lambda$. For each value of $\lambda$, the $Y$ observations are transformed according to (3.36) and ANOVA model (16.2) is fitted and the error sum of squares $SSE$ is obtained. The value of $\lambda$ that minimizes $SSE$ is the maximum likelihood estimate of $\lambda$. As we saw in regression, $SSE$ as a function of $\lambda$ is often flat in the neighborhood of the maximum likelihood estimate $\hat{\lambda}$, so that a meaningful value of $\lambda$ in the neighborhood may be chosen for the transformation in preference to the maximum likelihood value.

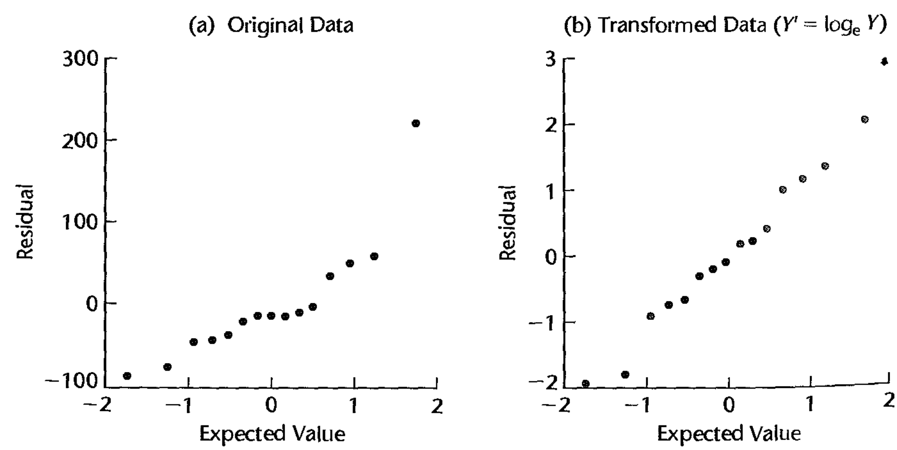
**Example**

The Box-Cox procedure was applied in the Servo-Data example of Table 18.5 by using 21 equally spaced values of $\lambda$ between $-1$ and 1. For each value of $\lambda$, the $Y$ observations were transformed according to (3.36) and $SSE$ for ANOVA model (16.2) was calculated. A portion of the results is shown in Table 18.6. The smallest $SSE$ is obtained with $\lambda = .1$. However, note that $SSE$ does not change much between $-.10$ and .20. Hence, the parameter $\lambda = 0$ may be preferred because it leads to the meaningful logarithmic transformation. This is also the transformation selected according to the simple guides. Normal probability plots of the residuals for the original and transformed data ($Y' = \log_e Y$) are shown in Figure 18.8. The normality assumption appears to be much more reasonable for the transformed data ($r = .991$). Also, the variances of the transformed data are much more stable now ($s_1^2 = 1.742$, $s_2^2 = 1.974$, $s_3^2 = .817$) as compared to the variances for the original data in Table 18.5.

**TABLE 18.6** Calculations for Box-Cox Procedure—Servo-Data Example.

| $\lambda$ | $SSE$ (in thousands) | $\lambda$ | $SSE$ (in thousands) |
|---|---|---|---|
| −1.0 | 203.7 | .10 | 15.3 |
| −.80 | 95.1 | .20 | 15.6 |
| −.60 | 48.7 | .40 | 18.7 |
| −.40 | 28.3 | .60 | 26.4 |
| −.20 | 19.2 | .80 | 42.6 |
| −.10 | 17.0 | 1.0 | 76.2 |
| .00 | 15.7 | | |

**FIGURE 18.8** Normal Probability Plots for Original and Transformed Data—Servo-Data Example.



A single factor ANOVA was performed on $Y'$, the logarithm of the $Y$ observations. The resulting $F$ test for equality of treatment means was:

$$F^* = \frac{MSTR}{MSE} = \frac{5.7264}{1.5112} = 3.789$$

For $\alpha = .10$, we require $F(.90; 2, 12) = 2.81$. Since $F^* = 3.789 > 2.81$, we conclude $H_a$, that the three means are not equal. The $P$-value of the test is .053. The transformed means for the three groups are 3.413, 2.797, and 4.437, respectively. The Bonferroni pairwise comparison procedure was then conducted at the .10 level, with $s^2\{\hat{D}\} = .6045$, $s\{\hat{D}\} = .7775$, $B = t(.9833; 12) = 2.402$, and $Bs\{\hat{D}\} = 1.868$. The resulting 90 percent Bonferroni pairwise confidence intervals are:

$$-2.984 \leq \mu_2 - \mu_1 \leq .752$$
$$-.884 \leq \mu_3 - \mu_1 \leq 2.892$$
$$.272 \leq \mu_3 - \mu_2 \leq 4.008$$

Therefore, we conclude that location 3 has longer average time computer failures than location 2.

### Comments

1. It is wise policy, as mentioned for regression, to check the residuals after a transformation has been applied to make sure that the transformation has been effective in both stabilizing the variances and making the distribution of the error terms reasonably normal.

2. When a transformation of the observations is required, one can work completely with the transformed data for testing the equality of factor level means. On the other hand, it is often desirable when making estimates of factor level effects to change a confidence interval based on the transformed variable back to an interval in the original variable for easier understanding of the significance of the results.

3. The variance stabilizing transformations (18.20), (18.21), (18.22), and (18.24) are obtained by using a Taylor series expansion for the variance of $Y$. An explanation of the approach may be found in Reference 18.3. ■

## 18.6 Effects of Departures from Model

In preceding sections, we considered how residual analysis and other diagnostic techniques can be helpful in assessing the appropriateness of the ANOVA model for the data at hand. We also discussed the use of transformations for both stabilizing the variance and obtaining an error distribution more nearly normal. The question now arises: what are the effects of any remaining departures from the model on the inferences made? A thorough review of the many studies investigating these effects has been made by Scheffé (Ref. 18.4). Here, we summarize the findings.

### Nonnormality

For the fixed ANOVA model I, lack of normality is not an important matter, provided the departure from normality is not extreme. It may be noted in this connection that *kurtosis* of the error distribution (either more or less peaked than a normal distribution) is more important than skewness of the distribution in terms of the effects on inferences.

The point estimators of factor level means and contrasts are unbiased whether or not the populations are normal. The $F$ test for the equality of factor level means is but little affected by lack of normality, either in terms of the level of significance or power of the test. Hence, the $F$ test is a *robust* test against departures from normality. For instance, while the specified level of significance might be .05, the actual level for a nonnormal error distribution might be .04 or .065. Typically, the achieved level of significance in the presence of nonnormality is slightly higher than the specified one, and the achieved power of the test is slightly less than the calculated one. Single interval estimates of factor level means and contrasts and the Scheffé multiple comparison procedure also are not much affected by lack of normality, provided that the sample sizes are not extremely small.

For the random ANOVA model II (to be discussed in Chapter 25), lack of normality has more serious implications. The estimators of the variance components are still unbiased, but the actual confidence coefficient for interval estimates may be substantially different from the specified one.

## Unequal Error Variances

When the error variances are unequal, the $F$ test for the equality of means with the fixed ANOVA model is only slightly affected if all factor level sample sizes are equal or do not differ greatly. Specifically, unequal error variances then raise the actual level of significance slightly higher than the specified level. Similarly, the Scheffé multiple comparison procedure based on the $F$ distribution is not affected to any substantial extent by unequal variances when the sample sizes are equal or are approximately the same. Thus, the $F$ test and related analyses are robust against unequal variances when the sample sizes are approximately equal. Single comparisons between factor level means, on the other hand, can be substantially affected by unequal variances, so that the actual and specified confidence coefficients may differ markedly in these cases.

The use of equal sample sizes for all factor levels not only tends to minimize the effects of unequal variances on inferences with the $F$ distribution but also simplifies calculational procedures. Thus, here at least, simplicity and robustness go hand in hand.

For the random ANOVA model II, unequal error variances can have pronounced effects on inferences about the variance components, even with equal sample sizes.

## Nonindependence of Error Terms

Lack of independence of the error terms can have serious effects on inferences in the analysis of variance, for both fixed and random ANOVA models. Since this defect is often difficult to correct, it is important to prevent it in the first place whenever feasible. The use of randomization in those stages of a study that are likely to lead to correlated error terms can be a most important insurance policy. In the case of observational data, however, randomization may not be possible. Here, in the presence of correlated error terms, it may be possible to modify the model. For instance, in the earlier discussion based on Figure 18.3, we noted that inclusion in the model of a linear term for the learning effect of the analyst might remove the correlation of the error terms.

Modification of the model because of correlated error terms may also be necessary in experimental studies. In one case, the experimenter asked each of 10 subjects to give ratings to four new flavors of a fruit syrup and to the standard flavor, on a scale from 0 to 100. When the single-factor analysis of variance model was applied, the experimenter found high degrees of correlation in the residuals for each subject. The experimenter thereupon modified the model to a repeated measures design model (Chapter 27). As described in Chapter 15, this latter type of model is intended for situations where the same subject is given each of the different treatments and differences between subjects are expected.

## 18.7 Nonparametric Rank $F$ Test

When transformations are not successful in bringing the distributions of the error terms close enough to normality to meet the robustness properties of the standard inference procedures, a nonparametric inference procedure can be useful. Nonparametric procedures do not depend on the distribution of the error terms; often the only requirement is that the distribution is continuous. The nonparametric procedure considered here assumes that the $r$ populations under study are continuous distributions that differ only with respect to location. Thus it provides a test for differences in population means or medians, assuming that the shapes of the populations (i.e., variances, skewness, kurtosis, etc.) are identical.

The test procedure is very simple. All $n_T$ observations are ranked from 1 to $n_T$ in ascending order. Then, the usual $F^*$ test statistic in (16.55) is calculated, but now based on the ranks, and the $F$ test is carried out in the ordinary manner.

### Test Procedure

The $Y_{ij}$ observations first need to be ranked in ascending order from 1 to $n_T$. We shall let $R_{ij}$ denote the rank of $Y_{ij}$. In the case of ties among some observations, each of the tied observations is given the mean of the ranks involved. For instance, if two observations are tied for what would otherwise have been the third- and fourth-ranked positions, each would be given the mean value 3.5.

To test whether the treatment means are equal, the usual $F^*$ test statistic is obtained based on the ranks $R_{ij}$. This test statistic is now denoted by $F_R^*$:

$$F_R^* = \frac{MSTR}{MSE} \tag{18.25}$$

where:

$$MSTR = \frac{\sum n_i(\bar{R}_{i\cdot} - \bar{R}_{\cdot\cdot})^2}{r-1} \tag{18.25a}$$

$$MSE = \frac{\sum\sum(R_{ij} - \bar{R}_{i\cdot})^2}{n_T - r} \tag{18.25b}$$

$$\bar{R}_{i\cdot} = \frac{\sum_j R_{ij}}{n_i} \tag{18.25c}$$

$$\bar{R}_{\cdot\cdot} = \frac{\sum\sum R_{ij}}{n_T} = \frac{(n_T+1)}{2} \tag{18.25d}$$

Note that $\bar{R}_{\cdot\cdot}$, the overall mean of the ranks, is a constant for any given total number of cases $n_T$.

When the treatment means are the same, test statistic $F_R^*$ follows approximately the $F(r-1, n_T - r)$ distribution provided that the sample sizes $n_i$ are not very small. To test the alternatives:

$$\begin{aligned} &H_0\colon \mu_1 = \mu_2 = \cdots = \mu_r \\ &H_a\colon \text{not all } \mu_i \text{ are equal} \end{aligned} \tag{18.26a}$$

the appropriate decision rule to control the Type I error at $\alpha$ is:

$$\begin{aligned} &\text{If } F_R^* \leq F(1-\alpha; r-1, n_T - r), \text{ conclude } H_0 \\ &\text{If } F_R^* > F(1-\alpha; r-1, n_T - r), \text{ conclude } H_a \end{aligned} \tag{18.26b}$$

**Example**

In the Servo-Data example of Table 18.5, we noted earlier that the logarithmic transformation of $Y$ improves considerably the appropriateness of the assumptions of normality and constancy of the error variance. If the search for a transformation of $Y$ had not been successful, or as an alternative to the transformation approach, we could use the nonparametric rank $F$ test. To use this test, we first rank the data in Table 18.5 from 1 to 15. The ranks are shown in Table 18.5. Note, incidentally, from Table 18.5 that the rank transformation has helped to stabilize the variances of the transformed observations (i.e., the ranks) for the three treatments. We now calculate $SSTR$ and $SSE$ as follows:

$$SSTR = 5[(8.4 - 8.0)^2 + (4.8 - 8.0)^2 + (10.8 - 8.0)^2] = 91.20$$

$$SSE = (2 - 8.4)^2 + (13 - 8.4)^2 + \cdots + (8 - 10.8)^2 = 188.80$$

Note that the overall mean $\bar{R}_{..}$ here is $(n_T + 1)/2 = (15 + 1)/2 = 8.0$. The $F_R^*$ test statistic is therefore:

$$F_R^* = \frac{91.20}{3-1} \div \frac{188.8}{15-3} = 2.90$$

For $\alpha = .10$, we require $F(.90; 2, 12) = 2.81$. Since $F_R^* = 2.90 > 2.81$, we conclude $H_a$. The $P$-value of the test is .094.

Recall that when we conducted the standard $F$ test based on the logarithmic transformation of $Y$, which was suggested both by the simple guides and the Box-Cox procedure, we found that it leads to the same conclusion here; but its $P$-value—.053—is considerably smaller. Thus, both tests show that the mean time between computer failures differs for the three locations.

### Comment

The *Kruskal-Wallis test* (Ref. 18.5), a widely used nonparametric test for testing the equality of treatment means, is based on a test statistic that is equivalent to the rank $F$ test statistic. The Kruskal-Wallis test statistic, denoted by $X_{KW}^2$, is also based on the ranks $R_{ij}$ from 1 to $n_T$ and is defined as follows:

$$X_{KW}^2 = \frac{SSTR}{\dfrac{SSTO}{n_T - 1}} \tag{18.27}$$

where:

$$SSTO = \sum\sum (R_{ij} - \bar{R}_{..})^2 \tag{18.27a}$$

Instead of using the $F$ distribution approximation, the Kruskal-Wallis test uses a chi-square distribution approximation. If the $n_i$ are reasonably large (five or more is the usual advice), $X_{KW}^2$ is approximately a $\chi^2$ random variable with $r - 1$ degrees of freedom when all treatment means are equal. The decision rule therefore is:

$$\begin{aligned}&\text{If } X_{KW}^2 \leq \chi^2(1-\alpha; r-1), \text{ conclude } H_0\\&\text{If } X_{KW}^2 > \chi^2(1-\alpha; r-1), \text{ conclude } H_a\end{aligned} \tag{18.28}$$

The $F_R^*$ and $X_{KW}^2$ test statistics are equivalent, being related as follows:

$$F_R^* = \frac{(n_T - r)X_{KW}^2}{(r-1)(n_T - 1 - X_{KW}^2)} \tag{18.29}$$

■

## ultiple Pairwise Testing Procedure

If the rank $F$ test (or the Kruskal-Wallis test) leads to the conclusion that the factor level means $\mu_i$ are not equal, it is frequently desired to obtain information about the comparative magnitudes of these means based on the ranked data. A large-sample testing analogue of the Bonferroni pairwise comparison procedure discussed in Section 17.7, based on the ranks of the observations, may be employed for this purpose, provided that the sample sizes are not too small. Testing limits for all $g = r(r-1)/2$ pairwise tests using the mean ranks $\bar{R}_{i\cdot}$ are set up as follows for family level of significance $\alpha$:

$$(\bar{R}_{i\cdot} - \bar{R}_{i'\cdot}) \pm B\left[\frac{n_T(n_T+1)}{12}\left(\frac{1}{n_i} + \frac{1}{n_{i'}}\right)\right]^{1/2} \tag{18.30}$$

where:

$$B = z(1 - \alpha/2g) \tag{18.30a}$$

$$g = \frac{r(r-1)}{2} \tag{18.30b}$$

If the testing limits include zero, we conclude that the corresponding treatment means $\mu_i$ and $\mu_{i'}$ do not differ. If the testing limits do not include zero, we conclude that the two corresponding treatment means differ. On the basis of all pairwise tests, we then set up groups of treatment means whose members do not differ according to the simultaneous testing procedure. In this way, we obtain information about the comparative magnitudes of the treatment means $\mu_i$.

**Example**

For the Servo-Data example in Table 18.5, we wish to ascertain, if possible, which location has the longest mean time between computer failures based on the rank data. For a family significance level of $\alpha = .10$ and $g = r(r-1)/2 = 3(2)/2 = 3$ pairwise tests, we require $B = z(.9833) = 2.13$. Since all treatment sample sizes are equal, we need to calculate the right term in (18.30) only once:

$$B\left[\frac{n_T(n_T+1)}{12}\left(\frac{1}{n_i} + \frac{1}{n_{i'}}\right)\right]^{1/2} = 2.13\left[\frac{15(16)}{12}\left(\frac{1}{5} + \frac{1}{5}\right)\right]^{1/2} = 6.02$$

Hence, the testing limits for the three pairwise tests are:

| | | | |
|---|---|---|---|
| Locations 1 and 2: | $(8.4 - 4.8) \pm 6.02$ | or | $-2.4$ and $9.6$ |
| Locations 3 and 2: | $(10.8 - 4.8) \pm 6.02$ | or | $-.02$ and $12.0$ |
| Locations 3 and 1: | $(10.8 - 8.4) \pm 6.02$ | or | $-3.6$ and $8.4$ |

Since no test shows a significant difference, we obtain only one grouping:

| Group 1 |
|---|
| Location 1 |
| Location 2 |
| Location 3 |

Note that zero is just inside the lower boundary of the testing limits for locations 2 and 3.

Recall that when the Bonferroni pairwise comparison procedure was conducted on the logarithm of the responses, we concluded that a significant difference existed between the means of locations 2 and 3. Thus here, and in general for small sample sizes, the simple transformations discussed in Section 18.5 are often preferred to the rank transformation because the resulting ANOVA tests are less conservative and tend to have greater statistical power than those associated with the rank transformation.

## 18.8 Case Example—Heart Transplant

In heart transplant surgery, the similarity of the donor's tissue type and that of the recipient is of importance because large differences may increase the probability that the transplanted heart is rejected. Table 18.7 shows a portion of the survival times (in days) obtained from an observational study of 39 patients following heart transplant surgery. The data are grouped into three categories, according to the degree of mismatch between the donor tissue and the recipient tissue. Investigators would like to determine if the mean survival time changes with the degree of mismatch. The alternatives to be tested are:

$$H_0: \mu_1 = \mu_2 = \mu_3$$
$$H_a: \text{not all } \mu_i \text{ are equal}$$

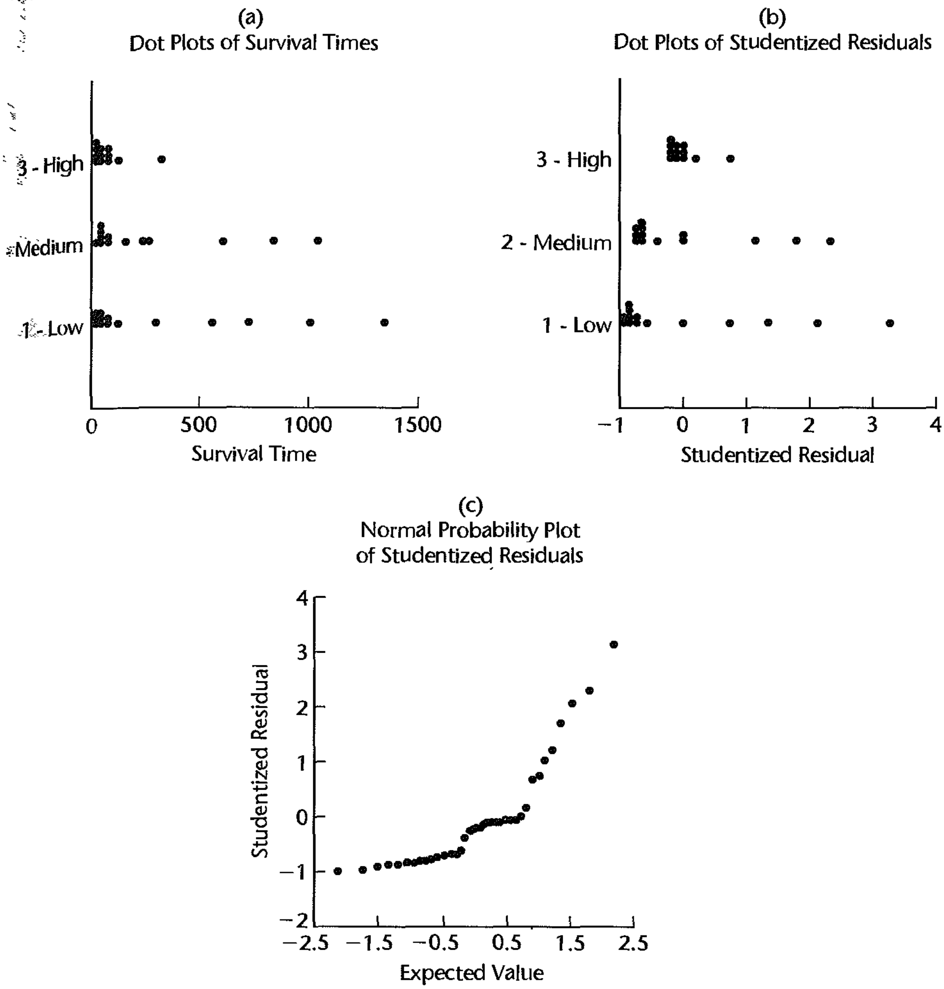
A SYSTAT dot plot of the data by mismatch category is provided in Figure 18.9a. The plot suggests that average survival time may decrease with higher degree of mismatch. An initial fit of analysis of variance model (16.2) was made and the studentized residuals were

**TABLE 18.7 Survival Times of Patients Following Heart Transplant Surgery—Heart Transplant Example.**

| | Degree of Tissue Mismatch ($i$) | | |
|---|---|---|---|
| Case $j$ | Low $i=1$ | Medium $i=2$ | High $i=3$ |
| 1 | 44 | 15 | 3 |
| 2 | 551 | 280 | 136 |
| 3 | 127 | 1,024 | 65 |
| ... | ... | ... | ... |
| 12 | 47 | 836 | 48 |
| 13 | 994 | 51 | |
| 14 | 26 | | |

*Source:* M. L. Puri and P. K. Sen, *Nonparametric Methods in General Linear Models* (New York: John Wiley & Sons, 1985).

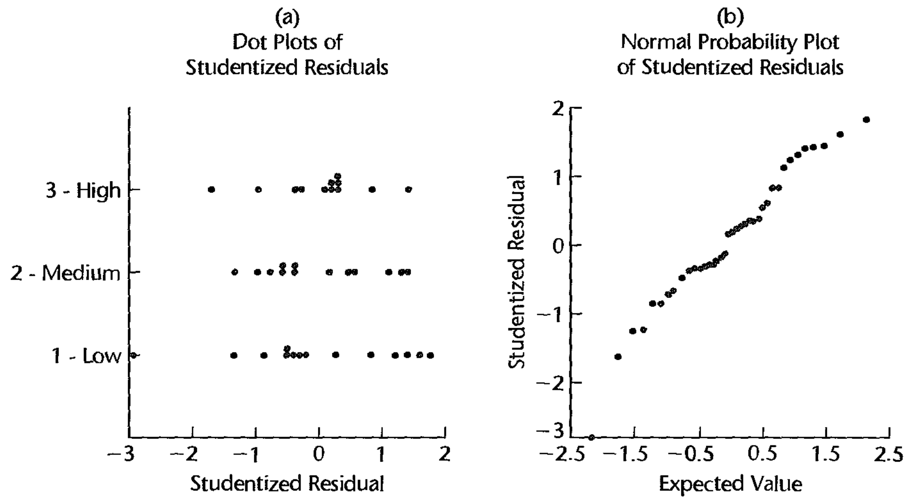
**FIGURE 18.9 SYSTAT Diagnostic Plots—Heart Transplant Example.**




obtained for diagnostic purposes. Two residual plots are presented in Figures 18.9b and 18.9c. The dot plot of the studentized residuals in Figure 18.9b shows that the distribution of the residuals is positively skewed. It also suggests that the error variance may be smaller in the high mismatch group. The Brown-Forsythe test in (18.12) was conducted to examine the constancy of the error variance. The Brown-Forsythe test statistic is $F^*_{BF} = 1.91$ and the $P$-value is .163, supporting constancy of the error variance. On the other hand, the positive skewness of the residuals is confirmed by the upward-curving shape of the normal probability plot in Figure 18.9c and the correlation test for normality ($r = .895$; for $\alpha = .05$, the interpolated critical value in Table B.6 is .971).

A transformation of the response variable was therefore investigated. The Box-Cox procedure led to the maximum likelihood estimate $\hat{\lambda} = .06$, which suggested the logarithmic transformation ($\lambda = 0$). The new response variable $Y' = \log_e Y$ was therefore obtained

**FIGURE 18.10 Diagnostic Plots and ANOVA Table for Transformed Data—Heart Transplant Example.**




(c)
ANOVA Table

DEP VAR: LOGTIME N: 39 MULTIPLE R: 0.282 SQUARED MULTIPLE R: 0.080

ANALYSIS OF VARIANCE

| SOURCE | SUM-OF-SQUARES | DF | MEAN-SQUARE | F-RATIO | P |
|---|---|---|---|---|---|
| CATEGORY | 7.611 | 2 | 3.806 | 1.560 | 0.224 |
| ERROR | 87.834 | 36 | 2.440 | | |

and ANOVA model (16.2) was fitted to this transformed variable. Two plots of studentized residuals are shown in Figure 18.10. A dot plot of the studentized residuals is presented in Figure 18.10a. Notice that the distribution of the residuals now appears to be symmetric, with constant variance. The normality of the distribution of the error terms is supported by the normal probability plot in Figure 18.10b and the correlation test for normality ($r = .982 > .971$).

The residual dot plot in Figure 18.10a shows the possible presence of an outlier in the low tissue mismatch category (studentized residual $= -2.99$). For this case the studentized deleted residual is $-3.40$. The Bonferroni critical value for the outlier test is $t(1-.05/2(39); 36) = t(.999359; 36) = 3.49$. Since $|-3.40| = 3.40 \leq 3.49$, we conclude that this case is not an outlier.

It therefore appears that the logarithmic transformation was successful so that ANOVA model (16.2) is appropriate for the transformed survival times. The ANOVA table for the transformed data is shown in Figure 18.10c. We see that $F^* = 1.56$ and that the $P$-value for the test is .224. For $\alpha = .10$, we therefore conclude $H_0$, that the mean survival time for heart transplant patients with the characteristics of those included in the study does not depend on the degree of tissue mismatch.

## ed eferences


18.1. Hartley, H. O. "Testing the Homogeneity of a Set of Variances," *Biometrika* 31 (1940), pp. 249–255.

18.2. Brown, M. B., and A. B. Forsythe. "Robust Tests for Equality of Variances," *Journal of the American Statistical Association* 69 (1974), pp. 364–67.

18.3. Snedecor, G. W., and W. G. Cochran. *Statistical Methods.* 8th ed. Ames, Iowa: Iowa State University Press, 1989.

18.4. Scheffé, H. *The Analysis of Variance.* New York: John Wiley & Sons, 1959.

18.5. Kruskal, W. H., and W. A. Wallis. "Use of Ranks on One-Criterion Variance Analysis," *Journal of the American Statistical Association* 47 (1952), pp. 583–621 (corrections appear in Vol. 48, pp. 907–11).


## oblems

18.1. Refer to Figures 18.3 and 18.4. What feature of the residual sequence plots enables you to diagnose that in one case the error variance changes over time whereas in the other case the effect is of a different nature? Could you make a diagnosis about time effects from a residual dot plot?

18.2. A student proposed in class that deviations of the observations $Y_{ij}$ around the estimated overall mean $\bar{Y}_{..}$ be plotted to assist in evaluating the appropriateness of ANOVA model (16.2). Would these deviations be helpful in studying the independence of the error terms? The constancy of the variance of the error terms? The normality of the error terms? Discuss.

18.3. A consultant discussing ANOVA applications in a seminar stated: "Sometimes I find that treatment effects in an experiment do not show up through differences in the treatment means. Hence, it is important to compare the residual plots for the treatments." A member of the audience asked: "I don't think I understood your point regarding differences in treatment means being explored using residual plots." Discuss.

*18.4. Refer to **Productivity improvement** Problem 16.7.

a. Prepare aligned residual dot plots by factor level. What departures from ANOVA model (16.2) can be studied from these plots? What are your findings?

b. Prepare a normal probability plot of the residuals. Also obtain the coefficient of correlation between the ordered residuals and their expected values under normality. Does the normality assumption appear to be reasonable here?

c. Obtain the studentized deleted residuals and conduct the Bonferroni outlier test; use $\alpha = .01$. State the alternatives, decision rule, and conclusion.

d. The economist wishes to investigate whether location of the firm's home office is related to productivity improvement. The home office locations are as follows (U: U.S.; E: Europe):

| | *j* | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *i* | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| 1 | U | E | E | E | E | U | U | U | U | | | |
| 2 | E | E | E | E | U | U | U | U | U | E | E | E |
| 3 | E | U | E | U | U | E | | | | | | |

Prepare aligned residual dot plots by factor level in which the location of the home office is identified. Does it appear that ANOVA model (16.2) could be improved by adding location of home office as a second factor? Explain.

18.5. Refer to **Questionnaire color** Problem 16.8.

a. Prepare aligned residual dot plots by color. What departures from ANOVA model (16.2) can be studied from these plots? What are your findings?

b. Prepare a normal probability plot of the residuals. Also obtain the coefficient of correlation between the ordered residuals and their expected values under normality. Does the normality assumption appear to be reasonable here?

c. The observations within each factor level are in geographic sequence. Prepare residual sequence plots. What can be studied from these plots? What are your findings?

d. Obtain the studentized deleted residuals and conduct the Bonferroni outlier test; use $\alpha = .025$. State the alternatives, decision rule, and conclusion.

18.6. Refer to **Rehabilitation therapy** Problem 16.9.

a. Obtain the residuals and prepare aligned residual dot plots by factor level. What departures from ANOVA model (16.2) can be studied from these plots? What are your findings?

b. Prepare a normal probability plot of the residuals. Also obtain the coefficient of correlation between the ordered residuals and their expected values under normality. Does the normality assumption appear to be reasonable here?

c. The observations within each factor level are in time order. Prepare residual sequence plots and analyze them. What are your findings?

d. Obtain the studentized deleted residuals and conduct the Bonferroni outlier test; use $\alpha = .01$. State the alternatives, decision rule, and conclusion.

*18.7. Refer to **Cash offers** Problem 16.10.

a. Obtain the residuals and prepare aligned residual dot plots by factor level. What departures from ANOVA model (16.2) can be studied from these plots? What are your findings?

b. Prepare a normal probability plot of the residuals. Also obtain the coefficient of correlation between the ordered residuals and their expected values under normality. Does the normality assumption appear to be reasonable here?

c. The observations within each factor level are in time order. Prepare residual sequence plots and interpret them. What are your findings?

d. Obtain the studentized deleted residuals and conduct the Bonferroni outlier test; use $\alpha = .025$. State the alternatives, decision rule, and conclusion.

e. An executive in the consumer organization has been told that used-car dealers in the region tend to make lower cash offers during weekends (Friday evening through Sunday) than at other times. The times when offers were obtained are as follows (W: weekend; O: other time):

| | $j$ | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $i$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| 1 | O | O | W | O | W | O | W | O | W | O | W | W |
| 2 | O | W | W | O | W | O | W | O | O | W | W | O |
| 3 | O | W | O | W | O | O | O | W | W | W | O | W |

Prepare aligned residual dot plots by factor level in which the time of the offer is identified. Does it appear that ANOVA model (16.2) could be improved by adding time of offer as a second factor? Explain.

*18.8. Refer to **Filling machines** Problem 16.11.

a. Obtain the residuals and prepare aligned residual dot plots by machine. What departures from ANOVA model (16.2) can be studied from these plots? What are your findings?

b. Prepare a normal probability plot of the residuals. Also obtain the coefficient of correlation between the ordered residuals and their expected values under normality. Does the normality assumption appear to be reasonable here?

c. The observations within each factor level are in time order. Prepare residual sequence plots and interpret them. What are your findings?

d. Obtain the studentized deleted residuals and conduct the Bonferroni outlier test; use $\alpha = .01$. State the alternatives, decision rule, and conclusion.

18.9. Refer to **Premium distribution** Problem 16.12.

a. Obtain the residuals and prepare aligned residual dot plots by agent. What departures from ANOVA model (16.2) can be studied from these plots? What are your findings?

b. Prepare a normal probability plot of the residuals. Also obtain the coefficient of correlation between the ordered residuals and their expected values under normality. Does the normality assumption appear to be reasonable here?

c. The observations within each factor level are in time order. Prepare residual sequence plots and interpret them. What are your findings?

d. Obtain the studentized deleted residuals and conduct the Bonferroni outlier test; use $\alpha = .025$. State the alternatives, decision rule, and conclusion.

18.10. **Computerized game.** Four teams competed in 20 trials of a computerized business game. Each trial involved a new game, the objective for each team being to maximize profits in the given trial. A researcher fitted ANOVA model (16.2) to determine whether or not the mean profits for the four teams are the same and obtained the following residuals:

| | *j* | | | | | | |
|---|---|---|---|---|---|---|---|
| *i* | 1 | 2 | 3 | ... | 18 | 19 | 20 |
| 1 | .10 | .28 | .10 | ... | .10 | .28 | .28 |
| 2 | −1.44 | −1.44 | −1.12 | ... | 1.02 | 1.18 | 1.51 |
| 3 | −.93 | −.70 | −.81 | ... | .54 | .43 | .65 |
| 4 | −.15 | .11 | .25 | ... | .11 | .25 | .38 |

The residuals for each team are given in time order. Construct appropriate residual plots to study whether the error terms are independent from trial to trial for each team. What are your findings?

*18.11. Refer to **Productivity improvement** Problem 16.7. Examine by means of the Brown-Forsythe test whether or not the treatment error variances are equal; use $\alpha = .05$. State the alternatives, decision rule, and conclusion. What is the $P$-value of the test?

18.12. Refer to **Rehabilitation therapy** Problem 16.9. Examine by means of the Brown-Forsythe test whether or not the treatment error variances are equal; use $\alpha = .10$. State the alternatives, decision rule, and conclusion. What is the $P$-value of the test?

*18.13. Refer to **Cash offers** Problem 16.10. Assume that the error terms are approximately normally distributed.

a. Examine by means of the Hartley test whether or not the treatment error variances are equal; use $\alpha = .01$. State the alternatives, decision rule, and conclusion. What is the $P$-value of the test?

b. Would you reach the same conclusion as in part (a) with the Brown-Forsythe test?

*18.14. Refer to **Filling machines** Problem 16.11. Assume that the error terms are approximately normally distributed.

a. Examine by means of the Hartley test whether or not the treatment error variances are equal; use $\alpha = .01$. State the alternatives, decision rule, and conclusion. What is the $P$-value of the test?

b. Would you reach the same conclusion as in part (a) with the Brown-Forsythe test statistic?

18.15. **Helicopter service.** An operations analyst in a sheriff's department studied how frequently their emergency helicopter was used during the past year, by time of day (shift 1: 2 A.M.–8 A.M.; shift 2: 8 A.M.–2 P.M.; shift 3: 2 P.M.–8 P.M.; shift 4: 8 P.M.–2 A.M.). Random samples of size 20 for each shift were obtained. The data follow (in time order):

| | | | | $j$ | | | |
|---|---|---|---|---|---|---|---|
| $i$ | 1 | 2 | 3 | ... | 18 | 19 | 20 |
| 1 | 4 | 3 | 5 | ... | 4 | 1 | 6 |
| 2 | 0 | 2 | 0 | .. | 2 | 2 | 0 |
| 3 | 2 | 1 | 0 | | 0 | 2 | 4 |
| 4 | 5 | 2 | 4 | | 5 | 2 | 3 |

Since the data are counts, the analyst was concerned about the normality and equal variances assumptions of ANOVA model (16.2).

a. Obtain the fitted values and residuals for ANOVA model (16.2).

b. Prepare suitable residual plots to study whether or not the error variances are equal for the four shifts. What are your findings?

c. Test by means of the Brown-Forsythe test whether or not the treatment error variances are equal: use $\alpha = .10$. What is the $P$-value of the test? Are your results consistent with the diagnosis in part (b)?

d. For each shift, calculate $\bar{Y}_{i.}$ and $s_i$. Examine the three relations found in the table on page 791 and determine the transformation that is most appropriate here. What do you conclude?

e. Use the Box-Cox procedure to find an appropriate power transformation of $Y$, first adding the constant 1 to each $Y$ observation. Evaluate $SSE$ for the values of $\lambda$ given in Table 18.6. Does $\lambda = .5$, a square-root transformation, appear to be reasonable, based on the Box-Cox procedure?

18.16. Refer to **Helicopter service** Problem 18.15. The analyst decided to apply the square root transformation $Y' = \sqrt{Y}$ and examine its effectiveness.

a. Obtain the transformed response data, fit ANOVA model (16.2), and obtain the residuals.

b. Prepare suitable plots of the residuals to study the equality of the error variances of the transformed response variable for the four shifts. Also obtain a normal probability plot and the coefficient of correlation between the ordered residuals and their expected values under normality. What are your findings? Does the transformation appear to have been effective?

c. Test by means of the Brown-Forsythe test whether or not the treatment error variances for the transformed response variable are equal: use $\alpha = .10$. State the alternatives, decision rule, and conclusion. Are your findings in part (b) consistent with your conclusion here?

*18.17. **Winding speeds.** In a completely randomized design to study the effect of the speed of winding thread (1: slow; 2: normal; 3: fast; 4: maximum) onto 75-yard spools, 16 runs of 10,000 spools each were made at each of the four winding speeds. The response variable is the number of thread breaks during the production run. The results (in time order) are as follows:

| | $j$ | | | | | | |
|---|---|---|---|---|---|---|---|
| $i$ | 1 | 2 | 3 | ... | 14 | 15 | 16 |
| 1 | 4 | 3 | 2 | ... | 2 | 3 | 4 |
| 2 | 7 | 6 | 4 | ... | 4 | 7 | 6 |
| 3 | 12 | 6 | 14 | ... | 13 | 10 | 14 |
| 4 | 17 | 15 | 7 | ... | 19 | 9 | 23 |

Since the responses are counts, the researcher was concerned about the normality and equal variances assumptions of ANOVA model (16.2).

a. Obtain the fitted values and residuals for ANOVA model (16.2).

b. Prepare suitable residual plots to study whether or not the error variances are equal for the four winding speeds. What are your findings?

c. Test by means of the Brown-Forsythe test whether or not the treatment error variances are equal; use $\alpha = .05$. What is the $P$-value of the test? Are your results consistent with the diagnosis in part (b)?

d. For each winding speed, calculate $\bar{Y}_{i\cdot}$ and $s_i$. Examine the three relations found in the table on page 791 and determine the transformation that is most appropriate here. What do you conclude?

e. Use the Box-Cox procedure to find an appropriate power transformation of $Y$. Evaluate $SSE$ for the values of $\lambda$ given in Table 18.6. Does $\lambda = 0$, a logarithmic transformation, appear to be reasonable, based on the Box-Cox procedure?

*18.18. Refer to **Winding speeds** Problem 18.17. The researcher decided to apply the logarithmic transformation $Y' = \log_{10} Y$ and investigate its effectiveness.

a. Obtain the transformed response data, fit ANOVA model (16.2), and obtain the residuals.

b. Prepare suitable plots of the residuals to study the equality of the error variances of the transformed response variable for the four winding speeds. Also obtain a normal probability plot and the coefficient of correlation between the ordered residuals and their expected values under normality. What are your findings about the effectiveness of the transformation?

c. Test by means of the Brown-Forsythe test whether or not the treatment error variances for the transformed response variable are equal; use $\alpha = .05$. State the alternatives, decision rule, and conclusion. Are your findings in part (b) consistent with your conclusion here?

18.19. Refer to **Helicopter service** Problem 18.15. Assume that ANOVA model (18.13) is appropriate. Use weighted least squares with the untransformed data to test for the equality of the shift means; control the $\alpha$ risk at .05. State the alternatives, full and reduced regression models, decision rule, and conclusion.

*18.20. Refer to **Winding speeds** Problem 18.17. Assume that ANOVA model (18.13) is appropriate. Use weighted least squares with the untransformed data to test for the equality of the winding thread speed means; use $\alpha = .01$. State the alternatives, full and reduced regression models, decision rule, and conclusion.

18.21. Why is the nonparametric rank $F$ test a nonparametric test?

18.22. Explain why the limits in (18.30) are testing limits and not confidence limits.

*18.23. Refer to **Productivity improvement** Problem 16.7.

a. Conduct the nonparametric rank $F$ test; use $\alpha = .05$. State the alternatives, decision rule, and conclusion.

b. What is the $P$-value of the test in part (a)?

c. Does the conclusion in part (a) differ from the one in Problem 16.7e?

d. Do the data suggest that a nonparametric test is needed here?

e. Conduct multiple pairwise tests based on the ranked data to group the three types of firms according to mean productivity improvement. Use family level of significance $\alpha = .10$. Describe your findings.

*18.24. Refer to **Cash offers** Problem 16.10.

a. Conduct the nonparametric rank $F$ test; use $\alpha = .01$. State the alternatives, decision rule, and conclusion.

b. What is the $P$-value of the test in part (a)?

c. Does the conclusion in part (a) differ from the one in Problem 16.10e?

d. Do the data suggest that a nonparametric test is needed here?

e. Conduct multiple pairwise tests based on the ranked data to group the three age categories according to mean cash offer. Use family level of significance $\alpha = .10$. Describe your findings.

18.25. **Telephone communications.** A management consultant was engaged by a firm to improve the cost-effectiveness of its communications. As part of the study, the consultant selected 10 home-office executives at random from each of the (1) sales, (2) production, and (3) research and development divisions, and studied the communications of these executives during the past 10 weeks in great detail. Among other data, the consultant obtained the following information on weekly dollar costs of long-distance telephone calls to branch offices by the executives:

| | | | | | $j$ | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $i$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 1 | 666 | 920 | 495 | 602 | 1,499 | 960 | 796 | 343 | 894 | 813 |
| 2 | 488 | 362 | 156 | 546 | 216 | 542 | 345 | 291 | 516 | 126 |
| 3 | 391 | 450 | 609 | 910 | 705 | 472 | 645 | 496 | 763 | 1,309 |

The consultant decided to employ a nonparametric approach to test whether or not the mean telephone expenses for the three divisions are equal.

a. What feature of the data may have suggested the use of a nonparametric test?

b. Conduct the nonparametric rank $F$ test, controlling the risk of Type I error at $\alpha = .05$. State the alternatives, decision rule, and conclusion. What is the $P$-value of the test?

c. Conduct multiple pairwise tests based on the ranked data to group the three divisions according to mean telephone expenditures; use family level of significance $\alpha = .05$. Describe your findings.

## Exercises

18.26. Refer to Figure 18.3. Modify ANOVA model (16.2) to include a linear trend term for the time effect. Is this modified model still an ANOVA model? A linear model?

18.27. Show that $n_T(n_T + 1)/12$ in (18.30) is the sample variance of the consecutive integers 1 to $n_T$.

18.28. Show that test statistics (18.25) and (18.27) are related according to (18.29).

## Projects

18.29. Refer to the **SENIC** data set in Appendix C.1 and Project 16.42.

a. Obtain the residuals and prepare aligned residual dot plots by region. Are any serious departures from ANOVA model (16.2) suggested by your plots?

b. Obtain a normal probability plot of the residuals and calculate the coefficient of correlation between the ordered residuals and their expected values under normality. Is the normality assumption reasonable here?

c. Examine by means of the Brown-Forsythe test whether or not the geographic region error variances are equal; use $\alpha = .05$. State the alternatives, decision rule, and conclusion. What is the $P$-value of the test?

18.30. Refer to the **SENIC** data set in Appendix C.1. A test of whether or not mean length of stay (variable 2) is the same in the four geographic regions (variable 9) is desired, but concern exists about the normality and equal variances assumptions of ANOVA model (16.2).

a. Obtain the residuals and plot them against the fitted values to study whether or not the error variances are equal for the four geographic regions. What are your findings?

b. For each geographic region, calculate $\bar{Y}_{i\cdot}$ and $s_i$. Examine the three relations found in the table on page 791 and determine the transformation that is the most appropriate one here. What do you conclude?

c. Use the Box-Cox procedure to find an appropriate power transformation of $Y$. Evaluate $SSE$ for the values of $\lambda$ given in Table 18.6. Does $\lambda = -1$, a reciprocal transformation, appear to be reasonable, based on the Box-Cox procedure?

d. Use the reciprocal transformation $Y' = 1/Y$ to obtain transformed response data.

e. Fit ANOVA model (16.2) to the transformed data and obtain the residuals. Plot these residuals against the fitted values to study the equality of the error variances of the transformed response variable for the four regions. Also obtain a normal probability plot of the residuals and the coefficient of correlation between the ordered residuals and their expected values under normality. What are your findings?

f. Examine by means of the Brown-Forsythe test whether or not the geographic region variances for the transformed response variable are equal; use $\alpha = .01$. State the alternatives, decision rule, and conclusion. What is the $P$-value of the test?

g. Assume that ANOVA model (16.2) is appropriate for the transformed response variable. Test whether or not the mean length of stay in the transformed units is the same in the four geographic regions. Control the $\alpha$ risk at .01. State the alternatives, decision rule, and conclusion. What is the $P$-value of the test?

18.31. Refer to the **CDI** data set in Appendix C.2 and Project 16.44.

a. Obtain the residuals and prepare aligned residual dot plots by region. Are any serious departures from ANOVA model (16.2) suggested by your plots?

b. Obtain a normal probability plot of the residuals and calculate the coefficient of correlation between the ordered residuals and their expected values under normality. Is the normality assumption reasonable here?

c. Examine by means of the Brown-Forsythe test whether or not the geographic region error variances are equal; use $\alpha = .01$. State the alternatives, decision rule, and conclusion. What is the $P$-value of the test?

18.32. Refer to the **Market share** data set in Appendix C.3 and Project 16.45.

a. Obtain the residuals and prepare aligned residual dot plots by factor-level combinations. Are any serious departures from ANOVA model (16.2) suggested by your plots?

b. Obtain a normal probability plot of the residuals and calculate the coefficient of correlation between the ordered residuals and their expected values under normality. Is the normality assumption reasonable here?

c. Examine by means of the Brown-Forsythe test whether or not the factor level error variances are equal; use $\alpha = .05$. State the alternatives, decision rule, and conclusion. What is the $P$-value of the test?

18.33. Refer to the **SENIC** data set in Appendix C.1 and Project 16.42.

a. Use the nonparametric rank $F$ test to determine whether or not the mean infection risk is the same in the four regions; control the level of significance at $\alpha = .05$. State the alternatives, decision rule, and conclusion. What is the $P$-value of the test?

b. Is your conclusion in part (a) the same as that obtained in Project 16.42? Is the nonparametric test more reasonable here?

c. Use the multiple pairwise testing procedure (18.30) to group the regions; employ family significance level $\alpha = .10$. What are your findings?

18.34. Refer to the **CDI** data set in Appendix C.2 and Project 16.44.

a. Use the nonparametric rank $F$ test to determine whether or not the mean crime rate is the same in the four regions; control the level of significance at $\alpha = .05$. State the alternatives, decision rule, and conclusion. What is the $P$-value of the test?

b. Is your conclusion in part (a) the same as that obtained in Project 16.44? Is the nonparametric test more reasonable here?

c. Use the multiple pairwise testing procedure (18.30) to group the regions; employ family significance level $\alpha = .05$. What are your findings?

18.35. Refer to the **Market share** data set in Appendix C.3 and Project 16.45.

a. Use the nonparametric rank $F$ test to determine whether or not the mean average monthly share is the same for the four factor combinations; control the level of significance at $\alpha = .05$. State the alternatives, decision rule, and conclusion. What is the $P$-value of the test?

b. Is your conclusion in part (a) the same as that obtained in Project 16.45? Is the nonparametric test more reasonable here?

c. Use the multiple pairwise testing procedure (18.30) to group the factor combinations; employ family significance level $\alpha = .05$. What are your findings?

18.36. Obtain the exact sampling distribution of the nonparametric rank $F_R^*$ test statistic in (18.25) when $H_0$ holds, for the case $r = 2$ and $n_i \equiv 2$. [*Hint:* What does the equality of the treatment means imply about the arrangement of the ranks 1, 2, 3, 4?]

18.37. Three populations are being studied; each is uniform between 300 and 800.

a. Generate 10 random observations from each of the three uniform populations and calculate the $F_R^*$ test statistic (18.25).

b. Repeat part (a) 500 times.

c. Calculate the mean and standard deviation of the 500 test statistics. How do these values compare with the characteristics of the relevant $F$ distribution?

d. What proportion of the 500 test statistics obtained in part (b) is less than $F(.90; 2, 27)$? What proportion is less than $F(.99; 2, 27)$? How do these proportions agree with theoretical expectations?

## Case Studies

18.38. Refer to the **Prostate cancer** data set in Appendix C.5 and Case Study 16.49. Check to see whether concern exists about the assumption of normality and equal variances for the ANOVA model that you decided upon in Case Study 16.49. Document the steps taken in your assessment of these concerns. Is a transformation indicated here? If yes, what transformation is recommended? Why?

18.39. Refer to the **Real estate sales** data set in Appendix C.7 and Case Study 16.50. Check to see whether concern exists about the assumption of normality and equal variances for the ANOVA model that you decided upon in Case Study 16.50. Document the steps taken in your assessment of these concerns. Is a transformation indicated here? If yes, what transformation is recommended? Why?

18.40. Refer to the **Ischemic heart disease** data set in Appendix C.9 and Case Study 16.51. Check to see whether concern exists about the assumption of normality and equal variances for the ANOVA model that you decided upon in Case Study 16.51. Document the steps taken in your assessment of these concerns. Is a transformation indicated here? If yes, what transformation is recommended? Why?