Chapter 17

Analysis of Factor Level Means

17.1 Introduction

In Chapter 16, we discussed the F test for determining whether or not the factor level means μ_i differ. This is a preliminary test to establish whether detailed analysis of the factor level means is warranted. When this test leads to the conclusion that the factor level means μ_i are equal, and ANOVA model (16.2) is appropriate, no relation between the factor and the response variable is present and usually no further analysis of factor means is therefore indicated. On the other hand, when the F test leads to the conclusion that the factor level means μ_i differ, a relation between the factor and the response variable is present. In this latter case, a thorough analysis of the nature of the factor level means is usually undertaken. This is done in two principal ways:

1. Analysis of the factor level means of interest using estimation techniques.
2. Statistical tests concerning the factor level means of interest.

Often, the analysis of factor level means combines the two approaches. For instance, a two-sided confidence interval may be constructed initially for an effect of interest. A test concerning this effect is then carried out either by determining whether or not the confidence interval contains the hypothesized value or by constructing the appropriate test statistic.

When many related comparisons are to be made, testing often precedes estimation. This occurs, for instance, when each factor level effect is compared with every other one and the number of factor levels is not small. Here, statistical tests are often performed first to determine the *active* or statistically significant set of comparisons. Estimation techniques are then used to construct confidence intervals for the active comparisons.

Special simultaneous estimation and testing procedures, called multiple comparison procedures, are required when a series of interval estimates or tests are performed. These multiple comparison procedures preserve the overall confidence coefficient $1 - \alpha$, or the overall significance level α, for the family of inferences.

We first discuss three simple graphical methods for displaying the factor level means. Much of the remainder of the chapter is devoted to a consideration of important multiple comparison procedures. In Section 16.10 we introduced methods for determining sample

TABLE 17.1 Summary of Results—Kenton Food Company Example.

	Package Design (i)				
	1	2	3	4	Total
n_i	5	5	4	5	19
$Y_{i\cdot}$	73	67	78	136	354
$\bar{Y}_{i\cdot}$	14.6	13.4	19.5	27.2	18.63

Source of Variation	*SS*	*df*	*MS*
Between designs	588.22	3	196.07
Error	158.20	15	10.55
Total	746.42	18	

Package Design	Characteristics
1	3 colors, with cartoons
2	3 colors, without cartoons
3	5 colors, with cartoons
4	5 colors, without cartoons

sizes in single-factor studies based on the power approach. This chapter concludes with a discussion of the estimation approach to sample size planning.

Throughout this chapter, we continue to assume the usual single-factor ANOVA model. The cell means version of this model was given in (16.2):

$$Y_{ij} = \mu_i + \varepsilon_{ij} \tag{17.1}$$

where:

μ_i are parameters

ε_{ij} are independent $N(0, \sigma^2)$

Our discussion of the analysis of factor means will be illustrated by two examples. The first is the Kenton Food Company example. Data for this example are provided in Table 16.1 on page 686, and the ANOVA table is displayed in Figure 16.5 on page 695. For convenience, we repeat the main results in Table 17.1. The second example, the rust inhibitor example, is described next.

Example

In a study of the effectiveness of different rust inhibitors, four brands (A, B, C, D) were tested. Altogether, 40 experimental units were randomly assigned to the four brands, with 10 units assigned to each brand. A portion of the results after exposing the experimental units to severe weather conditions is given in coded form in Table 17.2a. The higher the coded value, the more effective is the rust inhibitor. This study is a completely randomized design, where the levels of the single factor correspond to the four rust inhibitor brands.

The analysis of variance is shown in Table 17.2b. For level of significance $\alpha = .05$ for testing whether or not the four rust inhibitors differ in effectiveness, we require

17.2
d
sis of
re
ts—Rust
tor
ple (data
coded).

(a) Data

	Rust Inhibitor Brand			
	A	B	C	D
j	$i = 1$	$i = 2$	$i = 3$	$i = 4$
1	43.9	89.8	68.4	36.2
2	39.0	87.1	69.3	45.2
3	46.7	92.7	68.5	40.7
...	...	...	...	...
8	38.9	88.1	65.2	38.7
9	43.6	90.8	63.8	40.9
10	40.0	89.1	69.2	39.7
$\bar{Y}_{i.}$	43.14	89.44	67.95	40.47

$\bar{Y}_{..} = 60.25$

(b) Analysis of Variance

Source of Variation	SS	df	MS
Between brands	15,953.47	3	5,317.82
Error	221.03	36	6.140
Total	16,174.50	39	

$F(.95; 3, 36) = 2.87$. Using the mean squares from Table 17.2b, we obtain the test statistic:

$$F^* = \frac{MSTR}{MSE} = \frac{5{,}317.82}{6.140} = 866.1$$

Since $F^* = 866.1 > 2.87$, we conclude that the four rust inhibitors differ in effectiveness. The P-value of the test is 0+. We therefore wish to analyze the nature of the factor level effects, particularly whether one rust inhibitor is substantially more effective than the others.

17.2 Plots of Estimated Factor Level Means

Before undertaking formal analysis of the nature of the factor level effects, it is usually helpful to examine these factor effects informally from a plot of the estimated factor level means $\bar{Y}_{i.}$. We shall take up three types of plots: (1) a line plot, (2) a bar graph, and (3) a main effects plot. All three plots are appropriate whether the sample sizes n_i are equal or not.

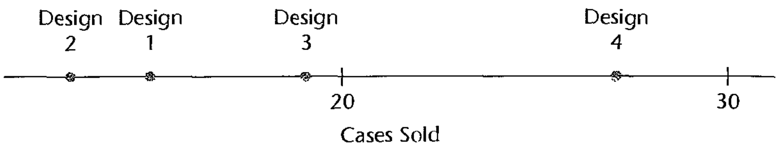
Line Plot

A line plot of the estimated factor level means simply shows the positions of the $\bar{Y}_{i.}$ on a line scale. It is a very simple, but effective, device for indicating when one or several factor level means may differ substantially from the others.

Example

In Figure 17.1 we present a line plot of the estimated factor level means $\bar{Y}_{i.}$ for the Kenton Food Company example. It is clear from Figure 17.1 that design 4 led by far to the highest

FIGURE 17.1 Line Plot of Estimated Factor Level Means—Kenton Food Company Example.



mean sales in the study, and that package designs 1 and 2 led to the smallest mean sales which did not differ much from each other. The purpose of the formal inference procedures to be taken up shortly is to determine whether the pattern noted here reflects underlying differences in the factor level means μ_i or is simply the result of random variation.

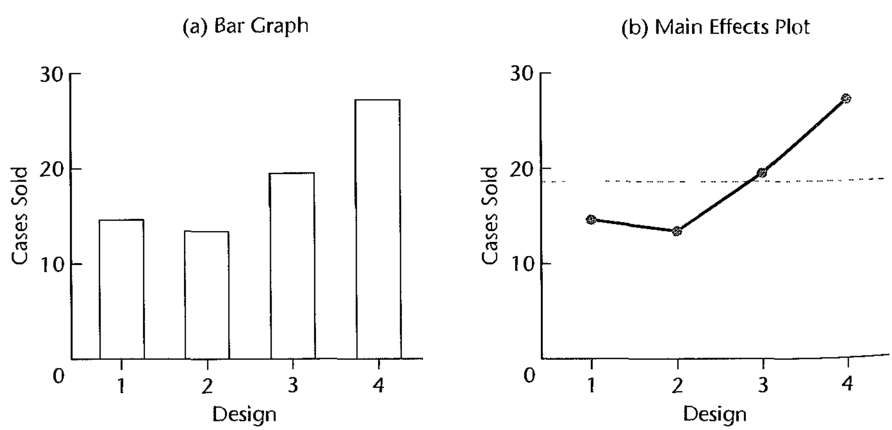
Bar Graph and Main Effects Plot

Bar graphs and main effects plots are frequently used to display the estimated factor level means in two dimensions. Both can be used to compare the magnitudes of different factor level means. In a bar graph, vertical bars are used to display the estimated factor level means. In a main effects plot, a scatter plot of the estimated factor level means is provided, and the plot symbols are connected by straight lines, to visibly highlight potential trends in the cell means. Note that these trend lines are not particularly meaningful for qualitative factors. For this reason, main effects plots are most appropriate for quantitative factors. In some packages, the main effects plot also displays the overall mean using a horizontal line, permitting visual comparisons of the factor-level means with the overall mean.

Example

A bar graph and a main effects plot of the estimated factor level means for the Kenton Food Company example are displayed in Figure 17.2. Because package design is a qualitative factor, the bar graph in Figure 17.2a is the recommended graphic here. An advantage of the main effects plot in Figure 17.2b is that it permits a visual comparison of the estimated factor level means and the overall mean. Here it shows that designs 3 and 4 had higher mean sales than the overall mean, while designs 1 and 2 both had smaller means sales than the overall mean.

FIGURE 17.2 MINITAB Bar Graph and Main Effects Plot of Estimated Factor Level Means—Kenton Food Company Example.



Comments

1. In Section 16.7 we defined the difference of the factor level mean and the overall mean as the factor level effect. In our discussion of multifactor studies in Chapter 19 and beyond, we shall refer to factor level effects as main effects. For this reason, the plot in Figure 17.2b is frequently referred to as a main effects plot.

2. None of the three plots provides information on the standard errors. Without such information, we cannot easily tell whether differences between factor level means are statistically significant. Later in this chapter, we shall enhance all three plots by including the information on the standard errors.

3. The normal probability plot introduced in Chapter 3 can also be used to compare the estimated factor level means. A normal probability plot is appropriate when the sample sizes n_i are equal and the number of factors r is sufficiently large. We recommend that a normal probability plot of factor level means be considered if $r \geq 10$. ■

17.3 Estimation and Testing of Factor Level Means

Inferences for factor level means are generally concerned with one or more of the following:

1. A single factor level mean μ_i
2. A difference between two factor level means
3. A contrast among factor level means
4. A linear combination of factor level means

We discuss each of these types of inferences in turn.

Inferences for Single Factor Level Mean

Estimation. An unbiased point estimator of the factor level mean μ_i is given in (16.16):

$$\hat{\mu}_i = \bar{Y}_{i\cdot} \tag{17.2}$$

This estimator has mean and variance:

$$E\{\bar{Y}_{i\cdot}\} = \mu_i \tag{17.3a}$$

$$\sigma^2\{\bar{Y}_{i\cdot}\} = \frac{\sigma^2}{n_i} \tag{17.3b}$$

The latter result follows because (16.43) indicates that $\bar{Y}_{i\cdot} = \mu_i + \bar{\varepsilon}_{i\cdot}$, the sum of a constant plus a mean of n_i independent ε_{ij} error terms, each of which has variance σ^2. Further, $\bar{Y}_{i\cdot}$ is normally distributed because the error terms ε_{ij} are independent normal random variables.

The estimated variance of $\bar{Y}_{i\cdot}$ is denoted by $s^2\{\bar{Y}_{i\cdot}\}$ and is obtained as usual by replacing σ^2 in (17.3b) by the unbiased point estimator MSE:

$$s^2\{\bar{Y}_{i\cdot}\} = \frac{MSE}{n_i} \tag{17.4}$$

The estimated standard deviation $s\{\bar{Y}_{i\cdot}\}$ is the positive square root of (17.4).

It can be shown that:

$$\frac{\bar{Y}_{i\cdot} - \mu_i}{s\{\bar{Y}_{i\cdot}\}} \text{ is distributed as } t(n_T - r) \text{ for ANOVA model (17.1)} \tag{17.5}$$

where the degrees of freedom are those associated with MSE. The result (17.5) follows from the definition of t in (A.44) since: (1) $\bar{Y}_{i\cdot}$ is normally distributed and (2) MSE/σ^2 is distributed independently of $\bar{Y}_{i\cdot}$ as $\chi^2(n_T - r)/(n_T - r)$ according to the following theorem:

$$\text{For ANOVA model (17.1), } SSE/\sigma^2 \text{ is distributed as } \chi^2 \text{ with } n_T - r \text{ degrees of freedom, and is independent of } \bar{Y}_{1\cdot}, \ldots, \bar{Y}_{r\cdot} \tag{17.6}$$

It follows directly from (17.5) that the $1 - \alpha$ confidence limits for μ_i are:

$$\bar{Y}_{i\cdot} \pm t(1 - \alpha/2; n_T - r)s\{\bar{Y}_{i\cdot}\} \tag{17.7}$$

Testing. The confidence interval based on the limits in (17.7) can be used to test a hypothesis of the form:

$$\begin{aligned} H_0&: \mu_i = c \\ H_a&: \mu_i \neq c \end{aligned} \tag{17.8}$$

where c is an appropriate constant. We conclude H_0, at level of significance α, when c is contained in the confidence interval, and we conclude H_a when the confidence interval does not contain c. Equivalently, one can compute the test statistic:

$$t^* = \frac{\bar{Y}_{i\cdot} - c}{s\{\bar{Y}_{i\cdot}\}} \tag{17.9}$$

Test statistic t^* follows a t distribution with $n_T - r$ degrees of freedom when H_0 is true, according to (17.5). Consequently, we conclude H_0 whenever $|t^*| \leq t(1 - \alpha/2; n_T - r)$; otherwise, we conclude H_a.

Example

In the Kenton Food Company example, the sales manager wished to estimate mean sales for package design 1 with a 95 percent confidence interval. Using the results from Table 17.1, we have:

$$\bar{Y}_{1\cdot} = 14.6 \qquad n_1 = 5 \qquad MSE = 10.55$$

We require $t(.975; 15) = 2.131$. Finally, we need $s\{\bar{Y}_{1\cdot}\}$. We have:

$$s^2\{\bar{Y}_{1\cdot}\} = \frac{MSE}{n_1} = \frac{10.55}{5} = 2.110$$

so that $s\{\bar{Y}_{1\cdot}\} = 1.453$. Hence, we obtain the confidence limits $14.6 \pm 2.131(1.453)$ and the 95 percent confidence interval is:

$$11.5 \leq \mu_1 \leq 17.7$$

Thus, we estimate with confidence coefficient .95 that the mean sales per store for package design 1 are between 11.5 and 17.7 cases.

Graphical Displays. One way to enhance a bar graph or the main effects plot of factor level means is to display the confidence limits in (17.7) for each factor level mean. Figure 17.3 provides two such plots. Figure 17.3a contains a *bar-interval graph*, in which the 95 percent confidence limits are superimposed on a bar graph of the treatment means. Figure 17.3b contains an *interval plot*, in which the 95 percent confidence limits for each factor level

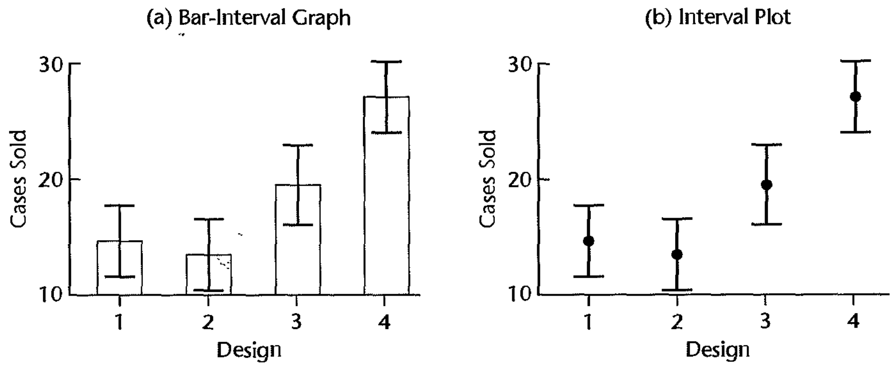


FIGURE 17.3 Bar-Interval Graph and Interval Plot—Kenton Food Company Example.

mean are displayed. Many investigators prefer to simply display limits that correspond to plus-or-minus one standard error—that is, $\bar{Y}_{i\cdot} \pm s\{\bar{Y}_{i\cdot}\}$.

Inferences for Difference between Two Factor Level Means

Estimation. Frequently two treatments or factor levels are to be compared by estimating the difference D between the two factor level means, say, μ_i and $\mu_{i'}$:

$$D = \mu_i - \mu_{i'} \tag{17.10}$$

Such a difference between two factor level means is called a *pairwise comparison*. A point estimator of D in (17.10), denoted by $\hat{D}$, is:

$$\hat{D} = \bar{Y}_{i\cdot} - \bar{Y}_{i'\cdot} \tag{17.11}$$

This point estimator is unbiased:

$$E\{\hat{D}\} = \mu_i - \mu_{i'} \tag{17.12}$$

Since $\bar{Y}_{i\cdot}$ and $\bar{Y}_{i'\cdot}$ are independent, the variance of $\hat{D}$ follows from (A.31b):

$$\sigma^2\{\hat{D}\} = \sigma^2\{\bar{Y}_{i\cdot}\} + \sigma^2\{\bar{Y}_{i'\cdot}\} = \sigma^2\left(\frac{1}{n_i} + \frac{1}{n_{i'}}\right) \tag{17.13}$$

The estimated variance of $\hat{D}$, denoted by $s^2\{\hat{D}\}$, is given by:

$$s^2\{\hat{D}\} = MSE\left(\frac{1}{n_i} + \frac{1}{n_{i'}}\right) \tag{17.14}$$

Finally, $\hat{D}$ is normally distributed by (A.40) because $\hat{D}$ is a linear combination of independent normal variables.

It follows from these characteristics, theorem (17.6), and the definition of t in (A.44) that:

$$\frac{\hat{D} - D}{s\{\hat{D}\}} \text{ is distributed as } t(n_T - r) \text{ for ANOVA model (17.1)} \tag{17.15}$$

Hence, the $1 - \alpha$ confidence limits for D are:

$$\hat{D} \pm t(1 - \alpha/2; n_T - r)s\{\hat{D}\} \tag{17.16}$$

Testing. There is often interest in testing whether two factor level means are the same. The alternatives here are of the form:

$$\begin{aligned} H_0&: \mu_i = \mu_{i'} \\ H_a&: \mu_i \neq \mu_{i'} \end{aligned} \tag{17.17}$$

The alternatives in (17.17) can be stated equivalently as follows:

$$\begin{aligned} H_0&: \mu_i - \mu_{i'} = 0 \\ H_a&: \mu_i - \mu_{i'} \neq 0 \end{aligned} \tag{17.17a}$$

Conclusion H_0 is reached at the α level of significance if zero is contained within the confidence limits (17.16); otherwise, conclusion H_a is reached. An equivalent procedure is based on the test statistic:

$$t^* = \frac{\hat{D}}{s\{\hat{D}\}} \tag{17.18}$$

Conclusion H_0 is reached if $|t^*| \leq t(1 - \alpha/2; n_T - r)$; otherwise, H_a is concluded.

Example

For the Kenton Food Company example, package designs 1 and 2 used 3-color printing and designs 3 and 4 used 5-color printing, as shown in Table 17.1. We wish to estimate the difference in mean sales for 5-color designs 3 and 4 using a 95 percent confidence interval. That is, we wish to estimate $D = \mu_3 - \mu_4$. From Table 17.1, we have:

$$\bar{Y}_{3.} = 19.5 \qquad n_3 = 4 \qquad MSE = 10.55$$
$$\bar{Y}_{4.} = 27.2 \qquad n_4 = 5$$

Hence:

$$\hat{D} = \bar{Y}_{3.} - \bar{Y}_{4.} = 19.5 - 27.2 = -7.7$$

The estimated variance of $\hat{D}$ is:

$$s^2\{\hat{D}\} = MSE\left(\frac{1}{n_3} + \frac{1}{n_4}\right) = 10.55\left(\frac{1}{4} + \frac{1}{5}\right) = 4.748$$

so that the estimated standard deviation of $\hat{D}$ is $s\{\hat{D}\} = 2.179$. We require $t(.975; 15) = 2.131$. The confidence limits therefore are $-7.7 \pm 2.131(2.179)$, and the desired 95 percent confidence interval is:

$$-12.3 \leq \mu_3 - \mu_4 \leq -3.1$$

Thus, we estimate with confidence coefficient .95 that the mean sales for package design 3 fall short of those for package design 4 by somewhere between 3.1 and 12.3 cases per store.

Note from Table 17.1 that the only difference between package designs 3 and 4 is the presence of cartoons; both designs used 5-color printing. The sales manager may therefore wish to test whether the addition of cartoons affects sales for 5-color designs. The alternatives here are:

$$H_0: \mu_3 - \mu_4 = 0$$
$$H_a: \mu_3 - \mu_4 \neq 0$$

Since the hypothesized difference zero in H_0 is not contained within the 95 percent confidence limits -12.3 and -3.1, we conclude H_a, that the presence of cartoons has an effect. We could also obtain test statistic (17.18):

$$t^* = \frac{\hat{D}}{s\{\hat{D}\}} = \frac{-7.7}{2.179} = -3.53$$

Since $|t^*| = 3.53 > t(.975; 15) = 2.131$, we conclude H_a. The two-sided P-value for this test is .003.

Inferences for Contrast of Factor Level Means

A *contrast* is a comparison involving two or more factor level means and includes the previous case of a pairwise difference between two factor level means in (17.10). A contrast will be denoted by L, and is defined as a linear combination of the factor level means μ_i where the coefficients c_i sum to zero:

$$L = \sum_{i=1}^{r} c_i \mu_i \qquad \text{where } \sum_{i=1}^{r} c_i = 0 \tag{17.19}$$

Illustrations of Contrasts. In the Kenton Food Company example, package designs 1 and 2 used 3-color printing and designs 3 and 4 used 5-color printing, as shown in Table 17.1. Also, package designs 1 and 3 utilized cartoons while no cartoons were utilized in designs 2 and 4. The following contrasts here may be of interest:

1. Comparison of the mean sales for the two 3-color designs:

$$L = \mu_1 - \mu_2$$

Here, $c_1 = 1$, $c_2 = -1$, $c_3 = 0$, $c_4 = 0$, and $\sum c_i = 0$.

2. Comparison of the mean sales for the 3-color and 5-color designs:

$$L = \frac{\mu_1 + \mu_2}{2} - \frac{\mu_3 + \mu_4}{2}$$

Here, $c_1 = 1/2$, $c_2 = 1/2$, $c_3 = -1/2$, $c_4 = -1/2$, and $\sum c_i = 0$.

3. Comparison of the mean sales for designs with and without cartoons:

$$L = \frac{\mu_1 + \mu_3}{2} - \frac{\mu_2 + \mu_4}{2}$$

Here, $c_1 = 1/2$, $c_2 = -1/2$, $c_3 = 1/2$, $c_4 = -1/2$, and $\sum c_i = 0$.

4. Comparison of the mean sales for design 1 with average sales for all four designs:

$$L = \mu_1 - \frac{\mu_1 + \mu_2 + \mu_3 + \mu_4}{4}$$

Here, $c_1 = 3/4$, $c_2 = -1/4$, $c_3 = -1/4$, $c_4 = -1/4$, and $\sum c_i = 0$.

Note that the first contrast is simply a pairwise comparison. In the second and third contrasts, averages of several factor level means are compared. The fourth contrast is the factor effect τ_1 defined by (16.60) and (16.63).

The averages used here are unweighted averages of the means μ_i; these are ordinarily the averages of interest. In special cases one might be interested in weighted averages of the μ_i to describe the mean response for a group of several factor levels. For example, if both 3-color and 5-color designs were to be employed, with 3-color printing used three times as often as 5-color printing, the comparison of the effect of cartoons versus no cartoons might be based on the contrast:

$$L = \frac{3\mu_1 + \mu_3}{4} - \frac{3\mu_2 + \mu_4}{4}$$

Here, $c_1 = 3/4$, $c_2 = -3/4$, $c_3 = 1/4$, $c_4 = -1/4$, and $\sum c_i = 0$.

Estimation. An unbiased estimator of a contrast L is:

$$\hat{L} = \sum_{i=1}^{r} c_i \bar{Y}_{i.} \tag{17.20}$$

Since the $\bar{Y}_{i.}$ are independent, the variance of $\hat{L}$ according to (A.31) is:

$$\sigma^2\{\hat{L}\} = \sum_{i=1}^{r} c_i^2 \sigma^2\{\bar{Y}_{i.}\} = \sum_{i=1}^{r} c_i^2 \left(\frac{\sigma^2}{n_i}\right) = \sigma^2 \sum_{i=1}^{r} \frac{c_i^2}{n_i} \tag{17.21}$$

An unbiased estimator of this variance is:

$$s^2\{\hat{L}\} = MSE \sum_{i=1}^{r} \frac{c_i^2}{n_i} \tag{17.22}$$

$\hat{L}$ is normally distributed by (A.40) because it is a linear combination of independent normal random variables. It can be shown by theorem (17.6), the characteristics of $\hat{L}$ just mentioned, and the definition of t that:

$$\frac{\hat{L} - L}{s\{\hat{L}\}} \text{ is distributed as } t(n_T - r) \text{ for ANOVA model (17.1)} \tag{17.23}$$

Consequently, the $1 - \alpha$ confidence limits for L are:

$$\hat{L} \pm t(1 - \alpha/2; n_T - r)s\{\hat{L}\} \tag{17.24}$$

Testing. The confidence interval based on the limits in (17.24) can be used to test a hypothesis of the form:

$$\begin{aligned} H_0&: L = 0 \\ H_a&: L \neq 0 \end{aligned} \tag{17.25}$$

H_0 is concluded at the α level of significance if zero is contained in the interval; otherwise H_a is concluded. An equivalent procedure is based on the test statistic:

$$t^* = \frac{\hat{L}}{s\{\hat{L}\}} \tag{17.26}$$

If $|t^*| \leq t(1 - \alpha/2; n_T - r)$, H_0 is concluded; otherwise, H_a is concluded.

Example

In the Kenton Food Company example, the mean sales for the 3-color designs are to be compared to the mean sales for the 5-color designs with a 95 percent confidence interval. We wish to estimate:

$$L = \frac{\mu_1 + \mu_2}{2} - \frac{\mu_3 + \mu_4}{2}$$

The point estimate is (see data in Table 17.1):

$$\hat{L} = \frac{\bar{Y}_{1\cdot} + \bar{Y}_{2\cdot}}{2} - \frac{\bar{Y}_{3\cdot} + \bar{Y}_{4\cdot}}{2} = \frac{14.6 + 13.4}{2} - \frac{19.5 + 27.2}{2} = -9.35$$

Since $c_1 = 1/2$, $c_2 = 1/2$, $c_3 = -1/2$, and $c_4 = -1/2$, we obtain:

$$\sum \frac{c_i^2}{n_i} = \frac{(1/2)^2}{5} + \frac{(1/2)^2}{5} + \frac{(-1/2)^2}{4} + \frac{(-1/2)^2}{5} = .2125$$

and:

$$s^2\{\hat{L}\} = MSE \sum \frac{c_i^2}{n_i} = 10.55(.2125) = 2.242$$

so that $s\{\hat{L}\} = 1.50$.

For a 95 percent confidence interval, we require $t(.975; 15) = 2.131$. The confidence limits for L therefore are $-9.35 \pm 2.131(1.50)$, and the desired 95 percent confidence interval is:

$$-12.5 \leq L \leq -6.2$$

Therefore, we conclude with confidence coefficient .95 that mean sales for the 3-color designs fall below those for the 5-color designs by somewhere between 6.2 and 12.5 cases per store.

To test the hypothesis of no difference in mean sales for the 3-color and 5-color designs:

$$H_0: L = 0$$
$$H_a: L \neq 0$$

at the $\alpha = .05$ level of significance, we simply note that the hypothesized value zero is not contained in the 95 percent confidence interval. Hence, we conclude H_a, that the mean sales differ. To obtain a P-value of the test, test statistic (17.26) must be obtained. We find:

$$t^* = \frac{-9.35}{1.50} = -6.23$$

and the corresponding two-sided P-value is 0+.

Comment

Many single-factor analysis of variance programs permit the user to specify a contrast of interest and then will furnish the t^* test statistic or the equivalent F^* test statistic. ■

Inferences for Linear Combination of Factor Level Means

Occasionally, we are interested in a linear combination of the factor level means that is not a contrast. For example, suppose that the Kenton Food Company will use all four package designs, one in each of its four major marketing regions, and that these marketing regions account for 35, 28, 12, and 25 percent of sales, respectively. In that case, there might be interest in the overall mean sales per store for all regions:

$$L = .35\mu_1 + .28\mu_2 + .12\mu_3 + .25\mu_4$$

Note that this linear combination is of the form $L = \sum c_i \mu_i$ but that the coefficients c_i sum to 1.0, not to zero as they must for a contrast.

We define a *linear combination of the factor level means* μ_i as:

$$L = \sum_{i=1}^{r} c_i \mu_i \tag{17.27}$$

with no restrictions on the coefficients c_i. Confidence limits and test statistics for a linear combination L are obtained in exactly the same way as those for a contrast by means of (17.24) and (17.26), respectively. Point estimator (17.20) and estimated variance (17.22) are still applicable when $\sum c_i \neq 0$.

Single Degree of Freedom Tests. The alternatives for tests concerning a factor level mean in (17.8), a difference between two factor level means in (17.17a), and a contrast of factor level means in (17.25) are all special cases of a test concerning a linear combination of factor level means:

$$H_0: \sum c_i \mu_i = c$$
$$H_a: \sum c_i \mu_i \neq c$$

where the c_i and c are appropriate constants. Test statistics (17.9), (17.18), and (17.26) can each be converted to an equivalent F^* test statistic by means of the relation in (A.50a):

$$F^* = (t^*)^2$$

Test statistic F^* follows the $F(1, n_T - r)$ distribution when H_0 holds. Note that the numerator degrees of freedom are always one. Hence, these tests are often referred to as *single-degree-of-freedom tests*. The t^* version of the test statistic is more versatile because it can also be used for one-sided tests while the F^* version cannot.

17.4 Need for Simultaneous Inference Procedures

The procedures for estimating and testing factor level means discussed up to this point have two important limitations:

1. The confidence coefficient $1-\alpha$ for the estimation procedures described is a statement confidence coefficient and applies only to a particular estimate, not to a series of estimates. Similarly, the specified Type I error rate, α, applies only to a particular test and not to a series of tests.

2. The confidence coefficient $1-\alpha$ and the specified significance level α are appropriate only if the estimate or test was not suggested by the data.

The first limitation is familiar from regression analysis. It is particularly serious for analysis of variance models because frequently many different comparisons are of interest here, and one needs to piece the different findings together. Consider the very simple case where three different advertisements are being compared for their effectiveness in stimulating sales. The following estimates of their comparative effectiveness have been obtained, each with a 95 percent statement confidence coefficient:

$$59 \leq \mu_2 - \mu_1 \leq 62$$
$$-2 \leq \mu_3 - \mu_1 \leq 3$$
$$58 \leq \mu_2 - \mu_3 \leq 64$$

It would be natural here to piece the different comparisons together and conclude that advertisement 2 leads to highest mean sales, while advertisements 1 and 3 are substantially less effective and do not differ much among themselves. One would therefore like a family confidence coefficient for this family of statements, to provide known assurance that the set of conclusions is correct.

The same concern for assurance of correct conclusions exists when the inferences involve tests. An analysis of factor means by testing procedures usually involves several single-degree-of-freedom tests to answer related questions. For instance, the sales manager of the Kenton Food Company might wish to know both whether the number of colors has an effect on mean sales and whether the use of cartoons has an effect. Whenever several tests are conducted, both the level of significance and the power, insofar as the family of tests is concerned, are affected. Consider, for example, three different t tests, each conducted with $\alpha = .05$. The probability that each of the tests will lead to conclusion H_0 when indeed H_0 is correct in each case, assuming independence of the tests, is $(.95)^3 = .857$. Thus, the level of significance that at least one of the three tests leads to conclusion H_a when H_0 holds in each case would be $1 - .857 = .143$, not .05. We see then that the level of significance and power for a *family* of tests is not the same as that for an *individual* test. Actually, the t^* statistics are dependent when they all are based on the same sample data and use the same *MSE* value. It is often therefore more difficult to determine the actual level of significance and power for a family of tests.

The second limitation of the procedures for estimating or testing factor level means discussed so far, namely, that the estimate or test must not be suggested by the data, is an important one in exploratory investigations where many new questions are often suggested once the data are being analyzed. The process of studying effects suggested by the data is sometimes called *data snooping*. One form of data snooping is to investigate comparisons where the effect appears to be large from the sample data, for example, testing whether there is a difference between the two treatment means corresponding to the smallest and largest estimated factor level means $\bar{Y}_{i.}$. Choosing the test in this manner implies a larger significance level than the nominal level used in constructing the decision rule. For example, it can be shown for a study with six factor levels that if the analyst will always compare the smallest and largest estimated factor level means by using the confidence limits (17.16) with a 95 percent confidence coefficient, the interval estimate will not contain zero and therefore suggest a real effect 40 percent of the time when indeed there is no difference between any of the factor level means (Ref. 17.1). Hence, the α level for the test is .40, not .05. With a larger number of factor levels, the likelihood of an erroneous indication of a real effect, i.e., the actual α level, would be even greater. The reason for the higher actual level of significance here is that a family of tests is being conducted implicitly since the analyst does not know in advance which estimated factor level means will be the extreme ones. The situation here is analogous to that in Chapter 10 where the test to determine whether the largest absolute residual is an outlier considers the family of tests for each of the n residuals.

One solution to this problem of making comparisons that are suggested by initial analysis of the data is to use a multiple comparison procedure where the family of inferences includes all the possible inferences that can be anticipated to be of potential interest after the data are examined. For instance, in an investigation where five factor level means are being studied, it is decided in advance that principal interest is in three pairwise comparisons. However, it is also agreed that other pairwise comparisons that will appear interesting should be studied as well. In this case, the family of *all* pairwise comparisons can be used as the basis for obtaining an appropriate family confidence coefficient or significance level for the comparisons suggested by the data.

In the next three sections, we shall discuss three multiple comparison procedures for analysis of variance models that permit the family confidence coefficient and the family α risk to be controlled. Two of these procedures, the Tukey and Scheffé procedures, allow data snooping to be undertaken naturally without affecting the confidence coefficient or significance level. The other procedure, the Bonferroni procedure, is applicable only when the effects to be investigated are identified in advance of the study.

17.5 Tukey Multiple Comparison Procedure

The Tukey multiple comparison procedure that we will consider here applies when:

> The family of interest is the set of all pairwise comparisons of factor level means; in other words, the family consists of estimates of all pairs $D = \mu_i - \mu_{i'}$ or of all tests of the form:
>
> $$H_0\colon \mu_i - \mu_{i'} = 0$$
> $$H_a\colon \mu_i - \mu_{i'} \neq 0$$

When all sample sizes are equal, the family confidence coefficient for the Tukey method is exactly $1 - \alpha$ and the family significance level is exactly α. When the sample sizes are not equal, the family confidence coefficient is greater than $1 - \alpha$ and the family significance level is less than α. In other words, the Tukey procedure is conservative when the sample sizes are not equal.

Studentized Range Distribution

The Tukey procedure utilizes the *studentized range distribution.* Suppose that we have r independent observations $Y_1, \ldots, Y_r$ from a normal distribution with mean μ and variance σ^2. Let w be the range for this set of observations; thus:

$$w = \max(Y_l) - \min(Y_l) \tag{17.28}$$

Suppose further that we have an estimate s^2 of the variance σ^2 which is based on ν degrees of freedom and is independent of the Y_l. Then, the ratio w/s is called the *studentized range.* It is denoted by:

$$q(r, \nu) = \frac{w}{s} \tag{17.29}$$

where the arguments in parentheses remind us that the distribution of q depends on r and ν. The distribution of q has been tabulated, and selected percentiles are presented in Table B.9.

This table is simple to use. Suppose that $r = 5$ and $\nu = 10$. The 95th percentile is then $q(.95; 5, 10) = 4.65$, which means:

$$P\left\{\frac{w}{s} = q(5, 10) \leq 4.65\right\} = .95$$

Thus, with five normal Y observations, the probability is .95 that their range is not more than 4.65 times as great as an independent sample standard deviation based on 10 degrees of freedom.

Simultaneous Estimation

The Tukey multiple comparison confidence limits for all pairwise comparisons $D = \mu_i - \mu_{i'}$ with family confidence coefficient of at least $1 - \alpha$ are as follows:

$$\hat{D} \pm Ts\{\hat{D}\} \tag{17.30}$$

where:

$$\hat{D} = \overline{Y}_{i\cdot} - \overline{Y}_{i'\cdot} \tag{17.30a}$$

$$s^2\{\hat{D}\} = s^2\{\overline{Y}_{i\cdot}\} + s^2\{\overline{Y}_{i'\cdot}\} = MSE\left(\frac{1}{n_i} + \frac{1}{n_{i'}}\right) \tag{17.30b}$$

$$T = \frac{1}{\sqrt{2}}q(1 - \alpha; r, n_T - r) \tag{17.30c}$$

Note that the point estimator $\hat{D}$ in (17.30a) and the estimated variance in (17.30b) are the same as those in (17.11) and (17.14) for a single pairwise comparison. Thus, the only difference between the Tukey confidence limits (17.30) for simultaneous comparisons and those in (17.16) for a single comparison is the multiple of the estimated standard deviation.

The family confidence coefficient $1 - \alpha$ pertaining to the multiple pairwise comparisons refers to the proportion of correct families, each consisting of all pairwise comparisons, when repeated sets of samples are selected and all pairwise confidence intervals are calculated each time. A family of pairwise comparisons is considered to be correct if every pairwise comparison in the family is correct. Thus, a family confidence coefficient of $1 - \alpha$ indicates that all pairwise comparisons in the family will be correct in $(1 - \alpha)100$ percent of the repetitions.

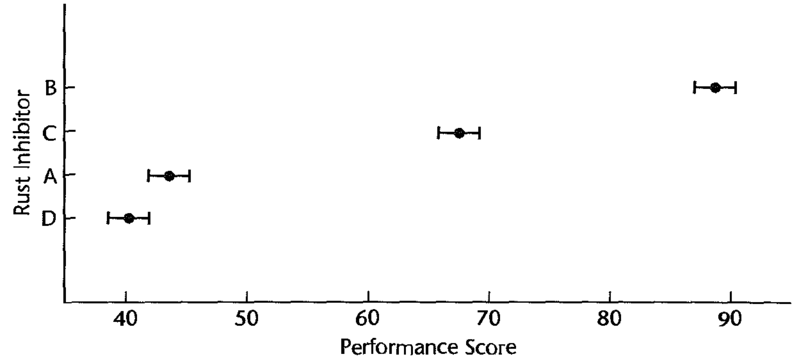
Simultaneous Testing

When we wish to conduct a family of tests of the form:

$$\begin{aligned} H_0&: \mu_i - \mu_{i'} = 0 \\ H_a&: \mu_i - \mu_{i'} \neq 0 \end{aligned} \tag{17.31}$$

for all pairwise comparisons, the family of confidence intervals based on (17.30) may be utilized for this purpose. We simply determine for each interval whether or not zero is contained in the interval. If zero is contained, conclusion H_0 is reached; otherwise, H_a is concluded. By following this procedure, the family level of significance will not exceed α.

FIGURE 17.4 Paired Comparison Plot—Rust Inhibitor Example.



Equivalently, the pairwise tests can be conducted directly by calculating for each pairwise comparison the test statistic:

$$q^* = \frac{\sqrt{2}\hat{D}}{s\{\hat{D}\}} \tag{17.32}$$

where $\hat{D}$ and $s^2\{\hat{D}\}$ are given in (17.30). Conclusion H_0 in (17.31) is reached if $|q^*| \leq q(1-\alpha; r; n_T - r)$; otherwise, H_a is concluded.

A *paired comparison plot* provides still another means of conducting all pairwise tests with the Tukey procedure when all sample sizes are equal, i.e., when $n_i \equiv n$. This plot provides a graphic means of making all pairwise comparisons. Around each estimated treatment mean $\overline{Y}_{i.}$ is plotted an interval whose limits are:

$$\overline{Y}_{i.} \pm \frac{1}{2}Ts\{\hat{D}\} \tag{17.33}$$

When the intervals overlap on this plot, the formal test leads to the conclusion that the two treatment means do not differ. When the intervals do not overlap, the formal test leads to the conclusion that the two treatment means differ. In addition, the paired comparison plot shows the direction of the difference.

Figure 17.4 provides an illustration of a paired comparison plot for the rust inhibitor example. There is no overlap between the intervals for rust inhibitors B and C, indicating that the mean performances differ for these two rust inhibitors. Figure 17.4 in addition shows that rust inhibitor B is superior to C since its interval is considerably to the right of that for C, thus providing directional information about the difference in mean performance for the two rust inhibitors. We discuss this plot in greater detail on page 750.

Example 1—Equal Sample Sizes

In the rust inhibitor example in Table 17.2, it was desired to estimate all pairwise comparisons by means of the Tukey procedure, using a family confidence coefficient of 95 percent. Since $r = 4$ and $n_T - r = 36$, we find the required percentile of the studentized range distribution from Table B.9 to be $q(.95; 4, 36) = 3.814$. Hence, by (17.30c), we obtain:

$$T = \frac{1}{\sqrt{2}}(3.814) = 2.70$$

TABLE 17.3 Simultaneous Confidence Intervals and Tests for Pairwise Differences Using the Tukey Procedure—Rust Inhibitor Example.

	Test		
Confidence Interval	H_0	H_a	q^*
$43.3 \leq \mu_2 - \mu_1 \leq 49.3$	$\mu_2 = \mu_1$	$\mu_2 \neq \mu_1$	58.99
$21.8 \leq \mu_3 - \mu_1 \leq 27.8$	$\mu_3 = \mu_1$	$\mu_3 \neq \mu_1$	31.61
$-.3 \leq \mu_1 - \mu_4 \leq 5.7$	$\mu_1 = \mu_4$	$\mu_1 \neq \mu_4$	3.40
$18.5 \leq \mu_2 - \mu_3 \leq 24.5$	$\mu_2 = \mu_3$	$\mu_2 \neq \mu_3$	27.37
$46.0 \leq \mu_2 - \mu_4 \leq 52.0$	$\mu_2 = \mu_4$	$\mu_2 \neq \mu_4$	62.39
$24.5 \leq \mu_3 - \mu_4 \leq 30.5$	$\mu_3 = \mu_4$	$\mu_3 \neq \mu_4$	35.01

Further, we need $s\{\hat{D}\}$. Using (17.30b), we find for any pairwise comparison since equal sample sizes were employed:

$$s^2\{\hat{D}\} = MSE\left(\frac{1}{n_i} + \frac{1}{n_{i'}}\right) = 6.140\left(\frac{1}{10} + \frac{1}{10}\right) = 1.23$$

so that $s\{\hat{D}\} = 1.11$. Hence, we obtain for each pairwise comparison:

$$Ts\{\hat{D}\} = 2.70(1.11) = 3.0$$

To illustrate the calculation of the pairwise confidence limits, consider the estimation of the difference between the treatment means for rust inhibitors A and B, $\mu_2 - \mu_1$:

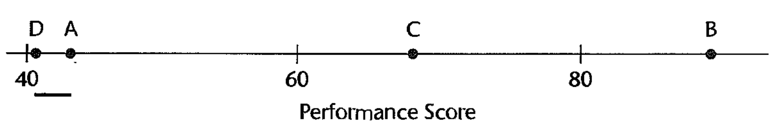
$$\hat{D} = \bar{Y}_{2.} - \bar{Y}_{1.} = 89.44 - 43.14 = 46.3$$

The confidence limits from (17.30) therefore are 46.3 ± 3.0 and the confidence interval is:

$$43.3 \leq \mu_2 - \mu_1 \leq 49.3$$

The complete family of pairwise confidence intervals is listed in the left column of Table 17.3. The pairwise comparisons indicate that all but one of the differences (D and A) are statistically significant (confidence interval does not cover zero).

We incorporate this information in a line plot of the estimated factor level means by underlining nonsignificant comparisons.



The line between D and A indicates that there is no clear evidence whether D or A is the better rust inhibitor. The absence of a line signifies that a difference in performance has been found and the location of the points indicates the direction of the difference. Thus, the multiple comparison procedure permits us to infer with a 95 percent family confidence coefficient for the chain of conclusions that B is the best inhibitor (better by somewhere between 18.5 and 24.5 units than the second best), C is second best, and A and D follow substantially behind with little or no difference between them.

The same conclusions are obtained if we carry out all pairwise tests using the simultaneous testing procedure based on test statistic (17.32). For example, to test:

$$H_0: \mu_2 - \mu_1 = 0$$
$$H_a: \mu_2 - \mu_1 \neq 0$$

we require the test statistic:

$$q^* = \frac{\sqrt{2}(89.44 - 43.14)}{1.11} = 58.99$$

Because $|q^*| = 58.99 > q(.95; 4, 36) = 3.814$, we conclude H_a, that the two treatment means differ. The test statistics q^* for the family of all pairwise tests are listed in the right column of Table 17.3. The absolute values of all test statistics exceed 3.814 except for one, so that all differences are found to be statistically significant except for that involving μ_1 and μ_4 (A and D). For this case, $|q^*| = 3.40$ does not exceed the critical value 3.814.

Figure 17.4 presents a paired comparison plot for the rust inhibitor example. Here are plotted the estimated treatment means $\bar{Y}_{i.}$ with the comparison intervals based on (17.33). For example, for rust inhibitor A, we have from earlier:

$$\bar{Y}_{1.} = 43.14 \qquad T = 2.70 \qquad s\{\hat{D}\} = 1.11$$

so that the comparison limits in (17.33) are:

$$43.14 \pm \frac{1}{2}(2.70)(1.11) \qquad \text{or} \qquad 41.64 \quad \text{and} \quad 44.64$$

We readily see that only the intervals for A and D overlap, that rust inhibitor B is clearly best, that rust inhibitor C is second best, and that rust inhibitors A and D are the least effective.

Example 2—Unequal Sample Sizes

In the Kenton Food Company example in Table 17.1, the sales manager was interested in the comparative performance of the four package designs. The analyst developed all pairwise comparisons by means of the Tukey procedure with a family confidence coefficient of at least 90 percent. Since the sample sizes are not equal here, the estimated standard deviation $s\{\hat{D}\}$ must be recalculated for each pairwise comparison. To compare designs 1 and 2, for instance, we obtain:

$$\hat{D} = \bar{Y}_{1.} - \bar{Y}_{2.} = 14.6 - 13.4 = 1.2$$

$$s^2\{\hat{D}\} = MSE\left(\frac{1}{n_1} + \frac{1}{n_2}\right) = 10.55\left(\frac{1}{5} + \frac{1}{5}\right) = 4.22$$

$$s\{\hat{D}\} = 2.05$$

For a 90 percent family confidence coefficient, we require $q(.90; 4, 15) = 3.54$ so that we obtain:

$$T = \frac{1}{\sqrt{2}}(3.54) = 2.50$$

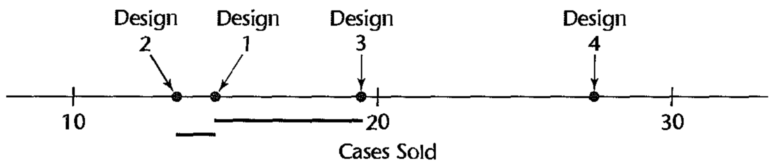
Hence, the confidence limits are $1.2 \pm 2.50(2.05)$ and the confidence interval for $\mu_1 - \mu_2$ is:

$$-3.9 \leq \mu_1 - \mu_2 \leq 6.3$$

In the same way, we obtain the other five confidence intervals:

$$-.6 = (19.5 - 14.6) - 2.50(2.18) \leq \mu_3 - \mu_1 \leq (19.5 - 14.6) + 2.50(2.18) = 10.4$$
$$7.5 = (27.2 - 14.6) - 2.50(2.05) \leq \mu_4 - \mu_1 \leq (27.2 - 14.6) + 2.50(2.05) = 17.7$$
$$.7 = (19.5 - 13.4) - 2.50(2.18) \leq \mu_3 - \mu_2 \leq (19.5 - 13.4) + 2.50(2.18) = 11.6$$
$$8.7 = (27.2 - 13.4) - 2.50(2.05) \leq \mu_4 - \mu_2 \leq (27.2 - 13.4) + 2.50(2.05) = 18.9$$
$$2.3 = (27.2 - 19.5) - 2.50(2.18) \leq \mu_4 - \mu_3 \leq (27.2 - 19.5) + 2.50(2.18) = 13.2$$

We summarize the comparative performance by a line plot, indicating each nonsignificant difference by a rule.



We can conclude with at least 90 percent family confidence that design 4 is clearly the most effective design. However, the small-scale study does not permit a complete ordering among the other three designs. Design 3 is more effective than design 2 but may not be more effective than design 1, which in turn may not be more effective than design 2.

Often, the results of the family of pairwise tests are summarized by setting up groups of factor levels whose means do not differ according to the single degree of freedom tests. For the Kenton Food Company example, there are three such groups:

Group 1		Group 2		Group 3	
Design 4	$\bar{Y}_{4\cdot} = 27.2$	Design 3	$\bar{Y}_{3\cdot} = 19.5$	Design 1	$\bar{Y}_{1\cdot} = 14.6$
		Design 1	$\bar{Y}_{1\cdot} = 14.6$	Design 2	$\bar{Y}_{2\cdot} = 13.4$

Comments

1. When the Tukey procedure is used with unequal sample sizes, it is sometimes called the *Tukey-Kramer procedure*.

2. When not all pairwise comparisons are of interest, the confidence coefficient for the family of comparisons under consideration will be greater than the specification $1 - \alpha$ used in setting up the Tukey intervals. Similarly, the family significance level for simultaneous testing will be less than α.

3. The Tukey procedure can be used for data snooping as long as the effects to be studied on the basis of preliminary data analysis are pairwise comparisons.

4. The Tukey procedure can be modified to handle general contrasts of factor level means. We do not discuss this modification since the Scheffé method (to be discussed next) is to be preferred for this situation.

5. To derive the Tukey simultaneous confidence intervals for the case when all sample sizes are equal, i.e., when $n_i \equiv n$ so that $n_T = rn$, consider the deviations:

$$(\bar{Y}_{1\cdot} - \mu_1), \ldots, (\bar{Y}_{r\cdot} - \mu_r) \tag{17.34}$$

and assume that ANOVA model (17.1) applies. The deviations in (17.34) are then independent variables (because the error terms are independent), they are normally distributed (because the error terms are independent normal variables), they have the same expectation zero (because μ_i is subtracted from $\bar{Y}_{i\cdot}$), and they have the same variance σ^2/n. Further. MSE/n is an estimator of σ^2/n that is independent of the deviations $(\bar{Y}_{i\cdot} - \mu_i)$ per theorem (17.6). Thus, it follows from the definition of the studentized range q in (17.29) that:

$$\frac{\max(\bar{Y}_{i\cdot} - \mu_i) - \min(\bar{Y}_{i\cdot} - \mu_i)}{\sqrt{\dfrac{MSE}{n}}} \sim q(r, n_T - r) \tag{17.35}$$

where $n_T - r$ is the number of degrees of freedom associated with MSE, $\max(\bar{Y}_{i\cdot} - \mu_i)$ is the largest deviation, and $\min(\bar{Y}_{i\cdot} - \mu_i)$ is the smallest deviation.

In view of (17.35), we can write the following probability statement:

$$P\left\{\frac{\max(\bar{Y}_{i\cdot} - \mu_i) - \min(\bar{Y}_{i\cdot} - \mu_i)}{\sqrt{\dfrac{MSE}{n}}} \leq q(1 - \alpha; r, n_T - r)\right\} = 1 - \alpha \tag{17.36}$$

Note now that the following inequality holds for *all* pairs of factor levels i and i':

$$|(\bar{Y}_{i\cdot} - \mu_i) - (\bar{Y}_{i'\cdot} - \mu_{i'})| \leq \max(\bar{Y}_{i\cdot} - \mu_i) - \min(\bar{Y}_{i\cdot} - \mu_i) \tag{17.37}$$

The absolute value at the left is needed since the factor levels i and i' are not ordered so that we may be subtracting the larger deviation from the smaller. To put this another way, we are merely concerned here with the difference between the two factor level deviations regardless of direction.

Since inequality (17.37) holds for all pairs of factor levels i and i', it follows from (17.36) that the probability:

$$P\left\{\left|\frac{(\bar{Y}_{i\cdot} - \mu_i) - (\bar{Y}_{i'\cdot} - \mu_{i'})}{\sqrt{\dfrac{MSE}{n}}}\right| \leq q(1 - \alpha; r, n_T - r)\right\} = 1 - \alpha \tag{17.38}$$

holds for all $r(r-1)/2$ pairwise comparisons among the r factor levels. By rearranging the inequality in (17.38), using the definitions of $s^2\{\hat{D}\}$ in (17.30b) and of T in (17.30c), and noting that for the equal sample size case $s^2\{\hat{D}\}$ becomes:

$$s^2\{\hat{D}\} = MSE\left(\frac{1}{n} + \frac{1}{n}\right) = \frac{2MSE}{n} \qquad \text{when } n_i \equiv n$$

we obtain the Tukey multiple comparison confidence limits in (17.30).

6. When the Tukey multiple comparison procedure is used for testing pairwise differences as in (17.31), the tests are sometimes called *honestly significant difference tests*.

7. The pairwise comparison plot can be used as an approximate plot when the sample sizes are not equal, provided that the sample sizes do not differ greatly. For this case. the comparison limits

should be obtained as follows:

$$\bar{Y}_{i\cdot} \pm \frac{1}{2}q(1-\alpha; r, n_T - r)s\{\bar{Y}_{i\cdot}\} \tag{17.39}$$

The limits in (17.39) are identical to those in (17.33) when the sample sizes are equal. ■

17.6 Scheffé Multiple Comparison Procedure

The Scheffé multiple comparison procedure was encountered previously for regression models. It is also applicable for analysis of variance models. It applies for analysis of variance models when:

The family of interest is the set of all possible contrasts among the factor level means:

$$L = \sum c_i \mu_i \qquad \text{where } \sum c_i = 0 \tag{17.40}$$

In other words, the family consists of estimates of all possible contrasts L or of tests concerning all possible contrasts of the form:

$$H_0: L = 0$$
$$H_a: L \neq 0$$

Thus, infinitely many statements belong to this family. The family confidence level for the Scheffé procedure is exactly $1-\alpha$, and the family significance level is exactly α, whether the factor level sample sizes are equal or unequal.

Simultaneous Estimation

We noted earlier that an unbiased estimator of L is:

$$\hat{L} = \sum c_i \bar{Y}_{i\cdot} \tag{17.41}$$

for which the estimated variance is:

$$s^2\{\hat{L}\} = MSE \sum \frac{c_i^2}{n_i} \tag{17.42}$$

The Scheffé confidence intervals for the family of contrasts L are of the form:

$$\hat{L} \pm Ss\{\hat{L}\} \tag{17.43}$$

where:

$$S^2 = (r-1)F(1-\alpha; r-1, n_T - r) \tag{17.43a}$$

and $\hat{L}$ and $s\{\hat{L}\}$ are given by (17.41) and (17.42), respectively. If we were to calculate the confidence intervals in (17.43) for all conceivable contrasts, then in $(1-\alpha)100$ percent of repetitions of the experiment, the entire set of confidence intervals in the family would be correct.

Note that the simultaneous confidence limits in (17.43) differ from those for a single confidence limit in (17.24) only with respect to the multiple of the estimated standard deviation.

Simultaneous Testing

Tests involving contrasts of the form:

$$H_0: L = 0$$
$$H_a: L \neq 0 \tag{17.44}$$

can be carried out by examination of the corresponding Scheffé confidence intervals based on (17.43). H_0 is concluded at the α family level of significance if the confidence interval includes zero; otherwise H_a is concluded. An equivalent direct testing procedure for the alternatives in (17.44) uses the test statistic:

$$F^* = \frac{\hat{L}^2}{(r-1)s^2\{\hat{L}\}} \tag{17.45}$$

Conclusion H_0 in (17.44) is reached at the α family significance level if $F^* \leq F(1-\alpha; r-1, n_T - r)$; otherwise, H_a is concluded.

Example

In the Kenton Food Company example, interest centered on estimating the following four contrasts with family confidence coefficient .90:

Comparison of 3-color and 5-color designs:

$$L_1 = \frac{\mu_1 + \mu_2}{2} - \frac{\mu_3 + \mu_4}{2}$$

Comparison of designs with and without cartoons:

$$L_2 = \frac{\mu_1 + \mu_3}{2} - \frac{\mu_2 + \mu_4}{2}$$

Comparison of the two 3-color designs:

$$L_3 = \mu_1 - \mu_2$$

Comparison of the two 5-color designs:

$$L_4 = \mu_3 - \mu_4$$

Consider first the estimation of L_1. Earlier, we found:

$$\hat{L}_1 = -9.35$$
$$s\{\hat{L}_1\} = 1.50$$

Since $r - 1 = 3$ and $n_T - r = 15$ (Table 17.1), we have:

$$S^2 = (r-1)F(1-\alpha; r-1, n_T - r) = 3F(.90; 3, 15) = 3(2.49) = 7.47$$

so that $S = 2.73$. Hence, the 90 percent confidence limits for L_1 by the Scheffé multiple comparison procedure are $-9.35 \pm 2.73(1.50)$ and the desired confidence interval is:

$$-13.4 \leq L_1 \leq -5.3$$

In similar fashion, we obtain the other desired confidence intervals, and the entire set is:

$$-13.4 \leq L_1 \leq -5.3$$
$$-7.3 \leq L_2 \leq .8$$
$$-4.4 \leq L_3 \leq 6.8$$
$$-13.7 \leq L_4 \leq -1.7$$

Note that the confidence interval for L_1 does not include zero. Hence, if we wished to test H_0: $L_1 = 0$ versus H_a: $L_1 \neq 0$, we would conclude H_a, that the mean sales for 3-color and 5-color designs differ. The confidence interval provides additional information, however; namely, that mean sales for 5-color designs exceed mean sales for 3-color designs, by somewhere between 5.3 and 13.4 cases per store.

Any chain of conclusions derived from the set of confidence intervals has associated with it family confidence coefficient .90. The principal conclusions drawn by the sales manager were as follows: 5-color designs lead to higher mean sales than 3-color designs, the increase being somewhere between 5 and 13 cases per store. No overall effect of cartoons in the package design is indicated, although the use of a cartoon in 5-color designs leads to lower mean sales than when no cartoon is used.

Comments

1. If in the Kenton Food Company example we had wished to estimate a single contrast with statement confidence coefficient .90, the required t value would have been $t(.95; 15) = 1.753$. This t value is smaller than the Scheffé multiple $S = 2.73$, so that the single confidence interval would be somewhat narrower. The increased width of the interval with the Scheffé procedure is the price paid for a known confidence coefficient for a family of statements and a chain of conclusions drawn from them, and for the possibility of making comparisons not specified in advance of the data analysis.

2. Since applications of the Scheffé procedure never involve all conceivable contrasts, the confidence coefficient for the finite family of statements actually considered will be greater than $1 - \alpha$ so that $1 - \alpha$ serves as a guaranteed lower bound. Similarly, the significance level for the finite family of tests considered will be less than α. For this reason, it has been suggested that lower confidence levels and higher significance levels be used with the Scheffé procedure than would ordinarily be employed. Confidence coefficients of 90 percent and 95 percent and significance levels of $\alpha = .10$ and $\alpha = .05$ with the Scheffé procedure are frequently mentioned.

3. The Scheffé procedure can be used for a wide variety of data snooping since the family of statements contains all possible contrasts. ■

Comparison of Scheffé and Tukey Procedures

1. If only pairwise comparisons are to be made, the Tukey procedure gives narrower confidence limits and is therefore the preferred method.

2. The Scheffé procedure has the property that if the F test of factor level equality indicates that the factor level means μ_i are not equal, the corresponding Scheffé multiple comparison procedure will find at least one contrast (out of all possible contrasts) that differs significantly from zero (the confidence interval does not cover zero). It may be, though, that this contrast is not one of those that has been estimated.

17.7 Bonferroni Multiple Comparison Procedure

The Bonferroni multiple comparison procedure was encountered earlier for regression models. It is also applicable for analysis of variance models when:

> The family of interest is a particular set of pairwise comparisons, contrasts, or linear combinations that is specified by the user in advance of the data analysis.

The Bonferroni procedure is applicable whether the factor level sample sizes are equal or unequal and whether inferences center on pairwise comparisons, contrasts, linear combinations, or a mixture of these.

Simultaneous Estimation

We shall denote the number of statements in the family by g and treat them all as linear combinations since pairwise comparisons and contrasts are special cases of linear combinations. The Bonferroni inequality (4.4) then implies that the confidence coefficient is at least $1 - \alpha$ that the following confidence limits for the g linear combinations L are all correct:

$$\hat{L} \pm Bs\{\hat{L}\} \tag{17.46}$$

where:

$$B = t(1 - \alpha/2g; n_T - r) \tag{17.46a}$$

Simultaneous Testing

When we wish to conduct a series of tests of the form:

$$H_0\colon L = 0$$
$$H_a\colon L \neq 0$$

we can use either the confidence intervals based on (17.46) or the test statistics:

$$t^* = \frac{\hat{L}}{s\{\hat{L}\}} \tag{17.47}$$

If $|t^*| \leq t(1 - \alpha/2g; n_T - r)$, we conclude H_0; otherwise, H_a is concluded.

Example

The sales manager of the Kenton Food Company is interested in estimating the following two contrasts with family confidence coefficient .975:

Comparison of 3-color and 5-color designs:

$$L_1 = \frac{\mu_1 + \mu_2}{2} - \frac{\mu_3 + \mu_4}{2}$$

Comparison of designs with and without cartoons:

$$L_2 = \frac{\mu_1 + \mu_3}{2} - \frac{\mu_2 + \mu_4}{2}$$

Earlier we found:

$$\hat{L}_1 = -9.35 \qquad s\{\hat{L}_1\} = 1.50$$
$$\hat{L}_2 = -3.25 \qquad s\{\hat{L}_2\} = 1.50$$

For a 97.5 percent family confidence coefficient with the Bonferroni method, we require:

$$B = t[1 - .025/2(2); 15] = t(.99375; 15) = 2.84$$

We can now complete the confidence intervals for the two contrasts. For L_1, we have confidence limits $-9.35 \pm 2.84(1.50)$, which lead to the confidence interval:

$$-13.6 \leq L_1 \leq -5.1$$

Similarly, we obtain the other confidence interval:

$$-7.5 \leq L_2 \leq 1.0$$

These confidence intervals have a guaranteed family confidence coefficient of 97.5 percent, which means that in at least 97.5 percent of repetitions of the experiment, both intervals will be correct.

Again, we would conclude from this family of estimates that mean sales for 5-color designs are higher than those for 3-color designs (by somewhere between 5 and 14 cases per store), and that no overall effect of cartoons in the package design is indicated.

The Scheffé multiple for a 97.5 percent family confidence coefficient in this case would have been:

$$S^2 = 3F(.975; 3, 15) = 3(4.15) = 12.45$$

or $S = 3.53$, as compared to the Bonferroni multiple $B = 2.84$. Thus, the Scheffé procedure here would have led to wider confidence intervals than the Bonferroni procedure.

Comment

It is not necessary that all comparisons be estimated with statement confidence coefficients $1 - \alpha/g$ for the Bonferroni family confidence coefficient to be $1 - \alpha$. Different statement confidence coefficients may be used, depending upon the importance of each statement, provided that $\alpha_1 + \alpha_2 + \cdots + \alpha_g = \alpha$. ■

Comparison of Bonferroni Procedure with Scheffé and Tukey Procedures

1. If all pairwise comparisons are of interest, the Tukey procedure is superior to the Bonferroni procedure, leading to narrower confidence intervals. If not all pairwise comparisons are to be considered, the Bonferroni procedure may be the better one at times.

2. The Bonferroni procedure will be better than the Scheffé procedure when the number of contrasts of interest is about the same as the number of factor levels, or less. Indeed, the number of contrasts of interest must exceed the number of factor levels by a considerable amount before the Scheffé procedure becomes better.

3. All three procedures are of the form "estimator $\pm$ multiplier $\times$ SE." The only difference among the three procedures is the multiplier. In any given problem, one may compute the Bonferroni multiple as well as the Scheffé multiple and, when appropriate, the Tukey multiple, and select the one that is smallest. This choice is proper since it does not depend on the observed data.

4. The Bonferroni multiple comparison procedure does not lend itself to data snooping unless one can specify in advance the family of inferences in which one may be interested

and provided this family is not large. On the other hand, the Tukey and Scheffé procedures involve families of inferences that lend themselves naturally to data snooping.

5. Other specialized multiple comparison procedures have been developed. For example, Dunnett's procedure (Ref. 17.2) performs pairwise comparisons of each treatment against a control treatment only whereas Hsu's procedure (Ref. 17.3) selects the "best" treatment and identifies those treatments that are worse than the "best."

Analysis of Means

One use of the Bonferroni simultaneous testing procedure is in the analysis of means (ANOM), introduced by Ott (Ref. 17.4). ANOM is an alternative to the standard F test for the equality of treatment means. It is conducted by testing H_0: $\tau_1 = 0$ versus H_a: $\tau_1 \neq 0$, H_0: $\tau_2 = 0$ versus H_a: $\tau_2 \neq 0$, and so on for all treatment effects τ_i. The statistics employed are the r estimated treatment effects defined in (16.75b):

$$\hat{\tau}_i = \bar{Y}_{i\cdot} - \hat{\mu}_{\cdot} \qquad i = 1, \ldots, r \tag{17.48}$$

where $\hat{\mu}_{\cdot}$ is the least squares mean given in (16.75a):

$$\hat{\mu}_{\cdot} = \frac{\sum \bar{Y}_{i\cdot}}{r} \tag{17.48a}$$

The estimated variance of $\hat{\tau}_i$ is obtained by (17.22) since $\hat{\tau}_i$ is a contrast of the estimated treatment means $\bar{Y}_{i\cdot}$:

$$s^2\{\hat{\tau}_i\} = \frac{MSE}{n_i}\left(\frac{r-1}{r}\right)^2 + \frac{MSE}{r^2}\sum_{h\neq i}\frac{1}{n_h} \tag{17.49}$$

Simultaneous testing by the Bonferroni procedure can be carried out by setting up for each treatment effect the confidence interval using (17.46) and noting whether or not the interval contains zero. The results are sometimes summarized in an *analysis of means plot*. It is easy to show that a contrast $\hat{\tau}_i = \bar{Y}_{i\cdot} - \hat{\mu}_{\cdot}$ is inside (outside) one of the Bonferroni contrast intervals whenever the cell mean $\bar{Y}_{i\cdot}$ is inside (outside) the limits $\hat{\mu}_{\cdot} \pm t(1 - \alpha/2r; n_T - r)s\{\hat{\tau}_i\}$. In an analysis of means plot, the cell means are plotted along with the indicated limits and the least squares mean $\hat{\mu}_{\cdot}$ in (17.48a). If any of the cell means fall above (below) these limits, the conclusion is drawn that the cell mean is larger (smaller) than the overall mean.

ANOM is similar to ANOVA for detecting the differences between cell means. However, an important difference between ANOVA and ANOM is that the former tests whether the cell means are different from each other, whereas the latter tests whether the cell means are different from the overall mean. Various enhancements for the analysis of means have been provided, including those in References 17.5 and 17.6.

Example

In Figure 17.5 we present a MINITAB ANOM plot for the Kenton Food Company example using $\alpha = .05$. We conclude that the mean of sales for design 4 is greater than the overall unweighted mean (16.63), while the mean of sales for both design 1 and design 2 are less than the overall unweighted mean. Note that MINITAB bases its ANOM procedure on the weighted mean $\hat{\mu}_{\cdot} = \bar{Y}_{\cdot\cdot}$, rather than the least squares mean in (17.48a).

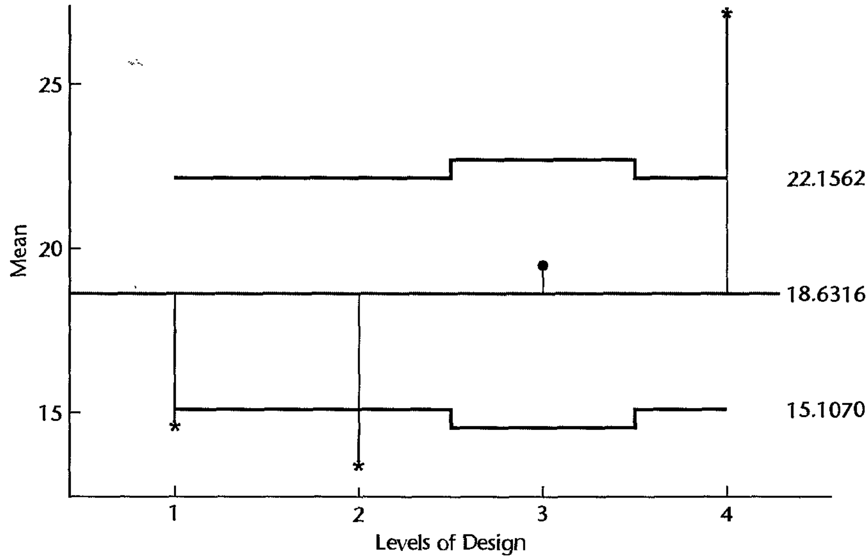


FIGURE 17.5 Analysis of Means Plot—Kenton Food Company Example.

17.8 Planning of Sample Sizes with Estimation Approach

In Section 16.10 we considered the planning of sample sizes using the power approach. We now take up another approach, the estimation approach to planning sample sizes, which may be used either in conjunction with the control of Type I and Type II errors or by itself. The essence of the approach is to specify the major comparisons of interest and to determine the expected widths of the confidence intervals for various sample sizes, given an advance planning value for the standard deviation σ. The approach is iterative, starting with an initial judgment of needed sample sizes. This initial judgment may be based on the needed sample sizes to control the risks of Type I and Type II errors when these have been obtained previously. If the anticipated widths of the confidence intervals based on the initial sample sizes are satisfactory, the iteration process is terminated. If one or more widths are too great, larger sample sizes need to be tried next. If the widths are narrower than they need be, smaller sample sizes should be tried next. This process is continued until those sample sizes are found that yield satisfactory anticipated widths for the important confidence intervals. We proceed to illustrate the estimation approach to planning sample sizes with two examples.

Example 1—Equal Sample Sizes

We are to plan sample sizes for the snow tires example discussed in Section 16.10 by means of the estimation approach; the sample sizes for each tire brand are to be equal, that is, $n_i \equiv n$. Management wishes three types of estimates:

1. A comparison of the mean tread lives for each pair of brands:

$$\mu_i - \mu_{i'}$$

2. A comparison of the mean tread lives for the two high-priced brands (1 and 4) and the two low-priced brands (2 and 3):

$$\frac{\mu_1 + \mu_4}{2} - \frac{\mu_2 + \mu_3}{2}$$

3. A comparison of the mean tread lives for the national brands (1, 2, and 4) and the local brand (3):

$$\frac{\mu_1 + \mu_2 + \mu_4}{3} - \mu_3$$

Management further has indicated that it wishes a family confidence coefficient of .95 for the entire set of comparisons.

We first need a planning value for the standard deviation of the tread lives of tires. Suppose that from past experience we judge the standard deviation to be approximately $\sigma = 2$ (thousand miles). Next, we require an initial judgment of needed sample sizes and shall consider $n = 10$ as a starting point.

We know from (17.21) that the variance of an estimated contrast $\hat{L}$ when $n_i \equiv n$ is:

$$\sigma^2\{\hat{L}\} = \frac{\sigma^2}{n} \sum c_i^2 \qquad \text{when } n_i \equiv n$$

Hence, given $\sigma = 2$ and $n = 10$, the anticipated values of the standard deviations of the required estimators are:

Contrast	Anticipated Variance	Anticipated Standard Deviation
Pairwise comparisons	$\frac{(2)^2}{10}[(1)^2 + (-1)^2] = .80$	.89
High- and low-priced brands	$\frac{(2)^2}{10}\left[\left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^2 + \left(-\frac{1}{2}\right)^2 + \left(-\frac{1}{2}\right)^2\right] = .40$	.63
National and local brands	$\frac{(2)^2}{10}\left[\left(\frac{1}{3}\right)^2 + \left(\frac{1}{3}\right)^2 + \left(\frac{1}{3}\right)^2 + (-1)^2\right] = .53$	.73

We shall employ the Scheffé multiple comparison procedure and therefore require the Scheffé multiple S in (17.43a) for $r = 4$, $n_T = 10(4) = 40$, and $1 - \alpha = .95$:

$$S^2 = (r-1)F(1-\alpha; r-1, n_T - r) = 3F(.95; 3, 36) = 3(2.87) = 8.61$$

or $S = 2.93$. Hence, the anticipated widths of the confidence intervals are:

Contrast	Anticipated Width of Confidence Interval $= \pm S\sigma\{\hat{L}\}$
Pairwise comparisons	$\pm 2.93(.89) = \pm 2.61$ (thousand miles)
High- and low-priced brands	$\pm 2.93(.63) = \pm 1.85$ (thousand miles)
National and local brands	$\pm 2.93(.73) = \pm 2.14$ (thousand miles)

Management was satisfied with these anticipated widths. However, it was decided to increase the sample sizes from 10 to 15 in case the actual standard deviation of the tread lives of tires is somewhat greater than the anticipated value $\sigma = 2$ (thousand miles).

Example 2—Unequal Sample Sizes

In the snow tires example, suppose that tire brand 4 is the snow tire presently used and is to serve as the basis of comparison for the other brands. The comparisons of interest therefore are $\mu_1 - \mu_4$, $\mu_2 - \mu_4$, and $\mu_3 - \mu_4$. The sample size for brand 4 is to be twice as large as for the other brands in order to improve the precision of the three pairwise comparisons. The desired precision, with a family confidence coefficient of .90, is to be ± 1 (thousand miles). The Bonferroni procedure will be used to provide assurance as to the family confidence level.

We know from (17.13) that the variance of an estimated difference $\hat{L}_i = \bar{Y}_{i\cdot} - \bar{Y}_{4\cdot}$ (the difference is now denoted more generally by $\hat{L}$) is for $i = 1, 2, 3$:

$$\sigma^2\{\hat{L}_i\} = \sigma^2\left(\frac{1}{n_i} + \frac{1}{n_4}\right)$$

We shall denote the sample sizes for brands 1, 2, and 3 by n and for brand 4 by $2n$. Hence, the variance of $\hat{L}_i$ becomes:

$$\sigma^2\{\hat{L}_i\} = \sigma^2\left(\frac{1}{n} + \frac{1}{2n}\right) = \frac{3\sigma^2}{2n}$$

Using again the planning value $\sigma = 2$ and an initial sample size $n = 10$, we find $\sigma^2\{\hat{L}_i\} = .60$ and $\sigma\{\hat{L}_i\} = .77$. For $\alpha = .10$ and $g = 3$ comparisons, the Bonferroni multiple is $B = t(.9833; 46) = 2.19$. Note that $n_T = 3(10) + 20 = 50$ for the first iteration; hence $n_T - r = 50 - 4 = 46$. The anticipated width of the confidence intervals therefore is $2.19(.77) = \pm 1.69$. This is larger than the specified width ± 1.0, so a larger sample size needs to be tried next.

We shall try $n = 30$ next. We find that $\sigma\{\hat{L}_i\} = .45$ now, and the Bonferroni multiple will be $B = t(.9833; 146) = 2.15$. Hence, the anticipated width of the confidence intervals for $n = 30$ is $2.15(.45) = \pm .97$. This is slightly smaller than the specified width ± 1.0. However, since the planning value for σ may not be entirely accurate, management may decide to use 30 tires for each of the new brands and 60 tires for brand 4, the presently used snow tires.

Comment

Since one cannot be certain that the planning value for the standard deviation is correct, it is advisable to study a range of values for the standard deviation before making a final decision on sample size. ■

17.9 Analysis of Factor Effects when Factor Is Quantitative

When the factor under investigation is quantitative, the analysis of factor effects can be carried beyond the point of multiple comparisons to include a study of the nature of the response function. Consider an experimental study undertaken to investigate the effect on sales of the price of a product. Five different price levels are investigated (78 cents, 79 cents, 85 cents, 88 cents, and 89 cents), and the experimental unit is a store. After a preliminary test of whether mean sales differ for the five price levels studied, the analyst might use multiple comparisons to examine whether "odd pricing" at 79 cents actually leads to higher sales than "even pricing" at 78 cents, as well as other questions of interest. In addition, the analyst may wish to study whether mean sales are a specified function of price, in the range of prices studied in the experiment. Further, once the relation has been established, the analyst may wish to use it for estimating sales volumes at various price levels not studied.

The methods of regression analysis discussed earlier are, of course, appropriate for the analysis of the response function. Since the single-factor studies discussed in this chapter almost always involve replications at the different factor levels, the lack of fit of a specified response function can be tested. For this purpose, the analysis of variance error sum of squares in (16.29) serves as the pure error sum of squares in (3.16), the two being identical. We illustrate this relation in the following example.

Example

In a study to reduce raw material costs in a glassworks firm, an operations analyst collected the experimental data in Table 17.4 on the number of acceptable units produced from equal amounts of raw material by 28 entry-level piecework employees who had received special training as part of the experiment. Four training levels were used (6, 8, 10, and 12 hours), with seven of the employees being assigned at random to each level. The higher the number of acceptable pieces, the more efficient is the employee in utilizing the raw material. This study is a single-factor completely randomized design with four factor levels.

Preliminary Analysis. The analyst first tested whether or not the mean number of acceptable pieces is the same for the four training levels. ANOVA model (17.1) was employed:

$$Y_{ij} = \mu_i + \varepsilon_{ij} \tag{17.50}$$

The alternative conclusions and appropriate test statistic are:

$$H_0\colon \mu_1 = \mu_2 = \mu_3 = \mu_4$$

$$H_a\colon \text{not all } \mu_i \text{ are equal}$$

$$F^* = \frac{MSTR}{MSE}$$

TABLE 17.4 **Data—Piecework Trainees Example.**

Treatment (hours of training) i		Employee (j) 1	2	3	4	5	6	7
1	6 hours	40	39	39	36	42	43	41
2	8 hours	53	48	49	50	51	50	48
3	10 hours	53	58	56	59	53	59	58
4	12 hours	63	62	59	61	62	62	61

The SPSS^X output for single-factor ANOVA is shown in Figure 17.6. Residual analysis (to be discussed in Chapter 18) showed ANOVA model (17.50) to be apt. Therefore, the analyst proceeded with the test, using $\alpha = .05$. The decision rule is:

$$\text{If } F^* \leq F(.95; 3, 24) = 3.01, \text{ conclude } H_0$$
$$\text{If } F^* > 3.01, \text{ conclude } H_a$$

FIGURE 17.6 SPSS^X Computer Output—Piecework Trainees Example.

	GROUP	COUNT (n_i)	MEAN ($\bar{Y}_{i\cdot}$)	STANDARD DEVIATION
	GRP01	7	40.0000	2.3094
Treatment →	GRP02	7	49.8571	1.7728
	GRP03	7	56.5714	2.6367
	GRP04	7	61.4286	1.2724
	TOTAL	28	51.9643	8.4129

ANALYSIS OF VARIANCE

SOURCE	D F	SUM OF SQUARES	MEAN SQUARES
BETWEEN GROUPS	3	***SSTR*** → 1808.6778	602.8926 ← ***MSTR***
WITHIN GROUPS	24	***SSE*** → 102.2856	4.2619 ← ***MSE***
TOTAL	27	***SSTO*** → 1910.9634	

F RATIO	F PROB.
141.461	0.0000
↑ ***F****	↑ ***P-value***

MULTIPLE RANGE TEST

TUKEY-HSD PROCEDURE
RANGES FOR THE 0.050 LEVEL -

3.90 ← ***q(.95; 4, 24)***

HOMOGENEOUS SUBSETS

SUBSET 1

GROUP	GRP01
MEAN	40.0000

SUBSET 2

GROUP	GRP02
MEAN	49.8571

SUBSET 3

GROUP	GRP03
MEAN	56.5714

SUBSET 4

GROUP	GRP04
MEAN	61.4286

From Figure 17.6, we have:

$$F^* = \frac{MSTR}{MSE} = \frac{602.8926}{4.2619} = 141.5$$

Since $F^* = 141.5 > 3.01$, the analyst concluded H_a, that training level effects differed and that further analysis of them is warranted. The P-value for the test statistic is 0+, as shown in Figure 17.6.

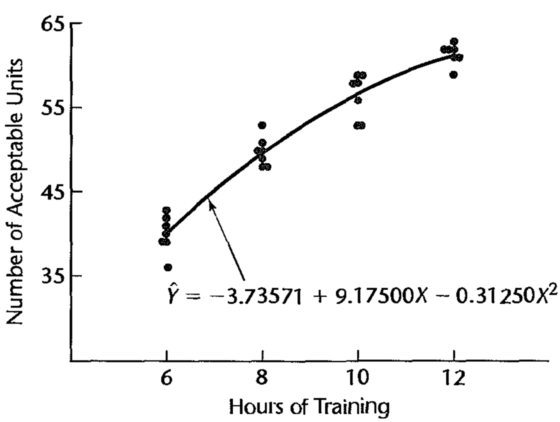
Investigation of Treatment Effects. The analyst's interest next centered on multiple comparisons of all pairs of treatment means. A Tukey multiple comparison option in the SPSS^X computer package was used. It gave the output shown in the lower portion of Figure 17.6. This output presents the results of single-degree-of-freedom tests conducted by means of the Tukey multiple comparison procedure for all pairwise comparisons. (The confidence intervals for the pairwise comparisons are not shown in the output.) All factor levels for which the test concludes that the pairwise means are equal are placed in the same group. This form of summary of single-degree-of-freedom tests was illustrated earlier for the Kenton Food Company example. When a group contains only one factor level, as is the case for all groups in the output of Figure 17.6, the implication is that all single-degree-of-freedom tests involving this factor level and each of the other factor levels lead to conclusion H_a, that the two factor level means being compared are not equal.

Two points should be noted in particular from the results in Figure 17.6: (1) All pairwise factor level differences are statistically significant. (2) There is some indication that differences between the means for adjoining factor levels diminish as the number of hours of training increases; that is, diminishing returns appear to set in as the length of training is increased.

Estimation of Response Function. These findings were in accord with the analyst's expectations that the treatment means μ_i would most likely follow a quadratic response function with respect to training level. The scatter plot in Figure 17.7 supports this expectation. The analyst now wished to investigate this point further by fitting a quadratic regression model. The model to be fitted and tested is:

$$Y_{ij} = \beta_0 + \beta_1 x_i + \beta_{11} x_i^2 + \varepsilon_{ij} \tag{17.51}$$

FIGURE 17.7 Scatter Plot and Fitted Quadratic Response Function—Piecework Trainees Example.



where Y_{ij} and ε_{ij} are defined as earlier, the βs are regression parameters, and x_i denotes the number of hours of training in the ith training level (X_i) centered around $\bar{X} = 9$, i.e., $x_i = X_i - 9$.

A portion of the data for the regression analysis is given in Table 17.5. Regressing Y on x and x^2 yielded the estimated regression function:

$$\hat{Y} = 53.52679 + 3.55000x - .31250x^2 \tag{17.52}$$

The analysis of variance for regression model (17.51) is shown in Table 17.6a. For completeness, we repeat in Table 17.6b the analysis of variance for ANOVA model (17.50).

TABLE 17.5 Illustration of Data for Regression Analysis—Piecework Trainees Example.

i	j	Y_{ij}	x_i	x_i^2
1	1	40	$6-9=-3$	9
1	2	39	$6-9=-3$	9
...	...	...	...	...
2	1	53	$8-9=-1$	1
2	2	48	$8-9=-1$	1
...	...	...	...	...
4	6	62	$12-9=3$	9
4	7	61	$12-9=3$	9

TABLE 17.6 Analyses of Variance—Piecework Trainees Example.

(a) Regression Model (17.51)

Source of Variation	SS	df	MS
Regression	1,808.100	2	904.05
Error	102.864	25	4.11
Total	1,910.964	27	

(b) Analysis of Variance Model (17.50)

Source of Variation	SS	df	MS
Treatments	1,808.678	3	602.89
Error	102.286	24	4.26
Total	1,910.964	27	

(c) ANOVA for Lack of Fit Test

Source of Variation	SS	df	MS
Regression	1,808.100	2	904.05
Error	102.864	25	4.11
Lack of fit	.578	1	.58
Pure error	102.286	24	4.26
Total	1,910.964	27	

Since the data contain replicates, the analyst could test regression model (17.51) for lack of fit, utilizing the fact that the ANOVA error sum of squares in (16.29) is identical to the regression pure error sum of squares in (3.16). Both measure variation around the mean of the Y observations at any given level of X (i.e., around the estimated treatment mean $\bar{Y}_{i\cdot}$). Hence, the lack of fit sum of squares can be readily obtained from previous results:

$$SSLF = \underset{\text{(Table 17.6a)}}{SSE} - \underset{\text{(Table 17.6b)}}{SSPE} = 102.864 - 102.286 = .578 \tag{17.53}$$

Since there are $c = r = 4$ levels of X here and $p = 3$ parameters in the regression model, $SSLF$ has associated with it $c - p = 4 - 3 = 1$ degree of freedom. Hence, we obtain $MSLF = .578/1 = .578$. Table 17.6c contains the analysis of variance for the regression model, with the error sum of squares and degrees of freedom broken down into lack of fit and pure error components.

The alternative conclusions (6.68a) for the test of lack of fit here are:

$$H_0: E\{Y\} = \beta_0 + \beta_1 x + \beta_{11} x^2$$
$$H_a: E\{Y\} \neq \beta_0 + \beta_1 x + \beta_{11} x^2$$

and test statistic (6.68b) is:

$$F^* = \frac{MSLF}{MSPE}$$

For $\alpha = .05$, decision rule (6.68c) becomes:

$$\text{If } F^* \leq F(.95; 1, 24) = 4.26, \text{ conclude } H_0$$
$$\text{If } F^* > 4.26, \text{ conclude } H_a$$

We calculate the test statistic from Table 17.6c:

$$F^* = \frac{.58}{4.26} = .136$$

Since $F^* = .136 \leq 4.26$, the analyst concluded that the quadratic response function is a good fit. Consequently, the fitted regression function in (17.52) was used in further evaluation of the relation between mean number of acceptable pieces produced and level of training, after expressing the fitted response function in the original predictor variable X (number of hours of training):

$$\hat{Y} = -3.73571 + 9.17500X - .31250X^2$$

Figure 17.7 displays this fitted response function.

Cited References

17.1. Cochran, W. G., and G. M. Cox. *Experimental Designs*. 2nd ed. New York: John Wiley & Sons, 1957, p. 74.

17.2. Dunnett, C. W. "A Multiple Comparison Procedure for Comparing Several Treatments with a Control." *Journal of the American Statistical Association* 50 (1955), pp. 1096–1121.

17.3. Hsu, J. C. *Multiple Comparisons: Theory and Methods*. London: Chapman & Hall, 1996.

17.4. Ott, E. R. "Analysis of Means—A Graphical Procedure." *Industrial Quality Control* 24 (1967), pp. 101–109.

17.5. Nelson, L. S. "Exact Critical Values for Use with the Analysis of Means," *Journal of Quality Technology* 15 (1983), pp. 40–44.

17.6. Nelson, P. R. "Additional Uses for the Analysis of Means and Extended Tables of Critical Values," *Technometrics* 35 (1993), pp. 61–71.

oblems

17.1. Refer to **Premium distribution** Problem 16.12. A student, asked to give a class demonstration of the use of a confidence interval for comparing two treatment means, proposed to construct a 99 percent confidence interval for the pairwise comparison $D = \mu_5 - \mu_3$. The student selected this particular comparison because the estimated treatment means $\bar{Y}_{5\cdot}$ and $\bar{Y}_{3\cdot}$ are the largest and smallest, respectively, and stated: "This confidence interval is particularly useful. If it does not straddle zero, it indicates, with significance level $\alpha = .01$, that the factor level means are not equal."

a. Explain why the student's assertion is not correct.

b. How should the confidence interval be constructed so that the assertion can be made with significance level $\alpha = .01$?

17.2. A trainee examined a set of experimental data to find comparisons that "look promising" and calculated a family of Bonferroni confidence intervals for these comparisons with a 90 percent family confidence coefficient. Upon being informed that the Bonferroni procedure is not applicable in this case because the comparisons had been suggested by the data, the trainee stated: "This makes no difference. I would use the same formulas for the point estimates and the estimated standard errors even if the comparisons were not suggested by the data." Respond.

17.3. Consider the following linear combinations of interest in a single-factor study involving four factor levels:

$$\text{(i)} \quad \mu_1 + 3\mu_2 - 4\mu_3$$

$$\text{(ii)} \quad .3\mu_1 + .5\mu_2 + .1\mu_3 + .1\mu_4$$

$$\text{(iii)} \quad \frac{\mu_1 + \mu_2 + \mu_3}{3} - \mu_4$$

a. Which of the linear combinations are contrasts? State the coefficients for each of the contrasts.

b. Give an unbiased estimator for each of the linear combinations. Also give the estimated variance of each estimator assuming that $n_i \equiv n$.

17.4. A single-factor ANOVA study consists of $r = 6$ treatments with sample sizes $n_i \equiv 10$.

a. Assuming that pairwise comparisons of the treatment means are to be made with a 90 percent family confidence coefficient, find the T, S, and B multiples for the following numbers of pairwise comparisons in the family: $g = 2, 5, 15$. What generalization is suggested by your results?

b. Assuming that contrasts of the treatment means are to be estimated with a 90 percent family confidence coefficient, find the S and B multiples for the following numbers of contrasts in the family: $g = 2, 5, 15$. What generalization is suggested by your results?

17.5. Consider a single-factor study with $r = 5$ treatments and sample sizes $n_i \equiv 5$.

a. Find the T, S, and B multiples if $g = 2$, 5, and 10 pairwise comparisons are to be made with a 95 percent family confidence coefficient. What generalization is suggested by your results?

b. What would be the T, S, and B multiples for sample sizes $n_i \equiv 20$? Does the generali[zation] obtained in part (a) still hold?

17.6. In making multiple comparisons, why is it appropriate to use the multiple comparison p[roce]dure that leads to the tightest confidence intervals for the sample data obtained? Discuss.

17.7. For a single-factor study with $r = 2$ treatments and sample sizes $n_i \equiv 10$, find the T, S, [and] B multiples for $g = 1$ pairwise comparison with a 99 percent family confidence coeffici[ent]. What generalization is suggested by your results?

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

a. Prepare a line plot of the estimated factor level means $\bar{Y}_{i\cdot}$. What does this plot suggest reg[ard]ing the effect of the level of research and development expenditures on mean productivi[ty] improvement?

b. Estimate the mean productivity improvement for firms with high research and developme[nt] expenditures levels; use a 95 percent confidence interval.

c. Obtain a 95 percent confidence interval for $D = \mu_2 - \mu_1$. Interpret your interval estimat[e].

d. Obtain confidence intervals for all pairwise comparisons of the treatment means; use th[e] Tukey procedure and a 90 percent family confidence coefficient. State your findings an[d] prepare a graphic summary by underlining nonsignificant comparisons in your line plot in part (a).

e. Is the Tukey procedure employed in part (d) the most efficient one that could be used here? Explain.

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

a. Prepare a bar-interval graph of the estimated factor level means $\bar{Y}_{i\cdot}$, where the interval[s] correspond to the confidence limits in (17.7) with $\alpha = .05$. What does this plot sugges[t] about the effect of color on the response rate? Is your conclusion in accord with the test result in Problem 16.8c?

b. Estimate the mean response rate for blue questionnaires; use a 90 percent confidence interval.

c. Test whether or not $D = \mu_3 - \mu_2 = 0$; use $\alpha = .10$. State the alternatives, decision rule, an[d] conclusion. In light of the result for the ANOVA test in Problem 16.8e, is your conclusion surprising? Explain.

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

a. Prepare a line plot of the estimated factor level means $\bar{Y}_{i\cdot}$. What does this plot suggest about the effect of prior physical fitness on the mean time required in therapy?

b. Estimate with a 99 percent confidence interval the mean number of days required in therapy for persons of average physical fitness.

c. Obtain confidence intervals for $D_1 = \mu_2 - \mu_3$ and $D_2 = \mu_1 - \mu_2$; use the Bonferroni procedure with a 95 percent family confidence coefficient. Interpret your results.

d. Would the Tukey procedure have been more efficient to use in part (c)? Explain.

e. If the researcher also wished to estimate $D_3 = \mu_1 - \mu_3$, still with a 95 percent family confidence coefficient, would the B multiple in part (c) need to be modified? Would this also be the case if the Tukey procedure had been employed?

f. Test for all pairs of factor level means whether or not they differ; use the Tukey procedure with $\alpha = .05$. Set up groups of factor levels whose means do not differ.

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

a. Prepare a main effects plot of the estimated factor level means $\bar{Y}_{i\cdot}$. What does this plot suggest regarding the effect of the owner's age on the mean cash offer?

b. Estimate the mean cash offer for young owners; use a 99 percent confidence interval.

c. Construct a 99 percent confidence interval for $D = \mu_3 - \mu_1$. Interpret your interval estimate.

d. Test whether or not $\mu_2 - \mu_1 = \mu_3 - \mu_2$; control the α risk at .01. State the alternatives, decision rule, and conclusion.

e. Obtain confidence intervals for all pairwise comparisons between the treatment means; use the Tukey procedure and a 90 percent family confidence coefficient. Interpret your results and provide a graphic summary by preparing a paired comparison plot. Are your conclusions in accord with those in part (a)?

f. Would the Bonferroni procedure have been more efficient to use in part (e) than the Tukey procedure? Explain.

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

a. Prepare a main effects plot of the estimated factor level means $\bar{Y}_{i\cdot}$. What does this plot suggest regarding the variation in the mean fills for the six machines?

b. Construct a 95 percent confidence interval for the mean fill for machine 1.

c. Obtain a 95 percent confidence interval for $D = \mu_2 - \mu_1$. Interpret your interval estimate.

d. Prepare a paired comparison plot and interpret it.

e. The consultant is particularly interested in comparing the mean fills for machines 1, 4, and 5. Use the Bonferroni testing procedure for all pairwise comparisons among these three treatment means with family level of significance $\alpha = .10$. Interpret your results and provide a graphic summary by preparing a line plot of the estimated factor level means with nonsignificant differences underlined. Do your conclusions agree with those in part (a)?

f. Would the Tukey testing procedure have been more efficient to use in part (e) than the Bonferroni testing procedure? Explain.

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

a. Prepare an interval plot of the estimated factor level means $\bar{Y}_{i\cdot}$, where the intervals correspond to the confidence limits in (17.7) with $\alpha = .10$. What does this plot suggest about the variation in the mean time lapses for the five agents?

b. Test for all pairs of factor level means whether or not they differ; use the Tukey procedure with $\alpha = .10$. Set up groups of factor levels whose means do not differ. Use a paired comparison plot to summarize the results.

c. Construct a 90 percent confidence interval for the mean time lapse for agent 1.

d. Obtain a 90 percent confidence interval for $D = \mu_2 - \mu_1$. Interpret your interval estimate.

e. The marketing director wishes to compare the mean time lapses for agents 1, 3, and 5. Obtain confidence intervals for all pairwise comparisons among these three treatment means; use the Bonferroni procedure with a 90 percent family confidence coefficient. Interpret your results and present a graphic summary by preparing a line plot of the estimated factor level means with nonsignificant differences underlined. Do your conclusions agree with those in part (a)?

f. Would the Tukey procedure have been more efficient to use in part (e) than the Bonferroni procedure? Explain.

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

a. Estimate the difference in mean productivity improvement between firms with low or moderate research and development expenditures and firms with high expenditures; use a 95 percent confidence interval. Employ an unweighted mean for the low and moderate expenditures groups. Interpret your interval estimate.

b. The sample sizes for the three factor levels are proportional to the population sizes. The economist wishes to estimate the mean productivity gain last year for all firms in the population. Estimate this overall mean productivity improvement with a 95 percent confidence interval.

c. Using the Scheffé procedure, obtain confidence intervals for the following comparisons with 90 percent family confidence coefficient:

$$D_1 = \mu_3 - \mu_2 \qquad D_3 = \mu_2 - \mu_1$$
$$D_2 = \mu_3 - \mu_1 \qquad L_1 = \frac{\mu_1 + \mu_2}{2} - \mu_3$$

Interpret your results and describe your findings.

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

a. Estimate the contrast $L = (\mu_1 - \mu_2) - (\mu_2 - \mu_3)$ with a 99 percent confidence interval. Interpret your interval estimate.

b. Estimate the following comparisons using the Bonferroni procedure with a 95 percent family confidence coefficient:

$$D_1 = \mu_1 - \mu_2 \qquad D_3 = \mu_2 - \mu_3$$
$$D_2 = \mu_1 - \mu_3 \qquad L_1 = D_1 - D_3$$

Interpret your results and describe your findings.

c. Would the Scheffé procedure have been more efficient to use in part (b) than the Bonferroni procedure? Explain.

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

a. Estimate the contrast $L = (\mu_3 - \mu_2) - (\mu_2 - \mu_1)$ with a 99 percent confidence interval. Interpret your interval estimate.

b. Estimate the following comparisons with a 90 percent family confidence coefficient; employ the most efficient multiple comparison procedure:

$$D_1 = \mu_2 - \mu_1 \qquad D_3 = \mu_3 - \mu_1$$
$$D_2 = \mu_3 - \mu_2 \qquad L_1 = D_2 - D_1$$

Interpret your results.

*17.17. Refer to **Filling machines** Problem 16.11. Machines 1 and 2 were purchased new five years ago, machines 3 and 4 were purchased in a reconditioned state five years ago, and machines 5 and 6 were purchased new last year.

a. Estimate the contrast:

$$L = \frac{\mu_1 + \mu_2}{2} - \frac{\mu_3 + \mu_4}{2}$$

with a 95 percent confidence interval. Interpret your interval estimate.

b. Estimate the following comparisons with a 90 percent family confidence coefficient; use the most efficient multiple comparison procedure:

$$D_1 = \mu_1 - \mu_2 \qquad L_1 = \frac{\mu_1 + \mu_2}{2} - \frac{\mu_3 + \mu_4}{2}$$
$$D_2 = \mu_3 - \mu_4 \qquad L_2 = \frac{\mu_1 + \mu_2}{2} - \frac{\mu_5 + \mu_6}{2}$$
$$D_3 = \mu_5 - \mu_6 \qquad L_3 = \frac{\mu_1 + \mu_2 + \mu_5 + \mu_6}{4} - \frac{\mu_3 + \mu_4}{2}$$
$$L_4 = \frac{\mu_1 + \mu_2 + \mu_3 + \mu_4}{4} - \frac{\mu_5 + \mu_6}{2}$$

Interpret your results. What can the consultant learn from these results about the differences between the six filling machines?

17.18. Refer to **Premium distribution** Problem 16.12. Agents 1 and 2 distribute merchandise only, agents 3 and 4 distribute cash-value coupons only, and agent 5 distributes both merchandise and coupons.

a. Estimate the contrast:

$$L = \frac{\mu_1 + \mu_2}{2} - \frac{\mu_3 + \mu_4}{2}$$

with a 90 percent confidence interval. Interpret your interval estimate.

b. Estimate the following comparisons with 90 percent family confidence coefficient; use the Scheffé procedure:

$$D_1 = \mu_1 - \mu_2 \qquad L_1 = \frac{\mu_1 + \mu_2}{2} - \mu_5$$

$$D_2 = \mu_3 - \mu_4 \qquad L_2 = \frac{\mu_3 + \mu_4}{2} - \mu_5$$

$$L_3 = \frac{\mu_1 + \mu_2}{2} - \frac{\mu_3 + \mu_4}{2}$$

Interpret your results.

c. Of all premium distributions, 25 percent are handled by agent 1, 20 percent by agent 2, 20 percent by agent 3, 20 percent by agent 4, and 15 percent by agent 5. Estimate the overall mean time lapse for premium distributions with a 90 percent confidence interval.

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

a. Use the analysis of means procedure to test for equality of treatment effects, with family significance level .05. Which treatments have the strongest effects?

b. Using the results in part (a), obtain the analysis of means plot. What additional information does this plot provide in comparison with the main effects plot in Problem 17.12a?

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

a. Use the analysis of means procedure to test for equality of treatment effects, with family significance level .10. Which treatments have the strongest effects?

b. Using the results in part (a), obtain the analysis of means plot. What additional information does this plot provide in comparison with the interval plot in Problem 17.13a?

17.21. Refer to **Solution concentration** Problem 3.15. Suppose the chemist initially wishes to employ ANOVA model (16.2) to determine whether or not the concentration of the solution is affected by the amount of time that has elapsed since preparation.

a. State the analysis of variance model.

b. Prepare a main effects plot of the estimated factor level means $\bar{Y}_{i.}$. What does this plot suggest about the relation between the solution concentration and time?

c. Obtain the analysis of variance table.

d. Test whether or not the factor level means are equal; use $\alpha = .025$. State the alternatives, decision rule, and conclusion.

e. Make pairwise comparisons of factor level means between all adjacent lengths of time; use the Bonferroni procedure with a 95 percent family confidence coefficient. Are your conclusions in accord with those in part (b)? Do your results suggest that the regression relation is not linear?

17.22. A market researcher stated in a seminar: "The power approach to determining sample sizes for analysis of variance problems is not meaningful: only the estimation approach should be used. We never conduct a study where all treatment means are expected to be equal, so we are always interested in a variety of estimates." Discuss.

17.23. Refer to **Questionnaire color** Problem 16.8. Suppose estimates of all pairwise comparisons are of primary importance. What would be the required sample sizes if the precision of all pairwise comparisons is to be ± 3.0, using the Tukey procedure with a 95 percent family confidence coefficient?

17.24. Refer to **Rehabilitation therapy** Problem 16.9. Suppose primary interest is in estimating the two pairwise comparisons:

$$L_1 = \mu_1 - \mu_2 \qquad L_2 = \mu_3 - \mu_2$$

What would be the required sample sizes if the precision of each comparison is to be ± 3.0 days, using the most efficient multiple comparison procedure with a 95 percent family confidence coefficient?

*17.25. Refer to **Filling machines** Problem 16.11. Suppose primary interest is in estimating the following comparisons:

$$L_1 = \mu_1 - \mu_2 \qquad L_3 = \frac{\mu_1 + \mu_2}{2} - \frac{\mu_3 + \mu_4}{2}$$

$$L_2 = \mu_3 - \mu_4 \qquad L_4 = \frac{\mu_1 + \mu_2 + \mu_3 + \mu_4}{4} - \frac{\mu_5 + \mu_6}{2}$$

What would be the required sample sizes if the precision of each of these comparisons is not to exceed $\pm .08$ ounce, using the best multiple comparison procedure with a 95 percent family confidence coefficient?

17.26. Refer to **Premium distribution** Problem 16.12. Suppose primary interest is in estimating the following comparisons:

$$L_1 = \mu_1 - \mu_2 \qquad L_3 = \frac{\mu_1 + \mu_2}{2} - \mu_5$$

$$L_2 = \mu_3 - \mu_4 \qquad L_4 = \frac{\mu_1 + \mu_2}{2} - \frac{\mu_3 + \mu_4}{2}$$

What would be the required sample sizes if the precision of each of the estimated comparisons is not to exceed ± 1.0 day, using the most efficient multiple comparison procedure with a 90 percent family confidence coefficient?

17.27. Refer to **Rehabilitation therapy** Problem 16.9. Suppose that primary interest is in comparing the below-average and above-average physical fitness groups, respectively, with the average physical fitness group. Thus, two comparisons are of interest:

$$L_1 = \mu_1 - \mu_2 \qquad L_2 = \mu_3 - \mu_2$$

Assume that a reasonable planning value for the error standard deviation is $\sigma = 4.5$ days.

a. It has been decided to use equal sample sizes (n) for the below-average and above-average groups. If twice this sample size ($2n$) were to be used for the average physical fitness group, what would be the required sample sizes if the precision of each pairwise comparison is to be ± 2.5 days, using the Bonferroni procedure and a 90 percent family confidence coefficient?

b. Repeat the calculations in part (a) if the sample size for the average physical fitness group is to be: (1) n and (2) $3n$, all other specifications remaining the same.

c. Compare your results in parts (a) and (b). Which design leads to the smallest total sample size here?

17.28. Refer to **Rehabilitation therapy** Problem 16.9. A biometrician has developed a scale for physical fitness status, as follows:

Physical Fitness Status	Scale Value
Below average	83
Average	100
Above average	121

a. Using this physical fitness status scale, fit first-order regression model (1.1) for regressing number of days required for therapy (Y) on physical fitness status (X).

b. Obtain the residuals and plot them against X. Does a linear regression model appear to fit the data?

c. Perform an F test to determine whether or not there is lack of fit of a linear regression function; use $\alpha = .05$. State the alternatives, decision rule, and conclusion.

d. Could you test for lack of fit of a quadratic regression function here? Explain.

*17.29. Refer to **Filling machines** Problem 16.11. A maintenance engineer has suggested that the differences in mean fills for the six machines are largely related to the length of time since a machine last received major servicing. Service records indicate these lengths of time to be as follows (in months):

Filling Machine	Number of Months	Filling Machine	Number of Months
1	.4	4	5.3
2	3.7	5	1.4
3	6.1	6	2.1

a. Fit second-order polynomial regression model (8.2) for regressing amount of fill (Y) on number of months since major servicing (X).

b. Obtain the residuals and plot them against X. Does a quadratic regression function appear to fit the data?

c. Perform an F test to determine whether or not there is lack of fit of a quadratic regression function; use $\alpha = .01$. State the alternatives, decision rule, and conclusion.

d. Test whether or not the quadratic term in the response function can be dropped from the model; use $\alpha = .01$. State the alternatives, decision rule, and conclusion.

Exercises

17.30. Show that when $r = 2$ and $n_i \equiv n$, q defined in (17.35) is equivalent to $\sqrt{2}|t^*|$, where t^* is defined in (A.65) in Appendix A.

17.31. Starting with (17.38), complete the derivation of (17.30).

17.32. Show that when $r = 2$, S^2 defined in (17.43a) is equivalent to $[t(1 - \alpha/2; n_T - r)]^2$.

17.33. Show that the estimated variance of $\hat{\tau}_i$ in (17.48) is given by (17.49).

17.34. (Calculus needed.) Refer to **Rehabilitation therapy** Problem 16.9. The sample sizes for the below-average, average, and above-average physical fitness groups are to be n, kn, and n, respectively. Assuming that ANOVA model (16.2) is appropriate, find the optimal value of k to minimize the variances of $\hat{L}_1 = \bar{Y}_{1\cdot} - \bar{Y}_{2\cdot}$ and $\hat{L}_2 = \bar{Y}_{3\cdot} - \bar{Y}_{2\cdot}$ for a given total sample size n_T.