Chapter 16

Single-Factor Studies

In the last chapter, we presented a general introduction to the design of experimental and observational studies. In this and the next two chapters, we shall focus on the design and analysis of single-factor studies. This includes the development of single-factor analysis of variance (ANOVA) model, the analysis and interpretation of factor level means, assessment of model adequacy, and the use of remedial measures when necessary.

In this chapter, we briefly review the design of single-factor studies and the associated linear models, then discuss the relation between regression and analysis of variance. In the next few sections we introduce in detail the single-factor ANOVA model and the associated F test for equality of factor level means. We then consider alternative formulations of the ANOVA model, followed by a regression approach to the single-factor ANOVA model. In the last few sections, we consider a nonparametric randomization test as an alternative to the ANOVA test, and, finally, we present two methods for the planning of sample sizes in single-factor studies.

16.1 Single-Factor Experimental and Observational Studies

Single-factor experimental and observational studies are the most basic form of comparative studies used in practice. In a single-factor experimental study, the treatments correspond to the levels of the factor, and randomization is used to assign the treatments to the experimental units. In the following we present three examples of single-factor studies. The first two examples are experimental studies, and the third is a cross-sectional observational study. We then briefly review the approach described in Chapter 15 for modeling a single-factor study.

Example 1 A hospital research staff wished to determine the best dosage level for a standard type of drug therapy to treat a medical condition. In order to compare the effectiveness of three dosage levels, 30 patients with the medical problem were recruited to participate in a pilot study. Each patient was randomly assigned to one of the three drug dosage levels. Randomization was performed in such a way that an equal number of patients ended up being evaluated for each drug dosage level, i.e., with exactly 10 patients studied in each drug dosage level group. This is an example of completely randomized design, based on a single, three-level quantitative factor. This particular design is said to be *balanced,* because each treatment is replicated the same number of times.

Example 2

In an experiment to investigate absorptive properties of four different formulations of a paper towel, five sheets of paper towel were randomly selected from each of the four types (formulation 1, formulation 2, formulation 3, and formulation 4) of paper towel. Twenty 6-ounce beakers of water were prepared, and the twenty paper towel sheets were randomly assigned to the beakers. Paper towels were then fully submerged in the beaker water for 10 seconds, withdrawn, and the amount of water absorbed by each paper towel sheet was determined. This is an example of a completely randomized design, based on a single, four-level qualitative factor.

Example 3

Four machines in a plant were studied with respect to the diameters of ball bearings they produced. The purpose of the study was to determine whether substantial differences in the diameters of ball bearings existed between the machines. If so, the machines would need to be calibrated. This is an example of an observational study, as no randomization of treatments to experimental units occurred.

As we noted in Chapter 15, although the first two examples are experimental studies and the third is an observational study, the methods used for statistical analysis are generally the same. If the single factor has r levels, one approach to constructing a linear statistical model employs $r-1$ indicator variables as predictors. Then the response for the jth replicate of the ith treatment or factor level is modeled:

$$Y_{ij} = \beta_0 + \beta_1 X_{ij1} + \cdots + \beta_{r-1} X_{ij,r-1} + \varepsilon_{ij}$$

where:

$$X_{ij1} = \begin{cases} 1 & \text{if treatment 1} \\ 0 & \text{otherwise} \end{cases}$$

$$X_{ij2} = \begin{cases} 1 & \text{if treatment 2} \\ 0 & \text{otherwise} \end{cases}$$

$$\cdots$$

$$X_{ij,r-1} = \begin{cases} 1 & \text{if treatment } r-1 \\ 0 & \text{otherwise} \end{cases}$$

Recall that because all of the predictors are indicator variables, this model is sometimes referred to as an *analysis of variance* model.

For the first example, we have an alternative. Because the factor—dosage level—is quantitative with three levels, we could also model its effect using a second-order (or lower-order) polynomial regression model, as described in Section 8.1. Specifically, two choices for the first example are:

$$Y_{ij} = \beta_0 + \beta_1 X_{ij1} + \beta_2 X_{ij2} + \varepsilon_{ij} \qquad \text{ANOVA Model}$$

where:

$$X_{ij1} = \begin{cases} 1 & \text{if treatment 1} \\ 0 & \text{otherwise} \end{cases}$$

$$X_{ij2} = \begin{cases} 1 & \text{if treatment 2} \\ 0 & \text{otherwise} \end{cases}$$

or, employing second-order polynomial model (8.1):

$$Y_{ij} = \beta_0 + \beta_1 x_{ij} + \beta_{11} x_{ij}^2 + \varepsilon_{ij} \qquad \text{Regression Model}$$

where:

x_{ij} = centered dosage level amount for the ijth case

In the next section, we discuss the choice between the two types of models.

16.2 Relation between Regression and Analysis of Variance

Regression analysis, as we have seen, is concerned with the statistical relation between one or more predictor variables and a response variable. Both the predictor and response variables in ordinary regression models are quantitative. The regression function describes the nature of the statistical relation between the mean response and the levels of the predictor variable(s).

We encountered the use of analysis of variance in our consideration of regression. It was used there for a variety of tests concerning the regression coefficients, the fit of the regression model, and the like. The analysis of variance is actually much more general than its use with regression models indicated. Analysis of variance models are a basic type of statistical model. They are concerned, like regression models, with the statistical relation between one or more predictor variables and a response variable. Like regression models, analysis of variance models are appropriate for both observational data and data based on formal experiments. Further, as in the usual regression models, the response variable for analysis of variance models is a quantitative variable. Analysis of variance models differ from ordinary regression models in two key respects:

1. The explanatory or predictor variables in analysis of variance models may be qualitative (gender, geographic location, plant shift, etc.).
2. If the predictor variables are quantitative, no assumption is made in analysis of variance models about the nature of the statistical relation between them and the response variable. Thus, the need to specify the nature of the regression function encountered in ordinary regression analysis does not arise in analysis of variance models.

Illustrations

Figure 16.1 illustrates the essential differences between regression and analysis of variance models for the case where the predictor variable is quantitative. Shown in Figure 16.1a is the regression model for a pricing study involving three different price levels, X = \$50, \$60, \$70. Note that the XY plane has been rotated from its usual position so that the Y axis faces the viewer. For each level of the predictor variable, there is a probability distribution of sales volumes. The means of these probability distributions fall on the regression curve, which describes the statistical relation between price and mean sales volume.

The analysis of variance model for the same study is illustrated in Figure 16.1b. The three price levels are treated as separate populations, each leading to a probability distribution of sales volumes. The quantitative differences in the three price levels and their statistical relation to expected sales volume are not considered by the analysis of variance model.

FIGURE 16.1 Relation between Regression and Analysis of Variance Models.

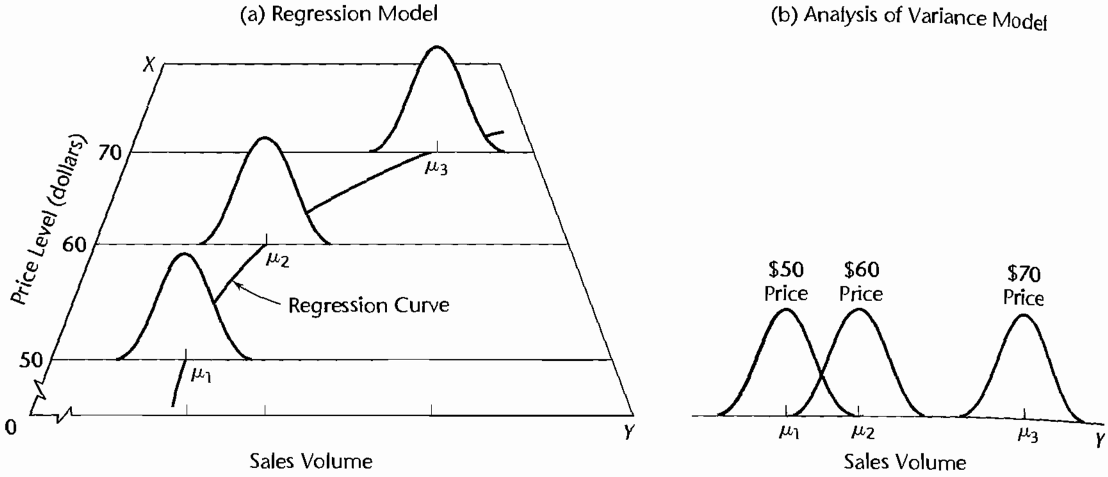


FIGURE 16.2 Analysis of Variance Model Representation —Incentive Pay Example.

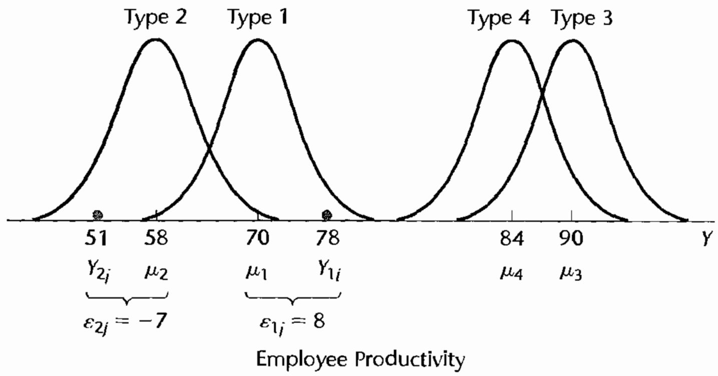


Figure 16.2 illustrates the analysis of variance model for a study of the effects of four different types of incentive pay systems on employee productivity. Here, each type of incentive pay system corresponds to a different population, and there is associated with each a probability distribution of employee productivities (Y). Since type of incentive pay system is a qualitative variable, Figure 16.2 does not contain a corresponding regression model representation.

Choice between Two Types of Models

As we have seen in Chapter 8, regression analysis can handle qualitative predictor variables by means of indicator variables. When indicator variables are so used with regression models, the regression results will be identical to those obtained with analysis of variance models. The reason why analysis of variance exists as a distinct statistical methodology is that the structure of the predictor indicator variables permits computational simplifications that are explicitly recognized in the statistical procedures for the analysis of variance.

Hence, there is no fundamental choice between regression and analysis of variance models when the predictor variables are qualitative.

On the other hand, there is a choice in modeling when the predictor variables are quantitative. One possibility is to recognize the quantitative nature of the predictor variables explicitly; this can only be done by a regression model. The other possibility is to set up classes for each quantitative variable and then employ either indicator variables in a regression model or an analysis of variance model. As we mentioned in Chapter 8, the strategy of setting up classes for quantitative variables is sometimes followed in large-scale studies as a means of obtaining a nonparametric regression fit when there is substantial doubt about the nature of the statistical relation. Here again, analysis of variance models and regression models with indicator variables will lead to identical results.

3 Single-Factor ANOVA Model

sic Ideas

The basic elements of the ANOVA model for a single-factor study are quite simple. Corresponding to each factor level, there is a probability distribution of responses. For example, in a study of the effects of four types of incentive pay on employee productivity, there is a probability distribution of employee productivities for each type of incentive pay. The ANOVA model assumes that:

1. Each probability distribution is normal.
2. Each probability distribution has the same variance.
3. The responses for each factor level are random selections from the corresponding probability distribution and are independent of the responses for any other factor level.

Figure 16.2 illustrates these conditions. Note the normality of the probability distributions and the constant variability. The probability distributions differ only with respect to their means. Differences in the means therefore reflect the essential factor level effects, and it is for this reason that the analysis of variance focuses on the mean responses for the different factor levels.

The analysis of the sample data from the factor level probability distributions usually proceeds in two steps:

1. Determine whether or not the factor level means are the same.
2. If the factor level means differ, examine how they differ and what the implications of the differences are.

In this chapter, we consider step 1, the testing procedure for determining whether or not the factor level means are the same. In the next chapter, we take up the analysis of the factor level means when the means differ.

Cell Means Model

Before stating the ANOVA model for single-factor studies, we need to develop some notation. We shall denote by r the number of levels of the factor under study (e.g., $r = 4$ types of incentive pay), and we shall denote any one of these levels by the index i ($i = 1, \ldots, r$). The number of cases for the ith factor level is denoted by n_i, and the total number of cases in the study is denoted by n_T, where:

$$n_T = \sum_{i=1}^{r} n_i \tag{16.1}$$

This notation differs from that used earlier for regression models, where the subscript i identifies the case or trial.

For analysis of variance models we shall always use the last subscript to represent the case or trial for a given factor level or treatment. Here, the index j will be used to identify the given case or trial for a particular factor level. We shall let Y_{ij} denote the value of the response variable in the jth trial for the ith factor level. For instance, Y_{ij} is the productivity of the jth employee in the ith incentive plan, or the sales volume of the jth store featuring the ith type of shelf display. Since the number of cases or trials for the ith factor level is denoted by n_i, we have $j = 1, \ldots, n_i$.

The ANOVA model can now be stated as follows:

$$Y_{ij} = \mu_i + \varepsilon_{ij} \tag{16.2}$$

where:

Y_{ij} is the value of the response variable in the jth trial for the ith factor level or treatment

μ_i are parameters

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

$i = 1, \ldots, r;\ j = 1, \ldots, n_i$

This model is called the *cell means model* for reasons to be explained shortly. This model may be used for data from observational studies or for data from experimental studies based on a completely randomized design.

Important Features of Model

1. The observed value of Y in the jth trial for the ith factor level or treatment is the sum of two components: (a) a constant term μ_i and (b) a random error term ε_{ij}.

2. Since $E\{\varepsilon_{ij}\} = 0$, it follows that:

$$E\{Y_{ij}\} = \mu_i \tag{16.3}$$

Thus, all responses or observations Y_{ij} for the ith factor level have the same expectation μ_i, and this parameter is the mean response for the ith factor level or treatment.

3. Since μ_i is a constant, it follows from (A.16a) that:

$$\sigma^2\{Y_{ij}\} = \sigma^2\{\varepsilon_{ij}\} = \sigma^2 \tag{16.4}$$

Thus, all observations have the same variance, regardless of factor level.

4. Since each ε_{ij} is normally distributed, so is each Y_{ij}. This follows from (A.36) because Y_{ij} is a linear function of ε_{ij}.

5. The error terms are assumed to be independent. Hence, the error term for the outcome on any one trial has no effect on the error term for the outcome of any other trial for the

same factor level or for a different factor level. Since the ε_{ij} are independent, so are the responses Y_{ij}.

6. In view of these features, ANOVA model (16.2) can be restated as follows:

$$Y_{ij} \text{ are independent } N(\mu_i, \sigma^2) \tag{16.5}$$

mple

Suppose that ANOVA model (16.2) is applicable to the earlier incentive pay study illustration and that the parameters are as follows:

$$\mu_1 = 70 \qquad \mu_2 = 58 \qquad \mu_3 = 90 \qquad \mu_4 = 84 \qquad \sigma = 4$$

Figure 16.2 contains a representation of this model. Note that employee productivities for incentive pay type 1 according to this model are normally distributed with mean $\mu_1 = 70$ and standard deviation $\sigma = 4$.

Suppose that in the jth trial of incentive pay type 1, the observed productivity is $Y_{1j} = 78$. In that case, the error term value is $\varepsilon_{1j} = 8$, for we have:

$$\varepsilon_{1j} = Y_{1j} - \mu_1 = 78 - 70 = 8$$

Figure 16.2 shows this observation Y_{1j}. Note that the deviation of Y_{1j} from the mean μ_1 represents the error term ε_{1j}. This figure also shows the observation $Y_{2j} = 51$, for which the error term value is $\varepsilon_{2j} = -7$.

The ANOVA Model Is a Linear Model

ANOVA model (16.2) is a linear model because it can be expressed in matrix terms in the form (6.19), i.e., as $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$. We illustrate this for a study involving $r = 3$ treatments, and for which $n_1 = n_2 = n_3 = 2$. $\mathbf{Y}$, $\mathbf{X}$, $\boldsymbol{\beta}$, and $\boldsymbol{\varepsilon}$ are then defined as follows here:

$$\mathbf{Y} = \begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{21} \\ Y_{22} \\ Y_{31} \\ Y_{32} \end{bmatrix} \qquad \mathbf{X} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \qquad \boldsymbol{\beta} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} \qquad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{21} \\ \varepsilon_{22} \\ \varepsilon_{31} \\ \varepsilon_{32} \end{bmatrix} \tag{16.6}$$

Note the simple structure of the $\mathbf{X}$ matrix and that the $\boldsymbol{\beta}$ vector consists of the means μ_i.

To see that these matrices yield ANOVA model (16.2), recall from (6.20) that the vector of expected values $E\{Y_{ij}\}$ is given by $\mathbf{E}\{\mathbf{Y}\} = \mathbf{X}\boldsymbol{\beta}$. We thus obtain:

$$\mathbf{E}\{\mathbf{Y}\} = \begin{bmatrix} E\{Y_{11}\} \\ E\{Y_{12}\} \\ E\{Y_{21}\} \\ E\{Y_{22}\} \\ E\{Y_{31}\} \\ E\{Y_{32}\} \end{bmatrix} = \mathbf{X}\boldsymbol{\beta} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_1 \\ \mu_2 \\ \mu_2 \\ \mu_3 \\ \mu_3 \end{bmatrix} \tag{16.7}$$

This indicates properly that $E\{Y_{ij}\} = \mu_i$. Hence, ANOVA model (16.2)—$Y_{ij} = \mu_i + \varepsilon_{ij}$—in matrix form is given by $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$:

$$\mathbf{Y} = \begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{21} \\ Y_{22} \\ Y_{31} \\ Y_{32} \end{bmatrix} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} = \begin{bmatrix} \mu_1 \\ \mu_1 \\ \mu_2 \\ \mu_2 \\ \mu_3 \\ \mu_3 \end{bmatrix} + \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{21} \\ \varepsilon_{22} \\ \varepsilon_{31} \\ \varepsilon_{32} \end{bmatrix} \tag{16.8}$$

Since the error terms in the model have the same structure as those in general linear regression model (6.19)—namely, independence and constant variance—the variance-covariance matrix of the error terms in the ANOVA model is the same as in (6.19):

$$\boldsymbol{\sigma}^2\{\boldsymbol{\varepsilon}\} = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix} = \sigma^2\mathbf{I} \tag{16.9}$$

In addition, like for general linear regression model (6.19), the variance-covariance matrix of the Y responses is the same as that of the error terms:

$$\boldsymbol{\sigma}^2\{\mathbf{Y}\} = \sigma^2\mathbf{I} \tag{16.10}$$

When ANOVA model (16.2) is expressed as a linear model, as in (16.8), it can be seen why it is called the cell means model, because the $\boldsymbol{\beta}$ vector contains the means of the "cells"—here factor levels. In Section 16.7 we discuss an equivalent ANOVA model called the factor effects model, where the $\boldsymbol{\beta}$ vector contains components of the factor level means.

Interpretation of Factor Level Means

Observational Data. In an observational study, the factor level means μ_i correspond to the means for the different factor level populations. For instance, in a study of the productivity of employees in each of three shifts operated in a plant, the populations consist of the employee productivities for each of the three shifts. The population mean μ_1 is the mean productivity for employees in shift 1, and μ_2 and μ_3 are interpreted similarly. The variance σ^2 refers to the variability of employee productivities within a shift.

Experimental Data. In an experimental study, the factor level mean μ_i stands for the mean response that would be obtained if the ith treatment were applied to all units in the population of experimental units about which inferences are to be drawn. Similarly, the variance σ^2 refers to the variability of responses if any given experimental treatment were applied to the entire population of experimental units. For instance, in a completely randomized design to study the effects of three different training programs on employee productivity, in which 90 employees participate, a third of these employees is assigned at random to each of the three programs. The mean μ_1 here denotes the mean productivity if training program 1 were given to each employee in the population of experimental units; the means μ_2 and μ_3 are interpreted correspondingly. The variance σ^2 denotes the variability in productivities if any one training program were given to each employee in the population of experimental units.

Distinction between ANOVA Models I and II

We shall consider two single-factor analysis of variance models. For brevity, we shall refer to these as ANOVA models I and II. ANOVA model I, which was stated in (16.2), applies to such cases as a comparison of five different advertisements or a comparison of four different rust inhibitors, where the conclusions pertain to just those factor levels included in the study. ANOVA model II, to be discussed in Chapter 25, applies to a different type of situation, namely, where the conclusions extend to a population of factor levels of which the levels in the study are a sample. Consider, for instance, a company that owns several hundred retail stores throughout the country. Seven of these stores are selected at random, and a sample of employees from each store is then chosen and asked in a confidential interview for an evaluation of the management of the store. The seven stores in the study constitute the seven levels of the factor under study, namely, retail store. In this case, however, management is not just interested in the seven stores included in the study but wishes to generalize the study results to all of the retail stores it owns. Another example when ANOVA model II is applicable is when three machines out of 75 in a plant are selected at random and their daily output is studied for a period of 10 days. The three machines constitute the three factor levels in this study, but interest is not just in the three machines in the study but in all machines in the plant.

Thus, the essential difference between situations where ANOVA models I and II are applicable is that model I is relevant when the factor levels are chosen because of intrinsic interest in them (e.g., five different advertisements) and they are not considered to be a sample from a larger population. ANOVA model II is appropriate when the factor levels constitute a sample from a larger population (e.g., three machines out of 75) and interest is in this larger population. Thus, ANOVA model I is also referred as the *fixed effects* model, and ANOVA model II is called the *random effects* model. In this and the next two chapters, we focus on ANOVA model I. For brevity, we omit the word "fixed" or "model I" and simply refer to the model as the ANOVA model.

Comment

The ANOVA model (16.2) for single-factor studies, like any other statistical model, is not likely to be met exactly by any real-world situation. However, it will be met approximately in many cases. As we shall note later, the statistical procedures based on ANOVA model (16.2) are quite robust, so that even if the actual conditions differ substantially from those of the model, the statistical analysis may still be an appropriate approximation. ■

16.4 Fitting of ANOVA Model

The parameters of ANOVA model (16.2) are ordinarily unknown and must be estimated from sample data. As with normal error regression models, the method of least squares and the method of maximum likelihood lead to the same estimators of the model parameters μ_i in normal error ANOVA model (16.2). Before turning to these estimators, we shall describe an example to be used in this chapter and the next, and we shall develop needed additional notation.

Example

The Kenton Food Company wished to test four different package designs for a new breakfast cereal. Twenty stores, with approximately equal sales volumes, were selected as the experimental units. Each store was randomly assigned one of the package designs, with each

TABLE 16.1 Number of Cases Sold by Stores for Each of Four Package Designs—Kenton Food Company Example.

Package Design	Store (j)					Total	Mean	Number of Stores
	1	2	3	4	5			
i	Y_{i1}	Y_{i2}	Y_{i3}	Y_{i4}	Y_{i5}	$Y_{i.}$	$\bar{Y}_{i.}$	n_i
1	11	17	16	14	15	73	14.6	5
2	12	10	15	19	11	67	13.4	5
3	23	20	18	17		78	19.5	4
4	27	33	22	26	28	136	27.2	5
All designs						$Y_{..} = 354$	$\bar{Y}_{..} = 18.63$	19

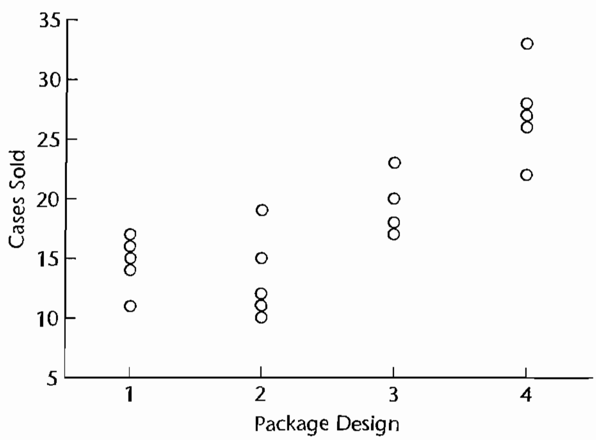


FIGURE 16.3 JMP Scatter Plot of Number of Cases Sold by Package Design—Kenton Food Company Example.

package design assigned to five stores. A fire occurred in one store during the study period, so this store had to be dropped from the study. Hence, one of the designs was tested in only four stores. The stores were chosen to be comparable in location and sales volume. Other relevant conditions that could affect sales, such as price, amount and location of shelf space, and special promotional efforts, were kept the same for all of the stores in the experiment. Sales, in number of cases, were observed for the study period, and the results are recorded in Table 16.1. This study is a completely randomized design with package design as the single, four-level factor. ♦

Figure 16.3 contains a JMP scatter plot of the number of cases sold versus package design number. We readily see that designs 3 and 4 led to the largest sales, and that designs 1 and 2 led to smaller sales. We also see that the variability in store sales appears to be about the same for the four designs, consistent with ANOVA model (16.2). To make more formal inferences, we first need to develop some additional notation.

Notation

As explained earlier, Y_{ij} represents the observation or response for the jth sample unit for the ith factor level. For the Kenton Food Company example, Y_{ij} denotes the number of cases sold by the jth store assigned to the ith package design. For instance, Y_{11} represents the sales of the first store assigned package design 1. For our example, $Y_{11} = 11$ cases. Similarly, sales of the second store assigned package design 3 are $Y_{32} = 20$ cases.

The total of the observations for the ith factor level is denoted by $Y_{i.}$:

$$Y_{i.} = \sum_{j=1}^{n_i} Y_{ij} \tag{16.11}$$

Note that the dot in $Y_{i.}$ indicates an aggregation over the j index; in our example, the aggregation is over all stores assigned to the ith package design. For instance, the total sales for all stores assigned package design 1 are, according to Table 16.1, $Y_{1.} = 73$ cases. Similarly, total sales for all stores assigned package design 4 are $Y_{4.} = 136$ cases.

The sample mean for the ith factor level is denoted by $\bar{Y}_{i.}$:

$$\bar{Y}_{i.} = \frac{\sum_{j=1}^{n_i} Y_{ij}}{n_i} = \frac{Y_{i.}}{n_i} \tag{16.12}$$

In our example, the mean number of cases sold by stores assigned package design 1 is $\bar{Y}_{1.} = 73/5 = 14.6$. Note that the dot in the subscript $\bar{Y}_{1.}$ indicates that the averaging is done over j (stores).

The total of all observations in the study is denoted by $Y_{..}$:

$$Y_{..} = \sum_{i=1}^{r} \sum_{j=1}^{n_i} Y_{ij} \tag{16.13}$$

where the two dots indicate aggregation over both the j and i indexes (in our example, over all stores for any one package design and then over all package designs). In our example, the total sales for all stores for all designs are $Y_{..} = 354$.

Finally, the overall mean for all responses is denoted by $\bar{Y}_{..}$:

$$\bar{Y}_{..} = \frac{\sum_i \sum_j Y_{ij}}{n_T} = \frac{Y_{..}}{n_T} \tag{16.14}$$

The two dots here indicate that the averaging is done over both i and j. For our example, we have from Table 16.1 that $\bar{Y}_{..} = 354/19 = 18.63$. Note that the overall mean (16.14) can be written as a weighted average of the factor level means in (16.12):

$$\bar{Y}_{..} = \sum_{i=1}^{r} \frac{n_i}{n_T} \bar{Y}_{i.} \tag{16.14a}$$

Least Squares and Maximum Likelihood Estimators

According to the least squares criterion, the sum of the squared deviations of the observations around their expected values must be minimized with respect to the parameters. For ANOVA model (16.2), we know from (16.3) that the expected value of observation Y_{ij} is $E\{Y_{ij}\} = \mu_i$. Hence, the quantity to be minimized is:

$$Q = \sum_i \sum_j (Y_{ij} - \mu_i)^2 \tag{16.15}$$

Now (16.15) can be written as follows:

$$Q = \sum_j (Y_{1j} - \mu_1)^2 + \sum_j (Y_{2j} - \mu_2)^2 + \cdots + \sum_j (Y_{rj} - \mu_r)^2 \tag{16.15a}$$

Note that each of the parameters appears in only one of the component sums in (16.15a). Hence, Q can be minimized by minimizing each of the component sums separately. It is well known that the sample mean minimizes a sum of squared deviations. Hence, the least squares estimator of μ_i, denoted by $\hat{\mu}_i$, is:

$$\hat{\mu}_i = \bar{Y}_{i.} \tag{16.16}$$

Thus, the *fitted value* for observation Y_{ij}, denoted by $\hat{Y}_{ij}$ for regression models, is simply the corresponding factor level sample mean here:

$$\hat{Y}_{ij} = \bar{Y}_{i.} \tag{16.17}$$

The same estimators are obtained by the method of maximum likelihood. The likelihood function here corresponds to that in (1.26) for the normal error simple linear regression model, except that the regression model expected value $\beta_0 + \beta_1 X_i$ is replaced here by μ_i:

$$L(\mu_1, \ldots, \mu_r, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left[-\frac{1}{2\sigma^2} \sum_i \sum_j (Y_{ij} - \mu_i)^2\right] \tag{16.18}$$

Maximizing this likelihood function with respect to the parameters μ_i is equivalent to minimizing the sum $\sum\sum(Y_{ij} - \mu_i)^2$ in the exponent, which is the least squares criterion in (16.15).

Example

For the Kenton Food Company example, the least squares and maximum likelihood estimates of the model parameters are as follows according to Table 16.1:

Parameter	Estimate
μ_1	$\hat{\mu}_1 = \bar{Y}_{1.} = 14.6$
μ_2	$\hat{\mu}_2 = \bar{Y}_{2.} = 13.4$
μ_3	$\hat{\mu}_3 = \bar{Y}_{3.} = 19.5$
μ_4	$\hat{\mu}_4 = \bar{Y}_{4.} = 27.2$

Thus, the mean sales per store with package design 1 are estimated to be 14.6 cases for the population of stores under study, and the fitted value for each of the observations for package design 1 is $\hat{Y}_{1j} = \bar{Y}_{1.} = 14.6$. Similarly, the mean sales for package design 2 are estimated to be 13.4 cases per store, and the fitted values for each response for this package design is $\hat{Y}_{2j} = \bar{Y}_{2.} = 13.4$.

Comments

1. The least squares and maximum likelihood estimators in (16.16) have all of the desirable properties mentioned in Chapter 1 for the regression estimators. For example, they are minimum variance unbiased estimators.

2. To derive the least squares estimator of μ_i, we need to minimize, with respect to μ_i, the ith component sum of squares in (16.15a):

$$Q_i = \sum_j (Y_{ij} - \mu_i)^2 \tag{16.19}$$

Differentiating with respect to μ_i, we obtain:

$$\frac{dQ_i}{d\mu_i} = \sum_j -2(Y_{ij} - \mu_i)$$

When we set this derivative equal to zero and replace the parameter μ_i by the least squares estimator $\hat{\mu}_i$, we obtain the result in (16.16):

$$-2\sum_{j=1}^{n_i}(Y_{ij} - \hat{\mu}_i) = 0$$

$$\sum_j Y_{ij} = n_i\hat{\mu}_i$$

$$\hat{\mu}_i = \bar{Y}_{i\cdot}$$ ■

esiduals

Residuals are highly useful for examining the aptness of ANOVA models. The residual e_{ij} is again defined, as for regression models, as the difference between the observed and fitted values:

$$e_{ij} = Y_{ij} - \hat{Y}_{ij} = Y_{ij} - \bar{Y}_{i\cdot} \tag{16.20}$$

Thus, a residual here represents the deviation of an observation from its estimated factor level mean.

An important property of the residuals for ANOVA model (16.2) is that they sum to zero for each factor level i:

$$\sum_j e_{ij} = 0 \qquad i = 1, \ldots, r \tag{16.21}$$

As for regression analysis, residuals for ANOVA models are useful for examining the appropriateness of the ANOVA model. We shall discuss this use of residuals in Chapter 18.

Example

Table 16.2 contains the residuals for the Kenton Food Company example. For instance, from Table 16.1, we find:

$$e_{11} = Y_{11} - \bar{Y}_{1\cdot} = 11 - 14.6 = -3.6$$

$$e_{21} = Y_{21} - \bar{Y}_{2\cdot} = 12 - 13.4 = -1.4$$

Note from Table 16.2 that the residuals sum to zero for each factor level, as expected.

TABLE 16.2 Residuals—Kenton Food Company Example.

Package Design	Store (j)					
i	1	2	3	4	5	Total
1	−3.6	2.4	1.4	−.6	.4	0
2	−1.4	−3.4	1.6	5.6	−2.4	0
3	3.5	.5	−1.5	−2.5		0
4	−.2	5.8	−5.2	−1.2	.8	0
All designs						0

16.5 Analysis of Variance

Just as the analysis of variance for a regression model partitions the total sum of squares into the regression sum of squares and the error sum of squares, so a corresponding partitioning exists for ANOVA model (16.2).

Partitioning of *SSTO*

The total variability of the Y_{ij} observations, not using any information about factor levels, is measured in terms of the total deviation of each observation, i.e., the deviation of Y_{ij} around the overall mean $\bar{Y}_{..}$:

$$Y_{ij} - \bar{Y}_{..} \tag{16.22}$$

When we utilize information about the factor levels, the deviations reflecting the uncertainty remaining in the data are those of each observation Y_{ij} around its respective estimated factor level mean $\bar{Y}_{i.}$:

$$Y_{ij} - \bar{Y}_{i.} \tag{16.23}$$

The difference between the deviations (16.22) and (16.23) reflects the difference between the estimated factor level mean and the overall mean:

$$(Y_{ij} - \bar{Y}_{..}) - (Y_{ij} - \bar{Y}_{i.}) = \bar{Y}_{i.} - \bar{Y}_{..} \tag{16.24}$$

Note from (16.24) that we can decompose the total deviation $Y_{ij} - \bar{Y}_{..}$ into two components:

$$\underbrace{Y_{ij} - \bar{Y}_{..}}_{\substack{\text{Total}\\\text{deviation}}} = \underbrace{\bar{Y}_{i.} - \bar{Y}_{..}}_{\substack{\text{Deviation of}\\\text{estimated}\\\text{factor level}\\\text{mean around}\\\text{overall mean}}} + \underbrace{Y_{ij} - \bar{Y}_{i.}}_{\substack{\text{Deviation}\\\text{around}\\\text{estimated}\\\text{factor}\\\text{level mean}}} \tag{16.25}$$

Thus, the total deviation $Y_{ij} - \bar{Y}_{..}$ can be viewed as the sum of two components:

1. The deviation of the estimated factor level mean around the overall mean.
2. The deviation of Y_{ij} around its estimated factor level mean, which is simply the residual e_{ij} according to (16.20).

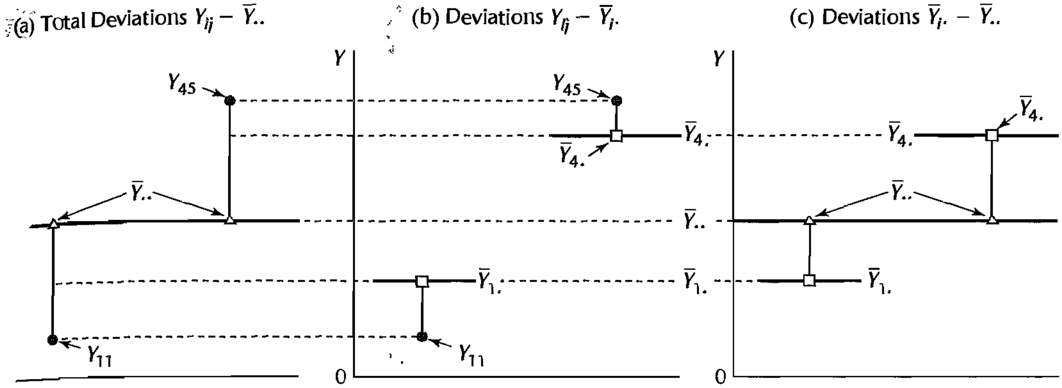
Figure 16.4 illustrates this decomposition for the Kenton Food Company example for two of the observations, Y_{11} and Y_{45}.

When we square both sides in (16.25) and then sum, the cross products on the right drop out and we obtain:

$$\sum_i \sum_j (Y_{ij} - \bar{Y}_{..})^2 = \sum_i n_i(\bar{Y}_{i.} - \bar{Y}_{..})^2 + \sum_i \sum_j (Y_{ij} - \bar{Y}_{i.})^2 \tag{16.26}$$

The term on the left measures the total variability of the Y_{ij} observations and is denoted, as

FIGURE 16.4 Illustration of Partitioning of Total Deviations $Y_{ij} - \bar{Y}_{..}$—Kenton Food Company Example (not drawn to scale; only observations Y_{11} and Y_{45} are shown).



for regression, by *SSTO* for *total sum of squares:*

$$SSTO = \sum_i \sum_j (Y_{ij} - \bar{Y}_{..})^2 \tag{16.27}$$

The first term on the right in (16.26) will be denoted by *SSTR*, standing for *treatment sum of squares:*

$$SSTR = \sum_i n_i(\bar{Y}_{i.} - \bar{Y}_{..})^2 \tag{16.28}$$

The second term on the right in (16.26) will be denoted by *SSE*, standing for *error sum of squares:*

$$SSE = \sum_i \sum_j (Y_{ij} - \bar{Y}_{i.})^2 = \sum_i \sum_j e_{ij}^2 \tag{16.29}$$

Thus, (16.26) can be written equivalently:

$$SSTO = SSTR + SSE \tag{16.30}$$

The correspondence to the regression decomposition in (2.50) is readily apparent.

The total sum of squares for the analysis of variance model is therefore made up of these two components:

1. *SSE:* A measure of the random variation of the observations around the respective estimated factor level means. The less variation among the observations for each factor level, the smaller is *SSE*. If $SSE = 0$, the observations for any given factor level are all the same, and this holds for all factor levels. The more the observations for each factor level differ among themselves, the larger will be *SSE*.

2. *SSTR:* A measure of the extent of differences between the estimated factor level means, based on the deviations of the estimated factor level means $\bar{Y}_{i.}$ around the overall mean $\bar{Y}_{..}$. If all estimated factor level means $\bar{Y}_{i.}$ are the same, then $SSTR = 0$. The more the estimated factor level means differ, the larger will be *SSTR*.

Example

The analysis of variance breakdown of the total sum of squares for the Kenton Food Company example in Table 16.1 is obtained as follows, using (16.27), (16.28), and (16.29):

$$SSTO = (11 - 18.63)^2 + (17 - 18.63)^2 + (16 - 18.63)^2 + \cdots + (28 - 18.63)^2$$
$$= 746.42$$
$$SSTR = 5(14.6 - 18.63)^2 + 5(13.4 - 18.63)^2 + 4(19.5 - 18.63)^2 + 5(27.2 - 18.63)^2$$
$$= 588.22$$
$$SSE = (11 - 14.6)^2 + (17 - 14.6)^2 + (16 - 14.6)^2 + \cdots + (28 - 27.2)^2$$
$$= 158.20$$

Thus, the decomposition of *SSTO* is:

$$746.42 = 588.22 + 158.20$$
$$SSTO = SSTR + SSE$$

Note that much of the total variation in the observations is associated with variation between the estimated factor level means.

Comments

1. To prove (16.26), we begin by considering (16.25):

$$Y_{ij} - \bar{Y}_{..} = (\bar{Y}_{i.} - \bar{Y}_{..}) + (Y_{ij} - \bar{Y}_{i.})$$

Squaring both sides we obtain:

$$(Y_{ij} - \bar{Y}_{..})^2 = (\bar{Y}_{i.} - \bar{Y}_{..})^2 + (Y_{ij} - \bar{Y}_{i.})^2 + 2(\bar{Y}_{i.} - \bar{Y}_{..})(Y_{ij} - \bar{Y}_{i.})$$

When we sum over all sample observations in the study (i.e., over both i and j), we obtain:

$$\sum_i \sum_j (Y_{ij} - \bar{Y}_{..})^2 = \sum_i \sum_j (\bar{Y}_{i.} - \bar{Y}_{..})^2 + \sum_i \sum_j (Y_{ij} - \bar{Y}_{i.})^2 + \sum_i \sum_j 2(\bar{Y}_{i.} - \bar{Y}_{..})(Y_{ij} - \bar{Y}_{i.}) \tag{16.31}$$

The first term on the right in (16.31) equals:

$$\sum_i \sum_j (\bar{Y}_{i.} - \bar{Y}_{..})^2 = \sum_i n_i (\bar{Y}_{i.} - \bar{Y}_{..})^2 \tag{16.32}$$

since $(\bar{Y}_{i.} - \bar{Y}_{..})^2$ is constant when summed over j; hence, n_i such terms are picked up for the summation over j.

The third term on the right in (16.31) equals zero:

$$\sum_i \sum_j 2(\bar{Y}_{i.} - \bar{Y}_{..})(Y_{ij} - \bar{Y}_{i.}) = 2 \sum_i (\bar{Y}_{i.} - \bar{Y}_{..}) \sum_j (Y_{ij} - \bar{Y}_{i.}) = 0 \tag{16.33}$$

This follows because $\bar{Y}_{i.} - \bar{Y}_{..}$ is constant for the summation over j; hence, it can be brought in front of the summation sign over j. Further, $\sum_j (Y_{ij} - \bar{Y}_{i.}) = 0$ for all i, since the sum of the deviations around the arithmetic mean is always zero.

Thus, (16.31) reduces to (16.26).

2. The squared estimated factor level mean deviations $(\bar{Y}_i. - \bar{Y}..)^2$ in *SSTR* in (16.28) are weighted by the number of cases n_i for that factor level. The reason is that for each observation Y_{ij} at factor level i, the deviation component $\bar{Y}_i. - \bar{Y}..$ is the same. ■

kdown of Degrees of Freedom

Corresponding to the decomposition of the total sum of squares, we can also obtain a breakdown of the associated degrees of freedom.

SSTO has $n_T - 1$ degrees of freedom associated with it. There are altogether n_T deviations $Y_{ij} - \bar{Y}..$, but one degree of freedom is lost because the deviations are not independent in that they must sum to zero; i.e., $\sum\sum(Y_{ij} - \bar{Y}..) = 0$.

SSTR has $r - 1$ degrees of freedom associated with it. There are r estimated factor level mean deviations $\bar{Y}_i. - \bar{Y}..$, but one degree of freedom is lost because the deviations are not independent in that the weighted sum must equal zero; i.e., $\sum n_i(\bar{Y}_i. - \bar{Y}..) = 0$.

SSE has $n_T - r$ degrees of freedom associated with it. This can be readily seen by considering the component of *SSE* for the ith factor level:

$$\sum_{j=1}^{n_i}(Y_{ij} - \bar{Y}_i.)^2 \tag{16.34}$$

The expression in (16.34) is the equivalent of a total sum of squares considering only the ith factor level. Hence, there are $n_i - 1$ degrees of freedom associated with this sum of squares. Since *SSE* is a sum of component sums of squares such as the one in (16.34), the degrees of freedom associated with *SSE* are the sum of the component degrees of freedom:

$$(n_1 - 1) + (n_2 - 1) + \cdots + (n_r - 1) = n_T - r \tag{16.35}$$

xample For the Kenton Food Company example, for which $n_T = 19$ and $r = 4$, the degrees of freedom associated with the three sums of squares are as follows:

SS	*df*
SSTO	19 − 1 = 18
SSTR	4 − 1 = 3
SSE	19 − 4 = 15

Note that degrees of freedom, like sums of squares, are additive:

$$18 = 3 + 15$$

Mean Squares

The mean squares, as usual, are obtained by dividing each sum of squares by its associated degrees of freedom. We therefore have:

$$MSTR = \frac{SSTR}{r - 1} \tag{16.36a}$$

$$MSE = \frac{SSE}{n_T - r} \tag{16.36b}$$

Here, $MSTR$ stands for *treatment mean square* and MSE, as before, stands for *error mean square*.

Example

For the Kenton Food Company example, we obtain from earlier results:

$$MSTR = \frac{588.22}{3} = 196.07$$

$$MSE = \frac{158.20}{15} = 10.55$$

Note that the two mean squares do not add to $SSTO/(n_T - 1) = 746.42/18 = 41.47$. Thus, the mean squares here, as in regression, are not additive.

Analysis of Variance Table

The breakdowns of the total sum of squares and degrees of freedom, together with the resulting mean squares, are presented in an ANOVA table such as Table 16.3. The ANOVA table for the Kenton Food Company example is presented in Figure 16.5 which contains the JMP output for single-factor analysis of variance. Note that the output contains the overall mean response ($\bar{Y} = 18.63158$), the number of observations, the ANOVA table, and the estimated factor level means $\bar{Y}_{i.}$. In this table, the line for the treatments source of variation is labeled "Package Design." The results in the JMP output are shown to more decimal places than we have shown, but are consistent with our calculations. Note also that the JMP ANOVA table shows the degrees of freedom column before the sum of squares column. The columns labeled "Std Error," "Lower 95%," and "Upper 95%" will be discussed in Chapter 17.

Expected Mean Squares

The expected values of MSE and $MSTR$ can be shown to be as follows:

$$E\{MSE\} = \sigma^2 \tag{16.37a}$$

$$E\{MSTR\} = \sigma^2 + \frac{\sum n_i(\mu_i - \mu_.)^2}{r - 1} \tag{16.37b}$$

where:

$$\mu_. = \frac{\sum n_i \mu_i}{n_T} \tag{16.37c}$$

is referred to as the weighted mean. These expected values are shown in the $E\{MS\}$ column of Table 16.3.

TABLE 16.3 ANOVA Table for Single-Factor Study.

Source of Variation	SS	df	MS	E{MS}
Between treatments	$SSTR = \sum n_i(\bar{Y}_{i.} - \bar{Y}_{..})^2$	$r - 1$	$MSTR = \frac{SSTR}{r-1}$	$\sigma^2 + \frac{\sum n_i(\mu_i - \mu_.)^2}{r-1}$
Error (within treatments)	$SSE = \sum\sum(Y_{ij} - \bar{Y}_{i.})^2$	$n_T - r$	$MSE = \frac{SSE}{n_T - r}$	σ^2
Total	$SSTO = \sum\sum(Y_{ij} - \bar{Y}_{..})^2$	$n_T - 1$		

GURE 16.5 rtion of JMP tput for gle-Factor alysis of Variance— enton Food mpany xample.

Oneway Anova

Summary of Fit

Rsquare	0.788055
Adj Rsquare	0.745666
Root Mean Square Error	3.247563
Mean of Response	18.63158
Observations (or Sum Wgts)	19

Analysis of Variance

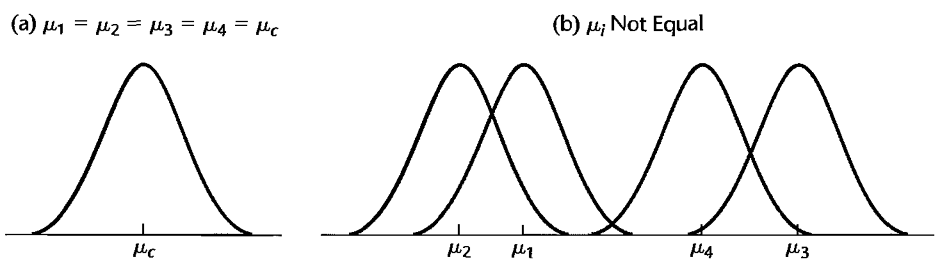
Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Package Design	3	588.22105	196.074	18.5911	<.0001
Error	15	158.20000	10.547		
C. Total	18	746.42105			

Means for Oneway Anova

Level	Number	Mean	Std Error	Lower 95%	Upper 95%
1	5	14.6000	1.4524	11.504	17.696
2	5	13.4000	1.4524	10.304	16.496
3	4	19.5000	1.6238	16.039	22.961
4	5	27.2000	1.4524	24.104	30.296

Std Error uses a pooled estimate of error variance

FIGURE 16.6 Sampling Distributions of $\bar{Y}_{i\cdot}$ for Four Treatments $(n_i \equiv n)$.



Two important features of the expected mean squares deserve attention:

1. *MSE* is an unbiased estimator of σ^2, the variance of the error terms ε_{ij}, whether or not the factor level means μ_i are equal. This is intuitively reasonable since the variability of the observations within each factor level is not affected by the magnitudes of the estimated factor level means for normal populations.

2. When all factor level means μ_i are equal and hence equal to the weighted mean $\mu_{\cdot}$, then $E\{MSTR\} = \sigma^2$ since the second term on the right in (16.37b) becomes zero. Hence, *MSTR* and *MSE* both estimate the error variance σ^2 when all factor level means μ_i are equal. When, however, the factor level means are not equal, *MSTR* tends on the average to be larger than *MSE*, since the second term in (16.37b) will then be positive. This is intuitively reasonable, as illustrated in Figure 16.6 for four treatments. The situation portrayed there assumes that all sample sizes are equal, i.e., $n_i \equiv n$. When all μ_i are equal, then all $\bar{Y}_{i\cdot}$ follow the same sampling distribution, with common mean μ_c and variance σ^2/n; this is portrayed in

Figure 16.6a. When the μ_i are not equal, on the other hand, the $\bar{Y}_{i\cdot}$ follow different sampling distributions, each with the same variability σ^2/n but centered on different means μ_i. One such possibility is shown in Figure 16.6b. Hence, the $\bar{Y}_{i\cdot}$ will tend to differ more from each other when the μ_i differ than when the μ_i are equal, and consequently *MSTR* will tend to be larger when the factor level means are not the same than when they are equal. This property of *MSTR* is utilized in constructing the statistical test discussed in the next section to determine whether or not the factor level means μ_i are the same. If *MSTR* and *MSE* are of the same order of magnitude, this is taken to suggest that the factor level means μ_i are equal. If *MSTR* is substantially larger than *MSE*, this is taken to suggest that the μ_i are not equal.

Comments

1. To find the expected value of *MSE*, we first note that *MSE* can be expressed as follows:

$$MSE = \frac{1}{n_T - r} \sum_i \sum_j (Y_{ij} - \bar{Y}_{i\cdot})^2$$

$$= \frac{1}{n_T - r} \sum_i \left[(n_i - 1) \frac{\sum_j (Y_{ij} - \bar{Y}_{i\cdot})^2}{n_i - 1} \right] \tag{16.38}$$

Now let us denote the ordinary sample variance of the observations for the ith factor level by s_i^2:

$$s_i^2 = \frac{\sum_j (Y_{ij} - \bar{Y}_{i\cdot})^2}{n_i - 1} \tag{16.39}$$

Hence, (16.38) can be expressed as follows:

$$MSE = \frac{1}{n_T - r} \sum_i (n_i - 1) s_i^2 \tag{16.40}$$

Since it is well known that the sample variance (16.39) is an unbiased estimator of the population variance, which in our case is σ^2 for all factor levels, we obtain:

$$E\{MSE\} = \frac{1}{n_T - r} \sum_i (n_i - 1) E\{s_i^2\}$$

$$= \frac{1}{n_T - r} \sum_i (n_i - 1)\sigma^2$$

$$= \sigma^2$$

2. We shall derive the expected value of *MSTR* for the special case when all sample sizes n_i are the same, namely, when $n_i \equiv n$. The general result in (16.37b) becomes for this special case:

$$E\{MSTR\} = \sigma^2 + \frac{n \sum (\mu_i - \mu_\cdot)^2}{r - 1} \qquad \text{when } n_i \equiv n \tag{16.41}$$

Further, when all factor level sample sizes are n, *MSTR* as defined in (16.28) and (16.36a) becomes:

$$MSTR = \frac{n \sum (\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^2}{r - 1} \qquad \text{when } n_i \equiv n \tag{16.42}$$

To derive (16.41), consider the model formulation for Y_{ij} in (16.2):

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

Averaging the Y_{ij} for the ith factor level, we obtain:

$$\bar{Y}_{i\cdot} = \mu_i + \bar{\varepsilon}_{i\cdot} \tag{16.43}$$

where $\bar{\varepsilon}_{i\cdot}$ is the average of the ε_{ij} for the ith factor level:

$$\bar{\varepsilon}_{i\cdot} = \frac{\sum_j \varepsilon_{ij}}{n} \tag{16.44}$$

Averaging the Y_{ij} over all factor levels, we obtain:

$$\bar{Y}_{\cdot\cdot} = \mu_{\cdot} + \bar{\varepsilon}_{\cdot\cdot} \tag{16.45}$$

where $\mu_{\cdot}$, which is defined in (16.37c), becomes for $n_i \equiv n$:

$$\mu_{\cdot} = \frac{n\sum \mu_i}{nr} = \frac{\sum \mu_i}{r} \qquad \text{when } n_i \equiv n \tag{16.46}$$

and $\bar{\varepsilon}_{\cdot\cdot}$ is the average of all ε_{ij}:

$$\bar{\varepsilon}_{\cdot\cdot} = \frac{\sum\sum \varepsilon_{ij}}{nr} \tag{16.47}$$

Since the sample sizes are equal, we also have:

$$\bar{Y}_{\cdot\cdot} = \frac{\sum \bar{Y}_{i\cdot}}{r} \qquad \bar{\varepsilon}_{\cdot\cdot} = \frac{\sum \bar{\varepsilon}_{i\cdot}}{r} \tag{16.48}$$

Using (16.43) and (16.45), we obtain:

$$\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot} = (\mu_i + \bar{\varepsilon}_{i\cdot}) - (\mu_{\cdot} + \bar{\varepsilon}_{\cdot\cdot}) = (\mu_i - \mu_{\cdot}) + (\bar{\varepsilon}_{i\cdot} - \bar{\varepsilon}_{\cdot\cdot}) \tag{16.49}$$

When we square $\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot}$ and sum over the factor levels, we obtain:

$$\sum(\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^2 = \sum(\mu_i - \mu_{\cdot})^2 + \sum(\bar{\varepsilon}_{i\cdot} - \bar{\varepsilon}_{\cdot\cdot})^2 + 2\sum(\mu_i - \mu_{\cdot})(\bar{\varepsilon}_{i\cdot} - \bar{\varepsilon}_{\cdot\cdot}) \tag{16.50}$$

We now wish to find $E\{\sum(\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^2\}$, and therefore need to find the expected value of each term on the right in (16.50):

a. Since $\sum(\mu_i - \mu_{\cdot})^2$ is a constant, its expectation is:

$$E\left\{\sum(\mu_i - \mu_{\cdot})^2\right\} = \sum(\mu_i - \mu_{\cdot})^2 \tag{16.51}$$

b. Before finding the expectation of the second term on the right, consider first the expression:

$$\frac{\sum(\bar{\varepsilon}_{i\cdot} - \bar{\varepsilon}_{\cdot\cdot})^2}{r-1}$$

This is an ordinary sample variance, since $\bar{\varepsilon}_{\cdot\cdot}$ is the sample mean of the r terms $\bar{\varepsilon}_{i\cdot}$ per (16.48). We further know that the sample variance is an unbiased estimator of the variance of the variable, in this case the variable being $\bar{\varepsilon}_{i\cdot}$. But $\bar{\varepsilon}_{i\cdot}$ is just the mean of n independent error terms ε_{ij} by (16.44). Hence:

$$\sigma^2\{\bar{\varepsilon}_{i\cdot}\} = \frac{\sigma^2\{\varepsilon_{ij}\}}{n} = \frac{\sigma^2}{n}$$

Therefore:

$$E\left\{\frac{\sum(\bar{\varepsilon}_{i\cdot} - \bar{\varepsilon}_{\cdot\cdot})^2}{r-1}\right\} = \frac{\sigma^2}{n}$$

so that:

$$E\left\{\sum(\bar{\varepsilon}_{i\cdot} - \bar{\varepsilon}_{\cdot\cdot})^2\right\} = \frac{(r-1)\sigma^2}{n} \tag{16.52}$$

c. Since both $\bar{\varepsilon}_{i\cdot}$ and $\bar{\varepsilon}_{\cdot\cdot}$ are means of ε_{ij} terms, all of which have expectation 0, it follows that:

$$E\{\bar{\varepsilon}_{i\cdot}\} = 0 \qquad E\{\bar{\varepsilon}_{\cdot\cdot}\} = 0$$

Hence:

$$E\left\{2\sum(\mu_i - \mu_\cdot)(\bar{\varepsilon}_{i\cdot} - \bar{\varepsilon}_{\cdot\cdot})\right\} = 2\sum(\mu_i - \mu_\cdot)E\{\bar{\varepsilon}_{i\cdot} - \bar{\varepsilon}_{\cdot\cdot}\} = 0 \tag{16.53}$$

We have thus shown, by (16.51), (16.52), and (16.53), that:

$$E\left\{\sum(\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^2\right\} = \sum(\mu_i - \mu_\cdot)^2 + \frac{(r-1)\sigma^2}{n}$$

But then (16.41) follows at once:

$$E\{MSTR\} = E\left\{\frac{n\sum(\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^2}{r-1}\right\} = \frac{n}{r-1}\left[\sum(\mu_i - \mu_\cdot)^2 + \frac{(r-1)\sigma^2}{n}\right]$$

$$= \sigma^2 + \frac{n\sum(\mu_i - \mu_\cdot)^2}{r-1} \qquad \text{when } n_i \equiv n$$

■

16.6 F Test for Equality of Factor Level Means

It is customary to begin the analysis of a single-factor study by determining whether or not the factor level means μ_i are equal. If, for instance, the four package designs in the Kenton Food Company example lead to the same mean sales volumes, there is no need for further analysis, such as to determine which design is best or how two particular designs compare in stimulating sales.

Thus, the alternative conclusions we wish to consider are:

$$\begin{aligned} &H_0\colon \mu_1 = \mu_2 = \cdots = \mu_r \\ &H_a\colon \text{not all } \mu_i \text{ are equal} \end{aligned} \tag{16.54}$$

Test Statistic

The test statistic to be used for choosing between the alternatives in (16.54) is:

$$F^* = \frac{MSTR}{MSE} \tag{16.55}$$

Note that *MSTR* here plays the role corresponding to *MSR* for a regression model.

Large values of F^* support H_a, since *MSTR* will tend to exceed *MSE* when H_a holds, as we saw from (16.37). Values of F^* near 1 support H_0, since both *MSTR* and *MSE* have the same expected value when H_0 holds. Hence, the appropriate test is an upper-tail one.

ution of F^*

When all treatment means μ_i are equal, each response Y_{ij} has the same expected value. In view of the additivity of sums of squares and degrees of freedom, Cochran's theorem (2.61) then implies:

$$\text{When } H_0 \text{ holds, } \frac{SSE}{\sigma^2} \text{ and } \frac{SSTR}{\sigma^2} \text{ are independent } \chi^2 \text{ variables}$$

It follows in the same fashion as for regression:

$$\text{When } H_0 \text{ holds, } F^* \text{ is distributed as } F(r-1, n_T - r)$$

When H_a holds, that is, when the μ_i are not all equal, F^* does *not* follow the F distribution. Rather, it follows a complex distribution called the *noncentral F distribution*. We shall make use of the noncentral F distribution when we discuss the power of the F test in Section 16.10.

Comment

SSTR and *SSE* are independent even if all μ_i are not equal. *SSTR* is solely based on the estimated factor level means $\bar{Y}_{i.}$. On the other hand, *SSE* reflects the variability within the factor level samples, and this within-sample variability is not affected by the magnitudes of the estimated factor level means when the error terms are normally distributed. ■

onstruction of Decision Rule

Usually, the risk of making a Type I error is controlled in constructing the decision rule. This provides protection against making further, more detailed, analyses of the factor effects when in fact there are no differences in the factor level means. The Type II error can also be controlled, as we shall see later in Section 16.10, through sample size determination.

Since we know that F^* is distributed as $F(r-1, n_T - r)$ when H_0 holds and that large values of F^* lead to conclusion H_a, the appropriate decision rule to control the level of significance at α is:

$$\begin{aligned} &\text{If } F^* \leq F(1-\alpha; r-1, n_T - r), \text{ conclude } H_0 \\ &\text{If } F^* > F(1-\alpha; r-1, n_T - r), \text{ conclude } H_a \end{aligned} \tag{16.56}$$

where $F(1-\alpha; r-1, n_T - r)$ is the $(1-\alpha)100$ percentile of the appropriate F distribution.

Example

For the Kenton Food Company example, we wish to test whether or not mean sales are the same for the four package designs:

$$\begin{aligned} &H_0\text{: } \mu_1 = \mu_2 = \mu_3 = \mu_4 \\ &H_a\text{: not all } \mu_i \text{ are equal} \end{aligned}$$

Management wishes to control the risk of making a Type I error at $\alpha = .05$. We therefore require $F(.95; 3, 15)$, where the degrees of freedom are those shown in Figure 16.5. From Table B.4 in Appendix B, we find $F(.95; 3, 15) = 3.29$. Hence, the decision rule is:

$$\begin{aligned} &\text{If } F^* \leq 3.29, \text{ conclude } H_0 \\ &\text{If } F^* > 3.29, \text{ conclude } H_a \end{aligned}$$

Using the data in the ANOVA table in Figure 16.5, we obtain the test statistic:

$$F^* = \frac{MSTR}{MSE} = \frac{196.07}{10.55} = 18.6$$

Since $F^* = 18.6 > 3.29$, we conclude H_a, that the factor level means μ_i are not equal, or that the four different package designs do not lead to the same mean sales volume. Thus, we conclude that there is a relation between package design and sales volume.

The P-value for the test statistic is the probability $P\{F(3, 15) > F^* = 18.6\}$, which is .00003. This P-value again indicates that the data from the experiment are not consistent with all designs having the same effect on sales volume.

The conclusion of a relation between package design and sales volume did not surprise the sales manager of the Kenton Food Company. The study was conducted in the first place because the sales manager expected the four package designs to have different effects on sales volume and was interested in finding out the nature of these differences. In the next chapter, we discuss the second stage of the analysis, namely, how to study the nature of the factor level means when differences exist.

Comments

1. If there are only two factor levels so that $r = 2$, it can easily be shown that the test employing F^* in (16.55) is the equivalent of the two-population, two-sided t test in Table A.2a. The F test here has $(1, n_T - 2)$ degrees of freedom, and the t test has $n_1 + n_2 - 2$ or $n_T - 2$ degrees of freedom; thus both tests lead to equivalent critical regions. For comparing two population means, the t test generally is to be preferred since it can be used to conduct both two-sided and one-sided tests (Table A.2); the F test can be used only for two-sided tests.

2. Since the F test for testing the alternatives (16.54) is a test of a linear statistical model, it can be obtained by the general linear test approach explained in Section 2.8:

a. The full model is ANOVA model (16.2):

$$Y_{ij} = \mu_i + \varepsilon_{ij} \qquad \text{Full model} \tag{16.57}$$

Fitting the full model by either the method of least squares or the method of maximum likelihood leads to the fitted values $\hat{Y}_{ij} = \bar{Y}_{i\cdot}$, per (16.17), and to the resulting error sum of squares:

$$SSE(F) = \sum\sum(Y_{ij} - \hat{Y}_{ij})^2 = \sum\sum(Y_{ij} - \bar{Y}_{i\cdot})^2$$

$SSE(F)$ has $df_F = n_T - r$ degrees of freedom associated with it because r parameter values $(\mu_1, \ldots, \mu_r)$ have to be estimated.

b. The reduced model under H_0 is:

$$Y_{ij} = \mu_c + \varepsilon_{ij} \qquad \text{Reduced model} \tag{16.58}$$

where μ_c is the common mean for all factor levels. Fitting the reduced model leads to the estimator $\hat{\mu}_c = \bar{Y}_{\cdot\cdot}$ so that all fitted values are $\hat{Y}_{ij} \equiv \bar{Y}_{\cdot\cdot}$, and the resulting error sum of squares is:

$$SSE(R) = \sum\sum(Y_{ij} - \hat{Y}_{ij})^2 = \sum\sum(Y_{ij} - \bar{Y}_{\cdot\cdot})^2$$

The degrees of freedom associated with $SSE(R)$ are $df_R = n_T - 1$ because one parameter (μ_c) had to be estimated.

c. Since, according to (16.27) and (16.29), respectively:

$$SSE(R) = SSTO$$

$$SSE(F) = SSE$$

and since by (16.30) $SSTO - SSE = SSTR$, the general linear test statistic (2.70) becomes here:

$$F^* = \frac{SSE(R) - SSE(F)}{df_R - df_F} \div \frac{SSE(F)}{df_F}$$

$$= \frac{SSTO - SSE}{(n_T - 1) - (n_T - r)} \div \frac{SSE}{n_T - r} = \frac{SSTR}{r - 1} \div \frac{SSE}{n_T - r} = \frac{MSTR}{MSE}$$ ■

6.7 Alternative Formulation of Model

actor Effects Model

At times, an alternative but completely equivalent formulation of the single-factor ANOVA model in (16.2) is used. This alternative formulation is called the *factor effects model*. With this alternative formulation, the treatment means μ_i are expressed in an equivalent fashion by means of the identity:

$$\mu_i \equiv \mu. + (\mu_i - \mu.) \tag{16.59}$$

where $\mu.$ is a constant that can be defined to fit the purpose of the study. We shall denote the difference $\mu_i - \mu.$ by τ_i:

$$\tau_i = \mu_i - \mu. \tag{16.60}$$

so that (16.59) can be expressed in equivalent fashion as:

$$\mu_i \equiv \mu. + \tau_i \tag{16.61}$$

The difference $\tau_i = \mu_i - \mu.$ is called the *ith factor level effect* or the *ith treatment effect*.

The ANOVA model in (16.2) can now be stated equivalently as follows:

$$Y_{ij} = \mu. + \tau_i + \varepsilon_{ij} \tag{16.62}$$

where:

$\mu.$ is a constant component common to all observations

τ_i is the effect of the ith factor level (a constant for each factor level)

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

$i = 1, \ldots, r;\ j = 1, \ldots, n_i$

ANOVA model (16.62) is called a factor effects model because it is expressed in terms of the factor effects τ_i, in distinction to the cell means model (16.2), which is expressed in terms of the cell (treatment) means μ_i.

Factor effects model (16.62) is a linear model, like the equivalent cell means model (16.2). We shall demonstrate this in the next section.

Definition of $\mu.$

The splitting up of the factor level mean μ_i into two components, an overall constant $\mu.$ and a factor level or treatment effect τ_i, depends on the definition of $\mu.$, which can be defined in many ways. We now explain two basic ways to define $\mu.$.

Unweighted Mean. Often, a definition of $\mu.$ as the unweighted average of all factor level means μ_i is found to be useful:

$$\mu. = \frac{\sum_{i=1}^{r} \mu_i}{r} \tag{16.63}$$

This definition implies that:

$$\sum_{i=1}^{r} \tau_i = 0 \tag{16.64}$$

because by (16.60) we have:

$$\sum \tau_i = \sum (\mu_i - \mu.) = \sum \mu_i - r\mu.$$

and by (16.63) we have:

$$\sum \mu_i = r\mu.$$

Thus, the definition of the overall constant $\mu.$ in (16.63) implies a restriction on the τ_i, in this case that their sum must be zero.

Example

For the earlier incentive pay example in Figure 16.2, we have $\mu_1 = 70$, $\mu_2 = 58$, $\mu_3 = 90$, and $\mu_4 = 84$. When $\mu.$ is defined according to (16.63), we obtain:

$$\mu. = \frac{70 + 58 + 90 + 84}{4} = 75.5$$

Hence:

$$\tau_1 = 70 - 75.5 = -5.5$$

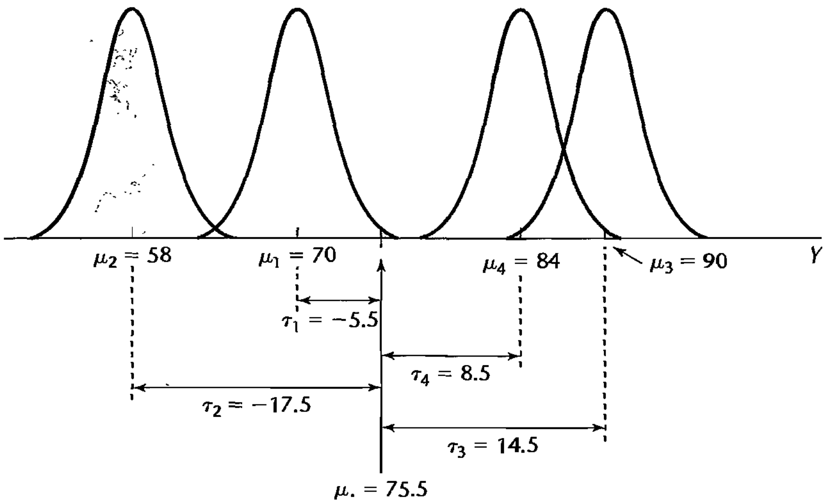
$$\tau_2 = 58 - 75.5 = -17.5$$

$$\tau_3 = 90 - 75.5 = 14.5$$

$$\tau_4 = 84 - 75.5 = 8.5$$

The first treatment effect $\tau_1 = -5.5$, for instance, indicates that the mean employee productivity for incentive pay type 1 is 5.5 units less than the average productivity for all four types of incentive pay. Figure 16.7 provides an illustration of these treatment effects.

FIGURE 16.7 Illustration of Treatment Effects—Incentive Pay Example.



Weighted Mean The constant $\mu_.$ can also be defined as some weighted average of the factor level means μ_i:

$$\mu_. = \sum_{i=1}^{r} w_i \mu_i \qquad \text{where } \sum_{i=1}^{r} w_i = 1 \tag{16.65}$$

Note that the w_i are weights defined so that their sum is 1. The restriction on the τ_i implied by definition (16.65) is:

$$\sum_{i=1}^{r} w_i \tau_i = 0 \tag{16.66}$$

This follows in the same fashion as (16.64).

The choice of weights w_i should depend on the meaningfulness of the resulting overall mean $\mu_.$. We present now two examples where different weightings are appropriate: (1) weighting according to a known measure of importance and (2) weighting according to sample size.

Example 1

A car rental firm wanted to estimate the average fuel consumption (in miles per gallon) for its large fleet of cars, which consists of 50 percent compacts, 30 percent sedans, and 20 percent station wagons. Here, a meaningful measure of $\mu_.$ might be in terms of overall mean fuel consumption:

$$\mu_. = .5\mu_1 + .3\mu_2 + .2\mu_3 \tag{16.67}$$

where μ_1, μ_2, and μ_3 are the mean fuel consumptions for the three types of cars in the fleet. An estimate of $\mu_.$ here is:

$$\hat{\mu}_. = .5\bar{Y}_{1.} + .3\bar{Y}_{2.} + .2\bar{Y}_{3.} \tag{16.68}$$

Example 2

When exact weights are unknown, the subgroup sample sizes may be useful as weights of relative importance. For instance, the proportions of households in a city with no children, one child, and more than one child are not known. A random sample of n_T households was selected, which contained n_1 households with no child, n_2 households with one child, and n_3 households with more than one child. For testing whether mean entertainment expenditures are the same for the three types of households, use of the proportions n_1/n_T, n_2/n_T, and n_3/n_T as weights might be meaningful. The resulting definition of the overall entertainment expenditures constant $\mu.$ would then be:

$$\mu. = \frac{n_1}{n_T}\mu_1 + \frac{n_2}{n_T}\mu_2 + \frac{n_3}{n_T}\mu_3 \tag{16.69}$$

This quantity would be estimated by $\bar{Y}..$:

$$\hat{\mu}. = \frac{n_1}{n_T}\bar{Y}_1. + \frac{n_2}{n_T}\bar{Y}_2. + \frac{n_3}{n_T}\bar{Y}_3. = \bar{Y}.. \tag{16.70}$$

When all sample sizes are equal, $\mu.$ as defined in (16.69) reduces to the unweighted mean (16.63).

Test for Equality of Factor Level Means

Since the factor effects model (16.62) is equivalent to the cell means model (16.2), the test for equality of factor level means uses the same test statistic F^* in (16.55). The only difference is in the statement of the alternatives. For the cell means model (16.2), the alternatives are as specified in (16.54):

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

For the factor effects model (16.62), these same alternatives in terms of the factor effects are:

$$\begin{aligned} &H_0: \tau_1 = \tau_2 = \cdots = \tau_r = 0 \\ &H_a: \text{not all } \tau_i \text{ equal zero} \end{aligned} \tag{16.71}$$

The equivalence of the two forms can be readily established. The equality of the factor level means $\mu_1 = \mu_2 = \cdots = \mu_r$ implies that all τ_i are equal. The equalities of the τ_i follow from (16.61) since the constant term $\mu.$ is common to all factor level effects τ_i. The equality of the factor level means in turn implies that all $\tau_i = 0$, whether the restriction on the τ_i is of the form in (16.64) or (16.66). In either case, the restriction can be satisfied in only one way given the equality of the τ_i, namely, that $\tau_i \equiv 0$. Thus, it is equivalent to state that all factor level means μ_i are equal or that all factor level effects τ_i equal zero.

16.8 Regression Approach to Single-Factor Analysis of Variance

We noted earlier that cell means model (16.2) is a linear model, and that we can obtain test statistic F^* for testing the equality of the factor level means μ_i by means of the general linear test (2.70). We shall now explain the regression approach to single-factor analysis of variance for three alternative models: (1) the factor effects model with unweighted mean, (2) the factor effects model with weighted mean, and (3) the cell means model. It is important to emphasize that the choice of model affects the definition of the model parameters, and not the outcome of the test for equality of factor level means.

Factor Effects Model with Unweighted Mean

To state ANOVA model (16.62):

$$Y_{ij} = \mu. + \tau_i + \varepsilon_{ij}$$

as a regression model, we need to represent the parameters $\mu.$, $\tau_1, \ldots, \tau_r$ in the model. However, constraint (16.64) for the case of equal weightings:

$$\sum_{i=1}^{r} \tau_i = 0$$

implies that one of the r parameters τ_i is not needed since it can be expressed in terms of the other $r - 1$ parameters. We shall drop the parameter τ_r, which according to constraint (16.64) can be expressed in terms of the other $r - 1$ parameters τ_i as follows:

$$\tau_r = -\tau_1 - \tau_2 - \cdots - \tau_{r-1} \tag{16.72}$$

Thus, we shall use only the parameters $\mu.$, $\tau_1, \ldots, \tau_{r-1}$ for the linear model.

To illustrate how a linear model is developed with this approach, consider a single-factor study with $r = 3$ factor levels when $n_1 = n_2 = n_3 = 2$. The $\mathbf{Y}$, $\mathbf{X}$, $\boldsymbol{\beta}$, and $\boldsymbol{\varepsilon}$ matrices for this case are as follows:

$$\mathbf{Y} = \begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{21} \\ Y_{22} \\ Y_{31} \\ Y_{32} \end{bmatrix} \qquad \mathbf{X} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & -1 & -1 \\ 1 & -1 & -1 \end{bmatrix} \qquad \boldsymbol{\beta} = \begin{bmatrix} \mu. \\ \tau_1 \\ \tau_2 \end{bmatrix} \qquad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{21} \\ \varepsilon_{22} \\ \varepsilon_{31} \\ \varepsilon_{32} \end{bmatrix} \tag{16.73}$$

Note that the vector of expected values, $\mathbf{E}\{\mathbf{Y}\} = \mathbf{X}\boldsymbol{\beta}$, yields the following:

$$\mathbf{E}\{\mathbf{Y}\} = \begin{bmatrix} E\{Y_{11}\} \\ E\{Y_{12}\} \\ E\{Y_{21}\} \\ E\{Y_{22}\} \\ E\{Y_{31}\} \\ E\{Y_{32}\} \end{bmatrix} = \mathbf{X}\boldsymbol{\beta} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & -1 & -1 \\ 1 & -1 & -1 \end{bmatrix} \begin{bmatrix} \mu. \\ \tau_1 \\ \tau_2 \end{bmatrix} = \begin{bmatrix} \mu. + \tau_1 \\ \mu. + \tau_1 \\ \mu. + \tau_2 \\ \mu. + \tau_2 \\ \mu. - \tau_1 - \tau_2 \\ \mu. - \tau_1 - \tau_2 \end{bmatrix} \tag{16.74}$$

Since $\tau_3 = -\tau_1 - \tau_2$ according to (16.72), we see that $E\{Y_{31}\} = E\{Y_{32}\} = \mu. + \tau_3$. Thus, the above $\mathbf{X}$ matrix and $\boldsymbol{\beta}$ vector representation provides in all cases the appropriate expected values:

$$E\{Y_{ij}\} = \mu. + \tau_i$$

The illustration in (16.73) indicates how we need to define in general the multiple regression model so that it is the equivalent of the single-factor ANOVA model (16.62). Note that we require indicator variables that take on values 0, 1, or −1. This coding was discussed in Section 8.1. While this coding is not as simple as a 0, 1 coding, it is desirable here because it leads to regression coefficients in the $\boldsymbol{\beta}$ vector that are the parameters in the factor effects ANOVA model, i.e., $\mu.$, $\tau_1 \ldots, \tau_{r-1}$.

We shall let X_{ij1} denote the value of indicator variable X_1 for the jth case from the ith factor level, X_{ij2} the value of indicator variable X_2 for this same case, and so on, using altogether $r - 1$ indicator variables in the model. The multiple regression model then is as follows:

$$Y_{ij} = \mu. + \tau_1 X_{ij1} + \tau_2 X_{ij2} + \cdots + \tau_{r-1} X_{ij,r-1} + \varepsilon_{ij} \qquad \text{Full model} \tag{16.75}$$

where:

$$X_{ij1} = \begin{cases} 1 & \text{if case from factor level 1} \\ -1 & \text{if case from factor level } r \\ 0 & \text{otherwise} \end{cases}$$

$$\vdots$$

$$X_{ij,r-1} = \begin{cases} 1 & \text{if case from factor level } r-1 \\ -1 & \text{if case from factor level } r \\ 0 & \text{otherwise} \end{cases}$$

Note how the ANOVA model parameters play the role of regression function parameters in (16.75); the intercept term is $\mu.$, and the regression coefficients are $\tau_1, \tau_2, \ldots, \tau_{r-1}$.

The least squares estimator of $\mu.$ is the average of the cell sample means:

$$\hat{\mu}. = \frac{\sum_{i=1}^{r} \bar{Y}_{i\cdot}}{r} \tag{16.75a}$$

Note that this quantity is generally not the same as the overall mean $\bar{Y}_{\cdot\cdot}$ unless the cell sample sizes are equal. Also, the least squares estimator of the ith factor effect is:

$$\hat{\tau}_i = \bar{Y}_{i\cdot} - \hat{\mu}. \tag{16.75b}$$

To test the equality of the treatment means μ_i by means of the regression approach, we state the alternatives in the equivalent formulation (16.71), noting that τ_r must equal zero when $\tau_1 = \tau_2 = \cdots = \tau_{r-1} = 0$ according to (16.72):

$$\begin{aligned} &H_0\colon \tau_1 = \tau_2 = \cdots = \tau_{r-1} = 0 \\ &H_a\colon \text{not all } \tau_i \text{ equal zero} \end{aligned} \tag{16.76}$$

Note that H_0 states that all regression coefficients in regression model (16.75) are zero, and the reduced model is therefore:

$$Y_{ij} = \mu. + \varepsilon_{ij} \qquad \text{Reduced model} \tag{16.77}$$

Thus, we employ the usual test statistic (6.39b) for testing whether or not there is a regression relation:

$$F^* = \frac{MSR}{MSE} \tag{16.78}$$

Example

To test the equality of mean sales for the four cereal package designs in the Kenton Food Company example by means of the regression approach, we shall employ the regression model:

$$Y_{ij} = \mu. + \tau_1 X_{ij1} + \tau_2 X_{ij2} + \tau_3 X_{ij3} + \varepsilon_{ij} \tag{16.79}$$

where:

$$X_{ij1} = \begin{cases} 1 & \text{if case from factor level 1} \\ -1 & \text{if case from factor level 4} \\ 0 & \text{otherwise} \end{cases}$$

$$X_{ij2} = \begin{cases} 1 & \text{if case from factor level 2} \\ -1 & \text{if case from factor level 4} \\ 0 & \text{otherwise} \end{cases}$$

$$X_{ij3} = \begin{cases} 1 & \text{if case from factor level 3} \\ -1 & \text{if case from factor level 4} \\ 0 & \text{otherwise} \end{cases}$$

A portion of the data in Table 16.1 is repeated in Table 16.4a, together with the coding of the indicator variables X_1, X_2, and X_3. For observation Y_{11}, for instance, note that $X_1 = 1$, $X_2 = 0$, and $X_3 = 0$; hence, we obtain from (16.79):

$$E\{Y_{11}\} = \mu. + \tau_1$$

TABLE 16.4 Regression Approach to the Analysis of Variance—Kenton Food Company Example.

(a) Data for Regression Model (16.79)

i	j	Y_{ij}	X_{ij1}	X_{ij2}	X_{ij3}
1	1	11	1	0	0
1	2	17	1	0	0
1	3	16	1	0	0
1	4	14	1	0	0
1	5	15	1	0	0
2	1	12	0	1	0
...	...	...	...	...	...
4	4	26	−1	−1	−1
4	5	28	−1	−1	−1

(b) Fitted Regression Function

$$\hat{Y} = 18.675 - 4.075X_1 - 5.275X_2 + .825X_3$$

(c) Regression Analysis of Variance Table

Source of Variation	*SS*	*df*	*MS*
Regression	*SSR* = 588.22	3	*MSR* = 196.07
Error	*SSE* = 158.20	15	*MSE* = 10.55
Total	*SSTO* = 746.42	18	

Similarly, for observation Y_{45} we have $X_1 = -1$, $X_2 = -1$, and $X_3 = -1$; hence:

$$E\{Y_{45}\} = \mu. - \tau_1 - \tau_2 - \tau_3 = \mu. + \tau_4$$

since $\tau_4 = -\tau_1 - \tau_2 - \tau_3$.

Note that we employ the following codings in the indicator variables for cases from each of the four factor levels:

	Coding		
Factor Level	X_1	X_2	X_3
1	1	0	0
2	0	1	0
3	0	0	1
4	−1	−1	−1

A computer run of the multiple regression of Y on X_1, X_2, and X_3 yielded the fitted regression function and analysis of variance table presented in Tables 16.4b and 16.4c. Test statistic (16.78) therefore is:

$$F^* = \frac{MSR}{MSE} = \frac{196.07}{10.55} = 18.6$$

This is the same test statistic obtained earlier based on the analysis of variance calculations. Indeed, the analysis of variance table in Table 16.4c obtained with the regression approach is the same as the one in Figure 16.5 obtained with the analysis of variance approach except that the treatment sum of squares and mean square are called the regression sum of squares and mean square in Table 16.4c. From this point on, the test procedure based on the regression approach parallels the analysis of variance test procedure explained earlier.

Note that in the fitted regression function in Table 16.4b, the intercept term $\hat{\mu}. = 18.675$ is the unweighted average of the estimated factor level means $\bar{Y}_{i\cdot}$, not the overall mean $\bar{Y}_{\cdot\cdot}$, because $\mu.$ was defined as the unweighted average of the factor level means μ_i. The regression coefficient $b_1 = \hat{\tau}_1 = \bar{Y}_{i\cdot} - \hat{\mu}. = 14.6 - 18.675 = -4.075$ is simply the difference between the estimated mean in the first cell and the unweighted overall mean. b_2 and b_3 represent similar differences between the estimated factor level mean and the overall unweighted mean.

Comment

The regression approach is not utilized generally for ordinary analysis of variance problems. The reason is that the **X** matrix for analysis of variance problems usually is of a very simple structure, as we have seen earlier. This simple structure permits computational simplifications that are explicitly recognized in the statistical procedures for analysis of variance. We take up the regression approach to analysis of variance here, and in later chapters, for two principal reasons. First, we see that analysis of variance models are encompassed by the general linear statistical model (6.19). Second, the regression approach is very useful for analyzing some multifactor studies when the structure of the **X** matrix is not simple. ■

actor Effects Model with Weighted Mean

When the factor effects model (16.62) is used with a weighted mean, a modification of the coding scheme in (16.75) is required. The new coding scheme leads to changes in the definitions of the regression coefficients. We describe the new coding scheme and summarize the changes in the context of the proportional sample size weights, $w_i = n_i/n_T$.

When the constant $\mu.$ is the weighted average of the factor level means using proportional sample size weights, we have, from (16.65):

$$\mu. = \sum_{i=1}^{r} w_i \mu_i = \sum_{i=1}^{r} \frac{n_i}{n_T} \mu_i \tag{16.80a}$$

From (16.66), the restriction on the τ_i is:

$$\sum_{i=1}^{r} \frac{n_i}{n_T} \tau_i = 0$$

Solving for τ_r, we find:

$$\tau_r = -\frac{n_1}{n_r}\tau_1 - \frac{n_2}{n_r}\tau_2 - \cdots - \frac{n_{r-1}}{n_r}\tau_{r-1} \tag{16.80b}$$

This leads to the weighted model:

$$Y_{ij} = \mu. + \tau_1 X_{ij1} + \tau_2 X_{ij2} + \cdots + \tau_{r-1} X_{ij,r-1} + \varepsilon_{ij} \qquad \text{Full model} \tag{16.81}$$

where:

$$X_{ij1} = \begin{cases} 1 & \text{if case from factor level 1} \\ -\dfrac{n_1}{n_r} & \text{if case from factor level } r \\ 0 & \text{otherwise} \end{cases}$$

$$\vdots$$

$$X_{ij,r-1} = \begin{cases} 1 & \text{if case from factor level } r-1 \\ -\dfrac{n_{r-1}}{n_r} & \text{if case from factor level } r \\ 0 & \text{otherwise} \end{cases}$$

Note that if all cell sample sizes are equal, the mean $\mu.$ is the unweighted mean, and the coding scheme above is the same as the unweighted coding scheme used in (16.75), since $-n_i/n_r = -1$ for $i = 1, \ldots, r-1$.

When the sample sizes are not all equal, as noted in (16.70), the least squares estimate of the weighted mean $\mu.$ is the overall mean $\bar{Y}_{..}$, and the least squares estimate of the ith factor effect τ_i is $\bar{Y}_{i.} - \bar{Y}_{..}$.

Example

In the Kenton Food Company example, weighted mean model (16.81) is:

$$Y_{ij} = \mu. + \tau_1 X_{ij1} + \tau_2 X_{ij2} + \tau_3 X_{ij3} + \varepsilon_{ij} \tag{16.82}$$

where:

$$X_{ij1} = \begin{cases} 1 & \text{if case from factor level 1} \\ -\frac{5}{5} & \text{if case from factor level 4} \\ 0 & \text{otherwise} \end{cases}$$

$$X_{ij2} = \begin{cases} 1 & \text{if case from factor level 2} \\ -\frac{5}{5} & \text{if case from factor level 4} \\ 0 & \text{otherwise} \end{cases}$$

$$X_{ij3} = \begin{cases} 1 & \text{if case from factor level 3} \\ -\frac{4}{5} & \text{if case from factor level 4} \\ 0 & \text{otherwise} \end{cases}$$

The fitted regression function is:

$$\hat{Y} = 18.63 - 4.03X_1 - 5.23X_2 + .87X_3$$

and the following relations hold:

$$\hat{\mu}_. = b_0 = \bar{Y}_{..} = 18.63$$
$$\hat{\tau}_1 = b_1 = \bar{Y}_{1.} - \bar{Y}_{..} = 14.6 - 18.63 = -4.03$$
$$\hat{\tau}_2 = b_2 = \bar{Y}_{2.} - \bar{Y}_{..} = 13.4 - 18.63 = -5.23$$
$$\hat{\tau}_3 = b_3 = \bar{Y}_{3.} - \bar{Y}_{..} = 19.5 - 18.63 = .87$$
$$\hat{\tau}_4 = -\frac{n_1}{n_4}\hat{\tau}_1 - \frac{n_2}{n_4}\hat{\tau}_2 - \frac{n_3}{n_4}\hat{\tau}_3 = 8.56.$$

A general linear test of the alternatives:

$$H_0: \tau_1 = \tau_2 = \tau_3 = 0$$
$$H_a: \text{not all } \tau_i = 0$$

is conducted using the full model in (16.82) and forming the reduced model by setting $\tau_1 = \tau_2 = \tau_3 = 0$ in full model (16.82). The test statistic (16.78) for the presence of a regression relation again yields:

$$F^* = \frac{MSR}{MSE} = \frac{196.07}{10.55} = 18.6$$

As expected, the results are identical to those obtained earlier for the ANOVA F test.

Cell Means Model

When the analysis of variance test is to be conducted by means of the regression approach based on the cell means model (16.2):

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

the $\boldsymbol{\beta}$ vector can be defined to contain all r treatment means μ_i:

$$\boldsymbol{\beta} = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_r \end{bmatrix} \tag{16.83}$$

and r indicator variables $X_1, X_2, \ldots, X_r$ are utilized, each defined as a 0, 1 variable as illustrated in Chapter 8:

$$\begin{aligned} X_1 &= \begin{cases} 1 & \text{if case from factor level 1} \\ 0 & \text{otherwise} \end{cases} \\ &\vdots \\ X_r &= \begin{cases} 1 & \text{if case from factor level r} \\ 0 & \text{otherwise} \end{cases} \end{aligned} \tag{16.84}$$

The regression model therefore is:

$$Y_{ij} = \mu_1 X_{ij1} + \mu_2 X_{ij2} + \cdots + \mu_r X_{ijr} + \varepsilon_{ij} \qquad \text{Full model} \tag{16.85}$$

with the μ_i playing the role of regression coefficients.

The $\mathbf{X}$ matrix with this approach contains only 0 and 1 entries. For example, for $r = 3$ factor levels with $n_1 = n_2 = n_3 = 2$ cases, the $\mathbf{X}$ matrix (observations in order Y_{11}, Y_{12}, Y_{21}, etc.) and $\boldsymbol{\beta}$ vector would be as follows:

$$\mathbf{X} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \qquad \boldsymbol{\beta} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix}$$

Note that regression model (16.85) has no intercept term. When a computer regression package is to be employed for this case, it is important that a fit with no intercept term be specified.

The ANOVA table obtained with regression model (16.85) is different from the one with the single-factor ANOVA model in (16.2) because the regression model (16.85) has no intercept term. Thus, the F test obtained with the regression model cannot be used to test the equality of factor level means. The test of whether the factor level means are equal, i.e., $\mu_1 = \mu_2 = \cdots = \mu_r$, asks only whether or not the regression coefficients in (16.83) are equal, not whether or not they equal zero. Hence, we need to fit the full model and then the reduced model to conduct this test. The reduced model when H_0: $\mu_1 = \cdots = \mu_r$ holds is:

$$Y_{ij} = \mu_c + \varepsilon_{ij} \qquad \text{Reduced model} \tag{16.86}$$

where μ_c is the common value of all μ_i under H_0. The $\mathbf{X}$ matrix here consists simply of a column of 1s. The $\mathbf{X}$ matrix and $\boldsymbol{\beta}$ vector for the reduced model in our example would be:

$$\mathbf{X} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \qquad \boldsymbol{\beta} = [\mu_c]$$

After the full and reduced models are fitted and the error sums of squares are obtained for each fit, the usual general linear test statistic (2.70) is then calculated.

Example

For the Kenton Food Company example, the regression fit for the cell means model in (16.85) is:

$$\hat{Y} = 14.6X_1 + 13.4X_2 + 19.5X_3 + 27.2X_4$$

It can be readily seen that the coefficient of X_i is equal to the estimated factor level mean $\bar{Y}_{i\cdot}$ for $i = 1, \ldots, 4$.

A general linear test of the alternatives:

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$$
$$H_a: \text{not all } \mu_i \text{ are equal}$$

is conducted using the full and reduced models in (16.85) and (16.86). Here we again find that $SSE(R) = 746.42$ and that $SSE(F) = 158.2$. From (2.70) we have:

$$F^* = \frac{746.42 - 158.2}{4 - 1} \div \frac{158.2}{19 - 4} = 18.6$$

This demonstrates that the test for equality of means using the regression approach is, as expected, the same as that obtained earlier for the ANOVA F test.

16.9 Randomization Tests

Randomization can provide the basis for making inferences without requiring assumptions about the distribution of the error terms ε. Consider factor effects model (16.62) for a single-factor study:

$$Y_{ij} = \mu_{\cdot} + \tau_i + \varepsilon_{ij}$$

Rather than assume that the ε_{ij} are independent normal random variables with mean zero and constant variance σ^2, we shall now consider each ε_{ij} to be a fixed effect associated with the experimental unit. In this framework, we view the n_T experimental units to be a finite population, and associated with each unit is the unit-specific effect ε_{ij}. When randomization assigns this experimental unit to treatment i, the observed response will be $Y_{ij} = \mu_{\cdot} + \tau_i + \varepsilon_{ij}$. The response Y_{ij} is still a random variable, but under the randomization view the randomness arises because the treatment effect τ_i is the result of a random assignment of the experimental unit to treatment i.

If there are no treatment effects, that is, if all $\tau_i = 0$, then the response $Y_{ij} = \mu_{\cdot} + \varepsilon_{ij}$ depends only on the experimental unit. Since with randomization the experimental unit is equally likely to be assigned to any treatment, the observed response Y_{ij}, if there are no treatment effects, could with equal likelihood have been observed for any of the treatments. Thus, when there are no treatment effects, randomization will lead to an assignment of the finite population of n_T observations Y_{ij} to the treatments such that all treatment combinations of observations are equally likely. This, in turn, leads to an exact sampling distribution of the test statistic under H_0: $\tau_i \equiv 0$, sometimes termed the *randomization distribution* of the test statistic. Percentiles of the randomization distribution can then be used to test for the presence of factor effects. This use of the randomization distribution provides the basis of a nonparametric test for treatment effects.

To illustrate the concept of a randomization distribution, consider a single-factor experiment consisting of two treatments and two replications. In this experiment, the alternatives of interest are:

$$H_0: \tau_1 = \tau_2 = 0$$

$$H_a: \text{not both } \tau_1 \text{ and } \tau_2 \text{ equal zero}$$

Test statistic F^* in (16.55) will be used to conduct the test. The sample results are:

Treatment 1 Y_{1j}	Treatment 2 Y_{2j}
3	8
7	10

For these data, $F^* = 3.20$.

Since the treatments are assigned to experimental units at random, it would have been just as likely, if there are no treatment effects, to have observed 3 and 8 for treatment 1 and 7 and 10 for treatment 2. In that event, the test statistic would have been $F^* = 1.06$. In fact, any division of the four observations into two groups of size two is equally likely with randomization if there are no treatment effects. Because this experiment is small, we can easily list all $4!/(2!2!) = 6$ possible outcomes of the experiment, assuming no treatment effects are present:

Randomization	Treatment 1	Treatment 2	F^*	Probability
1	3, 7	8, 10	3.20	1/6
2	3, 8	7, 10	1.06	1/6
3	3, 10	8, 7	.08	1/6
4	8, 7	3, 10	.08	1/6
5	7, 10	3, 8	1.06	1/6
6	8, 10	3, 7	3.20	1/6

The last two columns give the randomization distribution of test statistic F^* under H_0. Randomization assures us that, when H_0 is true, each possible value of the test statistic has probability 1/6. From the randomization distribution, we see that the P-value for the test is the probability:

$$P\{F^* \geq 3.20\} = \frac{2}{6} = .33$$

This P-value is somewhat different than the usual (normal theory) P-value:

$$P\{F(1, 2) \geq 3.20\} = .22$$

In this instance, because the sample sizes are very small, the F distribution does not provide a particularly good approximation to the exact sampling distribution of F^* under H_0. However, both empirical and theoretical studies have shown that the F distribution is a good approximation to the exact randomization distribution when the sample sizes are not small. Thus, randomization alone can justify the F test as a good approximate test, without requiring any assumption of independent, normal error terms. We shall next demonstrate the use of the randomization test in a more realistic setting.

Comments

1. Because of the discreteness of the randomization distribution, it is conservative to define the P-value as the probability of equaling or exceeding the observed value of the test statistic when H_0 holds. For continuous sampling distributions, it does not matter whether the P-value is defined as the probability of exceeding the observed value of the test statistic or as the probability of equaling or exceeding it. For instance, $P\{F(1, 2) > 3.20\} = P\{F(1, 2) \geq 3.20\}$. When more than one treatment combination yields the value of the test statistic F^*, some authors suggest that the P-value be calculated as $P\{F > F^*\} + P\{F = F^*\}/2$. This leads to a less conservative P-value.

2. The randomization test is sometimes referred to as a *permutation* test, although permutation tests are also applied to nonrandomized studies. Because of the conservativeness of permutation (or randomization) tests for small samples, their virtues continue to be debated in the literature. See Reference 16.1. ■

Example

A manufacturer of children's plastic toys considered the introduction of statistical process control (SPC) and engineering process control (EPC) in order to reduce the volume of scrap and rework at each of its nine manufacturing plants. To assess the effects of these quality practices, a single-factor experiment was conducted for a six-month period. The treatments were:

Treatment i	Quality Practice
1	None (control group)
2	SPC
3	Both SPC and EPC

The three treatments were each randomly assigned to three of the nine available plants. The response of interest was the reduction in the defect rate at the end of the six-month trial period. The results are given in the first row (randomization 1) in Table 16.5. Management wishes to test whether or not the mean reduction in the defect rate is the same for the three

TABLE 16.5 Randomization Samples and Test Statistics—Quality Control Example.

Randomization	Treatment 1	Treatment 2	Treatment 3	F^*	Probability
1	1.1, .5, −2.1	4.2, 3.7, .8	3.2, 2.8, 6.3	4.39	1/1,680
2	1.1, .5, −2.1	4.2, 3.7, 3.2	.8, 2.8, 6.3	3.74	1/1,680
3	1.1, .5, −2.1	4.2, 3.7, 2.8	3.2, .8, 6.3	3.67	1/1,680
	...	...	...	...	...
1,680	3.2, 2.8, 6.3	4.2, 3.7, .8	1.1, .5, −2.1	4.39	1/1,680

treatments:

$$H_0\colon \tau_1 = \tau_2 = \tau_3 = 0$$
$$H_a\colon \text{not all } \tau_i \text{ equal zero}$$

The risk of a Type I error is to be controlled at $\alpha = .10$. We shall now conduct this test by obtaining the exact randomization distribution.

In this experimental study, there are $9!/(3!3!3!) = 1{,}680$ possible combinations of assigning the nine experimental units to the three treatments. A computer program was utilized to enumerate these 1,680 combinations and to calculate the F^* statistic for each. A partial listing of results is presented in Table 16.5.

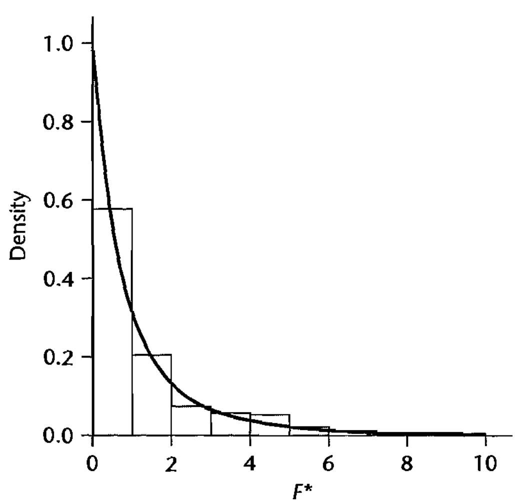
Of the 1,680 possible values of the test statistic F^*, 120 were equal to or greater than the observed value 4.39. Thus, from the randomization distribution we find:

$$P\text{-value} = P\{F^* \geq 4.39\} = \frac{120}{1{,}680} = .071$$

Since $.071 < \alpha = .10$, we conclude that the mean reduction in the defect rate is not the same for the three treatments.

Even though the sample sizes are not very large here, the exact randomization distribution is well approximated by the F distribution. Figure 16.8 shows both the randomization

FIGURE 16.8 Randomization Distribution of F^* and Corresponding F Distribution—Quality Control Example.



distribution in the form of a histogram and the density function for the corresponding F distribution, $F(2, 6)$. Note how well the F distribution approximates the randomization distribution. The P-value according to the F distribution is $P\{F(2, 6) \geq 4.39\} = .067$. This is very close to the randomization P-value of .071.

16.10 Planning of Sample Sizes with Power Approach

For analysis of variance studies, as for other statistical studies, it is important to plan the sample sizes so that needed protection against both Type I and Type II errors can be obtained, or so that the estimates of interest have sufficient precision to be useful. This planning is necessary for both observational and experimental studies to ensure that the sample sizes are large enough to detect important differences with high probability. At the same time, the sample sizes should not be so large that the cost of the study becomes excessive and that unimportant differences become statistically significant with high probability. Planning of sample sizes is therefore an integral part of the design of a study.

We shall generally assume in our discussion of planning sample sizes that all treatments are to have equal sample sizes, reflecting that they are about equally important. Indeed, when major interest lies in pairwise comparisons of all treatment means, it can be shown that equal sample sizes maximize the precision of the comparisons. Another reason for equal sample sizes is that certain departures from the assumed ANOVA model are less troublesome if all factor levels have the same sample size, as noted earlier.

There will be times, however, when unequal sample sizes are appropriate. For instance, when four experimental treatments are each to be compared to a control, it may be reasonable to make the sample size for the control larger. We shall comment later on the planning of sample sizes for such a case.

Planning of sample sizes can be approached in terms of (1) controlling the risks of making Type I and Type II errors, (2) controlling the widths of desired confidence intervals, or (3) a combination of these two. The procedures for planning sample sizes that we shall discuss here are applicable to both observational studies and to experimental studies based on a completely randomized single-factor design. In later chapters, we shall consider the planning of sample sizes for other study designs. In this section, we consider planning of sample sizes with the power approach, which permits controlling the risks of making Type I and Type II errors. In Section 16.11 we discuss planning of sample sizes when the best treatment is to be identified. Later, in Section 17.8, we take up planning of sample sizes to control the precision of estimates of important effects. We shall consider planning of sample sizes for multifactor studies in Section 24.7.

Before we can discuss planning of sample sizes with the power approach, we need to consider the power of the F test.

Power of F Test

By the power of the F test for a single-factor study, we refer to the probability that the decision rule will lead to conclusion H_a, that the treatment means differ, when in fact H_a holds. Specifically, the power is given by the following expression for the cell means model (16.2):

$$\text{Power} = P\{F^* > F(1 - \alpha; r - 1, n_T - r) | \phi\} \tag{16.87}$$

where ϕ is the *noncentrality parameter*, that is, a measure of how unequal the treatment means μ_i are:

$$\phi = \frac{1}{\sigma}\sqrt{\frac{\sum n_i(\mu_i - \mu.)^2}{r}} \tag{16.87a}$$

and:

$$\mu. = \frac{\sum n_i \mu_i}{n_T} \tag{16.87b}$$

When all factor level samples are of equal size n, the parameter ϕ becomes:

$$\phi = \frac{1}{\sigma}\sqrt{\frac{n}{r}\sum(\mu_i - \mu.)^2} \qquad \text{when } n_i \equiv n \tag{16.88}$$

where:

$$\mu. = \frac{\sum \mu_i}{r} \tag{16.88a}$$

Power probabilities are determined by utilizing the noncentral F distribution since this is the sampling distribution of F^* when H_a holds. The resulting calculations are quite complex. We present a series of tables in Appendix Table B.11 that can be used readily to look up power probabilities directly. The proper table to use depends on the number of factor levels and the level of significance employed in the decision rule. Specifically, Table B.11 is used as follows:

1. Each page refers to a different ν_1, the number of degrees of freedom for the numerator of F^*. For ANOVA model (16.2), $\nu_1 = r - 1$, or the number of factor levels minus one. Table B.11 contains power tables for $\nu_1 = 2, 3, 4, 5$, and 6, as shown at the top of each page.

2. Two levels of significance, denoted by α, are presented in Table B.11, namely, $\alpha = .05$ and $\alpha = .01$. The upper table on each page refers to $\alpha = .05$ and the lower table to $\alpha = .01$.

3. Within each table, the rows refer to different values of ν_2, the degrees of freedom for the denominator of F^*. The columns refer to different values of ϕ, the noncentrality parameter defined in (16.87a). For ANOVA model (16.2), $\nu_2 = n_T - r$.

Examples

1. Consider the case where $\nu_1 = 2$, $\nu_2 = 10$, $\phi = 3$, and $\alpha = .05$. We then find from Table B.11 (p. 1337) that the power is $1 - \beta = .98$.

2. Suppose that for the Kenton Food Company example, the analyst wishes to determine the power of the decision rule in the example on page 699 when there are substantial differences between the factor level means. Specifically, the analyst wishes to consider the case when $\mu_1 = 12.5$, $\mu_2 = 13$, $\mu_3 = 18$, and $\mu_4 = 21$. The weighted mean in (16.87b) therefore is:

$$\mu. = \frac{5(12.5) + 5(13) + 4(18) + 5(21)}{19} = 16.03$$

Thus, the specified value of ϕ is:

$$\phi = \frac{1}{\sigma}\left[\frac{5(-3.53)^2 + 5(-3.03)^2 + 4(1.97)^2 + 5(4.97)^2}{4}\right]^{1/2}$$

$$= \frac{1}{\sigma}(7.86)$$

Note that we still need to know σ, the standard deviation of the error terms ε_{ij} in the model. Suppose that from past experience it is known that $\sigma = 3.5$ cases approximately. Then we have:

$$\phi = \frac{1}{3.5}(7.86) = 2.25$$

Further, we have for this example:

$$\nu_1 = r - 1 = 3 \qquad \nu_2 = n_T - r = 15 \qquad \alpha = .05$$

Table B.11 on page 1338 indicates that the power is $1 - \beta = .91$. In other words, there are 91 chances in 100 that the decision rule, based on the sample sizes employed, will lead to the detection of differences in the mean sales volumes for the four package designs when the differences are the ones specified earlier.

Comments

1. Any given value of ϕ encompasses many different combinations of factor level means μ_i. Thus, in the Kenton Food Company example, the means $\mu_1 = 12.5$, $\mu_2 = 13$, $\mu_3 = 18$, $\mu_4 = 21$ and the means $\mu_1 = 21$, $\mu_2 = 12.5$, $\mu_3 = 18$, $\mu_4 = 13$ lead to the same value of $\phi = 2.25$ and hence to the same power.

2. The larger ϕ—that is, the larger the differences between the factor level means—the greater the power and hence the smaller the probability of making a Type II error for a given risk α of making a Type I error. Also, the smaller the specified α risk, the smaller is the power for any given ϕ, and hence the larger the risk of a Type II error.

3. Since many single-factor studies are undertaken because of the expectation that the factor level means differ and it is desired to investigate these differences, the α risk used in constructing the decision rule for determining whether or not the factor level means are equal is often set relatively high (e.g., .05 or .10 instead of .01) so as to increase the power of the test.

4. The power table for $\nu_1 = 1$ is not reproduced in Table B.11 since this case corresponds to the comparison of two population means. As noted previously, the F test is the equivalent of the two-sided t test for this case, and the power tables for the two-sided t test presented in Table B.5 can then be used, with noncentrality parameter:

$$\delta = \frac{|\mu_1 - \mu_2|}{\sigma\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} \tag{16.89}$$

and degrees of freedom $n_1 + n_2 - 2$. ■

Use of Table B.12 for Single-Factor Studies

The power approach in planning sample sizes can be implemented by use of the power tables for F tests presented in Table B.11. A trial-and-error process is required, however, with these tables. Instead, we shall use other tables that furnish the appropriate sample sizes directly. Table B.12 presents sample size determinations that are applicable when all treatments are to have equal sample sizes and all effects are fixed.

The planning of sample sizes for single-factor studies with fixed factor levels using Table B.12 is done in terms of the noncentrality parameter (16.88) for equal sample sizes. However, instead of requiring a direct specification of the levels of μ_i for which it is important to control the risk of making a Type II error, Table B.12 only requires a specification of the minimum range of factor level means for which it is important to detect differences between the μ_i with high probability. This minimum range is denoted by Δ:

$$\Delta = \max(\mu_i) - \min(\mu_i) \tag{16.90}$$

The following three specifications need to be made in using Table B.12:

1. The level α at which the risk of making a Type I error is to be controlled.

2. The magnitude of the minimum range Δ of the μ_i which is important to detect with high probability. The magnitude of σ, the standard deviation of the probability distributions of Y, must also be specified since entry into Table B.12 is in terms of the ratio:

$$\frac{\Delta}{\sigma} \tag{16.91}$$

3. The level β at which the risk of making a Type II error is to be controlled for the specification given in 2. Entry into Table B.12 is in terms of the power $1 - \beta$.

When using Table B.12, four α levels are available at which the risk of making a Type I error can be controlled ($\alpha = .2, .1, .05, .01$). The Type II error risk can be controlled at one of four β levels ($\beta = .3, .2, .1, .05$) through the specification of the power $1 - \beta$. Table B.12 provides necessary sample sizes for studies consisting of $r = 2, \ldots, 10$ factor levels or treatments.

Example

A company owning a large fleet of trucks wishes to determine whether or not four different brands of snow tires have the same mean tread life (in thousands of miles). It is important to conclude that the four brands of snow tires have different mean tread lives when the difference between the means of the best and worst brands is 3 (thousand miles) or more. Thus, the minimum range specification is $\Delta = 3$. It is known from past experience that the standard deviation of the tread lives of these tires is $\sigma = 2$ (thousand miles), approximately. Management would like to control the risks of making incorrect decisions at the following levels:

$$\alpha = .05$$

$$\beta = .10 \qquad \text{or} \qquad \text{Power} = 1 - \beta = .90$$

Entering Table B.12 for $\Delta/\sigma = 3/2 = 1.5$, $\alpha = .05$, $1 - \beta = .90$, and $r = 4$, we find $n = 14$. Hence, 14 snow tires of each brand need to be tested in order to control the risks of making incorrect decisions at the desired levels.

Specification of Δ/σ Directly. Table B.12 can also be used when the minimum range is specified directly in units of the standard deviation σ. Let the specification of Δ in this case be $k\sigma$ so that we have by (16.91):

$$\frac{\Delta}{\sigma} = \frac{k\sigma}{\sigma} = k$$

Hence, Table B.12 is entered directly for the specified value k with this approach.

Example

Suppose it is specified in the snow tires example that it is important to detect differences between the mean tread lives if the range of the mean tread lives is $k = 2$ standard deviations or more. Suppose also that the other specifications are:

$$\alpha = .10$$
$$\beta = .05 \qquad \text{or} \qquad \text{Power} = 1 - \beta = .95$$

From Table B.12, we find for $k = 2$ and $r = 4$ that $n = 9$ tires will need to be tested for each brand in order that the specified risk protection will be achieved.

Comment

While specifying Δ/σ directly does not require an advance planning value of the standard deviation σ, this is not of as much advantage as it might seem because a meaningful specification of Δ in units of σ will frequently require knowledge of the approximate magnitude of the standard deviation. ∎

Some Further Observations on Use of Table B.12

1. The exact specification of Δ/σ has great effect on the sample sizes n when Δ/σ is small, but it has much less effect when Δ/σ is large. For instance, when $r = 3$, $\alpha = .05$, and $\beta = .10$, we have from Table B.12:

Δ/σ	n
1.0	27
1.5	13
2.0	8
2.5	6

Thus, unless Δ/σ is quite small, one need not be too concerned about some imprecision in specifying Δ/σ.

2. Reducing either the specified α or β risks or both increases the required sample sizes. For instance, when $r = 4$, $\alpha = .10$, and $\Delta/\sigma = 1.25$, we have:

β	$1-\beta$	n
.20	.80	13
.10	.90	16
.05	.95	20

3. A moderate error in the advance planning value of σ can cause a substantial miscalculation of required sample sizes. For instance, when $r = 5$, $\alpha = .05$, $\beta = .10$, and $\Delta = 3$, we have:

σ	Δ/σ	n
1	3.0	5
2	1.5	15
3	1.0	32

In view of the usual approximate nature of the advance planning value of σ, it is generally desirable to investigate the needed sample sizes for a range of likely values of σ before deciding on the sample sizes to be employed.

4. Table B.12 is based on the noncentrality parameter ϕ in (16.88) even though no specification is made of the individual factor level means μ_i for which it is important to conclude that the factor level means differ. To see how Table B.12 utilizes the noncentrality parameter ϕ, consider again the snow tires example where $r = 4$ brands are to be tested and a minimum range of $\Delta = 3$ (thousand miles) of the four mean tread lives μ_i is to be detected with high probability. The following are some possible sets of values of the μ_i, each of which has range $\Delta = 3$:

Case	μ_1	μ_2	μ_3	μ_4	$\sum(\mu_i - \mu.)^2$
1	24	27	25	26	5.00
2	25	25	26	23	4.75
3	25	25	25	28	6.75
4	25	25	26.5	23.5	4.50

The term $\sum(\mu_i - \mu.)^2$ of the noncentrality parameter ϕ in (16.88) differs for each of these four possibilities and hence the power differs, even though the range is the same in all cases. Note that the term $\sum(\mu_i - \mu.)^2$ is the smallest for case 4, where two factor level means are at $\mu.$ and the other two are equally spaced around $\mu.$. It can be shown that for a given range Δ, the term $\sum(\mu_i - \mu.)^2$ is minimized when all but two factor level means are at $\mu.$ and the two remaining factor level means are equally spaced around $\mu.$. Thus, we have:

$$\min \sum_{i=1}^{r}(\mu_i - \mu.)^2 = \left(\frac{\Delta}{2}\right)^2 + \left(-\frac{\Delta}{2}\right)^2 + 0 + \cdots + 0 = \frac{\Delta^2}{2} \qquad \textbf{(16.92)}$$

Since the power of the test varies directly with $\sum(\mu_i - \mu.)^2$, use of (16.92) in calculating Table B.12 ensures that the power is at least $1 - \beta$ for any combination of μ_i values with range Δ.

16.11 Planning of Sample Sizes to Find "Best" Treatment

There are occasions when the chief purpose of the study is to ascertain the treatment with the highest or lowest mean. In the snow tires example, for instance, it may be desired to determine which of the four brands has the longest mean tread life.

Table B.13, developed by Bechhofer, enables us to determine the necessary sample sizes so that with probability $1-\alpha$ the highest (lowest) estimated treatment mean is from the treatment with the highest (lowest) population mean. We need to specify the probability $1-\alpha$, the standard deviation σ, and the smallest difference λ between the highest (lowest) and second highest (second lowest) treatment means that it is important to recognize. Table B.13 assumes that equal sample sizes are to be used for all r treatments.

Example

Suppose that in the snow tires example, the chief objective is to identify the brand with the longest mean tread life. There are $r = 4$ brands. We anticipate, as before, that $\sigma = 2$ (thousand miles). Further, we are informed that a difference $\lambda = 1$ (thousand miles) between the highest and second highest brand means is important to recognize, and that the probability is to be $1 - \alpha = .90$ or greater that we identify correctly the brand with the highest mean tread life when $\lambda \geq 1$.

The entry in Table B.13 is $\lambda\sqrt{n}/\sigma$. For $r = 4$ and probability $1 - \alpha = .90$, we find from Table B.13 that $\lambda\sqrt{n}/\sigma = 2.4516$. Hence, since the λ specification is $\lambda = 1$, we obtain:

$$\frac{(1)\sqrt{n}}{2} = 2.4516$$

$$\sqrt{n} = 4.9032 \qquad \text{or} \qquad n = 25$$

Thus, when the mean tread life for the best brand exceeds that of the second best by at least 1 (thousand miles) and when $\sigma = 2$ (thousand miles), sample sizes of 25 tires for each brand provide an assurance of at least .90 that the brand with the highest estimated mean $\bar{Y}_{i\cdot}$ is the brand with the highest population mean.

Comment

If the planning value for the standard deviation is not accurate, the probability of identifying the population with the highest (lowest) mean correctly is, of course, affected. This is no different from the other approaches, where a misjudgment of the standard deviation affects the risks of making a Type II error. ■

Cited Reference

16.1. Berger, V. W. "Pros and Cons of Permutation Tests in Clinical Trials," *Statistics in Medicine* 19 (2000), pp. 1319–1328.

Problems

16.1. Refer to Figure 16.1a. Could you determine the mean sales level when the price level is \$68 if you knew the true regression function? Could you make this determination from Figure 16.1b if you only knew the values of the parameters μ_1, μ_2, and μ_3 of ANOVA model (16.2)? What distinction between regression models and ANOVA models is demonstrated by your answers?

16.2. A market researcher, having collected data on breakfast cereal expenditures by families with 1, 2, 3, 4, and 5 children living at home, plans to use an ordinary regression model to estimate the mean expenditures at each of these five family size levels. However, the researcher is undecided between fitting a linear or a quadratic regression model, and the data do not give clear evidence in favor of one model or the other. A colleague suggests: "For your purposes you might simply use an ANOVA model." Is this a useful suggestion? Explain.

16.3. In a study of intentions to get flu-vaccine shots in an area threatened by an epidemic, 90 persons were classified into three groups of 30 according to the degree of risk of getting flu. Each group was together when the persons were asked about the likelihood of getting the shots, on a probability scale ranging from 0 to 1.0. Unavoidably, most persons overheard the answers of nearby respondents. An analyst wishes to test whether the mean intent scores are the same for the three risk groups. Consider each assumption for ANOVA model (16.2) and explain whether this assumption is likely to hold in the present situation.

16.4. A company, studying the relation between job satisfaction and length of service of employees, classified employees into three length-of-service groups (less than 5 years, 5–10 years, more than 10 years). Suppose $\mu_1 = 65$, $\mu_2 = 80$, $\mu_3 = 95$, and $\sigma = 3$, and that ANOVA model (16.2) is applicable.

a. Draw a representation of this model in the format of Figure 16.2.

b. Find $E\{MSTR\}$ and $E\{MSE\}$ if 25 employees from each group are selected at random for intensive interviewing about job satisfaction. Is $E\{MSTR\}$ substantially larger than $E\{MSE\}$ here? What is the implication of this?

16.5. In a study of length of hospital stay (in number of days) of persons in four income groups, the parameters are as follows: $\mu_1 = 5.1$, $\mu_2 = 6.3$, $\mu_3 = 7.9$, $\mu_4 = 9.5$, $\sigma = 2.8$. Assume that ANOVA model (16.2) is appropriate.

a. Draw a representation of this model in the format of Figure 16.2.

b. Suppose 100 persons from each income group are randomly selected for the study. Find $E\{MSTR\}$ and $E\{MSE\}$. Is $E\{MSTR\}$ substantially larger than $E\{MSE\}$ here? What is the implication of this?

c. If $\mu_2 = 5.6$ and $\mu_3 = 9.0$, everything else remaining the same, what would $E\{MSTR\}$ be? Why is $E\{MSTR\}$ substantially larger here than in part (b) even though the range of the factor level means is the same?

16.6. A student asks: "Why is the F test for equality of factor level means not a two-tail test since any differences among the factor level means can occur in either direction?" Explain, utilizing the expressions for the expected mean squares in (16.37).

*16.7. **Productivity improvement.** An economist compiled data on productivity improvements last year for a sample of firms producing electronic computing equipment. The firms were classified according to the level of their average expenditures for research and development in the past three years (low, moderate, high). The results of the study follow (productivity improvement is measured on a scale from 0 to 100). Assume that ANOVA model (16.2) is appropriate.

		j											
	i	1	2	3	4	5	6	7	8	9	10	11	12
1	Low	7.6	8.2	6.8	5.8	6.9	6.6	6.3	7.7	6.0			
2	Moderate	6.7	8.1	9.4	8.6	7.8	7.7	8.9	7.9	8.3	8.7	7.1	8.4
3	High	8.5	9.7	10.1	7.8	9.6	9.5						

a. Prepare aligned dot plots of the data. Do the factor level means appear to differ? Does the variability of the observations within each factor level appear to be approximately the same for all factor levels?

b. Obtain the fitted values.

c. Obtain the residuals. Do they sum to zero in accord with (16.21)?

d. Obtain the analysis of variance table.

e. Test whether or not the mean productivity improvement differs according to the level of research and development expenditures. Control the α risk at .05. State the alternatives, decision rule, and conclusion.

f. What is the P-value of the test in part (e)? How does it support the conclusion reached in part (e)?

g. What appears to be the nature of the relationship between research and development expenditures and productivity improvement?

16.8. **Questionnaire color.** In an experiment to investigate the effect of color of paper (blue, green, orange) on response rates for questionnaires distributed by the "windshield method" in supermarket parking lots, 15 representative supermarket parking lots were chosen in a metropolitan area and each color was assigned at random to five of the lots. The response rates (in percent) follow. Assume that ANOVA model (16.2) is appropriate.

	i	j=1	2	3	4	5
1	Blue	28	26	31	27	35
2	Green	34	29	25	31	29
3	Orange	31	25	27	29	28

a. Prepare aligned dot plots of the data. Do the factor level means appear to differ? Does the variability of the observations within each factor level appear to be approximately the same for all factor levels?

b. Obtain the fitted values.

c. Obtain the residuals.

d. Obtain the analysis of variance table.

e. Conduct a test to determine whether or not the mean response rates for the three colors differ. Use level of significance $\alpha = .10$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

f. When informed of the findings, an executive said: "See? I was right all along. We might as well print the questionnaires on plain white paper, which is cheaper." Does this conclusion follow from the findings of the study? Discuss.

16.9. **Rehabilitation therapy.** A rehabilitation center researcher was interested in examining the relationship between physical fitness prior to surgery of persons undergoing corrective knee surgery and time required in physical therapy until successful rehabilitation. Patient records in the rehabilitation center were examined, and 24 male subjects ranging in age from 18 to 30 years who had undergone similar corrective knee surgery during the past year were selected for the study. The number of days required for successful completion of physical therapy and the prior physical fitness status (below average, average, above average) for each patient follow.

	i	j=1	2	3	4	5	6	7	8	9	10
1	Below Average	29	42	38	40	43	40	30	42		
2	Average	30	35	39	28	31	31	29	35	29	33
3	Above Average	26	32	21	20	23	22				

Assume that ANOVA model (16.2) is appropriate.

a. Prepare aligned dot plots of the data. Do the factor level means appear to differ? Does the variability of the observations within each factor level appear to be approximately the same for all factor levels?

b. Obtain the fitted values.

c. Obtain the residuals. Do they sum to zero in accord with (16.21)?

d. Obtain the analysis of variance table.

e. Test whether or not the mean number of days required for successful rehabilitation is the same for the three fitness groups. Control the α risk at .01. State the alternatives, decision rule, and conclusion.

f. Obtain the P-value for the test in part (e). Explain how the same conclusion reached in part (e) can be obtained by knowing the P-value.

g. What appears to be the nature of the relationship between physical fitness status and duration of required physical therapy?

*16.10. **Cash offers.** A consumer organization studied the effect of age of automobile owner on size of cash offer for a used car by utilizing 12 persons in each of three age groups (young, middle, elderly) who acted as the owner of a used car. A medium price, six-year-old car was selected for the experiment, and the "owners" solicited cash offers for this car from 36 dealers selected at random from the dealers in the region. Randomization was used in assigning the dealers to the "owners." The offers (in hundred dollars) follow. Assume that ANOVA model (16.2) is applicable.

		j											
	i	1	2	3	4	5	6	7	8	9	10	11	12
1	Young	23	25	21	22	21	22	20	23	19	22	19	21
2	Middle	28	27	27	29	26	29	27	30	28	27	26	29
3	Elderly	23	20	25	21	22	23	21	20	19	20	22	21

a. Prepare aligned dot plots of the data. Do the factor level means appear to differ? Does the variability of the observations within each factor level appear to be approximately the same for all factor levels?

b. Obtain the fitted values.

c. Obtain the residuals.

d. Obtain the analysis of variance table.

e. Conduct the F test for equality of factor level means; use $\alpha = .01$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

f. What appears to be the nature of the relationship between age of owner and mean cash offer?

*16.11. **Filling machines.** A company uses six filling machines of the same make and model to place detergent into cartons that show a label weight of 32 ounces. The production manager has complained that the six machines do not place the same amount of fill into the cartons. A consultant requested that 20 filled cartons be selected randomly from each of the six machines and the content of each carton carefully weighed. The observations (stated for convenience as deviations from 32.00 ounces) follow. Assume that ANOVA model (16.2) is applicable.

	j						
i	1	2	3	...	18	19	20
1	−.14	.20	.07	...	.07	−.01	−.19
2	.46	.11	.12	...	.02	.11	.12
3	.21	.78	.32	...	.50	.20	.61
4	.49	.58	.52	...	.42	.45	.20
5	−.19	.27	.06	...	.14	.35	−.18
6	.05	−.05	.28	...	.35	−.09	.05

a. Prepare aligned box plots of the data. Do the factor level means appear to differ? Does the variability of the observations within each factor level appear to be approximately the same for all factor levels?

b. Obtain the fitted values.

c. Obtain the residuals. Do they sum to zero in accord with (16.21)?

d. Obtain the analysis of variance table.

e. Test whether or not the mean fill differs among the six machines; control the α risk at .05. State the alternatives, decision rule, and conclusion. Does your conclusion support the production manager's complaint?

f. What is the P-value of the test in part (e)? Is this value consistent with your conclusion in part (e)? Explain.

g. Based on the box plots obtained in part (a), does the variation between the mean fills for the six machines appear to be large relative to the variability in fills between cartons for any given machine? Explain.

16.12. **Premium distribution.** A soft-drink manufacturer uses five agents (1, 2, 3, 4, 5) to handle premium distributions for its various products. The marketing director desired to study the timeliness with which the premiums are distributed. Twenty transactions for each agent were selected at random, and the time lapse (in days) for handling each transaction was determined. The results follow. Assume that ANOVA model (16.2) is appropriate.

	j						
i	1	2	3	...	18	19	20
1	24	24	29		27	26	25
2	18	20	20	..	26	22	21
3	10	11	8	...	9	11	12
4	15	13	18	...	17	14	16
5	33	22	28	...	26	30	29

a. Prepare aligned box plots of the data. Do the factor level means appear to differ? Does the variability of the observations within each factor level appear to be approximately the same for all factor levels?

b. Obtain the fitted values.

c. Obtain the residuals. Do they sum to zero in accord with (16.21)?

d. Obtain the analysis of variance table.

e. Test whether or not the mean time lapse differs for the five agents; use $\alpha = .10$. State the alternatives, decision rule, and conclusion.

f. What is the P-value of the test in part (e)? Explain how the same conclusion as in part (e) can be reached by knowing the P-value.

g. Based on the box plots obtained in part (a), does there appear to be much variation in the mean time lapse for the five agents? Is this variation necessarily the result of differences in the efficiency of operations of the five agents? Discuss.

16.13. Refer to **Questionnaire color** Problem 16.8. Explain how you would make the random assignments of supermarket parking lots to colors in this single-factor study. Make all appropriate randomizations.

16.14. Refer to **Cash offers** Problem 16.10. Explain how you would make the random assignments of dealers to "owners" in this single-factor study. Make all appropriate randomizations.

16.15. Refer to Problem 16.4. What are the values of τ_1, τ_2, and τ_3 if the ANOVA model is expressed in the factor effects formulation (16.62), and $\mu.$ is defined by (16.63)?

16.16. Refer to Problem 16.5. What are the values of τ_i if the ANOVA model is expressed in the factor effects formulation (16.62), and $\mu.$ is defined by (16.63)?

16.17. Refer to **Premium distribution** Problem 16.12. Suppose that 25 percent of all premium distributions 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.

a. Obtain a point estimate of $\mu.$ when the ANOVA model is expressed in the factor effects formulation (16.62) and $\mu.$ is defined by (16.65), with the weights being the proportions of premium distribution handled by each agent.

b. State the alternatives for the test of equality of factor level means in terms of factor effects model (16.62) for the present case. Would this statement be affected if $\mu.$ were defined according to (16.63)? Explain.

*16.18. Refer to **Productivity improvement** Problem 16.7. Regression model (16.75) is to be employed for testing the equality of the factor level means.

a. Set up the **Y**, **X**, and $\boldsymbol{\beta}$ matrices.

b. Obtain $\mathbf{X}\boldsymbol{\beta}$. Develop equivalent expressions of the elements of this vector in terms of the cell means μ_i.

c. Obtain the fitted regression function. What is estimated by the intercept term?

d. Obtain the regression analysis of variance table.

e. Conduct the test for equality of factor level means; use $\alpha = .05$. State the alternatives, decision rule, and conclusion.

16.19. Refer to **Questionnaire color** Problem 16.8. Regression model (16.75) is to be employed for testing the equality of the factor level means.

a. Set up the **Y**, **X**, and $\boldsymbol{\beta}$ matrices.

b. Obtain $\mathbf{X}\boldsymbol{\beta}$. Develop equivalent expressions of the elements of this vector in terms of the cell means μ_i.

c. Obtain the fitted regression function. What is estimated by the intercept term?

d. Obtain the regression analysis of variance table.

e. Conduct the test for equality of factor level means; use $\alpha = .10$. State the alternatives, decision rule, and conclusion.

16.20. Refer to **Rehabilitation therapy** Problem 16.9. Regression model (16.81) is to be employed for testing the equality of the factor level means.

a. Set up the **Y**, **X**, and $\boldsymbol{\beta}$ matrices.

b. Obtain $\mathbf{X}\boldsymbol{\beta}$. Develop equivalent expressions of the elements of this vector in terms of the cell means μ_i.

c. Obtain the fitted regression function. What is estimated by the intercept term?

d. Obtain the regression analysis of variance table.

e. Conduct the test for equality of factor level means; use $\alpha = .01$. State the alternatives, decision rule, and conclusion.

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

a. Fit regression model (16.75) to the data. What is estimated by the intercept term?

b. Obtain the regression analysis of variance table and test whether or not the factor level means are equal; use $\alpha = .01$. State the alternatives, decision rule, and conclusion.

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

a. Fit the full regression model (16.85) to the data. Why would a fitted regression model containing an intercept term not be proper here?

b. Fit the reduced model (16.86) to the data.

c. Use test statistic (2.70) for testing the equality of the factor level means; employ level of significance $\alpha = .01$.

16.23. Refer to Example 1 on page 717. Find the power of the test if $\alpha = .01$, everything else remaining unchanged. How does this power compare with that in Example 1?

16.24. Refer to Example 2 on page 717. The analyst is also interested in the power of the test when $\mu_1 = \mu_2 = 13$ and $\mu_3 = \mu_4 = 18$. Assume that $\sigma = 3.5$.

a. Obtain the power of the test if $\alpha = .05$.

b. What would be the power of the test if $\alpha = .01$?

*16.25. Refer to **Productivity improvement** Problem 16.7. Obtain the power of the test in Problem 16.7e if $\mu_1 = 7.0$, $\mu_2 = 8.0$, and $\mu_3 = 9.0$. Assume that $\sigma = .9$.

16.26. Refer to **Rehabilitation therapy** Problem 16.9. Obtain the power of the test in Problem 16.9e if $\mu_1 = 37$, $\mu_2 = 35$, and $\mu_3 = 28$. Assume that $\sigma = 4.5$.

*16.27. Refer to **Cash offers** Problem 16.10. Obtain the power of the test in Problem 16.10e if the mean cash offers are $\mu_1 = 22$, $\mu_2 = 28$, and $\mu_3 = 22$. Assume that $\sigma = 1.6$.

16.28. Why do you think that the approach to planning sample sizes to find the best treatment by means of Table B.13 does not consider the risk of an incorrect identification when the best two treatment means are the same or practically the same?

*16.29. Consider a single-factor study where $r = 5$, $\alpha = .01$, $\beta = .05$, and $\sigma = 10$, and equal treatment sample sizes are desired by means of the approach in Table B.12.

a. What are the required sample sizes if $\Delta = 10, 15, 20, 30$? What generalization is suggested by your results?

b. What are the required sample sizes for the same values of Δ as in part (a) if $\alpha = .05$, all other specifications remaining the same? How do these sample sizes compare with those in part (a)?

16.30. Consider a single-factor study where $r = 6$, $\alpha = .05$, $\beta = .10$, and $\Delta = 50$, and equal treatment sample sizes are desired by means of the approach in Table B.12.

a. What are the required sample sizes if $\sigma = 50, 25, 20$? What generalization is suggested by your results?

b. What are the required sample sizes for the same values of σ as in part (a) if $r = 4$, all other specifications remaining the same? How do these sample sizes compare with those in part (a)?

16.31. Consider a single-factor study where $r = 5$, $1 - \alpha = .95$, and $\sigma = 20$, and equal sample sizes are desired by means of the approach in Table B.13.

a. What are the required sample sizes if $\lambda = 20, 10, 5$? What generalization is suggested by your results?

b. What are the required sample sizes for the same values of λ as in part (a) if $\sigma = 30$, all other specifications remaining the same? How do these sample sizes compare with those in part (a)?

16.32. Refer to **Questionnaire color** Problem 16.8. Suppose that the sample sizes have not yet been determined but it has been decided to sample the same number of supermarket parking lots for each questionnaire color. A reasonable planning value for the error standard deviation is $\sigma = 3.0$.

a. What would be the required sample sizes if: (1) differences in the response rates are to be detected with probability .90 or more when the range of the treatment means is 4.5, and (2) the α risk is to be controlled at .05?

b. If the sample sizes determined in part (a) were employed, what would be the minimum power of the test for treatment mean differences (using $\alpha = .05$) when the range of the treatment means is 6.0?

c. Suppose the chief objective is to identify the color with the highest mean response rate. The probability should be at least .99 that the best color is recognized correctly when the difference between the response rates for the best and second best colors is 1.5 percent points or more. What are the required sample sizes?

16.33. Refer to **Rehabilitation therapy** Problem 16.9. Suppose that the sample sizes have not yet been determined but it has been decided to use the same number of patients for each physical fitness group. Assume that a reasonable planning value for the error standard deviation is $\sigma = 4.5$ days.

a. What would be the required sample sizes if: (1) differences in the mean times for the three physical fitness categories are to be detected with probability .80 or more when the range of the treatment means is 5.63 days, and (2) the α risk is to be controlled at .01?

b. If the sample sizes determined in part (a) were employed, what would be the power of the test for treatment mean differences when $\mu_1 = 37$, $\mu_2 = 32$, and $\mu_3 = 28$?

c. Suppose the chief objective is to identify the physical fitness group with the smallest mean required time for therapy. The probability should be at least .90 that the correct group is identified when the mean required time for the second best group differs by 2.0 days or more. What are the required sample sizes?

*16.34. Refer to **Filling machines** Problem 16.11. Suppose that the sample sizes have not yet been determined but it has been decided to sample the same number of cartons for each filling machine. Assume that a reasonable planning value for the error standard deviation is $\sigma = .15$ ounce.

a. What would be the required sample sizes if: (1) differences in the mean amount of fill for the six filling machines are to be detected with probability .70 or more when the range of the treatment means is .15 ounce, and (2) the α risk is to be controlled at .05?

b. For the sample sizes determined in part (a), what would be the power of the test if $\mu_1 = .09$, $\mu_2 = .18$, $\mu_3 = .30$, $\mu_4 = .20$, $\mu_5 = .10$, and $\mu_6 = .20$?

c. Suppose the chief objective is to identify the filling machine with the smallest mean fill. The probability should be at least .95 that the filling machine with the smallest mean fill is recognized correctly when the filling machine with the next smallest mean fill differs by .10 ounce or more. What are the required sample sizes?

16.35. Refer to **Premium distribution** Problem 16.12. Suppose that the sample sizes have not yet been determined but it has been decided to sample the same number of premium distributions for each agent. Assume that a reasonable planning value for the error standard deviation is $\sigma = 3.0$ days.

a. What would be the required sample sizes if: (1) differences in the mean time lapse for the five agents are to be detected with probability .95 or more when the range of the treatment means is 3.75 days, and (2) the α risk is to be controlled at .10?

b. Suppose the chief objective is to identify the best agent, i.e., the one with the smallest mean time lapse. The probability should be at least .90 that the best agent is recognized correctly when the mean time lapse for the second best agent differs by 1.0 day or more. What are the required sample sizes?

Exercises

16.36. (Calculus needed.) State the likelihood function for ANOVA model (16.2) when $r = 3$ and $n_i \equiv 2$ and obtain the maximum likelihood estimators.

16.37. Show that when test statistic t^* in Table A.2a is squared, it is equivalent to the F^* test statistic (16.55) for $r = 2$.

16.38. Derive the restriction in (16.66) when the constant $\mu.$ is defined according to (16.65).

16.39. a. Obtain the least squares estimators of the regression coefficients in full regression model (16.85). What is $SSE(F)$ here?

b. Obtain the least squares estimator of $\mu.$ in reduced regression model (16.86). What is $SSE(R)$ here?

16.40. A completely randomized experiment is to be conducted involving $r = 3$ treatments, with $n = 2$ experimental trials for each treatment. Because the normality of the error terms is strongly in doubt, the test for treatment effects based on the F^* test statistic in (16.55) is to be carried out by means of the randomization distribution.

a. Determine the number of ways that the six experimental units can be divided into three groups of size two. How many unique F^* statistics are possible?

b. Using the results in part (a), what is the smallest P-value that is possible with the randomization test? What does this suggest about the adequacy of the planned sample size?

16.41. (Calculus needed.) Given $\mu_1 = 0$, $\mu_3 = 1$, and $0 \leq \mu_2 \leq 1$, show that $\sum(\mu_i - \mu.)^2$ is minimized when $\mu_2 = .5$, where $\mu. = (\mu_1 + \mu_2 + \mu_3)/3$.

Projects

16.42. Refer to the **SENIC** data set in Appendix C.1. Test whether or not the mean infection risk (variable 4) is the same in the four geographic regions (variable 9); use $\alpha = .05$. Assume that ANOVA model (16.2) is applicable. State the alternatives, decision rule, and conclusion.

16.43. Refer to the **SENIC** data set in Appendix C.1. The effect of average age of patient (variable 3) on mean infection risk (variable 4) is to be studied. For purposes of this ANOVA study, average age is to be classified into four categories: Under 50.0, 50.0–54.9, 55.0–59.9, 60.0 and over. Assume that ANOVA model (16.2) is applicable. Test whether or not the mean infection risk differs for the four age groups. Control the α risk at .10. State the alternatives, decision rule, and conclusion.

16.44. Refer to the **CDI** data set in Appendix C.2. The effect of geographic region (variable 17) on the crime rate (variable 10 ÷ variable 5) is to be studied. Assume that ANOVA model (16.2) is applicable. Test whether or not the mean crime rates for the four geographic regions differ; use $\alpha = .05$. State the alternatives, decision rule, and conclusion.

16.45. Refer to the **Market share** data set in Appendix C.3. Test whether or not the average monthly market share (variable 2) is the same for the four factor-level combinations associated with the two levels of each factor for discount price (variable 5) and package promotion (variable 6); use $\alpha = .05$. Assume that model (16.2) is applicable. State the alternatives, decision rule, and conclusion.

16.46. Consider a test involving H_0: $\mu_1 = \mu_2 = \mu_3$. Five observations are to be taken for each factor level, and level of significance $\alpha = .05$ is to be employed in the test.

a. Generate five random normal observations when $\mu_1 = 100$ and $\sigma = 12$ to represent the observations for treatment 1. Repeat this for the other two treatments when $\mu_2 = \mu_3 = 100$ and $\sigma = 12$. Finally, calculate F^* test statistic (16.55).

b. Repeat part (a) 100 times.

c. Calculate the mean of the 100 F^* statistics.

d. What proportion of the F^* statistics lead to conclusion H_0? Is this consistent with theoretical expectations?

e. Repeat parts (a) and (b) when $\mu_1 = 80$, $\mu_2 = 60$, $\mu_3 = 160$, and $\sigma = 12$. Calculate the mean of the 100 F^* statistics. How does this mean compare with the mean obtained in part (c) when $\mu_1 = \mu_2 = \mu_3 = 100$? Is this result consistent with the expectation in (16.37b)?

f. What proportion of the 100 test statistics obtained in part (e) lead to conclusion H_a? Does it appear that the test has satisfactory power when $\mu_1 = 80$, $\mu_2 = 60$, and $\mu_3 = 160$?

16.47. A completely randomized experiment involving $r = 2$ treatments was carried out, based on $n = 3$ experimental trials for each treatment. The test for equality of the treatment means is to be carried out by means of the randomization distribution of the F^* test statistic (16.55).

a. Determine the number of ways that the six experimental units can be divided into two groups of size three each. How many unique F^* statistics are possible?

b. For the sample results:

j:	1	2	3
Y_{1j}:	23	34	78
Y_{2j}:	17	29	23

obtain the randomization distribution of the test statistic F^* and the P-value of the randomization test.

c. Obtain the P-value of the normal-theory F^* statistic for the sample results in part (b). How does this P-value compare with the one from the randomization test in part (b)? What does this suggest about the appropriateness of the F distribution here if the error terms are far from normally distributed?

16.48. A completely randomized psychological reinforcement experiment was conducted in which a standard treatment and an experimental treatment were each applied to four subjects. The sample results are:

j:	1	2	3	4
Y_{1j} (standard treatment):	16	14	18	16
Y_{2j} (experimental treatment):	12	15	13	12

The test for equality of treatment means is to be carried out by means of the randomization distribution of the F^* test statistic (16.55), with $\alpha = .10$.

a. Obtain the randomization distribution of the test statistic F^* and carry out the indicated test. State the alternatives, decision rule, and conclusion. What is the P-value of the randomization test?

b. For the randomization distribution in part (a), determine the proportion of F^* values that exceed $F(.90; 1, 6)$, the proportion of F^* values that exceed $F(.95; 1, 6)$, and the proportion that exceed $F(.99; 1, 6)$.

c. How do the proportions obtained in part (b) compare with the probabilities for the normal error model? Discuss.

Case Studies

16.49. Refer to the **Prostate cancer** data set in Appendix C.5. Carry out a one-way analysis of variance of this data set, where the response of interest is PSA level (variable 2) and the single factor is Gleason score (variable 9). The analysis should consider transformations of the response variable. Document steps taken in your analysis, and justify your conclusions.

16.50. Refer to the **Real estate sales** data set in Appendix C.7. Carry out a one-way analysis of variance of this data set, where the response of interest is sales price (variable 2) and the single factor is number of bedrooms (variable 4). Recode the number of bedrooms into four categories: 0–2, 3, 4, and greater than or equal to 5. The analysis should consider transformations of the response variable. Document steps taken in your analysis, and justify your conclusions.

16.51. Refer to the **Ischemic heart disease** data set in Appendix C.9. Carry out a one-way analysis of variance of this data set, where the response of interest is total cost (variable 2) and the single factor is total number of interventions (variable 5). Recode the number of interventions into six categories: 0, 1, 2, 3–4, 5–7, and greater than or equal to 8. The analysis should consider transformations of the response variable. Document steps taken in your analysis, and justify your conclusions.

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
$Y_{i.}$	73	67	78	136	354
$\bar{Y}_{i.}$	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 for 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.

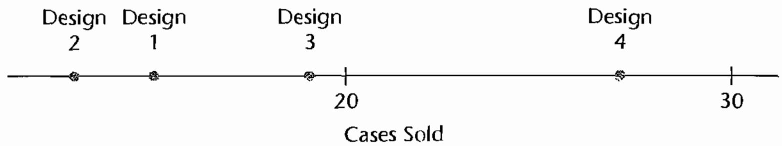
Line Plot

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.

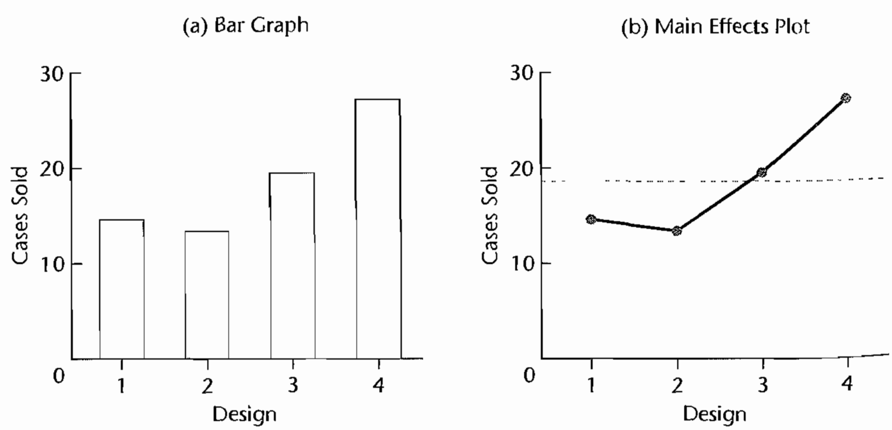
Bar Graph and Main Effects Plot

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:

For ANOVA model (17.1), SSE/σ^2 is distributed as χ^2 with $n_T - r$ degrees of freedom, and is independent of $\bar{Y}_{1\cdot}, \ldots, \bar{Y}_{r\cdot}$. **(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:

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

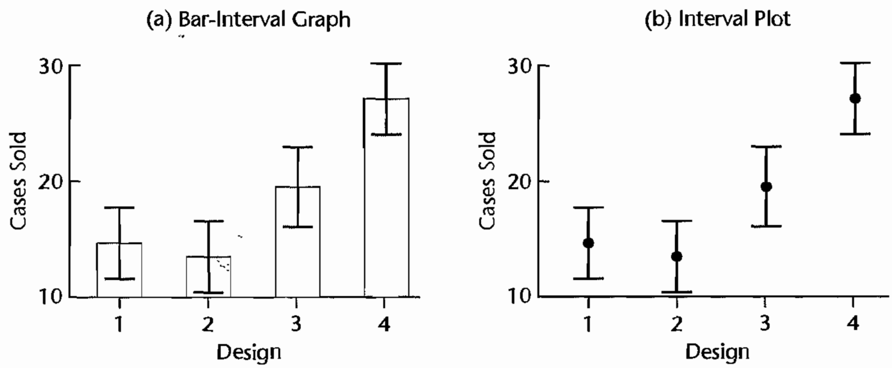
so that $s\{\bar{Y}_{1\cdot}\} = 1.453$. Hence, we obtain the confidence limits $14.6 \pm 2.131(1.453)$ and the 95 percent confidence interval is:

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

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:

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

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

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

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:

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

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:

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

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:

$$H_0\text{: } \sum c_i \mu_i = c$$
$$H_a\text{: } \sum c_i \mu_i \neq c$$

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:
>
> $$H_0: \mu_i - \mu_{i'} = 0$$
> $$H_a: \mu_i - \mu_{i'} \neq 0$$

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:

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

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_i. 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.

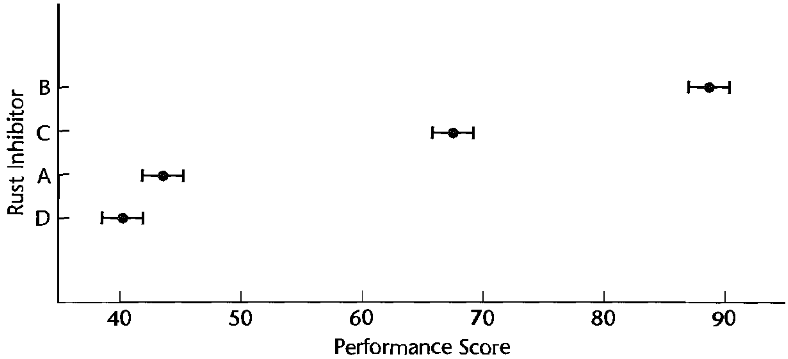
Simultaneous Testing

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:

$$q^* = \frac{\sqrt{2}\hat{D}}{s\{\hat{D}\}} \quad \textbf{(17.32)}$$

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 $\bar{Y}_{i\cdot}$ is plotted an interval whose limits are:

$$\bar{Y}_{i\cdot} \pm \frac{1}{2}Ts\{\hat{D}\} \quad \textbf{(17.33)}$$

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, $\mu_2 - \mu_1$:

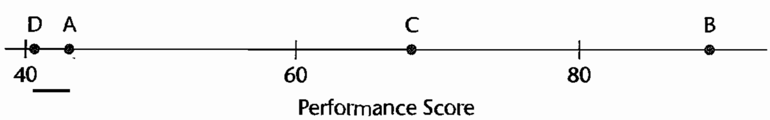
$$\hat{D} = \bar{Y}_{2\cdot} - \bar{Y}_{1\cdot} = 89.44 - 43.14 = 46.3$$

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 $\bar{Y}_{i\cdot}$ with the comparison intervals based on (17.33). For example, for rust inhibitor A, we have from earlier:

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

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:

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

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

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

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

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

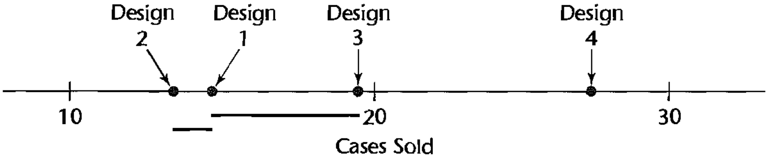
Hence, the confidence limits are $1.2 \pm 2.50(2.05)$ and the confidence interval for $\mu_1 - \mu_2$ is:

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

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:

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

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:

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

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

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

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:

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

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:

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

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):

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

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

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

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

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

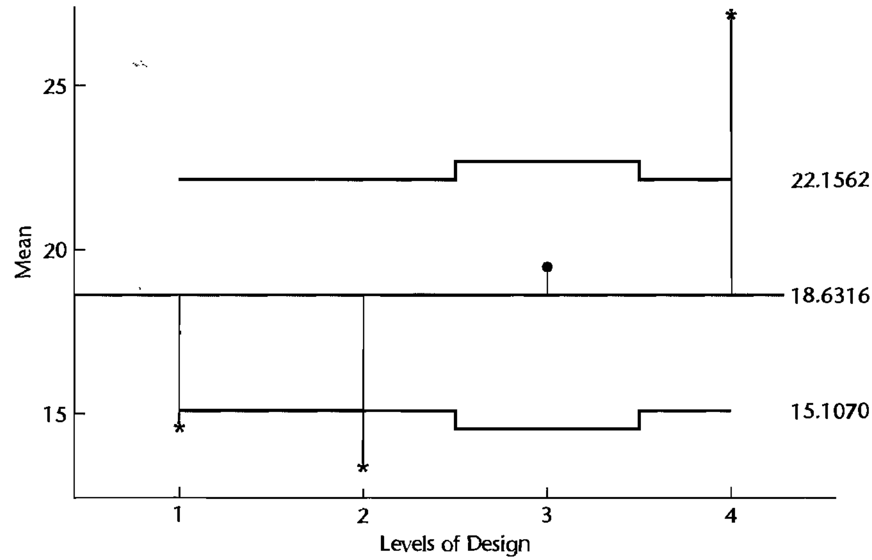
Simultaneous testing by the Bonferroni procedure can be carried out by setting up for each treatment effect the confidence interval using (17.46) and noting whether or not the interval contains zero. The results are sometimes summarized in an *analysis of means plot*. It is easy to show that a contrast $\hat{\tau}_i = \bar{Y}_i - \hat{\mu}.$ is inside (outside) one of the Bonferroni contrast intervals whenever the cell mean $\bar{Y}_{i\cdot}$ is inside (outside) the limits $\hat{\mu}. \pm t(1 - \alpha/2r; n_T - r)s\{\hat{\tau}_i\}$. In an analysis of means plot, the cell means are plotted along with the indicated limits and the least squares mean $\hat{\mu}.$ 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 $\hat{\mu}. = \bar{Y}_{..}$, 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
Treatment →	GRP01	7	40.0000	2.3094
	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.

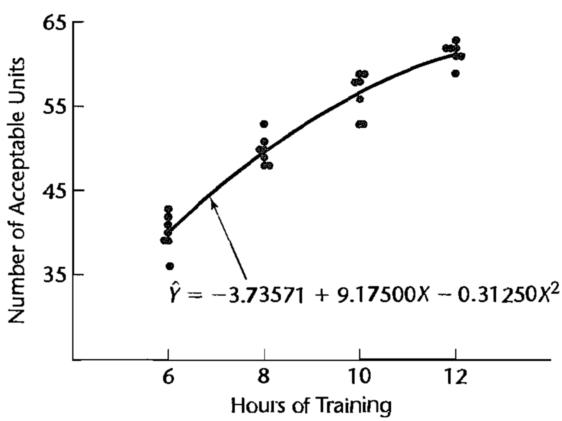
Investigation of Treatment Effects. The analyst's interest next centered on multiple comparisons of all pairs of treatment means. A Tukey multiple comparison option in the SPSS[X] computer package was used. It gave the output shown in the lower portion of Figure 17.6. This output presents the results 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:

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

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 $\bar{Y}_{5.}$ and $\bar{Y}_{3.}$ 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 estima[te].

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 ±.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.

Projects

17.35. Refer to the **SENIC** data set in Appendix C.1 and Project 16.42. Obtain confidence intervals for all pairwise comparisons between the four regions; use the Tukey procedure and a 90 percent family confidence coefficient. Interpret your results and state your findings. Prepare a line plot of the estimated factor level means and underline all nonsignificant comparisons.

17.36. Refer to the **CDI** data set in Appendix C.2 and Project 16.44. Obtain confidence intervals for all pairwise comparisons between the four regions; use the Tukey procedure and a 90 percent family confidence coefficient. Interpret your results and state your findings. Prepare a line plot of the estimated factor level means and underline all nonsignificant comparisons.

17.37. Refer to the **Market share** data set in Appendix C.3 and Project 16.45. Obtain confidence intervals for all pairwise comparisons among the four factor levels; use the Tukey procedure and a 95 percent family confidence coefficient. Interpret your results and state your findings. Prepare a line plot of the estimated factor level means, underscoring all nonsignificant comparisons.

17.38. Refer to Project 16.46e.

a. For each replication, construct confidence intervals for all pairwise comparisons among the three treatment means; use the Tukey procedure with a 95 percent family confidence coefficient. Then determine whether all confidence intervals for the replication are correct, given that $\mu_1 = 80$, $\mu_2 = 60$, and $\mu_3 = 160$.

b. For what proportion of the 100 replications are all confidence intervals correct? Is this proportion close to theoretical expectations? Discuss.

Case Studies

17.39. Refer to the **Prostate cancer** data set in Appendix C.5 and Case Study 16.49. Obtain confidence intervals for all pairwise comparisons among the three Gleason score levels; use the Tukey procedure and a 95 percent family confidence coefficient. Interpret your results and state your findings. Prepare a line plot of the estimated factor level means, underscoring all nonsignificant comparisons.

17.40. Refer to the **Real estate sales** data set in Appendix C.7 and Case Study 16.50. Obtain confidence intervals for all pairwise comparisons among the four number-of-bedroom categories; use the Tukey procedure and a 90 percent family confidence coefficient. Interpret your results and state your findings. Prepare a line plot of the estimated factor level means, underscoring all nonsignificant comparisons.

17.41. Refer to the **Ischemic heart disease** data set in Appendix C.9 and Case Study 16.51. Obtain confidence intervals for all pairwise comparisons among the six number-of-intervention categories; use the Tukey procedure and a 90 percent family confidence coefficient. Interpret your results and state your findings. Prepare a line plot of the estimated factor level means, underscoring all nonsignificant comparisons.