sums of squares are then calculated using this estimated value in place of the missing datum, except that the groups SS is slightly biased, a situation corrected by subtracting the quantity

$$\text{bias} = \frac{[B_j - (a-1)\hat{X}_{ij}]^2}{a(a-1)}. \tag{12.30}$$

Also, 1 must be subtracted from the total DF, which means that the remainder DF is also reduced by 1.

If there are more than one missing data, but not too many (say, no more than 10% of all data are missing, but in no case more than $a-1$ data missing), then the following procedure may be used. Insert into the data array estimates of all but one of the missing data. Such estimates may be guesses; one might use as a guess, the mean of the known data for the group in which the missing datum would be located, or even the mean of all known data. Then the remaining missing datum is estimated by Equation 12.29, where the group and block sums include all data except that one being estimated. Then that estimate is placed in the data array and one of the other estimates is recalculated using Equation 12.29. This is done for each missing datum in turn. Then, the process is repeated for each of the missing data, thus arriving at somewhat better estimates of all of them, This iterative, or repetitive, process is engaged in until there is no further change in the estimated values. Typically, only two (or sometimes three) cycles of the process are required.

The total SS, blocks SS, and groups SS are then computed using all data, both known and estimated. But this groups SS is biased and should be corrected in the following manner (Glen and Kramer, 1958). Use Equation 12.30 to determine the bias for each block having one missing datum. Then, use Equation 12.31 to determine the bias for each block containing two or more data:

$$\text{bias} = \sum^{m_j} \hat{X}_{ij}^2 + \frac{B_j^2}{a - m_j} - \frac{(B_j + \sum^{m_j} \hat{X}_{ij})^2}{a}, \tag{12.31}$$

where m_j is the number of missing data in block j (m_j must be less than a), and each summation is performed over all m_j missing data in that block.

Finally, the results of all the above bias calculations are added and this sum is subtracted from the groups SS. Using the total SS, the blocks SS, and the corrected groups SS, the remainder SS is calculated. The total DF and the remainder DF each must be reduced by the number of missing data estimated; the groups DF and blocks DF are as usual (i.e., $a-1$ and $b-1$, respectively).

12.5 Repeated-Measures Experimental Designs

A *repeated-measures* experimental design, also called a *within-subjects* or *treatment-by-subject* design, is one in which multiple measurements on the same experimental subject comprise the replicate data.*

*Kepner and Robinson (1988) discuss nonparametric methods for this experimental design.

In Example 10.1, the desire was to test the null hypothesis of no effect of four different diets on the body weight of animals. If, in the completely randomized experimental design for that hypothesis, there were four independent samples of animals (say, five animals per sample), then all members of a given sample would be fed one of the four diets. In a repeated-measures design there would not be independent samples of animals; instead, each of five animals would have its body weight recorded after being maintained on one of the diets, then each of them would be weighed after being fed another diet for a time, and so on. The array of data would look like the tabulation in Section 12.4, for the randomized complete block design, but where k columns of data represent the dietary treatments and n rows are the repeated measures on the several animals (referred to as "subjects"); the column (group) totals are denoted as G_i, the row (subject) totals as S_j, and the grand total of the data as $\sum\sum X_{ij}$ (which also is $\sum G_i$ and $\sum S_j$):

	Diets				
Subjects	1	2	3	4	*Totals*
1	X_{11}	X_{21}	X_{31}	X_{41}	S_1
2	X_{12}	X_{22}	X_{32}	X_{42}	S_2
3	X_{13}	X_{23}	X_{33}	X_{43}	S_3
4	X_{14}	X_{24}	X_{34}	X_{44}	S_4
5	X_{15}	X_{25}	X_{35}	X_{45}	S_5
Totals	G_1	G_2	G_3	G_4	$\sum\sum X_{ij}$

In this experimental design, the total sum of squares and degrees of freedom are computed as in Table 12.5 where $a = k$ and $b = n$. Then the total SS is apportioned into two parts: the sum of squares among subjects and the sum of squares within subjects. The measure of variability among subjects is

$$\text{subjects SS} = \frac{\sum_{j=1}^{n} S_j^2}{k} - C, \tag{12.32}$$

and the variability within subjects is

$$\text{within-subjects SS} = \text{total SS} - \text{subjects SS}. \tag{12.33}$$

The total degrees of freedom, $N - 1$, are partitioned between

$$\text{subjects DF} = n - 1 \tag{12.34}$$

and

$$\begin{aligned} \text{within-subjects DF} &= \text{total DF} - \text{subjects DF} \\ &= n(k - 1). \end{aligned} \tag{12.35}$$

Then, the within-subjects variability is partitioned into that due to the experimental treatment:

$$\text{groups SS} = \frac{\sum_{i=1}^{k} G_i^2}{n} - C, \tag{12.36}$$

and that unaccounted for by the treatments and the subjects:

$$\text{remainder SS} = \text{within-subjects SS} - \text{groups SS}. \tag{12.37}$$

And the associated degrees of freedom are:

$$\text{groups DF} = k - 1 \tag{12.38}$$

and

$$\begin{aligned} \text{remainder DF} &= \text{within-subjects DF} - \text{groups DF} \\ &= (k-1)(n-1). \end{aligned} \tag{12.39}$$

The "remainder" may also be referred to as the "groups × subjects interaction."

The null hypothesis of interest is that of no difference among the means of the treatment groups, and for this F = (groups MS)/(remainder MS) is employed. It is generally inadvisable, on the basis of the underlying statistical theory, to attempt to test for difference among subjects; and that is typically not an object of performing such an experiment.

Example 12.5 shows the results of an experiment in which blood cholesterol was measured in seven subjects after each subject had been treated with one of each of three drugs, one drug at a time (with time allowed between exposure to each drug to allow its effects to disappear from the animal). The hypothesis of interest is that the mean blood cholesterol level is the same regardless of treatment.

EXAMPLE 12.5 A repeated-measures analysis of variance. The data are blood cholesterol concentrations, in milligrams of cholesterol per deciliter of blood (mg/dl).

H_0: The mean cholesterol level is the same in animals on all three drugs (i.e., $\mu_1 = \mu_2 = \mu_3$).

H_A: The mean cholesterol level is not the same in animals on all three drugs.

Subjects	Drug 1	Drug 2	Drug 3	*Totals* (S_j)
1	164	152	178	494
2	202	181	222	605
3	143	136	132	411
4	210	194	216	620
5	228	219	245	692
6	173	159	182	514
7	161	157	165	483
Totals (G_i)	1281	1198	1340	3819

EXAMPLE 12.5 (continued)

$$\sum_{i=1}^{k}\sum_{j=1}^{n} X_{ij} = 3819$$

$$\sum_{i=1}^{k}\sum_{j=1}^{n} X_{ij}^2 = 715,393$$

$$C = \frac{(3819)^2}{21} = 694,512.43$$

$$\text{total SS} = \sum_{i=1}^{k}\sum_{j=1}^{n} X_{ij}^2 - C$$

$$= 715,393 - 694,512.43 = 20,880.57$$

$$\text{subjects SS} = \frac{\sum_{j=1}^{n} S_j^2}{k} - C$$

$$= \frac{(494)^2 + (605)^2 + \cdots + (483)^2}{3} - 694,512.43$$

$$= 713,243.67 - 694,512.43 = 18,731.24$$

$$\text{within-subjects SS} = \text{total SS} - \text{subjects SS}$$

$$= 20,880.57 - 18,731.24 = 2,149.33$$

$$\text{drugs SS} = \frac{\sum_{i=1}^{k} G_i^2}{n} - C$$

$$= \frac{(1281)^2 + (1198)^2 + (1340)^2}{7} - C$$

$$= 695,966.43 - 694,512.43 = 1,454.00$$

$$\text{remainder SS} = \text{within-subjects SS} - \text{drugs SS}$$

$$= 2,149.33 - 1,454.00 = 695.33$$

Analysis of Variance Table

Source of variation	SS	DF	MS
Total	20880.57	20	
Subjects	18731.24	6	
Within-Subjects	2149.33	14	
Drugs	1454.00	2	727.00
Remainder	695.33	12	57.94

To test H_0, $F = \dfrac{\text{drugs MS}}{\text{remainder MS}} = \dfrac{727.00}{57.94} = 12.6$

$F_{0.05(1),2,12} = 3.89$

Therefore, reject H_0.

$$0.001 < P < 0.0025 \quad [P = 0.0011]$$

The hypothesis of Example 12.5 could have been tested by using a total of twenty-one animals, randomly allocated seven to each of the drug treatment groups, by employing the one-factor ANOVA of Section 10.1. Or, the rows in the data tabulation of Example 12.5 could have represented siblings, or individuals with some factor other than genetics in common, in which case the randomized complete block design of Section 12.4 would have been appropriate. The repeated-measures design will be more powerful than the one-way ANOVA if there is consequential relationship among the data in each row, and it will be more powerful than the randomized block design if there is a stronger relationship within subjects than within blocks. And the repeated-measures design may be advantageous in its economical requirement for fewer experimental subjects. The computational results of this repeated-measures analysis are identical to those that would result from the same data subjected to a randomized block analysis.

The repeated-measures design is disadvantageous if there are effects of the sequence in which the treatments (diets in the present example) are administered to the subjects. Another disadvantage arises if insufficient time is allowed between the administration of different treatments to avoid *carryover* effects of the previous treatment. Carryover effect may often be counteracted by what is known as *counterbalancing*, whereby, to the extent possible, each subject receives the treatments in a different sequence.

In a repeated-measures experiment, we presume that there are correlations among the repeated measurements, in this case, weight measurements of the same individual on different diets. That is, it is reasonable to suppose that an individual that is heavier than its colleagues on one diet will also be heavier on another diet. But for the probability of a calculated F to be determined from tabled values of F, there should be equal correlations among pairs of groups of data. That is, the correlation between the data of groups 1 and 2 should be the same as the correlation between the data of groups 1 and 3, the same as that between the data of groups 2 and 3, and so on. This characteristic, referred to as *compound symmetry*, coupled with the usual ANOVA assumption of equality of group variances, is related to what statisticians call *sphericity* (Huynh and Feldt, 1970), or *circularity* (Rouanet and Lépine, 1970), an underlying assumption of repeated-measures ANOVA. Violation of the latter assumption is, unfortunately, common but difficult to test for, and results in a Type I error greater than the specified α. Box (1954) and Geisser and Greenhouse (1958) described an approximation procedure, tedious to apply, that does not have this assumption but that usually results in an α that is too low. (See also Crowder and Hand, 1990: 54–58; Girden, 1992: 19–26; Glantz and Slinker, 1990: 400–404; Keppel, 1991: 352–353; Maxwell and Delaney, 1990: 476–477.) An alternative procedure for analyzing repeated-measures experimental designs, one that does not depend upon the sphericity assumption, is *multivariate analysis of variance* (see Chapter 16)—abbreviated MANOVA—which has gained in popularity with the availability of computer programs to handle the relatively complex computations (Girden, 1992: 22–26; Maxwell and Delaney, 1990: Chapters 13 and 14; O'Brien and Kaiser, 1985; Stevens, 1996: Chapter 13).

Two Within-Subjects Factors. An experimental design might have two factors of interest—say, diet and exercise regimen—with the replication being multiple observations (e.g., body weight) made on the same experimental subject. That is, the body weight

of n animals would be measured at each of the ab combinations of a diets and b exercise regimens. This would be analyzed as a three-factor ANOVA, with subjects (i.e., animals) as the third (random-effects) factor. An appropriate computer program will yield the necessary sums of squares; the pertinent degrees of freedom would be $a - 1$ for diets, $b - 1$ for exercise, $(a - 1)(b - 1)$ for the interaction of diet × exercise (i.e., factor A × factor B interaction), $(a - 1)(n - 1)$ for the diet × subject (i.e., $A \times S$) interaction, $(b - 1)(n - 1)$ for the exercise × subject (i.e., $B \times S$) interaction, and $(a-1)(b-1)(n-1)$ for the diet × exercise × subject ($A \times B \times S$) interaction. The needed mean squares would be obtained by dividing each sum of squares by its associated degrees of freedom; and the hypothesis tests for diet (factor A), exercise treatment (factor B), and the $A \times B$ interaction would be tested as indicated in the second example of Appendix Section A.3.

12.6 Multiple Comparisons and Confidence Intervals in Two-Factor Analysis of Variance

If a two-factor analysis of variance reveals a significant effect among levels of a fixed-effects factor having more than two levels, then we can determine between which levels the difference(s) occur(s). This may be done using the Tukey test (Section 11.1) or the Newman-Keuls test (Section 11.2). The appropriate SE is calculated by Equation 11.2, substituting for n the number of data in each level (i.e., there are bn data in each level of factor A and an data in levels of factor B); s^2 is the within-cells MS: and ν is the within-cells degrees of freedom. If there is no replication in the experiment, then we are obliged to use the remainder MS in place of the within-cells MS, and to use the remainder DF as ν.

The calculation of confidence limits for the population mean estimated by each significantly different level mean can be performed by the procedures of Section 11.3, as can the computation of confidence limits for differences between pairs of significantly different level means.

If it is desired to compare a control mean to each of the other level means, Dunnett's test described in Section 11.4, may be used. Section 11.4 also shows how to calculate confidence limits for the difference between such means. Scheffé's procedure for multiple contrasts (Section 11.5) may also be applied to the levels of a factor, where the critical value in Equation 11.17 employs either a or b in place of k (depending, respectively, on whether the levels of factor A or B are being examined), and the within-cells DF is used in place of $N - k$. In all references to Chapter 11, n in the standard error computation is to be replaced by the number of data per level, and s^2 and ν are the within-cells MS and DF, respectively.

Multiple comparison testing and confidence interval determination are appropriate for levels of a fixed-effects factor but are not used with random-effects factors.

If Interaction Is Significant. On concluding that there is a significant interaction between factors A and B, it is generally not meaningful to test for differences among