Chapter 27

Repeated Measures and Related Designs

In this chapter we take up repeated measures designs—designs that are widely used in the behavioral and life sciences. We begin by considering some basic elements of repeated measures designs. We then take up single-factor repeated measures designs, after which we consider two-factor experiments with repeated measures on both one factor and on two factors. We conclude this chapter with an introduction to split-plot designs, which include two-factor repeated measures designs with repeated measures on one factor.

27.1 Elements of Repeated Measures Designs

Description of Designs

Repeated measures designs utilize the same subject (person, store, plant, test market, etc.) for each of the treatments under study. The subject therefore serves as a block, and the experimental units within a block may be viewed as the different occasions when a treatment is applied to the subject. A repeated measures study may involve several treatments or only a single treatment that is evaluated at different points in time. Subjects used in repeated measures studies in the behavioral and life sciences include persons, households, observers, and experimental animals. At other times the subjects in repeated measures designs are stores, test markets, cities, and plants. We shall refer to all of these study units used in repeated measures designs as *subjects*.

Three examples of repeated measures designs follow.

1. Fifteen test markets are to be used to study each of two different advertising campaigns. In each test market, the order of the two campaigns will be randomized, with a sufficient time lapse between the two campaigns so that the effects of the initial campaign will not carry over into the second campaign. The subjects in this study are the test markets.

2. Two hundred persons who have persistent migraine headaches are each to be given two different drugs and a placebo, for two weeks each, with the order of the drugs randomized for each person. The subjects in the study are the persons with migraine headaches.

3. In a weight loss study, 100 overweight persons are to be given the same diet and their weights measured at the end of each week for 12 weeks to assess the weight loss over

time. Here the subjects are the overweight persons, who are observed repeatedly to provide information about the effects of a single treatment over time.

Each of these studies involves a *repeated measures design* because the same subject is measured repeatedly. This key characteristic distinguishes this type of design from the designs considered earlier.

Advantages and Disadvantages

A principal advantage of repeated measures designs is that they provide good precision for comparing treatments because all sources of variability between subjects are excluded from the experimental error. Only variation within subjects enters the experimental error, since any two treatments can be compared directly for each subject. Thus, one may view the subjects as serving as their own controls. Another advantage of a repeated measures design is that it economizes on subjects. This is particularly important when only a few subjects (e.g., stores, plants, test markets) can be utilized for the experiment. Also, when interest is in the effects of a treatment over time, as when the shape of the learning curve for a new process operation is to be studied, it is usually desirable to observe the same subject at different points in time rather than observing different subjects at the specified points in time.

Repeated measures designs have a serious potential disadvantage, however, namely, that there may be several types of interference. One type of interference is an *order effect*, which is connected with the position in the treatment order. For instance, in evaluating five different advertisements, subjects may tend to give higher (or lower) ratings for advertisements shown toward the end of the sequence than at the beginning. Another type of interference is connected with the preceding treatment or treatments. For instance, in evaluating five different soup recipes, a bland recipe may get a higher (or lower) rating when preceded by a highly spiced recipe than when preceded by a blander recipe. This type of interference is called a *carryover effect*.

Various steps can be taken to minimize the danger of interference effects. Randomization of the treatment orders for each subject independently will make it more reasonable to analyze the data as if the error terms are independent. Allowing sufficient time between treatments is often an effective means of reducing carryover effects. It may be desirable at times to balance the order of treatment presentations and sometimes even the number of times each treatment is preceded by any other treatment. Latin square designs and crossover designs (discussed in Chapter 28) are helpful to this end.

How to Randomize

The randomization of the order of the treatments assigned to a subject is straightforward. For each subject, a random permutation is used to define the treatment order, and independent permutations are selected for the different subjects.

Comment

Designs with repeated measures, discussed here, need to be distinguished from designs with repeated observations, discussed in Section 26.7. In repeated measures designs, several or all of the treatments are applied to the same subject. Designs with repeated observations, on the other hand, are designs where several observations on the response variable are made for a given treatment applied to an experimental unit. It is possible to develop a repeated measures design with repeated observations, as when a given subject is exposed to each of the treatments under study and a number of observations are made at the end of each treatment application. ■

27.2 Single-Factor Experiments with Repeated Measures on All Treatments

We first consider repeated measures designs where the treatments are based on a single factor, as in the examples in Section 27.1. Almost always, the subjects in repeated measures designs (persons, stores, test markets, experimental animals) are viewed as a random sample from a population. Hence, *in all of the models for repeated measures designs to be presented in this chapter, the effects of subjects will be viewed as random.*

Figure 27.1 contains the layout for a single-factor experiment with repeated measures on all treatments. Here, there are five subjects and four treatments, with the order of treatments independently randomized for each subject. Notice that this layout corresponds to the one in Figure 21.1 for a randomized complete block design. Indeed, as we shall see next, the models for single-factor repeated measures designs are formally the same as the ones for randomized block designs, with blocks now considered to be subjects.

Model

When treatment effects are fixed, a model often appropriate for a single-factor repeated measures design is the following additive model:

$$Y_{ij} = \mu.. + \rho_i + \tau_j + \varepsilon_{ij} \tag{27.1}$$

where:

$\mu..$ is a constant
ρ_i are independent $N(0, \sigma_\rho^2)$
τ_j are constants subject to $\sum \tau_j = 0$
ε_{ij} are independent $N(0, \sigma^2)$
ρ_i and ε_{ij} are independent
$i = 1, \ldots, s; j = 1, \ldots, r$

FIGURE 27.1 Layout for Single-Factor Repeated Measures Design ($s = 5, r = 4$).

	Treatment Order			
	1	2	3	4
Subject 1	T_4	T_3	T_2	T_1
2	T_3	T_4	T_1	T_2
3	T_4	T_3	T_1	T_2
4	T_2	T_1	T_4	T_3
5	T_1	T_2	T_4	T_3

Note that repeated measures model (27.1) is identical to randomized block model (25.67) with random block effects, except that $n_b = s$.

Hence, we know from Section 25.5 that repeated measures model (27.1) assumes the following about the observations Y_{ij}:

$$E\{Y_{ij}\} = \mu.. + \tau_j \tag{27.2a}$$

$$\sigma^2\{Y_{ij}\} = \sigma_Y^2 = \sigma_\rho^2 + \sigma^2 \tag{27.2b}$$

$$\sigma\{Y_{ij}, Y_{ij'}\} = \sigma_\rho^2 = \omega\sigma_Y^2 \qquad j \neq j' \tag{27.2c}$$

$$\sigma\{Y_{ij}, Y_{i'j'}\} = 0 \qquad i \neq i' \tag{27.2d}$$

where ω is the coefficient of correlation between any two observations for the same subject:

$$\omega = \frac{\sigma_\rho^2}{\sigma_Y^2} \tag{27.2e}$$

Thus, repeated measures model (27.1) assumes that in advance of the random trials, any two treatment observations Y_{ij} and $Y_{ij'}$ for a given subject are correlated in the same fashion for all subjects. This key assumption implies, as we saw in (25.71), that the variance-covariance matrix of the observations Y_{ij} for any given subject has compound symmetry. Any two observations from different subjects in advance of the random trials are independent according to model (27.1).

Equally important, we know from Chapter 25 that repeated measures model (27.1) assumes that, once the subjects have been selected, any two observations for a given subject are independent. Thus, model (27.1) assumes that there are no interference effects in the repeated measures study, such as order effects or carryover effects from one treatment to the next.

Comment

If interaction effects between subjects and treatments are present, interaction model (25.74) can be employed. As we noted in Chapter 25, both the additive and interaction models lead to the same procedures for making inferences about the treatment effects. ■

Analysis of Variance and Tests

Since repeated measures model (27.1) is the same as randomized complete block model (25.67), the analysis of variance and the test for treatment effects will be the same as before.

Analysis of Variance. The ANOVA sums of squares for repeated measures model (27.1) are the same as in (21.6), but the names of two of the sums of squares are usually changed for repeated measures applications. The sum of squares for blocks in (21.6a) will now be called the *sum of squares for subjects,* and the interaction sum of squares between blocks and treatments in (21.6c) will now be called the *interaction sum of squares between treatments and subjects*. These two sums of squares will be denoted, respectively, by *SSS* and *SSTR.S*. Thus, the analysis of variance decomposition for single-factor repeated measures model (27.1) is:

$$SSTO = SSS + SSTR + SSTR.S \tag{27.3}$$

TABLE 27.1 ANOVA Table for Single-Factor Repeated Measures Design—ANOVA Model (27.1) with Subject Effects Random and Treatment Effects Fixed.

Source of Variation	*SS*	*df*	*MS*	$E\{MS\}$
Subjects	*SSS*	$s-1$	*MSS*	$\sigma^2 + r\sigma_\rho^2$
Treatments	*SSTR*	$r-1$	*MSTR*	$\sigma^2 + s\frac{\sum \tau_j^2}{r-1}$
Error	*SSTR.S*	$(r-1)(s-1)$	*MSTR.S*	σ^2
Total	*SSTO*	$sr-1$		

where:

$$SSTO = \sum_i \sum_j (Y_{ij} - \bar{Y}_{\cdot\cdot})^2 \tag{27.3a}$$

$$SSS = r \sum_i (\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^2 \tag{27.3b}$$

$$SSTR = s \sum_j (\bar{Y}_{\cdot j} - \bar{Y}_{\cdot\cdot})^2 \tag{27.3c}$$

$$SSTR.S = \sum_i \sum_j (Y_{ij} - \bar{Y}_{i\cdot} - \bar{Y}_{\cdot j} + \bar{Y}_{\cdot\cdot})^2 \tag{27.3d}$$

Note that no error sum of squares is present because there are no replications here.

Table 27.1 contains the analysis of variance table for repeated measures model (27.1). It is the same as the ANOVA table in Table 25.8 for additive randomized block model (25.67), except for the change in notation. Note again that in the absence of interactions between treatments and subjects, the interaction mean square *MSTR.S* is an unbiased estimator of the error variance σ^2.

Comment

In repeated measures studies, *SSTR* and *SSTR.S* are sometimes combined into a *within-subjects sum of squares SSW:*

$$SSW = SSTR + SSTR.S \tag{27.4}$$

which can be shown to equal:

$$SSW = \sum_i \sum_j (Y_{ij} - \bar{Y}_{i\cdot})^2 \tag{27.4a}$$

Hence, the ANOVA decomposition in (27.3) can also be expressed as follows:

$$SSTO = \underbrace{SSS}_{\substack{\text{Between-}\\ \text{subjects}\\ \text{variability}}} + \underbrace{SSW}_{\substack{\text{Within-}\\ \text{subjects}\\ \text{variability}}} \tag{27.5}$$

■

Test for Treatment Effects. As the $E\{MS\}$ column in Table 27.1 indicates, the appropriate statistic for the test on treatment effects:

$$H_0: \text{all } \tau_j = 0$$
$$H_a: \text{not all } \tau_j \text{ equal zero} \qquad (27.6a)$$

is:

$$F^* = \frac{MSTR}{MSTR.S} \qquad (27.6b)$$

When H_0 holds, F^* follows the F distribution, and the decision rule for controlling the Type I error at α is:

$$\text{If } F^* \leq F[1-\alpha; r-1, (r-1)(s-1)], \text{ conclude } H_0$$
$$\text{If } F^* > F[1-\alpha; r-1, (r-1)(s-1)], \text{ conclude } H_a \qquad (27.6c)$$

Example

In a wine-judging competition, four Chardonnay wines of the same vintage were judged by six experienced judges. Each judge tasted the wines in a blind fashion, i.e., without knowing their identities. The order of the wine presentation was randomized independently for each judge. To reduce carryover and other interference effects, the judges did not drink the wines and rinsed their mouths thoroughly between tastings. Each wine was scored on a 40-point scale; the higher the score, the greater is the excellence of the wine. The data for this competition are presented in Table 27.2. A plot of the wine scores for each judge is shown in Figure 27.2. We see that there are some distinct differences in ratings between judges but that the ratings for wines 3 and 4 are consistently best and for wine 1 generally worst. We also see that the rating curves for the judges do not appear to exhibit substantial departures from being parallel. Hence, an additive model appears to be appropriate.

The six judges are considered to be a random sample from the population of possible judges, while the four wines tasted are of interest in themselves. Hence, single-factor repeated measures model (27.1) was expected to be appropriate, with the effects of subjects (judges) considered random and the effects of treatments (wines) considered fixed. As

TABLE 27.2 Data—Wine-Judging Example (ratings on a scale of 0 to 40).

Judge	Wine (j)				
i	1	2	3	4	$\bar{Y}_{i\cdot}$
1	20	24	28	28	25
2	15	18	23	24	20
3	18	19	24	23	21
4	26	26	30	30	28
5	22	24	28	26	25
6	19	21	27	25	23
$\bar{Y}_{\cdot j}$	20.00	22.00	26.67	26.00	$23.67 = \bar{Y}_{\cdot\cdot}$

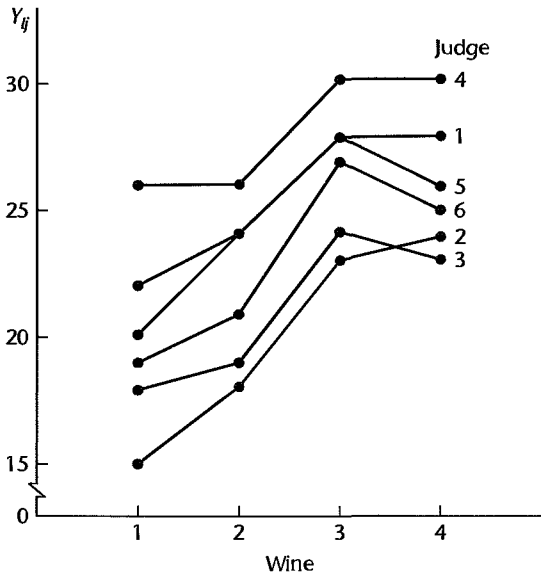


FIGURE 27.2 Plot of Wine Scores for Each Judge—Wine-Judging Example.

Factor	Type	Levels	Values					
Judge	random	6	1	2	3	4	5	6
Wine	fixed	4	1	2	3	4		

Analysis of Variance for Rating

Source	DF	SS	MS	F	P
Judge	5	173.333	34.667	32.50	0.000
Wine	3	184.000	61.333	57.50	0.000
Error	15	16.000	1.067		
Total	23	373.333			

FIGURE 27.3 MINITAB ANOVA Table for Single-Factor Repeated Measures Design—Wine-Judging Example.

we shall see later, additional diagnostic analysis supports the appropriateness of ANOVA model (27.1).

Figure 27.3 contains MINITAB ANOVA output for the wine-judging data in Table 27.2. To test for treatment effects:

$$H_0: \tau_1 = \tau_2 = \tau_3 = \tau_4 = 0$$

$$H_a: \text{not all } \tau_j \text{ equal zero}$$

we use the results of Table 27.3:

$$F^* = \frac{MSTR}{MSTR.S} = \frac{61.333}{1.067} = 57.5$$

For level of significance $\alpha = .01$, we require $F(.99; 3, 15) = 5.42$. Since $F^* = 57.5 > 5.42$, we conclude H_a, that the mean wine ratings for the four wines differ. The P-value for this test is 0+.

TABLE 27.3 Estimated Within-Subjects Variance-Covariance Matrix between Treatment Observations—Wine-Judging Example.

		j'			
		1	2	3	4
j	1	14.000	11.000	9.200	8.200
	2		10.000	8.200	7.600
	3			7.067	6.200
	4				6.800

Comments

1. As we noted in Chapter 25 (in Comment 2 on p. 1065), a conservative test for treatment effects should be used if the assumptions of compound symmetry in repeated measures model (27.1) are not met (i.e., if either the variances of the observations for different treatments for a given subject are not the same for all subjects or if the correlations between any two treatment observations for a given subject are not the same for all treatment pairs and for all subjects). In repeated measures studies, the compound symmetry assumption will be violated, for instance, if repeated responses over time are more highly correlated for observations closer together than for observations further apart in time.

2. When the treatment effects are random, test statistic (27.6b) and decision rule (27.6c) are still appropriate for testing treatment effects.

3. The efficiency of the repeated measures design in the wine-judging example, relative to a completely randomized design where each judge is used to assess a single wine, can be measured by means of (21.14). Using the results in Figure 27.3 with $n_b = s$, we obtain:

$$\hat{E} = \frac{(s-1)MSS + s(r-1)MSTR.S}{(sr-1)MSTR.S} = \frac{5(34.667) + 6(3)(1.067)}{23(1.067)} = 7.85$$

Thus, almost eight times as many replications per treatment would have been required with a completely randomized design in which each judge rates a single wine as in the repeated measures design to achieve the same precision for any estimated contrast.

4. When a single-factor repeated measures design involves $r = 2$ treatments, the F^* statistic in (27.6b) is equivalent to the two-sided t test for paired observations based on test statistic (A.69).

5. Occasionally, a formal test for subject effects is desired:

$$H_0: \sigma_\rho^2 = 0$$
$$H_a: \sigma_\rho^2 > 0$$

Table 27.1 indicates that the appropriate test statistic for repeated measures model (27.1) is $F^* = MSS/MSTR.S$. ■

Evaluation of Appropriateness of Repeated Measures Model

Since repeated measures model (27.1) is equivalent to randomized block model (25.67), the earlier discussion on diagnostics for randomized block models is entirely applicable here. In particular, a plot of the responses Y_{ij} by subject, as in Figure 27.2, can be examined for indications of serious lack of parallelism, which would suggest that additive model (27.1) may not be appropriate.

Residual sequence plots by subject can be helpful for studying constancy of the error variance and presence of interference effects. The residuals for repeated measures models (27.1) are the same as in (21.5):

$$e_{ij} = Y_{ij} - \bar{Y}_{i\cdot} - \bar{Y}_{\cdot j} + \bar{Y}_{\cdot\cdot} \tag{27.7}$$

A normal probability plot of the estimated residuals in (27.7) can be helpful for evaluating whether the residuals are normally distributed.

In addition to these graphic diagnostics, the estimated within-subjects variance-covariance and correlation matrices for the treatment observations Y_{ij} can be examined for appropriateness of the repeated measures model. A typical entry in the variance-covariance matrix is the estimated within-subjects covariance between observations for treatments j and j':

$$\frac{\sum_{i=1}^{s}(Y_{ij} - \bar{Y}_{j})(Y_{ij'} - \bar{Y}_{j'})}{s - 1} \tag{27.8}$$

The estimated within-subjects variance-covariance matrix should show variances of the same order of magnitude, and all of the covariances should be of similar magnitude. Of course, estimated variances and covariances tend to be subject to large sampling errors unless the sample sizes are very large. Hence, moderate differences in variances and covariances should be viewed as likely to be the result of sampling errors.

The estimated correlation matrix should show approximately similar coefficients of correlation between pairs of treatment observations within a subject.

Finally, the Tukey test described in Section 20.2 can be conducted to examine the appropriateness of the additive model. This test will need to be interpreted here as conditional on the subjects actually used in the repeated measures study.

Example

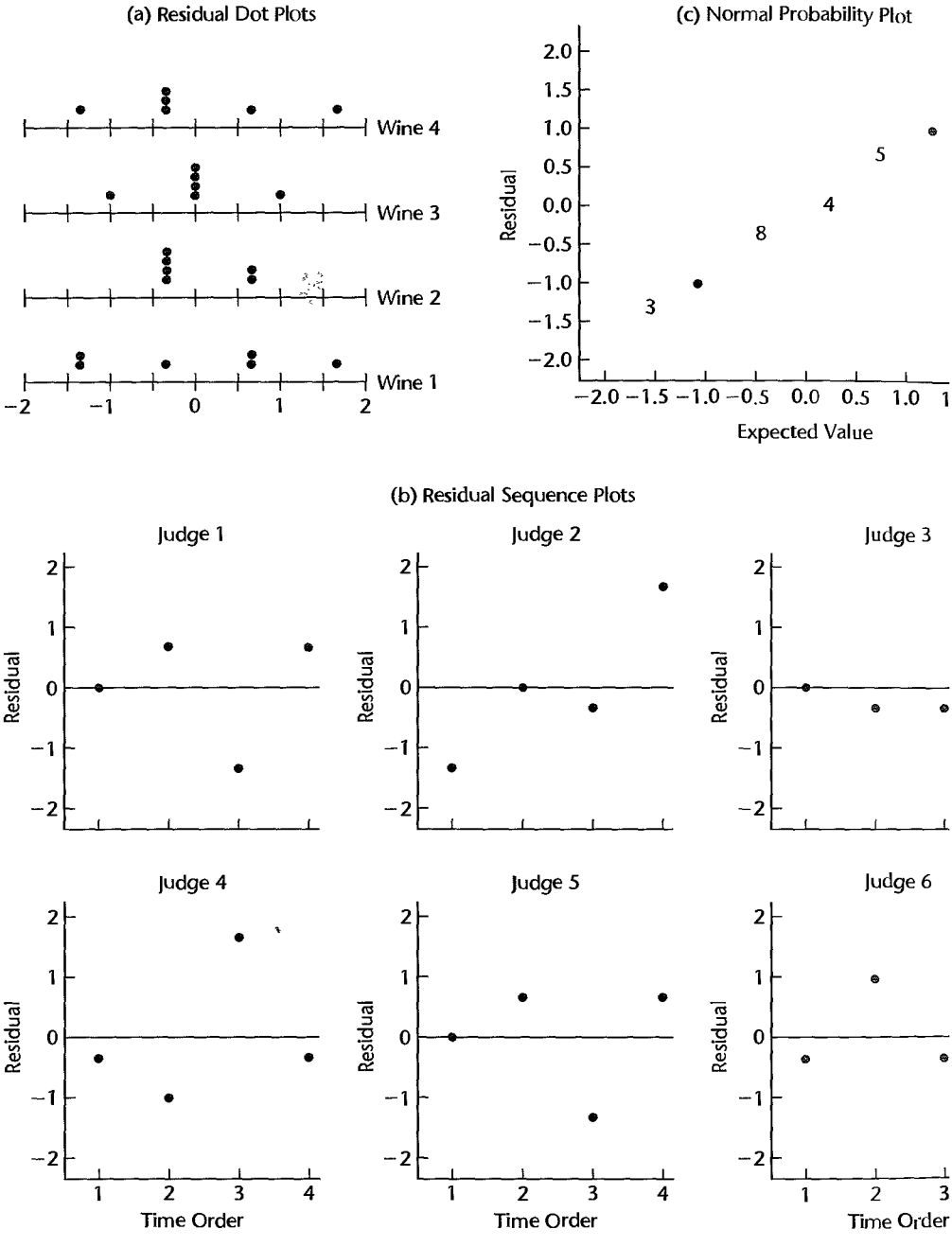
For the wine-judging example, the residuals were obtained from (27.7), and are presented in Figure 27.4a in SAS/GRAPH aligned dot plots by wine. These plots support the assumption of constant error variance. Figure 27.4b presents residual sequence plots for each judge, where the residuals are plotted in the order in which the wines were tasted by the judge. These plots do not indicate any correlations of the error terms within a judge, and thus suggest that no interference effects are present. Finally, a normal probability plot of the residuals is presented in Figure 27.4c. This plot shows evidence of the effects of the rounded nature of the data, but does not suggest any major departure from normality. The correlation between the ordered residuals and their expected values under normality is .993, which also suggests that lack of normality is not a problem here.

Table 27.3 presents the estimated within-subjects variance-covariance matrix for the treatment observations. The differences found there could easily arise from sampling errors.

As we noted earlier, the plot of the responses by subject in Figure 27.2 also supports the appropriateness of model (27.1), since the plots for the judges are reasonably parallel. Thus, there is no indication of interactions between subjects and treatments.

On the basis of these and other diagnostics, it was concluded that repeated measures model (27.1) is reasonably appropriate for the data in the wine-judging example.

FIGURE 27.4 SAS/GRAPH Diagnostic Residual Plots—Wine-Judging Example.



Analysis of Treatment Effects

The analysis of treatment effects for single-factor repeated measures model (27.1) proceeds in exactly the same fashion as described in Section 21.5 for randomized block designs with fixed treatment effects. The multiples in (21.9) for setting up confidence intervals are applicable here as they stand. The mean square used in estimating the variance of the estimated contrast is still the interaction mean square, which is now denoted by *MSTR.S*. We shall illustrate the estimation procedures by an example.

Example

In the wine-judging example, it was desired to compare all treatment means $\mu_{\cdot j}$ pairwise, with a 95 percent family confidence coefficient. Here $\mu_{\cdot j}$ is the mean rating of wine j averaged over judges. The Tukey procedure was utilized for this purpose. Using (17.30) with *MSE* replaced by *MSTR.S* and the estimated pairwise difference denoted by $\hat{L}$, we obtain using the results in Figure 27.3:

$$s^2\{\hat{L}\} = MSTR.S\left(\frac{1}{s} + \frac{1}{s}\right) = 1.067\left(\frac{2}{6}\right) = .3557$$

Using (21.9b), we find for a 95 percent family confidence coefficient:

$$T = \frac{1}{\sqrt{2}}q(.95; 4, 15) = \frac{1}{\sqrt{2}}(4.08) = 2.885$$

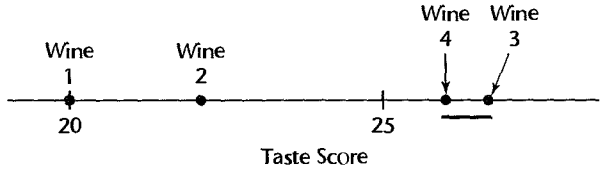
Hence:

$$Ts\{\hat{L}\} = 2.885\sqrt{.3557} = 1.72$$

Thus we obtain for the pairwise comparisons (see Table 27.2 for the $\bar{Y}_{j}$):

$$-2.39 = (26.00 - 26.67) - 1.72 \leq \mu_{\cdot 4} - \mu_{\cdot 3} \leq (26.00 - 26.67) + 1.72 = 1.05$$
$$2.28 = (26.00 - 22.00) - 1.72 \leq \mu_{\cdot 4} - \mu_{\cdot 2} \leq (26.00 - 22.00) + 1.72 = 5.72$$
$$4.28 = (26.00 - 20.00) - 1.72 \leq \mu_{\cdot 4} - \mu_{\cdot 1} \leq (26.00 - 20.00) + 1.72 = 7.72$$
$$2.95 = (26.67 - 22.00) - 1.72 \leq \mu_{\cdot 3} - \mu_{\cdot 2} \leq (26.67 - 22.00) + 1.72 = 6.39$$
$$4.95 = (26.67 - 20.00) - 1.72 \leq \mu_{\cdot 3} - \mu_{\cdot 1} \leq (26.67 - 20.00) + 1.72 = 8.39$$
$$.28 = (22.00 - 20.00) - 1.72 \leq \mu_{\cdot 2} - \mu_{\cdot 1} \leq (22.00 - 20.00) + 1.72 = 3.72$$

We display these results graphically as follows:



We conclude from these pairwise comparisons that wines 3 and 4 are judged best, and do not differ significantly from each other. Wines 1 and 2 are judged to be inferior to wines 3 and 4, with wine 1 receiving a mean rating significantly lower than that for wine 2. The family confidence coefficient of .95 applies to the entire set of comparisons.

TABLE 27.4
Ranked Data for Coffee Sweeteners in a Repeated Measures Design—Coffee Sweeteners Example.

Subject	Sweetener (j)				
i	A	B	C	D	E
1	5	1	2	4	3
2	4	2	1	5	3
3	3	2	1	4	5
4	5	2	3	4	1
5	4	1	2	3	5
6	4	1	3	5	2
$\bar{R}_{\cdot j}$	4.17	1.50	2.00	4.17	3.17

Comment

When the treatments are time order positions, as when process rework is observed for a new manufacturing process at periodic intervals. the nature of the time effect may be analyzed by developing an appropriate regression model. ■

Ranked Data

In repeated measures studies, the observations are frequently ranks, as when a number of tasters are each asked to rank recipes or when several university admissions officers are each asked to rank applicants for admission. When the data in a repeated measures study are ranks, the nonparametric rank F test described in Comment 3 on page 900 may be used for testing whether the treatment means are equal. No new principles are involved, so we shall proceed directly to an example.

Example

Six subjects were each asked to rank five coffee sweeteners according to their taste preferences, with rank 5 assigned to the most preferred sweetener. The data are presented in Table 27.4 and suggest that a sweetener effect may be present. For example, no judge ranked sweetener B higher than 2 (not preferred).

Test statistic (21.7b) for the ranked data here is:

$$F_R^* = \frac{9.00}{1.20} = 7.5$$

For level of significance $\alpha = .05$, we need $F(.95; 4, 20) = 2.87$. Since $F_R^* = 7.5 > 2.87$, we conclude that the five sweeteners are not equally liked. The P-value of the test is .0007.

Multiple Pairwise Testing Procedure

Just as in the case of the rank F test for single-factor studies (Section 18.7), we can use a large-sample testing analog of the Bonferroni pairwise comparison procedure to obtain information about the comparative magnitudes of the treatment means for repeated measures designs when the rank F test (or the Friedman test) indicates that the treatment means differ. Testing limits for all $g = r(r-1)/2$ pairwise comparisons using the mean ranks $\bar{R}_{\cdot j}$ are set up as follows for family level of significance α:

$$\bar{R}_{\cdot i} - \bar{R}_{\cdot j'} \pm B\left[\frac{r(r+1)}{6s}\right]^{1/2} \tag{27.9}$$

where:

$$B = z(1 - \alpha/2g) \tag{27.9a}$$

$$g = \frac{r(r-1)}{2} \tag{27.9b}$$

If the testing limits include zero, we conclude that the corresponding treatment means $\mu_{\cdot j}$ and $\mu_{\cdot j'}$ do not differ. If the testing limits do not include zero, we conclude that the two corresponding treatment means differ. We can then set up groups of treatments whose means do not differ according to this simultaneous testing procedure.

Example

We now wish to make all pairwise tests by means of (27.9) with family level of significance $\alpha = .20$ for the coffee sweeteners example. For $r = 5$, we have $g = 5(4)/2 = 10$ and obtain:

$$B = z[1 - .20/2(10)] = z(.99) = 2.326$$

Thus, the right term in (27.9) for $s = 6$ and $r = 5$ is:

$$B\left[\frac{r(r+1)}{6s}\right]^{1/2} = 2.326\left[\frac{5(6)}{6(6)}\right]^{1/2} = 2.12$$

We note from Table 27.4 that the pairs of mean ranks whose difference does not exceed 2.12 are (B, C), (B, E), (C, E), (A, E), (D, E), and (A, D). Hence, we can set up two groups, within which the treatment means do not differ:

Group 1		Group 2	
Sweetener B	$\bar{R}_{\cdot 2} = 1.50$	Sweetener E	$\bar{R}_{\cdot 5} = 3.17$
Sweetener C	$\bar{R}_{\cdot 3} = 2.00$	Sweetener A	$\bar{R}_{\cdot 1} = 4.17$
Sweetener E	$\bar{R}_{\cdot 5} = 3.17$	Sweetener D	$\bar{R}_{\cdot 4} = 4.17$

Thus, we conclude with family level of significance of .20 that sweeteners A and D are preferred to sweeteners B and C, and that it is not clear whether sweetener E belongs in the preferred group or in the other group.

Comments

1. The rank F test can also be used for repeated measures designs where the observations are not ranked, in case the distribution of the error terms departs far from normality. Ranks of the observations Y_{ij} are then assigned within each subject, and the rank F test is carried out in the usual manner.

2. The test statistic F_R^* is related to Kendall's coefficient of concordance W in the following way:

$$W = \frac{F_R^*}{F_R^* + n - 1} \tag{27.10}$$

The coefficient of concordance W is a measure of the agreement of the rankings of the s subjects. It equals 1 if there is perfect agreement, and equals 0 if there is no agreement, that is, if all treatments receive the same mean ranking. For the coffee sweeteners example in Table 27.4, the coefficient of concordance W is:

$$W = \frac{7.5}{7.5 + 6 - 1} = .60$$

This measure indicates that a fair amount of agreement exists between the subjects. ■

27.3 Two-Factor Experiments with Repeated Measures on One Factor

Description of Design

In many two-factor studies, repeated measures can only be made on one of the two factors. Consider, for instance, an experimenter who wished to study the effects of two types of incentives (factor A) on a person's ability to solve problems. The researcher also wanted to study two types of problems (factor B)—abstract and concrete problems. Each experimental subject could be asked to do each type of problem, but could not be exposed to more than one type of incentive stimulus because of potential interference effects. Thus, the design the experimenter utilized may be represented schematically as shown in Figure 27.5.

In a two-factor experiment with repeated measures on one factor, two randomizations generally need to be employed. First, the level of the nonrepeated factor (A, in Figure 27.5) needs to be randomly assigned to the subjects. Second, the order of the levels of the repeated factor (B, in Figure 27.5) needs to be randomized independently for all subjects.

Since s subjects are randomly assigned incentive stimulus A_1 and s subjects are randomly assigned incentive stimulus A_2, as far as factor A is concerned the experiment is a completely randomized one. On the other hand, as far as factor B (type of problem) is concerned, each subject is a block. Thus, for factor B, the experiment is a randomized complete block design, with block effects random. We call this experimental design a *two-factor experiment with repeated measures on factor B*.

In the experiment depicted in Figure 27.5, comparisons between factor A level means involve differences between groups of subjects as well as differences associated with the two factor A levels. On the other hand, comparisons between factor B level means at the same level of factor A are based on the same subject, and hence only involve differences associated with the two factor B levels. Thus, for these latter comparisons, each subject serves as its own control. The main effects of factor A are therefore said to be confounded

FIGURE 27.5
Layout for Two-Factor Design with Random Assignments of Factor *A* Level to Subjects and Repeated Measures on Factor *B*.

Incentive Stimulus	Subject	Treatment Order 1	Treatment Order 2
A_1	1	A_1B_1	A_1B_2
	⋮	⋮	⋮
	s	A_1B_1	A_1B_2
A_2	$s+1$	A_2B_2	A_2B_1
	⋮	⋮	⋮
	$2s$	A_2B_1	A_2B_2

with differences between groups of subjects, whereas the main effects of factor B are free of such confounding. It is for this reason that tests on factor B main effects will generally be more sensitive than tests on the main effects for factor A.

Comments

1. A two-factor experiment with repeated measures on one factor may be viewed as an incomplete block design. With reference to the repeated measures design in Figure 27.5, there are four treatments (A_1B_1, A_1B_2, A_2B_1, and A_2B_2) and one-half of the blocks (subjects) contain treatments A_1B_1 and A_1B_2 while the other half of the blocks contain treatments A_2B_1 and A_2B_2.

2. When the factor on which repeated measures are taken is time, randomization of the levels of the repeated factor is impossible. Consider, for instance, a study of two different advertising campaigns in which the effect on sales is to be measured in 10 test markets during four consecutive months. Here, the only randomization required is for assigning the advertising campaigns to the test markets. Similarly, when the nonrepeated factor is a characteristic of the subject, such as age of subject, no randomization is involved for that factor. ■

Model

The development of a model for a two-factor experiment with repeated measures on one factor is only a little more complex than for earlier cases. As before, we shall develop the model for random subject effects and fixed factor A and factor B effects. Let, as usual, α_j and β_k denote the factor A and factor B main effects, respectively, $(\alpha\beta)_{jk}$ the AB interaction effect, and ρ the subject (block) main effect. We do need to recognize, however, that the subject effect in this design is nested within factor A. Therefore, we will denote this effect by $\rho_{i(j)}$. As before, we assume that there are no interactions between treatments and subjects, although this condition is not essential here. A model that incorporates the above specifications is as follows for a balanced study, where the number of subjects receiving each level of factor A is the same:

$$Y_{ijk} = \mu\ldots + \rho_{i(j)} + \alpha_j + \beta_k + (\alpha\beta)_{jk} + \varepsilon_{ijk} \qquad \textbf{(27.11)}$$

where:

$\mu\ldots$ is a constant
$\rho_{i(j)}$ are independent $N(0, \sigma_\rho^2)$
α_j are constants subject to $\sum \alpha_j = 0$
β_k are constants subject to $\sum \beta_k = 0$
$(\alpha\beta)_{jk}$ are constants subject to $\sum_j (\alpha\beta)_{jk} = 0$ for all k and $\sum_k (\alpha\beta)_{jk} = 0$ for all j
ε_{ijk} are independent $N(0, \sigma^2)$
$\rho_{i(j)}$ and ε_{ijk} are independent
$i = 1, \ldots, s; j = 1, \ldots, a; k = 1, \ldots, b$

The observations Y_{ijk} for repeated measures model (27.11) have the following properties:

$$E\{Y_{ijk}\} = \mu\ldots + \alpha_j + \beta_k + (\alpha\beta)_{jk} \qquad \textbf{(27.12a)}$$

$$\sigma^2\{Y_{ijk}\} = \sigma_Y^2 = \sigma_\rho^2 + \sigma^2 \qquad \textbf{(27.12b)}$$

$$\sigma\{Y_{ijk}, Y_{ijk'}\} = \sigma_\rho^2 \qquad k \neq k' \qquad \textbf{(27.12c)}$$

$$\sigma\{Y_{ijk}, Y_{i'j'k'}\} = 0 \qquad i \neq i' \text{ and/or } j \neq j' \qquad \textbf{(27.12d)}$$

Note that the observations Y_{ijk} have constant variance. In addition, in advance of the random trials any two observations for different levels of factor B for the same subject have constant covariance, for all subjects, while observations for different subjects are independent. Also, all observations are assumed to be normally distributed.

Once the subjects have been selected, repeated measures model (27.11) assumes that any two observations for the same subject are independent, that is, that there are no interference effects.

Analysis of Variance and Tests

Analysis of Variance. The ANOVA sums of squares for repeated measures model (27.11) can be obtained by means of the rules in Appendix D. The sum of squares that is used for estimating the error variance turns out to be the interaction sum of squares $SSB.S(A)$. The ANOVA sums of squares are shown in Table 27.5. Also shown there are the degrees of freedom for each sum of squares.

Tests for Factor Effects. The expected mean squares for the analysis of variance in Table 27.5 are given in Table 27.6. These expected mean squares can be obtained by means of the rules in Appendix D.

It is clear from the expected mean squares in Table 27.6 that the test for AB interaction effects:

$$\begin{aligned} &H_0\text{: all } (\alpha\beta)_{jk} = 0 \\ &H_a\text{: not all } (\alpha\beta)_{jk} \text{ equal zero} \end{aligned} \tag{27.13a}$$

uses the test statistic:

$$F^* = \frac{MSAB}{MSB.S(A)} \tag{27.13b}$$

TABLE 27.5 Analysis of Variance for Two-Factor Experiment with Repeated Measures on Factor B—Model (27.11).

Source of Variation	SS	df
Factor A	$SSA = bs\sum_j(\bar{Y}_{j\cdot} - \bar{Y}_{\cdot\cdot})^2$	$a-1$
Factor B	$SSB = as\sum_k(\bar{Y}_{\cdot k} - \bar{Y}_{\cdot\cdot})^2$	$b-1$
AB interactions	$SSAB = s\sum_j\sum_k(\bar{Y}_{jk} - \bar{Y}_{j\cdot} - \bar{Y}_{\cdot k} + \bar{Y}_{\cdot\cdot})^2$	$(a-1)(b-1)$
Subjects (within factor A)	$SSS(A) = b\sum_i\sum_j(\bar{Y}_{ij\cdot} - \bar{Y}_{j\cdot})^2$	$a(s-1)$
Error	$SSB.S(A) = \sum_i\sum_j\sum_k(Y_{ijk} - \bar{Y}_{jk} - \bar{Y}_{ij\cdot} + \bar{Y}_{j\cdot})^2$	$a(s-1)(b-1)$
Total	$SSTO = \sum_i\sum_j\sum_k(Y_{ijk} - \bar{Y}_{\cdot\cdot})^2$	$abs-1$

TABLE 27.6 Expected Mean Squares for a-Factor Experiment with Repeated Measures on Factor *B*—Model (27.11) (*A*, *B* fixed, subjects random).

Source of Variation	*MS*	*E*{*MS*}
Factor *A*	*MSA*	$\sigma^2 + b\sigma^2_{\rho} + bs\dfrac{\sum \alpha_j^2}{a-1}$
Factor *B*	*MSB*	$\sigma^2 + as\dfrac{\sum \beta_k^2}{b-1}$
AB interactions	*MSAB*	$\sigma^2 + s\dfrac{\sum\sum (\alpha\beta)_{jk}^2}{(a-1)(b-1)}$
Subjects (within factor *A*)	*MSS*(*A*)	$\sigma^2 + b\sigma^2_{\rho}$
Error	*MSB.S*(*A*)	σ^2

and the decision rule for controlling the Type I error at α is:

$$\begin{aligned} &\text{If } F^* \leq F[1-\alpha;(a-1)(b-1),a(s-1)(b-1)], \text{ conclude } H_0 \\ &\text{If } F^* > F[1-\alpha;(a-1)(b-1),a(s-1)(b-1)], \text{ conclude } H_a \end{aligned} \tag{27.13c}$$

The test for factor A main effects:

$$\begin{aligned} &H_0\text{: all } \alpha_j = 0 \\ &H_a\text{: not all } \alpha_j \text{ equal zero} \end{aligned} \tag{27.14a}$$

uses the test statistic:

$$F^* = \frac{MSA}{MSS(A)} \tag{27.14b}$$

and the decision rule for controlling the Type I error at α is:

$$\begin{aligned} &\text{If } F^* \leq F[1-\alpha;a-1,a(s-1)], \text{ conclude } H_0 \\ &\text{If } F^* > F[1-\alpha;a-1,a(s-1)], \text{ conclude } H_a \end{aligned} \tag{27.14c}$$

Finally, the test for factor B main effects:

$$\begin{aligned} &H_0\text{: all } \beta_k = 0 \\ &H_a\text{: not all } \beta_k \text{ equal zero} \end{aligned} \tag{27.15a}$$

uses the test statistic:

$$F^* = \frac{MSB}{MSB.S(A)} \tag{27.15b}$$

and the decision rule for controlling the Type I error at α is:

$$\begin{aligned} &\text{If } F^* \leq F[1-\alpha;b-1,a(s-1)(b-1)], \text{ conclude } H_0 \\ &\text{If } F^* > F[1-\alpha;b-1,a(s-1)(b-1)], \text{ conclude } H_a \end{aligned} \tag{27.15c}$$

Comments

1. When the assumption of compound symmetry in repeated measures model (27.11) is not met, the conservative test discussed in Comment 2 on page 1065 should be employed.

2. When the study is not balanced (i.e., when the number of subjects within each level of factor A is not the same), the tests described here are no longer appropriate. Instead, the methods for unbalanced mixed and random effects models discussed in Section 25.7 can be employed. ■

Evaluation of Appropriateness of Repeated Measures Model

Our earlier discussion on evaluating the appropriateness of a repeated measures model applies here also. The residuals for repeated measures model (27.11) are:

$$e_{ijk} = Y_{ijk} - \bar{Y}_{jk} - \bar{Y}_{ij\cdot} + \bar{Y}_{i\cdot} \tag{27.16}$$

A special feature of repeated measures model (27.11) also warrants attention. This model requires that the variance between subjects, σ_ρ^2, be constant for all levels of factor A. This assumption can be examined by dot plots of the estimated subject effects $\bar{Y}_{ij\cdot} - \bar{Y}_{i\cdot}$ for each level of factor A.

We can also conduct a formal test of the equality of the between-subjects variances by noting that the variation between subjects within factor A, $SSS(A)$, can be decomposed into components for each factor A level:

$$SSS(A) = SSS(A_1) + SSS(A_2) + \cdots + SSS(A_a) \tag{27.17}$$

where:

$$SSS(A_j) = b\sum_i (\bar{Y}_{ij\cdot} - \bar{Y}_{i\cdot})^2 \tag{27.17a}$$

Each component sum of squares has $n - 1$ degrees of freedom associated with it. We can therefore test the equality of the between-subjects variances by means of the Hartley test statistic (18.8) or the Brown-Forsythe test statistic (18.12). For the latter test, d_{ij} in (18.11) is defined as the absolute difference between the estimated mean, $\bar{Y}_{ij\cdot}$, and the median of the estimated means $\bar{Y}_{1j\cdot}, \ldots, \bar{Y}_{aj\cdot}$.

Similarly, the error variation, $SSB.S(A)$, can be decomposed into components for each factor A level:

$$SSB.S(A) = SSB.S(A_1) + SSB.S(A_2) + \cdots + SSB.S(A_a) \tag{27.18}$$

where:

$$SSB.S(A_j) = \sum_i \sum_k (Y_{ijk} - \bar{Y}_{jk} - \bar{Y}_{ij\cdot} + \bar{Y}_{j\cdot})^2 \tag{27.18a}$$

Each component has $(s - 1)(b - 1)$ degrees of freedom associated with it. The Hartley or Brown-Forsythe tests can be conducted here also, this time to test for the equality of the error variance σ^2 for the different factor A levels.

The Hartley test assumes normality and is sensitive to this assumption. Hence, the appropriateness of the normality assumption should be established first before the Hartley test is employed. Unlike the Hartley test, the Brown-Forsythe test is robust and relatively insensitive to departures from normality.

Analysis of Factor Effects: Without Interaction

When the two factors do not interact or the interactions are not important, the main effects may be analyzed in a straightforward fashion. The relevant mean square to be used in the estimated variance of an estimated contrast of factor A level means for repeated measures model (27.11) is $MSS(A)$ because this mean square is the denominator of the appropriate F^* statistic for testing factor A main effects. Similarly, the mean square for estimating contrasts of factor B level means is $MSB.S(A)$.

The multiples for the estimated standard deviation of an estimated contrast of factor A or factor B level means are as follows:

Main A Effect	Main B Effect	
Single comparison		
$t[1-\alpha/2; a(s-1)]$	$t[1-\alpha/2; a(s-1)(b-1)]$	**(27.19a)**
Tukey procedure (for pairwise comparisons)		
$T=\frac{1}{\sqrt{2}}q[1-\alpha; a, a(s-1)]$	$T=\frac{1}{\sqrt{2}}q[1-\alpha; b, a(s-1)(b-1)]$	**(27.19b)**
Scheffé procedure		
$S^2=(a-1)F[1-\alpha; a-1, a(s-1)]$		
	$S^2=(b-1)F[1-\alpha; b-1, a(s-1)(b-1)]$	**(27.19c)**
Bonferroni procedure		
$B=t[1-\alpha/2g; a(s-1)]$	$B=t[1-\alpha/2g; a(s-1)(b-1)]$	**(27.19d)**

Note from Table 27.6 that the analysis of factor B effects can be carried out more precisely than that for factor A effects. The reason is that comparisons among factor A levels utilize $MSS(A)$, which involves the variability among the subjects as well as the experimental error, while comparisons among factor B levels utilize $MSB.S(A)$, which involves only experimental error.

Example 1

A national retail chain wanted to study the effects of two advertising campaigns (factor A) on the volume of sales of athletic shoes over time (factor B). Ten similar test markets (subjects, S) were chosen at random to participate in this study. The two advertising campaigns (A_1 and A_2) were similar in all respects except that a different national sports personality was used in each. Sales data were collected for three two-week periods (B_1: two weeks prior to campaign; B_2: two weeks during which campaign occurred; B_3: two weeks after campaign was concluded). The experiment was conducted during a six-week period when sales of athletic shoes are usually quite stable.

The data on sales (coded) are presented in Table 27.7, and are plotted in Figure 27.6 by test market for each advertising campaign. There is no evidence in Figure 27.6 of any interactions between the test markets and the treatments. In general, sales tended to increase during each advertising campaign, and then tended to decline to previous or lower levels than just before the campaign.

TABLE 27.7
Data—Athletic Shoes Sales Example.

Advertising Campaign	Test Market	Time Period $k=1$	$k=2$	$k=3$
$j=1$	$i=1$	958	1,047	933
	$i=2$	1,005	1,122	986
	$i=3$	351	436	339
	$i=4$	549	632	512
	$i=5$	730	784	707
$j=2$	$i=1$	780	897	718
	$i=2$	229	275	202
	$i=3$	883	964	817
	$i=4$	624	695	599
	$i=5$	375	436	351

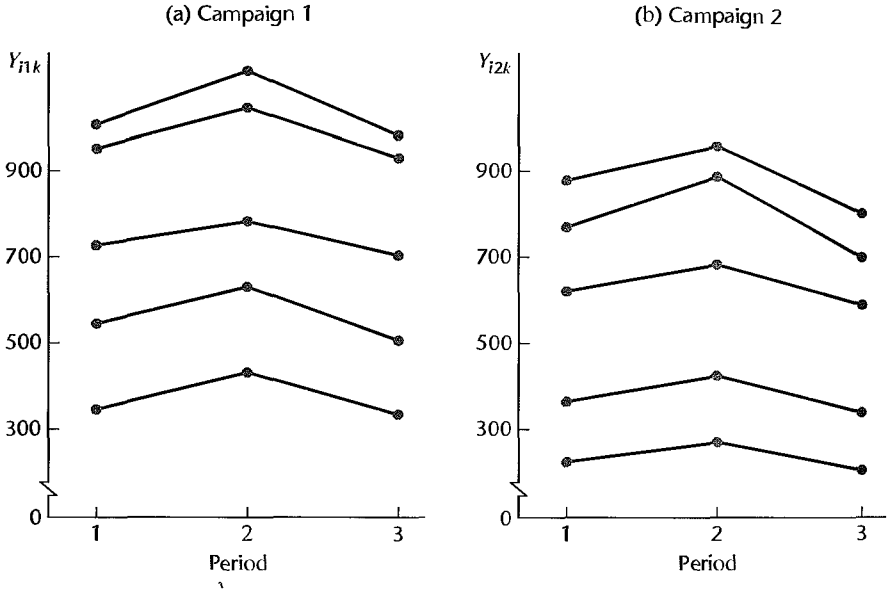


FIGURE 27.6
Plots of Sales Data by Test Market and Campaign—Athletic Shoes Sales Example.

From Figure 27.6 and other diagnostic analyses (not shown), it was concluded that repeated measures model (27.11) is appropriate here. Figure 27.7 contains the MINITAB output for the fit of this model.

First we wish to test for campaign-time interaction effects:

$$H_0: \text{all } (\alpha\beta)_{jk} = 0$$

$$H_a: \text{not all } (\alpha\beta)_{jk} \text{ equal zero}$$

We use the results from Figure 27.7 in test statistic (27.13b):

$$F^* = \frac{MSAB}{MSB.S(A)} = \frac{196}{358} = .55$$

FIGURE 27.7
MTAB Output for ANOVA—Athletic Shoes Sales Example.

Factor	Type	Levels	Values
A	fixed	2	1 2
S(A)	random	5	1 2 3 4 5
B	fixed	3	1 2 3

Analysis of Variance for Y

Source	DF	SS	MS	F	P
A	1	168151	168151	0.73	0.417
S(A)	8	1833681	229210	640.31	0.000
B	2	67073	33537	93.69	0.000
A*B	2	391	196	0.55	0.589
Error	16	5727	358		
Total	29	2075023	71553		

Source	Variance Component	Error Term	Expected Mean Square (using restricted model)
1 A		2	(5) + 3(2) + 15Q[1]
2 S(A)	76284.0	5	(5) + 3(2)
3 B		5	(5) + 10Q[3]
4 A*B		5	(5) + 5Q[4]
5 Error	358.0		(5)

MEANS

A	N	Y
1	15	739.40
2	15	589.67

B	N	Y
1	10	648.40
2	10	728.80
3	10	616.40

For level of significance $\alpha = .05$, we require $F(.95; 2, 16) = 3.63$. Since $F^* = .55 \leq 3.63$, we conclude H_0, that no significant interaction effects are present. The P-value for the test is .59.

Next we wish to test for advertising campaign main effects:

$$H_0: \text{all } \alpha_j = 0$$

$$H_a: \text{not all } \alpha_j \text{ equal zero}$$

We use the results from Figure 27.7 in test statistic (27.14b):

$$F^* = \frac{MSA}{MSS(A)} = \frac{168{,}151}{229{,}210} = .73$$

For level of significance $\alpha = .05$, we require $F(.95; 1, 8) = 5.32$. Since $F^* = .73 \leq 5.32$, we conclude H_0, that no advertising campaign main effects exist. The P-value for the test is .42. Thus, either of the two national sports personalities is equally effective in the advertising campaign.

Finally, we wish to test for time period effects:

$$H_0\text{: all } \beta_k = 0$$
$$H_a\text{: not all } \beta_k \text{ equal zero}$$

Using the results from Figure 27.7 in test statistic (27.15b), we obtain:

$$F^* = \frac{MSB}{MSB.S(A)} = \frac{33{,}537}{358} = 93.7$$

For level of significance $\alpha = .05$, we require $F(.95; 2, 16) = 3.63$. Since $F^* = 93.7 > 3.63$, we conclude H_a, that period main effects exist. The P-value for the test is 0+.

To examine the nature of the time period effects, we shall conduct pairwise comparisons of mean sales for the three time periods:

$$L = \mu_{\cdot\cdot k} - \mu_{\cdot\cdot k'}$$

The Tukey procedure will be employed, with a 99 percent family confidence coefficient. We require:

$$T = \frac{1}{\sqrt{2}} q(.99; 3, 16) = \frac{1}{\sqrt{2}}(4.78) = 3.38$$
$$s^2\{\hat{L}\} = \frac{2MSB.S(A)}{as} = \frac{2(358)}{2(5)} = 71.60$$

Hence, $Ts\{\hat{L}\} = 3.38\sqrt{71.60} = 28.6$.

The point estimates of the changes in mean sales, based on the estimated factor B level means $\bar{Y}_{\cdot\cdot k}$ in Figure 27.7, are:

$$\hat{L}_1 = \bar{Y}_{\cdot\cdot 2} - \bar{Y}_{\cdot\cdot 1} = 728.8 - 648.4 = 80.4$$
$$\hat{L}_2 = \bar{Y}_{\cdot\cdot 3} - \bar{Y}_{\cdot\cdot 1} = 616.4 - 648.4 = -32.0$$
$$\hat{L}_3 = \bar{Y}_{\cdot\cdot 3} - \bar{Y}_{\cdot\cdot 2} = 616.4 - 728.8 = -112.4$$

and the desired confidence intervals therefore are:

$$52 \le \mu_{\cdot\cdot 2} - \mu_{\cdot\cdot 1} \le 109$$
$$-61 \le \mu_{\cdot\cdot 3} - \mu_{\cdot\cdot 1} \le -3$$
$$-141 \le \mu_{\cdot\cdot 3} - \mu_{\cdot\cdot 2} \le -84$$

We conclude with family confidence coefficient .99 that the two advertising campaigns lead to an immediate increase in mean sales of between 52 and 109 (8 to 17 percent), but that mean sales in the following period fall below those for the period preceding the campaign by somewhere between 3 and 61 (.5 to 9 percent).

Analysis of Factor Effects: With Interaction

When interactions exist between the two factors, the analysis of factor effects becomes considerably more complex. As we saw in Chapter 19, page 848, when interaction effects are important, attention usually focuses on simple effects. To compare simple main effects of the repeated measure factor B, the appropriate error term for these pairwise comparisons remains $MSB.S(A)$, the same as when there is no interaction. However, the appropriate

error term used for the pairwise comparisons of the simple main effects for factor A needs to be modified from that used without interaction in comparing main effects of factor A. For each level of factor B considered individually, the analysis reduces to a single-factor experiment in which there are no repeated measures. Hence, the mean square within treatments is the appropriate error term to make pairwise comparisons among the treatment effects within each level of factor B. This mean square is a weighted average of $MSB.S(A)$ and $MSS(A)$ where the weights are the corresponding degrees of freedom:

$$MS(\text{Within Treatments}) = \frac{a(b-1)(s-1)MSB.S(A) + a(s-1)MSS(A)}{ab(s-1)}$$

Note that $MS(\text{Within Treatments})$ is a linear combination of mean squares whose expectations are not necessarily the same. Stated differently, $MS(\text{Within Treatments})$ represents a pooling of what will often be heterogeneous sources of variability.

To employ this error term as a basis for pairwise comparisons among the simple main effects, we employ the Satterthwaite procedure. The correspondences to (25.26) for $\hat{L} = MS(\text{Within Treatments})$ are:

$$MS_1 = MSB.S(A) \qquad MS_2 = MSS(A) \qquad c_1 = \frac{a(b-1)(s-1)}{ab(s-1)} \qquad c_2 = \frac{a(s-1)}{ab(s-1)}$$

Substitution of these values into (25.28) leads to the Satterthwaite adjusted degrees of freedom:

$$df_{adj} = \frac{[SSB.S(A) + SSS(A)]^2}{\dfrac{[SSB.S(A)]^2}{a(b-1)(s-1)} + \dfrac{[SSS(A)]^2}{a(s-1)}} \tag{27.20}$$

We will now illustrate the analysis of factor effects in the presence of interactions with an example.

Example 2

During exercise, blood flow increases in some parts of the body in response to metabolic demand. Using radioactive microspheres, an experiment was conducted to determine in which of five parts of the body (factor B) this occurs. Microspheres distribute in tissue as a function of blood flow; i.e., the greater the blood flow to a part of the body, the more microspheres (and radioactivity) it will contain. The experiment was designed to compare blood flow in five different parts of the body (factor B) between the resting control condition (factor A_1) and during exercise (factor A_2). Tissues were examined in the following parts of the body: bone, brain, skin, muscle, and heart. The experiment was conducted by injecting a total of eight rats (subjects) intravenously with radioactive microspheres. After the microspheres were injected, four rats were exercised on a treadmill for 15 minutes (factor A_2) and the other four rats were placed on the treadmill, but the treadmill was not turned on (factor A_1). At the end of the 15-minute period, the rats were sacrificed and tissues in the five parts were harvested and the radioactivity in the tissues was measured. The data for this blood flow experiment are presented in Table 27.8 and plotted in Figure 27.8 by body part for each exercise condition.

On the basis of Figure 27.8 and other diagnostic analyses (not shown), it was decided that repeated measure model (27.11) is appropriate here. Table 27.9 contains the analysis of variance table based on repeated measures model (27.11).

TABLE 27.8 Data—Blood Flow during Exercise Example.*

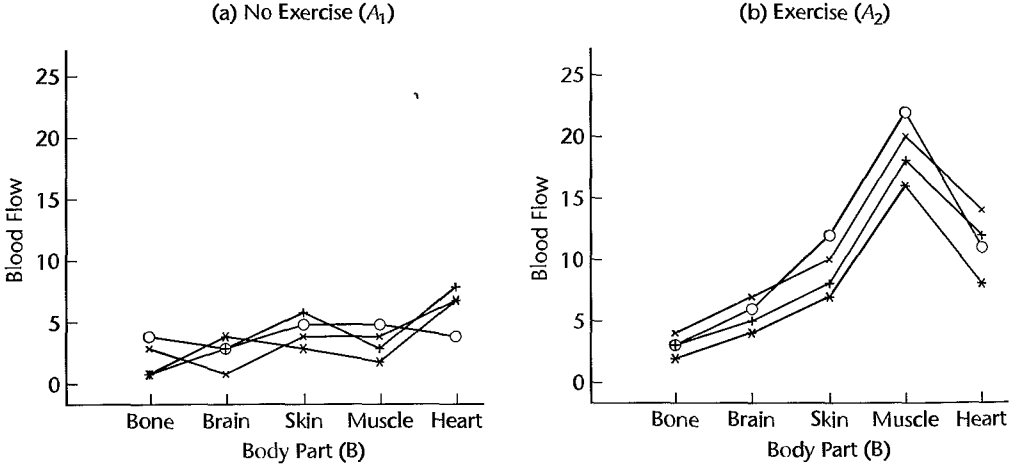
Exercise Condition		Body Part $k=1$ (Bone)	$k=2$ (Brain)	$k=3$ (Skin)	$k=4$ (Muscle)	$k=5$ (Heart)
(No Exercise)	$i=1$	4	3	5	5	4
	$i=2$	1	3	6	3	8
$j=1$	$i=3$	3	1	4	4	7
	$i=4$	1	4	3	2	7
(Exercise)	$i=1$	3	6	12	22	11
	$i=2$	3	5	8	18	12
$j=2$	$i=3$	4	7	10	20	14
	$i=4$	2	4	7	16	8

*Adapted from F.J. Gordon, *Analysis of Variance: Designs, Computations, and Multiple Comparisons*. Department of Pharmacology, Emory University School of Medicine, 2003.

TABLE 27.9 Analysis of Variance Table—Blood Flow during Exercise Example.

Source of Variation	*SS*	*df*	*MS*	*F**	*P*-value
A	324.9000	1	324.9000	44.104	.0006
S(*A*)	44.2000	6	7.3667		
B	389.5000	4	97.3750	49.936	.0000
AB	262.1000	4	65.5250	33.603	.0000
B.S(*A*)	46.8000	24	1.9500		
Total	1067.5000	39			

FIGURE 27.8 Plot of Exercise Condition by Body Part for Each Rat—Blood Flow during Exercise Example.



First we wish to test for exercise by body part interaction effects:

$$H_0\text{: all } (\alpha\beta)_{jk} = 0$$

$$H_a\text{: not all } (\alpha\beta)_{jk} \text{ equal zero}$$

We use the results from Table 27.9 as the test statistic (27.18a):

$$F^* = \frac{MSAB}{MSB.S(A)} = \frac{65.5250}{1.9500} = 33.603$$

For level of significance $\alpha = .05$, we require $F(.95; 4, 24) = 2.776$. Since $F^* = 33.6 > 2.776$, we conclude H_a, suggesting that interaction effects are present. The P-value for the test is 0+.

Next, because of the presence of a strong interaction effect, we wish to compare simple main effects of the repeated measures factor B (body part). We shall conduct pairwise comparisons of mean blood flows among body parts separately within the exercise and no exercise conditions; namely,

No Exercise	Exercise
$D_1 = \mu_{.11} - \mu_{.12}$	$D_{11} = \mu_{.21} - \mu_{.22}$
$D_2 = \mu_{.11} - \mu_{.13}$	$D_{12} = \mu_{.21} - \mu_{.23}$
$D_3 = \mu_{.11} - \mu_{.14}$	$D_{13} = \mu_{.21} - \mu_{.24}$
$D_4 = \mu_{.11} - \mu_{.15}$	$D_{14} = \mu_{.21} - \mu_{.25}$
$D_5 = \mu_{.12} - \mu_{.13}$	$D_{15} = \mu_{.22} - \mu_{.23}$
$D_6 = \mu_{.12} - \mu_{.14}$	$D_{16} = \mu_{.22} - \mu_{.24}$
$D_7 = \mu_{.12} - \mu_{.15}$	$D_{17} = \mu_{.22} - \mu_{.25}$
$D_8 = \mu_{.13} - \mu_{.14}$	$D_{18} = \mu_{.23} - \mu_{.24}$
$D_9 = \mu_{.13} - \mu_{.15}$	$D_{19} = \mu_{.23} - \mu_{.25}$
$D_{10} = \mu_{.14} - \mu_{.15}$	$D_{20} = \mu_{.24} - \mu_{.25}$

The Tukey procedure will be employed, with a 90 percent confidence coefficient, for each exercise condition. Then to combine these two Tukey procedures, a Bonferroni adjustment will be made for each exercise condition. Thus, we require

$$T = \frac{1}{\sqrt{2}} q(.95; 5, 24) = \frac{4.17}{\sqrt{2}} = 2.95$$

$$s^2\{\hat{D}\} = \frac{2MSB.S(A)}{s} = \frac{2(1.95)}{4} = .975$$

where .95 is used in the T argument instead of .90 to incorporate the Bonferroni adjustment for the two conditions. Hence, $Ts\{\hat{D}\} = 2.95\sqrt{.975} = 2.91$. Table 27.10 lists the cell means by exercise group and body part.

Any means within an exercise group that differ by more than 2.91 units are concluded to be significantly different from one another at the .10 level of significance. Therefore, for the no exercise group, heart is significantly different from bone, brain, and muscle. For the exercise group: heart is significantly different from bone, brain, and muscle; muscle is significantly different from bone, brain, skin, and heart; and skin is significantly different from bone, brain, and muscle.

TABLE 27.10 Treatment Means by Exercise Group and Body Part—Blood Flow during Exercise Example.

	$k = 1$ (Bone)	$k = 2$ (Brain)	$k = 3$ (Skin)	$k = 4$ (Muscle)	$k = 5$ (Heart)
$j = 1$ (No exercise)	2.25	2.75	4.50	3.50	6.50
$j = 2$ (Exercise)	3.00	5.50	9.25	19.00	11.25

To examine simple main effects of the nonrepeated measure factor A (exercise) for each level of B (body part), we shall conduct the five pairwise comparisons of mean blood flows between the two exercise groups within each body part; namely,

$$D_1 = \mu_{\cdot 11} - \mu_{\cdot 21}$$
$$D_2 = \mu_{\cdot 12} - \mu_{\cdot 22}$$
$$D_3 = \mu_{\cdot 13} - \mu_{\cdot 23}$$
$$D_4 = \mu_{\cdot 14} - \mu_{\cdot 24}$$
$$D_5 = \mu_{\cdot 15} - \mu_{\cdot 25}$$

The Tukey procedure will be employed using a 95 percent confidence coefficient for each body part with a Bonferroni adjustment for the five body parts. The within-treatment sum of squares is

$$SS(\text{Within Treatments}) = SSB.S(A) + SSS(A) = 46.8000 + 44.2000 = 91.0000$$

The approximate Satterthwaite adjusted degrees of freedom from (27.20) are:

$$df_{adj} = \frac{[46.8000 + 44.2000]^2}{\frac{(46.8000)^2}{2(4)(3)} + \frac{(44.2000)^2}{2(3)}} = \frac{8281.0000}{416.8667} = 19.86$$

Being conservative, we use $df_{adj} = 19$ associated with $MS(\text{Within Treatments})$, where

$$MS(\text{Within Treatments}) = \frac{91.0000}{30} = 3.033$$

Thus, we require

$$T = \frac{1}{\sqrt{2}} q(.99; 2.\ 19) = \frac{4.05}{\sqrt{2}} = 2.86$$

$$s^2\{\hat{D}\} = \frac{2MS(\text{Within Treatments})}{s} = \frac{2(3.033)}{4} = 1.52$$

Hence, $Ts\{\hat{D}\} = 2.86\sqrt{1.52} = 3.53$. Any means within body parts that differ by more than 3.53 units are significantly different from one another at the .10 level of significance. Therefore, we conclude that average blood flow for skin, muscle, and heart differ significantly between exercise groups.

FIGURE 27.9
Layout for Blocked Repeated Measures Design with Random Assignments of Factor *A* Level to Subjects and Repeated Measures on Factor *B*.

		Treatment Order	
		1	2
Block 1	Subject 1	A_2B_1	A_2B_2
	Subject 2	A_1B_2	A_1B_1
Block 2	Subject 3	A_1B_2	A_1B_1
	Subject 4	A_2B_2	A_2B_1
⋮	⋮	⋮	
Block n_b	Subject $2n_b - 1$	A_1B_1	A_1B_2
	Subject $2n_b$	A_2B_2	A_2B_1

Blocking of Subjects in Repeated Measures Designs

As already noted, comparisons among factor B effects can usually be carried out with greater precision than those for factor A effects because the latter involve between-subject variability as well as experimental error. To improve the precision of factor A comparisons, it is often helpful to block the subjects by some appropriate characteristic(s) so that the subjects within a block are homogeneous. Figure 27.9 illustrates the blocking of subjects in connection with the repeated measures design of Figure 27.5. Altogether, n_b blocks are used, each consisting of two similar subjects. One subject in each block is assigned at random to factor level A_1, the other is assigned to factor level A_2. In the second stage of randomization, each subject is randomly assigned the order of the two levels of factor B, namely, type of problem. Thus, the only difference between the repeated measures designs in Figures 27.9 and 27.5 is the blocking of the subjects for purposes of studying factor A effects more precisely. Note that for this layout, the number of subjects is $s = 2n_b$.

When there is a choice between which of the two factors should be the one on which repeated measures are taken (factor B), it should be the one for which more precise estimates are required. The reason is that even with blocking, the variability between subjects within a block will usually be greater than the variability within a subject.

27.4 Two-Factor Experiments with Repeated Measures on Both Factors

In Section 27.2 we considered single-factor repeated measures studies. The model for these designs can be extended when the treatments follow a factorial structure. For example, consider a study where four treatments are employed that represent two levels of each of two factors. Figure 27.10 depicts the layout for such a design when four subjects are utilized in the study. Note that the order of the treatments is randomized within each subject. When the treatments represent a factorial structure, we can explore as usual interaction effects as well as the main effects for the two factors. The design in Figure 27.10 is said to represent

FIGURE 27.10 Layout for Two-Factor Repeated Measures Design with Repeated Measures on Both Factors ($s = 4, a = 2, b = 2$).

	Treatment Order			
	1	2	3	4
Subject 1	A_1B_2	A_2B_2	A_1B_1	A_2B_1
2	A_2B_1	A_1B_2	A_2B_2	A_1B_1
3	A_2B_2	A_1B_1	A_2B_1	A_1B_2
4	A_1B_1	A_2B_1	A_1B_2	A_2B_2

repeated measures on both factors because each subject receives all treatments defined by the factorial structure.

Model

When both factor effects are fixed, the subjects constitute a random sample, and there are repeated measures on both factors, a model frequently appropriate is given by:

$$Y_{ijk} = \mu... + \rho_i + \alpha_j + \beta_k + (\alpha\beta)_{jk} + (\rho\alpha)_{ij} + (\rho\beta)_{ik} + \varepsilon_{ijk} \qquad \textbf{(27.21)}$$

where:

$\mu...$ is a constant

ρ_i are independent $N(0, \sigma_\rho^2)$

α_j are constants subject to $\sum \alpha_j = 0$

β_k are constants subject to $\sum \beta_k = 0$

$(\alpha\beta)_{jk}$ are constants subject to $\sum_j (\alpha\beta)_{jk} = 0$ for all k and $\sum_k (\alpha\beta)_{jk} = 0$ for all j

$(\rho\beta)_{ik}$ are $N\left(0, \frac{b-1}{b}\sigma_{\rho\beta}^2\right)$ subject to the restrictions $\sum_k (\rho\beta)_{ik} = 0$ for all i

$\sigma\{(\rho\beta)_{ik}, (\rho\beta)_{ik'}\} = -\frac{1}{b}\sigma_{\rho\beta}^2$ for $k \neq k'$

$(\rho\alpha)_{ij}$ are $N\left(0, \frac{a-1}{a}\sigma_{\rho\alpha}^2\right)$ subject to the restrictions $\sum_j (\rho\alpha)_{ij} = 0$ for all i

$\sigma\{(\rho\alpha)_{ij}, (\rho\alpha)_{ij'}\} = -\frac{1}{a}\sigma_{\rho\alpha}^2$ for $j \neq j'$

ρ_i, $(\rho\alpha)_{ij}$ and $(\rho\beta)_{ik}$ are pairwise independent

ε_{ijk} are independent $N(0, \sigma^2)$ and independent of ρ_i, $(\rho\alpha)_{ij}$ and $(\rho\beta)_{ik}$

$i = 1, \ldots, s; j = 1, \ldots, a; k = 1, \ldots, b$

Note that two of the interaction terms in the model are random since the factor ρ_i is a random effect and that all sums of effects over the fixed factor levels are zero.

The observations Y_{ijk} for repeated measures model (27.21) have the following properties:

$$E\{Y_{ijk}\} = \mu... + \alpha_j + \beta_k + (\alpha\beta)_{jk} \qquad \textbf{(27.22a)}$$

$$\sigma^2\{Y_{ijk}\} = \sigma_Y^2 = \sigma_\rho^2 + \frac{a-1}{a}\sigma_{\rho\alpha}^2 + \frac{b-1}{b}\sigma_{\rho\beta}^2 + \sigma^2 \qquad \textbf{(27.22b)}$$

Model (27.21) is an extension of the single-factor repeated measures model (27.1), where the treatment effect τ_j is now decomposed into factor A and factor B main effects and an AB interaction effect. However, separate first-order treatment-by-subject interaction terms are assumed to exist.

Once the subjects have been selected, repeated measures model (27.21), like the earlier repeated measures model (27.1), assumes that all of the treatment observations for a given subject are independent—that is, that there are no interference effects.

Analysis of Variance and Tests

Analysis of Variance. The ANOVA sums of squares for model (27.21) and the expected mean squares can be obtained readily by following the rules in Appendix D. The sum of squares for estimating the error variance terms reflects the interactions between treatments and subjects. Table 27.11 presents the ANOVA decomposition, degrees of freedom, and expected mean squares for two-factor repeated measures model (27.21).

Tests for Factor Effects. It is clear from the expected mean squares column in Table 27.11a that the test for AB interaction effects:

$$\begin{aligned} &H_0\text{: all } (\alpha\beta)_{jk} = 0 \\ &H_a\text{: not all } (\alpha\beta)_{jk} \text{ equal zero} \end{aligned} \tag{27.23a}$$

uses the test statistic:

$$F^* = \frac{MSAB}{MSABS} \tag{27.23b}$$

and the decision rule for controlling the Type I error at α is:

$$\begin{aligned} &\text{If } F^* \leq F[1-\alpha; (a-1)(b-1), (a-1)(b-1)(s-1)], \text{ conclude } H_0 \\ &\text{If } F^* > F[1-\alpha; (a-1)(b-1), (a-1)(b-1)(s-1)], \text{ conclude } H_a \end{aligned} \tag{27.23c}$$

The test for factor A main effects:

$$\begin{aligned} &H_0\text{: all } \alpha_j = 0 \\ &H_a\text{: not all } \alpha_j \text{ equal zero} \end{aligned} \tag{27.24a}$$

uses the test statistic:

$$F^* = \frac{MSA}{MSAS} \tag{27.24b}$$

and the decision rule for controlling the Type I error at α is:

$$\begin{aligned} &\text{If } F^* \leq F[1-\alpha; a-1, (a-1)(s-1)], \text{ conclude } H_0 \\ &\text{If } F^* > F[1-\alpha; a-1, (a-1)(s-1)], \text{ conclude } H_a \end{aligned} \tag{27.24c}$$

Similarly, the test for factor B main effects:

$$\begin{aligned} &H_0\text{: all } \beta_k = 0 \\ &H_a\text{: not all } \beta_k \text{ equal zero} \end{aligned} \tag{27.25a}$$

uses the test statistic:

$$F^* = \frac{MSB}{MSBS} \tag{27.25b}$$

TABLE 27.11 ANOVA Table and Sums of Squares for Two-Factor Repeated Measures Design with Repeated Measures on Both Factors—Subjects Random, Factors *A* and *B* Fixed.

(a) ANOVA Table

Source of Variation	*SS*	*df*	*MS*	$E\{MS\}$
Subjects(S)	*SSS*	$s-1$	*MSS*	$\sigma^2 + ab\sigma^2_\rho$
Factor *A*	*SSA*	$a-1$	*MSA*	$\sigma^2 + b\sigma^2_{\rho\alpha} + bs\dfrac{\sum\alpha_j^2}{a-1}$
Factor *B*	*SSB*	$b-1$	*MSB*	$\sigma^2 + a\sigma^2_{\rho\beta} + as\dfrac{\sum\beta_k^2}{b-1}$
AB interactions	*SSAB*	$(a-1)(b-1)$	*MSAB*	$\sigma^2 + s\dfrac{\sum\sum(\alpha\beta)^2_{jk}}{(a-1)(b-1)}$
AS interactions	*SSAS*	$(a-1)(s-1)$	*MSAS*	$\sigma^2 + b\sigma^2_{\rho\alpha}$
BS interactions	*SSBS*	$(b-1)(s-1)$	*MSBS*	$\sigma^2 + a\sigma^2_{\rho\beta}$
Error	*SSABS*	$(a-1)(b-1)(s-1)$	*MSABS*	σ^2
Total	*SSTO*	$abs-1$		

(b) Sums of Squares

$$SSS = ab\sum_i(\bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot\cdot\cdot})^2$$

$$SSA = sb\sum_j(\bar{Y}_{\cdot j\cdot} - \bar{Y}_{\cdot\cdot\cdot})^2$$

$$SSB = sa\sum_k(\bar{Y}_{\cdot\cdot k} - \bar{Y}_{\cdot\cdot\cdot})^2$$

$$SSAB = s\sum_j\sum_k(\bar{Y}_{\cdot jk} - \bar{Y}_{\cdot j\cdot} - \bar{Y}_{\cdot\cdot k} + \bar{Y}_{\cdot\cdot\cdot})^2$$

$$SSAS = b\sum_i\sum_j(\bar{Y}_{ij\cdot} - \bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot j\cdot} + \bar{Y}_{\cdot\cdot\cdot})^2$$

$$SSBS = a\sum_i\sum_k(\bar{Y}_{i\cdot k} - \bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot\cdot k} + \bar{Y}_{\cdot\cdot\cdot})^2$$

$$SSABS = \sum_i\sum_j\sum_k(Y_{ijk} - \bar{Y}_{ij\cdot} - \bar{Y}_{i\cdot k} - \bar{Y}_{\cdot jk} + \bar{Y}_{i\cdot\cdot} + \bar{Y}_{\cdot j\cdot} + \bar{Y}_{\cdot\cdot k} - \bar{Y}_{\cdot\cdot\cdot})^2$$

and the decision rule for controlling the Type I error at α is:

$$\begin{aligned} &\text{If } F^* \le F[1-\alpha; b-1, (b-1)(s-1)], \text{ conclude } H_0 \\ &\text{If } F^* > F[1-\alpha; b-1, (b-1)(s-1)], \text{ conclude } H_a \end{aligned} \qquad (27.25c)$$

Comments

1. When the effects of either factor *A* or factor *B* are random, the expected mean squares can be found by employing the rules in Appendix D. In turn, these expected mean squares will identify the appropriate test statistics.

2. Conservative *F* tests described in Section 25.5 should be used when the assumption of compound symmetry in repeated measures model (27.21) is not met.

3. Repeated measures model (27.21) assumes that treatments and subjects interact. If treatments and subjects do not interact, it can be shown that the treatment by subject interaction sum of squares is made up of three components:

$$SSTR.S = SSAS + SSBS + SSABS$$

Thus, it is possible to pool the first-order interactions in the model (the factor A by subject interactions and the factor B by subject interactions) with the second-order interactions (the factor A by factor B by subject interactions). When the repeated measures model does not allow for interactions between treatments and subjects, the analysis of factor effects becomes somewhat easier. However, in many cases, *MSABS* tends to be considerably smaller than either *MSAS* or *MSBS*, justifying the use of separate error terms. ■

Evaluation of Appropriateness of Repeated Measures Model

Our earlier discussion on the evaluation of the appropriateness of repeated measures model (27.1) applies here as well. In particular, residual sequence plots by subject should be constructed to examine whether interference effects are present and whether the error variance is constant. Plots of the observations by subject should be utilized to see whether the assumption of no treatment by subject interactions is appropriate.

Analysis of Factor Effects

If factors A and B do not interact or interact only in an unimportant fashion, the analysis of factor A and factor B main effects proceeds as usual. For the analysis of either factor A or factor B main effects, either *MSAS* or *MSBS*, respectively, will be used in the estimated variance of the estimated contrast since this mean square is the denominator of the F^* test statistic for testing factor A or factor B main effects.

The multiples for the estimated standard deviation of an estimated contrast of factor A or factor B level means are as follows:

Main *A* Effect	Main *B* Effect	
Single comparison		
$t[1-\alpha/2;(a-1)(s-1)]$	$t[1-\alpha/2;(b-1)(s-1)]$	**(27.26a)**
Tukey procedure (for pairwise comparisons)		
$T=\frac{1}{\sqrt{2}}q[1-\alpha;a,(a-1)(s-1)]$	$T=\frac{1}{\sqrt{2}}q[1-\alpha;b,(b-1)(s-1)]$	**(27.26b)**
Scheffé procedure		
$S^2=(a-1)F[1-\alpha;a-1,(a-1)(s-1)]$	$S^2=(b-1)F[1-\alpha;b-1,(b-1)(s-1)]$	**(27.26c)**
Bonferroni procedure		
$B=t[1-\alpha/2g;(a-1)(s-1)]$	$B=t[1-\alpha/2g;(b-1)(s-1)]$	**(27.26d)**

If strong interactions between factors A and B exist that cannot be made unimportant by some simple transformation, the analysis of the factor effects should be performed in terms of the treatment means $\mu_{\cdot jk}$, which are averaged over subjects. This analysis is similar to that in Section 27.3 for a two-factor study with interaction. The pooled mean square

MSTR.S will be used in estimating the variance of any estimated contrast of the treatment means. The degrees of freedom associated with *MSTR.S* will need to be estimated using the Satterthwaite procedure discussed before in Chapter 25, page 1043.

Example

A clinician studied the effects of two drugs used either alone or together on the blood flow in human subjects. Twelve healthy middle-aged males participated in the study and they are viewed as a random sample from a relevant population of middle-aged males. The four treatments used in the study are defined as follows:

A_1B_1	placebo (neither drug)
A_1B_2	drug B alone
A_2B_1	drug A alone
A_2B_2	both drugs A and B

The 12 subjects received each of the four treatments in independently randomized orders. The response variable is the increase in blood flow from before to shortly after the administration of the treatment. The treatments were administered on successive days. This wash-out period prevented any carryover effects because the effect of each drug is short-lived. The experiment was conducted in a double-blind fashion so that neither the physician nor the subject knew which treatment was administered when the change in blood flow was measured.

Table 27.12 contains the data for this study. A negative entry denotes a decrease in blood flow. Figure 27.11 contains the MINITAB output for the fit of repeated measures model (27.21). Included in the output are the expected mean squares for the specified ANOVA model. As explained in Chapter 25, each term in an expected mean square is represented in the MINITAB output by (1) the numeric code, in parentheses, for the variance of the model term and (2) the preceding number, which is the numerical multiple. When the model term is fixed, the letter Q is used in the printout to show that the variance is replaced by the sum of squared effects divided by degrees of freedom. For example, the expected value of *MSA* as shown in Figure 27.11 is:

$$(7) + 2(5) + 24Q[2] = \sigma^2 + 2\sigma^2_{\rho\alpha} + 24\frac{\sum \alpha_j^2}{2-1}$$

which corresponds, of course, to the factor A expected mean square shown in Table 27.11a.

TABLE 27.12
Data—Blood Flow Example.

Subject	Treatment			
i	A_1B_1	A_1B_2	A_2B_1	A_2B_2
1	2	10	9	25
2	−1	8	6	21
3	0	11	8	24
...	...	...	...	...
10	−2	10	10	28
11	2	8	10	25
12	−1	8	6	23

FIGURE 27.11 MINITAB Output for ANOVA—Blood Flow Example.

(a) MINITAB Output

Analysis of Variance for Flow

Source	DF	SS	MS	F	P
Subject	11	258.50	23.50	20.68	0.000
A	1	1587.00	1587.00	775.87	0.000
B	1	2028.00	2028.00	524.89	0.000
A*B	1	147.00	147.00	129.36	0.000
Subject*A	11	22.50	2.05	1.80	0.172
Subject*B	11	42.50	3.86	3.40	0.027
Error	11	12.50	1.14		
Total	47	4098.00			

Source	Variance Component	Error Term	Expected Mean Square for Each Term (using restricted model)
1 Subject	5.5909	7	(7) + 4(1)
2 A		5	(7) + 2(5) + 24Q[2]
3 B		6	(7) + 2(6) + 24Q[3]
4 A*B		7	(7) + 12Q[4]
5 Subject*A	0.4545	7	(7) + 2(5)
6 Subject*B	1.3636	7	(7) + 2(6)
7 Error	1.1364		(7)

(b) SAS Output

Source	DF	Type III SS	Mean Square	F Value	Pr > F
a	1	1587.000000	1587.000000	775.87	<.0001
Error(a)	11	22.500000	2.045455		

Source	DF	Type III SS	Mean Square	F Value	Pr > F
b	1	2028.000000	2028.000000	524.89	<.0001
Error(b)	11	42.500000	3.863636		

Source	DF	Type III SS	Mean Square	F Value	Pr > F
a*b	1	147.0000000	147.0000000	129.36	<.0001
Error(a*b)	11	12.5000000	1.1363636		

	N	Mean	Std Dev	Minimum	Maximum
a1b1	12	0.5000000	2.1105794	−2.0000000	4.0000000
a1b2	12	10.0000000	3.1908961	5.0000000	16.0000000
a2b1	12	8.5000000	2.0225996	6.0000000	12.0000000
a2b2	12	25.0000000	3.4377583	20.0000000	31.0000000

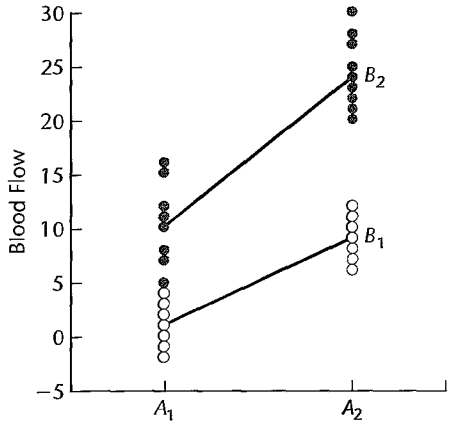


FIGURE 27.12 Interaction Plot with Responses Superimposed—Blood Flow Example.

Various diagnostics were utilized to see if repeated measures model (27.21) is appropriate for the data in Table 27.12. The results (not shown here) supported the appropriateness of this model. The clinician expected the two drugs to interact in increasing the blood flow. To test for interaction effects:

$$H_0\text{: all } (\alpha\beta)_{jk} = 0$$
$$H_a\text{: not all } (\alpha\beta)_{jk} \text{ equal zero}$$

we use test statistic (27.23b) and the results from Figure 27.11:

$$F^* = \frac{MSAB}{MSABS} = \frac{147.000}{1.1364} = 129.36$$

For level of significance $\alpha = .01$, we require $F(.99; 1, 11) = 9.65$. Since $F^* = 129.36 > 9.65$, we conclude H_a, that interaction effects exist. The P-value for this test is 0+.

Figure 27.12 contains an interaction plot of the estimated treatment means, with the responses superimposed. Substantial interaction effects are evident. To study the nature of the interaction effects, the clinician wished to compare the joint use of the two drugs with the use of each drug alone, drug A with drug B, and each drug with no drug. Thus, the following pairwise comparisons are to be made:

$$L_1 = \mu_{\cdot 22} - \mu_{\cdot 21} \qquad L_4 = \mu_{\cdot 21} - \mu_{\cdot 11}$$
$$L_2 = \mu_{\cdot 22} - \mu_{\cdot 12} \qquad L_5 = \mu_{\cdot 12} - \mu_{\cdot 11}$$
$$L_3 = \mu_{\cdot 21} - \mu_{\cdot 12}$$

Point estimates of these pairwise comparisons are ($\bar{Y}_{jk}$ values are in Figure 27.11b):

$$\hat{L}_1 = 25.0 - 8.5 = 16.5 \qquad \hat{L}_4 = 8.5 - .5 = 8.0$$
$$\hat{L}_2 = 25.0 - 10.0 = 15.0 \qquad \hat{L}_5 = 10.0 - .5 = 9.5$$
$$\hat{L}_3 = 8.5 - 10.0 = -1.5$$

The estimated variance of each estimate $\hat{L}$ is given in (17.22), with the relevant mean square here being *MSABS*. Hence, we have:

$$s^2\{\hat{L}\} = MSABS\left(\frac{1}{s} + \frac{1}{s}\right) = 1.1364\left(\frac{2}{12}\right) = .1894$$

and $s\{\hat{L}\} = .435$. Using the Bonferroni procedure with a 95 percent family confidence coefficient, we require $B = t[1 - (.05)/2(5); 11] = t(.995; 11) = 3.106$. Hence, $t(.995; 11)s\{\hat{L}\} = 3.106(.435) = 1.35$ and the desired confidence intervals with a 95 percent family confidence coefficient are:

$$15.15 \leq \mu_{\cdot 22} - \mu_{\cdot 21} \leq 17.85 \qquad 6.65 \leq \mu_{\cdot 21} - \mu_{\cdot 11} \leq 9.35$$
$$13.65 \leq \mu_{\cdot 22} - \mu_{\cdot 12} \leq 16.35 \qquad 8.15 \leq \mu_{\cdot 12} - \mu_{\cdot 11} \leq 10.85$$
$$-2.85 \leq \mu_{\cdot 21} - \mu_{\cdot 12} \leq -.15$$

It is clear from these results that either drug A alone or drug B alone leads to an increase in blood flow, and that the combination of the two drugs leads to a substantial additional increase in blood flow as compared to when either drug is used alone. Finally, a significant difference exists in the mean effects of the two drugs used alone.

Comments

1. Repeated measures designs are discussed in more detail in References 27.1 and 27.2.

2. In economics and econometrics, repeated measurement data over time are commonly referred to as *panel data*. The process of combining cross-sectional data and data over time to form a panel is called pooling. See References 27.3 and 27.4 for a discussion of these models and their analyses.

3. Another area of application for repeated measurement data is referred to as growth curve model analyses. Here separate regression models are fit to each subject over time. See Reference 27.5 for a discussion of these models and their analyses. ■

27.5 Regression Approach to Repeated Measures Designs

When the repeated measures study is balanced and the treatment effects are fixed, the analysis of variance model can be expressed in the form of a regression model with indicator variables for purposes of obtaining the various sums of squares and conducting tests for treatment effects. Repeated measures models (27.1) and (27.21) can be stated in the form of a regression model as explained in Section 23.4 for randomized block designs. Repeated measures model (27.11), which also involves nested effects, can be expressed in the form of a regression model by including suitable indicator variables as explained in Section 26.6 on page 1105.

When the repeated measures study is not balanced, as, for instance, when there are missing observations, the tests based on the expected mean squares in Tables 27.1, 27.6, and 27.11 are no longer appropriate. Methods for analyzing unbalanced mixed and random effects models are discussed in Section 25.7.

27.6 Split-Plot Designs

Split-plot designs are frequently used in field, laboratory, industrial, and social science experiments. The repeated measures design in Figure 27.5 for a study with repeated measures on one factor is a type of split-plot design. We shall discuss split-plot designs only for two-factor studies, but these designs can be extended to apply when three or more factors are under investigation.

Split-plot designs were originally developed for agricultural experiments. Consider an investigation to study the effects of two irrigation methods (factor A) and two fertilizers (factor B) on yield of a crop, using four available fields as experimental units. In a completely randomized design, four treatments (A_1B_1, A_1B_2, A_2B_1, A_2B_2) would then be assigned at random to the four fields. Since there are four treatments and just four experimental units, there will be no degrees of freedom for estimation of error, as shown in the following abbreviated ANOVA table, listing source of variation and degrees of freedom only:

Source of Variation	Degrees of Freedom
Factor A (irrigation methods)	1
Factor B (fertilizer types)	1
AB interactions	1
Error	0
Total	3

If the fields could be subdivided into smaller experimental units, replicates of each factor-level combination could be obtained and the error variance could then be estimated. Unfortunately, in this investigation it is not possible to apply different irrigation methods (factor A) in areas smaller than a field, although different fertilizer types (factor B) could be applied in relatively small areas. A split-plot design can accommodate this situation.

In a split-plot design, each of the two irrigation methods is randomly assigned to two of the four fields, which are usually called *whole plots*. In turn, each whole plot is then subdivided into two or more smaller areas called *split plots*, and the two fertilizers are then randomly assigned to the split plots within each whole plot. The key feature of split-plot designs is the use of two (or more) distinct levels of randomization. At the first level of randomization, the whole-plot treatments are randomly assigned to whole plots; at the second level, the split-plot treatments are randomly assigned to split plots.

The layout for the agricultural experiment example is shown in Figure 27.13. Note that this layout is conceptually identical to the layout for the two-factor repeated measures design in Figure 27.5. The fields in Figure 27.13 correspond to the subjects in Figure 27.5, and the split plots correspond to the occasions on which treatments can be applied to a subject. Consequently, the split-plot model here is the same as in (27.11):

$$Y_{ijk} = \mu\ldots + \rho_{i(j)} + \alpha_j + \beta_k + (\alpha\beta)_{jk} + \varepsilon_{ijk} \qquad \textbf{(27.27)}$$

For the split-plot agricultural experiment example, α_j denotes the main effect of the jth irrigation method (jth whole-plot treatment) and β_k denotes the main effect of the kth

FIGURE 27.13 Layout for Two-Factor Split-Plot Experiment—Agricultural Experiment Example (factor *A* is whole-plot treatment and factor *B* is split-plot treatment).

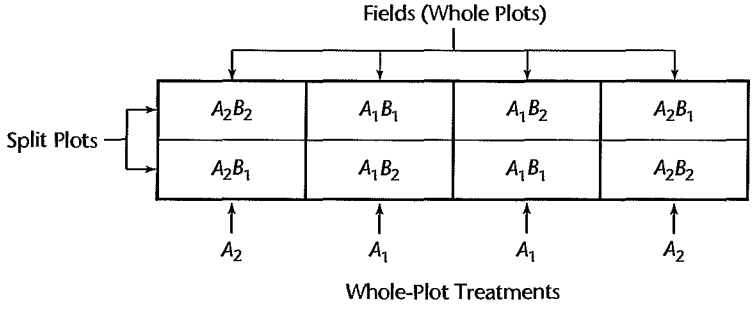


TABLE 27.13 ANOVA Table for Two-Factor Split-Plot Experiment.

Source of Variation	*SS*	*df*	*MS*
Whole plots			
Factor *A*	*SSA*	$a-1$	*MSA*
Whole-plot error	*SSW(A)*	$a(s-1)$	*MSW(A)*
Split plots			
Factor *B*	*SSB*	$b-1$	*MSB*
AB interactions	*SSAB*	$(a-1)(b-1)$	*MSAB*
Split-plot error	*SSB.W(A)*	$a(s-1)(b-1)$	*MSB.W(A)*
Total	*SSTO*	$abs-1$	

fertilizer type (kth split-plot treatment). Also, $\rho_{i(j)}$ denotes the effect of the ith whole plot, nested within the jth level of factor *A* (irrigation method).

Some computer packages produce special ANOVA tables that list the whole-plot effects and split-plot effects separately. Table 27.13 illustrates such a table. These tables serve as a reminder that the denominator of the *F* test for the whole-plot treatments is given by the error mean square for whole plots and that the denominator of the *F* test for the split-plot treatments and for the interactions between the whole-plot and split-plot treatments is given by the split-plot error mean square, as shown in Table 27.13. Note that this table is simply a rearrangement of the ANOVA table in Table 27.5 for a two-factor study with repeated measures on one factor. *SSS(A)* is now denoted by *SSW(A)* and *SSB.S(A)* is now denoted by *SSB.W(A)*. The expected mean squares are the same as in Table 27.6.

Comments

1. Whenever subjects can receive all treatments in a two-factor study without interference effects, a repeated measures design with repeated measures on both factors might be preferable, because the factor effects for both factors may be estimated more precisely than in a split-plot design.

2. Split-plot designs are useful in industrial experiments when one factor requires larger experimental units than another. Consider, for instance, a study of the effects of two additives (factor *A*) and two different containers (factor *B*) for prolonging the shelf life of a milk product. Here, it is easier to make larger batches of the milk product with a given additive, whereas the different containers can be used with smaller batches.

3. Split-plot designs may be viewed as a type of incomplete block design where the whole plots are considered to be the blocks, with each whole plot being given only some of the full set of treatments. Incomplete block designs are discussed in Chapter 28.

4. A wide variety of split-plot designs has been developed. For instance, split-plot designs can involve more than two stages of randomization. In a split-split-plot experiment, three stages of randomization are generally involved. Whole plots are divided into split plots and split plots are further divided into split split plots. Three treatments are then assigned to the various levels of experimental units, using three distinct stages of randomization. References 27.2 and 27.6 provide further information about these designs. ■

Cited References

27.1. Winer. B. J., D. R. Brown, and K. M. Michels. *Statistical Principles in Experimental Design.* 3rd ed. New York: McGraw-Hill Book Co., 1991.

27.2. Koch, G. G., J. D. Elashoff, and I. A. Amara. "Repeated Measurements—Design and Analysis," in *Encyclopedia of Statistical Sciences,* vol. 8, eds. S. Kotz and N. L. Johnson. New York: John Wiley & Sons, 1988, pp. 46–73.

27.3. Pindyck, R. S., and D. L. Rubinfeld. *Econometric Models and Economic Forecasts.* 4th ed. Boston: Irwin/McGraw-Hill, 1998.

27.4. Hsiao. C., *Analysis of Panel Data.* Cambridge: Cambridge University Press, 1986.

27.5. Graybill, F. A., *Theory and Application of the Linear Model.* Boston: Duxbury Press, 1976.

27.6. Dean, A., and D. Voss. *Design and Analysis of Experiments.* New York: Springer-Verlag, 1999.

Problems

27.1. A serious potential problem with repeated measures designs is associated with carryover effects. Describe some steps that can be taken to minimize this problem.

27.2. In designing a two-factor repeated measures study with repeated measures on one factor, does it matter which of the two factors is included as the repeated measures factor? Explain fully.

27.3. **Blood pressure.** The relationship between the dose of a drug that increases blood pressure and the actual amount of increase in mean diastolic blood pressure was investigated in a laboratory experiment. Twelve rabbits received in random order six different dose levels of the drug, with a suitable interval between each drug administration. The increase in blood pressure was used as the response variable. The data on blood pressure increase follow.

Rabbit	Dose (j)						Rabbit	Dose (j)					
i	.1	.3	.5	1.0	1.5	3.0	i	.1	.3	.5	1.0	1.5	3.0
1	21	21	23	35	36	48	7	9	12	17	22	33	40
2	19	24	27	36	36	46	8	20	20	30	30	38	41
3	12	25	27	26	33	40	9	18	18	27	31	42	49
4	9	17	18	27	34	39	10	8	12	11	24	26	31
5	7	10	19	25	31	38	11	18	22	25	32	38	38
6	18	26	26	29	39	44	12	17	23	26	28	34	35

a. Obtain the residuals for repeated measures model (27.1) and plot them against the fitted values. Also prepare a normal probability plot of the residuals. What do you conclude about the appropriateness of model (27.1)?

b. Prepare aligned residual dot plots by dose level. Do these plots support the assumption of constancy of the error variance? Discuss.

c. Plot the observations Y_{ij} for each rabbit in the format of Figure 27.2. Does the assumption of no interactions between subjects (rabbits) and treatments appear to be reasonable here?

d. Conduct the Tukey test for additivity, conditional on the rabbits actually selected; use $\alpha = .005$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

27.4. Refer to **Blood pressure** Problem 27.3. Assume that repeated measures model (27.1) is appropriate.

a. Obtain the analysis of variance table.

b. Test whether or not the mean increase in blood pressure differs for the various dose levels; use $\alpha = .01$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

c. Analyze the effects of the six dose levels by comparing the means for successive dose levels using the Bonferroni procedure with a 95 percent family confidence coefficient. State your findings and summarize them by a suitable line plot.

d. According to the estimated efficiency measure (21.14), how effective was the repeated measures design here as compared to a completely randomized design?

27.5. Refer to **Blood pressure** Problems 27.3 and 27.4.

a. Develop a regression model in which the subject effects are represented by 1, −1, 0 indicator variables and the dose effect is represented by linear, quadratic, and cubic terms in $x = X - \bar{X}$, where X is the dose level. For instance, the x value for the first dose level ($X = .1$) is $x = .1 - 1.07 = -.97$.

b. Fit the regression model to the data.

c. Obtain the residuals and plot them against the fitted values. Does the model utilized appear to provide a reasonable fit?

d. Test whether or not the cubic effect is required in the model; use $\alpha = .05$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

27.6. **Grapefruit sales.** A supermarket chain studied the relationship between grapefruit sales and the price at which grapefruits are offered. Three price levels were studied: (1) the chief competitor's price, (2) a price slightly higher than the chief competitor's price, and (3) a price moderately higher than the chief competitor's price. Eight stores of comparable size were randomly selected for the study. Sales data were collected for three one-week periods, with the order of the three price levels randomly assigned for each store. The experiment was conducted during a time period when sales of grapefruits are usually quite stable, and no carryover effects were anticipated for this product. Data on store sales of grapefruits during the study period follow (data coded).

Store	Price level (j)		
i	1	2	3
1	62.1	61.3	60.8
2	58.2	57.9	55.1
...	...	...	...
7	46.8	43.2	41.5
8	51.2	49.8	47.9

a. Obtain the residuals for repeated measures model (27.1) and plot them against the fitted values. Also prepare a normal probability plot of the residuals. What do you conclude about the appropriateness of model (27.1)?

b. Prepare aligned residual dot plots by price level. Do these plots support the assumption of constancy of the error variance? Discuss.

c. Plot the observations Y_{ij} for each store in the format of Figure 27.2. Does the assumption of no interactions between subjects (stores) and treatments appear to be reasonable here?

d. Conduct the Tukey test for additivity, conditional on the stores actually selected; use $\alpha = .01$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

*27.7. Refer to **Grapefruit sales** Problem 27.6. Assume that repeated measures model (27.1) is appropriate.

a. Obtain the analysis of variance table.

b. Test whether or not the mean sales of grapefruits differ for the three price levels; use $\alpha = .05$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

c. Analyze the effects of the three price levels by estimating all pairwise comparisons of the price level means. Use the most efficient multiple comparison procedure with a 95 percent family confidence coefficient. State your findings and summarize them by a suitable line plot.

d. According to the estimated efficiency measure (21.14), how effective was the repeated measures design compared to a completely randomized design?

27.8. Refer to **Blood pressure** Problem 27.3. A consultant is concerned about the validity of the model assumptions and suggests that the study should be analyzed by means of the nonparametric rank F test. Rank the data within each rabbit and perform the rank F test; use $\alpha = .01$. State the alternatives, decision rule, and conclusion. Comment on the consultant's concern here.

*27.9. Refer to **Grapefruit sales** Problem 27.6. It has been suggested that the nonparametric rank F test should be used here. Rank the data within each store and perform the rank F test; use $\alpha = .05$. State the alternatives, decision rule, and conclusion. Is your conclusion the same as that obtained in Problem 27.7b?

27.10. **Truth in advertising.** A consumer research organization showed five different advertisements to 10 subjects and asked each to rank them in order of truthfulness. A rank of 1 denotes the most truthful. The results were:

Subject	Advertisement (j)				
i	A	B	C	D	E
1	3	1	2	5	4
2	4	2	1	3	5
3	4	2	3	1	5
4	3	1	2	5	4
5	4	1	2	5	3

Subject	Advertisement (j)				
i	A	B	C	D	E
6	4	2	1	3	5
7	4	1	2	3	5
8	5	1	3	2	4
9	4	2	3	1	5
10	5	1	2	3	4

a. Do the subjects perceive the five advertisements as having equal truthfulness? Conduct the nonparametric rank F test using level of significance $\alpha = .05$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

b. Use the multiple pairwise testing procedure (27.9) to group the five different advertisements according to mean perceived truthfulness; employ family significance level $\alpha = .10$. Summarize your findings.

c. Obtain the coefficient of concordance (27.10) and interpret this measure.

27.11. **Incentive stimulus.** Refer to the example in Section 27.3 about the effects of two types of incentives (factor A) on a person's ability to solve two types of problems (factor B);

the repeated measures design is illustrated in Figure 27.5. Twelve persons were randomly selected and assigned in equal numbers to the two incentive groups. The order of the two types of problems was then randomized independently for each person. The problem-solving ability scores follow (the higher the score, the greater the ability to solve problems).

Incentive Stimulus	Subject	Problem Type: Abstract ($k = 1$)	Problem Type: Concrete ($k = 2$)
$j = 1$	$i = 1$	10	18
	$i = 2$	14	19
	$i = 3$	17	18
	$i = 4$	8	12
	$i = 5$	12	14
	$i = 6$	15	20
$j = 2$	$i = 1$	16	35
	$i = 2$	19	32
	$i = 3$	22	37
	$i = 4$	20	33
	$i = 5$	24	39
	$i = 6$	21	32

a. Obtain the residuals for repeated measures model (27.11) and plot them against the fitted values. Also prepare a normal probability plot of the residuals. What do you conclude about the appropriateness of model (27.11)?

b. Plot the problem-solving ability scores by incentive stimulus and problem type, in the format of Figure 27.6. What do you conclude about the appropriateness of model (27.11)? Discuss.

12. Refer to **Incentive stimulus** Problem 27.11. Assume that repeated measures model (27.11) is appropriate.

a. Obtain the analysis of variance table.

b. Plot the data and the estimated treatment means in the format of Figure 27.12. Does it appear that interaction effects are present? That main effects are present?

c. Test whether or not the two factors interact; use $\alpha = .05$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

d. The following comparisons between problem types are of interest:

$$L_1 = \mu_{\cdot 11} - \mu_{\cdot 12} \qquad L_2 = \mu_{\cdot 21} - \mu_{\cdot 22}$$

Estimate these comparisons by means of confidence intervals. Use the Tukey procedure with a 90 percent family confidence coefficient for each problem type. Then combine these two Tukey procedures with a Bonferroni adjustment for each problem type. State your findings.

e. The following comparisons between incentive stimuli are of interest:

$$L_3 = \mu_{\cdot 11} - \mu_{\cdot 21} \qquad L_4 = \mu_{\cdot 12} - \mu_{\cdot 22}$$

Estimate these comparisons by means of confidence intervals. Use the Tukey procedure with a 90 percent family confidence coefficient for each incentive stimulus. Then combine these two Tukey procedures with a Bonferroni adjustment for each incentive stimulus. State your findings.

*27.13. **Store displays.** A repeated measures study was conducted to examine the effects of two different store displays for a household product (factor A) on sales in four successive time periods (factor B). Eight stores were randomly selected, and four were assigned at random to each display. The sales data (coded) follow.

Type of Display	Store	Time Period $k=1$	$k=2$	$k=3$	$k=4$
$j=1$	$i=1$	956	953	938	1,049
	$i=2$	1,008	1,032	1,025	1,123
	$i=3$	350	352	338	438
	$i=4$	412	449	385	532
$j=2$	$i=1$	769	766	739	859
	$i=2$	880	875	860	915
	$i=3$	176	185	168	280
	$i=4$	209	223	217	301

a. Obtain the residuals for repeated measures model (27.11) and plot them against the fitted values. Also prepare a normal probability plot of the residuals. What do you conclude about the appropriateness of model (27.11)?

b. Plot the sales data by type of display and time period, in the format of Figure 27.6. What do you conclude about the appropriateness of model (27.11)? Discuss.

*27.14. Refer to **Store displays** Problem 27.13. The experimenter wished to explore further the appropriateness of repeated measures model (27.11).

a. Conduct a formal test of the constancy of the between-subjects variances. Use (27.17) and perform the Hartley test, with $\alpha = .01$. State the alternatives, decision rule, and conclusion.

b. Decompose the error variation $SSB.S(A)$ into components using (27.18), and perform the Hartley test for the constancy of the error variance σ^2 for the different factor A levels; use $\alpha = .01$. State the alternatives, decision rule, and conclusion.

*27.15. Refer to **Store displays** Problem 27.13. Assume that repeated measures model (27.11) is appropriate.

a. Obtain the analysis of variance table.

b. Plot the data and the estimated treatment means in the format of Figure 27.12. Does it appear that interaction effects are present? That main effects are present?

c. Test whether or not the two factors interact; use $\alpha = .025$. State the alternatives, decision rule, and conclusion. What is the P-value for the test?

d. Test separately whether or not display and time main effects are present; use $\alpha = .025$ for each test. State the alternatives, decision rule, and conclusion for each test. What is the P-value for each test?

e. To study the nature of the factor A and factor B main effects, estimate the following pairwise comparisons:

$$L_1 = \mu_{\cdot 1 \cdot} - \mu_{\cdot 2 \cdot} \qquad L_3 = \mu_{\cdot \cdot 2} - \mu_{\cdot \cdot 3}$$
$$L_2 = \mu_{\cdot \cdot 1} - \mu_{\cdot \cdot 2} \qquad L_4 = \mu_{\cdot \cdot 3} - \mu_{\cdot \cdot 4}$$

Use the Bonferroni procedure with a 90 percent family confidence coefficient. State your findings.

27.16. **Calculator efficiency.** To test the efficiency of its new programmable calculator, a computer company selected at random six engineers who were proficient in the use of both this calculator and an earlier model and asked them to work out two problems on both calculators. One of the problems was statistical in nature, the other was an engineering problem. The order of the four calculations was randomized independently for each engineer. The length of time (in minutes) required to solve each problem was observed. The results follow (type of problem is factor A and calculator model is factor B):

		$j = 1$ Statistical Problem		$j = 2$ Engineering Problem	
Engineer i		$k = 1$ New Model	$k = 2$ Earlier Model	$k = 1$ New Model	$k = 2$ Earlier Model
1	Jones	3.1	7.5	2.5	5.1
2	Williams	3.8	8.1	2.8	5.3
3	Adams	3.0	7.6	2.0	4.9
4	Dixon	3.4	7.8	2.7	5.5
5	Erickson	3.3	6.9	2.5	5.4
6	Maynes	3.6	7.8	2.4	4.8

a. Obtain the residuals for repeated measures model (27.21) and plot them against the fitted values. Also prepare a normal probability plot of the residuals. What do you conclude about the appropriateness of model (27.21)?

b. Prepare aligned residual dot plots by treatment ignoring the factorial nature of the treatments. Do these plots support the assumption of constancy of the error variance? Discuss.

27.17. Refer to **Calculator efficiency** Problem 27.16. Assume that repeated measures model (27.21) is appropriate.

a. Obtain the analysis of variance table.

b. Plot the data and the estimated treatment means in the format of Figure 27.12. Does it appear that treatment interaction effects are present?

c. Test whether or not the two treatment factors interact; use $\alpha = .01$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

d. It is desired to study the nature of the interaction effects by considering the three comparisons:

$$L_1 = \mu_{\cdot 12} - \mu_{\cdot 11} \qquad L_3 = L_2 - L_1$$
$$L_2 = \mu_{\cdot 22} - \mu_{\cdot 21}$$

Obtain confidence intervals for these comparisons; use the Bonferroni procedure with a 95 percent family confidence coefficient. State your findings.

*27.18. **Migraine headaches.** Two experimental pain killer drugs for relief of migraine headaches were studied at a major medical center. Ten persistent migraine sufferers were randomly selected for a pilot study and received in random order each of the four treatment combinations, with a suitable interval between drug administrations. The decrease in pain intensity was used as the response variable. The four treatments used in the study are defined as follows: A_1B_1 = low dose of both drugs; A_1B_2 = low dose of drug A, high dose of drug B; A_2B_1 = high dose of drug A, low dose of drug B; A_2B_2 = high dose of both drugs. The data

on reduction in pain intensity follow (the higher the score, the greater the reduction in pain).

Person	A_1 ($j = 1$)		A_2 ($j = 2$)	
i	B_1 ($k = 1$)	B_2 ($k = 2$)	B_1 ($k = 1$)	B_2 ($k = 2$)
1	1.6	3.4	2.7	4.3
2	2.3	5.1	4.2	6.5
3	4.2	5.3	4.6	6.0
.		...	...	...
8	6.0	7.2	6.3	7.3
9	1.2	1.4	1.3	1.7
10	2.7	3.0	3.0	3.1

a. Obtain the residuals for repeated measures model (27.21) and plot them against the fitted values. Also prepare a normal probability plot of the residuals. What do you conclude about the appropriateness of model (27.21)?

b. Prepare aligned residual dot plots by treatment ignoring the factorial nature of the treatments. Do these plots support the assumption of constancy of the error variance? Discuss.

*27.19. Refer to **Migraine headaches** Problem 27.18. Assume that repeated measures model (27.21) is appropriate.

a. Obtain the analysis of variance table.

b. Plot the data and the estimated treatment means in the format of Figure 27.12. Does it appear that treatment interaction effects are present? That main effects are present?

c. Test whether or not the two treatment factors interact; use $\alpha = .005$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

d. Test separately whether or not factor A and factor B main effects are present; use $\alpha = .05$ for each test. State the alternatives, decision rule, and conclusion for each test. What is the P-value for each test?

e. Estimate the following comparisons by means of confidence intervals:

$$L_1 = \mu_{\cdot 21} - \mu_{\cdot 11} \qquad L_3 = \mu_{\cdot 21} - \mu_{\cdot 12}$$
$$L_2 = \mu_{\cdot 12} - \mu_{\cdot 11} \qquad L_4 = \mu_{\cdot 22} - \mu_{\cdot 11}$$

Use the Bonferroni procedure and family confidence coefficient .95. Summarize your findings.

27.20. **Wheat yield.** Refer to the split-plot agricultural experiment of Section 27.6, for which the layout is shown in Figure 27.13. The results of this experiment to investigate the effects of two irrigation methods (factor A) and two fertilizers (factor B) on wheat yield follow for the 10 fields used in the study.

Irrigation Method j:	1					2				
Field i:	1	2	3	4	5	1	2	3	4	5
Fertilizer $k = 1$:	43	40	31	27	36	63	52	45	47	54
$k = 2$:	48	43	36	30	39	70	53	48	51	57

a. Obtain the residuals for split-plot model (27.27) and plot them against the fitted values. Also prepare a normal probability plot of the residuals. What do you conclude about the appropriateness of model (27.27)?

b. Plot the wheat yield data by irrigation method and type of fertilizer in the format of Figure 27.6. What do you conclude about the appropriateness of model (27.27)? Discuss.

27.21. Refer to **Wheat yield** Problem 27.20. Assume that split-plot model (27.27) is appropriate.

a. Obtain the analysis of variance table.

b. Plot the data and the estimated treatment means in the format of Figure 27.12. Does it appear that interaction effects are present? That main effects are present?

c. Test whether or not the two factors interact; use $\alpha = .05$. State the alternatives, decision rule, and conclusion. What is the P-value for the test?

d. Test separately whether or not factor A and factor B main effects are present; use $\alpha = .05$. State the alternatives, decision rule, and conclusion for each test. What is the P-value for each test?

e. To study the nature of the factor A and factor B main effects, estimate the following pairwise comparisons:

$$L_1 = \mu_{\cdot 1 \cdot} - \mu_{\cdot 2 \cdot} \qquad L_2 = \mu_{\cdot \cdot 1} - \mu_{\cdot \cdot 2}$$

Use the Bonferroni procedure with a 90 percent family confidence coefficient. State your findings.

Exercise

27.22. Derive the total sum of squares breakdown in (27.5).

Projects

27.23. Refer to **Blood pressure** Problem 27.3. Obtain the estimated within-subjects variance-covariance matrix using (27.8). Are the estimated variances and covariances of the same orders of magnitude? Is the compound symmetry assumption reasonable here?

27.24. Refer to **Grapefruit sales** Problem 27.6. Obtain the estimated within-subjects variance-covariance matrix using (27.8). Are the variances and covariances roughly of the same order of magnitude? Is the compound symmetry assumption reasonably satisfied here?

27.25. Refer to the **Drug effect experiment** data set in Appendix C.12. Consider only Part I of the study and observation unit 1 for each drug dosage level; i.e., include only observations for which variable 2 equals 1 and variable 6 equals 1. Treat the 12 rats as subjects and ignore the classification of the rats into the three initial lever press rate groups. Assume that the subjects (rats) have random effects and that the treatments (dosage levels) have fixed effects.

a. State the additive repeated measures model for this study.

b. Obtain the residuals and plot them against the fitted values. Also prepare a normal probability plot of the residuals. What do you conclude about the appropriateness of the model employed?

c. Plot the responses for each rat in the format of Figure 27.2. Does the assumption of no interactions between subjects (rats) and treatments appear to be appropriate?

27.26. Refer to the **Drug effect experiment** data set in Appendix C.12 and Project 27.25.

a. Obtain the analysis of variance table.

b. Test whether or not the drug dosage level affects the mean lever press rate; use $\alpha = .05$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

c. Analyze the effects of the four dosage levels by comparing the mean responses for each pair of successive dosage levels; use the Bonferroni procedure with a 90 percent family confidence coefficient. State your findings.

d. Fit a regression model in which the subject effects are represented by $1, -1, 0$ indicator variables and the dosage effect is represented by linear and quadratic terms in $x = X - \bar{X}$, where X is the dosage level. Assume that there are no interactions between subjects and treatments.

e. Obtain the residuals and plot them against the fitted values. Does the regression model appear to provide a good fit? Discuss.

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

27.27. Refer to the **Drug effect experiment** data set in Appendix C.12. Consider the combined study. Assume that subjects (rats) and observation units have random effects, and that factor A (initial lever press rate), factor B (dosage level), and factor C (reinforcement schedule) have fixed effects. Also assume that there are no interactions between subjects and treatments.

a. Use rules (D.1) and (D.6) in Appendix D to develop the model for this experiment.

b. Fit the model in part (a), obtain the residuals, and plot them against the fitted values. Also prepare a normal probability plot of the residuals. What do you conclude about the appropriateness of your model?

27.28. Refer to the **Drug effect experiment** data set in Appendix C.12 and Project 27.27. Assume that the model in Project 27.27a is appropriate.

a. Use an appropriate statistical package to obtain the analysis of variance table and the expected mean squares.

b. Test whether or not ABC interactions are present; use $\alpha = .01$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

c. For each reinforcement schedule, plot the estimated treatment means against dosage level with different curves for the three initial lever press rate groups, in the format of Figure 24.5. Examine your plots for the nature of the interaction effects and report your findings.

27.29. Consider a repeated measures design study with $s = 3$ and $r = 3$, where each subject ranks all treatments (with no ties allowed).

a. Develop the exact sampling distribution of F_R^* when H_0 holds. [*Hint:* All ranking permutations for a subject are equally likely under H_0 and all subjects are assumed to act independently.]

b. How does the 90th percentile of the exact sampling distribution obtained in part (a) compare with $F(.90; 2, 4)$? What is the implication of this?