The estimate of $\sigma_\beta^2$ is found from the mean square for blocks adjusted for treatments. In general, for a balanced incomplete block design, this mean square is

$$MS_{\text{Blocks(adjusted)}} = \frac{\left( \frac{k \sum_{i=1}^{a} Q_i^2}{\lambda a} + \sum_{j=1}^{b} \frac{y_{.j}^2}{k} - \sum_{i=1}^{a} \frac{y_{i.}^2}{r} \right)}{(b-1)} \tag{4-46}$$

and its expected value [which is derived in Graybill (1961)] is

$$E[MS_{\text{Blocks(adjusted)}}] = \sigma^2 + \frac{a(r-1)}{b-1}\sigma_\beta^2$$

Thus, if $MS_{\text{Blocks(adjusted)}} > MS_E$, the estimate of $\hat{\sigma}_\beta^2$ is

$$\hat{\sigma}_\beta^2 = \frac{[MS_{\text{Blocks(adjusted)}} - MS_E](b-1)}{a(r-1)} \tag{4-47}$$

and if $MS_{\text{Blocks(adjusted)}} \leq MS_E$, we set $\hat{\sigma}_\beta^2 = 0$. This results in the combined estimator

$$\tau_i^* = \begin{cases} \dfrac{kQ_i(\hat{\sigma}^2 + k\hat{\sigma}_\beta^2) + \left( \sum_{j=1}^{b} n_{ij} y_{.j} - kr\bar{y}_{..} \right)\hat{\sigma}^2}{(r-\lambda)\hat{\sigma}^2 + \lambda a(\hat{\sigma}^2 + k\hat{\sigma}_\beta^2)}, & \hat{\sigma}_\beta^2 > 0 \quad \text{(4-48a)} \\ \dfrac{y_{i.} - (1/a)y_{..}}{r}, & \hat{\sigma}_\beta^2 = 0 \quad \text{(4-48b)} \end{cases}$$

We now compute the combined estimates for the data in Example 4-5. From Table 4-24 we obtain $\hat{\sigma}^2 = MS_E = 0.65$ and $MS_{\text{Blocks(adjusted)}} = 22.03$. (Note that in computing $MS_{\text{Blocks(adjusted)}}$ we make use of the fact that this is a symmetric design. In general, we must use Equation 4-46.) Because $MS_{\text{Blocks(adjusted)}} > MS_E$, we use Equation 4-47 to estimate $\sigma_\beta^2$ as

$$\hat{\sigma}_\beta^2 = \frac{(22.03 - 0.65)(3)}{4(3-1)} = 8.02$$

Therefore, we may substitute $\hat{\sigma}^2 = 0.65$ and $\hat{\sigma}_\beta^2 = 8.02$ into Equation 4-48a to obtain the combined estimates listed below. For convenience, the intrablock and interblock estimates are also given. In this example, the combined estimates are close to the intrablock estimates because the variance of the interblock estimates is relatively large.

| Parameter | Intrablock Estimate | Interblock Estimate | Combined Estimate |
|---|---|---|---|
| $\tau_1$ | −1.12 | 10.50 | −1.09 |
| $\tau_2$ | −0.88 | −3.50 | −0.88 |
| $\tau_3$ | −0.50 | −0.50 | −0.50 |
| $\tau_4$ | 2.50 | −6.50 | 2.47 |

## 4-5 PROBLEMS

4-1. A chemist wishes to test the effect of four chemical agents on the strength of a particular type of cloth. Because there might be variability from one bolt to another, the chemist decides to use a randomized block design, with the bolts of cloth considered as blocks. She selects five bolts and applies all four chemicals in random order to each bolt. The resulting tensile strengths follow. Analyze the data from this experiment (use $\alpha = 0.05$) and draw appropriate conclusions.

| | Bolt | | | | |
|---|---|---|---|---|---|
| Chemical | 1 | 2 | 3 | 4 | 5 |
| 1 | 73 | 68 | 74 | 71 | 67 |
| 2 | 73 | 67 | 75 | 72 | 70 |
| 3 | 75 | 68 | 78 | 73 | 68 |
| 4 | 73 | 71 | 75 | 75 | 69 |

4-2. Three different washing solutions are being compared to study their effectiveness in retarding bacteria growth in 5-gallon milk containers. The analysis is done in a laboratory, and only three trials can be run on any day. Because days could represent a potential source of variability, the experimenter decides to use a randomized block design. Observations are taken for four days, and the data are shown here. Analyze the data from this experiment (use $\alpha = 0.05$) and draw conclusions.

| | Days | | | |
|---|---|---|---|---|
| Solution | 1 | 2 | 3 | 4 |
| 1 | 13 | 22 | 18 | 39 |
| 2 | 16 | 24 | 17 | 44 |
| 3 | 5 | 4 | 1 | 22 |

4-3. Plot the mean tensile strengths observed for each chemical type in Problem 4-1 and compare them to an appropriately scaled $t$ distribution. What conclusions would you draw from this display?

4-4. Plot the average bacteria counts for each solution in Problem 4-2 and compare them to a scaled $t$ distribution. What conclusions can you draw?

4-5. Consider the hardness testing experiment described in Section 4-1. Suppose that the experiment was conducted as described and that the following Rockwell C-scale data (coded by subtracting 40 units) obtained:

| | Coupon | | | |
|---|---|---|---|---|
| Tip | 1 | 2 | 3 | 4 |
| 1 | 9.3 | 9.4 | 9.6 | 10.0 |
| 2 | 9.4 | 9.3 | 9.8 | 9.9 |
| 3 | 9.2 | 9.4 | 9.5 | 9.7 |
| 4 | 9.7 | 9.6 | 10.0 | 10.2 |

(a) Analyze the data from this experiment.

(b) Use the Fisher LSD method to make comparisons among the four tips to determine specifically which tips differ in mean hardness readings.

(c) Analyze the residuals from this experiment.

4-6. A consumer products company relies on direct mail marketing pieces as a major component of its advertising campaigns. The company has three different designs for a new brochure and wants to evaluate their effectiveness, as there are substantial differences in costs between the three designs. The company decides to test the three designs by mailing 5000 samples of each to potential customers in four different regions of the country. Since there are known regional differences in the customer base, regions are considered as blocks. The number of responses to each mailing is as follows.

| | Region | | | |
|---|---|---|---|---|
| Design | NE | NW | SE | SW |
| 1 | 250 | 350 | 219 | 375 |
| 2 | 400 | 525 | 390 | 580 |
| 3 | 275 | 340 | 200 | 310 |

(a) Analyze the data from this experiment.

(b) Use the Fisher LSD method to make comparisons among the three designs to determine specifically which designs differ in the mean response rate.

(c) Analyze the residuals from this experiment.

4-7. The effect of three different lubricating oils on fuel economy in diesel truck engines is being studied. Fuel economy is measured using brake-specific fuel consumption after the engine has been running for 15 minutes. Five different truck engines are available for the study, and the experimenters conduct the following randomized complete block design.

| | Truck | | | | |
|---|---|---|---|---|---|
| Oil | 1 | 2 | 3 | 4 | 5 |
| 1 | 0.500 | 0.634 | 0.487 | 0.329 | 0.512 |
| 2 | 0.535 | 0.675 | 0.520 | 0.435 | 0.540 |
| 3 | 0.513 | 0.595 | 0.488 | 0.400 | 0.510 |

4-19. An industrial engineer is investigating the effect of four assembly methods (*A, B, C, D*) on the assembly time for a color television component. Four operators are selected for the study. Furthermore, the engineer knows that each assembly method produces such fatigue that the time required for the last assembly may be greater than the time required for the first, regardless of the method. That is, a trend develops in the required assembly time. To account for this source of variability, the engineer uses the Latin square design shown below. Analyze the data from this experiment ($\alpha = 0.05$) and draw appropriate conclusions.

| Order of Assembly | Operator 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1 | $C = 10$ | $D = 14$ | $A = 7$ | $B = 8$ |
| 2 | $B = 7$ | $C = 18$ | $D = 11$ | $A = 8$ |
| 3 | $A = 5$ | $B = 10$ | $C = 11$ | $D = 9$ |
| 4 | $D = 10$ | $A = 10$ | $B = 12$ | $C = 14$ |

4-20. Suppose that in Problem 4-18 the observation from batch 3 on day 4 is missing. Estimate the missing value from Equation 4-24, and perform the analysis using the value.

4-21. Consider a $p \times p$ Latin square with rows ($\alpha_i$), columns ($\beta_k$), and treatments ($\tau_j$) fixed. Obtain least squares estimates of the model parameters $\alpha_i$, $\beta_k$, and $\tau_j$.

4-22. Derive the missing value formula (Equation 4-24) for the Latin square design.

4-23. ***Designs involving several Latin squares.*** [See Cochran and Cox (1957), John (1971).] The $p \times p$ Latin square contains only $p$ observations for each treatment. To obtain more replications the experimenter may use several squares, say $n$. It is immaterial whether the squares used are the same or different. The appropriate model is

$$y_{ijkh} = \begin{matrix} \mu + \rho_h + \alpha_{i(h)} \\ + \tau_j + \beta_{k(h)} \\ + (\tau\rho)_{jh} + \epsilon_{ijkh} \end{matrix} \begin{cases} i = 1,2,\ldots,p \\ j = 1,2,\ldots,p \\ k = 1,2,\ldots,p \\ h = 1,2,\ldots,n \end{cases}$$

where $y_{ijkh}$ is the observation on treatment $j$ in row $i$ and column $k$ of the $h$th square. Note that $\alpha_{i(h)}$ and $\beta_{k(h)}$ are the row and column effects in the $h$th square, $\rho_h$ is the effect of the $h$th square, and $(\tau\rho)_{jh}$ is the interaction between treatments and squares.

(a) Set up the normal equations for this model, and solve for estimates of the model parameters. Assume that appropriate side conditions on the parameters are $\Sigma_h \hat{\rho}_h = 0$, $\Sigma_i \hat{\alpha}_{i(h)} = 0$, and $\Sigma_k \hat{\beta}_{k(h)} = 0$ for each $h$, $\Sigma_j \hat{\tau}_j = 0$, $\Sigma_j (\widehat{\tau\rho})_{jh} = 0$ for each $h$, and $\Sigma_h (\widehat{\tau\rho})_{jh} = 0$ for each $j$.

(b) Write down the analysis of variance table for this design.

4-24. Discuss how the operating characteristics curves in the Appendix may be used with the Latin square design.

4-25. Suppose that in Problem 4-18 the data taken on day 5 were incorrectly analyzed and had to be discarded. Develop an appropriate analysis for the remaining data.

4-26. The yield of a chemical process was measured using five batches of raw material, five acid concentrations, five standing times (*A, B, C, D, E*), and five catalyst concentrations ($\alpha, \beta, \gamma, \delta, \epsilon$). The Graeco-Latin square that follows was used. Analyze the data from this experiment (use $\alpha = 0.05$) and draw conclusions.

| Batch | Acid Concentration 1 | 2 | 3 |
|---|---|---|---|
| 1 | $A\alpha = 26$ | $B\beta = 16$ | $C\gamma = 19$ |
| 2 | $B\gamma = 18$ | $C\delta = 21$ | $D\epsilon = 18$ |
| 3 | $C\epsilon = 20$ | $D\alpha = 12$ | $E\beta = 16$ |
| 4 | $D\beta = 15$ | $E\gamma = 15$ | $A\delta = 22$ |
| 5 | $E\delta = 10$ | $A\epsilon = 24$ | $B\alpha = 17$ |

| Batch | Acid Concentration 4 | 5 |
|---|---|---|
| 1 | $D\delta = 16$ | $E\epsilon = 13$ |
| 2 | $E\alpha = 11$ | $A\beta = 21$ |
| 3 | $A\gamma = 25$ | $B\delta = 13$ |
| 4 | $B\epsilon = 14$ | $C\alpha = 17$ |
| 5 | $C\beta = 17$ | $D\gamma = 14$ |

4-27. Suppose that in Problem 4-19 the engineer suspects that the workplaces used by the four operators may represent an additional source of variation. A fourth factor, workplace ($\alpha, \beta, \gamma, \delta$) may be introduced and another experiment conducted, yielding the Graeco-Latin square that follows. Analyze the data from this experiment (use $\alpha = 0.05$) and draw conclusions.

| Order of Assembly | Operator 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1 | $C\beta = 11$ | $B\gamma = 10$ | $D\delta = 14$ | $A\alpha = 8$ |
| 2 | $B\alpha = 8$ | $C\delta = 12$ | $A\gamma = 10$ | $D\beta = 12$ |
| 3 | $A\delta = 9$ | $D\alpha = 11$ | $B\beta = 7$ | $C\gamma = 15$ |
| 4 | $D\gamma = 9$ | $A\beta = 8$ | $C\alpha = 18$ | $B\delta = 6$ |

4-28. Construct a $5 \times 5$ hypersquare for studying the effects of five factors. Exhibit the analysis of variance table for this design.

4-29. Consider the data in Problems 4-19 and 4-27. Suppressing the Greek letters in 4-27, analyze the data using the method developed in Problem 4-23.

4-30. Consider the randomized block design with one missing value in Problem 4-15. Analyze this data by using the exact analysis of the missing value problem discussed in Section 4-1.4. Compare your results to the approximate analysis of these data given from Problem 4-15.

4-31. An engineer is studying the mileage performance characteristics of five types of gasoline additives. In the road test he wishes to use cars as blocks; however, because of a time constraint, he must use an incomplete block design. He runs the balanced design with the five blocks that follow. Analyze the data from this experiment (use $\alpha = 0.05$) and draw conclusions.

| Additive | Car 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| 1 | | 17 | 14 | 13 | 12 |
| 2 | 14 | 14 | | 13 | 10 |
| 3 | 12 | | 13 | 12 | 9 |
| 4 | 13 | 11 | 11 | 12 | |
| 5 | 11 | 12 | 10 | | 8 |

4-32. Construct a set of orthogonal contrasts for the data in Problem 4-31. Compute the sum of squares for each contrast.

4-33. Seven different hardwood concentrations are being studied to determine their effect on the strength of the paper produced. However, the pilot plant can only produce three runs each day. As days may differ, the analyst uses the balanced incomplete block design that follows. Analyze the data from this experiment (use $\alpha = 0.05$) and draw conclusions.

| Hardwood Concentration (%) | Days 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 2 | 114 | | | |
| 4 | 126 | 120 | | |
| 6 | | 137 | 117 | |
| 8 | 141 | | 129 | 149 |
| 10 | | 145 | | 150 |
| 12 | | | 120 | |
| 14 | | | | 136 |

| Hardwood Concentration (%) | Days 5 | 6 | 7 |
|---|---|---|---|
| 2 | 120 | | 117 |
| 4 | | 119 | |
| 6 | | | 134 |
| 8 | | | |
| 10 | 143 | | |
| 12 | 118 | 123 | |
| 14 | | 130 | 127 |

4-34. Analyze the data in Example 4-5 using the general regression significance test.

4-35. Prove that $k\sum_{i=1}^{a} Q_i^2/(\lambda a)$ is the adjusted sum of squares for treatments in a BIBD.

4-36. An experimenter wishes to compare four treatments in blocks of two runs. Find a BIBD for this experiment with six blocks.

4-37. An experimenter wishes to compare eight treatments in blocks of four runs. Find a BIBD with 14 blocks and $\lambda = 3$.

4-38. Perform the interblock analysis for the design in Problem 4-31.

4-39. Perform the interblock analysis for the design in Problem 4-33.

4-40. Verify that a BIBD with the parameters $a = 8$, $r = 8$, $k = 4$, and $b = 16$ does not exist.

4-41. Show that the variance of the intrablock estimators $\{\hat{\tau}_i\}$ is $k(a - 1)\sigma^2/(\lambda a^2)$.

4-42. ***Extended incomplete block designs.*** Occasionally, the block size obeys the relationship $a < k < 2a$. An extended incomplete block design consists of a single replicate of each treatment in each block along with an incomplete block design with $k^* = k - a$. In the balanced case, the incomplete block design will have parameters $k^* = k - a$, $r^* = r - b$, and $\lambda^*$. Write out the statistical analysis. (*Hint:* In the extended incomplete block design, we have $\lambda = 2r - b + \lambda^*$.)

Table 5-22 Analysis of Variance for the Radar Detection Experiment Run as a 3 × 2 Factorial in a Latin Square

| Source of Variation | Sum of Squares | Degrees of Freedom | General Formula for Degrees of Freedom | Mean Square | $F_0$ | $P$-Value |
|---|---|---|---|---|---|---|
| Ground clutter, $G$ | 571.50 | 2 | $a - 1$ | 285.75 | 28.86 | <0.0001 |
| Filter type, $F$ | 1469.44 | 1 | $b - 1$ | 1469.44 | 148.43 | <0.0001 |
| $GF$ | 126.73 | 2 | $(a - 1)(b - 1)$ | 63.37 | 6.40 | 0.0071 |
| Days (rows) | 4.33 | 5 | $ab - 1$ | 0.87 | | |
| Operators (columns) | 428.00 | 5 | $ab - 1$ | 85.60 | | |
| Error | 198.00 | 20 | $(ab - 1)(ab - 2)$ | 9.90 | | |
| Total | 2798.00 | 35 | $(ab)^2 - 1$ | | | |

The ANOVA is summarized in Table 5-22. We have added a column to this table indicating how the number of degrees of freedom for each sum of squares is determined.

## 5-7 PROBLEMS

5-1. The yield of a chemical process is being studied. The two most important variables are thought to be the pressure and the temperature. Three levels of each factor are selected, and a factorial experiment with two replicates is performed. The yield data follow:

| Temperature (°C) | Pressure (psig) 200 | 215 | 230 |
|---|---|---|---|
| 150 | 90.4 | 90.7 | 90.2 |
| | 90.2 | 90.6 | 90.4 |
| 160 | 90.1 | 90.5 | 89.9 |
| | 90.3 | 90.6 | 90.1 |
| 170 | 90.5 | 90.8 | 90.4 |
| | 90.7 | 90.9 | 90.1 |

(a) Analyze the data and draw conclusions. Use $\alpha = 0.05$.

(b) Prepare appropriate residual plots and comment on the model's adequacy.

(c) Under what conditions would you operate this process?

5-2. An engineer suspects that the surface finish of a metal part is influenced by the feed rate and the depth of cut. He selects three feed rates and four depths of cut. He then conducts a factorial experiment and obtains the following data:

| Feed Rate (in/min) | Depth of Cut (in) 0.15 | 0.18 | 0.20 | 0.25 |
|---|---|---|---|---|
| | 74 | 79 | 82 | 99 |
| 0.20 | 64 | 68 | 88 | 104 |
| | 60 | 73 | 92 | 96 |
| | 92 | 98 | 99 | 104 |
| 0.25 | 86 | 104 | 108 | 110 |
| | 88 | 88 | 95 | 99 |
| | 99 | 104 | 108 | 114 |
| 0.30 | 98 | 99 | 110 | 111 |
| | 102 | 95 | 99 | 107 |

(a) Analyze the data and draw conclusions. Use $\alpha = 0.05$.

(b) Prepare appropriate residual plots and comment on the model's adequacy.

(c) Obtain point estimates of the mean surface finish at each feed rate.

(d) Find the $P$-values for the tests in part (a).

5-3. For the data in Problem 5-2, compute a 95 percent confidence interval estimate of the mean difference in response for feed rates of 0.20 and 0.25 in/min.

5-4. An article in *Industrial Quality Control* (1956, pp. 5–8) describes an experiment to investigate the effect of the type of glass and the type of phosphor on the

5-26. A manufacturer of laundry products is investigating the performance of a newly formulated stain remover. The new formulation is compared to the original formulation with respect to its ability to remove a standard tomato-like stain in a test article of cotton cloth using a factorial experiment. The other factors in the experiment are the number of times the test article is washed (1 or 2), and whether or not a detergent booster is used. The response variable is the stain shade after washing (12 is the darkest, 0 is the lightest). The data are shown in the following table.

| | Number of Washings | | Number of Washings | |
|---|---|---|---|---|
| | 1 | | 2 | |
| **Formulation** | **Booster** | | **Booster** | |
| | Yes | No | Yes | No |
| New | 6, 5 | 6, 5 | 3, 2 | 4, 1 |
| Original | 10, 9 | 11, 11 | 10, 9 | 9, 10 |

(a) Conduct an analysis of variance. Using $\alpha = 0.05$, what conclusions can you draw?

(b) Investigate model adequacy by plotting the residuals.