INSY 7300-6 Reference: Chapter 4 of Montgomery's 8th Ed.

By S. Maghsoodloo

THE RANDOMIZED COMPLETE BLOCK DESIGN (RCBD)

Consider an experiment where it is desired to determine if there are significant differences among 3 computer languages (Java, VB, and C^{++}) in coding a complex problem. Obviously we need human subjects to design such an experiment so that the 3 coding methods can be analyzed for efficiency. Suppose we select 10 programmers at random who are equally competent at all 3 languages to design 3 programs each, and as the response variable we measure the length of time that each programmer takes to complete a set of workable codes. Note that each programmer forms a Block and acts as his/her own control, and the experiment would be an Incomplete Block Design only if some of the programmers could not code in at least one of the 3 languages. The data are provided in Table 2 below.

(BLK = Block = Programmer)

Table 2. (The layout for a RCBD experiment with 10 programmers and 3 treatments)

	BLOCK 1	2	3	4	5	6	7	8	9	10	$y_{i.}$
Java	2.5 Hrs	3.2	3.1	3.7	2.4	2.1	3.3	3.5	1.9	5.3	31.0
VB	3.1	2.5	2.8	4.7	2.9	2.9	4.4	4.4	2.5	4.6	34.8
C^{++}	2.2	3.7	2.7	3.9	2.4	1.7	2.6	3.0	1.8	4.1	28.1
$y_{.j}$	7.8 Hrs.	9.4	8.6	12.3	7.7	6.7	10.3	10.9	6.2	14.0	$y_{..}$ = 93.9

Before developing the LAM (linear additive model) and the corresponding E(SS) for the model, we first obtain the ANOVA Table and draw appropriate conclusions. Note that randomization is performed within each block (i.e., each programmer is assigned one of the 3 languages at a random sequence), and therefore, blocks form a restriction on randomization. In essence, blocks are not to be thought of as a factor (i.e., the objective is not to determine if there are significant differences amongst blocks).

As before, the CF = $93.9^2/30$ = 293.907 (with 1 *df*); USS = $2.5^2 + 3.1^2 + 2.2^2 ... + 4.1^2$ =

319.030 (with 30 *df*) $\rightarrow$ SS_T = SS(Total) = 25.123 (with 29 *df*). The treatment SS is computed using the subtotals pertaining to the 3 languages, i.e., SS(Treatments) = $(31^2 + 34.8^2 + 28.1^2)/10$ – CF = 2.258 (with 2 *df*); similarly SS(Blocks) = $(7.8^2 + 9.4^2 + ... + 14.0^2)/3$ – CF =18.749667 (with 9 *df*), and thus SS(Residuals) = SS(Unexplained) = 4.115333 (with 18 *df*). The orthogonal partition of Total SS is given in ANOVA Table 3, where the *P-value* = $\Pr(F_{2,18} \geq 4.93812) = 0.01951$, which implies that we can reject H_0 at any LOS α larger or equal to 1.952%.

Table 3. (The ANOVA for the data of Table 2)

Source of Variation	DF	SS	MS	F_0	$F_{0.05,2,18}$ The threshold, or Critical value	*P-value*
Total	29	25.123			of F-statistic	
Languages Programmers	2 9	2.258 18.749667	1.129	4.93812 Because F_0 exceeds 3.5546,	3.5546	0.01951
Residuals	18	4.115333	0.22863	then reject H_0		

The above ANOVA Table clearly indicates that there are significant differences among the 3 programming languages (the Treatments) at the LOS 0.05 because the *P-value* = 0.01951 < 0.05, i.e., we must reject the null hypothesis H_0: $\tau_1 = \tau_2 = \tau_3 = 0$ at any level of significance set larger than 0.01951.

Exercise 7. Analyze the above experiment as a completely randomized design (CRD), i.e., the 30 experiments were run completely in random order. State how many subjects are needed for the corresponding CRD to complete the entire experiment. ANS: $F_0 = 1.3332$.

We now develop the LAM for a RCBD. As before we start with an identity!

$$y_{ij} \equiv y_{ij} + \mu - \mu + \mu_{i.} - \mu_{i.} + \mu_{.j} - \mu_{.j}$$

$$\equiv \mu + (\mu_{i.} - \mu) + (\mu_{.j} - \mu) + (y_{ij} - \mu_{i.} - \mu_{.j} + \mu) \equiv \mu + \tau_i + \beta_j + \in_{ij} \quad (5)$$

The term β_j is called the effect of the j^{th} block, and the definitions of the other 3 terms on

the RHS of Eq. (5) should be self-explanatory! Again, we replace the model parameters in Eq. (5)

by their corresponding point unbiased estimators in order to obtain:

$$y_{ij} \equiv \bar{y}_{..} + (\bar{y}_{i.} - \bar{y}_{..}) + (\bar{y}_{.j} - \bar{y}_{..}) + (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..}) \quad \rightarrow$$

$$y_{ij} - \bar{y}_{..} \equiv (\bar{y}_{i.} - \bar{y}_{..}) + (\bar{y}_{.j} - \bar{y}_{..}) + (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..}) \qquad (6)$$

Or: $$y_{ij} - \bar{y}_{..} \equiv \hat{\tau}_i + \hat{\beta}_j + e_{ij}$$

We 1st square both sides of Eq. (6) and then sum from i = 1 to "*a*" (the number of treatments), and also sum from j =1 to "*b*" (the no. of blocks); this leads to

$$\sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij} - \bar{y}_{..})^2 \equiv b\sum_{i=1}^{a}(\bar{y}_{i.} - \bar{y}_{..})^2 + a\sum_{j=1}^{b}(\bar{y}_{.j} - \bar{y}_{..})^2 + \sum\sum(y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2$$

Or: $$SS_T = SS_{Total} = b\sum_{i=1}^{a}\hat{\tau}_i^2 + a\sum_{j=1}^{b}\hat{\beta}_j^2 + \sum\sum e_{ij}^2 \qquad (7)$$

Exercise 8. Show that Eq. (7) is indeed an identity because the double sums of the 3 cross-product terms on the RHS vanish. As a 5-point bonus problem, show that $V(e_{ij}) = V(y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..}) = (N - a - b + 1)\sigma_{\epsilon}^2/N$.

Exercise 9. For the mixed-effects (i.e., fixed treatments but random blocks), show that

$E(SS_{Treatments}) = \sigma^2(a-1) + b\sum_{i=1}^{a}\tau_i^2$, $E(SS_{Blocks}) = (b-1)(\sigma^2 + a\sigma_\beta^2)$, and $E(SS_E) = (a-1)(b-1)\sigma^2$, where σ_β^2 represents the variance of the random blocks.

It should be quite clear that identity (7) is fundamental to the analysis of RCBD experiments because it breaks down the Total SS on the LHS into 3 orthogonal (i.e., additive) components on the RHS of Eq. (7): the 1st SS on the RHS is due to Treatments, the 2nd is due to Blocks, and the 3rd SS is due to Residuals. The *df* of Treatments is (a –1), that of Blocks is (b –1), and hence the *df* of residuals is (ab –1) – (a –1) – (b –1) = (a –1)(b –1), where $ab = N$ = total number of experimental runs. It must be noted that the REPEATED-MEASURES DESIGN (RMSD) explained on pp. 677-679 of Montgomery(8e) is simply a RCBD where blocks are human subjects just like the above example with experimental layout in Table 2. Table 15.26 on page 679 of Montgomery 1st breaks down the Total SS (with an –1 *df*) into 2 orthogonal components- one between blocks (with n –1 *df*) and the other within subjects with n(a–1) *df*. Then, the within subjects SS is orthogonally broken down into two further additive components: $SS_{Treatments}$ with (a – 1) *df* and $SS_{Residuals}$ with (a – 1)(n – 1) *df*.

Exercise 10. Work problems 4.3 and 4.21; 4.21 on p.179 involves a RMSD (Repeated-Measures).

ANALYSIS OF a RCBD EXPERIMENT WITH a FEW MISSING VALUES

Montgomery(8e) proposes 2 different procedures for analyses: the approximate method, and the exact method using the GRST approach. We first discuss the exact method. To this end, consider the coded (by 85) data of Table 4.7 on page 154 of Montgomery's 8th ed., where the value of y_{24} is missing so that there are only N = 23 (instead of 24) experimental runs and as a result Treatments (Extrusion Pressure) and Blocks (Batches of Resin) are no longer orthogonal. For your convenience I am reproducing Table 4.7 at the bottom of page 154 of Montgomery(8e) below.

Table 4.7 (The coded by 85 data for Table 4.7 on page 154 of Montgomery's 8th Ed.)

Extrusion Pressure	BLK 1	2	3	4	5	6	Coded $y_{i.}$
8500 PSI	5.3	4.2	13.2	8.9	2.4	12.9	46.90
8700	7.5	4.5	5.6	x	2.0	10.8	30.40
8900	0.5	5.8	4.6	1.2	3.0	8.4	23.50
9100	–2.5	4.5	0.6	2.4	– 6.1	5.7	4.60
Coded $y_{.j}$	10.8	19.0	24.0	12.5	1.30	37.8	$y_{..}$ = 105.4

The non-orthogonality (or obliqueness) of the above design is due to the fact that SS(Total) = 938.220 – $105.40^2/23$ = 455.21304, SS(Treatments) = 163.995, SS_{BLKS} = 190.1189, and SS(Residuals) can be computed from equation (7) as follows, where in the double sum below j ≠4 when i = 2: $SS_{Residuals}$ =

$$\sum_{i=1}^{4}\sum_{j=1}^{6}(y_{ij}-\bar{y}_{i.}-\bar{y}_{.j}+\bar{y}_{..})^2 = (5.3 - 7.8167 - 2.70 + 4.5826)^2 + \ldots + (5.7 - 0.767 - 9.450 + 4.5826)^2 =$$

102.2929 $\rightarrow$ SS(Treatments) + SS_{BLKS} + SS_{RES} = 456.4068 > SS(Total) = SS_T = 455.21304. Therefore, Treatments and Blocks are no longer orthogonal (i.e., additive), and we must resort to other means to compute the three SS's. Note that the above formula for computing $SS_{Residuals}$ is no longer accurate because in the unbalanced case

$$SS_{RES} = \sum_{i=1}^{4}\sum_{j=1}^{6}(y_{ij}-\hat{y}_{ij})^2 = \sum_{i=1}^{4}\sum_{j=1}^{6}[y_{ij}-(\hat{\mu}+\hat{\tau}_i+\hat{\beta}_j)]^2,$$ where $\hat{\mu}$, $\hat{\tau}_i$ & $\hat{\beta}_j$ are obtained as follows.

We now apply the GRST to the model : $y_{ij} = \mu + \tau_i + \beta_j + \in_{ij}$. Only for the sake of convenience, we will not put hats on the estimators.

μ : $23\mu + 6\tau_1 + 5\tau_2 + 6\tau_3 + 6\tau_4 + 4\beta_1 + 4\beta_2 + 4\beta_3 + 3\beta_4 + 4\beta_5 + 4\beta_6 = y_{..} = 105.40$

τ_1 : $6\mu + 6\tau_1 + \beta_1 + \beta_2 + \beta_3 + \beta_4 + \beta_5 + \beta_6 = y_{1.} = 46.90$

τ_2 : $5\mu + 5\tau_2 + \beta_1 + \beta_2 + \beta_3 + \beta_5 + \beta_6 = y_{2.} = 30.40$

τ_3 : $6\mu + 6\tau_3 + \beta_1 + \beta_2 + \beta_3 + \beta_4 + \beta_5 + \beta_6 = y_{3.} = 23.50$

Note that the above 4 equations are independent. Next we write 5 LS Normal equations for β_j's (the 6th will not be independent of other five and the equation for μ):

β_1 : $4\mu + \tau_1 + \tau_2 + \tau_3 + \tau_4 + 4\beta_1 = y_{.1} = 10.80$

β_2 : $4\mu + \tau_1 + \tau_2 + \tau_3 + \tau_4 + 4\beta_2 = y_{.2} = 19.00$

β_3 : $4\mu + \tau_1 + \tau_2 + \tau_3 + \tau_4 + 4\beta_3 = y_{.3} = 24.00$

β_4 : $3\mu + \tau_1 + \tau_3 + \tau_4 + 3\beta_4 = y_{.4} = 12.50$

β_5 : $4\mu + \tau_1 + \tau_2 + \tau_3 + \tau_4 + 4\beta_5 = y_{.5} = 1.30$

The above heterogeneous system of 9 LSNEs have 11 unknowns; thus we must impose 2 constraint equations in order to obtain unique solutions.

Constraint 1: $\tau_1 + \tau_2 + \tau_3 + \tau_4 = 0$

Constraint 2: $\beta_1 + \beta_2 + \beta_3 + \beta_4 + \beta_5 + \beta_6 = 0.$

The solution to the above system of 11 equations with 11 unknowns can easily be obtained using Matlab or MS Excel . I used Matlab and obtained the following solutions to 5 decimals:

$\hat{\mu}$ = 4.64500, $\hat{\tau}_1$ = 3.17167, $\hat{\tau}_2$ = 1.43500, $\hat{\tau}_3$ = −0.72833, $\hat{\tau}_4$ = −3.87833 , $\hat{\beta}_1$ = −1.94500, $\hat{\beta}_2$ = 0.10500, $\hat{\beta}_3$ = 1.35500, $\hat{\beta}_4$ = 0.00000, $\hat{\beta}_5$ = − 4.32000 and $\hat{\beta}_6$ = 4.80500. These unbiased LS estimates lead to $R(\mu, \tau, \beta) = \hat{\mu}\, y_{..} + \sum_{i=1}^{4} \hat{\tau}_i\, y_{i.} + \sum_{j=1}^{6} \hat{\beta}_j\, y_{.j}$ = 836.5240. Hence, SS(RES) = USS − $R(\mu, \tau, \beta)$ = 938.220 − 836.5240 = 101.6960.

In order to test the null hypothesis $H_0 : \tau_1 = \tau_2 = \tau_3 = \tau_4 = 0$, completely disregard the 3 LSNEs for τ_i's and put $\tau_1 = \tau_2 = \tau_3 = \tau_4 = 0$ in the remaining 6 LSNEs:

$$
\begin{aligned}
\mu &: \quad 23\mu + 4\beta_1 + 4\beta_2 + 4\beta_3 + 3\beta_4 + 4\beta_5 + 4\beta_6 = y_{..} = 105.40 \\
\beta_1 &: \quad 4\mu + 4\beta_1 = y_{.1} = 10.80 \\
\beta_2 &: \quad 4\mu + 4\beta_2 = y_{.2} = 19.00 \\
\beta_3 &: \quad 4\mu + 4\beta_3 = y_{.3} = 24.00 \\
\beta_4 &: \quad 3\mu + 3\beta_4 = y_{.4} = 12.50 \\
\beta_5 &: \quad 4\mu + 4\beta_5 = y_{.5} = 1.30
\end{aligned}
$$

Because there are 7 unknowns in the above system, we need one more independent equation to obtain unique solutions. Thus, we impose the constraint:

$$\beta_1 + \beta_2 + \beta_3 + \beta_4 + \beta_5 + \beta_6 = 0.$$

The solution to this last system of normal equations is $\hat{\mu}$ = 4.56528, $\hat{\beta}_1$ = −1.86528, $\hat{\beta}_2$ = 0.18472, $\hat{\beta}_3$ = 1.43472, $\hat{\beta}_4$ = −0.39861, $\hat{\beta}_5$ = − 4.24028 and $\hat{\beta}_6$ = 4.88472. These lead to $R(\mu, \beta) = \hat{\mu}\, y_{..} + \sum_{j=1}^{6} \hat{\beta}_j\, y_{.j}$ = 673.1258 → $SS_{\text{Treatments}}$(adjusted) = $R(\mu, \tau, \beta) - R(\mu, \beta)$ = 836.52400 − 673.12583 = 163.39817, which is identical to the Minitab exact ANOVA output listed below under AdjSS.

General Linear Model: y versus EP, BLKs

Factor	Type	Levels	Values
EP	fixed	4	8500, 8700, 8900, 9100
BLKs	fixed	6	1, 2, 3, 4, 5, 6

Analysis of Variance for y, using Adjusted SS for Tests

Source	DF	Seq SS	Adj SS	Adj MS	F_0	*P-Value*
EP	3	163.995	163.398	54.466	7.50	0.003
BLKs	5	189.522	189.522	37.904	5.22	0.007
Error	14	101.696	101.696	7.264		
Total	22	455.213				

S = 2.69518   R-Sq = 77.66%   R-Sq(adj) = 64.89%

In order to verify the value of SS(BLKS) in the above ANOVA table, we have to hypothesize that $\beta_1 = \beta_2 = \ldots = \beta_6 = 0$ and put these hypothesized values in the original LSNEs and solve the remaining 5 equations in 5 unknowns simultaneously, as given below.

$$\mu: \; 23\mu + 6\tau_1 + 5\tau_2 + 6\tau_3 + 6\tau_4 = 105.40$$

$$\tau_1: \; 6\mu + 6\tau_1 + \qquad = 46.90$$

$$\tau_2: \; 5\mu + \quad 5\tau_2 + \qquad = 30.40$$

$$\tau_3: \; 6\mu + \qquad 6\tau_3 + \quad = 23.50$$

$$\text{Constraint:} \quad \tau_1 + \tau_2 + \tau_3 + \tau_4 = 0.00$$

The solutions to the above system is $\hat{\mu}$ = 4.6450, $\hat{\tau}_1$ = 3.1717, $\hat{\tau}_2$ = 1.4350, $\hat{\tau}_3$ = –0.7283, $\hat{\tau}_4$ = –3.8783. These estimates yield $R(\mu, \tau) = \hat{\mu}\, y_{..} + \sum_{i=1}^{4} \hat{\tau}_i\, y_{i.}$ = 647.0020, and hence SS(BLKS) = 836.5240 – 47.0020 = 189.5220. Note that Montgomery's Table 4.8, atop p.155, does not match that of Minitab's listed above. Even if, we predict the value of x = $\hat{y}_{24}$ as x = $\bar{y}_{2.}$ = 91.08 and repeat an approximate ANOVA analysis, Minitab will give the following output, which is an approximate ANOVA Table.

Source	DF	Seq SS	Adj SS	Adj MS	F_0	P
EPm	3	166.144	166.144	55.381	8.17	0.002
Blk	5	189.522	189.522	37.904	5.59	0.004
Error	15	101.696	101.696	6.780		
Total	23	457.362	457.3618			

S = 2.60379   R-Sq = 77.76%   R-Sq(adj) = 65.91%, which still does not match Table 4.8 of Montgomery's 8th edition.

Exercise 11. Work problem 4.19, on page 179 of Montgomery's 8th edition using only the approximate method; obtain the ANOVA Table but omit all other parts of this problem.

If the F-test for a RCBD rejects H_0: $\tau_i = 0$ for all i, then Tukey's pair-wise comparisons can be applied to determine which 2 treatments differ significantly.

THE LATIN SQUARE DESIGN (LSQD)

This is a design where there are 2 block restrictions on randomization, i.e., row and column restrictions on randomization. The letters A, B, C, … represent different treatments of the same factor. A standard LSQD is one where the 1st row and 1st column are written in alphabetical order. The example in Table 4.9 on page 159 of Montgomery's 8th ed. represents a standard LSQD with row restriction (Batches of raw material) and column restriction (Operators). Note that the five treatments of the factor "Rocket Propellant Formulations" A, B, C, D, and E are in alphabetical order in the 1st row and column and hence this LSQD is of the standard type. A Latin Square is of the standard type as long as both 1st row and column are in alphabetical order and every letter (or treatment) must appear exactly once in every row and column.

Fortunately, in a LSQD, as long as there are no missing observations, rows, columns and treatments are orthogonal, and hence we can break down the total observed deviation $y_{ijk} - \bar{y}_{...}$, where the dependent variable y is burning rate, as follows:

$$y_{ijk} - \bar{y}_{...} \equiv (\bar{y}_{i..} - \bar{y}_{...}) + (\bar{y}_{.j.} - \bar{y}_{...}) + (\bar{y}_{..k} - \bar{y}_{...}) + (y_{ijk} - \bar{y}_{i..} - \bar{y}_{.j.} - \bar{y}_{..k} + 2\bar{y}_{...})$$

where $\hat{\tau}_k \equiv (\bar{y}_{..k} - \bar{y}_{...})$ is the effect of the k^{th} treatment. As before, squaring both sides of the above identity 1st and then summing over all i, j, and k gives rise to the orthogonal partition of Total SS, i.e.,

$$SS_T = SS(Rows) + SS(Columns) + SS(Treatments) + SS(Residuals).$$

Since a $p \times p$ LSQD has p^2 observations, and Row, Column restrictions and treatments all carry exactly $p - 1$ *df*, then Residuals must have $p^2 - 1 - 3(p - 1) = p^2 - 3p + 2 = (p - 2)(p - 1)$ *df*. You should study the Example 4.3 on pp. 161-162 of Montgomery(8e) for which $p = 5$ and *df* of Residuals is $5^2 - 1 - 3(5 - 1) = 12$, which is listed under the Error term in the ANOVA Table 4.12, atop page 162.

Exercise 12. Work problem 4.23 on page 180 of Montgomery(8e).

BALANCED INCOMPLETE BLOCK DESIGNS (BIBDs)

Incomplete block designs (INBDs) occur frequently both in manufacturing and science applications when the block size k is smaller than the number of treatments "a". This implies that each treatment will be tested only in r blocks, where $r < b$ = the total number of blocks. For example, suppose we have 6 brands of passenger tires whose qualities are to be compared. First, a car (the block) can have a maximum of $k = 4$ tires, and secondly, for obvious reasons, it is wrong to compare tires on front wheel with those in the rear wheel; so in essence our block sizes is only $k = 2$.

Yet, as another example, suppose we wish to compare 5 programming languages (Fortran, C, Java, C^{++}, and Visual Basic) for efficiency, but all of our human subjects are versatile at only 3 languages, i.e., our block sizes are only $k = 3$, but we have $a = 5$ treatments. The question is then how do we design an experiment that would lead to a balanced INBD (i.e., the number of times, λ, that any 2 treatments appear together in all the blocks are the same for all pairs of treatments in the entire experiment). Since each treatment appears in $r < b$ blocks and there are "a" treatments, then the total number of experimental runs is $a \times r$. Further, since there are a total of "b" blocks each of size $k < a$, then the total number of observations must also equal to $b \times k$, and thus $a \times r$ must equal to $b \times k$.

To determine the design layout we must argue that for all balanced INBDs λ must equal to $r(k-1)/(a-1)$, or else we will not have a balanced INBD! The argument for $\lambda = r(k-1)/(a-1)$ is given atop page 169 of Montgomery's 8th edition and after we design the problem in the above paragraph with $a = 5$ programming codes and $k = 3$, I will try to illustrate as to why $\lambda = r(k-1)/(a-1)$. Before designing the experiment, we must remind the reader that randomization will take place only within each block because this is a block design and not a CRD. For the case of $a = 5$ and $k = 3$, our first task is to determine r such that λ will be a positive integer; $\lambda = r(k-1)/(a-1) = r(3-1)/(5-1) = r/2 \rightarrow r = 2, 4, 6, 8, \ldots \rightarrow$ At $r = 2$, $a \times r = 10$ which must equal to $b \times k = 3b \rightarrow$ implying that the number of blocks $b = 10/3$, which is impossible. Next we try $r = 4$: $\rightarrow a \times r = 20 = 3b$, which is again impossible. Finally, we try $r = 6$: $\rightarrow a \times r = 30 = 3b$, which shows that the minimum number of programmers (or blocks) needed to conduct this experiment is $b = 30/3 = 10$. We could use the time taken by each programmer to code and debug the same complex problem as our response variable y, measured in hours. One possible layout of the above DOE is given below.

Blocks / Treats	1	2	3	4	5	6	7	8	9	10	$y_{i.}$
Fortran		5.5	4.6		3.9		5.2	6.3		4.8	30.3
C		7.6*		8.5	5.6*	4.1	6.9*		7.2		39.9
Java	4.1		3.9*	5.6	4.7*	4.0				5.0*	27.3
C^{++}	3.8	4.1*		5.1				4.8*	5.9	4.2*	27.9
VB	5.6		5.2*			4.9	6.5*	6.8*	7.6		36.6
$y_{.j}$	13.5	17.2	13.7	19.2	14.2	13.0	18.6	17.9	20.7	14.0	162

Note that $a = 5$, $r = 6$, $a\times r = 30$, $b = 10$, $k = 3$, $b\times k = 30 \rightarrow a\times r = b\times k$, $\lambda = r(k-1)/(a-1) = 3$, i.e., each pair of treatments (such as Fortran and C) appear exactly 3 times together in all the 10 blocks. The design is balanced because $\lambda = 3$ for every pair of treatments. By now it should be clear to as to why $\lambda = r(k-1)/(a-1)$. If you consider the treatment 1 "Fortran" which appears in blocks 2, 3, 5, 7, 8 & 10 (r = 6 out of the 10 BLKs), then in the same 6 blocks (2, 3, 5, 7, 8, 10), because block size is k, there are $r(k-1) = 12$ (please count the * observations for yourself) other observations taken with the other $a-1 = 4$ treatments. Those asterisked observations are the ones with which Fortran is being paired with the other $(a-1)$ treatments in all the 10 blocks. Since the design is balanced, the same 12* observations must represent the grand total number of times that treatment 1 appears together with all the other treatments in the blocks 2, 3, 5, 7, 8, and 10. But this total by definition of λ must also equal to $\lambda(a-1)$, and thus $\lambda(a-1) = r(k-1) = 12$.

In order to obtain the ANOVA Table, we can compute the SS(Total) and SS(Blocks) as before:

$\longrightarrow$ SS(Total) = SS_T = USS – CF = 921.90 – $162^2/30$ = 47.10, SS(Blocks) =($13.5^2 + 17.2^2 + \ldots + 14.0^2$)/3 – CF = 23.84. However, since treatments do not appear in every block, then we must adjust for this non-orthogonality according to Equations (4.34) and (4.35) on pages 169 of Montgomery(8e). The use of Equation (4.35) gives: $Q_1 = 30.3 - (17.2 + 13.7 + 14.2 + 18.6 + 17.9 + 14.0)/k = -1.566666$, $Q_2 = 39.9 -$

(17.2 + 19.2 + 14.2 + 13.0+ 18.6 +20.7)/3 = 5.6000. Similarly, you shouldverify that $Q_3 = -1.9$, $Q_4 = -6.266666$,

$Q_5 = 4.133333$, and $\sum_{i=1}^{5} Q_i \equiv 0$. Equation (4.34) on p. 169 of Montgomery(8e) now gives $SS_{Treatments}$(adjusted) =

$k(Q_1^2 + Q_2^2 + \ldots + Q_5^2)/(\lambda \times 5) = 18.756$. The ANOVA Table for the design matrix near the bottom of page 34 of these notes is provided below.

Source of Variation	*df*	SS	MS	The critical value of F is $F_{0.05,4,16}$
Total	29	47.10		$F_{0.05,4,16} = 3.01$
Blocks (Unadjusted) Blocks (Adjusted)	9	23.84 21.8360		
Treatments (adjusted)	4	18.756	4.6890	$F_0 = 16.6572$
Residuals	16	4.504	0.2815	*P-value* = $0.0^4 1473$

Note that in the above BIBD, the SS(BLKS) cannot be directly adjusted for treatments using formulas (4.38 & 4.39) on page 171 of Montgomery(8e) because the number of times that each pair of blocks appear in all the treatments is not the same for all pairs of blocks. For example, blocks 2 & 3 appear together only with Fortran (their $\lambda = 1$) while blocks 2 & 5 appear together in Fortran and C (i.e., their $\lambda = 2$). However, the Blocks can be adjusted for treatments only if the design is symmetric such as Table 4.22 on p. 168 of Montgomery(8e) and Problem 4.40 on p. 181, where $a = b = 5$.   The above ANOVA table clearly shows that there are significant differences amongst treatments, and that we must strongly reject the null hypothesis $H_0 : \tau_1 = \tau_2 = \tau_3 = \tau_4 = \tau_5 = 0$ at a LOS as small as the *P-value* = $\hat{\alpha}$ = 0.00001473 (or larger).

<u>Exercise 13.</u>   Work problem 4.40 on page 181 of Montgomery(8e) using the method described on pp. 33-36 of my notes.

Errata for Chapter 4 of Montgomery's 8th Edition

1. Page 154, at the last paragraph, change y'_{ij} to $y'_{..}$.

2. Page 155, replace the entire ANOVA Table 4.8 atop page 155 with either one of the two listed below from Minitab, where in the approximate case the missing response x = y_{24} has been replaced with 91.08.

General Linear Model:

Approximate ANOVA for y versus EP, BLKs

Source	DF	Seq SS	Adj SS	Adj MS	F_0	P
EPm	3	166.144	166.144	55.381	8.17	0.002
BLKs	5	189.522	189.522	37.904	5.59	0.004
Error	15	101.696	101.696	6.780		
Total	23	457.362	457.3618			

S = 2.60379   R-Sq = 77.76%   R-Sq(adj) = 65.91%

Or: The Exact ANOVA from Minitab's GLM: y versus EP, BLKs

Factor	Type	Levels	Values
EP	fixed	4	8500, 8700, 8900, 9100
BLKs	fixed	6	1, 2, 3, 4, 5, 6

Analysis of Variance for y, using Adjusted SS for Tests

Source	DF	Seq SS	Adj SS	Adj MS	F	P
EP	3	163.995	163.398	54.466	7.50	0.003
BLKs	5	189.522	189.522	37.904	5.22	0.007
Error	14	101.696	101.696	7.264		
Total	22	455.213				

3. Page 160, Table 4.10, in the equation for $SS_{Treatments}$ modify its CF from $y^2_{..}/N$ to $y^2_{...}/N$.

4. Page 171, change the Eq. (4.38) for Q'_j to the following

$$Q'_j = y_{.j} - \frac{1}{r}\sum_{i=1}^{a} n_{ij} y_{i.} = y_{.j} - \frac{1}{3}\sum_{i=1}^{a} n_{ij} y_{i.}$$