STATISTICAL MODELLING

IV. Randomized Complete Block Design (RCBD)

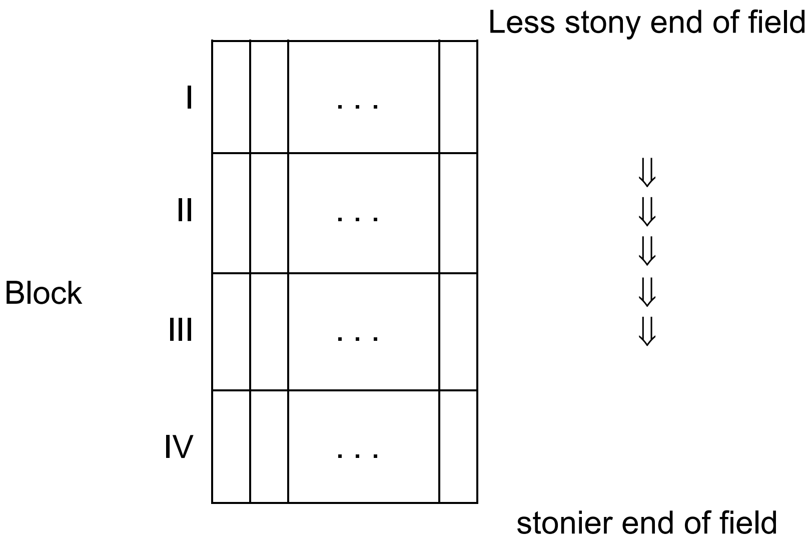
(ref. Mead and Curnow, sec. 5.1–5.3; Cochran and Cox, sec. 4.2)

IV.A Design of an RCBD

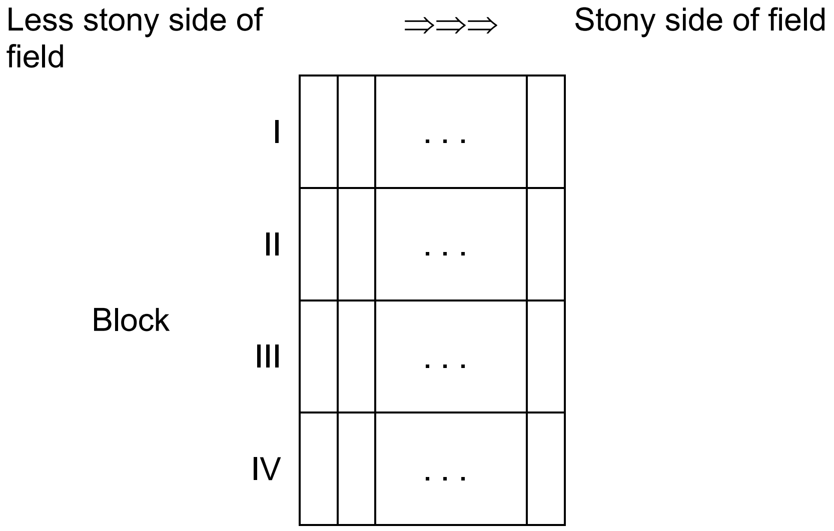
Definition IV.6: A **randomized complete block design** is one in which the number of experimental units per block is equal to the number of treatments and every treatment occurs once and only once in each block, the order of treatments within a block being randomized. ■

We will use *b* to denote the number of blocks and *t* to denote both the number of units in each block and the number of treatments. The total number of observations will be denoted $n = bt$.

In the RCBD the units are grouped into blocks such that the units in a block are as similar as possible. This would lead to different blocks. In the context of field experiments, where the units are called plots, this means placing plots parallel to the trend and blocks perpendicular to it as illustrated in the following diagram:



It is clear that the blocks will be quite different and the plots similar. However, suppose that I had thought there was a fertility trend in a particular direction, say down the field as before, and so had laid out the trial in an RCBD to take this into account. In the event it turns out that I had got it wrong and the trend was across the field. The situation would be as follows:



Clearly, Blocks would be similar and plots different. In fact this experiment can be less sensitive than a CRD. So getting it wrong can be costly.

a) Obtaining a layout for an RCBD in R

The general set of expressions for obtaining an RCBD layout is given in Appendix B, *Randomized layouts and sample size computations in R*.

To use these expressions to generate a layout for a particular case, you will need to substitute the actual values for *b*, *t* and *n* and the actual names for *Blocks*, *Units, Treats* and the data frame to contain them. Also, the labels for the treatments are optional. The crucial feature, that makes this design different from a completely randomized design, is that there are two unrandomized factors indexing the units and there is nesting between these factors: *Units* are nested within *Blocks*. This is because our aim is to randomize the treatments to the *Units* **within** each *Block*. The `nested.factors` argument to `fac.layout` is used to specify this.

Example IV.1 Penicillin yield

In this example the effects of four treatments (A, B, C and D) on the yield of penicillin are to be investigated. It was known that corn steep liquor, an important raw material in producing penicillin, is highly variable from one blending of it to another. So, to ensure that the results of the experiment apply to more than one blend, several blends are to be used in the experiment. Thus the trial was conducted using the same blend in four flasks and randomizing the four treatments to these four flasks. Altogether five blends were utilized.

The names to be used for the blocks, units and treatments for this example are Blends, Flask and Treat, respectively. Also, $b = 5$ and $t = 4$ so that $n = 20$. Substituting these values into the expressions above, yield the following expressions to be used in this case:

```
> b <- 5
> t <- 4
> n <- b*t
> RCBDPen.unit <- list(Blend=b, Flask=t)
> RCBDPen.nest <- list(Flask = "Blend")
> Treat <- factor(rep(1:t, times=b), labels=c("A","B","C","D"))
> data.frame(fac.gen(RCBDPen.unit), Treat) #basic systematic layout
   Blend Flask Treat
1      1     1     A
2      1     2     B
3      1     3     C
4      1     4     D
5      2     1     A
6      2     2     B
7      2     3     C
8      2     4     D
9      3     1     A
10     3     2     B
11     3     3     C
12     3     4     D
13     4     1     A
14     4     2     B
15     4     3     C
16     4     4     D
17     5     1     A
18     5     2     B
19     5     3     C
20     5     4     D
> RCBDPen.lay <- fac.layout(unrandomized = RCBDPen.unit,
+                           nested.factors = RCBDPen.nest,
+                           randomized = Treat, seed = 311)
> RCBDPen.lay
   Units Permutation Blend Flask Treat
1      1          11     1     1     C
2      2          12     1     2     B
3      3          10     1     3     D
4      4           9     1     4     A
5      5          13     2     1     C
6      6          15     2     2     D
7      7          16     2     3     B
8      8          14     2     4     A
9      9           8     3     1     D
```

```
10    10          7    3    2    C
11    11          5    3    3    A
12    12          6    3    4    B
13    13         17    4    1    A
14    14         19    4    2    D
15    15         20    4    3    B
16    16         18    4    4    C
17    17          4    5    1    A
18    18          2    5    2    D
19    19          1    5    3    B
20    20          3    5    4    C
```

Note that the layout is given in what is termed **standard order** for Blend then Flask in that the values of the first factor change slowest and the last change fastest. From the layout we see that, for the first blend, the Treatments are to be done in the order C, B, D, A. ■

IV.B Indicator-variable models and estimation for an RCBD

a) Maximal model

Generally, the RCBD involves *b* blocks in each of which *t* treatments are observed so that there are $n = b \times t$ observations in all. The maximal model used for an RCBD is:

$$\psi_{\text{B+T}} = E\left[\mathbf{Y}\right] = \mathbf{X}_{\text{B}}\boldsymbol{\beta} + \mathbf{X}_{\text{T}}\boldsymbol{\tau} \text{ and } \text{var}\left[\mathbf{Y}\right] = \sigma^2\mathbf{I}_n.$$

where **Y** is the *n*-vector of random variables for the response variable observations,

$\boldsymbol{\beta}$ is the *b*-vector of parameters specifying a different mean response for each block,

$\mathbf{X}_{\text{B}}$ is the $n \times b$ matrix indicating the block from which an observation came,

$\boldsymbol{\tau}$ is the *t*-vector of parameters specifying a different mean response for each treatment,

$\mathbf{X}_{\text{T}}$ is the $n \times t$ matrix indicating the observations that received each of the treatments.

Our model also involves assuming $\mathbf{Y} \sim N\left(\psi_{\text{B+T}}, \mathbf{V}\right)$.

The model for the expectation is still of the form $E\left[\mathbf{Y}\right] = \mathbf{X}\boldsymbol{\theta}$, with $\mathbf{X} = \left[\mathbf{X}_{\text{B}} \quad \mathbf{X}_{\text{T}}\right]$ and $\boldsymbol{\theta}' = \left[\boldsymbol{\beta}' \quad \boldsymbol{\tau}'\right]$.

It can be show that for this model $\hat{\psi}_{\text{B+T}} = \overline{\mathbf{B}} + \overline{\mathbf{T}} - \overline{\mathbf{G}}$ where $\overline{\mathbf{B}}, \overline{\mathbf{T}}$ and $\overline{\mathbf{G}}$ are the *n*-vectors of block, treatment and grand means, respectively.

Note that $\overline{\mathbf{B}} = \mathbf{M}_{\text{B}}\mathbf{Y}$**,** $\overline{\mathbf{T}} = \mathbf{M}_{\text{T}}\mathbf{Y}$ **and** $\overline{\mathbf{G}} = \mathbf{M}_{\text{G}}\mathbf{Y}$ are the block, treatment and grand mean operators, respectively. So once again the estimators of the expected values are functions of means.

Suppose the data has been arranged in the vector **Y** in nonrandomized order with all the observations for a block placed together. This is called standard order for blocks then treatments.

In this case, the mean operators are:

$$\mathbf{M}_{\mathrm{G}} = n^{-1}\mathbf{J}_b \otimes \mathbf{J}_t = n^{-1}\mathbf{J}_n$$
$$\mathbf{M}_{\mathrm{B}} = t^{-1}\mathbf{I}_b \otimes \mathbf{J}_t$$
$$\mathbf{M}_{\mathrm{T}} = b^{-1}\mathbf{J}_b \otimes \mathbf{I}_t$$

where $\otimes$ is called the direct product operator and,
if $\mathbf{A}_r$ and $\mathbf{B}_c$ are square matrices of order r and c, respectively,

$$\mathbf{A}_r \otimes \mathbf{B}_c = \begin{bmatrix} a_{11}\mathbf{B} & \cdots & a_{1r}\mathbf{B} \\ \vdots & \ddots & \vdots \\ a_{r1}\mathbf{B} & \cdots & a_{rr}\mathbf{B} \end{bmatrix}.$$

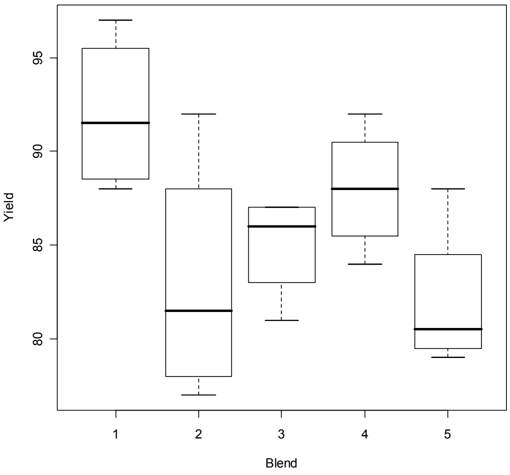
Note that the mean operators are simpler than for the CRD because blocks and treatments must be equally replicated and so the divisor can be taken out as a factor leaving matrices whose elements that are all zero or one.

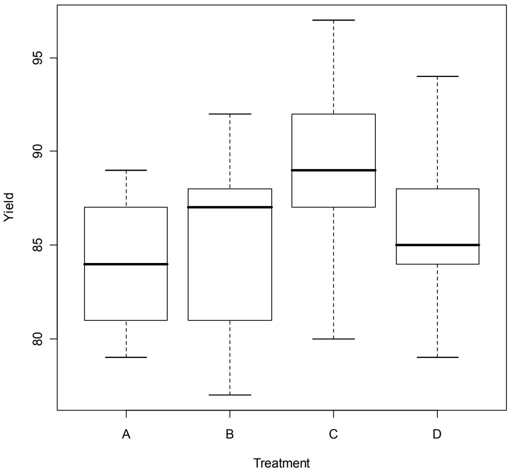
Example IV.1 Penicillin yield (continued)

The yields of penicillin, in nonrandom order, were:

		Treatment			
		A	B	C	D
	1	89	88	97	94
	2	84	77	92	79
Blend	3	81	87	87	85
	4	87	92	89	84
	5	79	81	80	88

The boxplots for an initial exploration of the data are as follows:





It would appear that there are yield mean differences between the blends and so the blocking was advisable. There also appears to be yield mean differences between the treatments.

If the yields are in standard order for Blend then Treatment, as for the systematic layout that is generated prior to randomization, then the vectors and matrices in the expectation model are:

$$\mathbf{y}=\begin{bmatrix}89\\88\\97\\94\\84\\77\\92\\79\\81\\87\\87\\85\\87\\92\\89\\84\\79\\81\\80\\88\end{bmatrix},\quad \mathbf{X}_{\mathrm{B}}=\begin{bmatrix}1&0&0&0&0\\1&0&0&0&0\\1&0&0&0&0\\1&0&0&0&0\\0&1&0&0&0\\0&1&0&0&0\\0&1&0&0&0\\0&1&0&0&0\\0&0&1&0&0\\0&0&1&0&0\\0&0&1&0&0\\0&0&1&0&0\\0&0&0&1&0\\0&0&0&1&0\\0&0&0&1&0\\0&0&0&1&0\\0&0&0&0&1\\0&0&0&0&1\\0&0&0&0&1\\0&0&0&0&1\end{bmatrix},\quad \boldsymbol{\beta}=\begin{bmatrix}\beta_1\\\beta_2\\\beta_3\\\beta_4\\\beta_5\end{bmatrix},\quad \mathbf{X}_{\mathrm{T}}=\begin{bmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\\1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\\1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\\1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\\1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{bmatrix},\quad \boldsymbol{\tau}=\begin{bmatrix}\tau_1\\\tau_2\\\tau_3\\\tau_4\end{bmatrix}$$

■

In the case of the example, the data have been arranged as prescribed so that the operators are as follows:

$$\mathbf{M}_{\mathrm{G}}=20^{-1}\mathbf{J}_5\otimes\mathbf{J}_4=\frac{1}{20}\begin{bmatrix}\mathbf{J}_4&\mathbf{J}_4&\mathbf{J}_4&\mathbf{J}_4&\mathbf{J}_4\\\mathbf{J}_4&\mathbf{J}_4&\mathbf{J}_4&\mathbf{J}_4&\mathbf{J}_4\\\mathbf{J}_4&\mathbf{J}_4&\mathbf{J}_4&\mathbf{J}_4&\mathbf{J}_4\\\mathbf{J}_4&\mathbf{J}_4&\mathbf{J}_4&\mathbf{J}_4&\mathbf{J}_4\\\mathbf{J}_4&\mathbf{J}_4&\mathbf{J}_4&\mathbf{J}_4&\mathbf{J}_4\end{bmatrix}$$

$$=\frac{1}{20}\begin{bmatrix}1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1\\1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1\\1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1\\1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1\\1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1\\1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1\\1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1\\1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1\\1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1\\1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1\\1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1\\1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1\\1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1\\1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1\\1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1\\1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1\\1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1\\1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1\\1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1\\1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1\end{bmatrix}$$

$$\mathbf{M}_{\mathrm{B}} = 4^{-1}\mathbf{I}_5 \otimes \mathbf{J}_4 = \frac{1}{4}\begin{bmatrix} \mathbf{J}_4 & \mathbf{0}_{4\times4} & \mathbf{0}_{4\times4} & \mathbf{0}_{4\times4} & \mathbf{0}_{4\times4} \\ \mathbf{0}_{4\times4} & \mathbf{J}_4 & \mathbf{0}_{4\times4} & \mathbf{0}_{4\times4} & \mathbf{0}_{4\times4} \\ \mathbf{0}_{4\times4} & \mathbf{0}_{4\times4} & \mathbf{J}_4 & \mathbf{0}_{4\times4} & \mathbf{0}_{4\times4} \\ \mathbf{0}_{4\times4} & \mathbf{0}_{4\times4} & \mathbf{0}_{4\times4} & \mathbf{J}_4 & \mathbf{0}_{4\times4} \\ \mathbf{0}_{4\times4} & \mathbf{0}_{4\times4} & \mathbf{0}_{4\times4} & \mathbf{0}_{4\times4} & \mathbf{J}_4 \end{bmatrix}$$

$$= \frac{1}{4}\begin{bmatrix}
1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1
\end{bmatrix}$$

$$\mathbf{M}_{\mathrm{T}} = 5^{-1}\mathbf{J}_5 \otimes \mathbf{I}_4 = \frac{1}{5}\begin{bmatrix} \mathbf{I}_4 & \mathbf{I}_4 & \mathbf{I}_4 & \mathbf{I}_4 & \mathbf{I}_4 \\ \mathbf{I}_4 & \mathbf{I}_4 & \mathbf{I}_4 & \mathbf{I}_4 & \mathbf{I}_4 \\ \mathbf{I}_4 & \mathbf{I}_4 & \mathbf{I}_4 & \mathbf{I}_4 & \mathbf{I}_4 \\ \mathbf{I}_4 & \mathbf{I}_4 & \mathbf{I}_4 & \mathbf{I}_4 & \mathbf{I}_4 \\ \mathbf{I}_4 & \mathbf{I}_4 & \mathbf{I}_4 & \mathbf{I}_4 & \mathbf{I}_4 \end{bmatrix}$$

$$= \frac{1}{5}\begin{bmatrix}
1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\
1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\
1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\
1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\
1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1
\end{bmatrix}$$

The estimators are:

$$\bar{\mathbf{B}} = \begin{bmatrix} \bar{B}_1 \\ \bar{B}_1 \\ \bar{B}_1 \\ \bar{B}_1 \\ \bar{B}_2 \\ \bar{B}_2 \\ \bar{B}_2 \\ \bar{B}_2 \\ \bar{B}_3 \\ \bar{B}_3 \\ \bar{B}_3 \\ \bar{B}_3 \\ \bar{B}_4 \\ \bar{B}_4 \\ \bar{B}_4 \\ \bar{B}_4 \\ \bar{B}_5 \\ \bar{B}_5 \\ \bar{B}_5 \\ \bar{B}_5 \end{bmatrix}, \quad \bar{\mathbf{T}} = \begin{bmatrix} \bar{T}_A \\ \bar{T}_B \\ \bar{T}_C \\ \bar{T}_D \\ \bar{T}_A \\ \bar{T}_B \\ \bar{T}_C \\ \bar{T}_D \\ \bar{T}_A \\ \bar{T}_B \\ \bar{T}_C \\ \bar{T}_D \\ \bar{T}_A \\ \bar{T}_B \\ \bar{T}_C \\ \bar{T}_D \\ \bar{T}_A \\ \bar{T}_B \\ \bar{T}_C \\ \bar{T}_D \end{bmatrix} \text{ and } \bar{\mathbf{G}} = \begin{bmatrix} \bar{G} \\ \bar{G} \\ \bar{G} \\ \bar{G} \\ \bar{G} \\ \bar{G} \\ \bar{G} \\ \bar{G} \\ \bar{G} \\ \bar{G} \\ \bar{G} \\ \bar{G} \\ \bar{G} \\ \bar{G} \\ \bar{G} \\ \bar{G} \\ \bar{G} \\ \bar{G} \\ \bar{G} \\ \bar{G} \end{bmatrix}$$

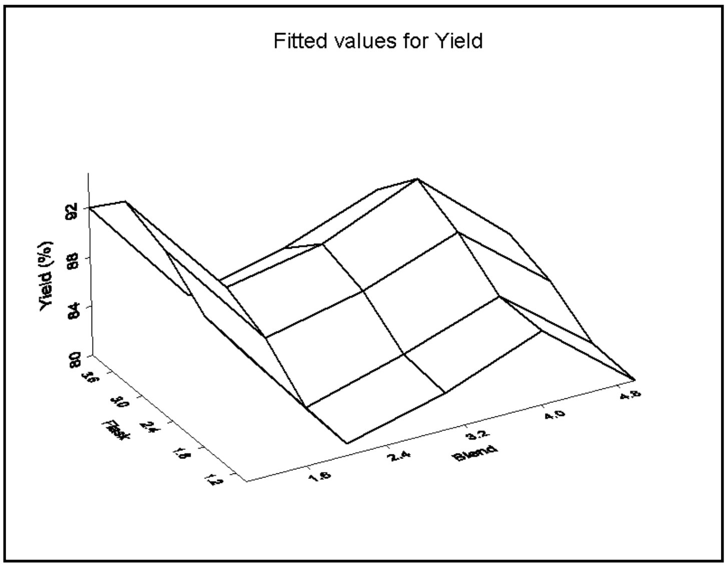
The means are given in the following table:

		Treatment				
		A	B	C	D	Means
Blend	1	89	88	97	94	92
	2	84	77	92	79	83
	3	81	87	87	85	85
	4	87	92	89	84	88
	5	79	81	80	88	82
	Means	84	85	89	86	86

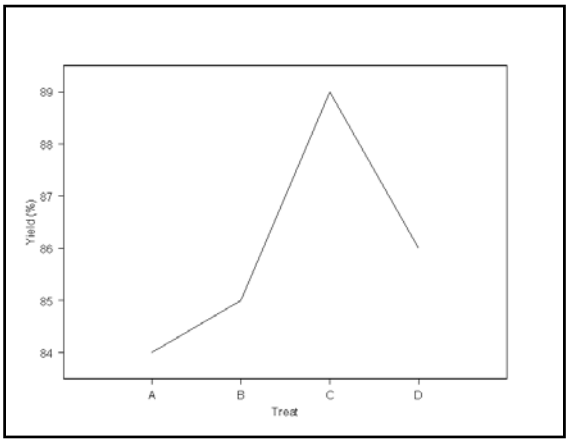
so that the estimates are:

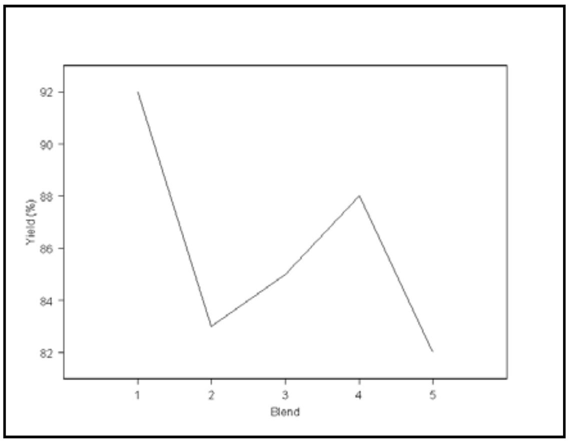
$$\bar{\mathbf{b}} = \begin{bmatrix} 92 \\ 92 \\ 92 \\ 92 \\ 83 \\ 83 \\ 83 \\ 83 \\ 85 \\ 85 \\ 85 \\ 85 \\ 88 \\ 888 \\ 88 \\ 88 \\ 82 \\ 82 \\ 82 \\ 82 \end{bmatrix}, \quad \bar{\mathbf{t}} = \begin{bmatrix} 84 \\ 85 \\ 89 \\ 86 \\ 84 \\ 85 \\ 89 \\ 86 \\ 84 \\ 85 \\ 89 \\ 86 \\ 84 \\ 85 \\ 89 \\ 86 \\ 84 \\ 85 \\ 89 \\ 86 \end{bmatrix}, \quad \bar{\mathbf{g}} = \begin{bmatrix} 86 \\ 86 \\ 86 \\ 86 \\ 86 \\ 86 \\ 86 \\ 86 \\ 86 \\ 86 \\ 86 \\ 86 \\ 86 \\ 86 \\ 86 \\ 86 \\ 86 \\ 86 \\ 86 \\ 86 \end{bmatrix} \text{ and } \boldsymbol{\psi}_{B+T} = \bar{\mathbf{b}} + \bar{\mathbf{t}} - \bar{\mathbf{g}} = \begin{bmatrix} 90 \\ 91 \\ 95 \\ 92 \\ 81 \\ 82 \\ 86 \\ 83 \\ 83 \\ 84 \\ 88 \\ 85 \\ 86 \\ 87 \\ 91 \\ 88 \\ 80 \\ 81 \\ 85 \\ 82 \end{bmatrix}$$

Now these fitted values are different for every combination of block and treatment. However they are additive as illustrated in the following diagram.



In one direction, these exhibit the same trend as the corresponding means, these trends being illustrated in the diagram below. If the additive model is to apply, the above surface should describe the differences between Blend-Treatment mean combinations, except for random variations around it.





■

b) Alternative expectation models

There are 4 possible different models for the expectation that we consider:

$$\boldsymbol{\psi}_{\mathrm{G}} = \mathbf{X}_{\mathrm{G}}\mu \qquad \text{(no treatment or block differences)}$$

$$\boldsymbol{\psi}_{\mathrm{B}} = \mathbf{X}_{\mathrm{B}}\boldsymbol{\beta} \qquad \text{(block differences only)}$$

$$\boldsymbol{\psi}_{\mathrm{T}} = \mathbf{X}_{\mathrm{T}}\boldsymbol{\tau} \qquad \text{(treatment differences only)}$$

$$\boldsymbol{\psi}_{\mathrm{B+T}} = \mathbf{X}_{\mathrm{B}}\boldsymbol{\beta} + \mathbf{X}_{\mathrm{T}}\boldsymbol{\tau} \qquad \text{(block and treatment differences)}$$

We note that $\mathcal{C}(\mathbf{X}_{\text{G}}) \subset \mathcal{C}(\mathbf{X}_{\text{B}}) \subset \mathcal{C}([\mathbf{X}_{\text{B}} \quad \mathbf{X}_{\text{T}}])$ and that $\mathcal{C}(\mathbf{X}_{\text{G}}) \subset \mathcal{C}(\mathbf{X}_{\text{T}}) \subset \mathcal{C}([\mathbf{X}_{\text{B}} \quad \mathbf{X}_{\text{T}}])$. Consequently, the model $\boldsymbol{\psi}_{\text{G}} = \mathbf{X}_{\text{G}}\mu$ is marginal to $\boldsymbol{\psi}_{\text{B}} = \mathbf{X}_{\text{B}}\boldsymbol{\beta}$, $\boldsymbol{\psi}_{\text{T}} = \mathbf{X}_{\text{T}}\boldsymbol{\tau}$ and $\boldsymbol{\psi}_{\text{B+T}} = \mathbf{X}_{\text{B}}\boldsymbol{\beta} + \mathbf{X}_{\text{T}}\boldsymbol{\tau}$ or $\boldsymbol{\psi}_{\text{G}} \leq \boldsymbol{\psi}_{\text{B}}, \boldsymbol{\psi}_{\text{T}}, \boldsymbol{\psi}_{\text{B+T}}$ and both $\boldsymbol{\psi}_{\text{B}} = \mathbf{X}_{\text{B}}\boldsymbol{\beta}$ and $\boldsymbol{\psi}_{\text{T}} = \mathbf{X}_{\text{T}}\boldsymbol{\tau}$ are marginal to $\boldsymbol{\psi}_{\text{B+T}} = \mathbf{X}_{\text{B}}\boldsymbol{\beta} + \mathbf{X}_{\text{T}}\boldsymbol{\tau}$ or $\boldsymbol{\psi}_{\text{B}}, \boldsymbol{\psi}_{\text{T}} \leq \boldsymbol{\psi}_{\text{B+T}}$.

Also note that, similar to the CRD, the models ψ_{B} and ψ_{T} can be obtained from $\psi_{\text{B+T}}$ by setting either $\boldsymbol{\beta}$ or $\boldsymbol{\tau}$ equal to zero and that ψ_{G} can be obtained from ψ_{B} and ψ_{T} by setting $\boldsymbol{\beta} = \mathbf{1}\mu$ and $\boldsymbol{\tau} = \mathbf{1}\mu$, respectively.

The estimators of the expected values under the different models are:

$$\begin{aligned}
\hat{\boldsymbol{\psi}}_{\text{G}} &= \bar{\mathbf{G}} && (\text{no treatment or block differences})\\
\hat{\boldsymbol{\psi}}_{\text{B}} &= \bar{\mathbf{B}} && (\text{block differences only})\\
\hat{\boldsymbol{\psi}}_{\text{T}} &= \bar{\mathbf{T}} && (\text{treatment differences only})\\
\hat{\boldsymbol{\psi}}_{\text{B+T}} &= \bar{\mathbf{B}} + \bar{\mathbf{T}} - \bar{\mathbf{G}} && (\text{block and treatment differences})
\end{aligned}$$

That is they are all functions of the three mean vectors for this design.

IV.C Hypothesis testing using the ANOVA method for an RCBD

An analysis of variance will be used to choose between the four alternative expectation models for an RCBD.

a) Analysis of the penicillin example

Example IV.1 Penicillin yield (continued)

The hypothesis test for the example RCBD is as follows:

Step 1: Set up hypotheses

a) H_0: $\tau_1 = \tau_2 = \tau_3 = \tau_4$ (or $\mathbf{X}_{\text{T}}\boldsymbol{\tau}$ not required in model)
H_1: not all population Treatment means are equal

b) H_0: $\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5$ (or $\mathbf{X}_{\text{B}}\boldsymbol{\beta}$ not required in model)
H_1: not all population Blend means are equal

Set $\alpha = 0.05$.

Step 2: Calculate test statistics

The analysis of variance table for a RCBD is:

Source	df	SSq	MSq	F	Prob
Blends	4	264	66.0	3.50	0.041
Flasks [Blends]	15	296			
Treatments	3	70	23.3	1.24	0.339
Residual	12	226	18.8		
Total	19	560			

Note that Flasks[Blends] in this table means "Flasks within Blends".

Step 3: Decide between hypotheses

It would appear that there are significant differences between the blends but not between the treatments so that the expectation model that best describes the response appears to be $\boldsymbol{\psi}_B = \mathbf{X}_B\boldsymbol{\beta}$.

In our RCBD example there were significant differences between the blends so that the blocking based on blends has been effective. If a CRD had been used, that is the four treatments randomized to 20 flasks irrespective of blends, then the residual sum of squares would have been approximately the sum of the blend and residual sum of squares from the RCBD, viz. 490 and the mean square 490/16 = 30.625. That is, the residual mean square would have been twice as large and the experiment much less sensitive or not able to detect as small a difference. ■

As to the general question of whether the blocking has been effective, it turns out that it will if the units within a block are as similar as possible which necessarily leads to block differences. That this is the case can be seen as follows: suppose I have four plots to group into two blocks of two and that two plots are reasonably similar while the other two plots are similar to each other but quite different to the first pair. What happens if we use two dissimilar plots to form a block thus:

x + + x ——> + x + x

Clearly, the blocks are similar; but the treatments will apply to quite different plots so that σ^2 will be large and large differences between the treatments will be required before the effect of treatments will be greater than the uncontrolled variation.

Suppose on the other hand that similar plots are grouped together:

x + + x ——> + + x x

Clearly, the blocks are quite different and there are only small differences between the plots so that σ^2 will be small. Smaller differences between the treatments will be able to be detected.

b) Sums of squares for the analysis of variance

In this section we will use the generic names of Blocks, Units and Treatments for the factors in an RCBD.

The estimators of the sum of squares for the RCBD ANOVA are the sums of squares of the following vectors:

$$\begin{array}{ll}
\text{Total or Units SSq:} & \mathbf{D}_{\mathrm{G}} = \mathbf{Y} - \bar{\mathbf{G}} \\
\text{Blocks SSq:} & \mathbf{B}_{e} = \bar{\mathbf{B}} - \bar{\mathbf{G}} \\
\text{Units[Blocks] SSq:} & \mathbf{D}_{\mathrm{B}} = \mathbf{Y} - \bar{\mathbf{B}} \\
\text{Treatments SSq:} & \mathbf{T}_{e} = \bar{\mathbf{T}} - \bar{\mathbf{G}} \\
\text{Residual SSq:} & \mathbf{D}_{\mathrm{B+T}} = \mathbf{Y} - \bar{\mathbf{B}} - \bar{\mathbf{T}} + \bar{\mathbf{G}} = \mathbf{Y} - \mathbf{B}_{e} - \mathbf{T}_{e} - \bar{\mathbf{G}}
\end{array}$$

where, as in chapter 3, *Completely Randomized Design*, the **D**s are *n*-vectors of **d**eviations from **Y** and the vectors with the *e* subscripts are *n*-vectors of ***e***ffects.

From section IV.B, *Models and estimation for an RCBD*, we have that

$$\begin{aligned}
\bar{\mathbf{G}} &= \mathbf{M}_{\mathrm{G}}\mathbf{Y} = n^{-1}\left(\mathbf{J}_b \otimes \mathbf{J}_t\right)\mathbf{Y} \\
\bar{\mathbf{B}} &= \mathbf{M}_{\mathrm{B}}\mathbf{Y} = t^{-1}\left(\mathbf{I}_b \otimes \mathbf{J}_t\right)\mathbf{Y} \\
\bar{\mathbf{T}} &= \mathbf{M}_{\mathrm{T}}\mathbf{Y} = b^{-1}\left(\mathbf{J}_b \otimes \mathbf{I}_t\right)\mathbf{Y}
\end{aligned}$$

and let $\mathbf{Y} = \mathbf{M}_{\mathrm{U}}\mathbf{Y} = \left(\mathbf{I}_b \otimes \mathbf{I}_t\right)\mathbf{Y}$.

It can be shown that the sums of squares for the analysis of variance are given by

$$\begin{array}{rll}
\mathbf{D}'_{\mathrm{G}}\mathbf{D}_{\mathrm{G}} = \left(\mathbf{Y} - \bar{\mathbf{G}}\right)'\left(\mathbf{Y} - \bar{\mathbf{G}}\right) = \mathbf{Y}'\mathbf{Q}_{\mathrm{U}}\mathbf{Y} & \text{with } \mathbf{Q}_{\mathrm{U}} = \mathbf{M}_{\mathrm{U}} - \mathbf{M}_{\mathrm{G}} \\
\mathbf{B}'_{e}\mathbf{B}_{e} = \left(\bar{\mathbf{B}} - \bar{\mathbf{G}}\right)'\left(\bar{\mathbf{B}} - \bar{\mathbf{G}}\right) = \mathbf{Y}'\mathbf{Q}_{\mathrm{B}}\mathbf{Y} & \text{with } \mathbf{Q}_{\mathrm{B}} = \mathbf{M}_{\mathrm{B}} - \mathbf{M}_{\mathrm{G}} \\
\mathbf{D}'_{\mathrm{B}}\mathbf{D}_{\mathrm{B}} = \left(\mathbf{Y} - \bar{\mathbf{B}}\right)'\left(\mathbf{Y} - \bar{\mathbf{B}}\right) = \mathbf{Y}'\mathbf{Q}_{\mathrm{BU}}\mathbf{Y} & \text{with } \mathbf{Q}_{\mathrm{BU}} = \mathbf{M}_{\mathrm{U}} - \mathbf{M}_{\mathrm{B}} \\
\mathbf{T}'_{e}\mathbf{T}_{e} = \left(\bar{\mathbf{T}} - \bar{\mathbf{G}}\right)'\left(\bar{\mathbf{T}} - \bar{\mathbf{G}}\right) = \mathbf{Y}'\mathbf{Q}_{\mathrm{T}}\mathbf{Y} & \text{with } \mathbf{Q}_{\mathrm{T}} = \mathbf{M}_{\mathrm{T}} - \mathbf{M}_{\mathrm{G}} \\
\mathbf{D}'_{\mathrm{B+T}}\mathbf{D}_{\mathrm{B+T}} = \left(\mathbf{Y} - \bar{\mathbf{B}} - \bar{\mathbf{T}} + \bar{\mathbf{G}}\right)'\left(\mathbf{Y} - \bar{\mathbf{B}} - \bar{\mathbf{T}} + \bar{\mathbf{G}}\right) = \mathbf{Y}'\mathbf{Q}_{\mathrm{BU_{Res}}}\mathbf{Y} & \text{with } \mathbf{Q}_{\mathrm{BU_{Res}}} = \mathbf{M}_{\mathrm{U}} - \mathbf{M}_{\mathrm{T}} - \mathbf{M}_{\mathrm{B}} + \mathbf{M}_{\mathrm{G}}
\end{array}$$

All the **M**s and **Q**s are symmetric and idempotent.

So the analysis of variance table is constructed as follows:

Source	df	SSq	MSq	F	p
Blocks	b–1	$\mathbf{Y}'\mathbf{Q}_{\mathrm{B}}\mathbf{Y}$	$\dfrac{\mathbf{Y}'\mathbf{Q}_{\mathrm{B}}\mathbf{Y}}{b-1}=s_{\mathrm{B}}^2$	$s_{\mathrm{B}}^2 \big/ s_{\mathrm{BU_{Res}}}^2$	p_{B}
Units[Blocks]	$b(t$–1)	$\mathbf{Y}'\mathbf{Q}_{\mathrm{BU}}\mathbf{Y}$			
Treatments	t–1	$\mathbf{Y}'\mathbf{Q}_{\mathrm{T}}\mathbf{Y}$	$\dfrac{\mathbf{Y}'\mathbf{Q}_{\mathrm{T}}\mathbf{Y}}{t-1}=s_{\mathrm{T}}^2$	$s_{\mathrm{T}}^2 \big/ s_{\mathrm{BU_{Res}}}^2$	p_{T}
Residual	(b–1)(t–1)	$\mathbf{Y}'\mathbf{Q}_{\mathrm{BU_{Res}}}\mathbf{Y}$	$\dfrac{\mathbf{Y}'\mathbf{Q}_{\mathrm{BU_{Res}}}\mathbf{Y}}{(b-1)(t-1)}=s_{\mathrm{BU_{Res}}}^2$		
Total	bt–1	$\mathbf{Y}'\mathbf{Q}_{\mathrm{U}}\mathbf{Y}$			

Again, this partition of the sums of squares has a geometrical interpretation, with the **Q** matrices specifying orthogonal projections. The matrix $\mathbf{Q}_{\mathrm{U}}$ orthogonally projects the data vector into the bt–1 dimensional part of the bt-dimensional data space that is orthogonal to equiangular line. This is partitioned, by $\mathbf{Q}_{\mathrm{B}}$ and $\mathbf{Q}_{\mathrm{BU}}$, into two subspaces: a) the b–1 dimensional part of the b-dimensional Block space that is orthogonal to equiangular line and b) b(t–1) dimensional Units[Blocks] space. The latter space is then partitioned, by $\mathbf{Q}_{\mathrm{T}}$ and $\mathbf{Q}_{\mathrm{BU_{Res}}}$, into two subspaces: a) the t–1 dimensional part of the t-dimensional Treatment space that is orthogonal to equiangular line and b) the (b–1)(t–1) Residual subspace. Here the Block and Treatment spaces are the column spaces of the matrices $\mathbf{X}_{\mathrm{B}}$ and $\mathbf{X}_{\mathrm{T}}$, respectively. That is, the Units space is divided into the three orthogonal subspaces, the Blocks, Treatments and Residual subspaces that are orthogonal to the equiangular line.

Example IV.1 Penicillin yield (continued)

The means and effects needed for the analysis are given in the following table:

		Treatment					
		A	B	C	D	Means	Effects
Blend	1	89	88	97	94	92	6
	2	84	77	92	79	83	-3
	3	81	87	87	85	85	-1
	4	87	92	89	84	88	2
	5	79	81	80	88	82	-4
	Means	84	85	89	86	86	
	Effects	-2	-1	3	0		0

The vectors for computing the sums of squares are given in the following table.

		Total				Residual
		Flask	Blend	Flask[Blend]	Treat	Flask[Blend]
	Yield	deviations	Effects	deviations	effects	deviations
Treat	$\mathbf{y}$	$\mathbf{d}_{\mathrm{G}} = \mathbf{Q}_{\mathrm{U}}\mathbf{y}$	$\mathbf{b}_e = \mathbf{Q}_{\mathrm{B}}\mathbf{y}$	$\mathbf{d}_{\mathrm{B}} = \mathbf{Q}_{\mathrm{BF}}\mathbf{y}$	$\mathbf{t}_e = \mathbf{Q}_{\mathrm{T}}\mathbf{y}$	$\mathbf{d}_{\mathrm{B+T}} = \mathbf{Q}_{\mathrm{BF}_{\mathrm{Res}}}\mathbf{y}$
		$= \mathbf{y} - \bar{\mathbf{g}}$	$= \bar{\mathbf{b}} - \bar{\mathbf{g}}$	$= \mathbf{y} - \bar{\mathbf{b}}$	$= \bar{\mathbf{t}} - \bar{\mathbf{g}}$	$= \mathbf{y} - \bar{\mathbf{t}} - \bar{\mathbf{b}} + \bar{\mathbf{g}}$
A	89	3	6	-3	-2	-1
B	88	2	6	-4	-1	-3
C	97	11	6	5	3	2
D	94	8	6	2	0	2
A	84	-2	-3	1	-2	3
B	77	-9	-3	-6	-1	-5
C	92	6	-3	9	3	6
D	79	-7	-3	-4	0	-4
A	81	-5	-1	-4	-2	-2
B	87	1	-1	2	-1	3
C	87	1	-1	2	3	-1
D	85	-1	-1	0	0	0
A	87	1	2	-1	-2	1
B	92	6	2	4	-1	5
C	89	3	2	1	3	-2
D	84	-2	2	-4	0	-4
A	79	-7	-4	-3	-2	-1
B	81	-5	-4	-1	-1	0
C	80	-6	-4	-2	3	-5
D	88	2	-4	6	0	6
SSq		560	264	296	70	226

That is, the Units SSq is $\mathbf{Y}'\mathbf{Q}_{\mathrm{U}}\mathbf{Y} = 560$, the Blend SSq is $\mathbf{Y}'\mathbf{Q}_{\mathrm{B}}\mathbf{Y} = 264$, the Flask[Blend] SSq is $\mathbf{Y}'\mathbf{Q}_{\mathrm{BF}}\mathbf{Y} = 296$, the Treatments SSq is $\mathbf{Y}'\mathbf{Q}_{\mathrm{T}}\mathbf{Y} = 70$ and the Residual SSq is $\mathbf{Y}'\mathbf{Q}_{\mathrm{BF}_{\mathrm{Res}}}\mathbf{Y} = 226$.

Note that $\mathbf{y} = \bar{\mathbf{g}} + \mathbf{b}_{\mathrm{e}} + \mathbf{t}_{\mathrm{e}} + \mathbf{d}_{\mathrm{B+T}}$. That is, the data vector **y** has been decomposed into 4 orthogonal vectors, the first on the equiangular line and the other 3 into the the three orthogonal subspaces, the Blocks, Treatments and Residual subspaces that are orthogonal to the equiangular line. ■

c) Expected mean squares

To justify our choice of test statistic, we want to work out the expected values of the mean squares in the analysis of variance table under the four alternative expectation models. They are summarized in the following table:

Source	df	MSq	E[MSq]			
			$\boldsymbol{\psi}_{\mathrm{B+T}}$	$\boldsymbol{\psi}_{\mathrm{T}}$	$\boldsymbol{\psi}_{\mathrm{B}}$	$\boldsymbol{\psi}_{\mathrm{G}}$
Blocks	b–1	$\dfrac{\mathbf{Y}'\mathbf{Q}_{\mathrm{B}}\mathbf{Y}}{b-1}$	$\sigma^2+q_{\mathrm{B}}(\boldsymbol{\psi})$	σ^2	$\sigma^2+q_{\mathrm{B}}(\boldsymbol{\psi})$	σ^2
Units[Blocks]	$b(t$–1)					
Treatments	t–1	$\dfrac{\mathbf{Y}'\mathbf{Q}_{\mathrm{T}}\mathbf{Y}}{t-1}$	$\sigma^2+q_{\mathrm{T}}(\boldsymbol{\psi})$	$\sigma^2+q_{\mathrm{T}}(\boldsymbol{\psi})$	σ^2	σ^2
Residual	(b–1)(t–1)	$\dfrac{\mathbf{Y}'\mathbf{Q}_{\mathrm{BU_{Res}}}\mathbf{Y}}{(b-1)(t-1)}$	σ^2	σ^2	σ^2	σ^2
Total	bt–1					

$$q_{\mathrm{B}}(\boldsymbol{\psi})=\frac{\boldsymbol{\psi}'\mathbf{Q}_{\mathrm{B}}\boldsymbol{\psi}}{b-1}=\sum_{i=1}^{b}t\left(\beta_i-\bar{\beta}_{.}\right)^2\Big/(b-1) \text{ and } q_{\mathrm{T}}(\boldsymbol{\psi})=\frac{\boldsymbol{\psi}'\mathbf{Q}_{\mathrm{T}}\boldsymbol{\psi}}{t-1}=\sum_{j=1}^{t}b\left(\tau_j-\bar{\tau}_{.}\right)^2\Big/(t-1)$$

Once again numerator of $q_{\mathrm{B}}(\boldsymbol{\psi})$ is the sum of squares of $\mathbf{Q}_{\mathrm{B}}\boldsymbol{\psi}=(\mathbf{M}_{\mathrm{B}}-\mathbf{M}_{\mathrm{G}})\boldsymbol{\psi}$ and of $q_{\mathrm{T}}(\boldsymbol{\psi})$ is the sum of squares of $\mathbf{Q}_{\mathrm{T}}\boldsymbol{\psi}=(\mathbf{M}_{\mathrm{T}}-\mathbf{M}_{\mathrm{G}})\boldsymbol{\psi}$ where $\boldsymbol{\psi}$ depends on the model under which the expected mean squares are being computed. The expressions for $q_{\mathrm{B}}(\boldsymbol{\psi})$ and $q_{\mathrm{T}}(\boldsymbol{\psi})$ given below the above table are for the maximal model. Expressions under the reduced models can be obtained substituting the values of the β_is and τ_js under the reduced model. That is, set them equal to 0 or to μ as described in section IV.B. Also, given these expressions, the population means of the mean squares could be computed if knew the β_is, τ_js and σ^2.

It is clear from these expected mean squares that if the Treatments F is not significant then a model not involving $\mathbf{X}_{\mathrm{T}}\boldsymbol{\tau}$ is required, as those models are the ones for which $q_{\mathrm{T}}(\boldsymbol{\psi})$ is zero. Similarly, if the Blocks F is not significant then a model not involving $\mathbf{X}_{\mathrm{B}}\boldsymbol{\beta}$ is required. In the case where both are not significant, then the minimal model adequately describes the data.

Generally, we will only present the expected mean squares under the maximal model, realizing that $q(\boldsymbol{\psi})$ is zero under the null hypothesis that removes the term from the model.

The expected mean squares under the maximal model indicate that the Residual mean square estimates the uncontrolled variation in the experiment, that is, the variation arising from uncontrolled differences between units within the same block, both treatment and block differences having been eliminated; indeed, the block and treatment means calculated from the residual vector are zero.

Again, an intuitive feel for the fact that these expected mean squares are correct can be gained by considering the differences that will potentially contribute to differences between the block and treatment means. The following is the data and associated means:

		Treatment				
		A	B	C	D	Means
Blend	1	89	88	97	94	92
	2	84	77	92	79	83
	3	81	87	87	85	85
	4	87	92	89	84	88
	5	79	81	80	88	82
	Means	84	85	89	86	86

Two treatment means will differ because of the different treatments involved and because of the different runs (the units in this example) involved in the observations from which the means are calculated; but block differences will not contribute to treatment mean differences as all treatments involve the same set of blocks. Similarly, two block means will differ because they are in different blocks and involve different runs, but treatments will not contribute to block mean differences. The expected mean squares reflect this fact. The Treatment F again involves the question "Is the variance of the treatment means greater than can be expected from uncontrolled differences between the runs?" The Block F value involves a similar question.

d) Summary of the hypothesis test

We summarize the ANOVA-based hypothesis test for a RCBD involving t treatments in b blocks with a total of $n = bt$ observed units.

Step 1: Set up hypotheses

a) H_0: $\tau_1 = \tau_2 = \ldots = \tau_t$ (or $\mathbf{X}_{\mathrm{T}}\boldsymbol{\tau}$ not required in model)
H_1: not all population Treatment means are equal

b) H_0: $\beta_1 = \beta_2 = \ldots = \beta_b$ (or $\mathbf{X}_{\mathrm{B}}\boldsymbol{\beta}$ not required in model)
H_1: not all population Block means are equal

Set α.

Step 2: Calculate test statistics

The analysis of variance table for an RCBD is:

Source	df	MSq	E[MSq]	F	p
Blocks	b–1	$\frac{\mathbf{Y}'\mathbf{Q}_{\mathrm{B}}\mathbf{Y}}{b-1} = s^2_{\mathrm{B}}$	$\sigma^2 + q_{\mathrm{B}}(\boldsymbol{\psi})$	$s^2_{\mathrm{B}}/s^2_{\mathrm{BU_{Res}}}$	p_{B}
Units[Blocks]	b(t–1)				
Treatments	t–1	$\frac{\mathbf{Y}'\mathbf{Q}_{\mathrm{T}}\mathbf{Y}}{t-1} = s^2_{\mathrm{T}}$	$\sigma^2 + q_{\mathrm{T}}(\boldsymbol{\psi})$	$s^2_{\mathrm{T}}/s^2_{\mathrm{BU_{Res}}}$	p_{T}
Residual	(b–1)(t–1)	$\frac{\mathbf{Y}'\mathbf{Q}_{\mathrm{BU_{Res}}}\mathbf{Y}}{(b-1)(t-1)} = s^2_{\mathrm{BU_{Res}}}$	σ^2		
Total	bt–1				

Step 3: Decide between hypotheses

If $\Pr\{F \geq F_O\} = p \leq \alpha$ then the evidence suggests that the null hypothesis should be rejected.

e) Comparison with traditional two-way ANOVA

As for the analysis for the CRD, the above analysis of variance table and the traditional two-way ANOVA table are essentially the same — at any rate the values of the F-statistics are exactly the same. As illustrated in the table below, the two tables have in common three sources that are labelled differently but the tables differ in that our table includes the line Units[Blocks]. Units[Blocks] reflects differences between units within a block and the sums of squares for this source is partitioned into Treatments and Residual sums of squares.

Source	df	Source in two-way ANOVA
Blocks	$b-1$	Between Blocks
Units[Blocks]	$b(t-1)$	
Treatments	$t-1$	Between Treatments
Residual	$(b-1)(t-1)$	Error
Total	$bt-1$	Total

The advantage of the table we have presented is that it exhibits the confounding in the experiment. The indenting of Treatments under Units[Blocks] signifies that treatment differences are confounded or “mixed-up” with unit differences as a result of the randomization of treatments to units within blocks. Also, the Residual reflects differences between the units within a block once the treatment differences have been removed or subtracted off. That is, it represents inherent variability in the units used in the experiment; the expected mean squares for this analysis will confirm this. This contrasts with the usual interpretation of the Error source in the traditional table, which is that it arises from differences in the treatment responses between the blocks. In our analysis we are assuming that the treatment responses are the same in each block, an assumption that we will need to check. It is claimed that the analysis presented here gives a clearer indication of the origins of the sources of variation that are affecting the response variable.

f) Computation of the ANOVA in R

The expressions for analyzing a randomized complete block design are summarized in Appendix C, *Analysis of designed experiments in R*. Here we examine the expressions for obtaining the analysis of variance and their output for the example.

Example IV.1 Penicillin yield (continued)

First the data is entered into a data frame so that it contains the factors Blend, Flask and Treat and the numeric vector Yield as illustrated in the figure below.

As for the completely randomized design, you can use the `aov` function, either with or without the `Error` as part of the model. In this experiment the uncontrolled variation is made up of Blend differences and difference between Flasks within Blends — the latter we denote Flasks[Blends]. R provides a shorthand for this: Blend/Flask is a shorthand that expands to Blend + Blend:Flask. Outputs for both are given below.

```
> RCBDPen.dat
   Blend Flask Treat Yield
1      1     1     A    89
2      1     2     B    88
3      1     3     C    97
4      1     4     D    94
5      2     1     A    84
6      2     2     B    77
7      2     3     C    92
8      2     4     D    79
9      3     1     A    81
10     3     2     B    87
11     3     3     C    87
12     3     4     D    85
13     4     1     A    87
14     4     2     B    92
15     4     3     C    89
16     4     4     D    84
17     5     1     A    79
18     5     2     B    81
19     5     3     C    80
20     5     4     D    88
> RCBDPen.aov <- aov(Yield ~ Blend + Treat + Error(Blend/Flask), RCBDPen.dat)
> summary(RCBDPen.aov)

Error: Blend
      Df Sum Sq Mean Sq
Blend  4    264      66

Error: Blend:Flask
          Df  Sum Sq Mean Sq F value Pr(>F)
Treat      3  70.000  23.333  1.2389 0.3387
Residuals 12 226.000  18.833
> #Compute Blend F and p
> Blend.F <- 66/18.833
> Blend.p <- 1-pf(Blend.F, 4, 12)
> data.frame(Blend.F,Blend.p)
   Blend.F   Blend.p
1 3.504487 0.0407441
> RCBDPen.NoError.aov <- aov(Yield ~ Blend + Treat, RCBDPen.dat)
> summary(RCBDPen.NoError.aov)
            Df  Sum Sq Mean Sq F value  Pr(>F)
Blend        4 264.000  66.000  3.5044 0.04075
Treat        3  70.000  23.333  1.2389 0.33866
Residuals   12 226.000  18.833
```

The ANOVA table from the expression that includes `Error` in the model resembles our table while that without is like the traditional ANOVA table. The advantage of the latter is that it includes the F and p-values for Blend. However, the instructions for manually computing these F and p values are also given. It is controversial whether a test of Blend should be performed. We prefer the form that includes the `Error` function, but note that Blend occurs both outside and inside the `Error` function. This is necessary to get correct fitted vales for the diagnostic checking described in section IV.D, *Diagnostic Checking*.

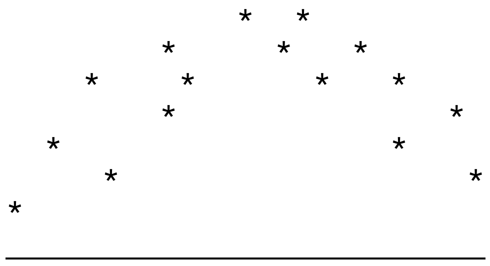
IV.D Diagnostic checking

Again, we have assumed a model on which the analysis outlined above is based, namely, that $\mathbf{Y} \sim \mathrm{N}\left(\boldsymbol{\psi}, \sigma^2\mathbf{I}\right)$ where, for the maximal model, $\boldsymbol{\psi}_{\mathrm{B+T}} = E\left[\mathbf{Y}\right] = \mathbf{X}_{\mathrm{B}}\boldsymbol{\beta} + \mathbf{X}_{\mathrm{T}}\boldsymbol{\tau}$. For this model to be appropriate requires a similar set of behaviours as for the CRD:

a) the response is operating additively (see section *IV.B, Indicator variable models and estimation for an RCBD*) as specified by the maximal model, that is, that a treatment has about the same additive effect on each unit;
b) that the variability of the units within the block are the same for each block;
c) each observation displays the covariance implied by the model (independence for Blocks fixed and equal correlation within blocks for Blocks random); and
d) that the response of the units is normally distributed.

The same set of diagnostic plots as for the analysis of a CRD can be used. Thus, we can obtain Residual-versus-fitted-values and Normal probability plots.

A particular pattern to look out for in the Residual-versus-fitted-values plot for this type of design is evidence of a curvilinear relationship, that is, a plot such as the following:



systematic trend in residuals

Such a plot indicates that there is nonadditivity between the blocks and treatments such as for higher units the treatments tend to have greater effects than for lower units. Such nonadditivity, or interaction, may be transformable by take logs, square root or reciprocals of the data and analyzing these. Another type of block-treatment interaction would occur where say a particular blend had a poison in it that affected only process B, then only the observation corresponding to that particular combination of blend and treatment would be affected and it would be extremely low leading to an extreme residual.

It is possible to test for transformable nonadditivity using Tukey's one-degree-of-freedom-for-nonadditivity, a test that can be used with any design or in a regression situation that involves an additive expectation model, i.e. a model with at least two terms that are summed. It involves detecting whether or not there is a curvilinear relationship between the residuals and fitted values. It is calculated by:

1. squaring the fitted values
2. obtaining the residuals, $\mathbf{e}_2$, from fitting the model to the squared fitted values
3. regressing the original residuals, $\mathbf{e}$, on the new residuals, $\mathbf{e}_2$.

For this, and subsequent designs, diagnostic checking should be based on the two plots and, where possible, this one degree-of-freedom. It cannot be computed for a completely randomized design involving a single terms for treatments as this is not a model with additive terms.

An R function, `tukey.1df`, for calculating the one-degree-of-freedom-for-nonadditivity is available in the `dae` library. It's usage is:

```
tukey.1df(aov.obj, data, error.term="within")
```

where `aov.obj` is an `aov` object or `aovlist` object created from a call to `aov`,
`data` is optional and is a `data.frame` containing the original response variable and factors used in the call to `aov`, and
`error.term` is the `error.term` whose residuals are to be tested for nonadditivity.

Example IV.1 Penicillin yield (continued)

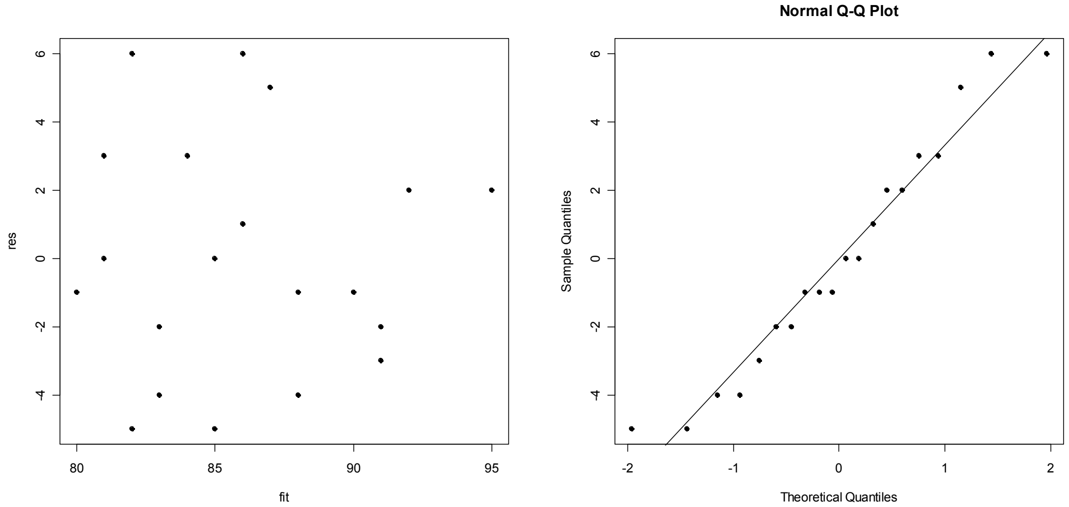
For the example the R output and plots are:

```
> #
> # Diagnostic checking
> #
> res <- resid.errors(RCBDPen.aov)
> fit <- fitted.errors(RCBDPen.aov)
> data.frame(Blend,Flask,Treat,Yield,res,fit)
   Blend Flask Treat Yield           res fit
1      1     1     A    89 -1.000000e+00  90
2      1     2     B    88 -3.000000e+00  91
3      1     3     C    97  2.000000e+00  95
4      1     4     D    94  2.000000e+00  92
5      2     1     A    84  3.000000e+00  81
6      2     2     B    77 -5.000000e+00  82
7      2     3     C    92  6.000000e+00  86
8      2     4     D    79 -4.000000e+00  83
9      3     1     A    81 -2.000000e+00  83
10     3     2     B    87  3.000000e+00  84
11     3     3     C    87 -1.000000e+00  88
12     3     4     D    85 -2.392617e-15  85
13     4     1     A    87  1.000000e+00  86
14     4     2     B    92  5.000000e+00  87
15     4     3     C    89 -2.000000e+00  91
16     4     4     D    84 -4.000000e+00  88
17     5     1     A    79 -1.000000e+00  80
18     5     2     B    81 -2.614662e-15  81
19     5     3     C    80 -5.000000e+00  85
20     5     4     D    88  6.000000e+00  82
> plot(fit, res, pch=16)
> qqnorm(res, pch=16)
> qqline(res)
> tukey.1df(RCBDPen.aov, RCBDPen.dat, error.term="Blend:Flask")
$Tukey.SS
[1] 2.001082

$Tukey.F
[1] 0.0982679

$Tukey.p
[1] 0.7597822

$Devn.SS
[1] 223.9989
```



From these plots, it would appear that there is no serious departure from the assumptions.

The analysis of variance table incorporating the one-degree-of-freedom is:

Source	df	SSq	MSq	E[MSq]	F	Prob
Blends	4	264	66.0	$\sigma^2 + q_{\mathrm{B}}(\psi)$	3.50	0.041
Flasks[Blends]	15	296				
Treatments	3	70	23.3	$\sigma^2 + q_{\mathrm{T}}(\psi)$	1.24	0.339
Residual	12	226	18.8	σ^2		
Nonadditivity	1	2.0	2.0		0.10	0.760
Deviation	11	224	20.4			
Total	19	560				

The hypotheses for the one-degree-of-freedom is:

H_0: Blends and Treatments are additive
H_1: Blends and Treatments are nonadditive

The null hypothesis cannot be rejected and there is no evidence of transformable nonadditivity.

IV.E Treatment differences

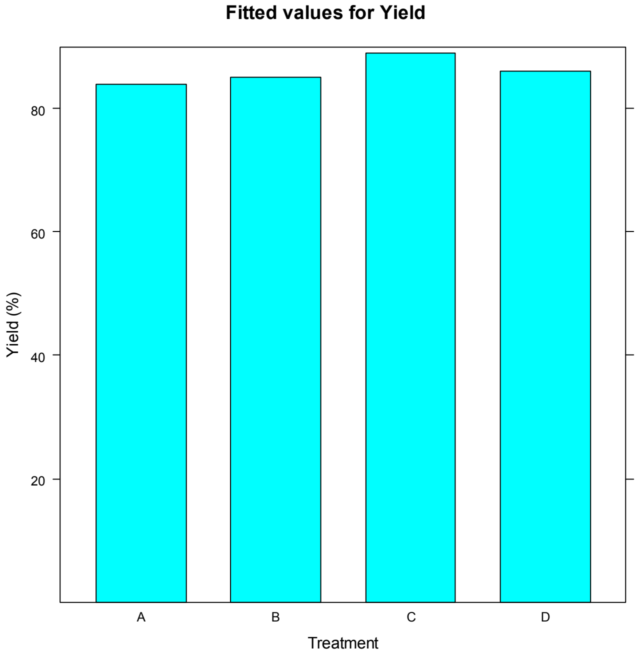
For the purposes of the scientist the effect of the blocks are not of primary interest. Rather, attention is likely to be focused on treatment differences which can be investigated using the treatment means. The discussion of multiple comparisons and submodels for the analysis of a CRD applies here also.

Example IV.1 Penicillin yield (continued)

The treatment means are:

Treatment			
A	B	C	D
84	85	89	86

As the treatment levels are qualitative a multiple comparison procedure would be used to examine the differences between the treatments if they were significant. However they are not significantly different so that we shall not apply such a procedure. The following bar chart illustrates this conclusion.



IV.F Fixed versus random effects

a) Another maximal model for the RCBD

Another point about the RCBD is that there are two alternative maximal models for it. In matrix terms, the original maximal model is

$$E[\mathbf{Y}] = \mathbf{X}_{\mathrm{B}}\boldsymbol{\beta} + \mathbf{X}_{\mathrm{T}}\boldsymbol{\tau} \text{ and } \mathbf{V} = \sigma^{\mathbf{2}}\mathbf{I}_n$$

and another maximal model is:

$$E[\mathbf{Y}] = \mathbf{X}_{\mathrm{T}}\boldsymbol{\tau} \text{ and } \mathbf{V} = \sigma^{\mathbf{2}}\mathbf{I}_n + \sigma_{\mathrm{B}}^2(\mathbf{I}_b \otimes \mathbf{J}_t) = \sigma^{\mathbf{2}}\mathbf{M}_{\mathrm{U}} + t\sigma_{\mathrm{B}}^2\mathbf{M}_{\mathrm{B}}.$$

Note that in part b) of this section we gave $\mathbf{M}_{\mathrm{U}} = \left(\mathbf{I}_b \otimes \mathbf{I}_t\right)$ and $\mathbf{M}_{\mathrm{B}} = t^{-1}\mathbf{I}_b \otimes \mathbf{J}_t$.

The difference between the two models is that, in the alternative model, $\boldsymbol{\beta}$ has been dropped from the expectation of an observation and the covariance of observations from different units in the same block is σ_{B}^2, rather than being zero. The variance matrices of the observations for the two models when b=3, t=4 is shown below.

Variance-covariance matrix for the RCBD — Blocks fixed

						Block						
			I				II				III	
Unit	1	2	3	4	1	2	3	4	1	2	3	4
1	σ^2	0	0	0	0	0	0	0	0	0	0	0
2	0	σ^2	0	0	0	0	0	0	0	0	0	0
3	0	0	σ^2	0	0	0	0	0	0	0	0	0
4	0	0	0	σ^2	0	0	0	0	0	0	0	0
1	0	0	0	0	σ^2	0	0	0	0	0	0	0
2	0	0	0	0	0	σ^2	0	0	0	0	0	0
3	0	0	0	0	0	0	σ^2	0	0	0	0	0
4	0	0	0	0	0	0	0	σ^2	0	0	0	0
1	0	0	0	0	0	0	0	0	σ^2	0	0	0
2	0	0	0	0	0	0	0	0	0	σ^2	0	0
3	0	0	0	0	0	0	0	0	0	0	σ^2	0
4	0	0	0	0	0	0	0	0	0	0	0	σ^2

— Blocks random

						Block						
			I				II				III	
Unit	1	2	3	4	1	2	3	4	1	2	3	4
1	$\sigma^2 + \sigma_{\mathrm{B}}^2$	σ_{B}^2	σ_{B}^2	σ_{B}^2	0	0	0	0	0	0	0	0
2	σ_{B}^2	$\sigma^2 + \sigma_{\mathrm{B}}^2$	σ_{B}^2	σ_{B}^2	0	0	0	0	0	0	0	0
3	σ_{B}^2	σ_{B}^2	$\sigma^2 + \sigma_{\mathrm{B}}^2$	σ_{B}^2	0	0	0	0	0	0	0	0
4	σ_{B}^2	σ_{B}^2	σ_{B}^2	$\sigma^2 + \sigma_{\mathrm{B}}^2$	0	0	0	0	0	0	0	0
1	0	0	0	0	$\sigma^2 + \sigma_{\mathrm{B}}^2$	σ_{B}^2	σ_{B}^2	σ_{B}^2	0	0	0	0
2	0	0	0	0	σ_{B}^2	$\sigma^2 + \sigma_{\mathrm{B}}^2$	σ_{B}^2	σ_{B}^2	0	0	0	0
3	0	0	0	0	σ_{B}^2	σ_{B}^2	$\sigma^2 + \sigma_{\mathrm{B}}^2$	σ_{B}^2	0	0	0	0
4	0	0	0	0	σ_{B}^2	σ_{B}^2	σ_{B}^2	$\sigma^2 + \sigma_{\mathrm{B}}^2$	0	0	0	0
1	0	0	0	0	0	0	0	0	$\sigma^2 + \sigma_{\mathrm{B}}^2$	σ_{B}^2	σ_{B}^2	σ_{B}^2
2	0	0	0	0	0	0	0	0	σ_{B}^2	$\sigma^2 + \sigma_{\mathrm{B}}^2$	σ_{B}^2	σ_{B}^2
3	0	0	0	0	0	0	0	0	σ_{B}^2	σ_{B}^2	$\sigma^2 + \sigma_{\mathrm{B}}^2$	σ_{B}^2
4	0	0	0	0	0	0	0	0	σ_{B}^2	σ_{B}^2	σ_{B}^2	$\sigma^2 + \sigma_{\mathrm{B}}^2$

Notice that, for Blocks random, the covariance between units from the same block is non-zero and is equal for all blocks.

One says that, in the original model, Blocks is fixed whereas, in the revised model, Blocks is random.

Definition IV.2: A factor will be designated as **random** if it is considered appropriate to use a probability distribution function to describe the distribution of effects associated with the population set of levels. ■

Definition IV.3: A factor will be designated as **fixed** if it is considered appropriate to have the effects associated with the population set of levels for the factor differ in an arbitrary manner, rather than being distributed according to a regularly-shaped probability distribution function. ■

As far as the model is concerned, random effects are modelled using terms in the variation model and fixed effects are modelled using terms in the expectation model. So when we are deciding whether a factor is random or fixed, we are choosing which mathematical model best describes the population distribution for the response variable. The above definitions provide us with a basis for making the choice. One needs to consider the population set of levels and how the set of response variable effects corresponding to this set of levels behaves. To be classified as random, we require that the set of population levels is large in number and that the effects are "well-behaved" so that a regularly-shaped probability distribution function with some variance is appropriate for describing them. On the other hand, fixed effects do not have the restrictions that are placed on random effects. There might be a small or a large number of levels in the population and their effects do not have to conform to a regularly-shaped probability distribution function because the model allows for arbitrary differences between them.

It is clear that, if it is anticipated that the effects of a factor will display a systematic trend, then this must be modelled using an expectation model, perhaps involving polynomial submodels. Also, the factor for a small set of treatments that are to be compared would be modelled using a term in the expectation model. In both cases, it seems inappropriate to model the effects as being, say normally distributed, with some variance σ_{T}^2 — the pattern in the treatment effects may well be quite irregular and there is no interest in the form of this distribution.

However, the effects from individual units treated alike (for example, animals, plots of land, runs of a chemical reactor) are anticipated to arise randomly and the effects could well follow a probability distribution, say a normal distribution. Hence it is appropriate to model them via a term in the variation model.

Notwithstanding any of this, you must always model terms to which other terms have been randomized as random effects. For example, because Treatments are randomized to Units (within Blocks) in an RCBD, Units must be a random factor.

In practice

- Random if
  1. large number of population levels and
  2. random behaviour
- Fixed if
  i. small or large number of population levels and
  ii. systematic or other non-random behaviour

What about Block effects in the RCBD? It could be either depending on the anticipated effects of the blocks. For the Block factor to be random, the effects associated with the population set of blocks would have to be capable of being described using a probability distribution, such as the normal probability distribution. Otherwise they should be designated as fixed.

Suppose the blocks are groups of plots and are contiguous; if it is anticipated that there might be some systematic trend between the plots, such as a fertility trend, a term in the expectation model would be more appropriate than a term in the variation model. The distribution of block effects cannot be regarded as a random sample — they display a systematic pattern. The factor Blocks should be designated as fixed.

However, suppose each block is in a separate location to other blocks and could be regarded as a random sample of all blocks obtained by dividing up the whole area under study. It seems likely that the population block effects could be described by a probability distribution such as the normal distribution and the factor Blocks could be designated as random. If there is some doubt about this, it is safest to not make the assumption of some probability distribution and to designate the factor as fixed.

Example IV.1 Penicillin yield (continued)

Should Blends be designated as fixed or random? It was said at the outset that it was expected that there would be a lot of variability from blend to blend — that is why the RCBD was employed. However, a systematic pattern in the average yields of the blends cannot be anticipated. Rather, it seems reasonable that the effects of the population set of blends can be described by a probability distribution. So Blends should be a random factor. In this case the analysis needs to be revised, using a call to `aov` in which Blends is not included outside the `Error` function.

```
RCBDPen.aov <- aov(Yield ~ Treat + Error(Blend/Flask), RCBDPen.dat)
```

This will change the fitted values and Tukey's one-degree-of-freedom-for-nonadditivity will no longer be applicable. ■

b) Estimation and analysis of variance for Blocks random

The estimator of the expected values under the model $\boldsymbol{\psi}_{\mathrm{T}} = E\left[\mathbf{Y}\right] = \mathbf{X}_{\mathrm{T}}\boldsymbol{\tau}$ and $\mathbf{V} = \sigma^{2}\mathbf{M}_{\mathrm{U}} + t\sigma_{\mathrm{B}}^{2}\mathbf{M}_{\mathrm{B}}$ are $\hat{\boldsymbol{\psi}}_{\mathrm{T}} = \overline{\mathbf{T}} = \mathbf{M}_{\mathrm{T}}\mathbf{Y}$, the same as for the model $\boldsymbol{\psi}_{\mathrm{T}} = E\left[\mathbf{Y}\right] = \mathbf{X}_{\mathrm{T}}\boldsymbol{\tau}$ and $\mathbf{V} = \sigma^{2}\mathbf{I}_{n}$.

The form of the analysis of variance table for the RCBD is the same irrespective of whether Blocks is fixed or random. However, the expected mean squares differ, as shown in the following table. Also, as the expectation model is no longer involves the sum of two terms, Tukey's one-degree-of-freedom for nonadditivity is no longer applicable.

		E[MSq]	
Source	df	Blocks Fixed	Blocks Random
Blocks	b-1	$\sigma^2 + q_{\text{B}}(\psi)$	$\sigma^2 + t\sigma_{\text{B}}^2$
Units[Blocks]	b(t-1)		
Treatments	t-1	$\sigma^2 + q_{\text{T}}(\psi)$	$\sigma^2 + q_{\text{T}}(\psi)$
Residual	(b-1)(t-1)	σ^2	σ^2
Total	bt-1		

In the case of the RCBD, the ramifications of the difference between the two sets of expected mean squares is small. All that has happened is that the term $q_{\text{B}}(\boldsymbol{\psi}) = \frac{\boldsymbol{\psi}'\mathbf{Q}_{\text{B}}\boldsymbol{\psi}}{b-1} = \sum t\left(\beta_i - \bar{\beta}_{.}\right)^2 \big/ (b-1)$ has become $t\sigma_{\text{B}}^2$. Also the hypotheses for Blocks becomes:

$$\text{H}_0\text{: } \sigma_{\text{B}}^2 = 0$$

$$\text{H}_1\text{: } \sigma_{\text{B}}^2 \neq 0$$

That is, the test based on this hypothesis asks can the σ_{B}^2 term be dropped from variance model $\mathbf{V} = \sigma^{\mathbf{2}}\mathbf{M}_{\text{U}} + t\sigma_{\text{B}}^2\mathbf{M}_{\text{B}}$? The F-statistic for testing this hypothesis is again the ratio of the Block and Residual mean squares. Thus the test for both fixed and random block effects are the same. This is not always the case.

IV.G Generalized randomized complete block design

(ref. Steel and Torrie, sec. 9.8; Addelman, 1969, Amer. Stat.)

A generalized randomized complete block design is the same as the ordinary randomized complete block design, except that each treatment occurs more than once in a block — we deal only with the case in which all treatment occur the same number of times. As before we let b be the number of blocks and t the number of treatments. In addition let k denote the number of units per block and g the number of times a treatment occurs in a block — that is, $k = t \times g$ and $n = b \times k$. The R expressions for obtaining a layout for this design is given in Appendix B, *Randomized layouts and sample size computations in R*.

The advantage of this design is that you end up with more degrees of freedom for the Residual when compared to the standard randomized complete block design. Also, you can test for Block:Treatment interaction, as is discussed in chapter VI, *Determining the analysis of variance table*. The disadvantage of the design is that it has larger blocks so that it is likely that the units within a block will be less homogeneous than would be the case if a standard randomized complete block design with smaller blocks were employed.

The model for the generalized RCBD, without the Block:Treatment interaction, is virtually the same as that for the RCBD so that, in this case, the analyses of variance are similar. Thus, depending on whether Blocks are fixed or random the maximal model, would be chosen from the following two alternatives:

Blocks fixed and Plots random: $E[\mathbf{Y}] = \mathbf{X}_\mathrm{B}\boldsymbol{\beta} + \mathbf{X}_\mathrm{T}\boldsymbol{\tau}$ and $\mathbf{V} = \sigma^2\mathbf{I}_n$ and

Blocks and Plots random: $E[\mathbf{Y}] = \mathbf{X}_\mathrm{T}\boldsymbol{\tau}$ and

$$\mathbf{V} = \sigma^2\mathbf{I}_n + \sigma^2_\mathrm{B}\left(\mathbf{I}_b \otimes \mathbf{J}_k\right) = \sigma^2\mathbf{M}_\mathrm{U} + k\sigma^2_\mathrm{B}\mathbf{M}_\mathrm{B}\,.$$

For Blocks and Plots random, the analysis of variance table is

Source	df	SSq	E[MSq]
Blocks	$b-1$	$\mathbf{Y}'\mathbf{Q}_\mathrm{B}\mathbf{Y}$	$\sigma^2_\mathrm{BU} \quad +k\sigma^2_\mathrm{B}$
Units[Blocks]	$b(k-1)$	$\mathbf{Y}'\mathbf{Q}_\mathrm{BU}\mathbf{Y}$	
Treatments	$t-1$	$\mathbf{Y}'\mathbf{Q}_\mathrm{T}\mathbf{Y}$	$\sigma^2_\mathrm{BU} \quad +q_\mathrm{T}(\boldsymbol{\psi})$
Residual	$b(k-1)-(t-1)$	$\mathbf{Y}'\mathbf{Q}_{\mathrm{BU}_\mathrm{Res}}\mathbf{Y}$	σ^2_BU
Total	$bk-1$		

For only Plots random, only the expected mean square for Blocks changes from that in the above analysis table — it becomes $\sigma^2_\mathrm{BU} + q_\mathrm{B}(\boldsymbol{\psi})$.

The R expressions for analysing data from an experiment based on this design are the same as for the standard randomized complete block design.

Example IV.2 Design for a wheat experiment

For example, suppose four treatments are to be compared when applied to a new variety of wheat. The researcher wants to employ a generalized randomized complete block design with 12 plots in each of 2 blocks so that each treatment is replicated 3 times in each block. Hence, $b = 2$, $t = 4$ and $s = 3$ so that $k = 4 \times 3 = 12$ and $n = 2 \times 12 = 24$. A possible layout for this experiment is shown in the table given below.

Layout for a generalized randomized complete block experiment

		Plots											
		1	2	3	4	5	6	7	8	9	10	11	12
Blocks	I	C	D	D	C	B	B	A	A	D	A	B	C
	II	D	A	D	C	A	D	B	A	B	B	C	C

The yield of wheat from each plot was measured.

The model for the example with Blocks and Plots random is:

$$E\left[\mathbf{Y}\right]=\mathbf{X}_{\mathrm{T}}\tau \text{ and } \mathbf{V}=\sigma^{\mathbf{2}}\mathbf{I}_{24}+\sigma_{\mathrm{B}}^{2}\left(\mathbf{I}_{2}\otimes\mathbf{J}_{12}\right)=\sigma^{\mathbf{2}}\mathbf{M}_{\mathrm{U}}+12\sigma_{\mathrm{B}}^{2}\mathbf{M}_{\mathrm{B}}.$$

The corresponding analysis of variance table

Source	df	SSq	E[MSq]	
Blocks	1	$\mathbf{Y}'\mathbf{Q}_{\mathrm{B}}\mathbf{Y}$	σ_{BP}^{2}	$+12\sigma_{\mathrm{B}}^{2}$
Plots[Blocks]	22	$\mathbf{Y}'\mathbf{Q}_{\mathrm{BP}}\mathbf{Y}$		
Treatments	3	$\mathbf{Y}'\mathbf{Q}_{\mathrm{T}}\mathbf{Y}$	σ_{BP}^{2}	$+q_{\mathrm{T}}\left(\psi\right)$
Residual	19	$\mathbf{Y}'\mathbf{Q}_{\mathrm{BP}_{\mathrm{Res}}}\mathbf{Y}$	σ_{BP}^{2}	
Total	23			

Note that a randomized complete block design with $b=6$ and $t=4$ would also have $n=6\times4=24$ but would have $\left(b-1\right)\left(t-1\right)=5\times3=15$ Residual degrees of freedom.■

IV.H Summary

In this chapter we have:

- described how to design an experiment using a randomized complete block design and a generalized randomized complete block design;
- formulated a linear model using indicator variables to describe the results from a completely randomized design; given the estimators of the parameters in the linear model, the expected values and the random errors as functions of **M** or mean operator matrices;
- outlined an hypothesis test for choosing between four expectation models using the ANOVA hypothesis test procedure outlined in chapter I;
  - the partition of the total sums of squares was given with the sums of squares expressed as the sums of squares of the elements of vectors and as quadratic forms where the matrices of the quadratic forms, **Q** matrices, are symmetric idempotents;
  - the expected mean squares under the alternative expectation models are used to justify the choice of F test statistic;
- shown how to obtain a layout and the analysis of variance in R;
- discussed procedures for checking the adequacy of the proposed models;
- introduced the concept of fixed and random factors and outlined two different models based on whether Blocks is fixed or random.

IV.I Exercises

IV.1 Given that $\mathbf{D}_{\mathrm{B}} = \mathbf{Y} - \overline{\mathbf{B}}$, $\mathbf{T}_{e} = \overline{\mathbf{T}} - \overline{\mathbf{G}}$ and $\mathbf{D}_{\mathrm{B+T}} = \mathbf{Y} - \overline{\mathbf{B}} - \overline{\mathbf{T}} + \overline{\mathbf{G}}$, show that $\mathbf{D}_{\mathrm{B}} - \mathbf{T}_{e} = \mathbf{D}_{\mathrm{B+T}}$. Also show that $\mathbf{Y}'\mathbf{Q}_{\mathrm{BU}}\mathbf{Y} - \mathbf{Y}'\mathbf{Q}_{\mathrm{T}}\mathbf{Y} = \mathbf{Y}'\mathbf{Q}_{\mathrm{BU}_{\mathrm{Res}}}\mathbf{Y}$.

IV.2 Prove that $\left(\overline{\mathbf{T}} - \overline{\mathbf{G}}\right)'\left(\overline{\mathbf{T}} - \overline{\mathbf{G}}\right) = \mathbf{Y}'\mathbf{Q}_{\mathrm{T}}\mathbf{Y}$. You are given that $\mathbf{Q}_{\mathrm{T}}$ is symmetric and idempotent.

IV.3 The highway department wants to study four different types of paving for possible use on interstate highways. Three different locations, with different weather conditions and traffic patterns, are chosen. Each section constitutes a block and is to be divided into four strips and the four paving types, labelled A–D, randomly assigned to the strips within a location. Use R to obtain a randomized layout for the experiment, using a `seed` of 832 and storing it in a data frame named `RCBDHway.lay`.

Suppose that the trial is conducted and the amount of wear after one year is measured. The data obtained is given in the following table.

		Strip			
		1	2	3	4
Location	1	D	A	B	C
		40.2	50.0	38.0	49.7
	2	C	D	B	A
		48.5	32.8	39.3	42.7
	3	A	C	D	B
		51.9	53.5	51.1	46.3

Combine the data frame `RCBDHway.lay` with a vector `Wear`, that contains the values from the above table, to form a new data frame named (or copy using `RCBDHway.dat`. Use R to obtain boxplots for Locations and Types and to perform an analysis of variance of the data, including diagnostic checking.

IV.4 An experiment was conducted to study the effects of temperature on the life (in hours) of a component. An RCBD was employed with five ovens forming the blocks. Four temperatures were randomly assigned to four runs within each oven. The following results were recorded:

		Temperature (degrees)			
		200	300	400	500
Oven	I	340	324	307	274
	II	361	338	312	281
	III	346	328	298	276
	IV	358	332	315	285
	V	343	321	294	269

What is the response variable for this experiment?

There are two factors other than Temperature in this experiment. What are they? Use `fac.gen` to generate these two factors in a data frame.

Then add Temperature and Life to the data frame and produce the boxplots for an initial exploration of the data.

Assuming Oven is random, perform an analysis of variance on the data using R, including diagnostic checking and an appropriate examination of mean differences. In particular what temperature would you recommend be used with this component.

IV.5 In evaluating insecticides, the numbers of living adult plum curculios emerging from separate caged areas of treated soil were observed. The results are shown in the table below.

		Insecticide					
		Lindane	Dieldrin	Aldrin	EPN	Chlordane	Check
Block	1	14	7	6	95	37	212
	2	6	1	1	133	31	172
	3	8	0	1	86	13	202
	4	36	15	4	115	69	217

This data, including the factors, is contained in *RCBDInse.dat.rda* available from the Statistical Modelling resources web site. Open this file and then perform an analysis of variance on the data using R, including diagnostic checking and the appropriate examination of mean differences.