DESIGN AND MIXED-MODEL ANALYSIS OF EXPERIMENTS

V. Randomized Complete Block Design (RCBD)

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

The RCBD was defined in the last chapter. We want to consider its analysis.

Example V.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 by using the same blend in four flasks and randomizing the four treatments to these four flasks. Altogether five blends were utilized and the layout of the experiment was as follows:

		Flask			
		1	2	3	4
Blend	1	A	C	B	D
	2	D	B	C	A
	3	B	A	D	C
	4	A	C	B	D
	5	C	D	A	B

The yields of penicillin, in nonrandom order, were:

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

V.A Maximal model for an RCBD

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:

$$\psi = E[\mathbf{Y}] = \mathbf{X}_B\beta + \mathbf{X}_T\tau \text{ and } \text{var}[\mathbf{Y}] = \sigma^2\mathbf{I}.$$

where $\mathbf{Y}$ is the n-vector of response variables,

β is the b-vector of parameters specifying a different mean response for each block,

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

τ is the t-vector of parameters specifying a different mean response for each treatment,

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

Example V.1 Penicillin yield (continued)

For the example, $b = 5$ and $t = 4$ so that $n = 20$. If the yields are placed in a vector with all those for the same blend placed together, then the vectors and matrices in the expectation model are:

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

■

The model for the expectation is still of the form $E[\mathbf{Y}]=\mathbf{X}\theta$, with $\mathbf{X}=[\mathbf{X}_B \quad \mathbf{X}_T]$ and $\theta'=[\beta' \quad \tau']$. However, the model is not of full rank because the sums of the columns of both $\mathbf{X}_B$ and $\mathbf{X}_T$ are equal to $\mathbf{1}_n$. Consequently, the rank($\mathbf{X}$) is $b+t-1$. Such a model is referred to as a **less than full rank model**. The differences between this model and the full rank models that we have considered previously are as follows:

1. In the full rank model it is assumed that the parameters specified in the model are unique. That is there exists exactly one set of real numbers $\{\theta_0,\theta_1,\ldots,\theta_p\}$ that describes the system. In the less than full rank model there are infinitely many sets of real numbers that describe the system. For our example there are infinitely many choices for $\{\beta_1,\ldots,\beta_b,\tau_1,\ldots,\tau_t\}$. The model parameters are said to be **nonidentifiable**.
2. In the full rank model $\mathbf{X}'\mathbf{X}$ is nonsingular and so $(\mathbf{X}'\mathbf{X})^{-1}$ exists and the normal equations have the one solution $\hat{\theta}=(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$. In a less than full rank model thee are infinitely many solutions to the normal equations.
3. In the full rank model all linear functions of $\{\theta_0,\theta_1,\ldots,\theta_p\}$ can be estimated unbiasedly. In the less than full rank model this is not the case: some functions can while others cannot. It will be our task to identify which functions have estimated values that are the same regardless of the estimators used to estimate them.

To illustrate that the parameters are nonidentifiable, suppose an experiment with 2 blocks and 2 treatments is conducted and the following information is known about the parameters:

$$\beta_1 + \tau_1 = 10$$
$$\beta_1 + \tau_2 = 15$$
$$\beta_2 + \tau_1 = 12$$
$$\beta_2 + \tau_2 = 17$$

Then the parameters are not identifiable for

1. if $\beta_1 = 5$, then $\tau_1 = 5$, $\tau_2 = 10$ and $\beta_2 = 7$
2. if $\beta_1 = 6$, then $\tau_1 = 4$, $\tau_2 = 9$ and $\beta_2 = 8$

Clearly we can pick a value for any one of the parameters and then find values of the other parameters that satisfy the above equations and so there are infinitely many possible values for the parameters. However, no matter what values are taken for β_1, β_2, τ_1 and τ_2, the value of $\beta_2 - \beta_1 = 2$ and of $\tau_2 - \tau_1 = 5$; these functions are invariant.

V.B Estimation of the parameters of the maximal model for an RCBD

To estimate the parameters of a less than full rank model we need to extend the estimation theory presented in section II. This will involve obtaining solutions to the normal equations using generalized inverses.

a) Estimation for the less than full rank model

Introduction to generalized inverses

Suppose that we have a system of n linear equations in q unknowns such as $\mathbf{Ax} = \mathbf{y}$ where $\mathbf{A}$ is an $n \times q$ matrix of real numbers, $\mathbf{x}$ is a q-vector of unknowns and $\mathbf{y}$ is a n-vector of real numbers. There are 3 possibilities:

1. the system is inconsistent and has no solution;
2. the system is consistent and has exactly one solution;
3. the system is consistent and has many solutions.

Consider the following equations:

$$x_1 + 2x_2 = 7$$
$$3x_1 + 6x_2 = 21$$

They are consistent in that the solution of one will also satisfy the second because the second equation is just 3 times the first. However, the following equations are inconsistent in that it is impossible to satisfy both at the same time:

$$x_1 + 2x_2 = 7$$
$$3x_1 + 6x_2 = 24$$

Generally to be consistent any linear relations on the left of the equations must also be satisfied by the righthand side of the equation.

Theorem V.1: Let $E[\mathbf{Y}] = \mathbf{X}\theta$ be a linear model for the expectation. Then the system of normal equations

$$(\mathbf{X}'\mathbf{X})\theta = \mathbf{X}'\mathbf{Y}$$

is consistent.

Proof: not given. ■

Definition V.1: Let **A** be an $n \times q$ matrix. A $q \times n$ matrix $\mathbf{A}^-$ such that

$$\mathbf{A}\mathbf{A}^-\mathbf{A} = \mathbf{A}$$

is called a **generalised inverse** for **A**. ■

Any matrix **A** has a generalized inverse but it is not unique unless **A** is nonsingular, in which case $\mathbf{A}^- = \mathbf{A}^{-1}$.

An algorithm for finding generalized inverses

To find a generalized inverse $\mathbf{A}^-$ for an $n \times q$ matrix **A** of rank *r*:

1. Find any $r \times r$ minor **M**
2. Find $\mathbf{M}^{-1}$
3. Replace **M** in **A** with $\left(\mathbf{M}^{-1}\right)'$
4. Replace all other entries in **A** with zeros
5. Transpose the resulting matrix.

Properties of generalized inverses

Let **A** be an $n \times q$ matrix of rank *r* with $n \geq q \geq r$. Then

1. $\mathbf{A}^-\mathbf{A}$ and $\mathbf{A}\mathbf{A}^-$ are idempotent.
2. $rank\left(\mathbf{A}^-\mathbf{A}\right) = rank\left(\mathbf{A}\mathbf{A}^-\right) = r$.
3. If $\mathbf{A}^-$ is a generalized inverse of **A**, then $\left(\mathbf{A}^-\right)'$ is a generalized inverse of $\mathbf{A}'$; that is, $\left(\mathbf{A}^-\right)' = (\mathbf{A}')^-$.
4. $\mathbf{A} = \mathbf{A}(\mathbf{A}'\mathbf{A})^-(\mathbf{A}'\mathbf{A})$ and $\mathbf{A}' = (\mathbf{A}'\mathbf{A})(\mathbf{A}'\mathbf{A})^-\mathbf{A}'$.

5. $\mathbf{A}(\mathbf{A}'\mathbf{A})^{-}\mathbf{A}'$ is unique, symmetric and idempotent as it is invariant to the choice of a generalized inverse. Furthermore $rank\left(\mathbf{A}(\mathbf{A}'\mathbf{A})^{-}\mathbf{A}'\right) = r$.

Generalized inverses and the normal equations

Theorem V.2: Let $\mathbf{Ax} = \mathbf{y}$ be consistent. Then $\mathbf{x} = \mathbf{A}^{-}\mathbf{y}$ is a solution to the system where $\mathbf{A}^{-}$ is any generalized inverse for **A**.

Proof: not given. ■

Applying this theorem to the normal equations $(\mathbf{X}'\mathbf{X})\theta = \mathbf{X}'\mathbf{Y}$, it is easy to see that

$$\hat{\theta} = (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{Y}$$

is a solution of the normal equations. Any generalized inverse will generate a solution, however, different generalized inverses result in different solutions.

Theorem V.3: Let $\mathbf{Ax} = \mathbf{y}$ be consistent and let $\mathbf{A}^{-}$ be a any generalized inverse for **A**. Then

$$\mathbf{x}_0 = \mathbf{A}^{-}\mathbf{y} + \left(\mathbf{I} - \mathbf{A}^{-}\mathbf{A}\right)\mathbf{z}$$

is a solution to the system where **z** is an arbitrary p-vector.

Proof: not given. ■

In the context of linear models this theorem implies that every vector of the form

$$\hat{\theta} = (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{Y} + \left\{\mathbf{I}_q - (\mathbf{X}'\mathbf{X})^{-}(\mathbf{X}'\mathbf{X})\right\}\mathbf{z}$$

where $(\mathbf{X}'\mathbf{X})^{-}$ is a generalized inverse for $\mathbf{X}'\mathbf{X}$ and **z** is arbitrary, is a solution to the normal equations. If $\mathbf{z} = \mathbf{0}$, we get the solution obtained via theorem V.2 and if $\mathbf{X}'\mathbf{X}$ is of full rank then we obtain the usual least squares estimators. It can also be proved that any solution can be written in this form.

Estimability

In the less than full rank model the problem we face is that θ cannot be estimated uniquely as it changes for the many different choice of $(\mathbf{X}'\mathbf{X})^{-}$. To overcome this problem we consider those linear functions of θ that are invariant to the choice of $(\mathbf{X}'\mathbf{X})^{-}$ and hence to the particular $\hat{\theta}$ obtained. That is, we consider functions $\mathbf{t}'\theta$ whose estimators $\mathbf{t}'\hat{\theta}$ are invariant to different $\hat{\theta}$s in that their values remain the same regardless of which solution to the normal equations is used. This property is true for all estimable functions.

Definition V.2: Let $E[\mathbf{Y}] = \mathbf{X}\theta$ and $\text{var}[\mathbf{Y}] = \sigma^2\mathbf{I}_n$ where **X** is $n \times q$ of rank $r \le q$. A function $\mathbf{t}'\theta$ is said to be estimable if there exists a vector **c** such that $E[\mathbf{c}'\mathbf{Y}] = \mathbf{t}'\theta$. ■

This definition is saying that, to be estimable, there must be some linear combination of the random variables $Y_1, Y_2, \ldots, Y_n$ that is a linear unbiased estimator for $\mathbf{t}'\theta$.

The next theorem is a Gauss-Markoff theorem applicable to the less than full rank model, with the full rank model being a special case (see theorem II.4).

Theorem V.4: Let $E[\mathbf{Y}] = \mathbf{X}\theta$ and $\text{var}[\mathbf{Y}] = \sigma^2\mathbf{I}_n$ where **X** is $n \times q$ of rank $r \le q$. Let $\mathbf{t}'\theta$ be estimable. Then the best linear unbiased estimator for $\mathbf{t}'\theta$ is $\mathbf{t}'\hat{\theta}$ where $\hat{\theta}$ is any solution to the normal equations. This estimator is invariant to the choice of $\hat{\theta}$.

Proof: not given. ■

So, provided we restrict our attention to estimable functions, our estimators will have the nice properties of being unique and BLU. However, there remains the problem of how to identify estimable functions. The following theorems give us some important cases that will cover the quantities we are most interest.

Theorem V.5: Let $E[\mathbf{Y}] = \mathbf{X}\theta$ and $\text{var}[\mathbf{Y}] = \sigma^2\mathbf{I}_n$ where **X** is $n \times q$ of rank $r \le q$. Each of the elements of $\mathbf{X}\theta$ is estimable.

Proof: not given. ■

That is, a fitted value which is a linear function $\mathbf{x}_i'\theta$, where $\mathbf{x}_i'$ is the q-vector that is the ith row of **X**, is estimable.

Theorem V.6: Let $\mathbf{t}_1'\theta, \mathbf{t}_2'\theta, \ldots, \mathbf{t}_k'\theta$ be a collection of estimable functions. Let $z = a_1\mathbf{t}_1'\theta + a_2\mathbf{t}_2'\theta + \ldots + a_k\mathbf{t}_k'\theta$ be a linear combination of these functions. Then z is estimable and the best linear unbiased estimator for z is $\hat{z} = a_1\mathbf{t}_1'\hat{\theta} + a_2\mathbf{t}_2'\hat{\theta} + \ldots + a_k\mathbf{t}_k'\hat{\theta}$ where $\hat{\theta}$ is any solution to the normal equations.

Proof: not given. ■

That is, a linear combination of estimable functions is itself estimable.

b) Estimation of the parameters for an RCBD

In the next theorem, because each of the parameters in $\theta' = \left[\beta' \quad \tau'\right]$ is not estimable, we consider the estimators for ψ — theorem V.5 tells us that these are estimable.

Theorem V.7: Let **Y** be a *n*-vector of jointly-distributed random variables with

$$\psi = E\left[\mathbf{Y}\right] = \mathbf{X}_B\beta + \mathbf{X}_T\tau \text{ and } \text{var}\left[\mathbf{y}\right] = \sigma^2\mathbf{I}.$$

where β is the *b*-vector of parameters specifying a different mean response for each block,

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

τ is the *t*-vector of parameters specifying a different mean response for each treatment,

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

Then $\hat{\psi} = \mathbf{B} + \mathbf{T} - \mathbf{G}$ where **B**, **T** and **G** are the *n*-vectors of block, treatment and grand means, respectively.

Proof: According to theorem V.2,

$$\hat{\theta} = \left(\mathbf{X}'\mathbf{X}\right)^{-}\mathbf{X}'\mathbf{Y}$$

Now for $\mathbf{X} = \left[\mathbf{X}_B \quad \mathbf{X}_T\right]$, $\mathbf{X}'\mathbf{X} = \begin{bmatrix}\mathbf{X}_B' \\ \mathbf{X}_T'\end{bmatrix}\left[\mathbf{X}_B \quad \mathbf{X}_T\right] = \begin{bmatrix}\mathbf{X}_B'\mathbf{X}_B & \mathbf{X}_B'\mathbf{X}_T \\ \mathbf{X}_T'\mathbf{X}_B & \mathbf{X}_T'\mathbf{X}_T\end{bmatrix} = \begin{bmatrix}t\mathbf{I}_b & \mathbf{J}_{b\times t} \\ \mathbf{J}_{t\times b} & b\mathbf{I}_t\end{bmatrix}$

To obtain the generalized inverse of $\mathbf{X}'\mathbf{X}$ our algorithm tells us that we must take a minor of order $b+t-1$ and obtain its inverse. So we delete any row and column and find its inverse. Suppose we omit the last row and column. Then we require the inverse of

$$\mathbf{Z}'\mathbf{Z} = \begin{bmatrix}\mathbf{X}_B'\mathbf{X}_B & \mathbf{X}_B'\mathbf{X}_{T-1} \\ \mathbf{X}_{T-1}'\mathbf{X}_B & \mathbf{X}_{T-1}'\mathbf{X}_{T-1}\end{bmatrix} = \begin{bmatrix}t\mathbf{I}_b & \mathbf{J}_{b\times(t-1)} \\ \mathbf{J}_{(t-1)\times b} & b\mathbf{I}_{(t-1)}\end{bmatrix}$$

where $\mathbf{X}_{T-1}$ is the $n\times(t-1)$ matrix formed by taking the first $t-1$ columns of $\mathbf{X}_T$.

Our generalized inverse will then be the transpose of the inverse of the minor with a row and column of zeros added. Now for a partitioned symmetric matrix **M** where

$$\mathbf{M} = \begin{bmatrix}\mathbf{A} & \mathbf{B} \\ \mathbf{B}' & \mathbf{D}\end{bmatrix} \text{ then } \mathbf{M}^{-1} = \left(\mathbf{M}^{-1}\right)' = \begin{bmatrix}\mathbf{U} & \mathbf{V} \\ \mathbf{V}' & \mathbf{W}\end{bmatrix} = \begin{bmatrix}\mathbf{A}^{-1} - \mathbf{V}\mathbf{B}'\mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{B}\mathbf{W} \\ -\mathbf{W}\mathbf{B}'\mathbf{A}^{-1} & \left(\mathbf{D} - \mathbf{B}'\mathbf{A}^{-1}\mathbf{B}\right)^{-1}\end{bmatrix}$$

In our case

$$\mathbf{W} = \left(\mathbf{D} - \mathbf{B}'\mathbf{A}^{-1}\mathbf{B}\right)^{-1} = \left(b\mathbf{I}_{(t-1)} - \mathbf{J}_{(t-1)\times b}\tfrac{1}{t}\mathbf{I}_b\mathbf{J}_{b\times(t-1)}\right)^{-1} = \left(b\mathbf{I}_{(t-1)} - \tfrac{b}{t}\mathbf{J}_{(t-1)}\right)^{-1} = \tfrac{1}{b}\left(\mathbf{I}_{(t-1)} + \mathbf{J}_{(t-1)}\right)$$

$$\mathbf{V} = -\mathbf{A}^{-1}\mathbf{B}\mathbf{W} = -\tfrac{1}{t}\mathbf{I}_b\mathbf{J}_{b\times(t-1)}\,\tfrac{1}{b}\left(\mathbf{I}_{(t-1)} + \mathbf{J}_{(t-1)}\right) = -\tfrac{1}{bt}\left(\mathbf{J}_{b\times(t-1)} + (t-1)\mathbf{J}_{b\times(t-1)}\right) = -\tfrac{1}{b}\mathbf{J}_{b\times(t-1)}$$

$$\mathbf{U} = \mathbf{A}^{-1} - \mathbf{V}\mathbf{B}'\mathbf{A}^{-1} = \tfrac{1}{t}\mathbf{I}_b + \tfrac{1}{b}\mathbf{J}_{b\times(t-1)}\mathbf{J}_{(t-1)\times b}\tfrac{1}{t}\mathbf{I}_b = \tfrac{1}{t}\mathbf{I}_b + \tfrac{t-1}{bt}\mathbf{J}_b$$

Consequently,

$$\mathbf{Z}'\mathbf{Z}^{-1} = \begin{bmatrix} \tfrac{1}{t}\mathbf{I}_b + \tfrac{t-1}{bt}\mathbf{J}_b & -\tfrac{1}{b}\mathbf{J}_{b\times(t-1)} \\ -\tfrac{1}{b}\mathbf{J}_{(t-1)\times b} & \tfrac{1}{b}\left(\mathbf{I}_{(t-1)} + \mathbf{J}_{(t-1)}\right) \end{bmatrix}$$

and

$$\left(\mathbf{X}'\mathbf{X}\right)^{-} = \begin{bmatrix} \tfrac{1}{t}\mathbf{I}_b + \tfrac{(t-1)}{bt}\mathbf{J}_b & -\tfrac{1}{b}\mathbf{J}_{b\times(t-1)} & \mathbf{0}_{b\times 1} \\ -\tfrac{1}{b}\mathbf{J}_{(t-1)\times b} & \tfrac{1}{b}\left(\mathbf{I}_{(t-1)} + \mathbf{J}_{(t-1)}\right) & \mathbf{0}_{(t-1)\times 1} \\ \mathbf{0}_{1\times b} & \mathbf{0}_{1\times(t-1)} & 0 \end{bmatrix}$$

Also,

$$\mathbf{X}'\mathbf{Y} = \begin{bmatrix} \mathbf{X}'_{\mathrm{B}} \\ \mathbf{X}'_{\mathrm{T}} \end{bmatrix}\mathbf{Y} = \begin{bmatrix} \mathbf{X}'_{\mathrm{B}} \\ \mathbf{X}'_{\mathrm{T}-1} \\ \mathbf{x}'_t \end{bmatrix}\mathbf{Y} = \begin{bmatrix} \mathbf{X}'_{\mathrm{B}}\mathbf{Y} \\ \mathbf{X}'_{\mathrm{T}-1}\mathbf{Y} \\ \mathbf{x}'_t\mathbf{Y} \end{bmatrix} = \begin{bmatrix} t\mathbf{B} \\ b\mathbf{T}_{t-1} \\ b\mathrm{T}_t \end{bmatrix}$$

where $\mathbf{X}_{\mathrm{T}\text{-}1}$ is the $n\times(t-1)$ matrix formed by taking the first t columns of $\mathbf{X}_{\mathrm{T}}$, $\mathbf{x}_t$ is column t of $\mathbf{X}_{\mathrm{T}}$, $\mathbf{B}$ is the b-vector of block means of $\mathbf{Y}$, $\mathbf{T}_{t-1}$ is the $(t-1)$-vector of the first t–1 treatment means of $\mathbf{Y}$ and T_t is the mean of $\mathbf{Y}$ for treatment t. That is, $\mathbf{X}'\mathbf{Y}$ is a vector of totals — we express them in terms of means for notational clarity.

Hence the nonunique estimators of θ are

$$\begin{aligned} \hat{\theta} &= \left(\mathbf{X}'\mathbf{X}\right)^{-}\mathbf{X}'\mathbf{Y} \\ &= \begin{bmatrix} \tfrac{1}{t}\mathbf{I}_b + \tfrac{(t-1)}{bt}\mathbf{J}_b & -\tfrac{1}{b}\mathbf{J}_{b\times(t-1)} & \mathbf{0}_{b\times 1} \\ -\tfrac{1}{b}\mathbf{J}_{(t-1)\times b} & \tfrac{1}{b}\left(\mathbf{I}_{(t-1)} + \mathbf{J}_{(t-1)}\right) & \mathbf{0}_{(t-1)\times 1} \\ \mathbf{0}_{1\times b} & \mathbf{0}_{1\times(t-1)} & 0 \end{bmatrix}\begin{bmatrix} t\mathbf{B} \\ b\mathbf{T}_{t-1} \\ bT_t \end{bmatrix} \\ &= \begin{bmatrix} \left(\mathbf{I}_b + \tfrac{(t-1)}{b}\mathbf{J}_b\right)\mathbf{B} - \mathbf{J}_{b\times(t-1)}\mathbf{T}_{t-1} \\ -\tfrac{t}{b}\mathbf{J}_{(t-1)\times b}\mathbf{B} + \left(\mathbf{I}_{(t-1)} + \mathbf{J}_{(t-1)}\right)\mathbf{T}_{t-1} \\ 0 \end{bmatrix} \end{aligned}$$

Now $\mathbf{J}_b\mathbf{B} = b\bar{Y}\mathbf{1}_b$, $\mathbf{J}_{(t-1)\times b}\mathbf{B} = b\bar{Y}\mathbf{1}_{(t-1)}$, $\mathbf{J}_{(t-1)}\mathbf{T}_{t-1} = \left(t\bar{Y} - T_t\right)\mathbf{1}_{(t-1)}$, and $\mathbf{J}_{b\times(t-1)}\mathbf{T}_{t-1} = \left(t\bar{Y} - T_t\right)\mathbf{1}_b$ where $\bar{Y}$ is the grand mean of all the observations.

Therefore

$$\hat{\theta} = \begin{bmatrix} \left(\mathbf{I}_b + \frac{(t-1)}{b}\mathbf{J}_b\right)\mathbf{B} - \mathbf{J}_{b\times(t-1)}\mathbf{T}_{t-1} \\ -\frac{t}{b}\mathbf{J}_{(t-1)\times b}\mathbf{B} + \left(\mathbf{I}_{(t-1)} + \mathbf{J}_{(t-1)}\right)\mathbf{T}_{t-1} \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} \mathbf{B} + (t-1)\bar{Y}\mathbf{1}_b - \left(t\bar{Y} - T_t\right)\mathbf{1}_b \\ -t\bar{Y}\mathbf{1}_{(t-1)} + \mathbf{T}_{t-1} + \left(t\bar{Y} - T_t\right)\mathbf{1}_{(t-1)} \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} \mathbf{B} + T_t\mathbf{1}_b - \bar{Y}\mathbf{1}_b \\ \mathbf{T}_{t-1} - T_t\mathbf{1}_{(t-1)} \\ 0 \end{bmatrix}$$

As a result the estimator for the fitted values is

$$\Psi = \mathbf{X}\hat{\theta}$$

$$= \begin{bmatrix} \mathbf{X}_B & \mathbf{X}_{T-1} & \mathbf{x}_t \end{bmatrix} \begin{bmatrix} \mathbf{B} + T_t\mathbf{1}_b - \bar{Y}\mathbf{1}_b \\ \mathbf{T}_{t-1} - T_t\mathbf{1}_{(t-1)} \\ 0 \end{bmatrix}$$

$$= \mathbf{X}_B\left(\mathbf{B} + T_t\mathbf{1}_b - \bar{Y}\mathbf{1}_b\right) + \mathbf{X}_{T-1}\left(\mathbf{T}_{t-1} - T_t\mathbf{1}_{(t-1)}\right)$$

Now each row of $\begin{bmatrix} \mathbf{X}_B & \mathbf{X}_{T-1} \end{bmatrix}$ corresponds to an observation, say the observation in the *i*th block that received the *j*th treatment. In the row for this observation the element in the *i*th column of $\mathbf{X}_B$ will be 1, as will the element in the *j*th column of $\mathbf{X}_{T-1}$; all other elements in the row will be zero. Note that if the treatment is treatment *t*, the only nonzero element will be the *i*th column of $\mathbf{X}_B$. So the fitted value for the observation in the *i*th block that received the *j*th treatment is the sum of *i*th element of $\mathbf{B} + T_t\mathbf{1}_b - \bar{Y}\mathbf{1}_b$ and the *j*th element of $\mathbf{T}_{t-1} - T_t\mathbf{1}_{(t-1)}$, unless the treatment is treatment *t* when it will be the *i*th element of $\mathbf{B} + T_t\mathbf{1}_b - \bar{Y}\mathbf{1}_b$.

The fitted value for the observation with treatment *t* in the *i*th block is $B_i + T_t - \bar{Y}$ and the fitted value for the observation with the *j*th treatment in the *i*th block is

$$B_i + T_t - \bar{Y} + T_j - T_t = B_i + T_j - \bar{Y}.$$

Clearly, $\hat{\Psi} = \mathbf{B} + \mathbf{T} - \mathbf{G}$. ■

Note that $\mathbf{B} = \mathbf{P}_{\mathrm{B}}\mathbf{Y}$, $\mathbf{T} = \mathbf{P}_{\mathrm{T}}\mathbf{Y}$, $\mathbf{G} = \mathbf{P}_{\mathrm{G}}\mathbf{Y}$ and $\mathbf{P}_i = \mathbf{X}_i\left(\mathbf{X}_i'\mathbf{X}_i\right)^{-1}\mathbf{X}_i'$ where $\mathbf{X}_{\mathrm{i}} = \mathbf{X}_{\mathrm{B}}$, $\mathbf{X}_{\mathrm{T}}$, $\mathbf{X}_{\mathrm{G}}$. That is, $\mathbf{P}_{\mathrm{B}}$, $\mathbf{P}_{\mathrm{T}}$ and $\mathbf{P}_{\mathrm{G}}$ are the block, treatment and grand mean operators, respectively. So once again the fitted values are functions of means.

Further, if the data in the vector **Y** has been arranged in nonrandomized order with all the observations for a block placed together, the operators are:

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

where, if $\mathbf{A}_r$ and $\mathbf{B}_c$ are square matrices of order r and c, respectively,

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

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

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

$$\mathbf{P}_B = 4^{-1}\mathbf{I}_5 \otimes \mathbf{J}_4 = \frac{1}{4}\begin{bmatrix}
1&1&1&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
1&1&1&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
1&1&1&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
1&1&1&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&1&1&1&1&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&1&1&1&1&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&1&1&1&1&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&1&1&1&1&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&1&1&1&1&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&1&1&1&1&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&1&1&1&1&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&1&1&1&1&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&1&1&1&1&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&1&1&1&1&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&1&1&1&1&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&1&1&1&1&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&1&1&1\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&1&1&1\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&1&1&1\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&1&1&1
\end{bmatrix}$$

$$\mathbf{P}_T = 5^{-1}\mathbf{J}_5 \otimes \mathbf{I}_4 = \frac{1}{5}\begin{bmatrix}
1&0&0&0&1&0&0&0&1&0&0&0&1&0&0&0&1&0&0&0\\
0&1&0&0&0&1&0&0&0&1&0&0&0&1&0&0&0&1&0&0\\
0&0&1&0&0&0&1&0&0&0&1&0&0&0&1&0&0&0&1&0\\
0&0&0&1&0&0&0&1&0&0&0&1&0&0&0&1&0&0&0&1\\
1&0&0&0&1&0&0&0&1&0&0&0&1&0&0&0&1&0&0&0\\
0&1&0&0&0&1&0&0&0&1&0&0&0&1&0&0&0&1&0&0\\
0&0&1&0&0&0&1&0&0&0&1&0&0&0&1&0&0&0&1&0\\
0&0&0&1&0&0&0&1&0&0&0&1&0&0&0&1&0&0&0&1\\
1&0&0&0&1&0&0&0&1&0&0&0&1&0&0&0&1&0&0&0\\
0&1&0&0&0&1&0&0&0&1&0&0&0&1&0&0&0&1&0&0\\
0&0&1&0&0&0&1&0&0&0&1&0&0&0&1&0&0&0&1&0\\
0&0&0&1&0&0&0&1&0&0&0&1&0&0&0&1&0&0&0&1\\
1&0&0&0&1&0&0&0&1&0&0&0&1&0&0&0&1&0&0&0\\
0&1&0&0&0&1&0&0&0&1&0&0&0&1&0&0&0&1&0&0\\
0&0&1&0&0&0&1&0&0&0&1&0&0&0&1&0&0&0&1&0\\
0&0&0&1&0&0&0&1&0&0&0&1&0&0&0&1&0&0&0&1\\
1&0&0&0&1&0&0&0&1&0&0&0&1&0&0&0&1&0&0&0\\
0&1&0&0&0&1&0&0&0&1&0&0&0&1&0&0&0&1&0&0\\
0&0&1&0&0&0&1&0&0&0&1&0&0&0&1&0&0&0&1&0\\
0&0&0&1&0&0&0&1&0&0&0&1&0&0&0&1&0&0&0&1
\end{bmatrix}$$

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

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

$\psi = \mathbf{X}_G\mu$ (no treatment or block differences)

$\psi = \mathbf{X}_B\beta$ (block differences only)

$\psi = \mathbf{X}_T\tau$ (treatment differences only)

$\psi = \mathbf{X}_B\beta + \mathbf{X}_T\tau$ (block and treatment differences)

An analysis of variance will be used to choose between these models. However, to get the sums of squares in the analysis of variance, not all these models need to be fitted. One of the following sequence of models needs to be fitted:

$$\begin{array}{lll} \psi = \mathbf{X}_G\mu & & \psi = \mathbf{X}_G\mu \\ \psi = \mathbf{X}_B\beta & \text{OR} & \psi = \mathbf{X}_T\tau \\ \psi = \mathbf{X}_B\beta + \mathbf{X}_T\tau & & \psi = \mathbf{X}_B\beta + \mathbf{X}_T\tau \end{array}$$

We note that $\mathcal{C}(\mathbf{X}_G) \subset \mathcal{C}(\mathbf{X}_B) \subset \mathcal{C}([\mathbf{X}_B \quad \mathbf{X}_T])$ and that $\mathcal{C}(\mathbf{X}_G) \subset \mathcal{C}(\mathbf{X}_T) \subset \mathcal{C}([\mathbf{X}_B \quad \mathbf{X}_T])$. Consequently, the model $\psi = \mathbf{X}_G\mu$ is marginal to $\psi = \mathbf{X}_B\beta$, $\psi = \mathbf{X}_T\tau$ and $\psi = \mathbf{X}_B\beta + \mathbf{X}_T\tau$ and both $\psi = \mathbf{X}_B\beta$ and $\psi = \mathbf{X}_T\tau$ are marginal to $\psi = \mathbf{X}_B\beta + \mathbf{X}_T\tau$.

The analysis of variance is constructed by considering the residual sums of squares (or unscaled deviances) after fitting each of the models in one of the above sequences is obtained. Let's call these D(μ), D(β), D(τ) and D(β,τ). The test statistics on which the hypothesis testing is based can then be conveniently computed in an analysis of variance table. The hypothesis test is a follows:

Step 1: Set up hypotheses

a) H_0: $\tau_1 = \tau_2 = ... = \tau_t$ (or $\mathbf{X}_T\tau$ not required in model)
H_1: at least one pair of population treatment means is different

b) H_0: $\beta_1 = \beta_2 = ... = \beta_b$ (or $\mathbf{X}_B\beta$ not required in model)
H_1: at least one pair of population block means is different

Step 2: Calculate test statistics

The analysis of variance table for an RCBD is:

Source	df	SSq	MSq		F
Blocks	b–1	R($\beta\|\mu$)†	R($\beta\|\mu$)/(b–1)	$\left(=s_B^2\right)$	s_B^2/s_R^2
Blocks.Plots	b(t–1)	D(β)			
Treatments	t–1	R($\tau\|\mu$)†	R($\tau\|\mu$)/(t–1)	$\left(=s_T^2\right)$	s_T^2/s_R^2
Residual	(b–1)(t–1)	D(β,τ)	D(β,τ)/(b–1)(t–1)	$\left(=s_R^2\right)$	
Total	bt–1	D(μ)			

†R($\beta|\mu$) = D(μ) – D(β) = R($\beta|\tau$) = D(τ) – D(β,τ)
R($\tau|\mu$) = D(μ) – D(τ) = R($\tau|\beta$) = D(β) – D(β,τ)

Step 3: Decide between hypotheses

If $P(F \geq F_{calc}) \leq 0.05$ then the evidence suggests that the null hypothesis should be rejected.

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 Blocks.Plots. Blocks.Plots reflects differences between plots 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
Blocks.Plots	$b(t-1)$	
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 Blocks.Plots signifies that treatment differences are confounded or "mixed-up" with plot differences as a result of the randomization of treatments to plots within blocks. Also, the Residual reflects differences between the plots within a block once the treatment differences have been removed or subtracted off. That is, it represents inherent variability in the plots 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.

In the following sections expressions for and properties of the quantities in the analysis of variance table for the RCBD are derived. Note that the most natural sequence of models for this table is where Blocks are added to the model followed by Treatments. We have given the Treatment SSqs as $R(\tau|\mu)$; this relies on the fact that $R(\tau|\mu) = R(\tau|\beta)$ which we will prove in the next section. It means that the order of fitting the Blocks and Treatments is immaterial as far as the Reduction sums of squares for Blocks and Treatments is concerned.

a) Expressions for the sums of squares

Computation by mean operators

The general recursive procedure based on mean operators that was described in section III.C, when applied to the RCBD, would proceed as follows:

1. $\mathbf{P}_G$ is applied to $\mathbf{y}$ to form $\mathbf{P}_G\mathbf{y} = \mathbf{g}$;
   $\mathbf{g}$ subtracted from $\mathbf{y}$ to form $\mathbf{R}_G\mathbf{y} = \mathbf{y} - \mathbf{g} = \mathbf{e}_G$;

2. $\mathbf{P}_B$ is applied to $\mathbf{R}_G\mathbf{y}$ to form $\mathbf{P}_B\mathbf{R}_G\mathbf{y} = \mathbf{b}_e$;
   $\mathbf{b}_e$ is subtracted from $\mathbf{e}_G$ to form $\mathbf{R}_B\mathbf{R}_G\mathbf{y} = \mathbf{e}_G - \mathbf{b}_e = \mathbf{e}_B$.
3. $\mathbf{P}_T$ is applied to $\mathbf{R}_B\mathbf{R}_G\mathbf{y}$ to form $\mathbf{P}_T\mathbf{R}_B\mathbf{R}_G\mathbf{y} = \mathbf{t}_e$;
   $\mathbf{t}_e$ is subtracted from $\mathbf{e}_B$ to form $\mathbf{R}_T\mathbf{R}_B\mathbf{R}_G\mathbf{y} = \mathbf{e}_B - \mathbf{t}_e = \mathbf{e}_{B+T}$.

That is, the vectors involved in this analysis are:

$$\begin{aligned}
\mathbf{g} &= \mathbf{P}_G\mathbf{y} \\
\mathbf{e}_G &= \mathbf{R}_G\mathbf{y} \\
\mathbf{b}_e &= \mathbf{P}_B\mathbf{R}_G\mathbf{y} \\
\mathbf{e}_B &= \mathbf{R}_B\mathbf{R}_G\mathbf{y} \\
\mathbf{t}_e &= \mathbf{P}_T\mathbf{R}_B\mathbf{R}_G\mathbf{y} \\
\mathbf{e}_{B+T} &= \mathbf{R}_T\mathbf{R}_B\mathbf{R}_G\mathbf{y}
\end{aligned}$$

with $\mathbf{e}_G = \mathbf{y} - \mathbf{g}$, $\mathbf{e}_B = \mathbf{y} - \mathbf{g} - \mathbf{b}_e$ and $\mathbf{e}_{B+T} = \mathbf{y} - \mathbf{g} - \mathbf{b}_e - \mathbf{t}_e$.

The first two steps are precisely the same as for the CRD except that the Blocks terms is being fitted instead of the Treatments term and all blocks are equally-replicated with t replicates. It is clear that theorems III.2–4 can be used to establish the following expressions for sums of squares from the analysis for the RCBD:

$$\begin{aligned}
D(\mu) &= (\mathbf{Y} - \mathbf{G})'(\mathbf{Y} - \mathbf{G}) = \mathbf{Y}'\mathbf{R}_G\mathbf{Y} \\
D(\beta) &= (\mathbf{Y} - \mathbf{B})'(\mathbf{Y} - \mathbf{B}) = \mathbf{Y}'\mathbf{R}_B\mathbf{R}_G\mathbf{Y} \\
R(\beta \mid \mu) &= \mathbf{B}_e'\mathbf{B}_e = \mathbf{Y}'\mathbf{P}_B\mathbf{R}_G\mathbf{Y}
\end{aligned}$$

where $\mathbf{B}$ and $\mathbf{B}_e$ are the n-vectors of block means and effects respectively.

Further, it must be that $D(\beta,\tau) = (\mathbf{Y} - \mathbf{B} - \mathbf{T} + \mathbf{G})'(\mathbf{Y} - \mathbf{B} - \mathbf{T} + \mathbf{G})$ as this is clearly the sums of squares of the residuals. However, it is not clear that $D(\beta,\tau) = \mathbf{E}_{B+T}'\mathbf{E}_{B+T}$ nor that $R(\tau|\beta) = R(\tau|\mu) = \mathbf{T}_e'\mathbf{T}_e$, as might be supposed from the above recursive procedure. A crucial relationship for deriving these results is that $\mathbf{P}_B\mathbf{P}_T = \mathbf{P}_T\mathbf{P}_B = \mathbf{P}_G$ — this relationship is included amongst those proved in the following lemma which shows that the result of multiplying any pair of different $\mathbf{P}$ matrices in either order is $\mathbf{P}_G$.

Lemma V.1: Let $\mathbf{P}_G = n^{-1}\mathbf{J}_b \otimes \mathbf{J}_t$, $\mathbf{P}_B = t^{-1}\mathbf{I}_b \otimes \mathbf{J}_t$ and $\mathbf{P}_T = b^{-1}\mathbf{J}_b \otimes \mathbf{I}_t$. Then

$$\mathbf{P}_B\mathbf{P}_T = \mathbf{P}_T\mathbf{P}_B = \mathbf{P}_B\mathbf{P}_G = \mathbf{P}_G\mathbf{P}_B = \mathbf{P}_T\mathbf{P}_G = \mathbf{P}_G\mathbf{P}_T = \mathbf{P}_G.$$

Proof: Note that $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = \mathbf{AB} \otimes \mathbf{CD}$ provided $\mathbf{A}$ and $\mathbf{B}$, as well as $\mathbf{C}$ and $\mathbf{D}$, are conformable. Hence

$$\mathbf{P}_B\mathbf{P}_T = \left(t^{-1}\mathbf{I}_b \otimes \mathbf{J}_t\right)\left(b^{-1}\mathbf{J}_b \otimes \mathbf{I}_t\right) = \left(bt\right)^{-1}\mathbf{J}_b \otimes \mathbf{J}_t = \mathbf{P}_G,$$
$$\mathbf{P}_T\mathbf{P}_B = \left(b^{-1}\mathbf{J}_b \otimes \mathbf{I}_t\right)\left(t^{-1}\mathbf{I}_b \otimes \mathbf{J}_t\right) = \left(bt\right)^{-1}\mathbf{J}_b \otimes \mathbf{J}_t = \mathbf{P}_G,$$
$$\mathbf{P}_B\mathbf{P}_G = \left(t^{-1}\mathbf{I}_b \otimes \mathbf{J}_t\right)\left(n^{-1}\mathbf{J}_b \otimes \mathbf{J}_t\right) = t\left(tn\right)^{-1}\mathbf{J}_b \otimes \mathbf{J}_t = \mathbf{P}_G,$$

and

$$\mathbf{P}_T\mathbf{P}_G = \left(b^{-1}\mathbf{J}_b \otimes \mathbf{I}_t\right)\left(n^{-1}\mathbf{J}_b \otimes \mathbf{J}_t\right) = b\left(bt\right)^{-1}\mathbf{J}_b \otimes \mathbf{J}_t = \mathbf{P}_G,$$

Similarly, $\mathbf{P}_G\mathbf{P}_B = \mathbf{P}_G\mathbf{P}_T = \mathbf{P}_G$. ■

Expression for D(β,τ)

Theorem V.8: Let $\mathbf{Y}$ be a *n*-vector of jointly-distributed random variables with $\psi = E[\mathbf{Y}] = \mathbf{X}_B\beta + \mathbf{X}_T\tau$ and $\text{var}[\mathbf{y}] = \sigma^2\mathbf{I}$. Also let $\mathbf{R}_T = \mathbf{I} - \mathbf{P}_T$, $\mathbf{R}_B = \mathbf{I} - \mathbf{P}_B$ and $\mathbf{R}_G = \mathbf{I} - \mathbf{P}_G$ where $\mathbf{P}_T$, $\mathbf{P}_B$ and $\mathbf{P}_G$ are as defined in lemma V.1. Then the estimator for the residual sum of squares is given by

$$D(\beta,\tau) = \left(\mathbf{Y} - \hat{\psi}\right)'\left(\mathbf{Y} - \hat{\psi}\right) = \mathbf{Y}'\mathbf{R}_T\mathbf{R}_B\mathbf{R}_G\mathbf{Y} = \mathbf{Y}'\mathbf{R}_T\mathbf{R}_B\mathbf{Y} = \mathbf{E}'_{B+T}\mathbf{E}_{B+T}$$

where $\mathbf{R}_T\mathbf{R}_B\mathbf{R}_G$ is symmetric idempotent and $\mathbf{E}_{B+T} = \mathbf{R}_T\mathbf{R}_B\mathbf{R}_G\mathbf{Y}$.

Proof: From theorem V.7, the estimators of the residuals are

$$\begin{aligned}\mathbf{Y} - \hat{\psi} &= \mathbf{Y} - \mathbf{B} - \mathbf{T} + \mathbf{G}\\ &= \mathbf{Y} - \mathbf{P}_B\mathbf{Y} - \mathbf{P}_T\mathbf{Y} + \mathbf{P}_G\mathbf{Y}\\ &= \left(\mathbf{I}_n - \mathbf{P}_B - \mathbf{P}_T + \mathbf{P}_G\right)\mathbf{Y}\end{aligned}$$

Now, recalling from lemma V.1 that $\mathbf{P}_G = \mathbf{P}_B\mathbf{P}_T$,

$$\begin{aligned}\mathbf{I}_n - \mathbf{P}_B - \mathbf{P}_T + \mathbf{P}_G &= \mathbf{I}_n - \mathbf{P}_B - \mathbf{P}_T + \mathbf{P}_B\mathbf{P}_T\\ &= \left(\mathbf{I}_n - \mathbf{P}_T\right)\left(\mathbf{I}_n - \mathbf{P}_B\right)\\ &= \mathbf{R}_T\mathbf{R}_B\end{aligned}$$

But $\mathbf{P}_B\mathbf{P}_G = \mathbf{P}_G$ so that

$$\mathbf{R}_B\mathbf{R}_G = \left(\mathbf{I} - \mathbf{P}_B\right)\left(\mathbf{I} - \mathbf{P}_G\right) = \mathbf{I} - \mathbf{P}_B - \mathbf{P}_G + \mathbf{P}_B\mathbf{P}_G = \mathbf{I} - \mathbf{P}_B - \mathbf{P}_G + \mathbf{P}_G = \mathbf{I} - \mathbf{P}_B = \mathbf{R}_B$$

and

$$\mathbf{I}_n - \mathbf{P}_B - \mathbf{P}_T + \mathbf{P}_G = \mathbf{R}_T\mathbf{R}_B = \mathbf{R}_T\mathbf{R}_B\mathbf{R}_G$$

Now to prove that $\mathbf{R}_T\mathbf{R}_B\mathbf{R}_G$ is symmetric idempotent. We note that for any two residual operators $\mathbf{R}_i$ and $\mathbf{R}_j$, $\mathbf{P}_i\mathbf{P}_j = \mathbf{P}_j\mathbf{P}_i$ implies $\mathbf{R}_i\mathbf{R}_j = \mathbf{R}_j\mathbf{R}_i$. But, from lemma V.1, this

is true for all three residual operators in $\mathbf{R}_T\mathbf{R}_B\mathbf{R}_G$. Also, each of these residual operators is symmetric. So

$$\left(\mathbf{R}_T\mathbf{R}_B\mathbf{R}_G\right)' = \mathbf{R}_G\mathbf{R}_B\mathbf{R}_T = \mathbf{R}_B\mathbf{R}_G\mathbf{R}_T = \mathbf{R}_B\mathbf{R}_T\mathbf{R}_G = \mathbf{R}_T\mathbf{R}_B\mathbf{R}_G$$

and $\mathbf{R}_T\mathbf{R}_B\mathbf{R}_G$ is symmetric. Also,

$$\mathbf{R}_T\mathbf{R}_B\mathbf{R}_G\mathbf{R}_T\mathbf{R}_B\mathbf{R}_G = \mathbf{R}_B\mathbf{R}_T\mathbf{R}_T\mathbf{R}_G\mathbf{R}_B\mathbf{R}_G = \mathbf{R}_B\mathbf{R}_T\mathbf{R}_B\mathbf{R}_G\mathbf{R}_G = \mathbf{R}_T\mathbf{R}_B\mathbf{R}_B\mathbf{R}_G = \mathbf{R}_T\mathbf{R}_B\mathbf{R}_G$$

and so $\mathbf{R}_T\mathbf{R}_B\mathbf{R}_G$ is idempotent.

Consequently, the estimator for the residual sum of squares for fitting the maximal model is given by

$$D(\beta,\tau) = \left(\mathbf{Y} - \hat{\psi}\right)'\left(\mathbf{Y} - \hat{\psi}\right) = \mathbf{Y}'\mathbf{R}_T\mathbf{R}_B\mathbf{R}_G\mathbf{Y} = \mathbf{Y}'\mathbf{R}_T\mathbf{R}_B\mathbf{Y} = \mathbf{E}'_{B+T}\mathbf{E}_{B+T}$$

■

Expression for R(τ|β)

Theorem V.9: Let $D(\beta) = \mathbf{Y}'\mathbf{R}_B\mathbf{R}_G\mathbf{Y}$ and $D(\beta,\tau) = \mathbf{Y}'\mathbf{R}_T\mathbf{R}_B\mathbf{R}_G\mathbf{Y}$ be the estimators for the residual sums of squares for the models $E[\mathbf{Y}] = \mathbf{X}_B\beta$ and $E[\mathbf{Y}] = \mathbf{X}_B\beta + \mathbf{X}_T\tau$, respectively. Then, for $R(\tau|\beta) = D(\beta) - D(\beta,\tau)$,

$$R(\tau|\beta) = \mathbf{Y}'\mathbf{P}_T\mathbf{R}_B\mathbf{R}_G\mathbf{Y} = \mathbf{Y}'\mathbf{P}_T\mathbf{R}_G\mathbf{Y} = \mathbf{T}_e'\mathbf{T}_e$$

where $\mathbf{T}_e = \mathbf{P}_T\mathbf{R}_B\mathbf{R}_G\mathbf{Y} = \mathbf{P}_T\mathbf{R}_G\mathbf{Y}$.

Proof:

$$\begin{aligned} R(\tau|\beta) &= D(\beta) - D(\beta,\tau) \\ &= \mathbf{Y}'\mathbf{R}_B\mathbf{R}_G\mathbf{Y} - \mathbf{Y}'\mathbf{R}_T\mathbf{R}_B\mathbf{R}_G\mathbf{Y} \\ &= \mathbf{Y}'\left(\mathbf{I} - \mathbf{R}_T\right)\mathbf{R}_B\mathbf{R}_G\mathbf{Y} \\ &= \mathbf{Y}'\mathbf{P}_T\mathbf{R}_B\mathbf{R}_G\mathbf{Y} \end{aligned}$$

Also,

$$\begin{aligned} R(\tau|\beta) &= \mathbf{Y}'\mathbf{P}_T\mathbf{R}_B\mathbf{R}_G\mathbf{Y} \\ &= \mathbf{Y}'\mathbf{P}_T\left(\mathbf{I} - \mathbf{P}_B\right)\mathbf{R}_G\mathbf{Y} \\ &= \mathbf{Y}'\left(\mathbf{P}_T - \mathbf{P}_T\mathbf{P}_B\right)\mathbf{R}_G\mathbf{Y} \\ &= \mathbf{Y}'\left(\mathbf{P}_T - \mathbf{P}_G\right)\mathbf{R}_G\mathbf{Y} \\ &= \mathbf{Y}'\left(\mathbf{P}_T\mathbf{R}_G - \mathbf{P}_G\mathbf{R}_G\right)\mathbf{Y} \\ &= \mathbf{Y}'\mathbf{P}_T\mathbf{R}_G\mathbf{Y} \qquad \text{since } \mathbf{P}_G\mathbf{R}_G = \mathbf{0} \end{aligned}$$

■

The implication of $R(\tau|\beta) = \mathbf{Y}'\mathbf{P}_T\mathbf{R}_B\mathbf{R}_G\mathbf{Y} = \mathbf{Y}'\mathbf{P}_T\mathbf{R}_G\mathbf{Y}$ is that $R(\tau|\beta) = R(\tau|\mu)$. It can be similarly proved that $R(\beta|\tau) = R(\beta|\mu)$.

Summary of sums of squares

The estimators for the sums of squares in the analysis of the RCBD are:

$$\begin{aligned}
D(\mu) &= \mathbf{E}_G'\mathbf{E}_G &&= \mathbf{Y}'\mathbf{R}_G\mathbf{Y} \\
R(\beta|\mu) = R(\beta|\tau) &= \mathbf{B}_e'\mathbf{B}_e &&= \mathbf{Y}'\mathbf{P}_B\mathbf{R}_G\mathbf{Y} = \mathbf{Y}'\mathbf{P}_B\mathbf{R}_T\mathbf{R}_G\mathbf{Y} \\
D(\beta) &= \mathbf{E}_B'\mathbf{E}_B &&= \mathbf{Y}'\mathbf{R}_B\mathbf{R}_G\mathbf{Y} \\
R(\tau|\beta) = R(\tau|\mu) &= \mathbf{T}_e'\mathbf{T}_e &&= \mathbf{Y}'\mathbf{P}_T\mathbf{R}_B\mathbf{R}_G\mathbf{Y} = \mathbf{Y}'\mathbf{P}_T\mathbf{R}_G\mathbf{Y} \\
D(\beta,\tau) &= \mathbf{E}_{B+T}'\mathbf{E}_{B+T} &&= \mathbf{Y}'\mathbf{R}_T\mathbf{R}_B\mathbf{R}_G\mathbf{Y} = \mathbf{Y}'\mathbf{R}_T\mathbf{R}_B\mathbf{Y}
\end{aligned}$$

b) Degrees of freedom

The following theorem establishes that the degrees of freedom of the sums of squares are as given in the analysis of variance table.

Theorem V.10: Let $D(\mu) = \mathbf{Y}'\mathbf{R}_G\mathbf{Y}$, $R(\beta|\mu) = \mathbf{Y}'\mathbf{P}_B\mathbf{R}_G\mathbf{Y}$, $D(\beta) = \mathbf{Y}'\mathbf{R}_B\mathbf{R}_G\mathbf{Y}$, $R(\tau|\mu) = \mathbf{Y}'\mathbf{P}_T\mathbf{R}_G\mathbf{Y}$ and $D(\beta,\tau) = \mathbf{Y}'\mathbf{R}_T\mathbf{R}_B\mathbf{R}_G\mathbf{Y}$. The degrees of freedom of D(μ), R($\beta|\mu$), D(β), R($\tau|\mu$) and D(β,τ) are n–1, b–1, n–b, t–1 and (b–1)(t–1), respectively, where n is the number of observations, b is the number of blocks and t is the number of treatments.

Proof: The matrices $\mathbf{R}_G$, $\mathbf{P}_B\mathbf{R}_G$, $\mathbf{R}_B\mathbf{R}_G$ and $\mathbf{P}_T\mathbf{R}_G$ are symmetric and idempotent by virtue of being special cases of operators with these properties that were given in lemma III.1. Also, theorem V.8 showed that $\mathbf{R}_T\mathbf{R}_B\mathbf{R}_G$ is symmetric and idempotent. Consequently, the ranks of all these matrices are equal to their trace.

We also note that $trace(\mathbf{A}\otimes\mathbf{B}) = trace(\mathbf{A})\,trace(\mathbf{B})$ so that

$$trace(\mathbf{P}_G) = trace\left(\{bt\}^{-1}\mathbf{J}_b \otimes \mathbf{J}_t\right) = \{bt\}^{-1}bt = 1$$

$$trace(\mathbf{P}_B) = trace\left(t^{-1}\mathbf{I}_b \otimes \mathbf{J}_t\right) = t^{-1}bt = b$$

and

$$trace(\mathbf{P}_T) = trace\left(b^{-1}\mathbf{J}_b \otimes \mathbf{I}_t\right) = b^{-1}bt = t$$

It was established in theorem III.5 that $trace(\mathbf{R}_G) = n-1$.

$$trace(\mathbf{P}_B\mathbf{R}_G) = trace(\mathbf{P}_B - \mathbf{P}_G) = b-1$$

$$trace(\mathbf{R}_B\mathbf{R}_G) = trace\left((\mathbf{I}-\mathbf{P}_B)(\mathbf{I}-\mathbf{P}_G)\right) = trace(\mathbf{I}-\mathbf{P}_B-\mathbf{P}_G+\mathbf{P}_G) = n-b$$

$$trace(\mathbf{P}_T\mathbf{R}_G) = trace(\mathbf{P}_T - \mathbf{P}_G) = t-1$$
$$trace(\mathbf{R}_T\mathbf{R}_B) = trace(\mathbf{I} - \mathbf{P}_B - \mathbf{P}_T + \mathbf{P}_G) = n-b-t+1 = bt-b-t+1 = (b-1)(t-1) \quad \blacksquare$$

c) Expected mean squares

As for the one way analysis of variance, the expected mean squares can be worked out for the four expectation models that are being considered:

$$\psi = \mathbf{X}_G\mu \qquad (\text{no treatment or block differences})$$
$$\psi = \mathbf{X}_B\beta \qquad (\text{block differences only})$$
$$\psi = \mathbf{X}_T\tau \qquad (\text{treatment differences only})$$
$$\psi = \mathbf{X}_B\beta + \mathbf{X}_T\tau \qquad (\text{block and treatment differences})$$

The sum of squares for which the expected mean squares are required are: $R(\beta|\mu)$, $R(\tau|\mu)$ and $D(\beta,\tau)$. These are given by

$$R(\beta \mid \mu) = \mathbf{B}_e'\mathbf{B}_e = \mathbf{Y}'\mathbf{P}_B\mathbf{R}_G\mathbf{Y}$$
$$R(\tau \mid \mu) = \mathbf{T}_e'\mathbf{T}_e = \mathbf{Y}'\mathbf{P}_T\mathbf{R}_G\mathbf{Y}$$
$$D(\beta,\tau) = \mathbf{E}_{B+T}'\mathbf{E}_{B+T} = \mathbf{Y}'\mathbf{R}_T\mathbf{R}_B\mathbf{Y}$$

Expected mean squares under maximal model

Theorem V.11: Let $\psi = E[\mathbf{Y}] = \mathbf{X}_B\beta + \mathbf{X}_T\tau$, $\mathbf{V}_Y = \sigma^2\mathbf{I}_n$, $R(\beta \mid \mu) = \mathbf{Y}'\mathbf{P}_B\mathbf{R}_G\mathbf{Y}$, $R(\tau \mid \mu) = \mathbf{Y}'\mathbf{P}_T\mathbf{R}_G\mathbf{Y}$ and $D(\beta,\tau) = \mathbf{Y}'\mathbf{R}_T\mathbf{R}_B\mathbf{Y}$ where $\mathbf{R}_T$, $\mathbf{R}_B$ and $\mathbf{R}_G$ are as defined in theorem V.8. Then,

$$E\left[R(\beta \mid \mu)/(b-1)\right] = \sigma^2 + f_B(\psi)$$
$$E\left[R(\tau \mid \mu)/(t-1)\right] = \sigma^2 + f_T(\psi),$$
$$E\left[D(\beta,\tau)/\{(b-1)(t-1)\}\right] = \sigma^2$$

where $f_B(\psi) = \sum_{i=1}^{b} t(\beta_i - \bar{\beta}_.)^2 \Big/ (b-1)$, $\bar{\beta}_. = \sum_{i=1}^{b}\beta_i/b$, β_i is the ith element of the b-vector β, $f_T(\psi) = \sum_{j=1}^{t} b(\tau_j - \bar{\tau}_.)^2 \Big/ (t-1)$, $\bar{\tau}_. = \sum_{j=1}^{t}\tau_j/t$, τ_j is the jth element of the t-vector τ, b is the number of blocks and t is the number of treatments.

Proof: For $E\left[R(\beta \mid \mu)/(b-1)\right]$, we first use theorem II.11 to show that

$$E\left[R\left(\beta\,|\,\mu\right)/(b-1)\right] = E\left[\mathbf{Y}'\mathbf{P}_{\mathrm{B}}\mathbf{R}_{\mathrm{G}}\mathbf{Y}\right]/(b-1)$$
$$= \left\{trace\left(\mathbf{P}_{\mathrm{B}}\mathbf{R}_{\mathrm{G}}\sigma^2\mathbf{I}_n\right) + \left(\mathbf{X}_{\mathrm{B}}\beta + \mathbf{X}_{\mathrm{T}}\tau\right)'\mathbf{P}_{\mathrm{B}}\mathbf{R}_{\mathrm{G}}\left(\mathbf{X}_{\mathrm{B}}\beta + \mathbf{X}_{\mathrm{T}}\tau\right)\right\}\Big/\{b-1\}$$
$$= \left\{\sigma^2 trace\left(\mathbf{P}_{\mathrm{B}}\mathbf{R}_{\mathrm{G}}\right) + \left(\ \mathbf{X}_{\mathrm{B}}\beta + \mathbf{X}_{\mathrm{T}}\tau\right)'\mathbf{P}_{\mathrm{B}}\mathbf{R}_{\mathrm{G}}\left(\mathbf{X}_{\mathrm{B}}\beta + \mathbf{X}_{\mathrm{T}}\tau\right)\right\}\Big/\{b-1\}$$

Now from theorem V.10, $trace\left(\mathbf{P}_{\mathrm{B}}\mathbf{R}_{\mathrm{G}}\right) = b-1$.

Also,

$$\mathbf{P}_{\mathrm{B}}\mathbf{R}_{\mathrm{G}}\left(\mathbf{X}_{\mathrm{B}}\beta + \mathbf{X}_{\mathrm{T}}\tau\right) = \mathbf{P}_{\mathbf{B}}\mathbf{R}_{\mathrm{G}}\mathbf{X}_{\mathrm{B}}\beta + \mathbf{P}_{\mathrm{B}}\mathbf{R}_{\mathrm{G}}\mathbf{X}_{\mathrm{T}}\tau$$
$$= \left(\mathbf{P}_{\mathrm{B}} - \mathbf{P}_{\mathrm{G}}\right)\mathbf{X}_{\mathrm{B}}\beta + \left(\mathbf{P}_{\mathrm{B}} - \mathbf{P}_{\mathrm{G}}\right)\mathbf{X}_{\mathrm{T}}\tau$$

Now from lemma III.2 $\mathbf{P}_{\mathrm{B}}\mathbf{X}_{\mathrm{B}} = \mathbf{X}_{\mathrm{B}}$ and we can write $\mathbf{X}_{\mathrm{B}}$ and $\mathbf{X}_{\mathrm{T}}$ in terms of direct products as follows:

$$\mathbf{X}_{\mathrm{B}} = \mathbf{I}_b \otimes \mathbf{1}_t \text{ and } \mathbf{X}_{\mathrm{T}} = \mathbf{1}_b \otimes \mathbf{I}_t$$

Consequently,

$$\left(\mathbf{P}_{\mathrm{B}} - \mathbf{P}_{\mathrm{G}}\right)\mathbf{X}_{\mathrm{B}} = \mathbf{X}_{\mathrm{B}} - \left(\{bt\}^{-1}\mathbf{J}_b \otimes \mathbf{J}_t\right)\mathbf{I}_b \otimes \mathbf{1}_t$$
$$= \mathbf{X}_{\mathrm{B}} - b^{-1}\mathbf{J}_b \otimes \mathbf{1}_t$$

and

$$\left(\mathbf{P}_{\mathrm{B}} - \mathbf{P}_{\mathrm{G}}\right)\mathbf{X}_{\mathrm{T}} = \left(t^{-1}\mathbf{I}_b \otimes \mathbf{J}_t - \{bt\}^{-1}\mathbf{J}_b \otimes \mathbf{J}_t\right)\mathbf{1}_b \otimes \mathbf{I}_t$$
$$= t^{-1}\mathbf{1}_b \otimes \mathbf{J}_t - \{bt\}^{-1} b\mathbf{1}_b \otimes \mathbf{J}_t$$
$$= \mathbf{0}_{bt\times t}$$

so that

$$\mathbf{P}_{\mathrm{B}}\mathbf{R}_{\mathrm{G}}\left(\mathbf{X}_{\mathrm{B}}\beta + \mathbf{X}_{\mathrm{T}}\tau\right) = \left(\mathbf{P}_{\mathrm{B}} - \mathbf{P}_{\mathrm{G}}\right)\mathbf{X}_{\mathrm{B}}\beta + \left(\mathbf{P}_{\mathrm{B}} - \mathbf{P}_{\mathrm{G}}\right)\mathbf{X}_{\mathrm{T}}\tau$$
$$= \left(\mathbf{X}_{\mathrm{B}} - b^{-1}\mathbf{J}_b \otimes \mathbf{1}_t\right)\beta$$
$$= \mathbf{X}_{\mathrm{B}}\beta - \bar{\beta}_{\cdot}\mathbf{1}_{bt}$$

Hence the expected mean square is

$$E\left[R\left(\beta\,|\,\mu\right)/(b-1)\right] = \left\{\sigma^2 trace\left(\mathbf{P}_{\mathrm{B}}\mathbf{R}_{\mathrm{G}}\right) + \left(\mathbf{X}_{\mathrm{B}}\beta + \mathbf{X}_{\mathrm{T}}\tau\right)'\mathbf{P}_{\mathrm{B}}\mathbf{R}_{\mathrm{G}}\left(\mathbf{X}_{\mathrm{B}}\beta + \mathbf{X}_{\mathrm{T}}\tau\right)\right\}\Big/\{b-1\}$$
$$= \left\{\sigma^2(b-1) + \left(\ \mathbf{X}_{\mathrm{B}}\beta - \bar{\beta}_{\cdot}\mathbf{1}_{bt}\right)'\left(\mathbf{X}_{\mathrm{B}}\beta - \bar{\beta}_{\cdot}\mathbf{1}_{bt}\right)\right\}\Big/\{b-1\}$$
$$= \sigma^2 + \sum_{i=1}^{b} t\left(\beta_i - \bar{\beta}_{\cdot}\right)^2\Big/(b-1)$$
$$= \sigma^2 + f_{\mathrm{B}}\left(\psi\right)$$

The proof that $E\left[R(\tau \mid \mu)/(t-1)\right] = \sigma^2 + f_T(\psi)$ is similar to that for $E\left[R(\beta \mid \mu)/(b-1)\right] = \sigma^2 + f_B(\psi)$ and is left as an exercise for you.

$$\begin{aligned} E\left[\mathrm{D}(\beta,\tau)/\{(b-1)(t-1)\}\right] &= E\left[\mathbf{Y}'\mathbf{R}_T\mathbf{R}_B\mathbf{R}_G\mathbf{Y}\right]/\{(b-1)(t-1)\} \\ &= \frac{\left\{\sigma^2 trace(\mathbf{R}_T\mathbf{R}_B) + (\mathbf{X}_B\beta + \mathbf{X}_T\tau)'\mathbf{R}_T\mathbf{R}_B\mathbf{R}_G(\mathbf{X}_B\beta + \mathbf{X}_T\tau)\right\}}{\{(b-1)(t-1)\}} \\ &= \sigma^2 + (\mathbf{X}_B\beta + \mathbf{X}_T\tau)'\mathbf{R}_T\mathbf{R}_B(\mathbf{X}_B\beta + \mathbf{X}_T\tau)/\{(b-1)(t-1)\} \end{aligned}$$

Now, using lemma III.2, $\mathbf{R}_B\mathbf{X}_B = \mathbf{0}_{bt\times t}$ and $\mathbf{R}_T\mathbf{X}_T = \mathbf{0}_{bt\times t}$ so that

$$\begin{aligned} &\mathbf{R}_T\mathbf{R}_B(\mathbf{X}_B\beta + \mathbf{X}_T\tau) \\ &= \mathbf{R}_T\mathbf{R}_B\mathbf{X}_T\tau \\ &= \mathbf{R}_B\mathbf{R}_T\mathbf{X}_T\tau \\ &= \mathbf{0}_{bt\times 1} \end{aligned}$$

Hence, $E\left[\mathrm{D}(\beta,\tau)/\{(b-1)(t-1)\}\right] = \sigma^2$ as claimed. ■

We summarize the results in the following table:

Source	df	MSq	E[MSq]	F
Blocks	b–1	R(β|μ)/(b–1)	$\sigma^2 + f_B(\psi)$ †	s_B^2/s_R^2
Blocks.Plots	$b(t$–1)			
Treatments	t–1	R(τ|μ)/(t–1)	$\sigma^2 + f_T(\psi)$ †	s_T^2/s_R^2
Residual	(b–1)(t–1)	D(β,τ)/(b–1)(t–1)	σ^2	
Total	bt–1			

† $f_B(\psi) = \sum t(\beta_i - \bar{\beta}_.)^2/(b-1)$

$f_T(\psi) = \sum b(\tau_i - \bar{\tau}_.)^2/(t-1)$

These indicate that the Residual mean square estimates the uncontrolled variation in the experiment, that is the variation arising from uncontrolled differences between plots 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 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 plots 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 plots, 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 plots?". The Block F value involves a similar question.

Now as to whether the blocking has been effective, it turns out that it will if the plots 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.

Expected mean squares under other models

We now state the theorems that give the expected mean squares under the treatment-differences-only, block-differences-only and no-block-treatment-differences models.

Theorem V.12: Let $\psi = E[\mathbf{Y}] = \mathbf{X}_T\tau$, $\mathbf{V}_Y = \sigma^2\mathbf{I}_n$, $R(\beta \mid \mu) = \mathbf{Y}'\mathbf{P}_B\mathbf{R}_G\mathbf{Y}$, $R(\tau \mid \mu) = \mathbf{Y}'\mathbf{P}_T\mathbf{R}_G\mathbf{Y}$ and $\mathrm{D}(\beta,\tau) = \mathbf{Y}'\mathbf{R}_T\mathbf{R}_B\mathbf{Y}$ where $\mathbf{R}_T$, $\mathbf{R}_B$ and $\mathbf{R}_G$ are as defined in theorem V.8. Then,

$$E\left[R(\beta \mid \mu)/(b-1)\right] = \sigma^2$$
$$E\left[R(\tau \mid \mu)/(t-1)\right] = \sigma^2 + f_T(\psi)$$
$$E\left[\mathrm{D}(\beta,\tau)/\{(b-1)(t-1)\}\right] = \sigma^2$$

where $f_T(\psi) = \sum_{j=1}^{t} b(\tau_j - \bar{\tau}_\cdot)^2 \Big/ (t-1)$, $\bar{\tau}_\cdot = \sum_{j=1}^{t}\tau_j \Big/ t$, τ_j is the jth element of the t-vector τ, b is the number of blocks and t is the number of treatments.

Proof: left as an exercise for you. ■

Note that this theorem implies that $\psi'\mathbf{P}_B\mathbf{R}_G\psi = \mathbf{0}$ so that $\mathbf{P}_B\mathbf{R}_G\psi = \mathbf{P}_B\mathbf{R}_G\mathbf{X}_T\tau = \mathbf{0}$.

Theorem V.13: Let $\psi = E[\mathbf{Y}] = \mathbf{X}_B\beta$, $\mathbf{V}_Y = \sigma^2\mathbf{I}_n$, $R(\beta \mid \mu) = \mathbf{Y}'\mathbf{P}_B\mathbf{R}_G\mathbf{Y}$, $R(\tau \mid \mu) = \mathbf{Y}'\mathbf{P}_T\mathbf{R}_G\mathbf{Y}$ and $\mathrm{D}(\beta,\tau) = \mathbf{Y}'\mathbf{R}_T\mathbf{R}_B\mathbf{Y}$ where $\mathbf{R}_T$, $\mathbf{R}_B$ and $\mathbf{R}_G$ are as defined in theorem V.8. Then,

$$E\left[R(\beta \mid \mu)/(b-1)\right] = \sigma^2 + f_B(\psi)$$
$$E\left[R(\tau \mid \mu)/(t-1)\right] = \sigma^2$$
$$E\left[\mathrm{D}(\beta,\tau)/\{(b-1)(t-1)\}\right] = \sigma^2$$

where $f_B(\psi) = \sum_{i=1}^{b} t(\beta_i - \bar{\beta}_\cdot)^2 \Big/ (b-1)$, $\bar{\beta}_\cdot = \sum_{i=1}^{b}\beta_i \Big/ b$, β_i is the ith element of the b-vector β, b is the number of blocks and t is the number of treatments.

Proof: left as an exercise for you. ■

Note that this theorem implies that $\psi'\mathbf{P}_T\mathbf{R}_G\psi = \mathbf{0}$ so that $\mathbf{P}_T\mathbf{R}_G\psi = \mathbf{P}_T\mathbf{R}_G\mathbf{X}_B\beta = \mathbf{0}$.

Theorem V.14: Let $\psi = E[\mathbf{Y}] = \mathbf{X}_G\mu$, $\mathbf{V}_Y = \sigma^2\mathbf{I}_n$, $R(\beta \mid \mu) = \mathbf{Y}'\mathbf{P}_B\mathbf{R}_G\mathbf{Y}$, $R(\tau \mid \mu) = \mathbf{Y}'\mathbf{P}_T\mathbf{R}_G\mathbf{Y}$ and $\mathrm{D}(\beta,\tau) = \mathbf{Y}'\mathbf{R}_T\mathbf{R}_B\mathbf{Y}$ where $\mathbf{R}_T$, $\mathbf{R}_B$ and $\mathbf{R}_G$ are as defined in theorem V.8. Then,

$$E\left[R(\beta \mid \mu)/(b-1)\right] = \sigma^2$$
$$E\left[R(\tau \mid \mu)/(t-1)\right] = \sigma^2$$
$$E\left[\mathrm{D}(\beta,\tau)/\{(b-1)(t-1)\}\right] = \sigma^2$$

where b is the number of blocks and t is the number of treatments.

Proof: left as an exercise for you. ■

Note that this theorem implies that $\psi'\mathbf{P}_B\mathbf{R}_G\psi = \psi'\mathbf{P}_T\mathbf{R}_G\psi = \mathbf{0}$ so that $\mathbf{P}_T\mathbf{R}_G\psi = \mathbf{P}_T\mathbf{R}_G\mathbf{X}_G\mu = \mathbf{0} = \mathbf{P}_B\mathbf{R}_G\psi = \mathbf{P}_B\mathbf{R}_G\mathbf{X}_G\mu$.

From these theorems it is clear that removing a term from the expectation model results in the corresponding expectation function of ψ being removed from the expected mean squares.

d) Distribution of the F statistic

We next derive the sampling distribution of the F-statistics for testing block and treatment differences.

Theorem V.15: Let $E[\mathbf{Y}] = \mathbf{X}_G\mu$ or $E[\mathbf{Y}] = \mathbf{X}_B\beta$, $\mathbf{V}_Y = \sigma^2\mathbf{I}_n$, $s_T^2 = R(\tau|\mu)/(t-1) = \mathbf{Y}'\mathbf{P}_T\mathbf{R}_G\mathbf{Y}/(t-1)$ and $s_R^2 = D(\beta,\tau)/\{(b-1)(t-1)\} = \mathbf{Y}'\mathbf{R}_T\mathbf{R}_B\mathbf{Y}/\{(b-1)(t-1)\}$ where $\mathbf{P}_T$, $\mathbf{R}_G$, $\mathbf{R}_T$ and $\mathbf{R}_B$ are as defined in lemma V.1 and theorem V.8. Then, the ratio of these two mean squares, given by

$$F_{(t-1),(b-1)(t-1)} = \frac{s_T^2}{s_R^2}$$

is distributed as a Snedecor's F with $(t-1)$ and $(b-1)(t-1)$ degrees of freedom.

Proof: We have only to show that $\mathbf{Y}'\mathbf{P}_T\mathbf{R}_G\mathbf{Y}$ and $\mathbf{Y}'\mathbf{R}_T\mathbf{R}_B\mathbf{Y}$ are independent quadratic forms and that $E[\mathbf{P}_T\mathbf{R}_G\mathbf{Y}] = \mathbf{0}$ and $E[\mathbf{R}_T\mathbf{R}_B\mathbf{Y}] = \mathbf{0}$ under the two expectation models. Then theorem II.6 is satisfied and theorem II.17 can be invoked to obtain the distribution of the test statistic.

Since the matrices of the two quadratic forms are idempotent, we have only to establish that $\mathbf{P}_T\mathbf{R}_G\mathbf{R}_T\mathbf{R}_B = \mathbf{0}$ so that two of the three conditions in theorem II.16 have been satisfied. It is left for you to establish this condition.

From the notes under theorems V.13 and V.14, $\mathbf{P}_T\mathbf{R}_G E[\mathbf{Y}] = \mathbf{P}_T\mathbf{R}_G\psi = \mathbf{0}$ under the two models given in the statement of the theorem. Also, as $E[\mathrm{D}(\beta,\tau)/\{(b-1)(t-1)\}] = \sigma^2$ under any model, $\mathbf{R}_T\mathbf{R}_B E[\mathbf{Y}] = \mathbf{R}_T\mathbf{R}_B\psi = \mathbf{0}$ under any model.

Therefore theorem II.7 can be invoked to produce the desired result. ■

Theorem V.16: Let $E[\mathbf{Y}] = \mathbf{X}_G\mu$ or $E[\mathbf{Y}] = \mathbf{X}_T\tau$, $\mathbf{V}_Y = \sigma^2\mathbf{I}_n$, $s_B^2 = R(\beta|\mu)/(t-1) = \mathbf{Y}'\mathbf{P}_B\mathbf{R}_G\mathbf{Y}/(b-1)$ and $s_R^2 = D(\beta,\tau)/\{(b-1)(t-1)\} = \mathbf{Y}'\mathbf{R}_T\mathbf{R}_B\mathbf{Y}/\{(b-1)(t-1)\}$ where $\mathbf{P}_B$, $\mathbf{R}_G$, $\mathbf{R}_T$ and $\mathbf{R}_B$ are as defined in lemma V.1 and theorem V.8. Then, the ratio of these two mean squares, given by

$$F_{(b-1),(b-1)(t-1)} = \frac{s_B^2}{s_R^2}$$

is distributed as a Snedecor's F with $(b-1)$ and $(b-1)(t-1)$ degrees of freedom.

Proof: parallels that for the treatment differences. ■

e) Analysis of variance table

Gathering together the results from the previous sections, the analysis of variance table for an RCBD is:

Source	df	SSq	MSq	E[MSq]	F
Blocks	b–1	$R(\beta\|\mu)^\dagger = \mathbf{b}_e'\mathbf{b}_e$	$R(\beta\|\mu)/(b-1)$	$\sigma^2 + f_B(\psi)^\dagger$	s_B^2/s_R^2
Blocks.Plots	$b(t-1)$	$D(\beta) = \mathbf{e}_B'\mathbf{e}_B$			
Treatments	t–1	$R(\tau\|\mu)^\dagger = \mathbf{t}_e'\mathbf{t}_e$	$R(\tau\|\mu)/(t-1)$	$\sigma^2 + f_T(\psi)^\dagger$	s_T^2/s_R^2
Residual	$(b-1)(t-1)$	$D(\beta,\tau) = \mathbf{e}_{B+T}'\mathbf{e}_{B+T}$	$D(\beta,\tau)/(b-1)(t-1)$	σ^2	
Total	bt–1	$D(\mu) = \mathbf{e}_G'\mathbf{e}_G$			

$^\dagger R(\beta|\mu) = D(\mu) - D(\beta) = R(\beta|\tau) = D(\tau) - D(\beta,\tau)$
$R(\tau|\mu) = D(\mu) - D(\tau) = R(\tau|\beta) = D(\beta) - D(\beta,\tau)$

f) Analysis of the penicillin example

Example V.1 Penicillin yield (continued)

We first carry out the recursive procedure described above.

Step 1: Application of grand mean operator to original data vector

The decomposition of the original data vector is achieved by applying the grand mean operator $\mathbf{P}_G$ to $\mathbf{y}$ to form $\mathbf{P}_G\mathbf{y} = \mathbf{g}$ and then subtracting $\mathbf{g}$ from $\mathbf{y}$ to form $\mathbf{R}_G\mathbf{y} = \mathbf{y} - \mathbf{g} = \mathbf{e}_G$.

Observations		Grand mean		Flask Deviations
y		**g**		$\mathbf{e}_G$
89		86		3
88		86		2
97		86		11
94		86		8
84		86		–2
77		86		–9
92		86		6
79		86		–7
81		86		–5
87	=	86	+	1
87		86		1
85		86		–1
87		86		1
92		86		6
89		86		3
84		86		–2
79		86		–7
81		86		–5
80		86		–6
88		86		2

Step 2: Application of block mean operator to Flasks deviations vector

The decomposition of the Flasks deviation vector is achieved by applying the block mean operator $\mathbf{P}_B$ to $\mathbf{e}_G = \mathbf{R}_G\mathbf{y}$ to form $\mathbf{P}_B\mathbf{R}_G\mathbf{y} = \mathbf{b}_e$ and then subtracting $\mathbf{b}_e$ from $\mathbf{e}_G$ to form $\mathbf{R}_B\mathbf{R}_G\mathbf{y} = \mathbf{e}_G - \mathbf{b}_e = \mathbf{e}_B$.

Flasks Deviations		Blend Effects		Blend.Flasks Deviations
$\mathbf{e}_G$		$\mathbf{b}_e$		$\mathbf{e}_B$
3		6		–3
2		6		–4
11		6		5
8		6		2
–2		–3		1
–9		–3		–6
6		–3		9
–7		–3		–4
–5		–1		–4
1	=	–1	+	2
1		–1		2
–1		–1		0
1		2		–1
6		2		4
3		2		1
–2		2		–4
–7		–4		–3
–5		–4		–1
–6		–4		–2
2		–4		6

This amounts to a partitioning of the total variation according to the natural variation in the experiment as follows:

Observations		Grand mean		Blend Effects		Blend.Flasks Deviations
$\mathbf{y}$		$\mathbf{g}$		$\mathbf{b}_e$		$\mathbf{e}_B$
89		86		6		–3
88		86		6		–4
97		86		6		5
94		86		6		2
84		86		–3		1
77		86		–3		–6
92		86		–3		9
79		86		–3		–4
81		86		–1		–4
87		86		–1		2
87	=	86	+	–1	+	2
85		86		–1		0
87		86		2		–1
92		86		2		4
89		86		2		1
84		86		2		–4
79		86		–4		–3
81		86		–4		–1
80		86		–4		–2
88		86		–4		6

Step 3: Application of treatment mean operator to Blends.Flasks deviations vector

The next task is to further isolate components of interest to the experimenter by applying the treatment mean operator $\mathbf{P}_T$ to $\mathbf{e}_B = \mathbf{R}_B\mathbf{R}_G\mathbf{y}$ to form $\mathbf{P}_T\mathbf{R}_B\mathbf{R}_G\mathbf{y} = \mathbf{t}_e$ and then subtracting $\mathbf{t}_e$ from $\mathbf{e}_B$ to form $\mathbf{R}_T\mathbf{R}_B\mathbf{R}_G\mathbf{y} = \mathbf{e}_B - \mathbf{t}_e = \mathbf{e}_{B+T}$.

Blend.Flasks Deviations		Treatment Effects		Blend.Flasks Deviations
$\mathbf{e}_B$		$\mathbf{t}_e$		$\mathbf{e}_{B+T}$
–3		–2		–1
–4		–1		–3
5		3		2
2		0		2
1		–2		3
–6		–1		–5
9		3		6
–4		0		–4
–4		–2		–2
2		–1		3
2	=	3	+	–1
0		0		0
–1		–2		1
4		–1		5
1		3		–2
–4		0		–4
–3		–2		–1
–1		–1		0
–2		3		–5
6		0		6

Thus the full decomposition of the data vector is:

Observ-ations		Grand Mean		Blend Effects		Blend.Flasks Deviations		Treatment Effects		Blend.Flasks Deviations
$\mathbf{y}$		$\mathbf{g}$		$\mathbf{b}_e$		$\mathbf{e}_B$		$\mathbf{t}_e$		$\mathbf{e}_{B+T}$
89		86		6		−3		−2		−1
88		86		6		−4		−1		−3
97		86		6		5		3		2
94		86		6		2		0		2
84		86		−3		1		−2		3
77		86		−3		−6		−1		−5
92		86		−3		9		3		6
79		86		−3		−4		0		−4
81		86		−1		−4		−2		−2
87		86		−1		2		−1		3
87	=	86	+	−1	+	2	=	3	+	−1
85		86		−1		0		0		0
87		86		2		−1		−2		1
92		86		2		4		−1		5
89		86		2		1		3		−2
84		86		2		−4		0		−4
79		86		−4		−3		−2		−1
81		86		−4		−1		−1		0
80		86		−4		−2		3		−5
88		86		−4		6		0		6

That is, $\mathbf{y} = \mathbf{g} + \mathbf{b}_e + \mathbf{t}_e + \mathbf{e}_{B+T}$.

The fitted values for a model are obtained by adding together the vectors in the above decomposition that correspond to the term(s) in the model, as well as those corresponding to terms marginal to those in the model. Hence the fitted values, $\widehat{E[\mathbf{Y}]} = \hat{\psi}$, for the four possible models are as follows:

$$\begin{aligned} \psi &= \mathbf{X}_G\mu & \hat{\psi} &= \mathbf{g} \\ \psi &= \mathbf{X}_B\beta & \hat{\psi} &= \mathbf{g} + \mathbf{b}_e \\ \psi &= \mathbf{X}_T\tau & \hat{\psi} &= \mathbf{g} + \mathbf{t}_e \\ \psi &= \mathbf{X}_B\beta + \mathbf{X}_T\tau & \hat{\psi} &= \mathbf{g} + \mathbf{b}_e + \mathbf{t}_e \end{aligned}$$

For the first three models, the fitted values are the vectors of the grand mean, block means and treatment means, respectively. The fitted values under the maximal model, for the example, are:

$$\begin{bmatrix}\hat{\psi}\end{bmatrix} = \begin{bmatrix}\mathbf{g}\end{bmatrix} + \begin{bmatrix}\mathbf{b}_e\end{bmatrix} + \begin{bmatrix}\mathbf{t}_e\end{bmatrix}$$

$$\begin{bmatrix}90\\91\\95\\92\\81\\82\\86\\83\\83\\84\\88\\85\\86\\87\\91\\88\\80\\81\\85\\82\end{bmatrix} = \begin{bmatrix}86\\86\\86\\86\\86\\86\\86\\86\\86\\86\\86\\86\\86\\86\\86\\86\\86\\86\\86\\86\end{bmatrix} + \begin{bmatrix}6\\6\\6\\6\\-3\\-3\\-3\\-3\\-1\\-1\\-1\\-1\\2\\2\\2\\2\\-4\\-4\\-4\\-4\end{bmatrix} + \begin{bmatrix}-2\\-1\\3\\0\\-2\\-1\\3\\0\\-2\\-1\\3\\0\\-2\\-1\\3\\0\\-2\\-1\\3\\0\end{bmatrix}$$

The sums of squares for the analysis of variance are obtained by summing the squares of the elements of the vectors $\mathbf{e}_G$, $\mathbf{b}_e$, $\mathbf{e}_B$ $\mathbf{t}_e$ and $\mathbf{e}_{B+T}$. The sums of squares are 560, 264, 296, 70 and 226, respectively.

The hypothesis test for the example RCBD is as follows:

Step 1: Set up hypotheses

a) H_0: $\tau_A = \tau_B = \tau_C = \tau_D$
H_1: at least one pair of population treatment means is different

b) H_0: $\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5$
H_1: at least one pair of population blend means is different

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
Blends.Flasks	15	296			
Treatments	3	70	23.3	1.24	0.339
Residual	12	226	18.8		
Total	19	560			

Step 3: Decide between hypotheses

It would appear that there are significant differences between the blends but not between the treatments.

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.

V.D Computation in Genstat

a) Obtaining a layout for an RCBD in Genstat

The layout for a randomized complete block design can be obtained using the *Stats > Design > Select Design* command in Genstat. Having selected this command, you need to choose *orthogonal hierarchical designs (randomized blocks, split-plots)* in answer to the question *Which type of design would you like to generate?*. You will then be asked a series of questions to which you should respond as follows:

How many block factors (or strata) does your design have?	2
What would you like to call block factor 1?	name for blocks
How many treatment factors are to be applied to the units indexed by Blocks	0
How many replicates are there of Blocks?	*b*
What would you like to call block factor 2?	name for units
How many treatment factors are to be applied to the units indexed by Blocks	1
What would you like to call treatment factor 1 (in this stratum)?	name for treatments
How many levels does treatment factor Treatments have?	*t*
Seed for randomization (0 for none)?	6-digit number
Do you want to print the design?	yes
Do you want to check the design by ANOVA	yes

Example V.1 Penicillin yield (continued)

The names to be used for the blocks, units and treatments for this example are Blend, Flask and Treat, respectively. Also, $b = 5$ and $t = 4$. Using these values in making the responses suggested above results in the following Genstat output.

```
Genstat 5  Release 4.1  (PC/Windows NT)              21 March 2000 20:40:42
Copyright 1998, Lawes Agricultural Trust (Rothamsted Experimental Station)
                ____________________________________________________

                Genstat 5 Fourth Edition - (for Windows)
                Genstat 5 Procedure Library Release PL11
                ____________________________________________________
```

```
  3  DESIGN

 *** Treatment combinations on each unit of the design ***


 Blend   1   2   3   4   5
 Flask
     1   3   2   3   1   1
     2   1   3   4   4   3
     3   2   1   2   3   2
     4   4   4   1   2   4

 Treatment factors are listed in the order: Treat



3.....................................................................

***** Analysis of variance *****
Source of variation     d.f.

Blend stratum              4

Blend.Flask stratum
Treat                      3
Residual                  12

Total                     19
```

So with the first blend, the Treatments are done in the order C, A, B, D. The analysis of variance is as expected for an RCBD.

b) Doing the ANOVA in Genstat

Example V.1 Penicillin yield (continued)

First the data is entered into a spreadsheet so that the spreadsheet contains the factors Blend, Flask and Treat and the variate Yield as illustrated in the following output.

```
Genstat 5  Release 4.1  (PC/Windows NT)              21 March 2000 20:46:33
Copyright 1998, Lawes Agricultural Trust (Rothamsted Experimental Station)

                ____________________________________________

                Genstat 5 Fourth Edition - (for Windows)
                Genstat 5 Procedure Library Release PL11
                ____________________________________________

  3  "Data taken from File: D:/ANALYSES/LM/ONEFAC/RCDBPEN.GSH"
  4  DELETE [redefine=yes] Blend,Flask,Treat,Yield
  5  FACTOR [modify=yes;nvalues=20;levels=5] Blend
  6  READ Blend; frepresentation=ordinal

   Identifier    Values   Missing    Levels
        Blend        20         0         5

  8  FACTOR [modify=yes;nvalues=20;levels=4] Flask
  9  READ Flask; frepresentation=ordinal

   Identifier    Values   Missing    Levels
        Flask        20         0         4
```

```
 11  FACTOR [modify=yes;nvalues=20;levels=4] Treat
 12  READ Treat; frepresentation=ordinal

   Identifier    Values   Missing    Levels
        Treat        20         0         4

 14  VARIATE [nvalues=20] Yield
 15  READ Yield

   Identifier   Minimum      Mean   Maximum    Values   Missing
        Yield     77.00     86.00     97.00        20         0

 17
 18  PRINT Blend,Flask,Treat,Yield

      Blend       Flask       Treat       Yield
          1           1           1       89.00
          1           2           2       88.00
          1           3           3       97.00
          1           4           4       94.00
          2           1           1       84.00
          2           2           2       77.00
          2           3           3       92.00
          2           4           4       79.00
          3           1           1       81.00
          3           2           2       87.00
          3           3           3       87.00
          3           4           4       85.00
          4           1           1       87.00
          4           2           2       92.00
          4           3           3       89.00
          4           4           4       84.00
          5           1           1       79.00
          5           2           2       81.00
          5           3           3       80.00
          5           4           4       88.00
```

Then, the Genstat instructions for analyzing the data for the example are as follows:

```
"**** Mean-operator analysis ****"

BLOCK Blend/Flask               "Factors describing physical set up"
TREAT Treat                     "Factor assigned to flasks"
ANOVA [FPROB=Y; PSE=LSD] Yield
CALC F=66/18.83 &  Prob=1-FPROB(F; 4; 12) : PRINT F,Prob
APLOT METHOD=fit,normal
```

Note the statements on the line after the ANOVA statement to compute the Blend F and p values.

The output produced by these Genstat commands is:

```
 20  "**** Mean-operator analysis ****"
 21
 22  BLOCK Blend/Flask               "Factors describing physical set up"
 23  TREAT Treat                     "Factor assigned to flasks"
 24  ANOVA [FPROB=Y; PSE=LSD] Yield
```

```
24.....................................................................

***** Analysis of variance *****

Variate: Yield

Source of variation     d.f.       s.s.       m.s.    v.r.  F pr.

Blend stratum              4     264.00      66.00    3.50

Blend.Flask stratum
Treat                      3      70.00      23.33    1.24  0.339
Residual                  12     226.00      18.83

Total                     19     560.00


***** Tables of means *****

Variate: Yield

Grand mean  86.00

    Treat        1        2        3        4
             84.00    85.00    89.00    86.00


*** Least significant differences of means (5% level) ***

Table               Treat
rep.                    5
d.f.                   12
l.s.d.              5.980


  25  CALC F=66/18.83 &  Prob=1-FPROB(F; 4; 12) : PRINT F,Prob

          F        Prob
      3.505     0.04073
```

V.E Diagnostic checking

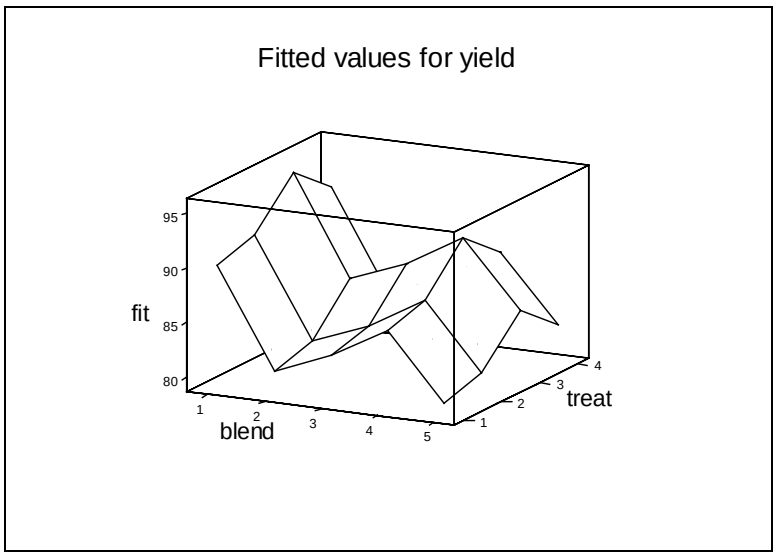
Again, we have assumed a model on which the analysis outlined above is based, namely, that $\mathbf{Y} \sim \mathrm{N}\left(\psi, \sigma^2\mathbf{I}\right)$ where, for the maximal model, $\psi = E\left[\mathbf{Y}\right] = \mathbf{X}_{\mathrm{B}}\beta + \mathbf{X}_{\mathrm{T}}\tau$. For this model to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

a) the response is operating additively, that is, that the treatments have 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.

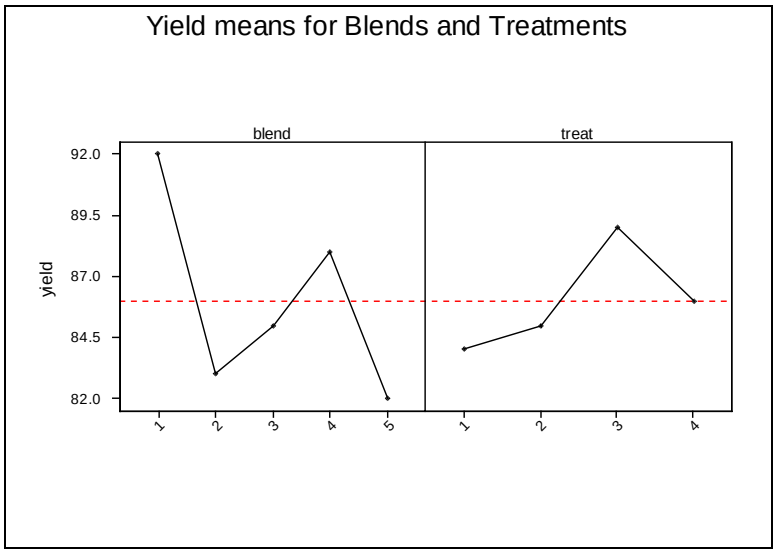
The need for and additive response comes from the maximal model for this case which is additive in Block and Treatment parameters:

$$\psi = E[\mathbf{Y}] = \mathbf{X}_B\beta + \mathbf{X}_T\tau$$

Its fitted values, $\hat{\psi} = \mathbf{B} + \mathbf{T} - \mathbf{G}$, display the additive pattern specified by this model and we hope these are an adequate description of the data. The additive pattern in the fitted values are illustrated in the following diagram.

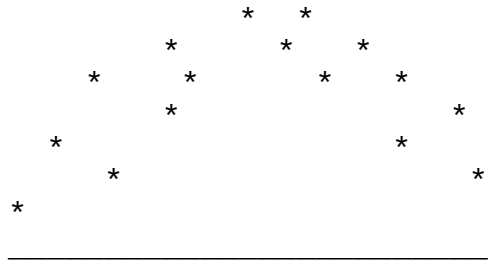


In one direction, these exhibit the same trend as the corresponding means, these trends being illustrated in the diagram below. If the additive model is to apply the above surface should describe the differences between Blend-Treatment mean combinations, except for random variations around it. Any nonrandom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



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 (except a completely randomized design) or in a regression situation. 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, **e**, on the new residuals, $\mathbf{e}_2$.

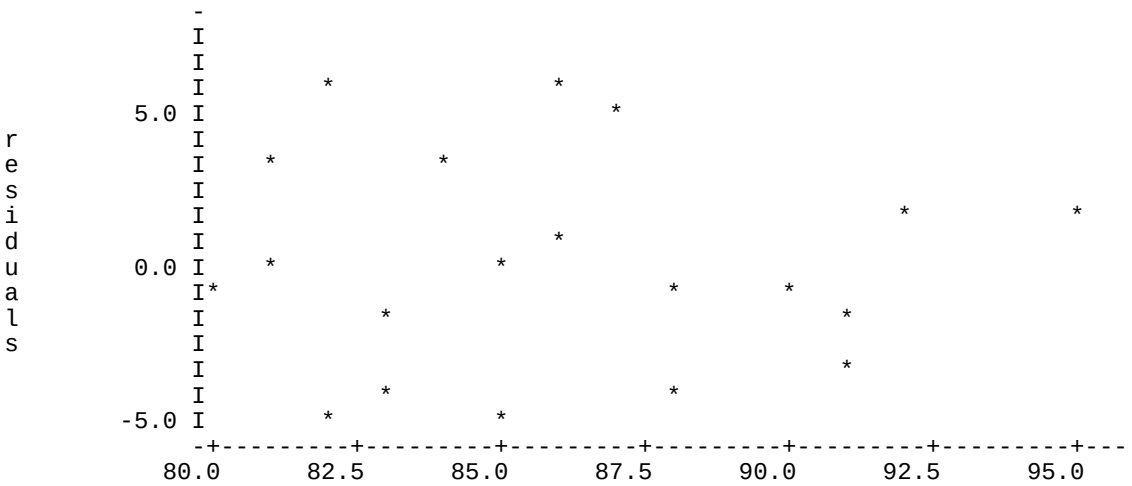
The Genstat instructions for calculating the one-degree-of-freedom-for-nonadditivity are:

```
"
**** Tukey''s one-degree-of-freedom-for-non-additivity.
**** It is the term designated covariate in the following analysis
"
AKEEP [FIT=Fit]
CALC ResSq=Fit*Fit
ANOVA [PRINT=*]  ResSq; RES=ResSq
COVAR ResSq                              "A computational trick"
ANOVA [PRINT=A; FPROB=Y] Yield
```

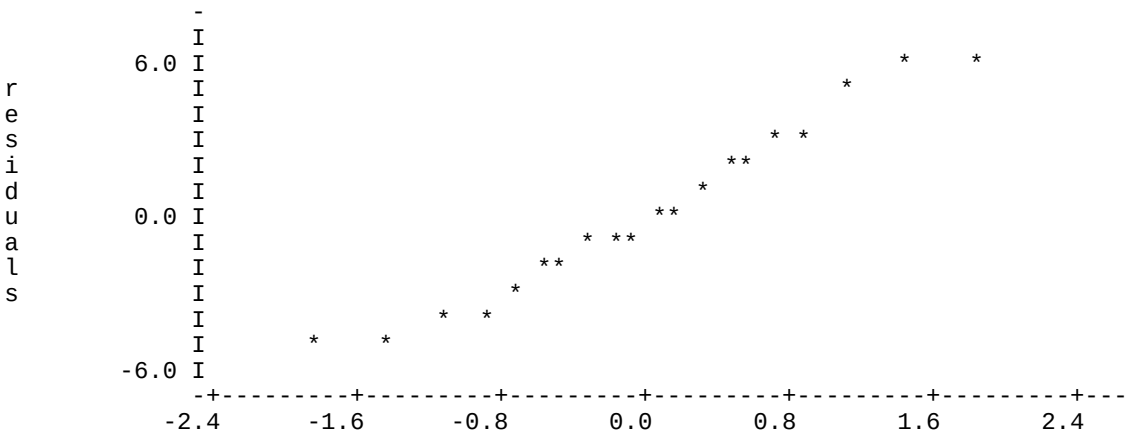
Example V.1 Penicillin yield (continued)

For the example the Genstat instructions and plots are:

```
26  APLOT METHOD=fit,normal
```



Normal plot




```
  27  "
 -28  **** Tukey''s one-degree-of-freedom-for-non-additivity.
 -29  **** It is the term designated covariate in the following analysis
 -30  "
  31  AKEEP [FIT=Fit]
  32  CALC ResSq=Fit*Fit
  33  ANOVA [PRINT=*]  ResSq; RES=ResSq
  34  COVAR ResSq                              "A computational trick"
  35  ANOVA [PRINT=A; FPROB=Y] Yield
```

```
35.............................................................

***** Analysis of variance (adjusted for covariate) *****

Variate: Yield
Covariate: ResSq

Source of variation     d.f.       s.s.       m.s.    v.r. cov.ef.  F pr.

Blend stratum              4     264.00      66.00    3.24

Blend.Flask stratum
Treat                      3      70.00      23.33    1.15    1.00  0.374
Covariate                  1       2.00       2.00    0.10          0.760
Residual                  11     224.00      20.36            0.92

Total                     19     560.00
```

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	$\sigma^2 + f_B(\psi)$	3.50	0.041
Blends.Flasks	15	296				
Treatments	3	70	23.3	$\sigma^2 + f_T(\psi)$	1.24	0.339
Residual	12	226	18.8	σ^2		
Nonadditivity	1	2	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 not additive

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

V.F 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 V.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.

V.G Fixed versus random effects

a) Alternative models for the RCBD

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

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

and the alternative model is:

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

The difference between the two models is that, in the alternative model, β has been dropped from the expectation of an observation and the covariance of observations from different plots in the same block is σ^2_{B}, 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			
Plot	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			
Plot	1	2	3	4	1	2	3	4	1	2	3	4
1	$\sigma^2+\sigma_B^2$	σ_B^2	σ_B^2	σ_B^2	0	0	0	0	0	0	0	0
2	σ_B^2	$\sigma^2+\sigma_B^2$	σ_B^2	σ_B^2	0	0	0	0	0	0	0	0
3	σ_B^2	σ_B^2	$\sigma^2+\sigma_B^2$	σ_B^2	0	0	0	0	0	0	0	0
4	σ_B^2	σ_B^2	σ_B^2	$\sigma^2+\sigma_B^2$	0	0	0	0	0	0	0	0
1	0	0	0	0	$\sigma^2+\sigma_B^2$	σ_B^2	σ_B^2	σ_B^2	0	0	0	0
2	0	0	0	0	σ_B^2	$\sigma^2+\sigma_B^2$	σ_B^2	σ_B^2	0	0	0	0
3	0	0	0	0	σ_B^2	σ_B^2	$\sigma^2+\sigma_B^2$	σ_B^2	0	0	0	0
4	0	0	0	0	σ_B^2	σ_B^2	σ_B^2	$\sigma^2+\sigma_B^2$	0	0	0	0
1	0	0	0	0	0	0	0	0	$\sigma^2+\sigma_B^2$	σ_B^2	σ_B^2	σ_B^2
2	0	0	0	0	0	0	0	0	σ_B^2	$\sigma^2+\sigma_B^2$	σ_B^2	σ_B^2
3	0	0	0	0	0	0	0	0	σ_B^2	σ_B^2	$\sigma^2+\sigma_B^2$	σ_B^2
4	0	0	0	0	0	0	0	0	σ_B^2	σ_B^2	σ_B^2	$\sigma^2+\sigma_B^2$

Notice that, for Blocks random, the covariance between plots 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 V.3: 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 V.4: 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 deviations of the treatment means from the grand mean 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 not anticipated to be systematically different 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.

Notwithswtanding 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 Plots (within Blocks) in an RCBD, Plots must be a random factor.

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 V.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.

b) Generalized least squares estimation for $\mathbf{V} \neq \sigma^2\mathbf{I}$

In this section we investigate the estimation of the expectation model parameters when we assume only that the variance matrix **V** is positive definite.

Definition V.5: A matrix **A** is positive definite if a quadratic form in the matrix $\mathbf{y}'\mathbf{A}\mathbf{y} > 0$ for $\mathbf{y} \neq \mathbf{0}$.

Theorem V.17: A symmetric matrix **A** is positive definite if and only if all its eigenvalues are positive.

Proof: not given ■

One implication of this theorem is that the inverse of a positive definite matrix must exist.

Theorem V.18: Let **Y** be a random vector whose distribution is multivariate normal and so its distribution function is

$$f(\mathbf{y};\theta,\mathbf{V}) = (2\pi)^{-n/2}|\mathbf{V}|^{-1/2}\exp\left\{-(\mathbf{y}-\psi)'\mathbf{V}^{-1}(\mathbf{y}-\psi)\right\}$$

where $\psi = E[\mathbf{Y}] = \mathbf{X}\theta$, $\text{var}[\mathbf{Y}] = \mathbf{V}$, **X** is an $n \times q$ matrix, θ is a $q \times 1$ vector of unknown parameters, **V** is an $n \times n$ positive definite matrix and $n \geq q$. Then, the maximum likelihood estimator for θ is denoted by $\tilde{\theta}$ and is given by

$$\tilde{\theta} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y}$$

Proof: Given the distribution function for **Y**, the likelihood function $L(\xi;\mathbf{y})$ is

$$L(\xi;\mathbf{y}) = (2\pi)^{-n/2}|\mathbf{V}|^{-1/2}\exp\left\{-(\mathbf{y}-\psi)'\mathbf{V}^{-1}(\mathbf{y}-\psi)\right\}$$

so that the log likelihood, $\ell = \ln\left[L(\xi;\mathbf{y})\right]$, is

$$\begin{aligned}
\ell &= \ln\left[L(\xi;\mathbf{y})\right] \\
&= \ln\left[(2\pi)^{-n/2}|\mathbf{V}|^{-1/2}\exp\left\{-(\mathbf{y}-\psi)'\mathbf{V}^{-1}(\mathbf{y}-\psi)\right\}\right] \\
&= -\frac{n}{2}\ln(2\pi) - \frac{1}{2}\ln(|\mathbf{V}|) - (\mathbf{y}-\psi)'\mathbf{V}^{-1}(\mathbf{y}-\psi)
\end{aligned}$$

The maximum likelihood estimates of θ are then obtained by maximizing ℓ with respect to θ, that is, by differentiating ℓ with respect to θ and setting the result equal to zero.

Clearly, the maximum likelihood estimates will be the solution of

$$\frac{\partial\left\{\left(\mathbf{y}-\psi\right)'\mathbf{V}^{-1}\left(\mathbf{y}-\psi\right)\right\}}{\partial\theta}=\frac{\partial\left\{\left(\mathbf{y}-\mathbf{X}\theta\right)'\mathbf{V}^{-1}\left(\mathbf{y}-\mathbf{X}\theta\right)\right\}}{\partial\theta}=\mathbf{0}.$$

Now,

$$\frac{\partial\left\{\left(\mathbf{y}-\mathbf{X}\theta\right)'\mathbf{V}^{-1}\left(\mathbf{y}-\mathbf{X}\theta\right)\right\}}{\partial\theta}=\frac{\partial\left\{\mathbf{y}'\mathbf{V}^{-1}\mathbf{y}-2\theta'\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}+\theta'\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\theta\right\}}{\partial\theta}$$

and applying the rules for differentiation given theorem II.1, we obtain

$$\begin{aligned}\frac{\partial\left\{\mathbf{y}'\mathbf{V}^{-1}\mathbf{y}-2\theta'\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}+\theta'\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\theta\right\}}{\partial\theta}&=-2\left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}\right)+\left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)\theta+\left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)'\theta\\&=-2\left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}\right)+2\left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)\theta\end{aligned}$$

Setting this derivative to zero we obtain

$$-2\left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}\right)+2\left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)\theta=0 \text{ or } \left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)\theta=\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$$

Hence, $\tilde{\theta}=\left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y}$ as claimed ■

Now the estimator is a not the normal least squares as it does not involve the variance matrix **V**. However, using the above as a lead, we define a generalized least squares estimator that does.

Definition V.6: Let **Y** be a random vector with $\psi=E\left[\mathbf{Y}\right]=\mathbf{X}\theta$ and $\text{var}\left[\mathbf{Y}\right]=\mathbf{V}$ where **X** is an $n\times q$ matrix, θ is a $q\times 1$ vector of unknown parameters, **V** is an $n\times n$ positive definite matrix and $n\geq q$. Then the generalized least squares estimator of θ is the estimator that minimizes the "sum of squares"

$$\left(\mathbf{y}-\mathbf{X}\theta\right)'\mathbf{V}^{-1}\left(\mathbf{y}-\mathbf{X}\theta\right)$$

■

Theorem V.19: Let **Y** be a random vector with $\psi = E[\mathbf{Y}] = \mathbf{X}\theta$ and $\text{var}[\mathbf{Y}] = \mathbf{V}$ where **X** is an $n \times q$ matrix, θ is a $q \times 1$ vector of unknown parameters, **V** is an $n \times n$ positive definite matrix and $n \geq q$. Then the generalized least squares estimator for θ is denoted by $\hat{\theta}$ and is given by

$$\hat{\theta} = \left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y}$$

Proof: see proof for theorem V.18 ■

Now the properties of these estimators parallel those for the normal least squares. In particular, they are BLUE as the next theorem states.

Theorem V.20: Let **Y** be a random vector with $\psi = E[\mathbf{Y}] = \mathbf{X}\theta$ and $\text{var}[\mathbf{Y}] = \mathbf{V}$ where **X** is an $n \times q$ matrix, θ is a $q \times 1$ vector of unknown parameters, **V** is an $n \times n$ positive definite matrix and $n \geq q$. The generalized least squares estimator $\hat{\theta} = \left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y}$ for θ is the best linear unbiased estimator.

Proof: not given ■

We now derive the estimator for τ in the model for the RCBD where Blocks are random. However, before doing so we derive the following useful lemma.

Lemma V.2: Let **A** and **B** be two symmetric idempotent matrices such that **I** = **A** + **B** and **AB** = **0**. Then $(a\mathbf{A} + b\mathbf{B})^{-1} = \frac{1}{a}\mathbf{A} + \frac{1}{b}\mathbf{B}$ where *a* and *b* are two real numbers.

Proof: left as an exercise for you ■

Theorem V.21: Let **Y** be an $n \times 1$ random vector in which the Y_i are arranged such that those from the same block occur consecutively in treatment order. Also, let $\psi = E[\mathbf{Y}] = \mathbf{X}_\mathrm{T}\tau = (\mathbf{1}_b \otimes \mathbf{I}_t)\tau$ and $\mathbf{V} = \sigma^2\mathbf{I}_n + \sigma_B^2(\mathbf{I}_b \otimes \mathbf{J}_t)$. Then the generalized least squares estimator of τ is given by

$$\hat{\tau} = \left(\mathbf{X}_\mathrm{T}'\mathbf{X}_\mathrm{T}\right)^{-1}\mathbf{X}_\mathrm{T}'\mathbf{Y} = \tfrac{1}{b}\mathbf{X}_\mathrm{T}'\mathbf{Y}$$

Proof: From theorem V.19, the generalized least squares estimator of τ is given by

$$\hat{\tau} = \left(\mathbf{X}_\mathrm{T}'\mathbf{V}^{-1}\mathbf{X}_\mathrm{T}\right)^{-1}\mathbf{X}_\mathrm{T}'\mathbf{V}^{-1}\mathbf{Y}$$

However, we obtain the following expression for $\mathbf{V}^{-1}$ by noting that $\mathbf{P}_\mathrm{B} = t^{-1}(\mathbf{I}_b \otimes \mathbf{J}_t)$ and by subtracting and adding $\sigma^2\mathbf{P}_\mathrm{B}$:

$$\begin{aligned}\mathbf{V}^{-1} &= \left\{\sigma^2\mathbf{I}_n + \sigma_{\mathrm{B}}^2(\mathbf{I}_b \otimes \mathbf{J}_t)\right\}^{-1} \\ &= \left(\sigma^2\mathbf{I}_n + t\sigma_{\mathrm{B}}^2\mathbf{P}_{\mathrm{B}}\right)^{-1} \\ &= \left\{\sigma^2(\mathbf{I}_n - \mathbf{P}_{\mathrm{B}}) + \left(\sigma^2 + t\sigma_{\mathrm{B}}^2\right)\mathbf{P}_{\mathrm{B}}\right\}^{-1} \\ &= \left\{\sigma^2\mathbf{R}_{\mathrm{B}} + \left(\sigma^2 + t\sigma_{\mathrm{B}}^2\right)\mathbf{P}_{\mathrm{B}}\right\}^{-1}\end{aligned}$$

Now $\mathbf{R}_{\mathrm{B}}$ and $\mathbf{P}_{\mathrm{B}}$ are symmetric and idempotent with $\mathbf{I} = \mathbf{R}_{\mathrm{B}} + \mathbf{P}_{\mathrm{B}}$ and $\mathbf{R}_{\mathrm{B}}\mathbf{P}_{\mathrm{B}} = \mathbf{0}$, so that lemma V.2 applies and

$$\begin{aligned}\mathbf{V}^{-1} &= \left\{\sigma^2\mathbf{R}_{\mathrm{B}} + \left(\sigma^2 + t\sigma_{\mathrm{B}}^2\right)\mathbf{P}_{\mathrm{B}}\right\}^{-1} \\ &= \frac{1}{\sigma^2}\mathbf{R}_{\mathrm{B}} + \frac{1}{\sigma^2 + t\sigma_{\mathrm{B}}^2}\mathbf{P}_{\mathrm{B}} \\ &= \frac{1}{\sigma^2}\mathbf{I} + \left(\frac{1}{\sigma^2 + t\sigma_{\mathrm{B}}^2} - \frac{1}{\sigma^2}\right)\mathbf{P}_{\mathrm{B}} \\ &= \frac{1}{\sigma^2}\left\{\mathbf{I} - \frac{t\sigma_{\mathrm{B}}^2}{\sigma^2 + t\sigma_{\mathrm{B}}^2}\mathbf{P}_{\mathrm{B}}\right\}\end{aligned}$$

Consequently, on noting that $(\mathbf{A} \otimes \mathbf{B})' = \mathbf{A}' \otimes \mathbf{B}'$,

$$\begin{aligned}\mathbf{X}_{\mathrm{T}}'\mathbf{V}^{-1} &= \frac{1}{\sigma^2}(\mathbf{1}_b \otimes \mathbf{I}_t)'\left\{\mathbf{I} - \frac{t\sigma_{\mathrm{B}}^2}{\sigma^2 + t\sigma_{\mathrm{B}}^2}\mathbf{P}_{\mathrm{B}}\right\} \\ &= \frac{1}{\sigma^2}\left\{\mathbf{X}_{\mathrm{T}}' - \frac{\sigma_{\mathrm{B}}^2}{\sigma^2 + t\sigma_{\mathrm{B}}^2}(\mathbf{1}_b' \otimes \mathbf{I}_t)(\mathbf{I}_b \otimes \mathbf{J}_t)\right\} \\ &= \frac{1}{\sigma^2}\left\{\mathbf{X}_{\mathrm{T}}' - \frac{\sigma_{\mathrm{B}}^2}{\sigma^2 + t\sigma_{\mathrm{B}}^2}(\mathbf{1}_b' \otimes \mathbf{J}_t)\right\}\end{aligned}$$

and

$$\begin{aligned}\mathbf{X}_{\mathrm{T}}'\mathbf{V}^{-1}\mathbf{X}_{\mathrm{T}} &= \frac{1}{\sigma^2}\left\{\mathbf{X}_{\mathrm{T}}' - \frac{\sigma_{\mathrm{B}}^2}{\sigma^2 + t\sigma_{\mathrm{B}}^2}(\mathbf{1}_b' \otimes \mathbf{J}_t)\right\}\mathbf{X}_{\mathrm{T}} \\ &= \frac{1}{\sigma^2}\left\{\mathbf{X}_{\mathrm{T}}'\mathbf{X}_{\mathrm{T}} - \frac{\sigma_{\mathrm{B}}^2}{\sigma^2 + t\sigma_{\mathrm{B}}^2}(\mathbf{1}_b' \otimes \mathbf{J}_t)(\mathbf{1}_b \otimes \mathbf{I}_t)\right\} \\ &= \frac{1}{\sigma^2}\left\{(\mathbf{1}_b' \otimes \mathbf{I}_t)(\mathbf{1}_b \otimes \mathbf{I}_t) - \frac{\sigma_{\mathrm{B}}^2}{\sigma^2 + t\sigma_{\mathrm{B}}^2}(\mathbf{1}_b' \otimes \mathbf{J}_t)(\mathbf{1}_b \otimes \mathbf{I}_t)\right\} \\ &= \frac{1}{\sigma^2}\left\{(\mathbf{1}_b'\mathbf{1}_b \otimes \mathbf{I}_t) - \frac{\sigma_{\mathrm{B}}^2}{\sigma^2 + t\sigma_{\mathrm{B}}^2}(\mathbf{1}_b'\mathbf{1}_b \otimes \mathbf{J}_t)\right\} \\ &= \frac{b}{\sigma^2}\left\{\mathbf{I}_t - \frac{\sigma_{\mathrm{B}}^2}{\sigma^2 + t\sigma_{\mathrm{B}}^2}\mathbf{J}_t\right\}\end{aligned}$$

To find the inverse of this note that $\mathbf{I}_t - \frac{1}{t}\mathbf{J}_t$ and $\frac{1}{t}\mathbf{J}_t$ are two symmetric idempotent matrices whose sum is $\mathbf{I}$ and whose product is zero. Now

$$\begin{aligned}
\left(\mathbf{X}_{\mathrm{T}}'\mathbf{V}^{-1}\mathbf{X}_{\mathrm{T}}\right)^{-1} &= \frac{\sigma^2}{b}\left\{\mathbf{I}_t - \tfrac{1}{t}\mathbf{J}_t + \tfrac{1}{t}\mathbf{J}_t - \frac{t\sigma_{\mathrm{B}}^2}{\sigma^2 + t\sigma_{\mathrm{B}}^2}\tfrac{1}{t}\mathbf{J}_t\right\}^{-1} \\
&= \frac{\sigma^2}{b}\left\{\mathbf{I}_t - \tfrac{1}{t}\mathbf{J}_t + \left(1 - \frac{t\sigma_{\mathrm{B}}^2}{\sigma^2 + t\sigma_{\mathrm{B}}^2}\right)\tfrac{1}{t}\mathbf{J}_t\right\}^{-1} \\
&= \frac{\sigma^2}{b}\left\{\mathbf{I}_t - \tfrac{1}{t}\mathbf{J}_t + \left(\frac{\sigma^2}{\sigma^2 + t\sigma_{\mathrm{B}}^2}\right)^{-1}\tfrac{1}{t}\mathbf{J}_t\right\} \\
&= \frac{\sigma^2}{b}\left\{\mathbf{I}_t + \left(\frac{\sigma^2 + t\sigma_{\mathrm{B}}^2}{\sigma^2} - 1\right)\tfrac{1}{t}\mathbf{J}_t\right\} \\
&= \frac{\sigma^2}{b}\left\{\mathbf{I}_t + \frac{\sigma_{\mathrm{B}}^2}{\sigma^2}\mathbf{J}_t\right\}
\end{aligned}$$

and

$$\begin{aligned}
\hat{\tau} &= \left(\mathbf{X}_{\mathrm{T}}'\mathbf{V}^{-1}\mathbf{X}_{\mathrm{T}}\right)^{-1}\mathbf{X}_{\mathrm{T}}'\mathbf{V}^{-1}\mathbf{Y} \\
&= \frac{\sigma^2}{b}\left\{\mathbf{I}_t + \frac{\sigma_{\mathrm{B}}^2}{\sigma^2}\mathbf{J}_t\right\}\frac{1}{\sigma^2}\left\{\mathbf{X}_{\mathrm{T}}' - \frac{\sigma_{\mathrm{B}}^2}{\sigma^2 + t\sigma_{\mathrm{B}}^2}\left(\mathbf{1}_b' \otimes \mathbf{J}_t\right)\right\}\mathbf{Y} \\
&= \frac{1}{b}\left\{\mathbf{I}_t + \frac{\sigma_{\mathrm{B}}^2}{\sigma^2}\mathbf{J}_t\right\}\left\{\mathbf{X}_{\mathrm{T}}'\mathbf{Y} - \frac{\sigma_{\mathrm{B}}^2}{\sigma^2 + t\sigma_{\mathrm{B}}^2}\left(\mathbf{1}_b' \otimes \mathbf{J}_t\right)\mathbf{Y}\right\} \\
&= \frac{1}{b}\mathbf{X}_{\mathrm{T}}'\mathbf{Y} + \frac{1}{b}\left\{\begin{array}{l}\dfrac{\sigma_{\mathrm{B}}^2}{\sigma^2}\mathbf{J}_t\mathbf{X}_{\mathrm{T}}'\mathbf{Y} - \dfrac{\sigma_{\mathrm{B}}^2}{\sigma^2 + t\sigma_{\mathrm{B}}^2}\left(\mathbf{1}_b' \otimes \mathbf{J}_t\right)\mathbf{Y} \\ \qquad\qquad - \dfrac{\sigma_{\mathrm{B}}^2}{\sigma^2}\dfrac{\sigma_{\mathrm{B}}^2}{\sigma^2 + t\sigma_{\mathrm{B}}^2}\mathbf{J}_t\left(\mathbf{1}_b' \otimes \mathbf{J}_t\right)\mathbf{Y}\end{array}\right\} \\
&= \frac{1}{b}\mathbf{X}_{\mathrm{T}}'\mathbf{Y} + \frac{1}{b}\left\{\begin{array}{l}\dfrac{\sigma_{\mathrm{B}}^2}{\sigma^2}\left(\mathbf{1}_b' \otimes \mathbf{J}_t\right)\mathbf{Y} - \dfrac{\sigma_{\mathrm{B}}^2}{\sigma^2 + t\sigma_{\mathrm{B}}^2}\left(\mathbf{1}_b' \otimes \mathbf{J}_t\right)\mathbf{Y} \\ \qquad\qquad - t\dfrac{\sigma_{\mathrm{B}}^2}{\sigma^2}\dfrac{\sigma_{\mathrm{B}}^2}{\sigma^2 + t\sigma_{\mathrm{B}}^2}\left(\mathbf{1}_b' \otimes \mathbf{J}_t\right)\mathbf{Y}\end{array}\right\} \quad \text{as } \mathbf{J}_t = 1 \otimes \mathbf{J}_t \\
&= \frac{1}{b}\mathbf{X}_{\mathrm{T}}'\mathbf{Y} + \frac{1}{b}\frac{\sigma_{\mathrm{B}}^2}{\sigma^2\left(\sigma^2 + t\sigma_{\mathrm{B}}^2\right)}\left\{\sigma^2 + t\sigma_{\mathrm{B}}^2 - \sigma^2 - t\sigma_{\mathrm{B}}^2\right\}\left(\mathbf{1}_b' \otimes \mathbf{J}_t\right)\mathbf{Y} \\
&= \frac{1}{b}\mathbf{X}_{\mathrm{T}}'\mathbf{Y}
\end{aligned}$$

■

So the fitted values under this model are $\mathbf{T} = \mathbf{P}_T\mathbf{Y}$ where $\mathbf{P}_T = \mathbf{X}_T\left(\mathbf{X}_T'\mathbf{X}_T\right)^{-1}\mathbf{X}_T'$, the same as for the model $E[\mathbf{Y}] = \mathbf{X}_T\tau$ and $\mathbf{V} = \sigma^2\mathbf{I}$.

c) Analysis of variance for Blocks random

We can no longer specify our analysis of variance by obtaining the reduction sums of squares for terms in the expectation model and the residual after all terms in the expectation model have been fitted. Since the Blocks term is no longer in the expectation model, it would not occur in the analysis of variance. Rather, we use the analysis of variance table for the model with Blocks fixed to determine the sources in the analysis of variance table; the expected mean squares for these sources are then obtained under the appropriate model.

We now consider the expected mean squares for the analysis of variance previously formulated for the RCBD, but under the model when Blocks are random.

Theorem V.22: Let $\psi = E[\mathbf{Y}] = \mathbf{X}_{\mathrm{T}}\tau$, $\mathbf{V} = \sigma^2\mathbf{I}_n + \sigma_B^2(\mathbf{I}_b \otimes \mathbf{J}_t)$, $R(\beta \mid \mu) = \mathbf{Y}'\mathbf{P}_{\mathrm{B}}\mathbf{R}_{\mathrm{G}}\mathbf{Y}$, $R(\tau \mid \mu) = \mathbf{Y}'\mathbf{P}_{\mathrm{T}}\mathbf{R}_{\mathrm{G}}\mathbf{Y}$ and $\mathrm{D}(\beta,\tau) = \mathbf{Y}'\mathbf{R}_{\mathrm{T}}\mathbf{R}_{\mathrm{B}}\mathbf{Y}$ where $\mathbf{R}_{\mathrm{T}}$, $\mathbf{R}_{\mathrm{B}}$ and $\mathbf{R}_{\mathrm{G}}$ are as defined in theorem V.8. Then,

$$E\left[R(\beta \mid \mu)/(b-1)\right] = \sigma^2 + t\sigma_{\mathrm{B}}^2$$
$$E\left[R(\tau \mid \mu)/(t-1)\right] = \sigma^2 + f_{\mathrm{T}}(\psi),$$
$$E\left[\mathrm{D}(\beta,\tau)/\{(b-1)(t-1)\}\right] = \sigma^2$$

where $f_{\mathrm{T}}(\psi) = \sum_{j=1}^{t} b(\tau_j - \bar{\tau}_\cdot)^2 \Big/ (t-1)$, $\bar{\tau}_\cdot = \sum_{j=1}^{t}\tau_j \Big/ t$, τ_j is the jth element of the t-vector τ, b is the number of blocks and t is the number of treatments.

Proof: For $E\left[R(\beta \mid \mu)/(b-1)\right]$, we first use theorem II.11 to show that

$$\begin{aligned} E\left[R(\beta \mid \mu)/(b-1)\right] &= E[\mathbf{Y}'\mathbf{P}_{\mathrm{B}}\mathbf{R}_{\mathrm{G}}\mathbf{Y}]/(b-1) \\ &= \left\{trace\left(\mathbf{P}_{\mathrm{B}}\mathbf{R}_{\mathrm{G}}\left\{\sigma^2\mathbf{I}_n + \sigma_{\mathrm{B}}^2(\mathbf{I}_b \otimes \mathbf{J}_t)\right\}\right) + (\mathbf{X}_{\mathrm{T}}\tau)'\mathbf{P}_{\mathrm{B}}\mathbf{R}_{\mathrm{G}}(\mathbf{X}_{\mathrm{T}}\tau)\right\}\Big/\{b-1\} \\ &= \left\{\sigma^2 trace(\mathbf{P}_{\mathrm{B}}\mathbf{R}_{\mathrm{G}}) + \sigma_{\mathrm{B}}^2 trace\left(\mathbf{P}_{\mathrm{B}}\mathbf{R}_{\mathrm{G}}\{\mathbf{I}_b \otimes \mathbf{J}_t\}\right) + (\mathbf{X}_{\mathrm{T}}\tau)'\mathbf{P}_{\mathrm{B}}\mathbf{R}_{\mathrm{G}}(\mathbf{X}_{\mathrm{T}}\tau)\right\}\Big/\{b-1\} \end{aligned}$$

Now in proving theorem V.11 it was demonstrated that $\mathbf{P}_{\mathrm{B}}\mathbf{R}_{\mathrm{G}}(\mathbf{X}_{\mathrm{T}}\tau) = \mathbf{0}_{bt\times t}$ and from theorem V.10 $trace(\mathbf{P}_{\mathrm{B}}\mathbf{R}_{\mathrm{G}}) = b-1$.

Also, from lemma V.I, $\mathbf{P}_{\mathrm{B}} = t^{-1}\mathbf{I}_b \otimes \mathbf{J}_t$ so that

$$\mathbf{P}_{\mathrm{B}}\mathbf{R}_{\mathrm{G}}\{\mathbf{I}_b \otimes \mathbf{J}_t\} = t\mathbf{P}_B\mathbf{R}_{\mathrm{G}}\mathbf{P}_{\mathrm{B}} = t\mathbf{P}_{\mathrm{B}}(\mathbf{I}-\mathbf{P}_{\mathrm{G}})\mathbf{P}_{\mathrm{B}} = t(\mathbf{P}_{\mathrm{B}}-\mathbf{P}_{\mathrm{G}})\mathbf{P}_{\mathrm{B}} = t(\mathbf{P}_{\mathrm{B}}-\mathbf{P}_{\mathrm{G}}) = t\mathbf{P}_{\mathrm{B}}\mathbf{R}_{\mathrm{G}}$$

Hence the expected mean square is

$$E\left[R(\beta\,|\,\mu)/(b-1)\right]=\left\{\sigma^2 trace(\mathbf{P}_B\mathbf{R}_G)+\sigma_B^2 trace\left(\mathbf{P}_B\mathbf{R}_G\{\mathbf{I}_b\otimes\mathbf{J}_t\}\right)+(\ \mathbf{X}_T\tau)'\mathbf{P}_B\mathbf{R}_G(\mathbf{X}_T\tau)\right\}\Big/\{b-1\}$$
$$=\left\{\sigma^2(b-1)+\sigma_B^2 trace(t\mathbf{P}_B\mathbf{R}_G)+\mathbf{0}_{bt\times t}\right\}\Big/\{b-1\}$$
$$=\left\{\sigma^2(b-1)+t\sigma_B^2(b-1)\right\}\Big/\{b-1\}$$
$$=\sigma^2+t\sigma_B^2$$

The proof that $E\left[R(\tau\,|\,\mu)/(t-1)\right]=\sigma^2+f_T(\psi)$ is left as an exercise for you.

Recalling that in theorem V.8 it was proved that $\mathbf{R}_T\mathbf{R}_B\mathbf{R}_G=\mathbf{R}_T\mathbf{R}_B$, we have

$$E\left[\mathrm{D}(\beta,\tau)/\{(b-1)(t-1)\}\right]$$
$$=E\left[\mathbf{Y}'\mathbf{R}_T\mathbf{R}_B\mathbf{R}_G\mathbf{Y}\right]/\{(b-1)(t-1)\}$$
$$=\frac{\left\{trace\left(\mathbf{R}_T\mathbf{R}_B\left\{\sigma^2\mathbf{I}_n+\sigma_B^2(\mathbf{I}_b\otimes\mathbf{J}_t)\right\}\right)+(\ \mathbf{X}_T\tau)'\mathbf{R}_T\mathbf{R}_B\mathbf{R}_G(\mathbf{X}_T\tau)\right\}}{\{(b-1)(t-1)\}}$$
$$=\frac{\left\{\sigma^2(b-1)(t-1)+t\sigma_B^2 trace(\mathbf{R}_T\mathbf{R}_B\mathbf{P}_B)+(\ \mathbf{X}_T\tau)'\mathbf{R}_T\mathbf{R}_B\mathbf{R}_G(\mathbf{X}_T\tau)\right\}}{\{(b-1)(t-1)\}}$$

Now, it can be shown that $\mathbf{R}_T\mathbf{R}_B\mathbf{P}_B=\mathbf{0}$ and that $\mathbf{R}_T\mathbf{R}_B\mathbf{R}_G(\mathbf{X}_T\tau)=\mathbf{0}_{bt\times 1}$ so that

$$E\left[\mathrm{D}(\beta,\tau)/\{(b-1)(t-1)\}\right]$$
$$=\frac{\left\{\sigma^2(b-1)(t-1)+t\sigma_B^2 trace(\mathbf{R}_T\mathbf{R}_B\mathbf{P}_B)+(\ \mathbf{X}_T\tau)'\mathbf{R}_T\mathbf{R}_B\mathbf{R}_G(\mathbf{X}_T\tau)\right\}}{\{(b-1)(t-1)\}}$$
$$=\frac{\left\{\sigma^2(b-1)(t-1)+0+0\right\}}{\{(b-1)(t-1)\}}$$
$$=\sigma^2$$

as claimed. ■

The difference between the two models leads to differences in the expected mean squares. The two sets of expected mean squares are shown in the following table:

		E[MSq]	
Source	df	Blocks Fixed	Blocks Random
Blocks	*b*-1	$\sigma^2+f_B(\psi)$	$\sigma^2+t\sigma_B^2$

Blocks.Plots	$b(t\text{-}1)$		
Treatments	t-1	$\sigma^2 + \mathrm{f}_\mathrm{T}(\psi)$	$\sigma^2 + \mathrm{f}_\mathrm{T}(\psi)$
Residual	$(b\text{-}1)(t\text{-}1)$	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 $f_\mathrm{B}(\psi) = \sum t(\beta_i - \bar{\beta}_\cdot)^2 / (b-1)$ has become $t\sigma_\mathrm{B}^2$. Also the hypotheses for Blocks becomes:

$$\mathrm{H}_0\text{: } \sigma_\mathrm{B}^2 = 0$$
$$\mathrm{H}_1\text{: } \sigma_\mathrm{B}^2 \neq 0$$

The F-statistic for testing this hypothesis is again the ratio of the Block and Residual mean squares. The sampling distribution of this F-statistic is a Snedecor's F with $(b-1)$ and $(b-1)(t-1)$ degrees of freedom. This result is given by theorem V.16 because under the null hypothesis of $\sigma_\mathrm{B}^2 = 0$ the expectation and variance models are the same as under the null hypothesis for fixed block effects, namely all block effects equal to zero.

Thus the test for both fixed and random block effects are the same. This is not always the case.

V.H Sample size

a) Type I and II errors

It is important to keep in mind that your conclusion about the null hypothesis is not 100% certain to be correct. There is always the chance that you are wrong, although in some cases the chance is so small that you will be virtually 100% certain. Remember, however, that very unusual events do happen; for example, Baum and Scheuer (1976) report a case of the same person winning the same lottery twice, the chance of doing so being 500 million to 1.

The possible outcomes — the conclusion (or verdict) reached as a result of a hypothesis test — are illustrated in the following table:

Relationship between reality and verdict for $\mathbf{H_0}$

	H_0 reality	
H_0 verdict	true	false
not rejected	correct (innocent cleared)	type II error (guilty cleared)

rejected	type I error (innocent convicted)	correct (guilty convicted)

Thus there are two types of errors can be made in performing a hypothesis test.

Definition V.7: A **type I error** is made when the null hypothesis is true and it is rejected. The probability of a type I error is designated α and is:

$$P\{H_0 \text{ rejected} \mid \mathrm{H}_0 \text{ true}\} = \alpha .$$

■

Definition V.8: A **type II error** is made when the null hypothesis is false and it is not rejected. The probability of a type II error is designated β and is:

$$P\{H_0 \text{ not rejected} \mid \mathrm{H}_0 \text{ false}\} = \beta$$

■

Using the analogy that deciding between the null and alternative hypotheses in a hypothesis test is the same as deciding the verdict in a trial, a type I error is the same as convicting an innocent person while a type II error is the same as failing to convict a guilty person.

In a hypothesis test we set the probability of a type I error at α, which is called the significance level, α; that is we set the level of risk we are prepared to take in making a type I error. But what about the probability of a Type II error, β? Rather than consider β, we often consider $1-\beta$, which is called the power of the test.

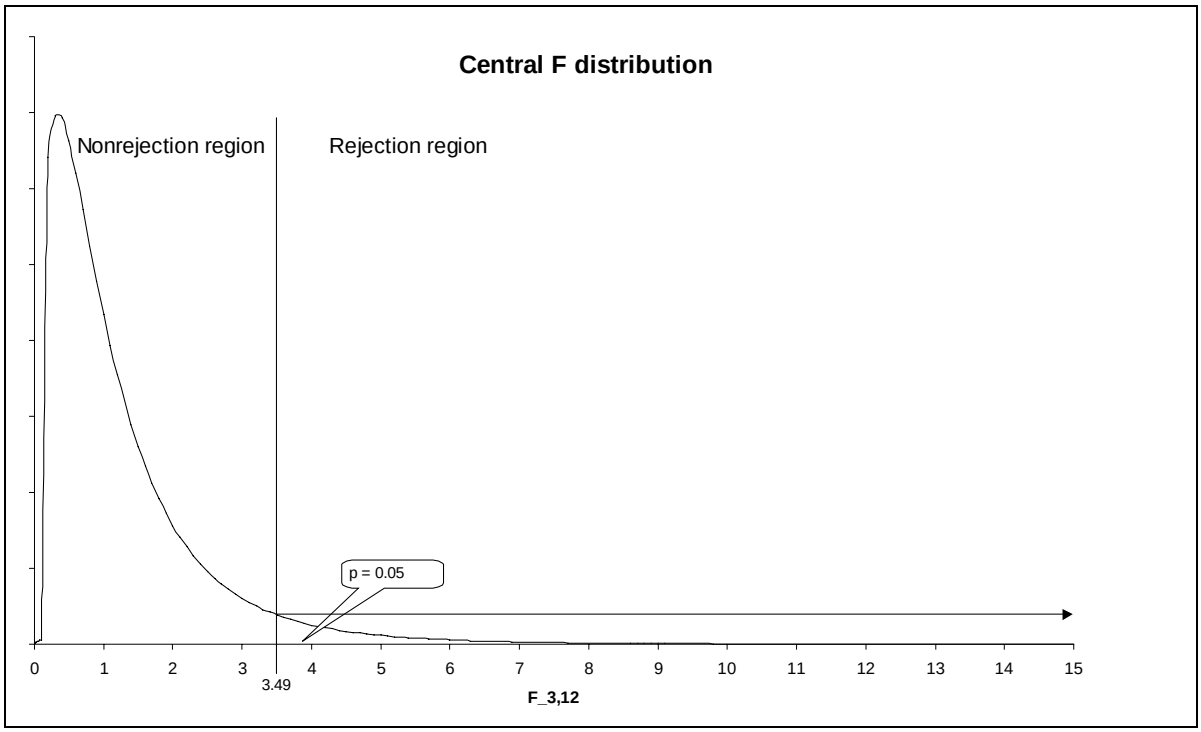
b) Power of a hypothesis test about expectation model terms

Definition V.9: The **power** of a hypothesis test is the probability of rejecting the null hypothesis when it is false is:

$$1-\beta = P\{H_0 \text{ rejected} \mid \mathrm{H}_0 \text{ false}\}$$

■

Now to compute this probability means that we need to know the condition under which the null hypothesis is rejected. In general the null hypothesis is rejected when the computed value of the probability of the F statistics from the analysis of variance is less than α. This will occur whenever the observed value of the F is greater than the α% value from the Snedecor's F distribution as illustrated in the following diagram for $F_{0.05,3,12}$. In this case any observed value of the test statistic greater than 3.49 would result in a p-value of less than 0.05 and so be rejected at the 5% level.



The next quantity that needs to be specified is how false is the null hypothesis; that is, how big a difference was expected in the treatments, relative to the magnitude of the uncontrolled variation. For the CRD and RCBD this is given by

$$\lambda = \frac{(t-1)f_{\mathrm{T}}(\psi)}{\sigma^2} = \frac{b\sum_{j=1}^{t}(\tau_j - \bar{\tau}_\cdot)^2}{\sigma^2}$$

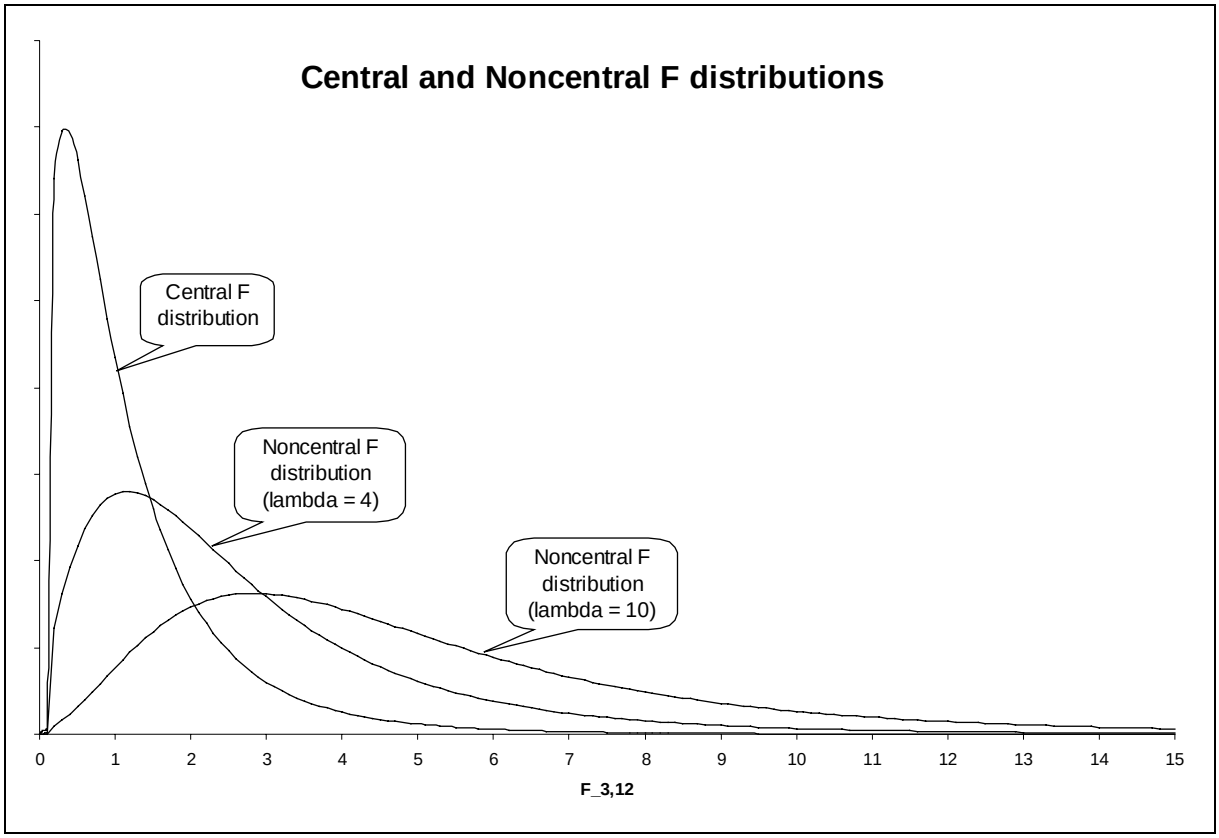
provided the number of replicates for all treatments is *b*. It also applies to the soon to be discussed Latin square, except that $b=t$ for this case. However, to use this formula requires that we know the values of the τs which is unlikely as estimating these is the purpose of the experiment. If they were known, the experiment would be irrelevant. A way to overcome this is to specify Δ, the difference such that if any two population means differ by this amount, the null hypothesis should be rejected. Then it can be shown that a general formula for the minimum value of λ is

$$\lambda = \frac{m\Delta^2}{2\sigma^2}$$

where *m* is the number of values used in computing one of the means being compared.

To compute the power we need the probability of getting a value of F greater than that of the α% value from Snedecor's F distribution when there is a difference between the treatments of the order specified by Δ (or λ). Clearly, Snedecor's distribution cannot be used to compute this probability as it is the distribution that applies when the null hypothesis is true: there is no difference between the τs in the population. We need a distribution of F for when the null hypothesis is false. This is

provided by the noncentral F distribution, a modification of the (central) Snedecor's F distribution to incorporate a noncentrality parameter λ. The shape of the noncentral F distribution depends on ν_1, ν_2 and λ — for the central F distribution $\lambda = 0$.



Distribution for:

$\lambda = 0$ is distribtion of $F_{3,12}$ when H_0 true
$\lambda = 4$ is distribtion of $F_{3,12}$ when H_0 is not true
$\lambda = 10$ is distribtion of $F_{3,12}$ when H_0 is even less true

Specifically, to compute the power of an analysis of variance test for a fixed factor:

$$\begin{aligned} 1-\beta &= 1-P\{H_0 \text{ not rejected} \mid H_0 \text{ false}\} \\ &= P\{H_0 \text{ rejected} \mid H_0 \text{ false}\} \\ &= P\{F_{\nu_1,\nu_2,\lambda} \geq F_{\alpha,\nu_1,\nu_2}\} \end{aligned}$$

where F_{α,ν_1,ν_2} is the F value from the central F distribution such that $P\{F_{\nu_1,\nu_2} \geq F_{\alpha,\nu_1,\nu_2}\} = \alpha$

Two ways to compute probabilities for the noncentral F distribution :

1. use function NCF from the Excel add-in, *PiFace.xla*.

   The add-in is available in the *G:\Disciplina\Genstat* directory in the Departmento de Cienĉias Florestais PC pool and from the web site for the subject in a self-extracting archive.

   To load use *Ferramentaos > Suplementos... > Procurar* or double-click on it in the Explorer. Ignore the two error messages that occur while loading it.

   The add-in needs to be reinstalled each time you reboot the computer in a pool — if you own your own computer, you can permanently install the add-in on it.

   Need to supply the function NCF with F, DF1, DF2 and λ.

2. Use tables or graphs

To compute the power we will use an Excel worksheet that has been set up in the file ANOVAPower.xls. It contains the columns given in the following table and the row of cells under these column headings needs to be modified as indicated.

Column heading	Modification to cell under heading
sample size (r)	enter a value
alpha	enter a value
DF numerator	enter a value
DF denominator	enter the formula '=a*(A5-b)' where a and b are constants that you work out for your proposed design
central F	leave this formula as it is '=FINV(B5,C5,D5)'
no. values in a mean (m)	enter the formula '=c*A5' where c is a constant that you work out for your proposed design
delta	enter a value
standard deviation	enter a value
lambda	leave this formula as it is '=F5*G5*G5/2/H5/H5'
power	leave this formula as it is '=1-NCF(E5,C5,D5,I5)'

Note use of the functions FINV and NCF: =FINV(α, ν_1, ν_2) computes the central F values (F_{α,ν_1,ν_2}) and =1-NCF(F_{α,ν_1,ν_2} , ν_1, ν_2, λ) computes $1 - \beta$;.

Once you have set up the worksheet as described, obtain the value for the power from the cell below the power heading.

Example V.1 Penicillin yield (continued)

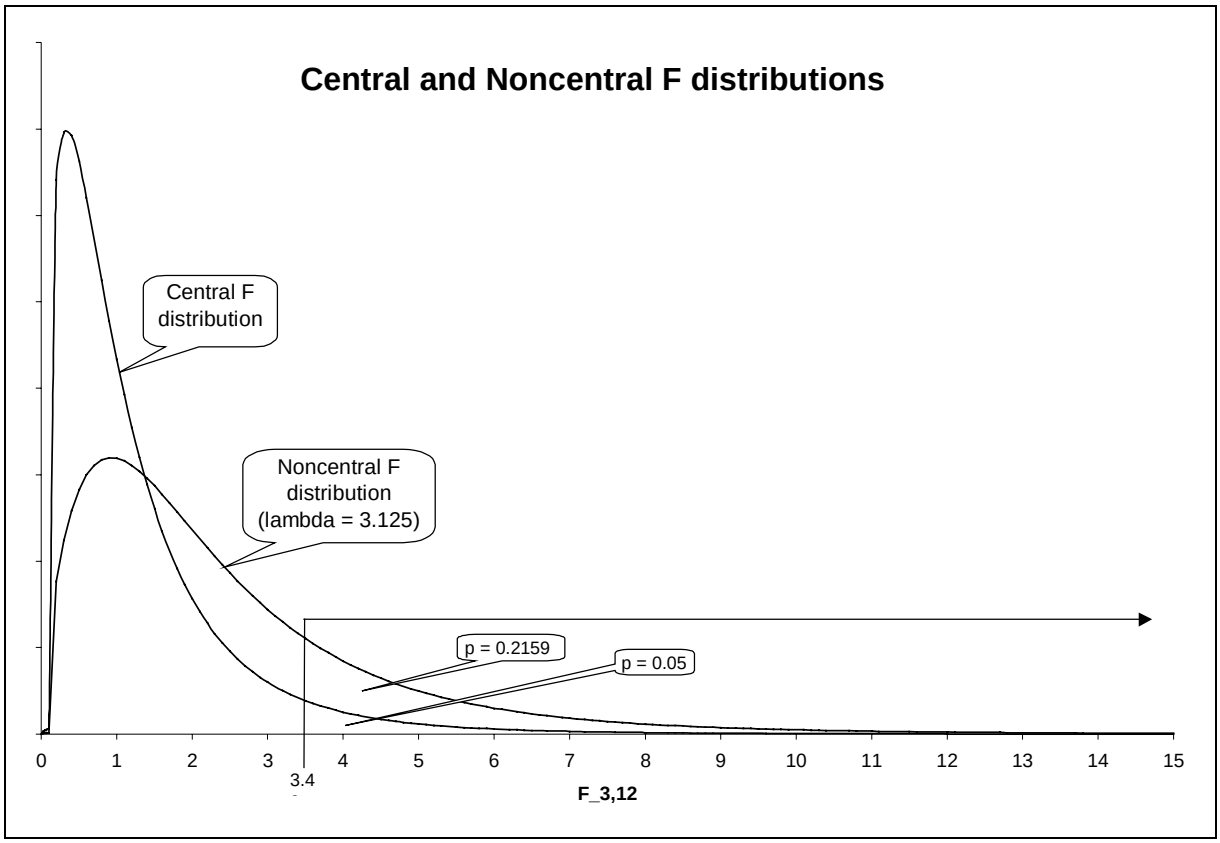
Suppose it was expected that the minimum difference between a pair of treatment means is 5 and that $\alpha = 0.05$. In the analysis of variance for this experiment, the Residual MSq was 18.83 so we will take $\sigma^2 \approx 20$. The entries in the Excel worksheet should be completed as follows:

Column heading	Cell contents
sample size (r)	5
alpha	0.05
DF numerator	3
DF denominator	=3*(A5-1)
central F	=FINV(B5,C5,D5)
no. values in a mean (m)	=1*A5
delta	5
standard deviation	=sqrt(20)
lambda	=F5*G5*G5/2/H5/H5
power	=1-NCF(E5,C5,D5,I5)

The completed worksheet has the following values

sample size (r)	alpha	DF numer-ator	DF denomin-ator	central F	no. values in a mean (m)	delta	standard deviation	lambda	power
5	0.05	3	12	3.4903	5	5	4.472136	3.125	0.2159

The power for detecting a minimum difference of 5 in a pair of treatment means when $\sigma = 4.47$ is 0.2159. Note that $\lambda = 3.125$. The situation is illustrated in the following diagram. If there was no difference between the treatments, the probability of rejecting the null hypothesis is determined from the central F distribution to be 0.05. However, if the minimum size of the treatment difference is as measured by $\lambda = 3.125$, the probability of rejecting the null hypothesis is determined from the noncentral F distribution to be 0.2159.



The power of the hypothesis test being 0.22 is not good because we do not have a high chance of correctly rejecting the null hypothesis. How can we improve things? ■

For a fixed number of treatments, the power will increase as the noncentrality parameter and the residual degrees of freedom increase. Examination of the formula for the noncentrality parameter leads us to conclude that, for a fixed number of treatments, the noncentrality parameter will increase if

1. the number of replicates, r;
2. the size of the differences between the treatments, as measured by Δ, is increased;
3. the uncontrolled variation, σ^2, is decreased.

So we can get better power by taking more observations, by increasing the differences between the treatments or by decreasing the uncontrolled variation. The latter can be done either by improving the protocols used in conducting the experiment or by employing a design that reduces the uncontrolled variation affecting treatment comparisons. There may be little avenue for improving the protocols — hopefully, the experimenter is already being as careful as possible. It might be possible to improve the design — for example, if a CRD is being used, it might be possible to change to an RCBD which has the effect of removing block difference from σ^2 and so reducing it. However, most often it is the number of replicates that will have to be increased.

c) Computing the required sample size for the CRD and RCBD

Computing the required sample size is achieved by computing the power for different sample sizes until the smallest sample size that has at least the required power is identified. Given the discussion on the power of an hypothesis test, in order to compute the sample size you will have to specify:

1. the significance level, α;
2. the power desired, $1 - \beta$;
3. the number of treatments to be investigated
4. the minimum size of the differences between a pair of treatment means, as measured by Δ;
5. the uncontrolled variation, σ^2, expected.

In working out values for these often the results from a previous experiment you or another researcher has run will be useful.

Once you have set up the worksheet as described examine the value computed for the power. If it is a) too high, decrease the sample size, or b) too low, increase the sample size, continuing in both cases until you have identified the smallest sample size that has at least the required power.

Example V.1 Penicillin yield (continued)

We now determine the number of replicates required to achieve a power of 0.80 in detecting $\Delta = 5$ with $\alpha = 0.05$. We continue to take $\sigma^2 \approx 20$. Because 5 replicates achieved a power of only 0.22 and is clearly much smaller than required, we take $r = 15$ as our first guess and find that this would provide a power of 0.6860. Next we increase r to 20 and this indicates a power of 0.8289 will be achieved. As this is in excess of the required power, r is reduced to 19 and the power reduces to 0.8056. As this is still in excess of the required power, r is reduced to 18 and the power reduces to 0.7798. Clearly, a sample size of 19 is required to achieve a power of at least 0.80 — fewer replicates will have less power. The values in the worksheet for this number of replicates is as follows:

sample size (r)	alpha	DF numer-ator	DF denomin-ator	central F	no. values in a mean (m)	delta	standard deviation	lambda	power
19	0.05	3	54	2.7758	19	5	4.472136	11.875	0.8056

■