STATISTICAL MODELLING

V. Latin squares designs (LS)

(Mead sec.8.1)

V.A Design of Latin squares

In the RCBD we arranged for the isolation of one set of differences, thereby eliminating their effect on treatment differences. Thus, in the penicillin example, we arranged for the elimination of blend differences. In a field trial, if there was a trend in a particular direction, this trend could be isolated by having blocks run perpendicular to the trend. However, suppose moisture is varying across the field and the stoniness down the field. The RCBD cannot be used to eliminate both sources of variability — a Latin square design is required.

Definition V.1: A **Latin square design** is one in which each treatment occurs once and only once in each row and each column so that the numbers of rows, columns and treatments are all equal. ■

Clearly, the total number of observations is $n = t^2$.

Example V.1 Fertilizer experiment

Suppose there are five different fertilizers to be compared; a Latin square design for this would be as follows:

5 x 5 Latin square

		Column				
		1	2	3	4	5
	I	A	B	C	D	E
	II	C	D	E	A	B
Row	III	E	A	B	C	D
	IV	B	C	D	E	A
	V	D	E	A	B	C

Less stony of field ⇓ ⇓ ⇓ Stonier end of field

Less moisture ⇒ ⇒ ⇒ More moisture

(Fertilizers A, B, C, D, E)

■

Even if one has not identified trends in two directions, a Latin square may be employed to guard against the problem of putting the blocks in the wrong direction. Latin squares may also be used when there are two different kinds of blocking variables — for example animals and times. The general principle is that one is interested in maximizing row and column differences so as to minimize the amount of uncontrolled variation affecting treatment comparisons.

The major disadvantage with the Latin square is that you are restricted to having the number of replicates equal to the number of treatments. Several fundamentally different Latin squares exist for a particular t — for $t = 4$ there are three different squares. A collection of Latin squares for $t = 3,4, ..., 9$ is given in Appendix 8A of Box, Hunter and Hunter. To randomize these designs appropriately involves the following:

1. randomly select one of the designs for a value of t;
2. randomly permute the rows and then the columns;
3. randomly assign letters to treatments.

a) Obtaining a layout for a Latin square in R

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

To use these expressions to generate a layout for a particular case, you will need to substitute the actual values for t and n and the actual names for *Rows*, *Columns, Treats* and the data frame to contain them. Also, you will need to put in an unrandomized Latin square layout and, optionally, the labels for the treatments. Note that, like the randomized complete block design the Latin square involves two unrandomized factors, *Rows* and *Columns* say, that index the units. However, for the Latin square, as the *Rows* and *Columns* are to be randomized independently, they are not nested (they are crossed), and so the `nested.factors` argument is not set and so will be `NULL`.

Example V.2 Pollution effects of petrol additives

Suppose that four cars and four drivers are employed in a study of possible differences between four petrol additives as far as their effect on pollution is concerned. Even if the cars are identical models, consistent differences are likely to occur between individual cars. Even though each driver may do their best to drive in the manner required by the test, consistent differences are likely to occur between the drivers. It would be desirable to isolate both car-to-car and driver-to-driver differences. A 4×4 Latin square would enable this to be done.

The names to be used for the rows, columns and treats for this example are Cars, Drivers and Additives, respectively. Also, $t = 4$ and a suitable design was obtained from Box, Hunter and Hunter. Substituting these into the general expressions in Appendix B, *Randomized layouts and sample size computations in R*, yields the following expressions to be used in this case:

```
> t <- 4
> n <- t*t
> LSPolut.unit <- list(Drivers=t, Cars=t)
> Additives <- factor(c(1,2,3,4, 4,3,2,1, 2,4,1,3, 3,1,4,2),
+                                              labels=c("A","B","C","D"))
> LSPolut.lay <- fac.layout(unrandomized=LSPolut.unit, randomized=Additives,
+                           seed=941)
> remove("Additives")
> LSPolut.lay
   Units Permutation Drivers Cars Additives
1      1          11       1    1         B
2      2          12       1    2         D
3      3          10       1    3         C
4      4           9       1    4         A
5      5           7       2    1         A
6      6           8       2    2         B
7      7           6       2    3         D
8      8           5       2    4         C
9      9          15       3    1         D
10    10          16       3    2         C
11    11          14       3    3         A
12    12          13       3    4         B
13    13           3       4    1         C
14    14           4       4    2         A
15    15           2       4    3         B
16    16           1       4    4         D
```

Thus the randomized layout is:

4 x 4 Latin square

		Car			
		1	2	3	4
	I	B	D	C	A
	II	A	B	D	C
Drivers	III	D	C	A	B
	IV	C	A	B	D

(Additives A, B, C, D)

■

V.B Indicator-variable models and estimation for a Latin square

There are a number of different maximal models for this design depending on whether each of the factors Rows, Columns and Treatments are to be regarded as fixed or random. As for the randomized complete block design, it happens that the analysis of variance for the Latin square is essentially unaffected by which model is used.

a) Maximal model

Generally, the Latin square involves t rows and columns so that there are $n = t^2$ observations in all. The maximal model for the observations from a Latin square, when Rows, Columns and Treatments are fixed, is

$$\psi_{\mathrm{R+C+T}} = E[\mathbf{Y}] = \mathbf{X}_{\mathrm{R}}\beta + \mathbf{X}_{\mathrm{C}}\delta + \mathbf{X}_{\mathrm{T}}\tau \text{ and } \mathbf{V} = \sigma^{\mathbf{2}}\mathbf{I}_{t^2},$$

where $\mathbf{Y}$ is the t^2-vector random variables for the response variable observations

- β is the t-vector of parameters specifying a different mean response for each row,
- $\mathbf{X}_{\mathrm{R}}$ is the $t^2 \times t$ matrix indicating the row from which an observation came,
- δ is the t-vector of parameters specifying a different mean response for each column,
- $\mathbf{X}_{\mathrm{C}}$ is the $t^2 \times t$ matrix indicating the column from which an observation came,
- τ is the t-vector of parameters specifying a different mean response for each treatment,
- $\mathbf{X}_{\mathrm{T}}$ is the $t^2 \times t$ matrix indicating the observations that received each of the treatments.

Our model also involves assuming $\mathbf{Y} \sim N(\psi_{\mathrm{R+C+T}}, \mathbf{V})$.

Example V.3 A 3×3 Latin square

For example, suppose that a 3×3 Latin square with the following arrangement of treatments was being considered:

3 x 3 Latin square

		Column		
		1	2	3
	I	A	B	C
Row	II	C	A	B
	III	B	C	A

Suppose the random vector Y is ordered in standard order for Rows then Columns.

Then, for this example,

$$\mathbf{X}_{\mathrm{R}}=\begin{bmatrix}1&0&0\\1&0&0\\1&0&0\\0&1&0\\0&1&0\\0&1&0\\0&0&1\\0&0&1\\0&0&1\end{bmatrix},\ \mathbf{X}_{\mathrm{C}}=\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\\1&0&0\\0&1&0\\0&0&1\\1&0&0\\0&1&0\\0&0&1\end{bmatrix},\text{ and }\mathbf{X}_{\mathrm{T}}=\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\\0&0&1\\1&0&0\\0&1&0\\0&1&0\\0&0&1\\1&0&0\end{bmatrix}$$

■

Note that for the general systematic layout $\mathbf{X}_{\mathrm{R}}=\mathbf{I}_t\otimes\mathbf{1}_t$ and $\mathbf{X}_{\mathrm{C}}=\mathbf{1}_t\otimes\mathbf{I}_t$, but that $\mathbf{X}_{\mathrm{T}}$ cannot be written as a direct product.

Then $\hat{\psi}_{\mathrm{R+C+T}}=\bar{\mathbf{R}}+\bar{\mathbf{C}}+\bar{\mathbf{T}}-2\bar{\mathbf{G}}$ where $\bar{\mathbf{R}}$, $\bar{\mathbf{C}}$, $\bar{\mathbf{T}}$ and $\bar{\mathbf{G}}$ are the t^2-vectors of row, column, treatment and grand means, respectively.

Also, note that $\bar{\mathbf{R}}=\mathbf{M}_{\mathrm{R}}\mathbf{Y}$, $\bar{\mathbf{C}}=\mathbf{M}_{\mathrm{C}}\mathbf{Y}$, $\bar{\mathbf{T}}=\mathbf{M}_{\mathrm{T}}\mathbf{Y}$, and $\bar{\mathbf{G}}=\mathbf{M}_{\mathrm{G}}\mathbf{Y}$. That is, $\mathbf{M}_{\mathrm{R}}$, $\mathbf{M}_{\mathrm{C}}$, $\mathbf{M}_{\mathrm{T}}$ and $\mathbf{M}_{\mathrm{G}}$ are the row, column, treatment and grand mean operators, respectively. So once again the estimators of the expected values are functions of means.

Further, if the data in the vector **Y** has been arranged in standard order for Rows then Columns, the operators are:

$$\mathbf{M}_{\mathrm{G}}=t^{-2}\mathbf{J}_t\otimes\mathbf{J}_t$$
$$\mathbf{M}_{\mathrm{R}}=t^{-1}\mathbf{I}_t\otimes\mathbf{J}_t$$
$$\mathbf{M}_{\mathrm{C}}=t^{-1}\mathbf{J}_t\otimes\mathbf{I}_t$$

In this case it is not possible to write $\mathbf{M}_{\mathrm{T}}$ as a direct product of **I** and **J** matrices as the treatments will not be in a systematic order expressible in this form.

Example V.3 A 3×3 Latin square (continued)

For this example,

$$\mathbf{M}_{\mathrm{G}} = \tfrac{1}{9}\mathbf{J}_3 \otimes \mathbf{J}_3 = \tfrac{1}{9}\begin{bmatrix} 1&1&1&1&1&1&1&1&1 \\ 1&1&1&1&1&1&1&1&1 \\ 1&1&1&1&1&1&1&1&1 \\ 1&1&1&1&1&1&1&1&1 \\ 1&1&1&1&1&1&1&1&1 \\ 1&1&1&1&1&1&1&1&1 \\ 1&1&1&1&1&1&1&1&1 \\ 1&1&1&1&1&1&1&1&1 \\ 1&1&1&1&1&1&1&1&1 \end{bmatrix}$$

$$\mathbf{M}_{\mathrm{R}} = \tfrac{1}{3}\mathbf{I}_3 \otimes \mathbf{J}_3 = \tfrac{1}{3}\begin{bmatrix} 1&1&1&0&0&0&0&0&0 \\ 1&1&1&0&0&0&0&0&0 \\ 1&1&1&0&0&0&0&0&0 \\ 0&0&0&1&1&1&0&0&0 \\ 0&0&0&1&1&1&0&0&0 \\ 0&0&0&1&1&1&0&0&0 \\ 0&0&0&0&0&0&1&1&1 \\ 0&0&0&0&0&0&1&1&1 \\ 0&0&0&0&0&0&1&1&1 \end{bmatrix}$$

$$\mathbf{M}_{\mathrm{C}} = \tfrac{1}{3}\mathbf{J}_3 \otimes \mathbf{I}_3 = \tfrac{1}{3}\begin{bmatrix} 1&0&0&1&0&0&1&0&0 \\ 0&1&0&0&1&0&0&1&0 \\ 0&0&1&0&0&1&0&0&1 \\ 1&0&0&1&0&0&1&0&0 \\ 0&1&0&0&1&0&0&1&0 \\ 0&0&1&0&0&1&0&0&1 \\ 1&0&0&1&0&0&1&0&0 \\ 0&1&0&0&1&0&0&1&0 \\ 0&0&1&0&0&1&0&0&1 \end{bmatrix}$$

$$\mathbf{X}_{\mathrm{T}} = \begin{bmatrix} 1&0&0 \\ 0&1&0 \\ 0&0&1 \\ 0&0&1 \\ 1&0&0 \\ 0&1&0 \\ 0&1&0 \\ 0&0&1 \\ 1&0&0 \end{bmatrix} \text{ and}$$

$$\mathbf{M}_{\mathrm{T}} = \frac{1}{3}\begin{bmatrix} 1&0&0&0&1&0&0&0&1 \\ 0&1&0&0&0&1&1&0&0 \\ 0&0&1&1&0&0&0&1&0 \\ 0&0&1&1&0&0&0&1&0 \\ 1&0&0&0&1&0&0&0&1 \\ 0&1&0&0&0&1&1&0&0 \\ 0&1&0&0&0&1&1&0&0 \\ 0&0&1&1&0&0&0&1&0 \\ 1&0&0&0&1&0&0&0&1 \end{bmatrix}$$

■

b) Alternative expectation models

There are 8 possible different models for the expectation when Rows, Columns and Treatments are considered fixed:

$$\psi_G = \mathbf{X}_G\mu \qquad \text{(no treatment, row or column differences)}$$
$$\psi_R = \mathbf{X}_R\beta \qquad \text{(row differences only)}$$
$$\psi_C = \mathbf{X}_C\delta \qquad \text{(column differences only)}$$
$$\psi_{R+C} = \mathbf{X}_R\beta + \mathbf{X}_C\delta \qquad \text{(additive row and column)}$$
$$\psi_T = \mathbf{X}_T\tau \qquad \text{(treatment differences only)}$$
$$\psi_{R+T} = \mathbf{X}_R\beta + \mathbf{X}_T\tau \qquad \text{(additive row and treatment differences)}$$
$$\psi_{C+T} = \mathbf{X}_C\delta + \mathbf{X}_T\tau \qquad \text{(additive column and treatment differences)}$$
$$\psi_{R+C+T} = \mathbf{X}_R\beta + \mathbf{X}_C\delta + \mathbf{X}_T\tau \qquad \text{(additive row, column and treatment differences)}$$

We note the following marginality relations between the models:

$$\begin{aligned}
\boldsymbol{\psi}_G &\leq \boldsymbol{\psi}_R, \boldsymbol{\psi}_C, \boldsymbol{\psi}_{R+C}, \boldsymbol{\psi}_T, \boldsymbol{\psi}_{R+T}, \boldsymbol{\psi}_{C+T}, \boldsymbol{\psi}_{R+C+T} \\
\boldsymbol{\psi}_R &\leq \boldsymbol{\psi}_{R+C}, \boldsymbol{\psi}_{R+T}, \boldsymbol{\psi}_{R+C+T} \\
\boldsymbol{\psi}_C &\leq \boldsymbol{\psi}_{R+C}, \boldsymbol{\psi}_{C+T}, \boldsymbol{\psi}_{R+C+T} \\
\boldsymbol{\psi}_T &\leq \boldsymbol{\psi}_{R+T}, \boldsymbol{\psi}_{C+T}, \boldsymbol{\psi}_{R+C+T} \\
\boldsymbol{\psi}_{R+C}, \boldsymbol{\psi}_{R+T}, \boldsymbol{\psi}_{C+T} &\leq \boldsymbol{\psi}_{R+C+T}
\end{aligned}.$$

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

$$\begin{aligned}
\hat{\boldsymbol{\psi}}_G &= \overline{\mathbf{G}} \\
\hat{\boldsymbol{\psi}}_R &= \overline{\mathbf{R}} \\
\hat{\boldsymbol{\psi}}_C &= \overline{\mathbf{C}} \\
\hat{\boldsymbol{\psi}}_{R+C} &= \overline{\mathbf{R}} + \overline{\mathbf{C}} - \overline{\mathbf{G}} \\
\hat{\boldsymbol{\psi}}_T &= \overline{\mathbf{T}} \\
\hat{\boldsymbol{\psi}}_{R+T} &= \overline{\mathbf{R}} + \overline{\mathbf{T}} - \overline{\mathbf{G}} \\
\hat{\boldsymbol{\psi}}_{C+T} &= \overline{\mathbf{C}} + \overline{\mathbf{T}} - \overline{\mathbf{G}} \\
\hat{\psi}_{R+C+T} &= \overline{\mathbf{R}} + \overline{\mathbf{C}} + \overline{\mathbf{T}} - 2\overline{\mathbf{G}}
\end{aligned}$$

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

V.C Hypothesis testing using the ANOVA method for a Latin square

An analysis of variance will be used to choose between the eight alternative expectation models for a Latin square.

a) Analysis of an example

Example V.2 Pollution effects of petrol additives (continued)

The data for the car pollution experiment, the reductions in the amount of nitrous oxides produced, are as follows:

4 x 4 Latin square

		Car 1	2	3	4
	I	B	D	C	A
		20	20	17	15
	II	A	B	D	C
		20	27	23	26
Drivers	III	D	C	A	B
		20	25	21	26
	IV	C	A	B	D
		16	16	15	13

(Additives A, B, C, D)

The hypothesis test for the example is:

Step 1: Set up hypotheses

a) H_0: $\tau_A = \tau_B = \tau_C = \tau_D$ (or $\mathbf{X}_A\tau$ not required in model)
H_1: not all population Additives means are equal

b) H_0: $\beta_I = \beta_{II} = \beta_{III} = \beta_{IV}$ (or $\mathbf{X}_D\beta$ not required in model)
H_1: not all population Drivers means are equal

c) H_0: $\delta_1 = \delta_2 = \delta_3 = \delta_4$ (or $\mathbf{X}_C\delta$ not required in model)
H_1: not all population Cars means are equal

Set $\alpha = 0.05$.

Step 2: Calculate test statistics

Source	df	MSq	E[MSq]	F	Prob
Drivers	3	72.00	$\sigma^2 + q_D(\boldsymbol{\psi})$	27.0	<0.001
Cars	3	8.00	$\sigma^2 + q_C(\boldsymbol{\psi})$	3.0	0.117
Drivers#Cars	9				
Additives	3	13.33	$\sigma^2 + q_A(\boldsymbol{\psi})$	5.0	0.045
Residual	6	2.67	σ^2		
Total	15				

Note that Drivers#Cars refers to the "interaction between Drivers and Cars" — it contrasts with Cars{Drivers] or Drivers[Cars] and will be explained in chapter VII. R does not distinguish between these: all are Drivers:Cars.

Step 3: Decide between hypotheses

It would appear that there were differences between drivers but not cars and that there are differences between the additives. The model that best describes the data would appear to be $\psi_{\mathrm{D+A}} = \mathbf{X}_{\mathrm{D}}\beta + \mathbf{X}_{\mathrm{A}}\tau$, the model for additive Drive and Additive effects. ■

b) Sums of squares for the analysis of variance

In this section we will use the generic names of Rows, Columns and Treatments for the factors in a Latin square.

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

Total or Units SSq:	$\mathbf{D}_{\mathrm{G}} = \mathbf{Y} - \overline{\mathbf{G}}$
Rows SSq:	$\mathbf{R}_e = \overline{\mathbf{R}} - \overline{\mathbf{G}}$
Columns SSq:	$\mathbf{C}_e = \overline{\mathbf{C}} - \overline{\mathbf{G}}$
Rows#Columns SSq:	$\mathbf{D}_{\mathrm{R+C}} = \mathbf{Y} - \overline{\mathbf{R}} - \overline{\mathbf{C}} + \overline{\mathbf{G}}$
Treatments SSq:	$\mathbf{T}_e = \overline{\mathbf{T}} - \overline{\mathbf{G}}$
Residual SSq:	$\mathbf{D}_{\mathrm{R+C+T}} = \mathbf{Y} - \overline{\mathbf{R}} - \overline{\mathbf{C}} - \overline{\mathbf{T}} + 2\overline{\mathbf{G}} = \mathbf{Y} - \mathbf{R}_e - \mathbf{C}_e - \mathbf{T}_e - \overline{\mathbf{G}}$

where the **D**s are n-vectors of **d**eviations from **Y** and the vectors with the e subscripts are n-vectors of **e**ffects. From section V.B, *Models and estimation for a Latin square*, we have that

$$\begin{aligned}
\overline{\mathbf{G}} &= \mathbf{M}_{\mathrm{G}}\mathbf{Y} = t^{-2}\left(\mathbf{J}_t \otimes \mathbf{J}_t\right)\mathbf{Y} \\
\overline{\mathbf{R}} &= \mathbf{M}_{\mathrm{R}}\mathbf{Y} = t^{-1}\left(\mathbf{I}_t \otimes \mathbf{J}_t\right)\mathbf{Y} \\
\overline{\mathbf{C}} &= \mathbf{M}_{\mathrm{C}}\mathbf{Y} = t^{-1}\left(\mathbf{J}_t \otimes \mathbf{I}_t\right)\mathbf{Y} \\
\overline{\mathbf{T}} &= \mathbf{M}_{\mathrm{T}}\mathbf{Y}
\end{aligned}$$

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

$$\mathbf{D}'_{\mathrm{G}}\mathbf{D}_{\mathrm{G}} = \left(\mathbf{Y} - \overline{\mathbf{G}}\right)'\left(\mathbf{Y} - \overline{\mathbf{G}}\right) = \mathbf{Y}'\mathbf{Q}_{\mathrm{U}}\mathbf{Y} \qquad \text{with } \mathbf{Q}_{\mathrm{U}} = \mathbf{M}_{\mathrm{U}} - \mathbf{M}_{\mathrm{G}}$$

$$\mathbf{R}'_e\mathbf{R}_e = \left(\overline{\mathbf{R}} - \overline{\mathbf{G}}\right)'\left(\overline{\mathbf{R}} - \overline{\mathbf{G}}\right) = \mathbf{Y}'\mathbf{Q}_{\mathrm{R}}\mathbf{Y} \qquad \text{with } \mathbf{Q}_{\mathrm{R}} = \mathbf{M}_{\mathrm{R}} - \mathbf{M}_{\mathrm{G}}$$

$$\mathbf{C}'_e\mathbf{C}_e = \left(\overline{\mathbf{C}} - \overline{\mathbf{G}}\right)'\left(\overline{\mathbf{C}} - \overline{\mathbf{G}}\right) = \mathbf{Y}'\mathbf{Q}_{\mathrm{C}}\mathbf{Y} \qquad \text{with } \mathbf{Q}_{\mathrm{C}} = \mathbf{M}_{\mathrm{C}} - \mathbf{M}_{\mathrm{G}}$$

$$\mathbf{D}'_{\mathrm{R+C}}\mathbf{D}_{\mathrm{R+C}} = \left(\mathbf{Y} - \overline{\mathbf{R}} - \overline{\mathbf{C}} + \overline{\mathbf{G}}\right)'\left(\mathbf{Y} - \overline{\mathbf{R}} - \overline{\mathbf{C}} + \overline{\mathbf{G}}\right) = \mathbf{Y}'\mathbf{Q}_{\mathrm{RC}}\mathbf{Y}$$

$$\text{with } \mathbf{Q}_{\mathrm{RC}} = \mathbf{M}_{\mathrm{U}} - \mathbf{M}_{\mathrm{R}} - \mathbf{M}_{\mathrm{C}} + \mathbf{M}_{\mathrm{G}}$$

$$\mathbf{T}'_e\mathbf{T}_e = \left(\overline{\mathbf{T}} - \overline{\mathbf{G}}\right)'\left(\overline{\mathbf{T}} - \overline{\mathbf{G}}\right) = \mathbf{Y}'\mathbf{Q}_{\mathrm{T}}\mathbf{Y} \qquad \text{with } \mathbf{Q}_{\mathrm{T}} = \mathbf{M}_{\mathrm{T}} - \mathbf{M}_{\mathrm{G}}$$

$$\mathbf{D}'_{\mathrm{R+C+T}}\mathbf{D}_{\mathrm{R+C+T}} = \left(\mathbf{Y} - \overline{\mathbf{R}} - \overline{\mathbf{C}} - \overline{\mathbf{T}} + 2\overline{\mathbf{G}}\right)'\left(\mathbf{Y} - \overline{\mathbf{R}} - \overline{\mathbf{C}} - \overline{\mathbf{T}} + 2\overline{\mathbf{G}}\right) = \mathbf{Y}'\mathbf{Q}_{\mathrm{RC_{Res}}}\mathbf{Y}$$

$$\text{with } \mathbf{Q}_{\mathrm{RC_{Res}}} = \mathbf{M}_{\mathrm{U}} - \mathbf{M}_{\mathrm{R}} - \mathbf{M}_{\mathrm{C}} - \mathbf{M}_{\mathrm{T}} + 2\mathbf{M}_{\mathrm{G}}$$

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

So the analysis of variance table is constructed as follows:

Source	df	SSq	MSq	F	p
Rows	t–1	$\mathbf{Y}'\mathbf{Q}_{\mathrm{R}}\mathbf{Y}$	$\dfrac{\mathbf{Y}'\mathbf{Q}_{\mathrm{R}}\mathbf{Y}}{t-1}=s_{\mathrm{R}}^2$	$s_{\mathrm{R}}^2 / s_{\mathrm{RC_{Res}}}^2$	p_{R}
Columns	t–1	$\mathbf{Y}'\mathbf{Q}_{\mathrm{C}}\mathbf{Y}$	$\dfrac{\mathbf{Y}'\mathbf{Q}_{\mathrm{C}}\mathbf{Y}}{t-1}=s_{\mathrm{C}}^2$	$s_{\mathrm{C}}^2 / s_{\mathrm{RC_{Res}}}^2$	p_{C}
Rows#Columns	$(t–1)^2$	$\mathbf{Y}'\mathbf{Q}_{\mathrm{RC}}\mathbf{Y}$			
Treatments	t–1	$\mathbf{Y}'\mathbf{Q}_{\mathrm{T}}\mathbf{Y}$	$\dfrac{\mathbf{Y}'\mathbf{Q}_{\mathrm{T}}\mathbf{Y}}{t-1}=s_{\mathrm{T}}^2$	$s_{\mathrm{T}}^2 / s_{\mathrm{RC_{Res}}}^2$	p_{T}
Residual	(t–1)(t–2)	$\mathbf{Y}'\mathbf{Q}_{\mathrm{RC_{Res}}}\mathbf{Y}$	$\dfrac{\mathbf{Y}'\mathbf{Q}_{\mathrm{RC_{Res}}}\mathbf{Y}}{(t-1)(t-2)}=s_{\mathrm{RC_{Res}}}^2$		
Total	t^2–1	$\mathbf{Y}'\mathbf{Q}_{\mathrm{U}}\mathbf{Y}$			

There is, as usual, a geometric interpretation of this analysis. The t^2–1 dimensional part of the t^2-dimensional data space that is orthogonal to equiangular line is partitioned into the t–1 dimensional Rows and Columns subspaces and the $(t–1)^2$ dimensional Rows#Columns subspace. These three subspaces are mutually orthogonal. Finally the Rows#Columns subspace is subdivided into two orthogonal subspaces: the t–1 dimensional Treatments subspace and the (t–1)(t–2) dimensional Residual subspace.

Example V.2 Pollution effects of petrol additives (continued)

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

Addit-ive	Reduct NO $\mathbf{y}$	Total Driver∧Car deviations $\mathbf{d}_{\mathrm{G}}$	Driver Effects $\mathbf{r}_e$	Car Effects $\mathbf{c}_e$	Driver#Car deviations $\mathbf{d}_{\mathrm{R+C}}$	Additive effects $\mathbf{t}_e$	Residual Driver#Car deviations $\mathbf{d}_{\mathrm{R+C+T}}$
B	20	0	-2	-1	3	2	1
D	20	0	-2	2	0	-1	1
C	17	-3	-2	-1	0	1	-1
A	15	-5	-2	0	-3	-2	-1
A	20	0	4	-1	-3	-2	-1
B	27	7	4	2	1	2	-1
D	23	3	4	-1	0	-1	1
C	26	6	4	0	2	1	1
D	20	0	3	-1	-2	-1	-1
C	25	5	3	2	0	1	-1
A	21	1	3	-1	-1	-2	1
B	26	6	3	0	3	2	1
C	16	-4	-5	-1	2	1	1
A	16	-4	-5	2	-1	-2	1
B	15	-5	-5	-1	1	2	-1
D	13	-7	-5	0	-2	-1	-1
SSq		296	216	24	56	40	16

That is, the Units SSq is $\mathbf{Y}'\mathbf{Q}_{\mathrm{U}}\mathbf{Y}=296$, the Driver SSq is $\mathbf{Y}'\mathbf{Q}_{\mathrm{D}}\mathbf{Y}=216$, the Car SSq is $\mathbf{Y}'\mathbf{Q}_{\mathrm{C}}\mathbf{Y}=24$, the Driver#Car SSq is $\mathbf{Y}'\mathbf{Q}_{\mathrm{DC}}\mathbf{Y}=56$, the Additives SSq is $\mathbf{Y}'\mathbf{Q}_{\mathrm{A}}\mathbf{Y}=40$ and the Residual SSq is $\mathbf{Y}'\mathbf{Q}_{\mathrm{RC}_{\mathrm{Res}}}\mathbf{Y}=16$.

Note that $\mathbf{y}=\overline{\mathbf{g}}+\mathbf{r}_{\mathrm{e}}+\mathbf{c}_{\mathrm{e}}+\mathbf{t}_{\mathrm{e}}+\mathbf{d}_{\mathrm{R+C+T}}$.

c) Expected mean squares

To justify our choice of test statistics, we want to work out the expected values of the mean squares in the analysis of variance table under the eight alternative expectation models. However, to save space we will work out the expected means squares under the maximal model and identify which terms in the expected mean squares go to zero under alternative models.

The expected means squares when both Rows and Columns effects are fixed are given in the following table:

Source	df	MSq	E[MSq]	F
Rows	t–1	$\dfrac{\mathbf{Y}'\mathbf{Q}_{\mathrm{R}}\mathbf{Y}}{t-1}=s_{\mathrm{R}}^2$	$\sigma^2+q_{\mathrm{R}}(\psi)$	$s_{\mathrm{R}}^2\big/s_{\mathrm{RC}_{\mathrm{Res}}}^2$
Columns	t–1	$\dfrac{\mathbf{Y}'\mathbf{Q}_{\mathrm{C}}\mathbf{Y}}{t-1}=s_{\mathrm{C}}^2$	$\sigma^2+q_{\mathrm{C}}(\psi)$	$s_{\mathrm{C}}^2\big/s_{\mathrm{RC}_{\mathrm{Res}}}^2$
Rows#Columns	$(t–1)^2$			
Treatments	t–1	$\dfrac{\mathbf{Y}'\mathbf{Q}_{\mathrm{T}}\mathbf{Y}}{t-1}=s_{\mathrm{T}}^2$	$\sigma^2+q_{\mathrm{T}}(\psi)$	$s_{\mathrm{T}}^2\big/s_{\mathrm{RC}_{\mathrm{Res}}}^2$
Residual	(t–1)(t–2)	$\dfrac{\mathbf{Y}'\mathbf{Q}_{\mathrm{RC}_{\mathrm{Res}}}\mathbf{Y}}{(t-1)(t-2)}=s_{\mathrm{RC}_{\mathrm{Res}}}^2$	σ^2	
Total	t^2–1			

$$q_{\mathrm{R}}(\psi)=\frac{\psi'\mathbf{Q}_{\mathrm{R}}\psi}{t-1}=\sum_{i=1}^{t}t\left(\beta_i-\bar{\beta}_{.}\right)^2\Big/(t-1),\ q_{\mathrm{C}}(\psi)=\frac{\psi'\mathbf{Q}_{\mathrm{C}}\psi}{t-1}=\sum_{j=1}^{t}t\left(\delta_j-\bar{\delta}_{.}\right)^2\Big/(t-1)$$

and $q_{\mathrm{T}}(\psi)=\dfrac{\psi'\mathbf{Q}_{\mathrm{T}}\psi}{t-1}=\sum_{k=1}^{t}t\left(\tau_k-\bar{\tau}_{.}\right)^2\Big/(t-1)$

Given the expressions in the above table, the population means of the mean squares could be computed if knew the β_is, δ_js, τ_ks and σ^2. Each of the terms $q_{\mathrm{R}}(\psi)$, $q_{\mathrm{C}}(\psi)$ and $q_{\mathrm{T}}(\psi)$ will equal zero when the terms $\mathbf{X}_{\mathrm{R}}\boldsymbol{\beta}$, $\mathbf{X}_{\mathrm{C}}\boldsymbol{\delta}$ and $\mathbf{X}_{\mathrm{T}}\boldsymbol{\tau}$, respectively, are removed from the model. Hence a significant F value for a line indicates that the corresponding term should be included in the model.

An alternative analysis would be to designate that both Rows and Columns are random. The model in this case would be that

$$\begin{aligned}\psi_{\mathrm{T}} &= E[\mathbf{Y}] = \mathbf{X}_{\mathrm{T}}\boldsymbol{\tau} \text{ and}\\ \mathbf{V} &= \sigma^2\mathbf{M}_{\mathrm{U}} + t\sigma_{\mathrm{R}}^2\mathbf{M}_{\mathrm{R}} + t\sigma_C^2\mathbf{M}_{\mathrm{C}}\\ &= \sigma^2\mathbf{I}_t\otimes\mathbf{I}_t + \sigma_{\mathrm{R}}^2 t\times t^{-1}\mathbf{I}_t\otimes\mathbf{J}_t + \sigma_C^2 t\times t^{-1}\mathbf{J}_t\otimes\mathbf{I}_t,\\ &= \sigma^2\mathbf{I}_t\otimes\mathbf{I}_t + \sigma_{\mathrm{R}}^2\mathbf{I}_t\otimes\mathbf{J}_t + \sigma_C^2\mathbf{J}_t\otimes\mathbf{I}_t\end{aligned}$$

This model allows for equal covariance between units from the same row and also between different units from the same column.

Example V.3 A 3×3 Latin square (continued)

For this example,

$$\psi_{\mathrm{T}} = E[\mathbf{Y}] = \mathbf{X}_{\mathrm{T}}\boldsymbol{\tau} = \begin{bmatrix}1&0&0\\0&1&0\\0&0&1\\0&0&1\\1&0&0\\0&1&0\\0&1&0\\0&0&1\\1&0&0\end{bmatrix}\begin{bmatrix}\tau_1\\\tau_2\\\tau_3\end{bmatrix}$$

$$\begin{aligned}\mathbf{V} &= \sigma^2\mathbf{M}_{\mathrm{U}} + t\sigma_{\mathrm{R}}^2\mathbf{M}_{\mathrm{R}} + t\sigma_C^2\mathbf{M}_{\mathrm{C}}\\ &= \sigma^2\mathbf{I}_t\otimes\mathbf{I}_t + \sigma_{\mathrm{R}}^2\mathbf{I}_t\otimes\mathbf{J}_t + \sigma_C^2\mathbf{J}_t\otimes\mathbf{I}_t\\ &= \sigma^2\begin{bmatrix}1&0&0&0&0&0&0&0&0\\0&1&0&0&0&0&0&0&0\\0&0&1&0&0&0&0&0&0\\0&0&0&1&0&0&0&0&0\\0&0&0&0&1&0&0&0&0\\0&0&0&0&0&1&0&0&0\\0&0&0&0&0&0&1&0&0\\0&0&0&0&0&0&0&1&0\\0&0&0&0&0&0&0&0&1\end{bmatrix} + \sigma_{\mathrm{R}}^2\begin{bmatrix}1&1&1&0&0&0&0&0&0\\1&1&1&0&0&0&0&0&0\\1&1&1&0&0&0&0&0&0\\0&0&0&1&1&1&0&0&0\\0&0&0&1&1&1&0&0&0\\0&0&0&1&1&1&0&0&0\\0&0&0&0&0&0&1&1&1\\0&0&0&0&0&0&1&1&1\\0&0&0&0&0&0&1&1&1\end{bmatrix}\\ &\quad + \sigma_C^2\begin{bmatrix}1&0&0&1&0&0&1&0&0\\0&1&0&0&1&0&0&1&0\\0&0&1&0&0&1&0&0&1\\1&0&0&1&0&0&1&0&0\\0&1&0&0&1&0&0&1&0\\0&0&1&0&0&1&0&0&1\\1&0&0&1&0&0&1&0&0\\0&1&0&0&1&0&0&1&0\\0&0&1&0&0&1&0&0&1\end{bmatrix}\end{aligned}$$

This shows that, under this model, $\mathrm{var}[Y_{ijk}] = \sigma^2 + \sigma_{\mathrm{R}}^2 + \sigma_{\mathrm{C}}^2$, $\mathrm{cov}[Y_{ijk}, Y_{ij'k}] = \sigma_{\mathrm{R}}^2$, $\mathrm{cov}[Y_{ijk}, Y_{i'jk}] = \sigma_{\mathrm{C}}^2$ and $\mathrm{cov}[Y_{ijk}, Y_{i'j'k}] = 0$ for $i \neq i'$ and $j \neq j'$. ■

The expected means squares under this model, that is when both Rows and Columns effects are random, are given in the following table:

Source	df	MSq	E[MSq]	F
Rows	t–1	$\dfrac{\mathbf{Y}'\mathbf{Q}_{\mathrm{R}}\mathbf{Y}}{t-1}=s_{\mathrm{R}}^2$	$\sigma^2+t\sigma_{\mathrm{R}}^2$	$s_{\mathrm{R}}^2 / s_{\mathrm{RC_{Res}}}^2$
Columns	t–1	$\dfrac{\mathbf{Y}'\mathbf{Q}_{\mathrm{C}}\mathbf{Y}}{t-1}=s_{\mathrm{C}}^2$	$\sigma^2+t\sigma_{\mathrm{C}}^2$	$s_{\mathrm{C}}^2 / s_{\mathrm{RC_{Res}}}^2$
Rows#Columns	$(t–1)^2$			
Treatments	t–1	$\dfrac{\mathbf{Y}'\mathbf{Q}_{\mathrm{T}}\mathbf{Y}}{t-1}=s_{\mathrm{T}}^2$	$\sigma^2+q_{\mathrm{T}}(\psi)$	$s_{\mathrm{T}}^2 / s_{\mathrm{RC_{Res}}}^2$
Residual	(t–1)(t–2)	$\dfrac{\mathbf{Y}'\mathbf{Q}_{\mathrm{RC_{Res}}}\mathbf{Y}}{(t-1)(t-2)}=s_{\mathrm{RC_{Res}}}^2$	σ^2	
Total	t^2–1			

The alternative expectation model is $\psi_{\mathrm{G}}=\mathbf{X}_{\mathrm{G}}\mu$ and under this model $q_{\mathrm{T}}(\psi)=0$. Alternative variance models involve setting $\sigma_{\mathrm{R}}^2=0$ and/or $\sigma_{\mathrm{C}}^2=0$ and this will result in the one(s) set to zero being dropped from the expected mean square. This exactly parallels what happens when both are fixed.

d) Summary of the hypothesis test

We summarize the ANOVA-based hypothesis test for a Latin square involving t treatments with a total of $n=t^2$ observed units and with Rows, Columns and Treatments fixed.

Step 1: Set up hypotheses

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

b) H_0: $\beta_1 = \beta_2 = ... = \beta_t$ (or $\mathbf{X}_{\mathrm{R}}\boldsymbol{\beta}$ not required in model)
H_1: not all population row means are equal

c) H_0: $\gamma_1 = \gamma_2 = ... = \gamma_t$ (or $\mathbf{X}_{\mathrm{C}}\boldsymbol{\gamma}$ not required in model)
H_1: not all population columns means are equal

Set α.

Step 2: Calculate test statistics

The analysis of variance table for a Latin square is:

Source	df	MSq	E[MSq]	F	p
Rows	t–1	$\dfrac{\mathbf{Y}'\mathbf{Q}_{\mathrm{R}}\mathbf{Y}}{t-1}=s_{\mathrm{R}}^2$	$\sigma^2+q_{\mathrm{R}}(\psi)$	$s_{\mathrm{R}}^2 / s_{\mathrm{RC_{Res}}}^2$	p_{R}
Columns	t–1	$\dfrac{\mathbf{Y}'\mathbf{Q}_{\mathrm{C}}\mathbf{Y}}{t-1}=s_{\mathrm{C}}^2$	$\sigma^2+q_{\mathrm{C}}(\psi)$	$s_{\mathrm{C}}^2 / s_{\mathrm{RC_{Res}}}^2$	p_{C}
Rows#Columns	(t–1)^2				
Treatments	t–1	$\dfrac{\mathbf{Y}'\mathbf{Q}_{\mathrm{T}}\mathbf{Y}}{t-1}=s_{\mathrm{T}}^2$	$\sigma^2+q_{\mathrm{T}}(\psi)$	$s_{\mathrm{T}}^2 / s_{\mathrm{RC_{Res}}}^2$	p_{T}
Residual	(t–1)(t–2)	$\dfrac{\mathbf{Y}'\mathbf{Q}_{\mathrm{RC_{Res}}}\mathbf{Y}}{(t-1)(t-2)}=s_{\mathrm{RC_{Res}}}^2$	σ^2		
Total	t^2–1				

Step 3: Decide between hypotheses

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

e) Comparison with traditional Latin-square ANOVA table

Again, the above analysis of variance table and the traditional Latin-square 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 four sources that are labelled differently but the tables differ in that our table includes the line Rows#Columns. Rows#Columns reflects the variation of column differences from row to row or the differences in row-column combinations after overall row and overall column differences have been removed — it is uncontrolled variation in the units. The Rows#Columns sum of squares is partitioned into Treatments and Residual sums of squares.

Source	df	Source in Latin-square ANOVA
Rows	t–1	Between Rows
Columns	t–1	Between Columns
Rows#Columns	(t–1)^2	
Treatments	t–1	Between Treatments
Residual	(t–1)(t–2)	Error
Total	t^2–1	Total

Again, the advantage of the table we have presented is that it exhibits the confounding in the experiment. The indenting of Treatments under Rows#Columns signifies that treatment differences are confounded or “mixed-up” with row-column differences, adjusted for overall row and overall column differences, as a result of the randomization of treatments to row-column combinations. This is not obvious from the traditional table.

f) Computation of ANOVA and diagnostic checking in R

Diagnostic checking is the same as for the RCBD, so it is included here with the ANOVA.

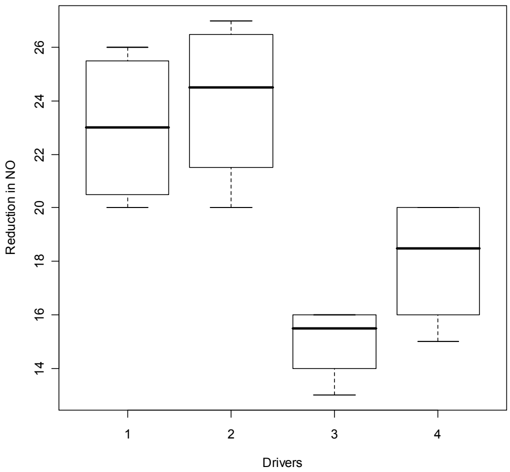
First the `data.frame` that contains the factors and response variable needs to be set up and attached. Then, you obtain initial boxplots and, as with the other designs, use the `aov` function, either with or without the `Error` as part of the model. In this experiment the uncontrolled variation is made up of Drivers, Cars and Drivers#Cars. R provides a shorthand for this: Drivers*Cars is a shorthand that expands to Drivers + Cars + Drivers:Cars, the latter being equivalent to Drivers#Cars.

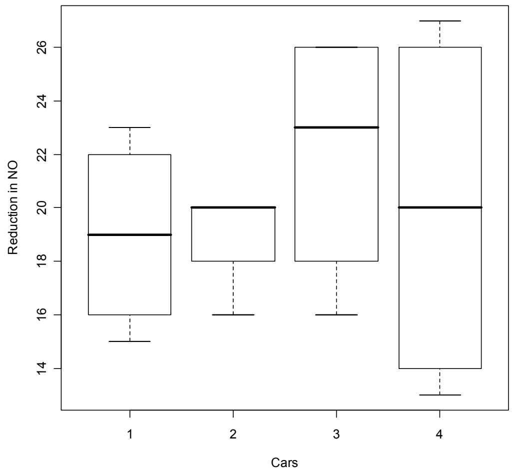
The output, including the expressions for the analysis with `Error` and doing the diagnostic checking for this data, is given below.

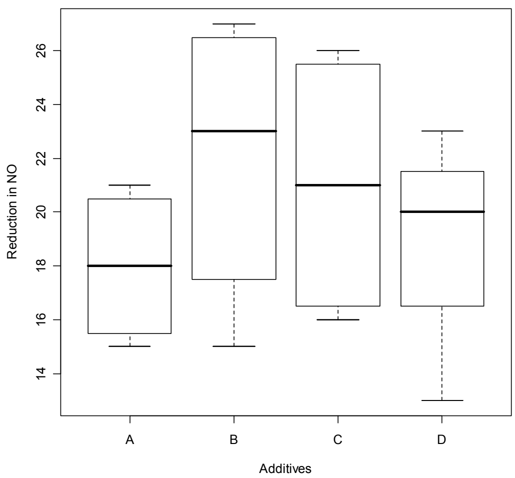
```
> load("LSPolut.dat.rda")
> attach(LSPolut.dat)
> boxplot(split(Reduct.NO, Drivers), xlab="Drivers", ylab="Reduction in NO")
> boxplot(split(Reduct.NO, Cars), xlab="Cars", ylab="Reduction in NO")
> boxplot(split(Reduct.NO, Additives), xlab="Additives", ylab="Reduction in NO")
> LSPolut.aov <- aov(Reduct.NO ~ Drivers + Cars + Additives + Error(Drivers*Cars),
+                                                                 LSPolut.dat)
> summary(LSPolut.aov)

Error: Drivers
        Df Sum Sq Mean Sq
Drivers  3    216      72

Error: Cars
     Df Sum Sq Mean Sq
Cars  3     24       8

Error: Drivers:Cars
          Df Sum Sq Mean Sq F value Pr(>F)
Additives  3 40.000  13.333       5 0.0452
Residuals  6 16.000   2.667
> #Compute Drivers and Cars Fs and p-values
> Drivers.F <- 72/2.667
> Drivers.p <- 1-pf(Drivers.F, 3, 6)
> Cars.F <- 8/2.667
> Cars.p <- 1-pf(Cars.F, 3, 6)
> data.frame(Drivers.F,Drivers.p,Cars.F,Cars.p)
  Drivers.F    Drivers.p   Cars.F    Cars.p
1  26.99663 0.0006989578 2.999625 0.1169842
> #
> # Diagnostic checking
> #
> res <- resid.errors(LSPolut.aov)
> fit <- fitted.errors(LSPolut.aov)
> data.frame(Drivers,Cars,Additives,Reduct.NO,res,fit)
   Drivers Cars Additives Reduct.NO res fit
1        1    1         B        20   1  19
2        1    2         D        20   1  19
3        1    3         C        17  -1  18
4        1    4         A        15  -1  16
5        2    1         A        20  -1  21
6        2    2         B        27  -1  28
7        2    3         D        23   1  22
```

```
8         2    4         C         26   1  25
9         3    1         D         20  -1  21
10        3    2         C         25  -1  26
11        3    3         A         21   1  20
12        3    4         B         26   1  25
13        4    1         C         16   1  15
14        4    2         A         16   1  15
15        4    3         B         15  -1  16
16        4    4         D         13  -1  14
> plot(fit, res, pch=16)
> qqnorm(res, pch=16)
> qqline(res)
> tukey.1df(LSPolut.aov, LSPolut.dat, error.term = "Drivers:Cars")
$Tukey.SS
[1] 4.54224

$Tukey.F
[1] 1.982167

$Tukey.p
[1] 0.2181923

$Devn.SS
[1] 11.45776
```

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







There are marked differences between the drivers, but not the cars. It is not clear whether there are differences between the additives. Perhaps there is variance heterogeneity.

The hypothesis test for the example is:

Step 1: Set up hypotheses

a) H_0: $\tau_A = \tau_B = \tau_C = \tau_D$ (or $\mathbf{X}_A\tau$ not required in model)
H_1: not all population Additives means are equal

b) H_0: $\beta_I = \beta_{II} = \beta_{III} = \beta_{IV}$ (or $\mathbf{X}_D\beta$ not required in model)
H_1: not all population Drivers means are equal

c) H_0: $\delta_1 = \delta_2 = \delta_3 = \delta_4$ (or $\mathbf{X}_C\delta$ not required in model)
H_1: not all population Cars means are equal

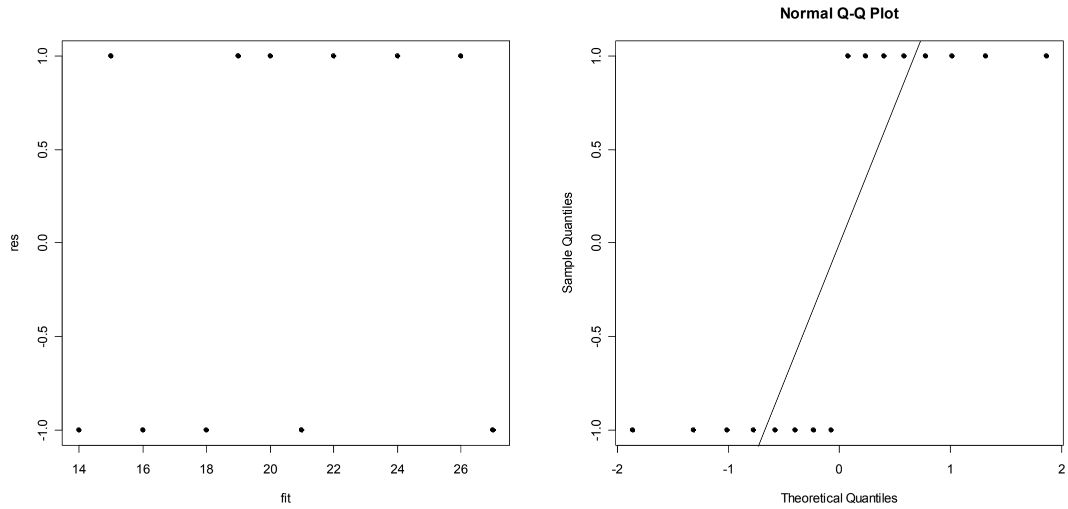
Set $\alpha = 0.05$.

Step 2: Calculate test statistics

Source	df	MSq	E[MSq]	F	Prob
Drivers	3	72.00	$\sigma^2 + q_D(\boldsymbol{\psi})$	27.0	<0.001
Cars	3	8.00	$\sigma^2 + q_C(\boldsymbol{\psi})$	3.0	0.117
Drivers#Cars	9				
Additives	3	13.33	$\sigma^2 + q_A(\boldsymbol{\psi})$	5.0	0.045
Residual	6	2.67	σ^2		
Nonadditivity	1	4.54		1.98	0.218
Deviation	5	2.29			
Total	15				

Step 3: Decide between hypotheses

It would appear that there were differences between drivers but not cars and that there are differences between the additives. The model that best describes the data would appear to be $\psi_{D+A} = \mathbf{X}_D\beta + \mathbf{X}_A\tau$, the model for additive Drive and Additive effects. The test for transformable nonadditivity is nonsignificant; the residuals-versus-fitted-values plot below indicates that the residuals are either –1 or 1 which is a reflection of the artificial nature of the data; further, the Normal Probability Plot below indicates that the data are not normal. Clearly, the example can only considered to be illustrative of the techniques.



V.D Diagnostic checking

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

a) the response is operating additively, that is, that a treatment has about the same additive effect on each unit;
b) that the variability of the units is the same for all row-column combinations;
c) each observation displays the covariance implied by the model (independence for Rows and Columns fixed; equal correlation within rows (columns) for Rows (Columns) random); and
d) that the response of the units is normally distributed.

The need for an additive response comes from the maximal model for this case which is additive in Rows, Columns and Treatment parameters:

$$\psi_{\mathrm{R+C+T}} = E\left[\mathbf{Y}\right] = \mathbf{X}_{\mathrm{R}}\beta + \mathbf{X}_{\mathrm{C}}\delta + \mathbf{X}_{\mathrm{T}}\tau$$

The fitted values, obtained using the estimator of the expected values $\hat{\psi}_{\mathrm{R+C+T}} = \overline{\mathbf{R}} + \overline{\mathbf{C}} + \overline{\mathbf{T}} - 2\overline{\mathbf{G}}$, will display the additive pattern specified by this model and we hope these are an adequate description of the data.

As noted in the previous section, the diagnostic checking is the same as for the RCBD. It is based on the residual-versus-fitted-values and normal probability plots, as well as Tukey's one-degree-of-freedom-for-nonadditivity.

V.E Treatment differences

For the purposes of the scientist the effects of rows and columns 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 and RCBD applies here also.

Example V.2 Pollution effects of petrol additives (continued)

In this example the Additives were significantly different and so we use the Tukey's HSD procedure to investigate the Additives differences

```
> #
> # multiple comparisons
> #
> model.tables(LSPolut.aov, type="means")
Tables of means
Grand mean

20

 Drivers
Drivers
 1  2  3  4
23 24 15 18

 Cars
Cars
 1  2  3  4
19 19 22 20

 Additives
Additives
 A  B  C  D
18 22 21 19
> q <- qtukey(0.95, 4, 6)
> q
[1] 4.895599
```

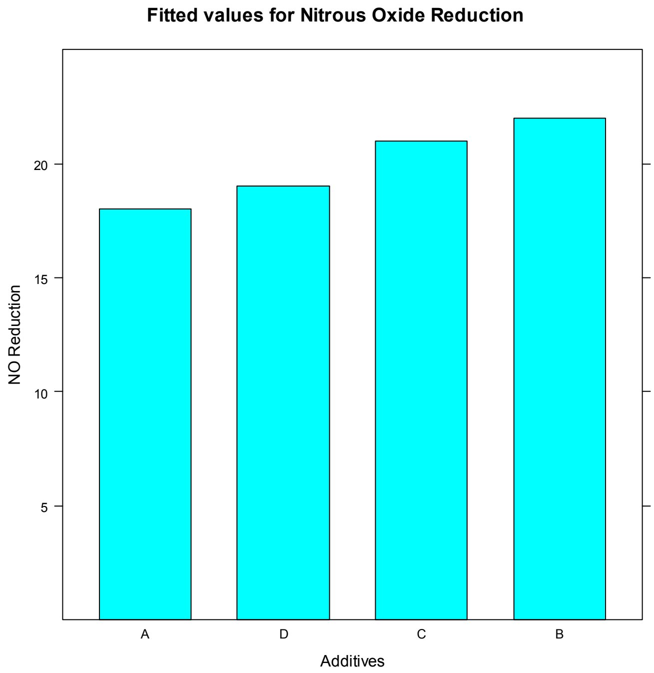
The critical value for Tukey's HSD procedure is as follows:

$$w(5\%) = \frac{4.895599}{\sqrt{2}} \times \sqrt{\frac{2.667 \times 2}{4}} = \frac{4.895599}{\sqrt{2}} \times 1.1548 = 4.00$$

Comparing the differences between the Additives means in the above output with Tukey;s HSD, it is concluded that only the difference between A and B are significant. We produce a bar chart to illustrate the differences.

```
> #
> # Plotting Treat means
> #
> LSPolut.tab <- model.tables(LSPolut.aov, type="means")
> LSPolut.Adds.Mean <- data.frame(Adds.lev = levels(Additives),
+                          Adds.Mean = as.vector(LSPolut.tab$tables$Additives))
> LSPolut.Adds.Mean <- LSPolut.Adds.Mean[order(LSPolut.Adds.Mean$Adds.Mean),]
> #use factor to order bars
> LSPolut.Adds.Mean$Adds.lev <-factor(LSPolut.Adds.Mean$Adds.lev,
+                                    levels=LSPolut.Adds.Mean$Adds.lev)
> barchart(Adds.Mean ~ Adds.lev, xlab="Additives", ylim=c(0,25),
+         ylab="NO Reduction", main="Fitted values for Nitrous Oxide Reduction",
```

```
+          data=LSPolut.Adds.Mean)
```



V.F Design of sets of Latin squares

(Mead sec.8.1)

A problem that occurs with Latin squares is that, because two sources of variation are isolated, the residual degrees of freedom are often small (< 10); thus, a reliable estimate of the uncontrolled variation is not obtained. To overcome this problem several squares can be used. However, there are several ways in which the squares can be repeated. In the case of Example V.2, *Pollution effects of petrol additives*, the Latin square could be repeated using:

1. using the same drivers and cars in each replicate;
2. using the same drivers but new cars (or the same cars but new drivers); or
3. using new cars and drivers.

In general, one can have as many squares as one likes. However, for space reasons only layouts for 2 squares will be presented. The general expressions for randomizing the various cases are given in Appendix B, *Randomized layouts and sample size computations in R*.

Case 1 — same Drivers and Cars

This case involves a complete repetition of the experiment, say on consecutive mornings with the same 4 Drivers and 4 Cars on the two occasions. There is no re-randomization of the square for the second morning — this is necessary if the crossed relationship between Occasions and the other factors is to be preserved.

Layout (r=2)

	Occasions							
	1				2			
Cars	1	2	3	4	1	2	3	4
Drivers								
1	A	B	C	D	A	B	C	D
2	C	D	A	B	C	D	A	B
3	D	C	B	A	D	C	B	A
4	B	A	D	C	B	A	D	C

Case 2 — same Cars different Drivers

In this case the experiment is repeated on a different occasion with same 4 cars on both occasions, but with the drivers on one occasion unconnected with those on the other. As a result the rows of the square, but not the columns, are rerandomized on the second occasion.

Layout (r=2)

	Occasions							
	1				2			
Cars	1	2	3	4	1	2	3	4
Drivers								
1	C	A	B	D	D	B	A	C
2	A	C	D	B	A	C	D	B
3	B	D	C	A	C	A	B	D
4	D	B	A	C	B	D	C	A

Note that the order in which the additives are tested by the second driver on occasion 1 is the same as for the fourth driver on occasion 2; that is, the second row of the square on occasion 1 is the same as the fourth row on occasion 2.

Case 3 — different Drivers and Cars

In this case, not only are the drivers on different occasions unconnected, but so are the cars as the cars used on the second occasion are completely different to those used on the first occasion. As a result the rows **and** columns of the square are rerandomized on the second occasion.

Layout (r=2)

	Occasions							
	1				2			
Cars	1	2	3	4	1	2	3	4
Drivers								
1	B	A	C	D	D	B	C	A
2	C	D	B	A	A	C	B	D
3	A	B	D	C	B	D	A	C
4	D	C	A	B	C	A	D	B

V.G Hypothesis tests for sets of Latin squares

(Mead sec.8.1)

We shall determine the analysis of variance for each of the three cases presented in the previous section, using the techniques discussed in chapter VI, *Determining the analysis of variance table*. I will give the degrees of freedom for the general case of r squares, even though layouts were presented for only two squares. In determining the expected mean squares it will be assumed that the unrandomized factors are to be classified as random factors and the randomized factors as fixed factors. You will find that these analyses differ in form to those given in Appendix 8B of Box, Hunter and Hunter. It is important because it affects the composition of the Residual and therefore potentially may alter conclusions made about treatment differences.

a) Case 1 — same Drivers and Cars

Layout (r=2)

	Occasions							
	1				2			
Cars	1	2	3	4	1	2	3	4
Drivers								
1	A	B	C	D	A	B	C	D
2	C	D	A	B	C	D	A	B
3	D	C	B	A	D	C	B	A
4	B	A	D	C	B	A	D	C

A. Description of pertinent features of the study

1. Observational unit – a car with a driver on a occasion
2. Response variable – Reduction
3. Unrandomized factors – Occasions, Drivers, Cars
4. Randomized factors – Additives
5. Type of study – Sets of Latin squares

B. The experimental structure

Structure	Formula
unrandomized	*2* Occasions**4* Drivers**4* Cars
randomized	*4* Additives

For this structure to be appropriate requires that the same square without re-randomization be used for each occasion; otherwise, some factors would be nested (because randomization would be within Occasions).

C. Sources derived from the structure formulae

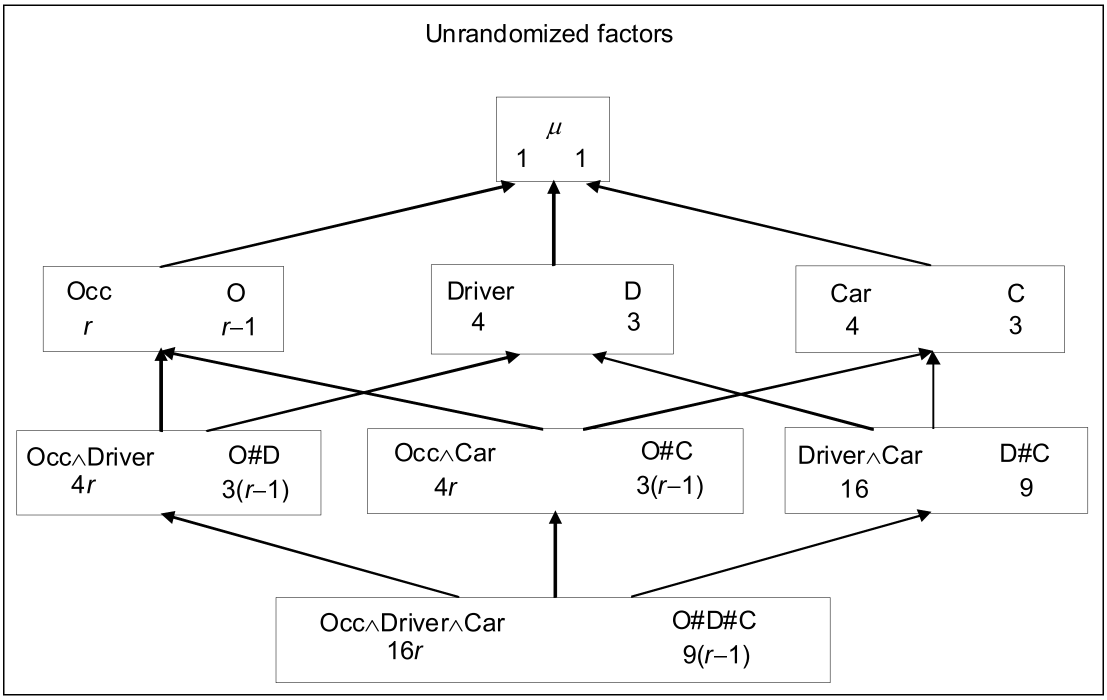
Occasions*Drivers*Cars = (Occasions+Drivers+Occasions#Drivers)*Cars

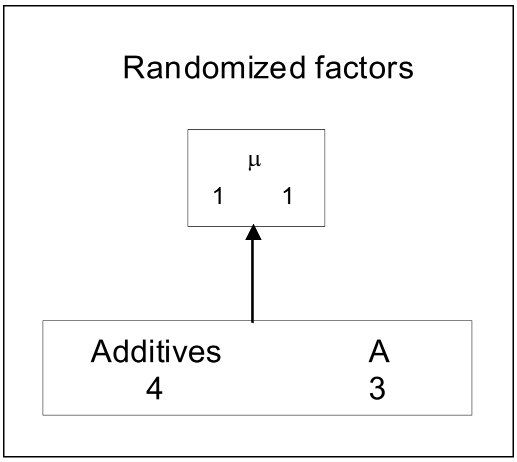
= Occasions + Drivers + Occasions#Drivers
+ Cars + Occasions#Cars + Drivers#Cars
+Occasions#Drivers#Cars

Additives = Additives

D. Degrees of freedom and sums of squares

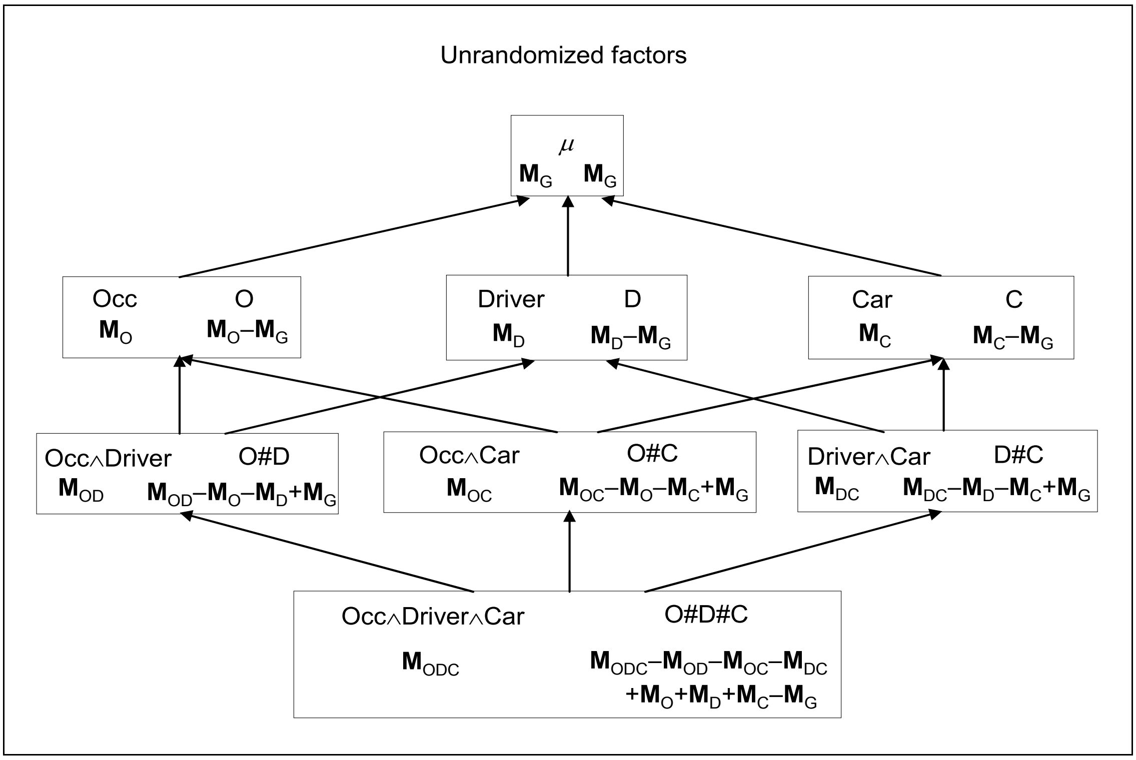
The Hasse diagrams, with degrees of freedom, for this study are:

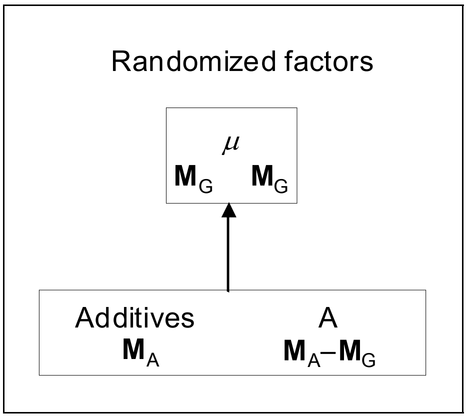




Alternatively, as all the factors in the unrandomized structure are crossed, the rule for a set of crossed factors can be used. That is, the degrees of freedom of any source can be calculated by taking the number of levels minus one for each factor in the source and multiplying these together. For example, since Occasions has r levels and Drivers has 4 levels, the degrees of freedom of Occasions#Drivers is $(r-1)(4-1) = 3(r-1)$.

The Hasse diagrams, with **M** and **Q** matrices, for this study are:





E. The analysis of variance table

Source	df	SSq
Occasions	$r-1$	$\mathbf{Y}'\mathbf{Q}_{\mathrm{O}}\mathbf{Y}$
Drivers	3	$\mathbf{Y}'\mathbf{Q}_{\mathrm{D}}\mathbf{Y}$
Cars	3	$\mathbf{Y}'\mathbf{Q}_{\mathrm{C}}\mathbf{Y}$
Occasions#Drivers	$3(r-1)$	$\mathbf{Y}'\mathbf{Q}_{\mathrm{OD}}\mathbf{Y}$
Occasions#Cars	$3(r-1)$	$\mathbf{Y}'\mathbf{Q}_{\mathrm{OC}}\mathbf{Y}$
Drivers#Cars	9	$\mathbf{Y}'\mathbf{Q}_{\mathrm{DC}}\mathbf{Y}$
Additives	3	$\mathbf{Y}'\mathbf{Q}_{\mathrm{A}}\mathbf{Y}$
Residual	6	$\mathbf{Y}'\mathbf{Q}_{\mathrm{DC_{Res}}}\mathbf{Y}$
Occasions#Drivers#Cars	$9(r-1)$	$\mathbf{Y}'\mathbf{Q}_{\mathrm{ODC}}\mathbf{Y}$
Total	$16r-1$	

F. Maximal expectation and variation models

Assume the randomized factor is a fixed factor and that all the unrandomized factors are random factors. Then the expectation term is Additives. The variation terms are: Occasions, Drivers, Occasions∧Drivers, Cars, Occasions∧Cars, Drivers∧Cars and Occasions∧Drivers∧Cars.

The expectation model is

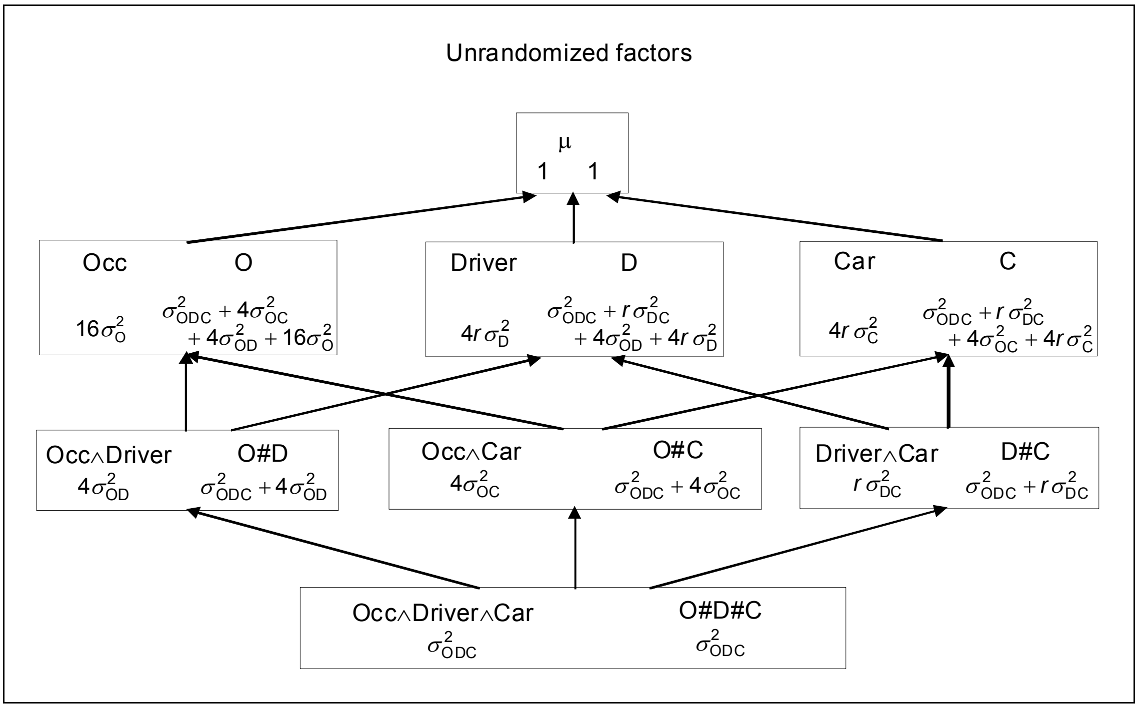
$$\mathrm{E}[\mathbf{Y}] = \text{Additives}$$

and the variation model is

$$\begin{aligned}\mathrm{var}[\mathbf{Y}] = &\ \text{Occasions} + \text{Drivers} + \text{Occasions}\wedge\text{Drivers} + \text{Cars} + \text{Occasions}\wedge\text{Cars} \\ &+ \text{Drivers}\wedge\text{Cars} + \text{Occasions}\wedge\text{Drivers}\wedge\text{Cars}\end{aligned}$$

G. The expected mean squares.

The Hasse diagram, with contributions to expected mean squares, for the unrandomized factors in this study is:



The single randomized factor Additive will contribute $q_{A}(\boldsymbol{\psi})$ to the expected mean square for its source.

Source	df	SSq	E[MSq]				
Occ	r–1	$\mathbf{Y'Q_{O}Y}$	σ^2_{ODC}		$+4\sigma^2_{OC}$	$+4\sigma^2_{OD}$	$+16\sigma^2_{O}$
Drivers	3	$\mathbf{Y'Q_{D}Y}$	σ^2_{ODC}	$+r\sigma^2_{DC}$		$+4\sigma^2_{OD}$	$+4r\sigma^2_{D}$
Cars	3	$\mathbf{Y'Q_{C}Y}$	σ^2_{ODC}	$+r\sigma^2_{DC}$	$+4\sigma^2_{OC}$		$+4r\sigma^2_{C}$
Occ#Drivers	3(r–1)	$\mathbf{Y'Q_{OD}Y}$	σ^2_{ODC}			$+4\sigma^2_{OD}$	
Occ#Cars	3(r–1)	$\mathbf{Y'Q_{OC}Y}$	σ^2_{ODC}		$+4\sigma^2_{OC}$		
Drivers#Cars	9	$\mathbf{Y'Q_{DC}Y}$					
Additive	3	$\mathbf{Y'Q_{A}Y}$	σ^2_{ODC}	$+r\sigma^2_{DC}$			$+q_{A}(\boldsymbol{\psi})$
Residual	6	$\mathbf{Y'Q_{DC_{Res}}Y}$	σ^2_{ODC}	$+r\sigma^2_{DC}$			
Occ#Drivers#Cars	9(r–1)	$\mathbf{Y'Q_{ODC}Y}$	σ^2_{ODC}				
Total	16r–1						

Hypothesis tests for Additive, Occ#Drivers and Occ#Cars are straightforward: Additive is compared with the Residual for Drivers#Cars and Occ#Drivers and Occ#Cars with Occ#Drivers#Cars. Tests for Occ, Drivers and Cars are more difficult because there is no single mean square that has the same expectations as one of these, except for the single component to be tested. For example, to test $\sigma_{\mathrm{O}}^2 = 0$, need to form an F as follows:

$$F = \left(\text{Occ MSq + Occ\#Drivers\#Cars MSq}\right) / \left(\text{Occ\#Drivers MSq + Occ\#Cars MSq}\right)$$

The tests for Drivers and Cars are similar. However, there is a problem with the degrees of freedom for such F ratios. They are given by Satterthwaite's approximation which we do not cover in this course.

b) Case 2 — same Cars different Drivers

Layout (r=2)

	Occasions							
	1				2			
Cars	1	2	3	4	1	2	3	4
Drivers								
1	C	A	B	D	D	B	A	C
2	A	C	D	B	A	C	D	B
3	B	D	C	A	C	A	B	D
4	D	B	A	C	B	D	C	A

A. Description of pertinent features of the study

1. Observational unit – a car with a driver on an occasion
2. Response variable – Reduction
3. Unrandomized factors – Occasions, Drivers, Cars
4. Randomized factors – Additives
5. Type of study – Sets of Latin squares

B. The experimental structure

Structure	Formula
unrandomized	(*2* Occasions/*4* Drivers)***4* Cars
randomized	*4* Additives

C. Sources derived from the structure formulae

(Occasions/Drivers)*Cars = (Occasions+Drivers[Occasions])*Cars

= Occasions+ Drivers[Occasions]
+Cars+Occasions#Cars
+Drivers#Cars[Occasions]

Additives = Additives

D. Degrees of freedom and sums of squares

Left as an exercise

E. The analysis of variance table

Source	df	SSq
Occasions	r–1	$\mathbf{Y}'\mathbf{Q}_{\mathrm{O}}\mathbf{Y}$
Drivers[Occasions]	3r	$\mathbf{Y}'\mathbf{Q}_{\mathrm{OD}}\mathbf{Y}$
Cars	3	$\mathbf{Y}'\mathbf{Q}_{\mathrm{C}}\mathbf{Y}$
Occasions#Cars	3(r–1)	$\mathbf{Y}'\mathbf{Q}_{\mathrm{OC}}\mathbf{Y}$
Drivers#Cars[Occasions]	9r	$\mathbf{Y}'\mathbf{Q}_{\mathrm{ODC}}\mathbf{Y}$
Additives	3	$\mathbf{Y}'\mathbf{Q}_{\mathrm{A}}\mathbf{Y}$
Residual	9r–3	$\mathbf{Y}'\mathbf{Q}_{\mathrm{ODC}_{\mathrm{Res}}}\mathbf{Y}$
Total	16r–1	

F. Maximal expectation and variation models

Assume the randomized factor is a fixed factor and that all the unrandomized factors are random factors. Then the expectation term is Additive. The variation terms are: Occasions, Occasions$\wedge$ Drivers, Cars, Occasions$\wedge$Cars, Occasions$\wedge$Drivers$\wedge$Cars.

The expectation model is

$$\mathrm{E}[\mathrm{Y}] = \text{Additives}$$

and the variation model is

$$\mathrm{var}[\mathrm{Y}] = \text{Occasions} + \text{Occasions}\wedge\text{ Drivers} + \text{Cars} + \text{Occasions}\wedge\text{Cars} + \text{Occasions}\wedge\text{Drivers}\wedge\text{Cars}$$

G. The expected mean squares

The Hasse diagram, with contributions to expected mean squares, for the unrandomized factors in this study is left as an exercise. The single randomized factor Additive will contribute $q_{\mathrm{A}}(\boldsymbol{\psi})$ to the expected mean square for its source. The analysis of variance table with expected mean squares is:

Source	df	SSq	E[MSq]			
Occasions	$r-1$	$\mathbf{Y}'\mathbf{Q}_{\mathrm{O}}\mathbf{Y}$	σ^2_{ODC}	$+4\sigma^2_{\mathrm{OC}}$	$+4\sigma^2_{\mathrm{OD}}$	$+16\sigma^2_{\mathrm{O}}$
Drivers[Occasions]	$3r$	$\mathbf{Y}'\mathbf{Q}_{\mathrm{OD}}\mathbf{Y}$	σ^2_{ODC}		$+4\sigma^2_{\mathrm{OD}}$	
Cars	3	$\mathbf{Y}'\mathbf{Q}_{\mathrm{C}}\mathbf{Y}$	σ^2_{ODC}	$+4\sigma^2_{\mathrm{OC}}$		$+4r\,\sigma^2_{\mathrm{C}}$
Occasions#Cars	$3(r-1)$	$\mathbf{Y}'\mathbf{Q}_{\mathrm{OC}}\mathbf{Y}$	σ^2_{ODC}	$+4\sigma^2_{\mathrm{OC}}$		
Drivers#Cars[Occasions]	$9r$	$\mathbf{Y}'\mathbf{Q}_{\mathrm{ODC}}\mathbf{Y}$				
Additives	3	$\mathbf{Y}'\mathbf{Q}_{\mathrm{A}}\mathbf{Y}$	σ^2_{ODC}			$+q_{\mathrm{A}}(\boldsymbol{\psi})$
Residual	$9r-3$	$\mathbf{Y}'\mathbf{Q}_{\mathrm{ODC_{Res}}}\mathbf{Y}$	σ^2_{ODC}			
Total	$16r-1$					

Hypothesis tests for Drivers[Occasions], Cars, Occasions#Cars and Additives are straightforward: Drivers[Occasions], Occasions#Cars and Additives are compared with the Residual for Drivers#Cars[Occasions] and Cars with Occasions#Cars. The test for Occasions, $\sigma^2_{\mathrm{O}} = 0$, involves an F as follows:

$$F = \frac{(\text{Occasions MSq + Residual MSq})}{(\text{Drivers[Occasions] MSq + Occasions\#Cars MSq})}$$

Again, there is a problem with the degrees of freedom for this F ratio. They are given by Satterthwaite's approximation which we do not cover in this course.

c) Case 3 — different Drivers and Cars

Layout (r=2)

	Occasions							
	1				2			
Cars	1	2	3	4	1	2	3	4
Drivers								
1	B	A	C	D	D	B	C	A
2	C	D	B	A	A	C	B	D
3	A	B	D	C	B	D	A	C
4	D	C	A	B	C	A	D	B

A. Description of pertinent features of the study

1. Observational unit – a car with a driver on an occasion
2. Response variable – Reduction
3. Unrandomized factors – Occasions, Drivers, Cars
4. Randomized factors – Additives
5. Type of study – Sets of Latin squares

B. The experimental structure

Structure	Formula
unrandomized	*2* Occasions/(*4* Drivers**4* Cars)
randomized	*4* Additives

C. Sources derived from the structure formulae

Occasions/(Drivers*Cars) = Occasions/(Drivers + Cars + Drivers#Cars)
= Occasions + Drivers[Occasions]
+ Cars[Occasions] + Drivers#Cars[Occasions]

Additives = Additives

D. Degrees of freedom

Left as an exercise

E. The analysis of variance table

Source	df	SSq
Occasions	r–1	$\mathbf{Y}'\mathbf{Q}_{\mathrm{O}}\mathbf{Y}$
Drivers[Occasions]	3r	$\mathbf{Y}'\mathbf{Q}_{\mathrm{OD}}\mathbf{Y}$
Cars[Occasions]	3r	$\mathbf{Y}'\mathbf{Q}_{\mathrm{OC}}\mathbf{Y}$
Drivers#Cars[Occasions]	9r	$\mathbf{Y}'\mathbf{Q}_{\mathrm{ODC}}\mathbf{Y}$
Additives	3	$\mathbf{Y}'\mathbf{Q}_{\mathrm{A}}\mathbf{Y}$
Residual	9r–3	$\mathbf{Y}'\mathbf{Q}_{\mathrm{ODC}_{\mathrm{Res}}}\mathbf{Y}$
Total	16r–1	

F. Maximal expectation and variation models

Assume the randomized factor is a fixed factor and that all the unrandomized factors are random factors. Then the expectation term is Additives. The variation terms are: Occasions, Occasions∧ Drivers, Occasions∧Cars, Occasions∧Drivers∧Cars.

The expectation model is

E[Y] = Additives

and the variation model is

var[Y] = Occasions + Occasions∧ Drivers + Occasions∧Cars + Occasions∧Drivers∧Cars

G. The expected mean squares.

The Hasse diagram, with contributions to expected mean squares, for the unrandomized factors in this study is left as an exercise. The single randomized factor Additive will contribute $q_{\mathrm{A}}(\boldsymbol{\psi})$ to the expected mean square for its source. The analysis of variance table with expected mean squares is:

Source	df	SSq	E[MSq]			
Occasions	$r-1$	$\mathbf{Y}'\mathbf{Q}_{\mathrm{O}}\mathbf{Y}$	σ^2_{ODC}	$+4\sigma^2_{\mathrm{OC}}$	$+4\sigma^2_{\mathrm{OD}}$	$+16\sigma^2_{\mathrm{O}}$
Drivers[Occasions]	$3r$	$\mathbf{Y}'\mathbf{Q}_{\mathrm{OD}}\mathbf{Y}$	σ^2_{ODC}		$+4\sigma^2_{\mathrm{OD}}$	
Cars[Occasions]	$3r$	$\mathbf{Y}'\mathbf{Q}_{\mathrm{OC}}\mathbf{Y}$	σ^2_{ODC}	$+4\sigma^2_{\mathrm{OC}}$		
Drivers#Cars[Occasions]	$9r$	$\mathbf{Y}'\mathbf{Q}_{\mathrm{ODC}}\mathbf{Y}$				
Additives	3	$\mathbf{Y}'\mathbf{Q}_{\mathrm{A}}\mathbf{Y}$	σ^2_{ODC}			$+q_{\mathrm{A}}(\boldsymbol{\psi})$
Residual	$9r-3$	$\mathbf{Y}'\mathbf{Q}_{\mathrm{ODC}_{\mathrm{Res}}}\mathbf{Y}$	σ^2_{ODC}			
Total	$16r-1$					

Hypothesis tests for Drivers[Occasions], Cars[Occasions] and Additives are straightforward: they are all compared with the Residual for Drivers:Cars[Occasions]. The test for Occasions, $\sigma^2_{\mathrm{O}} = 0$, involves an F as follows:

$$F = \frac{(\text{Occasions MSq} + \text{Residual MSq})}{(\text{Drivers[Occasions] MSq} + \text{Cars[Occasions] MSq})}$$

Again, there is a problem with the degrees of freedom for this F ratio. They are given by Satterthwaite's approximation which we do not cover in this course.

d) Comparison of Latin square analyses

One square only

Unrandomized Structure	Car*Driver

Source	df
Car	3
Driver	3
Driver#Car	9
Additives	3
Residual	6

Case 1
Same drivers and cars

Source	df
Occ	r–1
Driver	3
Car	3
Occ#Driver	3(r–1)
Occ#Car	3(r–1)
Driver#Car	9
Additives	3
Residual	6
Occ#Driver#Car	9(r–1)
Total	16r–1

Case 2
Same cars
Different drivers

Source	Df
Occ	r–1
Driver[Occ]	3r
Cars	3
Occ#Car	3(r–1)
Drive#Car[Occ]	9r
Additives	3
Residual	9r–3
Total	16r–1

Case 3
Different
drivers and cars

Source	df
Occ	r–1
Driver[Occ]	3r
Car[Occ]	3r
Drive#Car[Occ]	9r
Additives	3
Residual	9r–3
Total	16r–1

Unrandomized structure

Case 1	Case 2	Case 3
2 Occasions**4* Driver**4* Car	(*2* Occ/*4* Driver)**4* Car	*2* Occ/(*4* Driver**4* Car)

e) Computation of ANOVA in R

The analysis of one of these experiments in R is obtained by using a *model formula* in which the explanatory variables are specified to be

Additives + Error(unrandomized structure formula)

In addition, fixed terms in the unrandomized structure formula need to be also included with the terms outside the `Error` function. For further information see Appendix C, *Analysis of designed experiments in R*.

V.H Summary

In this chapter we have:

- described how to design an experiment using a Latin square design and a set of Latin square designs;
- formulated a linear model using indicator variables to describe the results from a Latin square design; given the estimators of the parameters in the linear model, the expected values and the random errors as functions of **M** or mean operator matrices;
- outlined an ANOVA hypothesis test for choosing between expectation models;
  - the partition of the total sums of squares was given with the sums of squares expressed as the sums of squares of the elements of vectors and as quadratic forms where the matrices of the quadratic forms, **Q** matrices, are symmetric idempotents;
  - discussed the differences between the analyses where Rows and Columns are fixed and where they are random;
  - the expected mean squares under the alternative expectation models are used to justify the choice of F test statistic;
- shown how to obtain a layout and the analysis of variance in R;
- discussed procedures for checking the adequacy of the proposed models;
- subsequent to the hypothesis test, examined treatment differences in detail;
- used the rules from chapter VI, *Determining the analysis of variance table*, for formulating the analyses for sets of Latin square designs.

V.I Exercises

V.1 It is desired to run a wine-tasting experiment in which the differences between six wines are to be evaluated by scoring them on a 20-point scale. It is decided to have 6 expert judges evaluate the wines by evaluating a glass of wine on each of six consecutive occasions. It is desired to be able to isolate judge differences in scoring and differences between occasions so that a Latin square is to be employed.

A standard Latin square layout is given below.

A	B	C	D	E	F
B	A	F	E	C	D
C	F	B	A	D	E
D	C	E	B	F	A
E	D	A	F	B	C
F	E	D	C	A	B

Use R to obtain a randomized layout for the experiment, using a `seed` of 559 to set the random number generator seed.

V.2 In the lecture, a Latin square example was discussed that involved 4 drivers and 4 cars in testing the effects of 4 additives on the pollution produced by the cars. Discuss the circumstances in which the factors Drivers and Cars are likely to be regarded as fixed and those in which they are likely to be regarded as random.

V.3 The following data are from a Latin square experiment designed to investigate the moisture content of turnip greens. The experiment involved the measurement of the percent moisture content of five leaves of different sizes from each of five plants. The treatments were time of measurement in days since the beginning of the experiment.

		Plant				
		1	2	3	4	5
	A	5 6.67	2 5.40	3 7.32	1 4.92	4 4.88
	B	4 7.15	5 4.77	2 8.53	3 5.00	1 6.16
Leaf Size	C	1 8.29	4 5.40	5 8.50	2 7.29	3 7.83
(A = smallest,	D	3 8.95	1 7.54	4 9.99	5 7.85	2 5.83
E = largest)	E	2 9.62	3 6.93	1 9.68	4 7.08	5 8.51

Classify the factors Leaf Size and Plant as either fixed or random.

The factors Size, Plant and Time and the Moisture contents have been saved in the data.frame file *LSTurn.dat.rda* available from the available from the Statistical Modelling resources web site.

Analyze the data using R, including diagnostic checking and the examination of mean differences. Note that If you designated any one of the factors Leaf Size and Plant as random, the corresponding function $q(\boldsymbol{\psi})$ should be replaced by $5\sigma_S^2$ or $5\sigma_P^2$ in the E[MSq] for the line for the factor that is random (5 is the number of replicates of each Leaf Size and of each Plant).

V.4 The following layout is that appropriate to an experiment in which the same four drivers and the same four cars are used to repeat testing of four petrol additives. The assignment of additives to cars and drivers was accomplished using a Latin square on each of the two occasions the testing is conducted.

	Occasions							
	1				2			
Car	1	2	3	4	1	2	3	4
Driver								
1	A	B	C	D	A	B	C	D
2	C	D	A	B	C	D	A	B
3	D	C	B	A	D	C	B	A
4	B	A	D	C	B	A	D	C

Use R to verify that the analysis given in class, assuming all unrandomized factors are random, is the appropriate one for this experiment. You will need to set up the factor information, generate some random numbers from say a normal distribution using the `rnorm` function, and use the `aov` and `summary` functions to analyse the random data. For the `aov` function, the model formula should include an `Error` function whose argument is the unrandomized structure. Because all unrandomized factors are random, it is not necessary to include, outside the `Error` function, any terms from inside the `Error` function.