STATISTICAL MODELLING

IX. Split-plot experiments

(Cochran and Cox, sec. 7.3; Mead, ch. 14; Mead and Curnow, sec. 6.7)

IX.A Design of split-plot experiments

When discussing the topic of confounding, our goal was to leave main effects unconfounded, or perhaps, only partially confounded and with most of the information about main effects confounded with less variable units.

On occasion, however, it is desirable to have main effects confounded with more variable units such as large plots. This particular class of designs is called split-plot designs for reasons that will become obvious. Their defining attribute is that there is randomization to two different physical entities such that some main effects are randomized to the more variable entities.

Definition IX.1: The **standard split-plot design** is one in which two factors, say A and B with *a* and *b* levels, respectively are assigned as follows: one of the factors, A say, is randomized according to a RCBD with say *r* blocks and each of its *ra* plots, called the **main plots**, is split into *b* **subplots** (or split-plots) and the levels of B randomized independently in each subplot. Altogether the experiment involves $n = rab$ subplots. ■

That is, the generic factor names for this design are Blocks, MainPlots, SubPlots, A and B.

The split-plot principle is very flexible and can be used to generate a large number of different types of experiments. For example, the main plots could be arranged in any of a CRD, RCBD, Latin square, BIBD, Youden Square and each plot of the particular design used then subdivided into subplots.

The subplots may utilize more complicated designs as well. For example, the main plots may be arranged in a RCBD each of which are subdivided in such a way as to allow a Latin Square to be placed in each main plot.

Also, subplots can be split into subsubplots and subsubplots into ... Nor is one restricted to applying just one factor to each type of unit. More than one factor can be randomized to main plots, more than one to subplots and so on.

The standard split-plot design is nearly the simplest possibility; only a CRD in the main plots would be simpler.

The split-plot design is useful in the following situations:

1. When the physical attributes of a factor require the use of larger units of experimental material than other factors.

   For example, land preparation treatments usually require to be performed on larger areas of land than do the sowing of different varieties (due to the different pieces of equipment). Temperature control for say storage purposes involves the use of relatively large chambers in which several samples can usually be stored. Different processing runs are often of a minimum size such that their produce can be readily subdivided for the application of further treatments. Also, some factors are relatively hard to change. For example, the temperature of a production operation is often difficult to change so that it might be better to change it less often by making it a main-plot factor.

2. When it is desired to incorporate an additional factor into an experiment.

   When an experiment, such as an agricultural experiment is run over a period of time, it sometimes occurs after the trial has been set up that it would be advantageous to incorporate an additional factor. This can be achieved by splitting the existing plots into enough subplots so that the levels of the new factor(s) can be randomized to the subplots in each subplot.

3. When it is expected that the differences amongst the levels of certain factors are larger than amongst those of other factors.

   The levels of the factors with larger differences are randomized to main plots. One effect of this may be to increase the precision of comparisons between the levels of the other factors.

4. When it is desired to ensure greater precision between some factors than others.

   Irrespective of the size of the differences between the main plot treatment factors, it is desired to increase the precision of some factors by assigning them to subplots. On the other hand, it may be that one is less interested in the main effects of some factors, in which case these factors should be assigned to main plots. A particular example of such factors is "noise" factors which are often included in production experiments to ensure the product works under a range of conditions; one is usually more interested in product differences and the interaction of products and the "noise" factors.

Note that the last two of these situations are utilising the anticipated greater variability of main plots relative to subplots. That is, we are expecting the larger units to be more variable than the smaller units. This will be expressed in the variation model for these experiments, and hence in the expected mean squares.

In describing the type of study, you need to identify the main plot and subplot design.

IX.B The standard split-plot experiment

a) Designing a standard split-plot experiment

In the standard split-plot, the main-plot treatment factor, A, is randomized to main plots and the subplot treatment factor, B, is randomized to subplots. The expressions for using R to obtain a layout for such an experiment are given in Appendix B, *Randomized layouts and sample size computations in R*.

Example IX.1 Production rate experiment

Suppose that one is interested in comparing 3 methods of work organization and 3 sources of raw material on the production rate of a certain product. It is decided that four factories are to be used in the experiment and that each factory is to be divided into three areas. The methods of work organization are to be assigned at random to areas. Each area is to be subdivided into 3 parts and the source of raw material for each part is obtained by randomizing the three sources to the three parts. The design, obtained using R, is illustrated in the following schematic diagram:

	Factories											
	1			2			3			4		
	Parts			Parts			Parts			Parts		
Area	1	2	3	1	2	3	1	2	3	1	2	3
I	2 C	2 A	2 B	2 C	2 B	2 A	2 B	2 A	2 C	3 B	3 C	3 A
II	3 B	3 A	3 C	1 A	1 B	1 C	3 C	3 B	3 A	1 C	1 A	1 B
III	1 B	1 C	1 A	3 A	3 B	3 C	1 B	1 A	1 C	2 B	2 A	2 C

Numerals inside boxes represent methods of organisation and A–C represent source of raw material

b) Determining the analysis of variance table

For these experiments we will not derive the analysis in full; rather the rules for determining the analysis of variance table will be relied upon to establish the analysis. In general, the analysis for these experiments is complicated by the fact that there is no analysis with $\mathbf{V} = \sigma^2\mathbf{I}$. Consequently, generalized least squares must be employed in deriving their analysis.

Example IX.1 Production rate experiment (continued)

a) Description of pertinent features of the study

1. Observational unit – a part
   Variables (incl. factors) are? Ans. Production, Factories, Areas, Parts, Methods, Sources
2. Response variable – Production
3. Unrandomized factors – Factories, Areas, Parts
4. Randomized factors – Methods, Sources
5. Type of study – Standard split-plot with main plots in an RCBD and subplots completely randomized within an area

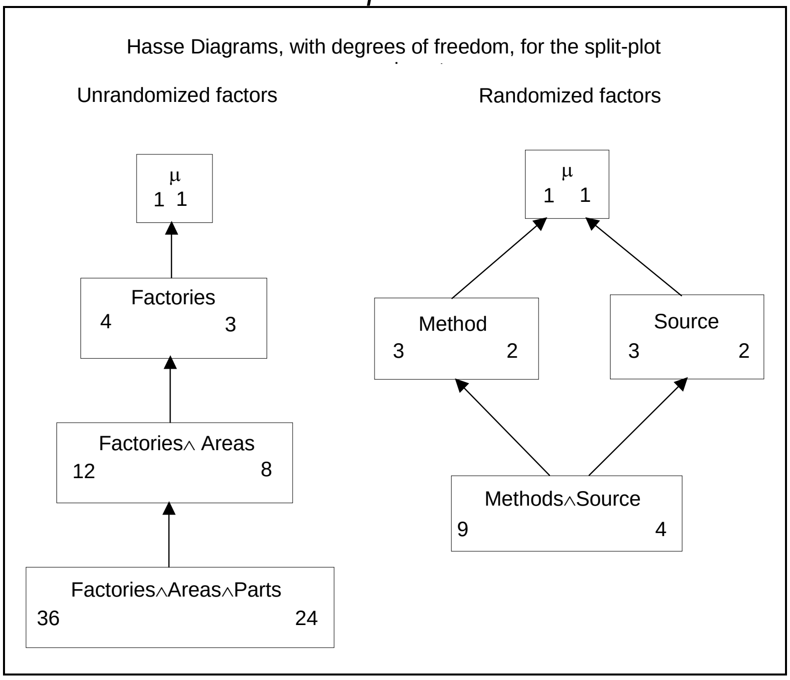
b) The experimental structure

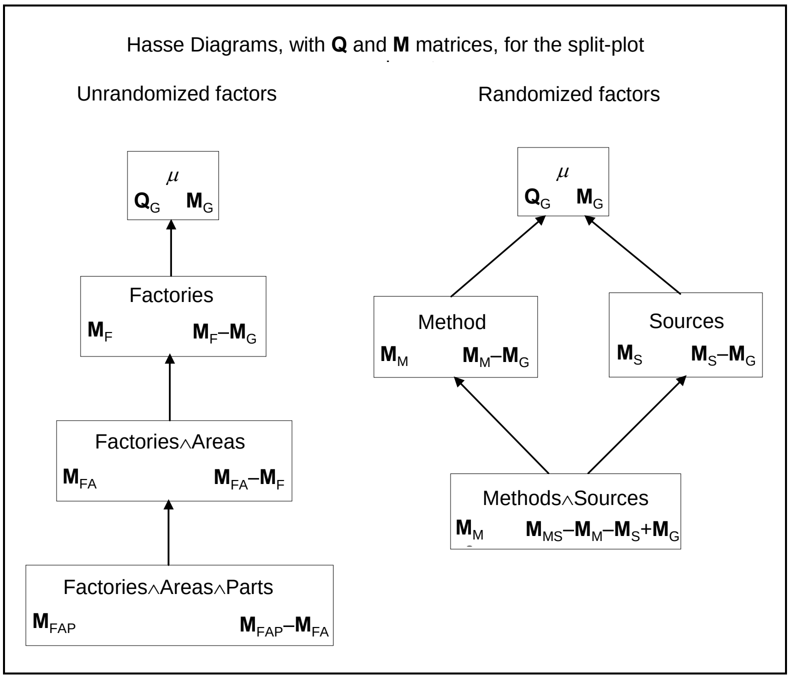
Structure	Formula
unrandomized	*4* Factories/*3* Areas/*3* Parts
randomized	*3* Methods**3* Sources

c) Sources derived from the structure formulae

Factories/Areas/Parts = (Factories + Areas[Factories])/Parts
= Factories + Areas[Factories] + Parts[Factories∧Areas]
and Methods*Sources = Methods + Sources + Methods#Sources

d) Degrees of freedom and sums of squares





Note, that in working out the degrees of freedom for the sources from the randomized structure, the rule for a set of crossed factors can be used. That is, for each factor in the source, calculate the number of levels minus one and multiply these together.

e) The analysis of variance table

Enter the sources for the study, their degrees of freedom and quadratic forms, into the analysis of variance table below.

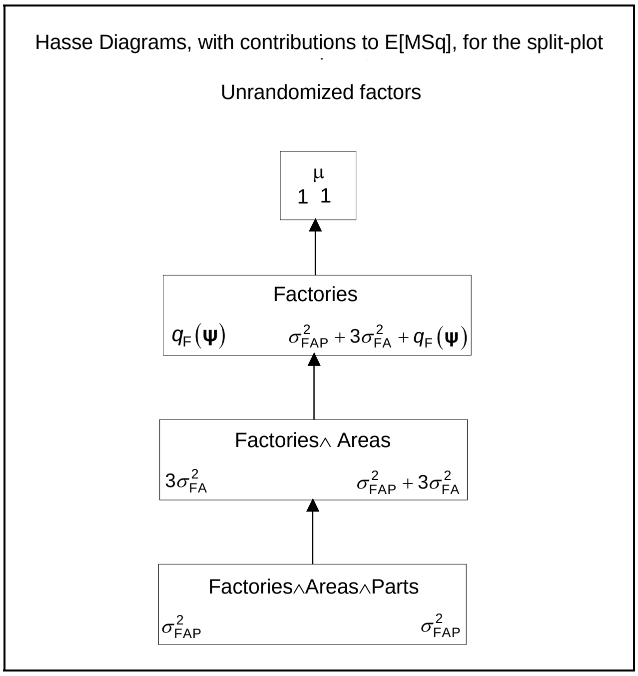
f) Maximal expectation and variation models

Clearly, the factors Methods is fixed and, as it is most likely that there are only a limited number of Sources, Sources is also likely to be fixed. Suppose that the Factories in the experiment are the only four that are owned by the company conducting the experiment — thus Factories should be taken to be fixed. The factors Areas and Parts must be taken as random as they had Methods and Source randomized to them — the observed areas and parts will be taken to be representative of a large number of areas and parts that could conceptually be set up in the factories. Hence, the maximal models for this experiment are:

$\psi = \mathrm{E}[\mathrm{Y}]$ = Factories + Methods∧Sources
and var[Y] = Factories∧Areas + Factories∧Areas∧Parts.

g) The expected mean squares.

The contributions of the randomized factors will all be of the form $q_F(\boldsymbol{\psi})$. The Hasse diagram, with the contributions to the expected mean squares of the unrandomized factors, is as follows:



The analysis of variance table for this experiment has the following form:

Source	df	SSq	E[MSq]		
Factories	3	$\mathbf{Y}'\mathbf{Q}_F\mathbf{Y}$	σ^2_{FAP}	$+3\sigma^2_{FA}$	$+q_F(\boldsymbol{\psi})$
Areas[Factories]	8	$\mathbf{Y}'\mathbf{Q}_{FA}\mathbf{Y}$			
Methods	2	$\mathbf{Y}'\mathbf{Q}_M\mathbf{Y}$	σ^2_{FAP}	$+3\sigma^2_{FA}$	$+q_M(\boldsymbol{\psi})$
Residual	6	$\mathbf{Y}'\mathbf{Q}_{FA_{Res}}\mathbf{Y}$	σ^2_{FAP}	$+3\sigma^2_{FA}$	
Parts[Factories∧Areas]	24	$\mathbf{Y}'\mathbf{Q}_{FAP}\mathbf{Y}$			
Sources	2	$\mathbf{Y}'\mathbf{Q}_S\mathbf{Y}$	σ^2_{FAP}		$+q_S(\boldsymbol{\psi})$
Methods#Sources	4	$\mathbf{Y}'\mathbf{Q}_{MS}\mathbf{Y}$	σ^2_{FAP}		$+q_{MS}(\boldsymbol{\psi})$
Residual	18	$\mathbf{Y}'\mathbf{Q}_{FAP_{Res}}\mathbf{Y}$	σ^2_{FAP}		

The design chosen has resulted in the differences between methods of organization being confounded with areas differences whereas differences between sources of raw material are confounded with parts (the areas do not differ in source of raw material). This has been achieved by splitting the main plots (areas in this case) into subplots (parts in this case). This is evident from the above table.

Now it is likely that Areas will be more variable than parts of an area so that differences between Methods may well be estimated with less precision than Source differences.

c) Analysis of the example

Example IX.1 Production rate experiment (continued)

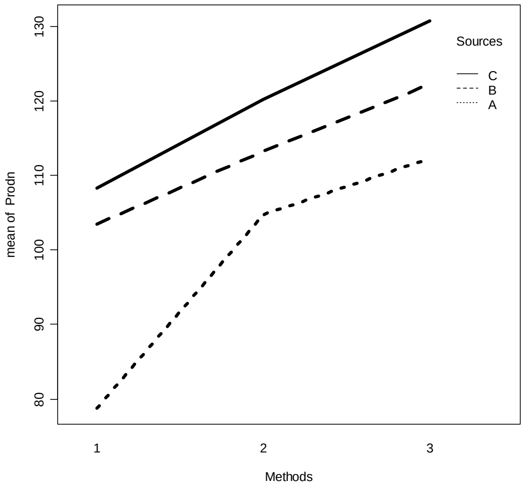
The data for the experiment is as follows:

Factories	Areas	Parts	Methods	Sources	Prod
1	1	1	2	C	119
		2		A	121
		3		B	110
	2	1	3	B	98
		2		A	78
		3		C	122
	3	1	1	B	87
		2		C	100
		3		A	80
2	1	1	2	C	126
		2		B	131
		3		A	109
	2	1	1	A	73
		2		B	114
		3		C	114
	3	1	3	A	116
		2		B	136
		3		C	133
3	1	1	2	B	94
		2		A	99
		3		C	123
	2	1	3	C	132
		2		B	133
		3		A	136
	3	1	1	B	109
		2		A	102
		3		C	105
4	1	1	3	B	122
		2		C	136
		3		A	119
	2	1	1	C	114
		2		A	60
		3		B	104
	3	1	2	B	118
		2		A	90
		3		C	113

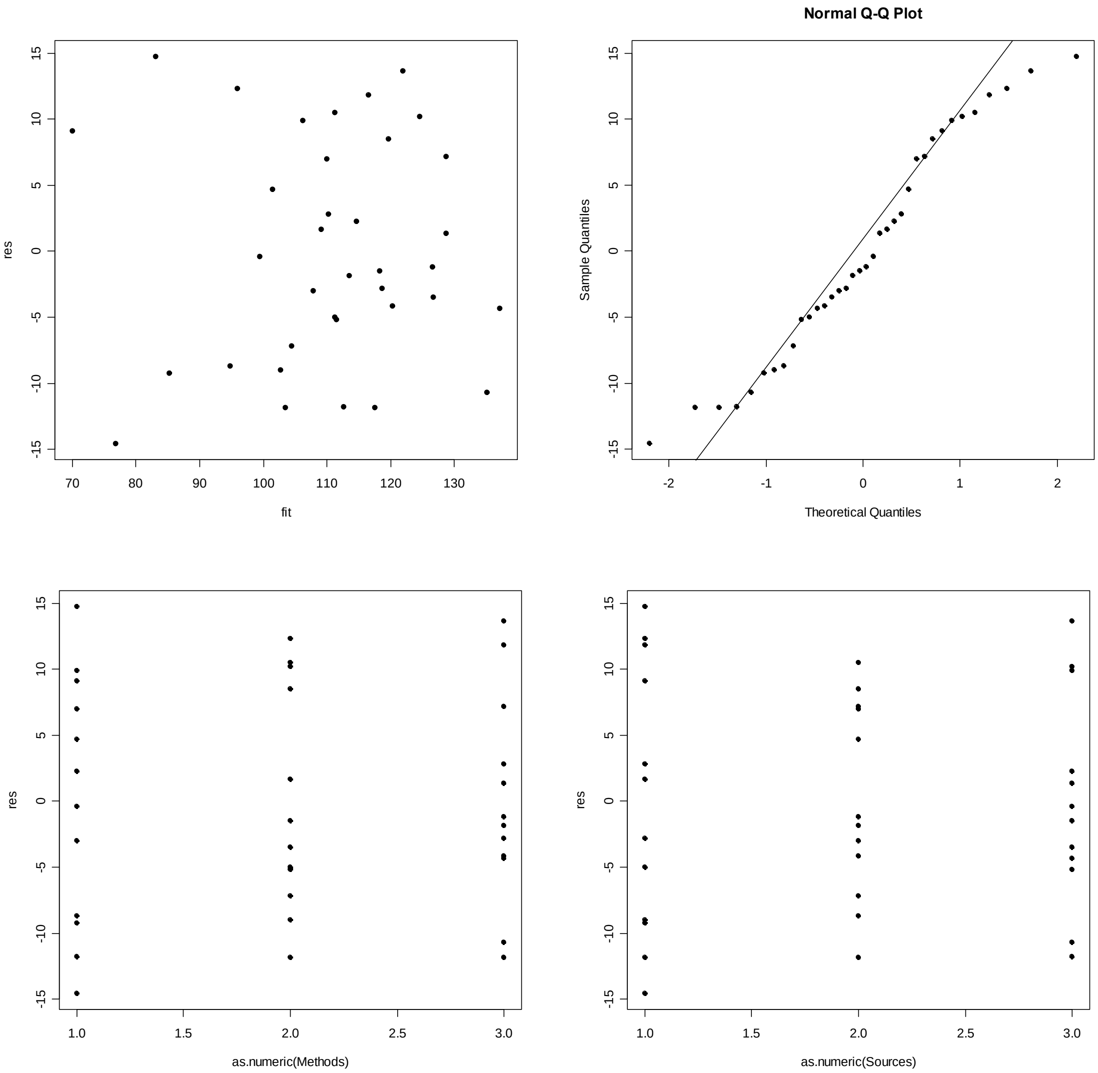
The R expressions for analyzing this experiment are summarized in Appendix C. The R output file from the expressions are as follows:

```
> attach(SPLProd.dat)
> interaction.plot(Methods, Sources, Prodn, lwd = 4)
> SPLProd.aov <- aov(Prodn ~ Factories + Methods * Sources +
+                         Error(Factories/Areas/Parts), SPLProd.dat)
```

```
> summary(SPLProd.aov)

Error: Factories
          Df  Sum Sq Mean Sq
Factories  3 1272.22  424.07

Error: Factories:Areas
          Df Sum Sq Mean Sq F value Pr(>F)
Methods    2 3820.7  1910.4   6.052 0.0364
Residuals  6 1893.9   315.7

Error: Factories:Areas:Parts
                Df  Sum Sq Mean Sq F value  Pr(>F)
Sources          2 2805.72 1402.86 10.2447 0.00107
Methods:Sources  4  369.44   92.36  0.6745 0.61829
Residuals       18 2464.83  136.94
> #Compute Factories and Areas[Factories] Fs and p-values
> Factories.F <- 424.07/315.7
> Factories.p <- 1-pf(Factories.F, 3, 6)
> Factories.Areas.F <- 315.7/136.94
> Factories.Areas.p <- 1-pf(Factories.Areas.F, 6, 18)
> data.frame(Factories.F,Factories.p,Factories.Areas.F,Factories.Areas.p)
  Factories.F Factories.p Factories.Areas.F Factories.Areas.p
1    1.343269   0.3459115          2.305389        0.07924927
> #
> # Diagnostic checking
> #
> tukey.1df(SPLProd.aov, SPLProd.dat, "Factories:Areas:Parts")
$Tukey.SS
[1] 3.437533

$Tukey.F
[1] 0.02374184

$Tukey.p
[1] 0.879358

$Devn.SS
[1] 2461.396

> res <- resid.errors(SPLProd.aov)
> fit <- fitted.errors(SPLProd.aov)
> plot(fit, res, pch=16)
> qqnorm(res, pch=16)
> qqline(res)
> plot(as.numeric(Methods), res, pch=16)
> plot(as.numeric(Sources), res, pch=16)
```



Looks like there is an interaction.



The analysis of variance table for the example is:

Source	df	MSq	E[MSq]	F	Prob
Factories	3	424.1	$\sigma^2_{\mathrm{FAP}} + 3\sigma^2_{\mathrm{FA}} + q_{\mathrm{F}}(\boldsymbol{\psi})$	1.34	0.346
Areas[Factories]	8				
Methods	2	1910.4	$\sigma^2_{\mathrm{FAP}} + 3\sigma^2_{\mathrm{FA}} + q_{\mathrm{M}}(\boldsymbol{\psi})$	6.05	0.036
Residual	6	315.7	$\sigma^2_{\mathrm{FAP}} + 3\sigma^2_{\mathrm{FA}}$	2.31	0.079
Parts[Factories∧Areas]	24				
Sources	2	1402.9	$\sigma^2_{\mathrm{FAP}} + q_{\mathrm{S}}(\boldsymbol{\psi})$	10.24	0.001
Methods#Sources	4	92.4	$\sigma^2_{\mathrm{FAP}} + q_{\mathrm{MS}}(\boldsymbol{\psi})$	0.67	0.618
Residual	18	136.9	σ^2_{FAP}		
Nonadditivity	1	3.4		0.02	0.879
Deviations	17	144.8			

The residual-versus-fitted-values, residuals-versus-factors and normal probability plots appear to be satisfactory. The data is not exhibiting transformable nonadditivity as Tukey's test for nonadditivity is not significant.

The interaction is not significant ($p = 0.618$) so that we need to examine the main effects to see which of these affects the response. There are significant differences between both the methods ($p = 0.036$) and the sources of material ($p = 0.001$), but not between factories. Note also that the Areas would appear to be more variable, although not significantly, than the parts within an area ($p = 0.079$).

So the fitted model for the expectation is $\psi = \mathrm{E}[Y] =$ Methods + Sources.

d) Treatment differences for the standard split-plot

As for other experiments treatment differences are investigated using multiple comparisons or polynomial submodels. Of these, only multiple comparisons is different to the ordinary factorial experiment and we investigate its use in the next section.

Multiple comparisons

As usual we base the multiple comparisons on Tukey's HSD which, as usual, is given by $HSD = \left(q/\sqrt{2}\right)s.e.d.$ However, their calculation is slightly more complicated for split-plot experiments because the standard errors for the difference vary. The formulae for the standard errors of difference for a split-plot design in which one factor (A) is randomized to whole-plots and another (B) to subplots (as in the example) are given in the table below. The degrees of freedom those for the Residual Mean square in the s.e.d.; for the last case, involving two Mean squares, the degrees of freedom are calculated using Satterthwaite's procedure, a procedure that is not covered in these notes.

Table of means	s.e.d.	
Two overall A means (e.g. Variety means)	$\sqrt{\dfrac{2s^2_{\mathrm{M_{Res}}}}{rb}}$	where *rb* is the no. of replications in an A mean
Two overall B means (e.g. Fertilizer means)	$\sqrt{\dfrac{2s^2_{\mathrm{S_{Res}}}}{ra}}$	where *ra* is the no. of replications in a B mean
Two interaction means		
– at the same level of A	$\sqrt{\dfrac{2s^2_{\mathrm{S_{Res}}}}{r}}$	where *r* is the no. of replications in an A∧B mean
– not at the same level of A	$\sqrt{\dfrac{2\left[(b-1)s^2_{\mathrm{S_{Res}}}+s^2_{\mathrm{M_{Res}}}\right]}{rb}}$	

Note: $s^2_{\mathrm{M_{Res}}}$ & $s^2_{\mathrm{S_{Res}}}$ are the Main-plot and Split-plot Residual MSqs, respectively, from the ANOVA table.

In terms of using R, as usual, we use `model.tables` to produce the tables of means and `qtukey` to produce the standardized range. Then the rest is done manually.

Example IX.1 Production rate experiment (continued)

Here are the tables of means and studentized range values for the example.

```
> SPLProd.means <- model.tables(SPLProd.aov, type="means")
> SPLProd.means
Tables of means
Grand mean

110.4444

 Factories
Factories
     1      2      3      4
101.67 116.89 114.78 108.44

 Methods
Methods
     1      2      3
 96.83 112.75 121.75

 Sources
Sources
     A      B      C
 98.58 113.00 119.75

 Methods:Sources
       Sources
Methods A      B      C
      1  78.75 103.50 108.25
```

```
      2 104.75 113.25 120.25
      3 112.25 122.25 130.75
> qtukey(0.95, 3, 6)
[1] 4.339195
> qtukey(0.95, 3, 18)
[1] 3.609304
```

The two tables of means of interest for the example are those for Methods and Sources. The tables, with Tukey's HSD, are:

Methods		
1	2	3
96.8	112.8	121.8

$$s.e.d. = \sqrt{\frac{2s^2_{\mathrm{FA}_{\mathrm{Res}}}}{rb}} = \sqrt{\frac{2 \times 315.7}{4 \times 3}} = 7.25$$

$$w(5\%) = \frac{q_{3,6,0.05}}{\sqrt{2}}\, s.e.d. = \frac{4.34}{\sqrt{2}} 7.25 = 22.21$$

Sources		
1	2	3
98.6	113.0	119.8

$$s.e.d. = \sqrt{\frac{2s^2_{\mathrm{FAP}_{\mathrm{Res}}}}{ra}} = \sqrt{\frac{2 \times 136.9}{4 \times 3}} = 4.78$$

$$w(5\%) = \frac{q_{3,18,0.05}}{\sqrt{2}}\, s.e.d. = \frac{3.61}{\sqrt{2}} 4.78 = 12.20$$

Tukey's HSD indicates that Methods 1 and 3 are different and that Method 2 is not different to either; also, Source 1 is lower than the other two.

IX.C Systematic or Unreplicated Main Plots

(Cochran and Cox, sec. 7.3)

As was pointed out at the outset split-plot designs are often used where one of the treatments requires larger units to apply. Sometimes the difficulty in applying the larger units means that it is not possible to randomize the treatments to the main plots or even to replicate them. In this section we look at the consequences of these two variations in the layout of a standard split-plot experiment.

Example IX.2 Varieties applied systematically

Suppose an experiment is to be run to investigate the effect on the yield of five varieties of wheat and 2 levels of fertilizer. The experiment is to involve three blocks. However, it was know that the varieties would ripen in a particular order and to facilitate harvest the varieties were assigned to the main plots in the same order within each block. The following diagram shows the layout with V_1 being the earliest variety and V_5 the latest variety to ripen.

Layout for systematic main plots in a split-plot experiment

Variety

Block		V_1	V_2	V_3	V_4	V_5
	I	N	Y	Y	N	N
		Y	N	N	Y	Y
		V_1	V_2	V_3	V_4	V_5
	II	N	N	Y	Y	N
		Y	Y	N	N	Y
		V_1	V_2	V_3	V_4	V_5
	III	N	N	N	Y	N
		Y	Y	Y	N	Y

N = No; Y = Yes

The experimental structure is:

Structure	Formula
unrandomized	(*3* Blocks**5* Plots)/*2* Subplots
randomized	*5* Variety**2* Fertilizer

In this structure we have Plots crossed with Blocks. This is appropriate because the first plot in each block has in common that they are in the same relative position in the block and that they will be harvested first.

The analysis of variance table for this experiment has the following form:

Source	df	SSq	E[MSq]			
Blocks	2	$\mathbf{Y}'\mathbf{Q}_{\mathrm{B}}\mathbf{Y}$	σ^2_{BPS}	$+2\sigma^2_{\mathrm{BP}}$		$+10\sigma^2_{\mathrm{B}}$
Plots	4	$\mathbf{Y}'\mathbf{Q}_{\mathrm{P}}\mathbf{Y}$				
Variety	4	$\mathbf{Y}'\mathbf{Q}_{\mathrm{V}}\mathbf{Y}$	σ^2_{BPS}	$+2\sigma^2_{\mathrm{BP}}$	$+6\sigma^2_{\mathrm{P}}$	$+q_{\mathrm{V}}(\psi)$
Blocks#Plots	8	$\mathbf{Y}'\mathbf{Q}_{\mathrm{BP}}\mathbf{Y}$	σ^2_{BPS}	$+2\sigma^2_{\mathrm{BP}}$		
Subplots[Blocks∧Plots]	15	$\mathbf{Y}'\mathbf{Q}_{\mathrm{BPS}}\mathbf{Y}$				
Fertilizer	1	$\mathbf{Y}'\mathbf{Q}_{\mathrm{F}}\mathbf{Y}$	σ^2_{BPS}			$+q_{\mathrm{F}}(\psi)$
Variety#Fertilizer	4	$\mathbf{Y}'\mathbf{Q}_{\mathrm{VF}}\mathbf{Y}$	σ^2_{BPS}			$+q_{\mathrm{VF}}(\psi)$
Residual	10	$\mathbf{Y}'\mathbf{Q}_{\mathrm{BPS}_{\mathrm{Res}}}\mathbf{Y}$	σ^2_{BPS}			

The implications of this analysis are that it is not possible to separate plot variability from Variety differences. While you could test Variety differences against the

Blocks#Plots line it is not possible to determine whether these differences involve Variety as well as overall Plot differences. However, it is possible to examine fertilizer differences both within varieties and averaged over all varieties.

Example IX.3 Unreplicated irrigation treatments

Irrigation treatments usually require large areas and it is often difficult to change between irrigated and nonirrigated areas. For example, in an experiment to investigate the effects of irrigation on the yield of different varieties of tomatoes, one area of tomatoes is irrigated and a second is not. Within each area eight varieties are randomized to 16 plots so that each variety occurs twice. In each plot is grown five tomato bushes and the yield of each is obtained.

The experimental structure is:

Structure	Formula
unrandomized	*2* Areas/*16* Plots/*5* Bushes
randomized	*2* Irrigations**8* Varieties

The analysis of variance table for this experiment has the following form:

Source	df	SSq	E[MSq]
Areas	1	$\mathbf{Y}'\mathbf{Q}_{\mathrm{A}}\mathbf{Y}$	
Irrigations	1	$\mathbf{Y}'\mathbf{Q}_{\mathrm{I}}\mathbf{Y}$	$\sigma^2_{\mathrm{APB}} + 5\sigma^2_{\mathrm{AP}} + 80\sigma^2_{\mathrm{A}} + q_{\mathrm{I}}(\psi)$
Plots[Areas]	30	$\mathbf{Y}'\mathbf{Q}_{\mathrm{AP}}\mathbf{Y}$	
Variety	7	$\mathbf{Y}'\mathbf{Q}_{\mathrm{V}}\mathbf{Y}$	$\sigma^2_{\mathrm{APB}} + 5\sigma^2_{\mathrm{AP}} + q_{\mathrm{V}}(\psi)$
Variety#Irrigations	7	$\mathbf{Y}'\mathbf{Q}_{\mathrm{VI}}\mathbf{Y}$	$\sigma^2_{\mathrm{APB}} + 5\sigma^2_{\mathrm{AP}} + q_{\mathrm{VI}}(\psi)$
Residual	16	$\mathbf{Y}'\mathbf{Q}_{\mathrm{AP}_{\mathrm{Res}}}\mathbf{Y}$	$\sigma^2_{\mathrm{APB}} + 5\sigma^2_{\mathrm{AP}}$
Bushes[Areas∧Plots]	128	$\mathbf{Y}'\mathbf{Q}_{\mathrm{APB}}\mathbf{Y}$	σ^2_{APB}

Like the split-plot experiment with main-plot treatments applied systematically, in unreplicated experiments it is not possible to separate plot variability from Irrigation differences. While you could test Irrigation differences against the residual for Plots[Areas] it is not possible to determine whether these differences involve Irrigation differences as well as Plot differences. However, it is possible to examine variety differences both within irrigations and averaged over both irrigation treatments.

IX.D A Complex Split-Plot Experiment

As was mentioned in the introduction the split-plot principle is very flexible. It can involve several levels of splitting. At each level a range of designs can be employed to assign the treatments to that level and as many factors as desired can be applied. Because of this the number of possible designs is limitless and quite complex designs are possible. To demonstrate the flexibility of such designs and the utility of the approach we use in determining the analysis of variance table, we next consider a more complicated split-plot design.

Example IX.4 Grazing experiment

An experiment is conducted to investigate the effects on pasture composition of different patterns of grazing. The different patterns were specified by three treatment factors. The three treatment factors were:

- Period: the length of the period for which plots were grazed; 3, 9 or 18 days;
- Spring grazing: the number of cycles of grazing in Spring; either 2 periods of grazing, with long gaps, or 4 periods of grazing, with short gaps;
- Summer grazing: the number of cycles in Summer; either 2 or 4 periods of grazing, with long or short gaps, respectively.

The experimental design involved the assignment of Periods to plots using a 3 × 3 Latin square. Each plot was split into 2 SubRows × 2 SubColumns; the 2 numbers of grazing cycles for Spring and Summer randomized to the SubRows and SubColumns, respectively, within each plot. The measured response was the percentage covered by the principal grass. The field layout was as given in the table below.

Layout for the grazing experiment

Period		Summer			Summer			Summer	
	18	2	4	**9**	4	2	**3**	4	2
Spring	4			2			2		
	2			4			4		
	9	4	2	**3**	2	4	**18**	2	4
Spring	2			2			2		
	4			4			4		
	3	2	4	**18**	2	4	**9**	4	2
Spring	2			2			2		
	4			4			4		

a) Description of pertinent features of the study

1. Observational unit – a subrow-subcolumn combination
2. Response variable – %Covered by Principal Grass
3. Unrandomized factors – Rows, Columns, Subrows, Subcolumns
4. Randomized factors – Periods, Spring, Summer
5. Type of study – Split-plot with main plots in an LS and subplots arranged in rows and columns with factors randomized to them

b) The experimental structure

Structure	Formula
unrandomized	(*3* Rows**3* Columns)/(*2* Subrows**2* Subcolumns)
randomized	*3* Periods**2* Spring**2* Summer

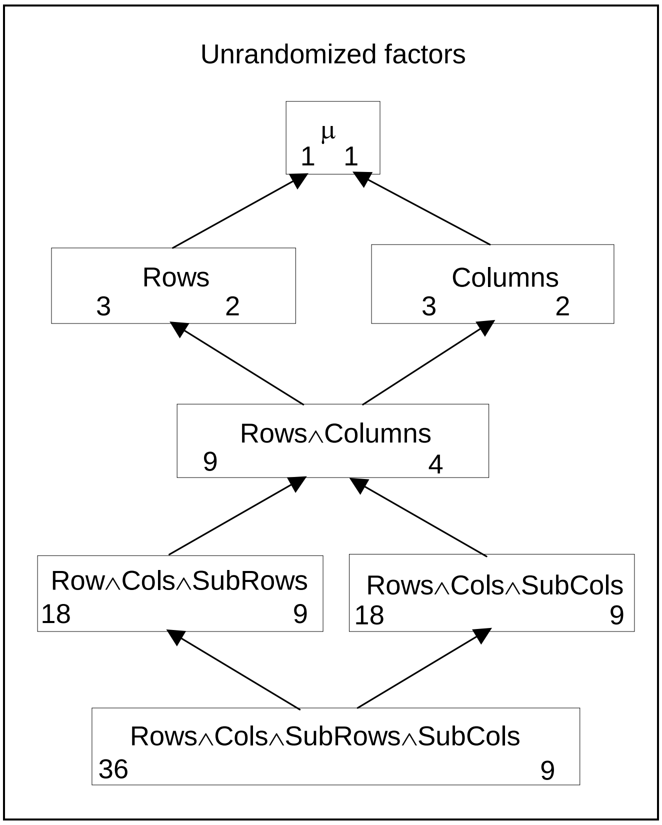
c) Sources derived from the structure formulae

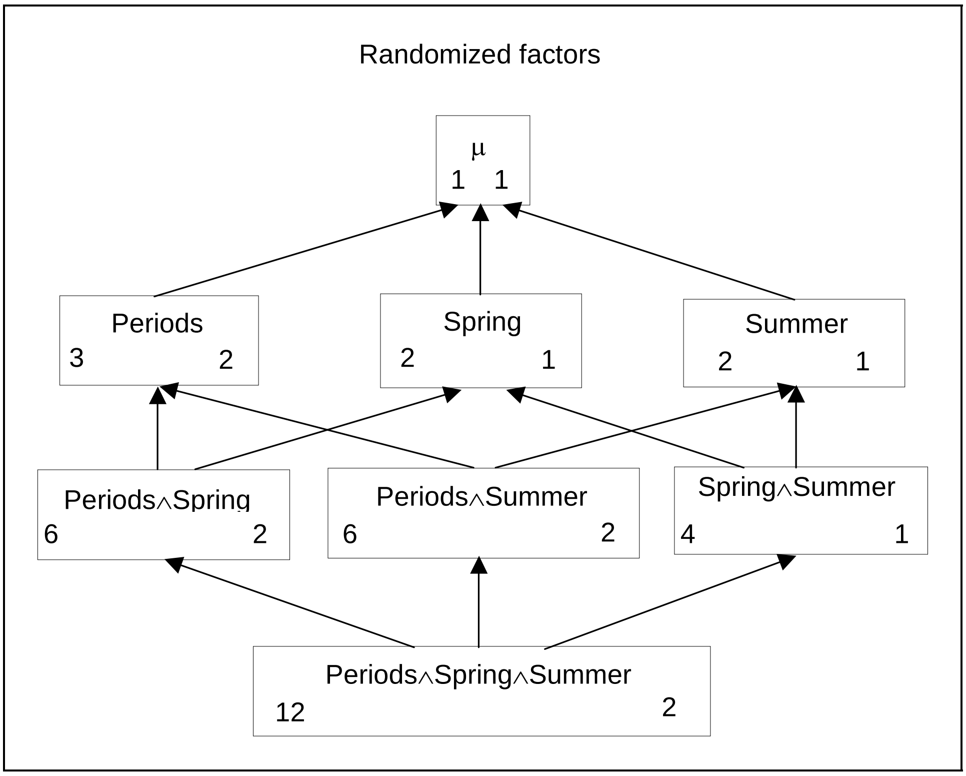
(Rows*Columns)/(Subrows*Subcolumns)
= (Rows + Columns + Rows#Columns)
/(Subrows + Subcolumns + Subrows#Subcolumns)
= Rows + Columns + Rows#Columns
+ Subrows[Rows∧Columns] + Subcolumns[Rows∧Columns]
+ Subrows#Subcolumns[Rows∧Columns]

and Periods*Spring*Summer
= (Periods + Spring + Periods#Spring)*Summer
= Periods + Spring + Periods#Spring
+ Periods#Summer + Spring#Summer + Periods#Spring#Summer

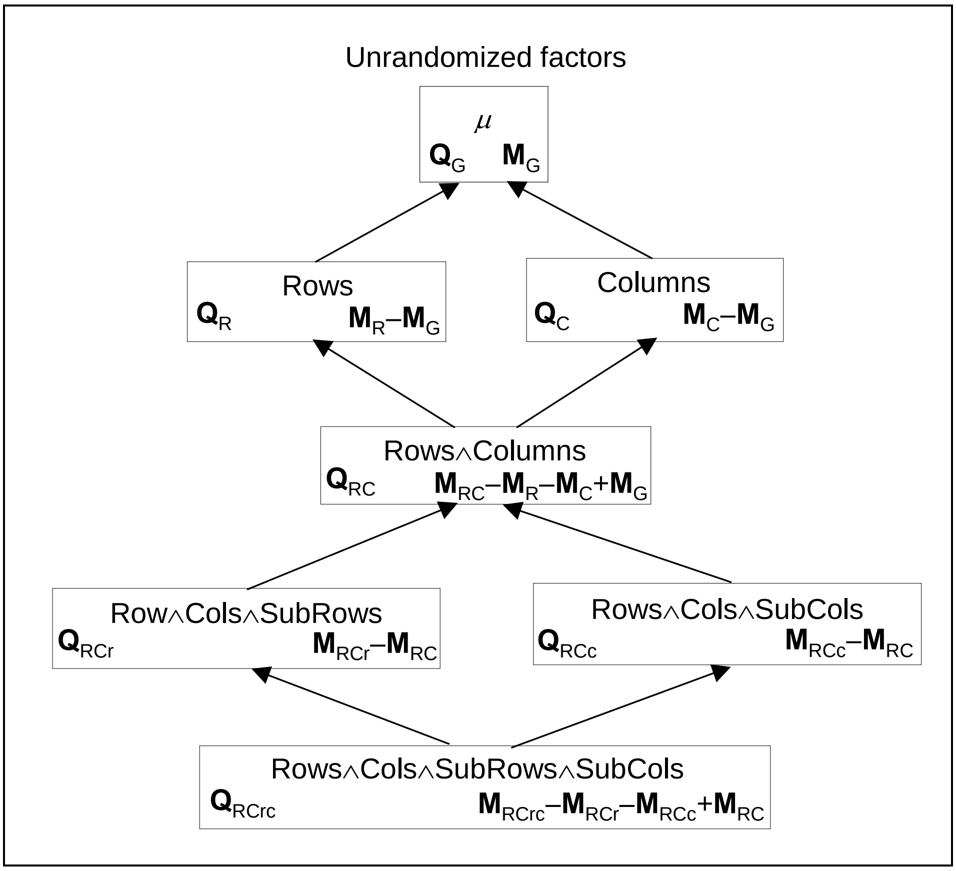
d) Degrees of freedom and sums of squares

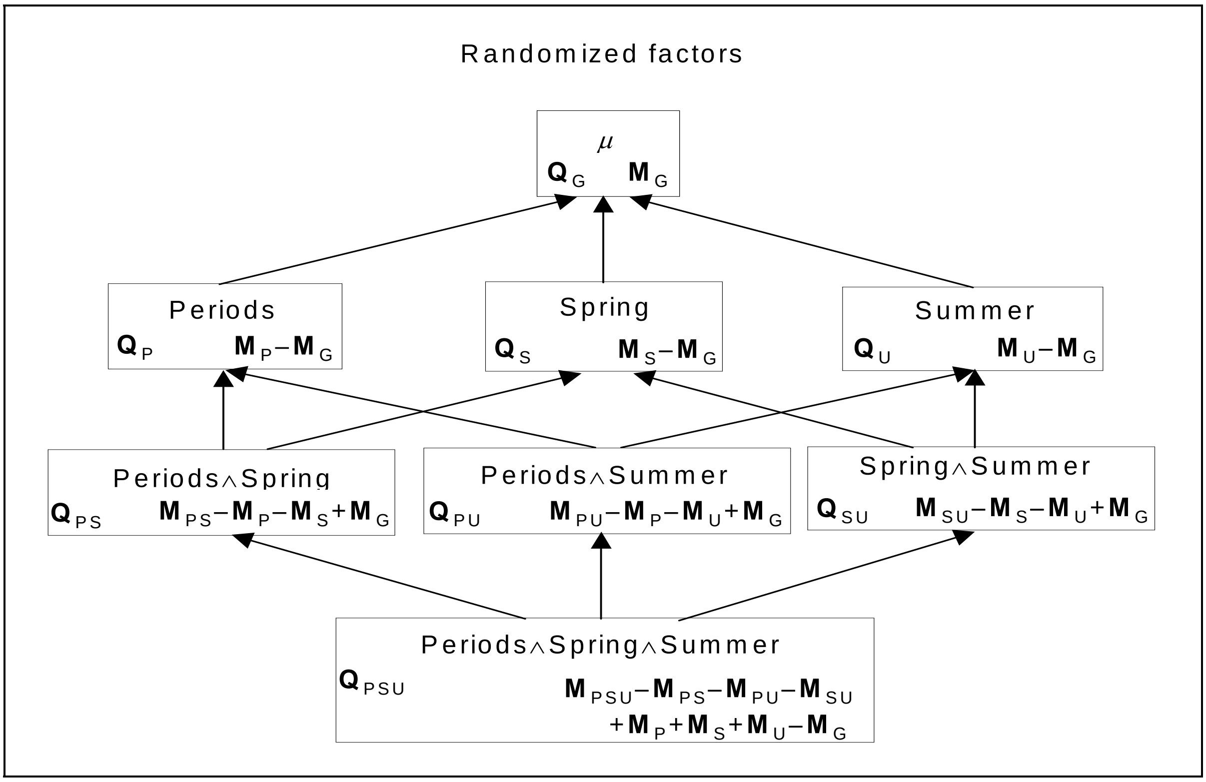
Hasse diagrams of generalized-factor marginalities, including the degrees of freedom, for the grazing experiment





Hasse diagrams of generalized-factor marginalities, including the Q and M matrices, for the grazing experiment





Note, that in working out the degrees of freedom for the sources from the randomized structure, the rule for a set of crossed factors can be used. That is, for each factor in the source, calculate the number of levels minus one and multiply these together.

e) The analysis of variance table

Enter the sources for the study, their degrees of freedom and quadratic forms, into the analysis of variance table below.

f) Maximal expectation and variation models

The models for this experiment, based on the unrandomized factors being random factors and the randomized factors being fixed factors, are:

ψ = E[Y] = Periods∧Spring∧Summer

and var[Y] = Rows + Columns + Rows∧Columns
+ Rows∧Columns∧Subrows + Rows∧Columns∧Subcolumns
+ Rows∧Columns∧Subrows∧Subcolumns

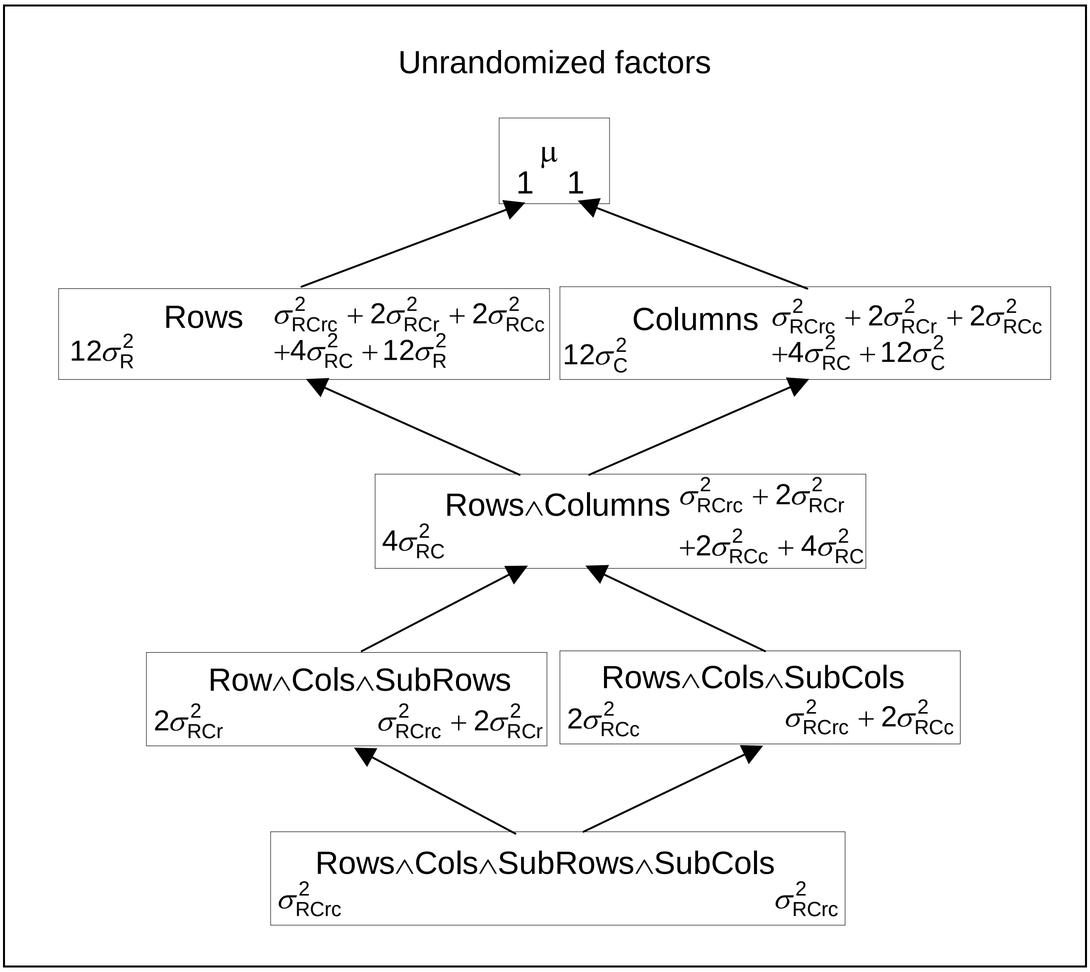
g) The expected mean squares.

The contributions of the randomized factors will all be of the form $q_F(\boldsymbol{\psi})$. The variance components and their multipliers are:

$$\sigma^2_{\mathrm{RCrc}},\ 2\sigma^2_{\mathrm{RCr}},\ 2\sigma^2_{\mathrm{RCc}}, 4\sigma^2_{\mathrm{RC}}, 12\sigma^2_{\mathrm{R}},\ 12\sigma^2_{\mathrm{C}}$$

The Hasse diagram, with the contributions to the expected mean squares of the unrandomized factors, is as follows:

Hasse diagrams of generalized-factor marginalities, including contributions of unrandomized factor to the E[MSq]s, for the grazing experiment



The analysis of variance table for this experiment has the following form:

Source	df	SSq	E[MSq]				
Rows	2	$\mathbf{Y}'\mathbf{Q}_{\mathrm{R}}\mathbf{Y}$	σ^2_{RCrc}	$+2\sigma^2_{\mathrm{RCr}}$	$+2\sigma^2_{\mathrm{RCc}}$	$+4\sigma^2_{\mathrm{RC}}$	$+12\sigma^2_{\mathrm{R}}$
Columns	2	$\mathbf{Y}'\mathbf{Q}_{\mathrm{C}}\mathbf{Y}$	σ^2_{RCrc}	$+2\sigma^2_{\mathrm{RCr}}$	$+2\sigma^2_{\mathrm{RCc}}$	$+4\sigma^2_{\mathrm{RC}}$	$+12\sigma^2_{\mathrm{C}}$
Rows#Columns	4	$\mathbf{Y}'\mathbf{Q}_{\mathrm{RC}}\mathbf{Y}$					
Periods	2	$\mathbf{Y}'\mathbf{Q}_{\mathrm{P}}\mathbf{Y}$	σ^2_{RCrc}	$+2\sigma^2_{\mathrm{RCr}}$	$+2\sigma^2_{\mathrm{RCc}}$	$+4\sigma^2_{\mathrm{RC}}$	$+q_{\mathrm{P}}(\psi)$
Residual	2	$\mathbf{Y}'\mathbf{Q}_{\mathrm{RC}_{\mathrm{Res}}}\mathbf{Y}$	σ^2_{RCrc}	$+2\sigma^2_{\mathrm{RCr}}$	$+2\sigma^2_{\mathrm{RCc}}$	$+4\sigma^2_{\mathrm{RC}}$	
Subrows[Rows∧Columns]	9	$\mathbf{Y}'\mathbf{Q}_{\mathrm{RCr}}\mathbf{Y}$					
Spring	1	$\mathbf{Y}'\mathbf{Q}_{\mathrm{S}}\mathbf{Y}$	σ^2_{RCrc}	$+2\sigma^2_{\mathrm{RCr}}$			$+q_{\mathrm{S}}(\psi)$
Periods#Spring	2	$\mathbf{Y}'\mathbf{Q}_{\mathrm{PS}}\mathbf{Y}$	σ^2_{RCrc}	$+2\sigma^2_{\mathrm{RCr}}$			$+q_{\mathrm{PS}}(\psi)$
Residual	6	$\mathbf{Y}'\mathbf{Q}_{\mathrm{RCr}_{\mathrm{Res}}}\mathbf{Y}$	σ^2_{RCrc}	$+2\sigma^2_{\mathrm{RCr}}$			
Subcols[Rows∧Columns]	9	$\mathbf{Y}'\mathbf{Q}_{\mathrm{RCc}}\mathbf{Y}$					
Summer	1	$\mathbf{Y}'\mathbf{Q}_{\mathrm{U}}\mathbf{Y}$	σ^2_{RCrc}		$+2\sigma^2_{\mathrm{RCc}}$		$+q_{\mathrm{U}}(\psi)$
Periods#Summer	2	$\mathbf{Y}'\mathbf{Q}_{\mathrm{PU}}\mathbf{Y}$	σ^2_{RCrc}		$+2\sigma^2_{\mathrm{RCc}}$		$+q_{\mathrm{PU}}(\psi)$
Residual	6	$\mathbf{Y}'\mathbf{Q}_{\mathrm{RCc}_{\mathrm{Res}}}\mathbf{Y}$	σ^2_{RCrc}		$+2\sigma^2_{\mathrm{RCc}}$		
Subrows#Subcols[Rows∧Cols]	9	$\mathbf{Y}'\mathbf{Q}_{\mathrm{RCrc}}\mathbf{Y}$					
Spring#Summer	1	$\mathbf{Y}'\mathbf{Q}_{\mathrm{SU}}\mathbf{Y}$	σ^2_{RCrc}				$+q_{\mathrm{SU}}(\psi)$
Periods#Spring#Summer	2	$\mathbf{Y}'\mathbf{Q}_{\mathrm{PSU}}\mathbf{Y}$	σ^2_{RCrc}				$+q_{\mathrm{PSU}}(\psi)$
Residual	6	$\mathbf{Y}'\mathbf{Q}_{\mathrm{RCrc}_{\mathrm{Res}}}\mathbf{Y}$	σ^2_{RCrc}				

This design has problems in that the Residuals have low degrees of freedom. An alternative design would be to assign the four Spring-Summer combinations completely at random to the four subplots in each plot; this would have resulted in more degrees of freedom for the subplot residual. However, this would halve the size of the unit to which the Spring and Summer treatments were applied. It may be that practical restrictions on the size of plots prevented the use of the smaller units and forced the adoption of the design presented here.

The data, in field order, is given in the table below.

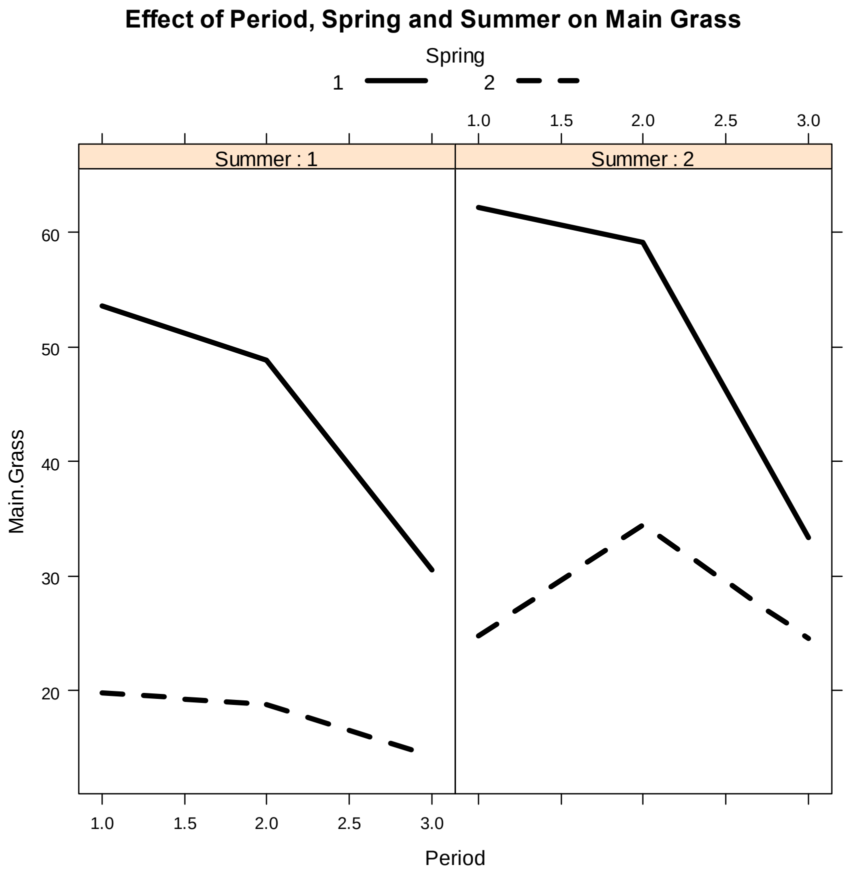
Results for the grazing experiment

			SubRows			
			1		2	
Row	Column	SubColumn	1	2	1	2
	1		12.5	26.2	33.4	44.2
1	2		59.2	47.6	49.9	15.8
	3		55.0	35.9	27.3	18.3
	1		56.2	52.3	27.5	25.1
2	2		67.7	62.2	24.1	27.5
	3		28.0	29.4	19.5	29.9
	1		57.2	69.5	16.9	19.5
3	2		30.3	26.6	11.0	17.6
	3		61.9	46.5	26.2	15.4

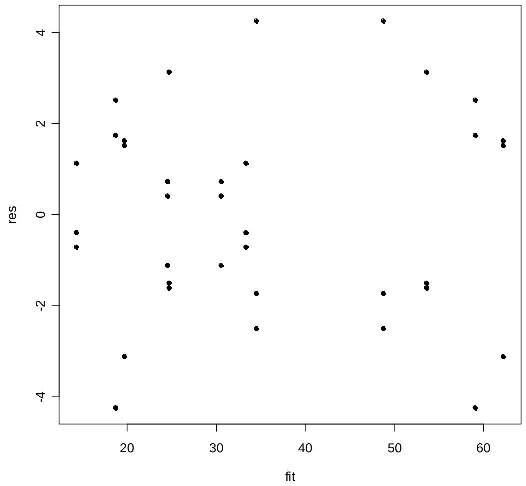
This data has been analyzed in R with quadratic trends being fitted to the Period means. Note that all terms involving the quantitative factor Period have been included in the `list` for the `split` argument. The output is given below.

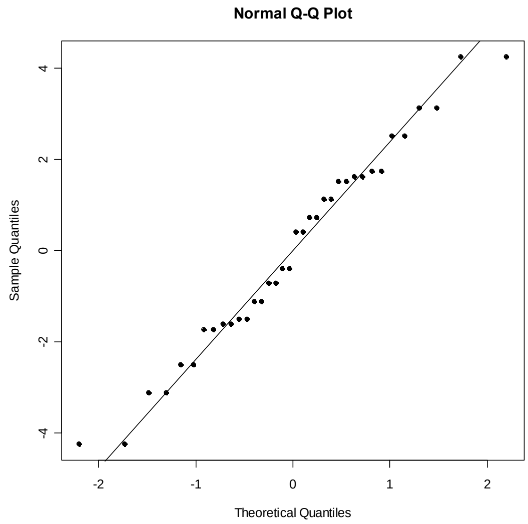
```
> attach(SPLGrass.dat)
> interaction.ABC.plot(Main.Grass, x.factor=Period, trace.factor=Summer,
+                      groups.factor=Spring, data=SPLGrass.dat,
+                      title="Effect of Period, Spring and Summer on Main Grass")
> Period.lev <- c(3, 9, 18)
> SPLGrass.dat$Period <- ordered(SPLGrass.dat$Period, levels=Period.lev)
> contrasts(SPLGrass.dat$Period) <- contr.poly(3, scores=Period.lev)
> contrasts(SPLGrass.dat$Period)
           .L         .Q
3  -0.65561007  0.4866643
9  -0.09365858 -0.8111071
18  0.74926865  0.3244428
>
> SPLGrass.aov <- aov(Main.Grass ~ Period * Spring * Summer +
+                       Error((Rows*Columns)/(SubRows*SubColumns)), SPLGrass.dat)
> summary(SPLGrass.aov, split = list(Period = list(L = 1, Q = 2),
+                                    "Period:Spring" = list(L = 1, Q = 2),
+                                    "Period:Summer" = list(L = 1, Q = 2),
+                                    "Period:Spring:Summer" = list(L = 1, Q = 2)))

Error: Rows
          Df Sum Sq Mean Sq F value Pr(>F)
Residuals  2 107.62   53.81

Error: Columns
          Df  Sum Sq Mean Sq F value Pr(>F)
Residuals  2 121.202  60.601

Error: Rows:Columns
            Df  Sum Sq Mean Sq F value Pr(>F)
Period       2 1677.43  838.72  7.8103 0.1135
  Period: L  1 1397.15 1397.15 13.0107 0.0690
  Period: Q  1  280.28  280.28  2.6100 0.2476
Residuals    2  214.77  107.39

Error: Rows:Columns:SubRows
                   Df Sum Sq Mean Sq F value    Pr(>F)
Spring              1 5697.7  5697.7 71.5247 0.0001493
Period:Spring       2  822.2   411.1  5.1603 0.0496865
  Period:Spring: L  1  820.6   820.6 10.3008 0.0183791
  Period:Spring: Q  1    1.6     1.6  0.0199 0.8923796
Residuals           6  478.0    79.7

Error: Rows:Columns:SubColumns
                   Df Sum Sq Mean Sq F value  Pr(>F)
Summer              1 696.08  696.08 11.3621 0.01503
Period:Summer       2  80.98   40.49  0.6609 0.55030
  Period:Summer: L  1   1.89    1.89  0.0309 0.86622
  Period:Summer: Q  1  79.08   79.08  1.2909 0.29922
Residuals           6 367.58   61.26

Error: Rows:Columns:SubRows:SubColumns
                          Df  Sum Sq Mean Sq F value Pr(>F)
Spring:Summer              1  21.314  21.314  0.7236 0.4276
Period:Spring:Summer       2  52.071  26.035  0.8839 0.4609
  Period:Spring:Summer: L  1  41.233  41.233  1.3998 0.2815
  Period:Spring:Summer: Q  1  10.838  10.838  0.3679 0.5664
Residuals                  6 176.733  29.456
> #
> # Diagnostic checking
> #
> tukey.1df(SPLGrass.aov, SPLGrass.dat, "Rows:Columns:SubRows:SubColumns")
```

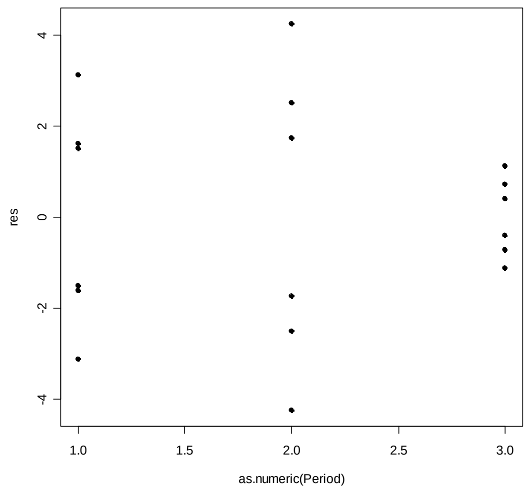
```
** Warning - there appears to be extremely little non-linear variation so that
   the values for Tukey.SS are unstable and the results below may be unreliable.
   Only use if at least two non-interacting factors above the same Residual
   in the analysis.
$Tukey.SS
[1] 73.92478

$Tukey.F
[1] 3.595264

$Tukey.p
[1] 0.1164419

$Devn.SS
[1] 102.8085

> res <- resid.errors(SPLGrass.aov)
> fit <- fitted.errors(SPLGrass.aov)
> plot(fit, res, pch=16)
>
> qqnorm(res, pch=16)
> qqline(res)
> plot(as.numeric(Period), res, pch=16)
> plot(as.numeric(Spring), res, pch=16)
> plot(as.numeric(Summer), res, pch=16)
```

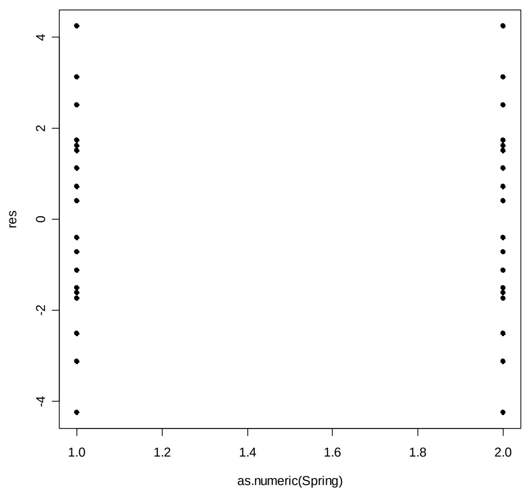


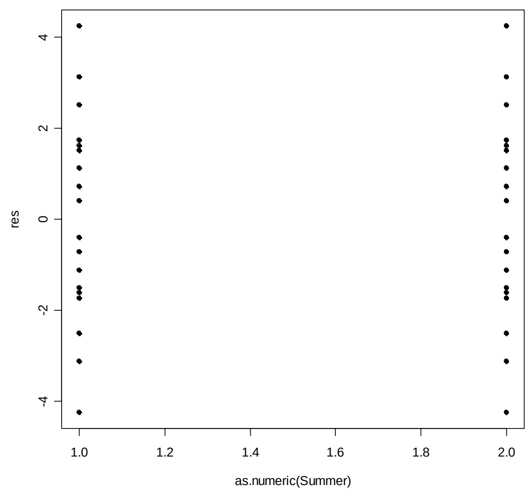
Maybe there is an interaction between Period and the other two factors.











The diagnostic checking indicates that the assumptions have been met because a) the residual-versus-fitted-values plot has fairly even distribution of points as one goes from left to right, b) the residuals-versus-factors plots show similar spread for the levels of each of the factors, c) the normal probability plot displays a roughly straight line pattern and c) Tukey's test-for-nonadditivity is not significant (indeed there is so little variation that a warning message has been output — in this case the warning message is incorrect).

From the first analysis of variance table in the output we determine the fitted model that best describes the data. Firstly, the Quadratic term for Spring#Summer#Periods is non-significant as is the Lin source for this three-factor interaction so that there is no three-factor interaction. For the two-factor interactions only the Lin term for Spring#Periods is significant. Given that this term involving both Spring and Periods is significant, only the main effect for Summer has to be examined. The F value for Summer is significant. Hence, the fitted expectation model is:

$$\psi = \mathrm{E}[\mathrm{Y}] = \mathrm{Summer} + \mathrm{Spring}\wedge\mathrm{Periods}_{\mathrm{Lin}}$$

The interpretation of this model is that there is a linear trend in the % of the principal grass over the Periods that differs between the two numbers of spring grazing cycles and that there is an overall difference between the two numbers of summer grazing cycles that is independent of the other two treatment factors, Periods and Spring.

The trends in the period means for the two numbers of spring cycles are illustrated in the interaction diagram given above.

The equations for the fitted straight lines describing the trend in the period means for the different numbers of spring grazing cycles can only be obtained using regression. This has been done at the end of the output and the fitted equations are,

```
> #
> # Fitted equation
> #
> Pe <- as.numeric(as.vector(Period))
> SPLGrass.lm <- lm(Main.Grass ~ Spring/Pe)
> coef(SPLGrass.lm)
(Intercept)     Spring4  Spring2:Pe  Spring4:Pe
 65.8008772 -40.6508772  -1.7850877  -0.2361111
```

$$\text{For 2 spring cycles,} \%\mathit{MainGrass} = 65.80 - 1.785\,\mathit{Period}$$
$$\text{For 4 spring cycles,} \%\mathit{MainGrass} = 25.15 - 0.236\,\mathit{Period}$$

That is, the percentage of the main grass decreases as the number of periods of grazing increases, but the decrease is greater for two spring cycles than for 4 spring cycles.

The Summer means are given in the table below.

```
> #
> # tables of means
> #
> SPLGrass.means <- model.tables(SPLGrass.aov, type="means")
> SPLGrass.means$tables$Summer
Summer
       2        4
30.97222 39.76667
```

Summer table of means

Summer	
2	4
30.97	39.77

Clearly, 4 cycles in summer, irrespective of the number of Spring cycles and the length of grazing, result in a larger percentage of the main grass surviving.

IX.E Summary

In this chapter we have:

- described how to design factorial experiments using the split-plot principle and outlined when split-plot designs should be employed;
- given a detailed description of the standard split-plot design;
- used the rules from chapter VI, *Determining the analysis of variance table*, to formulate the ANOVA hypothesis test for choosing between expectation models in split-plot experiment and outlined the test;
  - the partition of the total sums of squares was given with the sums of squares expressed as the sums of squares of the elements of vectors and as quadratic forms where the matrices of the quadratic forms, **Q** matrices, are symmetric idempotents;
  - the expected mean squares under the alternative expectation models are used to justify the choice of F test statistic;
- shown how to obtain a layout and the analysis of variance in R;
- described the modifications to multiple comparison procedures required in the context of split-plot experiments;
- discussed the implications of systematically applying treatments to main plots and of not replicating the main-plot treatments;
- outlined the analysis of a more complex split-plot experiment.

IX.F Exercises

IX.1 An experiment was designed to evaluate alternative processes for pigment dispersion in aqueous media. Six solutions of each of two specific dispersion liquids were prepared. The pigment Phthalocyanine Blue was dispersed in half of the solutions for each liquid, chosen at random, using a roller mill and the other half for each liquid using a Manton-Gaulin mill.

The 12 solutions were then mixed by weight with a polystyrene paint. This was done within one half-hour of mixing for half of each solution and after a week for the other half of the solution. The degree of dispersion was then measured by the reflectance in a spectrophotometer at 660μ. The data are as follows:

Liquid	Solut	Halfsol	Mill	Time	Reflect
1	1	1	1	2	44.0
		2		1	43.5
	2	1	2	2	44.0
		2		1	44.5
	3	1	1	2	43.0
		2		1	42.5
	4	1	2	2	45.0
		2		1	45.5
	5	1	2	1	45.0
		2		2	45.5
	6	1	1	2	44.5
		2		1	44.0
2	1	1	2	1	44.5
		2		2	43.5
	2	1	1	2	43.5
		2		1	43.0
	3	1	1	1	43.0
		2		2	43.5
	4	1	2	1	43.0
		2		2	43.5
	5	1	2	1	43.0
		2		2	43.0
	6	1	1	1	44.0
		2		2	44.0

What are the features of the study?

1. Observational unit ______________________________
2. Response variable ______________________________
3. Unrandomized factors ______________________________
4. Randomized factors ______________________________
5. Type of study ______________________________

What is the experimental structure for this experiment?

Structure	Formula
unrandomized	
randomized	

What are the sources derived from the experimental structure? Write out the Hasse diagrams that include the degrees of freedom for each structure formula. Also those that include the **M** and **Q** matrices.

What are the expectation and variation models and the contributions of the unrandomized and randomized factors to the expected mean squares based on Solut and Halfsol being random and the rest of the factors being fixed?

Write down the analysis of variance table, including the degrees of freedom, sums of squares and the expected mean squares for the lines in it.

Source	df	E[MSq]
Total		

Analyze the data using R, including diagnostic checking and the examination of treatment differences.

IX.2 An experiment on celery is to be conducted to investigate the effect on the yield of three methods of seedling propagation, two levels of nutrient and four harvest dates. The six combinations of propagation methods and nutrients are to be applied to main plots using a completely randomized design with three replicates of each treatment combination. The harvest dates are to be randomized to the four subplots within each main plot.

What are the features of the study?

1. Observational unit ____________________
2. Response variable ____________________
3. Unrandomized factors ____________________
4. Randomized factors ____________________
5. Type of study ____________________

What is the experimental structure for this experiment?

Structure	Formula
unrandomized	
randomized	

What are the terms derived from the experimental structure? Write out the Hasse diagram that includes the degrees of freedom for the first structure formula and those that include the **M** and **Q** matrices for both structure formulae.

What are the expectation and variation models based on all unrandomized factors being random and all randomized factors being fixed?

Write down the analysis of variance table, including the degrees of freedom, the sums of squares and the expected mean squares for the lines in it.

Source	df	E[MSq]
Total		

Obtain a randomized layout for the experiment in R using the seed 445.

IX.3 An experiment was conducted to look at the effect of irrigation and canopy type on the number of shoot per node of grape vines. The area consisted of 3 rows each containing 56 vines. Each row formed a rep and was divided into 2 halves which are called Columns. Each column was divided into 2 plots to which the irrigation treatments (no irrigation, irrigation) were randomized. Each plot was subdivided into 2 subplots to which 2 canopy treatments were randomized. Each subplot consisted of 7 vines. The layout for the experiment is given in the table below.

Layout for a vineyard experiment

	Columns	1				2			
	Plots	1		2		1		2	
Reps	Subplots	1	2	1	2	1	2	1	2
1	Irrigation	Yes		No		No		Yes	
	Canopy	L	S	S	L	S	L	L	S
2	Irrigation	No		Yes		Yes		No	
	Canopy	L	S	L	S	S	L	S	L
3	Irrigation	No		Yes		No		Yes	
	Canopy	L	S	L	S	L	S	S	L

S = Severe Pruning, Low Trellis
L = Light Pruning, High Trellis

The number of shoots and the number of nodes was measured on each vine and the resulting data, in randomized order, are given in the table at the end of the question.

What are the features of the study?

1. Observational unit ______________________
2. Response variable ______________________
3. Unrandomized factors ______________________
4. Randomized factors ______________________
5. Type of study ______________________

What is the experimental structure for this experiment?

Structure	Formula
unrandomized	
randomized	

What are the terms derived from the experimental structure? Write out the Hasse diagrams for the first structure formula only. Use the rule for completely crossed structures to get the degrees of freedom for the second structure formula.

What are the expectation and variation models based on all unrandomized factors being random and all randomized factors being fixed?

Write down the analysis of variance table, including the sums of squares and the expected mean squares for the lines in it.

Source	df	E[MSq]
Total		

The data and factors have been saved in *SplRejuv.dat.rda* that is available from the course's web site. Use R to analyse the shoots per node (ShotPNod) data, including diagnostic checking. What are the fitted models?

Results for a vineyard experiment

				Vine													
				1		2		3		4		5		6		7	
Reps	Cols	Plot	SubP	Sh	N	Sh	N	Sh	N	Sh	N	Sh	N	Sh	N	Sh	N
1	1	1	1	46	36	42	36	40	36	46	36	37	36	39	36	39	36
			2	55	52	64	72	69	72	60	72	71	72	60	72	67	72
		2	1	41	36	47	36	48	36	45	36	40	36	41	36	48	36
			2	66	72	71	72	50	72	56	72	52	72	56	62	65	62
	2	1	1	51	36	37	36	33	36	39	36	28	26	39	36	38	36
			2	60	72	43	52	56	72	62	72	55	62	50	50	53	52
		2	1	50	36	39	36	47	36	39	36	47	36	32	36	34	36
			2	55	72	50	72	66	72	59	62	59	72	56	62	57	72
2	1	1	1	53	36	40	36	50	36	36	36	38	36	49	36	40	36
			2	68	62	54	62	57	72	68	72	67	72	65	72	62	72
		2	1	50	36	41	36	47	36	58	36	51	36	56	36	47	36
			2	55	72	50	72	66	72	71	72	54	62	55	72	45	62
	2	1	1	45	36	45	36	52	36	52	36	52	36	44	36	47	36
			2	67	62	52	62	66	72	57	62	59	62	48	62	49	62
		2	1	59	72	71	67	53	62	68	62	49	52	58	62	59	62
			2	48	36	37	36	38	26	48	36	33	36	43	36	40	36
3	1	1	1	40	36	46	36	42	36	47	36	50	36	49	36	49	36
			2	82	72	57	72	68	72	68	72	76	60	60	36	73	68
		2	1	36	38	50	36	45	36	44	36	35	36	61	36	51	36
			2	63	72	76	59	62	63	90	74	65	72	52	45	78	72
	2	1	1	40	36	49	36	48	36	53	36	49	36	32	36	36	28
			2	63	69	61	52	52	61	65	72	80	62	70	65	67	69
		2	1	47	36	53	36	54	36	47	36	41	36	53	36	48	36
			2	64	72	70	82	73	72	64	72	72	72	75	72	76	72