STATISTICAL MODELLING

PRACTICAL IX
SOLUTIONS

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 - a half-solution

   Variables (incl. factors) are?
   Ans. Reflectance, Liquid, Solution, Halfsol, Mill, Time

2. Response variable - Reflectance
3. Unrandomized factors - Liquid, Solution, Halfsol
4. Randomized factors - Mill, Time
5. Type of study - Split-plot with mainplots in an RCBD and subplots completely randomized

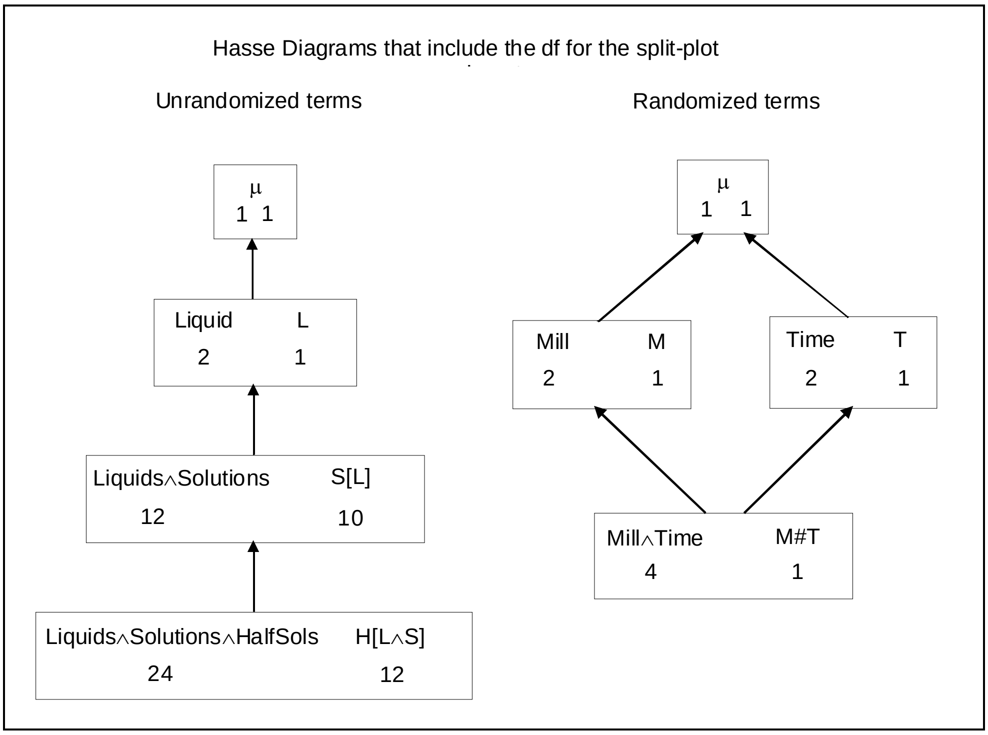
What is the experimental structure for this experiment?

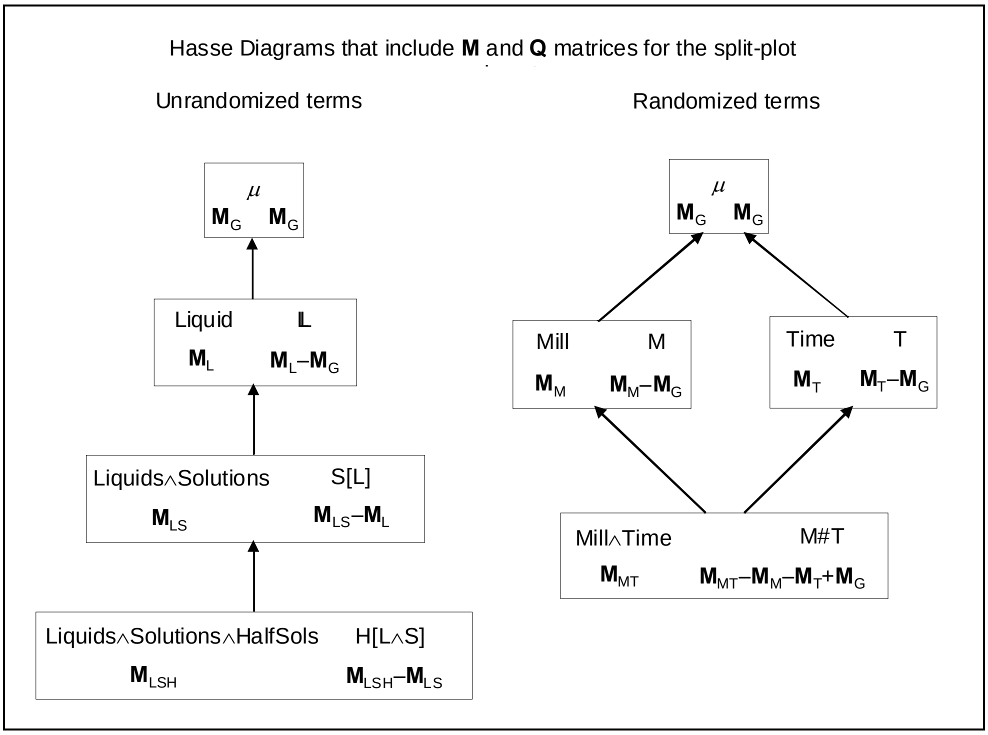
Structure	Formula
unrandomized	*2* Liquids/*6* Solutions/*2* Halfsol
randomized	*2* Mills**2* Time

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.

Liquids/Solutions/Halfsol = (Liquids+Solutions[Liquids])/Halfsol
= Liquids+Solutions[Liquids]
+Halfsol[Liquids Solutions]

*Mills*Time = Mills+Time+Mills#Time*



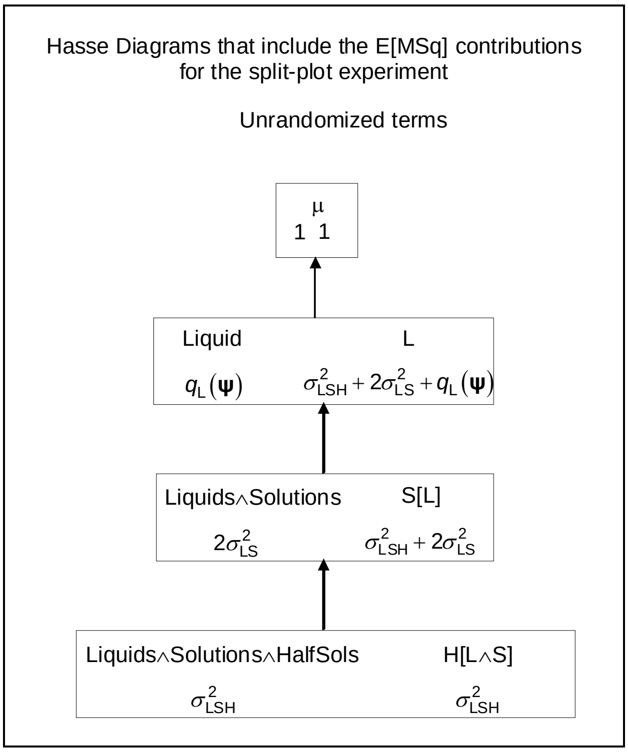


Hasse Diagrams that include **M** and **Q** matrices for the split-plot

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?

ψ = *E[Y] = Liquids + Mills∧Time and*
Var[Y] = Liquids ∧Solutions+ Liquids∧Solutions ∧Halfsol

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:*



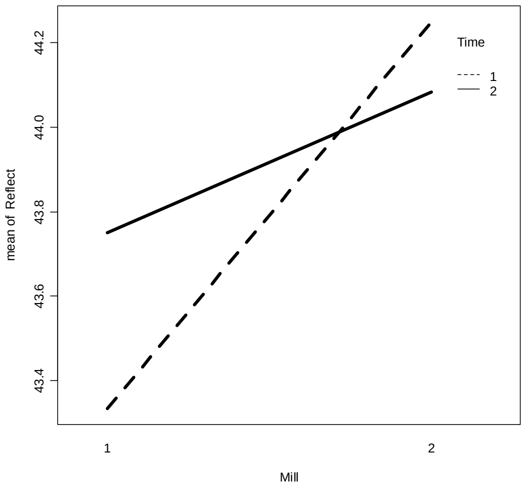
Write down the analysis of variance table, including the degrees of freedom, sums of squares and the expected mean squares for the lines in it.

Source	df	SSq	E[MSq]
Liquids	1	$\mathbf{Y}'\mathbf{Q}_{\mathrm{L}}\mathbf{Y}$	$\sigma^2_{\mathrm{LSH}} + 2\sigma^2_{\mathrm{LS}} + q_{\mathrm{L}}(\boldsymbol{\psi})$
Solutions[Liquids]	10	$\mathbf{Y}'\mathbf{Q}_{\mathrm{LS}}\mathbf{Y}$	
Mills	1	$\mathbf{Y}'\mathbf{Q}_{\mathrm{M}}\mathbf{Y}$	$\sigma^2_{\mathrm{LSH}} + 2\sigma^2_{\mathrm{LS}} + q_{\mathrm{M}}(\boldsymbol{\psi})$
Residual	9	$\mathbf{Y}'\mathbf{Q}_{\mathrm{LS_{Res}}}\mathbf{Y}$	$\sigma^2_{\mathrm{LSH}} + 2\sigma^2_{\mathrm{LS}}$
Halfsol[Liquids∧Solutions]	12	$\mathbf{Y}'\mathbf{Q}_{\mathrm{LSH}}\mathbf{Y}$	
Time	1	$\mathbf{Y}'\mathbf{Q}_{\mathrm{T}}\mathbf{Y}$	$\sigma^2_{\mathrm{LSH}} + q_{\mathrm{T}}(\boldsymbol{\psi})$
Mills#Time	1	$\mathbf{Y}'\mathbf{Q}_{\mathrm{MT}}\mathbf{Y}$	$\sigma^2_{\mathrm{LSH}} + q_{\mathrm{MT}}(\boldsymbol{\psi})$
Residual	10	$\mathbf{Y}'\mathbf{Q}_{\mathrm{LSH_{Res}}}\mathbf{Y}$	σ^2_{LSH}
Total	23	$\mathbf{Y}'\mathbf{Q}_{\mathrm{U}}\mathbf{Y}$	

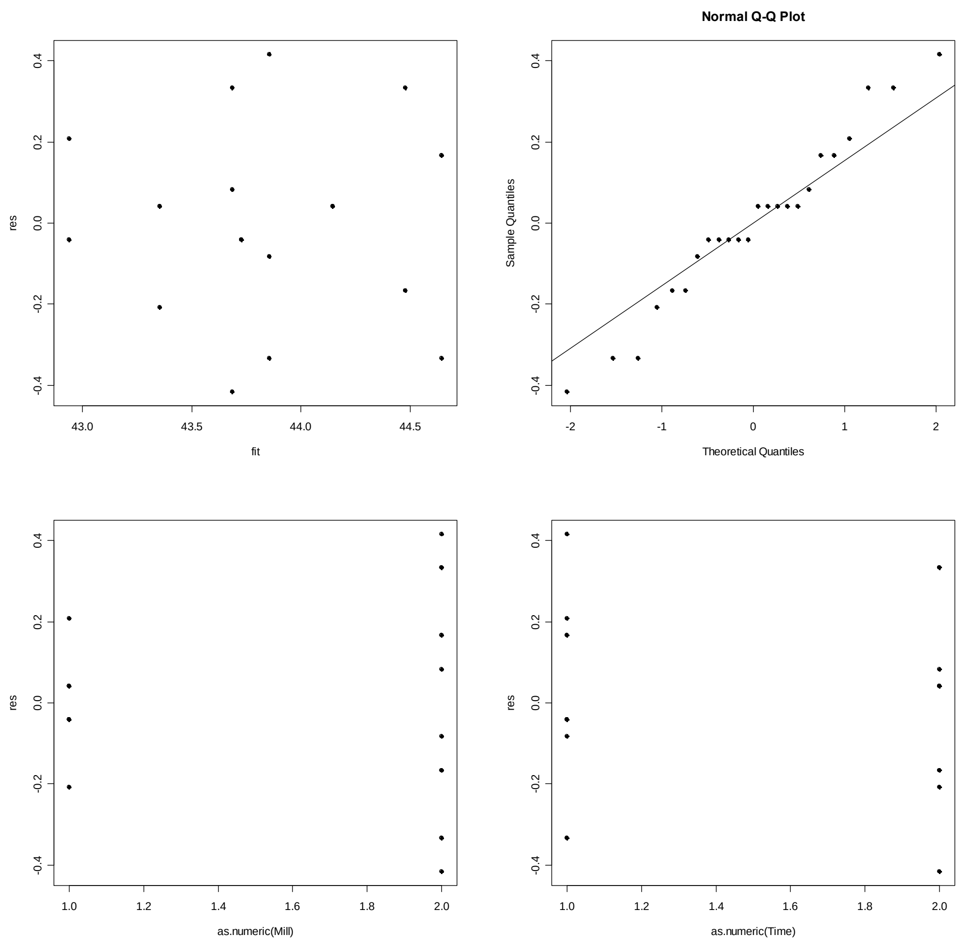
The R output file for computing the analysis of variance table is:

```
> attach(SPLPigm.dat)
> SPLPigm.dat
   Liquid Solution HalfSoln Mill Time Reflect
1       1        1        1    1    2    44.0
2       1        1        2    1    1    43.5
3       1        2        1    2    2    44.0
4       1        2        2    2    1    44.5
5       1        3        1    1    2    43.0
6       1        3        2    1    1    42.5
7       1        4        1    2    2    45.0
8       1        4        2    2    1    45.5
9       1        5        1    2    1    45.0
10      1        5        2    2    2    45.5
11      1        6        1    1    2    44.5
12      1        6        2    1    1    44.0
13      2        1        1    2    1    44.5
14      2        1        2    2    2    43.5
15      2        2        1    1    2    43.5
16      2        2        2    1    1    43.0
17      2        3        1    1    1    43.0
18      2        3        2    1    2    43.5
19      2        4        1    2    1    43.0
20      2        4        2    2    2    43.5
21      2        5        1    2    1    43.0
22      2        5        2    2    2    43.0
23      2        6        1    1    1    44.0
24      2        6        2    1    2    44.0
> interaction.plot(Mill, Time, Reflect, lwd=4)
> SPLPigm.aov <- aov(Reflect ~ Liquid + Mill * Time +
+                     Error(Liquid/Solution/HalfSoln), SPLPigm.dat)
> summary(SPLPigm.aov)

Error: Liquid
       Df Sum Sq Mean Sq
Liquid  1 3.7604  3.7604

Error: Liquid:Solution
          Df Sum Sq Mean Sq F value Pr(>F)
Mill       1 2.3437  2.3437  2.4786 0.1499
Residuals  9 8.5104  0.9456

Error: Liquid:Solution:HalfSoln
          Df  Sum Sq Mean Sq F value  Pr(>F)
Time       1 0.09375 0.09375  0.9184 0.36050
Mill:Time  1 0.51042 0.51042  5.0000 0.04933
Residuals 10 1.02083 0.10208
> #
> # Diagnostic checking
> #
> tukey.1df(SPLPigm.aov, SPLPigm.dat, "Liquid:Solution:HalfSoln")
$Tukey.SS
[1] 0.01795977

$Tukey.F
[1] 0.1611748

$Tukey.p
[1] 0.6974428

$Devn.SS
[1] 1.002874
> res <- resid.errors(SPLPigm.aov)
> fit <- fitted.errors(SPLPigm.aov)
> plot(fit, res, pch=16)
> qqnorm(res, pch=16)
> qqline(res)
> plot(as.numeric(Mill), res, pch=16)
> plot(as.numeric(Time), res, pch=16)
```

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

43.85417

 Liquid
Liquid
    1     2
44.25 43.46

 Mill
Mill
    1     2
43.54 44.17

 Time
Time
    1     2
43.79 43.92

 Mill:Time
    Time
Mill 1     2
   1 43.33 43.75
   2 44.25 44.08
> qtukey(0.95, 4, 9)
[1] 4.41489
```



Looks like there is an interaction between Mill and Time.



Source	df	MSq	E[MSQ]	F	Prob
Liquids	1	3.760	$\sigma^2_{LSH} + 2\sigma^2_{LS} + q_L(\boldsymbol{\psi})$	3.98	0.0772
Solutions[Liquids]	10				
Mills	1	2.344	$\sigma^2_{LSH} + 2\sigma^2_{LS} + q_M(\boldsymbol{\psi})$	2.48	0.1498
Residual	9	0.946	$\sigma^2_{LSH} + 2\sigma^2_{LS}$	9.26	0.0009
Halfsol[Liquids Solutions]	12				
Time	1	0.094	$\sigma^2_{LSH} + q_T(\boldsymbol{\psi})$	0.92	0.3601
Time#Mills	1	0.510	$\sigma^2_{LSH} + q_{MT}(\boldsymbol{\psi})$	5.00	0.0493
Residual	10	0.102	σ^2_{LSH}		
Non-additivity	1	0.018		0.16	0.6974
Deviations	9	0.111			
Total	23				

The residual-versus-fitted-values plots and the normal probability plot appear to be satisfactory. Tukey's test for nonadditivity is not significant. However, there is some evidence of a difference in variability between the two Mills.

In this analysis, the Time#Mills interaction is significant so that the expectation model that appears to best describe the data is $\psi = \text{E}[\text{Y}] = \text{Time}\wedge\text{Mills}$. *Also, the variability between solutions is greater then between half-solutions taken from the same solution.*

		Mill		
		1	2	
Time	1	43.33	44.25	HSD(5%) = 0.58
	2	43.75	44.08	(same Mill)

However, in this experiment, interactions between the factor Liquids and the randomized factors are likely to be of interest to the experimenter. One could revise the experimental structure to take this into account.

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 – a subplot
2. Response variable – Yield
3. Unrandomized factors – MainPlots, Subplots
4. Randomized factors – Propagation, Nutrient, Harvest
5. Type of study – Split-plot with main-plots in a CRD and subplots completely randomized

What is the experimental structure for this experiment?

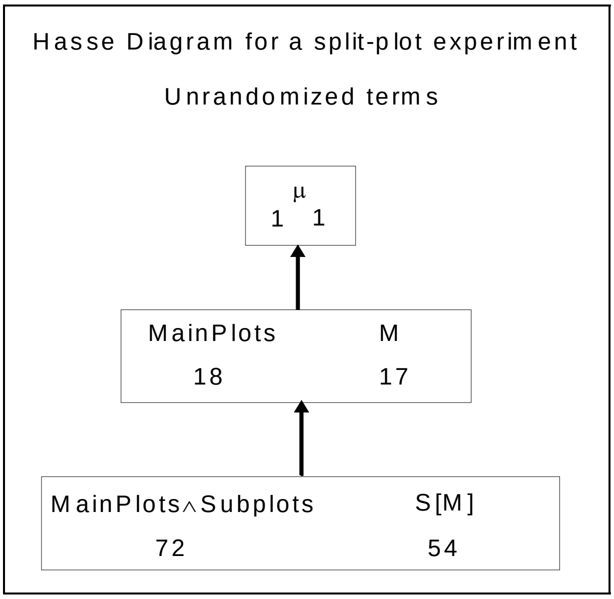
Structure	Formula
unrandomized	*18* MainPlots/*4* Subplots
randomized	*3* Propagation**2* Nutrient**4* Harvests

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.

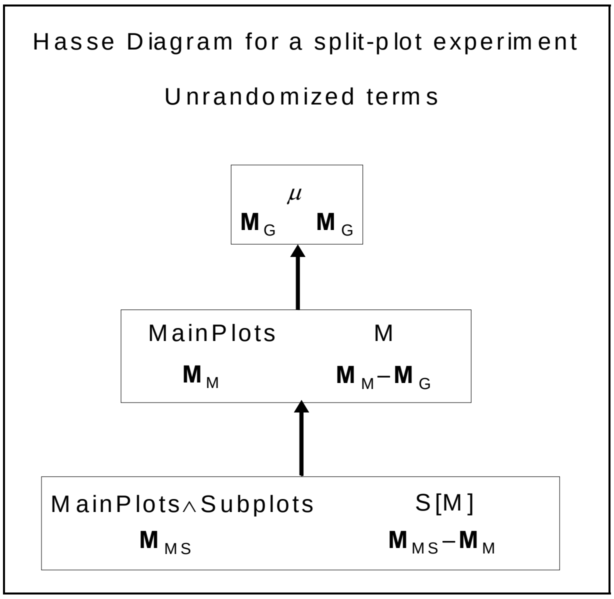
Plots/Subplots = MainPlots + Subplots[MainPlots]

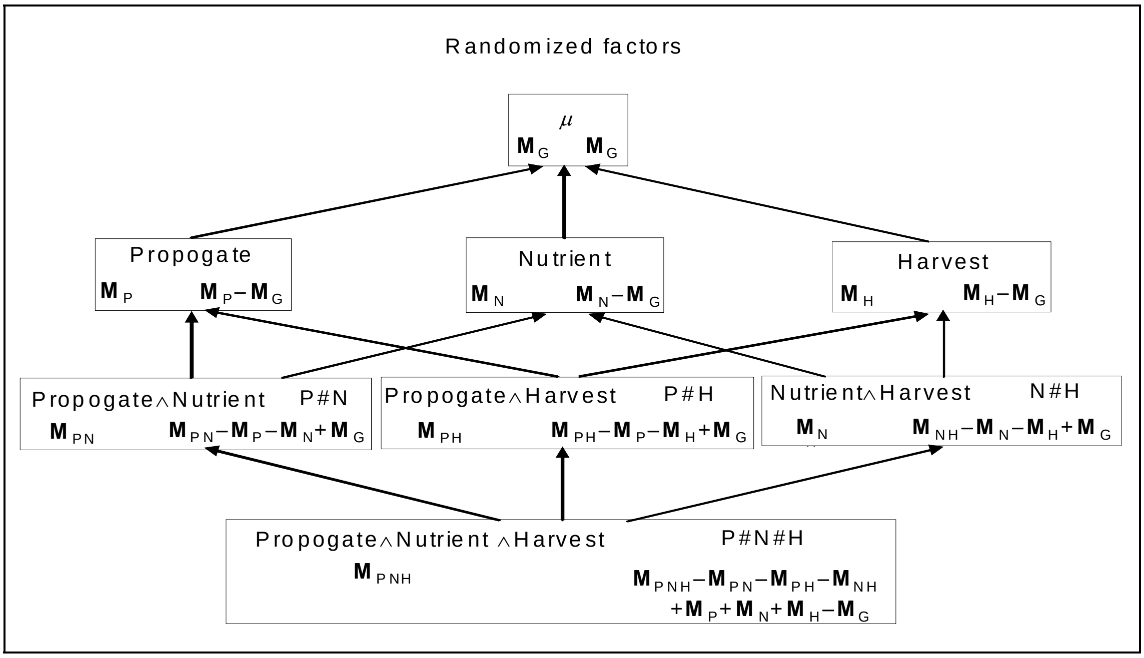
*Propagation*Nutrient*Harvests*
*= (Propagation + Nutrient + Propagation#Nutrient) *Harvests*
= Propagation + Nutrient + Propagation#Nutrient
+ Propagation#Harvests + Nutrient#Harvests
+ Propagation#Nutrient#Harvests

The Hasse diagram that includes the degrees of freedom is as follows The degrees of freedom for the randomized factors can be obtained using the rule for completely crossed structures.



The Hasse diagrams that include the ***Q*** *and* ***M*** *matrices for both structure formulae are as follows.*



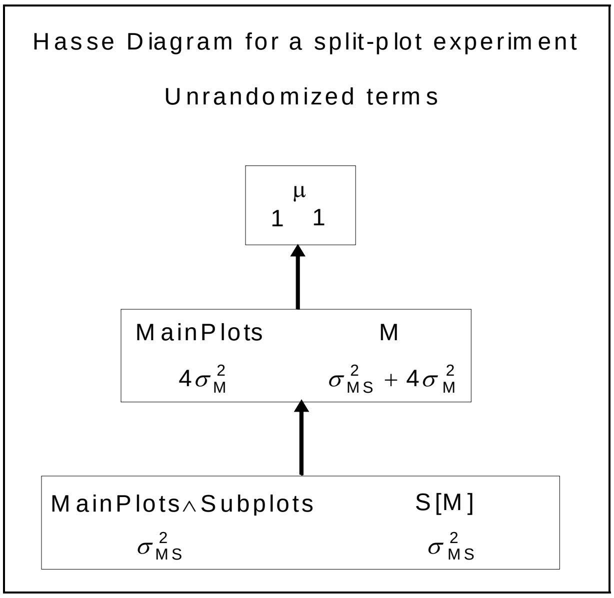


What are the expectation and variation models based on all unrandomized factors being random and all randomized factors being fixed?

$$\psi = E[Y] = Propagation{\wedge}Nutrient{\wedge}Harvests \text{ and}$$
$$Var[Y] = MainPlots + MainPlots{\wedge}Subplots$$

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.

The contributions of the randomized factors to the expected mean squares 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:



Source	df	SSq	E[MSq]
MainPlots	17	$\mathbf{Y}'\mathbf{Q}_{\mathrm{M}}\mathbf{Y}$	
Propagation	2	$\mathbf{Y}'\mathbf{Q}_{\mathrm{P}}\mathbf{Y}$	$\sigma^2_{\mathrm{MS}} \quad +4\sigma^2_{\mathrm{M}} \quad +q_{\mathrm{P}}\left(\psi\right)$
Nutrient	1	$\mathbf{Y}'\mathbf{Q}_{\mathrm{N}}\mathbf{Y}$	$\sigma^2_{\mathrm{MS}} \quad +4\sigma^2_{\mathrm{M}} \quad +q_{\mathrm{N}}\left(\psi\right)$
Propagation#Nutrient	2	$\mathbf{Y}'\mathbf{Q}_{\mathrm{PN}}\mathbf{Y}$	$\sigma^2_{\mathrm{MS}} \quad +4\sigma^2_{\mathrm{M}} \quad +q_{\mathrm{PN}}\left(\psi\right)$
Residual	12	$\mathbf{Y}'\mathbf{Q}_{\mathrm{M}_{\mathrm{Res}}}\mathbf{Y}$	$\sigma^2_{\mathrm{MS}} \quad +4\sigma^2_{\mathrm{M}}$
Subplots[MainPlots]	54	$\mathbf{Y}'\mathbf{Q}_{\mathrm{MS}}\mathbf{Y}$	
Harvests	3	$\mathbf{Y}'\mathbf{Q}_{\mathrm{H}}\mathbf{Y}$	$\sigma^2_{\mathrm{MS}} \quad +q_{\mathrm{H}}\left(\psi\right)$
Propagation#Harvests	6	$\mathbf{Y}'\mathbf{Q}_{\mathrm{PH}}\mathbf{Y}$	$\sigma^2_{\mathrm{MS}} \quad +q_{\mathrm{PH}}\left(\psi\right)$
Nutrient#Harvests	3	$\mathbf{Y}'\mathbf{Q}_{\mathrm{NH}}\mathbf{Y}$	$\sigma^2_{\mathrm{MS}} \quad +q_{\mathrm{NH}}\left(\psi\right)$
Propagation#Nutrient#Harvests	6	$\mathbf{Y}'\mathbf{Q}_{\mathrm{PNH}}\mathbf{Y}$	$\sigma^2_{\mathrm{MS}} \quad +q_{\mathrm{PNH}}\left(\psi\right)$
Residual	36	$\mathbf{Y}'\mathbf{Q}_{\mathrm{MS}_{\mathrm{Res}}}\mathbf{Y}$	σ^2_{MS}
Total	71	$\mathbf{Y}'\mathbf{Q}_{\mathrm{U}}\mathbf{Y}$	

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

```
> r <- 3
> a <- 6
> b <- 4
> n <- r*a*b
> fnames <- list(Propogation = 1:3, Nutrient = 1:2, Harvests = 1:b)
> SPLCeler.ran <- fac.gen(generate = fnames, times = r)
> Random.Order <- order(rep(runif(r*a), each=b), runif(n))
> SPLCeler.unit <- list(MainPlots=r*a, Subplots=b)
> SPLCeler.nest <- list(Subplots = "MainPlots")
> SPLCeler.lay <- fac.layout(unrandomized = SPLCeler.unit,
+                            nested.factors = SPLCeler.nest,
+                            randomized = SPLCeler.ran, seed = 445)
> SPLCeler.lay
```

	Units	Permutation	MainPlots	Subplots	Propogation	Nutrient	Harvests
1	1	46	1	1	3	1	1
2	2	48	1	2	3	1	2
3	3	45	1	3	3	1	4
4	4	47	1	4	3	1	3
5	5	70	2	1	2	1	1
6	6	71	2	2	2	1	3
7	7	69	2	3	2	1	4
8	8	72	2	4	2	1	2
9	9	5	3	1	3	2	2
10	10	8	3	2	3	2	3
11	11	6	3	3	3	2	4
12	12	7	3	4	3	2	1
13	13	23	4	1	1	1	2
14	14	22	4	2	1	1	1
15	15	24	4	3	1	1	3
16	16	21	4	4	1	1	4
17	17	38	5	1	2	1	4
18	18	39	5	2	2	1	3
19	19	40	5	3	2	1	2
20	20	37	5	4	2	1	1
21	21	43	6	1	2	2	4
22	22	44	6	2	2	2	2
23	23	42	6	3	2	2	1
24	24	41	6	4	2	2	3
25	25	14	7	1	2	1	4
26	26	13	7	2	2	1	3
27	27	15	7	3	2	1	1
28	28	16	7	4	2	1	2
29	29	52	8	1	3	1	1
30	30	49	8	2	3	1	2
31	31	50	8	3	3	1	3
32	32	51	8	4	3	1	4
33	33	20	9	1	1	1	2
34	34	19	9	2	1	1	4
35	35	18	9	3	1	1	1
36	36	17	9	4	1	1	3
37	37	62	10	1	3	1	4
38	38	61	10	2	3	1	1
39	39	64	10	3	3	1	2
40	40	63	10	4	3	1	3
41	41	1	11	1	3	2	4
42	42	2	11	2	3	2	3
43	43	4	11	3	3	2	1
44	44	3	11	4	3	2	2
45	45	12	12	1	1	1	3
46	46	9	12	2	1	1	1
47	47	10	12	3	1	1	4
48	48	11	12	4	1	1	2
49	49	35	13	1	1	2	2
50	50	33	13	2	1	2	3
51	51	36	13	3	1	2	4
52	52	34	13	4	1	2	1
53	53	67	14	1	2	2	2
54	54	68	14	2	2	2	1
55	55	66	14	3	2	2	4
56	56	65	14	4	2	2	3
57	57	27	15	1	3	2	4
58	58	28	15	2	3	2	1
59	59	26	15	3	3	2	3
60	60	25	15	4	3	2	2
61	61	54	16	1	2	2	2
62	62	53	16	2	2	2	1
63	63	56	16	3	2	2	4
64	64	55	16	4	2	2	3
65	65	29	17	1	1	2	4
66	66	30	17	2	1	2	3
67	67	31	17	3	1	2	1

68	68	32	17	4	1	2	2
69	69	58	18	1	1	2	3
70	70	60	18	2	1	2	1
71	71	59	18	3	1	2	2
72	72	57	18	4	1	2	4