STATISTICAL MODELLING

PRACTICAL VI
SOLUTIONS

VI.1 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

What are the features of this experiment?

1.	Observational unit	a leaf
2.	Response variable	Moisture Content
3.	Unrandomized factors	Plant, Leaf Size
4.	Randomized factors	Time
5.	Type of study	Latin square

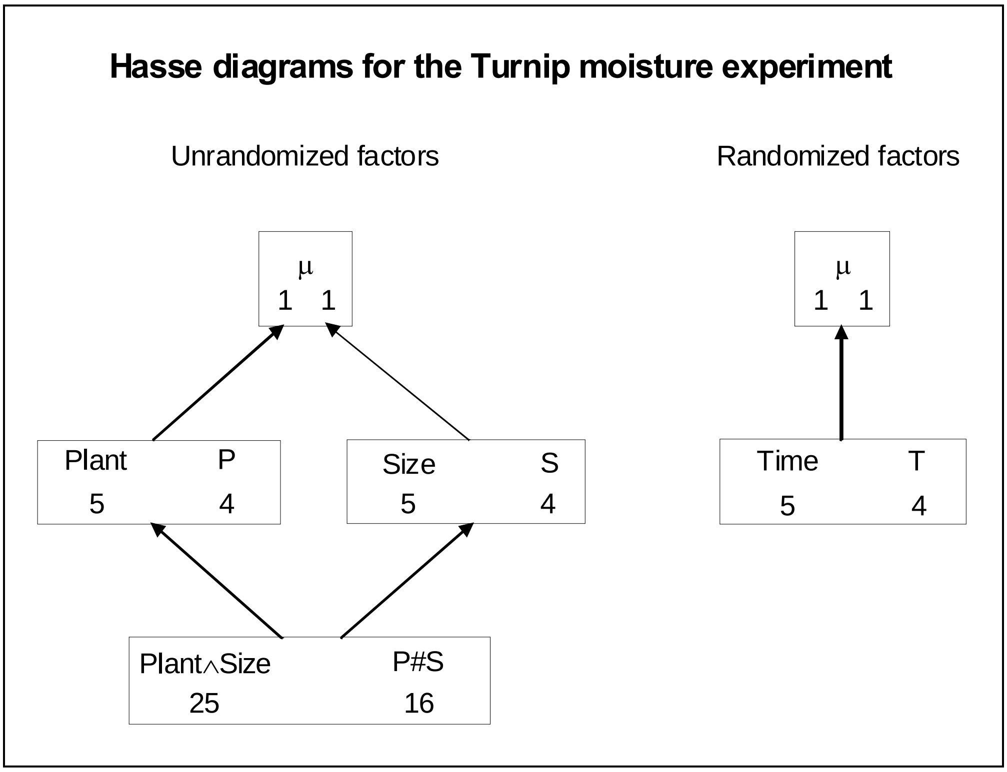
What is the experimental structure for this experiment?

Structure	*Formula*
unrandomized	*5 Plant*5 Size*
randomized	*5 Time*

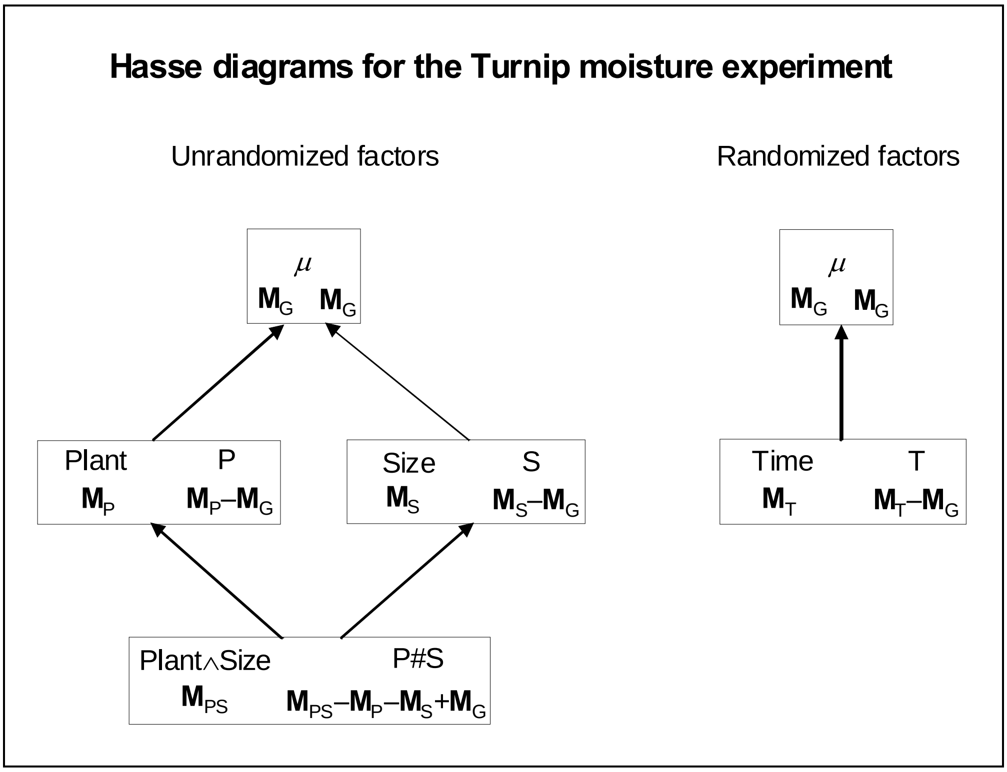
What are the Hasse diagrams of generalized-factor marginalities, with degrees of freedom and with **M** and **Q** matrices, for this study?

*The sources in the analysis are: Plant*Size $=$ Plant + Size + Plant#Size and Time $=$ Time.*

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



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



Derive the maximal expectation and variation models for this study?

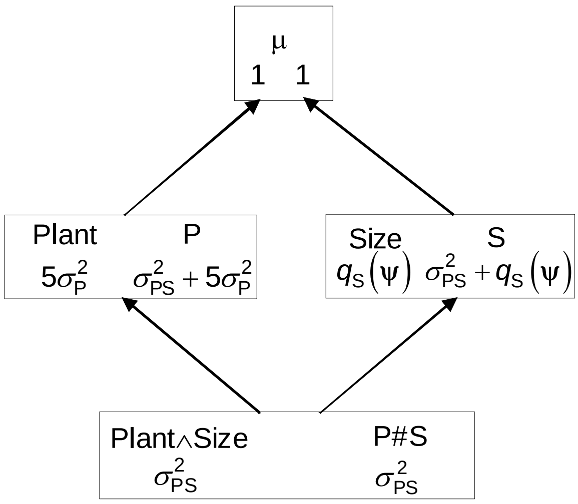
In this example, Plant is likely to be a random factor and Size and Time to be fixed factors. In particular, Size may well show a systematic trend and this would be modelled used expectation terms.

Hence, the maximal models are: ψ = *E[Y] = Size + Time and var[Y] = Plant + Plant$\wedge$Size.*

Determine the expected mean squares for this study, using where appropriate the Hasse diagrams of generalized-factor marginalities.

Hasse diagrams for the Turnip moisture experiment

Unrandomized factors



In addition the randomized factor Time will contribute $q_{\mathrm{T}}(\psi)$.

Give the analysis of variance table, including the degrees of freedom, sums of squares and expected mean squares.

Source	df	SSq	E[MSq]
Plant	4	$\mathbf{Y}'\mathbf{Q}_{\mathrm{P}}\mathbf{Y}$	$\sigma^2_{\mathrm{PS}} + 5\sigma^2_{\mathrm{P}}$
Size	4	$\mathbf{Y}'\mathbf{Q}_{\mathrm{S}}\mathbf{Y}$	$\sigma^2_{\mathrm{PS}} + q_{\mathrm{S}}(\psi)$
Plant#Size	16	$\mathbf{Y}'\mathbf{Q}_{\mathrm{PS}}\mathbf{Y}$	
Time	4	$\mathbf{Y}'\mathbf{Q}_{\mathrm{T}}\mathbf{Y}$	$\sigma^2_{\mathrm{PS}} + q_{\mathrm{T}}(\psi)$
Residual	12	$\mathbf{Y}'\mathbf{Q}_{\mathrm{PS}_{\mathrm{Res}}}\mathbf{Y}$	σ^2_{PS}
Total	24		

$q_{\mathrm{S}}(\psi) = 5\sum\left(\beta_j - \bar{\beta}_.\right)^2 \big/ 4$, $q_{\mathrm{T}}(\psi) = 5\sum\left(\tau_k - \bar{\tau}_.\right)^2 \big/ 4$

VI.2 A chemist has four different containers of soil. He wants to determine whether the moisture contents of these four soils differs. He randomly selects 10 samples from each container and determines the moisture content of each sample.

What are the features of this experiment?

1. Observational unit — a sample
2. Response variable — Moisture content
3. Unrandomized factors — Soil, Sample
4. Randomized factors — none
5. Type of study — SRS

What is the experimental structure for this experiment?

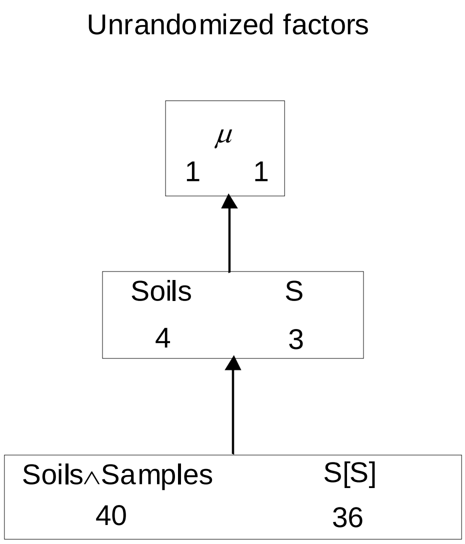
Structure	*Formula*
unrandomized	*4 Soils/10 Samples*
randomized	—

What are the Hasse diagrams of generalized-factor marginalities, with degrees of freedom and with **M** and **Q** matrices, for this study?

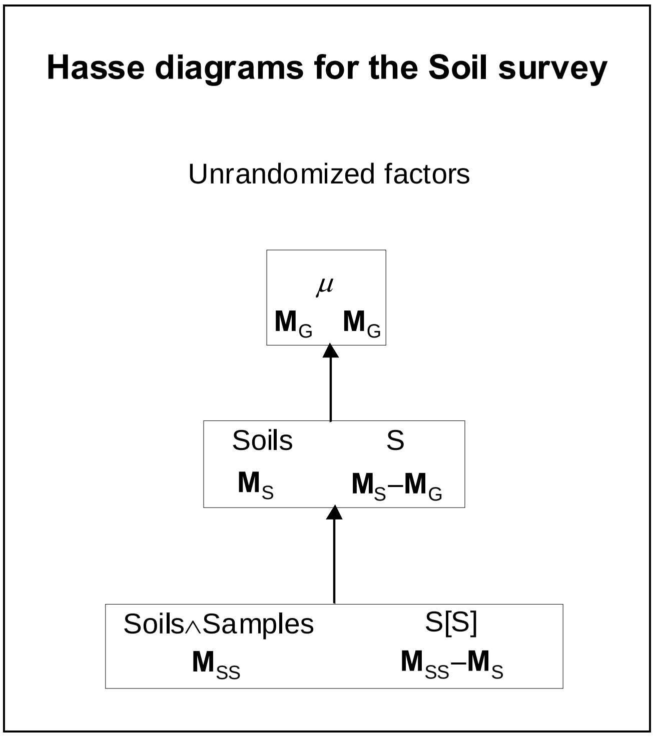
The sources in the analysis are: Soils/Samples = Soils + Samples[Soils].

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

Hasse diagrams for the Soil survey



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



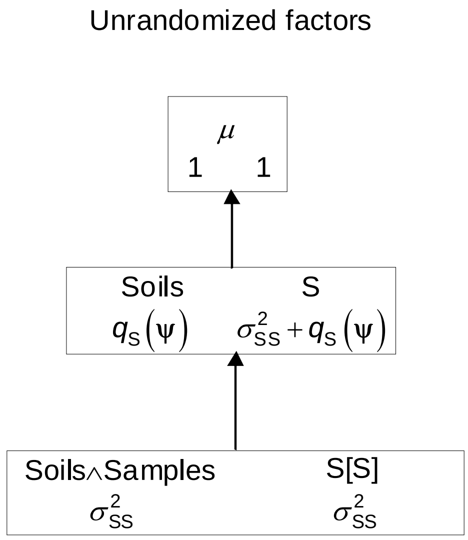
Derive the maximal expectation and variation models for this study?

In this survey Samples are likely to be random as these samples are meant to be representative of a larger population of samples and one envisages using a probability distribution function to model sample variability. I ti s not clear whether soils are fixed or random. Do we have four soils that difffer in an arbitrary way or are these 4 soils representative of a large group of similar soils? Assume they are fixed.

Based on this the maximal models would be: ψ = *E[Y]* = *Soils and var[Y]* = *Soils*$\wedge$*Samples.*

Determine the expected mean squares for this study, using where appropriate the Hasse diagrams of generalized-factor marginalities.

Hasse diagrams for the Soil survey



Give the analysis of variance table, including the degrees of freedom, sums of squares and expected mean squares.

Source	df	SSq	E[MSq]
Soils	3	$\mathbf{Y}'\mathbf{Q}_{\mathrm{S}}\mathbf{Y}$	$\sigma^2_{\mathrm{SS}} + q_{\mathrm{S}}(\psi)$
Samples[Soils]	36	$\mathbf{Y}'\mathbf{Q}_{\mathrm{SS}}\mathbf{Y}$	σ^2_{SS}
Total	39		

VI.3 In an experiment to investigate the yield (ℓ/hr) of machine producing a chemical, four randomly selected machines were operated at five different temperatures for an hour and the yield measured. The order in which each machine was operated at the different temperatures was randomized for each machine.

What are the features of this experiment?

1. Observational unit — a machine at a time
2. Response variable — Yield
3. Unrandomized factors — Machine, Time
4. Randomized factors — Temperature
5. Type of study — RCBD

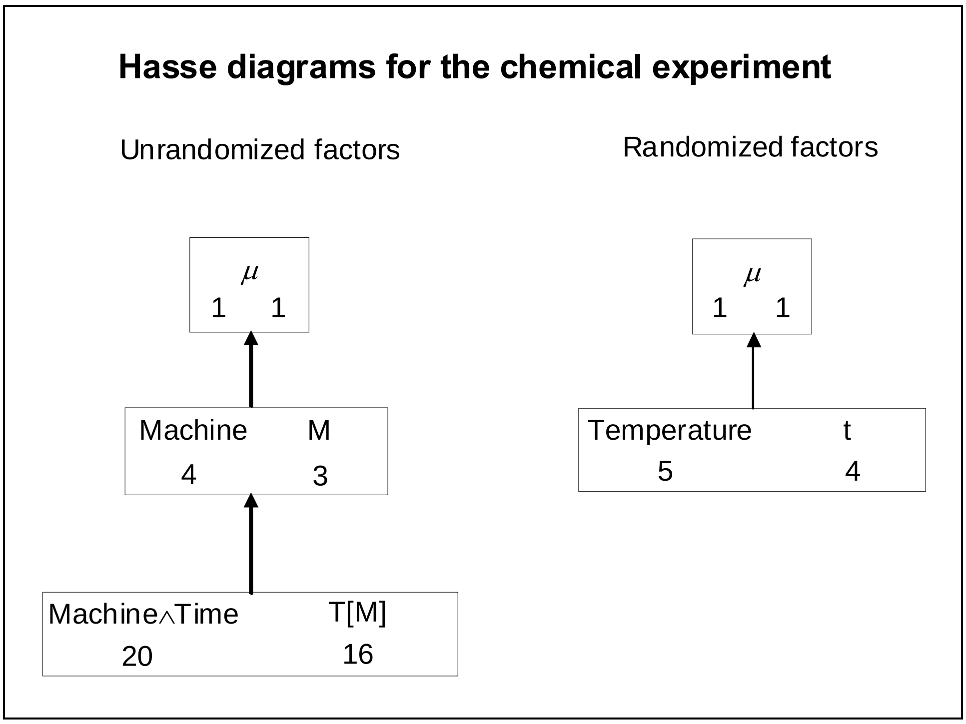
What is the experimental structure for this experiment?

Structure	*Formula*
unrandomized	*4 Machine/ 5 Time*
randomized	*5 Temperature*

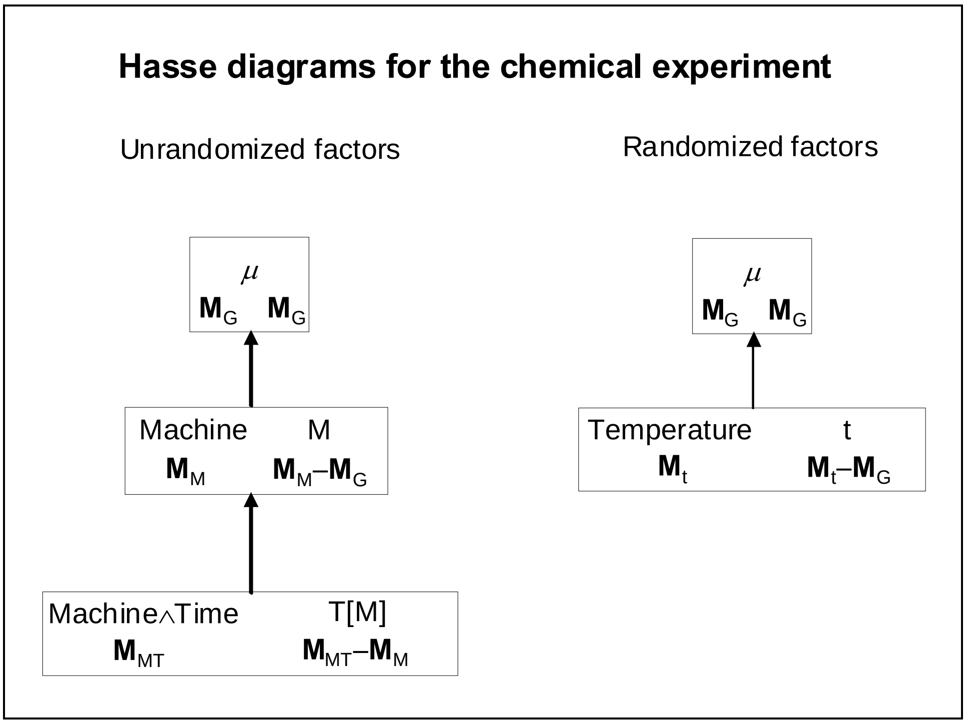
What are the Hasse diagrams of generalized-factor marginalities, with degrees of freedom and with **M** and **Q** matrices, for this study?

The sources in the analysis are: Machine/Time $=$ Machine + Time[Machine] and Temperature $=$ Temperature.

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



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

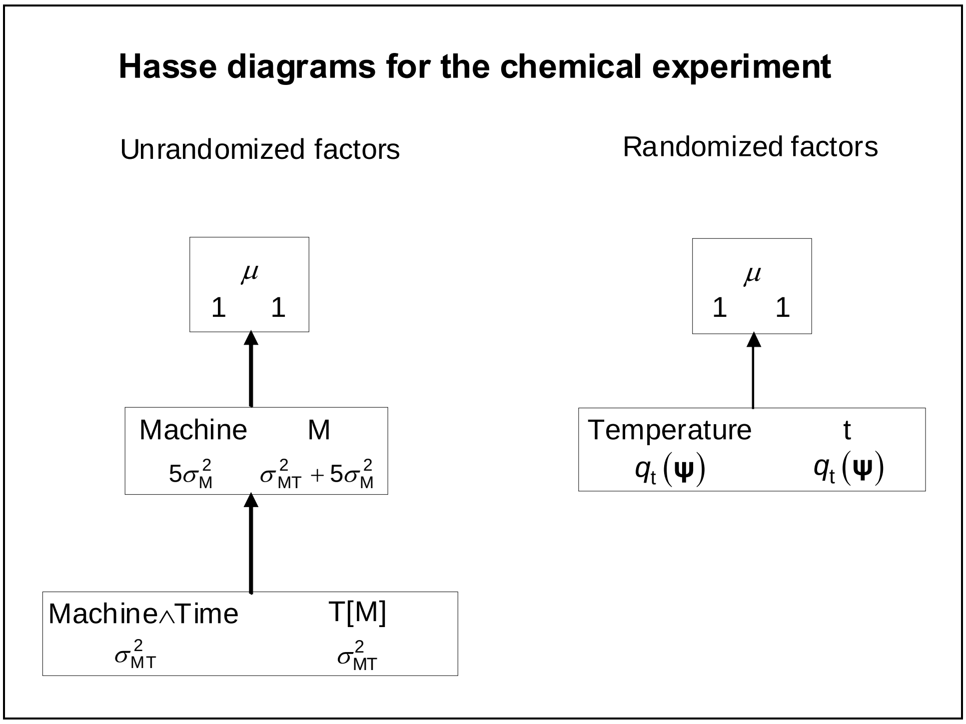


Derive the maximal expectation and variation models for this study?

In this experiment Machine and Time are likely to be random. It seems reasonable to regard all machines as having the same mean value and that different machine vary haphazardly around this mean value. Also, having randomized to the times within a machine, it would appear that we are not anticipating a systematic difference between the times and so again it would be appropriate to model the times within a machine as having the same mean value and the different times as varying haphazardly about the mean value. The factor Temperature would be fixed.

Based on this the maximal models would be: $\psi = E[Y] =$ *Temperature and var[Y]* $=$ *Machine + Machine∧Time.*

Determine the expected mean squares for this study, using where appropriate the Hasse diagrams of generalized-factor marginalities.



Give the analysis of variance table, including the degrees of freedom, sums of squares and expected mean squares.

Source	df	SSQ	E[MSq]
Machine	3	$\mathbf{Y}'\mathbf{Q}_{M}\mathbf{Y}$	$\sigma^2_{MT} + 5\sigma^2_{M}$
Time[Machine]	16	$\mathbf{Y}'\mathbf{Q}_{MT}\mathbf{Y}$	
Temperature	4	$\mathbf{Y}'\mathbf{Q}_{t}\mathbf{Y}$	$\sigma^2_{MT} + q_t(\boldsymbol{\psi})$
Residual	12	$\mathbf{Y}'\mathbf{Q}_{MT_{Res}}\mathbf{Y}$	σ^2_{MT}
Total	19		

$q_t(\boldsymbol{\psi}) = 4\sum(\tau_k - \bar{\tau}_.)^2/4$

VI.4 A study is to be conducted to compare two methods of measuring the concentration of a certain component of a liquid product. Three factories are selected from those that routinely determine the concentration of the component. A sample of the product is obtained and divided into 3 lots of 4 portions and each lot is randomly assigned to be sent to one of the factories. At each factory the concentration of their 4 portions is determined using both methods. The order in which a portion is tested using a particular method is completely randomized.

What are the features of this experiment?

1. Observational unit — a test at a factory
2. Response variable — Concentration
3. Unrandomized factors — Factories, Tests
4. Randomized factors — Lots, Portions, Methods
5. Type of study — RCBD with many randomized factors

What is the experimental structure for this experiment?

Structure	*Formula*
unrandomized	*3 Factories/ 8 Tests*
randomized	*(3 Lots/4 Portions)*2 Methods*

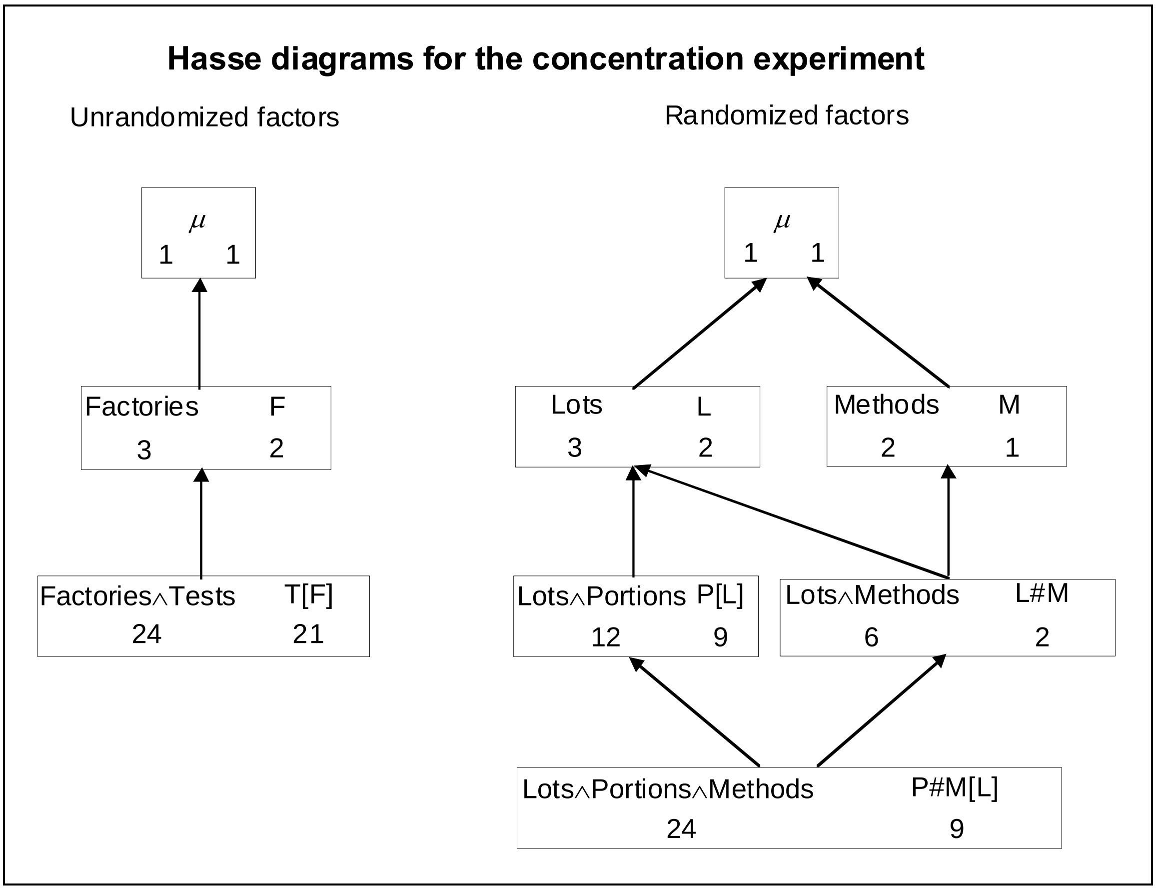
What are the Hasse diagrams of generalized-factor marginalities, with degrees of freedom and with **M** and **Q** matrices, for this study?

The sources in the analysis will be:

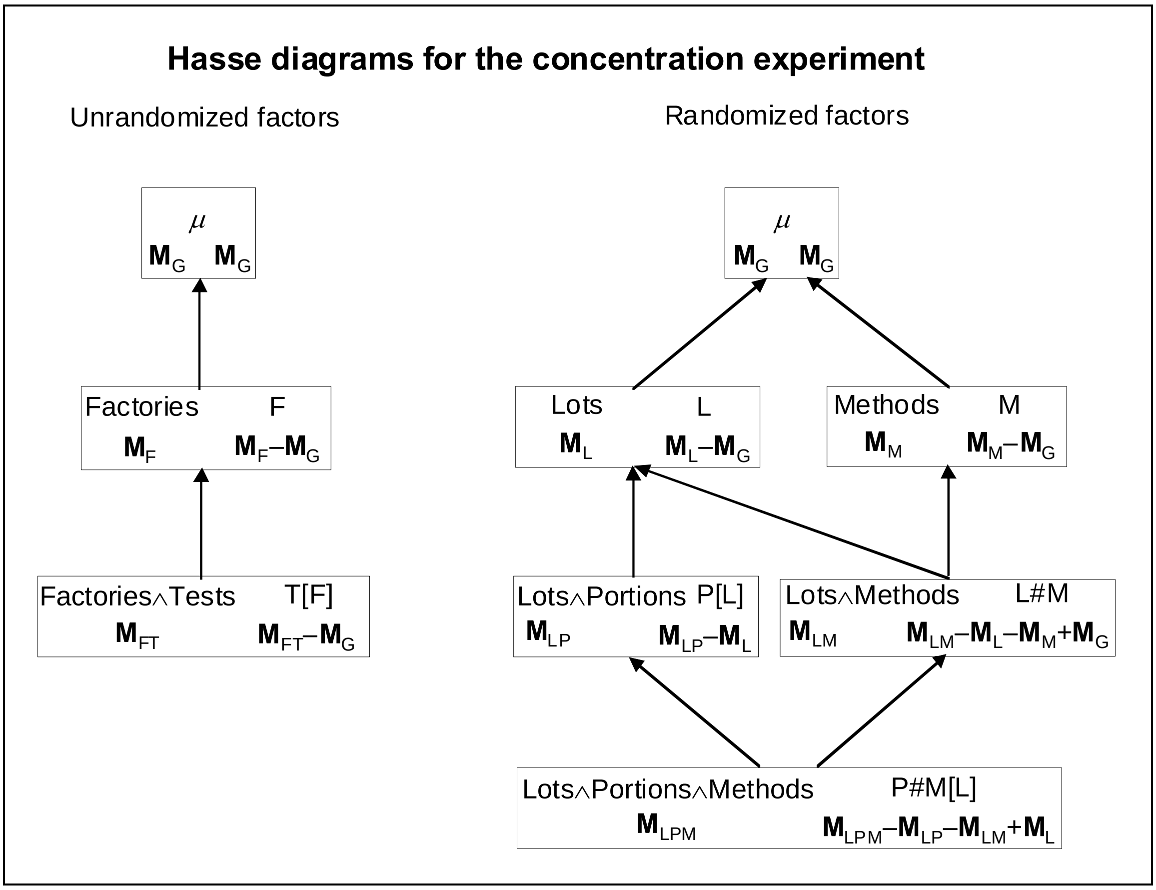
Factories/Tests = Factories + Tests[Factories]

*(Lots/Portions)*Methods*
*= (Lots + Portions[Lots])*Methods*
= Lots + Portions[Lots] + Methods + Lots#Methods + Portions#Methods[Lots]

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



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



Derive the maximal expectation and variation models for this study?

In this experiments the random factors would be Factories, Tests, Lots and Portions and the fixed factors would be Methods.

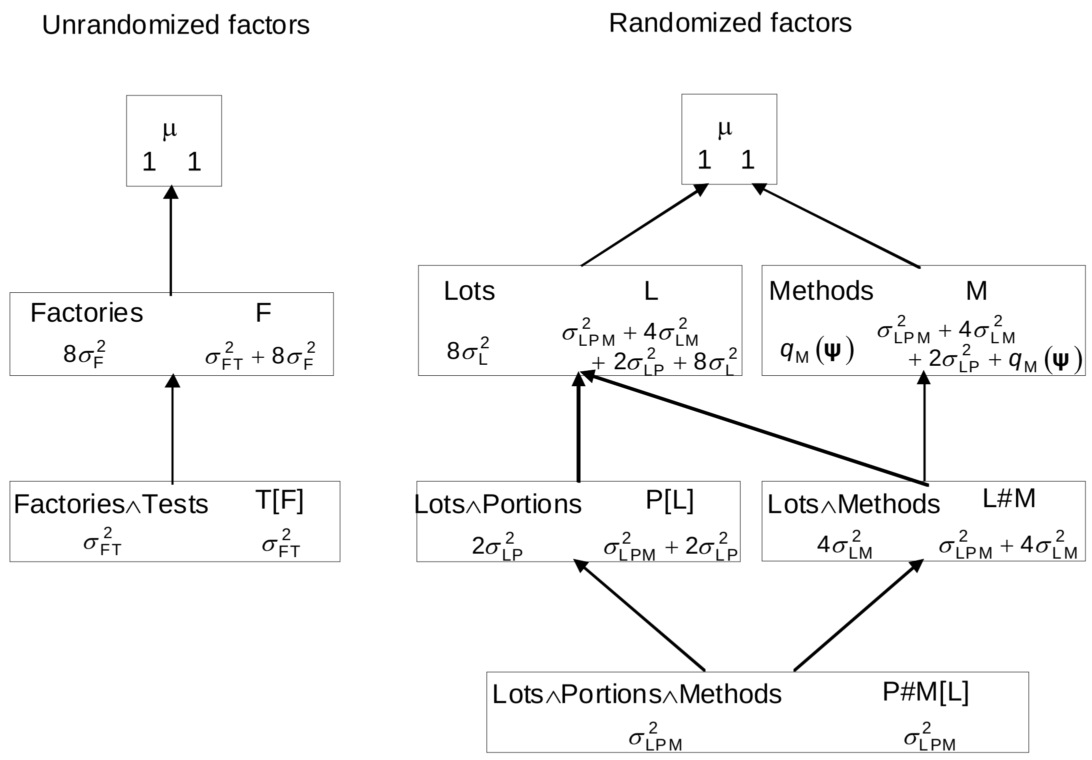
Hence the maximal models would be:

$\psi = E[Y] =$ *Methods and*

$var[Y] =$ *Factories + Factories∧Tests*
+ Lots + Lots∧Portions + Lots∧Methods + Lots∧Portions∧Methods

Determine the expected mean squares for this study, using where appropriate the Hasse diagrams of generalized-factor marginalities.

Hasse diagrams for the concentration experiment



Give the analysis of variance table, including the degrees of freedom, sums of squares and expected mean squares.

Source	df		SSq	E[MSq]
Factories	2		$\mathbf{Y}'\mathbf{Q}_{F}\mathbf{Y}$	
Lots	2		$\mathbf{Y}'\mathbf{Q}_{L}\mathbf{Y}$	$\sigma^2_{FT}+8\sigma^2_{F}+\sigma^2_{LPM}+4\sigma^2_{LM}+2\sigma^2_{LP}+8\sigma^2_{L}$
Tests[Factories]	21		$\mathbf{Y}'\mathbf{Q}_{FT}\mathbf{Y}$	
Portions[Lots]		9	$\mathbf{Y}'\mathbf{Q}_{LP}\mathbf{Y}$	$\sigma^2_{FT}+\sigma^2_{LPM}+2\sigma^2_{LP}$
Methods		1	$\mathbf{Y}'\mathbf{Q}_{M}\mathbf{Y}$	$\sigma^2_{FT}+\sigma^2_{LPM}+4\sigma^2_{LM}+q_{M}(\boldsymbol{\psi})$
Lots#Methods		2	$\mathbf{Y}'\mathbf{Q}_{LM}\mathbf{Y}$	$\sigma^2_{FT}+\sigma^2_{LPM}+4\sigma^2_{LM}$
Portions#Methods[Lots]		9	$\mathbf{Y}'\mathbf{Q}_{LPM}\mathbf{Y}$	$\sigma^2_{FT}+\sigma^2_{LPM}$
Total	24			

$$q_{M}(\boldsymbol{\psi})=12\sum\left(\tau_k-\overline{\tau}_{.}\right)^2\Big/1$$

So which ratio of mean squares would be used to test for an average difference between the two methods?

Take the ratio of the Methods and Lots#Methods mean squares to make the test. Notice that the degrees of freedom for this test are very low.

VI.5 Adapt the R expressions for randomizing a randomized complete block design to obtain a randomized layout and dummy analysis for the experiment described in exercise VI.4.

```
> b <- 3
> t <- 8
> n <- b*t
> Fac2Factory.unit <- list(Factories=b, Tests=t)
> Fac2Factory.nest <- list(Tests = "Factories")
> Fac2Factory.ran <- fac.gen(list(Lots = 3, Portions = 4, Methods = 2))
> Fac2Factory.lay <- fac.layout(unrandomized = Fac2Factory.unit,
+                               nested.factors = Fac2Factory.nest,
+                               randomized = Fac2Factory.ran, seed = 1015)
> Fac2Factory.lay
   Units Permutation Factories Tests Lots Portions Methods
1      1          19         1     1    2        1       2
2      2          17         1     2    2        1       1
3      3          22         1     3    2        2       1
4      4          21         1     4    2        4       2
5      5          20         1     5    2        4       1
6      6          23         1     6    2        3       1
7      7          24         1     7    2        3       2
8      8          18         1     8    2        2       2
9      9           2         2     1    3        1       2
10    10           1         2     2    3        3       1
11    11           3         2     3    3        2       2
12    12           8         2     4    3        3       2
13    13           6         2     5    3        1       1
14    14           7         2     6    3        2       1
15    15           5         2     7    3        4       2
16    16           4         2     8    3        4       1
17    17          13         3     1    1        1       2
18    18           9         3     2    1        4       2
19    19          14         3     3    1        1       1
20    20          11         3     4    1        3       1
21    21          10         3     5    1        2       2
22    22          12         3     6    1        2       1
23    23          16         3     7    1        3       2
24    24          15         3     8    1        4       1
> # add a column y with random normal data and analyze
> Fac2Factory.dat <- Fac2Factory.lay
> Fac2Factory.dat$y <- rnorm(n)
> Fac2Factory.aov <- aov(y ~ (Lots/Portions)*Methods + Error(Factories/Tests),
Fac2Factory.dat)
> summary(Fac2Factory.aov)

Error: Factories
     Df Sum Sq Mean Sq
Lots  2 5.7115  2.8558

Error: Factories:Tests
                      Df  Sum Sq Mean Sq
Methods                1  0.3194  0.3194
Lots:Portions          9 15.0825  1.6758
Lots:Methods           2  2.2787  1.1393
Lots:Methods:Portions  9 13.6808  1.5201
```

This confirms the form of the previously derived analysis.