STATISTICAL MODELLING

PRACTICAL VII
SOLUTIONS

VII.1 A completely randomized experiment was conducted to investigate the effect of vitamin B_{12} (0, 5 mg) and antibiotics (0, 40 mg) fed to swine. The response was the average daily gain in weight.

		Vitamin B_{12}	
		0	5
		1.30	1.26
	0	1.19	1.21
		1.08	1.19
Antibiotics			
		1.05	1.52
	40	1.00	1.56
		1.05	1.55

What are the components of this experiment?

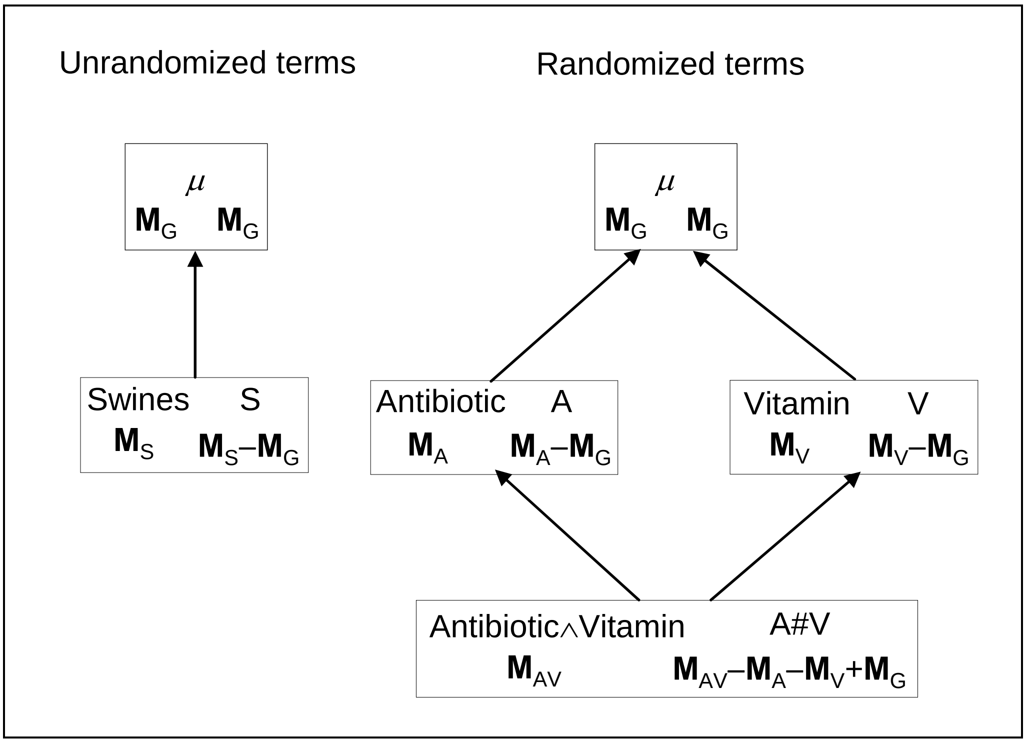
1. Observational unit – a swine
2. Response variable – Weight gain
3. Unrandomized factors – Swines
4. Randomized factors – Antibiotics, Vitamin
5. Type of study – Factorial CRD

What is the experimental structure for this experiment?

Structure	Formula
unrandomized	*12* Swines
randomized	*2* Antibiotic**2* Vitamin

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

Hasse diagrams, with M and Q matrices, for two-factor CRD



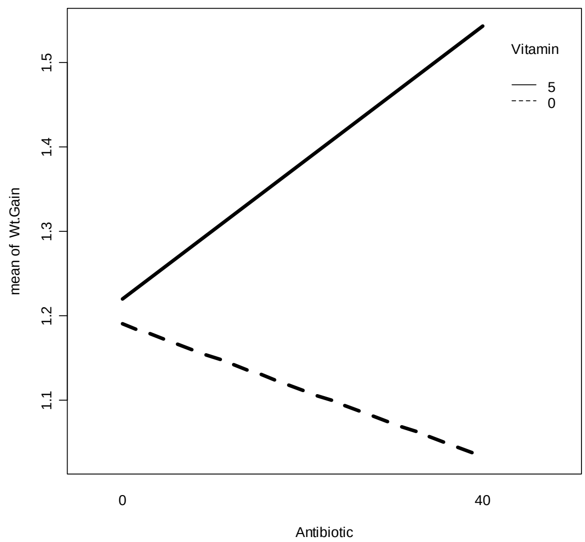
What are degrees of freedom, sums of squares and expected mean squares for the lines in the analysis of variance table based on all unrandomized factors being random and all randomized factors being fixed?

As there is only one unrandomized term and this is a variation term, the unrandomized factors will contribute σ_S^2 *to the expected mean squares. As all randomized factors are fixed, all the unrandomized terms will contribute a quadratic function of the form* $q_F(\psi)$.

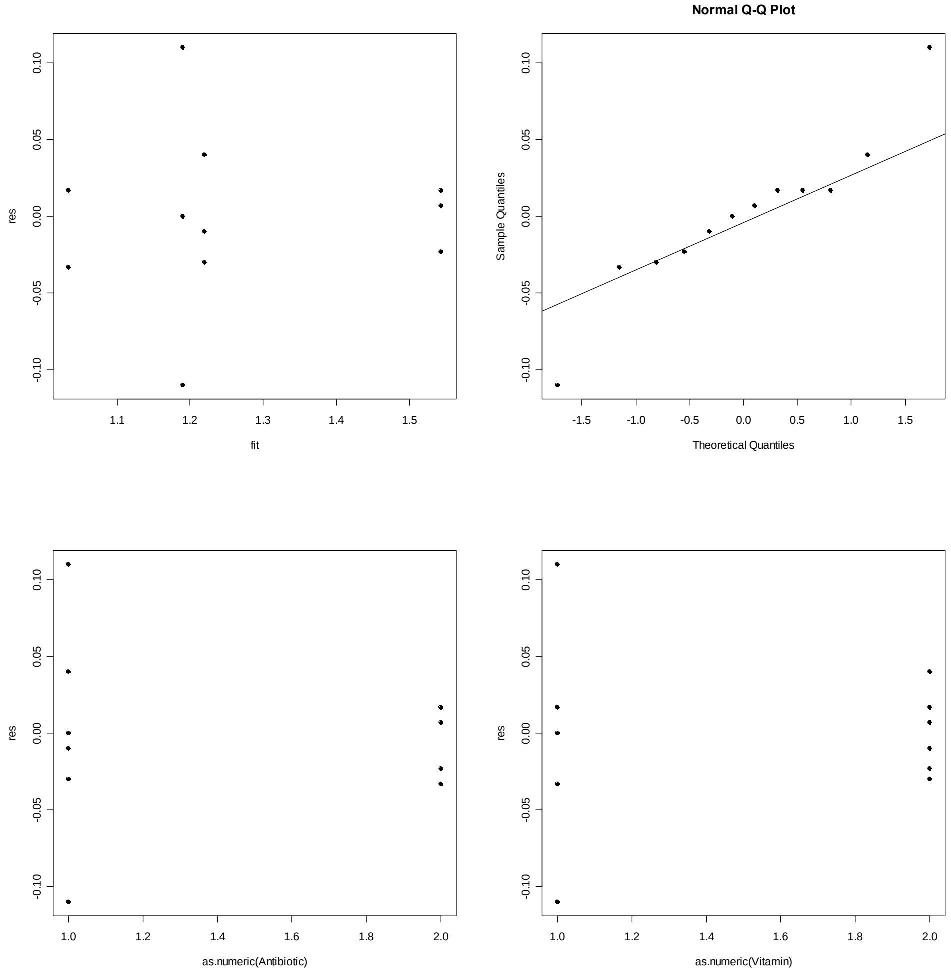
Source	df	SSq	E[MSq]
Swines	11	$\mathbf{Y}'\mathbf{Q}_S\mathbf{Y}$	
Antibiotic	1	$\mathbf{Y}'\mathbf{Q}_A\mathbf{Y}$	$\sigma_S^2 + q_A(\psi)$
Vitamin	1	$\mathbf{Y}'\mathbf{Q}_V\mathbf{Y}$	$\sigma_S^2 + q_V(\psi)$
Antibiotic#Vitamin	1	$\mathbf{Y}'\mathbf{Q}_{AV}\mathbf{Y}$	$\sigma_S^2 + q_{AV}(\psi)$
Residual	8	$\mathbf{Y}'\mathbf{Q}_{S_{Res}}\mathbf{Y}$	σ_S^2
Total	11	$\mathbf{Y}'\mathbf{Q}_S\mathbf{Y}$	

Obtain the usual analysis for a two-factor factorial experiment using R, including diagnostic checking. Also, examine treatment differences using multiple comparison procedures on the appropriate table(s) of means with a view to identifying the levels combinations of the factors that produce the maximum weight gain.

```
> attach(Fac2Swine.dat)
> Fac2Swine.dat
   Swine Antibiotic Vitamin Wt.Gain
1      1          0       0    1.30
2      2          0       5    1.26
3      3          0       0    1.19
4      4          0       5    1.21
5      5          0       0    1.08
6      6          0       5    1.19
7      7         40       0    1.05
8      8         40       5    1.52
9      9         40       0    1.00
10    10         40       5    1.56
11    11         40       0    1.05
12    12         40       5    1.55
> interaction.plot(Antibiotic, Vitamin, Wt.Gain, lwd=4)
> Fac2Swine.aov <- aov(Wt.Gain ~ Antibiotic * Vitamin + Error(Swine),
Fac2Swine.dat)
> summary(Fac2Swine.aov)

Error: Swine
                   Df   Sum Sq  Mean Sq F value    Pr(>F)
Antibiotic          1 0.020833 0.020833  5.6818 0.0442922
Vitamin             1 0.218700 0.218700 59.6455 5.622e-05
Antibiotic:Vitamin  1 0.172800 0.172800 47.1273 0.0001290
Residuals           8 0.029333 0.003667
> #
> # Diagnostic checking
> #
> res <- resid.errors(Fac2Swine.aov)
> fit <- fitted.errors(Fac2Swine.aov)
> plot(fit, res, pch=16)
> qqnorm(res, pch=16)
> qqline(res)
> plot(as.numeric(Antibiotic), res, pch = 16)
> plot(as.numeric(Vitamin), res, pch = 16)
```



There appears to be an interaction between Antibiotic and Vitamin.



As this is a CRD we cannot perform Tukey's nonadditivity test on the Residual.

The residuals-versus-fitted-values plot displays an appropriate pattern except that one treatment (no Vitamin, no Antibiotic) appears to have a somewhat higher spread compared to the others. This is reflected in the Normal Probability plot where there are two extreme residuals. That is, the problem is with two of the three observations for this treatment combination. Otherwise the plot looks satisfactory. There are insufficient observations to conclude that one of the treatments is more variable than the others and without further evidence of the reason for these outliers, it is advisable to leave them in the analysis.

The significance test for the example is a follows:

Step 1: Set up hypotheses

a) H_0: there is no interaction between Antibiotic and Vitamin

$$\left((\alpha\rho)_{ij} - \overline{(\alpha\rho)}_{i.} - \overline{(\alpha\rho)}_{.j} + \overline{(\alpha\rho)}_{..} = 0 \quad \text{for all i,j}\right)$$

H_1: there is an interaction between Antibiotic and Vitamin

$$\left((\alpha\rho)_{ij} - \overline{(\alpha\rho)}_{i.} - \overline{(\alpha\rho)}_{.j} + \overline{(\alpha\rho)}_{..} \neq 0 \quad \text{for some i,j}\right)$$

b) H_0: $\alpha_0 = \alpha_{40}$
H_1: the population Antibiotic means are different

c) H_0: $\rho_0 = \rho_5$
H_1: the population Vitamin means are different

Step 2: Calculate test statistics

The analysis of variance table for a two-factor CRD, with random factors being the unrandomized factors and fixed factors the randomized factors, is:

Source	df	SSq	MSq	E[MSq]	F	Prob
Swine	11	0.4417				
Antibiotic	1	0.0208	0.0208	$\sigma_S^2 + q_A(\psi)$	5.68	0.0443
Vitamin	1	0.2187	0.2187	$\sigma_S^2 + q_V(\psi)$	59.65	<0.001
Antibiotic#Vitamin	1	0.1728	0.1728	$\sigma_S^2 + q_{AV}(\psi)$	47.13	<0.001
Residual	8	0.0293	0.0037	σ_S^2		

Step 3: Decide between hypotheses

There is a significant interaction between Antibiotic and Vitamin level so that it appears that the expectation model that best describes the data is the maximal model $\psi_{AV} = \mathbf{X}_{AV}(\alpha\beta)$.

Because there is a significant interaction the differences between the means for each combination of Vitamin and Antibiotic are to be examined to identify the levels combinations that maximize the weight gain.

```
> #
> # multiple comparisons
> #
> Fac2Swine.tab <- model.tables(Fac2Swine.aov, type="means")
> Fac2Swine.tab$tables$"Antibiotic:Vitamin"
          Vitamin
Antibiotic 0        5
        0  1.190000 1.220000
        40 1.033333 1.543333
> q <- qtukey(0.95, 4, 8)
> q
[1] 4.52881
```

Tukey's HSD is

$$w(5\%) = \frac{4.52881}{\sqrt{2}} \times \sqrt{\frac{0.0037 \times 2}{3}} = 0.16$$

From the output above we see that the greatest weight gain will be achieved with Antibiotic is at 40 and Vitamin B_{12} is at 5. The weight gain is 1.54 for this combination and it significantly different for the weight gain for the three other treatment combinations.

VII.2 Examination of the interaction plot of swine weight gain in exercise VII.1 suggests that there might be a response when both Antibiotic is at 40 and Vitamin B_{12} is at 5 and not a significant difference between the other three combinations. To investigate this possibility set up an analysis with a nested factorial structure that examines the divergence between two groups of treatments thought to be different and differences within the groups where no divergence is expected.

```
> Fac2Swine.dat$Both <- factor(1 + as.numeric(Antibiotic == "40" &
+                                             Vitamin == "5"))
> Fac2Swine.dat$Treats <- fac.combine(list(Antibiotic, Vitamin))
> attach(Fac2Swine.dat)
> Fac2Swine.dat
   Swine Antibiotic Vitamin Wt.Gain Both Treats
1      1          0       0    1.30    1      1
2      2          0       5    1.26    1      2
3      3          0       0    1.19    1      1
4      4          0       5    1.21    1      2
5      5          0       0    1.08    1      1
6      6          0       5    1.19    1      2
7      7         40       0    1.05    1      3
8      8         40       5    1.52    2      4
9      9         40       0    1.00    1      3
10    10         40       5    1.56    2      4
11    11         40       0    1.05    1      3
12    12         40       5    1.55    2      4
> Fac2Swine.aov <- aov(Wt.Gain ~ Both/Treats + Error(Swine), Fac2Swine.dat)
> summary(Fac2Swine.aov)

Error: Swine
            Df  Sum Sq Mean Sq F value    Pr(>F)
Both         1 0.35204 0.35204 96.0121 9.879e-06
Both:Treats  2 0.06029 0.03014  8.2212   0.01148
Residuals    8 0.02933 0.00367
```

The analysis indicates that our proposed model does not fit the data as there are differences between the three other treatment combinations.

VII.3 An experiment was conducted to investigate the effect of temperature and copper content on the warping of copper plates. Copper plates were produced using each of the combinations of temperature and copper content on one day and this was repeated on a second day; the order of the temperature-copper content combinations was randomized to the 16 production runs used each day in the experiment. The amount of warping of the copper plates produced was measured and the results are given in the following table.

	Temperature (°C)							
	50		75		100		125	
Day	1	2	1	2	1	2	1	2
Copper content (%)								
40	17	20	12	9	16	12	21	17
60	16	21	18	13	18	21	23	21
80	24	22	17	12	25	23	23	22
100	28	27	27	31	30	30	29	31

What are the components of this experiment?

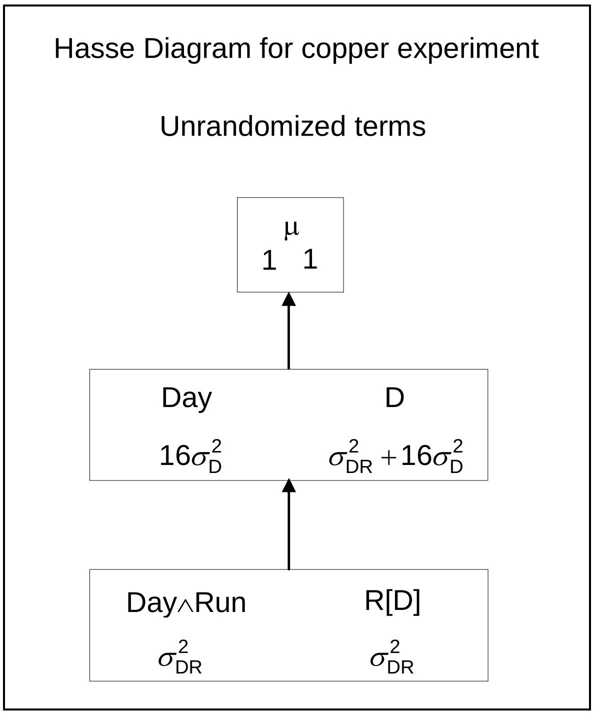
1. Observational unit – a run
2. Response variable – Amount of warping
3. Unrandomized factors – Day, Run
4. Randomized factors – Copper, Temperature
5. Type of study – Two-factor RCBD

What is the experimental structure for this experiment?

Structure	Formula
unrandomized	*2* Day/*16* Run
randomized	*4* Copper**4* Temperature

What are the degrees of freedom, sums of squares and expected mean squares for the lines in the analysis of variance table based on all unrandomized factors being random and all randomized factors being fixed?

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:*



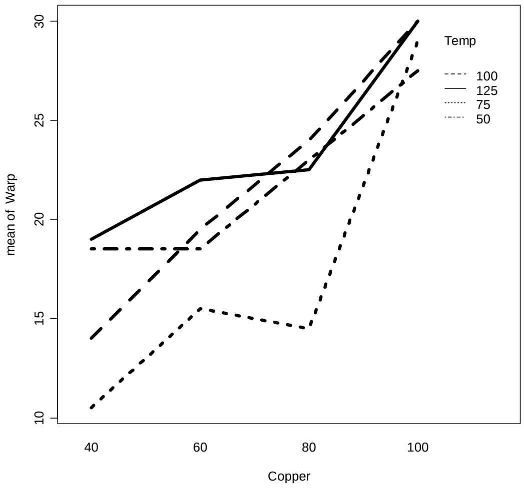
Source	df	SSq	E[MSq]
Day	1	$\mathbf{Y}'\mathbf{Q}_{\mathrm{D}}\mathbf{Y}$	$\sigma^2_{\mathrm{DR}} + 16\sigma^2_{\mathrm{D}}$
Run[Day]	30	$\mathbf{Y}'\mathbf{Q}_{\mathrm{DR}}\mathbf{Y}$	
Copper	3	$\mathbf{Y}'\mathbf{Q}_{\mathrm{C}}\mathbf{Y}$	$\sigma^2_{\mathrm{DR}} + q_{\mathrm{C}}(\boldsymbol{\psi})$
Temp	3	$\mathbf{Y}'\mathbf{Q}_{\mathrm{T}}\mathbf{Y}$	$\sigma^2_{\mathrm{DR}} + q_{\mathrm{T}}(\boldsymbol{\psi})$
Copper#Temp	9	$\mathbf{Y}'\mathbf{Q}_{\mathrm{CT}}\mathbf{Y}$	$\sigma^2_{\mathrm{DR}} + q_{\mathrm{CT}}(\boldsymbol{\psi})$
Residual	15	$\mathbf{Y}'\mathbf{Q}_{\mathrm{DR}_{\mathrm{Res}}}\mathbf{Y}$	σ^2_{DR}
Total	31	$\mathbf{Y}'\mathbf{Q}_{\mathrm{U}}\mathbf{Y}$	

Analyze the data using R, including diagnostic checking and obtaining a fitted equation and surface for an appropriate polynomial submodel.

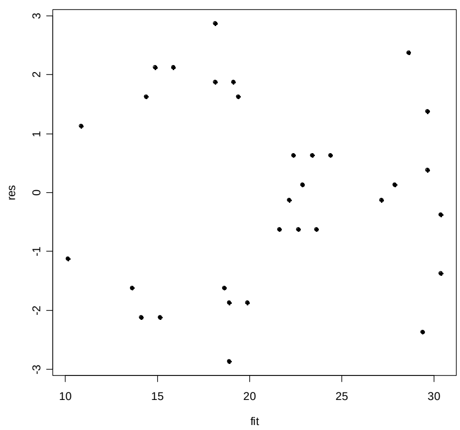
```
> attach(Fac2Copp.dat)
> Fac2Copp.dat
   Day Run Copper Temp Warp Warp.fit
1    1   1     40   50   17  18.2775
2    2   1     40   50   20  18.2775
3    1   2     40   75   12  10.5675
4    2   2     40   75    9  10.5675
5    1   3     40  100   16  15.0075
6    2   3     40  100   12  15.0075
7    1   4     40  125   21  19.5975
8    2   4     40  125   17  19.5975
9    1   5     60   50   16  19.0925
10   2   5     60   50   21  19.0925
11   1   6     60   75   18  13.5225
12   2   6     60   75   13  13.5225
13   1   7     60  100   18  18.0025
14   2   7     60  100   21  18.0025
15   1   8     60  125   23  20.5325
16   2   8     60  125   21  20.5325
17   1   9     80   50   24  22.2825
18   2   9     80   50   22  22.2825
19   1  10     80   75   17  18.8525
20   2  10     80   75   12  18.8525
21   1  11     80  100   25  23.3725
22   2  11     80  100   23  23.3725
23   1  12     80  125   23  23.8425
24   2  12     80  125   22  23.8425
25   1  13    100   50   28  27.8475
26   2  13    100   50   27  27.8475
27   1  14    100   75   27  26.5575
28   2  14    100   75   31  26.5575
29   1  15    100  100   30  31.1175
30   2  15    100  100   30  31.1175
31   1  16    100  125   29  29.5275
32   2  16    100  125   31  29.5275
> interaction.plot(Copper, Temp, Warp, lwd=4)
```

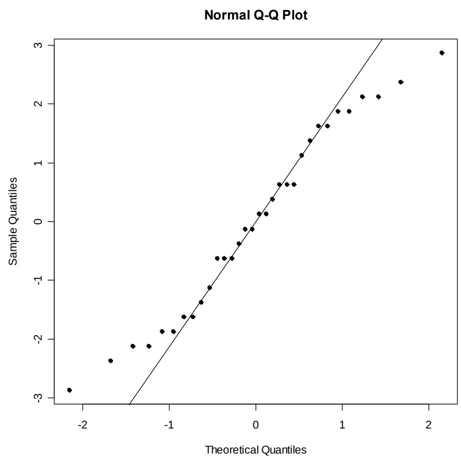
```
> #
> # Set up to fit polynomials
> #
> Copper.lev <- seq(40, 100, 20)
> Fac2Copp.dat$Copper <- ordered(Fac2Copp.dat$Copper, levels=Copper.lev)
> contrasts(Fac2Copp.dat$Copper) <- contr.poly(4, scores=Copper.lev)
> contrasts(Fac2Copp.dat$Copper)
            .L   .Q         .C
40  -0.6708204  0.5 -0.2236068
60  -0.2236068 -0.5  0.6708204
80   0.2236068 -0.5 -0.6708204
100  0.6708204  0.5  0.2236068
> Temp.lev <- seq(50, 125, 25)
> Fac2Copp.dat$Temp <- ordered(Fac2Copp.dat$Temp, levels=Temp.lev)
> contrasts(Fac2Copp.dat$Temp) <- contr.poly(4, scores=Temp.lev)
> contrasts(Fac2Copp.dat$Temp)
            .L   .Q         .C
50  -0.6708204  0.5 -0.2236068
75  -0.2236068 -0.5  0.6708204
100  0.2236068 -0.5 -0.6708204
125  0.6708204  0.5  0.2236068
> Fac2Copp.aov <- aov(Warp ~ Day + Copper * Temp + Error(Day/Run), Fac2Copp.dat)
> summary(Fac2Copp.aov, split = list(
+         Copper = list(L=1, Q=2, Dev=3),
+         Temp = list(L=1, Q= 2, Dev=3),
+         "Copper:Temp" = list(L.L=1, L.Q=2, Q.L=4, Q.Q=5, Dev=c(3,6:9))))

Error: Day
    Df Sum Sq Mean Sq
Day  1    4.5     4.5

Error: Day:Run
                   Df Sum Sq Mean Sq  F value    Pr(>F)
Copper              3 805.75  268.58  50.6761 4.391e-08
  Copper: L         1 739.60  739.60 139.5472 5.356e-09
  Copper: Q         1  45.12   45.12   8.5142 0.0106025
  Copper: Dev       1  21.03   21.03   3.9670 0.0649314
Temp                3 162.00   54.00  10.1887 0.0006562
  Temp: L           1  32.40   32.40   6.1132 0.0258714
  Temp: Q           1  72.00   72.00  13.5849 0.0022024
  Temp: Dev         1  57.60   57.60  10.8679 0.0048924
Copper:Temp         9 101.75   11.31   2.1331 0.0935649
  Copper:Temp: L.L  1   0.08    0.08   0.0151 0.9038498
  Copper:Temp: L.Q  1   2.02    2.02   0.3821 0.5457662
  Copper:Temp: Q.L  1  44.10   44.10   8.3208 0.0113440
  Copper:Temp: Q.Q  1   0.13    0.13   0.0236 0.8799928
  Copper:Temp: Dev  5  55.42   11.08   2.0913 0.1232238
Residuals          15  79.50    5.30
> #Compute Day F and p
> Day.F <- 4.5/5.3
> Day.p <- 1-pf(Day.F, 1, 15)
> data.frame(Day.F,Day.p)
      Day.F     Day.p
1 0.8490566 0.3714026
> #
> # Diagnostic checking
> #
> res <- resid.errors(Fac2Copp.aov)
> fit <- fitted.errors(Fac2Copp.aov)
>
> plot(fit, res, pch=16)
> qqnorm(res, pch=16)
> qqline(res)
> plot(as.numeric(Day), res, pch = 16)
> plot(as.numeric(Copper), res, pch = 16)
> plot(as.numeric(Temp), res, pch = 16)
> tukey.1df(Fac2Copp.aov, Fac2Copp.dat, error.term="Day:Run")
$Tukey.SS
```

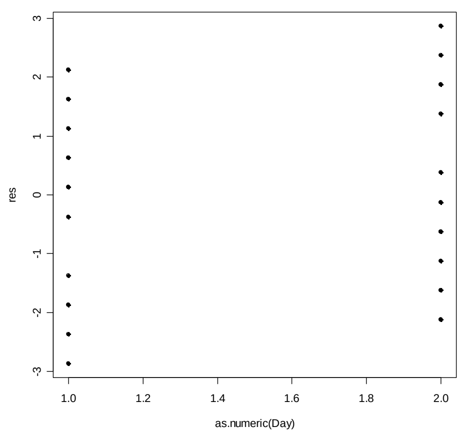
```
[1] 16.91468

$Tukey.F
[1] 3.783723

$Tukey.p
[1] 0.07211823

$Devn.SS
[1] 62.58532
```

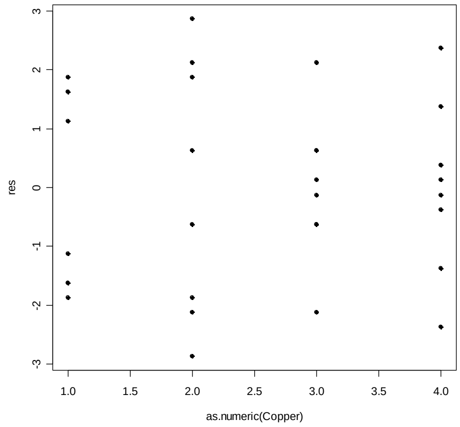


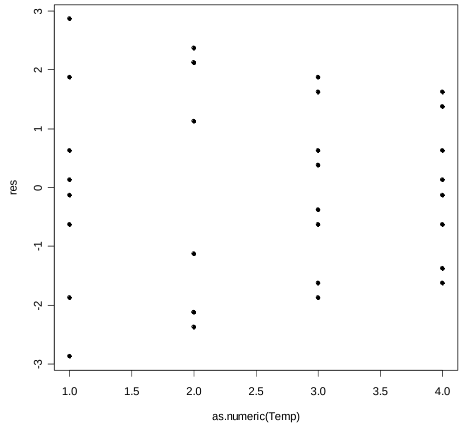
There appears to be an interaction between Temperature and Copper.











The residuals-versus-fitted-values plot is satisfactory as are the residuals-versus-Day and residuals-versus-Copper plots. However, the residuals-versus-Temperature plot is displaying a tendency for variability to decrease as Temperature increases. The difference is small and, as the treatments are equally replicated, will have little effect on the analysis. Tukey's one-degree-of-freedom-for-nonadditivity is not significant so that the residuals are not displaying nonadditivity. The normal probability plot is displaying some evidence of nonnormality, but it is not sufficient to warrant taking action.

The following ANOVA table was constructed using R and includes the partitioning of the two quantitative factors into Linear and Quadratic trends.

Source	df	SSq	MSq	F	Prob
Day	1	4.50	4.50	0.85	0.371
Run[Day]	30				
Copper	3	805.75	268.58	50.68	<.001
Linear	1	739.60	739.60	139.55	<.001
Quadratic	1	45.12	45.12	8.51	0.011
Deviations	1	21.03	21.03	3.97	0.065
Temperature	3	162.00	54.00	10.19	<.001
Linear	1	32.40	32.40	6.11	0.026
Quadratic	1	72.00	72.00	13.58	0.002
Deviations	1	57.60	57.60	10.87	0.005
Copper#Temperature	9	101.75	11.31	2.13	0.094
$A_{Linear}\#B_{Linear}$	1	0.08	0.08	0.02	0.904
$A_{Quadratic}\#B_{Linear}$	1	2.02	2.02	0.38	0.546
$A_{Linear}\#B_{Quadratic}$	1	44.10	44.10	8.32	0.011
$A_{Quadratic}\#B_{Quadratic}$	1	0.13	0.13	0.02	0.880
Deviations	5	55.42	11.08	2.09	0.123
Residual	15	79.50	5.30		
Nonadditivity	1	16.92	16.92	3.78	0.072
Deviations	14	62.58	4.47		
Total	31	1153.50			

This analysis indicates that the interaction can be explained by a linear-quadratic interaction, with the Copper main effect being quadratic and that the Temperature main effect significantly deviates from a quadratic. Consequently a model involving polynomials with at most squared terms on each factor does not adequately describe the data.

However, on noting that deviations for temperature have only 1 degree of freedom, it follows that a cubic will fit the temperature means perfectly. So we will investigate the fitting of a quadratic Copper main effect, cubic Temperature and a linear-quadratic interaction. The fitted equation can be obtained using S-the Plus `lm` *function.*

```
> # get fitted equation using power terms
> #
> Cu <- as.numeric(as.vector(Copper))
> Cu2 <- Cu * Cu
> Te <- as.numeric(as.vector(Temp))
> Te2 <- Te * Te
> Te3 <- Te2 * Te
> LinLin <- Cu * Te
> LinQuad <- LinLin * Te
> Fac2Copp.lm <- lm(Warp ~ Cu + Cu2 + Te + Te2 +Te3 + LinLin + LinQuad)
> coef(Fac2Copp.lm)
 (Intercept)        Cu         Cu2         Te        Te2        Te3     LinLin
                                                                             LinQuad
144.80250000 -0.7851250  0.00296875 -4.1946000  0.0418800 -0.0001280  0.0147800
                                                                          -0.0000840
```

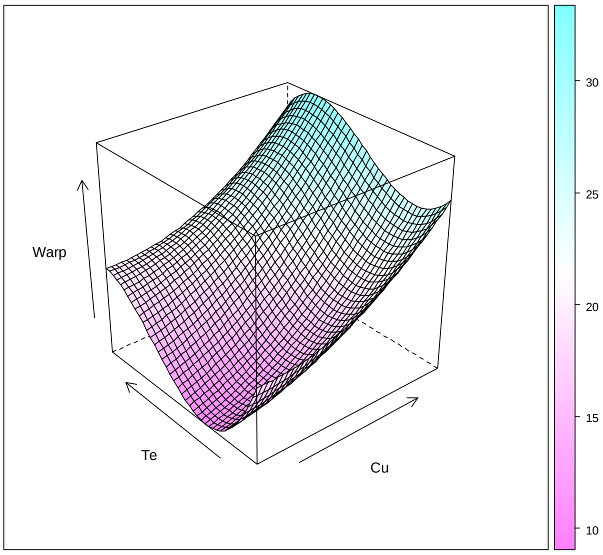
That is, the Warp response to copper and temperature can be approximated by the following function of the copper and temperature levels.

$$E[y_{km}] = 144.8 - 0.785x_{\alpha_i} + 0.00297x_{\alpha_i}^2 - 4.19x_{\beta_j} + 0.0419x_{\beta_j}^2 - 0.0001280x_{\beta_j}^3$$
$$+ 0.01478x_{\alpha_i}x_{\beta_j} - 0.000084x_{\alpha_i}x_{\beta_j}^2$$

However, it is dangerous to assume that this function adequately describes the response over the whole range of observed temperature values as we have fitted a cubic term to the temperatures that will fit the observed points exactly. The fitted cubic, having been constrained to go through the observed temperatures, has its shape between observed temperatures determined by the need to fit through the observed data points and may not reflect the actual response in these segments. Indeed examination of the interaction plot reveals that the Temperature trend for Copper at 100 is nearly flat, yet on the fitted surface below we see that it indicates a wavy trend in the temperature for Copper at 100.

The fitted surface can be plotted using the following expressions. Note the use of `poly` *function in the model. This makes the use of* `predict` *much simpler because we do not have to supply values for all 7 independent variables but we did not use it to get the coefficients because it does not give the natural coefficients.*

```
> #
> # get fitted surface using orthogonal polynomials
> #
> Fac2Copp.lm <- lm(Warp ~ poly(Cu, 2) + poly(Te,3)
+                            + poly(Cu, 1) * poly(Te, 2), singular.ok=T)
> coef(Fac2Copp.lm)
            (Intercept)             poly(Cu, 2)1             poly(Cu, 2)2
poly(Te, 3)1             poly(Te, 3)2
              21.125000                27.195588                 6.717514
5.692100                 8.485281
           poly(Te, 3)3              poly(Cu, 1)             poly(Te, 2)1
poly(Te, 2)2 poly(Cu, 1):poly(Te, 2)1
              -7.589466                       NA                       NA
NA                 1.600000
poly(Cu, 1):poly(Te, 2)2
             -37.565942
> Fac2Copp.grid <- list(Cu = seq(min(Cu), max(Cu), length = 40),
+                        Te = seq(min(Te), max(Te), length = 40))
> Fac2Copp.surf <- expand.grid(Fac2Copp.grid)
> Fac2Copp.surf$Warp <- as.vector(predict(Fac2Copp.lm, Fac2Copp.surf))
Warning message:
prediction from a rank-deficient fit may be misleading in:
predict.lm(Fac2Copp.lm, Fac2Copp.surf)
> wireframe(Warp ~ Cu*Te, data= Fac2Copp.surf, drape=TRUE)
```



VII.4 The following are data on the number of units produced per day by different operators in different machines. The order of the operator-machine combinations was randomized to the days in a particular period. The whole process was repeated in a second period with re-randomization of the operator-machine combinations. The first observation for each combination in the following table is for the first period and the second for the second period.

	Operator							
Machine	A		B		C		D	
1	18	17	16	18	17	20	27	27
2	17	13	18	18	20	16	28	23
3	16	17	17	19	20	16	31	30
4	15	17	21	22	16	16	31	24
5	17	18	16	18	14	13	28	22

What are the components of this experiment?

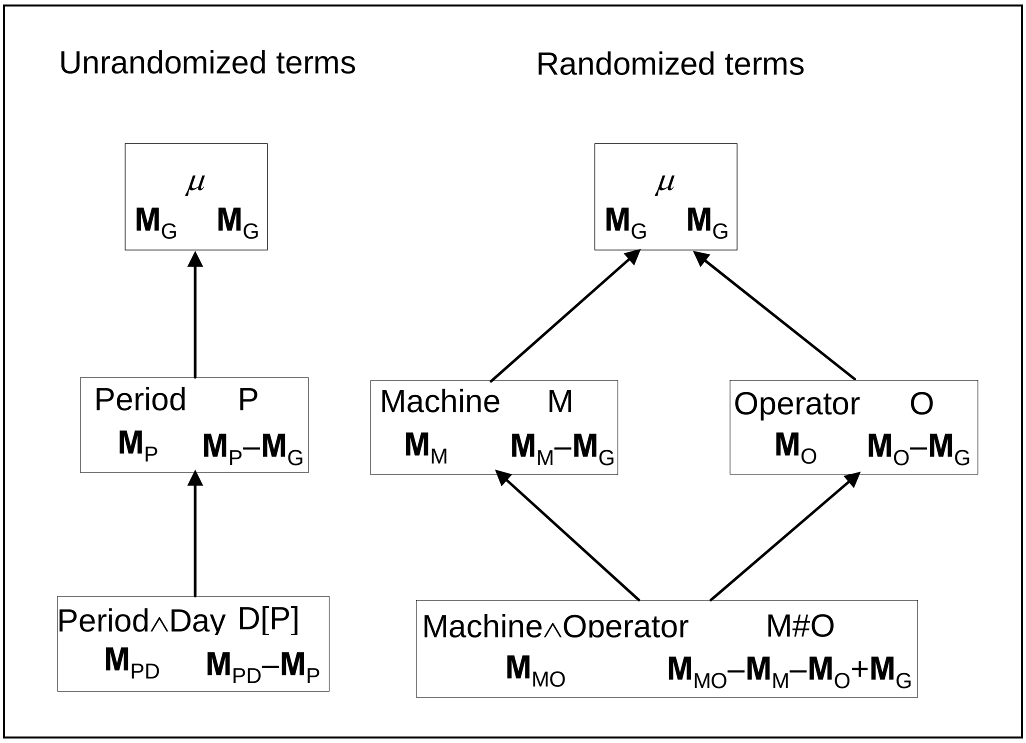
1. Observational unit — an observation made on a day
2. Response variable — No. units produced
3. Unrandomized factors — Period, Day
4. Randomized factors — Machine, Operator
5. Type of study — Factorial RCBD

What is the experimental structure for this experiment?

Structure	Formula
unrandomized	*2* Period/*20* Day
randomized	*5* Machine**4* Operator

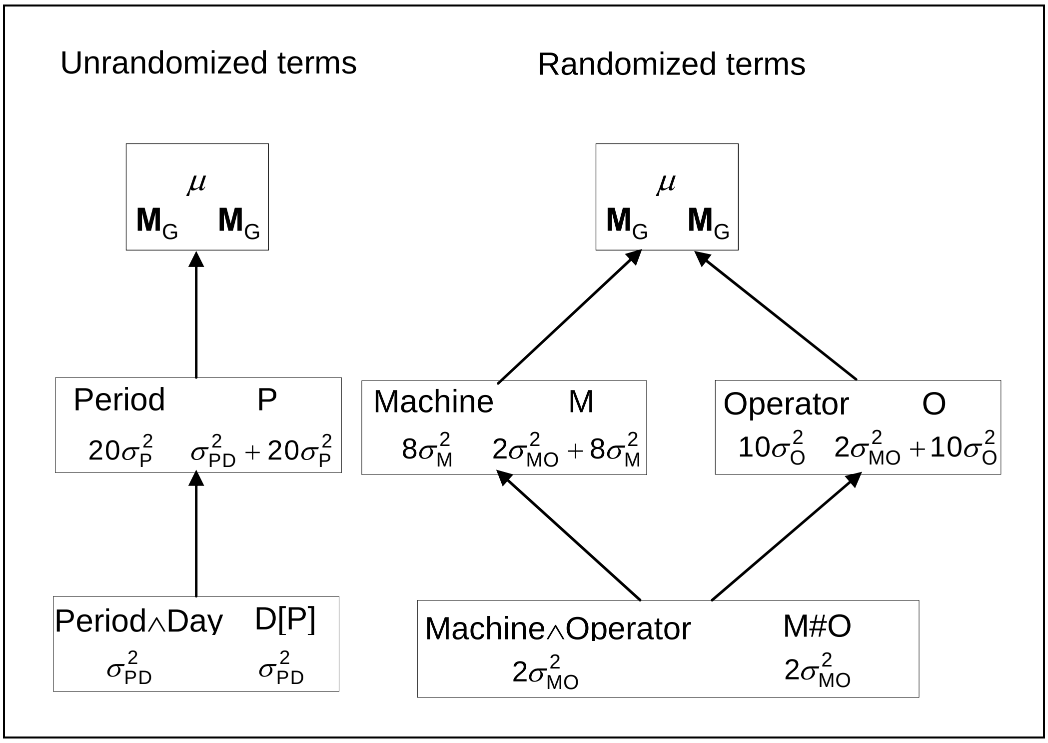
What are the Hasse diagrams of generalized-factor marginalities, with **M** and **Q** matrices, for this study?

Hasse diagrams, with M and Q matrices



What are the degrees of freedom, sums of squares and expected mean squares for the lines in the analysis of variance table based on all factors being random?

Hasse diagrams, with expected mean squares



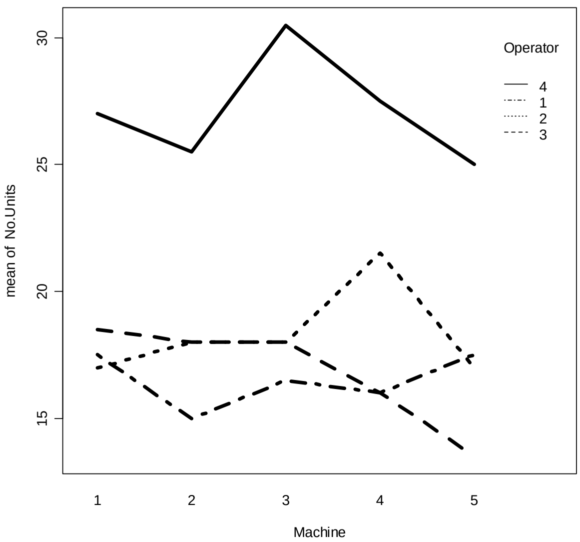
Source	df	SSq	E[MSq]
Period	1	$\mathbf{Y}'\mathbf{Q}_{\mathrm{P}}\mathbf{Y}$	$\sigma^2_{\mathrm{PD}} + 20\sigma^2_{\mathrm{P}}$
Day[Period]	38	$\mathbf{Y}'\mathbf{Q}_{\mathrm{PD}}\mathbf{Y}$	
Machine	4	$\mathbf{Y}'\mathbf{Q}_{\mathrm{M}}\mathbf{Y}$	$\sigma^2_{\mathrm{PD}} + 2\sigma^2_{\mathrm{MO}} + 8\sigma^2_{\mathrm{M}}$
Operator	3	$\mathbf{Y}'\mathbf{Q}_{\mathrm{O}}\mathbf{Y}$	$\sigma^2_{\mathrm{PD}} + 2\sigma^2_{\mathrm{MO}} + 10\sigma^2_{\mathrm{O}}$
Machine#Operator	12	$\mathbf{Y}'\mathbf{Q}_{\mathrm{MO}}\mathbf{Y}$	$\sigma^2_{\mathrm{PD}} + 2\sigma^2_{\mathrm{MO}}$
Residual	19	$\mathbf{Y}'\mathbf{Q}_{\mathrm{PD}_{\mathrm{Res}}}\mathbf{Y}$	σ^2_{PD}
Total	39	$\mathbf{Y}'\mathbf{Q}_{\mathrm{U}}\mathbf{Y}$	

The data has been saved in *Fac2Prod.dat.rda* and is available from the web site. Analyze the data using R, including diagnostic checking.

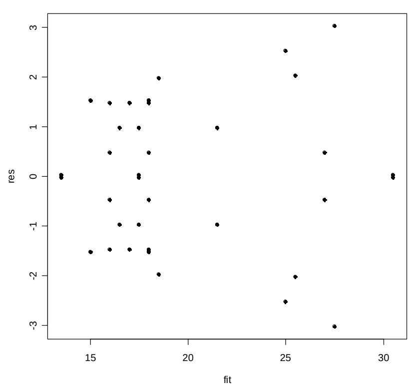
Note that from the above ANOVA table, we see that the correct denominator for the Machine and Operator Fs is Machine#Operator. These will have to be recomputed using R or Excel after obtaining the results of the `aov` *function.*

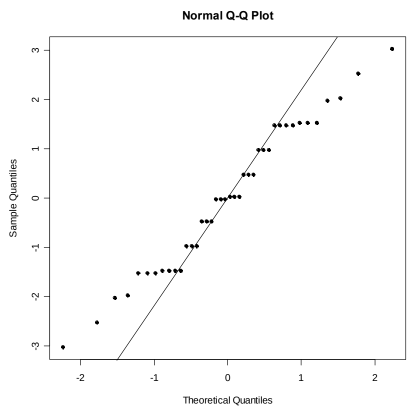
```
> attach(Fac2Prod.dat)
> interaction.plot(Machine, Operator, No.Units, lwd=4)
> Fac2Prod.aov <- aov(No.Units ~ Machine * Operator + Error(Period/Day),
+                                                                Fac2Prod.dat)
> summary(Fac2Prod.aov)

Error: Period
          Df Sum Sq Mean Sq F value Pr(>F)
Residuals  1  9.025   9.025

Error: Period:Day
                 Df Sum Sq Mean Sq F value    Pr(>F)
Machine           4  31.40    7.85   1.745    0.1818
Operator          3 753.67  251.22  55.844 1.297e-09
Machine:Operator 12  77.20    6.43   1.430    0.2352
Residuals        19  85.48    4.50
> #Compute Period F and p
> Period.F <- 9.025/4.50
> Period.p <- 1-pf(Period.F, 1, 19)
> data.frame(Period.F,Period.p)
  Period.F  Period.p
1 2.005556 0.1729098
> #
> # recalculate main effect F and p-values
> #
> MSq <- c(7.85,251.22)
> df.num <- c(4,3)
> Fvalue <- MSq/6.43
> pvalue <- 1-pf(Fvalue, df.num, 12)
> data.frame(MSq,df.num,Fvalue,pvalue)
     MSq df.num    Fvalue       pvalue
1   7.85      4  1.220840 3.526651e-01
2 251.22      3 39.069984 1.805285e-06
>
> #
> # Diagnostic checking
> #
```

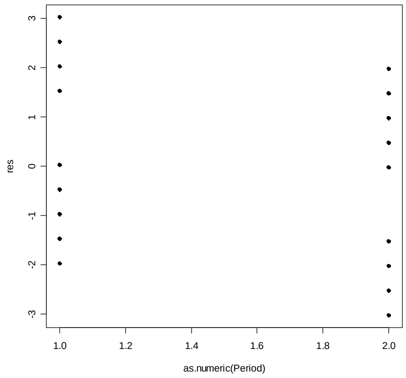
```
> res <- resid.errors(Fac2Prod.aov)
> fit <- fitted.errors(Fac2Prod.aov)
> plot(fit, res, pch=16)
> qqnorm(res, pch=16)
> qqline(res)
> plot(as.numeric(Period), res, pch = 16)
> plot(as.numeric(Machine), res, pch = 16)
> plot(as.numeric(Operator), res, pch = 16)
> tukey.1df(Fac2Prod.aov, Fac2Prod.dat, error.term="Period:Day")
** 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] 13.54856

$Tukey.F
[1] 3.390605

$Tukey.p
[1] 0.08211073

$Devn.SS
[1] 71.92644
```

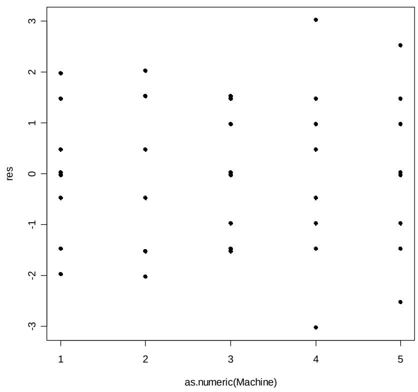


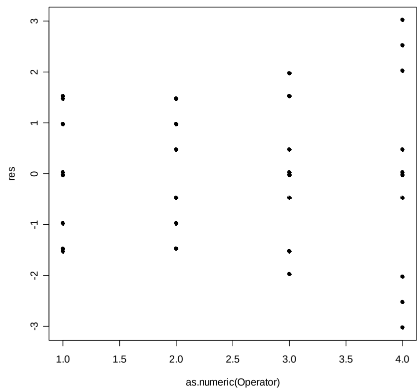
There may be an interaction.











Step 1: Set up hypotheses

a) H_0: $\sigma_O^2 = 0$

H_1: $\sigma_O^2 \neq 0$

b) H_0: $\sigma_M^2 = 0$

H_1: $\sigma_M^2 \neq 0$

c) H_0: $\sigma_{MO}^2 = 0$

H_1: $\sigma_{MO}^2 \neq 0$

Step 2: Calculate test statistics

The analysis of variance table for a factorial RCBD:

Source	df	MSq	E[MSq]	F	Prob
Period	1	9.02	$\sigma^2_{PD} + 20\sigma^2_{P}$	2.01	0.173
Day[Period]	38				
Machine	4	7.85	$\sigma^2_{PD} + 2\sigma^2_{MO} + 10\sigma^2_{O}$	1.22	0.353
Operators	3	251.22	$\sigma^2_{PD} + 2\sigma^2_{MO} + 8\sigma^2_{M}$	39.05	<0.000
Machine#Operators	12	6.43	$\sigma^2_{PD} + 2\sigma^2_{MO}$	1.43	0.235
Residual	19	4.50	σ^2_{PD}		
Total	39				

Step 3: Decide between hypotheses

The interaction between Machine and Operators is not significant; that is, the production of an Operator did not vary across Machines. However, there was variability between Operators.

The diagnostic checking is not altogether conclusive in this case. The Residual-versus-fitted-values-plot gives evidence of slight variance heterogeneity with higher variance at higher production rates. From the residuals-versus-Machine and residuals-versus-Operator plots it would appear that this is due to one Operator being more variable than the others. The normal probability plot is showing a straight-line trend and so the normality assumption appears justified. Again, the differences in variability are not large — note that in this case, where the factors are random, the equal replication does not allow us to conclude that the variance heterogeneity will have little effect on the analysis. Also, Tukey's test for nonadditivity is not appropriate in this case as there are not two noninteracting factors involved in the maximal expectation model.

VII.5 The yields of an undesirable by-product of a process were measured from 12 runs in which 2 different catalysts and 2 different pressures were used in a random order; that is each combination was replicated 3 times. This experiment was repeated at two different laboratories. The data, given below, are the percentage of by-product produced and it is available in the file *Fac2ByPr.dat.rda* from the web site.

	Catalyst	I		II	
	Laboratory	A	B	A	B
Pressure	High	53	27	40	45
		43	45	32	12
		45	57	29	69
	Low	42	32	61	54
		95	27	24	60
		60	98	11	26

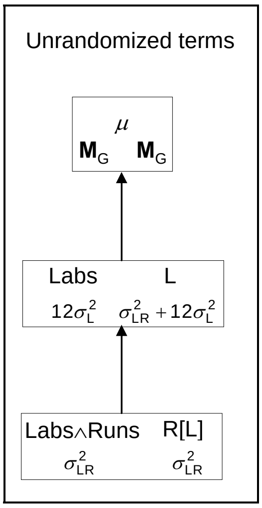
The components of this experiment are:

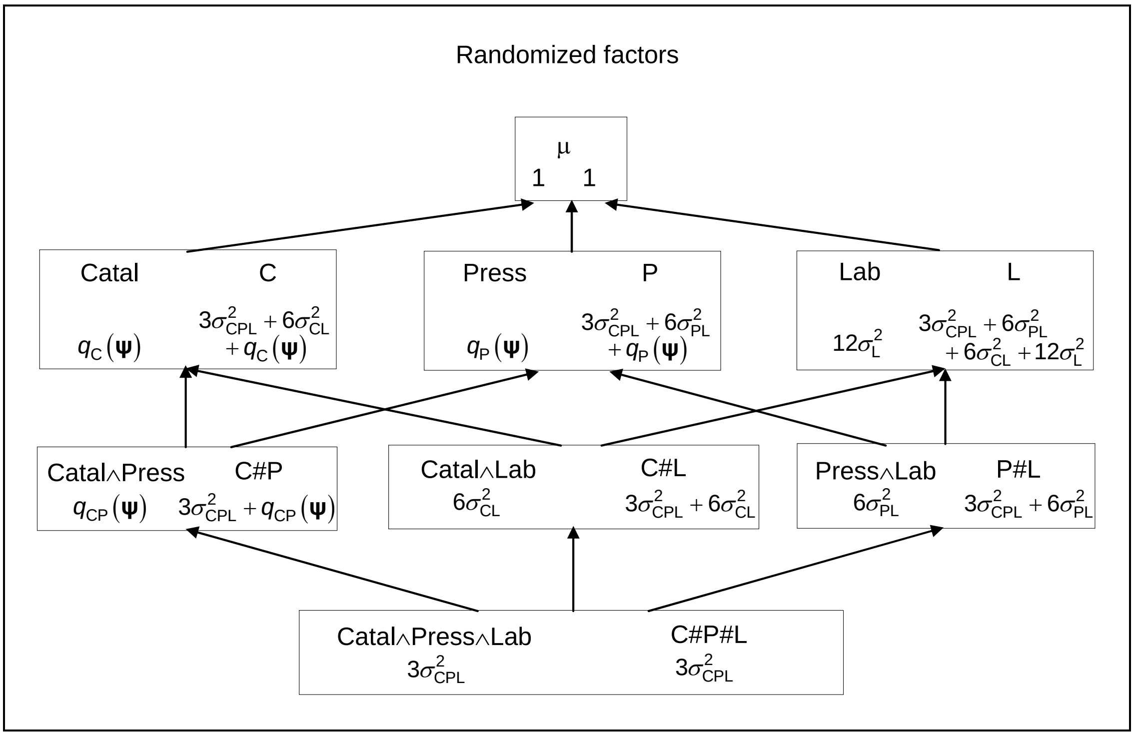
1. Observational unit – a run
2. Response variable – % By-product
3. Unrandomized factors – Laboratories, Runs
4. Randomized factors – Catalyst, Pressure
5. Type of study – Two-factor RCBD

In this experiment interactions between laboratories and treatments are likely to be of interest so that the experimental structure for this experiment would be:

Structure	Formula
unrandomized	Labs/Runs
randomized	Catalyst*Pressure*Labs

What are the degrees of freedom, sums of squares and expected mean squares for the lines in the analysis of variance table based on all unrandomized factors being random and all randomized factors being fixed?



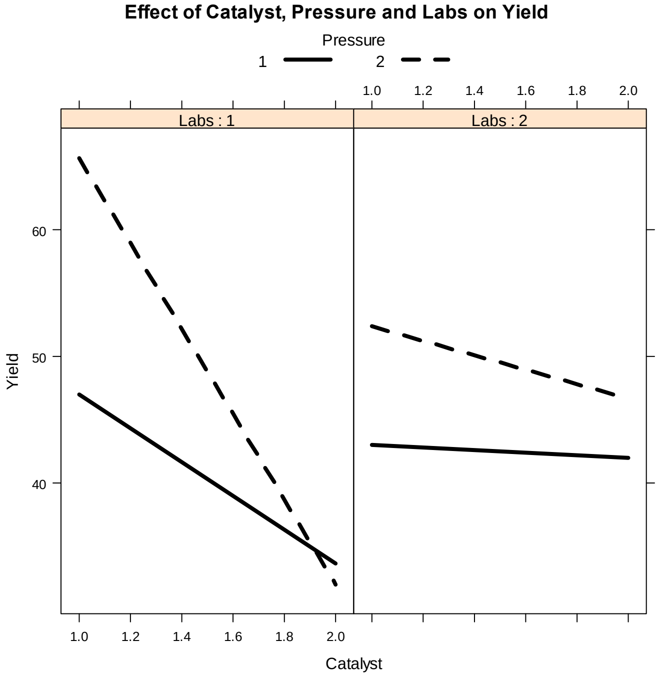


Source	df	SSq	E[MSq]
Labs	1	$\mathbf{Y}'\mathbf{Q}_{L}\mathbf{Y}$	$\sigma^2_{LR}+3\sigma^2_{CPL}+6\sigma^2_{CL}+6\sigma^2_{PL}+12\sigma^2_{L}$
Runs[Labs]	22	$\mathbf{Y}'\mathbf{Q}_{LR}\mathbf{Y}$	
Catalyst	1	$\mathbf{Y}'\mathbf{Q}_{C}\mathbf{Y}$	$\sigma^2_{LR}+3\sigma^2_{CPL}+6\sigma^2_{CL}+q_C(\psi)$
Pressure	1	$\mathbf{Y}'\mathbf{Q}_{P}\mathbf{Y}$	$\sigma^2_{LR}+3\sigma^2_{CPL}+6\sigma^2_{PL}+q_P(\psi)$
Catalyst#Pressure	1	$\mathbf{Y}'\mathbf{Q}_{CP}\mathbf{Y}$	$\sigma^2_{LR}+3\sigma^2_{CPL}+q_{CP}(\psi)$
Catalyst#Labs	1	$\mathbf{Y}'\mathbf{Q}_{CL}\mathbf{Y}$	$\sigma^2_{LR}+3\sigma^2_{CPL}+6\sigma^2_{CL}$
Pressure#Labs	1	$\mathbf{Y}'\mathbf{Q}_{PL}\mathbf{Y}$	$\sigma^2_{LR}+3\sigma^2_{CPL}+6\sigma^2_{PL}$
Catalyst#Pressure#Labs	1	$\mathbf{Y}'\mathbf{Q}_{CPL}\mathbf{Y}$	$\sigma^2_{LR}+3\sigma^2_{CPL}$
Residual	16	$\mathbf{Y}'\mathbf{Q}_{LR_{Res}}\mathbf{Y}$	σ^2_{LR}

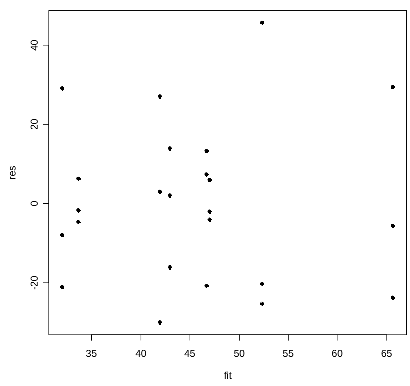
Analyze the data using R, including diagnostic checking and the examination of treatment differences. Note that in producing the exploratory interaction plots, because there are 3 factors in the randomized structure, an interaction plot for two of the factors should be produced for each level of the third factor. In this case, an interaction plot of Catalyst by Pressure for each Lab seems the natural choice. Use the nonstandard function `interaction.ABC.plot` from the *dae* library.

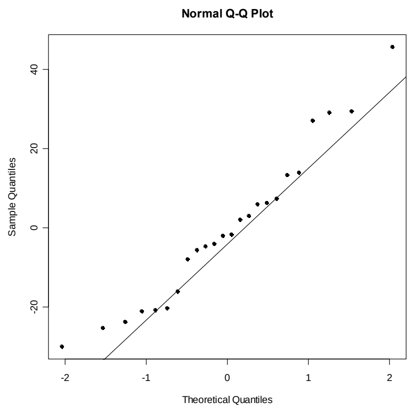
```
> attach(Fac2ByPr.dat)
> Fac2ByPr.dat
   Labs Runs Pressure Catalyst Yield
 1    1    1        1        1    53
 2    2    1        1        1    27
 3    1    2        1        2    40
 4    2    2        1        2    45
 5    1    3        1        1    43
 6    2    3        1        1    45
 7    1    4        1        2    32
 8    2    4        1        2    12
 9    1    5        1        1    45
10    2    5        1        1    57
11    1    6        1        2    29
12    2    6        1        2    69
13    1    7        2        1    42
14    2    7        2        1    32
15    1    8        2        2    61
16    2    8        2        2    54
17    1    9        2        1    95
18    2    9        2        1    27
19    1   10        2        2    24
20    2   10        2        2    60
21    1   11        2        1    60
22    2   11        2        1    98
23    1   12        2        2    11
24    2   12        2        2    26
> interaction.ABC.plot(Yield, Catalyst, Pressure, Labs, data=Fac2ByPr.dat,
+                       title="Effect of Catalyst, Pressure and Labs on Yield")
> Fac2ByPr.aov <- aov(Yield ~ Catalyst * Pressure * Labs + Error(Labs/Runs),
+                                                                 Fac2ByPr.dat)
> summary(Fac2ByPr.aov)

Error: Labs
     Df Sum Sq Mean Sq
Labs  1 12.042  12.042

Error: Labs:Runs
                       Df Sum Sq Mean Sq F value Pr(>F)
Catalyst                1 1080.0  1080.0  1.9609 0.1805
Pressure                1  360.4   360.4  0.6543 0.4304
Catalyst:Pressure       1  234.4   234.4  0.4255 0.5235
Catalyst:Labs           1  610.0   610.0  1.1076 0.3082
Pressure:Labs           1    3.4     3.4  0.0061 0.9386
Catalyst:Pressure:Labs  1   92.0    92.0  0.1671 0.6881
Residuals              16 8812.7   550.8
>
> #Compute Labs F and p
> Labs.F <- 12.042/550.8
> Labs.p <- 1-pf(Labs.F, 1, 16)
> data.frame(Labs.F,Labs.p)
      Labs.F    Labs.p
1 0.02186275 0.8843001
> #
> # Diagnostic checking
> #
> res <- resid.errors(Fac2ByPr.aov)
> fit <- fitted.errors(Fac2ByPr.aov)
> plot(fit, res, pch=16)
> qqnorm(res, pch=16)
> qqline(res)
> plot(as.numeric(Labs), res, pch = 16)
> plot(as.numeric(Catalyst), res, pch = 16)
> plot(as.numeric(Pressure), res, pch = 16)
> tukey.1df(Fac2ByPr.aov, Fac2ByPr.dat, error.term="Labs:Runs")
** 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
```

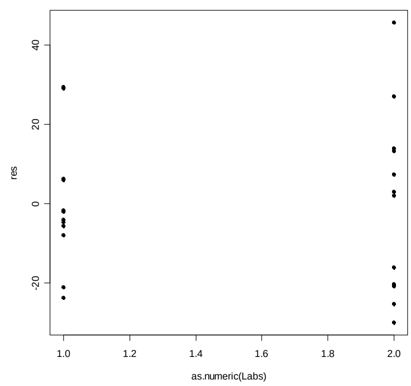
```
 in the analysis.
$Tukey.SS
[1] 392.5961

$Tukey.F
[1] 0.6993934

$Tukey.p
[1] 0.4161118

$Devn.SS
[1] 8420.07
```

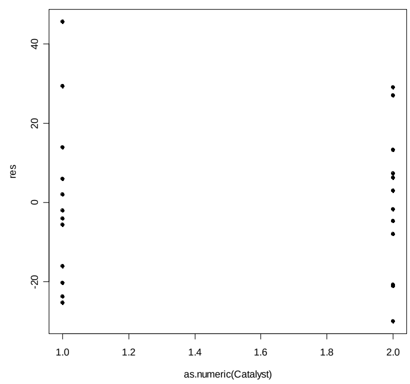


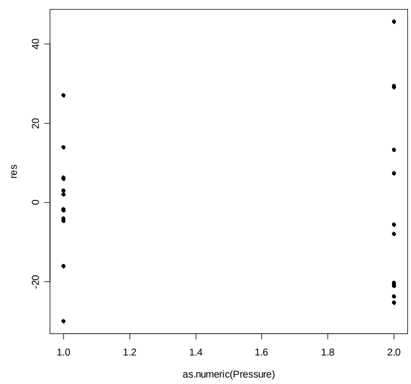
The interaction plots for the two labs appear to indicate a different interaction for the two labs.











The residuals-versus-fitted-values, residuals-versus-factor and normal probability plots appear to be satisfactory. Tukey's one-degree-of-freedom-for-nonadditivity is not appropriate.

Step 1: Set up hypotheses

a) H_0: *there is no interaction between Pressure and Catalyst*

$$\left((\alpha\beta)_{ij} - \overline{(\alpha\beta)}_{i.} - \overline{(\alpha\beta)}_{.j} + \overline{(\alpha\beta)}_{..} = 0 \quad \text{for all i,j}\right)$$

H_1: *there is an interaction between Pressure and Catalyst*

$$\left((\alpha\beta)_{ij} - \overline{(\alpha\beta)}_{i.} - \overline{(\alpha\beta)}_{.j} + \overline{(\alpha\beta)}_{..} \neq 0 \quad \text{for some i,j}\right)$$

b) H_0: $\alpha_1 = \alpha_2$

H_1: $\alpha_1 \neq \alpha_2$

c) H_0: $\beta_1 = \beta_2$

H_1: $\beta_1 \neq \beta_2$

d) H_0: $\sigma^2_{\text{PCL}} = 0$

H_1: $\sigma^2_{\text{PCL}} \neq 0$

e) H_0: $\sigma^2_{\text{PL}} = 0$

H_1: $\sigma^2_{\text{PL}} \neq 0$

f) H_0: $\sigma^2_{CL} = 0$
H_1: $\sigma^2_{CL} \neq 0$

g) H_0: $\sigma^2_{L} = 0$
H_1: $\sigma^2_{L} \neq 0$

Step 2: Calculate test statistics

Source	df	MSq	E[MSq]	F	p
Labs	1	12.0	$\sigma^2_{LR} + 3\sigma^2_{CPL} + 6\sigma^2_{CL} + 6\sigma^2_{PL} + 12\sigma^2_{L}$	0.02	0.884
Runs[Labs]	22				
Catal	1	1080.0	$\sigma^2_{LR} + 3\sigma^2_{CPL} + 6\sigma^2_{CL} + q_C(\psi)$	1.96	0.181
Press	1	360.4	$\sigma^2_{LR} + 3\sigma^2_{CPL} + 6\sigma^2_{PL} + q_P(\psi)$	0.65	0.430
Catal#Press	1	234.4	$\sigma^2_{LR} + 3\sigma^2_{CPL} + q_{CP}(\psi)$	0.43	0.523
Catal#Labs	1	610.0	$\sigma^2_{LR} + 3\sigma^2_{CPL} + 6\sigma^2_{CL}$	1.11	0.308
Press#Labs	1	3.4	$\sigma^2_{LR} + 3\sigma^2_{CPL} + 6\sigma^2_{PL}$	0.01	0.939
Catal#Press#Labs	1	92.0	$\sigma^2_{LR} + 3\sigma^2_{CPL}$	0.17	0.688
Residual	16	550.8	σ^2_{LR}		

Step 3: Decide between hypotheses

For P#L#C interaction
The P#L#C interaction is not significant.

For P#L, P#C and L#C interactions
According to the expected mean squares, the two-factor interactions should be tested against the three-factor interaction. However, as the three-factor interaction is not significant and has only one degree of freedom, it is preferable to test the two-factor interactions against the Residual. They are all are not significant.

For P, L and C
For similar reasons as the two-factor interactions, it is preferable to test the main effects against the Residual. The main effects are not significant.