CHAPTER 13

Section 13-2

13-1 a) Because factor df = total df – error df = 19-16 = 3 (and the degrees of freedom equals the number of levels minus one), 4 levels of the factor were used.

b) Because the total df = 19, there were 20 trials in the experiment. Because there are 4 levels for the factor, there were 5 replicates of each level.

c) From part (a), the factor df = 3
MS(Error) = 407.5/16 = 25.5, f = MS(Factor)/MS(Error) = 39.1/25.5 = 1.53.
From Appendix Table VI, 0.1 < P-value < 0.25

d) We fail to reject H_0. There are not significance differences in the factor level means at $\alpha = 0.05$.

13-2 a) Because the factor was tested over 4 levels and total degrees of freedom is 31, total number of observations is 31 + 1 = 32. Hence, each level has 32/4 = 8 replicates.

b) Because the factor was tested over 4 levels there are 3 degrees of freedom for factor. Because there are 31 total degrees of freedom, df(Error) = 28.

Because the MS(Factor)= 330.4716, the SS(Factor) = 3(330.4716) = 991.4148.
Because the F statistic equals MS(Factor)/MS(Error) = 4.13 = 330.4716/MS(Error). Therefore, MS(Error) = 80.0173.

Therefore, SS(Error)/df(Error) = MS(Error) = 80.0173. Therefore, SS(Error) = 28(80.0173) = 2240.484.

Therefore, SS(Total) = SS(Factor) + SS(Error) = 3231.899.

The P-value corresponds to an F = 4.13 with 3 numerator and 28 denominator degrees of freedom and this equal 0.02.

c) Because the P-value = 0.02 < 0.05, there are significant differences among the mean levels of the factor at significance level 0.05.

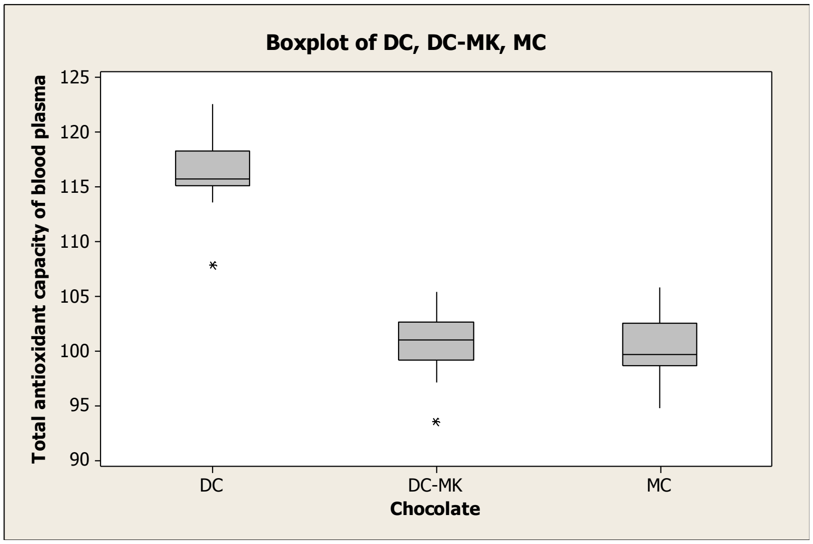
13-3 a) Because there are 29 total degrees of freedom there are 30 observations. Because there are 5 degrees of freedom for treatments there are 6 treatments. Therefore, there are 5 replicates for each treatment.

b) df(Error) = 24, MS(Error) = SS(Error)/df(Error) = 31.05/24 = 1.294
SS(Treatments) = SS(Total) – SS(Error) = 66.34 – 31.05 = 35.29
MS(Treatments) = SS(Treatments)/df(Treatments) = 35.29/5 = 7.058
F = MS(Treatments)/MS(Error) = 7.058/1.294=5.454
P-value = 0.002 from software < 0.01

c) Factor means differ significantly at significance level 0.01

d) Estimate of σ^2 = MS(Error) = 1.294.

13-4 a)



The box plots indicate that the different types of chocolate affect the total antioxidant capacity of blood plasma, especially the dark chocolate.

b) The computer result is shown below.

One-way ANOVA: DC, DC-MK, MC

```
Source  DF      SS     MS      F      P
Factor   2  1952.6  976.3  93.58  0.000
Error   33   344.3   10.4
Total   35  2296.9

S = 3.230   R-Sq = 85.01%   R-Sq(adj) = 84.10%
```

Because the P-value < 0.01 we reject H_0 and conclude that the type of chocolate has an effect on cardiovascular health at $\alpha = 0.05$ or $\alpha = 0.10$.

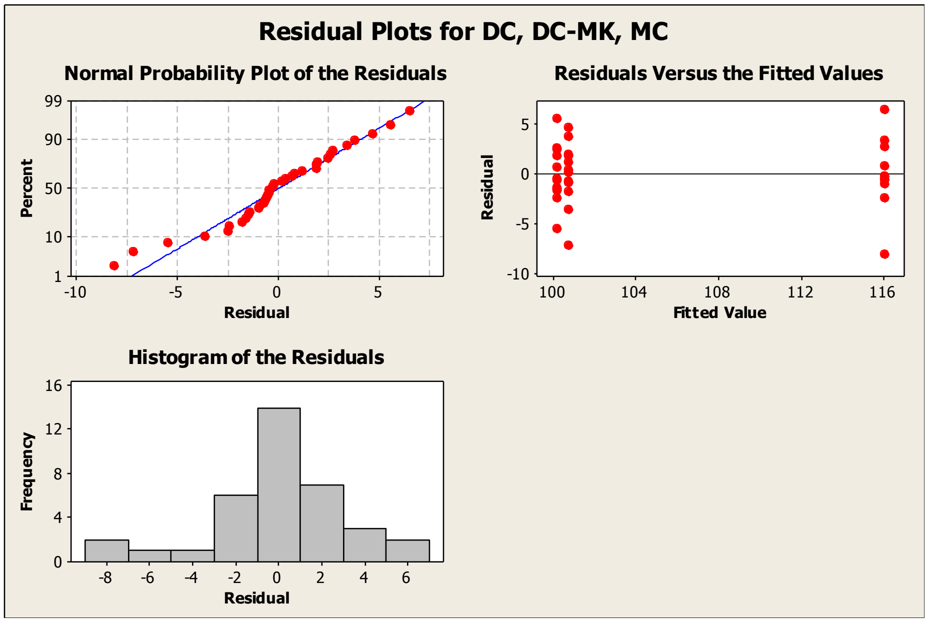
c) The computer result is shown below.

```
Fisher 95% Individual Confidence Intervals
All Pairwise Comparisons

Simultaneous confidence level = 88.02%

DC subtracted from:

         Lower   Center    Upper     -+---------+---------+---------+--------
DC+MK  -18.041  -15.358  -12.675      (---*----)
MC     -18.558  -15.875  -13.192     (----*---)
                                     -+---------+---------+---------+--------
                                  -18.0     -12.0      -6.0       0.0

DC+MK subtracted from:

     Lower  Center  Upper     -+---------+---------+---------+--------
MC  -3.200  -0.517  2.166                              (---*----)
                              -+---------+---------+---------+--------
                           -18.0     -12.0      -6.0       0.0
```

The top intervals show differences between the mean antioxidant capacity for DC+MK – DC and MC – DC. Because these intervals are entirely within the negative range (do not include zero) there are significant differences between

DC+MK and DC, and MC and DC. This implies that dark chocolate increases the mean antioxidant capacity of the subjects' blood plasma.

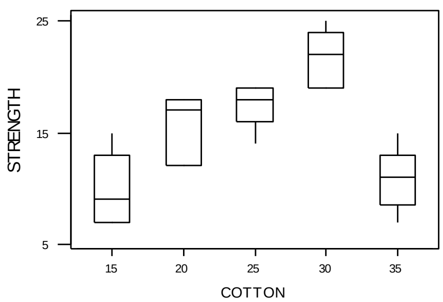
d) The normal probability plot and the residual plots show that the model assumptions are reasonable.



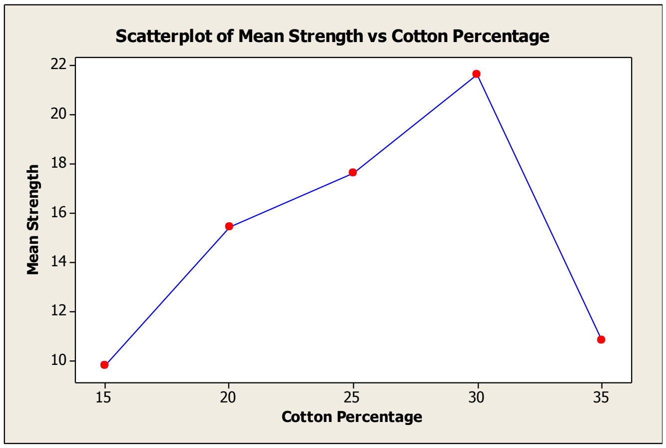
13-5 a) Analysis of Variance for STRENGTH

Source	DF	SS	MS	F	P
COTTON	4	475.76	118.94	14.76	0.000
Error	20	161.20	8.06		
Total	24	636.96			

Reject H_0 and conclude that cotton percentage affects mean breaking strength.

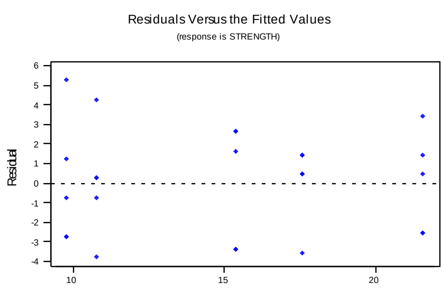


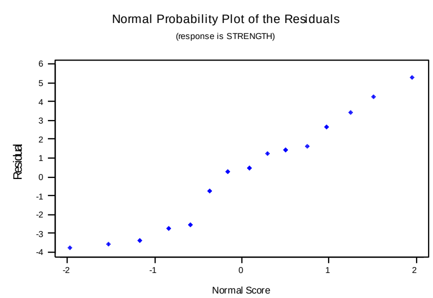
b) Tensile strength seems to increase up to 30% cotton and declines at 35% cotton.

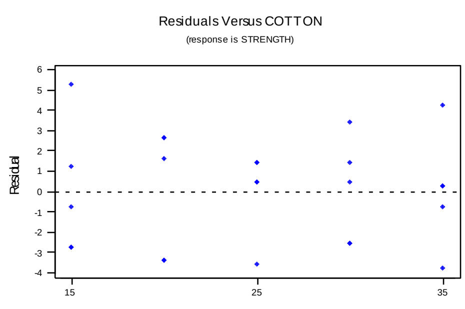



```
                                  Individual 95% CIs For Mean
                                  Based on Pooled StDev
Level        N      Mean     StDev  ------+---------+---------+---------+
15           5     9.800     3.347  (-----*----)
20           5    15.400     3.130              (----*----)
25           5    17.600     2.074                  (----*----)
30           5    21.600     2.608                          (----*----)
35           5    10.800     2.864    (-----*----)
                                    ------+---------+---------+---------+
Pooled StDev =     2.839               10.0      15.0      20.0      25.0
```

c) The normal probability plot and the residual plots show that the model assumptions are reasonable.



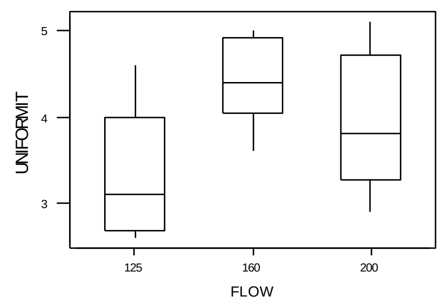




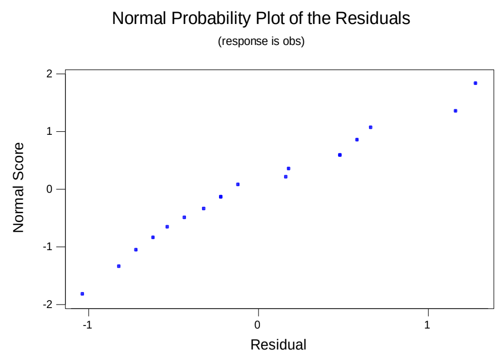
13-6 a) Analysis of Variance for FLOW

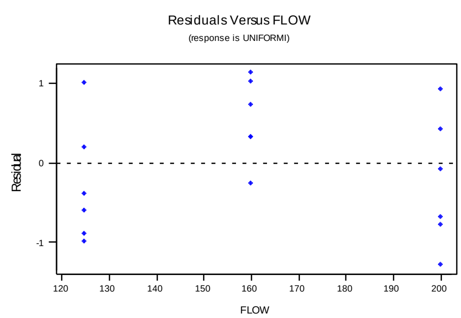
Source	DF	SS	MS	F	P
FLOW	2	3.6478	1.8239	3.59	0.053
Error	15	7.6300	0.5087		
Total	17	11.2778			

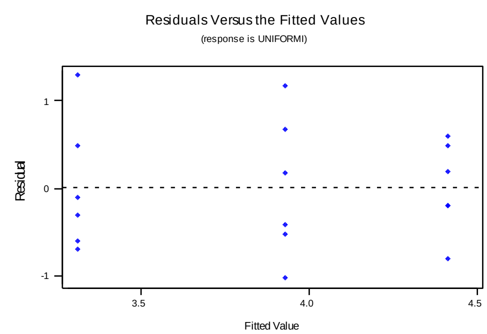
Fail to reject H_0. There is no evidence that flow rate affects etch uniformity.



b) Residuals are acceptable.







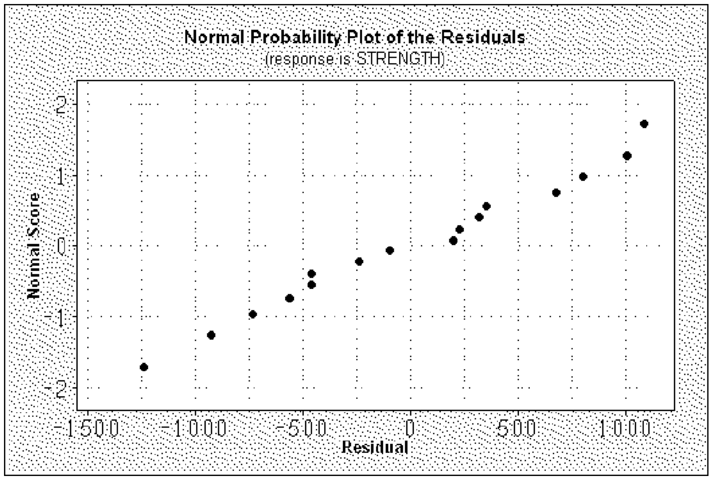
13-7 a) Analysis of Variance for STRENGTH

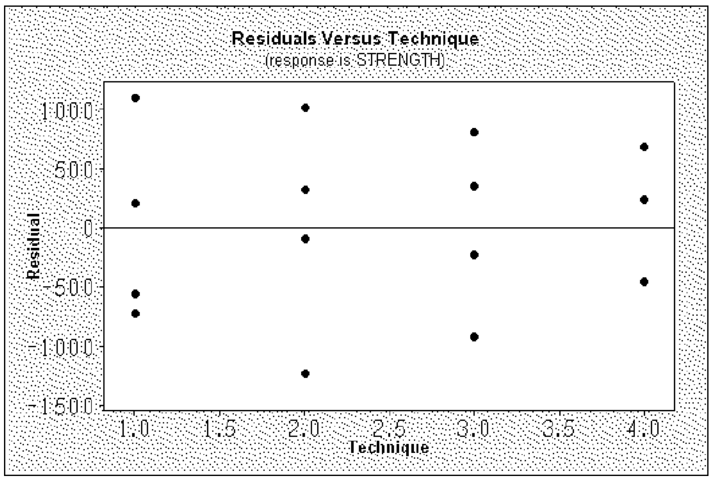
Source	DF	SS	MS	F	P
Technique	3	23103659	7701220	12.59	0.001
Error	12	7337806	611484		
Total	15	30441465			

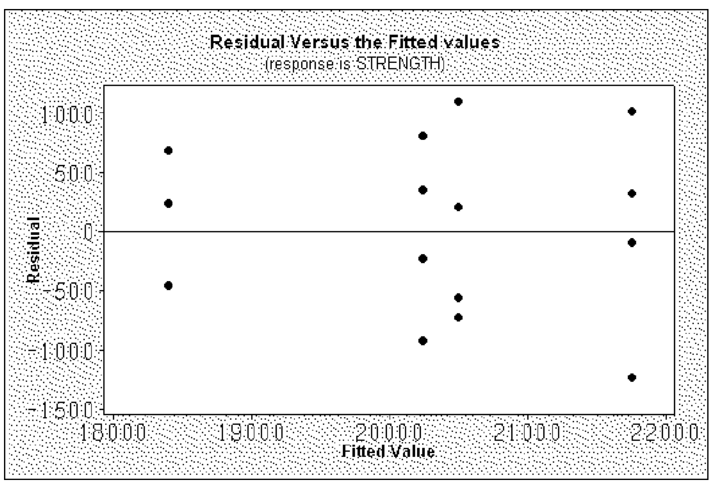
Reject H_0. Techniques affect the mean strength of the concrete.

b) *P*-value = 0.001

c) Residuals are acceptable





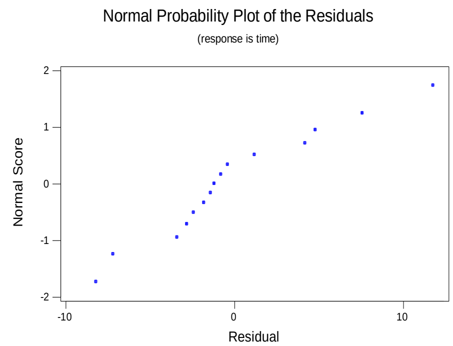


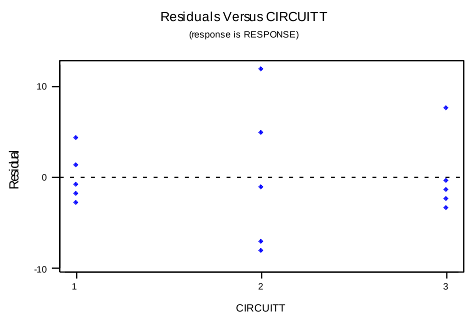
13-8 a) Analysis of Variance for CIRCUIT TYPE

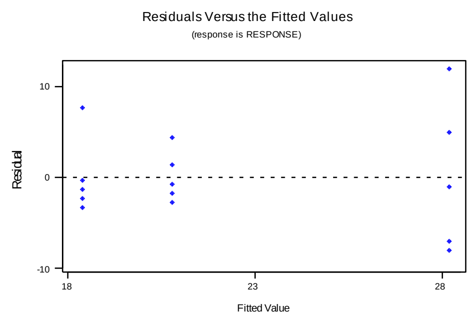
Source	DF	SS	MS	F	P
CIRCUITT	2	260.9	130.5	4.01	0.046
Error	12	390.8	32.6		
Total	14	651.7			

Reject H_0

b) There is some indication of greater variability in circuit two. There is some curvature in the normal probability plot.







c) 99% Confidence interval on the mean of circuit type 3.

$$\bar{y}_3 - t_{0.005,12}\sqrt{\frac{MS_E}{n}} \le \mu_i \le \bar{y}_3 + t_{0.005,12}\sqrt{\frac{MS_E}{n}}$$

$$18.4 - 3.055\sqrt{\frac{32.6}{5}} \le \mu_1 \le 18.4 + 3.055\sqrt{\frac{32.6}{5}}$$

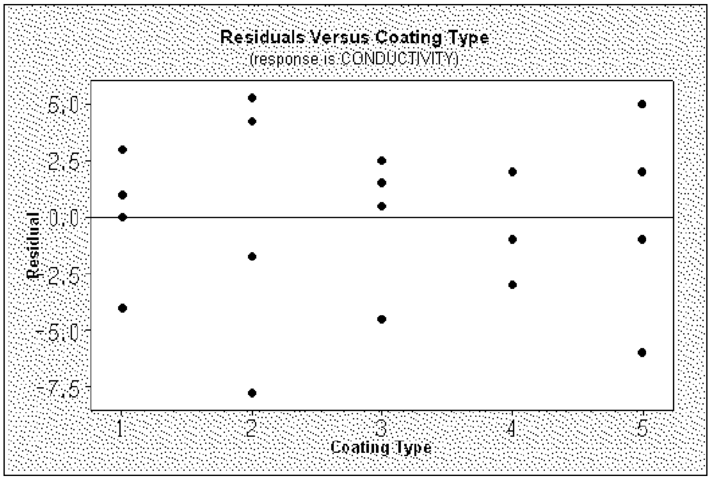
$$10.60 \le \mu_1 \le 26.20$$

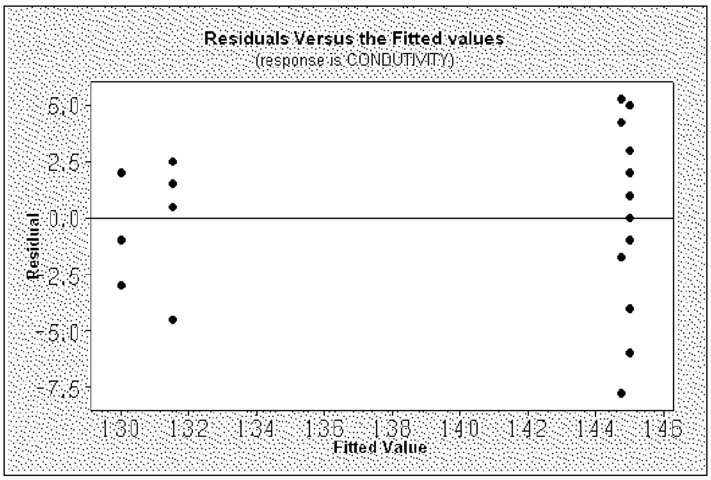
13-9 a) Analysis of Variance for CONDUCTIVITY

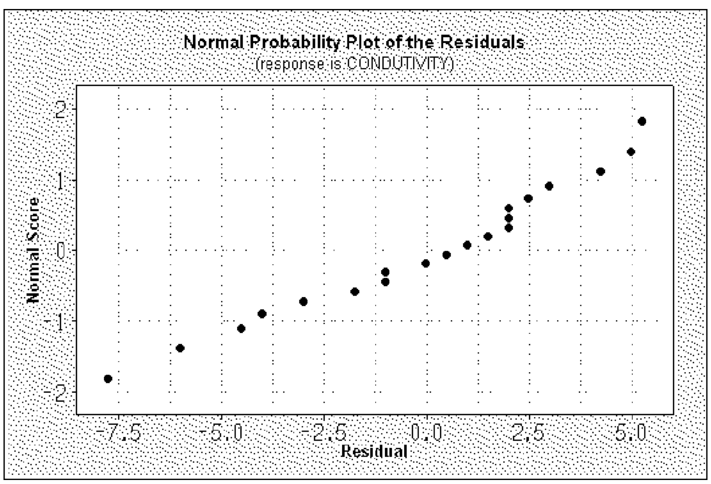
Source	DF	SS	MS	F	P
Coating Type	4	968.0	242.0	14.65	0.000
Error	15	247.8	16.5		
Total	19	1215.8			

Reject H_0, P-value $\cong 0$

b) There is some indication of that the variability of the response may be increasing as the mean response increases. There appears to be an outlier on the normal probability plot.







c) 95% Confidence interval on the mean of coating type 1

$$\bar{y}_1 - t_{0.025,15}\sqrt{\frac{MS_E}{n}} \le \mu_i \le \bar{y}_1 + t_{0.015,15}\sqrt{\frac{MS_E}{n}}$$

$$145.00 - 2.131\sqrt{\frac{16.2}{4}} \le \mu_1 \le 145.00 + 2.131\sqrt{\frac{16.2}{4}}$$

$$140.71 \le \mu_1 \le 149.29$$

99% confidence interval on the difference between the means of coating types 1 and 4.

$$\bar{y}_1 - \bar{y}_4 - t_{0.005,15}\sqrt{\frac{2MS_E}{n}} \le \mu_1 - \mu_4 \le \bar{y}_1 - \bar{y}_4 + t_{0.005,15}\sqrt{\frac{2MS_E}{n}}$$

$$(145.00 - 129.25) - 2.947\sqrt{\frac{2(16.2)}{4}} \le \mu_1 - \mu_4 \le (145.00 - 129.25) - 2.947\sqrt{\frac{2(16.2)}{4}}$$

$$7.36 \le \mu_1 - \mu_4 \le 24.14$$

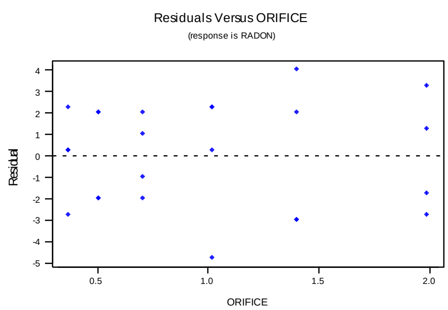
13-10 a) Analysis of Variance for ORIFICE

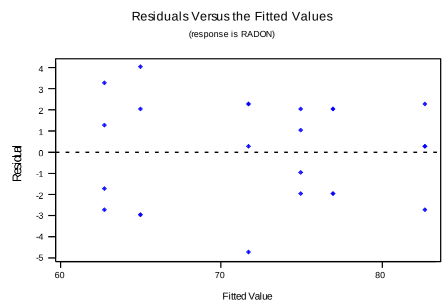
Source	DF	SS	MS	F	P
ORIFICE	5	1133.37	226.67	30.85	0.000
Error	18	132.25	7.35		
Total	23	1265.63			

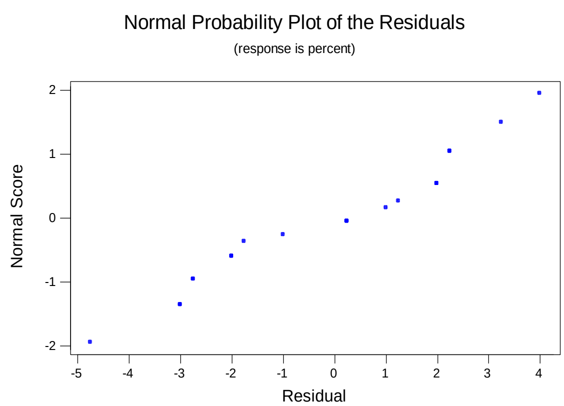
Reject H_0

b) P-value $\cong 0$

c)







d) 95% CI on the mean radon released when diameter is 1.99

$$\bar{y}_5 - t_{0.025,18}\sqrt{\frac{MS_E}{n}} \le \mu_i \le \bar{y}_5 + t_{0.025,18}\sqrt{\frac{MS_E}{n}}$$

$$62.75 - 2.101\sqrt{\frac{7.35}{4}} \le \mu_1 \le 62.75 + 2.101\sqrt{\frac{7.35}{4}}$$

$$59.90 \le \mu_1 \le 65.60$$

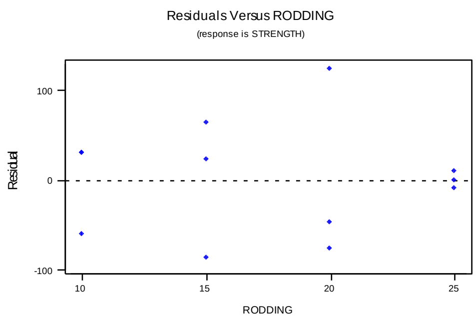
13-11 a) Analysis of Variance for STRENGTH

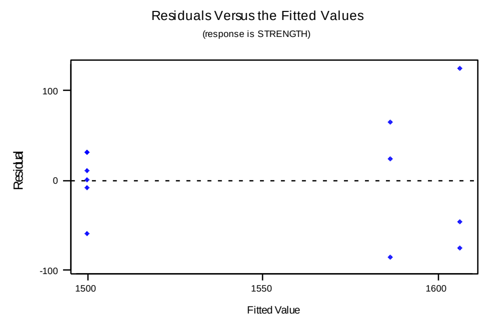
Source	DF	SS	MS	F	P
RODDING	3	28633	9544	1.87	0.214
Error	8	40933	5117		
Total	11	69567			

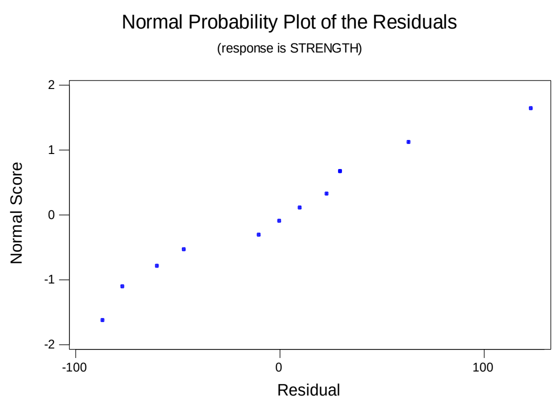
Fail to reject H_0

b) *P-value* = 0.214

c) The residual plot indicates some concern with nonconstant variance. The normal probability plot looks acceptable.







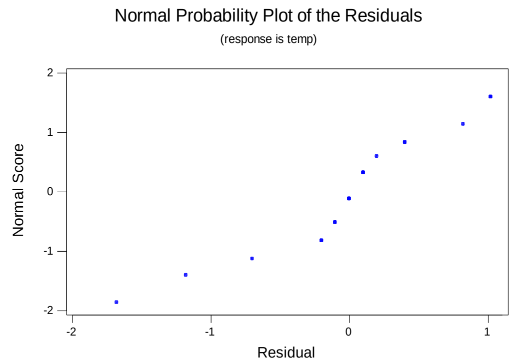
13-12 a) Analysis of Variance of PREPARATION METHOD

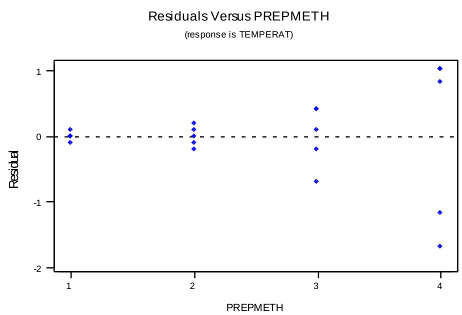
Source	DF	SS	MS	F	P
PREPMETH	3	22.124	7.375	14.85	0.000
Error	16	7.948	0.497		
Total	19	30.072			

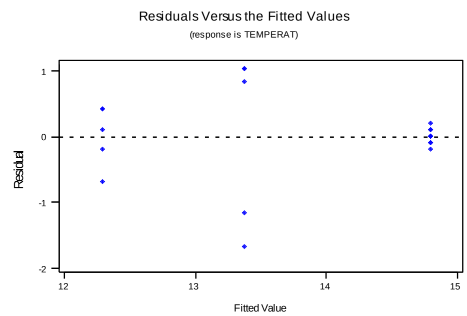
Reject H_0

b) *P*-value $\cong$ 0

c) There are some differences in the amount variability at the different preparation methods and there is some curvature in the normal probability plot. There are also some potential problems with the constant variance assumption apparent in the fitted value plot.







d) 95% Confidence interval on the mean of temperature for preparation method 1

$$\bar{y}_1 - t_{0.025,16}\sqrt{\frac{MS_E}{n}} \le \mu_i \le \bar{y}_1 + t_{0.015,16}\sqrt{\frac{MS_E}{n}}$$

$$14.8 - 2.120\sqrt{\frac{0.497}{5}} \le \mu_3 \le 14.8 + 2.120\sqrt{\frac{0.497}{5}}$$

$$14.13 \le \mu_1 \le 15.47$$

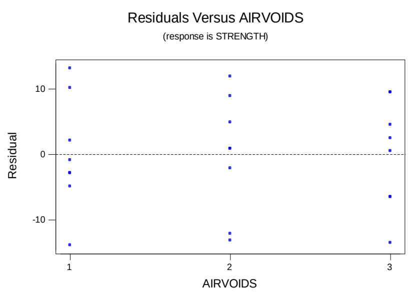
13-13 a) Analysis of Variance for STRENGTH

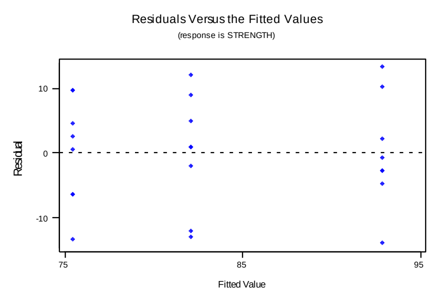
Source	DF	SS	MS	F	P
AIRVOIDS	2	1230.3	615.1	8.30	0.002
Error	21	1555.8	74.1		
Total	23	2786.0			

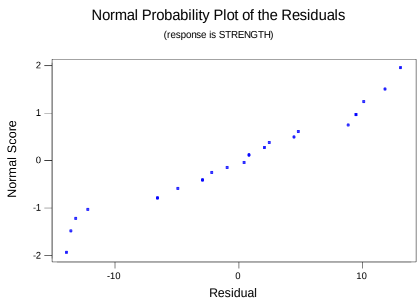
Reject H_0

b) P-value = 0.002

c) The residual plots indicate that the constant variance assumption is reasonable. The normal probability plot has some curvature in the tails but appears reasonable.







d) 95% Confidence interval on the mean of retained strength where there is a high level of air voids

$$\bar{y}_3 - t_{0.025,21}\sqrt{\frac{MS_E}{n}} \le \mu_i \le \bar{y}_3 + t_{0.015,21}\sqrt{\frac{MS_E}{n}}$$

$$75.5 - 2.080\sqrt{\frac{74.1}{8}} \le \mu_3 \le 75.5 + 2.080\sqrt{\frac{74.1}{8}}$$

$$69.17 \le \mu_1 \le 81.83$$

e) 95% confidence interval on the difference between the means of retained strength at the high level and the low levels of air voids.

$$\bar{y}_1 - \bar{y}_3 - t_{0.025,21}\sqrt{\frac{2MS_E}{n}} \le \mu_1 - \mu_3 \le \bar{y}_1 - \bar{y}_3 + t_{0.025,21}\sqrt{\frac{2MS_E}{n}}$$

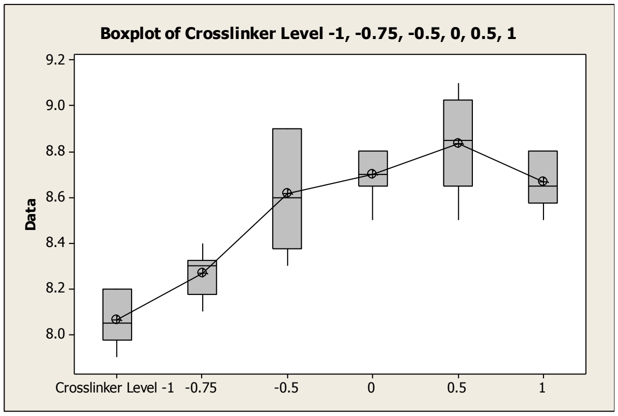
$$(92.875 - 75.5) - 2.080\sqrt{\frac{2(74.1)}{8}} \le \mu_1 - \mu_4 \le (92.875 - 75.5) + 2.080\sqrt{\frac{2(74.1)}{8}}$$

$$8.42 \le \mu_1 - \mu_4 \le 26.33$$

13-14 a)

ANOVA

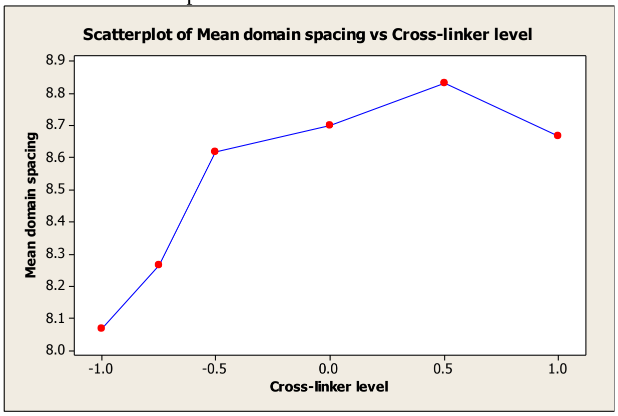
Source	DF	SS	MS	F	P
Factor	5	2.5858	0.5172	18.88	0.000
Error	30	0.8217	0.0274		
Total	35	3.4075			



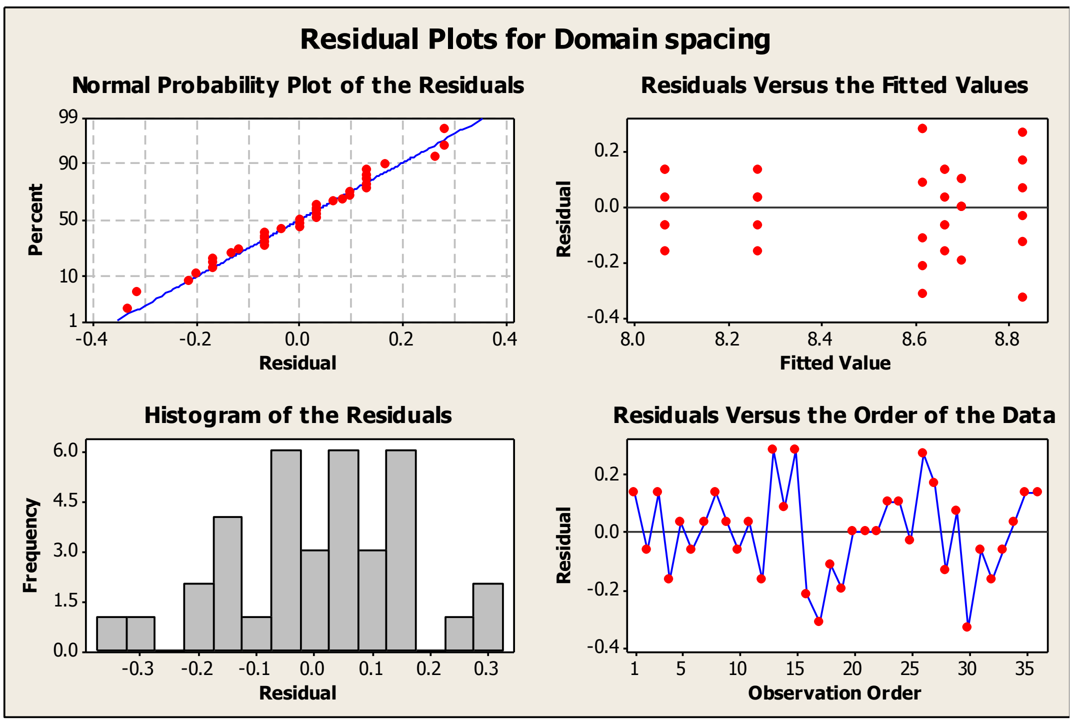
Yes, the box plot and ANOVA show that there is a difference in the cross-linker level.

b) Anova table in part (a) showed the p-value = 0.000 < α = 0.01. Therefore there is at least one level of cross-linker is different. The variability due to random error is SS_E = 0.8217

c) Domain spacing seems to increase up to the 0.5 cross-linker level and declines once cross-linker level reaches 1.

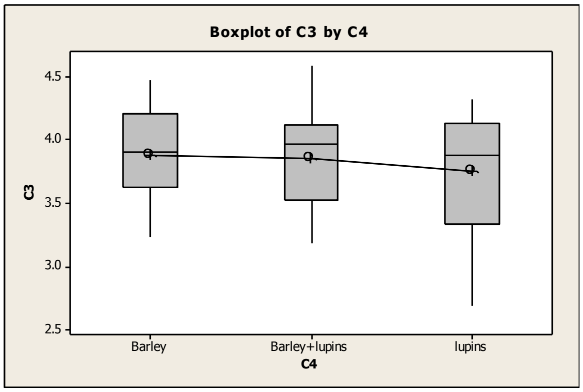


d) The normal probability plot and the residual plots show that the model assumptions are reasonable.



13-15 a) No, the diet does not affect the protein content of cow's milk.

Comparative boxplots

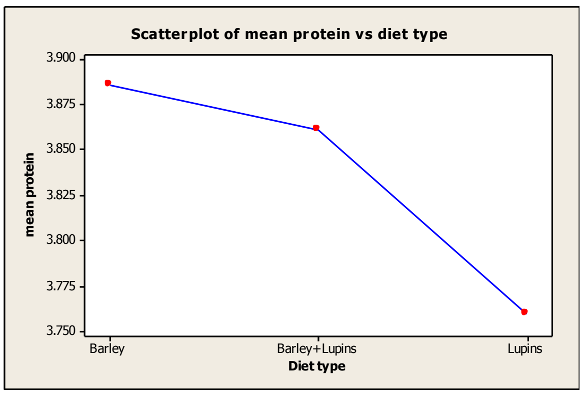


ANOVA

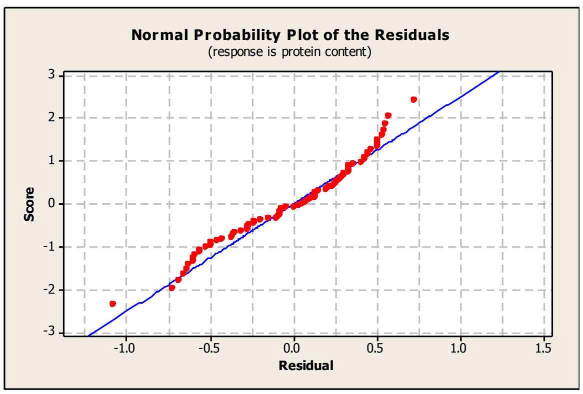
```
Source  DF      SS     MS     F      P
C4       2   0.235  0.118  0.72  0.489
Error   76  12.364  0.163
Total   78  12.599

       S = 0.4033   R-Sq = 1.87%   R-Sq(adj) = 0.00%
```

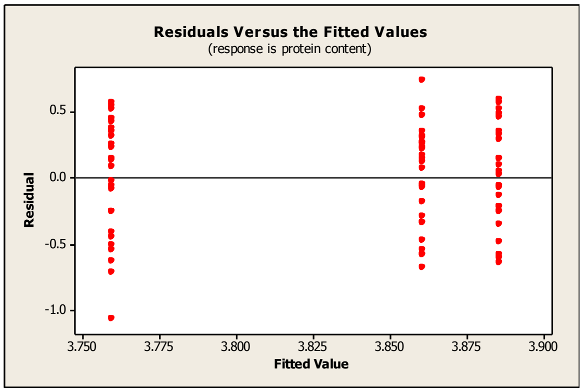
b) P-value = 0.489. The variability due to random error is SS_E = 0.146.

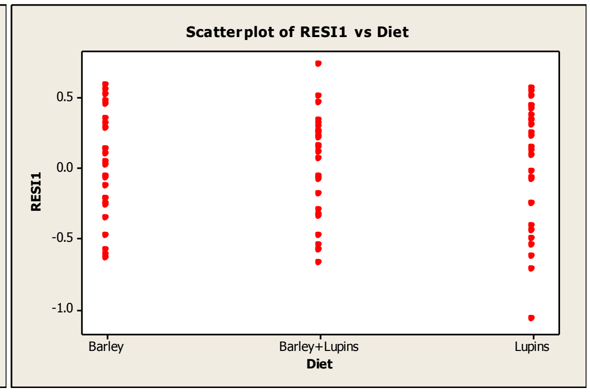
c) The Barley diet has the highest average protein content and lupins the lowest.



d) Based on the residual plots, no violation of the ANOVA assumptions is detected.







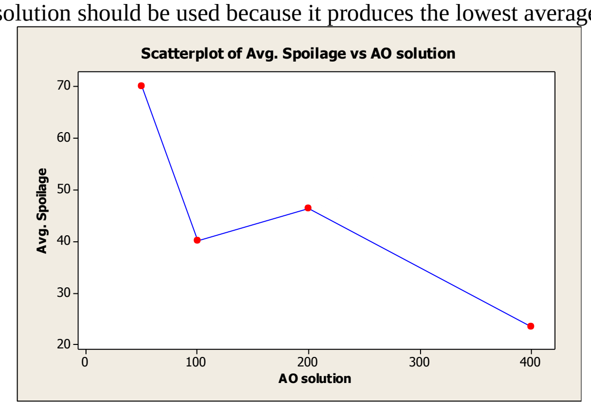
13-16 a) From the analysis of variance shown below, $F_{0.1,3,8} = 2.93 < F_0 = 3.43$, so there is difference in the spoilage percentage when using different AO solutions.

ANOVA

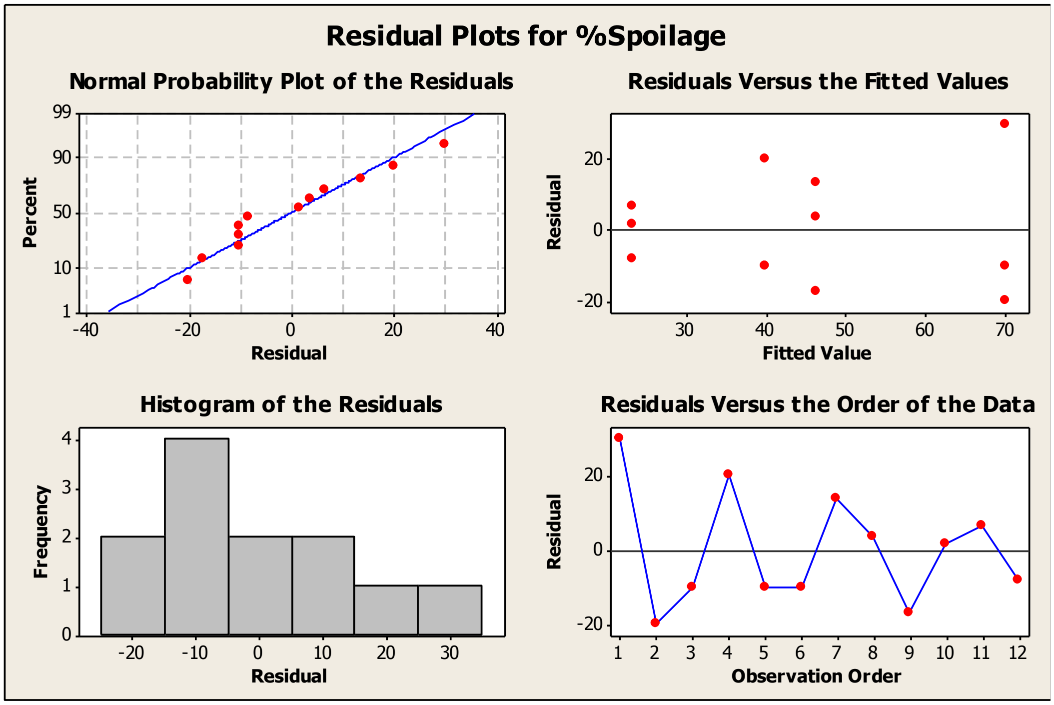
Source	DF	SS	MS	F	P
AO solutions	3	3364	1121	3.43	0.073
Error	8	2617	327		
Total	11	5981			

b) From the table above, the P-value = 0.073 and the variability due to random error is $SS_E = 2617$.

c) A 400ppm AO solution should be used because it produces the lowest average spoilage percentage.



d) The normal probability plot and the residual plots show that the model assumptions are reasonable.



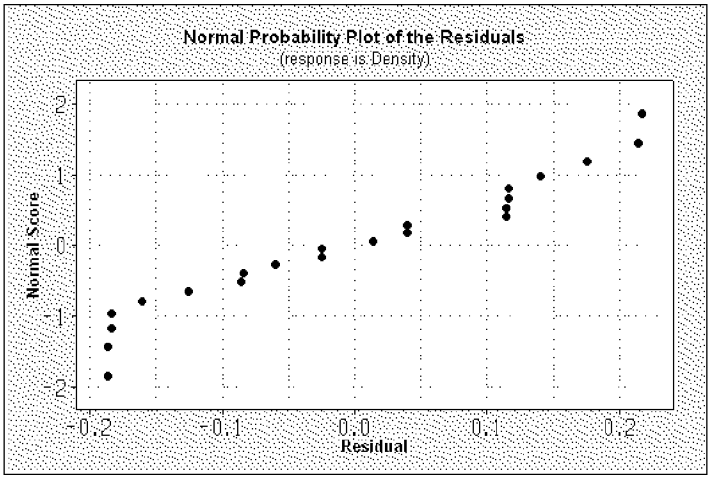
13-17 a) Analysis of Variance for TEMPERATURE

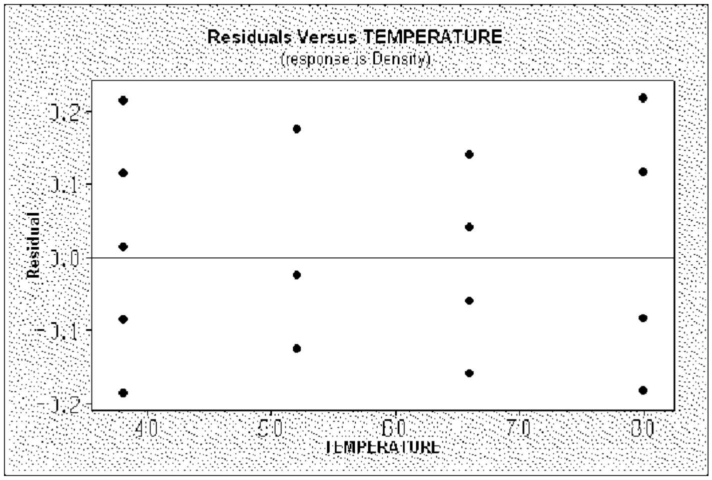
Source	DF	SS	MS	F	P
TEMPERATURE	3	0.1272	0.0424	1.93	0.162
Error	18	0.3964	0.0220		
Total	21	0.5236			

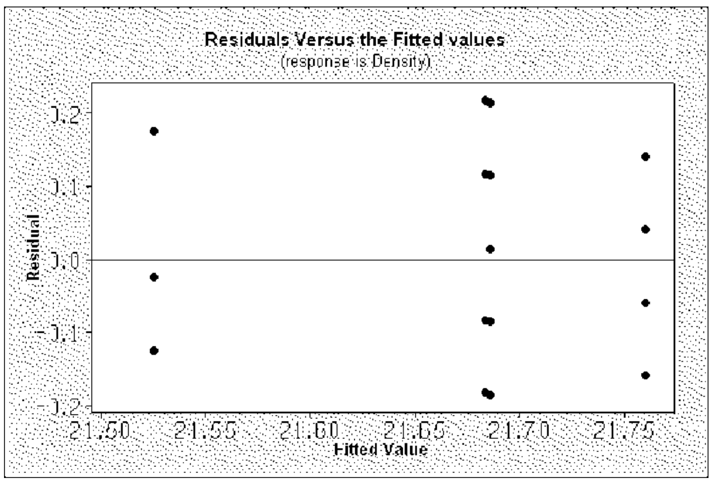
Fail to reject H_0

b) *P*-value = 0.162

c) Residuals are acceptable







13-18 a)

```
Source      DF      SS     MS      F      P
Chocolate    2  1952.6  976.3  93.58  0.000
Error       33   344.3   10.4
Total       35  2296.9

S = 3.230   R-Sq = 85.01%   R-Sq(adj) = 84.10%


                          Individual 95% CIs For Mean Based on
                          Pooled StDev
Level   N    Mean  StDev   ---+---------+---------+---------+------
DC     12  116.06   3.53                                (---*---)
DC+MK  12  100.70   3.24    (--*---)
MC     12  100.18   2.89   (--*---)
                           ---+---------+---------+---------+------
                          100.0     105.0     110.0     115.0

Pooled StDev = 3.23


Grouping Information Using Fisher Method

Chocolate   N     Mean  Grouping
DC         12  116.058  A
DC+MK      12  100.700    B
```

```
MC          12  100.183    B

Means that do not share a letter are significantly different.


Fisher 95% Individual Confidence Intervals
All Pairwise Comparisons among Levels of Chocolate

Simultaneous confidence level = 88.02%


Chocolate = DC subtracted from:

Chocolate    Lower   Center    Upper
DC+MK      -18.041  -15.358  -12.675
MC         -18.558  -15.875  -13.192

Chocolate     -+---------+---------+---------+--------
DC+MK           (---*----)
MC            (----*---)
              -+---------+---------+---------+--------
           -18.0     -12.0      -6.0       0.0


Chocolate = DC+MK subtracted from:

Chocolate   Lower  Center  Upper     -+---------+---------+--------+--------
MC         -3.200  -0.517  2.166                                (---*----)
                                     -+---------+---------+--------+--------
                                  -18.0     -12.0      -6.0       0.0
```

b) The standard error of a mean is $3.230/12^{1/2} = 0.932$. From the graphical method, group DC is significantly different from the others and this agrees with Fisher's method.

13-19
```
Fisher's pairwise comparisons
Family error rate = 0.264
Individual error rate = 0.0500
Critical value = 2.086
Intervals for (column level mean) - (row level mean)
                 15          20          25          30
      20      -9.346
              -1.854
      25     -11.546      -5.946
              -4.054       1.546
      30     -15.546      -9.946      -7.746
              -8.054      -2.454      -0.254
      35      -4.746       0.854       3.054       7.054
               2.746       8.346      10.546      14.546
```

Significant differences are detected between levels 15 and 20, 15 and 25, 15 and 30, 20 and 30, 20 and 35, 25 and 30, 25 and 35, and 30 and 35.

13-20
```
Fisher's pairwise comparisons
Family error rate = 0.117
Individual error rate = 0.0500
Critical value = 2.131

Intervals for (column level mean) - (row level mean)
           125       160
160    -1.9775
       -0.2225
250    -1.4942     -0.3942
```

```
       0.2608       1.3608
```

There are significant differences between levels 125 and 160.

13-21
```
Fisher's pairwise comparisons
Family error rate = 0.184
Individual error rate = 0.0500
Critical value = 2.179
Intervals for (column level mean) - (row level mean)
                   1           2           3
       2        -360
                 -11
       3        -137          48
                 212         397
       4         130         316          93
                 479         664         442
```

Significance differences between levels 1 and 2, 1 and 4, 2 and 3, 2 and 4, and 3 and 4.

13-22
```
Fisher's pairwise comparisons

Family error rate = 0.0251
Individual error rate = 0.0100

Critical value = 3.055

Intervals for (column level mean) - (row level mean)

             1           2

 2     -18.426
         3.626

 3      -8.626      -1.226
        13.426      20.826
```

No significant differences at α = 0.01.

13-23
```
Fisher's pairwise comparisons

Family error rate = 0.0649
Individual error rate = 0.0100
Critical value = 2.947
Intervals for (column level mean) - (row level mean)

             1           2           3           4

 2      -8.642
         8.142
 3       5.108       5.358
        21.892      22.142
 4       7.358       7.608      -6.142
        24.142      24.392      10.642
 5      -8.642      -8.392     -22.142     -24.392
         8.142       8.392      -5.358      -7.608
```

Significant differences between 1 and 3, 1 and 4, 2 and 3, 2 and 4, 3 and 5, 4 and 5.

13-24
```
Fisher's pairwise comparisons
 Family error rate = 0.189
Individual error rate = 0.0500
Critical value = 2.120
```

```
Intervals for (column level mean) - (row level mean)
                  1           2           3
      2     -0.9450
             0.9450
      3      1.5550      1.5550
             3.4450      3.4450
      4      0.4750      0.4750     -2.0250
             2.3650      2.3650     -0.1350
```

There are significant differences between levels 1 and 3, 4; 2 and 3, 4; and 3 and 4.

13-25

```
Fisher's pairwise comparisons
Family error rate = 0.118
Individual error rate = 0.0500
Critical value = 2.080
Intervals for (column level mean) - (row level mean)
                   1           2
       2       1.799
              19.701
       3       8.424      -2.326
              26.326      15.576
```

Significant differences between levels 1 and 2; and 1 and 3.

13-26

a) $LSD = t_{0.025,25}\sqrt{\dfrac{2MS_E}{b}} = 2.042\sqrt{\dfrac{2\times 0.0274}{6}} = 0.1952$

```
Fisher 95% Individual Confidence Intervals
All Pairwise Comparisons among Levels of Cross-linker level

Simultaneous confidence level = 65.64%

Cross-linker level = -0.5 subtracted from:

Cross-linker
level            Lower   Center    Upper
-0.75          -0.5451  -0.3500  -0.1549
-1             -0.7451  -0.5500  -0.3549
0              -0.1118   0.0833   0.2785
0.5             0.0215   0.2167   0.4118
1              -0.1451   0.0500   0.2451

Cross-linker
level          ---------+---------+---------+---------+
-0.75                  (---*---)
-1                 (---*---)
0                               (---*---)
0.5                               (---*---)
1                              (---*---)
               ---------+---------+---------+---------+
                     -0.50      0.00      0.50      1.00


Cross-linker level = -0.75 subtracted from:

Cross-linker
level            Lower   Center    Upper
-1             -0.3951  -0.2000  -0.0049
0               0.2382   0.4333   0.6285
0.5             0.3715   0.5667   0.7618
```

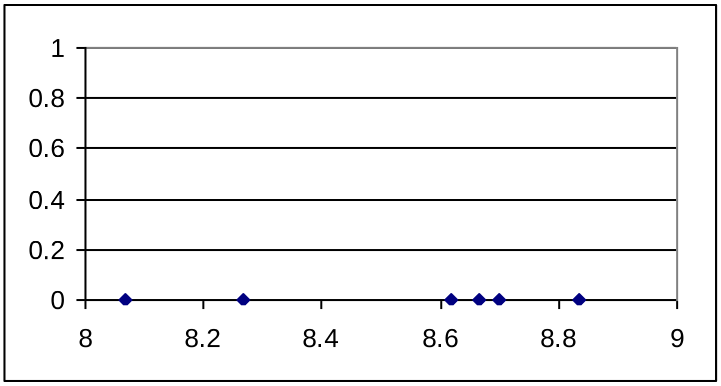
```
1              0.2049   0.4000   0.5951

Cross-linker
level          ---------+---------+---------+---------+
-1                        (---*---)
0                                      (---*---)
0.5                                      (---*---)
1                                     (---*---)
               ---------+---------+---------+---------+
                     -0.50      0.00      0.50      1.00


Cross-linker level = -1 subtracted from:

Cross-linker
level           Lower  Center   Upper  ---------+---------+---------+---------+
0              0.4382  0.6333  0.8285                             (---*---)
0.5            0.5715  0.7667  0.9618                               (---*---)
1              0.4049  0.6000  0.7951                            (---*---)
                                  ---------+---------+---------+---------+
                                        -0.50      0.00      0.50      1.00


Cross-linker level = 0 subtracted from:

Cross-linker
level            Lower   Center   Upper
0.5            -0.0618   0.1333  0.3285
1              -0.2285  -0.0333  0.1618

Cross-linker
level          ---------+---------+---------+---------+
0.5                              (---*---)
1                            (---*---)
               ---------+---------+---------+---------+
                     -0.50      0.00      0.50      1.00


Cross-linker level = 0.5 subtracted from:

Cross-linker
level            Lower   Center   Upper
1              -0.3618  -0.1667  0.0285

Cross-linker
level          ---------+---------+---------+---------+
1                          (---*---)
               ---------+---------+---------+---------+
                     -0.50      0.00      0.50      1.00
```

Cross-linker levels -0.5, 0, 0.5 and 1 are not detected to differ. Cross-linker levels -0.75 and -1 are not detected to differ from one other, but both are significantly different to the others.

b) The mean values are
8.0667, 8.2667, 8.6167, 8.7, 8.8333, 8.6667

$$\hat{\sigma}_{\bar{X}} = \sqrt{\frac{MS_E}{b}} = \sqrt{\frac{0.0274}{6}} = 0.0676$$

The width of a scaled normal distribution is 6(0.0676) = 0.405



With a scaled normal distribution over this plot, the conclusions are similar to those from the LSD method.

13-27 a) There is no significant difference in protein content between the three diet types.

```
Fisher 99% Individual Confidence Intervals
All Pairwise Comparisons among Levels of C4

Simultaneous confidence level = 97.33%

C4 = Barley subtracted from:

C4                Lower   Center   Upper
Barley+lupins   -0.3207  -0.0249  0.2709
lupins          -0.4218  -0.1260  0.1698

C4              -------+---------+---------+---------+--
Barley+lupins       (-----------*-----------)
lupins          (-----------*-----------)
                -------+---------+---------+---------+--
                    -0.25      0.00      0.25      0.50

C4 = Barley+lupins subtracted from:

C4        Lower   Center   Upper  -------+---------+---------+---------+--
lupins  -0.3911  -0.1011  0.1889   (-----------*-----------)
                                  -------+---------+---------+---------+--
                                      -0.25      0.00      0.25      0.50
```

b) The mean values are: 3.886, 3.8611, 3.76 (barley, b+l, lupins)

```
From the ANOVA the estimate of σ can be obtained
Source  DF      SS     MS     F      P
C4       2   0.235  0.118  0.72  0.489
Error   76  12.364  0.163
Total   78  12.599

S = 0.4033   R-Sq = 1.87%   R-Sq(adj) = 0.00%
The minimum sample size could be used to calculate the standard error of a
sample mean
```

$$\hat{\sigma}_{\bar{X}} = \sqrt{\frac{MS_E}{b}} = \sqrt{\frac{0.163}{25}} = 0.081$$

The graph would not show any differences between the diets.

13-28 $\bar{\mu} = 57.5$, $\tau_1 = -2.5$, $\tau_2 = 2.5$, $\tau_3 = -2.5$, $\tau_4 = 2.5$.

$$\Phi^2 = \frac{n\left(\sum_{i=1}^{a} \tau_i^2\right)}{a\sigma 2} = \frac{n(25)}{4(25)} = \frac{n}{4}, \qquad a-1=3 \qquad a(n-1)=4(n-1)$$

Various choices for *n* yield:

n	Φ^2	Φ	$a(n-1)$	Power = $1-\beta$
8	2	1.4	28	0.54
16	4	2	60	0.91

Therefore, *n* = 16 is needed.

13-29 $\bar{\mu} = 192$, $\tau_1 = -17$, $\tau_2 = -12$, $\tau_3 = 3$, $\tau_4 = 8$, $\tau_5 = 18$.

$$\Phi^2 = \frac{n(830)}{5(100)} = 1.66n, \qquad a-1=4 \qquad a(n-1)=5(n-1)$$

Various choices for *n* yield:

n	Φ^2	Φ	$a(n-1)$	Power = $1-\beta$
4	6.64	2.6	15	0.915
5	8.3	2.9	20	0.983

Therefore, *n* = 5 is needed.

<u>Section 13-3</u>

13-30 a)

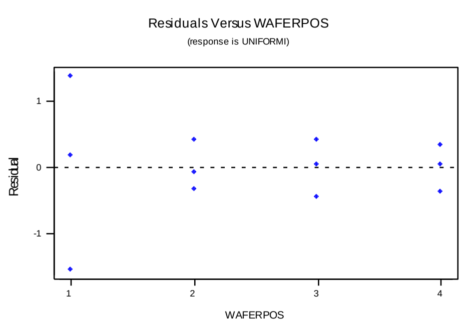
```
Analysis of Variance for UNIFORMITY
Source      DF        SS        MS       F       P
WAFERPOS     3    16.220     5.407    8.29   0.008
Error        8     5.217     0.652
Total       11    21.437
```

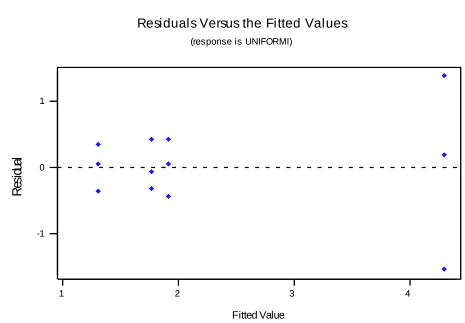
Reject H_0, and conclude that there are significant differences among wafer positions.

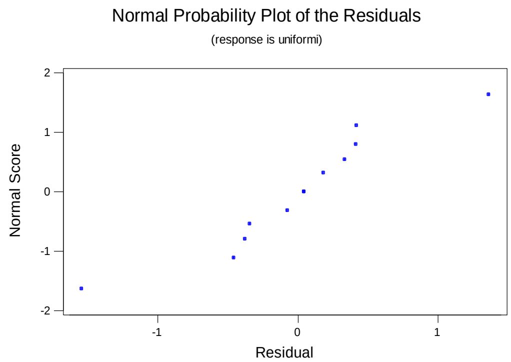
b) $\hat{\sigma}_\tau^2 = \dfrac{MS_{Treatments} - MS_E}{n} = \dfrac{5.407 - 0.652}{3} = 1.585$

c) $\hat{\sigma}^2 = MS_E = 0.652$

d) Greater variability at wafer position 1. There is some slight curvature in the normal probability plot.







13-31 a) Analysis of Variance for OUTPUT

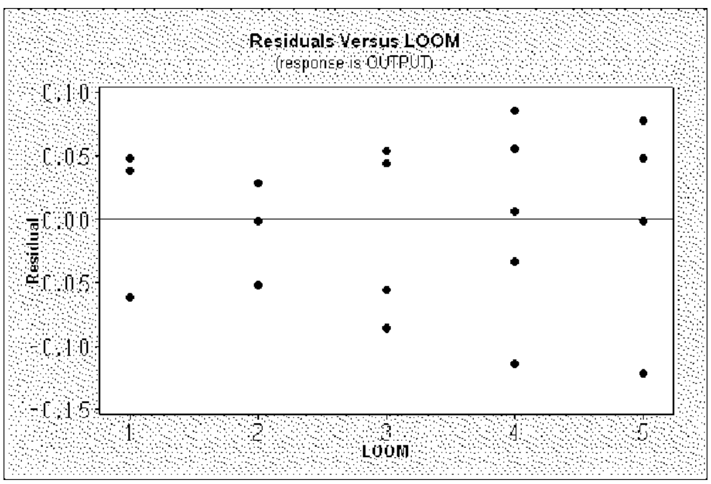
Source	DF	SS	MS	F	P
LOOM	4	0.10074	0.02519	6.11	0.002
Error	20	0.08248	0.00412		
Total	24	0.18322			

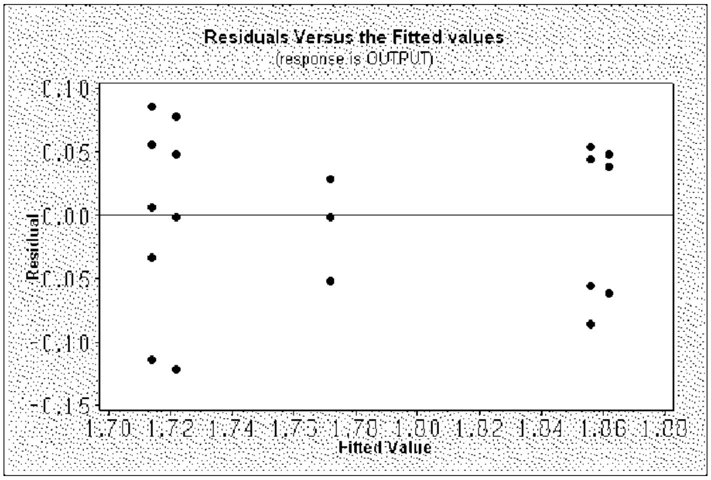
Reject H_0, there are significant differences among the looms.

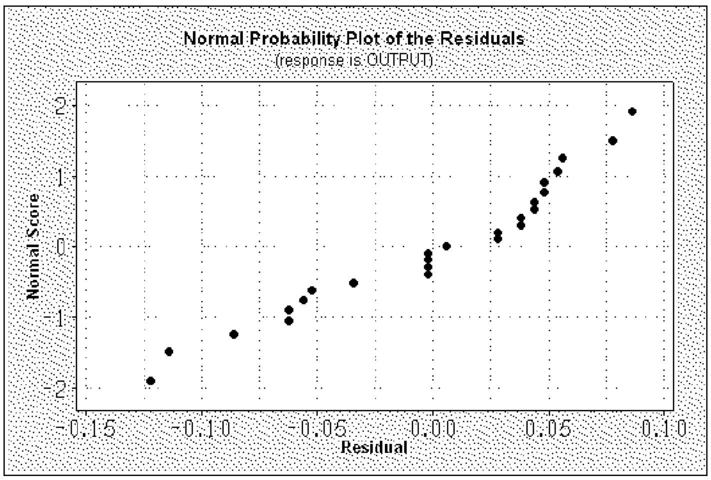
b) $\hat{\sigma}_{\tau}^2 = \dfrac{MS_{Treatments} - MS_E}{n} = \dfrac{0.02519 - 0.00412}{5} = 0.00421$

c) $\hat{\sigma}^2 = MS_E = 0.00412$

d) Residuals are acceptable







13-32 a) Yes, the different batches of raw material significantly affect mean yield at α = 0.01 because the P-value is small.

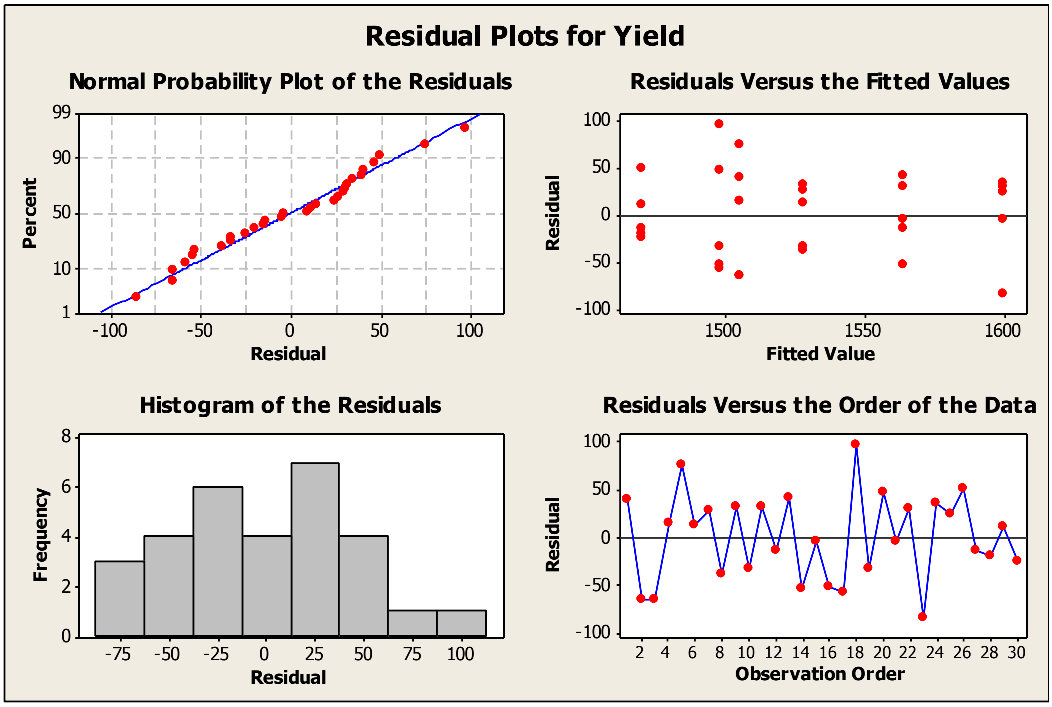
Source	DF	SS	MS	F	P
Batch	5	56358	11272	4.60	0.004
Error	24	58830	2451		
Total	29	115188			

b) Variability between batches

$$\hat{\sigma}_\tau^2 = \frac{MS_{Treatments} - MS_E}{n} = \frac{11272 - 2451}{5} = 1764.2$$

c) Variability within batches $\hat{\sigma}^2 = MSE = 2451$

d) The normal probability plot and the residual plots show that the model assumptions are reasonable.



13-33 a) Analysis of Variance for BRIGHTNENESS

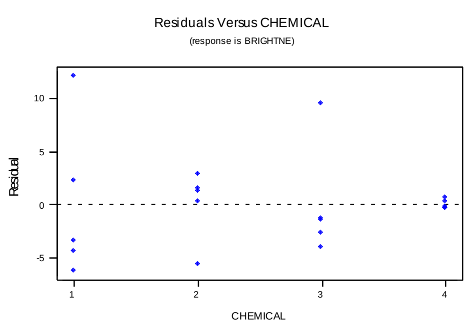
Source	DF	SS	MS	F	P
CHEMICAL	3	54.0	18.0	0.75	0.538
Error	16	384.0	24.0		
Total	19	438.0			

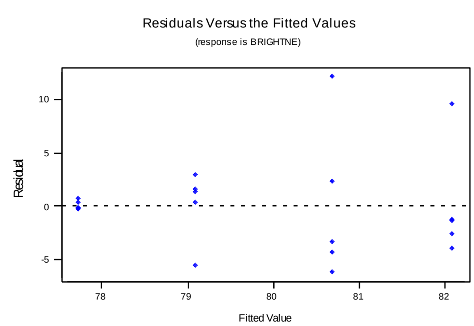
Fail to reject H_0, there is no significant difference among the chemical types.

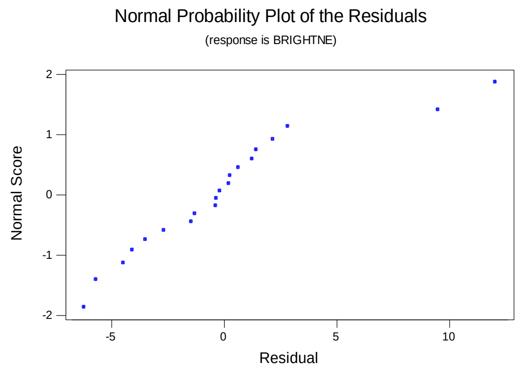
b) $\hat{\sigma}_\tau^2 = \dfrac{18.0 - 24.0}{5} = -1.2$ set equal to 0

c) $\hat{\sigma}^2 = 24.0$

d) Variability is smaller in chemical 4. There is some curvature in the normal probability plot.







13-34 a) $\hat{\sigma}^2_{total} = \hat{\sigma}^2_{position} + \hat{\sigma}^2 = 2.237$

b) $\dfrac{\hat{\sigma}^2_{position}}{\hat{\sigma}^2_{total}} = 0.709$

c) It could be reduced to 0.6522. This is a reduction of approximately 71%.

13-35 a) Instead of testing the hypothesis that the individual treatment effects are zero, we are testing whether there is variability in protein content between all diets.

$$H_0 : \sigma^2_\tau = 0$$
$$H_1 : \sigma^2_\tau \neq 0$$

b) The statistical model is

$$y = \mu + \tau_i + \varepsilon_{ij} \begin{cases} i = 1,2,...,a \\ j = 1,2,...,n \end{cases}$$
$$\varepsilon_i \sim N(0,\sigma^2) \text{ and } \tau_i \sim N(0,\sigma^2_\tau)$$

c) The **last TWO observations were omitted** from two diets to generate equal sample sizes with n = 25.

ANOVA: Protein versus DietType

```
Analysis of Variance for Protein

Source     DF       SS      MS     F      P
DietType    2   0.2689  0.1345  0.82  0.445
Error      72  11.8169  0.1641
Total      74  12.0858


S = 0.405122   R-Sq = 2.23%   R-Sq(adj) = 0.00%
```

$$\sigma^2 = MS_E = 0.1641$$
$$\sigma^2_\tau = \frac{MS_{tr} - MS_E}{n} = \frac{0.1345 - 0.1641}{25} = -0.001184$$

Section 13-4

13-36 a) $MS_{factor} = \dfrac{SS_{factor}}{DF_{factor}}$, $DF_{factor} = \dfrac{SS_{factor}}{MS_{factor}} = \dfrac{178.957}{59.652} = 3$

The levels of the factor = DF for the factor + 1 = 3 + 1 = 4. Therefore, 4 levels of the factor are used in this experiment. This can also be obtained from the result that the error degrees of freedom equal the product of the degrees of freedom for factor and block.

Let dfF and dfB denote the degrees of freedom for factors and blocks, respectively. Therefore, the total degrees of freedom = 15 = dfT + dfB + (dfT)(dfB) = dfT + 3 + 3(dfT). Therefore dfT = 3.

b) Because the number of blocks = DF of block +1 = 3 + 1 =4. There are 4 blocks used in this experiment.

c) From part a), DF factor = 3.

$$F = \frac{MS_{factor}}{MS_{error}} = \frac{59.652}{6.113} = 9.758$$

P-value =0.003

$DF_{error} = DF_{Total} - DF_{Factor} - DF_{Block} = 15 - 3 - 3 = 9.$

$MS_{block} = \dfrac{SS_{block}}{DF_{block}}$, $SS_{block} = MS_{block}\, DF_{block} = 6.113\ (9) = 55.017$

d) Because the P-value < 0.01, we reject H_0. There are significance differences in the factor level means at $\alpha = 0.05$ or $\alpha = 0.01$.

13-37 The output from computer software follows.

```
Source  DF       SS        MS       F      P
Factor   2  1952.64  976.322  147.35  0.000
Block   11   198.54   18.049    2.72  0.022
Error   22   145.77    6.626
Total   35  2296.95

S = 2.574   R-Sq = 93.65%   R-Sq(adj) = 89.90%
```

Because the *P*-value for the factor is near zero, there are significant differences in the factor level means at $\alpha = 0.05$ or $\alpha = 0.01$.

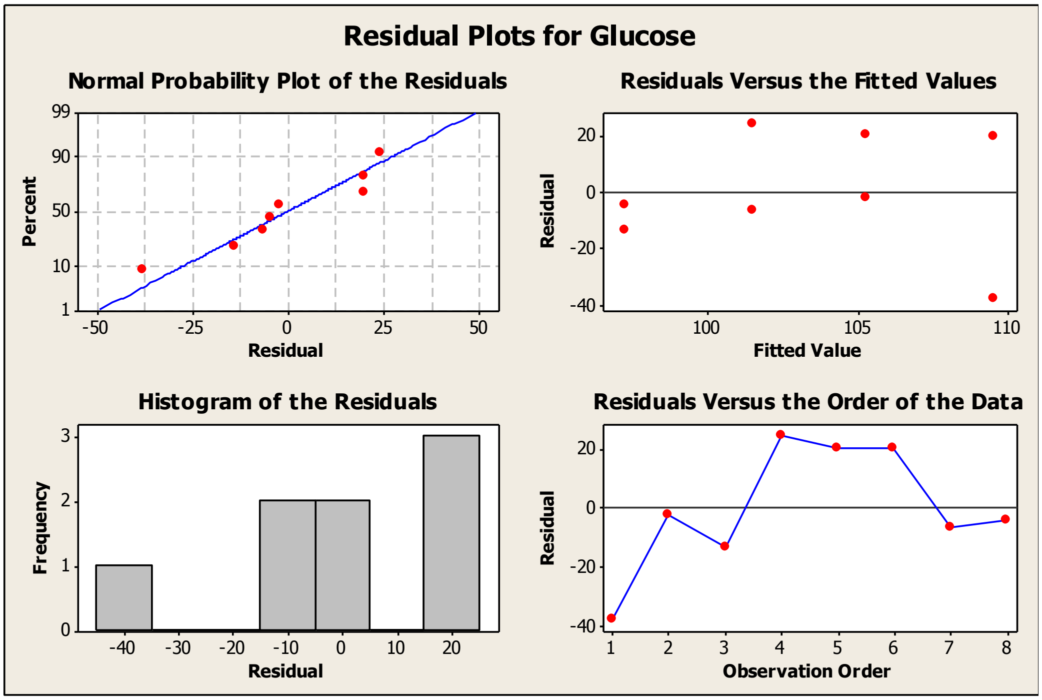
13-38 a) Analysis of variance for Glucose

```
Source  DF       SS       MS     F      P
Time     1    36.13   36.125  0.06  0.819
Min      1   128.00  128.000  0.21  0.669
Error    5  3108.75  621.750
Total    7  3272.88
```

No, there is no effect of exercise time on the average blood glucose.

b) *P*-value = 0.819

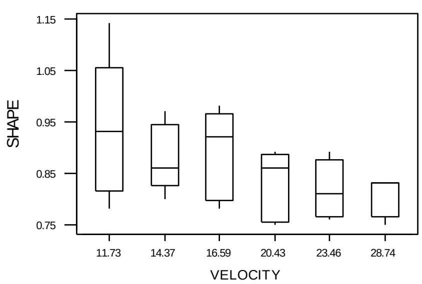
c) The normal probability plot and the residual plots show that the model assumptions are reasonable.

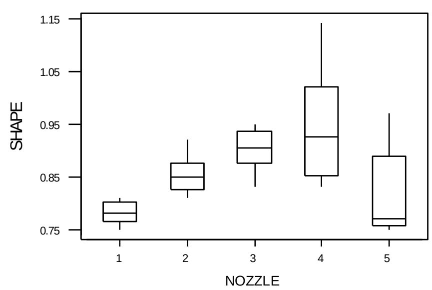


13-39 a) Analysis of Variance for SHAPE

Source	DF	SS	MS	F	P
NOZZLE	4	0.102180	0.025545	8.92	0.000
VELOCITY	5	0.062867	0.012573	4.39	0.007
Error	20	0.057300	0.002865		
Total	29	0.222347			

Reject H_0, nozzle type affects shape measurement.



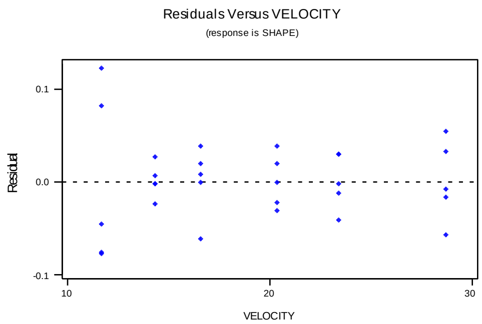


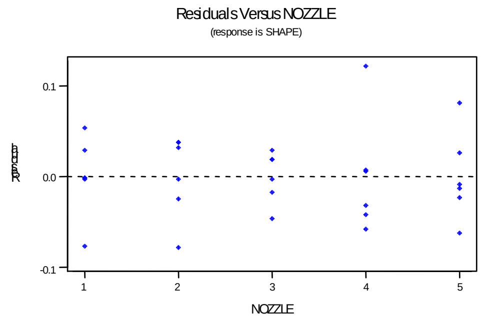
b) Fisher's pairwise comparisons
Family error rate = 0.268
Individual error rate = 0.0500
Critical value = 2.060
Intervals for (column level mean) - (row level mean)

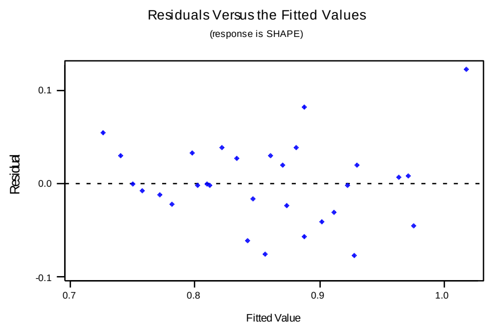
	1	2	3	4
2	-0.15412			
	0.01079			
3	-0.20246	-0.13079		
	-0.03754	0.03412		
4	-0.24412	-0.17246	-0.12412	
	-0.07921	-0.00754	0.04079	
5	-0.11412	-0.04246	0.00588	0.04754
	0.05079	0.12246	0.17079	0.21246

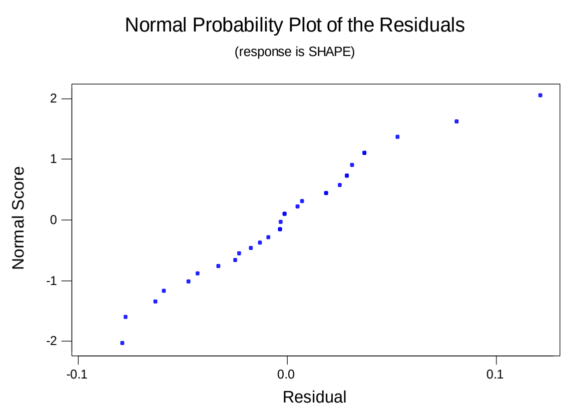
```
There are significant differences between levels 1 and 3, 4; 2 and 4; 3 and 5;
and 4 and 5.
```

c) The residual analysis shows that there is some inequality of variance. The normal probability plot is acceptable.









13-40 a) Analysis of Variance of HARDNESS

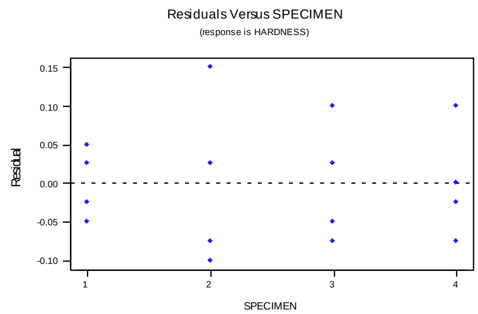
Source	DF	SS	MS	F	P
TIPTYPE	3	0.38500	0.12833	14.44	0.001
SPECIMEN	3	0.82500	0.27500	30.94	0.000
Error	9	0.08000	0.00889		
Total	15	1.29000			

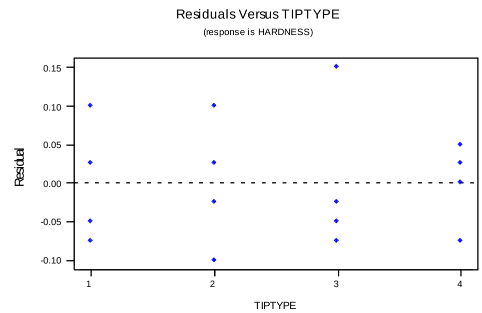
Reject H_0, and conclude that there are significant differences in hardness measurements between the tips.

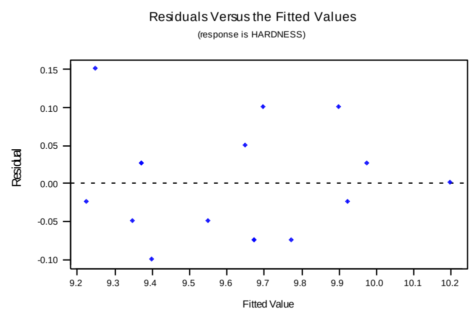
b)
```
Fisher's pairwise comparisons
Family error rate = 0.184
Individual error rate = 0.0500
Critical value = 2.179
```

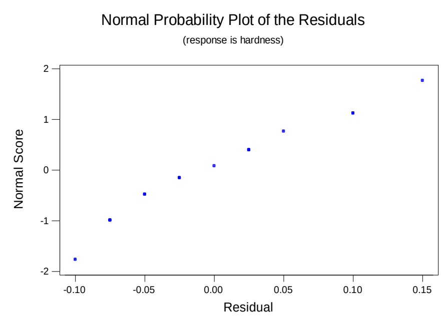
```
Intervals for (column level mean) - (row level mean)
            1           2           3
2     -0.4481
       0.3981
3     -0.2981     -0.2731
       0.5481      0.5731
4     -0.7231     -0.6981     -0.8481
       0.1231      0.1481     -0.0019

Significant difference between tip types 3 and 4
```

c) Residuals are acceptable.







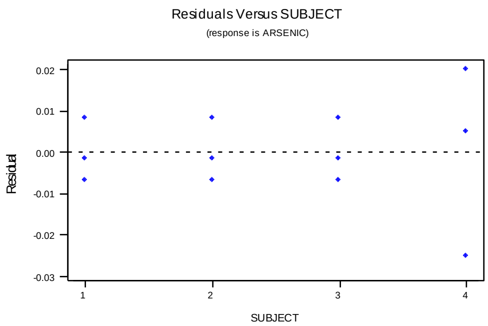


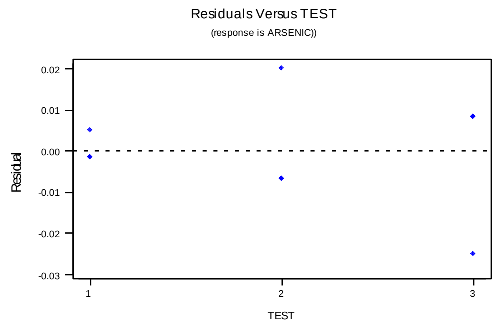
13-41 a) Analysis of Variance for ARSENIC

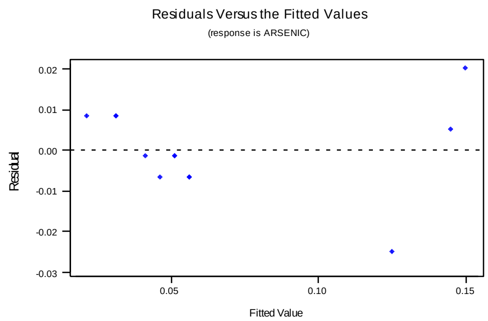
Source	DF	SS	MS	F	P
TEST	2	0.0014000	0.0007000	3.00	0.125
SUBJECT	3	0.0212250	0.0070750	30.32	0.001
Error	6	0.0014000	0.0002333		
Total	11	0.0240250			

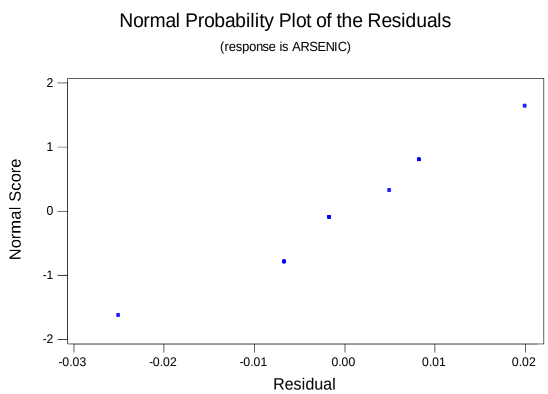
Fail to reject H_0, there is no evidence of differences between the tests.

b) Some indication of variability increasing with the magnitude of the response.









13-42 a) Analysis of Variance of PROPECTIN

Source	DF	SS	MS	F	P
STORAGE	3	1972652	657551	4.33	0.014
LOT	8	1980499	247562	1.63	0.169
Error	24	3647150	151965		
Total	35	7600300			

Reject H_0, and conclude that the storage times affect the mean level of propectin.

b) P-value = 0.014

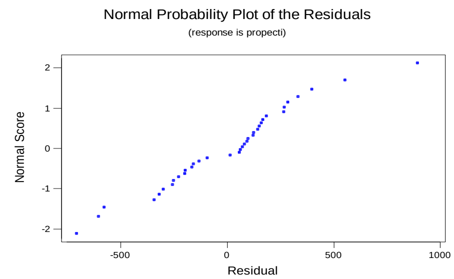
c)

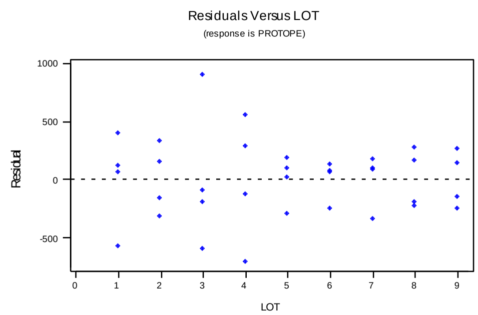
```
Fisher's pairwise comparisons
Family error rate = 0.196
Individual error rate = 0.0500
Critical value = 2.037
Intervals for (column level mean) - (row level mean)
                0            7          14
 7         -171
            634
14         -214         -445
            592          360
```

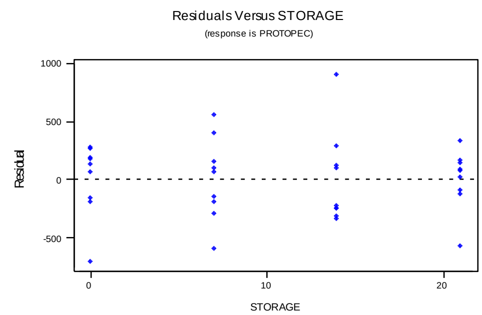
21	239	8	50
	1045	813	856

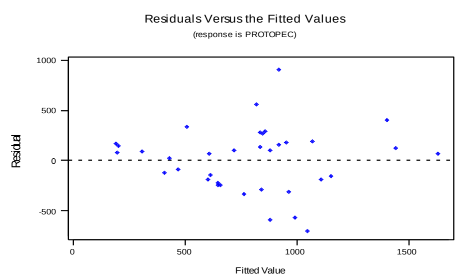
There are differences between 0 and 21 days; 7 and 21 days; and 14 and 21 days. The propectin levels are significantly different at 21 days from the other storage times so there is evidence that the mean level of propectin decreases with storage time. However, differences such as between 0 and 7 days and 7 and 14 days were not significant so that the level is not simply a linear function of storage days.

d) Observations from lot 3 at 14 days appear unusual. Otherwise, the residuals are acceptable.









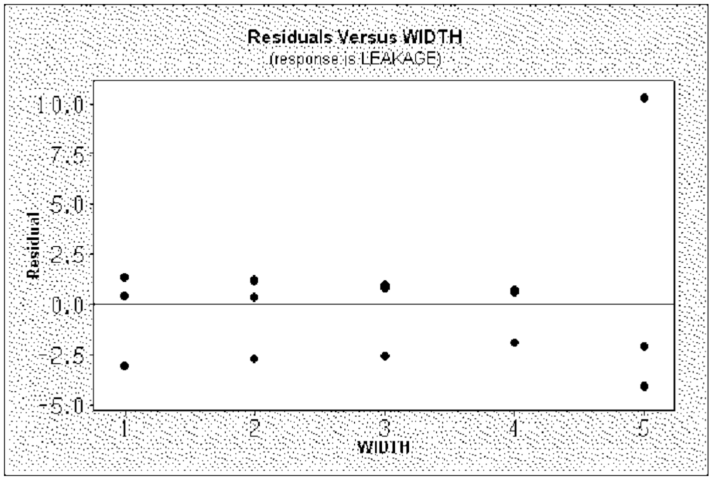
13-43 A version of the electronic data file has the reading for length 4 and width 5 as 2. It should be 20.

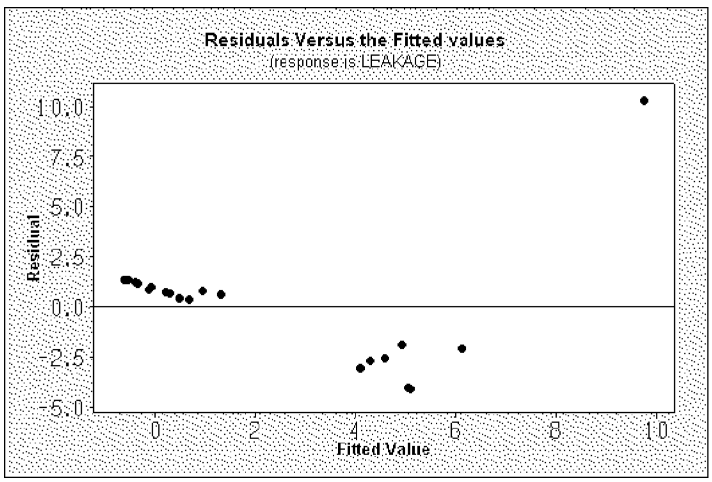
a) Analysis of Variance for LEAKAGE

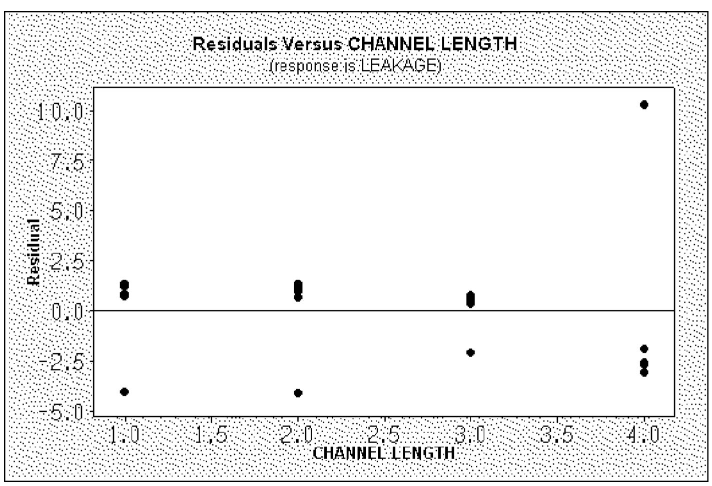
Source	DF	SS	MS	F	P
WIDTH	4	90.577	22.6443	1.51	0.261
CHANNEL LENGTH	3	73.668	24.5560	1.64	0.233
Error	12	179.987	14.9989		
Total	19	344.232			

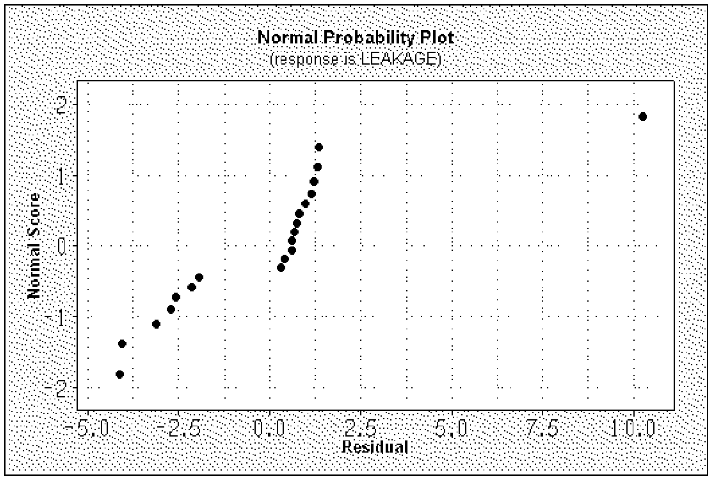
Fail to reject H_0, mean leakage voltage does not depend on the channel length.

b) One unusual observation in width 5, length 4. There are some problems with the normal probability plot, including the unusual observation.


Residuals Versus WIDTH
(response is LEAKAGE)
Residual
10.0
7.5
5.0
2.5
0.0
-2.5
-5.0
1
2
3
4
5
WIDTH


Residuals Versus the Fitted values
(response is LEAKAGE)
Residual
10.0
7.5
5.0
2.5
0.0
-2.5
-5.0
0
2
4
6
8
10
Fitted Value


Residuals Versus CHANNEL LENGTH
(response is LEAKAGE)
Residual
10.0
7.5
5.0
2.5
0.0
-2.5
-5.0
1.0
1.5
2.0
2.5
3.0
3.5
4.0
CHANNEL LENGTH



c) Analysis of Variance for LEAKAGE VOLTAGE

Source	DF	SS	MS	F	P
WIDTH	4	6.737	1.68425	3.84	0.031
CHANNEL LENGTH	3	8.388	2.79600	6.37	0.008
Error	12	5.267	0.43892		
Total	19	20.392			

Reject H_0. And conclude that the mean leakage voltage does depend on channel length. By removing the data point that was erroneous, the analysis results in a conclusion. The erroneous data point that was an obvious outlier had a strong effect the results of the experiment.

Supplemental Exercises

13-44 a) Note that df(Factor) = df(Total) – df(Error) = 19 – 15 = 4. Because the number of levels for a factor = df(Factor) + 1, 5 levels were used in the experiment.

b) Total number of observations = df(Total) + 1 = 19 + 1 = 20.
Because there are 5 levels used in this experiment, the number of replicates = 20/5 = 4.

c) From part (a), the df(Factor) = 4.
SS(Factor) = SS(Total) – SS(Error) = 326.2 – 167.5 = 158.7.
MS(Factor) = SS(Factor)/DF(Factor) = 158.7/4 = 39.675.
MS(Error) = SS(Error)/DF(Error) = 167.5/15 = 11.167.
F = MS(Factor)/MS(Error) = 39.675/11.167 = 3.553
0.025 < P-value < 0.05.

d) Because the P-value < α = 0.05 we reject H_0 for α = 0.05. There are significance differences in the factor level means at α = 0.05. Because the P-value > α = 0.01 we fail to reject H_0 for α = 0.01. There are not significance differences in the factor level means at α = 0.01.

13-45 a) Because MS = SS/df(Factor), df(Factor) = SS/MS = 126.880/63.4401 = 2. The number of levels = df(Factor) + 1 = 2 + 1 = 3. Therefore, 3 levels of the factor were used.

b) Because df(Total) = df(Factor) + df(Block) + df(Error)
11 = 2 + df(Block) + 6. Therefore, df(Block) = 3. Therefore, 4 blocks were used in the experiment.

c) From parts (a) and (b), df(Factor) = 3 and df(Block) = 2
SS(Error) = df(Error)MS(Error) = (6)3.1567 = 18.9402
F = MS(Factor)/MS(Error) = 63.4401/3.1567 = 20.097
P-value = 0.002

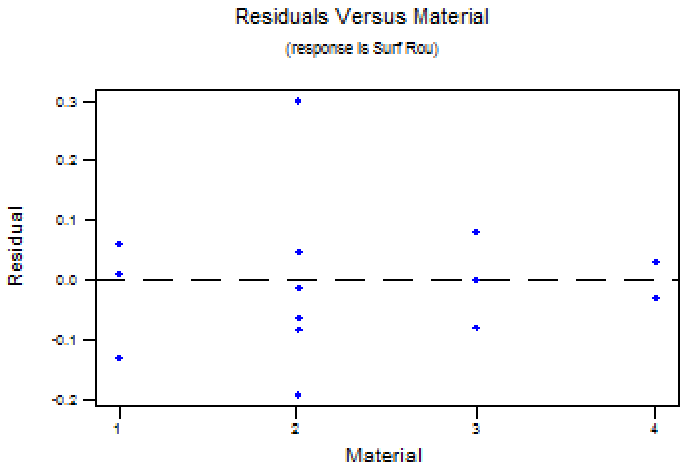
d) Because the P-value < 0.01 we reject H_0. There are significant differences in the factor level means at α = 0.05 or α = 0.01.

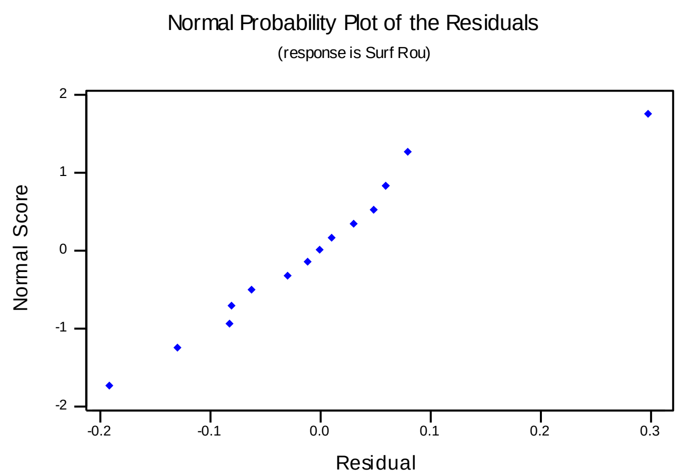
13-46 a)

```
Analysis of Variance for SURFACE ROUGNESS
Analysis of Variance for y
Source      DF        SS        MS         F        P
Material     3    0.2402    0.0801      4.96    0.020
Error       11    0.1775    0.0161
Total       14    0.4177
```

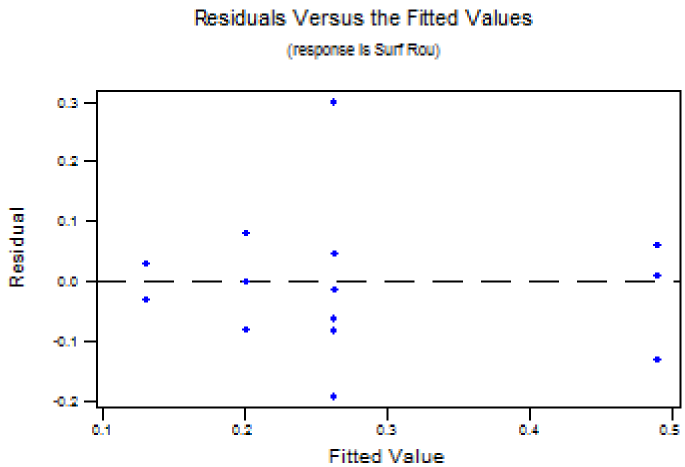
Fail to reject H_0

b) One observation is an outlier.





c) There appears to be a problem with constant variance. This may be due to the outlier in the data.



d) 99% confidence interval on the difference in the means of EC10 and EC1

$$\bar{y}_1 - \bar{y}_4 - t_{0.005,11}\sqrt{\frac{MS_E}{n_1} + \frac{MS_E}{n_2}} \le \mu_1 - \mu_3 \le \bar{y}_1 - \bar{y}_4 + t_{0.005,11}\sqrt{\frac{MS_E}{n_1} + \frac{MS_E}{n_2}}$$

$$(0.490 - 0.130) - 3.106\sqrt{\frac{(0.0161)}{4} + \frac{(0.0161)}{2}} \le \mu_1 - \mu_4 \le (0.490 - 0.130) + 3.106\sqrt{\frac{(0.0161)}{4} + \frac{(0.0161)}{2}}$$

$$0.0187 \le \mu_1 - \mu_4 \le 0.7013$$

13-47 a) Analysis of Variance for RESISTANCE

Source	DF	SS	MS	F	P
ALLOY	2	10941.8	5470.9	76.09	0.000
Error	27	1941.4	71.9		
Total	29	12883.2			

Reject H_0, the type of alloy has a significant effect on mean contact resistance.

b)
```
Fisher's pairwise comparisons
 Family error rate = 0.119
 Individual error rate = 0.0500
 Critical value = 2.052
 Intervals for (column level mean) - (row level mean)
                  1            2
      2      -13.58
               1.98
      3      -50.88       -45.08
              -35.32      -29.52
```

There are differences in the mean resistance for alloy types 1 and 3; and types 2 and 3.

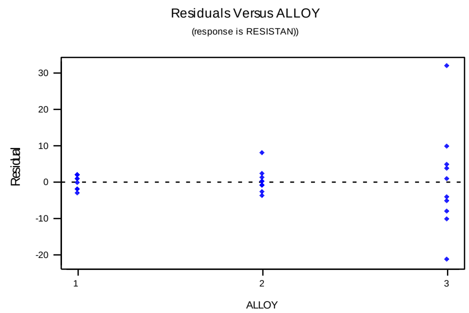
c) 95% confidence interval on the mean contact resistance for alloy 3

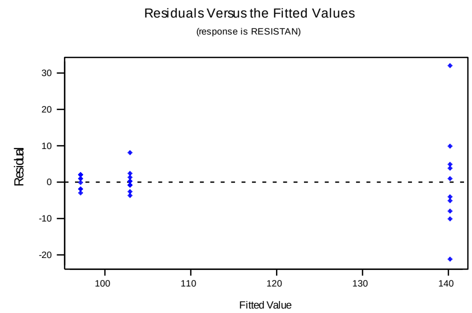
$$\bar{y}_3 - t_{0.025,27}\sqrt{\frac{MS_E}{n}} \le \mu_i \le \bar{y}_3 + t_{0.025,27}\sqrt{\frac{MS_E}{n}}$$

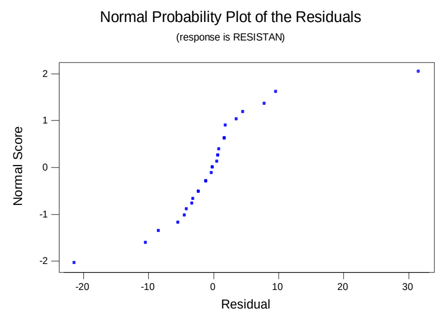
$$140.4 - 2.052\sqrt{\frac{71.9}{10}} \le \mu_3 \le 140.4 + 2.052\sqrt{\frac{71.9}{10}}$$

$$134.898 \le \mu_1 \le 145.902$$

d) Variability of the residuals increases with the response. The normal probability plot has some curvature in the tails, indicating a problem with the normality assumption. A transformation of the response should be conducted.







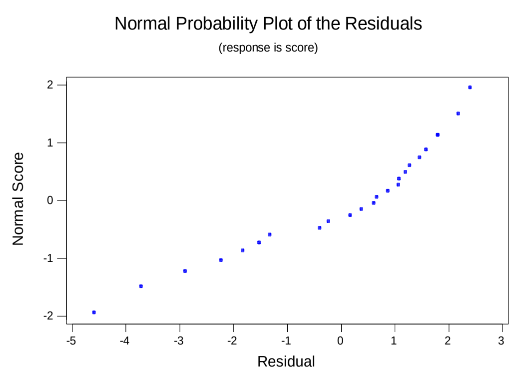
13-48 a) Analysis of Variance for SCORE

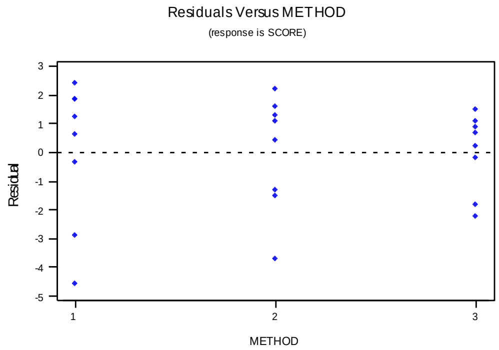
Source	DF	SS	MS	F	P
METHOD	2	13.55	6.78	1.68	0.211
Error	21	84.77	4.04		
Total	23	98.32			

Fail to reject H_0

b) *P*-value = 0.211

c) There is some curvature in the normal probability plot. There appears to be some differences in the variability for the different methods. The variability for method one is larger than the variability for method 3.





d.) $$\hat{\sigma}_{\tau}^{2} = \frac{MS_{Treatments} - MS_E}{n} = \frac{6.78 - 4.04}{8} = 0.342$$

$$\hat{\sigma}^2 = MS_E = 4.04$$

13-49 a)Analysis of Variance for VOLUME

Source	DF	SS	MS	F	P
TEMPERATURE	2	16480	8240	7.84	0.007
Error	12	12610	1051		
Total	14	29090			

Reject H_0.

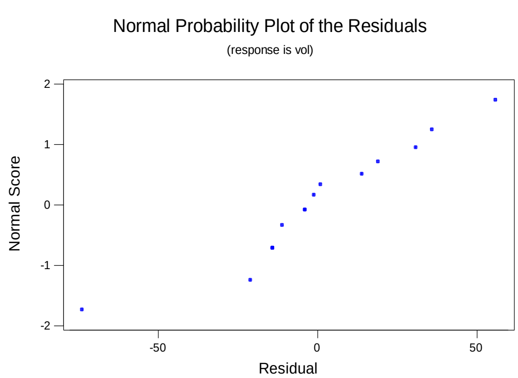
b) *P-value* = 0.007

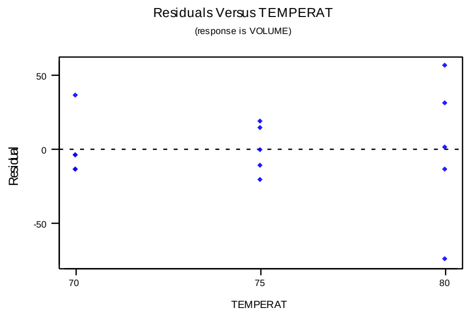
c) Fisher's pairwise comparisons

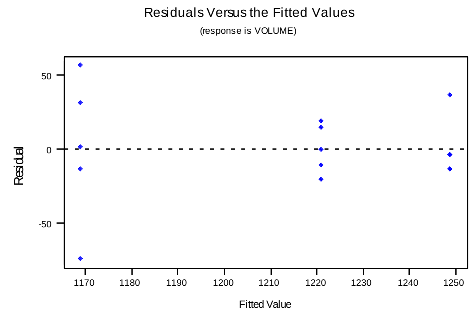
```
Family error rate = 0.116
Individual error rate = 0.0500
Critical value = 2.179
Intervals for (column level mean) - (row level mean)
                70             75
    75        -16.7
               72.7
    80         35.3           7.3
              124.7          96.7
```

There are significant differences in the mean volume for temperature levels 70 and 80; and 75 and 80. The highest temperature results in the smallest mean volume.

d) There are some relatively small differences in the variability at the different levels of temperature. The variability decreases with the fitted values. There is an unusual observation on the normal probability plot.







13-50 a) Analysis of Variance of Weight Gain

Source	DF	SS	MS	F	P
MEANWEIG	2	0.2227	0.1113	1.48	0.273
AIRTEMP	5	10.1852	2.0370	27.13	0.000
Error	10	0.7509	0.0751		
Total	17	11.1588			

Reject H_0 and conclude that the air temperature has an effect on the mean weight gain.

b) Fisher's pairwise comparisons

Family error rate = 0.314
Individual error rate = 0.0500

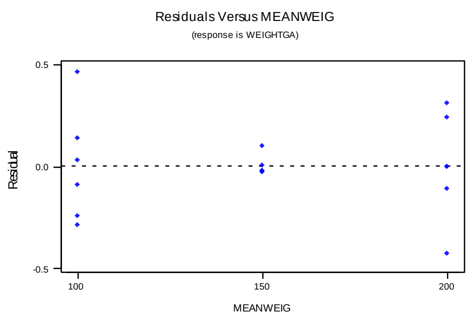
Critical value = 2.179

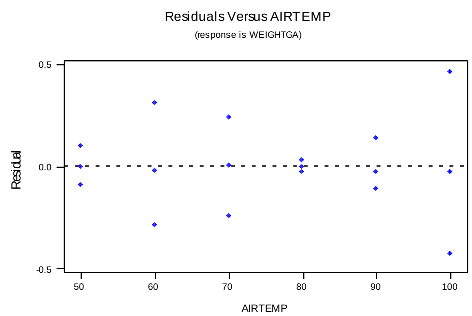
Intervals for (column level mean) - (row level mean)

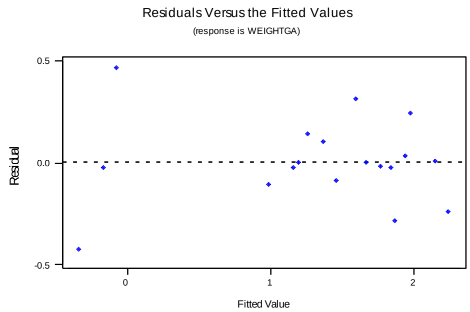
	50	60	70	80	90
60	-0.9101 0.1034				
70	-1.2901 -0.2766	-0.8868 0.1268			
80	-0.9834 0.0301	-0.5801 0.4334	-0.2001 0.8134		
90	-0.3034 0.7101	0.0999 1.1134	0.4799 1.4934	0.1732 1.1868	
100	1.0266 2.0401	1.4299 2.4434	1.8099 2.8234	1.5032 2.5168	0.8232 1.8368

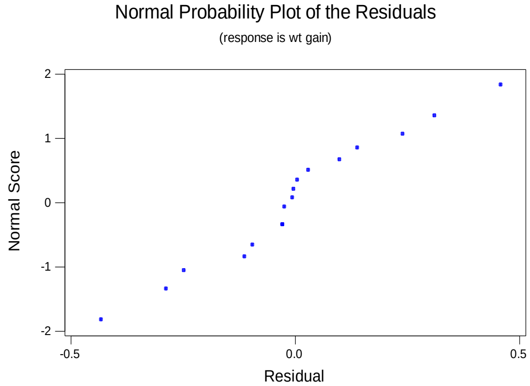
There are significant differences in the mean air temperature levels 50 and 70, 100; 60 and 90, 100; 70 and 90, 100; 80 and 90, 100; and 90 and 100. The mean of temperature level 100 is different from all the other temperatures.

c) There appears to be some problems with the assumption of constant variance.







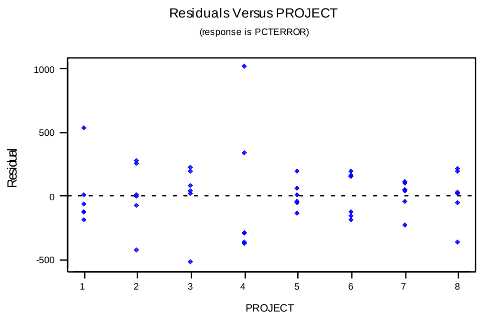


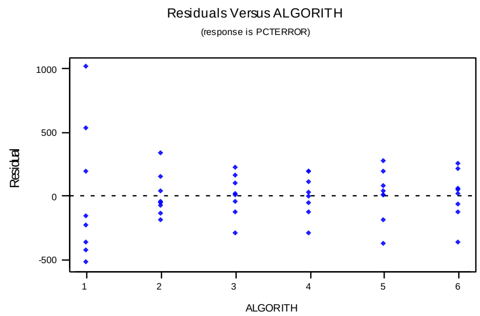
13-51 a) Analysis of Variance for PCTERROR

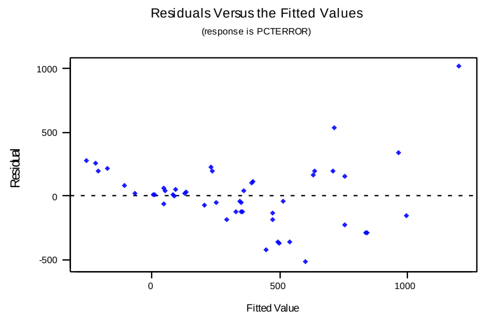
Source	DF	SS	MS	F	P
ALGORITH	5	2825746	565149	6.23	0.000
PROJECT	7	2710323	387189	4.27	0.002
Error	35	3175290	90723		
Total	47	8711358			

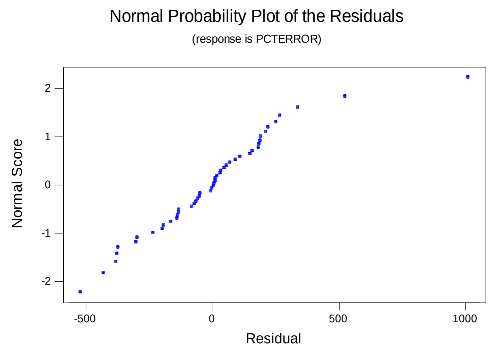
Reject H_0, the algorithms are significantly different.

b) The residuals look acceptable, except there is one unusual point.



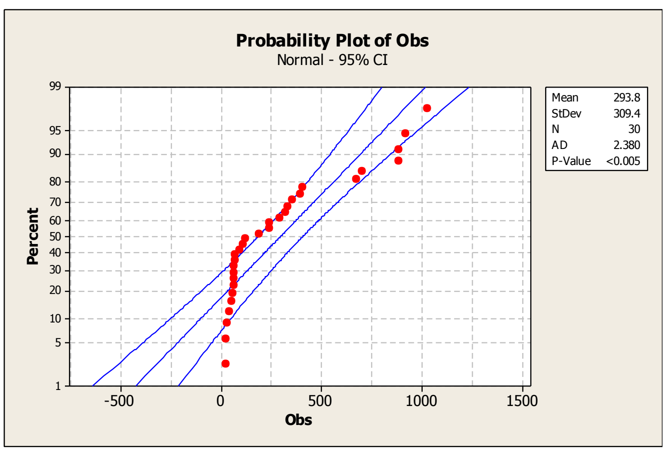




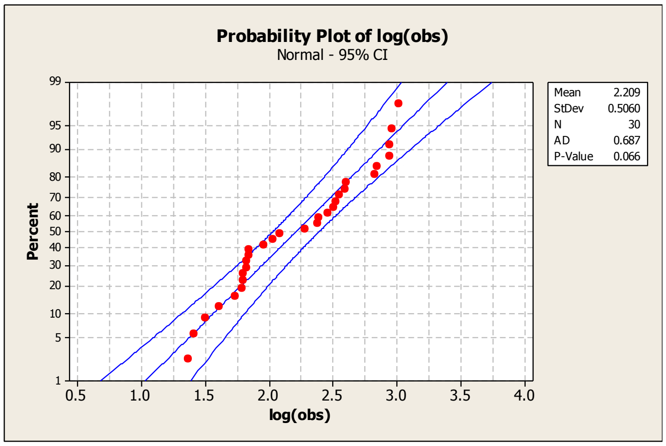


c) The best choice is algorithm 5 because it has the smallest mean and a low variability.

13-52 a) The normal probability plot shows that the normality assumption is not reasonable.



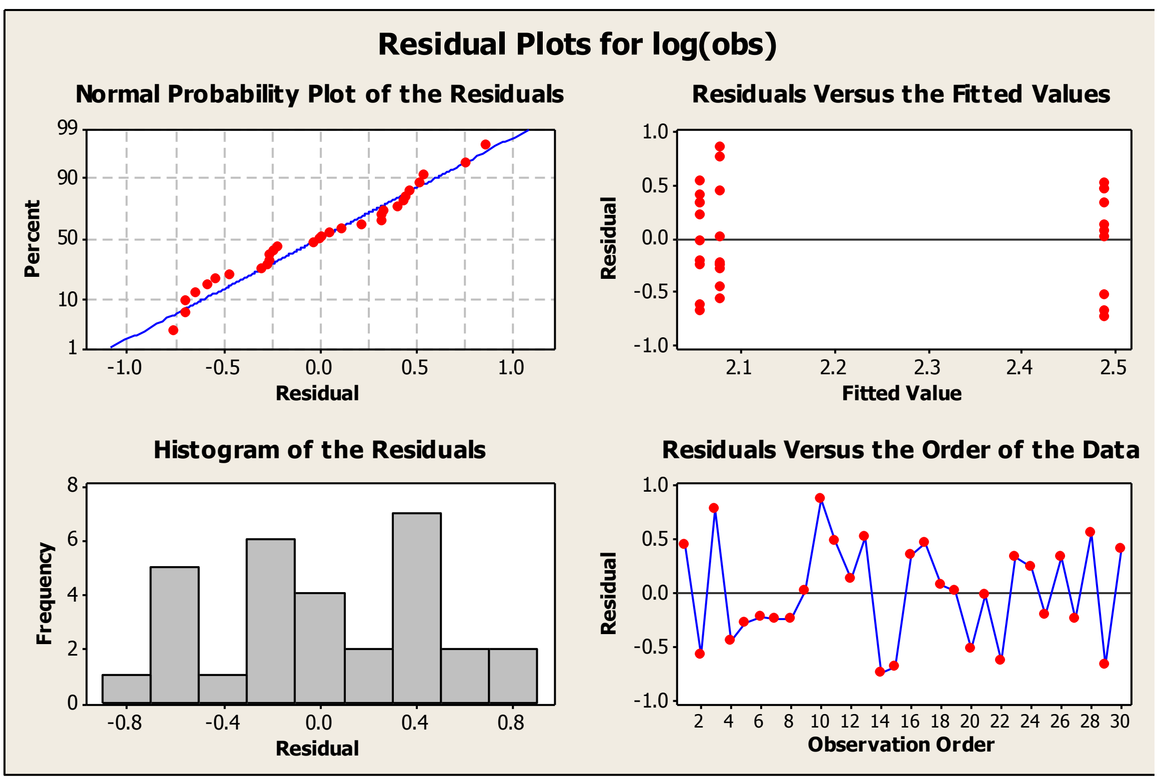
b) The normal probability plot shows that the normality assumption is reasonable.



Source	DF	SS	MS	F	P
Treatments	2	1.188	0.594	2.57	0.095
Error	27	6.237	0.231		
Total	29	7.425			

There is evidence to support the claim that the treatment means differ at α = 0.1 for the transformed data since the P-value = 0.095.

c) The normal probability plot and the residual plots show that the model assumptions are reasonable.



13-53 a) μ = 2.4, Φ^2 = 0.391, Φ = 0.625

Numerator degrees of freedom = $a-1=4=v_1$

Denominator degrees of freedom = $a(n-1)=15=v_2$

From Chart Figure 13-6, β ≈ 0.8 and the power = 1 – β = 0.2

b)

n	Φ^2	Φ	a(n – 1)	β	Power = 1 – β
50	4.89	2.211	245	0.03	0.90

The sample size should be approximately $n = 50$.

13-54 a) $\mu = (1+5+8+4)/4 = 4.5$ and

$$\Phi^2 = \frac{5[(1-4.5)^2 + (5-4.5)^2 + (8-4.5)^2 + (4-4.5)^2]}{4(4)} = 7.81$$

$$\Phi = 2.8$$

Numerator degrees of freedom = $a - 1 = 3 = \nu_1$

Denominator degrees of freedom = $a(n-1) = 16 = \nu_2$

From Figure 13-6, $\beta = 0.06$ and the power = $1 - \beta = 0.94$

b)

n	Φ^2	Φ	a(n – 1)	β	Power = 1 – β
5	7.81	2.8	16	0.06	0.94
4	6.25	2.5	12	0.18	0.82

The sample size should be approximately $n = 5$.

Mind Expanding Exercises

13-55 $MS_E = \dfrac{\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \bar{y}_i)^2}{a(n-1)}$ and $y_{ij} = \mu + a_i + \varepsilon_{ij}$. Then $y_{ij} - \bar{y}_i = \varepsilon_{ij} - \bar{\varepsilon}_{i.}$ and

$\dfrac{\sum_{j=1}^{n}(\varepsilon_{ij} - \bar{\varepsilon}_{i.})}{n-1}$ is recognized to be the sample variance of the independent random variables $\varepsilon_{i1}, \varepsilon_{i2}, \ldots, \varepsilon_{in}$.

Therefore, $E\left[\dfrac{\sum_{j=1}^{n}(\varepsilon_{ij} - \bar{\varepsilon}_{i.})^2}{n-1}\right] = \sigma^2$ and $E(MS_E) = \sum_{i=1}^{a}\dfrac{\sigma^2}{a} = \sigma^2$.

The development would not change if the random effects model had been specified because $y_{ij} - \bar{y}_i = \varepsilon_{ij} - \bar{\varepsilon}_{i.}$ for this model also.

13-56 The two sample t-test rejects equality of means if the statistic
$t = \dfrac{|\bar{y}_1 - \bar{y}_2|}{s_p\sqrt{\frac{1}{n}+\frac{1}{n}}}$ is too large.

The ANOVA F-test rejects equality of means if $F = \dfrac{n\sum_{i=1}^{2}(\bar{y}_{i.} - \bar{y}_{..})^2}{MS_E}$ is too large.

Now, $F = \dfrac{\frac{n}{2}(\bar{y}_{1.} - \bar{y}_{2.})^2}{MS_E} = \dfrac{(\bar{y}_{1.} - \bar{y}_{2.})^2}{MS_E\frac{2}{n}}$ and $MS_E = s_p^2$.

Consequently, $F = t^2$. Also, the distribution of the square of a t random variable with a(n – 1) degrees of freedom is an F distribution with 1 and $a(n-1)$ degrees of freedom. Therefore, if the critical value for a two-sided t-test of size α is t_0, then the tabulated F value for the F test above is t_0^2. Therefore, $t > t_0$ whenever $F = t^2 > t_0$ and the two tests are identical.

13-57 $MS_E = \dfrac{\sum_{i=1}^{2} \sum_{j=1}^{n} (y_{ij} - \bar{y}_{i.})^2}{2(n-1)}$ and $\dfrac{\sum_{j=1}^{n} (y_{ij} - \bar{y}_{i.})^2}{n-1}$ is recognized as the sample standard deviation calculated from the data from population *i*. Then, $MS_E = \dfrac{s_1^2 + s_2^2}{2}$ which is the pooled variance estimate used in the t-test.

13-58 $V(\sum_{i=1}^{a} c_i Y_{i.}) = \sum_{i=1}^{a} c_i^2 V(Y_{i.})$ from the independence of $Y_{1.}, Y_{2.}, \ldots, Y_{a.}$.

Also, $V(Y_{i.}) = n_i \sigma_i^2$. Then, $V(\sum_{i=1}^{a} c_i Y_{i.}) = \sigma^2 \sum_{i=1}^{a} c_i^2 n_i$

13-59 If b, c, and d are the coefficients of three orthogonal contrasts, it can be shown that

$$\frac{(\sum_{i=1}^{4} b_i y_{i.})^2}{\sum_{i=1}^{a} b_i^2} + \frac{(\sum_{i=1}^{a} c_i y_{i.})^2}{\sum_{i=1}^{a} c_i^2} + \frac{(\sum_{i=1}^{a} d_i y_{i.})^2}{\sum_{i=1}^{a} d_i^2} = \sum_{i=1}^{a} y_{i.}^2 - \frac{(\sum_{i=1}^{a} y_{i.})^2}{a}$$

always holds. Upon dividing both sides by n, we have

$Q_1^2 + Q_2^2 + Q_3^2 = \sum_{i=1}^{a} \dfrac{y_i^2}{n} - \dfrac{y_{..}^2}{N}$ which equals $SS_{\text{Treatments}}$.

The equation above can be obtained from a geometrical argument. The square of the distance of any point in four-dimensional space from the zero point can be expressed as the sum of the squared distance along four orthogonal axes. Let one of the axes be the 45 degree line and let the point be ($y_{1.}, y_{2.}, y_{3.}, y_{4.}$). The three orthogonal contrasts are the other three axes. The square of the distance of the point from the origin is $\sum_{i=1}^{a} y_{i.}^2$ and this equals the sum of the squared distances along each of the four axes.

13-60 Because $\Phi^2 = \dfrac{n\sum_{i=1}^{a} (\mu_i - \bar{\mu})^2}{a\sigma^2}$, we only need to shows that $\dfrac{D^2}{2} \le \sum_{i=1}^{a} (\mu_i - \bar{\mu})^2$.

Let μ_1 and μ_2 denote the means that differ by D. Now, $(\mu_1 - x)^2 + (\mu_2 - x)^2$ is minimized for x equal to the mean of μ_1 and μ_2. Therefore, $(\mu_1 - \dfrac{\mu_1 + \mu_2}{2})^2 + (\mu_2 - \dfrac{\mu_1 + \mu_2}{2})^2 \le (\mu_1 - \bar{\mu})^2 + (\mu_2 - \bar{\mu})^2 \le \sum_{i=1}^{a} (\mu_i - \bar{\mu})^2$

Then, $\left(\dfrac{\mu_1 - \mu_2}{2}\right)^2 + \left(\dfrac{\mu_2 - \mu_1}{2}\right)^2 = \dfrac{D^2}{4} + \dfrac{D^2}{4} = \dfrac{D^2}{2} \le \sum_{i=1}^{a} (\mu_i - \bar{\mu})^2$.

13-61 $MS_E = \dfrac{\sum_{i=1}^{a} \sum_{j=1}^{n} (y_{ij} - \bar{y}_{i.})^2}{a(n-1)} = \dfrac{\sum_{i=1}^{a} s_i^2}{a}$ where $s_i^2 = \dfrac{\sum_{j=1}^{n} (y_{ij} - \bar{y}_{i.})^2}{n-1}$.

Because s_i^2 is the sample variance of $y_{i1}, y_{i2}, \ldots, y_{in}$, $\frac{(n-1)S_i^2}{\sigma^2}$ has a chi-square distribution with n-1 degrees of freedom. Then, $\frac{a(n-1)MS_E}{\sigma^2}$ is a sum of independent chi-square random variables. Consequently, $\frac{a(n-1)MS_E}{\sigma^2}$ has a chi-square distribution with a(n – 1) degrees of freedom. Consequently,

$$P(\chi^2_{1-\frac{\alpha}{2},a(n-1)} \le \frac{a(n-1)MS_E}{\sigma^2} \le \chi^2_{\frac{\alpha}{2},a(n-1)}) = 1-\alpha$$

$$= P\left(\frac{a(n-1)MS_E}{\chi^2_{\frac{\alpha}{2},a(n-1)}} \le \sigma^2 \le \frac{a(n-1)MS_E}{\chi^2_{1-\frac{\alpha}{2},a(n-1)}}\right)$$

Using the fact that a(n – 1) = N – a completes the derivation.

13-62 From the previous exercise, $\frac{(N-a)MS_E}{\sigma^2}$ has a chi-square distribution with $N - a$ degrees of freedom. Now,

$V(\overline{Y}_{i.}) = \sigma_\tau^2 + \frac{\sigma^2}{n}$ and mean square treatment = MS_T is n times the sample variance of $\overline{y}_{1.}, \overline{y}_{2.}, \ldots, \overline{y}_{a.}$.

Therefore, $\frac{(a-1)MS_T}{n(\sigma_\tau^2 + \frac{\sigma^2}{n})} = \frac{(a-1)MS_T}{n\sigma_\tau^2 + \sigma^2}$ has a chi-squared distribution with $a - 1$ degrees of freedom. Using the independence of MS_T and MS_E, we conclude that $\left(\frac{MS_T}{n\sigma_\tau^2 + \sigma^2}\right) / \left(\frac{MS_E}{\sigma^2}\right)$ has an $F_{(a-1),(N-a)}$ distribution.

Therefore,

$$P(f_{1-\frac{\alpha}{2},a-1,N-a} \le \frac{MS_T}{MS_E}\frac{\sigma^2}{n\sigma_\tau^2 + \sigma^2} \le f_{\frac{\alpha}{2},a-1,N-a}) = 1-\alpha$$

$$= P\left(\frac{1}{n}\left[\frac{1}{f_{\frac{\alpha}{2},a-1,N-a}}\frac{MS_T}{MS_E} - 1\right] \le \frac{\sigma_\tau^2}{\sigma^2} \le \frac{1}{n}\left[\frac{1}{f_{1-\frac{\alpha}{2},a-1,N-a}}\frac{MS_T}{MS_E} - 1\right]\right)$$

by an algebraic solution for $\frac{\sigma_\tau^2}{\sigma^2}$ and $P(L \le \frac{\sigma_\tau^2}{\sigma^2} \le U)$.

13-63 a) As in the previous exercise, $\frac{MS_T}{MS_E}\frac{\sigma^2}{n\sigma_\tau^2 + \sigma^2}$ has an $F_{(a-1),(N-a)}$ distribution.

and

$$1-\alpha = P(L \le \frac{\sigma_\tau^2}{\sigma^2} \le U)$$

$$= P(\frac{1}{U} \le \frac{\sigma^2}{\sigma_\tau^2} \le \frac{1}{L})$$

$$= P(\frac{1}{U} + 1 \le \frac{\sigma^2}{\sigma_\tau^2} + 1 \le \frac{1}{L} + 1)$$

$$= P(\frac{L}{L+1} \le \frac{\sigma_\tau^2}{\sigma^2 + \sigma_\tau^2} \le \frac{U}{U+1})$$

b) $1-\alpha = P(L \le \frac{\sigma_\tau^2}{\sigma^2} \le U)$

$$= P(L+1 \le \frac{\sigma_\tau^2 + 1}{\sigma^2} \le U+1)$$

$$= P(L+1 \le \frac{\sigma_\tau^2 + \sigma^2}{\sigma^2} \le U+1)$$

$$= P(\frac{1}{U+1} \le \frac{\sigma^2}{\sigma^2+\sigma_\tau^2} \le \frac{1}{L+1})$$

Therefore, $\left(\frac{1}{U+1}, \frac{1}{L+1}\right)$ is a confidence interval for $\frac{\sigma^2}{\sigma_\tau^2+\sigma^2}$

13-64 $MS_T = \frac{\sum_{i=1}^{a} n_i(\bar{y}_{i.} - \bar{y}_{..})^2}{a-1}$ and for any random variable *X*, $\mathrm{E}(\mathrm{X}^2) = \mathrm{V}(\mathrm{X}) + [\mathrm{E}(\mathrm{X})]^2$.

Then,

$$E(MS_T) = \frac{\sum_{i=1}^{a} n_i \{V(\bar{Y}_{i.} - \bar{Y}_{..}) + [E(\bar{Y}_{i.} - \bar{Y}_{..})]^2\}}{a-1}$$

Now, $\bar{Y}_{1.} - \bar{Y}_{..} = (\frac{1}{n_1} - \frac{1}{N})Y_{11} + \ldots + (\frac{1}{n_1} - \frac{1}{N})Y_{1n_1} - \frac{1}{N}Y_{21} - \ldots - \frac{1}{N}Y_{2n_2} - \ldots - \frac{1}{N}Y_{a1} - \ldots - \frac{1}{N}Y_{an_a}$

and

$$V(\bar{Y}_{1.} - \bar{Y}_{..}) = \left((\tfrac{1}{n_1} - \tfrac{1}{N})^2 n_1 + \frac{N-n_1}{N^2}\right)\sigma^2 = (\tfrac{1}{n_1} - \tfrac{1}{N})\sigma^2$$

$$E(\bar{Y}_{1.} - \bar{Y}_{..}) = (\tfrac{1}{n_1} - \tfrac{1}{N})n_1\tau_1 - \frac{n_2}{N}\tau_2 - \ldots - \frac{n_a}{N}\tau_a = \tau_1 \text{from the constraint}$$

Then,

$$E(MS_T) = \frac{\sum_{i=1}^{a} n_i\{(\frac{1}{n_i} - \frac{1}{N})\sigma^2 + \tau_i^2\}}{a-1} = \frac{\sum_{i=1}^{a} [(1-\frac{n_i}{N})\sigma^2 + n_i\tau_i^2]}{a-1}$$

$$= \sigma^2 + \frac{\sum_{i=1}^{a} n_i\tau_i^2}{a-1}$$

Because $E(MS_E) = \sigma^2$, this does suggest that the null hypothesis is as given in the exercise.

13-65 a) If *A* is the accuracy of the interval, then $t_{\frac{\alpha}{2},a(n-1)}\sqrt{\frac{2MS_E}{n}} = A$

Squaring both sides yields $t^2_{\frac{\alpha}{2},a(n-1)} \frac{2MS_E}{n} = A^2$

Also, $t^2_{\frac{\alpha}{2},a(n-1)} = F_{\alpha,1,a(n-1)}$. Then,

$$n = \frac{2MS_E F_{\alpha,1,a(n-1)}}{A^2}$$

b) Because *n* determines one of the degrees of freedom of the tabulated F value on the right-side of the equation in part (a), some approximation is needed. Because the value for a 95% confidence interval based on a normal distribution is 1.96, we approximate $t_{\frac{\alpha}{2},a(n-1)}$ by 2 and we approximate

$t^2_{\frac{\alpha}{2},a(n-1)} = F_{\alpha,1,a(n-1)}$ by 4.

Then, $n = \dfrac{2(4)(4)}{4} = 8$. With n = 8, $a(n$ – 1) = 35 and $F_{0.05,1,35} = 4.12$.

The value 4.12 can be used for F in the equation for n and a new value can be computed for n as

$$n = \frac{2(4)(4.12)}{4} = 8.24 \cong 8$$

Because the solution for n did not change, we can use n = 8. If needed, another iteration could be used to refine the value of n.