RESEARCH SCHOOL OF
FINANCE, ACTUARIAL STUDIES AND STATISTICS
College of Business & Economics, The Australian National University

GENERALISED LINEAR MODELLING
(STAT3015/STAT7030)

Model solutions to Assignment 1 for 2015

Question 1

(a) This is a randomised complete block experimental design with two design factors and a blocking factor. The design or fixed experimental factors are:

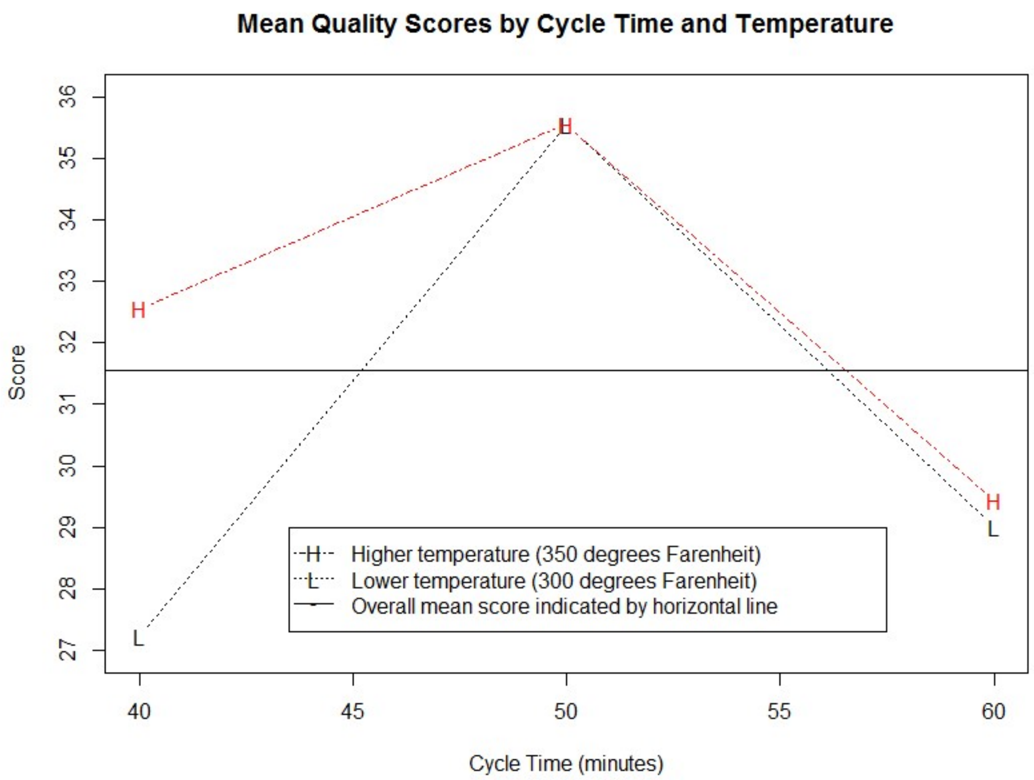
> **ctime**, the cycle times, with three levels – 40, 50 and 60 minutes; and
> **temp**, the temperatures, with two levels – 300 and 350 °F.

Together, the levels of two design factors form 3 × 2 = 6 experimental treatments, which are the combinations of interest.

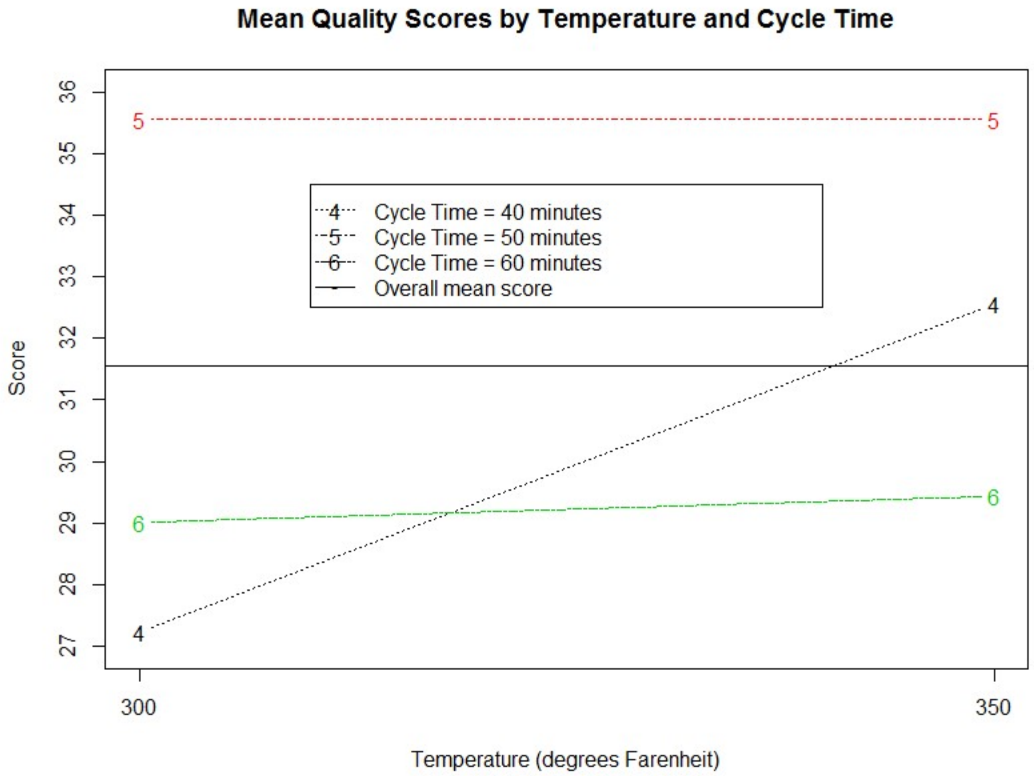
The blocking factor is **operator**, which also has three levels – operators 1, 2 and 3, who were presumably chosen at random from a larger population of possible operators. The different operators represent a known, but unwanted source of random variation in the experimental outcomes, the **score**s.

The experiment is designed to control for variation between operators by getting each operator to repeat the entire experiment for 3 complete replications each (i.e. repeat the complete experiment within each block). So a quick check of the data shows that there are 3 × 3 = 9 replications of each of the main treatments, which gives a total sample size of 6 × 9 = 54. **(3 marks)**

(b)



Question 1, part b continued



It appears that a cycle time of 50 minutes gives the best scores, regardless of whether the temperature is set at low or high. The results for the other two cycle times are more mixed and do appear to depend on the temperature setting – the apparent "crossing" on the graph for these cycle times may result in a significant interaction when we try to model these results.

What we don't know until we have fitted some model to the results is whether the 2 combination means for a cycle time of 50 minutes are significantly better than all of the other treatment combinations, including the mean for a cycle time of 40 minutes at a temperature of 350 °F, which is the only other combination that appears to be above the overall mean. **(3 marks)**

(c) Treating operator as a fixed effect, the underlying population model that best fits the description in the question is:

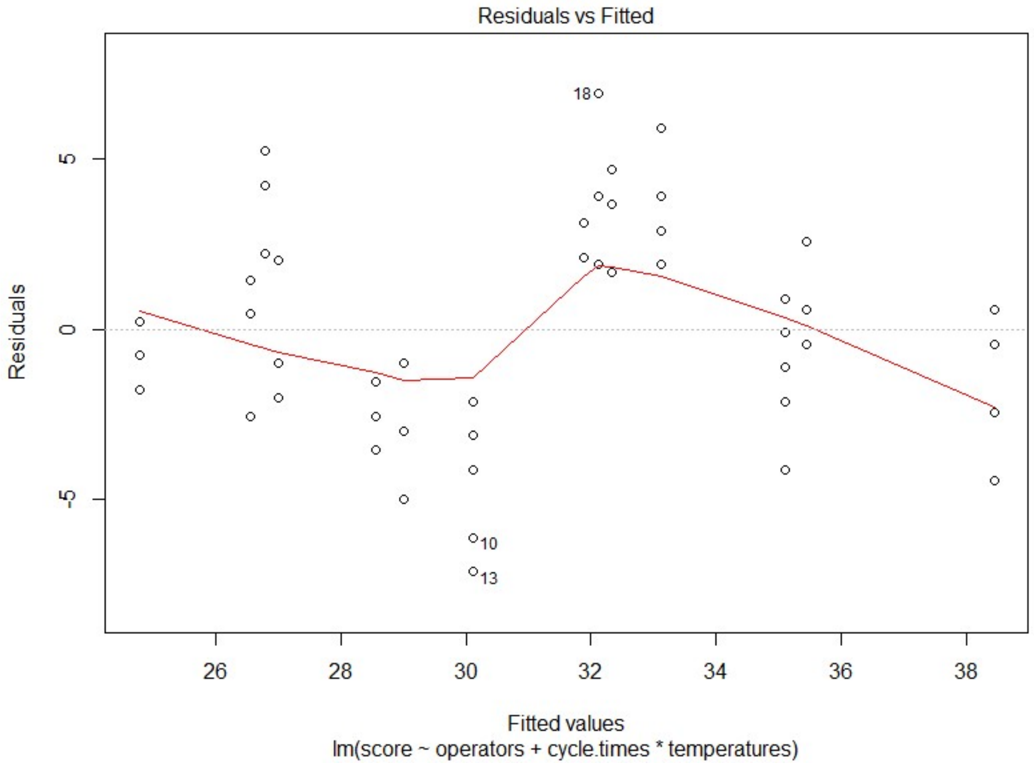
$$Y_{ijkl} = \mu + \delta_j + \eta_k + \gamma_l + \lambda_{kl} + \varepsilon_{ijkl}$$

with constraints: $\sum_j \delta_j = 0$, $\sum_k \eta_k = 0$, $\sum_l \gamma_l = 0$, $\sum_k \lambda_{kl} = 0$, $\sum_l \lambda_{kl} = 0$ and

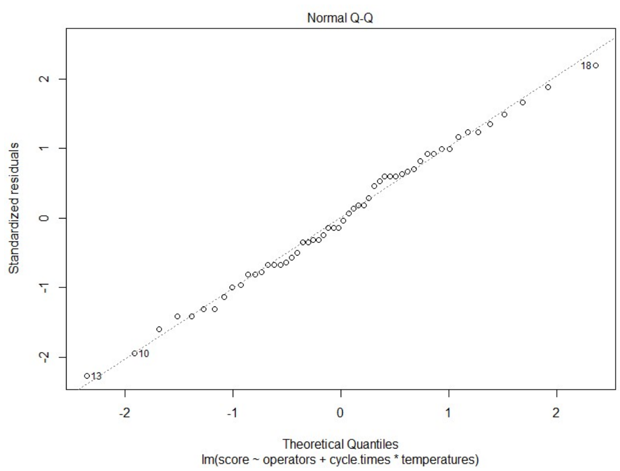
assuming the errors (ε_{ijkl}) are independently and identically distributed $N(0, \sigma^2)$,

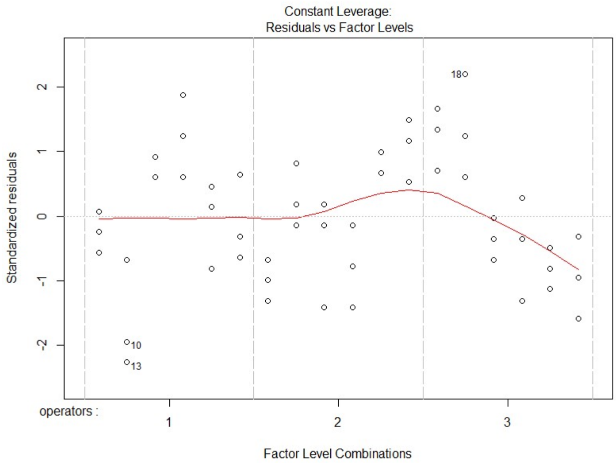
where Y is the quality **score**,
j indicates the levels of **operator** = {1, 2, 3},
k indicates the levels of **ctime** = {40 minutes, 50 minutes, 60 minutes},
l indicates the levels of **temp** = {300 °F, 350 °F} and
i = 1, 2, 3 replications for all 18 combinations of j, k & l.

Question 1, part c continued



The above residual plot has an obvious pattern which suggests the assumption of independence of the errors is not correct, which in turn suggests there is some source of variability in the data that is not accounted for by the model.





The normal quantile plot does not suggest any additional problems with the model. However, as a leverage plot, I have chosen the default plot (#5), rather than a bar plot of Cook's distances, as it suggests what is going on with the above main residual plot.

The model described in the question is a standard model for this experimental design, and assumes no interactions between the blocking variable and the main experimental factors. However, some of the operators (notably operator 3) do appear to be responding differently to some of the different treatment combinations, which brings this assumption into question. **(6 marks)**

Question 1 continued

(d) The Analysis of Variance table for the model in part (c) is:

```
> anova(ClothDye.lm)
Analysis of Variance Table

Response: score
                        Df Sum Sq Mean Sq F value    Pr(>F)    
operators                2 261.33 130.667 11.3172 0.0001007 ***
cycle.times              2 436.00 218.000 18.8812 1.031e-06 ***
temperatures             1  50.07  50.074  4.3370 0.0428819 *  
cycle.times:temperatures 2  78.81  39.407  3.4131 0.0414850 *  
Residuals               46 531.11  11.546                      
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The very small p-value associated with the operator term ($F_{2,46} = 11.3$, $p = 0.0001$) suggests that there are significant differences between the three operators. Attempting to control for these differences is the reason we included this term in the model, though the residual plots suggest that we have yet not been completely successful in controlling for all unwanted sources of variability. These problems cast doubts about using this model to address our main question of interest.

Ignoring these problems for the moment, the interaction term ($F_{2,46} = 3.4$, $p = 0.042$) between cycle times and temperatures is significant at $\alpha = 0.05$. This suggests that there are significant differences between the six treatment combinations (above and beyond the combined additive effects of the two main factors). The main effects terms for both cycle times ($F_{2,46} = 18.9$, $p = 0.000001$) and temperatures ($F_{1,46} = 4.3$, $p = 0.043$) are also significant.

To understand the differences between the treatment combinations, it does not really help to examine the model coefficients, as this is a fairly complicated model (regardless of how you have parameterised the model). The following parameter estimates table uses sum contrasts, consistent with the sum constraints used in the description of the model in part (c):

```
> round(summary(ClothDye.lm)$coef,5)
                          Estimate Std. Error  t value Pr(>|t|)
(Intercept)               31.55556    0.46240 68.24313  0.00000
operators1                -2.44444    0.65393 -3.73808  0.00051
operators2                 2.88889    0.65393  4.41773  0.00006
cycle.times1              -1.66667    0.65393 -2.54869  0.01421
cycle.times2               4.00000    0.65393  6.11685  0.00000
temperatures1             -0.96296    0.46240 -2.08254  0.04288
cycle.times1:temperatures1 -1.70370   0.65393 -2.60533  0.01232
cycle.times2:temperatures1  0.96296   0.65393  1.47258  0.14768
```

It is easier to examine the plots of the means for the treatment combinations shown in part (b). Judging by these plots, the significant interaction is a result of the relationship between the high and low temperatures at a cycle time of 40 minutes being completely different to the fairly consistent pattern shown for the other two cycle times.

As we are dealing with a balanced experimental design (the same number of replications for all treatment combinations) and as the model uses sum parameters, the estimated intercept parameter shown above equals the overall mean score. I have added this overall mean as a horizontal line on the plots shown in part (b).

The clients (the people who conducted the experiment and collected these data) are presumably interested in which combinations produce a better score than the overall mean, so it is fairly obvious that using either temperature with a cycle time of 50 minutes will produce a higher mean score than all of the other combinations; the mean scores for which are either not different from, or lower than, the overall mean.

Question 1, part d continued

To confirm this we could do hypotheses tests for each of the 6 different treatment combinations, to see if the mean score for each combination is significantly different from the overall mean:

H_0: $\mu_m = \mu$ vs H_A: $\mu_m \neq \mu$

where μ is the overall mean **score**,
m = 1, 2, 3, 4, 5, 6 indicates the six combinations of **ctime** and **temp**
= {(40 mins, 300 °F), (50 mins, 300 °F), (60 mins, 300 °F),
(40 mins, 350 °F), (50 mins, 350 °F), (60 mins, 350 °F)}, and
μ_m is the mean **score** for combination m

For example, a test of $\mu = \mu_1$ is equivalent to testing:

$$\mu_1 = \frac{\mu_1 + \mu_2 + \mu_3 + \mu_4 + \mu_5 + \mu_6}{6} \Rightarrow \tfrac{5}{6}\mu_1 - \tfrac{1}{6}\mu_2 - \tfrac{1}{6}\mu_3 - \tfrac{1}{6}\mu_4 - \tfrac{1}{6}\mu_5 - \tfrac{1}{6}\mu_6 = 0$$

This is a linear combination of the μ_m's. To calculate appropriate standard errors for this test, we can use the approach described on page 5 of the brick (also used in the *S-Plus* example shown on pages 7 and 8). I have used the *R* commands shown in the appendix to produce the following table:

```
> round(means.tests,5)
                     lvl.mns      est      se    tstat  pvalue
(40 mins, 300 degF)  27.22222 -4.33333 1.03396 -4.19102 0.00012
(50 mins, 300 degF)  35.55556  4.00000 1.03396  3.86864 0.00034
(60 mins, 300 degF)  29.00000 -2.55556 1.03396 -2.47163 0.01721
(40 mins, 350 degF)  32.55556  1.00000 1.03396  0.96716 0.33852
(50 mins, 350 degF)  35.55556  4.00000 1.03396  3.86864 0.00034
(60 mins, 350 degF)  29.44444 -2.11111 1.03396 -2.04178 0.04693
```

The above table includes a total of 6 comparisons, so we could use the Bonferroni correction to adjust our level of significance to account for the fact we are making multiple comparisons. Using an adjusted significance level of $\alpha / 6 = 0.05 / 6 = 0.0083$, the means for the cycle times of 50 minutes (in combination with either of the two temperatures) are the only combinations that produce mean scores that are significantly higher than the overall mean.

However, we really only need to make one comparison to decide if the two means with a cycle time of 50 minutes are significantly higher than all the rest, and that is to compare the mean for either of these two combinations with the next best combination mean, the mean for a cycle time of 40 minutes at 350 °F. Using a similar approach to the above to find a 95% confidence interval for the difference in the means for a cycle time of 50 minutes and a cycle time of 40 minutes (both at a temperature of 350 °F):

$$\mu_5 - \mu_4 \pm t_{46}(0.975)\, s\sqrt{\tfrac{1}{n_5} + \tfrac{1}{n_4}} \approx (35.556 - 32.556) \pm (2.013)(3.398)\sqrt{\tfrac{2}{9}} \approx (0.987, 5.013)$$

As this confidence interval is entirely positive, it suggests that a cycle time of 50 minutes does produce a significantly better score. So, assuming the experimenters are interested in which cycle times and temperatures produce the highest scores, a cycle time of 50 minutes appears to be optimal. However, the significant interaction between cycle time and temperature, which affects the 40 and 60 minutes cycle times, suggests there might be similar effects with other temperatures (even at a cycle time of 50 minutes), that were not covered in this experiment. Remembering that this is probably just one in a continuing series of experiments, I would recommend that the clients adopt a cycle time of 50 minutes, but further experiment with more temperature levels at this cycle time to find the optimal temperature. **(7 marks)**

Question 1 continued

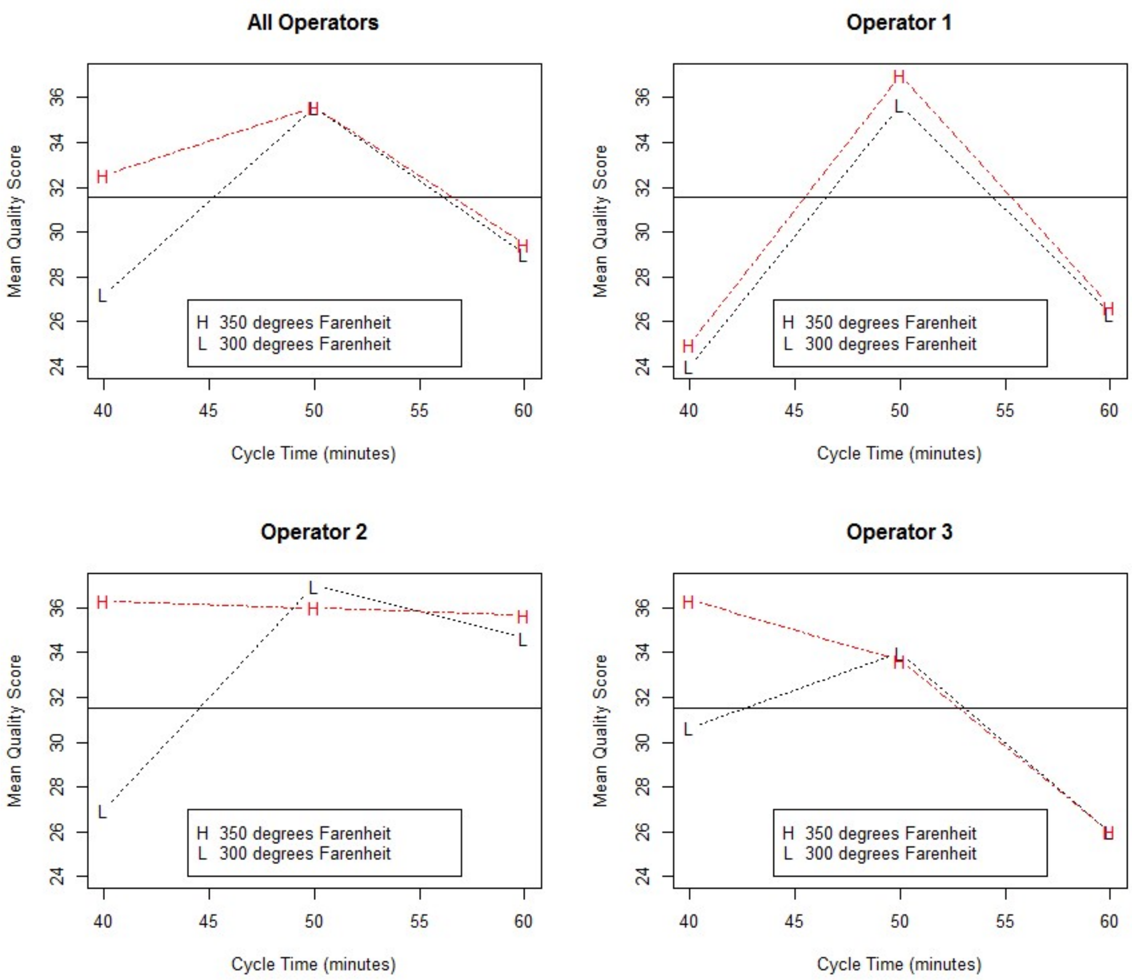
(e) Given the comments on the residual plots in part (c), an obvious alternative model is one that includes interactions between the operators and the main factors of interest; the cycle times and temperatures. The residual plots (not shown, but the commands are included in the appendix) for this model are much closer to being a random scatter of points without the obvious pattern that appeared in the residual plot in part (c). This suggests the model in part (c) was missing a significant interaction between the operators and one or both of the other factors.

The Analysis of Variance table for this expanded model is:

```
> anova(ClothDye.lm2)
Analysis of Variance Table

Response: score
                                   Df Sum Sq Mean Sq F value    Pr(>F)    
operators                           2 261.33 130.667 39.8644 7.439e-10 ***
cycle.times                         2 436.00 218.000 66.5085 8.141e-13 ***
temperatures                        1  50.07  50.074 15.2768 0.0003934 ***
operators:cycle.times               4 355.67  88.917 27.1271 1.982e-10 ***
operators:temperatures              2  11.26   5.630  1.7175 0.1938948    
cycle.times:temperatures            2  78.81  39.407 12.0226 0.0001002 ***
operators:cycle.times:temperatures  4  46.19  11.546  3.5226 0.0158701 *  
Residuals                          36 118.00   3.278                      
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

The three way interaction term ($F_{4,36} = 11.6$, $p = 0.016$) is significant at $\alpha = 0.05$, which suggests that different operators produce different results at different combinations of cycle times and temperatures. Again this is best illustrated by a series of factor plots:



Question 1, part e continued

We can examine differences between the combinations using t tests or confidence intervals. For example, a 95% confidence interval for the difference for operator 3 between the combinations (40 mins, 350 °F) and (50 mins, 350 °F) is:

$$(36.33333-33.66667)\pm t_{residual\,df=36}(0.975)\hat{\sigma}\sqrt{\tfrac{1}{3}+\tfrac{1}{3}}\approx 2.67\pm 3.00=(-0.33,5.67)$$

As this confidence interval includes 0, it suggests that the apparent difference between these two combinations is not significant.

As this is a balanced experiment (there are 3 replications for each of the 18 different combinations), the width of the confidence interval shown above (3.00) will be the same for comparing any two of the 18 combinations. This quantity is often calculated when analysing experimental designs and is called Fisher's least significant difference (LSD). If the means for two combinations do not differ by more than this amount, then they are not significantly different.

This expanded model is a better fit than the model in part (c) and is therefore a more appropriate model to use to answer the main research question. However, the answers suggested by the above factor plots are consistent with the earlier analysis. All three operators produce good results at a cycle time of 50 minutes (with almost no differences due to the different temperatures) and (applying the LSD) no operator produces significantly better results at any other cycle time. So, I would still recommend to the clients that they use a cycle time of 50 minutes, but that they also need to continue their experiments to further examine differences due to different operators and possibly also increase the range of temperatures that they examine. **(4 marks)**

(f) The description of the model is much the same as in part (c):

$$Y_{ijkl}=\mu+\delta_j+\eta_k+\gamma_l+\lambda_{kl}+\varepsilon_{ijkl}$$

We are now assuming that the operator term δ_j is a series of random effects which are independently and identically distributed N(0, σ_δ^2) and which are also independent of the residual variation in the model (the errors), which are independently and identically distributed N(0, σ_ε^2). **(3 marks)**

(g) The Analysis of Variance table for this mixed effects model is:

```
> anova(ClothDye.lmer)
Analysis of Variance Table
                        Df Sum Sq Mean Sq F value
cycle.times              2 436.00 218.000 18.8812
temperatures             1  50.07  50.074  4.3370
cycle.times:temperatures 2  78.81  39.407  3.4131
```

Apart from changes in the formatting of this table (different *R* functions were written by different statisticians who have different tastes and preferences when it comes to presenting output), the F statistics are identical to those shown in the part (d), so the earlier analysis of the fixed effects has not changed. **(2 marks)**

Question 1 continued

(h) The summary output for this mixed effects model is shown below:

```
> summary(ClothDye.lmer)
Linear mixed model fit by REML ['lmerMod']
Formula: score ~ (1 | factor(operators)) + cycle.times * temperatures

REML criterion at convergence: 280.2

Scaled residuals:
     Min       1Q   Median       3Q      Max
-2.15635 -0.65019 -0.04997  0.68292  2.01582

Random effects:
 Groups            Name        Variance Std.Dev.
 factor(operators) (Intercept)  6.618   2.573
 Residual                      11.546   3.398
Number of obs: 54, groups:  factor(operators), 3

Fixed effects:
                           Estimate Std. Error t value
(Intercept)                 31.5556     1.5556  20.286
cycle.times1                -1.6667     0.6539  -2.549
cycle.times2                 4.0000     0.6539   6.117
temperatures1               -0.9630     0.4624  -2.083
cycle.times1:temperatures1  -1.7037     0.6539  -2.605
cycle.times2:temperatures1   0.9630     0.6539   1.473

Correlation of Fixed Effects:
            (Intr) cycl.1 cycl.2 tmprt1 cy.1:1
cycle.tims1  0.000
cycle.tims2  0.000 -0.500
temperatrs1  0.000  0.000  0.000
cycl.tms1:1  0.000  0.000  0.000  0.000
cycl.tms2:1  0.000  0.000  0.000  0.000 -0.500
```

$\sigma_\delta^2 = 6.618$ is the estimated variance component associated with the operators, and

$\sigma_\varepsilon^2 = 11.546$ is the estimated residual (within operator) variance, so:

$\sigma_\delta^2 / \left(\sigma_\delta^2 + \sigma_\varepsilon^2\right) = 6.618/(6.618 + 11.546) \approx 36\%$

A reasonable proportion (36%) of the overall variability is due to variation between blocks (operators), so blocking has been effective. **(2 marks)**