

STAT 581 - Problem Set 6

Alex Towell (atowell@siue.edu)

Problem 1

A soft drink bottler is interested in studying the effects on a filling process. A factorial experiment is run using three factors: percent carbonation (in %), operating pressure (in psi), and line speed (in bpm). The response variable is deviation from the target fill height. The experiment consists of $n = 2$ runs for each of the $abc = 12$ treatment combinations. The data is available on Blackboard as an Excel File.

Preliminary analysis

The three-factor factorial experiment with $a = 3$ levels of the *carbonation* factor (in %), denoted by A , $b = 2$ levels of the operating *pressure* (in psi), denoted by B , and $c = 2$ levels of the line *speed* (in bpm), denoted by C . Each treatment combination has $n = 2$ replicates for a total of $N = abcn = 24$ observations. The response variable y is deviation from the target fill height.

We describe the observations in the three-factor factorial experiment by the fixed-effects interaction model where treatments are defined to be deviations from the overall mean,

$$Y_{ijkl} = \mu + \tau_i + \beta_j + \gamma_k + (\tau\beta)_{ij} + (\tau\gamma)_{ik} + (\beta\gamma)_{jk} + (\tau\beta\gamma)_{ijk} + \epsilon_{ijkl} \begin{cases} i = 1, \dots, a = 3 \\ j = 1, \dots, b = 2 \\ k = 1, \dots, c = 2 \\ \ell = 1, \dots, n = 2 \end{cases}$$

where μ is the overall main effect, τ_i is the effect of the i -th level of factor A , β_j is the effect of the j -th level of factor B , and γ_k is the effect of the k -th level of factor C . The interaction effects may also be observed using the same labeling, e.g., $(\tau\gamma)_{ik}$ is the interaction effect of the i -th level of factor A and the k -th level of factor C .

(a)

Perform tests for all main effects and for all interaction effects. State the F -statistic and p -value for each of the main effect tests, and for the test of an interaction between carbonation and pressure.

```
library("readxl")
h6.data = read_excel("handout6data.xlsx")

A = as.factor(na.omit(h6.data$carbon))
B = as.factor(na.omit(h6.data$pressure))
C = as.factor(na.omit(h6.data$speed))
y = na.omit(h6.data$deviation)

contrasts(A) = contr.sum # we use these in the additional analysis section
contrasts(B) = contr.sum # to make it easier to provide explicit estimates of
contrasts(C) = contr.sum # the parameters

x.mod = aov(y ~ A*B*C)
summary(x.mod)
```

```
##           Df Sum Sq Mean Sq F value    Pr(>F)
## A           2 252.75   126.38  178.412 1.19e-09 ***
## B           1  45.37    45.37   64.059 3.74e-06 ***
## C           1  22.04    22.04   31.118 0.00012 ***
## A:B         2   5.25     2.63    3.706 0.05581 .
## A:C         2   0.58     0.29    0.412 0.67149
## B:C         1   1.04     1.04    1.471 0.24859
## A:B:C       2   1.08     0.54    0.765 0.48687
## Residuals  12   8.50     0.71
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The tests for each of the main effects and a test of an interaction between carbon (A) and pressure (B) is given by the following table:

| effects | F^* | p -value |
|--------------|-------|------------|
| A | 178.4 | .000 |
| B | 64.1 | .000 |
| C | 31.1 | .000 |
| $A \times B$ | 3.7 | .056 |

(b)

Explain how an interaction plot is used in studying an interaction effect.

If an interaction plot is parallel, or nearly so, then there is no need to include the interaction term.

In the 3-factor ANOVA model, a 3-factor interaction occurs when a 2-factor interaction depends on the level of the 3rd factor.

(c)

Determine fitted values for the main effects model, as well as fitted values for the model including the carbonation / pressure interaction. Plot the fitted values for each model using an interaction plot. Comment on the decision of whether or not to include the interaction term.

From the earlier results, we saw that factors A (carbon), B (pressure), and C (speed) were statistically significant, but only the interaction between factor A and B (carbon and pressure) had any significance.

Additive model

The additive model is given by

$$Y_{ijk\ell}^{(R)} = \mu + \tau_i + \beta_j + \gamma_k + \epsilon_{ijk\ell}.$$

Thus,

$$E(Y_{ijk}^{(R)}) = \mu + \tau_i + \beta_j + \gamma_k.$$

We fit the additive model with:

```
mod.R = aov(y ~ A+B+C)
```

We define the estimated means for the models with:

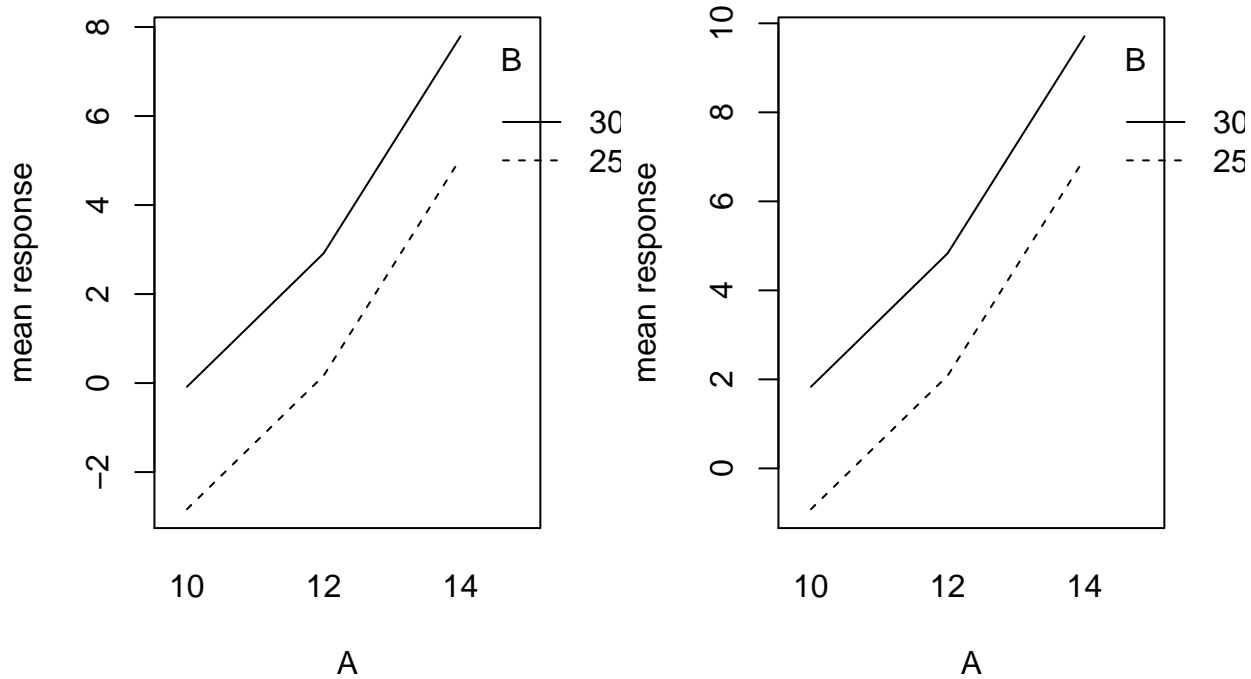
```
fits.R = predict(mod.R)
```

The interaction plots for the additive where the left and right plots are respectively for factor C at levels 1 and 2.

```

par(mfrow=c(1,2))
for (i in 1:nlevels(C))
{
  interaction.plot(A[C==levels(C)[i]],B[C==levels(C)[i]],
    fits.R[C==levels(C)[i]],
    xlab="A",
    ylab="mean response",
    trace.label="B")
}

```



Interaction model

The model that includes interactions between factors A (carbon) and B (pressure) is given by

$$Y_{ijk\ell}^{(F)} = \mu + \tau_i + \beta_j + \gamma_k + (\tau\beta)_{ij} + \epsilon_{ijk\ell}.$$

Thus,

$$E(Y_{ijk}^{(F)}) = \mu + \tau_i + \beta_j + \gamma_k + (\tau\beta)_{ij}.$$

We fit the interaction model with:

```
mod.F = aov(y ~ A+B+C+A:B)
```

We define the estimated means for the models with:

```
fits.F = predict(mod.F)
```

The interaction plots for the interaction model where the left and right plots are respectively for factor C at levels 1 and 2.

```

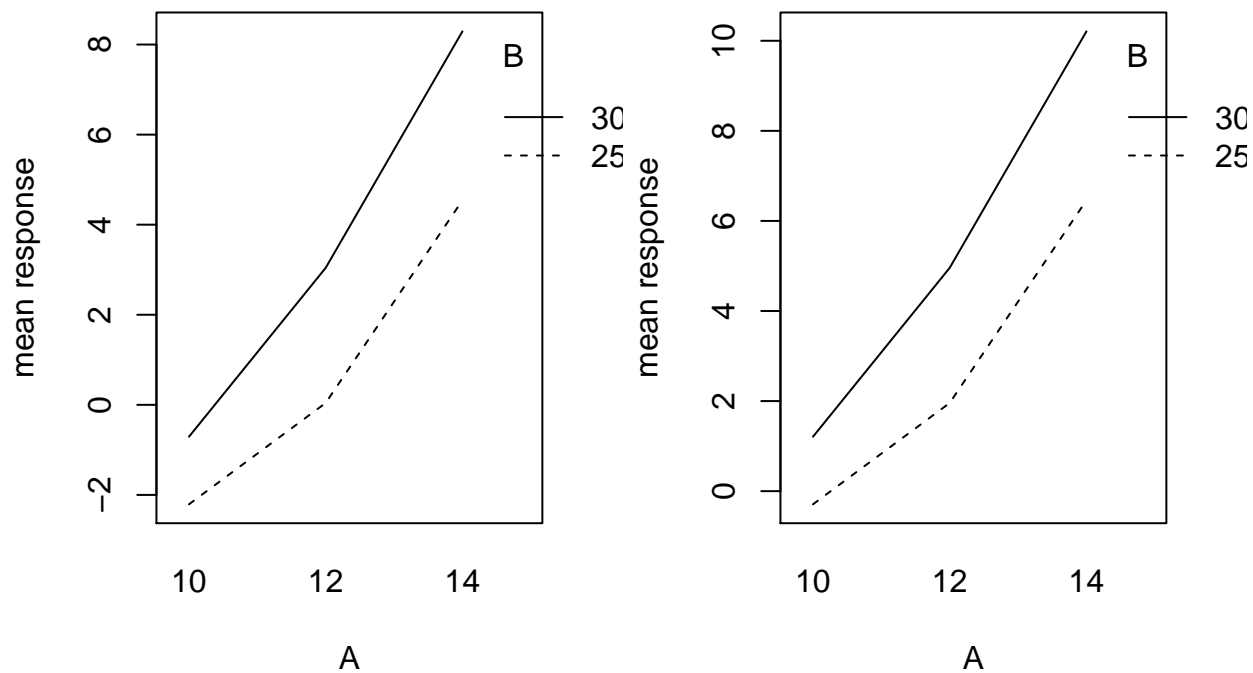
par(mfrow=c(1,2))
for (i in 1:nlevels(C))
{
  interaction.plot(A[C==levels(C)[i]],B[C==levels(C)[i]],
    fits.F[C==levels(C)[i]],
    xlab="A",

```

```

      ylab="mean response",
      trace.label="B")
}

```



Decision

The interaction model fits show only a slight difference from the additive model. As a general rule, we prefer simpler models to complex models, and so we decide to leave out the interaction term.

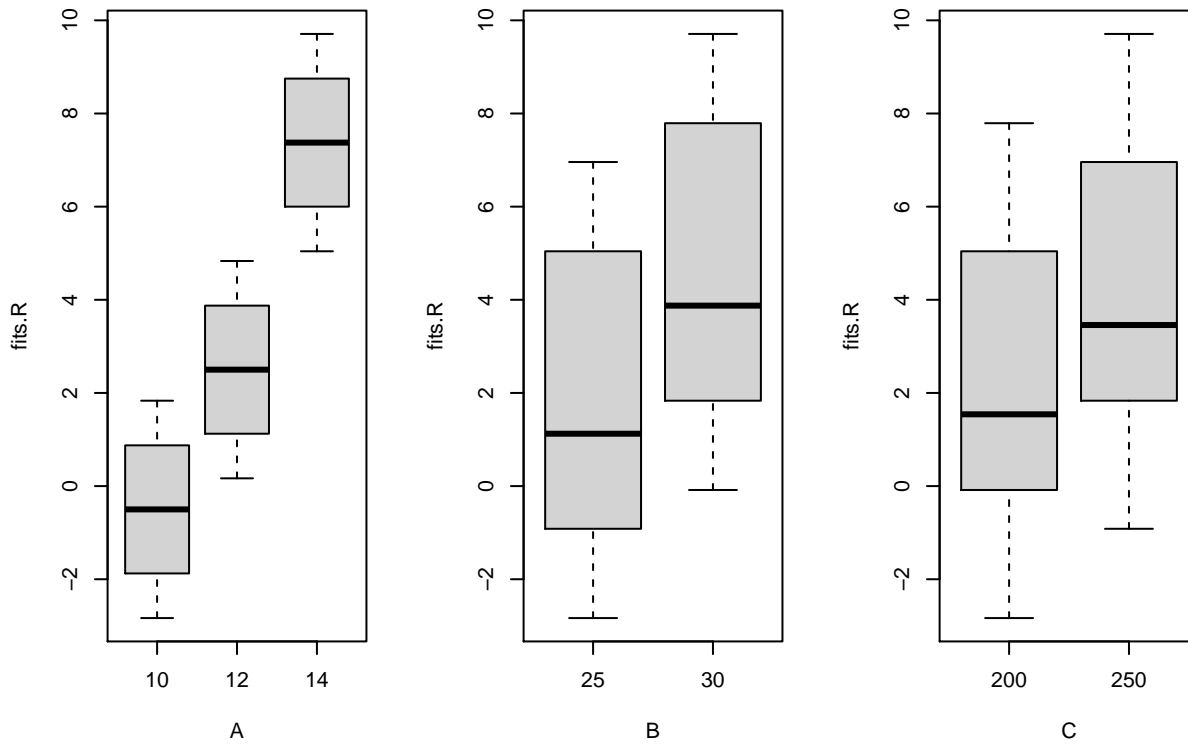
(d)

Create box plots showing the main effects. Provide an overall conclusion, stated in the context of the problem.

```

par(mfrow=c(1,3))
plot(fits.R ~ A)
plot(fits.R ~ B)
plot(fits.R ~ C)

```



We see that each factor has a positive effect on fill height. The factor with the biggest marginal effect on fill height is the percent of carbonation (factor A).

Problem 2

An engineer is interested in the effects of cutting speed, tool geometry, and cutting angle on the life (in hours) of a machine tool. Two levels of each factor are set, and $n = 3$ runs of a 2^3 design are completed. The data is available on Blackboard as an Excel File.

(a)

Perform tests for all main effects and for all interaction effects. State the F -statistic and p -value for each test of an effect deemed to be important.

```
data2 = read_excel("handout6data.xlsx")
```

```
A = as.factor(na.omit(data2$spd))
B = as.factor(na.omit(data2$geo))
C = as.factor(na.omit(data2$angle))
y = na.omit(h6.data$life)
```

```
contrasts(A) = contr.sum
contrasts(B) = contr.sum
contrasts(C) = contr.sum
```

```
saturated.mod = aov(y ~ A*B*C)
summary(saturated.mod)
```

| ## | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|--------|----|--------|---------|---------|--------------|
| ## A | 1 | 0.7 | 0.7 | 0.022 | 0.883680 |
| ## B | 1 | 770.7 | 770.7 | 25.547 | 0.000117 *** |
| ## C | 1 | 280.2 | 280.2 | 9.287 | 0.007679 ** |
| ## A:B | 1 | 16.7 | 16.7 | 0.552 | 0.468078 |

```
## A:C          1  468.2   468.2  15.519 0.001172 **
## B:C          1   48.2    48.2   1.597 0.224475
## A:B:C        1   28.2    28.2   0.934 0.348282
## Residuals    16  482.7    30.2
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Factors B and C are important effects and the interaction between A and C is also important. Thus, A , B , C , and $A : C$ should be included in the model (we include A because of the importance of $A : C$).

| effects | F^* | p -value |
|---------|--------|------------|
| A | .022 | .884 |
| B | 25.547 | .000 |
| C | 9.287 | .008 |
| $A : C$ | 15.519 | .001 |

(b)

Fit a reduced model with the main effects, and the statistically significant interaction effect.

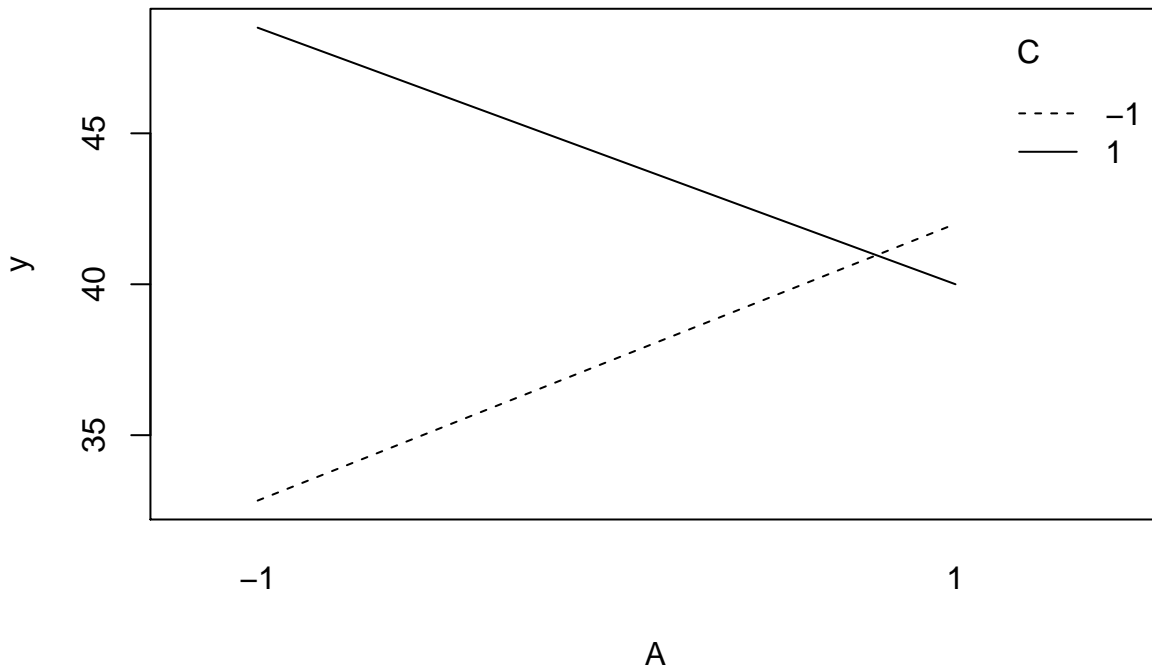
```
reduced.mod = aov(y ~ A+B+C+A:C)
summary(reduced.mod)
```

```
##          Df Sum Sq Mean Sq F value    Pr(>F)
## A          1    0.7      0.7    0.022 0.883641
## B          1  770.7   770.7   25.436 7.22e-05 ***
## C          1  280.2   280.2    9.247 0.006724 **
## A:C         1  468.2   468.2   15.452 0.000897 ***
## Residuals   19  575.7    30.3
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

(i)

Create a plot for the interaction effect. Provide an interpretation, stated in the context of the problem.

```
# define the estimated means for the reduced model for plotting (randomness has
# been smoothed)
fits.r = predict(reduced.mod)
interaction.plot(A,C,fits.r,ylab="y")
```

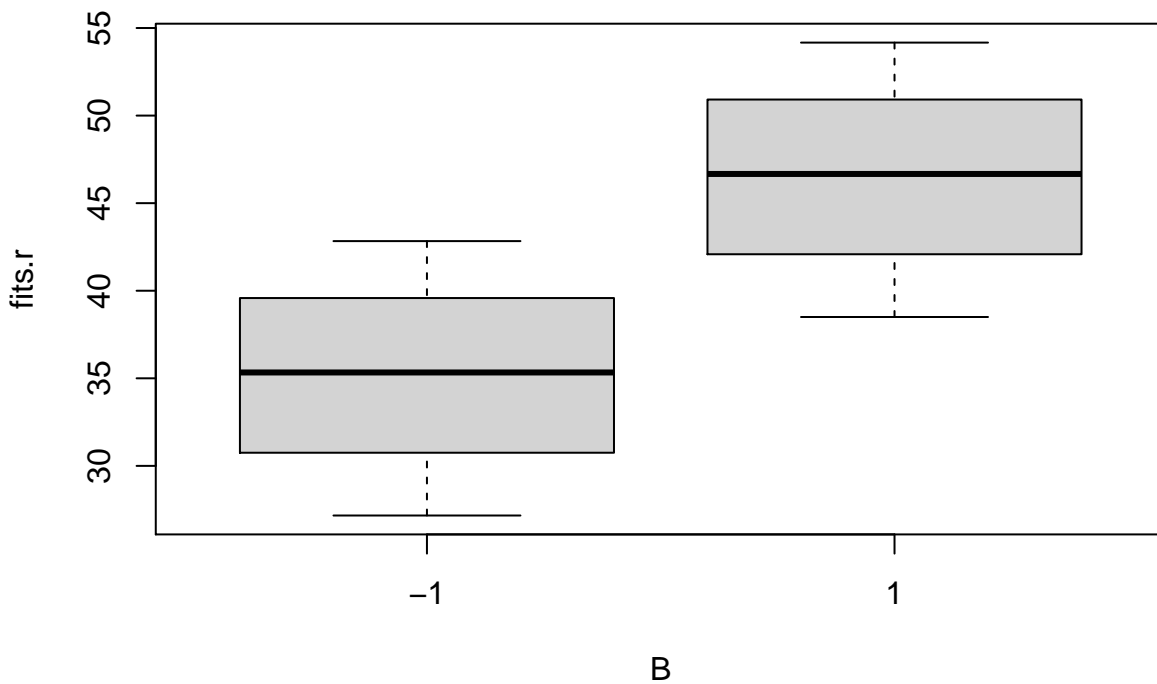


Interpretation The experiment finds that when cutting speed (factor A) is low, cutting angle has a positive effect on the life of a machine tool (response).

(ii)

Create a box plot showing the main effect for the remaining factor. Provide an interpretation, stated in the context of the problem.

```
plot(fits.r ~ B)
```

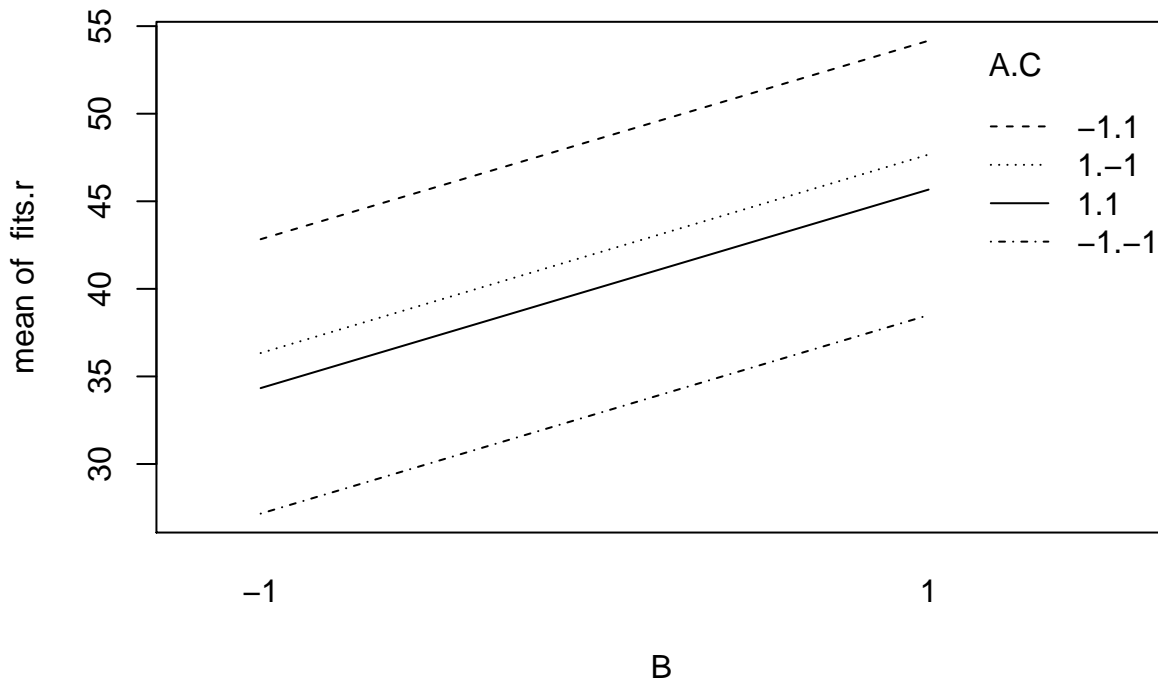


Interpetation Tool geometry (factor B) has a positive effect on the life of a machine tool.

(iii)

Create a plot of the fitted values for the reduced model. Which setting of the factors should be used if the goal is to maximize tool life. Explain how this result follows from the interaction plot and the main effects plot.

```
A.C = interaction(A,C)
interaction.plot(B,A.C, fits.r)
```



Maximum setting In this characterization experiment, we find that the best combination of factor levels is given by a low cutting speed (A^-), a high tool geometry (B^+), and a high cutting angle (C^+).

Explanation First, we observe that A^- and C^+ has the best interaction effect (its corresponding line is above all of the others). Second, we see that B has a positive effect, i.e., as we go from B^- to B^+ , the response increases, and thus B^+ is better. Taking these two facts together, we see that the best combination in this experiment is (A^-, B^+, C^+) .

We may also just look at the highest point on the graph and see that (A^-, B^+, C^+) has the highest mean response. However, the interaction plot reveals more, particularly if in follow-up experiments where we use what we learned in this experiment to narrow down our search for an optimal combination (where optimality may be with respect to, say, mean response or minimal variability in the response).

Appendix

Problem 1: additional analysis of the additive model

The model parameter estimates are given by:

```
dummy.coef(mod.R)
```

```
## Full coefficients are
##
## (Intercept):      3.125
## A:               10      12      14
##                -3.625  -0.625  4.250
## B:               25      30
##                -1.375   1.375
```



```
## C:                200        250
##                -0.9583333  0.9583333
```

Thus, the estimated mean response $\hat{Y}_{ijk}^{(R)}$ is given by

$$\begin{aligned}\hat{Y}_{ijk}^{(R)} = & 3.125 - I(i=1)3.625 - I(i=2)6.25 + I(i=3)4.250 \\ & - I(j=1)1.375 + I(j=2)1.375 \\ & - I(k=1)9.58\bar{3} + I(k=2)9.58\bar{3}.\end{aligned}$$

In R code, this may be brute-force implemented as:

```
yhat.R = function(i,j,k)
{
  y = 3.125
  if (i == 1) y = y - 3.625; if (i == 2) y = y - .625; if (i == 3) y = y + 4.25
  if (j == 1) y = y - 1.375; if (j == 2) y = y + 1.375
  if (k == 1) y = y - .95833333; if (k == 2) y = y + .95833333
  y
}
```

We print out all of the estimates of the mean response with:

```
for (i in 1:3)
  for (j in 1:2)
    for (k in 1:2)
      cat("Y(",c(i,j,k),") = ",yhat.R(i,j,k),"\n")
```

```
## Y( 1 1 1 ) = -2.833333
## Y( 1 1 2 ) = -0.9166667
## Y( 1 2 1 ) = -0.08333333
## Y( 1 2 2 ) = 1.833333
## Y( 2 1 1 ) = 0.1666667
## Y( 2 1 2 ) = 2.083333
## Y( 2 2 1 ) = 2.916667
## Y( 2 2 2 ) = 4.833333
## Y( 3 1 1 ) = 5.041667
## Y( 3 1 2 ) = 6.958333
## Y( 3 2 1 ) = 7.791667
## Y( 3 2 2 ) = 9.708333
```

Problem 1: plots of interactions that were deemed insignificant

The data showed no statistically significant interaction effects between factors A and B and between factors B and C . Thus, their interaction plots are, as expected, parallel:

```
par(mfrow=c(1,2))
interaction.plot(A,B,fits.r,ylab="y")
interaction.plot(B,C,fits.r,ylab="y")
```

