HW4

Luke Beebe

2024-04-06

5.1 T/F

An interaction effect in the model from a factorial experiment involving quantitative factors is a way of incorporating curvature into the response surface model representation of the results.

False, curvature can be represented by adding higher-order terms.

5.2

A factorial experiment may be conducted as a RCBD by running each replicate of the experiment in a unique block

True, this helps to control potential sources of variation.

5.3

If an interaction effect in a factorial experiment is significant, the main effects of the factors involved in that interaction are difficult to interpret individually.

True, interpreting the main effects of the individual factors is challenging because the effect of one factor depends on the level of the other factor(s) involved.

5.7

The following output was obtained from a computer program that performed a two-factor ANOVA on a factorial experiment.

Two-way ANOVA: y versus A, B

Source	DF	SS	MS	F	P
A	1	?	0.0002	?	?
B	?	180.378	?	?	?
Interaction	3	8.479	?	?	0.932
Error	8	158.797	?		
Total	15	347.653			

a. *Fill in the blanks in the ANOVA table. You can use bounds on the P-values.*

Two-way ANOVA: y versus A, B

Source	DF	SS	MS	F	P
A	1	?0.0002	0.0002	?0.00001	?0.998
B	?3	180.378	? 60.126	?3.029	?0.094
Interaction	3	8.479	? 2.826	?0.142	0.932
Error	8	158.797	?19.8496		
Total	15	347.653			

b. *How many levels were used for factor B?*

4

c. *How many replicates of the experiment were performed?*

4

d. *What conclusions would you draw about this experiment?*

We do not reject the null. None of the factors are statistically significant from one another.

5.9

An engineer suspects that the surface finish of a metal part is influenced by the feed rate and the depth of cut. He selects three feed rates and four depths of cut. He then conducts a factorial experiment and obtains the following data:

```
surface <- c(74,79,82,99,
             64,68,88,104,
             60,73,92,96,

             92,98,99,104,
             86,104,108,110,
             88,88,95,99,

             99,104,108,114,
             98,99,110,111,
             102,95,99,107)
feed_rate <- c(0.2,0.2,0.2,0.2,
               0.2,0.2,0.2,0.2,
               0.2,0.2,0.2,0.2,

               0.25,0.25,0.25,0.25,
               0.25,0.25,0.25,0.25,
               0.25,0.25,0.25,0.25,

               0.3,0.3,0.3,0.3,
               0.3,0.3,0.3,0.3,
               0.3,0.3,0.3,0.3)
depth <- c(0.15,0.18,0.2,0.25,
           0.15,0.18,0.2,0.25,
           0.15,0.18,0.2,0.25,

           0.15,0.18,0.2,0.25,
           0.15,0.18,0.2,0.25,
           0.15,0.18,0.2,0.25,

           0.15,0.18,0.2,0.25,
           0.15,0.18,0.2,0.25,
           0.15,0.18,0.2,0.25)
data <- data.frame(surface, feed_rate, depth)
```

a. *Analyze the data and draw conclusions. Use $\alpha = 0.05$.*

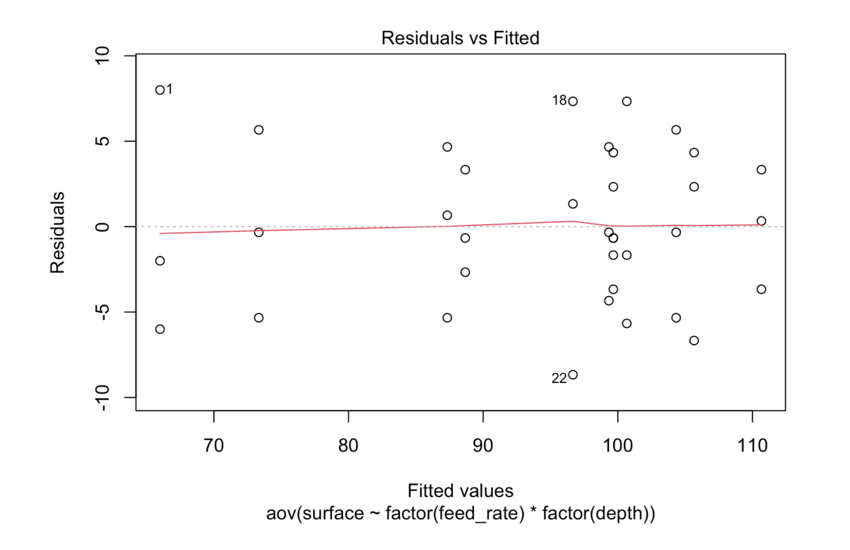
```
model <- aov(surface ~ factor(feed_rate)*factor(depth), data)
summary(model)
```

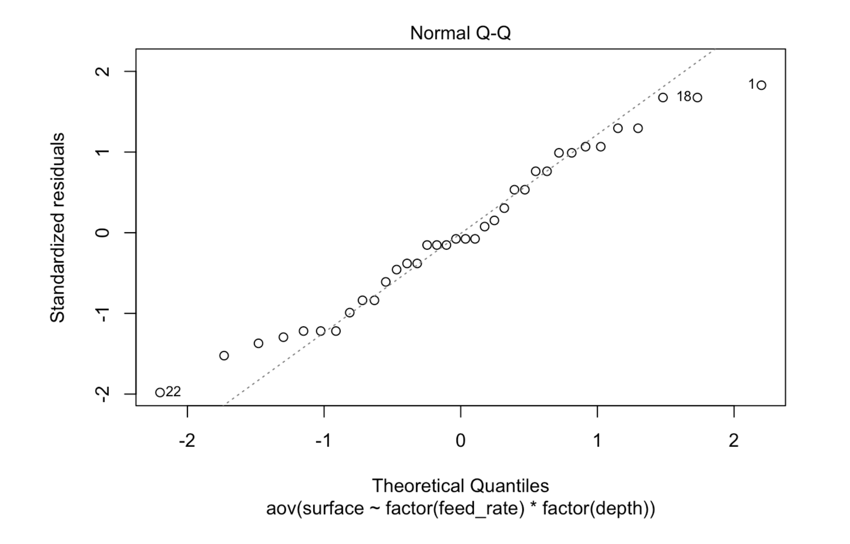
```
##                                 Df Sum Sq Mean Sq F value   Pr(>F)    
## factor(feed_rate)                2 3160.5  1580.2  55.018 1.09e-09 ***
## factor(depth)                    3 2125.1   708.4  24.663 1.65e-07 ***
## factor(feed_rate):factor(depth)  6  557.1    92.8   3.232    0.018 *  
## Residuals                       24  689.3    28.7                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

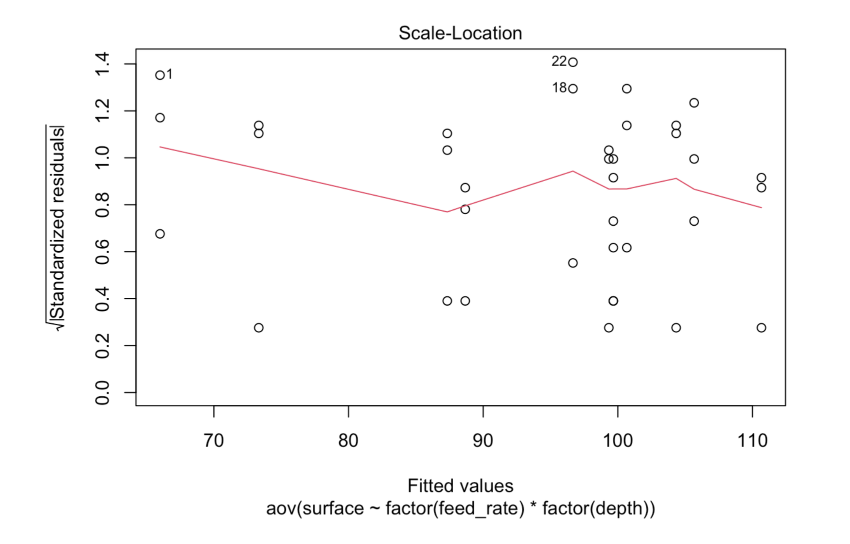
The feed_rate, depth, and interaction term feed_rate:depth are all significant.

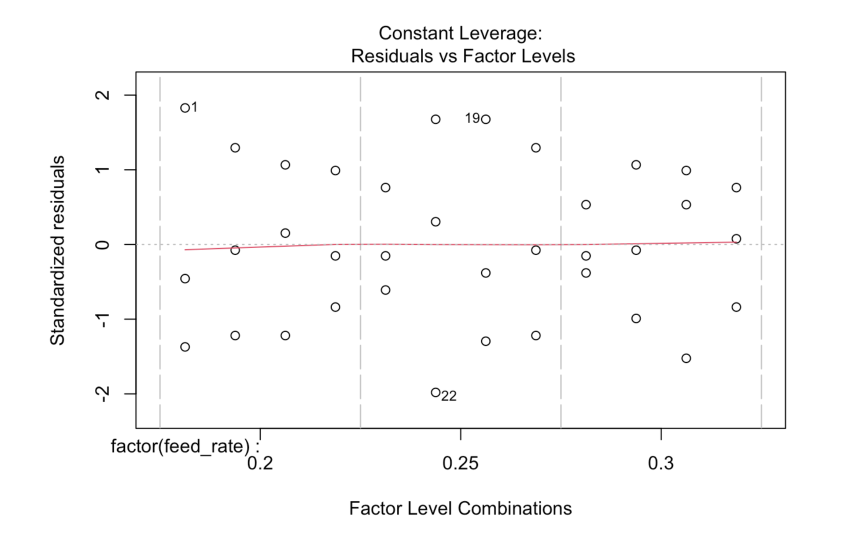
b. *Prepare appropriate residual plots and comment on the model's adequacy.*

```
plot(model)
```









Nothing looks too wonky with our residuals.

c. *Obtain point estimates of the mean surface finish at each feed rate.*

```
df = NULL
for(rate in unique(feed_rate)){
  avg = sum(data[feed_rate==rate,]$surface)/length(data[feed_rate==rate,]$surface)
  df <- rbind(df, data.frame(rate, avg))
}
ggplot(df, aes(x=factor(rate), y=avg, group=1)) + geom_line(linetype="dashed") + geom_point()
```

d. *Find the P-values for the tests in part (a).*

feed_rate, p = 1.09e-09
depth, p = 1.65e-07
interaction, p = 0.018

5.10

For the data in Problem 5.9, compute a 95 percent confidence interval estimate of the mean difference in response for feed rates of 0.20 and 0.25 in/min.

```
t.test(data[feed_rate==0.2,]$surface, data[feed_rate==0.25,]$surface, paired=T)
```

```
## 
##  Paired t-test
## 
## data:  data[feed_rate == 0.2, ]$surface and data[feed_rate == 0.25, ]$surface
## t = -5.3742, df = 11, p-value = 0.0002253
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
##  -22.552742  -9.447258
## sample estimates:
## mean difference 
##             -16
```

The confidence interval estimate of mean difference in response for feed rates of 0.2 and 0.25 in/min is approximately **(-22.55, -9.45)**. Meaning that there is a significant difference.

5.17

Use Tukey's test to determine which levels of the pressure factor are significantly different for the data in Problem 5.8.

```
yield <- c(90.4, 90.7, 90.2,
           90.2, 90.6, 90.4,

           90.1, 90.5, 89.9,
           90.3, 90.6, 90.1,

           90.5, 90.8, 90.4,
           90.7, 90.9, 90.1)
temp <- c(150, 150, 150,
          150, 150, 150,

          160, 160, 160,
          160, 160, 160,

          170, 170, 170,
          170, 170, 170)
pressure <- c(200, 215, 230,
              200, 215, 230,

              200, 215, 230,
              200, 215, 230,

              200, 215, 230,
              200, 215, 230)
data <- data.frame(yield, temp, pressure)
model <- aov(yield ~ factor(temp)*factor(pressure), data)
TukeyHSD(model)$`factor(pressure)`
```

```
##               diff        lwr         upr        p adj
## 215-200  0.3166667  0.1017380  0.53159536 0.0066518030
## 230-200 -0.1833333 -0.3982620  0.03159536 0.0944905140
## 230-215 -0.5000000 -0.7149287 -0.28507131 0.0002951232
```

The pressure of **215-200** and **230-215** are significantly different.

5.23

The percentage of hardwood concentration in raw pulp, the vat pressure, and the cooking time of the pulp are being investigated for their effects on the strength of paper. Three levels of hardwood concentration, three levels of pressure, and two cooking times are selected. A factorial experiment with two replicates is conducted, and the following data are obtained:

```
strength <- c(196.6,197.7,199.8,
              196.0,196.0,199.4,

              198.5,196.0,198.4,
              197.2,196.9,197.6,

              197.5,195.6,197.4,
              196.6,196.2,198.1,


              198.4,199.6,200.6,
              198.6,200.4,200.9,

              197.5,198.7,199.6,
              198.1,198.0,199.0,

              197.6,197.0,198.5,
              198.4,197.8,199.8)

hardwood_conc <- c(rep(2,6),
                   rep(4,6),
                   rep(8,6),

                   rep(2,6),
                   rep(4,6),
                   rep(8,6))

pressure <- c(rep(c(400, 500, 600), 12))

cooking_time <- c(rep(3, 18),
                  rep(4, 18))

data <- data.frame(strength, hardwood_conc, pressure, cooking_time)
```

a. *Analyze the data and draw conclusions. Use $\alpha = 0.05$.*

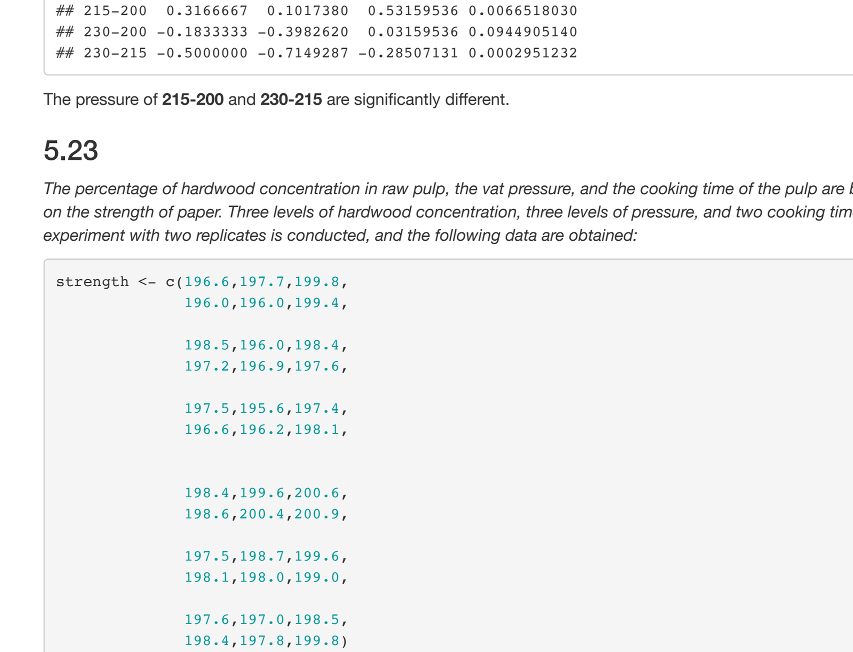
```
model <- aov(strength ~ factor(hardwood_conc)*factor(pressure) + factor(pressure):factor(cooking_time) + factor(h
ardwood_conc):factor(cooking_time) + factor(cooking_time), data)
summary(model)
```

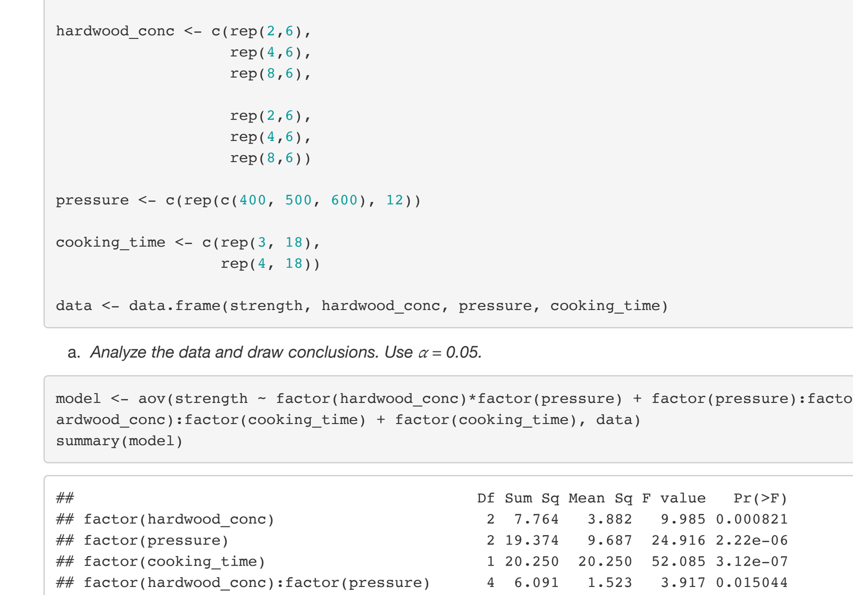
```
##                                           Df Sum Sq Mean Sq F value   Pr(>F)
## factor(hardwood_conc)                      2  7.764   3.882   9.985 0.000821
## factor(pressure)                           2 19.374   9.687  24.916 2.22e-06
## factor(cooking_time)                       1 20.250  20.250  52.085 3.12e-07
## factor(hardwood_conc):factor(pressure)     4  6.091   1.523   3.917 0.015044
## factor(pressure):factor(cooking_time)      2  2.195   1.097   2.823 0.081047
## factor(hardwood_conc):factor(cooking_time)  2  2.082   1.041   2.677 0.091070
## Residuals                                 22  8.553   0.389                 
##                                                   
## factor(hardwood_conc)                      ***
## factor(pressure)                           ***
## factor(cooking_time)                       ***
## factor(hardwood_conc):factor(pressure)     *  
## factor(pressure):factor(cooking_time)      .  
## factor(hardwood_conc):factor(cooking_time) .  
## Residuals                                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

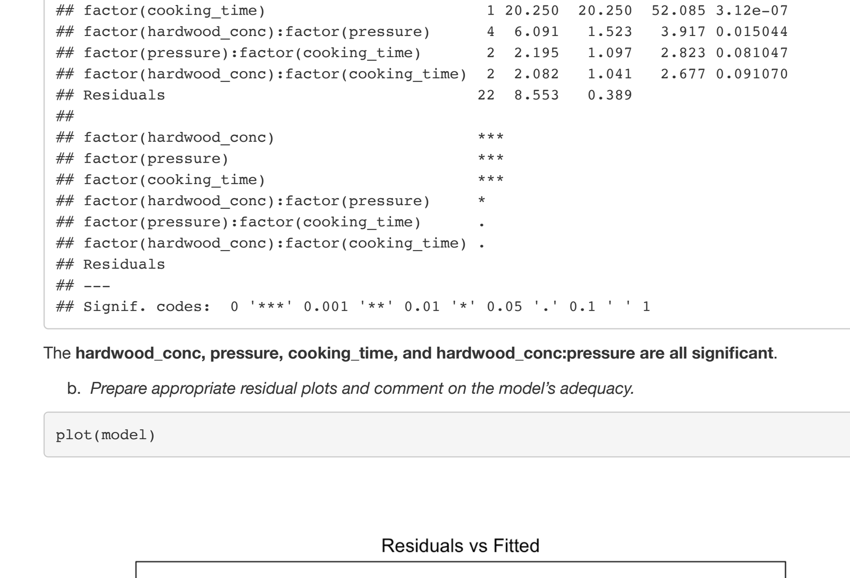
The **hardwood_conc, pressure, cooking_time, and hardwood_conc:pressure are all significant**.

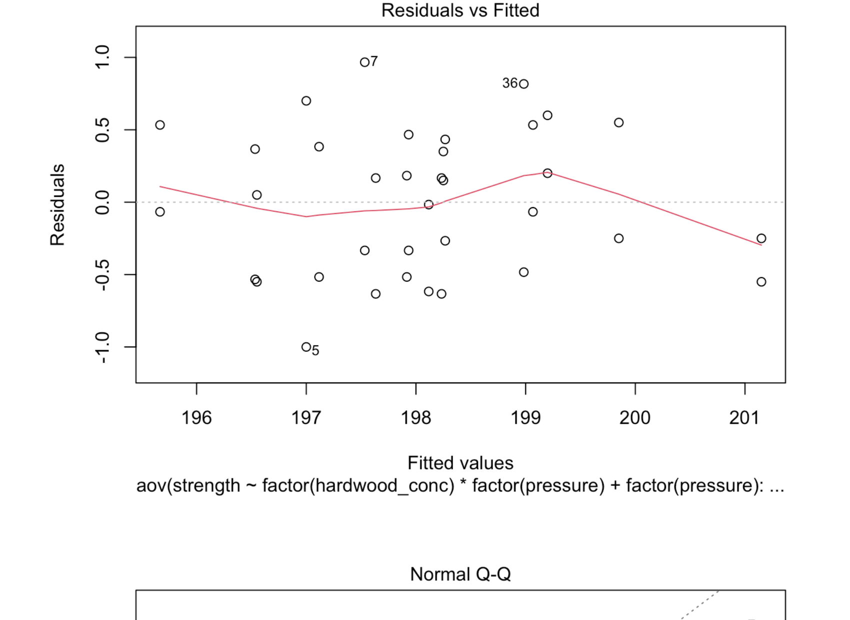
b. *Prepare appropriate residual plots and comment on the model's adequacy.*

```
plot(model)
```









Nothing looks too wonky with our residuals.

c. *Under what set of conditions would you operate this process? Why?*

Linear relationship, homoscedastic, independent and independent, identically distributed samples. These are the conditions to run a valid ANOVA model.

5.39

Reconsider the experiment in Problem 5.9. Suppose that this experiment had been conducted in three blocks, with each replicate a block. Assume that the observations in the data table are given in order, that is, the first observation in each cell comes from the first replicate, and so on. Reanalyze the data as a factorial experiment in blocks and estimate the variance component for blocks. Does it appear that blocking was useful in this experiment?

```
surface <- c(74,79,82,99,
             64,68,88,104,
             60,73,92,96,

             92,98,99,104,
             86,104,108,110,
             88,88,95,99,

             99,104,108,114,
             98,99,110,111,
             102,95,99,107)
feed_rate <- c(0.2,0.2,0.2,0.2,
               0.2,0.2,0.2,0.2,
               0.2,0.2,0.2,0.2,

               0.25,0.25,0.25,0.25,
               0.25,0.25,0.25,0.25,
               0.25,0.25,0.25,0.25,

               0.3,0.3,0.3,0.3,
               0.3,0.3,0.3,0.3,
               0.3,0.3,0.3,0.3)
depth <- c(0.15,0.18,0.2,0.25,
           0.15,0.18,0.2,0.25,
           0.15,0.18,0.2,0.25,

           0.15,0.18,0.2,0.25,
           0.15,0.18,0.2,0.25,
           0.15,0.18,0.2,0.25,

           0.15,0.18,0.2,0.25,
           0.15,0.18,0.2,0.25,
           0.15,0.18,0.2,0.25)
block <- c(rep(1,4),
           rep(2,4),
           rep(3,4),

           rep(1,4),
           rep(2,4),
           rep(3,4),

           rep(1,4),
           rep(2,4),
           rep(3,4))
data <- data.frame(surface, feed_rate, depth, block)
model <- aov(surface ~ feed_rate*depth + block)
summary(model)
```

```
##                 Df Sum Sq Mean Sq F value   Pr(>F)    
## feed_rate        1 2970.4  2970.4  95.328 5.66e-11 ***
## depth            1 2042.3  2042.3  65.543 3.84e-09 ***
## block            1  140.2   140.2   4.498 0.042029 *  
## feed_rate:depth  1  413.2   413.2  13.262 0.000978 ***
## Residuals       31  965.9    31.2                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Yes, it does appear that blocking was useful in this experiment. There is a statistically significant difference in variance between random error and the blocks.