STAT 581 - Problem Set 5

Alex Towell (atowell@siue.edu)

Problem 1

A factorial experiment is used to develop a nitride etch process on a single wafer plasma etching tool. The design factors are the gap between the electrodes and the RF power applied to the cathode. The response variable is the etch rate for silicon nitride. The data is available on Blackboard as an Excel File.

(a)

Provide a general definition of an interaction effect.

An interaction effect occurs when the effect of one factor depends on the level of the other factors.

(b)

State the two factor with interaction ANOVA model. Compute the estimates of the model parameters.

Two-factor with interaction ANOVA model

We consider a two-factor factorial experiment with $a = 2$ levels of the *gap* factor, denoted by A, and $b = 2$ levels of the *RP power* factor, denoted by B, for a total of $ab = 4$ treatment combinations. Each treatment combination has $n = 4$ replicates for a total of $N = nab = 16$ observations.

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

$$Y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \epsilon_{ijk} \begin{cases} i = 1, \ldots, a \\ j = 1, \ldots, b \\ k = 1, \ldots, n \end{cases}$$

where μ is the overall main effect, τ_i is the effect of the i-th level of factor A, β_j is effect of the j-th level of factor B, $(\tau\beta)_{ij}$ is the interaction effect of the i-th level of factor A and the j-th level of factor B, and ϵ_{ijk} is the random error component.

Compute the estimates of the model parameters

A set of estimators for these model parameters are given by

$$\begin{aligned} \hat{\mu} &= \bar{y}_{\cdots} \\ \hat{\tau}_i &= \bar{y}_{i\cdot\cdot} - \hat{\mu} \\ \hat{\beta}_j &= \bar{y}_{\cdot j\cdot} - \hat{\mu} \\ \widehat{(\tau\beta)}_{ij} &= \bar{y}_{ij\cdot} - \hat{\mu} - \hat{\tau}_i - \hat{\beta}_j, \end{aligned}$$

and thus an estimator of the k-th observation given factor A at level i and factor B at level j is given by

$$\hat{y}_{ijk} = \hat{\mu} + \hat{\tau}_i + \hat{\beta}_j + \widehat{(\tau\beta)}_{ij},$$

which simplifies to

$$\hat{y}_{ijk} = \bar{y}_{ij\cdot}.$$

We perform these computations using R:

```
library("readxl")
data = read_excel("handout5data.xlsx")
gap = as.factor(na.omit(data$gap))
rfp = as.factor(na.omit(data$rfp))
rate = na.omit(data$rate)

# use contrasts to define parameter restrictions for the two way ANOVA model
contrasts(gap) = contr.sum
contrasts(rfp) = contr.sum

# we are still using aov to perform ANOVA computations.
# for two way ANOVA, we include two factors plus an interaction in the model statement
two.way.mod = aov(rate ~ gap+rfp+gap:rfp)

# the command below is used to print the model parameter estimates
dummy.coef(two.way.mod)
```

```
## Full coefficients are
##
## (Intercept):     776.0625
## gap:                  0.8       1.2
##                   50.8125  -50.8125
## rfp:                  275       325
##                 -153.0625  153.0625
## gap:rfp:          0.8:275   1.2:275  0.8:325  1.2:325
##                  -76.8125   76.8125  76.8125 -76.8125
```

We see that $\hat{\tau}_1 = 50.8125$. By the constraint that $\tau_1 + \tau_2 = 0$, $\hat{\tau}_2 = -50.8125$. The remaining parameter estimates can be observed directly on the R output, e.g., $\widehat{(\tau\beta)}_{21} = 76.8125$.

Remark

Assuming the model is appropriate, would it be reasonable to say that an estimator of the mean response given factor A at level i and factor B at level j is given by

$$\hat{E}(Y_{ij}|A = i, B = j) = \bar{y}_{ij\cdot}?$$

Since we may be primarily just be looking for evidence of effects, coming up with a more sophisticated, explanatory model may not be warranted at this stage of the analysis.

If we find some evidence that there are some factors or interactions thereof that effect the response, we may then, if deemed useful, try to use a more restrictive model. For instance, a general linear regression model may not only provide a more accurate and precise estimator of the conditional response, but also more explanatory power (e.g., the coefficients in a linear regression model have an obvious interpretation).

(c)

Test for an interaction effect. Compute the test statistic F_{AB} and the p-value. Provide an interpretation, stated in the context of the problem.

We test for an interaction effect with:

```r
summary(two.way.mod)
```

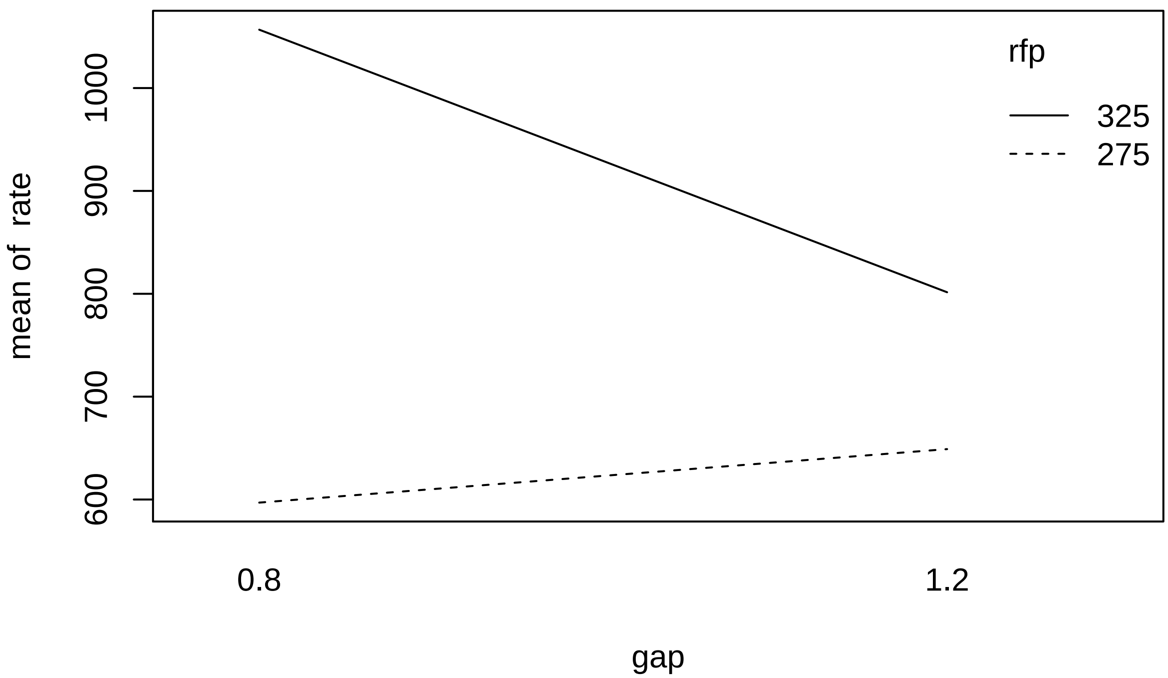
```
##             Df Sum Sq Mean Sq F value   Pr(>F)    
## gap          1  41311   41311   23.77 0.000382 ***
## rfp          1 374850  374850  215.66 4.95e-09 ***
## gap:rfp      1  94403   94403   54.31 8.62e-06 ***
## Residuals   12  20858    1738                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We see that $F_{AB} = 54.31$ which has a p-value of .000. Thus, the experiment finds an interaction effect between the gap between the electrodes and the RF power applied to the cathode on the etch rate for silicon nitride.

(d)

Graph an interaction plot.

```r
# the command below is used to create an interaction plot
# the x.factor will be plotted on the horizontal axis.
# the trace.factor will be plotted as separate curves
interaction.plot(x.factor=gap,trace.factor=rfp,response=rate)
```



This interaction effect demonstrates an interaction effect, since the effect of one factor depends on the level of the other factor.

(e)

Perform pairwise comparisons using the Fisher LSD method to investigate differences between the treatment combinations. Provide the grouping information.

Fisher LSD

```r
library("multcomp")

# because of the interaction effect, we will need to perform comparisons on
```

```r
# treatment combinations.
# comparisons on factor effects may be misleading since a factor effect depends
# on the level of the other factor.

# create one factor with a*b=4 levels
combined = interaction(gap,rfp)

# compute Fisher comparisons between treatment (4 choose 2)=6 combinations.
comb.mod = aov(rate~combined)
comparisons = glht(comb.mod,linfct=mcp(combined="Tukey"))
summary(comparisons,test=univariate())
```

```
##
##   Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: aov(formula = rate ~ combined)
##
## Linear Hypotheses:
##                         Estimate Std. Error t value Pr(>|t|)
## 1.2.275 - 0.8.275 == 0    52.00      29.48   1.764 0.103166
## 0.8.325 - 0.8.275 == 0   459.75      29.48  15.595 2.49e-09 ***
## 1.2.325 - 0.8.275 == 0   204.50      29.48   6.937 1.57e-05 ***
## 0.8.325 - 1.2.275 == 0   407.75      29.48  13.831 9.79e-09 ***
## 1.2.325 - 1.2.275 == 0   152.50      29.48   5.173 0.000232 ***
## 1.2.325 - 0.8.325 == 0  -255.25      29.48  -8.658 1.66e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Univariate p values reported)
```

Grouping information

```r
cld(summary(comparisons,test=univariate()))
```

```
## 0.8.275 1.2.275 0.8.325 1.2.325
##     "a"     "a"     "c"     "b"
```

(f)

Provide an overall interpretation to the analysis, stated in the context of the problem.

The experiment finds that the gap has no effect on etch rate when RF power is set at 275, but that an RF power of 325 leads to greater overall etch rates.

Note that the optimal treatment combination in this experiment is a gap of \$0.8 and an RF power of 325.

Problem 2

The factors that affect the breaking strength of a synthetic fiber are being studied. A factorial experiment is run using the four production machines and the three operators. The data is available on Blackboard as an Excel File.

(a)

Fit the interaction model. Test for an interaction effect and for main effects. Compute the test statistics F_A, F_B, F_{AB}, and the p-values. Provide an interpretation, stated in the context of the problem.

Let

$$\begin{aligned} A &:= \text{mach} \\ B &:= \text{op} \\ \text{response} &:= \text{str}. \end{aligned}$$

We fit the interaction model

$$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \epsilon_{ijk}$$

with the following R code:

```
mach = as.factor(na.omit(data$machine))
op = as.factor(na.omit(data$operator))
str = na.omit(data$strength)

# fit the interaction model, where mach*op denotes mach + op + mach:op
x.mod = aov(str~mach*op)
summary(x.mod)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## mach         3  12.46    4.15   1.095 0.388753
## op           2 160.33   80.17  21.143 0.000117 ***
## mach:op      6  44.67    7.44   1.963 0.150681
## Residuals   12  45.50    3.79
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We see that

$$\begin{aligned} F_A &= 1.095 \\ F_B &= 21.143 \\ F_{AB} &= 7.44 \end{aligned}$$

and

$$\begin{aligned} p_A &= .388753 \\ p_B &= .000117 \\ p_{AB} &= .150681. \end{aligned}$$

Interpretation

The experiment finds that the operator has an effect on strength and the machine has no effect on strength. In addition, the experiment finds there is no interaction effect.

(b)

Fit the additive model. Perform pairwise comparisons using the Fisher LSD method to investigate differences between operators. Provide the grouping information.

Fitting the additive model

Having seen some evidence that interaction effects are unimportant, we fit the additive model

$$y_{ijk} = \mu + \tau_i + \beta_j + \epsilon_{ijk}$$

with the following R code:

```
a.mod = aov(str~mach+op)
summary(a.mod)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## mach         3  12.46    4.15   0.829 0.495098
## op           2 160.33   80.17  16.004 0.000101 ***
## Residuals   18  90.17    5.01
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Fisher LSD comparisons for operator

```
compare.op = glht(a.mod, linfct=mcp(op="Tukey"))
summary(compare.op,test=univariate())
```

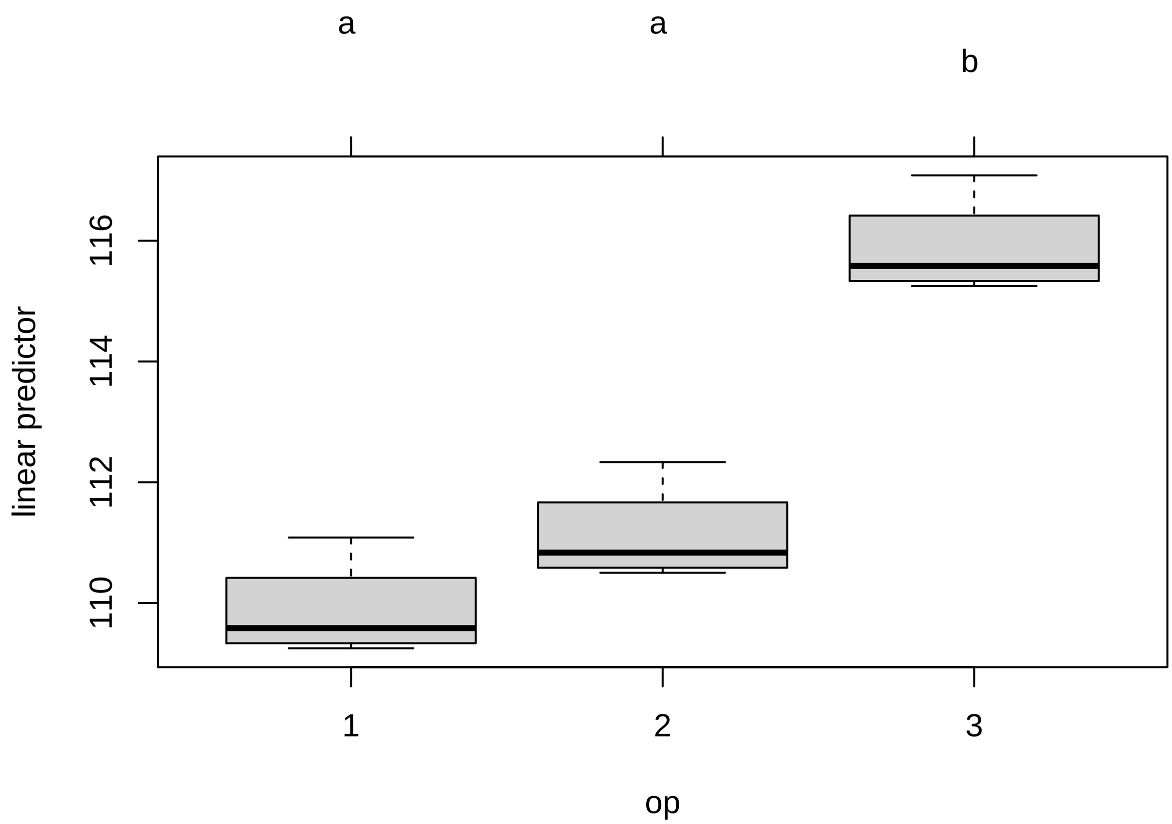
```
##
##   Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: aov(formula = str ~ mach + op)
##
## Linear Hypotheses:
##            Estimate Std. Error t value Pr(>|t|)
## 2 - 1 == 0    1.250      1.119   1.117 0.278683
## 3 - 1 == 0    6.000      1.119   5.362 4.27e-05 ***
## 3 - 2 == 0    4.750      1.119   4.245 0.000487 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Univariate p values reported)
```

Grouping information

```
cld(summary(compare.op,test=univariate()))
```

```
##   1   2   3
## "a" "a" "b"
```

```
plot(cld(summary(compare.op,test=univariate())))
```



From the grouping information, we say that

$$\mu_3 < \mu_1 < \mu_2$$

and

$$\mu_3 < \mu_2 < \mu_1$$

are compatible with the data.

(c)

Compute confidence intervals for the between operator differences.

```
confint(compare.op,calpha=univariate_calpha())
```

```
##
##   Simultaneous Confidence Intervals
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: aov(formula = str ~ mach + op)
##
## Quantile = 2.1009
## 95% confidence level
##
##
## Linear Hypotheses:
##            Estimate lwr     upr
## 2 - 1 == 0  1.2500  -1.1011  3.6011
## 3 - 1 == 0  6.0000   3.6489  8.3511
## 3 - 2 == 0  4.7500   2.3989  7.1011
```

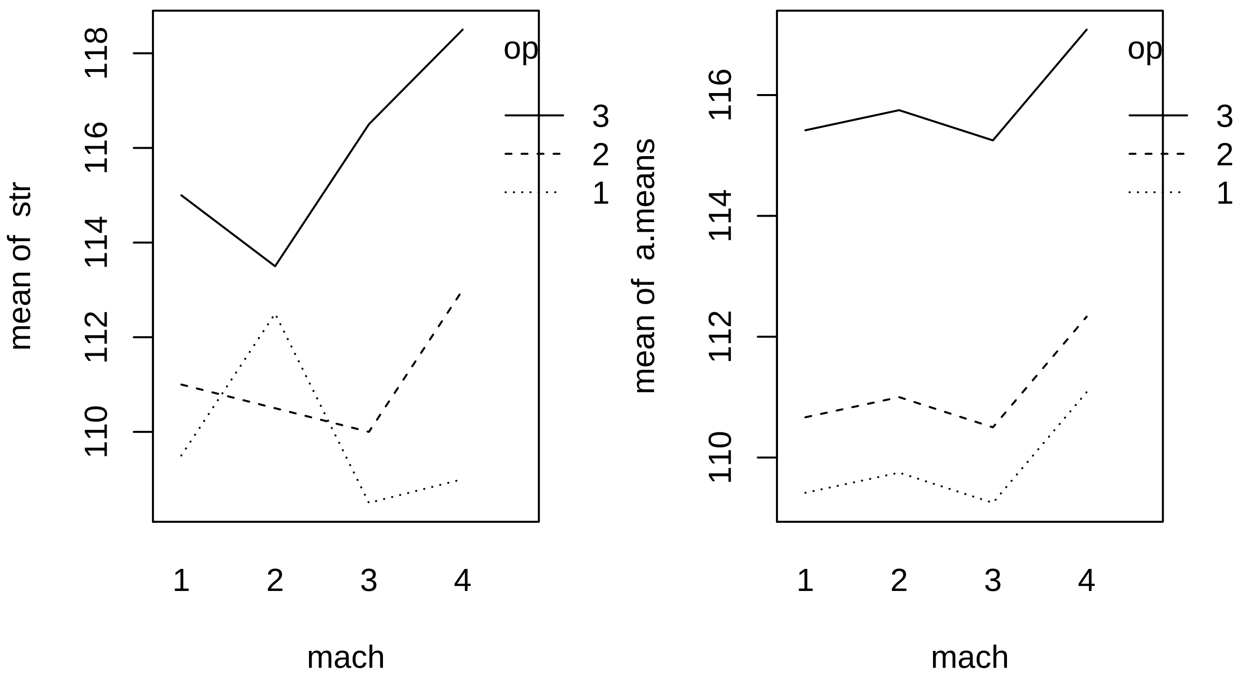
(d)

Graph an interaction plot for both the interaction model and the additive model. Explain how the analysis is providing a simplification to the observed data.

```r
par(mfrow=c(1,2))

# interaction model
interaction.plot(mach,op,str)

# additive model
# compute estimated means under the additive model
a.means = predict(a.mod)
interaction.plot(mach,op,a.means)
```



Explanation

The model smooths over the randomness in the data, simplifying the analysis.