STAT 571B - Homework 5
Brian Hallmark

5.25. An article in the *IEEE Transactions on Electron Devices* (Nov. 1986, pp. 1754) describes a study on polysilicon doping. The experiment shown below is a variation of their study. The response variable is base current.

R code:

```
p1=read.table("hw5_p1.txt",header=T)
p1.lm=aov(Current~Doping*Temp,data=p1)
summary(p1.lm)
par(mfrow=c(1,2))
plot(Current~Doping*Temp)
```

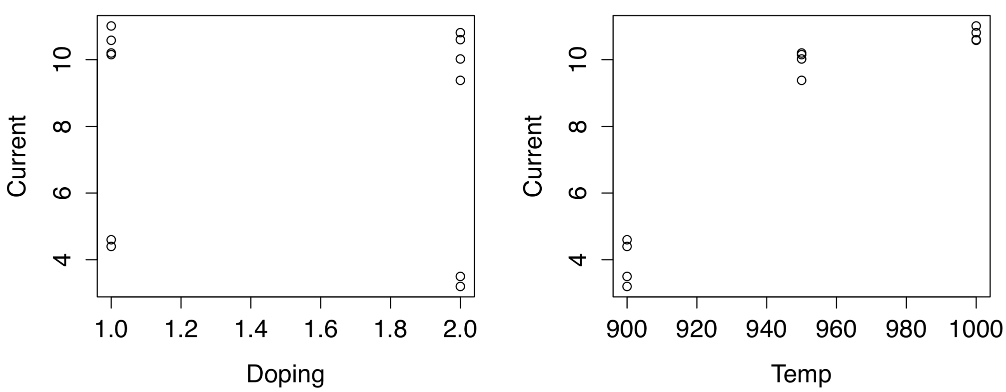
(a) Is there evidence (with $\alpha = 0.05$) indicating that either polysilicon doping level or anneal temperature affect base current?

```
             Df  Sum Sq  Mean Sq  F value    Pr(>F)
Doping        1    0.98     0.98    0.426  0.532422
Temp          1   93.16    93.16   40.447  0.000218 ***
Doping:Temp   1    0.56     0.56    0.244  0.634667
Residuals     8   18.43     2.30
```

Temperature is significant but doping and the interaction term are not.

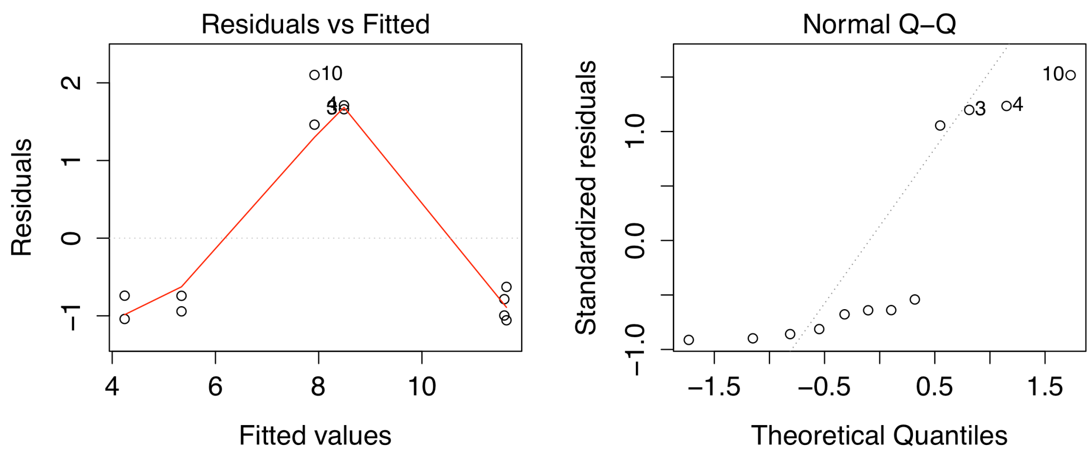
(b) Prepare graphical displays to assist in interpretation of this experiment.

These plots indicate why that is – current is not different at the two levels of doping, but it increases with increasing temperature



(c) Analyze the residuals and comment on model adequacy.

The diagnostic plots are troubling and indicate the model assumptions are not valid



(d) Is the model $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{22} x_2^2 + \beta_{12} x_1 x_2 + \varepsilon$ supported by this experiment (x_1 = doping level, x_2 = temperature)? Estimate the parameters in this model and plot the response surface.

R code:

```
p1.lm2<-lm(Current~Doping+Temp+I(Temp^2)+Doping:Temp,data=p1)
summary(p1.lm2)
summary.lm(p1.lm2)
plot(p1.lm2)
```

Output:

```
Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept) -9.775e+02  5.296e+01 -18.457 3.40e-07 ***
Doping      -1.064e+01  3.213e+00  -3.312   0.0129 *
Temp         2.028e+00  1.113e-01  18.221 3.71e-07 ***
I(Temp^2)   -1.040e-03  5.852e-05 -17.771 4.41e-07 ***
Doping:Temp  1.060e-02  3.379e-03   3.137   0.0164 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2389 on 7 degrees of freedom
Multiple R-squared: 0.9965,     Adjusted R-squared: 0.9944
F-statistic: 493.7 on 4 and 7 DF,  p-value: 1.175e-08
```

These results indicate a strong fit (R^2) and all the parameters are significant. NOTE: the parameters are for doping values of 1 and 2 not the original values.

I was not able to produce a nice response surface in R.

6.1. An engineer is interested in the effects of cutting speed (*A*), tool geometry (*B*), and cutting angle (C) on the life (in hours) of a machine tool. Two levels of each factor are chosen, and three replicates of a 2^3 factorial design are run. The results are as follows:

R code:

```
p2=read.table("hw5_p2.txt",header=T)
p2.lm=aov(Life~A*B*C,data=p2)
summary(p2.lm)
summary.lm(p2.lm)
plot.design(Life~A*B*C,data=p2)
p2.effects=p2.lm$coefficients*2
p2.effects
p2.lm2=lm(Life~A+B+C+A:C,data=p2)
summary(p2.lm2)
plot(p2.lm2)
```

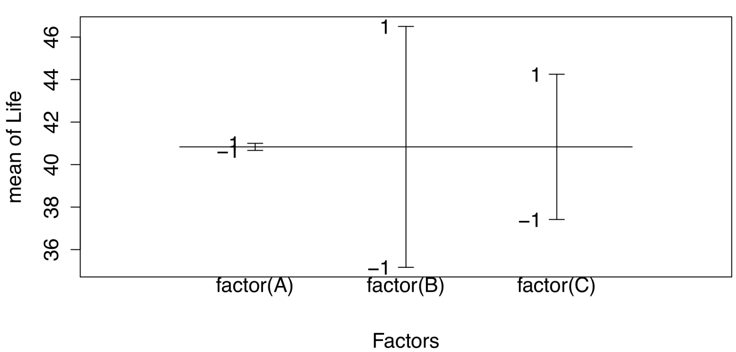
(a) Estimate the factor effects. Which effects appear to be large?

I computed this by hand and then verified it using the R code by taking the regression coefficients * 2

A	0.333
B	11.333
C	6.833
AB	-1.667
AC	-8.833
BC	-2.833
ABC	-2.167

The largest effects are B, C and AC.

This can also be seen with the plot.design() function:



(b) Use the analysis of variance to confirm your conclusions for part (a).

R ANOVA output:

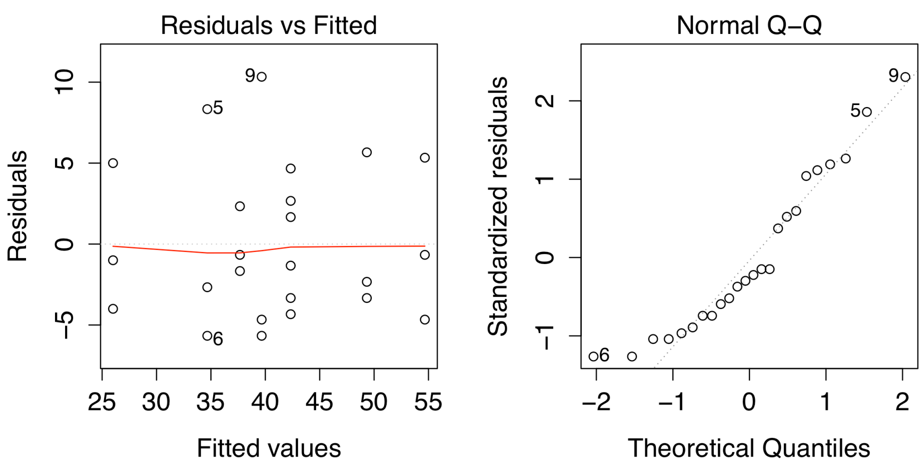
```
Df Sum Sq Mean Sq F value   Pr(>F)
A             1    0.7     0.7   0.022 0.883680
B             1  770.7   770.7  25.547 0.000117 ***
C             1  280.2   280.2   9.287 0.007679 **
```

```
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
```
This confirms that the only significant factors are B, C and the AC interaction.

(c) Write down a regression model for predicting tool life (in hours) based on the results of this experiment.

We can use the output from `summary.lm()` to write the regression equation:
$\hat{y} = 40.833 + 5.667x_2 + 3.417x_3 - 4.417x_2x_3$

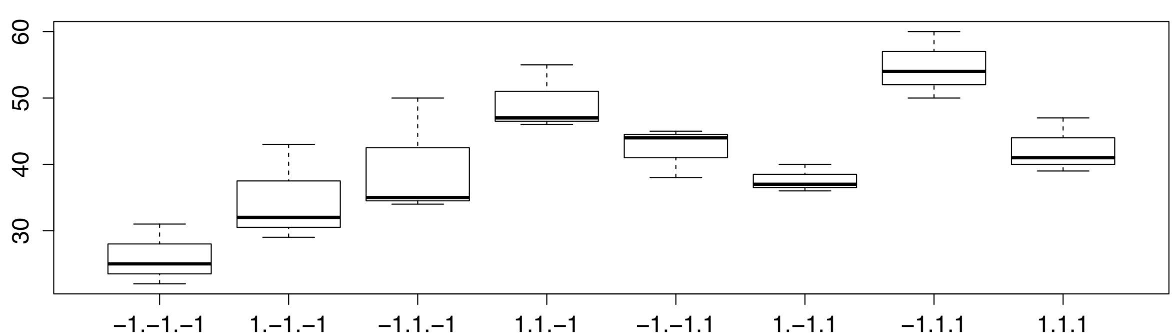
(d) Analyze the residuals. Are there any obvious problems?



There are no huge significant departures from normality or equal variances.

(e) Based on the analysis of main effects and interaction plots, what levels of *A*, *B*, and *C* would you recommend using?

I would recommend using the low level for A and the high levels for B and C, since that produces the longest life (as shown below in the boxplot of combinations).

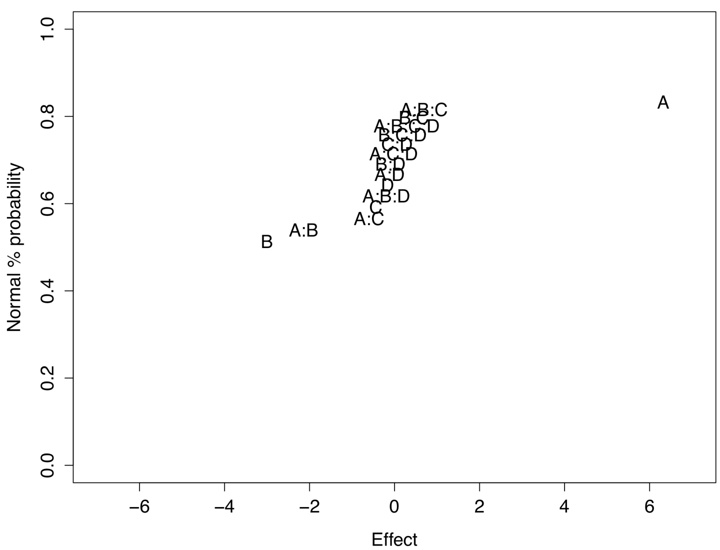


6.34. Resistivity on a silicon wafer is influenced by several factors. The results of a 2^4 factorial experiment performed during a critical process step is shown below.

(a) Estimate the factor effects. Plot the effect estimates on a normal probability plot and select a tentative model.

R code:

```
p3=read.table("hw5_p3.txt",header=T)
p3.lm=aov(Resistivity~A*B*C*D,data=p3)
summary.lm(p3.lm)
p3.effects=p3.lm$coefficients*2
p3.effects=p3.effects[-1]
x=sort(p3.effects)
length(x)
i=1:15
ii=(i-.5)/15
pnorm(ii)
y=pnorm(ii)
plot(x,y,ylim=c(0,1),pch="")
text(x,y,names(x))
```




```
       A         B         C         D       A:B       A:C       B:C       A:D
 6.32125 -3.00375 -0.44125 -0.15875 -2.13875 -0.59625  0.45875 -0.11375
     B:D       C:D     A:B:C     A:B:D     A:C:D     B:C:D   A:B:C:D
-0.09375  0.05875  0.68875 -0.19375 -0.02125  0.18875  0.28375
```

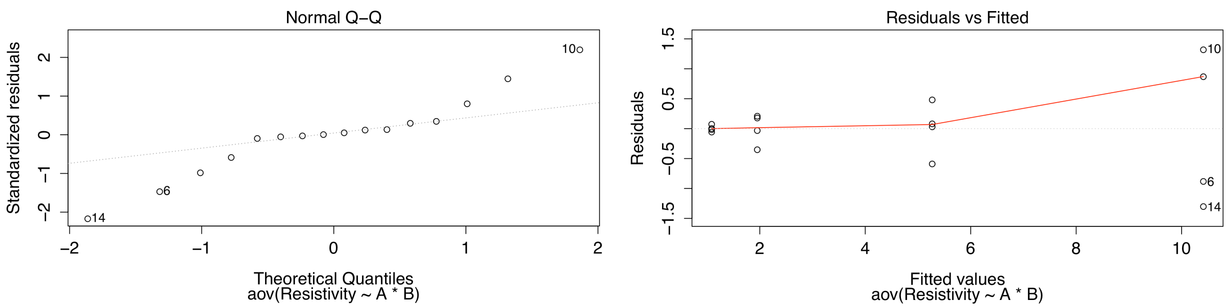
The largest effects are for factors A, B and the interaction AB.

(b) Fit the model identified in part (a) and analyze the residuals. Is there any indication of model inadequacy?

R code:

```
p3.lm=aov(Resistivity~A*B,data=p3
summary.lm(p3.lm)
summary(p3.lm)
plot(p3.lm)
```

<- NOTE: the ANOVA table indicated A, B and AB significant



Yes, the variance is not constant – it has a cone like shape. The QQ-plot also indicates non-normality.

(c) Repeat the analysis from parts (a) and (b) using ln(*y*) as the response variable. Is there any indication that the transformation has been useful?

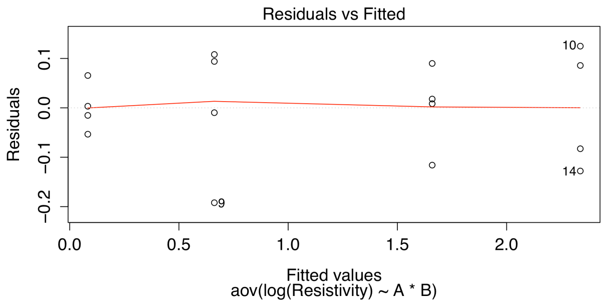
R code:

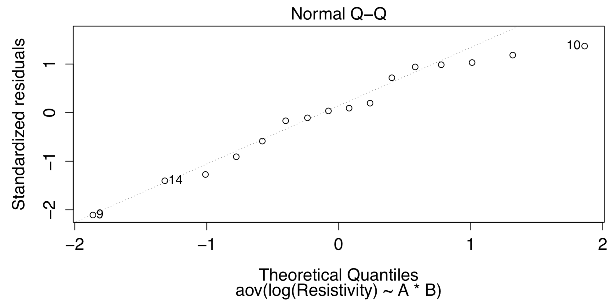
```
p3.lm3=aov(log(Resistivity)~A*B,data=p3)
summary(p3.lm3)
summary.lm(p3.lm3)
plot(p3.lm3)
```

Yes, the log transformation has changed things. Now the interaction term is now longer significant:

```
             Df Sum Sq Mean Sq F value   Pr(>F)
A             1 10.572  10.572  954.04 8.33e-13 ***
B             1  1.580   1.580  142.61 5.10e-08 ***
A:B           1  0.010   0.010    0.88    0.367
Residuals    12  0.133   0.011
```

It also made the residuals approximately normally distributed.





(d) Fit a model in terms of the coded variables that can be used to predict the resistivity.

Taking the output from summary.lm(p3.lm3) I get the equation:

$$\ln(y) = 1.185 + 0.813x_1 - 0.314x_2$$

7.1 Consider the experiment described in Problem 6.1. Analyze this experiment assuming that each replicate represents a block of a single production shift.

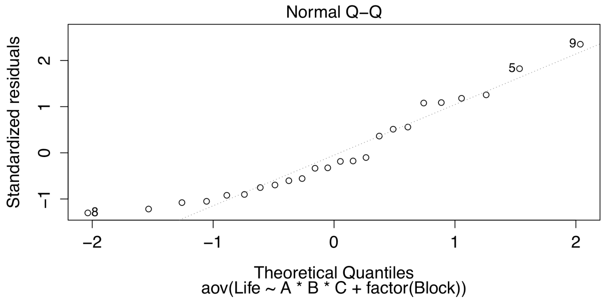
R code:

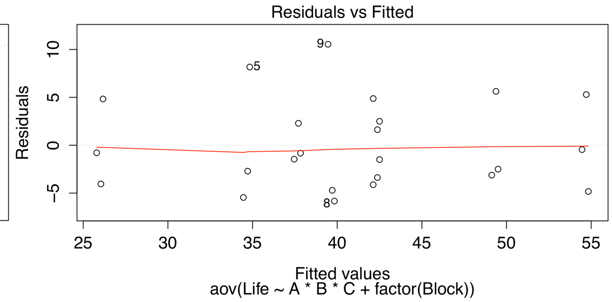
```
p4=read.table("hw5_p4.txt",header=T)
p4.lm=aov(Life~A*B*C+factor(Block),data=p4)
summary(p4.lm)
plot(p4.lm)
```

ANOVA Output:

```
              Df Sum Sq Mean Sq F value   Pr(>F)    
A              1    0.7     0.7   0.019 0.891320    
B              1  770.7   770.7  22.381 0.000322 ***
C              1  280.2   280.2   8.136 0.012789 *  
factor(Block)  2    0.6     0.3   0.008 0.991571    
A:B            1   16.7    16.7   0.484 0.497998    
A:C            1  468.2   468.2  13.596 0.002438 ** 
B:C            1   48.2    48.2   1.399 0.256623    
A:B:C          1   28.2    28.2   0.818 0.381072    
Residuals     14  482.1    34.4                     
```

The blocking does not change the results. Factors B, C and the interaction AC are still significant. The residuals and qqplot indicate nothing too unusual.





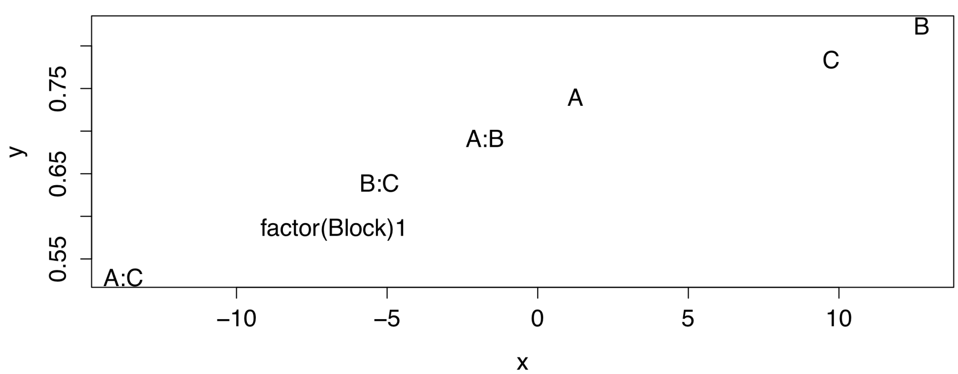
7.4. Consider the data from the first replicate of Problem 6.1. Suppose that these observations could not all be run using the same bar stock. Set up a design to run these observations in two blocks of four observations each with *ABC* confounded. Analyze the data.

Here we recode the data (in Excel) from 6.1 to match this new blocking/confounding

R code:

```
p5=read.table("hw5_p5.txt",header=T)
p5.lm=aov(Life~A*B*C+factor(Block),data=p5)
summary(p5.lm)
summary.lm(p5.lm)
p5.effects=p5.lm$coefficients*2
p5.effects=p5.effects[-1]
x=sort(p5.effects)
i=1:7
ii=(i-.5)/7
y=pnorm(ii)
plot(x,y,pch="")
text(x,y,names(x))
p5.lm2=aov(Life~A+B+C+A:C+factor(Block),data=p5)
summary(p5.lm2)
summary.lm(p5.lm2)
```

From the normal probability effects plot below, the most extreme effects appear to be B, C and AC. Thus we make a new model with these terms (and the blocking). I kept A to preserve hierarchy, but that doesn't appear to matter. Note that from this plot, all appear to lie on a line, which suggests we should not be surprised by the results further down.




```
> p5.lm2=aov(Life~A+B+C+A:C+factor(Block),data=p5)
> summary(p5.lm2)

               Df Sum Sq Mean Sq F value Pr(>F)
A               1    3.1     3.1   0.102 0.7797
B               1  325.1   325.1  10.616 0.0827 .
C               1  190.1   190.1   6.208 0.1303
factor(Block)   1   91.1    91.1   2.976 0.2267
A:C             1  378.1   378.1  12.347 0.0723 .
Residuals       2   61.3    30.6
F-statistic:  6.45 on 5 and 2 DF,  p-value: 0.1397
```

In this case the overall F statistic is not significant and none of the individual terms are significant. We cannot reject the null hypothesis that what we see is due to chance/noise.