Chapter 18-19 Homework

David Oniani

May 1, 2021

## Problem 18.6

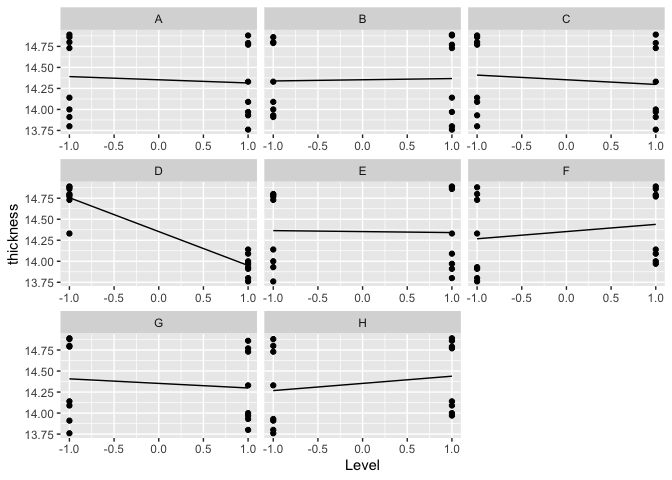
Silicon wafers for integrated circuits are grown in a device called a susceptor, and a response of interest is the thickness of the silicon. Eight factors, each at two levels, were believed to contribute: rotation method (A), wafer code (B), deposition temperature (C), deposition time (D), arsenic flow rate (E), HCl etch temperature (F), HCl flow rate (G), and nozzle position (H). A design was run with generators I = ABCD = BCEF = ACEG = BCEH.

### Main Effect plots

str (pr18.6)

## 'data.frame': 16 obs. of 9 variables:  
## $ A : num -1 -1 -1 -1 -1 -1 -1 -1 1 1 ...  
## $ B : num -1 -1 -1 -1 1 1 1 1 -1 -1 ...  
## $ C : num -1 -1 1 1 -1 -1 1 1 -1 -1 ...  
## $ D : num -1 -1 1 1 1 1 -1 -1 1 1 ...  
## $ E : num -1 1 -1 1 -1 1 -1 1 -1 1 ...  
## $ F : num -1 1 1 -1 1 -1 -1 1 -1 1 ...  
## $ G : num -1 1 1 -1 -1 1 1 -1 1 -1 ...  
## $ H : num -1 1 1 -1 1 -1 -1 1 -1 1 ...  
## $ thickness: num 14.8 14.9 14 13.9 14.1 ...

ggplot(pivot\_longer (pr18.6, cols = 1:8, names\_to = "Factor", values\_to = "Level"),   
 aes(x=Level, y=thickness)) + geom\_point() +  
 stat\_summary (fun=mean, geom="line", aes(group=Factor)) +  
 facet\_wrap(~Factor, scales = 'free\_x')



What do you notice in these plots?

The plot shows the main effects for each of the eight factors. Bigger slopes correspond to bigger main effects. Based on the plot, it appears that factors B and E have the smallest main effects (i.e., lines are approximately parallel to the x-axis [i.e., line is flat]). Factors A (negative direction) and C (negative direction) seem to have main effects bigger than factors B and E as the lines are less parallel to the x-axis than for factors B and E (considering absolute values/magnitudes of the effects as the direction for some effects is positive and for some effects, it is negative). Factors F, G, and H come next with the lines that are clearly not parallel to the x-axis. Finally, factor D has the biggest slope (i.e., take the absolute value of the slope) and has the largest main effect. That being said, the direction of this main effect is negative. As for the largest positive main effect, it seems to be factor H. Therefore, we have the following tentative main effects ordering (based on absolute values/magnitudes of the effects): E, B, C, A, G, F, H, D.

### Fit a model with all main effects and 2-factor interactions:

fit18.6 = lm (thickness ~ (A+B+C+D+E+F+G+H)^2, data=pr18.6)  
summary (fit18.6)

##   
## Call:  
## lm(formula = thickness ~ (A + B + C + D + E + F + G + H)^2, data = pr18.6)  
##   
## Residuals:  
## 1 2 3 4 5 6 7 8   
## -0.04812 0.04812 -0.04812 0.04812 0.04812 -0.04812 0.04812 -0.04812   
## 9 10 11 12 13 14 15 16   
## 0.04812 -0.04812 0.04812 -0.04812 -0.04812 0.04812 -0.04812 0.04812   
##   
## Coefficients: (22 not defined because of singularities)  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 14.353125 0.048125 298.247 0.00213 \*\*  
## A -0.038125 0.048125 -0.792 0.57348   
## B 0.014375 0.048125 0.299 0.81521   
## C -0.055625 0.048125 -1.156 0.45406   
## D -0.403125 0.048125 -8.377 0.07564 .   
## E -0.011875 0.048125 -0.247 0.84599   
## F 0.085625 0.048125 1.779 0.32598   
## G -0.054375 0.048125 -1.130 0.46123   
## H NA NA NA NA   
## A:B 0.015625 0.048125 0.325 0.80014   
## A:C -0.046875 0.048125 -0.974 0.50838   
## A:D 0.025625 0.048125 0.532 0.68851   
## A:E 0.014375 0.048125 0.299 0.81521   
## A:F 0.004375 0.048125 0.091 0.94228   
## A:G -0.010625 0.048125 -0.221 0.86167   
## A:H NA NA NA NA   
## B:C NA NA NA NA   
## B:D NA NA NA NA   
## B:E 0.029375 0.048125 0.610 0.65112   
## B:F NA NA NA NA   
## B:G NA NA NA NA   
## B:H NA NA NA NA   
## C:D NA NA NA NA   
## C:E NA NA NA NA   
## C:F NA NA NA NA   
## C:G NA NA NA NA   
## C:H NA NA NA NA   
## D:E NA NA NA NA   
## D:F NA NA NA NA   
## D:G NA NA NA NA   
## D:H NA NA NA NA   
## E:F NA NA NA NA   
## E:G NA NA NA NA   
## E:H NA NA NA NA   
## F:G NA NA NA NA   
## F:H NA NA NA NA   
## G:H NA NA NA NA   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.1925 on 1 degrees of freedom  
## Multiple R-squared: 0.9874, Adjusted R-squared: 0.8115   
## F-statistic: 5.613 on 14 and 1 DF, p-value: 0.3206

# See if H and F are aliased with the C, D, E, and F interactions.  
alias\_check\_h = lm(H ~ B \* C \* E, data=pr18.6)  
summary(alias\_check\_h)

## Warning in summary.lm(alias\_check\_h): essentially perfect fit: summary may be  
## unreliable

##   
## Call:  
## lm(formula = H ~ B \* C \* E, data = pr18.6)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -4.441e-16 -1.220e-16 0.000e+00 1.220e-16 4.441e-16   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 6.163e-33 6.650e-17 0.000e+00 1   
## B 8.875e-32 6.650e-17 0.000e+00 1   
## C 4.725e-32 6.650e-17 0.000e+00 1   
## E -7.241e-32 6.650e-17 0.000e+00 1   
## B:C -3.747e-32 6.650e-17 0.000e+00 1   
## B:E 6.598e-32 6.650e-17 0.000e+00 1   
## C:E -1.207e-32 6.650e-17 0.000e+00 1   
## B:C:E 1.000e+00 6.650e-17 1.504e+16 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2.66e-16 on 8 degrees of freedom  
## Multiple R-squared: 1, Adjusted R-squared: 1   
## F-statistic: 3.231e+31 on 7 and 8 DF, p-value: < 2.2e-16

alias\_check\_f = lm(F ~ B \* C \* E, data=pr18.6)  
summary(alias\_check\_f)

## Warning in summary.lm(alias\_check\_f): essentially perfect fit: summary may be  
## unreliable

##   
## Call:  
## lm(formula = F ~ B \* C \* E, data = pr18.6)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -4.441e-16 -1.220e-16 0.000e+00 1.220e-16 4.441e-16   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 6.163e-33 6.650e-17 0.000e+00 1   
## B 8.875e-32 6.650e-17 0.000e+00 1   
## C 4.725e-32 6.650e-17 0.000e+00 1   
## E -7.241e-32 6.650e-17 0.000e+00 1   
## B:C -3.747e-32 6.650e-17 0.000e+00 1   
## B:E 6.598e-32 6.650e-17 0.000e+00 1   
## C:E -1.207e-32 6.650e-17 0.000e+00 1   
## B:C:E 1.000e+00 6.650e-17 1.504e+16 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2.66e-16 on 8 degrees of freedom  
## Multiple R-squared: 1, Adjusted R-squared: 1   
## F-statistic: 3.231e+31 on 7 and 8 DF, p-value: < 2.2e-16

Note that the H main effect is not estimable. Also note that none of the interaction effects involving H are estimable.

Use the defining contrast, BCEH = BCEF, and defining contrast algebra to show that the main effect for H is confounded (aliased) with the main effect for F.

Given BCEH = BCEF, let us multiply both sides by BCE. We get, BCEBCEH = BCEBCEF. Applying the commutative property, we get BBCCEEH = BBCCEEF. Using the fact that the same letters cancel each other, we get H = F and we conclude that the main effect for H is confounded (aliased) with the main effect for F.

Alternatively, we could do this without using defining contrast algebras. We can take a look at summaries for the models H ~ B \* C \* E and F ~ B \* C \* E. In both cases, only one interaction B:C:E is significant with the Estimate of 1 and a very significant p-value of less than 2e-16. In both cases, we also get the R-squared value of 1. Therefore, we can conclude that the main effect for H is confounded (aliased) with the main effect for F.

### Second model

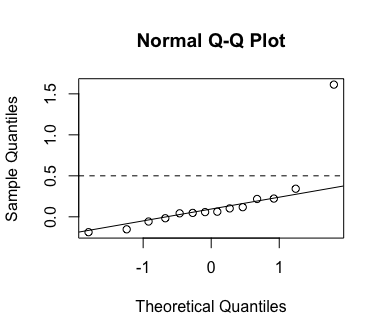
Remove H and refit. Include the A:B:E interaction, which is estimable.

fit18.6b = lm (thickness ~ (A+B+C+D+E+F+G)^2 + A:B:E, data=pr18.6)  
summary (fit18.6b)

##   
## Call:  
## lm(formula = thickness ~ (A + B + C + D + E + F + G)^2 + A:B:E,   
## data = pr18.6)  
##   
## Residuals:  
## ALL 16 residuals are 0: no residual degrees of freedom!  
##   
## Coefficients: (14 not defined because of singularities)  
## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) 14.353125 NA NA NA  
## A -0.038125 NA NA NA  
## B 0.014375 NA NA NA  
## C -0.055625 NA NA NA  
## D -0.403125 NA NA NA  
## E -0.011875 NA NA NA  
## F 0.085625 NA NA NA  
## G -0.054375 NA NA NA  
## A:B 0.015625 NA NA NA  
## A:C -0.046875 NA NA NA  
## A:D 0.025625 NA NA NA  
## A:E 0.014375 NA NA NA  
## A:F 0.004375 NA NA NA  
## A:G -0.010625 NA NA NA  
## B:C NA NA NA NA  
## B:D NA NA NA NA  
## B:E 0.029375 NA NA NA  
## B:F NA NA NA NA  
## B:G NA NA NA NA  
## C:D NA NA NA NA  
## C:E NA NA NA NA  
## C:F NA NA NA NA  
## C:G NA NA NA NA  
## D:E NA NA NA NA  
## D:F NA NA NA NA  
## D:G NA NA NA NA  
## E:F NA NA NA NA  
## E:G NA NA NA NA  
## F:G NA NA NA NA  
## A:B:E 0.048125 NA NA NA  
##   
## Residual standard error: NaN on 0 degrees of freedom  
## Multiple R-squared: 1, Adjusted R-squared: NaN   
## F-statistic: NaN on 15 and 0 DF, p-value: NA

Make a normal Q-Q plot of the effects.

effs18.6b = fit18.6b$effects [2:fit18.6$rank]  
qqnorm (effs18.6b)  
qqline (effs18.6b)  
abline (h=0.5, lty=2)



effs18.6b [abs (effs18.6b) > 0.5]

## D   
## 1.6125

Which effects are significant?

There is only one effect that is significant. It is the effect D and its value is 1.6125.

### What effects are confounded with factor D?

Use defining contrast algebra for each of these relations to determine the effects (main or interaction effects) that are confounded with the main effect of D (hint - multiply both sides by ABC in each expression):

* ABCD = BCEF
* ABCD = ACEG
* ABCD = BCEH

Given ABCD = BCEF, let us multiply both sides by ABC. After regrouping, we get AABBCCD = ABBCCEF which gives us D = AEF. Thus, we got that the three-way interaction AEF is confounded with the main effect of D.

Given ABCD = ACEG, let us multiply both sides by ABC. After regrouping, we get AABBCCD = AABCCEG which gives us D = BEG. Thus, we got that the three-way interaction BEG is confounded with the main effect of D.

Given ABCD = BCEH, let us multiply both sides by ABC. After regrouping, we get AABBCCD = ABBCCEH which gives us D = AEH. Thus, we got that the three-way interaction AEH is confounded with the main effect of D.

## Problem 19.5

Curing time and temperature affect the shear strength of an adhesive that bonds galvanized steel bars. The following experiment was repeated on 2 separate days. Twenty-four pieces of steel are obtained by random sampling from warehouse stock. These are grouped into twelve pairs; the twelve pairs are glued and then cured with one of nine curing treatments assigned at random. The treatments are the three by three factorial combinations of temperature (375, 400, and 450 degrees F, coded -1, 0, 2) and time (30, 35, or 40 seconds, coded -1, 0, 1). Four pairs were assigned to the center point, and one pair to all other conditions. The response is shear strength (in psi, data from Khuri 1992):

|  |  |  |  |
| --- | --- | --- | --- |
| Temp. | Time | Day 1 | Day 2 |
| -1 | -1 | 1226 | 1213 |
| 0 | -1 | 1898 | 1961 |
| 2 | -1 | 2142 | 2184 |
| -1 | 0 | 1472 | 1606 |
| 0 | 0 | 2010 | 2450 |
| 0 | 0 | 1882 | 2355 |
| 0 | 0 | 1915 | 2420 |
| 0 | 0 | 2106 | 2240 |
| 2 | 0 | 2352 | 2298 |
| -1 | 1 | 1491 | 2298 |
| 0 | 1 | 2078 | 2531 |
| 2 | 1 | 2531 | 2609 |
|  |  |  |  |

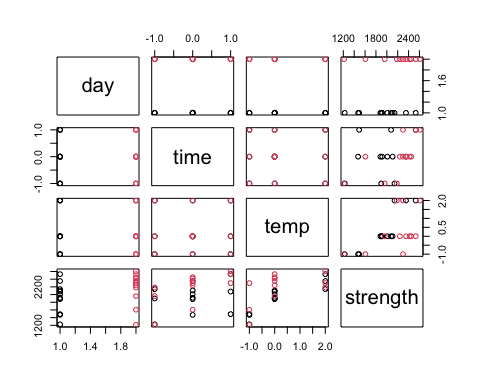
str (pr19.5)

## 'data.frame': 24 obs. of 4 variables:  
## $ day : int 1 1 1 1 1 1 1 1 1 1 ...  
## $ time : int -1 0 2 -1 0 0 0 0 2 -1 ...  
## $ temp : int -1 -1 -1 0 0 0 0 0 0 1 ...  
## $ strength: int 1226 1898 2142 1472 2010 1882 1915 2106 2352 1491 ...

tmp\_time = pr19.5$time  
tmp\_temp = pr19.5$temp  
pr19.5$temp = tmp\_time  
pr19.5$time = tmp\_temp  
str(pr19.5)

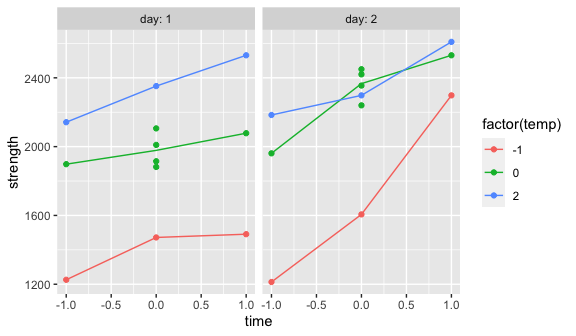
## 'data.frame': 24 obs. of 4 variables:  
## $ day : int 1 1 1 1 1 1 1 1 1 1 ...  
## $ time : int -1 -1 -1 0 0 0 0 0 0 1 ...  
## $ temp : int -1 0 2 -1 0 0 0 0 2 -1 ...  
## $ strength: int 1226 1898 2142 1472 2010 1882 1915 2106 2352 1491 ...

plot (pr19.5, col=pr19.5$day)



Plot the data:

ggplot (pr19.5, aes (x=time, y=strength, color=factor(temp))) + geom\_point() +  
 facet\_wrap (vars(day), labeller="label\_both") +  
 stat\_summary (fun=mean, geom="line", aes(group=temp))



(Note: No interpretation needed for the plot above.)

### Response surface model:

Fit a second-order response surface model and include day as a block effect.

fit19.5 = rsm (strength ~ day + FO(time, temp) + SO(time, temp), data=pr19.5)  
summary (fit19.5)

##   
## Call:  
## rsm(formula = strength ~ day + FO(time, temp) + SO(time, temp),   
## data = pr19.5)  
##   
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1775.333 112.645 15.7604 1.409e-11 \*\*\*  
## day 255.167 66.434 3.8409 0.0013093 \*\*   
## time 255.214 48.625 5.2486 6.535e-05 \*\*\*  
## temp 432.903 52.667 8.2197 2.519e-07 \*\*\*  
## time:temp -37.143 37.665 -0.9861 0.3378908   
## time^2 -12.750 70.464 -0.1809 0.8585521   
## temp^2 -165.681 36.092 -4.5906 0.0002602 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Multiple R-squared: 0.8799, Adjusted R-squared: 0.8375   
## F-statistic: 20.76 on 6 and 17 DF, p-value: 5.967e-07  
##   
## Analysis of Variance Table  
##   
## Response: strength  
## Df Sum Sq Mean Sq F value Pr(>F)  
## day 1 390660 390660 14.7524 0.0013093  
## FO(time, temp) 2 2240209 1120104 42.2982 2.514e-07  
## TWI(time, temp) 1 25752 25752 0.9725 0.3378908  
## PQ(time, temp) 2 641107 320554 12.1050 0.0005386  
## Residuals 17 450179 26481   
## Lack of fit 3 38532 12844 0.4368 0.7301256  
## Pure error 14 411647 29403   
##   
## Stationary point of response surface:  
## time temp   
## 9.6870884 0.2205951   
##   
## Eigenanalysis:  
## eigen() decomposition  
## $values  
## [1] -10.52705 -167.90350  
##   
## $vectors  
## [,1] [,2]  
## time -0.9929124 0.1188488  
## temp 0.1188488 0.9929124

Which effects are significant?

The main effects day, time, and temp are significant with the p-values of approximately 0.001, 6.535e-05, and 2.514e-07 respectively. The intercept is significant as well, with the p-value of approximately 1.409e-11. Finally, temp^2 is also significant with the p-value of approximately 0.0003.

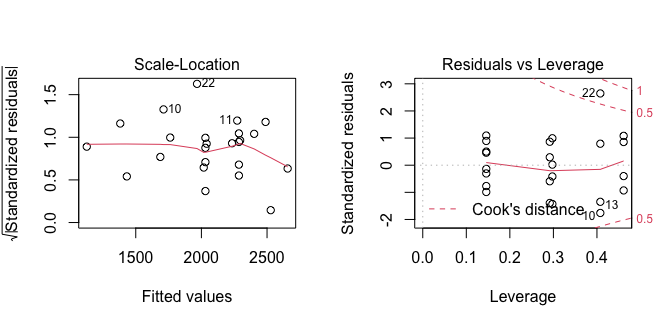
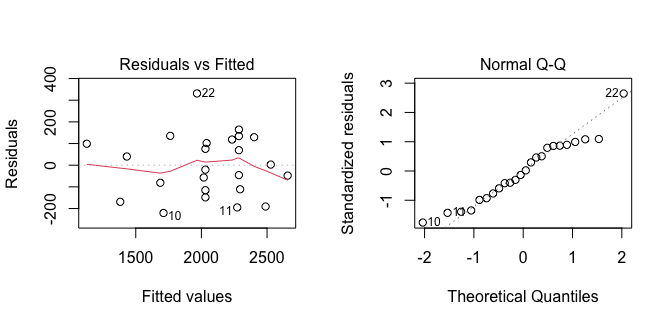
### Test for lack of fit

Based on the output above, is there significant lack of fit? Include the p-value for the lack of fit test.

There is not a significant lack of fit since the p-value is approximately 0.730, which is way above the 0.05 cutoff.

### Check the residuals:

par (mfrow=c(1,2))  
plot (fit19.5)



Interpret the residual plots:

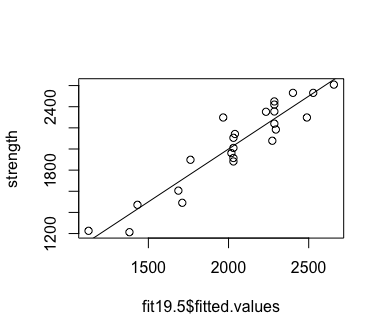
Residuals vs Fitted plot shows that the variance is equal since the points are scattered randomly above and below the reference line at 0. That being said, the red line does not follow the dotted line. However, this does not change the fact that we have equal variance.

Normal Q-Q plot shows that the distribution is approximately normal since most of the points follow the dotted line. There is an evidence of some non-normality at both tails of the Normal Q-Q (slightly heavier right tail), but, overall, the plot does not show non-normal patterns.

Scale-Location plot shows equal variance and, to some degree, further corroborates the hypothesis from Residuals vs Fitted plot.

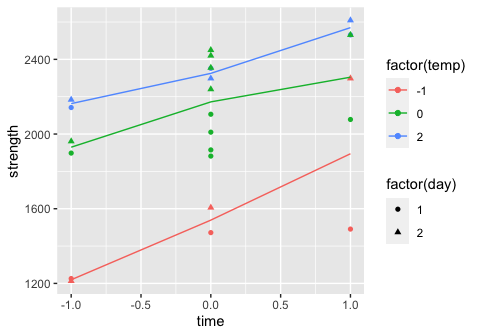
Residuals vs Leverage does not show any noticeable outliers. While we do see the Cook’s contours, none of the points are beyond them.

plot (strength ~ fit19.5$fitted.values, data=pr19.5)  
abline (0, 1)



### Analysis conclusion

ggplot (pr19.5, aes (x=time, y=strength, color=factor(temp), shape=factor(day))) +   
 geom\_point() +  
 stat\_summary (fun=mean, geom="line", aes(group=temp))



Given these modeling results, and the data plot just above, what values of time and temperature would you choose to maximize strength. Stay within the range of values tested for those factors. You may state a specific value for time and for temperature or a range of values. Use the original scales for time and temperature, as stated at the beginning of the problem.

The highest value for strength is approximately 2550. This value is obtained when time is 1 (coded) and temperature is 2 (coded). Recall, that time time was coded -1, 0, 1 for 30, 35, and 40 seconds respectively and the temperature was coded -1, 0, and 2 for 375, 400, and 450 degrees F. Hence, in order to maximize the strength, we would choose 40 seconds and 450 degrees F.