

Homework 7

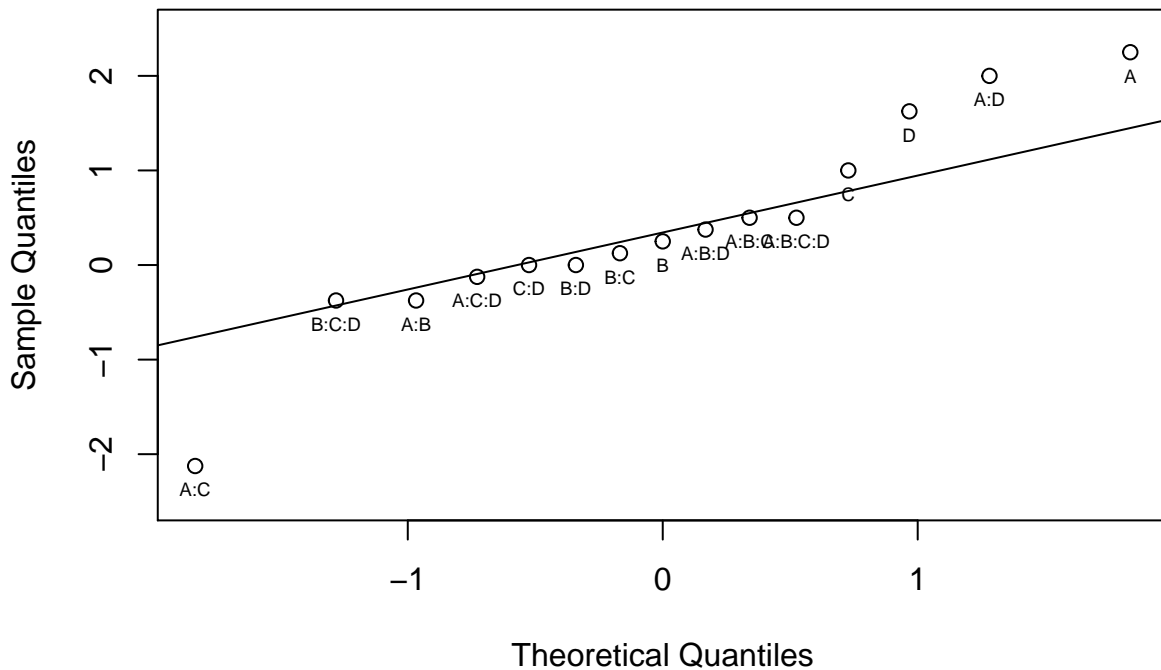
Cody Frisby

March 28, 2016

1)

A) Significant Factors

Normal Plot of the Effects



This plot shows that A, C, D, AC, and AD are effects of interest to us. Let's fit this model and look at the residuals from it.

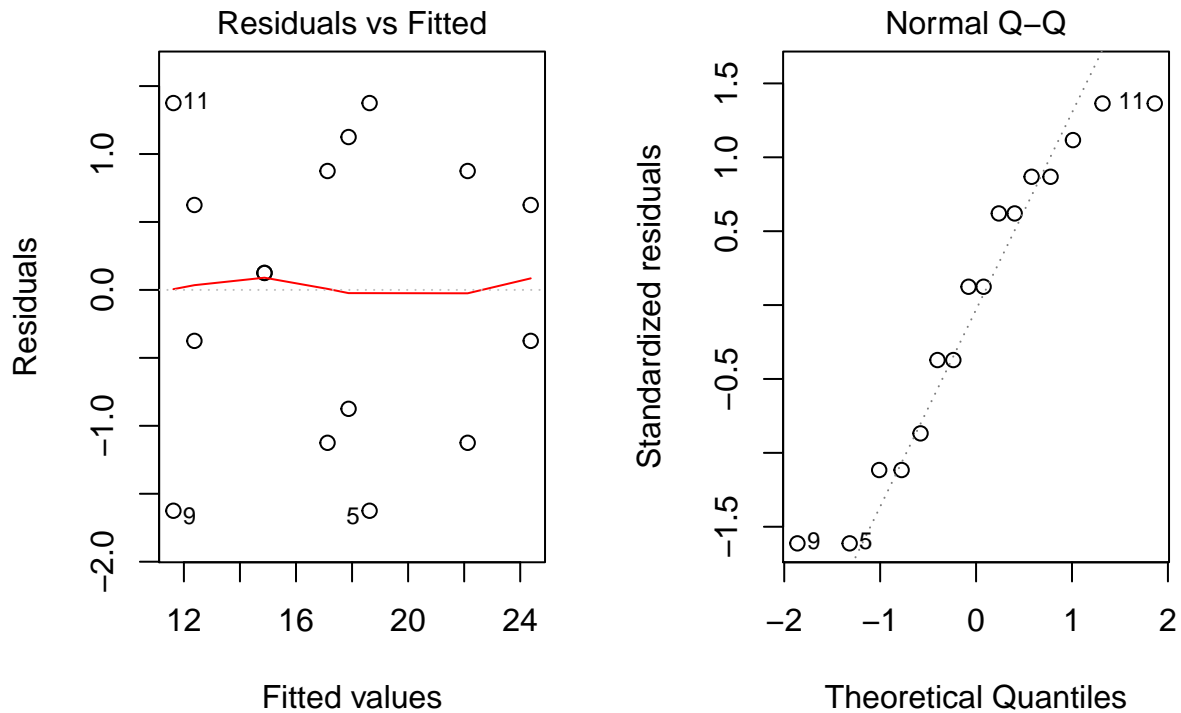
B) Model from effects from part A Here, I display an ANOVA table, which will have an estimate of the error since we are fitting a reduced model

$$yield = A + C + D + AC + AD$$

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
A	1	81.00	81.000	49.846154	0.0000346
C	1	16.00	16.000	9.846154	0.0105485
D	1	42.25	42.250	26.000000	0.0004647
A:C	1	72.25	72.250	44.461539	0.0000558
A:D	1	64.00	64.000	39.384615	0.0000919
Residuals	10	16.25	1.625	NA	NA

A, C, D, AC, & AD all have a significant effect on yield.

C) Residual Analysis Here, we take a look at the residuals from our model and see if the model assumptions hold up.



It looks like there are no major concerns with the residuals. Our model assumptions hold up.

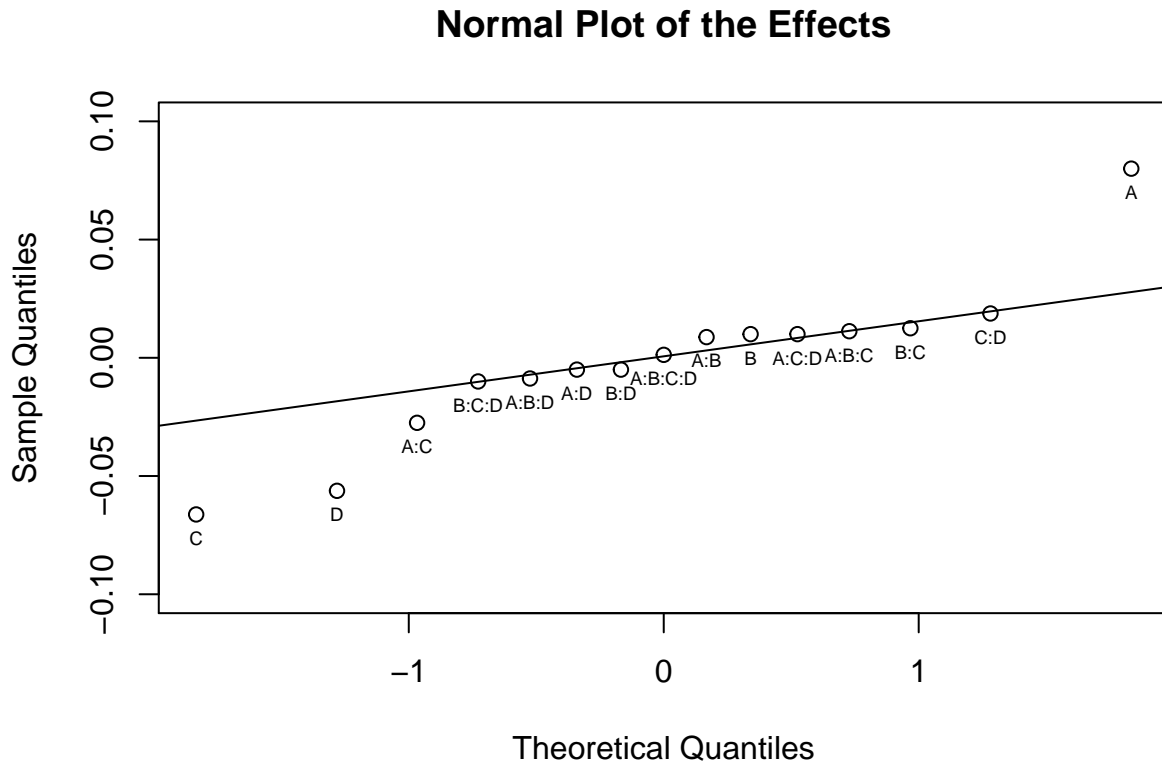
```
ix <- 2:5
prob1[ix] <- lapply(prob1[ix], as.factor)
#plot.design(prob1)
```

D) 2^3 Design with Replicates? Yes, if we remove B as a factor, this design becomes a 2^3 design with two replicates per level/factor combination.

E) Maximize Yield To maximize yield I would suggest running with A “high”, C “low”, and D “high”. Our model predicts that this combination will give us the most desirable result.

2) Semiconductor Experiment

A) Which Factors? The factors that are significant can be found by fitting a full factorial model and then plotting the effects and visually identifying the “outliers”.

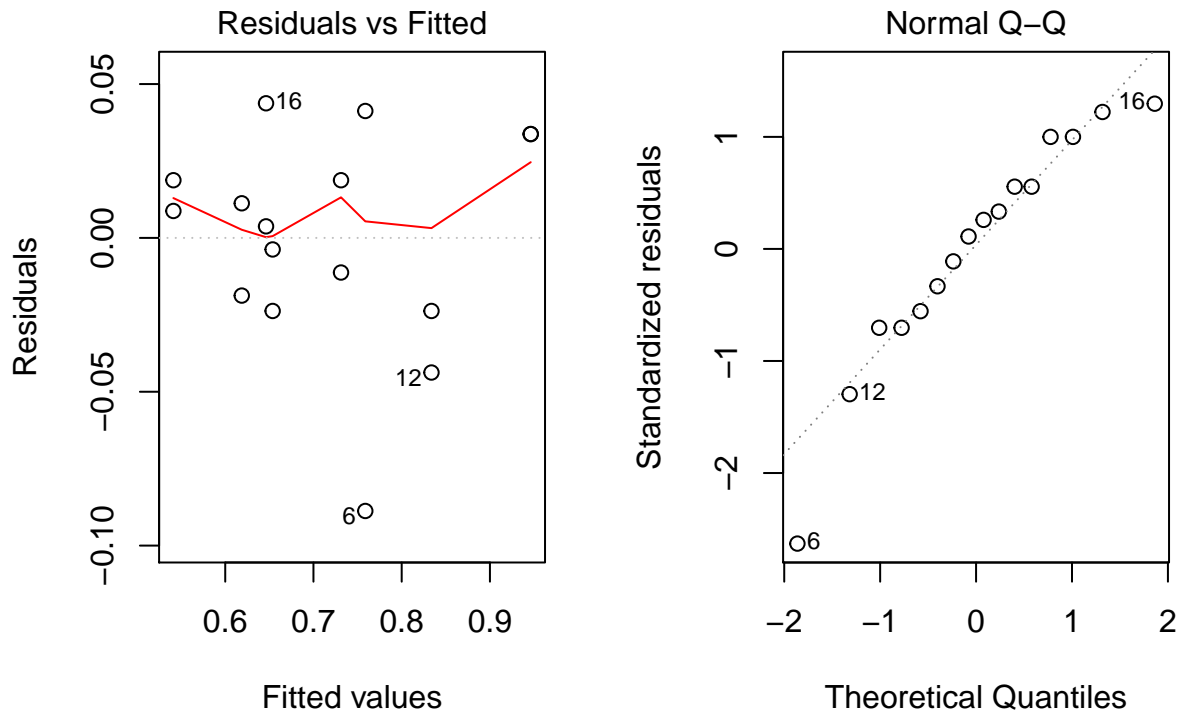


Here, A, C, D, & AC appear to be of interest to us.

B) Residuals ANOVA table from our fitted model is

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
A	1	0.102400	0.1024000	61.805213	0.0000077
C	1	0.070225	0.0702250	42.385460	0.0000437
D	1	0.050625	0.0506250	30.555556	0.0001786
A:C	1	0.012100	0.0121000	7.303155	0.0205671
Residuals	11	0.018225	0.0016568	NA	NA

And here we plot the residuals.



Normality assumption appears to be OK, but the equality of variances is of a little concern. Although, sample size is small, and the scatter does appear to be random “enough”. Assumptions are OK.

R code:

```

prob1 <- read.csv("~/Documents/STAT4100/data/hw6-1.csv")
# fit the fully model
fit.all <- aov(yield ~ .^4, data = prob1)
# plot the effects
tmp <- qqnorm(coef(fit.all)[-1], ylim = c(-2.5, 2.5),
              main = "Normal Plot of the Effects")
qqline(coef(fit.all)[-1])
text(tmp$x, tmp$y, names(coef(fit.all)[-1]), pos=1, cex=0.6)
fit <- aov(yield ~ A*C+A*D, data = prob1) # this is the model
anova(fit)
ix <- 2:5
prob1[ix] <- lapply(prob1[ix], as.factor)
plot.design(prob1)

```