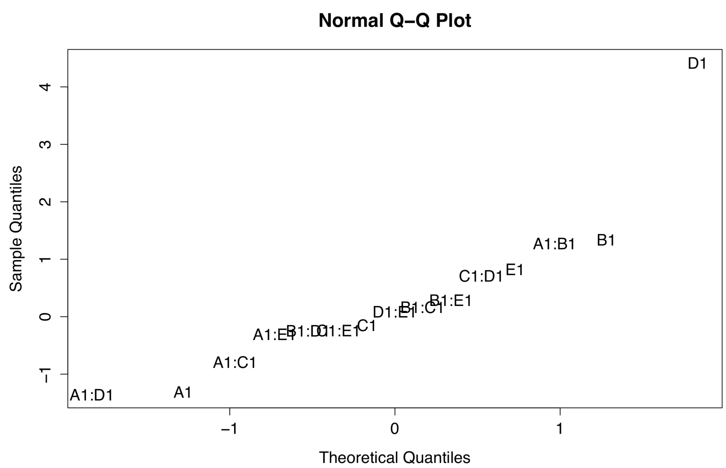
STAT 571B - Homework 6
Brian Hallmark

8.6.

(a) Prepare a normal probability plot of the effects. Which effects seem active?

R code:

```
p1<-read.table("hw6p1.txt",header=T)
p1$A=factor(p1$A);p1$B=factor(p1$B);p1$C=factor(p1$C)
p1$D=factor(p1$D);p1$E=factor(p1$E)
p1.lm1=lm(Color~(A+B+C+D+E)^2,data=p1)
anova(p1.lm1)
p1.lm1$effects
p1.effects=p1.lm1$effects[-1]
p1.qq=qqnorm(p1.effects,type="n")
text(p1.qq$x,p1.qq$y,names(p1.effects))
```



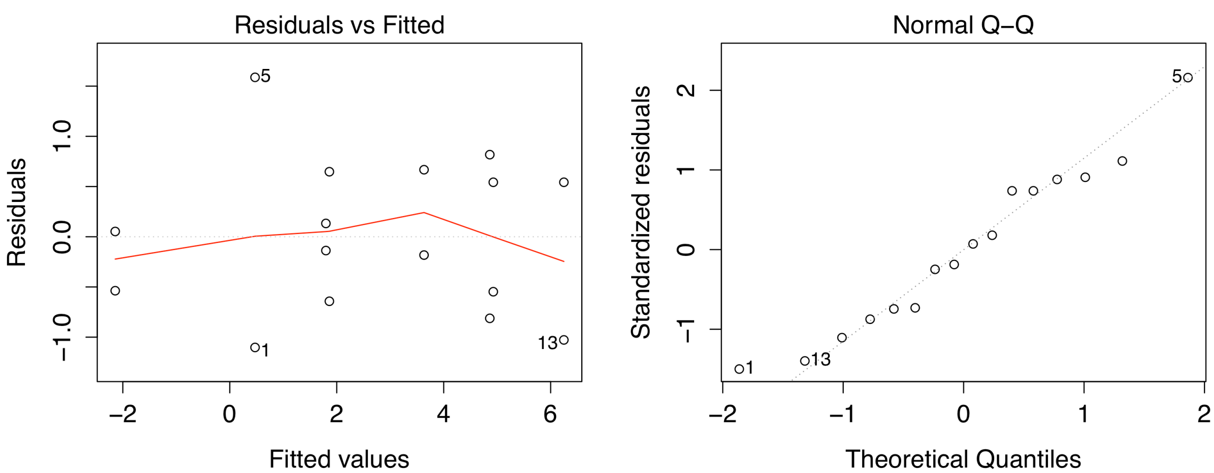
Compiled Output from R:

	df	Sum Sq	Mean Sq	% contribution	Effect/2
A	1	6.86	6.86	**5.982**	**-1.31**
B	1	7.18	7.18	**6.261**	**1.34**
C	1	0.09	0.09	0.078	-0.1475
D	1	78.15	78.15	**68.146**	**4.42**
E	1	2.74	2.74	2.389	0.8275
A:B	1	6.5	6.5	**5.668**	**1.275**
A:C	1	2.48	2.48	2.163	-0.7875
A:D	1	7.34	7.34	**6.400**	**-1.355**
A:E	1	0.37	0.37	0.323	-0.3025
B:C	1	0.11	0.11	0.096	0.1675
B:D	1	0.24	0.24	0.209	-0.245
B:E	1	0.33	0.33	0.288	0.2875
C:D	1	2.03	2.03	1.770	0.7125
C:E	1	0.23	0.23	0.201	-0.24
D:E	1	0.03	0.03	0.026	0.0875

The normal probability plot, effect sizes and % contribution indicate that factors A,B,D and interactions AB and AD are significant. Thus we can run a new model with only these factors in part b. NOTE: I forgot to multiple the parameter estimates by 2 to get the correct effect sizes but the plot is the same.

(b) Calculate the residuals. Construct a normal probability plot of the residuals and plot the residuals versus the fitted values. Comment on the plots.

R code:
p1.lm2=lm(Color~A+B+D+A:B+A:D,data=p1)
summary.aov(p1.lm2)
par(mfrow=c(1,2))
plot(p1.lm2)



These plots appear satisfactory.

(c) If any factors are negligible, collapse the 2^{5-1} design into a full factorial in the active factors. Comment on the resulting design, and interpret the results.

We can collapse the design into a 2^3 full factorial experiment with 2 replicates.

```
p1d2=read.table("hw6p1d2.txt",header=T)
p1d2.lm1=aov(Color~A*B*D,data=p1d2)
summary(p1d2.lm1)
```

ANOVA table:

```
            Df Sum Sq Mean Sq F value   Pr(>F)    
A            1   6.86    6.86   6.716   0.0320 *  
B            1   7.18    7.18   7.027   0.0292 *  
D            1  78.15   78.15  76.456 2.29e-05 ***
A:B          1   6.50    6.50   6.362   0.0357 *  
A:D          1   7.34    7.34   7.185   0.0279 *  
B:D          1   0.24    0.24   0.235   0.6409    
A:B:D        1   0.23    0.23   0.225   0.6476    
Residuals    8   8.18    1.02                     
```

Here the factors A, B, D and interactions AB and AD are significant.

8.7.

(a) Write out the alias structure for this design. What is the resolution of this design?

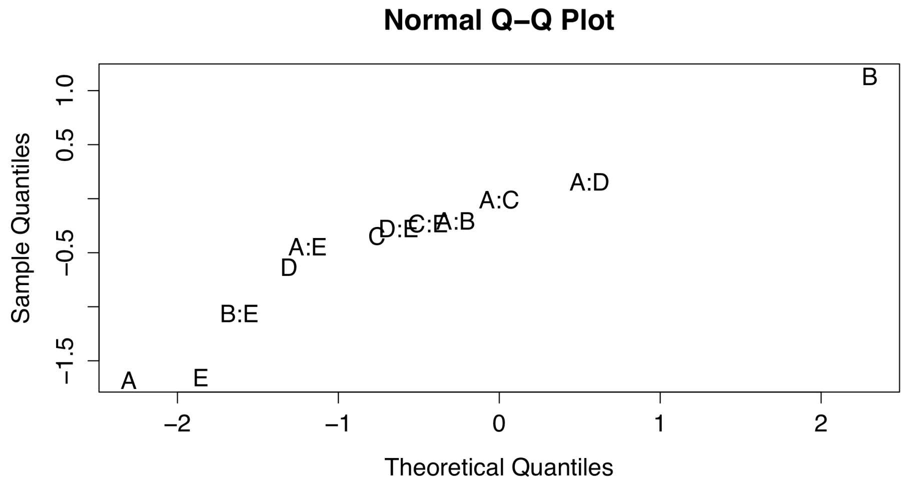
I=ABCD, resolution IV

A	*	ABCD	=	BCD
B	*	ABCD	=	ACD
C	*	ABCD	=	ABD
D	*	ABCD	=	ABC
E	*	ABCD	=	ABCDE
AB	*	ABCD	=	CD
AC	*	ABCD	=	BD
AD	*	ABCD	=	BC
AE	*	ABCD	=	BCDE
BC	*	ABCD	=	AD
BD	*	ABCD	=	AC
BE	*	ABCD	=	ACDE
CD	*	ABCD	=	AB
CE	*	ABCD	=	ABDE
DE	*	ABCD	=	ABCE

(b) Analyze the data. What factors influence the mean free height?

R code:

```
p2=read.table("hw6p2.txt",header=T)
p2.lm2=aov(I~A*B*C*D*E,data=p2)
summary.lm(p2.lm2)
summary(p2.lm2)
plot(p2.lm2)
p2.effects=p2.lm1$effects[-1]
p2.effects=p2.effects*2
p2.qq=qqnorm(p2.effects,type="n")
text(p2.qq$x,p2.qq$y,names(p2.effects))
```



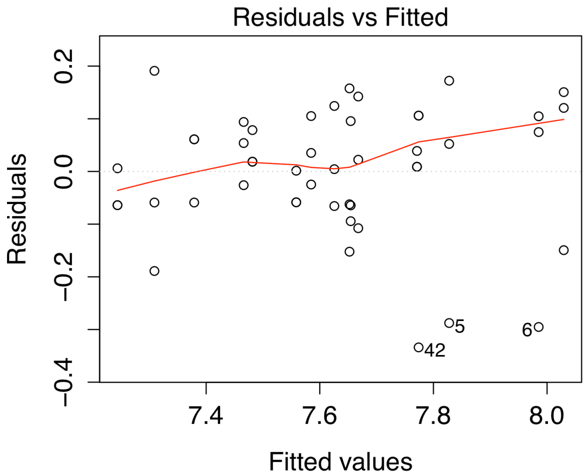
ANOVA Table:

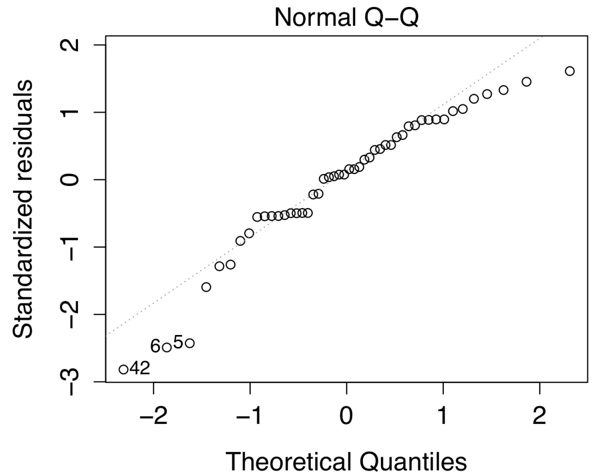
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
A	1	0.7033	0.7033	35.888	1.12e-06	***
B	1	0.3218	0.3218	16.420	0.000302	***
C	1	0.0295	0.0295	1.506	0.228774	
D	1	0.0999	0.0999	5.099	0.030893	*
E	1	0.6840	0.6840	34.906	1.42e-06	***
A:B	1	0.0105	0.0105	0.536	0.469451	
A:C	1	0.0000	0.0000	0.001	0.975515	
B:C	1	0.0063	0.0063	0.322	0.574603	
A:E	1	0.0488	0.0488	2.489	0.124500	
B:E	1	0.2806	0.2806	14.319	0.000640	***
C:E	1	0.0130	0.0130	0.664	0.421343	
D:E	1	0.0188	0.0188	0.959	0.334662	
A:B:E	1	0.0001	0.0001	0.003	0.959204	
A:C:E	1	0.0046	0.0046	0.235	0.631251	
B:C:E	1	0.0426	0.0426	2.174	0.150128	
Residuals	32	0.6271	0.0196			

Together, these results (normal plot, ANOVA results, effect sizes) indicate that factors A,B,E and interaction BE are significant. Although D appears significant in the ANOVA table, the effect size is small and it is not an outlier on the normal probability plot.

(d) Analyze the residuals from this experiment, and comment on your findings.

The residuals from the first model have some issues:



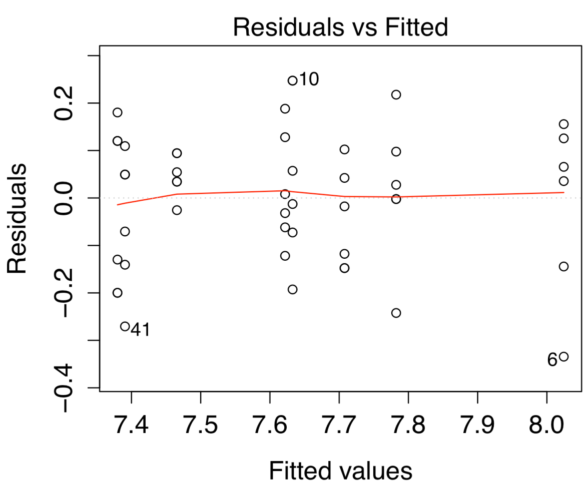


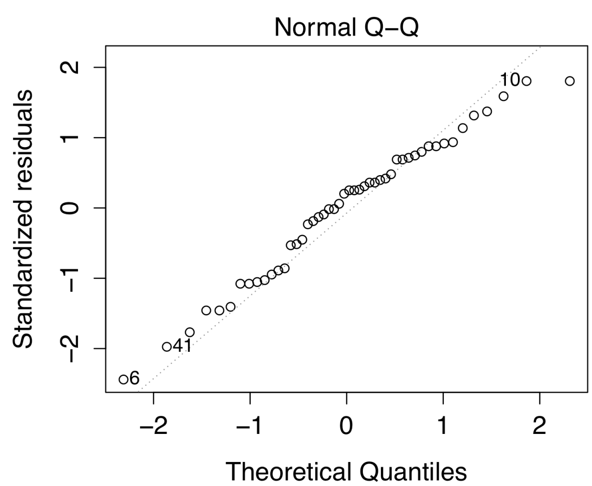
If we run the new model with the significant factors we get

```
p2.lm2=aov(I~A+B+E+B:E,data=p2)
summary(p2.lm2)
plot(p2.lm2)
```

ANOVA table:

```
            Df Sum Sq Mean Sq F value   Pr(>F)
A            1 0.7033  0.7033   33.56 7.31e-07 ***
B            1 0.3218  0.3218   15.35 0.000315 ***
E            1 0.6840  0.6840   32.64 9.54e-07 ***
B:E          1 0.2806  0.2806   13.39 0.000687 ***
Residuals   43 0.9011  0.0210
```





(e) Is this the best possible design for five factors in 16 runs? Specifically, can you find a fractional design for five factors in 16 runs with a higher resolution than this one?

No, this is not the best design since if the generator was set to I=ABCDE it would have resolution V.

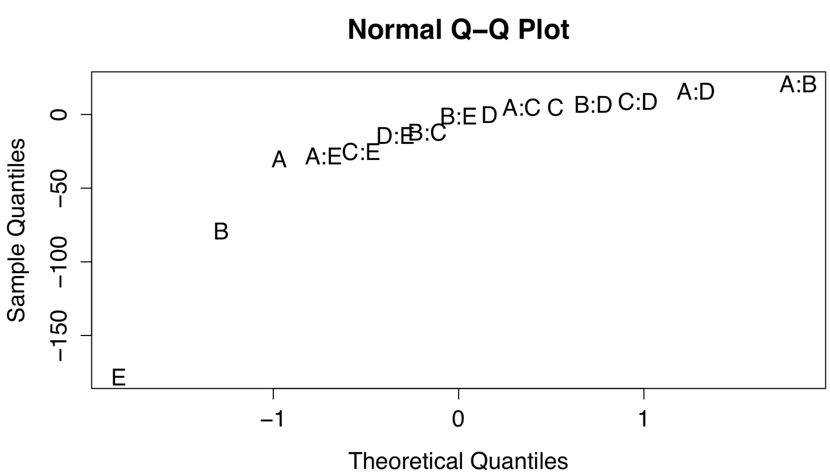
8.28.

(a) What type of design has been used? Identify the defining relation and the alias relationships.

I=-ABCDE

A	*	-ABCDE	=	-BCDE
B	*	-ABCDE	=	-ACDE
C	*	-ABCDE	=	-ABDE
D	*	-ABCDE	=	-ABCE
E	*	-ABCDE	=	-ABCD
AB	*	-ABCDE	=	-CDE
AC	*	-ABCDE	=	-BDE
AD	*	-ABCDE	=	-BCE
AE	*	-ABCDE	=	-BCD
BC	*	-ABCDE	=	-ADE
BD	*	-ABCDE	=	-ACE
BE	*	-ABCDE	=	-ACD
CD	*	-ABCDE	=	-ABE
CE	*	-ABCDE	=	-ABD
DE	*	-ABCDE	=	-ABC

(b) Estimate the factor effects and use a normal probability plot to tentatively identify the important factors.



This plot suggests that factors A, B and E are significant and that perhaps some interactions are significant.

A summary of the results calculated in R and Excel are:

	Df	Sum Sq	Mean Sq	% SS	Effect
A	1	225	225	**2.17**	**-30**
B	1	1560	1560	**15.06**	**-79**
C	1	6	6	0.06	5
D	1	0	0	0.00	0
E	1	7921	7921	**76.45**	**-178**
A:B	1	110	110	**1.06**	**21**
A:C	1	6	6	0.06	5
B:C	1	36	36	0.35	-12
A:D	1	64	64	0.62	16
B:D	1	12	12	0.12	7
C:D	1	20	20	0.19	9
A:E	1	196	196	**1.89**	**-28**
B:E	1	0	0	0.00	-1
C:E	1	156	156	**1.51**	**-25**
D:E	1	49	49	0.47	-14

This also suggests that A, B and E are the main effects and that some interactions could be important

(c) Perform an appropriate statistical analysis to test the hypothesis that the factors identified in part above have a significant effect on the yield of peanut oil.

```
p3.lm2=aov(y~A*B*E,data=p3)
summary(p3.lm2)
```

ANOVA Table:

```
            Df Sum Sq Mean Sq F value   Pr(>F)
A            1    225     225   5.455 0.047750 *
B            1   1560    1560  37.824 0.000274 ***
E            1   7921    7921 192.024 7.11e-07 ***
A:B          1    110     110   2.673 0.140721
A:E          1    196     196   4.752 0.060882 .
B:E          1      0       0   0.006 0.939859
A:B:E        1     20      20   0.491 0.503387
Residuals    8    330      41
```

These results indicate that factors A, B and E are important but the interactions are not.

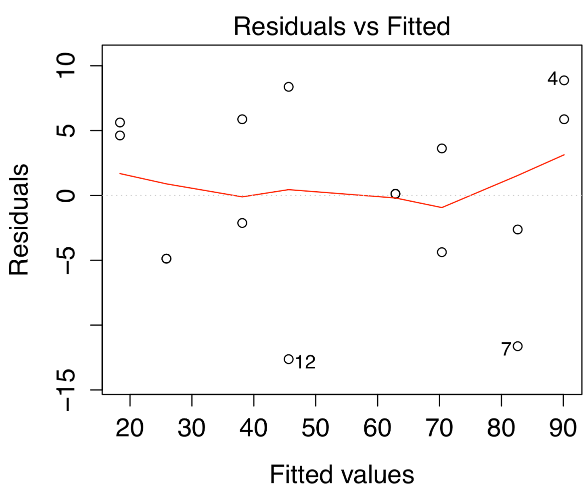
(d) Fit a model that could be used to predict peanut oil yield in terms of the factors that you have identified as important.

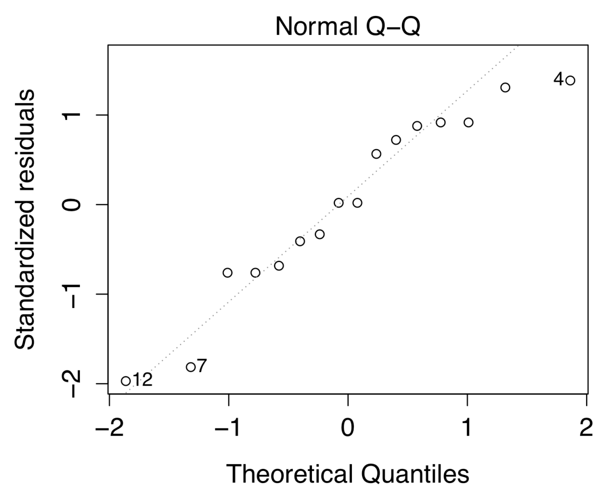
```
p3.lm3=aov(y~A+B+E,data=p3)
summary.lm(p3.lm3)
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   54.250      1.849  29.333 1.54e-12 ***
A              3.750      1.849   2.028 0.065404 .
B              9.875      1.849   5.339 0.000177 ***
E            -22.250      1.849 -12.030 4.70e-08 ***
---
```

The coefficients in this regression model allows us to estimate peanut oil as a function of A, B and E

(e) Analyze the residuals from this experiment and comment on model adequacy.





These plots appear okay. The model appears adequate but further testing is probably prudent to be sure.