

Montgomery 7.5

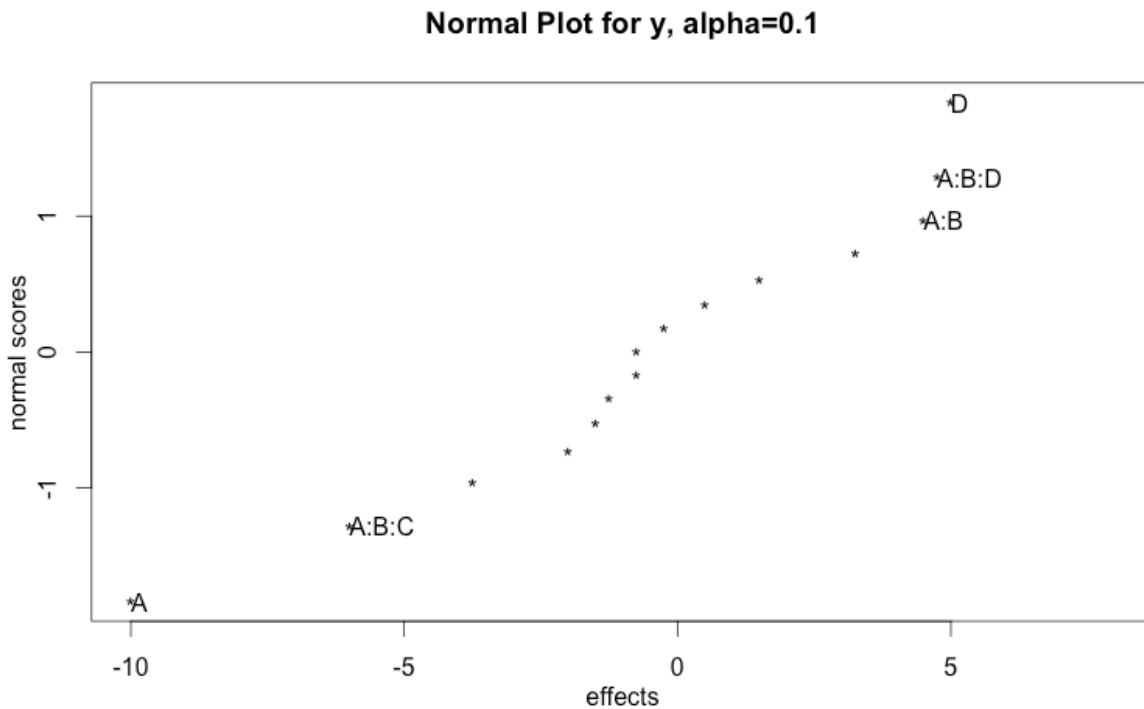
Consider the data from the first replicate of Problem 6.7. Construct a design with two blocks of eight observations each with ABCD confounded. Analyze the data.

```
> levels = c(-1, 1)
> y = c(90, 74, 81, 83, 77, 81, 88, 73, 98, 72, 87, 85, 99, 79, 87, 80)
> df = expand.grid(A=levels, B=levels, C=levels, D=levels)
> df$block = df$A * df$B * df$C * df$D
> n = 1; A = df$A; B = df$B; C = df$C; D = df$D;
> AB = A*B; AC = A*C; AD = A*D; BC = B*C;
> BD = B*D; CD = C*D;
> ABC = A*B*C; ABD = A*B*D; BCD = B*C*D; ABCD = ABC*D
> effects = data.frame(e=t(y %*%
cbind(A,B,C,D,AB,AC,AD,BC,BD,CD,ABC,ABD,BCD,ABCD)/(4*n)))
> effects
```

	e
A	-20.0
B	-1.5
C	-1.5
D	10.0
AB	9.0
AC	1.0
AD	-7.5
BC	-2.5
BD	-3.0
CD	3.0
ABC	-12.0
ABD	9.5
BCD	-4.0
ABCD	6.5

From the above, we can see that factors A, D, AB, ABC, and ABD are the most significant. A plot can also help visualize this, though I found for every problem in this homework, I needed to run the "DanielPlot" with $\alpha = 0.01$ and not the default 0.05.

```
> df[] = lapply(df, factor)
> df$y = y
> lmod = aov(y ~ block + A * B * C * D, df)
> library(FrF2)
> DanielPlot(lmod, alpha=0.1)
```

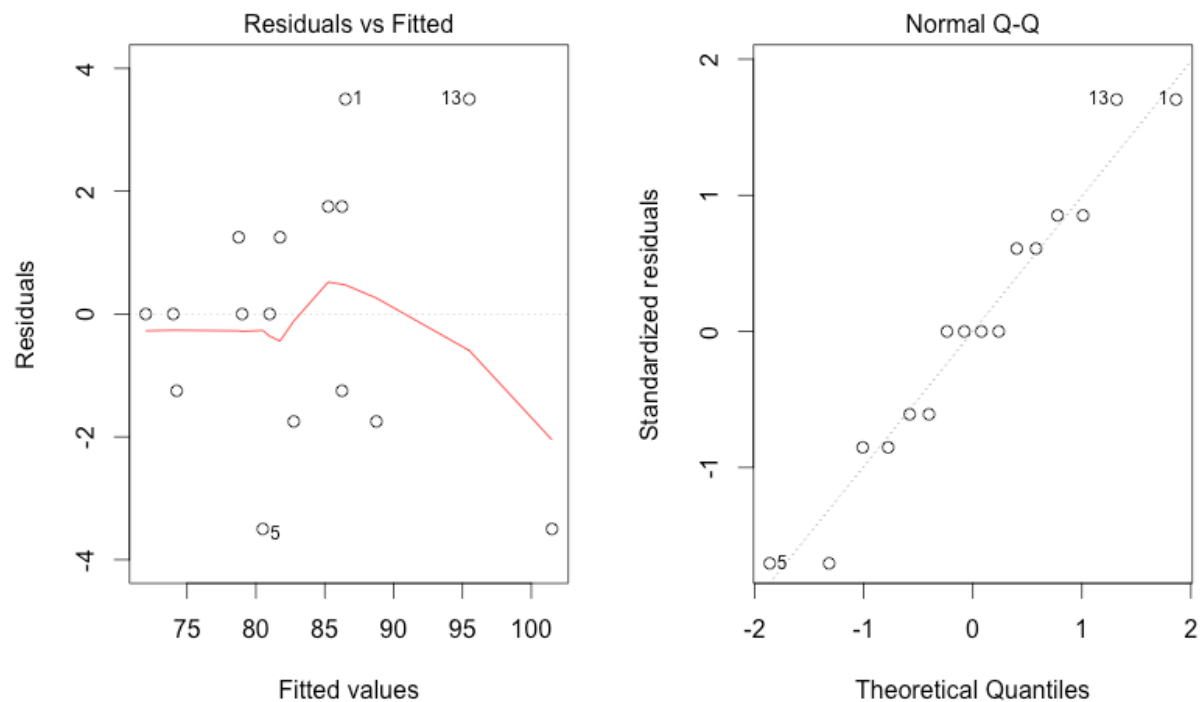


Reducing the analysis to the most significant factors, I get the following:

```
> amod = aov(y ~ A + D + A:B + A:B:C + A:B:D, df)
> options(show.signif.stars=FALSE)
> summary(amod)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
A	1	400.0	400.0	23.704	0.00823
D	1	100.0	100.0	5.926	0.07165
A:B	2	83.3	41.6	2.467	0.20049
A:B:C	4	153.5	38.4	2.274	0.22288
A:D:B	3	155.5	51.8	3.072	0.15333
Residuals	4	67.5	16.9		

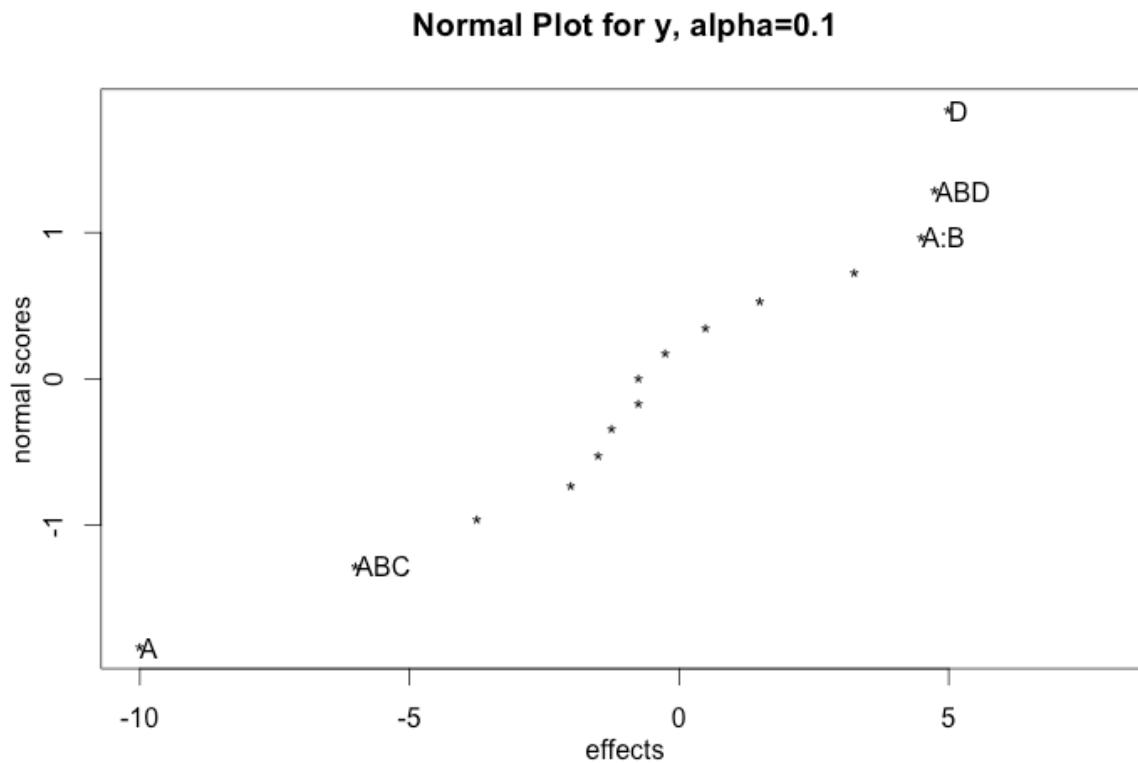
The p-values A falls below $\alpha=0.05$, therefore we reject the null hypothesis and state that A does significantly affect the outcome. Looking at plots of residuals vs fitted and the normal QQ, the fit of the data looks acceptably distributed.



Montgomery 7.6

Repeat Problem 7.5 assuming that four blocks are required. Confound ABD and ABC (and consequently CD) with blocks.

```
> library(FrF2)
> levels = c(-1, 1)
> df = expand.grid(A=levels, B=levels, C=levels, D=levels)
> df$ABC = df$A * df$B * df$C
> df$ABD = df$A * df$B * df$D
> df[] = lapply(df, factor)
> df$y = c(90, 74, 81, 83, 77, 81, 88, 73, 98, 72, 87, 85, 99, 79, 87, 80)
> lmod = lm(y ~ ABC + ABD + A * B * C * D, df)
> DanielPlot(lmod, alpha=0.1)
```

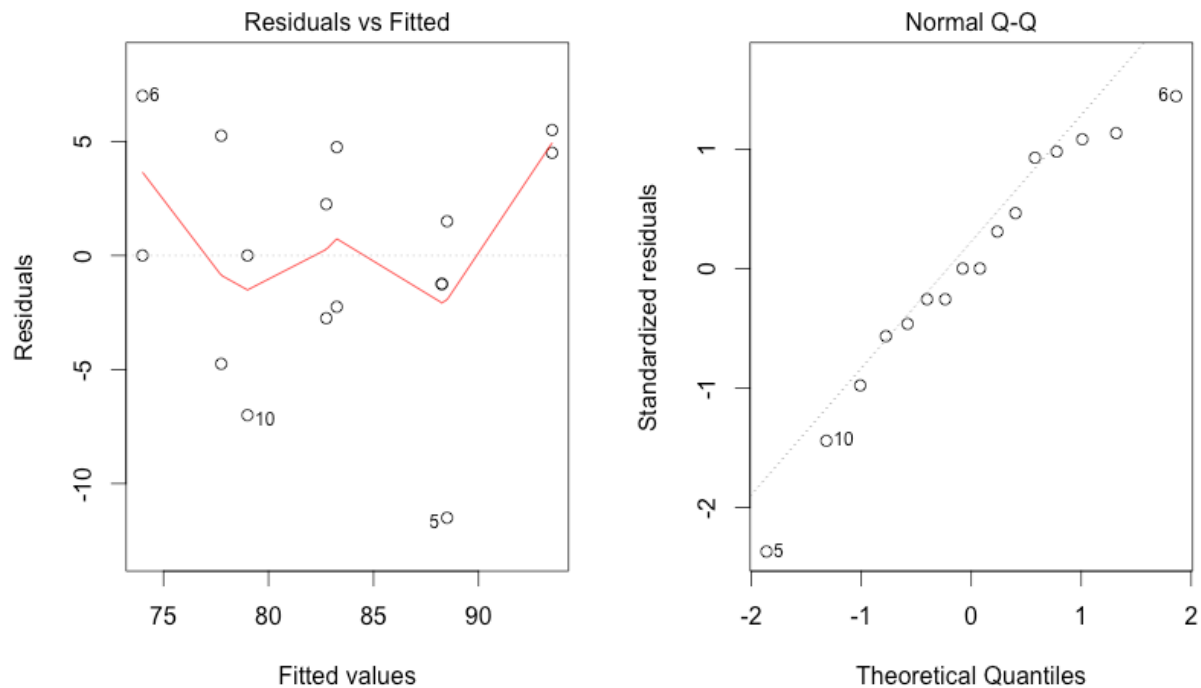


The above plot again shows factors A and ABC to be significant. Reducing the model to those:

```
> amod = aov(y ~ A + D + A:B, df)
> options(show.signif.stars=FALSE)
> summary(amod)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
A	1	400.0	400.0	11.687	0.00574
D	1	100.0	100.0	2.922	0.11542
A:B	2	83.3	41.6	1.216	0.33330
Residuals	11	376.5	34.2		

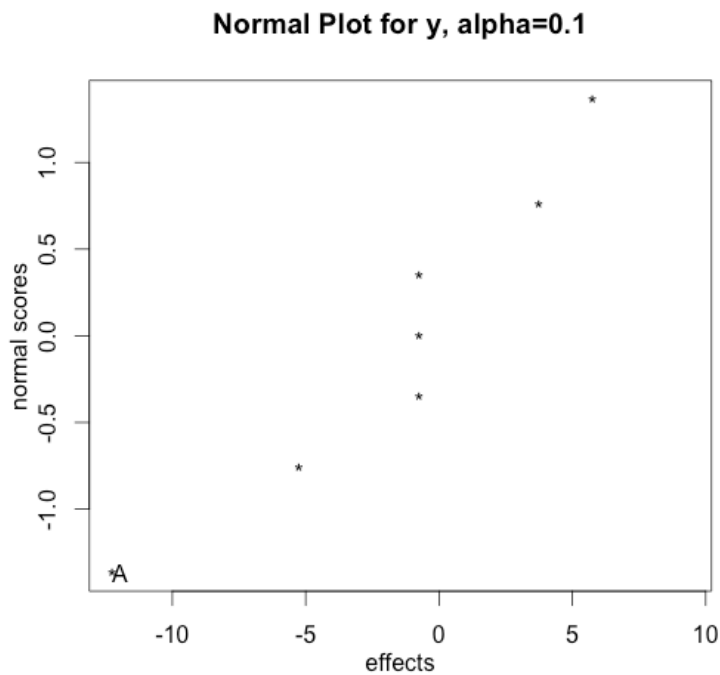
Only factor A's p-value is less than $\alpha=0.05$, so we reject the null hypothesis and state that it does have an effect on the outcome. The res/fit and QQ plots look acceptably distributed.



Montgomery 8.1

Suppose that in the chemical process development experiment described in Problem 6.7, it was only possible to run a one-half fraction of the 2^4 design. Construct the design and perform the statistical analysis, using the data from replicate I.

```
> library(FrF2)
> df = FrF2(nruns=8, nfactors=4, randomize=F)
> df$y = c(74, 81, 83, 73, 83, 81, 88, 90)
> df
  A B C D y
1 -1 -1 -1 -1 74
2  1 -1 -1  1 81
3 -1  1 -1  1 83
4  1  1 -1 -1 73
5 -1 -1  1  1 83
6  1 -1  1 -1 81
7 -1  1  1 -1 88
8  1  1  1  1 90
class=design, type= FrF2
> DanielPlot(amod, autolab=T, alpha=0.1)
```



The above plot shows that only factor A is significant. Reducing the model to just this, we find that A falls below $\alpha=0.05$ and reject the null hypothesis to state that A does significantly affect the outcome.

```
> a2 = aov(y ~ A, df)
> summary(a2)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
A	1	300.1	300.12	11.79	0.0139
Residuals	6	152.8	25.46		

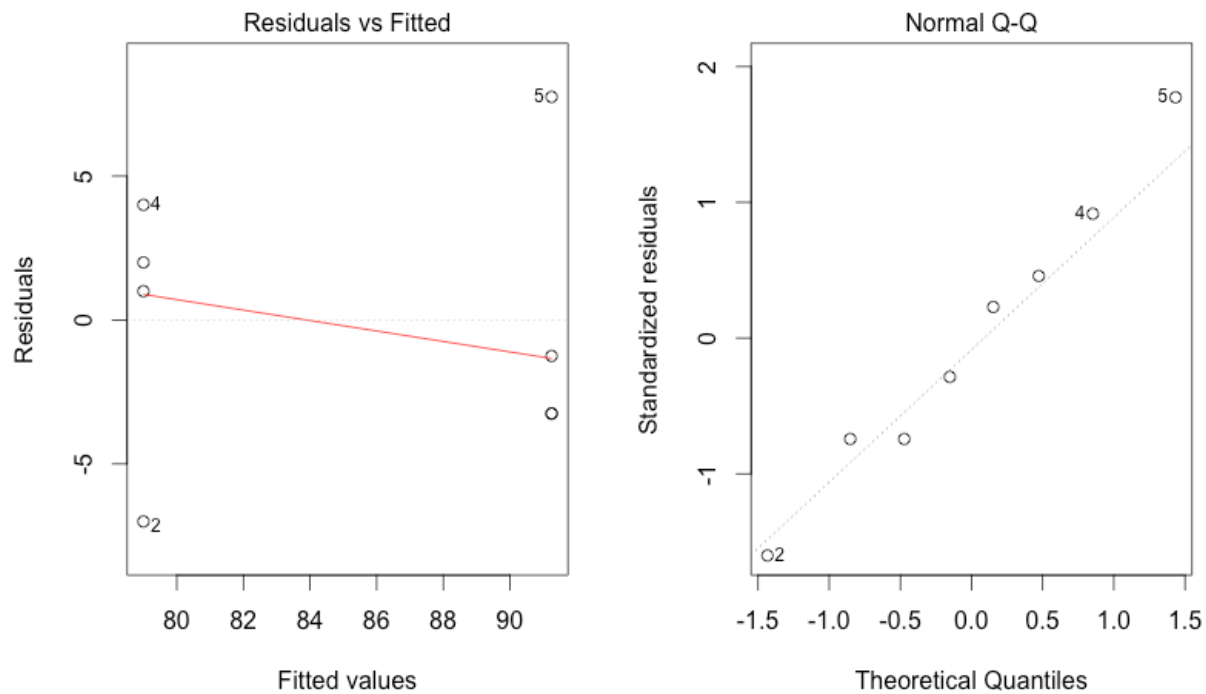
Using the coefficients, we find that the regression formula is:

```
> coef(a2)
```

	A1
(Intercept)	85.125
	-6.125

$$y = 85.125 - 6.125(A)$$

Inspection of res/fit and QQ plots looks fine:



Montgomery 8.49

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

```
> levels = c(-1,1)
> df = expand.grid(A=levels, B=levels, C=levels, D=levels)
> df$E = df$A * df$B * df$C * df$D * -1
> df[] = lapply(df, factor)
> df$y = c(63,21,36,99,24,66,71,54,23,74,80,33,63,21,44,96)
> df
```

	A	B	C	D	E	y
1	-1	-1	-1	-1	-1	63
2	1	-1	-1	-1	1	21
3	-1	1	-1	-1	1	36
4	1	1	-1	-1	-1	99
5	-1	-1	1	-1	1	24
6	1	-1	1	-1	-1	66
7	-1	1	1	-1	-1	71
8	1	1	1	-1	1	54
9	-1	-1	-1	1	1	23
10	1	-1	-1	1	-1	74
11	-1	1	-1	1	-1	80
12	1	1	-1	1	1	33
13	-1	-1	1	1	-1	63

```

14  1 -1  1  1  1 21
15 -1  1  1  1  1 44
16  1  1  1  1 -1 96

```

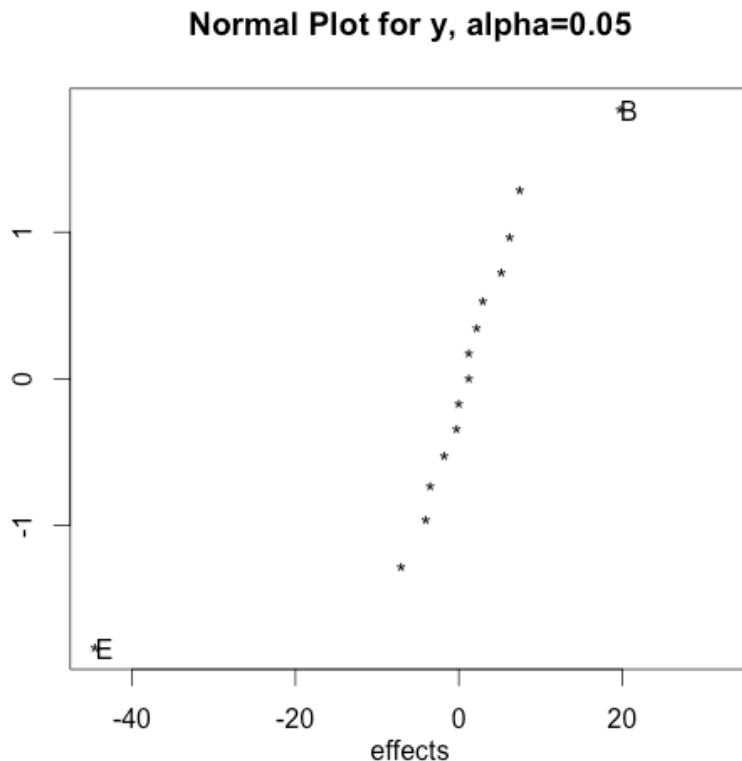
This is a 2^5-1 factorial design where E has been confounded with $-(ABCD)$.

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

```

> lmod = lm(y ~ A*B*C*D*E, df)
> DanielPlot(lmod)

```



The above plot shows factors B and E to be significant.

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

```

> options(show.signif.stars=FALSE)
> reduced.amod = aov(y ~ B + E, df)
> summary(reduced.amod)

```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
B	1	1560	1560	23.0	0.000349
E	1	7921	7921	116.8	7.26e-08
Residuals	13	882	68		

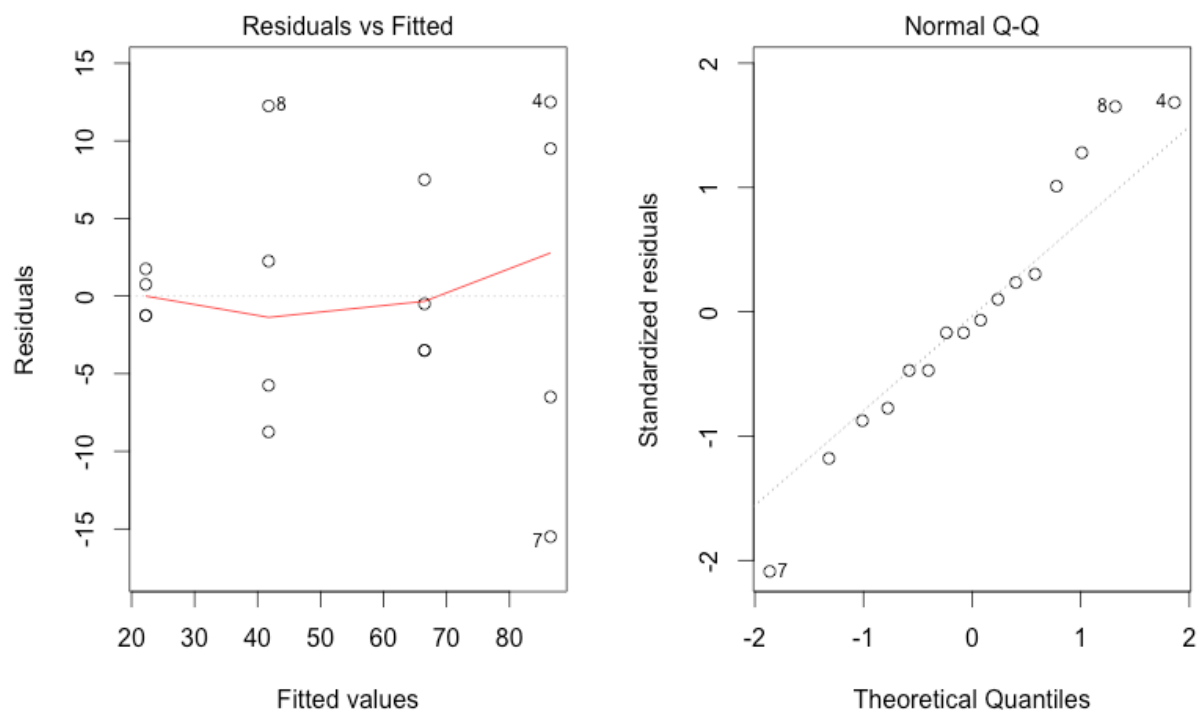
ANOVA of a reduced model shows that p-values for factors B and E fall well below $\alpha=0.05$, therefore we reject the null hypothesis and state that both have an effect.

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

```
> coef(reduced.amod)
(Intercept)      B1      E1
    66.625    19.750   -44.500
```

$$y = 66.625 + 19.75(B) - 44.5(E)$$

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



The residuals are evenly distributed on both sides of zero, and the normal QQ plot shows they mostly hew to the line with just three samples being labeled as outliers; therefore, we accept that the data is normally distributed.