STAT571
Homework 3
Ken Youens-Clark

4.1

Source	DF	SS	MS	F	P
Treatment	4	1010.56	**252.64**	29.84	**< 0.0001**
Block	**5**	**323.82**	64.765	**7.65**	**0.0004**
Error	20	169.33	**8.4665**		
Total	29	1503.71			

```
Block DF     = Total DF - Error DF - Treatment DF = 29 - 20 - 4 = 5
Block SS     = Total SS - Error SS - Treatment SS
             = 1503.71 - 169.33 - 1010.56 = 323.82
Treatment MS = Treatment SS / Treatment DF = 1010.56 / 4 = 252.64
Error MS     = Error SS / Error DF = 169.33 / 20 = 8.4665
Block F      = Block MS / Error MS = 64.765 / 8.4665 = 7.65
```

4.12

a) Is there evidence to support the claim that the treatments differ?

This appears to be a one-way ANOVA based purely on treatment (drug) and response (from genes):

```
> dat = read.csv("~/work/stat571/hw03/m4.12.dat")
> amod = aov(response ~ treatment, data=dat)
> summary(amod)
            Df  Sum Sq Mean Sq F value Pr(>F)
treatment    2  538442  269221   3.249 0.0544 .
Residuals   27 2236974   82851
```

$P > 0.05$, so we fail to reject the null hypothesis therefore, there is no difference.

If we add blocking on observations (patients):

```
> dat = read.csv("~/work/stat571/hw03/m4.12.dat")
> dat$obs = factor(rep(1:10, 3))
> amod = aov(response ~ treatment + obs, data=dat)
> summary(amod)
            Df  Sum Sq Mean Sq F value Pr(>F)
treatment    2  538442  269221   3.681 0.0457 *
```

```
obs          9  920632  102292   1.399 0.2597
Residuals   18 1316343   73130
```

Then we reject H_0 as the p-value (0.0457) is < 0.05 and say that there is a difference.

Blocking does not seem justified because patients are randomly chosen, i.e., not selected based on any similarity like age, race, genetics, etc., but it seems as if it must be required since this chapter is about blocking and not just one-way ANOVA.

b) Check normality.

QQ plot of unblocked is shown on the left, blocked on the right. Blocking definitely improves normality of data.
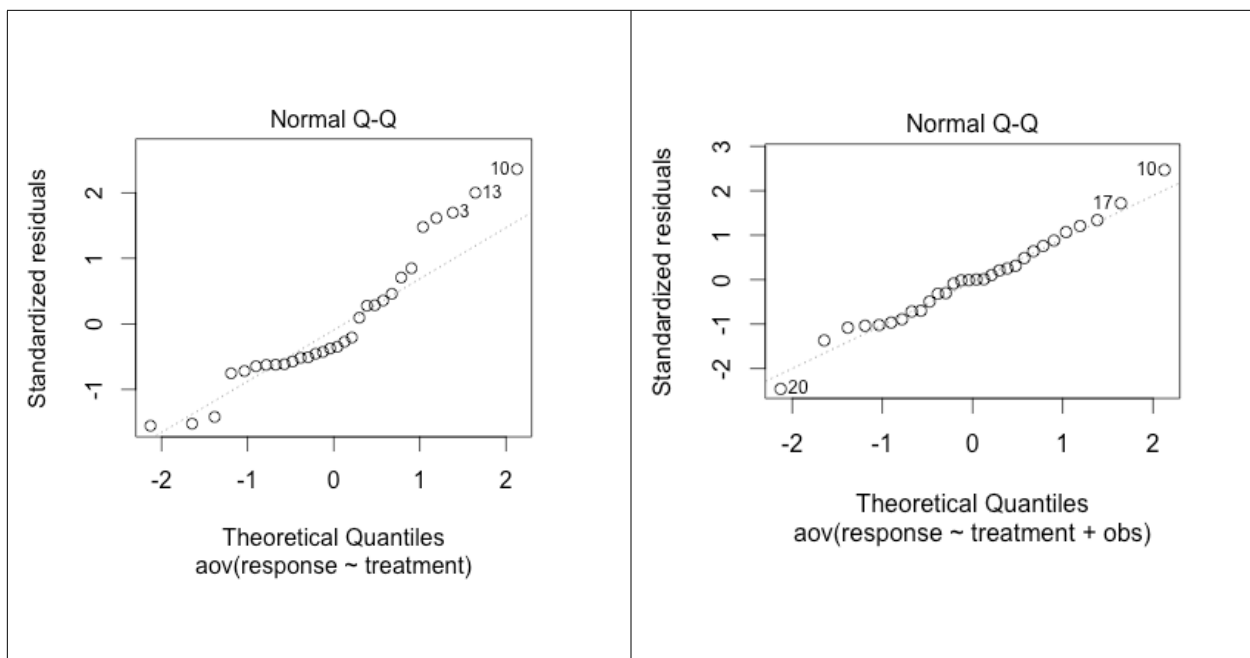

c) Take log of raw data. Is there evidence to support treatment means differ?

```
> dat$response.log = log(dat$response)
> amod.log = aov(response.log ~ treatment + obs, data=dat)
> summary(amod.log)
            Df Sum Sq Mean Sq F value Pr(>F)
treatment    2  6.297   3.148   3.094  0.070 .
obs          9 14.749   1.639   1.610  0.186
Residuals   18 18.319   1.018
```

As the p-value $0.07 < 0.05$, we fail to reject the null hypothesis and say that there is no evidence to support a difference in treatment means.

d) Analyze residuals from transformed data.

Normal Q-Q
Im(response.log ~ treatment + obs)
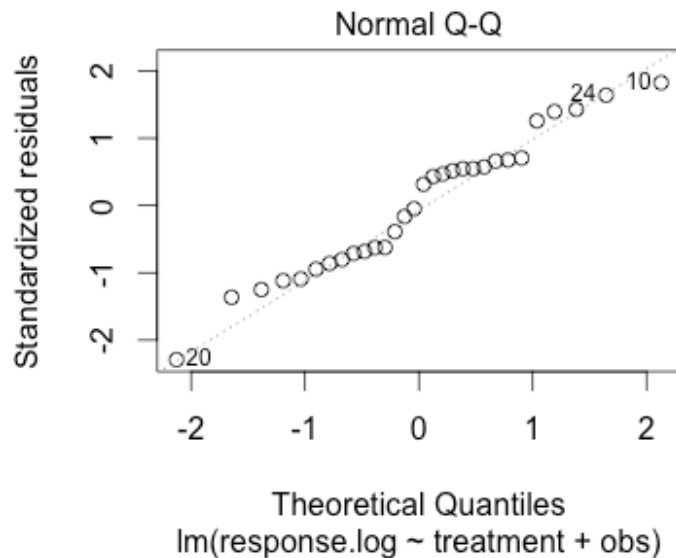
Three outliers (20, 24, 10) are highlighted by R, and the points vary widely from the predicted, so normality of logged data is not supported.

4.16

Assuming that chemical types and bolts are fixed, estimate the model parameters t_i and b_i in problem 4.3.

```
> dat = read.csv("~/work/stat571/hw03/4.3.long.dat")
> dat$bolt = factor(rep(1:5, 4))
> head(dat)
  treatment response bolt
1        c1       73    1
2        c1       68    2
3        c1       74    3
4        c1       71    4
5        c1       67    5
6        c2       73    1
> amod = aov(response ~ treatment + bolt, data=dat)
> summary(amod)
            Df Sum Sq Mean Sq F value   Pr(>F)
treatment    3  12.95    4.32   2.376    0.121
bolt         4 157.00   39.25  21.606 2.06e-05 ***
Residuals   12  21.80    1.82
```

4.20

4.28

To estimate block variance from 4.12:

[MS(Blocks) – MS(Error)]/a = (102292 – 73130) / 3 = 9720.6667