

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

b) How many blocks were used in this experiment?

Six blocks (DF + 1)

c) What conclusions can you draw?

Low p-value means we should reject the null hypothesis, therefore there is a difference in treatments. Blocking seems to have been a good choice as the F statistic suggests there was a good deal of variability there.

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)
```

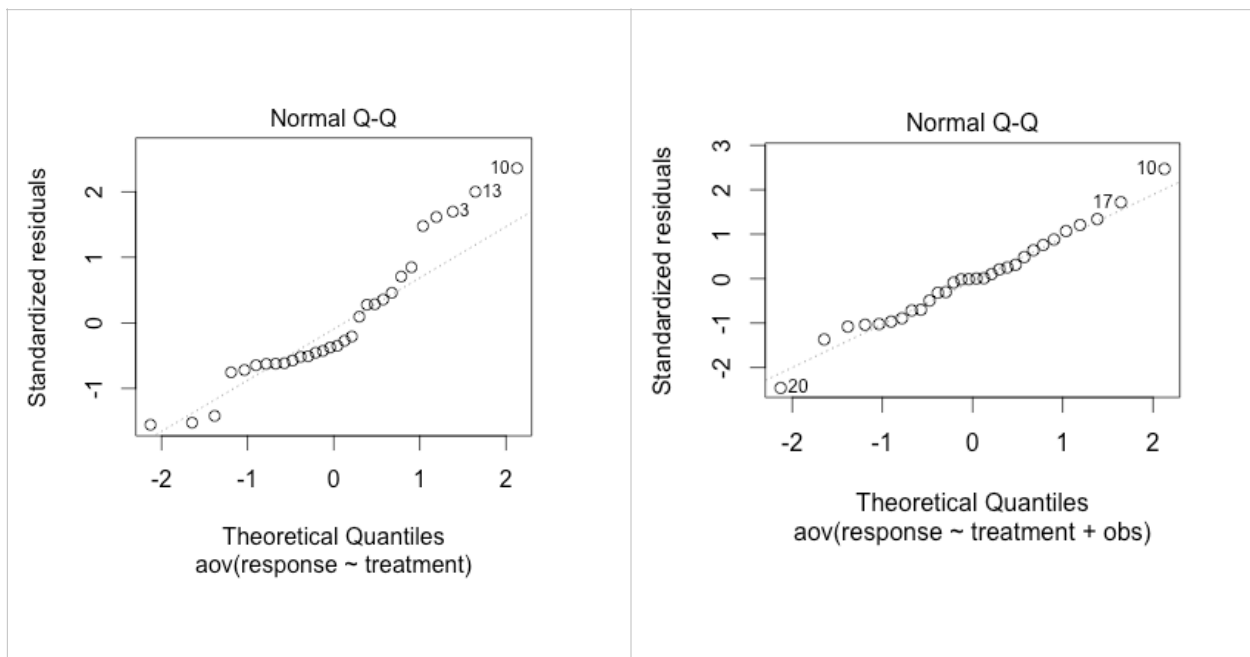
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
treatment	2	538442	269221	3.681	0.0457 *
obs	9	920632	102292	1.399	0.2597
Residuals	18	1316343	73130		

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?

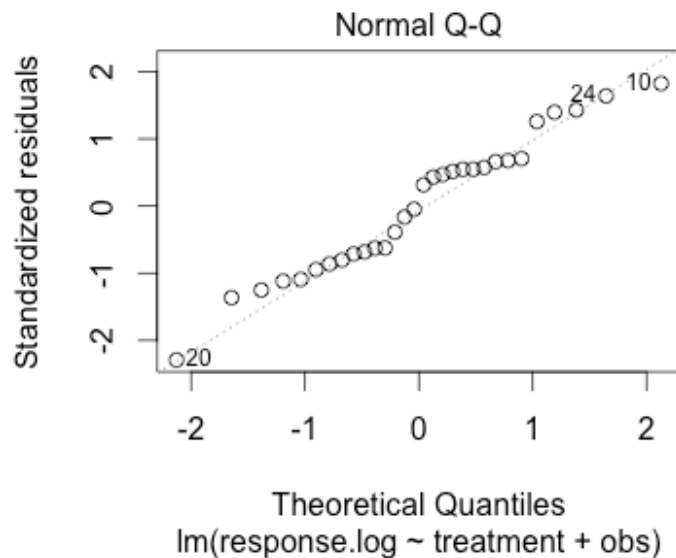
```
> dat$response.log = log(dat$response)
```

```
> 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.



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.

4.19

Suppose Tip 2/Coupon 3 is missing:

Tip	Coupon			
	1	2	3	4
1	9.3	9.4	9.6	10.0
2	9.4	9.3	9.8	9.9
3	9.2	9.4	9.5	9.7
4	9.7	9.6	10.0	10.2

```

a = # treatments (tips) = 4
b = # blocks (coupons) = 4
y'[i.] = 9.4 + 9.3 + 9.9 = 28.6 (row)
y'[,j] = 9.6 + 9.5 + 10.0 = 29.1 (column)
y'.. = 144.2
x      = (ay'[i.] + by'[,j] - y'..) / (a - 1)(b - 1) =
        = [4(28.6) + 4(29.1) - 144.2] / (3)(3) = 114.4 + 116.4 - 144.2 / 9
        = 86.6/9 = 9.6

```

Entering this into R:

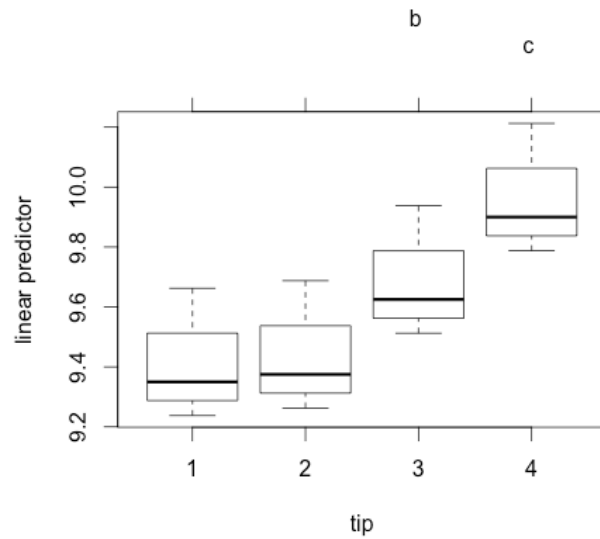
```

> tip = factor(c(rep(1,4), rep(2,4), rep(3, 4), rep(4,4)))
> coupon = factor(1:4)
> response = c(9.3, 9.4, 9.2, 9.7,
+             9.4, 9.3, 9.4, 9.6,
+             9.6, 9.6, 9.5, 10.0,
+             10.0, 9.9, 9.7, 10.2)
> df = data.frame(tip, coupon, response)
> matrix(df$response, 4, 4)
      [,1] [,2] [,3] [,4]
[1,]  9.3  9.4  9.6 10.0
[2,]  9.4  9.3  9.6  9.9
[3,]  9.2  9.4  9.5  9.7
[4,]  9.7  9.6 10.0 10.2
> amod = aov(response ~ tip + coupon, data=df)
> summary(amod)
              Df Sum Sq Mean Sq F value    Pr(>F)
tip              3  0.7925  0.26417    38.04 1.94e-05 ***
coupon           3  0.4025  0.13417    19.32 0.000292 ***
Residuals       9  0.0625  0.00694
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

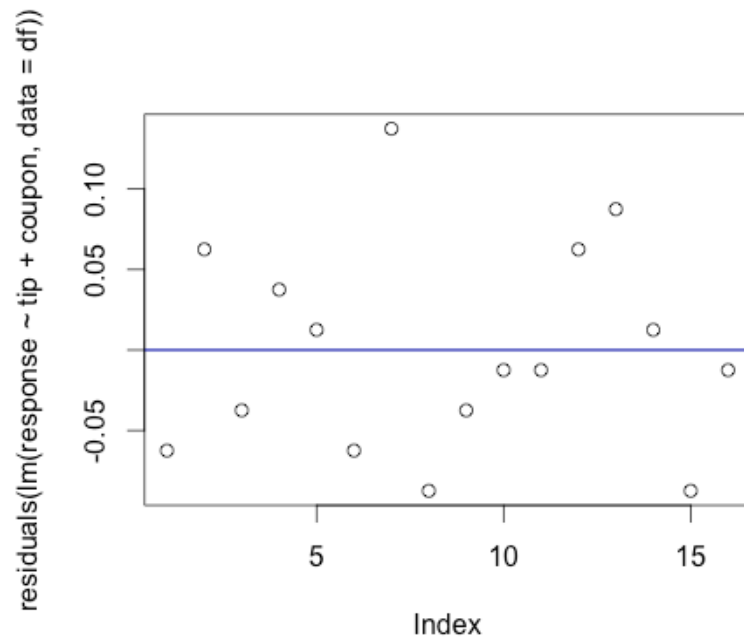
We reject the null hypothesis as the p-value (1.94e-05) is so low.

b) Tips 1 & 2 group together, 3 & 4 separately:



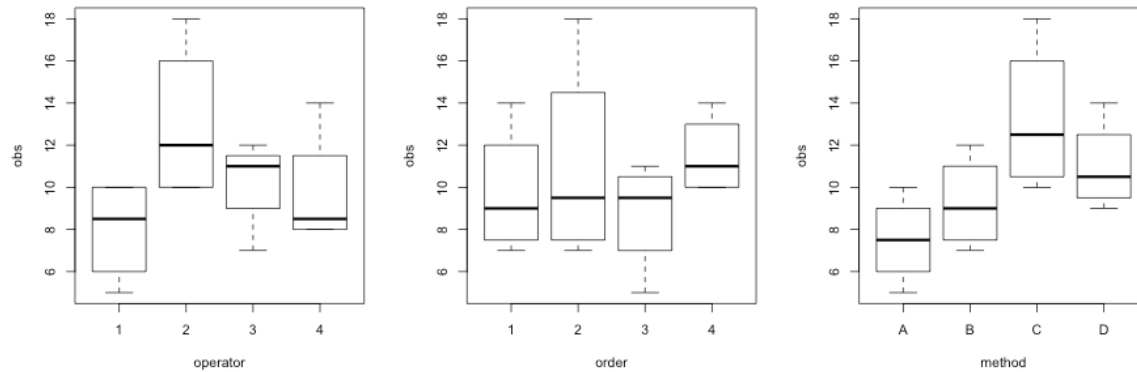
c) Inspection of residuals seems to show randomness associated with normally distributed data.

```
> plot(residuals(lm(response ~ tip + coupon, data=df)))
> abline(0, 0, col='blue')
```



4.23

Boxplots of the three factors (operator, order, method) show large variability by operator, moderate variability by method, and little by order.



A three-way ANOVA supports this:

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
operator	3	51.5	17.167	9.810	0.00993 **
order	3	18.5	6.167	3.524	0.08852 .
method	3	72.5	24.167	13.810	0.00421 **
Residuals	6	10.5	1.750		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Only "order" is found to support the null hypothesis of no difference by order, but p-values for both "operator" and "method" indicate rejection of the null hypothesis.

Analysis of residuals and QQ plot show basically normal data with three values being slightly off.

