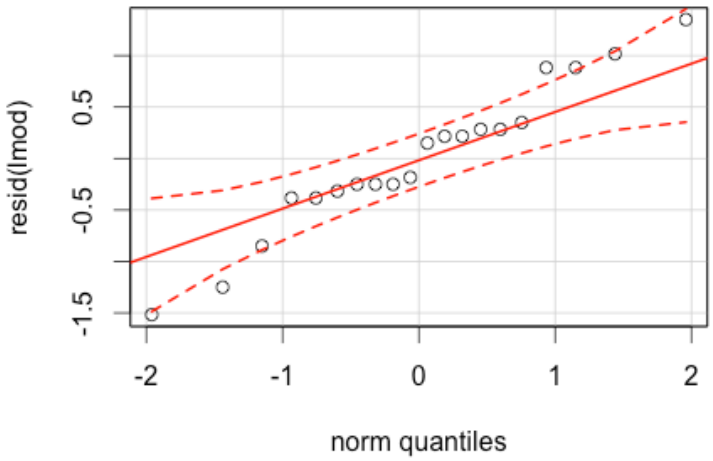
Ken Youens-Clark
STAT571B
Homework 4

1. Montgomery 4.40

```
> library(car)
> data = read.csv(file.path("~/work/stat571/hw04/4.40.dat"))
> data$additive = factor(data$additive)
> data$car = factor(data$car)
> lmod = lm(y~additive+car, data)
>
> # not very normal
> qqPlot(resid(lmod))
```




```
> # Shapiro says p=0.58
> shapiro.test(data$y)

      Shapiro-Wilk normality test

data:  data$y
W = 0.96194, p-value = 0.5833

> summary(aov(y~additive+car, data))
            Df Sum Sq Mean Sq F value  Pr(>F)
additive     4  31.70   7.925   8.703 0.00203 **
car          4  35.23   8.808   9.673 0.00132 **
Residuals   11  10.02   0.911
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

The QQ plot of the response (y) does not look normally distributed, but the Shapiro-Wilk reports a very high p-value (almost .6), so we'll accept the data. ANOVA shows

that the gasoline additive has a p-value 0.002 << α=0.05, so we reject the null hypothesis and state it has a significant affect on mileage performance.

2. Montgomery 5.1

Source	DF	SS	MS	F	P
A	1	0.322	**0.322**	**0.0367**	**0.8513**
B	**2**	80.554	40.2771	4.59	**0.0331**
Interaction	**2**	**45.348**	**22.674**	**2.5833**	**0.1167**
Error	12	105.327	8.7773		
Total	17	231.551			

A MS = A SS / A DF = 0.322 / 1 = 0.322
A F = A MS / Error MS = 0.322 / 8.7773 = 0.0367
A P = A F (A DF, Error DF) = 0.0367 (1, 12) = 0.8513
B DF = B SS / B MS = F 80.554 / 40.2771 ~= 2
B P = B F (B DF, Error DF) = F 4.59 (2, 12) = 0.0331
Interaction DF = $(a - 1)(b - 1)$ = (2-1)(3-1) = 1 * 2 = 2
Interaction SS = Total - Error - A - B = 231.551 - 105.327 - 80.554 - 0.322
Interaction MS = Interaction SS / DF = 45.348 / 2 = 22.674
Interaction F = Interaction MS / Error MS = 22.674 / 8.7773 = 2.5833
Interaction P = Interaction F (Interaction DF, Error DF) = F 2.5833 (2, 12) = 0.1167

b) How many levels were used for factor B?

B DF + 1 = 2 + 1 = 3

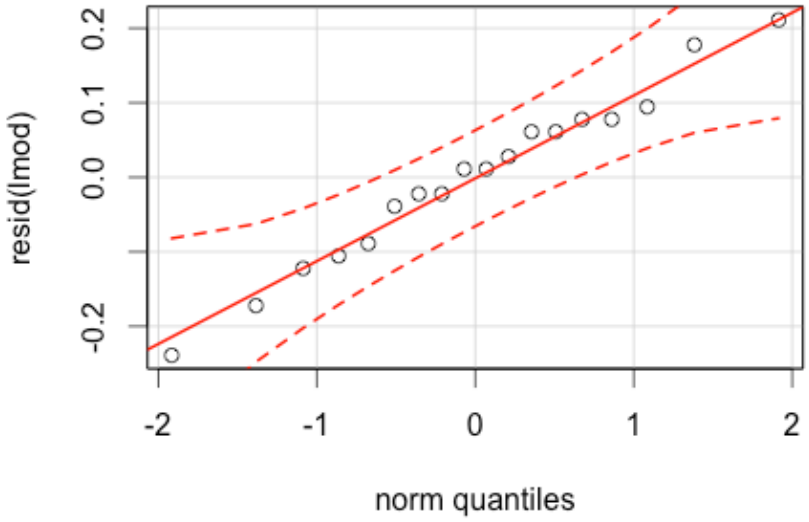
c) How many replicates of the experiment were performed?

N = Total DF + 1 = 18
N = abn (# factor A * # factor B * # replicates)
a = 2
b = 3
18 = 2 * 3 * n
18 = 6n
3 = n

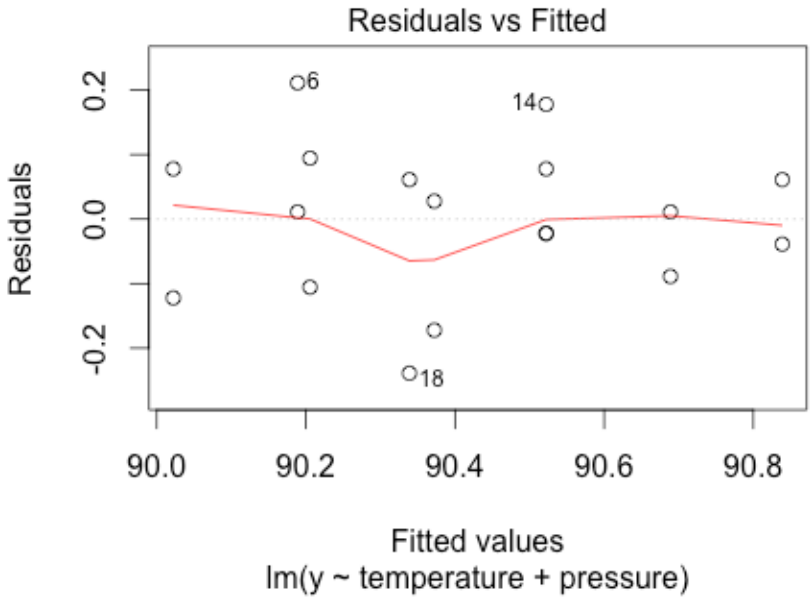
d) Given α=0.05, the null hypothesis is accepted for Factor A and the interaction of AB as their p-values are 0.85 and 0.11, respectively, but rejected for Factor B as it falls below (0.03). We can state that A and AB have no significant effect on outcomes, but B does.

3. Montgomery 5.3

```
> library(car)
> dat = read.csv("~/work/stat571/hw04/5.3.dat")
> dat$temperature = factor(dat$temperature)
> dat$pressure = factor(dat$pressure)
> shapiro.test(dat$y)

      Shapiro-Wilk normality test

data:  dat$y
W = 0.97363, p-value = 0.8625

> qqPlot(resid(lm(y~temperature+pressure, dat)))
```



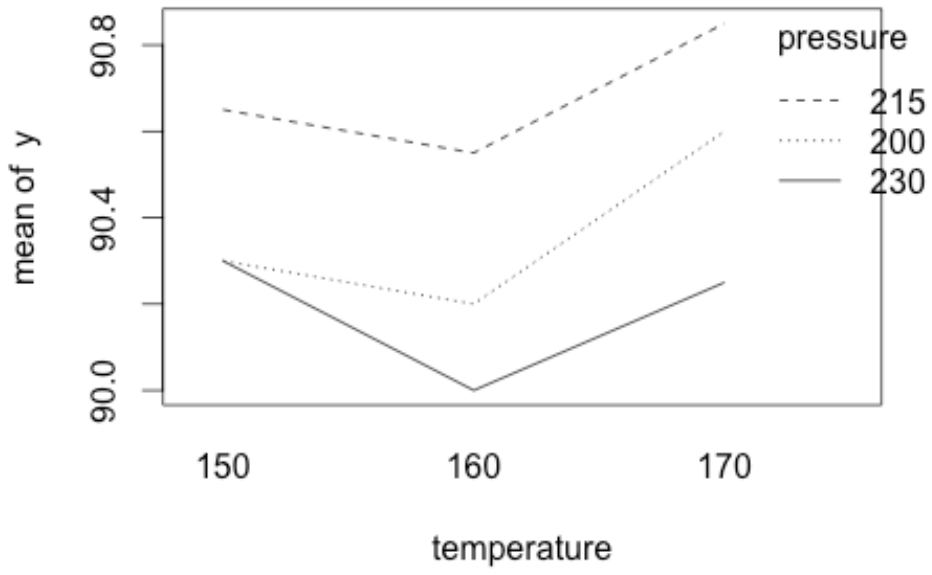

```
> lmod = lm(y~temperature+pressure+temperature*pressure, data=dat)
> summary(aov(lmod))
                    Df Sum Sq Mean Sq F value   Pr(>F)
temperature          2 0.3011  0.1506   8.469 0.008539 **
pressure             2 0.7678  0.3839  21.594 0.000367 ***
temperature:pressure 4 0.0689  0.0172   0.969 0.470006
Residuals            9 0.1600  0.0178
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```



a) The p-values for both temperature and pressure fall well below α=0.05, therefore we reject the null hypothesis and state that both factors have a significant affect on the yield. The p-value for the interaction of temperature and pressure (0.47) is well above 0.05, so it is found not to be a significant. Converging lines in the plot below show interaction with temperature is greatest when the pressure is 200.

b) The above QQ plot shows the data looks normally distributed. The residuals vs fitted plot and the very high p-value (0.8625) from the Shapiro test also confirms this.

c) Based on the interaction plot below, I would run at the highest yield given by a temperature of 170C and a pressure of 215.



4. Montgomery 5.15

First we can look for the significance of the two factors (row & column):

```
> dat = read.csv("~/work/stat571/hw04/5.15.dat")
> dat$Row = factor(dat$Row)
> dat$Col = factor(dat$Col)
> amod = aov(y~Row+Col, data=dat)
> summary(amod)
            Df Sum Sq Mean Sq F value   Pr(>F)
Row          2  580.5  290.25  60.399 0.000106 ***
Col          3   28.9    9.64   2.006 0.214717
Residuals    6   28.8    4.81
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

We can see that the Row factor is significant ($p << 0.05$) but Col is not ($p\ 0.2 > 0.05$). Next we use Tukey's test for non-additivity:

```
> library(dae)
> amod = aov(y ~ Row + Col + Error(Row/Col), data=dat)
> tukey.1df(amod, dat, error.term="Row:Col")
$Tukey.SS
[1] 3.540525

$Tukey.F
[1] 0.6999073

$Tukey.p
[1] 0.440953

$Devn.SS
[1] 25.29281
```

The reported p-value is 0.441 which is less than α=0.05, so we fail to reject the null hypothesis and therefore conclude there is no evidence of interaction.

5. Montgomery 5.21

An ANOVA of the data blocking on day and accounting for temperature and pressure effects on yield:

```
> dat = read.csv("~/work/stat571/hw04/5.21.dat")
> dat$day = factor(dat$day)
> dat$pressure = factor(dat$pressure)
> lmod = lm(y~day+temp+pressure+temp*pressure, dat)
> summary(aov(lmod))
              Df Sum Sq Mean Sq F value   Pr(>F)
day            1  13.01   13.01  24.480  0.00112 **
temp           2  99.85   49.93  93.981 2.78e-06 ***
pressure       2   5.51    2.75   5.184  0.03599 *
temp:pressure  4   4.45    1.11   2.095  0.17331
Residuals      8   4.25    0.53
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

The above output shows that both temperature and pressure have p-values < 0.05, so we reject the null hypotheses for these factors and state that it they have a significant effect on the yield. The interaction of temp/pressure has a p-value 0.17 > 0.05, so we accept the null hypothesis and state this does not affect yield.

Check of normal data via QQ plot and residuals/fitted show no problems.