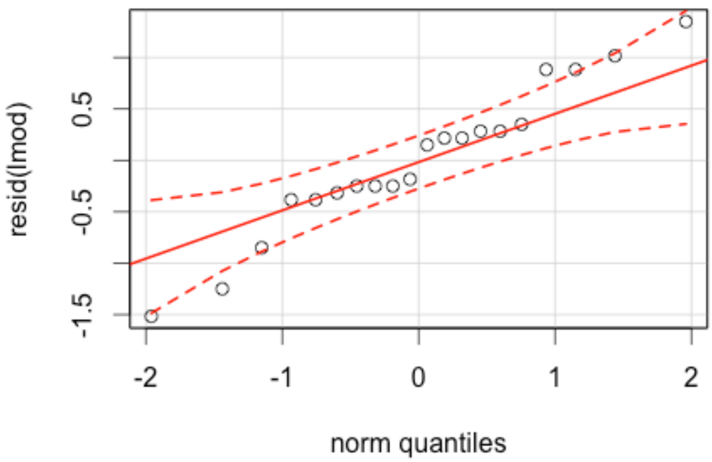
Ken Youens-Clark
STAT571B
Homework 4

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

> anova(lmod)
Analysis of Variance Table

Response: y
          Df Sum Sq Mean Sq F value   Pr(>F)
additive   4 31.700  7.9250   8.703 0.002026 **
car        4 35.233  8.8083   9.673 0.001321 **
Residuals 11 10.017  0.9106
---
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 neither the gasoline additive nor the car have a p-value above a significance level at **α**=0.05, so we reject the null hypothesis and state that both additive and car have a significant affect on mileage performance.

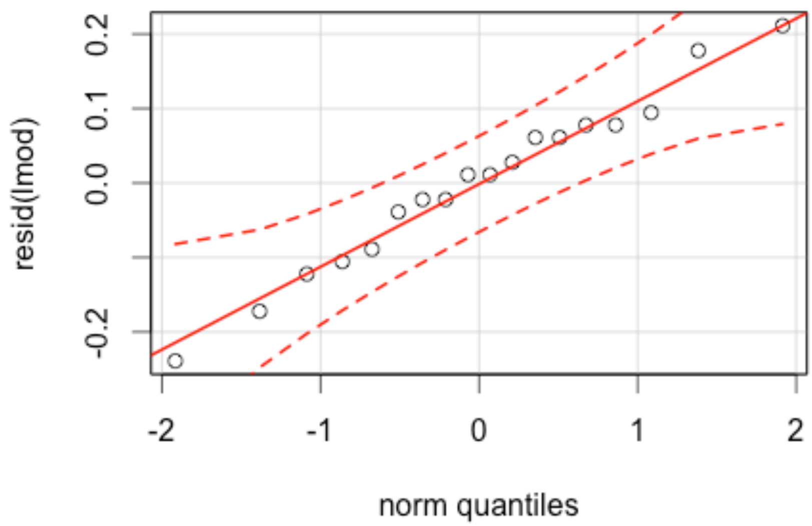
2. Montgomery 5.1

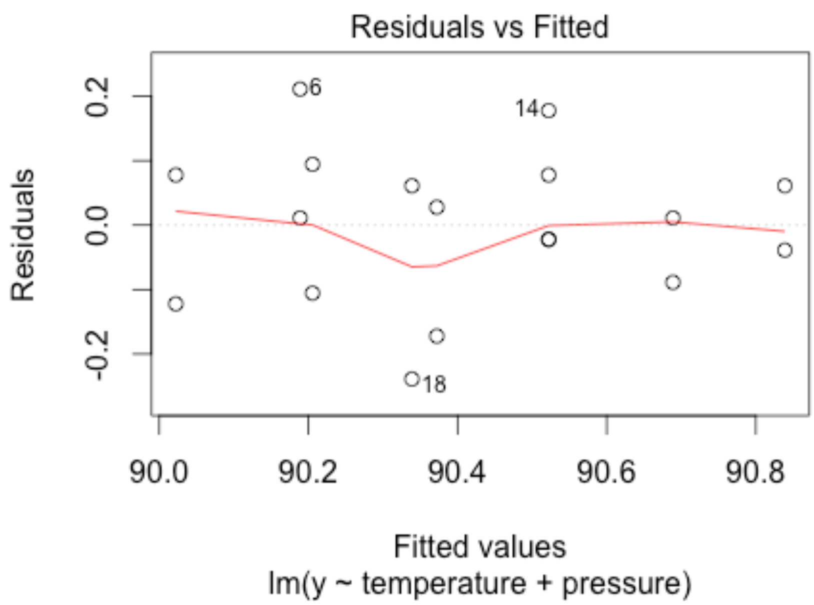
Source	DF	SS	MS	F	P
A	1	0.322	**0.322**	**0.0031**	**0.9565**
B	**2**	80.554	40.2771	4.59	**0.0331**
Interaction	**2**	**45.348**	**22.674**	**2.5833**	**0.1167**
Error	12	105.327	8.7773		
Total	17	231.551			

A MS = A SS / A DF = 0.322 / 1 = 0.322
A F = A MS / Error MS = 0.322 / 105.327 = 0.0031
A P = A F (A DF, Error DF) = 0.0031 (1, 12) = 0.9565
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

3. Montgomery 5.3

```
> library(car)
> dat = read.csv("~/work/stat571/hw04/5.5.dat")
> dat$temperature = factor(dat$temperature)
> dat$pressure = factor(dat$pressure)
>
> lmod = lm(y~temperature+pressure, data=dat)
> qqPlot(resid(lmod))
```





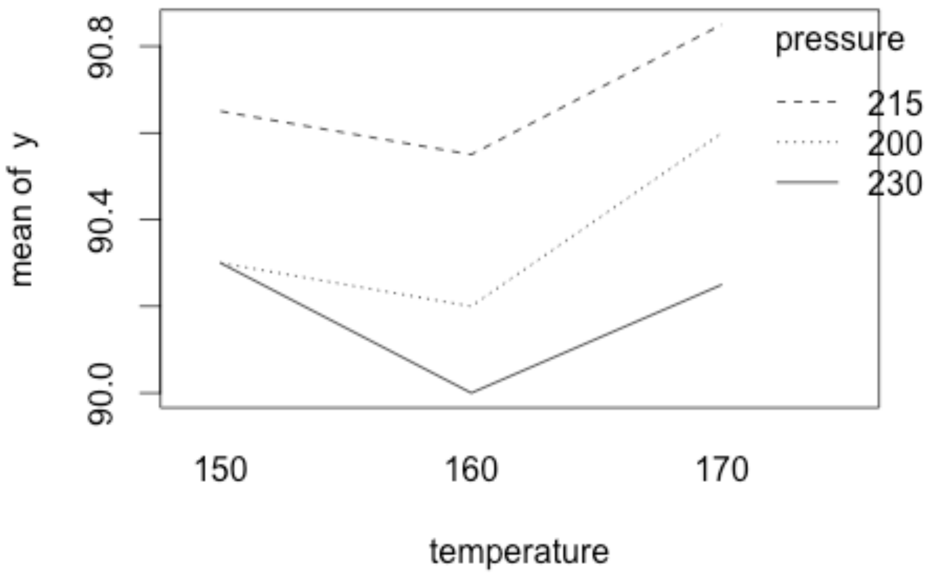

```
> shapiro.test(dat$y)

	Shapiro-Wilk normality test

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

> summary(aov(lmod))
            Df Sum Sq Mean Sq F value   Pr(>F)    
temperature  2 0.3011  0.1506   8.551  0.00426 ** 
pressure     2 0.7678  0.3839  21.803 7.03e-05 ***
Residuals   13 0.2289  0.0176                     
---
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.
b) The above QQ plot shows the data looks very normally distributed. The residuals vs fitted plot also 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

To test for non-additivity, we can use Tukey's 1-degree-of-freedom test:

```
> dat = read.csv("~/work/stat571/hw04/5.15.dat")
> dat$Row = factor(dat$Row)
> dat$Col = factor(dat$Col)
> library(dae)
> tukey.1df(aov(y~Row+Col, data=dat), dat)
```

This reports an F value of infinity and p value of 0, so we must reject the null hypothesis and state that there is additivity/interaction between the rows and columns.

Another package ("additivityTests") also reports that we ought to reject the null hypothesis:

```
> dat = matrix(c(36,18,30,39,20,37,36,22,33,32,20,34), nrow=3)
> dat
     [,1] [,2] [,3] [,4]
[1,]   36   39   36   32
```

```
[2,]   18   20   22   20
[3,]   30   37   33   34
> tukey.test(dat, alpha=0.05)

Tukey test on 5% alpha-level:

Test statistic: 0.6999
Critival value: 6.608
The additivity hypothesis cannot be rejected.
```

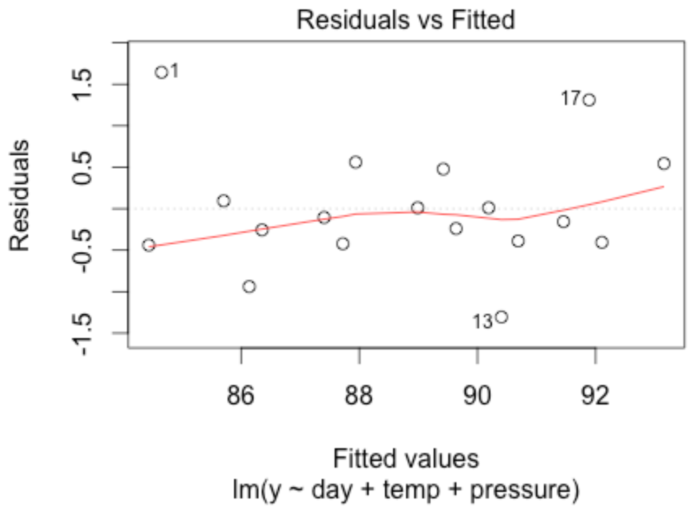
5. Montgomery 5.21

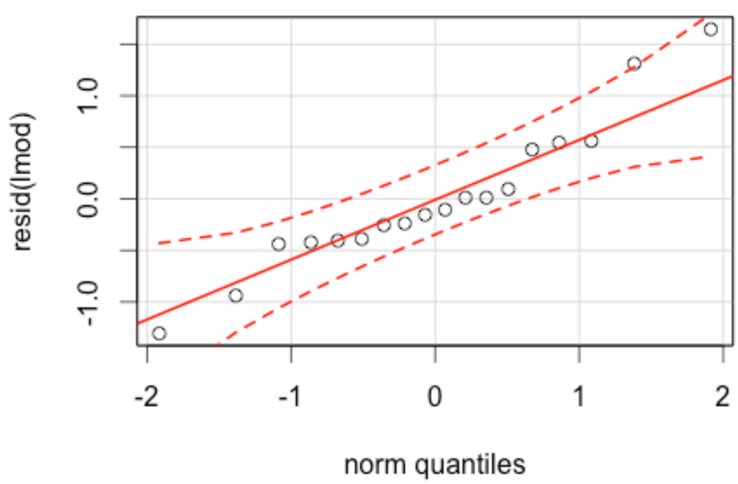
An ANOVA of the data blocking on day and accounting for temperature and pressure effects on yield:

```
> 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, dat)
> summary(aov(lmod))
            Df Sum Sq Mean Sq F value   Pr(>F)    
day          1  13.01   13.01  17.933  0.00116 ** 
temp         2  99.85   49.93  68.848 2.65e-07 ***
pressure     2   5.51    2.75   3.797  0.05275 .  
Residuals   12   8.70    0.73                     
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Shows that only pressure is found to have little effect on the yield whereas temperature has a significant effect and blocking by day is a good choice as that also seems to influence the outcome though not as much as temperature.

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


Residuals vs Fitted
Residuals
1.5
0.5
-0.5
-1.5
1
17
13
86
88
90
92
Fitted values
lm(y ~ day + temp + pressure)


resid(lmod)
1.0
0.0
-1.0
-2
-1
0
1
2
norm quantiles