STAT 571B, Homework 2
Ken Youens-Clark

1. Montgomery 3.3

One-way ANOVA

```
Source      DF      SS      MS      F       P
Factor       3   36.15   <MST>    <F>      <P>
Error    <DFE>   <SSE>   <MSE>
Total       19  196.04

DFE = N - a           = 20 - 4          = 16
SSE = SS(Total) / SST = 196.04 - 36.15  = 69.89
MST = SST / DFT       = 196.04 / 3      = 65.347
MSE = SSE / DFE       = 69.89 / 16      = 4.368
F   = MST / MSE       = 65.347 / 4.368  = 14.96
P   = F (Signal, Noise) = F (DFT, DFE)  = F (14.96, 3, 16) = < .00001
```

2. Montgomery 3.22 (skip part d)

a) Do the three circuits have the same response time?

```
> dat = read.delim("~/work/stat571/hw02/3.22.dat")
> dat$circuit = factor(dat$circuit) # to turn integers into factors
> amod = aov(response ~ circuit, data=dat)
> amod.sum = unlist(summary(amod))
> amod.sum['Pr(>F)1']
    Pr(>F)1
0.003785956
```

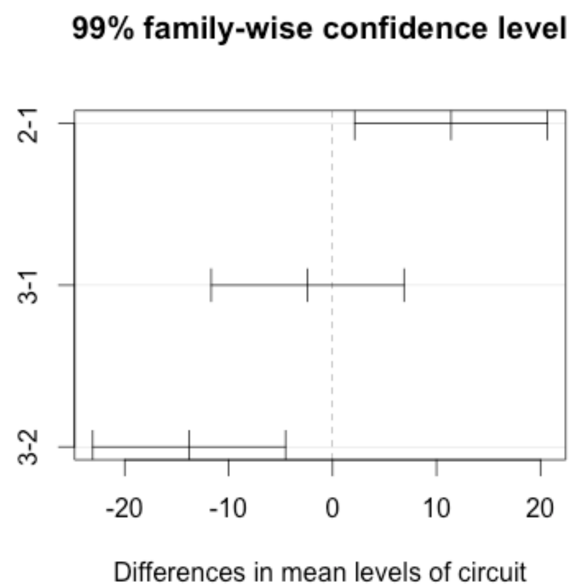
P << α = 0.01 ∴ reject H0 (no difference), response times are different

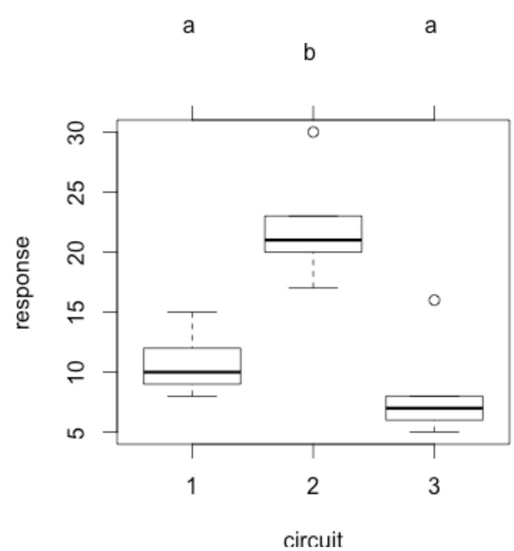
b) Tukey's test

```
> TukeyHSD(amod, conf.level=0.99)
  Tukey multiple comparisons of means
    99% family-wise confidence level

Fit: aov(formula = response ~ circuit, data = dat)

$circuit
     diff        lwr       upr     p adj
2-1  11.4   2.123163 20.676837 0.0023656
3-1  -2.4 -11.676837  6.876837 0.6367043
3-2 -13.8 -23.076837 -4.523163 0.0005042


> library("multcomp")
```

```
> tmod <- glht(amod, linfct = mcp(circuit="Tukey"))
> summary(tmod)

       Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: Tukey Contrasts

Fit: aov(formula = response ~ circuit, data = dat)

Linear Hypotheses:
           Estimate Std. Error t value Pr(>|t|)
2 - 1 == 0     11.4        2.6   4.385   0.0024 **
3 - 1 == 0     -2.4        2.6  -0.923   0.6367
3 - 2 == 0    -13.8        2.6  -5.308   <0.001 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
(Adjusted p values reported -- single-step method)
```

The combination 3-1 is the most significantly similar (overlap “0” in plot below, left). The others are different.





c) graphical comparison

```
> mean(dat[dat$circuit == 1, "response"])
[1] 10.8
> mean(dat[dat$circuit == 2, "response"])
[1] 22.2
> mean(dat[dat$circuit == 3, "response"])
[1] 8.4
> summary(amod)
            Df Sum Sq Mean Sq F value   Pr(>F)
```

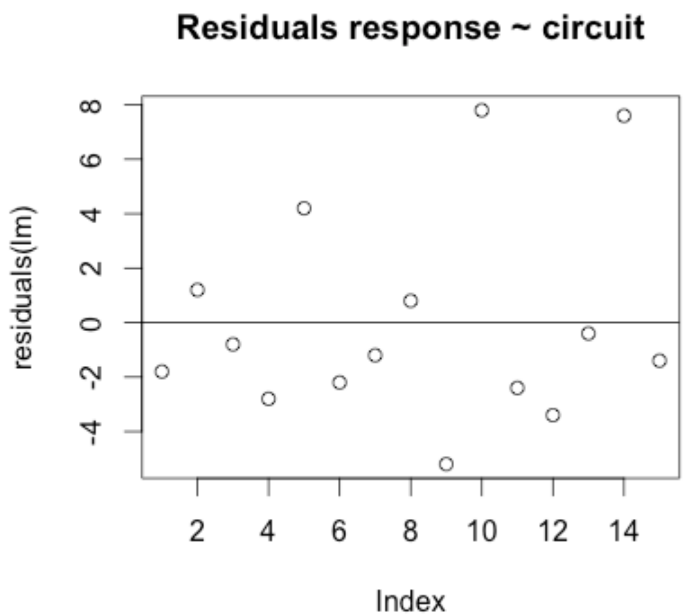
```
circuit      2  543.6   271.8   16.08 0.000402 ***
Residuals   12  202.8    16.9
```

scale = sqrt(MSE/n) = sqrt(16.9/15) = 1.06 * 3SD = 3.09

(see attached drawing)

e) Choose either circuit 1 or 3 as they have the lower response times which would be desirable for a shutoff valve.

f) The plot of residuals shows a random distribution of residuals, so basic analysis of variance assumptions are satisfied.



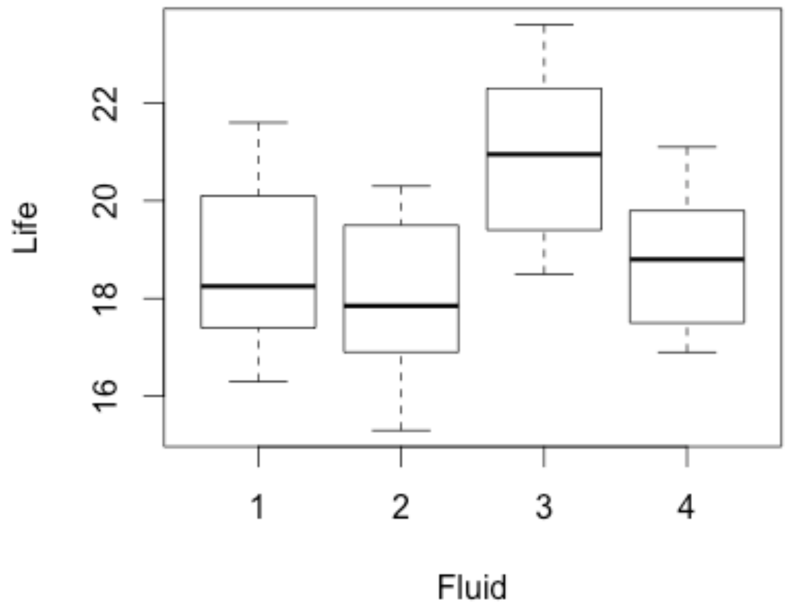
3. Montgomery 3.23

a) Do the fluids differ?

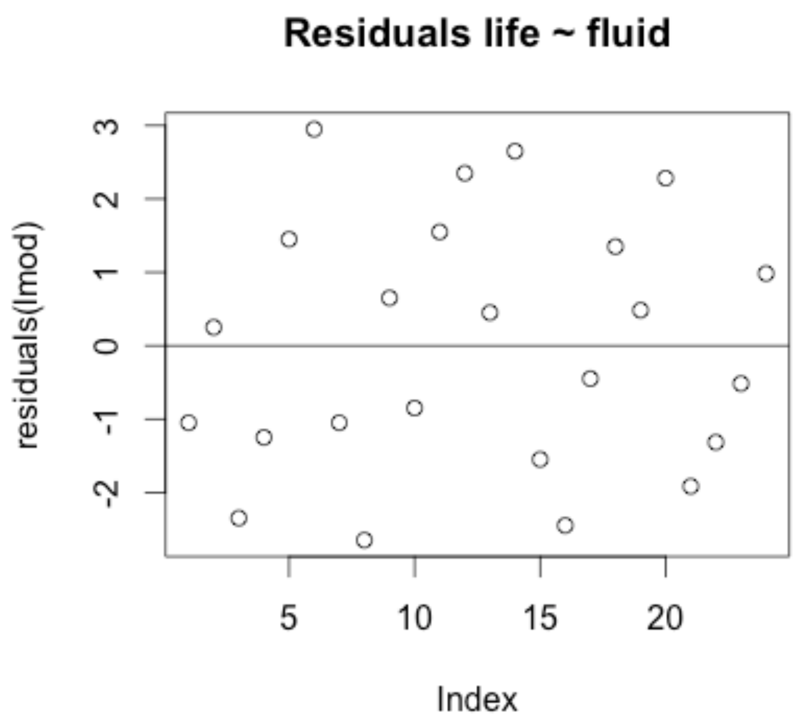
```
> dat = read.delim("~/work/stat571/hw02/3.23.long.dat")
> dat$fluid = factor(dat$fluid) # to turn integers into factors
> amod = aov(life ~ fluid, data=dat)
> amod.sum = unlist(summary(amod))
> amod.sum['Pr(>F)1']
   Pr(>F)1
0.05246316
```

P > α = 0.05 ∴ fail to reject null hypothesis, fluids are not significantly different

b) I would select fluid 3 as it has the longest life.



c) Plot of residuals is random, so basic analysis of variance satisfied.



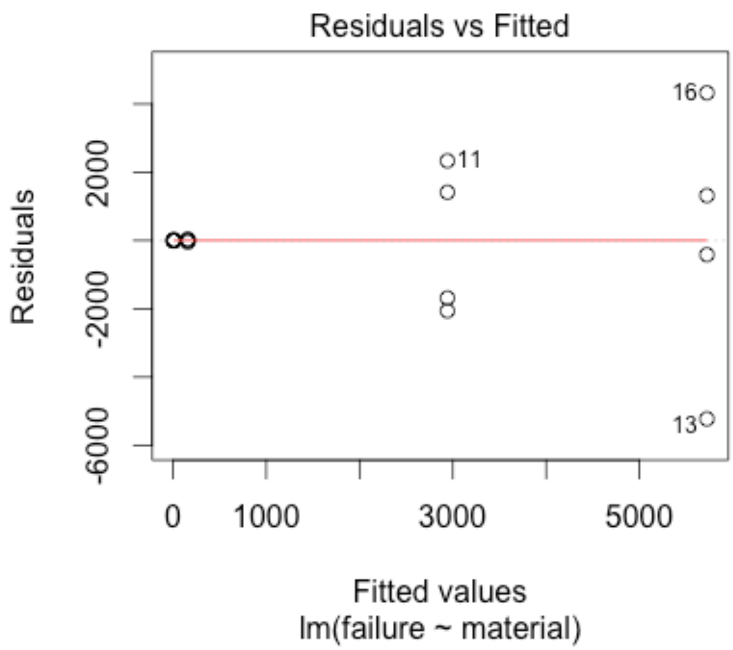
4. Montgomery 3.28

a) Do all five materials have the same effect on mean failure time?

```
> dat = read.delim("~/work/stat571/hw02/3.28.dat")
> dat$material = factor(dat$material)
> amod = aov(failure ~ material, data=dat)
> amod.sum = unlist(summary(amod))
> amod.sum['Pr(>F)1']
    Pr(>F)1
0.003785956
```

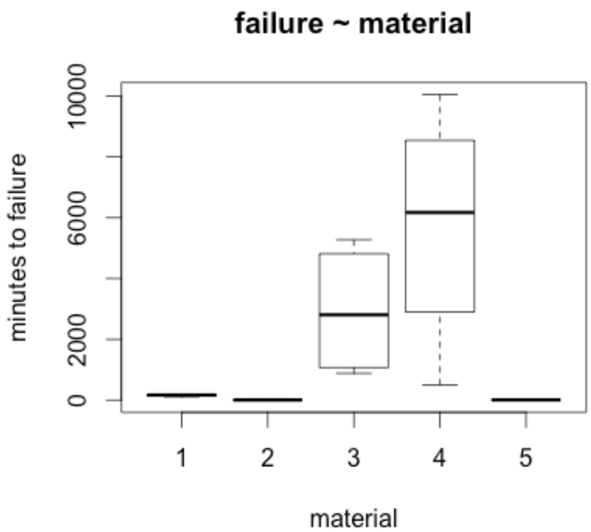
Very small p-value to support null hypothesis ∴ there is a difference.

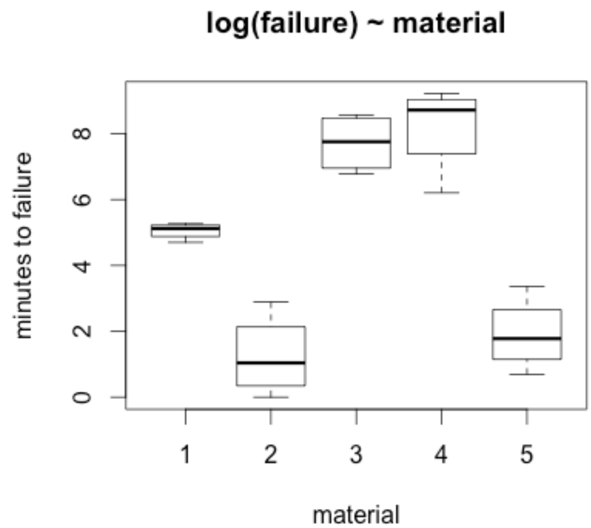
b) Plot of residuals vs predicted shows poor variance (opening funnel to right).



c) Transform "failure" by log

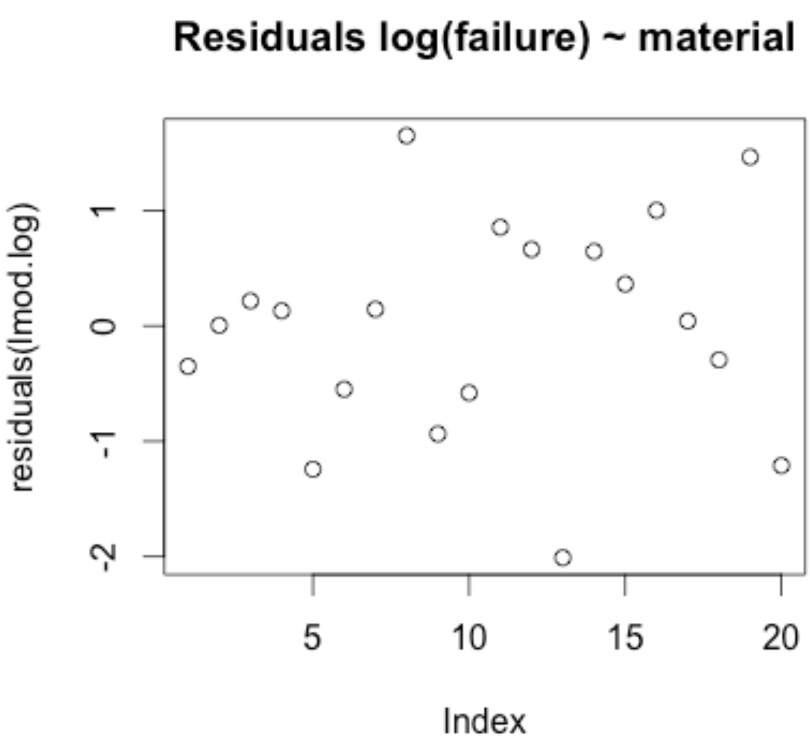
```
> dat$log.failure = log(dat$failure)
> boxplot(failure ~ material, data=dat, main = 'failure ~ material', ylab =
'minutes to failure', xlab = 'material')
> boxplot(log.failure ~ material, data=dat, main = 'log(failure) ~ material',
ylab = 'minutes to failure', xlab = 'material')
```





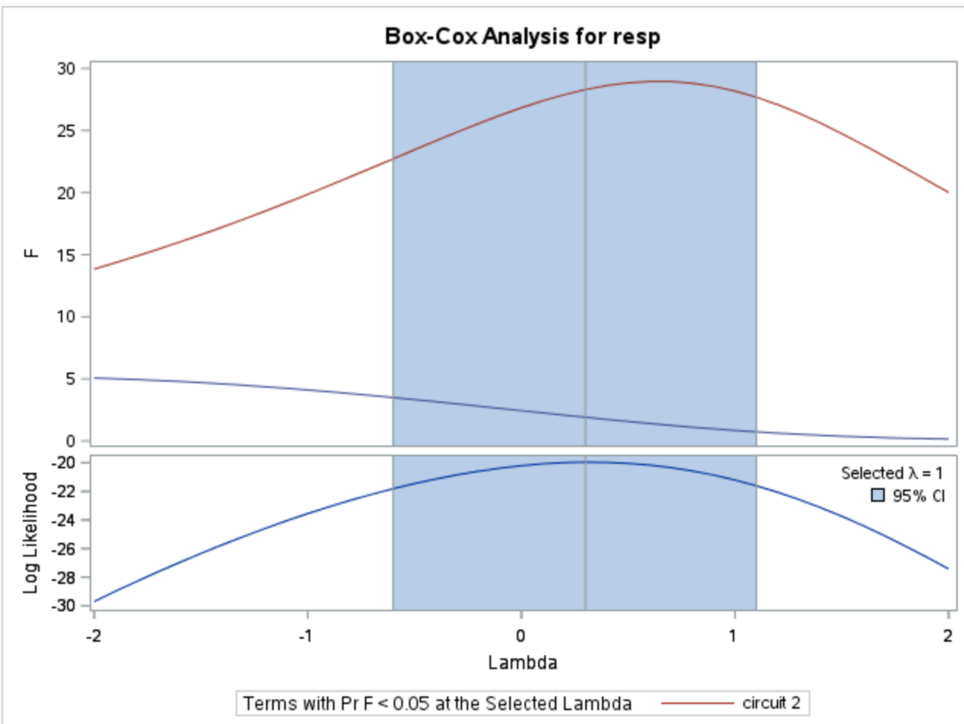
The above boxplots show the normalizing effect of transforming the minutes-to-failure data by natural logarithm. ANOVA analysis and scatterplots further support this change by an increase p-value and better distribution of residuals, respectively.

```
> amod2 = aov(log.failure ~ material, data=dat)
> amod2.sum = unlist(summary(amod2))
> amod2.sum['Pr(>F)1']
     Pr(>F)1
1.176093e-07
```



SAS also seems to suggest that log transformation is most appropriate:

```
data circuits;
input circuit resp @@;
datalines;
1 9 1 12 1 10 1 8 1 15
2 20 2 21 2 23 2 17 2 30
3 6 3 5 3 8 3 16 3 7
;
run;

proc transreg data=circuits;
model boxcox(resp/convenient lambda=-2.0 to 2.0 by 0.1)=class(circuit);
run;
```



5. Montgomery 3.51

```
> dat = read.delim("~/work/stat571/hw02/3.23.long.dat")
> dat$fluid = factor(dat$fluid)
> kruskal.test(life ~ fluid, data=dat)

      Kruskal-Wallis rank sum test

data:  life by fluid
Kruskal-Wallis chi-squared = 6.2177, df = 3, p-value = 0.1015
```

High p-value, so fail to reject null hypothesis.