One-Way ANOVA Handout

1 Initialization

```
> library(NCStats)
> library(multcomp)  # for multiple comparison methods
```

2 Raspberry Example

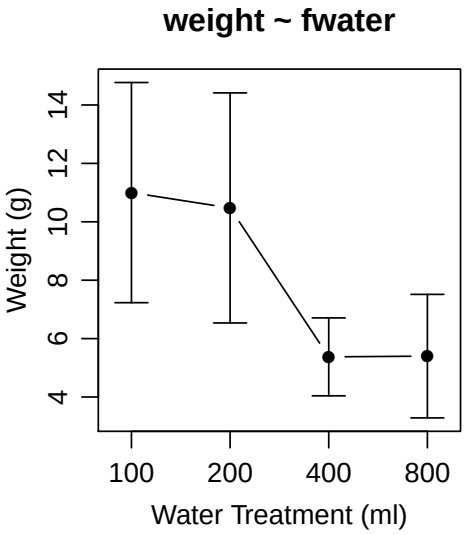
You must change the directory to where the following file is located.

```
> rasp <- read.table("Raspberry.txt",head=TRUE)
> str(rasp)

'data.frame':        16 obs. of  2 variables:
 $ water : int  100 100 100 100 200 200 200 200 400 400 ...
 $ weight: num  8.1 10.9 11.1 13.9 12.2 11.5 11.4 6.8 6.5 5.5 ...

> rasp$fwater <- factor(rasp$water)
> str(rasp)

'data.frame':        16 obs. of  3 variables:
 $ water : int  100 100 100 100 200 200 200 200 400 400 ...
 $ weight: num  8.1 10.9 11.1 13.9 12.2 11.5 11.4 6.8 6.5 5.5 ...
 $ fwater: Factor w/ 4 levels "100","200","400",..: 1 1 1 1 2 2 2 2 3 3 ...
```

2.1 Fitting the Linear Model

```
> lm1 <- lm(weight~fwater,data=rasp)
> anova(lm1)

Analysis of Variance Table

Response: weight
          Df  Sum Sq Mean Sq F value   Pr(>F)
fwater     3 115.043  38.348  10.793 0.001004
Residuals 12  42.635   3.553

> summary(lm1)

Call:
lm(formula = weight ~ fwater, data = rasp)

Residuals:
    Min      1Q  Median      3Q     Max
-3.6750 -0.5500  0.1125  1.0500  2.9000

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  11.0000     0.9425  11.672 6.58e-08
fwater200    -0.5250     1.3328  -0.394  0.70057
```

```
fwater400    -5.6250     1.3328  -4.220  0.00119
fwater800    -5.6000     1.3328  -4.202  0.00123

Residual standard error: 1.885 on 12 degrees of freedom
Multiple R-squared: 0.7296,        Adjusted R-squared: 0.662
F-statistic: 10.79 on 3 and 12 DF,  p-value: 0.001004

> fitPlot(lm1,xlab="Water Treatment (ml)",ylab="Weight (g)")
```

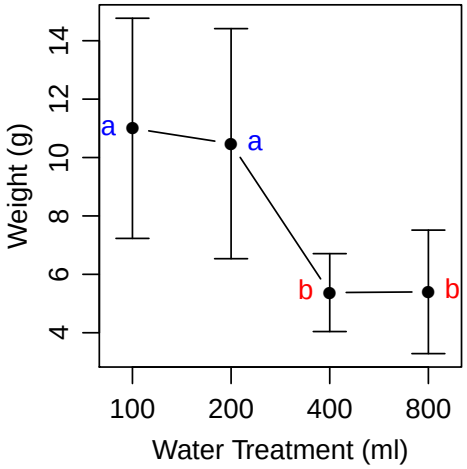


2.2 Multiple Comparison Tests

```
> rasp.mc <- glht(lm1, mcp(fwater = "Tukey"))
> summary(rasp.mc)

         Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: Tukey Contrasts


Fit: lm(formula = weight ~ fwater, data = rasp)

Linear Hypotheses:
               Estimate Std. Error t value Pr(>|t|)
200 - 100 == 0   -0.525      1.333  -0.394  0.97832
400 - 100 == 0   -5.625      1.333  -4.220  0.00546
800 - 100 == 0   -5.600      1.333  -4.202  0.00586
400 - 200 == 0   -5.100      1.333  -3.826  0.01100
800 - 200 == 0   -5.075      1.333  -3.808  0.01147
800 - 400 == 0    0.025      1.333   0.019  1.00000
(Adjusted p values reported -- single-step method)

> confint(rasp.mc)

         Simultaneous Confidence Intervals
```

```
Multiple Comparisons of Means: Tukey Contrasts


Fit: lm(formula = weight ~ fwater, data = rasp)

Quantile = 2.9679
95% family-wise confidence level


Linear Hypotheses:
               Estimate lwr     upr
200 - 100 == 0 -0.5250  -4.4807  3.4307
400 - 100 == 0 -5.6250  -9.5807 -1.6693
800 - 100 == 0 -5.6000  -9.5557 -1.6443
400 - 200 == 0 -5.1000  -9.0557 -1.1443
800 - 200 == 0 -5.0750  -9.0307 -1.1193
800 - 400 == 0  0.0250  -3.9307  3.9807
```

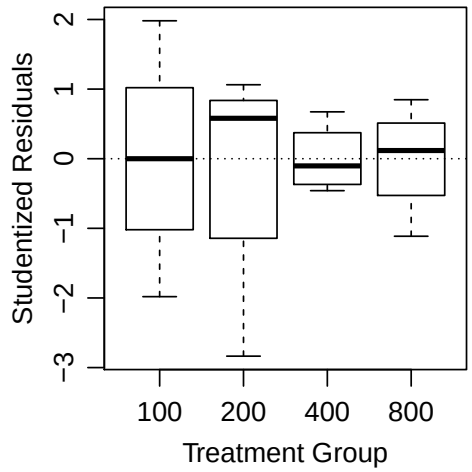
```
> fitPlot(lm1,xlab="Water Treatment (ml)",ylab="Weight (g)",main="")
> addSigLetters(lm1,lets=c("a","a","b","b"),pos=c(2,4,2,4),col=c("blue","blue","red","red"))
```



2.3 Checking the Assumptions

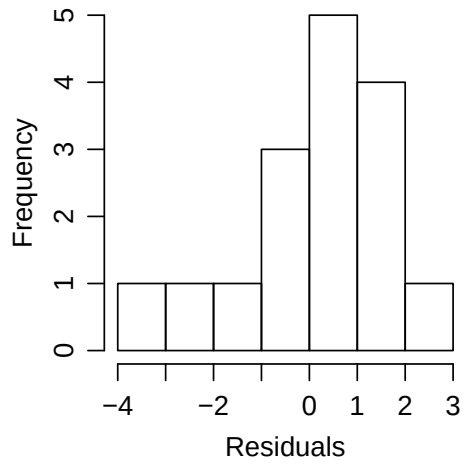
```
> leveneTest(lm1)

Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  3  0.3256 0.8069
      12
```

```
> residPlot(lm1)
```




```
> adTest(lm1$residuals)

        Anderson-Darling normality test

data:  lm1$residuals
A = 0.4308, p-value = 0.2688
```

```
> hist(lm1$residuals,xlab="Residuals",main="")
```




```
> outlierTest(lm1)

No Studentized residuals with Bonferonni p < 0.05
Largest |rstudent|:
   rstudent unadjusted p-value Bonferonni p
8 -2.836044           0.016196      0.25914
```

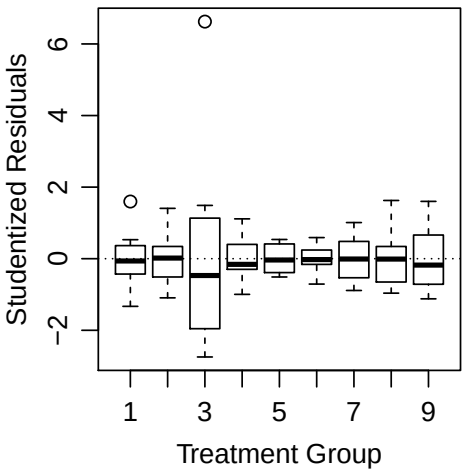
3 Benthic Infaunal Example

It is assumed that the initialization steps shown in Section 1 have already been followed and that the working directory has been changed to where the external data file is located.

```
> ben <- read.table("BenthicInfaunal.txt",head=TRUE)
> ben$fsite <- factor(ben$site)
> str(ben)

'data.frame':        72 obs. of  3 variables:
 $ site     : int  1 1 1 1 1 1 1 1 2 2 ...
 $ abundance: num  14.4 20.4 21.2 17.6 29 ...
 $ fsite    : Factor w/ 9 levels "1","2","3","4",..: 1 1 1 1 1 1 1 1 2 2 ...

> lm2 <- lm(abundance~fsite,data=ben)
```

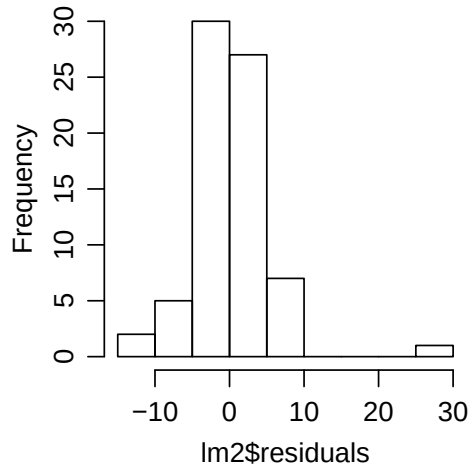
3.1 Assumption Checking with Possible Transformations

```
> leveneTest(lm2)

Levene's Test for Homogeneity of Variance (center = median)
      Df F value   Pr(>F)
group  8  3.2452 0.003726
      63

> residPlot(lm2)
```




```
> adTest(lm2$residuals)

        Anderson-Darling normality test

data:  lm2$residuals
A = 1.6389, p-value = 0.0002996
```

```
> hist(lm2$residuals,main="")
```



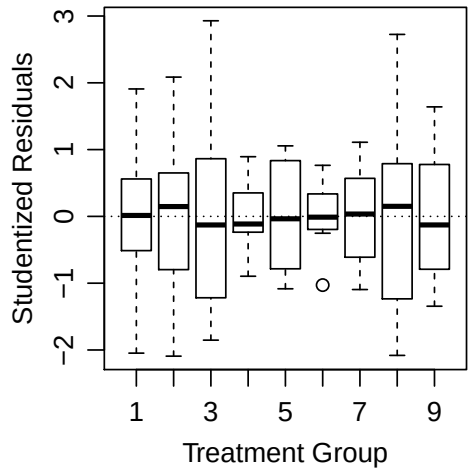

```
> outlierTest(lm2)

   rstudent unadjusted p-value Bonferonni p
20 6.624666         9.5554e-09   6.8799e-07
```

The following function was used to determine that a log transformation would most likely lead to the assumptions being met. This function cannot be illustrated in a handout because it requires interactions from the user.

```
> transChooser(lm2,show.stats=TRUE)
```

```
> ben$logab <- log(ben$abundance)
> lm3 <- lm(logab~fsite,data=ben)
> leveneTest(lm3)

Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  8  1.5339 0.1636
      63
```

```
> residPlot(lm3)
```



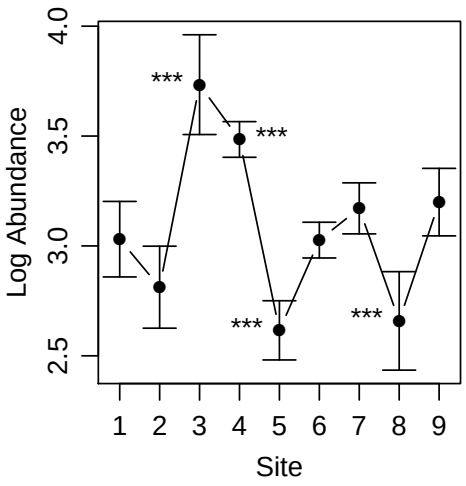

```
> adTest(lm3$residuals)

        Anderson-Darling normality test

data:  lm3$residuals
A = 0.3323, p-value = 0.5062

> outlierTest(lm3)

No Studentized residuals with Bonferonni p < 0.05
Largest |rstudent|:
   rstudent unadjusted p-value Bonferonni p
20 2.928889           0.004754      0.34229
```

3.2 Model Summarization

```
> anova(lm3)

Analysis of Variance Table

Response: logab
          Df Sum Sq Mean Sq F value    Pr(>F)
fsite      8 8.6683 1.08353  29.066 < 2.2e-16
Residuals 63 2.3485 0.03728

> ben.mc <- glht(lm3, mcp(fsite = "Dunnett"))
> summary(ben.mc)

         Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: Dunnett Contrasts


Fit: lm(formula = logab ~ fsite, data = ben)
```

```
Linear Hypotheses:
            Estimate Std. Error t value Pr(>|t|)
2 - 1 == 0 -0.218435   0.096537  -2.263  0.14558
3 - 1 == 0  0.703189   0.096537   7.284  < 0.001
4 - 1 == 0  0.453836   0.096537   4.701  < 0.001
5 - 1 == 0 -0.414859   0.096537  -4.297  < 0.001
6 - 1 == 0 -0.004238   0.096537  -0.044  1.00000
7 - 1 == 0  0.140280   0.096537   1.453  0.57978
8 - 1 == 0 -0.371867   0.096537  -3.852  0.00197
9 - 1 == 0  0.168668   0.096537   1.747  0.37988
(Adjusted p values reported -- single-step method)
```

```
> fitPlot(lm3,ylab="Log Abundance",xlab="Site",main="")
> addSigLetters(lm3,lets=c("","","***","***","***","","","***",""),pos=c(2,4,2,4,2,2,4,2,4))
```




```
> logdiff <- confint(ben.mc)$confint
> logdiff
```

```
         Estimate         lwr        upr
2 - 1 -0.21843454 -0.48181178  0.0449427
3 - 1  0.70318863  0.43981139  0.9665659
4 - 1  0.45383639  0.19045915  0.7172136
5 - 1 -0.41485933 -0.67823657 -0.1514821
6 - 1 -0.00423765 -0.26761489  0.2591396
7 - 1  0.14028047 -0.12309677  0.4036577
8 - 1 -0.37186732 -0.63524456 -0.1084901
9 - 1  0.16866808 -0.09470916  0.4320453
attr(,"conf.level")
[1] 0.95
attr(,"calpha")
[1] 2.728245
attr(,"error")
[1] 8.127829e-05
```

```
> exp(logdiff)
```

```
       Estimate       lwr       upr
2 - 1 0.8037761 0.6176633 1.0459679
3 - 1 2.0201841 1.5524144 2.6289010
4 - 1 1.5743404 1.2098050 2.0487168
5 - 1 0.6604332 0.5075112 0.8594333
6 - 1 0.9957713 0.7652024 1.2958147
7 - 1 1.1505965 0.8841781 1.4972914
8 - 1 0.6894457 0.5298059 0.8971878
9 - 1 1.1837272 0.9096375 1.5404049
attr(,"conf.level")
[1] 0.95
attr(,"calpha")
[1] 2.728245
attr(,"error")
[1] 8.127829e-05
```