One-Way Indicator Variable Regression Handout

1 Initialization

```
> library(NCStats)
```

2 Salmon Sperm Example

2.1 Data Preparation

You must change the directory to where the following file is located. I also removed three outliers as discussed in the simple linear regression handout.

```
> ss <- read.table("SalmonSperm.txt",head=TRUE)
> ss1 <- ss[-c(1,10,11),]
> str(ss1)

'data.frame':        11 obs. of  3 variables:
 $ step.len : num  2.94 3 3.02 3.17 3.18 3.2 3.27 3.31 3.72 3.84 ...
 $ fert.succ: num  3 2.2 7 7 13.5 10.4 6.7 12.8 37.8 30 ...
 $ mat      : Factor w/ 2 levels "Adult","Parr": 2 2 1 2 1 1 1 1 2 2 ...

> xlbl <- "Sperm Tail End Piece Length (um)"
> ylbl <- "Fertilization Success"
```

2.2 Fitting the Linear Model

```
> lm1 <- lm(fert.succ~step.len*mat,data=ss1)
> residPlot(lm1,legend="topleft")
```
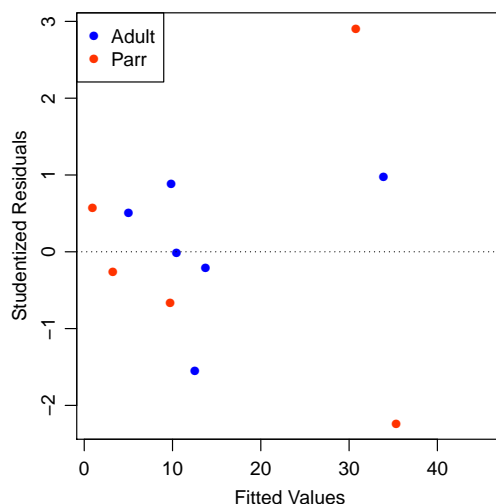
```
> adTest(lm1$residuals)

        Anderson-Darling normality test

data:  lm1$residuals
A = 0.1784, p-value = 0.8932
```

2.3   Model Exploration and Summarization

```
> summary(lm1)

Call:
lm(formula = fert.succ ~ step.len * mat, data = ss1)

Residuals:
    Min      1Q  Median      3Q     Max
-5.8463 -1.8829 -0.0417  2.0147  7.0466

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      -85.769     20.266  -4.232  0.00388
step.len          30.066      6.066   4.956  0.00164
matParr          -25.661     27.273  -0.941  0.37809
step.len:matParr   8.155      8.148   1.001  0.35022

Residual standard error: 4.547 on 7 degrees of freedom
Multiple R-squared: 0.9139,        Adjusted R-squared: 0.877
F-statistic: 24.78 on 3 and 7 DF,  p-value: 0.0004207

> confint(lm1)

                     2.5 %     97.5 %
(Intercept)      -133.68997 -37.84811
step.len           15.72132  44.41036
matParr           -90.15097  38.82930
step.len:matParr  -11.11233  27.42328

> fitPlot(lm1,interval="c",xlab=xlbl,ylab=ylbl,legend="topleft")
```
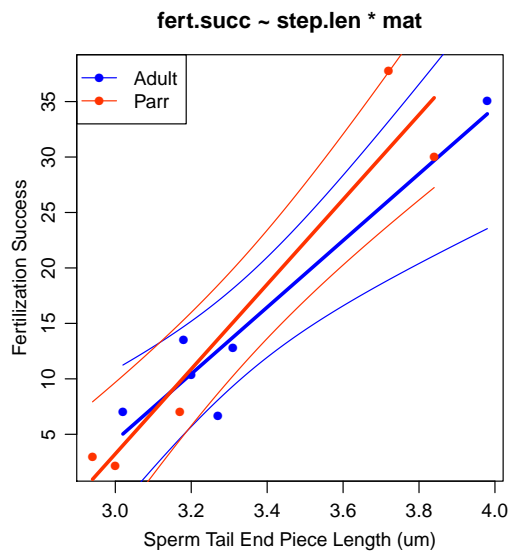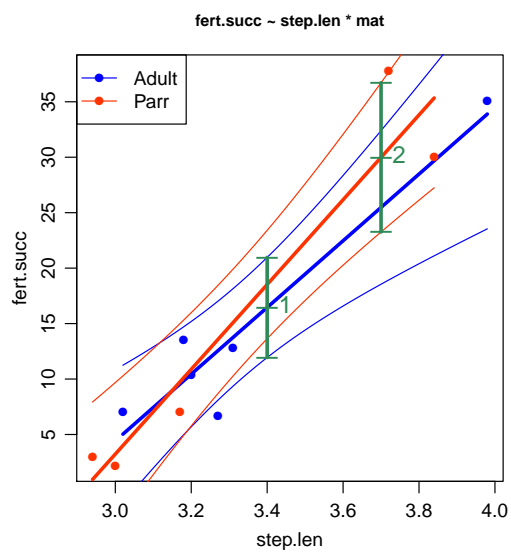
fert.succ ~ step.len * mat


```
> nd <- data.frame(step.len=c(3.4,3.7),mat=c("Adult","Parr"))
> predictionPlot(lm1,nd,interval="c",legend="topleft")

  obs step.len   mat      fit      lwr      upr
1   1      3.4 Adult 16.45483 11.94067 20.96899
2   2      3.7  Parr 29.98900 23.25902 36.71899
```

fert.succ ~ step.len * mat

3

2.4   ANOVA Demonstration

```
> anova(lm1)

Analysis of Variance Table

Response: fert.succ
             Df  Sum Sq Mean Sq F value    Pr(>F)
step.len      1 1510.23 1510.23 73.0316 5.966e-05
mat           1    6.11    6.11  0.2953    0.6037
step.len:mat  1   20.72   20.72  1.0017    0.3502
Residuals     7  144.75   20.68

> lm2 <- lm(fert.succ~step.len+mat,data=ss1)
> anova(lm2)

Analysis of Variance Table

Response: fert.succ
          Df  Sum Sq Mean Sq F value    Pr(>F)
step.len   1 1510.23 1510.23 73.0157 2.709e-05
mat        1    6.11    6.11  0.2952    0.6017
Residuals  8  165.47   20.68

> lm3 <- lm(fert.succ~step.len,data=ss1)
> anova(lm3)

Analysis of Variance Table

Response: fert.succ
          Df  Sum Sq Mean Sq F value   Pr(>F)
step.len   1 1510.23 1510.23   79.22 9.35e-06
Residuals  9  171.58   19.06

> anova(lm3,lm2,lm1)

Analysis of Variance Table

Model 1: fert.succ ~ step.len
Model 2: fert.succ ~ step.len + mat
Model 3: fert.succ ~ step.len * mat
  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1      9 171.57
2      8 165.47  1     6.106 0.2953 0.6037
3      7 144.75  1    20.715 1.0017 0.3502
```

3 Fish Energy Density Example

3.1 Data Preparation

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

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

'data.frame':       64 obs. of  3 variables:
 $ species: Factor w/ 4 levels "bayanchovy","bluefish",..: 2 2 2 2 2 2 2 2 2 2 ...
 $ dw     : int  39 34 34 32 31 30 30 29 26 25 ...
 $ ed     : int  10000 9000 8500 8100 7500 7100 7700 6100 5900 5500 ...
```

3.2 Assumption Checking and Diagnostics

```
> lm1 <- lm(ed~dw*species,data=FED)
> residPlot(lm1)
```
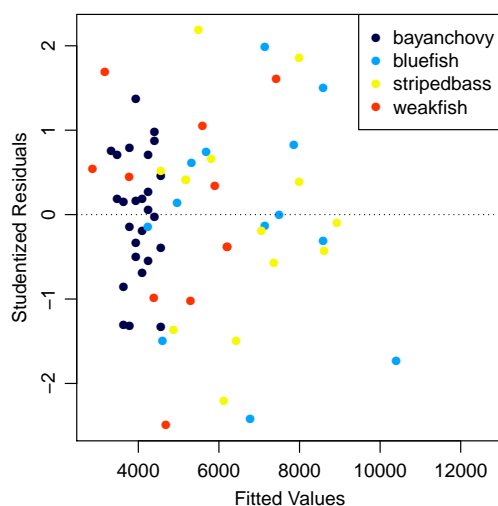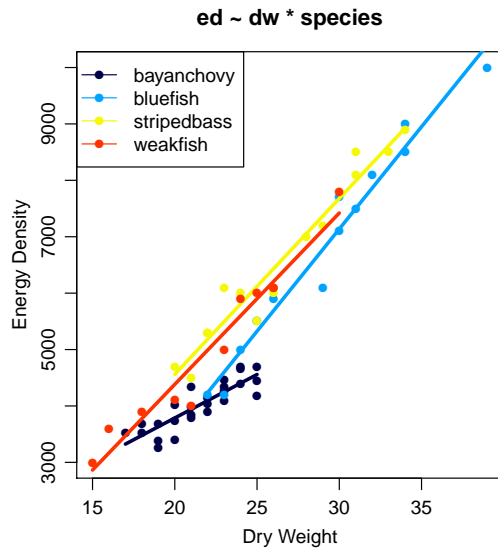


```
> adTest(lm1$residuals)

        Anderson-Darling normality test

data:  lm1$residuals
A = 0.353, p-value = 0.4549
```

```
> fitPlot(lm1,xlab="Dry Weight",ylab="Energy Density",legend="topleft")
```

ed ~ dw * species

3.3 Model Exploration and Summarization

```
> anova(lm1)

Analysis of Variance Table

Response: ed
            Df    Sum Sq   Mean Sq F value    Pr(>F)
dw           1 170693154 170693154 1858.966 < 2.2e-16
species      3  10592036   3530679   38.452 1.258e-13
dw:species   3   4105617   1368539   14.904 3.002e-07
Residuals   56   5142008     91822

> compSlopes(lm1)

Multiple comparison control procedures used: fdr

Multiple Slope Comparisons
              comparison     diff      lwr       upr   raw.p    adj.p
1     bluefish-bayanchovy 208.3095  145.628 270.99117 0.00000 0.00000
2 stripedbass-bayanchovy 157.6206   94.298 220.94336 0.00001 0.00003
3    weakfish-bayanchovy 149.6038   83.209 215.99852 0.00003 0.00006
4    stripedbass-bluefish -50.6890 -101.086  -0.29227 0.04873 0.05848
5       weakfish-bluefish -58.7057 -112.912  -4.49955 0.03430 0.05145
6   weakfish-stripedbass  -8.0168  -62.963  46.92958 0.77116 0.77116

Slope Information
        level slopes    lwr    upr raw.p adj.p
1  bayanchovy 154.19 102.23 206.15     0     0
4    weakfish 303.79 262.46 345.13     0     0
3 stripedbass 311.81 275.61 348.01     0     0
2    bluefish 362.50 327.44 397.56     0     0


> FED1 <- Subset(FED,species!="bayanchovy")
> lm2 <- lm(ed~dw*species,data=FED1)
> anova(lm2)
```

```
Analysis of Variance Table

Response: ed
          Df    Sum Sq   Mean Sq  F value    Pr(>F)
dw         1 104962559 104962559 838.6208 < 2.2e-16
species    2   2583571   1291785  10.3210 0.0003476
dw:species 2    556352    278176   2.2225 0.1247887
Residuals 32   4005150    125161

> compIntercepts(lm2)

Tukey HSD on adjusted means assuming parallel lines.
              comparison    diff     lwr     upr     p.adj
1 stripedbass-bluefish    631.40  291.11  971.69 0.00018433
2    weakfish-bluefish    506.47  144.53  868.42 0.00441609
3 weakfish-stripedbass  -124.92 -480.89  231.05 0.66939670

Mean adjusted values at a covariate value of 26.11
   bluefish stripedbass    weakfish
     5828.7      6460.1      6335.1


> fitPlot(lm2,xlab="Dry Weight",ylab="Energy Density",legend="topleft")
```


ed ~ dw * species