

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.

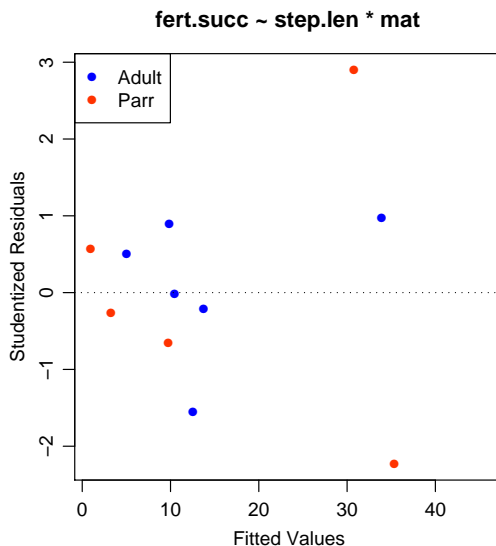
```
> ss <- read.table("SalmonSperm.txt",head=TRUE)
> ss1 <- ss[-c(1,10,11),]
> str(ss1)

'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 ...

> attach(ss1)
> xlbl <- "Sperm Tail End Piece Length (um)"
> ylbl <- "Fertilization Success"
```

2.2 Fitting the Linear Model

```
> lm1 <- lm(fert.succ~step.len*mat)
> residual.plot(lm1,legend="topleft")
```



```
> ncv.test(lm1)

Non-constant Variance Score Test
Variance formula: ~ fitted.values
Chisquare = 2.186415    Df = 1    p = 0.1392331
```

```
> ad.test(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)
```

```
Residuals:
```

```
      Min       1Q   Median       3Q      Max
-5.84627 -1.88289 -0.04166  2.01470  7.04657
```

```
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
```

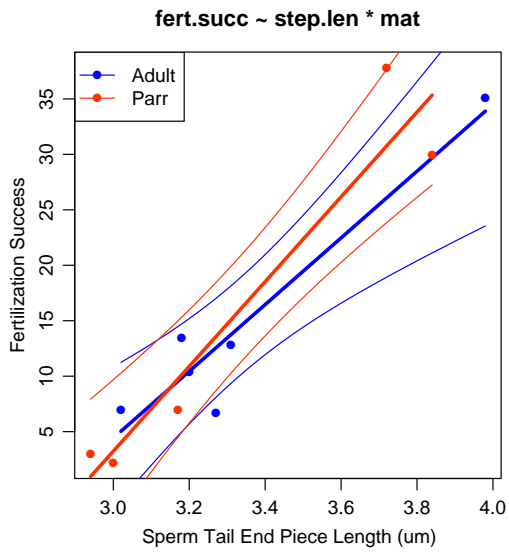
```
Multiple R-squared: 0.9139, Adjusted R-squared: 0.877
```

```
F-statistic: 24.78 on 3 and 7 DF, p-value: 0.0004207
```

```
> confint(lm1)
```

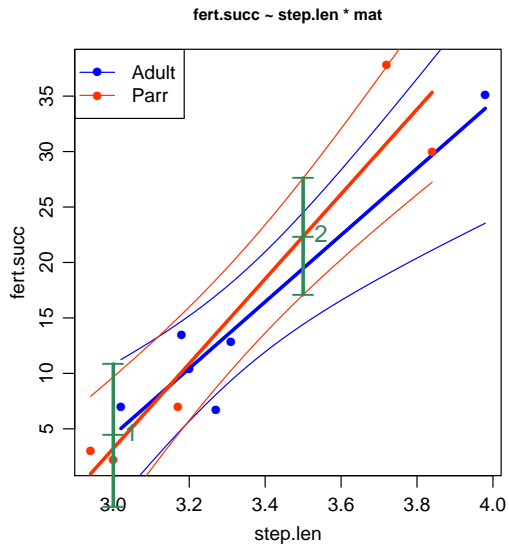
```
              2.5 %    97.5 %
(Intercept) -133.68997 -37.84811
step.len     15.72132  44.41036
matParr     -90.15097  38.82930
step.len:matParr -11.11233  27.42328
```

```
> fit.plot(lm1, interval="c", xlab=xlbl, ylab=ylbl, legend="topleft")
```



```
> nd <- data.frame(step.len=c(3,3.5),mat=c("Adult", "Parr"))
> prediction.plot(lm1,nd,interval="c",legend="topleft")
```

obs	step.len	mat	fit	lwr	upr
1	1	3.0 Adult	4.428491	-1.992451	10.84943
2	2	3.5 Parr	22.344739	17.083054	27.60642



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		
Total	10	1681.81			

```
> lm2 <- lm(fert.succ~step.len+mat)
```

```
> 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		
Total	10	1681.81			

```
> lm3 <- lm(fert.succ~step.len)
```

```
> 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		
Total	10	1681.81			

```
> 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	171.575				
2	8	165.469	1	6.106	0.2953	0.6037
3	7	144.754	1	20.715	1.0017	0.3502

```
> detach(ss1)
```

3 Fish Energy Density Example

3.1 Data Preparation

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

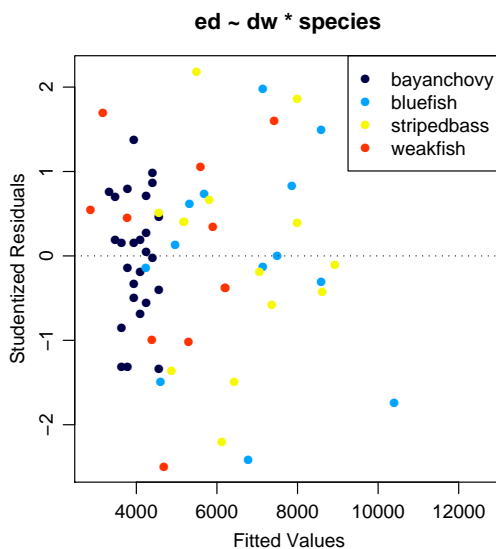
```
> FED <- read.table("FishEnergyDensity.txt",head=TRUE)
> str(FED)

'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 ...

> attach(FED)
```

3.2 Assumption Checking and Diagnostics

```
> lm1 <- lm(ed~dw*species)
> residual.plot(lm1)
```



```
> ncv.test(lm1)
```

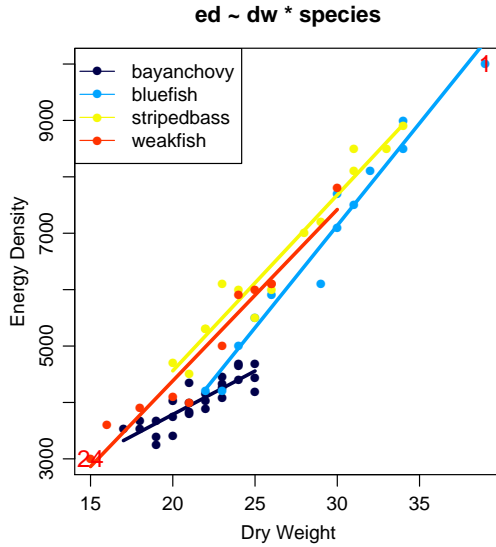
```
Non-constant Variance Score Test
Variance formula: ~ fitted.values
Chisquare = 1.728625   Df = 1   p = 0.1885866
```

```
> ad.test(lm1$residuals)
```

```
Anderson-Darling normality test
```

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

```
> fit.plot(lm1,xlab="Dry Weight",ylab="Energy Density",legend="topleft")
> highlight(ed~dw,pts=c(1,24))
```



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		
Total	63	190532816			

```
> comp.slopes(lm1)
```

Multiple comparison control procedures used: fdr

Multiple Slope Comparisons

	comparison	diff	lwr	upr	raw.p	adj.p
1	bluefish-bayanchovy	208.30952	145.62788	270.99117	0.00000	0.00000
2	stripedbass-bayanchovy	157.62055	94.29774	220.94336	0.00001	0.00003
3	weakfish-bayanchovy	149.60379	83.20906	215.99852	0.00003	0.00006
4	stripedbass-bluefish	-50.68898	-101.08569	-0.29227	0.04873	0.05848
5	weakfish-bluefish	-58.70573	-112.91192	-4.49955	0.03430	0.05145
6	weakfish-stripedbass	-8.01676	-62.96310	46.92958	0.77116	0.77116

Slope Information

	level	slopes	lwr	upr	raw.p	adj.p
1	bayanchovy	154.1905	102.2341	206.1469	0	0
4	weakfish	303.7943	262.4567	345.1318	0	0
3	stripedbass	311.8110	275.6128	348.0092	0	0
2	bluefish	362.5000	327.4355	397.5645	0	0

```

> detach(FED)
> FED1 <- Subset(FED,species!="bayanchovy")
> attach(FED1)
> lm2 <- lm(ed~dw*species)
> 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		
Total	37	112107632			

```

> comp.intercepts(lm2)

```

Tukey HSD on adjusted means assuming parallel lines.

	comparison	diff	lwr	upr	p.adj
1	stripedbass-bluefish	631.3980	291.1082	971.6878	0.0001843341
2	weakfish-bluefish	506.4749	144.5318	868.4180	0.0044160914
3	weakfish-stripedbass	-124.9231	-480.8922	231.0461	0.6693967026

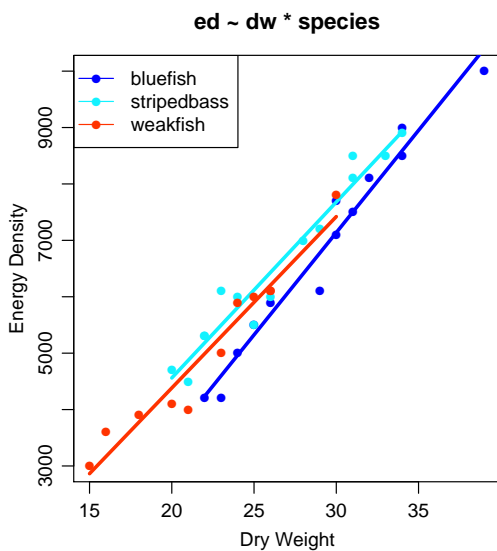
Mean adjusted values at a covariate value of 26.1052631578947

	bluefish	stripedbass	weakfish
	5828.663	6460.061	6335.138

```

> fit.plot(lm2,xlab="Dry Weight",ylab="Energy Density",legend="topleft")

```



```

> detach(FED1)

```