

Indicator Variable Regression

R Handout

Derek H. Ogle

Laysan Teal Monitoring

Monitoring the abundance of animal populations can be expensive. Researchers often attempt to find methods for accurately predicting abundance from simple and inexpensive metrics. In one study, Reynolds *et al.* (2017) compared the abundance of Laysan Teal (*Anas laysanensis*) as estimated from the time-intensive Lincoln-Petersen mark-resight method to the maximum annual observed count of Laysan Teal made bi-monthly on Midway Atoll. They wanted to determine if there was a significantly positive relationship between the mark-resight abundance estimate and the maximum annual count, and if that relationship differed between the breeding and non-breeding seasons.

```
> options(show.signif.stars=FALSE)
> library(NCStats)
> setwd("C:/aaaWork/Web/GitHub/NCMTH207/modules/IVRegression")
```

```
> df <- read.csv("LaysanTeal.csv")
> str(df)
```

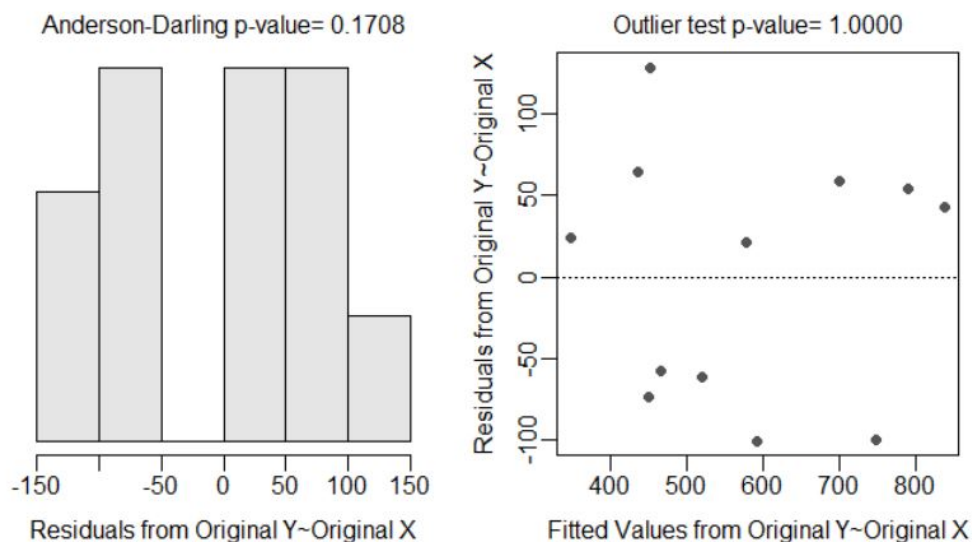
```
'data.frame': 12 obs. of 3 variables:
 $ max.count : int 75 130 281 309 340 213 274 284 276 361 ...
 $ mr.estimate: int 502 459 647 844 879 374 376 409 582 492 ...
 $ season : Factor w/ 2 levels "breeding","non-breeding": 1 1 1 1 1 2 2 2 2 2 ...
```

```
> xlbl <- "Maximum Count"
> ylbl <- "Lincoln-Petersen Abundance"
```

Fitting the Linear Model

```
> lm1 <- lm(mr.estimate~max.count*season,data=df)
```

```
> transChooser(lm1)
```



Lecture Support – ANOVA Demonstration

```
> lm1 <- lm(mr.estimate~max.count*season,data=df)
> anova(lm1)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
max.count	1	83421	83421	10.6370	0.011503
season	1	190750	190750	24.3225	0.001147
max.count:season	1	359	359	0.0457	0.835991
Residuals	8	62740	7843		

```
> lm2 <- lm(mr.estimate~max.count+season,data=df)
> anova(lm2)
```

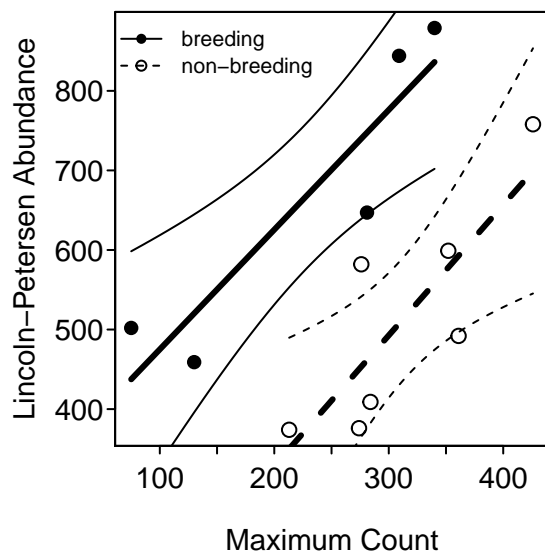
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
max.count	1	83421	83421	11.899	0.0072813
season	1	190750	190750	27.207	0.0005521
Residuals	9	63099	7011		

```
> lm3 <- lm(mr.estimate~max.count,data=df)
> anova(lm3)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
max.count	1	83421	83421	3.2862	0.09994
Residuals	10	253849	25385		

Model Exploration and Summarization

```
> fitPlot(lm1,interval="confidence",xlab=xlbl,ylab=ybl,legend="topleft",cex.legend=0.7)
```



```
> cbind(Ests=coef(lm1), confint(lm1))
```

```

              Ests      2.5 %      97.5 %
(Intercept)  324.6129857  106.7406493  542.485322
max.count    1.5047886    0.6333927    2.376185
seasonnon-breeding -323.9985782 -757.0781733  109.081017
max.count:seasonnon-breeding  0.1355129  -1.3255575  1.596583

```

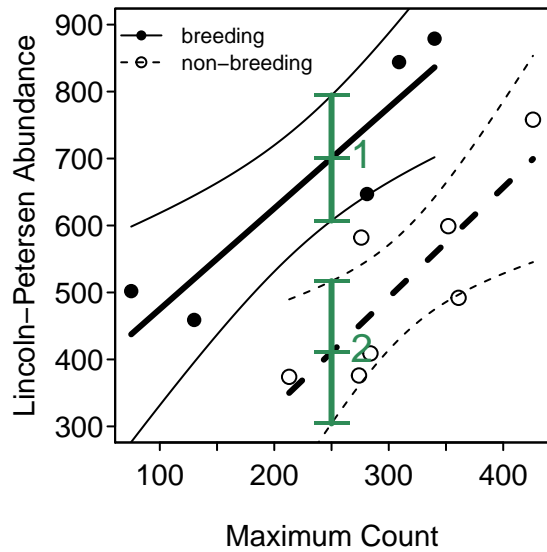
```
> ( nd <- data.frame(max.count=c(250,250), season=c("breeding","non-breeding")) )
```

```

max.count  season
1         250  breeding
2         250 non-breeding

```

```
> predictionPlot(lm1,nd,interval="confidence",xlab=xlbl,ylab=ylbl,ylim=c(300,900),
  legend="topleft",cex.leg=0.7)
```



```

obs max.count  season  fit  lwr  upr
1   1      250  breeding 700.8101 607.3089 794.3114
2   2      250 non-breeding 410.6898 304.4186 516.9610

```

Fish Energy Density

Hartman and Brandt (1995) examined the relationship between energy density (J/g wet weight) and percent dry weight for four species of fish – Bay Anchovy (*Anchoa mitchilli*), Bluefish (*Pomatomus saltatrix*), Striped Bass (*Morone saxatilis*), and Weakfish (*Cynoscion regalis*). They wanted to describe the relationship between these two variables and determine if there were any differences among species.

```
> FED <- read.csv("FishEnergyDensity.csv")
> 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 ...
```

```
> FED$ed2 <- FED$ed/1000
```

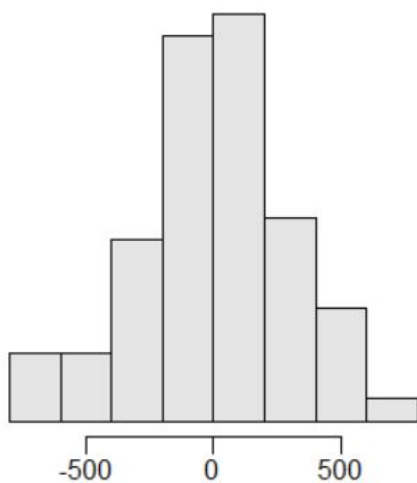
Model Fitting and Analysis

```
> lm1 <- lm(ed2~dw*species,data=FED)
```

```
> transChooser(lm1)
```

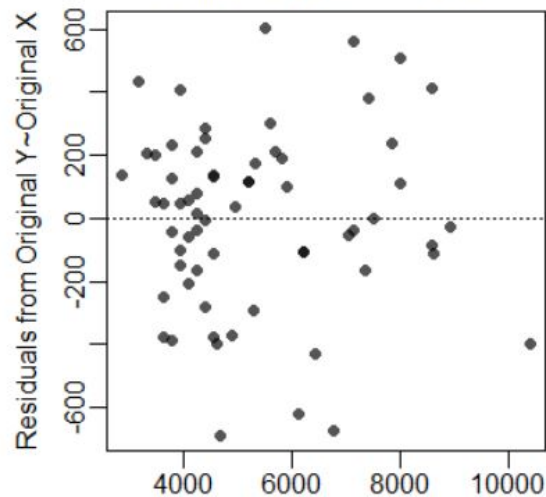


Anderson-Darling p-value= 0.4549



Residuals from Original Y~Original X

Outlier test p-value= 1.0000



Fitted Values from Original Y~Original X

```
> anova(lm1)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
dw	1	170.693	170.693	1858.966	< 2.2e-16
species	3	10.592	3.531	38.452	1.258e-13
dw:species	3	4.106	1.369	14.904	3.002e-07
Residuals	56	5.142	0.092		

```
> compSlopes(lm1)
```

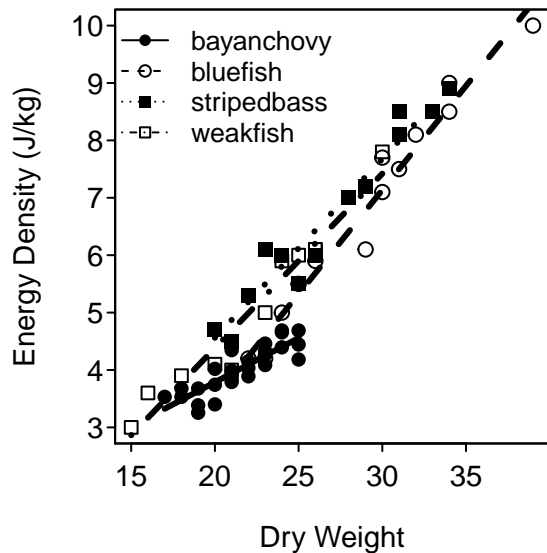
Multiple Slope Comparisons (using the 'holm' adjustment)

	comparison	diff	95% LCI	95% UCI	p.unadj	p.adj
1	bluefish-bayanchovy	0.20831	0.14563	0.27099	0.00000	0.00000
2	stripedbass-bayanchovy	0.15762	0.09430	0.22094	0.00001	0.00005
3	weakfish-bayanchovy	0.14960	0.08321	0.21600	0.00003	0.00012
4	stripedbass-bluefish	-0.05069	-0.10109	-0.00029	0.04873	0.10290
5	weakfish-bluefish	-0.05871	-0.11291	-0.00450	0.03430	0.10290
6	weakfish-stripedbass	-0.00802	-0.06296	0.04693	0.77116	0.77116

Slope Information (using the 'holm' adjustment)

	level	slopes	95% LCI	95% UCI	p.unadj	p.adj
1	bayanchovy	0.15419	0.10223	0.20615	0	0
4	weakfish	0.30379	0.26246	0.34513	0	0
3	stripedbass	0.31181	0.27561	0.34801	0	0
2	bluefish	0.36250	0.32744	0.39756	0	0

```
> fitPlot(lm1,xlab="Dry Weight",ylab="Energy Density (J/kg)",legend="topleft",cex.legend=0.8)
```



Follow-Up Model Fitting and Analysis

```
> FED1 <- filterD(FED,species!="bayanchovy")
> lm2 <- lm(ed2~dw*species,data=FED1)
> anova(lm2)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
dw	1	104.963	104.963	838.6208	< 2.2e-16
species	2	2.584	1.292	10.3210	0.0003476
dw:species	2	0.556	0.278	2.2225	0.1247887
Residuals	32	4.005	0.125		

```
> compIntercepts(lm2)
```

Warning: Removed an interaction term from 'mdl' (i.e., assumed parallel lines) to test intercepts.

Tukey HSD on means adjusted assuming parallel lines

	comparison	diff	95% LCI	95% UCI	p.adj
1	stripedbass-bluefish	0.6313980	0.2911082	0.9716878	0.0001843341
2	weakfish-bluefish	0.5064749	0.1445318	0.8684180	0.0044160914
3	weakfish-stripedbass	-0.1249231	-0.4808922	0.2310461	0.6693967026

Mean ed2 when dw=26.10526

bluefish	stripedbass	weakfish
5.828663	6.460061	6.335138

```
> compIntercepts(lm2,common.cov=0)
```

Warning: Removed an interaction term from 'mdl' (i.e., assumed parallel lines) to test intercepts.

Tukey HSD on means adjusted assuming parallel lines

	comparison	diff	95% LCI	95% UCI	p.adj
1	stripedbass-bluefish	0.6313980	0.2911082	0.9716878	0.0001843341
2	weakfish-bluefish	0.5064749	0.1445318	0.8684180	0.0044160914
3	weakfish-stripedbass	-0.1249231	-0.4808922	0.2310461	0.6693967026

Mean ed2 when dw=0

bluefish	stripedbass	weakfish
-2.752451	-2.121053	-2.245976

```
> fitPlot(lm2,xlab="Dry Weight",ylab="Energy Density (J/kg)",legend="topleft",cex.legend=0.8)
```

