Simple Linear Regression

R Handout

Derek H. Ogle

Oak Gall Wasps

Prior and Hellman (2010) investigated the impact of invasive Oak Gall Wasps (*Neuroterus saltatorius*) on a native butterfly. They hoped to use the amount of damage shown on oak leaves as a surrogate for the density of the wasps' galls. To examine this they recorded the density of wasp galls (number per leaf) and the percentage leaf damage for ten trees. Is leaf damage a reliable predictor of wasp gall density?

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

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

```
'data.frame':   10 obs. of  2 variables:
 $ damage : num  6.9 9.2 15.6 18.3 20.9 28 43.1 42.9 52 40.9
 $ density: int  163 326 296 459 115 936 610 1196 1262 1631
```

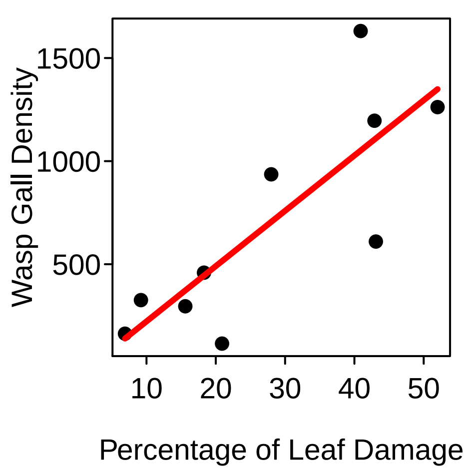
```
> xlbl <- "Percentage of Leaf Damage"
> ylbl <- "Wasp Gall Density"
```

Lecture Support I – Model Fitting and Simple Predictions

```
> ( lm1 <- lm(density~damage,data=df) )
```

```
Coefficients:
(Intercept)       damage
     -45.70        26.82
```

```
> fitPlot(lm1,xlab=xlbl,ylab=ylbl)
```



```
> predict(lm1,data.frame(damage=35))
```

```
       1
893.0516
```

Lecture Support II – Sampling Variability

```
> summary(lm1)
```

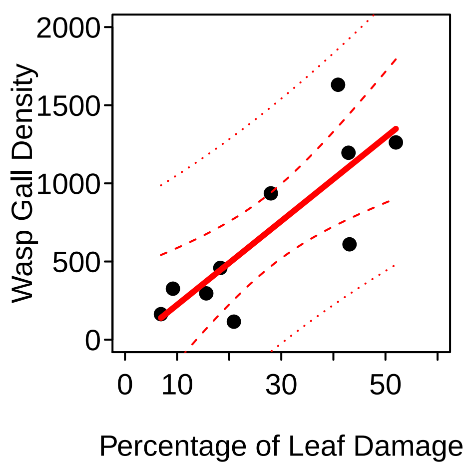
```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -45.703    213.941  -0.214  0.83618
damage        26.822      6.763   3.966  0.00414

Residual standard error: 323.6 on 8 degrees of freedom
Multiple R-squared: 0.6628, Adjusted R-squared: 0.6207
F-statistic: 15.73 on 1 and 8 DF,  p-value: 0.004144
```

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

```
                 Ests      2.5 %    97.5 %
(Intercept) -45.70283 -539.05161 447.64595
damage       26.82156   11.22529  42.41783
```

```
> fitPlot(lm1,ylab=ylbl,xlab=xlbl,ylim=c(0,2000),xlim=c(0,60),interval="both")
```



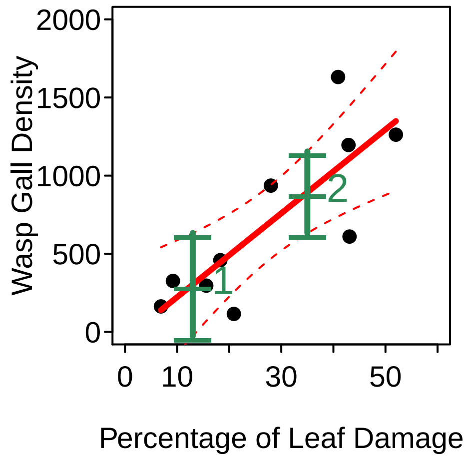

```
> predict(lm1,data.frame(damage=35),interval="confidence")
```

```
       fit     lwr      upr
1 893.0516 631.604 1154.499
```

```
> predict(lm1,data.frame(damage=35),interval="prediction")
```

```
       fit      lwr      upr
1 893.0516 102.4161 1683.687
```

```
> predictionPlot(lm1,data.frame(damage=c(13,35)),ylab=ylbl,xlab=xlbl,
                 ylim=c(0,2000),xlim=c(0,60),interval="confidence")
```




```
  obs damage      fit       lwr       upr
1   1     13 302.9774 -26.88797  632.8428
2   2     35 893.0516 631.60398 1154.4993
```

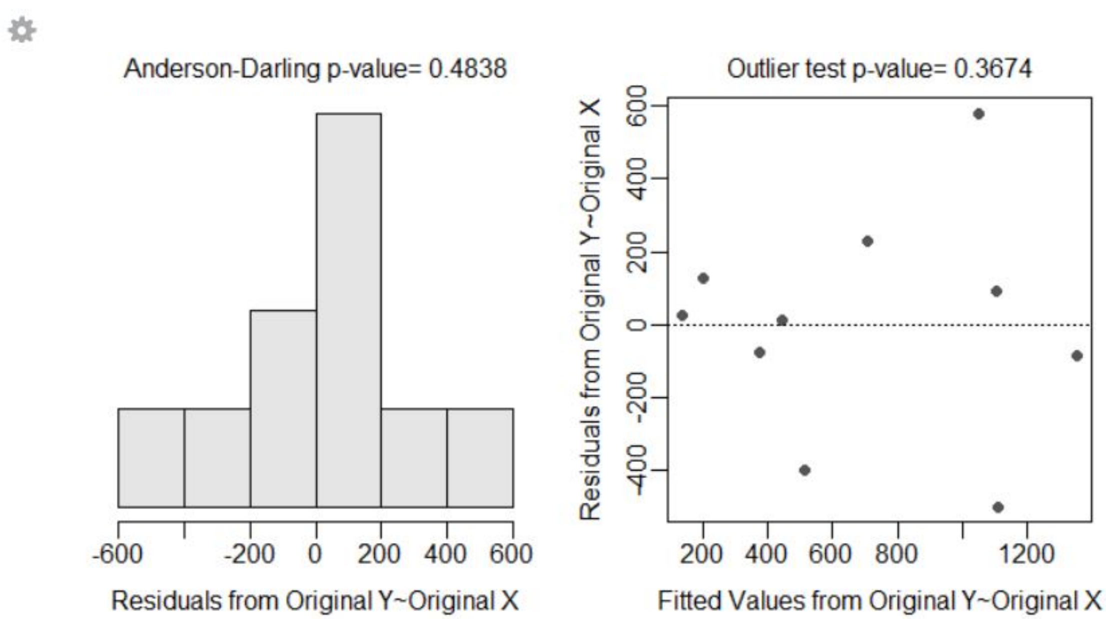
Lecture Support III – Model Comparisons

```
> anova(lm1)
```

```
          Df  Sum Sq Mean Sq F value   Pr(>F)
damage     1 1646594 1646594  15.727 0.004144
Residuals  8  837587  104698
```

Lecture Support IV – Assumption Checking

```
> transChooser(lm1)
```



Petrels

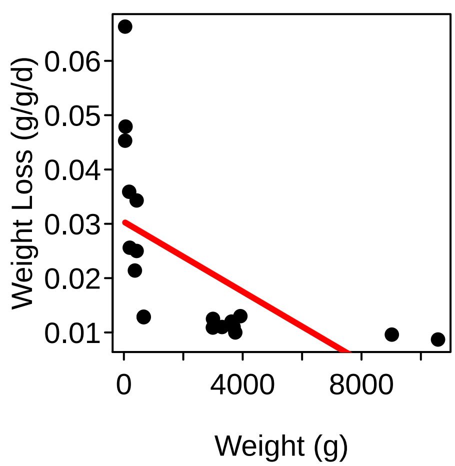
Croxall (1982) examined the weight loss of adult petrels during periods of egg incubation. He examined 13 species but some had measurements for both sexes such that 19 measurements are found in `Petrels.csv`. For each measurement the mean initial weight (g) and mean weight lost (g/g/d) were recorded. Determine if the mean initial weight significant explains variability in mean weight lost.

```
> petrels <- read.csv("Petrels.csv")
> str(petrels)
```

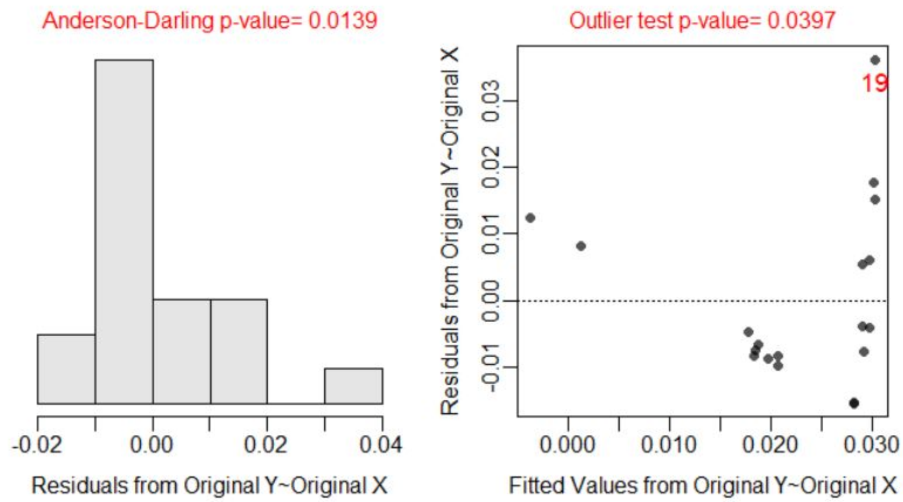
```
'data.frame':   19 obs. of  4 variables:
 $ species    : Factor w/ 13 levels "Diomedea chrysostoma",..: 2 2 4 4 1 1 3 3 3 9 ...
 $ sex        : Factor w/ 4 levels "both","female",..: 3 2 3 2 3 2 3 2 1 3 ...
 $ weight     : int  10577 9022 3922 3694 3751 3624 3305 3000 2996 668 ...
 $ weight.loss: num  0.0087 0.0096 0.013 0.011 0.01 0.012 0.011 0.0125 0.0109 0.0128 ...
```

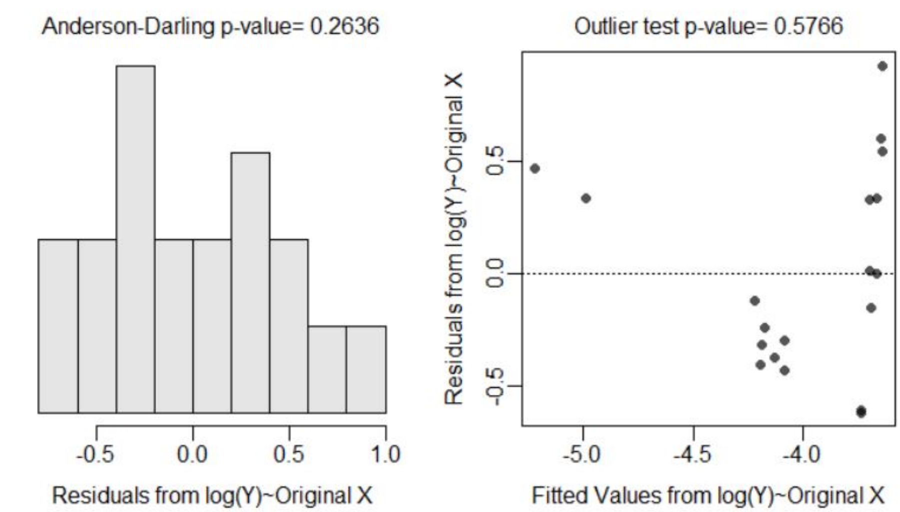
Analysis

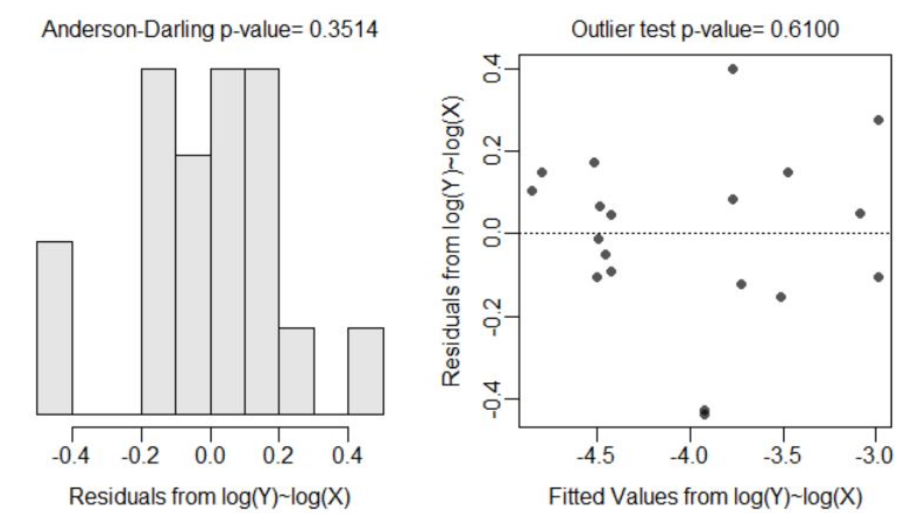
```
> lm1 <- lm(weight.loss~weight,data=petrels)
> fitPlot(lm1,xlab="Weight (g)",ylab="Weight Loss (g/g/d)")
```



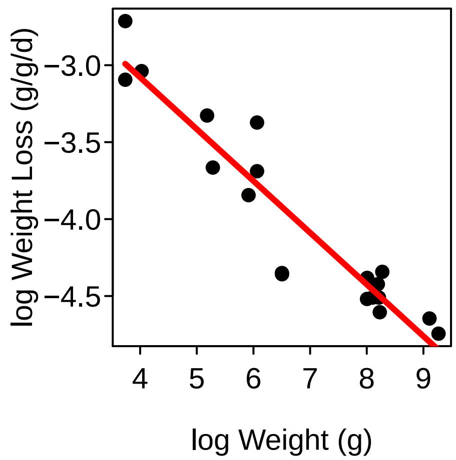
```
> transChooser(lm1)
```







```r
> petrels$log.wt <- log(petrels$weight)
> petrels$log.wtloss <- log(petrels$weight.loss)
> lm2 <- lm(log.wtloss~log.wt,data=petrels)
> fitPlot(lm2,xlab="log Weight (g)",ylab="log Weight Loss (g/g/d)")
```




```r
> anova(lm2)
```

```
          Df Sum Sq Mean Sq F value    Pr(>F)
log.wt     1 6.5113  6.5113  140.65 1.204e-09
Residuals 17 0.7870  0.0463
```

```r
> summary(lm2)
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.73403    0.19792  -8.761 1.04e-07
log.wt      -0.33632    0.02836 -11.860 1.20e-09

Residual standard error: 0.2152 on 17 degrees of freedom
Multiple R-squared: 0.8922, Adjusted R-squared: 0.8858
F-statistic: 140.6 on 1 and 17 DF,  p-value: 1.204e-09
```

```r
> cbind(Ests=coef(lm2),confint(lm2))
```

```
                  Ests      2.5 %     97.5 %
(Intercept) -1.7340329 -2.1516113 -1.3164546
log.wt      -0.3363196 -0.3961507 -0.2764885
```

```r
> ( p.log.wtloss <- predict(lm2,data.frame(log.wt=log(5000)),interval="confidence") )
```

```
        fit       lwr       upr
1 -4.598532 -4.746569 -4.450495
```

```r
> exp(p.log.wtloss)*exp(anova(lm2)[2,3]/2)
```

```
         fit         lwr        upr
1 0.01030234 0.008884726 0.01194614
```