Math 158, Spring 2009
Jo Hardin
Multiple Regression V – R code
Residuals & Leverage

Consider the multiple regression model about the oak trees from Exam 1:

$$\begin{aligned} E[Y] &= \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_1 \cdot X_2 \\ Y &= \text{log range (ln km}^2 \times 100\text{)}, \ \mathit{ln.range} \\ X_1 &= \text{log size of acorn (ln cm}^3\text{)}, \ \mathit{ln.size} \\ X_2 &= \text{binary variable on 2 locations}, \ \mathit{location} \\ X_3 &= \text{Tree height (m)}, \ \mathit{Tree\_Height} \end{aligned}$$

```
> acorn <- read.table("acorns.txt", header=T, sep="\t")
> attach(acorn)
> ln.size <- log(Acorn_size)
> ln.range <- log(Range)
```

- ```
  > acorn.lm <- lm(ln.range ~ ln.size*location + Tree_Height)
  > summary(acorn.lm)

   lm(formula = ln.range ~ ln.size * location + Tree_Height)
                    Estimate Std. Error t value Pr(>|t|)
   (Intercept)       5.54826    0.52061  10.657 2.23e-12 ***
   ln.size          -0.12237    0.31058  -0.394   0.6960
   location          2.32115    0.51203   4.533 6.86e-05 ***
   Tree_Height       0.03422    0.01996   1.715   0.0954 .
   ln.size:location  0.81995    0.36705   2.234   0.0322 *

   Residual standard error: 0.9881 on 34 degrees of freedom
   Multiple R-squared: 0.709,      Adjusted R-squared: 0.6748
  ```

- How much does #39 change the results/signficance?

  ```
  > acorn39.lm<-lm(ln.range[-39]~ln.size[-39]*location[-39]+Tree_Height[-39])
  > summary(acorn39.lm)
  lm(formula = ln.range[-39] ~ ln.size[-39] * location[-39] + Tree_Height[-39])
                              Estimate Std. Error t value Pr(>|t|)
   (Intercept)                 5.63989    0.40907  13.787 3.01e-15 ***
   ln.size[-39]                0.19247    0.25276   0.761   0.4518
   location[-39]               2.31394    0.40188   5.758 1.97e-06 ***
   Tree_Height[-39]            0.02895    0.01570   1.844   0.0742 .
   ln.size[-39]:location[-39]  0.51534    0.29526   1.745   0.0902 .

   Residual standard error: 0.7756 on 33 degrees of freedom
   Multiple R-squared: 0.7607,     Adjusted R-squared: 0.7317
  ```
  ```

- How much do the predictions for # 39 change?

```
> acorn[39,]
                     Species     Region Range Acorn_size Tree_Height
39 Quercus tomentella Engelm California    13        7.1          18

> predict.lm(acorn.lm,data.frame(ln.size=log(7.1),location=0,Tree_Height=18),
             interval="pred")
       fit      lwr      upr
1 5.924453 3.766785 8.082121
> exp(predict.lm(acorn.lm,data.frame(ln.size=log(7.1),location=0,Tree_Height=18),
             interval="pred"))
       fit      lwr     upr
1 374.0738 43.24082 3236.09
> predict.lm(acorn.lm,data.frame(ln.size=log(7.1),location=0,Tree_Height=18),
             interval="conf")
       fit      lwr      upr
1 5.924453 5.135152 6.713754

> predict.lm(acorn39.lm,data.frame(ln.size=log(7.1),location=0,Tree_Height=18),

             interval="pred")
       fit      lwr      upr
1 6.538306 4.822318 8.254294
> exp(predict.lm(acorn39.lm,data.frame(ln.size=log(7.1),location=0,Tree_Height=18),
             interval="pred"))
      fit      lwr      upr
 1 691.115 124.2528 3844.098
> predict.lm(acorn39.lm,data.frame(ln.size=log(7.1),location=0,Tree_Height=18),
             interval="conf")
       fit      lwr      upr
1 6.538306 5.863828 7.212784
```

- Residual & Leverage plots:

```
> par(mfrow=c(2,3))
> plot(fitted(acorn.lm), rstudent(acorn.lm), xlab="fitted", ylab="studentized
         resid",pch=19,cex=.75)
> abline(h=0)
> plot(rstudent(acorn.lm) ~ hatvalues(acorn.lm),pch=19,cex=.75,
         xlab="leverage",ylab="studentized resid")
> abline(h=0)

> plot(cooks.distance(acorn.lm),pch=19,cex=.75,ylab="Cook's Distance",xlab="")
> plot(dffits(acorn.lm),pch=19,cex=.75,ylab="DFFITS",xlab="")
> plot(dfbeta(acorn.lm)[,5],pch=19,cex=.75,ylab="DFBETAS: interaction",xlab="")
> plot(dfbeta(acorn.lm)[,3],pch=19,cex=.75,ylab="DFBETAS: location",xlab="")
```