(Very) Quick Introduction to Linear Models in R

Derek H. Ogle, Northland College

5-Mar-2015

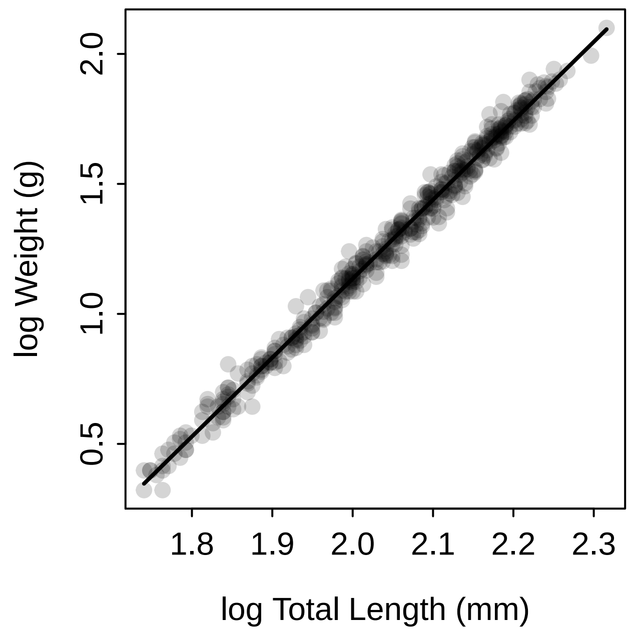
Preliminaries

```
> # clears objects in R workspace
> rm(list = ls())

> # load needed packages
> library(FSA)         # for headtail(), filterD(), fitPlot()
> library(dplyr)       # for mutate(),%>%
> library(multcomp)    # for glht(), mcp()

> ruf <- read.csv("RuffeSLRH.csv") %>%
   mutate(logW=log10(wt),logL=log10(tl),fYear=factor(year))
> headtail(ruf)
     fishID year month day  tl   wt      logW     logL fYear
1      1092 1988     6   1  71  6.0 0.7781513 1.851258  1988
2      1097 1988     6   1  74  6.0 0.7781513 1.869232  1988
3      1132 1988     6   1  75  6.0 0.7781513 1.875061  1988
9949     99 2007     9  20 115 17.9 1.2528530 2.060698  2007
9950     88 2007     9  20 120 18.6 1.2695129 2.079181  2007
9951     60 2007     9  20 134 24.6 1.3909351 2.127105  2007
```

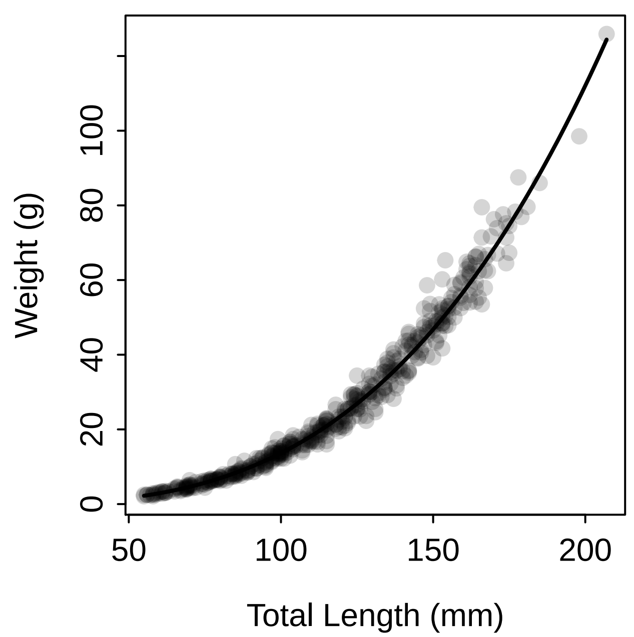
Simple Linear Regession

```
> ruf90 <- filterD(ruf,year==1990)
> fit1 <- lm(logW~logL,data=ruf90)
> coef(fit1)
(Intercept)        logL
  -4.936979    3.036223

> confint(fit1)
                2.5 %    97.5 %
(Intercept) -4.996232 -4.877726
logL         3.007323  3.065123

> anova(fit1)
           Df Sum Sq Mean Sq F value    Pr(>F)
logL        1 69.661  69.661   42628 < 2.2e-16
Residuals 455  0.744   0.002
Total     456 70.405

> ( tmp <- range(ruf90$logL) )
[1] 1.740363 2.315970

> xs <- seq(tmp[1],tmp[2],length.out=99)
> xs[1:10]
 [1] 1.740363 1.746236 1.752110 1.757983 1.763857 1.769730 1.775604 1.781478 1.787351 1.793225

> ys <- predict(fit1,data.frame(logL=xs))
> ys[1:10]
        1         2         3         4         5         6         7         8         9        10
0.3471506 0.3649840 0.3828174 0.4006508 0.4184842 0.4363176 0.4541510 0.4719844 0.4898178 0.5076512
```

```
> plot(logW~logL,data=ruf90,pch=19,col=rgb(0,0,0,1/6),ylab="log Weight (g)",xlab="log Total Length (mm)")
> lines(ys~xs,lwd=2)
```

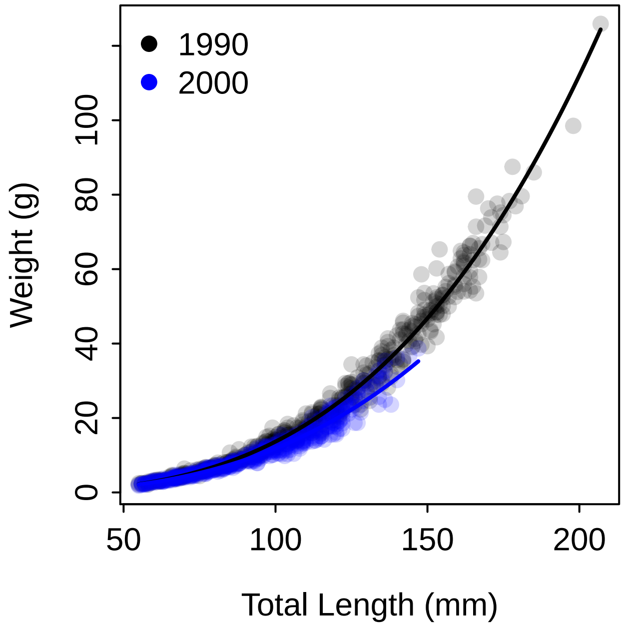



```
> plot(wt~tl,data=ruf90,pch=19,col=rgb(0,0,0,1/6),ylab="Weight (g)",xlab="Total Length (mm)")
> btxs <- 10^xs
> btys <- 10^ys
> lines(btys~btxs,lwd=2)
```



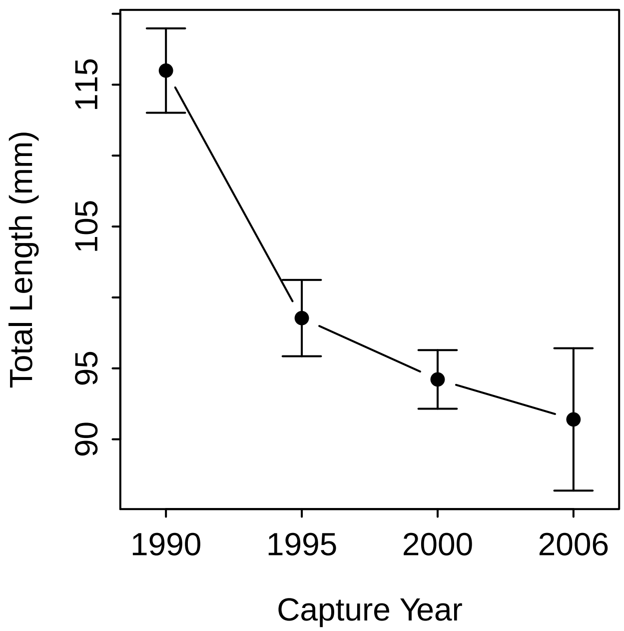
Dummy Variable Regession (aka ANCOVA)

```
> ruf9000 <- filterD(ruf,year %in% c(1990,2000))
> fit2 <- lm(logW~logL*fYear,data=ruf9000)
> anova(fit2)
            Df  Sum Sq Mean Sq   F value    Pr(>F)
logL         1 127.429 127.429 74533.553 < 2.2e-16
fYear        1   0.757   0.757   443.043 < 2.2e-16
logL:fYear   1   0.112   0.112    65.463 1.952e-15
Residuals  882   1.508   0.002
Total      885 129.807

> coef(fit2)
   (Intercept)           logL      fYear2000 logL:fYear2000
    -4.9369786      3.0362230      0.3319951     -0.1976742

> confint(fit2)
                    2.5 %     97.5 %
(Intercept)    -4.9975070 -4.8764502
logL            3.0067013  3.0657446
fYear2000       0.2362067  0.4277836
logL:fYear2000 -0.2456248 -0.1497235

> tmp <- ruf9000 %>% group_by(fYear) %>% summarize(min=min(tl,na.rm=TRUE),max=max(tl,na.rm=TRUE))
> tmp
Source: local data frame [2 x 3]

  fYear min max
1  1990  55 207
2  2000  55 147

> # base plot
> clrs <- c(rgb(0,0,0,1/6),rgb(0,0,1,1/6))
> plot(wt~tl,data=ruf9000,pch=19,col=clrs[ruf9000$fYear],ylab="Weight (g)",xlab="Total Length (mm)")
>
> # plot line for 1990
> tmpx <- seq(tmp$min[1],tmp$max[1],length.out=99)
> tmpy <- 10^(predict(fit2,data.frame(logL=log10(tmpx),fYear=factor(1990))))
> lines(tmpy~tmpx,lwd=2)
>
> # plot line for 2000
> tmpx <- seq(tmp$min[2],tmp$max[2],length.out=99)
> tmpy <- 10^(predict(fit2,data.frame(logL=log10(tmpx),fYear=factor(2000))))
> lines(tmpy~tmpx,col="blue",lwd=2)
>
> # add a legend
> legend("topleft",c("1990","2000"),pch=19,col=c("black","blue"),bty="n")
```



1-way ANOVA

```
> ruf2 <- filterD(ruf,year %in% c(1990,1995,2000,2006))
> fit3 <- lm(tl~fYear,data=ruf2)
> anova(fit3)
            Df Sum Sq Mean Sq F value    Pr(>F)
fYear        3 120795   40265   57.25 < 2.2e-16
Residuals 1218 856640     703
Total     1221 977435

> mc1 <- glht(fit3,mcp(fYear="Tukey"))
> summary(mc1)
                 Estimate Std. Error t value Pr(>|t|)
1995 - 1990 == 0  -17.451      1.969  -8.864   <0.001
2000 - 1990 == 0  -21.777      1.783 -12.215   <0.001
2006 - 1990 == 0  -24.596      4.651  -5.288   <0.001
2000 - 1995 == 0   -4.326      1.994  -2.169    0.120
2006 - 1995 == 0   -7.145      4.736  -1.509    0.410
2006 - 2000 == 0   -2.819      4.662  -0.605    0.925

> cld(mc1)
1990 1995 2000 2006
 "b"  "a"  "a"  "a"

> fitPlot(fit3,ylab="Total Length (mm)",xlab="Capture Year")
```



Application Assignment

Create a script that performs the following tasks:

1. Load the `BLGLW.CSV` data into an R data.frame.
2. Determine if there is a significant relationship between the weight and length of Bluegill in these data.
3. Determine if there is a significant difference between sexes in the relationship between the weight and length of Bluegill.
4. Determine if there is a significant difference among the three lakes in the mean weight of Bluegill.

Save your script!