# (Very) Quick Introduction to Linear Models in R

*Derek H. Ogle, Northland College*

*5-Mar-2015*

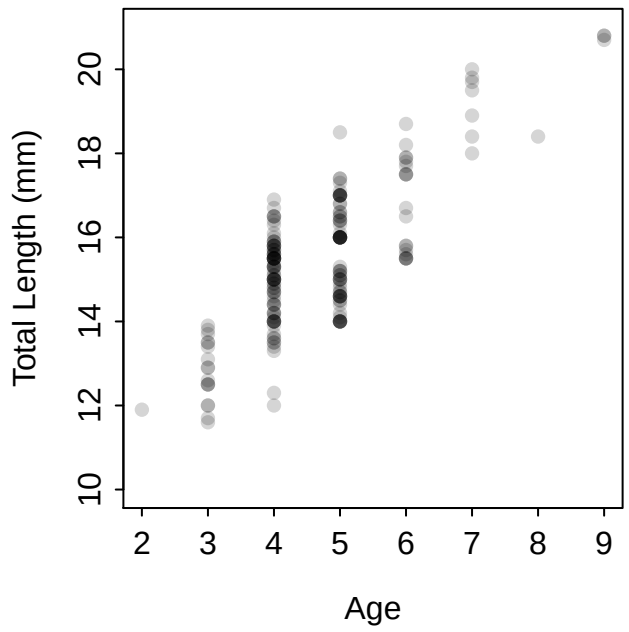
## Preliminaries

```
> source("03_SummarizeAgeData.R")
```

```
> ls()
 [1] "ages"         "BGSpr"        "BGSprLC"      "brks"         "clr"          "clrs"
 [7] "crap"         "d1"           "fn"           "freq"         "hook1"        "LCblg"
[13] "LCblgPREF"    "lmM"          "lmMF"         "rcum"         "Spr"          "SprLC"
[19] "sturgWts"     "tmp"          "wae.aged"     "waeF.fnl"     "waeF.sumlen" "waeM.fnl"
[25] "waeM.sumlen"
```

## Simple Linear Regession

```
> plot(Length.or.Lower.Length.IN~Age..observed.annuli.,data=waeM.fnl,pch=16,col=rgb(0,0,0,1/6),
      xlab="Age",ylab="Total Length (mm)",ylim=c(10,21))
```



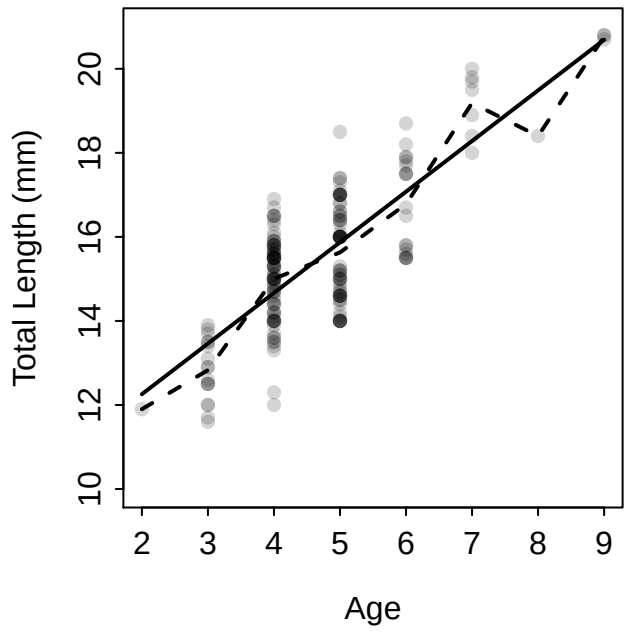

```
> lmM <- lm(Length.or.Lower.Length.IN~Age..observed.annuli.,data=waeM.fnl)
> coef(lmM)
         (Intercept) Age..observed.annuli.
            9.846584              1.204990
```

```
> confint(lmM)
                         2.5 %    97.5 %
(Intercept)           9.194574 10.498594
Age..observed.annuli. 1.068303  1.341677
```

```
> anova(lmM)
Analysis of Variance Table

Response: Length.or.Lower.Length.IN
                       Df Sum Sq Mean Sq F value    Pr(>F)
Age..observed.annuli.   1 343.94  343.94  302.32 < 2.2e-16
Residuals             193 219.57    1.14
```

```
> ages <- waeM.sumlen$Age..observed.annuli.
> ( waeM.sumlen %<>% mutate(predL=predict(lmM,data.frame(Age..observed.annuli.=ages))) )
Source: local data frame [8 x 7]

  Age..observed.annuli.   n     mean         sd  min  max    predL
1                     2   1 11.90000        NaN 11.9 11.9 12.25656
2                     3 17 12.82941 0.73719460 11.6 13.9 13.46155
3                     4 81 15.00247 0.96319459 12.0 16.9 14.66654
4                     5 66 15.63030 1.12221891 14.0 18.5 15.87153
5                     6 19 16.77895 1.12476118 15.5 18.7 17.07653
6                     7  7 19.18571 0.76469726 18.0 20.0 18.28152
7                     8  1 18.40000        NaN 18.4 18.4 19.48651
8                     9  3 20.76667 0.05773503 20.7 20.8 20.69150
```

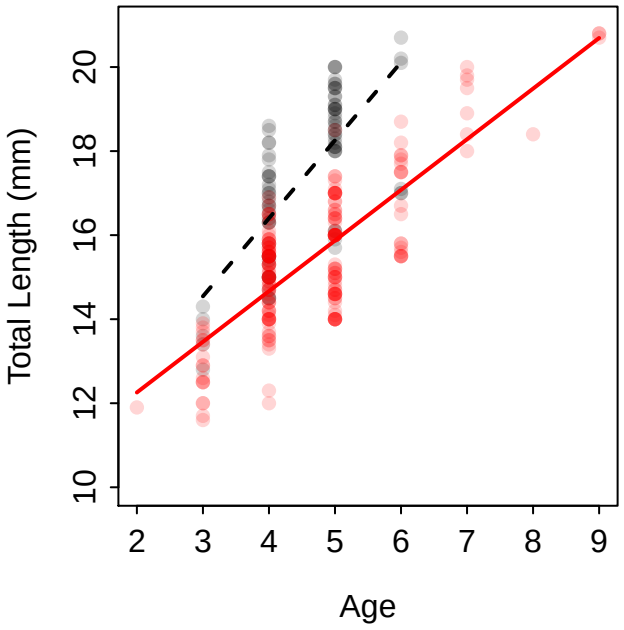
```
> plot(Length.or.Lower.Length.IN~Age..observed.annuli.,data=waeM.fnl,pch=16,col=rgb(0,0,0,1/6),
     xlab="Age",ylab="Total Length (mm)",ylim=c(10,21))
> lines(mean~Age..observed.annuli.,data=waeM.sumlen,lwd=2,lty=2)
> lines(predL~Age..observed.annuli.,data=waeM.sumlen,lwd=2,lty=1)
```



## Dummy Variable Regession (aka ANCOVA)

```
> wae <- rbind(waeF.fnl,waeM.fnl)
> levels(wae$Gender)
[1] "F" "M"

> lmMF <- lm(Length.or.Lower.Length.IN~Age..observed.annuli.*Gender,data=wae)
> coef(lmMF)
                  (Intercept)         Age..observed.annuli.                      GenderM
                    8.9762355                    1.8561514                    0.8703481
Age..observed.annuli.:GenderM
                   -0.6511612

> confint(lmMF)
                                   2.5 %    97.5 %
(Intercept)                    7.4368672 10.515604
Age..observed.annuli.          1.5194958  2.192807
GenderM                       -0.8165676  2.557264
Age..observed.annuli.:GenderM -1.0175733 -0.284749

> anova(lmMF)
Analysis of Variance Table

Response: Length.or.Lower.Length.IN
                             Df Sum Sq Mean Sq F value    Pr(>F)
Age..observed.annuli.         1 440.82  440.82 344.707 < 2.2e-16
Gender                        1 260.34  260.34 203.577 < 2.2e-16
Age..observed.annuli.:Gender  1  15.65   15.65  12.238  0.000545
Residuals                   278 355.51    1.28

> ages <- waeF.sumlen$Age..observed.annuli.
> ( waeF.sumlen %<>% mutate(predL2=predict(lmMF,data.frame(Age..observed.annuli.=ages,Gender="F"))) )
Source: local data frame [4 x 7]

  Age..observed.annuli.  n     mean        sd  min  max   predL2
1                     3  5 13.62000 0.5761944 12.8 14.3 14.54469
2                     4 38 16.39474 1.1790852 14.5 18.7 16.40084
3                     5 38 18.60526 1.1212800 15.7 20.7 18.25699
4                     6  6 18.71667 1.5484401 17.0 20.2 20.11314

> ages <- waeM.sumlen$Age..observed.annuli.
> ( waeM.sumlen %<>% mutate(predL2=predict(lmMF,data.frame(Age..observed.annuli.=ages,Gender="M"))) )
Source: local data frame [8 x 8]

  Age..observed.annuli.  n     mean         sd  min  max    predL   predL2
1                     2  1 11.90000        NaN 11.9 11.9 12.25656 12.25656
2                     3 17 12.82941 0.73719460 11.6 13.9 13.46155 13.46155
3                     4 81 15.00247 0.96319459 12.0 16.9 14.66654 14.66654
4                     5 66 15.63030 1.12221891 14.0 18.5 15.87153 15.87153
5                     6 19 16.77895 1.12476118 15.5 18.7 17.07653 17.07653
6                     7  7 19.18571 0.76469726 18.0 20.0 18.28152 18.28152
7                     8  1 18.40000        NaN 18.4 18.4 19.48651 19.48651
8                     9  3 20.76667 0.05773503 20.7 20.8 20.69150 20.69150
```

```
> clr <- c(rgb(0,0,0,1/6),rgb(1,0,0,1/6))
> plot(Length.or.Lower.Length.IN~Age..observed.annuli.,data=wae,pch=16,col=clr[Gender],
      xlab="Age",ylab="Total Length (mm)",ylim=c(10,21))
> lines(predL2~Age..observed.annuli.,data=waeF.sumlen,lwd=2,lty=2,col="black")
> lines(predL2~Age..observed.annuli.,data=waeM.sumlen,lwd=2,lty=1,col="red")
```




## 1-way ANOVA

```
> lm3 <- lm(Age..observed.annuli.~Gender,data=wae)
> anova(lm3)
Analysis of Variance Table

Response: Age..observed.annuli.
           Df  Sum Sq Mean Sq F value Pr(>F)
Gender      1   0.922  0.9218  0.9198 0.3383
Residuals 280 280.596  1.0021
```

## Application Assignment

Create a script that performs the following tasks:

1. Load your FM data into R.

**Save your script!**