R Handout - R Introduction/Review — Dr. Derek Ogle

Northland College, MN AFS — Dec 2013

Preliminaries

```
> library(FSA)
```

Reading Data

```
> setwd("C:/aaaWork/Web/fishR/courses/MNAFS2013/CourseMaterial/")
> d <- read.csv("MnFats.csv",header=TRUE)
> str(d)
'data.frame': 305 obs. of  6 variables:
 $ unit: Factor w/ 1 level "MN-1": 1 1 1 1 1 1 1 1 1 1 ...
 $ year: int  2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 ...
 $ len : int  310 363 373 381 394 394 396 401 406 409 ...
 $ wt  : int  240 330 370 490 470 490 460 490 540 650 ...
 $ sex : Factor w/ 3 levels "F","M","UNK": 1 1 2 2 2 2 1 2 2 1 ...
 $ age : int  9 10 17 10 11 14 11 15 13 15 ...
> view(d)
    unit year len   wt sex age
56  MN-1 2000 396  510   M  11
148 MN-1 2003 800 4230   M  27
151 MN-1 2003 371  390   M   7
153 MN-1 2003 406  520   M   7
168 MN-1 2003 572 1650   M  16
294 MN-1 2006 422  748   F  16
> nrow(d)
[1] 305
```

Parts of the Data

```
> d[5,]
  unit year len  wt sex age
5 MN-1 2000 394 470   M  11
> d[c(5,11,17),]
   unit year len  wt sex age
5  MN-1 2000 394 470   M  11
11 MN-1 2000 411 570   M  14
17 MN-1 2000 467 810   F  17
> d$age
```

```
  [1]   9 10 17 10 11 14 11 15 13 15 14 13 12 11 11 15 17 20 14 12 17 13 12 17 14 14
 [27]  12 12 18 11 12 17 17 13 20 15 18 13 13 17 16 17 19 24 16 23 23  6  9 10 10  9
 [53]   9 10 10 11 10 10 12 11 13 11 13 13 17 18 12 15 17 16 19 15 19 12 16 19 18 16
 [79]  19 19 13 13 13 11 12 13 14 12 13 14 15 14 21 13 17 16 19 15 15 16 24 19 19 33
[105]  12 12 13 16 17 14 20 19 18 21 27 21 25  7 12 13 16 12 15 11 14 20 14  9 14 15
[131]  14 18 19 18 15 21 15 25 23 27 24 24 25 20 25 19 23 27 26 12  7  7  7 11 14 11
[157]  16 17 13 14 12 22 14 16 18 16 23 16 19 12 12 16 14 12 13 17 17 11 13 15 16 20
[183]  11 15 23 23 20 24 22 22  7  6  9 12 12 16 14 13 13 16 15 15 14 17 18 18  9  7
[209]  10 15 11 11 10  9 11 10 15 13 13 18 14 13 11 15 18 10 13 12 16 28 17 10 14 16
[235]  20 21 16 20 21 28 22  8 17 10 17 11 16 16 12 16 15 18 15 14 13 14 27 21 22 20
[261]  18 18 15 18 18 18 16 23 26 11 21 24 22 17 19 18 26  7 11 15 10 15 11 14 17 17
[287]  23 20 10  9 13 13 11 16  7 11 32  8  7 12 13 16 17 13 18
> d$age[c(5,11,17)]
[1] 11 14 17
```

```
> d03 <- Subset(d,year==2003)
> nrow(d03)
[1] 86
> dmale03 <- Subset(d,year==2003 & sex=="M")
> nrow(dmale03)
[1] 52
> d0003 <- Subset(d,year==2000 | year==2003)
> nrow(d0003)
[1] 190
> dMF <- Subset(d,sex!="UNK")
> nrow(dMF)
[1] 304
> dgt500 <- Subset(d,len>500)
> nrow(dgt500)
[1] 131
```

Adding Variables

```
> d$loglen <- log(d$len)
> d$logwt <- log(d$wt)
> view(d)
    unit year len   wt sex age loglen logwt
3   MN-1 2000 373  370   M  17  5.922 5.914
45  MN-1 2000 648 2460   M  16  6.474 7.808
80  MN-1 2000 668 3330   F  19  6.504 8.111
110 MN-1 2003 455  900   M  14  6.120 6.802
255 MN-1 2006 465  771   F  13  6.142 6.648
305 MN-1 2006 589 1905   F  18  6.378 7.552
> d$fyear <- factor(d$year)
> str(d)
'data.frame': 305 obs. of  9 variables:
 $ unit  : Factor w/ 1 level "MN-1": 1 1 1 1 1 1 1 1 1 1 ...
```

```
 $ year  : int   2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 ...
 $ len   : int   310 363 373 381 394 394 396 401 406 409 ...
 $ wt    : int   240 330 370 490 470 490 460 490 540 650 ...
 $ sex   : Factor w/ 3 levels "F","M","UNK": 1 1 2 2 2 2 1 2 2 1 ...
 $ age   : int   9 10 17 10 11 14 11 15 13 15 ...
 $ loglen: num   5.74 5.89 5.92 5.94 5.98 ...
 $ logwt : num   5.48 5.8 5.91 6.19 6.15 ...
 $ fyear : Factor w/ 3 levels "2000","2003",..: 1 1 1 1 1 1 1 1 1 1 ...
> d <- lencat(~len,data=d,startcat=290,w=10)
> view(d)
    unit year len   wt sex age loglen logwt fyear LCat
1   MN-1 2000 310  240   F   9  5.737 5.481  2000  310
21  MN-1 2000 475  880   M  17  6.163 6.780  2000  470
79  MN-1 2000 668 3220   F  19  6.504 8.077  2000  660
101 MN-1 2000 615 2370   M  24  6.422 7.771  2000  610
125 MN-1 2003 470  900   F  14  6.153 6.802  2003  470
126 MN-1 2003 470  920   M  20  6.153 6.824  2003  470
```

Simple Summaries

```
> Summarize(~age,data=d,digits=2)
      n     mean       sd      min       Q1   median       Q3      max percZero
 305.00    15.48     4.79     6.00    12.00    15.00    18.00    33.00     0.00
> Summarize(age~fyear,data=d,digits=2)
  fyear   n  mean   sd min Q1 median   Q3 max percZero
1  2000 104 14.66 4.01   6 12     14 17.0  33        0
2  2003  86 16.84 5.10   7 13     16 20.8  27        0
3  2006 115 15.20 5.02   6 11     15 18.0  32        0
```
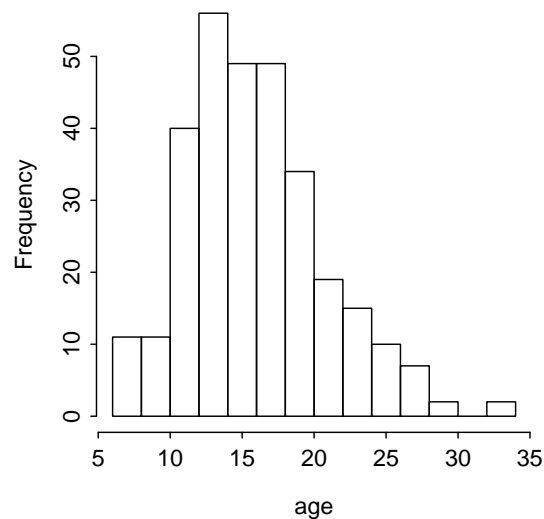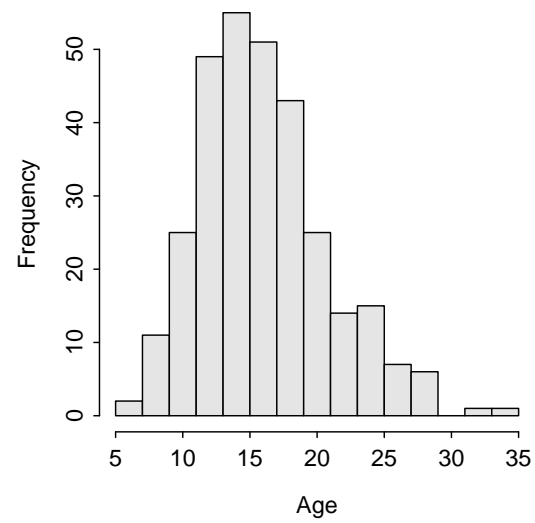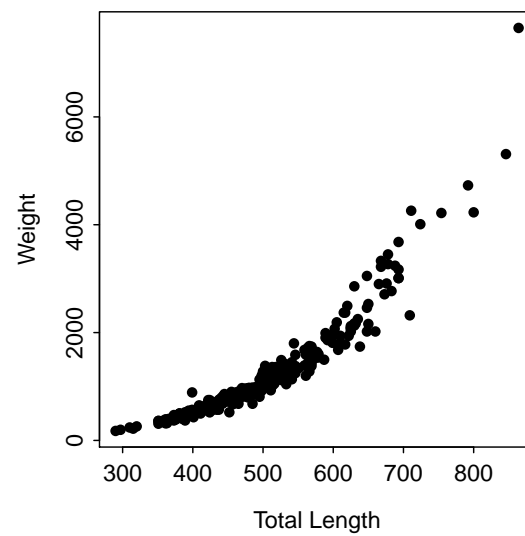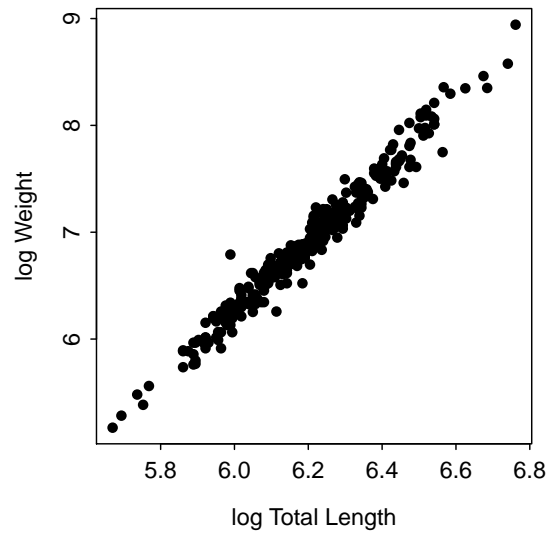
```
> hist(~age,data=d)
```

```
> hist(~age,data=d,xlab="Age",breaks=seq(5,35,2),col="gray90")
```
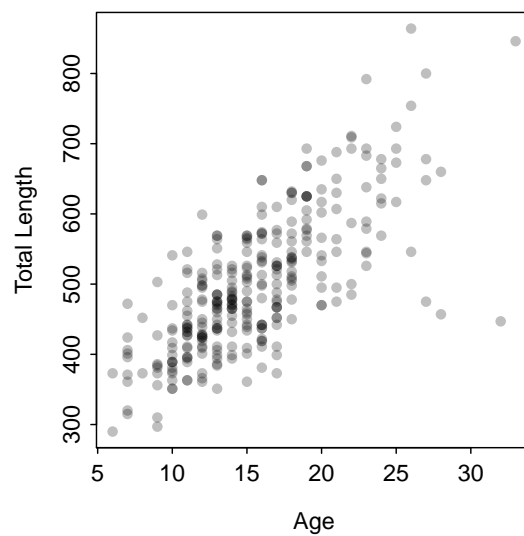


```
> plot(wt~len,data=d,xlab="Total Length",ylab="Weight",pch=16)
```



```
> plot(logwt~loglen,data=d,xlab="log Total Length",ylab="log Weight",pch=16)
```

```
> plot(len~age,data=d,xlab="Age",ylab="Total Length",pch=16,col=rgb(0,0,0,0.25))
```


Function Types, Regression Example

```
> lm1 <- lm(logwt~loglen,data=d)
> coef(lm1)
(Intercept)      loglen
    -13.187       3.241
> summary(lm1)

Call:
lm(formula = logwt ~ loglen, data = d)

Residuals:
    Min      1Q  Median      3Q     Max
-0.3713 -0.0660 -0.0049  0.0628  0.5665
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -13.1875     0.2062   -64.0   <2e-16
loglen        3.2413     0.0333    97.4   <2e-16

Residual standard error: 0.11 on 303 degrees of freedom
Multiple R-squared:  0.969,Adjusted R-squared:  0.969
F-statistic: 9.49e+03 on 1 and 303 DF,  p-value: <2e-16
> ( p1 <- predict(lm1,data.frame(loglen=log(400))) )
    1
6.233
> exp(p1)
    1
509.2
```

Adding Fitted Line to a Plot

```
> plot(wt~len,data=d,xlab="Total Length",ylab="Weight",pch=16)
> (loga <- coef(lm1)[1])
(Intercept)
     -13.19
> ( a <- exp(loga) )
(Intercept)
  1.874e-06
> ( b <- coef(lm1)[2] )
loglen
 3.241
> curve(a*x^b,from=250,to=900,col="red",lwd=2,add=TRUE)
```