# R Handout - Basics and Terminology

Dr. Derek Ogle
Northland College

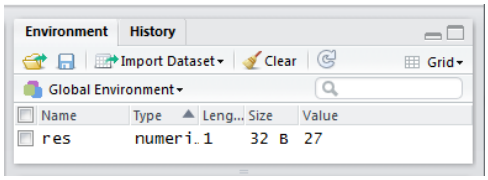
Mar 2014, Vermont CFWRU Workshop

## Load Necessary Packages

```
> library(FSA)      # for mrClosed
```

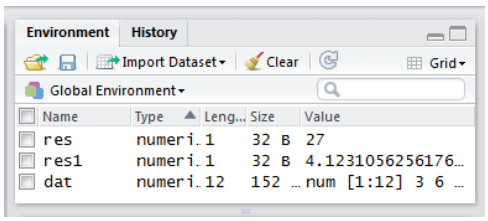
## Expressions & Assignments

```
> 3+4*2             # this is an expression
[1] 11
> res <- 3+4*2      # but this is an assignment
> res               #   to see what was assigned to memory
[1] 11
> ( res <- 9+3*6 ) # assign AND view
[1] 27
```




## Functions & Arguments

```
> sqrt(17)
[1] 4.123
> ( res1 <- sqrt(17) )
[1] 4.123
>
> dat <- c(3,6,8,3,5,6,2,7,6,8,2,10)
> mean(dat)
[1] 5.5
> mean(dat,trim=0.1)
[1] 5.4
```



## Types of Functions

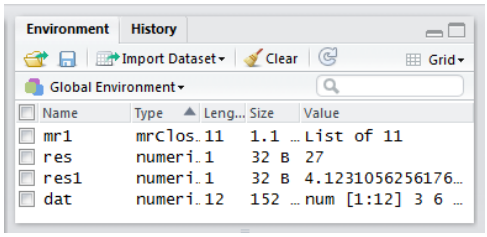
```
> mr1 <- mrClosed(M=346,n=184,m=49,type="Chapman")
> summary(mr1)
```

*Used Chapman's modification of the Petersen method with M=346, n=184, and m=49.*

```
        N
[1,] 1283
> confint(mr1)
```

*The binomial method was used.*

```
     95% LCI 95% UCI
[1,]    1025    1636
```




## Vectors & Data Types

```
> ( lake <- c("Star","Twin","Long","Deep") )
[1] "Star" "Twin" "Long" "Deep"
> ( numSpec <- c(4,8,7,3) )
[1] 4 8 7 3
> ( maxDepth <- c(6.5,7.8,3.8,25.6) )
[1]  6.5  7.8  3.8 25.6
> ( springFed <- c(TRUE,FALSE,FALSE,TRUE) )
[1]  TRUE FALSE FALSE  TRUE
```

```
> lake[1]
[1] "Star"
> lake[2]
[1] "Twin"
> lake[c(1,2)]
[1] "Star" "Twin"
> lake[-1]
[1] "Twin" "Long" "Deep"
> lake[c(TRUE,FALSE,FALSE,TRUE)]
[1] "Star" "Deep"
> lake=="Star"
[1]  TRUE FALSE FALSE FALSE
> maxDepth[lake=="Star"]
[1] 6.5
> numSpec[maxDepth<7]
[1] 4 7
```

## Data.frames

```
> # Put previous vectors into a data.frame.  For realistic sizes of data sets
> #   I would enter data externally and read into R ... more on this later
> ( df <- data.frame(lake,numSpec,maxDepth,springFed) )
  lake numSpec maxDepth springFed
1 Star       4      6.5      TRUE
2 Twin       8      7.8     FALSE
3 Long       7      3.8     FALSE
4 Deep       3     25.6      TRUE
> df[1,1]
[1] Star
Levels: Deep Long Star Twin
> df[1,]
  lake numSpec maxDepth springFed
1 Star       4      6.5      TRUE
> df[c(1,2),]
  lake numSpec maxDepth springFed
1 Star       4      6.5      TRUE
2 Twin       8      7.8     FALSE
> df[-1,]
  lake numSpec maxDepth springFed
2 Twin       8      7.8     FALSE
3 Long       7      3.8     FALSE
4 Deep       3     25.6      TRUE
> df[,2]
[1] 4 8 7 3
> df[,"numSpec"]
[1] 4 8 7 3
```

```
> str(df)
'data.frame': 4 obs. of  4 variables:
 $ lake     : Factor w/ 4 levels "Deep","Long",..: 3 4 2 1
 $ numSpec  : num  4 8 7 3
 $ maxDepth : num  6.5 7.8 3.8 25.6
 $ springFed: logi  TRUE FALSE FALSE TRUE
> df$numSpec
[1] 4 8 7 3
> df$numSpec[1]
[1] 4
> mean(df$numSpec)
[1] 5.5
```

# R Handout - Data Frames

**Dr. Derek Ogle**
Northland College

Mar 2014, Vermont CFWRU Workshop

## Load Necessary Packages

```
> library(FSA)      # for Subset(), view(), lencat()
```

## Reading External Data

When beginning with R it is easiest to load external data with the following steps:

1. Enter data in an external software (e.g., spreadsheet or database) and save as a "comma separated values" (CSV) file.
2. Start an initial script in RStudio. Save this script to the **exact same folder** as the CSV file.
3. Use the *"Session"*, *"Set Working Directory ..."*, *"To Source File Location"* menu items to print the appropriate `setwd()` function to the *Console* pane.
4. Copy the appropriate `setwd()` function from the *Console* pane to your script.
5. Use `read.csv()` to load the external file into the R environment (described below).
6. Use `str()`, `head()`, `view()`[1], or view the file from the *Environment* tab (upper-right pane of RStudio) to make sure the data appears proper.

```
> setwd("C:/aaaWork/Web/fishR/courses/Vermont2014/CourseMaterial/") # Derek's computer only
> d <- read.csv("Data/MNBCData.csv",header=TRUE)
> str(d)
'data.frame': 2422 obs. of  20 variables:
 $ species: Factor w/ 9 levels "BLC","BLG","LMB",..: 1 1 1 1 1 1 1 1 1 1 ...
 $ gear   : Factor w/ 5 levels "All","GN","GN, TN",..: 5 5 5 5 5 5 5 5 3 3 ...
 $ lake   : Factor w/ 24 levels "Bean Lake","Benton",..: 2 2 2 2 2 2 3 3 16 16 ...
 $ yearcap: int  2006 2006 2006 2006 2006 2006 2006 2006 2006 2006 ...
 $ fish   : int  67 43 45 44 42 41 63 78 56 55 ...
 $ agecap : int  1 2 4 4 4 4 4 4 1 1 ...
 $ lencap : int  108 198 258 247 249 235 278 284 135 127 ...
 $ anu1   : num  1.15 1.42 1.52 1.28 1.43 ...
 $ anu2   : num  2.28 3.06 3.28 3.35 2.53 ...
 $ anu3   : num  NA 3.88 4.27 3.86 3.36 ...
 $ anu4   : num  NA NA 4.88 4.11 3.68 ...
 $ anu5   : num  NA NA 5.66 4.34 3.99 ...
 $ anu6   : num  NA NA NA NA NA NA NA NA NA NA ...
 $ anu7   : num  NA NA NA NA NA NA NA NA NA NA ...
 $ anu8   : num  NA NA NA NA NA NA NA NA NA NA ...
 $ anu9   : num  NA NA NA NA NA NA NA NA NA NA ...
 $ anu10  : num  NA NA NA NA NA NA NA NA NA NA ...
 $ anu11  : num  NA NA NA NA NA NA NA NA NA NA ...
 $ anu12  : num  NA NA NA NA NA NA NA NA NA NA ...
 $ radcap : num  2.28 3.88 5.66 4.34 3.99 ...
> view(d)
```

[1] This requires the `FSA` package.

```
     species gear        lake yearcap fish agecap lencap   anu1  anu2  anu3  anu4
127      BLG   TN        Long    2006   29      3    162 0.9370 1.966 3.034 3.536
407      YEP  All   Bean Lake    1998   72      2    237 1.4616 3.344 4.228    NA
1058     BLC  All   Iowa Lake    1998  120      4    295 0.9971 2.453 3.697 4.831
1277     BLC  All  Lake Sarah    1998   46      2    193 1.4169 3.555 3.753    NA
1809     BLC  All   Long Lake    1998  128      1    107 0.5355 1.660    NA    NA
2334     BLC  All Timber Lake    1998   38      2    192 0.9008 3.043 3.789    NA
      anu5 anu6 anu7 anu8 anu9 anu10 anu11 anu12 radcap
127     NA   NA   NA   NA   NA    NA    NA    NA  3.536
407     NA   NA   NA   NA   NA    NA    NA    NA  4.228
1058 5.303   NA   NA   NA   NA    NA    NA    NA  5.303
1277    NA   NA   NA   NA   NA    NA    NA    NA  3.753
1809    NA   NA   NA   NA   NA    NA    NA    NA  1.660
2334    NA   NA   NA   NA   NA    NA    NA    NA  3.789
> nrow(d)
[1] 2422
```

## Subsets of the Data

```
> d[5,]
  species gear   lake yearcap fish agecap lencap  anu1  anu2 anu3  anu4  anu5 anu6
5     BLC   TN Benton    2006   42      4    249 1.431 2.534 3.36 3.676 3.992   NA
  anu7 anu8 anu9 anu10 anu11 anu12 radcap
5   NA   NA   NA    NA    NA    NA  3.992
> d[c(5,11,17),]
   species   gear   lake yearcap fish agecap lencap  anu1  anu2 anu3  anu4  anu5
5      BLC     TN Benton    2006   42      4    249 1.431 2.534 3.36 3.676 3.992
11     BLC GN, TN   Long    2006   54      1    123 1.452 2.185   NA    NA    NA
17     BLC GN, TN   Long    2006   41      1    118 1.221 1.928   NA    NA    NA
   anu6 anu7 anu8 anu9 anu10 anu11 anu12 radcap
5    NA   NA   NA   NA    NA    NA    NA  3.992
11   NA   NA   NA   NA    NA    NA    NA  2.185
17   NA   NA   NA   NA    NA    NA    NA  1.928
> d$age[1:25]
 [1] 1 2 4 4 4 4 4 4 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 3
```

```
> levels(d$species)
[1] "BLC" "BLG" "LMB" "NOP" "PMK" "SMB" "WAE" "WHC" "YEP"
> levels(d$lake)
 [1] "Bean Lake"         "Benton"            "Bingham"           "Bingham Lake"
 [5] "Buff Lake"         "Cottonwood Lake"   "Fish Lake"         "Fox Lake"
 [9] "Hills Reservoir"   "Iowa Lake"         "Kinbrae"           "Lake Okamanpeedan"
[13] "Lake Sarah"        "Lake Shetek"       "Lake Yankton"      "Long"
[17] "Long Lake"         "Okabena Lake"      "Rock Lake"         "South Silver"
[21] "South Silver Lake" "Summit Lake"       "Talcot"            "Timber Lake"
```

```
> dBLC <- Subset(d,species=="BLC")
> xtabs(~species,data=dBLC)

species
BLC
563

> dBLCTL <- Subset(d,species=="BLC" & lake=="Talcot")
> xtabs(~species+lake,data=dBLCTL)

       lake
species Talcot
    BLC     68

> dBLCBLG <- Subset(d,species=="BLC" | species=="BLG")
> xtabs(~species,data=dBLCBLG)

species
BLC BLG
563 174

> d2 <- Subset(d,species!="BLC")
> xtabs(~species,data=d2)

species
BLG LMB NOP PMK SMB WAE WHC YEP
174  27  72  75  42 550 175 744

> dPred <- Subset(d,species %in% c("LMB","NOP","SMB","WAE"))
> xtabs(~species,data=dPred)

species
LMB NOP SMB WAE
 27  72  42 550

> dgt500 <- Subset(d,lencap>=500)
> nrow(dgt500)

[1] 173

> min(dgt500$lencap)

[1] 500
```

## Adding Variables I

```
> d$lenin <- d$lencap/25.4
> d$loglen <- log(d$lencap)
> view(d)

     species  gear              lake yearcap fish agecap lencap   anu1  anu2  anu3
151      NOP    TN   Hills Reservoir    2006   26      3    511 1.3490 2.039 2.401
244      WAE GN,TN            Talcot    2006   27      2    313 1.4650 3.207 4.255
1233     YEP   All Lake Okamanpeedan    1998   44      2    139 1.2602 2.458 2.745
1488     BLC   All       Lake Shetek    1998  131      4    268 0.9118 2.651 3.961
1799     YEP   All      Lake Yankton    1998  170      3    254 1.1317 2.928 4.344
2344     BLC   All       Timber Lake    1998   94      2    224 1.1785 3.656 4.432
      anu4  anu5 anu6 anu7 anu8 anu9 anu10 anu11 anu12 radcap  lenin loglen
151  2.875    NA   NA   NA   NA   NA    NA    NA    NA  2.875 20.118  6.236
244     NA    NA   NA   NA   NA   NA    NA    NA    NA  4.255 12.323  5.746
1233    NA    NA   NA   NA   NA   NA    NA    NA    NA  2.745  5.472  4.934
1488 4.645 4.964   NA   NA   NA   NA    NA    NA    NA  4.964 10.551  5.591
1799 4.693    NA   NA   NA   NA   NA    NA    NA    NA  4.693 10.000  5.537
2344    NA    NA   NA   NA   NA   NA    NA    NA    NA  4.432  8.819  5.412
```

```
> # Create a year factor (categorical) variable
> d$fyearcap <- factor(d$yearcap)
> str(d)
'data.frame': 2422 obs. of  23 variables:
 $ species : Factor w/ 9 levels "BLC","BLG","LMB",..: 1 1 1 1 1 1 1 1 1 1 ...
 $ gear    : Factor w/ 5 levels "All","GN","GN, TN",..: 5 5 5 5 5 5 5 5 3 3 ...
 $ lake    : Factor w/ 24 levels "Bean Lake","Benton",..: 2 2 2 2 2 2 3 3 16 16 ...
 $ yearcap : int  2006 2006 2006 2006 2006 2006 2006 2006 2006 2006 ...
 $ fish    : int  67 43 45 44 42 41 63 78 56 55 ...
 $ agecap  : int  1 2 4 4 4 4 4 4 1 1 ...
 $ lencap  : int  108 198 258 247 249 235 278 284 135 127 ...
 $ anu1    : num  1.15 1.42 1.52 1.28 1.43 ...
 $ anu2    : num  2.28 3.06 3.28 3.35 2.53 ...
 $ anu3    : num  NA 3.88 4.27 3.86 3.36 ...
 $ anu4    : num  NA NA 4.88 4.11 3.68 ...
 $ anu5    : num  NA NA 5.66 4.34 3.99 ...
 $ anu6    : num  NA NA NA NA NA NA NA NA NA NA ...
 $ anu7    : num  NA NA NA NA NA NA NA NA NA NA ...
 $ anu8    : num  NA NA NA NA NA NA NA NA NA NA ...
 $ anu9    : num  NA NA NA NA NA NA NA NA NA NA ...
 $ anu10   : num  NA NA NA NA NA NA NA NA NA NA ...
 $ anu11   : num  NA NA NA NA NA NA NA NA NA NA ...
 $ anu12   : num  NA NA NA NA NA NA NA NA NA NA ...
 $ radcap  : num  2.28 3.88 5.66 4.34 3.99 ...
 $ lenin   : num  4.25 7.8 10.16 9.72 9.8 ...
 $ loglen  : num  4.68 5.29 5.55 5.51 5.52 ...
 $ fyearcap: Factor w/ 2 levels "1998","2006": 2 2 2 2 2 2 2 2 2 2 ...
> levels(d$fyearcap)
[1] "1998" "2006"
```

```
> # Create a length categorization variable
> d <- lencat(~lencap,data=d,startcat=75,w=25)
> view(d)
     species  gear        lake yearcap fish agecap lencap   anu1  anu2  anu3  anu4
161      NOP GN,TN     Kinbrae    2006   28      2    628 1.1290 2.552 3.035    NA
951      WAE   All    Fox Lake    1998  108      3    415 2.6896 4.362 5.059 5.456
1096     WHC   All   Iowa Lake    1998  106      2    254 0.6068 3.138 3.956    NA
1102     WHC   All   Iowa Lake    1998   91      3    292 0.8688 3.026 3.983 4.300
1373     WAE   All  Lake Sarah    1998  151      6    585 2.1288 4.105 5.638 6.319
1605     YEP   All Lake Shetek    1998   40      3    203 0.9115 2.230 3.072 3.410
      anu5  anu6  anu7 anu8 anu9 anu10 anu11 anu12 radcap  lenin loglen fyearcap
161     NA    NA    NA   NA   NA    NA    NA    NA  3.035 24.724  6.443     2006
951     NA    NA    NA   NA   NA    NA    NA    NA  5.456 16.339  6.028     1998
1096    NA    NA    NA   NA   NA    NA    NA    NA  3.956 10.000  5.537     1998
1102    NA    NA    NA   NA   NA    NA    NA    NA  4.300 11.496  5.677     1998
1373 6.717 7.015 7.168   NA   NA    NA    NA    NA  7.168 23.031  6.372     1998
1605    NA    NA    NA   NA   NA    NA    NA    NA  3.410  7.992  5.313     1998
     LCat
161   625
951   400
1096  250
1102  275
1373  575
1605  200
```

```
> xtabs(~species+LCat,data=d)
       LCat
species  75 100 125 150 175 200 225 250 275 300 325 350 375 400 425 450 475 500 525
    BLC   0  26  39  78  94  91  84  67  56  17   7   4   0   0   0   0   0   0   0
    BLG   6  16  25  65  48  10   4   0   0   0   0   0   0   0   0   0   0   0   0
    LMB   0   0   0   0   0   0   0   0   2   3   3   3   4   3   3   5   1   0   0
    NOP   0   0   0   0   0   0   0   0   0   0   0   2   2   6   3   2   5   3   6
    PMK   1  20  28  18   7   1   0   0   0   0   0   0   0   0   0   0   0   0   0
    SMB   1   3   0   0   9   7   0   5   6   6   2   0   0   0   2   1   0   0   0
    WAE   0   0   0   0  11  18  18  17  28  56  54  48  54  30  35  27  33  37  26
    WHC   0   5   2  15  37  50  23  18  24   1   0   0   0   0   0   0   0   0   0
    YEP   0   4  66 141 152 158 119  59  35  10   0   0   0   0   0   0   0   0   0
       LCat
species 550 575 600 625 650 675 700 725 750 775 800 825
    BLC   0   0   0   0   0   0   0   0   0   0   0   0
    BLG   0   0   0   0   0   0   0   0   0   0   0   0
    LMB   0   0   0   0   0   0   0   0   0   0   0   0
    NOP   2   6  10   9   4   4   1   2   2   0   2   1
    PMK   0   0   0   0   0   0   0   0   0   0   0   0
    SMB   0   0   0   0   0   0   0   0   0   0   0   0
    WAE  22  13   5   9   2   4   3   0   0   0   0   0
    WHC   0   0   0   0   0   0   0   0   0   0   0   0
    YEP   0   0   0   0   0   0   0   0   0   0   0   0
```

# Back-Calculation Example

```
> # Focus hereafter on Talcot Lake Black Crappie in 2006 (only year sampled)
> #   and eliminate several variables not used (for illustration & simplicity)
> dBLC <- Subset(d,species=="BLC" & lake=="Talcot" & yearcap==2006,
                 select=-c(gear,yearcap,lenin,loglen,fyearcap,LCat))
> levels(dBLC$species)
[1] "BLC"
> levels(dBLC$lake)
[1] "Talcot"
```

## Reshaping

Some definitions:

- *"Wide" Data* – Rows contains repeated measurements on same individuals. This is currently the cases ... each row has multiple scale radii measurements for the same fish in the `anuX` variables.
- *"Long" Data* – Each row contains only one measurement from an individual. Thus, multiple measurements on the same fish will be in multiple rows.

The `reshape()` function can be used to change the shape of a data frame from wide to long, or vice versa. Within this function several items have to be defined when moving from "wide" to "long" format.

- `idvar` – The single variable name that identifies an individual (a fish in this case).
- `varying` – A vector of names for the variables that contain the repeated measurements (i.e., the variables names containing the scale radius measurements).
- `v.names` – A single name for repeated measurements variable in the long format. This will usually be very closely related to the common portion of the names in `varying`.
- `timevar` – A single name for the labels for the repeated measurements values in the long format. This is likely a descriptive name for specific portion of the names in `varying`.
- `times` – A vector of values for the repeated measurements in the long format (i.e., the ages corresponding to the radial measurements in this case).

```
> # list the variables that contain the repeated measurements
> varying1 <- c("anu1","anu2","anu3","anu4","anu5","anu6",
                "anu7","anu8","anu9","anu10","anu11","anu12")
> # this is an alternative to the above that eliminates repetitive typing
> ( varying2 <- which(grepl("anu",names(dBLC))) )
 [1]  6  7  8  9 10 11 12 13 14 15 16 17
>
> # do the reshaping
> ldBLC <- reshape(dBLC,direction="long",
                   idvar="fish",          # what identifies unique fish
                   varying=varying1,      # declare the repeated measurements
                   v.names="anu",         # name for repeat meas in long format
                   timevar="age",         # name of var that identifies the repeat
                   times=1:12)            # values in timevar for repeat
```

```
>
> view(ldBLC)
      species   lake fish agecap lencap radcap age   anu
149.2     BLC Talcot  149      1    128  2.369   2 2.369
38.4      BLC Talcot   38      1    168  3.363   4    NA
81.6      BLC Talcot   81      2    193  3.668   6    NA
84.8      BLC Talcot   84      1    127  2.469   8    NA
48.10     BLC Talcot   48      1    144  2.855  10    NA
49.12     BLC Talcot   49      1    137  2.756  12    NA
>
> ldBLC[ldBLC$fish==165,]                # example for one fish
       species   lake fish agecap lencap radcap age   anu
165.1      BLC Talcot  165      5    276  5.159   1 1.757
165.2      BLC Talcot  165      5    276  5.159   2 3.265
165.3      BLC Talcot  165      5    276  5.159   3 4.308
165.4      BLC Talcot  165      5    276  5.159   4 4.728
165.5      BLC Talcot  165      5    276  5.159   5 5.045
165.6      BLC Talcot  165      5    276  5.159   6 5.159
165.7      BLC Talcot  165      5    276  5.159   7    NA
165.8      BLC Talcot  165      5    276  5.159   8    NA
165.9      BLC Talcot  165      5    276  5.159   9    NA
165.10     BLC Talcot  165      5    276  5.159  10    NA
165.11     BLC Talcot  165      5    276  5.159  11    NA
165.12     BLC Talcot  165      5    276  5.159  12    NA
>
> # remove all of the NAs
> ldBLC <- Subset(ldBLC,!is.na(anu))
> ldBLC[ldBLC$fish==165,]                # same example for one fish
    species   lake fish agecap lencap radcap age   anu
67      BLC Talcot  165      5    276  5.159   1 1.757
135     BLC Talcot  165      5    276  5.159   2 3.265
172     BLC Talcot  165      5    276  5.159   3 4.308
193     BLC Talcot  165      5    276  5.159   4 4.728
208     BLC Talcot  165      5    276  5.159   5 5.045
216     BLC Talcot  165      5    276  5.159   6 5.159
>
> # remove the "plus" growth
> ldBLC <- Subset(ldBLC,agecap-age>=0)
> ldBLC[ldBLC$fish==165,]                # same example for one fish
    species   lake fish agecap lencap radcap age   anu
67      BLC Talcot  165      5    276  5.159   1 1.757
104     BLC Talcot  165      5    276  5.159   2 3.265
125     BLC Talcot  165      5    276  5.159   3 4.308
140     BLC Talcot  165      5    276  5.159   4 4.728
148     BLC Talcot  165      5    276  5.159   5 5.045
```

## Adding Variables II

```
> k <- 35  # use Carlander intercept of k=35 mm
> ldBLC <- within(ldBLC, {
   bcFL <- (anu/radcap)*(lencap-k)+k
 })
> view(ldBLC)
```

```
    species   lake fish agecap lencap radcap age   anu   bcFL
42      BLC Talcot   61      2    213  4.259   1 0.842  70.19
45      BLC Talcot   77      2    218  4.384   1 0.980  75.91
88      BLC Talcot  143      3    241  4.736   2 3.186 173.58
101     BLC Talcot   80      5    290  5.023   2 2.710 172.58
121     BLC Talcot   30      5    279  5.023   3 3.787 218.96
140     BLC Talcot  165      5    276  5.159   4 4.728 255.87
```

```
> Summarize(bcFL~age,data=ldBLC)
```

```
Warning:  To continue, variable(s) on RHS of formula were converted to a factor.
  age  n   mean     sd   min    Q1 median    Q3 max percZero
1   1 68  89.32 14.167  63.8  78.4   88.7  98.9 117        0
2   2 37 173.03 16.442 134.0 162.0  171.0 184.0 217        0
3   3 21 226.02 11.139 205.0 221.0  227.0 231.0 249        0
4   4 15 255.28  8.737 247.0 250.0  252.0 261.0 273        0
5   5  8 277.12  8.414 266.0 271.0  276.0 285.0 288        0
```

# R Handout - Simple Statistical Summaries

**Dr. Derek Ogle**
Northland College

Mar 2014, Vermont CFWRU Workshop

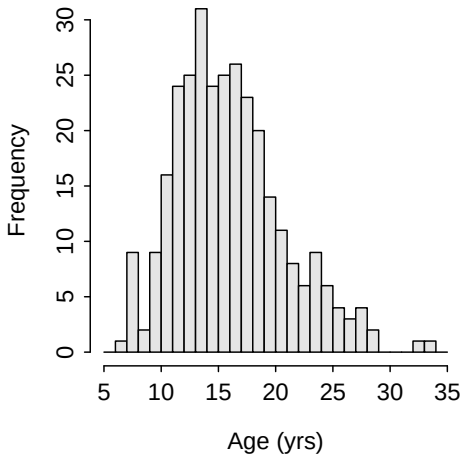
## Preliminaries

```
> library(FSA)      # for Subset(), Summarize(), hist(), fact2num()
> library(plotrix) # for plotCI()
```

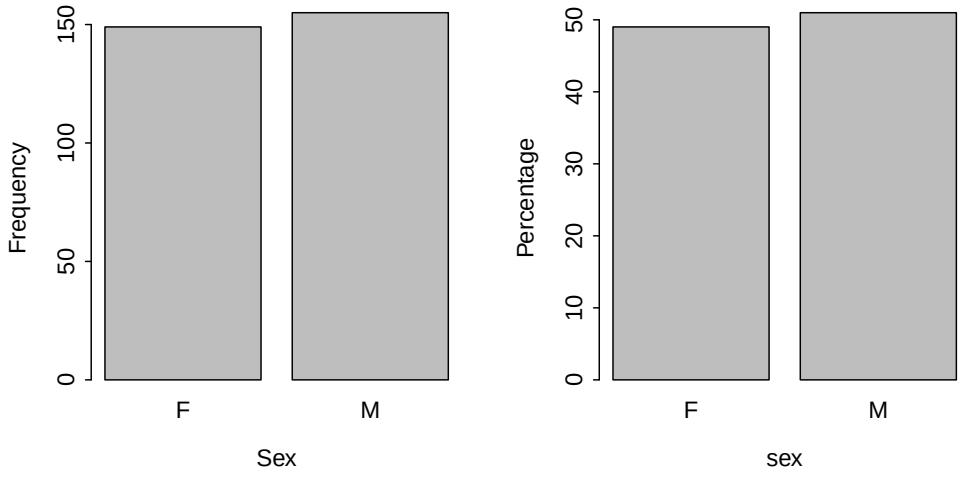
```
> setwd("C:/aaaWork/Web/fishR/courses/Vermont2014/CourseMaterial/") # Derek's Computer
> d <- read.csv("Data/MnFats.csv",header=TRUE)
> d <- Subset(d,sex!="UNK")  # removed one unknown sex individual (for simplicity)
> str(d)
'data.frame': 304 obs. of  6 variables:
 $ unit: Factor w/ 1 level "MN-1": 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/ 2 levels "F","M": 1 1 2 2 2 2 1 2 2 1 ...
 $ age : int  9 10 17 10 11 14 11 15 13 15 ...
> d <- within(d, {
             fyear <- factor(year)
             loglen <- log(len)
             logwt <- log(wt)
             } )
> view(d)
    unit year len   wt sex age logwt loglen fyear
65  MN-1 2000 498  920   F  18 6.824  6.211  2000
69  MN-1 2000 564 1720   M  16 7.450  6.335  2000
161 MN-1 2003 500 1260   M  22 7.139  6.215  2003
165 MN-1 2003 538 1170   F  16 7.065  6.288  2003
167 MN-1 2003 572 1650   M  16 7.409  6.349  2003
238 MN-1 2006 630 2130   M  21 7.664  6.446  2006
```

## Simple Univariate Summaries

```
> Summarize(~age,data=d,digits=2)
       n     mean       sd      min       Q1   median       Q3      max percZero
  304.00    15.51     4.77     6.00    12.00    15.00    18.00    33.00     0.00
> hist(~age,data=d,xlab="Age (yrs)",breaks=seq(5,35,1))
```

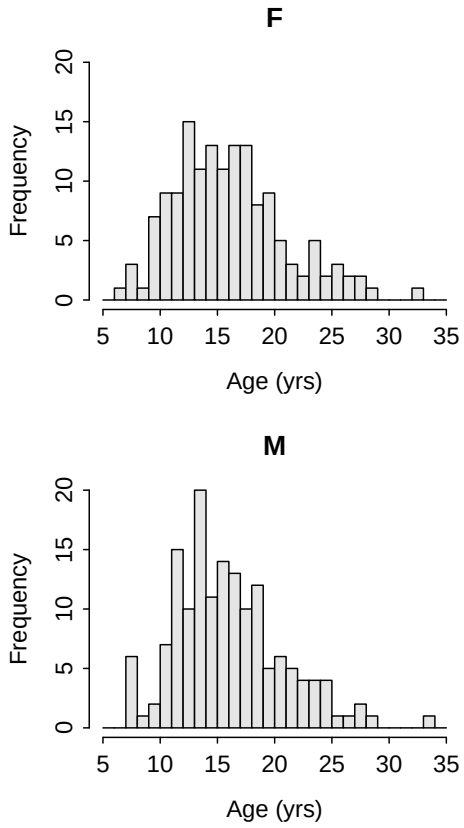



```
> ( sextbl <- xtabs(~sex,data=d) )
sex
  F   M
149 155
> prop.table(sextbl)*100
sex
    F     M
49.01 50.99
> barplot(sextbl,xlab="Sex",ylab="Frequency")                    # Left
> barplot(prop.table(sextbl)*100,xlab="sex",ylab="Percentage")   # Right
```

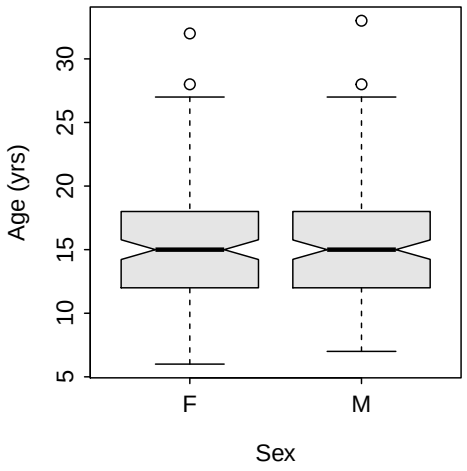


## Simple Bivariate Summaries

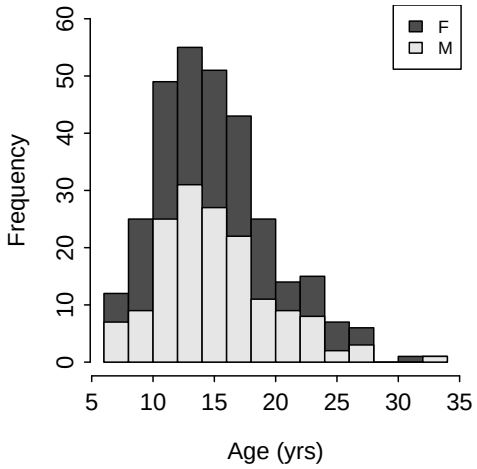
```
> Summarize(age~sex,data=d,digits=2)
  sex   n  mean   sd min Q1 median Q3 max percZero
1   F 149 15.52 4.85   6 12     15 18  32        0
2   M 155 15.50 4.71   7 12     15 18  33        0
> hist(age~sex,data=d,xlab="Age (yrs)",breaks=seq(5,35,1),nrow=2,ncol=1)
```




```
> boxplot(age~sex,data=d,xlab="Sex",ylab="Age (yrs)",col="gray90",notch=TRUE)
```

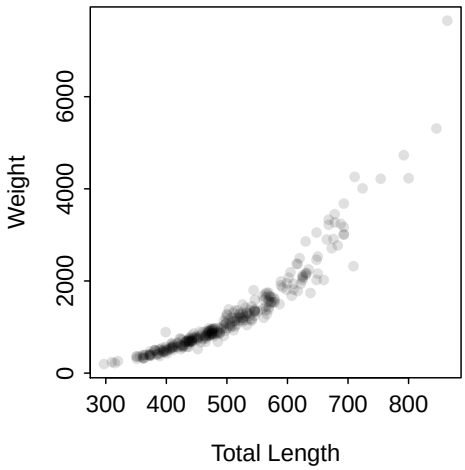



```
> histStack(age~sex,data=d,xlab="Age (yrs)",breaks=seq(5,35,1),ylim=c(0,60),
          col="gray.colors",legend="topright")
```

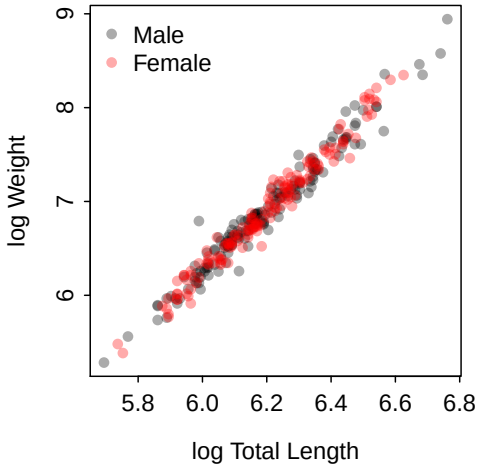



```
> agesextbl <- xtabs(~sex+age,data=d)
> round(prop.table(agesextbl)*100,1)
   age
sex   6   7   8   9  10  11  12  13  14  15  16  17  18  19  20  21  22  23  24  25
  F 0.3 1.0 0.3 2.3 3.0 3.0 4.9 3.6 4.3 3.6 4.3 4.3 2.6 3.0 1.6 1.0 0.7 1.6 0.7 1.0
  M 0.0 2.0 0.3 0.7 2.3 4.9 3.3 6.6 3.6 4.6 4.3 3.3 3.9 1.6 2.0 1.6 1.3 1.3 1.3 0.3
   age
sex  26  27  28  32  33
  F 0.7 0.7 0.3 0.3 0.0
  M 0.3 0.7 0.3 0.0 0.3
> round(prop.table(agesextbl,margin=1)*100,1)
   age
sex    6    7    8    9   10   11   12   13   14   15   16   17   18   19   20   21
  F  0.7  2.0  0.7  4.7  6.0  6.0 10.1  7.4  8.7  7.4  8.7  8.7  5.4  6.0  3.4  2.0
  M  0.0  3.9  0.6  1.3  4.5  9.7  6.5 12.9  7.1  9.0  8.4  6.5  7.7  3.2  3.9  3.2
   age
sex   22   23   24   25   26   27   28   32   33
  F  1.3  3.4  1.3  2.0  1.3  1.3  0.7  0.7  0.0
  M  2.6  2.6  2.6  0.6  0.6  1.3  0.6  0.0  0.6
```

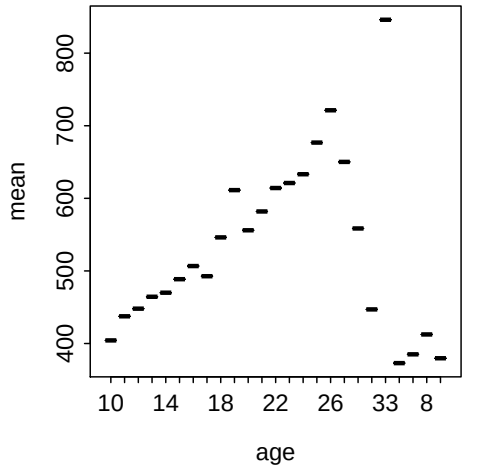
```
> plot(wt~len,data=d,xlab="Total Length",ylab="Weight",pch=16,col=rgb(0,0,0,1/8))
```



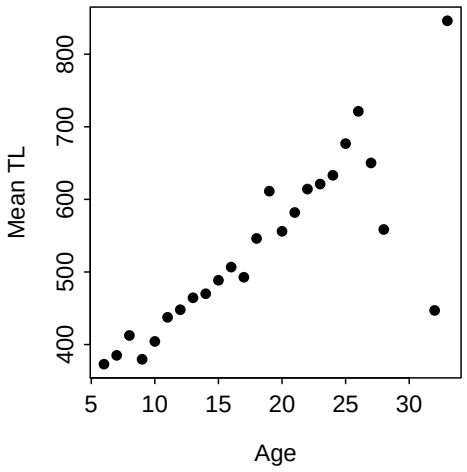
```
> colM <- rgb(0,0,0,1/3)
> colF <- rgb(1,0,0,1/3)
> plot(logwt~loglen,data=Subset(d,sex=="M"),pch=16,col=colM,
      xlab="log Total Length",ylab="log Weight")
> points(logwt~loglen,data=Subset(d,sex=="F"),pch=16,col=colF)
> legend("topleft",c("Male","Female"),pch=16,bty="n",col=c(colM,colF))
```



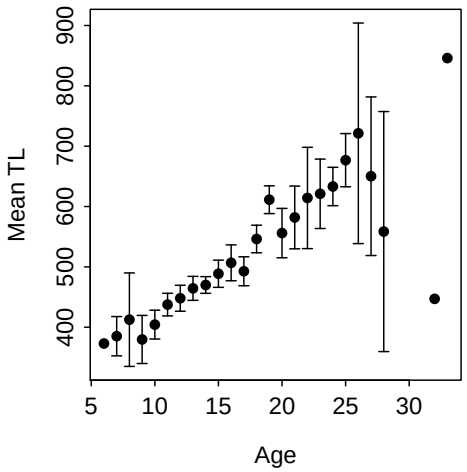

```
> lenAtAge <- Summarize(len~age,data=d,digits=1)
Warning:  To continue, variable(s) on RHS of formula were converted to a factor.
> str(lenAtAge)
'data.frame': 25 obs. of  10 variables:
 $ age     : Factor w/ 25 levels "10","11","12",..: 22 23 24 25 1 2 3 4 5 6 ...
 $ n       : num  1 9 2 9 16 24 25 31 24 25 ...
 $ mean    : num  373 385 412 380 404 ...
 $ sd      : num  NA 49.8 55.9 61 48.7 46.8 54.8 56.4 34.6 57.5 ...
 $ min     : num  373 315 373 297 351 363 361 351 394 361 ...
 $ Q1      : num  373 361 393 356 375 406 419 436 462 450 ...
 $ median  : num  373 396 412 381 389 434 429 465 475 490 ...
 $ Q3      : num  373 406 432 386 421 457 495 485 488 526 ...
 $ max     : num  373 472 452 503 541 546 599 569 526 569 ...
 $ percZero: num  0 0 0 0 0 0 0 0 0 0 ...
> plot(mean~age,data=lenAtAge)  # NO GOOD!!
```



```
> plot(mean~fact2num(age),data=lenAtAge,pch=16,xlab="Age",ylab="Mean TL")  # GOOD!!
```




```
> lenAtAge <- within(lenAtAge, {
   LCI <- mean-1.96*sd/sqrt(n)
   UCI <- mean+1.96*sd/sqrt(n)
 })
> head(lenAtAge)
  age  n  mean   sd min  Q1 median  Q3 max percZero   UCI   LCI
1   6  1 373.0   NA 373 373    373 373 373        0    NA    NA
2   7  9 385.1 49.8 315 361    396 406 472        0 417.6 352.6
3   8  2 412.5 55.9 373 393    412 432 452        0 490.0 335.0
4   9  9 379.7 61.0 297 356    381 386 503        0 419.6 339.8
5  10 16 404.3 48.7 351 375    389 421 541        0 428.2 380.4
6  11 24 437.5 46.8 363 406    434 457 546        0 456.2 418.8
> with(lenAtAge,plotCI(fact2num(age),mean,ui=UCI,li=LCI,
      pch=16,xlab="Age",ylab="Mean TL"))
```



# R Handout - Simple Linear Regression

**Dr. Derek Ogle**
Northland College

Mar 2014, Vermont CFWRU Workshop

## Preliminaries

```
> library(FSA)      # for Subset(), residPlot(), fitPlot()
```

```
> setwd("C:/aaaWork/Web/fishR/courses/Vermont2014/CourseMaterial/") # Derek's Computer
> d <- read.csv("Data/MnFats.csv",header=TRUE)
> d <- Subset(d,sex!="UNK")  # removed one unknown sex individual (for simplicity)
> d <- within(d, { fyear <- factor(year)
                   loglen <- log(len)
                   logwt <- log(wt)
                 } )
> view(d)
```

```
    unit year len   wt sex age logwt loglen fyear
8   MN-1 2000 401  490   M  15 6.194  5.994  2000
82  MN-1 2000 399  488   F  13 6.190  5.989  2000
94  MN-1 2000 526 1490   F  17 7.307  6.265  2000
156 MN-1 2003 442  690   M  16 6.537  6.091  2003
234 MN-1 2006 566 1280   M  20 7.155  6.339  2006
302 MN-1 2006 559 1678   F  17 7.425  6.326  2006
```

```
> clr <- rgb(0,0,0,1/4)
```

## Model Fitting

```
> lm1 <- lm(wt~len,data=d)
> names(lm1)
```

```
 [1] "coefficients"  "residuals"     "effects"       "rank"          "fitted.values"
 [6] "assign"        "qr"            "df.residual"   "xlevels"       "call"
[11] "terms"         "model"
```

```
> coef(lm1)
```

```
(Intercept)         len
  -3108.526       8.643
```

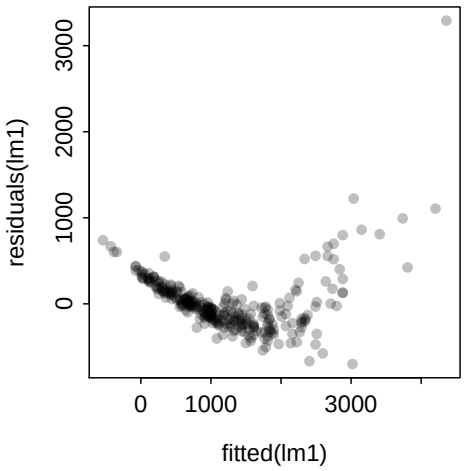
```
> residuals(lm1)[1:10]   # only show first 10 residuals
```

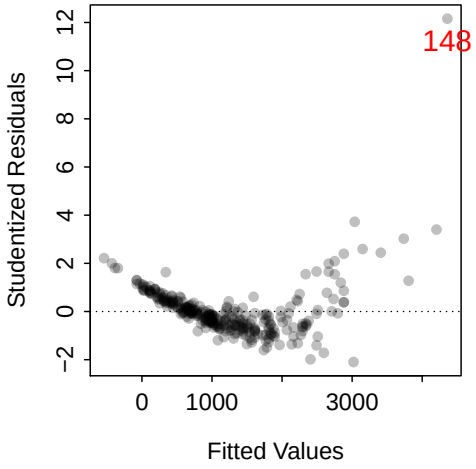
```
    1     2     3     4     5     6     7     8     9    10
669.1 301.0 254.5 305.4 173.0 193.0 145.7 132.5 139.3 223.4
```

```
> fitted(lm1)[1:10]      # only show first 10 fitted values
```

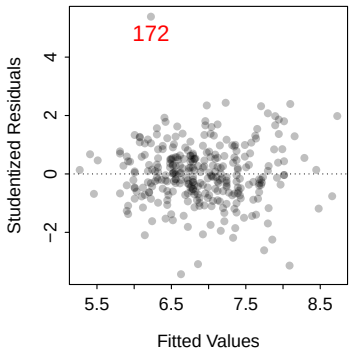
```
      1       2       3       4       5       6       7       8       9      10
-429.08   29.02  115.45  184.60  296.96  296.96  314.25  357.47  400.68  426.61
```

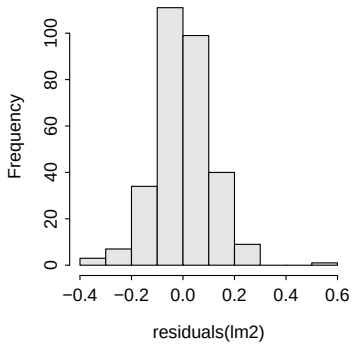
```
> plot(residuals(lm1)~fitted(lm1),pch=16,col=clr)   # manual construction -- Left
> residPlot(lm1,col=clr)                            # auto construction -- Right
```

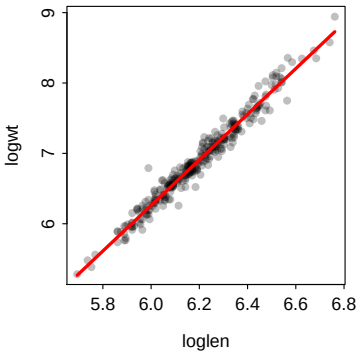



```
> lm2 <- lm(logwt~loglen,data=d)
> residPlot(lm2,col=clr)    # Left
> hist(~residuals(lm2))     # Middle
> fitPlot(lm2,col.pt=clr)   # Right
```



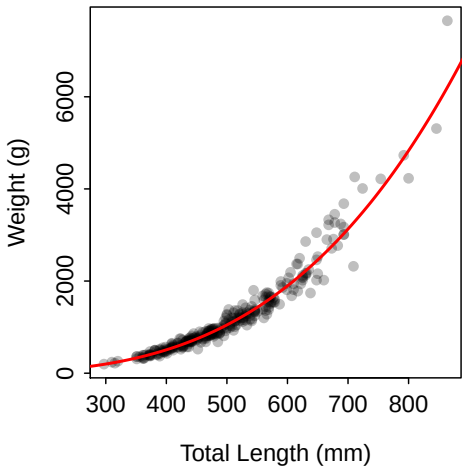




```
> d[172,]                    # Outlier?
    unit year len  wt sex age logwt loglen fyear
172 MN-1 2003 399 890   M  14 6.791  5.989  2003
> d[d$len>395 & d$len<405,]  # Fish w/ similar len
    unit year len  wt sex age logwt loglen fyear
7   MN-1 2000 396 460   F  11 6.131  5.981  2000
8   MN-1 2000 401 490   M  15 6.194  5.994  2000
55  MN-1 2000 396 510   M  11 6.234  5.981  2000
56  MN-1 2000 399 460   M  10 6.131  5.989  2000
82  MN-1 2000 399 488   F  13 6.190  5.989  2000
151 MN-1 2003 396 460   M   7 6.131  5.981  2003
172 MN-1 2003 399 890   M  14 6.791  5.989  2003
227 MN-1 2006 401 430   M  12 6.064  5.994  2006
242 MN-1 2006 399 567   F  17 6.340  5.989  2006
277 MN-1 2006 401 522   M   7 6.258  5.994  2006
> d1 <- d[-172,]            # Remove the fish
```

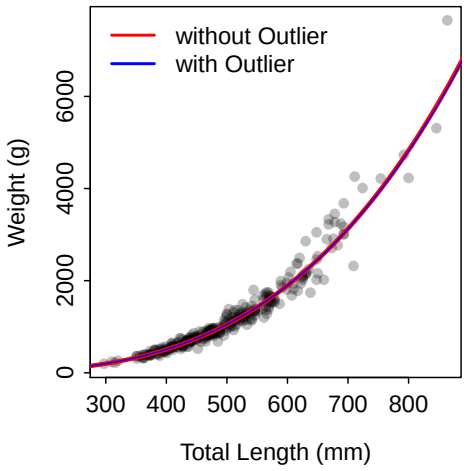
## Model Fitting

```
> lm3 <- lm(logwt~loglen,data=d1)
> anova(lm3)
Analysis of Variance Table

Response: logwt
           Df Sum Sq Mean Sq F value Pr(>F)
loglen      1  112.5     112   10097 <2e-16
Residuals 301    3.4       0
> summary(lm3)


Call:
lm(formula = logwt ~ loglen, data = d1)

Residuals:
    Min      1Q  Median      3Q     Max
-0.3687 -0.0639 -0.0046  0.0639  0.2668

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -13.2516     0.2006   -66.1   <2e-16
loglen        3.2514     0.0324   100.5   <2e-16

Residual standard error: 0.106 on 301 degrees of freedom
Multiple R-squared:  0.971,Adjusted R-squared:  0.971
F-statistic: 1.01e+04 on 1 and 301 DF,  p-value: <2e-16
> coef(lm3)
(Intercept)      loglen
    -13.252       3.251
> confint(lm3)
              2.5 %  97.5 %
(Intercept) -13.646 -12.857
loglen        3.188   3.315
> # Predict weight for 400 mm individual
> ( p1 <- predict(lm3,data.frame(loglen=log(400)),interval="prediction") )
    fit   lwr   upr
1 6.229 6.021 6.437
> exp(p1)
    fit   lwr   upr
1 507.2 411.8 624.8
```

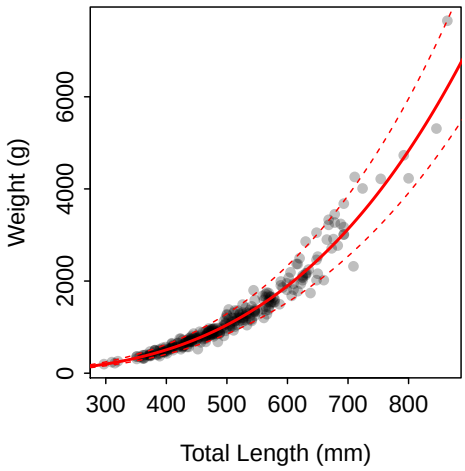
```
> plot(wt~len,data=d1,xlab="Total Length (mm)",ylab="Weight (g)",pch=16,col=clr)
> ( cf <- coef(lm3) )
(Intercept)      loglen
    -13.252       3.251
> curve(exp(cf[1])*x^cf[2],from=275,to=900,col="red",lwd=2,add=TRUE)
```




```
> plot(wt~len,data=d1,xlab="Total Length (mm)",ylab="Weight (g)",pch=16,col=clr)
> curve(exp(cf[1])*x^cf[2],from=275,to=900,col="red",lwd=3,add=TRUE)
> ( cfOut <- coef(lm2) )
(Intercept)      loglen
     -13.18        3.24
> curve(exp(cfOut[1])*x^cfOut[2],from=275,to=900,col="blue",lwd=1,add=TRUE)
> legend("topleft",c("without Outlier","with Outlier"),col=c("red","blue"),lwd=2,bty="n")
```




```
> # Predict weight for all lengths b/w 275 and 900 mm
> xs <- seq(275,900,1)
> pW <- exp(predict(lm3,data.frame(loglen=log(xs)),interval="prediction"))
> pW[1:5,]  # first five rows
    fit   lwr   upr
1 150.0 121.4 185.3
2 151.8 122.9 187.5
3 153.6 124.3 189.7
4 155.4 125.8 191.9
5 157.2 127.3 194.2
> plot(wt~len,data=d1,xlab="Total Length (mm)",ylab="Weight (g)",pch=16,col=clr)
> lines(pW[,"fit"]~xs,col="red",lwd=2)
> lines(pW[,"lwr"]~xs,col="red",lty=2)
> lines(pW[,"upr"]~xs,col="red",lty=2)
```


Weight (g)
0
2000
4000
6000
300
400
500
600
700
800
Total Length (mm)

# R Handout - One-Way ANOVA

**Dr. Derek Ogle**
Northland College

Mar 2014, Vermont CFWRU Workshop

## Preliminaries

```
> library(FSA)      # for Subset(), wrVal(), psdVal()
> library(car)      # for outlierTest(), leveneTest()
> library(multcomp) # for glht(), mcp()
> library(plotrix)  # for plotCI(), cld()
```

```
> setwd("C:/aaaWork/Web/fishR/courses/Vermont2014/CourseMaterial/") # Derek's Computer
> d <- read.csv("Data/Keuska99.csv",header=TRUE)
> str(d)
'data.frame': 2391 obs. of  10 variables:
 $ species  : Factor w/ 4 levels "BG","LMB","WAE",..: 1 1 1 1 1 1 1 1 1 1 ...
 $ date     : Factor w/ 12 levels "3/31/1999","4/1/1999",..: 6 6 6 6 4 4 6 6 6 6 ...
 $ geartype : Factor w/ 2 levels "BOOM SHOCKER",..: 1 1 1 1 1 1 1 1 1 1 ...
 $ sex      : Factor w/ 3 levels "F","M","U": NA NA NA NA NA NA NA NA NA NA ...
 $ inches   : num  0.9 1 1 1 1 1 1.1 1.1 1.1 1.1 ...
 $ pounds   : num  0.001 0.001 0.001 0.001 0.001 0.001 0 0 0.001 0.001 ...
 $ mm       : int  23 25 25 25 25 25 28 28 28 28 ...
 $ grams    : int  0 0 0 0 1 1 0 0 0 0 ...
 $ age      : int  NA NA NA NA NA NA NA NA NA NA ...
 $ agestruct: Factor w/ 2 levels "SCALE","SPINE": NA NA NA NA NA NA NA NA NA NA ...
> levels(d$species)
[1] "BG"  "LMB" "WAE" "YEP"
>
> # Focus on LMB and remove some variables (only to make the handout easier to read)
> lmb <- Subset(d,species=="LMB",select=c("species","geartype","mm","grams"))
>
> # Identify which fish had both mm and grams recorded, show first 10
> complete.cases(lmb[,c("mm","grams")])[1:10]
 [1]  TRUE FALSE  TRUE  TRUE FALSE FALSE FALSE  TRUE  TRUE  TRUE
>
> # Retain only those with both measures
> lmb <- Subset(lmb,complete.cases(lmb[,c("mm","grams")]))
> str(lmb)
'data.frame': 541 obs. of  4 variables:
 $ species : Factor w/ 1 level "LMB": 1 1 1 1 1 1 1 1 1 1 ...
 $ geartype: Factor w/ 2 levels "BOOM SHOCKER",..: 1 1 1 1 1 1 1 1 1 1 ...
 $ mm      : int  81 91 91 107 107 112 132 137 137 137 ...
 $ grams   : int  6 8 8 9 14 15 28 24 24 36 ...
> view(lmb)
    species     geartype  mm grams
79      LMB BOOM SHOCKER 178    67
153     LMB BOOM SHOCKER 241   172
289     LMB BOOM SHOCKER 284   304
306     LMB     FYKE NET 292   297
415     LMB BOOM SHOCKER 315   428
512     LMB BOOM SHOCKER 361   600
```

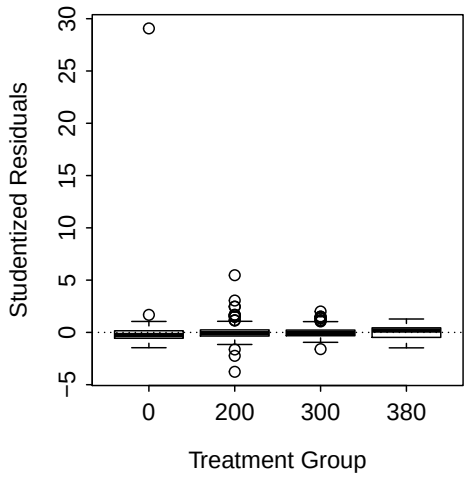
## Data Prep for Relative Weight Analysis

```
> ( wsLMB <- wsVal("Largemouth Bass") )
           species  units   type ref    int slope quad min.len max.len measure
75 Largemouth Bass metric linear  75 -5.528 3.273   NA     150      NA      TL
   method comment       source
75    RLP    none Henson, 1991
> lmb1 <- Subset(lmb,mm>=150)
>
> lmb1 <- within(lmb1,{
           Ws <- 10^(wsLMB$int)*mm^wsLMB$slope
           Wr <- grams/Ws*100
         })
> view(lmb1)
    species     geartype  mm grams     Wr     Ws
69      LMB BOOM SHOCKER 180    69  96.68  71.37
145     LMB BOOM SHOCKER 246   169  85.18 198.39
288     LMB BOOM SHOCKER 290   371 109.13 339.96
363     LMB BOOM SHOCKER 307   380  92.76 409.64
437     LMB     FYKE NET 330   519 100.02 518.91
450     LMB BOOM SHOCKER 335   516  94.66 545.09
```

```
> ( wsPSD <- psdVal("Largemouth Bass") )
     zero     stock   quality preferred memorable    trophy
        0       200       300       380       510       630
> lmb1 <- lencat(~mm,data=lmb1,breaks=wsPSD)
> view(lmb1)
    species     geartype  mm grams    Wr     Ws LCat
10      LMB BOOM SHOCKER 152    40 97.47  41.04    0
178     LMB BOOM SHOCKER 257   213 93.04 228.93  200
271     LMB BOOM SHOCKER 284   293 92.29 317.48  200
314     LMB BOOM SHOCKER 297   303 82.43 367.56  200
325     LMB BOOM SHOCKER 300   331 87.14 379.86  300
467     LMB     FYKE NET 343   516 87.63 588.87  300
> xtabs(~LCat,data=lmb1)
LCat
  0 200 300 380 510
 80 242 184  18   3
>
> lmb1 <- lencat(~mm,data=lmb1,breaks=c(0,200,300,380,1000))
> view(lmb1)
    species     geartype  mm grams     Wr     Ws LCat LCat1
20      LMB     FYKE NET 157    47 103.02  45.62    0     0
77      LMB BOOM SHOCKER 191    81  93.47  86.66    0     0
104     LMB     FYKE NET 226   138  91.81 150.31  200   200
264     LMB BOOM SHOCKER 282   290  93.48 310.22  200   200
390     LMB BOOM SHOCKER 312   411  95.16 431.88  300   300
513     LMB BOOM SHOCKER 399  1009 104.45 966.02  380   380
> xtabs(~LCat1,data=lmb1)
LCat1
  0 200 300 380
 80 242 184  21
```

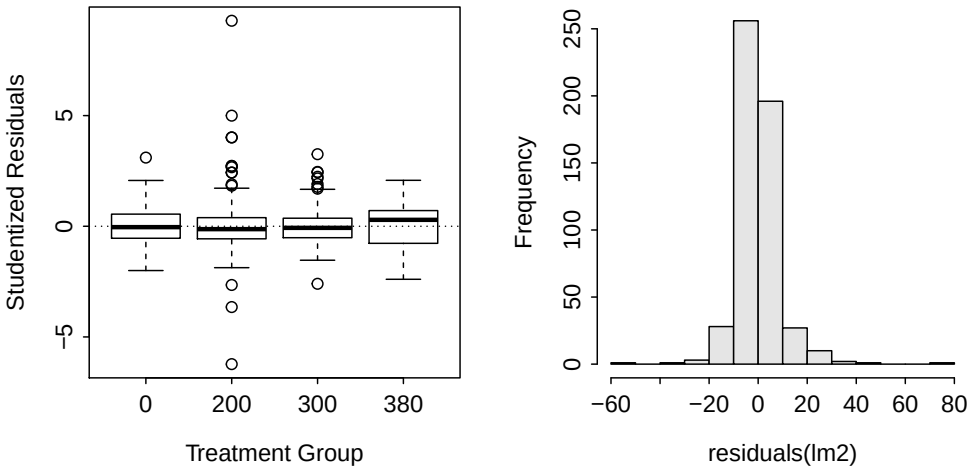
## One-Way ANOVA of Wr by PSD Category

```
> lm1 <- lm(Wr~LCat1,data=lmb1)
> residPlot(lm1)
```




```
> outlierTest(lm1)
    rstudent unadjusted p-value Bonferonni p
80     29.07         3.266e-111   1.721e-108
122     5.47          6.986e-08    3.682e-05
> lmb1[80,]                           # Outlier?
   species     geartype  mm grams  Wr   Ws LCat LCat1
80     LMB BOOM SHOCKER 198   350 359 97.5    0     0
> lmb1[lmb1$mm>=195 & lmb1$mm<=205,]  # Fish w/ similar lengths
   species     geartype  mm grams     Wr     Ws LCat LCat1
78     LMB BOOM SHOCKER 196    87  92.25  94.31    0     0
79     LMB     FYKE NET 196    92  97.55  94.31    0     0
80     LMB BOOM SHOCKER 198   350 358.99  97.50    0     0
81     LMB     FYKE NET 203    97  91.69 105.79  200   200
> lmb2 <- lmb1[-80,]                  # Remove the fish
```

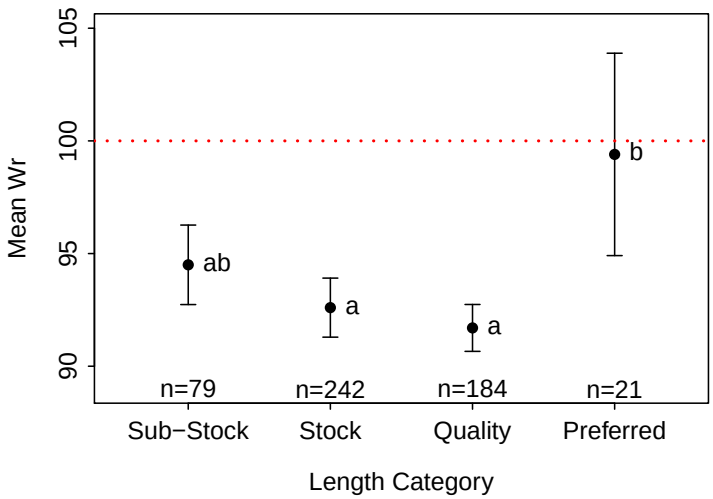
```
> lm2 <- lm(Wr~LCat1,data=lmb2)
> leveneTest(lm2)
Levene's Test for Homogeneity of Variance (center = median)
       Df F value Pr(>F)
group   3    1.72   0.16
      522
> residPlot(lm2)        # Left
> hist(~residuals(lm2)) # Right
```




```
> anova(lm2)
Analysis of Variance Table

Response: Wr
           Df Sum Sq Mean Sq F value Pr(>F)
LCat1       3   1342     447    5.47  0.001
Residuals 522  42673      82
> mc1 <- glht(lm2,mcp(LCat1="Tukey"))
> summary(mc1)


 Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: Tukey Contrasts


Fit: lm(formula = Wr ~ LCat1, data = lmb2)

Linear Hypotheses:
               Estimate Std. Error t value Pr(>|t|)
200 - 0 == 0     -1.862      1.172   -1.59   0.3660
300 - 0 == 0     -2.743      1.216   -2.26   0.1004
380 - 0 == 0      4.883      2.220    2.20   0.1139
300 - 200 == 0   -0.882      0.884   -1.00   0.7368
380 - 200 == 0    6.745      2.057    3.28   0.0055
380 - 300 == 0    7.626      2.083    3.66   0.0013
(Adjusted p values reported -- single-step method)
> cld(mc1)
   0  200  300  380
"ab"  "a"  "a"  "b"
> confint(mc1)


 Simultaneous Confidence Intervals

Multiple Comparisons of Means: Tukey Contrasts


Fit: lm(formula = Wr ~ LCat1, data = lmb2)
```

```
Quantile = 2.54
95% family-wise confidence level


Linear Hypotheses:
               Estimate lwr    upr
200 - 0 == 0   -1.862   -4.838  1.115
300 - 0 == 0   -2.743   -5.833  0.346
380 - 0 == 0    4.883   -0.756 10.522
300 - 200 == 0 -0.882   -3.128  1.365
380 - 200 == 0  6.745    1.519 11.970
380 - 300 == 0  7.626    2.336 12.917
```

```
> ( sumWr <- Summarize(Wr~LCat1,data=lmb2,digits=1) )
  LCat1   n mean   sd  min   Q1 median    Q3 max percZero
1     0  79 94.5  8.0 76.6 89.6   94.1  99.4 122        0
2   200 242 92.6 10.4 38.4 87.5   91.4  96.1 170        0
3   300 184 91.7  7.2 68.4 87.1   91.1  95.0 121        0
4   380  21 99.4 10.5 78.3 92.6  102.0 106.0 118        0
> sumWr <- within(sumWr, {
   LCI <- mean-1.96*sd/sqrt(n)
   UCI <- mean+1.96*sd/sqrt(n)
 })
> sumWr
  LCat1   n mean   sd  min   Q1 median    Q3 max percZero    UCI   LCI
1     0  79 94.5  8.0 76.6 89.6   94.1  99.4 122        0  96.26 92.74
2   200 242 92.6 10.4 38.4 87.5   91.4  96.1 170        0  93.91 91.29
3   300 184 91.7  7.2 68.4 87.1   91.1  95.0 121        0  92.74 90.66
4   380  21 99.4 10.5 78.3 92.6  102.0 106.0 118        0 103.89 94.91
> with(sumWr,plotCI(1:4,mean,ui=UCI,li=LCI,pch=16,xlim=c(0.5,4.5),xaxt="n",
                    ylim=c(89,105),xlab="Length Category",ylab="Mean Wr"))
> axis(1,1:4,c("Sub-Stock","Stock","Quality","Preferred"))
> abline(h=100,col="red",lty=3,lwd=2)
> text(1:4,sumWr$mean,c("ab","a","a","b"),pos=c(4,4,4,4))
> text(1:4,89,paste("n=",sumWr$n,sep=""))
```



# R Handout - Nonlinear Models

**Dr. Derek Ogle**
Northland College

Mar 2014, Vermont CFWRU Workshop

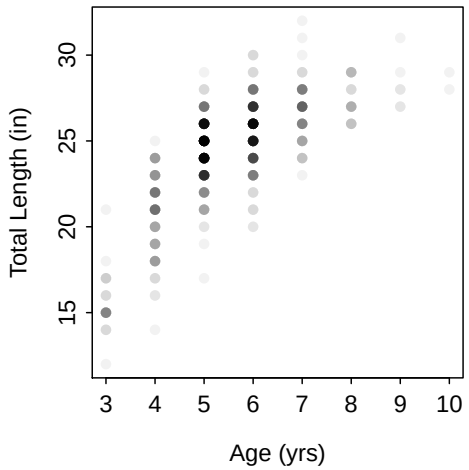
## Preliminaries

```
> library(FSA)      # for Subset(), vbModels(), vbStarts(), vbFuns(), confint()
> library(nlstools) # for nlsBoot()
```

```
> setwd("C:/aaaWork/Web/fishR/courses/Vermont2014/CourseMaterial/") # Derek's Computer
> d <- read.csv("Data/TroutBR.csv",header=TRUE)
> str(d)
'data.frame': 851 obs. of  3 variables:
 $ tl     : int  16 16 17 17 17 17 17 17 17 17 ...
 $ age    : int  4 4 2 3 3 3 3 3 3 4 ...
 $ species: Factor w/ 2 levels "Brown","Rainbow": 1 1 1 1 1 1 1 1 1 1 ...
> rbt <- Subset(d,species=="Rainbow")
> str(rbt)
'data.frame': 627 obs. of  3 variables:
 $ tl     : int  12 14 14 14 14 15 15 15 15 15 ...
 $ age    : int  3 3 3 3 4 3 3 3 3 3 ...
 $ species: Factor w/ 1 level "Rainbow": 1 1 1 1 1 1 1 1 1 1 ...
```

```
> # Declare some constants
> xlbl <- "Age (yrs)"
> ylbl <- "Total Length (in)"
> clr <- rgb(0,0,0,1/20)
```

```
> plot(tl~age,data=rbt,xlab=xlbl,ylab=ylbl,pch=16,col=clr)
```



# Fit Typical Model

```
> vbModels()
```

**FSA von Bertalanffy Parametrizations**

Original: $E(L_t) = L_\infty - (L_\infty - L_0)\, e^{-Kt}$

Typical: $E(L_t) = L_\infty\left(1 - e^{-K(t - t_0)}\right)$

GallucciQuinn: $E(L_t) = \frac{\omega}{K}\left(1 - e^{-K(t - t_0)}\right)$

Mooij: $E(L_t) = L_\infty - (L_\infty - L_0)\, e^{-\frac{\omega}{L_\infty} t}$

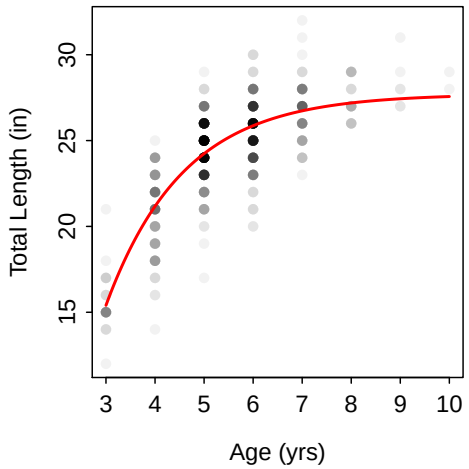
Schnute: $E(L_t) = L_1 + (L_2 - L_1)\frac{1 - e^{-K(t - t_1)}}{1 - e^{-K(t_2 - t_1)}}$

Francis: $E(L_t) = L_1 + (L_3 - L_1)\frac{1 - r^{2\frac{t - t_1}{t_3 - t_1}}}{1 - r^2}$

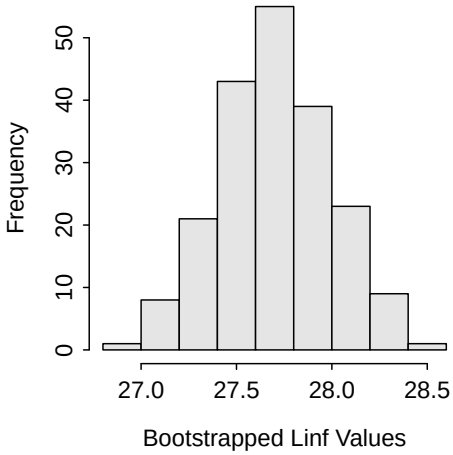
where $r = \frac{L_3 - L_2}{L_2 - L_1}$

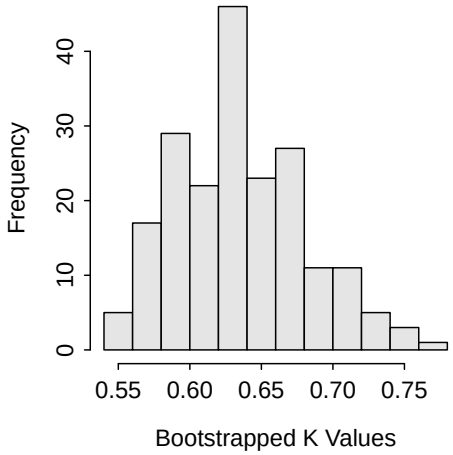
```
> ( svb1 <- vbStarts(tl~age,data=rbt,type="typical") )
$Linf
[1] 28.67

$K
[1] 0.5242

$t0
[1] -1.429
> fit1 <- nls(tl~Linf*(1-exp(-K*(age-t0))),data=rbt,start=svb1)
> summary(fit1)

Formula: tl ~ Linf * (1 - exp(-K * (age - t0)))

Parameters:
     Estimate Std. Error t value Pr(>|t|)
Linf  27.7118     0.2838    97.6   <2e-16
K      0.6324     0.0425    14.9   <2e-16
t0     1.7169     0.1016    16.9   <2e-16

Residual standard error: 1.78 on 624 degrees of freedom

Number of iterations to convergence: 5
Achieved convergence tolerance: 5.38e-08
> confint(fit1)
Waiting for profiling to be done...
       2.5%   97.5%
Linf 27.192 28.3280
K     0.550  0.7192
t0    1.493  1.8999
```

```
> ( cf <- coef(fit1) )
   Linf       K      t0
27.7118  0.6324  1.7169
> plot(tl~age,data=rbt,xlab=xlbl,ylab=ylbl,pch=16,col=clr)
> curve(cf["Linf"]*(1-exp(-cf["K"]*(x-cf["t0"]))),
       from=3,to=10,n=500,lwd=2,col="red",add=TRUE)
```

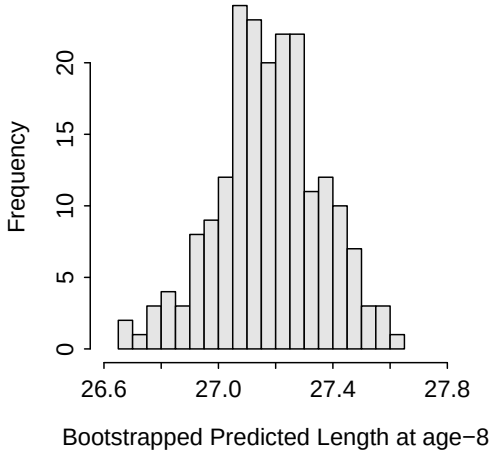



```
> boot1 <- nlsBoot(fit1,niter=200)   # niter should be nearer 1000
> ests1 <- boot1$coefboot
> ests1[1:5,]  # first five rows
      Linf      K    t0
[1,] 27.72 0.6334 1.743
[2,] 28.19 0.5566 1.522
[3,] 28.27 0.5645 1.591
[4,] 27.84 0.6338 1.752
[5,] 27.69 0.6158 1.670
> hist(~ests1[,"Linf"],xlab="Bootstrapped Linf Values")
> hist(~ests1[,"K"],xlab="Bootstrapped K Values")
```



```
> confint(boot1)
     95% LCI 95% UCI
Linf 27.1020 28.2683
K     0.5625  0.7303
t0    1.5242  1.9057
>
> predict(fit1, data.frame(age=8))
[1] 27.19
> pv <- ests1[,"Linf"]*(1-exp(-ests1[,"K"]*(8-ests1[,"t0"])))
> hist(~pv,breaks=20,xlim=c(26.6,27.8),xlab="Bootstrapped Predicted Length at age-8")
```




```
> quantile(pv,c(0.025,0.975))
 2.5% 97.5%
26.79 27.52
```

# R Handout - Indicator Variable Regression

Dr. Derek Ogle
Northland College

Mar 2014, Vermont CFWRU Workshop

## Preliminaries

```
> library(FSA)      # for Subset(), fitPlot()
```

## Lake Trout Data

Curtis (1990) examined the population dynamics related to the recovery of an offshore lake trout population near Stannard Rock, Lake Superior. Relative abundance of lake trout greater than 43.2 cm long was recorded as the CPE (fish caught per 50,000 m of 114.3-mm-mesh gill net) of each age group in each year.

| | Age-Group | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Year | VI | VII | VIII | IX | X | XI | XII | XIII | XIV |
| 1959 | 64 | 219 | 241 | 121 | 33 | 9 | 1 | 0.5 | 1 |
| 1963 | 129 | 339 | 331 | 192 | 70 | 16 | 0.5 | 0.5 | 0.5 |
| 1964 | 149 | 524 | 515 | 201 | 63 | 18 | 2 | 0.5 | 0.5 |
| 1965 | 75 | 379 | 501 | 328 | 133 | 39 | 11 | 1 | 0.5 |
| 1966 | 149 | 488 | 459 | 172 | 64 | 22 | 5 | 0.5 | 0.5 |
| 1967 | 63 | 368 | 287 | 130 | 55 | 19 | 6 | 0.5 | 0.5 |
| 1968 | 50 | 215 | 259 | 141 | 55 | 18 | 5 | 1 | 0.5 |
| 1969 | 45 | 150 | 153 | 76 | 23 | 6 | 0.5 | 0.5 | 0.5 |
| 1973 | 101 | 759 | 1268 | 1116 | 491 | 141 | 40 | 4 | 0.5 |
| 1974 | 151 | 733 | 1114 | 1092 | 571 | 163 | 50 | 9 | 5 |
| 1975 | 109 | 901 | 1517 | 1606 | 1076 | 342 | 117 | 12 | 7 |
| 1976 | 53 | 604 | 1204 | 1560 | 1146 | 396 | 156 | 18 | 10 |
| 1977 | 157 | 867 | 1343 | 1410 | 1031 | 417 | 192 | 17 | 7 |
| 1978 | 89 | 735 | 1307 | 1623 | 1150 | 445 | 198 | 18 | 14 |
| 1979 | 29 | 299 | 718 | 1268 | 1195 | 585 | 300 | 36 | 14 |

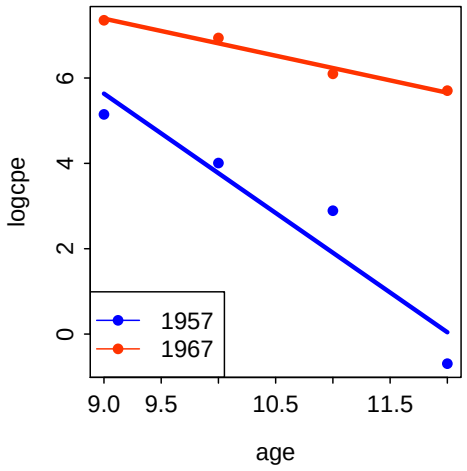
```
> ages <- 9:12
> yc67 <- c(1560,1031,445,300)
> yc64 <- c(1116,571,342,156)
> yc57 <- c(172,55,18,0.5)
> yc54 <- c(192,63,39,5)
> d <- data.frame(yc=factor(rep(c(1967,1964,1957,1954),each=4)),
                  age=rep(ages,times=4),
                  cpe=c(yc67,yc64,yc57,yc54))
> d <- within(d,logcpe <- log(cpe))
> d
     yc age    cpe  logcpe
1  1967   9 1560.0  7.3524
2  1967  10 1031.0  6.9383
3  1967  11  445.0  6.0981
```

```
4  1967  12  300.0  5.7038
5  1964   9 1116.0  7.0175
6  1964  10  571.0  6.3474
7  1964  11  342.0  5.8348
8  1964  12  156.0  5.0499
9  1957   9  172.0  5.1475
10 1957  10   55.0  4.0073
11 1957  11   18.0  2.8904
12 1957  12    0.5 -0.6931
13 1954   9  192.0  5.2575
14 1954  10   63.0  4.1431
15 1954  11   39.0  3.6636
16 1954  12    5.0  1.6094
```

## Model Fitting

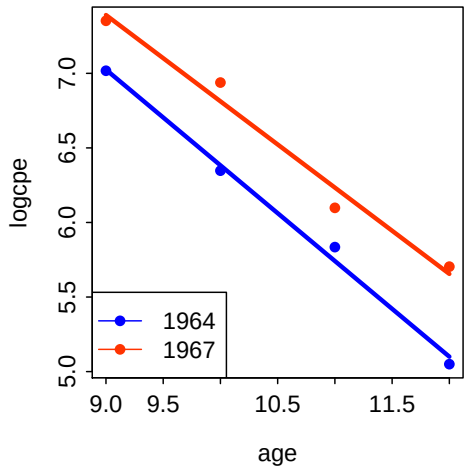
```
> # Compare 1957 and 1967 year-classes
> lm1 <- lm(logcpe~age*yc,data=Subset(d,yc %in% c(1957,1967)))
> anova(lm1)
Analysis of Variance Table

Response: logcpe
          Df Sum Sq Mean Sq F value Pr(>F)
age        1  14.91   14.91   32.41 0.0047
yc         1  27.16   27.16   59.03 0.0015
age:yc     1   4.13    4.13    8.98 0.0401
Residuals  4   1.84    0.46
> summary(lm1)

Call:
lm(formula = logcpe ~ age * yc, data = Subset(d, yc %in% c(1957,
    1967)))

Residuals:
      1       2       3       4       5       6       7       8
-0.0386  0.1258 -0.1358  0.0486 -0.4864  0.2374  0.9843 -0.7353

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   22.409      3.203    7.00   0.0022
age           -1.864      0.303   -6.14   0.0036
yc1967        -9.810      4.530   -2.17   0.0963
age:yc1967     1.285      0.429    3.00   0.0401

Residual standard error: 0.678 on 4 degrees of freedom
Multiple R-squared:  0.962,Adjusted R-squared:  0.933
F-statistic: 33.5 on 3 and 4 DF,  p-value: 0.00272
> confint(lm1)
                2.5 % 97.5 %
(Intercept)  13.51506 31.303
age          -2.70615 -1.022
yc1967      -22.38791  2.768
age:yc1967    0.09413  2.476
```

```
> fitPlot(lm1,legend="bottomleft")
```




```
> # Compare 1964 and 1967 year-classes
> lm2 <- lm(logcpe~age*yc,data=Subset(d,yc %in% c(1964,1967)))
> anova(lm2)
Analysis of Variance Table

Response: logcpe
          Df Sum Sq Mean Sq F value  Pr(>F)
age        1   3.72    3.72  293.92 6.8e-05
yc         1   0.42    0.42   33.53  0.0044
age:yc     1   0.01    0.01    0.78  0.4265
Residuals  4   0.05    0.01
> summary(lm2)

Call:
lm(formula = logcpe ~ age * yc, data = Subset(d, yc %in% c(1964,
    1967)))

Residuals:
       1        2        3        4        5        6        7        8
-0.03863  0.12583 -0.13576  0.04856 -0.00721 -0.03578  0.09320 -0.05021

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  12.7987     0.5314   24.08  1.8e-05
age          -0.6416     0.0503  -12.75  0.00022
yc1967       -0.2001     0.7515   -0.27  0.80324
age:yc1967    0.0629     0.0712    0.88  0.42650

Residual standard error: 0.113 on 4 degrees of freedom
Multiple R-squared:  0.988,Adjusted R-squared:  0.979
F-statistic:  109 on 3 and 4 DF,  p-value: 0.000271
> confint(lm2)
              2.5 %  97.5 %
(Intercept) 11.3233 14.2741
age         -0.7813 -0.5018
yc1967      -2.2866  1.8865
age:yc1967  -0.1347  0.2605
```

```
> fitPlot(lm2,legend="bottomleft")
```




```
> # Fit without the insignificant interaction term as a demonstration
> lm2a <- lm(logcpe~age+yc,data=Subset(d,yc %in% c(1964,1967)))
> anova(lm2a)
Analysis of Variance Table

Response: logcpe
          Df Sum Sq Mean Sq F value  Pr(>F)
age        1   3.72    3.72   307.3 1.1e-05
yc         1   0.42    0.42    35.1   0.002
Residuals  5   0.06    0.01
> summary(lm2a)

Call:
lm(formula = logcpe ~ age + yc, data = Subset(d, yc %in% c(1964,
    1967)))

Residuals:
      1       2       3       4       5       6       7       8
-0.0858  0.1101 -0.1200  0.0958  0.0400 -0.0200  0.0775 -0.0974

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  12.4683     0.3695   33.74  4.3e-07
age          -0.6101     0.0348  -17.53  1.1e-05
yc1967        0.4608     0.0778    5.92    0.002

Residual standard error: 0.11 on 5 degrees of freedom
Multiple R-squared:  0.986,Adjusted R-squared:  0.98
F-statistic:  171 on 2 and 5 DF,  p-value: 2.49e-05
> confint(lm2a)
              2.5 %  97.5 %
(Intercept) 11.5184 13.4182
age         -0.6995 -0.5206
yc1967       0.2607  0.6608
```

# R Handout - K-Pass Removal

**Dr. Derek Ogle**
Northland College

Mar 2014, Vermont CFWRU Workshop

## Preliminaries

```
> library(FSA)      # for removal()
> setwd("C:/aaaWork/Web/fishR/courses/Vermont2014/CourseMaterial/") # Derek's Computer
```

## Single Removal Estimates

```
> d <- c(71,48,40)
> res <- removal(d,type="CarleStrub")
> summary(res)
The Carle & Strub's K-Pass Removal Method method was used.
   Estimate Std. Error
No 252.0000    46.3817
p    0.2809     0.0719
> confint(res)
   95% LCI  95% UCI
No  161.09 342.9065
p     0.14   0.4218
```

## Multiple Removal Estimates – Data in Wide Format

```
> d <- read.csv("Data/JonesStockwell2.csv",header=TRUE)
> head(d)
  species site   age0 first second third
1 rainbow  A15 Age-1+    17     10     7
2 rainbow   A9  Age-0    17     12     8
3 rainbow  A10  Age-0    25     10     8
4 rainbow   B1  Age-0    69     43    23
5 rainbow   B1 Age-1+     6      2     1
6 rainbow Bio1  Age-0    38     20    20
```

```
> ( res <- apply(d[,4:6],MARGIN=1,FUN=removal,type="CarleStrub",just.ests=TRUE) )
         [,1]    [,2]    [,3]      [,4]   [,5]      [,6]    [,7]      [,8]    [,9]
No    41.0000 48.0000 48.0000 165.00000 9.0000 109.00000 31.0000 930.00000 31.0000
p      0.4304  0.3776  0.5119   0.42994 0.6923   0.33766  0.4737   0.26271  0.5185
No.se  7.0691 10.5594  4.7448  14.21881 0.6903  20.56207  4.7616 102.15821  3.6697
p.se   0.1303  0.1334  0.1037   0.06499 0.1726   0.09617  0.1382   0.03914  0.1275
         [,10]   [,11]     [,12]   [,13]     [,14]     [,15]     [,16]   [,17]
No    252.0000 11.0000 341.00000 46.0000 145.00000 148.00000 124.00000 14.0000
p       0.2809  0.6471   0.42266  0.4270   0.41786   0.36066   0.33969  0.5200
No.se  46.3817  1.0198  21.35046  7.6421  14.33083  20.63412  21.63915  2.4450
p.se    0.0719  0.1700   0.04584  0.1238   0.07094   0.07865   0.08978  0.1892
       [,18]  [,19]  [,20] [,21]   [,22]   [,23]   [,24]     [,25]   [,26]  [,27]
No    3.0000 3.0000 7.0000     2 14.0000 38.0000 24.0000 140.00000 24.0000 4.0000
p     0.7500 0.7500 0.7000     1  0.5417  0.4714  0.5897   0.71354  0.7059 0.8000
No.se 0.2659 0.2659 0.5783     0  2.1563  5.3417  2.1307   2.35888  1.0292 0.2052
p.se  0.2659 0.2659 0.1928   NaN  0.1820  0.1254  0.1276   0.04197  0.1029 0.2052
       [,28]    [,29]    [,30]  [,31] [,32]   [,33]     [,34]     [,35]   [,36]  [,37]
No    8.0000 13.0000 15.0000 2.0000     1 13.0000 546.00000 192.00000 17.0000 1.0000
p     0.6667  0.8125  0.3235 0.3333     1  0.5217   0.09486   0.38083  0.2326 0.5000
No.se 0.7687  0.3308  8.3835 2.8664     0  2.3325 487.43490  20.69024 17.5787 0.7338
p.se  0.1922  0.1103  0.2673 0.7166   NaN  0.1957   0.09356   0.06628  0.3133 0.7338
       [,38] [,39]  [,40]
No    5.0000     1 6.0000
p     0.5556     1 0.6000
No.se 1.1886     0 1.0024
p.se  0.2972   NaN 0.2506
```

```
> # transpose the result and make as a data.frame, add specific info from d, add CIs
> res <- data.frame(t(res))
> res <- cbind(d[,1:3],res)
> res <- within(res,{
   No.LCI <- No-1.96*No.se
   No.UCI <- No+1.96*No.se
 })
> head(res)                  # first 6 rows
  species site   age0  No      p   No.se    p.se No.UCI  No.LCI
1 rainbow  A15 Age-1+  41 0.4304  7.0691 0.13027  54.86  27.145
2 rainbow   A9  Age-0  48 0.3776 10.5594 0.13343  68.70  27.304
3 rainbow  A10  Age-0  48 0.5119  4.7448 0.10367  57.30  38.700
4 rainbow   B1  Age-0 165 0.4299 14.2188 0.06499 192.87 137.131
5 rainbow   B1 Age-1+   9 0.6923  0.6903 0.17257  10.35   7.647
6 rainbow Bio1  Age-0 109 0.3377 20.5621 0.09617 149.30  68.698
```

# R Handout - Age Bias and Precision

Dr. Derek Ogle
Northland College

Mar 2014, Vermont CFWRU Workshop

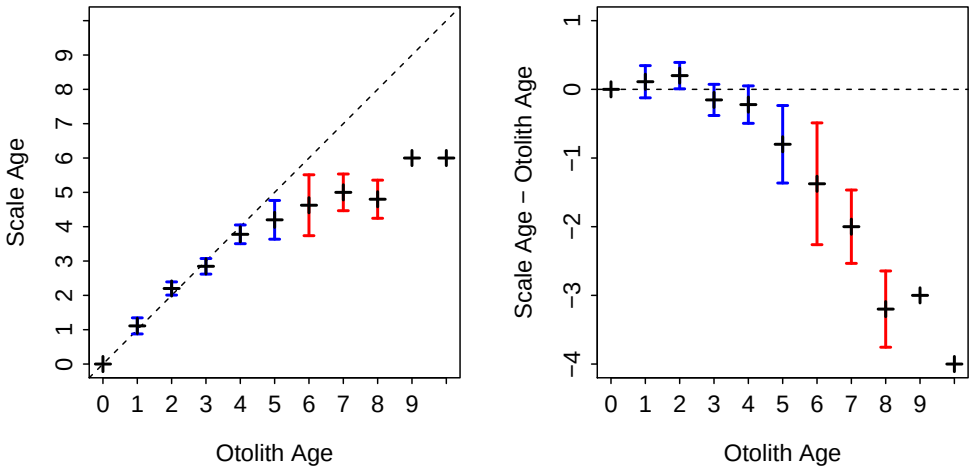
## Preliminaries

```
> library(FSA)      # for ageBias(), agePrecision()
> setwd("C:/aaaWork/Web/fishR/courses/Vermont2014/CourseMaterial/")
```

## Lake Huron Alewife – Otolith-Scale Bias?

```
> d <- read.csv("Data/AlewifeLH.csv",header=TRUE)
> str(d)
'data.frame': 104 obs. of  2 variables:
 $ otoliths: int  0 0 1 1 1 1 1 1 1 1 ...
 $ scales  : int  0 0 0 1 1 1 1 1 1 1 ...
> ab1 <- ageBias(otoliths~scales,data=d,col.lab="Otolith Age",row.lab="Scale Age")
> summary(ab1,what="symmetry",flip.table=TRUE)
Raw agreement table (square & flipped)
         Otolith Age
Scale Age  0  1  2  3  4  5  6  7  8  9 10
       10  -  -  -  -  -  -  -  -  -  -  -
       9   -  -  -  -  -  -  -  -  -  -  -
       8   -  -  -  -  -  -  -  -  -  -  -
       7   -  -  -  -  -  -  -  -  -  -  -
       6   -  -  -  -  -  -  2  1  -  1  2
       5   -  -  -  -  1  4  2  5  4  -  -
       4   -  -  -  - 12  4  3  1  1  -  -
       3   -  -  4 11  5  2  1  -  -  -  -
       2   -  3 16  2  -  -  -  -  -  -  -
       1   - 14  -  -  -  -  -  -  -  -  -
       0   2  1  -  -  -  -  -  -  -  -  -

Bowker's (Hoenig's) Test of Symmetry
 df chi.sq        p
 16  34.47 0.004698
> summary(ab1,what="bias")
Summary of Scale Age by Otolith Age
 otoliths  n min max mean     SE      t   adj.p   sig   LCI  UCI
        0  2   0   0 0.00     NA     NA      NA FALSE    NA   NA
        1 18   0   2 1.11 0.1111   1.00 0.33138 FALSE 0.877 1.35
        2 20   2   3 2.20 0.0918   2.18 0.16836 FALSE 2.008 2.39
        3 13   2   3 2.85 0.1041  -1.48 0.33098 FALSE 2.619 3.07
        4 18   3   5 3.78 0.1292  -1.72 0.31112 FALSE 3.505 4.05
        5 10   3   5 4.20 0.2494  -3.21 0.05354 FALSE 3.636 4.76
        6  8   3   6 4.62 0.3750  -3.67 0.04801  TRUE 3.738 5.51
        7  7   4   6 5.00 0.2182  -9.16 0.00071  TRUE 4.466 5.53
        8  5   4   5 4.80 0.2000 -16.00 0.00071  TRUE 4.245 5.36
        9  1   6   6 6.00     NA     NA      NA FALSE    NA   NA
       10  2   6   6 6.00     NA     NA      NA FALSE    NA   NA
```

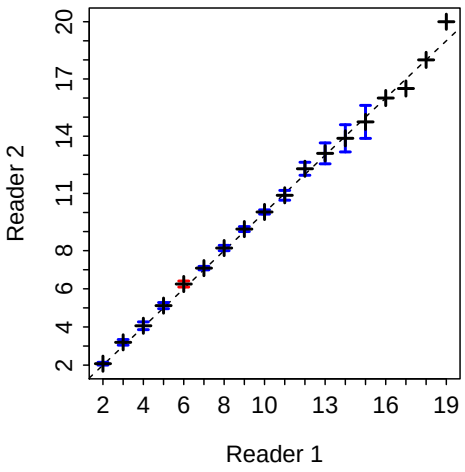
```
> plot(ab1)                    # LEFT
> plot(ab1,difference=TRUE)    # RIGHT
```

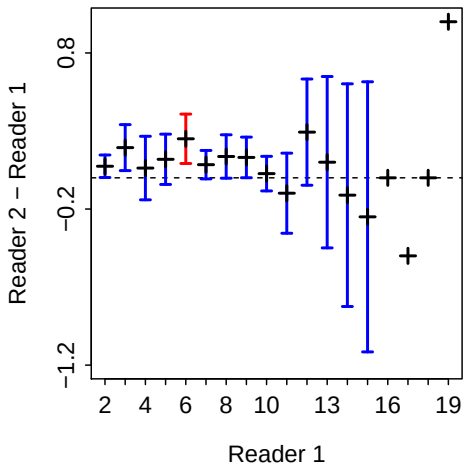



## Striped Bass – Reader Precision?

```
> sb <- read.csv("Data/StripedBass4.csv",header=TRUE)
> str(sb)

'data.frame': 1202 obs. of  2 variables:
 $ reader1: int  2 2 2 2 2 2 2 2 2 2 ...
 $ reader2: int  2 2 2 2 2 2 2 2 2 2 ...

> ap1 <- agePrecision(reader1~reader2,data=sb)
> summary(ap1,what="agreement")

Percentage of fish by differences in ages between pairs of assignments
      0       1       2       3       4       5
61.8136 30.3661  6.7388  0.7488  0.1664  0.1664

> summary(ap1,what="precision")

Precision summary statistics
    n R   CV   APE PercAgree
 1202 2 3.98 2.815     61.81
```

```
> ab2 <- ageBias(reader1~reader2,data=sb,col.lab="Reader 1",row.lab="Reader 2")
> summary(ab2,what="symmetry",flip.table=TRUE)

Raw agreement table (square & flipped)
        Reader 1
Reader 2   2   3   4   5   6   7   8   9  10  11  12  13  14  15  16  17  18  19  20
      20   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   1   -
      19   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -
      18   -   -   -   -   -   -   -   -   -   -   -   1   -   -   -   -   1   -   -
      17   -   -   -   -   -   -   -   -   -   -   -   -   -   1   -   1   -   -   -
      16   -   -   -   -   -   -   -   -   -   -   -   -   -   -   2   1   -   -   -
      15   -   -   -   -   -   -   -   -   -   -   1   2   2   3   -   -   -   -   -
      14   -   -   -   -   -   -   -   -   -   2   6   8   5   4   -   -   -   -   -
      13   -   -   -   -   -   -   1   -   -   3   5   8   1   -   -   -   -   -   -
      12   -   -   -   -   -   -   -   1  17  13  23   9   1   -   -   -   -   -   -
      11   -   -   -   -   -   1   1   4  22  25   4   1   -   -   -   -   -   -   -
      10   -   -   -   -   -   2  15  51 144  24   2   1   -   -   -   -   -   -   -
      9    -   -   -   -   1   1  29  89  32   4   -   -   -   -   -   -   -   -   -
      8    -   -   -   -   3  21  97  25   9   -   -   -   -   -   -   -   -   -   -
      7    -   -   -   3  23 149  38   5   -   -   -   -   -   -   -   -   -   -   -
      6    -   -   -   6  51  15   2   -   -   -   -   -   -   -   -   -   -   -   -
      5    -   -   5  45  10   1   -   1   -   -   -   -   -   -   -   -   -   -   -
      4    -   6  25   5   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -
      3    4  25   1   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -
      2   50   -   1   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -

Bowker's (Hoenig's) Test of Symmetry
 df chi.sq         p
 37  72.69 0.0004127
> plot(ab2)    # Left
> plot(ab2,difference=TRUE,ylim=c(-1.2,1))   # Right
```



# R Handout - Mark-Recapture

**Dr. Derek Ogle**
Northland College

Mar 2014, Vermont CFWRU Workshop

## Preliminaries

```
> library(FSA)  # for mrClosed(), capHistSum()
> setwd("C:/aaaWork/Web/fishR/courses/Vermont2014/CourseMaterial/") # Derek's Computer
```

## Single Census, Closed Population

```
> bg <- read.csv("Data/BluegillJL.csv",header=TRUE)
> view(bg)
    first second
24      1      0
79      1      0
124     1      0
167     1      0
242     0      1
273     0      1
> bgtbl <- xtabs(~first+second,data=bg)
> addmargins(bgtbl)
     second
first   0   1 Sum
  0     0  81  81
  1   187   9 196
  Sum 187  90 277
> mr1 <- mrClosed(M=196,n=90,m=9,type="Chapman")
> summary(mr1)
        N
[1,] 1792
> confint(mr1)
     95% LCI 95% UCI
[1,]     990    3504
```

```
> bgch <- capHistSum(bg)
> bgch$caphist


 01  10  11
 81 187   9
> mr2 <- mrClosed(bgch,type="Chapman")
> summary(mr2)
        N
[1,] 1792
```

## Single Census, Closed Population, Size Classes

The following table appeared in the grey literature article "A mark recapture experiment to estimate the escapement of chinook salmon in the Keta River, 2000".

**Table 1.–Numbers of chinook salmon marked in the Keta River and inspected for marks on the spawning grounds in 2000, by length group.**

| | 550–659 mm | ≥660 mm | Total |
|---|---|---|---|
| Event 1: Released with marks (*M*) | 70 | 200 | **270** |
| Event 2: Captured (*C*) | 52 | 172 | **224** |
| Recaptured (*R*) | 9 | 37 | **46** |
| *R*/C | 17.3% | 21.5% | **20.5%** |

```
> mr3 <- mrClosed(M=c(70,200),n=c(52,172),m=c(9,37),type="Chapman",
                  labels=c("550-659 mm",">=660 mm"))
> summary(mr3,incl.SE=TRUE,incl.all=TRUE)
              N    SE
550-659 mm  375  94.7
>=660 mm    914 116.6
All        1289 150.2
> confint(mr3)
           95% LCI 95% UCI
550-659 mm     228     639
>=660 mm       700    1216
```

## Multiple Census, Closed Population, Capture History Data

```
> np <- read.csv("Data/PikeNYPartial1.csv",header=TRUE)
> view(np)
```

```
     id first second third fourth
7  2007     1      0     0      0
12 2012     1      0     0      0
14 2014     1      0     0      0
28 2028     0      1     0      0
44 2044     0      0     1      0
46 2046     0      0     1      0
```

```
> npch <- capHistSum(np,cols=-1)
> npch$caphist
```

```
0001 0010 0011 0100 0101 0110 1000 1001 1010 1100
   5    8    2   12    1    2   21    1    2    3
```

```
> npch$sum
```

```
   n m  R  M
1 27 0 27  0
2 18 3 18 27
3 14 4 14 42
4  9 4  0 52
```

```
> mr4 <- mrClosed(npch,type="Schnabel")
> summary(mr4)
```

```
       N
[1,] 128
```

```
> confint(mr4)
```

```
     95% LCI 95% UCI
[1,]      75     238
```

## Multiple Census, Closed Population, Summarized Data

```
> mr5 <- mrClosed(n=c(16,19,16),m=c(0,7,7),R=c(16,19,0),type="Schnabel")
> summary(mr5)
```

```
      N
[1,] 50
```

```
> confint(mr5)
```

```
     95% LCI 95% UCI
[1,]      31      87
```

# R Handout - Age-Length Key

**Dr. Derek Ogle**
Northland College

Mar 2014, Vermont CFWRU Workshop

## Preliminaries

```
> library(FSA)      # for Subset(), view(), Summarize(), lencat(), ageKey(), fact2num()
> library(plotrix)  # for histStack()
```

```
> setwd("C:/aaaWork/Web/fishR/courses/Vermont2014/CourseMaterial/") # Derek's Computer
> d <- read.csv("Data/SpotVA2.csv",header=TRUE)
> str(d)
'data.frame': 403 obs. of  2 variables:
 $ tl : num  10.6 7.1 12.3 9.7 11.2 8.9 12.6 7.6 10 7 ...
 $ age: int  1 1 3 2 3 1 3 1 1 1 ...
> view(d)
     tl age
16  8.2   1
40  8.9   1
174 8.5  NA
270 8.9  NA
281 8.1  NA
321 7.9  NA
```

```
> sp.len <- Subset(d,is.na(age))
> str(sp.len)
'data.frame': 331 obs. of  2 variables:
 $ tl : num  9.6 9.4 9.1 9.4 9.6 9 8.2 9.8 10.7 9.1 ...
 $ age: int  NA NA NA NA NA NA NA NA NA NA ...
> sp.age <- Subset(d,!is.na(age))
> str(sp.age)
'data.frame': 72 obs. of  2 variables:
 $ tl : num  10.6 7.1 12.3 9.7 11.2 8.9 12.6 7.6 10 7 ...
 $ age: int  1 1 3 2 3 1 3 1 1 1 ...
```

## Constructing and Applying the Age-Length Key

```
> Summarize(~tl,data=sp.age,digits=1)
       n     mean       sd      min       Q1   median       Q3      max percZero
    72.0     10.3      2.1      6.3      8.7     10.3     12.0     13.9      0.0
> sp.age.mod <- lencat(~tl,data=sp.age,startcat=6,w=1)
> view(sp.age.mod)
     tl age LCat
2   7.1   1    7
16  8.2   1    8
40 12.5   2   12
```

```
41  8.8   1    8
55 10.3   1   10
66 13.3   3   13
```

```
> ( AL.raw <- xtabs(~LCat+age,data=sp.age.mod) )
```

```
    age
LCat  0  1  2  3  4
  6   2  0  0  0  0
  7   0 10  0  0  0
  8   1  9  0  0  0
  9   0  8  2  0  0
  10  0  9  1  0  0
  11  0  1  3  6  0
  12  0  1  4  4  1
  13  0  0  0  8  2
```

```
> ( AL.key <- prop.table(AL.raw,margin=1) )
```

```
    age
LCat   0   1   2   3   4
  6  1.0 0.0 0.0 0.0 0.0
  7  0.0 1.0 0.0 0.0 0.0
  8  0.1 0.9 0.0 0.0 0.0
  9  0.0 0.8 0.2 0.0 0.0
  10 0.0 0.9 0.1 0.0 0.0
  11 0.0 0.1 0.3 0.6 0.0
  12 0.0 0.1 0.4 0.4 0.1
  13 0.0 0.0 0.0 0.8 0.2
```

```
> sp.len.mod <- ageKey(AL.key,age~tl,data=sp.len)
```

```
Warning:  The maximum observed length in the length sample (13.8) is greater than the
largest length category in the age-length key (13).  Thus,
the last length category will be treated as all-inclusive.
```

```
> view(sp.len.mod)
```

```
      tl age
21  10.8   1
112  9.5   2
165  9.1   2
251 10.4   1
252  9.5   1
266 10.7   1
```

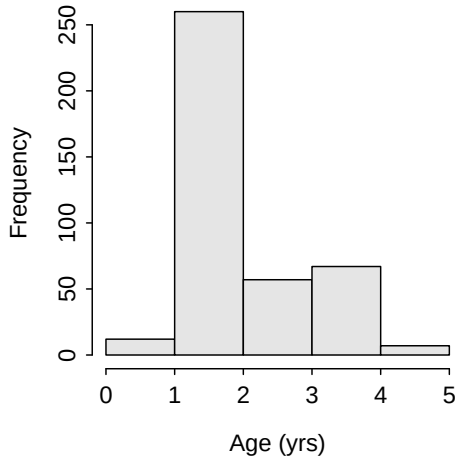
```
> sp.comb <- rbind(sp.age,sp.len.mod)
> str(sp.comb)
```

```
'data.frame': 403 obs. of  2 variables:
 $ tl : num  10.6 7.1 12.3 9.7 11.2 8.9 12.6 7.6 10 7 ...
 $ age: num  1 1 3 2 3 1 3 1 1 1 ...
```

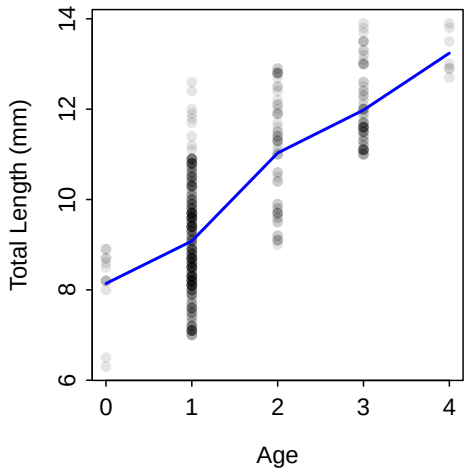
## Summarizing Final Results

```
> agefreq <- xtabs(~age,data=sp.comb)
> round(prop.table(agefreq)*100,1)

age
   0    1    2    3    4
 3.0 64.5 14.1 16.6  1.7
```

```
> ( sp.sum <- Summarize(tl~age,data=sp.comb,digits=2) )

Warning:  To continue, variable(s) on RHS of formula were converted to a factor.
  age   n  mean   sd  min    Q1 median   Q3  max percZero
1   0  12  8.14 0.86  6.3  8.15   8.35  8.7  8.9        0
2   1 260  9.08 1.18  7.0  8.20   8.95  9.9 12.6        0
3   2  57 11.03 1.23  9.0  9.80  11.30 12.0 12.9        0
4   3  67 11.98 0.82 11.0 11.40  11.70 12.4 13.9        0
5   4   7 13.24 0.48 12.7 12.90  13.00 13.6 13.9        0
> hist(~age,data=sp.comb,breaks=0:5,xlab="Age (yrs)")
```




```
> plot(tl~age,data=sp.comb,ylab="Total Length (mm)",xlab="Age",pch=16,col=rgb(0,0,0,1/10))
> lines(mean~fact2num(age),data=sp.sum,col="blue",lwd=2)
```



```
> histStack(tl~age,data=sp.comb,xlab="Total Length",ylim=c(0,100),
            col="gray.colors",legend="topright")
Warning:  z was converted to a factor
```