Size Structure I

Derek H. Ogle, Northland College

4-Mar-2015

Preliminaries

```
> library(fishWiDNR)   # for setDBClasses()
> library(dplyr)       # for filter(), select(), mutate(), group_by(), summarize()
> library(FSA)         # for Summarize(), hist(), expandCounts(), filterD()
> library(lubridate)   # for month()
```

```
> setwd("C:/aaaWork/Web/fishR/Courses/WiDNR_Statewide_2015/Day1_IntroR_FMData")
> d <- read.csv("SAWYER_fish_raw_data_012915.csv",stringsAsFactors=FALSE,na.strings=c("-","NA",""))
> d <- setDBClasses(d,type="RDNR")
> d <- expandCounts(d,~Number.of.Fish,~Length.or.Lower.Length.IN+Length.Upper.IN,new.name="Len")
> d <- mutate(d,Mon=month(Survey.Begin.Date,label=TRUE))
> d <- select(d,Species,Waterbody.Name,Survey.Year,Gear,Survey.Begin.Date,Mon,Len)
```

```
> Spr <- filterD(d,Survey.Year==2013,Mon %in% c("Apr","May","Jun"))
> BGSpr <- filterD(Spr,Species=="BLUEGILL")
> BGSprLC <- filterD(BGSpr,Waterbody.Name=="LAKE CHETAC",Gear=="BOOM SHOCKER")
```

So . . .

- `Spr` has all species sampled from all water bodies in the Spring of 2013.
- `BGSpr` has only Bluegill sampled from all water bodies in the Spring of 2013.
- `BGSprLC` has only Bluegill sampled with boom shockers from Lake Chetac in the Spring of 2013.

. . . and they all look similar to this . . .

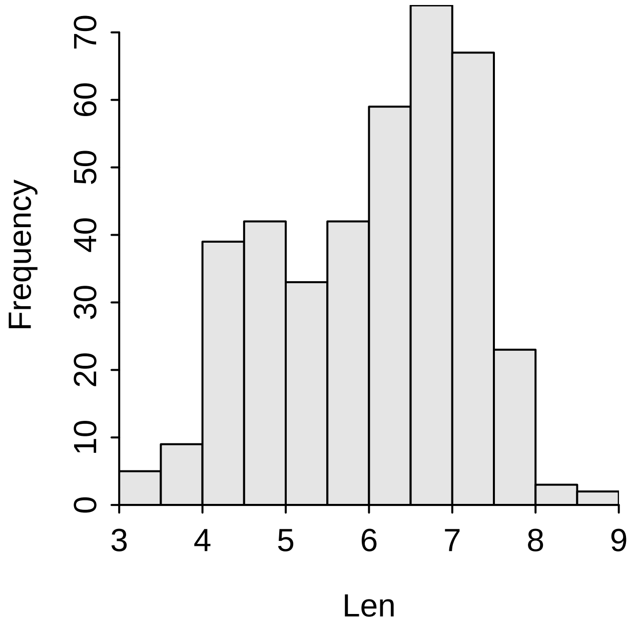
```
   Species Waterbody.Name Survey.Year         Gear Survey.Begin.Date Mon Len
1 BLUEGILL    LAKE CHETAC        2013 BOOM SHOCKER        2013-05-09 May 4.0
2 BLUEGILL    LAKE CHETAC        2013 BOOM SHOCKER        2013-05-09 May 4.7
3 BLUEGILL    LAKE CHETAC        2013 BOOM SHOCKER        2013-05-09 May 4.7
4 BLUEGILL    LAKE CHETAC        2013 BOOM SHOCKER        2013-05-09 May 7.3
5 BLUEGILL    LAKE CHETAC        2013 BOOM SHOCKER        2013-05-09 May 7.4
6 BLUEGILL    LAKE CHETAC        2013 BOOM SHOCKER        2013-05-09 May 6.6
```

Very Simple Summaries

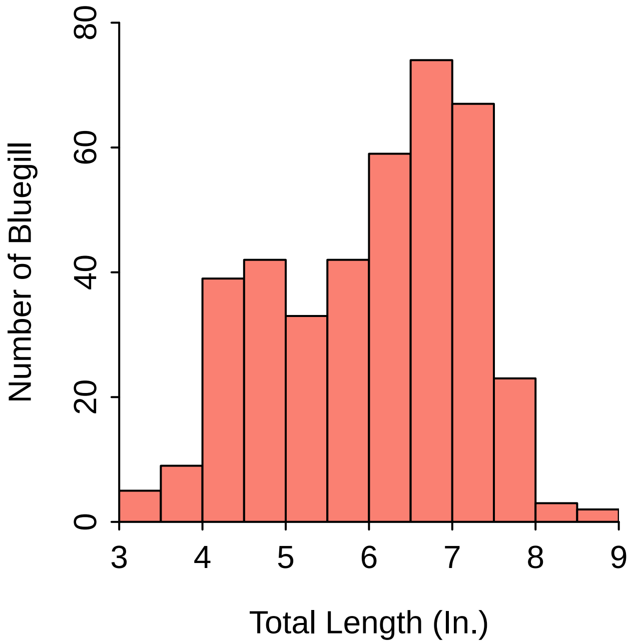
```
> Summarize(~Len,data=BGSprLC,digits=2)
       n     mean       sd      min       Q1   median       Q3      max percZero
  398.00     5.98     1.16     3.00     5.00     6.20     6.90     8.90     0.00
```

Length Frequency Histograms

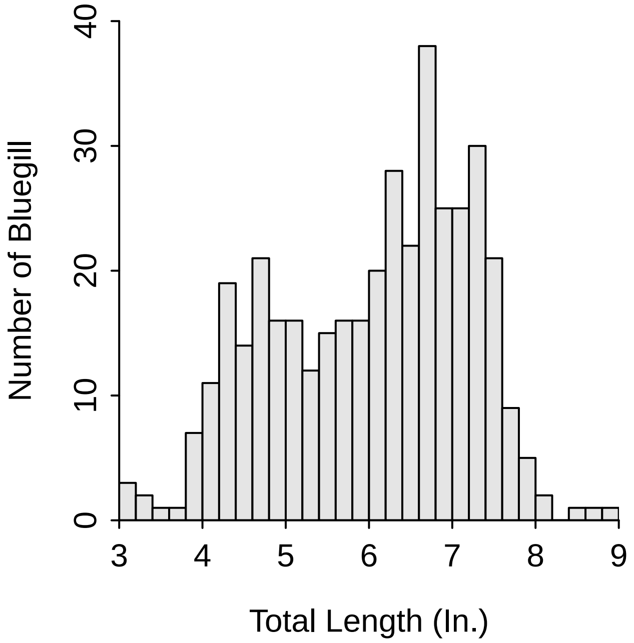
```
> hist(~Len,data=BGSprLC)
```



```
> hist(~Len,data=BGSprLC,xlab="Total Length (In.)",ylab="Number of Bluegill",
      xlim=c(3,9),ylim=c(0,80),col="salmon")
```



```
> hist(~Len,data=BGSprLC,xlab="Total Length (In.)",ylab="Number of Bluegill",
      xlim=c(3,9),ylim=c(0,40),breaks=seq(3,9,0.2))
```



Multiple Summaries at Once

```
> BGSpr <- group_by(BGSpr,Waterbody.Name)
> summarize(BGSpr,n=n(),meanLen=mean(Len))                   # see use of na.rm=TRUE below
Source: local data frame [11 x 3]

      Waterbody.Name   n  meanLen
1     BLACK DAN LAKE 599       NA
2       CONNORS LAKE 198       NA
3       DURPHEE LAKE 603       NA
4         GREEN LAKE 144 6.567361
5        LAKE CHETAC 589       NA
6      LAKE CHIPPEWA 746       NA
7  LAKE OF THE PINES 303       NA
8    LOWER CLAM LAKE  35 4.554286
9         MOOSE LAKE   1       NA
10        ROUND LAKE 414       NA
11    WHITEFISH LAKE  72       NA
```

```
> summarize(BGSpr,n=n(),valid_n=sum(!is.na(Len)),
            meanLen=mean(Len,na.rm=TRUE),sdLen=sd(Len,na.rm=TRUE),
            minLen=min(Len,na.rm=TRUE),maxLen=max(Len,na.rm=TRUE)  )
Source: local data frame [11 x 7]

      Waterbody.Name   n valid_n  meanLen     sdLen minLen maxLen
1     BLACK DAN LAKE 599     241 4.352697 0.9151520    2.1    7.0
2       CONNORS LAKE 198     108 5.155556 1.1018534    1.7    7.0
3       DURPHEE LAKE 603     574 6.603136 0.5071123    1.4    7.9
4         GREEN LAKE 144     144 6.567361 1.1392446    2.8    8.4
5        LAKE CHETAC 589     400 5.979250 1.1819420    2.0    8.9
6      LAKE CHIPPEWA 746     181 5.758011 1.1447001    3.7    8.0
7  LAKE OF THE PINES 303      90 5.000000 1.1646478    1.7    6.8
8    LOWER CLAM LAKE  35      35 4.554286 1.0042096    2.7    6.2
9         MOOSE LAKE   1       0      NaN       NaN     NA     NA
10        ROUND LAKE 414     309 5.070874 1.3018442    1.8    8.7
11    WHITEFISH LAKE  72      67 4.392537 1.3614067    2.1    7.4

> BGSpr <- filterD(BGSpr,Len>=3)
> summarize(BGSpr,n=n(),valid_n=sum(!is.na(Len)),
            meanLen=round(mean(Len,na.rm=TRUE),2),sdLen=round(sd(Len,na.rm=TRUE),2),
            minLen=min(Len,na.rm=TRUE),maxLen=max(Len,na.rm=TRUE),
            PSDQ=perc(Len,6,digits=0),PSD7=perc(Len,7,digits=0),PSDP=perc(Len,8,digits=0)  )
Source: local data frame [10 x 10]

      Waterbody.Name   n valid_n meanLen sdLen minLen maxLen PSDQ PSD7 PSDP
1     BLACK DAN LAKE 236     236    4.39  0.89    3.0    7.0    4    1    0
2       CONNORS LAKE 102     102    5.32  0.89    3.0    7.0   28    1    0
3       DURPHEE LAKE 573     573    6.61  0.46    4.5    7.9   94   21    0
4         GREEN LAKE 142     142    6.62  1.06    3.0    8.4   79   44    6
5        LAKE CHETAC 399     399    5.99  1.17    3.0    8.9   57   24    2
6      LAKE CHIPPEWA 181     181    5.76  1.14    3.7    8.0   44   20    1
7  LAKE OF THE PINES  83      83    5.23  0.87    3.0    6.8   20    0    0
8    LOWER CLAM LAKE  34      34    4.61  0.97    3.0    6.2   12    0    0
9         ROUND LAKE 296     296    5.18  1.21    3.0    8.7   25    9    2
10    WHITEFISH LAKE  59      59    4.65  1.25    3.0    7.4   15    8    0

> Spr <- group_by(Spr,Waterbody.Name,Species)
> summarize(Spr,n=n(),valid_n=sum(!is.na(Len)),
            meanLen=round(mean(Len,na.rm=TRUE),2),sdLen=round(sd(Len,na.rm=TRUE),2)  )
Source: local data frame [122 x 6]
Groups: Waterbody.Name

   Waterbody.Name                Species   n valid_n meanLen sdLen
1  BLACK DAN LAKE         BLACK BULLHEAD   2       0      NA    NA
2  BLACK DAN LAKE          BLACK CRAPPIE 402     402    6.89  1.42
3  BLACK DAN LAKE               BLUEGILL 599     241    4.35  0.92
4  BLACK DAN LAKE        LARGEMOUTH BASS  76      76   11.01  3.15
5  BLACK DAN LAKE            MUSKELLUNGE  38      15   34.88  7.35
6  BLACK DAN LAKE          NORTHERN PIKE   8       8   22.91  5.82
7  BLACK DAN LAKE            PUMPKINSEED  43      31    4.61  1.12
8  BLACK DAN LAKE PUMPKINSEED X BLUEGILL  13       9    5.36  1.01
9  BLACK DAN LAKE              ROCK BASS   4       4    4.40  2.23
10 BLACK DAN LAKE                WALLEYE 180     180   10.74  5.02
..            ...                    ... ...     ...     ...   ...
```

Application Assignment

Create a script that performs the following tasks:

1. Load and prepare your FM data in R (**HINT:** *use all or some of your scripts from previous application assignments*).
2. Reduce your data.frame to one year and several (4 or more) fish of interest. Call this the *original data.frame.*
3. Reduce the *original data.frame* to one water body and species of interest.
   - Compute summary stastistics for the length variable.
   - Construct a length frequency histogram.
   - Does your description of the length frequency change dramatically with different bin widths?
4. Reduce the *original data.frame* to only one species.
   - Efficiently construct summary statistics for the length variable for each water body. Include PSD values that are of interest to you (**HINT**: *use, for example,* `psdVal("Largemouth Bass",units="in")` *to find Gabelhouse lengths for a particular species*).
5. (*Time Permitting*) Re-create the summary statistics for one species in each water body but include calculations of the median and first and third quartiles (**HINT**: *use, for example,* `quantile(x,0.50,na.rm=TRUE)` *to compute the median (i.e., 50% quantile) of the data in* `x`.).
6. (*Time Permitting*) Compute summary statistics of the length variable for each water body AND each of the several species of interest to you. Save the summary statistics to an object and write the results to a CSV file.

Save your script!