

# fishR Vignette - Size Structure Analysis

Dr. Derek Ogle, Northland College

January 27, 2014

## 1 Background

Neumann and Allen (2007) note that

Size structure is one of the most commonly used fisheries assessment tools. The size structure of a fish population at any point in time can be considered a snapshot that reflects the interactions of the dynamic rates of recruitment, growth, and mortality. Thus, length-frequency data provide valuable insight into the dynamics of fish populations and help identify problems such as inconsistent year-class strength, slow growth, or excessive mortality.

This vignette will focus on the two most common methods of assessing the size structure of a fish population: length frequency histograms and size structure indices.

Before loading these data, the `FSA` and `FSAdata` packages maintained by the author should be loaded with

```
> library(FSA)
> library(FSAdata)
```

Throughout this vignette, the total length data from Bluegill (*Lepomis macrochirus*) and Largemouth Bass (*Micropterus salmoides*) sampled in 2007 from a small catch-and-release lake in northern Wisconsin will be examined. These data (along with other data) can be found in the `InchLake1` data frame in the `FSAdata` package. The entire data frame is loaded into R and the structure and the first six rows of data are viewed with

```
> data(InchLake1)
> str(InchLake1)

'data.frame': 4894 obs. of 5 variables:
 $ netID : int 36 36 36 36 37 37 37 37 37 37 ...
 $ netType: Factor w/ 6 levels "angling","miniNC",...: 1 1 1 1 1 1 1 1 1 1 ...
 $ year : int 2007 2007 2007 2007 2007 2007 2007 2007 2007 2007 ...
 $ species: Factor w/ 10 levels "Black Crappie",...: 6 6 6 6 2 2 1 1 1 1 ...
 $ length : num 12.1 13.3 13.5 12.4 6 7.9 11.8 11.7 11.7 11.7 ...

> head(InchLake1)

  netID netType year species length
1    36 angling 2007 Largemouth Bass 12.1
2    36 angling 2007 Largemouth Bass 13.3
3    36 angling 2007 Largemouth Bass 13.5
4    36 angling 2007 Largemouth Bass 12.4
5    37 angling 2007 Bluegill 6.0
6    37 angling 2007 Bluegill 7.9
```

The Bluegill and Largemouth Bass data from 2007 are isolated with

```
> bg <- Subset(InchLake1,year==2007 & species=="Bluegill")
> str(bg)
```

```

'data.frame': 2686 obs. of 5 variables:
 $ netID : int 37 37 38 38 38 38 38 38 38 38 ...
 $ netType: Factor w/ 5 levels "angling","miniNC",...: 1 1 5 5 5 5 5 5 5 5 ...
 $ year : int 2007 2007 2007 2007 2007 2007 2007 2007 2007 2007 ...
 $ species: Factor w/ 1 level "Bluegill": 1 1 1 1 1 1 1 1 1 1 ...
 $ length : num 6 7.9 2 1.9 2 2.6 2.1 3 2 3 ...

> lmb <- Subset(InchLake1,year==2007 & species=="Largemouth Bass")
> str(lmb)

'data.frame': 161 obs. of 5 variables:
 $ netID : int 36 36 36 36 37 37 37 37 37 37 ...
 $ netType: Factor w/ 5 levels "angling","miniNC",...: 1 1 1 1 1 1 1 1 1 1 ...
 $ year : int 2007 2007 2007 2007 2007 2007 2007 2007 2007 2007 ...
 $ species: Factor w/ 1 level "Largemouth Bass": 1 1 1 1 1 1 1 1 1 1 ...
 $ length : num 12.1 13.3 13.5 12.4 8.3 12.6 9.3 7.6 9.8 12.5 ...

```

Note, from the data frame structure, that the number of observations in each subsetted data frame are reduced from the original data frame and that, in each instance, the *species* variable has only one level – the level of the species of interest. These subsetted data frames will be used in the examples below.

## 2 Length Frequencies

A length frequency shows the number (i.e., frequency) of fish in each of several length categories that cover the range of observed lengths. The width of the length categories is important, as too few intervals may mask important information in the length frequency, whereas too many intervals does not provide a useful summary. [Anderson and Neumann \(1996\)](#) suggested 1-cm (0.5-in) intervals for fish that reach 30 cm (12 in), 2-cm (1-in) intervals for fish that reach 60 cm (24 in), and 5-cm (2-in) intervals for fish that reach 150 cm (60 in) maximum length. Of course, these are suggestions and you should use your best judgment when constructing the intervals.

Histograms are constructed in R with `hist()`. By default, `hist()` will use interval breaks and widths that are “pretty” and right-inclusive – e.g., an 8 cm fish will be included in the 6-8 cm interval rather than the 8-10 cm interval. This is not what most fisheries biologists desire. Thus, the `FSA` package contains a modified `hist()` that will use left-inclusive bins by default. The examples herein use the modified version of `hist()` which requires a formula notation as the first argument and a corresponding `data=` argument.

For most fisheries applications, R will have to be forced to use the interval breaks that you want. The exact interval breaks are set in R by putting the lower values of the breaks into a vector and assigning that vector to the `breaks=` argument in `hist()`. For example, the range of values observed for Bluegills in Inch Lake was between 1.0 and 9.2 inches as found with

```

> Summarize(~length,data=bg,digits=2)

```

n	mean	sd	min	Q1	median	Q3	max	percZero
2686.00	1.81	1.37	1.00	1.20	1.40	1.70	9.20	0.00

With this range of values it makes sense to construct length intervals that begin at 1 in and continue in 0.5-in intervals up to 9.5 in. This vector of numbers is constructed with `seq()` as follows<sup>1</sup>

```

> ( bg.brks <- seq(1,9.5,0.5) )
[1] 1.0 1.5 2.0 2.5 3.0 3.5 4.0 4.5 5.0 5.5 6.0 6.5 7.0 7.5 8.0 8.5 9.0 9.5

```

<sup>1</sup>The “extra” parentheses in this command force R to print the result at the same time that it is being assigned to an object. Without the parentheses you result would be assigned but not made visible to you until you typed the name of the assigned object.

With this vector of interval breaks the length frequency histogram for Bluegills (Figure 1) is constructed with<sup>2</sup>

```
> hist(~length,data=bg,breaks=bg.brks,main="",xlab="Length (in)",col="gray90")
```

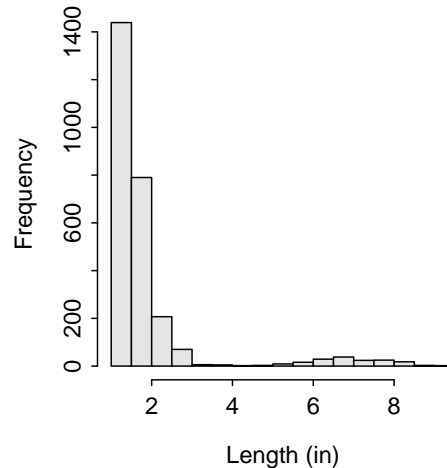


Figure 1. Length frequency for Inch Lake Bluegill.

From Figure 1 it is very apparent that a very large number of Bluegills less than approximately 3 inches were captured in Inch Lake. These fish can be explored a bit more closely by isolating fish less than 4 inches, creating a set of interval breaks that are 0.1-in wide<sup>3</sup>, and creating a new length frequency histogram (Figure 2) as follows

```
> bg.small <- Subset(bg,length <= 4)
> bg.brks2 <- seq(1,4,0.1)
> hist(~length,data=bg.small,breaks=bg.brks2,main="",xlab="Length (in)",col="gray90")
```

These small fish are likely one-year-old (sampled just after ice-out), there is some weak evidence for a slight bimodality with one mode of fish near 1.2 inches and another near 1.5 inches, and a sharp drop-off in abundance after 1.7 inches (Figure 2).

The larger fish can be examined (Figure 3) more closely in a similar manner with

```
> bg.lrg <- Subset(bg,length >= 4)
> bg.brks3 <- seq(4,9.5,0.2)
> hist(~length,data=bg.lrg,breaks=bg.brks3,main="",xlab="Length (in)",col="gray90")
```

Finally, you can easily create length frequency histograms separated by levels of a factor variable by including a formula of the form `length~factor` where `length` is generically the length variable and `factor` is generically the variable that you want to separate by. By default the same y-axis limits will be used for each graph. This may be instructive in some instances but, in this instance, one gear captured most of the fish such that all of the other histograms are non-informative. Different y-axis limits can be used by setting `same.ylim=FALSE`. For example, length frequency histograms separated by gear type (Figure 4) are constructed with

<sup>2</sup>Note that my figures may look different than yours as I have changed the default settings for the graphics with `par()`. In particular, I used `par(mar=c(3.5,3.5,1,1),mgp=c(2.1,0.4,0),tcl=-0.2)` for most graphics. See the “Basic Plotting in R” vignette.

<sup>3</sup>I felt comfortable making such narrow intervals here because the sample size is so large

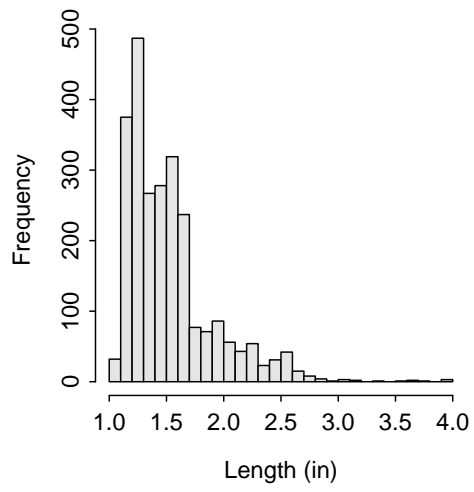


Figure 2. Length frequency for Inch Lake Bluegill less than 4 inches.

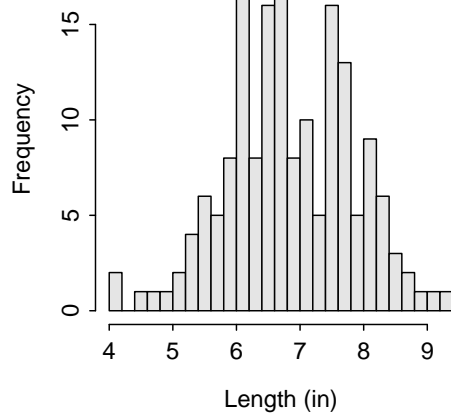


Figure 3. Length frequency for Inch Lake Bluegill greater than 4 inches.

```
> hist(length~netType,data=bg,breaks=bg.brks,same.ylim=FALSE,xlab="Length (in)",col="gray90")
```

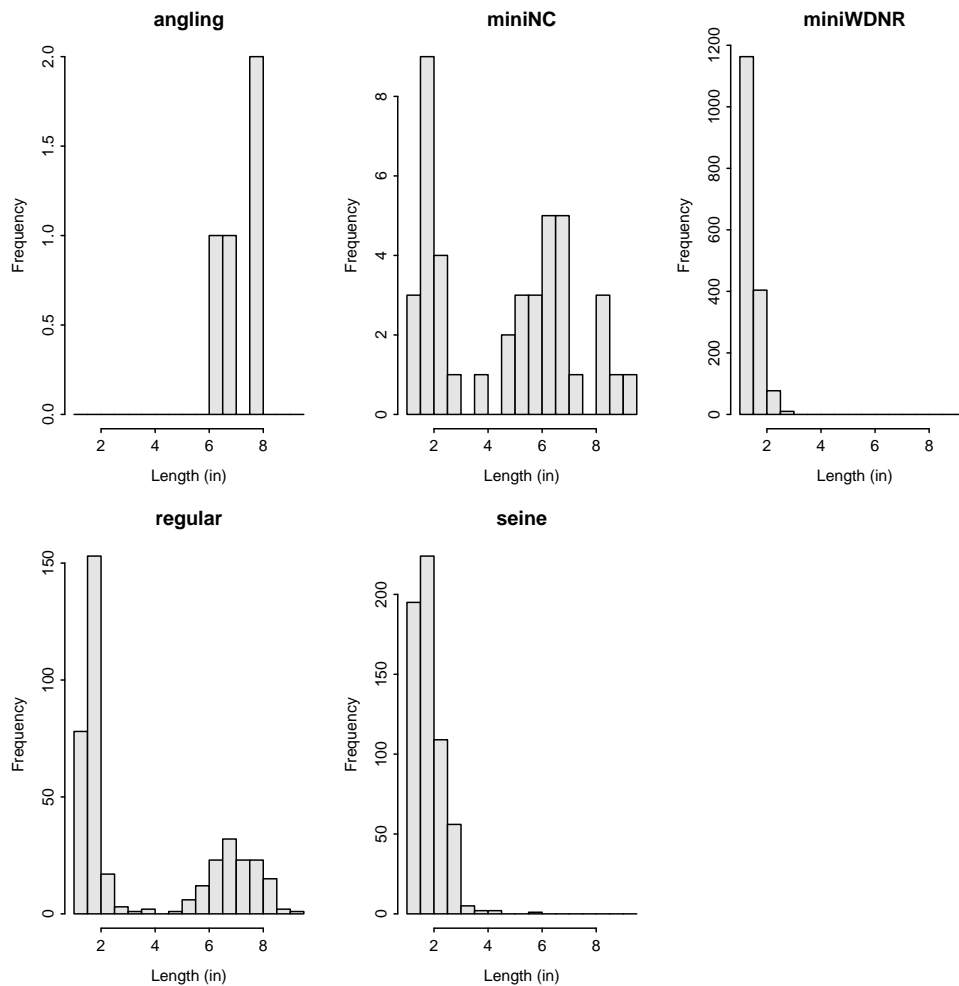


Figure 4. Length frequency for Inch Lake Bluegill separated by gear type.

### 3 Proportional Size Distribution (*PSD*)

Gabelhouse (1984b) introduced the “five-cell length categorization system” that defines lengths for so-called “stock”, “quality”, “preferred”, “memorable”, and “trophy” size individuals of a wide variety of species<sup>4</sup>. For example, stock-sized fish are at least 20% of the world-record length with the idea that fish smaller than this size have very little recreational value and that fish larger than this size are generally mature and available to most gears (Gabelhouse 1984b). Additionally, quality-sized fish are at least 36% of world-record length and are generally fish that most anglers would like to catch (Gabelhouse 1984b). It should be noted that all quality-sized fish are also stock-sized fish – i.e., the quality-sized fish are a subset of the stock-sized fish. Indeed, all fish in the “higher” categories also belong to the previous categories.

The minimum lengths for the five-cells in Gabelhouse’s system can be obtained with `psdVal()`. This function requires the name of the species, contained in quotes and spelled correctly<sup>5</sup>, as its first argument. There are

<sup>4</sup>The list of known species for which length categories have been defined can be seen with `psdVal()`.

<sup>5</sup>Spelled correctly means spelled as it appears in the function. You can see a list of all species stored in the function by typing `psdVal()`, without any arguments.

two other optional arguments to `psdVal()`. The `units=` argument controls whether the results are returned in millimeters (`"mm"`; default), centimeters (`"cm"`), or inches (`"in"`). Finally, `incl.zero=` controls whether the list of values returned will include a zero as the first value in the vector (`=TRUE`; default) or not (`=FALSE`). Including a zero in the vector is particularly useful in the common situation where the data set contains fish that are smaller than the stock size. Thus, a vector containing the five-cell length, in inches, categories (and zero) for Bluegills is obtained with

```
> ( bg.cuts <- psdVal("Bluegill",units="in") )
      zero      stock  quality preferred memorable   trophy
        0         3       6         8         10        12
```

Length frequency data can be summarized into a few useful statistics called proportional size distribution indices ( $PSD - X$ ; Guy *et al.* 2007)<sup>6</sup>. A  $PSD - X$  index is defined as the proportion of stock-sized fish that are also greater than some other larger size category. Specifically,

$$PSD - X = \frac{\text{Number of fish } \geq \text{specified length}}{\text{Number of fish } \geq \text{stock length}} * 100$$

The  $X$  in this formula is replaced with a letter or descriptor specific to the length specification in the numerator. For example,  $PSD - Q$  would have a numerator that is the “number of fish greater than or equal to ‘quality’ size” and would be interpreted as the proportion of stock-sized fish that are also quality-sized fish.  $PSD - Q$  is the most common calculation and is often just referred to as  $PSD$  (Guy *et al.* 2007). Similarly,  $PSD - P$  would use the preferred-size length in the numerator and is the proportion of stock-sized fish that are also preferred-size. Finally,  $PSD - 7$  would use 7 inches in the numerator and is the proportion of stock-sized fish that are greater than or equal to 7 inches. The  $PSD - X$  is a percentage but it is commonly referred to as a proportion and shown without the percentage sign.

A confidence interval for the  $PSD - X$  value is computed using binomial distribution theory. The binomial distribution theory is appropriate in this case because each fish of at least stock-size is recorded as to having also been of the specific size in the numerator or not (i.e., two choices) and the number of “successes” (i.e., of the specific size) is of interest to the biologist.

### 3.1 $PSD - X$ Calculation in R

Calculation of the  $PSD - X$  is simplified in R through the use of several functions (in addition to `psdVal()` as defined above). First, `lencat()` is used to convert the actual length measurements into a new variable that contains the length categories, which can then be summarized to obtain the number of fish in each length category. For  $PSD - X$  calculations, `lencat()` requires three arguments – a formula of the form `~length` as the first argument, the name of the data frame containing the length data as the second argument, and a vector containing the breaks (i.e., minimum values) for the length categories in the `breaks=` argument. The new variable appended to the data frame will be called `LCat` unless a different name is specified in `vname=`. The result from `lencat()` should be saved to a data frame object (either already existing or new).

The length categories, using the breaks previously stored in `bg.cuts`, for computing common  $PSD - X$  indices for Bluegill are appended to the `bg` data frame with

```
> bg <- lencat(~length,data=bg,breaks=bg.cuts)
> str(bg)
'data.frame': 2686 obs. of 6 variables:
 $ netID : int 37 37 38 38 38 38 38 38 38 38 ...
 $ netType: Factor w/ 5 levels "angling","miniNC",...: 1 1 5 5 5 5 5 5 5 5 ...
 $ year : int 2007 2007 2007 2007 2007 2007 2007 2007 2007 2007 ...
```

<sup>6</sup>These indices were formerly called the proportional stock density and relative stock density (RSD).

```

$ species: Factor w/ 1 level "Bluegill": 1 1 1 1 1 1 1 1 1 1 ...
$ length : num  6 7.9 2 1.9 2 2.6 2.1 3 2 3 ...
$ LCat   : Factor w/ 5 levels "0","3","6","8",...: 3 3 1 1 1 1 1 2 1 2 ...
> head(bg)
  netID netType year species length LCat
5     37 angling 2007 Bluegill   6.0    6
6     37 angling 2007 Bluegill   7.9    6
52    38  seine 2007 Bluegill   2.0    0
53    38  seine 2007 Bluegill   1.9    0
54    38  seine 2007 Bluegill   2.0    0
55    38  seine 2007 Bluegill   2.6    0

```

The frequency of individuals in each length category is then obtained with `table()` as follows

```

> ( bgtbl <- table(bg$LCat) )

  0    3    6    8   10
2506  41  116   23    0

```

Thus, there are 2506 fish between 0 and 3 inches and 41 fish between 3 and 6 inches. It is more instructive, however, to know how many total fish are greater than each of the length categories – e.g., how many Bluegills are larger than 6 inches and how many are larger than 3 inches. A reverse cumulative table can be obtained by submitting the saved `table()` object to `rcumsum()` as follows

```

> ( bgrcum <- rcumsum(bgtbl) )

  0    3    6    8   10
2686  180  139   23    0

```

From this, it is seen that there are 139 Bluegills larger than 6 inches (i.e., “quality or better”) and 180 Bluegills larger than 3 inches (i.e., “stock or better”).

The computation of  $PSD - Q$  is now straightforward with

```

> ( bg.psdq <- bgrcum["6"]/bgrcum["3"]*100 )

  6
77.22

```

A confidence interval for the “proportion of successes” (i.e., the  $PSD$  value as a proportion, in this case) is computed with `binCI()`. This function requires two arguments – the number of “successes” (i.e., number of quality-sized fish) and the total number of “trials” (i.e., number of stock-sized fish). The result is usually multiplied by 100 to match the scale of the  $PSD - X$  calculation. For example<sup>7</sup>,

```

> ( bg.psdq.ci <- binCI(bgrcum["6"],bgrcum["3"])*100 )

 95% LCI 95% UCI
 70.56  82.74

```

Thus, one is 95% confident that  $PSD$  for the population of Bluegills is between 71 and 83.

The four common  $PSD - X$  values<sup>8</sup> can be quickly computed by dividing the number of fish in each of the quality-sized or greater categories by the number of fish in the stock-sized category. For example,

<sup>7</sup>The “zero” and “stock” categories need to be removed from reverse cumulative sum table before making this calculation.

<sup>8</sup>The four common values are  $PSD$ ,  $PSD - P$ ,  $PSD - M$ , and  $PSD - T$ , though fewer may be presented if the categories contained zero fish.

```
> ( bg.psd <- bgrcum[-c(1,2)]/bgrcum["3"]*100 )
      6      8      10
77.22 12.78  0.00
```

Thus, for example, the  $PSD - P$  is 13.

### 3.2 PSD for Any Length

PSD calculations do not need to be restricted to only those lengths in the five-cell length categorization system. For example, suppose that there is a 7-inch minimum size limit on Bluegill. The biologist may then be interested in finding the proportion of stock-sized Bluegills that can be legally harvested (i.e., greater than or equal to 7 inches), or the  $PSD - 7$ . The  $PSD - 7$  is computed with the same concepts as shown previously, with the exception that the additional length(s) of interest must be added to the cut-off breaks by including the lengths in `addLens=` in `psdVal()`. For example, the  $PSD - 7$  for Inch Lake Bluegill is computed with

```
> ( bg.cuts2 <- psdVal("Bluegill",units="in",addLens=7) )
      zero      stock      quality      7 preferred memorable      trophy
      0         3         6         7         8         10         12
> bg <- lencat(~length,data=bg,breaks=bg.cuts2,vname="LCat2")
> bgtbl2 <- table(bg$LCat2)
> ( bgrcum2 <- rcumsum(bgtbl2) )

      0      3      6      7      8      10
2686 180 139  72  23  0
> bgrcum2["7"]/bgrcum2["3"]*100
 7
40
> ( bg.psd7 <- binCI(bgrcum2["7"],bgrcum2["3"])*100 )
95% LCI 95% UCI
33.12  47.29
```

Thus, it appears that between 33% and 47% of stock-length Bluegill in Inch Lake are of a legally harvestable length.

## 4 Convenience Functions

The FSA package has several functions to make the calculations of  $PSD - X$  indices more convenient. The first of these functions is `psdCalc()`, which will compute all of the common  $PSD - X$  indices, along with the  $PSD - X$  for any additional lengths provided by the user, with associated confidence intervals. This function requires the same formula and corresponding data frame as described for `lencat()`, the same `species=` and `units=` described for `psdVal()`, and an optional `digits=` that controls the printed number of decimal places for the result. For example, the calculations for Inch Lake Bluegill are obtained with

```
> psdCalc(~length,data=bg,species="Bluegill",units="in",digits=1)
      Value 95% LCI 95% UCI
PSD-Q 77.2    70.6    82.7
```



PSD-P	12.8	8.7	18.4
PSD-M	0.0	0.0	2.1

The following illustrates the inclusion of the 7-inch group using `addLens=` similar to what was illustrated previously

```
> psdCalc(~length,data=bg,species="Bluegill",units="in",digits=1,addLens=7)
      Value 95% LCI 95% UCI
PSD-Q  77.2   70.6   82.7
PSD-7  40.0   33.1   47.3
PSD-P  12.8    8.7   18.4
PSD-M   0.0    0.0    2.1
```

The `psdPlot()` can be used to construct a length frequency histogram with shaded bars representing the stock-size and greater fish and the five-cell length category lengths marked with vertical lines. This function requires the same formula, `data=`, `species=`, and `units=` described for `psdCalc()` and the same `startcat=` and `w=` described for `lencat()`. Other aspects of the plot are controlled as usual but it should be noted that if the range of values on the x-axis is limited with `xlim=`, then the minimum value in `xlim=` must be less than the stock length for the species. An example of this plot (Figure 5) was constructed for Inch Lake Bluegill with

```
> psdPlot(~length,data=bg,species="Bluegill",units="in",w=0.5,xlim=c(2.5,9.5),ylim=c(0,80))
```

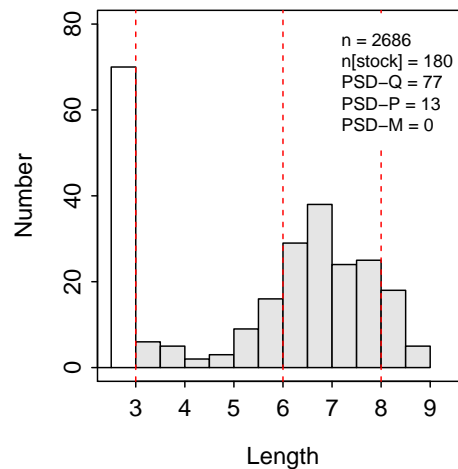


Figure 5. Length frequency histogram for Inch Lake Bluegill with the five-cell length categories and the various  $PSD - X$  values computed.

Finally, a fisheries scientist is often faced with the task of computing the  $PSD - X$  values for a wide variety of species across a variety of groups (e.g., years, gear, lakes). This process is made more efficient by using `psdDataPrep()` to create the length categorization variable at one time for all species in a data frame for which the five-cell system has been defined. As might be expected, `psdDataPrep()` has some limitations. For example, the species names in the original data frame must be spelled exactly as they appear in the `InchLake1` data frame, primarily full common names with the first letter capitalized only on the first name for species with compound names. For example, the species names in the original `InchLake1` data frame are not in the correct format (largely, the second names are capitalized)

```
> levels(InchLake1$species)
[1] "Black Crappie"      "Bluegill"           "Bluntnose Minnow"  "Fathead Minnow"
[5] "Iowa Darter"       "Largemouth Bass"   "Pumpkinseed"       "Rusty Crayfish"
[9] "Tadpole Madtom"    "Yellow Perch"
```

The `recodeSpecies()` can be used to convert species names. This function is flexible but in its simplest form it will simply “re-capitalize” the names to match the required form. More complicated examples are illustrated in the `recodeSpecies()` help. A simple conversion for the `InchLake1` data frame is completed with (note that the new variable name is the same as the old variable name but with a “1” appended)

```
> InchLake1 <- recodeSpecies(~species,data=InchLake1)
> levels(InchLake1$species1)
[1] "Black Crappie"      "Bluegill"           "Bluntnose Minnow"  "Fathead Minnow"
[5] "Iowa Darter"       "Largemouth Bass"   "Pumpkinseed"       "Rusty Crayfish"
[9] "Tadpole Madtom"    "Yellow Perch"
```

The `psdDataPrep()` requires three arguments – a formula of the form `length~species` in the first argument, the data frame in `data=`, and the units of measurement in `units=`. Species for which no known five-cell length categories exist can be eliminated from the resulting data frame with `remove.other=TRUE`. Removing the species without five-cell categories is not a requirement but it will speed-up the process and make the resulting tables less cluttered. The `InchLake1` data frame can be appended with the length categorization variable with

```
> InchLake1 <- psdDataPrep(length~species1,data=InchLake1,units="in",remove.other=TRUE)
Warning: The following species do not have known Gabelhouse five-cell length categories:
Bluntnose Minnow, Fathead Minnow, Iowa Darter, Rusty Crayfish, Tadpole Madtom
> head(InchLake1)
  netID netType year species length species1 PSDcat
7     37 angling 2007 Black Crappie  11.8 Black Crappie preferred
8     37 angling 2007 Black Crappie  11.7 Black Crappie preferred
9     37 angling 2007 Black Crappie  11.7 Black Crappie preferred
10    37 angling 2007 Black Crappie  11.7 Black Crappie preferred
11    37 angling 2007 Black Crappie  11.4 Black Crappie preferred
82    39 seine  2007 Black Crappie   1.3 Black Crappie      zero
```

The data could then be subsetted and summarized as described previously. However, the following code (somewhat explained with comments) can be used to construct the common *PSD – X* values for each species and year combination

```
> # make a frequency table by species and length category
> ( t1 <- with(InchLake1,table(interaction(species1,year),PSDcat)) )
      PSDcat
      zero stock preferred memorable quality
Black Crappie.2007      5      2      12      4      0
Bluegill.2007      2506     41      23      0     116
Largemouth Bass.2007   22     55       7      0      77
Pumpkinseed.2007       7      1       1      0       5
Yellow Perch.2007      0      0       6      0       6
Black Crappie.2008     0      4       5     10       0
Bluegill.2008      337     53     96      0     259
Largemouth Bass.2008   6     28       6      0      50
Pumpkinseed.2008      0      0       0      0       1
Yellow Perch.2008     17      0       5      1       6
```

```

> # do a reverse cum sum ignoring the zero (i.e., sub-stock) group
> ( t1r <- t(apply(t1[,-1],MARGIN=1,FUN=rcumsum)) )

      stock preferred memorable quality
Black Crappie.2007      18         16          4          0
Bluegill.2007          180        139         116         116
Largemouth Bass.2007  139         84          77          77
Pumpkinseed.2007       7          6           5           5
Yellow Perch.2007     12          12           6           6
Black Crappie.2008     19         15          10           0
Bluegill.2008          408        355         259         259
Largemouth Bass.2008  84         56          50           50
Pumpkinseed.2008      1          1           1           1
Yellow Perch.2008     12         12           7           6

> # Divide all greater than stock groups by the stock group -- i.e., PSD
> ( t1p <- round(t1r[,-1]/t1r[,1]*100,1) )

      preferred memorable quality
Black Crappie.2007      88.9        22.2         0.0
Bluegill.2007          77.2        64.4        64.4
Largemouth Bass.2007  60.4        55.4        55.4
Pumpkinseed.2007     85.7        71.4        71.4
Yellow Perch.2007    100.0        50.0        50.0
Black Crappie.2008    78.9        52.6         0.0
Bluegill.2008         87.0        63.5        63.5
Largemouth Bass.2008  66.7        59.5        59.5
Pumpkinseed.2008    100.0       100.0       100.0
Yellow Perch.2008    100.0        58.3        50.0

```

## 5 $PSD - X$ Interpretation

As a **very general** rule-of-thumb, a population of fish is said to be “in balance” if the  $PSD$  value is between 30 and 70. Different ranges have been proposed for different species (Table 1). The idea behind these target ranges is that if the  $PSD$  value is very low then there are few large fish in the population, whereas a very large  $PSD$  indicates that there are few small fish in the population. These target ranges should be used with caution, though, as an “in balance”  $PSD$  can be produced from a wide variety of size structures. For example, three populations of Bluegills may have the same  $PSD$  (and number of quality fish) but one population may have all quality fish between 6 and 7 inches, one population may have all quality fish greater than 9 inches (with fish between 6 and 9 inches absent), and the third population may have all quality fish spread more evenly between 6 and 9 inches. The reaction of the fish manager will likely be different for these three populations, but a simple look at the  $PSD$  would not detect the differences. Thus, the  $PSD$  is a useful summary, but it should not be used without also examining length frequency histograms (Section 2 and Section 4) or other  $PSD - X$  values.

A  $PSD$  value should also not be interpreted without considering other species and specific management goals. For example, Willis *et al.* (1993) proposed target  $PSD$  and  $PSD - P$  values for Bluegills and Largemouth Bass under three different management scenarios (Table 2).

A so-called “tic-tac-toe” graph is one method of visually representing size structure indices for a predator and a prey. A tic-tac-toe graph plots the  $PSD - X$  value for the predator on the x-axis and the prey on the y-axis. Target ranges are highlighted on each axis with shading such that there are vertical and horizontal stripes that “cut” the graph into nine squares. The “center square” is where the predator and the prey would both fall into their respective target ranges. Construction of a tic-tac-toe graph for a “balanced” management plan for Bluegills and Largemouth Bass from Inch Lake is shown below (Figure 6).

Table 1. Suggested target ranges for size structure indices for a variety of species. From Willis *et al.* (1993).

Species	<i>PSD</i>	<i>PSD - P</i>	<i>PSD - M</i>	Reference
Largemouth Bass	40-70	10-40	0-10	Gabelhouse (1984b)
Bluegill	20-60	5-20	0-10	Anderson (1985)
Crappies (midwestern ponds)	30-60	>10	-	Gabelhouse (1984a)
Crappies (Kansas reservoirs)	40-70	10-40	0-10	Willis (1984)
White bass	40-70	10-40	0-10	Willis (1984)
Walleye	30-60	-	-	Anderson and Weithman (1978)
Northern pike	30-60	-	-	Anderson and Weithman (1978)
Yellow perch	30-60	-	-	Anderson and Weithman (1978)

Table 2. Suggested target ranges for size structure indices for Largemouth Bass and Bluegill under three different management options. From Willis *et al.* (1993).

Option	Largemouth Bass			Bluegill	
	<i>PSD</i>	<i>PSD - P</i>	<i>PSD - M</i>	<i>PSD</i>	<i>PSD - P</i>
Panfish	20-40	0-10	-	50-80	10-30
Balance	40-70	10-40	0-10	20-60	5-20
Big bass	50-80	30-60	10-25	10-50	0-10

The *PSD* for Bluegill captured in Inch Lake in May 2007 was computed previously. Similar *PSD* calculations for Largemouth Bass captured during the same time frame are computed with

```
> lmb.cuts <- psdVal("Largemouth Bass",units="in")
> lmb <- lenecat(~length,data=lmb,breaks=lmb.cuts,vname="LCat2")
> ( lmbrcum2 <- rcumsum(table(lmb$LCat2)) )

  0   8  12  15  20
161 139  84   7   0

> ( lmb.psdq <- lmbrcum2["12"]/lmbrcum2["8"]*100 )

  12
60.43

> ( lmb.psdq.ci <- binCI(lmbrcum2["12"],lmbrcum2["8"])*100 )

 95% LCI 95% UCI
 52.13  68.18
```

Thus, one is 95% confident that the *PSD* value for the population of Largemouth Bass is between 52 and 68.

The base tic-tac-toe graph is constructed with `tictactoe()`. All arguments to this function are optional. However, the default axis labels are changed with `xlab=` and `ylab=`, respectively. More importantly, the default “in balance” target ranges are changed with the `predbal=` and the `preybal=` arguments. Both of these arguments requires a vector of length two that indicates the lower and upper values of the target range. The default values for both `predbal=` and `preybal=` is `c(30,70)`.

After the base tic-tac-toe graph has been constructed, the size structure indices for a predator-prey pair are added to the graph with `tictactoeAdd()`. This function requires a vector of length two corresponding to the numerator and denominator of the index calculation for the predator as the first argument and a similar vector for the prey as the second argument. For example, a tic-tac-toe graph, with modified “in balance” ranges, of the *PSD* values for the Bluegill and Largemouth Bass in Inch Lake in 2007 (Figure 6) is constructed with

```
> tictactoe(xlab="Largemouth Bass",ylab="Bluegill",predbal=c(40,70),preybal=c(20,60))
> tictactoeAdd(c(lmbrcum2["12"],lmbrcum2["8"]),c(bgrcum["6"],bgrcum["3"]),pt.col="blue")
```

Predator PSD was 60 with a 95% CI of (52.1,68.2).  
Prey PSD was 77 with a 95% CI of (70.6,82.7).

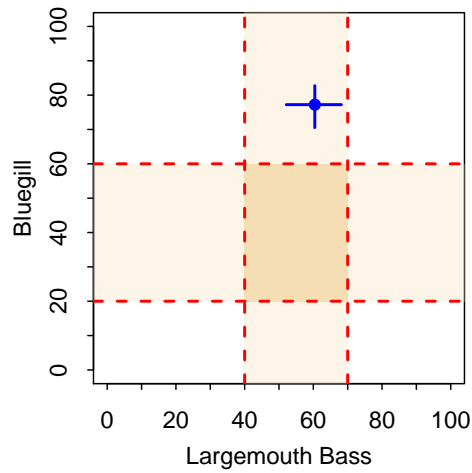


Figure 6. Tic-tac-toe graph of Bluegill and Largemouth Bass *PSD* values for Inch Lake in May 2007. The shaded target ranges represent a “balanced” management option.

From this, and assuming that the predator and prey are “in balance” if both *PSD* values are in the center shaded box, it is seen that the bass-Bluegill system is out of balance. It appears that there are more quality-sized Bluegill than would be expected in a balanced predator-prey system. Of course, other statistics should be looked at before making final conclusions about this population. For example, bass in Inch Lake likely also prey on an abundant population of bluntnose minnows (*Pimephales notatus*) and yellow perch (*Perca flavescens*).

[Neumann and Allen \(2007\)](#) provide a much more thorough discussion regarding the interpretation of length frequency histograms and size structure indices.

## References

- Anderson, R. and R. Neumann. 1996. Length, weight, and associated structural indices. in, Murphy, B.R. and D.W. Willis, editors Fisheries Techniques, second edition, American Fisheries Society, Bethesda, Maryland pp. 447–481. [2](#)
- Anderson, R. O. 1985. Managing Ponds for Good Fishing. University of Missouri Extension Division, Columbia, MO. [12](#)
- Anderson, R. O. and A. S. Weithman. 1978. The concept of balance for coolwater fish populations. American Fisheries Society Special Publications 11:371–381. [12](#)
- Gabelhouse, J., D. W. 1984a. An assessment of crappie stock in small midwestern private impoundments. North American Journal of Fisheries Management 4:371–384. [12](#)
- Gabelhouse, J., D. W. 1984b. A length-categorization system to assess fish stocks. North American Journal of Fisheries Management 4:273–285. [5](#), [12](#)
- Guy, C., R. M. Neumann, D. W. Willis, and R. Anderson. 2007. Proportional size distribution (PSD): A further refinement of population size structure index terminology. Fisheries 32:348. [6](#)
- Neumann, R. M. and M. S. Allen. 2007. Size Structure, chapter 9, pp. 375–421. American Fisheries Society. [1](#), [13](#)
- Willis, D. W. 1984. A statewide summary of sampling data for white crappie, walleye, white bass, and largemouth bass collected in kansas reservoirs. Comprehensive planning option project fw-9-p-3, Kansas Fish and Game Commission. [12](#)
- Willis, D. W., B. R. Murphy, and C. S. Guy. 1993. Stock density indices: Development, use, and limitations. Reviews in Fisheries Science 1:203–222. [11](#), [12](#)

# Reproducibility Information

## Version Information

- **Compiled Date:** Mon Jan 27 2014
- **Compiled Time:** 3:07:51 PM
- **Code Execution Time:** 3.34 s

## R Information

- **R Version:** R version 3.0.2 (2013-09-25)
- **System:** Windows, i386-w64-mingw32/i386 (32-bit)
- **Base Packages:** base, datasets, graphics, grDevices, methods, stats, utils
- **Other Packages:** FSA\_0.4.3, FSAdata\_0.1.4, gdata\_2.13.2, knitr\_1.5.15
- **Loaded-Only Packages:** bitops\_1.0-6, car\_2.0-19, caTools\_1.16, cluster\_1.14.4, evaluate\_0.5.1, formatR\_0.10, Formula\_1.1-1, gplots\_2.12.1, grid\_3.0.2, gtools\_3.1.1, highr\_0.3, Hmisc\_3.13-0, KernSmooth\_2.23-10, lattice\_0.20-24, MASS\_7.3-29, multcomp\_1.3-1, mvtnorm\_0.9-9996, nlme\_3.1-113, nnet\_7.3-7, plotrix\_3.5-2, quantreg\_5.05, sandwich\_2.3-0, sciplot\_1.1-0, SparseM\_1.03, splines\_3.0.2, stringr\_0.6.2, survival\_2.37-4, tools\_3.0.2, zoo\_1.7-10
- **Required Packages:** FSA, FSAdata and their dependencies (car, gdata, gplots, Hmisc, knitr, multcomp, nlme, plotrix, quantreg, sciplot)