

# AIFFD Chapter 15 - Community Indices, Parameters, and Comparisons

*Derek H. Ogle*

## Contents

15.1 Sample Data Set and Structural Indices . . . . .	2
15.2 Estimation of Species Richness by Rarefaction . . . . .	2
15.3 Calculation of Jaccard's and Simple Matching Coefficients . . . . .	2
15.4 Calculation of Spearman's Rank Correlation and Kendall's Tau Similarity Coefficients . . . . .	2
15.5 Calculation of a Cophenetic Correlation Coefficient . . . . .	3
15.6 K-Means Clustering Analysis . . . . .	3
15.7 Principal Components Analysis . . . . .	3
References . . . . .	5

---

This document contains R versions of the boxed examples from **Chapter 15** of the “Analysis and Interpretation of Freshwater Fisheries Data” book. Some sections build on descriptions from previous sections, so each section may not stand completely on its own. More thorough discussions of the following items are available in linked vignettes:

- the use of linear models in R in the [preliminaries vignette](#),
- differences between and the use of type-I, II, and III sums-of-squares in the [preliminaries vignette](#), and
- the use of “least-squares means” is found in the [preliminaries vignette](#).

The following additional packages are required to complete all of the examples (with the required functions noted as a comment and also noted in the specific examples below).

```
> library(FSA) #
```

In addition, external tab-delimited text files are used to hold the data required for each example. These data are loaded into R in each example with `read.table()`. Before using `read.table()` the working directory of R must be set to where these files are located on **your** computer. The working directory for all data files on **my** computer is set with

```
> setwd("C:/aaaWork/Web/fishR/BookVignettes/aiffd2007")
```

In addition, I prefer to not show significance stars for hypothesis test output, reduce the margins on plots, alter the axis label positions, and reduce the axis tick length. In addition, contrasts are set in such a manner as to force R output to match SAS output for linear model summaries. All of these options are set with

```
> options(width=90,continue=" ",show.signif.stars=FALSE,
           contrasts=c("contr.sum","contr.poly"))
> par(mar=c(3.5,3.5,1,1),mgp=c(2.1,0.4,0),tcl=-0.2)
```

## 15.1 Sample Data Set and Structural Indices

During a 2-week period in 1988, a survey of the fishes of the Kankakee River, Illinois, was conducted using a boat-mounted electrofisher. Six sites (stations) were each sampled eight times with effort standardized among samples (Peterson 1989; Kwak 1993).

```
> d <- read.table("data/Box15_1.txt",header=TRUE)
> str(d)
```

```
'data.frame':  34 obs. of  8 variables:
 $ Species : Factor w/ 34 levels "Banded_darter",...: 17 9 5 7 8 12 18 22 25 26..
 $ SpecCode: Factor w/ 34 levels "BAD","BKS","BLD",...: 15 11 5 7 8 12 18 22 25..
 $ Station1: int  6 164 42 0 13 0 10 0 8 1 ...
 $ Station2: int  7 90 33 0 58 0 11 22 89 22 ...
 $ Station3: int  0 6 29 0 10 0 10 4 5 4 ...
 $ Station4: int  4 6 3 1 14 0 0 0 15 1 ...
 $ Station5: int  26 432 44 15 36 7 2 0 8 5 ...
 $ Station6: int  5 194 35 0 13 0 0 2 35 1 ...
```

NOT CONVERTED YET

## 15.2 Estimation of Species Richness by Rarefaction

A cumulative sample of fishes from station 1 of the Kankakee River, Illinois, included 648 individuals representing 26 species (see [Box 15.1](#)). Below, we estimate the expected species richness from a sample of 100 individuals.

NOT CONVERTED YET

## 15.3 Calculation of Jaccard's and Simple Matching Coefficients

To calculate both Jaccard's and simple matching coefficients, first determine the number of species present and absent for both stations and the number of species occurring at one station but not another. Using the summary data for stations 1 and 2 on the Kankakee River, Illinois (see [Box 15.1](#)),

- number of species present at both stations is  $p = 18$ ,
- number of species absent at both stations is  $a = 4$ , and
- number of species present at one station but not the other is  $m = 12$ .

NOT CONVERTED YET

## 15.4 Calculation of Spearman's Rank Correlation and Kendall's Tau Similarity Coefficients

Species total abundances from stations 1 and 2 of the Kankakee River, Illinois ([Box 15.1](#)), are used to illustrate the calculation of Spearman's rank and Kendall's tau coefficients.

NOT CONVERTED YET

## 15.5 Calculation of a Cophenetic Correlation Coefficient

Fish assemblages of the Kankakee River, Illinois (Box 15.1), were clustered with the single-linkage method (Box 15.8). Below, we illustrate calculation of the matrix cophenetic correlation coefficient. The values in the cophenetic matrix are estimated from the single-linkage dendrogram (Box 15.8) by tracing the path connecting each pair of assemblages. For example, when tracing the linkage from station 1 upward through the tree and downward to station 2, the greatest distance is 177.45. The remaining values are similarly estimated and are included in the cophenetic matrix below.

NOT CONVERTED YET

## 15.6 K-Means Clustering Analysis

The following R program (analysis initially done in SAS) performs k-means clustering with PROC FASTCLUS on the summary fish abundance data for six stations on the Kankakee River, Illinois (Box 15.1). Three-letter codes are used in place of species names (see Box 15.11 for key to codes). The number of clusters,  $k = 3$ , is specified by the MAXCLUSTERS command. Note that this example is for illustration only. The K-means clustering procedure should be used only when the number of assemblages (samples) exceeds 30.

## 15.7 Principal Components Analysis

```
> dt <- t(d[, -c(1:2)])           # transpose only the count portion
> colnames(dt) <- d$SpecCode      # use species codes to label columns (variable names)
> dt
```

	LOG	GZS	BLM	BUM	CAP	HOC	MIS	RDS	RYS	SAS	SFS	STS	SUM	BLR	GOR	NHS	SHR
Station1	6	164	42	0	13	0	10	0	8	1	19	45	0	0	34	5	35
Station2	7	90	33	0	58	0	11	22	89	22	24	32	0	0	0	2	0
Station3	0	6	29	0	10	0	10	4	5	4	3	69	0	4	35	10	8
Station4	4	6	3	1	14	0	0	0	15	1	2	14	0	0	36	7	2
Station5	26	432	44	15	36	7	2	0	8	5	23	51	4	1	9	2	35
Station6	5	194	35	0	13	0	0	2	35	1	8	14	1	0	55	7	22

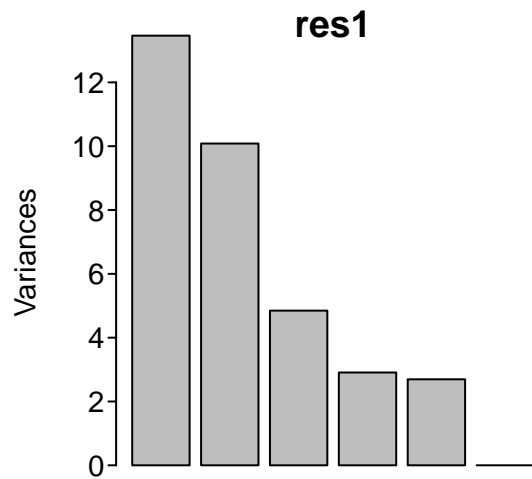
	QLL	RVR	SVR	SAB	BKS	BLG	GSF	LMB	LOS	OSF	ROB	SMB	BAD	BLD	JOD	LOP	SLD
Station1	5	2	8	0	4	2	1	0	35	1	30	143	4	1	0	25	5
Station2	2	0	1	3	10	0	42	3	94	0	3	59	0	9	6	0	0
Station3	1	2	1	0	13	0	6	2	26	0	15	195	0	1	3	51	6
Station4	4	0	3	0	9	2	3	0	39	8	31	151	1	0	0	42	7
Station5	17	0	14	3	10	1	7	8	48	31	27	165	0	4	2	24	2
Station6	14	5	4	0	6	0	1	0	37	3	62	204	1	0	0	7	2

```
> res1 <- prcomp(dt, retx=TRUE, center=TRUE, scale.=TRUE)
> summary(res1)
```

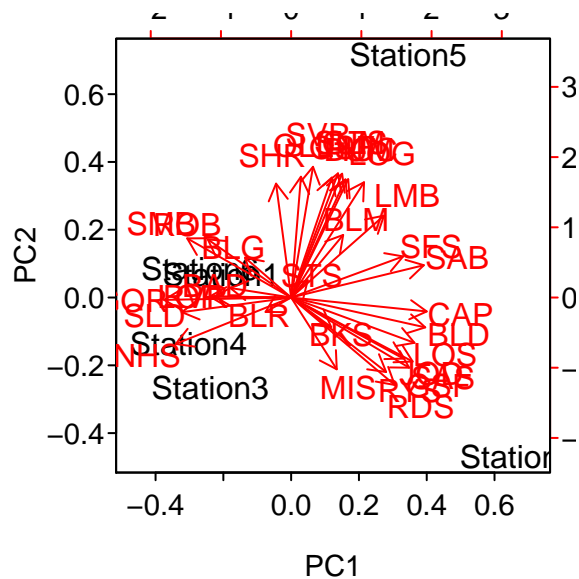
Importance of components:

	PC1	PC2	PC3	PC4	PC5	PC6
Standard deviation	3.669	3.1755	2.2017	1.70541	1.64181	6.416e-16
Proportion of Variance	0.396	0.2966	0.1426	0.08554	0.07928	0.000e+00
Cumulative Proportion	0.396	0.6926	0.8352	0.92072	1.00000	1.000e+00

```
> screeplot(res1)
```



```
> biplot(res1)
```



```
> pc1 <- res1$rotation[,1] # first PCA loadings
> pc2 <- res1$rotation[,2]
> pc1s <- predict(res1)[,1] # first PCscore for each individual
```

---

Reproducibility Information  
 Compiled Date: Fri May 15 2015  
 Compiled Time: 5:34:34 PM  
 Code Execution Time: 0.07 s

R Version: R version 3.2.0 (2015-04-16)  
System: Windows, i386-w64-mingw32/i386 (32-bit)  
Base Packages: base, datasets, graphics, grDevices, methods, stats, utils  
Required Packages: FSA and its dependencies (car, dplyr, gdata, gplots,  
Hmisc, knitr, lmtest, multcomp, plotrix, relax, sciplot)  
Other Packages: FSA\_0.6.13, knitr\_1.10.5, popbio\_2.4, quadprog\_1.5-5,  
rmarkdown\_0.6.1, TTR\_0.22-0, xts\_0.9-7, zoo\_1.7-12  
Loaded-Only Packages: acepack\_1.3-3.3, assertthat\_0.1, bitops\_1.0-6,  
car\_2.0-25, caTools\_1.17.1, cluster\_2.0.1, codetools\_0.2-11,  
colorspace\_1.2-6, DBI\_0.3.1, digest\_0.6.8, dplyr\_0.4.1, evaluate\_0.7,  
foreign\_0.8-63, formatR\_1.2, Formula\_1.2-1, gdata\_2.16.1,  
ggplot2\_1.0.1, gplots\_2.17.0, grid\_3.2.0, gridExtra\_0.9.1,  
gtable\_0.1.2, gtools\_3.4.2, highr\_0.5, Hmisc\_3.16-0, htmltools\_0.2.6,  
KernSmooth\_2.23-14, lattice\_0.20-31, latticeExtra\_0.6-26, lme4\_1.1-7,  
lmtest\_0.9-33, magrittr\_1.5, MASS\_7.3-40, Matrix\_1.2-0, mgcv\_1.8-6,  
minqa\_1.2.4, multcomp\_1.4-0, munsell\_0.4.2, mvtnorm\_1.0-2,  
nlme\_3.1-120, nloptr\_1.0.4, nnet\_7.3-9, parallel\_3.2.0, pbkrtest\_0.4-2,  
plotrix\_3.5-11, plyr\_1.8.2, proto\_0.3-10, quantreg\_5.11,  
RColorBrewer\_1.1-2, Rcpp\_0.11.6, relax\_1.3.15, reshape2\_1.4.1,  
rpart\_4.1-9, sandwich\_2.3-3, scales\_0.2.4, sciplot\_1.1-0, SparseM\_1.6,  
splines\_3.2.0, stringi\_0.4-1, stringr\_1.0.0, survival\_2.38-1,  
TH.data\_1.0-6, tools\_3.2.0, yaml\_2.1.13

---

## References