PN-ABK-246
75231

WORKING PAPER SERIES

FISHERIES STOCK ASSESSMENT

TITLE XII
Collaborative Research Support Program

Fisheries Stock Assessment CRSP Management Office
International Programs, College of Agriculture
The University of Maryland, College Park, Maryland 20742

In cooperation with the United States Agency for International Development (Grant No. DAN-4146-G-SS-5071-00) the Fisheries Stock Assessment CRSP involves the following participating institutions:

The University of Maryland—Center for Environmental and Estuarine Studies
The University of Rhode Island—International Center for Marine Resource Development
The University of Washington—Center for Quantitative Sciences
The University of Costa Rica—Centro de Investigación en Ciencias del Mar y Limnología
The University of the Philippines—Marine Science Institute (Diliman)—College of Fisheries (Visayas)

In collaboration with The University of Delaware; The University of Maryland—College of Business and Management; The University of Miami; and The International Center for Living Aquatic Resources Management (ICLARM).

WORKING PAPER SERIES

Working Paper No. 62

THE LOGARITHMIC SERIES DISTRIBUTION AND SOME APPLICATIONS TO THE ANALYSIS OF TROPICAL MULTISPECIES FISHERIES DATA

by
Saul B. Saila and James E. McKenna

Graduate School of Oceanography
The University of Rhode Island

August 1989

The Fisheries Stock Assessment CRSP (sponsored by part by USAID Grant No. Dan-4146-G-SS-5071-00) is intended to support collaborative research between U.S. and developing countries' universities and institutions on fisheries stock assessment and management strategies.

This Working Paper has been produced by The University of Rhode Island research component in collaboration with The University of the Philippines and in association with the International Center for Living Aquatic Resources Management. Additional copies are also available from:

International Center for Marine Resource Development
The University of Rhode Island
126 Woodward Hall
Kingston, Rhode Island 02881

The Logarithmic Series Distribution and Some Applications to the Analysis of Tropical Multispecies Fisheries Data

Saul B. Saila
and James E. McKenna
Graduate School of Oceanography
The University of Rhode Island
Narragansett, RI 02882-1197

Introduction

The logarithmic series distribution was first described by Fisher, Corbet and Williams (1943). They considered the relative numbers of individuals of different species obtained when sampling at random from an animal population. This problem was examined in relation to the distribution of butterflies on the Malay peninsula and to data on the number of moths of different species caught in light traps. Williams (1964) provided many later applications for which this distribution was found to be useful. Recently, Dial and Marzluff (1989) demonstrated that hollow curve distributions, as exemplified by the logarithmic series, are common in nature and are a result of assemblages being dominated by one or a few taxa.

The logarithmic series distribution is a special case of the more general power series distribution. Noack (1952) and Khatri (1959) studied the analytic properties of this distribution and Patil (1961) investigated estimation procedures for it.

One of the several available derivations for the logarithmic series distribution is provided in Appendix I. It is hoped that this material will provide a general overview of the logarithmic series distribution, and that it will provide some rationale for an interest in its applications as they relate to tropical multispecies stock assessment methodologies. Of special relevance are the suitability and fit of the log-series distribution to tropical multispecies fish abundance data—as well as the potential utility of the index of diversity (α), which is derived from this distribution—as well as Mountford's index of similarity (I), also derived from this distribution.

The concern here is with changes in multispecies assemblages exemplified by the distribution of species (group) abundances. This is thought to be useful because these changes permit some investigation of species balance independent of which species (group) may be present. In choosing a statistic that characterizes this distribution as a measure of diversity, it is important to recognize that any such statistic should be independent of sample size and should show relatively small variability in replicated temporal samples. The latter property would then allow good spatial discrimination among sampling locations.

General Properties

The logarithmic series is derived mathematically from expansion of ln (1+x). To avoid negative numbers in the series, it may be written as:

$$- \ln (1-x) - x + x^2/2 + x^3/3 + x^4/4$$

Since there are no logarithms of negative numbers, it follows that x must be less than unity.

In using the log-series to express the frequency distribution of species containing different numbers of individuals, it is written as:

$$n_1, n_1 x/2, n_1 x/3, n_1 x/4 \ldots$$

with the successive terms being numbers of species with 1, 2, 3 ... individuals. No zero term in included.

The series in this form has two parameters, (n_1) which is the number of species with one individual and (x) which is a number less than 1. The series has an infinite number of terms. It is discontinuous with only integer values of individuals per species and is convergent. The sum of all species to infinity is:

$$S = \frac{n_1}{x} [-\ln (1-x)]$$

The corresponding series of individuals (individuals of all species in the same abundance class) is:

$$n_1, n_1x, n_1x^2, n_1x^3, n_1x^4 \ldots$$

This is a geometric series with a constant multiple x. As $x < 1$ the series is convergent, and the sum to infinity (N) or the total number of individuals is:

$$N = n_1 (1-x), \text{ or } n_1 = N/(1-x)$$

The ratio of the number of species to the number of individuals is:

$$\frac{S}{N} = \frac{1-x}{x} [-\ln(1-x)]$$

Thus, for any average number of individuals per species (the reciprocal of the above), there is only one value of x, and from this n_1 can be obtained. If N and S are known, the series is fixed.

If sampling by individuals is done from a population graduated by a log-series distribution, then the smaller the sample the smaller the average number of individuals per species and the smaller the value of x. The larger the sample the larger the average number of individuals per species and the larger the value of x. However, in all random samples of any size from one population, the ratio of n_1 to x is a constant and is called the index of diversity or α.

$$n_1/x = \alpha \text{ or } n_1 = \alpha x.$$

With increasing sample size, the value of x increases and approaches unity. With increasing sample size, the number of species with one individual increases and gradually approaches the value of α, but never (theoretically) exceeds it.

The log-series can also be written as:

$$\alpha x, \frac{\alpha x^2}{2}, \frac{\alpha x^3}{3}, \text{ etc.}$$

where αx is the number of species (groups) represented by one individual, $\frac{\alpha x^2}{2}$ is the number of species represented by two individuals, etc. The sum of this, the total number of species is:

$$S = -\alpha \ln (1-x)$$

α is a property of the sampled population, and x is a property of the sample and depends on its size. Some definitions of symbols are:

x = sampling parameter of log-series distribution: $0 < x < 1$

α = parameter of log-series distribution: $\alpha > 0$ (diversity index)

S = the number of species encountered

N = the number of individuals encountered

n_1 = the number of species encountered with one individual.

Some relations among the above include:

1) $\alpha = n_1/x$

2) $S = \alpha[-\ln(1-x)]$

3) $x = N(N+\alpha)$

4) $\alpha = N(1-x)/x$

5) $n_1 = N\alpha/(N+\alpha)$

6) $n = N(1-x)$

7) $N/S = x/(1-x)[-\ln(1-x)]$

8) $S = \alpha \ln(1+N/\alpha)$

For times when N/S is large in relation to 1, then:

(9) $S = \alpha \ln(N/\alpha)$

The large sample variance of α is given by Anscombe (1950) as:

(10) $$\mathrm{var}(\alpha) = \frac{\alpha}{-\ln(1-x)}$$

Fit to Samar Sea Fisheries Data

Various models have been applied to data sets involving diverse taxa in an effort to find appropriate descriptions of species-abundance relationships. The comprehensive works of Pielou (1975, 1977) and Williams (1964) describe many of these models. This brief review only relates to applications of the log-series model to fisheries and other aquatic applications of this model. Based on a study of benthic macro-invertebrates,

Shepard (1984) suggested that the logarithmic series distribution is the most appropriate for studying species-abundance patterns in aquatic ecosystems. Kobayashi (1987) developed a new index of similarity which is independent of sample size, and it is based on the assumption that the communities from which samples are derived are described by the log-series distribution. Koch (1987) examined large marine data sets and concluded that the species-occurrence frequency distribution for such data sets fit the log-series distribution well. Wolda (1981) stated that Fisher's α is the best available measure of species diversity. In the case of multispecies trawl data, it is not expected that any one simple mathematical model will fit all data with complete fidelity, especially in view of sampling variability. However, if Fisher's logarithmic series distribution is found to provide reasonable fits to available empirical data on the distribution of species (groups) in standardized trawl samples from tropical multispecies fisheries, this might provide a probability distribution based model for quantifying species-abundance patterns in these complex ecosystems. Although use of the logarithmic series must be judged on its utility for given applications, there seems to be increasing evidence for use of this distribution to describe species abundance relations.

Taylor, Kempton, and Woiwood (1976) made an exhaustive study of macrolepidoperta collected at several sites, and showed that although the log-series model was not an ideal description of population structure, diversity as measured by the parameter α generally behaved more predictably and consistently than other diversity statistics tested. In the above study, α was found to be a normally distributed property of the populations studied.

The data used in this study consisted of the Samar Sea, Philippines trawl survey data base. These data include a total of 267 individual trawl hauls which included more than two million marine organisms (mostly fish) identified to 179 taxonomic groups, primarily at a species level. Temporal replications were obtained by sampling at ten quarterly intervals using 28 sampling stations. Description of the gear, standardized hauls

processing procedures and other details are given in Saeger (1983). The Samar Sea data set is believed to be one of the larger tropical fisheries data sets examined in this detail to date.

Clearly, sampling errors ascribed to a single sample depend on the population it is supposed to represent. Following Taylor, Kempton, and Woiwood (1976), in order to allow for a certain amount of natural population variability in addition to intrinsic sampling variation, we assume that the number of species S, rather than the number of individuals N, in replicate samples follows a Poisson distribution. This is in contrast to Fisher et al. (1943) who assumed that the number of individuals summed over all species should conform to a Poisson model. Empirical evidence for the validity of the assumption is presented in Table 1. This table shows the mean and the variance-to-mean ratios of quarterly cruise totals for both the number of species and the number of individuals caught at 28 trawl stations for which at least five successive hauls were available. It is evident from Table 1 that the variance of N (the number of individuals captured) increases roughly as the square of the mean and is substantially greater than the mean in all instances. This provides some evidence that Fisher's assumption is not valid for individuals sampled over time. In contrast to the variance of N, the variance of S (species) is quite close to the mean (see last column), suggesting that the total number of species behaves as a Poisson variate. This justifies the assumption made earlier that the number of species, not individuals, conforms to the Poisson model.

It is desirable to extend the assumption that the total number of species S is a Poisson variate to the case where the number of species (groups) with a particular abundance r (n_r) is also a Poisson variate. As Anscombe (1950) has pointed out, it then follows that the total number of species ($S = \Sigma n_r$) is a Poisson variate. Table 2 illustrates the distribution of species (group) frequencies for ten cruises covering stations 1-5 in the Samar Sea. Note (last column) that each cruise involved between 17 - 75 x 10^3

Table 1. Mean and variance-to-mean ratio of individuals and species caught at 28 stations in the Samar Sea, Philippines

		No. of individuals (N)		No. of species (S)	
Station No.	Sample Size	$\bar{x}$	$s^2/\bar{x}$	$\bar{x}$	$s^2/\bar{x}$
1	10	15213	21821	38	2.01
2	9	6434	5152	36	1.12
3	10	6118	2876	33	2.04
4	10	8130	8152	37	1.29
5	10	5165	1979	33	1.23
6	10	6970	2360	31	0.96
7	10	6276	2646	31	0.59
8	10	7270	5734	29	0.43
9	10	7017	4008	30	1.15
10	5	37222	26601	24	1.31
11	10	34892	33560	29	1.54
11	10	17508	1133	30	0.96
13	9	7817	4955	33	1.28
14	10	5111	4509	36	1.17
15	10	5170	1342	38	0.41
16	10	8654	14415	37	1.04
17	10	6829	4255	38	0.34
18	10	5486	6330	40	0.48
19	10	7372	8477	33	1.11
20	10	5886	19258	34	0.70
21	9	8272	5300	34	1.18
22	10	9125	8843	39	1.16
23	10	3872	1254	38	2.20
24	9	12852	14327	34	0.74
25	9	10527	5573	33	1.58
26	10	13573	13137	23	0.46
27	10	13460	8668	33	0.97
28	10	10020	2298	31	2.50

Table 2. Distribituon of species frequencies for successive cruises in the Salmar Sea, 1997-99

	Abundance Class											Total	
Cruise#	1	2+	4+	8+	16+	32+	64+	128+	256+	512+	1024+	spp.	No's.
97	8	15	15	13	10	22	15	11	4	8	0	67	17913
98	28	28	23	22	15	17	12	13	8	3	5	77	22053
99	7	18	18	19	21	18	17	12	7	5	6	62	24424
100	12	22	19	20	21	9	14	14	7	6	5	68	70296
102	8	20	24	23	27	17	17	12	13	5	5	80	34038
104	6	6	11	19	21	19	12	8	5	10	7	63	49354
106	2	15	30	15	25	20	22	25	11	11	6	83	31721
107	3	0	50	22	17	25	16	22	7	9	13	80	47677
108	6	21	16	24	31	21	13	10	7	8	7	71	30873
110	6	0	9	38	37	27	24	22	18	12	13	78	75608*
$\bar{x}$	8.5	14.5	21.5	21.5	22.5	19.5	16.2	14.9	8.7	7.7	6.7	72.9	40396
s^2	53.4	90.7	138.9	45.6	62.1	24.5	16.4	34.5	17.6	8.4	14.9	58.3	
$s^2/\bar{x}$	6.2	6.2	6.5	2.1	2.8	1.3	1.0	2.3	2.0	1.1	2.2	0.8	

*minimum value because station #2 had no total available

individuals. The data have been grouped into abundance classes of approximately equal range on a logarithmic scale. The pattern of variation shown by cruises should show independent Poisson variation. However, the variation in the first three abundance clauses (namely, 1, 2^+, 4^+) is considerably higher than expected. It is believed that this situation is the result of sampling variability for rare species. The analysis of Table 2 suggests that those species which occur as individuals or as a very few individuals in this type of sample are subject to a great deal of sampling variability, as they probably have a highly contagious spatial distribution and/or are not vulnerable to the gear used. Koch (1987) has shown that for those species which occur infrequently, sample size effect may be very large relative to the faunal patterns reported. At this point it is suggested that the rare species be eliminated before formal testing of specific hypotheses. Whether sampling strategies could be refined to permit more consistent representation of rare species is not known at present. Koch (1987) developed methods to test for sample size effects on distributional patterns of species. However, the paleontological studies he addressed dealt exclusively with non-motile species. In our case the number of species (groups) with r individuals in a sample (n_r) tend to show independent Poisson variation with expectations (E_r) described by a logarithmic series distribution, except in the case of the first three abundance classes, which show considerable variability.

Computer programs for calculating the parameters of the log-series distribution are found in the Appendix. The method used for calculating the parameters of the log-series distribution and their variances are based on Shepard (1984). That is, for a sample with N individuals and s species, the value of α is calculated from:

$$\alpha\,[\exp -s/\alpha - 1] - N = 0$$

by means of an iterative procedure. To obtain the expected number of species with n individuals each (s_n), the following relation is used:

$$s_n = (\alpha/n)\,x^n.$$

However, it is first necessary to calculate

$$x = N(N + \alpha).$$

The variance associated with s and α are:

$$\text{var}(s) = \alpha \ln [(2N + \alpha)/(N + \alpha)] - [\alpha^2 N/(N + \alpha)^2]$$

$$\text{var}(\alpha) = [\alpha^3 \{(N + \alpha)^2 \ln [(2N + \alpha)/(N + \alpha)] - \alpha N\}]$$

For testing the fit of observed to expected number of species, a test described by Fisher is used. A value of k is estimated as follows: A value of the difference:

$$[\alpha n (1 + 1/2 + 1/3 + \ldots + \frac{1}{n-1})] - \frac{s^2}{2\alpha}$$

is evaluated. In the expression αn is the number of observed species with n individuals each and s is the number of species. Solution of the above in the first term represents the score of observed values and the second term represents the expected total score. Significance is tested by estimating the standard deviation by

$$\text{s.d.} = [0.908 (\log N/s)^{1.673} s]^{-2}$$

Non-significance is accepted if the difference is less than its standard deviation. If this test indicates a significant difference between observed and expected scores, the difference is divided by its variance to obtain a value of k which provides an exact fit.

Table 3 illustrates estimates of the log-series distribution parameters to data from three cruises chosen at random from the Samar Sea file. These three cruises are 97, 98, and 108. The values of α range from less than 2 to more than 8. The goodness-of-fit of the distribution to the data is demonstrated by high values of x and by the values of k. Values of k less than unity are considered to provide reasonably good fits to the data. It is clear from examination of the last column that most of the calculated k values are greater than unity. However, if the data are truncated by removing the most variable abundance class (namely, those with seven or less individuals) then the recalculated k values are significantly reduced and most of them are less than unity. The parameter k varies inversely with the unevenness in species abundance represented by each term in the series

Table 3. Log-series distribution parameters for three cruises selected at random and for 28 stations sampled during each cruise, Samar Sea, Philippines. Note that some station data were incomplete and were listed as "missing data."

Cruise/Station	N	S	α	SEα	x	k
97-01	3180	26	3.874	.286	.999	0.210
97-02	5384	35	5.014	.311	.999	0.181
97-03	1346	20	3.331	.303	.998	2.996
97.04	2251	28	4.504	.328	.998	3.025
97.05	5772	29	3.984	.264	.000	2.934
97-06	3904	26	3.740	.198	.999	4.835
97-07	3388	28	4.180	.198	.998	0.166
97-08	2002	28	4.608	.351	.997	N.S.
97-09	2920	19	2.723	219	.999	1.818
97-10	10936	31	3.905	237	.999	1.371
97-11	105220	19	1.724	109	.999	1.842
97-12	51855	23	2.294	140	.999	2.417
97-13	1628	29	5.012	388	.999	7.269
97-14	878	36	7.557	.601	.991	9.962
97-15	2396	35	5.810	.398	.997	0.233
97-16	1160	30	5.624	.453	.995	1.078
97-17 (missing data)						
97-18	1767	33	6.760	.421	.997	1.971
97-19	3087	23	3.372	.262	.999	2.987
97-20	4102	29	4.114	.290	.999	3.336
97-21	1156	22	3.855	.346	.997	N.S.
97-22	3587	37	5.747	.366	.998	0.893
97-23	8114	31	4.081	.255	.999	1.774
97-24	8283	33	4.373	.266	.999	2.834
97-25	18018	29	3.379	.202	.999	3.208
97-26	7663	26	3.263	.222	.999	5.478
97-27	36397	26	2.738	.162	.999	4.276
97-28	18292	23	2.596	.171	.999	N.S.

Table 3. (continued)

Cruise/Station	N	S	α	SEα	x	k
98-01	11196	40	5.213	.285	.999	N.S.
98-02	2029	39	6.848	.462	.997	1.882
98-03	1748	42	7.744	.521	.996	0.212
98-04	3754	36	5.518	.353	.998	N.S.
98-05	3326	38	5.015	.383	.998	1.008
98-06	4234	32	4.703	.311	.998	2.245
98-07	2511	38	6.352	.420	.997	3.371
98-08	5153	26	3.574	.251	.999	4.891
98-09	3436	30	4.522	.314	.998	N.S.
98-10	36101	25	2.613	.158	.999	0.462
98-11	6545	33	4.536	.282	.999	1.543
98-12	760	23	4.474	.422	.994	7.153
98-13	4231	30	4.361	.295	.999	3.282
98-14	1799	36	6.396	.450	.992	5.472
98-15	2591	40	6.713	.434	.997	5.619
98-16	2250	36	6.086	.416	.997	7.902
98-17	1156	42	8.456	.616	.993	10.881
98-18	3244	37	5.855	.378	.978	8.407
98-19	5331	32	4.524	.291	.999	8.926
98-20	11851	34	4.251	.249	.999	N.S.
98-21	3029	28	4.264	.301	.998	1.409
98-22	3345	32	4.903	.333	.998	2.301
98-23	3444	33	5.057	.338	.998	N.S.
98-24	6016	31	4.275	.275	.999	1.092
98-25	4123	33	5.963	.366	.998	2.183
98-26	6428	36	5.032	.303	.999	3.084
98-27	9988	39	4.421	.224	.999	3.192
98-28	6781	33	4.511	.280	.999	5.646

Table 3. (continued)

Cruise/Station	N	S	α	SEα	x	k
108-01	2429	42	7.214	.462	.997	0.138
108-02	2588	40	6.715	.434	.997	1.525
108-03	3277	35	5.471	.360	.998	0.250
108-04	12678	40	5.118	.276	.999	.983
108-05	9901	39	5.159	.288	.999	2.200
108-06	11211	36	4.618	.263	.999	3.453
108-07	9347	35	4.594	.269	.999	4.857
108-08	14075	36	4.467	.249	.999	5.149
108-09	4751	33	4.781	.308	.999	9.196
108-10 (missing data)						
108-11	23172	36	4.176	.223	.999	6.169
108-12	14958	38	4.713	.256	.999	0.130
108-13	7935	36	4.867	.287	.999	1.480
108-14	6076	40	5.743	.334	.999	2.694
108-15	373614	37	5.739	.365	.998	1.527
108-16	8205	35	4.686	.278	.999	4.715
108-17	10200	38	4.660	.251	.999	0.110
108-18	3199	42	6.827	.421	.998	1.613
108-19	5561	40	5.830	.342	.999	2.474
108-20	7244	38	5.201	.301	.999	3.537
108-21	9678	37	4.822	.278	.999	4.424
108-22	9840	43	5.779	.310	.999	4.856
108-23	2595	47	8.151	.497	.997	0.163
108-24	4463	42	6.415	.379	.998	1.166
108-25	27249	27	2.958	.176	.999	1.920
108-26	19768	36	4.164	.231	.999	2.434
108-27	8599	34	4.500	.269	.999	4.524
108-28	8780	38	5.100	.291	.999	5.181

for which the distribution is a continuous function. When the abundances of each species are roughly similar, k is expected to have a high value. When the abundances vary greatly, k is expected to have a small value. This approach to describing the fit of the log-series model is identical to that done by Shepard (1984) for stream macro invertebrates. It is thought that this method is more appropriate than the usual chi-square goodness-of-fit test. The reason is that that power of the chi-square test is low, and samples may be accepted in fitting the logarithmic series using this test when in fact the fit is not good. Krebs (1989) has recommended that so-called Whittaker plots be made to demonstrate conformity to a log-series model. These plots involve plotting relative abundance (percent) on a logarithmic scale on the y-axis and species in rank order on the x-axis. The result is a relatively straight line for the logarithmic series, if the data conform to the distribution.

The log-series fitting program described in the appendix provides an option to calculate the theoretical Whittaker plot for a logarithmic series. This methodology is described by Krebs (1989) and involves solving the following equation:

$$R = \alpha E_1\left(n \ln\left(1 + \frac{\alpha}{N}\right)\right)$$

where R = species in rank order (x-axis of plot)

α = Fisher's index of diversity

n = number of individuals expected for a specified value of R (y-axis of plot)

N = total number of individuals in sample

E_1 = standard exponential integral (Abramowitz and Stegun 1964, Ch. 5)

Solving the above equation for n using integer values of R provides values for the expected Whittaker plot. These can be compared with the original data.

Figure 1 illustrates the distribution of alpha for 267 stations from the Samar Sea. The fit of the normal distribution to the observed frequency distribution looks reasonably good. The null hypothesis is not rejected at the 0.05 probability level. The evidence at

hand suggests that use of the α diversity statistic with symmetric intervals for comparative studies of tropical multispecies fish assemblages seems justified to some extent.

Taylor, Kempton, and Woiwood (1976) also suggested that the logarithmic series parameters α and x may be given biological interpretations. They indicated from their work that the parameter α remains reasonably constant over time for the same station, but that x changes somewhat more. From the Samar Sea data, a two-way randomized block analysis of variance design was used to compare temporal versus spatial variability of α.

The analysis consisted of 15 stations x 10 samples at each station taken at quarterly intervals. These stations and samples were chosen to permit orthogonal comparisons. Table 4 illustrates the results of this analysis. From this table it is evident that both temporal and spatial variability were statistically significant. This would be expected, since the area was in the early stages of recovery from a trawling restriction on larger vessels. However, the variation within stations over the ten quarterly samples seemed to be somewhat smaller than that observed over the 15 stations. This tends to support the premise that α seems somewhat less variable within assemblages at a given location (environment) than among locations.

Table 4. Analysis of variance table for a 15 x 10 randomized block experiment testing for the temporal and spatial changes in the parameter α. Samar Sea data for stations 1--5 over 10 cruises.

Source of Variation	d.f.	Sum of Squares	Mean Square	F_o	Prob ($F > F_o$)
among stations	14	84.260	6.019	4.935	0.05E-04
within stations	9	46.270	5.141	4.215	0.15E-03
error	116	153.671	1.220		
Total	149	284.200			

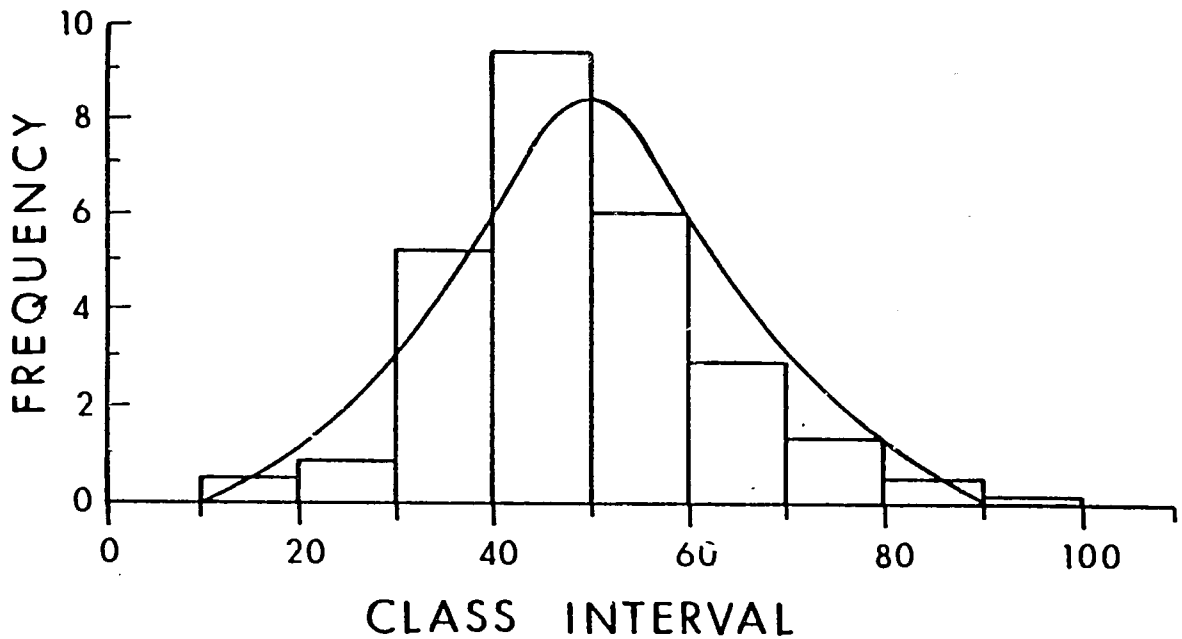

Figure 1. Distribution of $\hat{\alpha}$ for 267 stations in the Samar Sea, Philippines with fitted normal curve.

Chi-square (d.F. = 5) = 11.0197

The Diversity Index (α) Based on the Log-Series Distribution

Fisher et al. (1943) have demonstrated that for all samples (or summation of samples) taken from the same population by the same methods, $\alpha = n_1/x$ is constant. These symbols have been previously defined as:

x = sampling parameter of the log-series distribution $0 < x < 1$

α = parameter of the log-series distribution: also Fisher's diversity index

n_1 = the number of species encountered with one individual.

Fisher's demonstration of the above is of significance to studies of multispecies fisheries and to ecology in general. Because α is high in populations which have many groups (species) relative to the number of individuals and low in populations which have a small number of groups (species) relative to the number of individuals, Williams (1947) has termed α an index of diversity. It measures the extent to which individuals are apportioned into groups (species).

Some characteristics of this index of diversity listed by Williams (1944) are briefly mentioned below.

a) If many samples are taken from the same population, not only will each have the same index of diversity but also the combined sample from two or more original samples will have the same index of diversity.

b) For samples of differing sizes from the same population (x), the sampling parameter of the log-series distribution approaches unity as the sample size increases. Also, since $n_1 = \alpha x$, it is evident that α, the index of diversity, is the theoretical limiting value of n_1. That is, α is the maximum number of species (groups) with one individual obtained from one sample regardless of its size.

c) If two samples of considerable size are taken from the same populations and the larger population is p times the smaller in size, then the number of species in the

larger sample will be $\alpha \ln p$ more than in the smaller sample. This follows from relationship 8) of the General Properties section where:

$$S = \alpha \ln (1 + N/\alpha)$$

If N is very large relative to α, we have that:

$$S = \alpha \ln (N/\alpha).$$

Therefore, if two samples from the same population contain N and PN species (groups):

$$S_{PN} - S_N = \alpha (\ln PN/\alpha - \ln N/\alpha) = \alpha \ln P.$$

By way of illustration, if one sample is twice as big as the other, it will contain $\alpha \ln 2 = .69$ α more species. Similarly, if the size of the sample is multiplied by $e = 2.718$, the number of species added $= \alpha$. This relation may be useful in multispecies trawl surveys using two trawl hauls whose time durations (hence areas swept) are in the ratio of 1 to 1.65 (which is $\sqrt{e}$). From this, the average increase in species between samples of the two sizes should be a direct measure of the index of diversity (α). To the best of our knowledge, this approach has not be tested in the field with fisheries data and may be worthy of future consideration.

The index of diversity (α) may also be used in a comparison of different populations under certain restrictive assumptions and conditions. Williams (1947) presents results which follow, but these been modified by substitution of an example from a tropical multispecies fishery. Specifically, the method relates to the question of the number of species (groups) that would be common to two areas if they were from the same population. A necessary assumption is that the density of the population be the same for the two areas. That is, the number of individuals (but not species) must be proportional to the area under consideration. Also, the two areas must be of different size. Let the two areas be of size a and size b, and the number of species in each be S_a and S_b. If the areas are from the same population, then they will have the same index of diversity. If the samples are large, the 1 in the following equation can be neglected:

$$S = a \ln (1 + N/\alpha).$$

Let S_{ab} be the total number of species in the two samples. The increase in species by adding b to a is:

$$S_{ab} - S_a = \alpha \ln \left(\frac{a+b}{a}\right) - \alpha \ln \left(\frac{a}{\alpha}\right) = \alpha \ln \left(\frac{a+b}{a}\right) \quad (1)$$

Similarly, the increase in species by adding a to b is:

$$S_{ab} - S_b = \alpha \ln \left(\frac{a+b}{b}\right) \quad (2)$$

The unknowns S_{ab} and α may then be calculated by solving the two equations (Equation 1 and Equation 2) for the two unknowns. It is evident from the two equations above (Equation 1 and Equation 2) that it is required that the area sizes a and b must be different in order for the model to work. The number of species (groups) expected to be common to both samples is:

$$\text{Species (groups) in common} = S_a + S_b - S_{ab}$$

The following example is taken from trawl survey data acquired from the Samar Sea, Philippines, to demonstrate the application of the above-mentioned method. Stations 1-4 are located in the shallow north area of the Samar Sea. These stations had 56 species or species groups in an assumed area of 50 square kilometers. Stations 24-25, located in the shallow south portion of the Samar Sea had 39 species in an assumed area of 25 square kilometers. In this case, the proportions of the two areas are correct, but the absolute sizes have been assumed. However, this does not affect the results. On the assumption of identity of origin, we have that:

$$S_{ab} - 56 = \alpha \ln \left(\frac{75}{50}\right)$$

$$S_{ab} - 39 = \alpha \ln \left(\frac{75}{25}\right)$$

Hence, $S_{ab} = 66$, $\alpha = 24.52$, and the expected number common to the two areas is 29. The actual number of species (groups) observed to be in common was 25. This result does

not suggest any significant deviation from the expected number. However, no formal statistical test of significance is available. In addition, it should be recalled that this exercise is based on the assumption that the number of individuals be proportional to the two areas under consideration. This assumption is clearly difficult to test in practice.

Siromoney (1962) introduced the concept of entropy for the logarithmic series distribution. He considered two types of logarithmic series distributions (Type I and Type II). Type I involves randomization by groups and Type II involves randomization by units.

Type I or randomization by groups may be illustrated by a sample of parasitized fish, classified according to the number of parasites on each fish. An increase in the sample of fish will not add any parasites to the fish already counted. That is, the new units (individual parasites) will fall into the new groups exclusively. In the case of Type II or randomization by units, consider the number of fish caught in a trawl haul which are classified according to species (groups). In this case, an increase in the sample may add new individuals to species already represented. That is, new units may fall into old groups in this case.

Our further interest is restricted to the Type II situation. Consider the following problem. Let f_r represent the frequency of species (groups) of exactly r fish each in a population of fish classified according to species (groups). The Type II distribution equates F_r to $\frac{\alpha B^r}{r}$; $r = 1,2,3,\ldots$ where B is a constant. This is a distribution on the set of all species. Thus, if a fish caught belongs to a species with exactly r fish, its probability is:

$$\frac{r}{\text{total number of fish}} = \frac{r}{\sum rf_r} = \frac{r(1-B)}{\alpha\ B},$$

and there are $\frac{\alpha B^r}{r}$ species with exactly r fish. Let H_α denote the entropy of this distribution. H_α measures the average uncertainty per fish with reference to the species.

Siromoney showed that H_α increases with B, when α is fixed and also that H_α is greater for larger samples from the same biological population.

Assuming a logarithmic series distribution for species abundance, Kempton and Taylor (1979) have shown that the expected number of species in a sample of size m is:

$$S_m(\alpha) = \sum_{i=1}^{m} \frac{\alpha}{\alpha + i - 1}$$

The above authors conclude that using $S_m(\alpha)$ from the log-series model enables all information in the sample to be used, with the consequent reduction in temporal variability and an increase in site (location) discrimination.

Summary of the Log-Series Model and Diversity Statistics

Assuming that the number of species (<u>not</u> the number of individuals) in replicate samples follows a Poisson distribution, and using Fisher's log-series distribution as a description for the expected number of species (E_r) with r individuals:

$$E_r = \alpha X^r/r, \quad r = 1,2,3, \ldots$$

where $\alpha > 0$ and $0 < x < 1$. These are parameters of the log-series distribution. This is a simple model that gives a description of species frequency distributions.

The expected number of species in the sample is given by:

$$E(S) = -\alpha \ln(1 - X).$$

The expected number of individuals in the sample is:

$$E(N) = \alpha X/(1 - X)$$

The maximum likelihood estimate $\hat{\alpha}$ of α is given by the solution of:

$$S = \hat{\alpha} \ln(1 + N/\alpha).$$

The variance of $\hat{\alpha}$ for large samples is:

$$\mathrm{var}(\hat{\alpha}) = \alpha / -\ln(1 - X)$$

Note: This variance can be used for comparing diversities at different sites or times.

The basic model is: the number of species with r individuals in a sample n_r, show independent Poisson variation with expectations E_r, which are described by a logarithmic-series distribution.

The information statistic H, proposed by Shannon and Weaver, is a measure of the information content per symbol in a coded message.

Expressed in terms of species abundance, it is defined as:

$$H = \ln N - 1/N \sum_{i=1}^{\infty} (r/n_r)\, n_r$$

Asymptotically, for large sample size:

$$H \approx V + \ln (\alpha + {}^1/_2)$$

where V = 0.577721, which is Euler's constant, with a relative error less than 10^{-4} when $\alpha \geq 5$. Thus, for data fitting the log-series model, the statistic

$$I = 0.561\, e^H - 0.5$$

may be equated with α with small bias when the sample size is large. The expression e^H is equivalent to the minimum number of species the sample must contain for the information statistic to have value H.

Definition of Symbols

S = total number of species encountered in replicate samples: $S = \sum n_r$.

N = total number of individuals in replicate samples.

n_r = number of species with a particular abundance r.

E_r = expected number of species with r individuals.

α = parameter of log-series distribution: $\alpha > 0$ (diversity index).

X = parameter of log-series distribution: $0 < X < 1$ (sampling parameter).

E(S) = expected number of species in the sample.

$\hat{\alpha}$ = estimate of α

H = The Shannon/Weaver information statistic.

Logarithmic Series Index of Similarity and Its Application to Classification of Stations

Mountford (1962) derived an index of similarity termed I, which is based on the logarithmic series distribution. The material which follows is largely taken from the above reference.

Let: a = the number of species in a sample from one site

b = the number of species in a sample from a second site

j = the number of species common to both sites.

An index of similarity based on the above three quantities must be a combination of these quantities and should be:

a) independent of sample size, and

b) should increase with increasing j and decrease with increasing a and b.

In order to meet these requirements, it is necessary to postulate a theoretical distribution of the species frequencies. The logarithmic series distribution (Fisher, Corbet and Williams, 1943) is postulated as:

$$\alpha x, \frac{\alpha x^2}{2}, \frac{\alpha x^3}{3}, \ldots$$

where αx^n is the number of species with n individuals, and α is a constant for all samples from the same population. The constant α is construed as an index of diversity. The relation between the expected number of individuals (N) and the expected number of species (S) in a sample from a log-series population with the constant α is:

$$S = \alpha \ln (1 + \frac{N}{\alpha})$$

If two samples having A and B individuals and a and b species respectively, are taken from the same population, then:

$$a = \alpha \ln (1 + \frac{A}{\alpha}) \tag{3}$$

$$b = \alpha \ln (1 + \frac{B}{\alpha}) \tag{4}$$

If j is the number of species common to both samples, then if the two samples are considered as one large joint sample, the total number of individuals is now A + B and the total number of species is a + b - j. Then:

$$a + b - j = \alpha \ln \{1 + \frac{(A + B)}{\alpha}\} \tag{5}$$

Taking antilogarithms and eliminating A and B from the above three equations gives:

$$\begin{aligned} e^{a/\alpha} + e^{b/\alpha} &= 1 + \frac{A}{\alpha} + 1 + \frac{B}{\alpha} \\ &= 1 + \{1 + \frac{(A + B)}{\alpha}\} \\ &= 1 + e^{(a+b-j)/\alpha} \end{aligned} \tag{6}$$

The similarity index $I = {}^{1}/\alpha$ and it is the positive root of the equation:

$$e^{aI} + e^{bI} = 1 + e^{(a+b-j)I}, \tag{7}$$

which is the defining relationship for I. Equation (7) can be solved iteratively for I. A good approximation to I is obtained from:

$$\frac{2j}{2ab-(a+b)j}$$

Since α is a constant for the population, then I, which is the inverse of α is independent of sample size for samples drawn from the same population.

Mountford's method for classifying sampling stations into groups of similar stations uses both an index of similarity between a pair of stations and an index of

imilarity between two groups of stations. The index of similarity between station B and a roup composed of stations A_1 and A_2 is defined as:

$$I(A_1A_2, B) = \frac{I(A_1B) + I(A_2B)}{2}$$

vhere $I(A_1B)$ is the index of similarity between the pair of stations A_1 and B. In general erms, the index of similarity between a station B and a group composed of m stations:

$A_1, A_2, \ldots A_m$; B is:

$$I(A_1, A_2, \ldots A_m; B = \frac{I(A_1B) + I(A_2B) + \ldots + I(A_mB)}{m}$$

The index between a group composed of stations A_1 and A_2 and a second group composed of stations B_1 and B_2 is:

$$I(A_1\ A_2; B_1\ B_2) + \frac{I(A_1B_1) + I(A_1B_2) + I(A_2B_1) + I(A_2B_2)}{4}$$

n general terms, the index between groups $A_1, A_2, \ldots A_m$ and $B_2, B_2, \ldots B_m$ is:

$$\frac{1}{mn} \sum_{i=1}^{m} \sum_{i=1}^{n} I(A_iB_j)$$

The following rule is used for classification. From a derived table of indices of similarity select the biggest value. The pair corresponding to this value is combined to form a single group. The indices of similarity between this new group and each of the other stations are evaluated according to the definition of the index between groups of sites.

Shepard (1984) illustrated the application of Mountford's index I to testing the statistical significance of clusters. This problem will be considered in total in a forthcoming report.

Appendix I

Derivation of the Logarithmic Series Distribution

Fisher et al (1943) are responsible for the first mathematical derivation of the logarithmic series distribution. Alternative derivations of this distribution are due to many authors. In this material the derivation by Kendall and Stuart (1958), in which the logarithmic series distribution is considered as a limiting form of the negative binomial distribution, will be briefly described. This derivation is considered intuitively appealing, especially for the application used herein.

Let q represent the probability for the presence of an attribute and p represent the probability for its absence. Then consider the Pascal form of the negative binomial:

$$\binom{x + r - 1}{x} p^r q^x$$

where x is the number of attributes present and r is the number of attributes absent. .

Now consider the above form as a model for the number x of different species found in standardized trawl hauls, for example. If the trawl gear is at all effective, the tow is of reasonable duration and the gear is not torn or broken, there will be at least one species present. Under this assumption, the complete absence of the attribute (x = 0) can be excluded.

We are then concerned with a negative binomial truncated at zero, with frequencies proportional to:

$$p^r \left[rq, \frac{r(r+1)}{2!} q^2, \frac{r(r+1)(r+2)}{3!} q^3, \ldots\right]$$

Since the total frequency is $1-p^r$, the distribution may be written as:

$$\frac{rp^r}{1-p^r} \left[q, \frac{r+1}{2!} q^2, \frac{(r+1)(r+2)}{3!} q^3 \ldots\right] \qquad \text{(A1)}$$

The case where r = o will occur if the size of the samples is large enough so that it is virtually certain that every species in the area is present in at least one of the trawl hauls. Then, it is possible to consider the limiting form, with r = o of Equation (A1):

$$\lim_{r \to o} \frac{r}{p^{-r} - 1} = \lim_{r \to o} \frac{r}{e^{-r\ln P} - 1}$$

Applying L'Hospital's rule we have that:

$$\lim_{r \to o} = \lim_{r \to o} \frac{1}{(-\ln P)e^{-r\ln P}} = \frac{1}{-\ln P}$$

Therefore, the limiting form of Equation (A1) is as follows if we set r = o in the expression in the brackets:

$$\frac{1}{\ln(1-q)} \left[q, \frac{q^2}{2}, \frac{q^3}{3}, \ldots\right]$$

For x = 1, 2, 3, ..., the frequency is the coefficient of t^x in the frequency generating function:

$$P(t) = \frac{1}{\ln(1-q)} \ln (1 - qt) \qquad \text{(A2)}$$

Equation (A2) is an expression for the logarithmic series distribution.

Appendix II

Log-Series Fitting Program—Manual of Operations

The Basic Programs

The LOGSERS.BAS and LOGSRFIT.BAS programs are designed to fit species abundance data from one or more sampling sites to the log-series model. LOGSERS.BAS makes use of only those features that are available as part of the standard BASICA/GWBASIC programming language and must be run within that environment. LOGSRFIT.BAS is the QUICK BASIC (Microsoft v.4.0) version. The structured nature of this version allows one to understand the algorithm more easily than by studying the code of the BASICA version. It is a faster version, which may also be compiled by the QUICK BASIC compiler. This is a distinct advantage when dealing with many sample locations or requesting Whittaker plots.

Program Operations

Data Requirements—These programs are designed to satisfy two end user needs for examining the fit of species abundance data to the log-series model. The first is the quick generation of the parameters of the log-series model fitted to the data from a single sampling site. The second is the generation of these parameters for each of a large number of sampling sites.

Data for multiple sampling sites must be contained in an ASCII file using the following format:

LINE 1	a title (in quotation marks)
LINE 2	the number of sites and the total number of species from all sites (separated by a comma)
LINE 3	label for each column of data (separated by commas; quotation marks are not necessary) (e.g., STATION,SPP1,SPP2,SPP3,SPP4, etc.)
LINE 4	(and all following lines) site id, abundance of species 1, abundance of species 2, etc. (separated by a comma)

If a species were absent from a particular location, a zero should be used to represent its abundance.

Data for a single sampling site may be contained in a data file formatted as described above or entered through the keyboard at run time. There are two options for processing the data from a single site, if those data are entered through the keyboard. Species abundance data can be entered as the user is prompted. This will produce the same output as that when data is read from a file. The other option is to enter only the number of species and total number of individuals. This will provide estimates of the α and x parameters, but will not estimate k (goodness of fit descriptor).

Program Flow—The initial screen gives the program name and a brief description of its use. A menu will appear next, asking the user if his/her data are from multiple sites or a single site only. The user responds with 1 or 2, accordingly.

Choosing single site data leads to a second menu, which provides the user with the choice of entering species abundance data or simply the total number of species and individuals. If species abundance data are used (option 1), the program will ask for the number of species at that site. It will then proceed to prompt the user for the abundance of each species, one at a time. If only number of species and individuals are used (option 2), then the program will request only those two values. Choosing multiple sites data will be followed by a description of the necessary file format and a prompt for the name of the data file.

The first output is to the screen. A hard coy can be acquired by using SHIFT-PrtSc to dump the screen to a printer. This initial output includes estimates of the α and x parameters—as well as the variance and standard error of α (Figure 1).

The initial output is followed by a statement giving the user the option of generating the data for a Whittaker plot (linearized form of the theoretical distribution that best fits the data) for that site. This is a time consuming routine and should only be chosen when it is truly necessary (the code for this routine was taken from Krebs, 1989). If this option is

chosen, the user is prompted for the name of the file in which to record this data. A different name should be used each time this option is chosen, or else the file will be overwritten. The output consists of a list of the species rank, the abundance and proportion at that rank (predicted by the log-series model), and the observed abundance at that site (Figure 2).

If the input data were simply the total number of individuals and species, then the program will end here. For other data, the program proceeds with a second output screen. This output includes the estimate of k (Figure 3). A hard copy can be acquired by using SHIFT-PrtSc.

If the data set included more than one site, the calculations will be repeated for the next site (after a pause). For each site the two output screens and the prompt for Whittaker plots will be generated.

After all of the sites have been processed, the program provides three options for the display or storage of the output from each site. The results for each site may be listed to the screen (choosing S), sent to the printer (choosing P), or stored in a file (choosing F). If storage in a file is chosen, the user will be prompted for a file name. These data are stored in simple ASCII form and may be edited or printed at a later time with a word processor (Figure 4). This output includes a summary of N, S, α, k, and a breakdown for each site, of the number of species that fall into each abundance class.

References

Abramowitz, M. and Stegun, I.A. (eds). 1964. Handbook of mathematical functions with formulas, graphs, and mathematical tables. Applied Math. Series 55. National Bureau of Standards, Washington, DC.

Anscombe, F.J. 1950. Sampling theory of the negative binomial and logarithmic series distributions. Biometrika 37:358-382.

Dial, K.P., and J.M. Marzluff. 1989. Non-random diversification within taxanomic assemblages. Syst. Zool. 38(1):26-37.

Fisher, R.A., A.S. Corbet, and C.B. Williams. 1943. The relation between the number of species and the number of individuals in a random sample of an animal population. Jour. of Animal Ecol. 12:42-58.

Kempton, R.A., and L.R. Taylor. 1979. Some observations on the yearly variability of species abundance at a site and the consistency of measures of diversity. pp 3-22. In: G.P. Patil and M. Rosenzweig (eds.). Contemporary Quantitative Ecology and Related Econometrics. International Cooperative Publishing House, Fairland, Maryland.

Khatri, C.G. 1959. On certain properties of power series distributions. Biometrika 46:480--490.

Kobayashi, S. 1987. Heterogeneity ratio: A measure of Beta-diversity and its use in community classification. Ecol. Res. 2:101-111.

Koch, C.F. 1987. Prediction of sample size effects on the measured temporal and geographic distribution pattern of species. Paleobiology 13(1):100-107.

Krebs, C.A. 1989. Ecological methodology. Harper and Row, New York.

Mountford, M.D. 1962. An index of similarity and its application to classificatory problems. In: Progress in Soil Ecology. P.W. Murphy (ed.). pp. 43-50. Butterworths, London.

Noack, A. 1950. A class of random variables with discrete distributions. Am Math. Stat. 21:127-132.

Patil, G.P. 1961. Asymptotic bias and variance of ratio estimates in generalized power series distributions and certain applications. Sankhya A. 23-269-280.

Pielou, E.C. 1975. Ecological Diversity. John Wiley and Sons, New York.

_____. 1977. Mathematical Ecology. John Wiley and Sons, New York.

Saeger, J. 1983. Results of the Samar Sea trawl survey. Dept. Mar. Fish. Tech. Rep. 3, 1-191.

Shepard, R.B. 1984. The log-series distribution and Mountford's similarity index as a basis for the study of stream benthic community structure. Freshw. Biol.

14:53-71.

Siromoney, G. 1961. Entropy of logarithmic series distributions. Sankhya A. 24:419-420.

Taylor, L.R., R.A. Kempton, and I.P. Woiwood. 1976. Diversity statistics and the log-series model. Jour. of Animal Ecol. 45:255-272.

Williams, C.B. 1944. Some applications of the logarithmic series and the index of diversity to ecological problems. Jour. of Animal Ecol. 36:1-44.

_____. 1947. The generic relations of species in sample ecological communities. Jour. of Animal Ecol. 16:11-18.

_____. 1964. Patterns in the balance of nature. Academic Press, London.

Wolda, H. 1981. Similarity indices, sample size and diversity. Oecologia 50:296-302.

```
10  '  ***** LOGSERS.BAS *****
20  '
30 PRINT "        ****** LOGSERS.BAS *******"
40 PRINT
50 PRINT "                by Jim McKenna 18 November 1988"
60 PRINT : PRINT
70 PRINT "  This program will read data on the abundance of each of a set"
80 PRINT " of species from a number of stations and will print the station #,"
90 PRINT " total # of individuals, total # of species, Fisher's  , X, the "
100 PRINT " frequency of species in abundance classes, and the associated K "
110 PRINT " (from a negative binomial distribution)."
120  '
130 PRINT : PRINT
140 PRINT "Strike any key to continue ...": WHILE INKEY$ = "": WEND
150 DIM J(500), M(500):  ' Internal to LOGSER
160 DIM WH(1000):   ' Part of Krebs' code for Whittaker plot data
170  '
180 PRINT TAB(10); "****************************************"
190 PRINT TAB(10); "*"; TAB(50); "*"
200 PRINT TAB(10); "*"; TAB(21); "1) SINGLE STATION"; TAB(50); "*"
210 PRINT TAB(10); "*"; TAB(50); "*"
220 PRINT TAB(10); "*"; TAB(21); "2) MULTIPLE STATIONS"; TAB(50); "*"
230 PRINT TAB(10); "*"; TAB(23); "(requires an input file)"; TAB(50); "*"
240 PRINT TAB(10); "*"; TAB(50); "*"
250 PRINT TAB(10); "****************************************"
260 PRINT
270 INPUT "Please enter your choice"; ANS4%
280 ON ANS4% GOTO 290, 320
290 GOSUB 3900:   'SINGLE-station input routine
300 GOTO 340
310  '
320 PRINT " Go to INPUT subroutine"
330 GOSUB 500:   ' INPUT
340 PRINT : PRINT
350 PRINT " Compute   and K for each station"
360 FOR ICOUNT = 1 TO ST
370    GOSUB 1450:   ' SUB. to compute
380  '  Prompt with request for Whittaker Plot data (Krebs 1987, ch. 10)
390    INPUT "Do you want the predicted abundances for a Whittaker plot";
ANS5$
400    IF ANS5$ = "Y" OR ANS5$ = "y" THEN GOSUB 2710
410    GOSUB 1820:  ' SUB. to compute K
420 NEXT ICOUNT
430 PRINT
440 PRINT " Go to OUTPUT subroutine"
450 GOSUB 2350
460 PRINT
470 PRINT "   ******** END ************"
480 END
490  '
500 PRINT : PRINT " Subroutine to read data and produce counts and
frequencies."
510 PRINT
520 PRINT " The data must be in the following form: "
```

```
530 PRINT " LINE 1: a title in quotes"
540 PRINT " LINE 2: # of stations, # of species"
550 PRINT " LINE 3: STATION and a label for each other column"
560 PRINT " STATION 1, abund SPP 1, abund SPP 2, ..., abund SPP S"
570 PRINT " STATION 2, abund SPP 1, abund SPP 2, ..., abund SPP S"
580 PRINT " ."
590 PRINT " ."
600 PRINT " ."
610 PRINT " STATION N, ..."
620 PRINT
630 INPUT "Enter the name of the data file "; FIL1$
640 OPEN "I", 1, FIL1$
650 INPUT #1, TITL$
660 INPUT #1, ST, S
670  ' READ COLUMN LABELS
680 FOR I% = 1 TO S + 1
690         INPUT #1, X$
700 NEXT I%
710  '
720  ' Dimension arrays
730  '
740 DIM STX(ST), CLASS(ST, 2, S), TOTN(ST), TOTS(ST), ALPHA(ST), VAR(ST)
750 DIM KSTAT(ST), MXCL(ST), KSIG$(ST)
760  '
770 PRINT " Initializing arrays to zeros ..."
780 GOSUB 1070
790  '
800 PRINT " Filling arrays with data ..."
810  '
820 FOR I = 1 TO ST
830    MAXCLASS = 1:   ' MAX. # of abund. classes found for this station so
far
840    INPUT #1, STX(I):   'Read the station #
850    FOR J = 1 TO S
860       INPUT #1, NUM:   ' Read abundance value for species J
870       TOTN(I) = TOTN(I) + NUM
880       IF NUM <= 0 THEN GOTO 1000
890          TOTS(I) = TOTS(I) + 1
900          FOR K = 1 TO MAXCLASS
905   'Count # of spp in this abund. class
910             IF NUM = CLASS(I, 1, K) THEN CLASS(I, 2, K) = CLASS(I, 2, K) +
1: GOTO 1000
950          NEXT K
960          CLASS(I, 1, MAXCLASS) = NUM:  ' Make new abundance class.
970          CLASS(I, 2, MAXCLASS) = 1
980          MAXCLASS = MAXCLASS + 1:  ' Increment max. number of abund.
classes
1000 NEXT J
1010    MXCL(I) = MAXCLASS - 1:  ' # of abund. classes for this station.
1020    GOSUB 1290:   ' Sort abundance classes
1030 NEXT I
1040  '
1050 RETURN
1060  '
```

```
1070  ' Sub. to initialize arrays
1080 FOR I = 1 TO ST
1090     STX(I) = 0
1100     TOTN(I) = 0
1110     TOTS(I) = 0
1120     ALPHA(I) = 0
1130     KSTAT(I) = 0
1140     MXCL(I) = 0
1150     FOR J = 1 TO 2
1160        FOR K = 1 TO S
1170           CLASS(I, J, K) = 0
1180        NEXT K
1190     NEXT J
1200 NEXT I
1210  '
1220 FOR I = 1 TO 500
1230     J(I) = 0
1240     M(I) = 0
1250 NEXT I
1260  '
1270 RETURN
1280  '
1290 PRINT ">";
1300 FOR II = 1 TO MXCL(I) - 1
1310     FOR JJ = II + 1 TO MXCL(I)
1320        IF CLASS(I, 1, JJ) > CLASS(I, 1, II) GOTO 1400
1330           TEMPC = CLASS(I, 1, II)
1340           TEMPF = CLASS(I, 2, II)
1350           CLASS(I, 1, II) = CLASS(I, 1, JJ)
1360           CLASS(I, 2, II) = CLASS(I, 2, JJ)
1370           CLASS(I, 1, JJ) = TEMPC
1380           CLASS(I, 2, JJ) = TEMPF
1400     NEXT JJ
1410 NEXT II
1420  '
1430 RETURN
1440  '
1450 PRINT " *****   *****":  ' ******* LOGSER ******
1460 J(1) = 50
1470 M(1) = 1
1480  '
1490 FOR I = 1 TO 260
1500        L = (M(I) + J(I)) / 2
1510        K = L * (EXP(TOTS(ICOUNT) / L) - 1) - TOTN(ICOUNT)
1520        IF (M(I) * (EXP(TOTS(ICOUNT) / M(I)) - 1) - TOTN(ICOUNT)) * K < 0
GOTO 1560
1530        M(I + 1) = L
1540        J(I + 1) = J(I)
1550        GOTO 1580
1560        M(I + 1) = M(I)
1570        J(I + 1) = L
1580        IF ABS(M(I + 1) - J(I + 1)) < .000001 GOTO 1600
1590 NEXT I
1600        ALPHA(ICOUNT) = L
```

```
1610  '
1620 PRINT "STATION #                    :"; STX(ICOUNT)
1630 PRINT "Number of Individuals (N)    :"; TOTN(ICOUNT)
1640 PRINT "Number of species (S)        :"; TOTS(ICOUNT): PRINT
1650 PRINT "                             :"; ALPHA(ICOUNT)
1660 X = TOTN(ICOUNT) / (TOTN(ICOUNT) + L)
1670  ' Calc. of variance from equation in Krebs (1987 - ch.10, eq'n 10.12)
1680 VAR(ICOUNT) = -ALPHA(ICOUNT) / (LOG(1! - X))
1690 PRINT "Variance of                  :"; VAR(ICOUNT)
1700 PRINT "  X                          :"; X
1710 C = SQR(TOTN(ICOUNT))
1720 PRINT "Standard Error of N          :"; C
1730 D = LOG((2 * TOTN(ICOUNT) + L) / (TOTN(ICOUNT) + L))
1740 E = SQR((L * D) - ((L ^ 2) * TOTN(ICOUNT)) / (TOTN(ICOUNT) + L) ^ 2)
1750 PRINT "Standard Error of S          :"; E
1760 F = SQR(((L ^ 3) * (((TOTN(ICOUNT) + L) ^ 2) * D - L * TOTN(ICOUNT))) /
((TOTS(ICOUNT) * TOTN(ICOUNT) + TOTS(ICOUNT) * L - TOTN(ICOUNT) * L) ^ 2))
1770 PRINT "Standard Error of            :"; F
1780 PRINT : PRINT
1790 PRINT "Strike any key to continue ..."; WHILE INKEY$ = "": WEND
1800 RETURN
1810  '
1820 PRINT " ***** K ***** ":  ' ********* KTEST *********
1830 Z = 0
1840 L = 0
1850 B = 0
1860 Q = MXCL(ICOUNT):   ' MAX. # of classes for this station
1870  '
1880  'Calculations:
1890  ' Determine if the first abund. class is one or not
1900 IF CLASS(ICOUNT, 1, 1) > 1 THEN START = 1 ELSE START = 2
1950  '
1960 FOR P = START TO Q
1970    X = 0
1980    FOR I = 1 TO CLASS(ICOUNT, 1, P) - 1
1990       T = (1 / I) + X
2000       X = T
2010    NEXT I
2020    Z = CLASS(ICOUNT, 2, P) * T + B
2030    B = Z
2040 NEXT P
2050  '
2060 C = (TOTS(ICOUNT) ^ 2) / (2 * ALPHA(ICOUNT))
2070 D = B - C
2080 I1 = ((.43429 * LOG(TOTN(ICOUNT) / TOTS(ICOUNT))) ^ 1.673) * .908 *
TOTS(ICOUNT)
2090 I2 = SQR(I1)
2100  '
2110 PRINT " Total # of individuals      :"; TOTN(ICOUNT)
2120 PRINT " Total # of species          :"; TOTS(ICOUNT)
2130 PRINT "                             :"; ALPHA(ICOUNT)
2140 PRINT
2150 PRINT " Observed K score            :"; B
2160 PRINT " Expected K score            :"; C
```

```
2170 PRINT " Difference æ std. Err.      :"; ABS(D); "æ"; I2
2180 IF ABS(D) < I2 GOTO 2270
2190    K = ABS(D / I1)
2200    PRINT " K value ...              :"; K
2210    IF K > .5 THEN KSIG$(ICOUNT) = "#" ELSE KSIG$(ICOUNT) = " "
2220    GOTO 2300
2270    PRINT " K is sufficiently close to zero!"
2280    KSIG$(ICOUNT) = "*"
2300 KSTAT(ICOUNT) = ABS(D / I1)
2310  PRINT "Strike any key to continue ...": WHILE INKEY$ = "": WEND
2320  '
2330 RETURN
2340  '
2350 PRINT " Output :....."
2360 INPUT "Would you like the output sent to the screen (1), the printer (2),
or a file (3)"; DESTIN%
2370 ON DESTIN% GOTO 2390, 2400, 2410
2380  '
2390  OPEN "O", 2, "SCRN:": GOTO 2430
2400  OPEN "O", 2, "LPT1:": GOTO 2430
2410  INPUT "Enter output file name :", FIL2$
2420  OPEN "O", 2, FIL2$
2430  INPUT "ENTER a title : ", TITL$
2440 PRINT #2, DATE$, TIME$
2450 PRINT #2, TITL$
2460 PRINT #2, " "
2470 PRINT #2, TAB(5); "STATION"; TAB(15); "N"; TAB(20); "S"; TAB(27); " ";
TAB(38); "K"
2480 FOR I = 1 TO ST
2490         PRINT #2, TAB(7); STX(I); TAB(13); TOTN(I); TAB(20); TOTS(I);
TAB(24);
2500         PRINT #2, USING "##.#### "; ALPHA(I);
2510         PRINT #2, TAB(35);
2520         PRINT #2, USING "#.#### "; KSTAT(I);
2530         PRINT #2, KSIG$(I)
2540
2550         PRINT #2, "ABUND. CLASS  :"; TAB(17);
2560         FOR J = 1 TO MXCL(I)
2570                 PRINT #2, USING " #### "; CLASS(I, 1, J);
2580         NEXT J
2590         PRINT #2, " "
2600         PRINT #2, "    # OF TAXA :"; TAB(17);
2610         FOR J = 1 TO MXCL(I)
2620                 PRINT #2, USING " #### "; CLASS(I, 2, J);
2630         NEXT J
2640         PRINT #2, " "
2650         PRINT #2, " "
2660 NEXT I
2670  '
2680  '
2690 RETURN
2700  '
2710  '      PROGRAM LOGSERIE.FOR   Version 1.1
2720  '      THIS PROGRAM CALCULATES THE PARAMETERS OF THE LOGARITHMIC SERIES
```

```
2730  '      OF FISHER, CORBET AND WILLIAMS (1943) USING TABLE 146 AND THE
2740  '      RECURSIVE EQUATION ON PAGE 309 IN WILLIAMS (1964).
2750  '      PROGRAMMMED BY C. KREBS IN JANUARY 1987.
2760  'This is simply the subprogram of the above program that calculates the
values
2770  ' for Whittaker Plots
2780  '      THEORETICAL WHITTAKER PLOT VALUES
2790  ' Initialize WH() array
2800          FOR I% = 1 TO 1000
2810                  WH(I%) = 0
2820          NEXT I%
2830        PRINT
2840         ' PRINT "* Toggle PRINTER on if you have more than 20 species !!"
2850        PRINT
2860        ZZUM# = 0!
2870  INPUT "Would you like the data for the Whittaker plot to go to a file
(F), the printer (P), or the screen only (S)"; ANS7$
2880        IF ANS7$ = "p" OR ANS7$ = "P" THEN FILX$ = "LPT1:"
2910        IF ANS7$ = "S" OR ANS7$ = "s" THEN FILX$ = "SCRN:"
2940        IF ANS7$ = "F" OR ANS7$ = "f" THEN INPUT "Enter a file name:",
FILX$
3010        OPEN "O", 3, FILX$
3020        PRINT "       DATA FOR THEORETICAL WHITTAKER PLOT OF SPECIES
ABUNDANCES"
3030        PRINT #3, "       DATA FOR THEORETICAL WHITTAKER PLOT OF SPECIES
ABUNDANCES"
3040        PRINT : PRINT
3050        KK% = MXCL(ICOUNT)
3060        XN# = TOTN(ICOUNT)
3070        IF (XN# < 500!) THEN XUMP# = 1!
3080        IF (XN# >= 500 AND XN# < 1000!) THEN XUMP# = 2!
3090        IF (XN# >= 1000! AND XN# < 2000) THEN XUMP# = 5!
3100        IF (XN# > 2000!) THEN XUMP# = 10!
3110        Y9# = LOG(1! + (ALPHA(ICOUNT) / XN#))
3120        FOR I% = 1 TO KK%
3130          XX% = I%
3140          X3# = XN# / (XX% + 1!)
3150          X3# = INT(X3#)
3160          N8% = 0
3170          I45% = 1
3180          I47% = 0
3190          Y# = X3# * Y9#
3200          IF (I47% = I45%) GOTO 3440
3210          IF (X3# < 300!) THEN XUMP# = .1
3220          IF (X3# < 500 AND X3# >= 300!) THEN XUMP# = 1!
3230          IF (X3# >= 500 AND X3# < 1000!) THEN XUMP# = 2!
3240          IF (X3# >= 1000! AND X3# < 2000) THEN XUMP# = 5!
3250          IF (X3# > 2000!) THEN XUMP# = 10!
3260  ' statement to carry proper argument to subroutine
3270          XSUB# = Y#
3280          GOSUB 3730:   'SUBroutine SEI#
3290          Z2# = ALPHA(ICOUNT) * SEI#
3300          N8% = N8% + 1
3310          DD = Z2# - XX%
```

```
3320          IF (N8% > 3000) GOTO 3440
3330          IF (DD > .001) GOTO 3360
3340          IF (DD < -.001) GOTO 3390
3350          GOTO 3440
3360          X3# = X3# + XUMP#
3370          I45% = 99
3380          GOTO 3190
3390          IF (X3# - XUMP# > 0!) GOTO 3410
3400          GOTO 3440
3410          I47% = 99
3420          X3# = X3# - XUMP#
3430          GOTO 3190
3440          PRINT I%;
3450          X3# = X3# - XUMP#
3460          WH(I%) = X3#
3470       NEXT I%
3480          PRINT
3490        PRINT "  SPECIES RANK    NUMBER OF INDIVIDUALS       PROPORTIONS
DATA"
3500        PRINT #3, "  SPECIES RANK    NUMBER OF INDIVIDUALS
PROPORTIONS      DATA"
3510        SUM# = 0!
3520     FOR I% = 1 TO KK%
3530          SUM# = SUM# + WH(I%)
3540     NEXT I%
3550        CORR# = XN# / SUM#
3560        ICONT% = 0
3570     FOR I% = 1 TO KK%
3580        ICONT% = ICONT% + 1
3590        WH(I%) = WH(I%) * CORR#
3600        Z6# = WH(I%) / XN#
3610        PRINT TAB(9); I%; TAB(24);
3612        PRINT USING "#####.#"; WH(I%);
3614        PRINT TAB(48);
3616        PRINT USING "#.###^^^^"; Z6#;
3618        PRINT TAB(63); CLASS(ICOUNT, 1, KK% - (I% - 1))
3620        PRINT #3, TAB(9); I%; TAB(24);
3622        PRINT #3, USING "#####.#"; WH(I%);
3624        PRINT #3, TAB(48);
3626        PRINT #3, USING "#.###^^^^"; Z6#;
3628        PRINT #3, TAB(63); CLASS(ICOUNT, 1, KK% - (I% - 1))
3630        IF (ICONT% >= 24) THEN PRINT "....Hit < ENTER > to continue
......": WHILE INKEY$ = "": WEND
3670        IF (ICONT% >= 24) THEN ICONT% = 0
3680       NEXT I%
3690          PRINT
3700          CLOSE 3
3710        RETURN
3720  '
3730  '       DOUBLE PRECISION FUNCTION SEI(XSUB)
3740  '      STANDARD EXPONENTIAL INTEGRAL FUNCTION
3750  '      APPROXIMATION FORMULAE FROM ABRAMOWITZ AND STEGUN (1964) CHAP. 5
3760        A# = -.57721566#: A1# = .99999193#: A2# = -.24991055#: A3# =
.05519968#
```

```
3770        A4# = -9.760039999999999D-03: A5# = .00107857#
3780        AA# = 2.334733: AB# = .250621: B1# = 3.330657: B2# = 1.681534
3790        ZSUB# = XSUB#
3800        SEI# = 0!
3810        IF (ZSUB# > 0! AND ZSUB# <= 1!) GOTO 3840
3820        IF (ZSUB# > 1!) GOTO 3860
3830        GOTO 3880
3840       SEI# = -LOG(ZSUB#) + A# + (A1# * ZSUB#) + (A2# * ZSUB# ^ 2) + (A3# *
ZSUB# ^ 3) + (A4# * ZSUB# ^ 4) + (A5# * ZSUB# ^ 5)
3850        RETURN
3860       SEI# = (1! / (ZSUB# * EXP(ZSUB#))) * ((ZSUB# ^ 2) + (AA# * ZSUB#) +
AB#) / ((ZSUB# ^ 2) + (B1# * ZSUB#) + B2#)
3870        RETURN
3880       RETURN
3890  '
3900  ' subroutine to allow keyboard entry of data for a single station
3910 PRINT
3920  ' Initialize variables
3930 ST = 1
3940 DIM STX(ST), TOTN(ST), TOTS(ST), ALPHA(ST), VAR(ST)
3950 DIM KSTAT(ST), MXCL(ST), KSIG$(ST)
3960  '
3970 PRINT TAB(10); "****************************************"
3980 PRINT TAB(10); "*"; TAB(50); "*"
3990 PRINT TAB(10); "*"; TAB(21); "1) Species Abundance"; TAB(50); "*"
4000 PRINT TAB(10); "*"; TAB(21); "   Data"; TAB(50); "*"
4010 PRINT TAB(10); "*"; TAB(50); "*"
4020 PRINT TAB(10); "*"; TAB(21); "2) # of Species & "; TAB(50); "*"
4030 PRINT TAB(10); "*"; TAB(21); "   # of Individuals only "; TAB(50); "*"
4040 PRINT TAB(10); "*"; TAB(50); "*"
4050 PRINT TAB(10); "****************************************"
4060 PRINT
4070 INPUT "Please enter your choice"; ANS4%
4080 ON ANS4% GOTO 4100, 4350
4090  '
4100 INPUT "Enter the number of species"; S
4110 DIM CLASS(ST, 2, S)
4120    MAXCLASS = 1:   ' MAX. # of abund. classes found for this station so
far
4130    STX(1) = 1:  'SET the station #
4140    FOR J = 1 TO S
4150         PRINT "Enter the number of individuals of species"; J; ":",
4160        INPUT NUM:    ' Read abundance value for species J
4170        TOTN(1) = TOTN(1) + NUM
4180        IF NUM <= 0 GOTO 4300
4190           TOTS(1) = TOTS(1) + 1
4200           FOR K = 1 TO MAXCLASS
4215  'Count # of spp in this abund. class
4210              IF NUM = CLASS(1, 1, K) THEN CLASS(1, 2, K) = CLASS(1, 2, K)
+ 1: GOTO 4300
4250           NEXT K
4260           CLASS(1, 1, MAXCLASS) = NUM:  ' Make new abundance class.
4270           CLASS(1, 2, MAXCLASS) = 1
```

```
          MAXCLASS = MAXCLASS + 1:  ' Increment max. number of abund.
ses
  NEXT J
    MXCL(1) = MAXCLASS - 1:  ' # of abund. classes for this station.
    GOSUB 1290:   ' Sort abundance classes
 RETURN
  '
 INPUT "Enter the number of species:", S
         TOTS(1) = S
)        MXCL(1) = S:   ' SET max number of classes to the number of
cies
)                       ' - this might not be correct.
0      INPUT "Enter the total number of individuals:", TOTN(1)
0      ICOUNT = 1
0  '
0 GOSUB 1450:   ' Calc.   and X
0  '  Prompt with request for Whittaker Plot data (Krebs 1987, ch. 10)
0    INPUT "Do you want the predicted abundances for a Whittaker plot";
5$
0    IF ANS5$ = "Y" OR ANS5$ = "y" THEN GOSUB 2710
0    STOP
0 RETURN
```

```
'  ***** LOGSRFIT.BAS *****
'
PRINT "        ****** LOGSRFIT.BAS *******"
PRINT
PRINT "                by Jim McKenna 18 November 1988"
PRINT : PRINT
PRINT "  This program will read data on the abundance of each of a set"
PRINT " of species from a number of stations and will print the station #,"
PRINT " total # of individuals, total # of species, Fisher's  , X, the "
PRINT " frequency of species in abundance classes, and the associated K "
PRINT " (from a negative binomial distribution)."
'
PRINT : PRINT
PRINT "Strike any key to continue ...": WHILE INKEY$ = "": WEND
DIM J(500), M(500): ' Internal to LOGSER
DIM WH(1000):  ' Part of Krebs' code for Whittaker plot data
'
PRINT TAB(10); "****************************************"
PRINT TAB(10); "*"; TAB(50); "*"
PRINT TAB(10); "*"; TAB(21); "1) SINGLE STATION"; TAB(50); "*"
PRINT TAB(10); "*"; TAB(50); "*"
PRINT TAB(10); "*"; TAB(21); "2) MULTIPLE STATIONS"; TAB(50); "*"
PRINT TAB(10); "*"; TAB(23); "(requires an input file)"; TAB(50); "*"
PRINT TAB(10); "*"; TAB(50); "*"
PRINT TAB(10); "****************************************"
PRINT
INPUT "Please enter your choice"; ANS4%
ON ANS4% GOTO 10, 20
10 GOSUB 8000:  'SINGLE-station input routine
GOTO 25
'
20 PRINT " Go to INPUT subroutine"
GOSUB 1000:  ' INPUT
25 PRINT : PRINT
PRINT " Compute   and K for each station"
FOR ICOUNT = 1 TO ST
   GOSUB 2000:  ' SUB. to compute
'  Prompt with request for Whittaker Plot data (Krebs 1987, ch. 10)
   INPUT "Do you want the predicted abundances for a Whittaker plot"; ANS5$
   IF ANS5$ = "Y" OR ANS5$ = "y" THEN GOSUB 5000
   GOSUB 3000: ' SUB. to compute K
NEXT ICOUNT
PRINT
PRINT " Go to OUTPUT subroutine"
GOSUB 4000
PRINT
PRINT "   ******** END ************"
END
'
1000 PRINT : PRINT " Subroutine to read data and produce counts and
frequencies."
PRINT
PRINT " The data must be in the following form: "
PRINT " LINE 1: a title in quotes"
```

```
PRINT " LINE 2: # of stations, # of species"
PRINT " LINE 3: STATION and a label for each other column"
PRINT " STATION 1, abund SPP 1, abund SPP 2, ..., abund SPP S"
PRINT " STATION 2, abund SPP 1, abund SPP 2, ..., abund SPP S"
PRINT " ."
PRINT " ."
PRINT " ."
PRINT " STATION N, ..."
PRINT
INPUT "Enter the name of the data file "; FIL1$
OPEN "I", 1, FIL1$
INPUT #1, TITL$
INPUT #1, ST, S
' READ COLUMN LABELS
FOR I% = 1 TO S + 1
        INPUT #1, X$
NEXT I%
'
' Dimension arrays
'
DIM STX(ST), CLASS(ST, 2, S), TOTN(ST), TOTS(ST), ALPHA(ST), VAR(ST)
DIM KSTAT(ST), MXCL(ST), KSIG$(ST)
'
PRINT " Initializing arrays to zeros ..."
GOSUB 1500
'
PRINT " Filling arrays with data ..."
'
FOR I = 1 TO ST
   MAXCLASS = 1:  ' MAX. # of abund. classes found for this station so far
   INPUT #1, STX(I):  'Read the station #
   FOR J = 1 TO S
      INPUT #1, NUM:  ' Read abundance value for species J
      TOTN(I) = TOTN(I) + NUM
      IF NUM > 0 THEN
         TOTS(I) = TOTS(I) + 1
         FOR K = 1 TO MAXCLASS
            IF NUM = CLASS(I, 1, K) THEN
               CLASS(I, 2, K) = CLASS(I, 2, K) + 1: 'Count # of spp in this
abund. class
               GOTO 1111
            END IF
         NEXT K
         CLASS(I, 1, MAXCLASS) = NUM: ' Make new abundance class.
         CLASS(I, 2, MAXCLASS) = 1
         MAXCLASS = MAXCLASS + 1: ' Increment max. number of abund. classes
      END IF
1111 NEXT J
   MXCL(I) = MAXCLASS - 1: ' # of abund. classes for this station.
   GOSUB 1800:  ' Sort abundance classes
NEXT I
'
RETURN
'
```

```
.500 ' Sub. to initialize arrays
'OR I = 1 TO ST
   STX(I) = 0
   TOTN(I) = 0
   TOTS(I) = 0
   ALPHA(I) = 0
   KSTAT(I) = 0
   MXCL(I) = 0
   FOR J = 1 TO 2
      FOR K = 1 TO S
         CLASS(I, J, K) = 0
      NEXT K
   NEXT J
NEXT I

FOR I = 1 TO 500
   J(I) = 0
   M(I) = 0
NEXT I

RETURN

1800 PRINT ">";
FOR II = 1 TO MXCL(I) - 1
   FOR JJ = II + 1 TO MXCL(I)
      IF CLASS(I, 1, JJ) < CLASS(I, 1, II) THEN
         TEMPC = CLASS(I, 1, II)
         TEMPF = CLASS(I, 2, II)
         CLASS(I, 1, II) = CLASS(I, 1, JJ)
         CLASS(I, 2, II) = CLASS(I, 2, JJ)
         CLASS(I, 1, JJ) = TEMPC
         CLASS(I, 2, JJ) = TEMPF
      END IF
   NEXT JJ
NEXT II

RETURN

2000 PRINT " *****   *****": ' ******* LOGSER ******
J(1) = 50
M(1) = 1
'
FOR I = 1 TO 260
      L = (M(I) + J(I)) / 2
      K = L * (EXP(TOTS(ICOUNT) / L) - 1) - TOTN(ICOUNT)
      IF (M(I) * (EXP(TOTS(ICOUNT) / M(I)) - 1) - TOTN(ICOUNT)) * K < 0 GOTO
2440
      M(I + 1) = L
      J(I + 1) = J(I)
      GOTO 2460
2440  M(I + 1) = M(I)
      J(I + 1) = L
2460  IF ABS(M(I + 1) - J(I + 1)) < .000001 GOTO 2480
NEXT I
```

```
2480  ALPHA(ICOUNT) = L
'
PRINT "STATION #                    :"; STX(ICOUNT)
PRINT "Number of Individuals (N)    :"; TOTN(ICOUNT)
PRINT "Number of species (S)        :"; TOTS(ICOUNT): PRINT
PRINT "                             :"; ALPHA(ICOUNT)
X = TOTN(ICOUNT) / (TOTN(ICOUNT) + L)
' Calc. of variance from equation in Krebs (1987 - ch.10, eq'n 10.12)
VAR(ICOUNT) = -ALPHA(ICOUNT) / (LOG(1! - X))
PRINT "Variance of                  :"; VAR(ICOUNT)
PRINT "  X                          :"; X
C = SQR(TOTN(ICOUNT))
PRINT "Standard Error of N          :"; C
D = LOG((2 * TOTN(ICOUNT) + L) / (TOTN(ICOUNT) + L))
E = SQR((L * D) - ((L ^ 2) * TOTN(ICOUNT)) / (TOTN(ICOUNT) + L) ^ 2)
PRINT "Standard Error of S          :"; E
F = SQR(((L ^ 3) * (((TOTN(ICOUNT) + L) ^ 2) * D - L * TOTN(ICOUNT))) /
((TOTS(ICOUNT) * TOTN(ICOUNT) + TOTS(ICOUNT) * L - TOTN(ICOUNT) * L) ^ 2))
PRINT "Standard Error of            :"; F
PRINT : PRINT
PRINT "Strike any key to continue ...": WHILE INKEY$ = "": WEND
RETURN
'
3000 PRINT " ***** K ***** ": ' ********* KTEST *********
Z = 0
L = 0
B = 0
Q = MXCL(ICOUNT):  ' MAX. # of classes for this station
'
'Calculations:
' Determine if the first abund. class is one or not
IF CLASS(ICOUNT, 1, 1) > 1 THEN
   START = 1
  ELSE
   START = 2
END IF
'
FOR P = START TO Q
   X = 0
   FOR I = 1 TO CLASS(ICOUNT, 1, P) - 1
      T = (1 / I) + X
      X = T
   NEXT I
   Z = CLASS(ICOUNT, 2, P) * T + B
   B = Z
NEXT P
'
C = (TOTS(ICOUNT) ^ 2) / (2 * ALPHA(ICOUNT))
D = B - C
I1 = ((.43429 * LOG(TOTN(ICOUNT) / TOTS(ICOUNT))) ^ 1.673) * .908 *
TOTS(ICOUNT)
I2 = SQR(I1)
'
PRINT " Total # of individuals      :"; TOTN(ICOUNT)
```

```
PRINT " Total # of species          :"; TOTS(ICOUNT)
PRINT "                             :"; ALPHA(ICOUNT)
PRINT
PRINT " Observed K score            :"; B
PRINT " Expected K score            :"; C
PRINT " Difference æ std. Err.      :"; ABS(D); "æ"; I2
IF ABS(D) >= I2 THEN
   K = ABS(D / I1)
   PRINT " K value ...              :"; K
   IF K > .5 THEN
      KSIG$(ICOUNT) = "#"
     ELSE
      KSIG$(ICOUNT) = " "
   END IF
 ELSE
   PRINT " K is sufficiently close to zero!"
   KSIG$(ICOUNT) = "*"
END IF
KSTAT(ICOUNT) = ABS(D / I1)
 PRINT "Strike any key to continue ...": WHILE INKEY$ = "": WEND
'
RETURN
'
4000 PRINT " Output :....."
INPUT "Would you like the output sent to the screen (1), the printer (2), or a
file (3)"; DESTIN%
ON DESTIN% GOTO 4100, 4110, 4120
'
4100 OPEN "O", 2, "SCRN:": GOTO 4150
4110 OPEN "O", 2, "LPT1:": GOTO 4150
4120 INPUT "Enter output file name :", FIL2$
     OPEN "O", 2, FIL2$
4150 INPUT "ENTER a title : ", TITL$
PRINT #2, DATE$, TIME$
PRINT #2, TITL$
PRINT #2, " "
PRINT #2, TAB(5); "STATION"; TAB(15); "N"; TAB(20); "S"; TAB(27); " ";
TAB(38); "K"
FOR I = 1 TO ST
        PRINT #2, TAB(7); STX(I); TAB(13); TOTN(I); TAB(20); TOTS(I); TAB(24);
        PRINT #2, USING "##.#### "; ALPHA(I);
        PRINT #2, TAB(35);
        PRINT #2, USING "#.#### "; KSTAT(I);
        PRINT #2, KSIG$(I)

        PRINT #2, "ABUND. CLASS  :"; TAB(17);
        FOR J = 1 TO MXCL(I)
                PRINT #2, USING " #### "; CLASS(I, 1, J);
        NEXT J
        PRINT #2, " "
        PRINT #2, "    # of TAXA :"; TAB(17);
        FOR J = 1 TO MXCL(I)
                PRINT #2, USING " #### "; CLASS(I, 2, J);
        NEXT J
```

```
        PRINT #2, " "
        PRINT #2, " "
NEXT I
'
'
RETURN
'
5000 '     PROGRAM LOGSERIE.FOR   Version 1.1
'     THIS PROGRAM CALCULATES THE PARAMETERS OF THE LOGARITHMIC SERIES
'     OF FISHER, CORBET AND WILLIAMS (1943) USING TABLE 146 AND THE
'     RECURSIVE EQUATION ON PAGE 309 IN WILLIAMS (1964).
'     PROGRAMMMED BY C. KREBS IN JANUARY 1987.
'This is simply the subprogram of the above program that calculates the values
' for Whittaker Plots
'     THEORETICAL WHITTAKER PLOT VALUES
' Initialize WH() array
        FOR I% = 1 TO 1000
                WH(I%) = 0
        NEXT I%
      PRINT
      ' PRINT "* Toggle PRINTER on if you have more than 20 species !!"
      PRINT
      ZZUM# = 0!
5005  INPUT "Would you like the data for the Whittaker plot to go to a file
(F), the printer (P), or the screen only (S)"; ANS7$
      IF ANS7$ = "p" OR ANS7$ = "P" THEN
         FILX$ = "LPT1:"
        ELSE
         IF ANS7$ = "S" OR ANS7$ = "s" THEN
             FILX$ = "SCRN:"
            ELSE
             IF ANS7$ = "F" OR ANS7$ = "f" THEN
                INPUT "Enter a file name:", FILX$
              ELSE
                GOTO 5005
             END IF
         END IF
      END IF
      OPEN "O", 3, FILX$
      PRINT "       DATA FOR THEORETICAL WHITTAKER PLOT OF SPECIES ABUNDANCES"
      PRINT #3, "       DATA FOR THEORETICAL WHITTAKER PLOT OF SPECIES
ABUNDANCES"
      PRINT : PRINT
      KK% = MXCL(ICOUNT)
      XN# = TOTN(ICOUNT)
      IF (XN# < 500!) THEN XUMP# = 1!
      IF (XN# >= 500 AND XN# < 1000!) THEN XUMP# = 2!
      IF (XN# >= 1000! AND XN# < 2000) THEN XUMP# = 5!
      IF (XN# > 2000!) THEN XUMP# = 10!
      Y9# = LOG(1! + (ALPHA(ICOUNT) / XN#))
      FOR I% = 1 TO KK%
        XX% = I%
        X3# = XN# / (XX% + 1!)
        X3# = INT(X3#)
```

```
        N8% = 0
        I45% = 1
        I47% = 0
5043    Y# = X3# * Y9#
        IF (I47% = I45%) GOTO 5049
        IF (X3# < 300!) THEN XUMP# = .1
        IF (X3# < 500 AND X3# >= 300!) THEN XUMP# = 1!
        IF (X3# >= 500 AND X3# < 1000!) THEN XUMP# = 2!
        IF (X3# >= 1000! AND X3# < 2000) THEN XUMP# = 5!
        IF (X3# > 2000!) THEN XUMP# = 10!
  statement to carry proper argument to subroutine
        XSUB# = Y#
        GOSUB 5100:  'SUBroutine SEI#
        Z2# = ALPHA(ICOUNT) * SEI#
        N8% = N8% + 1
        DD = Z2# - XX%
        IF (N8% > 3000) GOTO 5049
        IF (DD > .001) GOTO 5045
        IF (DD < -.001) GOTO 5047
        GOTO 5049
5045    X3# = X3# + XUMP#
        I45% = 99
        GOTO 5043
5047    IF (X3# - XUMP# > 0!) GOTO 5048
        GOTO 5049
5048    I47% = 99
        X3# = X3# - XUMP#
        GOTO 5043
5049    PRINT I%;
        X3# = X3# - XUMP#
        WH(I%) = X3#
5050 NEXT I%
        PRINT
      PRINT " SPECIES RANK    NUMBER OF INDIVIDUALS       PROPORTIONS
DATA"
      PRINT #3, "  SPECIES RANK    NUMBER OF INDIVIDUALS       PROPORTIONS
DATA"
      SUM# = 0!
   FOR I% = 1 TO KK%
5060    SUM# = SUM# + WH(I%)
   NEXT I%
      CORR# = XN# / SUM#
      ICONT% = 0
   FOR I% = 1 TO KK%
      ICONT% = ICONT% + 1
      WH(I%) = WH(I%) * CORR#
      Z6# = WH(I%) / XN#
      PRINT TAB(9); I%; TAB(24);
      PRINT USING "#####.#"; WH(I%);
      PRINT TAB(48);
      PRINT USING "#.###^^^^"; Z6#;
      PRINT TAB(63); CLASS(ICOUNT, 1, KK% - (I% - 1))
      PRINT #3, TAB(9); I%; TAB(24);
      PRINT #3, USING "#####.#"; WH(I%);
```

```
        PRINT #3, TAB(48);
        PRINT #3, USING "#.###^^^^"; Z6#;
        PRINT #3, TAB(63); CLASS(ICOUNT, 1, KK% - (I% - 1))
        IF (ICONT% >= 24) THEN
          PRINT "....Hit < ENTER > to continue ......"
          WHILE INKEY$ = "": WEND
        END IF
        IF (ICONT% >= 24) THEN ICONT% = 0
 5065 NEXT I%
          PRINT
          CLOSE 3
        RETURN
'
'        DOUBLE PRECISION FUNCTION SEI(XSUB)
'       STANDARD EXPONENTIAL INTEGRAL FUNCTION
'       APPROXIMATION FORMULAE FROM ABRAMOWITZ AND STEGUN (1964) CHAP. 5
 5100   A# = -.57721566#: A1# = .99999193#: A2# = -.24991055#: A3# = .05519968#
        A4# = -9.760039999999999D-03: A5# = .00107857#
        AA# = 2.334733: AB# = .250621: B1# = 3.330657: B2# = 1.681534
        ZSUB# = XSUB#
        SEI# = 0!
        IF (ZSUB# > 0! AND ZSUB# <= 1!) GOTO 5110
        IF (ZSUB# > 1!) GOTO 5120
        GOTO 5150
 5110 SEI# = -LOG(ZSUB#) + A# + (A1# * ZSUB#) + (A2# * ZSUB# ^ 2) + (A3# *
 ZSUB# ^ 3) + (A4# * ZSUB# ^ 4) + (A5# * ZSUB# ^ 5)
        RETURN
 5120 SEI# = (1! / (ZSUB# * EXP(ZSUB#))) * ((ZSUB# ^ 2) + (AA# * ZSUB#) + AB#)
 / ((ZSUB# ^ 2) + (B1# * ZSUB#) + B2#)
        RETURN
 5150 RETURN
'
 8000 ' subroutine to allow keyboard entry of data for a single station
 PRINT
 ' Initialize variables
 ST = 1
 DIM STX(ST), TOTN(ST), TOTS(ST), ALPHA(ST), VAR(ST)
 DIM KSTAT(ST), MXCL(ST), KSIG$(ST)
'
 PRINT TAB(10); "*****************************************"
 PRINT TAB(10); "*"; TAB(50); "*"
 PRINT TAB(10); "*"; TAB(21); "1) Species Abundance"; TAB(50); "*"
 PRINT TAB(10); "*"; TAB(21); "   Data"; TAB(50); "*"
 PRINT TAB(10); "*"; TAB(50); "*"
 PRINT TAB(10); "*"; TAB(21); "2) # of Species & "; TAB(50); "*"
 PRINT TAB(10); "*"; TAB(21); "   # of Individuals only "; TAB(50); "*"
 PRINT TAB(10); "*"; TAB(50); "*"
 PRINT TAB(10); "*****************************************"
 PRINT
 INPUT "Please enter your choice"; ANS4%
 ON ANS4% GOTO 8010, 8020
'
 8010 INPUT "Enter the number of species"; S
 DIM CLASS(ST, 2, S)
```

```
    MAXCLASS = 1:  ' MAX. # of abund. classes found for this station so far
    STX(1) = 1: 'SET the station #
    FOR J = 1 TO S
         PRINT "Enter the number of individuals of species"; J; ":",
       INPUT NUM:   ' Read abundance value for species J
       TOTN(1) = TOTN(1) + NUM
       IF NUM > 0 THEN
          TOTS(1) = TOTS(1) + 1
          FOR K = 1 TO MAXCLASS
             IF NUM = CLASS(1, 1, K) THEN
                CLASS(1, 2, K) = CLASS(1, 2, K) + 1: 'Count # of spp in this
abund. class
                GOTO 8111
             END IF
          NEXT K
          CLASS(1, 1, MAXCLASS) = NUM: ' Make new abundance class
          CLASS(1, 2, MAXCLASS) = 1
          MAXCLASS = MAXCLASS + 1: ' Increment max. number of abund. classes
       END IF
8111 NEXT J
   MXCL(1) = MAXCLASS - 1: ' # of abund. classes for this station.
   GOSUB 1800:  ' Sort abundance classes
RETURN

8020 INPUT "Enter the number of species:", S
        TOTS(1) = S
        MXCL(1) = S:  ' SET max number of classes to the number of species
                      ' - this might not be correct.
     INPUT "Enter the total number of individuals:", TOTN(1)
     ICOUNT = 1

GOSUB 2000:  ' Calc.   and X
   Prompt with request for Whittaker Plot data (Krebs 1987, ch. 10)
   INPUT "Do you want the predicted abundances for a Whittaker plot"; ANS5$
   IF ANS5$ = "Y" OR ANS5$ = "y" THEN GOSUB 5000
   STOP
RETURN
```

Fig. 1

Example of multiple site data set.
Abundance of demersal species in the Samar Sea, Philippines
Cruise 98, stations 1 & 5

```
"Samar Sea cruise 98, species abundance by numbers - stations 1 & 5"
2,77
STATION,S1,S2,S3,S4,S5,S6,S7,S8,S9,S10,S11,S12,S13,S14,S15,S16,S17,S18,S19,S20,S21
S22,S23,S24,S25,S26,S27,S28,S29,S30,S31,S32,S33,S34,S35,S36,S37,S38,S39,S40
S41,S42,S43,S44,S45,S46,S47,S48,S49,S50,S51,S52,S53,S54,S55,S56,S57,S58,S59,S60
S61,S62,S63,S64,S65,S66,S67,S68,S69,S70,S71,S72,S73,S74,S75,S76,S77
1,2138,1921,1743,1733, 367, 440, 314, 262, 256, 225, 220, 183, 175, 141, 110, 103
  67,  47,  42,  34,  31,  26,  21,  21,  16,  10,   7,   6,   5,   5,   3,   5
   3,   3,   3,   2,   1,   1,   1,   1,   0,   0,   0,   0,   0,   0,   0,   0
   0,   0,   0,   0,   0,   0,   0,   0,   0,   0,   0,   0,   0,   0,   0,   0
   0,   0,   0,   0,   0,   0,   0,   0,   0,   0,   0,   0,   0
5,1611, 678, 293, 137, 118,  92,  91,  55,  45,  37,  19,  16,  13,  13,  11,  11
  11,   9,   7,   6,   5,   5,   5,   5,   5,   5,   5,   3,   3,   3,   2,   2
   1,   1,   1,   1,   1,   1,   0,   0,   0,   0,   0,   0,   0,   0,   0,   0
   0,   0,   0,   0,   0,   0,   0,   0,   0,   0,   0,   0,   0,   0,   0,   0
   0,   0,   0,   0,   0,   0,   0,   0,   0,   0,   0,   0,   0,   0
```

Fig. 2

First output screen

```
 ***** α *****
STATION #                    : 1
Number of Individuals (N)    : 11196
Number of species (S)        : 40

  α                          : 5.213393
Variance of α                : .6794909
  X                          : .9995346
Standard Error of N          : 105.8112
Standard Error of S          : 1.900002
Standard Error of α          : .2847294
```

Fig. 3

DATA FOR THEORETICAL WHITTAKER PLOT OF SPECIES ABUNDANCES

SPECIES RANK	NUMBER OF INDIVIDUALS	PROPORTIONS	DATA
1	2999.541	.267911844911129	2138
2	1904.799	.170132107204861 1	1921
3	1350.958	.120664345106511 3	1745
4	1010.63	9.026710371240286D-02	1733
5	773.8246	6.911617059629276D-02	867
6	606.8959	5.420649107033204D-02	440
7	480.0813	4.267976492895063D-02	314
8	384.4537	3.433848466563114D-02	262
9	309.2711	2.762335272676905D-02	256
10	250.2637	2.235295390094677D-02	225
11	203.1613	.0181458838797757	220
12	165.5053	1.478253794210474D-02	183
13	135.2252	1.207799253910088D-02	175
14	110.6383	9.881994391220213D-03	141
15	90.71086	8.102078998289009D-03	110
16	74.40621	6.645785267330060D-03	103
17	61.07779	5.455828715374945D-03	67
18	50.20801	4.484459773776104D-03	47
19	41.27927	3.686965553333384D-03	43
20	34.03275	3.039723934979045D-03	34
21	27.95085	2.496503344087882D-03	31
22	23.03357	2.057306441937969D-03	26
23	20.18672	1.803029998-73619[illegible]D-03	21
24	17.72808	1.583430204020097D-03	16
25	15.58825	1.386946123447865D-03	10
26	13.45781	1.202019847857607D-03	7
27	11.51678	1.028651550609157D-03	6
28	9.834355	8.783989535045741D-04	5
29	6.158323	7.251464395779054D-04	3
30	6.599502	5.874517762271706D-04	2
31	5.044677	4.507571128744382D-04	1

Fig. 4

Second output screen

```
***** K *****
Total # of individuals      : 11196
Total # of species          : 40
  α                         : 5.213393

Observed K score            : 161.8404
Expected K score            : 153.4509
Difference ± std. Err.      : 8.389435 ± 12.73978
K is sufficiently close to zero!
```

Fig. 5

Summary output

06-26-1989 13:49:29

LOGSERIES fit of demersal species abundance from cruise 98 in the Samar Sea

STATION	N	S	α	K	
1	11196	40	5.2134	0.0517	*

CLASS :	1	2	3	5	6	7	10	16	21	26	31	34	42	47	67	103	110	141	175	183	220	225	256	262	314	440	867	1733	1745	1921	2138
FREQUENCY :	4	1	3	4	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

STATION	N	S	α	K
5	3327	38	6.0152	0.1359

CLASS :	1	2	3	5	6	7	9	11	13	16	19	37	45	55	91	92	118	137	293	678	1611
FREQUENCY :	6	2	3	7	1	1	1	3	2	1	1	1	1	1	1	1	1	1	1	1	1

* indicates a significantly good fit to the logseries model