International Council for
the Exploration of the Sea

C.M. 1991/D:9
Statistics Committee

Analysis of groundfish survey data: Combining the GLM and delta approaches.

by
Gunnar Stefánsson
Marine Research Institute
P.O. Box 1390
Skúlagata 4
121 Reykjavík
Iceland

ABSTRACT

This paper describes a new method for the analysis of groundfish survey data by explicitly incorporating zero and nonzero values into a single model. This is done by using a model which modifies the delta-distribution approach to fit into the GLM framework and uses maximum likelihood to estimate parameters. No assumptions of homogeneity are used for the structure of the zero or nonzero values. The method is primarily applicable to fixed-station designs, although extensions to other designs are possible. The maximum likelihood estimation reduces to fitting a GLM to 0/1 values and another GLM to the positive abundance values. The new model is tested on the Icelandic groundfish survey data and results from different models are compared on the basis of tuned VPA runs.

1. Introduction

Several entirely different approaches have existed for the analysis of groundfish survey data. These can be classified according to assumptions on spatial distribution of the species and according to assumptions on the probability distribution of the measurements.

Most methods assume a homogeneous population, at least within some strata. Thus, within each stratum the assumption is that all the measurements are of the same average population. When stations are randomized every year, this assumption is true to some extent, although it usually wastes information and does in no way acknowledge the fact that there is always an underlying spatial pattern to the fish density, often with some year-to-year consistency.

The analysis then boils down to evaluating an average within each stratum and integrating these averages to obtain a stock index for the whole region.

Probably the most common single method for the analysis is the stratified analysis of Cochran (1977). Alternatives include the so-called delta-distribution (Aitchison, 1955, Pennington, 1983), where the zero values are treated separately and the positive values are assumed to follow a lognormal distribution. As before, no spatial pattern is allowed within the strata. This delta-distribution would be better named the delta-lognormal distribution, as it is perfectly feasible to use a similar delta-gamma distribution (Steinarsson and Stefánsson, 1986).

Entirely different approaches have also been tried, including the use of log-linear models and kriging. In both cases, the underlying spatial distribution is explicitly modelled. However, both of these methods have some problems with zero values. In particular, when the data from each tow is split into age groups, a large number of zero values occur. Many of these may exist simply because the tows occur far away from the potential location of this particular age group. Other zero values may be important indicators of a small stock size. Thus, one should consider models where these two types of zero values automatically influence the biomass indices in the right ways. It is usually not possible to limit exactly the area of interest and this may have severe effects on the stock estimates for some procedures of analysis. The fact that the data are best analyzed in an age-disaggregated fashion compounds problems inherent in log-transforms (Myers and Pepin, 1987), since disaggregation will likely lead to many low abundance values, if there are several age groups in the stock of interest.

The current approach is to use maximum likelihood where an explicit formula is written down for the probability distribution of catch at each station. This distribution can incorporate all the considerations mentioned above.

Data on cod from the annual Icelandic groundfish survey will be used as an example throughout this paper.

2. The distribution

Typical histograms of age-disaggregated catches per towing mile from a trawl survey are given in figs. 1-3. There is a large number of zero values, and a heavy tail, so for clarity numbers per tow are presented on a log-scale. In fig. 2, the zero values seem to be an extension of the distribution. For some stocks and transforms the zero values may stand out as a separate peak (fig. 3). For this example, most of the zero values occur in a part of the ocean where this age group has almost never been seen.

When zero values are eliminated, the data are close to lognormal (fig 1), indicating that something like a lognormal or gamma density may be appropriate for the positive values (or possibly a negative binomial for the entire data set). It should be noted that the distribution of positive values is almost always found to be heavy-tailed - it is in no way a unique feature of this particular plot. The fact that the values per tow can also be very small is due solely to the use of age-length keys, since otherwise the smallest number is one fish per tow, which for this survey reduces to about 0.25 fish per towing mile. Plots for age groups 2 and 3 show a very similar behaviour and are therefore omitted.

When a small yearclass appears, its distribution may change from the average in a number of ways. The density may stay constant at many points, but the extent of the spatial distribution may diminish. This density change would result in the positive part of the histogram having the same mean, but the number of zero values would increase. In the exact opposite case the spatial distribution stays the same but the density goes down at each point, though never to zero. These different types of changes have been investigated e.g. by Myers and Stokes (1989).

To model this, the number of fish caught at a station may be taken to follow a distribution with a discrete probability of zero and some density for positive values. Thus the c.d.f. becomes:

$$(1) \quad P[Y_{st} \leq \omega] = (1-p_{st}) + p_{st}F_{st}(\omega)$$

where F_{st} is a continuous c.d.f., typically describing the distribution of positive values.

When p_{st} is taken to be a constant within a stratum and F_{st} is a constant lognormal distribution within the stratum, this is the usual delta-lognormal method. If p_{st} is taken as the constant one, the data are discrete and F_{st} is the negative binomial, we obtain another well known approach. If zero values are thrown out, p_{st} is set to 1 and F_{st} is taken to be a gamma density with a parametrized mean, this reduces to a generalized linear model.

From now on a gamma density will be assumed for the positive values. The usual formula,

$$(2) \quad \frac{y^{\alpha} e^{y/\beta}}{\beta^{\alpha}\,\Gamma(\alpha)}$$

for the gamma density will be reparametrized by α and the mean, $\mu = \alpha\beta$.

The likelihood corresponding to the above c.d.f. is given by:

$$(3) \quad L = \prod_{s,t:y_{st}=0} (1-p_{st}) \prod_{s,t:y_{st}>0} p_{st} \frac{y^{\alpha} e^{y\alpha/\mu}}{\left(\frac{\mu}{\alpha}\right)^{\alpha} \Gamma(\alpha)}$$

By denoting by n_{st} the number of repetitions of station s in year t and r_{st} the number of positive values at this station, the above likelihood can be written as:

$$(4) \quad L = \prod_{s,t} (1-p_{st})^{n_{st}-r_{st}} p_{st}^{r_{st}} \prod_{s,t:y_{st}>0} \frac{y_{st}^{\alpha} e^{y_{st}\alpha/\mu_{st}}}{\left(\frac{\mu_{st}}{\alpha}\right)^{\alpha} \Gamma(\alpha)}$$

Note that in almost all surveys, $n_{st} = 1$ and r_{st} is either 1 or 0. When repeated tows are performed at each station, the above formula must be understood to incorporate each station only once in the left part, but every positive number occurs once in the second product.

In the above formula there are two distinct components, the probability of a nonzero value and the distribution of the nonzero values. These can therefore be modelled and fitted separately.

The fitted (unconditional) mean value at each station is given by

(5) $\hat{p}_{st}\,\hat{\mu}_{st}$

It should be noted that a model very similar to this one has been used for meteorological applications (Coe and Stern, 1982) and for cod stomach content data (Waiwood et al, 1991).

3. Models for the mean of positive values

A natural model for the positive values is a gamma density where the mean at each station is modelled to include various effects such as an indicator for the general area, the wind speed, depth, time of day, etc. The effects will be taken as multiplicative, as that would seem to be the most natural model for this data.

It has been found, however, that it is often necessary to include quite fine details of the area. Thus when fitting a model with large regions it is found that a further subdivision is needed, sometimes all the way down to the station level. This detail is of course not possible when stations are randomized every year, but it is quite natural when fixed stations have been used for a number of years (it is also possible with partial replacement).

Thus a reasonable model incorporating fine spatial details and the annual abundance becomes:

(6) $\mu_{st} = e^{\eta_t + \gamma_s}$

Using this model alone (for all the data points) causes several problems. Firstly, it is impossible to fit the gamma density using zero values, as the likelihood becomes zero. Secondly, if zero values are omitted, there is a total lack of information of whether the distribution shrinks when a small yearclass is observed. Finally, if a log-transform such as $log(y+c)$ or a negative binomial is used, it is quite hard to determine which zeroes should be included and which should not. Incorrect inclusion may affect results considerably. Incorrect choice of c may also create havoc, which will be noticed in strange residual plots.

It follows from the above comments that explicit models for zero values is at least of potentially great benefit. It must be noted that for stations that are always zero, there is no information to estimate γ_s when this model is used. This will cause no problems in practice, since the estimated effect at that station will be zero as the value of p_{st} must become zero.

4. Models for the probabilities

The usual model for a probability is the logit-link model, and following the arguments in the preceding section, a basic model can be written as:

$$(7) \quad p_{st} = \frac{1}{1+e^{-(\varsigma_t+\xi_s)}}$$

There may be stations with no fish in any year and stations where fish appear in all years. In this instance the usual maximum likelihood procedure (which needs the constraint $0<p_{st}<1$) must be replaced by defining the MLE, $\hat{\xi}_s$, as the value giving

$$(8) \quad \lim_{\xi_s \to \hat{\xi}_s} \ln L(\tilde{\xi}, \tilde{\varsigma}) = \sup_{\xi_s} \ln L(\tilde{\xi}, \tilde{\varsigma}).$$

where $\tilde{\xi} = (\xi_1, \xi_2, ...)$ and $\tilde{\varsigma} = (\varsigma_1, \varsigma_2, ...)$.

It is obvious that the limit is not attained but it does correspond to $\xi_s \to -\infty$ for all-zero stations and $\xi_s \to \infty$ for all-positive stations. This causes no problems as it simply means that $\hat{p}_{st} \to 0$ or 1 for these stations, and this is numerically implemented by setting ξ_s to a very small or large value.

Hence for a station that is always zero, the fitted value at the station will become $E[\hat{Y}_{st}] = \hat{p}_{st}\hat{\mu}_{st} = 0$.

In the actual estimation procedure, groups of stations corresponding to factor levels which give zeroes (or ones) for all years are omitted from the fitting procedure. After fitting is completed, computations are performed by setting the corresponding p_{st} to 0 or 1 for all years. This is intuitively reasonable and is justified by the above redefinition.

5. Analysis using the full model

The full model is most naturally fitted in two phases using GLIM, as this program will fit each distribution easily. Computer time is the only limitation, and this is of course a trivial cost compared to the cost of the survey. For testing this model, cod data in the Icelandic groundfish survey is used. This survey covers over 500 stations each year and has been in operation since 1985. Using only the station and year effects, the models therefore include over 500 parameters and some thousands of observations.

The basic survey data is age-disaggregated along the lines described in Pálsson and Stefánsson (1991), yielding the numbers per age group per station per year, along with various items recorded at each station.

The actual model fitted included the above-mentioned station- and year effect, as well as factors describing the diurnal variation, the wind direction, wind speed and depth (which may vary in spite of the fixed station design, due to minor deviations in location).

The estimate of the "surface", $\hat{p}_{st}\hat{\mu}_{st}$ is quite reasonable, regardless of which collections of zero values are included in the data. Thus, even if large regions of zero values are included, these simply enforce $\hat{p}_{st} = 0$ over that region and have no effect at all outside it if a station effect is used.

Variance estimates are affected by an incorrect choice of region, however, since with an irrelevant region of zero values, there will seem to be too many degrees of freedom in the data, with all the excess data giving a perfect fit.

Having obtained the full surface, it remains to compute annual indices of abundance. For this both the zero and nonzero values have to be used, since each set of data contains its own piece of information. The response at each station is estimated with

$$(9)\quad p_{st}\mu_{st} = \frac{e^{\eta_t+\gamma_s}}{1+e^{-(\varsigma_t+\xi_s)}}$$

and the ratios of these values between two years will vary from one station to the next. This introduces a new twist to the usual GLM models of abundance, since usually it is sufficient to simply read off the year effects or at most compute the predicted value at a single station. The exception is when there is an interaction with year, in which case integration over the regions has to be used. In the present setting it is also the case that some form of integration over the whole region has to be used.

Integration here is simply performed by computing an average based on weighting the fitted value at each station with the inverse number of stations in the statistical square.

6. Comparison of alternative indices

Several indices have been computed for the Icelandic groundfish survey in previous years. These include a stratified mean, SM, as described in Pálsson et al (1989), a geometric mean (Pálsson, 1984) and indices based on multiplicative models (MI), based on log-transforming before fitting an ANCOVA model, as in Myers and Pepin, 1986 and Stefánsson, 1988. The current model yields several indices, namely the year effects τ_t and ξ_t along with the integrated index. The index used by Pálsson (GMR) is based only on a subarea (the juvenile ground. Finally, there are two further trivial indices to be considered, the geometric mean of all observations and the simple arithmetic mean over all stations.

The basic indices are given in in tables 1-7 for age groups 1-7. The reduced-data geometric mean index (GMR) was only available for ages 1-4, as this index has only been used earlier for recruitment estimation. The BI and GI indices (year effects in binomial and gamma-based models, respectively) are presented only for comparative purposes, since they should not be used unless combined through integration. The integrated index from the gamma-binomial model will be denoted GB.

It should be noted that the age-length keys used differ somewhat between the different methods. Since the GMR and SM indices have been in use for some time, they are based on different, fixed areas used for computing age-length keys. The other indices considered (GB, AM, GM and MI) are all based on the same areas. The same linear terms are used in the models for GI, BI (components of GB) and MI.

A basic comparison between the indices consists of simply computing their correlations, for each age group. The correlation matrices between indices are given in tables 8-14. It is obvious that for the most part, the indices are all measuring the same effects - the correlations tend to be quite high. This is to be expected, due to the large number of stations in the survey.

The data series available consists of the years 1985-1991, and since the VPA for this period has not converged, a comparison based on one tuned VPA cannot be used to compare the indices. The approach chosen is to give each index the "benefit of doubt", by fitting the best (logged) VPA to each log-index separately and computing an R^2 value between the logged stock estimate and log-index, based on that fit.

The VPA fitting procedure used is roughly as described in Stefánsson 1988, since that method produces the VPA for each index series which gives the minimum sum of squares, thus potentially giving the highest value of R^2. Indices for ages 1 and 2 are used as measures of the abundance of age group 3, since only ages 3+ appear in the catches.

More specifically, the selection pattern in the terminal year is fixed (iteratively) to be equal to the average of the 3 preceding years and the fishing mortality on the oldest fish (age 14) is fixed to be equal to the average of the fishing mortality for ages 10-12. This setting reduces the minimization to only one variable, the terminal fishing mortality multiplier. This overall mortality is estimated as the value which gives the minimal overall residual sum of squares from the regression lines for each age group of the log-index on logged VPA stock numbers.

A slight complication appears in that a migration of the 1984 yearclass is known to have occurred in 1990 (and probably also in 1991), from Greenland waters into the Icelandic area. This migration is simply assumed to be 28 millions each year and is subtracted in the VPA computations (this number is based on VPA-based least-squares estimates of the migrations and the tagging data available, but its computation is outside the scope of this paper).

The correlations with VPA are given in table 15, based on using age groups 1-4 for fitting, with each of the main indices separately. Table 16 gives the results based on fitting to age groups 1-7.

It is immediately seen that the correlations with VPA are generally quite good and it is not at all trivial to distinguish between the various types of indices. Formal statistical tests are not appropriate since the different R^2-values are correlated.

7. Discussion and conclusions

The approach considered is based on an intuitively appealing model. Results obtained are close to those obtained by other methods, but the ad-hoc nature of most other methods of analysis is eliminated by using an explicit model for zero and non-zero values. The model yields indices which are free of problems usually associated with zero values, such as those involving the definition of an appropriate area for the analysis and those related to log-transforming values which can be arbitrarily close to zero.

Variance estimates for the resulting parameters are available, but should be viewed with caution, since the real variances of interest are those related to prediction capabilities and the degrees of freedom vary depending on the inclusion of zero-catch tows. The actual variances of interest are probably better obtained by tuning VPAs with the indices, as in Anon. (1990).

This model has considerable potential for the analysis of groundfish survey data, since it can incorporate several relevant properties of fish distributions, including changes in density and range. The usual qualities of GLMs, specifically the potential for incorporating effects such as diurnal variations, are also available.

It must be noted, however, that here, as in Anon. (1990), the actual type of analysis considered does not seem to be of great consequence to the predictive power of the numbers obtained, since even simple methods of analysis yield results fairly consistent with VPA results for the data set considered. This conclusion is, however, likely to be dependent on the number of stations in the survey.

REFERENCES

Aitchison, J. 1955. On the distribution of a positive random variable having a discrete probability mass at the origin. J. Amer. Stat. Assoc. 50:901-908.

Anon. 1990. Report of the Working Group on Methods of Fish Stock Assessments. ICES C.M. 1990/Assess: 15. 95pp mimeo.

Cochran, W.G. 1977. Sampling techniques. 3rd ed. John Wiley & Sons, Inc., New York 330 p.

Coe, R. And Stern, R. D. 1982. Fitting models to daily rainfall data. J. Appl. Meteor. 21:1024-1031.

Myers, R. A. and Pepin, P. 1986 The estimation of population size from research surveys using regression models. ICES CM 1986/D:9

Myers, R. A., and Pepin, P. 1987. The robustness of lognormal based estimators of abundance. ICES C.M. 1987/D:2

Myers, R. A. and Stokes, K. 1989. Density-dependent habitat utilization of groundfish and the improvement of research surveys. ICES C.M. 1989/D:15.

Pálsson, Ó. K. 1984. Studies on recruitment of cod and haddock in Icelandic waters. ICES C. M. 1984/D:6

Pálsson, Ó. K., Jónsson, E., Schopka, S. A., Stefánsson, G. and Steinarsson, B. Æ. 1989. Icelandic groundfish survey data used to improve precision in stock assessments. J. Northw. Atl. Fish. Sci. 9:53-72.

Pálsson, Ó. K. and Stefánsson, G. 1991. Spatial distribution of Icelandic cod in March 1985-1991. ICES C.M. 1991/G:63

Pennington, M. 1983. Efficient estimators of abundance, for fish and plankton surveys. Biometrics, 39: 281-286.

Stefánsson, G. 1988. A statistical analysis of Icelandic trawler reports, 1973-1987. ICES C.M. 1988/D:13.

Steinarsson, B. Æ. and Stefánsson, G. 1986. Comparison of random and fixed trawl stations in Icelandic groundfish surveys and some computational considerations. ICES C.M. 1986/D:13.

Waiwood, K. G., Smith, S. J. and Petersen, M. R. 1991. Feeding of cod (Gadus Morhua) at low temperatures. Can. J. Fish. Aquat. Sci. 48:824-831.

Table 1. Indices for 1-group cod.

Year	AM	GM	MI	GI	BI	GB	SM	GMR
----	--	--	--	--	--	--	--	--
85	3.3119	1.3335	1	0.1919	0.1647	451.8	19.5	1.8
86	3.8045	1.6639	1.1873	0.2429	0.2399	635.3	17.2	1.6
87	1.1470	0.5873	0.6206	0.0537	0.0761	96.2	3.92	.5
88	0.9828	0.4258	0.6046	0.0545	0.0505	83.0	3.5	.4
89	1.3984	0.7202	0.6839	0.0585	0.1387	130.2	4.8	.7
90	1.8112	0.8038	0.7877	0.0997	0.1355	219.4	7.0	.9
91	0.6859	0.3781	0.5991	0.0848	0.0255	89.9	4.9	.6

AM=Aritmetic mean of all observations
GM=Geometric mean of all observations
MI=Multiplicative model index
GI=Gamma model year effect
BI=Bernoulli model year effect
GB=Gamma+Bernoulli model integrated index
SM=Stratified mean
GMR=Geometric mean of reduced set of observations

Table 2. Indices for 2-group cod.

Year	AM	GM	MI	GI	BI	GB	SM	GMR
--	--	--	--	--	--	--	----	---
85	22.423	3.2531	1	0.3403	0.0351	396.6	111.7	4.6
86	12.687	3.7730	1.1535	0.5138	0.1582	822.8	62.7	5.3
87	6.664	2.5497	0.7740	0.2212	0.0501	273.0	30.5	3.1
88	1.769	0.8888	0.4010	0.0732	0.0017	37.1	7.7	1.1
89	4.911	1.7930	0.6642	0.2181	0.0051	165.0	19.7	2.7
90	3.131	1.1974	0.5583	0.1317	0.0039	91.4	15.9	1.7
91	4.771	1.9388	0.7207	0.2341	0.0132	221.6	22.4	2.3

AM=Aritmetic mean of all observations
GM=Geometric mean of all observations
MI=Multiplicative model index
GI=Gamma model year effect
BI=Bernoulli model year effect
GB=Gamma+Bernoulli model integrated index
SM=Stratified mean
GMR=Geometric mean of reduced set of observations

Table 3. Indices for 3-group cod.

Year	AM	GM	MI	GI	BI	GB	SM	GMR
----	--	--	--	--	--	----	----	----
85	7.887	2.429	1	0.7085	0.1049	255.2	43.2	3.8
86	23.594	5.748	1.9804	2.1997	0.8643	3522.2	114.7	9.6
87	22.781	6.744	2.1278	2.5888	0.8854	4241.4	118	11.1
88	20.430	6.080	1.7552	1.5989	0.6897	2240.3	93.2	8.2
89	5.048	1.494	0.7591	0.4677	0.0623	116.3	25.4	2.7
90	5.118	1.504	0.8446	0.7597	0.0769	220.8	29.6	2.6
91	5.119	1.747	0.9034	0.5259	0.3037	396.8	29.1	3.

AM=Aritmetic mean of all observations
GM=Geometric mean of all observations
MI=Multiplicative model index
GI=Gamma model year effect
BI=Bernoulli model year effect
GB=Gamma+Bernoulli model integrated index
SM=Stratified mean
GMR=Geometric mean of reduced set of observations

Table 4. Indices for 4-group cod.

Year	AM	GM	MI	GI	BI	GB	SM	GMR
85	12.179	3.6080	1	0.0719	0.9895	2006.0	70.9	5.0
86	5.501	1.9609	0.6290	0.0340	0.9922	949.8	27.7	2.9
87	20.910	5.2070	1.3600	0.1147	0.9982	3251.1	100.7	7.9
88	30.810	8.1416	1.7419	0.1517	0.9979	4305.0	134.5	9.5
89	15.316	3.1714	0.9407	0.0772	0.9880	2149.2	90.5	5.5
90	2.891	0.9912	0.5037	0.0230	0.9302	596.5	14.4	1.6
91	6.390	2.1227	0.7856	0.0499	0.9833	1387.2	35.9	3.4

AM=Aritmetic mean of all observations
GM=Geometric mean of all observations
MI=Multiplicative model index
GI=Gamma model year effect
BI=Bernoulli model year effect
GB=Gamma+Bernoulli model integrated index
SM=Stratified mean
GMR=Geometric mean of reduced set of observations

Table 5. Indices for 5-group cod.

Year	AM	GM	MI	GI	BI	GB	SM
85	17.983	4.703	1	0.7087	1.0000	2282.7	93.6
86	4.968	2.027	0.5139	0.2524	1.0000	810.5	25.6
87	5.200	1.717	0.4889	0.2050	1.0000	659.9	24.2
88	25.006	6.091	1.1435	0.8606	1.0000	2771.2	98.8
89	14.895	3.476	0.7929	0.6520	1.0000	2098.6	81.3
90	5.337	2.049	0.6004	0.4481	1.0000	1442.9	28.1
91	3.169	1.347	0.4563	0.1989	1.0000	640.4	16.4

AM=Aritmetic mean of all observations
GM=Geometric mean of all observations
MI=Multiplicative model index
GI=Gamma model year effect
BI=Bernoulli model year effect
GB=Gamma+Bernoulli model integrated index
SM=Stratified mean

Table 6. Indices for 6-group cod.

Year	AM	GM	MI	GI	BI	GB	SM
85	5.497	2.009	1	0.5425	1.0000	581.2	30.1
86	7.240	2.877	1.2251	0.7387	1.0000	793.3	34.2
87	2.754	1.283	0.7589	0.3054	1.0000	328.1	14
88	2.070	1.048	0.6647	0.2300	1.0000	247.0	10.4
89	7.614	2.250	1.1190	0.7063	1.0000	759.3	43.2
90	6.390	2.623	1.3036	1.0598	1.0000	1140.0	35.2
91	3.649	1.665	0.9578	0.5102	1.0000	550.0	20.4

AM=Aritmetic mean of all observations
GM=Geometric mean of all observations
MI=Multiplicative model index
GI=Gamma model year effect
BI=Bernoulli model year effect
GB=Gamma+Bernoulli model integrated index
SM=Stratified mean

Table 7. Indices for 7-group cod.

Year	AM	GM	MI	GI	BI	GB	SM
85	3.6515	1.3491	1	0.5548	1.0000	471.7	21.9
86	1.5645	0.8970	0.7769	0.3153	1.0000	268.7	8.8
87	2.7088	1.3528	0.9936	0.5345	1.0000	455.3	15.2
88	1.8700	0.9513	0.8310	0.3406	1.0000	290.4	9.1
89	0.8282	0.4539	0.6348	0.1603	1.0000	136.7	5.3
90	3.0302	1.4137	1.0760	0.7030	1.0000	600.2	16.7
91	3.7801	1.7746	1.2062	0.9115	1.0000	785.6	24.1

AM=Aritmetic mean of all observations
GM=Geometric mean of all observations
MI=Multiplicative model index
GI=Gamma model year effect
BI=Bernoulli model year effect
GB=Gamma+Bernoulli model integrated index
SM=Stratified mean

Table 8. Correlations between indices for age group 1.

AM	1.0000							
GM	0.9931	1.0000						
MI	0.9883	0.9902	1.0000					
GI	0.9527	0.9455	0.9729	1.0000				
BI	0.9202	0.9477	0.9240	0.8169	1.0000			
GB	0.9835	0.9841	0.9968	0.9847	0.8967	1.0000		
SM	0.9544	0.9269	0.9299	0.9506	0.7839	0.9368	1.0000	
GMR	0.9562	0.9347	0.9345	0.9390	0.8229	0.9321	0.9908	1.0000
Variable	AM	GM	MI	GI	BI	GB	SM	GMR

Table 9. Correlations between indices for age group 2.

AM	1.0000							
GM	0.8139	1.0000						
MI	0.8041	0.9884	1.0000					
GI	0.7028	0.9503	0.9747	1.0000				
BI	0.4442	0.8251	0.8247	0.8872	1.0000			
GB	0.6281	0.9237	0.9376	0.9779	0.9613	1.0000		
SM	0.9990	0.8070	0.7996	0.6985	0.4470	0.6283	1.0000	
GMR	0.8385	0.9865	0.9860	0.9601	0.8144	0.9243	0.8310	1.0000
Variable	AM	GM	MI	GI	BI	GB	SM	GMR

Table 10. Correlations between indices for age group 3.

AM	1.0000							
GM	0.9854	1.0000						
MI	0.9905	0.9868	1.0000					
GI	0.9588	0.9488	0.9812	1.0000				
BI	0.9641	0.9556	0.9769	0.9460	1.0000			
GB	0.9597	0.9496	0.9830	0.9925	0.9692	1.0000		
SM	0.9950	0.9818	0.9972	0.9795	0.9665	0.9782	1.0000	
GMR	0.9844	0.9845	0.9969	0.9844	0.9687	0.9878	0.9941	1.0000
Variable	AM	GM	MI	GI	BI	GB	SM	GMR

Table 11. Correlations between indices for age group 4.

AM	1.0000							
GM	0.9829	1.0000						
MI	0.9890	0.9924	1.0000					
GI	0.9958	0.9834	0.9963	1.0000				
BI	0.6420	0.6418	0.6510	0.6508	1.0000			
GB	0.9955	0.9841	0.9967	0.9999	0.6558	1.0000		
SM	0.9780	0.9366	0.9579	0.9752	0.6902	0.9739	1.0000	
GMR	0.9876	0.9685	0.9881	0.9947	0.7115	0.9951	0.9790	1.0000
Variable	AM	GM	MI	GI	BI	GB	SM	GMR

Table 12. Correlations between indices for age group 5.

AM	1.0000						
GM	0.9919	1.0000					
MI	0.9857	0.9927	1.0000				
GI	0.9548	0.9497	0.9719	1.0000			
BI	0.0000	0.0000	0.0000	0.0000	1.0000		
GB	0.9548	0.9497	0.9720	1.0000	0.0000	1.0000	
SM	0.9700	0.9523	0.9653	0.9590	0.0000	0.9589	1.0000
Variable	AM	GM	MI	GI	BI	GB	SM

Table 13. Correlations between indices for age group 6.

AM	1.0000						
GM	0.9282	1.0000					
MI	0.9044	0.9717	1.0000				
GI	0.8200	0.8966	0.9676	1.0000			
BI	0.0000	0.0000	0.0000	0.0000	1.0000		
GB	0.8188	0.8956	0.9672	1.0000	0.0000	1.0000	
SM	0.9829	0.8712	0.8832	0.8246	0.0000	0.8236	1.0000
Variable	AM	GM	MI	GI	BI	GB	SM

Table 14. Correlations between indices for age group 7.

AM	1.0000						
GM	0.9515	1.0000					
MI	0.9474	0.9924	1.0000				
GI	0.9133	0.9709	0.9876	1.0000			
BI	0.0000	0.0000	0.0000	0.0000	1.0000		
GB	0.9103	0.9693	0.9862	0.9999	0.0000	1.0000	
SM	0.9867	0.9339	0.9306	0.9149	0.0000	0.9131	1.0000
Variable	AM	GM	MI	GI	BI	GB	SM

Table 15. Squared correlations between tuned VPA results and indices.
Age groups 1-4 used for tuning.

Age	AM	GM	MI	GB	SM	GMR
1	.64	.51	.61	.66	.75	.68
2	.96	.86	.83	.79	.96	.85
3	.82	.77	.84	.89	.86	.82
4	.78	.77	.77	.78	.70	.81
5	.94	.91	.90	.91	.93	
6	.77	.78	.82	.86	.77	
7	.49	.61	.66	.71	.57	
Ave 1-4	.80	.73	.76	.78	.82	.79
Ave 1-7	.77	.74	.78	.80	.79	

Table 16. Squared correlations between tuned VPA results and indices.
Age groups 1-7 used for tuning.

Age	AM	GM	MI	GB	SM
1	.65	.51	.61	.64	.80
2	.95	.86	.83	.80	.95
3	.83	.77	.84	.89	.87
4	.78	.77	.77	.78	.70
5	.95	.91	.90	.91	.93
6	.78	.78	.82	.86	.78
7	.49	.61	.66	.71	.58
Ave 1-4	.80	.73	.76	.78	.83
Ave 1-7	.78	.74	.78	.80	.80

Figure 1. Frequency of log-numbers of 1-group cod.
Zero values omitted

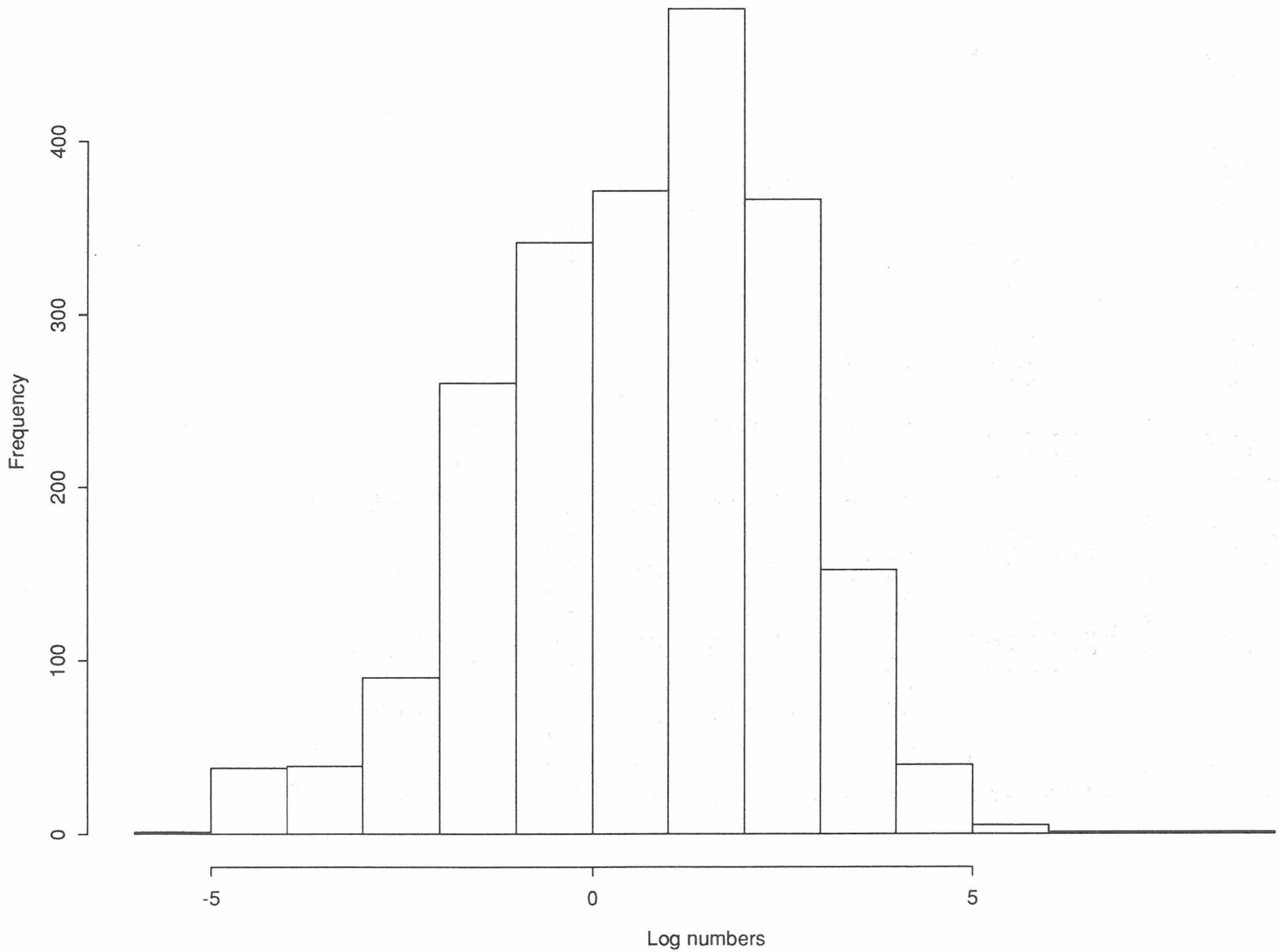

Figure 2. Frequency of log(numbers+1) of 1-group cod.

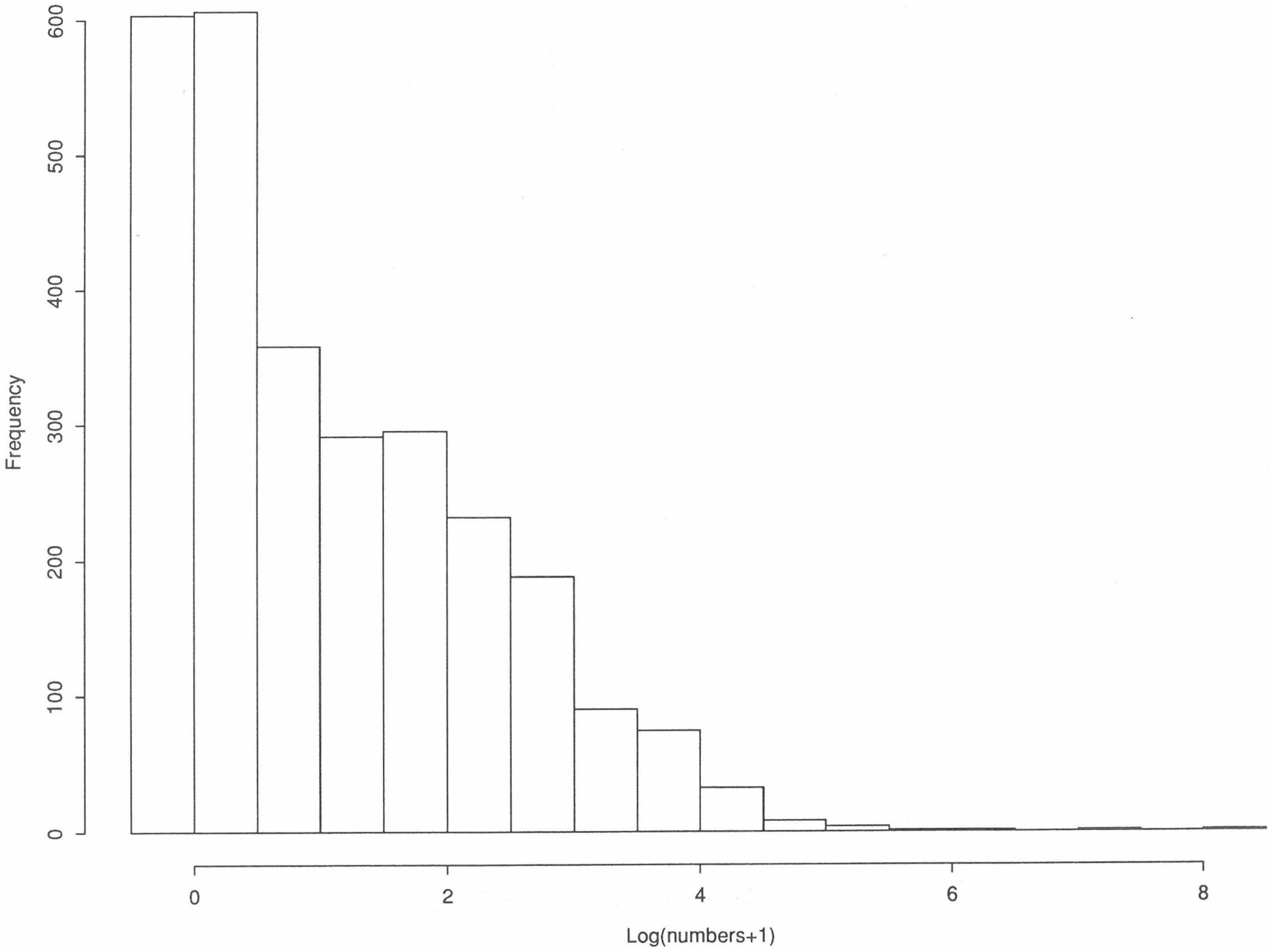

Figure 3. Frequency of log(numbers+0.001) of 1-group cod.