Fol. 41 D

International Counsil for
the Exploration of the Sea

C.M. 1990/D:28 Ref H
Statistics Committee
Ref. Pelagic fish Committee

SCHOOLING-BY-SIZE IN THE BARENTS SEA CAPELIN STOCK

by

Harald Gjøsæter and Knut Korsbrekke[1]

Institute of Marine Research
P.O. Box 1870 Nordnes
N-5024 Bergen
NORWAY

ABSTRACT

This paper analyses the biological samples of capelin taken by the Institute of Marine Research, Bergen, on annual stock assesment surveys in the Barents Sea from 1972 to 1989. The dependency between mean length of 2–year-olds and the mean age in the rest of the sample is demonstrated with the use of linear regression. The data from 1974 and 1978 are further analysed using a stepwise linear regression. The variation in mean length is best explained using the mean length of the capelin the 2–year-olds are schooling with.

[1] Authorship equal

INTRODUCTION

Like most other pelagic fish species, the capelin (*Mallotus villosus*) is sometimes found to form schools. This is particularly significant during migration, e.g. when the mature capelin move towards the coasts to spawn. To a lesser degree, schooling also takes place during non migratory phases of the capelins life cycle. In the Barents Sea, this is observed during September, when the capelin form feeding concentrations in the central and northern parts of the ocean, and the horizontal movements are limited. During this feeding season, the capelin is observed to form discrete schools at intermediate depths during daytime, and to disperse into loosely defined schools or layers during nighttime. At this time, a survey to assess the stock size and its geographical distribution is carried out annually by IMR in Bergen and PINRO in Murmansk. Acoustic methods are used, where both integration of echo signals and trawl sampling are used to calculate the stock size distributed on age and length groups. When analysing the biological data from a number of these surveys, it was found that there seemed to be a dependency between the mean length in each age group and the mean age in each sample, indicating the precence of a phenomenon called "schooling-by-size". This was most noticable with the two-year-olds, where the smallest individuals seemed to school together with the one-year-olds, while the largest individuals were found together with the three years old and older fish. Because of the possible consequences of division of the stock size estimate on age groups, and of an optimal sampling strategy, we decided to analyse this phenomenon more carefully. In this paper we analyse the precence of a dependency between age distribution and size distribution within samples, and try to explain such dependencies by linear regression models.

MATERIALS AND METHODS

The basic data consists of various parameters measured on individual specimens of capelin in trawl samples, including age and length, together with data on the sample itself, viz. geographical position, depth, number of individuals etc. Before these data were used in the analysis, mean values for length within age groups and age and length within samples were calculated. Data from the years 1972 to 1989 were included. All data were sampled at joint Norwegian/USSR surveys to assess the stock size and the geographical distribution of the Barents Sea capelin. These multiship surveys are carried out in September-October each year, and the resulting stock size estimates form the basis for the management advices concerning this stock given by ICES. The 1974 and 1978 data were choosen for a more detailed analysis, because the samples were particularly abundant these years, and a different dependency (slope) between the mean length of 2–year-olds and the mean age in the rest of the sample was demonstrated.

When analysing the dependency between the mean length of the 2–year group and the mean age (excluding 2–year-olds) in a sample, we used an ordinary least square regression. Noting the mean length as L_2 and the mean age as A_o we assume that the regression line is on the form :

$$L_2 = \beta_0 + \beta_1 A_o + \varepsilon$$

The F-ratio were calculated as :

$$F = \frac{MS_{reg}}{s^2}$$

F follows a F-distribution under the null-hypothesis that β_1=0 and that the errors ϵ are independent and normally distributed N(0,σ^2).

The analysis of the 1974 and the 1978 data was done using a stepwise linear regression. In this way we could identify the most interesting variables for further analysis. For both years we used L_2, the mean length of the 2–year-olds as the dependent variable. The independent variables were L_o, the mean length in the sample excluding 2–year-olds, A_o, the mean age in the sample excluding 2–year-olds, N_1, the number of 1–year-olds, N_3, the number of 3–year-olds, N_4, the number of 4–year-olds and N_3+N_4, the number of both 3 and 4–year-olds. All observations were weighted with the square root of N_2. N_2 is the number of 2–year-olds in the sample.

RESULTS

Initially we wanted to investigate if there was any dependency between the mean length of 2–year-olds and the mean age of capelin they were schooling with. The result for our initial regression analyses is shown in table 1. The table contains the estimated intercept (β_0) and slope (β_1). The table also include the test observator F with its degrees of freedom and probability value. The degrees of freedom is 1,n-2 were n is the number of samples (observations) that year. As one can see there is significant dependencies (at the 5% level) between L_2 and A_o all years except 1988 and 1989. (Note the relative low number of samples in 1989). The percentage of variation explained (multiple R^2) varies from 1.56 (1988) to 63.54 (1975).

Our 1974 data consists of 63 observations and our 1978 data of 59 observations. This is after 2 of the 1974 observations and 1 of the 1978 observations were deleted because of missing data. Weighted mean, estimated standard deviation (STD) and coefficient of variation (CV) is given in tables 2 and 3. The weighted correlation matrix is given in tables 4 and 5. Detailed results from each step in the regression analysis is given in tables 6 and 7. A relatively high degree of fit is demonstrated from the fact that the percentage of variation explained (multiple R^2) is as high as 66.56% in the 1974 regression and 71.15% in 1978. The higher degree of fit in 1978 is in accordance with the fact that the F-to-enter (or F-to remove) is higher. This can also be expected from the correlation matrix with a higher correlation between the dependent and the independent variables in 1978.

DISCUSSION

One can see from table 1 that the slope is varying from 0.2103 (1988) to 1.6659 (1972). Very much of this variation can be explained from changes in the age composition. A very strong 2–year old year class will dominate the material and little of the variation may be explained from "who" they are schooling with. Similarly will a relative high abundance of both 1–year and 3–year-olds give the 2–year-olds larger possibilities to "choose" who to school with. Small (short) 2–year-olds tends to school with 1–year-olds while large (long) 2–year old "choose" to school with older capelin.

The more detailed regression analysis of the 1974 and the 1978 data shows us that the mean length of the capelin the 2–year-olds are schooling with, gives a better fit than the mean age. This indicates that capelin form schools with individuals of approximately the same length. The mechanism behind this selection, is probably the need for approximately equal swimming speed among the individuals in a school, and swimming speed is coupled to body length. In addition to the mean length of capelin not in the 2–year-olds group the number of 1–year-olds and 3–year-olds gave additional explanation of the variation in 1974 and 1978 respectively. The explanatory effect was significant, but not very large compared to the mean length.

These findings have obvious effects on survey strategies and on the use of age-length keys. The method used up to now for obtaining a stock size estimate partitioned on age groups, is based on the assumptions that the length and age distribution obtained by sampling within a subarea are representative for that subarea. The results reffered to in this paper may question both.

Firstly, the total length distribution may be biased by chance alone when the number of samples within a subarea becomes very small. This will lead to errors in the total stock size estimate in numbers, since the conversion factor between echo-values and number of fish is length-dependent for capelin.

Secondly, the partition of the total number of fish on age groups may also be wrong, again if the number of samples become small.

However, if the number of samples are large, and if the probability of sampling a school of one particular length-age composition reflects a true proportion of a population, then the problem caused by schooling-by-size may be negligible.

Further work will be undertaken to investigate if some kind of correcting age-length keys based on conditional age or length distributions can diminish this problem even when a small number of samples are taken within one subarea.

TABLE 1.

Year	Intercept	Slope	F	d.f.	P (Tail)
1972	8.10922	1.6659	10.493	1,11	0.0079
1973	10.13791	0.9919	25.559	1,30	0.0000
1974	11.13375	0.3441	5.500	1,63	0.0222
1975	9.54897	1.0689	43.569	1,25	0.0000
1976	11.59829	0.4609	16.663	1,49	0.0002
1977	10.25114	0.8975	85.693	1,57	0.0000
1978	9.89622	0.8761	64.699	1,58	0.0000
1979	10.57554	0.7165	66.589	1,51	0.0000
1980	11.66978	0.6708	58.866	1,68	0.0000
1981	11.72500	0.5275	13.056	1,63	0.0006
1982	11.23060	0.7615	43.917	1,59	0.0000
1983	10.54208	1.3012	60.517	1,75	0.0000
1984	10.44154	1.0943	31.399	1,64	0.0000
1985	11.45771	0.5704	20.740	1,64	0.0000
1986	10.22968	1.4297	33.782	1,31	0.0000
1987	11.57195	1.0394	6.710	1,14	0.0214
1988	13.49643	0.2103	0.587	1,37	0.4483
1989	12.93821	0.8259	2.853	1,16	0.1106

TABLE 2. (1974)

Variable name	Weighted mean	STD	CV
L2	11.9138	1.7678	0.148382
Lo	12.2750	3.5245	0.287124
Ao	2.3403	1.4382	0.614531
N1	9.6436	27.1187	2.812086
N3	16.7536	26.9448	1.608300
N4	0.7125	4.3727	6.137393
N3+N4	17.4660	29.4377	1.685427

TABLE 3. (1978)

Variable name	Weighted mean	STD	CV
L2	11.8884	2.6135	0.219832
Lo	12.8117	5.2714	0.411450
Ao	2.3138	2.1087	0.911358
N1	6.4747	29.2522	4.517892
N3	7.5557	23.8279	3.153643
N4	0.7756	3.5241	4.543744
N3+N4	8.3313	26.7253	3.207836

TABLE 4. (1974)

	L2	Lo	Ao	N1	N3	N4	N3+N4
L2	1.0000						
Lo	0.6573	1.0000					
Ao	0.2631	0.7957	1.0000				
N1	0.0246	-0.5677	-0.8475	1.0000			
N3	0.4788	0.8065	0.6980	-0.4375	1.0000		
N4	0.4684	0.6494	0.3694	-0.1943	0.5154	1.0000	
N3+N4	0.5079	0.8347	0.6938	-0.4293	0.9919	0.6203	1.0000

TABLE 5. (1978)

	L2	Lo	Ao	N1	N3	N4	N3+N4
L2	1.0000						
Lo	0.7880	1.0000					
Ao	0.7304	0.9426	1.0000				
N1	-0.5379	-0.6839	-0.6661	1.0000			
N3	0.7442	0.6562	0.5346	-0.3026	1.0000		
N4	0.6010	0.5677	0.4537	-0.2552	0.7982	1.0000	
N3+N4	0.7427	0.6599	0.5364	-0.3034	0.9968	0.8435	1.0000

TABLE 6. (1974)

STEP NO. 0

ANOVA :	SS	df	MS	
Residual	193.75668	62	3.125108	
Variable	**Coefficient**	**F-to-remove**	**Variable (not in eq.)**	**F-to-enter**
Intercept	11.91383			
			Lo	46.41
			Ao	4.54
			N1	0.04
			N3	18.15
			N4	17.14
			N3+N4	21.20

STEP NO. 1

ANOVA :	SS	df	MS	F Ratio
Regression	83.714729	1	83.71473	46.41
Residual	110.04195	61	1.803966	
Variable	**Coefficient**	**F-to-remove**	**Variable (not in eq.)**	**F-to-enter**
Intercept	7.86681			
Lo	0.32969	46.41	Ao	28.79
			N1	41.89
			N3	0.81
			N4	0.32
			N3+N4	0.59

STEP NO. 2

ANOVA :	SS	df	MS	F Ratio
Regression	128.95567	2	64.47784	59.70
Residual	64.801003	60	1.080017	
Variable	**Coefficient**	**F-to-remove**	**Variable (not in eq.)**	**F-to-enter**
Intercept	5.44610			
Lo	0.49684	119.29	Ao	1.01
N1	0.03826	41.89	N3	2.09
			N4	1.25
			N3+N4	2.75

TABLE 7. (1978)

STEP NO. 0

ANOVA :	SS	df	MS	
Residual	396.15189	58	6.830205	

Variable	Coefficient	F-to-remove	Variable (not in eq.)	F-to-enter
Intercept	11.88844			
			Lo	93.35
			Ao	65.18
			N1	23.20
			N3	70.74
			N4	32.24
			N3+N4	70.14

STEP NO. 1

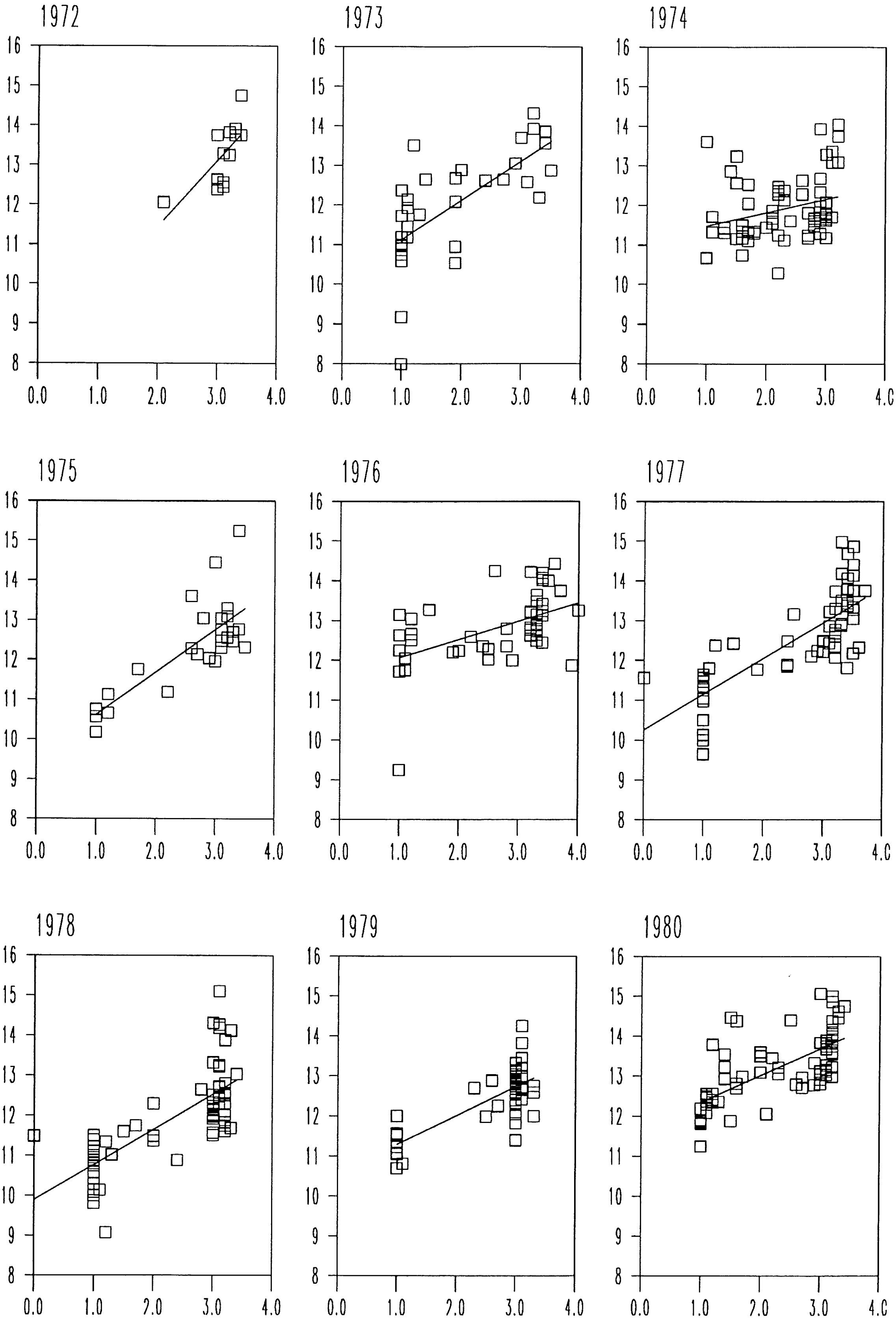
ANOVA :	SS	df	MS	F Ratio
Regression	245.96025	1	245.9603	93.35
Residual	150.19164	57	2.634941	

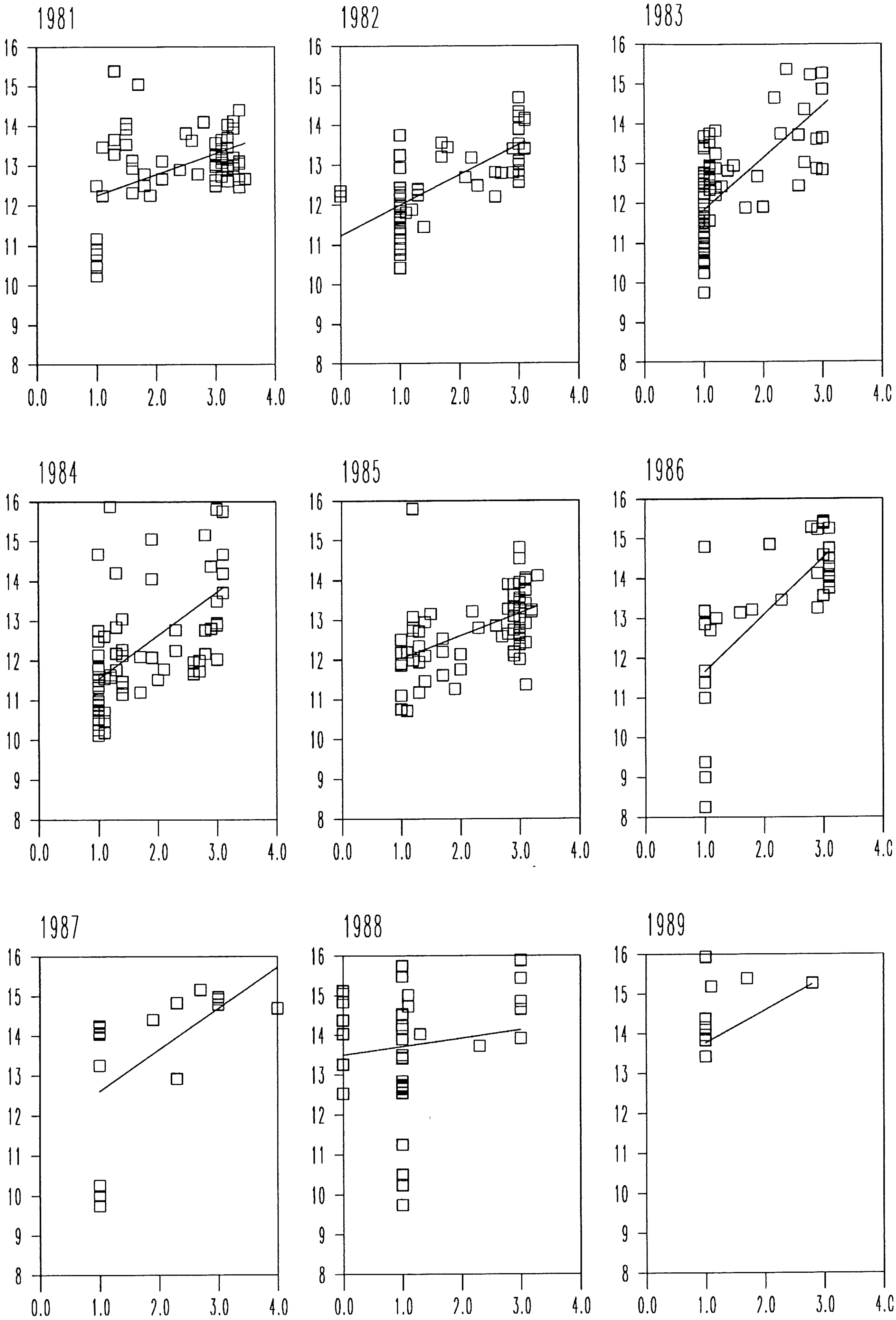
Variable	Coefficient	F-to-remove	Variable (not in eq.)	F-to-enter
Intercept	6.88347			
Lo	0.39066	93.35	Ao	0.20
			N1	0.00
			N3	17.58
			N4	5.67
			N3+N4	16.91

STEP NO. 2

ANOVA :	SS	df	MS	F Ratio
Regression	281.84915	2	140.9246	69.04
Residual	114.30274	56	2.041120	

Variable	Coefficient	F-to-remove	Variable (not in eq.)	F-to-enter
Intercept	8.21540			
Lo	0.26089	30.60	Ao	0.87
N3	0.04375	17.58	N1	1.29
			N4	0.14
			N3+N4	0.14


1972
1973
1974
1975
1976
1977
1978
1979
1980
16 15 14 13 12 11 10 9 8
0.0 1.0 2.0 3.0 4.0


1981
1982
1983
1984
1985
1986
1987
1988
1989