International Council for the Exploration of the Sea

C.M. 1974/F:35
Demersal Fish (Northern) Committee

A NOTE ON A POSSIBLE SOURCE OF BIAS IN THE ESTIMATION OF MEAN LENGTH AT AGE

by

M.D. Nicholson and D.W. Armstrong

Marine Laboratory, Aberdeen, Scotland.

Introduction

When estimating age composition by use of an age-length key the individual length frequency distributions at each age, which are produced as an intermediate step, are often utilised to provide estimates of mean lengths at each age.

If over a particular length range one year-class dominates those adjacent to it, the bulk of the sample for age determination from this length range will be drawn from this year-class. This introduces a possible source of bias to estimates of mean lengths at age, such that mean lengths may be over-estimated for year-classes older than the dominant one, and underestimated for the younger year-classes.

An attempt to assess, using a computer simulation approach, the magnitude of this bias is presented in this paper and is part of a series of studies currently being carried out at the Marine Laboratory, Aberdeen on the properties of fish sampling procedures.

Method

A computer program was written to generate at random a prescribed number of fish lengths (and their ages) from a specified population. These form the primary sample. These lengths are grouped into a length frequency distribution from which, using length intervals as strata, a stratified sub-sample is drawn for age. The level of subsampling is the same at each length interval. If any particular cell frequency is below this level, all available ages in this cell are noted. Continuing as for a normal age-length key, estimates of mean lengths at age are then produced.

This process was repeated 100 times, with primary sample size, number aged per cell and the sampled population held constant. These 100 replicates provide an empirical distribution of the estimates of mean lengths at age, from which the average estimated mean length and standard deviation of estimated mean length at each age can be calculated.

The effect of varying the primary sample size, the number of fish aged per cell or the structure of the sampled population can be assessed by examining the average mean lengths and standard deviations calculated from the resulting empirical distributions.

Analysis and Results

For this study, the population from which samples were drawn is composed of Sprat (Sprattus sprattus (L.)) of three age groups, with lengths at age following a Normal distribution. The population is fully defined by a mean length, standard deviation and percentage composition for each age group. Realistic values of mean lengths and standard deviations were obtained from landings in Scotland during 1970/1971. These are

	Mean length	Standard deviation
Age 1	7.0 cm	0.59
Age 2	10.4 cm	0.95
Age 3	11.4 cm	1.05

Figure 1 shows the distributions of length at age (with equal percentage compositions) using these parameters. As can be seen, the length distribution of the 1 group overlaps very little with that of the 2 group, whereas those of the 2 and 3 groups overlap considerably. By varying the relative sizes of the year classes (i.e. by progressively increasing the percentage composition of the 2 group) it is possible from this simple population to determine any effect which the presence of a dominant year-class has on estimates of mean length at age, and also the extent to which such an effect might be increased when the respective length distributions overlap greatly.

Three sets of values of percentage composition were examined. Denoting the percentage composition of the jth age group by p_j, these are

$p_1 = 33.3\%$ $p_2 = 33.3\%$ $p_3 = 33.3\%$

$p_1 = 25\%$ $p_2 = 50\%$ $p_3 = 25\%$

and $p_1 = 15\%$ $p_2 = 70\%$ $p_3 = 15\%$

The number of length measurements taken in the primary sample is not a factor being studied in this paper, and was kept constant throughout at 500. This number was large enough to ensure an adequate number per cell over most of the length range when fish were grouped by half centimetre intervals, which is standard practice for sprat.

The number subsampled for age per half centimetre length group was varied, taking the values 2, 3, 5, 7 and 10. Combining each of these values with each of the three sets of values of percentage compositions, a total of fifteen computer runs were made, each run producing for each age an empirical distribution of 100 estimated mean lengths. From these distributions, the average and standard deviations of an estimate of mean length were calculated (Tables 1 and 2).

An indication of any bias may be seen in a systematic difference between the estimates of mean length and the population mean length being estimated. Figure 2 shows the average estimates given in Table 1 plotted for each age against the number aged per half centimetre group, together ith the true mean lengths being estimated.

Although the estimates for the 1 and 2 groups do tend to fall below their true mean lengths, the discrepancies are very small. By comparison, the amount by which estimates from the 3 group tend to exceed their true mean is larger. This must be attributed to the extent to which the length distribution of the 3 group is overlapped by the dominant 2 group.

As would be expected, especially for the 3 group, the differences between true and estimated mean lengths increase as the percentage composition of the 2 group increases. They change very little for the 2 group. Only when the number aged per half centimetre becomes small does mean length tend to be underestimated, due to the overlapping 3 group.

The tendency for the mean length of the 3 group to be overestimated increases as the number of fish aged per half centimetre length group decreases; a possible limiting factor to be considered when deciding minimum sampling levels for age determination.

Further insight is obtained from the standard deviations of an estimated mean length. These are plotted in a similar way in Figure 3.

The standard deviations of estimated mean length for the 3 group are in general greater than those observed for the 1 and 2 groups (but equal to those of the 2 group when each has equal percentage composition), and the increase in variability when percentage composition and number aged per half centimetre are decreased is much greater than for similar changes in the 1 group.

With a percentage composition of 15% and ageing 2 fish per half centimetre, the standard deviation of an estimate of mean length for the 3 group is approximately six times that of the 1 group. Again, this must be attributed to the extent to which the length distribution of the 3 group is overlapped by that of the dominant 2 group. Under these conditions the standard deviation of an estimate of mean length for the 3 group is $\pm$ 0.67. For this value there is a 13% chance that an estimate of mean length will be in error by at least one centimetre.

Conclusions

It would obviously be unwise to extend any quantitative result beyond the context of this limited study. Even the extent to which these results might be relevant to more complex situations requires further examination. However, the results of this analysis suggest that estimates of mean lengths at age derived from age-length keys will be slightly biased, that the extent of this bias is increased in the presence of a dominant age-group and further increased when the length distributions of adjacent age-groups overlap. Also, estimates obtained in the presence of a dominant age-group and with greatly overlapping length distributions will be much less accurate.

The implications from this at this stage are mainly precautionary. For example, an observed relationship between year-class density and growth could be spurious, being simply the effect of bias in estimates of mean length. Mainly however, this study indicates that when designing a sampling scheme for some species, such factors as brood fluctuations and growth must be taken into consideration, and when this information is not available, sampling levels should be set as high as possible.

Table 1 Average of estimated mean lengths

	Percentage Composition	Number of fish aged per half centimetre				
		2	3	5	7	10
AGE 1+	33.3	7.0036	6.9934	7.0008	6.9979	6.9919
	25.0	6.9951	6.9935	6.9892	6.9976	6.9966
	15.0	6.9924	6.9810	6.9818	6.9712	6.9864
AGE 2+	33.3	10.3569	10.3761	10.3940	10.3821	10.3855
	50.0	10.3711	10.3915	10.3941	10.3874	10.3788
	70.0	10.3750	10.3966	10.3978	10.3922	10.3940
AGE 3+	33.3	11.4661	11.4326	11.3982	11.4165	11.4261
	25.0	11.5129	11.5073	11.4619	11.4498	11.4534
	15.0	11.6344	11.5658	11.4773	11.4596	11.4360

Table 2 Standard deviations of estimated mean lengths

	Percentage Composition	Number of fish aged per half centimetre				
		2	3	5	7	10
AGE 1+	33.3	0.0574	0.0499	0.0489	0.0436	0.0379
	25.0	0.0740	0.0599	0.0591	0.0487	0.0542
	15.0	0.1123	0.0968	0.0770	0.0757	0.0693
AGE 2+	33.3	0.2171	0.1828	0.1561	0.1324	0.1207
	50.0	0.1729	0.1322	0.1233	0.0972	0.0855
	70.0	0.1135	0.1052	0.0801	0.0682	0.0661
AGE 3+	33.3	0.2395	0.1992	0.1676	0.1408	0.1309
	25.0	0.3388	0.2900	0.2174	0.1941	0.1843
	15.0	0.6659	0.5215	0.3337	0.2647	0.2579

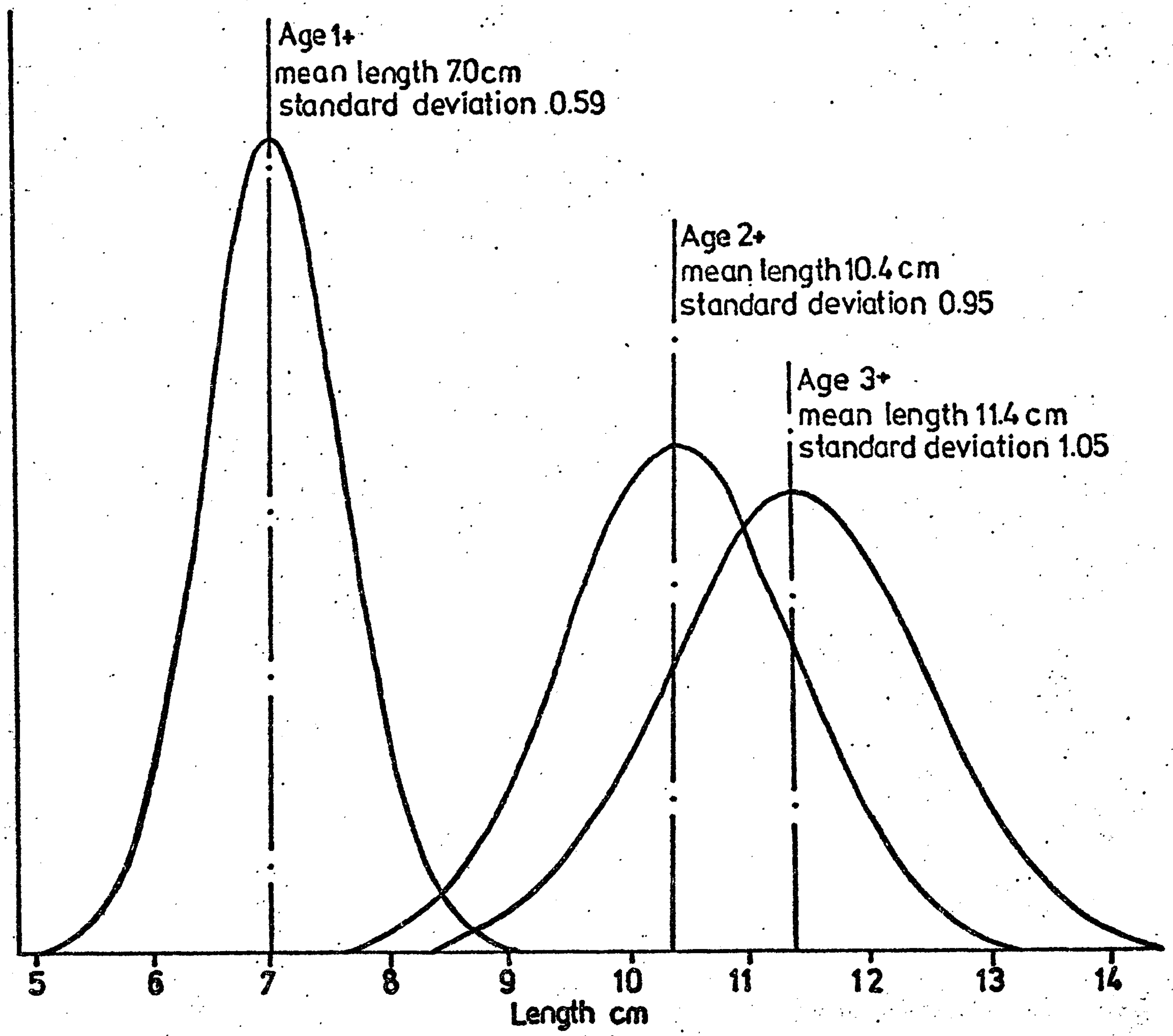


Figure 1 Distribution of length at age for Sprats (*Sprattus sprattus* (L.)) showing the amounts by which the three age groups overlap.

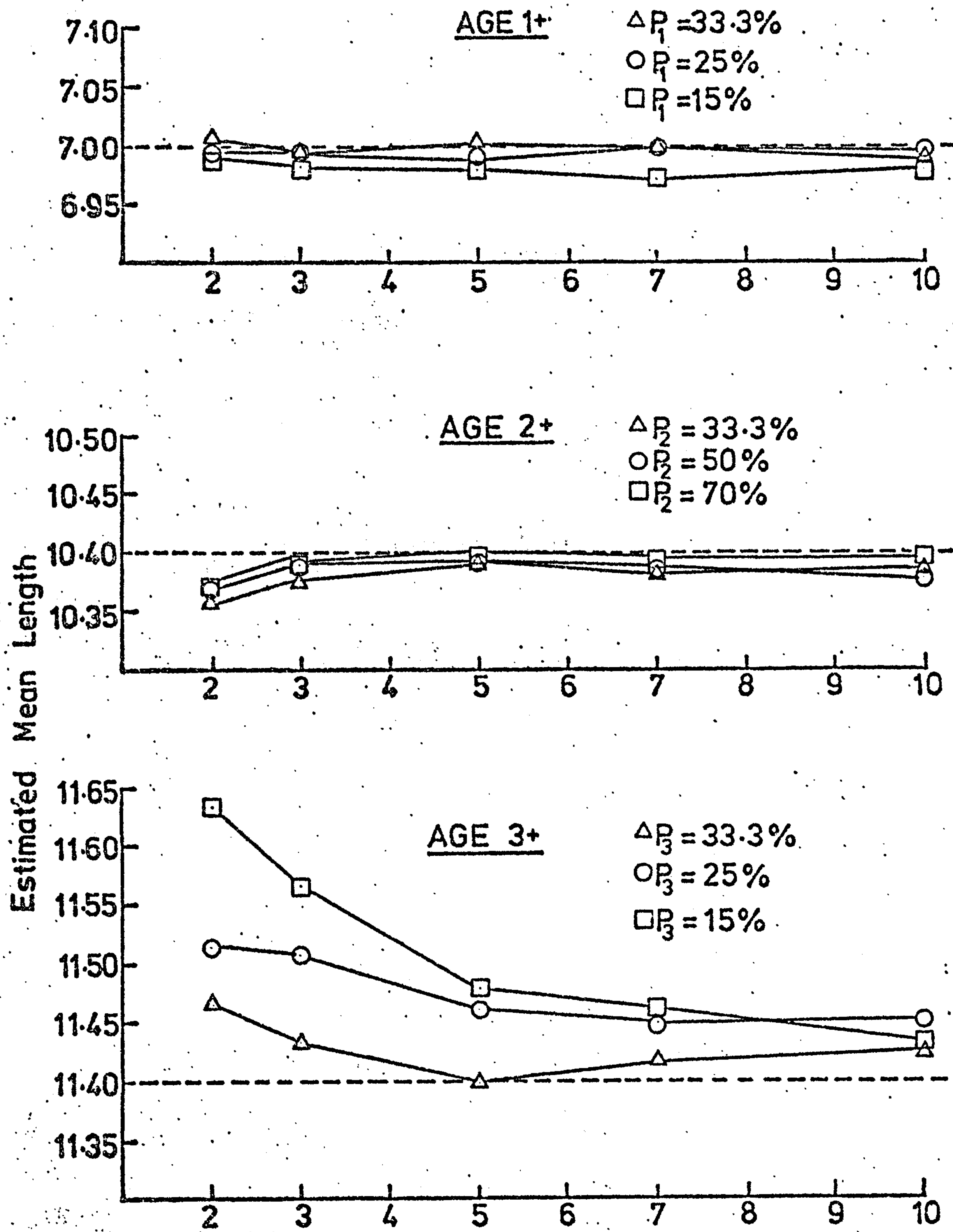
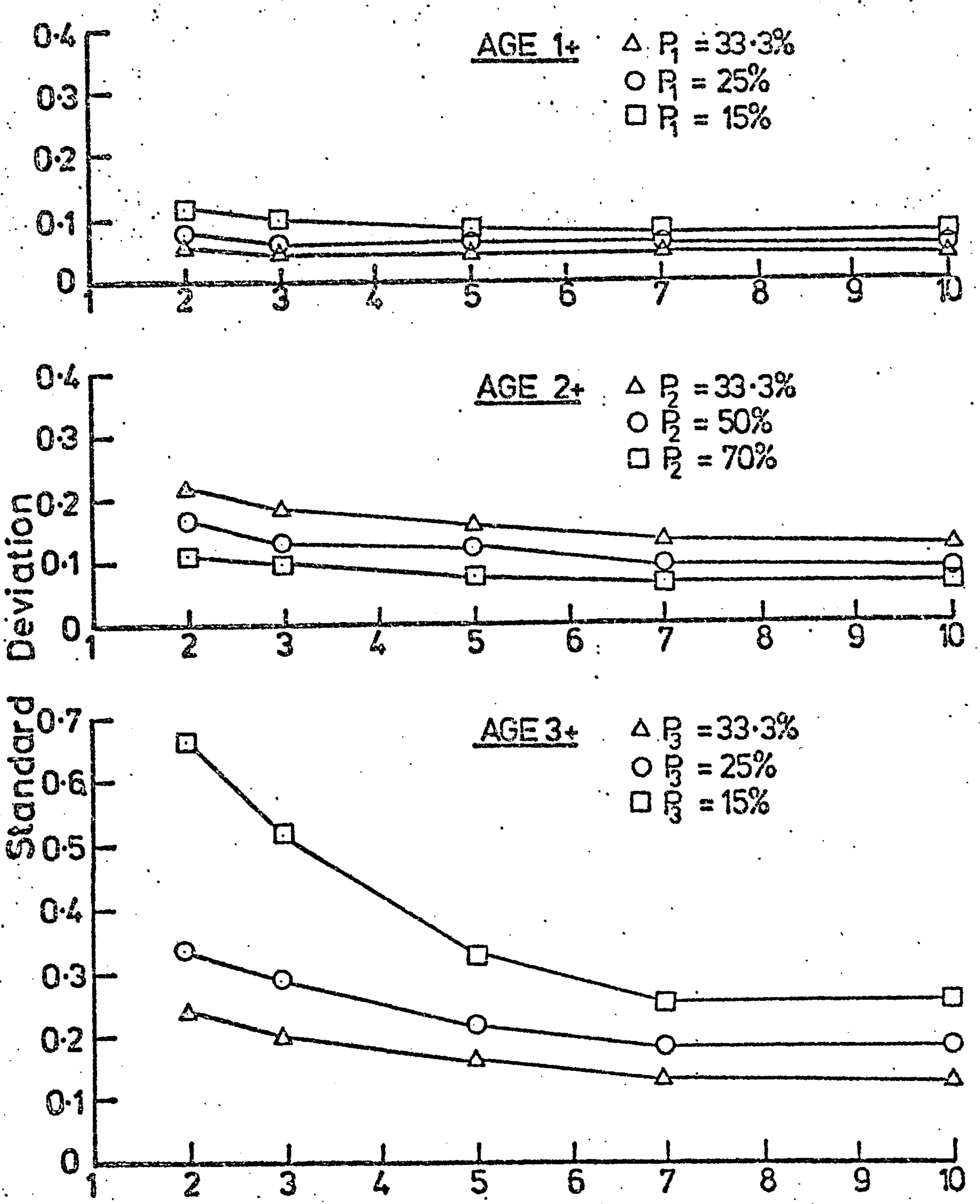


Figure 2 Showing for each age group the average estimated mean lengths plotted against number of fish aged per half centimetre for varying percentage composition by age. Broken lines show the true mean lengths.



<u>Figure 3</u> Showing for each age group the standard deviation of an estimate of mean length plotted against number of fish aged per half centimetre for varying percentage composition by age.