The United Nations University
FISHERIES TRAINING PROGRAMME
P.O. BOX 1390, SKÚLAGATA 4
120 Reykjavik, Iceland

Final Project, 2008

EVALUATION OF THE SAMPLING SCHEME FOR CHUB MACKEREL (*Scomber japonicus* ,Houttuyn, 1782) IN THE INSHORE FISHERY IN GHANA.

Samantha Vida Osei
Ministry of Fisheries Ghana
P.O. Box M37, Accra
abeaman82@yahoo.com

Supervisor
Gudmundur Thordarson
Marine Research Institute (MRI)
gudthor@hafro.is

ABSTRACT

Management decisions in fisheries are largely based on estimates from fisheries data collection programmes. The aim of a sampling strategy is to sample a catch to mirror the population of interest. This study was conducted using both an ANOVA (Analysis of Variance) and a block bootstrapped approach to evaluate a preferred sampling scheme to collect length frequency data of chub mackerel (*Scomber japonicus*) from the inshore fishery in Ghana. An extensive data set on haddock (*Melanogrammus aeglefinus*) in Icelandic waters was use to evaluate and develop the sampling scheme for chub mackerel. In general the study showed that for a gain in precision in a sample, the number of samples is more important than the number of measurements in each sample. Thus to get a representative sample of the population; the sampling should be spread out, as fish caught in clusters contain less information about the population than an equal number caught at random. It is suggested from the study that due to constraints by logistics and cost of sampling in the fishery, sample size of 30 with 30 individual measurements in a sample should be considered the absolute minimum number of samples in a sampling scheme for chub mackerel annually.

TABLE OF CONTENTS

1.0 INTRODUCTION 4

1.1 Justification of present study 5

2.0 LITERATURE REVIEW 6

2.1 Biology and distribution of Chub mackerel 6

2.2 Scomber japonicus fishery 6

2.3 Sampling strategies 8

2.4 Evaluation of sampling design in fisheries research 10

3.0 METHODOLOGY 12

3.1 Study site 12

3.2 Data collection 12

3.3 Data analysis 13

4.0 RESULTS 16

4.1 ANOVA based analysis of effects of *S* and *M* 16

4.1.1 Chub mackerel data 16

4 .1.2 Haddock data 17

4.1.3. Block bootstrap of different combinations of sample sizes and measurements 18

5.0 DICUSSION 21

5.1 Variance in Chub mackerel data 21

5.2 Variance and length frequency distribution of the Haddock data 21

5.3 Variability within different combinations of sample sizes and measurements 22

6.0 CONCLUSION AND RECOMMENDATIONS 23

6.1 Conclusion 23

6.2 Recommendations. 23

Acknowledgements 23

REFERENCES 25

APPENDICES 29

LIST OF FIGURES AND TABLES

FIGURES

Figure 1. Marine capture fisheries production: top ten species in 2004 (fao, 2006)..... 7
Figure 2. Global capture production for *scomber japonicus* (fao fishery statistic, 2004)..... 7
Figure 3. Catches of chub mackerel in the inshore fisheries from 1998-2007 in the coastal waters of ghana. 8
Figure 4. Coastal map of ghana showing 200m contour line..... 12
Figure 5. Length frequency distribution of *s.japonicus* from 2003 – 2007. 13
Figure 6. 0 A Contour plot of variance for different combinations of sample size (*s*) and number of measurements (*m*) of chub mackerel from 2003 -2007..... 16
Figure 7.0 Contour plots of variance for different combinations of sample size (s) and number of measurements (m) of haddock. (a) variance in 200 bootstrapped samples, (b) length frequency distribution of (a), (c) variance in 26 bootstrapped samples and (d) length frequency distribution of (c). 18
Figure 8: 0 Comparison of length distribution of haddock data (200 bootstrapped sample, (red) with combinations of bootstrapped samples sizes and measurements (blue). (a) sample size = 5 and measurements =30, (b) sample size = 30 and measurements = 30, . 20

TABLES

Table 1.0 Estimates Of Sum Of Squares For Different Combinations Of Sample Sizes and Number Of Measurements In The 200 Bootstrapped Haddock Data From 2000 – 2006 Commercial Data..... 19

1.0 INTRODUCTION

The coastal zone of Ghana lies between Cape Palmas to about 2^{o} longitude East (Longhurst, 1962; Williams, 1968) and this zone is part of the Central West African Upwelling Zone. The area is characterized by a seasonal major coastal upwelling from July to September and a minor one from December to February. During this period, warm coastal water is replaced by cold nutrient rich water from off the continental shelf. This then results in increased productivity (Wiafe, 2002). The major fishing seasons in Ghana in the months of the upwelling as fishermen have a high catch rate harvest during these times.

The coastline of Ghana extends to about 550km in length, it has an exclusive economic zone of 24,300km^2 with a narrow continental shelve of approximately 80km. There are four fishing regions namely western, central, greater Accra and Volta regions with about 300 fishing villages along the entire coastline.

Marine fisheries are an important sector of the economy. It contributes 3% of the nation's gross domestic product (GDP) and about 5% of the agriculture GDP. Fisheries are now the country's most important non-traditional export and earn over US$119million from the export of fish and fisheries products in 2003 (Directorate of Fisheries unpublished data). Fish is the cheapest and preferred source of animal protein in Ghana with a per capita consumption of fish 25kg per annum and about 75% of total production is consumed locally.

The fleets operating in the marine fishery are industrial, semi-industrial or inshore and the artisanal fleets. The artisanal fishery is mainly by wooden dugout canoe of less than 10m of which 50% are motorized. In the semi-industrial or inshore fisheries, the fleet is made of wood from 8 – 37m in length and motorized. The industrial fisheries have steel hulled vessels of about 30m in length for the trawlers, 30m long shrimpers and 49 – 60m in the tuna fishery. These fleets exploit both the pelagic and demersal resources in the coastal waters of Ghana. The pelagic resource constitutes about 80% of the total landings in the country. Sardinellas (*Sardinella aurita and Sardinella madrensis*) and Chub mackerel (*Scomber japonicus*) are the most important small pelagic fish species. Over the past ten years, there have been wide fluctuations in the catches of marine species in the Ghanaian coastal waters averaging about 250,000 t annually (Marine Fisheries Research, 2007). However the biomass estimate for pelagic resource have decreased from about 54,000t in 2005 to 48,000 t in 2007 (Marine Fisheries Research, 2007). Attention should be paid to this trend in decline of the pelagic resource by implementing appropriate management measures to prevent further decline and to sustain the stocks.

Chub mackerel is an important pelagic species which constitutes about 30% of the total annual inshore catch. Despite the importance of this species to the pelagic marine fishery and to the economy, there is limited knowledge on its growth and population dynamics. The fundamental factor underlying the inability to manage the stock is the lack of research on the biology, ecology and its response to exploitation.

The first step in managing the *S. japonicus* resource in the Ghanaian coastal waters is to develop an effective sampling scheme to assess and monitor the resource. The primary goal of this project is to use sample length frequency data from the inshore fishery:

- Evaluate the sampling scheme in the inshore fishery using an ANOVA-based method proposed by *Helle and Pennington (2004)*. As there is limited data on *S. japonicus*

from Ghana, data on haddock caught in Icelandic waters is used to test the reliability of the method when projecting outside of the range of the data used.

- Suggest suitable sampling strategy for data collections in the fishery, using both the ANOVA method and block bootstrap.

1.1 Justification of present study

The process of fish growth is influenced by both biotic and abiotic factors in a complex manner. Chub mackerel is also reported as one of the stocks that could be affected directly or indirectly by climate change (Rothschild, 1996). To properly manage a living resource, knowledge on the population dynamics and biology must be understood. Growth parameters and mortality rates are important parameters for assessing sustainable exploitation level of pelagic species (CADIMA, 2000). Lorenzo and Gonzalez (1993) emphasized the importance of regional studies to enhance more understanding of the behaviour of the Chub mackerel in other less known geographical areas, in their study of this species off the Canary Islands.
Management advice on fisheries is often based on length frequency data from commercial landings or surveys and with other information such as catch and effort measurement. Majority of information about the life cycle of this species comes mainly from the Pacific region and other parts of the world (Schaefer, 1980). However, in spite of the economic importance and widespread occurrence of the Chub mackerel in the coastal waters of Ghana during the upwelling season, little is known about its biology, distribution, ecology, and population dynamics in the region. Length frequency and biological data on this species in Ghana is limited and therefore estimation of the status of this stock becomes difficult. The aim of this work therefore is to use recent length frequency data to evaluate a preferred sampling scheme for S. *japonicus* in the inshore fishery which would be sufficient to use for estimating its population parameters in the coastal waters of Ghana.

2.0 LITERATURE REVIEW

2.1 Biology and distribution of Chub mackerel

Scomber japonicus (Houttuyn, 1782), commonly called Chub mackerel of the scombridae family is primarily a coastal pelagic species, found within a depth range of 0 – 300m (Collette and Naunen, 1983) but is usually most abundant at around 50 -200m in subtropical waters of about 10 – 27^{o}C (Castro *et al*, 2000). It has a wide distribution over the continental shelves of the tropical and subtropical regions of the Pacific, Indian and Atlantic Ocean and adjacent seas (Collette and Naunen, 1983).

There is a geographical variation in the growth rate of Chub mackerel within the different zones of its distribution. The growth rate of the species in the North West Mediterranean, North East and South West Atlantic are significantly different (Perrotta *et al*, 2005). FAO, 2001 estimated the life span of this species in Ghana to be five to six years. It moves to open waters to feed at night but stays near the bottom during the day. During the reproductive season, adults migrate from deeper shelf-break waters to shallow coastal areas to spawn (Collette and Nauen, 1983; Cousseau *et al*., 1987; Catro and Santana, 2000; Perrotta *et al*., 2001). Chub mackerel spawns in batches at a water temperature of 15^{o}C to 20^{o}C. Yamada *et al*; (1998) estimated that Chub mackerel in Japanese waters spawned every 5.7 days (6.3 times) over a period of 36 days. Chub mackerel starts schooling at an early stage of its life. Schooling is often by size and it initiates at approximately 3cm of length, however they also form schools with species like *Sarda chiliensis*, *Trachurus symmetricus* and *Sardinops sagax* (Collette, 1995).

2.2 Scomber japonicus fishery

S. japonicus is highly exploited species worldwide (Figure 1). It is of high economic value and placed sixth in the total world nominal marine catches ranking (FAO, 2004). The global landing reached its peak (3 million tonnes) in 1978 and has been fluctuating over the years between 1.5 – 2.0 metric tonnes (Figure 2). Purse seining is presently the most predominant method used in fishing S. japonicus. However there are other types of gears like lampara nets, set nets, traps nets, gill net and even trawls used in this fishery (FAO, 2009). Over its total distribution, the countries with the largest catch of this resource are China and Peru with an average of 500,000mt each.

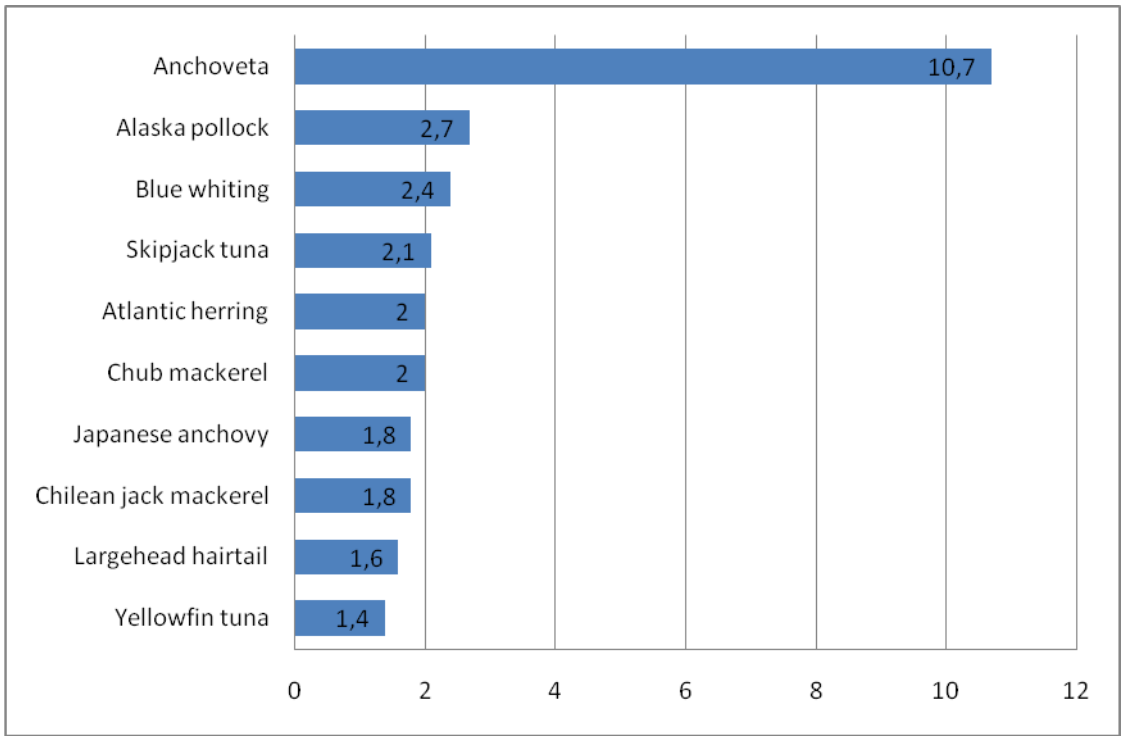

Figure 1: Marine capture fisheries production: Top ten Species in 2004 (FAO, 2006)

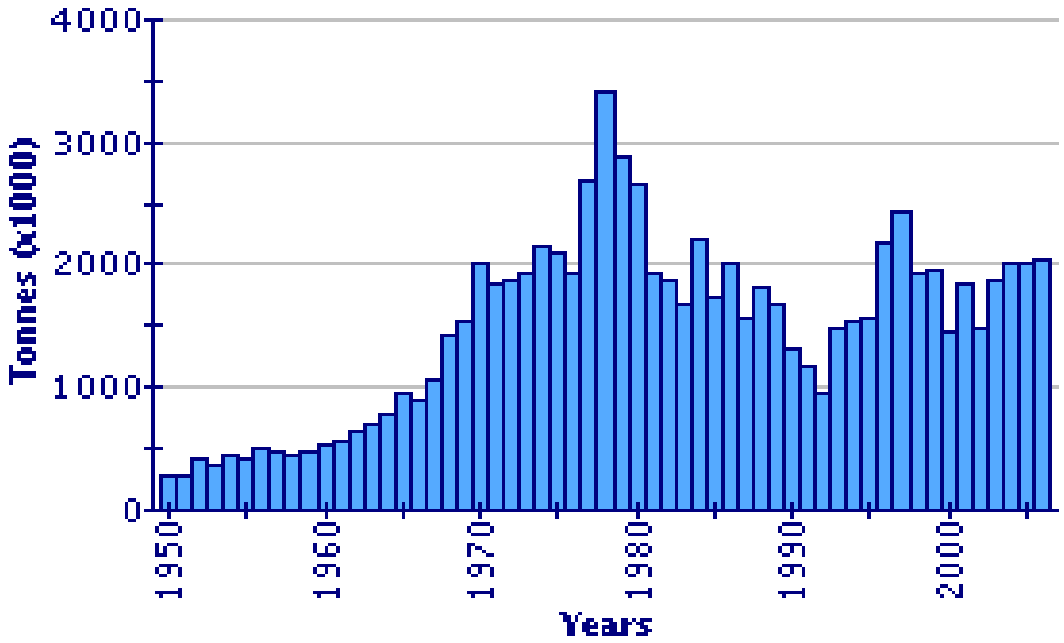

Figure 2: Global Capture Production for *Scomber japonicus* (FAO Fishery Statistic, 2004)

Chub mackerel locally called 'saman' is an important pelagic resource in the coastal waters of Ghana. It is one of the most valuable species in terms of economic value, abundance and quality. Its importance to the fishing industry led to the setting up of a pilot cannery at Osu, Accra in past years (Koranteng, 1995). Chub mackerel fishery is seasonal and coincides with

the upwelling seasons in the region. The species is more abundant in the major and minor upwelling seasons because of the availability of food during these seasons but catches are low for the rest of the year. It is mostly fished with a purse seine net with a mesh size of 44mm in the inshore fishery at an average length of 10 – 26cm, but grows to over 30cm in length (fork length). In the inshore fishery it is one of the main target species and constitutes about 27% of the total inshore catch in 2005 (Marine Fisheries Research Division, unpublished data). It is locally marketed smoked, fried or fresh.

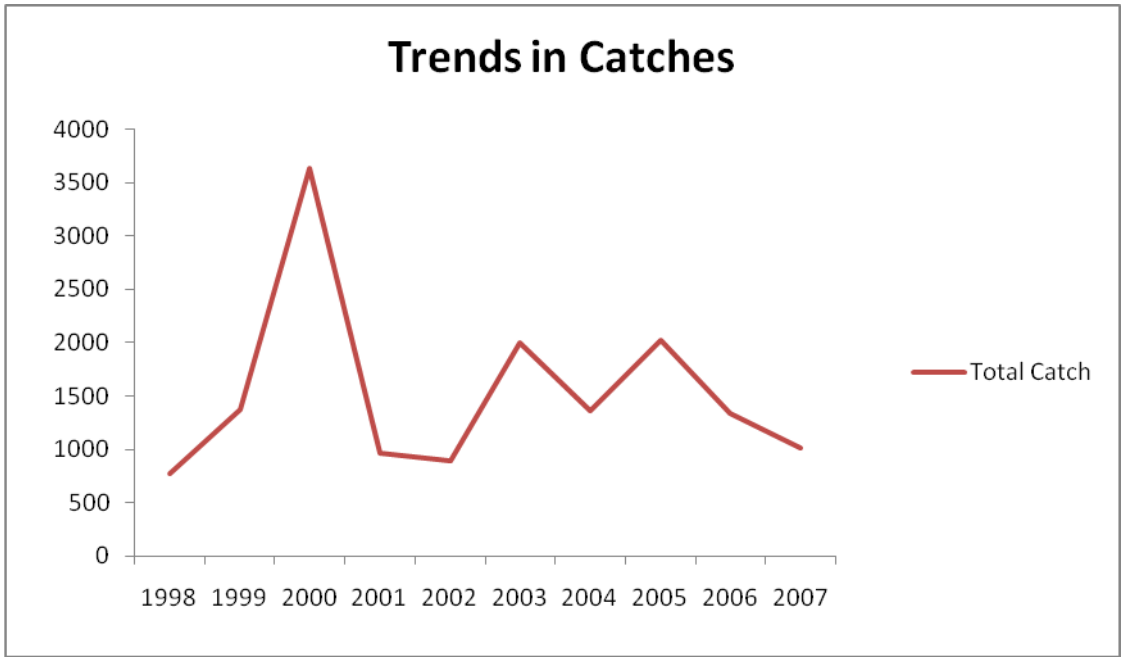

Figure 3: Catches of chub mackerel in the inshore fisheries from 1998-2007 in the coastal waters of Ghana.

2.3 Sampling strategies

Sampling is stated as selecting a portion of a population which will be a representation of the whole population in a research area (Lohr, 1999). The objective of sampling in fisheries research is to obtain data from stocks and on their exploitation, analyse the characteristics of the resources, the effect of exploitation on the abundance of resources and to determine sustainable exploitation levels now and for posterity (FAO, 2005).

The basic terms population, sample and sampling associated with the sampling process must be distinguished. There may be little or no knowledge on the population of interest but the elements of the population should be well defined. The population can be finite or infinite. A population is termed finite when the total number or size of the population is known and infinite when the population is too large and cannot be estimated. Variance, standard deviation, co-efficient of variation and the range are some common measures of dispersion of the values of the characteristic in a population (FAO, 2005).

A sample represents a subset of a controlled size of a population. When a sample is large it can be grouped in to classes (FAO, 2005) and can thus be termed as absolute frequency,

relative frequency and cumulative frequency depending on how the elements are grouped in each class. It is from the sample data that the characteristics of the population are estimated. Statistics of location and dispersion are values calculated from the sample data. Mean, median and mode are examples of statistic of location whilst range, variance, standard deviation and co-efficient of variation are examples of statistic of dispersion (Gallucci *et al*, 1996).

Sampling is the process of selecting individual observations intended to yield some knowledge about the population of interest, especially for the purpose of statistical inference. Statistical inference is carried out with predefined precision based on the values of the characteristics of interest in the sample selected and the properties of the sampling (Kish, 1995).

The aim of a sampling strategy is to sample the catch in a way that represents the population from which the samples were taken. The sampling methodology and sampling frequency determines the quality of a sampling strategy (John, 2003). A very essential ingredient needed for sound fishery research is reliable data from landing ports or research surveys. The collection of data is based on an overall strategy which should clearly define which vessels are sampled in the fleet, which events on the vessels are sampled and what catch is sampled from a fishing event. In addition assessment of the baseline information available on the fishery, evaluation of the sampling procedure of the fishery, assessment of the operational considerations for the fishery and the strategy design for formulating a good sampling plan is indispensable (FAO, 2002).

The manner in which a sample is selected is an important factor in determining how useful the sample will be for making inference about the population from which it is selected (*Helle and Pennington,2004*). Simple random sampling, stratified random sampling, cluster sampling and sometimes multispecies sampling which combines several of the basic methods are different methods for selecting a sample in fisheries research (FAO, 2005).

Simple random sampling is not frequently used in fisheries research but as part of more complex methods. Selection of samples with this method is in two ways - with or without replacement, however sampling without replacement is commonly used. Each individual is chosen entirely by chance and each member of the population has an equal chance of being included in the sample. (Gallucci *et al*, 1996)

In stratified sampling, the population is divided into groups called strata. A sample is then drawn from within these strata by simple random sampling. This is generally used when the population is heterogeneous and produces estimates with smaller variance than simple random sampling (Gallucci *et al*, 1996). The theory of stratified sampling deals with the properties of sampling distribution of the estimators and with different types of allocation of the sample sizes to obtain the maximum precision (FAO, 2005). This method is usually used in biological sampling of landings and scientific surveys (FAO, 2002).

Cluster sampling divides the population into groups, or clusters. A number of clusters are selected randomly to represent the population, and then all elements within selected clusters are observed in the sample. In cluster sampling, only a few clusters are sampled. Hence no elements from non-selected clusters are included in the sample (Australian Bureau of Statistics, 2004). Partitioning the clusters in such a way that they all have similar mean values increases the precision of the estimates. This differs from stratified sampling, where some units are selected from each group. This method of sampling has been used in fisheries to

estimate landings per trip from data of artisanal fisheries with many landing sites and a small number of vessels operating from each site. Cluster sampling is a method that may prove to be cost effective for artisanal fisheries (Gallucci *et al*, 1996).

Multi-stage sampling is like cluster sampling, but involves selecting a random sample within each chosen cluster, rather than including all units in the cluster. At each stage there is a random selection of the sampling units. Thus, multi-stage sampling involves selecting a sample in at least two stages. In the first stage, large groups or clusters are selected (Australian Bureau of Statistics, 2004). These elements are designed to contain more population units than are required for the final sample. Population units are chosen from selected clusters to derive a final sample in the second stage. The process of choosing population units within clusters continues until the final sample is achieved if more than two stages are used.

All sampling methods rely on random sampling; however it is practically impossible to get a perfectly random sample. Bias is a reflection of the difficulty in obtaining a truly representative sample (Lohr, 1999). Bias sampling occurs when certain members of the population have no probability of being selected, leading to the under or over estimation of the population parameter in the sample data. However, if the degree of underrepresentation is small the sample is dealt with as a reasonable approximation to a random sample. Some common examples of bias sampling are under-coverage, voluntary response bias, sampling error, non-response and response bias (Kalton, 1983).

Fisheries research is often concerned with the estimation of the population mean and totals and also the proportion of the population that shares some characteristics of interest (FAO, 2005). Selection of samples with adequate criterion makes it possible to measure with precision the conclusions or inference about that population.

2.4 Evaluation of sampling design in fisheries research

Due to the importance of fisheries to many nations and organization, rules and regulations for management are implemented to conserve the fishery for posterity. Many fisheries organisation's relay on estimates from data collection programmes to formulate management decisions to sustain their fisheries. The number of individuals to include in a research study, the sample size of the study, is an important consideration in the sampling design of much fisheries research. Again determination of the sample size is important for economic reasons because samples that are too large may waste time, resources and money whilst an under-sized sample may not have the capability to produce useful results.

Although there is a wide range of approaches to sampling and the type of data collected, few studies have been conducted to address the effect of sample design and sample size in estimating the parameters related to fisheries (Goodyear 1995, Brouwer and Griffith 2005, Yamaguchi and Matsuishi 2007).

Goodyear (1995) advocated that the use of length stratified sampling for growth estimates and for developing models of mean length at age should be avoided in his study on the influence of sampling protocol on the estimation of the mean lengths at age using computer simulation of a population of red grouper (*Epinephelus morio*)

Folmer and Pennington (2000) studied a statistical evaluation of the design and precision of a shrimp trawl survey off West Greenland. In their study various statistical techniques were used to estimate two indices of shrimp abundance and their precision, and also to determine the effective sample sizes for estimates of length-frequency distributions. They concluded that the surveys produce a fairly precise abundance index and that given the relatively small size; reducing tow duration to 15 minutes would increase the overall survey precision.

The precision of age, size, growth and mortality parameters at samples of 25 to 1000 using bootstrapped population samples were discussed by Kritzer *et al* (2001) in their study on the effects of sample size and population structure on the precision of demographic parameter estimates of four reef fishes; *Cephalopholis cyanostigma, Lethrinus miniatus, lutjanus carponotatus and Plectropomus leopardus*. From the study they suggested that estimation of parameters such as mean length, mean age and modal age for less commercially important species can be estimated with high precision at small sample sizes when compared to other parameters. In addition they also suggested the use of extant data sets for species similar to the focal species in case of limited data situation to approximate the population in question. They also added that when the research population is substantially different from the proxy population an additional sampling is required.

Helle and Pennington's (2004) studies on survey design considerations for estimating length compositions of the commercial catch of some deep-water species in the north east Atlantic revealed that the precision of estimates of length distributions is a function of both how the fish was sampled and the number of fish sampled. The study showed an efficient sampling design is to collect relatively small samples from many vessels as possible to gain precision as fish caught together tend to be more similar than those in the entire population.

Investigations by Brouwer and Griffiths (2005) on the influence of sample design on estimates of growth and mortality in *Argyrozona argyrozona* and the effects of either using random or stratified sampling procedures showed that random and stratified sampling produced similar results for growth, fishing mortality and spawning biomass per recruit. Thus both the random and stratified estimates revealed that a minimum of 300 random samples or at least 10 fish per 2cm size class (ie n = 193) were necessary to provide reliable estimates of growth, fishing mortality and spawning biomass-per-recruit. But to optimise the trade-off between cost and parameter precision, stratified sampling of *Argyrozona argyrozona* during the spawning season is preferred.

Studies by Jurajda *et al* (2008) on the evaluation of sampling methods in floodplain lakes including whole-lake sampling showed that the accurate representation of the fish community using just one sampling method and strategy is not feasible even in a small floodplain lake. In addition the study showed that in regarding the ability to capture representative samples, the behaviour of particular fish species seems to be a more significant factor than fish size.

3.0 METHODOLOGY

3.1 Study site

The study considered the coastal zone of Ghana which lies between Cape Palmas to about 2^{o} longitude East (Longhurst, 1962; Williams, 1968).

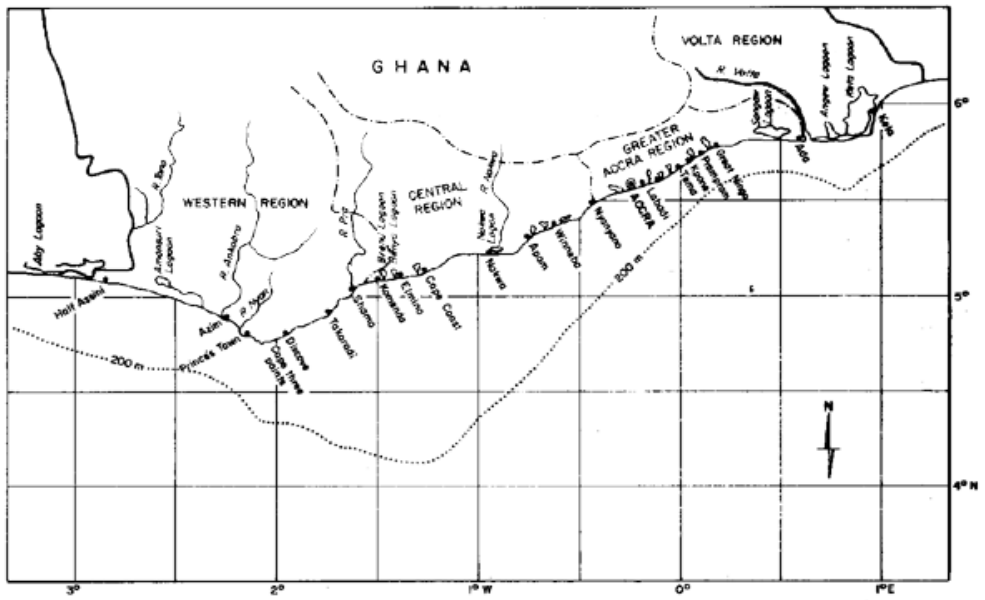

Figure 4: Coastal map of Ghana showing 200m contour line

3.2 Data collection

Length (forked cm) frequency data sampled from commercial landings at Tema Fishing Harbour within the major upwelling months in Ghana (July, August and September) from Inshore Vessels will be used in this study. Purse seining net of 40mm mesh size is the main gear used for fishing this stock in the 30-50m contour depth. A sample is collected randomly from a vessel landing within the specified day for sampling and this is done once, twice or thrice a month for the major fishing season within a year. The number of length-measured chub mackerel in each sample ranges from 45 to 240, a total of 447 to 917 are measured a year and in all 3371 fishes were measured from 2003 to 2007. Fork lengths measurements of individual fish were taken using graduated board and measurements of the lengths were taken from the snout to the end of the forked portion of the caudal fin, to the nearest centimetre below. Individual lengths were grouped by months and then by years.

Twenty six samples with an average of about four samples in a year are collected annually and are used in this study. This species is exploited at a size of about 8 – 30 cm but mostly fished at an average size range of 11 – 17cm (Figure 4).

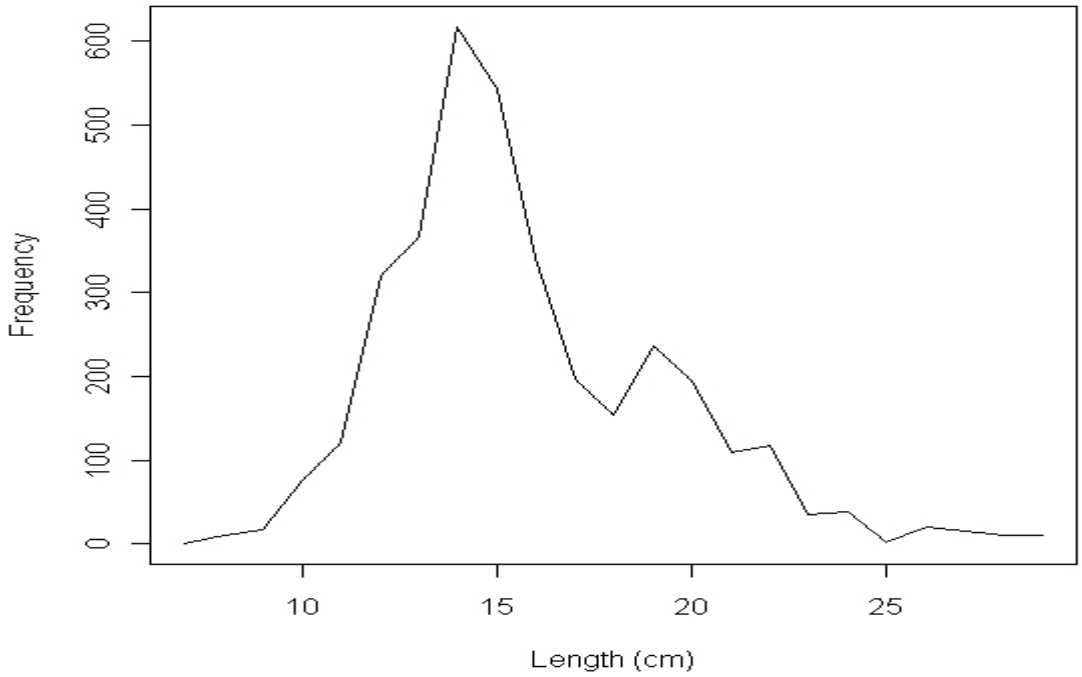

Figure 5: Length frequency distribution of *S.japonicus* from 2003 – 2007.

3.3 Data analysis

Variance component analysis (one-way anova) was used to quantify the sources of variability and based on these estimates an efficient sampling scheme can be devised. The model is:

$$y_{sj} = \alpha_s + \varepsilon_{sj}$$

where y_{sj} is individual length measurement from a unique sample. α_s is a factor in the model and denotes an individual sample and ε_{sj} is an error term. The model was implemented in R (R Development Core Team, 2005). Assuming that a sample is randomly chosen then the components will be independent and therefore the variance of *y* is given by

$$V|y_{sj}| = \delta_A^2 + \delta_E^2 .$$

Where δ_E^2 is the mean square error from the model and δ_A^2 is defined as

$$\frac{MSR - MSE}{n}$$

Where MSR and MSE are the between and within mean squared errors. *n* is the number of samples. If sample size is unequal then *n* is replaced by *n'* which is defined as

$$\acute{n} = \frac{1}{r-1}\left[\left(\sum n_i\right) - \frac{\Sigma n_i^2}{\Sigma n_i}\right]$$

where *r* is the total number of samples.

The variance component model for this data is a completely nested model and therefore analysis of variance techniques were used to estimate the variance components (See Neter *et al*., 1996).

To assess various sampling schemes it is assumed that *S* samples are taken and *m* fish are measured in each *S*. In addition it is also assumed that the samples and fish measured are completely random. Then the variance of the estimator of the mean length is given by

$$V_{\bar{y}_{sj}} = \frac{\delta_A^2}{S} + \frac{\delta_E^2}{SM}$$

where S and SM are sample size and total samples respectively.
The value of *V* was calculated for all combinations of *S* and *M* from 1 to 100.

3.3.1 Using Haddock (Melanogrammus aeglefinus, Linnaeus, 1758) as a proxy data

The method described above makes it possible to explore the effects of increasing the number of samples and/or increasing the number of measurements. Therefore in theory it is possible to explore the decrease in variance by increasing the number of samples way beyond the actual number of samples used in the ANOVA analysis. As the chub mackerel data only contained 26 samples it was of interest to investigate if the results from 26 samples were different when using a much larger data set containing many more samples.
To test this length frequency commercial catch data of haddock from 2007 collected from trawlers in Icelandic waters was used containing 200 samples, each containing more than 100 measurements. First the ANOVA model was applied to the whole data set and the changes in *V* where calculated for all combinations of *S* and *M*. Then a random sub-sample containing 26 samples was taken from the haddock data set and the analysis redone. The results were then compared visually.
Haddock belongs to the gadoid genus (cod-like species) and the fish is wide spread throughout the deeper waters of the temperate northern Atlantic, and shoals at depths less than 300m with preference for between 75 and 125m depths and are usually between 50 - 65cm in catches (Icelandic Fisheries, 2009).
Haddock differs from chub mackerel in terms of its environment, distribution and biology. Haddock is a demersal species whereas chub mackerel is pelagic, sizes of chub mackerel in commercial catches ranges between 10 – 30cm in the coastal waters of Ghana and sizes haddock in catches from Icelandic waters is between 50 – 65cm. However haddock data is characteristic of that of the chub mackerel in terms of within sample correlation, thus some samples contain only small fish where as others contain only large individuals as seen in the samples collected on chub mackerel. Hence it is utterly important to use the haddock data which is more reliable and abundant to estimate the preferred sampling scheme for chub mackerel.

In spite of these differences it is fully justified to use the haddock data to test if the projections outside of the data range produced from the method hold. As both the chub mackerel and haddock data sets contain similar within sample correlations the result from the analysis of the haddock data should hold for the chub mackerel data.

A block bootstrapped technique was used to explore how length distributions produced by different sampling schemes compared to the actual length distribution (the whole data set) would differ. The technique is a statistical method that uses random re-sampling of data from

an original data set. Random samples were bootstrap (Efron *and* Tibshirani, 1993) from an original haddock data set and the bootstrapping routine was carried out in R (R Development Core Team 2007).

Each resulting bootstrapped sample combination was compared with the original data by estimating the sum of squares between the bootstrapped sample combinations and the original combinations.

4.0 RESULTS

4.1 ANOVA based analysis of effects of *S* and *M*

4.1.1 Chub mackerel data

The analysis of changes in variance, with different combinations of number of samples (*S*) and measurements (*M*), of the chub mackerel data showed a marked decrease in variance with increasing number of samples (*S*). Thus a fourfold decrease is realised in variance or from 2 down to 0.5 when the number of measurements is maintained at 20 and *S* is increased from 3 to 10. Similarly the variance decrease from 0.5 to 0.1 when *S* is increased from 10 to 30 and measurement is maintained at 20 (Figure 6).
However the variance does not decrease a lot with increased number of measurements (*M*), as long as the number of measurements is around 20 or more. An exception to this observation is when *S* is more than 50 then there seem to be a decrease in variance with increasing *M*. However there is though little reduction in variance if *M* exceeds 40 (Figure 6).
The results suggested that more samples were needed than actually were available for chub mackerel.

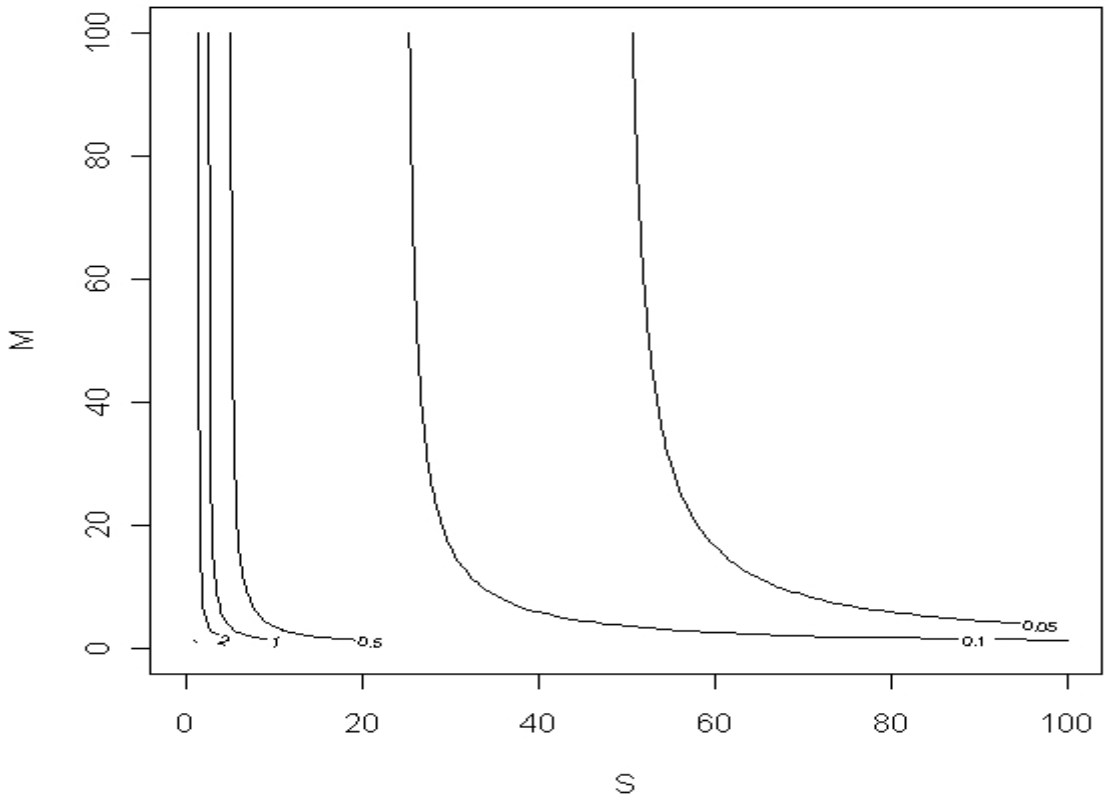

Figure 6: A contour plot of variance for different combinations of sample size (*S*) and number of measurements (*M*) of chub mackerel from 2003 -2007.

4 .1.2 Haddock data

There does not seem to be much difference between the length distributions of the original haddock data which contains 200 individual samples, each with more than 100 measurements and its subsample which contains 26 samples. (Figure 7 (b and d)). For example both datasets show a distinct peak at 45cm length.

The results of the variance analysis for both 200 and the 26 bootstrapped samples of haddock are almost identical (Figure 7a and 7c). Thus the variance decreases with increase in the number of samples (*S*). Keeping the number of measurement constant at 20, there is a fourfold increased decrease in variance (from 2 to 0.5) by increasing *S* from 3 to 10. Again a fivefold decrease in variance (0.5 to 0.1) is realised by increasing *S* from 10 to 30. Variance does not decrease a lot when the number of measurements (*M*) is increased as long as measurements are around 20. Nevertheless when *S* is more than 50, there tend to be some decrease in variance with increase in *M*. But when *M* is over 40 there is little reduction in variance. These results are similar to those obtained from the Chub mackerel data.

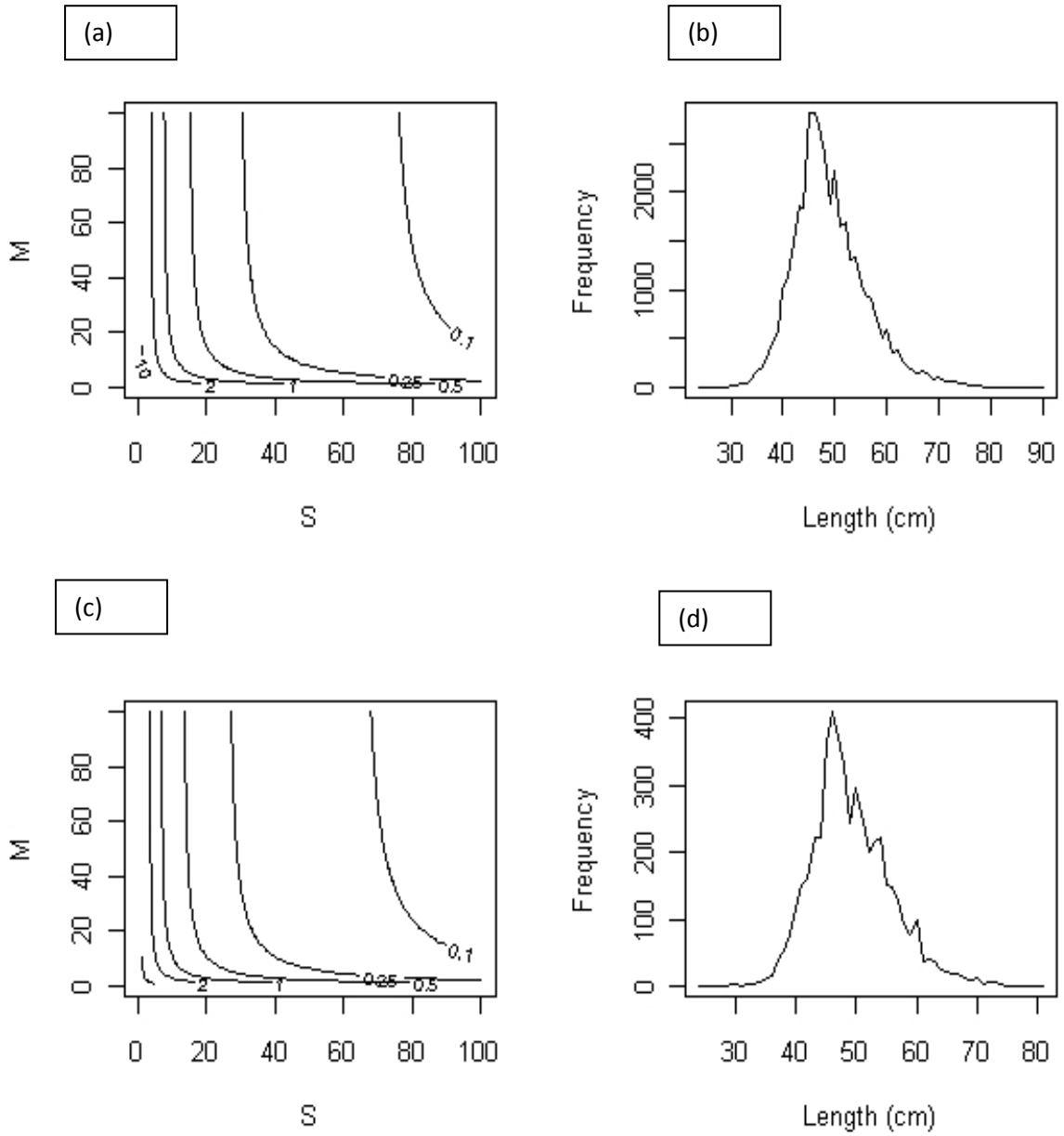

Figure 7: Contour plots of variance for different combinations of sample size (S) and number of measurements (M) of haddock. (a) Variance in 200 bootstrapped samples, (b) length frequency distribution of (a), (c) Variance in 26 bootstrapped samples and (d) length frequency distribution of (c).

4.1.3. Block bootstrap of different combinations of sample sizes and measurements

Sum of squares analysis for different combinations of sample sizes and measurements showed significant decrease in variance with increasing sample sizes when *M* is kept constant (Table 1.0). On the other hand if *S* is kept constant and *M* is varied little or no change in variance is observed. This is seen in combinations of 20 and 30 *S* with *M* of 30 having variance of 0.0018 and 0.0023 respectively. The difference in this is only 0.0005 which is insignificant. This implies the decrease of variance in the samples above *S* of 20 with a constant *M* of 30 is gradual and might not affect the variance in the sample.

Table 1: Estimates of Sum of Squares for different combinations of sample sizes and number of measurements in the 200 bootstrapped haddock data from 2000 – 2006 commercial data. Examples of length distributions are shown in Figure 8 for the *S* and *M* combinations in italics

Sample size (*S*)	Measurements (*M*)	Sum of Squares
1	1	0.9941
5	*5*	*0.0400*
5	30	0.0059
10	10	0.0153
10	30	0.0051
20	5	0.0065
20	10	0.0049
20	30	0.0018
20	40	0.0023
20	80	0.0009
30	*30*	*0.0022*
40	30	0.0015
80	30	0.0009

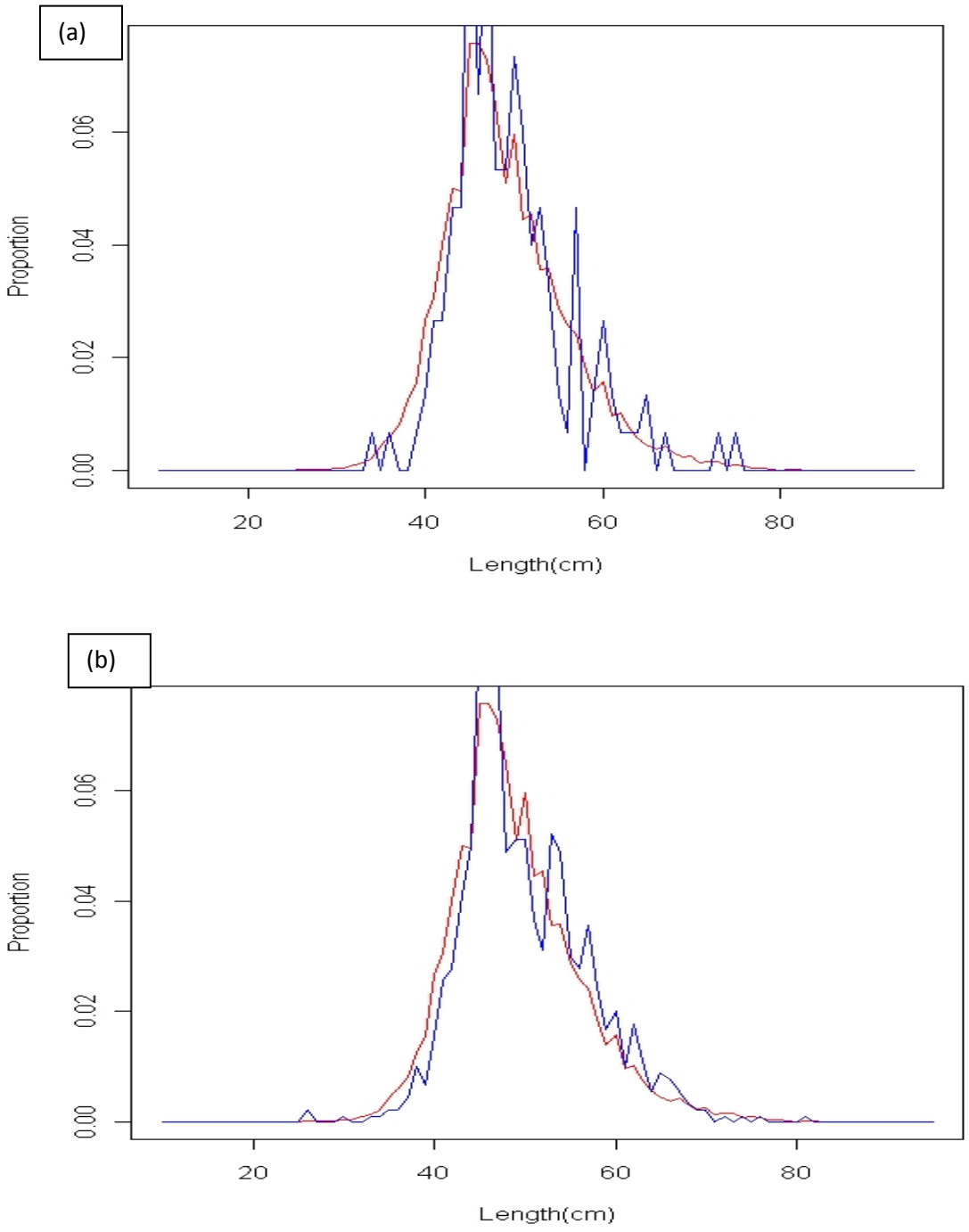

Figure 8: 0 Comparison of length distribution of Haddock data (red) with combinations of bootstrapped samples sizes and measurements (blue). (a) sample size = 5 and measurements =30, (b) sample size = 30 and measurements = 30.

5.0 DISCUSSION

In an ideal survey, the sample population should be a scaled down version of the population, mirroring every characteristics of the whole population but this ideal situation is rarely met exactly (Lohr, 1999). A good sample would reproduce the characteristics of interest in a population as closely as possible. However, minimum data requirements will vary depending on which parameter is being estimated and the nature of the sample, even with the same fish species. Thus at small sample sizes, parameters such as mean length and age can confidently be estimated.

Availability of reliable data is of great importance in the assessment of exploited stocks and hence for management decisions. But often there are limited resources for data collection and therefore compromises in data collection in terms of accuracy have to be accepted. In formulating a sampling scheme for a given resources it is expedient to get a sample that is representative of the population given the resource available.

5.1 Variance in Chub mackerel data

Based on variance component analysis Helle and Pennington (2004) predicted that the only way to decrease variance in a sample would be to increase the number of boats that collected samples which in this case is the sample sizes. It has been shown from the variance analysis of the Chub mackerel data that the variance decreases with increasing sample sizes. Thus by increasing the number of samples threefold and keeping the number of measurements at 20, there is a four to fivefold decrease in variance. This suggests that in order to improve precision in estimates derived from catch data, increasing the number of samples is of uttermost importance.

5.2 Variance and length frequency distribution of the Haddock data

The results from the variance analysis in the Chub mackerel data suggested that more samples were needed than actually were available. Therefore it was of interest to see if the findings from the analysis would change if more samples were used. To test this, a data set from commercial catches of Haddock in Icelandic waters which contained 200 individual samples, each sample containing at least 100 measurements was used. Two runs of the analysis were done, one using all the available data and then with a subsample from the full data set containing 26 samples.

The results obtained from the 26 samples were very identical to that of the 200 samples, which suggests that the results from the analysis of the Chub mackerel data would not change markedly if more samples were available.

Also an objective of sampling method is to estimate the length distribution of the species population. That is the length distribution is a description of the relative abundance of individuals in the population (Folmer and Pennington, 2000). However the fish sampled were not a random sample of individuals from the entire commercial catch, but were selected from a number of clusters. The basic problem with cluster sampling is that fish caught together tend to be more similar in length than those in the entire population (Helle and Pennington, 2004).

The length distributions in this study from both bootstrapped Haddock samples (200 and 26) were very much similar. This result is supported by observations made by Pennington *et al*., (2002) that the practical implication of positive intra-cluster correlation is that a sample of animals caught in clusters generally contain less information on the population structure than equal number of fish caught at random.

5.3 Variability within different combinations of sample sizes and measurements

In order to give some idea of the results, in terms of difference between the population length distribution and the distribution of a certain sampling scheme a block bootstrap was done on the Haddock data. This method is to test the accuracy of the estimates of mean lengths in the distribution. The population length distribution was assumed to be the same as the length distribution from the whole data set. The difference between the two distributions was measured by calculating the sum of squares. The results from the ANOVA and the block bootstrap are similar but the block bootstrap makes it possible to compare the overall distribution whereas the ANOVA method compares only the mean length of the samples. The results from the block bootstrap imply that a sample size of 30 and 30 measurements is a more adequate representation of the Haddock data than the other.

Pennington and Vølstad (1994) showed that a way to determine how much information is contained in a sample collected in clusters is to calculate the effective sample size. This is termed as the number of individuals needed to be sampled at random to obtain the same precision for a population estimate as that achieved by the cluster sample (Folmer and Pennington, 2000).

From the study, to have a better sampling scheme that would represent the population of Chub mackerel, emphasis should be placed on obtaining more samples rather than large samples to decrease the variation in the samples. Hence one must sample more and more to gain less and less variation in the sample. This seems to support observations made by Helle and Pennington (2004) that to gain precision in a sample one needs to spread out the sampling.

6.0 CONCLUSION AND RECOMMENDATIONS

6.1 Conclusion

The study has shown that the number of samples is very important in a data collection scheme and a way to increase sampling precision significantly would be to increase the number of samples collected each year. In contrast an increase in the number of measurements does not increase accuracy. Logistics and cost is of great concern when devising a sampling scheme. The most expensive part of a sampling programme is obtaining each sample and therefore a trade-off is to be set at obtaining a representative sample at the estimated sampling cost. For an optimal sampling scheme, this study suggests that a sample size of 30 with 30 measurements in each sample would be the absolute minimum sample requirement for the Chub mackerel during the upwelling season annually from each landing site.

It is evident from this study the need to develop sampling strategies for fisheries data collection programmes. The bootstrap technique employed here to evaluate and devise a sampling scheme seems to be a laudable idea.

6.2 Recommendations

A holistic study of the biology, ecology and distribution of Chub mackerel in the coastal waters of Ghana and also within the Eastern Central Atlantic region is recommended. Thus tagging studies should be conducted to ascertain the migratory routes of this species and also studies on the spawning season and grounds is vital in understanding the behaviour of this species.

Again, samples from data collection programmes should be sold back to the market after all the necessary information has been collected. So that some of the budget allocated for data collection in the Marine Research Institute in Tema, Ghana can be recycled. It might also be sensible to record more information from at least some of the fish measured, such as weight, sex and maturity. Furthermore ageing should be given serious consideration, either by collection of otoliths or scales.

Further studies are recommended to critically assess the sampling scheme used for small pelagics in the Ghanaian marine fishery. In addition, the study advocates sampling from the two fishing harbours thus Tema and Takoradi and all the landing sites along the entire stretch of the coastline of Ghana so as to mark out similarities and differences if there are between species in the east and west coast and also to ensure an extensive and reliable estimate of the status of the most exploited species in the fisheries.

A good centralized data base for storage of all fisheries data from surveys should be set up and must be easily accessible for analysis.

ACKNOWLEDGEMENT

I am most grateful to God for his blessings, favour and an opportunity for me to part-take in the United Nations Fisheries Training Programme.
A dozen thanks to Gudmundur Thordarson, my supervisor for his patience and assistance in my project work. Not forgetting GunnAr Stefansson who was always there to help me whenever I needed it. I say thank you to Gudni Magnus Eiriksson for his guidance and inputs into this work. I also thank the scientist and crew at the Marine Research Institute, Iceland, who were involved in the sampling and preparation of Haddock data used in this study.
I appreciate Mr. Pierre Coussey, Ministry of Fisheries, Ghana for being a father and an inspiration to me. I extend my sincere gratitude to Mr. Paul Banaman and Mr. Samuel Quartey of Ministry of Fisheries Research Institute, Tema for their immense help. A big thanks to my friends especially Benjamin Akyempon and family and for their prayers and encouragement.

I appreciate the staff of the United Nations University, most especially Thor for granting me a fellowship to participate in the UNU Fisheries Training programme in Iceland. Thanks to all 2008 fellows for imparting my life positively.
God bless you all.

REFERENCES

Australian Bureau of Statistics, 2004, **Themes: Multistage sampling**, viewed 22nd January, 2008. < http://www.abs.gov.au/websitedbs/d3310116.>

Brouwer, S.B. and Griffiths, M.H. 2005. *Influence of sampling design on estimates of growth and mortality in Argyrozona argyrozona (Pisces: Sparidae)*. Fisheries Research 74, 44-54.

Castro Hernández, J.J. and A.T. Santana Ortega, 2000. *Synopsis of biological data on the chub mackerel* (*Scomber japonicus* Houttuyn, 1782). FAO Fish. Synop. 157. 77 p. FAO, Rome.

CADIMA, E. 2000. *Manual de avaliacao de recursos pesqueiros*. Document tecnico sobre as pescas. FAO Fish. Tech. Pap., v.393, p. 10-125.

Collette, B.B. 1995 *Scombridae. Atunes, bacoretas, bonitos, caballas, estorninos, melva, etc*. p. 1521-1543. In W. Fischer, F. Krupp, W. Schneider, C. Sommer, K.E. Carpenter and V.

Niem (eds.) Guia FAO para Identification de Especies para lo Fines de la Pesca. Pacifico Centro-Oriental. 3 Vols. FAO, Rome.

Collette, B.B. and C.E. Nauen 1983 FAO species catalogue. Vol. 2. *Scombrids of the world. An annotated and illustrated catalogue of tunas, mackerels, bonitos and related species known to date.* FAO Fish. Synop. 125(2) 137 p.

Cousseau M. B., Angelescou V., Perrota R. G. Algunas caracteristicas de la estructura y comportamiento migratorio de los cardumenes de caballa (*Scomber japonicus marplatensis*) en la plataforma bonaerense (Mar Argentino), periodo 1965–1984. Revista de Investigación y Desarrollo Pesquero (INIDEP) (1987) 7:21–42.

Efron, B. and Tibshirani, R.J. 1993. *An introduction to the Bootstrap*. Chapman Hall: New York.

FAO, 2002. *Guidelines for developing an at-sea fishery observer programme*. FAO Fisheries Technical Paper No.414, Rome, 2002.

FAO, 2004. *Review of the state of world fisheries resources*. FAO Fisheries Technical Paper No.335, FAO, Rome, 2001, 136pp

FAO, 2005. *Sampling methods applied to fisheries science:a manual* FAO Fisheries Technical Paper No.434, FAO, Rome, 2005.

FAO, 2006. *The state of the world's fisheries and aquaculture*. FAO Fisheries and Aquaculture Department. Food and Agriculture of the United Nations, Rome 2007

FAO, 2009. Species Fact Sheets. Rome, FAO. [February 17, 2009] <http://www.fao.org/fishery/species/3277>

Folmer, O. and Pennington M., 2000. *A statistical evaluation of the design and precision of the shrimp trawl survey off West Greenland.* Fisheries Research 49 (2000) 165-178.

Gallucci, V.F., Saila S.B., Gustafson, D.J., Rothschild, J.B., (1996). *Stock Assessment: Quantitative Methods and Applications for Small-Scale Fisheries.* Lewis Publishers, United States of America.

Gluyas-Millan, M.G., Castonguay, M., *and* Quinonez-Velazquez C., 1997. *Growth of juvenile Pacific mackerel, Scomber japonicus in the Gulf of California.* Scientia marina., 62 (3): 225-231

Goodyear, C.P. 1995. *Mean size at age: An evaluation of sampling strategies with simulated Red Grouper data. Transactions of the America Fisheries Society* 124, 746-755.

Helle, K. and Pennington, M., (2004). *Survey design considerations for estimating the length composition of the commercial catch of some deep-water species in the northeast Atlantic.* Fisheries Research 70 (2004) 55-60

Houttuyn, M. (1782). *Beschryving van eenige Japansce visschen, en andere zee-schepzelen.* Ver. Holl. Maatsch. Wet. Haarlem 20(2):311-350.

Icelandic Fisheries, 2009. Information Centre of the Icelandic Ministry of Fisheries and Agriculture [February, 2009] http://www.fisheries.is/the-main-species/other-codfishes/haddock/nr/276, 02/2009.

John Eng, MD. (2003). *Sample size estimation: How many individuals should be studied*. Radiology and Radiologist Research Statistical analysis, 227:309-313.

Jurajda P., Janac M., White, S.M. and Ondračková M., (2008). *Small – but not easy: Evaluation of sampling methods in floodplain lake including whole-lake sampling*. Fisheries Research 96 (2009) 102-108.

Linnæus, C. 1758. *Systema naturæ per regna tria naturæ, secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis*. Tomus I. Editio decima, reformata. - pp. [1-4], 1-824. Holmiæ. (Salvius).

Lohr, L.S. (1999). *Sampling: Design and Analysis*. Brooks/Cole Publishing Company, USA

Longhurst, A. R., 1962. *A review of the Oceanography of the Gulf of Guinea*. Bull. IFAN, ser. 12, 633-663.

Lorenzo, J. A. *and* J. González. 1993. Las aves de El Médano (Tenerife- Islas Canarias). Asociación. Tinerfeña de Amigos de la Naturaleza. 192 pp.

Lorenzo, N.J.M., Gonzales, P.J.M. *and* Ramos, A.G. 1995. *Growth of the chub mackerel Scomber japonicus (Pisces Scombridae)* off the Canary Islands. Sci. Mar., 59(3-4): 287-291.

Kalton G.(1983). *Introduction to Survey Sampling. Series: Quantitative Applications in Social Sciences.* SAGE Publication Inc. Newbury Park, Carlifornia.

Kish, L., (1995). *Survey Sampling*. John Wiley and Sons Inc. Third Avenue, New York

Koranteng A.K., (1995). The Ghanaian Fishery For Sardinellas. Fisheries Department, Research and Utilization Branch, Tema, Ghana.

Krizter, J.P., Davies, C.R. and Mapstone, B.D. 2001. *Characterizing fish populations: effects of sample size and population structure on precision of demographic parameter estimates*. Canadian Journal of fisheries and Aquatic Sciences 58, 1557-1568.

Marine Fisheries Research (2007). *Marine Research Bulletin*. Volume 1, Issue 1. Ministry of Fisheries Ghana.

Maigret, J. and B. Ly, 1986. *Les poissons de mer de Mauritanie*. Science Nat., Compiègne. 213 p.

Murua, H. and F. Saborido-Rey, 2003. *Female reproductive strategies of marine fish species of the North Atlantic*. J. Northwest Atl. Fish. Sci. 33:23-31.

Neter, J., Kutner, M. H., Nachtsheim, C. J., *and* Wasserman, W. 1996, Applied Linear Regression Models (New York: Irwin)

Pennington, M. And Vølstad, J.H., 1994. Assessing the effective of intra-haul correlation and variable density on estimates of population characteristics from marine surveys. Biometrics 50, 725-732.

Pennington, M., Burmeister, L.-M., Hjellvik, V., 2002. Assessing the precision of frequency distributions estimated from trawl-survey samples. Fish. Bull. 100, 74–81.

Perrotta G.R., Carvalho N., *and* Isidro E., 2005. Comparative study on growth of chub mackerel (*Scomber japonicus* Houttuyn, 1782)from three different regions: NW Mediterranean, NE and SW Atlantic.

Perrota, R.G.; Viñas, M.; Hernandez, D. *and* Trigali, L., (2001) Temperature conditions in the Argentine chub mackerel (*Scomber japonicus*) fishing ground: implications for fishery management. **Fisheries Oceanography**., v.10, n. 3, p. 275-283.

R Development Core Team (2005). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Rothschild, B. 1996. *How Bountiful are Ocean Fisheries*? *CONSEQUENCES*: 2. No1. (www. gcrio.ciesin.org/CONSEQUENCES/winter96/oceanfish.html)

Schaefer, K.M., 1980. Synopsis of Biological Data on the Chub mackerel, *Scomber japonicus* Houttun. 1782, in the Pacific Ocean, Inter-Amer. Trop. Tuna Comm., Spec. Rep., 2: 395-446.
Searle, S.R., Casella, G., McCulloch, C.E., 1992. Variance Components. Wiley, New York.

Sotiris K., Tserpes G. and Tsimenidis N. 2000. *Aspects on the demography of Chub Mackerel (Scomber japonicus Houttuyn, 1782) in the Hellenic Seas.* Belg. J. Zool., 130 (supplement 1): 3-7

Wiafe, G. 2002: *Spatial and Temporal Dynamics of Plankton communities in the Gulf of Guinea Ecosystems*. Ph.D Thesis. University of Ghana.188pp.

Williams, F., 1968. *Report on the Guinean Trawling Survey.* Publ. Sci. Tech. Res. Com. OAU, Vols. I and II

Yamada T., Aoki I., Mitani I., 1998. *Spawning time, spawning frequency and fecundity of Japanese chub mackerel, Scomber japonicus in the waters around the Izu Islands, Japan*. Elsevier Fisheries Research 38 (1) 83-89

Yamaguchi, H. and Matsuishi, T. 2007. Effects of sampling errors on abundance estimates from virtual population analysis for Walleye Pollock in northern waters of seas of japan. Fisheries Science 73, 1061-1069.

APPENDICES

R code used for variance component estimate of chub mackerel

```
Scombdat<-read.table("datafinal.txt", header=T)
Scombdat[1:4,]
id<-paste(Scombdat$year,".",Scombdat$month,".",Scombdat$day, sep="")
Scombdat<-cbind(id,Scombdat)
Scombdat[1:4,]

# Getting rid of extra columns
Scombdat<-Scombdat[,c(1,5,6)]
Scombdat[1:4,]

#Expanding the data
id.2<-rep(Scombdat$id,Scombdat$no)
le.2<-rep(Scombdat$Classint,Scombdat$no)
Scombdat2<-data.frame(id=id.2, le=le.2)
## Variance analysis
fm<-aov(le~factor(id),Scombdat2)
sm<-summary(fm)
nvec<-tapply(Scombdat2$id,Scombdat2$id,length)
n<-(sum(nvec)-sum(nvec^2)/sum(nvec))/(length(nvec)-1)
MS<-sm[[1]][,3]
MSR<-MS[1]
MSE<-MS[2]
sigmahatA<-(MSR-MSE)/n
sigmahatE<-MSE

sigmahatA
sigmahatE

sm

#Analysis of variance components
sigmahatA
sigmahatE
S<-1:100 # No.samples
M<-1:100 # Measurements
var.cont<-NULL
for(i in S){
Vy<-sigmahatA/i + sigmahatE/(i*M)
var.cont<-rbind(var.cont, Vy)
}
var.cont[1:5,1:10]
```

```
plot(c(0,100),c(0,100),xlab="S", ylab="M", type="n")
contour(S,M, var.cont, add=T,levels=c(10,2,1,0.5,0.1,0.05,0.01))

# Length Frequency Distribution Plot of Chub mackerel data
table(Scombdat2$le)
tt<-table(Scombdat2$le)
plot(as.numeric(names(tt)),tt,xlab="Length (cm)", ylab="Frequency", type="l")



### R code for Bootstrapping 200 samples from original Haddock data
proxydat<-read.table("dat.txt",header=T)
kk<-proxydat

# Trimming the old row.names off
 kk<-kk[,2:3]

# Fixing the random number generator, so you can get the same result everytime
 set.seed(7)
# Extracting the individual samples names
 id.all<-unique(kk$id)
 #Getting a random sample, Here I choose 2 samples.
#This you will have to change!!!
id.boot<-sample(id.all,size=200, replace=F)
 id.boot
# Extracting the bootsrapped samples from the main data
 dat.boot<-kk[!is.na(match(kk$id, id.boot)),]

##### Analysis of variance #######
fm<-aov(le~factor(id),dat.boot)
sm<-summary(fm)
nvec<-tapply(dat.boot$id,dat.boot$id,length)
n<-(sum(nvec)-sum(nvec^2)/sum(nvec))/(length(nvec)-1)
MS<-sm[[1]][,3]
MSR<-MS[1]
MSE<-MS[2]
sigmahatA<-(MSR-MSE)/n
sigmahatE<-MSE
sigmahatA
sigmahatE
S<-1:100 # No.samples
M<-1:100 # Measurements
```

```
var.cont<-NULL
for(i in S){
Vy<-sigmahatA/i + sigmahatE/(i*M)
var.cont<-rbind(var.cont, Vy)
}
plot(c(0,100),c(0,100),xlab="S", ylab="M", type="n")

## Bootstrapping 26 samples from 200 Bootstrapped data.
# From script Haddock200samples
# dat.boot
 set.seed(7)
id.all<-unique(dat.boot$id)
id.boot<-sample(id.all,size=26, replace=F)
dat.boot26<-dat.boot[!is.na(match(dat.boot$id, id.boot)),]
length(unique(dat.boot26$id))

fm26<-aov(le~factor(id),dat.boot26)
sm26<-summary(fm26)
nvec26<-tapply(dat.boot26$id,dat.boot26$id,length)
n26<-(sum(nvec26)-sum(nvec26^2)/sum(nvec26))/(length(nvec26)-1)
MS26<-sm26[[1]][,3]
MSR26<-MS26[1]
MSE26<-MS26[2]
sigmahatA26<-(MSR26-MSE26)/n26
sigmahatE26<-MSE26

S<-1:100 # No.samples
M<-1:100 # Measurements
var.cont26<-NULL
for(i in S){
Vy<-sigmahatA26/i + sigmahatE26/(i*M)
var.cont26<-rbind(var.cont26, Vy)
}

par(mfrow=c(2,2))
plot(c(0,100),c(0,100),xlab="S", ylab="M", type="n")
contour(S,M, var.cont, add=T,levels=c(10,2,1,0.5,0.1,0.05,0.25,0.01))
tt<-table(dat.boot$le)
plot(as.numeric(names(tt)),tt,xlab="Length (cm)", ylab="Frequency", type="l")
plot(c(0,100),c(0,100),xlab="S", ylab="M", type="n")
contour(S,M, var.cont26, add=T,levels=c(10,2,1,0.5,0.1,0.05,0.25,0.01))
tt26<-table(dat.boot26$le)
plot(as.numeric(names(tt26)),tt26,xlab="Length (cm)", ylab="Frequency", type="l")
```

```
##### Distribution Plot #####
table(dat.boot$le)
tt<-table(dat.boot$le)
plot(as.numeric(names(tt)),tt,xlab="Length (cm)", ylab="Frequency", type="l")

### Bootstrapping different combinations of sample sizes and numbers and Calculation of
sum of squares
m.id<-unique(dat.boot$id)
all.dat<-table(dat.boot$le)/length(dat.boot$le)
le.vec<-10:95
le.dist<-rep(0,length(le.vec))
names(le.dist)<-le.vec
id.tab<-names(all.dat)
le.dist[id.tab]<-all.dat
all.dat<-le.dist
set.seed(3)
S<-80
M<-30
S.boot<-sample(m.id,size=S,replace=T)
MS.le.boot<-NULL
for(i in S.boot){
        init.dat<-dat.boot$le[dat.boot$id==i]
        M.boot<-sample(init.dat, size=M, replace=F)
        MS.le.boot<-c(MS.le.boot,M.boot)
}
MS.dat<-table(MS.le.boot)/length(MS.le.boot)
MS.dat

le.dist<-rep(0,length(le.vec))
names(le.dist)<-le.vec
id.tab<-names(MS.dat)
le.dist[id.tab]<-MS.dat
MS.dat<-le.dist

SS<-sum((all.dat-MS.dat)^2)

plot(le.vec, all.dat,col="red", xlab="Length(cm)",ylab="Proportion", type="l")
lines(le.vec, MS.dat,col="blue")
```