SCRS/1983/059 Col.Vol.Sci.Pap. ICCAT, 20 (1): 68-74 (1984)



# EFFECTS OF SAMPLE SIZE ON THE ACCURACY OF LENGTH-FREQUENCY SAMPLING OF TUNAS TRANSSHIPPED TO PUERTO RICO

*A. L. Coan, N. W. Bartoo*

SUMMARY

The effects of sample size (number of fish per sample and number of vessels sampled) on the accuracy of length-frequency measurements taken from 1980 Tema-based baitboat catches of yellowfin and bigeye tunas transshipped to Puerto Rico are evaluated using a Bootstrap or Monte Carlo simulation. Results indicate that increasing the number of vessels or vessel trips sampled has a greater effect on the accuracy of length-frequency measurements than increasing the number of fish per sample. Length-frequency samples taken from 1980 transshipments of Tema-based baitboats are accurate (95% confidence interval) to within ±1.4% for yellowfin tuna and ±1.1% for bigeye tuna of the actual percent catch in any one length interval. Estimates of undersized yellowfin and bigeye tuna catches in these transshipments are accurate to within ±2.1% and ±1.4%, respectively. Based on these results, the U.S. sampling in Puerto Rico is adequate for sampling of Tema-based baitboat transshipments.

RESUME

L'influence du volume de l'échantillon (nombre de poissons par échantillon et nombre de bateaux échantillonnés) sur le degré de précision des mensurations de fréquence de taille relevées sur les transbordements à Puerto Rico en 1980 de prises d'albacore et thon obèse effectuées par les canneurs basés à Téma, est évalué au moyen d'une simulation Bootstrap, ou Monte Carlo. Les résultats indiquent que la précision des mensurations de fréquences et de longueur plus sensible à un accroissement du nombre de bateaux ou sorties échantillonnés qu'à celui du nombre de poissons par échantillon. Les échantillons de fréquence de taille prélevés sur les transbordements de 1980 des canneurs basés à Téma sont exacts (intervalle de confiance à 95%) à plus ou moins 14% près pour l'albacore et plus ou moins 1,1% près pour le thon obèse par rapport au pourcentage réel de la prise pour tout intervalle de longueur. Les estimations des prises d'albacore et thon obèse sous taille dans ces transbordements sont exacts à plus ou moins 2,1% et 1,4% près, respectivement. Ces résultats permettent de penser que l'échantillonnage américain à Puerto Rico est adéquat en ce qui concerne les transbordements des canneurs basés à Téma.

RESUMEN

Mediante una simulación Bootstrap o Monte Carlo, se evalúan los efectos del tamaño de la muestra (número de peces por muestra, y número de peces muestreados) respecto a la exactitud de las medidas de frecuencias de talla de rabil y patudo, capturados por barcos de cebo con base en Tema (1980) y transbordados a Puerto Rico. Los resultados indican que incrementar el número de barcos o viajes muestreados, tiene efectos más importantes sobre la exactitud de las medidas de frecuencia de talla que aumentar el número de peces por muestra. Las muestras de frecuencias de talla obtenidas de los transbordos efectuados en 1980, en barcos de cebo con base en Tema, son precisos (95% de intervalo de confianza), entre ± 1.4% para rabil y ± 1.1% para patudo, del actual porcentaje de captura en cualquier intervalo de talla.

Las estimaciones en los transbordos, de las capturas de peces con talla inferior a la reglamentada, se ajustan a cifras comprendidas entre ± 2.1% y ± 1.4% respectivamente. En base a estos resultados, el muestreo norteamericano en Puerto Rico resulta adecuado para muestrear los transbordos de los barcos de cebo con base en Tema.

## INTRODUCTION

There is no simple solution to the question: how big a sample size is needed to adequately sample sizes of fish in a fishery catch? The sample size depends upon the manner in which the data are treated or combined. A sub-sample of X individual vessels participating in a fishery are sampled for length frequency of catches with a sample size of N fish per vessel per trip is a typical scenario. The resulting X samples are combined to produce a single length frequency which is then considered representative of the fishery catch. The number of vessels to sample, X, and the number of fish per sample, N, are, of course, the two parameters which can be varied and will control the adequacy of sampling. The parameters which produce an adequate length sampling for one fishery or species with its peculiar age frequency may not produce an adequate sampling for another species. When choosing sampling parameters, the shape or number of modes in the length frequencies as well as variation in length frequency between catches of vessels must be considered.

Our analysis of the effects of the number of vessels sampled and the number of fish per sample on length frequencies is done specifically for the Tema-based baitboat fishery catches of Eastern Atlantic yellowfin and bigeye tunas transshipped to and sampled in Puerto Rico. Currently, approximately 100 fish per sample are taken from as many of these transshipments as possible. Data from the 1980 fishing season are used in a Bootstrap or Monte Carlo simulation which examines the accuracy of expected composite length frequencies with variations in N and X. Confidence intervals are generated around estimated length distributions and estimated catches of fish less than 55 cm (the regulation minimum size for bigeye and yellowfin tunas in the Atlantic).

Bootstrap simulations of this type are very useful in planning sampling programs. Confidence intervals can be generated by the simulations. These confidence intervals can be used to develop sampling programs, that have limited resources, at a known level of accuracy. Further, because of the intractability of variance estimates associated with multi-modal distributions, simulations offer the only practical approximation of variance estimates.

## PROCEDURE

The 41 yellowfin tuna and 20 bigeye tuna individual length-frequency samples, taken from Tema-based baitboat transshipments during 1980, are assumed to be accurate reflections of the population on board individual vessels or vessel trips. This assumption is valid since it is not our intent to evaluate how well the vessels sampled the population but rather to examine how well these vessels were sampled. Each individual length frequency is weighted by the total catch on each vessel. The final composite length frequency is then a simple summation of all 41 length frequencies for yellowfin and 20 length frequencies for bigeye. We consider this the "actual" composite length frequency for each species.

The two parameters under consideration, X and N, are examined separately: (1) The number of fish, N, in each sample length frequency is varied while the number of vessels, X, is fixed at several values. (2) The number of vessels sampled, X, is varied while N is fixed at several values.

For various combinations of X and N, fish are randomly "sampled" with replacement from each vessel. The X vessels to be sampled are also chosen randomly with replacement. Length frequencies are weighted by the total catch of each vessel and combined over vessels to give a single composite length frequency with sample sizes X and N. This single simulated sample can be compared to the "actual" composite length frequency. However, such a comparison only gives one observation of the expected deviation from the "actual" composite length frequency. It is more useful to observe the expected distribution of a large number of simulated length frequencies. This distribution is the estimated probability distribution for any single length frequency simulated. Further, approximate confidence intervals (C.I.) for length frequencies and for estimates of undersize fish in the catches may be taken from the distributions. For each combination of X and N, 1000 simulations are used to generate each probability distribution.

RESULTS AND CONCLUSIONS

An example of the expected probability distributions, each the result of 1,000 simulated sampling experiments with N=100, X=41 is shown for yellowfin tuna (Figure 1) and N=100, X=20 for bigeye tuna (Figure 2). Distributions are generated with the number of vessels sampled, X, varied between 20, 40, 80 and 120 vessels and fixed at 41 for yellowfin tuna and 20 for Bigeye tuna. The number of fish per sample, N, for both yellowfin and bigeye is varied in each distribution between 50, 100, 150 or 200. Comparison of the distributions shows that the spread or range of the distribution at any particular length interval generally becomes narrower or tighter as the number of fish in each sample increases. Note also the greater spread in the distributions at certain peaks in the length frequency, such as at 49 cm for yellowfin tuna and 50 cm for bigeye tuna.

When evaluating, through Monte Carlo simulations, the effects of sample sizes on length-frequency sampling, it is useful to concentrate on a particular length interval where the distribution of expected results is greatest. This interval has the greatest uncertainty. As the spread of the distribution (approximate 100% C.I.) is reduced at this widest location, the spread of the distribution will improve at all locations or lengths, and the accuracy of the sampling at all length intervals will improve. In this analysis the length interval at 49 cm for yellowfin tuna, and 50 cm for bigeye tuna is judged to show the greatest spread in the expected distribution. This is best seen in Figure 1 for yellowfin tuna and Figure 2 for bigeye tuna.

Figures 3A and 3B demonstrate the effect of different numbers of fish per sample on the range of the approximate 95 % C.I. around the point estimate of the mean percent catch at 49 cm for yellowfin tuna and 50 cm for bigeye tuna. Several different levels of number of vessels sampled are shown in the figures. For both species, increasing the number of fish per sample with any given number of vessels sampled usually reduces the spread of the distribution. However, in some cases increasing the number of fish per sample increased the spread. The greatest reduction in the range of the 95% C.I. produced by increasing the sample size from 50 fish to 200 fish is 1.0% for yellowfin tuna and 0.5% for bigeye tuna.

Figures 4A and 4B demonstrate the effect of varying numbers of vessels sampled on the range of the 95% C.I. around the point estimate of the mean percent catch at 49 cm for yellowfin tuna and 50 cm for bigeye tuna. Several levels of sample sizes are shown. As the number of vessels sampled increases, the range of the 95% C.I. decreases for both species.

The effect of number of fish per sample is relatively slight compared to the effect of number of vessels sampled. This is because the variance of fish sizes between vessels is greater than the variance within vessels. The greatest reduction in the 95% C.I. produced by increasing the number of vessels sampled from 20 to 120 is 4.7% for yellowfin tuna and 4.1% for bigeye tuna.

Lastly, the effects of increasing sample size and number of vessels sampled on the resulting estimates of undersized fish is evaluated for both species (Figures 5A and 5B). Again, increasing sample size has less of an effect than increasing number of vessels sampled. Increasing sample sizes from 50 fish to 200 fish will produce only a 3.0% reduction in the range of the approximate 95% C.I. around the point estimate of the undersized yellowfin tuna catch; a 2.0% decrease in the 95% C.I. around the estimated undersized bigeye tuna catch. Increasing the number of vessels sampled from 20 to 120 produces a 30% reduction in the 95% C.I. around the point estimate of the undersized yellowfin catch; a 26% reduction in the 95% C.I. for bigeye tuna.

Thus, at least in this example, increasing the number of vessels or vessel trips sampled, X, has the most effect on improving the accuracy of the resulting length-frequency distribution and estimates of undersized fish. Thus, the best sampling strategy, for yellowfin and bigeye transshipments of Tema-based baitboat catches to Puerto Rico, is to keep the number of fish sampled at the current limit of approximately 100 and to sample as many individual vessel transshipments as possible.

The accuracy of the length frequencies, obtained from the 1980 sampling of yellowfin and bigeye tuna transshipments of Tema-based baitboats in Puerto Rico, can be assessed by running the simulation on the actual 41 yellowfin tuna samples and 20 bigeye samples (X=41 and X=20 sampled with no replacement, N=100 sampled with replacement). The results show that the length frequencies are accurate (95% C.I.) to within $\pm1.4\%$ of the actual percent catch in any one length interval for yellowfin tuna and $\pm1.1\%$ for bigeye tuna. The estimated percent of undersized fish obtained from these length frequencies are accurate (95% C.I.) to within $\pm2.1\%$ for yellowfin tuna and $\pm1.4\%$ for bigeye tuna.

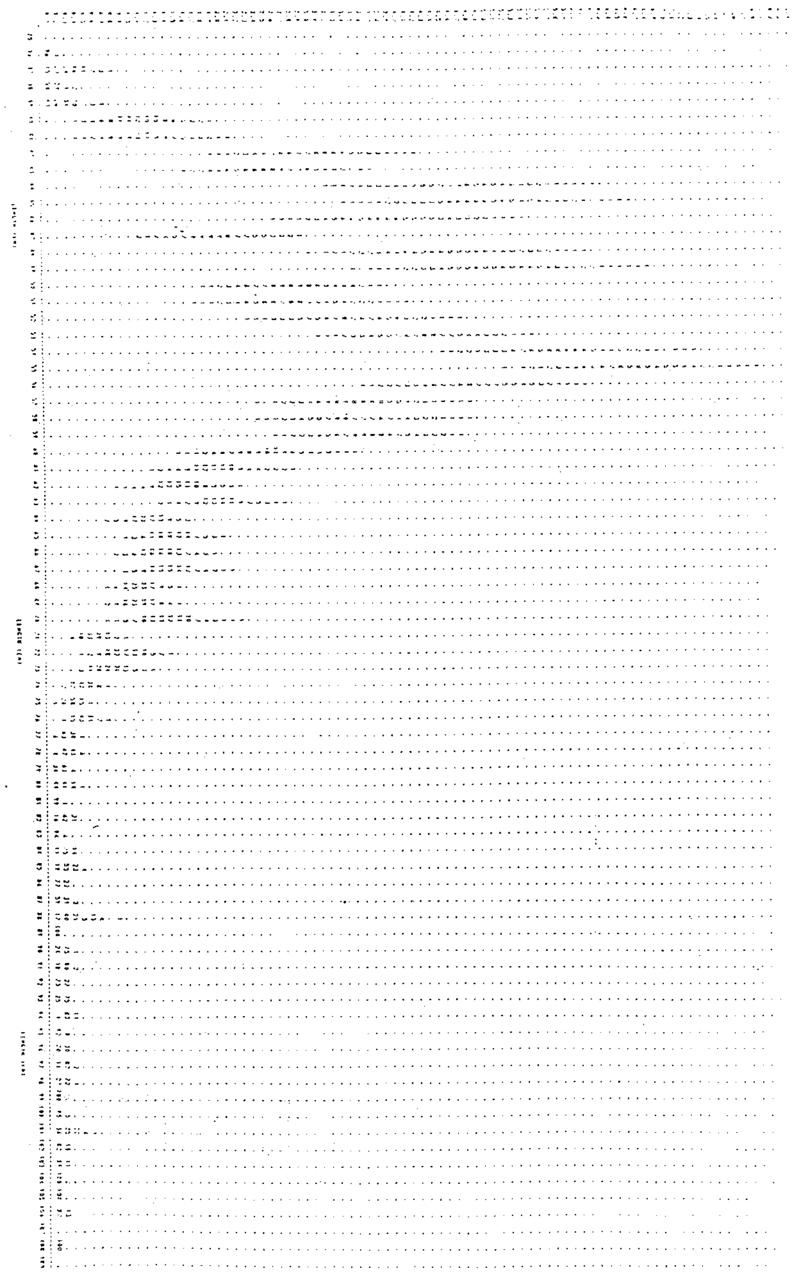

Figure 1. Distribution of 1,000 yellowfin tuna length-frequency distributions. Each distribution is the result of sampling 41 Tema-based baitboat vessel transshipments with a sample size of 100 fish per vessel.

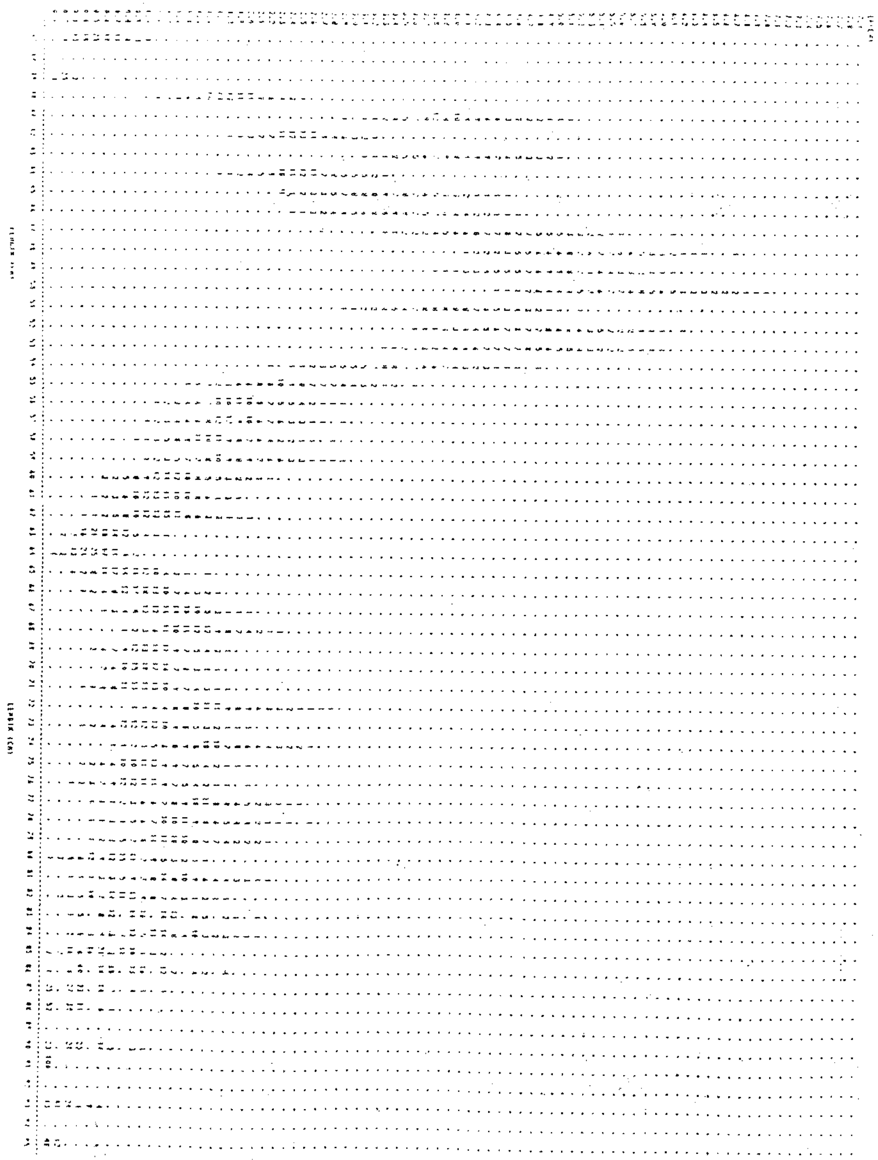

Figure 2. Distribution of 1,000 bigeye tuna length-frequency distributions. Each distribution is the result of sampling 20 Tema-based baitboat vessel transshipments with a sample size of 100 fish per sample.

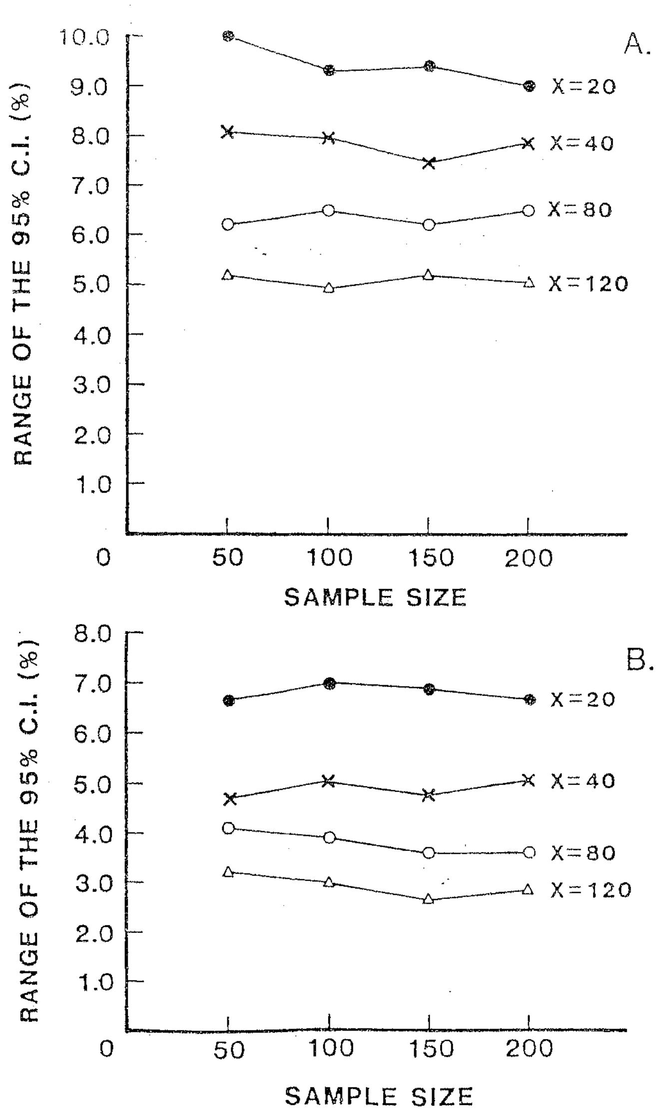


Figure 3. Changes in the range of the 95% confidence interval (C.I.) around the point estimate of the mean percent catch in the (A) 49 cm length interval for yellowfin tuna and the (B) 50 cm length interval for bigeye tuna for various sample sizes and vessels sampled, X.

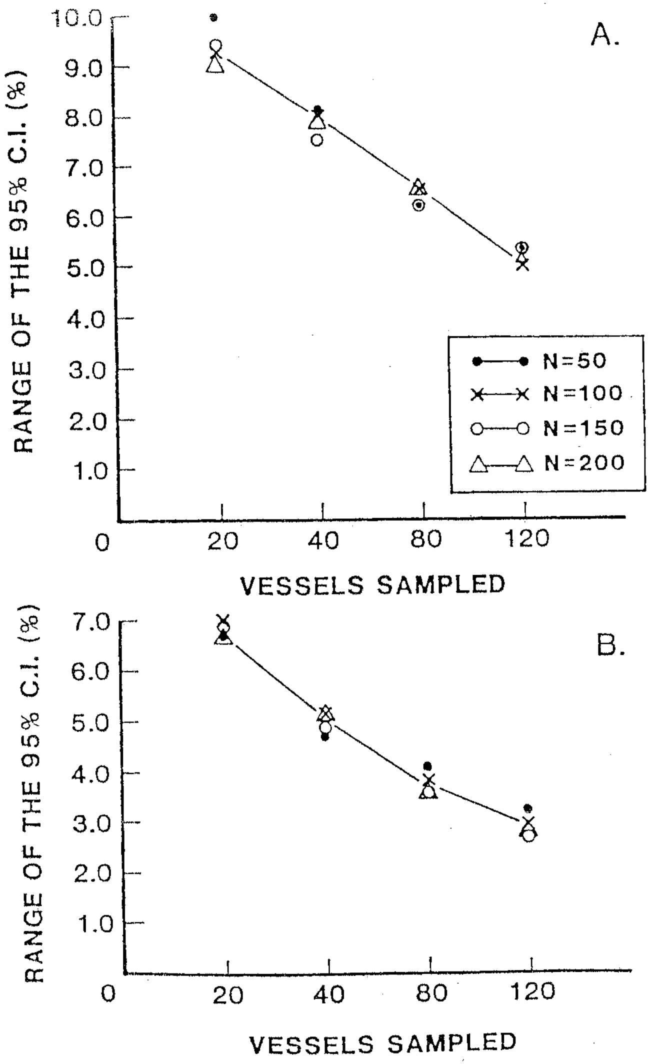


Figure 4. Changes in the range of the 95% confidence interval (C.I.) around the point estimate of the mean percent catch in the (A) 49 cm length interval for yellowfin tuna and the (B) 50 cm length interval for bigeye tuna for various numbers of vessels sampled and sample sizes, N.

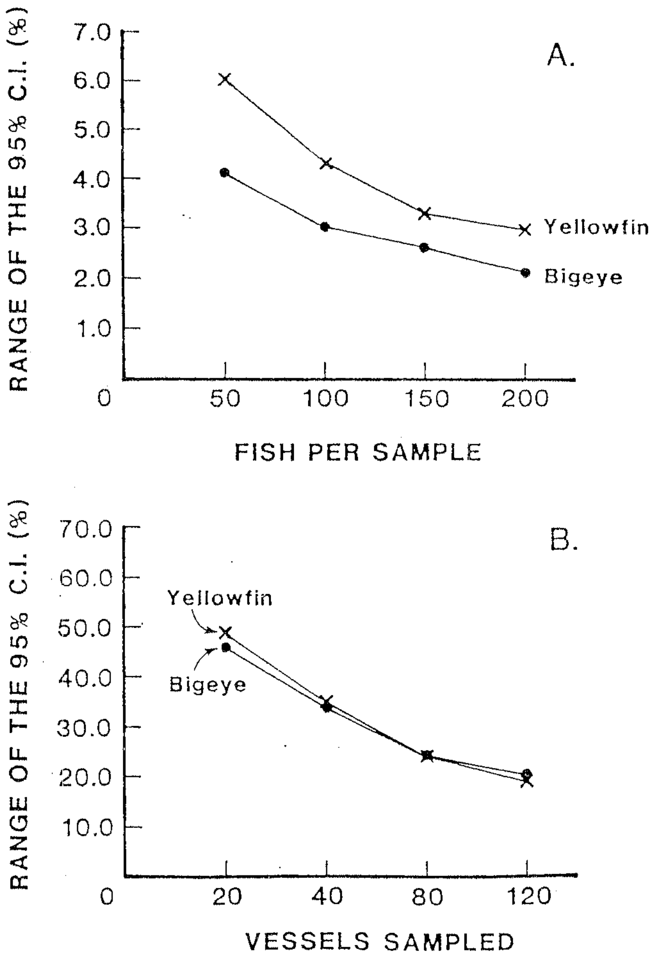


Figure 5. A) Changes in the range of the 95% confidence interval (C.I.) around the point estimate of the percent undersized yellowfin and bigeye tuna at various sample sizes and number of vessels sampled fixed at 41 for yellowfin tuna and 20 for bigeye tuna. B) Changes in the range of the 95% C.I. around the point estimate of the percent undersized yellowfin and bigeye tuna at various numbers of vessels sampled and sample size fixed at 100.