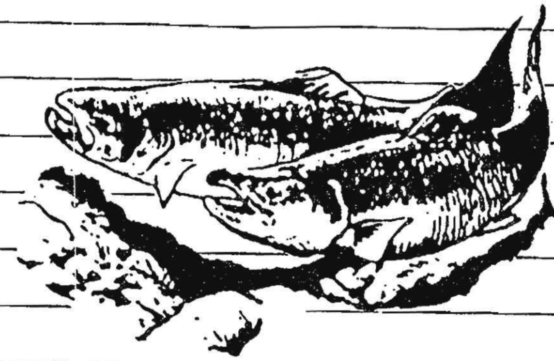
Research
Information bulletin
U.S. DEPARTMENT OF THE INTERIOR
NATIONAL BIOLOGICAL SURVEY
NO. 24
DATE 1994



How Many Fish Should Be Used in a Composite Sample?

In this bulletin, we provide guidelines for determining sample sizes necessary to work with composite samples. The need to form composite samples is sometimes associated with the high cost of analytical measurements, as in the analysis of contaminant concentrations in fish tissue. In these situations, researchers form composite samples to obtain a relatively low-cost estimate of the average measure of contaminant concentration. One of the following questions invariably arises:

1. If we want to obtain reliable estimates of the mean and variance, and we can analyze only X samples, how many fish should we collect to form those X composites? or

2. If we collect Y individual fish, how many composites should be formed, and how many fish should be combined to form each composite sample?

To answer these questions, we used a simulation approach and examined the tradeoffs in precision of the estimates of mean and variance. Our base population consisted of 195 striped bass (*Morone saxatilis*) for which we knew individual contaminant concentrations. We simulated the formation of composite samples using bootstrap sampling, a common technique for statistical analyses.

Form Composite Samples in the Correct Manner

In a previous bulletin (RIB No. 23), we demonstrated the correct technique of forming composite samples: each fish is homogenized and equal-sized subsamples from each homogenate are combined to form a single composite sample. The contaminant concentration in the composite estimates the average of the concentrations in the individual fish.

For this study, we used the bootstrap technique to randomly sample fish from the base population and calculated contaminant concentrations for each simulated composite sample. We varied either the number of fish collected while holding the number of composites fixed, or we varied the number of composites while holding the number of fish collected constant. We repeated each combination 1,000 times. We then examined the mean contaminant concentration and the variance of the variance, a measure of the relative precision of the variance. The true mean was estimated directly from the base population and was helpful in interpreting our results.

Collect More Fish With a Fixed Number of Samples

In the first example, we attempted to answer the question, If an investigator can process only

10 analytical samples, how many fish should be collected to form the 10 composites? To answer this question, we varied the number of fish in each composite from 1 to 50. Note that the product of the number of composite samples (10) and the number of fish per composite (1–50) is equal to the number of fish collected (10–500). We then estimated the mean and variance for collections of 10–500 fish, corresponding to composite samples of 1–50 fish each.

The true mean concentration is estimated equally well by all composite sizes (Fig. 1a), but the precision of the variance estimate is increased as more fish are collected (Fig. 1b). Ten composite samples, each formed from many fish, will yield better estimates of the variance and hence higher precision than 10 composites formed from few fish. Based on Figure 1b, and given 10 composites, we estimate that about 100 to 200 fish should be collected to obtain a reliable estimate of the variance for our striped bass example.

More Composite Samples With a Fixed Number of Fish

In the second example, we addressed the question, Given 100 fish, how many composite samples can be formed without losing reliability? We simulated the formation of 2–100 composite samples from 100 fish, thereby varying the number of fish per composite between 1 and 50. Because we worked in whole fish units, we allowed n, the number of fish collected, to equal the product of the number of composites and the number of fish per composite; this implied that n was close to 100, but sometimes as low as 96 (for example, 12 composites, each formed from 8 fish, require 96 fish).

An examination of the means estimated from different numbers of composite samples in relation to the true mean gave no clear indication that any particular case we simulated was preferred (Fig. 2a). This is expected because n is fixed. However, as the number of composite samples approaches the number of fish collected (i.e., a composite is formed from a single fish), we expect the average variance to approach the true variance and its precision to increase. The precision of the variance estimate was clearly related to the number of composite samples in an exponential manner (Fig. 2b). This indicated that given a collection of 100 individuals, the variance estimate stabilizes between 10 and 20 composite samples. For these data, we recommend forming composites with no more than 5–10 fish, when a collection of 100 fish is available for analysis.

An Aid in Determining Adequate Sample Sizes

As we demonstrated in our first example, if the purpose is to estimate a mean without regard to its precision, then a few composites of a few fish each could be formed. For a fixed number of composite samples, more fish per composite are necessary when a reliable estimate of the variance is also required. Our second example clearly demonstrates that the best composite sample is formed of a single individual—but this is no different from directly sampling the population and analyzing each sample separately. We realize this may not be practical when costs are considered. In that case, we recommend (1) determining the number of individuals that can be sampled for a fixed cost, and (2) using the approach in example 2 to estimate the precision of the variance for composites formed of varying numbers of fish. Of course, this requires knowing something (namely, the mean and variance) of the population one wishes to sample. These estimates are sometimes available in the literature. Our approach should provide researchers with an objective method for determining the necessary sample size when forming composite samples.

For further information contact:

Mary C. Fabrizio, Anthony M.Frank, or
Jacqueline F. Savino
Great Lakes Center
1451 Green Road
Ann Arbor, MI 48105
(313) 994-3331

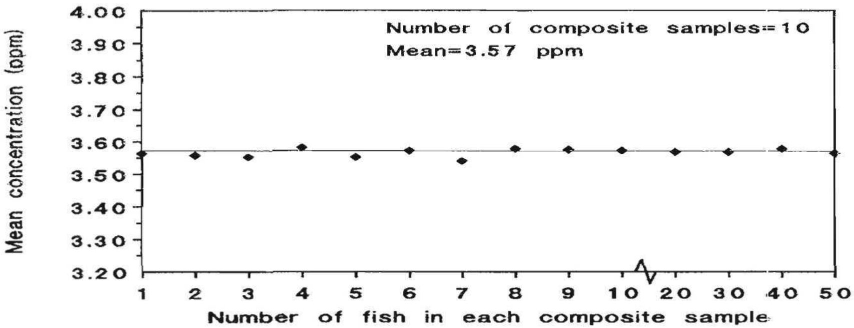


Fig. 1a. Mean contaminant concentration vs. number of fish per composite.

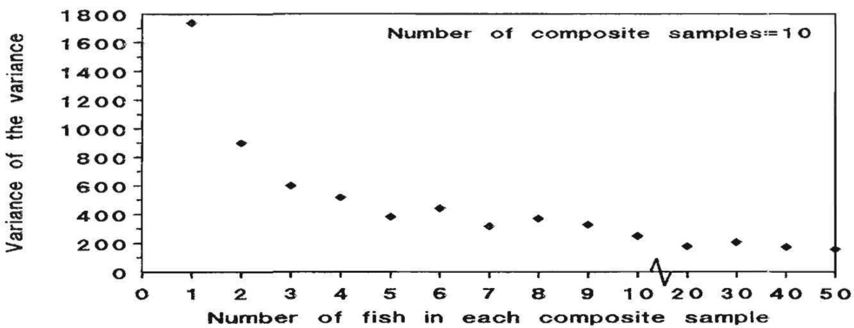


Fig. 1b. Precision of the variance vs. number of fish per composite.

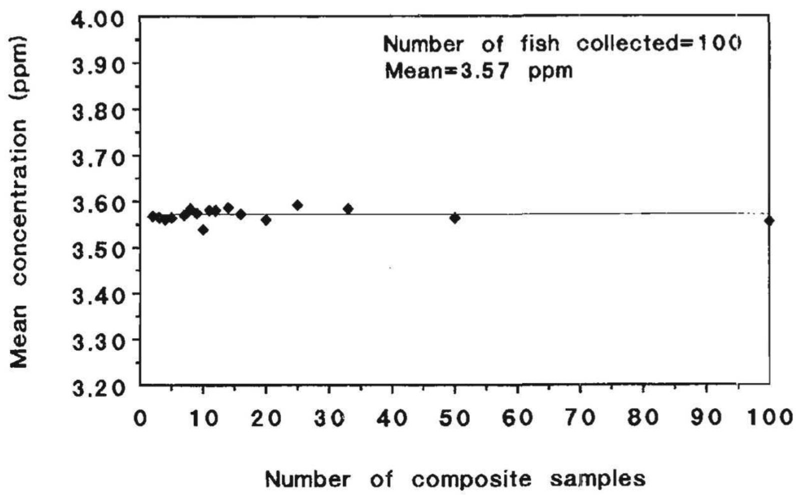


Fig. 2a. Mean contaminant concentration vs. number of composite samples.

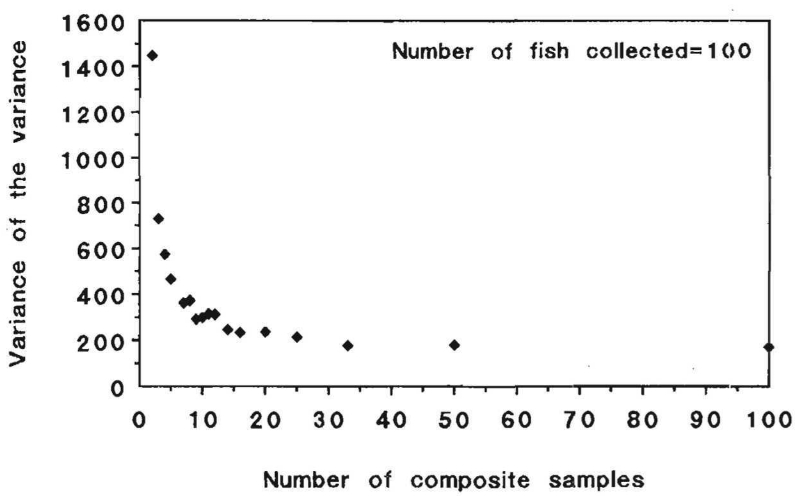


Fig. 2b. Precision of the variance vs. number of composite samples.