Bull. Inst. Zool., Academia Sinica 14(2): 61-70 (1975)

TWO-STAGE SAMPLING IN STOMACH ANALYSIS OF FISH

YAO-SUNG LIN

Department of Zoology National Taiwan University Taipei, Taiwan 107 Republic of China

Received for publication, Nov. 4, 1975

ABSTRACT

Yao-Sung Lin (1975). Two-Stage Sampling in Stomach Analysis of Fish. Bull. Inst. Zool., Academia Sinica, 14(2): 61-70. A comparison of precision was made in the estimation processes of mean number of zooplankton per fish stomach with simple random sampling and two-stage sampling. Formulae are developed for the variance of the estimated mean based on the enumeration of subsamples. The variance equation shows that the precision of the mean depends on the number of fish in the samples and also the number of organisms counted. Given a moderate sample of fish, two-stage sampling of stomach contents will cost about half the time required for simple random sampling. Under the same cost, the two-stage sampling would increase the precision from simple random sampling when estimating the amount of zooplankton in fish stomach.

Abundance of food organisms in stomachs of fish is generally estimated by examining a representative sample of the fish population. A point estimator based on such a sample is subject to sampling error due to variability in number of organisms consumed by individual fish. In a simple random sampling where stomachs of individual fish are examined, the variance will be inversely related to sample size. Although precision in point estimation may require processing large numbers of fish, since variation in the number of organisms in stomachs of some fish can be rather great, however, a form of two-stage sampling might increase efficiency.

and the second

In two-stage sampling a subsample of organisms from pooled stomach contents of individual fish are counted. Pooling reduces the time required to count organisms but the subsampling procedure adds additional sampling errors beyond that caused by the non-uniform distribution of stomach contents in fish of field samples. The present study evaluates the reduction in processing time and the magnitude of the error in two-stage sampling compared to simple random sampling.

17

 \cdot \cdot

MATERIALS AND METHODS

Young-of-the-year yellow perch, *Perca flavescens* (Mitchill), were caught in an 18-ft bottom trawl on five dates in July-September 1973 near Shackelton Point, Oneida Lake. A random sample of 80 fish from one haul on each date was preserved in 10% formalin and brought to the laboratory. Generally four groups of 20 fish each were randomly selected for sequential analysis. Stomachs of individual fish were removed and all zooplanktons counted (*Daphnia pulex* predominated). After counting, the stomach contents were flushed into a container until all 20 fish had been examined. The process was

. . .

61

repeated with successive groups of 20 fish until stomachs of all 80 fish were analyzed. The stomach contents in the four containers were than diluted by a volume of water and became the pooled or primary sampling units. From each of these primary units three 1 ml subsamples or secondary units were drawn with replacement with a Hensen-Stempel pipet and the number of organisms were counted in a Sedgewick-Rafter chamber. Number of organisms in a secondary unit was multiplied by a conversion factor V/F (where V is the volume of the primary unit and F is the number of fish in the primary unit) to estimate the mean number of organisms per fish stomach in the primary unit. An identical subsampling and estimation procedure was followed if the originl sample of 80 fish was pooled into two or one (rather than 4) primary units of 40 or 80 (rather than 20) fish.

In two-stage sampling, the primary unit is the well-mixed dilution of stomach contents of fish. While the secondary unit (subsample) is defined as a sample of the primary unit. The volume of the secondary unit was always 1 ml and that of the primary unit was dependent upon the abundance of organisms in stomachs and the number of fish in the primary unit. In the present study, pooled stomach contents were diluted so that count of organisms in the secondary unit was about 40 to 80 organisms except for the July 31 sample in which the count per ml subsample averaged slightly less than 20 organisms. Fish taken on July 31 contained few zooplanktons, therefore it was impossible to obtain a count of more than 40 organisms without reducing the volume of the primary unit to a rather small amount. Such small volume of the primary unit would cause difficulty in subsampling with the Hensen-Stempel pipet.

For convenience in later discussion, the author defined the three types of two-stage sampling which varies in number of primary unit as type 1, 2 and 3. The number of fish in each primary unit of type 1, 2 and 3 was 80, 40 and 20 respectively. The diluted volume of pooled stomach contents in the primary unit was proportional to the number of fish in the unit. As a result, the mean density per primary unit was expected to be the same in the three types of two-stage sampling for the same 80 fish.

Time spent on each phase of the procedure (dissecting, pooling and counting) was monitored to determine the efficiency in terms of cost and variance of two-stage sampling over simple random sampling.

STATISTICAL CONSIDERATION

Variance derivations

In simple random sampling, all organisms in the stomach of individual fish are counted and the variance of the mean is

$$V(\bar{X}_{ran}) = \frac{S^2}{N} \tag{1}$$

where

N=number of fish in the sample S^2 =variance of stomach contents

In two-stage sampling, the estimated overall

mean number of organisms is

$$\bar{\bar{y}} = \frac{\sum_{n=1}^{n} y}{nm} = \frac{\sum_{n=1}^{n} \bar{y}}{n}$$

where

- n = number of primary unit
- m=number of subsamples per primary unit
- y=number of organisms per subsample
- \bar{y} = mean number of organisms per subsample
- \bar{y} =mean number of organisms per primary unit

The estimated variance of \overline{y} assuming random sampling is⁽²⁾

$$V(\bar{y}) = \frac{S_b^2}{n} + \frac{S_w^2}{nm}$$

where

 S_b^2 = variance among primary unit means

 S_w^2 = variance among subsamples within primary units

The estimate of the mean number of organisms per fish is given by

$$\overline{\overline{X}}_{tss} = \left(\frac{V}{F}\right) \overline{\overline{y}}$$

The calculated variance of $\overline{\overline{X}}_{tss}$ is accordingly

$$V(\overline{\overline{X}}_{tss}) = \left(\frac{V}{F}\right)^2 V(\overline{\overline{y}})$$
$$= \left(\frac{V}{F}\right)^2 \left(\frac{S_b^2}{n} + \frac{S_w^2}{nm}\right)$$
$$= \frac{\sum_{i=1}^{n} \left(\frac{V}{F}\right)^2 (\overline{y} - \overline{\overline{y}})^2}{n(n-1)} + \left(\frac{V}{F}\right)^2 \frac{S_w^2}{nm} \quad (2)$$

Since the mean number of organisms per m1 (\bar{y}) multiplied by the conversion factor (V/F) is an estimate of mean number of organisms per stomach of fish in the primary unit (\bar{X}) , and the overall mean number of organisms in all primary units (\bar{y}) , multiplied by (V/F) is an estimate of the mean number of organisms per fish stomach in all the primary units (\bar{X}) , equation (2) could be rewritten as:

$$V(\overline{\overline{X}}_{tss}) = \frac{\sum_{i=1}^{n} (\overline{X} - \overline{\overline{X}})^{2}}{n(n-1)} + \left(\frac{V}{F}\right)^{2} \frac{S_{w}^{2}}{nm}$$
$$= \frac{S_{\overline{X}}^{2}}{n} + \left(\frac{V}{F}\right)^{2} \frac{S_{w}^{2}}{nm} \qquad (3)$$

where

2014

ŝ

 \overline{X} =mean number of organisms per fish in the primary unit

 \overline{X} = mean number of organisms per fish in the sample

 $S_{\bar{X}}^2$ = variance of the mean number of organisms among the primary units

Miura and Naka⁽⁵⁾ demonstrated that the

number of several zooplanktons in subsamples of 0.2 ml taken from a volume of 20 ml was Poisson distributed. To determine if the number of D. *pulex* in subsamples was distributed similarly, the author prepared six 40 ml samples in which the density of D. *pulex* varied from 2-64 per ml. On the basis of 10 counts of 1 ml subsamples from each of the six samples, the sample mean and variances of the number organisms per unit volume were calculated.

Mean number of *D. pulex* per ml and the variance of counts increased proportionally as anticipated in sampling from a Poisson distribution (Table 1). Counts were tested for departure from a Poisson by Chi-square in the following manner⁽⁷⁾

$$X_{m-1}^2 = \frac{S_1^2(m-1)}{\overline{y}} = \frac{\sum_{j=1}^m (y-\overline{y})^2}{\overline{y}}$$

where

 \overline{y} =mean number of organisms per subsample S_1^2 =variance of organisms among the sub-

samples

m=number of subsamples

The pooled X^2 is computed within each sample and is distributed with 54 degrees of freedom (Table 1). The X^2 of 53.64, insignificant at the 0.05% level confirms that counts from subsamples follows a Poisson distribution.

If it is assumed that the fish in the primary unit are a random sample of F drawn from the

TABLE 1

Mean and variance in a series of *D. pulex* and Chi-square test of departure from Poisson distribution for each sampling series of known⁺concentration

| Statistic | | | | | | |
|--------------------------------------------------|------|------|------|-------|-------|------|
| | 2 | 4 | 8 | 16 | 32 | 64 |
| Number of subsamples (m) | 10 | 10 | 10 | 10 | 10 | 10 |
| Mean number per subsample (y) | 2.6 | 3.4 | 9.0 | 15.7 | 30.2 | 62.1 |
| Variance (S ²) | 1.56 | 2.71 | 7.56 | 20.67 | 39.51 | 67.9 |
| ∦2 ,2,2,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1 | 5.4 | 7.20 | 7.56 | 11.88 | 11.79 | 9.81 |

Pooled Chi-square=53.64 (d. f.=.54)

population, then the variance $(S_{\overline{x}}^2)$ of the mean number of organisms per fish in the primary unit would be

$$S_{\bar{X}}^2 = \frac{S_{\bar{X}}^2}{F}$$

Substituting this term into equation (3) yields the following variances for the two-stage sampling scheme:

$$V(\overline{\overline{X}}_{tss}) = \frac{S^2}{nF} + \left(\frac{V}{F}\right)^2 \frac{S_w^2}{nm}$$
$$= \frac{S^2}{M} + \left(\frac{V}{F}\right)^2 \frac{S_w^2}{nm} \tag{4}$$

where M=total number of fish in the sample.

Equation (4) shows that the first variance component depends on the total number of fish in the sample and the second variance component depends on the total number of subsamples, the product of nm. If we let K=nm, then equation (4) would be written as

$$V(\bar{\bar{X}}_{tss}) = \frac{S^2}{M} + \frac{1}{K} \left(\frac{V}{F}\right)^2 S_w^2$$
 (5)

If individual organisms are scattered throughout a large sampled volume, statistical theory would predict that the number of organisms per unit of volume would follow a Poisson distribution.

Since a characteristic of the Poisson distribution is that the variance equals the mean, the variance (S_{2}^{*}) of the subsample within the primary unit could be approximated by the mean number of organisms per subsample (\bar{y}) . Thus equation (5) may be approximated by

$$V(\overline{\overline{X}}_{tss}) = \frac{S^2}{M} + \frac{1}{K} \left(\frac{V}{F}\right)^2 \overline{y}$$
 (6)

Since $\overline{X} = \overline{y}(V/F)$, substituting this into equation (6) gives:

$$V(\overline{\overline{X}}_{tss}) = \frac{S^2}{M} + \frac{1}{K} \left(\frac{V}{F}\right) \overline{X}$$
(7)

$$=\frac{S^2}{M}+\frac{X^2}{K\overline{y}}$$
(8)

The above equation shows that the overall variance of the mean in two-stage sampling has two-components. The first component (S^2/M) takes the same expression as the variance of

simple random sampling (S^2/N) , the size of which depends on the population variance (S^2) and is inversely proportional to the number of fish in the sample. If the number of fish sampled was fixed for both sampling schemes, it is conceivable that the two-stage sampling would be less precise since only a fraction of the total organisms would be counted. Given a fixed cost, the number of fish that can be examined in two-stage sampling (M) will be larger than that of simple random sampling (N). Hence in this case it is clear that the first variance component of twostage sampling will be smaller than the variance of the mean of simple random sampling. The difference $(S^2/N - S^2/M)$ could be considered as the profit of two-stage sampling. On the other hand, the second variance component, which results from incomplete counting of organisms in stomachs and does not exist in simple random sampling, could be considered as the loss of twostage sampling. Finally, the net gain in precision from two-stage sampling depends simply on the difference between profit and loss.

The values of S^2 and \overline{X} depend on the fish population, whereas the product of $K\overline{y}$ is the expected number of organisms counted in twostage sampling. Hence, equation (8) shows that the precision of the estimate depends on the number of fish in the samples and also the number of organisms counted. When the total number of organisms counted is given, the number of primary units and the total number of subsamples have no effect on the precision of the two-stage sampling.

For a fixed sample size M and K, it is apparent that the estimated variance could be reduced simply by increasing the density of organisms (\overline{y}) in the primary unit. Since time spent on counting a higher density of organisms in the subsample is rather short, increasing the density of organisms in the primary unit is a very efficient method of increasing the precision of two-stage sampling. The density of the primary unit, however, must not be so great as to violate the assumption that the number of organisms in the subsample is Poisson distributed.

Optimum allocation

The size of M and K can be chosen arbitraily by the sampler. They may be allocated to minimize the variance, $V(\overline{X}_{tss})$ for a specified cost of taking the sample M and K or alternately to minimize the cost for a specific value of $V(\overline{X_{tss}})$.

The best allocation of samples to stages is⁽⁴⁾

$$\frac{M}{K} = \frac{S}{\overline{X}} \sqrt{\frac{t_2}{t_1} \overline{y}} = CV \sqrt{\frac{t_2}{t_1} \overline{y}}$$
(9)

where

- CV = coefficient of variation of the number of organisms per fish.
 - t_1 = time required to dissect out a fish stomach content and empty contents into a container.
 - t_2 = time required to take subsamples and count food organisms.

Equation (9) suggests two rules for conducting two-stage sampling. For a given cost or a

specific variance, one should take a large sample of fish if variation of stomach contents is large or if time required to dissect out stomachs and pool contents is brief.

RESULTS

Comparison between mean number of Daphnia pulex and their variance in stomachs of 80 fish indicated that the value of the mean was much less than that of the variance (Table 2). However, a significant linear relationship between log of mean and log of variance was found. The relationship can be described by the general equation $S^2 = 8.531 \overline{X}^{1.541}$; r (correlation coefficient) is 0.934 (p < 0.05). This equation implies that the variance of Daphnia pulex in fish stomachs increased at a greater rate than that of the mean which is characteristic of overdispersion.

The average time required to examine an

| I AB | LE 2 | |
|-------|------|--|
| • | | |

Mean, variance and coefficient of variation of D. pulex in stomachs of eighty young yellow perch collected from Oneida Lake, 1973

| Date | Mean number per fish | Variance | Coefficient of variation |
|--------------|----------------------|----------|--------------------------|
| July 31 | 59.9 | 4289.16 | 1.09 |
| August 16 | 415.9 | 52881.97 | 0.55 |
| August 31 | 283.2 | 62311.40 | 0.88 |
| September 5 | 114.5 | 10862.30 | 0.91 |
| September 19 | 251.8 | 81451.79 | 0.13 |

TABLE 3

Average time (seconds) required to examine individual fish stomachs (t_0) , remove and empty stomach contents (t_1) , subsample and count food organisms (t_2) , and to dilute pooled stomach contents to desired volume (t_3)

| Data | | | | | |
|----------|----------------|-----------------------|----------------|----|--|
| Date | t ₀ | <i>t</i> ₁ | t ₂ | 18 | |
| 31 July | 267 | 167 | 95 | 70 | |
| 16 Aug. | 611 | 270 | 175 | 70 | |
| 31 Aug. | 423 | 183 | 120 | 70 | |
| 5 Sept. | 292 | 155 | 130 | 70 | |
| 19 Sept. | 474 | 180 | 160 | 70 | |

individual fish in simple random sampling ranged from about 4.5 to 10 minutes, and the total cost in processing 80 fish with 3, 6 or 12 subsampling units in two-stage sampling ranged from 3.9 to 6.4 hours (Table 3). The cost varied with stomach contents and size of fish. If the same time spent on two-stage sampling was allotted for simple random sampling, only 32 to 55 fish (N) could have been examined. The average value is 42 fish, which is about half the number of fish processed in two-stage sampling. Therefore, the main advantage for conducting twostage sampling is to obtain a larger, more representative sample of the fish population which would increase the precision of the estimator as compared to that of simple random sampling.

In two-stage sampling, the first variance component depends on the total number of fish in the sample, rather than on the number of primary units and number of fish per unit initially drawn from the population. The second variance component depends on the product of $\overline{y}K$ or the total number of subsamples for a fixed concentration of the primary unit. Hence, to distinguish three variations (types) of the twostage sampling scheme, only the total number of fish and the number of subsamples were listed (Table 4).

Comparison of the three types of two-stage sampling indicated that relative precision increased from type 1 to type 3. This increase was probably related to the allocation ratio of M to K. For example to process the sample of fish collected on August 16 type 1 two-stage sampling required 354 minutes (Table 4). Given the same cost, 36 fish could have been examined by the simple random sampling scheme. The profit $(S^2/36 - S^2/80) = 807.92$ from two-stage

TABLE 4

| Expected variance (| of simple random | sampling and | twostage | sampling for a |
|---------------------|-------------------|--------------|----------|----------------|
| given cost | which varied with | the types of | sampling | scheme |

| | | | Two-stage sampling* | | | Simple random sampling | | Relative | |
|------|-------|------|---------------------|----|---------------|---------------------------|-------------|-----------|--|
| Date | Cost | Туре | | K | S^{s}_{tss} | М | S_{ran}^2 | precision | |
| 7/31 | 13715 | 1 | 18.4 | 3 | 118.61 | 51 | 84.10 | 0.71 | |
| | 14070 | 2 | 18.4 | 6 | 86.11 | 53 | 80.93 | 0.94 | |
| | 14780 | 3 | 18.4 | 12 | 69.86 | 55 | 77.98 | 1.12 | |
| 8/16 | 22195 | 1 | 69.3 | 3 | 1563.33 | 36 | 1468.94 | 0.94 | |
| | 22790 | 2 | 69.3 | 6 | 1112.18 | 37 | 1429.24 | 1.29 | |
| | 23980 | 3 | 69.3 | 12 | 886.60 | 39 | 1355.95 | 1.53 | |
| 8/31 | 15070 | 1 | 47.2 | 3 | 1392.06 | 36 | 1730.87 | 1.24 | |
| | 15500 | 2 | 47.2 | 6 | 1085.48 | 37 | 1684.09 | 1.71 | |
| | 16360 | 3 | 47.2 | 12 | 932.18 | 39 | 1597.33 | 1.71 | |
| 9/ 5 | 12860 | 3 | 45.6 | 3 | 231.61 | 44 | 246.87 | 1.07 | |
| | 13320 | 6 | 45.6 | 6 | 183.69 | 46 | 236.14 | 1.29 | |
| | 14240 | 12 | 45.6 | 12 | 159.74 | 49 | 221.68 | 1.39 | |
| 9/19 | 14950 | 3 | 50.4 | 3 | 1437.48 | 32 | 2545.37 | 1.77 | |
| | 15500 | 6 | 50.4 | 6 | 1227.81 | 33 | 2468.24 | 2.01 | |
| | 16600 | 12 | 50.4 | 12 | 1122.98 | 35 | 2327.18 | 2.07 | |

*: number of fish in two-stage sampling is 80

K=number of subsamples examined

M=number of fish that could have been examined in simple random sampling

 $S_{tss}^2 =$ variance of two-stage sampling

 S_{ran}^2 = variance of simple random sampling

y=mean number of organisms per subsample

sampling is less than the loss $(\overline{X}^2/K\overline{y}) = 902.3$ caused by subsampling error. Hence, two-stage sampling is less precise. In type 2, doubling the number of subsamples cost only 10 minutes more than type 1, whereas this extra time could only have added one more fish to the simple random sampling. Thus, in this case the profit of two-stage sampling (845.92) become larger than that of loss (415.15). As a result, there is a gain in precision when type 2 of two-stage sampling is employed. Relative precision increased further when type 3 two-stage sampling was employed but to a lesser extent. This shows that when the fish sample is already large, a further increase in number of fish has little effect on the precision of the estimator. An increase in the number of subsamples has most effect on the precision if the number of subsamples is small originally.

The relative precision could have been improved if optimum allocation had been applied. Under optimum allocation the ratio of M to K

1.6 1.4 Precision 1.2 R 1.0 Relative 08 0,6 0,4 10 8 6 2 ō Time(Hours)

Fig. 1. Theoretical effect of cost (time) on relative precision on sample of July 31, 1973.

A=Two stage sampling with total of 3 subsampling units.

B=Two stage sampling with total of 12 subsampling units.

would have ranged from 4 to 8 with relative precision ranging from 1.18 to 2.11 (Table 5).

| | | Tabl | е 5 | | in an Ca |
|-----------|--------|--------|----------|--------|----------|
| Expected | ratio | of fis | h (M) | to n | umber |
| of subsan | ples | (K) an | d relati | ive pr | ecision |
| when the | two | stage | sampli | ng is | under |
| optimum | alloca | ation | | | |

| Date | Ratio of <i>M</i> to <i>K</i> | Relative precision | |
|----------|----------------------------------|--------------------|--|
| July 31 | 4 | 1.18 | |
| Aug. 16 | 4 | 1.62 | |
| Aug. 31 | 5 | 1.78 | |
| Sept. 5 | 6 | 1.43 | |
| Sept. 19 | 8, | 2.11 | |

In theory there are upper and lower limits of cost for the gain in precision from two-stage sampling when the number of subsamples (K) is constant⁽⁴⁾. To demonstrate this phenomenon for actual stomach analyses, the data from two



Theoretical effect of cost (time) on relative Fig. 2. precision on sample of August 16, 1973. A=Two stage sampling with total of 3 subsampling units.

B=Two stage sampling with total of 12 subsampling units.

sampling dates (July 31 and August 16) were chosen to calculate the relative precision as a function of cost. Given a cost (C) and number subsamples (K), the relative precision can be obtained by substituting the appropriate values into equation⁽⁴⁾. A plot of the relative precision against the cost is shown in Fig. 1 and Fig. 2. Relative precision is less than 1 when the cost is very low and initially increases dramatically with increasing cost. At higher cost, a maximum value of relative precision will be obtained, after which the relative precision becomes a decreasing function of cost.

Precision of estimation based on 3 subsamples is higher than that based on 12 subsamples when the fixed cost is low. The reverse is true when the cost is high. Take for example the sample of August 16 (Fig. 2). Precision of two-stage sampling with 3 subsample units is higher than that with 12 units when the cost is below 2 hours. The reverse is true when the cost is above 2 hours.

DISCUSSION

Although the precision of two-stage sampling with n primary units and m subsamples per unit is the same as that with one primary unit and mn subsamples, the disadvantage of the latter is that there is no estimation of sampling error. In the present study, however, the sampling error of the mean could be estimated from the linear relationship between log mean and log variance of stomach contents. The general application of this procedure, however, to other situations still needs to be justified.

It has been shown that the precision of subsampling depends on the total number of organisms counted. For a fixed number of organisms counted, the number of subsamples to be taken is inversely proportional to the concentration of the primary unit. It is evident that a higher concentration will need fewer subsamples and hence require less time. There is limitation, however, on the maximum concentration of zooplankton allowed. There should not be so many organisms that difficulty is encountered in counting and the assumption of. Poisson distribution should not be violated.

Ricker⁽⁶⁾ stated that when more than about. 200 organisms are counted on a medium-sized slide, the eye becomes confused and exact enumeration is not possible. Frolander⁽³⁾ found that counts of approximately 400 organisms per ml. in a petri dish gives no difficulty. R. L. Noble (personal communication) believes that dilution of a sample to 50-200 organisms per ml and counting in a Sedgewick-Rafter chamber results. in satisfactory estimation of zooplankton composition. However, the maximum number of zooplankton that one can count without difficulty related not only to the size of the plate but also. to size of the zooplankton. Difficulty may arise when counting over 200 D. pulex in a Sedgewick-Rafter chamber whereas larger numbers of smallsized zooplankton such as Bosmina, Chydorus and Cyclops may be counted easily.

A basic assumption on the variance equation (8) of the two-stage sampling was that the subsampling counts were Poisson distributed. Although counts of *D. pulex* from subsamples in the present study appeared to follow a Poisson distribution, this may not be true for some zooplanktons. Organisms with long appendages often clump together and can be overdispersed particularly when they occur in large numbers.

When zooplankton are overdispersed, Cassie⁽¹⁾ suggested that the variance of subsample counts will be in the form of, $\sigma^2 = u + cu^2$, where u is the mean number of organisms in the sample and c is a constant which depends on the characteristic of the sampling technique. Using sample mean \overline{X} as an estimate of population uand substituting the variance expression here into equation (5) gives the overall variance of the estimation as:

$$V(\overline{\overline{X}}_{tss}) = \frac{S^2}{M} + \frac{\overline{X}^2}{K\overline{y}} + \frac{c\overline{X}^2}{K}$$

The last term of the equation arises from the overdispersion of organisms in the primary unit, whereas the middle term of the right-hand side arises from random distribution of the organisms. It is clear that an increase in the density of primary unit reduces only the variance arising from random distribution, whereas an increase in the number of subsamples decreases both variance components of subsampling. Hence, for a fixed number of organisms counted, the precision of the estimator could be improved by employing a large number of subsamples (low density of primary unit). However, a large number of subsamples will cost more time and effort which defeats the purpose of two-stage sampling. In addition, the subsampling error of an estimator in the situation when organisms are overdispersed in the primary unit will be much larger than in the case of a Poisson disttibution.

As shown in equation (8), the subsampling error is expressed as $\overline{X^2}/K\overline{y}$. By using the coefficient of variation (*CV*), the subsampling error may be expressed in relative terms, as percentage of error of the estimator contributed by subsampling procedures. The basic relationship is

$$CV = \frac{100\sigma}{\overline{X}} = \frac{100}{\sqrt{K\overline{y}}}$$

where $K\overline{y}$ is the total number of organisms counted. The equation shows that the expected coefficient of variation is 10 percent for 100 organisms counted and 5 percent for 400 organisms counted. It is clear that counting beyond 100 will not do much to improve the precision of the estimator. For practical purposes, a count of 100 to 400 organisms may be satisfactory regardless of the abundance of organisms in stomach or in the primay unit.

On the basis of the present study, one can conclude that two-stage sampling generally is better than simple random sampling when estimating the amount of zooplankton in fish stomachs. Given a moderate sample of fish, twostage sampling of stomach contents will cost about only half the time required for simple random sampling. If the organisms couted are Poisson distributed, the expected additional sampling error will be inversely proportional to the total number of organisms couned. However, when the organisms counted are overdispersed in the subsamples, the expected subsampling error will be inversely proportional to total number of organisms counted and total number of subsamples. In conducting the subsamples, therefore, it is desirable to know beforehand whether the number of organisms in the subsample is Poisson distributed or not. If the assumption of Poisson distribution seems reasonable fewer subsamples with high density should be applied to save time and effort, otherwise, a larger number of subsamples will be required to reduce the subsampling errors caused by overdispersion.

Acknowledgments: I wish to thank Dr. Douglas S. Robson, Cornell University, for suggestion of the study, guidance throughout the investigation and advice on statistical analysis.

REFERENCES

- Cassie, R.E. (1971). Sampling statistics. In: W.T. Edmondson and G.G. Winberg (ed.). A manual on methods for the assessment of secondary productivity in fresh waters. I.B.P. Handbook No. 17. Blackwell, Oxford 358pp.
- Cochran, W.G. (1963). Sampling techniques. John Wiley and Sons, Inc., New York 413pp.
- Frolander, H. F. (1968). Statistical variation in zooplankton numbers from subsampling with Stempel pipette. J. Water Pollution Contr. Fed. 40 (suppl) R82-88.
- 4. Lin, Y.S. (1975). Food and growth of young yellow perch during the pelagic and demersal stages in Oneida Lake. Ph.D. Thesis, Cornell University, Ithaca, N.Y. 96pp. (unpublished).
- Miura, T. and K. Naka (1970). Some problems in a numerical treatment of netted plankton samples. Japan. Jour. Limnol. 31(3): 83-95.
- Ricker, W.E. (1937). Statistical treatment of sampling processes useful in the enumeration of plankton organisms. Arch. Hydrobiol. 31: 68-87.
- Southwood, T. R. E. (1966). Ecological methods. Methuen & Co Ltd., London 391pp.

YAO-SUNG LIN

二段抽樣法在魚胃含物的分析

林 曜 松

本研究是比較簡單逢機抽樣法及二段抽樣法在估計平均魚胃中浮游生物含量的準確度。作者依據二 段抽樣法理論推算出平均胃含量的變方公式。依據此公式可推知:變方大小是與樣品數及計算的浮游生 物總數有關。當取同樣大小的樣品數時,二段抽樣法比簡單逢機抽樣法可省時一倍。在花費相同時間下 ,前者比後者準確度要高。