Abstract—Longitudinal surveys of anglers or boat owners are widely used in recreational fishery management to estimate total catch over a fishing season. Survey designs with repeated measures of the same random sample over time are effective if the goal is to show statistically significant differences among point estimates for successive time intervals. However, estimators for total catch over the season that are based on longitudinal sampling will be less precise than stratified estimators based on successive independent samples. Conventional stratified variance estimators would be negatively biased if applied to such data because the samples for different time strata are not independent. We formulated new general estimators for catch rate, total catch, and respective variances that sum across time strata but also account for correlation stratum samples. A case study of the Japanese recreational fishery for ayu (*Plecoglossus altivelis*) showed that the conventional stratified variance estimate of total catch was about 10% of the variance estimated by our new method. Combining the catch data for each angler or boat owners throughout the season reduced the variance of the total catch estimate by about 75%. For successive independent surveys based on random independent samples, catch, and variance estimators derived from combined data would be the same as conventional stratified estimators when sample allocation is proportional to strata size. We are the first to report annual catch estimates for ayu in a Japanese river by formulating modified estimators for day-permit anglers.

Manuscript accepted 14 September 2001.
Fish. Bull. 100:228–243 (2002).

Longitudinal logbook survey designs for estimating recreational fishery catch, with application to ayu (*Plecoglossus altivelis*)

Shuichi Kitada

Tokyo University of Fisheries
4-5-7 Konan, Minato
Tokyo 108-8477, Japan
Email address: kitada@tokyo-u-fish.ac.jp

Kiyoshi Tezuka

Nakagawa Branch
Tochigi Prefectural Fisheries Experimental Station
Ogawa, Nasu
Tochigi 324-0501, Japan

Angler surveys are widely used in fishery management to estimate recreational catch, and there is an extensive body of literature on this subject (see Guthrie et al., 1991). Pollock et al. (1994) published a manual on angler survey methods and their applications in fishery management. The first purpose of our study is to make two very important points for the designers of recreational fishery surveys: 1) longitudinal surveys taking repeated measures on the same random sample of anglers or boats over time are better than successive independent surveys if the goal is to determine significant trends in catch and fishing effort, and 2) stratified surveys, or successive independent surveys based on random independent samples of anglers or boats, are better than longitudinal surveys if the goal is to obtain precise estimates of annual total catch and fishing effort. If longitudinal fisheries data sets are used to estimate annual catches, then correlations between monthly sample observations must be taken into account when evaluating the precision of catch estimates. This problem is not addressed in the literature, and the variance estimation procedures for this situation are unclear. The second purpose of our study is to estimate the annual catch of ayu (*Plecoglossus altivelis*) in a river because no estimates have been reported in Japan.

In our study, we formulated a new method for accurate variance estimation with longitudinal fishery data, exemplified by a case study of the recreational fishery for ayu in Nakagawa River in Tochigi Prefecture, Japan (Fig. 1). Annual catch estimates based on sums of monthly estimates were compared with those based on combined data for each angler or boat throughout the fishing season. We demonstrate how use of a design with repeated measures facilitates determination of significant seasonal trends in catches, and show the usefulness of combined (nonstratified) data analysis. We also estimate the total annual catch of the ayu fishery by formulating modified estimators for day-permit anglers (anglers who are granted permits to take fish for one day).

Materials and methods

Case study of ayu

We used longitudinal data collected for the Japanese ayu fishery to compare estimators of effort and catch and their associated variance estimators. Ayu is the most popular target species of recreational anglers in rivers in Japan. In the Nakagawa River (Fig. 1), the upstream run of wild juvenile ayu from the coast begins in late March to early April, and is completed by early July. Ayu mature and spawn from September to November, and then die after spawning. Cooperatives release both hatchery-produced and wild juveniles caught

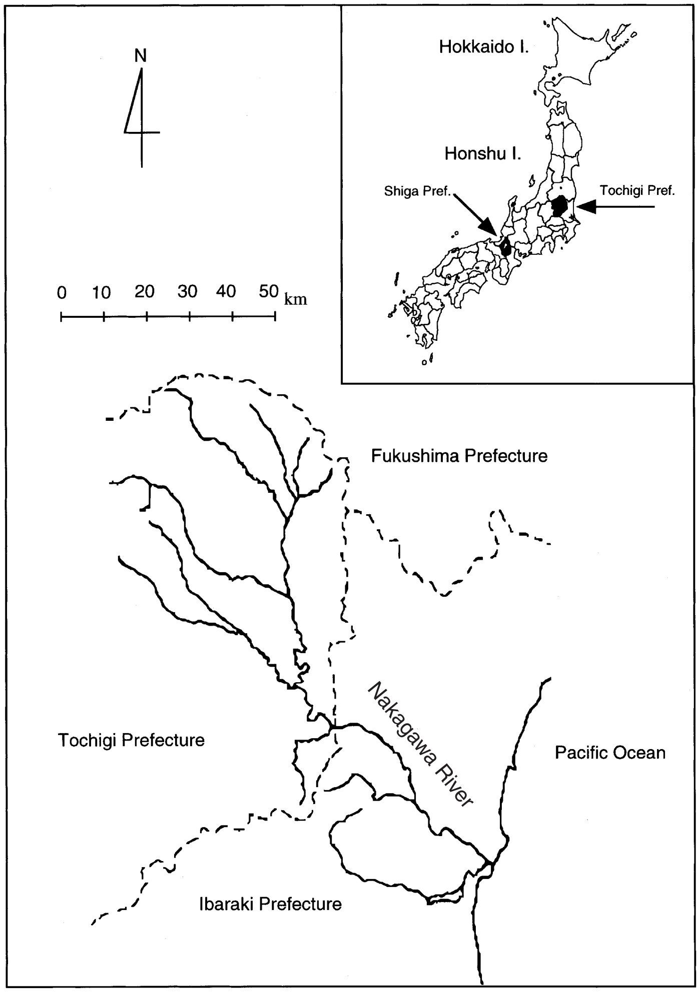

Figure 1

Location of the angler survey. Bold lines in the lower figure indicate the area of the Nakagawa River where the survey was conducted.

in Biwako Lake, Shiga Prefecture (Fig. 1) from early April to the end of May. Thus, recreational anglers catch both naturally recruited wild juveniles and transplanted wild juveniles from Biwako Lake and hatchery-produced ayu released by the cooperatives. Both hatchery and wild stocks consist of a single year class that recruits in the spring. The river fishing season for ayu begins on June 1st and closes at the end of October. To estimate the annual ayu catch by recreational anglers in the Nakagawa River, we conducted a longitudinal log book survey in 1993.

Sampling procedure

There are four cooperatives that set fishing rights on the Nakagawa River in Tochigi Prefecture. Fishing permits for ayu are sold at the cooperatives and fishing tackle shops, and these permits are valid over the entire Nakagawa River in the prefecture. The cooperatives record the total number of season- and day-permits sold, and a complete list of season-permit anglers is available. An *a priori* sample size of 120 anglers (an expected sampling fraction of about 0.5% of the total number of season-permit anglers) was allocated to the four cooperatives in proportion to the number of season-permits sold (Table 1). Anglers who possess a permit (season or day) can fish for ayu over the whole Nakagawa area, regardless of where the permit was purchased. Hence, we treated the samples as if they were drawn from the population by simple random sampling, even though they were drawn by stratified random sampling of cooperatives.

We asked the cooperatives to select samples randomly, but the samples were drawn arbitrarily. The selection, however, was not a purposive sampling; therefore we treated them as random samples. The sampled season-permit anglers were asked to record catch data throughout the fishing season, including each fishing date, the number of ayu caught, and the fishing site, on a printed form, which was returned after the fishing season was over. To estimate the total catch in weight, we also surveyed the body weights of ayu in recreational catches by month. The

Table 1

The number of permits sold in 1993, sample size, and the number of logbooks returned by the four fishermen's cooperative associations along the Nakagawa River. n = number of anglers sampled.

Fishermen's cooperative associations	Number of season permits sold	Number of day permits	n	Number of logbooks returned
Hokubu	11,314	6520	70	64
Nanbu	6391	1946	30	21
Chuo	1911	231	10	10
Motegi	2346	369	10	9
Total	21,962	9066	120	104

primary sampling unit in a population of season-permit anglers or boat owners (i.e. party-boat owners or personal-boat owners) was an angler or a boat owner, and the secondary sampling unit was a fishing day. We selected anglers or boat owners by simple random sampling without replacement from a list of anglers or boat owners, and asked a sample of anglers or boat owners to record catch data on all fishing days throughout the survey. Because all the secondary sampling units were surveyed, we regarded this as a single-stage cluster sampling procedure (Cochran, 1977).

Estimation of total catch by month

Estimation for season-permit anglers or boat owners The principal notations for estimation of the total catch for season-permit anglers or party-boat owners are as follows (Cochran, 1977; Pollock et al., 1994):

- N = total number of sampling units (season-permit angler or party-boat or personal-boat owner; known number);
- n = sample size drawn from the population N;
- M_k = total number of fishing days in the population in kth month (to be estimated);
- M_{ik} = total number of fishing days in kth month of selected ith sample;
- $\overline{M}_k$ = mean number of fishing days per sampling unit in kth month (to be estimated);
- C_{ik} = number of fish caught by ith sample in kth month;
- R_k = catch rate of kth month (to be estimated);
- $C_k^{(s)}$ = total catch by season-permit anglers or by party-boat owners or personal-boat owners in kth month (to be estimated).

The catch rate is the number of fish caught per sampling unit each day. For logbook surveys, the ratio of the mean is the preferable estimator of catch rate (Jones et al.,1995; Pollock et al., 1997).

The ratio and the variance for the catch rate is estimated by

$$\hat{R}_k = \frac{\sum_{i=1}^{n} C_{ik}}{\sum_{i=1}^{n} M_{ik}}, \tag{1}$$

$$\tilde{V}(\hat{R}_k) \simeq \frac{N}{M_k^2}\frac{N-n}{n(n-1)}\sum_{i=1}^{n}(C_{ik} - M_{ik}\hat{R}_k)^2. \tag{2}$$

This variance estimator is obtained by dividing Equation 6.9 in Cochran (1977, p. 155) by the total number of fishing days in the kth month M_k. In Equation 2, M_k is unknown; hence we approximated the variance estimate by using the estimator of M_k as follows (Thompson, 1992, p. 62):

$$\hat{V}(\hat{R}_k) \simeq \frac{N-n}{N\hat{\overline{M}}_k^2 n(n-1)}\sum_{i=1}^{n}(C_{ik} - M_{ik}\hat{R}_k)^2, \tag{3}$$

where $\hat{M}_k = N\hat{\overline{M}}_k$. The variance of $\hat{M}_k$ is estimated by $\hat{V}(\hat{M}_k) = N^2\hat{V}(\overline{M}_k)$, where $\overline{M}_k$ is the mean number of fishing days per sampling unit in kth month. The estimator and the variance are as follows (Cochran, 1977, p. 249, from Eq. 9A.2):

$$\hat{\overline{M}}_k = \frac{1}{n}\sum_{i=1}^{n} M_{ik}, \tag{4}$$

$$\hat{V}(\hat{\overline{M}}_k) = \frac{N-n}{Nn(n-1)}\sum_{i=1}^{n}(M_{ik} - \hat{\overline{M}}_k)^2. \tag{5}$$

The total number of fish caught in kth month is estimated by

$$\begin{aligned}\hat{C}_k^{(s)} &= \hat{M}_k\hat{R}_k = N\hat{\overline{M}}_k\hat{R}_k \\ &= N\frac{\sum_{i=1}^{n} M_{ik}}{n}\frac{\sum_{i=1}^{n} C_{ik}}{\sum_{i=1}^{n} M_{ik}} = \frac{N}{n}\sum_{i=1}^{n} C_{ik}.\end{aligned} \tag{6}$$

This results in an unbiased estimator. When the total number of fishing days M_k is unknown, the ratio estimator coincides with the unbiased estimator. The variance is evaluated by (Cochran, 1977, p. 249, from Eq. 9A.2):

$$\hat{V}(\hat{C}_k^{(s)}) = \frac{N(N-n)}{n(n-1)} \sum_{i=1}^{n} (C_{ik} - \hat{\bar{C}}_k)^2 \tag{7}$$

where $\hat{\bar{C}}_k = \sum_{i=1}^{n} C_{ik} / n.$

Estimation for day-permit anglers We estimated the catch of day-permit anglers separately. The notations for day-permit anglers catch estimation are as follows:

D = total number of day-permits sold through the season (known);
D_k = total number of day-permits sold in k th month (known);
d_k = total number of day-permit anglers who returned logbooks in kth month;
$R_k^{(d)}$ = catch rate of kth month for day-permit anglers;
$C_k^{(d)}$ = total catch by day-permit anglers in kth month (to be estimated).

The estimator of total number of fish caught by day-permit anglers in kth month is

$$\hat{C}_k^{(d)} = D_k \hat{R}_k^{(d)}, \tag{8}$$

where the catch rate for day-permit anglers, estimated by the sample mean, is

$$\hat{R}_k^{(d)} = \frac{1}{d_k} \sum_{i=1}^{d_k} c_{ik}. \tag{9}$$

The variance estimator of $\hat{C}_k^{(d)}$ is (Cochran, 1977, p. 26, from Eq. 2.20):

$$\hat{V}(\hat{C}_k^{(d)}) = D_k^2 \hat{V}(\hat{R}_k^{(d)}) = \frac{D_k(D_k - d_k)}{d_k(d_k-1)} \sum_{i=1}^{d_k} (C_{ik} - \hat{R}_k^{(d)})^2. \tag{10}$$

Total monthly catch of season- and day-permit anglers The total catch by season- and day-permit anglers in kth month is C_k, which is estimated by

$$\hat{C}_k = \hat{C}_k^{(s)} + \hat{C}_k^{(d)}. \tag{11}$$

These two total estimates are obtained from independent samples (season- and day-permit anglers); therefore the variance is estimated by adding the variances:

$$\hat{V}(\hat{C}_k) = \hat{V}(\hat{C}_k^{(s)}) + \hat{V}(\hat{C}_k^{(d)}). \tag{12}$$

The total catches in weight in kth month are estimated by using the mean body weight of the species in kth month ($\overline{w}_k$) estimated from the survey of individual body weights by

$$\hat{W}_k = \hat{W}_k^{(s)} + \hat{W}_k^{(d)} = \hat{C}_k \hat{\bar{w}}_k, \tag{13}$$

where $\hat{W}_k^{(s)}$ and $\hat{W}_k^{(d)}$ are the total catches in weight for season- and day-permit anglers, given by $\hat{W}_k^{(s)} = \hat{C}_k^{(s)} \hat{\bar{w}}_k$ and $\hat{W}_k^{(d)} = \hat{C}_k^{(d)} \hat{\bar{w}}_k$. In general, season-permit and day-permit anglers fish in the same location; therefore we assumed the same mean body weight for both kinds of anglers. The samples for estimating C_k and $\overline{w}_k$ are independent; therefore the variances are estimated by using Goodman's (1960) formula:

$$\hat{V}(\hat{W}_k) = \hat{V}(\hat{W}_k^{(s)}) + \hat{V}(\hat{W}_k^{(d)}) \tag{14}$$

where $\hat{V}(\hat{W}_k^{(s)}) = \hat{\bar{w}}_k^2 \hat{V}(\hat{C}_k^{(s)}) + \hat{C}_k^{(s)2} \hat{V}(\hat{\bar{w}}_k) + \hat{V}(\hat{C}_k^{(s)}) \hat{V}(\hat{\bar{w}}_k)$

and $\hat{V}(\hat{W}_k^{(d)}) = \hat{\bar{w}}_k^2 \hat{V}(\hat{C}_k^{(d)}) + \hat{C}_k^{(d)2} \hat{V}(\hat{\bar{w}}_k) + \hat{V}(\hat{C}_k^{(d)}) \hat{V}(\hat{\bar{w}}_k).$

The mean body weight of fish in kth month is estimated from a survey of individual body weights (w_{ik} of l_k fish caught on the fishing grounds). The estimator for kth month and its variance estimator are

$$\hat{\bar{w}}_k = \sum_{i=1}^{l_k} w_{ik} / l_k$$

and $$\hat{V}(\hat{\bar{w}}_k) = \sum_{i=1}^{l_k} (w_{ik} - \hat{\bar{w}}_k)^2 / \left(l_k(l_k - 1)\right).$$

The total fishing days in kth month is estimated as the sum of the fishing days estimates of the season-permit anglers and the day-permit anglers by

$$\hat{M}_{Tk} = \hat{M}_k + D_k \tag{15}$$

and the variance is estimated by $\hat{V}(\hat{M}_{Tk}) = N^2 \hat{V}(\hat{\bar{M}}_k)$ because D_k is known.

Estimation of annual catch

Method 1 (based on monthly estimates) The annual catch is estimated by summing monthly catch estimates over the entire fishing season (K months). The point estimator is

$$\hat{C} = \sum_{k=1}^{K} \hat{C}_k = \sum_{k=1}^{K} \hat{C}_k^{(s)} + \sum_{k=1}^{K} \hat{C}_k^{(d)} = \hat{C}^{(s)} + \hat{C}^{(d)}. \tag{16}$$

When the same sample of anglers reports catches throughout the season, the sampling is not independent in each month, and monthly catch estimates are auto-correlated. Taking this correlation into account, the variance estimator is

$$\begin{aligned} \hat{V}(\hat{C}) &= \hat{V}(\hat{C}^{(s)}) + \hat{V}(\hat{C}^{(d)}) \\ &= \sum_{k=1}^{K} \hat{V}(\hat{C}_k^{(s)}) + 2 \sum_{k \le k'}^{K} \widehat{\mathrm{Cov}}(\hat{C}_k^{(s)}, \hat{C}_{k'}^{(s)}) \\ &+ \sum_{k=1}^{K} \hat{V}(\hat{C}_k^{(d)}) + 2 \sum_{k \le k'}^{K} \widehat{\mathrm{Cov}}(\hat{C}_k^{(d)}, \hat{C}_{k'}^{(d)}). \end{aligned} \tag{17}$$

The covariance between two total estimates of season-permit anglers is estimated by (see Appendix 1, from Cochran, 1977, p. 25)

$$\widehat{\mathrm{Cov}}(\hat{C}_k^{(s)},\hat{C}_{k'}^{(s)})=\frac{N(N-n)}{n(n-1)}\sum_{i=1}^{n}(C_{ik}-\hat{\bar{C}}_k)(C_{ik'}-\hat{\bar{C}}_{k'}).\quad(18)$$

The fourth term of Equation 17 is equal to 0 if a different sample of day-permit anglers is drawn in each month.

The total catch in weight is estimated by

$$\hat{W}=\sum_{k=1}^{K}\hat{W}_k=\sum_{k=1}^{K}\hat{W}_k^{(s)}+\sum_{k=1}^{K}\hat{W}_k^{(d)}=\hat{W}^{(s)}+\hat{W}^{(d)}\quad(19)$$

and the variance estimator is similar to Equation 17 but has a slightly different covariance which is

$$\widehat{\mathrm{Cov}}(\hat{W}_k^{(s)},\hat{W}_{k'}^{(s)})\sim\hat{\bar{w}}_k\hat{\bar{w}}_{k'}\widehat{\mathrm{Cov}}(\hat{C}_k^{(s)},\hat{C}_{k'}^{(s)}).\quad(20)$$

This covariance estimator was derived by the delta method (Seber, 1982, p. 7, see Appendix 2), which coincides with a covariance when $\hat{\bar{w}}_k$ and $\hat{\bar{w}}_{k'}$ are constant.

The mean annual catch rate is estimated for season-permit anglers or boats by

$$\hat{R}^{(s)}=\frac{\hat{C}^{(s)}}{\hat{M}^{(s)}},\quad(21)$$

where $\hat{M}^{(s)}=\sum_{k=1}^{K}\hat{M}_k=N\sum_{k=1}^{K}\hat{\bar{M}}_k$.

Here $\hat{\bar{M}}_k$ is given by Equation 4. The total effort is estimated by $\hat{M}^{(s)}+D$. The approximate variance of $\hat{V}(\hat{R}^{(s)})$ is given by the delta method (Seber, 1982, p. 7; Appendix 3), that is

$$\hat{V}(\hat{R}^{(s)})\simeq\frac{1}{\hat{M}^{(s)2}}\left\{\begin{array}{l}\hat{V}(\hat{C}^{(s)})+\left(\dfrac{\hat{C}^{(s)}}{\hat{M}^{(s)}}\right)^2\hat{V}(\hat{M}^{(s)})\\-2\dfrac{\hat{C}^{(s)}}{\hat{M}^{(s)}}(\widehat{\mathrm{Cov}}(\hat{C}^{(s)},\hat{M}^{(s)})\end{array}\right\},\quad(22)$$

where $\hat{V}(\hat{C}^{(s)})$ is given by Equation 17, and the variance of the total number of fishing days is estimated by

$$\hat{V}(\hat{M}^{(s)})=N^2\left\{\sum_{k=1}^{K}\hat{V}(\hat{\bar{M}}_k)+2\sum_{k\le k'}^{K}\widehat{\mathrm{Cov}}(\hat{\bar{M}}_k,\hat{\bar{M}}_{k'})\right\}.\quad(23)$$

Here $\hat{V}(\hat{\bar{M}}_k)$ is given by Equation 5 and the covariance of two sample means is estimated by (Cochran, 1977, p. 25)

$$\widehat{\mathrm{Cov}}(\hat{\bar{M}}_k,\hat{\bar{M}}_{k'})=\frac{N-n}{Nn(n-1)}\sum_{i=1}^{n}(M_{ik}-\hat{\bar{M}}_k)(M_{ik'}-\hat{\bar{M}}_{k'}).\quad(24)$$

The estimator of covariance between $\hat{C}^{(s)}$ and $\hat{M}^{(s)}$ is similar to Equation 18 (see Appendix 1):

$$\widehat{\mathrm{Cov}}(\hat{C}^{(s)},\hat{M}^{(s)})=\frac{N(N-n)}{n(n-1)}\sum_{i=1}^{n}(C_i-\hat{\bar{C}})(M_i-\hat{\bar{M}}),\quad(25)$$

where C_i and M_i are the number of fish caught and the number of fishing days of ith season-permit angler or boat owner throughout the season, respectively, and

$$\hat{\bar{C}}_k=\sum_{i=1}^{n}C_{ik}/n.$$

and

$$\hat{\bar{M}}=\sum_{i=1}^{n}M_i/n.$$

The mean number of fishing days per sampling unit (season-permit angler or boat) is estimated by

$$\hat{\bar{M}}=\frac{\hat{M}^{(s)}}{N}$$

and the variance estimator is

$$\hat{V}(\hat{\bar{M}})=\frac{1}{N^2}\hat{V}(\hat{M}^{(s)}),$$

where $\hat{M}^{(s)}$ and $\hat{V}(\hat{M}^{(s)})$ are already derived from Equation 21 and Equation 23, respectively.

The catch rate of day-permit anglers over the season is estimated by

$$\hat{R}^{(d)}=\frac{\hat{C}^{(d)}}{D}$$

and the variance is

$$\hat{V}(\hat{R}^{(d)})=\frac{1}{D^2}\hat{V}(\hat{C}^{(d)}),$$

where $\hat{C}^{(d)}$ and $\hat{V}(\hat{C}^{(d)})$ are given in Equation 8 and Equation 10, respectively.

Method 2 (based on total annual catch of each angler or boat owner) Another procedure for estimating annual catch is to use annual data rather than monthly catch for individual anglers or boats. The advantages of this procedure are that covariances between months do not have to be considered and estimators are much less complicated than those obtained using method 1. Equations derived for monthly estimation can be used without modification for this procedure.

Modified estimators for day-permit anglers

We could not conduct a survey of day-permit anglers, so we substituted $\hat{R}_k$ for $\hat{R}_k^{(d)}$ in Equation 8. In addition, the total number of day permits sold in kth month (D_k) was unknown. Hence, we slightly modified the procedure for estimating D_k by $D\hat{p}_k$, where p_k is the proportion of day permits sold in kth month to the annual total number of day permits sold (D).

Day permits issued by the cooperatives are sold mainly in fishing tackle shops. We selected four tackle shops and surveyed the total number of day permits sold at the selected jth fishing tackle shop (D_j), and the total number of day permits sold at the selected jth fishing tackle shop in kth month (D_{jk}). The proportion of day permits sold in kth month was estimated by

$$\hat{p}_k = \frac{\sum_{j=1}^{h} D_{jk}}{\sum_{j=1}^{h} D_j},$$

where h is the number of fishing tackle shops selected from a total of H shops. The evaluation of the variance of $\hat{p}_k$ was similar to Equation 3:

$$\tilde{V}(\hat{p}_k) = \frac{H(H-h)}{D^2 h(h-1)} \sum_{j=1}^{h} (D_{jk} - D_j \hat{p}_k)^2.$$

Some day permits, however, were sold at the fishing sites, and the above variance estimator was not appropriate for this situation. Assuming $\Sigma_{j=1}^{h} D_j$ was selected by simple random sampling from D, we evaluated the variance by (Cochran, 1977, p. 52, Eq. 3.8)

$$\hat{V}(\hat{p}_k) = \frac{D - \sum_{j=1}^{h} D_j}{D\left(\sum_{j=1}^{h} D_j - 1\right)} \hat{p}_k (1 - \hat{p}_k).$$

The modified estimator for the total number of fish caught by day-permit anglers in kth month is

$$\hat{C}_k^{(d)} = D\hat{p}_k \hat{R}_k. \tag{26}$$

Here $\hat{p}_k$ and $\hat{R}_k$ are independent because these are estimated from different survey data. The variance of revised $\hat{C}_k^{(d)}$ was estimated by using Goodman's (1960) method:

$$\hat{V}(\hat{C}_k^{(d)}) = D^2 \left\{ \hat{R}_k^2 \hat{V}(\hat{p}_k) + \hat{p}_k^2 \hat{V}(\hat{R}_k) + \hat{V}(\hat{p}_k)\hat{V}(\hat{R}_k) \right\}.$$

The total fishing days in kth month was estimated by Equation 15 but in this case D_k was unknown. The modified estimator was

$$\hat{M}_{Tk} = \hat{M}_k + D\hat{p}_k$$

and the variance was slightly revised as

$$\hat{V}(\hat{M}_{Tk}) = N^2 \hat{V}(\hat{\bar{M}}_k) + D^2 \hat{V}(\hat{p}_k).$$

The annual catch was estimated by Equation 16, substituting Equation 8 by Equation 26. In this case, we estimated $\hat{C}_k^{(d)}$ and $\hat{C}_{k'}^{(d)}$ from the same sample of season-permit anglers. Hence, the fourth term of the covariance in Equation 17 must be considered. The approximate covariance is estimated by (see Appendix 4)

$$\widehat{\mathrm{Cov}}(\hat{C}_k^{(d)}, \hat{C}_{k'}^{(d)}) \simeq D^2 \hat{p}_k \hat{p}_{k'} \widehat{\mathrm{Cov}}(\hat{R}_k, \hat{R}_{k'}), \tag{27}$$

where the covariance between $\hat{R}_k$ and $\hat{R}_{k'}$ is

$$\widehat{\mathrm{Cov}}(\hat{R}_k, \hat{R}_{k'}) \cong \frac{1}{N^4} \begin{pmatrix} \frac{1}{\hat{\bar{M}}_k \hat{\bar{M}}_{k'}} \widehat{\mathrm{Cov}}(\hat{C}_k^{(s)}, \hat{C}_{k'}^{(s)}) \\ -\frac{\hat{C}_{k'}^{(s)}}{\hat{\bar{M}}_k \hat{\bar{M}}_{k'}^2} \widehat{\mathrm{Cov}}(\hat{C}_k^{(s)}, \hat{\bar{M}}_{k'}) \\ -\frac{\hat{C}_k^{(s)}}{\hat{\bar{M}}_k^2 \hat{\bar{M}}_{k'}} \widehat{\mathrm{Cov}}(\hat{\bar{M}}_k, \hat{C}_{k'}^{(s)}) \\ +\frac{\hat{C}_k^{(s)} \hat{C}_{k'}^{(s)}}{\hat{\bar{M}}_k^2 \hat{\bar{M}}_{k'}^2} \widehat{\mathrm{Cov}}(\hat{\bar{M}}_k, \hat{\bar{M}}_{k'}) \end{pmatrix}.$$

Here $\widehat{\mathrm{Cov}}(\hat{C}_k^{(s)}, \hat{C}_{k'}^{(s)})$ and $\widehat{\mathrm{Cov}}(\hat{\bar{M}}_k, \hat{\bar{M}}_{k'})$ are already given by Equation 18 and Equation 24. The other covariance components are

$$\widehat{\mathrm{Cov}}(\hat{C}_k^{(s)}, \hat{\bar{M}}_{k'}) = \frac{N-n}{n^2(n-1)} \sum_{i=1}^{n} (C_{ik} - \hat{\bar{C}}_k)(M_{ik'} - \hat{\bar{M}}_{k'})$$

$$\widehat{\mathrm{Cov}}(\hat{\bar{M}}_k, \hat{C}_{k'}^{(s)}) = \frac{N-n}{n^2(n-1)} \sum_{i=1}^{n} (C_{ik'} - \hat{\bar{C}}_{k'})(M_{ik} - \hat{\bar{M}}_k).$$

The annual catch in weight was estimated by Equation 19, substituting Equation 8 with Equation 26. The covariance in the fourth term of $\hat{V}(\hat{W})$ in Equation 17 was estimated with Equation 20 by

$$\widehat{\mathrm{Cov}}(\hat{W}_k^{(d)}, \hat{W}_{k'}^{(d)}) \simeq \hat{\bar{w}}_k \hat{\bar{w}}_{k'} \widehat{\mathrm{Cov}}(\hat{C}_k^{(d)}, \hat{C}_{k'}^{(d)})$$

where $\widehat{\mathrm{Cov}}(\hat{C}_k^{(d)}, \hat{C}_{k'}^{(d)})$ is given in Equation 27.

Results

In 1993, 21,962 season permits and 9066 day permits were sold (Table 1). The total number of day permits sold at the four fishing tackle shops was 4776, and the number (proportion, $\hat{p}_k$) sold was 2732 (0.572) in June, 1,189 (0.249) in July, 716 (0.150) in August, 124 (0.026) in September, and 15 (0.003) in October.

We received 104 logbooks from the 120 season-permit anglers sampled, a return rate of 86.7%. In addition, two anglers voluntarily submitted logbooks, but we did not include these unsolicited returns because they were not randomly selected. The modes of the catch rates by the sampled anglers were from five to ten fish per month (Fig. 2). The histograms show a large variation in the catch rate among season-permit anglers. The peak fishing season was from June to July. In September, the number of anglers decreased, and the fishing season ended in October. The

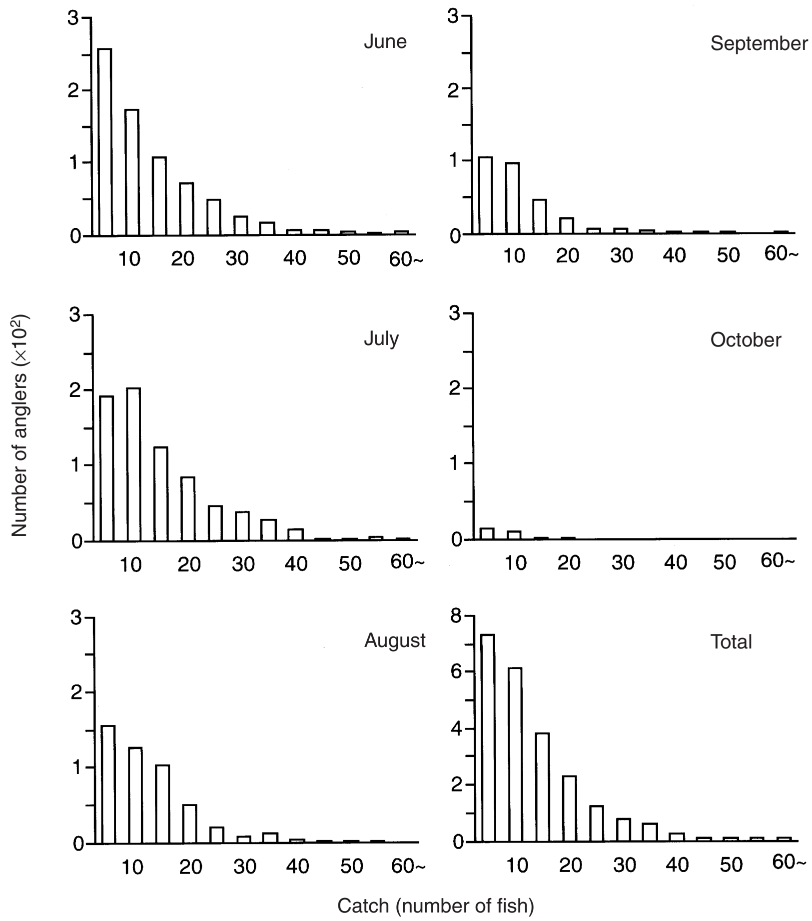

Figure 2

Distributions of the catch rate of ayu by month for 104 sampled season-permit anglers in the Nakagawa River.

modes of the number of fishing days per season-permit angler were five for all months. The variation in the number of fishing days among anglers was also large (Fig. 3).

Monthly plots of the total number of fish caught versus the number of fishing days showed linear relationships (Fig. 4). The variation in Figure 4 indicates differences in the skill of the anglers. The monthly number of anglers decreased over the fishing season. Figure 5 shows the monthly changes in the total number of fishing days, the total number of fish caught, and the catch rate for the 104 sampled anglers. The decline in number of fish caught was largely due to the decrease in fishing days. The change in catch rate indicated a decline in the abundance of the stock.

The mean body weight of ayu was greatest in June (Fig. 6) and was affected by a method of fishing for ayu called "Tomo-zuri" angling, which takes advantage of the attacking behavior of ayu when another fish enters its territory. Anglers attach a "call" fish (a live ayu) above a treble hook that snares the territorial wild fish, as it attacks the "call" fish. Because larger individuals establish territories earlier than smaller ayu, fish caught in June were predominantly the larger individuals.

Reflecting the monthly trend in the number of fishing days, 89% of the total annual catches of season-permit anglers and 98% of those of day-permit anglers were taken from June to August (Table 2). The catch by day-permit anglers was substantially smaller than anticipated, estimated at about 2% of season-permit anglers' catch in both numbers and total weight. CVs ranged from 7% to 12% in June and July for all parameters; however, they were higher in August and September, ranging from 10% to 20%. In October, CVs exceeded 43% for total catch in number and weight. The decreasing precision of the monthly catch rate estimates was caused by the decrease in anglers (n_k) (Figs. 4 and 5).

The CVs of annual estimates of $\overline{M}$ and $\hat{M}_T$ by method 1 were about 7%, but that of $\hat{R}^{(s)}$ was about 20% (Table 2) because we evaluated the covariance terms for the number of catches and fishing days between months; those were

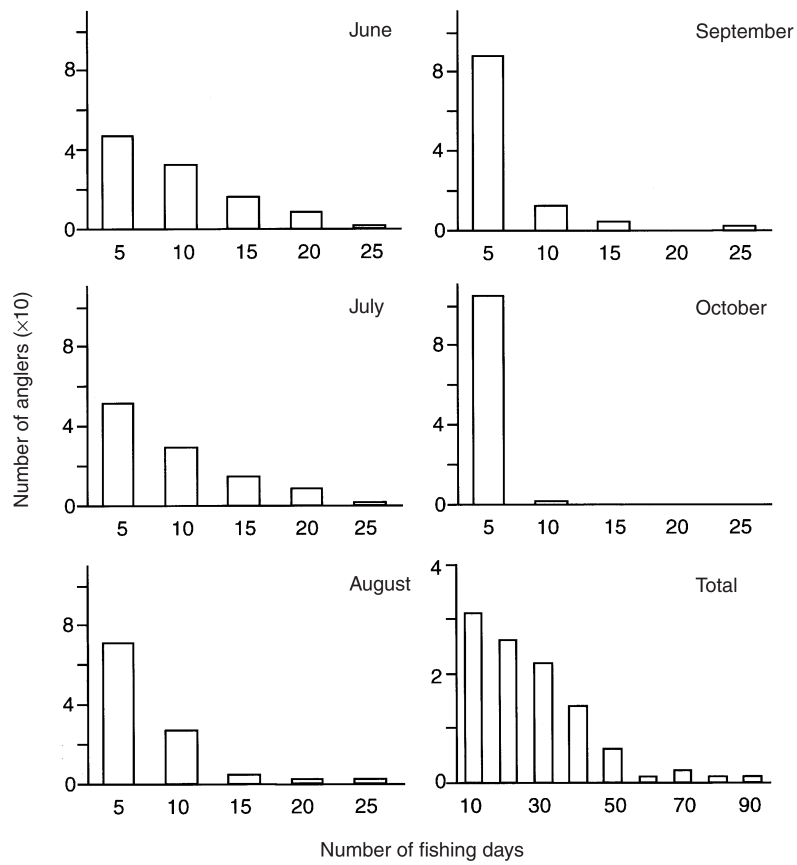

Figure 3

Distributions of the ayu fishing days per angler by month for 104 sampled season-permit anglers in the Nakagawa River.

$\widehat{\text{Cov}}(\hat{C}_k^{(s)}, \hat{C}_{k'}^{(s)})$ in $\hat{V}(\hat{C}^{(s)})$ and $\widehat{\text{Cov}}(\hat{\overline{M}}_k, \hat{\overline{M}}_{k'})$ in Equation 22. The CVs of $\hat{C}$ and $\hat{W}$ were also evaluated at about 21% and were strongly affected by the covariances between months in Equation 17. The variance of the total number estimate $\hat{V}(\hat{C})$ was 1.2604×10^{12}, and variance by neglecting the covariance term in Equation 17 was 1.2230×10^{11}. The CV of $\hat{C}$ without the covariance term was 6.53%. If we neglect the covariance, the variance is substantially underestimated. The variance was 10.31 times larger when the covariance term was included.

We obtained similar point estimates of annual catch by method 2 (Total$^{\text{meth. 2}}$ in Table 2). The CVs of $\hat{\overline{M}}$, $\hat{M}_T$, $\hat{C}$, and $\hat{W}$ for day-permit anglers were about 7%, but that of $\hat{R}^{(s)}$ was reduced from 19.7% to 6.6% by not considering the covariance terms. The CVs of $\hat{C}$ and $\hat{W}$ dropped about 10% from 21% without the covariance. Similar point estimates and smaller variance estimates were obtained. The variance estimate of the annual catch obtained by method 1 with covariances (1.2604×10^{12}) was 4.11 times larger than that by method 2 (3.0667×10^{11}).

The relationship between the sample size and the precision of the annual catch estimate for season-permit anglers was examined. We calculated the values $\hat{V}(\hat{C})$ for various values of n by using Equation 7. To obtain precision over the season for CVs of $\hat{C}^{(s)}\left(=\sqrt{\hat{V}(\hat{C})}/\hat{C}\right)$ below 10%, a sample size of 120 or more is required (Table 3).

A high positive correlation in catches between adjacent months was detected (Table 4). We mapped anglers (objects) and fishing days (categories) into a two-dimensional graph by correspondence analysis (Hayashi, 1950; Benzécri, 1992) using the function "pq3.prcomp" in S version 4 (Chambers and Hastie, 1992). Correspondence analysis showed the relations between rows and columns in a frequency table graphically as points in a common low-dimensional space (Clausen, 1998). Both objects (rows) and categories (columns) of variables are represented as points in such a way that an object is relatively close to its category and relatively far from other categories (Leeuw and van Rijckevorsel 1988). For example, the 72nd angler fished 10 days in June, five days in July, seven days in August, three days in September, one day in October, and this angler was mapped closed to June, reflecting the month of his highest fishing effort (Fig. 7). The results suggest several fishing patterns with high catch seasons in June–July, July–

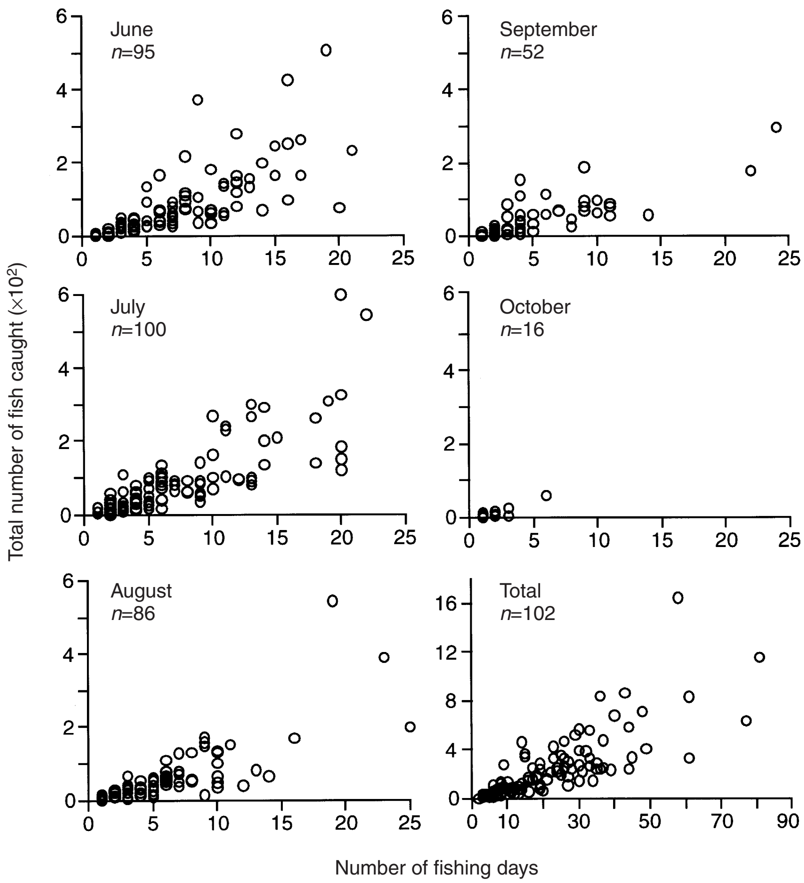

Figure 4

Relationship between the number of fishing days and the number of fish caught by 104 sampled season-permit anglers in the Nakagawa River.

August, August–September, and September–October, resulting high correlations between adjoining months, and large covariances between distant months (Table 4).

The prime advantage of a longitudinal study is its effectiveness for studying change, and a repeated measures analysis of variance can be applied to a complete data set with a constant correlation (Diggle et al., 1994). However, our data set was incomplete because the number of anglers who fish in each month changed (see n in Fig. 4) and had a different correlation structure among month (Table 4). We tested the differences between successive monthly catch estimates of season-permit anglers by using a parametric bootstrapping method. In the central limit theorem, the sample distribution of a monthly total catch estimate can be regarded as a normal with the mean $C_k^{(s)}$ and the variance $\hat{V}(\hat{C}_k^{(s)})$. Based on the two point estimates, variance estimates and the correlation coefficient between successive two months, we generated 10,000 bivariate normal random variables (Gentle, 1998). The means and 95% confidence intervals of the differences between two monthly total catch estimates were –226,561 [–650,404~203,080] for June and July, 870,720 [455,091~1,277,402] for July and August, 470,594 [161,013~783,168] for August and September, and 537,488 [290,905~782,727] for September and October. Significance levels were corrected for multiple testing by using the Bonferroni ajustment factor (Sokal and Rohlf, 1995).The confidence interval for June and July straddled 0, showing no significant difference. On the other hand, three other confidence intervals did not include 0; therefore the monthly differences were statistically significant (Fig. 8).

Discussion

Bias and source of variation

The estimate of the total annual catch of ayu by the recreational fishery was the first obtained in Japan and was much larger than expected. The total number of day-permits sold was 9066, and was quite small (1.9%) compared with the estimated total number of anglers (477,520).

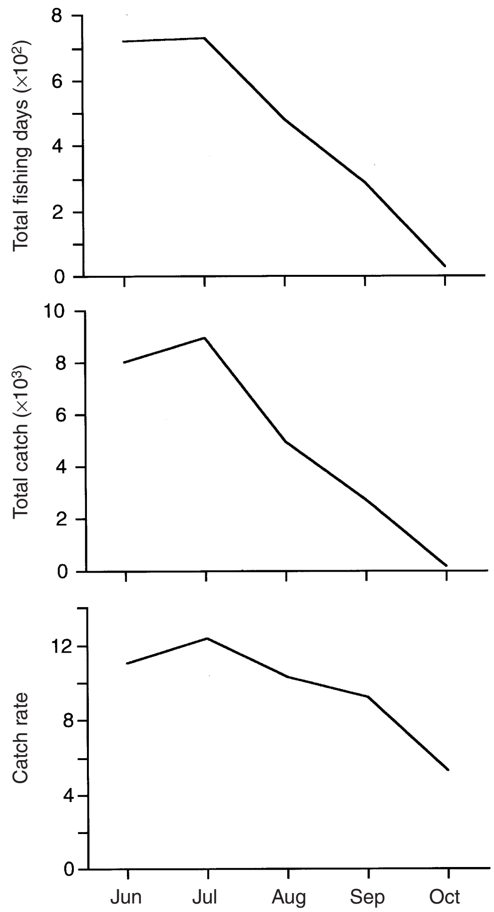

Figure 5

Monthly changes in the number of fishing days, total catch, and catch rate for 104 sampled anglers in the Nakagawa River.

Although the difference in the catch rate between day- and season-permit anglers was unknown, the influence of this bias on the total catch estimates would be minor. In order to check the bias, however, one could conduct a logbook survey of day-permit anglers. Sixteen anglers of the total sample (13%) in our study did not return logbooks and therefore may have caused a bias in our estimates; however no attempt was made to evaluate the difference between nonrespondents and respondents. The angler sample was drawn arbitrarily by the cooperatives but was not a random sample in the strictest sense. If cooperative anglers tended to be selected, this could have been a source of bias.

The source of variation in total catch is the variation in the catch of the sampling unit, including differences in fishing days, skill of the anglers, and the number of anglers that a party boat could accommodate. A stratified sampling scheme based on categories of anglers or boats is effective for this situation. The weakest point in the use of logbook surveys, perhaps, is that the catch data are reported by those who catch the fish and by boat owners with monetary interests. To what extent the anglers might have exaggerated or under-reported their catch is not known. Party boat owners may record lower than actual catches to reduce taxes. To examine this possible source of bias, on-site surveys should be conducted. For the ayu fishery in the Nakagawa River, an access point survey may be practical (Pollock et al., 1994). When comparatively complete lists of boat owners and anglers are available, logbook surveys based on these lists, combined with on-site surveys, are appropriate.

Longitudinal and stratified survey designs

Longitudinal surveys taking repeated measures on the same random sample over time are better than successive independent surveys if the designer's goal is to show statistically significant differences in the estimates between time intervals. Monthly estimates showing seasonal trends

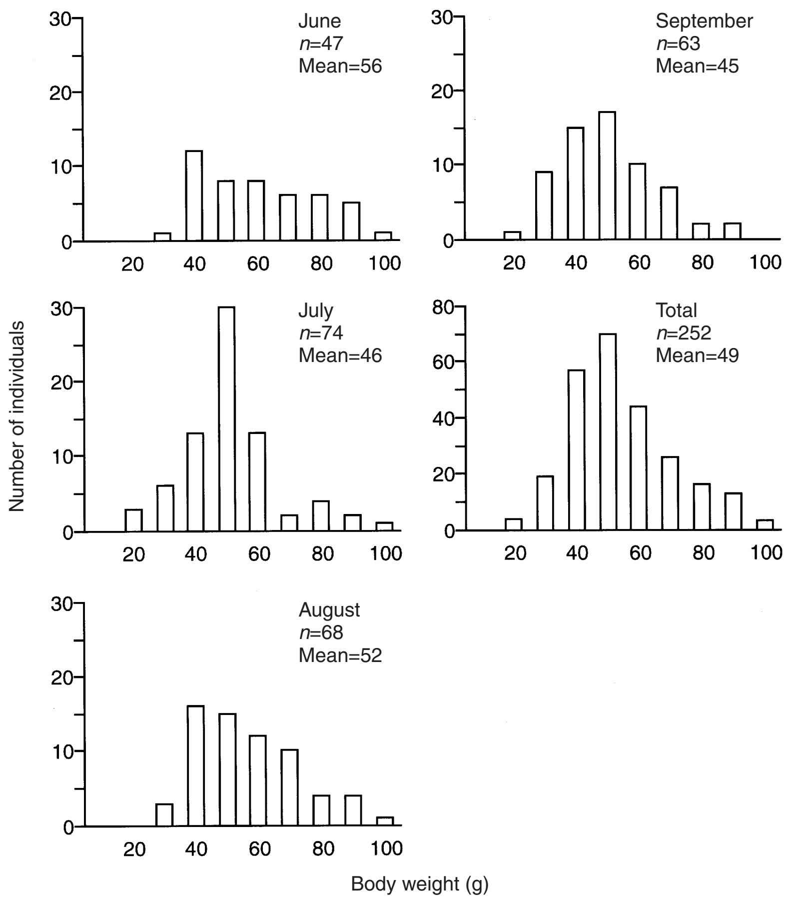

Figure 6
Distributions of body weight by month for ayu caught in the Nakagawa River.

can be obtained by the equations derived in our study. In such repeated measure designs, the most precise estimates of annual catch are obtained by method 2. On the other hand, stratified surveys, or successive independent surveys between time intervals based on random samples of anglers or boats, are better than longitudinal surveys if the designer's goal is to obtain precise estimates of total effort, or catch (or total effort and catch), over the entire season. Stratification by month would improve the precision of annual estimates even more if estimates varied greatly across months. Stratified sampling allows independent monthly estimates, and monthly estimates can be summed to produce precise estimates over time. In the absence of correlations between monthly sample observations, the estimated variance of annual estimates can be obtained simply by adding the estimated variances of the monthly estimates. The estimated variances of annual estimates stratified by month would be considerably less than those of annual estimates based on repeated monthly observations of a one-time annual sample.

If method 2 is used to analyze data obtained by such independent surveys, how would the precision of the estimator compare with the precision of a stratified estimator? For simplicity, we consider a population that is divided into two subpopulations of N_1, N_2 units, respectively. The stratified estimator of the population total and the respective variance are

$$\hat{Y} = N_1\bar{y}_1 + N_2\bar{y}_2,$$
$$\hat{V}(\hat{Y}) = N_1^2\hat{V}(\bar{y}_1) + N_2^2\hat{V}(\bar{y}_2),$$

where $\bar{y}_1$ and $\bar{y}_2$ are the sample means for sample sizes of n_1 and n_2. On the other hand, those obtained by method 2 are

$$\hat{Y}_c = \frac{N_1+N_2}{n_1+n_2}(n_1\bar{y}_1 + n_2\bar{y}_2),$$
$$\hat{V}(\hat{Y}_c) = \left\{\frac{n_1(N_1+N_2)}{n_1+n_2}\right\}^2 \hat{V}(\bar{y}_1) + \left\{\frac{n_2(N_1+N_2)}{n_1+n_2}\right\}^2 \hat{V}(\bar{y}_2).$$

Subtracting Y_c from Y, we have

Table 2

Estimated parameters and coefficients of variation (in parentheses). Total$^{\text{meth. 1}}$ = total estimated by summing monthly estimates (method 1). Total$^{\text{meth. 2}}$ = total estimated by combining data throughout the season (method 2). $R^{(s)}$ = catch rate for season-permit anglers. $\overline{M}$ = mean number of fishing days per season-permit anglers. M_T = total number of fishing days (season+day-permit anglers). $C^{(s)}$ = total catch in number for season-permit anglers. $C^{(d)}$ = total catch in number for day-permit anglers. C = total catch in number (season+day-permit anglers). $W^{(s)}$ = total catch in weight for season-permit anglers. $W^{(d)}$ = total catch in weight for day-permit anglers. W = total catch in weight (season+day-permit anglers).

Parameter	June	July	August	September	October	Total$^{\text{meth. 1}}$	Total$^{\text{meth. 2}}$
$R^{(s)}$	11.12	12.37	10.29	9.29	5.35	11.04	11.04
	(0.087)	(0.075)	(0.115)	(0.187)	(1.495)	(0.197)	(0.066)
$\overline{M}$	6.92	7.05	4.62	2.80	0.28	21.66	21.65
	(0.073)	(0.075)	(0.101)	(0.153)	(0.291)	(0.073)	(0.073)
M_T	157,231	157,047	102,722	61,687	6,152	484,839	484,628
	(0.071)	(0.074)	(0.100)	(0.152)	(0.290)	(0.072)	(0.071)
$C^{(s)}$	1,690,440	1,914,495	1,042,772	570,800	32,732	5,251,241	5,251,241
	(0.116)	(0.115)	(0.154)	(0.179)	(0.437)	(0.214)	(0.105)
$C^{(d)}$	57,658	27,915	13,982	2,186	152	101,894	100,108
	(0.087)	(0.077)	(0.118)	(0.197)	(1.528)	(0.056)	(0.066)
C	1,748,099	1,942,410	1,056,755	572,987	32,884	5,353,135	5,351,349
	(0.112)	(0.114)	(0.152)	(0.178)	(0.435)	(0.210)	(0.103)
$W^{(s)}(t)$	94.64	88.30	53.26	25.85	1.48	263.53	253.90
	(0.125)	(0.122)	(0.159)	(0.184)	(0.439)	(0.213)	(0.107)
$W^{(d)}(t)$	3.23	1.29	0.71	0.10	0.01	5.34	4.84
	(0.099)	(0.086)	(0.124)	(0.202)	(1.530)	(0.065)	(0.069)
$W(t)$	97.86	89.59	53.98	25.95	1.49	268.86	258.74
	(0.122)	(0.120)	(0.157)	(0.183)	(0.437)	(0.209)	(0.105)

Table 3

Coefficient of variations of the total catch estimate $\hat{C}^{(s)}$ (method 2) for various sample sizes (number of anglers).

n	CV	n	CV	n	CV
10	0.3401	110	0.1025	230	0.0709
20	0.2405	120	0.0982	250	0.0680
30	0.1963	130	0.0943	300	0.0621
40	0.1700	140	0.0909	400	0.0538
50	0.1521	150	0.0878	500	0.0481
60	0.1388	160	0.0850	600	0.0439
70	0.1285	170	0.0824	700	0.0406
80	0.1202	180	0.0802	800	0.0380
90	0.1134	190	0.0780	900	0.0358
100	0.1075	200	0.0760	1000	0.0340

Table 4

Estimated variance-covariance matrix (×10^{10}) for the monthly estimates of catch (number) by season-permit anglers in number by Equation 17 (lower diagonal) and the correlation coefficient $r(C_{ik}, C_{ik'})$ (upper diagonal, in bold font). The diagonal component refers to $\hat{V}(\hat{C}_k^{(s)})$ and the lower half refers to $\widehat{\mathrm{Cov}}(\hat{C}_k^{(s)}, \hat{C}_{k'}^{(s)})$.

Month	Jun	Jul	Aug	Sep	Oct
Jun	3.837	**0.67**	**0.43**	**0.30**	**–0.03**
Jul	6.011	4.862	**0.65**	**0.49**	**0.08**
Aug	9.422	4.434	2.578	**0.65**	**0.28**
Sep	10.681	6.258	1.456	1.038	**0.44**
Oct	10.725	6.336	1.522	0.069	0.020

$$\hat{Y} - \hat{Y}_c = \frac{n_2 N_1 - n_1 N_2}{n_1 + n_2}(\bar{y}_1 - \bar{y}_2),$$

showing that the two methods provide different estimates with the extent of the difference, depending upon the sizes of the strata, the sample sizes, and the estimates of the sample means. According to Cochran (1977), method 2 works well enough if the sample allocation is proportional because a simple random sample distributes itself approximately proportionally among strata. With proportional allocation $N_1/n_1 = N_2/n_2$; therefore the difference of the two estimators is 0. In our case study, the annual

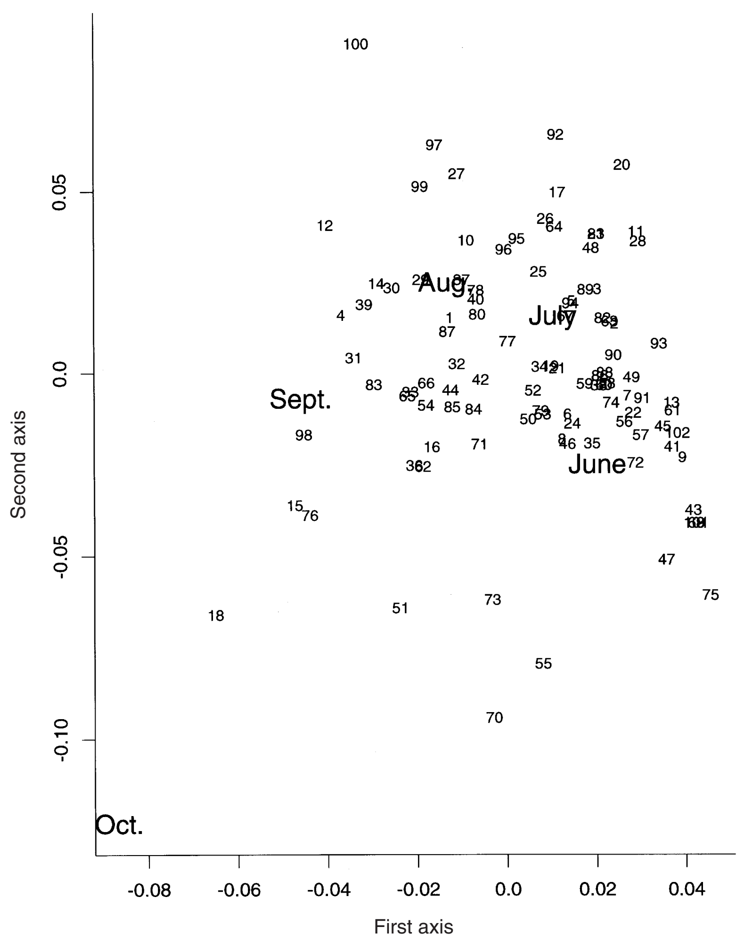

Figure 7

Plots of the determined quantities for sample (anglers) and category (months) from the correspondence analysis for the fishing days of 104 sampled anglers. The numbers in the figure refer to the number of anglers that were sampled .

catch estimates for the season-permit anglers $C^{(s)}$ were the same value for method 1 and method 2. The population size N and the sample size n were in proportion constant throughout the season; therefore $N_i/n_i = N/n$ was same for all strata. The ratio of the variances is

$$\frac{\hat{V}(\hat{Y})}{\hat{V}(\hat{Y}_c)} = \frac{(n_1+n_2)^2\left\{N_1^2\hat{V}(\bar{y}_1)+N_2^2\hat{V}(\bar{y}_2)\right\}}{(N_1+N_2)^2\left\{n_1^2\hat{V}(\bar{y}_1)+n_2^2\hat{V}(\bar{y}_2)\right\}},$$

which also shows that the precision of both methods depends on the sizes of the strata, the sample sizes, and the variance of the sample means. If the allocation is proportional, the variance ratio becomes 1 and the two variances coincide. The objective of stratified surveys is to obtain precise total effort or catch estimates (or both) over the entire season. A proportional sample allocation is recommended, which allows a simple calculation with method 2 and improves precision of estimates at the same time.

Acknowledgements

We extend our gratitude to the anonymous referees and to John V. Merriner, Katherine Myers, and Sharyn Matriotti for their critical readings that greatly improved the manuscript. We also thank John M. Hoenig, John B. Pearce, Yashushi Taga, and Geoff Gordon for their comments on earlier versions of the manuscript. A portion of this study was funded by the Fisheries Agency of Japan (Japan Sea-Farming Association).

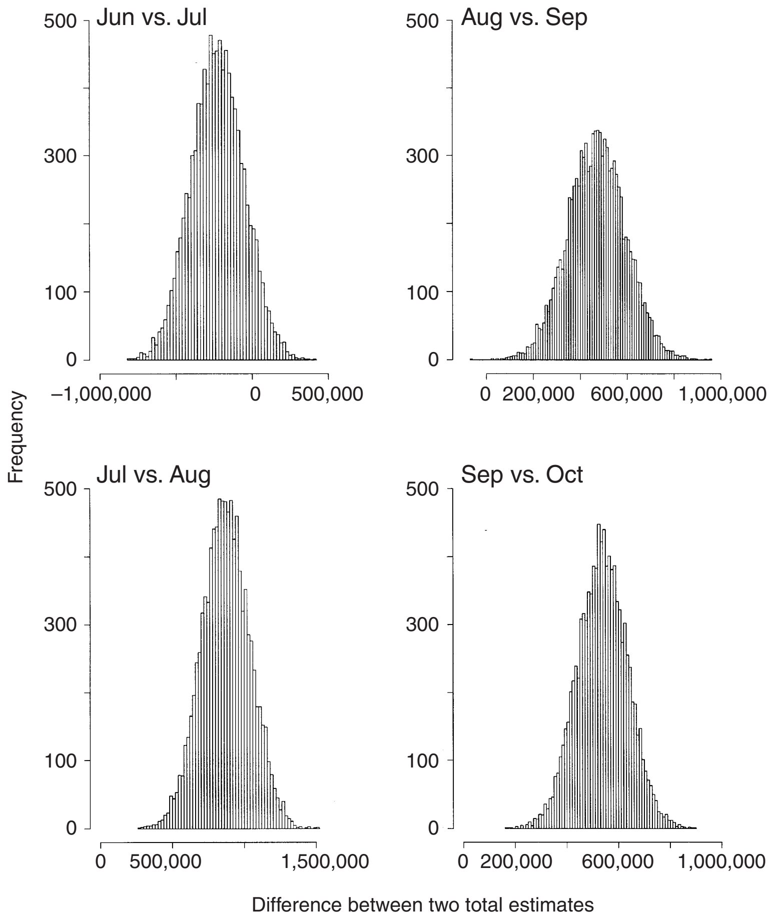

Figure 8

Bootstrap distributions of diffences between successive monthly ayu catch estimates.

Literature cited

Benzécri, J. P.
1992. Correspondence analysis handbook. Dekker, New York, NY, 665 p.

Chambers J. M., and, T. J. Hastie (eds.)
1992. Statistical models in S. Chapman and Hall, New York, NY, 608 p.

Clausen, S.-E.
1998. Applied correspondence analysis: an introduction. SAGE Publications, Thousand Oaks, CA, 69 p.

Cochran, W. G.
1977. Sampling techniques, third ed. John Wiley and Sons, New York, NY, 413 p.

Diggle, P. J, K-Y. Liang, and S. L. Zeger.
1994. Analysis of longitudinal data. Oxford Univ. Press, Oxford, 253 p.

Gentle, J. E.
1998. Random number generation and monte carlo methods. Springer, New York, NY, 247 p.

Goodman, L. A.
1960. On the exact variance of products. J. Am. Stat. Assoc. 55:708–713.

Guthrie, D., J. M. Hoenig, M. Holliday, C. M. Jones, M. J. Mills, S. A. Moberly, K. H. Pollock, and D. R. Talhelm (eds).
1991. Creel and angler surveys in fisheries management. Am. Fish. Soc. Symp. 12, Bethesda, MD, 528 p.

Hayashi, C.
1950. On the prediction of phenomena from mathematical statistic point of view. Annals Inst. Stat. Math. 3:69–98.

Jones, C. M., D. S. Robson, H. D. Lakkis, and J. Kressel.
1995. Properties of catch rates used in analysis of angler surveys. Trans. Am. Fish. Soc. 124: 911–928.

Leeuw, J. de, and J. L. A. van Rijckevorsel.
1988. Beyond homogeneity analysis. *In* Component and correspondence analysis (J. L. A. van Rijckevorsel and J. de Leeuw, eds.), p. 56–57. John Wiley and Sons, New York, NY.

Pollock, K. H., J. M. Hoenig, C. M. Jones, D. S. Robson, and C. J. Greene.
1997. Catch rate estimation for roving and access point surveys. N. Am. J. Fish. Manag. 17:11–19.

Pollock, K. H., C. M. Jones, and T. L. Brown.
1994. Angler survey methods and their applications in fisheries management. Am. Fish. Soc. Special Publication 25, Bethesda, MD, 371 p.
Seber, G. A. F.
1982. The estimation of animal abundance and related parameters, second ed. Griffin, London, 654 p.
Sokal, R. R., and F. J. Rohlf.
1995. Biometry, third ed. Freeman and Company, New York, NY, 887 p.
Thompson, S. K.
1992. Sampling. John Wiley and Sons, New York, NY, 343 p.

Appendix 1: The covariance of two estimators from sample means

First we consider the covariance of two total estimates. Let X_i and Y_i be simple random samples (i=1, . . . , n) from a population of size N with mean μ_x and μ_y, and $\overline{X}$ and $\overline{Y}$ be two sample means.

Cochran (1977, p .25) derived the covariance of two-sample mean, that is

$$\mathrm{Cov}(\overline{X},\overline{Y}) = \frac{N-n}{N}\frac{1}{n}\mathrm{Cov}(X,Y)$$
$$= \frac{N-n}{N^2 n}\sum_{i=1}^{N}(X_i-\mu_x)(Y_i-\mu_u).$$

This is estimated by

$$\widehat{\mathrm{Cov}}(\overline{X},\overline{Y}) = \frac{N-n}{N}\frac{1}{n}\widehat{\mathrm{Cov}}(X,Y)$$
$$= \frac{N-n}{Nn(n-1)}\sum_{i=1}^{n}(X_i-\overline{X})(Y_i-\overline{Y}).$$

The covariance between two population total estimators is defined by

$$\mathrm{Cov}(\hat{X},\hat{Y}) = \mathrm{Cov}(N\overline{X},N\overline{Y}) = E(N\overline{X}-N\mu_x)(N\overline{Y}-N\mu_y)$$
$$= N^2\mathrm{Cov}(\overline{X},\overline{Y}) = \frac{N(N-n)}{n}\mathrm{Cov}(X,Y).$$

This is estimated by

$$\widehat{\mathrm{Cov}}(\hat{X},\hat{Y}) = \frac{N(N-n)}{n(n-1)}\sum_{i=1}^{n}(X_i-\overline{X})(Y_i-\overline{Y}).$$

For the monthly total catches, we get

$$\widehat{\mathrm{Cov}}(\hat{C}_k^{(s)},\hat{C}_{k'}^{(s)}) = \frac{N(N-n)}{n(n-1)}\sum_{i=1}^{n}(C_{ik}-\hat{\overline{C}}_k)(C_{ik'}-\hat{\overline{C}}_{k'}).$$

Appendix 2: Approximate covariance between $\hat{W}_k^{(s)}$ and $\hat{W}_{k'}^{(s)}$

Taylor's series of $\hat{W}_k$ with respect to the random variables is obtained by

$$\hat{W}_k \simeq C_k\overline{w}_k + \overline{w}_k(\hat{C}_k - C_k) + C_k(\hat{\overline{w}}_k - \overline{w}_k).$$

From Taylor's series (mentioned above), the approximate covariance is obtained.

$$\mathrm{Cov}(\hat{W}_k^{(s)},\hat{W}_{k'}^{(s)}) = E(\hat{W}_k^{(s)} - W_k^{(s)})(\hat{W}_{k'}^{(s)} - W_{k'}^{(s)})$$
$$\simeq E\left[\overline{w}_k(\hat{C}_k^{(s)} - C_k^{(s)}) + C_k^{(s)}(\hat{\overline{w}}_k - \overline{w}_k)\right]$$
$$\left[\overline{w}_{k'}(\hat{C}_{k'}^{(s)} - C_{k'}^{(s)}) + C_{k'}^{(s)}(\hat{\overline{w}}_{k'} - \overline{w}_{k'})\right]$$
$$= \overline{w}_k\overline{w}_{k'}\mathrm{Cov}(\hat{C}_k^{(s)},\hat{C}_{k'}^{(s)}) + \overline{w}_k C_{k'}^{(s)}\mathrm{Cov}(C_k^{(s)},\hat{\overline{w}}_{k'})$$
$$+ C_k^{(s)}\overline{w}_{k'}\mathrm{Cov}(\hat{\overline{w}}_k,\hat{C}_{k'}^{(s)}) + C_k^{(s)}C_{k'}^{(s)}\mathrm{Cov}(\hat{\overline{w}}_k,\hat{\overline{w}}_{k'}).$$

Here $\hat{C}_{(.)}^{(s)}$ and $\hat{\overline{w}}_{(.)}$ are independent, and both w_k and $w_{k'}$ are estimated from different samples. Therefore Cov $(\hat{C}_k^{(s)},\hat{\overline{w}}_{k'}) = \mathrm{Cov}(\hat{\overline{w}}_k,\hat{C}_{k'}^{(s)}) = \mathrm{Cov}(\hat{\overline{w}}_k,\hat{\overline{w}}_{k'}) = 0$, then we get the covariance as only the first term.

Appendix 3: Approximate variance of $\hat{R}$

Taylor's series of $\hat{R}$ with respect to $\hat{C}$ and $\hat{M}$ is obtained by

$$\hat{R} \simeq \frac{C}{M} + \frac{1}{M}(\hat{C}-C) - \frac{C}{M^2}(\hat{M}-M).$$

Then the approximate variance is obtained by

$$V(\hat{R}) = E(\hat{R}-R)^2 \simeq \frac{1}{\hat{M}^2}V(\hat{C}) + \frac{\hat{C}^2}{\hat{M}^4}V(\hat{M})$$
$$-2\frac{\hat{C}}{\hat{M}^3}\mathrm{Cov}(\hat{C},\hat{M}).$$

Appendix 4: Covariance between $\hat{C}_k^{(d)}$ *and* $\hat{C}_{k'}^{(d)}$

By expanding $\hat{C}_k^{(d)}$ and $\hat{C}_{k'}^{(d)}$ we get

$$\hat{C}_k^{(d)} \simeq Dp_kR_k + DR_k(\hat{p}_k - p_k) + Dp_k(\hat{R}_k - R_k).$$
$$\hat{C}_{k'}^{(d)} \simeq Dp_{k'}R_{k'} + DR_{k'}(\hat{p}_{k'} - p_{k'}) + Dp_{k'}(\hat{R}_{k'} - R_{k'});$$

then the approximate covariance is given by

$$\mathrm{Cov}(\hat{C}_k^{(d)},\hat{C}_{k'}^{(d)}) = E(\hat{C}_k^{(d)} - Dp_kR_k)(\hat{C}_{k'}^{(d)} - Dp_{k'}R_{k'})$$
$$\simeq D^2\begin{bmatrix} R_kR_{k'}\mathrm{Cov}(\hat{p}_k,\hat{p}_{k'}) + R_kp_{k'}\mathrm{Cov}(\hat{p}_k,\hat{R}_{k'}) \\ +p_kR_{k'}\mathrm{Cov}(\hat{R}_k,\hat{p}_{k'}) + p_kp_{k'}\mathrm{Cov}(\hat{R}_k,\hat{R}_{k'}) \end{bmatrix}.$$

If the first three covariance components are equal to 0 because of independent sampling, then we have

$$\mathrm{Cov}(\hat{C}_k^{(d)},\hat{C}_{k'}^{(d)}) \simeq D^2p_kp_{k'}\mathrm{Cov}(\hat{R}_k,\hat{R}_{k'}).$$

Here we can write $\hat{R}_k$ and $\hat{R}_{k'}$ as the ratio of two random variables from Equation 1 by

$$\hat{R}_k = \frac{\hat{C}_k^{(s)}}{N\hat{\overline{M}}_k}.$$

By using a method similar to that given in Appendix 3, we get

$$\mathrm{Cov}(\hat{R}_k, \hat{R}_{k'}) \simeq \frac{1}{N^4}\begin{pmatrix} \frac{1}{\hat{\overline{M}}_k \hat{\overline{M}}_{k'}} \mathrm{Cov}(\hat{C}_k^{(s)}, \hat{C}_{k'}^{(s)}) \\ -\frac{\hat{C}_{k'}^{(s)}}{\hat{\overline{M}}_k \hat{\overline{M}}_{k'}^2} \mathrm{Cov}(\hat{C}_k^{(s)}, \hat{\overline{M}}_{k'}) \\ -\frac{\hat{C}_k^{(s)}}{\hat{\overline{M}}_k^2 \hat{\overline{M}}_{k'}} \mathrm{Cov}(\hat{\overline{M}}_k, \hat{C}_{k'}^{(s)}) \\ +\frac{\hat{C}_k^{(s)}\hat{C}_{k'}^{(s)}}{\hat{\overline{M}}_k^2 \hat{\overline{M}}_{k'}^2} \mathrm{Cov}(\hat{\overline{M}}_k, \hat{\overline{M}}_{k'}) \end{pmatrix}.$$

Here $\widehat{\mathrm{Cov}}(\hat{C}_k^{(s)}, \hat{C}_{k'}^{(s)})$ and $\widehat{\mathrm{Cov}}(\hat{\overline{M}}_k, \hat{\overline{M}}_{k'})$ are already given by Equation 18 and Equation 24. Hence we can estimate $\mathrm{Cov}(\hat{C}_k^{(s)}, \hat{\overline{M}}_{k'})$ and $\mathrm{Cov}(\hat{\overline{M}}_k, C_{k'}^{(s)})$. These are the covariances between a total estimate and a sample mean. By a method similar to that in Appendix 1, we have

$$\mathrm{Cov}(\hat{X}, \overline{Y}) = E(N\overline{X} - N\mu_x)(\overline{Y} - \mu_y)$$
$$= N\mathrm{Cov}(\overline{X}, \overline{Y}) = \frac{N(N-n)}{n}\mathrm{Cov}(X, Y).$$

The covariance is estimated by

$$\widehat{\mathrm{Cov}}(\hat{X}, \overline{Y}) = \frac{N-n}{n(n-1)}\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y}).$$

For our case, the two covariances are as follows;

$$\widehat{\mathrm{Cov}}(\hat{C}_k^{(s)}, \hat{\overline{M}}_{k'}) = \frac{N-n}{n^2(n-1)}\sum_{i=1}^{n}(C_{ik} - \hat{\overline{C}}_k)(M_{ik'} - \hat{\overline{M}}_{k'})$$

$$\widehat{\mathrm{Cov}}(\hat{\overline{M}}_k, \hat{C}_{k'}^{(s)}) = \frac{N-n}{n^2(n-1)}\sum_{i=1}^{n}(C_{ik'} - \hat{\overline{C}}_{k'})(M_{ik} - \hat{\overline{M}}_k).$$