Erratum

Statistical estimation of mean values of fish stock indicators from trawl surveys

John COTTER[a]

FishWorld Science, Lowestoft, NR33 7LH, UK

Redefining the "geographic" mean

I reconsider the definition of the "geographic" mean for an indicator, y, such as fish length, given as Eq. (2) in the original paper. It can be re-written as

$$\bar{Y}_{\text{geog}} = \frac{\sum_{i=1}^{A} \sum_{j=1}^{n_i} y_{ij}}{N}$$

revealing that, among the A all-inclusive "small plots" in the region, those having the most fish contribute most to the geographic mean, as is to be expected, and that the geographic and population means (Eq. (1)) are equal: $\bar{Y}_{\text{geog}} = \bar{Y}_{\text{pop}} = N^{-1} \sum_{j=1}^{N} y_j$. Concerning the estimators, $\bar{y}_{\text{stn}}$ (Eq. (4)) for $\bar{Y}_{\text{geog}}$ is not consistent, whereas $\bar{y}_{\text{fish}}$ (Eq. (3)) for $\bar{Y}_{\text{pop}}$ is consistent meaning that $\bar{y}_{\text{fish}} \to \bar{Y}_{\text{pop}}$ as the number of fish caught increases. Consistency is a reassuring property for an estimator to have.

An alternative definition for a geographic mean is

$$\bar{Y}_{\text{geog}}^{*} = \frac{\sum_{i=1}^{A} n_i^{-1} \sum_{j=1}^{n_i} y_{ij}}{A}, \quad n_i > 0 \text{ for all } i.$$

In contrast, it weights the small plots equally without regard to the number of fish in each provided that they are all inhabited. With this new definition, $\bar{Y}_{\text{geog}}^{*} \neq \bar{Y}_{\text{pop}}$ but inspection of Eq. (4) shows that $\bar{y}_{\text{stn}}$ is a consistent estimator for $\bar{Y}_{\text{geog}}^{*}$, meaning that, if trawl towing paths are assumed equivalent to small plots and do not overlap, $\bar{y}_{\text{stn}} \to \bar{Y}_{\text{geog}}^{*}$ as the number of populated, fished stations (n_{stn} in Eq. (4)) is increased. $\bar{Y}_{\text{geog}}^{*}$ is therefore comparable with $\bar{Y}_{\text{pop}}$ in having a consistent estimator.

Sampling efficiency is a good reason to prefer $\bar{Y}_{\text{geog}}^{*}$ and $\bar{y}_{\text{stn}}$ to $\bar{Y}_{\text{pop}}$ and $\bar{y}_{\text{fish}}$. $\bar{y}_{\text{stn}}$ is biased slightly upwards of $\bar{Y}_{\text{geog}}^{*}$ when the trawl fails to detect habitation at some stations that are, consequently, excluded from n_{stn}. However, the station means that are averaged by $\bar{y}_{\text{stn}}$ always have fixed, equal weights that introduce no variance of their own to the estimate. On the other hand, $\bar{Y}_{\text{pop}}$, is estimated by $\bar{y}_{\text{fish}}$ which, re-written from Eq. (3) as

$$\bar{y}_{\text{fish}} = \frac{\sum_{i=1}^{n_{\text{stn}}} n_i \bar{y}_i}{n_{\text{fish}}},$$

can be seen to have the station weights, n_i / n_{fish}. Such weights are abundance-based random variables likely to add high

[a] Corresponding author: john.cotter@phonecoop.coop

variation to that of the station means, $\bar{y}_i$. Dependence between numerator and denominator of the weights, and between the n_i and the y_{ij} measured at the i'th station, e.g. when small fish are most numerous, would add further variability from survey to survey. Sampling efficiency of $\bar{y}_{\text{fish}}$ is therefore likely to be poor. In support of this reasoning, $\bar{y}_{\text{stn}}$ usually gave the highest, and $\bar{y}_{\text{fish}}$ the most variable estimate in Figure 1. $\bar{y}_{\text{mix}}$ (Eq. (7)) offers a compromise between the precision of $\bar{y}_{\text{stn}}$ and the catch-weighted $\bar{y}_{\text{fish}}$ for reasons given in the paper.

The weighting of sampling strata, referred to in the discussion (Cotter 2009, p. 132), is also affected by the new definition, $\bar{Y}_{\text{geog}}^*$. When using $\bar{y}_{\text{stn}}$, strata would best be weighted by their inhabited areas, not their total areas, to avoid bias of the stratified mean estimate of $\bar{Y}_{\text{geog}}^*$.

Reference

Cotter J., 2009, Statistical estimation of mean values of fish stock indicators from trawl surveys. Aquat. Living Resour. 22, 127–133.