Variance and CV Estimators for the Discard Ratios and Fleet-total Discards in the WCGOP Bycatch Reports

Yong-Woo Lee*

Fishery Observation Science Program
Fisheries Monitoring and Analysis Division
Northwest Fisheries Science Center
National Oceanic and Atmospheric Administration

(last update on Oct. 7, 2015)

Part 1.1: Estimation of variance for the discard ratios and fleet-total discards per stratum

For the fleet-total expanded discards (or bycatch) estimation of fish or invertebrate species, WCGOP uses a deterministic ratio estimator approach. The effort metric is the retained weight of target species or target species groups (e.g., sable fish, pink shrimp, or FMP-listed groundfish species-except hake). Target species are fishery sector specific, so can be different between the specific sectors.

The ratio of discard (or bycatch) of a species in a given stratum is estimated by division of discarded amount of a species with the amount of retained target species.

$$\hat{R} = \frac{\sum_{h=1}^{n} d_h}{\sum_{h=1}^{n} w_h}$$

where,
$\hat{R}$ = Estimated discard ratio (of a species in a given stratum)
h = Haul identifier
n = Total number of hauls (in a given stratum)
d_h = Observed discarded weight (or count)of a species (in a given stratum)
w_h = Observered retained wieght of target species (in a given statum)

Once the discard ratio of a species is estimated in a stratum, then the fleet-total discards of a species in that stratum can be estimated subsequently based on the expansion with the total

* If any erratum is found, please contact the author via email (Yong-Woo.Lee@noaa.gov)

landed weight of target species (as the measure of total effort). Information on total landings is available from landing receipts (a.k.a fish tickets).

$$\widehat{D}_s = \hat{R}_s \cdot K_s$$

where,
$s =$ strata identifier
$\widehat{D} =$ Estimated total discards (of a species per given stratum)
$\hat{R} =$ Estimated discard ratio (of a species per given stratum)
$K =$ Total landed weight of target species (per given stratum)

We can estimate the variance (v) of total discards ($\widehat{D}$) based on the estimated variance of bycatch ratio ($\hat{R}$) from a sample. Note that K (total landed weight of target species) is assumed to be known without uncertainty. The following formula is based on the general approximate variance formula for ratio estimate (Cochran 1977).

$$v(\hat{R}) = \frac{N-n}{nN} \cdot \frac{1}{(\overline{w})^2} \cdot \left\{ {s_d}^2 + \hat{R}^2 \cdot {s_w}^2 + 2 \cdot \hat{R} \cdot \rho \cdot s_d \cdot s_w \right\}$$

where,
$\hat{R} =$ Estimated discard ratio
$N =$ Total number of hauls in a stratum
$n =$ Number of observed hauls
$\overline{w} =$ Average retained weight of target species from observed hauls
$s_d =$ Sample standard deviation of discard
$s_w =$ Sample standard deviation of retained target species
$\rho =$ Correlation coefficient between d and w

Another approximation for the variance estimate of discard ratio is possible. Ratio can be considered as the product of two variables (i.e., the numerator and the inverse of denominator). Thus, we can apply the method of unbiased exact variance of a product of two variables (Goodman 1960), in conjunction with the Taylor series variance approximation approach (a.k.a Delta method) on the denominator of the ratio, to derive a variance estimator for discard ratio, as described in Pitkitch et al. (1998). This is supposedly an enhanced version of variance estimator of the ratio from a sample.

$$v(\hat{R})^* = \left(\frac{\bar{d}}{\overline{w}}\right)^2 \cdot \left\{ \frac{{s_{\bar{d}}}^2}{\bar{d}^2} + \frac{{s_{\overline{w}}}^2}{\overline{w}^2} - \left(\frac{{s_{\bar{d}}}^2}{\bar{d}^2} \cdot \frac{{s_{\overline{w}}}^2}{\overline{w}^2} \right) \right\}$$

It should be noted that, the latter method is adopted in the WCGOP reports.

Once the variance of discard ratio is available, then the variance (v) of fleet-total discards of a species per given strata can be estimated as follows. Note that the total landed weight (K) of target species is assumed to be a known quantity without error.

$$v(\hat{D}) = v(\hat{R} \cdot K) = K^2 \cdot v(\hat{R})$$

Part 1.2: Estimation of CV for the discard ratios and fleet-total discards per stratum

Coefficient of Variation (CV) is a standardized measure of dispersion and defined as the ratio of the standard deviation (σ) and the mean (μ) of the population. CV is estimated from a sample with the sample standard deviation (as an estimate of σ) and the sample average (as an estimate of μ). Thus, the CV of the discard ratio per given stratum can be estimated based on the estimated variance of $\hat{R}$ as follows.

$$CV(\hat{R}) = \frac{\sqrt{v(\hat{R})}}{\hat{R}} = \frac{s(\hat{R})}{\hat{R}}$$

The CV of the fleet-total discards per strata can be easily estimated, based on the estimated variance of $\hat{D}$, as described in the previous section.

$$CV(\hat{D}) = \frac{\sqrt{v(\hat{D})}}{\hat{D}} = \frac{\sqrt{v(K \cdot \hat{R})}}{K \cdot \hat{R}} = \frac{\sqrt{K^2 \cdot v(\hat{R})}}{K \cdot \hat{R}} = \frac{K \cdot \sqrt{v(\hat{R})}}{K \cdot \hat{R}} = \frac{\sqrt{v(\hat{R})}}{\hat{R}} = \frac{s(\hat{R})}{\hat{R}}$$

One thing must be noted here is that the estimated CV of the discard ratio and the estimated CV of fleet-total discards for a given species are identical at the individual stratum level.

Part 2: Estimation of variance and CV for the total discards in combined strata

Oftentimes, the fleet-total expanded discards have to be reported as the combined total across multiple strata. A good example would be the report of coast-wide total discards of a species across the fishery management areas or states. The question of interest is how to estimate the variance and the CV for the combined sum of discards across multiple strata?

According to a statistical theory, the variance of sum of two random variables is the sum of variance of each variable and the covariance between the two variables.

$$V(X_1 + X_2) = V(X_1) + V(X_2) + COV(X_1, X_2)$$

The formula can be generalized to the cases of more than two variables by extending the above.

$$V\left(\sum_{i \neq j} V(X_i + X_j)\right) = \sum_{i \neq j} V(X_i) + V(X_j) + COV(X_i, X_j)$$

If the variables are not independent, then we have to take into account of the covariance components in the variance estimation. However, if the variables are assumed to be independent (or covariance terms are negligible), then the variance of the sum of variables becomes the sum of variances from each variable, because the covariance terms cancel out as zeros under the independence assumption between the variables.

Likewise, we can apply this theory to derive the estimator for the variance of the sum of fleet-total discards from combined strata, based on the assumption of independence, as follows.

$$v\left(\sum_{s} \widehat{D}_s\right) = \sum_{s} v(\widehat{D}_s) = v(\widehat{D}_1) + v(\widehat{D}_2) + \cdots + v(\widehat{D}_s)$$

The above formula can be rewritten in terms of discard ratio and total landed amount, because the fleet-wide total discard is the product of estimated discard ratio ($\hat{R}$) and the total landed weight of target species (*K*).

$$v\left(\sum_{s} \widehat{D}_s\right) = v\left(\sum_{s} \hat{R}_s \cdot K_s\right) = \sum_{s} v(\hat{R}_s \cdot K_s) = v(\hat{R}_1 \cdot K_1) + v(\hat{R}_2 \cdot K_2) + \cdots + v(\hat{R}_s \cdot K_s)$$

Note that the total landed weight (*K*) has no uncertainty (i.e., no error is assumed), thus it can be moved outside of the variance function as a constant. Thus, another way to express the variance formula for the sum of total discards across the multiple strata is,

$$v\left(\sum_{s} \widehat{D}_s\right) = \sum_{s} {K_s}^2 \cdot v(\hat{R}_s) = {K_1}^2 \cdot v(\hat{R}_1) + {K_2}^2 \cdot v(\hat{R}_2) + \cdots + {K_s}^2 \cdot v(\hat{R}_s)$$

Based on the estimated variance, as above, for the sum of total discards for combined strata, an estimator for the CV of the total discards across multiple strata can be easily formulated by treating $\sum D$ as a random variable.

$$CV\left(\sum_s \widehat{D}_s\right) = \frac{\sqrt{v\left(\sum \widehat{D}_s\right)}}{\sum \widehat{D}_s} = \frac{\sqrt{v(\widehat{D}_1) + v(\widehat{D}_2) + \cdots + v(\widehat{D}_s)}}{\widehat{D}_1 + \widehat{D}_2 + \cdots + \widehat{D}_s}$$

$$= \frac{\sqrt{{K_1}^2 \cdot v(\hat{R}_1) + {K_2}^2 \cdot v(\hat{R}_2) + \cdots + {K_s}^2 \cdot v(\hat{R}_s)}}{\widehat{D}_1 + \widehat{D}_2 + \cdots + \widehat{D}_s}$$

$$= \frac{\sqrt{{K_1}^2 \cdot s(\hat{R}_1)^2 + {K_2}^2 \cdot s(\hat{R}_2)^2 + \cdots + {K_s}^2 \cdot s(\hat{R}_s)^2}}{\widehat{D}_1 + \widehat{D}_2 + \cdots + \widehat{D}_s}$$

An estimator for the CV of the sum of ratios across multiple strata can be formulated as well by treating $\sum R$ as a random variable.

$$CV\left(\sum_s \hat{R}_s\right) = \frac{\sqrt{v\left(\sum \hat{R}_s\right)}}{\sum \hat{R}_s} = \frac{\sqrt{v(\hat{R}_1) + v(\hat{R}_2) + \cdots + v(\hat{R}_s)}}{\hat{R}_1 + \hat{R}_2 + \cdots + \hat{R}_s}$$

One interesting thing to note is that it was shown in the previous section that CV($\widehat{D}$) and CV($\hat{R}$) are identical at individual stratum level. However, when the multiple strata are combined, the CV estimates, CV($\sum \widehat{D}_s$) and CV($\sum \hat{R}_s$), of the quantities of $\sum \widehat{D}_s$ and $\sum \hat{R}_s$ are not identical, except one special case where the total landed weight of target species (K) are identical between all strata ($K_1 = K_2 = K_3 = \ldots = K_s$).

References

Cochran, W.G. 1977. Sampling techniques. Wiley, New York, pp. 428.

Goodman, L.A. 1960. On the exact variance of products. Journal of the American Statistical Association 55: 708–713.

Pikitch E.K., J.R. Wallace, E.A. Babcock, D.L. Erickson, M. Saelens, G. Oddsson. 1998. Pacific halibut bycatch in the Washington, Oregon, and California groundfish and shrimp trawl fisheries. North American Journal of Fisheries Management. 18(3): 569-586.