CM 2006/K :16

Raising procedures for discards : Sampling theory

Joël Vigneau

1. Introduction

The “best” raising procedure for discards in fishery science is a method everyone talks about but nobody has ever seen it. The specific workshop on discards (Anon. 2003) proposes formulas that are too simplistic as they do not take into account the within trip variance. The literature is full of attempts and results but they are always specific to particular case studies (Stratoudakis et al, 2001, Cotter et al, 2001) or based on inconsistent hypothesis as demonstrated by Rochet and Trenkel (2005). When some generalities are finally found (Borges *et al.*, 2005), the formulas developed do not correspond to multistage sampling because of the unavailability of essential parameters and because the aim of the study is “*not to identify optimum raising procedures”*. The object here is not to discuss the accuracy of such or such document but to build step by step the reasoning for estimating the total volume of discards, in weight or in number, from the hypothesis to the associated variance. With the objective of clarity, the exact reference to the chapters of the two books always cited (Cochran, 1977 and Thomson, 1992) will be provided and when possible the reference to formulas specified in Cochran’s sampling theory.

Keywords : discards, raising procedure, multistage sampling

Joël Vigneau: Ifremer, avenue du Général de Gaulle, 14520 Port-en-Bessin, France. [Tel: (33)2 31 51 56 00, e-mail : Joel.Vigneau@ifremer.fr]. Stéphanie Mahévas : Ifremer, rue de

2. The basis – ground implementation

The implementation of on-board observers for discards or for total catch estimates generally follows the same protocol. The stratification used is the quarter for the temporal dimension, and métier as defined by the *ad hoc* expert group (Anon. 2005) for the technical dimension. For an accurate discussion on the best stratification for discards purpose, see Tamsett *et al.* (1999). In each of the strata, observers stroll around harbours or contact captains by phone, and try to arrange an observation trip regarding the métier practised by the vessel, the weather forecast and the availability of the observer. In theory, it is then the choice of a vessel followed by the choice of a fishing trip. In theory then, the vessel should be drawn from a list of vessels practising a given métier at a given quarter, and the trip should be drawn from a list of trips operated by the given vessel. This approach has been developed by Tamsett *et al.* (1999) and formulas developed for a three-stage sampling can be found in Wang *et al.* (*in press*). The respect of random is extremely difficult in on-board observers programmes due to all sorts of impediments which description is not the subject of this paper. Cochran (1977) specifies that if the variable of interest, here the volume of discards of a given species, is randomly distributed in the population, the non-random status of the drawing is of less importance. It is hardly the case here if we believe that discarding conforms to some underlying hidden rules. Therefore, the respect of random drawing is the first step of a good discard sampling.

Here is a suggestion to approximate as much as possible the drawing of fishing trips at random in the population of fishing trips. First, create a list of vessels practising a given métier per month, based on information from preceding year. For a quarter, draw at random a list of vessel*month and contact the captains on the list to check (i) that they will practise the right métier that particular month of the quarter and (ii) whether they are willing to take an observer on board. Such a list should contain much more vessels*month than expected to allow impossibilities to sample. The next on the list would then be chosen in replacement. The population of fishing trips are considered independent (no vessel effect) and sample trips will be considered as the primary sampling units (Stratoudakis *et al.*, 1999).

Once on board, the observer will chose hauls to sample. In Cochran's theory, the hauls should be taken randomly. Once again, there is no predefined list of hauls upon which one can base a random sample. The idea is then to make a kind of systematic sampling, spreading the samples equally during the day and the night and equally among the fishing days of the trip. The random hypothesis is ensured by the random distribution of the variable of interest in this particular case (same vessel, same trip, same geographical area, …). The hauls are considered as the secondary units.

Once a haul is chosen, another sub-sampling may occur by dividing the catch into boxes, and when a box is chosen, a sub-sample of one species may occur for counting, weighing and/or measuring. These levels are not considered here as they would make this document too much complex. Moreover, it is known that these levels do not account for much in the total variance (Tamsett *et al.* 1999)if the on-board observer pays great attention in dividing into as much equal parts as possible.

By construction, the sampling for discards follow a multi-stage sampling with fishing trips chosen with equal probabilities. For educational purpose, this document will explain only the two-stage sampling.

In the following formulas, the correction of finite population will be used at the trip level and will not be used to estimate the variance between trips. The reason being that the proportion of hauls sampled during a trip is well over the theoretical threshold of 5% and the proportion of trips sampled is well below.

3. Notations

	Population	Sample
Number of trips	N	n
Volume of discards in a haul *j* of a trip *i*	Y_{ij}	y_{ij}
Reference to a trip	i (i = 1, …, N)	i (i = 1, …, n)
Number of hauls in a trip *i*	M_i	m_i
Reference to a haul	j (j = 1, …, M_i)	j (j = 1, …, m_i)
Haul level		
Mean volume of discards per haul in a trip *i*	$\overline{Y}_i = \frac{1}{M_i}\sum_{j=1}^{M_i} Y_{ij}$	$\overline{y}_i = \frac{1}{m_i}\sum_{j=1}^{m_i} y_{ij}$
Mean volume of discards per haul	$\overline{\overline{Y}} = \frac{1}{N}\sum_{i=1}^{N} \overline{Y}_i$	$\overline{\overline{y}} = \frac{1}{n}\sum_{i=1}^{n} \overline{y}_i$
Trip level		
Total volume of discards in a trip *i*	$Y_i = M_i \overline{Y}_i$	$\hat{y}_i = M_i \overline{y}_i$
Mean volume of discards per trip	$\overline{Y} = \frac{1}{N}\sum_{i=1}^{N} \overline{Y}_i$	$\overline{y} = \frac{1}{n}\sum_{i=1}^{n} \overline{y}_i$
Variance (between hauls and within trip *i*)	$S_{2i}^2 = \frac{\sum_{j=1}^{M_i}\left(Y_{ij} - \overline{Y}_i\right)^2}{M_i - 1}$	$s_{2i}^2 = \frac{\sum_{j=1}^{m_i}\left(y_{ij} - \overline{y}_i\right)^2}{m_i - 1}$

4. First alternative – Use of sampling theory

Assumptions

A1 : The trips are drawn at random with equal probability $p = 1/N$

4.1. Number of trips is known at the population level

Formulas can be found in Cochran (1977) chapter 11.7 and Thompson (2002) chapter 13.1

	Population	Sample
Total volume of discards (in a stratum)	$Y = \sum_{i=1}^{N} Y_i$	$\hat{y}_I = \frac{N}{n}\sum_{i=1}^{n} \hat{y}_i$

$$\hat{y}_I = \frac{N}{n}\sum_{i=1}^{n} \hat{y}_i = \frac{N}{n}\sum_{i=1}^{n} M_i \overline{y}_i = \frac{N}{n}\sum_{i=1}^{n}\left[M_i \sum_{j=1}^{m_i} \frac{y_{ij}}{m_i} \right] \quad (11.21)$$

the associated variance is

$$Var(\hat{y}_I) = \frac{N^2}{n}\frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{n-1} + \frac{N}{n}\sum_{i=1}^{n}\frac{M_i^2(1-m_i/M_i)s_{2i}^2}{m_i} \quad (11.24)$$

An unbiased estimate of the total volume of discards at the level of one stratum is $$\hat{y}_I = \frac{N}{n}\sum_{i=1}^{n}\hat{y}_i$$ ant its associated variance is $$Var(\hat{y}_I) = \frac{N^2}{n}\frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{n-1} + \frac{N}{n}\sum_{i=1}^{n}\frac{M_i^2(1-m_i/M_i)s_{2i}^2}{m_i}$$

4.2. Number of hauls is known at the population level

Formulas can be found in Cochran (1977) chapter 11.8 and Thompson (2002) chapter 13.1

	Population	Sample
Total volume of discards (in a stratum)	$Y = M_0\frac{\sum_{i=1}^{N}Y_i}{\sum_{i=1}^{N}M_i}$	$\hat{y}_{II} = M_0\frac{\sum_{i=1}^{n}\hat{y}_i}{\sum_{i=1}^{n}M_i}$

$$\hat{y}_{II} = M_0\frac{\sum_{i=1}^{n}\hat{y}_i}{\sum_{i=1}^{n}M_i} = M_0\frac{\sum_{i=1}^{n}M_i\bar{y}_i}{\sum_{i=1}^{n}M_i} = M_0\frac{\sum_{i=1}^{n}\frac{M_i}{m_i}\sum_{j=1}^{m_i}y_{ij}}{\sum_{i=1}^{n}M_i} \quad (11.25)$$

the associated variance is

$$Var(\hat{y}_{II}) = \frac{N^2}{n}\frac{\sum_{i=1}^{n}M_i^2(\bar{y}_i - \bar{\bar{y}})^2}{n-1} + \frac{N}{n}\sum_{i=1}^{n}\frac{M_i^2(1-m_i/M_i)}{m_i}s_{2i}^2 \quad (11.30)$$

There is the possibility to avoid the need of knowing N, the total number of trips at the population level, by replacing N by $M_0/\bar{M}$, $\bar{M}$ being the mean number of fishing operations per trip. $\bar{M}$ may be known at the population level, but then, N is known and there is no need to use another formula, or may be estimated from the sample.

It comes then

$$Var(\hat{y}_{II}) = \frac{M_0^2}{n\bar{M}^2}\frac{\sum_{i=1}^{n}M_i^2(\bar{y}_i - \bar{\bar{y}})^2}{n-1} + \frac{M_0}{n\bar{M}}\sum_{i=1}^{n}\frac{M_i^2(1-m_i/M_i)}{m_i}s_{2i}^2$$

An unbiased estimate of the total volume of discards at the level of one stratum is

$$\hat{y}_{II} = M_0 \frac{\sum_{i=1}^{n} M_i \bar{y}_i}{\sum_{i=1}^{n} M_i}$$

and its associated variance is

$$Var(\hat{y}_{II}) = \frac{N^2}{n} \frac{\sum_{i=1}^{n} M_i^2 \left(\bar{y}_i - \bar{\bar{y}}\right)^2}{n-1} + \frac{N}{n} \sum_{i=1}^{n} \frac{M_i^2 (1 - m_i / M_i)}{m_i} s_{2i}^2$$

or

$$Var(\hat{y}_{II}) = \frac{M_0^2}{n\bar{M}^2} \frac{\sum_{i=1}^{n} M_i^2 \left(\bar{y}_i - \bar{\bar{y}}\right)^2}{n-1} + \frac{M_0}{n\bar{M}} \sum_{i=1}^{n} \frac{M_i^2 (1 - m_i / M_i)}{m_i} s_{2i}^2$$

5. Second alternative – Volume of discards is proportional to an auxiliary variable

Formulas can be found in Cochran (1977) chapter 11.12 and Thompson (2002) chapter 14.1

Let X be the auxiliary variable. Auxiliary variables can be fishing time, landings of all the species or a component of the landings (small female nephrops,)

	Population	Sample
Auxiliary variable (X)		
Mean per haul in a trip *i*	$\bar{X}_i = \frac{1}{M_i} \sum_{j=1}^{M_i} X_{ij}$	$\bar{x}_i = \frac{1}{m_i} \sum_{j=1}^{m_i} x_{ij}$
Total in a trip *i*	$X_i = M_i \bar{X}_i$	$\hat{x}_i = M_i \bar{x}_i$
Trip level		
Variance (between hauls or within trip *i*)	$s_{2i}^2 = \frac{1}{m_i - 1} \sum_{j=1}^{m_i} \left[\left(y_{ij} - \hat{R} x_{ij} \right) - \left(\bar{y}_i - \hat{R} \bar{x}_i \right) \right]^2$	

Ratio estimates

	Population	Sample
Mean ratio	$R = \frac{\sum_{i=1}^{N} M_i \bar{Y}_i}{\sum_{i=1}^{N} M_i \bar{X}_i}$	$\hat{R} = \frac{\sum_{i=1}^{n} M_i \bar{y}_i}{\sum_{i=1}^{n} M_i \bar{x}_i}$

$$\hat{R} = \frac{\sum_{i=1}^{n} M_i \overline{y}_i}{\sum_{i=1}^{n} M_i \overline{x}_i} = \frac{\sum_{i=1}^{n} \frac{M_i}{m_i} \sum_{j=1}^{m_i} y_{ij}}{\sum_{i=1}^{n} \frac{M_i}{m_i} \sum_{j=1}^{m_i} x_{ij}}$$

the associated variance is

$$Var(\hat{R}) = \frac{N^2}{X^2 n} \frac{\sum_{i=1}^{n} \left(\hat{y}_i - \hat{R}\hat{x}_i\right)^2}{n-1} + \frac{N}{X^2 n} \sum_{i=1}^{n} \frac{M_i^2 (1 - m_i / M_i)}{m_i} s_{2i}^2$$

An unbiased estimate of the ratio between the volume of discards and an auxiliary variable at the level of one stratum is

$$\hat{R} = \frac{\sum_{i=1}^{n} M_i \overline{y}_i}{\sum_{i=1}^{n} M_i \overline{x}_i}$$

ant its associated variance is

$$Var(\hat{R}) = \frac{N^2}{X^2 n} \frac{\sum_{i=1}^{n} \left(\hat{y}_i - \hat{R}\hat{x}_i\right)^2}{n-1} + \frac{N}{X^2 n} \sum_{i=1}^{n} \frac{M_i^2 (1 - m_i / M_i)}{m_i} s_{2i}^2$$

Total discards

	Population	Sample
Total volume of discards (in a stratum)	$Y = XR = X \frac{\sum_{i=1}^{N} Y_i}{\sum_{i=1}^{N} X_i}$	$\hat{y}_{III} = X\hat{R} = X \frac{\sum_{i=1}^{n} \hat{y}_i}{\sum_{i=1}^{n} \hat{x}_i}$

$$\hat{y}_{III} = X\hat{R} = X \frac{\sum_{i=1}^{n} \hat{y}_i}{\sum_{i=1}^{n} \hat{x}_i} = X \frac{\sum_{i=1}^{n} M_i \overline{y}_i}{\sum_{i=1}^{n} M_i \overline{x}_i} = X \frac{\sum_{i=1}^{n} \frac{M_i}{m_i} \sum_{j=1}^{m_i} y_{ij}}{\sum_{i=1}^{n} \frac{M_i}{m_i} \sum_{j=1}^{m_i} x_{ij}}$$

the associated variance is

$$Var(\hat{y}_{III}) = X^2 \operatorname{var}(\hat{R}) = \frac{N^2}{n} \frac{\sum_{i=1}^{n} \left(\hat{y}_i - \hat{R}\hat{x}_i\right)^2}{n-1} + \frac{N}{n} \sum_{i=1}^{n} \frac{M_i^2 (1 - m_i / M_i)}{m_i} s_{2i}^2 \qquad (11.51)$$

An unbiased estimate of the total volume of discards at the level of one stratum is

$$\hat{y}_{III} = X\hat{R} = X\frac{\sum_{i=1}^{n} M_i \overline{y}_i}{\sum_{i=1}^{n} M_i \overline{x}_i}$$

ant its associated variance is

$$Var(\hat{y}_{III}) = X^2 \operatorname{var}(\hat{R}) = \frac{N^2}{n}\frac{\sum_{i=1}^{n}\left(\hat{y}_i - \hat{R}\hat{x}_i\right)^2}{n-1} + \frac{N}{n}\sum_{i=1}^{n}\frac{M_i^2(1-m_i/M_i)}{m_i}s_{2i}^2$$

6. Discussion

The objective of the paper is to provide discard sampling users with recipes that are handy, practical and guarantee without bias. The statisticians who want to do more complex analysis should always compare their results to one of these methods.

One of the outcomes of this study has been to show

- the need to estimate the total number of trips as a measure of effort, within the DCR
- that the estimates of the total number of hauls at the population level would improve the accuracy of the discard estimates with the same sampling intensity. Moreover, the knowledge of the total number of hauls at the population levels takes into account the vessels practising more than one métier during their fishing trips.

Actually, what is operated on-board fishing vessels is a 5 stages sampling (vessels, trips, hauls, boxes, counting/measuring of one species). Considering that the box level have a very minor contribution to the overall variance (Tamsett *et al.* 1999), there is little chance that the counting level account for more. It remains then a three-stage sampling. Although, the complexity of statistical development increases, three-stage sampling remains workable (Wang *et al. in press*) but the problem is that there is on average one single trip per vessel sampled. If some experience of replication of trips by vessels sampled exists, it would be very relevant to investigate the importance of the vessel effect in discarding pattern (Allen *et al.* 2002).

The heterogeneity of fishing trip duration within a stratum is likely to have an impact on the achieved precision of the estimates. The formulas derived from the multi-stage sampling are all based on the mean of total discards per sampled trip, thus averaging different magnitude of discards. Any linear or non linear linkage between the volume of discards and fishing time would increase the variance if the heterogeneity of trip duration is too high, e.g. mixing one day trips and 15 days trips and only the respect of the random process guarantees the absence of bias. Knowing the difficulty to guarantee the random process in on-board sampling programmes, it is wise to avoid too much heterogeneity in the trip duration at the moment of designing the stratification. Moreover, *a posteriori* investigation of the data may give valuable information for designing optimum stratification, and this kind of analysis should always be carried out before applying any raising methodology.

At the moment of elaborating a sampling design, multi-stage sampling would give more precise result than simple random sampling only if variance within trip is big and variance between trips is low, on a relative scale. For discard sampling, all the studies have shown the opposite pattern, but field implementation provides un-escapable constraints.

The use of the ratio to another variable does not request a formal linkage with the volume of discards. The higher the correlation will be between these two variables, the better will be the precision achieved. The correlation between the two variables should therefore always be estimated and the precision should be compared to the precision obtained using sampling theory.

If there is an agreement on the need of such a document, then we should think about the way to reach this agreement among statisticians involved in discard estimates. One solution would be to set the basis of such document in PGCCDBS and propose the writing of a paper for the next ICES ASC meeting. Extension of this approach to other fields requiring sampling procedures should be encouraged.

7. References

Allen, M., Kilpatrick, D., Armstrong, M., Briggs, R., Course, G., Pérez, N. 2002. Multistage cluster sampling design and optimal sample sizes for estimation of fish discards from commercial trawlers. Fish. Res. 55: 11–24.

Anonymous. 2003. Workshop on Discard Sampling Methodology and Raising Procedures. Charlottenlund, Denmark. 2 – 4 September, 28 pp.

Anonymous. 2005. Report of the Ad Hoc Meeting of independent experts on Fleet-Fishery based sampling - Nantes, mai 2005, 34 pp. + annex.

Borges L., A.F. Zuur, E. Rogan, R. Officer. 2005. Choosing the best sampling unit and auxiliary variable for discards estimations. Fisheries Research 75. p. 29-39

Cochran, W.G. (1977) Sampling Techniques. Third Edition. Ed. John Wiley & Sons

Pitcher, T., Watson, R., Forrest, R., Valtýsson, H., and Guénette, S.

Rochet, M.J., V.M. Trenkel. 2005. Factors for the variability of discards : assumptions and field evidence. Can. J. Fish. Aquat. Sci. **62** : p. 224-235

Stratoudakis, Y, Fryer, R.J., Cook, R.M. and Pierce, G.J. (1999) Fish discarded from Scottish demersal vessels: Estimators of total discards and annual estimates for targetted gadoids. *ICES. J. Mar. Sci.*, **56**, 592-605.

Stratoudakis, Y., Fryer, R.J., Cook, R.M., Pierce, G.J., and Coull, K.A. 2001. Fish bycatch and discarding in Nephrops trawlers in the Firth of Clyde (west of Scotland). Aquat. Living Resour. 14: 283–291.

Tamsett D., G. Janacek. 1999. Sampling trips for measuring discards in commercial fishing based on multilevel modelling of measurements in the North Sea from NE England. Fisheries Research 42. p. 103-115.

Tamsett D., G. Janacek, M. Emberton, B. Lart, G. Course. 1999. Onboard sampling for measuring discards in commercial fishing based on multilevel modelling of measurements in the Irish Sea from NW England and N Wales. Fisheries Research 42. p. 117-126

Thompson, S. K. 1992. Sampling. Wiley, New York

Wang, J., G. Gao, Y. Fan, L. Chen, S. Liu, Y. Jin and J. Yu. *In press*. The estimation of sample size in multi-stage sampling ant its application in medical survey. Applied Math. And Computation