Calculation of variation coefficients for different direct and indirect estimators used in Eustat economic surveys

Índice

1. Introduction

What follows is the presentation of various classic design-based estimators, or estimators assisted by basic models used in different Eustat economic surveys. The design philosophy states that the variable under study (from now on y) is treated as a fixed and unknown quantity. Randomness is given by the means of selecting the sample, which can be done using various procedures (simple random sampling, stratified random sampling, etc.). The success of design-based estimation methods lies in how representative is the sample of the population to be studied.

2. Direct estimators

A population $U = (1, ..., N)$ is defined which represents the whole Autonomous Community of the Basque Country and strata are defined, combining, for example, Provinces and CNAE (National Code of Economic Activities) classification. Following this, it is assumed that in each one of the strata a simple random sample is made. In this context, defining the strata a priori and selecting the sample at random within one of them, classic design-based direct expansion estimators can be used, given that the sampling n_h is fixed in each stratum.

Following this, without loss of generality, it is assumed that the basic sampling units are established.

Thus,

- y_{hj}: value of variable y in establishment j ($j = 1, ..., N_h$) of stratum h ($h = 1, ..., H$);
- Y_h: population total (of y) in stratum h;
- Y: population total (of y) in the A.C. of the Basque Country;
- N_h: population size of stratum h;
- n_h: size of the sample in stratum h;
- $\pi_{hj} = \frac{n_h}{N_h}$: probability inclusion of unit j in stratum h;
- $\omega_{hj} = \frac{1}{\pi_{hj}}$: sampling weight of unit j in stratum h;
- $\bar{y}_h = \frac{\sum_{j=1}^{n_h} y_{hj}}{n_h}$: sampling mean in stratum h;

- $s^2_{yh} = \frac{\sum_{j=1}^{n_h}(y_{hj}-\bar{y}_h)^2}{n_h-1}$: sampling variance in stratum h;
- $\hat{t}_{yh} = \frac{N_h}{n_h}\sum_{j=1}^{n_h} y_{hj} = N_h\bar{y}_h$: estimator of the total in stratum h;
- $\widehat{\text{var}}(\hat{t}_{yh}) = \left(1 - \frac{n_h}{N_h}\right)N_h^2\frac{s^2_{yh}}{n_h}$: variance estimator of the total in stratum h;
- $\bar{y}_{str} = \frac{\hat{t}_{str.y}}{N} = \sum_{h=1}^{H}\frac{N_h}{N}\bar{h}_h = \frac{\sum_{h=1}^{H}\sum_{j=1}^{n_h} w_{hj}y_{hj}}{\sum_{h=1}^{H}\sum_{j=1}^{n_h} w_{hj}}$: stratified sampling mean;
- $\hat{t}_{str.y} = \sum_{h=1}^{H}\hat{t}_{yh} = \sum_{h=1}^{H} N_h\bar{y}_h = \sum_{h=1}^{H}\sum_{j=1}^{n_h} w_{hj}y_{hj}$: stratified estimator of the total.

Given that the sampling is made independently in each stratum, the stratified variance of the total is given by

$$\text{var}(\hat{t}_{str.y}) = \sum_{h=1}^{H}\left(1 - \frac{n_h}{N_h}\right)N_h^2\frac{s^2_{yh}}{n_h},$$

the standard error is

$$\text{e.e}(\hat{t}_{str.y}) = \sqrt{\sum_{h=1}^{H}\left(1 - \frac{n_h}{N_h}\right)N_h^2\frac{s^2_{yh}}{n_h}},$$

and the variation coefficient

$$\text{c.v}(\hat{t}_{str.y}) = \frac{\text{e.e}(\hat{t}_{str.y})}{\hat{t}_{str.y}}.$$

Properties: estimators $\bar{y}_{str}$ and $\hat{t}_{str}$ are unconditionally unbiased.

2.1. Horvitz-Thompson Estimator

The Horvitz-Thompson estimator of the population total $Y_h = \sum_{j=1}^{N_h} y_{hj}$ of variable y in stratum h is given by

$$\hat{t}_{yh.\text{HT}} = \sum_{j=1}^{n_h} w_{hj}y_{hj},$$

where $w_{hj} = 1/\pi_{hj}$ are the sampling weights of the j-th unit in stratum h and π_{hj} is its probability of inclusion (or sampling fraction). In a simple random sample with n_h units selected from the total N_h, w_{hj}, $j = 1, ..., n_h$. In this case an unbiased estimator of the variance of the Horvitz-Thompson is given by the following expression:

$$\widehat{\text{var}}(\hat{t}_{yh.\text{HT}}) = N_h^2\Big(1 - \frac{n_h}{N_h}\Big)\frac{1}{n_h}\widehat{\text{var}}(y_{hj}) = N_h^2\Big(1 - \frac{n_h}{N_h}\Big)\frac{1}{n_h}\frac{\sum_{j=1}^{n_h}(y_{hj} - \bar{y}_h)^2}{n_h - 1}.$$

The Horvitz-Thompson estimator is a direct estimator and does not make use of any type of auxiliary information, or in other words, it uses only the information obtained in the sample and the sampling weights for its calculation. When the sampling size is small it is not an appropriate estimator, even though the design is unbiased, given that it is a highly unstable estimator and its variance can be great in these cases.

2.2. Generalised regression estimator (GREG)

The generalised regression estimator is an estimator which uses information auxiliary to the variable, for example, x to estimate variable y. It differs from the regular regression estimator in that it introduces weightings in the estimation of the coefficients of the model (normally the sampling weights). This type of estimator uses regression models as a means of obtaining estimators which are consistent from the point of view of design. They require the sample to be random. They were proposed fundamentally by Särndal, Swensson and Wrettman (1989). The generalised regression estimator of the total Y_h in stratum h is given by

$$\hat{t}_{yh.\text{GREG}} = \sum_{j=1}^{N_h}\hat{y}_{hj} + \sum_{j=1}^{n_h}\omega_{hj}(y_{hj} - \hat{y}_{hj}), \tag{1}$$

where $\hat{y}_{hj}$, $j = 1, ..., N_h$ are the values forecast by a model given in stratum h. The term $\sum_{j=1}^{n_h}\omega_{hj}(y_{hj} - \hat{y}_{hj})$ can be interpreted as a regression adjustment given the estimator provided by the model. The effect is an important reduction in variance, especially when the relationship between y and x is very strong. If the chosen model is a linear regression model, $y_{hj} = x'_{hj}\beta_h + \epsilon_{hj}$, with $\text{var}(\epsilon_{hj}) = \sigma_h^2$ and $x_{hj} = (1, x_{hj1}, ..., x_{hjk})'$, then $\hat{y}_{hj} = x'_{hj}\hat{\beta}_{h.\text{GREG}}$, where

$$\hat{\beta}_{h.\text{GREG}} = \left(\sum_{j=1}^{n_h}\omega_{hj}x'_{hj}x_{hj}c_{hj}\right)^{-1}\sum_{j=1}^{n_h}\omega_{hj}x'_{hj}y_{hj}c_{hj},$$

and c_{hj} are specified constants. Expression (1) could also be written as:

$$\hat{t}_{yh.\text{GREG}} = \hat{t}_{yh.\text{HT}} + (X_h - \hat{t}_{xh.\text{HT}})'\hat{\beta}_{h.\text{GREG}},$$

where $\hat{t}_{yh.\mathrm{HT}} = \sum_{j=1}^{n_h} \omega_{hj} y_{hj}$, is the Horvitz-Thompson estimator of Y_h, and $\hat{t}_{yh.\mathrm{HT}} = \sum_{j=1}^{n_h} \omega_{hj} x_{hj}$ is the Horvitz-Thompson estimator of X_h. In effect, both expressions coincide, since,

$$\begin{aligned} \hat{t}_{yh.\mathrm{GREG}} &= \textstyle\sum_{j=1}^{N_h} x'_{hj} \hat{\beta}_{h.\mathrm{GREG}} + \sum_{j=1}^{n_h} \omega_{hj} (y_{hj} - x'_{hj}) \hat{\beta}_{h.\mathrm{GREG}} \\ &= \hat{t}_{yh.\mathrm{HT}} + (X_h - \hat{t}_{xh.\mathrm{HT}})' \hat{\beta}_{h.\mathrm{GREG}}. \end{aligned}$$

The generalised regression estimator can also be expressed as a linear weighting upon y_j so that

$$\hat{t}_{yh.\mathrm{GREG}} = \sum_{j=1}^{n_h} \omega^*_{hj} y_{hj} = \sum_{j=1}^{n_h} \omega_{hj} g_{hj} y_{hj},$$

where weights $\omega^*_{hj} = \omega_{hj} = g_{hj}$ with $\omega_{hj} = 1/\pi_{hj}$,

$$g_{hj} = 1 + \Big(\sum_{j=1}^{N_h} x_{hj} - \sum_{j=1}^{n_h} \omega_{hj} x_{hj} \Big)' T_h^{-1} x_{hj} c_{hj} = 1 + (X_h - \hat{t}_{xh.\mathrm{HT}})' T_h^{-1} x_{hj} c_{hj},$$

and

$$T_h = \sum_{j=1}^{n_h} \omega_{hj} x_{hj} x'_{hj} c_{hj}.$$

The value of g_{hj} is close to the unit in most cases. The larger the sample, the closer it will be to the unit. It is relatively rare to find g_{hj} that are greater than 4 or less than 0. The weights ω^*_j are called calibrated weights since these weights, applied to x_j reproduce exactly the total population of x_j, which is to say

$$\sum_{j=1}^{n_h} \omega^*_{hj} x_{hj} = \sum_{j=1}^{N_h} x_{hj} = X_h.$$

The variance of the GREG estimator is given by

$$\mathrm{var}(\hat{t}_{yh.\mathrm{GREG}}) = \sum_{j=1}^{N_h} \sum_{k=1}^{N_h} \Big(\frac{\omega_j \omega_k}{\omega_{jk}} - 1 \Big) \epsilon_j \epsilon_k,$$

where $\epsilon_j = y_j - x'_j \beta_{h.\mathrm{GREG}}$ and is estimated by the expression:

$$\hat{\mathrm{var}}(\hat{t}_{yh.\mathrm{GREG}}) = \sum_{j=1}^{n_h} \sum_{k=1}^{n_h} (\omega_{hj} \omega_{hj} - \omega_{hjk})(g_{hj} \hat{\epsilon}_j), (g_{kh} \hat{\epsilon}_j)$$

where $\hat{\epsilon}_j = y_j - x_j'\hat{\beta}_{h.\text{GREG}}$. In the case of simple random sampling, this expression takes the form:

$$\widehat{\text{var}}(\hat{t}_{yh.\text{GREG}}) = N_h^2(1 - \frac{n_h}{N_h})\frac{1}{n_h}\widehat{\text{var}}(g_h\hat{\epsilon}), \tag{2}$$

where $g_h = (g_1, ..., g_{n_h})$ and $\hat{\epsilon} = (\hat{\epsilon}_1, ..., \hat{\epsilon}_{n_h})$.

2.2.1. Direct ratio indicator

When there is a single auxiliary variable, regression goes to the origin and the regression model is heterocedastic, with weights $c_{hj} = x_{hj}$, the GREG estimator is a direct ratio estimator. The above-defined values g_{hj} are, in this case, constant for all observations $j = 1, ..., n_h$ and are given by

$$g_h = 1 + \frac{X_h - \hat{t}_{x.\text{HT}}}{\hat{t}_{x.\text{HT}}},$$

where $\hat{t}_{x.\text{HT}} = \sum_{j=1}^{n_h} \omega_{hj}x_{hj}$ is a number and not a vector. Furthermore,

$$\hat{\beta}_{h.\text{D}} = \frac{\sum_{j=1}^{n_h} \omega_{hj}y_{hj}}{\sum_{j=1}^{n_h} \omega_{hj}x_{hj}},$$

and so $\hat{t}_{x.\text{GREG}} = \hat{t}_{x-\text{D}} = X'\hat{\beta}_{h.\text{D}} = (\sum_{j=1} N_h x_{hj})'\hat{\beta}_{\text{D}}$. In official statistics this estimator is frequently expressed as

$$\hat{t}_{yh.\text{D}} = \frac{\sum_{j=1}^{N_h} x_{hj}}{\sum_{j=1}^{n_h} \omega_{hj}x_{hj}} \sum_{j=1}^{n_h} \omega_{hj}y_{hj} = \frac{X_h}{\hat{t}_{xh.\text{HT}}}\hat{t}_{yh.\text{HT}} = (\text{FE})\hat{t}_{yh.\text{HT}},$$

where FE is the exponent that does not depend on the variable to be estimated. Note that this exponent coincides with g_h of the GREG estimator. If it is a small domain, in that there are few sample observations falling within this domain, the estimator is highly unstable. It is a direct indicator that uses information solely from its own domain. Its variance is $O(1/n_h)$, therefore quite large. It is obtained as a specific case of expression (2), from which we deduce

$$\widehat{\text{var}}(\hat{t}_{yh.\text{D}}) \approx N_h^2(1 - \frac{n_h}{N_h})\frac{1}{n_h}(\frac{X_h}{\hat{t}_{x.\text{HT}}})^2\widehat{\text{var}}_h(\hat{\epsilon}),$$

where $\widehat{\text{var}}_h(\hat{\epsilon}) = \frac{\sum_{j=1}^{n_h}(\hat{\epsilon}_{hj} - \hat{\epsilon}_j)^2}{n_h - 1}$ is the sampling variance of the remainders of model $y_{hj} = \beta_h x_{hj} + \epsilon_{hj}$, from $j = 1, ..., n_h$ with $\text{var}(\epsilon_{hj}) = \sigma^2 x_{hj}$. In other words, the remainders are obtained directly on calculating $\hat{\epsilon}_{hj} = y_{hj} - \hat{y}_{hj} = y_{hj} - x'_{hj}\hat{\beta}_{h.\mathrm{D}}$. As bias is considered to be virtually nil, the mean squared error of this estimator approaches because of its variance, which is to say $\mathrm{MSE}(\hat{t}_{yh.\mathrm{D}}) \approx \mathrm{var}(\hat{t}_{yh.\mathrm{D}})$. In this case the estimator of its variation coefficient is estimated using the expression

$$\widehat{\mathrm{c.v.}}(\hat{t}_{yh.\mathrm{D}}) = \frac{\widehat{\mathrm{e.e.}}(\hat{t}_{yh.\mathrm{D}})}{\hat{t}_{yh.\mathrm{D}}},$$

where $\widehat{\mathrm{e.e.}}(\hat{t}_{yh.\mathrm{D}}) = \sqrt{\widehat{\mathrm{var}}(\hat{t}_{yh.\mathrm{D}})}$.

3. Indirect estimators

Indirect estimators use information from other domains to estimate the total or mean of a variable in a given domain.

3.1. Indirect ratio estimator or synthetic estimator

Let us suppose that auxiliary information from other domains is used so that

$$\hat{\beta} = \frac{\sum_{h=1}^{H}\sum_{j=1}^{n_h}\omega_{hj}y_{hj}}{\sum_{h=1}^{H}\sum_{j=1}^{n_h}\omega_{hj}x_{hj}}.$$

Notice that $\hat{\beta}$ uses global sampling information, both for values of y and of x and is, therefore, more stable. The estimator takes 'borrowed information' from the other domains. The variance of $\hat{\beta}$ is of the order $O(1/n)$, therefore much smaller. The estimator of the total in stratum h is known as a synthetic estimator and is given by:

$$\hat{t}_{yh.\mathrm{SYN}} = X_h\hat{\beta} = X_h\frac{\sum_{h=1}^{H}\sum_{j=1}^{n_h}\omega_{hj}y_{hj}}{\sum_{h=1}^{H}\sum_{j=1}^{n_h}\omega_{hj}x_{hj}}.$$

This estimator can be rather biased, but its variance is small when the size of the global sample $n = \sum_{h=1}^{H} n_h$ is large. This estimator can be especially unbiased when the $\hat{\beta}_{h.\mathrm{D}}$ of the various strata resemble one another, which is to say that they resemble the domain which contains them. Using these hypotheses and not others, the use of the synthetic estimator is recommended, since it is then a stable estimator and can become virtually unbiased. Its variance is given by

$$\widehat{\text{var}}(\hat{t}_{yh.\text{SYN}}) \approx N^2(1-\frac{n}{N})\frac{1}{n}(\frac{X_h}{\sum_{h=1}^{H}\sum_{j=1}^{n_h}\omega_{hj}x_{hj}})^2\widehat{\text{var}}(\epsilon),$$

where $\widehat{\text{var}}(\epsilon)$ is the sampling variance of the remainders of the heterocedastic model $y_j = \beta x_j + \epsilon_j$ with $\text{var}(\epsilon) = \sigma^2 x_j$, $j = 1, ..., N$, at a population level which is to say that the remainders are calculated in the whole sample, not only in the stratum under study.

Särndal and Hidiroglou (1989) provide an approximation to the bias of the synthetic estimator where $E(\hat{t}_{yh.\text{SYN}}) - t_{yh.\text{SYN}} \approx -\sum_{j=1}^{N}\epsilon_j$ where $\epsilon_j = y_j - x'_j\hat{\beta}$. Then the estimator will be approximately unbiased if it is verified that $\sum_{j=1}^{N}\epsilon_k = 0$. This condition is not normally satisfied. If the model does not adjust well to the domain of interest, the sum of the remainders could be far from zero, indicating a considerable bias. Otherwise, we could expect a limited bias. Therefore, it is preferable to estimate the mean squared error for the accuracy of the estimator. It is given by

$$\text{MSE}(\hat{t}_{yh.\text{SYN}}) = \text{var}(\hat{t}_{yh.\text{SYN}}) + (\text{bias}_{yh})^2,$$

and is estimated by the expression:

$$\widehat{\text{MSE}}(\hat{t}_{yh.\text{SYN}}) = \widehat{\text{var}}(\hat{t}_{yh.\text{SYN}}) + (\sum_{j=1}^{n_h}\hat{\epsilon}_j)^2,$$

where $\hat{\epsilon}_j = y_j - x'_j\hat{\beta}$, $j = 1, ..., n$ are the remainders obtained from the estimated model with all the sampling data, although in each case only the specifics of this stratum are added.

The estimator of its variation coefficient is given by

$$\widehat{\text{c.v.}}(\hat{t}_{yh.\text{SYN}}) = \frac{\widehat{\text{rmse}}(\hat{t}_{yh.\text{SYN}})}{\hat{t}_{yh.\text{SYN}}},$$

where, $\widehat{\text{rmse}}(\hat{t}_{yh.\text{SYN}}) = \sqrt{\widehat{\text{MSE}}(\hat{t}_{yh.\text{SYN}})}$.

4. Compound estimators

The compound estimator is created to compensate for the bias of the indirect indicator, compared to the instability of the direct estimator. It is given by

$$\hat{t}_{yh.\text{C}} = \phi_h\hat{t}_{yh.\text{D}} + (1-\phi_h)\hat{t}_{yh.\text{I}},$$

where $0 \leq \phi_h \leq 1$, $\hat{t}_{yh.\mathrm{D}}$ is a direct estimator and $\hat{t}_{yh.\mathrm{I}}$ is an indirect estimator. Many of the estimators proposed in literature under designs or models are compound estimators. Among the design-based compound estimators, we could take the example of $\hat{t}_{yh.\mathrm{D}}$ as a direct ratio estimator and $\hat{t}_{yh.\mathrm{I}}$ as an indirect ratio estimator or synthetic estimator.

Pfefferman (2002) puts forward the compound estimator:

$$\hat{t}_{yh.\mathrm{C}} = \phi_h \hat{t}_{yh.\mathrm{D}} + (1-\phi_h)\hat{t}_{yh.\mathrm{SYN}}, \qquad \phi_h = \frac{n_h}{N_h}. \tag{3}$$

This choice of ϕ_h is particularly appropriate for populations which are not very large, since otherwise the n_h/N_h quotient would not necessarily favour the direct estimator when n_h grows. With these weights, the weight of the direct or indirect estimator is greater according to its sampling representation. In other words, the greater the sampling fraction, the greater the contribution of the direct estimator. When the sample is represented to a lesser extent, the weight of the indirect estimator is greater in the compound estimator. It could also be true that $n_h = 1$ and $N_h = 1$ in which case the compound estimator would be equal to the direct one.

The mean squared error of the compound estimator (3) is given by

$$\begin{aligned}\mathrm{MSE}(\hat{t}_{yc.\mathrm{C}}) \;\approx\; & \phi_h^2 \mathrm{MSE}(\hat{t}_{yd.\mathrm{D}}) + (1-\phi_h)^2 \mathrm{MSE}(\hat{t}_{yd.\mathrm{SYN}}) \\ & +2\phi_h(1-\phi_h)E[(\hat{t}_{yh.\mathrm{D}} - Y_h)(\hat{t}_{yh.\mathrm{SYN}} - Y_h)].\end{aligned}$$

Its estimation is not simple, since it is possible that the third term of this sum, which is to say the co-variance, is not small. A possible approximation is given by

$$\begin{aligned}& E[(\hat{t}_{yh.\mathrm{D}} - Y_h)(\hat{t}_{yh.\mathrm{SYN}} - Y_h)] = \\ & = E[\hat{t}_{yh.\mathrm{D}}\hat{t}_{yh.\mathrm{SYN}}] - Y_h E[\hat{t}_{yh.\mathrm{D}}] - Y_h E[\hat{t}_{yh.\mathrm{SYN}}] + Y_h^2 \\ & \approx E[\hat{t}_{yh.\mathrm{D}}\hat{t}_{yh.\mathrm{SYN}}] - Y_h^2 - Y_h(Y_h + \mathrm{bias}_{\mathrm{h.SYN}}) + Y_h^2 \\ & = E[\hat{t}_{yh.\mathrm{D}}\hat{t}_{yh.\mathrm{SYN}}] - Y_h^2 - Y_h(\mathrm{bias}_{\mathrm{h.SYN}}) \\ & \approx E[\hat{t}^2_{yh.\mathrm{D}}] - Y_h(\mathrm{bias}_{h.\mathrm{SYN}}) \\ & = \mathrm{MSE}(\hat{t}_{yh.\mathrm{D}}) - Y_h(\mathrm{bias}_{h.\mathrm{SYN}}).\end{aligned}$$

Note that $E[\hat{t}_{yh.\mathrm{D}}\hat{t}_{yh.\mathrm{SYN}}]$ could be approximated to with $E[\hat{t}^2_{yh.\mathrm{D}}]$ since the co-variance is different to zero only for the common terms of both estimators, which is to say for the terms that intervene in the calculation of the direct estimator.

Substituting $\mathrm{MSE}(\hat{t}_{yh.\mathrm{D}})$ for its estimator, Y_h for its synthetic estimator $\hat{t}_{yh.\mathrm{SYN}}$ and $\mathrm{bias}_{h.\mathrm{SYN}}$ for its estimator, we obtain the estimator of the mean squared error as

$$
\begin{aligned}
\widehat{\mathrm{MSE}}(\hat{t}_{yh.\mathrm{C}}) &\approx \phi_h^2 \widehat{\mathrm{MSE}}(\hat{t}_{yh.\mathrm{D}}) + (1-\phi_h)^2 \widehat{\mathrm{MSE}}(\hat{t}_{yh.\mathrm{SYN}}) \\
&+ 2\phi_h(1-\phi_h)[\widehat{\mathrm{MSE}}(\hat{t}_{yh.\mathrm{D}}) - \hat{t}_{yh.\mathrm{SYN}}(\mathrm{bias}_{h-\mathrm{SYN}})].
\end{aligned}
$$

The term $\widehat{\mathrm{MSE}}(\hat{t}_{yh.\mathrm{D}}) - \hat{t}_{yh.\mathrm{SYN}}(\mathrm{bias}_{h.\mathrm{SYN}})$ could turn out to be negative, in which case it could be approximated to zero and thus $\widehat{\mathrm{MSE}}(\hat{t}_{yh.\mathrm{C}}) \approx \phi_h^2 \widehat{\mathrm{MSE}}(\hat{t}_{yh.\mathrm{D}}) + (1-\phi_h)^2 \widehat{\mathrm{MSE}}(\hat{t}_{yh.\mathrm{SYN}})$. When $n = 0$ is taken $\widehat{\mathrm{MSE}}(\hat{t}_{yh.\mathrm{C}}) = \widehat{\mathrm{MSE}}(\hat{t}_{yh.\mathrm{SYN}})$ Since then the compound estimator is equal to the synthetic estimator, independently of the value of the direct indicator, which will have been obtained by aggregating to higher levels, since there is no sample in the CNAE itself.

The estimator of its variation coefficient is given by

$$
\widehat{\mathrm{c.v}}(\hat{t}_{yh.\mathrm{C}}) = \frac{\widehat{\mathrm{rmse}}(\hat{t}_{yh.\mathrm{C}})}{\hat{t}_{yh.\mathrm{C}}},
$$

where $\widehat{\mathrm{rmse}}(\hat{t}_{yh.\mathrm{C}}) = \sqrt{\widehat{\mathrm{MSE}}(\hat{t}_{yh.\mathrm{C}})}$.

Referencias

[1] Prasad, N.G.N and Rao, J.N.K. (1990). The Estimation of Mean Squared Error of Small Area Estimators. *Journal of the American Statistical Association*, **85**, 163-171.

[2] Pfeffermann, D. (2002). Snall Area Estimations-New Developments and Directions. *International Statistical Review*, **70**, 125-143.

[3] Rao, J.N.K. *Small Area Estimation*. Wiley, New York, 2003.

[4] Särndal, C.E., Hidirogloy, M.A. (1989). Small Domain Estimation: A Conditional Analysis. *Journal of the American Statistical Association*, **84**, 266-275.

[5] Särndal, C.E., Swensson, B. and Wretman, J.H. (1989). The Weighted Technique for Estimating the Variance of the General Regression Estimatorof the Finite Population Total. *Biometrica*, **76**, 3, 527-537.

[6] Särndal, C.E., Swensson, B. and Wretman, J.H. *Model Assisted Survey Sampling*. Springer-Verlag, New York, 1992.