Sampling Error Estimation in Stratified Surveys

Ricardo Cao, José A. Vilar, Juan M. Vilar, Ana K. López

Department of Mathematics, Campus de Elviña, University of A Coruña, A Coruña, Spain
Email: juan.vilar@udc.es

ABSTRACT

Many operations carried out by official statistical institutes use large-scale surveys obtained by stratified random sampling without replacement. Variables commonly examined in this type of surveys are binary, categorical and continuous, and hence, the estimates of interest involve estimates of proportions, totals and means. The problem of approximating the sampling relative error of this kind of estimates is studied in this paper. Some new jackknife methods are proposed and compared with plug-in and bootstrap methods. An extensive simulation study is carried out to compare the behavior of all the methods considered in this paper.

1. Introduction

Many of the operations carried out by official statistics institutions are based on surveys performed on a finite population using stratified random sampling without replacement. These surveys provide information on three types of variables: binary, categorical with more than two modalities, and continuous quantitative variables. Usually, most of the variables are binary, *i.e.* they yield two possible responses commonly coded with 1 and 0, and the aim is to estimate $p = P(1)$. A typical example is the variable indicating the presence or absence of a particular characteristic of interest in the study population, e.g. if a person is employed or unemployed, a household has Internet access or not, etc. In the case of categorical variables with more than two possible responses, say $A_i, i = 1, \cdots, I, I > 2$, the aim is to estimate the proportion of each of the answers, $p_i = P(A_i)$. For continuous quantitative variables, X, the objective is to estimate the mean of the variable, $\mu = E(X)$.

Stratified sampling is an appropriate method when several homogeneous and mutually exclusive strata or subpopulations are identified in the population. Stratification can contribute to improve the representativeness of the sample by reducing sampling error. The bigger the differences between the strata, the greater the gain in accuracy. Moreover, some strata can be occasionally small in size but big in importance in the study. In these cases, an exhaustive sampling is recommended, *i.e.* all the individuals of these strata will be part of the sample.

These strata are called self-represented.

Once that a population quantity θ, such as a mean or a proportion, has been estimated, it is important to obtain an accurate estimate of the sampling error to assess the reliability of our estimator $\hat{\theta}$. The sampling error of $\hat{\theta}$ can be presented in absolute terms, using the standard deviation of $\hat{\theta}$, namely

$$E_{abs}(\hat{\theta}) = \sqrt{\mathrm{Var}(\hat{\theta})},$$

or in relative terms, using the variation coefficient of the estimator given by

$$E_{rel}(\hat{\theta}) = \sqrt{\frac{\mathrm{Var}(\hat{\theta})}{E(\hat{\theta})}}.$$

This work addresses the problem of estimating sampling errors in surveys carried out by stratified sampling from finite populations. Some new proposals to obtain sampling relative error estimates are introduced and various methods based on plug-in, jackknife and bootstrap techniques are compared. This research originated from the collaboration with the Galician Official Statistical Institute (IGE), the agency in charge of the official statistics in the Autonomous Community of Galicia, a region in the Northwest of Spain. Specifically, IGE was interested in assessing different criteria to approximate the sampling relative error for a variety of statistics produced by the Annual Company Survey (ACS), which is conducted by IGE in Galicia.

About 200 variables are recorded in the Galician ACS, a survey including nearly 4000 companies in a total population of 65,000. The sample is obtained by stratified random sampling, with strata having different weights and some of them being self-represented. The strata are constructed by regarding the combination of three variables: the level of research performed in the company, the size of the company, measured in terms of the number of paid employees, and the main activity of the company.

The rest of this work is divided into the following sections. Notation used some basic concepts are introduced in Section 2. Section 3 describes the different jackknife procedures used to estimate the relative error of the estimators. The bootstrap methods of the error estimation are described in Section 4. Some results from a broad simulation study carried out to compare the behaviour of all the methods studied are presented in Section 5. Finally, the main conclusions of the work are established in Section 6.

2. Notation and Basic Concepts

From now on, N, L and N_i will denote the population size, the number of strata and the population size of the i-th stratum, $i = 1, \cdots, L$, respectively. Given a variable of interest, X, x_{ij} is the j-th element of the i-th population stratum, with $j = 1, \cdots, N_i, i = 1, \cdots, L$. When the variable in study is quantitative (continuous or discrete), it will make sense to speak about the population mean,

$$\mu = \frac{1}{N} \sum_{i=1}^{L} \sum_{j=1}^{N_i} x_{ij} = \sum_{i=1}^{L} \frac{N_i}{N} \bar{x}_{i\cdot}, \qquad (2.1)$$

and the population total,

$$\tau = \sum_{i=1}^{L} \sum_{j=1}^{N_i} x_{ij} = N\mu = \sum_{i=1}^{L} N_i \bar{x}_{i\cdot} = \sum_{i=1}^{L} x_{i\cdot}, \qquad (2.2)$$

being $\bar{x}_i = \frac{1}{N_i} \sum_{j=1}^{N_i} x_{ij}$ and $x_{i\cdot} = \sum_{j=1}^{N_i} x_{ij}$.

Let $\{X_{ij}, j = 1, \cdots, n_i, i = 1, \cdots, L\}$ be a stratified random sample without replacement of X, of size $n = \sum_{i=1}^{L} n_i$, n_i being the sample size within the i-th stratum. Denoting by $F_i = N_i / n_i$ the elevation factors of each stratum, the unbiased estimators of the mean μ and the total τ are obtained as follows.

$$\hat{\mu} = \sum_{i=1}^{L} \frac{N_i}{N} \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij} = \frac{1}{N} \sum_{i=1}^{L} F_i n_i \bar{X}_{i\cdot}$$
$$= \frac{1}{N} \sum_{i=1}^{L} F_i X_{i\cdot}, \qquad (2.3)$$

and

$$\hat{\tau} = N\hat{\mu} = \sum_{i=1}^{L} N_i \bar{X}_{i\cdot} = \sum_{i=1}^{L} F_i n_i \bar{X}_{i\cdot} = \sum_{i=1}^{L} F_i X_{i\cdot}, \qquad (2.4)$$

where $\bar{X}_{i\cdot} = \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij}$ and $X_{i\cdot} = \sum_{j=1}^{n_i} X_{ij}$.

Note that, by definition, $N_i = n_i$ for the strata self-represented in the sample, and hence the elevation factor of these strata is $F_i = 1$.

The unbiasedness property of $\hat{\mu}$ and $\hat{\tau}$ as estimators of μ and τ, respectively, follows from its construction as convex linear combinations of sample means (see (2.3) and (2.4)). So, from (2.1) and (2.2) follows

$$E(\hat{\mu}) = \frac{1}{N} \sum_{i=1}^{L} F_i n_i E(\bar{X}_{i\cdot}) = \frac{1}{N} \sum_{i=1}^{L} F_i n_i \bar{x}_{i\cdot}$$
$$= \sum_{i=1}^{L} \frac{N_i}{N} \bar{x}_{i\cdot} = \mu,$$

and

$$E(\hat{\tau}) = NE(\hat{\mu}) = N\mu = \tau.$$

Some simple calculations, although long (see, for example, [1]), yield the variance of these estimators. Specifically,

$$\mathrm{Var}(\hat{\mu}) = \frac{1}{N^2} \sum_{i=1}^{L} F_i^2 n_i^2 \frac{N_i - n_i}{N_i n_i} \sigma_i^2$$
$$= \frac{1}{N^2} \sum_{i=1}^{L} \frac{N_i (N_i - n_i)}{n_i} \sigma_i^2 \qquad (2.5)$$

and

$$\mathrm{Var}(\hat{\tau}) = \mathrm{Var}(N\hat{\mu}) = \sum_{i=1}^{L} \frac{N_i (N_i - n_i)}{n_i} \sigma_i^2, \quad (2.6)$$

σ_i^2 being the variance of the finite population in the i-th stratum, given as

$$\sigma_i^2 = \frac{1}{N_i - 1} \sum_{j=1}^{N_i} (x_{ij} - \bar{x}_{i\cdot})^2.$$

Replacing the population variances σ_i^2, in (2.5) and (2.6) by the corresponding sample variances corrected by their degrees of freedom,

$$S_i^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_{i\cdot})^2,$$

we obtain plug-in estimators for the variance of $\hat{\mu}$ and $\hat{\tau}$. Specifically,

$$\widehat{\mathrm{Var}}_{PI}(\hat{\mu}) = \frac{1}{N^2} \sum_{i=1}^{L} \frac{N_i (N_i - n_i)}{n_i} S_i^2$$

$$\widehat{\mathrm{Var}}_{PI}(\hat{\tau}) = \sum_{i=1}^{L} \frac{N_i (N_i - n_i)}{n_i} S_i^2. \qquad (2.7)$$

In the particular case of a binary variable we have

$$\sigma_i^2 = \frac{N_i}{N_i-1}\left(\frac{1}{N_i}\sum_{j=1}^{N_i} x_{ij}^2 - \bar{x}_{i\cdot}^2\right) = \frac{N_i}{N_i-1} p_i(1-p_i),$$

that is estimated using

$$S_i^2 = \frac{n_i}{n_i-1}\left(\frac{1}{N_i}\sum_{j=1}^{n_i} X_{ij}^2 - \bar{X}_{i\cdot}^2\right) = \frac{n_i}{n_i-1} P_i(1-P_i),$$

where p_i denotes the true proportion of ones in the i-th stratum and P_i the corresponding sampling proportion. Using the new expressions for σ_i^2 and S_i^2 in (2.6) and (2.7), respectively, we obtain the variance of $\hat{\tau}$ and its plug-in estimator for binary variables, which are reduced to

$$\mathrm{Var}(\hat{\tau}) = \sum_{i=1}^{L}\frac{N_i^2(N_i-n_i)}{n_i(N_i-1)} p_i(1-p_i),$$

$$\widehat{\mathrm{Var}}_{PI}(\hat{\tau}) = \sum_{i=1}^{L}\frac{N_i(N_i-n_i)}{n_i-1} P_i(1-P_i).$$

Going back to the general case in (2.5), it is deduced that the absolute and relative sampling errors of estimator $\hat{\mu}$ take the form

$$E_{abs}(\hat{\mu}) = \sqrt{\mathrm{Var}(\hat{\mu})} = \frac{1}{N}\left(\sum_{i=1}^{L}\frac{N_i(N_i-n_i)}{n_i}\sigma_i^2\right)^{1/2}$$
$$= \frac{1}{N} E_{abs}(\hat{\tau}),$$

$$E_{rel}(\hat{\mu}) = \frac{1}{\tau}\left(\sum_{i=1}^{L}\frac{N_i(N_i-n_i)}{n_i}\sigma_i^2\right)^{1/2} = E_{rel}(\hat{\tau}),$$

Hence, the plug-in estimators of these errors are

$$\hat{E}_{abs,PI}(\hat{\mu}) = \frac{1}{N}\left(\sum_{i=1}^{L}\frac{N_i(N_i-n_i)}{n_i}S_i^2\right)^{1/2}$$
$$= \frac{1}{N}\hat{E}_{abs,PI}(\hat{\tau}), \quad (2.8)$$

$$\hat{E}_{rel,PI}(\hat{\tau}) = \frac{1}{\tau}\left(\sum_{i=1}^{L}\frac{N_i(N_i-n_i)}{n_i}S_i^2\right)^{1/2} = \hat{E}_{rel,PI}(\hat{\tau}), \quad (2.9)$$

As $\tau = N\mu$ and $\hat{\tau} = N\hat{\mu}$, the relative errors of $\hat{\tau}$ and $\hat{\mu}$ as well as of their estimations are the same. Hence, the rest of the study will focus on the estimation of the relative error of parameter τ, the population total.

3. Jackknife Estimation

The jackknife method is a general estimation procedure introduced by [2] that has been widely used to estimate the bias and the standard error of a statistic. It is well known that the jackknife technique leads to a reduction in the bias. Furthermore, the jackknife method is basically a resampling procedure, and hence one can estimate the accuracy of an estimator without assuming previous hypotheses on the distribution of the population.

Let $\hat{\theta}$ be an estimator of a parameter θ based on the sample $X_1, X_2, \cdots, X_n$, then the jackknife estimator of the variance of $\hat{\theta}$ is defined as

$$\widehat{\mathrm{Var}}_{JACK}(\hat{\theta}) = \frac{n-1}{n}\sum_{i=1}^{n}\left(\hat{\theta}_{(i)} - \hat{\theta}_{(\cdot)}\right)^2,$$

where $\hat{\theta}_{(i)}$ is the jackknife *pseudovalue*, that is the estimator calculated using the whole sample except for the i-th observation, $X_1, \cdots, X_{i-1}, X_{i+1}, \cdots, X_n$, and $\hat{\theta}_{(\cdot)}$ is the mean of the jackknife pseudovalues,

$$\hat{\theta}_{(\cdot)} = \frac{1}{n}\sum_{i=1}^{n}\hat{\theta}_{(i)}.$$

In a stratified sampling, the jackknife pseudovalues can be constructed following one of the two possible criteria: either removing a sample value at each iteration or removing a stratum at each iteration. Application of these criteria leads to two different jackknife estimators for the variance of $\hat{\mu}$ and $\hat{\tau}$. Moreover, according to (2.4), $\hat{\tau}$ can be expressed as a linear combination of independent statistics $(X_{i\cdot})$, each one being separately constructed from the subsample of each stratum. Consequently, the variance of $\hat{\tau}$ can be calculated as a linear combination of variances of statistics constructed at stratum level. If these variances are previously estimated by jackknife in each stratum, then there will be a third way of using the jackknife to approximate the variance of $\hat{\tau}$. Each of the three jackknife proposals are described more in detail below.

3.1. Jackknife Leaving One Sample Value out

Each jackknife pseudovalue is constructed by removing a single data value from the overall sample $X_{ij}, j=1,\cdots,n_i, i=1,\cdots,L$. Thus, the pseudovalue obtained when eliminating the *s*-th observation of the *r*-th stratum, X_{rs}, takes the form

$$\hat{\tau}_{(rs)}^{(1)} = \sum_{i=1,i\neq r}^{L} F_i X_{i\cdot} + F_r' X_{r\cdot}^{(s)} = \hat{\tau} + F_r' X_{r\cdot}^{(s)} - F_r X_{r\cdot},$$

where $F_r' = \frac{N_r}{n_r-1}$ and $X_{i\cdot}^{(s)} = \sum_{j=1,j\neq s}^{n_i} X_{ij}$.

On the other hand we have that

$$F_r' X_{r\cdot}^{(s)} - F_r X_{r\cdot}$$
$$= \frac{N_r}{n_r-1}\sum_{j=1,j\neq s}^{n_i} X_{rj} - \frac{N_r}{n_r}\sum_{j=1}^{n_i} X_{rj}$$
$$= \left(\frac{N_r}{n_r-1} - \frac{N_r}{n_r}\right)\sum_{j=1,j\neq s}^{n_i} X_{rj} - \frac{N_r}{n_r} X_{rs}$$
$$= \frac{N_r}{(n_r-1)n_r} X_{r\cdot} - \frac{N_r}{n_r}\left(\frac{1}{n_r-1}+1\right)X_{rs}$$
$$= \frac{N_r}{n_r-1}\left(\bar{X}_{r\cdot} - X_{rs}\right),$$

and hence

$$\hat{\tau}_{(rs)}^{(1)} = \hat{\tau} + \frac{N_r}{n_r - 1}\left(\bar{X}_{r\cdot} - X_{rs}\right).$$

By averaging all the pseudovalues, we obtain

$$\hat{\tau}_{(\cdot)}^{(1)} = \frac{1}{n}\sum_{r=1}^{L}\sum_{s=1}^{n_r}\hat{\tau}_{(rs)}^{(1)} = \frac{1}{n}\sum_{r=1}^{L}\sum_{s=1}^{n_r}\left(\hat{\tau} + \frac{N_r}{n_r - 1}\left(\bar{X}_{r\cdot} - X_{rs}\right)\right)$$

$$= \hat{\tau} + \frac{1}{n}\sum_{r=1}^{L}\frac{N_r}{n_r - 1}\left(n_r\bar{X}_{r\cdot} - X_{r\cdot}\right) = \hat{\tau}.$$

Hence, the jackknife estimator of $\operatorname{Var}(\hat{\tau})$ is given by

$$\widehat{\operatorname{Var}}_{JACK,1}(\hat{\tau}) = \frac{n-1}{n}\sum_{r=1}^{L}\sum_{s=1}^{n_r}\left(\frac{N_r}{n_r - 1}\left(\bar{X}_{r\cdot} - X_{rs}\right)\right)^2$$

$$= \frac{n-1}{n}\sum_{r=1}^{L}\frac{n_r N_r^2}{(n_r - 1)^2}\left(\frac{1}{n_r}\sum_{s=1}^{n_r}X_{rs}^2 - \bar{X}_{r\cdot}^2\right)$$

$$= \frac{n-1}{n}\sum_{r=1}^{L}\frac{N_r^2}{n_r - 1}S_r^2,$$

and the first variant of the jackknife estimator for the relative error is

$$\hat{E}_{rel,JACK,1}(\hat{\tau}) = \frac{1}{\hat{\tau}}\left(\frac{n-1}{n}\sum_{r=1}^{L}\frac{N_r^2}{n_r - 1}S_r^2\right)^{1/2}. \qquad (3.1)$$

3.2. Jackknife Leaving One Stratum out

Here we propose to calculate each pseudovalue removing all the observations of one stratum. Thus, the r-th jackknife pseudovalue is based on the original sample without the observations of stratum r, *i.e.*

$$\hat{\tau}_{(r)}^{(2)} = \frac{N}{N - N_r}\sum_{i=1,i\neq r}^{L}F_iX_{i\cdot} = \frac{N}{N - N_r}\left(\hat{\tau} - F_rX_{r\cdot}\right).$$

Now, two variants of the jackknife estimator are introduced by considering different ways of averaging the pseudovalues $\hat{\tau}_{(r)}^{(2)}$. First, we use a weighted mean, where each pseudovalue is weighted by the population size of the stratum removed in the calculation. Thus, we have

$$\hat{\tau}_{(\cdot)}^{(2A)} = \sum_{r=1}^{L}\frac{N_r}{N}\hat{\tau}_{(r)}^{(2)} = \sum_{r=1}^{L}\frac{N_r}{N - N_r}\left(\hat{\tau} - F_rX_{r\cdot}\right)$$

$$= \hat{\tau}\sum_{r=1}^{L}\frac{N_r}{N - N_r} - \sum_{r=1}^{L}\frac{N_r}{N - N_r}F_rX_{r\cdot}.$$

Then, the jackknife estimator of $\operatorname{Var}(\hat{\tau})$ takes the form

$$\widehat{\operatorname{Var}}_{JACK,2A}(\hat{\tau}) = \sum_{r=1}^{L}\frac{N_r(N - N_r)}{N^2}\left(\hat{\tau}_{(r)}^{(2)} - \hat{\tau}_{(\cdot)}^{(2A)}\right)^2$$

$$= \sum_{r=1}^{L}\frac{N_r(N - N_r)}{N^2}\left(\frac{N(\hat{\tau} - F_rX_{r\cdot})}{N - N_r} - \sum_{i=1}^{L}\frac{N_i(\hat{\tau} + F_iX_{i\cdot})}{N - N_i}\right)^2.$$

Note that, in this case, it seems that a more simplified expression cannot be achieved. The jackknife estimator of the relative error is then calculated as

$$\hat{E}_{rel,JACK,2A}(\hat{\tau}) = \frac{1}{\hat{\tau}}\left(\widehat{\operatorname{Var}}_{JACK,2A}(\hat{\tau})\right)^{1/2}. \qquad (3.2)$$

An alternative variant of the jackknife leaving a stratum out is obtained if all the strata contribute with the same weight in the estimation, *i.e.* the pseudovalues are directly averaged as follows

$$\hat{\tau}_{(\cdot)}^{(2B)} = \frac{1}{L}\sum_{r=1}^{L}\hat{\tau}_{(r)}^{(2)} = \frac{N}{L}\sum_{r=1}^{L}\frac{1}{N - N_r}\left(\hat{\tau} - F_rX_{r\cdot}\right)$$

$$= \frac{N}{L}\left(\hat{\tau}\sum_{i=1}^{L}\frac{1}{N - N_r} - \sum_{r=1}^{L}\frac{F_rX_r}{N - N_r}\right).$$

Using $\hat{\tau}_{(\cdot)}^{(2B)}$, the jackknife estimator of $\operatorname{Var}(\hat{\tau})$ becomes

$$\widehat{\operatorname{Var}}_{JACK,2B}(\hat{\tau})$$

$$= \frac{L-1}{L}\sum_{r=1}^{L}\left(\hat{\tau}_{(r)}^{(2)} - \hat{\tau}_{(\cdot)}^{(2B)}\right)^2$$

$$= \frac{L-1}{L}\sum_{r=1}^{L}\left(\frac{N(\hat{\tau} - F_rX_{r\cdot})}{N - N_r} - \frac{N}{L}\sum_{i=1}^{L}\frac{\hat{\tau} + F_iX_{i\cdot}}{N - N_i}\right)^2,$$

not admitting a simpler explicit expression either.

The jackknife estimator of the relative error with this criterion is

$$\hat{E}_{rel,JACK,2B}(\hat{\tau}) = \frac{1}{\hat{\tau}}\left(\widehat{\operatorname{Var}}_{JACK,2B}(\hat{\tau})\right)^{1/2}. \qquad (3.3)$$

3.3. Jackknife within Each Stratum

According to (2.4), the variance of $\hat{\tau}$ can be expressed as a linear combination of the variances of the sample means within each stratum

$$\operatorname{Var}(\hat{\tau}) = \sum_{i=1}^{L}N_i^2\operatorname{Var}(\bar{X}_{i\cdot}). \qquad (3.4)$$

Hence, a new jackknife approximation to the variance of $\hat{\tau}$ can be obtained by estimating each $\operatorname{Var}(\bar{X}_{i\cdot})$ with the jackknife method and replacing these estimators in (3.4). For the jackknife estimator of $\operatorname{Var}(\bar{X}_{i\cdot})$, the pseudovalues are defined as

$$u_i^{(s)} = \bar{X}_{i\cdot}^{(s)} = \frac{1}{n_i - 1}\sum_{j=1,j\neq s}^{n_i}X_{ij} = \frac{1}{n_i - 1}X_{i\cdot}^{(s)},$$

for $s = 1, 2, \cdots, n_i$, and their mean is given by

$$u_i^{(\cdot)} = \frac{1}{n_i}\sum_{s=1}^{n_i}u_i^{(s)} = \frac{1}{n_i(n_i - 1)}\sum_{s=1}^{n_i}\sum_{j=1,j\neq s}^{n_i}X_{ij}$$

$$= \frac{1}{n_i(n_i - 1)}\sum_{j=1}^{n_i}(n_i - 1)X_{ij} = \bar{X}_{i\cdot}.$$

Then, the jackknife estimator of the variance of the sample mean of the i-th stratum is

$$\widehat{\mathrm{Var}}_{JACK}\left(\bar{X}_{i\cdot}\right)=\frac{n_i-1}{n_i}\sum_{s=1}^{n_i}\left(u_i^{(s)}-u_i^{(\cdot)}\right)^2$$

$$=\frac{n_i-1}{n_i}\sum_{s=1}^{n_i}\left(\sum_{j=1,j\neq s}^{n_i}\frac{X_{ij}}{(n_i-1)n_i}-\frac{X_{is}}{n_i}\right)^2$$

$$=\frac{n_i-1}{n_i}\sum_{s=1}^{n_i}\left(\sum_{j=1}^{n_i}\frac{X_{ij}}{(n_i-1)n_i}-\frac{X_{is}}{(n_i-1)}\right)^2$$

$$=\frac{n_i-1}{n_i}\sum_{s=1}^{n_i}\left(\sum_{j=1}^{n_i}\frac{X_{ij}}{(n_i-1)n_i}\right)^2+\left(\frac{X_{is}}{(n_i-1)}\right)^2$$

$$-\frac{2X_{is}}{(n_i-1)}\sum_{j=1}^{n_i}\frac{X_{ij}}{(n_i-1)n_i}$$

$$=\frac{1}{(n_i-1)n_i^2}\left(\sum_{j=1}^{n_i}X_{ij}\right)^2+\frac{1}{n_i(n_i-1)}\sum_{s=1}^{n_i}X_{is}^2$$

$$-\frac{2}{(n_i-1)n_i^2}\left(\sum_{j=1}^{n_i}X_{ij}\right)^2$$

$$=\frac{1}{n_i(n_i-1)}\sum_{s=1}^{n_i}X_{is}^2-\frac{1}{(n_i-1)n_i^2}\left(\sum_{j=1}^{n_i}X_{ij}\right)^2$$

$$=\frac{1}{n_i-1}\left(\frac{1}{n_i}\sum_{j=1}^{n_i}X_{ij}^2-\bar{X}_{i\cdot}^2\right)$$

$$=\frac{1}{(n_i-1)n_i}\sum_{j=1}^{n_i}\left(X_{ij}-\bar{X}_{i\cdot}\right)^2=\frac{S_i^2}{n_i},$$

and using these previous jackknife estimations we obtain

$$\widehat{\mathrm{Var}}_{JACK,3}\left(\hat{\tau}\right)=\sum_{i=1}^{L}\frac{N_i^2}{n_i}S_i^2.$$

The corresponding jackknife estimation of the relative error is given by

$$\hat{E}_{rel,JACK,3}\left(\hat{\tau}\right)=\frac{1}{\hat{\tau}}\left(\sum_{i=1}^{L}\frac{N_i^2}{n_i}S_i^2\right)^{1/2}. \qquad (3.5)$$

4. Bootstrap Estimation

An alternative resampling method often used to estimate the variance and the relative sampling error of the estimators is the bootstrap or self-sufficient estimation method. As any resampling procedure, including jackknife, the bootstrap takes advantage of not requiring hypotheses on the underlying distribution. The bootstrap method was introduced by Efron (see [3-5]) and has been widely treated in the literature. The basic idea of bootstrap consists in *estimating* the underlying population and then drawing out a number of resamples from the estimated population. By extracting a large number B of these resamples (in the order of one or several thousand), B

bootstrap replications of a particular estimator $\hat{\theta}$ can be obtained and used to approximate the estimator variance and relative error.

According to (2.4), the estimator $\hat{\tau}$ can be written as the sum of the independent random variables $F_iX_{i\cdot}, i=1,\cdots,L$. Hence, the bootstrap can be either applied to the global population, in order to directly estimate the variance of $\hat{\tau}$, or to each of the strata, in order to estimate the variance of each statistic $F_iX_{i\cdot}$ and then approximate the variance of $\hat{\tau}$ as the sum of the variances of $F_iX_{i\cdot}$. In particular, any bootstrap resampling plan ensuring the independence among different strata is valid to be used on the whole sample or stratum by stratum, although the latter is indeed more efficient computationally. Note also that the statistic $F_iX_{i\cdot}$ will have zero variance in self-represented strata, where the sampling is exhaustive $\left(F_i=1\right)$, and therefore it is not necessary to draw out bootstrap resamples in these strata.

A detailed description of two bootstrap resampling procedures to approximate the population total in a fixed stratum i, with $i\in\{1,2,\cdots,L\}$, is provided below.

4.1. Bickel and Freedman Bootstrap Method

The proposal by Bickel and Freedman (see [6]) consists in estimating the underlying population from a mixture of two distinct and equal-sized finite populations.

If F_i is an integer number, then the bootstrap algorithm proceeds as follows.

BF.1 The estimated subpopulation for the i-th stratum is constructed by grouping F_i identical copies of the sample of the stratum, that is the empirical population is given by

$$\hat{\Psi}_i=\left\{\overbrace{X_{i1},\cdots,X_{in_i}}^{1},\overbrace{X_{i1},\cdots,X_{in_i}}^{2},\cdots,\overbrace{X_{i1},\cdots,X_{in_i}}^{F_i}\right\}.$$

BF.2 A bootstrap resample of size n_i, $\left\{X_{i1}^*,\cdots,X_{in_i}^*\right\}$, is selected at random and without replacement from $\hat{\Psi}_i$.

BF.3 A bootstrap estimate for the population total in the i-th stratum, $X_{i\cdot}^*$, is calculated from the bootstrap resample derived in the previous step.

BF.4 Steps BF.2 and BF.3 are repeated a large number, B, of times ($B=1000$, 2000 or 5000, for example), thereby obtaining a set of B bootstrap replicates of the total estimator, $\left\{X_{i\cdot}^{*(1)},\cdots,X_{i\cdot}^{*(B)}\right\}$. Then, the bootstrap variance of the total, $\mathrm{Var}^*\left(X_{i\cdot}^*\right)$, is approximated by means of

$$\widehat{\mathrm{Var}}^*\left(X_{i\cdot}^*\right)=\frac{1}{B}\sum_{j=1}^{B}\left(X_{i\cdot}^{*(j)}-\overline{X_{i\cdot}^{*(\cdot)}}\right)^2,$$

$$\text{with } \overline{X_{i\cdot}^{*(\cdot)}}=\frac{1}{B}\sum_{j=1}^{B}X_{i\cdot}^{*(j)}. \qquad (4.1)$$

When the elevation factor F_i is not integer, Steps BF.1 and BF.2 are modified as follows.

BF.1' Consider $F_i = K_i + R_i$, being $K_i = [F_i]$, where $[\cdot]$ denotes the integer part function. Thus, $N_i = K_i n_i + r_i$, with $r_i = R_i n_i$ and $1 \le r_i < n_i$.

Two empirical finite subpopulations $\hat{\Psi}_i^{(1)}$ and $\hat{\Psi}_i^{(2)}$ are now considered for the i-th stratum, which are formed, respectively, by K_i and $(K_i + 1)$ identical copies of the observed sample, that is

$$\hat{\Psi}_i^{(1)} = \left\{ \overbrace{X_{i1}, \cdots, X_{in_i}}^{1}, \overbrace{X_{i1}, \cdots, X_{in_i}}^{2}, \cdots, \overbrace{X_{i1}, \cdots, X_{in_i}}^{K_i} \right\},$$

$$\hat{\Psi}_i^{(2)} = \left\{ \overbrace{X_{i1}, \cdots, X_{in_i}}^{1}, \overbrace{X_{i1}, \cdots, X_{in_i}}^{2}, \cdots, \overbrace{X_{i1}, \cdots, X_{in_i}}^{K_i}, \overbrace{X_{i1}, \cdots, X_{in_i}}^{K_i+1} \right\}$$

BF.2' Define $\alpha_i = \left[\left(1 - \dfrac{r_i}{n_i} \right) \left(1 - \dfrac{r_i}{N_i - 1} \right) \right]$. A bootstrap resample of size n_i, $\{X_{i1}^*, \cdots, X_{in_i}^*\}$, is selected randomly and without replacement from $\hat{\Psi}_i^{(1)}$ with probability α_i, and from $\hat{\Psi}_i^{(2)}$ with probability $1 - \alpha_i$.

The specific selection of α_i ensures that the mean and the variance in the resampling of a bootstrap observation X_{ij}^* are equal to the expected quantities when the size of the bootstrap population is N_i. Notice that the first resampling algorithm is a particular case of the previous one. In fact, if F_i is an integer, *i.e.* $R_i = 0$, then $r_i = 0, K_i = F_i, \hat{\Psi}_i^{(1)} = \hat{\Psi}_i, \alpha_i = 1$ and both algorithms become identical.

4.2. Booth, Butler and Hall Bootstrap Method

The bootstrap procedure proposed by Booth, Butler and Hall (see [7]) is based on completing the population $\hat{\Psi}_i^{(1)}$ estimated in Step BF.1' by adding a random subsample of the observed sample. In this way, one avoids using the two finite bootstrap populations $\hat{\Psi}_i^{(1)}$ and $\hat{\Psi}_i^{(2)}$. More precisely, Steps BF.1' and BF.2' of the Bickel and Freedman algorithm are modified as follows.

BBH.1 As in Step BF.1', consider $F_i = K_i + R_i$, where $K_i = [F_i]$ and $r_i = R_i n_i$, with $0 \le r_i < n_i$.

The finite subpopulation estimated for the i-th stratum is given by $\hat{\Psi}_i^{(4)} = \hat{\Psi}_i^{(1)} \bigcup \hat{\Psi}_i^{(3)}$, where $\hat{\Psi}_i^{(1)}$ is constructed as in Step BF.1', *i.e.*

$$\hat{\Psi}_i^{(1)} = \left\{ \overbrace{X_{i1}, \cdots, X_{in_i}}^{1}, \overbrace{X_{i1}, \cdots, X_{in_i}}^{2}, \cdots, \overbrace{X_{i1}, \cdots, X_{in_i}}^{K_i} \right\},$$

and $\hat{\Psi}_i^{(3)}$ is formed by a random subsample of size r_i generated without replacement from $\{X_{i1}, \cdots, X_{in_i}\}$.

BBH.2 A bootstrap resample of size n_i, $\{X_{i1}^*, \cdots, X_{in_i}^*\}$, is selected at random and without replacement from $\hat{\Psi}_i^{(4)}$.

The bootstrap estimation of the variance of the statistic $\hat{\tau}$ is now obtained using (2.4) and the bootstrap approximations $\widehat{\text{Var}}^*(X_i^*)$ calculated with one of the algorithms previously described. Specifically,

$$\widehat{\text{Var}}^*(\hat{\tau}^*) = \sum_{i=1}^{L} F_i^2 \widehat{\text{Var}}^*(X_{i\cdot}^*).$$

Considering that

$$E^*(\hat{\tau}^*) = \sum_{i=1}^{L} F_i E^*(X_{i\cdot}^*) = \sum_{i=1}^{L} F_i X_{i\cdot} = \hat{\tau},$$

the bootstrap relative error is given by

$$\hat{E}_{rel}^*(\hat{\tau}^*) = \frac{1}{\hat{\tau}} \left(\sum_{i=1}^{L} F_i^2 \widehat{\text{Var}}^*(X_{i\cdot}^*) \right)^{1/2}. \tag{4.2}$$

Note that the bootstrap variance for an exhaustive stratum is equal to zero, and hence the sum in (4.2) can be restricted to the non-exhaustive strata. Thus, we can write

$$\hat{E}_{rel}^*(\hat{\tau}^*) = \frac{1}{\hat{\tau}} \left(\sum_{i=1, F_i \neq 1}^{L} F_i^2 \widehat{\text{Var}}^*(X_{i\cdot}^*) \right)^{1/2}.$$

This clearly reduces the computational effort because no bootstrap resamples are generated from self-represented strata.

5. Simulation Study

An extensive simulation study was performed to compare the estimators of the relative error of the population mean, $\varphi = E_{rel}(\hat{\mu})$, described in the previous sections. Specifically, the following estimators $\hat{\varphi}$ were computed: the plug-in estimator $\hat{E}_{rel,PI}(\hat{\mu})$ given in (2.9); the jackknife estimators $\hat{E}_{rel,JACK,1}(\hat{\mu})$ (leaving out one value), $\hat{E}_{rel,JACK,2A}(\hat{\mu})$ and $\hat{E}_{rel,JACK,2B}(\hat{\mu})$ (both leaving out one stratum), and $\hat{E}_{rel,JACK,2B}(\hat{\mu})$ (jackknife in each stratum) given in (3.1), (3.2), (3.3) and (3.5), respectively; and the bootstrap estimators derived from (4.1) using the resampling algorithms by Bickel and Freedman (BF estimator) and by Booth, Butler and Hall (BBH estimator).

As the self-represented strata do not affect to the sampling error, corrected versions of the jackknife estimators were also constructed by omitting the pseudovalues associated with these strata. These corrected versions are referred as the original ones but adding the letter "C". Note that this correction is implicit by construction in the case of the plug-in and bootstrap estimators.

Three different types of response variables were simulated in our experiments:

Binary variables. The response is "1" or "0" and the

parameter of interest is $p = P("1")$, which was chosen to take values close to 0.50, 0.30, 0.15 and 0.05.

Multinomial variables. Variables taking four possible results denoted by A_1, A_2, A_3 and A_4 were considered. Here, the parameter of interest is the vector $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3, \theta_4)$, with $\theta_i = P(A_i)$, for $i = 1, 2, 3, 4$. Specifically, $\boldsymbol{\theta}$ was selected to take the value $(0.58, 0.24, 0.12, 0.06)$.

Continuous variables. Response variables generated from three possible absolutely continuous distributions: uniform, normal and exponential.

Different scenarios of high and low variability between the mean responses of the population strata were simulated. In all our experiments, the population size was $N = 5000$ data values, classified in $L = 10$ strata so that two of these strata were self-represented in the sampling. The first experiments were carried out with a sample size $n = 400$. Thus, the ratio between the sample and population sizes mimics the one in the Galician ACS conducted by IGE and that initially motivated the present work.

Our first results come from an experiment with binary response. The specific parameter values used to generate the population and the sample are shown in **Table 1**.

Table 1 summarizes the main features of the experiment. Thus, for instance, stratum L1 is formed by 1200 observations taking two possible values, "0" and "1", that have been randomly generated from Bernoulli trials with $P("1") = p_1 = 0.30$. Note that all the values p_i are around 0.5, but there is a high variability between them. The overall population consists of 5000 observations and satisfies that $\mu = p = P("1") = 0.5072$ and $\varphi = E_{rel}(\hat{\mu}) = 0.05135$.

Under this population design, $M = 1000$ samples of size $n = 400$ were generated, and hence 1000 estimates $\hat{\varphi}$ were obtained with each studied method. In the case of the bootstrap procedures, a set of $B = 1000$ bootstrap replicates was considered to compute each estimator. The behaviour of each estimation procedure was examined by using the following values:

i)	$E(\hat{\varphi}) = M^{-1}\sum_{r=1}^{M}\hat{\varphi}_r$	(iv)	$MSE(\hat{\varphi}) = \left[\text{Bias}(\hat{\varphi})\right]^2 + \text{Var}(\hat{\varphi})$
ii)	$\text{Bias}(\hat{\varphi}) = E(\hat{\varphi}) - \varphi$	(v)	$RMSE(\hat{\varphi}) = \sqrt{MSE(\hat{\varphi})}$
iii)	$sd(\hat{\varphi}) = \sqrt{\text{Var}(\hat{\varphi})}$	(vi)	$Effic(\hat{\varphi}) = MSE(\hat{\varphi})/MSE(\hat{\varphi}_{PI})$

The quantity $Effic(\hat{\varphi})$ measures the efficiency of each estimator $\hat{\varphi}$ with respect to the plug-in estimator, $\hat{\varphi}_{PI}$. Thus, $Effic(\hat{\varphi}) \leq 1$ means that the considered estimator $\hat{\varphi}$ presents better behavior than the plug-in estimator in terms of mean squared error. Results from our first experiment are shown in **Table 2**, where different scales have been used to obtain a more intuitive comparison.

Similar experiments were carried out using different sample sizes $n = \sum_{i=1}^{L} n_i'$, where the new sample size in the i-th stratum, $i = 1, \cdots, L$, was determined by $n_i' = k n_i$, with n_i being the sample size of the first experiment (see **Table 1**) and $k = 0.25, 0.30, 2, 3$ and 4. Excluded from this rule are the two self-represented strata for which obviously $n_i' = N_i$ in all cases. The values of $RMSE(\hat{\varphi})$ and $Effic(\hat{\varphi})$ obtained with these new sample sizes are presented in **Tables 3** and **4**, respectively.

New trials of our first experiment were run with different values of parameter $p = P("1")$. **Table 5** presents results of $RMSE(\hat{\varphi})$ obtained for different values of p.

Tables 2-**5** provide a sample of the results obtained in our extensive numerical study with binary response. Some interesting conclusions derived from this study are stated below.

- The plug-in method gave good results and was often the most efficient procedure. Moreover, it has the advantage of being computationally fast. It is also observed that its efficiency improves with increasing sample size.

Table 1. Parameters in the first experiment with binary response: population size (N_i), sample size (n_i), elevation factor (F_i) and $p_i = P("1")$ for the *i*-th stratum $(Li), i = 1, \cdots, 10$.

	Stratum										
	L1	L2	L3	L4	L5	L6	L7	L8	L9	L10	Total
N_i	1200	1000	800	600	450	350	300	240	40	20	5000
n_i	65	50	80	32	46	18	28	21	40	20	400
F_i	18.46	20	10	18.75	9.78	19.44	10.71	11.43	1	1	
p_i	0.30	0.60	0.68	0.40	0.45	0.55	0.50	0.75	0.50	0.69	

Table 2. Results for the experiment with binary response conducted under parameters in Table 1. Total population parameters: $p = P(\text{"1"}) = 0.5072$ and $\varphi = E_{rel}(\hat{\mu}) = 0.05135$. Results based on $M = 1000$ trials with sample size $n = 400$.

Method	Measure Scale	$E(\hat{\varphi})$ $\times 10^{-1}$	$Bias(\hat{\varphi})$ $\times 10^{-2}$	$sd(\hat{\varphi})$ $\times 10^{-2}$	$RMSE(\hat{\varphi})$ $\times 10^{-2}$	$MSE(\hat{\varphi})$ $\times 10^{-3}$	$Effic(\hat{\varphi})$ $\times 1$
Plug-in		0.5148	0.0127	0.2597	0.2600	0.0068	1.0000
Jackknife 1		0.5382	0.2472	0.2472	0.3692	0.0136	1.4198
Jackknife 1.C		0.5380	0.2450	0.2741	0.3677	0.0135	1.4140
Jackknife 2A		0.6042	0.9064	1.3101	1.5931	0.2538	6.1268
Jackknife 2A.C		0.6132	0.9963	1.3224	1.6557	0.2741	6.3676
Jackknife 2B		1.4605	9.4694	2.8553	9.8906	9.7283	38.0378
Jackknife 2B.C		1.4547	9.4121	2.8302	9.8356	9.6740	37.8263
Jackknife 3		0.5301	0.1662	0.2821	0.3274	0.0107	1.2592
Jackknife 3.C		0.5299	0.1641	0.2820	0.3263	0.0106	1.2548
Bootstrap BF		0.5199	0.0635	0.2777	0.2849	0.0081	1.0957
Bootstrap BBH		0.5131	0.0047	0.2707	0.2707	0.0073	1.0412

Table 3. Values of $RMSE(\hat{\varphi}) \times 10^{-2}$ for different sample sizes.

Method	$n = 145$	$n = 230$	$n = 400$	$n = 740$	$n = 1080$	$n = 1420$
Plug-in	1.1840	0.5366	0.2600	0.127	0.075	0.0052
Jackknife 1	1.5743	0.7060	0.3692	0.298	0.329	0.0380
Jackknife 1.C	1.5737	0.7051	0.3677	0.296	0.325	0.0376
Jackknife 2A	4.3400	2.0697	1.5931	2.572	3.177	0.3616
Jackknife 2A.C	4.2781	2.0474	1.6557	2.660	3.267	0.3707
Jackknife 2B	8.5497	9.0836	9.8906	10.957	11.401	1.1773
Jackknife 2B.C	8.4640	9.0140	9.8356	10.912	11.359	1.1732
Jackknife 3	1.2161	0.6026	0.3274	0.280	0.318	0.0366
Jackknife 3.C	1.2158	0.6021	0.3263	0.277	0.315	40.0362
Bootstrap BF	1.2351	0.5643	0.2849	0.160	0.171	0.0157
Bootstrap BBH	1.2605	0.5715	0.2707	0.137	0.090	0.0072

Table 4. Results of $Effic(\hat{\varphi})$ for different sample sizes.

Method	$n = 145$	$n = 230$	$n = 400$	$n = 740$	$n = 1080$	$n = 1420$
Plug-in	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Jackknife 1	1.3296	1.3156	1.4198	2.3526	4.3525	7.3621
Jackknife 1.C	1.3290	1.3141	1.4140	2.3306	4.3041	7.2785
Jackknife 2A	3.6656	3.8572	6.1268	20.2704	42.0836	70.0182
Jackknife 2A.C	3.6133	3.8156	6.3678	20.9679	43.2797	7107821
Jackknife 2B	7.2211	16.9287	38.0378	86.3640	151.0166	227.9469
Jackknife 2B.C	7.1487	16.7990	37.8263	86.0093	150.4603	227.1530
Jackknife 3	1.0271	1.1230	1.2592	2.2079	4.2156	7.0899
Jackknife 3.C	1.0269	1.1220	1.2548	2.1869	4.1686	7.0090
Bootstrap BF	1.0432	1.0516	1.0957	1.2631	2.2631	3.0474
Bootstrap BBH	1.0646	1.0707	1.0412	1.0806	1.1895	1.3961

Table 5. Results of $RMSE(\hat{\boldsymbol{\varphi}}) \times 10^{-2}$ for different values of $p = P("1")$.

Method	$p = 0.050$	$p = 0.148$	$p = 0.298$	$p = 0.507$
Plug-in	0.2812	0.9716	0.4892	0.2600
Jackknife 1	0.3426	1.2228	0.6364	0.3692
Jackknife 1.C	0.3415	1.2171	0.6337	0.3677
Jackknife 2A	0.7920	7.0413	3.5333	1.5931
Jackknife 2A.C	0.7720	6.9848	3.4795	1.6557
Jackknife 2B	1.8147	6.4254	6.1898	9.8906
Jackknife 2B.C	1.7825	6.3505	6.1220	9.8356
Jackknife 3	0.3062	1.1173	0.5958	0.3274
Jackknife 3.C	0.3052	1.1127	0.5934	0.3263
Bootstrap BF	0.3132	1.0505	0.5259	0.2849
Bootstrap BBH	0.29184	1.0212	0.5065	0.2707

- Both bootstrap methods (BF and BBH) yield competitive results. The BBH bootstrap presents better behaviour than the BF bootstrap, especially for large sample sizes. For moderate or small sample sizes (less than 10% of the population size), the bootstrap methods behave similarly to the plug-in method in terms of efficiency, although with a higher computational cost (that, in any case, results to be acceptable).

- The jackknife 3 estimator (based on applying jackknife to each stratum) behaved similarly to the bootstrap estimators for small sample sizes $(n < 400)$ and worsened with increasing n. Jackknife 1 estimator (based on leaving out one sample datum) was competitive although yielded worse results than the jackknife 3 with small sample sizes. Both variants of the jackknife 2 estimator (based on leaving out one stratum) yielded much worse results for all the considered sample sizes. Therefore, it is not advisable to use the jackknife 2 estimators. In general, the worse behaviour of the jackknife-based estimators seems to be mainly due to the bias of the estimation, although for the jackknife 2 estimators poor results are also observed in terms of standard deviation.

- No substantial differences were observed between using the jackknife procedures and their corrected versions (.C), based on cancelling the exhaustive strata. Just a slight advantage for the corrected versions is observed.

- Similar conclusions are valid for the two scenarios of high and low variability of the response between different strata. In general, the plug-in estimator is the most efficient in both situations, although the differences between the plug-in and the other methods are smaller in the case of low variability.

- For small sample sizes $(n = 145)$, both two bootstrap methods and jackknife 3 led to a greater efficiency than the plug-in method.

- Results in **Table 5** allows us to conclude that prior comments are valid for binary variables regardless of the specific value taken by $p = P("1")$. In particular, it is observed that *RMSE* takes the smallest values when p is close to 0 or 0.5.

Next step in our simulation study is addressed to analyze the case of multiple response. Specifically, it is assumed that there are four mutually exclusive and exhaustive results for the response variable, let us say A_1, A_2, A_3 and A_4. As in the case of binary response, a population of $N = 5000$ observations divided in $L = 10$ strata was simulated. Data forming the j-th stratum were randomly generated from a multinomial distribution with parameter $\boldsymbol{\theta}^j = \left(\theta_1^j, \theta_2^j, \theta_3^j, \theta_4^j\right)$, where θ_i^j denotes the probability of event A_i in the j-th stratum, $\theta_i^j = P\left(A_i^j\right), i = 1, 2, 3, 4$. Different values of $\boldsymbol{\theta}^j$ were selected to set up scenarios with high and low variability of the responses between strata. In all cases, the overall population, which is formed by bringing together all strata, presents the theoretical parameter vector $\boldsymbol{\theta} = \left(\theta_1, \theta_2, \theta_3, \theta_4\right) = \left(0.58, 0.24, 0.12, 0.06\right)$, where $\theta_i = P\left(A_i\right), i = 1, 2, 3, 4$.

A total of $M = 1000$ samples of size $n = 400$ were randomly selected and each of them was used to construct an estimator $\hat{\boldsymbol{\varphi}}$ of the relative error in the sampling of $\hat{\boldsymbol{\theta}} = \left(\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3, \hat{\theta}_4\right)$. Again, two strata were considered to be exhaustive in the sampling. The theoretical error for the simulated population is $\boldsymbol{\varphi} = \left(0.05, 0.09, 0.15, 0.21\right)$. The set of estimates obtained with each of the studied methods $\left\{\hat{\boldsymbol{\varphi}}_r, r = 1, \cdots, M\right\}$ is then used to calculate the mean squared error and the efficiency of each procedure with respect to the plug-in method. Both quantities were simultaneously obtained for each estimated marginal component $\left(\hat{\varphi}_i, i = 1, 2, 3, 4\right)$ and jointly for the estimated vector $(\hat{\boldsymbol{\varphi}})$. In particular, the global $RMSE(\hat{\boldsymbol{\varphi}})$ is given by

$$RMSE(\hat{\boldsymbol{\varphi}}) = \sqrt{MSE(\hat{\boldsymbol{\varphi}})} = \left(\frac{1}{M} \sum_{r=1}^{M} (\hat{\boldsymbol{\varphi}}_r - \boldsymbol{\varphi})^t (\hat{\boldsymbol{\varphi}}_r - \boldsymbol{\phi}) \right)^{1/2}.$$

Results from this new simulation are shown in **Tables 6** and **7**.

Results from **Tables 6** and **7** show that the conclusions derived from simulations with binary response are equally valid for the case of multiple response.

Finally, we focus on the case of continuous response variables. Thus, new experiments were carried out to examine the performance of the different estimation procedures with responses generated from uniform, normal and exponential distributions. The simulation plan was designed by following the same outline as in previous experiments.

For instance, some experiments with normal and exponential responses were carried out considering the features summarized in **Table 8**. Note that these parameters lead to a situation of high variability between data from different strata. Some outcomes derived from these experiments are presented in **Tables 9** and **10**, for the Gaussian case, and in **Tables 11** and **12**, for the exponential case. Here, the target parameter in the overall population, $\varphi = E_{rel}(\hat{\mu})$, took the value $\varphi = 0.0177$ for the Gaussian response, and $\varphi = 0.05365$, for the exponential response.

Table 6. Results of $RMSE(\hat{\boldsymbol{\varphi}}) \times 10^{-2}$ for the experiment with multiple response and high response variability between strata. Results based on $M = 1000$ trials with sample size $n = 400$.

Method	$RMSE(\hat{\varphi}_1)$	$RMSE(\hat{\varphi}_2)$	$RMSE(\hat{\varphi}_3)$	$RMSE(\hat{\varphi}_4)$	$RMSE(\hat{\boldsymbol{\varphi}})$
Plug-in	0.2556	0.5974	1.3419	2.8867	3.2490
Jackknife 1	0.3344	0.7827	1.7149	3.5292	4.0150
Jackknife 1.C	0.3333	0.7799	1.7121	3.5243	4.0089
Jackknife 2A	2.8069	5.8041	8.7689	12.2577	16.3924
Jackknife 2A.C	2.7814	5.7443	8.7550	12.2467	16.3924
Jackknife 2B	1.7803	3.7279	5.4087	9.3790	11.5882
Jackknife 2B.C	1.7725	3.7293	5.3482	9.1843	11.4002
Jackknife 3	0.3157	0.7037	1.6276	3.2936	3.7539
Jackknife 3.C	0.3147	0.7014	1.6250	3.2890	3.7482
Bootstrap BF	0.2716	0.6176	1.3581	3.0913	3.4432
Bootstrap BBH	0.2556	0.5993	1.3701	3.2010	3.5424

Table 7. Results of $Effic(\hat{\boldsymbol{\varphi}})$ for the experiment with multiple response and high response variability between strata. Results based on $M = 1000$ trials with sample size $n = 400$.

Method	$n = 145$	$n = 230$	$n = 400$	$n = 740$	$n = 1080$	$n = 1420$
Plug-in	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Jackknife 1	1.1427	1.1322	1.2358	1.6634	2.4630	3.4143
Jackknife 1.C	1.1523	1.1315	1.2339	1.6572	2.4503	3.3943
Jackknife 2A	2.0393	3.2180	5.0454	6.6219	7.4788	7.1866
Jackknife 2A.C	2.0346	3.2084	5.0327	6.6123	7.4769	7.1893
Jackknife 2B	1.5625	2.1351	3.5667	6.5472	10.5012	13.8590
Jackknife 2B.C	1.5410	2.1142	3.5088	6.3559	10.1349	13.3493
Jackknife 3	0.9652	1.0451	1.1554	1.6130	2.3709	3.3767
Jackknife 3.C	0.9818	1.0442	1.1536	1.6071	2.3587	3.3573
Bootstrap BF	1.0101	0.9744	1.0598	1.1614	1.5541	1.8045
Bootstrap BBH	1.0056	1.0889	1.0903	1.0709	1.1188	1.1535

Table 8. Main features for experiments with normal and exponential responses: population size (N_i), sample size (n_i) and distribution parameters for the i-th stratum (Li).

	Stratum										
	L1	L2	L3	L4	L5	L6	L7	L8	L9	L10	Total
N_i	1200	1000	800	600	450	350	300	240	40	20	5000
n_i	65	50	80	32	46	18	28	21	40	20	400
	Normal										
μ_i	50	55	46	45	56	44	57	55	46	50	
σ_i	10	22	15	10	9	20	21	19	11	12	
	Exponential										
λ_i	0.10	0.11	0.09	0.10	0.11	0.10	0.11	0.09	0.10	0.08	

Table 9. Results from the experiment with Gaussian response conducted according to parameters in Table 8 and based on $M = 1000$ trials with sample size $n = 400$.

Measure	$E(\hat{\varphi})$	Bias$(\hat{\varphi})$	$sd(\hat{\varphi})$	$RMSE(\hat{\varphi})$	$MSE(\hat{\varphi})$	$Effic(\hat{\varphi})$
Method	Scale	$\times 10^{-1}$	$\times 10^{-2}$	$\times 10^{-2}$	$\times 10^{-2}$	$\times 10^{-3}$	$\times 1$

Method	$E(\hat{\varphi})$ $\times 10^{-1}$	Bias$(\hat{\varphi})$ $\times 10^{-2}$	$sd(\hat{\varphi})$ $\times 10^{-2}$	$RMSE(\hat{\varphi})$ $\times 10^{-2}$	$MSE(\hat{\varphi})$ $\times 10^{-3}$	$Effic(\hat{\varphi})$ $\times 1$
Plug-in	1.7691	−0.0086	0.1081	0.1085	0.0012	1.0000
Jackknife 1	1.8541	0.0756	0.1117	0.1349	0.0018	1.2441
Jackknife 1.C	1.8531	0.0753	0.1117	0.1348	0.0018	1.2425
Jackknife 2A	1.5536	−0.2441	0.3690	0.4425	0.0196	4.0799
Jackknife 2A.C	1.5360	−0.2418	0.3675	0.4399	0.0194	4.0563
Jackknife 2B	4.1834	2.4056	0.8390	2.5478	0.6491	23.4929
Jackknife 2B.C	4.1350	2.3573	0.8122	2.5021	0.6261	23.0715
Jackknife 3	1.8265	0.0487	0.1070	0.1175	0.0014	1.0837
Jackknife 3.C	1.8261	0.0484	0.1070	0.1174	0.0014	1.0826
Bootstrap BF	1.7781	0.0003	0.1062	0.1062	0.0011	0.9793
Bootstrap BBF	1.7714	−0.0064	0.1078	0.1080	0.0012	0.9959

Table 10. Results of $Effic(\hat{\varphi})$ for Gaussian response and different sample sizes.

Method	$n = 145$	$n = 230$	$n = 400$	$n = 740$	$n = 1080$	$n = 1420$
Plug-in	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Jackknife 1	1.2080	1.1241	1.2441	1.9405	3.3611	5.6684
Jackknife 1.C	1.2079	1.2237	1.2425	1.9327	3.3438	5.6372
Jackknife 2A	3.9205	4.5127	4.0799	7.5143	16.6298	31.3715
Jackknife 2A.C	3.9042	4.4950	4.0563	7.5016	16.6140	31.3358
Jackknife 2B	4.4365	10.2989	23.4929	54.3585	93.2693	145.4229
Jackknife 2B.C	4.3642	10.1115	23.0715	53.5200	91.8624	143.3255
Jackknife 3	0.9942	1.0336	1.0837	1.8723	3.2635	5.5708
Jackknife 3.C	0.9943	1.0335	1.0826	1.8648	3.2468	5.5406
Bootstrap BF	1.0140	0.9996	0.9793	1.0183	1.8710	2.0230
Bootstrap BBH	1.0409	1.0193	0.9959	1.0675	1.1031	1.2344

Table 11. Results from the experiment with Exponential response conducted according to parameters in Table 8 and based on $M = 1000$ trials with sample size $n = 400$.

Method	Measure Scale	$E(\hat{\varphi})$ $\times 10^{-1}$	$Bias(\hat{\varphi})$ $\times 10^{-2}$	$sd(\hat{\varphi})$ $\times 10^{-2}$	$RMSE(\hat{\varphi})$ $\times 10^{-2}$	$MSE(\hat{\varphi})$ $\times 10^{-3}$	$Effic(\hat{\varphi})$ $\times 1$
Plug-in		5.3474	−0.0179	0.2894	0.2899	0.08410	1.0000
Jackknife 1		5.5928	0.2275	0.2915	0.3698	0.1368	1.2756
Jackknife 1.C		5.5928	0.2252	0.2917	0.3685	0.1358	1.2710
Jackknife 2A		2.4534	−2.9119	0.8971	3.0470	9.2841	10.5098
Jackknife 2A.C		2.4865	−2.8788	0.9022	3.0168	9.1013	10.40556
Jackknife 2B		6.4268	1.0615	2.1040	2.3566	5.5537	8.1284
Jackknife 2B.C		6.3898	1.0245	2.0803	2.3402	5.4766	8.0719
Jackknife 3		5.5246	0.1593	0.3012	0.3407	0.1161	1.1751
Jackknife 3.C		5.5224	0.1571	0.3013	0.3398	0.1154	1.1719
Bootstrap BF		5.3963	0.0310	0.2991	0.3007	0.0904	1.0371
Bootstrap BBH		5.3442	−0.0211	0.2921	0.2928	0.0857	1.0100

Table 12. Results of $Effic(\hat{\varphi})$ for exponential response and different sample sizes.

Method	$n = 145$	$n = 230$	$n = 400$	$n = 740$	$n = 1080$	$n = 1420$
Plug-in	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Jackknife 1	1.1165	1.1730	1.2756	2.2765	4.0523	6.4474
Jackknife 1.C	1.1162	1.1720	1.2710	2.2561	4.0078	6.3742
Jackknife 2A	5.5146	8.1224	10.5098	12.4363	13.3740	12.6355
Jackknife 2A.C	5.4883	8.0736	10.4056	12.2327	13.0697	12.2639
Jackknife 2B	2.8584	4.8487	8.1284	15.7562	26.8622	38.1192
Jackknife 2B.C	2.8457	4.8210	8.0719	15.7186	26.8389	38.1956
Jackknife 3	0.9698	1.0215	1.1751	2.1765	3.9274	6.3672
Jackknife 3.C	0.9700	1.0215	1.1719	2.1572	3.8847	6.2964
Bootstrap BF	1.0320	1.0097	1.0371	1.2404	2.0056	2.8627
Bootstrap BBH	1.0358	1.0717	1.0100	1.0406	1.1994	1.2663

Results in **Tables 9-12** allow us to confirm that the different analyzed estimation procedures behaved as in the previous experiments. Moreover, an analogous behaviour was also observed with uniform response and with scenarios of low variability between strata. In short, the conclusions derived from our numerical study with binary or multinomial response can be extended to the case of continuous response variables, regardless of the generating probability distribution.

6. Final Conclusion

The present work deals with the problem of estimating the sampling relative error of point estimates derived from large sample surveys. It is specifically assumed that the survey's sampling design is the stratified random sampling without replacement, because this design is often considered in many surveys conducted by different official statistics institutions. Variables commonly ex-

amined in this type of surveys are binary, categorical and continuous, and hence, the estimates of interest involve estimates of proportions, totals and means. In this setting, several procedures to approximate the sampling relative error of this kind of estimates are proposed. Different estimation techniques are considered, including the natural estimation of plug-in type and more sophisticated methods based on jackknife and bootstrap methodologies. The behaviour of the different procedures proposed is examined and compared by means of an extensive simulation study. In general, the plug-in method presents good behaviour in all the analyzed situations, with the additional advantage of having a low computational cost. For small sample sizes, the jackknife estimator denoted by "jackknife 3", which is based on the prior application of the jackknife technique to each stratum, and the two bootstrap methods considered (particularly the bootstrap proposed by Booth, Butler and Hall) yield results similar

as those obtained with the plug-in estimator, and in some cases, even better. However, the estimators obtained with these methods have a higher (although acceptable) computational cost.

7. Acknowledgements

REFERENCES

[1] S. K. Thompson, "Sampling," Wiley, New York, 1992.

[2] M. Quenouille, "Approximate Tests of Correlation in Time Series," *Journal of the Royal Statistical Society*: *Series B* (*Statistical Methodology*), Vol. 11, No. 1, 1949, pp. 18-84.

[3] B. Efron, "Bootstrap Methods: Another Look at the Jackknife," *The Annals of Statistics*, Vol. 7, No. 1, 1979, pp. 1-26. doi:10.1214/aos/1176344552

[4] B. Efron, "The Jackknife, the Bootstrap, and Other Resampling Plans, Siam Monograph 38," Society of Industrial and Applied Mathematics CBMS-NSF Monographs, 1982.

[5] B. Efron and R. Tibshirani, "An Introduction to the Bootstrap," Chapman & Hall, New York, 1993.

[6] P. J. Bickel and D. A. Freedman, "Asymptotic Normality and the Bootstrap in Stratified Sampling," *The Annals of Statistics*, Vol. 12, No. 2, 1984, pp. 470-482. doi:10.1214/aos/1176346500

[7] J. G. Booth, R. W. Butler and P. Hall, "Bootstrap Methods for Finite Populations," *Journal of the American Statistical Association*, Vol. 89, No. 248, 1994, pp. 1282-1289. doi:10.1080/01621459.1994.10476868