STATISTICS IN TRANSITION new series, June 2018
Vol. 19, No. 2, pp. 201–218, DOI 10.21307/stattrans-2018-012

# UNIVARIATE SAMPLE SIZE DETERMINATION BY ALTERNATIVE COMPONENTS: ISSUES ON DESIGN EFFICIENCY FOR COMPLEX SAMPLES

**Ceylan Talu Yozgatligil**[1], **H. Öztaş Ayhan**[2]

## ABSTRACT

Sample size determination for any sample survey can be based on the desired objectives of the survey as well as the level of confidence of the desired estimates for some survey variables, the desired precision of the survey results and the size of the population. In addition to these, the cost of enumeration can also be considered as an important criterion for sample size determination. Recently, some international organizations have been using univariate sample size determination approaches for their multivariate sample designs. These approaches also included some design efficiency and error statistics for the determination of the univariate sample sizes. These should be used for determining the survey quality measures after the data collection, not before. The additional components of the classical sample size measure will create selection and representation bias of survey estimates, which is discussed in this article.

**Key words:** univariate sample size, representation bias, sample allocation, error statistics, design efficiency measures.

## 1. Introduction

Sample size determination for univariate cases has been commonly used for many years. Surveys which are based on large population sizes require other sample size determination methodologies than the univariate cases, because they are based on criteria for multivariate observations. Therefore, univariate sample size determination methodologies cannot satisfy the multivariable criteria. Recently, some national and international survey organizations have been using some modified univariate sample size determination formulas which have low efficiency. As a result, this can lead to under- or overrepresentation of the population by the selected sample. These modified formulas contain unnecessary components, such as *design effect* and *response* or *nonresponse rates*, etc. This article highlights the components which create selection and representation bias of survey estimates. Therefore, the methodological problem is not the concern of

[1] Department of Statistics, Middle East Technical University, 06800 Ankara, Turkey.
E-mail: ceylan@metu.edu.tr
[2] Department of Statistics, Middle East Technical University, 06800 Ankara, Turkey.
E-mail: oayhan@metu.edu.tr

this article. The main aim is to emphasize the correct usage of the sample size determination formulas for the multivariable case. Some researchers may want to follow the methodology (or formula) used by the respectful organizations, which may not fully (or correctly) represent the target population.

## 2. Classical Univariate Sample Size Measures

Sample size determination is an important aspect of the representativeness of the survey results. There are many approaches which can be taken. Generally, all surveys utilize too many variables. Some of these variables may be more important than others for decision-makers. The researcher generally wishes to satisfy the representation of several survey variables, which are important.

There are many studies on determination of sample size in different disciplines. Some of them propose a new methodology and some others gather the existing ones and compare their performances. Dell et al. (2002) discussed simple methods of estimating the number of animals needed for various types of variables and experiments. They showed that it is crucial to choose the power, the significance level, and the size of the effect to be detected, and to estimate the population variability of the variable being studied, and using a complicated design and statistical analysis usually results in the highest power to detect any difference. Shore (2008) addressed sample size determination relating to hypothesis testing, parameter estimation, relational modelling and optimal sampling. Sathien et al. (2010) gave a few suggestions regarding the methods to determine an optimum sample size in descriptive and analytical studies. Marshall et al. (2013) described basic requirements for sample size determination and the sample size determination methods to estimate a normal distribution mean, standard deviation, quantile, binomial proportion and Poisson occurrence rate. In his book, Ryan (2013) discussed many sample size calculation techniques with applications using software. Siddiqui (2013) presented the guidelines described in the literature as to determine the appropriate sample size for the various statistical techniques. Safo et al. (2015) compared the performance of the existing sample size method and the sample size method developed by the authors for lasso logistic regression. Placzek and Friede (2017) proposed methods for planning and analysing a multiple nested subgroups design and described sample size determination prior to the trial and sample size recalculation via a blinded review in an internal pilot study.

For the representation of survey results, a very important single survey variable (*univariate case*) can be chosen and the sample size only for this variable can be evaluated. Alternatively, two variables (*bivariate case*) may affect one another and the sample size determination may be based on the presence of these. Finally, several variables (*multivariate case*) may become very important to determine the minimum sample size, by utilizing multivariate information. One of the common and most practical solution to these problems is to select several independent variables (*univariate case*) and compute sample sizes for each of these separately and choose the largest computed sample size to satisfy all variables.

The use of several survey variables one at a time has some practical conveniences. On the other hand, the type of measurement scale of the survey variable also leads to the use of different test statistics as input information for the sample size determination model. Here, the case of the proportion and another case of the sample mean for the determination of sample size will be illustrated.

### 2.1. Determination of Sample Size for Proportion

For the test of the following hypothesis; $H_0 : \theta = \theta_0$ versus $H_1 : \theta = \theta_1$, the sample proportion (*p*) of success is distributed asymptotically as $N\{\theta, \theta(1-\theta)/n\}$ with the requirement that $\Pr\left[\,|p-\theta| \le d \mid \theta\right] \ge 1-\alpha$. This leads to the sample size estimation as: $n \ge \theta\left(1-\theta\right) \chi^2_{(1),\alpha} \big/ d^2$, where $\alpha$ is the level of significance and *d* is the level of tolerance of the estimate. For the unknown population proportion, we take $\theta$ = 0.5 and the sample size estimate will be: $n \ge 0.25\, \chi^2_{(1),\alpha} \big/ d^2$ . This naturally represents the worst case, which creates the maximum variance, to be on the safe side as the sample designer. If we have prior information about the population proportion, then consequently we can have relatively smaller sample size estimation.

Hence, the sample size can be determined as: $n \ge 0.25\, \chi^2_{(1),\alpha} \big/ d^2$ . The overall sampling fraction for this design will be, *f* = *n/N* = 1/F. Here, (*N*) is the total number of Housing Units (HUs) in the population, and (*n*) is the total number of Housing Units (HUs) in the selected sample. For self-weighting sample designs, the sampling fraction for any domain will be the same as any other domain in the design. Furthermore, this will also be equal to one another within any prefecture as well as the total population.

### 2.2. Determination of Sample Size for Frequency Type Variable

The frequentist case of sample size determination is concerned with the normal distribution with known variance. When the random variable *X* is distributed as $N\left(\mu, \sigma^2\right)$, the mean $\mu$ may be estimated with absolute error (*d)* and probability 1 – *α* by the sample mean (*m*) if

$$\Pr\left[\,|m-\mu| \le d \mid \mu\right] \ge 1-\alpha.$$

Since $\sqrt{n}(m-\mu)/\sigma \sim N(0,1)$, it follows that the above inequality is satisfied when the sample size (*n*) satisfies $n \ge \sigma^2\, Z^2_{\alpha/2} \big/ d^2$ . Here, tolerance level refers to $d = Z_{\alpha/2}\left(\hat{\sigma}/\sqrt{n}\right)$. We can also easily create an application for this case just like the previous one. If we take the same element variance value and the same tolerance level for this case, then the estimated sample size will be the same as before. Hence, the sample size is determined by

$$n \ge \sigma^2\, Z^2_{\alpha/2} \big/ d^2$$

formula, which is affected by the level of confidence, level of tolerance (desired error variance) and the element variance. Due to changes in these parameters will consequently result in differing outcomes.

## 3. Sample Design for the Survey

The sample design for the survey will be based on the latest available information on the population. The population will be stratified into domains (prefectures) and self–weighting samples will be selected for each domain.

### 3.1. Sampling Frame and Stratification

The latest population figures are based on the population projections for the survey time. The aggregated data from the urban–rural information for the available districts will be aggregated into several prefectures within a nested structure in defined geographic areas. Dividing the *total urban and rural population* $(M_h)$ for each domain (prefecture) by their Population Census *average household size* $(\overline{H}_h)$ of each prefecture, we can compute the *number of urban and rural Housing Units* $(N_h = M_h / \overline{H}_h)$ for the survey date. This calculation is based on the assumption that: the *average household size does not change significantly over the years*. This assumption is verified and used in many countries of the world.

In summary, Desu and Raghavarao (1990) and Adcock (1997) proposed the following measures for frequentist methods.

$n_0 \geq \sigma^2 Z^2_{\alpha/2} / d^2$ where $d$ is the *absolute error*; $d = Z_{\alpha/2} \dfrac{\hat{\sigma}}{\sqrt{n}}$ .

Alternatively, for studies aiming at the hypothesis testing

$$n^* = \sigma^2 \left(Z_{\alpha/2} + Z_\beta\right)^2 / d^2 .$$

For the studies with binary response, i.e. binomial distribution,

$$n_0 \geq \theta(1-\theta)\, \chi^2_{(1),\alpha} / d^2 .$$

The ultimate sample size is adjusted for the known population size as:

$$n = \frac{n_0}{1 + \frac{n_0}{N}} .$$

### 3.2. Measures of Design Efficiency

The following measures of the design efficiency are commonly used for many surveys, after the data collection. There are several measures of design efficiency in survey research. Basically, it is the ratio of sampling variances, which is based on two different sample designs. The comparison of the two variances has to be based on the same sample sizes for both designs.

### *i. Design efficiency*

Design efficiency is the ratio of two sampling variances for given sample designs (*Di; i=1,2*).

$$DesEff = \mathrm{var}(\bar{y}_{D1}) / \mathrm{var}(\bar{y}_{D2})$$

where *D2* is not based on Simple Random Sample (*SRS*) design.

### *ii. Design effect*

A design effect (*deff*) measures the relative increase or decrease in the variance of an estimator due to departures from simple random sampling. Kish (1965) presented *deff* as a convenient way of gauging the effect of clustering on an estimator of a mean (Henry and Valliant, 2015). Later work by Rao and Scott (1984) and others found that more complicated versions of *deff*'s were useful to adjust inferential statistics calculated from complex survey data (Sirken, 2002).

A specialized version of *deff* was proposed in Kish (1965), who addressed only the effect of using weights that are not all equal. Kish derived the “design effect due to weighting” for a case in which weights vary for reasons other than statistical efficiency (Henry and Valliant, 2015). There are also sample designs and estimators where having varying weights can be quite efficient.

Design effect is the ratio of two sampling variances for given sample designs.

$$deff = \mathrm{var}\,(\bar{y}_{D1}) / \mathrm{var}\,(\bar{y}_{SRS})$$

where *Design 2* is based on *SRS* only (Kish, 1965 & 1982). The original definition of the design effect is based on the sampling variance of a given complex design, which is compared with the SRS sampling variance of the same sample size. Theoretically, SRS has to be taken as an independent sample from the same population rather than adjusting the complex sample design boundaries as if it was selected as a SRS.

The efficient sample size calculations assume simple random sampling. If the sample design deviates from SRS, the efficient sample size will also vary. *deff* is a measure for the relative efficiency of an estimator under a studied sampling design. It is the direct way of measuring the effect of design on sampling variability. The planned sample size computation for the univariate case naturally corresponds to the “gross sample size”. After the data collection “net sample size” will be achieved. The difference can be reflected through the computation of the nonresponse amount. On the other hand, the *deff* computation will be based on the sampling variance of the existing data, which is collected from the net sample size. Naturally, this will not include the planned inclusion probabilities and the clustering for the complex sample design in particular.

### *iii. Design factor*

Design factor is the ratio of two standard errors for given sample designs (Kish, 1965).

$$deft = se\,(\bar{y}_{D1}) / se\,(\bar{y}_{SRS})$$

where *Design 2* is based on *SRS* only. Here, $deft = \sqrt{deff}$ and $deft^2 = deff$ .

*deft* is a measure of efficiency of a given sample design compared to a direct simple random sampling of individuals, defined as the ratio between the standard error using the given sample design and the standard error that would result if a simple random sample was used. A *deft* value of 1.0 indicates that the sample design is as efficient as the simple random sample.

### 3.3. Computed Error Statistics for the Analysis of Design Efficiency

The sample design efficiency for a given design will be compared with some error statistics, in order to show the data quality measures. These measures are based on the error statistics which are based on the complex multivariate designs when compared with the base design, which is SRS with replacement. The basic error statistics which are obtained for this comparison will be: *standard error, design factor, design effect, rate of homogeneity, cluster size*, etc. Some examples of these statistics are given in Table 1 below, which is based on the "2013 Turkey: Population and Health Survey" (HÜNEE, 2014). It is based on the complex sample survey design, which has 14,496 target sample households. The total sample household population was 41,476 persons. The household population consists of 78% urban and 22% rural domains. The aim of the presentation of these figures is merely to highlight the importance and usage of these error statistics. Here, the interpretation of the survey results is not intended to be the main purpose of this study.

**Table 1.** Sampling Related Error Statistics for Selected Survey Variables Turkey 2013

| Survey Variables | Ratio mean *r = y/x* | Standard error *se(r)* | Design factor *deft* | Relative error *se(r)/r* |
|---|---|---|---|---|
| Never married women | 0.275 | 0.006 | 1.277 | 0.021 |
| Currently married women | 0.683 | 0.006 | 1.276 | 0.009 |
| Number of live births | 1.667 | 0.020 | 1.133 | 0.012 |
| Number of living children | 2.919 | 0.050 | 1.252 | 0.017 |
| Wants no children | 0.474 | 0.007 | 1.202 | 0.015 |
| Ideal number of children | 2.721 | 0.019 | 1.507 | 0.007 |
| Total fertility rate (*3 years*) | 2.258 | 0.069 | 1.360 | 0.031 |
| Infant mortality rate (*5 years*) | 13.282 | 2.345 | 1.111 | 0.177 |

*Source:* HÜNEE (2014).

The purpose of computing these statistics is to compare the efficiency of the latest design used. On the other hand, some survey institutions are mistakenly proposing to include these error statistics into their selection procedures. They are utilizing a univariate sample size formula, which is combined with some of these error statistics as well as response or nonresponse rate components, in order to pre-adjust the sample size. This article has shown that the use of additional unrelated components will create the selection and representation bias for the estimation of selected population parameters.

## 4. Some Modified Sample Size Estimators

For large scale surveys, an ideal way of obtaining the required sample size should be based on multivariate sample size determination. Some international organizations are insisting on using univariate sample size determination methods with some modifications to the formulae in place. For this case, their argument is based on using the univariate sample size methodology for several design variables separately. Then, they intend to compensate for the missing components by adding some error statistics (*deff, nonresponse*, etc.) in advance which are based on complex sample designs. They also argue that adding these statistics to their modified sample size formulae will solve their methodological bias.

These error statistics are theoretically used for measuring the design efficiencies of their complex sample designs, when compared with the unrestricted design (i.e. SRS-WR). However, they are not proposed to be used prior to sample selection as an additional design component. Another important point is when these additional components are used within the desired sample selection formulae, they will naturally effect the overall sample selection probabilities in an undesired way, which will create sample selection bias. Consequently, it is not advised to use the modified univariate sample size determination formulas of this type. We would like to justify our argument by giving two different modified formulas in the following subsections.

Survey sampling statisticians are responsible for designing sample surveys and determining the ideal and unbiased sample results for their surveys. When they are comparing their survey results with several internationally organized surveys, where their sample selection was biased due to the use of undesired sample size formulation, which created biased results. Consequently, these methodological problems naturally concern survey sampling statisticians overall. In addition, naturally these issues have to be brought to the attention of survey methodology community.

### 4.1. Demographic and Health Surveys (DHS)

The DHS (2012) has used the following formula for calculating the final sample size in terms of the number of households while taking design effect and non-response into account in advance, and is given by:

$$n_{DHS} = deft^2 \frac{(1/P-1)}{\alpha^2} \Big/ (R_i \; R_h \; d).$$

The formula in terms of our notation is given by

$$n_{DHS} = deft^2 \frac{(1/p-1)}{d^2} \Big/ (R_i \; R_h \; e)$$

where

*n* is the sample size in households;

*deft** is the design effect (a default value of 1.5 is used for *deft* if not specified);

*p* is the estimated proportion;

*d* is the desired relative standard error, $\sqrt{p(1-p)/n}\Big/p$ ;

$R_i$ is the individual response rate;

$R_h$ is the *household gross response rate*; and

*e* is the number of eligible individuals per household.

(*): The symbol *deft* actually denotes the design factor, not the design effect. (Default value of *deft=1.5* is recommended in DHS manual as a special case. Naturally, this corresponds to *deff=2.25*). In practice, this can be an acceptable threshold value for complex clustered sample designs.

The *household gross response rate* is the number of households interviewed over the number selected. DHS reports the net household response rate, which is the number of households interviewed over the number valid households found in the field (i.e. excluding vacant and destroyed dwellings). The practical aspects of $R_h$ and $R_i$ rates are discussed in Ayhan (1981) for the Turkish Fertility survey data. Ayhan (1981) has used the WFS (1975) recommendations that the first visit to the household (or individual) plus the number of re-calls constitute total calls. For a household survey, 1 + 3 = 4 total calls, and for individual survey, 1 + 2 = 3 total calls are proposed as threshold values.

For a required precision with a relative standard error *α*, the net sample size (number of completed interviews) needed for a simple random sampling (SRS) is given by:

$$n_{SRS}=\frac{(1/p-1)}{\alpha^2}.$$

Since a simple random sampling is not feasible for DHS, the sample size for a complex survey with clustering such as DHS can be calculated by inflating the above calculated sample size by using a design effect (*deff=deft*$^2$).

A simple random sample would be a random selection of individuals or households directly from the target population. This is not feasible for DHS surveys because a list of all eligible individuals or households is not available.

## 4.2. Multiple Indicator Cluster Surveys (MICS)

Another survey which is based on the complex sample design is the MICS (2006). Methodological manuals of the United Nations Children's Fund (UNICEF), Statistics and Monitoring Division, propose using the modified univariate sample selection formulae for their multivariable surveys.

The sample size calculating formula for MICS is given by

$$n_{MICS}=\frac{4r(1-r)deff(1.1)}{(0.12r)^2\,p.n_h}.$$

The formula in terms of our notation is given by

$$n_{MICS}=\frac{\chi^2_{(1),\alpha}\,p(1-p)deff(r)}{(0.12p)^2\,k.n_h}$$

where

$n_{MICS}$ is the required sample size, expressed as number of households

$\chi^2_{(1),\alpha}$ is a factor to achieve the (1−$\alpha$) per cent level of confidence

*p* is the predicted or anticipated prevalence (coverage rate) for the indicator being estimated

*r* is the factor necessary to raise the sample size for nonresponse. (*for example, for 10% nonresponse rate r should be 1.1.*)

*deff* is the design factor

*0.12p* is the margin of error to be tolerated at the 95 per cent level of confidence, defined as 12 per cent of *p* (*12 per cent thus represents the relative sampling error of p*)

*k* is the proportion of the total population upon which the indicator *p* is based, and

$n_h$ is the average household size.

For the Multiple Indicator Cluster Surveys (MICS), *UNICEF* proposes $r$ = 1.1 as an early compensation for the nonresponse amount. This will correspond to 10% increase in the sample size before the data collection, which intends to compensate the same amount of loss in the collected sample following the data collection. This approach cannot be accepted due to several bias producing aspects. Firstly, nonresponse rate is a part of survey error, which should not be included as the sample selection component. Secondly, 10% nonresponse rate can be a lower bound threshold value for this error statistics. For many surveys, the nonresponse rates are higher than this in the literature. Recently, there has been even a tendency of increase in nonresponse rates for the sample surveys of some developed countries.

For the MICS methodology, relative sampling error (value of *0.12p*) has been used for margin of error in the previous formulae because it scales the margin of error to result in comparable accuracy regardless of whether a high coverage indicator or low coverage indicator is chosen as the key one for sample size determination.

Recently, UNICEF, Statistics and Monitoring Section has decided not to clarify the sample selection formulae by removing the related methodology from their website. Instead, they provided a “sample size determination” template electronically. This template is naturally based on the previously discussed methodology for a univariate sample size determination algorithm, for a multivariate complex sample design.

## 5. Design Efficiency of Alternative Sample Sizes

This section clearly shows the partition of the components which are based on the modified sample size formulas of the two international institutions. The bias which will be created for the estimation by using the unrelated sample size formulae is given for the examined two large scale surveys.

The formula proposed by DHS is:

$$n_{DHS} = \frac{deff \dfrac{\left( 1/p - 1 \right)}{d^2}}{R_i R_h e}$$

where $n_{DHS}$ is the sample size in HH's offered by DHS.

The formula used in this study is

$$n_{DHS} = \frac{[(1-p)/p]}{d^2} \frac{deff}{R_i.R_h.e}.$$

The formula proposed by MICS is

$$n_{MICS} = \frac{\chi^2_{(1),\alpha} p(1-p) deff(r)}{(0.12p)^2 k.n_h}$$

where $n_{MICS}$ is the sample size in HH's offered by MICS.

The formula used in this study is

$$n_{MICS} = \frac{\chi^2_{(1),\alpha} [p(1-p)]}{d^2} \frac{deff(r)}{k.n_h}.$$

The relationship between the classical sample size formulization and DHS's sample size formulization can be given as

$$n_{DHS} = n_C \frac{1}{\chi^2_{(1),\alpha} p^2} \frac{deff}{R_i.R_h.e},$$

where $n_C$ is the classical sample size formula for the binary response.

The relationship between the classical sample size formulization and MICS's sample size formulization can be given as

$$n_{MICS} = n_C \frac{deff.(r)}{k.n_h}.$$

## 6. Issues on Selection Bias Representation

A comparison of the outcomes for the classical sample size determination methods and modified sample size determination methods provides information on the population representation and related biases. If we compare the results, in terms of overall sampling fractions, the following comparison can be used.

Overall sampling fraction of the *classical estimate*:

$$f_{SRS} = \frac{n_{SRS}}{N} = \frac{1}{F} = \frac{\chi^2_{(1),\alpha} [p(1-p)]}{d^2} \Big/ N.$$

Overall sampling fraction of the *modified estimate of DHS*:

$$f_{DHS} = \frac{n_{DHS}}{N} = n_C \frac{1}{\chi^2_{(1),\alpha} p^2} \frac{deff}{R_i.R_h.e} \bigg/ N = f_{SRS} \frac{1}{\chi^2_{(1),\alpha} p^2} \frac{deff}{R_i.R_h.e}.$$

Overall sampling fraction of the *modified estimate of MICS*:

$$f_{MICS} = \frac{n_{MICS}}{N} = n_C \frac{deff.(r)}{k.n_h} \bigg/ N = f_{SRS} \frac{deff.(r)}{k.n_h}.$$

Selection bias of the estimates for DHS:

$$B(f_{DHS}) = f_{DHS} - f_{SRS} = f_{SRS} \left( \frac{1}{\chi^2_{(1),\alpha} p^2} \frac{deff}{R_i.R_h.e} - 1 \right),$$

where bias will be *0* if and only if $\frac{1}{\chi^2_{(1),\alpha} p^2} \frac{deff}{R_i.R_h.e} = 1$.

Selection bias of the estimates for MICS:

$$B(f_{MICS}) = f_{MICS} - f_{SRS} = f_{SRS} \left( \frac{deff.(r)}{k.n_h} - 1 \right),$$

where bias will be *0* if and only if $\frac{deff.(r)}{k.n_h} = 1$.

Selection bias formulas show that the sample size calculation used by the surveys affect the overall sample selection probabilities. Accordingly, it is not recommended to use the modified univariate sample size determination formulas of this type for the complex sample designs.

Sample size determination for a two stage cluster sampling is proposed by Desu and Raghavarao (1990), and Aliaga and Ren (2006). Hansen *et al.* (1953) has evaluated the general cost function model for the complex sample designs. For the multivariable designs, there is no established standard computation formula for sample sizes. Depending on the type of design, the related parameters constitute variables for complex sample designs.

## 7. Weighting Adjustment Procedures

After the determination of the univariate sample size, the actual SRS sample is selected from the population by using a randomization process. For the purpose of the allocation of complex sample survey designs, the selected sample is then reallocated to the proposed new sample design. The proposed design can be allocated to complex designs, which may be based on *equal allocation*, *proportional allocation, probability proportional to size (PPS) allocation, weighted PPS allocation, optimum allocation,* and *clustering*. A comparison of the sample allocation methods is summarized by Ayhan and Islam (2005). Under this approach, the following adjustment methodologies can be used after the data have been collected from a complex sampling plan.

### 7.1. Weighting Independent Stages

Data weighting methods have been covered by Kish (1992), Kalton and Flores-Cervantes (2003), and Ayhan (2003), and Alkaya et al. (2017) in detail. Several alternative approaches, such as cell weighting, ranking, linear weighting, GREG weighting and several others can be proposed (Vaillant et al., 2013; Brick, 2013).

#### 7.1.1. Adjustments for Design Weights

For complex or stratified sample designs, design weights have to be used for the adjustment of the sample selection probabilities if the sample design is not self-weighting. For self-weighting sample designs selection probabilities of each domain will be the same as the overall, that is $f = \frac{n}{N} = \frac{n_i}{N_i} = f_i \quad \forall i$. Then, the design weights are given by

$$w_i = \frac{1}{p_i}\left[\frac{n}{\sum\left(1/p_i\right)}\right], \; i = 1,2,\cdots,n$$

where $p_i$ is the selection probability of the domain and $\sum w_i = n$.

#### 7.1.2. Adjustments for Non-Response Weight

Non-response weights should be used as an error correction component after the data collection, not before. Non-response adjustment weights are used to compensate for the losses of non-response amounts when the overall non-response rate is greater than 10 per cent and the domain non-response rates are more than 5 per cent for any domain (WFS, 1977). Sample design outcomes other than the above restrictions do not require any weighting adjustments for the sample outcomes. Hence, the non-response weights are for each domain given by

$$w_j = \overline{R}/R_j \; , \; j = 1,2,\cdots,$$

here $\overline{R} = \frac{\sum n_j}{\sum\left(n_j/R_j\right)}$, where $R_j$ is the non-response rate for the domain (or strata), $\overline{R}$ is the average non-response rate overall domains, and $n_j$ is the domain size. These rates ($R_j$ and $\overline{R}$) are recommended by WFS (1977) and has been used in 42 WFS country surveys, including Turkish Fertility Survey 1980 (Ayhan, 1981).

### 7.1.3. Adjustments for Post-Stratification Weights

Computation of the post-stratification weights is required for each domain in order to avoid bias due to cross-tabulation of the data. Kalton and Flores-Cervantes (2003) have proposed an alternative combined adjustment methodology for sample surveys.

This procedure adjusts the sample weights so that the sample totals conform to the population totals on a cell-by-cell basis. The weights for each respondent (typically, the inverse of the probability of the case) in a weighting cell (or post-stratum) is multiplied by an adjustment factor (Tourangeau et al., 2013). Then, the weight formula is given as

$$w_{2ij} = \frac{N_j}{\sum^{n_j} w_{1ij}} w_{1ij},$$

in which $w_{2ij}$ is the adjusted or post-stratified weight, $w_{1ij}$ is the unadjusted weight, and the adjustment factor is the ratio between the population total for cell $j$ ($N_j$) and the sum of the unadjusted weights for the respondents in that cell.

Rather than using independent weighting and adjustment procedures for each stage of the weighting, alternative approaches can also be used. This is based on combined weighting methods, which take into account the conditional probability approach for the previous stages. As an alternative to the weighting independent stages, the combined weighting methods can be proposed to avoid bias for the sample estimates.

## 7.2. Combined Weighting Methods

Ayhan (2003) and Alkaya et al. (2017) have proposed the following combined weighting procedure for sample surveys. These weighting procedures are used in a sequential manner for each weighting component. The weights are proposed as products for each weighting stage in a combined way. Sample design may be self-weighting or non-self-weighting. Design weights have to be introduced for non-self-weighting designs in the following way.

The probability of selection of the overall sample is obtained simply by the sampling fraction of the selected sample $f = x/X = 1/F$ for the total sample. On the other hand, after using some method of stratification, the sampling fraction of any strata is $f_i = x_i / X_i = 1 / F_i$.

### 7.2.1. Design Weights

Design weights (Ayhan, 1991; Verma, 1991) for non-self-weighting sample designs can be computed for each domain $i$ with the same probability of selection $p_i$ (Ayhan, 2003).

For *combined ratio mean* $\theta = Y/X = \sum_i^H Y_i \Big/ \sum_i^H X_i$ where *i = 1 ,2*,.., *H*, here *H* represents the number of domains, estimated by $\hat{\theta} = y/x = \sum_i^H y_i \Big/ \sum_i^H x_i$ . On the other hand, for a *separate ratio mean* $\theta_w = \sum_i^H W_i\ \theta_i$ , estimated by

$$\hat{\theta}_w = \sum_i^H W_i\ \hat{\theta}_i = \sum_i^H W_i\ [y_i/x_i] \ ,$$

$$W_i = \left[ \sum_{i=1}^H x_i \Big/ \sum_{i=1}^H \{ x_i/[(X/x)\, p_i ]\} \right] \Big/ [(X/x)\, p_i ] = P_0/P_i$$

where $\sum_{i=1}^H (W_i\ x_i) = x$ .

Here, $P_0$ has been computed to adjust the overall weighted and unweighted sample to be the same.

### 7.2.2. Combined Weighting for Nonresponse

In addition, a weighting procedure for nonresponse is also essential for self-weighting and non-self-weighting sample design outcomes. The non-response rate is calculated as

$$W_i^* = R_0/R_i$$

where $R_i = x_i^*/x_i$ is the response rate in domain *i*.

The overall response rate $(R_0)$ for the design can be computed as

$$R_0 = \sum_{i=1}^I (W_i\ x_i) \Big/ \sum_{i=1}^I (W_i\ x_i\ /\ R_i),$$

where $R_0$ is used to adjust the sample sizes to be the same, $\sum_{i=1}^I \left( W_i\ \ W_i^*\ x_i \right) = x$ .

### 7.2.3. Combined Weighting for Post-Stratification

Finally, post-stratification of a complex sampling scheme requires additional weighting procedures for independent subclasses. The combined weight can be calculated by using the following weights:

$$W_i' = \left( W_i\ \ W_i^*\ X_i \right) / X \ ,$$

$$W_i = \left[ \sum_{i=1}^H x_i \Big/ \sum_{i=1}^H \{ x_i/[(X/x)\, p_i ]\} \right] \Big/ [(X/x)\, p_i ] \ \text{ and}$$

$$W_i^* = R_0 / R_i \quad ,$$

where $W_i'$ is the post-stratification weight, $W_i$ is the design weight, $W_i^*$ is the non-response weight, $R_0 = \sum_{i=1}^{I}(W_i \; x_i) \Big/ \sum_{i=1}^{I}(W_i \; x_i \, / \, R_i)$ and $R_i = x_i^* / x_i$ .

Consequently, $\sum_{i=1}^{I} \left( W_i \; W_i^* \; W_i' \, x_i \right) = x$ is the overall sample adjustment procedure for the combined weighted estimator. This is naturally provides the adjustment to the base variable *x* (Ayhan, 2003; Alkaya et al., 2017).

Alternative weighting adjustment procedures in multistage complex sample surveys are proposed by Ayhan et al. (2000) for adjusting the original selection probabilities by PPS procedures.

The next step in the analysis of the collected data is to compute the following error statistics for the proposed sampling design. This will provide information on how efficient the designed sample was when compared with the basic standard, which is SRS.

## 8. Discussion and Conclusions

In a multivariable survey design, the determination of the sample size is an important concept that has to be answered. Although there is no settled methodology, some prestigious organizations modify the formula for univariate sample size determination to be able to use it in the multivariable case. For this purpose, they included some factors such as *deff* or non-response rate in their sample size formulas. These factors have to be calculated after the sample has been collected as a data quality measure, not before. Hence, the modified univariate sample size methodologies of several survey institutions do not represent the corresponding population. The amount of bias involved in the formulas is clearly identified during the previous formulations. This paper highlights the importance of sample selection in a representative manner, to avoid the selection bias.

The ideal approach should be <u>not to determine</u> the sample size of the complex survey design as if it was based on the univariate case and use SRS assumptions. Consequently, representation bias enables the survey results not representing the corresponding population parameters.

For the complex designs, the suggested alternative strategy is to use weighting after SRS. For the purpose of allocation, the selected sample is reallocated to the proposed new sample design. As an alternative to the weighting independent stages, the combined weighting methods can be proposed to avoid bias for the sample estimates.

International survey organizations <u>should be responsible for following</u> recent developments in survey sampling theory and methods, in order to maintain themselves as reliable institutions.

## REFERENCES

ADCOCK, C. J., (1997). Sample Size Determination: A Review. The Statistician, 46 (2), pp. 261–283.

ALIAGA, A., REN, R., (2006). Optimal Sample Sizes for Two-Stage Cluster Sampling in Demographic and Health Surveys. DHS Working Papers, No. 30, Demographic and Health Research, ORC Macro.

ALKAYA, A., AYHAN, H. Ö., ESIN, A., (2017). Sequential Data Weighting Procedures for Combined Ratio Estimators in Complex Sample Surveys, Statistics in Transition, 18 (2), pp. 139–162.

AYHAN, H. Ö., (1981). Sources of Nonresponse and Nonresponse Bias in 1978 Turkish Fertility Survey. Turkish Journal of Population Studies, 2-3, pp. 104–148.

AYHAN, H. Ö., (1991). Post-stratification and weighting in sample surveys. Research Symposium '91, State Institute of Statistics, Ankara. pp. 1–11.

AYHAN, H. Ö., HANCIOGLU, A., TÜRKYILMAZ A. S., ÜNALAN, T., (2000). Sample Design, Implementation and Outcome. Chapter 13 in Push and Pull Factors of International Migration: Country Report – Turkey (H. Ö. Ayhan, et. al., eds). EUROSTAT Working Papers; Population and Social Conditions 3/2000/E/n$^0$ 8, Luxembourg: EUROSTAT Press, pp. 112–131.

AYHAN, H. Ö., (2003). Combined Weighting Procedures for Post-Survey Adjustment in Complex Sample Surveys. Bulletin of the International Statistical Institute 60, pp. 53–54.

AYHAN, H. Ö., ISLAM, M. Q., (2005). Sample Design and Allocation for Random Digit Dialling, Quality and Quantity, 39 (5), pp. 625 – 641.

BRICK, J. M., (2013). Unit Nonresponse and Weighting Adjustments: A Critical Review, Journal of Official Statistics, 29 (3), pp. 329–353.

DELL R. B., HALLERAN, S., RAMAKRISNAN, R., (2002). Sample Size Determination, ILAR J, 43 (4), pp. 207–213.

DESU, M. M., RAGHAVARAO, D., (1990). Sample Size Methodology, Boston: Academic Press.

DHS, (2012). Demographic and Health Surveys Methodology, Sampling and Household Listing Manual. ICF International, Calverton, Maryland USA, p. 109.

HANSEN, M. H., HURWITZ, W. N., MADOW, W. G., (1953). Sample Survey Methods and Theory, New York: Wiley.

HÜNEE, (2014). 2013 Turkey Population and Health Survey, Hacettepe University, Institute of Population Studies, Ankara, Turkey.

HENRY, K. A., VALLIANT, R., (2015). A Design Effect Measure for Calibration Weighting in Single-Stage Samples, Survey Methodology, 41 (2), pp. 315–331.

MARSHALL, B., CARDON, P., PODDAR A., FONTENOT, R., (2013). Does Sample Size Matter in Qualitative Research? A Review of Qualitative Interviews in is Research, Journal of Computer Information Systems, Vol. 54 (1), pp. 11–22.

MICS, (2006). Multiple Indicator Cluster Survey Manual 2005, Division of Policy and Planning, United Nations Children's Fund (UNICEF), New York, USA, p. 595.

KALTON, G., FLORES–CERVANTES, I., (2003). Weighting Methods, Journal of Official Statistics, 19 (2), pp. 81–97.

KISH, L., (1965). Survey Sampling. New York: John Wiley and Sons.

KISH, L., (1982). Design Effect. Encyclopedia of the Statistical Sciences, John Wiley and Sons, Inc. Vol. 2, pp. 347–348.

KISH, L., (1992). Weighting for Unequal $P_i$'s, Journal of Official Statistics, 8 (2), pp. 183–200.

PLACZEK, M., FRIEDE, T., (2017). Clinical trials with nested subgroups: Analysis, sample size determination and internal pilot studies, Statistical Methods in Medical Research 0 (0), pp. 1–18, DOI: 10.1177/0962280217696116.

RAO, J. N. K., SCOTT, A. J., (1984), On Chi-square Tests for Multiway Contingency Tables with Cell Proportions Estimated from Survey Data, Annals of Statistics 12, pp. 46–60.

RYAN, T. P., (2013). Sample Size Determination and Power, John Wiley and Sons Inc., Hoboken, New Jersey.

SAFO, S., SONG, X., DOBBIN, K. K., (2015). Sample Size Determination for Training Cancer Classifiers from Microarray and RNA-Seq Data, Ann. Appl. Stat. 9 (2), 1053—1075, DOI:10.1214/15-AOAS825, http://projecteuclid.org/euclid.aoas/1437397123.

SATHIAN, B., SREEDHARAN, J., BABOO, S., SHARAN, K., ABHILASH, E., RAJESH, E. (2010). Relevance of Sample Size Determination in Medical Research, Nepal Journal of Epidemiology, 1 (1), pp. 4–10, DOI: http://dx.doi.org/10.3126/nje.v1i1.4100.

SHORE, H., (2008). Sample-Size Determination. Encyclopedia of Statistics in Quality and Reliability, IV, John Wiley & Sons, Ltd.

SIDDIQUI, K. A., (2013). Heuristics for Sample Size Determination in Multivariate Statistical Techniques (December 6, 2013), World Applied Sciences Journal 27 (2), pp. 285–287, Available at SSRN: https://ssrn.com/abstract=2447286.

SIRKEN, M. G., (2002). Design Effects of Sampling Frames in Establishment Survey, Survey Methodology, 28 (2), pp. 183–190.

TOURANGEAU, R., CONRAD, F. G., COUPER, M. P., (2013). The Science of Web Surveys, Oxford University Press, New York.

VALLIANT, R., DEVER, J. A., KREUTER, F., (2013). Practical Tools for Designing and Weighting Survey Samples, Statistics for Social and Behavioral Sciences. Springer.

VERMA, V., (1991). Sampling Methods, Training Handbook, Statistical Institute for Asia and the Pacific, Tokyo.

WFS, (1975). WFS Basic Documentation, No. 5, Supervisor's Instruction, World Fertility Survey, London, U.K.

WFS, (1977). Guidelines for Country Report, No.1, WFS Basic Document, No. 8, World Fertility Survey, London, U.K., 186.