Variance Estimation of the Design Effect

Alberto Padilla
Banco de México

Abstract
Sample size determination is a crucial part of the planning process of a survey and it can be accomplished in different ways, some of them require information not available or that may be obtained with a substantial cost. Sample size calculation can be done by using the design effect estimator proposed by Kish. This estimator is also used as an efficiency measure for a probability sampling plan and to build confidence intervals. Even though the design effect estimator is widely used in practice, little is known about its statistical properties and there are no variance estimators available. In this paper we propose a method to estimate the variance. With this estimator it is possible to assess the precision of the estimators during the planning stage of a survey. An example using stratified sampling is given.

Key Words: Ratio estimator, simple random sampling, design effect, sample size; resampling method

1. Introduction

The design effect, $deff_K$, Kish (1965), is defined as the ratio of the variance of an estimator under a specific design to the variance of the estimator under simple random sampling without replacement, *srswor*. The estimator of the design effect is used for example in the computation of the sample size for complex sample designs and to build confidence intervals. To determine the sample size one must have an estimation of the $deff_K$ for the statistic of interest and a computation of the sample size under *srswor*. With these two elements at hand, setting aside adjustments for nonresponse, the computation of the sample size for a complex designs reduces to the product of the $deff_K$ and the sample size under *srswor*. Confidence intervals can be built by multiplying the standard deviation of a statistic, like a mean or total, computed under *srswor* with the square root of the estimated $deff_K$. From these examples, it can be seen that the $deff_K$ helps us to compare the efficiency of design plans with *srswor*. A plan would be more, equal or less efficient than *srswor* if the estimated $deff_K$ is less, equal or greater than one.

It is worth mentioning that in the *deff* estimator, the estimation of the variance under *srswor* is computed with the sample obtained from the complex design without considering the stratification, clustering, unequal probability sampling, etc. This method does not guarantee an unbiased estimation of the population variance under *srswor*. This problem has been analyzed by Rao (1962) who built unbiased estimators for the variance under *srswor* for three complex designs; Cochran (1977) illustrated Rao´s method for a stratified population and recently Gambino (2009) built an unbiased estimator in terms of the Horvitz-Thompson estimator (1952) of the population characteristic. The estimator of the $deff_K$ is a widely used quantity but little is known about its properties and to the author´s best knowledge there are no variance estimators available. We propose a resampling method to estimate the variance of the design effect, namely Sitter's bootstrap for complex designs, see Sitter (1992). The results obtained are promising and it seems to be an option to estimate the variance of the design effect estimator proposed by Gambino (2009).

The article is organized as follows. In section 2 we introduce definitions and notation used throughout the paper. An example of sample size determination is shown in section 3. Section 4 deals with the proposed resampling method for estimating the variance of the design effect and an example using stratified random sampling is given.

2. Notation and Definitions

Let U denote a finite population of N elements labeled as $k=1,\ldots,N$, $1<N$. It is customary to represent U with its labels k as $U=\{1,2,\ldots,k,\ldots,N\}$. The example used in this paper refers to stratified random sampling, hereinafter *strs*, so it is convenient to introduce some definitions and notation for this design.

Stratified random sampling: the variable of interest will be represented by y_{hi}, where i stands for the ith element of the population in hth stratum, with $i \in \{1,2,\cdots,N_h\}$. N_h and n_h denote the total number of elements in the population and the sample size in the hth stratum, $N=\sum_{h=1}^{H} N_h$ and $n=\sum_{h=1}^{H} n_h$, where H is the total number of strata in the population. The population mean will be denoted by $y_{st}=\sum_{h=1}^{H} W_h y_h$, where $W_h = N_h / N$ and $y_h = \sum_{i=1}^{N_h} y_{hi} / N_h$. The unbiased estimator of the population mean is computed as $\hat{y}_{st}=\sum_{h=1}^{H} W_h \hat{y}_h$, where $\hat{y}_h = \sum_{i=1}^{n_h} y_{hi} / n_h$. The population variance within strata shall be denoted by s_{hU}^2, its simple estimator as $\hat{s}_h^2$ and the variance of the stratified mean estimator, using *srswor* within strata, will be represented by $v_{strs}=\sum_{h=1}^{H} W_h^2 (1-n_h/N_h) s_{hU}^2 / n_h$. The formula for the variance estimator of this quantity is $\hat{v}_{mae}$ with $\hat{s}_h^2$ in place of s_{hU}^2 in the formula for v_{srswor}.

3. Example of sample size determination using deff

As it was mentioned above, the design effect, efd_K, Kish (1965), is defined as the ratio of the variance of an estimator under a specific design different from simple random sampling, $v_{alt}(\hat{\theta}_{alt})$, to the variance of the estimator under simple random sampling without replacement, $v_{srswor}(\hat{\theta}_{srswor})$. In this way, the design effect is computed with the following formula $deff_K(\hat{\theta}) = v_{alt}(\hat{\theta}_{alt}) / v_{srswor}(\hat{\theta}_{srswor})$, with $v_{srswor}(\hat{\theta}_{srswor}) > 0$. During the planning stage of a survey or a sampling plan, design effects are widely used to compute sample sizes. In this section, an example of sample size determination from an actual consumer survey will be presented.

Example 1: sample size computation using deff. In México, INEGI, the official statistics agency determines the sample size for different surveys using the design effect. In particular, for the 2008 Household Income and Consumption Survey, INEGI used as the target variable to compute the sample size, the household mean current income. They employed the following expression, for the adjusted sample size, *n*:

In this equation:

$$n = \frac{z_{\alpha}^{2} s^{2} \text{deff}}{r^{2} \overline{X}^{2} (1 - \text{tnr}) \text{PHV}} \tag{1}$$

- z_α = the αth quantile of a standard normal distribution, INEGI used a two-sided 90% confidence level.
- s^2 = estimation of the population variance between elements = 1’767,586,178.
- $\overline{X}$ = estimation of the mean for the household mean current income = 34,127 mexican pesos.
- deff = design effect estimator.
- r = maximum relative aceptable error = 4%.
- tnr = maximum nonresponse rate = 15%.
- PHV = mean number of households per dwelling = 1.02.

With all these values and using formula (1), the adjusted sample size was 9,711 dwellings, which was rounded to 10,000. The design effect, the estimator of the population variance within elements and the mean for the household mean current income were obtained from the 2006 Household Income and Consumption Survey. It was required that the estimations of the variables of interest hold for some states, so the final sample size was 35,146 dwellings.

Formula (1) can be expressed as follows:

$$n = \frac{z_{\alpha}^{2} s^{2} \text{DEFF}}{r^{2} \overline{X}^{2} (1 - \text{tnr}) \text{PHV}} = \frac{z_{\alpha}^{2} s^{2}}{r^{2} \overline{X}^{2}} \text{DEFF} \frac{1}{(1 - \text{tnr}) \text{PHV}} = n_{srswor} \text{DEFF} \frac{1}{(1 - \text{tnr}) \text{PHV}} \tag{2}$$

In this equation, n_{srswor} stands for the sample size obtained by *srswor*, which is easy to compute, see Cochran (1977). From the right side in equation (2) it can be easily seen that the design effect increases, decreases or left unaffected the sample size obtained by *srswor*.

4. Variance estimation of the design effect

As it was mentioned in the introduction, to the author’s best knowledge there is no variance estimator of the design effect. In survey sampling, the standard approach to obtain a variance estimator of an estimator like a mean, total or the design effect is as follows. First, one has to build an expression for the population variance of the estimator, in this case the design effect, then one has to find an estimator of the population variance. In the literature there are no expressions for the variance and variance estimators for the design effect, but it’s posible to estimate the variance using a resampling method. In Chaudhuri & Stenger (2005) one can find eight type of bootstraps for samples extracted from finite populations. Some of these methods were proposed by Sitter (1992) and one of them is the bootstrap used in this article.

Sitter (1992) proposed bootstrap estimators for three sample designs:
a) Stratified random sampling with *srswor* selection of elements within strata.
b) Two stage cluster sampling with equal or unequal sizes.
c) The Rao-Hartley-Cochran method for probability proportional to size sampling, see Rao et al. (1962).

In this article, Sitter proposed three methods to build confidence intervals. One of them is the percentile method, used in this paper due to its simplicity. Another method is a double bootstrap, a computer intensive method in which the bootstrap is applied two times to each sample. The third method, used by Sitter in its article, is a jackknife estimator of the variance applied to the sample and to the replicated subsample. The last two methods require a special study in order to compare them with the percentile method and to explore some aspects related to the required number of bootstrap samples that is closed to the nominal confidence for the design effect estimations.

In this work we use the method proposed by McCarthy & Snowden, see Chaudhuri & Stenger (2005), which is a special case of the extended bootstrap of Sitter (1992) for stratified random samples. This method is described next.

Without considering the fractional part of $n_h' = n_h - (1 - f_h)$ and $k_h = \frac{N_h}{n_h}(1 - \frac{1 - f_h}{n_h})$ with $f_h = n_h / N_h$, the method is applied as follows:

a) Replicate $(y_{h1}, \ldots, y_{hn_h})$ k_h times in a separate and independent way, *h=1,…,H*, to create *H* different pseudo-strata.
b) Extract a *srswors* of size n_h' from the *hth* pseudo-stratum and repeat independently this procedure for every *h=1,…,H*, creating in this way the bootstrap observations

$$s* = \{(y^*_{h1}, \ldots, y^*_{hn_h'}),\ h = 1, \ldots, H\}$$

and let $\hat{\theta}^* = \hat{\theta}(s^*.)$
c) Repeat stage (b) a large number of times, say *B*, and compute for every *bth* bootstrap sample, $\hat{\theta}^*_b,\ b = 1, \ldots, B$. With the B estimators at hand $\hat{\theta}^*_b$, compute the following quantities:

$$\hat{\theta}^*_B = \frac{1}{B}\sum_{b=1}^{B} \hat{\theta}^*_b \quad \text{y}$$
$$\hat{\text{v}}_{\text{BWO}} = \frac{1}{B-1}\sum_{b=1}^{B} (\hat{\theta}^*_b - \hat{\theta}^*_B)^2 \tag{3}$$

These two expressions give us the bootstrap estimators of a mean or total and its variance. The variance estimator *BWO* can also be used as a variance estimator for the estimator of the original sample. *BWO* stands for bootstrap in the case of sampling without replacement.

In this article, $\hat{\theta}^*_b$ may refer to a stratified estimator of the mean, a ratio estimator, like the design effect estimator, o a variance estimator, like the variance under *strs* or *srswor*.

We use Sitter's extended bootstrap, due to its ease of implementation, compared to the methods described in Chaudhuri & Stenger (2005). Nonetheless, we are not claiming that it has better performance than other bootstrap methods for estimating the variance of the design effect.

There is no closed expression available for the variance of the estimator of the design effect, $deff_G$, so we will illustrate how to compute the bootstrap estimator of the variance in a small stratified population.

<u>Example 2</u>: *stratified random sampling*. Based on Cochran's (1977) example, page 137, we simulated a small population with *120* elements and *5* strata. Tables 1 and table 2 below contain summary values of the simulated population and the population variances under *strs*, *srswor* and the design effect.

Table 1: Simulated population values

Stratum	N_h	n_h	W_h	y_h	s_{hU}^2
1	13	9	0.11	2.33	1.62
2	18	7	0.15	1.61	0.08
3	26	6	0.22	5.04	1.18
4	26	10	0.22	7.01	3.06
5	37	8	0.31	9.86	0.31
Population	120				3.44

Table 2: Population variances for *strs* and *srswor* as well as the design effect

Population quantity	Value
v_{srswor}	0.1762
v_{strs}	0.0196
$deff_K$	0.1114

In this example the following steps were applied to simulate Sitter's bootstrap:

a) We drew *5,000* samples, *strs*, of size *40* from the stratified population.
b) For every *strs*, we simulate *B=2,000* bootstrap samples with the above mentioned Sitter's method. This value was used upon recommendation in Stuart et al. (1999) for variance estimation in the case of independent random variables.
c) The design effect estimator was computed using $deff_G$, see Gambino (2009), as well as with the bootstrap. For every *strs*, the variance estimator of the design effect was obtained from $\hat{v}_{BWO}$, Chaudhuri & Stenger (2005) or Sitter (1992).

To the author's best knowledge, Sitter's bootstrap methods are not available in statistical softwares, so we wrote programs in *R*, *R* Development Core Team (2010), to extract the samples and apply the bootstrap. The 5,000 *strs* samples were drew in *R* using library *pps*, Gambino (2005) and the 95% two-sided intervals for the population design effect were obtained from the bootstrap histogram built with the estimated design effects. With this method and for every *strs*, the 2.5% and 97.5% percentiles of every bootstrap histogram are found, so one can determine whether or not the population design effect is contained inside the interval. This was done for the 5,000 *strs* samples, counting the number of times the interval contained the population design effect and dividing by 5,000. In this way, the coverage for the $deff_G$ estimators were obtained. The results of the simulation are found in the next table.

Table 3: Results from the 5,000 *strs* samples to estimate the variance of the design effect, 2,000 bootstrap simulations were generated for every *strs*

Estimator	Average of estimators (A)	Population Value (B)	Difference (%) = (A-B)/B
Population mean *strs BWO*	6.144	6.143	0.02
V*strs BWO*	0.018	0.0196	-7.60
V*srswor BWO*	0.176	0.176	-0.10
deff *BWO*	0.103	0.111	-7.80
deff_G *strs*	0.112	0.111	0.30
Std deviation deff BWO	0.021		
Std deviation deff_G *strs*	0.018		
Coverage deff_G *strs*=	90%		

In table 3, the estimators followed by *BWO* were obtained from the bootstrap simulations. In the above table, the bootstrap estimators of the population mean, population mean *strs BWO*, and the variance under *srswor*, V*srswor BWO*, had a relative difference smaller than 1% compared to the correspondent population value. On the other hand, the bootstrap estimators of the population value under *strs*, V*strs BWO*, and the population design effect efd_K, efd *BWO*, subestimate by 8% the correspondent population value. The deff estimator, $deff_G$, had a small bias, the relative difference to the design effect of the population, efd_K, was 0.3%.

It can be seen from table 3 that the square root of the bootstrap estimator of the variance of the design effect has a value of 0.021 which can be expressed as a coefficient of variation of 18.9%. There is a certain amount of variation around the population design effect, but it is not so bad considering the small population and sample size. In order to compare the result obtained from the variance estimator of the design effect generated by the bootstrap, we also compute the variance between the 5,000 $deff_G$ estimators. The square root of this estimator is 0.018 compared to 0.021 obtained with the bootstrap. The coverage, end of table 3, for the design effect obtained with the bootstrap turned out to be approximately 90%, which is below from the nominal of 95%. Sitter (1992) mentioned that the coverage can be improved using a method different from the percentile; nonetheless, we found a better coverage by increasing the number of bootstrap samples for every *strs*.

The first simulations were run with values of *B* similar to those recommended by Sitter in its article, Sitter used *B=300*, but Stuart et al. (1999), mentioned that at least 2,000 bootstrap simulations in the case of variance estimation for random samples are required. With this number of bootstrap simulations, the coverage improved substantially compare to the number of simulations proposed by Sitter. We did not find an article or a reference for the number of simulations recommended for the bootstrap in case of complex simple designs.

Below is a histogram for the 5,000 *strs* samples of the estimator $deff_G$.

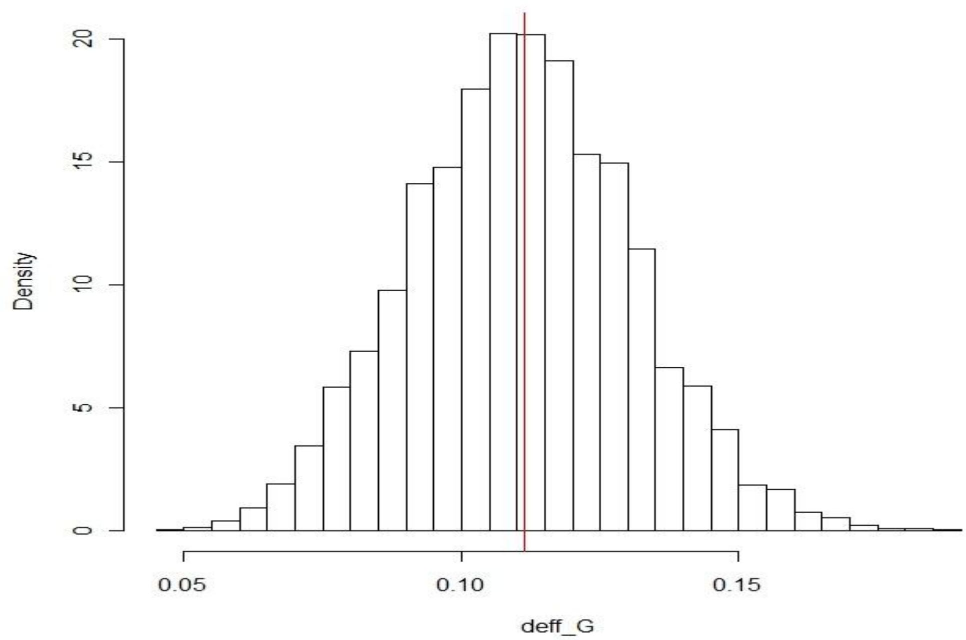

Figure 1: 5,000 estimators of $deff_G$

In this histogram, the red line corresponds to the population value of the design effect which is 0.1114. It seems to be a slight skewness; unfortunately, to the author's best knowledge, the distribution of the ratio of variance estimates like $deff_G$ is not known for finite population sampling. This was a simulation exercise, but in practice we only have one sample and the B bootstrap subsamples from the original *strs*. With the *strs* we compute the estimator of the design effect, $deff_G$, and with the B bootstrap subsamples we generate a histogram in order to obtain lower and upper limits from it with the percentile method. These values give us the interval for the design effect. In figure 2 there is a histogram of the design effect estimators built with the $B=2,000$ bootstrap subsamples obtained from the last *strs* sample of size $n=40$. This sample was selected just to illustrate a histogram one would build in practice. In this sample the estimator of the design effect is $deff_G=0.116$ and the square root of the variance between the bootstrap estimators of the design effect is equal to 0.022.

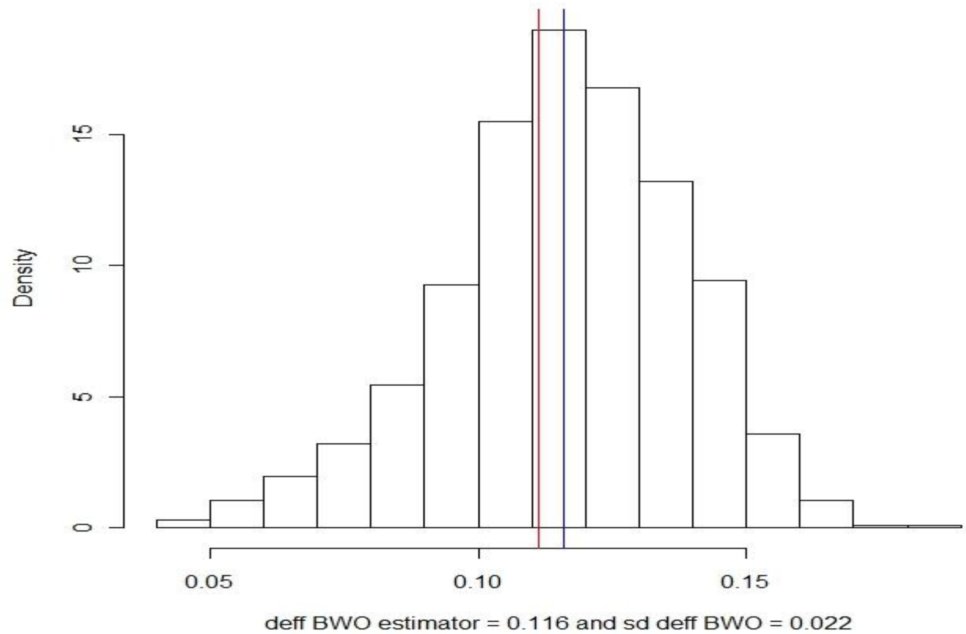

Figure 2: 2,000 bootstrap replicates of one *strs* sample

The red line corresponds to the population value of the design effect, 0.1114, and the blue one to the *deff BWO* estimator based on *2,000* bootstrap replicates. In this case, the bootstrap estimated variance was 0.000492 or a standard deviation of 0.022. It is worth noting that we have not investigated the causes of the asymmetry in figure 2 and it's a topic for future research.

5. Conclusions

We proposed to use the bootstrap for finite population samples, one of Sitter's methods (1992), to obtain a variance estimation of the design effect. The method was illustrated in a small stratified population. The results from the simulations suggest the feasibility to estimate the variance of the design effect from the bootstrap histogram. We used the estimator of the design effect proposed by Gambino (2009), which provides an unbiased estimator of the variance of *srswor*. It is necessary to improve the coverage of the bootstrap histogram. This can be done with Sitter's (1992) recommendations of variants of the bootstrap and it is a work for future research. At the time this work was done, Sitter's bootstraps were not available in *R* or in the *R survey* library, Lumley (2010), so we had to program it in R.

References

Chaudhuri, A. & Stenger, H. (2005) *Survey Sampling: theory and methods*, 2nd edn., Chapman & Hall/CRC.
Cochran, W.G. (1977) *Sampling Techniques*, 3rd edn. New York: Wiley.
Gambino, J.G. (2009) *Design effects caveat*, The American Statistician, pp. 141-145.

Gambino, J.G. (2005) *pps: Functions for PPS sampling*. R package version 0.94.
Horvitz, D.G. & Thompson, D. J. (1952) *A generalization of sampling without replacement from a finite universe*, Journal of the American Statistical Association 47, pp. 663-685.
INEGI, *Encuesta Nacional de Ingresos y Gastos de los Hogares 2008*. *Diseño Muestral*.
Kish, L. (1965) *Survey Sampling*, New York: Wiley & Sons.
Lumley, T. (2010) S*urvey: analysis of complex survey samples*. R package version 3.23-3.
Padilla, A.M., *Una cota para el sesgo relativo del efecto del diseño*, Memorias electrónicas en extenso de la 4ª Semana Internacional de la Estadística y la Probabilidad. Julio 2011, CD ISBN: 978-607-487-324-5.
Padilla, A.M., *A bound for the relative bias of the design effect*. ICES IV, Fourth International Conference on Establishment Surveys, June 11-14, 2012, Montréal, Québec, Canada.
Padilla Terán, A.M., *El efecto del diseño: sesgo y estimación de varianza*. Documentos de Investigación 2012-18, Banco de México.
R Development Core Team (2010). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.
Rao, J.N.K. (1962) *On the estimation of the relative efficiency of sampling procedures*, *Annals* of the *Institute* of *Statistical* Mathematics, pp. 143-150.
Rao, J.N.K., Hartley, H.O. & Cochran, W.G. (1962) *On a simple procedure of unequal probability sampling without replacement*, Journal of the Royal *Statistical* Society B 24, pp. 482-491.
Särndal, C.E., Swensson, B. & Wretman, J.H. (1992) *Model Assisted Survey Sampling*, Springer-Verlag, New York, 1992.
Sitter, R.R., (1992) *A resampling procedure for complex survey data*, Journal of the American Statistical Association, Vol. 87, pp. 755-765.
Stuart, A., Ord, K. & Arnold (1999) S. Kendall’s Advanced Theory of Statistics (sixth edn). Volume 2ª, Classical Inference and the Linear Model. Edward Arnold, London.
Thompson, M.E. (1997). *Theory of Sample Surveys*. Chapman & Hall, London.