METHODS FOR COMPARISON OF DESIGN EFFECT COMPONENTS ACROSS SURVEYS

Ibrahim S. Yansaneh, U.N. Statistics Division and John L. Eltinge, U.S. Bureau of Labor Statistics
I.S. Yansaneh, U.N. Statistics Division, DC2-1526, Two U.N. Plaza, New York, NY 10017; yansaneh@un.org

Key Words: Additive model; Average design effect; Demographic and Health Surveys (DHS); Generalized variance function; Interaction.

1. Introduction

1.1 Sample Design for Developing Country Surveys

International aid agencies often conduct surveys of social, economic and health conditions in developing countries. These surveys frequently use very similar sample designs and questionnaires across countries. For example, the Demographic and Health Surveys (DHS) series, sponsored by the United States Agency for International Development and conducted by ORC Macro, has been carried out in more than 50 countries beginning in 1984. The DHS series uses a common core questionnaire focused on fertility, fertility intentions and family planning; child health and mortality, nutritional status of women and children, access to services, etc. For some countries, optional modules are added to the core questionnaire. For some general background on the DHS, see Verma and Lê (1996), Lê and Verma (1997), and references cited therein.

The DHS series generally uses a stratified cluster sample design, in which strata are defined by the intersection of administrative region and urban/rural classification, usually with a separate stratum for the national capital region. Within strata, geographic area units are the primary sample units (PSUs), households are the secondary sample units, and individual women of reproductive age (normally 15-49 years) are the sample elements. The PSUs for the DHS are generally census enumeration areas, which are selected with probability proportional to size. The sizes of the PSUs range from about 100 to 300 households.

However, the DHS sample designs are not identical across countries, and may vary with respect to, e.g., the number of design strata, sampling rates within strata, sample sizes within primary sample units, and the number of primary sample units selected within strata. In addition, the underlying populations may vary with respect to their prevalence rates for specified health or social conditions, the degree of heterogeneity within and across strata or clusters, and the distribution of specific subpopulations within and across strata or clusters.

1.2 Definition, Interpretation and Exploratory Analysis of Design Effects

Because the DHS sample designs share some common features across countries, it is important to consider the extent to which information collected in previous DHS surveys may offer practical insight into ways to improve sample design efficiency for future DHS surveys. The present paper will focus on a measure of efficiency known as a univariate *design effect*, defined as:

$$\Delta = V_C(\hat{\theta}_C) / V_{SRS}(\hat{\theta}_{SRS}) \qquad (1.1)$$

where θ is a univariate parameter of interest; $\hat{\theta}_C$ is an approximately unbiased estimator of θ to be computed from a sample of size n obtained through a specific complex sample design C; $V_C(\hat{\theta}_C)$ is the variance of $\hat{\theta}_C$ evaluated with respect to the complex design; $\hat{\theta}_{SRS}$ is a standard approximately unbiased estimator of θ based on a with-replacement simple random sample, also of size n, selected from the same finite population; and $V_{SRS}(\hat{\theta}_{SRS})$ is the variance of $\hat{\theta}_{SRS}$ evaluated with respect to the same simple random sampling design.

In sample design work, design effects Δ are of practical interest for several reasons. First, for a given estimand θ, the estimator $\hat{\theta}_C$ based on data from the complex sample of size n has the same variance as the estimator $\hat{\theta}_{SRS}$ computed from data from a simple random sample of size n/Δ. For this reason, the ratio n/Δ is sometimes called the "effective sample size" for estimation of θ using data from the complex design C. For a general discussion of "effective sample size" calculations, see Kish (1965, 1995), Cochran (1977), Potthoff et al. (1992) and references cited therein.

Second, design effects provide an indication of specific classes of estimands for which a given complex design C may be especially efficient or inefficient. For a given class of estimands, severe problems with inefficiency may suggest a major reconsideration of the sample design, or omission of those estimands from the goals of the survey.

Third, for a given parameter θ, the design effect (1.1) will depend on several features of the specific underlying population, complex design C, and point estimator $\hat{\theta}_C$ under consideration. These features include the finite population means and variances within each stratum, the intracluster correlations encountered within each stage of clustering, the heterogeneity of selection probabilities, nonresponse effects, and deviations from standard probability weighting in the construction of a point estimator; see, e.g., Kish (1995), Gabler et al. (1999) and references cited therein. Consequently, to the extent that DHS surveys in several countries may have similar design features, one might expect those surveys also to have similar design effects.

The remainder of this paper will explore the design effects for the DHS. Section 2 will use simple regression methods to explore the extent to which variability in design effects may be associated with variability in underlying design or population characteristics. Section 3 applies the proposed methods to the comparison of DHS design effects for Ghana and Bangladesh.

2. Regression Approximations for Design Effects

Anecdotal discussion of design effects often centers on "average" design effects within classes of surveys and estimands, e.g., design effects for the estimated prevalence rates for relatively common items, based on data from household surveys with heavily clustered designs. Discussion of these averages may be informative to the extent that the true design effects are homogeneous within the specified class. Conversely, marked heterogeneity of estimated design effects between two groups would suggest that the distinction between those two groups is of serious practical interest.

A natural extension of this idea follows from regression modeling of estimated design effects as functions of simple classificatory predictor variables. For example, consider three binary classifications i, j and k, respectively, and let x_{Ai}, x_{Bj} and x_{Ck} equal indicator variables for these three classifications, respectively, e.g., x_{Ai}={1 if i = 1; 0 otherwise}. In addition, let $\hat{\Delta}_{ijk}$ equal the estimated design effect for an outcome variable in group (i, j, k). Then one could consider approximation of the true design effect Δ_{ijk} through, e.g., the simple additive model,

$$\Delta_{ijk} = \beta_0 + \beta_1 x_{Ai} + \beta_2 x_{Bj} + \beta_3 x_{Ck} \qquad (2.1)$$

or the model with one two-factor interaction,

$$\Delta_{ijk} = \beta_0 + \beta_1 x_{Ai} + \beta_2 x_{Bj} + \beta_3 x_{Ck} + \beta_4 x_{Ai} x_{Bj} , \qquad (2.2)$$

say. In the analysis of the regression results, the coefficient of determination R^2 would be of special interest, because it represents the proportion of variability in the observed $\hat{\Delta}_{ijk}$ that is associated with the relatively simple models (2.1) or (2.2). Conversely, standard significance testing results from ordinary least squares regression output would be viewed with a high degree of caution, since the outcome variables $\hat{\Delta}_{ijk}$ will not in general be independent across different values of the ordered triple (i, j, k).

3. Exploratory Analysis for Bangladesh and Ghana

To address the issues raised in Section 2, we carried out an analysis of design effects associated with several subpopulation means and proportions estimated from the DHS for two countries, Ghana and Bangladesh. Table 1 lists three continuous variables and five categorical variables from the DHS that are considered here. For the continuous variables (rows 1-3 of Table 1), Table 2 lists the estimates of the associated design effects (*DEFF*) for subpopulations defined by the intersection of three binary classifications: *GHABD*, *URBAN*, and *AGECAT*. In addition, Table 2 lists the average number of respondents per pseudo-primary sample unit within a specified stratum, for a given outcome variable (*PSUSIZE*); and the coefficient of variation (*CV*) of the survey weights for the subpopulation in question. Results are presented separately for the three continuous variables. Table 3 gives the corresponding results for the five categorical variables (rows 4-8 of Table 1).

In keeping with the reasoning outlined in Section 2, we used ordinary least squares regression to fit the 24 rows of data in Table 2 to the model,

$$
\begin{aligned}
DEFF &= \beta_0 + \beta_1 GHABD + \beta_2 ICHEVBN \\
&+ \beta_3 IIDEAL + \beta_4 URBAN + \beta_5 AGECAT \\
&+ \beta_6 PSUSIZE + \beta_7 CV
\end{aligned}
\quad (3.1),
$$

where *ICHEVBN* and *IIDEAL* are indicators of the specific continuous variable (respectively *CHEVBORN* and *IDEAL*) for which the design effect is computed, where the baseline group is based on *CHLIVING*. Table 4 displays the results from fitting model (3.1) both for the design effects and their natural logarithms. Note that the 24 rows in Table 2 are not all independent, so the standard assumptions for inference based on ordinary least squares regression are not satisfied. Consequently, we interpret the results in Table 4 primarily in terms of simple approximation or data reduction. For example, in keeping with standard results, we can view the unadjusted R^2 values as indicating the proportion of variability in the values of *DEFF* that are "explained" by the predictors in the simple additive model (3.1). Conversely, interpretation of formal inferential results (e.g., t statistics or p values) should be interpreted only as qualitative indicators of the dominant predictors.

Table 5 presents related results for a fit of the 40 rows of Table 3 to the model

$$
\begin{aligned}
DEFF &= \beta_0 + \beta_1 GHABD + \beta_2 IMARR \\
&+ \beta_3 ICNTRAC + \beta_4 IKNAIDS \\
&+ \beta_5 IELEC + \beta_6 URBAN \\
&+ \beta_7 AGECAT + \beta_8 PSUSIZE + \beta_9 CV
\end{aligned}
\quad (3.2)
$$

where *IMARR*, *ICNTRAC*, *IKNAIDS* and *IELECT* are indicators that identify the specific categorical variables (*MARRIED*, *CNTRACEP*, *KNOWAIDS* and *ELECTRIC*) used to compute a design effect, where the baseline group is based on *EDUC*.

In addition, we applied a stepwise selection option in SAS PROG REG to explore alternative forms of models (3.1) and (3.2) for *DEFF* and its natural logarithm. The selection procedure forced in the main effects *GHABD* and *URBAN*, and used a cutoff p-value of 0.25. In addition, the forward selection option for model (3.2) allowed inclusion of a two-factor interaction between *IELEC* and *URBAN*. Results from the final selected models are reported in the final columns of Tables 4 and 5.

Examination of Tables 4 and 5 leads to the following comments. First, note that for the fixed sets of main effects, Tables 4 and 5 report R^2 values that are greater than 50%, i.e., for each of the underlying main-effect models, the predictors "explain" over half of the variability in observed values in *DEFF* and the natural logarithm of *DEFF*. Second, each of the final model fits from the model selection procedure (which allowed one two-factor interaction for an expanded version of model (3.2)) had R^2 values that are greater than 60%. Third, note that the variable selection procedure led to a substantial reduction in the number of predictors included in the models. However, the specific selected predictors varied across different outcome variables. For example, for the count variables considered for model (3.1), the models for *DEFF* and ln(*DEFF*) both led to *GHABD*, *URBAN* and *PSUSIZE* in the final selected set of predictors. On the other hand, for the categorical outcomes considered in model (3.2), the final set of predictors for *DEFF* included the main effects *GHABD*, *URBAN*, *IMARR*, *ICNTRAC*, *IELEC* and the two factor interaction *URBELEC*, while the final set of predictors for ln(*DEFF*) included one additional main effect for the predictor *CV*. Fourth, the calculated design effects show a substantial amount of variablility across several factors, notably factors related to *URBAN*, *ELECTRIC*, and *MARRIED*. Therefore, it is important that design effects be examined separately across categories of these factors. In other words, specifying a uniformly applicable design effect across these factors would not be reasonable.

Acknowledgements

The data used in this analysis was provided by ORC Macro. The authors thank Mamdou Thiam of ORC-Macro, for providing helpful information on DHS sample designs and the contents of DHS datasets. The opinions expressed in this paper are those of the authors and do not necessarily represent the policies of either the United Nations or the U.S. Bureau of Labor Statistics.

References

Cochran, W.G. (1977). *Sampling Techniques, Third Edition.* New York: Wiley.

Gabler, S., Haeder, S. and Lahiri, P. (1999). "A Model Based Justification of Kish's Formula for Design Effects for Weighting and Clustering." *Survey Methodology* **25**, 105-106.

Kish, L. (1995). "Methods for Design Effects." *Journal of Official Statistics* **11**, 55-77.

Kish, L. (1965). *Survey Sampling.* New York: Wiley.

Potthoff, R.F., Woodbury, M.A. and Manton, K.G. (1992). "Equivalent Sample Size" and "Equivalent Eegrees of Freedom" Refinements for Inference Using Survey Weights Under Superpopulation Models." *Journal of the American Statistical Association* **87**, 383-396.

Lê, T and Verma, V (1997). An Analysis of Sample Designs and Sampling Errors Sampling Errors for the Demographic and Health Surveys. *DHS Analytical Reports*, No. 3. Calverton, Maryland: Macro International, Inc.

Verma, V. and Lê, T (1996). "An Analysis of Sampling Errors for the Demographic and Health Surveys." *International Statistical Review* **64**, 265-294.

Table 1: Variables Used in the Analysis

Variable	Description
CHEVBORN	Number of children ever born
CHLIVING	Number of living children
IDEAL	Ideal number of children
MARRIED	Is currently married (1=yes; 0=no)
CNTRACEP	Is using contraceptives (1=yes; 0=no)
EDUC	Has secondary or higher education (1=yes; 0=no)
KNOWAIDS	Has knowledge of AIDS (1=yes; 0=no)
ELECTRIC	Has access to electricity (1=yes; 0=no)
GHABD	Country indicator (1=Ghana, 0=Bangladesh)
CV	Coefficient of variation of the weights
PSUSIZE	Number of respondentsin a "pseudo" PSU
URBAN	Type of residence (1=Urban, 0=Rural)
AGECAT	Age category (=1 for AGE<30; 2 for 30+)

Yansaneh, I.S. and Eltinge, J.L. (2000). "Design Effect and Cost Issues for Surveys in Developing Countries." *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 770-775.

Table 2: Design effects and related predictors for continuous variables for Ghana and Bangladesh

DEFF	GHABD	ICHEVBN	IIDEAL	URBAN	AGECAT	PSUSIZE	CV
1.60613	1	1	0	0	1	79	32.36
1.65318	1	1	0	0	2	69	35.29
1.15592	1	1	0	1	1	39	18.86
1.37131	1	1	0	1	2	30	22.09
1.62783	1	0	0	0	1	79	32.36
1.69724	1	0	0	0	2	69	35.29
1.08824	1	0	0	1	1	39	18.86
1.36347	1	0	0	1	2	30	22.09
1.75113	1	0	1	0	1	73	32.36
1.70638	1	0	1	0	2	62	35.29
1.42922	1	0	1	1	1	37	18.86
1.35905	1	0	1	1	2	27	22.09
1.60497	0	1	0	0	1	147	23.63
2.28486	0	1	0	0	2	117	25.04
2.11912	0	1	0	1	1	78	51.87
2.48054	0	1	0	1	2	67	51.88
1.7011	0	0	0	0	1	147	23.63
2.2288	0	0	0	0	2	117	25.04
1.95774	0	0	0	1	1	78	51.87
1.83854	0	0	0	1	2	67	51.88
2.89087	0	0	1	0	1	140	23.63
2.72216	0	0	1	0	2	107	25.04
2.18093	0	0	1	1	1	77	51.87
1.38517	0	0	1	1	2	65	51.88

Table 3: Design effects for Categorical variables based on data for Ghana and Bangladesh

DEFF	GHABD	IMARR	ICNTRAC	IKNAIDS	IELEC	URBAN	AGECAT	PSUSIZE	CV
1.35542	1	1	0	0	0	0	1	79	32.36
1.46771	1	1	0	0	0	0	2	69	35.29
1.12588	1	1	0	0	0	1	1	39	18.86
1.61349	1	1	0	0	0	1	2	30	22.09
1.19166	1	0	1	0	0	0	1	79	32.36
1.67003	1	0	1	0	0	0	2	69	35.29
1.21659	1	0	1	0	0	1	1	39	18.86
1.69706	1	0	1	0	0	1	2	30	22.09
2.44456	1	0	0	0	0	0	1	79	32.36
2.64556	1	0	0	0	0	0	2	69	35.29
1.41367	1	0	0	0	0	1	1	39	18.86
1.76601	1	0	0	0	0	1	2	30	22.09
1.63005	1	0	0	1	0	0	1	79	32.36
2.47303	1	0	0	1	0	0	2	69	35.29
0.78754	1	0	0	1	0	1	1	39	18.86
0.89258	1	0	0	1	0	1	2	30	22.09
7.18133	1	0	0	0	1	0	1	79	32.36
5.48332	1	0	0	0	1	0	2	69	35.29
0.98051	1	0	0	0	1	1	1	39	18.86
1.42106	1	0	0	0	1	1	2	30	22.09
0.95676	0	1	0	0	0	0	1	147	32.36
1.14733	0	1	0	0	0	0	2	117	35.29
2.54298	0	1	0	0	0	1	1	78	18.86
1.15324	0	1	0	0	0	1	2	67	22.09
2.25473	0	0	1	0	0	0	1	147	23.63
2.27263	0	0	1	0	0	0	2	117	25.04
1.39986	0	0	1	0	0	1	1	78	51.87
2.53017	0	0	1	0	0	1	2	67	51.88
2.60954	0	0	0	0	0	0	1	147	23.63
2.43106	0	0	0	0	0	0	2	117	25.04
2.7239	0	0	0	0	0	1	1	78	51.87
4.14867	0	0	0	0	0	1	2	67	51.88
3.4713	0	0	0	1	0	0	1	147	23.63
3.7167	0	0	0	1	0	0	2	117	25.04
2.57032	0	0	0	1	0	1	1	78	51.87
3.66971	0	0	0	1	0	1	2	67	51.88
9.37351	0	0	0	0	1	0	1	147	23.63
8.06139	0	0	0	0	1	0	2	117	25.04
2.67955	0	0	0	0	1	1	1	78	51.87
2.87245	0	0	0	0	1	1	2	67	51.88

Table 4: Continuous Variables
(number of design effects = 24)

Predictors	Full Model		Final Model	
	DEFF	logDEFF	DEFF	logDEFF
Intercept	5.409 (2.232)	2.234 (1.141)	3.599 (0.744)	1.621 (0.375)
GHABD	1.715 (0.728)	-0.883 (0.372)	-1.138 (0.309)	-0.650 (0.156)
URBAN	-1.285 (0.615)	-0.666 (0.315)	-0.827 (0.308)	-0.494 (0.155)
ICHEVBN	0.096 (0.159)	0.046 (0.081)	-	-
IIDEAL	0.130 (0.171)	0.06 (0.087)	-	-
AGECAT	-0.230 (0.237)	-0.103 (0.121)	-	-
PSUSIZE	-0.021 (0.135)	-0.010 (0.007)	-0.011 (0.006)	-0.006 (0.003)
CV	-0.007 (0.009)	-0.001 (0.005)	-	-
R-squared	69%	72%	64%	69%

Table 5: Categorical Variables
(number of design effects = 40)

Predictors	Full Model		Final Model	
	DEFF	logDEFF	DEFF	logDEFF
Intercept	-1.038 (4.272)	-0.808 (1.799)	3.093 (0.242)	0.517 (0.231)
GHABD	0.178 (1.393)	0.130 (0.586)	-1.106 (0.223)	-0.322 (0.108)
URBAN	0.773 (1.214)	0.193 (0.511)	-0.155 (0.249)	-0.199 (0.112)
IMARR	-1.103 (0.361)	-0.575 (0.152)	-1.042 (0.305)	-0.499 (0.131)
ICNTRAC	-0.744 (0.361)	-0.343 (0.152)	-0.683 (0.305)	-0.267 (0.131)
IKNAIDS	-0.121 (0.361)	-0.151 (0.152)	-	-
IELEC	4.924 (0.460)	1.059 (0.194)	4.985 (0.412)	1.135 (0.177)
AGECAT	0.448 (0.429)	0.235 (0.181)	-	0.115 (0.096)
PSUSIZE	0.022 (0.026)	0.009 (0.011)	-	-
CV	0.022 (0.017)	0.016 (0.007)	-	0.012 (0.005)
URBELEC	-5.381 (0.571)	-1.289 (0.240)	-5.381 (0.556)	1.059 (0.194)
R-squared	90%	81%	89%	80%