

Post Two Way Stratification Analysis for Vaccinations in Sudan

Alia Osman Sayed Saad

Assistant professor of Social Statistics

Department of Sociology and Social Work, College of Arts, Imam Abdulrahman Bin Faisal University.

P. O. Box 1982, Dammam 31441, Saudi Arabia

alsaad@iau.edu.sa

Abstract

The present research deal with post two stratification analysis for vaccinations data of MICS5 in Sudan. In order to guarantee more precision for estimated indicators: percentage of Child ever given BCG, Polio, PENTA and Measles or MMR vaccination. State and mother education level are determined as two stratification variables since the correlation between them is small and the correlation with target indicators are some of size. So the two way stratification analysis for the four indicators gives more precise estimates the post stratification with high relative efficiency.

Introduction

Most complex survey as Multiple Indicators Cluster Survey (MICS) has a lot of information which can be used to improve the precision of survey estimates. The purpose of this research is to propose an estimator of two way stratifications. However, there are many different auxiliary information in MICS, so we must choose the suitable variables which will minimize the variance of the estimated indicators.

This research focuses on deep stratification analysis of survey for child health mainly vaccinations in MICS-5 for Sudan. Avoiding preventable diseases will provide safe and healthy life for children, which will reduce infant and under - five mortality rates. Expanded Program on Immunization (EPI) since 1974 saved the lives of millions of children in the world. However, there are millions of children are not reached. So, vaccine-preventable diseases account for more than 2 million child deaths every year. Unbiased and precise estimates of indicators as percentage of child who vaccinated against tuberculosis, diphtheria, pertussis, tetanus, polio, measles, hepatitis B, homophiles influenza type b, pneumonia/meningitis, rotavirus, and rubella, are useful for planning to achieve 100% vaccination in Sudan.

Data

This research use data obtained from the multiple indicators cluster survey, round 5. Which was carried by Central Bureau of Statistics Sudan in collaboration with Ministry of Health as part of global MICS program and technical support from United Nations Children's Fund (UNICEF).

Data from MICS 5 are used in this research as information on vaccination coverage that was collected by MICS questionnaire; such as Child ever given Bacillus – Calmette - Guerin (Tuberculosis) (BCG), Polio, PENTA vaccination (DPT+HepB+Hib), and Measles, Mumps, and Rubella (MMR) vaccination for all children under three years of age.

The child health data of vaccination in Sudan are used with the auxiliary variables such as state and mother education level. These two auxiliary variables are used to estimate the target indicators in this research by using the post two way stratification analysis. The two auxiliary variables are selected under two conditions that the correlation between these variables and target indicators are some of size, and the correlation between the two auxiliary variables is small (Thomsen 1976) which will used as stratifying variables. In addition, the variance of target indicators are minimum within strata and very high between the strata (Cohran 1977) and (Kish 1965).

Methodology

Two way stratification analysis is used to estimate the research indicator:

$$\bar{y}_{Tws} = \frac{1}{n} \sum_i \sum_j G_{ij} n_{ij} \bar{y}_{ij} \quad \dots\dots\dots (1)$$

where:

\bar{y}_{ij} : the sample mean in the ijth stratum

$y_{ij} = 1$ if ijth having a particular characteristic

and

$y_{ij} = 0$ if ijth not having a particular characteristic

G_{ij} : weighting factor define by

$$G_{ij} = \frac{n^2 W_{ij}}{n_i n_j} \quad \dots\dots\dots (2)$$

in case of proportional allocation when $W_{ij} = W_i W_j$ the variance of the estimator in equation (1) is: $V(\bar{y}_{Tws}) \approx \frac{1}{n^2} \sum_i \sum_j W_{ij} S_{ij}^2$ and the only restrictions on sample size are that, for estimates of means, the sample must be at least as large as the larger of L (i=1, 2, 3, ... L) or M (j=1, 2, 3, ... M), and, for estimates of variances, the sample must be at least as large as $2(L+M)$ (Bryant 1985).

$$V(\bar{y}_{Tws}) \approx \frac{1}{n^2} \sum_i \sum_j W_{ij} S_{ij}^2 = \frac{1}{n^2} \left[\sum_i W_i (\bar{y}_{i.} - \bar{Y})^2 + \sum_j W_j (\bar{y}_{.j} - \bar{Y})^2 \right] \quad \dots\dots\dots (4)$$

since we haven't information about the population size in each ijth stratum, and the sample size of this research is large (14732). Then we used the sample in two auxiliary variables state and mother's education level to estimate the ijth weight's W_{ij} .

Simulation

This research used the Mont Carlo method () to simulate weights W_{ij} . Uncertain inputs and fixed values are modeled with triangular distribution to simulate values of the two auxiliary variables. The process is repeated hundreds of thousands times.

State	Mother's Education Level			
	None	Primary	Secondary	Higher
Northern	0.017408	0.012396	0.005289	0.002137
River Nile	0.019063	0.013575	0.005791	0.002341
Red Sea	0.013574	0.009666	0.004124	0.001667
Kassala	0.021771	0.015503	0.006614	0.002673

Gadarif	0.02784	0.019824	0.008458	0.003418
Khartoum	0.022804	0.016238	0.006928	0.0028
Gezira	0.025647	0.018263	0.007791	0.003149
White Nile	0.024604	0.01752	0.007475	0.003021
Sinnar	0.027316	0.019452	0.008299	0.003354
Blue Nile	0.033301	0.023713	0.010117	0.004089
North Kordofan	0.025703	0.018303	0.007808	0.003156
South Kordofan	0.036219	0.025791	0.011003	0.004447
West Kordofan	0.023795	0.016944	0.007229	0.002922
North Darfor	0.031052	0.022112	0.009434	0.003813
West Darfor	0.027545	0.019615	0.008368	0.003382
South Darfor	0.032338	0.023028	0.009824	0.00397
Central Darfor	0.02791	0.019874	0.008479	0.003427
East Darfor	0.029691	0.021143	0.00902	0.003646

Main Results:

In order to calculate the indicators: percentage of Child ever given BCG, Polio, PENTA and Measles or MMR vaccination; we will find the the factor G_{ij} for each indicators for two way stratification estimator using equation (2).

G_{ij} for child ever given BCG vaccination (a)

State	Mother's Education Level			
	None	Primary	Secondary	Higher
Northern	0.98165	0.983599	0.969983	0.991267
River Nile	1.00152	1.003545	0.989475	1.011692
Red Sea	0.965881	0.967813	0.954372	0.97573
Kassala	1.012204	1.014225	1.000085	1.022274
Gadarif	1.000666	1.002628	0.988714	1.010579
Khartoum	1.003962	1.00593	0.991967	1.014012
Gezira	0.983522	0.985482	0.971682	0.993342
White Nile	0.989257	0.991211	0.977457	0.999154
Sinnar	1.002405	1.004429	0.990456	1.012437
Blue Nile	0.99784	0.999813	0.985913	1.007858
North Kordofan	1.016588	1.01862	1.004348	1.026778
South Kordofan	1.019384	1.021406	1.007153	1.029549
West Kordofan	0.989546	0.991505	0.977713	0.99956
North Darfor	1.002904	1.004908	0.990943	1.013013
West Darfor	1.010809	1.012846	0.998691	1.020889

South Darfor	1.002332	1.004346	0.990307	1.012203
Central Darfor	1.007778	1.009763	0.99571	1.017883
East Darfor	0.993558	0.995552	0.981653	1.003607

G_{ij} for child ever given Polio vaccination: (b)

State	Mother's Education Level			
	None	Primary	Secondary	Higher
Northern	0.98165	0.983599	0.969983	0.991267
River Nile	1.00152	1.003545	0.989475	1.011692
Red Sea	0.965881	0.967813	0.954372	0.97573
Kassala	1.012204	1.014225	1.000085	1.022274
Gadarif	1.000666	1.002628	0.988714	1.010579
Khartoum	1.003962	1.00593	0.991967	1.014012
Gezira	0.983522	0.985482	0.971682	0.993342
White Nile	0.989257	0.991211	0.977457	0.999154
Sinnar	1.002405	1.004429	0.990456	1.012437
Blue Nile	0.99784	0.999813	0.985913	1.007858
North Kordofan	1.016588	1.01862	1.004348	1.026778
South Kordofan	1.019384	1.021406	1.007153	1.029549
West Kordofan	0.989546	0.991505	0.977713	0.99956
North Darfor	1.002904	1.004908	0.990943	1.013013
West Darfor	1.010809	1.012846	0.998691	1.020889
South Darfor	1.002332	1.004346	0.990307	1.012203
Central Darfor	1.007778	1.009763	0.99571	1.017883
East Darfor	0.993558	0.995552	0.981653	1.003607

G_{ij} for child ever given PENTA vaccination: (c)

State	Mother's Education Level			
	None	Primary	Secondary	Higher
Northern	0.98165	0.983599	0.969983	0.991267
River Nile	1.00152	1.003545	0.989475	1.011692
Red Sea	0.965881	0.967813	0.954372	0.97573
Kassala	1.012204	1.014225	1.000085	1.022274
Gadarif	1.000666	1.002628	0.988714	1.010579
Khartoum	1.003962	1.00593	0.991967	1.014012
Gezira	0.983522	0.985482	0.971682	0.993342
White Nile	0.989257	0.991211	0.977457	0.999154

Sinnar	1.002405	1.004429	0.990456	1.012437
Blue Nile	0.99784	0.999813	0.985913	1.007858
North Kordofan	1.016588	1.01862	1.004348	1.026778
South Kordofan	1.019384	1.021406	1.007153	1.029549
West Kordofan	0.989546	0.991505	0.977713	0.99956
North Darfor	1.002904	1.004908	0.990943	1.013013
West Darfor	1.010809	1.012846	0.998691	1.020889
South Darfor	1.002332	1.004346	0.990307	1.012203
Central Darfor	1.007778	1.009763	0.99571	1.017883
East Darfor	0.993558	0.995552	0.981653	1.003607

G_{ij} for child ever given Measles or MMR vaccination: (d)

State	Mother's Education Level			
	None	Primary	Secondary	Higher
Northern	0.9816497	0.983599	0.969983	0.991267
River Nile	1.0015196	1.003545	0.989475	1.011692
Red Sea	0.9658809	0.967813	0.954372	0.97573
Kassala	1.0122043	1.014225	1.000085	1.022274
Gadarif	1.0006659	1.002628	0.988714	1.010579
Khartoum	1.0039625	1.00593	0.991967	1.014012
Gezira	0.9835223	0.985482	0.971682	0.993342
White Nile	0.989257	0.991211	0.977457	0.999154
Sinnar	1.0024054	1.004429	0.990456	1.012437
Blue Nile	0.9978402	0.999813	0.985913	1.007858
North Kordofan	1.016588	1.01862	1.004348	1.026778
South Kordofan	1.0193842	1.021406	1.007153	1.029549
West Kordofan	0.989546	0.991505	0.977713	0.99956
North Darfor	1.0029037	1.004908	0.990943	1.013013
West Darfor	1.010809	1.012846	0.998691	1.020889
South Darfor	1.0023322	1.004346	0.990307	1.012203
Central Darfor	1.0077785	1.009763	0.99571	1.017883
East Darfor	0.9935584	0.995552	0.981653	1.003607

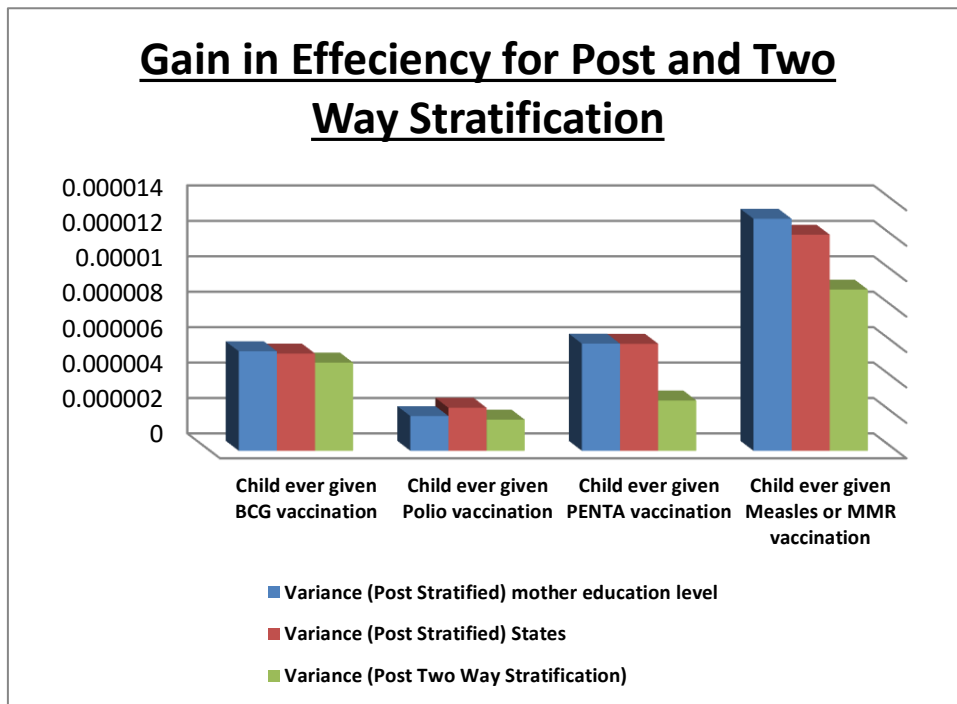
Then we calculate these indicators using equation (1) and their variances equation (4) and the post stratified mean equation (5) and it's variance equation (6):

$$\bar{y}_{ps} = \sum_i W_i \bar{y}_i \quad \dots\dots (5)$$

$$V(\bar{y}_{ps}) = \frac{(1-f)}{n} \sum_i W_i S_i^2 + \frac{1}{n^2} \sum_i (1-W_i) S_i^2 \quad \dots\dots (6)$$

Indicator	Child ever given BCG vaccination	Child ever given Polio vaccination	Child ever given PENTA vaccination	Child ever given Measles or MMR vaccination
Mean (Post Stratified) Mother education level	0.916	0.969	0.899	0.736
Mean (Post Stratified) States	0.918	0.968	0.898	0.74
Mean (Post Two Way Stratification)	0.79	0.72	0.83	0.75
Variance (Post Stratified) Mother education level	0.00000562	0.00000196	0.00000605	0.0000131
Variance (Post Stratified) States	0.00000501	0.00000243	0.00000604	0.0000122
Variance (Post Two Way Stratification)	0.00000498	0.00000176	0.00000284	0.00000091
Relative efficiency (Mother education level)	11%	10%	53%	31%
Relative efficiency (States)	9%	28%	53%	25%

As in table (6) find that the two way stratification estimators are more precise than the post stratified estimates. So we gain more efficiency due to two way stratification method.



Conclusion

In complex surveys, we are interest to increase the precision of the survey estimates. So using the auxiliary information in survey sampling will improve the precision of the survey estimates.

In this research, we used the two way stratification analysis method to obtain precise and accurate estimates. However implementation must be carefully particularly on selecting the stratification variables that will rich the estimate precise.

References

Cochran, W. G. (1977). Sampling Techniques, 3rd ed. Wiley, New York.

Edward C. Bryaat. (1955) An Analysis Of Some Two – Way Stratifications. Iowa State College.

Falorsi, P.D. and Righi, P. (2008). A balanced sampling approach for multi-way stratification designs for small area estimation. Survey Methodology, Statistics Canada.

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

Lohr, S. L. (1999). Sampling: Design and Analysis. Pacific Grove, Ca: Brooks – cole..

Singh, D. and Chaudhary, F. S. (1986). Theory and analysis of sample survey designs. New Age International (P) Limited Publishers.

Skinner, C. J., Holt, D. and Smith, T. M. F. (1989). Analysis of Complex Surveys. John Wiley & Sons Ltd.

Thomsen, I. (1976). The Effect Of Stratification When Two Stratified Variables Are Used.

Westfall, J.A., Patterson, P.L. and Coulston. J.W. (2011). Post-Stratified Estimation: Within-Strata and Total Sample Size Recommendations. NRC Research Press.