A Cost Effective Sampling Technique to Estimate Prevalence of Disease in Heterogeneous Population with a Generalized Estimator

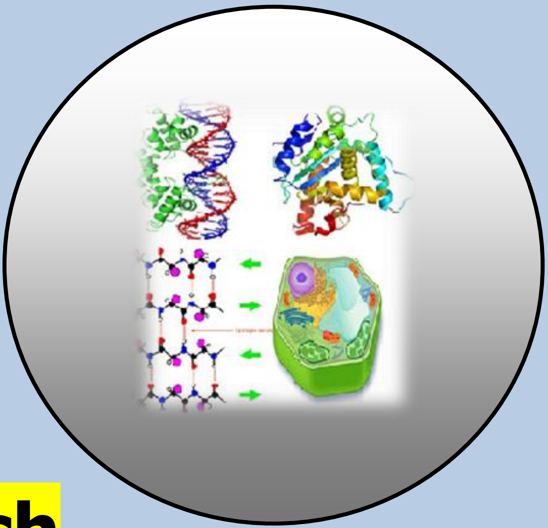
By

S. Mohd Haider Zaidi

ISSN 0970-4973 Print
ISSN 2319-3077 Online/Electronic

Global Impact factor of Journal: 0.756
Scientific Journals Impact Factor: 3.285
Index Copernicus International Value
IC Value of Journal 6.01 Poland, Europe

J. Biol. Chem. Research
Volume 31 (2) 2014 Pages No. 1267-1271



Journal of Biological and Chemical Research

An International Peer reviewed Journal of Life Sciences and Chemistry

Indexed, Abstracted and Cited: Index Copernicus International (Europe), Polish Ministry of Science and Higher Education (Poland, Europe) Research Bible (Japan), Scientific Journals Impact Factor Master List, Directory of Research Journals Indexing (DRJI), Indian Science. In, Database Electronic Journals Library (Germany), Open J-Gate, J Gate e-Journal Portal, Info Base Index, International Impact Factor Services (IIFS) (Singapore), Scientific Indexing Services (USA), International Institute of Organized Research (I2OR), Eye Source and

citefactor.org journals indexing

Directory Indexing of International Research Journals

Published by Society for Advancement of Sciences®

J. Biol. Chem. Research. Vol. 31, No. 2: 1267-1271, (2014)
(An International Peer reviewed Journal of Life Sciences and Chemistry)
Ms 32/1/240/2014
All rights reserved
ISSN 2319-3077 (Online/Electronic)
ISSN 0970-4973 (Print)

Dr. S.M.H. Zaidi
http:// www.sasjournals.com
http:// www.jbcr.in
jbiolchemres@gmail.com
info@jbcr.in

RESEARCH PAPER

Received: 25/10/2014 Revised: 01/12/2014 Accepted: 03/12/2014

A Cost Effective Sampling Technique to Estimate Prevalence of Disease in Heterogeneous Population with a Generalized Estimator

S. Mohd Haider Zaidi
Department of Statistics, Shia PG College, Sitapur Road, Lucknow – 226020, U.P., India

ABSTRACT

Most methods consider sampling only from a homogeneous population in which each person has the same probability of becoming infected. In this paper, we propose a stratified sampling technique with a new cost effective optimum allocation method to estimate the prevalence of the disease if it is present in a heterogeneous population with potentially different risk groups. An optimum allocation with a generalized class of estimators is discussed and compared with the Neyman optimum allocation. The proposed optimum allocation is better than the Neyman optimum allocation in the sense of having less approximate variance. The comparison shows that the proposed allocation is always better than Neyman allocation except the correlation coefficient between the main and auxiliary variables (ρ_h) in the hth stratum is equal to zero, in which case both are equally efficient.
Keywords: Sampling technique, Heterogeneous population, Prevalence and Optimum Allocation.

INTRODUCTION

Not much cost effective techniques are available to detect prevalence of a disease in case if it is present in a heterogeneous population with potentially different risk groups. Most methods consider sampling only from a homogeneous population in which each person has the same probability of becoming infected. Those methods are not good for heterogeneous population as sampling error rapidly increases with heterogeneity.

Sometimes a meta analysis with random effect model has been done when there is heterogeneity that cannot readily be explained. A random effects meta-analysis model involves an assumption that the effects being estimated in the different studies are not identical, but follow some distribution. The conventional choice of distribution is a normal distribution. But it is difficult to establish the validity of any distributional assumption, and this is a common criticism of random effects meta-analyses. However, simulations have shown that methods are relatively robust even under extreme distributional assumptions, both in estimating heterogeneity [**Kontopantelis et al., 2013**].

THE NEW COST EFFECTIVE OPTIMUM ALLOCATION METHOD

Starting with stratified sampling technique, which is best suited for heterogeneous population, let a population of size N be stratified into *h* non-overlapping *L* strata, the *h*th stratum with size N_h and also let

$$W_h = \frac{N_h}{N}; \bar{Y}_h = \frac{1}{N_h}\sum_{j=1}^{N_h} Y_{hj} \text{ and } S_{yh}^2 = \frac{1}{(N_h - 1)}\sum_{j=1}^{N_h}\left(Y_{hj} - \bar{Y}_h\right)^2$$

Where Y_{hj} is the value on the *j*th unit according to a character *y* under our study in the *h*th stratum, $j = 1, 2, ..., N_h; h = 1, 2, ..., L$. Further for a simple random sample of size n_h from the h^{th} stratum with its values $y_{h1}, y_{h2}, ..., y_{hj}, ...y_{hn_h}$, let

$$\bar{y}_h = \frac{1}{n_h}\sum_{j=1}^{n_h} y_{hj} \; ; \; s_{yh}^2 = \frac{1}{(n_h - 1)}\sum_{j=1}^{n_h}\left(y_{hj} - \bar{y}_h\right)^2 \text{ and } \bar{y}_{st} = \sum_{h=1}^{L} W_h \bar{y}_h$$

Using information on an auxiliary character x, proceeding on the lines of Srivastava (1971), consider the class of estimators

$$\bar{y}_g = \sum_{h=1}^{L} W_h \bar{y}_h g\left(\frac{\bar{x}_h}{\bar{X}_h}\right)$$

Or,

$$\bar{y}_g = \sum_{h=1}^{L} W_h \bar{y}_h g\left(u_h\right) \qquad \text{……(1)}$$

Where $g(1) = 1$, $\bar{x}_h$ is the mean of the values on x in the simple random sample from the *h*th stratum, $u_h = \frac{\bar{x}_h}{\bar{X}_h}$.

It may be noted here that the usual $\bar{y}_{st}$, separate ratio and product estimators are special cases of $\bar{y}_g$ for $g(u_h)=1$, u_h^{-1} and u_h respectively. Under certain conditions and using Taylor's expansion of $g(u_h)$ about the point $u_h=1$ from [a2] we have the approximate variance of $\bar{y}_g$.

$$V(\bar{y}_g)=\sum_{h=1}^{L}W_h^2\frac{(N_h-n_h)}{N_h n_h}\left[S_{yh}^2+R_h^2S_{xh}^2g_{1h}^2+2R_hg_{1h}S_{y\times h}\right] \quad ......(2)$$

where $R_h=\dfrac{\bar{Y}_h}{\bar{X}_h}$; $S_{xh}^2=\dfrac{1}{(N_h-1)}\sum_{j=1}^{N_h}\left(X_{hj}-\bar{X}_h\right)^2$;

X_{hj} being the value on the j^{th} unit in the h^{th} stratum according to the auxiliary character *x*.

$S_{y\times h}=\dfrac{1}{(N_h-1)}\sum_{j=1}^{N_h}\left(Y_{hj}-\bar{Y}_h\right)\left(X_{hj}-\bar{X}_h\right)=\rho_hS_{yh}S_{xh}$ and g_{1h} being the first partial derivative of $g(u_h)$ at the point *u* = 1.

It may be easily seen that $V(\bar{y}_g)$ is minimized for

$g_{1h}=-\dfrac{\rho_hS_{yh}}{R_hS_{xh}}=-\rho\dfrac{C_{yh}}{C_{xh}}$ where $C_{yh}=\dfrac{S_{yh}}{\bar{Y}_h}$, $C_{xh}=\dfrac{S_{xh}}{\bar{X}_h}$

And the minimum $V(\bar{y}_g)$ is given by,

$$V(\bar{y}_g^*)=\sum_{h=1}^{L}\frac{W_h^2(1-\rho_h^2)S_{yh}^2}{n_h}-\sum_{h=1}^{L}\frac{W_h^2(1-\rho_h^2)S_{yh}^2}{N_h} \quad ......(3)$$

SAMPLING WITH AN OPTIMUM ALLOCATION

Considering the cost function $C=C_0+\sum_{h=1}^{L}C_hn_h$ where C_0 and C_h are the overhead cost and cost per unit within h^{th} stratum respectively. Minimizing $V(\bar{y}_g)$ for the given cost restriction

$$C_1n_1+C_2n_2+...+C_Ln_L=C-C_0$$

By using Lagrange's method of multipliers, we have the optimum allocation

$$n_h=n\cdot\frac{W_h\left(1-\rho_h^2\right)^{1/2}S_{yh}/\sqrt{C_h}}{\sum_{h=1}^{L}W_h\left(1-\rho_h^2\right)^{1/2}S_{yh}/\sqrt{C_h}}\ ;\ h=1,2,...,L \quad ......(4)$$

In particular, if $C_h = C$ *i.e.* for the given cost function $C = C_0 + Cn$, the optimum allocation (4) reduces to

$$n_h = n \cdot \frac{W_h \left(1-\rho_h^2\right)^{1/2} S_{yh}}{\sum_{h=1}^{L} W_h \left(1-\rho_h^2\right)^{1/2} S_{yh}}\,;\; h = 1, 2, \ldots, L \qquad \text{......(5)}$$

Substituting the value of n_h from (5) in (3), we have

$$V\left(\bar{y}_g\right)_o = \frac{1}{n}\left[\sum_{h=1}^{L} W_h \left(1-\rho_h^2\right)^{1/2} S_{yh}\right]^2 - \frac{1}{N}\sum_{h=1}^{L} W_h^2 \left(1-\rho_h^2\right) S_{yh}^2 \qquad \text{......(6)}$$

Further we know [3] that the variance of $\bar{y}_{st}$ under Neyman optimum allocation

$$\left(n_h = n \cdot \frac{W_h S_{yh}}{\sum_{h=1}^{L} W_h S_{yh}}\,;\; h = 1, 2, \ldots, L \right)$$

$$V\left(\bar{y}_{st}\right)_N = \frac{1}{n}\left[\sum_{h=1}^{L} W_h S_{yh}\right]^2 - \frac{1}{N}\sum_{h=1}^{L} W_h^2 S_{yh}^2 \qquad \text{......(7)}$$

From (6) and (7) we conclude that the proposed allocation given by (5) [or (4)] is more efficient than Neyman optimum allocation except in the particular case when the correlation coefficient between the main and auxiliary variables $\left(\rho_h\right)$ in the *h*th stratum is equal to zero, in which case both are equally efficient.

Usually, in practice ρ_h the correlation coefficient between the study variable *y* and the auxiliary variable *x* in the *h*th stratum assumes high values, hence the proposed allocation behaves quite significantly better than the existing Neyman optimum allocation in almost all practical situations. It is to be pointed out that in any situation the proposed estimator is either equally or more efficient than Neyman allocation.

ILLUSTRATION

Taking the examples from Cochran (1977) [4] on page 167 and 172, the following table gives the relative efficiency of the proposed allocation over the Neyman allocation.

Examples from Cochran (1997)	$V\left(\bar{y}_{st}\right)_N$	$V\left(\bar{y}_{st}\right)_o$	Efficiency of proposed allocation over Neyman allocation
On page 167	3.9334	2.7424	143.17%
On page 172	37.4403	6.2058	603.31%

Seeing the last column of the above table, we infer that the efficiency of the proposed allocation as compared to Neyman allocation is very high.

ACKNOWLEDGEMENTS
Author is grateful to Principal Shia P.G. College, Lucknow, U.P., India for providing all facilities during course of present investigation.

REFERENCES
Kontopantelis, E.; Springate, D. A.; Reeves, D. (2013). Friede, Tim, ed. "A Re-Analysis of the Cochrane Library Data: The Dangers of Unobserved Heterogeneity in Meta-Analyses". PLoS ONE. 8 (7)

Srivastava, S.K. (1971) "Generalized Estimator for Mean of a Finite Population using Multi-Auxiliary Information." Journal of American Statistical Association, 66 *pp*. 404-407

Murthy, M.N. (1967): Sampling Theory and Methods, Statistical Publishing Society, Calcutta, India.

Cochran, W.G. (1977): Sampling Techniques, John Wiley & Sons, Inc., New York

Corresponding author: Dr. Syed Mohd Haider Zaidi, Department of Statistics, Shia P.G. College, Sitapur Road, Lucknow – 226020, U.P., India
Email: syedmohdhaiderzaidi@yahoo.com