ANNALS OF FAMILY MEDICINE

Online Supplementary Material

Wadland WC, Holtrop J, Weismantel D, Pathak PK, Fadel H, Powell J. Practice-based referral rates to a tobacco cessation quit line: assessing the impact of comparative feedback vs general reminders. *Ann Fam Med*. 2007;5(2):135-142.

http://www.annfammed.org/cgi/content/full/5/2/135/DC1

Supplemental Appendix. An Intraclass Correlation Model for a Physician-Based Sampling Scheme

Pramod K. Pathak[1,2,3]
David Weismantel[4]
William Wadland,[4] *et al*

[1]Department of Statistics and Probability, Michigan State University, East Lansing, Mich
[2]Department of Epidemiology, Michigan State University, East Lansing, Mich
[3]Department of Family and Community Medicine, Michigan State University, East Lansing, Mich
[4]Department of Family Practice, Michigan State University, East Lansing, Mich

ABSTRACT
We derive a closed-form formula for intraclass correlation for a sampling scheme based on physician-based research networks. The formula is applied to an illustrative example from our ongoing research on physician-assisted smoking-cessation programs. It provides a means of calculating the design effect of cluster-based sampling schemes compared with simple random sampling.

1. Statement of the Problem
Consider a population of N clusters (sites) of unequal sizes. Suppose that the jth cluster contains M_j units (clinicians), 1 ≤ j ≤ N. Suppose that a cluster is first drawn at random from the given population of N clusters. Then a pair of 2 units is drawn one-by-one therefrom by simple random sampling without replacement. Now let y_1 and y_2 denote the Y-variate values associated with the 2 units drawn, the respective number of subjects referred to by the 2 clinicians for smoking cessation program. Let ρ denote the coefficient of correlation between y_1 and y_2. We refer to ρ as the intraclass (cluster) correlation for the Y-variate values. It measures the intracluster association of Y-variate values.

It can be shown that

$$\begin{aligned} \mathrm{Var}(y_1) &= \mathrm{Var}(y_2) \\ &= \frac{1}{N}\Sigma\sigma^2_j + \frac{1}{N}\quad \Sigma\,(Y_{j.} - Y_{..})^2 \end{aligned} \tag{1.1}$$

where for each j, 1 ≤ j ≤, $Y_{j.}$ and σ^2_j respectively denote the mean and the variance of the Y-variate values in the jth cluster, $Y_{..} = N^{-1}\,\Sigma Y_{j.}$ being the unweighted average of the cluster means $Y_{j.}$.

And

$$\mathrm{Cov}(y_1, y_2) = -\frac{1}{N}\Sigma\frac{1}{(M_j-1)}\sigma^2{}_j + \frac{1}{N}\Sigma\,(Y_{j.} - Y_{..})^2 \tag{1.2}$$

Therefore, the intracluster correlation admits the following form:

$$\rho(y_1, y_2) = \frac{-\Sigma\,(M_j-1)^{-1}\sigma^2{}_j + \Sigma(Y_{j.} - Y_{..})^2}{\Sigma\sigma^2{}_j + \Sigma\,(Y_{j.} - Y_{..})^2} \tag{1.3}$$

2. Illustrative Example
The data for the smoking cessation consists of N = 87 sites and the variable of interest is the total number of referrals by each clinician (Y). Initial analysis of the data resulted in the following statistics:

Mean cluster size: $M_{.} = 3.59$

Between cluster variation: $\Sigma\,(Y_{j.} - Y_{..})^2 = 14580.13$

Total within cluster variation: $\Sigma\sigma^2{}_j = 1175.82$

Total per unit within cluster variation: $\Sigma\,[(\sigma^2{}_j)/((M_j - 1))] = 367.03$

The intracluster correlation is:

$$\rho = \frac{(-367.03 + 14580.13)}{(1175.82 + 14580.13)} = .902 \tag{2.1}$$

And the design effect of the survey for the smoking cessation study is:

Design effect: $1 + (M_{.} - 1)\ \rho = 3.34$

Reference

1. Pathak PK. An introduction to sampling theory and practice. Available at: http://www.stt.msu.edu/~pathakp/book0-5. Accessed 13 March 2007.