ON THE SAMPLE SIZE DETERMINATION FOR MULTIPLE VARIABLE SAMPLE SURVEYS

Dhirendra N. Ghosh, Science Applications International Corporation

In a multiple variable sample survey, if one wants to ensure that all the estimates satisfy a simultaneous confidence interval, i.e., the estimates all fall within prescribed limits with a certain probability (say 95%), the following equation must hold:

$$\text{Prob.}[|(\bar{x}_i - M_i)| \leq d_i, \ldots |(\bar{x}_k - M_k)| \leq d_k] = 0.95.$$

In order to determine the necessary sample size, one needs to know the variance-covariance matrix and calculate the multivariate normal probabilities where the integral for the ith variate has the limits

$$\left(\frac{-d_i}{\sigma_i/\sqrt{n_i}}, \frac{d_i}{\sigma_i/\sqrt{n_i}} \right)$$

However, multivariate normal tables for more than three variables are not known. Thus this method does not work for cases with more than three variables.

An approximation may be worked out in the following way. It is known that

$$T^2 = n\,[\bar{x}_k - M_k]\ \Sigma^{-1}\ [\bar{x}_i - \bar{M}_i, \ldots \bar{x}_k - M_k]^1$$

is distributed as X^2 with k degrees of freedom. If d_i is substituted for $\bar{x}_i - M_i$ and T^2 is equated to the tabulated values corresponding to a probability level, n, the sample size can be determined.

The objective of this paper is to examine the closeness of approximation achieved by this method. The following table presents the values of the sample size determined by both the above methods for several cases.

σ_1^2	σ_2^2	δ_1	δ_2	ρ	μ calculated from normal tables	η calculated from x^2 distribution
100	169	2	3	0	115	64
"	"	"	"	.10	112	75
"	"	"	"	.30	106	84
"	169	"	"	.65	100	104
"	"	"	"	.80	96	112
25	25	1	1	0	120	75
"	"	"	"	.10	110	100
"	"	"	"	.30	105	124
"	"	"	"	.65	102	130
"	"	"	"	.80	98	135

The above table shows that the sample size calculated from the x^2 tables do not provide a close approximation of the true sample size determined from normal tables.

REFERENCES

1. Ghosh, Dhirendra N: Determining the Sample Size For Multivariate Estimates Satisfying Simultaneous Confidence Intervals: 1984 Proceedings of the American Statistical Section on Survey Research Methods.
2. Goodman, Leo: On Simultaneous Confidence Intervals for Multinomial Proportion. Technometrics, Vol 7, 1965.
3. Moonan, William J: On the Problem of Sample Size for Multivariate Simple Random Sampling. Journal of Experimental Education. Vol 22, 1952.
4. Roy, S.N. and Bose, R.C.: Simultaneous Confidence Interval Estimation, Annals of Mathematical Statistics. Vol 24, 1955
5. Tables of the Bivariate Normal Distribution Function and Relation Function. NBS, 1959
6. Steck, G.P.: A Table for Computive Trivariate Normal Population, Ann. Math. Stat Vol 29, 1959