Jour. Ind. Soc. Ag. Statistics
44 *(3), 1992 : 256–259*

A Note on the Performance Confidence Coefficient of Investigators

S. Mohanty
Orissa University of Agriculture and Technology, Bhubaneswar
(Received : December 1990)

Summary

A method of sampling experimentation has been suggested to assess the performance confidence coefficient of the investigators in similar line as the coefficient of repeatability used in quantitative genetics. This coefficient will act as a yardstick of faith on the results of census or survey sampling.

Keywords : Random effect model, Normal distribution; Coefficient of repeatability.

Introduction

In the theory of sampling, it is generally assumed that to each unit U_i in the population, a value X_i, called the true value of the unit is attached and it is further assumed that whenever the U_i is selected, the character X is always measured by X_i. This assumption is an over simplification of the actual problem. The characters which are measured by physical measurements are not likely to differ from X_i. But characters for which the information is collected by interview and require personal judgement, are likely to differ from X_i's and also likely to change with the investigator collecting the information. Thus, there is always a possibility of obtaining different observations on the same unit by different investigators. These, observations can very well be assumed as observations on a random variable. Further, when an investigator is collecting information by interviewing different units, experience shows that the responses obtained by him can not be assumed to be uncorrelated as his personality and judgement affects the observations he produces. Similarly correlation also exist between the observation on a unit by different investigators. The reason that the investigator is brought into the picture is that the results of modern large scale surveys depend on the ability of the investigator specially trained for the purpose. Any observational error made by the investigator will inflate the error of the survey.

In this note a sampling scheme has been proposed, where multiple measurement of all the units of the sample are utilised to estimate a coefficient analogous to repeatability defined in quantitative genetics. This coefficient helps to study the reliability of the performance of the investigators.

When more than one measurement of the character X is made on each unit of the sample, the observed total variance can be partitioned in to variances *(i)* within the units and *(ii)* between the units. this subdivision serves to show how much is to be gained by the repetition of the measurements and throw some light on the environmental variation which usually affects the large scale surveys. Within the unit component is entirely due to bias of the investigators present in the measurement of the same unit and between component is partly due to true value of the units and partly due to environmental condition attached to the survey. The ratio of the between the unit component to the total observed variance measures the coefficient of intraclass correlation between the repeated measurements of the same unit and is known as coefficient of repeatability in quantitative genetics. The knowledge of the above ratio helps to measure the degree of true determination of the character under study in the large scale survey.

2. Sampling scheme

Let us select a sample of size n with replacement in which U_i occurs f_i ($\neq 1$) times, such that $\sum_{i=1}^{d} f_i = n$, d being the number of distinct units in the sample.

Let us select a random sample of m investigators out of M investigators trained for the survey. The distinct d units of the sample are distributed among the m investigators in such a manner that one investigator only makes one observation on U_i and the same unit is observed by f_i investigators. The observations thus collected are analyzed following the random effect model under one way classification.

The random effect model for the above experiment is given by :

$Y_{ij} = \mu + b_i + e_{ij}$, where

y_{ij} is the measurement of U_i on the j-th occasion, for j = 1, 2,.. . f_i, there being f_i measurements for U_i, i = 1, 2, . . ., d, μ is the over all mean, b_i is the random effect of U_i and e_{ij} is the error present in the measurement of the i-th unit on the j-th occasion.

Here, both b_i and e_{ij} are independent random variables normally distributed with mean o and variance σ_b^2 and σ_e^2 respectively. The following ANOVA table is obtained by analysing the data as per the given model.

Table 1

Source of variation	df	ss	EMS
Between the units	d–1	$\sum_{i=1}^{d} \frac{Y_{i.}^2}{r_i} - \frac{Y_{..}^n}{n}$	$\sigma_e^2 + \lambda\,\sigma_b^2$
Within the units	n–d	$\sum_i \sum_j Y_{ij}^2 - \sum_i \frac{Y_{i.}^2}{f_i}$	σ_e^2
Total	n–1	$\sum \sum Y_{ij}^2 - \frac{Y_{..}^2}{n}$	

The notations used in the Table-1 are as follows :

$$Y_{..} = \sum_i \sum_j Y_{ij}\,, \quad \sum_{j=1}^{f_i} Y_{ij} = Y_{i.} \qquad \text{and}$$

$$\lambda = \frac{n^2 - \sum_{i=1}^{d} f_i^2}{n(d-1)}$$

Equating the observed mean square with that of expected mean square, σ_b^2 and σ_e^2 are estimated.

The coefficient of repeatability is estimated as $\frac{\hat{\sigma}_b^2}{\hat{\sigma}_b^2 + \hat{\sigma}_e^2}$. When the estimated repeatability is very high, it means that most of the variation in the sample is between the units. From this one can conclude that error in measuring the same unit by different investigators is small. This tells about the reliability of the investigators in measuring the true character under study and may be named as the performance confidence coefficient of the investigators.

This experiment is to be carried out to calculate the performance confidence coefficient of the investigator before embarking on the actual survey work. If the coefficient is found to be low, then further training need to be given to the investigators to reduce error of measurement in the main survey work.

When the performance confidence coefficient is high, multiple measurements are not so useful as indicated above. However, when it is low, the gain due to multiple measurements are substantial. The gain however falls off rapidly as the number of measurements increases and is seldom worthwhile to make more than two measurements. Mohanty [1] has shown that repeating some of the units of the sample twice in such cases, the precision of the estimate can be increased always by using a linear estimate of mean per unit based all units and ratio or regression estimate based on repeated units.

This sampling experimentation may also be carried out alongwith the main survey on census to know the performance confidence coefficient of the investigators. This coefficient will act as a yardstick of faith on the results of the census or sample survey.

REFERENCES

[1] Mohanty, S., 1977. Sampling with repeated units. *Sankhya, Series C, Pt-1*, **39**, 43-46

[2] Narain, P., 1990. Statistical genetics. Wiley Eastern.