DOCUMENT RESUME

ED 110 495 TM 004 768

AUTHOR Pandey, Tej N.
TITLE Estimating the Standard Error of the Mean in Multiple Matrix Sampling When Items are Sampled With and Without Replacement.
PUB DATE [Apr 75]
NOTE 12p.; Paper presented at the Annual Meeting of the American Educational Research Association (Washington, D.C., March 30-April 3, 1975)

EDRS PRICE MF-$0.76 HC-$1.58 PLUS POSTAGE
DESCRIPTORS *Comparative Analysis; *Item Sampling; *Matrices; *Standard Error of Measurement; *Statistical Analysis
IDENTIFIERS *Multiple Matrix Sampling

ABSTRACT

Standard errors of pooled mean estimate in multiple matrix sampling were compared for two procedures. The data were from tests involving items with and without replacement. The two procedures involve the formulations of Madow and Lord, and Novick; the former permits sampling of item, with or without replacement, whereas the latter is to be used for item sampling without replacement. The results show that the two estimates give considerably differing error estimates of the pooled mean. (Author)

ED110495

ESTIMATING THE STANDARD ERROR OF THE MEAN IN MULTIPLE MATRIX SAMPLING WHEN ITEMS ARE SAMPLED BOTH WITH AND WITHOUT REPLACEMENT

Tej N. Pandey
California State Department of Education

Paper Presented At
American Educational Research Association Meeting, 1975
Washington, D.C.

TM004 768

Estimating the Standard Error of the Mean in Multiple Matrix Sampling When Items are Sampled With and Without Replacement

Tej N. Pandey
California State Department of Education

INTRODUCTION

Multiple matrix sampling is a procedure in which a universe of test items is subdivided into more than one test form with each form administered to a certain number of examinees. Although each examinee is administered only a portion of the test items in the total pool, the results from each form administered may be used to estimate the parameters of the matrix universe and associated standard errors. Several states, for example, California, Oregon, and New Mexico are using multiple matrix sampling procedure advantageously for their statewide assessment programs--providing group information at relatively lesser cost and testing time as compared to the traditional testing procedures.

A review of the matrix sampling literature dealing with the estimation of the mean and associated standard error indicates that the major emphasis in the matrix sampling item allocation designs is towards those which allocate an equal number of items to each form and items are sampled without replacement. The equations for estimating the standard error of the pooled mean under these assumptions have been given by Lord and Novick (1968) for dichotomus item scores and by Pandey and Shoemaker (1975) for polychotomus item scores. The available computer programs utilize either one or the other of these equations to compute the standard error of the pooled mean.

To date, not much attention has been given towards estimation of the standard error of the pooled mean involving item sampling designs allocating items both with and without replacement and possibly an unequal number of items in the forms. The necessity for such applications would be common in reading tests for lower grades, requiring simultaneous oral administration for a part of each test form and the remaining items may be unique item samples from the item pool. Furthermore, it is not uncommon in such area as reading comprehension where items are related to a passage; an equal number of items to each form may not always be possible.

This paper presents relevant equation from Madow (1972) for estimating the standard error of the pooled mean in matrix sampling. The equation was derived for cases of stratified samples of persons and items, with possibly unequal sizes of samples, and with possible overlap of samples. Madow's derivations utilize conditional variance theorem rather than polykays and bipolykays used by Lord and Novick. This paper presents the equation modified so as to be applicable for cases of items sampled both with and without replacement. Also, a more general equation for the estimation of standard error of the mean has been derived starting with Lord and Novick's formulations, which is shown to be equivalent to the Madow's equation for the special case of sampling of items without replacement.

For a typical data set, this paper compares the standard error of the pooled mean as computed using Madow's equation and those computed from Lord and Novick's equation using certain approximations to satisfy the assumptions underlying the equation.

NOTATIONS

Let us suppose a population U of N persons and a universe V of K items. It is proposed to estimate the average score, $\overline{X}$, that would be obtained if all N persons in U took all K items in V, and the standard error of the estimated mean. Furthermore, suppose the population U has been stratified into G strata U_1, U_2, ..., U_G, where U_g consists of N_g persons, $g=1, 2, \ldots, G$, and $\sum N_g=N$. Also, that the universe V has been stratified into D strata V_1, V_2, ..., V_D where V_d consists of K_d items, $d=1, 2, \ldots, D$, and $\sum K_d=K$.

Suppose that T samples, u_{tg} are selected by simple random sampling from U_g, and T samples v_{td} are selected by simple random sampling from V_d; the number of elements in u_{tg} is n_{tg} and the number of elements in v_{td} is k_{td}, $t=1, 2, \ldots, T$. Denote u_t the sample consisting of the elements of u_{t1}, u_{t2}, ..., u_{tG} and v_t the sample consisting of the elements of v_{t1}, v_{t2}, ..., v_{tD} having $n_t=n_{t1}+n_{t2}+\ldots+n_{tG}$ and $k_t=k_{t1}+k_{t2}+\ldots+k_{tD}$ elements respectively. The pair (u_t, v_t) is defined as the t^{th} stratified matrix sample.

COMPUTATIONAL FORMULAS

Following Madow (1972), the computational formula for the unbiased estimate of the mean through multiple matrix sampling is:

$$\overline{x} = \sum_t c_t \, \overline{x}_t ,$$

where c_1, c_2, ..., c_T are constants, and

$$\overline{x}_t = \sum_g \sum_d \frac{N_g}{N} \cdot \frac{K_d}{K} \quad \overline{x}_{tgd}$$

and

$$\overline{x}_{tgd} = \frac{1}{n_{tg}\, k_{td}} \sum_i \sum_j x_{tgdij} ,$$

$i=1, 2, \ldots, n_{tg}$; $j=1, 2, \ldots, k_{td}$, $t=1, 2, \ldots, T$.

Assuming no strata in the sampling of persons, but two strata for the sampling of items--one strata for the sampling of items with replacement, the other for the sampling of items without replacement--the standard error for the estimate of the mean in multiple matrix sampling is given by:

$$\mathrm{Var}\,(\bar{x}) = \frac{1}{T^2}\left[\sum_t\left(\frac{1}{n_{tg}} - \frac{1}{N}\right) - \frac{T(T-1)}{N}\right]\sigma^2_{\bar{x}_{i.}}$$
$$+ \left[\left(\frac{K_{d_1}}{K}\right)^2\left(\frac{1}{k_{td_1}} - \frac{1}{K}\right) + \left(\frac{K_{d2}}{K}\right)^2\frac{1}{T^2}\left\{\sum_t\left(\frac{1}{k_{td_2}} - \frac{1}{K}\right) - \frac{T(T-1)}{K}\right\}\right]\sigma^2_{\bar{x}_{.j}}$$
$$+ \left[\left(\frac{K_{d_1}}{K}\right)^2\frac{1}{T^2}\left\{\left(\frac{1}{k_{td_1}} - \frac{1}{K}\right)\sum_t\left(\frac{1}{n_{tg}} - \frac{1}{N}\right) - \frac{T(T-1)}{N}\left(\frac{1}{k_{td_1}} - \frac{1}{K}\right)\right\}\right.$$
$$\left. + \left(\frac{K_{d_2}}{K}\right)^2\frac{1}{T^2}\left\{\sum\left(\frac{1}{n_{tg}} - \frac{1}{N}\right)\left(\frac{1}{k_{td_2}} - \frac{1}{K}\right) + \frac{T(T-1)}{NK}\right\}\right]\sigma^2_{\bar{x}..}$$

where $\sigma^2_{\bar{x}_{i.}}$, $\sigma^2_{\bar{x}_{.j}}$, and $\sigma^2_{\bar{x}..}$ are the population variances associated due to person effect, item effect, and person x item effect in a linear model of test scores. The variance representing the standard error is a composite of the three variances--due to sampling of persons, sampling of items and an interaction term. Also, each of the latter two terms are shown to be composite of two terms--due to sampling of items with replacement and sampling of items without replacement. If the sampling of items is without replacement only, the above equation can be written as:

$$\mathrm{Var}\,(\bar{x}) = \left[\frac{1}{T^2}\sum_t\frac{1}{n_{tg}} - \frac{1}{N}\right]\sigma^2_{\bar{x}_{i.}} + \left[\frac{1}{T^2}\sum_t\frac{1}{k_{td}} - \frac{1}{K}\right]\sigma^2_{\bar{x}_{.j}}$$
$$+ \left[\frac{1}{T^2}\sum_t\frac{1}{n_{tg}\,k_{tg}} - \frac{1}{KT^2}\sum_t\frac{1}{n_{tg}} - \frac{1}{NT^2}\sum_t\frac{1}{k_{td}} + \frac{1}{NK}\right]\sigma^2_{\bar{x}..}$$

It is easy to note that for finite N, the first term representing the variance due to sampling of persons vanishes if the number of persons taking each form is equal. Similarly, for finite K, the second term representing the variance due to sampling of items vanishes if the number of items assigned to each form is equal. Therefore, when $N=T\,n_{tg}$ and $K=T\,k_{td}$ for t=1, 2,..., T; the computational formula for the standard error of the mean is given by:

$$\text{Var}\ (\bar{x}) = \frac{1}{NK}\left[T-1\right]\sigma^2_{x..}$$

The foregoing equations indicate that the standard error of the mean can be reduced by assigning equal number of items to forms and administering forms in a manner so that equal number of persons take each form.

Lord and Novick (1968) present the formulas for the standard error of the mean when the items are sampled without replacement for (a) items sampled inexhaustively (equation 11.12.3) and (b) items sampled exhaustively (equation 11.12.4). These equations are given for binary item scoring. It is shown that how relatively more general equation can be arrived at from Lord and Novick's equations (11.11.6) and (11.12.2). [Notations have been changed here for the sake of consistency.]

By the formula for the variance of a sum,

$$\text{Var}\ (\bar{x}) = \frac{1}{T^2}\left[\sum_t \text{Var}\ (\bar{x}_t) + \sum_{t \neq t'}\sum \text{Cov}\ (\bar{x}_t\ ,\ \bar{x}_{t'})\right]$$

Using equation (11.11.6), it can be shown that

$$\sum_t \text{Var}\ (\bar{x}_t) = \sigma^2_{x_{i.}}\sum_t\left(\frac{1}{n_t} - \frac{1}{N}\right) + \sigma^2_{x_{.j}}\sum_t\left(\frac{1}{K_t} - \frac{1}{K}\right) + \sigma^2_{x..}\sum_t\left(\frac{1}{n_t} - \frac{1}{N}\right)\left(\frac{1}{K_t} - \frac{1}{N}\right)$$

Also using equation (11.12.2), it can be shown that

$$\sum_{t \neq t'}\sum \text{Cov}\ (\bar{x}_t,\ \bar{x}_{t'}) = T(T-1)\left[\frac{1}{NK}\sigma^2_{x..} - \frac{1}{N}\sigma^2_{x_{i.}} - \frac{1}{K}\sigma^2_{x_{.j}}\right]$$

Combining (11.11.6) and (11.12.2), we get

$$\mathrm{Var}(\bar{x}) = \left[\frac{1}{T^2}\sum_t \frac{1}{n_t} - \frac{1}{N}\right]\sigma^2_{\bar{x}_{i.}} + \left[\frac{1}{T^2}\sum_t \frac{1}{k_t} - \frac{1}{k}\right]\sigma^2_{\bar{x}_{.j}}$$

$$+ \left[\frac{1}{T^2}\sum_t \frac{1}{n_t k_t} - \frac{1}{KT^2}\sum_t \frac{1}{n_t} - \frac{1}{NT^2}\sum_t \frac{1}{k_t} + \frac{1}{NK}\right]\sigma^2_{\bar{x}_{..}}$$

which is the same as the equation derived by Madow. In the opinion of the author, a computer program using the above equation will be more useful than programming for equations (11.12.3) and (11.12.4). The later are the special cases of the above equation.

STANDARD ERROR APPROXIMATION FROM LORD AND NOVICK'S EQUATION

The underlying assumptions of equation (11.12.4) for estimating the standard error of the mean are that items are sampled exhaustively and without replacement. However, for designs involving sampling of items both with and without replacement, approximate results can be obtained by inflating the size of the item universe, as if each of the item sampled with replacement is a unique item. For example, a multiple matrix sampling design involving T forms, in which k_1 items are with replacement and k_2 items are without replacement; the inflated finite item universe is

$$K = (k_1 + k_2)T.$$

However, it is to be noted that the unique item universe is only $k_1 + k_2T$.

DATA

The data was collected as part of the California Assessment Program involving the <u>Reading Test</u> for grades 2 and 3. The assessment item pool consisted of 212 items in a multiple choice format. The total number of items were divided into ten nearly parallel forms. When assigning items to forms, 12 items, involving oral presentation of the stimuli, were repeated across all ten forms. The remaining 20 items in each form were unique items. The test was administered to 457 second and third grade pupils in a typical California school district according to standardized testing procedures described in the manual. The standard error of the pooled mean were computed using the exact formula as well as approximations to Lord and Novick's formula. For approximate results, for finite item universe, the item universe was taken as 320 instead of 212. The results were computed for finite and infinite item universe as well as finite and infinite population. The results are given in Table 1.

RESULTS AND IMPLICATIONS

The purpose of two methods of computing the standard error is not to show if there are any differences in the two estimates, rather to show how trivial or large are the differences for a typical data set. The results of this investigation show that for data collected using the specified item sampling design, the estimates of the standard error of the mean as computed using approximations from Madow's formulations differ considerably from those computed from approximations of Lord and Novick's equations. If the total error in computing the pooled means is a composite of contributions due to sampling of persons, sampling of items, and an interaction term, the major differences appear in the term representing the error due to the sampling of items. This term is considerably overestimated from approximations of Lord and Novick's equation.

It should be emphasized that Lord and Novick's equation is recommended for use when item sampling designs are based upon sampling of items without replacement. The virtue of using this equation for approximations lies only in its computational simplicity. Based on the findings from this particular item sampling design, however, it is recommended that exact computational procedures be used when item sampling designs involve sampling of items both with and without replacement.

REFERENCES

Lord, F. M. and Novick, M. R. Statistical Theories of Mental Test Scores. Reading, Mass.: Addison-Wesley, 1968.

Madow, W. G. Multiple Matrix Sampling, with Stratification and Possibly Unequal Sample Sizes from the Strata, Pacific Education Evaluation Systems Monograph, 1972.

Pandey, T. N. and Shoemaker, D. M. Estimating Moments of Universe Scores and Associated Standard Errors in Multiple Matrix Sampling for all Item-Scoring Procedures, to appear in Educational and Psychological Measurement, Fall 1975.

Table 1

Comparison of the Standard Errors of the Pooled Mean

Sampling	Exact Equation	Approximation to Lord & Novick's Equation
Grade 2		
N finite, K finite	0.004	0.026
N finite, K infinite	0.014	0.028
N infinite, K finite	0.015	0.030
N infinite, K infinite	0.020	0.031
Grade 3		
N finite, K finite	0.003	0.020
N finite, K infinite	0.011	0.021
N infinite, K finite	0.012	0.023
N infinite, K infinite	0.016	0.024
Grade 2 + Grade 3		
N finite, K finite	0.002	0.021
N finite, K infinite	0.010	0.022
N infinite, K finite	0.010	0.023
N infinite, K infinite	0.014	0.024

Table 2

Number of Pupils Taking Each Form

	FORM 1	FORM 2	FORM 3	FORM 4	FORM 5	FORM 6	FORM 7	FORM 8	FORM 9	FORM 10	TOTAL
Grade 2	24	21	24	22	20	22	22	21	21	19	216
Grade 3	25	23	27	22	27	23	22	24	25	23	241
Grade 2 + Grade 3	49	44	51	44	47	45	44	45	46	42	457