Multiple matrix sampling is a psychometric procedure in which a set of test items is subdivided randomly into subtests of items with each subtest administered to different subgroups of examinees selected at random from the examinee population. Although each examinee receives only a proportion of the complete set of items, the statistical model employed permits the researcher to estimate the mean, variance and frequency distribution of test scores which would have been obtained by testing all examinees on all items. Contained herein is a detailed description of multiple matrix sampling. The topics covered range from an introductory discussion to the listing with expanded writeup of the computer program used to analyze the data. Throughout this Report an attempt has been made to keep the practitioner clearly in mind. (Author)
Principles and Procedures of Multi-Variable Matrix Sampling
PRINCIPLES AND PROCEDURES OF MULTIPLE MATRIX SAMPLING

David M. Shoemaker

ABSTRACT

Multiple matrix sampling is a psychometric procedure in which a set of test items is subdivided randomly into subtests of items with each subtest administered to different subgroups of examinees selected at random from the examinee population. Although each examinee receives only a proportion of the complete set of items, the statistical model employed permits the researcher to estimate the mean, variance and frequency distribution of test scores which would have been obtained by testing all examinees on all items. Contained herein is a detailed description of multiple matrix sampling. The topics covered range from an introductory discussion to the listing with expanded writeup of the computer program used to analyze the data. Throughout this Report an attempt has been made to keep the practitioner clearly in mind.
Contents

Acknowledgments
I. Introduction
II. Characteristics, Advantages, and Applications of Multiple Matrix Sampling
III. Procedural Guidelines in Multiple Matrix Sampling
IV. Computational Formulas in Multiple Matrix Sampling
V. Computer Simulation of Multiple Matrix Sampling
VI. Hypothesis Testing and Multiple Matrix Sampling
VII. Unique Applications of Multiple Matrix Sampling
Bibliography
Appendices
A. Computer Program for Estimating Test Parameters Through Multiple Matrix Sampling
B. Computer Program for Simulating Multiple Matrix Sampling
ACKNOWLEDGMENTS

I would like to thank Frederic M. Lord, Hobart G. Osburn and Howard L. Sullivan for reading in detail all of the manuscript and contributing several improvements. Any blunders, inaccuracies, or awkward passages that remain cannot be laid at their doorsteps. Appreciated to no minor degree is the encouragement given me in this endeavor by my good friend and colleague Masahito Okada. The Southwest Regional Laboratory for Educational Research and Development is to be thanked for the financial support given to several projects upon which portions of this report are based.
Introduction

Multiple matrix sampling or, more popularly, item-examinee sampling, is a psychometric procedure whose time has come. It is the Zeitgeist. Descriptions of multiple matrix sampling procedures and explorations into areas of application are scattered over a multitude of technical journals. There is no single book or article which describes, studies, and unifies all of this material. Yet there is a need for such a document both as a reference source and as a textbook.

Although statisticians have dealt for several decades with incomplete data problems in the design of experiments and data analysis, the psychometrician responsible primarily for the derivation of statistical procedures in multiple matrix sampling and the application of such procedures to problems in psychology and education is Frederic M. Lord. Lord and Novick discuss multiple matrix sampling in Chapter 11 in Statistical theories of mental test scores but the chapter does not encompass the degree of detail and depth of explanation required by the majority of educational research practitioners who desire to implement this research procedure in a particular investigation. This Technical Report has been designed to remedy this situation.

Throughout this Report an attempt has been made to keep the practitioner clearly in mind. The emphasis is on the why, when, and how to use multiple matrix sampling. The topics covered range from an introduction to multiple matrix sampling to the listing with expanded writeup of the computer program used to analyze the data. All discussions and guidelines contained in the monograph reflect theoretical and empirical results reported in the literature as well as personal experiences of the author in implementing multiple matrix sampling in a variety of applied situations.
II

Characteristics, Advantages, And Applications

Of Multiple Matrix Sampling

The majority of contemporary psychometric procedures reflect strongly the original impetus of the psychometric movement, that is, the measurement of individual differences. Historically, individual differences have been investigated, and appropriately so, using the matched-items model in which a single set of test items is administered in a standardized procedure to all, or a sample, of the examinee population under consideration. One exemplar of such methodology is the anthropometric laboratory of Sir Francis Galton established at the International Health Exhibition in England in 1884. Galton measured individuals ranging in age from five to eighty on such dimensions as standing height, sitting height, arm span, weight, breathing capacity and strength of pull "to supply information on the methods, practice, and uses of human measurement." Understandably so and undoubtedly for lack of a reasonable alternative, procedures appropriate for the assessment of individual differences have been transferred completely to investigations concerned primarily with the measurement of group differences. An example of a research design emphasizing the assessment of group differences is found in an investigation which contrasts treatment effects through administering each treatment to a group of examinees selected randomly from the examinee population. Given treatments A, B, and C, for example, the researcher is interested primarily in the behavior of group A as contrasted with group B as contrasted with group C. Differences among individual examinees are of little concern. The point to be made is simply this: the methodology employed successfully in the assessment of individual differences is neither the appropriate nor the most efficient methodology for group assessment. Multiple matrix sampling or, more popularly, item-examinee sampling, has been demonstrated theoretically and empirically to be the appropriate procedure for group assessment and a procedure superior to the matched-items model.

The matched-items model and the multiple matrix sampling model are contrasted readily by considering the data base which would be generated if the entire testable population of N examinees were administered the complete set of K test items. Such a data base is illustrated in Figure 2.1 and the arrangement is referred to commonly as an item-examinee matrix. Test items are scored dichotomously frequently and such is the case in Figure 2.1. For example, examinee 1 passed item 1, failed item 2, passed items 3 and 4, and failed item 5. Within the
Figure 2.1: Item-examinee matrix illustrating examinee-sampling.
### Figure 2.2: Item-examinee matrix illustrating multiple matrix sampling.

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>...</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>...</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Item-examinee sample 1

Item-examinee sample 2

Item-examinee sample 3
framework of the item-examinee matrix, the matched-items model used in
the assessment of individual differences is referred to as the examinee-
sampling model because all test items are administered to a subgroup of
examinees selected at random from the population of N examinees. By
contrast, multiple matrix sampling involves the joint sampling of examinee
subgroups and item subtests as illustrated in Figure 2.2. Data from item-
examinee sample 1 were obtained by administering a set of items selected
at random from the population of K test items and administering these
and only these items to a subgroup of examinees selected randomly from
the population of N examinees. Replicating this procedure produces
item-examinee samples 2 and 3 and suggests, concomitantly, the derivation
of the expression "multiple matrix sampling." Statistics obtained from
examinee-sampling and from multiple matrix sampling are used to estimate
parameters of the N by K item-examinee matrix. It must be remembered,
however, that the N by K item-examinee matrix illustrated in Figures
2.1 and 2.2 is a hypothetical matrix the parameters of which are estimated
from the subset of data gathered in practice through examinee-sampling or
multiple matrix sampling.

Advantages of Multiple Matrix Sampling

A concept important in discussing the advantages of multiple matrix
sampling and one mentioned frequently herein is the standard error of
estimate. Assume that two experimental procedures have been developed
for measuring weight and each procedure is used to obtain in a standard-
ized manner 1000 independent measurements of the weight of a given object.
Hypothetical measurements so acquired have been assembled into frequency
distributions and are given in Figure 2.3. The standard error of estimate
associated with procedure M is the standard deviation of the 1000 values
for the weight obtained using procedure M; the standard error of esti-
matting the weight for procedure E is determined identically. The
difference in standard errors of estimate depicted in Figure 2.3 illus-
trates an important advantage of multiple matrix sampling over examinee-
sampling in group assessment. Lord and Novick (1968) have demonstrated
algebraically that, when subtests are constructed by sampling items
without replacement from the K-item population, the standard error in
estimating the group mean test score using multiple matrix sampling is
less than the standard error obtained with examinee-sampling. Further-
more, the minimum standard error of estimate under multiple matrix
sampling is found by administering one item to each of K random samples
of examinees. A conclusion such as this is of major significance because
the parameter of primary importance in many investigations is the group
mean test score.

To clarify this point, consider how such a result could have been
determined empirically through post mortem item-examinee sampling. In
post mortem item-examinee sampling, an existing N by K item-examinee
data base is taken to be the population of scores and item scores from
item-examinee samples selected randomly from this base are used to
Figure 2.3: Hypothetical distributions of weight measurements resulting from 1000 replications of procedure M and 1000 replications of procedure E.
estimate parameters of interest. Although all examinees have responded to all items, in post mortem item-examinee sampling the investigator acts as if individual examinees had responded only to specific items. The standard error of estimate under examinee-sampling could be approximated, for example, by selecting at random 1000 examinee subgroups and testing each subgroup over \( K \) items. Data from each examinee subgroup provide an estimate of the mean score over \( N \) examinees and the standard deviation calculated over these 1000 estimates is the standard error of estimating the population mean under examinee-sampling. A single estimate of the mean test score under multiple matrix sampling is obtained, for example, by dividing randomly the set of \( K \) test items into \( t \) non-overlapping subtests containing \( K/t \) items each and administering each subtest to a subgroup of examinees selected at random from the population of \( N \) examinees. A single estimate of the population mean is obtained by pooling the \( t \) estimates obtained from each item-examinee sample. Replicating this procedure 1000 times provides 1000 pooled estimates of the population mean test score and hence the standard error of estimate associated with the particular item-examinee sampling plan used. (All computational formulas used in multiple matrix sampling are explained in detail in Chapter IV.)

The advantages of multiple matrix sampling have thus far been focused on the standard error of estimating the mean test score. Important also is the expected value or mean of the estimates of the population mean test score over replications. In Figure 2.3, the standard error of procedure \( M \) is less than the standard error of procedure \( E \); however, on the average, the values obtained using procedure \( E \) are more accurate than those obtained using procedure \( M \) (assuming that the true weight is the value on the abscissa indicated by the pointer). A consideration such as this prompts an examination of the mean estimate of the population mean test score obtained under multiple matrix sampling. The results of several empirical investigations (Johnson & Lord, 1958; Lord, 1962; Plumlee, 1964; Stufflebeam & Cook, 1967; Shoemaker, 1970a, 1970b) using post mortem item-examinee sampling support the conclusion that, on the average, estimates of the mean test score are extremely accurate. (Results such as these are to be expected since the mean of a random sample is always an unbiased estimator of the population mean and estimates of the mean test score obtained through multiple matrix sampling are no exception.) Shoemaker (1970b) has demonstrated that this conclusion is appropriate, additionally, for estimates of the population standard deviation.

In addition to the statistical advantages of multiple matrix sampling in estimating group achievement, there are other advantages of practical import: (a) The testing time per examinee is reduced under multiple matrix sampling. This is, indeed, an important consideration as the time necessary for testing \( K \) items per examinee is frequently difficult or impossible to obtain. (b) Under multiple matrix sampling, the costs of scoring each test are reduced. (c) Multiple matrix sampling as a procedure may be accepted more readily in certain situations than the matched-items design. In a company, for example, supervisors fearing that test results may be used against their employees may be assured more convincingly if
each employee takes only a part of a test and different employees take different parts. (d) Given a limited amount of available testing time per examinee, performance on a larger number of test items can be approximated through multiple matrix sampling than through a matched-items design. (e) With multiple matrix sampling it is possible to estimate simultaneously parameters of several tests. To the examinee, the test so constructed is merely another test; however, to the test constructor, the composite is a collection of several tests each having parameters estimated through multiple matrix sampling.

Limitations of Multiple Matrix Sampling

Although advantages of multiple matrix sampling are more numerous than limitations, the latter do exist. Estimating parameters through multiple matrix sampling assumes that the responses of an examinee to an item sample are exactly those which would have been obtained had the examinee responded to those items embedded in the K-item test. Although the data available (Sirotnik, 1970; Shoemaker, 1970c) suggest that multiple matrix sampling is relatively immune to a context effect, there is one important exception: using multiple matrix sampling to estimate parameters of speeded tests. In this case, an examinee's response is not independent of the context of the test and multiple matrix sampling should not be used.

An insidious variation of the context effect occurs when multiple matrix sampling is used to estimate parameters for a test which is impossible to administer in practice. For example, parameters of a 500-item vocabulary test designed for grade one students could be estimated readily through multiple matrix sampling by forming 25 subtests having 20 items each with each subtest administered to one class of grade one students. Although all students could respond appropriately to the 20-item test, data from each subtest would be used to estimate the results which would have been obtained had all grade one children taken the 500-item test. The problem is that no individual grade one student could have tolerated the 500-item test.

A potentially serious limitation of multiple matrix sampling is found in the logistics involved in giving different tests to different subgroups of examinees. If test items are administered individually, problems are minimal. If, however, each item requires oral instructions by the test administrator and different tests are to be distributed among the examinees in the testing room, serious problems occur. In this situation, the examinees must be segregated and isolated according to subtest before administering each test. If the instructions to each item are written on the test booklet, administering different tests to different examinees within the testing room is accomplished with relative ease.
REFERENCES


Procedural Guidelines in Multiple Matrix Sampling

Multiple matrix sampling as a procedure involves basically three steps: (a) a K-item test is subdivided through random or stratified-random sampling into subtests each having typically the same number of items, (b) each subtest is administered to a group of examinees selected randomly from the examinee population, and (c) test parameters are estimated from subtest results. Although the procedure is described easily, implementing it produces many interesting questions. For example: How many subtests should be formed? To how many examinees should each subtest be administered? Is it more appropriate to administer a few subtests containing a large number of items or a large number of subtests containing few items? These are only a few of the questions encountered frequently when using multiple matrix sampling. Described herein are general guidelines for answering these and other related questions.

Let $t$ denote the number of subtests, $k$ the number of items per subtest and $n$ the number of examinees to which each subtest is administered. A specific sampling plan is denoted by $(t/k/n)$. For example, $(2/25/60)$, $(10/5/60)$ and $(10/20/30)$ are three sampling plans which could be used to estimate the parameters of a 50-item test. With the first plan, 2 subtests are formed containing 25 items each with each subtest administered to 60 examinees; with the second plan, 10 tests with 5 items each with each subtest administered to 60 examinees; and with the third, 10 tests with 20 items each with each subtest administered to 30 examinees. The third plan introduces an important variable in multiple matrix sampling, namely, the procedure used to sample items in constructing subtests. With $(2/25/60)$ and $(10/5/60)$ subtests are formed by sampling test items without replacement from the pool of 50 items. With $(10/20/30)$, items are sampled without replacement for a given subtest but with replacement among subtests; consequently, an individual item will often be included in more than one subtest, but no item will be included twice in the same subtest. The rule is this: if the product $tk$ is less than or equal to $K$, the sampling of items for subtests is always without replacement; when $tk$ is greater than $K$, the sampling of items is without replacement for each subtest and with replacement between subtests. Selecting items for two subtests using the latter sampling procedure is demonstrated easily with a deck of cards numbered consecutively from 1 to $K$: (a) the deck of $K$ cards is shuffled thoroughly, (b) $k$ cards are selected at random from the deck with the numbers on the cards indicating those items to be included in subtest $i$, (c) the $k$ cards are returned to the deck, (d) the card deck is reshuffled, and
(e) k cards are selected at random for subtest j. Although a multitude of sampling plans are possible, it is generally the case that tk is equal to or greater than K.

Although constructing subtests having overlapping item subsets is desirable in that it increases the number of observations acquired by the sampling plan (and, hence, decreases generally the standard error of estimate associated with that sampling plan), it is of critical importance that, when tk is greater than K, tk be an integer multiple of K, and items are sampled randomly but subject to the restriction that each item appear with equal frequency among subtests. With (10/20/30), for example, the multiple is 4 and each of the 50 items should appear in exactly 4 subtests. Any deviation from this procedure results in a marked increase in the standard error of estimate.

An important characteristic of any sampling plan used in multiple matrix sampling is the number of observations acquired by that plan. Defining one observation as the score received by one examinee on one item, the number of observations acquired by a sampling plan is equal to the product tkn. For example, 3000 observations are acquired by (2/25/60) and by (10/5/60) while 6000 observations are acquired by (10/20/30). The number of observations per sampling plan is an important concept in multiple matrix sampling and one mentioned frequently herein.

In multiple matrix sampling, a variety of sampling plans are possible with the selection of a particular sampling plan being typically the result of both practical and statistical considerations. Determining the relative merits of individual sampling plans is accomplished readily through a consideration of the standard error of estimate for each parameter for each sampling plan. Shoemaker (1970a, 1970b, 1971a, 1971b) has determined empirically, through post mortem item-examinee sampling, standard errors of estimate for selected parameters as a function of variations in (a) the number of observations acquired by the sampling plan, (b) t, k, and n, (c) test reliability of the normative distribution of test scores, (d) the variance of item difficulty indices, and (e) degree of skewness in the normative test score distribution. The following are general guidelines in multiple matrix sampling resulting from these and other investigations (Shoemaker & Osburn, 1968; Osburn, 1969):

1. The number of observations acquired by the sampling plan is an important variable. In general, as the number of observations increases, the standard error of estimating parameters decreases. (The major exception to this guideline occurs when guideline 4 is not followed.)

2. Increasing the number of examinees per subgroup is least effective in reducing the standard error of estimate.
3. For normal normative distributions, increases in the number of items per subtest are most effective in reducing standard errors of estimate; for negatively-skewed distributions, increases in the number of subtests are most effective.

4. When \( t_k \) is greater than \( K \), \( t_k \) should be an integer multiple of \( K \) and items should be selected randomly but subject to the restriction that among subtests each item appears with equal frequency.

5. In general, fewer observations are required to estimate parameters of a skewed normative distribution than of a normal normative distribution.

6. If subtest items are being selected according to a stratified-random sampling plan instead of a random sampling plan, items should be stratified according to difficulty level and not according to content.

7. As the reliability of the normative distribution of test scores increases, it becomes increasingly difficult to estimate parameters. For this reason, it is true generally that a relatively large number of observations is required by the sampling plan when estimating parameters of a distribution having high reliability. This is true also when the variance of item difficulty indices is large.

8. If no information concerning the normative distribution of test scores is available, select a sampling plan having the number of subtests equal to the square root of the total number of test items (rounded to the nearest integer) with each subtest having approximately the same number of test items.

Guidelines such as these are concerned primarily with relative standard errors of estimate in multiple matrix sampling. Although Lord and Novick (1968, equation 11.12.3) have determined algebraically the standard error of estimating the mean proportion correct score in multiple matrix sampling given nonoverlapping random samples of dichotomously-scored items drawn without replacement from the item population, the standard error of estimate for any parameter using any and all sampling plans may be determined easily and effectively through use of the simulation model for multiple matrix sampling described in detail in Chapter V.

**Multiple Matrix Sampling Step by Step**

Step 1: Construct or select the \( K \)-item test. If possible, assemble the items into strata according to difficulty level.

Step 2: Determine the limitations and restrictions which must be imposed upon the test administration procedure.
Step 3: Select a sampling plan which is appropriate in view of the known characteristics of the normative distribution, the restrictions and limitations inherent in the test administration procedure, and guidelines 1 through 8.

Step 4: Administer subtests to examinees in a standardized procedure. Avoid confounding subtests with examinee subgroups, i.e., make every attempt to have examinee subgroups homogeneous.

Step 5: Compute estimates of parameters using equations 4.1, 4.2, 4.4, 4.5, 4.7 and 4.9 with the computer program given in Appendix A.
REFERENCES


Shoemaker, D. M. Standard errors of estimate in item-examinee sampling as a function of test reliability, variation in item difficulty indices and degree of skewness in the normative distribution. Unpublished manuscript, 1971. (b)
Computational Formulas in Multiple Matrix Sampling

Computational formulas used in multiple matrix sampling are applied easily in practice and are detailed and sequenced appropriately in the following application of the procedure. It should be noted initially that all formulas assume uniform item scoring procedures; for example, some items cannot be scored dichotomously and other trichotomously.

An Application of Multiple Matrix Sampling

A spelling program is being designed for kindergarten students and the word and rule content of this program is to be related closely to the reading program used by these students. Before constructing such a program it is necessary to determine the spelling proficiency of those students who have used the reading program but have not had formal spelling instruction on the related words. Although there were 78 unique words introduced in the particular reading program under consideration, technical considerations dictated that only words having regular spellings be included in the word population. As a result, the original word population was reduced from 78 words to 50 words. The modified word population was then subdivided through random sampling without replacement into 5 subtests containing 10 words each. (This is one of many procedures which could have been used. Alternative procedures are discussed in detail in Chapter III.)

Three kindergarten classes were selected randomly from the pool of 9 classes. Students within each class were divided at random into 5 groups and each group was assigned at random to one of the 5 subtests. Each test was administered individually. All items were scored dichotomously (1 = pass, 0 = fail) with the results of each subtest given in Tables 4.1 through 4.5.

Estimating Parameters From Subtest Results

In multiple matrix sampling, subtest results are of secondary interest. Of chief concern is the estimation of parameters, that is, the results which would have been obtained had all students been tested over the entire set of 50 items comprising the word population. The results of each subtest, however, can be used to provide estimates of parameters of interest. For example, from subtest 1 it is possible to
obtain an estimate of several parameters, i.e., $\mu$ (the population mean test score), $\sigma$ (the standard deviation of test scores), $\sigma^2$ (the variance of test scores), $\mu_3$ (the third moment about the arithmetic mean), $\mu_4$ (the fourth moment about the arithmetic mean), $\alpha_{21}$ (the coefficient of reliability), $g_1$ (the index of skewness), and $g_2$ (the degree of kurtosis). All of these parameters are not independent, but each can be estimated from the results of one subtest. In multiple matrix sampling, multiple subtests are used and, hence, multiple estimates of each parameter are obtained. A more accurate estimate of each parameter is obtained by combining or pooling the estimates obtained from each subtest.

Although it is possible to estimate several parameters, the majority of investigations are interested primarily in estimating $\mu$, $\sigma^2$, and $\alpha_{21}$. The appropriate formulas for estimating these parameters from subtest $i$ are

\[
\hat{\mu}_i = \frac{k_i T_i}{k_i}, \quad (4.1)
\]

\[
\hat{\sigma}_i^2 = \frac{n_i k_i (K - 1) s_i^2 - (K - k_i \sum v_i)}{k_i (k_i - 1)(n_i - 1)}, \quad (4.2)
\]

and,

\[
\hat{\alpha}_{21_i} = \frac{K}{k_i - 1} \left[ 1 - \left( \frac{\hat{\mu}_i}{\mu_i} - \frac{\hat{\sigma}_i^2}{\sigma_i^2} \right) \right], \quad (4.3)
\]

where,

$K =$ the total number of items in the population,

$k_i =$ the number of items in subtest $i$,

$n_i =$ the number of examinees receiving subtest $i$,
Table 4.1

Results And Computations For Subtest 1

<table>
<thead>
<tr>
<th>Examinee</th>
<th>Class</th>
<th>Item</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>02</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>03</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>04</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>05</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>06</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>07</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>08</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>09</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>13</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>15</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>16</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>17</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>18</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

-17-
Table 4.2
Results And Computations For Subtest 2

<table>
<thead>
<tr>
<th>Examinee</th>
<th>Class</th>
<th>Item</th>
<th>08</th>
<th>12</th>
<th>18</th>
<th>23</th>
<th>30</th>
<th>33</th>
<th>38</th>
<th>40</th>
<th>43</th>
<th>46</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>02</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>03</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>04</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>05</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>06</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>07</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>08</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>09</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>10</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>11</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>12</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>13</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>14</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
</tbody>
</table>

\[ p = 0.8571, p(1-p) = 0.1225 \]

\[ k = 50, k_2 = 10, n_2 = 14 \]

\[ s_2^2 = 3.0612, s_2 = 1.7523, \mu_2 = \frac{(50)(5.2857)}{10} = 26.4285 \]

\[ \Sigma T = 74, \Sigma p(1-p) = 1.5001, \Sigma a = 74 \]

\[ \Sigma a^2 = 434, \sigma_2^2 = \frac{(14)(50)((50 - 1)(3.0612) - (50 - 10)(1.5001))}{10(10 - 1)(14 - 1)} = 53.8430 \]

\[ \bar{T}_2 = 5.2857 \]
Table 4.3
Results And Computations For Subtest 3

<table>
<thead>
<tr>
<th>Examinee</th>
<th>Class</th>
<th>Item 01</th>
<th>Item 05</th>
<th>Item 06</th>
<th>Item 14</th>
<th>Item 15</th>
<th>Item 20</th>
<th>Item 25</th>
<th>Item 26</th>
<th>Item 29</th>
<th>Item 32</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>02</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>03</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>04</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>05</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>06</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>07</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>08</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>09</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>10</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>11</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>12</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

\[ p = 0.6154 \quad 0.6154 \quad 0.3846 \quad 0.6923 \quad 0.3846 \quad 0.5385 \quad 0.4615 \quad 0.6154 \quad 0.1538 \quad 0.2308 \]

\[ p(1 - p) = 0.2367 \quad 0.2367 \quad 0.2367 \quad 0.2130 \quad 0.2367 \quad 0.2485 \quad 0.2485 \quad 0.2367 \quad 0.1301 \quad 0.1775 \]

\[ K = 50 \]

\[ k_3 = 10 \]

\[ n_3 = 13 \]

\[ \sum p(1-p) = 2.2011 \]

\[ \sum \bar{X} = 61 \]

\[ \sum n^2 = 411 \]

\[ \bar{T}_3 = 4.6923 \]

\[ s_3 = 9.5976 \]

\[ k_3 = 10 \]

\[ n_3 = 13 \]

\[ \sum p(1-p) = 2.2011 \]

\[ \bar{T}_3 = 4.6923 \]

\[ s_3 = 9.5976 \]
Table 4.4
Results And Computations For Subtest 4

<table>
<thead>
<tr>
<th>Examinee</th>
<th>Class</th>
<th>Item</th>
<th>02</th>
<th>04</th>
<th>07</th>
<th>10</th>
<th>19</th>
<th>24</th>
<th>31</th>
<th>39</th>
<th>41</th>
<th>47</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>02</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>03</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>04</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>05</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>06</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>07</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>08</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>09</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>10</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>11</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>12</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>13</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

\[
p = 0.4615, 0.6154, 0.7692, 0.6923, 0.0769, 0.3846, 0.7692, 0.4615, 0.3846, 0.3077
\]

\[
p(1 - p) = 0.2485, 0.2367, 0.1775, 0.2130, 0.0710, 0.2367, 0.1775, 0.2485, 0.2367, 0.2130
\]

\[
K = 50
\]

\[
k_4 = 10
\]

\[
\mu_4 = \frac{(50)(4.9231)}{10} = 24.6155
\]

\[
\sigma_4^2 = \frac{(13)(50)((50 - 1)(10.6864) - (50 - 10)(2.0591))}{10(10 - 1)(13 - 1)} = 265.5789
\]

\[
\sum T = 64
\]

\[
\overline{T_4} = 4.9231
\]
Table 4.5
Results And Computations For Subtest 5

<table>
<thead>
<tr>
<th>Item</th>
<th>Examinee</th>
<th>Class</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>01</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>02</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>03</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>04</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>05</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>06</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>07</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>08</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>09</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Item</th>
<th>13</th>
<th>17</th>
<th>22</th>
<th>28</th>
<th>34</th>
<th>35</th>
<th>36</th>
<th>45</th>
<th>48</th>
<th>49</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

\[
p = 0.3333, \quad p(1 - p) = 0.2222
\]

\[
\begin{align*}
K &= 50 \\
k_5 &= 10 \\
\sigma^2 &= 6.9097 \\
\bar{\sigma}^2 &= 9.5835 \\
\bar{\sigma}^2 &= 169.3378 \\
\bar{T}_5 &= 1.9167
\end{align*}
\]
\( \bar{T}_i = \) the mean test score on subtest \( i \),

\( s^2 = \frac{n_i}{k_i} \), the variance of test scores on subtest \( i \),

and

\( \sum k_i \)

\( \sum v_i = \) the sum of the \( k_i \) item variances in subtest \( i \).

If items are scored dichotomously, the variance of item \( j \) is equal to \( p_j(1 - p_j) \) where \( p_j \) is the proportion of examinees answering item \( j \) correctly.

The computational formula for \( \hat{o^2} \) was derived from an associated formula given by Sirotnik (1970) in which it was assumed that the number of examinees and number of items in the population were both finite. Formula 4.2 is based on the assumption that the number of examinees in the population is infinite and that the number of items in the population is finite.

The results of each subtest provide an estimate of \( \mu \) and \( \sigma^2 \) and a pooled estimate of \( \mu \) and \( \sigma^2 \) is obtained by combining the \( t \) subtest estimates using

\[ \hat{\mu}_{\text{pooled}} = \frac{\sum O_i \hat{\mu}_i}{\sum O_i} \]  

(4.4)

and

\[ \hat{\sigma^2}_{\text{pooled}} = \frac{\sum O_i \hat{\sigma^2}_i}{\sum O_i} \]  

(4.5)

where,

\[ O_i = n_i k_i \]  

(4.6)
the number of observations obtained from subtest \( i \). If the total number of examinees \( \sum_{i} N_{i} = N \) is less than 500, \( \hat{\sigma}^{2}_{\text{pooled}} \) should be multiplied by \( (N - 1)/N \). Pooled estimates of \( \mu \) and \( \sigma^{2} \) for the word spelling project are given in Table 4.6. The pooled estimate of the mean test score on the 50-item test is 20.4287. On the basis of this result, the conclusion was made that kindergarten students can spell correctly approximately 40 per cent of words having regular spelling in the reading program without having had any formal spelling instruction.

Although individual estimates of the reliability of the 50-item test could have been obtained from each subtest and then combined into a single estimate, a simpler procedure for estimating \( \alpha_{21} \) is one using the pooled estimates of \( \mu \) and \( \sigma^{2} \). Specifically,

\[
\hat{\alpha}_{21} = \frac{\hat{\mu}_{\text{pooled}}}{\hat{\sigma}^{2}_{\text{pooled}}} = \frac{\hat{\mu}_{\text{pooled}}}{\hat{\sigma}^{2}_{\text{pooled}}} \left[ 1 - \frac{\hat{\mu}_{\text{pooled}}}{K - 1} \right].
\] (4.7)

For the word spelling test \( \hat{\alpha}_{21} \) for the 50-item test was estimated from 4.7 to be .9479. The exact computations are given in Table 4.7 where \( \hat{\alpha}_{21} \) is computed as an intermediate step in approximating the normative test score distribution with a probability distribution.

Direct Calculation of \( SE(\hat{\mu}_{\text{pooled}}) \)

A more meaningful interpretation of \( \hat{\mu}_{\text{pooled}} \) is possible if \( SE(\hat{\mu}_{\text{pooled}}) \) is known. Although \( SE(\hat{\mu}_{\text{pooled}}) \) and \( SE(\hat{\sigma}^{2}_{\text{pooled}}) \) may be determined for all sampling plans through use of the simulation model described in Chapter V, Lord and Novick (1968, equation 11.12.3) have derived an equation for determining the standard error of the mean proportion correct score given (a) items are scored dichotomously, (b) items are sampled randomly and without replacement from the item population, and (c) examinees are sampled randomly and without replacement from the examinee population. Restrictions (b) and (c) produce item subsets and examinee subgroups which are nonoverlapping, i.e., no item is found in more than one subtest and no examinee
<table>
<thead>
<tr>
<th>Subtest</th>
<th>Number of Observations</th>
<th>n</th>
<th>k</th>
<th>( \hat{\mu} )</th>
<th>( \hat{\sigma}^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>180</td>
<td>18</td>
<td>10</td>
<td>17.7780</td>
<td>158.1778</td>
</tr>
<tr>
<td>2</td>
<td>140</td>
<td>14</td>
<td>10</td>
<td>26.4285</td>
<td>53.8430</td>
</tr>
<tr>
<td>3</td>
<td>130</td>
<td>13</td>
<td>10</td>
<td>23.4615</td>
<td>230.0509</td>
</tr>
<tr>
<td>4</td>
<td>130</td>
<td>13</td>
<td>10</td>
<td>24.6155</td>
<td>265.5789</td>
</tr>
<tr>
<td>5</td>
<td>120</td>
<td>12</td>
<td>10</td>
<td>9.5835</td>
<td>169.3378</td>
</tr>
<tr>
<td></td>
<td>700</td>
<td>70</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[
\hat{\mu}_{\text{pooled}} = \frac{(180)(17.7780) + (140)(26.4285) + \ldots + (120)(9.5835)}{180 + 140 + \ldots + 120}
\]

\[= 20.4287\]

\[
\hat{\sigma}^2_{\text{pooled}} = \frac{(180)(158.1778) + (140)(53.8430) + \ldots + (120)(169.3378)}{180 + 140 + \ldots + 120}
\]

\[= 172.5178\]

\[N \leq 500\]

\[= 172.5178 \left( (70 - 1)/70 \right)\]

\[= 170.0533\]
is found in more than one subgroup. Equation 11.12.3 when modified to give the standard error of the mean test score is

\[
\text{VAR}(\hat{\mu}_{\text{pooled}}) = \frac{1}{tkn} \left[ \frac{1}{(K-1)(N-1)} \right] \\
\left[ \frac{K^2N\sigma_p^2}{\sigma^2} \{(K-k)(n-1) - kn(t-1)\} + \right. \\
\left. K\sigma^2 \{(N-n)(k-1) - kn(t-1)\} + \right. \\
\hat{\mu}(K - \hat{\mu}) \{(K-k)(N-n) + kn(t-1)\} \right],
\]

where \(K\) refers to the total number of test items, \(N\) to the total number of examinees, \(\sigma^2\) to the population variance, \(\sigma_p^2\) to the variance of item difficulty indices, and \(\hat{\mu}\) to the estimate of the population mean obtained from multiple matrix sampling.

In practice, \(\sigma^2\) and \(\sigma_p^2\) are estimated; \(t, k,\) and \(n\) are parameters defining the sampling plan. Of course, \(\text{SE}(\hat{\mu}_{\text{pooled}}) = \text{VAR}(\hat{\mu}_{\text{pooled}})^{1/2}\)

No equation is given by Lord and Novick for \(\text{SE}(\hat{\sigma}_{\text{pooled}})\) under multiple matrix sampling.

\section*{Approximating the Normative Distribution}

In addition to estimating individual parameters through multiple matrix sampling, it is possible to estimate the entire normative frequency distribution of test scores which would have been obtained by testing all students on all 50 items. The negative hypergeometric distribution has been shown by Keats and Lord (1962) to provide a reasonably good fit for a wide variety of test score distributions when the test score is the number of correct responses. The negative hypergeometric distribution is a function of the mean test score \(\hat{\mu}\), the variance of the test scores \(\sigma^2\) and the total number of items in the test \(K\). Lord (1962) and Shoemaker (1970) have demonstrated the negative hypergeometric distribution
with parameters estimated by multiple matrix sampling can be used satisfactorily to approximate normative distributions of number correct test scores. The formula for the negative hypergeometric distribution is

\[
h(T) = \frac{c(-K)_T (a)_T}{(-b)_T T!} \text{ for } T = 0, 1, 2, \ldots, K
\]

(4.9)

where,

\[
a = (-1 + 1/\hat{\sigma}_{21})_{\text{pooled}}
\]

\[
b = -a - 1 + K/\hat{\sigma}_{21}
\]

\[
c = \frac{b^K}{(a + b)^K}
\]

noting that,

\[
b^K = b(b - 1)(b - 2) \ldots (b - K + 1)
\]

\[
(a)_T = a(a + 1)(a + 2) \ldots (a + T - 1)
\]

\[
(a)_0 = b^0 = 1
\]

\[
T! = T(T - 1)(T - 2) \ldots (2)(1).
\]

Using estimates of \( \mu \) and \( \sigma^2 \) obtained from the word spelling project, the calculations necessary for approximating the normative
distribution on the 50-item test with the negative hypergeometric distribution are illustrated in Table 4.7 with complete results given in Table 4.9. The computations involved in estimating \( \mu \) and \( \sigma^2 \) and approximating the normative distribution by the negative hypergeometric distribution are more laborious than difficult. A computer program has been developed which performs all the necessary computations and output for the word spelling project is given in Tables 4.8 and 4.9. A detailed writeup and listing of the computer program is given in Appendix A.

An examination of the estimates of parameters given in Table 4.8 suggests that individual subtests were not equally difficult, particularly subtest 5. Although the words included in subtest 5 were selected randomly from the 50-word population and administered to subgroups of examinees selected at random from each class, the results merely confirm the well-known fact that extreme cases do occur through random sampling. An obvious advantage, then, of multiple matrix sampling over any individual item-examinee sample is that the estimates obtained in the former case are based on a composite and hence less subject to sampling extremities. Stated more precisely, the standard error associated with the pooled estimate of the mean test score is less than the standard error associated with any of the estimates of the mean obtained from subtests. The results for \( \mu \) and \( \sigma \) given in Chapter V illustrate adequately the difference in standard errors of estimate described here.

The relative frequencies given in Table 4.9 are actually the individual probabilities associated with all possible test scores. For example, the probability of an examinee spelling correctly 20 words out of 50 is .023. An equally appropriate interpretation is that 2.3 percent of the examinees in the population would spell correctly 20 words. As should be the case, the relative frequencies in Table 4.9 sum to unity. An estimate of the number of examinees receiving each test score is obtained by multiplying the total number of examinees in the population by the probability associated with each test score. For example, if there were 1000 students in the population of kindergarteners, 23 students would be expected to spell correctly 20 of the 50 words on the test.

Although equations 4.1 and 4.2 are appropriate for all item scoring procedures, the negative hypergeometric distribution is used only when the test score is the number of correct answers. This is, of course, the case when items are scored \( 1 = \text{pass} \) and \( 0 = \text{fail} \). When items are not scored dichotomously, the normative frequency distribution may be approximated by a Pearson curve using the first moment about the origin and the second, third and fourth moments about the mean. There are 12 curves in the family of Pearson curves and the procedure for selecting the appropriate curve and making the necessary calculations to approximate the normative distribution are given by Elderton (1938, pp. 38-127) and by Kendall (1952, pp. 137-145). Lord (1960) has suggested that a Pearson Type I curve may be an appropriate selection. It should be mentioned, however, that such a procedure is not a casual undertaking. Before such procedures can be used, computational formulas for estimating \( \mu_3 \) and \( \mu_4 \) must be derived. Guidelines for estimating these moments are given by Hooke (1956).
Table 4.7
Computations For Negative Hypergeometric Distribution

\[ a_{21} = \frac{50}{50 - 1} \times \left(1 - \frac{20.4286^2}{50}/170.0552\right) = .9479 \]

\[ a = (-1 + 1/.9479)(20.4287) = 1.1226 \]

\[ b = -1.1226 - 1 + 50/.9479 = 50.6250 \]

\[ c = \frac{50.6250^{[50]}}{51.7476^{[50]}} = \frac{50.6250(50.6250-1)(50.6250-2) \cdots (50.6250-49)}{51.7476(51.7476-1)(51.7476-2) \cdots (51.7476-49)} = .0214 \]

\[ h(0) = (.0214) \frac{(-50)_0(1.1226)_0}{(-50.6250)_0 0!} = (.0214) \frac{(1)(1)}{(1)(1)} = .0214 \]

\[ h(1) = (.0214) \frac{(-50)_1(1.1226)_1}{(-50.6250)_1 1!} = (.0214) \frac{(-50)(1.1226)}{(-50.6250)(1)} = .0237 \]

\[ h(2) = (.0214) \frac{(-50)_2(1.1226)_2}{(-50.6250)_2 2!} = (.0214) \frac{(-50)(-49)(1.1226)(2.1226)}{(-50.6250)(-49.6250)(2)} = .0248 \]
Table 4.8
Estimates Of Parameters For Word Spelling Project

<table>
<thead>
<tr>
<th>Sample</th>
<th>Estimate Of Parameter</th>
<th>Mean</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>17.7777770</td>
<td>158.1844600</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>26.4285710</td>
<td>53.8461540</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>23.4615380</td>
<td>230.0498600</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>24.6153840</td>
<td>265.5769300</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>9.5833331</td>
<td>169.3392200</td>
</tr>
</tbody>
</table>

POOLED MEAN = 20.4285710
POOLED VARIANCE = 170.0552200
Table 4.9
Estimated Relative Frequency Per Test Score On The 50-Item Test Using
The Negative Hypergeometric Distribution

<table>
<thead>
<tr>
<th>Score</th>
<th>Relative Frequency</th>
<th>Score</th>
<th>Relative Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>.0213564</td>
<td>26</td>
<td>.0216387</td>
</tr>
<tr>
<td>1</td>
<td>.0236785</td>
<td>27</td>
<td>.0211852</td>
</tr>
<tr>
<td>2</td>
<td>.0248133</td>
<td>28</td>
<td>.0207151</td>
</tr>
<tr>
<td>3</td>
<td>.0254953</td>
<td>29</td>
<td>.0202280</td>
</tr>
<tr>
<td>4</td>
<td>.0259319</td>
<td>30</td>
<td>.0197236</td>
</tr>
<tr>
<td>5</td>
<td>.0262115</td>
<td>31</td>
<td>.0192016</td>
</tr>
<tr>
<td>6</td>
<td>.0263806</td>
<td>32</td>
<td>.0186613</td>
</tr>
<tr>
<td>7</td>
<td>.0264667</td>
<td>33</td>
<td>.0181020</td>
</tr>
<tr>
<td>8</td>
<td>.0264872</td>
<td>34</td>
<td>.0175230</td>
</tr>
<tr>
<td>9</td>
<td>.0264544</td>
<td>35</td>
<td>.0169233</td>
</tr>
<tr>
<td>10</td>
<td>.0263766</td>
<td>36</td>
<td>.0163017</td>
</tr>
<tr>
<td>11</td>
<td>.0262602</td>
<td>37</td>
<td>.0156567</td>
</tr>
<tr>
<td>12</td>
<td>.0261101</td>
<td>38</td>
<td>.0149867</td>
</tr>
<tr>
<td>13</td>
<td>.0259298</td>
<td>39</td>
<td>.0142896</td>
</tr>
<tr>
<td>14</td>
<td>.0257223</td>
<td>40</td>
<td>.0135627</td>
</tr>
<tr>
<td>15</td>
<td>.0254900</td>
<td>41</td>
<td>.0128031</td>
</tr>
<tr>
<td>16</td>
<td>.0252347</td>
<td>42</td>
<td>.0120066</td>
</tr>
<tr>
<td>17</td>
<td>.0249578</td>
<td>43</td>
<td>.0111683</td>
</tr>
<tr>
<td>18</td>
<td>.0246608</td>
<td>44</td>
<td>.0102815</td>
</tr>
<tr>
<td>19</td>
<td>.0243444</td>
<td>45</td>
<td>.0093369</td>
</tr>
<tr>
<td>20</td>
<td>.0240095</td>
<td>46</td>
<td>.0083215</td>
</tr>
<tr>
<td>21</td>
<td>.0236568</td>
<td>47</td>
<td>.0072158</td>
</tr>
<tr>
<td>22</td>
<td>.0232868</td>
<td>48</td>
<td>.0059869</td>
</tr>
<tr>
<td>23</td>
<td>.0228997</td>
<td>49</td>
<td>.0045728</td>
</tr>
<tr>
<td>24</td>
<td>.0224960</td>
<td>50</td>
<td>.0028209</td>
</tr>
<tr>
<td>25</td>
<td>.0220756</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Computational Irregularities

In estimating parameters from subtests having a small number of items and examinees, it happens frequently that \( \hat{\sigma}^2 \) is equal to zero or is less than zero for one or more subtests. Although uninterpretable, estimates such as these should not be discarded or set equal to zero in computing \( \hat{\sigma}^2 \). It must be remembered that results of any subtest pooled are relatively unimportant; what is important is the accuracy of the pooled estimate of \( \sigma^2 \). Any procedure which ignores part of the data produces an estimate of \( \sigma^2 \) which is biased, i.e., it would not approach the true value even if the number of subtests was increased indefinitely. Sirotnik (1970) has verified empirically this conclusion.
REFERENCES


V

Computer Simulation of Multiple Matrix Sampling

In evaluating a particular sampling plan or contrasting the relative merits of several plans used in multiple matrix sampling, statistics of primary importance are the standard error of estimate and the mean estimate for each parameter given that sampling plan. For example, if an investigator were estimating parameters of a 50-item test using multiple matrix sampling, one sampling plan might be (5/20/30); another, (10/10/60). In the first sampling plan, the 50-item test is subdivided through random sampling without replacement within subtests and with replacement between subtests into 5 subtests containing 20 items each with each subtest administered to 30 examinees; in the second plan, similarly, 10 subtests containing 10 items each with each subtest administered to 60 examinees.

The sampling plan selected will be used only once in an investigation; yet, in selecting the particular plan to be used, the investigator must be aware of the standard error of estimate associated with each sampling plan under consideration. Lord and Novick (1968, equation 11.12.3) have derived algebraically the standard error of estimating the mean proportion correct score given nonoverlapping random samples of dichotomously-scored items drawn without replacement from the item population. No comparable equation is given by them for computing the standard error of estimating the population standard deviation under multiple matrix sampling. What is required, however, are equations for estimating standard errors of estimate per parameter for all potentially useful sampling plans, not just those plans involving nonoverlapping random samples of items from the item population. The computational difficulties in such a task are not minor; however, the results of such equations are approximated readily and to any desired degree of accuracy through the computer simulation model described herein. The remaining sections of this chapter are devoted to a detailed discussion of a simulation model for multiple matrix sampling. The reader uninterested in such matters can bypass safely this chapter without a loss of continuity. However, several of the guidelines for multiple matrix sampling given in Chapter III are based on results obtained through use of this model and, it must be stressed, that the results obtained are only as good as the simulation model used.
Simulated Post Mortem Sampling

The algorithm used within the model is described most appropriately as simulated post mortem multiple matrix sampling. In post mortem sampling, item-examinee samples are taken from an N by K data base obtained by testing N examinees over K items; in simulated post mortem sampling, the N by K data base is computer-generated by a simulation model. Generating data bases with prescribed parameters is essential in investigating hypotheses in multiple matrix sampling because it is difficult, if not impossible, to locate existing data bases having the necessary variation in test parameters. For example, if the standard error of estimate were being investigated as a function of variation in item difficulty indices for a given test reliability and test length, it would be difficult locating data bases with \( \sigma^2_p = .00, .05, \) and .08 all having \( \alpha_{20} = .80 \) and the same test length. Such a problem is, however, handled easily with a simulation model. As an overview, the computer program generates a data base, selects multiple item-examinee samples from this data base, performs all calculations necessary for estimating parameters, and replicates this procedure as many times as specified before computing the standard error of estimate and mean estimate per parameter over replications. The computer program is restricted to data bases having dichotomously-scored items and, in multiple matrix sampling, to subtests having an equal number of items and examinees.

Generation of Data Bases

In simulating multiple matrix sampling, generation of the data base is of primary importance. Although one procedure might be that of generating an N by K matrix and storing it in memory, a more appropriate procedure is one in which the item scores on the K-item test are generated for one and only one individual at a time. All that is stored in memory are the K item scores for one individual. The procedure, however, for generating item scores must be one such that, over any number of hypothetical examinees generated, the items and test scores have prescribed characteristics. In this procedure, the population of examinees N is countably infinite. The test parameters subject to manipulation within the program are: (a) K, the number of items in the item population, (b) \( \mu \), the mean test score over examinees, (c) \( \sigma^2 \), the variance of test scores over examinees, (d) \( \alpha_{20} \), the coefficient of reliability for the K-item test, (e) \( \sigma^2_p \), the variance of the item difficulty indices, where, the difficulty index \( p_i \) for item i is the proportion of examinees answering correctly item i, and (f) the degree of skewness in the distribution of test scores for examinees on the K-item test. In the computer program, values for K, \( \mu \) and \( \sigma^2 \) must be specified by the user. The maximum value for K is 150; \( \mu \) is, therefore, restricted to values \( 0 < \mu < K \). If \( \alpha_{20} \) is specified, \( \sigma^2 \) is determined by the well-known relationship
derived originally by Tucker (1949). If \( \sigma^2 \) is specified by the user, \( \alpha_{20} \) is determined consequently. Such an arrangement has been incorporated within the program to facilitate hypothesis testing where either \( \sigma^2 \) or \( \alpha_{20} \) is to be controlled across levels of \( K \). Of course, \( \bar{p} = \mu/K \) is determined once \( \mu \) has been specified. The degree of skewness in the normative distribution is simulated by using the lognormal or normal probability distribution functions to generate test score distributions. The lognormal distribution with two parameters is used to generate positively-skewed test score distributions while the three parameter lognormal distribution is used for negatively-skewed distributions. The lognormal distribution is described in detail by Aitchison and Brown (1957) and a detailed explanation of simulating stochastic variates with the lognormal distribution is given by Naylor, Balintfy, Burdick and Chu (1966). The normal density function is, of course, used to simulate normal test score distributions. Density functions for the two and three parameter lognormal probability distributions are, respectively,

\[
\Lambda (T | \mu, \sigma^2) = \frac{1}{T \sigma \sqrt{2\pi}} \exp \left[ -\frac{(\ln(T) - \mu)^2}{2\sigma^2} \right] \tag{5.2}
\]

\[
\Lambda (T' | T'=K-T, \mu, \sigma^2) = \frac{1}{T \sigma \sqrt{2\pi}} \exp \left[ -\frac{(\ln(T) - \mu)^2}{2\sigma^2} \right] \tag{5.3}
\]

for \( T = 0, 1, 2, \ldots, K \).

For the normal distribution, the density function is

\[
N(T|\mu, \sigma^2) = \frac{1}{\sigma \sqrt{2\pi}} \exp \left[ -\frac{(T - \mu)^2}{2\sigma^2} \right]. \tag{5.4}
\]
The constants \( \mu \) and \( \sigma^2 \) in 5.4 are equal, respectively, to the desired mean and variance in the normative distribution; however, in 5.2 and 5.3, \( \mu \) and \( \sigma^2 \) are a function of the desired mean and variance in the normative distribution. If the desired mean and variance of the normative distribution are denoted, respectively, by \( \alpha \) and \( \beta \), \( \mu \) and \( \sigma^2 \) in 5.2 and 5.3 are computed by

\[
\mu = \ln(\alpha) - \frac{\ln(\beta^2/\alpha^2 + 1)}{2} \tag{5.5}
\]

and

\[
\sigma^2 = \ln(\beta^2/\alpha^2 + 1). \tag{5.6}
\]

The appropriate derivations for 5.5 and 5.6 are given by Naylor, Balintfy, Burdick and Chu (1966). If \( z \) is a random normal deviate \( N(0,1) \), test scores \( T \) having lognormal distributions are generated by

\[
T_i = \exp(\mu + \sigma z_i) \quad i=1,2, \ldots, N \tag{5.7}
\]

for positively-skewed distributions, and

\[
T_i = K - \exp(\mu + \sigma z_i) \quad i=1,2, \ldots, N \tag{5.8}
\]

for negatively-skewed distributions. For normal distributions,

\[
T_i = \mu + \sigma z_i \quad i=1,2, \ldots, N \tag{5.9}
\]

The \( T \) scores computed in 5.7, 5.8 and 5.9 will be continuous variables. Because items are scored dichotomously, the \( T \) score must be rounded to the nearest integer value. The midpoint of each score interval is taken to be that point above which one-half of the area in that score interval is found. This point is found by integrating via trapezoid rule the area under the appropriate normal or lognormal curve. If the \( T \) score is
equal to or greater than the midpoint, the score is rounded up; if not, the score is rounded down.

Item scores are related to test scores. Specifically, if $X_{ij}$ is the item score for examinee $i$ on item $j$, $\sum_{j} X_{ij} = T_j$. Also, $\bar{p} = \bar{T}/K$. If $\sigma_p^2$ is greater than zero, individual item difficulty indices are generated by

$$p_i = \bar{p} + \sigma_p z_i \quad i=1,2, \ldots, K \quad (5.10)$$

where $z_i$ is a random normal deviate. When $\sigma_p^2$ is not equal to zero, the distribution of $p_i$ values will be approximately normal. If $\sigma_p^2$ is equal to zero, $p_i = \bar{p}$ for all values of $i$. With skewed distributions, $\sigma_p^2$ is typically $0 < \sigma_p^2 < .001$ and, because of this, $\sigma_p^2$ is set to zero for all skewed distributions generated by the simulation model. After the item difficulty indices have been generated within the program, deciding if an examinee passes or fails each item is relatively simple. Item difficulty indices are computed for all items generated. An examinee "passes" those items which will bring the computed item difficulty indices most closely to the desired item difficulty indices. For example, if the computed item difficulty for item $i$ were less than the desired item difficulty for item $i$, examinee $j$ would pass item $i$ if the computed difficulty were equal to or greater than the desired item difficulty, he "fails" item $i$. In the program, the desired item difficulty indices are sorted in descending order. If, in following the algorithm from the first through the $K$th item, $\sum_{j} X_{ij} \neq T_j$, the first $T_j - \sum_{j} X_{ij} = d$ items not already passed by examinee $j$ are scored by the program as items answered correctly by him.

The validity of the simulation model is found in its ability to generate the desired database. Two examples of data bases generated by the model are given in Tables 5.1 and 5.2. Although the discrepancies in Table 5.2 are minor, it should be noted that the magnitude of the discrepancies decreases with increases in $K$.

**Simulation of Multiple Matrix Sampling**

Subtests are constructed within the program by sampling at random items from the $K$-item population. For example, if $K$ equals 50 and a (5/10/30)
Table 5.1
Results Obtained From Simulation Model For 3000 Examinees When K = 20
With The Normative Distribution Distributed Normally

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Input</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>K</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>( \mu )</td>
<td>10.0000</td>
<td>10.0150</td>
</tr>
<tr>
<td>( \sigma^2 )</td>
<td>computed</td>
<td></td>
</tr>
<tr>
<td>( \alpha_{20} )</td>
<td>.8000</td>
<td>.7999</td>
</tr>
<tr>
<td>( \overline{p} )</td>
<td>.5000</td>
<td>.5010</td>
</tr>
<tr>
<td>( \sigma^2 )</td>
<td>.0800</td>
<td>.0799</td>
</tr>
<tr>
<td>N</td>
<td>3000</td>
<td>3000</td>
</tr>
</tbody>
</table>

Item Difficulty Indices

<table>
<thead>
<tr>
<th>Item</th>
<th>Input</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.987</td>
<td>.987</td>
</tr>
<tr>
<td>2</td>
<td>.901</td>
<td>.901</td>
</tr>
<tr>
<td>3</td>
<td>.857</td>
<td>.857</td>
</tr>
<tr>
<td>4</td>
<td>.782</td>
<td>.787</td>
</tr>
<tr>
<td>5</td>
<td>.750</td>
<td>.750</td>
</tr>
<tr>
<td>6</td>
<td>.737</td>
<td>.738</td>
</tr>
<tr>
<td>7</td>
<td>.702</td>
<td>.703</td>
</tr>
<tr>
<td>8</td>
<td>.634</td>
<td>.634</td>
</tr>
<tr>
<td>9</td>
<td>.565</td>
<td>.566</td>
</tr>
<tr>
<td>10</td>
<td>.518</td>
<td>.519</td>
</tr>
<tr>
<td>11</td>
<td>.458</td>
<td>.458</td>
</tr>
<tr>
<td>12</td>
<td>.442</td>
<td>.442</td>
</tr>
<tr>
<td>13</td>
<td>.430</td>
<td>.430</td>
</tr>
<tr>
<td>14</td>
<td>.374</td>
<td>.374</td>
</tr>
<tr>
<td>15</td>
<td>.275</td>
<td>.275</td>
</tr>
<tr>
<td>16</td>
<td>.208</td>
<td>.209</td>
</tr>
<tr>
<td>17</td>
<td>.139</td>
<td>.140</td>
</tr>
<tr>
<td>18</td>
<td>.113</td>
<td>.114</td>
</tr>
<tr>
<td>19</td>
<td>.112</td>
<td>.113</td>
</tr>
<tr>
<td>20</td>
<td>.026</td>
<td>.026</td>
</tr>
</tbody>
</table>

Obtained Frequency Dist.

<table>
<thead>
<tr>
<th>Score</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>16</td>
</tr>
<tr>
<td>1</td>
<td>17</td>
</tr>
<tr>
<td>2</td>
<td>36</td>
</tr>
<tr>
<td>3</td>
<td>47</td>
</tr>
<tr>
<td>4</td>
<td>96</td>
</tr>
<tr>
<td>5</td>
<td>136</td>
</tr>
<tr>
<td>6</td>
<td>182</td>
</tr>
<tr>
<td>7</td>
<td>225</td>
</tr>
<tr>
<td>8</td>
<td>307</td>
</tr>
<tr>
<td>9</td>
<td>266</td>
</tr>
<tr>
<td>10</td>
<td>328</td>
</tr>
<tr>
<td>11</td>
<td>327</td>
</tr>
<tr>
<td>12</td>
<td>256</td>
</tr>
<tr>
<td>13</td>
<td>210</td>
</tr>
<tr>
<td>14</td>
<td>191</td>
</tr>
<tr>
<td>15</td>
<td>122</td>
</tr>
<tr>
<td>16</td>
<td>104</td>
</tr>
<tr>
<td>17</td>
<td>67</td>
</tr>
<tr>
<td>18</td>
<td>38</td>
</tr>
<tr>
<td>19</td>
<td>16</td>
</tr>
<tr>
<td>20</td>
<td>13</td>
</tr>
</tbody>
</table>
Table 5.2
Results Obtained From Simulation Model For 3000 Examinees When K = 20 With The Normative Distribution Negatively-Skewed (Three Parameter Lognormal Distribution)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Input</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>K</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>$\mu$</td>
<td>17.5000</td>
<td>17.6150</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>computed</td>
<td></td>
</tr>
<tr>
<td>$\alpha_{20}$</td>
<td>.8000</td>
<td>.7570</td>
</tr>
<tr>
<td>$\frac{1}{p}$</td>
<td>.8750</td>
<td>.8810</td>
</tr>
<tr>
<td>$\sigma_{p}^2$</td>
<td>.0000</td>
<td>.0002</td>
</tr>
<tr>
<td>N</td>
<td>3000</td>
<td>3000</td>
</tr>
</tbody>
</table>

Item Difficulty Indices

<table>
<thead>
<tr>
<th>Item</th>
<th>Input</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.875</td>
<td>.930</td>
</tr>
<tr>
<td>2</td>
<td>.875</td>
<td>.908</td>
</tr>
<tr>
<td>3</td>
<td>.875</td>
<td>.888</td>
</tr>
<tr>
<td>4</td>
<td>.875</td>
<td>.876</td>
</tr>
<tr>
<td>5</td>
<td>.875</td>
<td>.876</td>
</tr>
<tr>
<td>6</td>
<td>.875</td>
<td>.876</td>
</tr>
<tr>
<td>7</td>
<td>.875</td>
<td>.876</td>
</tr>
<tr>
<td>8</td>
<td>.875</td>
<td>.876</td>
</tr>
<tr>
<td>9</td>
<td>.875</td>
<td>.876</td>
</tr>
<tr>
<td>10</td>
<td>.875</td>
<td>.876</td>
</tr>
<tr>
<td>11</td>
<td>.875</td>
<td>.876</td>
</tr>
<tr>
<td>12</td>
<td>.875</td>
<td>.876</td>
</tr>
<tr>
<td>13</td>
<td>.875</td>
<td>.876</td>
</tr>
<tr>
<td>14</td>
<td>.875</td>
<td>.876</td>
</tr>
<tr>
<td>15</td>
<td>.875</td>
<td>.876</td>
</tr>
<tr>
<td>16</td>
<td>.875</td>
<td>.876</td>
</tr>
<tr>
<td>17</td>
<td>.875</td>
<td>.876</td>
</tr>
<tr>
<td>18</td>
<td>.875</td>
<td>.876</td>
</tr>
<tr>
<td>19</td>
<td>.875</td>
<td>.875</td>
</tr>
<tr>
<td>20</td>
<td>.875</td>
<td>.875</td>
</tr>
</tbody>
</table>

Obtained Frequency Dist.

<table>
<thead>
<tr>
<th>Score</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>16</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>9</td>
<td>8</td>
</tr>
<tr>
<td>10</td>
<td>15</td>
</tr>
<tr>
<td>11</td>
<td>32</td>
</tr>
<tr>
<td>12</td>
<td>32</td>
</tr>
<tr>
<td>13</td>
<td>43</td>
</tr>
<tr>
<td>14</td>
<td>84</td>
</tr>
<tr>
<td>15</td>
<td>135</td>
</tr>
<tr>
<td>16</td>
<td>200</td>
</tr>
<tr>
<td>17</td>
<td>352</td>
</tr>
<tr>
<td>18</td>
<td>670</td>
</tr>
<tr>
<td>19</td>
<td>885</td>
</tr>
<tr>
<td>20</td>
<td>495</td>
</tr>
</tbody>
</table>
sampling plan were used, 5 subtests would be formed by sampling without replacement from the 50-item pool 10 items for each subtest. If (10/10/30) were used, 10 subtests would be formed containing 10 items each; however, the sampling plan for items requires sampling without replacement for each subtest but with replacement between subtests. In (10/10/30), several items will be common to more than one subtest. Taking (10/10/30) as an example, item scores on the K-item test would be generated by the program for 300 examinees. For subtest 1, the data from the first 30 examinees would be processed for only those items included in subtest 1. An identical procedure is followed for subtest 2 through subtest 10. The computations performed on each item-examinee sample are identical to those outlined in Chapter IV. If the user opts r replications of a particular sampling plan, r pooled estimates of each parameter will be produced and the standard error of estimate per parameter with that sampling plan is the standard deviation of the r pooled estimates for each parameter. Sample output for the (10/15/30) plan with 5 replications is given on page 42 through 49 for the normal normative distribution case.

Uses for the Simulation Model

It is anticipated that the computer program for simulating multiple matrix sampling described herein, and listed with expanded writeup in Appendix B, will facilitate readily a detailed examination of the relative merits of one or more sampling plans in multiple matrix sampling. In multiple matrix sampling questions asked frequently are "How do I do it?" and "If I sample this way, how accurate will the estimates be?" Questions such as these are answered easily through use of the simulation model. The results obtained from the program are reasonable to the degree that the normative distributions can be described adequately by the normal and lognormal probability distributions. It is commonly known that achievement test scores are frequently normally distributed. However, the scores on criterion-referenced tests, i.e., end-of-program tests, are frequently markedly negatively-skewed and resemble closely a three parameter lognormal distribution. It is anticipated that the simulation model will prove to be an asset in test theory and test construction courses permitting the student to have a working familiarity with sampling procedures used in multiple matrix sampling.
REFERENCES


PARAMETER INPUTTED

\[ \begin{array}{|c|c|}
\hline
\text{Parameter} & \text{Value} \\
\hline
\text{A} & 10.000 \\
\text{W} & 0.0 \\
\text{G} & 20 \\
\text{F} & 0.700 \\
\text{V1 (D)} & 0.0 \\
\hline
\end{array} \]

SAMPLE EXAMINATION SAMPLING PLAN

\[ \begin{array}{c}
\text{N} \\
\text{C1} \\
\text{C2} \\
\text{C3} \\
\text{S} \\
\text{N} \\
\end{array} \]

SWITCHES

\[ \begin{array}{c}
\text{ITEM-SAMPLING PLAN} \\
\text{HYPGEOMETRIC} \\
\text{NEGATIVE DISTRIBUTION} \\
\text{INTERMEDIATE PRIMICUT} \\
\end{array} \]

COMBINED STRENGTH = 3.86334
<table>
<thead>
<tr>
<th>N</th>
<th>0.0036</th>
<th>0.0039</th>
<th>0.0041</th>
<th>0.0044</th>
<th>0.0047</th>
<th>0.0050</th>
<th>0.0054</th>
<th>0.0057</th>
<th>0.0061</th>
<th>0.0064</th>
<th>0.0068</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0068</td>
<td>0.0072</td>
<td>0.0077</td>
<td>0.0082</td>
<td>0.0087</td>
<td>0.0092</td>
<td>0.0097</td>
<td>0.0103</td>
<td>0.0109</td>
<td>0.0115</td>
<td>0.0121</td>
</tr>
<tr>
<td>2</td>
<td>0.0121</td>
<td>0.0128</td>
<td>0.0135</td>
<td>0.0142</td>
<td>0.0149</td>
<td>0.0157</td>
<td>0.0165</td>
<td>0.0173</td>
<td>0.0182</td>
<td>0.0191</td>
<td>0.0200</td>
</tr>
<tr>
<td>3</td>
<td>0.0200</td>
<td>0.0210</td>
<td>0.0219</td>
<td>0.0230</td>
<td>0.0240</td>
<td>0.0252</td>
<td>0.0262</td>
<td>0.0273</td>
<td>0.0285</td>
<td>0.0297</td>
<td>0.0309</td>
</tr>
<tr>
<td>4</td>
<td>0.0309</td>
<td>0.0322</td>
<td>0.0333</td>
<td>0.0343</td>
<td>0.0361</td>
<td>0.0375</td>
<td>0.0383</td>
<td>0.0403</td>
<td>0.0417</td>
<td>0.0432</td>
<td>0.0447</td>
</tr>
<tr>
<td>5</td>
<td>0.0447</td>
<td>0.0462</td>
<td>0.0477</td>
<td>0.0493</td>
<td>0.0508</td>
<td>0.0524</td>
<td>0.0540</td>
<td>0.0556</td>
<td>0.0572</td>
<td>0.0588</td>
<td>0.0604</td>
</tr>
<tr>
<td>6</td>
<td>0.0604</td>
<td>0.0620</td>
<td>0.0637</td>
<td>0.0653</td>
<td>0.0659</td>
<td>0.0685</td>
<td>0.0701</td>
<td>0.0717</td>
<td>0.0734</td>
<td>0.0748</td>
<td>0.0764</td>
</tr>
<tr>
<td>7</td>
<td>0.0764</td>
<td>0.0779</td>
<td>0.0794</td>
<td>0.0809</td>
<td>0.0823</td>
<td>0.0836</td>
<td>0.0851</td>
<td>0.0865</td>
<td>0.0878</td>
<td>0.0891</td>
<td>0.0903</td>
</tr>
<tr>
<td>8</td>
<td>0.0903</td>
<td>0.0915</td>
<td>0.0926</td>
<td>0.0937</td>
<td>0.0948</td>
<td>0.0958</td>
<td>0.0967</td>
<td>0.0976</td>
<td>0.0984</td>
<td>0.0992</td>
<td>0.0999</td>
</tr>
<tr>
<td>9</td>
<td>0.0999</td>
<td>0.1005</td>
<td>0.1011</td>
<td>0.1016</td>
<td>0.1020</td>
<td>0.1024</td>
<td>0.1027</td>
<td>0.1030</td>
<td>0.1031</td>
<td>0.1032</td>
<td>0.1033</td>
</tr>
<tr>
<td>10</td>
<td>0.1033</td>
<td>0.1037</td>
<td>0.1031</td>
<td>0.1030</td>
<td>0.1027</td>
<td>0.1024</td>
<td>0.1020</td>
<td>0.1016</td>
<td>0.1011</td>
<td>0.1005</td>
<td>0.0999</td>
</tr>
<tr>
<td>11</td>
<td>0.0999</td>
<td>0.0992</td>
<td>0.0994</td>
<td>0.0976</td>
<td>0.0967</td>
<td>0.0958</td>
<td>0.0948</td>
<td>0.0937</td>
<td>0.0926</td>
<td>0.0915</td>
<td>0.0903</td>
</tr>
<tr>
<td>12</td>
<td>0.0891</td>
<td>0.0878</td>
<td>0.0865</td>
<td>0.0851</td>
<td>0.0838</td>
<td>0.0823</td>
<td>0.0806</td>
<td>0.0794</td>
<td>0.0776</td>
<td>0.0764</td>
<td>0.0764</td>
</tr>
<tr>
<td>13</td>
<td>0.0764</td>
<td>0.0748</td>
<td>0.0733</td>
<td>0.0717</td>
<td>0.0701</td>
<td>0.0685</td>
<td>0.0669</td>
<td>0.0653</td>
<td>0.0637</td>
<td>0.0620</td>
<td>0.0604</td>
</tr>
<tr>
<td>14</td>
<td>0.0604</td>
<td>0.0583</td>
<td>0.0572</td>
<td>0.0556</td>
<td>0.0540</td>
<td>0.0524</td>
<td>0.0508</td>
<td>0.0493</td>
<td>0.0477</td>
<td>0.0462</td>
<td>0.0447</td>
</tr>
<tr>
<td>15</td>
<td>0.0447</td>
<td>0.0432</td>
<td>0.0417</td>
<td>0.0403</td>
<td>0.0389</td>
<td>0.0375</td>
<td>0.0361</td>
<td>0.0348</td>
<td>0.0335</td>
<td>0.0322</td>
<td>0.0309</td>
</tr>
<tr>
<td>16</td>
<td>0.0309</td>
<td>0.0297</td>
<td>0.0285</td>
<td>0.0273</td>
<td>0.0262</td>
<td>0.0251</td>
<td>0.0240</td>
<td>0.0230</td>
<td>0.0219</td>
<td>0.0210</td>
<td>0.0200</td>
</tr>
<tr>
<td>17</td>
<td>0.0200</td>
<td>0.0191</td>
<td>0.0182</td>
<td>0.0173</td>
<td>0.0165</td>
<td>0.0157</td>
<td>0.0149</td>
<td>0.0142</td>
<td>0.0135</td>
<td>0.0128</td>
<td>0.0121</td>
</tr>
<tr>
<td>18</td>
<td>0.0121</td>
<td>0.0115</td>
<td>0.0109</td>
<td>0.0103</td>
<td>0.0097</td>
<td>0.0092</td>
<td>0.0087</td>
<td>0.0082</td>
<td>0.0077</td>
<td>0.0073</td>
<td>0.0068</td>
</tr>
<tr>
<td>19</td>
<td>0.0068</td>
<td>0.0064</td>
<td>0.0061</td>
<td>0.0057</td>
<td>0.0054</td>
<td>0.0050</td>
<td>0.0047</td>
<td>0.0044</td>
<td>0.0041</td>
<td>0.0039</td>
<td>0.0036</td>
</tr>
</tbody>
</table>

**AREAS:**

\[
\text{AREA} = 0.9933524
\]
<table>
<thead>
<tr>
<th>Rounding Rule</th>
<th>Score</th>
<th>Rounding</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.58</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1.57</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2.57</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>3.56</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>4.55</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>5.54</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>6.53</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>7.52</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>8.51</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>9.50</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>10.49</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>11.48</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>12.47</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>13.46</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>14.45</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>15.44</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>16.43</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>17.42</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>18.41</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>19.40</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>SUBTEST</td>
<td>ITEMS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>-------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2 8 15 20 3 19 13 9 14 7 4 6 1 12 16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>10 12 11 2 5 11 7 16 17 20 3 15 9 19 13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>7 9 13 19 2 3 18 10 20 12 11 4 14 6 17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>8 2 15 20 16 5 3 13 11 18 14 17 4 7 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>10 11 17 3 9 14 19 16 1 12 6 13 8 7 18</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>18 14 20 17 15 7 12 16 2 6 9 2 4 12 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>16 9 13 17 4 2 19 20 10 6 7 8 18 15 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>20 11 7 4 2 1 12 3 15 5 17 13 3 18 9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>20 2 12 14 15 16 1 6 8 9 7 5 4 13 19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>9 11 5 3 17 18 14 7 13 19 8 2 1 16 10</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Estimate of Parameter

<table>
<thead>
<tr>
<th>SAMPLE</th>
<th>MEAN</th>
<th>VARIANCE</th>
<th>STANDARD DEV</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10.66666</td>
<td>16.0153</td>
<td>4.0016</td>
</tr>
<tr>
<td>2</td>
<td>9.7333</td>
<td>10.158</td>
<td>3.1872</td>
</tr>
<tr>
<td>3</td>
<td>9.5555</td>
<td>14.554</td>
<td>3.7841</td>
</tr>
<tr>
<td>4</td>
<td>9.5999</td>
<td>14.318</td>
<td>3.7841</td>
</tr>
<tr>
<td>5</td>
<td>10.177</td>
<td>17.932</td>
<td>4.2347</td>
</tr>
<tr>
<td>6</td>
<td>9.3333</td>
<td>19.400</td>
<td>4.4046</td>
</tr>
<tr>
<td>7</td>
<td>11.111</td>
<td>16.088</td>
<td>4.1122</td>
</tr>
<tr>
<td>8</td>
<td>10.000</td>
<td>9.367</td>
<td>3.0526</td>
</tr>
<tr>
<td>10</td>
<td>10.799</td>
<td>17.084</td>
<td>3.7615</td>
</tr>
</tbody>
</table>

**Replication No. 1**

- **Est. of Mean**: 10.044
- **Est. of Standard Deviation**: 3.781
<table>
<thead>
<tr>
<th>Subtest</th>
<th>Items</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>17    18 1  2  10  4  12  5  8  15  20  9  14  17  7</td>
</tr>
<tr>
<td>2</td>
<td>14    4  10  1  20  10  2  5  11  13  6  17  8  7  15</td>
</tr>
<tr>
<td>3</td>
<td>6     4  12  14  18  7  3  19  11  20  8  2  5  1  16</td>
</tr>
<tr>
<td>4</td>
<td>8     10  7  20  17  2  3  16  19  5  14  13  4  12  15</td>
</tr>
<tr>
<td>5</td>
<td>19    6  7  13  4  11  10  13  1  15  14  17  20  5  2</td>
</tr>
<tr>
<td>6</td>
<td>14    8  6  1  17  11  18  12  19  7  2  16  4  15  3</td>
</tr>
<tr>
<td>7</td>
<td>5     7  14  8  18  20  19  1  17  12  2  15  16  4</td>
</tr>
<tr>
<td>8</td>
<td>13    12  14  17  2  5  10  20  8  9  13  3  4  6  7</td>
</tr>
<tr>
<td>9</td>
<td>9     14  3  12  8  11  13  15  17  10  4  2  7  19  6</td>
</tr>
<tr>
<td>10</td>
<td>8     19  6  5  20  1  7  15  9  4  11  12  10  13  17</td>
</tr>
</tbody>
</table>

**Estimate of Parameter**

<table>
<thead>
<tr>
<th>Sample</th>
<th>Mean</th>
<th>Variance</th>
<th>Standard Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10.6666660</td>
<td>16.0547180</td>
<td>4.0068340</td>
</tr>
<tr>
<td>3</td>
<td>11.022216</td>
<td>16.0278015</td>
<td>4.004732</td>
</tr>
<tr>
<td>4</td>
<td>8.9333215</td>
<td>13.312895</td>
<td>3.6485743</td>
</tr>
<tr>
<td>5</td>
<td>10.272204</td>
<td>11.9441700</td>
<td>3.4560337</td>
</tr>
<tr>
<td>6</td>
<td>10.844433</td>
<td>9.0143509</td>
<td>3.0229609</td>
</tr>
<tr>
<td>7</td>
<td>10.5333319</td>
<td>9.9500961</td>
<td>3.1543770</td>
</tr>
<tr>
<td>8</td>
<td>9.9111080</td>
<td>16.5915680</td>
<td>4.0722745</td>
</tr>
<tr>
<td>9</td>
<td>9.466653</td>
<td>17.9122477</td>
<td>4.2205734</td>
</tr>
<tr>
<td>10</td>
<td>10.5333319</td>
<td>18.1821594</td>
<td>4.2040543</td>
</tr>
</tbody>
</table>

Replication No. 2

Est. of Mean: 10.169
Est. of Standard Deviation: 3.397
<table>
<thead>
<tr>
<th>SUBTEST</th>
<th>ITEMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4 19 7 6 18 17 3 10 9 15 14 2 1 11 13</td>
</tr>
<tr>
<td>2</td>
<td>11 7 10 14 1 4 16 2 6 3 8 15 17 5 19</td>
</tr>
<tr>
<td>3</td>
<td>9 20 4 2 16 15 6 1 11 19 7 3 12 17 14</td>
</tr>
<tr>
<td>4</td>
<td>5 8 6 3 9 4 14 15 17 11 18 19 1 70 2</td>
</tr>
<tr>
<td>5</td>
<td>18 9 14 11 19 17 16 10 7 13 1 15 2 5 3</td>
</tr>
<tr>
<td>6</td>
<td>1 16 10 19 6 7 8 15 5 18 12 13 4 14 6</td>
</tr>
<tr>
<td>7</td>
<td>9 13 17 3 1 16 10 2 14 19 5 15 20 12 7</td>
</tr>
<tr>
<td>8</td>
<td>5 9 7 5 18 1 19 10 13 20 3 1 15 12 4</td>
</tr>
<tr>
<td>9</td>
<td>11 12 5 8 16 2 15 14 16 9 7 18 13 17 20</td>
</tr>
<tr>
<td>10</td>
<td>15 7 13 14 11 1 8 5 20 12 17 16 19 6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SAMPLE</th>
<th>MEAN</th>
<th>VARIANCE</th>
<th>STANDARD DEV</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10.7999983</td>
<td>7.2483969</td>
<td>2.0922846</td>
</tr>
<tr>
<td>2</td>
<td>9.2889970</td>
<td>21.0768351</td>
<td>4.5910597</td>
</tr>
<tr>
<td>3</td>
<td>10.6222200</td>
<td>9.0901139</td>
<td>3.0138531</td>
</tr>
<tr>
<td>4</td>
<td>9.8666649</td>
<td>14.9670533</td>
<td>3.8687277</td>
</tr>
<tr>
<td>5</td>
<td>9.1999989</td>
<td>21.3140544</td>
<td>4.6167145</td>
</tr>
<tr>
<td>6</td>
<td>10.5777760</td>
<td>14.8761902</td>
<td>3.8569660</td>
</tr>
<tr>
<td>7</td>
<td>9.4666653</td>
<td>15.2042278</td>
<td>3.9120817</td>
</tr>
<tr>
<td>8</td>
<td>9.4666653</td>
<td>16.3124237</td>
<td>4.0388641</td>
</tr>
<tr>
<td>9</td>
<td>10.0000000</td>
<td>16.4783630</td>
<td>4.0593548</td>
</tr>
<tr>
<td>10</td>
<td>10.2666655</td>
<td>10.9070606</td>
<td>3.3025837</td>
</tr>
</tbody>
</table>

REPLICA DATA NO. 3

EST. OF MEAN
9.356

EST. OF STANDARD DEVIATION
3.825
### Subtest Items

<table>
<thead>
<tr>
<th>Subtest</th>
<th>Items</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4 2 1 5 7 16 15 15 18 10 8 14 6 9 3</td>
</tr>
<tr>
<td>2</td>
<td>17 18 1 3 12 11 13 4 8 15 20 2 6 5 10</td>
</tr>
<tr>
<td>3</td>
<td>13 16 7 20 17 2 5 6 12 8 1 19 14 4</td>
</tr>
<tr>
<td>4</td>
<td>4 9 15 13 6 18 19 11 16 3 5 20 12 10 2</td>
</tr>
<tr>
<td>5</td>
<td>4 10 5 6 12 18 8 13 17 14 15 7 16 19 11</td>
</tr>
<tr>
<td>6</td>
<td>3 17 18 14 5 3 4 10 7 13 11 6 16 1 6</td>
</tr>
<tr>
<td>7</td>
<td>11 15 9 7 18 12 6 10 14 13 4 19 17 8 3</td>
</tr>
<tr>
<td>8</td>
<td>6 19 1 15 2 14 11 20 5 7 12 16 10 3</td>
</tr>
<tr>
<td>9</td>
<td>16 13 20 5 7 2 10 8 12 17 18 3 15 16 11</td>
</tr>
<tr>
<td>10</td>
<td>3 17 16 9 12 8 5 11 4 15 18 1 6 2 10</td>
</tr>
</tbody>
</table>

### Estimate of Parameter

#### Sample

<table>
<thead>
<tr>
<th>Sample</th>
<th>Mean</th>
<th>Variance</th>
<th>Standard Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10.2222204</td>
<td>22.0361176</td>
<td>4.6942635</td>
</tr>
<tr>
<td>2</td>
<td>10.4888868</td>
<td>18.7351990</td>
<td>4.3284169</td>
</tr>
<tr>
<td>3</td>
<td>11.5555534</td>
<td>14.8910809</td>
<td>3.858963</td>
</tr>
<tr>
<td>4</td>
<td>9.2444420</td>
<td>15.0130234</td>
<td>3.8745613</td>
</tr>
<tr>
<td>5</td>
<td>10.1333323</td>
<td>28.0562439</td>
<td>5.2861670</td>
</tr>
<tr>
<td>6</td>
<td>10.6222200</td>
<td>12.2294531</td>
<td>3.4570627</td>
</tr>
<tr>
<td>7</td>
<td>10.3898874</td>
<td>14.3166456</td>
<td>3.8492308</td>
</tr>
<tr>
<td>8</td>
<td>9.5555534</td>
<td>10.1718702</td>
<td>3.1853368</td>
</tr>
<tr>
<td>9</td>
<td>9.9555540</td>
<td>15.4686565</td>
<td>3.9330187</td>
</tr>
<tr>
<td>10</td>
<td>9.5111103</td>
<td>15.8550692</td>
<td>3.9818470</td>
</tr>
</tbody>
</table>

Repl. No. 4

Est. of Mean: 10.218
Est. of Standard Deviation: 4.083
### SUBTEST ITEMS

<table>
<thead>
<tr>
<th></th>
<th>18</th>
<th>9</th>
<th>15</th>
<th>14</th>
<th>8</th>
<th>4</th>
<th>17</th>
<th>11</th>
<th>20</th>
<th>12</th>
<th>13</th>
<th>5</th>
<th>19</th>
<th>16</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### ESTIMATE OF PARAMETER

<table>
<thead>
<tr>
<th>SAMPLE</th>
<th>MEAN</th>
<th>VARIANCE</th>
<th>STANDARD DEV</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10.7555552</td>
<td>11.9608207</td>
<td>3.4407557</td>
</tr>
<tr>
<td>2</td>
<td>10.0111086</td>
<td>21.6546105</td>
<td>4.6534405</td>
</tr>
<tr>
<td>3</td>
<td>10.0888892</td>
<td>17.2166230</td>
<td>4.1492970</td>
</tr>
<tr>
<td>4</td>
<td>10.5777776</td>
<td>14.3824806</td>
<td>3.7924232</td>
</tr>
<tr>
<td>5</td>
<td>10.6444411</td>
<td>21.9667206</td>
<td>4.6366668</td>
</tr>
<tr>
<td>6</td>
<td>10.9777756</td>
<td>13.3876837</td>
<td>3.7265178</td>
</tr>
<tr>
<td>7</td>
<td>11.0222115</td>
<td>17.3232323</td>
<td>4.1622000</td>
</tr>
<tr>
<td>8</td>
<td>10.2898870</td>
<td>12.2294531</td>
<td>3.4970627</td>
</tr>
<tr>
<td>9</td>
<td>10.1233236</td>
<td>15.0863810</td>
<td>3.8441190</td>
</tr>
<tr>
<td>10</td>
<td>11.1555558</td>
<td>14.1529202</td>
<td>3.7621699</td>
</tr>
</tbody>
</table>

### REPLICATION NO. 5

**EST. OF MEAN**

10.1646287

**EST. OF STANDARD DEVIATION**

3.301

**HERE POOLED MEAN OVER REPS**

10.1646287

**SE OF POOLED MEAN OVER REPS**

0.1643227

**HERE POOLED SD OVER REPS**

3.3049706

**SE OF POOLED SD OVER REPS**

0.1175046
VI

Hypothesis Testing and Multiple Matrix Sampling

Parameters estimated through multiple matrix sampling are integrated easily into a variety of hypothesis testing procedures. For example, one-sample and two-sample t-tests can be performed readily with estimates of $\mu$ and $\sigma^2$ obtained by multiple matrix sampling. Specifically,

$$t_{df=N-1} = \frac{\hat{\mu}_{\text{pooled}} - \mu_{\text{standard}}}{(\hat{\sigma}^2_{\text{pooled}}/N)^{1/2}} \quad (6.1)$$

and

$$t_{df=N_1+N_2-2} = \frac{\hat{\mu}_{1\text{pooled}} - \hat{\mu}_{2\text{pooled}} - (\mu_1 - \mu_2)}{\left[\frac{(N_1-1)\hat{\sigma}^2_{1\text{pooled}} + (N_2-1)\hat{\sigma}^2_{2\text{pooled}}}{N_1+N_2-2}\right]^{1/2} \left[\frac{1}{N_1} + \frac{1}{N_2}\right]} \quad (6.2)$$

The t-test for the difference between two independent means given in 6.2 can be extended to completely randomized and factorial analysis of variance designs where the dependent variable is a mean test score. Although analysis of variance designs with mean scores as the dependent variable are found infrequently in the literature, the frequently occurring circumstances in which mean scores are preferable to raw scores in such analyses are detailed most succinctly by Peckham, Glass and Hopkins (1969).
Consider the design in which the relative merits of four experimental training programs are being contrasted through end-of-program test scores obtained from students participating in each procedure. Through an analysis of pretest scores given to all students, ten classes have been selected for each training program such that, across training programs, the four groups of 10 classes are approximately homogeneous at the start of instruction. The mean achievement test score for each class is estimated easily through multiple matrix sampling. The statistical layout and sources of variation are given in Table 6.1. If an additional variable, such as school district, were added to the design, the statistical layout and sources of variation are modified slightly as seen in Table 6.2. After the measurement on the dependent variable is accomplished, computations in the analysis of variance proceed in the usual manner. The novelty herein is in estimating the class mean test score through multiple matrix sampling.

Testing homogeneity of variance hypotheses of the form \( \sigma_1^2 = \sigma_2^2 = \ldots = \sigma^2 \) is accomplished for two variances by

\[
F_{\text{pooled}}(N_1 - 1, N_2 - 1) = \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2}.
\]

and for more than two variances by, for example,

\[
F_{\text{max}} = \frac{\hat{\sigma}_{\text{largest}}^2}{\hat{\sigma}_{\text{smallest}}^2}.
\]

Tables for the \( F_{\text{max}} \) statistic have been constructed by Hartley and are given in Winer (1962, p. 653). Another simple test for homogeneity of variance developed by Cochran which lends itself to multiple matrix sampling is
and the necessary tables for the $C$ statistic are given in Winer (1962, p. 654). The procedures in 6.3, 6.4, and 6.5 are not the only tests possible, but they are used frequently and illustrate the concept.

The normative distribution approximated by the negative hypergeometric distribution with parameters estimated through multiple matrix sampling provides the basic data for several goodness-of-fit tests. For example, the Kolmogorov-Smirnov one-sample test (Siegel, 1956, pp. 47-52) provides a test of the hypothesis that the approximated distribution of scores came from a population of scores having a specified theoretical distribution. The test involves specifying the cumulative frequency distribution which would occur under the theoretical distribution and comparing that with the approximated cumulative frequency distribution. The cumulative frequency distribution is, of course, obtained readily after the individual frequencies have been determined by multiplying the number of examinees in the population by the relative frequency per test score approximated by the negative hypergeometric distribution. A simple extension of the Kolmogorov-Smirnov one-sample test is the Kolmogorov-Smirnov two-sample test (Siegel, 1956, pp. 127-136) which is concerned with the agreement between two approximated frequency distributions.

The tests of hypotheses mentioned herein do not constitute an exhaustive listing of statistical tests to which estimates of parameters obtained through multiple matrix sampling are applicable. The intent is merely that of suggesting the applicability of a novel technique to traditional hypothesis testing procedures. It should be noted that the $t$-tests given in 6.1 and 6.2 are to be considered conservative tests of the hypotheses under consideration. The standard errors of estimate given in the denominators are those for the matched-items design and there is evidence (Osburn, 1967) suggesting that the corresponding standard errors under multiple matrix sampling will be less. In the algebraic derivation supporting this conclusion, Osburn was considering a form of multiple matrix sampling in which $k$ items were selected at random from the population of items for each examinee.
Table 6.1
Statistical Layout For One-Way Analysis Of Variance Problem With The Dependent Variable Being A Mean Achievement Test Score Estimated Through Multiple Matrix Sampling

<table>
<thead>
<tr>
<th>Source Of Variation</th>
<th>Degrees Of Freedom</th>
</tr>
</thead>
<tbody>
<tr>
<td>Programs</td>
<td>3</td>
</tr>
<tr>
<td>Classes Within Programs</td>
<td>36</td>
</tr>
<tr>
<td>Total</td>
<td>39</td>
</tr>
</tbody>
</table>
Table 6.2

Statistical Layout For Factorial (Two-Way) Analysis Of Variance Problem With The Dependent Variable Being A Mean Achievement Test Score Estimated Through Multiple Matrix Sampling

<table>
<thead>
<tr>
<th>Source Of Variation</th>
<th>Degrees Of Freedom</th>
</tr>
</thead>
<tbody>
<tr>
<td>Programs</td>
<td>3</td>
</tr>
<tr>
<td>Districts</td>
<td>1</td>
</tr>
<tr>
<td>Programs x Districts</td>
<td>3</td>
</tr>
<tr>
<td>Classes Within Programs x Districts</td>
<td>32</td>
</tr>
<tr>
<td>Total</td>
<td>39</td>
</tr>
</tbody>
</table>
REFERENCES


Unique Applications of Multiple Matrix Sampling

Multiple matrix sampling has been used traditionally to estimate parameters of standardized tests where the total test score is equal to the sum of the item scores. For investigations focused primarily on group assessment, multiple matrix sampling has been demonstrated empirically to be an important and valuable procedure. Multiple matrix sampling, however, is applicable to a broader range of research problems than that suggested by the current literature. Four unique and important applications of multiple matrix sampling are described in this chapter. As is the case with most psychometric procedures and is certainly the case with multiple matrix sampling, the range of applications is determined solely by the degree of inventiveness in the individual researcher.

Design of Experiments

In the evaluation of instructional programs, the pre-post paradigm is used frequently and, as is traditionally the case, an individual test is administered to all examinees at both the start and end of instruction. Given an item population related to the instructional program under evaluation, a research design such as this is improved easily with the addition of multiple matrix sampling. In place of using the same test pre and post, random or stratified-random parallel tests are used with parameters for both tests estimated through multiple matrix sampling. A procedure such as this could be expanded further to include intermediate testing using additional parallel tests. An example of a design such as this and one demonstrating the concomitant benefits is given by Osburn and Shoemaker (1968). In the evaluation of instructional programs it should be noted that a researcher is seldom interested in individual test items, individual tests, or individual examinees but is interested primarily in group behavior over time with regard to some specified item population. As such, multiple matrix sampling in conjunction with random or stratified-random parallel tests is an ideal measurement procedure.

Estimation of Covariance and Correlation Matrices

Item and test covariance matrices (and, hence, correlation matrices) are estimated readily through multiple matrix sampling. A modified
sampling plan is required such that all possible pairs of items or tests are included in one or more subtests or subbatteries. For example, consider estimating the elements in a covariance matrix for a 5-item test. To compute the covariance of Item 1 with Item 4, there must be a subgroup of examinees responding to both Item 1 and Item 4. If the examinee subgroup is sampled randomly from the population of examinees, \( \text{COV}(1, 4) \) computed over those examinees is an estimate of \( \text{COV}(1, 4) \) which would have been obtained by testing all examinees over both items. All remaining entries in the covariance matrix are estimated identically. A test covariance matrix is determined similarly with items being replaced by tests. A procedure such as this sets the stage for multiple matrix sampling playing an important role in a variety of multivariate procedures as, for example, factor analysis. Although little has been done in this area, some important preliminary research and a few of the relevant equations for estimating parameters have been reported by Lord (1960), Ray, Hundleby and Goldstein (1962), Knapp (1968) and Timm (1970).

Questionnaires and Surveys

A perennial problem with questionnaires and surveys is the disappointingly low rate of completions or returns. Return rates of 20 to 30 per cent are not uncommon. Although examinees fail to return questionnaires for a multitude of reasons, one factor is undoubtedly the length of the questionnaire and the time required to complete all questions. If the measurement required is the proportion of examinees in each category, results can be approximated through multiple matrix sampling by administering questions selected randomly to a random sample of examinees. For example, if an 8-page questionnaire were to be administered to all elementary school teachers within a particular city, the questions contained therein could be divided into 8 subquestionnaires (each of which would require no more than the front of one piece of paper) with each subquestionnaire administered to a random sample of teachers. The time for completing each subquestionnaire is minimal and, as such, may increase the rate of returns. The point to be made is simply this: a little data from a large number of teachers is better than a lot of data from few teachers. It must be remembered, however, that questions within questionnaires are interrelated frequently (If "No" on Question 13, go to Question 20.) and complications such as these must be incorporated in constructing subquestionnaires.

Measurement in the Affective Domain

It is frequently the case that an investigator is interested in scaling the preferences or affect of a group of individuals for a particular set of objects. Although there are several procedures which could be used, the method of paired-comparisons is one encountered frequently in the literature (e.g., Snider (1960) and Holliman (1970). In the
Table 7.1

Example F-matrix And P-matrix Obtained By Method Of Paired-Comparisons For 6 Stimuli

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6</td>
<td>5</td>
<td>7</td>
<td>10</td>
<td>7</td>
<td></td>
<td>.353</td>
<td>.294</td>
<td>.412</td>
<td>.588</td>
<td>.412</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>11</td>
<td>9</td>
<td>10</td>
<td>12</td>
<td>10</td>
<td></td>
<td>.647</td>
<td>.529</td>
<td>.588</td>
<td>.706</td>
<td>.588</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>12</td>
<td>8</td>
<td>12</td>
<td>12</td>
<td>9</td>
<td></td>
<td>.706</td>
<td>.471</td>
<td>.706</td>
<td>.706</td>
<td>.529</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>10</td>
<td>7</td>
<td>5</td>
<td>14</td>
<td>6</td>
<td></td>
<td>.588</td>
<td>.412</td>
<td>.294</td>
<td>.824</td>
<td>.353</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>7</td>
<td>5</td>
<td>5</td>
<td>3</td>
<td>5</td>
<td></td>
<td>.412</td>
<td>.294</td>
<td>.294</td>
<td>.176</td>
<td>.294</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>10</td>
<td>7</td>
<td>8</td>
<td>11</td>
<td>12</td>
<td></td>
<td>.588</td>
<td>.412</td>
<td>.471</td>
<td>.647</td>
<td>.706</td>
<td></td>
</tr>
</tbody>
</table>
method of paired-comparisons, all possible combinations of the objects taking two at a time are presented individually and for each pair the examinee is asked to indicate his preference. For example, if 6 stimuli were being scaled by the method of paired-comparisons, the test so constructed would contain $6(6-1)/2 = 15$ items, for 12 stimuli, 66 items. After all pairs have been administered to all examinees, the preliminary analysis of the data involves the computation of the $F$-matrix and subsequent $P$-matrix. The $P$-matrix is the base from which the scale values per stimulus are computed and it is in estimating the values in the $P$-matrix that an application of multiple matrix sampling is found. Relevant preliminary research in this area has been reported by McCormick and Roberts (1952), McCormick and Bachus (1952) and Bursack and Cook (1970). If the $s$ stimuli are numbered consecutively from 1 to $s$, the $F$-matrix is an $s$ by $s$ matrix with entries denoting the frequency with which the column stimulus was judged more favorable than the row stimulus. An example of an $F$-matrix and associated $P$-matrix are given in Table 7.1. Dividing each entry in the $F$-matrix by the total number of examinees, which is in this case equal to 17, produces the corresponding entry in the $P$-matrix labeled appropriately as the proportion of examinees selecting the column stimulus over the row stimulus. In estimating the entries in the $P$-matrix through multiple matrix sampling, paired-comparisons are selected at random from the pool of all possible pairs and administered to samples of examinees selected randomly from the testable population.

Shoemaker (1971) using a post mortem item-examinee sampling design has explored systematically the feasibility of using multiple matrix sampling to estimate scale values obtained by the method of paired-comparisons. The major conclusions reached in this investigation were that (a) scale values can be approximated satisfactorily through multiple matrix sampling, and (b) the similarity between the estimated scale values and the normative scale values increases with increases in the number of observations acquired by the sampling plan, with the converse true. The specific procedure used to estimate the $P$-matrix from subtest results is detailed in the following 5 steps. Each step is illustrated with results from one replication of a (3/10/15) sampling plan. (In the Shoemaker investigation, the data base consisted of responses made by 407 primary grade students to a 15-item test designed to scale degree of affect to 6 stimuli.)

Step 1: Three subtests containing 10 items each are formed by sampling items randomly and without replacement within subtests but with replacement between subtests.

<table>
<thead>
<tr>
<th>Subtest</th>
<th>Items</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8 14 11 9 6 2 3 12 15 10</td>
</tr>
<tr>
<td>2</td>
<td>14 6 3 12 4 1 2 11 13 10</td>
</tr>
<tr>
<td>3</td>
<td>14 13 2 9 6 11 4 15 3 12</td>
</tr>
</tbody>
</table>
Step 2: Three subgroups of examinees containing 15 examinees each are formed by sampling randomly and without replacement from the 407-examinee population.

<table>
<thead>
<tr>
<th>Subgroup</th>
<th>Examinees</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>359 22 280 272 139 206 169 321 323 23 271 66 221 109 100</td>
</tr>
<tr>
<td>2</td>
<td>345 367 281 390 366 70 361 250 154 168 8 138 279 335 399</td>
</tr>
<tr>
<td>3</td>
<td>342 220 276 125 382 219 217 327 401 385 113 62 77 192 156</td>
</tr>
</tbody>
</table>

Step 3: Pairing subtest 1 with subgroup 1, an f-matrix is formed for each subtest using only the responses made by the corresponding examinee subgroup on the items contained in that subtest. Each f-matrix is constructed in conjunction with a link-matrix containing the code numbers of stimuli paired within each test item. For the data base considered herein, the link-matrix was:

<table>
<thead>
<tr>
<th>Test Item</th>
<th>Stimulus Pair</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>1 2</td>
</tr>
<tr>
<td>02</td>
<td>4 3</td>
</tr>
<tr>
<td>03</td>
<td>5 6</td>
</tr>
<tr>
<td>04</td>
<td>2 6</td>
</tr>
<tr>
<td>05</td>
<td>1 3</td>
</tr>
<tr>
<td>06</td>
<td>4 5</td>
</tr>
<tr>
<td>07</td>
<td>2 3</td>
</tr>
<tr>
<td>08</td>
<td>1 5</td>
</tr>
<tr>
<td>09</td>
<td>4 6</td>
</tr>
<tr>
<td>10</td>
<td>1 4</td>
</tr>
<tr>
<td>11</td>
<td>2 5</td>
</tr>
<tr>
<td>12</td>
<td>3 6</td>
</tr>
<tr>
<td>13</td>
<td>3 5</td>
</tr>
<tr>
<td>14</td>
<td>1 6</td>
</tr>
<tr>
<td>15</td>
<td>2 4</td>
</tr>
</tbody>
</table>

The f-matrices for the 3 subtests used in (3/10/15) are:

\[
\begin{bmatrix}
0 & 0 & 9 & 14 & 8 \\
0 & 0 & 8 & 15 & 0 \\
6 & 7 & 10 & 12 & 9 \\
1 & 0 & 0 & 3 & 2 \\
7 & 0 & 11 & 6 & 13
\end{bmatrix}
\]

f-matrix 1 =
Step 4: In pooling the f-matrices to obtain the P-matrix, an accounting-matrix is required to distinguish between items omitted in the construction of subtests and items to which all examinees in a particular subgroup responded identically. For the f-matrices given in step 3, the accounting-matrix is

\[
\begin{bmatrix}
0 & 1 & 0 & 2 & 1 & 3 \\
1 & 0 & 0 & 2 & 3 & 2 \\
0 & 0 & 0 & 3 & 2 & 3 \\
2 & 2 & 3 & 0 & 3 & 2 \\
1 & 3 & 2 & 3 & 0 & 3 \\
3 & 2 & 3 & 2 & 3 & 0
\end{bmatrix}
\]

Off-diagonal zeros are of critical importance in pooling subtest results. In each f-matrix, \(f(i,j) + f(j,i) = n\) for those stimulus pairs contained within the subtest and \(f(j,i)\), for example, could be zero for two reasons: (a) the item containing stimulus pair \((i,j)\) was not included in the subtest, or (b) all examinees in that particular subgroup selected stimulus \(i\) over stimulus \(j\). This distinction must be maintained in pooling the f-matrices to produce the P-matrix.

Step 5: The P-matrix is formed by pooling across corresponding entries in the f-matrices after each entry in the f-matrix has been divided by the number of examinees in the corresponding subgroup. The sum of proportions is then divided by the corresponding number in the accounting-matrix. As an example, consider computing the \((1,6)\) and \((5,1)\) entries in the P-matrix:
8/15 + 10/15 + 7/15
P(1,6) = \frac{8/15 + 10/15 + 7/15}{3} = .556

1/15 + no data + no data
P(5,1) = \frac{1/15 + no data + no data}{1} = .067

If the number of examinees per subgroup is unequal, the proportions are combined by a weighted arithmetic mean and the corresponding entry in the accounting-matrix is equal to the number of examinees for which data existed. Elements in the P-matrix are set equal to .5 if the corresponding entry in the accounting-matrix is equal to zero. In this example, the P-matrix is

\[
P\text{-matrix} =
\begin{bmatrix}
.500 & .867 & .500 & .467 & .933 & .556 \\
.133 & .500 & .500 & .567 & .889 & .400 \\
.500 & .500 & .500 & .333 & .400 & .222 \\
.533 & .433 & .667 & .500 & .800 & .667 \\
.067 & .111 & .600 & .200 & .500 & .133 \\
.444 & .600 & .778 & .333 & .867 & .500 \\
\end{bmatrix}
\]

After the P-matrix has been formed, scale values per stimulus are computed as if all examinees had responded to all items using computational procedures detailed in Edwards (1957). Using Thurstone's Model V scaling procedure, the resultant scale values from the P-matrix given in step 5 and those obtained from using all 407 examinees over all items are

<table>
<thead>
<tr>
<th></th>
<th>S_1</th>
<th>S_2</th>
<th>S_3</th>
<th>S_4</th>
<th>S_5</th>
<th>S_6</th>
</tr>
</thead>
<tbody>
<tr>
<td>(3/10/15)</td>
<td>.000</td>
<td>.442</td>
<td>.688</td>
<td>.172</td>
<td>1.184</td>
<td>.184</td>
</tr>
<tr>
<td>Norm(5)</td>
<td>.000</td>
<td>.075</td>
<td>.638</td>
<td>.215</td>
<td>1.023</td>
<td>.193</td>
</tr>
</tbody>
</table>
REFERENCES


BIBLIOGRAPHY


Cronbach, L. J. Course improvement through evaluation. Teachers College Record, 1963, 64, 672-683.


APPENDIX A

Listing And Expanded Writeup Of Computer Program For Estimating Test Parameters Through Multiple Matrix Sampling And For Approximating Normative Distributions With The Negative Hypergeometric Distribution
A Fortran IV Program For Estimating Test Parameters Through Multiple Matrix Sampling And For Approximating A Normative Distribution Of Test Scores With The Negative Hypergeometric Distribution

The negative hypergeometric distribution provides a reasonably good fit for a variety of test score distributions when the test score is the number of correct answers. The negative hypergeometric distribution is a function of the mean test score, the variance of the test scores and the total number of items in the test. The first two parameters may be approximated efficiently by multiple matrix sampling. Furthermore, the negative hypergeometric distribution with parameters estimated by multiple matrix sampling can be used satisfactorily to approximate a normative distribution of number correct test scores.

In multiple matrix sampling, a set of K test items is randomly divided into subsets of items. Each subset of items is then randomly assigned to a group of examinees. Although each examinee receives only a proportion of the complete set of test items, the statistical model permits one to estimate the mean and variance of the total test score distribution for all examinees over the complete set of test items. Multiple matrix sampling is an efficient procedure for approximating a normative distribution when it is not possible or is economically unfeasible to administer the complete set of K items to all examinees in the testable population.

The Fortran IV program which approximates the normative distribution with the negative hypergeometric distribution is relatively machine-independent and has been implemented easily on an IBM 7040, IBM S360/50, IBM S360/91 and a UNIVAC 1108. The program has been designed to approximate test score distributions involving at maximum 500 items. However, this restriction may be easily modified. The number of subtests and number of examinees per subtest are limited only by the amount of computer time available.

Organization Of Control Cards And Data Cards

<table>
<thead>
<tr>
<th>columns</th>
<th>(all integers right-justified)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Card Set 1 (1 card)</td>
<td>1-72</td>
</tr>
<tr>
<td>Card Set 2 (1 card)</td>
<td>1-5</td>
</tr>
<tr>
<td></td>
<td>6-10</td>
</tr>
<tr>
<td></td>
<td>11-15</td>
</tr>
<tr>
<td>columns</td>
<td>(all integers right-justified)</td>
</tr>
<tr>
<td>---------</td>
<td>------------------------------</td>
</tr>
<tr>
<td>51-55</td>
<td>Punch 00000 if there is only one format card by which all item scores are to be inputted</td>
</tr>
<tr>
<td></td>
<td>Punch 00001 if there is to be a different format card for each item-examinee sample within a data set</td>
</tr>
</tbody>
</table>

**Format Card Set**

(k cards, optional)

Standard Fortran IV format punched in columns 1-72 on each card and enclosed in parentheses for inputting item scores for each examinee in each item-examinee sample. The number of format cards may not exceed 9 for each item-examinee data set. The first card after the format cards must contain END OF FORMAT in columns 1-13.

Example: (5X,25F1.0) END OF FORMAT

**Data Card Set**

(k cards, optional)

The responses of each examinee per item-examinee sample should be sequenced by examinee group and within each group by examinee.

**Acceptable Input Data Structures**

<table>
<thead>
<tr>
<th>Plan 1</th>
<th>Plan 2</th>
<th>Plan 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fortran Source Deck Card Set 1</td>
<td>Fortran Source Deck Card Set 1</td>
<td>Fortran Source Deck Card Set 1</td>
</tr>
<tr>
<td>Card Set 2</td>
<td>Card Set 2</td>
<td>Card Set 2</td>
</tr>
<tr>
<td>Format Card Set</td>
<td>Data From Sample 1</td>
<td>Format Card Set 1</td>
</tr>
<tr>
<td>Data Cards</td>
<td></td>
<td>Data From Sample 1</td>
</tr>
<tr>
<td></td>
<td>Format Card Set 2</td>
<td>Format Card Set 2</td>
</tr>
<tr>
<td></td>
<td>Data From Sample 2</td>
<td>Data From Sample 2</td>
</tr>
<tr>
<td></td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td></td>
<td>Format Card Set t</td>
<td>Format Card Set t</td>
</tr>
<tr>
<td></td>
<td>Data From Sample t</td>
<td>Data From Sample t</td>
</tr>
<tr>
<td>Plan 4</td>
<td>Plan 5</td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>--------</td>
<td></td>
</tr>
<tr>
<td><strong>Fortran Source Deck</strong></td>
<td><strong>Fortran Source Deck</strong></td>
<td></td>
</tr>
<tr>
<td>Card Set 1</td>
<td>Card Set 1</td>
<td></td>
</tr>
<tr>
<td>Card Set 2</td>
<td>Card Set 2</td>
<td></td>
</tr>
<tr>
<td>Format Card Set</td>
<td>Format Card Set 1</td>
<td></td>
</tr>
<tr>
<td>Card with no. of examinees and items for subtest 1</td>
<td>Card with no. of examinees and items for subtest 1</td>
<td></td>
</tr>
<tr>
<td>Data from subtest 1</td>
<td>Data from subtest 1</td>
<td></td>
</tr>
<tr>
<td>Card with no. of examinees and items for subtest 2</td>
<td>Format Card Set 2</td>
<td></td>
</tr>
<tr>
<td>Data from subtest 2</td>
<td>Card with no. of examinees and items for subtest 2</td>
<td></td>
</tr>
<tr>
<td>Data from subtest 2</td>
<td>Data from subtest 2</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td></td>
</tr>
<tr>
<td>Card with no. of examinees and items for subtest t</td>
<td>Format Card Set t</td>
<td></td>
</tr>
<tr>
<td>Data from subtest t</td>
<td>Card with no. of examinees and items for subtest t</td>
<td></td>
</tr>
<tr>
<td>Data from subtest t</td>
<td>Data from subtest t</td>
<td></td>
</tr>
</tbody>
</table>

**Plan 1:** Mean and variance of test scores are inputted on card set 2. No item scores are required.

**Plan 2:** Mean and variance of test scores are to be estimated from item-examinee samples. All item scores in each item-examinee sample are organized in the same manner on the data card and are to be inputted with one format card.

**Plan 3:** Same as Plan 2 with exception that item scores for each item-examinee sample are not organized on data cards in same manner. Each sample requires an individual set of format cards describing how item scores are organized for that particular sample.

**Plan 4:** Same as Plan 2 with exception that number of examinees and number of items per subtest are not constant across subtests. Same format card is used for each data set.

**Plan 5:** Same as Plan 3 with exception that number of examinees and number of items per subtest are not constant across subtests. In addition, different format cards are used for each data set.
APPROXIMATION OF FREQUENCY DISTRIBUTION OF TEST SCORES
BY NEGATIVE HYPERGEOMETRIC DISTRIBUTION

REFERENCE
LORD, F.M. AND NOVICK, M.R. STATISTICAL THEORIES OF MENTAL TEST
SCORES. READING, MASS. ADDISON-WESLEY, 1968, CHAPTER 23.

DAVID M. SHOEMAKER

NTS = NUMBER OF ITEMS PER SUBTEST
NTP = NUMBER OF ITEMS IN TEST ITEM POPULATION
NSM = NUMBER OF SUBTESTS
NSS = NUMBER OF EXAMINEES PER SUBGROUP
NSP = NUMBER OF EXAMINEES IN EXAMINEE POPULATION
XBAR = ESTIMATE OF MEAN TEST SCORE
VAR = ESTIMATE OF TEST SCORE VARIANCE

COMMON DUMMY(500), P(500)
DIMENSION TITLE(18)

INPUT PROBLEM PARAMETERS

1000 READ (5, 1, END=5000) (TITLE(I), I=1, 18),
1NSM, NTS, NSS, NTP, NSP, XBAR, VAR, NGPH, NFMT
WRITE (6, 5) (TITLE(I), I=1, 18)
IF (IFIX(VAR*1000.) .NE. 0) WRITE (6, 9) XBAR, VAR

ESTIMATE MEAN AND VARIANCE FROM SUBTESTS

IF (IFIX(VAR*1000.). EQ. 0) CALL POOL(NSS, NTS, NSS, NTP, XBAR, VAR, NFMT)
IF (NSP .EQ. 0) NSP=1000.
WRITE (6, 10) NSP

COMPUTE PARAMETERS FOR NEGATIVE HYPERGEOMETRIC DISTRIBUTION

S=NTP
A21= (S/(S-1.))*(1.-XBAR*(S-XBAR)/(S*VAR))
IF (A21  .GT. 0.) GO TO 40
WRITE (6, 7) A21
GO TO 1000

40 CONTINUE
A=(-1.+1./A21)*XBAR
B=-A-1.+S/A21
SLOG1=0.
SLOG2=0.
C=A+B
DO 20 I=1,NTP
  SLOG1=SLOG1+ALOG10(B-I+1.)
  SLOG2=SLOG2+ALOG10(C-I+1.)
  C=10.**03LU(3A-SLOG2)
WRITE (6,3) A21,A,B,C
20 COMPUTE NEGATIVE HYPERGEOMETRIC DISTRIBUTION

N3=NTP+1
WRITE (6,4)
CK=0.
DO 100 I=1,N3
  K=I-1
  CALL NEGHGR (K,A,B,C,S,NSP,HX,HFX)
  CK=CK+HX
  F(I)=HX
WRITE (6,?) K,HX,HFX,CK
100 CONTINUE

C PLOT NEGATIVE HYPERGEOMETRIC DISTRIBUTION
C
IF ( NGPH .LE. 0 ) CALL PLOT (NTP)
GO TO 1000
5000 WRITE (6,8)
CALL EXIT
1 FORMAT (18A4/515,2F10.0,2I5)
2 FORMAT (110,3F30.7)
17///)
4 FORMAT (////5X,5HSCORE,22X,4HH(X),26X,6HN\=H(X),24X,6HCUM HX//)
5 FORMAT (1H1,18A4//)
7 FORMAT (43H KR21 NEGATIVE OR ZERO ..., DATA SET ABORTED,5X,6HKR21 =
110.4)
8 FORMAT (1H:,20X,19HALL INPUT PROCESSED)
9 FORMAT (/7H XBAR =F12.3//7H VAR =F12.3)
10 FORMAT (/4X,3HN =18)
END
SUBROUTINE NFGHGR (K,A,B,C,S,NSUB,HX,HFX)

NEGATIVE HYPERGEOMETRIC FUNCTION

IF ( K .EQ. 0 ) GO TO 150
SLUG1=0.
SLUG2=0.
SLUG3=0.
SLUG4=0.
DO 100 I=1,K
SLUG1=SLUG1+AL0G10(S-1+1.)
SLUG2=SLUG2+AL0G10(A+I-1.)
SLUG3=SLUG3+AL0G10(B+I-1.)
SLUG4=SLUG4+AL0G10(FLOAT(I))
HX=C*10.**(SLUG1+SLUG2-SLUG3-SLUG4)
HFX=HX*NSUB
RETURN
150 HX=C
GO TO 125
END

SUBROUTINE RDFMT(FMT)

SUBROUTINE FOR INPUTTING VARIABLE FORMAT

INPUT STRUCTURE
FORMAT (ENCLOSED IN PARENTHESES) COL 1-72
CONTINUE ON CARD 2 IF NECESSARY
CONTINUE ON CARD 3 IF NECESSARY
ETC.
MAXIMUM NUMBER OF FORMAT CARDS IS 9
'END OF FORMAT' NECESSARY ... PUNCH IN COLUMNS 1 - 15

DIMENSION FMT(200)
DATA END/3HEND/
N=1
DO 100 I=1,10
M=N+17
READ (5,1) (FMT(J),J=N,M)
IF ( FMT(N) .EQ. END ) RETURN
100 N=N+18
WRITE (6,2)
STOP
1 FORMAT (18A4)
2 FORMAT (3/H *** EXCESSIVE NUMBER OF FORMAT CARDS)
SUBROUTINE POOL (NSUB, NITEMSINSAM, NTP, XBAR, VAR, NFMT)

DETERMINATION OF POOLED ESTIMATE OF POPULATION MEAN TEST SCORE AND VARIANCE

COMMON P(500), X(500)
DIMENSION FMT(200)
IF ( NFMT .EQ. 0 ) CALL RDFMT(FMT)
WRITE (6,1)
NTEST=NSUB*NITEMS
SESTM=0.
SESTV=0.
NSM=0
SWGHT=0.
DO 1000 J=1,NSAM
IF ( NFMT .NE. 0 ) CALL RDFMT(FMT)
IF ( NTEST .EQ. 0 ) READ (5,6) NSUB,NITEMS
SY=0.
SY=0.
P(J)=0.
RD 1000 J=1,NSUB
READ (5,FMT) (X(K),K=1,NITEMS)
Y=0.
DO 110 K=1,NITEMS
P(K)=P(K)+X(K)
Y=Y+X(K)
SY=SY+Y
SY=SY+Y
SY=SY+Y
XBAR=SY/NSUB
VR=(SY+SY*SY/NSUB)/NSUB
SPQ=0.
DO 120 J=1,NITEMS
PP=P(J)/NSUB
SPQ=SPQ+PP*(1.-PP)
NSM=NSM+NSUB
SWGHT=NSUB*NITEMS
ESSTM=ESSTM+ESSTM*SWGHT
ESTV=ESTV+ESTV*SWGHT
SWGHT=SWGHT*SWGHT
WRITE (6,2) I,ESTM,ESTV
CONTINUE
XBAR=ESSTM/SWGHT
VAR=ESTV/SWGHT
IF ( NSM .LT. 500 ) VAR=VAR*(NSM-1.)/NSM
WRITE (6,3) XBAR,VAR
RETURN
FORMAT ('//24X,21HESTIMATE OF PARAMETER//5X,6HSAMPLE,10X,4HMEAN 1,16X,8HVARIANCE//')
FORMAT ('//110,2F20.7')
FORMAT ('//14H POOLED XBAR =F20.7//18H POOLED VARIANCE =F16.7//')
FORMAT (215)
END
SUBROUTINE PLOT (NITEMS)
REAL N
COMMON N(500), P(500)
DIMENSION BCD(10)
DATA BGD/1HU, 1H1, 1H2, 1H3, 1H4, 1H5, 1H6, 1H7, 1H8, 1H9/
DATA BLK, DOT, XX/1H #1H..1HX/

LOCATE MAXIMUM VALUE FOR H(X)

NNN=NITEMS+1
T=0.
DO 50 I=1, NNN
   IF ( P(I) .GT. T ) T=P(I)
50 CONTINUE

DETERMINATION OF APPROPRIATE SCALE FACTOR FOR H(X) PLOT

J=0
DO 60 I=1, 6
   K=I-1
   J=10.*K
   IF ( J .EQ. 0 ) GO TO 60
   J=K-1
   WRITE (6,3) J
60 CONTINUE

SCALE H(X) BEFORE PLOTTING

DO 75 I=1, NNN
   P(I)=P(I)*10.*J
75 CONTINUE

LABEL ORDI NATE

WRITE (6,1)
DO 500 I=1, 100
   N(I)=BLK
   N(101)=BCD(2)
   WRITE (6,2) (N(J), J=1, 101)
500 CONTINUE

DO 750 I=1, 101
   N(I)=BCD(I)
750 CONTINUE

CONTINUE
N(101)=BCU(1)
WRITE (6,2) (N(J),J=1,101)
DO 595 I=1,101
줄
595 N(I)=DOT
WRITE (6,2) (N(J),J=1,101)

C PLOT VALUES OF SCALED H(X)

DO 100 I=1,NNN
줄
100 N(I)=DOT
WRITE (6,2) (N(J),J=1,101)
RETURN

1 FORMAT (1H1,50X,40HPROPORTION OF POPULATION RECEIVING SCORE//).
2 FORMAT (14X,101A1)
3 FORMAT (///5X,19H(X) SCALED BY 10 **13,9H IN GRAPH//).
4 FORMAT (110,4X,101A1)
END
### Sample Output

<table>
<thead>
<tr>
<th>Card</th>
<th>column</th>
</tr>
</thead>
<tbody>
<tr>
<td>00000000001111111111222222222233333333333344444444455555555556</td>
<td></td>
</tr>
<tr>
<td>123456789012345678901234567890123456789012345678901234567890</td>
<td></td>
</tr>
</tbody>
</table>

**FIRST YEAR WORD SPELLING PROJECT**

SHOEMAKER/UKADA

000050000000000000000000000000000000

(10X, 10F1.0)

END OF FORMAT

0001800010

01 1 1    1011111110
02 1 1    1111101011
03 1 1    1110100000
04 1 1    1110000000
05 1 1    0000000000
06 1 1    1101010001
07 2 1    1110101010
08 2 1    1100100000
09 2 1    0000011010
10 2 1    0100000000
11 2 1    1111101010
12 2 1    1001001000
13 3 1    0100100000
14 3 1    0111100010
15 3 1    0000000000
16 3 1    0000000000
17 3 1    1101000110
18 2 1    0000000000

0001400010

01 1 2    1111110011
02 1 2    1011100000
03 1 2    1100101000
04 1 2    1111110000
05 1 2    1011100010
06 1 2    1110110000
07 2 2    1010100000
08 2 2    1111110010
09 2 2    1101111010
10 2 2    1111110000
11 3 2    0001110000
12 3 2    1111111100
13 3 2    1101000000
14 3 2    0011110000
Appendix B

Computer Program For Simulating Multiple Matrix Sampling
The computer program for simulating multiple matrix sampling is described in detail in Chapter V. A listing of the Fortran IV program is given in this appendix for those readers who may want to implement the model on the computer configuration available to them. The program given herein was written originally for a UNIVAC 1108 and a modified version has been implemented on an IBM S360/91. In modifying the program for the S360/91, the only changes made were those involving the uniform (.00 to .99) random number generator RUNIF. On the 1108, RUNIF is initialized by RINITL. Calling RINITL with BASE as the argument causes BASE to be used as the starting value or seed in the algorithm used by RUNIF in generating uniform random numbers. Because RINITL is specific to UNIVAC 1108, readers should consult the local computing center staff to determine the subprogram and calling procedures at that installation comparable to the RINITL/RUNIF system. The conversion process was relatively simple for the S360/91 and it is anticipated that such will be the case with other hardware and software systems. Input values to the program are made on one parameter card. The organization of the card is described at the beginning of the program listing. Examples of parameter cards are found on page 97 of this appendix. Sample output from the program is given in Chapter V.
COMPUTER SIMULATION OF ITEM-EXAMINEE SAMPLING
DAVID M. SHOEMAKER

PARAMETER CARD (THERE IS JUST ONE)

COLUMNS (ALL INTEGERS RIGHT-JUSTIFIED)

01-03 INTEGER NUMBER OF ITEMS IN TOTAL TEST

04-09 DESIRED MEAN TEST SCORE IN POPULATION
(MUST BE SPECIFIED, WITH DECIMAL POINT PUNCHED ON CARD)

10-15 DESIRED VARIANCE OF TEST SCORES IN POPULATION
(WITH DECIMAL POINT PUNCHED ON CARD)

NOTE ... IF VARIANCE IS OMITTED, RELIABILITY MUST BE SPECIFIED.

16-21 DESIRED VARIANCE OF ITEM DIFFICULTY INDICES OVER POPULATION OF EXAMINEES. THE ITEM DIFFICULTY INDEX FOR ITEM I IS THE PROPORTION OF EXAMINEES ANSWERING ITEM I CORRECTLY.
(MUST BE SPECIFIED WITH DECIMAL POINT PUNCHED ON CARD)

WITH SKewed DISTRIBUTIONS, VARIANCE OF ITEM DIFFICULTY INDICES IS ASSUMED TO BE EQUAL TO ZERO.

22-27 DESIRED RELIABILITY OF TEST SCORES IN POPULATION
(WITH DECIMAL POINT PUNCHED ON CARD)

NOTE ... IF RELIABILITY IS OMITTED, VARIANCE MUST BE SPECIFIED.

28-31 INTEGER NUMBER OF SUBTESTS IN ITEM-EXAMINEE SAMPLING

32-35 INTEGER NUMBER OF ITEMS PER SUBTEST
(CONSTANT ACROSS SUBTESTS)

36-39 INTEGER NUMBER OF EXAMINEES PER SUBTEST
(CONSTANT ACROSS SUBTESTS)

40-43 INTEGER NUMBER OF INDEPENDENT REPLICATIONS OF ITEM-EXAMINEE SAMPLING PLAN

44 SAMPLING PLAN FOR ITEMS
0 = SAMPLING WITH REPLACEMENT
(USED WHEN TK IS GREATER THAN K)

1 = SAMPLING WITHOUT REPLACEMENT
(USED WHEN TK IS LESS THAN OR EQUAL TO K)
2 = SAMPLING WITHOUT REPLACEMENT BUT SUBJECT TO RESTRICTION THAT ITEMS OCCUR WITH EQUAL FREQUENCY AMONG SUBTESTS (USED WHEN TK IS GREATER THAN K)

INTERMEDIATE PRINTOUT OPTION
0 = NO INTERMEDIATE PRINTOUT
1 = INTERMEDIATE PRINTOUT WANTED

NEGATIVE HYPERGEOMETRIC DISTRIBUTION OPTION
0 = NO NEG. HYPER. DIST. WANTED
1 = COMPUTE NEG. HYPER. DIST.

DEGREE OF SKEWNESS IN NORMATIVE DISTRIBUTION
1 = NORMALLY DISTRIBUTED
2 = POSITIVELY SKEWED
3 = NEGATIVELY SKEWED

SEED FOR UNIFORM RANDOM NUMBER GENERATOR (ODD NUMBER)

GENERATE ITEM DIFFICULTY INDICES
0 = GENERATE NEW ITEM DIFFICULTY INDICES
1 = USE ITEM DIFFICULTY INDICES GENERATED BY PREVIOUS DATA CARD

MAXIMUM NUMBER OF ITEMS IS 150 (EASILY MODIFIED, HOWEVER)
ITEMS SCORED DICHOTOMOUSLY

PROGRAM WILL PROCESS REPEATED PARAMETER CARDS (NUMBER LIMITED ONLY BY AMOUNT OF COMPUTER TIME ALLOCATED)

REAL N,M,MPOP
COMMON N(150),M(150),LT(3000)
COMMON /BLOCK1/ YVAR,YSU,MPOP,SPU,KPOP,KDIST,BASE,INTPRT
COMMON /BLOCK2/ RND(150)
COMMON /BLOCK3/ P(150),Q(150),NSUB

READ (5,1,END=5000) KPOP,MPOP,VPOP,PVAR,A20,NT, IPT, NSPT, NREPS,
1 ISAMP, INTPRT, NHPER, KDIST, BASE, ISAVE
IF ( KDIST .GT. 1 ) PVAR=0.
WRITE (6,2) BASE,MPOP,VPOP,KPOP,A20,PVAR,NT, IPT, NSPT, NREPS,
1 ISAMP, NHPER, KDIST, INTPRT

INITIALIZE RANDOM NUMBER GENERATOR (UNIQUE TO UCC)
CALL RINITL(BASE)
NSUB=0

C * CHECK ON PARAMETERS

IF ( ISAMP .NE. 1 ) GO TO 30
IF ( NT*PT .GT. KPOP ) GO TO 55
IF ( A20 .LT. 0. OR. A20 .GT. 1. ) GO TO 55
IF ( NDIS1 .GT. 3 .OR. NDISI .LT. 1 ) GO TO 55
IF ( IFIX(A20*1000.) .EQ. 0 .AND. IFIX(VPOP*1000.) .EQ. 0 ) GO TO 55
IF ( IFIX(MPOP*1000.) .GE. KPOP*1000 ) GO TO 55
IF ( PVAR .LT. .2 ) GO TO 70

55 WRITE (6,3)
GO TO 2000

C COMPUTE NECESSARY PARAMETERS

TEMP=MPOP*(KPOP-MPOP)-KPOP*KPOP*PVAR
IF ( IFIX(VPOP) .EQ. 0 ) VPOP=TEMP/(KPOP-(KPOP-1.)*A20)
SPOP=SORT(VPOP)
WRITE (6,13) SPOP

C IF ( ISAVE .EQ. 1 ) GO TO 102

C GENERATE ITEM DIFFICULTY INDICES (PROPORTION OF EXAMINEES ANSWERING ITEM CORRECTLY)

IF ( IFIX((KPOP-MPOP)*1000.) .EQ. 0 ) GO TO 55
D1=1000.
PBAR=MPOP/KPOP
DO 173 I=1,KPOP
173 Q(I)=0.
IF ( IFIX(PVAR*1000.) .GT. 0 ) GO TO 66
DO 65 I=1,KPOP
65 P(I)=PBAR
GO TO 102
66 PSD=SORT(PVAR)

DO 100 IJ=1,100
DO 74 I=1,KPOP
CALL RANUNU (Z)
Q(I)=Z*PSD +PBAR
IF ( Q(I) .LT. 0. ) Q(I)=0.
IF ( Q(I) .GT. 1. ) Q(I)=1.

C DETERMINE INITIAL MEAN AND VARIANCE OF GENERATED ITEM DIFFICULTY INDICES

SP=0.
SPP=0.
DO 81 I=1,KPOP
PP=Q(I)
SP=SP+PP
SPP = SP + PP
PVR = (SPP - SP * SP / KPOP) / KPOP
CVR = SQRT(PVAR / PVR)

SCALE VARIANCE OF ITEM DIFFICULTY INDICES TO STANDARD

SP = 0,
SPP = 0,
DO 82 I = 1, KPOP
Q(I) = Q(I) * CVR
IF (Q(I) .GT. 1.) Q(I) = 1.
P = Q(I)
SP = SP + PP

82 SCALE MEAN OF ITEM DIFFICULTY INDICES

SP = 0,
SPP = 0,
DO 84 I = 1, KPOP
Q(I) = Q(I) - D4
IF (Q(I) .LT. 0.) Q(I) = 0.
IF (Q(I) .GT. 1.) Q(I) = 1.
PP = Q(I)
SP = SP + PP

84 SCALE MEAN OF ITEM DIFFICULTY INDICES

PVR = (SPP - SP * SP / KPOP) / KPOP
PhiR = SP / KPOP
D2 = ABS(PVAR - PVR)
D3 = ABS(PBAR - PBH)
I1 = (D1 + .0005) * 1000.
I2 = (D2 + .0005) * 1000.
I3 = (D3 + .0005) * 1000.
IF (I2 .LE. 5 .AND. I3 .LE. 5) GO TO 103
IF (I2 .GE. I1) GO TO 100
D1 = D2
DO 90 I = 1, KPOP

V0 DO 104 I = 1, KPOP
104 P(I) = Q(I)
NSTOP = KPOP - 1
DO 110 I = 1, NSTOP
J = I + 1
DO 110 J = JJ, KPOP
IF (.IFIX(P(I) * 1000.) .GT. IFIX(P(J) * 1000.)) GO TO 110
TEMP = P(J)
P(J) = P(I)
P(I) = TEMP
110 CONTINUE

102 IF (INTPRT,EQ.1) WRITE (6,4) (P(I), I = 1, KPOP)

COMPUTATION OF CONSTANTS FOR GENERATION OF LOGNORMAL DISTRIBUTION

YSU = 0.
IF ( NDIST .EQ. 1 ) GO TO 111
IF ( NDIST .EQ. 3 ) MPOP=KPOP-MPOP
YVAR=ALOG(VPOP/(MPOP*MPOP)+1.)
YBAR=ALOG(MPOP)-YVAR/2.
YSD=SQR(T(YVAR)

COMPUTE ROUNING VALUES FOR EACH TEST SCORE INTERVAL

111 IF ( ISAVE .EQ. 0 ) CALL ROUND

REPLICATION OF ITEM-EXAMINEE SAMPLING PARADIGM

SX=0.
SX=0.
SX=0.
SX=0.
DO /000 IJK=1,NREPS
CALL ALLOC (NT, IPT, ISAMP)
IF ( INTPRT .EQ. 0 ) GO TO 113
WRITE (6,5)
J=0.
N1=NT*IPT
DO 112 I=1,N1,IPT
KK=I+IPT-1
J=J+1
112 WRITE (6,6) J, (LT(A), KFI, KK)
113 CALL POOL (NSPI, IPT, NT, XBAR, XVAR)
XSD=0.
IF ( XVAR .GT. 0. ) XSD=SQR(T(XVAR)
IF ( INTPRT .EQ. 1 ) WRITE (6,7) IJK, XBAR, XSD
SX=SX+XBAR
SX=SX+XSD
SX=SX+XBAR
SX=SX+SXXS+XSD

COMPUTATION OF CONSTANT FOR NEGATIVE HYPERGEOMETRIC DISTRIBUTION OPTION

IF ( NHPER .EQ. 0 ) GO TO 7000
A21=(KPOP/(KPOP-1.))*1.-XBAR*(KPOP-XBAR)/(KPOP-XVAR))
IF ( A21 .GT. 0. ) GO TO 120
WRITE (6,8) A21
GO TO 2000
120 A=(-1.*+1./A21)*XBAR
B=-A-1.**KPOP/A21
SLOG1=0.
SLOG2=0.
C=A+B
DO 140 I=1,KPOP
SLOGI=SLOGI+ALOG10(B=I+1.)
140 SLOG2=SLOG2+ALOG10(C=I+1.)
C=10.*SLOG1-SLOG2

GENERATION OF NEGATIVE HYPERGEOMETRIC DISTRIBUTION
WRITE (6,9)
N3=KP0P+1
CK=0.
L0 100 I=1,N3
   K=1
   (ALL NEGHUR(K,A,B,C,Hx))
   CK=CK+HX
   WRITE(6,10) K,Hx,CK
5000 CONTINUE

COMPUTE STANDARD ERRORS OF ESTIMATE OVER REPLICATIONS

BARS=SXS/NREPS
BAHM=SXM/NREPS
SES=SQRT((SXXS-SXS)*SXS/NREPS)/NREPS
SEM=SQRT((SXXM-SXM)*SXM/NREPS)/NREPS
   WRITE (6,11) BAHM,SEM,BARS,SES
GO TO 2000

EXIT GRACEFULLY.

5000 WRITE (6,12)
   CALL EXIT

FORMAT STATEMENTS

1 FORMAT (13,4F6.0,214,411,F6.0,12)
2 FORMAT (12H1PROBLEM NO.,F9.2//20H PARAMETERS INPUTTED///4X,4HMEAN,
      17X,F8.3//4X,SRHIVARIANCE,3X,F8.3//4X,1HK,10X,14//4X,SRHALPHA 2U,3X,
      2F8.3//4X,6HVAR(P),5X,F8.3//4X,27HITEM-EXAMINEE SAMPLING PLAN///
      34X,2HNT,5X,14 //4X,SHPFT,4X,14//4X,4HNSPT,3X,14//4X,9HNREPS,2X,
      414//9H SWITCHES///4X,1dHITEM-SAMPLING PLAN,8X,12//4X,23HNEGATIVE
      5 HYPERGEOMETRIC,3X,12//4X,22HNORMAL DISTRIBUTION,16//4X,21HINT
      6REPLICATE PRINTOUT,17//)

3 FORMAT (2H*** ERROR ON PARAMETER CARD///12X,2HOH//29H NEED MORE
      INFO ON PARAMETERS).

4 FORMAT (2H24H ITEM DIFFICULTY INDICES//(10F/.3))
5 FORMAT (///8H SUBTEST,5X,5HITEMS///)
6 FORMAT (1X,15.5X,2014//11X,2014)
7 FORMAT (///16H REPLICATION NO..15.5X,12HTEST OF MEAN,324.3/ 26X,26H
      1STI. OF STANDARD DEVIATION,F10.3///)
8 FORMAT (43H KR21 NEGATIVE OR ZERO ... DATA SET ABORTED,5X,6HKR21 =
      1(110.4)
9 FORMAT (///5X,5HSCORE,22X,4HH(X),26X,8HCUM H(X)//)
10 FORMAT (110.2F30.7)
11 FORMAT (30H AVERAGE POOLED MEAN OVER REPSF15,7//3X,27HSE OF POOLED
      1MEAN OVER REPSF15,7//30H AVERAGE POOLED SD OVER REPSF15,7//
      27HSE OF POOLED SD OVER REPSF15,7)
0 FORMAT (20H ALL INPUT PROCESSED)
1 FORMAT (7/71H COMPUTED SIGMA = 10.92///)
ND
SUBROUTINE POOL (NSPT, IPT, NT, XBAR, XVAR)
C DETERMINATION OF POOLED ESTIMATE OF POPULATION MEAN TEST SCORE AND
C VARIANCE

REAL MPOP
COMMON P(150), X(150), LT(3000)
COMMON /BLOCK1/ YBAR, YSD, MPOP, SP0, KPOP, NDIST, BASE, IPRT
DIMENSION TEST(150)
IF ( IPRT .EQ. 1 ) WRITE (6,1)
SESTM=0.
SESTV=0.
NSM=0
DO 1000 I=1, NT
SY=0.
SYY=0.
DO 508 K=1, IPT
P(K)=0.
ISTART=IPT*(I-1)+1
ISTOP=IPT*I
DO 500 J=1, NSPI
CALL DATA (TEST)
LL=0
DO 505 K=ISTART, ISTOP
KK=LT(K)
LL=LL+1
505
X(LL)=TEST(KK)
Y=0.
DO 510 K=1, IPT
T=X(K)
P(K)=P(K)+1
510
Y=Y+T
SY=SY+Y
SYY=SYY+Y*Y
XBR=SY/NSPT
VR=(SYY-SY*SY/NSPT)/NSPT
SPQ=0.
DO 520 J=1, IPT
PP=P(J)/NSPT
SPQ=SPQ+PP*(1.-PP)
NSM=NSM+1
ESTM=KPOP*XBR/IPT
ESTV=(NSPT*KPOP*((KPOP-1.)*VR-(KPOP-IPT)*SPQ))/(IPT*(IPT-1.)*
1(NSPT-1.))
SESTM=SESTM+ESTM
SESTV=SESTV+ESTV
ESIS=0.
IF ( ESTV .LT. 0. ) ESTS=-SQRT(ABS(ESTV))
IF ( IPRT .EQ. 1 ) WRITE (6,2) I, ESTM, ESTV, ESTS
1000 CONTINUE
XVAR=SESTM/NSM
XVAR=SESTV/NSM
N=NSPT*NSM
IF ( M ,LT, 500 ) XVAR=XVAR*(M-1.)/M
RETURN
FORMAT ( '///38X,22H ESTIMATE OF PARAMETER///5X,6HSAMPLE,10X,
14HMEAN,16X,8HVARIANCE,12X,12HSTANDARD DEVIATION ///' )
FORMAT ( 110,3F20.7)
END
SUBROUTINE DATA (X)

GENERATION OF ITEM SCORES AND TEST SCORE FOR HYPOTHETICAL EXAMINEE

REAL MPUP
INTEGER TSCORE
COMMON /BLOCK1/ YBAR, YSD, MPUP, SPOP, KPOP, NDIS, BASE, INTPR
COMMON /BLOCK2/ RND(150)
COMMON /BLOCK3/ P(150), Q(150), NSUB
DIMENSION X(150)

GENERATE TOTAL TEST SCORE

NSUB=NSUB+1
CALL RANDNU (7)
GO TO 10 (215, 220, 220), NDIS
10 TEMP=Z*SPOP+MPOP
215 GO TO 230
220 TEMP=EXP(Z*YSD+YBAR)
230 IF ( TEMP .LT. 0. ) TEMP=0.
IF ( TEMP .GT. FLOAT(KPOP) ) TEMP=KPOP
KK=TEMP+1.
IF ( KK .GT. KPOP ) KK=KPOP
TSCORE=TEMP+(KK-RND(KK))
IF ( TSCORE .LT. 0 ) TSCORE=0
IF ( TSCORE .GT. KPOP ) TSCORE=KPOP
IF ( NDIS .EQ. 3 ) TSCORE=KPOP-TSCORE

GENERATE ITEM SCORES FOR EXAMINEE

DO 240 J=1,KPOP
240 X(J)=0.
IF ( TSCORE .EQ. 0 ) GO TO 300
IF ( TSCORE .LT. KPOP ) GO TO 248
DO 242 J=1,KPOP
242 X(J)=1.
GO TO 300
248 KOUNT=0
DO 250 J=1,KPOP
250 IF ( IFIX(Q(J)*1000.) .GT. IFIX(P(J)*1000.) ) GO TO 250
KOUNT=KOUNT+1
IF ( KOUNT .GT. TSCORE ) GO TO 300
X(J)=1.
CONTINUE
DO 260 J=1,KPOP
260 IF ( IFIX(X(J)) .EQ. 1 ) GO TO 260
KOUNT=KOUNT+1
IF ( KOUNT .GT. TSCORE ) GO TO 300
X(J)=1.
CONTINUE
300 DO 320 J=1,KPOP
320 Q(J)=(Q(J)*(NSUB-1.)*X(J))/NSUB
RETURN
END
SUBROUTINE NEGHGR (K, A, B, C, HX)

NEGATIVE HYPERGEOMETRIC FUNCTION

REAL N, M, MPOP
COMMON N(150), M(150), LT(3000)
COMMON /BLOCK1/ YBAR, YSD, MPOP, SPOP, KPOP, NDIST, BASE, INTPRI
IF ( K .EQ. 0 ) GO TO 150
S = KPOP
SLOG1 = 0.
SLOG2 = 0.
SLOG3 = 0.
SLOG4 = 0.
DO 100 I = 1, K
SLOG1 = SLOG1 + ALOG10(S - I + 1.)
SLOG2 = SLOG2 + ALOG10(A + I - 1.)
SLOG3 = SLOG3 + ALOG10(B - I + 1.)
100 SLOG4 = SLOG4 + ALOG10( FLOAT(I) )
HX = C * 10. ** ( SLOG1 + SLOG2 - SLOG3 - SLOG4 )
RETURN
150 HX = C
RETURN
END
SUBROUTINE ALLOY (NT, IPT, ISAMP)

RANDOM ASSIGNMENT OF ITEMS TO SUBTESTS

REAL MPOP
COMMON X(300), LT(3000)
COMMON /BLOCK1/ YBAR, YSD, MPOP, SPOP, KPOP, NDIST, BASE, INTPT
DIMENSION L(150), KNTR(150)

DO 100 I = 1, KPOP
  KNTR(I) = 0
100 L(I) = I

NN = NT * IPT

IF ( ISAMP .EQ. 2 ) GO TO 200

K = 0
DO 150 I = 1, NN
  R = RUNIF(BASE)
  JJ = R * KPOP + 1.
  IF ( JJ .LT. 1 ) JJ = 1
  IF ( JJ .GT. KPOP ) JJ = KPOP
  IF ( L(JJ) .GT. 0 ) GO TO 170
  GO TO 165

165 LT(I) = L(JJ)
  K = K + 1
  IF ( ISAMP .NE. 1 ) GO TO 180
  L(JJ) = -L(JJ)
  GO TO 150

170 LT(I) = L(JJ)
  K = K + 1
  IF ( ISAMP .NE. 1 ) GO TO 180
  L(JJ) = -L(JJ)
  GO TO 150

180 IF ( K .LT. IPT ) GO TO 185
  K = 0
  DO 245 II = 1, KPOP

185 L(II) = IABS(L(II))
  GO TO 150

190 CONTINUE
RETURN

200 NMULT = NN / KPOP
  IF ( IFIX((FLOAT(NN) / KPOP) * 10.) .NE. IFIX(FLOAT(NMULT) * 10.) ) GO TO 400
  K = 0
  NSTOP = NN / IPT
  DO 300 I = 1, NSTOP

210 R = RUNIF(BASE)
  JJ = R * KPOP + 1.
  IF ( JJ .LT. 1 ) JJ = 1
  IF ( JJ .GT. KPOP ) JJ = KPOP
  IF ( L(JJ) .GT. 0 .AND. KNTR(JJ) .LT. NMULT ) GO TO 220
  GO TO 210

220 LT(I) = L(JJ)
  K = K + 1
  KNTR(JJ) = KNTR(JJ) + 1
  IF ( K .LT. IPT ) GO TO 250
  K = 0

RETURN
LC 230  J=1,KPOP
L(J)=IABS(L(J))
GO TO 300
L(JJ)=-L(JJ)
CONTINUE
GO 350  I=1,KPOP
IF ( KNTH(I) .EQ. NMULT ) GO TO 350
NSTUP=NSTOP+1
LT(NSTOP)=1
CONTINUE
RETURN
IF ( NN .GT. KPOP ) ISAMP=0
IF ( NN .EQ. KPOP ) ISAMP=1
WRITE (6,1) ISAMP
GO TO 130
FORMAT (29H TK NOT INTEGER MULTIPLE OF K//30H ITEM-SAMPLING SWITCH
RES*T TO15//)
END
SUBROUTINE RANDN((X)
REAL MPOP
COMMON /BLOCK1/, YBAH, YSD, MPOP, SPOP, KPOP, NDIST, BASE, INTPK
DIMENSION C(290), C1(90), C2(65), C3(45), C4(60), C5(10)
EQUIVALENCE (VAIL, C(200)), (C1(1), C(1)), (C2(1), C(91)), 
1 (C3(1), C(176)), (C4(1), C(221)), (C5(1), C(281))
DATA C1/
1 2,2,3,3,3,3,3,3,5,6,6,6,6,6,6, 
2 6,6,6,8,8,8,8,1,1,1,5,0,0,
3 0,0,0,0,0,0,1,1,1,1,1,1,
4 1,1,1,1,2,2,2,2,3,3,3,3, 
5 4,4,4,4,4,4,4,5,5,5,5,5, 
6 5,6,7,7,7,7,7,8,8,8,9, 
7 9,9,9,9,1,1,1,1,1,1,1,1,2,1,2, 
8 1,2,1,3,1,3,1,4,1,4,1,5,1,6,1,7,1,8,9, 
9 4,4,4,7,9,9,9,9,9,1,1,1,1,1,1,1,1,1/
DATA C2/
1 1,3,1,3,1,3,1,3,1,3,1,3,1,4,1,4,1,4,1,6,1,6,6,6, 
2 1,6,1,6,1,6,1,6,1,7,1,7,1,7,1,8,1,9,1,9,1,9, 
3 1,9,1,9,1,9,1,9,1,9,1,9,2,2,2,2,2,2,2,2,2, 
4 2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2, 
5 2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2, 
6 7,1,1,1,3,4,1,1,9,1,4,9,8,6, 
7 5,1,2,1,6,1,7,3,1,5,2,1,8,2,2,2,2, 
8 2,9,2,3,2,4,2,1,1,2,7,0,0,2,6,2,8,2,9, 
9 9,9,43216507, 94,46409288, 94,9496939, 95,92578378, 95,95566764 /
DATA C3/
1 1,0,5,4895620, 96,1388536, 96,4198279, 96,6788825, 96,9367756, 
2 971393598, 97,4474970, 97,6942627, 97,9212915, 98,1233554, 
3 98,3249373, 98,5020795, 98,6448314, 98,7806989, 98,9110419, 
4 99,0207369, 99,1260517, 99,2236259, 99,3158205, 99,4021949, 
5 99,4845636, 99,5501310, 99,5889759, 99,6268373, 99,7300203, 
6 94,2278196, 94,5572077, 94,8551446, 95,1165313, 95,4986329, 
7 95,6691427, 96,0485017, 96,3804134, 96,6571775, 96,9819670, 
8 97,1291678, 97,4201251, 97,6132812, 97,8422883, 98,0579525, 
9 98,3065206, 98,4242076, 98,6325151, 98,7141582, 98,8832851 /
DATA C4/
1 9,89490775, 99,0781611, 99,1730598, 99,3063286, 99,3813410, 
2 99,4262546, 99,5110801, 99,5805552, 99,6077866, 99,6413834, 
3 9,73,996,992,920,998,982,990,996,985,959, 
4 9,42,994,986,985,890,988,980,983,977,843, 
5 9,73,975,974,978,755,970,5,971,976,968,967, 
6 12,8,2,8052333, 6,9186539, 60,9,325579, 
7 4,6444448, 6,5,4086308, 10,1,111111,1,4,735714, 
8 16,6666666, 7,5104139, 5,5743498, 5,2288610,25, 
9 5,9645244, 4,3951201, 4,9208132, 3,9631786,33,333333 / 
DATA C5/
1 1,3,4427955, 3,7,748844, 3,6020289, 4,1690656,50, 
2 3,1592514,100,3,2956424,3,0324898,2,9143782 /
SGN=1,
U=RUNIF(BASE)
IF ( U .LT. .5 ) SGN=-1.
U=RUNIF(BASE)
IV2=1000.*U
IV1=IV2/10
V=100.*U-.1*FLOAT(IV2)
IF ( U .GE. .79 ) GO TO 10
X=(C(IV1+1)+V)*SGN
RETURN
10 IF ( U .GE. .94 ) GO TO 20
X=(C(IV2-7/0)+V)*SGN
RETURN
20 IF ( U .GE. VTAIL ) GO TO 30
J=1/0
21 J=J+1
IF ( U .GE. C(J) ) GO TO 21
IF ( U .LT. C(J+30) ) GO TO 23
U=RUNIF(BASE)
X=(C(J-30)+.1*U)*SGN
RETURN
23 U=RUNIF(BASE)
V=RUNIF(BASE)
U1=AMIN1(U,V)
U2=AMAX1(U,V)
IF ( U2 .GE. C(J+60) ) GO TO 25
24 X=(C(J-30)+.1*U1)*SGN
RETURN
25 W=-.5*(.1*U1-.1)*((.1*U1-.1)2 C(J-30)+.1*U1+.1)
IF ( (EXP(W)-1.)*C(J+90)-U2+U1 ) 23,23,24
30 U1=RUNIF(BASE)
U2=RUNIF(BASE)
S=U1+U1+U2+U2
IF ( S .GE. 1. ) GO TO 30
T=SORT((9.-2.*ALOG(S))/S)
IF ( U1+1 .LE. 3. ) GO TO 32
X=U1*T*SGN
RETURN
32 IF ( U2+1 .LT. 3. ) GO TO 30
X=U2*T*SGN
RETURN
END
SUBROUTINE ROUND

ROUNING SUBROUTINE FOR NORMAL AND LOGNORMAL FUNCTIONS. AREA
BETWEEN ADJACENT INTEGER SCORES IS COMPUTED BY MEANS OF THE
TRAPEZOID FORMULA. THE ROUNING VALUE IS THAT CONTINUOUS TEST
SCORE SUCH THAT ONE-HALF OF THE AREA WITHIN THE SCORE INTERVAL IS
ABOVE THAT POINT.

REAL MPop
COMMON /BLOCK1/ YBAR, YSD, MPOP, SPOP, KPOP, NDISt, BASE, IPRT
COMMON /BLOCK2/ RND(150)
DIMENSION Y(101)

DELTA=.01
YVAR=YSD*YSD
VPOP=SPOP*SPOP
PI=3.141592/
GO TO (40,50,50), NDISt
40 CND1=1./SQRT(2.*PI*VPOP)
CND2=2.*VPOP
GO TO 60
50 CLN1=1./SQRT(2.*PI*YVAR)
CLN2=2.*YVAR
60 CONTINUE

SARea=0.
IF (IPRT,EQ.1) WRITE (6,7)
DO 100 I=1,KPOP
  N=I-1
DO 190 J=1,101
  K=J-1
  X=N+K*DELTA
  GO TO (130,131,131), NDISt
130 Y(J)=CND1*EXP(-(X-MPOP)**2)/CND2
  GO TO 150
131 IF (X,GT,0.) GO TO 135
  Y(J)=0.
  GO TO 150
135 Y(J)=(CLN1/X)*EXP(-(ALOG(X)-YBAR)**2)/CLN2
150 CONTINUE
IF (IPRT,EQ,1) WRITE (6,5) N,(Y(J),J=1,101,10)
N1=Y(1)*10000.
N2=Y(101)*10000.
IF (N1,GT,0.OR. N2,GT,0.) GO TO 153
RND(1)=N+5
GO TO 100
153 AREA=0.
DO 155 J=2,100
155 AREA=AREA+Y(J)
100
- 96 -

\[
\text{AREA} = \delta \left( \frac{y(1) + y(101)}{2} \right) \times \text{AREA}
\]

\[
\sin A = \text{AREA} + \text{AREA}
\]

\[
P = 0,
\]

\[
160 \quad J = 1, 100
\]

\[
K = J + 1
\]

\[
P = P + \left( \frac{y(j) + y(k)}{2} \right) \times \delta / \text{AREA}
\]

\[
\text{IF} \left( P > 1.5 \right) \text{GO TO 160}
\]

\[
\text{KNO}(I) = N + J \times \delta
\]

\[
\text{GO TO 100}
\]

\[
100 \text{ CONTINUE}
\]

\[
\text{WRITE (6,3)}
\]

\[
\text{CALL EXIT}
\]

\[
100 \text{ CONTINUE}
\]

\[
\text{IF ( IPRT .EQ. 0 ) RETURN}
\]

\[
\text{WRITE (6,6) AREA}
\]

\[
\text{WRITE (6,1)}
\]

\[
\text{GO 200 I = 1, KPOP}
\]

\[
J = 1 - 1
\]

\[
600 \text{ WRITE (6,2) J, KNO(I), I}
\]

\[
\text{RETURN}
\]

\[
\text{FORMAT (/ / 22H ROUNING RULE } \text{SCORE,5X} \text{ROUND,5X} \text{SCORE/) }
\]

\[
\text{FORMAT (/ 17X,15,F10.2,5X,15/}
\]

\[
\text{FORMAT (/ 34H PROBLEM IN ROUND SUBR EXIT CALLED/) }
\]

\[
\text{FORMAT (/ 1X,13,F11.4/) }
\]

\[
\text{FORMAT (/ 7H AREA } = \text{F15.7/) }
\]

\[
\text{FORMAT (/ / 23H DISTRIBUTION ORDINATES/}
\]

\[
13X,1H3X,2H0,6X,2H1,1H6X,2H2,6X,2H3,6X,2H4,6X,2H5,26X,2H6,6X,2H7,6X,2H8,6X,2H9,5X,3H1.0/)
\]

\[
\text{END}
\]
## Examples of Parameter Cards

<table>
<thead>
<tr>
<th>Card</th>
<th>Column</th>
</tr>
</thead>
<tbody>
<tr>
<td>0000000011111111122222223333333344444444555555555666666666677777777778</td>
<td>123456789012345678901234567890123456789012345678901234567890</td>
</tr>
<tr>
<td>0210</td>
<td>030</td>
</tr>
</tbody>
</table>

```
<table>
<thead>
<tr>
<th>Card</th>
<th>Column</th>
</tr>
</thead>
<tbody>
<tr>
<td>80 76</td>
<td>.00</td>
</tr>
<tr>
<td>80 40</td>
<td>.05</td>
</tr>
<tr>
<td>80 76</td>
<td>.00</td>
</tr>
<tr>
<td>60 40</td>
<td>.05</td>
</tr>
<tr>
<td>60 40</td>
<td>.00</td>
</tr>
<tr>
<td>80 40</td>
<td>.00</td>
</tr>
<tr>
<td>40 20</td>
<td>.05</td>
</tr>
<tr>
<td>40 20</td>
<td>.00</td>
</tr>
</tbody>
</table>
```