ED 404 374 TM 026 460

AUTHOR Beaton, Albert E.; Gonzalez, Eugenio
TITLE NAEP Primer.
INSTITUTION Boston Coll., Chestnut Hill, MA. Center for the Study of Testing, Evaluation, and Educational Policy.
SPONS AGENCY National Center for Education Statistics (ED), Washington, DC.
PUB DATE 95
CONTRACT R999B20005
NOTE 110p.; Computer disk not available from EDRS.
PUB TYPE Guides - Non-Classroom Use (055) -- Machine-Readable Data Files (MRDF) (102)

EDRS PRICE MF01/PC05 Plus Postage.
DESCRIPTORS *Achievement Tests; *Data Analysis; Databases; *Educational Trends; Elementary Secondary Education; Grade 8; Mathematics Tests; Private Schools; Public Schools; *Research Utilization; Sampling; Test Construction; *Test Use
IDENTIFIERS *National Assessment of Educational Progress; *Statistical Package for the Social Sciences; Trial State Assessment (NAEP)

ABSTRACT

This guide to the National Assessment of Educational Progress (NAEP) is designed to help the secondary data analyst use the NAEP and to introduce some of the sophisticated technology used by the NAEP. The NAEP has been gathering information on American students since 1969. It samples populations that consist of all students in U.S. schools, both public and private, at grades 4, 8, and 12, as well as ages 9, 13, and 17. NAEP data are designed for measuring trends in student performance over time and for cross-sectional analyses of the correlates of performance. Since the introduction of the Trial State Assessments in 1990, the NAEP has also been used to compare the performances of students in participating states. All data collected by the NAEP are available for the secondary user. This primer, which assumes that the user has a working knowledge of the Statistical Package for the Social Sciences, gets the user started on the simplified database and introduces a few special features of the NAEP. The examples use a set of 1,000 eighth graders assessed in mathematics. These mini-files are used to illustrate several basic NAEP analyses. Five appendixes present file layouts and variable information, as well as a guide to using the attached primer computer disk. (Contains 28 figures, 2 tables, and 46 references.) (SLD)

NAEP PRIMER

Albert E. Beaton
Eugenio Gonzalez

The NAEP Primer

ALBERT E. BEATON
EUGENIO J. GONZALEZ

The Center for the Study of Testing, Evaluation, and Educational Policy
Boston College

The NAEP Primer was developed under grant number R999B20005 of the National Center for Education Statistics.

Center for the Study of Testing, Evaluation, and Educational Policy
Boston College
Chestnut Hill, MA 02167
United States
+ 1 (617) 552-4521

Table of Contents

1. INTRODUCTION

The purpose of this Primer is to make the data from the National Assessment of Educational Progress (NAEP), commonly known as "the Nation's Report Card," more accessible to secondary data analysts who are interested in examining their own questions about the status and accomplishments of students in American schools. The NAEP database is very large and complex, so much so as to be daunting. Although the NAEP database is very well documented, the potential user must make a substantial commitment of time and effort to understand what is available and how to use it. The purpose of this Primer is to help such users get started on a small but interesting portion of the NAEP data. Furthermore, it is intended to familiarize a secondary user with some of the sophisticated technology used by NAEP.

NAEP has been gathering data on the performance of American students since 1969. Over the years, it has gathered data about the performance of students not only in reading, writing, mathematics, and science but also in other areas such as citizenship, geography, history, and the arts. NAEP collected data annually until the 1979-1980 school year, but the data are now collected biennially. Not only has data been collected on students' performances but also on their backgrounds, on their attitudes, on their schools and, at times, on their teachers.

The populations in which NAEP samples consist of all students in American schools, both public and private, at grades 4, 8, and 12 *as well as* ages 9, 13, and 17. Until 1983, NAEP sampled only ages 9, 13, and 17 but since then it has have *also* sampled grades 4, 8, and 12 which are the grades in which most of the 9, 13, and 17 year old students are located. NAEP reports results by both age and grade.

NAEP data are designed for measuring trends in student performance over time and for extensive cross-sectional analyses of the correlates of performance. Since the introduction of the Trial State Assessments in 1990, NAEP has also been used to compare the performances of students in participating states.

All data collected by NAEP are available for secondary users, subject to the maintenance of the confidentiality of the participating student, districts, and the states. NAEP results can be reported at the state level for the Trial State Assessment data only, and regionally or nationally for the rest.

The full NAEP database contains the responses to each test item, indicators of student performance on the various subject matter scales and sub-scales, and responses to questionnaire items. It also contains information about schools and, when available, information about the teachers of the students in the NAEP sample. In cases where the response to an open-ended question is judged by more than one rater, the responses of all raters are included in the full data file. In fact, the data files contain all of the data necessary to reproduce the calculations that appear in the NAEP reports. The NAEP data files do not contain information that would uniquely identify its participants.

Understanding all the intricacies of the NAEP data files is a formidable task, despite its thorough and detailed documentation. Besides the vastness of the NAEP files, there are a number of design details and technical sophistications that can mislead a potential user. For example, sampling both age and grade requires the secondary data analyst to decide whether he or she wishes to use an age sample or a grade sample and then to remove the students who are not members of the selected population. Another issue is the use of sampling weights: the NAEP data base may have fifty or more sampling weights per individual in order to facilitate the computation of standard errors using the jackknife method. Another complicating feature is the use of plausible values of student performance rather than standard test scores. Each of these features, as well as others, require careful thought and some sophistication on the part of the secondary user.

This Primer is designed to get potential users started quickly on a small but interesting part of the NAEP database. We assume that the reader has a working knowledge of intermediate statistics including regression analysis and the analysis of variance. We also assume that the reader has a working knowledge of SPSS, a commonly available statistical system for mainframe and personal computers. The strategy is to get the user started quickly on a simplified database and introduce him or her to a few of the special features of NAEP.

The examples included in the Primer will focus on a sample of eighth grade students who were assessed in mathematics in 1990. Data from 1000 students have been selected from the NAEP 1990 national assessment file and placed in a mini-file on a floppy disk. Thirteen-year-olds who are not in the eighth grade have been excluded from the sample. There are two such mini-files, one that contains data appropriate for policy analysis and one that is appropriate for psychometric analyses.

Using these mini-files, we will introduce the reader to several basic analyses of NAEP data using the plausible values. All example analyses are written in SPSS, and the programs are supplied on the enclosed floppy disk. The floppy disk also contains a program for post-processing output from SPSS analysis to improve population estimates. These mini-files introduce the reader to some of the analysis methods that should be used with NAEP data. The SPSS command file used to create these mini-files is available on floppy disk so that potential users who have access to the full NAEP database can select other mini-files from different subject areas or different variables for analysis.

2. DESIGN OF NAEP

The National Assessment of Educational Progress (NAEP) is a large, Congressionally-mandated survey of what students in public and private schools in the United States know and can do. It is designed to monitor changes in performance over time and to permit extensive cross-sectional studies of the correlates of student performance.

NAEP has introduced a number of technical innovations in order to fulfill its mission efficiently and accurately. The sampling plan was initiated by the Research Triangle Institute (RTI) and was further developed by Westat, Inc. The sampling plan was designed to give every student in the country a known probability of being assessed. Since its beginning, NAEP has used innovative testing technology; for example, the assessment exercises were administered by a tape recorder to allow students who were poor in reading to show their skills in other subject areas. The design of NAEP was modified substantially (Messick, Beaton, and Lord, 1983) in the 1983 assessment when NAEP introduced a number of psychometric innovations such as Balanced Incomplete Block (BIB) spiraling, item response theory (IRT) scaling, and scale anchoring. The NAEP design is now extending the use of performance assessment and introducing student portfolio assessment. As the times have changed, NAEP has adapted itself while maintaining basic comparability with the past.

Understanding some of these features is essential to understanding how to use and interpret NAEP results. Since this Primer focuses on the 1990 assessment, the major features of the NAEP 1990 design are presented here. The NAEP 1990 design is described in considerable detail in an Overview of the National Assessment of Educational Progress (Beaton and Zwick, 1992), in The Design of the National Assessment of Educational Progress (Johnson, 1992), in The NAEP 1990 Technical Report (Johnson and Allen, 1992) and in the Technical Report of NAEP's 1990 Trial State Assessment (Koffler, 1991). The designs of previous years are described in the NAEP Technical Reports (Beaton, 1987, 1988; Johnson and Zwick, 1990).

BACKGROUND AND GOVERNANCE

The governance of NAEP is complex and has changed over the years since NAEP first collected data in 1969. The National Assessment of Educational Progress Improvement Act of 1988 (P.L. 100-297) was passed by the United States Congress and requires that reading and mathematics be assessed at least every two years and that writing and science be assessed at least every four years. Congress assigned responsibility for NAEP policy guidelines to an independent National Assessment Governing Board (NAGB), appointed by the Secretary of Education. NAGB is comprised of state governors, chief state school officers, various educational policy makers, teachers, and members of the general public. The Commissioner of the National Center for Education Statistics (NCES) is responsible for the administration of NAEP. In 1990, the operation of NAEP was contracted to the

Educational Testing Service (ETS), which subcontracted sampling and field operations to Westat, Inc., and subcontracted the printing and distribution of materials, and the scoring and data entry of student responses to National Computer Systems, Inc. (NCS). Congress also provided for a Technical Review Panel to review and report on the NAEP technology.

MEASUREMENT INSTRUMENTS

Since 1969, NAEP has collected data in numerous subject areas and for many different populations. The subject areas assessed include reading, writing, mathematics, and science, which have been regularly assessed, and other subjects such as art, history, and consumer skills that have been assessed only occasionally. In 1990, reading, mathematics, and science were assessed at grades four, eight, and twelve, and at ages 9, 13, and 17. Reading, writing, mathematics, and science were also assessed in separate samples to report long-term trends in educational achievement. The mathematical proficiency of eighth grade public school students was assessed for the first time at the state level in the 1990 Trial State Assessment.

For each subject area that is assessed, NAEP must create exercises that measure student proficiency and questions that probe the students' attitudes and practices in that area. General background and attitude questions must be reviewed and renewed for the Student Background Questionnaire. Questionnaires must be developed for school principals and, at times, for teachers. Questionnaires must also be developed for excluded students, that is, students unable to be assessed using the NAEP instruments. Also, the administrative procedures must be developed and administrative records kept. High quality assessment exercises, questionnaires, and other information are essential for NAEP to fulfill its mission.

The NAEP subject-matter assessment exercises are developed through a consensus approach. National committees of teachers and subject matter experts develop the objectives for the assessment in a subject area, which become the assessment specifications. Assessment exercises are written according to these specifications; they may be open-ended or multiple-choice, or even fairly long essays or performance tasks. Exercises are submitted for committee review for appropriateness, and are examined for ethnic and gender sensitivity. The items are then pre-tested on samples of students for empirical evidence of their adequacy. The items that survive the vetting processes are then placed in an item pool for use in the assessment. The development of the content-area frameworks and innovative assessment methods are described by Mullis (1992).

NAEP regularly develops a large number of assessment booklets, some for quite different purposes. The main NAEP samples, as well as the students in the Trial State Assessment are assessed in a single subject area using booklets that contain written instructions and items. Some subject areas require special administration such as, for example, mathematical estimation in which the items must be timed individually. Other booklets are used for measuring long-term trends; these booklets exclusively contain items that have been used in past assessments and must be administered using the same timings and instructions as in the past. We cannot attempt to cover all of the NAEP variations here, and so we will focus here only on the "main" NAEP instrumentation and sampling.

The main NAEP assessment materials are assembled into booklets using a system called Balanced Incomplete Block (BIB) spiraling. The purpose of BIB spiraling is to allow a large sampling of the subject matter within an area while also limiting the time demands on individual students. BIB spiraling also makes it possible to study the relationship between each pair of items in a subject area, as well as context effect. Under this system, many different assessment booklets are printed and thus students in the same assessment session may be assessed in different subject areas (e.g., mathematics or science), or receive different booklets in a single subject area with different but overlapping items.

The NAEP item pool is large since broad coverage is necessary in each subject area. Using the item pool, assessment "blocks" or testlets are formed. These blocks are then assembled into assessment booklets. Each subject matter block contains a number of student exercises and is separately timed. For 9-year-old and fourth grade students, the timing of these blocks in 1990 was set at ten minutes whereas fifteen minute blocks were developed for the students who are 13- or 17-years-old or in the eighth or twelfth grades. A five minute block of specific background questions in the subject area is prepared for each age and grade level. Another block of student background and attitude questions is also formed; students at age 9/grade 4 are allowed ten minutes for this block while students at other ages and grade are allowed six minutes. An assessment booklet is composed of the general background block, a subject-matter specific block, and, typically, three subject matter blocks. The actual assessment time is, therefore, 45 minutes for age 9/grade 4 $(10+5+3*10)$, and 56 minutes for age 13/grade 8 and age 17/grade 12 students $(6+5+3*15)$.

Since some NAEP scales cover more than one age or grade level, some items must be developed that are appropriate for more than one age/grade level. For example, an item might be used in the fourth and eighth grade level tests. This allows to make comparisons on performance across agre/grade levels.

BIB spiraling places assessment blocks into booklets so that each block is paired with each other block in one and only one booklet. This can be shown best by example. For many subject areas, NAEP develops seven blocks of items, which are labeled A, B, C, D, E, F, and G. Seven booklets are then formed as shown in Figure 2-1.

Figure 2-1 BIB Spiraling Design used in NAEP

Booklet	*BLOCKS*				
1	X	Y	A	B	D
2	X	Y	B	C	E
3	X	Y	C	D	F
4	X	Y	D	E	G
5	X	Y	E	F	A
6	X	Y	F	G	B
7	X	Y	G	A	C

Block X in Figure 2-1 contains the general background and background questions and block Y contains the subject-area specific background questions. Each booklet is shown to contain three different blocks containing subject-matter, and each of these blocks appears once as the first, second, and third block of exercises of some booklet. Note that each subject matter block is paired with each other subject matter block in exactly one booklet.

After the booklets are printed, they are then "spiraled", or rotated, into random sequences. In 1990, reading, mathematics, and science booklets were mixed together in a random sequence before being packaged for shipment. The packaging resulted in each booklet being placed first, last, or anywhere in-between in approximately the same number of packages.

In the 1984 and 1986 assessments, NAEP booklets included blocks from different subject areas and so a student might receive, for example, a reading, a mathematics, and a science block in the same booklet. The advantage of this was the ability to compute the correlation among the performances in different subject areas. Unfortunately, combining blocks from different subject areas required printing a very large number of booklets which were administered to a small number of students. It also meant that many students took only a few items in any subject area. Since 1988, NAEP has focused booklets in one subject area, although blocks from different subject areas may be spiraled together for special purposes. When the booklets contain blocks from only one subject area, NAEP calls it Focused BIB Spiraling.

We note that in forming blocks there are several constraints. In situations where the main sample is to be used for trend estimates, some blocks of items are simply copied into new assessment forms and mixed with blocks of new items. The 1990 mathematics assessment was designed to be the first in a new trend series, and thus is not so encumbered. However, the NAEP scales cover more than one age or grade level, and so some items must be developed that are appropriate at different levels; for example, an item might be used at the fourth and eighth grade levels. The formation of blocks, therefore, involves a number of different issues that must be balanced.

In some assessments at some grade and age levels, a teacher of a sampled student may be asked to complete a questionnaire about his or her background, teaching methods, and then questions about the particular students who are taught. For example, in 1990, mathematics teachers of eighth grade students who were assessed in mathematics were given such questionnaires. The principal of each sampled school was also given a questionnaire about the school's practices and facilities.

POPULATIONS AND SAMPLES

Initially, when NAEP was under the direction of the Education Commission of the States (ECS), NAEP sampled 9-, 13-, and 17-year-old students and also out-of-school 17-year-olds and adults. Assessment data collected before the year 1983 can be used to estimate the performance of age cohorts but not the performance of students in various grades in school. Since 1983, NAEP has not only sampled ages 9, 13, and 17, but also the grades that most of these students are in, although these populations overlap considerably. At present, these are grades four, eight, and twelve. The definitions of age as well as the times of year in which the assessments take place have changed over the years, and so NAEP collects data from several "long-term trend" samples that have the same population definitions as the earliest data. When using data from other years, the secondary data analyst must take care to assure that the data compared over time actually use the same population definitions.

The 1990 NAEP sampling procedures are presented in detail in Rust (1992) and Rust and Johnson (1992). For national samples, the student populations of the United States are assigned to a sampling frame consisting of primary sampling units (PSUs). The PSUs are the Census Bureau's Metropolitan Statistical Areas or counties. Adjacent small counties may be merged to form a larger PSU. PSUs are selected from the sampling frame with known probabilities. The sample is stratified to ensure that four national regions are adequately represented. Within each PSU, an exhaustive search is done to update the list of schools and the available information about them. This is especially important since lists of private schools are not always complete. When the school list and information is updated, schools are selected with probability proportional to size. Finally, a list of eligible students--because they are in either a NAEP age population or a grade population--is developed, and students are randomly selected from this list.

It is important to realize that not all students have an equal probability of selection. In order to have adequate sample sizes for policy analyses, private school students are selected at three times the rate of public school students, and students in schools with large minority enrollments are selected at twice the rate of other students. However, the probability of selection of each student is known, and so NAEP provides sampling weights so that the data may be used for population estimates.

Some of the students in the NAEP sample were deemed unable to be assessed because of a handicapping condition or limited English proficiency and were excluded from participation in NAEP. For these students, school personnel were asked to fill out a form containing some background information about the student and the reasons for exclusion. The information collected on these students and their sampling weights are included in the NAEP files.

FIELD ADMINISTRATION

The administration of NAEP for the national samples was done by professional staff employed by Westat. This staff contacted the schools, assured proper within-school sampling, administered the assessment, distributed the teacher questionnaires, and shipped the resultant data to National Computer Systems (NCS), the subcontractor for scoring and data entry. For the Trial State Assessment, Westat provided extensive training for assessment administrators, but the administration was done by personnel supplied by the state departments of education. To assure proper quality control, Westat made unannounced visits to 50% of the schools on the day of the assessment. The field operations and data collection for the Trial State Assessment are described in detail in Caldwell, Slobasky, Moore, and Ter Maat (1992).

SCORING AND DATA ENTRY

NAEP has many open-ended and essay exercises that must be scored before being entered into computer files. The professional scoring procedures are described in Foertsch, Gentile, Jenkins, Jones, and Whittington (1992). The database formation is described in Rogers, Freund, and Ferris (1992).

ANALYSIS

An overview of the analysis phase is described in Allen and Zwick (1990). The analysis phase begins with extensive checking of all input data. Each item is examined to assure that it falls within the appropriate range and various quality control checks are performed. The sampling weights are produced by Westat and described in Johnson, Rust, and Thomas (1990).

The exercises from the various subject areas are then scaled using item response theory (IRT). In mathematics, five sub-scales are developed:

1. number and operations,
2. measurement,
3. geometry,
4. data analysis and statistics, and
5. algebra and functions.

The scales are developmental in that they span the three NAEP age/grade levels. Using all available data, a likelihood distribution for each student's proficiency on each sub-scale is estimated. Note that this is not simply computing a test score; this procedure acknowledges the uncertainty associated with measurement. Different students receive different items and so a simple test score is not appropriate, especially since we wish to generalize to a much larger population of mathematical proficiencies. The probability distribution for each student represents possible or "plausible" values for a student's performance if we could measure that performance perfectly. From this distribution, five plausible values are selected at random to be used in calculations of estimates for the NAEP population distribution and its parameters.

Overall mathematics plausible values are developed as a weighted composite of the sub-scale plausible values. The scales are anchored for interpretation (Beaton and Allen, 1992). The methodology is explained in general by Mislevy, Beaton, Kaplan, and Sheehan (1992). The scaling of the NAEP 1990 mathematics data is described in Yamamoto and Jenkins (1990). The use of plausible values is discussed in more detail in chapter four of this Primer.

The estimates of population parameters are then made for questionnaire items, test items, and the proficiency scales and sub-scales. These are organized in books of tables called "almanacs." These almanacs contain one page per item or scale and give an estimate of the proportion of the national population that would have made each specific response. These almanacs are also available in CD-ROM format for more recent NAEP assessments. Estimates are also made for the sub-populations on which NAEP reports such as regions of the country, genders, racial/ethnic groupings, and so forth. Each population estimate in these tables is presented with its standard error. The standard errors are computed using the jackknife method. The estimation procedures used in NAEP are described in detail in Johnson and Rust (1992) and Johnson, Rust, and Thomas (1992).

Reporting

The NAEP results are typically reviewed by NAEP staff and authors who have expertise in the specific subject areas being reported. Authors of reports may be experts in the subject area being studied from universities, schools, government agencies, or NAEP staff. The results are interpreted and, as necessary, additional analyses may be requested. The final document is extensively reviewed and revised before final publication.

The NAEP Database

The NAEP database is developed as data arrives at ETS and data are checked. When the data entry is completed, the database is then carefully documented and prepared for use by secondary analysts.

As mentioned above, the database contains all information from whatever source. There are different files for different samples, such as main assessment in mathematics or science or for the special trend samples. There are also special files with information about the sample of students who were excluded because of a handicapping condition or limited English proficiency. The 1990 file is documented by Rogers, Kline, Johnson, Mislevy, and Rust (1992).

3. THE NAEP PRIMER MINI-FILES

Even though it is well documented, the full NAEP database is huge and can be overwhelming to potential users. It is composed of data amassed since 1969, and includes thousands of test items, millions of proficiency estimates, and huge amounts of information on student backgrounds and attitudes as well as on their schools and teachers. This NAEP Primer cannot cover all of the information in the full data base, describing each variable and ways to use it. Instead, it will focus on one set of data that was collected in the 1990 eighth grade assessment of mathematics. Even this data file might be too complex for use on a personal computer, despite faster and more powerful computer systems, and so we will introduce the reader to a subsample of the variables and of the student records that can be easily analyzed on a personal computer.

For the examples in this chapter, we have taken one mini-sample of 1,000 students from the NAEP 1990 Mathematics eighth-grade assessment. This sample is in the enclosed NAEP Primer disk, along with the information about their contents. The purpose of this mini-sample is to help to familiarize the user with the NAEP data and with the special procedures required to use them appropriately. This NAEP mini-sample may be used freely since variables that might be used to identify individual students, teachers, and schools have been carefully excluded. We note that this mini-sample is capable of producing proper parameter estimates of the performance of the students in American schools, although, of course, using the full NAEP database would produce more precise estimates.

The mini-sample is in the form of rectangular data files that are appropriate for entry into and analysis by SPSS, or other commonly available statistical systems. Such statistical systems are relatively easy to use and make available a large number of statistical procedures for parameter estimation and data analysis. The program that created these files is available on the accompanying disk and will be discussed in Chapter 5. The reader should be able to make mini-files from the complete data, tailored to his or her own needs, by modifying this program.

The sample is presented in two separate and distinct files. The first mini-file, M08PS1.DAT, is designed for policy analysts and others who are interested in estimating and examining how students perform in school. This file contains information about student proficiency (plausible values), student backgrounds and attitudes, and information from the teacher questionnaire. It does not contain responses to individual cognitive items in mathematics or any other subject areas. The second mini-file, M08MS1.DAT, is designed for measurement specialists who are interested in studying the psychometric properties of the items in NAEP assessment. The measurement file contains the actual student item responses, mathematics composite and sub-scale plausible values, as well as a few demographic variables.

The two mini-files are a self-weighted sample of 1,000 students from the full NAEP files. The mini-files contain eighth grade students only, since the 13-year-old students in the full

NAEP sample who are not in the eighth grade have been removed. Eighth graders are included in the file whether or not they are 13-years-old. The students in each mini-file are randomly sorted in order to make sub-sampling easy; that is, for example, the file can be divided into 10 consecutive mutually exclusive sub-samples of 100 each, where the first 100 students and each successive 100 students is also a self-weighted sample from the NAEP full file. The policy and measurement mini-samples contain the same students and can be merged to form a single file.

The differential sampling weights in the full NAEP data base *must* be used with the full NAEP files but *should not* be used with these mini-files. In the full NAEP files, each student is assigned a sampling weight that is used in estimating population parameters and a set of weights that may be used in estimating their standard errors. The student sampling weight is inversely proportional to the probability that the student was selected for the sample. In practice, this means that students who had a higher probability of being selected have lower sampling weights, and vice versa. For example, students from inner cities were oversampled for NAEP, thus the full sample has a disproportionally large number of inner city students, but this is compensated for in analyses by assigning those students lower sampling weights.

The self-weighting feature of the mini-files eliminates the need for differential sampling weights. By sub-sampling students proportionally to their sampling weights, each student in the mini-sample has the same probability of being selected from the NAEP population, and thus all students have, in principle, the same sampling weight. When all sampling weights are equal, they have no effect on parameter estimates, although they may be used for other purposes, as we shall see below.

Using a self-weighted subsample simplifies analyses but does not compensate for the fact that the NAEP sampling plan is complex, and is not a simple random sample of students. Since the students within a school tend to be more similar than students from different schools, a sample of 1,000 students contains less information when schools are sampled than would a same-sized simple random sample of students. The effective sample size can be estimated using the design effect, which is the ratio of the error variance of the implemented NAEP sample design to what the error variance would have been if a same-sized simple random sample of students had been used. The median design effect for NAEP has been estimated to be between 1.11 and 1.86 for item statistics for various sub-groups of this population (Johnson & Allen, 1992), although the design effect should be smaller for many other variables. The effective sample size is estimated to be the actual sample size divided by the design effect. In this way, we estimate that, for the NAEP mini-sample, the thousand students in the NAEP mini-sample are effectively equivalent to approximately 800 students in a simple random sample.

Using SPSS, we can use the design effect for exploratory purposes to adjust the sample size and standard errors that SPSS prints out for many statistical analyses. The suggestion is to assign a constant sampling weight of .8 to each student in the sample. This does not affect parameter estimates but does affect their standard errors. The sample sizes that SPSS reports will be 80% of the actual sample sizes and error variances will be enlarged by approximately 25% and the standard errors by about 5%. The values of Student t-statistics and their associated probability statistics will also be adjusted accordingly.

It should be stressed that estimating standard errors using this simple method does not give optimum results. A design effect is computed by averaging the ratio of the standard error estimated by the jackknife method to the standard error estimated assuming simple random sampling. In fact, the ratio varies substantially for different parameter estimates and so adjustment by the design effect may be substantially off for a particular parameter estimate. Although the authors believe that this method is adequate for many purposes, including exploratory analysis, the jackknife method can be expected to give better results when important data interpretations are involved.

NAEP File Conventions

In preparing the mini-files, we have tried to make as few changes as possible from the way the data are presented in the full data base. We have done this because we expect the reader to work back and forth between the mini-files and the full data base; for example, simplifying the variable labeling in the mini-file would not help someone who had to use both. We have kept the variables, their coding, and their labeling the same as in the full file.

The mini-files are organized by student records. There may or may not be other students from the same school in the file, depending on the selection during the sub-sampling. For the most part, the information in the file comes from an assessment booklet that is collected from the students. Some information, however, comes from a questionnaire given to a student's mathematics teacher, if the teacher completed the questionnaire, and from a school questionnaire. Other variables come from administrative records used by the NAEP contractors, the Educational Testing Service (ETS) and Westat, Inc.

Missing Values

As with all surveys, a data analyst must be concerned with missing data. Some variables can have no missing values; for example, the school code, booklet number, and the region of the country are present for all students. The plausible values that are used for estimating the mathematics proficiency of populations of students are available for all student in these files. Many other variables may have missing data.

NAEP typically distinguishes among several different types of missing or inappropriate data. There are several conventions for coding missing values but, unfortunately, there are occasional exceptions to the general rules. The reader is advised to look up each variable in the codebooks for exceptions. The conventions are:

- Blanks: If a student did not have an opportunity to have a response for a variable, the field for that variable is left blank. Such blank fields are common with BIB spiraling where a student is administered only a sample of items but the files contain spaces for all items. As mentioned above, blanks are also used for teacher variables if a student's teacher did not respond to a questionnaire. Derived variables such as parents' education may also be blank if one or more of its components are missing. Blanks are converted by SPSS to its system missing value.

- Sevens: NAEP codes an "I Don't Know" response as a field of sevens, that is, "7" if the variable is coded in a one character field, "77" for a two character field, and so forth.
- Eights: If a student is administered an item and skips it, or there is no response marked on the booklet, the field corresponding to the variable is coded with a field of eights.
- Nines: For item responses in mathematics (or in other subject areas), NAEP fills the field with nines if the student did not reach the item. That is, all omitted cognitive items after the last item to which the student responded are coded as nines.
- Zeros: If a student gives more than one response where there should be only one, the field is coded as zeros.

Additional codes are used to indicate illegible, illiterate, and off-task codes for open-ended items.

The ability to discriminate among various types of invalid or inappropriate responses can add increased information to data analyses, but it also results in a complication that must be addressed in each analysis. The user also needs to be wary because these codes do not apply to variables that have no missing data, such as the plausible values and items such as "Size and Type of Community."

VARIABLE NAMING CONVENTIONS

As we mentioned previously, for the purposes of this Primer we could have changed the labels for the variables in the mini-file to something simpler and easier to remember, but have decided not to do so. We wish to keep the labels here the same as the labels in the full file so that the user can easily work back and forth between the two files. For the same reasons, we have also kept the NAEP conventions for missing data.

The NAEP variable labeling system is necessarily complex because it must allow a unique identifier for each item and derived variable that was used over many NAEP years, many subject areas, and many assessment forms and questionnaires. Where possible, NAEP uses a simple identifier such as REGION, which has values of 1=NORTHEAST, 2=SOUTHEAST, 3=CENTRAL, 4=WEST, AND 5=TERRITORY. This value of five cannot be present in this mini-file because territories are not part of the national sample, although they may be part of other samples. Other variables such as DRACE are not as simple because they are derived from several sources. The variable DRACE may have the values of 1=WHITE, 2=BLACK, 3=HISPANIC, 4=ASIAN, 5=AMERICAN INDIAN, and 6=UNCLASSIFIED. This variable combines several sources of information to form a single indicator of a student's race. There are no missing data codes for DRACE, although some students are not classified. The information from which DRACE was derived is available in the full data base.

There are so many variables in NAEP that simple, short mnemonic identifiers are impossible. Instead, an eight character code is developed for each item. These codes consist of a letter followed by a six digit number which is followed by a letter. The variable coding scheme is shown in Figure 3-1.

Figure 3-1 NAEP Item Naming Conventions[2]

Field Name	A short name (of up to eight characters) that identifies the field. This name is used consistently across all documentation, SAS & SPSS-X control files, and catalog files to identify each field uniquely within a data file. In general, nonresponse data field names are abbreviations of the field descriptions. Field names associated with response data are formatted as follows:
Position 1	Identifies nature/source of the response data: B = Common background item within common background block S = Subject-related background or attitude item (usually found within reading, writing, mathematics, and science cognitive blocks in the 1984 and 1986 assessments) N = Cognitive item within cognitive block (including reading, writing, mathematics and science cognitive items used in the 1984 and 1986 assessments) C = School questionnaire item T = Teacher questionnaire item X = Excluded student questionnaire item K = Science cognitive or background item M = Math cognitive or background item R = Reading cognitive or background item W = Writing cognitive or background item E = Math or science cognitive item for long-term trend blocks
Position 2 - 5	Identify an exercise (student files) or question (school, teacher, excluded student files). If position 1 is S or N, a zero in position 2 signifies a reading item, a 2 signifies a mathematics item, 4 a science item, and 6 a computer item.
Position 6 - 7	Identify a part within an exercise (student file) or a part within a question (student, teacher, excluded student files).
Position 8	Identifies the block containing an item (Student files only) to avoid duplicated naming of items that occur in more than one block. The numeric designation (1 through 12) has been replaced by an alphabetic one (A through L). This position is blank for questionnaire items and all other variables.

THE POLICY MINI-FILES

The Policy mini-file contains a selection of variables from the NAEP eighth grade sample that was administered the mathematics assessment. The file contains most of the information on the student and teacher questionnaires, but it does not contain any variables from other sources that might be used to identify any individual or school system.

In any data analysis, it is important that the researcher fully understand the nature of any variable that is used and how it was derived. NAEP has so many different questionnaire forms and assessment booklets that tracing the genealogy of an item can be difficult. In fact, since some NAEP cognitive items are kept confidential for future use, all of the items are

[2] From Rogers, A.M., Kline, D.L., Johnson, E.G., Mislevy, R.M., & Rust, K.F. (1990). *National Assessment of Educational Progress 1988 public-use data tapes version 2.0 user guide.* Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress.

not readily available for inspection. We cannot give the full background of the items in the Policy files here, but we will make suggestions as to how to find more information.

The layout of the records in the policy mini-file is shown in Appendix A. For each variable in the mini-file, the layout shows its NAEP identification code, and a 40 character description of the variable. The record layout also shows the starting position, ending position and length of each variable in the record as well as the number of decimal places. For variables with value labels (i.e., labels associated with each possible value of the variable), the values and their labels are also shown. Continuous variables such as the plausible values and the student's age do not have value labels.

For the most part, these variable labels, variable descriptions, and the associated value labels are sufficient to describe a variable, but some are not. For example, the first two variables in the mini-file, YEAR and AGE, are constants in this sample. YEAR is the year of the assessment and, since this sample was taken from the 1990 assessment, the value of the YEAR is "90" for all observations. Since this mini-file is a sample from the age 13/Grade 8 population, the variable AGE listed here will be 13 for all students. The actual ages of the students are recorded in the variable "DAGE" which is in the 35th and 36th characters of each student record.

The variables BOOK and SCH indicate the booklet number and the school code respectively. The booklet number can be used to tell which blocks of mathematics items were assigned to a student. The school code uniquely identifies each school in the sample but gives no further information about the school's identity.

The next few variables give general information that is derived from the assessments booklet's cover, Westat administrative files, or are derived from other variables. The first two of these indicate whether the student has an Individualized Education Plan (IEP), or Limited English Proficiency (LEP). We note that most IEP and LEP students were excluded from the assessment and that some basic data on these excluded students is available in a separate excluded student file in the main database. A small number of IEP and LEP students were deemed able to sit for the assessment and are included in this sample.

The variable COHORT has the value of "2" for all students and is completely consistent with the AGE variable above. NAEP labels the age 9/Grade 4 population as Cohort 1, the age 13/Grade 8 population as Cohort 2, and the age 17/Grade 12 population as Cohort 3.

SCRID is a scrambled student booklet number. This number identifies the actual (and unique) booklet that each individual student used. Since the original booklet number is scrambled, this number cannot be used for individual identification. This variable can be used to merge the cases from the policy and the measurement file.

DGRADE is the grade in school for the students in this sample. All students in the main NAEP file that were not in the eighth grade have been removed and so the value of this variable is eight for all students.

The next two variables, DSEX and DRACE, are variables that are derived from other variables. For the most part these values are taken from the student questionnaire or the student booklet cover. If the values for these variables are not present, the information is taken from other available student information.

The next two variables, REGION and STOC, identify the region of the country in which the student attended school and the size and type of the community in which the school is located.

The variable SEASON is necessary to distinguish between students who were tested in the Winter of 1990 and those who were tested in the Spring of that year. Those tested in the Spring had a few more months of education before the assessment.

The next variable is WEIGHT. This field in the main data file is for a differential sampling weight but, since the mini-sample is self-weighting, the sampling weight has been coded to a constant. We have set its value to .8 for each student to compensate for the complex NAEP sampling design (see Chapter 4).

The next variables, PARED (Parents educational level) and HOMEEN2 (Home Environment-Reading Materials) are derived from other variables. The Parents Educational Level is the highest level attained by either of the two parents. HOMEEN2 is derived from the students' responses to the questions B000901A (Does your family get a newspaper regularly?), B000903A (Is there an Encyclopedia in your home?), B000904A (Are there more than 25 books in your home?), and B000905A (Does your family get magazines regularly?).

DAGE is the student's actual age in years, as computed from the WESTAT records used in selecting the sample.

SINGLEP is a derived variable indicating the number of parents living at home with a student.

SCHTYPE is the type of school which is derived from the principal questionnaire. The possible values in NAEP are Public School, Private School, Catholic School, Bureau of Indian Affair School, and Department of Defense School. This sample contains only Public, Private and Catholic School students.

PERCMAT is an indicator of the student's perception of mathematics.

The next variables give information about the type of teaching certificate that a teacher has (TCERTIF), the teacher's majors at both the undergraduate (TUNDMAJ) and the graduate level (TGRDMAJ), and the number of mathematics courses that the teacher has taken (TMATCRS). These are followed by indicators of the teacher's emphasis on numbers and operations (TEMPHNO) and on probability and statistics (TEMPHPS). Basically, these are indicators that are derived from the teacher questionnaire.

The next two variables come from the School Questionnaire. SPOLICY indicates the number of recent changes in school policy and SPROBS is an indicator of the problems in the school.

IEP/LEP is an indicator of whether the student is either an IEP or LEP student.

CALCUSE is an indicator of whether the student used a calculator appropriately on the items in the calculator blocks.

IDP is an indicator of the instructional dollars spent per pupil, which is taken from the Quality Education Data, Inc. database. This variable refers to the money spent on students for books and supplies, not the actual money spent on education, which would include salaries, building maintenance, and other administrative costs.

The variable CAI, which indicates the availability of micro-computer assisted instruction, comes from the same source.

The next section contains plausible values for the various sub-scales. There are five plausible values for each sub-scale (numbers and operations; measurement; geometry; data analysis and statistics; and algebra and functions) and then five plausible values for the composite score. There are no value labels since they are continuous variables. These plausible values will be discussed in detail in Chapter 4.

The next two variables, MTHLOG and MRPLOG, are preliminary IRT scale scores and will not be discussed further in this Primer.

The record then contains a number of items from the student questionnaire. These are identified by their NAEP eight character identification code. These items are transcribed directly from the student questionnaire.

The next set of items have identification codes beginning with M which indicates that these items are from the mathematics questionnaire that was administered to students who were assessed in mathematics. Students assessed in other subject areas would have been administered a questionnaire specific to that subject. These mathematics items address such issues as the use of textbooks, worksheets, calculators, computers, and other features of mathematics education.

The final section of each record in the policy file is a series of questions taken from the questionnaire that was administered to the mathematics teachers of the students in the sample. The item identification code begins with a "T." These items probe the teacher's teaching experiences and teaching practices. All items from the teacher section of a record will be blank if the teacher of that particular student did not fill out a questionnaire.

THE MEASUREMENT FILE

The layout for the records in the Measurement File is shown in Appendix B. The format of the layout is similar to that of the Policy file in Appendix A.

The Measurement mini-file contains only a few file identification and demographic variables for each subject. The file identification codes are the same as for the Policy mini-file. The demographic variables are gender, race/ethnicity, region of the country, parents' educational level, and the student's age.

The plausible values for each sub-scale and the composite scale are reported next. The rest of each record contains student responses to items in the mathematics assessment. Each item has a unique identification code and position on the record. The item description gives an indication as to what the item would be like in the actual assessment but not enough to completely destroy the item's confidentiality.

The items in the Measurement file are scored either right, wrong, omitted, not reached, or not administered to the student. For any student record, most of the item responses are coded as not administered (i.e., blank) since the BIB-Spiraling design (see Chapter 2) assigns only three blocks of assessment items to a student.

These item responses were derived from the data in the full NAEP database. The full database contains the actual responses of each student, i.e., which of the possible responses that a student selected for a multiple-choice item. For the convenience of the secondary analysts, we have scored these items using the scoring key as either right or wrong, depending on if the student gave a valid response.

The Measurement file contains only one rating for an extended response item. The main data file contains the ratings for all raters if more than one individual rated the item response.

4. Plausible Values

Introduction

As mentioned above, NAEP does not produce an ordinary test score to represent an individual student's performance. Instead it produces a set of five plausible values for each student in each of the assessed areas. Plausible values improve the estimation of population parameters but at the cost of additional computational requirements. In this chapter, we will present the rationale for plausible values, the rules for using them in data analyses, and present several examples using the NAEP Policy Mini-sample data that illustrate how to use plausible values in statistical analysis. The chapter will give the details of a general method for statistical inference using plausible values, followed by two short-cut procedures. The short-cut procedures will be useful only for certain types of parameter estimates. The first short-cut procedure will be exact, but limited to estimates of a single parameter, such as a mean or a regression coefficient; the second will be approximate, but appropriate for simultaneous parameter estimation, such as in an analysis of variance.

Plausible values were developed during the analysis of the 1983-84 NAEP data in order to improve estimates of population distributions. Under BIB spiraling in 1984, students were presented with three 14-minute blocks of exercises, each block consisting of either reading or writing items. A student might receive zero, one, two, or three reading blocks and the remaining blocks in a booklet, if any, would be writing blocks. Thus, the reading or writing proficiency of a student might be estimated from 14, 28, or 42 minutes of assessment exercises, and the resultant differences in measurement precision generated two problems. First, many students, especially those who received only one block of reading items, answered all of the items correctly or below the chance level. Since a maximum likelihood computer program (LOGIST , Wingersky, 1983) was used at first, proficiency estimates for students with either perfect scores or scores below the chance level could not be estimated. Secondly, the attempts by NAEP to estimate proficiency distributions were affected by the imprecision of measurement. Furthermore, the estimation was complicated by the fact that measurement precision varied substantially, depending on the number of blocks from a single subject area that was assigned to a student. Standard statistical procedures that use individual scores to make population parameter estimates would not be adequate to achieve consistent estimates of NAEP's population proficiency distributions. Something different needed to be done.

Before proceeding, it is important to remember that NAEP does not need scores for individual students, since scores are reported neither for students nor their teachers or schools. Thus, computing individual scores was not only unnecessary but would, in fact, *not* lead to consistent estimates of population proficiency distributions (Mislevy, Beaton, Kaplan, and Sheehan, 1992). The population distributions could have been estimated directly, but this approach would not allow NAEP's complex sampling structure to be accounted for in error estimates. Also, estimating population characteristics directly would not provide secondary data analysts with data files that are compatible with SPSS and other statistical systems. As an alternative, the concept of plausible values was developed.

Plausible values were introduced in the 1983-84 NAEP in two ways. First, using item response theory (IRT), plausible values were developed for the NAEP reading scale by Mislevy and Sheehan (1987). Secondly, using linear models, plausible values were developed for the NAEP writing scale by Beaton and Johnson (1987).

The general theory of the NAEP plausible values is attributable to Mislevy (Mislevy and Sheehan, 1987, 1989) based on the work of Rubin (Rubin, 1987, and Rubin, and Schneker 1986) on multiple imputations. We will not present the detailed theory of plausible values here, nor how they are constructed, since this is carefully and rigorously explained in Mislevy, Johnson, and Muraki (1992) and in the NAEP Technical Reports (Beaton, 1987, 1988, and Johnson & Allen, 1990).

RATIONALE OF PLAUSIBLE VALUES

Plausible values should *not* be considered individual test scores; they are not. A plausible value is usually not the best available statistic for estimating an individual's proficiency. Further, plausible values explicitly include a random component so that they are entirely inappropriate for individual decision-making. NAEP does not estimate individual scores since it does not need to and, in fact, it is legally forbidden to report the performance of individual students.

NAEP is designed to produce population estimates; that is, to produce estimates of how various populations of students collectively perform on its proficiency scales and subscales in various academic subject areas. The plausible values may be thought of as intermediate computations to simplify the estimation of population proficiency distributions, their parameters, and estimates of their error variances. There are other ways of making population estimates, but NAEP chose the plausible value method in order to make its data available to secondary data analysts who use commonly available statistical systems such as SPSS. Using plausible values properly, however, does require some extra work and thought on the part of the user as compared to working with individual test scores.

In order to expand the coverage of the subject areas that are assessed, NAEP uses a method of assigning items to assessment booklets called BIB spiraling (See Chapter 2). Given that each assessment booklet contains only a sample of items from a subject area, an individual's proficiency on allthe items can never be known precisely . Through BIB spiraling, NAEP assigns different blocks of assessment items to different students. Some students are given a fairly large number of items from a particular subject area or sub-area and others are given a few. As a result, the proficiency of some students can be well estimated while the estimates for others are less accurate. Using plausible values can improve inferences about population distributions by acknowledging and accounting for the lack of precision in the estimation process.

In analyzing its assessment booklets, NAEP does not attempt to characterize an individual's proficiency by a single number. A person who responded in a particular way to a sample of items might have scored better or worse with a different sample of items, and so different points on the proficiency scale are possible representations of the true proficiency of a particular individual. Given the error of measurement, producing a single score for each individual to be used for parameter estimation will often produce biased estimates of population parameters and overestimate their precision.

Basically, plausible values contain both the available information about a student's proficiency as well as information about the uncertainty or measurement imprecision about the proficiency estimated. Under the assumptions of item response theory, NAEP produces a distribution that represents the likelihood that an individual is at various points on the proficiency scale. A likelihood distribution for each subscale is estimated for each individual student in an assessment. The NAEP plausible values for an individual are randomly selected from his or her own distribution.

NAEP randomly selects five plausible values for each individual on each sub-scale in any subject area they were tested. The differences among the plausible values for an individual are indicative of the measurement uncertainty on the sub-scale. Within a sub-scale, and across the sample, one set of plausible values is as good as another. Each of these sets of plausible values is equally well designed to estimate the population parameters, although the estimates will differ somewhat. The difference in the estimates is attributable to measurement error, that is, the uncertainty that is included in the plausible values. Rubin's (1987) theory of multiple imputation requires that more than one plausible value be generated for each individual; the more the better. However, empirical evidence suggests that five is a reasonably good number of plausible values.

The main property of plausible values is that they produce consistent estimators of population proficiency distributions and their parameters; that is, the parameter estimates approach their true values as the sample size grows indefinitely large. If NAEP had used a single 'optimum' value for an individual's proficiency--say, the most likely or the average of possible proficiency scores--then population estimates made from these optimum values would not in general approach the true values as the sample size grew larger. The 'optimum' values for estimating individual proficiency would produce biased population parameter estimates. An example will be shown below.

It is important to note that analyzing the average of the five plausible values for an individual is not appropriate and *should be avoided.* The average of an individual's five plausible values may be a better estimate of the individual's proficiency, but it will not in general produce consistent population estimates, or estimates of their error variance. Using the average of an individual's plausible values to obtain parameter estimates will generally underestimate the variance of the proficiency distribution, resulting in biased parameter estimates.

Let us now compute some simple descriptive statistics to show some of the properties of plausible values.

EXAMPLE 4 -1. DESCRIPTIVE STATISTICS: COMPUTING BASIC STATISTICS WITH PLAUSIBLE VALUES

In this example, basic statistics such as means, standard deviations, correlations, and percentiles of NAEP mathematics proficiency scales are computed. The program that generated these results is labeled EX41A.SPS on the Primer Disk and is included in the

EXAMPLES subdirectory. Sampling weights were not used for this example since at this point we are only interested in describing the sample, not estimating population parameters.

The means, standard deviations, and minimum and maximum values of the five plausible values generated for the eighth grade mathematics composite scale and its five subscales (Numbers and Operations; Measurement; Geometry; Data Analysis and Statistics; and Algebra and Functions) are shown in Figure 4-1. Note that in Figure 4-1 the means and standard deviations of the different plausible values within a subscale are quite similar; in fact, the means are not significantly different. There is, of course, more variability in the tails of the distribution as evidenced by their extreme values.

Figure 4-1 Descriptive Statistics for All Mathematics Proficiency Scales

Variable	Mean	Std Dev	Minimum	Maximum	N	Label
MRPCMP1	264.55	35.95	150.00	370.89	1000	PLAUS. VALUE #1 (COMPOSITE)
MRPCMP2	264.96	35.57	159.02	375.37	1000	PLAUS. VALUE #2 (COMPOSITE)
MRPCMP3	264.76	36.50	161.01	369.09	1000	PLAUS. VALUE #3 (COMPOSITE)
MRPCMP4	265.64	35.81	162.52	362.79	1000	PLAUS. VALUE #4 (COMPOSITE)
MRPCMP5	265.30	36.01	153.38	358.11	1000	PLAUS. VALUE #5 (COMPOSITE)
MRPSCA1	268.97	34.47	166.20	370.70	1000	PLAUS. VALUE #1 (NUM & OPER)
MRPSCA2	268.74	33.79	168.98	364.21	1000	PLAUS. VALUE #2 (NUM & OPER)
MRPSCA3	268.98	35.23	175.18	365.68	1000	PLAUS. VALUE #3 (NUM & OPER)
MRPSCA4	269.68	34.06	163.64	357.21	1000	PLAUS. VALUE #4 (NUM & OPER)
MRPSCA5	269.11	34.41	167.01	355.15	1000	PLAUS. VALUE #5 (NUM & OPER)
MRPSCB1	260.76	42.57	118.30	384.17	1000	PLAUS. VALUE #1 (MEASUREMENT)
MRPSCB2	261.60	42.40	118.56	387.03	1000	PLAUS. VALUE #2 (MEASUREMENT)
MRPSCB3	260.91	42.65	130.08	399.99	1000	PLAUS. VALUE #3 (MEASUREMENT)
MRPSCB4	262.03	42.40	137.78	384.30	1000	PLAUS. VALUE #4 (MEASUREMENT)
MRPSCB5	261.78	41.90	133.56	386.49	1000	PLAUS. VALUE #5 (MEASUREMENT)
MRPSCC1	260.99	34.71	146.14	365.52	1000	PLAUS. VALUE #1 (GEOMETRY)
MRPSCC2	262.31	34.65	144.38	382.53	1000	PLAUS. VALUE #2 (GEOMETRY)
MRPSCC3	261.46	34.85	142.90	369.33	1000	PLAUS. VALUE #3 (GEOMETRY)
MRPSCC4	262.48	35.10	170.84	372.33	1000	PLAUS. VALUE #4 (GEOMETRY)
MRPSCC5	262.31	34.64	143.10	356.13	1000	PLAUS. VALUE #5 (GEOMETRY)
MRPSCD1	265.25	40.94	137.79	404.39	1000	PLAUS. VALUE #1 (DATA ANAL&STAT)
MRPSCD2	265.89	40.41	136.70	384.15	1000	PLAUS. VALUE #2 (DATA ANAL&STAT)
MRPSCD3	265.10	41.21	146.65	375.43	1000	PLAUS. VALUE #3 (DATA ANAL&STAT)
MRPSCD4	266.03	40.62	138.45	365.90	1000	PLAUS. VALUE #4 (DATA ANAL&STAT)
MRPSCD5	266.72	40.77	138.91	372.48	1000	PLAUS. VALUE #5 (DATA ANAL&STAT)
MRPSCE1	263.80	35.96	154.39	360.99	1000	PLAUS. VALUE #1 (ALG & FUNCTNS)
MRPSCE2	263.76	35.75	154.98	369.64	1000	PLAUS. VALUE #2 (ALG & FUNCTNS)
MRPSCE3	264.33	37.10	150.89	359.28	1000	PLAUS. VALUE #3 (ALG & FUNCTNS)
MRPSCE4	265.18	35.88	158.32	365.06	1000	PLAUS. VALUE #4 (ALG & FUNCTNS)
MRPSCE5	264.15	36.79	154.25	364.83	1000	PLAUS. VALUE #5 (ALG & FUNCTNS)

The correlations among the plausible values are shown in Figure 4-2, and the program for producing them are in the file EX41B.SPS on the NAEP Primer disk. These correlations are indicators of the measurement uncertainty, or measurement error, in the plausible estimation of a student's proficiency. If these correlations were equal to one, then there would be no measurement error and plausible values would be unnecessary. The plausible values that NAEP produces are generally highly correlated, indicating fairly good measurement; however, these high correlations cannot be expected to hold for homogeneous sub-populations. The median correlation among the plausible values for the composite scale (.927) is the highest, and all correlations for the composite scales are above 0.92. The median correlations among the subscales are more variable: .915 for Numbers and

Operations (46 items); .866 for Measurement (20 items); .861 for Geometry (26 items); .911 for Data Analysis and Statistics (19 items); and .908 for Algebra and Functions (25 items).

Figure 4-2 Correlations Between Mathematics Proficiency Scales

Composite

Correlations:	MRPCMP1	MRPCMP2	MRPCMP3	MRPCMP4	MRPCMP5
MRPCMP1	1.0000	.9202**	.9279**	.9241**	.9256**
MRPCMP2	.9202**	1.0000	.9241**	.9269**	.9282**
MRPCMP3	.9279**	.9241**	1.0000	.9300**	.9276**
MRPCMP4	.9241**	.9269**	.9300**	1.0000	.9296**
MRPCMP5	.9256**	.9282**	.9276**	.9296**	1.0000

Numbers and Operations

Correlations:	MRPSCA1	MRPSCA2	MRPSCA3	MRPSCA4	MRPSCA5
MRPSCA1	1.0000	.9047**	.9161**	.9116**	.9152**
MRPSCA2	.9047**	1.0000	.9097**	.9146**	.9104**
MRPSCA3	.9161**	.9097**	1.0000	.9177**	.9181**
MRPSCA4	.9116**	.9146**	.9177**	1.0000	.9145**
MRPSCA5	.9152**	.9104**	.9181**	.9145**	1.0000

Measurement

Correlations:	MRPSCB1	MRPSCB2	MRPSCB3	MRPSCB4	MRPSCB5
MRPSCB1	1.0000	.8552**	.8735**	.8664**	.8614**
MRPSCB2	.8552**	1.0000	.8597**	.8658**	.8743**
MRPSCB3	.8735**	.8597**	1.0000	.8721**	.8652**
MRPSCB4	.8664**	.8658**	.8721**	1.0000	.8815**
MRPSCB5	.8614**	.8743**	.8652**	.8815**	1.0000

Geometry

Correlations:	MRPSCC1	MRPSCC2	MRPSCC3	MRPSCC4	MRPSCC5
MRPSCC1	1.0000	.8574**	.8664**	.8577**	.8461**
MRPSCC2	.8574**	1.0000	.8578**	.8520**	.8632**
MRPSCC3	.8664**	.8578**	1.0000	.8667**	.8647**
MRPSCC4	.8577**	.8520**	.8667**	1.0000	.8643**
MRPSCC5	.8461**	.8632**	.8647**	.8643**	1.0000

Data Analysis and Statistics

Correlations:	MRPSCD1	MRPSCD2	MRPSCD3	MRPSCD4	MRPSCD5
MRPSCD1	1.0000	.9096**	.9127**	.9045**	.9172**
MRPSCD2	.9096**	1.0000	.9109**	.9120**	.9120**
MRPSCD3	.9127**	.9109**	1.0000	.9099**	.9148**
MRPSCD4	.9045**	.9120**	.9099**	1.0000	.9067**
MRPSCD5	.9172**	.9120**	.9148**	.9067**	1.0000

Algebra and Functions

Correlations:	MRPSCE1	MRPSCE2	MRPSCE3	MRPSCE4	MRPSCE5
MRPSCE1	1.0000	.8973**	.9029**	.9076**	.9052**
MRPSCE2	.8973**	1.0000	.9049**	.9092**	.9091**
MRPSCE3	.9029**	.9049**	1.0000	.9129**	.9089**
MRPSCE4	.9076**	.9092**	.9129**	1.0000	.9136**
MRPSCE5	.9052**	.9091**	.9089**	.9136**	1.0000

N of cases: 1000 1-tailed Signif: * - .01 ** - .001

Each set of plausible values is entered into statistical systems as a separate variable. Although the plausible values for a student are interchangeable, it is prudent as well as convenient to use each set of plausible value variables as a unit. This is because the plausible values are built in sets for estimating the interrelationships among NAEP subscales. Since the 1990 NAEP Assessment (Mazzeo, 1992), each set of plausible values has been developed for the several sub-scales in a subject area simultaneously and thus randomly selected from a multivariate distribution. The plausible values are paired to produce consistent estimates of the correlations among the sub-scales. Thus, to estimate the correlations between the NAEP Number and Operations and the Measurement sub-scales, the first plausible value on one sub-scale should be paired with the first plausible value on the other, the second with the second, and so forth. To illustrate this feature, the correlations between the plausible values for the Numbers and Operations and Measurement subscales are shown in Figure 4-3. Note that the correlations in the diagonal of this matrix--that is, between paired plausible values---are all somewhat higher than the off-diagonal correlations.

Figure 4-3 **Correlations between the Numbers and Operations (MRPSCA) and Measurement (MRPSCB) Proficiency Scales**

Correlations:	MRPSCB1	MRPSCB2	MRPSCB3	MRPSCB4	MRPSCB5
MRPSCA1	.9232**	.8614**	.8648**	.8655**	.8667**
MRPSCA2	.8487**	.9244**	.8514**	.8640**	.8707**
MRPSCA3	.8721**	.8669**	.9265**	.8763**	.8734**
MRPSCA4	.8633**	.8664**	.8634**	.9285**	.8727**
MRPSCA5	.8640**	.8632**	.8576**	.8702**	.9296**

N of cases: 1000 1-tailed Signif: * - .01 ** - .001

It was emphasized earlier that the average of the plausible values should not be used in place of individual plausible values. Some descriptive statistics may suggest why. For example, consider the estimation of selected percentiles. Figure 4-4 shows the mean, standard deviation, and the 2nd, 5th, 10th, 25th, 50th, 75th, 90th, 95th, and 98th percentiles for each of the five Measurement plausible values and also for the average of those five values. The Measurement subscale plausible values (MRPSCB1 through MRPSCB5) were used for demonstration because the intercorrelations among the plausible values are among the lowest of any of the subscales and so the statistical differences among them should be more obvious. The estimates of the second percentile that are computed from different sets of plausible values range from 170.52 to 174.18; the average of these five percentile estimates is 172.05. The estimate of the second percentile using the average of the five plausible values is 177.45, an estimate that is higher than any of the estimates from individual plausible values and is also closer to the estimated population mean on this scale. We also note that the 98th percentile computed from the average of the plausible values (338.28) is also closer to the population mean than the average of the estimates from individual plausible values (341.76). Usually, the estimated percentiles based on the average plausible value will be closer to the population mean than those estimates based on the individual plausible values. The standard deviation of the average plausible values is also lower than the standard deviation estimated from most of the individual sets of plausible values, except for MRPSCB5. *The average of the estimates made from different plausible values is the recommended estimate of the population parameter.* Indeed, in this table only the mean (261.42) is exactly the same whether estimated from the average of the plausible values or from the average of estimates made from the different plausible values.

Each set of plausible values is designed to give consistent estimates of population distributions and so any of them may be used for data analyses. However, using a single plausible value will generally give a consistent estimate of population parameters, but standard procedures of statistical inference will understate the estimate of the uncertainty associated with a parameter estimate. The uncertainty due to sampling individuals from a population of individuals will be estimated but the full uncertainty arising from sampling items from a population of items will not be completely included. To account for both uncertainties, analyses should be run at least twice, but preferably as many times as there are plausible values (usually five), so that error estimates that fully include both types of uncertainty are produced.

Figure 4-4 Descriptive Statistics for the Measurement Proficiency Scales (MRPSCB)

	MRPSCB1	*MRPSCB2*	*MRPSCB3*	*MRPSCB4*	*MRPSCB5*	*AVERAGE*	*MEANSCB*
MEAN	260.76	261.60	260.91	262.03	261.78	261.42	261.42
S.D.	42.57	42.40	42.65	42.40	41.90	42.38	40.08
Percentile							
2	171.22	171.04	170.52	173.28	174.18	172.05	177.45
5	190.86	188.22	189.83	188.36	192.05	189.86	195.19
10	205.10	205.87	206.23	204.66	206.29	205.63	208.67
25	231.00	233.94	231.20	231.48	232.56	232.04	233.03
50	262.12	261.65	263.14	262.87	264.01	262.76	262.08
75	290.87	293.08	291.71	293.84	291.52	292.20	291.06
90	317.79	314.79	314.56	315.14	313.45	315.15	312.15
95	330.70	327.18	329.50	329.53	328.70	329.12	324.62
98	340.64	339.76	345.89	342.91	339.61	341.76	338.28

THE GENERAL METHOD FOR USING PLAUSIBLE VALUES

Using plausible values to estimate population parameters and their error variances requires computations over and above what would be necessary if we could assume that the students' proficiency were measured without error. Analyses have to be run repeatedly, once for each plausible value, and the results of the several analyses synthesized into a single parameter estimate and its error variance. In this section, we will introduce a general method for using plausible values and a computer program to simplify the necessary calculations. In the following section, we will introduce two simple short-cut methods that are useful for some--but not all--common statistical analyses.

The details of the algorithms are fully described by Mislevy, Johnson, and Muraki (1992) and will not be detailed here. We will follow their notation here in order to aid readers who wish to investigate more fully the theory and usage of plausible values.

The general procedure for using plausible values is as follows:

1. Estimate a parameter (or parameters) repeatedly, each time using a different set of the M plausible values. The parameter(s) can be anything estimable from the data, such as a mean, the difference between means, or percentiles: In this section we will assume that the parameter(s) can be estimated using a standard statistical package. The estimation is done using each set of plausible values as if it were a vector of the students' true proficiencies, θ. If all of the (M=5) plausible values in the NAEP database are used, the parameter will be estimated five times, once using each set of plausible values. We will call the parameter(s) T, and its estimates t_m (m=1,2,...,M), where T and t_m may be vectors of length k.

2. Estimate the error variance for the parameter, each time using a different set of plausible values. Most statistical systems automatically produce an estimate of the

sampling variance (or standard error) for many parameter estimates under the assumption of simple random sampling. Since NAEP does not select a simple random sample of students, the use of the jackknife method is preferred for estimating error variances (see Chapter 6). Using the mini-sample in this chapter, we will use a weight of .8 for each observation in order to compensate for the fact that students were not selected by simple random sampling. Each error variance must be estimated five times, once using each plausible value. We will call these error variances $U_m (m=1,2,...,M)$, where U_m may be an error covariance matrix of order k by k.

In many statistical systems, steps (1) and (2) are run together, requiring just one pass over the data.

After each plausible value is analyzed, the parameter estimates and their error variances from the several analyses should be combined into a single parameter estimate and its error variance. COMBPV, a computer program for combining results is available on the NAEP Primer disk. The algorithm for combining the results of individual analyses into overall estimates is detailed in Mislevy, Johnson, and Muraki (1992).

COMBPV is a Microsoft QBASIC 4.5 program written for IBM-compatible personal computers. The disk contains two copies, COMBPV.BAS which is in ASCII format and ready to compile and COMBPV.EXE which is compiled and ready to execute on most DOS computers. The program is described in Appendix C. The use and output of COMBPV will be shown in the following example.

EXAMPLE 4-2. GENERAL METHOD: ESTIMATING GENDER DIFFERENCES IN MEASUREMENT PROFICIENCY

Let us demonstrate the recommended use of plausible values using SPSS and the program COMBPV. Let us say that we want to estimate the difference between the means of eighth grade male and female students on the NAEP Measurement proficiency scales in 1990, to estimate the standard error of this difference, and to test the hypothesis that the mean proficiency on the measurement scale is equal for males and females in this population. To do this, we will make the same statistical assumptions that would be used if the students' measurement proficiency were known without error. To estimate the gender difference, we will use the NAEP Policy mini-sample, which has the necessary data. We will use a weight of .8 for each observation. For simplicity, we will use only two plausible values, not the five that are available. Two is the minimum number for estimating the components of the error variance.

The first computational phase involves estimating the gender differences using each plausible value separately. For generality, we will use the SPSS REGRESSION command since the procedures used in this example will generalize to more complicated problems.

A copy of the SPSS commands for this example is shown in Figure 4-5 and is available in file EX42.SPS of the Primer disk.

Figure 4-5 **SPSS Code for estimating Gender differences on the Measurement Proficiency Scale (MRPSCB) using the General Method**

```
GET FILE = 'C:\PRIMER\M08PS1.SYS'
  / KEEP = WEIGHT DSEX MRPSCB1 MRPSCB2.
WEIGHT BY WEIGHT.
COMPUTE NEWSEX = DSEX.
RECODE NEWSEX (2=0) (1=1) .
VALUE LABELS NEWSEX 0 'FEMALE' 1 'MALE'.
FREQUENCY VARIABLES = DSEX NEWSEX.
REGRESS VARIABLES = MRPSCB1 MRPSCB2 NEWSEX
 / STATISTIC = DEFAULT
 / DEPENDENT = MRPSCB1 MRPSCB2
 / METHOD    = ENTER NEWSEX.
```

Only four variables from the Policy Mini-File are needed for this analysis:

WEIGHT: the sampling weight, which is the constant .8 for all students

DSEX: the students gender, with Males=1 and Females=2

MRPSCP1: the first measurement plausible value

MRPSCP2: the second measurement plausible value

The program has the following features:

- It applies the weight .8 to all observations in order to compensate for the fact that the students were not selected by simple random sampling, as explained in Chapter 5.
- It creates a new variable NEWSEX by recoding DSEX. NEWSEX is more convenient for use as a zero-one dummy variable with Females=0 and Males=1. As a result of this recoding, when NEWSEX is used as the independent variable in a regression analysis, the slope in the regression equation will be the difference between the male mean and the female mean on the dependent variable, in this case a plausible value. The regression slope, therefore, will be an estimate of the parameter of interest.
- Incidentally, the constant in the regression equations will be the mean of the females on the dependent variable. The SPSS program prints frequency distributions to check the recoding.
- It regresses each of the two measurement plausible values on NEWSEX. The program uses the SPSS default options.

The resulting regression analyses are shown in Figure 4-6 and the parameter estimates and their standard errors are shown in Table 4-1.

Table 4-1: **Parameter estimates, standard error and error variance from Figure 4-6.**

	Average	Std.err.	Error Variance
PV1	14.4970	2.9697	8.8191
PV2	11.0809	2.9753	8.8524

Given the coding of the dummy variable NEWSEX, the regression coefficients show that the male mean exceeded the female mean by 14.4970 scale points when using the first plausible value and by 11.0809 points when using the second. Both differences are large, compared to their standard errors. SPSS prints out Student t statistics and their associated probabilities but these are not appropriate for plausible values since they do not contain all of the error components.

Figure 4-6 **SPSS Output for estimating Gender differences on the Measurement Proficiency Scale (MRPSCB) using the General Method**

```
        * * * *   M U L T I P L E   R E G R E S S I O N   * * * *

Equation Number 1    Dependent Variable..   MRPSCB1   PLAUSIBLE VALUE #1 (MEASU

Block Number  1.  Method:  Enter      NEWSEX

Variable(s) Entered on Step Number
   1..    NEWSEX

Multiple R           .17028
R Square             .02900
Adjusted R Square    .02778
Standard Error     41.98392

Analysis of Variance
                    DF       Sum of Squares       Mean Square
Regression           1          42004.15745       42004.15745
Residual           798        1406594.15471        1762.64932

F =      23.83013       Signif F =  .0000

------------------ Variables in the Equation ------------------

Variable              B         SE B       Beta          T  Sig T

NEWSEX        14.496995     2.969715    .170283      4.882  .0000
(Constant)   253.699864     2.072427               122.417  .0000

End Block Number   1   All requested variables entered.

        * * * *   M U L T I P L E   R E G R E S S I O N   * * * *

Equation Number 2    Dependent Variable..   MRPSCB2   PLAUSIBLE VALUE #2 (MEASU

Block Number  1.  Method:  Enter      NEWSEX

Variable(s) Entered on Step Number
   1..    NEWSEX

Multiple R           .13071
R Square             .01708
Adjusted R Square    .01585
Standard Error     42.06313

Analysis of Variance
                    DF       Sum of Squares       Mean Square
Regression           1          24540.54985       24540.54985
Residual           798        1411907.17663        1769.30724

F =      13.87015       Signif F =  .0002

------------------ Variables in the Equation ------------------

Variable              B         SE B       Beta          T  Sig T

NEWSEX        11.080873     2.975319    .130707      3.724  .0002
(Constant)   256.203665     2.076338               123.392  .0000

End Block Number   1   All requested variables entered.
```

The recommended estimate of the difference between the mean of eighth grade males and females in measurement is simply the average of the two regression coefficients, $(14.4970 + 11.0809)/2 = 12.7889$.

33

The program COMBPV will do this computation as well as compute the error variance of the parameter estimate. The input and output to COMBPV for this example is shown in Figure 4-7. The input is in file EX42B.PAR are included in the accompanying diskette.

Figure 4-7 COMBPV Input and Output code for estimating Gender differences on the Measurement Proficiency Scale (MRPSCB) using the General Method

Input

```
EXAMPLE FOR FIGURE 4-2c
K = 1
M = 2
N = 800
PARAMETERS
NEWSEX , 0.0
PV1
14.496995
8.8191
PV2
11.080873
8.8524
```

Output

```
EXAMPLE FOR FIGURE 4-2c     08-06-1993    13:50:58

Number of Plausible Values      (M):  2
Number of Parameters            (K):  1
Number of Subjects              (N):  800

Parameter                     Hypothesized value
NEWSEX                            0.0000

PARAMETER ESTIMATES AND ERROR COVARIANCE MATRIX - PLAUSIBLE VALUE  1

Parameter    Estimate  | Error covariance matrix
NEWSEX        14.49699 |      8.8191

PARAMETER ESTIMATES AND ERROR COVARIANCE MATRIX - PLAUSIBLE VALUE  2

Parameter    Estimate  | Error covariance matrix
NEWSEX        11.08087 |      8.8524

AVERAGE SAMPLING ERROR  (U*)

NEWSEX
     8.83575

ERROR DUE TO IMPUTATION (BM)

NEWSEX
     5.83495

SUMMARY SECTION
---------------

AVERAGE PARAMETER ESTIMATES (T*) AND TOTAL ERROR COVARIANCE MATRIX (V)

Parameter   Estimate   | Total error covariance matrix
NEWSEX         12.7889 |     17.5882

SIGNIFICANCE TEST RESULTS
-------------------------

     T          DEGREES OF FREEDOM              P
   3.049          ( 1 ,   4.03)              0.0376
```

The input was copied from the SPSS output, which was shown in Figure 4-6, using a word processor that produced an ASCII file. The output includes the input for documentation, contains intermediate calculations (U^* and B_M, which are described in Appendix C), and,

BEST COPY AVAILABLE

finally, the parameter estimate (12.7889) and its error variance (17.5882). Since this is a one parameter test, a Student t statistic (3.049), its degrees of freedom (4.03), and associated probability (.0376) are presented (see last line of Figure 4–7).

Therefore, the hypotheses that the difference between the average proficiency of males and females is the result of sampling and measurement error can be rejected at the .05 level, but not at the .01 level. Note that if we had run this analysis using either plausible value alone as if it were an accurate measure of proficiency, we would have rejected the hypothesis of gender differences with virtual certainty, that is, with estimated probabilities of .0002 or less.

One Parameter Short-Cut Method for Using Plausible Values

The general method in the previous section requires copying computer output from the results of one set of analyses into a file for use in COMBPV, another computer program. In this section, we will introduce a short-cut method and demonstrate its use on the same problem that was shown in Example 4-2.

Before proceeding, it is important to note that the short-cut procedure used in this section is *not* a general one. First, it is appropriate only when one parameter is being estimated, such as a population mean or the difference between two population means, as in Example 4-2. Secondly, this short-cut method is appropriate only for linear statistics, such as proportions, means, and regression coefficients but is *not* appropriate for non-linear statistics, such as standard deviations, percentiles or correlation coefficients. The short-cut method shown here will, therefore, be appropriate for many commonly-used statistical analyses, but not all. The more complicated, general method presented above will be necessary for non-linear applications.

Let us say that we wish to estimate a population parameter T by regressing the student scores θ on an independent variable, x. Let us say further that the available data consist of a random sample of observations with measurements on two plausible values, y_1 and y_2, instead of θ and x. Using the general method, we would regress both y_1 and y_2, resulting in a parameter estimates for each plausible value, t_1 and t_2, and their error variances, U_1 and U_2. Using the COMBPV program, we would compute the recommended estimates of T and its standard error SQR(V) with Equation 4-1,

Equation 4-1

$$t = \sum t_m / 2$$

and its error variance V.

The short cut method uses a transformation of the data so that the estimated parameter can be read directly from the regression output without post-processing, and the error variance can be computed by simple addition. In its simplest form, this short-cut method uses only two plausible values, even though five are available in NAEP. The key idea is to transform these two plausible values into two new and different variables as follows

$$y' = (y_1 + y_2)/2$$

$$d' = (y_1 - y_2)/2$$

where y' is the average of the two plausible values and d' is one half of the difference between the two plausible values. These transformed variables are used instead of the original plausible values in estimating the parameter and its error variances.

To analyze the data, each of the new variables--instead of the original plausible values-- is regressed on the independent variable x. The result is two regression coefficients, $t_{y'}$ and $t_{d'}$, and their respective standard errors, $s_{y'}$ and $s_{d'}$.

The appropriate estimate of the population parameter is $t_{y'}$, since this is algebraically the average of the two regression coefficients produced by regressing each plausible value, y_1 and y_2, on x. The estimate of the parameter, therefore, can be read directly from the regression program output and does not require any additional computation.

The regression coefficient $t_{d'}$ is algebraically half of the difference between the two estimates made using the two plausible values individually, as in the last example.

The error variance of the population parameter estimate $t_{y'}$ can be shown to be the sum of the error variances of the two transformed parameters plus three times the square of $t_{d'}$. In this case, the best estimate of the error variance is computed by

Equation 4-2 $$V = S_{y'}^2 + S_{d'}^2 + (3 * t_{d'}^2)$$

and its standard error (s) by

Equation 4-3 $$s = \sqrt{V}$$

which is easily computed from computer output using a hand calculator.

The corresponding number of degrees of freedom can be computed using the formulae provided by Rust and Johnson (1992),

Equation 4-4 $$f_M = 3 * S_{d'}^2 / V$$

where $3 = 2 * (1 + M^{-1})$ when M=2, and

Equation 4-5 $$\upsilon = \frac{1}{f_M^2 + \frac{(1 - f_M)^2}{d}}$$

where *d* is the number of degrees of freedom under the usual statistical assumptions. Mislevy, Johnson, and Muraki (1992) suggest that if $(s_{d'})^2$ is large compared to V, the approximation for the degrees of freedom will have little effect.

The results using this short-cut method are identical to those using the general method, under the usual regression assumptions.

It is worth noting that, in many cases, the post-processing of regression results will not be necessary. If the regression output shows that the coefficient $t_{y'}$ is insignificant, that is, $t_{y'}/s_{y'}$ does not exceed the critical value in a Student t-table, then the additional calculations for estimating the standard error are not necessary. These extra calculations do not change the estimate of the parameter but can only enlarge its standard error and reduce the number of degrees of freedom, consequently reducing the value of the significance statistic and thereby reducing the estimated probability of its value occurring by chance.

EXAMPLE 4 -3. ONE PARAMETER SHORT-CUT METHOD: ESTIMATING GENDER DIFFERENCES IN MEASUREMENT PROFICIENCY

The SPSS program for Example 4-3 is shown in Figure 4-8 and its output in Figure 4-9. Note that this program differs from the program for Example 4-2 only in that the transformation of the plausible values is inserted.

Figure 4-8 SPSS Code for estimating Gender differences on the Measurement Proficiency Scale (MRPSCB) using the Short- Cut Method

```
GET FILE = 'C:\PRIMER\M08PS1.SYS'
  / KEEP = WEIGHT DSEX MRPSCB1 MRPSCB2.

WEIGHT BY WEIGHT.
COMPUTE NEWSEX = DSEX.
RECODE NEWSEX (2=0) (1=1) .
VALUE LABELS NEWSEX 0 'FEMALE' 1 'MALE'.
FREQUENCY VARIABLES = DSEX NEWSEX.

COMPUTE AVE1_2 = (MRPSCB1 + MRPSCB2) / 2.
COMPUTE DIF1_2 = (MRPSCB1 - MRPSCB2) / 2.

REGRESS VARIABLES = AVE1_2 DIF1_2 NEWSEX
 / STATISTIC = DEFAULT
 / DEPENDENT = AVE1_2 DIF1_2
 / METHOD    = ENTER NEWSEX.
```

The parameter estimates and their standard errors are shown in Table 4-2 below.

Table 4-2: Parameter estimates, standard error and error variance from Figure 4-9.

	Estimate	*Std.err.*	*Error Variance*
$b_{y'}$	12.7889	2.8609	8.1847
$b_{d'}$	1.7081	0.8070	0.6512

The regression coefficient $b_{y'}$=12.7889 is the estimated difference between the means of males and females, as in Example 4-2. The total error variance can be computed as

$$V = 2.8609^2 + 0.8070^2 + 3 * 1.7081^2 = 17.5888$$

and the standard error as

$$S_{y'} = \sqrt{17.5888} = 4.1939$$

Figure 4-9 SPSS Output for estimating Gender differences on the Measurement Proficiency Scale (MRPSCB) using the Short-Cut Method

```
        * * * *   M U L T I P L E   R E G R E S S I O N   * * * *

Equation Number 1    Dependent Variable..   AVE1_2

Block Number  1.  Method:  Enter     NEWSEX

Variable(s) Entered on Step Number
   1..    NEWSEX

Multiple R            .15630
R Square              .02443
Adjusted R Square     .02321
Standard Error      40.44535

Analysis of Variance
                    DF      Sum of Squares      Mean Square
Regression           1         32689.25379      32689.25379
Residual           798       1305389.51191       1635.82646

F =      19.98333       Signif F =  .0000

------------------ Variables in the Equation ------------------
Variable              B          SE B       Beta          T  Sig T

NEWSEX        12.788934    2.860885    .156301      4.470  .0000
(Constant)   254.951764    1.996480               127.701  .0000

End Block Number   1   All requested variables entered.

        * * * *   M U L T I P L E   R E G R E S S I O N   * * * *

Equation Number 2    Dependent Variable..   DIF1_2

Block Number  1.  Method:  Enter     NEWSEX

Variable(s) Entered on Step Number
   1..    NEWSEX

Multiple R            .07472
R Square              .00558
Adjusted R Square     .00434
Standard Error      11.40841

Analysis of Variance
                    DF      Sum of Squares      Mean Square
Regression           1           583.09986        583.09986
Residual           798        103861.15375        130.15182

F =       4.48015       Signif F =  .0346

------------------ Variables in the Equation ------------------
Variable              B          SE B       Beta          T  Sig T

NEWSEX         1.708061     .806969    .074719      2.117  .0346
(Constant)    -1.251901     .563147                -2.223  .0265

End Block Number   1   All requested variables entered.
```

To approximate the number of degrees of freedom

$$f_M = \frac{3 * 1.7081^2}{17.5888} = .4976$$

and

$$\upsilon = \frac{1}{.4976^2 + \frac{(1-.4976)^2}{798}} = 4.03$$

The Student t-statistic is

$$t_{df=4} = \frac{12.7889 - 0}{4.1939} = 3.049$$

which is statistically significant at the .05 level. These results are, of course, identical to those of the preceding example, except for rounding error.

The short-cut method displays the results in a way that is easier for interpretation; this advantage will become more evident in the more complicated examples. First, the recommended estimate of the population parameter is shown directly, although the associated standard error is understated. The difference between the parameter estimates using different plausible values is also shown directly; if this is large, then caution is necessary in interpreting the results. Finally, the components of the error variance are viewed separately – citing the numbers for the above parameters for this example would help the reader to follow it.

The reader may notice that the plausible values were averaged, despite the earlier warning not to do so. With linear statistics, an estimate based on the average of plausible values will be identical to the average of the five estimates, and so the averaging is possible with linear statistics, but not otherwise. But the standard error associated with the population estimate from the average of the plausible values is not optimum and requires additional components that are estimated from the average difference between the plausible values.

We note in passing that the significance test could be computed directly from the Analysis of Variance table that most regression programs routinely print out. In this case,

- The between mean square for y': 32689.2538
- The within mean square for y': 1635.8265
- The between mean square for d': 583.0999
- The within mean square for d': 130.1518

with an F statistic computed as

$$F = \frac{32689.2538}{1635.8266 + 130.1518 + 3 * 583.0999} = 9.2992$$

and, since there is but one degree of freedom for the numerator, $t = \sqrt{F}$ and in this case, $t = \sqrt{9.2992} = 3.049$, as before. The identity of the F and t statistics is exact when just one parameter is estimated. We will explore this method further in the next section when more than one parameter is estimated.

MULTI-PARAMETER SHORT-CUT APPROXIMATION FOR USING PLAUSIBLE VALUES

The general method requires more calculations when several parameters are estimated jointly along with their error covariance matrices. For example, a set of parameters may be estimated in a regression analysis, and it may be of interest whether some subset of the regression coefficients are significantly different from zero. The general method requires estimating the regression coefficients repeatedly, once for each plausible value, and computing a separate error covariance matrix for each set of regression coefficients. The several vectors of regression coefficients and error matrices must be entered into the COMBPV program. This short-cut method proposed here is simpler computationally, and, like the one parameter short-cut is appropriate only for linear statistics. However, this short-cut procedure is different in that it does *not* in general produce exactly the same significance test as the general method. We believe that the results will be close enough for most practical purposes.

This short-cut approximation uses the same device as the one-parameter method, that is, transforms two plausible values into new variables, the average plausible value y' and half of their difference d', that is

$$y' = (y_1 + y_2)/2$$

$$d' = (y_1 - y_2)/2 .$$

Analyses are run using these variables instead of using the two plausible values separately. The analysis program will typically produce parameter estimates, and an analysis of variance table.

The recommended parameter estimate will be the estimate from the analysis of y' since this estimate will be the average of the two separate estimates. This estimated parameter from this short-cut method will be exactly the same as that using the general method.

The approximate significance test can be computed from the ANOVA table. There will be four mean squares of interest:

- The between mean square for y': MSB(y')
- The within mean square for y': MSW(y')
- The between mean square for d': MSB(d')
- The within mean square for d': MSW(d')

The approximate F statistic can be computed as follows:

Equation 4-6

$$F = \frac{MSB(y')}{MSW(y') + MSW(d') + 3 * MSB(d')}$$

A simple example may help to illuminate the multi-parameter short-cut procedure. First, in Example 4-4 we will show a three variable regression problem and how to use the general method when more than one parameter is estimated. Second, we will do the same example using the multi-parameter short-cut method and show the differences.

EXAMPLE 4-4. MULTI- PARAMETER GENERAL METHOD: ESTIMATING REGRESSION COEFFICIENTS

Let us say that we wish to estimate the regression of the NAEP measurement subscale on three student questionnaire items. The variables are:

MRPSCB1: The first measurement plausible value.

MRPSCB1: The second measurement plausible value.

M810702B: Background question "Do you agree: all people use math in their jobs."

M810703B: Background question "Do you agree: I am good in math."

M810705B: Background question "Do you agree: Math is useful in solving everyday problems."

We note that the questionnaire variables M810702B to M810705B are Likert type items coded from STRONGLY AGREE=1 to STRONGLY DISAGREE=5. There are also some missing data, which are coded as an 8 or 9.

An SPSS program for running this analysis using the general method is shown in Figure 4-10 and is in file EX44.SPS on the Primer disk. The data are weighted by the constant .8 to adjust for non-random sampling. The missing code "8" is recoded to "9", and then "9" is declared a missing value. The program produces frequency distributions (not shown) for these variables in order to check for irregularities in the data. The program regresses each of two plausible values on the three student questionnaire items. Note that the statistics option of the REGRESSION procedure is used so that the covariances of the parameter estimates as well as the default statistics will be printed.

Figure 4-10 SPSS Code for Estimating Regression Coefficients on Multiple Parameters using the General Method

```
TITLE "EXAMPLE 4-4".
GET FILE = 'C:\PRIMER\M08PS1.SYS'.
WEIGHT BY WEIGHT.
RECODE M810702B M810703B M810705B (8=9).
MISSING VALUES M810702B M810703B M810705B (9).
FREQUENCY VARIABLES = M810702B M810703B M810705B.
REGRESS
   VARIABLES  = MRPSCB1 MRPSCB2
   M810702B M810703B M810705B
 / STATISTICS = DEFAULT BCOV
 / DEPENDENT  = MRPSCB1 TO MRPSCB2
 / METHOD     = ENTER M810702B TO M810705B.
```

The SPSS results are shown in Figure 4-11. Note that SPSS has produced the coefficients in reverse numerical order from how they were specified in the METHOD command. The covariance matrix of the regression coefficients is printed; however, it should be noted that SPSS prints the covariance on and below the diagonal with the correlations of the regression coefficients above the diagonal. We note that each plausible value produces a statistically significant coefficient for M810703B, but not for M810702B and M810705B, and an overall significant F statistic.

Figure 4-11 **SPSS Output for Estimating Regression Coefficients on Multiple Parameters using the General Method**

```
        * * * *   M U L T I P L E   R E G R E S S I O N   * * * *

Equation Number 1    Dependent Variable..   MRPSCB1   PLAUSIBLE VALUE #1 (MEASU

Variable(s) Entered on Step Number
   1..     M810705B  DO YOU AGREE: MATH USEFUL/SOLVING EVERYDAY PROBLEMS
   2..     M810703B  DO YOU AGREE: I AM GOOD IN MATH
   3..     M810702B  DO YOU AGREE: ALL PEOPLE USE MATH IN THEIR JOBS

Multiple R           .27541
R Square             .07585
Adjusted R Square    .07222
Standard Error     40.87238

Analysis of Variance
                    DF      Sum of Squares      Mean Square
Regression           3        104642.55007      34880.85002
Residual           763       1274964.57386       1670.55107

F =      20.87985       Signif F =  .0000

Var-Covar Matrix of Regression Coefficients (B)
Below Diagonal:  Covariance    Above:  Correlation

             M810705B   M810703B   M810702B

M810705B     2.65086    -.12309    -.35471
M810703B     -.29791    2.20969    -.19178
M810702B    -1.08580    -.53598    3.53477

------------------ Variables in the Equation ------------------

Variable              B         SE B        Beta          T  Sig T

M810705B      -2.005494    1.628146    -.046872     -1.232  .2184
M810703B     -10.960869    1.486502    -.267300     -7.374  .0000
M810702B        .967697    1.880098     .019804       .515  .6069
(Constant)   290.215246    4.736888                 61.267  .0000

End Block Number   1   All requested variables entered.

* * * *   M U L T I P L E   R E G R E S S I O N   * * * *

Equation Number 2    Dependent Variable..   MRPSCB2   PLAUSIBLE VALUE #2 (MEASU

Variable(s) Entered on Step Number
   1..     M810705B  DO YOU AGREE: MATH USEFUL/SOLVING EVERYDAY PROBLEMS
   2..     M810703B  DO YOU AGREE: I AM GOOD IN MATH
   3..     M810702B  DO YOU AGREE: ALL PEOPLE USE MATH IN THE

Multiple R           .26442
R Square             .06992
Adjusted R Square    .06626
Standard Error     40.73362

                                                              (continues...)
```

42

BEST COPY AVAILABLE

Figure 4-11 SPSS Output for Estimating Regression Coefficients on Multiple Parameters using the General Method (continued)

```
Analysis of Variance
                    DF      Sum of Squares      Mean Square
Regression           3         95197.64323      31732.54774
Residual           763       1266322.60002       1659.22773

F =      19.12489       Signif F =  .0000

Var-Covar Matrix of Regression Coefficients (B)
Below Diagonal:  Covariance    Above:  Correlation

            M810705B   M810703B   M810702B
M810705B     2.63289    -.12309    -.35471
M810703B     -.29589    2.19471    -.19178
M810702B    -1.07844    -.53235    3.51081

------------------ Variables in the Equation ------------------

Variable              B         SE B       Beta          T  Sig T
M810705B      -2.263857     1.622618   -.053260     -1.395  .1634
M810703B     -10.463685     1.481456   -.256864     -7.063  .0000
M810702B       1.609393     1.873715    .033155       .859  .3906
(Constant)   289.229429     4.720807                61.267  .0000

End Block Number   1   All requested variables entered.
```

The estimates of the regression coefficients and their covariance are combined to make final parameter estimates using COMBPV. The COMBPV input was copied from the SPSS output. The input to, and output from COMBPV, are shown in Figure 4-12 and 4-13. The average parameter estimates and the total error covariance matrix are shown near the bottom of the output (Figure 4-13).

Figure 4-12 COMBPV Input Code for Estimating Regression Coefficients on Multiple Parameters using the General Method

```
EXAMPLE 4-4 - THREE PARAMETERS
K = 3
M = 2
N = 800
PARAMETER ESTIMATES
M810705B , 0.0
M810703B , 0.0
M810702B , 0.0
PV1
-2.005494 , -10.960869 ,  .967697
 2.65086
-0.29791  ,   2.20969
-1.08580  ,  -0.53598   , 3.53477
PV2
-2.263857 , -10.463685 , 1.609393
 2.63289
-0.29589  ,   2.19471
-1.07844  ,  -0.53235   , 3.51081
```

Figure 4-13 **COMBPV Output Code for Estimating Regression Coefficients on Multiple Parameters using the General Method**

```
EXAMPLE M8107.PAR - THREE PARAMETERS -        08-12-1995    13:27:16

Number of Plausible Values        (M):  2
Number of Parameters              (K):  3
Number of Subjects                (N):  767

Parameter                      Hypothesized value
M810705B                           0.0000
M810703B                           0.0000
M810702B                           0.0000

PARAMETER ESTIMATES AND ERROR COVARIANCE MATRIX - PLAUSIBLE VALUE  1

Parameter     Estimate   | Error covariance matrix
M810705B      -2.00549   |     2.6509      -0.2979       -1.0858
M810703B     -10.96087   |    -0.2979       2.2097       -0.5360
M810702B       0.96770   |    -1.0858      -0.5360        3.5348

PARAMETER ESTIMATES AND ERROR COVARIANCE MATRIX - PLAUSIBLE VALUE  2

Parameter     Estimate   | Error covariance matrix
M810705B      -2.26386   |     2.6329      -0.2959       -1.0784
M810703B     -10.46369   |    -0.2959       2.1947       -0.5324
M810702B       1.60939   |    -1.0784      -0.5324        3.5108

AVERAGE SAMPLING ERROR  (U*)

M810705B        M810703B        M810702B
     2.64188        -0.29690        -1.08212
    -0.29690         2.20220        -0.53417
    -1.08212        -0.53417         3.52279


ERROR DUE TO IMPUTATION (BM)

M810705B        M810703B        M810702B
     0.03338        -0.06424        -0.08290
    -0.06423         0.12360         0.15952
    -0.08290         0.15951         0.20589

SUMMARY SECTION
---------------

AVERAGE PARAMETER ESTIMATES (T*) AND TOTAL ERROR COVARIANCE MATRIX (V)

Parameter    Estimate    | Total error covariance matrix
M810705B       -2.1347   |    2.6919    -0.3933    -1.2065
M810703B      -10.7123   |   -0.3932     2.3876    -0.2949
M810702B        1.2885   |   -1.2065    -0.2949     3.8316

SIGNIFICANCE TEST RESULTS
-------------------------

     F            DEGREES OF FREEDOM                P
   18.325           (  3 , 216.59)               0.0000
```

EXAMPLE 4-5. MULTI- PARAMETER SHORT-CUT METHOD: ESTIMATING REGRESSION COEFFICIENTS

The SPSS program for estimating the same regression equation as in the previous example is shown in Figure 4-14 and is in the file EX45.SPS on the Primer disk. The major difference from the Example 4-4 is that the two plausible values are transformed into the average plausible value and half of their difference.

Figure 4-14 SPSS Code for Estimating Regression Coefficients on Multiple Parameters using the Short- Cut Method

```
TITLE "EXAMPLE 4-5".
GET FILE = 'C:\PRIMER\M08PS1.SYS'.

WEIGHT BY WEIGHT.
RECODE M810702B M810703B M810705B (8=9).
MISSING VALUES M810702B M810703B M810705B (9).
FREQUENCY VARIABLES = M810702B M810703B M810705B.

COMPUTE AVE1_2 = (MRPSCB1 + MRPSCB2) / 2.
COMPUTE DIF1_2 = (MRPSCB1 - MRPSCB2) / 2.

REGRESS
   VARIABLES  =  AVE1_2 DIF1_2
   M810702B M810703B M810705B
 / STATISTICS = DEFAULT
 / DEPENDENT  = AVE1_2 DIF1_2
 / METHOD     = ENTER M810702B TO M810705B.
```

The SPSS output is shown in Figure 4-15. The top panel in this output has the results from regressing the average plausible value on the three questionnaire items. The estimated regression equation is

$$MP = 289.722 + (1.289 * M810702B) - (10.712 * M810703B) - (2.135 * M810705B)$$

where MP is Measurement Proficiency. The set of regression coefficients at the bottom of the page are half the difference between the two sets of parameter estimates. All of these coefficients are small compared to the sampling error.

SPSS output suggests that only the regression coefficient associated with M810703B is statistically significant from zero. If we wish to test the hypothesis that one of these coefficients is a random fluctuation from a population in which the true parameter value is zero, then the one parameter short-cut method described in the last section is appropriate, and it will produce the same parameter estimates and error variances as the general method.

The multi-parameter hypothesis that these three coefficients are simultaneously equal to zero in the population can be approximated by collecting four mean squares from the two ANOVA tables.

- The between mean square for y': MSB(y') = 33267.0762
- The within mean square for y' : MSW(y') = 1536.2916
- The between mean square for d': MSB(d') = 39.6227
- The within mean square for d' : MSW(d') = 128.5978

$$F = \frac{33267.0762}{1536.2916 + 128.5978 + 3 * 39.6227} = 18.6550$$

Figure 4-15 SPSS Output for Estimating Regression Coefficients on Multiple Parameters using the Short- Cut Method

```
        * * * *   M U L T I P L E   R E G R E S S I O N   * * * *

Equation Number 1    Dependent Variable..   AVE1_2

Variable(s) Entered on Step Number
   1..    M810705B  DO YOU AGREE: MATH USEFUL/SOLVING EVERYD
   2..    M810703B  DO YOU AGREE: I AM GOOD IN MATH
   3..    M810702B  DO YOU AGREE: ALL PEOPLE USE MATH IN THE

Multiple R            .28007
R Square              .07844
Adjusted R Square     .07482
Standard Error      39.19556

Analysis of Variance
                    DF      Sum of Squares      Mean Square
Regression           3         99801.22868      33267.07623
Residual           763       1172497.73348       1536.29158

F =      21.65414       Signif F =  .0000

------------------ Variables in the Equation ------------------
Variable                B          SE B        Beta          T   Sig T

M810705B       -2.134675    1.561350   -.051952     -1.367   .1720
M810703B      -10.712277    1.425517   -.272031     -7.515   .0000
M810702B        1.288545    1.802966    .027460       .715   .4750
(Constant)    289.722338    4.542554                63.780   .0000

End Block Number   1   All requested variables entered.

        * * * *   M U L T I P L E   R E G R E S S I O N   * * * *

Equation Number 2    Dependent Variable..   DIF1_2

Variable(s) Entered on Step Number
   1..    M810705B  DO YOU AGREE: MATH USEFUL/SOLVING EVERYD
   2..    M810703B  DO YOU AGREE: I AM GOOD IN MATH
   3..    M810702B  DO YOU AGREE: ALL PEOPLE USE MATH IN THE

Multiple R            .03478
R Square              .00121
Adjusted R Square    -.00272
Standard Error      11.34010

Analysis of Variance
                    DF      Sum of Squares      Mean Square
Regression           3           118.86797         39.62266
Residual           763         98145.85346        128.59782

F =        .30811       Signif F =  .8195

------------------ Variables in the Equation ------------------
Variable                B          SE B        Beta          T   Sig T

M810705B         .129182     .451731    .011313       .286   .7750
M810703B        -.248592     .412432   -.022715      -.603   .5469
M810702B        -.320848     .521636   -.024604      -.615   .5387
(Constant)       .492909    1.314256                  .375   .7077

End Block Number   1   All requested variables entered.
```

We note that this F-statistic is slightly different from the one computed using the general method, which was 18.325. We believe that this approximation will be close enough for exploratory analyses and for most practical purposes. It should be noted that if the F-statistic is insignificant for the average plausible value, it will also be insignificant when this approximation is used.

5. Creating Mini Files from NAEP Data

This chapter will describe how the NAEP mini-files were selected and how other types of mini-files can also be selected from the full NAEP Public Use Data Tapes. Selecting the mini-files from the NAEP Public Use Data Tapes is a two step process. The first step in this process is to compute the sum of the sampling weights for the members of the population or sub-population to be sampled. In this step the actual number of sample members in the files is also counted. The second step is to select members of the sample using a systematic sampling procedure with a random starting point. This systematic sampling procedure is based on the individual sampling weights rather than on the actual number of cases in the file or members of the population that is being sampled. The NAEP Public Use Data Tape will be required if the user wishes to select other mini- samples different from the one provided in the Primer Diskette.

There were several considerations in the selection of the mini-files from the main sample. Each student record in the main sample file has associated with it a sampling weight (WEIGHT) which should be used in calculating population estimates from the full sample data. As commonly used, the sampling weight assigned to each case in a sample is the reciprocal of the probability of selection for that particular case in the target population; that is, a student with a sampling weight of 200 had a probability of being selected into the sample of 1 over 200, or .005. That student, then, can be considered to represent approximately 200 students from the population. The sum of the sampling weights for members of a given sample is then equal to the population size that the sample represents. Individual students' sampling weights vary substantially due to intentional oversampling of certain sections of the population (i.e., inner city and private school students) and due to adjustments for nonresponses. It is important to note that in the case of the NAEP files, the sampling weight as defined above is multiplied by the sample size and divided by the population size. In this way, the sample weights are reduced so that the sum of the sampling weights for members of a given sample is made equal to the sample size. By transforming all sampling weights by the same factor the proportion of the population that each member of the sample represents is not altered.

One of the purposes of the NAEP mini-files is to select a subset of 1000 cases from the main sample. The cases for the mini-samples are selected in such a way that each member of the population sampled has an equal probability of being selected into the mini sample. This equal probability of selection avoids the need for using different individual sampling weights when doing statistical analysis using the mini files. To attain this end, it is important to select sample members with probability of selection proportional to their sampling weights in the main file. The selection of such a sample is what is explained below.

The general strategy for selecting the mini-files is that of a simple random sample, or spaced sample from the main file using the full individual sampling weights provided in the NAEP Public Use Data Tapes. The difference between this sampling procedure and the one generally known as systematic random sampling is that instead of selecting cases based on an interval of cases, the selection interval is based on the cumulative count of the individual sampling weights. This is, instead of selecting every Nth case in the file, we will be selecting every Nth accumulated weight in the file. This procedure requires that the total sum of the individual sampling weights for the members of the population from which the mini file is to be extracted will be computed in a first pass over the file. Following this, a random starting point is selected and every Nth weight equivalent case is then selected.

Each main file in the NAEP Public use Data Tapes contains different information that can be used to select cases or define subpopulations. In the case of the files used in this Primer, the students selected were only those who were actually in the 8th grade and who had a mathematics proficiency score assigned to them in the file. This is important to establish at the outset since the population from which the sample is desired needs to be defined prior to sampling, and the corresponding sum of the weights needs to be calculated for these cases in the sample. Depending on how the individual sampling weights were calculated, the sum of the weights could be equal to the estimated size of the population or to the actual number of cases in the sample --NAEP Public Use Data Tapes individual sampling weights adding to the latter. In either case, the end result of the selection process will be the same and the program presented later in this chapter does not need to be modified in any way. The sum of the sample weight will maintain the same relationship to the sample weights assigned to each one of the cases on the file.

General Method

The general method for selecting the cases for a mini sample file uses two passes over the NAEP Public Use Data Tapes files. The first pass over the file selects the NAEP Public Use Data Tapes sample members who are in the target population and sums their individual sampling weights. The sum of the sampling weights is then divided by the intended size of the mini-sample, which in our case is 1000. The resulting number is the size of the sampling interval. This sampling interval size corresponds to the cumulative distribution of sample weights for the target sample members, and not to the actual number of cases.

During the second pass over the file, the main sample is divided into 1000 segments and one student is selected in each segment. This is done by accumulating the sampling weights of the main sample members to isolate successive intervals and selecting one sample member from each interval.

Because the selection interval is based on the individual sampling weights, and not on the actual case count, the probability for students in the NAEP Public Use Data Tapes to be selected for the mini-file is proportional to their own sampling weight and thus inversely proportional to their probability of being selected for the full NAEP sample. Students with a high probability of being selected for the NAEP sample have a small probability of being selected for the mini-sample and vice versa. The result of this type of selection is that the students in the mini-files all have the same probability of selection from the targeted population. The larger their weight in the main sample, the greater the probability that they will be selected in the particular segment of cases where they happen to fall. The smaller their weight, the lesser the probability of them being selected in that particular segment.

Selecting one mini-sample member from each sampling interval also assures a broad representation of the students in the NAEP Public Use Data Tapes targeted population.

The student files in the NAEP Public Use Data Tapes are sorted by the school code. These files are sorted to permit direct match merging with the school, and excluded student files. Selecting the mini-samples from intervals ordered by school practically assures that all schools are represented in the mini-sample. That is, they will all have some probability of being selected into the mini file. The user may also choose, prior to selecting the cases, to sort them by any other variable which is deemed important for the research project, therefore assuring that the different groups in this variable are adequately sampled.

When performing this type of sampling the user must be aware that there still exists the possibility that a case with an extremely large sampling weight be selected twice into the mini- sample. This can only happen when the person's individual sampling weight is equal to or greater than the sampling interval being used. Only under these circumstances does the person stand the chance of being selected twice into the mini file. Otherwise, when the sampling interval is larger than the largest sampling weight for the members of the target population, none of the cases stand the possibility of being selected twice into the mini file. This may be a consideration in the selection of a sample size from the NAEP Public Use Data Tapes.

For our purposes we have selected the sample size to be 1000. The sample size of 1000 has been chosen out of convenience and because using this sample size ensures that none of the cases in the main file stand the chance of being selected twice for the mini- file. The sampling interval used was greater than any of the sampling weights for the cases. If a sample of 2000 had been selected instead then several members of the population could have been selected twice.

SELECTING THE CASES

Let us assume that we wish to select a mini-sample of 1000 cases from the main sample of 8th graders who have a mathematics proficiency score in the NAEP Public Use Data Tapes. To accomplish this the full NAEP Public Use Data Tapes must be available to you, as well as computer equipment necessary to handle these tapes. Selecting a mini-sample from the main NAEP file also requires knowledge of the content and format of the main files. This information can be obtained from the NAEP Public Use Data Tapes Users Guide.

In the example that follows we will demonstrate how the mini-sample file M08PS1.DAT -- included with the Primer Disk-- was created. This two step procedure is carried out using two SPSS programs which are detailed below. If the user chooses to do so, the sample can also be selected by merging the two programs into one and modifying it slightly[3]. We have

---

[3]If the user chooses to do so, the sample can be created with only one program by reading the file a first time, selecting the cases to sample from, aggregating the file to obtain the number of members of the population and the sum of their sampling weights, and then re-read the file a second time using the information derived from the AGGREGATE procedure. When the file is read the second time the sum of the weights is used to compute the interval for the selection of the sample.

chosen to carry out the selection process using the two programs instead of one because it illustrates better the process of selecting the sample and it uses less computer resources.

The procedures used for selecting the cases for the mini-sample take place within the INPUT PROGRAM and END INPUT PROGRAM commands of SPSS. These commands are only available on the mainframe, OS/2 and Windows version of SPSS, but not on the PC-DOS version. Because of the size of the data files in the NAEP Public Use Data Tapes, handling of the NAEP files using a PC is not recommended as the task can be very time and resource consuming.

STEP 1

The first pass over the NAEP Public Use Data Tapes file simply defines the targeted population and sums its sampling weights. The SPSS program used is shown in Figure 5-1. The program consists of :

- title information (Optional)
- definition of the NAEP Public Use Data Tapes data file and the variables that are needed to obtain the sum of the individual sampling weights and the number of members of the target population. In this case the only variables needed were WEIGHT, DGRADE and MRCPCM1.
- a first DESCRIPTIVE command which will give information about the total number of cases in the file.
- the SELECT command which is used to define the targeted population. In our case, we are selecting only those members of the 8th graders population which have a mathematics proficiency score.
- a second DESCRIPTIVE command which computes descriptive statistics for the variable WEIGHT. The statistic SUM needs to be explicitly requested, as it is not a default statistic reported by SPSS. Other statistics may be of interest such as minimum and maximum. The maximum will help determine if any of the cases stands the probability of being selected twice into the mini-sample. This was explained previously. Based on the maximum value for the variable WEIGHT, the user may reconsider the desired mini-sample size.

The results from this program can be seen on Figure 5-2. The sum of the weights for the members of the sample is 5991.58. Some other statistics worth noting is that there are the 6473 8th graders who have mathematics values, the average weight is .93, and the weights range between .20325 and 5.17654. If we divide the sum of the weights by the sample size of 1000 to obtain the sampling interval (5.99158), we can see that in this situation none of the cases will be sampled twice, so we do not need to concern ourselves over including the same case twice in the sample.

Figure 5-1 **SPSS Code to obtain the Number of Members of a Target population in the Main NAEP File and the Sum of their Individual Sampling Weights**

```
TITLE 'FIRST STEP IN SELECTING CASES FOR MINI FILE'

DATA LIST FILE = 'filename' NOTABLE /
   DGRADE 96-97 WEIGHT 177-183(5) MRPCMP1  1001-1005(2)

DESCRIPTIVE VARIABLES = ALL / STATISTICS = DEFAULT SUM
SELECT IF DGRADE = 8 AND NOT(MISSING(MRPCMP1))
DESCRIPTIVE VARIABLES = ALL / STATISTICS = DEFAULT SUM

FINISH
```

Figure 5-2 **SPSS Ouput with the Number of Members of a Target population in the Main NAEP File and the Sum of their Individual Sampling Weights**

```
    1  0
    2  0  DATA LIST FILE = 'TEMP:[SCRATCH.BCASSESS]Y21RMS2_MAT.DAT' NOTABLE
    3  0     / DGRADE 96-97 WEIGHT 177-183(5) MRPCMP1 1001-1005 (2)
    4  0
    5  0  DESCRIPTIVE VARIABLES = ALL
    6  0    / STATISTICS = DEFAULT SUM

Number of valid observations (listwise) =      8634.00

                                                                  Valid
Variable        Mean    Std Dev   Minimum   Maximum          Sum      N

DGRADE          7.73        .49         5         9     66761.00   8634
WEIGHT          1.00        .61    .20325   8.87082      8634.00   8634
MRPCMP1       259.80      33.48    149.28    370.23   2243078.02   8634

    7  0  SELECT IF DGRADE = 8 AND NOT(MISSING(MRPCMP1))
    8  0
    9  0  DESCRIPTIVE VARIABLES = ALL
   10  0    / STATISTICS = DEFAULT SUM
   11  0

Number of valid observations (listwise) =      6473.00

                                                                  Valid
Variable        Mean    Std Dev   Minimum   Maximum          Sum      N

DGRADE          8.00        .00         8         8     51784.00   6473
WEIGHT           .93        .52    .20325   5.17654      5991.58   6473
MRPCMP1       265.00      32.46    151.24    370.23   1715355.26   6473

   12  0  FINISH
```

A few other statistics that are worth noting is that prior to selecting the members of the target population there were a total of 8634 cases in the file. The mean weight was 1.00 and the sum of the weights in the sample was equal to the sample size. The students who were 13 years old but not in the eighth grade had higher weights on the average than the eighth graders.

It is also worth noting that this first step does not neccesarily require a computer run to count the cases in the file. This information can also be found in the NAEP Public Use Data Tapes Users Guide. For example, the sample size and sum of the weights for the 8th graders in the Mathematics file can be read directly from Table 7-2 in the Secondary User's guide. But this will not be the case if we were to sample from the members of a different, more specific population.

STEP 2

During the second step of the process the cases are actually selected. The SPSS program used for the second pass over the full NAEP data file is shown in Figure 5-3 and its output is shown in Figure 5-4. This program has the following features:

- Title information (this is optional).
- Sets the seed for the random number generator. This ensures that the same sequence of random numbers is generated each time the program is run, as long as the random number generator remains the same. The random number is used to sort the cases in the mini file. Once the cases are sorted, it allows the user to select the first, say 500 cases, compute statistics on this sub-sample, and then compare the results with those that are obtained with the remaining 500. Or in a classroom situation, the instructor may select 10 sets of sequentially selected samples of 100 cases each and compare the results. Because the mini sample has the characteristics of a simple random sample, these samples of 100 cases each will be statistically equivalent.
- Uses the INPUT PROGRAM procedure to read the main file and create the cases for the mini files.
- The first DATA LIST command specifies the name of the main file where the members of the target population are located, as well as variable names and their location in the file.
- The NUMERIC and LEAVE command define some temporary variables which will be used for the selection of the cases and that must remain unchanged across cases. The variables here defined are:
    - **WSUM:** The sum of the weights for the members of the sample, which was computed in the first pass through the file.
    - **INTV:** Length of the sampling interval.
    - **RN:** Randomly selected number, used for the selection of the specific case within a sampling interval. This number can be selected using a random number generator or --as it is done in this example-- made explicit. Changing this random number will result in a different set of cases being selected.
    - **V2SEL:** Stands for 'Value to select'. This corresponds to the cumulative sampling weight value to be selected within the sampling interval. Corresponds to (RN * INTV) plus the upper limit for the previous sampling interval. It is also equivalent to the previous selected value plus the sampling interval.
- A DO IF statement which processes the subsequent commands only for those cases which meet the specified condition. In our example the condition is that the cases be 8th graders with a non missing proficiency score in math.
- The following variables are then assigned a numeric value:

SEQNO keeps track of the corresponding sequence number of the case that is read in case that particular person case is read more than once into the mini sample.

WSUM to have the value of the sum of the weight of the members of the general file eligible to be in the mini-sample, or members of the sampled population.

INTV or the sampling interval size, which is equal to the sum of the weights divided by the sought sample size.

RN is defined to have the value of 0.6127884. This value is randomly chosen by the researcher prior to the selection process. Changing this value will result in a different set of cases being selected.

- A set of three variables is then computed for each case that is a member of the target population. There is a slightly different procedure if the case is the first or the second one.

LOWLMT is the lower limit in the cumulative distribution of sampling weights. Corresponds to the sum of the weight up to and including the prior selected sample member.

UPLMT is the upper limit of the cumulative distribution of sampling weights. Corresponds to the sum of the weights up to and including the selected sample member.

V2SEL is the cumulative weight value to be selected. It corresponds to the sampling interval in which the current case is included in.

SELVCTR This variable takes a value of 1 when the case is to be selected into the mini file and a value of zero when it is not.

- When the first case is read (SEQNO=1) it computes the values and evaluates the IF statement to see if the value to be selected is included in the interval corresponding to the current case. If it is included, then the variable SELVCTR (or selection vector) is set to one. All cases selected as members of the mini-file will have a value of 1 on this variable. For the rest of the cases (SEQNO > 1) it follows exactly the same procedure with two exceptions. First, given that the first case has no preceding case, the lower limit of the interval must be set to zero, whereas for the rest of the cases, the lower limit of the interval is set to the previous upper limit of the weight interval. Second, for the cases where SEQNO>1 it evaluates to see if the previous case was selected. If the previous case was selected, then it increases the V2SEL variable in the amount of one interval.

- The REPEATING DATA command may apply only when the value for the WEIGHT variable for the current case is greater than the sampling interval. A temporary variable is computed (#i) which returns the number of times that the sampling interval can be fitted above and beyond the selected value within the interval covered by the current case. If the interval goes beyond the upper limit of the interval, then the subsequent commands are not executed. But if the interval is contained, then the truncated value of the division (#i) will return the number of times the sampling interval can be included in the current case, and consequently the number of times the case needs to be re-read or repeated in the selected mini-sample. This is accomplished using the REPEATING DATA command which repeats the data contained in the case currently being read as many times as the sampling interval is included in the weight interval. The variable #DUMMY is a temporary scratch variable which is read for those cases only to make the procedure work. The subcommand OCCURS indicates how many times the case is to be repeated in the file.

- For this case the selection vector is set to 1, and at the end the value to select (V2SEL) is increased in the amount of the number of intervals contained within the case being read.
- The last commands end the case and the INPUT PROGRAM.
- Since the plausible values were revised for the 1990 data after the data files were first published, the user may need to update the proficiency scores with the newer one. This is accomplished by using the UPDATE command to replace the old values with the newer one. The user will not need to do this if the NAEP Public Use Data Tapes already contain the revised proficiency scores.
- The LIST command produces a listing of the first 50 cases which are members of the target population and that are in the main sample so that the user can manually check that the cases are being selected properly and that the program is producing the expected results. Careful attention must be placed to repeated cases. Figure 5-2b includes such listing and the user can see which cases are being selected to be in the mini- files.
- The cases for which the selection vector (SELVCTR) has been set to one are selected for the following procedures. There should be as many cases with SELVCTR = 1 as the requested sample size. This is further verified with the DESCRIPTIVE command which will provide descriptive statistics for the variables requested. The results from this procedure can also be used to verify further uses of the selected file, as well as to compare the characteristics of the selected sample with those of the main sample or of the general population.
- One last step before writing the cases to the file is to create a new variable called SORTER which will be a uniform random number assigned to each case. The cases will then be sorted using this variable. This will provide a randomly sequenced file as an end result. The cases in this file will not be ordered any more by any criteria by which they may have been ordered in the main data file. Another advantage of performing this final sort using the random number is that it will make it that much more difficult to trace back the cases selected to the main data file and consequently facilitating the opportunity of identifying individual subjects from the main data file.
- The last step in the selection of the file is writing the cases selected and the variables extracted to a raw data file that can then be read using other statistical systems, including SPSS. Notice that the variables are written to the new file in the same order in which they were read, but the column location has been changed to eliminate any blank spaces between the variables.

The programmer may be tempted to use SPSS system files to select the cases, but the INPUT PROGRAM as currently made available by SPSS does not work with SPSS system files.

A few comments on how these programs work:

- By setting the seed at the beginning of the program and forcing the random position within an interval to a particular random number (i.e., RN = 0.6127884), we assure that re-running this program will always result in the same mini-sample. The order of the sample members may differ if different random numbers are selected for the random sequencing.

- The programs described above were designed for simplicity of presentation and not for computer efficiency.

- We have chosen to select the mini-samples in a systematic manner with a random starting place. We could have selected a case at random within each interval by changing the definition of RN to be RN = Uniform(1) within each interval and redefining it within the selection loop.

- Alternatively, a simple random sample of the full file could be selected by generating 1000 random numbers in the range from 0 to WSUM, sorting them in ascending order, and then adjusting the program to select at these values instead of one case per interval. A simple random sample could not be guaranteed to span all primary sampling units and might have many multiply selected individuals.

- Besides showing how the NAEP mini file was created, describing these programs is also intended to encourage researchers to make mini-files to explore other populations of interest. For example, a sample of only children attending public schools could be selected, or a sample of only 13-year-old students. Again, the researcher should remember that when selecting different sample characteristics and restricting the sample size, the probability increases for the sample members to be selected more than once into the mini-samples.

- Other populations which may be sub-sampled are ages 9, 13 and 17. Sampling these populations would simply require changing the selection variables and the filenames where the samples are located. The NAEP data file is not adequate to estimate other age or grade populations; to illustrate, all of the 16-year-old in the NAEP sample are in grade 11 (except possibly for a few in grade 8) whereas the majority of 16-year-old are in grade 10, which was not sampled.

- These programs also suggest how other variables can be selected to create mini files. The full NAEP data files contain literally hundreds of student variables as well as more information about their teachers, schools, and communities. The full inventory of variables is too lengthy to be listed here but is available in the NAEP Public Use Data Tapes User's Guide. It should be noted, however, that much of the additional information is available for only random sub-samples of the full NAEP sample and thus variable selection should be done carefully, with due regard to the handling of missing data.

- Running the above program examples for selecting a sample from whom mathematics values are available requires only selecting on the availability of the

MRPCMP1 variable, which is in a given position in the NAEP Public Use Data Tapes, and some re-labeling of the output. Running the above programs for other grades requires identifying the appropriate file and assuring that the variables exist for those grade levels.

Figure 5-3 SPSS Code to Select cases for the NAEP Mini-Files

```
INPUT PROGRAM
DATA LIST FILE = 'filename' NOTABLE/
  YEAR         1-2        AGE         3-4        BOOK        5-6
  SCRID       54-59       DGRADE     96-97       WEIGHT    177-183(5)
        .
        .  (Include other relevant variables for the analysis...)
        .
  T031901  1383           T032001  1384          T032101  1385
NUMERIC wsum,intv,rn,v2sel
LEAVE   wsum,intv,rn,v2sel
+ do if (dgrade eq 8 and not missing(mrpcmp1))
+   COMPUTE SEQNO = seqno + 1
+   compute wsum = 5991.58
+   compute intv = wsum / 1000
+   compute rn   = 0.6127884
+   compute selvctr= 0
*          Here the SAMPLE selection takes place.
+   do if   (SEQNO = 1)
+      compute lowlmt = 0
+      compute uplmt  = lowlmt + weight
+      compute v2sel  = intv * rn
+      if (lowlmt le v2sel and uplmt ge v2sel) selvctr=1
+   end if
+   do if (seqno > 1)
+      compute lowlmt = lag(uplmt)
+      compute uplmt  = lowlmt + weight
+      compute v2sel  = lag(v2sel)
+      if (lag(selvctr)=1) v2sel = v2sel + intv
+      if (lowlmt le v2sel and uplmt ge v2sel) selvctr = 1
+   end if
+  compute #i = trunc((uplmt - v2sel)/intv)
+   do if (#i > 0)
+      repeating data file = 'filename' NOTABLE
         / occurs = #i / start = 1 / data = #dummy(a1)
+      compute v2sel = v2sel + (#i * intv)
+      compute selvctr = 1
+   end if
+ end case
+ end if
END INPUT PROGRAM
+ list variables = seqno weight lowlmt uplmt intv v2sel selvctr
    / cases from 1 to 200
select if selvctr=1
*
* Here the corrected plausible values replace the old plausible values!
*
sort cases by scrid
UPDATE FILE = *
     / FILE = '[bcassess.grade8]newpvs8.sys'
     / by scrid
*
SELECT IF SELVCTR = 1
compute sorter=uniform(10)
sort cases by sorter
WRITE OUTFILE = 'newfilename' NOTABLE
 / 1     YEAR        1-2        AGE         3-4        BOOK        5-6
           .
           . ( Include the rest of the variables for the analysis...)
           .
         T031901   203          T032001   204          T032101   205
         SCRID     210-215
EXECUTE
FINISH
```

56

Figure 5-4 SPSS Output after selecting population members for the Mini-Files

```
  1  0
  2  0  SET SEED = 28071964
  3  0
  4  0  INPUT PROGRAM
  5  0  DATA LIST FILE = 'TEMP:[SCRATCH.BCASSESS]Y21RMS2_MAT.DAT' NOTABLE/
  6  0    YEAR          1 -    2
                        .
                        .
                        .
154  0    T032101   1385
155  0
156  0  numeric wsum,intv,rn,v2sel,nsel
157  0  leave   wsum,intv,rn,v2sel,nsel
158  0  + do if (dgrade eq 8 and not missing(mrpcmpl))
159  1  +   COMPUTE SEQNO = seqno + 1
160  1  +   compute wsum = 5991.57579
161  1  +   compute intv = wsum / 1000
162  1  +   compute rn   = 0.6127884
163  1  +   compute selvctr= 0
164  1  *          Here the SAMPLE selection takes place.
165  1  +   do if   (SEQNO = 1)
166  2  +      compute lowlmt = 0
167  2  +      compute uplmt  = lowlmt + weight
168  2  +      compute v2sel  = intv * rn
169  2  +      if (lowlmt le v2sel and uplmt ge v2sel) selvctr=1
170  2  +   end if
171  1  +   do if (seqno > 1)
172  2  +      compute lowlmt = lag(uplmt)
173  2  +      compute uplmt  = lowlmt + weight
174  2  +      compute v2sel  = lag(v2sel)
175  2  +      if (lag(selvctr)=1) v2sel = v2sel + intv
176  2  +      if (lowlmt le v2sel and uplmt ge v2sel) selvctr = 1
177  2  +   end if
178  1  +  compute #i = trunc((uplmt - v2sel)/intv)
179  1  +   do if (#i > 0)
180  2  +      repeating data file = 'TEMP:[SCRATCH.BCASSESS]y21rms2_mat.dat' NOTABLE
181  2                    / occurs = #i
182  2                    / start  = 1
183  2                    / data   = #dummy(a1)
184  2  +      compute v2sel = v2sel + (#i * intv)
185  2  +      compute selvctr = 1
186  2  +   end if
187  1  + end case
188  1  + end if
189  0  END INPUT PROGRAM
190  0  EXECUTE

191  0  LIST VARIABLES = SEQNO WEIGHT LOWLMT UPLMT INTV V2SEL SELVCTR
            / CASES = 1 to 50

  SEQNO   WEIGHT   LOWLMT    UPLMT    INTV    V2SEL   SELVCTR

   1.00   .55781      .00      .56    5.99     3.67      .00
   2.00   .55781      .56     1.12    5.99     3.67      .00
   3.00   .75885     1.12     1.87    5.99     3.67      .00
     .
     .
     .
  48.00   .36535    28.43    28.79    5.99    33.63      .00
  49.00   .30477    28.79    29.10    5.99    33.63      .00
  50.00   .42450    29.10    29.52    5.99    33.63      .00

Number of cases read:  6,473    Number of cases listed:  50
```

(continues...)

57

BEST COPY AVAILABLE

Figure 5-4 SPSS Output after selecting population members for the Mini-Files (contiued)

```
 192  0  SELECT IF SELVCTR = 1
 193  0  *
 194  0  * Here the corrected plausible values replace the old plausible values!
 195  0  *
 196  0  sort cases by scrid
 197  0  UPDATE FILE = *
 198  0       / FILE = '[bcassess.grade8]newpvs8.sys'
 199  0       / by scrid
 200  0  *
 201  0  SELECT IF SELVCTR = 1
 202  0  descriptive variables = WEIGHT
 203  0  *

Number of valid observations (listwise) =      1000.00

                                                            Valid
Variable       Mean    Std Dev   Minimum   Maximum      N  Label
WEIGHT         1.21        .69    .22777   5.17654   1000

 204  0  compute sorter=uniform(10)
 205  0  sort cases by sorter
 206  0  compute weight = .8
 207  0  WRITE OUTFILE = '[BCASSESS.GRADE8]NEWM8PS1.DAT' /
 208  0    YEAR          1 -    2
                        .
                        .
                        .
 356  0    T032101    298
 357  0
 358  0  EXECUTE
 359  0
```

6. JACKKNIFE VARIANCE ESTIMATION IN NAEP

The purpose of this chapter is to describe the procedures used in computing jackknife variance estimates with the NAEP variables. The first part of the chapter briefly describes what the jackknife technique is, and why it is recommended to obtain estimates of the standard errors of statistics. Other alternate methods to the jackknife, such as the use of the design effect, are also presented. Towards the end of the chapter, SPSS code --which is included in the NAEP Secondary User's Guide (Rogers, et al, 1992)-- is presented and explained in detail. This code demonstrates how to compute jackknife variance estimates for NAEP variables using the information available in the main data files. Other examples of computing jackknife variance estimates are presented and their corresponding output is discussed.

Before continuing on the topic, we must warn the reader that the use of jackknife variance estimation techniques is computationally intensive and requires substantial computer time, as well as some post-processing after the variance estimates or its components are obtained. Commercially available statistical software such as SAS and SPSS do not compute jackknife variance estimates directly, but rather provide the user with commands that, when properly used, can produce them. Using such commands requires the user to write computer program code to deal specifically with the set of variables which is of interest. In the case of NAEP, to obtain the proper jackknife variance estimate requires the use of 57 different sets of sampling weights. Jackknife estimation is not recommended with the mini-files provided with this NAEP Primer. It is only recommended when using the full NAEP samples. When using the mini-files, the use of a weight of .8, which inflates the error variance estimate for the statistic of interest, is suggested. As it will be discussed later in this chapter, this weight of .8 provides a reasonable approximation to the values that would be obtained from jackknife variance estimate of the parameters.

ESTIMATING THE SAMPLING ERROR

Given that NAEP uses a complex sampling design in the selection of the students to be assessed, traditional statistical systems do not provide appropriate error variance estimates since they have been designed to deal specifically with samples that have been selected using simple random sampling techniques. This issue of obtaining inadequate variance estimates is dealt with in the mini files by using a sampling weight adjusted for the design effect.

As the reader should be aware at this point, there are two major error or variance components in the NAEP design. The first error component is the measurement error, which was explained in Chapter 4. The measurement error component is reflected in the use of two or more sets of plausible values to estimate the proficiency of populations or subgroups of the populations. The second major error component is related to the sampling error, and results from the particular sampling techniques employed by NAEP to obtain its samples.

This important error component --sampling error-- is related to the uncertainty in the estimation of population parameters of interest, and it results from the fact that the information about the estimate is obtained from a sample of the population, and not from the complete population. This sample from the population is selected to have certain properties and characteristics, and specific procedures are strictly followed --such as stratification and clustering-- that help obtain a representative sample from the population and, at the same time, allow for an efficient and economic sampling and data collection process.

One important way in which the complex sample design used by NAEP differs from that of a simple random sampling method is that the NAEP sampling procedure entails selecting a group of students from the same school, as well as clusters of schools from the same geographically defined primary sampling unit (PSU). As a consequence of this sampling procedure, the individual observations obtained from the subjects in the sample are not independent from one another as they would be if simple random sampling procedures had been used. When a particular school is selected into the sample with, say, twenty students in it, those twenty students will tend to be more alike than if 20 students had been chosen at random from the population. This similarity among the individuals selected has the effect of reducing the variation among the observations obtained from those individuals. Consequently, using standard formulas for estimating the standard errors of the sample statistics, such as means, percentiles, etc., would find that the standard error estimates would be smaller than those which would be obtained if appropriate procedures had been used. The standard error of a statistic, which is a measure of its variability, gives an indication of how precise the statistic would be in estimating the corresponding population parameter. This standard error of the statistic is also used to conduct significance tests and, if conventional simple random sampling statistical techniques were to be used without accommodation for the specific sampling design, statistically significant tests would occur at a higher rate than if the sampling design had been taken into consideration.

Given the importance and possible consequences of the studies that may be conducted with the NAEP data set, it is important to account for such underestimation of the error variance. To do so, it is necessary to compute the standard error of the statistic taking into account the implemented sampling design. There are several techniques available to accomplish this goal. Among them we find Hierarchical Linear Models, Bootstrapping methods, Balanced Repeated Replication, and Jackknife Repeated Replication (JRR). NAEP has traditionally used the jackknife method.

JACKKNIFE REPEATED REPLICATION (JRR) VARIANCE ESTIMATION

To account for the fact that there is some error involved in the way the sample is selected from the population, every statistic computed for the sample should be accompanied by a measure of the uncertainty, or sampling variability, associated with the corresponding statistic. This is equivalent to indicating how much the statistic would be expected to vary if the sampling procedure were to be repeated an indefinite number of times and the distribution of the statistic were constructed. For this reason, the particular sampling design used in selecting the sample in the first place must be taken into account when computing such measures of variability. If the data were to be treated as a simple random sample, without paying attention to the specific sampling design, the estimates of the sampling variability would tend to be underestimated.

As indicated earlier, there are alternatives that allow for the estimation of the sampling error for a statistic and that remove some of the conventions imposed by the methods of estimating variance for simple random samples. Such is the case of the JRR technique, which is considered a paired selection model because it assumes that the sampled population can be partitioned into $a/2$ strata, or Primary Sampling Units. This means that the Primary Sampling Units (PSUs) are paired by two independent selections. Following this first stage sampling, there may be any number of subsequent stages of selection that may involve equal or unequal probability of the corresponding elements. In such a way, the sample is constituted by H pairs of statistically equivalent samples. Each one of the elements within the pair in the sample can be substituted by the other element in the pair as they are considered to be statistically equivalent to each other. Differences between the elements in the pairs are considered to be part of the sampling error. Given this design, the JRR estimates of sampling variance are obtained as described below.

We assume the there are H strata each consisting of two ultimate PSUs. In the case of NAEP, this translated to there being 56 strata, each one of them containing 2 different, but equivalent and interchangeable samples. When computing a statistic "t" from the sample, the general formula for the JRR variance estimate of the statistics t is then given by the following equation:

Equation 6-1

$$JKK\,\text{var}_t = \frac{1}{2}\sum_{h=1}^{H}\left\{\left[t(J_h) - t(S)\right]^2 + \left[t(CJ_h) - t(S)\right]^2\right\},$$

where H is the number of pairs in the entire sample. The term $t(S)$ corresponds to the statistic computed for the whole sample, computed with any specific weights that may have been used to compensate for the unequal probability of selection of the different elements in the sample or any other post-stratification weight. The element $t(J_h)$ denotes the same statistic using the h^{th} jackknife replicate formed by including all cases not in the h^{th} stratum of the sample, removing all cases associated with one of the randomly selected PSUs of the pair within the h^{th} stratum, and including, twice, the elements associated with the other PSU in the h^{th} stratum. This is generally accomplished by zeroing out the weights for the cases of the element of the pair to be excluded from the replication, and multiplying by two the weights of the remaining element within the h^{th} pair. The element $t(CJ_h)$ denotes the h^{th} complement jackknife replicate formed in the same way as the h^{th} jackknife replicate with the eliminated and doubled elements of the pair interchanged.

As we can see from the formula above, the computation of the JRR variance estimate for any statistic from the NAEP files will require the computation of any statistic 113 times: once to obtain the statistic for the full sample, 56 times to obtain the statistics for each of the jackknife replicates (J_h), and 56 more times to obtain the statistics for each of the 56 complement jackknife replicates (CJ_h). But the cost involved in repeating the analysis for each of the complement jackknife replicates far outweighs the benefits gained from doing so. And for most practical purposes, when estimating linear statistics the jackknife replicate as well as its complement will be very similar. So the formula for estimating the JRR variance can be further reduced to:

Equation 6-2

$$JKK \text{ var}_t = \sum_{h=1}^{H} \left\{ \left[t(J_h) - t(S) \right]^2 \right\}$$

Notice that in this case the statistic for the jackknife complement ($t(CJ_h)$) does not need to be computed, and consequently the statistic needs to be computed only 57 times, instead of 113. The statistic is only computed for one of the elements of the pair of samples. This element is randomly chosen between the two elements of the pair. When using JRR techniques for the estimation of the sampling variability, the approach will approximately reflect the combined effect of the between and within PSU contributions to the variance.

A major expenditure of resources in the computation of a jackknife variance estimate occurs in the construction of the pseudo-replicates. This requires us to create a new set of weights for each pseudo-replicate sample and, when necessary, introduce the proper corrections to the weights because of non- response within the particular element of the pair.

Johnson (1987) indicates that the jackknife method is suitable for estimating sampling errors in the NAEP design because:

- it provides unbiased estimates of the sampling error arising from the complex sample selection procedure for linear estimates such as simple totals and means, and does so approximately for more complex estimates;
- reflects the component of sampling error introduced by the use of weighting factors that are dependent on the data actually obtained;
- it can be adapted readily to the estimation of sampling errors for parameters estimated using statistical modeling procedures, as well as for tabulation estimates such as totals and means; and
- once appropriate weights are derived and attached to each record, jackknifing can be used to estimate sampling errors.

This JRR procedure to estimate error variance will work well, for example, when estimating the proportion of boys and girls surveyed, or when estimating the amount of television watched by boys and girls across the nation. But when estimating statistics that are based on plausible values for the population or sub-groups of it, the computation of the standard error of a statistic needs some adjustments that were explained previously in Chapter 4. Because of the design of the cognitive item questionnaires used by NAEP, not all students respond to all of the items of the assessment. In fact, each surveyed individual responds to only about 3/7 of the total number of items included in the assessment. The plausible values for each of the respondents are estimated based on the information that is available from each of them. A random element is included in this plausible value to account for the uncertainty of the proficiency estimate. In this way, the uncertainty due to the measurement process is approximated, and is accounted for when estimating the variance of a statistic based on such proficiency scores. This uncertainty due to the measurement process must be accounted for when estimating the variance of a statistic.

DEGREES OF FREEDOM

When computing the error variance estimate for a statistic using the JRR techniques, the number of degrees of freedom will also vary from the number of degrees of freedom that would correspond to a simple random sample estimate. The effective number of degrees of freedom of the variance estimate of a statistic will, at most, be equal to the number of pairs used to form the pseudo replicates. The number of degrees of freedom is equal to the number of independent pieces of information used to estimate the variance. For the main assessment there are a total of 56 pieces of information (56 pairs of PSU) used to estimate the JRR variance, each of which provides at most 1 degree of freedom, regardless of the number of individuals within each pair. If the differences between the pairs are not normally distributed, or if some of the squared differences are considerably larger or smaller than others, then the degrees of freedom of the variance estimate will be less than the number of pairs used to obtain it.

An estimate of the effective number of degrees of freedom for the variance of a statistic comes from an approximation given by the formula:

Equation 6-3

$$df_{eff} = \left(\sum_{i=1}^{M} (t_i - t)^2 \right)^2 \Big/ \sum_{i=1}^{M} (t_i - t)^4 \ ,$$

where M is the number of pairs used for estimating the JRR variance estimate, t_i is the statistic obtained for the i^{th} pseudo replicate, and t is the statistic obtained for the full sample. For more details and a full explanation on the computation of the degrees of freedom, see Johnson and Rust (1992).

APPROXIMATIONS

A JRR estimate of the variability of a statistic based on one or more observed NAEP variables in the 1990 sample requires computing the statistic of interest 57 times. The first time is to obtain the value for the statistic, and 56 additional times, each to compute the contribution of each of the 56 sampling pairs to the variance of the estimates. When estimating the variability for a statistic that involves one of the proficiency scales, this procedure would have to be repeated for each of the imputed scores. In the case of NAEP, this implies repeating the above procedure five times. This also implies that the full implementation of the JRR to estimate the variance estimate for a statistic would require computing the statistic as many as 285 times. This would include 57 runs to obtain a variance estimate for each of the five sets of plausible values.

An alternative to this approach is to account approximately for the effects of the sampling design by using an inflation factor, called the design effect (DE), developed by Kish (1965) and extended by Kish and Frankel (1974). The DE for a statistic is defined as the ratio of the actual variance of the statistic taking the sampling design into account (i.e., computed via the JRR estimation procedure), over the simple random sample variance estimate based on the same number of elements. In the case of NAEP this would involve computing the JRR variance estimate, and dividing it over the simple random sample estimate that would be obtained by using standard statistical packages.

This DE may be used to adjust the error estimates based on simple random sampling assumptions, and to account approximately for the effect of the sampling design. In practice, this is generally achieved by dividing the total sample size by the design effect and using this effective sample size in the computation of the error variance. Another way in which this is implemented is by adjusting the sample weights by dividing each one of them by the design effect. When the JRR variance estimate is greater than the simple random sample estimate, the DE will be greater than 1 and the sampling weights are consequently deflated, thus resulting in a reduced sample size used to compute the statistics. It is important to note that the reduction of the effective sample size does not alter the linear statistic computed, but it does alter the estimate of it's error variance.

The value of the design effect will depend on the type of statistic computed and the particular variables considered in the analysis, as well as the clustering effects occurring among sampled elements and the effects of any variable weights resulting from variable overall sampling fractions. It is worth pointing out that in order to compute the DE, the JRR variance estimate needs to be computed, thus making it unnecessary to use the design effect since a "better" estimate has already been obtained. But in some cases, as it is suggested in this paper, instead of using a DE that is specific for each and every possible combination of variables, an average of the overall design effect may have already been computed for the survey, and this average design effect can then in turn be used to adjust the individual sampling weights. Since the design effects vary across the different possible analyses, using the average DE will in some cases underestimate, and in others overestimate the sampling variance, but on the average, the variance estimate would be expected to be reasonably unbiased.

There are several possible ways in which the standard errors for statistics can be computed. When no proficiency scales are involved, the computations of the JRR variance estimate are greatly reduced since only one set of 57 statistics needs to be computed. But when proficiency scales are involved, then more sophisticated and complicated analysis may need to be performed to obtain adequate results. NAEP (Rogers, et al, 1992) recommends the following alternatives when estimating variability of statistics:

- Full implementation (285 runs): this would involve obtaining JRR variance estimates for each of the five plausible values provided for the individuals, and then combining the results. Even though this would provide the best estimate of the variance of a statistic, it is time consuming and may even discourage researchers from trying to implement it. It is believed that the extra work necessary to obtain the corresponding variance estimates using this method far outweighs the benefits.

- Estimates based on five sets of plausible values, jackknife based on one set of plausible values (61 runs): this is the procedure used by NAEP in reporting proficiency scores at the national level. The estimate of the variance of a statistic is based on the JRR estimate of the variance of one of the plausible values, generally the first one, with a correction for imputation using the five sets of plausible values.

In the examples presented later in this chapter, this is the procedure that is implemented. An advantage of this method is that the computational requirements are significantly reduced by performing the JRR on only one set of plausible values, but some information may be lost by doing so. The amount of information lost in this case is believed to be negligible.

- Estimates based on five sets of plausible values, design effect for sampling variance. In this method, no JRR variance estimate is obtained, but rather the average DE reported for the NAEP survey is used. This is what we recommend to use when working with the mini-files included in this primer. By using the average design effect, the effective sample size is reduced, and the resulting variance estimates for the statistics are consequently inflated. The main advantage of this method is that it would only require computing the statistic of interest five times, and then correcting the variance estimate for the imputation. This makes this procedure computationally simpler than any of the previous methods. The main disadvantage is that by using the average of the DE for all of the analyses, the variance estimate will be over or underestimated, and only on the average will it be the correct one. Still, it does provide a good approximation when exploratory analysis is being performed on the data.

- Estimates based on M sets of plausible values, where $1 < M < 5$, design effect for sample variance. This is similar to the previous one, but less than five plausible values, and at least two, are used to approximate the error due to imputation. Some of the information due to the imputation of the proficiency scores is lost in the process, but the computations are simpler and less cumbersome. The variance estimates still include a component for the imputation and the uncertainty of the proficiency score.

- Estimates based on one set of plausible values, design effect for sampling variance. This is by all means the least computationally intensive of the methods, and generally the least accurate. Since only one plausible value is used, no information about the imputation process and the uncertainty of measurement is included in the estimate of the variance.

Under no circumstances --regardless of the approach taken to obtain variance estimates-- should the variance estimate be computed by using the average of all or any set of the plausible values. Variance estimates obtained in such a way will always be underestimated and will consequently lead to an inflated number of significant results. They must also account for the specific features of the sampling design, either by using the design effect to adjust the sampling weights, or by using the JRR variance estimation technique.

When obtaining the variance estimates in analysis that do not involve any of the proficiency scales, the analysis is greatly simplified because the measurement error term does not need to be included in the estimation. In this case, the maximum number of runs needed to obtain the JRR variance estimates will be 57, and 1 if the design effect is to be used to adjust the sampling weights.

OBTAINING JRR VARIANCE ESTIMATES FROM THE NAEP DATA

Even though it was said above that numerous analyses needed to be performed in order to obtain jackknife variance estimates, they can be easily obtained, in some cases, and for some statistics, using SPSS computer code that is included in the NAEP Secondary User's Guide (Roger, et al, 1992). This SPSS computer code is presented and explained below. This computer code has the useful feature that the estimates are obtained in one pass over the data file and the output directly gives all of the estimates necessary for the JRR estimation, without further processing on the part of the user. The code presented can be used in SPSS on the mainframe, or in the newer versions of SPSS for the Windows and OS/2 environment. Some of the commands necessary to simplify the analysis are not available in any of the versions of SPSS/PC. If the analysis is to be performed on the PC version of SPSS, then the computer code required increases with the number of different sampling weights used to obtain the estimates.

There are two different programs presented below, the first of which can be used to obtain the JRR variance estimates when no plausible values are involved. The second one is used when estimating the variance for the proficiency scores, which require a correction for imputation after the corresponding correction for sampling. A set of examples of the code with its corresponding output follows together with its explanation.

ANNOTATED COMMAND FILE

In this section, the commands necessary to compute the JRR using SPSS are detailed and described. The code presented in this section is taken from the NAEP Public Use Data Tapes User's Guide (Rogers, et al, 1992). Some minor modifications have been made to the original code. We must again point out that the JRR variance estimation should be done with the full NAEP file which contains the proper replicate sample weights. The set of 56 replicate sample weights are included in this file with the names SRWT01 to SRWT56, which stand for Student Replicate Weight, followed by the corresponding number (01 through 56, one for each pair of units). The first step in the analysis is to select the variables that will be included in the analysis. Following this are the rest of the commands that perform computation of the variance. Two examples are included in this section. The first one computes the mean number of hours that the 8th graders watch television, separated by gender. This example shows how to compute the JRR variance estimate when the dependent variable is assumed to be known with certainty. The second example computes the variance estimate of a statistic for a plausible value where there is sampling as well as measurement error to account for. This last example includes the correction for the JRR variance estimation as well as a correction for imputation. The output corresponding to each example will also be presented. For the first two examples, the means and their standard errors are computed. It is important at this point to remind the reader that when there are no sampling weights involved, the mean is mathematically defined as

Equation 6-4
$$\sum_{i=1}^{n} x_i \Big/ n ,$$

where n is the total sample size, and x_i is the value of the variable x for each of the individuals in the sample. When there are individual sampling weights to be used to compute the mean, then the mean is computed as

Equation 6-5
$$\sum_{i=1}^{n} (x_i * w_i) \Big/ \sum_{i=1}^{n} w_i ,$$

where w_i is the individual case weight. Understanding of these formulas will aid in understanding the computer code that follows.

EXAMPLE 6-1

In this example, the purpose is to estimate the number of hours that boys and girls in the 8th grade watch TV daily. The ultimate goal is to determine if there are differences between boys and girls in terms of the numbers of hours they watch TV daily. For that purpose we need the mean number of hours that each group watches TV, as well as their appropriate standard error of the mean. The variables necessary for this analysis are DSEX, B001801A, and the set of 57 replicate sampling weights (WEIGHT and SRWT01 to SRWT56) provided in the main file. The students that reported watching 6 or more hours of TV daily will be assigned a value of 6 hours per day. The analysis is conducted in the following way. The numbers in the paragraph correspond to those included in Figure 6-1.

1. The system file containing the variables for the 8th grade sample is read and the variables pertinent to the analysis are selected from it. The variable DGRADE is kept because it will be used to select only those students that are in the 8th grade.
2. The 8th graders, as well as those who have valid values recorded for the variable B001801A, which is the amount of TV watched daily, are selected from the files. Valid values for the variable B001801A are those between, and including, 1 and 7.
3. The term WTX is computed as the product of the value for the case on the variable B001801A times the individual full student sample weight. This term is the same as (w_i*x_i) described above and which will be used to compute the mean number of hours that the 8th graders watch TV.
4. As indicated previously, in order to compute the JRR variance estimation for a variable in the NAEP files, the statistic needs to be computed 57 times. In order to reduce the amount of code needed for the analysis, and to perform the analysis in one pass over the data file, vectors of variables are defined. The vector WT corresponds to the student replicate weights (56 in total), and the vector WX corresponds to the term w_i*x_i, necessary to compute each of the means of the replicate samples. In this case B001801A is multiplied by each of the student replicate weights.

Figure 6-1 Standard error computation: Jackknife Multiweight method (SPSS Commands)

```
1  get file = '[bcassess.grade8]jackex1.sys'
    / keep = dgrade dsex b001801a weight srwt01 to srwt56

2  select if dgrade = 8
   select if b001801a < 8 and b001801a > 0

3  compute wtx = weight * b001801a

4  vector wt = srwt01 to srwt56
   vector wx(56)
   loop #i = 1 to 56
   +  compute wx(#i) = wt(#i) * b001801a
   end loop

5  aggregate outfile = *
    / break = dsex
    / uwn = n(weight)
    / swt,sw1 to sw56 = sum(weight,srwt01 to srwt56)
    / swx,sx1 to sx56 = sum(wtx,wx1 to wx56)

6  compute xbar = swx / swt
   compute xvar = 0

7  vector sw = sw1 to sw56
   vector sx = sx1 to sx56
   loop #i = 1 to 56
   + compute #jrsm  = sx(#i) / sw(#i)
   + compute #diff  = #jrsm - xbar
   + compute xvar   = xvar + (#diff * #diff)
   end loop

8  compute xse = sqrt(xvar)

9  print format xse (f8.4)
   report format = list
    / variables = dsex (label),uwn,swt,xbar,xse
```

5. At this point, we have all of the elements necessary to compute the 57 means needed to compute the JRR. By using the command AGGREGATE, and the summary function SUM, SPSS obtains the sum of the weights (Sw_i) as well as the weighted sum of B001801A ($S(w_i*x_i)$) for the full sample, as well as for each of the pseudo replicate samples. The accumulated vectors are kept in the variables SWT (Sum of the weights for the full sample), SW1 to SW56 (the sum of the weights for each of the 56 pseudo-replicate samples), SWX (Sum of the weighted X for the full sample), and SX1 to SX56 (the sum of the weighted X for each of the 56 pseudo-replicate samples). The resulting file contains two records, one for each of the values of the variables DSEX. Each record contains a total of 116 variables.

6. The mean on the variable B001801A for each of the groups of the sample is obtained here by dividing the sum of the weighted x (SWX) by the sum of the weights (SWT). The accumulator for the JRR variance (XVAR) is also initialized to the value of zero. This step of initializing the variance is necessary in order to allow for the accumulation of the 56 variance elements to proceed in step 7. These steps are done automatically for each of the records or cases on the file, which correspond to the two values of the variable DSEX.

7. In this step the JRR variance estimate is obtained in the following way: The first compute statement gives the mean Jackknife replicate sample (JRSM). This is computed by dividing the corresponding terms for each of the 56 pseudo replicate samples. On the next statement, the difference (#DIFF) between the mean for the pseudo-replicate sample (#JRSM) and the mean for the whole group (XBAR) is

computed, then squared and added to the variable that accumulates the variance components (XVAR). This process is repeated a total of 56 times, one for each of the pseudo-replicate samples in the file.

8. Once the JRR variance estimate is obtained (XVAR), the standard error of the statistics (XSE), in this case the mean, is obtained by extracting the square root of the JRR variance (XVAR) of the statistics.

9. This final section of the computer code assigns a print format to the variables of interest, and produces a report where the labels for the variable DSEX are printed out. The unweighted n for each of the groups (UNW), sum of the weights (SWT), mean value for the variable B001801A (XBAR), and its standard error (XSE) are requested as part of the report. The resulting output is shown in Figure 6-2.

Figure 6-2 Standard error computation: Multiweight method

GENDER	UWN	SWT	XBAR	XSE
MALE	3206	2985.75	4.32	.0395
FEMALE	3238	2978.10	4.24	.0429

EXAMPLE 6-2

In this example, the mean proficiency score and its standard error is computed for 8th grade boys and girls separately. Some of the steps are very similar to those presented in the previous example, but with the added complexity that in this case we must include the error due to imputation in the estimation of the standard error of the mean. The procedure described below is that performed and recommended by NAEP, in which the statistic of interest is computed using all 5 plausible values, but the jackknife variance estimate is obtained based on only the first plausible value. This reduces the number of statistics that need to be computed from 285 to 61.

The computer code is presented in Figure 6-3, and is described below. The numbers preceding the paragraph correspond to those in Figure 6-4.

1. The system file containing the variables for the 8th grade sample is read and the variables pertinent to the analysis are selected from it. The variable DGRADE is kept because it will be used to select only those students that are in the 8th grade. Since we will be estimating the mean proficiency in the mathematics scale, all five composite plausible values (MRPCMP1 to MRPCMP5) are kept for the analysis.

2. The 8th graders are selected from the files, as well as all of those who do not have missing values on the composite scale. Even though all students should have a proficiency score as part of their record, this statement ensures that cases are to be excluded if the proficiency score is missing.

3. The term WTX is computed as the product of the value for the case on the first plausible value times the individual full student sample weight. This term is the same as (w_i*x_i) described above and which will be used to compute the JRR variance estimate. Since the JRR variance estimate will be based on only the first plausible value, this is only done using MRPCMP1.

Figure 6-3 **Standard error computation: Jackknife multiweight method with correction for Imputation (SPSS commands)**

```
1    get file = '[bcassess.grade8]jackexl.sys'
      / keep = dgrade dsex weight srwt01 to srwt56 mrpcmpl to mrpcmp5

2    select if dgrade = 8
     select if (not sysmis(MRPCMP1))

3    compute wtx = weight * MRPCMP1

4    vector wt = srwt01 to srwt56
     vector wx(56)
     loop #i = 1 to 56
     +  compute wx(#i) = wt(#i) * MRPCMP1
     end loop

5    vector value = mrpcmpl to mrpcmp5
     vector ws(5)
     loop #i = 1 to 5
     +  compute ws(#i) = value(#i) * weight
     end loop

6    aggregate outfile = *
      / break = dsex
      / uwn = n(weight)
      / swt,swl to sw56 = sum(weight,srwt01 to srwt56)
      / swx,sxl to sx56 = sum(wtx,wxl to wx56)
      / ssl to ss5      = sum(wsl to ws5)

7    compute xbar = swx / swt
     compute xvar = 0

8    vector sw = swl to sw56
     vector sx = sxl to sx56
     loop #i = 1 to 56
     + compute #jrsm  = sx(#i) / sw(#i)
     + compute #diff  = #jrsm - xbar
     + compute xvar   = xvar + (#diff * #diff)
     end loop

9    vector ss = ssl to ss5
     loop #i = 1 to 5
     + compute ss(#i) = ss(#i) / swt
     end loop
     compute pvmean = mean(ssl to ss5)

10   compute ssvar  = variance(ssl to ss5)
     compute xse    = sqrt(xvar + (6/5) * ssvar)

11   print format xvar,xse,pvmean (f8.4)
     report format = list
      / variables  = dsex (label),uwn,swt,pvmean,xbar,xse
```

4. As indicated previously, in order to compute the JRR variance estimation for a variable in the NAEP files, the statistic needs to be computed 57 times. In order to reduce the amount of code needed for the analysis, and to perform the analysis in one pass over the data file, vectors of variables are defined. The vector WT corresponds to the student replicate weights (56 in total), and the vector WX corresponds to the term w_i*x_i, necessary to compute each of the means of the replicate samples. In this case the value on the first plausible value (MRPCMP1) is multiplied by each of the student replicate weights.

5. The vectors for the weighted plausible values are then created in this step. These will be used to obtain the mean plausible value for the groups of interest. Again, as NAEP has suggested, all five plausible values will be used only in the estimation of the error due to imputation, but not in the estimation of the error due to sampling computed with the JRR procedure.

6. At this point, we have all of the elements necessary to compute the 57 means needed to compute the JRR variance estimate, as well as the five mean plausible values to compute the error due to imputation. By using the command AGGREGATE, and the summary function SUM, SPSS obtains the sum of the weights (Sw_i) as well as the weighted sum of each of the plausible values ($S(w_i*x_i)$) for the full sample, as well as for each of the pseudo replicate samples. The accumulated vectors are kept in the variables SWT (Sum of the weights for the full sample), SW1 to SW56 (the sum of the weights for each of the 56 pseudo-replicate samples), SWX (Sum of the weighted X for the full sample), SX1 to SX56 (the sum of the weighted X for each of the 56 pseudo-replicate samples), and SS1 to SS5 (sum of each of the weighted plausible values). In this example, the resulting file contains 2 records, one for each of the values of the variable DSEX. Each record contains a total of 121 variables.

7. The mean (XBAR) of the first plausible value (MRPCMP1) for each of the groups of the sample is obtained here by dividing the sum of the weighted x (SWX) by the sum of the weights (SWT) for the full sample. The accumulator for the JRR variance (XVAR) is also initialized to the value of zero during this step. This step of initializing the variance is necessary in order to allow for the accumulation of the 56 variance elements to proceed in step 8. These steps are done automatically for each of the records or cases on the file, which correspond to the two values of the variable DSEX.

8. In this step the JRR variance estimate is obtained in the following way: The first compute statement gives the mean for the Jackknife replicate sample (#JRSM). This is computed by dividing the corresponding terms for each of the 56 pseudo replicate samples. On the next statement, the difference (#DIFF) between the mean for the pseudo-replicate sample (#JRSM) and the mean for the whole group (XBAR) is computed, then squared and added to the variable that accumulates the variance components (XVAR). This step is repeated a total of 56 times, one for each of the pseudo-replicate samples in the file. The resulting term XVAR is the variance due to sampling. If we were working with a variable which was assumed to be known with certainty, we would stop here and use XVAR as the estimate of the variance for the statistic of interest. But since the statistic of interest in this case is the mean proficiency, which is known to be measured with uncertainty, and this uncertainty is reflected by the imputation process that yields the five plausible values, the error due to imputation must then be computed and added to the variance term. This is accomplished in steps 9 and 10.

9. Here, the mean of each of the five plausible values (SS1 to SS5) is computed. This will serve two purposes. First of all, the variance of the five means is used as a component of the variance due to imputation. This is done in step 10. Second of all, as the reader should remember, the statistic reported should be the mean of the statistics obtained with each one of the plausible values. Thus the variable PVMEAN is such a statistic, and is the one that should be presented in the final report.

10. Once the JRR variance estimate is obtained (XVAR), the standard error of the statistics (XSE), in this case the mean proficiency score in mathematics for boys and

girls in the 8th grade, is obtained by extracting the square root of the JRR variance estimate, plus 6/5 of the variance due to imputation. For more details on the explanation on the computation of the error due to imputation refer to Chapter 5 in this book. Thus the term XSE is our final estimate of the standard error of the mean proficiency score. Again, this error based on the sampling variance of the first plausible value and the measurement error from the set of five plausible values.

11. This final section of the computer code assigns a print format to the variables of interest, and produces a report where the labels for the variable DSEX are printed out. The unweighted n for each of the groups (UNW), sum of the weights (SWT), mean value for the combined plausible values (PVMEAN) as well as the mean value for the first plausible value (XBAR), and the standard error (XSE) of the mean are printed as part of the report. The resulting output is shown in Figure 6-4.

Figure 6-4 **Standard error computation: Jackknife Multiweight method with correction for imputation (SPSS Output)**

GENDER	UWN	SWT	PVMEAN	XBAR	XSE
MALE	3218	2997.64	265.5500	265.46	1.2666
FEMALE	3255	2993.94	264.3834	264.30	1.0577

EXAMPLE 6-3

In the examples presented above, the statistics of interest were the average number of hours the student watches TV, or the mean proficiency score in the mathematics composite scale. The statistics were computed for only two subgroups of the population and even though the analysis is more complicated than computing variance estimates based on the assumptions of simple random sampling, the code and the processing of the data is pretty straightforward. But the researcher may be interested in more complicated analysis where more than one grouping variable is of interest, and even when more than one statistic within those subgroups is of interest. This requires some more processing of the data and more complex computer code, but it can still be accomplished in one pass over the data. Several levels of aggregation may need to be performed to accomplish this as well as the creation of intermediary files.

This is what is shown in the following example. There are two sets of statistics that are of interest in this example. The first set of statistics of interest are the mean proficiency scores for boys and girls in the 8th grade, broken down by the amount of hours that each group watches television (B001801a). The second set of statistics is the proportion of student that fall under each of the categories of the variable B001801a (Frequency of watching TV), broken down by gender (DSEX). For each set of statistics we want to obtain the population estimate for the statistic as well as its corresponding standard error. The code necessary to perform such analysis is presented in Figure 6-5, and its corresponding output appears in Figure 6-6.

Figure 6-5 **Standard error computation: Jackknife Multiweight method for proportions and proficiency levels with correction for imputation (SPSS Commands)**

```
get file = 'system_file_for_example'

select if b001801a < 8
weight by weight
sort cases by dsex
split file by dsex
oneway variables = mrpcmpl by b001801a (1,7)
  / format = labels
  / statistics = descriptives
split file off
weight off

compute wtx = weight * mrpcmpl

vector wt = srwt01  to srwt56
vector wx(56)
loop #i = 1 to 56
+  compute wx(#i) = wt(#i) * mrpcmpl
end loop

vector pv = mrpcmpl to mrpcmp5
vector wpv(5)
loop #i = 1 to 5
+  compute wpv(#i) = weight * pv(#i)
end loop

aggregate outfile = *
 / break = dsex b001801a
 / uwn = n(weight)
 / swt,swl to sw56 = sum(weight,srwt01 to srwt56)
 / swx,sxl to sx56 = sum(wtx,wxl to wx56)
 / swpvl to swpv5  = sum(wpvl to wpv5)

aggregate outfile = '[bcassess.grade8]dsexsw.sys'
 / break = dsex
 / tuwn  = sum(uwn)
 / totsw ,totswl to totsw56 = sum(swt,swl to sw56)
 / totswx,totsxl to totsx56 = sum(swx,sxl to sx56)
 / totswpvl  to totswpv5   = sum(swpvl   to swpv5)

compute con = 1
aggregate outfile = '[bcassess.grade8]totsw.sys'
 / break = con
 / totsw,totswl to totsw56 = sum(swt,swl to sw56)

match files
 / file  = *
 / table = '[bcassess.grade8]dsexsw.sys'
    / drop = tuwn totswx totsxl to totsx56 totswpvl to totswpv5 con
 / by dsex

add files
 / file = *
 / file = '[bcassess.grade8]dsexsw.sys' / in = con
   / rename
      (totswx,totsxl to totsx56 = swx, sxl to sx56)
      (totsw ,totswl to totsw56 = swt, swl to sw56)
      (totswpvl to totswpv5     = swpvl to swpv5 )
      (tuwn                     = uwn)

match files
 / table = '[bcassess.grade8]totsw.sys'
 / file  = *
 / by con

recode b001801a (missing,sysmis=-88)(else = copy)
add value labels b001801a -88 'Total'
```

(contiues...)

BEST COPY AVAILABLE

Figure 6-5 **Standard error computation: Jackknife Multiweight method for proportions and proficiency levels with correction for imputation (continued)**

```
compute xvar = 0
compute xbar = swx / swt
compute pvar = 0
compute pbar = (swt / totsw) * 100

vector sw  = sw1 to sw56
vector sx  = sx1 to sx56
vector tsw = totsw1 to totsw56
loop #i = 1 to 56
+ compute #xdiff = (sx(#i) / sw(#i)) - xbar
+  compute xvar  = xvar + #xdiff * #xdiff
+ compute #pdiff = 100 * (sw(#i) / tsw(#i)) - pbar
+  compute pvar  = pvar + #pdiff * #pdiff
end loop

vector swpv= swpv1 to swpv5
vector pvbar(5)
loop #i = 1 to 5
+ compute pvbar(#i) = swpv(#i) / swt
end loop

compute meanpv = mean(pvbar1 to pvbar5)

compute pvvar = variance(pvbar1 to pvbar5)
compute xvar  = xvar + (6/5) * pvvar
compute xse   = sqrt(xvar)
compute pse   = sqrt(pvar)

sort cases by dsex b001801a

set width = 132
print format pse xse pbar xbar meanpv (f8.3)
report
 / format    = list automatic
 / variables = dsex (label), b001801a (label),
               uwn,swt,pbar,pse,meanpv,xbar,xse
```

Figure 6-6 **Standard error computation: Jackknife Multiweight method for proportions and proficiency levels with correction for imputation**

Simple Random Sample Variance Estimates

DSEX = MALES

GROUP	COUNT	MEAN	STANDARD DEVIATION	STANDARD ERROR	MINIMUM	MAXIMUM
NONE	55	275.9943	36.1229	4.8587	191.7700	343.6900
1 HOUR O	325	272.6858	33.8802	1.8777	168.7200	370.2300
2 HOURS	660	271.4032	34.9047	1.3577	172.8000	361.4600
3 HOURS	670	268.8143	32.3509	1.2498	169.9900	367.1800
4 HOURS	549	265.1473	30.6248	1.3065	151.2400	357.4900
5 HOURS	293	258.5682	31.9843	1.8680	164.9500	358.3100
6 HOURS	431	249.9964	30.1713	1.4527	164.9100	345.8100
TOTAL	2985	265.5430	33.3701	.6107	151.2400	370.2300

DSEX = FEMALES

GROUP	COUNT	MEAN	STANDARD DEVIATION	STANDARD ERROR	MINIMUM	MAXIMUM
NONE	28	262.6458	33.8056	6.3347	186.1800	323.0400
1 HOUR O	412	272.2171	32.2633	1.5889	186.1300	357.0000
2 HOURS	648	272.6547	29.7743	1.1688	169.5800	341.9400
3 HOURS	686	267.2114	29.5691	1.1282	169.8600	357.4900
4 HOURS	511	263.0063	28.9180	1.2788	180.4400	347.1600
5 HOURS	302	259.0167	27.4491	1.5792	171.4900	342.2300
6 HOURS	388	243.9385	30.6576	1.5564	168.8900	332.3200
TOTAL	2978	264.4614	31.2008	.5717	168.8900	357.4900

Jackknife Variance Estimates

GENDER (WESTAT)	HOW MUCH TV DO YOU USUALLY WATCH EACH DAY	UWN	SWT	PBAR	PSE	MEANPV	XBAR	XSE
MALE	Total	3206	2985.75	50.064	.672	265.649	265.543	1.252
MALE	NONE	57	55.27	1.851	.191	276.118	275.994	6.510
MALE	1 HOUR OR LESS	352	325.57	10.904	.614	272.687	272.686	2.074
MALE	2 HOURS	724	660.91	22.135	.817	271.779	271.403	1.830
MALE	3 HOURS	728	670.04	22.441	.816	268.864	268.814	1.430
MALE	4 HOURS	575	549.44	18.402	.817	265.613	265.147	1.664
MALE	5 HOURS	317	293.18	9.819	.615	258.639	258.568	2.371
MALE	6 HOURS OR MORE	453	431.34	14.446	.861	249.421	249.996	1.603
FEMALE	Total	3238	2978.10	49.936	.672	264.558	264.461	1.070
FEMALE	NONE	37	28.48	.956	.177	263.894	262.646	6.288
FEMALE	1 HOUR OR LESS	437	412.30	13.845	.748	272.450	272.217	2.825
FEMALE	2 HOURS	733	648.94	21.790	.988	272.704	272.655	1.426
FEMALE	3 HOURS	725	686.92	23.066	1.025	267.090	267.211	1.256
FEMALE	4 HOURS	551	511.34	17.170	.899	262.568	263.006	1.347
FEMALE	5 HOURS	338	302.10	10.144	.575	259.123	259.017	1.805
FEMALE	6 HOURS OR MORE	417	388.01	13.029	.853	244.969	243.939	1.604

7. BIBLIOGRAPHIC REFERENCES

Allen, N. L. (1992) Data analysis for the science assessment. *The NAEP 1990 Technical Report.* (pp. 275-287) Report No. 21-TR-20. Princeton, NJ: Educational Testing Service.

Allen, N. L., & Zwick, R. (1992). Overview of Part II: The analysis of 1990 NAEP data. *The NAEP 1990 Technical Report.* (pp. 143-157) Report No. 21-TR-20. Princeton, NJ: Educational Testing Service.

Beaton, A. E. (1987). *The NAEP 1983-1984 Technical Report* . Report No. 15-TR-20. Princeton, NJ: Educational Testing Service.

Beaton, A. E. (1988). Expanding the new design:*The NAEP 1985-1986 Technical Report* . Report No. 17-TR-20. Princeton, NJ: Educational Testing Service.

Beaton, A. E., & Allen, N. L. (1992). Interpreting scales through scale anchoring. *Journal of Educational Statistics,* 17(2), 191-204.

Beaton, A. E. & Johnson, E. G. (1990). The average response method of scaling. *Journal of Educational Statistics,* 15 (1), 9-38.

Beaton, A. E., & Johnson, E. G. (1992). Overview of the scaling methodology used in the national assessment. *Journal of Educational Measurement,* 29(2),. 163-176.

Beaton, A. E., & Zwick, R. (1992). Overview of the National Assessment of Educational Progress. *Journal of Educational Statistics,* 17(2), 95-109.

Caldwell, N., Slobasky, R. , Moore, D. & Ter Maat, J. (1992). Field operations and data collection. In E. G. Johnson and N. L. Allen (Eds.) *The NAEP 1990 Technical Report.* (pp. 89-106) Report No. 21-TR-20. Princeton, NJ: Educational Testing Service.

Donoghue, J. R., (1992). Data analysis for the reading assessment. *The NAEP 1990 Technical Report.* (pp. 215-225) Report No. 21-TR-20. Princeton, NJ: Educational Testing Service.

Foertsch, M. A., Jones, L. R., & Koffler, S.L. (1992). Developing the NAEP objectives, items, and background questions for the 1990 assessments of reading, mathematics, and science. *The NAEP 1990 Technical Report.* (pp. 29-49) Report No. 21-TR-20. Princeton, NJ: Education Testing Service.

Foertsch, M. A., Gentile, C. A., Jenkins, L. B., Jones, C., Jones, L. E. & Whittington, B. (1992). Professional Series. In E. G. Johnson and N. L. Allen (Eds.) *The NAEP 1990 Technical Report.* (pp. 115-126) Report No. 21-TR-20. Princeton, NJ: Educational Testing Service.

Frankel, M. (1983) Sampling theory. In P.H. Rossi, J.D. Wright & A. Anderson (Eds.), *Handbook of survey research.* San Diego, CA: Academic Press.

Grima, A. W. & Johnson, E. G. (1992). Data analysis for the writing assessment. *The NAEP 1990 Technical Report.* (pp. 303-321) Report No. 21-TR-20. Princeton, NJ: Education Testing Service.

Johnson, E. G. (1989). Considerations and techniques for the analysis of NAEP data. *Journal of Educational Statistics*, 14 (4), 303-334.

Johnson, E. G. (1992). Overview of Part I: The design and implementation of the 1990 NAEP. *The NAEP 1990 Technical Report.* (pp. 9-28) Report No. 21-TR-20. Princeton, NJ: Educational Testing Service.

Johnson, E. G. (1992). The Design of the National Assessment of Educational Progress. *Journal of Educational Measurement*, 29(2), 95-110.

Johnson, E. G. & Allen, N. (1992). *The NAEP 1990 Technical Report.* (pp. 115-126) Report No. 21-TR-20. Princeton, NJ: Educational Testing Service.

Johnson, E. G., & Rust, K. F. (1992). Weighting procedures and estimation of sampling variance. *The NAEP 1990 Technical Report.* (pp. 159-193). Report No. 21-TR-20. Princeton, NJ: Educational Testing Service.

Johnson, E. G., & Rust, K. F. (1992). Population inferences and variance estimation for NAEP data. *Journal of Educational Statistics*, 17(2), 175-190.

Johnson, E. G. & Zwick, R. (1990). *Focusing the new design: The NAEP 1988 Technical Report.* Report No. 19-TR-20. Princeton, NJ: Educational Testing Service.

Johnson, E. G. (1987). Estimation of uncertainty due to sampling variability. In A. E. Beaton (Ed), *The NAEP 1983-1984 Technical Report* (pp. 505-512). Report No. 15-TR-20. Princeton, New Jersey: Educational Testing Service.

Kennel, T., Mohn, L., Reynolds, L., Smrdel, D., Thayer, B., & Zaback, L. (1992). Processing of materials and data. *The NAEP 1990 Technical Report.* (pp. 107-114) Report No. 21-TR-20. Princeton, NJ: Education Testing Service.

Kish, L. (1965). *Survey Sampling*. New York: Wiley.

Kish, L. & Frankel, M. R. (1974). Inference from complex samples. *Journals of the Royal Statistical Society*, Series B, 36, 1-22.

Kline, D. L. (1992). Assessment instruments. *The NAEP 1990 Technical Report.* (pp. 71-88) Report No. 21-TR-20. Princeton, NJ: Education Testing Service.

Koffler, S. L. (1991). The technical report of NAEP's 1990 Trial State Assessment program (No. 21-TS-01). Washington, DC: National Center for Education Statistics.

Messick, S. J., Beaton, A. E., & Lord, F. M. (1983). *NAEP Reconsidered: A new design for a new era.* (NAEP report 83-1) Princeton, NJ: Educational Testing Service.

Mislevy, R. J. (1987). The reading data analysis: Introduction. In A. E. Beaton (Ed), *The NAEP 1983-1984 Technical Report* (pp. 239-244). Report No. 15-TR-20. Princeton, New Jersey: Educational Testing Service.

Mislevy, R. J. (1992). Scaling procedures. *The NAEP 1990 Technical Report.* (pp. 199-213) Report No. 21-TR-20. Princeton, NJ: Education Testing Service.

Mislevy, R. J., Beaton, A. E., Kaplan, B., & Sheehan, K. M. (1992). Estimating population characteristics from sparse matrix samples of item responses. *Journal of Educational Measurement*, 29(2), 133-162.

Mislevy, R. J., Johnson, E. G., & Muraki, E. (1992). Scaling procedures in NAEP. *Journal of Educational Statistics*, 17(2), 131-154.

Mislevy, R. J. & Sheehan, K. M. (1989). Information matrices in latent-variable models, *Journal of Educational Statistics*, 14 (4), 335-350.

Mislevy, R. J. & Sheehan, K. M. (1987). Marginal estimation procedures. In A.E. Beaton (Ed), *The NAEP 1983-1984 Technical Report* (pp. 293-360). Report No. 15-TR-20. Princeton, New Jersey: Educational Testing Service.

Mullis, I. V. (1992). Developing the NAEP content area frameworks and innovative assessment methods in the 1992 assessments of mathematics, reading and writing. *Journal of Educational Measurement*, 29(2), 111-132.

Rogers, A. M., Freund, D. S., & Ferris, J. J. (1992). Database creation quality control of data entry and database products. *The NAEP 1990 Technical Report*. (pp. 127-133) Report No. 21-TR-20. Princeton, NJ: Educational Testing Service.

Rogers, A.M., Kline, D.L., Johnson, E.G., Mislevy, R.M., & Rust, K.F. (1990). *National Assessment of Educational Progress 1988 public-use data tapes version 2.0 user guide.* Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress.

Rogers, A.M., Kline, D.L., Johnson, E.G., Mislevy, R.M., & Rust, K.F. (1992). *National Assessment of Educational Progress 1990 public-use data tapes version 2.0 user guide.* Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress.

Rubin, D. B. (1987). *Multiple imputations for non response in surveys*. New York, New York: John Wiley & Sons.

Rubin, D. B. & Schneker, N. (1986). Multiple imputation for interval estimation from simple random samples with ignorable non response, *Journal of the American Statistical Association*, 81, 366-374.

Rust, K.F. (1992). Sample Design. *The NAEP 1990 Technical Report*. (pp. 51-68) Report No. 21-TR-20. Princeton, NJ: Educational Testing Service.

Rust, K. F., & Johnson, E. G. (1992). Sampling and weighting in the national assessment. *Journal of Educational Statistics*, 17(2),111-130.

Wingersky, M. S. (1983). LOGIST: A program for computing maximum likelihood procedures for logistic test models. In R. K. Hambleton (Ed.), *Applications of item response theory* (pp. 45-56). Vancouver, Canada: Educational Research Institute of British Columbia.

Wingersky, M. S. , Barton, M. A. , & Lord, F. M. (1982). LOGIST user's guide. Princeton, NJ: Educational Testing Service.

Yamamoto, K.& Jenkins, F. (1992). Data analysis for the mathematics assessment. *The NAEP 1990 Technical Report*. (pp. 243-256) Report No. 21-TR-20. Princeton, NJ: Educational Testing Service.

Yamamoto, K. and Mazzeo, J. (1992). Item response theory scale linking in NAEP. *Journal of Educational Statistics*, 17(2), 155-174.

Appendices

Appendix A

File Layout and Variable Information for the Mathematics 8th Grade Policy File (M08PS1.DAT and M08PS1.SPS)

```
VARIABLE     START   END    LEN  DEC  VARIABLE LABELS
-------------------------------------------------------------------------------------

YEAR            1      2     2    0   ASSESSMENT YEAR

AGE             3      4     2    0   ASSESSMENT AGE

BOOK            5      6     2    0   BOOKLET NUMBER                       (BOOK COVER)

SCH             7      9     3    0   SCHOOL CODE                          (BOOK COVER)

IEP            10            1    0   INDIVIDUALIZED EDUCATION PLAN        (BOOK COVER)
                                         VALUE  LABEL
                                            1   YES
                                            2   NO

LEP            11            1    0   LIMITED ENGLISH PROFICIENCY          (BOOK COVER)
                                         VALUE  LABEL
                                            1   YES
                                            2   NO

COHORT         12            1    0   AGE/GRADE COHORT GROUP
                                         VALUE  LABEL
                                            1   AGE 09
                                            2   AGE 13
                                            3   AGE 17

SCRID          13     18     6    0   SCRAMBLED STUDENT BOOKLET NUMBER

DGRADE         19     20     2    0   DERIVED GRADE                            (WESTAT)
                                         VALUE  LABEL
                                            0   NOT GRADED
                                            1   GRADE 1
                                            2   GRADE 2
                                            3   GRADE 3
                                            4   GRADE 4
                                            5   GRADE 5
                                            6   GRADE 6
                                            7   GRADE 7
                                            8   GRADE 8
                                            9   GRADE 9
                                           10   GRADE 10
                                           11   GRADE 11
                                           12   GRADE 12
                                           40   SPECIAL EDUCATION

DSEX           21            1    0   GENDER                                   (WESTAT)
                                         VALUE  LABEL
                                            1   MALE
                                            2   FEMALE

DRACE          22            1    0   DERIVED RACE/ETHNICITY                   (WESTAT)
                                         VALUE  LABEL
                                            1   WHITE
                                            2   BLACK
                                            3   HISPANIC
                                            4   ASIAN
                                            5   AMERICAN INDIAN
                                            6   UNCLASSIFIED

REGION         23            1    0   REGION OF COUNTRY
                                         VALUE  LABEL
                                            1   NORTHEAST
                                            2   SOUTHEAST
                                            3   CENTRAL
                                            4   WEST
                                            5   TERRITORY

STOC           24            1    0   SIZE AND TYPE OF COMMUNITY               (WESTAT)
                                         VALUE  LABEL
                                            1   EXTREME RURAL
                                            2   LOW METROPOLITAN
```

BEST COPY AVAILABLE 80

					3 HIGH METROPOLITAN	
					4 MAIN BIG CITY	
					5 URBAN FRINGE	
					6 MEDIUM CITY	
					7 SMALL PLACE	
SEASON	25		1	0	SEASON OF ASSESSMENT	(WESTAT)
					VALUE LABEL	
					1 WINTER	
					2 SPRING	
WEIGHT	26	32	7	5	OVERALL STUDENT SAMPLE WEIGHT	(WESTAT)
PARED	33		1	0	PARENTS' EDUCATION LEVEL	(ETS)
					VALUE LABEL	
					1 DIDN'T FINISH HIGHSC	
					2 GRAD FROM HIGHSCHOOL	
					3 SOME ED AFTER HIGHSC	
					4 GRAD FROM COLLEGE	
					5 UNKNOWN	
					7 I DON'T KNOW	
					8 OMITTED	
HOMEEN2	34		1	0	HOME ENVIRONMENT - READING MATERIALS (OF 4)	(ETS)
					VALUE LABEL	
					1 0 - 2 TYPES	
					2 3 TYPES	
					3 4 TYPES	
					8 OMITTED	
DAGE	35	36	2	0	ACTUAL AGE	(ETS)
SINGLEP	37		1	0	HOW MANY PARENTS LIVE AT HOME	(ETS)
					VALUE LABEL	
					1 2 PARENTS AT HOME	
					2 1 PARENT AT HOME	
					3 NEITHER PARENT HOME	
					8 OMITTED	
SCHTYPE	38		1	0	SCHOOL TYPE	(PQ)
					VALUE LABEL	
					1 PUBLIC SCHOOL	
					2 PRIVATE SCHOOL	
					3 CATHOLIC SCHOOL	
					4 BIA SCHOOL	
					5 DOD SCHOOL	
PERCMAT	39		1	0	STUDENTS' PERCEPTION OF MATHEMATICS	(ETS)
					VALUE LABEL	
					1 STRONGLY AGREE	
					2 AGREE	
					3 UNDEC,DISAGR,STRDSGR	
TCERTIF	40		1	0	TEACHERS' TYPE OF TEACHING CERTIFICATE	(ETS)
					VALUE LABEL	
					1 MATH	
					2 EDUCATION	
					3 ELSE	
TUNDMAJ	41		1	0	TEACHERS' UNDERGRADUATE MAJOR	(ETS)
					VALUE LABEL	
					1 MATH	
					2 EDUCATION	
					3 ELSE	
TGRDMAJ	42		1	0	TEACHERS' GRADUATE MAJOR	(ETS)
					VALUE LABEL	
					1 MATH	
					2 EDUCATION	
					3 ELSE	
TMATCRS	43		1	0	TEACHERS' NUMBER OF MATH AREAS TAKEN COURSES	(ETS)
					VALUE LABEL	
					1 0 - 3	
					2 4 - 5	
					3 6 - 7	
TEMPHNO	44		1	0	TEACHER EMPHASIS - NUMBERS AND OPERATIONS	(ETS)
					VALUE LABEL	
					1 HEAVY EMPHASIS	

					2 MODERATE EMPHASIS	
					3 LITTLE/NO EMPHASIS	
TEMPHPS	45		1	0	TEACHER EMPHASIS - PROBABILITY AND STAT	(ETS)
					VALUE LABEL	
					1 HEAVY EMPHASIS	
					2 MODERATE EMPHASIS	
					3 LITTLE/NO EMPHASIS	
SPOLICY	46		1	0	CHANGES IN SCHOOL POLICY SINCE 1984-85	(ETS)
					VALUE LABEL	
					1 0 - 2	
					2 3 - 4	
					3 5 - 8	
SPROBS	47		1	0	PROBLEMS IN THE SCHOOL	(ETS)
					VALUE LABEL	
					1 MODERATE TO SERIOUS	
					2 MINOR	
					3 NOT A PROBLEM	
IEP/LEP	48		1	0	INDIVIDUAL EDUC PLAN OR LIMITED ENGLISH PROF	(ETS)
					VALUE LABEL	
					1 YES	
					2 NO	
CALCUSE	49		1	0	STUDENT USED CALCULATOR APPROPRIATELY	(ETS)
					VALUE LABEL	
					1 HIGH	
					2 OTHER	
					8 OMITTED	
IDP	50		1	0	INSTRUCTION DOLLARS PER PUPIL	(QED)
					VALUE LABEL	
					0 UNCLASSIFIED	
					1 UNDER $14.99	
					2 $15 TO $24.99	
					3 $25 TO $34.99	
					4 $35 TO $44.99	
					5 $45 TO $54.99	
					6 $55 TO $64.99	
					7 $65 TO $74.99	
					8 $75 TO $149.99	
					9 $150 AND UP	
CAI	51		1	0	MICRO-COMPUTER ASSISTED INSTRUCTION	(QED)
					VALUE LABEL	
					0 UNCLASSIFIED	
					1 YES	
					2 NO	
MRPSCA1	52	56	5	2	PLAUSIBLE NAEP MATH VALUE #1 (NUM & OPER)	(ETS)
MRPSCA2	57	61	5	2	PLAUSIBLE NAEP MATH VALUE #2 (NUM & OPER)	(ETS)
MRPSCA3	62	66	5	2	PLAUSIBLE NAEP MATH VALUE #3 (NUM & OPER)	(ETS)
MRPSCA4	67	71	5	2	PLAUSIBLE NAEP MATH VALUE #4 (NUM & OPER)	(ETS)
MRPSCA5	72	76	5	2	PLAUSIBLE NAEP MATH VALUE #5 (NUM & OPER)	(ETS)
MRPSCB1	77	81	5	2	PLAUSIBLE NAEP MATH VALUE #1 (MEASUREMENT)	(ETS)
MRPSCB2	82	86	5	2	PLAUSIBLE NAEP MATH VALUE #2 (MEASUREMENT)	(ETS)
MRPSCB3	87	91	5	2	PLAUSIBLE NAEP MATH VALUE #3 (MEASUREMENT)	(ETS)
MRPSCB4	92	96	5	2	PLAUSIBLE NAEP MATH VALUE #4 (MEASUREMENT)	(ETS)
MRPSCB5	97	101	5	2	PLAUSIBLE NAEP MATH VALUE #5 (MEASUREMENT)	(ETS)
MRPSCC1	102	106	5	2	PLAUSIBLE NAEP MATH VALUE #1 (GEOMETRY)	(ETS)
MRPSCC2	107	111	5	2	PLAUSIBLE NAEP MATH VALUE #2 (GEOMETRY)	(ETS)
MRPSCC3	112	116	5	2	PLAUSIBLE NAEP MATH VALUE #3 (GEOMETRY)	(ETS)
MRPSCC4	117	121	5	2	PLAUSIBLE NAEP MATH VALUE #4 (GEOMETRY)	(ETS)
MRPSCC5	122	126	5	2	PLAUSIBLE NAEP MATH VALUE #5 (GEOMETRY)	(ETS)
MRPSCD1	127	131	5	2	PLAUSIBLE NAEP MATH VALUE #1 (DATA ANAL&STAT)	(ETS)
MRPSCD2	132	136	5	2	PLAUSIBLE NAEP MATH VALUE #2 (DATA ANAL&STAT)	(ETS)
MRPSCD3	137	141	5	2	PLAUSIBLE NAEP MATH VALUE #3 (DATA ANAL&STAT)	(ETS)
MRPSCD4	142	146	5	2	PLAUSIBLE NAEP MATH VALUE #4 (DATA ANAL&STAT)	(ETS)
MRPSCD5	147	151	5	2	PLAUSIBLE NAEP MATH VALUE #5 (DATA ANAL&STAT)	(ETS)
MRPSCE1	152	156	5	2	PLAUSIBLE NAEP MATH VALUE #1 (ALG & FUNCTNS)	(ETS)
MRPSCE2	157	161	5	2	PLAUSIBLE NAEP MATH VALUE #2 (ALG & FUNCTNS)	(ETS)
MRPSCE3	162	166	5	2	PLAUSIBLE NAEP MATH VALUE #3 (ALG & FUNCTNS)	(ETS)
MRPSCE4	167	171	5	2	PLAUSIBLE NAEP MATH VALUE #4 (ALG & FUNCTNS)	(ETS)
MRPSCE5	172	176	5	2	PLAUSIBLE NAEP MATH VALUE #5 (ALG & FUNCTNS)	(ETS)

```
MRPCMP1    177   181   5  2   PLAUSIBLE NAEP MATH VALUE #1 (COMPOSITE)     (ETS)
MRPCMP2    182   186   5  2   PLAUSIBLE NAEP MATH VALUE #2 (COMPOSITE)     (ETS)
MRPCMP3    187   191   5  2   PLAUSIBLE NAEP MATH VALUE #3 (COMPOSITE)     (ETS)
MRPCMP4    192   196   5  2   PLAUSIBLE NAEP MATH VALUE #4 (COMPOSITE)     (ETS)
MRPCMP5    197   201   5  2   PLAUSIBLE NAEP MATH VALUE #5 (COMPOSITE)     (ETS)

MTHLOG     202   206   5  2   LOGIST NAEP MATH THETA (SINGLE SCALE)        (ETS)
MRPLOG     207   211   5  2   LOGIST NAEP MATH VALUE (SINGLE SCALE)        (ETS)

B003001A   212         1  0   WHICH RACE/ETHNICITY BEST DESCRIBES YOU
                                 VALUE  LABEL
                                     1  WHITE
                                     2  BLACK
                                     3  HISPANIC
                                     4  ASIAN/PACIFIC AMERIC
                                     5  AMER IND/ALASKA NATV
                                     6  OTHER
                                     8  OMITTED
                                     0  MULTIPLE RESPONSE

B003101A   213         1  0   IF HISPANIC, WHAT IS YOUR HISPANIC BACKGROUND
                                 VALUE  LABEL
                                     1  NOT HISPANIC
                                     2  MEX,MEX AMER,CHICANO
                                     3  PUERTO RICAN
                                     4  CUBAN
                                     5  OTHER SPANISH/HISPAN
                                     8  OMITTED
                                     0  MULTIPLE RESPONSE

B003201A   214         1  0   HOW OFTEN OTHER THAN ENGLISH SPOKEN IN HOME
                                 VALUE  LABEL
                                     1  NEVER
                                     2  SOMETIMES
                                     3  ALWAYS
                                     8  OMITTED
                                     0  MULTIPLE RESPONSE

B003501A   215         1  0   MOTHER'S EDUCATION LEVEL
                                 VALUE  LABEL
                                     1  DIDN'T FINISH HIGHSC
                                     2  GRAD FROM HIGHSCHOOL
                                     3  SOME ED AFTER HIGHSC
                                     4  GRAD FROM COLLEGE
                                     7  I DON'T KNOW
                                     8  OMITTED
                                     0  MULTIPLE RESPONSE

B003601A   216         1  0   FATHER'S EDUCATION LEVEL
                                 VALUE  LABEL
                                     1  DIDN'T FINISH HIGHSC
                                     2  GRAD FROM HIGHSCHOOL
                                     3  SOME ED AFTER HIGHSC
                                     4  GRAD FROM COLLEGE
                                     7  I DON'T KNOW
                                     8  OMITTED
                                     0  MULTIPLE RESPONSE

B000901A   217         1  0   DOES YOUR FAMILY GET A NEWSPAPER REGULARLY
                                 VALUE  LABEL
                                     1  YES
                                     2  NO
                                     7  I DON'T KNOW
                                     8  OMITTED
                                     0  MULTIPLE RESPONSE

B000903A   218         1  0   IS THERE AN ENCYCLOPEDIA IN YOUR HOME
                                 VALUE  LABEL
                                     1  YES
                                     2  NO
                                     7  I DON'T KNOW
                                     8  OMITTED
                                     0  MULTIPLE RESPONSE

B000904A   219         1  0   ARE THERE MORE THAN 25 BOOKS IN YOUR HOME
                                 VALUE  LABEL
                                     1  YES
                                     2  NO
                                     7  I DON'T KNOW
```

					8	OMITTED
					0	MULTIPLE RESPONSE
B000905A	220	1	0	DOES YOUR FAMILY GET MAGAZINES REGULARLY		
					VALUE	LABEL
					1	YES
					2	NO
					7	I DON'T KNOW
					8	OMITTED
					0	MULTIPLE RESPONSE
B001801A	221	1	0	HOW MUCH TELEVISION DO YOU USUALLY WATCH EACH DAY		
					VALUE	LABEL
					1	NONE
					2	1 HOUR OR LESS
					3	2 HOURS
					4	3 HOURS
					5	4 HOURS
					6	5 HOURS
					7	6 HOURS OR MORE
					8	OMITTED
					0	MULTIPLE RESPONSE
B003901A	222	1	0	HOW MUCH TIME EACH DAY IS SPENT ON HOMEWORK		
					VALUE	LABEL
					1	DON'T HAVE HOMEWORK
					2	DON'T USUALLY DO IT
					3	1/2 HR OR LESS
					4	1 HOUR
					5	2 HOURS
					6	MORE THAN 2 HOURS
					8	OMITTED
					0	MULTIPLE RESPONSE
B006701A	223	1	0	HOW OFTEN DOES SOMEONE AT HOME HELP WITH HOMEWORK		
					VALUE	LABEL
					1	ALMOST EVERY DAY
					2	ONCE OR TWICE A WEEK
					3	ONCE OR TWICE MONTH
					4	NEVER OR HARDLY EVER
					5	DON'T HAVE HOMEWORK
					8	OMITTED
					0	MULTIPLE RESPONSE
B001101A	224	1	0	HOW MANY PAGES READ IN SCHOOL AND FOR HOMEWORK		
					VALUE	LABEL
					1	MORE THAN 20
					2	16-20
					3	11-15
					4	6-10
					5	5 OR FEWER
					8	OMITTED
					0	MULTIPLE RESPONSE
S004001A	225	1	0	HOW MANY DAYS OF SCHOOL MISSED LAST MONTH		
					VALUE	LABEL
					1	NONE
					2	1 OR 2 DAYS
					3	3 OR 4 DAYS
					4	5 TO 10 DAYS
					5	MORE THAN 10 DAYS
					8	OMITTED
					0	MULTIPLE RESPONSE
B007001A	226	1	0	DO YOU AGREE: RULES FOR BEHAVIOR ARE STRICT		
					VALUE	LABEL
					1	STRONGLY AGREE
					2	AGREE
					3	DISAGREE
					4	STRONGLY DISAGREE
					8	OMITTED
					0	MULTIPLE RESPONSE
B007002A	227	1	0	DO YOU AGREE: I DON'T FEEL SAFE AT SCHOOL		
					VALUE	LABEL
					1	STRONGLY AGREE
					2	AGREE
					3	DISAGREE
					4	STRONGLY DISAGREE
					8	OMITTED

				0	MULTIPLE RESPONSE
B007003A	228	1	0	DO YOU AGREE: STUDENTS OFTEN DISRUPT CLASS	
				VALUE	LABEL
				1	STRONGLY AGREE
				2	AGREE
				3	DISAGREE
				4	STRONGLY DISAGREE
				8	OMITTED
				0	MULTIPLE RESPONSE
S003401A	229	1	0	DO YOU EXPECT TO GRADUATE FROM HIGH SCHOOL	
				VALUE	LABEL
				1	YES
				2	NO
				7	I DON'T KNOW
				8	OMITTED
				0	MULTIPLE RESPONSE
B005601A	230	1	0	DOES MOTHER OR STEPMOTHER LIVE AT HOME WITH YOU	
				VALUE	LABEL
				1	YES
				2	NO
				8	OMITTED
				0	MULTIPLE RESPONSE
B005701A	231	1	0	DOES FATHER OR STEPFATHER LIVE AT HOME WITH YOU	
				VALUE	LABEL
				1	YES
				2	NO
				8	OMITTED
				0	MULTIPLE RESPONSE
B006001A	232	1	0	DOES MOTHER OR STEPMOTHER WORK AT JOB FOR PAY	
				VALUE	LABEL
				1	YES,FULL-TIME
				2	YES,PART-TIME
				3	NO
				4	DON'T LIVE W/EITHER
				8	OMITTED
				0	MULTIPLE RESPONSE
B006201A	233	1	0	DOES FATHER OR STEPFATHER WORK AT JOB FOR PAY	
				VALUE	LABEL
				1	YES,FULL-TIME
				2	YES,PART-TIME
				3	NO
				4	DON'T LIVE W/EITHER
				8	OMITTED
				0	MULTIPLE RESPONSE
M810101B	234	1	0	IN MATH CLASS HOW OFTEN DO PROBLEMS FROM TEXTBOOKS	
				VALUE	LABEL
				1	ALMOST EVERY DAY
				2	SEVERAL TIMES A WEEK
				3	ABOUT ONCE A WEEK
				4	LESS THAN ONCE WEEK
				5	NEVER
				8	OMITTED
				0	MULTIPLE RESPONSE
M810102B	235	1	0	IN MATH CLASS HOW OFTEN DO PROBLEMS ON WORKSHEETS	
				VALUE	LABEL
				1	ALMOST EVERY DAY
				2	SEVERAL TIMES A WEEK
				3	ABOUT ONCE A WEEK
				4	LESS THAN ONCE WEEK
				5	NEVER
				8	OMITTED
				0	MULTIPLE RESPONSE
M810103B	236	1	0	IN MATH CLASS HOW OFTEN WORK IN SMALL GROUPS	
				VALUE	LABEL
				1	ALMOST EVERY DAY
				2	SEVERAL TIMES A WEEK
				3	ABOUT ONCE A WEEK
				4	LESS THAN ONCE WEEK
				5	NEVER
				8	OMITTED
				0	MULTIPLE RESPONSE

```
M810104B    237          1   0   IN MATH CLASS HOW OFTEN USE RULERS, BLOCKS, SOLIDS
                                    VALUE  LABEL
                                       1   ALMOST EVERY DAY
                                       2   SEVERAL TIMES A WEEK
                                       3   ABOUT ONCE A WEEK
                                       4   LESS THAN ONCE WEEK
                                       5   NEVER
                                       8   OMITTED
                                       0   MULTIPLE RESPONSE

M810105B    238          1   0   IN MATH CLASS HOW OFTEN DO YOU USE A CALCULATOR
                                    VALUE  LABEL
                                       1   ALMOST EVERY DAY
                                       2   SEVERAL TIMES A WEEK
                                       3   ABOUT ONCE A WEEK
                                       4   LESS THAN ONCE WEEK
                                       5   NEVER
                                       8   OMITTED
                                       0   MULTIPLE RESPONSE

M810106B    239          1   0   IN MATH CLASS HOW OFTEN DO YOU USE A COMPUTER
                                    VALUE  LABEL
                                       1   ALMOST EVERY DAY
                                       2   SEVERAL TIMES A WEEK
                                       3   ABOUT ONCE A WEEK
                                       4   LESS THAN ONCE WEEK
                                       5   NEVER
                                       8   OMITTED
                                       0   MULTIPLE RESPONSE

M810107B    240          1   0   IN MATH CLASS HOW OFTEN DO YOU TAKE MATH TESTS
                                    VALUE  LABEL
                                       1   ALMOST EVERY DAY
                                       2   SEVERAL TIMES A WEEK
                                       3   ABOUT ONCE A WEEK
                                       4   LESS THAN ONCE WEEK
                                       5   NEVER
                                       8   OMITTED
                                       0   MULTIPLE RESPONSE

M810108B    241          1   0   IN MATH CLASS HOW OFTEN WRITE REPORT OR DO PROJECT
                                    VALUE  LABEL
                                       1   ALMOST EVERY DAY
                                       2   SEVERAL TIMES A WEEK
                                       3   ABOUT ONCE A WEEK
                                       4   LESS THAN ONCE WEEK
                                       5   NEVER
                                       8   OMITTED
                                       0   MULTIPLE RESPONSE

M810201B    242          1   0   TEACHER EXPLAINS CALCULATOR USE TO SOLVE PROBLEMS
                                    VALUE  LABEL
                                       1   YES
                                       2   NO
                                       8   OMITTED
                                       0   MULTIPLE RESPONSE

M810301B    243          1   0   HOW OFTEN USE CALCULATOR IN MATH CLASS
                                    VALUE  LABEL
                                       1   ALMOST ALWAYS
                                       2   SOMETIMES
                                       3   NEVER
                                       8   OMITTED
                                       0   MULTIPLE RESPONSE

M810302B    244          1   0   HOW OFTEN USE CALCULATOR TO DO PROBLEMS AT HOME
                                    VALUE  LABEL
                                       1   ALMOST ALWAYS
                                       2   SOMETIMES
                                       3   NEVER
                                       8   OMITTED
                                       0   MULTIPLE RESPONSE

M810303B    245          1   0   HOW OFTEN USE CALCULATOR TO TAKE QUIZ OR TEST
                                    VALUE  LABEL
                                       1   ALMOST ALWAYS
                                       2   SOMETIMES
                                       3   NEVER
                                       8   OMITTED
```

```
                                    0   MULTIPLE RESPONSE

S208501B    246          1   0   DOES FAMILY OWN A CALCULATOR
                                   VALUE  LABEL
                                      1   YES
                                      2   NO
                                      8   OMITTED
                                      0   MULTIPLE RESPONSE

M810401B    247          1   0   HAVE YOU EVER USED A SCIENTIFIC CALCULATOR
                                   VALUE  LABEL
                                      1   YES
                                      2   NO
                                      8   OMITTED
                                      0   MULTIPLE RESPONSE

M810501B    248          1   0   WHAT KIND OF MATH CLASS ARE YOU TAKING THIS YEAR
                                   VALUE  LABEL
                                      1   NO MATH THIS YEAR
                                      2   EIGHTH-GRADE MATH
                                      3   PRE-ALGEBRA
                                      4   ALGEBRA
                                      5   OTHER
                                      8   OMITTED
                                      0   MULTIPLE RESPONSE

M810601B    249          1   0   HOW MUCH TIME SPENT EACH DAY ON MATH HOMEWORK
                                   VALUE  LABEL
                                      1   NONE
                                      2   15 MINUTES
                                      3   30 MINUTES
                                      4   45 MINUTES
                                      5   AN HOUR
                                      6   MORE THAN AN HOUR
                                      7   NOT TAKING MATH NOW
                                      8   OMITTED
                                      0   MULTIPLE RESPONSE

M810701B    250          1   0   DO YOU AGREE: I LIKE MATH
                                   VALUE  LABEL
                                      1   STRONGLY AGREE
                                      2   AGREE
                                      3   UNDECIDED
                                      4   DISAGREE
                                      5   STRONGLY DISAGREE
                                      8   OMITTED
                                      0   MULTIPLE RESPONSE

M810702B    251          1   0   DO YOU AGREE: ALL PEOPLE USE MATH IN THEIR JOBS
                                   VALUE  LABEL
                                      1   STRONGLY AGREE
                                      2   AGREE
                                      3   UNDECIDED
                                      4   DISAGREE
                                      5   STRONGLY DISAGREE
                                      8   OMITTED
                                      0   MULTIPLE RESPONSE

M810703B    252          1   0   DO YOU AGREE: I AM GOOD IN MATH
                                   VALUE  LABEL
                                      1   STRONGLY AGREE
                                      2   AGREE
                                      3   UNDECIDED
                                      4   DISAGREE
                                      5   STRONGLY DISAGREE
                                      8   OMITTED
                                      0   MULTIPLE RESPONSE

M810704B    253          1   0   DO YOU AGREE: MATH IS MORE FOR BOYS THAN FOR GIRLS
                                   VALUE  LABEL
                                      1   STRONGLY AGREE
                                      2   AGREE
                                      3   UNDECIDED
                                      4   DISAGREE
                                      5   STRONGLY DISAGREE
                                      8   OMITTED
                                      0   MULTIPLE RESPONSE

M810705B    254          1   0   DO YOU AGREE: MATH USEFUL/SOLVING EVERYDAY PROBLEM
                                   VALUE  LABEL
```

```
                                             1   STRONGLY AGREE
                                             2   AGREE
                                             3   UNDECIDED
                                             4   DISAGREE
                                             5   STRONGLY DISAGREE
                                             8   OMITTED
                                             0   MULTIPLE RESPONSE

M810801B    255              1   0   HOW MANY GRADES YOU ATTENDED SCHOOL IN THIS STATE
                                        VALUE  LABEL
                                             1   LESS THAN ONE GRADE
                                             2   1 - 2 GRADES
                                             3   3 - 5 GRADES
                                             4   MORE THAN 5 GRADES
                                             8   OMITTED
                                             0   MULTIPLE RESPONSE

T006001     256              1   0   WHAT IS YOUR GENDER
                                        VALUE  LABEL
                                             1   MALE
                                             2   FEMALE
                                             8   OMITTED
                                             0   MULTIPLE RESPONSE

T022801     257              1   0   WHICH BEST DESCRIBES YOU
                                        VALUE  LABEL
                                             1   AMER IND/ALASKA NATV
                                             2   ASIAN/PACIFIC AMERIC
                                             3   HISPANIC (ANY RACE)
                                             4   BLACK (NOT HISPANIC)
                                             5   WHITE (NOT HISPANIC)
                                             8   OMITTED
                                             0   MULTIPLE RESPONSE

T030001     258     259      2   0   HOW MANY YEARS TEACHING ELEM OR SECONDARY LEVEL

T030101     260     261      2   0   HOW MANY YEARS HAVE YOU TAUGHT MATHEMATICS

T030201     262              1   0   WHAT TYPE OF TEACHING CERTIFICATION DO YOU HAVE
                                        VALUE  LABEL
                                             1   NONE
                                             2   TEMP,PROB,PROV,EMERG
                                             3   REG CERT < HIGHEST
                                             4   HIGHEST CERT AVAIL
                                             8   OMITTED
                                             0   MULTIPLE RESPONSE

T030302     263              1   0   DO YOU HAVE STATE CERTIF FOR MID/JR HS EDUC (GEN)
                                        VALUE  LABEL
                                             1   YES
                                             2   NO
                                             3   NOT OFFERED IN STATE
                                             8   OMITTED
                                             0   MULTIPLE RESPONSE

T030303     264              1   0   DO YOU HAVE STATE CERTIF FOR MID/JUNIOR HS MATH
                                        VALUE  LABEL
                                             1   YES
                                             2   NO
                                             3   NOT OFFERED IN STATE
                                             8   OMITTED
                                             0   MULTIPLE RESPONSE

T023201     265              1   0   WHAT IS THE HIGHEST ACADEMIC DEGREE YOU HOLD
                                        VALUE  LABEL
                                             1   HIGH SCHOOL DIPLOMA
                                             2   ASSOC DEG/VOC CERT
                                             3   BACHELOR'S DEGREE
                                             4   MASTER'S DEGREE
                                             5   ED SPEC/PROF DIPLOMA
                                             6   DOCTORATE
                                             7   PROFESSIONAL DEGREE
                                             8   OMITTED
                                             0   MULTIPLE RESPONSE

T023301     266              1   0   UNDERGRADUATE MAJOR: EDUCATION
                                        VALUE  LABEL
                                             0   NO
                                             1   YES
                                             8   OMITTED
```

BEST COPY AVAILABLE

```
T023302    267    1   0   UNDERGRADUATE MINOR: EDUCATION
                            VALUE  LABEL
                              0    NO
                              1    YES
                              8    OMITTED

T023311    268    1   0   UNDERGRADUATE MAJOR: MATHEMATICS
                            VALUE  LABEL
                              0    NO
                              1    YES
                              8    OMITTED

T030401    269    1   0   COURSES TAKEN IN TEACHING ELEMENTARY MATH
                            VALUE  LABEL
                              1    NONE
                              2    1
                              3    2
                              4    3 OR MORE
                              8    OMITTED
                              0    MULTIPLE RESPONSE

T030402    270    1   0   COURSES TAKEN IN TEACHING MIDDLE SCHOOL MATH
                            VALUE  LABEL
                              1    NONE
                              2    1
                              3    2
                              4    3 OR MORE
                              8    OMITTED
                              0    MULTIPLE RESPONSE

T030403    271    1   0   COURSES TAKEN IN TEACHING ELEM/MID SCH GEOMETRY
                            VALUE  LABEL
                              1    NONE
                              2    1
                              3    2
                              4    3 OR MORE
                              8    OMITTED
                              0    MULTIPLE RESPONSE

T030404    272    1   0   COURSES TAKEN IN REMEDIAL/DEVELOPMENT MATH INSTRUC
                            VALUE  LABEL
                              1    NONE
                              2    1
                              3    2
                              4    3 OR MORE
                              8    OMITTED
                              0    MULTIPLE RESPONSE

T030405    273    1   0   COURSES TAKEN IN CALCULATOR/COMPUTER MATH INSTRUC
                            VALUE  LABEL
                              1    NONE
                              2    1
                              3    2
                              4    3 OR MORE
                              8    OMITTED
                              0    MULTIPLE RESPONSE

T030412    274    1   0   COURSES TAKEN IN APPLIED MATHEMATICS
                            VALUE  LABEL
                              1    NONE
                              2    1
                              3    2
                              4    3 OR MORE
                              8    OMITTED
                              0    MULTIPLE RESPONSE

T030413    275    1   0   COURSES TAKEN IN COMPUTER SCIENCE (GENERAL)
                            VALUE  LABEL
                              1    NONE
                              2    1
                              3    2
                              4    3 OR MORE
                              8    OMITTED
                              0    MULTIPLE RESPONSE

T030414    276    1   0   COURSES TAKEN IN COMPUTER PROGRAMMING
                            VALUE  LABEL
                              1    NONE
                              2    1
```

```
                              3    2
                              4    3 OR MORE
                              8    OMITTED
                              0    MULTIPLE RESPONSE
T030501      277        1   0   TIME SPENT ON IN-SERVICE EDUC IN MATH (LAST YEAR)
                           VALUE   LABEL
                              1    NONE
                              2    LESS THAN 6 HOURS
                              3    6-15 HOURS
                              4    16-35 HOURS
                              5    MORE THAN 35 HOURS
                              8    OMITTED
                              0    MULTIPLE RESPONSE
T030601      278        1   0   TRAINED TO TEACH STUDENTS WITH LIMITED ENG PROFIC
                           VALUE   LABEL
                              1    YES
                              2    NO
                              8    OMITTED
                              0    MULTIPLE RESPONSE
T030602      279        1   0   TRAINED TO TEACH STUDENTS FROM DIFFERENT CULTURES
                           VALUE   LABEL
                              1    YES
                              2    NO
                              8    OMITTED
                              0    MULTIPLE RESPONSE
T030603      280        1   0   TRAINED TO TEACH STUDENTS WITH DIF COGNITIVE STYLE
                           VALUE   LABEL
                              1    YES
                              2    NO
                              8    OMITTED
                              0    MULTIPLE RESPONSE
T030701      281        1   0   I HAVE GREAT FREEDOM IN DECISIONS ON MATH INSTRUCT
                           VALUE   LABEL
                              1    STRONGLY AGREE
                              2    AGREE
                              3    UNDECIDED
                              4    DISAGREE
                              5    STRONGLY DISAGREE
                              8    OMITTED
                              0    MULTIPLE RESPONSE
T030702      282        1   0   MY MATH CLASSES ARE FREQUENTLY INTERRUPTED
                           VALUE   LABEL
                              1    STRONGLY AGREE
                              2    AGREE
                              3    UNDECIDED
                              4    DISAGREE
                              5    STRONGLY DISAGREE
                              8    OMITTED
                              0    MULTIPLE RESPONSE
T030801      283        1   0   HOW WELL SUPPLIED BY SCHOOL WITH MATERIAL/RESOURCE
                           VALUE   LABEL
                              1    I GET ALL NEEDED
                              2    I GET MOST NEEDED
                              3    I GET SOME NEEDED
                              4    I GET NONE
                              8    OMITTED
                              0    MULTIPLE RESPONSE
T030901      284        1   0   ARE STUDENTS ASSIGNED TO THIS CLASS BY ABILITY
                           VALUE   LABEL
                              1    YES
                              2    NO
                              8    OMITTED
                              0    MULTIPLE RESPONSE
T031001      285        1   0   WHICH BEST DESCRIBES ABILITY OF STUDENTS IN CLASS
                           VALUE   LABEL
                              1    PRIMARILY HIGH
                              2    PRIMARILY AVERAGE
                              3    PRIMARILY LOW
                              4    WIDELY MIXED
                              8    OMITTED
                              0    MULTIPLE RESPONSE
```

```
T031101     286          1   0   TIME SPENT ON MATH INSTRUCTION PER WEEK (HOURS)

T031102     287    288   2   0   TIME SPENT ON MATH INSTRUCTION (MINUTES)

T031201     289          1   0   TIME STUDENTS SPEND ON MATH HOMEWORK EACH DAY
                                    VALUE  LABEL
                                        1  NONE
                                        2  15 MINUTES
                                        3  30 MINUTES
                                        4  45 MINUTES
                                        5  AN HOUR
                                        6  MORE THAN AN HOUR
                                        8  OMITTED
                                        0  MULTIPLE RESPONSE

T031401     290          1   0   HOW OFTEN STUDENTS DO MATH PROBLEMS FROM TEXTBOOK
                                    VALUE  LABEL
                                        1  ALMOST EVERY DAY
                                        2  SEVERAL TIMES A WEEK
                                        3  ABOUT ONCE A WEEK
                                        4  LESS THAN ONCE WEEK
                                        5  NEVER
                                        8  OMITTED
                                        0  MULTIPLE RESPONSE

T031402     291          1   0   HOW OFTEN STUDENTS DO MATH PROBLEMS ON WORKSHEETS
                                    VALUE  LABEL
                                        1  ALMOST EVERY DAY
                                        2  SEVERAL TIMES A WEEK
                                        3  ABOUT ONCE A WEEK
                                        4  LESS THAN ONCE WEEK
                                        5  NEVER
                                        8  OMITTED
                                        0  MULTIPLE RESPONSE

T031403     292          1   0   HOW OFTEN DO STUDENTS WORK IN SMALL GROUPS
                                    VALUE  LABEL
                                        1  ALMOST EVERY DAY
                                        2  SEVERAL TIMES A WEEK
                                        3  ABOUT ONCE A WEEK
                                        4  LESS THAN ONCE WEEK
                                        5  NEVER
                                        8  OMITTED
                                        0  MULTIPLE RESPONSE

T031406     293          1   0   HOW OFTEN DO STUDENTS USE COMPUTERS
                                    VALUE  LABEL
                                        1  ALMOST EVERY DAY
                                        2  SEVERAL TIMES A WEEK
                                        3  ABOUT ONCE A WEEK
                                        4  LESS THAN ONCE WEEK
                                        5  NEVER
                                        8  OMITTED
                                        0  MULTIPLE RESPONSE

T031408     294          1   0   HOW OFTEN TAKE TEACHER-GENERATED MATH TESTS
                                    VALUE  LABEL
                                        1  ALMOST EVERY DAY
                                        2  SEVERAL TIMES A WEEK
                                        3  ABOUT ONCE A WEEK
                                        4  LESS THAN ONCE WEEK
                                        5  NEVER
                                        8  OMITTED
                                        0  MULTIPLE RESPONSE

T031409     295          1   0   HOW OFTEN TAKE OTHER PUBLISHED TESTS
                                    VALUE  LABEL
                                        1  ALMOST EVERY DAY
                                        2  SEVERAL TIMES A WEEK
                                        3  ABOUT ONCE A WEEK
                                        4  LESS THAN ONCE WEEK
                                        5  NEVER
                                        8  OMITTED
                                        0  MULTIPLE RESPONSE

T031901     296          1   0   WHAT IS THE AVAILABILITY OF COMPUTERS FOR STUDENTS
                                    VALUE  LABEL
                                        1  NOT AVAILABLE
                                        2  DIFFICULT TO ACCESS
```

```
                                    3   AVAILABLE IN CLASS
                                    8   OMITTED
                                    0   MULTIPLE RESPONSE
T032001   297      1   0   DAYS PER WEEK COMPUTER USED FOR MATH CONCEPTS
                              VALUE LABEL
                                    1   NONE
                                    2   1
                                    3   2
                                    4   3
                                    5   4
                                    6   5
                                    8   OMITTED
                                    0   MULTIPLE RESPONSE
T032101   298      1   0   MINUTES PER WEEK STUDENT SPENDS USING COMPUTERS
                              VALUE LABEL
                                    1   NONE
                                    2   15 MINUTES
                                    3   30 MINUTES
                                    4   45 MINUTES
                                    5   AN HOUR
                                    6   MORE THAN AN HOUR
                                    8   OMITTED
                                    0   MULTIPLE RESPONSE
```

Appendix B

File Layout and Variable Information for the Mathematics 8th Grade Measurement File (M08MS1.DAT and M08MS1.SPS)

VARIABLE	START	END	LEN	DEC	VARIABLE LABELS	
YEAR	1	2	2		ASSESSMENT YEAR	
AGE	3	4	2		ASSESSMENT AGE	
BOOK	5	6	2		BOOKLET NUMBER	(BOOK COVER)
SCRID	7	12	6		SCRAMBLED STUDENT BOOKLET NUMBER	
NUMCOR	13	14	2		NUMBER OF ITEMS CORRECT IN BOOKLET	
PCTCOR	15	17	3		PERCENT CORRECT IN BOOKLET	
LOGITP	18	23	6	4	LOGIT PERCENT CORRECT IN BOOKLET	
ZSCORE	24	29	6	4	STANDARDIZED LOGIT PERCENT CORRECT IN BOOKLET	
DGRADE	30	31	2		DERIVED GRADE VALUE LEVEL 0 NOT GRADED 1 GRADE 1 2 GRADE 2 3 GRADE 3 4 GRADE 4 5 GRADE 5 6 GRADE 6 7 GRADE 7 8 GRADE 8 9 GRADE 9 10 GRADE 10 11 GRADE 11 12 GRADE 12 40 SPECIAL EDUCATION	(WESTAT)
DSEX	32		1		GENDER VALUE LEVEL 1 MALE 2 FEMALE	(WESTAT)
DRACE	33		1		DERIVED RACE/ETHNICITY VALUE LEVEL 1 WHITE 2 BLACK 3 HISPANIC 4 ASIAN 5 AMERICAN INDIAN 6 UNCLASSIFIED	(WESTAT)
REGION	34		1		REGION OF COUNTRY VALUE LEVEL 1 NORTHEAST 2 SOUTHEAST 3 CENTRAL 4 WEST 5 TERRITORY	
WEIGHT	35	41	7	5	OVERALL STUDENT SAMPLE WEIGHT	(WESTAT)
PARED	42		1		PARENTS' EDUCATION LEVEL VALUE LEVEL 1 DIDN'T FINISH HIGHSC 2 GRAD FROM HIGHSCHOOL 3 SOME ED AFTER HIGHSC 4 GRAD FROM COLLEGE 5 UNKNOWN 7 I DON'T KNOW 8 OMITTED	(ETS)
DAGE	43	44	2		ACTUAL AGE	(ETS)

MRPSCA1	45	49	5	2	PLAUSIBLE NAEP MATH VALUE #1 (NUM & OPER)	(ETS)
MRPSCA2	50	54	5	2	PLAUSIBLE NAEP MATH VALUE #2 (NUM & OPER)	(ETS)
MRPSCA3	55	59	5	2	PLAUSIBLE NAEP MATH VALUE #3 (NUM & OPER)	(ETS)
MRPSCA4	60	64	5	2	PLAUSIBLE NAEP MATH VALUE #4 (NUM & OPER)	(ETS)
MRPSCA5	65	69	5	2	PLAUSIBLE NAEP MATH VALUE #5 (NUM & OPER)	(ETS)
MRPSCB1	70	74	5	2	PLAUSIBLE NAEP MATH VALUE #1 (MEASUREMENT)	(ETS)
MRPSCB2	75	79	5	2	PLAUSIBLE NAEP MATH VALUE #2 (MEASUREMENT)	(ETS)
MRPSCB3	80	84	5	2	PLAUSIBLE NAEP MATH VALUE #3 (MEASUREMENT)	(ETS)
MRPSCB4	85	89	5	2	PLAUSIBLE NAEP MATH VALUE #4 (MEASUREMENT)	(ETS)
MRPSCB5	90	94	5	2	PLAUSIBLE NAEP MATH VALUE #5 (MEASUREMENT)	(ETS)
MRPSCC1	95	99	5	2	PLAUSIBLE NAEP MATH VALUE #1 (GEOMETRY)	(ETS)
MRPSCC2	100	104	5	2	PLAUSIBLE NAEP MATH VALUE #2 (GEOMETRY)	(ETS)
MRPSCC3	105	109	5	2	PLAUSIBLE NAEP MATH VALUE #3 (GEOMETRY)	(ETS)
MRPSCC4	110	114	5	2	PLAUSIBLE NAEP MATH VALUE #4 (GEOMETRY)	(ETS)
MRPSCC5	115	119	5	2	PLAUSIBLE NAEP MATH VALUE #5 (GEOMETRY)	(ETS)
MRPSCD1	120	124	5	2	PLAUSIBLE NAEP MATH VALUE #1 (DATA ANAL&STAT)	(ETS)
MRPSCD2	125	129	5	2	PLAUSIBLE NAEP MATH VALUE #2 (DATA ANAL&STAT)	(ETS)
MRPSCD3	130	134	5	2	PLAUSIBLE NAEP MATH VALUE #3 (DATA ANAL&STAT)	(ETS)
MRPSCD4	135	139	5	2	PLAUSIBLE NAEP MATH VALUE #4 (DATA ANAL&STAT)	(ETS)
MRPSCD5	140	144	5	2	PLAUSIBLE NAEP MATH VALUE #5 (DATA ANAL&STAT)	(ETS)
MRPSCE1	145	149	5	2	PLAUSIBLE NAEP MATH VALUE #1 (ALG & FUNCTNS)	(ETS)
MRPSCE2	150	154	5	2	PLAUSIBLE NAEP MATH VALUE #2 (ALG & FUNCTNS)	(ETS)
MRPSCE3	155	159	5	2	PLAUSIBLE NAEP MATH VALUE #3 (ALG & FUNCTNS)	(ETS)
MRPSCE4	160	164	5	2	PLAUSIBLE NAEP MATH VALUE #4 (ALG & FUNCTNS)	(ETS)
MRPSCE5	165	169	5	2	PLAUSIBLE NAEP MATH VALUE #5 (ALG & FUNCTNS)	(ETS)
MRPCMP1	170	174	5	2	PLAUSIBLE NAEP MATH VALUE #1 (COMPOSITE)	(ETS)
MRPCMP2	175	179	5	2	PLAUSIBLE NAEP MATH VALUE #2 (COMPOSITE)	(ETS)
MRPCMP3	180	184	5	2	PLAUSIBLE NAEP MATH VALUE #3 (COMPOSITE)	(ETS)
MRPCMP4	185	189	5	2	PLAUSIBLE NAEP MATH VALUE #4 (COMPOSITE)	(ETS)
MRPCMP5	190	194	5	2	PLAUSIBLE NAEP MATH VALUE #5 (COMPOSITE)	(ETS)

NUMBERS AND OPERATION SCALE

N276803C	195	1	59 + 46 + 82 + 68 = 255	(NO CALCULATOR)	(RATER 1)
N277602C	196	1	604 - 207 = 397	(NO CALCULATOR)	(RATER 1)
N286201C	197	1	24 DIVIDED BY 6 SHOWS HOW TO PACK BASEBALLS		
N274801C	198	1	.35 CHANGED TO A PERCENT IS 35%		
N258801C	199	1	125% OF 10 IS GREATER THAN 10		
N286602C	200	1	WRITE 3 3/10 AS 3.3		(RATER 1)
N275301C	201	1	OF NUMBERS GIVEN, 5 IS COMMON FACTOR OF 10 AND 15		
N260101C	202	1	COMPUTE +6, -12 =-6		
N286301C	203	1	.075 IS BETWEEN .07 AND .08		
M017401D	204	1	ADD WHOLE NUMBERS		
M017701D	205	1	IDENTIFY SOLUTION PROCEDURE		
M017901D	206	1	SOLVE MULTI-STEP STORY PROBLEM		
M018201D	207	1	SOLVE MULTI-STEP STORY PROBLEM		
M018401D	208	1	SOLVE STORY PROBLEM (DIVISION)		
M018501D	209	1	SOLVE STORY PROBLEM (FRACTIONS)		
M018601D	210	1	READ A SCALE DIAGRAM		
M020001E	211	1	APPLY PLACE VALUE		(RATER 1)
M020101E	212	1	APPLY PART-WHOLE RELATIONSHIP		(RATER 1)
M020501E	213	1	USE A NUMBER LINE GRAPH		(RATER 1)
M021901F	214	1	SOLVE STORY PROBLEM (MONEY)		
M022001F	215	1	ESTIMATE DISTANCE ON MAP		
M022301F	216	1	SOLVE STORY PROBLEM (REASONING)		
M022701F	217	1	UNDERSTAND WHEN TO ESTIMATE		
M022901F	218	1	APPLY PLACE VALUE		
M023001F	219	1	SOLVE STORY PROBLEM (REMAINDER)		
M023801F	220	1	ESTIMATE DECIMAL/FRACTION		
M015501G	221	1	IF 2/25 = N/500 THEN N = 40		
M015901G	222	1	FIGURE A BEST ILLUSTRATES THE STATEMENT		
M016501G	223	1	120 IS LEAST COMMON MULTIPLE OF 8, 12 AND 15		
M012431H	224	1	FIND CHECKBOOK BALANCE		
M012531H	225	1	SOLVE TWO-STEP STORY PROBLEM		
M012931H	226	1	INTERPRET A GIVEN RULE		
N202831H	227	1	INTERPRET REPRESENTATION OF FRACTION		
M011131H	228	1	SOLVE STORY PROBLEM (MULTIPLICATION)		
M013431H	229	1	APPLY DIVISION		
M013531H	230	1	USE SCIENTIFIC NOTATION		
M013631H	231	1	ORDER FRACTIONS		
M027031I	232	1	(150 / 3) + (6 X 2) = 62		
M027331I	233	1	PRODUCT OF 3.12 AND 8 CUBED = 1597.44		(RATER 1)
M027831I	234	1	OBJECT 30 LBS-EARTH WEIGHS 5 LBS ON MOON		(RATER 1)
M028031I	235	1	($14.95 + $5.85 + $9.70) X .06 = $32.33		

M028131I	236		1	12 DIVIDES N W/O REMAINDER, ALSO 2,3,4,6	(RATER 1)
M028231I	237		1	BEEF = $2.59 /LB - 0.93 LBS COST $2.41	
M028631I	238		1	MEAT COST: (214,964/52)X2.53 = $10458.83	(RATER 1)
M028731I	239		1	50 CENTS TO 60 CENTS - PERCENT INCREASE IS 20	
M028931I	240		1	IF 10.3/5.62 = N/4.78 THEN 8.76 IS CLOSEST TO N	

MEASUREMENT SCALE

N267201C	241		1	PENCIL LENGTH SHOWN IS 3 3/4 TO NEAREST 4TH INCH	
N265201C	242		1	USE CENTIMETER NOT M OR KM FOR PENCIL LENGTH	
N265901C	243		1	ONE LITER IS 1000 MILILITERS	
N252101C	244		1	PERIMETER OF RECTANGLE 8M X 5M IS 26 METERS	
M017501D	245		1	COMPARE WEIGHTS	
M018101D	246		1	APPLY CONCEPT OF PERIMETER	
M019101D	247		1	INTERPRET MEASUREMENT TOLERANCE	
M019201D	248		1	FIND TOTAL SURFACE AREA	
M020301E	249		1	READ A RULER	(RATER 1)
M022601F	250		1	COMPARE WEIGHTS	
M022801F	251		1	USE A RULER	(RATER 1)
M022802F	252		1	USE A RULER	(RATER 1)
M023401F	253		1	FIND AREA OF A RECTANGLE	
M023701F	254		1	USE A PROTRACTOR	(RATER 1)
M015401G	255		1	150 MINUTES = 2 1/2 HOURS	
M015701G	256		1	LIQUID LET OUT OF THE TUBE: 15 MILLILITERS	
M016201G	257		1	BOX 48 CUBIC INCHES-MEASUREMENT REPRESENTS VOLUME	
M012331H	258		1	APPLY MULTIPLICATION	
M013331H	259		1	IDENTIFY MEASUREMENT INSTRUMENT	
M027631I	260		1	MODEL: IF 15 FT = 3 INCHES, THEN 35 FT = 7 INCHES	

GEOMETRY SCALE

N253701C	261		1	2ND SET OF LINE SEGMENTS CANNOT MAKE A TRIANGLE	
N269901C	262		1	THE FOURTH FIGURE SHOWN IS NOT A PARALLELOGRAM	
N254602C	263		1	SECOND LINES SHOWN ARE PERPENDICULAR	
M017601D	264		1	APPLY TRANSFORMATIONAL GEOMETRY	
M018001D	265		1	APPLY PROPERTIES OF A CUBE	
M019001D	266		1	APPLY PROPERTIES OF A PARALLELOGRAM	
M019601D	267		1	APPLY PYTHAGOREAN THEOREM	
M019801E	268		1	DRAW AN OBTUSE ANGLE	(RATER 1)
M019901E	269		1	VISUALIZE A GEOMETRIC FIGURE	(RATER 1)
M020901E	270		1	DRAW A LINE OF SYMMETRY	(RATER 1)
M021001E	271		1	USE SIMILAR TRIANGLES	(RATER 1)
M021301E	272		1	USE TANGRAMS	(RATER 1)
M021302E	273		1	DRAW LINES TO FORM RECTANGLE	(RATER 1)
M022201F	274		1	DRAW GEOMETRIC FIGURE	(RATER 1)
M022501F	275		1	DRAW A GEOMETRIC FIGURE	(RATER 1)
M023101F	276		1	VISUALIZE A CUBE	
M015601G	277		1	STRAIGHT LINE CAN'T BE DRAWN ON SURFACE OF SPHERE	
M016301G	278		1	FLIP TRIANGLE OVER LINE L AND GET FIGURE E	
M016401G	279		1	DIST. BTWN MIDPOINT OF MN & MIDPOINT OF PQ = 30 CM	
M016601G	280		1	DIAGONAL MEASUREMENT OF TV SCREEN SHOWN IS 50 INCH	
M016701G	281		1	FIGURE A CONTAINS PERPENDICULAR LINE SEGMENTS	
M012731H	282		1	IDENTIFY TRIANGLE TYPE	
M012831H	283		1	FIND ANGLE IN TRIANGLE	
M027231I	284		1	THE LINE SEGMENT IS A DIAMETER IN CIRCLE A	
M027431I	285		1	FIGURE THAT HAS 2 CIRCULAR BASES - A CYLINDER	
M028331I	286		1	RATIO LENGTH SIDE EQUIL TRIANGLE TO PERIMETER 1:3	

DATA ANALYSIS AND STATISTICS SCALE

N250901C	287		1	80 BOXES OF ORANGES PICKED ON THURSDAY	(GRAPH)
N250902C	288	289	2	MORE LEMONS ON WED THAN ORANGES/GFRUIT	(GRAPH)
N250201C	290		1	BAG WITH 10 MARBLES BEST CHANCE TO GET RED ONE	
N263501C	291	292	2	AVERAGE AGE OF CHILDREN IS 7	
M017801D	293		1	INTERPRET PIE CHART DATA	
M018901D	294		1	FIND A MEDIAN	
M020201E	295		1	COMPLETE A BAR GRAPH	(RATER 1)
M020801E	296		1	LIST SAMPLE SPACE	(RATER 1)
M021101E	297		1	EXPLAIN SAMPLING BIAS	(RATER 1)
M023301F	298		1	SOLVE A PROBABILITY PROBLEM	
M023501F	299		1	FIND EXPECTED VALUE	
M023601F	300		1	INTERPRET A LINE GRAPH	
M015801G	301		1	AVERAGE WGHT 50 TOMATOES=2.36 COMBINED WGHT=118	
M016101G	302		1	9 CHIPS IN BAG - PROBABILITY DRAW EVEN CHIP = 4/9	
M017001G	303		1	15 GIRLS, 11 BOYS - PROBABILITY SELECT BOY = 11/26	
M012631H	304		1	INTERPRET CIRCLE GRAPH	
M013031H	305		1	FIND AN AVERAGE	(RATER 1)
M013131H	306		1	FIND A PROBABILITY	(RATER 1)
M028531I	307		1	MAKE A CIRCLE GRAPH TO ILLUSTRATE DATA	(RATER 1)

BEST COPY AVAILABLE

ALGEBRA AND FUNCTIONS SCALE

N256101C	308	1	THE VALUE OF N + 5 WHEN N = 3 IS 8	(RATER 1)
N264701C	309	1	X TIMES 1 = X TRUE WHEN ANY NO. SUBSTITUTED FOR X	
N255701C	310	1	2X + 3Y + 4X = 6X + 3Y	
M018301D	311	1	APPLY CONCEPT OF EQUALITY	
M018701D	312	1	SOLVE AN INEQUALITY	
M018801D	313	1	IDENTIFY COORDINATES ON A GRID	
M019301D	314	1	FIT EQUATION TO DATA	
M019701E	315	1	SOLVE A NUMBER SENTENCE	(RATER 1)
M020401E	316	1	COMPLETE A LETTER PATTERN	(RATER 1)
M021201E	317	1	GRAPH AN INEQUALITY	(RATER 1)
M022101F	318	1	COMPLETE A GEOMETRIC PATTERN	
M022401F	319	1	REPRESENT WORDS WITH SYMBOLS	
M023201F	320	1	EXTEND A NUMBER PATTERN	
M016001G	321	1	LEAST WHOLE NUMBER X FOR WHICH 2X > 11 IS 6	
M016801G	322	1	LENGTH OF RECTANGLE CAN BE EXPRESSED AS L - 3	
M016901G	323	1	IF PATTERN CONTINUES 100TH FIG. WILL HAVE 201 DOTS	
M016902G	324	1	EXPLAIN HOW GOT ANSWER FOR QUESTION 16	(RATER 1)
M012231H	325	1	USE ORDER OF OPERATIONS	
M013231H	326	1	EXTRAPOLATE NUMBER PATTERN	
M013731H	327	1	CONVERT TEMPERATURES	
M027131I	328	1	IF N + N + N = 60, THEN VALUE OF N = 20	
M027531I	329	1	3 X (BOX + 5) = 30 BOX = 5	
M027731I	330	1	TO GET 2ND NUMBER IN PAIRS: MULT. BY 2 AND ADD 1	
M027931I	331	1	COST TO RENT MOTORBIKE: FILL IN TABLE	(RATER 1)
M028431I	332	1	PLOT THE POINTS (5,2) ON THE GRID SHOWN	(RATER 1)

Appendix C

Instructions on using COMBPV.EXE

HOW TO USE COMBPV.EXE

COMBPV is an IBM-compatible personal computer program that is designed to combine the results of statistical analyses using different plausible values. We assume here that the data analyst has run an analysis several times, each time using a different set of plausible values. The statistical input to the program is the parameter estimates computed in the several analyses and also their error variances or covariances. The estimates may be of a single parameter or a vector of parameters. The error variances or covariances may be produced by standard statistical programs or by other techniques such as the jackknife. The output is an overall parameter estimate, its standard error or covariance, an F or t statistic and its number(s) of degrees of freedom, and its associated probability statistic.

COMBPV.EXE is a self standing QBASIC program. It needs no other program or file except for the information file from which the input data will be read. A word processor or other editing software is strongly suggested for creating the parameter file, but it is not required. In any case, the information file must be in ASCII format.

COMBPV works as follows. At the drive prompt, type COMBPV. The program will prompt the user to specify the file that contains the program information. Results will be placed on the computer screen. Upon completion of a run, the program will request a filename for the results, if they are to be saved, and ask if the user wishes to perform another run. The user must be careful when specifying an output file since if the file already exists, it will be rewritten, and the old contents will be lost.

COMBPV.EXE requires the following information in the information file in order to operate:

- a title for the analysis
- names of the parameters being estimated
- hypothesized values for the estimated parameter
- number of plausible values used in estimating the parameter (M)
- the number of degrees of freedom for the parameter estimates (N)
- number of parameters estimated (K)
- the estimated parameters
- error covariance matrix for the estimated parameters

COMBPV.EXE works by reading a program information file which contains the information necessary to compute the F or t statistic, the degrees of freedom, and its sampling probability. The program information file is set up in the following way:

Record 1: A title for the analysis to be performed. This can have up to 80 characters and may contain any characters, numbers, or letters. This title line is printed at the beginning of the output file together along with the date and time of the run.

Records 2 through 4: Each one of these lines must be written from the first column on. The first character of the line must be a K, M or N, followed by the "=" sign and the corresponding value. There must be a separate line for the number of parameters estimated (K), for the number of plausible values (M) and the number of degrees of freedom in the estimate of the parameters (N). These lines can be placed in any order, but must always be lines 2 through 4 of the parameter file. The letters K, N or M can be either lower or upper case. If the first letter of these lines are not either K, M or N, the program will automatically stop and a message will be displayed on the screen.

Record 5: This record could be left blank, or an identifying text could be included to make the file more readable for the user. The information in this line is not used by COMBPV. Please see examples below.

Records 6 through [6+(k-1)]: Each of these records will contain the name and hypothesized value for each of the K parameters being estimated. Each record begins with a parameter label of up to 10 characters followed by comma and the hypothesized value. The labels can be alphanumeric and may also have embedded blanks, but not commas. For example, in the case of testing a regression coefficient, this value would be 0.0 to the test the hypothesis that the regression coefficients are significantly different from zero. For precision purpose, the expected values for the parameters have to include at least one decimal place, even if it is zero. The program will look for as many parameter names and corresponding hypothesized values as were specified in the "K=" statement between lines 2 and 4.

Records [6+k] through end: The estimated parameters and their error covariance matrices must then be entered. The first line will contain identifying text followed in the next line by the K parameter estimates made by using the first plausible value. These parameter estimates must be separated by commas. The estimates, for precision purposes, should be written with adequate precision. The next K lines will contain the diagonal and the elements below the diagonal of the error covariance matrix of the estimates. The diagonal will contain the variances and the off diagonal elements of the covariance estimates. Even though all of the values in the matrix could be written on one line, or even each of the elements on a line each, it is recommended that they be written in the matrix form so as to facilitate checking the accuracy of the values. The element of this matrix will be read as follows: (1,1), / (2,1), (2,2), /(3,1), (3,2), (3,3), / etc. where the "/" symbol indicates a new record. The K+2 records containing the parameter estimates and their error covariances are repeated for the results from using the second plausible value, and so forth until the Mth set is entered.

Note that the line preceding the parameter estimates and their covariance matrix should be either be left blank, or a descriptive text can be included, such as PV#, to indicate the origin of the results and aid the user in reading the file (see example).

The program is written to recognize commas as delimiters for numerical and alphanumeric values, so these should not be used in any other way within the program information file.

The above is repeated for the estimates from each of the plausible values. The program will read as many sets of parameters and error covariance matrices as plausible values were specified in the "M=" record above.

When the program is run, the first prompt will ask the user to enter the name of the file containing the information necessary for the analysis. The name must be entered with its proper path and location on the disk. If the name of the file entered is not found, then the program will beep and prompt the user to enter a new file name.

When the program is running, and the proper information file is read, the information will be printed on the screen. This is a good time to check that the information that is being read in by the program is correct. After displaying any results on the screen, the computer will pause to allow the user to verify them. The process may be continued by pressing any one of the keys on the keyboard. If at any point the user detects that the program is reading the wrong file, or the information that was entered is incorrect, then pressing the keys Ctrl and Break simultaneously will automatically exit the user from the program. All of the results from COMBPV are then displayed on the screen for the user to check and take note of. After displaying the results on the screen, the user is asked if the results should be written to a file or not. A proper file name must then be specified. The user must be careful since specifying an already existing file will replace the old contents of the file with new ones, thus risking the loss of some information.

What follows is an example of an annotated program information file ready to be read with the COMBPV.EXE program. Other examples are included in the sample disk and can be identified by their extension .PAR at the end of their filename.

PARAMETER INFORMATION FILE (M8107.PAR)

```
EXAMPLE M8107.PAR - THREE PARAMETERS
K = 3
M = 2
N = 767
PARAMETER ESTIMATES
M810705B , 0.0
M810703B , 0.0
M810702B , 0.0
PV1
-2.005494 , -10.960869 ,   .967697
 2.65086
-0.29791   ,   2.20969
-1.08580   ,  -0.53598   , 3.53477
PV2
-2.263857 , -10.463685 , 1.609393
 2.63289
-0.29589   ,   2.19471
-1.07844   ,  -0.53235   , 3.51081
```

RESULT FILE:

COMBPV displays results on the screen, and also provides the user with the option of writing the results to a file on a disk. The output file proves to be useful since it can be attached to a document, printed, or inserted in the results section of an analysis.

The output printed to the screen while running COMBPV is exactly the same that is written to the output file. It is described in below, and the output of the parameter information file listed above is described.

DESCRIPTION OF THE SAMPLE OUTPUT

1. This line contains the title that was specified in the parameter file. It is followed by the time and date [of the internal clock of the computer] when the procedure was run.

2. The initial parameters are specified in the following lines. The number of plausible values, number of parameters tested, and the number of subjects or size of the sample used to obtain the parameter estimates.

3. This section prints out the names specified for the parameters, as well as their hypothesized values. The hypothesized value is that against which the obtained parameters are being tested.

4. This section contains the different parameter estimates obtained from the M sets of plausible values, together with their corresponding variance/covariance matrices. These parameters are those contained in the parameter file. The user may want to verify their accuracy and see if the parameters and elements of the matrix were read properly.

5. The U* matrix is the average sampling error. It is obtained by averaging each of the elements of the variance covariance matrix printed in section 4 above. It is the average error due to sampling.

6. The BM matrix contains the variance/covariance matrix due to imputation. In other words, the error component due to imputation.

7. In this section the summary of the results are presented. The matrix contains the average for each of the parameters being estimated, followed by the corresponding variance /covariance matrix of the estimates. This total error variance, includes the error due to sampling, as well as the error due to the imputation process.

8. The last part of the output contains the statistical test, reporting on the probability of obtaining the estimates for the parameters given their true or hypothesized value. If the probability value is less that 0.05, then the differences between the hypothesized and the observed are considered to be statistically significant at the 0.05 level. The significance tests reported are those for each of the individual parameters as well as the overall significance test.

SAMPLE OUTPUT FROM COMBPV
Parameter Information File: M8107.PAR

```
1   EXAMPLE M8107.PAR - THREE PARAMETERS -        08-06-1995     13:50:22

2   Number of Plausible Values        (M):  2
    Number of Parameters              (K):  3
    Number of Subjects                (N):  767

3   Parameter                    Hypothesized value
    M810705B                          0.0000
    M810703B                          0.0000
    M810702B                          0.0000

4   PARAMETER ESTIMATES AND ERROR COVARIANCE MATRIX - PLAUSIBLE VALUE  1

    Parameter     Estimate   | Error covariance matrix
    M810705B      -2.00549   |    2.6509     -0.2979      -1.0858
    M810703B     -10.96087   |   -0.2979      2.2097      -0.5360
    M810702B       0.96770   |   -1.0858     -0.5360       3.5348

    PARAMETER ESTIMATES AND ERROR COVARIANCE MATRIX - PLAUSIBLE VALUE  2

    Parameter     Estimate   | Error covariance matrix
    M810705B      -2.26386   |    2.6329     -0.2959      -1.0784
    M810703B     -10.46369   |   -0.2959      2.1947      -0.5324
    M810702B       1.60939   |   -1.0784     -0.5324       3.5108

5   AVERAGE SAMPLING ERROR  (U*)

    M810705B       M810703B       M810702B
         2.64188       -0.29690       -1.08212
        -0.29690        2.20220       -0.53417
        -1.08212       -0.53417        3.52279

6   ERROR DUE TO IMPUTATION (BM)

    M810705B       M810703B       M810702B
         0.03338       -0.06424       -0.08290
        -0.06423        0.12360        0.15952
        -0.08290        0.15951        0.20589

7   SUMMARY SECTION
    ---------------

    AVERAGE PARAMETER ESTIMATES (T*) AND TOTAL ERROR COVARIANCE MATRIX (V)

    Parameter    Estimate   | Total error covariance matrix
    M810705B      -2.1347   |    2.6919    -0.3933    -1.2065
    M810703B     -10.7123   |   -0.3932     2.3876    -0.2949
    M810702B       1.2885   |   -1.2065    -0.2949     3.8316

8   SIGNIFICANCE TESTS FOR INDIVIDUAL PARAMETERS
    --------------------------------------------
    Parameter          Estimate | Standard Error | T value |    DF    |  PROB.
    M810705B           -2.13468 |        1.64071 |   -1.30 |   216.59 | 0.1941
    M810703B          -10.71228 |        1.54518 |   -6.93 |   216.59 | 0.0000
    M810702B            1.28855 |        1.95745 |    0.66 |   216.59 | 0.5107

    OVERALL SIGNIFICANCE TEST RESULTS
    ---------------------------------

         F          DEGREES OF FREEDOM            P
      18.325         (  3 , 216.59)             0.0000
```

Appendix D

Contents of the Pimer Disk

This document lists the files contained in the NAEP Primer disk. Each file name is followed by a brief description of its contents.

C:\PRIMDISK\FILE.LST

The file containing this text.

c:\primdisk\combpv\combpv.exe
c:\primdisk\combpv\combpv.bas

These files contain the actual COMBPV program. The .BAS file contains the QBasic 4.5 source code. This is in plain text format so it can be examined and/or edited with any word processor or text editor. The .EXE file is the compiled version of the program. It is an executable file which can be run by typing COMBPV at the DOS prompt, from the corresponding sub directory.

c:\primdisk\combpv\combpv.txt
c:\primdisk\combpv\combpv.doc

Documentation on how to use the COMBPV program. The .DOC file is an MS Word for Windows formatted file. The .TXT file is a plain text file that can be read with any plain text editor or word processor.

c:\primdisk\combpv\m8107.par
c:\primdisk\combpv\ex42c.par

Two examples of parameter files that can be used with COMBPV.EXE. They correspond to examples in the analysis chapter of the Primer.

c:\primdisk\examples\ ....

This directory contains a set of 8 SPSS command files which were used in the examples included in the Primer. The file name corresponds to the example number followed by the extension .SPS indicating it is an SPSS command file.

c:\primdisk\layout\layout8p.txt
c:\primdisk\layout\layout8m.txt

These two files contain the file layout for the mini-files included in this diskette. The LAYOUT8P.TXT contains the layout for the 8th grade policy file and the LAYOUT8M contains the layout for the measurement file. They are both text files and can be printed directly from the DOS prompt or using a text editor or word processor. The layout files contain information about variable location, name, labels, and format.

c:\primdisk\minifile\m08ps1.sps
c:\primdisk\minifile\m08ms1.sps

This files contain the SPSS command files necessary to read the data contained in the policy mini-file as well as in the measurement mini-file. They contain DATA LIST specification, VARIABLE and VALUE labels. It is in plain text format so it can be read with any text editor, word processor, or directly included in SPSS.

c:\primdisk\minifile\m08ps1.dat

c:\primdisk\minifile\m08ms1.dat

These are the mini-data files. They variables are located as specified in the layout files. There are 1000 cases in each file. The measurement file is called M08MS1.DAT and the policy file is called M08PS1.DAT.

c:\primdisk\minifile\makemini.sps

This is the command file used to extract the cases for the NAEP Primer mini-files.

Appendix E

Q-Basic 4.5 Source Code for COMBPV.BAS

```
'
'=====================================================================
' COMBPV.BAS
'=====================================================================
'
DECLARE FUNCTION GETFILENAME$ (TEXT$)
DECLARE SUB WRITE2FILE (M!, K!, N!, TAU!(), T!(), U!(), IV$(), TSTAR!(), USTAR!(), BM!(),
V!(), F!, PF!, NU!, TITLE$)
DECLARE SUB PMAT2FILE (TEXT$, LABEL$(), A!(), MVAR!)
DECLARE SUB PMAT (TEXT$, LABEL$(), A!(), MVAR!)
DECLARE SUB MISLEVY (M!, K!, N!, TAU!(), T!(), U!(), IV$(), TITLE$)
DECLARE SUB SWP (A!(), MVAR!, K!, DET!)
DECLARE SUB READPAR (M!, K!, N!, TAU!(), T!(), U!(), IV$(), TITLE$)
DECLARE FUNCTION BETAI# (A!, B!, X!)
DECLARE FUNCTION BETACF# (A!, B!, X!)
DECLARE FUNCTION GAMMLN# (XX!)
DECLARE FUNCTION PROBF# (F!, DF1!, DF2!)
'
CLEAR
CLS
DIM SHARED FILENAME$
ON ERROR GOTO ERRORHANDLER
'
' INPUT NAME OF FILE CONTAINING THE PARAMETERS TO BE ANALYZED
'
FILENAME$ = GETFILENAME$("NAME OF THE FILE CONTAINING SPECIFICATIONS")
OPEN FILENAME$ FOR INPUT AS #1
CLS
'
' ==============================================================
'  THIS SECTION OF THE PROGRAM READS THE PARAMETERS TO BE USED
'  IN THE ANALYSIS.
' ==============================================================
'
LINE INPUT #1, TITLE$
PRINT TITLE$, DATE$, TIME$
PRINT
FOR I = 1 TO 3
LINE INPUT #1, RECORD$
  SPEC$ = UCASE$(MID$(RECORD$, 1, 1))
  SELECT CASE SPEC$
    CASE "K"
      NUMLEN = LEN(RECORD$) - INSTR(RECORD$, "=")
      K = VAL(RIGHT$(RECORD$, NUMLEN))
    CASE "N"
      NUMLEN = LEN(RECORD$) - INSTR(RECORD$, "=")
      N = VAL(RIGHT$(RECORD$, NUMLEN))
    CASE "M"
      NUMLEN = LEN(RECORD$) - INSTR(RECORD$, "=")
      M = VAL(RIGHT$(RECORD$, NUMLEN))
    CASE ELSE
      PRINT "CHARACTERS IN FIRST THREE LINES NOT RECOGNIZED"
      PRINT "PROGRAM WILL STOP": END
  END SELECT
NEXT I
'
PRINT "Number of Plausible Values      (M): "; M
PRINT "Number of Parameters            (K): "; K
PRINT "Number of Subjects              (N): "; N
DIM T(K, M), U(K, K, M), TAU(K), IV$(K)
PRINT
LINE INPUT #1, D$
PRINT "Parameter"; TAB(30); "Hypothesized value"
FOR I = 1 TO K
  INPUT #1, IV$(I), TAU(I)
  IV$ = MID$(IV$, 1, 10)
  PRINT IV$(I);
  PRINT USING "#####.####"; TAB(30); TAU(I)
NEXT I
'
' READ THE PARAMETERS AND ERROR COVARIANCE MATRICES
'
FOR I = 1 TO M
```

BEST COPY AVAILABLE

```
  LINE INPUT #1, D$
  FOR J = 1 TO K
    INPUT #1, T(J, I)
  NEXT J
  FOR J = 1 TO K
    FOR L = 1 TO J
      INPUT #1, U(J, L, I)
      U(L, J, I) = U(J, L, I)
    NEXT L
  NEXT J
NEXT I
'
' ... AND PRINT THEM TO CHECK THE INPUTED DATA FOR ACCURACY.
'
FOR Q = 1 TO M
PRINT
PRINT "PARAMETER ESTIMATES AND ERROR COVARIANCE MATRIX - PLAUSIBLE VALUE "; Q
PRINT
PRINT "Parameter    Estimate  | Error covariance matrix"
FOR I = 1 TO K
  PRINT IV$(I);
  PRINT USING "#####.#####"; TAB(12); T(I, Q);
  PRINT " | ";
  FOR J = 1 TO K
    PRINT USING " #####.####  "; U(J, I, Q);
  NEXT J
  PRINT
NEXT I
PRINT
PRINT "PRESS ANY KEY TO CONTINUE..."
DO UNTIL INKEY$ <> "": LOOP
NEXT Q
'
CALL MISLEVY(M, K, N, TAU(), T(), U(), IV$(), TITLE$)
'
CLOSE
INPUT "WOULD YOU LIKE TO RUN THIS PROGRAM AGAIN (Y/N)"; OK$
IF UCASE$(OK$) = "Y" THEN RUN
END
'
' ERROR HANDLER UTILITY
'
ERRORHANDLER:
BEEP
SELECT CASE ERR
  CASE 52, 53, 64, 75, 76
     PRINT
     PRINT "BAD FILE OR PATH NAME !! TRY AGAIN."
     PRINT
     FILENAME$ = GETFILENAME$("ENTER FILE NAME")
     RESUME
  CASE ELSE
     PRINT
     PRINT "UNEXPECTED ERROR HAS OCCURRED. PROGRAM WILL TERMINATE!!"
     END
END SELECT

FUNCTION BETACF# (A, B, X)
  CONST ITMAX = 100, EPS = .0000003
  AM = 1!
  BM = 1!
  AZ = 1!
  QAB = A + B
  QAP = A + 1!
  QAM = A - 1!
  BZ = 1! - QAB * X / QAP
 FOR M = 1 TO ITMAX
    EM = M
    TEM = EM + EM
    D = EM * (B - M) * X / ((QAM + TEM) * (A + TEM))
    AP = AZ + D * AM
    BP = BZ + D * BM
    D = -(A + EM) * (QAB + EM) * X / ((A + TEM) * (QAP + TEM))
    APP = AP + D * AZ
    BPP = BP + D * BZ
    AOLD = AZ
    AM = AP / BPP
    BM = BP / BPP
    AZ = APP / BPP
    BZ = 1!
```

```
    IF (ABS(AZ - AOLD) < EPS * ABS(AZ)) THEN EXIT FOR
 NEXT M
BETACF = AZ
END FUNCTION

FUNCTION BETAI# (A, B, X)
IF X < 0! OR X > 1! THEN
   PRINT "BAD ARGUMENT IN BETAI"
   GOTO 99
END IF
IF X = 0! OR X = 1! THEN
  BT = 0!
 ELSE
  BT = EXP(GAMMLN(A + B) - GAMMLN(A) - GAMMLN(B) + A * LOG(X) + B * LOG(1! - X))
END IF
IF (X < (A + 1!) / (A + B + 2!)) THEN
  BETAI = BT * BETACF(A, B, X) / A
  GOTO 99
ELSE
  BETAI = 1! - BT * BETACF(B, A, 1! - X) / B
  GOTO 99
END IF
99 END FUNCTION

FUNCTION GAMMLN# (XX)
DIM COF(6), STP, FPF, X, TMP, SER AS DOUBLE
COF(1) = 76.18009173#
COF(2) = -86.50532033#
COF(3) = 24.01409822#
COF(4) = -1.231739516#
COF(5) = .120858003#
COF(6) = -.536382#
STP = 2.50662827465#
FPF = 5.5#
X = XX - 1#
TMP = X + FPF
TMP = (X + .5#) * LOG(TMP) - TMP
SER = 1#
FOR J = 1 TO 6
  X = X + 1#
  SER = SER + COF(J) / X
NEXT J
GAMMLN = TMP + LOG(STP * SER)
END FUNCTION

FUNCTION GETFILENAME$ (TEXT$)
PRINT TEXT$;
INPUT TEMP$
GETFILENAME$ = TEMP$
END FUNCTION

SUB MISLEVY (M, K, N, TAU(), T(), U(), IV$(), TITLE$)
'
DIM TSTAR(K), USTAR(K, K), BM(K, K), V(K, K), STEP1(K), BMVIN(K, K)
'
' COMPUTE THE AVERAGE OF THE PARAMETER ESTIMATE (TSTAR)
'
FOR I = 1 TO K
    FOR J = 1 TO M
      TSTAR(I) = TSTAR(I) + T(I, J)
    NEXT J
  TSTAR(I) = TSTAR(I) / M
NEXT I
'
' NOW COMPUTE AVERAGE SAMPLING ERROR MATRIX (USTAR)
'
FOR I = 1 TO K
  FOR L = 1 TO K
    FOR J = 1 TO M
      USTAR(I, L) = USTAR(I, L) + U(I, L, J)
    NEXT J
    USTAR(I, L) = USTAR(I, L) / M
  NEXT L
NEXT I
CALL PMAT("AVERAGE SAMPLING ERROR (U*)", IV$(), USTAR(), K)
PRINT
PRINT "PRESS ANY KEY TO CONTINUE..."
DO UNTIL INKEY$ <> "": LOOP
'
' NOW COMPUTE ERROR DUE TO IMPUTATION (BM)
```

```
'
FOR I = 1 TO K
  FOR J = 1 TO K
    FOR L = 1 TO M
      BM(I, J) = BM(I, J) + (T(J, L) - TSTAR(J)) * (T(I, L) - TSTAR(J))
    NEXT L
    BM(I, J) = BM(I, J) / (M - 1)
  NEXT J
NEXT I
CALL PMAT("ERROR DUE TO IMPUTATION (BM)", IV$(), BM(), K)
PRINT
PRINT "PRESS ANY KEY TO CONTINUE..."
DO UNTIL INKEY$ <> "": LOOP
'
' NOW COMPUTE THE TOTAL COVARIANCE MATRIX
'
FOR I = 1 TO K
  FOR J = 1 TO K
    V(I, J) = USTAR(I, J) + (1 + 1 / M) * BM(I, J)
  NEXT J
NEXT I
'
' NOW COMPUTE THE INVERSE OF V (V IS DE-INVERTED LATER!)
'
DET = 1!
FOR I = 1 TO K
  CALL SWP(V(), K, I, DET)
NEXT I
'
' NOW COMPUTE THE F STATISTIC (F)
'
FOR I = 1 TO K
  FOR J = 1 TO K
    F = F + (TAU(J) - TSTAR(J)) * V(I, J) * (TAU(I) - TSTAR(I))
  NEXT J
NEXT I
'
' CORRECT THE F FOR THE NUMBER OF PARAMETERS (DIVIDE BY K)
'
F = F / K
'
' NOW COMPUTE THE DEGREES OF FREEDOM (NU)
'
FOR I = 1 TO K
  FOR J = 1 TO K
    FOR Q = 1 TO K
      BMVIN(I, J) = BMVIN(I, J) + BM(I, Q) * V(Q, J)
    NEXT Q
  NEXT J
NEXT I
'
' COMPUTE THE TRACE OF BM * INVERSE OF V (TRBMVIN)
'
FOR I = 1 TO K
  FOR J = 1 TO K
    TRBMVIN = TRBMVIN + BM(I, J) * V(J, I)
  NEXT J
NEXT I
FM = (1 + (1 / M)) * TRBMVIN / K
NU = 1 / ((FM ^ 2 / (M - 1)) + ((1 - FM) ^ 2 / (N - K)))
PF = PROBF(F, K, NU)
'
' NOW DE-INVERT THE  V MATRIX
'
DET = 1!
FOR I = 1 TO K
  CALL SWP(V(), K, I, DET)
NEXT I
'
' PRINT THE SIGNIFICANCE TEST RESULTS
' -----------------------------------
'
' PRINT THE AVERAGE PARAMETER AND TOTAL ERROR COVARIANCE MATRIX
'
PRINT "SUMMARY SECTION"
PRINT "---------------"
PRINT
PRINT "AVERAGE PARAMETER ESTIMATES (T*) AND TOTAL ERROR COVARIANCE MATRIX (V)"
PRINT
PRINT "Parameter   Estimate    | Total error covariance matrix"
```

```
FOR I = 1 TO K
  PRINT IV$(I);
  PRINT USING " #####.####"; TAB(12); TSTAR(I);
  PRINT " | ";
  FOR J = 1 TO K
   PRINT USING " #####.####"; V(I, J);
  NEXT J
  PRINT
NEXT I
'
' PRINT TEST FOR INDIVIDUAL PARAMETERS...
'
PRINT
PRINT "SIGNIFICANCE TESTS FOR INDIVIDUAL PARAMETERS"
PRINT "--------------------------------------------"
PRINT "Parameter        Estimate | Standard Error | T value |    DF    |    PROB."
TEMP$ = " ######.#####   |  ######.#####   | ####.## | #####.## |   #.####"
FOR I = 1 TO K
    STDERR = SQR(V(I, I))
    TVALUE = TSTAR(I) / STDERR
    TPROB = PROBF(TVALUE ^ 2, 1, NU)
    PRINT IV$(I); TAB(12);
    PRINT USING TEMP$; TSTAR(I); STDERR; TVALUE; NU; TPROB
NEXT I
PRINT
PRINT "OVERALL SIGNIFICANCE TEST RESULTS"
PRINT "---------------------------------"
PRINT
IF K = 1 THEN
    PRINT "     T          DEGREES OF FREEDOM              P   "
  TEMP$ = "#####.###         (###-,####.##)            #.####"
    PRINT USING TEMP$; SQR(F); K; NU; PF
  ELSE
    PRINT "     F          DEGREES OF FREEDOM              P   "
  TEMP$ = "#####.###         (###-,####.##)            #.####"
    PRINT USING TEMP$; F; K; NU; PF
END IF
'
' ASK USER IF RESULTS ARE TO BE PRINTED TO FILE AND DO SO IF REQUESTED
'
PRINT
INPUT "WOULD YOU LIKE THE RESULTS TO BE WRITEN TO A FILE (Y/N)"; OK$
OK$ = UCASE$(OK$)
SELECT CASE OK$
 CASE "Y"
   CALL WRITE2FILE(M, K, N, TAU(), T(), U(), IV$(), TSTAR(), USTAR(), BM(), V(), F, PF, NU,
TITLE$)
END SELECT

END SUB

SUB PMAT (TEXT$, LABEL$(), A(), MVAR)
'
' PRINTs a Square Matrix a()
'
PRINT
PRINT TEXT$
PRINT
FOR I = 1 TO MVAR
  PRINT LABEL$(I),
NEXT I
PRINT
FOR I = 1 TO MVAR
  FOR J = 1 TO MVAR
    PRINT USING "######.#####  "; A(I, J);
  NEXT J
  PRINT
NEXT I
PRINT
END SUB

SUB PMAT2FILE (TEXT$, LABEL$(), A(), MVAR)
'
' PRINTS a Square Matrix a() TO FILE #2
'
PRINT #2,
PRINT #2, TEXT$
PRINT #2,
FOR I = 1 TO MVAR
  PRINT #2, LABEL$(I),
```

```
NEXT I
PRINT #2,
FOR I = 1 TO MVAR
  FOR J = 1 TO MVAR
    PRINT #2, USING "######.#####  "; A(I, J);
  NEXT J
  PRINT #2,
NEXT I
PRINT #2,
END SUB

FUNCTION PROBF# (F, DF1, DF2)
X = DF2 / (DF2 + DF1 * F)
A = DF2 / 2!
B = DF1 / 2!
PROBF = BETAI(A, B, X)
END FUNCTION

SUB SWP (A(), MVAR, K, DET)
'
' Sweep Subroutine
'
DET = DET * A(K, K)
SELECT CASE A(K, K)
 CASE IS <= 0
        PRINT "The determinant is "; DET; " so no swept is done"
 CASE ELSE
      pivot = 1 / CDBL(A(K, K))
      A(K, K) = pivot
    FOR J = 1 TO MVAR
       IF J = K THEN GOTO 11
      FOR JP = 1 TO MVAR
       IF JP = K THEN GOTO 10
         A(J, JP) = CDBL(A(J, JP)) - (pivot * CDBL(A(J, K)) * CDBL(A(K, JP)))
10      NEXT JP
11    NEXT J
    FOR J = 1 TO MVAR
     IF J = K THEN GOTO 12
       A(K, J) = CDBL(A(K, J)) * pivot
       A(J, K) = CDBL(-A(J, K)) * pivot
12     NEXT J
  END SELECT
END SUB

SUB WRITE2FILE (M, K, N, TAU(), T(), U(), IV$(), TSTAR(), USTAR(), BM(), V(), F, PF, NU,
TITLE$)
'
' WRITES OUTPUT TO FILE WITH NAME ASSIGNED BY THE USER
'
FILENAME$ = GETFILENAME$("NAME OF THE OUTPUT FILE ")
OPEN FILENAME$ FOR OUTPUT AS #2
PRINT #2,
PRINT #2, TITLE$, DATE$, TIME$
PRINT #2,
'
PRINT #2, "Number of Plausible Values      (M): "; M
PRINT #2, "Number of Parameters            (K): "; K
PRINT #2, "Number of Subjects              (N): "; N
PRINT #2,
PRINT #2, "Parameter"; TAB(30); "Hypothesized value"
FOR I = 1 TO K
  PRINT #2, IV$(I);
  PRINT #2, USING "#####.####"; TAB(30); TAU(I)
NEXT I
'
' PRINT THEM TO CHECK THE INPUTED DATA FOR ACCURACY.
'
FOR Q = 1 TO M
PRINT #2,
PRINT #2, "PARAMETER ESTIMATES AND ERROR COVARIANCE MATRIX - PLAUSIBLE VALUE "; Q
PRINT #2,
PRINT #2, "Parameter    Estimate   | Error covariance matrix"
FOR I = 1 TO K
  PRINT #2, IV$(I);
  PRINT #2, USING "#####.#####"; TAB(12); T(I, Q);
  PRINT #2, " | ";
  FOR J = 1 TO K
    PRINT #2, USING " #####.####  "; U(J, I, Q);
  NEXT J
  PRINT #2,
```

BEST COPY AVAILABLE

```
NEXT I
NEXT Q
'
CALL PMAT2FILE("AVERAGE SAMPLING ERROR  (U*)", IV$(), USTAR(), K)
'
CALL PMAT2FILE("ERROR DUE TO IMPUTATION (BM)", IV$(), BM(), K)
'
PRINT #2, "SUMMARY SECTION"
PRINT #2, "---------------"
PRINT #2,
PRINT #2, "AVERAGE PARAMETER ESTIMATES (T*) AND TOTAL ERROR COVARIANCE MATRIX (V)"
PRINT #2,
PRINT #2, "Parameter   Estimate    | Total error covariance matrix"
FOR I = 1 TO K
  PRINT #2, IV$(I);
  PRINT #2, USING " #####.####"; TAB(12); TSTAR(I);
  PRINT #2, " | ";
  FOR J = 1 TO K
   PRINT #2, USING " #####.####"; V(I, J);
  NEXT J
  PRINT #2,
NEXT I
PRINT #2,
'
' PRINT TEST FOR INDIVIDUAL PARAMETERS...
'
PRINT #2,
PRINT #2, "SIGNIFICANCE TESTS FOR INDIVIDUAL PARAMETERS"
PRINT #2, "--------------------------------------------"
PRINT #2, "Parameter        Estimate | Standard Error | T value |    DF     |    PROB."
TEMP$ = " ######.#####  |  ######.#####  | ####.## | #####.## |  #.####"

FOR I = 1 TO K
    STDERR = SQR(V(I, I))
    TVALUE = TSTAR(I) / STDERR
    TPROB = PROBF(TVALUE ^ 2, 1, NU)
    PRINT #2, IV$(I); TAB(12);
    PRINT #2, USING TEMP$; TSTAR(I); STDERR; TVALUE; NU; TPROB
NEXT I
PRINT #2,
PRINT #2, "OVERALL SIGNIFICANCE TEST RESULTS"
PRINT #2, "---------------------------------"
PRINT #2,
IF K = 1 THEN
    PRINT #2, "     T          DEGREES OF FREEDOM              P    "
  TEMP$ = "#####.###          (###-,####.##)           #.####"
    PRINT #2, USING TEMP$; SQR(F); K; NU; PF
  ELSE
    PRINT #2, "     F          DEGREES OF FREEDOM              P    "
  TEMP$ = "#####.###          (###-,####.##)           #.####"
    PRINT #2, USING TEMP$; F; K; NU; PF
END IF
END SUB
```

BEST COPY AVAILABLE

BOSTON COLLEGE
Chestnut Hill, MA 02167