Contract No.: DAWS01-95-D-0029
MPR Reference No.: 8574-004

1998 Health Care Survey of DoD Beneficiaries:

Form A Codebook and User's Guide

August 1999

DRAFT

Submitted to:

United Healthcare
Global Consulting
12125 Technology Drive
Eden Prairie, MN 55343
(612) 833-7149

Project Officer:
Mary Zastrow

Submitted by:

Mathematica Policy Research, Inc.
600 Maryland Ave., SW, Suite 550
Washington, DC 20024-2512
(202) 484-9220

Project Director:
Myles Maxfield

PAGE IS INTENTIONALLY LEFT BLANK TO ALLOW FOR DOUBLE-SIDED COPYING

Contents

PAGE IS INTENTIONALLY LEFT BLANK TO ALLOW FOR DOUBLE-SIDED COPYING

Table of Variables

PAGE IS INTENTIONALLY LEFT BLANK TO ALLOW FOR DOUBLE-SIDED COPYING

Alphabetical Table of Variables

Chapter 1

Introduction

This Codebook and Users' Guide provides programmers and analysts with a tool to assist them in creating their own cross-tabulations and basic statistical estimates using the 1998 Health Care Survey of DoD Beneficiaries (HCSDB) Form A data set. It is intended for users wanting to create tables and to perform analyses other than those in the reports associated with this project.

Any user that wishes to recreate specific tables or charts from the analytic reports should also refer to "The 1998 Health Care Survey of DoD Beneficiaries: Technical Manual Form A"; that document outlines the procedures required to reproduce the report charts and specifies the variables used for each.

This chapter explains how to use this guide, reviews the survey, briefly describes the sample design, and concludes with a list of other documents on the HCSDB data that may be useful for policymakers, administrators, or others who may not be as familiar with computer systems.

How to Use This Guide

Chapter 2 describes the database conventions and types of variables in the database. This chapter explains the relationship of the raw survey data to the cleaned and constructed variables preferred for data analyses.

Chapter 3 provides table-making instructions in both SAS and SPSS, presenting the basic computer programming code needed to tabulate the data in SAS and the interactive steps for generating tables in SPSS. Either package may be used. While we assume that most users have some knowledge of computer systems and statistical processing, examples of how to create tables and the resulting output are given to simplify the process of tabulating the data. Because of the complex sample design, users interested in estimating standard errors will need to use a statistical package such as SUDAAN™ or WesVar PC® to develop estimates from the survey data. Sample programming code is included to estimate standard errors using methods that are appropriate for the complex sample design.

Chapter 4 is the codebook describing each variable in the database, including all possible values of the variable, weighted and unweighted frequency counts and percents for each value, and the values' interpretation or formatting. The codebook helps users assess the availability of certain measures, specify variables of interest, and identify all possible values of a variable. The variables are listed in the order of their position on the data file, where they are grouped according to source as follows:

Sampling variables used to place beneficiaries in appropriate strata

Information from the Defense Enrollment Eligibility Reporting System (DEERS) at the time of sampling

Updated DEERS and sampling information at the time of data collection

Questionnaire responses

Variables created during the fielding of the survey

Recoded questionnaire responses

Coding Scheme flags and missing value counts

Constructed variables for analysis

Weights

We also provide an alphabetical quick-reference list to help the user locate each variable.

Users who wish to know more about the technical aspects of the database creation, construction of new variables, or MPR's report production procedures should refer to "The 1998 Health Care Survey of DoD Beneficiaries: Technical Manual Form A," available from the TRICARE Management Activity Office.

What Is the HCSDB?

The HCSDB is an annual health care survey of active duty military personnel, retirees, and their adult family members. The child survey was not fielded in 1998. The survey is sponsored by the Assistant Secretary of Defense (Health Affairs) [OASD(HA)], under authority of the National Defense Authorization Act for Fiscal Year 1993 (P.L. 102-484). The adult Form A survey is intended to assess beneficiaries' satisfaction with and access to health care, knowledge of the TRICARE system, and use of preventive and other health care services.

The questionnaire, along with the coding scheme for each response category of each question, appears as Appendix A. A crosswalk between the 1998 questions and the questions from the 1994-95, 1996, and 1997 surveys appears in Appendix E.

The HCSDB covers the following topics:

Use of Health Care. In Section I, respondents are asked about their use of care in the past 12 months: inpatient nights, outpatient visits, and emergency room visits to military and civilian facilities. Civilian care is broken into visits paid for by TRICARE and visits paid for by private insurance, Medicare, or Medicaid. Data are also collected on the number of prescriptions written by civilian providers but filled at a military pharmacy.

Preventive Health Care and Health Habits. Section II collects information on the use of preventive health care services, including routine physical examinations, blood pressure readings, cholesterol screening, flu shots, and routine dental examinations. All women are asked about Pap smears; women that are or have been pregnant within the past 12 months are questioned about prenatal care. Women age 40 and over are asked about mammography and breast examination by a health care professional. Men are asked about prostate examinations. All respondents are asked whether they smoke or use chewing tobacco or snuff. Smokers are asked whether they have received smoking cessation counseling from a health care professional.

Understanding TRICARE. Section III assesses the respondent's level of understanding of TRICARE. Respondents are also asked about the source of their information on TRICARE.

Health Plan. Section IV collects data on TRICARE enrollment and the use of supplemental insurance coverage and/or other private insurance. TRICARE Prime enrollees are asked about their satisfaction with Prime and about the possibility of disenrolling. Those not currently enrolled are asked if they intend to enroll in the next 12 months. All respondents are asked about out-of-pocket expenses for medical care and health insurance coverage.

Satisfaction with Health Plan. Section V is designed to measure satisfaction with one's primary health plan. Respondents rate their personal doctor or nurse, specialist, and their health plan on a scale from

0 to 10 where 0 is the worst and 10 is the best. There are additional questions on problems with receiving necessary care, claims processing, finding and understanding written materials, customer service, processing paperwork, and resolving complaints.

Access to Health Care. Section VI collects information on where DoD beneficiaries received most of their care in the past 12 months (military or civilian facility) and evaluates access at these facilities by measuring the following:

- How often three or more telephone calls were required to make an appointment
- The length of time between making the appointment and the appointment itself for different types of care such as well-patient visits, specialty referrals, routine visits for minor illness or injury, and urgent care
- How often travel time to the facility was more than 30 minutes
- How often office waiting time to see the health care provider was more than 30 minutes

Satisfaction with Health Care. Sections VII and VIII contain a battery of questions about general and specific aspects of care at the facility used most; these questions cover topics such as availability of providers and their staff, convenience, and courtesy and respect shown by providers and their staff. Section VII questions are similar in content and format to questions in the Consumer Assessment of Health Plans Survey (CAHPS). CAHPS is a survey program sponsored by the Agency for Health Care Policy and Research (AHCPR), U.S. Department of Health and Human Services, and the Picker Institute. The program is designed to monitor the satisfaction and access to care of civilian health care plan beneficiaries.

Section VIII questions are mostly from the 1997 questionnaire to allow for trend analysis. There are questions on both military and civilian care; respondents who received care at military facilities are asked about their overall satisfaction and about satisfaction with specific aspects of military care. "Mirror" questions are asked of respondents who received care at civilian facilities.

Health Status. Section IX is designed to measure the respondent's self-perceived health status. A set of questions is based on the SF-12, a widely used scale that creates summary measures of physical and mental health status (Ware et.al. 1995). Respondents are also asked about the number of workdays missed because of health problems.

Facts About You. Section X collects demographic information, including income, operational deployment status, age, marital status, education, and race/ethnicity.

Sample Design Overview

The sample of beneficiaries for the HCSDB was drawn from an extract file of the DEERS database of military health system (MHS) beneficiaries with a reference date of July 29, 1998. The DEERS extract file includes all eligible MHS beneficiaries as follows:

- Everyone in the Uniformed Services and active duty (Army, Air Force, Navy, Marine Corps, Coast Guard, the Commissioned Corps of the Public Health Service, National Oceanic and Atmospheric Administration, Guard/Reserve personnel who are activated for a period in excess of 30 days, and other special categories of people who qualify for benefits)
- Those who retired from military careers
- Immediate family members of people in the previous two categories

- Surviving family members.

A stratified probability sample design was used to select DoD health care beneficiaries for the 1998 HCSDB. Strata were defined by a combination of enrollment status groups, and beneficiary groups, and geographic areas. Specific information on the sample design appears in, "The 1998 Health Care Survey of DoD Beneficiaries: Form A Sample Design", Mathematica Policy Research, Washington, D.C.

From a sample of 206,007 beneficiaries, 70,504 adult MHS beneficiaries completed and returned a 1998 HCSDB Form A questionnaire during the period November 1998 through June 1999, yielding a response rate of 35%. Information on developing response rates can be found in "The 1998 Health Care Survey of DoD Beneficiaries: Technical Manual Form A".

Other Documents on the 1998 HCSDB

This document is intended for programmers and analysts using the 1998 HCSDB Form A data. Following is a list of other documents that may be useful for broader audiences:

The 1998 Health Care Survey of DoD Beneficiaries: Form A Sample Design

The 1998 Health Care Survey of DoD Beneficiaries: Technical Manual Form A

The 1998 Health Care Survey of DoD Beneficiaries: National Executive Summary Report

The 1998 Health Care Survey of DoD Beneficiaries: Regional Reports

The 1998 Health Care Survey of DoD Beneficiaries: Catchment Reports

The 1998 Health Care Survey of DoD Beneficiaries: Medicare Subvention Demonstration Report

Any of these reports may be requested from the TRICARE Management Activity Office.

Chapter 2

Description of the HCSDB Form A Database

This chapter presents the procedures for developing the database, and presents the database file layout.

Variable Naming Conventions

The conventions used to name variables on the 1998 HCSDB Form A data file are listed below and summarized in Tables II.1.

Survey Variables. Survey variable names consist of up to eight alphanumeric characters that start with an alpha character ("H" for Form A survey variables), followed by a year designation ("98") and ending with three numbers and, if necessary, one alpha character to identify the relevant survey question. For example, the variable representing the first question on the Form A survey is given the name H98001.

Self-Reported Data. Defense Manpower Data Center (DMDC) standard demographic self-reporting variables on the Form A survey are prefixed with an “SR.” Survey variables with this naming convention include SRSEX (gender), SRAGE (age), and SRMARST (marital status).

Recoded Variables. Recoded variables with fewer than seven characters are suffixed with an "_R". For example, when the survey variable H98043 was modified during implementation of the Coding Scheme, a new variable H98043_R was created to represent the modified version of the variable H98043. When recoded variables have seven characters, the new variable name is suffixed with an “R.” For example, H98045AR-H98045DR represent the modified versions of the variables H98045A-H98045D.

Coding Scheme Flags and Counts. Coding Scheme flags, variables N1-N38M, reference the notes in the Coding Scheme for Form A. N5, for example, is set when checking the values of H98002 and H98003. See the Coding Scheme in Appendix B for more information. Coding Scheme counts are sums of missing value responses for each questionnaire; each of these variable names begin with the 4 characters “MISS”.

Constructed Independent Variables. Independent variables are prefixed with an "X." These include original survey variables modified as a result of data cleaning or recoding and newly constructed variables that did not previously exist on the survey file. For example, since the variable SRSEX was modified as a result of data cleaning and recoding, it was renamed XSEXA.

Constructed Dependent Variables. Dependent variables are given different prefixes depending on function. Healthy People 2000 variables, for example, are prefixed with an "HP,” SF12 scale variables (measures of health status) are prefixed with an "SF," demographic variables are prefixed with a “D,” and all other newly constructed dependent variables are prefixed with a "K."

Weighting Variables. Weighting variables are prefixed with a "W."

TABLE II.1

NAMING CONVENTIONS FOR 1998 HCSDB VARIABLES
(Variables Representing Survey Questions)

1st Character: Survey Type	2nd – 3rd Characters: Survey Year	4th – 6th Characters: Question #	Additional Characters: Additional Information
H= Health Beneficiaries (18 and Older, Form A)	98	001 to 120	A to S are used to label responses associated with a multiple response question ---------- __R denotes an edited variable

(Constructed Variables)

1st Characters: Variable Group	Additional Characters: Additional Information
SR=Self-reported demographic Data	Descriptive text, e.g., SRAGE
D=Constructed demographic data	Descriptive text, e.g., DAGE
N=Coding scheme notes	Number referring to Note, e.g., N2
X=Constructed independent variable	Descriptive text, e.g., XREGION
HP=Constructed Healthy Person 2000 variable	Descriptive text, e.g., HP_BP (had blood pressure screening in past two years and know the results)
SF12=SF-12 Health Status variables	Descriptive text, e.g., SF12PCS, SF12MCS (physical and mental health scores)
K=Constructed dependent variables	Descriptive text, e.g., KTOTINPT (total inpatient days)

Cleaning and Editing Conventions

Data quality procedures are found in the Coding Scheme tables. The complete Coding Scheme appears in Appendix B. It contains detailed instructions for all editing procedures used to correct data inconsistencies and errors. Editing procedures check for appropriate response values and consistent responses throughout the questionnaire. The steps to insure data quality include the following:

Initial Cleaning. Missing value flags were encoded when DRC created the SAS dataset:

- Skipped items were encoded with SAS missing value code of '.'.
- Multiple responses, where there should be a single response, were encoded with SAS missing value '.A'.
- Incomplete grid responses were encoded as SAS missing value '.I' with two exceptions: 1) If there was a response in the right column(s) and none in the left column(s), the missing grids were zero-filled; 2) if there was a response in the left column(s) and none in the right column(s), the field was right-adjusted and then zero-filled.

Data Cleaning and Recoding of Variables – Implementation of the Coding Scheme. Skip patterns were checked for consistency, and questions that were skipped legitimately were recoded with the SAS missing value of ".N"; questions that were answered, but should have been skipped, were recoded with a SAS missing value of ".C". When possible, variables were backward coded or forward coded to make all responses consistent within a sequence. Numeric values were checked, and values that were out of range were flagged with the SAS missing value of ".O".

Frequency Checks. Formatted and unformatted frequency tables for all variables in the 1998 HCSDB data file appear in Chapter 4 of this document. These frequency tables and other relevant cross tabulations were used to examine the range of values recorded for each data item to determine the type and magnitude of missing values. All value labels have been checked for accuracy.

Record Selection Criteria

Blank returns, nonrespondents, and any respondents found to be ineligible for MHS benefits were removed from the database. In addition, among eligible respondents with a non-blank questionnaire, a questionnaire must be "complete" to be included in the database.

To determine if a questionnaire is "complete", 18 key questions were chosen: 6, 7, 8, 9A, 9B, 11, 13, 26, 43, 53, 65, 67, 69A, 73, 84, 86, 88, 105, 115, 116, and 117. These key questions were adapted from the complete questionnaire rule developed by AHCPR for CAHPS surveys. At least 50 percent of these key items (nine or more) must be answered for it to be accepted as a complete questionnaire.

We retained 70,504 eligible respondents.

Weighting Procedures

The analysis of survey data from complex sample designs, such as the 1998 HCSDB, requires weights to do the following:

Compensate for variable probabilities of selection

Adjust for differential response rates

Improve the precision of the survey-based estimates through post-stratification [for details, see Brick and Kalton (1996) and references cited therein]

Sampling weights are equivalent to the reciprocal of the probability of each respondent's selection into the sample. Sampling weights are further adjusted for nonresponse within classes defined by sampling strata: a cross-classification of enrollment status, geographic area, and beneficiary group. These nonresponse-adjusted weights are then ratio-adjusted to population counts from the DEERS files to compensate for variations from the estimated population counts. To properly weight the data, an analyst should use the final weight WRWT98. Chapter 4 contains weighted and unweighted frequencies for each variable included in this data set.

Chapter
3

Programming Guide

This chapter is designed to help users create tables and variance estimates. Procedures for using SAS, SPSS, SUDAAN, and WesVarPC to create estimates are explained. Examples provided in the text are based on a preliminary version of the data which do not necessarily correspond to the final version of the 1998 HCSDB; layouts and formats are the same, but numeric results may differ.

How To Make a Table Using SAS

The 1998 HCSDB Form A dataset is in a Statistical Analysis System (SAS) format. SAS is a computer software system used for data management, summarization, and analysis. A format library for Form A is included along with the dataset. SAS can be run interactively or non-interactively (or batch mode), and the sample programs presented here can be run using either method. Special instructions are given later in the chapter for working interactively with the SAS Display Manager System in a Windows environment. All SAS programs generate a LOG and a LST file. The LOG file shows how SAS interprets your program and flags SAS syntax errors. The LST file shows the requested output.

File References, Libraries, and Options

SAS recognizes two types of datasets -- permanent and temporary. Permanent datasets, such as the HCSDB, are located through a LIBNAME that references the directory where the data is stored. For example, if the Form A dataset for 1998 is located on a CD-ROM in the subdirectory HCSDB98\FORMA, your LIBNAME statement must look like this:

LIBNAME INFORMA 'F:\HCSDB98\FORMA';

The Form A dataset can then be referred to as INFORMA.HCSDB98A, where INFORMA is the location of the file HCSDB98A.

A format library requires a LIBNAME LIBRARY statement that shows the location of the format library. For example, if the Form A format library is stored on your hard drive in a FMTLIB subdirectory, the LIBNAME statement should look like this:

LIBNAME LIBRARY 'C:\HCSDB98\FORMA\FMTLIB';

The OPTIONS statement controls page format and line length. A table with a “portrait” orientation might have this statement:

OPTIONS PS=79 LS=132;

A table with a “landscape” orientation that is left justified would have this OPTIONS statement:

OPTIONS PS=50 LS=175 NOCENTER;

DATA Step

The DATA step is used to create permanent or temporary datasets. It is also used to create new variables, modify existing variables, and limit the number of variables or observations. In a DATA step, you can do any or all of the following activities:

Construct new variables. For example, to construct a variable of active duty by sex:

```
/*  Active duty males   */
IF XSEXA = 1 AND BFGROUPP = 1 THEN XSEX_AD = 1;
*  Active duty females;
ELSE IF XSEXA = 2 and BFGROUPP = 1 THEN XSEX_AD = 2;
ELSE XSEX_AD = .;   /*  missing value  */
```

[Note: the two methods to insert comments: enclosed within /* */ or beginning with * and ending with a semicolon]

Modify existing variables. For example, if the self-reported age of a respondent is less than 18, make the age variable equal to a missing value:

```
IF SRAGE < 18 THEN SRAGE = .;
```

Limit the number of variables. Use a KEEP statement:

```
KEEP XREGION CACSMPL H98099A H98103A;
```

Limit the number of observations. Use a subsetting IF:

```
/*  Keep only region 3 observations  */
IF XREGION = 3;
```

Create a new temporary dataset. For example, CAC_1 is a temporary file of observations for only those respondents in catchment area 1:

```
LIBNAME INFORMA 'F:\HCSDB98\FORMA';
DATA CAC_1;

/*  Input file is HCSDB98A  */
SET INFORMA.HCSDB98A;
IF CACSMPL = 1;
RUN;
```

Create a new permanent dataset. For example, OUT.CAC_9901 is a permanent dataset only of Region 1 out-of-catchment respondents:

```
LIBNAME INFORMA 'F:\HCSDB98\FORMA';
LIBNAME OUT 'C:\HCSDB98\FORMA';
DATA OUT.CAC_9901;
SET INFORMA.HCSDB98A;
```

```
IF CACSMPL = 9901;
RUN;
```

PROC TABULATE

PROC TABULATE produces summary statistics in a table layout. The table can have up to three dimensions: page, row, and column. Within any dimension, multiple variables can be reported one after another or hierarchically. Useful statistics that are available in PROC TABULATE include:

N	number of observations with nonmissing values
NMISS	number of observations with missing values
MEAN	the arithmetic mean
SUM	the sum
PCTN	percent that one frequency represents of another frequency
PCTSUM	percent that one sum represents of another sum

The essential elements to execute PROC TABULATE are outlined below (items within < > are not required):

PROC TABULATE DATA=*your dataset* <option list>;
CLASS *class variables*;
VAR *analysis variables*;
TABLE << *page expression, > row expression, > column expression </ table options* >;
WEIGHT WRWT98;
RUN;

If the input file is to be limited to a specific population, a separate DATA step can precede the TABULATE, or a WHERE statement can be used within the TABULATE procedure. For example, to create a table from only respondents in catchment area 1, you would use the following statement after the PROC TABULATE statement:

```
WHERE CACSMPL = 1;
```

CLASS variables are any variables that are used for grouping; variables such as XREGION, XSEXA, and CACSMPL are good examples of class variables. Class variables can be either character or numeric and typically have a discrete number of values. Unless MISSING is specified in the options list in the PROC TABULATE state, any observations with a missing CLASS variable will be dropped from the table.

The VAR statement identifies all analysis variables for a table. Analysis variables must be numeric and can be either discrete or continuous. SAS excludes missing values when computing statistics such as means and percentages.

The WEIGHT statement identifies the numeric variable whose value is used for weighting each *analysis* variable. In the HCSDB for 1998, the weight variable is WRWT98.

The TABLE statement defines the table features. Every variable listed in this statement **must** be classified as either a class variable or an analysis variable in the CLASS or VAR statements. A comma separates each table dimension (page, row, and column). If there are three dimensions, the first is the page, the second is the row, and the last is the column. If there are only two dimensions, the first is the row and the second is the column. Tables with only one dimension are in column form. Each dimension expression is composed of the same following elements:

Analysis variables

Class variables

The universal class variable ALL, which summarizes the class variables in the same group or dimension

Keyword for the statistic to be performed, such as MEAN, SUM, or PCTSUM

A format modifier, which defines how to format values in cells. For example, F=8.2 will present values with a maximum of 8 positions and 2 digits to the right of the decimal.

Labels, which temporarily replace variable names and statistic keywords. These labels have the form ='*label*'; for example, XREGION='Region' or MEAN=' ' (to eliminate the word MEAN from the headings).

Crossing operator * (asterisk). The asterisk is used to cross elements within the same dimension. For example, you would use XENRLLMT*XSEXA to cross enrollment status by sex. The asterisk is also used to connect the statistic (e.g., MEAN, SUM) to the appropriate dimension; for example, to calculate the mean of respondents' satisfaction with the care received at military facilities, you would use H98099A *MEAN.

Denominator definitions are enclosed by < > (brackets).

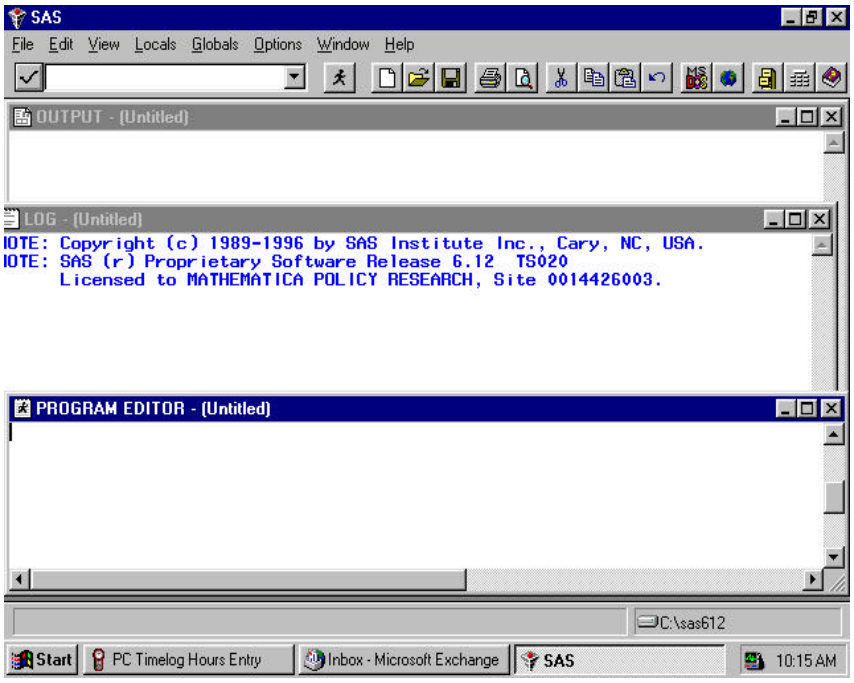
Concatenation operator is a single space between elements in a dimension. For example, to concatenate the satisfaction variables for military and civilian care, you would use H98099A H98103A.

Grouping is accomplished with parentheses. Below is an example of grouping, concatenation, and crossing within a single dimension:

(BFGROUPP ALL)*XSEXA

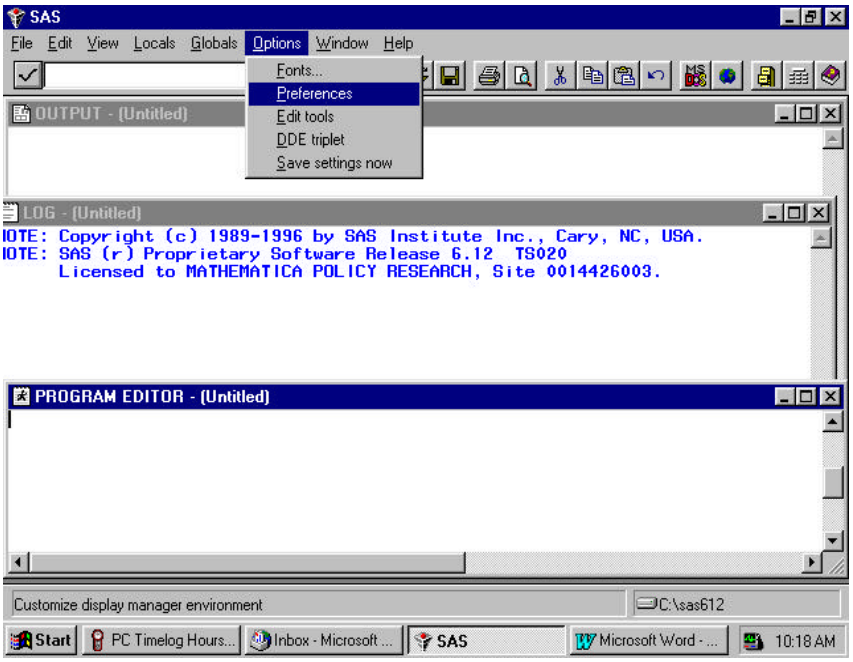
The SAS Display Manager System

The SAS Display Manager system provides an interactive tool for running SAS commands, like those given above, in the Windows environment. Double clicking the SAS icon on the desktop begins the SAS session. When you first enter the system, the following screen opens.

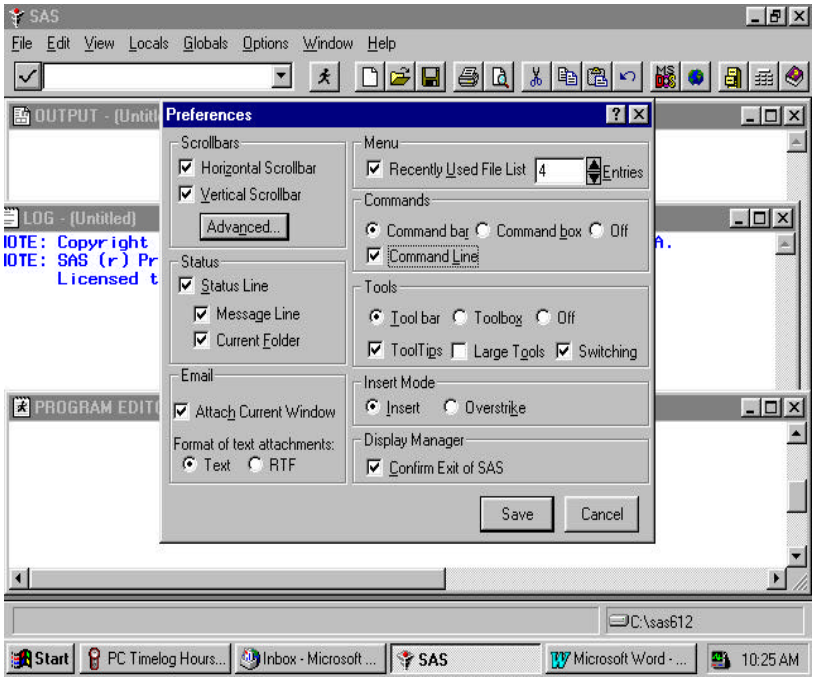


The screen is divided into three windows, each corresponding to an aspect of your SAS session. The **Program Editor** is the window where you compose, edit, and run your SAS commands. The **Log** window displays messages from the SAS system as well as your SAS statements as they are executed. Any error messages appear in the Log. The **Output** window displays the output tables requested in procedure commands written in the Program Editor. Toggling among the windows is accomplished by clicking anywhere in a given window. The cursor will jump to the selected window. Below are some options for customizing these screens by defining **Preferences**.

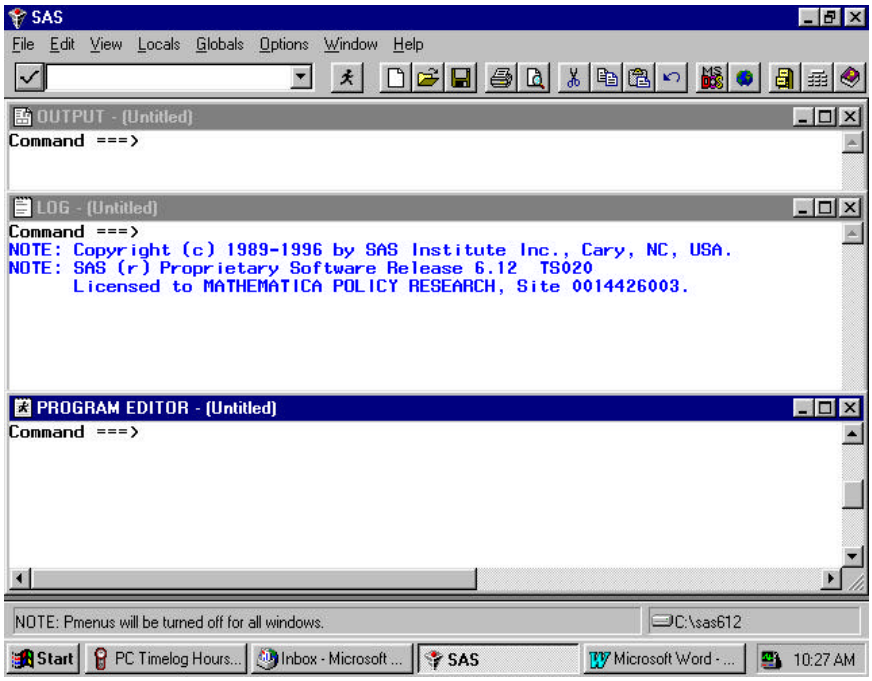
Clicking on **Options** results in the following screen.



Click on **Preferences** as highlighted above, and the following screen will open.



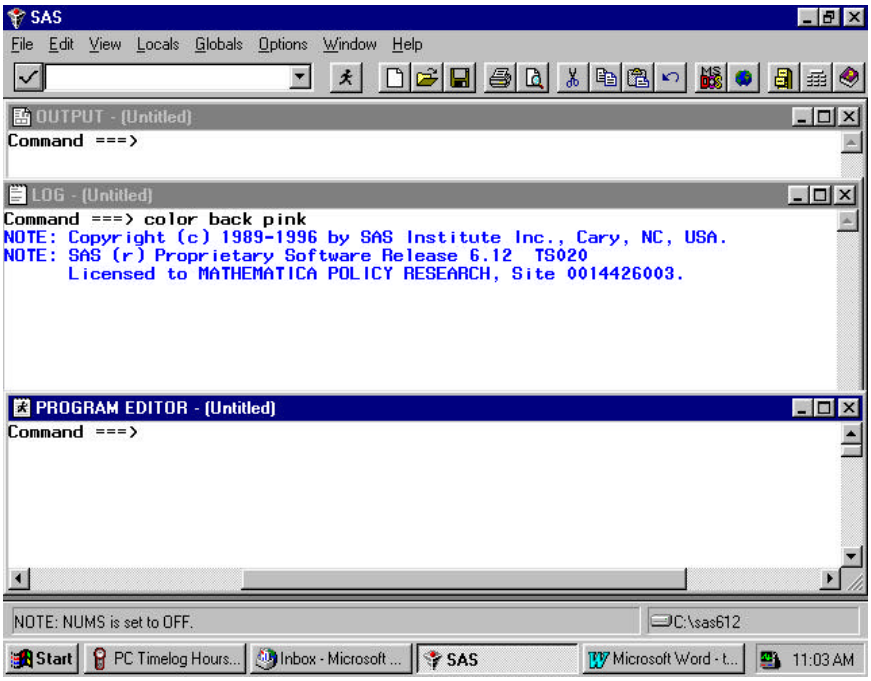
Many of these settings are system default options. To add a command line to the three windows, you would click in the box opposite **Command Line,** causing a check mark to appear in the box. Your screen should resemble the screen above. Click on **Save** and the screen will change to the following.



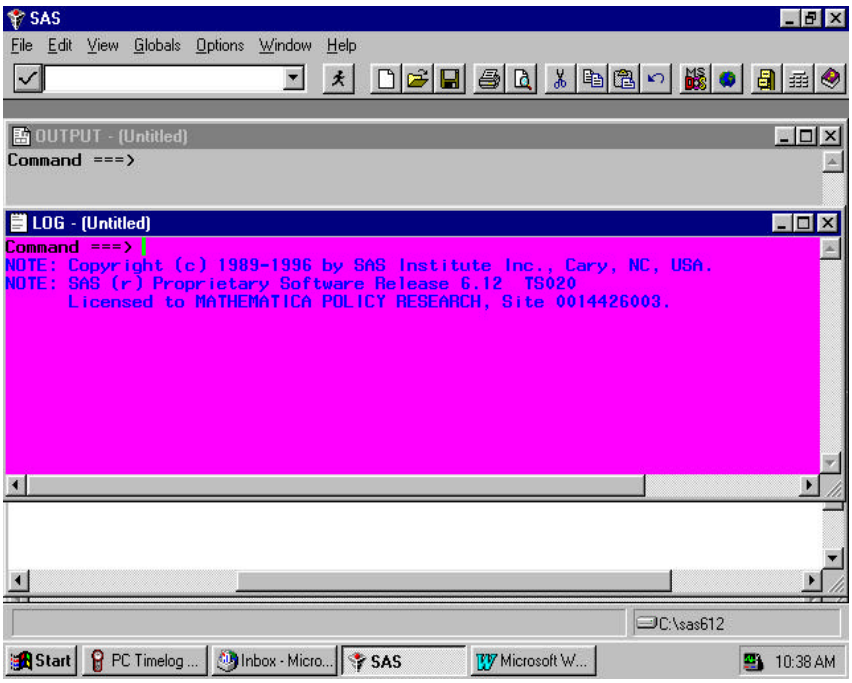
Each window shows the word **Command** followed by an arrow. Commands may be typed at this location. To arrive at the command line, depress the **Home** button on your keyboard. The cursor will appear opposite the arrow.

Toggling among the windows may be accomplished by typing the desired window name at the command line and pressing **Enter**. SAS recognizes **Pgm** as the abbreviated reference to the Program Editor and **Out** as a shortened name for the Output window. A few keystrokes allows you to navigate among the windows. For example, the command line lets you continue to customize our SAS session as follows.

In order to more easily distinguish between the SAS windows, it may be preferable to change the background color of selected windows. As an example, set the background color of the Log window to pink and the Output window to gray. Press the **Home** key to arrive at the command line. Type **Log** opposite the arrow to toggle to the Log window. Type the command, ***color back pink*** (or some other color) on the command line. Your screen will resemble the following.



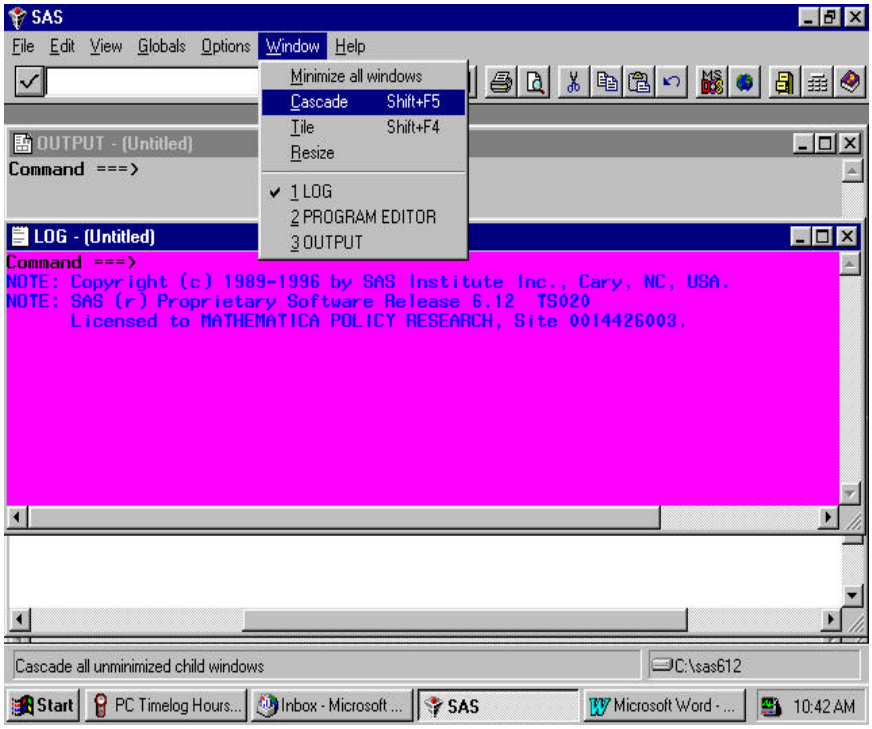
Press **Enter** to process the commands and the window will shade to pink. Toggle to the Output window by typing **Out** and keying **Enter**. Type ***color back gray*** and key **Enter**. These changes make it easier to distinguish between the windows at a glance. The screen looks like the following.



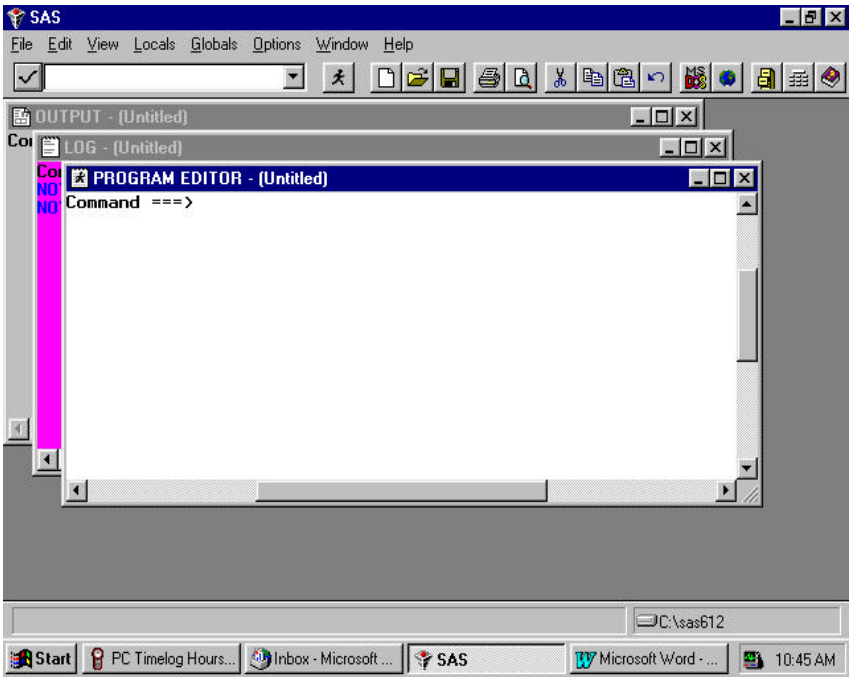
The following option also uses color to distinguish between windows.

All SAS statements for building and processing SAS datasets are typed into the Program editor. A SAS session may involve typing statements like the ones above for library reference, computing new variables, data steps, etc. Entering a long series of statements in such a small space may be awkward, so another arrangement for the windows may be preferable.

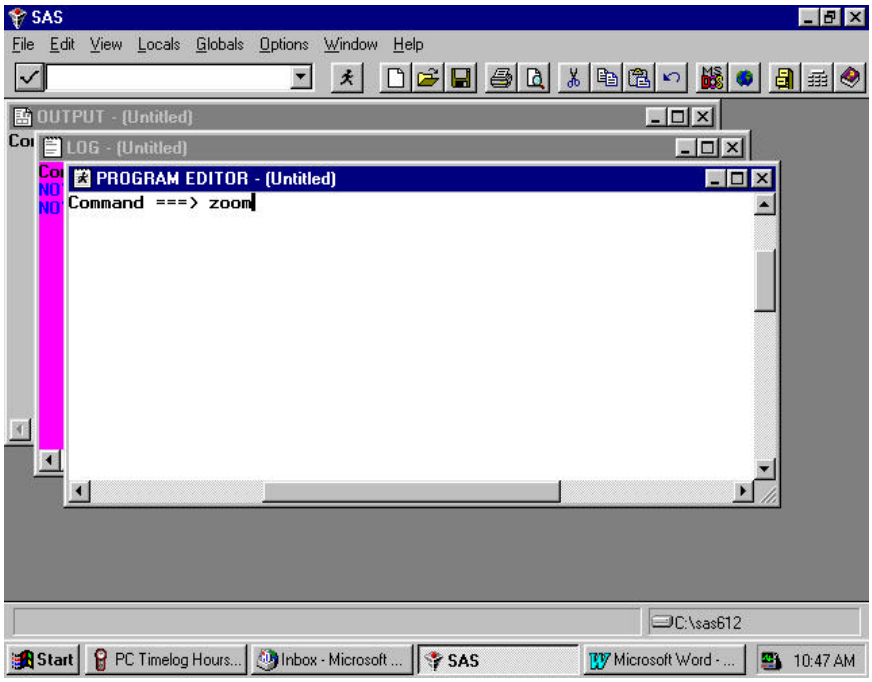
Cascading the windows is one option. To cascade the windows, open the **Window** menu, and choose **Cascade** as indicated in the following.



Clicking the option **Cascade** produces the following result.

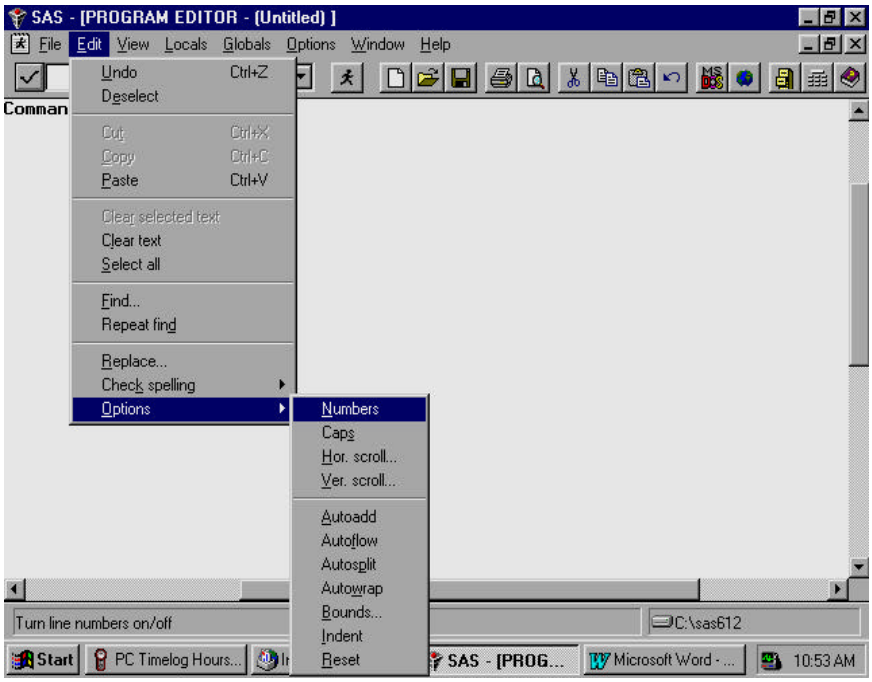


Each window is partly superimposed on the other. The colors distinguish between windows at a glance. With the Program Editor in front, SAS statements may be typed there with relative ease. As a final option, you can enlarge the Program Editor to fill the entire screen. At the command line, type ***zoom*** as in the following:

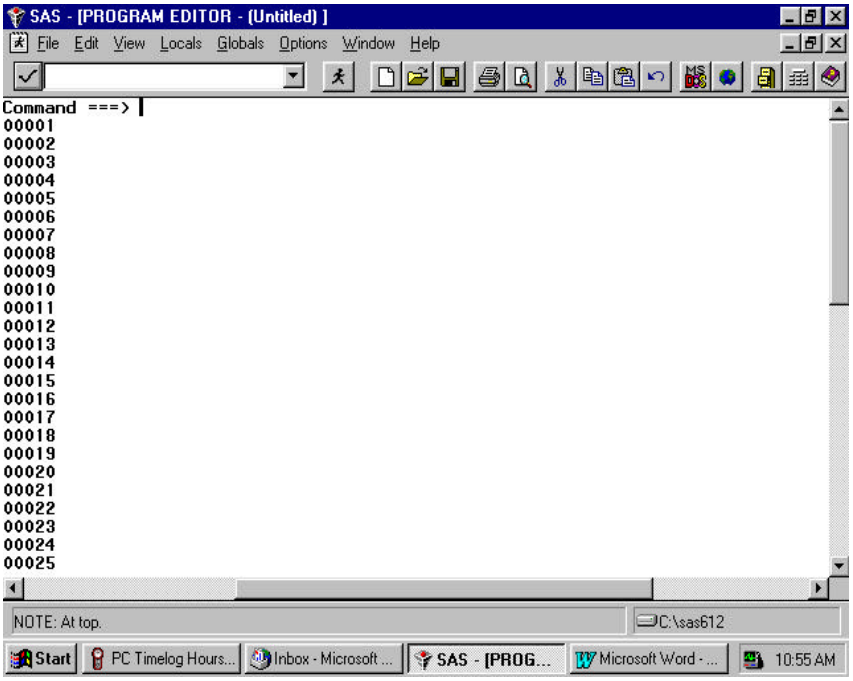


The window changes to fill the screen.

One more option for customizing screens is explained below. This involves adding line numbers to the editing environment in the Program Window. After adding the line numbers, many useful line-editing commands become available (see the SAS Manual). On the **Edit** menu choose **Options** and **Numbers** as in the following screen.



The line numbers appear at the left of the full screen Program Editor as in the screen below, and the SAS statements can be typed into the screen and edited.

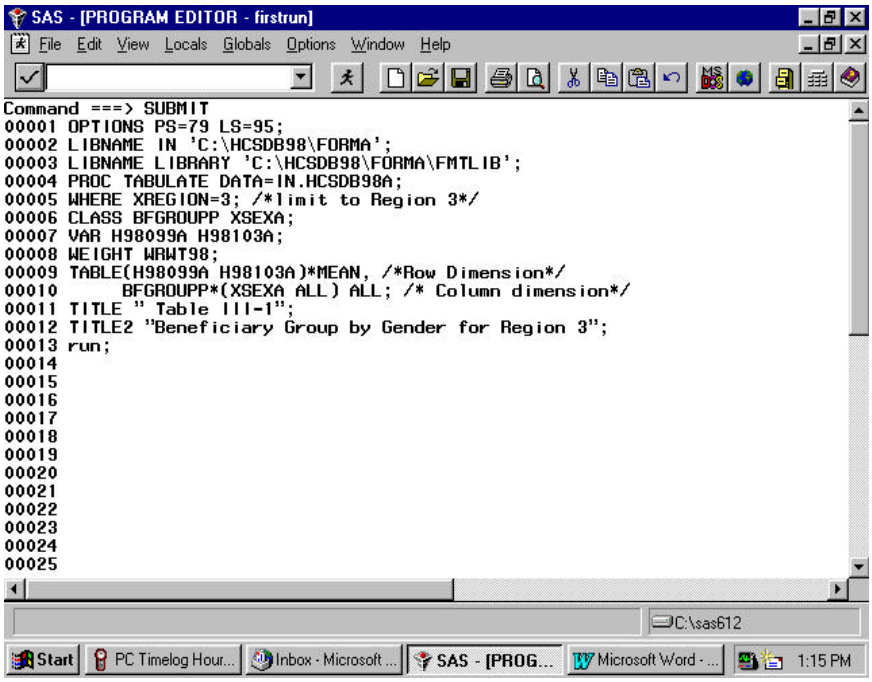


Below is an example of a PROC TABULATE to construct a table of satisfaction variables by beneficiary group by gender for respondents in region 3. Beneficiary group (BFGROUPP) and sex (XSEXA) are both class variables with a discrete number of values. The columns of the table are beneficiary group broken out by sex, a total for each beneficiary group, and a region total. The satisfaction variables (H98099A and H98103A) are the analysis variables appearing as the rows of the table. The statistic that we want to see is the weighted mean of these satisfaction variables for each group in the table and for the entire region as a whole.

Enter the following SAS statements into the **Program Editor.**

```
OPTIONS PS=79 LS=95;
LIBNAME IN 'C:\HCSDB98\FORMA';
LIBNAME LIBRARY 'C:\HCSDB98\FORMA\FMTLIB';
PROC TABULATE DATA=IN.HCSDB98A;
WHERE XREGION = 3;   /* limit to Region 3  */
CLASS BFGROUPP XSEXA;
VAR H98099A H98103A;
WEIGHT WRWT98;
TABLE (H98099A H98103A)*MEAN,    /* Row Dimension */
        BFGROUPP*(XSEXA ALL)  ALL;    /* Column Dimension */
TITLE "Table III-1";
TITLE2' Beneficiary Group by Gender for Region 3';
RUN;
```

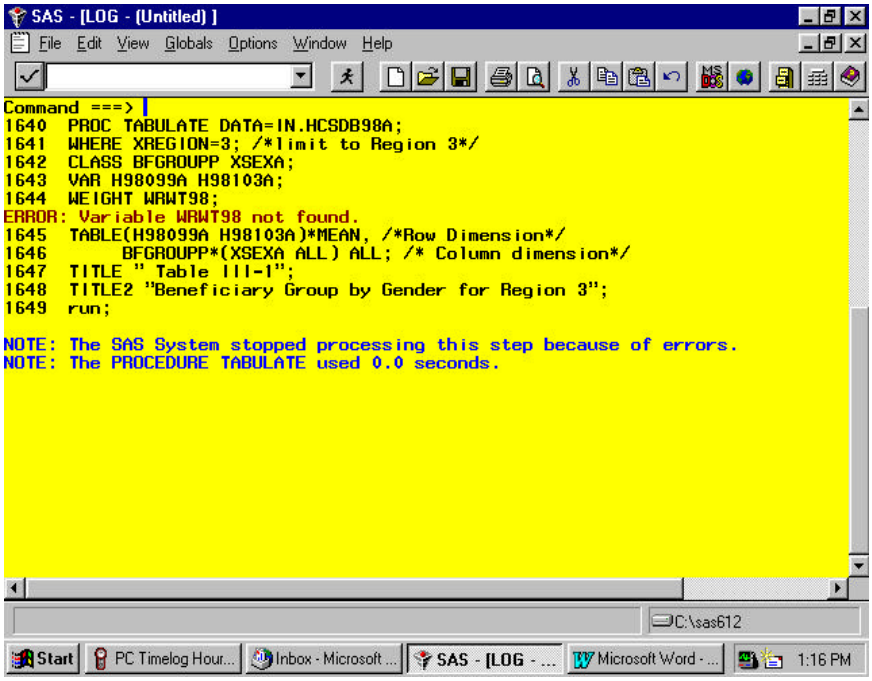
Key **Home** and type the command **SUBMIT** on the Command Line. **Submit** instructs the SAS system to process the commands written in the Program Editor. Your screen should resemble the following.



Enter the **Submit** command, and the SAS statements disappear from the Program Editor.

If a table is successfully produced, the Output window will open and the table will be displayed. If no output is produced, then SAS has encountered an error. SAS statements about the error can be seen and evaluated in the **Log** window. In *all* cases, the Log window should be carefully examined after SAS statements are processed. SAS may produce a table even if there are errors in the program, so the table may not be correct.

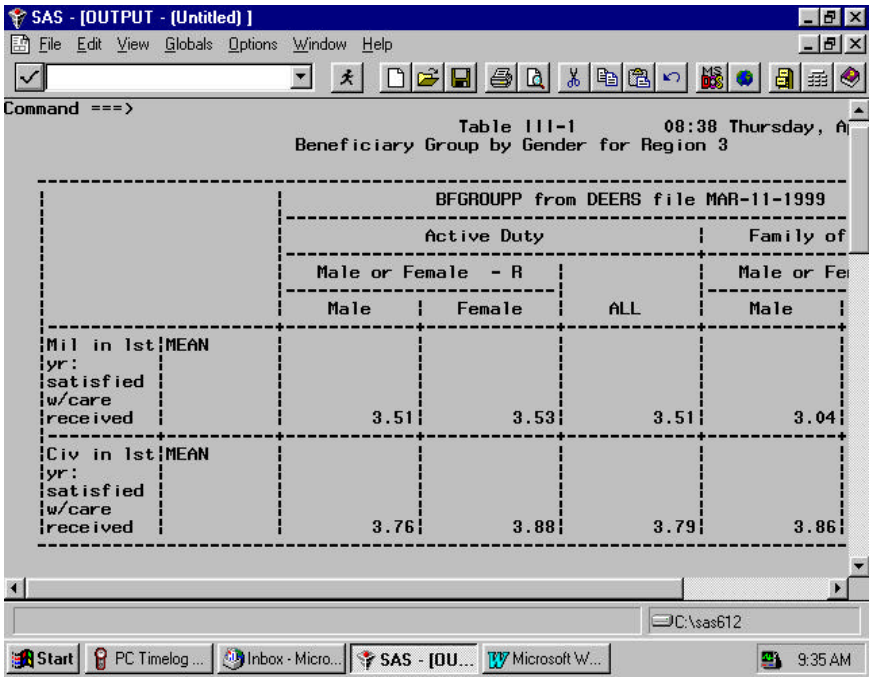
No table was produced for this run. The error is indicated in the Log Window as shown below.



The variable *WRWT98* was not found in the dataset. Type **Pgm** on the Command line to return to the Program Editor. Type **Recall** on the Command line and the program statements will reappear in the window.

You can correct the error by entering the SAS statements for computing the variable *WRWT98* into the program and rerunning the procedure. (WRWT98 is, in fact, included in the data set HCSDB98A but excluded here for the sake of the example.)

The corrected program produces the following output.



The result of this process is Table III.1.

Note that the TITLE statement defines the heading for each page. Titles of more than one line are entered as TITLE, TITLE2, etc.

Using Formats

The format library is the key to interpreting values of discrete variables. For example, in the program above, the format library found at C:\HCSDB98\FORMA\FMTLIB indicates that a value of 1 for XSEXA means male, and a value of 2 for XSEXA means female. Similarly, if BFGROUPP equals 1, the respondent is active duty; if BFGROUPP equals 2, the respondent is a family member of active duty personnel; if BFGROUPP equals 3, the respondent is an under-65 retiree or a survivor or one of their family members; if BFGROUPP equals 4, the respondent is a 65-or-over retiree, survivor, or one of their family members.

Since formats are associated with the variables in the HCSDB, formatting is automatic as long as SAS can locate the format library. Error messages will result if the LIBNAME LIBRARY statement is not present. If the format library is not available for some reason, use the statement

FORMAT _ALL_;

within the PROC TABULATE to prevent SAS from searching for the missing format library. The default formats in the format library were used to produce the table described in the previous section.

Table Appearance

Format modifiers and temporary labels improve the appearance of a table. In Table III.1, the values of the statistics are of the form x.xx. If each cell is defined to be six positions wide with two positions to the right of the decimal, there is adequate space plus some extra room to keep the table from looking crowded. This is done by crossing the statistic with the format modifier:

MEAN*F=6.2

Labels are attached to all variables in the HCSDB. You can use temporary labels to override the label within the SAS dataset. It is not always necessary to use both the variable label and the formatted values for each value of a class variable. In the previous example, the formatted values of BFGROUPP are active duty, family members of active duty, etc.—more beneficiary groups; the title also tells you that these are beneficiary groups. The table can be made attractive by deleting the heading for BFGROUPP by including a blank for the temporary label:

BFGROUPP=' '

Similarly, because the statistic being reported here is a mean, you do not need MEAN on each row. You can add or eliminate a label and include a format modifier to the same variable:

MEAN=' '*F=6.2

The headings for XSEXA and ALL can be improved:

XSEXA='Gender'
ALL='Group Total' for the ALL that is crossed with BFGROUPP
ALL='Total' for the Region 3 total

Table III.1
Beneficiary Group by Gender for Region 3

		BFGROUPP from DEERS file MAR-11-1999					
		Active Duty			Family of Active		
		Male or Female - R			Male or Female - R		
		Male	Female	ALL	Male	Female	
Mil in lst yr: satisfied w/care received	MEAN	3.51	3.53	3.51	3.04	3.53	
Civ in lst yr: satisfied w/care received	MEAN	3.76	3.88	3.79	3.86	4.00	

(CONTINUED)

		BFGROUPP from DEERS file MAR-11-1999				
					Ret/Surv/Fam 65+	
			Ret/Surv/Fam <65			
		Family of Active	Male or Female - R			Male or Female - R
		ALL	Male	Female	ALL	Male
Mil in lst yr: satisfied w/care received	MEAN	3.51	3.42	3.62	3.52	3.29
Civ in lst yr: satisfied w/care received	MEAN	4.00	4.11	4.05	4.08	4.29

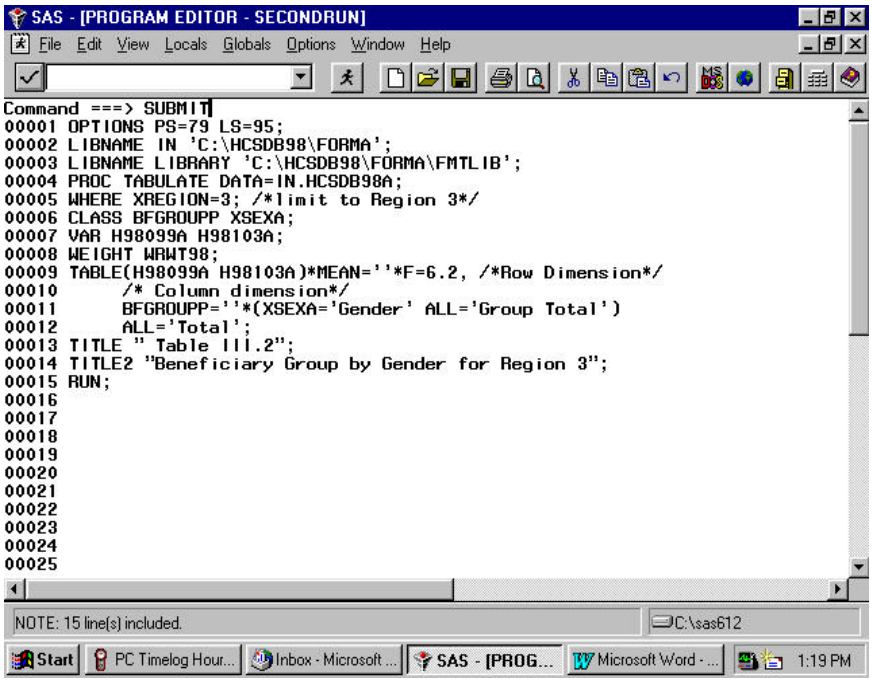
		BFGROUPP from DEERS file MAR-11-1999		
		Ret/Surv/Fam 65+		
		Male or Female - R		
		Female	ALL	ALL
Mil in lst yr: satisfied w/care received	MEAN	3.24	3.28	3.46
Civ in lst yr: satisfied w/care received	MEAN	4.24	4.26	4.14

The new program looks like this:

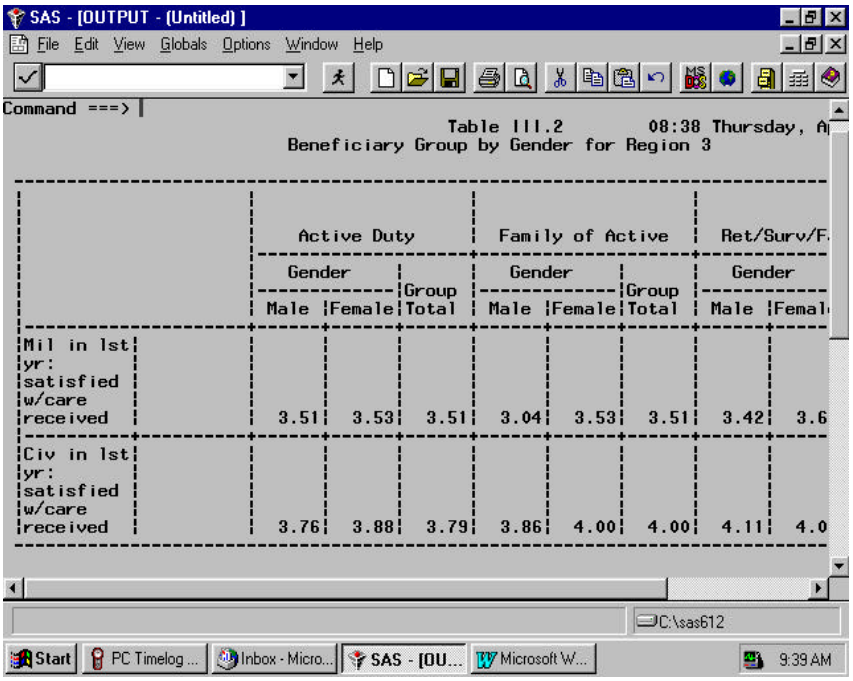
```
OPTIONS PS=79 LS=95;
LIBNAME IN 'C:\HCSDB98\FORMA';
LIBNAME LIBRARY 'C:\HCSDB98\FORMA\FMTLIB';
PROC TABULATE DATA=IN.HCSDB98A;
WHERE XREGION = 3;   /* limit to Region 3  */
CLASS BFGROUPP XSEXA;
VAR H98099A H98103A;
WEIGHT WRWT98;
TABLE (H98099A H98103A)*MEAN=' '*F=6.2,    /* Row Dimension */
            /* Column Dimension */
            BFGROUPP=' '*(XSEXA='Gender' ALL='Group Total')
            ALL='Total';


TITLE "Table III.2";
TITLE2 'Beneficiary Group by Gender for Region 3';
RUN;
```

Typing these statements into the Program Window produces the following screen.



After the **Submit** command is entered, the following table is displayed in the Output window.



The resulting output is in Table III.2.

Table III.2

Beneficiary Group by Gender for Region 3

	Active Duty			Family of Active			Ret/Surv/Fam <65			Ret/Surv/Fam 65+
	Gender		Group	Gender		Group	Gender		Group	Gender
	Male	Female	Total	Male	Female	Total	Male	Female	Total	Male
Mil in lst yr: satisfied w/care received	3.51	3.53	3.51	3.04	3.53	3.51	3.42	3.62	3.52	3.29
Civ in lst yr: satisfied w/care received	3.76	3.88	3.79	3.86	4.00	4.00	4.11	4.05	4.08	4.29

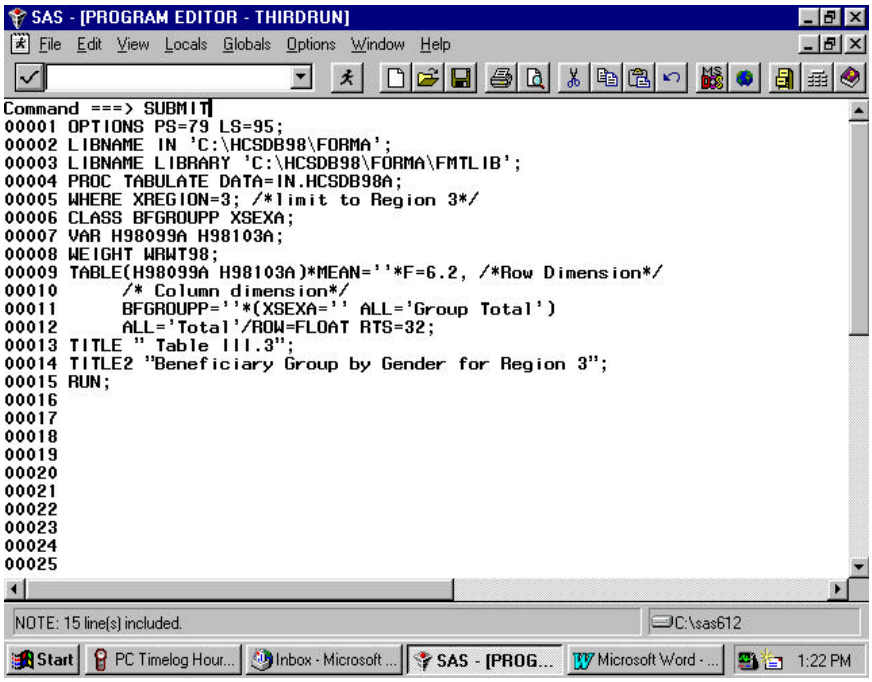
(CONTINUED)

	Ret/Surv/Fam 65+		
	Gender		
	Female	Group Total	Total
Mil in lst yr: satisfied w/care received	3.24	3.28	3.46
Civ in lst yr: satisfied w/care received	4.24	4.26	4.14

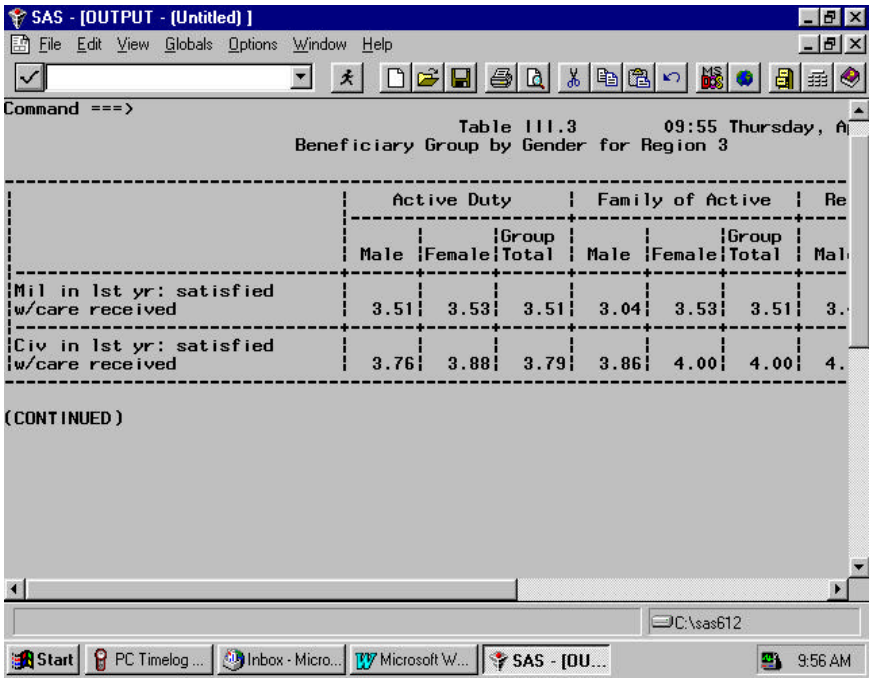
Although the label for MEAN is deleted, there is still a space in the table for this label. You can eliminate this blank space by using the TABLE option of ROW=FLOAT. SAS row headings are automatically allocated; you can override this by using the TABLE option of RTS=n where n is an integer value specifying the number of print positions to be used for row headings. If you decide that we don't need the label 'Gender' for XSEXA because 'male' and 'female' are self-explanatory, the revised program is as follows:

```
OPTIONS PS=79 LS=95;
LIBNAME IN 'C:\HCSDB98\FORMA';
LIBNAME LIBRARY 'C:\HCSDB98\FORMA\FMTLIB';
PROC TABULATE DATA=IN.HCSDB98A;
WHERE XREGION = 3;   /* limit to Region 3  */
CLASS BFGROUPP XSEXA;
VAR H98099A H98103A;
WEIGHT WRWT98;
TABLE (H98099A H98103A)*MEAN=' '*F=6.2,    /* Row Dimension */
        /* Column Dimension */
        BFGROUPP=' '*(XSEXA=' ' ALL='Group Total')
        ALL='Total' / ROW=FLOAT   RTS=32;
Title "Table III.3";
TITLE2' Beneficiary Group by Gender for Region 3';
RUN;
```

Typed into the Program Window, the revised program appears as follows.



The output table is displayed in the Output Window as follows.



The result is Table III.3.

Table III.3

Beneficiary Group by Gender for Region 3

	Active Duty			Family of Active			Ret/Surv/Fam <65		
	Male	Female	Group Total	Male	Female	Group Total	Male	Female	Group Total
Mil in lst yr: satisfied w/care received	3.51	3.53	3.51	3.04	3.53	3.51	3.42	3.62	3.52
Civ in lst yr: satisfied w/care received	3.76	3.88	3.79	3.86	4.00	4.00	4.11	4.05	4.08

(CONTINUED)

	Ret/Surv/Fam 65+			
	Male	Female	Group Total	Total
Mil in lst yr: satisfied w/care received	3.29	3.24	3.28	3.46
Civ in lst yr: satisfied w/care received	4.29	4.24	4.26	4.14

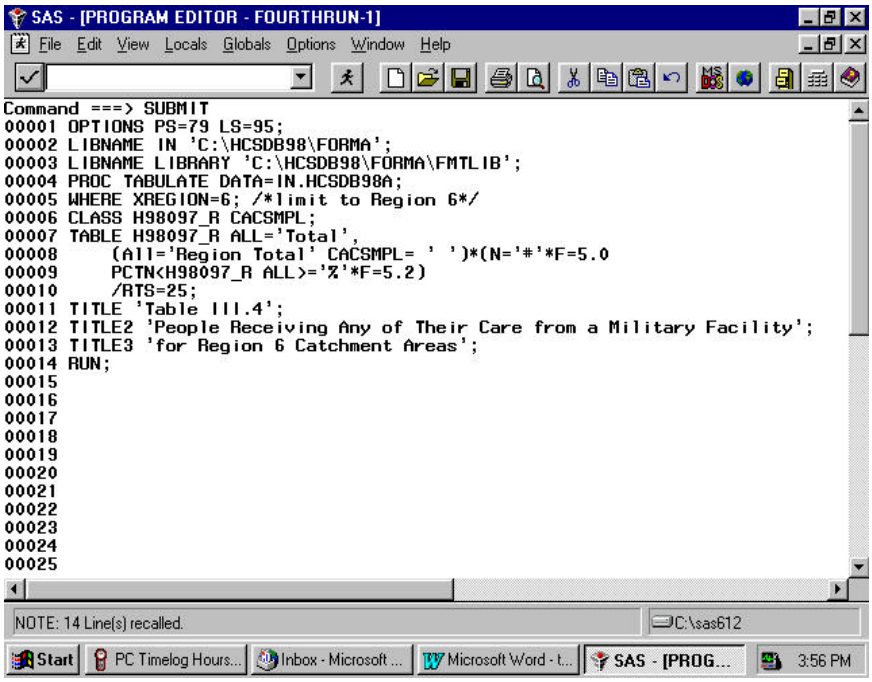
Calculating Percents

When calculating percentages, it is necessary to appropriately define the denominator. To calculate a column percentage, the denominator definition must include all *class* variables that define the *row*. For example, if you want to look at the percentage of people in your region and each of the catchment areas who answered yes (or no) to question 97, 'Did you receive any health care from a military facility or provider in the past 12 months?', your TABLE statement in the TABULATE procedure would look like this:

```
WHERE XREGION = 6;
TABLE H98097_R ALL='Total',
      (All='Region Total'  CACSMPL)*PCTN<H98097_R  ALL>='Percent';
```

Table III.4 includes a program and its output for calculating column percentages.

The program statements in the Program Editor appear as follows.



The submitted statements produce the following output.

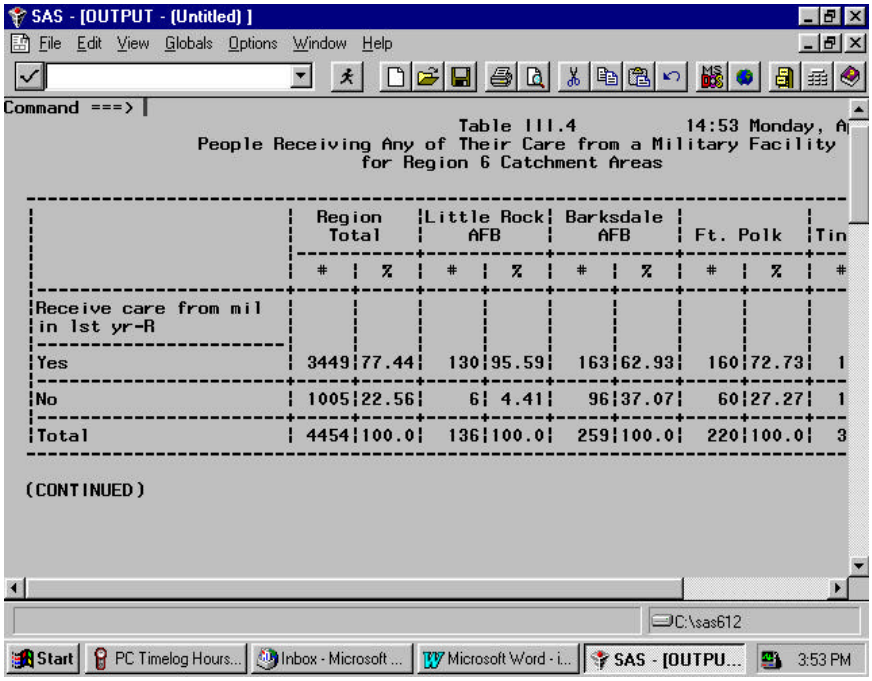


Table III.4

People Receiving Any of Their Care from a Military Facility
for Region 6 Catchment Areas

	Region Total		Little Rock AFB		Barksdale AFB		Ft. Polk		Tinker AFB	
	#	%	#	%	#	%	#	%	#	%
Receive care from mil in lst yr-R										
Yes	3449	77.44	130	95.59	163	62.93	160	72.73	192	61.34
No	1005	22.56	6	4.41	96	37.07	60	27.27	121	38.66
Total	4454	100.00	136	100.00	259	100.00	220	100.00	313	100.00

(CONTINUED)

	Altus AFB		Ft. Sill		Ft. Sam Houston		Ft. Hood		Dyess AFB	
	#	%	#	%	#	%	#	%	#	%
Receive care from mil in lst yr-R										
Yes	134	95.04	146	75.65	455	84.73	391	74.48	173	79.00
No	7	4.96	47	24.35	82	15.27	134	25.52	46	21.00
Total	141	100.00	193	100.00	537	100.00	525	100.00	219	100.00

(CONTINUED)

	Sheppard AFB		Laughlin AFB		Lackland AFB		NH Corpus Christ		Vance AFB	
	#	%	#	%	#	%	#	%	#	%
Receive care from mil in lst yr-R										
Yes	153	85.00	96	100.00	296	84.09	275	73.53	113	95.76
No	27	15.00	.	.	56	15.91	99	26.47	5	4.24
Total	180	100.00	96	100.00	352	100.00	374	100.00	118	100.00

(CONTINUED)

	Brooks AFB		Goodfellow AFB		Kelly AFB		Randolph Air Force Base		Out/Area-Reg 6	
	#	%	#	%	#	%	#	%	#	%
Receive care from mil in lst yr-R										
Yes	69	98.57	88	95.65	119	97.54	139	97.20	157	43.13
No	1	1.43	4	4.35	3	2.46	4	2.80	207	56.87
Total	70	100.00	92	100.00	122	100.00	143	100.00	364	100.00

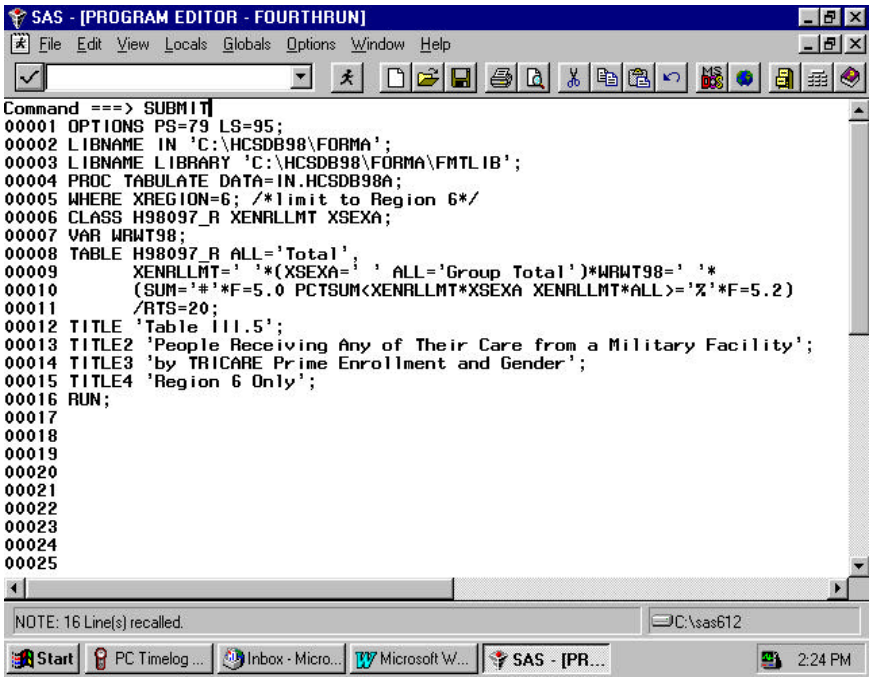
The statistic N is included with PCTN to make it easier to verify that the denominator definitions have been set up properly. After you check to see that the percentages are accurate, the N statistic can be removed. Note that the output for Table III.4 is unweighted. The N statistic (and PCTN statistic) is always unweighted even if a WEIGHT statement is included.

Similarly, if you want to look at the percentage of TRICARE enrollees (and non-enrollees) by gender who answered yes to question 97, this would be a row percentage. To calculate a row percentage, the denominator definition must include all *class* variables that define the c*olumn*. Your TABLE statement would look like this:

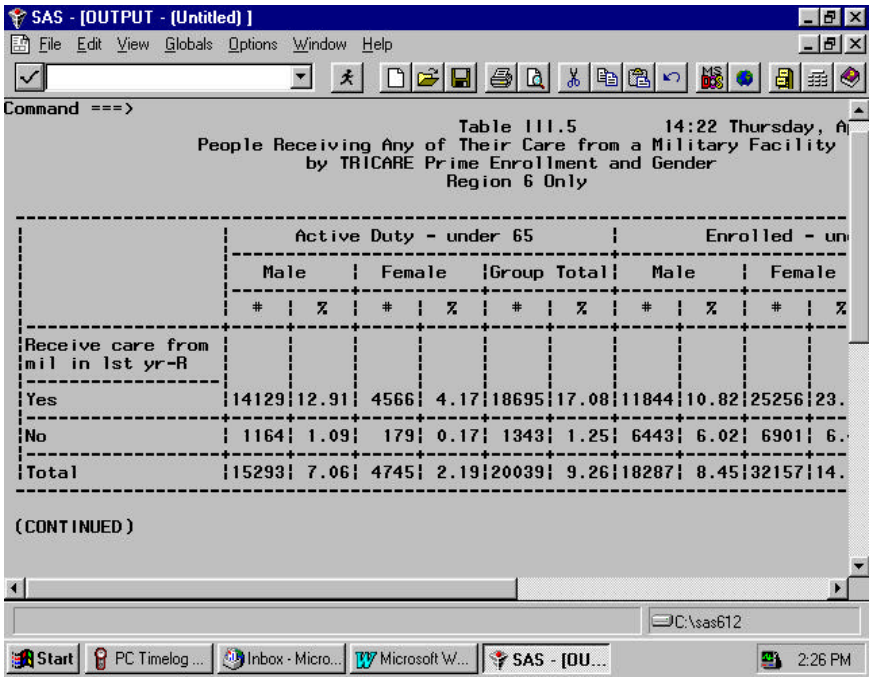
```
TABLE H98097_R ALL='Total',
      XENRLLMT *(XSEXA=' ' All='Group Total')*
      PCTN<XENRLLMT*XSEXA XENRLLMT*ALL>='Percent';
```

Notice that there are no parentheses used in the denominator definition. Because parenthetical groupings are not allowed in the denominator definition, all crossings and concatenations must be included. As noted above, the N and PCTN statistic are unweighted counts of CLASS variables. If you want to produce a weighted count and percentage for this table, you would include WRWT98 (the 1998 weight variable) as an analysis variable in the VAR statement and in the column crossing of the TABLE statement; the statistics to be generated should be specified as SUM and PCTSUM. A program and output to demonstrate weighted row percentages appears in Table III.5.

The following screen shows the new program typed into the Program Editor.



These commands produce the following output.



Here, as above, the SUM statistic is included to help determine the accuracy of the denominator definition.

Additional information about running SAS is available from the SAS Institute. Please consult the appropriate manuals for more detailed information.

See Table III.5 to view the entire table.

Table III.5

People Receiving Any of Their Care from a Military Facility
by TRICARE Prime Enrollment and Gender
Region 6 Only

	Active Duty - under 65						Enrolled - under 65					
	Male		Female		Group Total		Male		Female		Group Total	
	#	%	#	%	#	%	#	%	#	%	#	%
Receive care from mil in lst yr-R												
Yes	14129	12.91	4566	4.17	18695	17.08	11844	10.82	25256	23.08	37099	33.90
No	1164	1.09	179	0.17	1343	1.25	6443	6.02	6901	6.45	13344	12.46
Total	15293	7.06	4745	2.19	20039	9.26	18287	8.45	32157	14.85	50444	23.30

(CONTINUED)

	Not enrolled - under 65						Not enrolled - 65 or over					
	Male		Female		Group Total		Male		Female		Group Total	
	#	%	#	%	#	%	#	%	#	%	#	%
Receive care from mil in lst yr-R												
Yes	11490	10.50	9778	8.93	21269	19.43	20089	18.36	12288	11.23	32377	29.58
No	18261	17.06	32852	30.68	51113	47.74	28763	26.86	12503	11.68	41266	38.54
Total	29751	13.74	42630	19.69	72381	33.43	48852	22.56	24791	11.45	73643	34.01

How to Make a Table Using SPSS

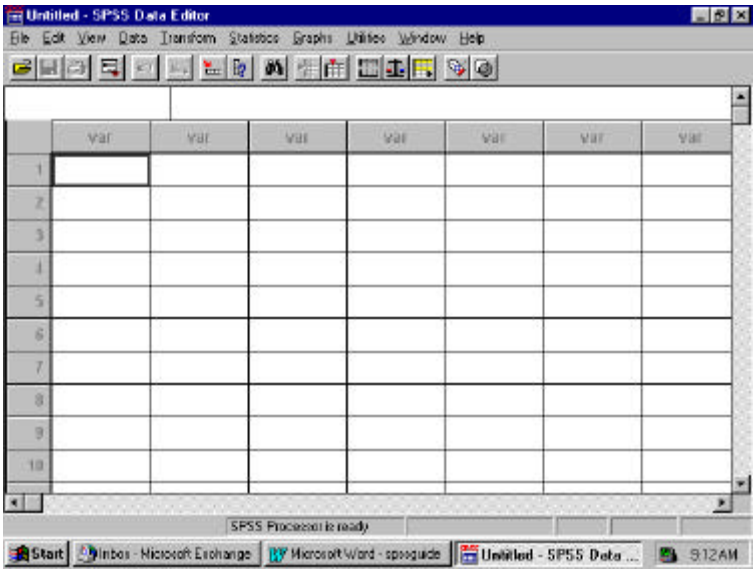
The 1998 HCSDB Form A dataset is in an SPSS format. SPSS is a computer software system used for data management, summarization, and analysis. SPSS can be run interactively, using menus, or in batch mode, using syntax commands. This guide instructs users on how to use SPSS dialog boxes to:

- Construct new variables
- Recode existing variables
- Select cases for analysis
- Weight cases for analysis
- Create customized tables

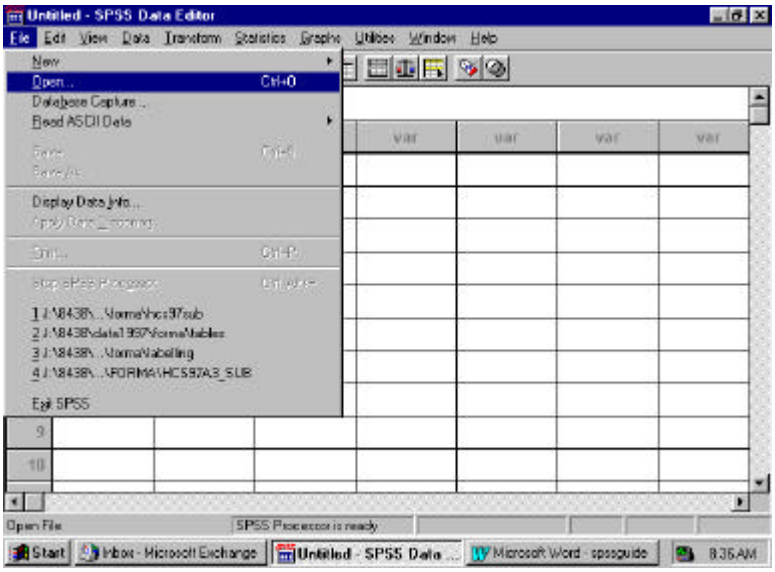
As you use the dialog boxes, you generate syntax automatically. This syntax may be pasted into a syntax file for future use or for modification.

Locating and opening the data file

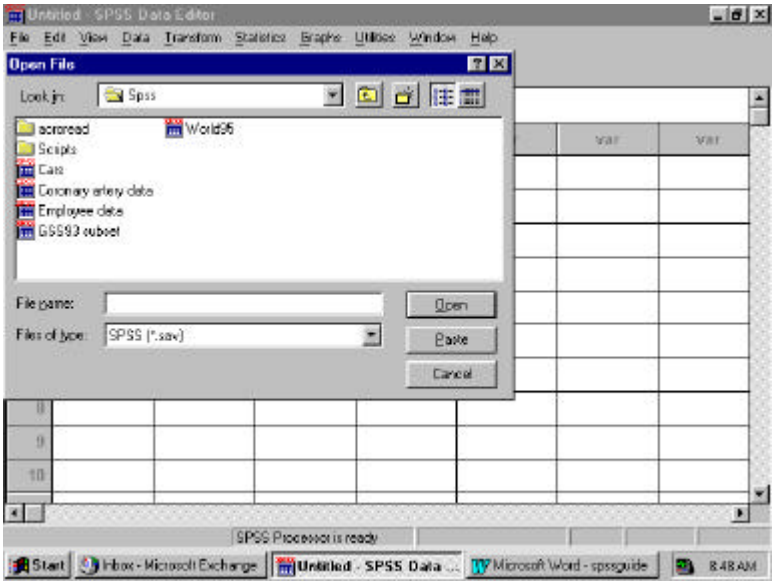
To begin an SPSS session, double click on the SPSS icon on your desktop. The Data Editor window will open and present a blank spreadsheet like the following screen:



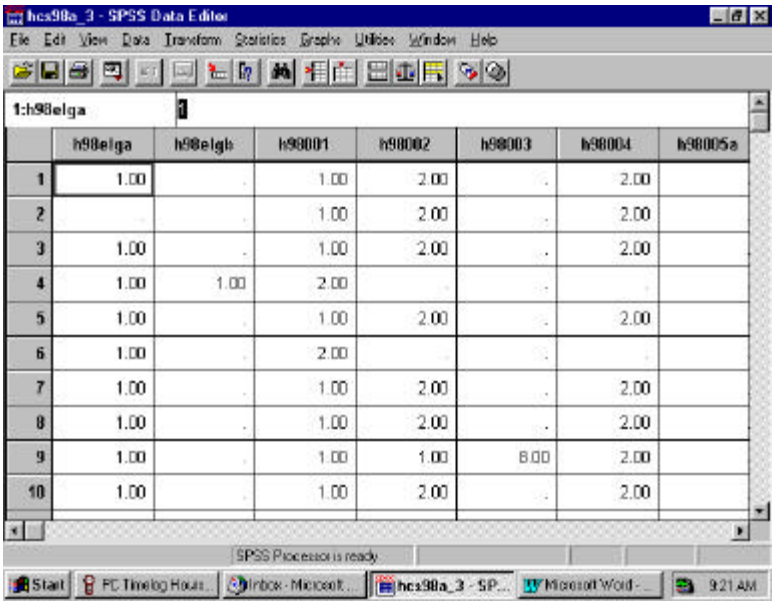
Click on **File** in the upper left corner to open the following menu:



Select the **Open** option or choose a file from the list displayed. **Open** produces the following screen:



If the file is not in this directory, navigate through your folders until you locate it. Mark the file and click **Open**. You will be returned to the spreadsheet Data Editor with the file on view. The 1998 HCSDB dataset has been opened and is displayed below.



Constructing new variables

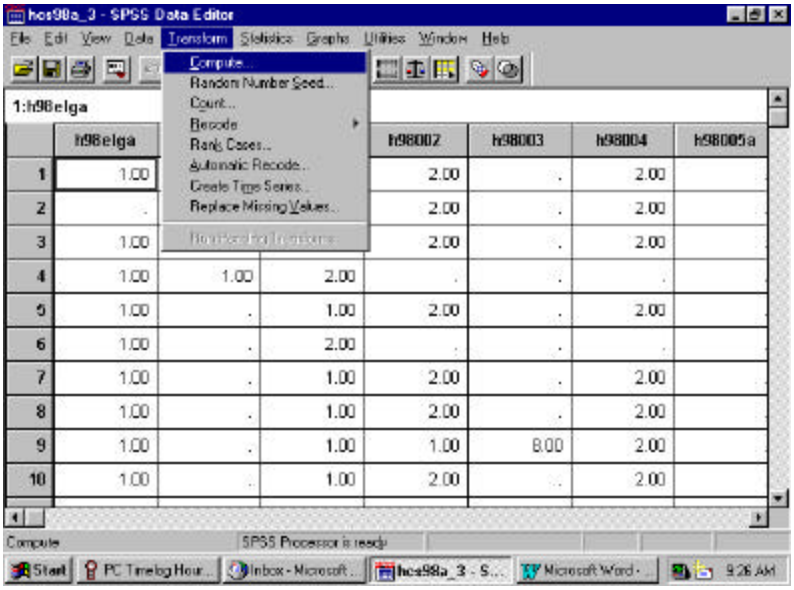
Data can be evaluated from many different aspects. It is sometimes useful to build new variables from combinations of the existing ones and to examine their distributions.

For example, the variable in the file for beneficiary group at the time of sampling is called **bfgroup,** and the variable for sex is **xsexa**. The value **1** for **bfgroup** indicates that the individual is on active duty. The relationships for constructing a new variable of active duty by sex are:

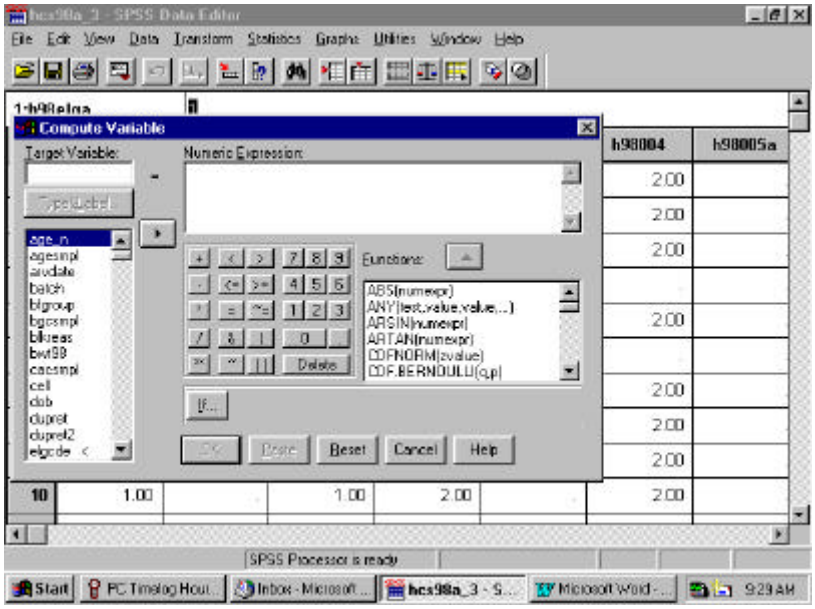
Active-duty-males: ***xsexa=1 and bfgroup=1***

Active-duty-females: ***xsexa=2 and bfgroup=1***

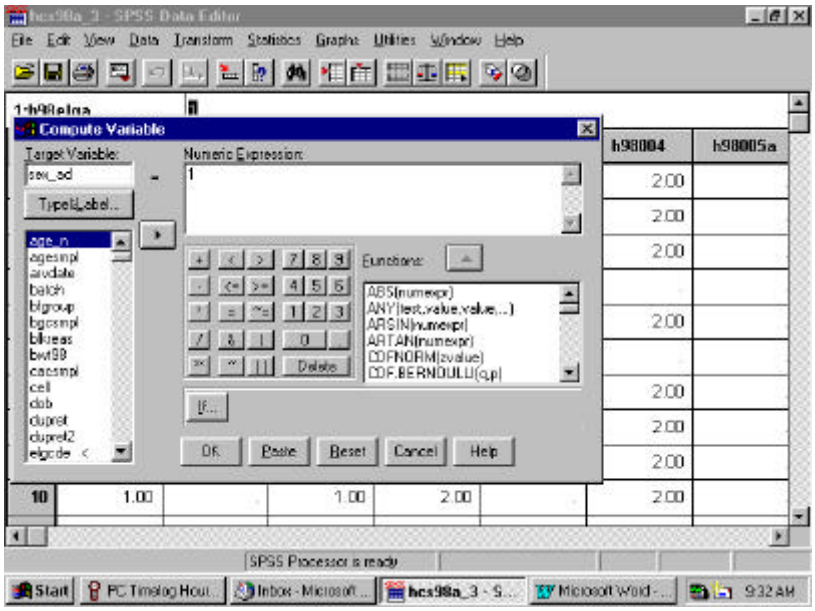
Open the **Transform** menu and select **Compute** as in the following:



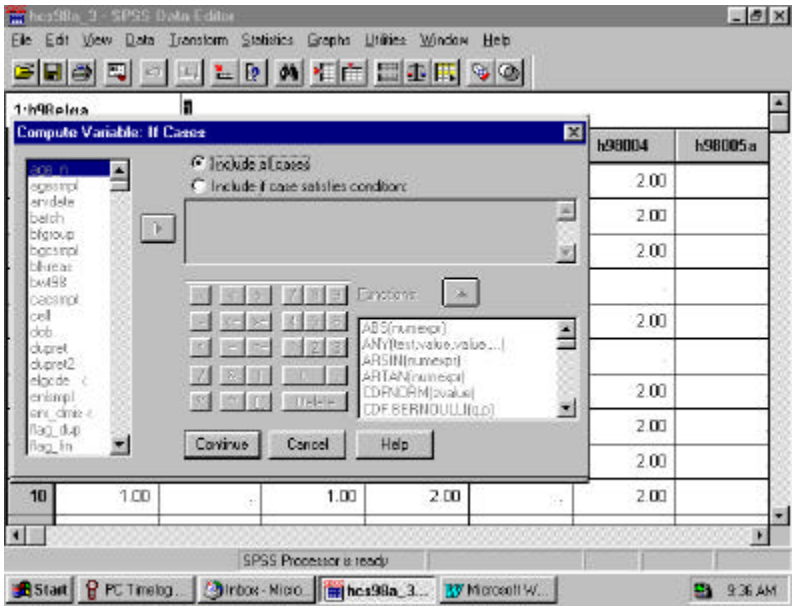
The following dialog box will open:



You can build the new variable in two steps to express the two conditions. The first task is to give the new variable a name and its first value. Enter the **Target Variable** slot and name the new variable ***sex_ad***. Next, assign the value **1** to **sex_ad** by entering it into the slot for **Numeric Expression.** Your screen should look like the following:

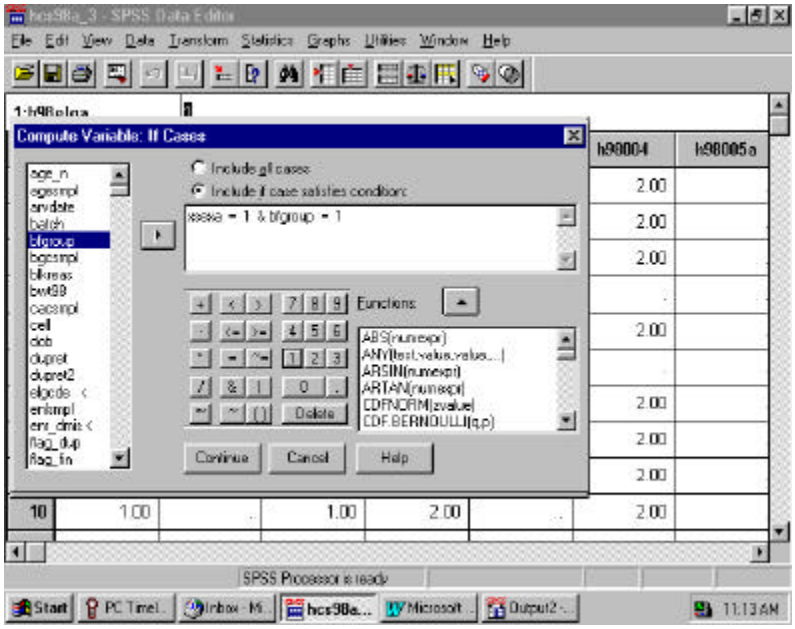


Once you we have assigned the value **1** to **sex_ad,** you can build the condition that qualifies the assignment. Click on **If..**and open the following dialog box:

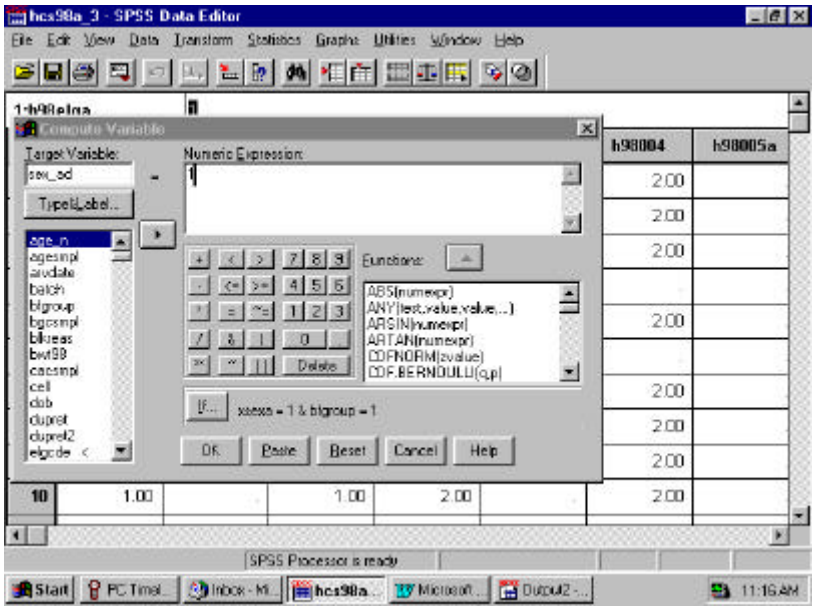


Click on the circle indicating **Include if case satisfies condition,** and the black dot will move to that circle. The slot underneath will open, ready for your input. Build the "if" condition. Write it directly into the slot or move the elements into the slot from the given options. Add the elements **xsexa = 1 & bfgroup = 1**.

The screen should resemble the following:



Click on **Continue** and return to the previous screen, which will now look like this:



Your condition will be written next to the **If** button. Click on **OK** to exit the dialog box, and the variable **sex_ad** will be created with its value set to **1**.

The next step is to build the second condition for the new variable, which will set it to the value **2**. Reopen the **Compute** dialog box. The commands you just gave still appear in the dialog box. Simply assign the value **2** to **sex_ad**, press **If**, and enter **2** for the value of **xsexa** . Click **Continue,** and finish with **OK**. The condition, ***if xsexa = 2 and bfgroup = 1,*** will be added to the new variable **sex_ad**.

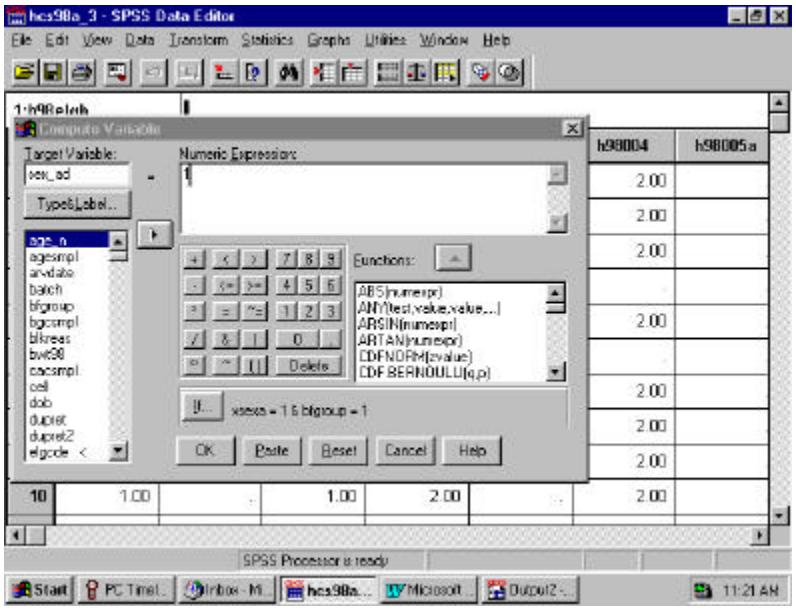
Once you have created a new variable, you may want to add it permanently to the dataset. The new variable is computed for each case in the file and added to the view in the Data Window after the last variable in the dataset. The variable name is the column heading.

Since the HCSDB data set carries Read-Only status to protect it from corruption, changes to it cannot be saved. At the end of the day, when the work session ends and you exit SPSS, the file will revert to its former status and the new variables will be lost. The solution is to save the dataset under a new name when you exit. Choose the **Save As** option on the **File** menu, and you will be prompted to name the file and to save it in a folder of your choosing. Give the file a new name and save it. Open the new expanded file anytime for processing.

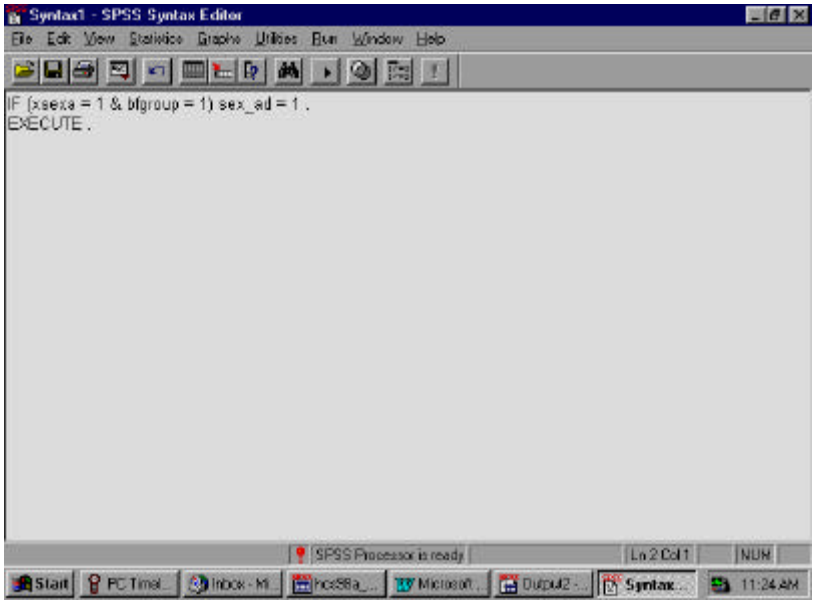
Suppose you do not want to use up your disk space for expanded copies of the dataset. Another option is to save the **syntax** you have generated in a file that can be run as it is needed. Syntax is a written instruction generated by the commands you give in a dialog box. These “sentences” can be saved in a file and executed when needed. This is the **batch mode** of processing syntax commands. Syntax files take up very little space.

Experienced SPSS programmers, who have mastered SPSS syntax, often prefer to work only in batch mode. This option is available to users who have not mastered the syntax language. You can ***paste*** the commands, generated interactively in the dialog box, onto a syntax file.

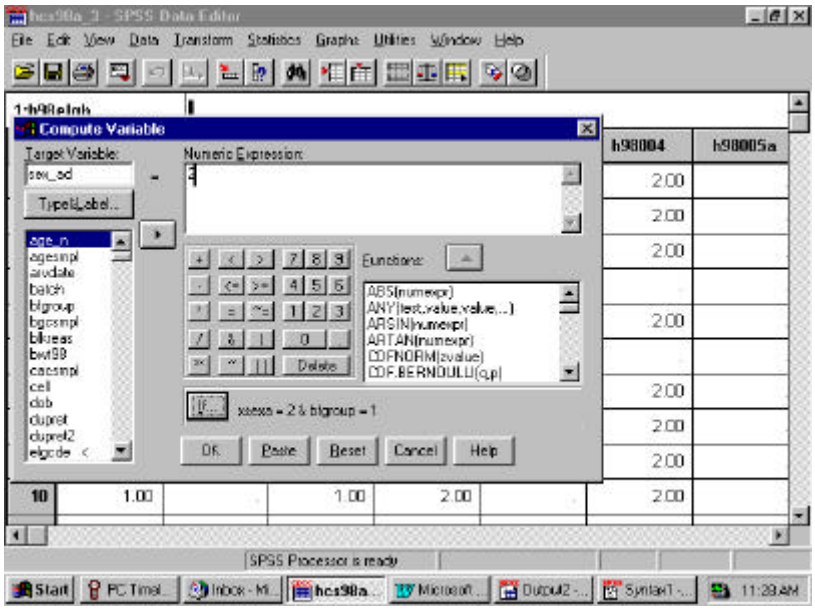
Recall the compute example for the new variable **sex_ad**. The screen below is the result of assigning **1** to **sex_ad** according to an **If** condition. You clicked on **OK** to set the value. Returning to the screen and clicking on **Paste** writes the command to a syntax file.



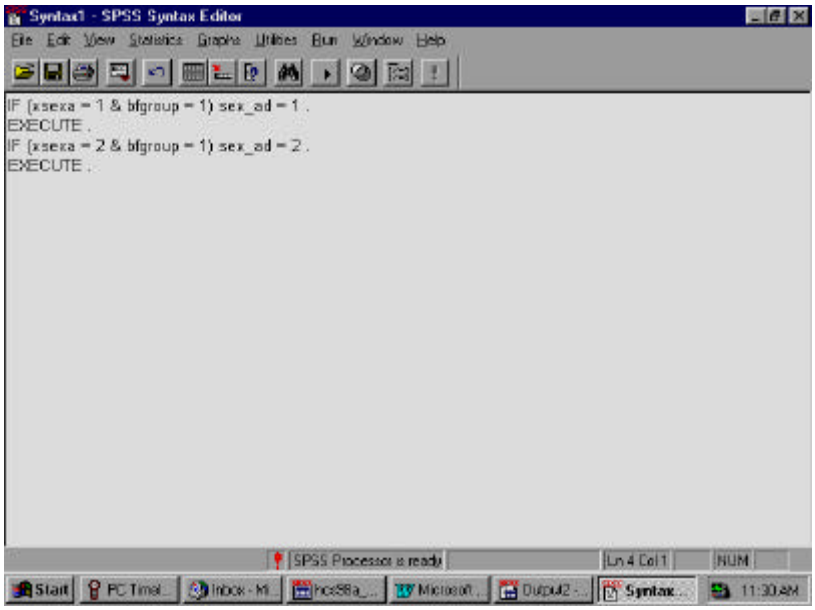
Click on the **Paste** button, and the syntax window below will open with the syntax written in it.



Now return to the compute dialog box.



Assign the value **2** to **sex_ad** as in the diagram above. Select **Paste,** and these commands will be appended to the syntax file.



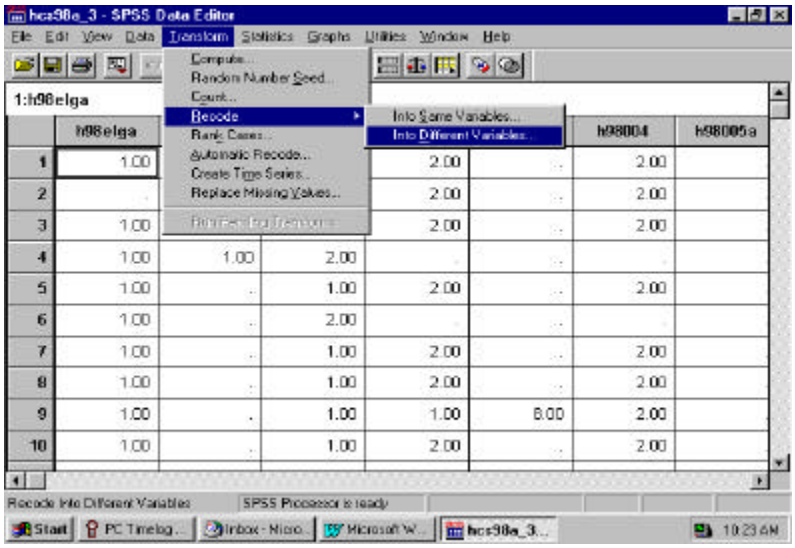
The results appear in the screen above. SPSS gives default names to syntax files, such as Syntax1, Syntax2, etc., as they are created. It is a good idea to save the syntax, re-naming the file using the **Save As** option on the **File** menu. Use a name that has some meaning to you, e.g., **New_computes**. The file will automatically receive the suffix .**sps**.

Another option for adding new variables to the dataset is to **Recode** existing variables **into** new variables. A common example involves **grouping** an age variable into age categories as shown below, using the variable **age_n,** which exists on our dataset. **Age_n** is coded in years from 18 to 98, which can be grouped into four age categories:

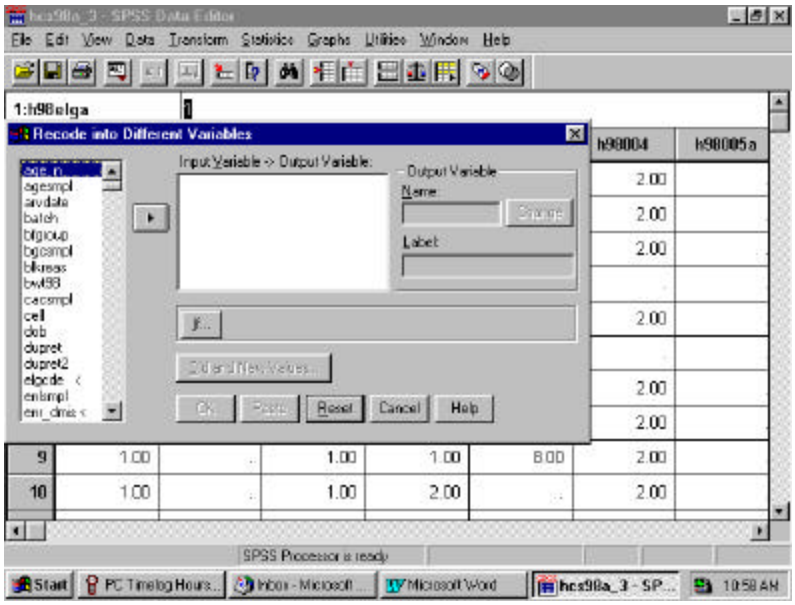
- 18 to 34 = 1 - label: “young adult”
- 35 to 49 = 2 - label: “mature adult”
- 50 to 65 = 3 – label: “middle-age”
- 66 to 98 = 4 – label: “senior citizen”

The new variable is called **age_grp**.

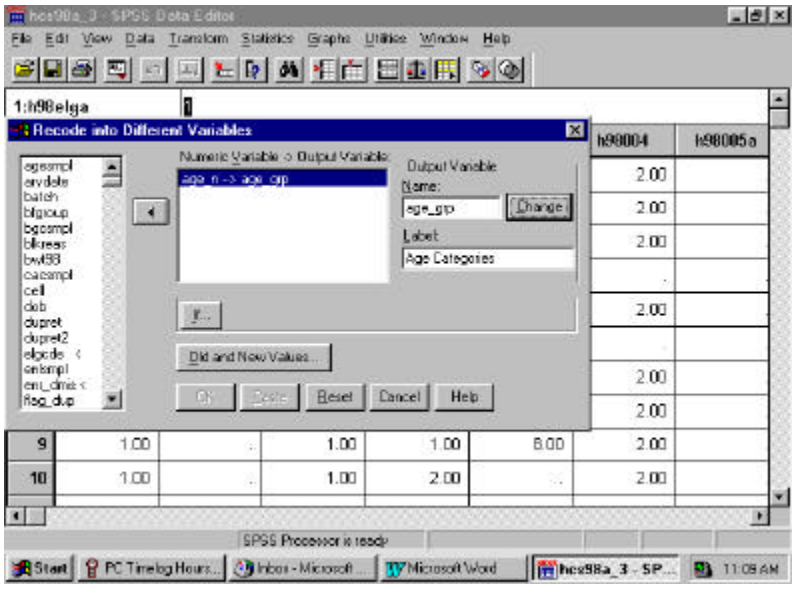
From the **Transform** menu, choose **Recode** and **Into Different Variables** as pictured below:



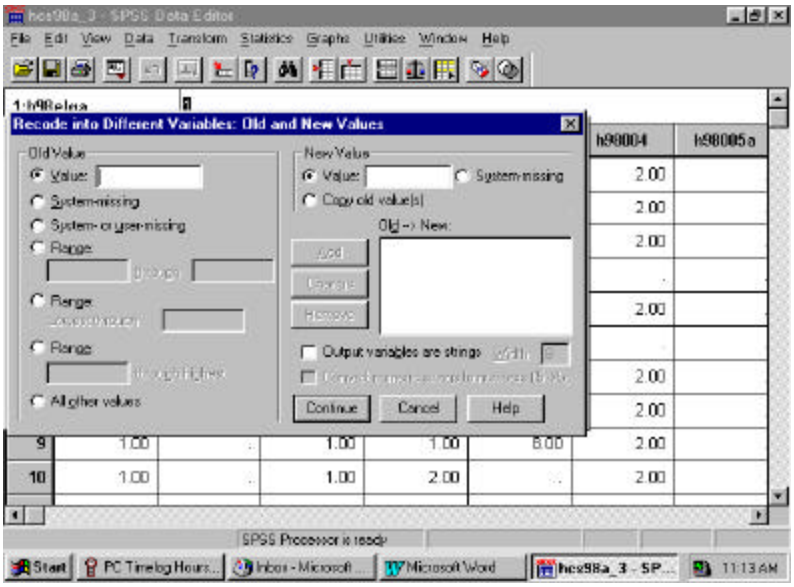
The following dialog box will open:



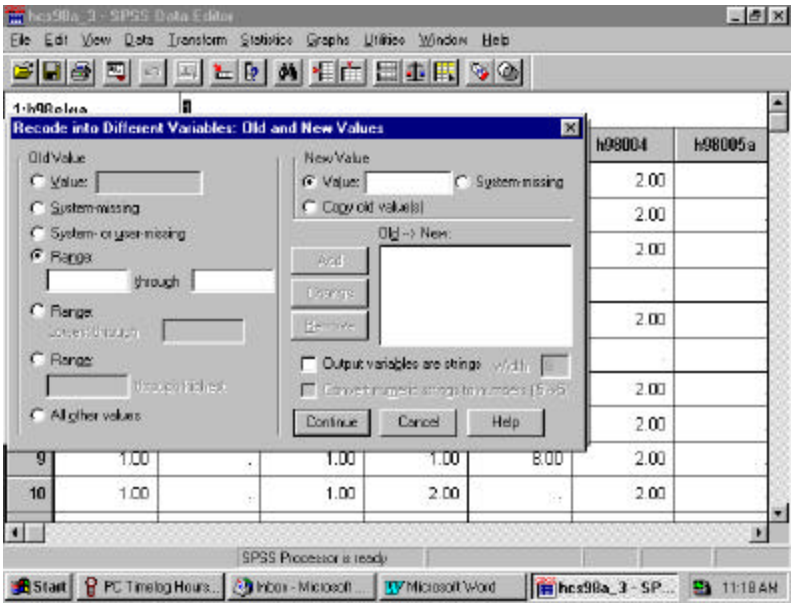
Move **age_n** from the variable list on the left to the box labeled **Input Variable -> Output Variable**. In the **Name** slot, enter the new variable name **age_grp**. Enter **Age Categories**, the variable label, in the **Label** slot. Click on **Change**. The dialog box should look like the one below.



Click on **Old and New Values,** and the following dialog box will open:

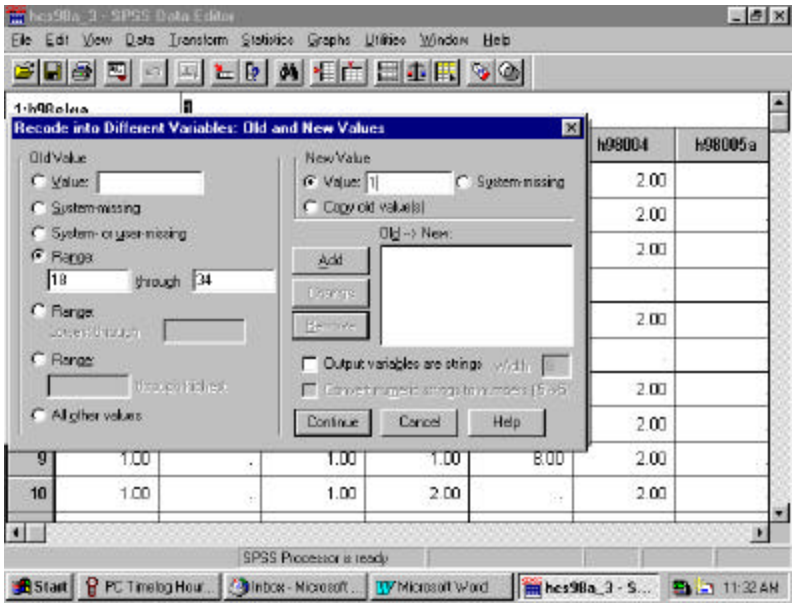


The next step in grouping the age variable is to specify the existing values of **age_n** to be recoded. To do this, click on the **Range** circle under **Old Value**.

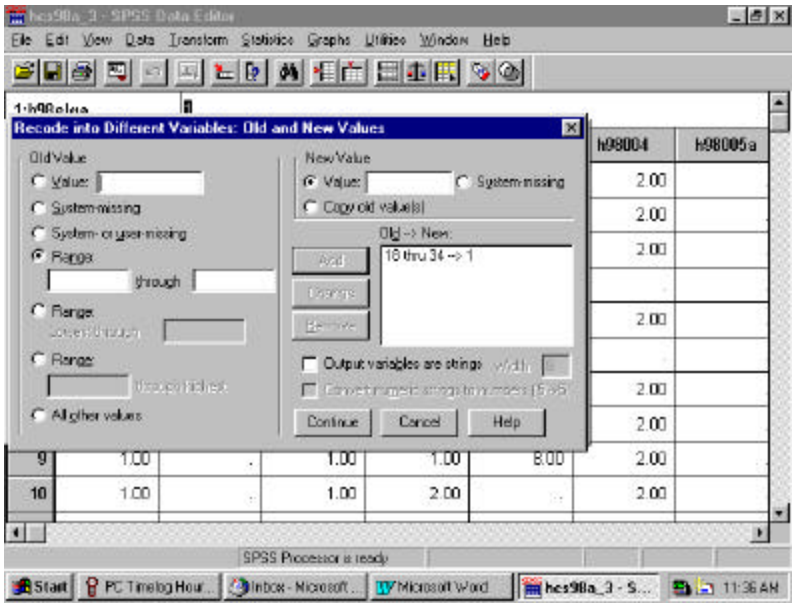


Once the appropriate slots are open, you have four ranges to enter.

First, enter 18 **through** 34 in the slots provided under **Range**. Next, enter the value **1** in the **Value** slot under **New Value**. **Add** is now illuminated.

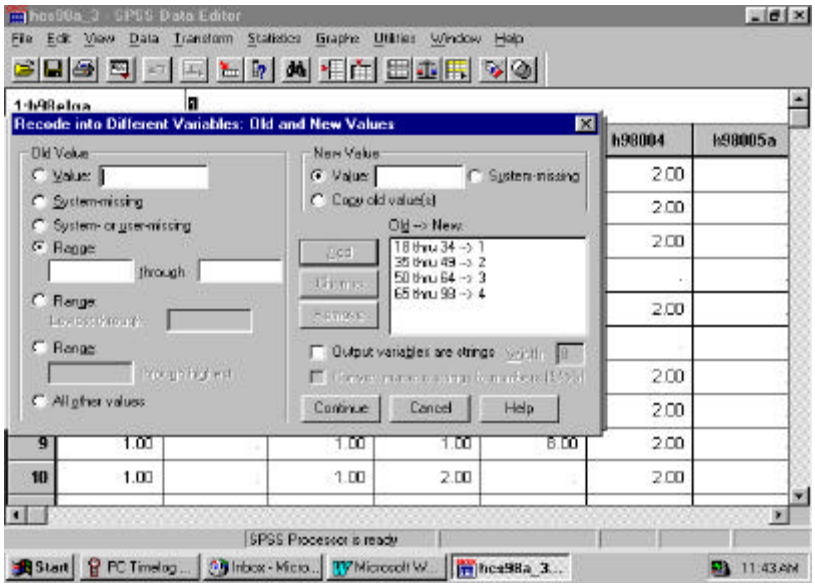


Clicking on **Add** produces the following result:

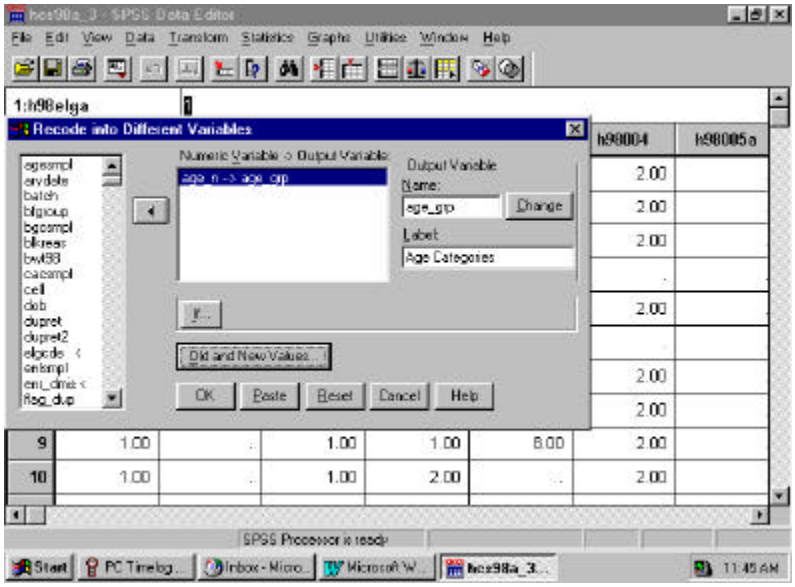


The specified range appears in the box labeled **Old -> New,** and the **Range** and **Value** slots have been cleared to permit additional entries.

The three remaining ranges are built in the same manner, adding each specification, until the dialog box looks like the one below.



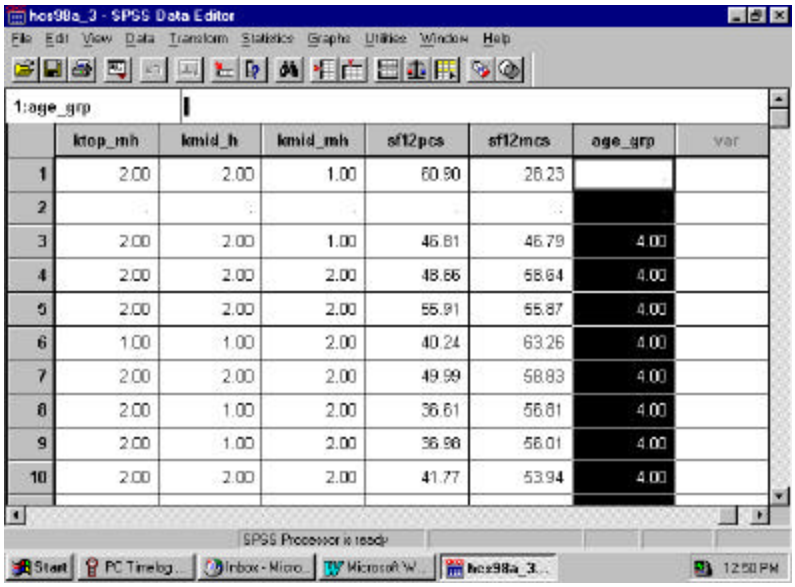
Click on **Continue** and return to the previous screen**.**



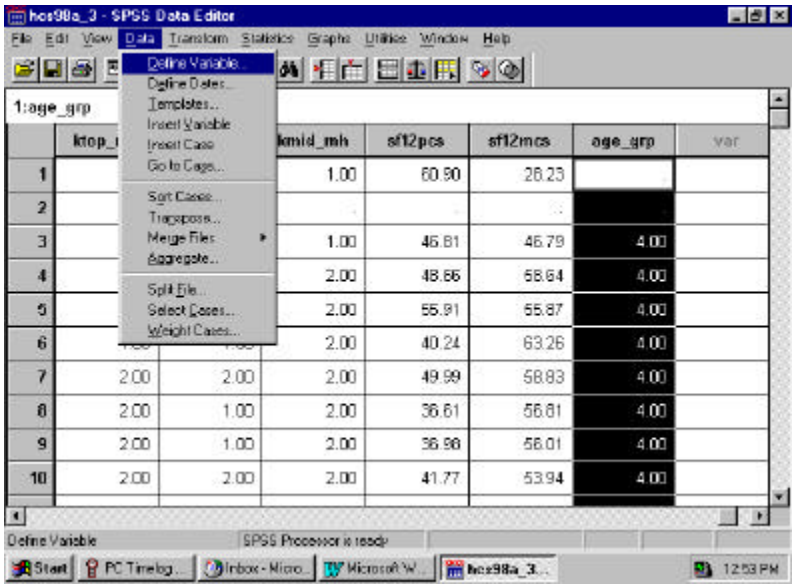
Click on **OK** to exit the screen. The new variable **age_grp** has been created. The **Recode** syntax can be pasted to a syntax file**.**

The final task is to create the value labels for the new variable **age_grp**. Labeling variables makes output from statistical reporting procedures much clearer and more elegant.

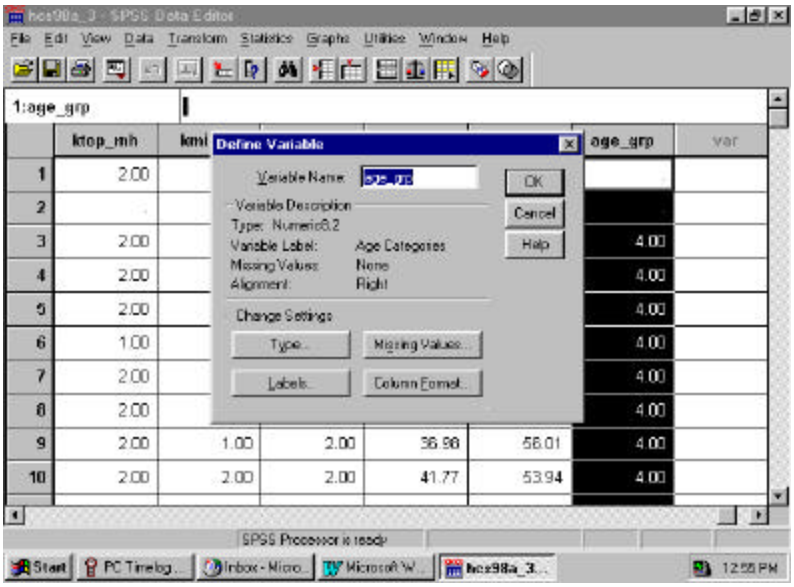
In the **Data Window,** go to the column for the new variable **age_grp** and double-click in the gray area containing the variable name. The entire column will darken indicating that it has been selected.



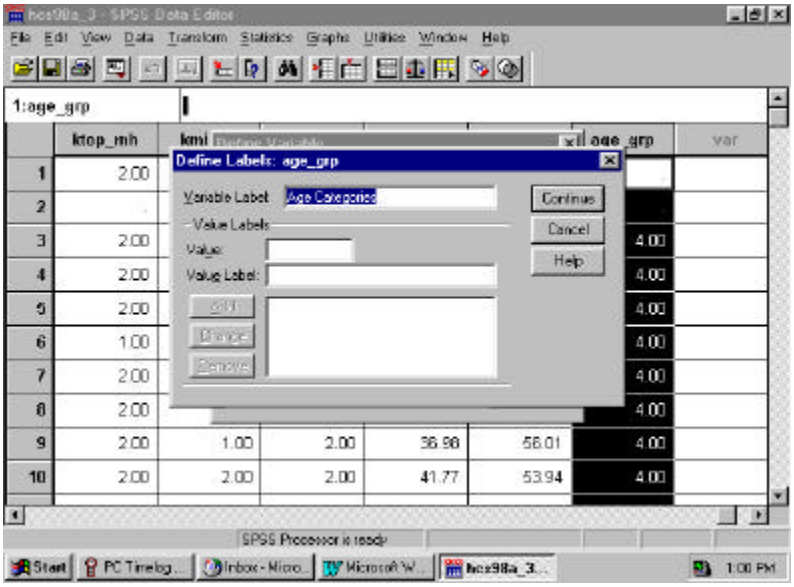
Click on the **Data** menu at the top of the screen to open the following dialog box.



Click on **Define Variable** to get to the following dialog box:

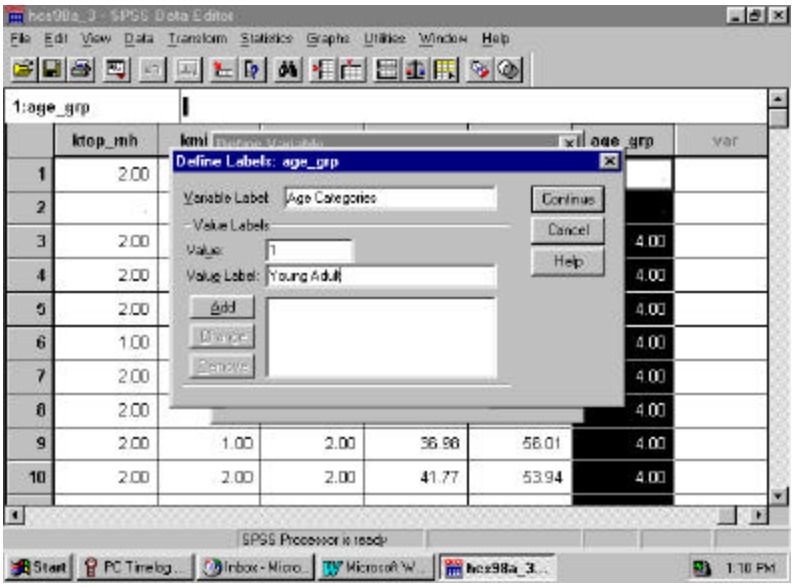


The characteristics of **age_grp** are displayed in this dialog box. Information about the variable type, its label, and its missing values appears here. Click on **Labels** to get to the following screen**:**

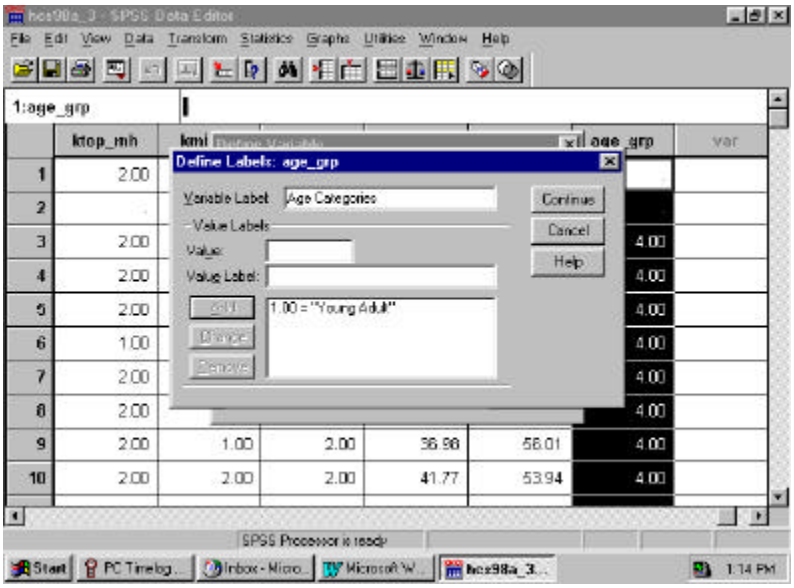


In the slot marked **Variable Label** is the label **Age Categories,** which was specified during the **Recode** process. If there is no label for the variable, enter one in this slot.

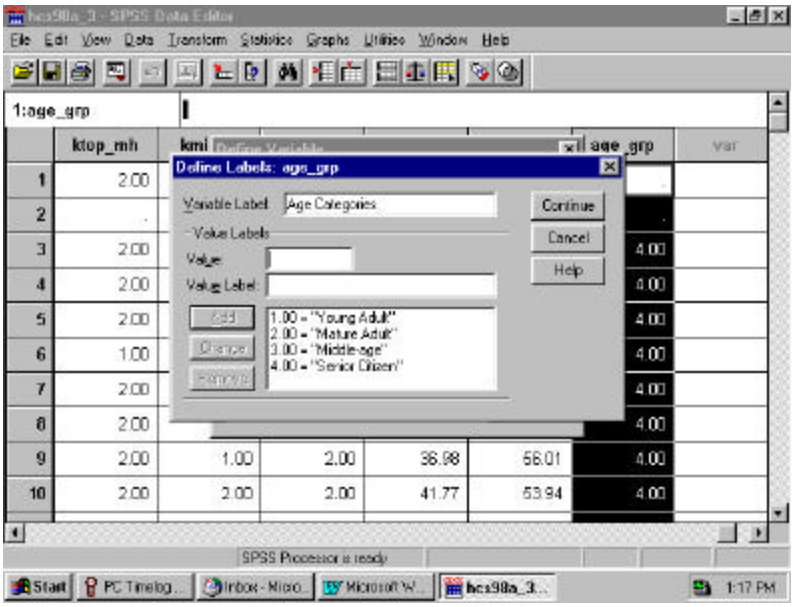
You can then begin to label the *values* of **age_grp**. Enter **1** in the slot marked **Value,** and enter the label **Young Adult** in the slot marked Value Label. The screen will look like the following:



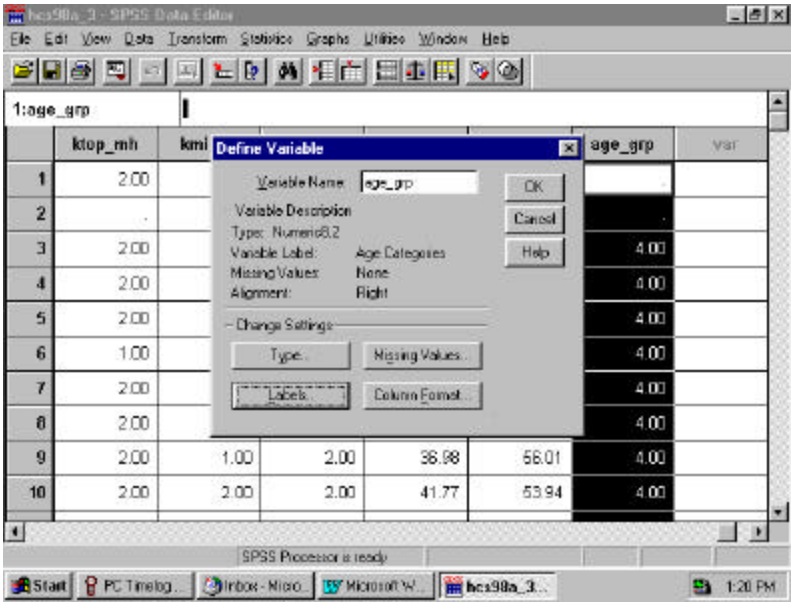
Add is now illuminated. Click on **Add** and the text of your command will appear in the central box, clearing the slots for further entries, as in the next screen.



Build the other three labels until the screen looks like the following:



Click continue, and return to the first screen**.**



Click on **OK** to exit. The labels have been added.

Limiting the Number of Variables

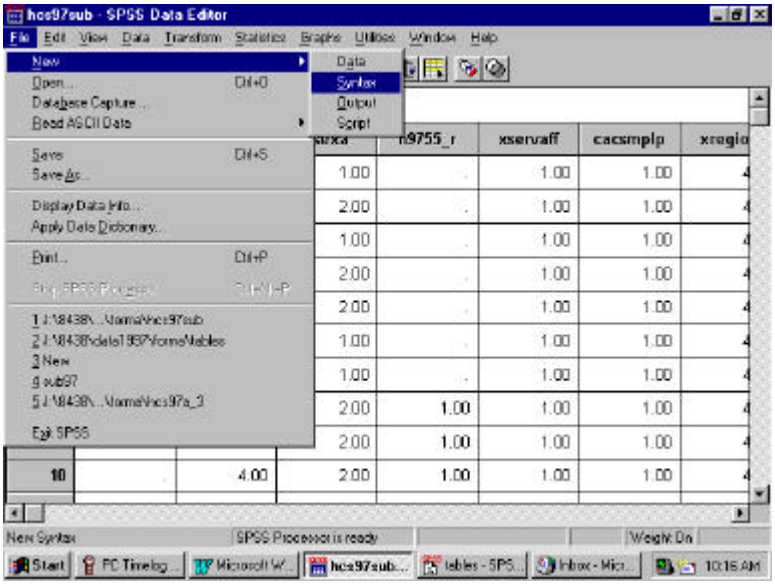
The HCSDB dataset contains many variables. To speed up software performance time, it may be desirable to limit the number of variables for analysis. There are ways to do this.

The first is to **Save** a subset of variables in a new file with a new name. This option is available only

through syntax. The **Keep** or **Drop** command lets you save a subset of variables. The choice of **Keep** or **Drop** is dependent on which list is shorter to write.

For example, suppose you want to run some procedures to evaluate overall satisfaction with medical services. You are also interested in the differences between military and civilian services, and in differences within these groupings by gender. Moreover, you want to look at regional differences and differences among catchment areas. You can do all the work on a subset of only seven variables, saving them in a separate file.

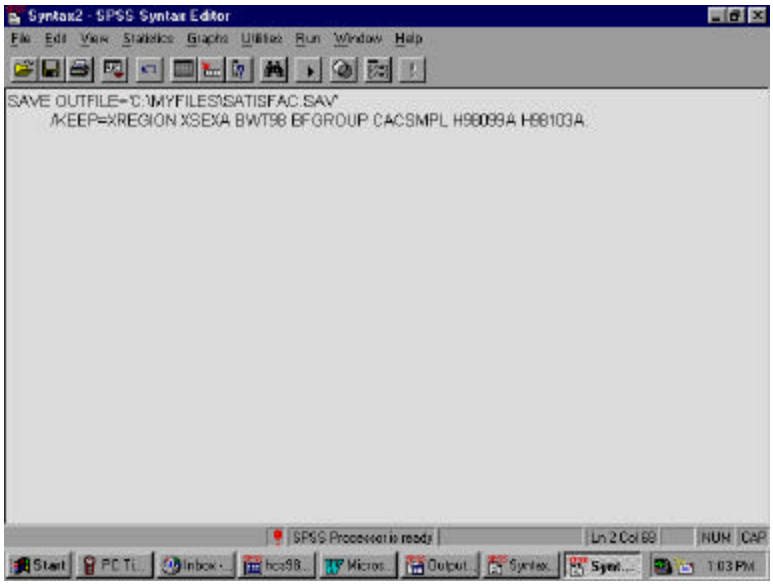
To write the syntax, open a syntax window. If you want to create a new syntax file, choose **New, Syntax** on the **File** menu as in the following:



A blank syntax window will open.

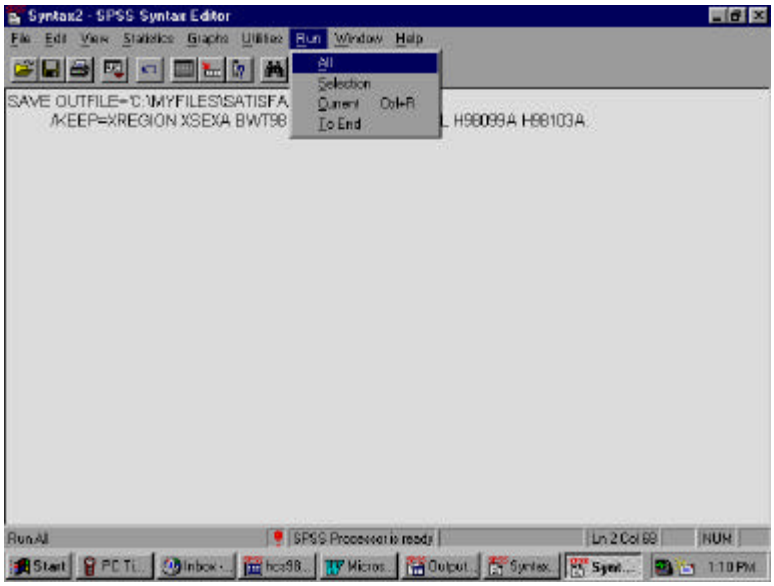
Write the following command, substituting the file name and directory specification:

Save outfile='c:\myfiles\satisfac.sav'/keep=xregion sexa cacsmpl bwt98 bfgroup h98099a h98103a., as in the following:



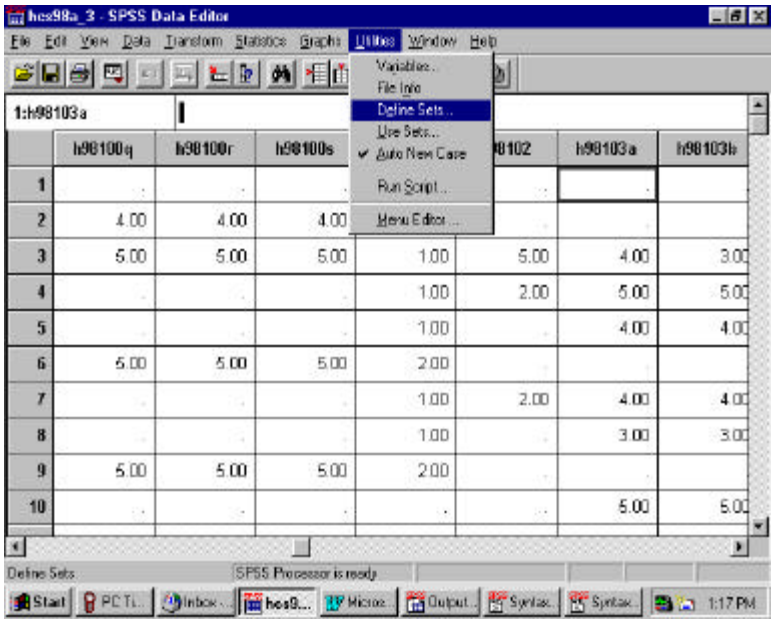
Upper case is optional. Be sure to enclose the entire file name in single quotes and to type a period at the end of the command.

Run the command by choosing the **Run** menu and selecting **All** from the choices.

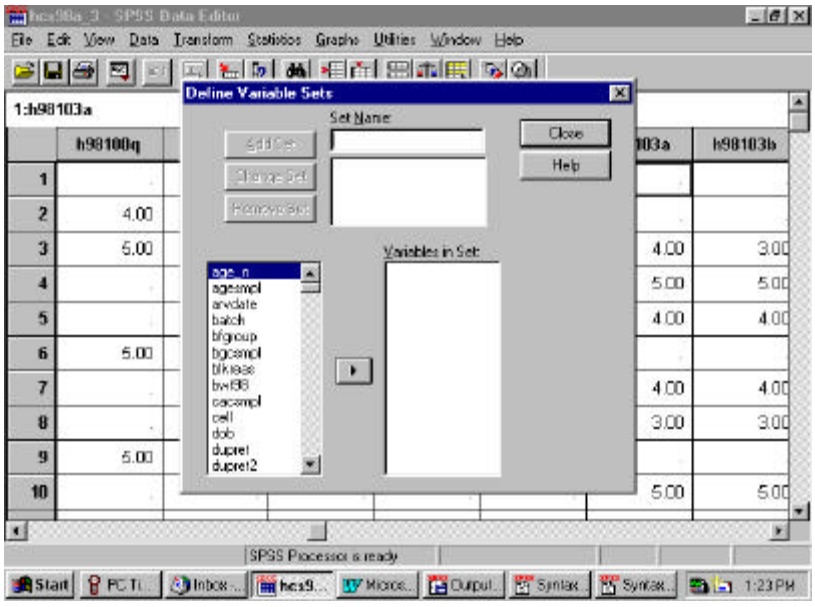


Open the new file according to the specifications at the beginning of this chapter.

The second way to limit the number of variables for analysis is to define a subset of *variables* that will appear in the dialog boxes for procedures. Using the **Utility** menu, define a subset of variables as in the following:

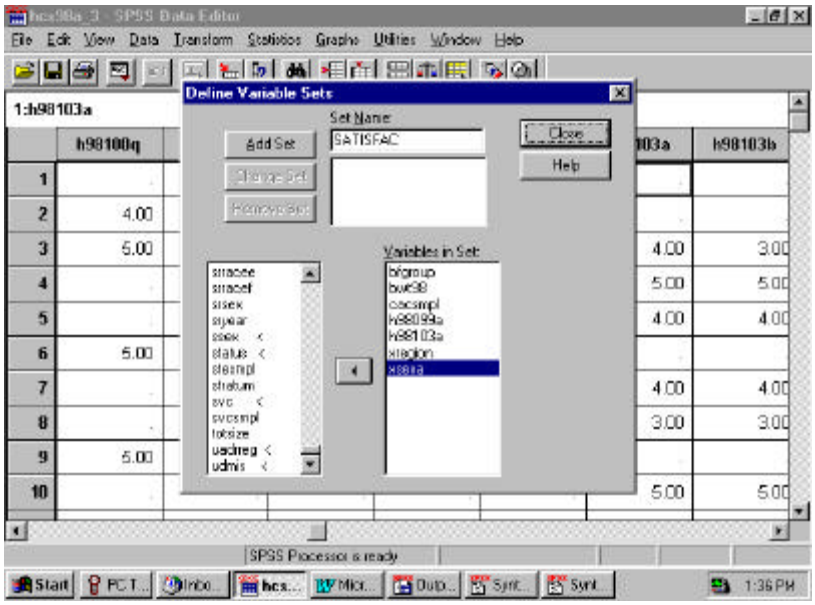


Select **Define Sets.**

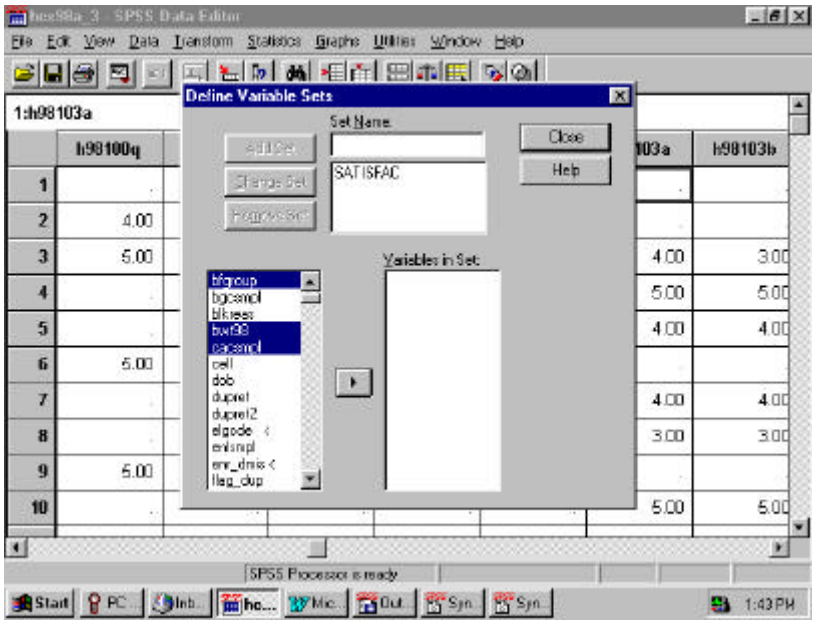


Insert a name for the subset of variables in the slot labeled **Set Name**. Move the variables you want to subset from the list on the left to the slot marked **Variables in Set**. By way of illustration, we will move the seven variables selected for the day's processing.

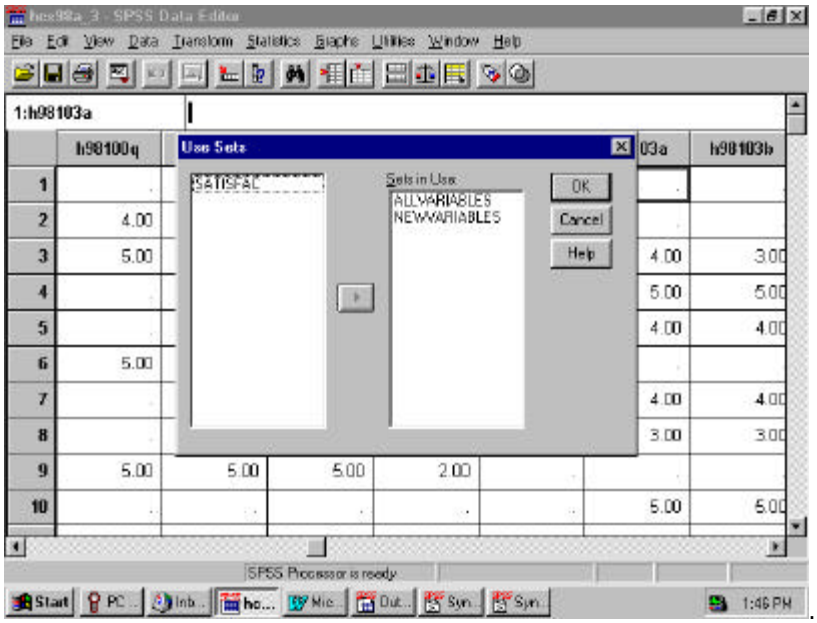
The screen should look like the following:



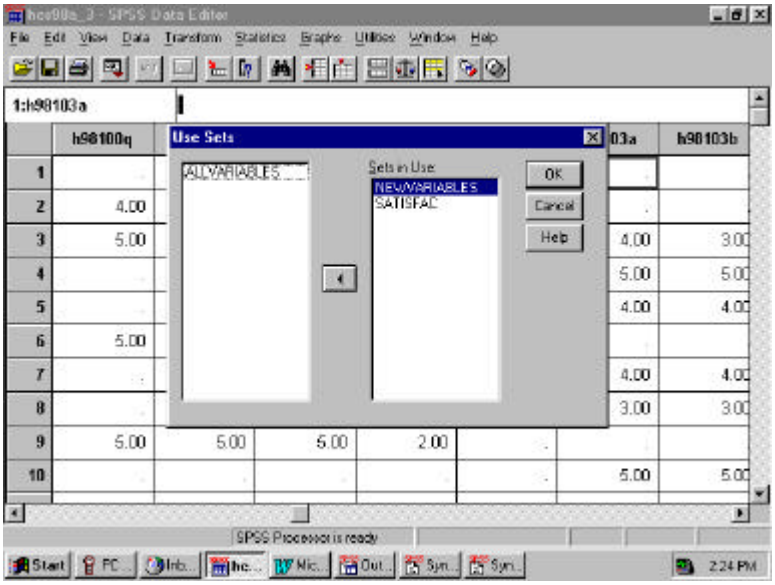
Click on **Add Set** to save the set specifications. The screen will change to the following:



The set is now available for use. To use the set, **Close** the dialog box, reopen the **Utilities** menu, choose **Use Sets…,** and receive this screen:



Move SATISFAC from the left slot to the right slot, which is labeled **Sets in Use**. Transfer ALLVARIABLES from the right to the left slot. Leave NEWVARIABLES where it is. **OK** saves this change.



Until you change this specification, only seven original variables and any new variables will appear in the dialog boxes associated with procedures.

Limiting the Number of Observations

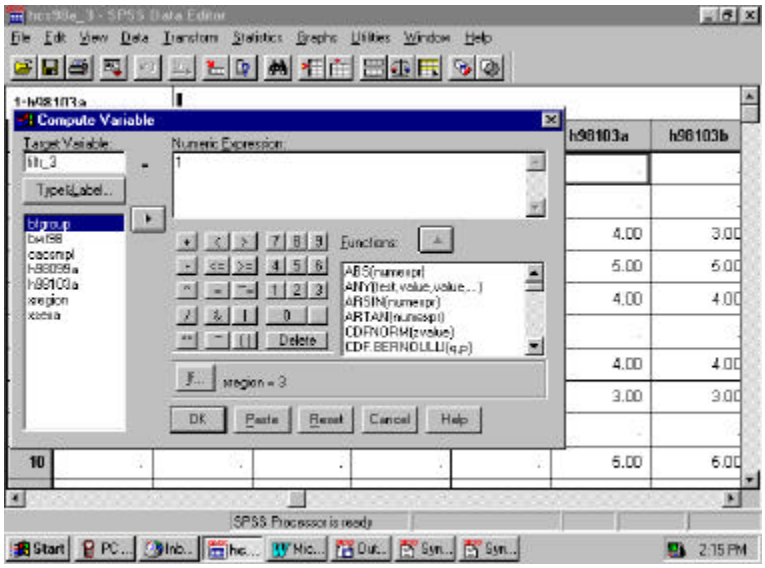
There are many ways to limit the number of observations available to statistical reporting procedures. The method illustrated here involves using **filter variables** with a menu-driven **Filter By** option. Using filters *deactivates* but does not *delete* cases from the file. A diagonal line appears next to the filtered cases in the **Data Window**.

The first task is to **compute** a filter variable for all the cases in the file.

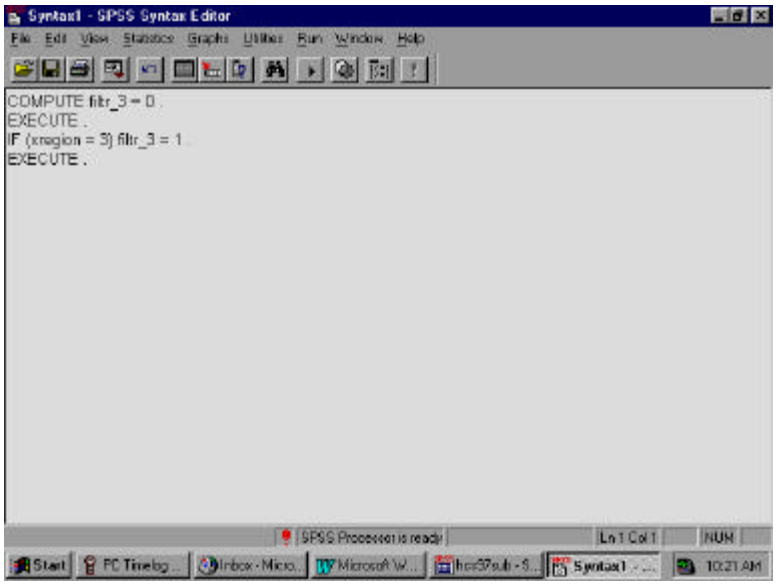
A filter variable has two values: **0** and **1**. The **1** indicates that the case will be included for procedures. The **0** flags the case for removal.

For example, suppose you want to produce a table for people who live in the Southeast, i.e., cases for which the variable **xregion = 3**. You would build a filter variable named **filtr_3**, which has the value **1** associated with the cases in the Southeast and **0** for all the other cases in the file. The logic is: if **xregion** = 3, then **filtr_3** = 1, else **filtr_3 =** 0.

The screen below shows the final step in computing the filter variable. The variable was first initialized to **0** in the same way as **0** was assigned to the new variable, sex_ad. Then, the "If" condition was built for setting the filter variable to **1**. You are now working with a subset of variables, allowing the dialog boxes to be used more efficiently**.**

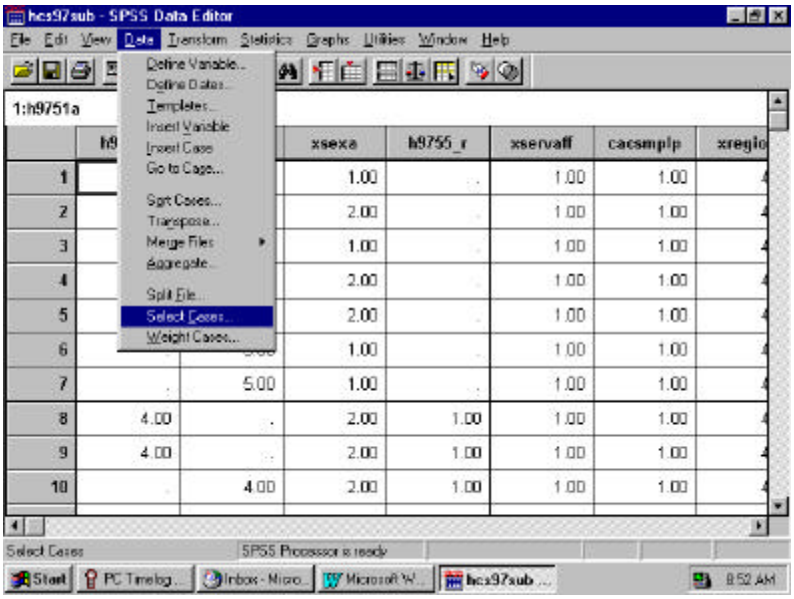


The screen that follows shows the syntax that was generated as you built the variable filtr_3**.**



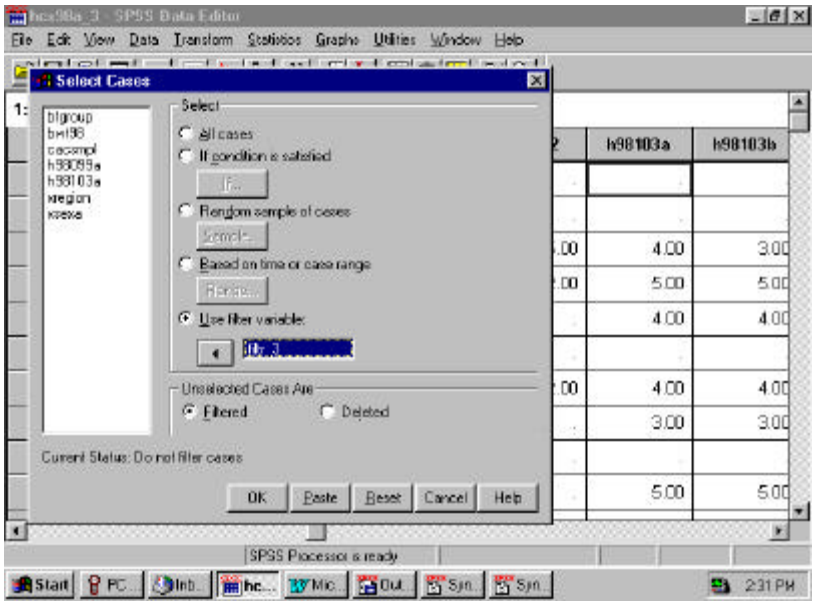
Once you build the filter variable, you can apply it for analyzing only those people from the Southeast.

Using the **Data** menu, choose **Select Cases**.

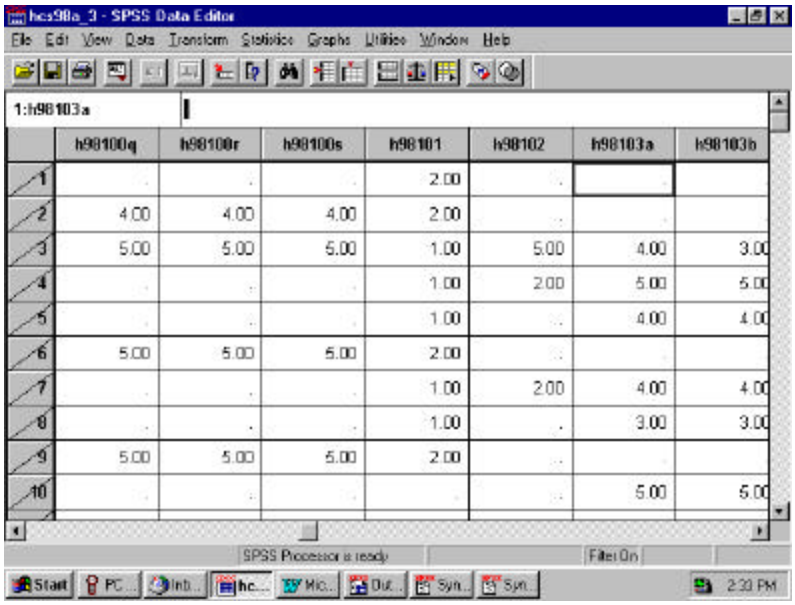


In the dialog box, check **Use filter variable**. Move the variable **filtr_3** from the variable list on the left side of the dialog box into the slot provided, as indicated below. Check that the option **Filtered** is checked under **Unselected Cases Are**. This is the default option.

Click **OK** and exit the dialog box.



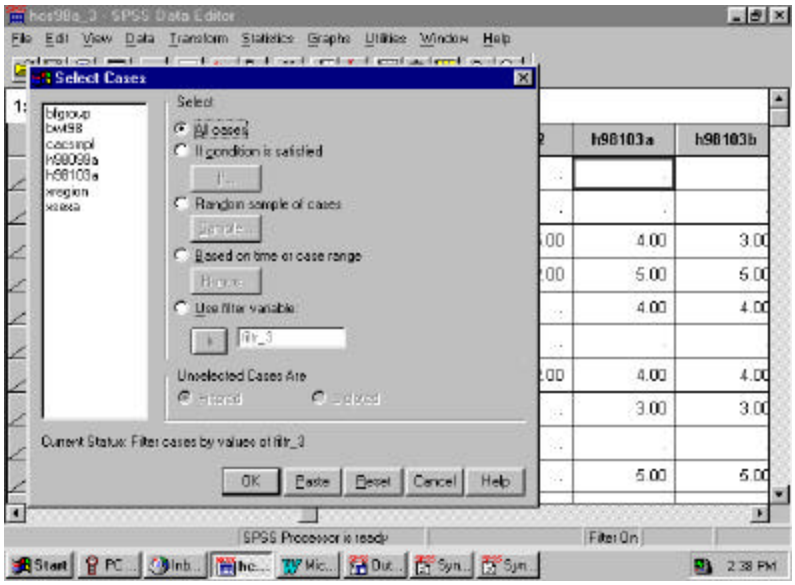
When you return to the **Data Window,** notice the slanting line next to some of the cases in the file. Those cases have been filtered out.



You can now produce tables for the subset of cases.

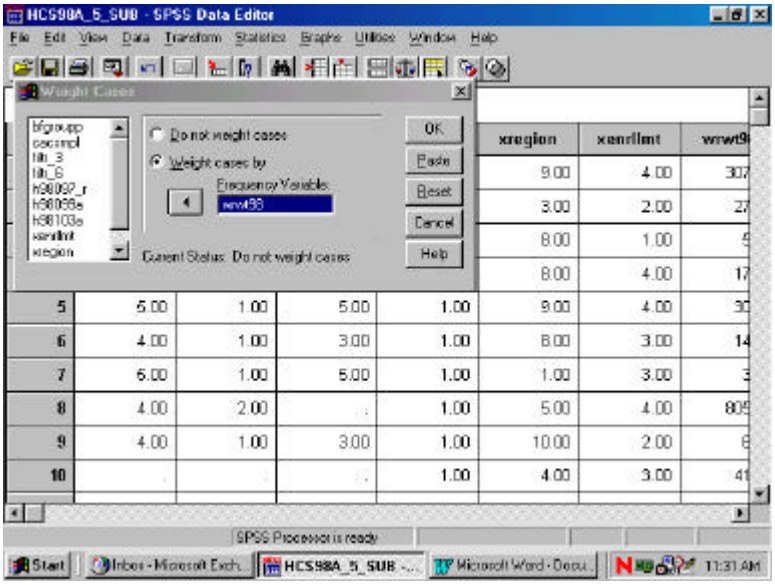
When using filter variables, it is important to check the filter status and to adjust it to fit the present need. Filtered cases are not available for procedures. Moreover, a filter is in effect until it is turned off or until another filter is activated. Check the status line at the bottom of the **Data Editor** window to see if a filter is activated. In the example above, **Filter On** is indicated on the status line. To see *which* filter is active, you must re-enter the **Select Cases** dialog box. There you can deactivate the filter or activate a new one.

To deactivate a filter, choose **All cases** and **OK** as in the screen below.

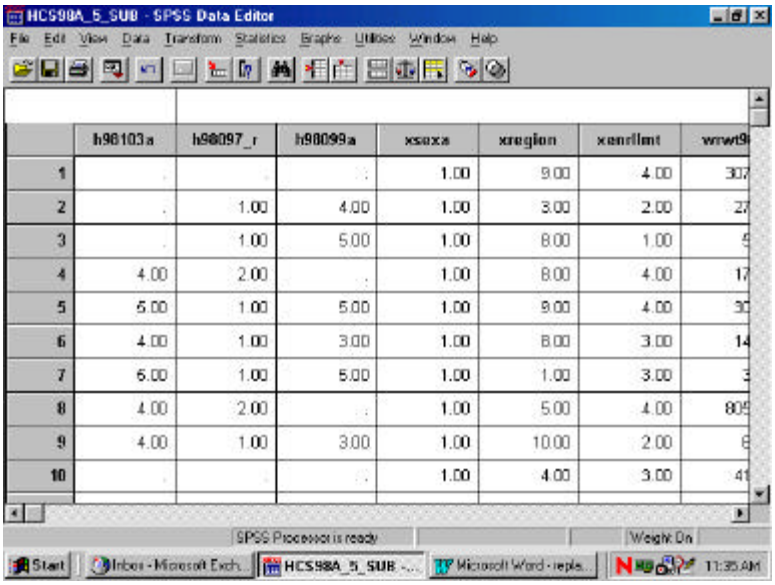


Weighting Data

The data file includes a weighting variable, **wrwt98**, which should be applied to all procedure runs. Again, using the **Data** menu, choose **Weight Cases**. In the dialog box, choose **Weight cases by**. Move the weight variable from the list on the left into the slot labeled **Frequency Variable** on the right as shown below:



Click on **OK** and exit the dialog box. The indication that the data is weighted appears on the status line near the bottom of the screen. As in the following screen, **Weight On** is specified there.



The status line indicates *if* the data is weighted. *Which* weight variable is in effect can only be checked by re-entering the **Weight Cases** dialog box. Weighting stays in effect until it is canceled or until another weight variable is activated.

Building Tables

Building tables starts with creating a new subset of variables that includes h98103a, h98099a, h98097_r, cacsmpl, bfgroupp, xsexa, xenrllmt, wrwt98, and filtr_3. The procedures ***Means*** and ***Crosstabs*** will probably meet most of your statistical reporting needs. SPSS also offers many options for editing the output tables themselves. Some of these options are explained here.

Calculating Means

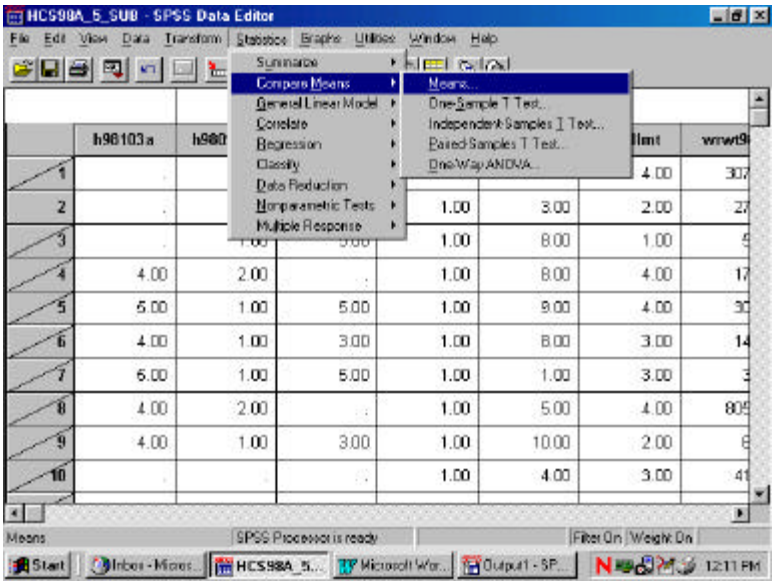
As an example, suppose you want to analyze the satisfaction variables and you want to focus on the Southeast (**xregion** = 3). Suppose you are also interested in overall differences in the mean satisfaction for care received in a *military* facility as opposed to the mean satisfaction with care received in a *civilian* facility for the Southeast region. Within this grouping, you want to examine the effects of the beneficiary group, **bfgroupp**, and sex, **xsexa**.

The satisfaction variables are **h98099a** – satisfaction with a military facility, and **h98103a** – satisfaction with a civilian facility. The statistic you want to see is the mean of the satisfaction variables for each group in our breakdown.

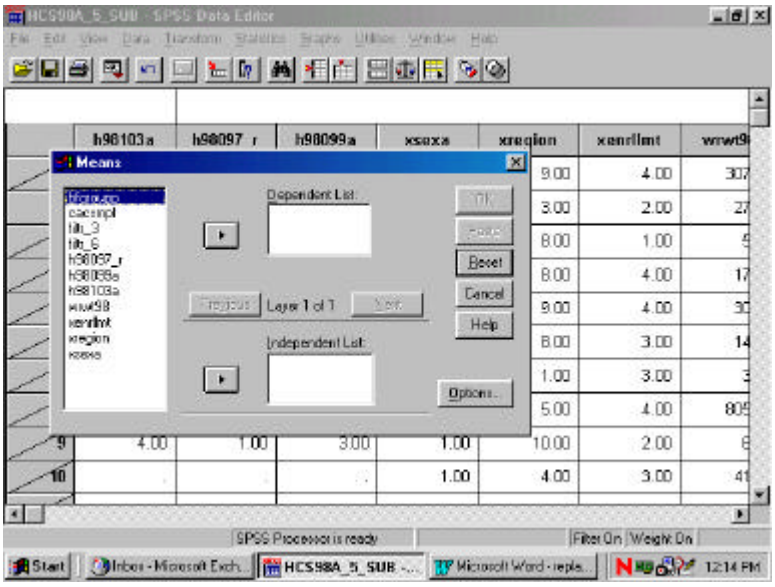
For this analysis, you can use the subset of variables defined above. The subset includes the weight variable, **wrwt98**, which you would activate for procedure runs. The subset also includes the new variable, **filtr_3**, which allows us to select only those cases in the Southeast.

Open the **Data** menu in the **Data Window**. In the **Weight Cases** dialog box, activate the weight variable **wrwt98**. Reopen the **Data** menu and, in the **Select Cases** dialog box, activate the filter variable, **filtr_3**. On the status line, **Filter On** and **Weight On** should appear.

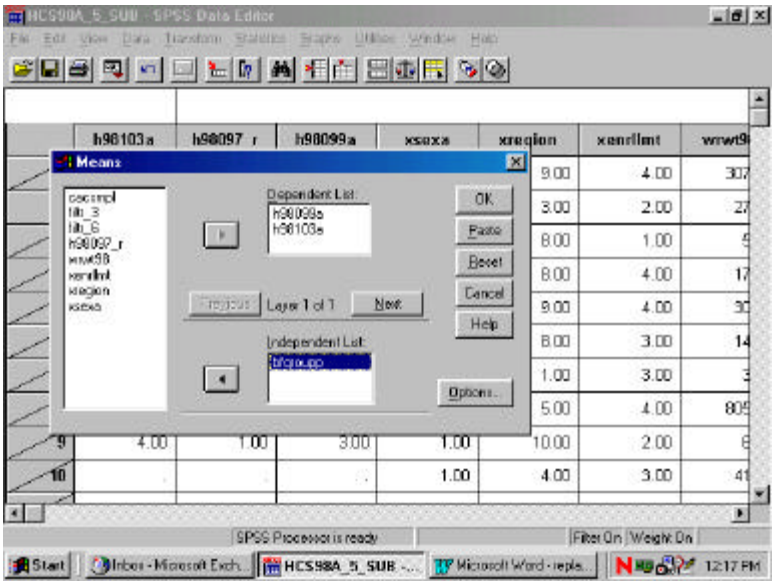
Open the **Statistics** menu in the **Data Window**. Choose **Compare Means** and **Means** from the options as illustrated below.



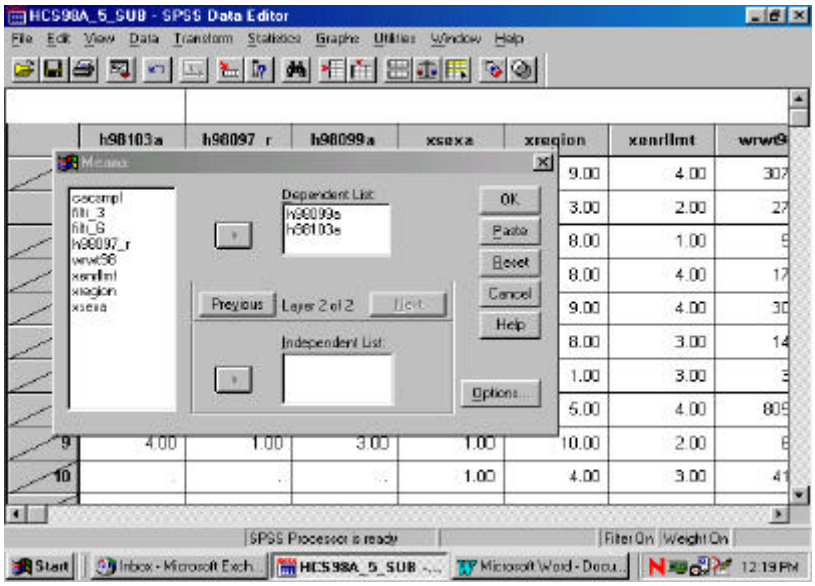
The dialog box for the Means procedure will open as in the following screen:



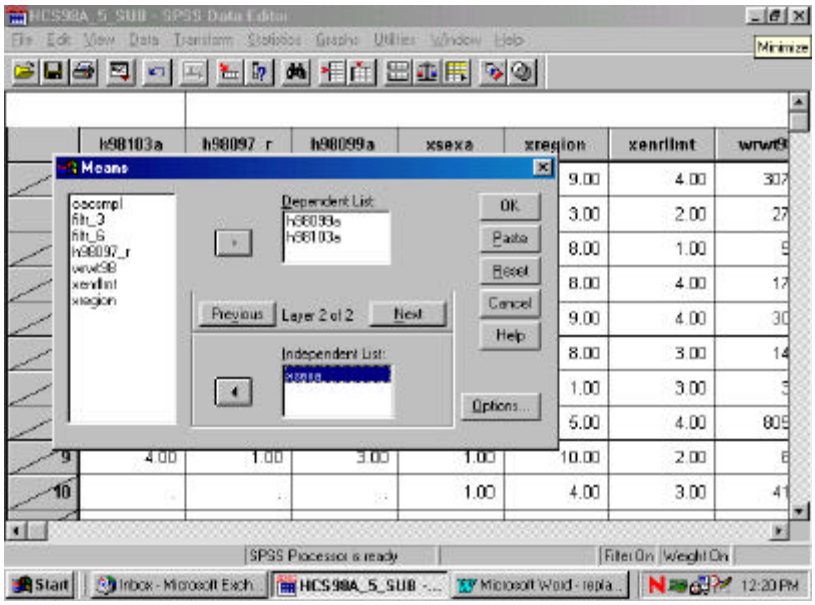
Move the satisfaction variables, **h98099a** and **h98103a**, from the variable list on the left to the box underneath **Dependent List**. These are the two analysis variables. Notice that **Layer 1 of 1** is specified in the middle of the dialog box. Move **bfgroupp** from the variable list on the left into the box under **Independent List**. **Bfgroupp** is the first grouping variable. The screen should look like the following:



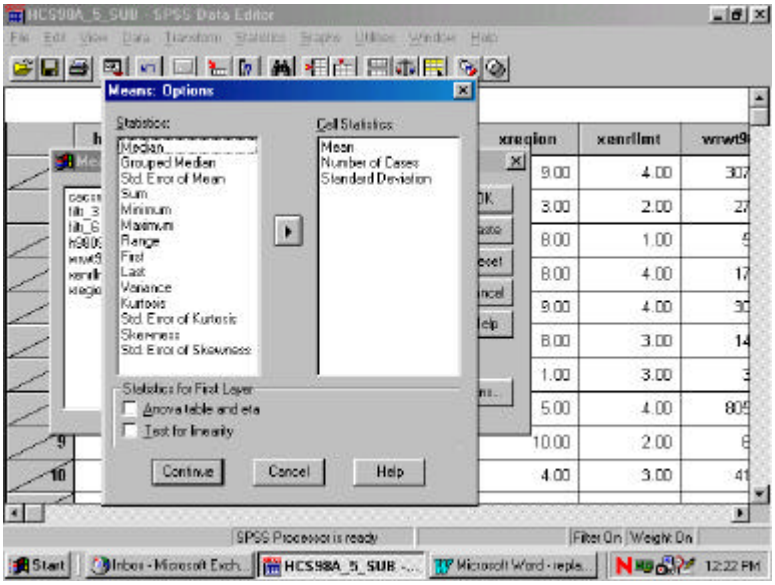
Click on **Nex**t in the center of the box to create a second layer. The following screen will open:



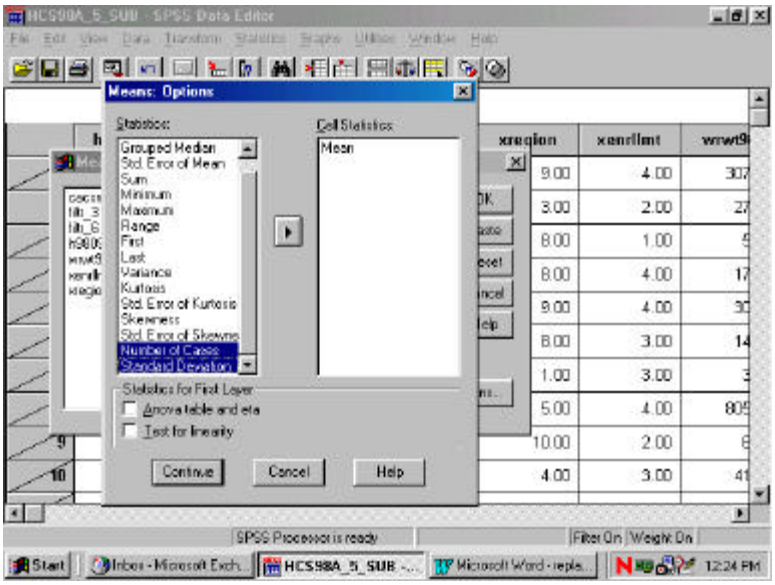
Notice that **Layer 2 of 2** is specified in the middle of the dialog box. Move **xsexa** from the variable list on the left into the box under **Independent List**. **Xsexa** is the second grouping variable. The screen should look like the following:



To set some options, click on **Options** and the following dialog box will open:

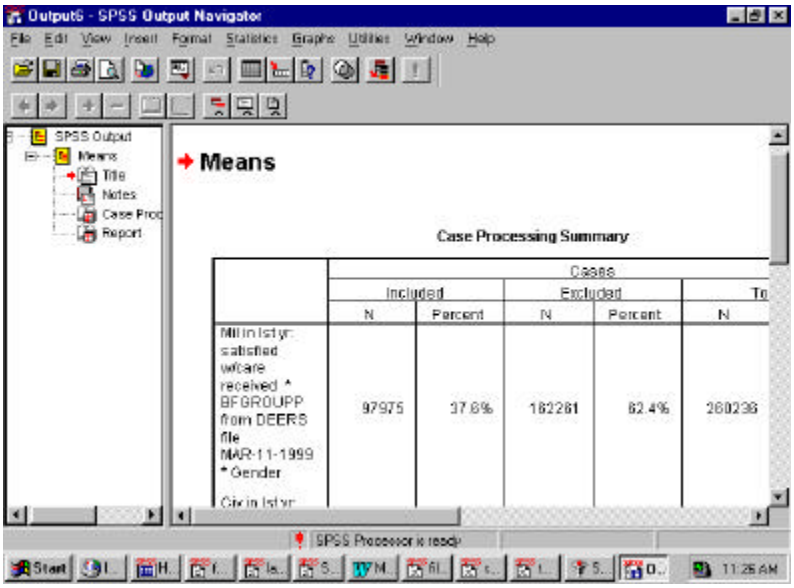


On the left of the box is a list of statistics, under **Statistics.** These are all the possible options for statistical output. In the box under **Cell Statistics** are the default output statistics for the analysis. In this case, **Mean** is the statistic of interest. Highlight **Number of Cases** and **Standard Deviation** and move them to the box at the left, removing them from the analysis, as follows:



Click on **Continue** and return to the previous screen. Click **OK**. The **Means** procedure will run. On the status line, ***Running Means*** will appear, and a counter for the number of cases processed will be activated.

When **Means** has finished processing, the **Output Navigator** window will open automatically. As the name suggests, the output window is not just for looking at output. A number of options are available for *navigating* through output, moving tables, and even editing the tables themselves.



The output is organized into two sections. On the left side is a navigating tool, which lists the components of the right side, the actual output. In the left pane, **Means** is indicated, and indented under it appear **Title**, **Notes**, **Case Processing Summary**, and **Report**. Clicking on **Means** highlights *and selects* all the elements. Lines appear around these elements in the right pane. The indenting indicates that the elements are hierarchically organized, with **Means** at the top. Clicking on any of the sub-elements selects just that element.

A closer look at the left pane reveals another feature. Hiding underneath the element icons are book

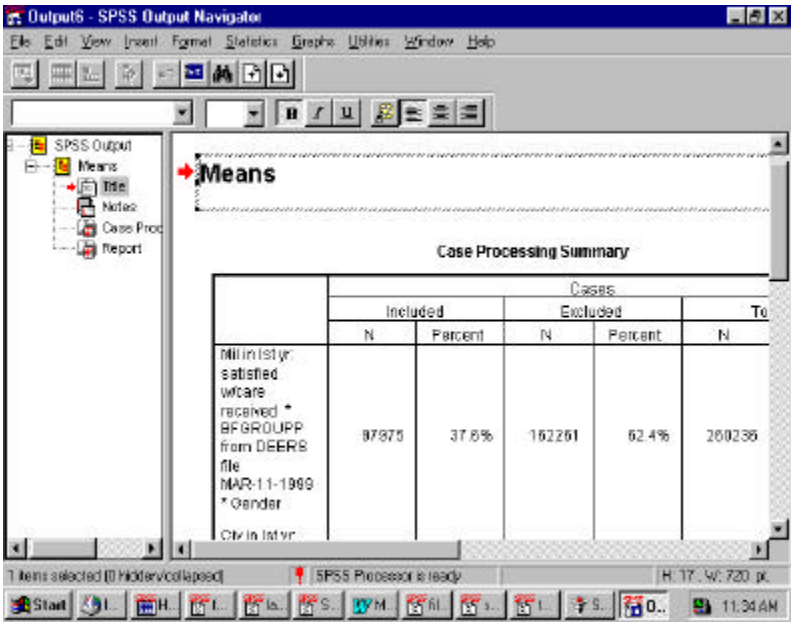
icons. The books are either open or closed. If a book is closed, the element is ***hidden***. Notice that the book under the **Notes** icon is closed. This is a default SPSS option. Double-clicking the icon will open the book, and the Notes will appear in the output. Double-clicking an **open** book will close it, and the physical element will *disappear* from the output. Closing a book and hiding the element does *not delete* the element.

It is possible to select elements in the right pane of the output. Simply click anywhere inside of the actual output element, and that element will be selected.

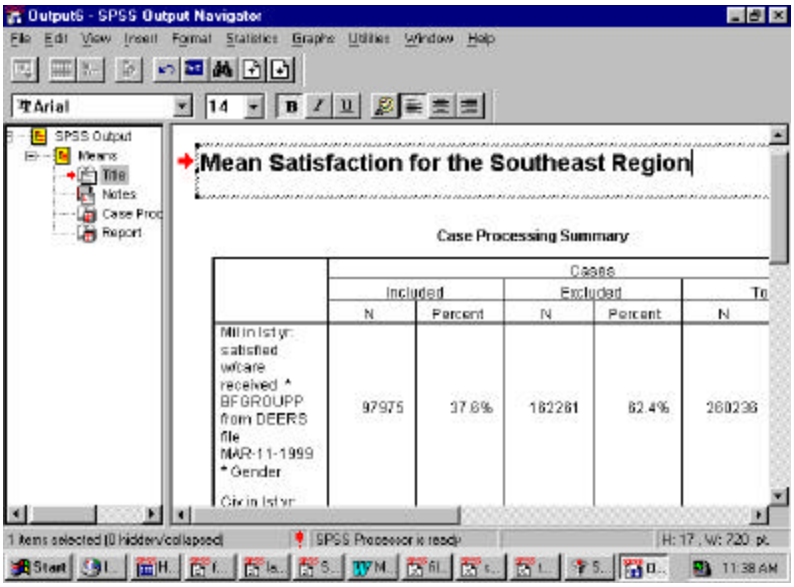
The output may contain many different procedures. The procedure name will be at the top of the list for each section in the left pane. The procedure name does not actually parallel physical output but indicates the category of the output elements.

As you click on each element in the left pane, you will notice that the screen jumps to the actual output of the element, in the right pane. When you click on the procedure name, you jump to the beginning of the next procedure output. This is a quick way to scroll through your output. It also lets you **delete, move**, and **edit** selected elements.

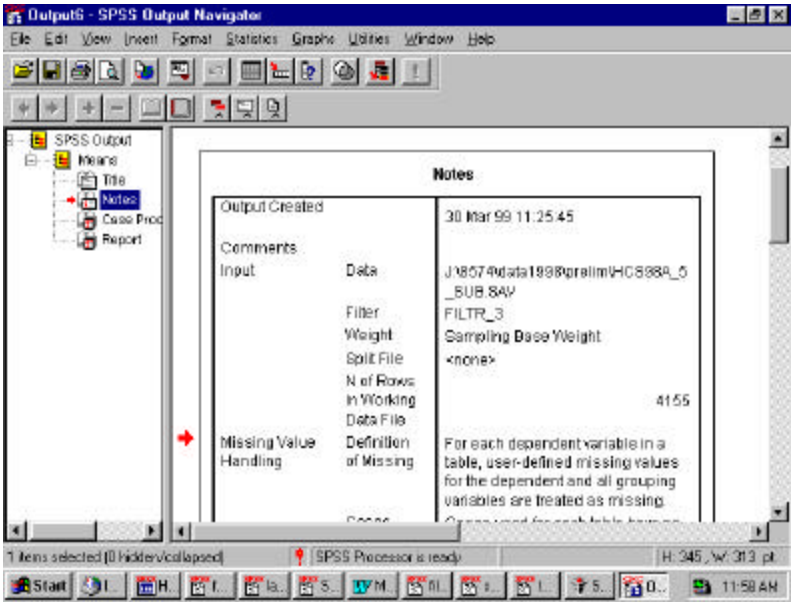
To **Edit** the **Title** element, **Means,** to create a more appropriate title, select the table title by clicking once on the **Title** icon in the left pane. A box now surrounds the title in the right pane. Double-click anywhere within this box, and a box appears around **Means**, as shown in the following screen.



You have entered the **edit** mode for this element, and the cursor appears inside the box. You can delete the word **Means** and write a title that relates to the information in the table. A possible title appears in the next screen. To exit edit mode, click anywhere outside the box. The change you made will be saved.



If you navigate to the next element, **Notes**, you see a closed book. Double-click this item, and the notes will appear as follows:



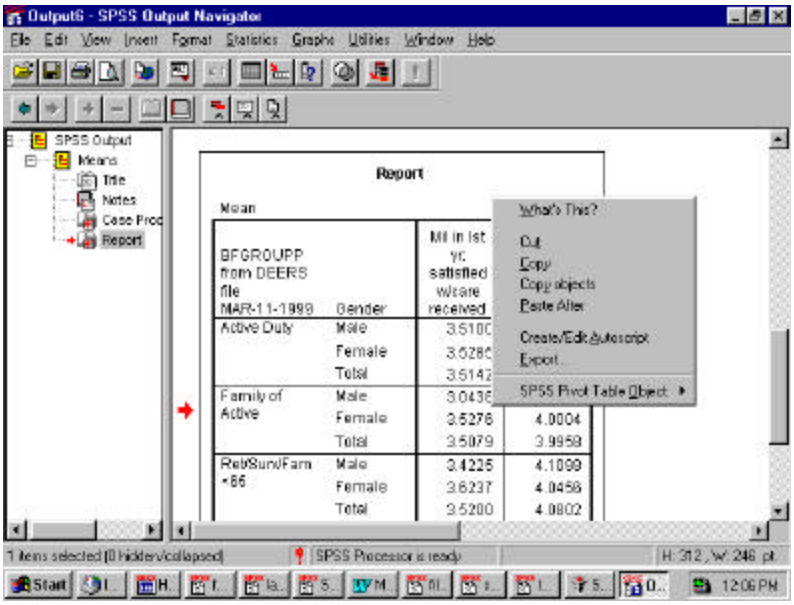
Decide if you want this information to appear in your report. If not, simply double-click the **Notes** icon, and the notes will again become hidden.

Navigate to **Case Processing Summary**. Double-click to bring up the Case Processing Summary table that gives useful information about the number of cases included in and the number of cases excluded from a given procedure. This information is important for the researcher but probably not necessary for the report, so you would delete this item after examining it.

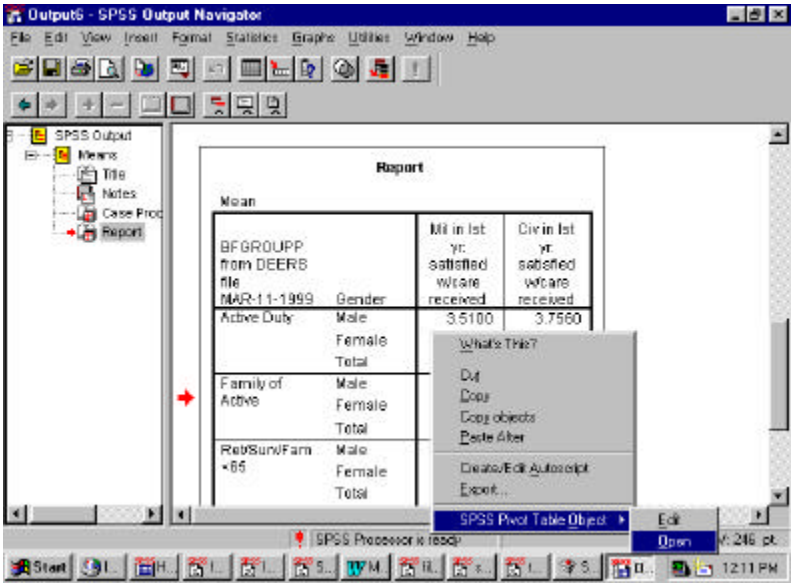
Navigate to **Report**. Double-click to see the actual table output from the procedure **Means**. You can

view this table by scrolling through the output. If the table is large, however, scrolling in the output window can be problematic. A better way to review the table is to open it as a **Pivot Table Object** in a special editor.

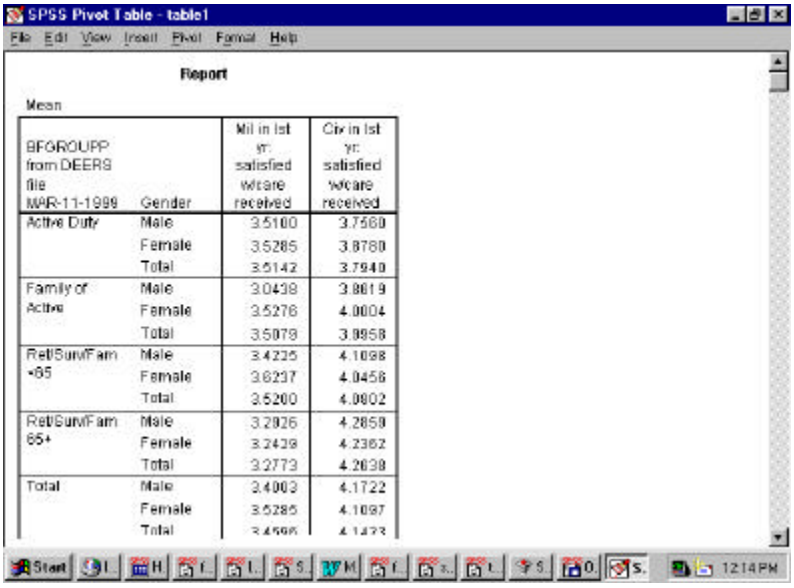
Select the table by clicking the **Report** icon or by clicking inside the table itself. A box will appear around the table. Insert the mouse pointer inside the table and right-click, opening the following dialog box:



Select **SPSS Pivot Table Object** and **Open** as pictured below:

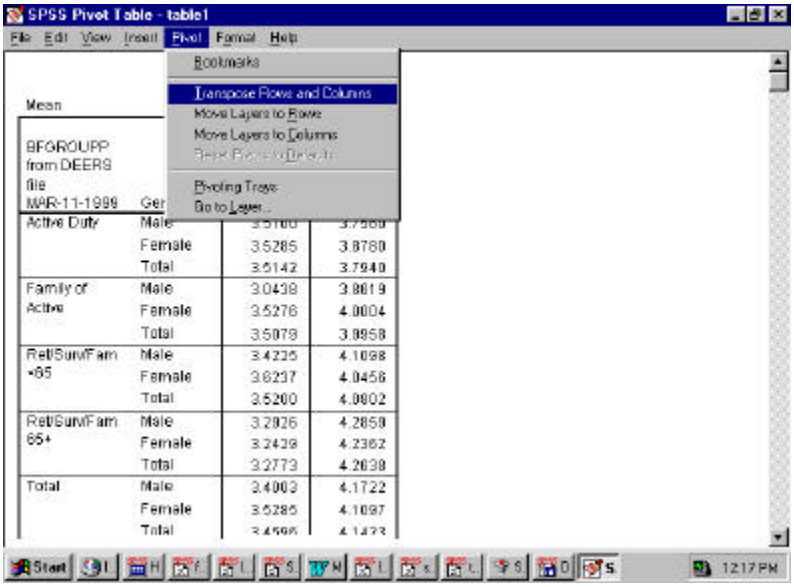


The table will appear in a new screen superimposed on the output. Maximize this screen as shown below.

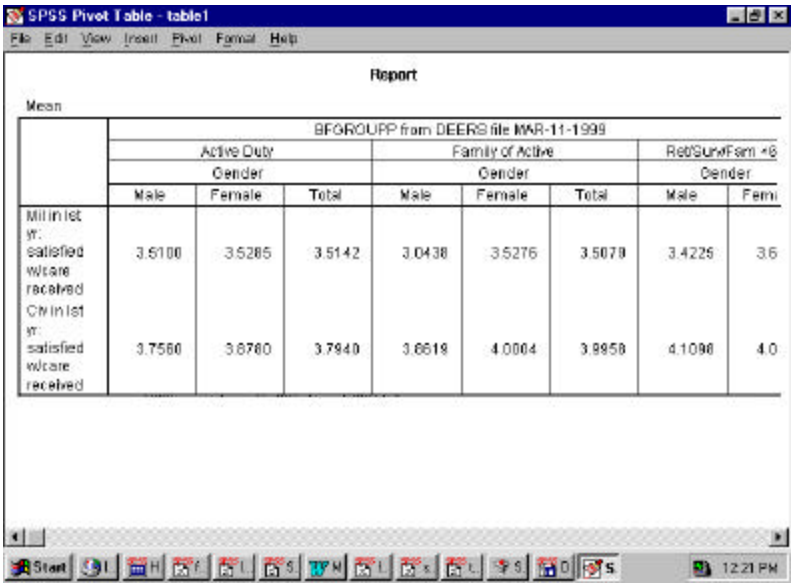


In this special editor, there are many options for formatting the table.

Suppose you want to change the table format from vertical to horizontal. Open the **Pivot** menu in the tool bar and choose **Transpose Rows and Columns** as shown below:

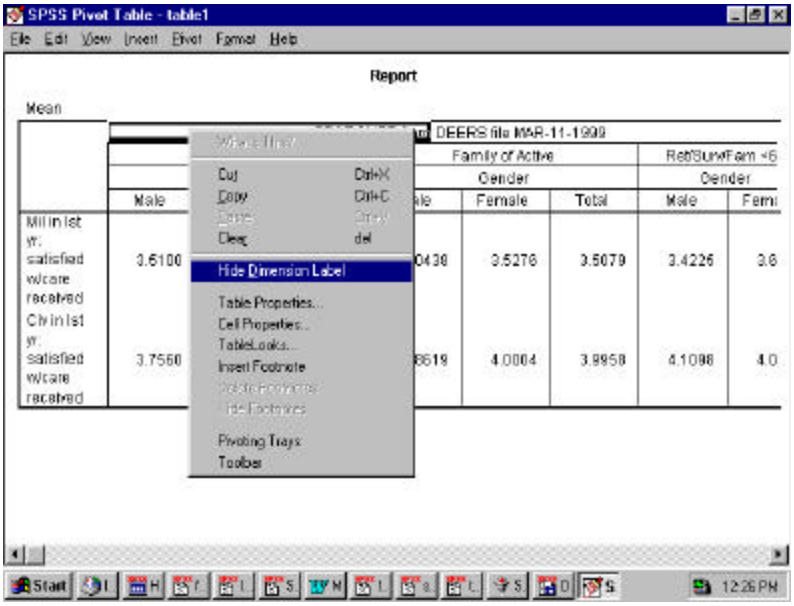


The rows and columns will be reversed as shown in the following screen. Though the table appears too wide in the viewer, it will fit the page when printed. You can do all the table editing in the left section of the table, and the changes will spread through the entire table.

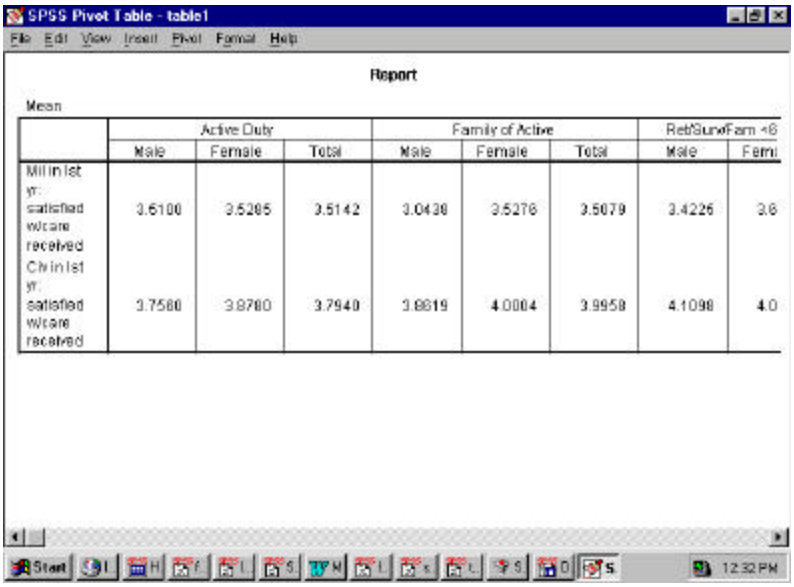


You would then notice that certain labels are redundant. The labels, **Bfgroupp from DEERS file** and **GENDER** are the **Variable Labels** for the variables. The information in these labels is echoed in the **Value Labels**, which are also reproduced in the table. You would delete the Variable Labels as follows.

Click inside the section of the table where the label, **Bfgroupp from DEERS file,** appears. Right-click to open a dialog box, choosing **Hide Dimension Label**, as illustrated below.



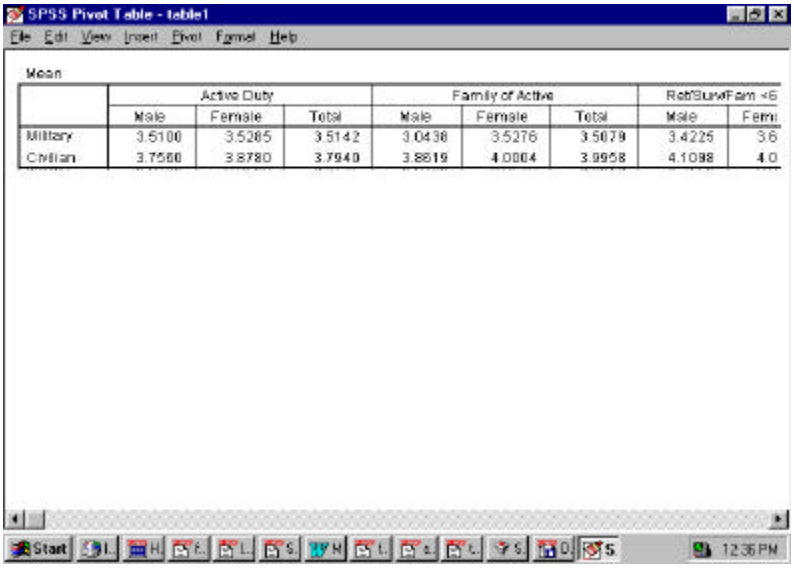
Click inside the table section labeled **GENDER** and repeat the above procedure. An improved table is shown in the following screen.



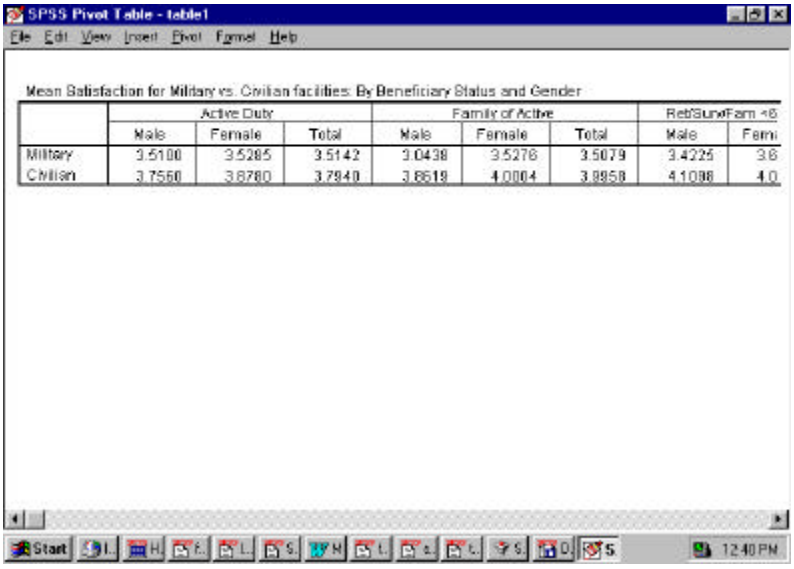
The mean values reported are formatted to allow space for the labels of the satisfaction variables. The spaces between the values are not pleasing to the eye. You can shorten these labels and add the lost information in another place, according to the following procedures:

Double-click on the label for military facility. Delete the text, entering only the word, **Military**. Do the same for the civilian label, entering only the word, **Civilian**.

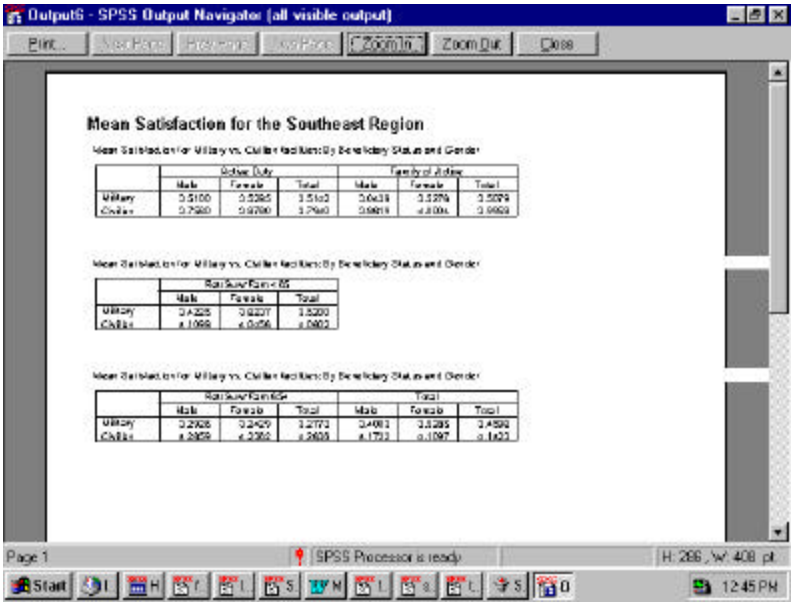
Double-click on the word, **Report,** in the center at the top of the table, right-click, and choose **Delete** from the dialog box.



The resulting table is much more readable. You can then add the deleted information to clarify the table output. Double-click on the label **Mean** at the top left corner of the table, opening the line for editing. Type in a new title for the table. The final result appears below.



After all the editing changes have been made, exit the Pivot Table editor and return to the output navigator. Click on the **File** menu and choose **Print Preview.** Zoom in on the page and review the appearance of the report. The page will appear as the page below.



Calculating Percents

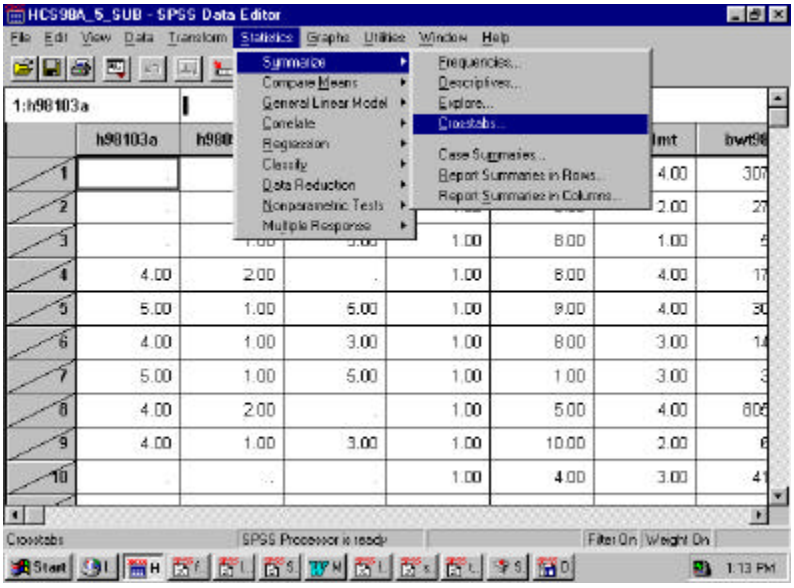
The **Crosstabs** procedure offers many options for analyzing data. The distribution of cases resulting from "crossing" one variable with another is often of interest. The number of cases, row percentages, column percentages, total percentages, and residuals are easily reproduced by **Crosstabs**. A full array of statistics is also available.

The examples given here involve examining relationships between variables, with a view toward the number of cases and the percent of cases in cells produced by "crossing" the variables.

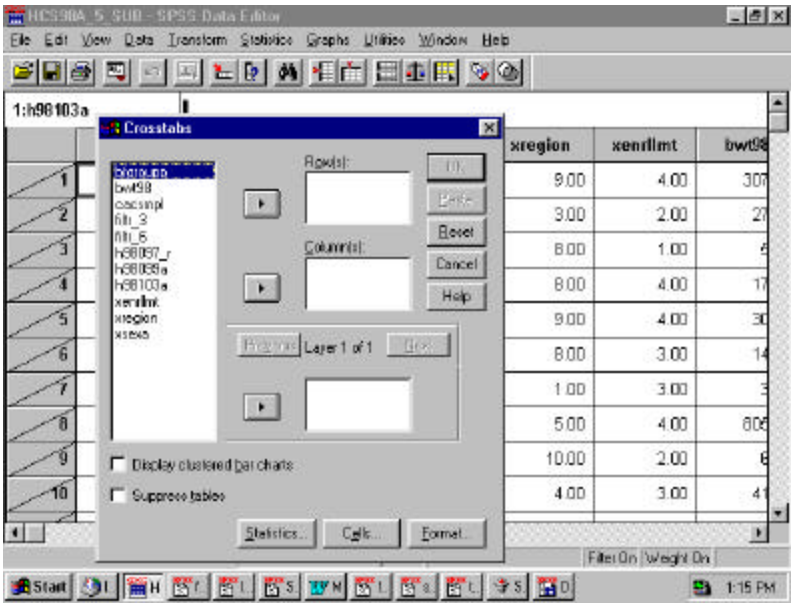
For example, suppose you want to see the percentage of people in the Southwest catchment areas who answered "yes" or "no" to the question, "Did you get most of your medical care from a military facility in the last 12 months?" The variables in this analysis are **cacsmpl** – the catchment area, and **h98097_r** – the question variable. The cases for the analysis are from the Southwest only.

The first task is to build a new filter variable, assigning **1** to the variable when **xregion = 6**. You would call the variable **filtr_6** and build it the same way you built the filter, **filtr_3**. Cases from the Southwest are selected when you activate the filter, and the other cases are filtered out. Check the status line for **Filter On**. **Weight On** should also appear. The cases should be weighted by **wrwt98**.

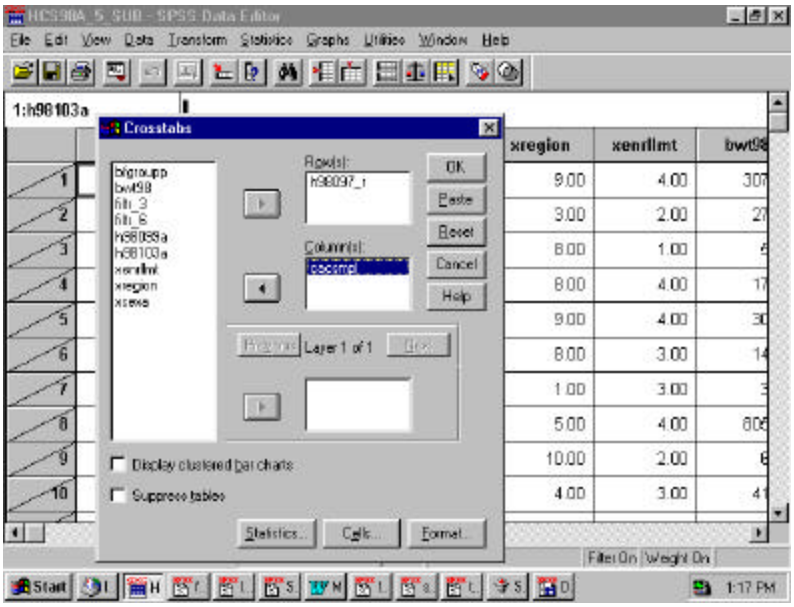
Next, open the **Statistics** menu in the **Data Window**, choosing **Summarize** and **Crosstabs** , as shown below.



The **Crosstabs** dialog box will open as follows:

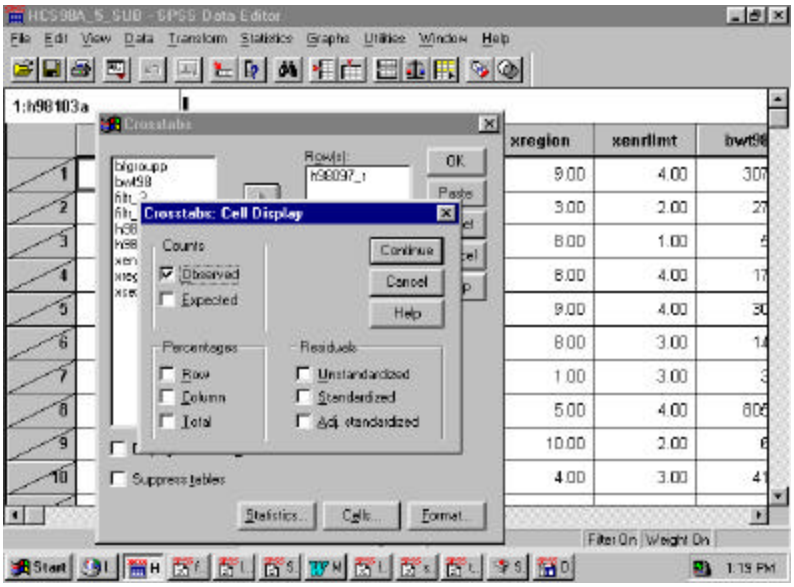


Move **h98097_r** from the variable list on the left into the box marked **Row(s)**:, and move the variable **cacsmpl** into the box marked **Column(s):**. The screen will resemble the following:



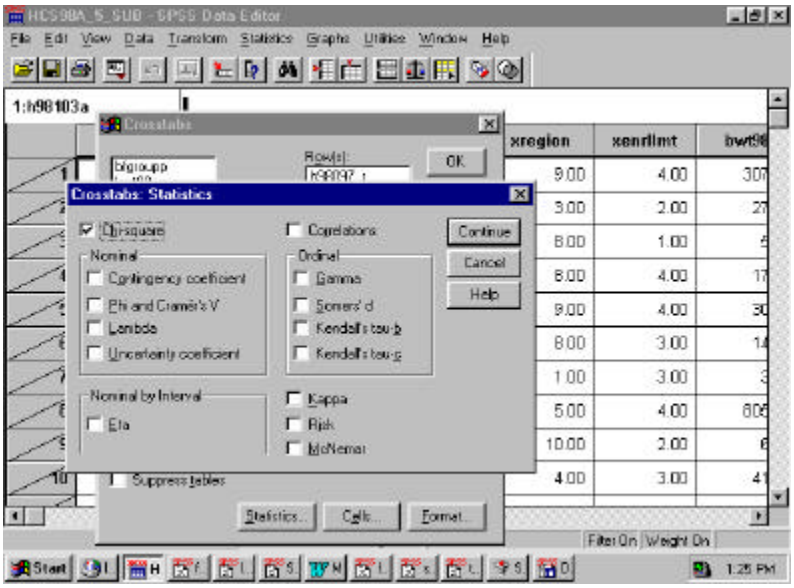
For this analysis, there are no **Layer** variables, so you can proceed to format the table cells.

Click on **Cells...** and open the following dialog box.



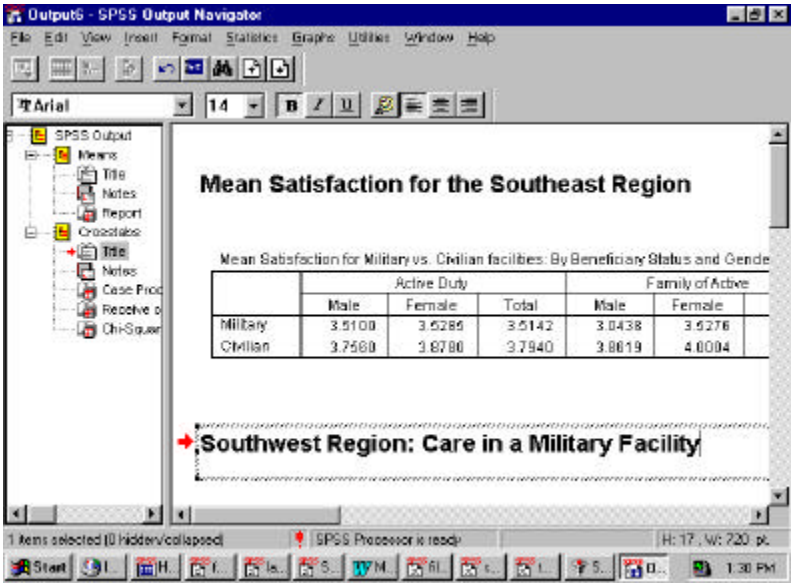
Under **Counts**, **Observed** is checked. This refers to the cell count, a statistic you want to see, so you would leave it checked. Under **Percentages**, check **Column** because you are interested in the percentage of people in each catchment area. Click **Continue** and return to the original screen.

Suppose you also want to see the chi-square statistic. Click on **Statistics**, and the following screen will open:



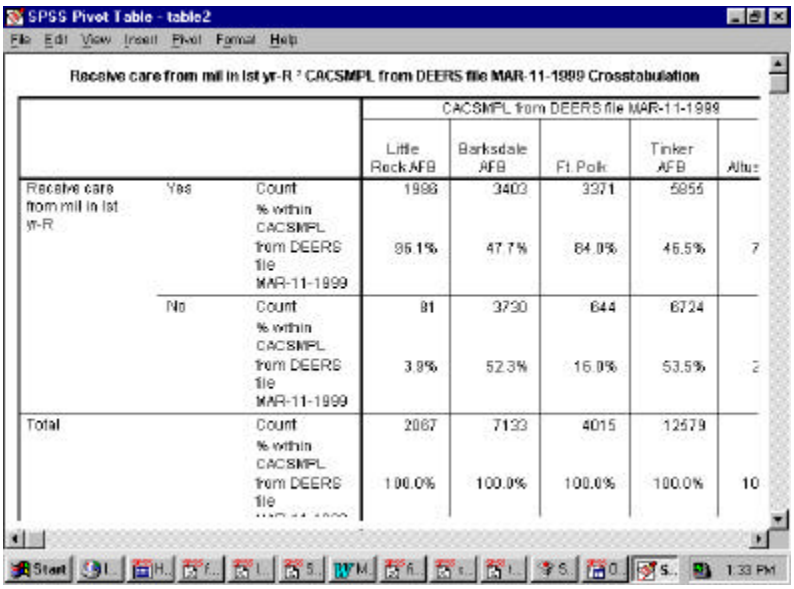
Check **Chi-square** as in the screen above, click **Continue** to return to the first screen, and click **OK** to run the procedure. ***Running Crosstabs*** will appear on the status line, together with the case counter.

When the run is completed, the output window will open, and you can proceed to reformat the table. For a given work session, SPSS appends new output to previous output--in our case, the **Means** procedure. As shown in the next screen, a second section now appears in the left pane, headed by the word **Crosstabs**. Navigate to the **Title** section and double-click inside the title box to change the text in the box to fit the table, as in the example below.

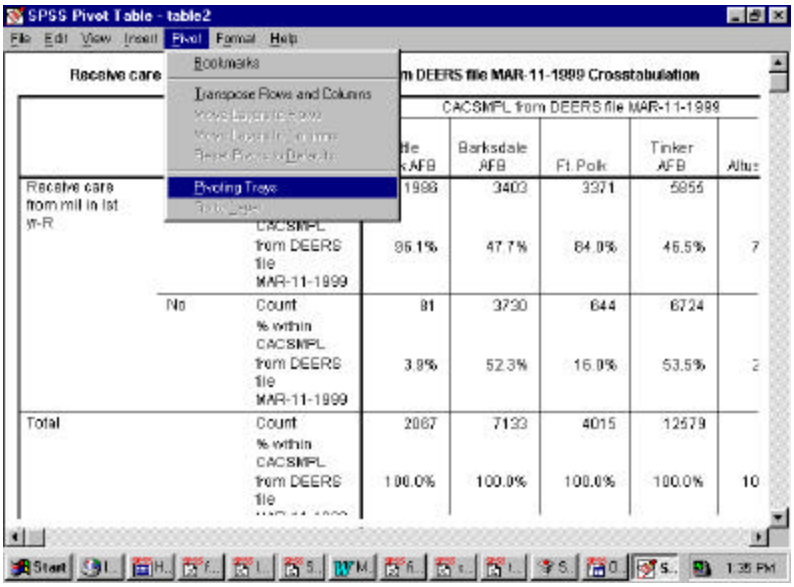


As you did for the **Means** procedure, you would again evaluate the **Notes** and examine the **Case Processing Summary**. Hide the **Notes** and delete the **Case Processing Summary** as you did before.

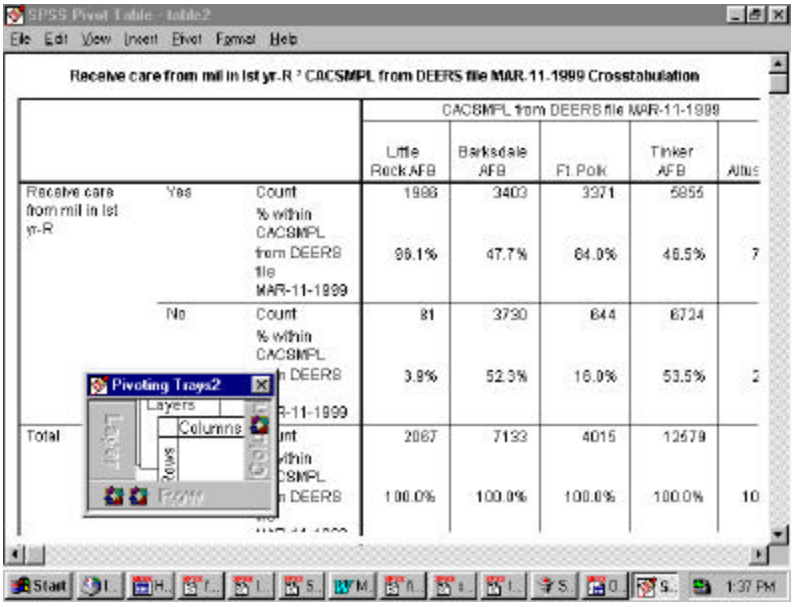
Navigate to the procedure icon. Follow the procedure for opening an **SPSS Pivot Table Object,** open the table in the special editor and maximize the screen as in the following:



The information you requested is in the table, but the table is hard to read. The first possibility is to realign the percent statistic, bringing it into the row dimension. To do this, open the **Pivot** menu and choose **Pivoting Trays**, as in the following screen:

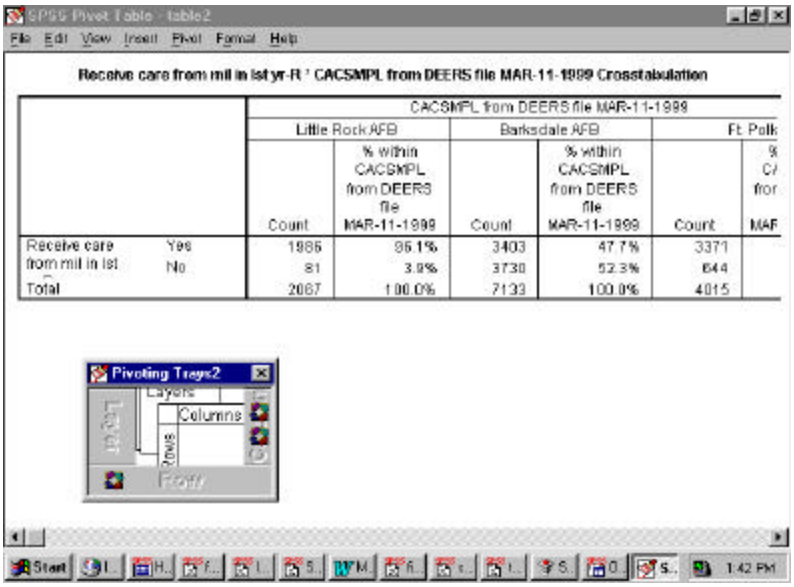


The pivoting tool will appear:



This tool reflects the table structure: rows, columns, and layers. The icons in the margins of the pivoting trays represent the table elements: the variables and the cell statistics. Place the mouse pointer on each icon and notice the element name appear. In this example, on the ROW axis, you would find the variable, **h98097_r** – received care from a military facility, and **Statistics** – the percent of people in each catchment area. On the column axis is the variable, **cacsmpl** – the catchment area.

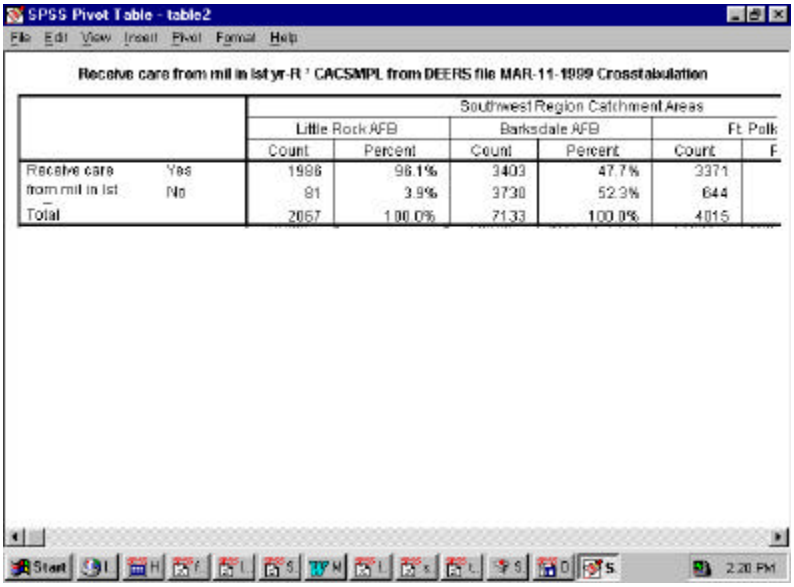
Place the mouse pointer on the **Statistics** icon. Click and drag the icon from the ROW to the COLUMN dimension. The table immediately reformats as in the following screen:



Close the pivoting tool and scroll from side to side in the table. Again, the table appears too wide, but the report will print properly. Notice that the table is much more readable.

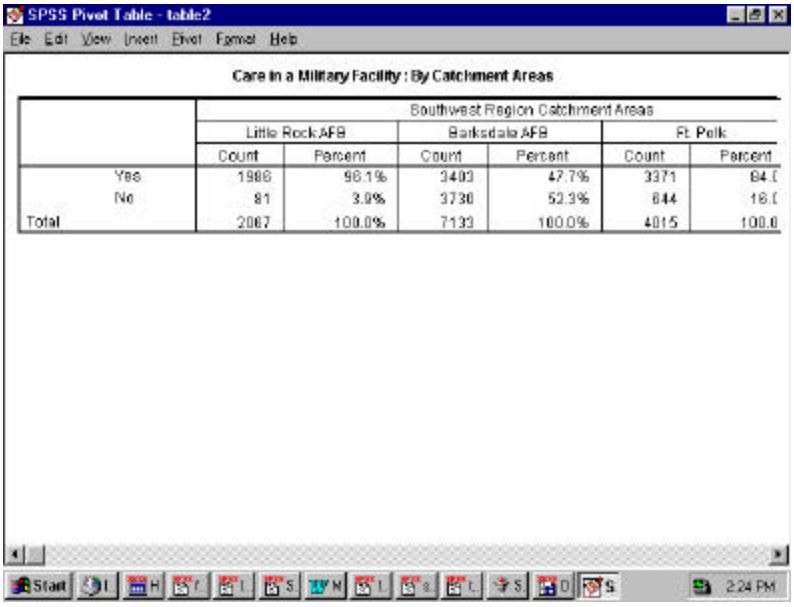
The label at the top of the table is the **Variable Label** for **cacsmpl**. Select it by double-clicking and edit it for clarity (see the screen below).

The table is now formatted to accommodate the long percent label, creating a lot of wasted space. Double-click this element, delete the text, and replace it with the word, “Percent”. The empty space disappears and the table appears as follows:



Next, notice that the label for **h98097_r** is awkward. Select and clear it.

Last, edit the text in the table label so that it better expresses the content of the table. The finished table appears as follows:

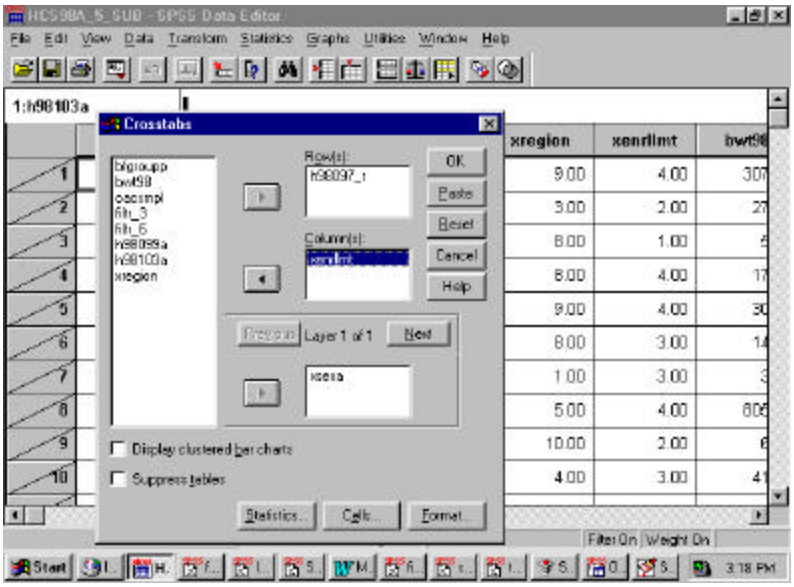


Check **Print Preview** to see if the table is acceptable.

The last example shows you how to add a **Layer** dimension to a **Crosstabs** analysis. Using the same row variable, **h98097_r**, suppose you want to look at the percentage of people by their enrollment status in TRICARE Prime, **xenrllmt**, who received care in a military facility. Suppose you are also interested in sex differences, **xsexa,** among the groupings. **Xsexa** is the **Layer** variable. You want to remain in the Southwest region, using **filtr_6** as the filter variable. The cases will be weighted by **wrwt98**.

The status line indicates **Weight On** and **Filter On**. Verify that both the weight and the filter variables are appropriate.

Once more, open the **Crosstabs** dialog box, enter the analysis variables, and set the **Cells** options, checking **Column** under **Percentages** until the dialog box looks like the following:



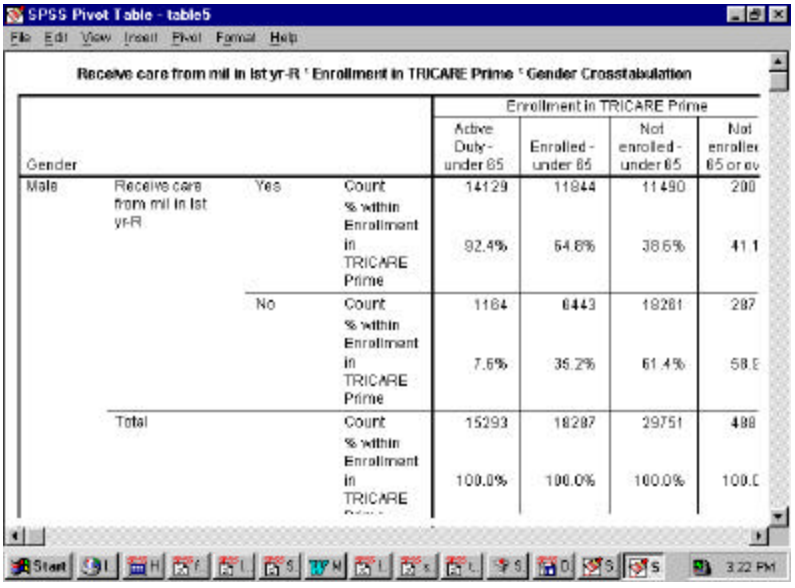
Do the following:

Run Crosstabs.

Edit the **Title** element in the **Output Navigator**.

Examine **Notes** and the **Case Processing Summary** to verify that the CrossTab ran as expected.

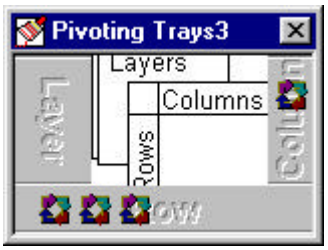
Open the table as an **SPSS Pivot Table Object,** and the following will appear:



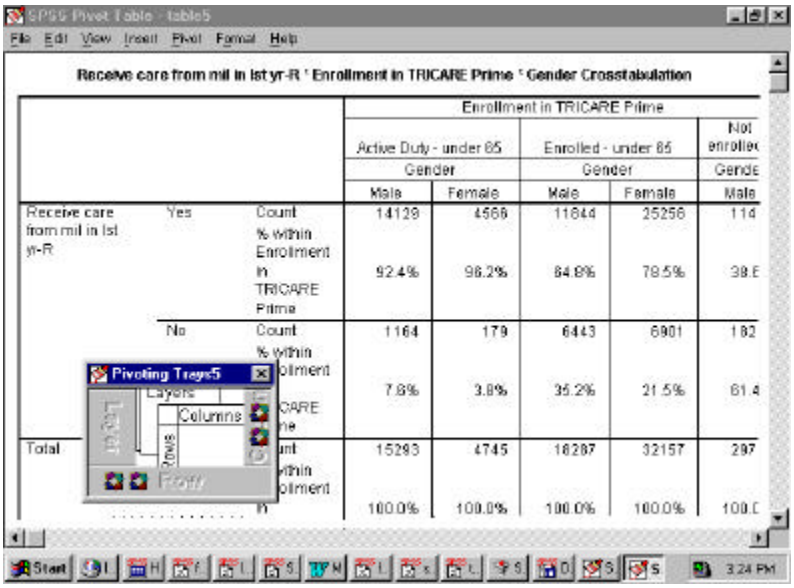
The table is difficult to read, but you can improve it by doing the following.

Select the **Pivot** menu to activate the **Pivoting Trays**. The table structure is reproduced in the tool as

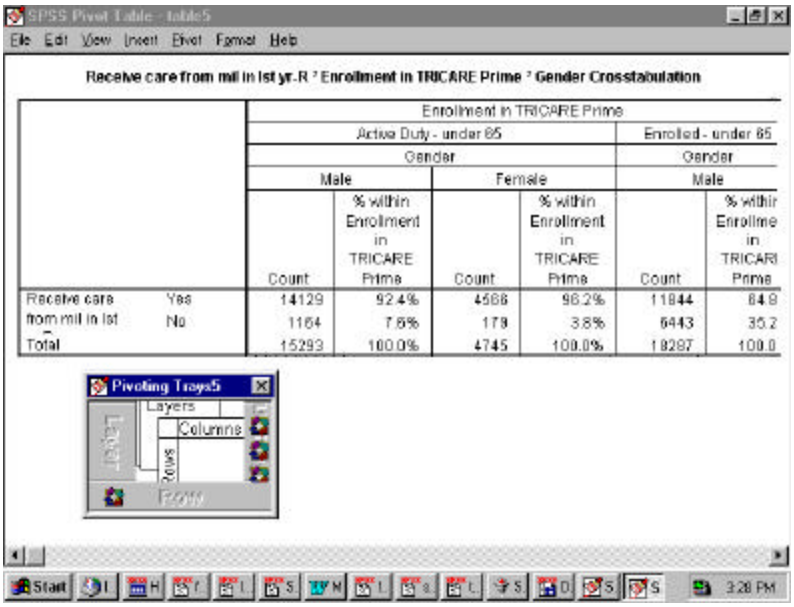
follows:



Place the mouse pointer on each small icon to find the second grouping variable, Gender, in the ROW dimension. Move it to the COLUMN dimension, and the table changes to the following:

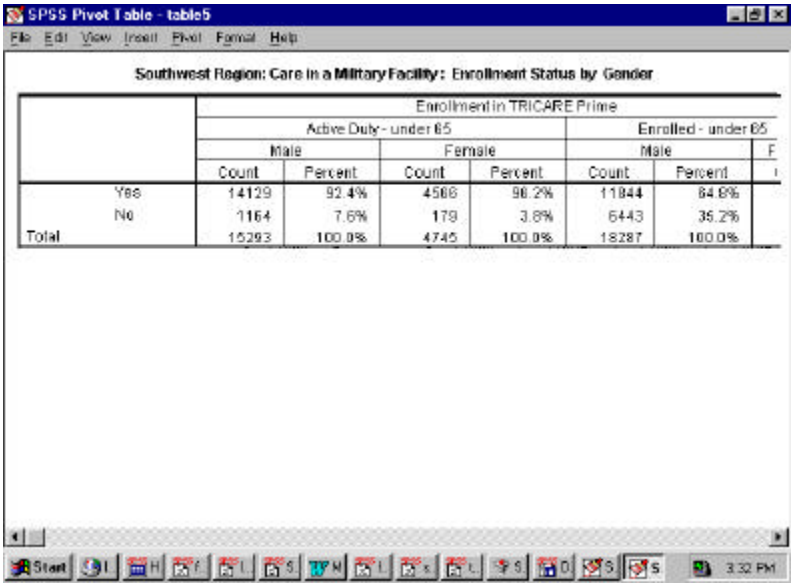


Then drag the **Statistics** icon to the COLUMN dimension to produce the following change:

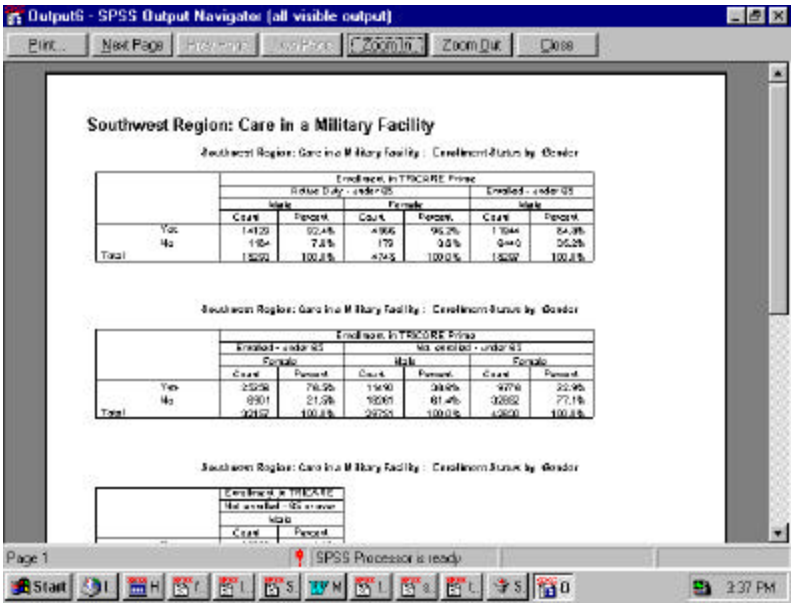


Close the **Pivoting Trays** to hide the dimension label, GENDER, in the table. Then, change the percent label to "Percent" and delete the label for **h98097_r** in the row dimension. Last, revise the label above the table to make it more informative.

The resulting table is both clear and informative.



The Print Preview, as in the view below, shows how the report will print.



Calculating Variances of Estimates

Sampling error occurs when estimates are derived from a sample rather than a complete census of the population. The sample used for a particular survey is only one of a large number of possible samples of the same size and design that could have been selected. Even if the same questionnaire and instructions were used, the estimates from each sample would differ from the others. The standard error (or square root of the variance) indicates the magnitude of the sampling error and thus measures the precision expected from a particular sample.

It is desirable to assess the accuracy of an estimate. It is customary to construct a confidence interval within which one is sufficiently sure the true population value lies. The standard error of a survey estimate measures the precision with which an estimate from one sample approximates the true population value. Thus, the standard error is used to construct a confidence interval for a survey parameter for an assessment regarding the accuracy of the estimate.

This section explains how to estimate standard errors or variances for estimators computed from the 1998 HCSDB. For a full discussion of variance estimation methods, see Wolter (1985) and references cited therein.

Variance Estimation Methods

To account for the sample design,[1] it is customary to use either Taylor series linearization or a resampling method for variance estimation. Neither variance estimation method is, in general, better so the choice of one or the other is largely a matter of convenience. To help users to estimate standard errors using Taylor series linearization or jackknife replication, the public release files for the 1998 HCSDB include the following variables:

[1]The 1998 HCSDB uses a stratified sampling design. For details, see D.S. Jang et al. "The 1998 Health Care Survey of DoD Beneficiaries: Form A Sample Design." Washington, DC: Mathematica Policy Research, November 1998.

The stratum variable and the final weight (STRATUM and WRWT98) for the Taylor series linearization method

Jackknife replicate weights (WRWT01 to WRWT40) for the jackknife replication method

There are two popular software packages for performing Taylor series linearization or the jackknife replication method: SUDAAN™ (Shah et al. 1996) and WesVarPC? (Brick et al. 1996), respectively.[2] The discussion below explains how SUDAAN and WesVarPC are used to calculate variance estimates using Taylor series linearization and jackknife replication methods.

Taylor Series Linearization Method

For most sample designs (including the 1998 HCSDB), design-based variance estimates for linear estimators of totals or means can be obtained with explicit formulas. However, nonlinear functions such as ratios do not have exact expressions for the variance. The Taylor series linearization method approximates the variance of a nonlinear estimator with the variances of the linear terms from the Taylor series expansion. Woodruff (1971) presented applications of this technique to sample surveys. Details on this method can also be found in "The 1998 Health Care Survey of DoD Beneficiaries: Technical Manual Form A".

To calculate variance estimates based on Taylor series linearization method with HCSDB's stratified sampling design, you need the stratum variable (STRATUM) as well as the final weight (WRWT98) specified for each data record. The public release files for the 1998 HCSDB include these variables: STRATUM and WRWT98.

SUDAAN incorporates the final analysis weight and the survey design to obtain estimates and their sampling errors. With a small overall sampling rate of about 3 percent, you can use the with-replacement design procedure (STRWR) in calculating standard errors.

All SUDAAN procedures require the following:

- The specification of sampling designs. The terminology for the stratified with-replacement sample design is DESIGN = STRWR.
- The data file sorted by the variable specified in the NEST statement. For the 1998 HCSDB, the data file for Form A must be sorted by STRATUM before using any SUDAAN procedure.

  A FILE TYPE appropriate for SUDAAN, if you use a stand-alone SUDAAN program. For example, some SUDAAN PC versions under Windows or MS-DOS accept only V6.02 through V6.04 SAS files, and FILE TYPE must be specified as SAS. SAS-callable SUDAAN is now available and can be used as a separate procedure directly in a SAS program with any available SAS file as input; FILE TYPE is not needed here.
- The WEIGHT variable for 1998, which is WRWT98.

The following program is an example of how to use SUDAAN to calculate variance estimates for a mean statistic. Suppose you want to estimate:

- Average level of overall satisfaction with military health care (H98099A) among all beneficiaries who ever had any health care from a military facility or provider in the past 12 months (H98097_R=1) for each region (XREGION)

[2]The latest version for SUDAAN 7.5 can also be used for replication methods including jackknife variance estimation.

```
PROC DESCRIPT DATA=HCSDB98 /*FILETYPE=SAS*/ DESIGN=STRWR;
    WEIGHT      WRWT98;
    NEST        STRATUM;
    SUBPOPN     H98097_R=1;
    SUBGROUP    XREGION;
    LEVELS      16;
    VAR         H98099A;
```

- A cross tabulation of respondents in region 3 whose health plan ratings fall into the ranges 0-6, 7-8, and 9-10 (KRATE_HP) by TRICARE enrollment (XENRLLMT)

The following program is an example of how to use SUDAAN to calculate variance estimates for column percentages or row percentages.

```
PROC CROSSTAB DATA=HCSDB98 /*FILETYPE=SAS*/ DESIGN=STRWR;
    WEIGHT      WRWT98;
    NEST        STRATUM;
    SUBPOPN     XREGION = 3;
    SUBGROUP    KRATE_HP XENRLLMT;
    LEVELS      3 5;
    TABLES      KRATE_HP*XENRLLMT;
```

From the above examples, users should note that**:**

- PROC DESCRIPT can be used to compute estimates of means and the corresponding standard errors.
- PROC CROSSTAB can be used to compute estimates of proportions and the corresponding standard errors.

For a more detailed and complete discussion of how to use SUDAAN, see Shah et al. (1996).

Jackknife Replication Method

Another popular way to estimate the variance is to use a resampling method such as jackknife replication, balanced repeated replication, random groups, or the bootstrap method. Like other replication methods, jackknife replication constructs a number of subsamples (replicates) from the full sample and computes the statistics of interest for each replicate (with the same formula as the full sample estimate). The mean square error of the replicate estimates around their corresponding full estimate provides an estimate of the sampling variance of the statistic of interest regardless of the functional form of the statistic.

There are 40 replicate weights (WRWT01-WRWT40) for the 1998 HCSDB in the public use file. Construction of these weights is described in the Form A Technical Manual. With the replicate weights, you can produce jackknife standard errors using in-house or custom written software, or you can use a publicly available software package such as WesVarPC or SUDAAN 7.5. Because WesVarPC 2.02 is available as freeware on the World Wide Web (http://www.westat.com/wesvarpc/index.html), the following example explains how it is used to produce jackknife variance estimates for statistics from the 1998 HCSDB.

Suppose you want to estimate average level of overall satisfaction with military health care (H98099A) among all beneficiaries who ever had any health care from a military facility or provider in the past 12 months (H98097_R=1) for each region (XREGION). You would use WesVarPc as follows.

Create a SAS V6.04 file, SAS Transport file, or ASCII file. WesVarPC has a restriction for the input data format. All files must be converted to one of these three types of files before being imported to WesVarPC.

Create a WesVarPC data file. From the ***Prep*** menu, choose the ***Import Data Files*** screen and import all variables for the analysis. For this example, input H98099A, H98097_R, and XREGION into the ***Variables*** box, WRWT01-WRWT40 into the ***Replicates*** box, and MPR_ID into the ***ID*** box. Also specify the replication method as JK1 on this screen.

Create a data file for the subpopulation. Specify the subpopulation by choosing the ***Subpop WesVarPC Data File*** from the ***Prep*** menu: H98097_R=1.

Calculate estimates. From the ***Tables*** menu, choose ***New*** and select the file created from the above procedure. Then, from the ***Table Request*** screen, specify H98099A as the ***Analysis*** variable, MEAN (H98099A) as the ***Compute Statistics***, and XREGION as the ***Table***.

The above steps can also be followed to produce standard errors. The WesVarPC user's manual (Brick et al. 1996) provides other possible methods for producing standard errors. The latest WesVarPC 3.0 is no longer freeware and can be purchased from SPSS.

Chapter

4

Codebook

This chapter describes every variable in the database. This codebook will also be helpful in identifying which data are available for various analyses, and what, if any, recoding of variables will benefit your needs. It may also be useful in reviewing output.

The variables are in order based on their order in the database. An alphabetical listing (see Table of Contents) is provided to assist in locating variables.

The codebook contains frequency distributions for both discrete and continuous variables. A discrete variable is one that has only a few values. A continuous variable may have many possible values.

Below are two examples of the presentation of variables in the codebook. For each variable, we include the variable name, definition, weighted and unweighted frequency distributions, and the format value for each value. The first example contains a frequency distribution for a discrete variable.

H98105[1]-	**In General, How Is Your Health[1]**				
	Unweighted		**Weighted**		
Value[2]	**Count[3]**	**Percent[4]**	**Count**	**Percent**	**Formatted Value[5]**
A	20	0.03	2445.25	0.09	Mult response
.	505	0.72	27573.59	0.99	No response
1	955	1.35	67692.89	2.42	Poor
2	5419	7.69	312408.50	11.16	Fair
3	21599	30.64	954897.70	34.12	Good
4	27905	39.58	1011837.00	36.16	Very Good
5	14101	20.00	421701.20	15.07	Excellent

The table below contains an example of a frequency distribution for a continuous variable: age. The

[1] variable name
[2] Variable description
[3] All possible values of the variable
[4] Frequency (count) of responses, including missing values and special codes
[5] Percentage of total responses represented by each value, including missing values and special codes
[6] Explanation of the response value codes

frequency does not list every possible value of age individually but instead shows several age ranges that together cover all possible values of age. You will notice that the last range representing ages "75 years and older" includes 2,341 respondents between the ages 75 and 99.

SRAGE [7] - Current age [8]

	Unweighted		Weighted		
Value[9]	**Count [10]**	**Percent**	**Count**	**Percent**	**Formatted Value**
.	862	1.22	42833.79	1.53	Missing data
A	20	0.03	741.58	0.03	Mult response
0	20	0.03	282.58	0.01	Less than 1 yr
1 – 2	8	0.01	94.70	0.00	01--02
3 – 5	11	0.02	596.58	0.02	03--05
6 – 12	18	0.03	508.30	0.02	06--12
13 -- 17	21	0.03	408.57	0.01	13--17
18 -- 34	24542	34.81	423165.30	15.12	18--34
35 -- 44	18348	26.02	398199.50	14.23	35--44
45 -- 54	10344	14.67	498295.00	17.81	45--54
55 -- 64	9235	13.10	656295.90	23.45	55--64
65 -- 74	4734	6.71	513534.40	18.35	65--74
75 -- 99	2341	3.32	263599.40	9.42	75 yrs and older

[7] Variable Name
[8] Variable Description
[9] Range of values show as minimum – maximum
[10] Frequency (count) of responses