SAS Analyst Tutorial

First time users of SAS Analyst should read **SAS Analyst: General Guidelines** to acquaint themselves with the Analyst environment, terminology and commands.

The following tutorial provides instructions on how to use Analyst to perform basic data manipulations and analyses (using the height/weight data set).

I. Getting Started

1. Create the temporary data set:

- Using either **Internet Explorer** or **Netscape 4.73 or higher**, go to http://www.umanitoba.ca/centres/mchpe/teaching/sasmanual/db_htwt.html. Copy the SAS program and then paste it in the Program Editor of SAS.

NOTE: In order to view and analyze the entire dataset (N=18), delete the following lines from the program:

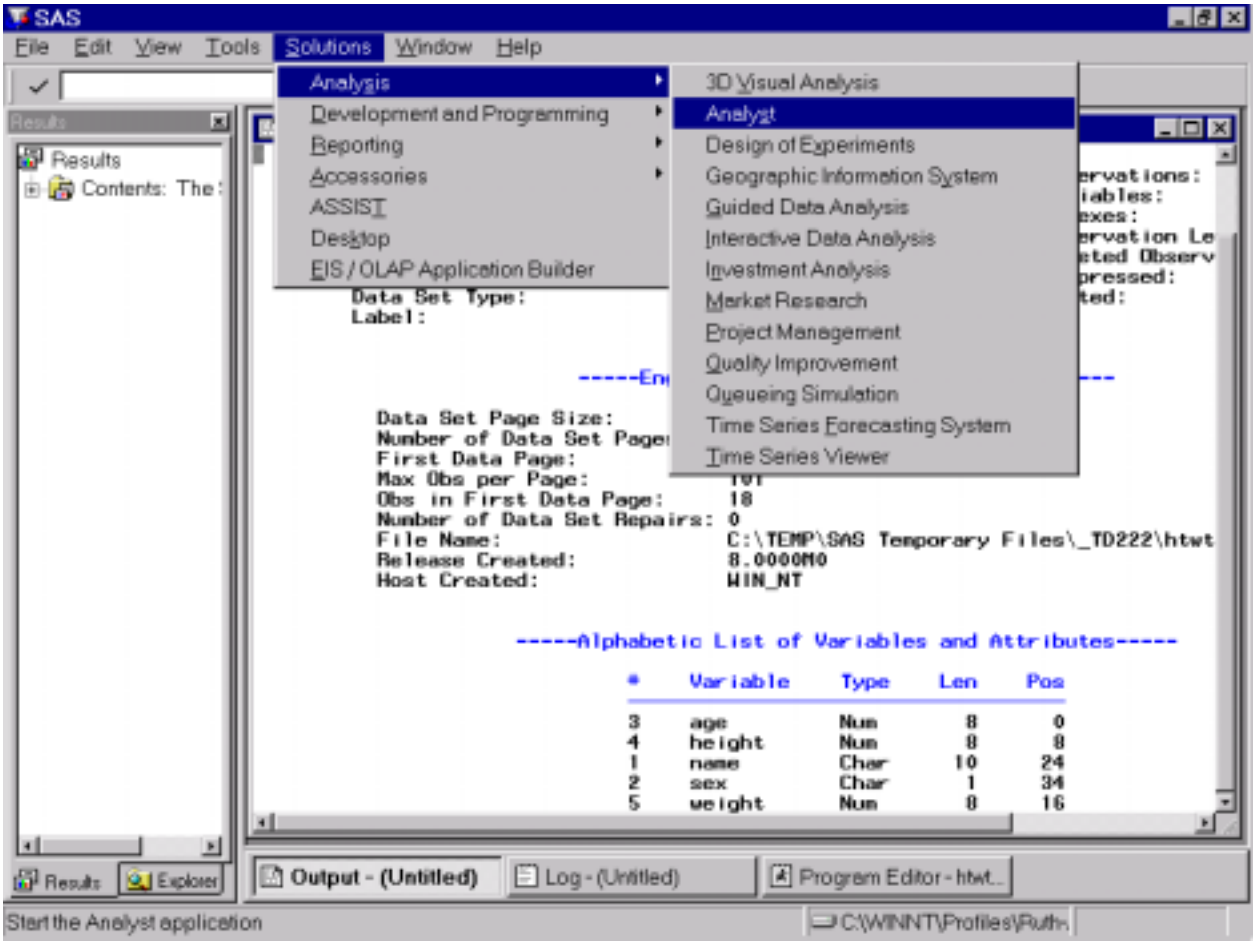
```
/* View the values of the 1st 10 observations */
proc print data=htwt(obs=10);  /* Begin a PROC step */
     run;                      /* End the PROC step */
```

NOTE: Users of Netscape 4.73 or higher: Delete the spaces at the start of each line of raw data so that they are aligned against the left margin. This ensures that the actual columns of data match column numbers identified in the INPUT line above.

- Submit the program: Click on the **SUBMIT** icon  at the top of the screen

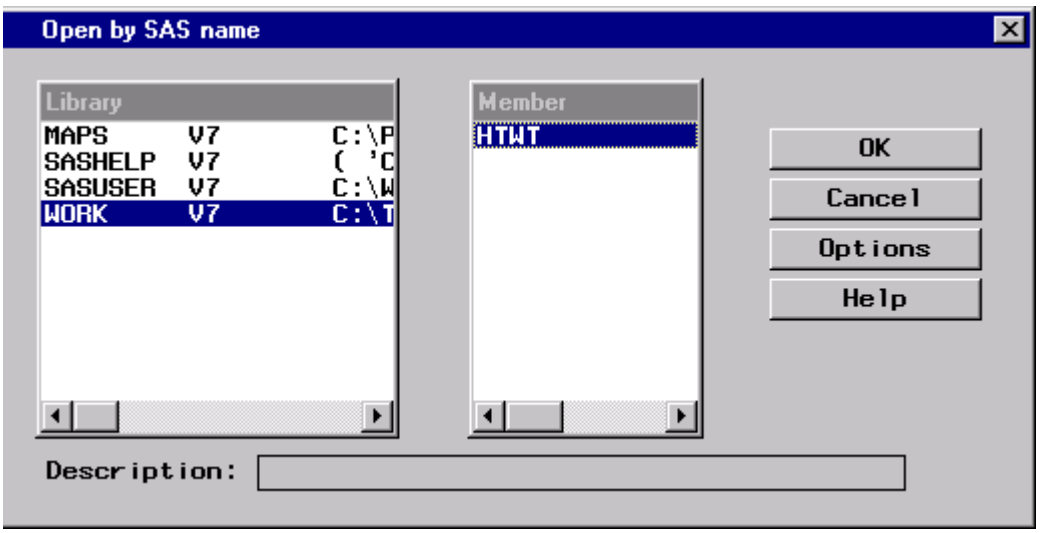
2. Launch Analyst:

From the menu at the top of the screen, select **SOLUTIONS → ANALYSIS → ANALYST**



3. Access the height/weight data set in Analyst:

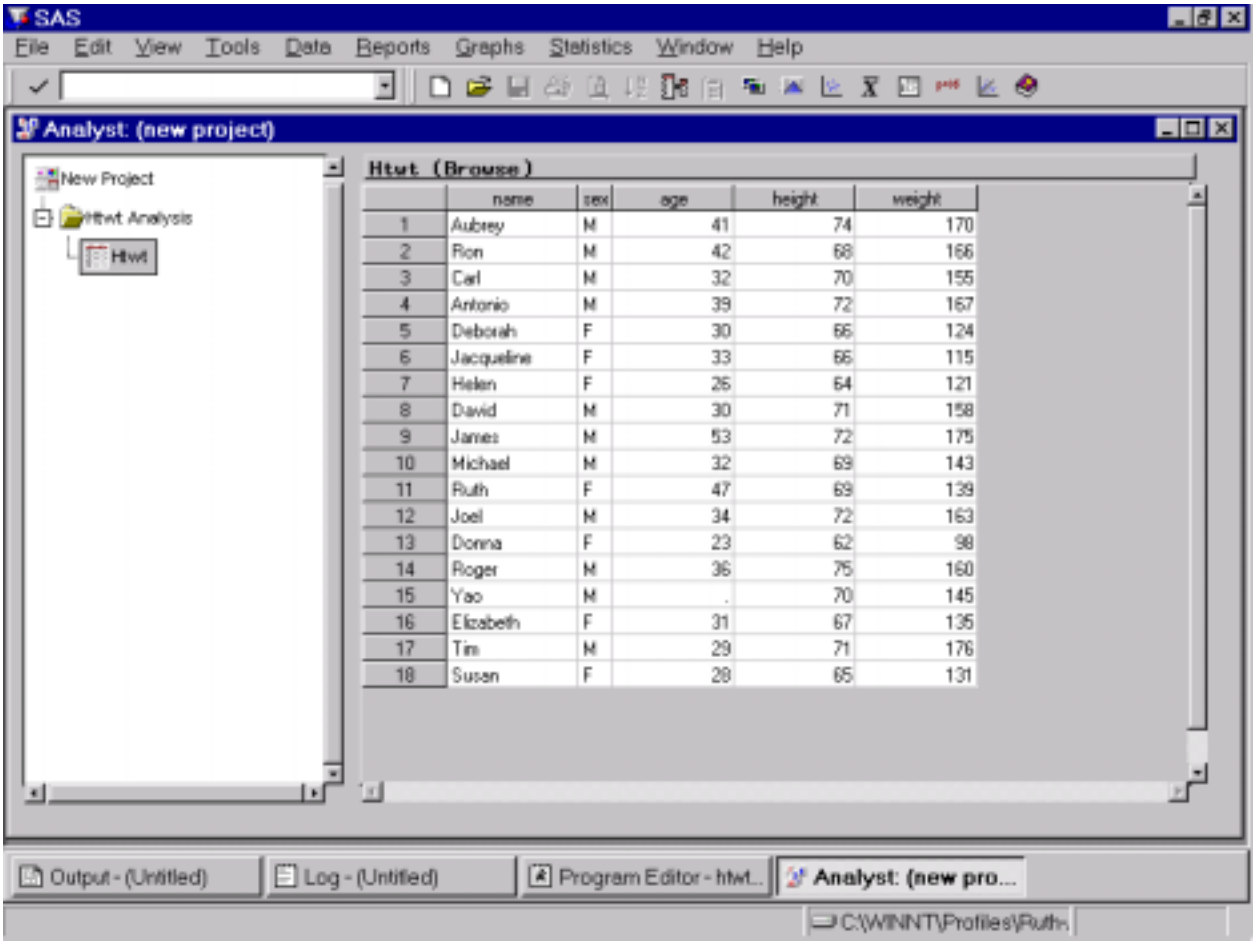
From the menu at the top of the screen, select **FILE → OPEN BY SAS NAME → WORK → HTWT → OK**.



*** Note:** The above window may appear slightly differently for users of SAS 8.

****Note:** At the start of every new session, the temporary data set must be re-created following the above steps.

The data set will look like:



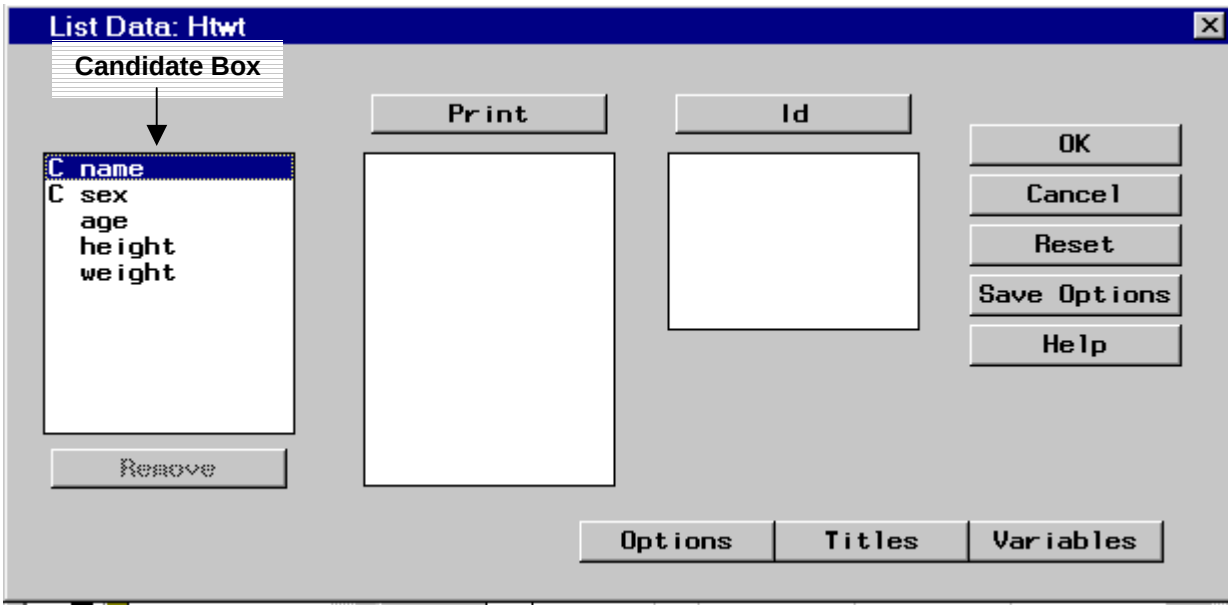
	name	sex	age	height	weight
1	Aubrey	M	41	74	170
2	Ron	M	42	68	166
3	Carl	M	32	70	155
4	Antonio	M	39	72	167
5	Deborah	F	30	66	124
6	Jacqueline	F	33	66	115
7	Helen	F	26	64	121
8	David	M	30	71	158
9	James	M	53	72	175
10	Michael	M	32	69	143
11	Ruth	F	47	69	139
12	Joel	M	34	72	163
13	Donna	F	23	62	98
14	Roger	M	36	75	160
15	Yao	M	.	70	145
16	Elizabeth	F	31	67	135
17	Tim	M	29	71	176
18	Susan	F	28	65	131

II. View the data: Analyst Procedures

Example 1: Listing all variables and all values

You can display the values for any of the variables and for any number of observations in the SAS data set.

1. From the menu at the top of the screen, select **REPORTS → LIST DATA**. The following dialogue box appears:



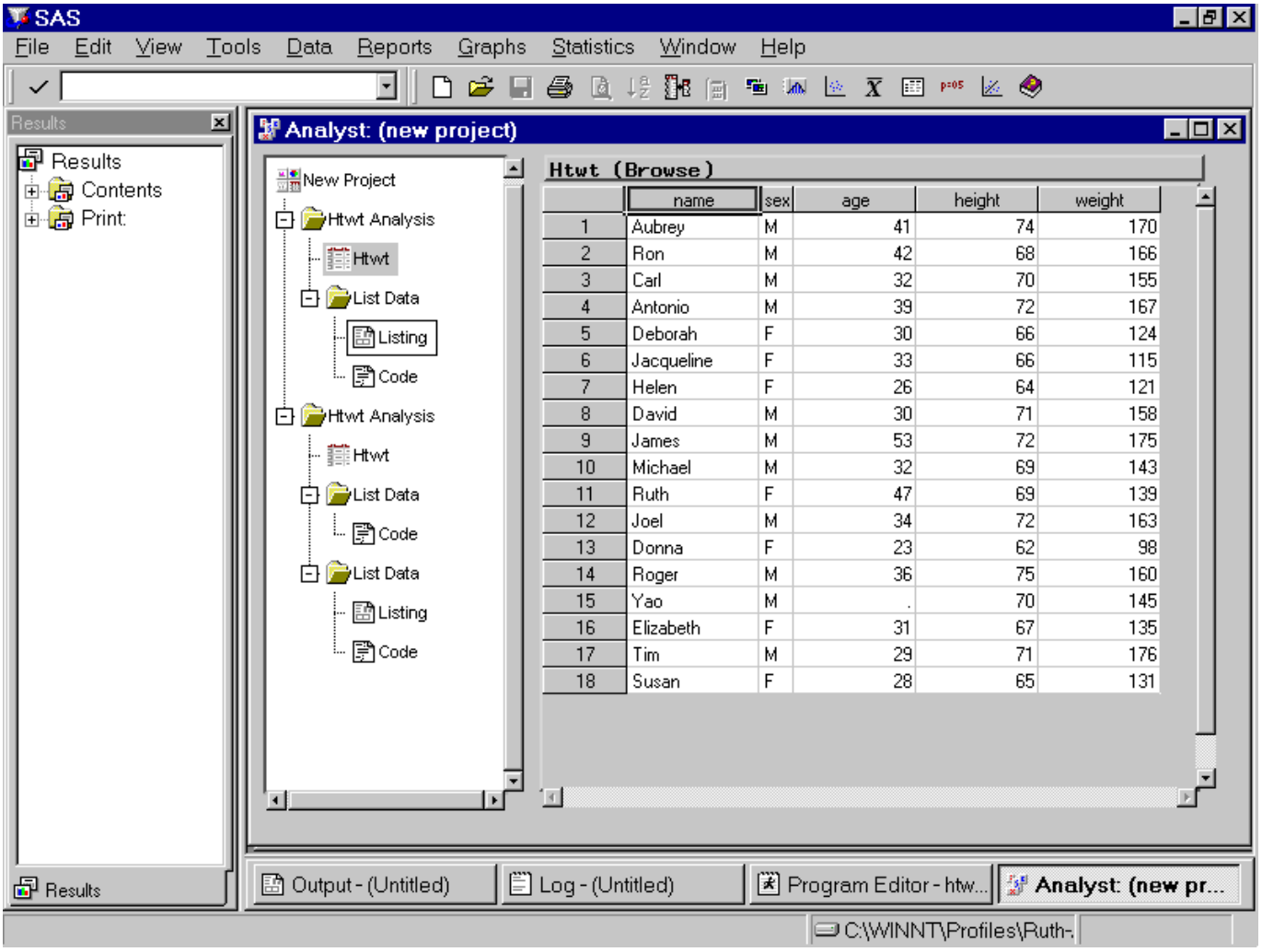
2. The **Candidate Box** lists all the variables in the data set. Select the variables you want to view and then click on **PRINT**. Note: if you accidentally select the wrong variable, just highlight it in the **Print Box** and click on **REMOVE.**

3. Click on **OPTIONS** - ensure that the following options are checked: *Column **name**s*, *Single*, and *observation number*.

4. Click on **TITLES** - these comprise the top three lines at the top of the output. Type in the titles - one on each line: View The Data: Example 1
   Variable List

5. Click **OK** when finished.

NB: The values of the variables identified in the **ID BOX** (ID variables) are used instead of observation numbers to identify the observations in the listing.
- leave blank for this example.



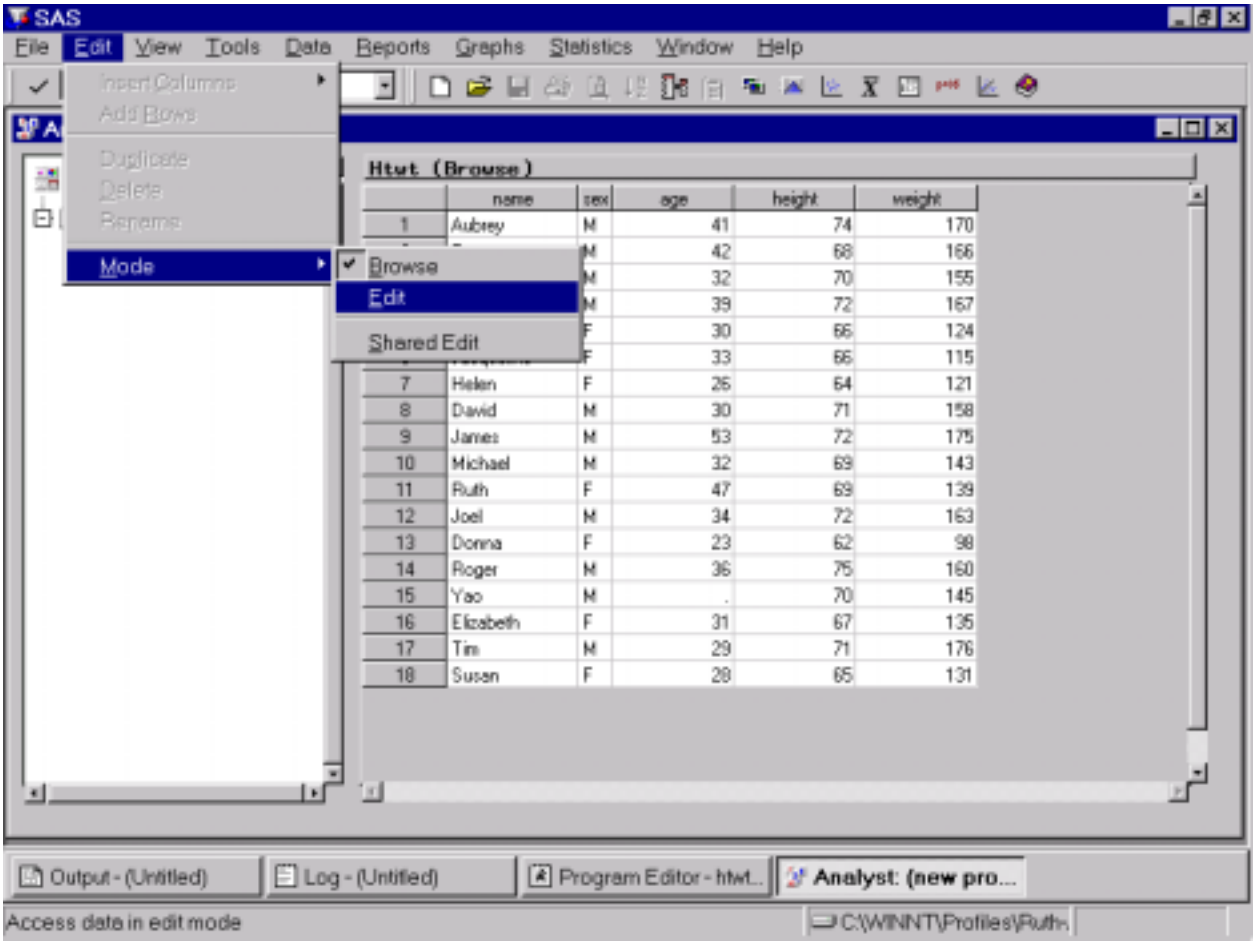
* To view the actual SAS code used to generate the output, double-click on **CODE** in the procedure tree next to the data listing,

Each new procedure performed is added to the tree. You can view any one at any time by right-clicking on one and choosing **VIEW** or **OPEN**.

Example 2: Illustrating Viewing Options

1. **Change mode from Browse to Edit:** From the top of the screen, select **EDIT → MODE → EDIT**.

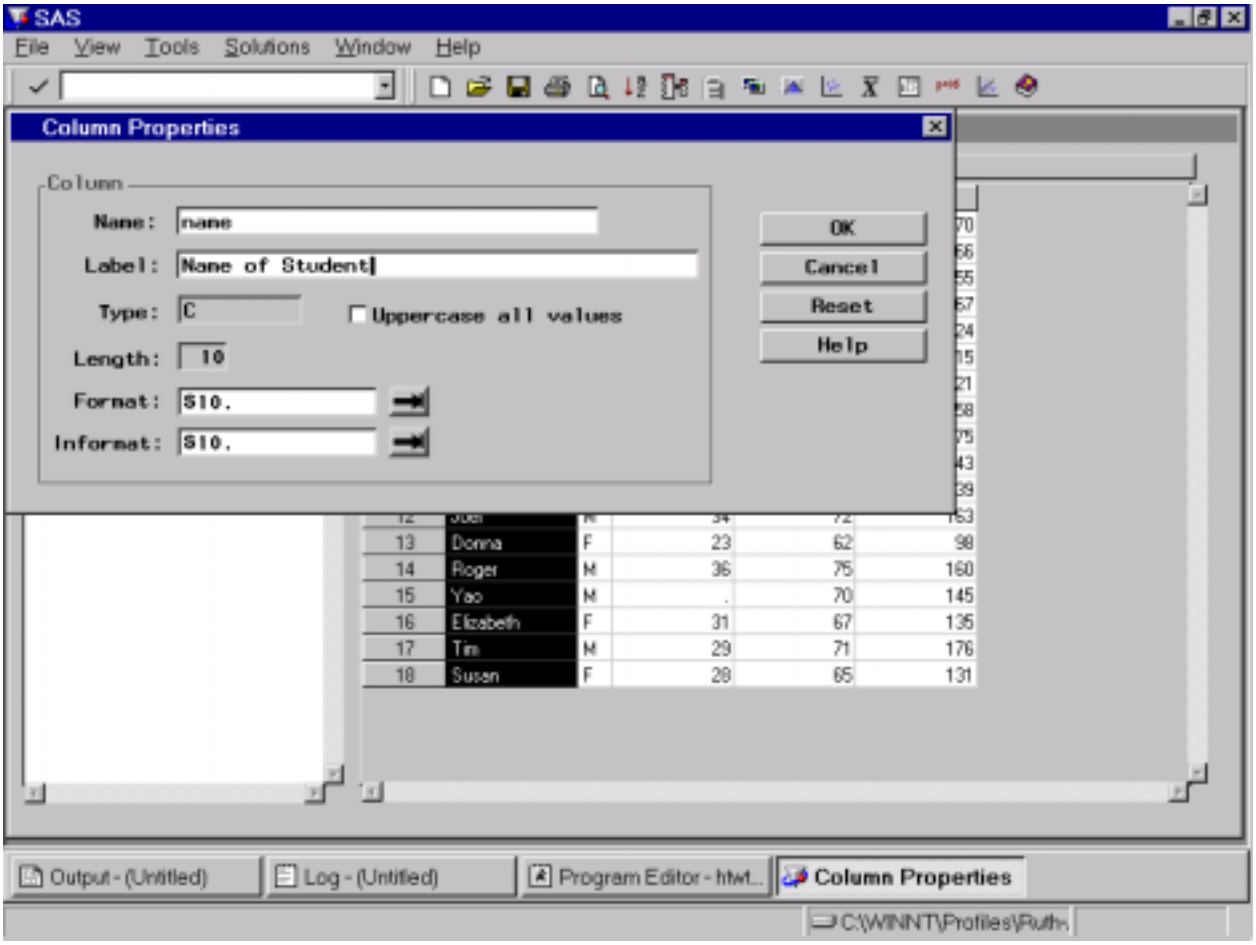
   This allows you to manipulate the data and analyze them in various ways.



2. **Add labels to the variable names:**
   More descriptive names (called labels) can be added for each variable. In this example, you will add labels to four variables as follows:

Variable Name	Label
Name	Name of Student
Weight	Weight in Pounds
Sex	Gender of Student
Age	*Age* of Student

- Right-click on the desired column and select **PROPERTIES**. In the dialogue window that pops up, type in the appropriate label. Click **OK.** Do this for each of the four variables.



3. **Work with only the first ten observations**:

   - **Select rows 11-18:** Click on the button that identifies row 11. Hold down the shift key and click on the button that identifies row 18. From the top of the screen, select **Edit → DELETE.**

4. **Print a list of the data values for *sex, age* and *weight***:

   You can print your data in a listing report. You can also specify the variables to be included in the report and some details about the report format.

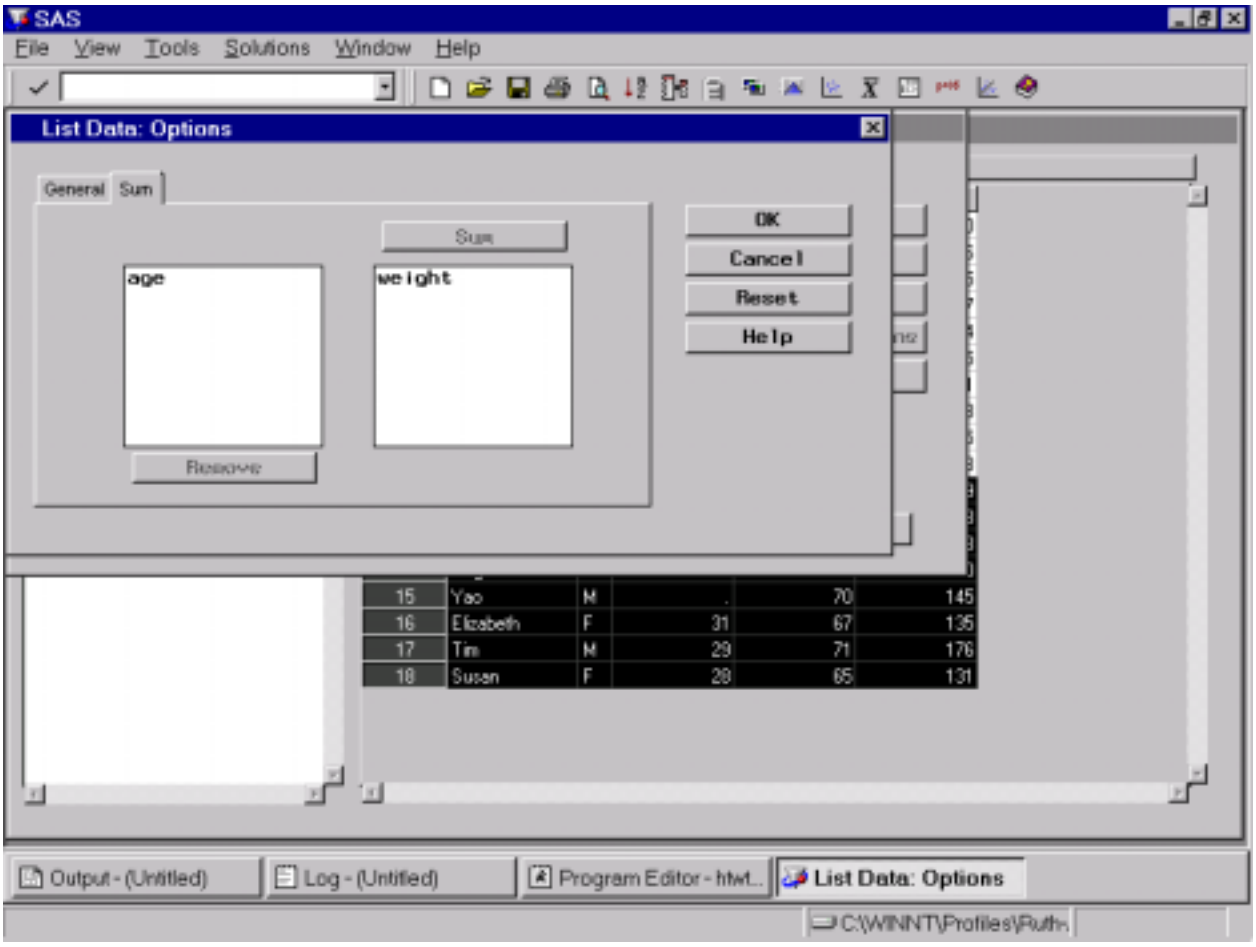
   - From the menu at the top of the screen, select **REPORTS → LIST DATA**. In the dialogue window that pops up, select ***sex, age,*** and ***weight*** in the **Candidate Box** and click on **PRINT**.

5. **Identify records by *name*** rather than by observation number:

   - Select ***name*** in the **Candidate Box** and click on **ID**.

6. **Sum the values for the *weight* variable:**

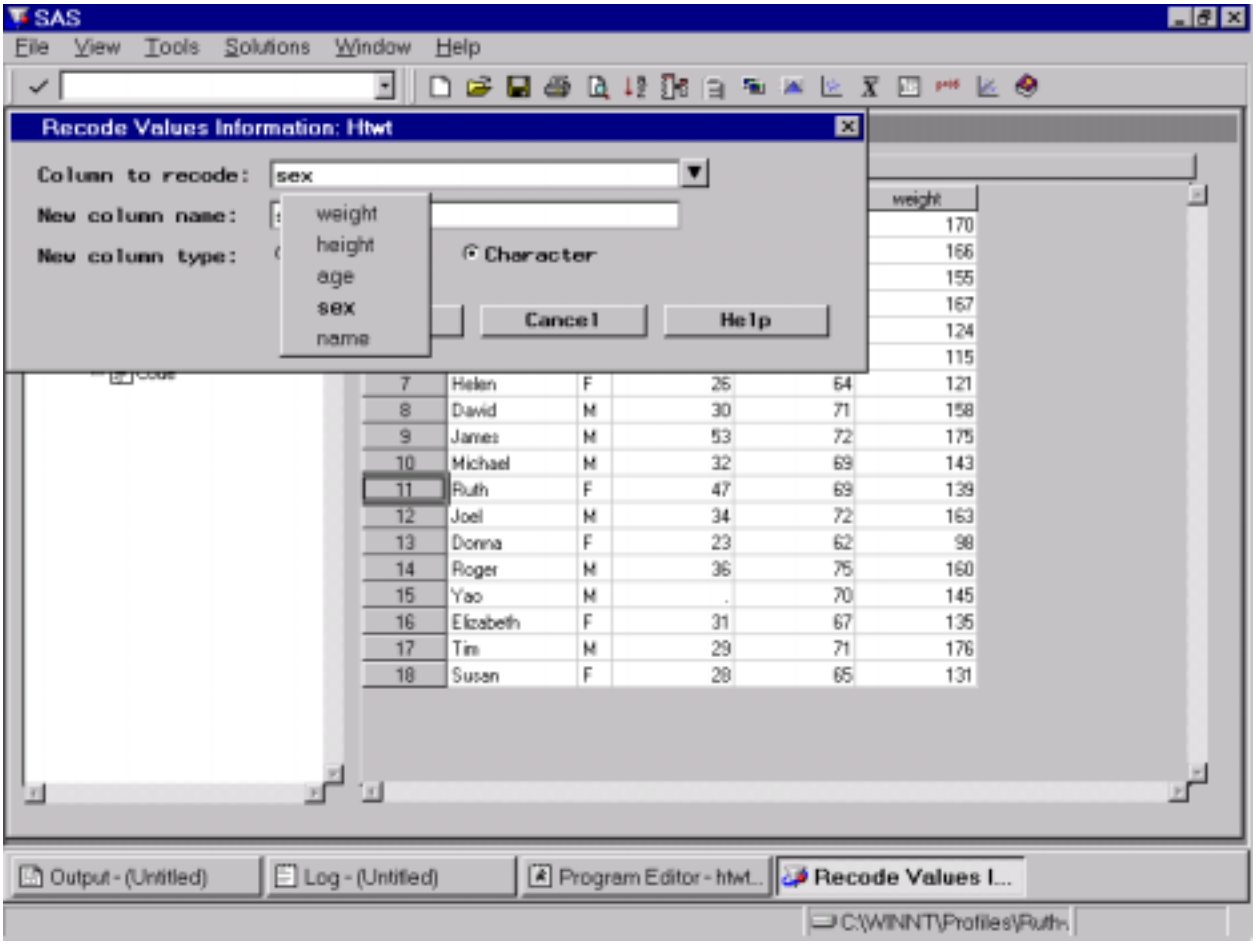
   - Click on **OPTIONS → SUM tab → WEIGHT → OK.**



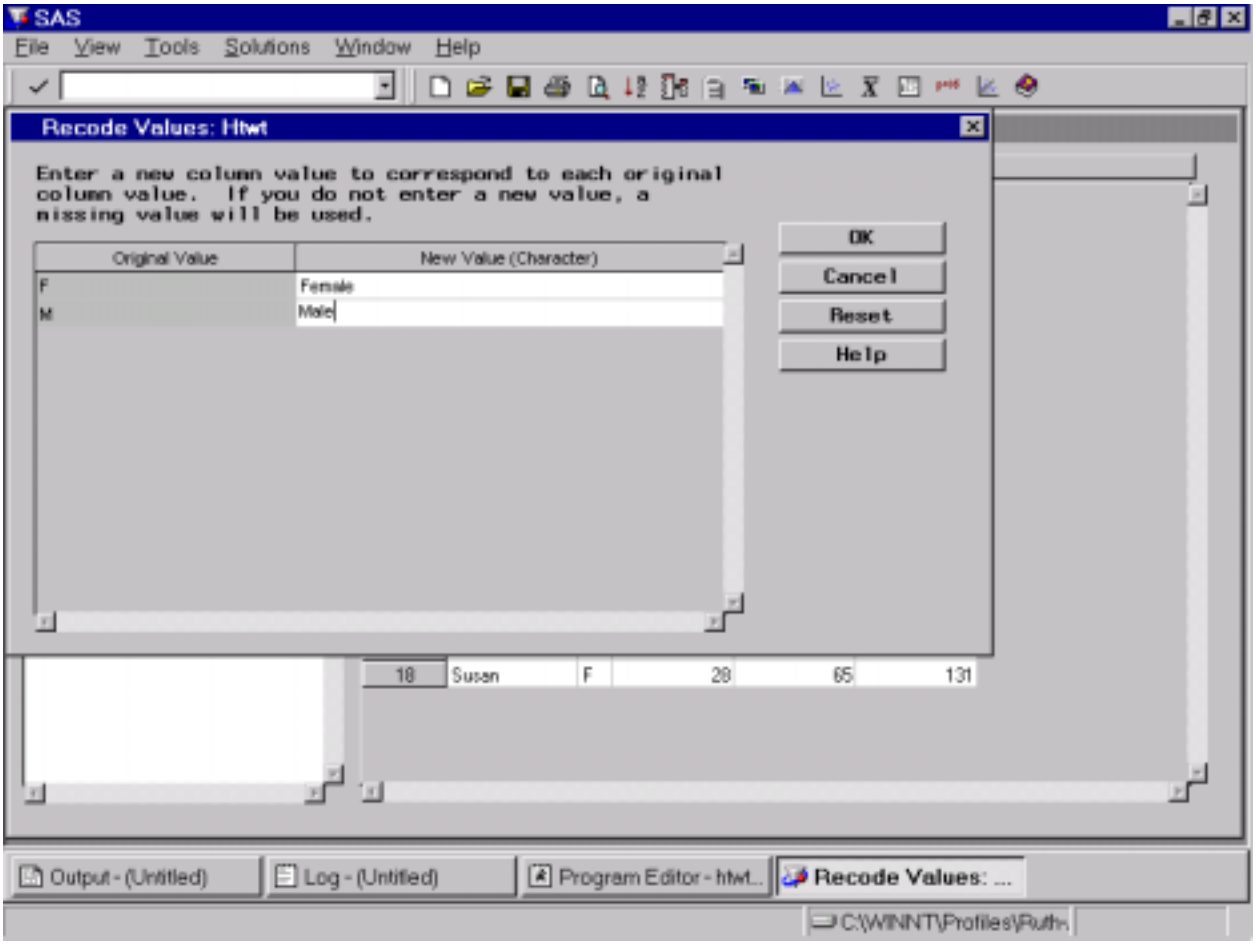
7. **Display *sex* with values of *male* and *female*** rather than with *M* and *F*:

   - From the menu at the top of the screen select **DATA → TRANSFORM → RECODE VALUES**

   - In the dialogue window that pops up, specify the variable to be recoded by clicking in the **COLUMN TO BE RECODED** line and selecting **SEX** from the list of variables. Since ***sex*** is a character variable, select this option under **COLUMN TYPE**.



- In the new dialogue box that opens, type in ***female*** and ***male*** beside F and M, respectively. Click **OK**. This will add a new column in the data listing entitled Sex_recoded.



Note:

If you want to change the name of the new column, right-click on new column and select **PROPERTIES.** In the dialogue window that pops up, replace the column name with the new name of your choice (e.g. *sex2*). Click **OK**

If you want to change the order of the columns, right-click on any column title and select **MOVE** from the pull-down menu. In the dialogue window that pops up, highlight the variable you wish to move and click on the up or down arrows as desired. As you click on the arrow, the variable will move up or down the list, as appropriate. Click **OK** when finished re-ordering the columns.

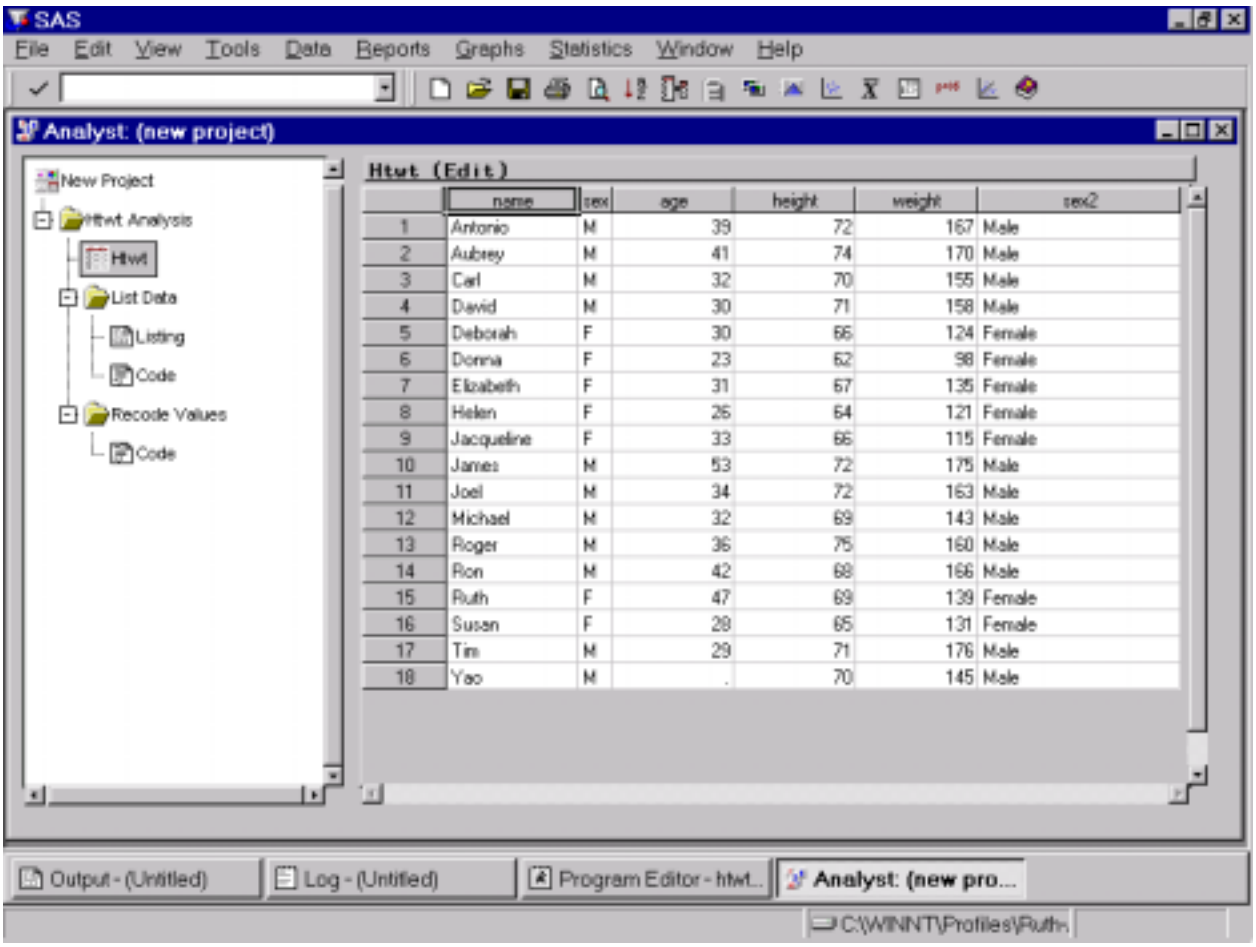
Example 3: Sorting the Data

These examples show you how to sort the data and then create a listing of the values of specified variables.

1. **Sort the data:**
   - Right-click on the column to be sorted (in this case, *name*) and select **SORT** from the pull-down menu.

   - If it is not selected already, highlight ***name*** in the **Candidate Box** and click on **SORT BY**. Click on the **ASCENDING/DESCENDING** button to choose the sort order. Click **OK** when finished.

     The data in the data list will now be listed in alphabetical order of name.



   - To sort by two variables, in the data listing, highlight both columns, right-click on them and select **SORT**. Follow the above instructions to sort them.

2. **List Data:**

   - Generate a listing report for ***sex*** and ***age*** and identify each observation by ***name***: From the menu at the top of the screen, select **REPORTS → LIST DATA**
   - In the dialogue window that pops up, select ***sex*** and ***age*** from the **Candidate Box** and click on **PRINT**.
   - Then select ***name*** from the **Candidate Box** and click on **ID**.

- Add titles to the report: Click on the **TITLES** button. Type in the following titles, one per line:
  View the Data: Example 3
  Data List for Sex and Age
  Sorted by Name
  Click **OK** when finished.

- The output should look like:

View the Data: Example 3 4
Data List for sex and age 10:54 Tuesday, August 22, 2000
Sorted by NAME

name	sex	age
Antonio	M	39
Aubrey	M	41
Carl	M	32
David	M	30
Deborah	F	30
Donna	F	23
Elizabeth	F	31
Helen	F	26
Jacqueline	F	33
James	M	53
Joel	M	34
Michael	M	32
Roger	M	36
Ron	M	42
Ruth	F	47
Susan	F	28
Tim	M	29
Yao	M	.

* **Note:** This does not need to be saved

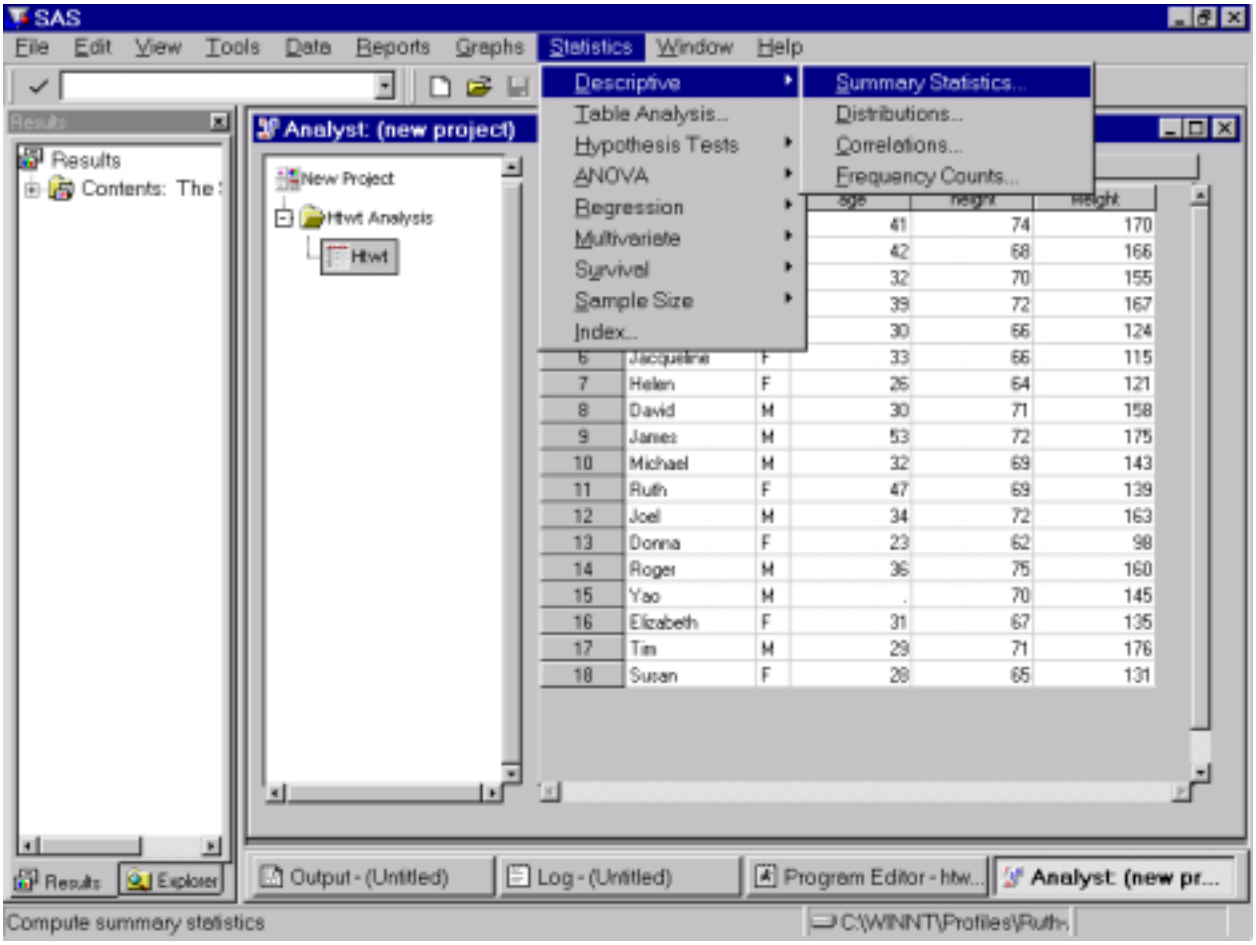
III. Explore the Data: Statistics for Numeric Data

Example 1: Calculating Means

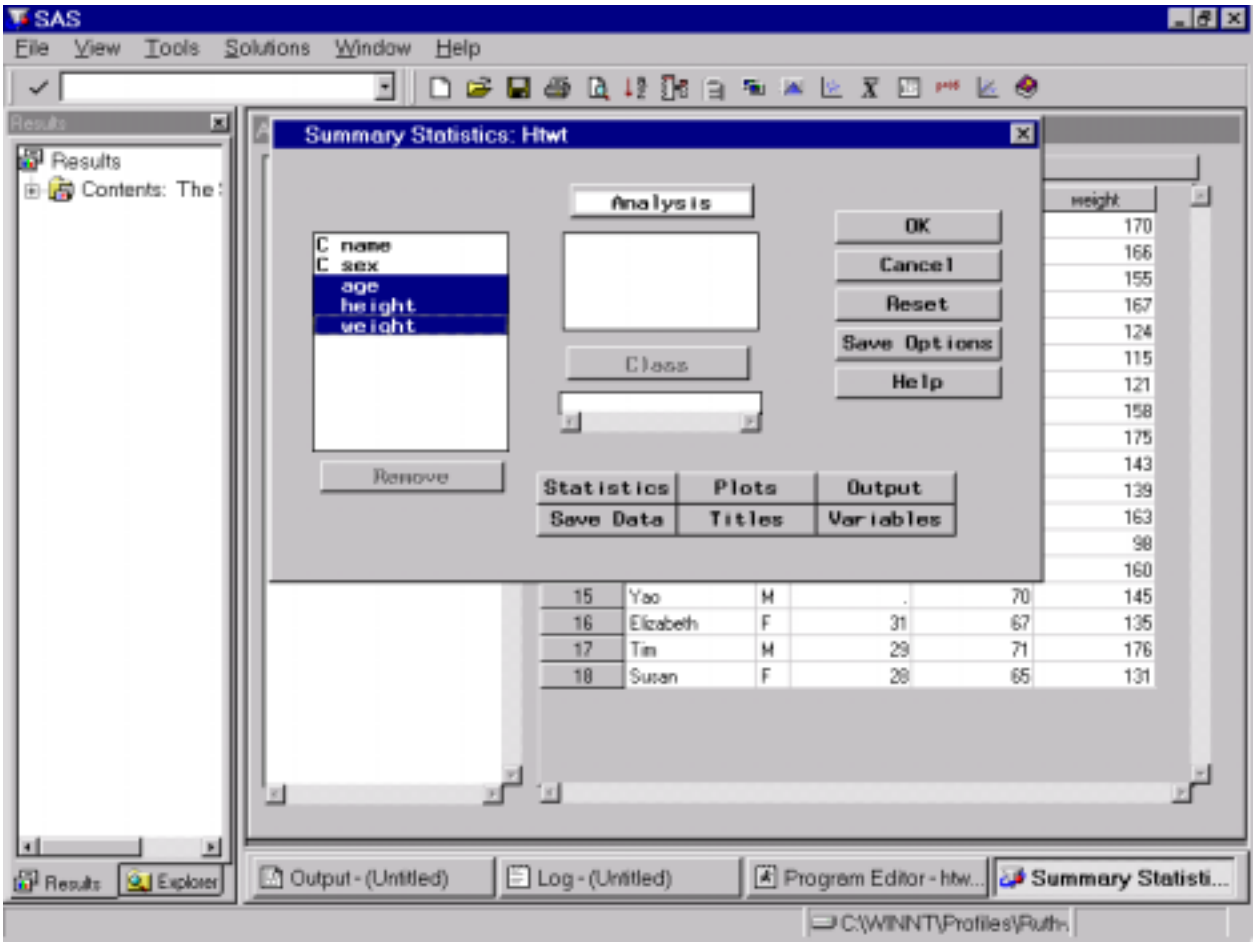
* If starting a brand new session, access the height/weight data set in Analyst:

- From the menu at the top of the screen, select **FILE → OPEN BY SAS NAME → WORK → HTWT → OK**.

1. **Generate descriptive statistics:**
   - From the menu at the top of the screen, select **STATISTICS → DESCRIPTIVE → SUMMARY STATISTICS**

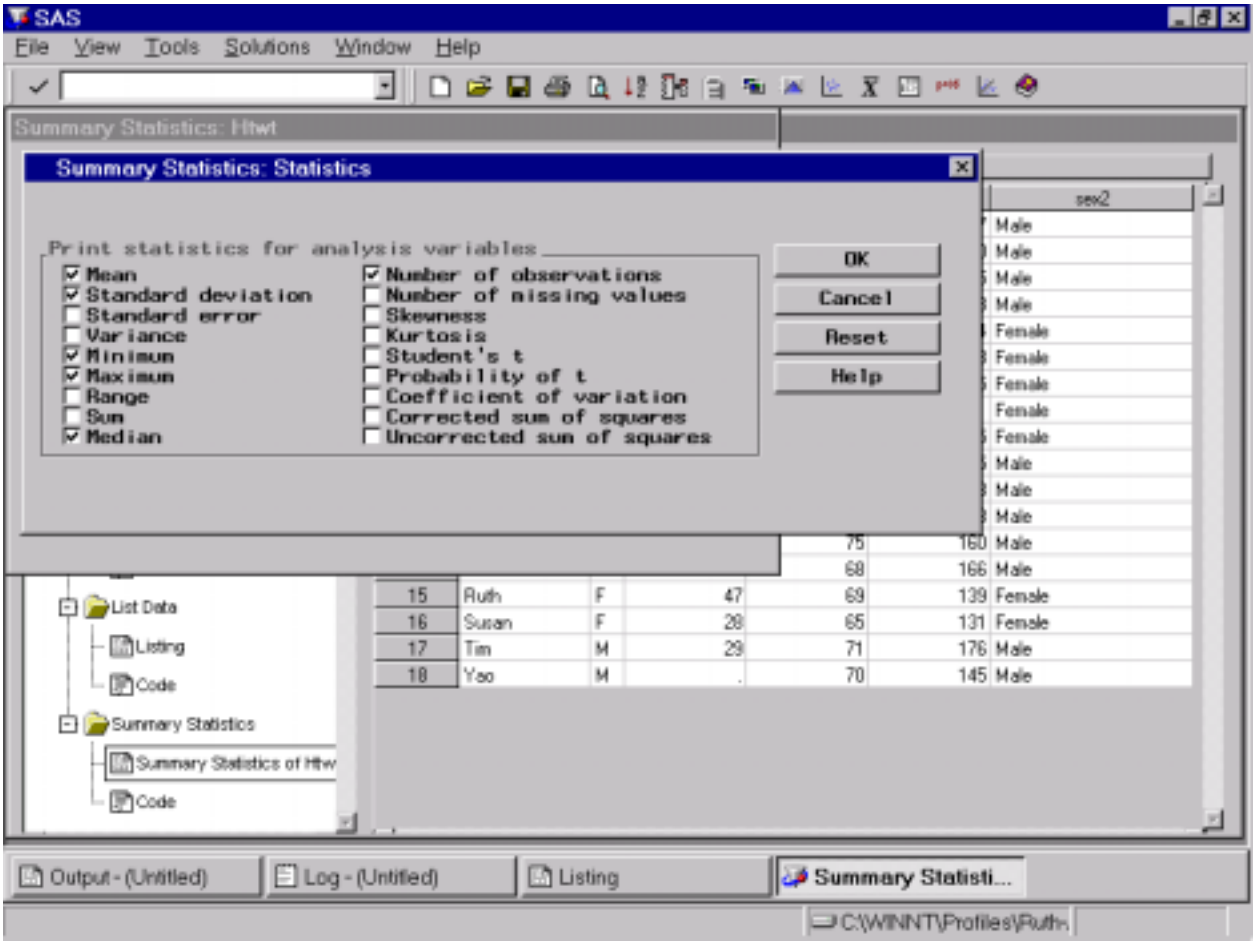


- From the **Candidate Box**, select the variables for which you wish to generate statistics (***age***, ***weight***, and ***height***) and click on **ANALYSIS**



- Click on the **STATISTICS** button to select the statistics you want generated.

  The basic statistics most often chosen are: *mean, standard deviation, minimum, maximum, and number of observations.* Click **OK** when finished.



2. Add titles to the output:
   - Click on the **TITLES** button and type in the titles, one per line:
     View the data: Example 1
     Calculating Means: Age, Weight & Height

   Click **OK** when finished.

3. Click **OK**.

The output should look like:

```
          View the Data: Example 1                                  4
         Data List for sex and age      10:54 Tuesday, August 22, 2000

                              Sorted by NAME

                        name          sex             age

                        Antonio        M               39
                        Aubrey         M               41
                        Carl           M               32
                        David          M               30
                        Deborah        F               30
                        Donna          F               23
                        Elizabeth      F               31
                        Helen          F               26
                        Jacqueline     F               33
                        James          M               53
                        Joel           M               34
                        Michael        M               32
                        Roger          M               36
                        Ron            M               42
                        Ruth           F               47
                        Susan          F               28
                        Tim            M               29
                        Yao            M                .

                   View the Data: Example 1                                  5
           Calculating Means: Age, Weight & Height
                                           10:54 Tuesday, August 22, 2000

                     The MEANS Procedure

Variable     N            Mean         Std Dev         Minimum         Maximum
ƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒ
age         17      34.4705882       7.7630346      23.0000000      53.0000000
height      18      69.0555556       3.5225696      62.0000000      75.0000000
weight      18     146.7222222      22.5409576      98.0000000     176.0000000
ƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒ
```

Example 2: Calculating Means: CLASS Option

This example demonstrates how to generate descriptive statistics with other options.

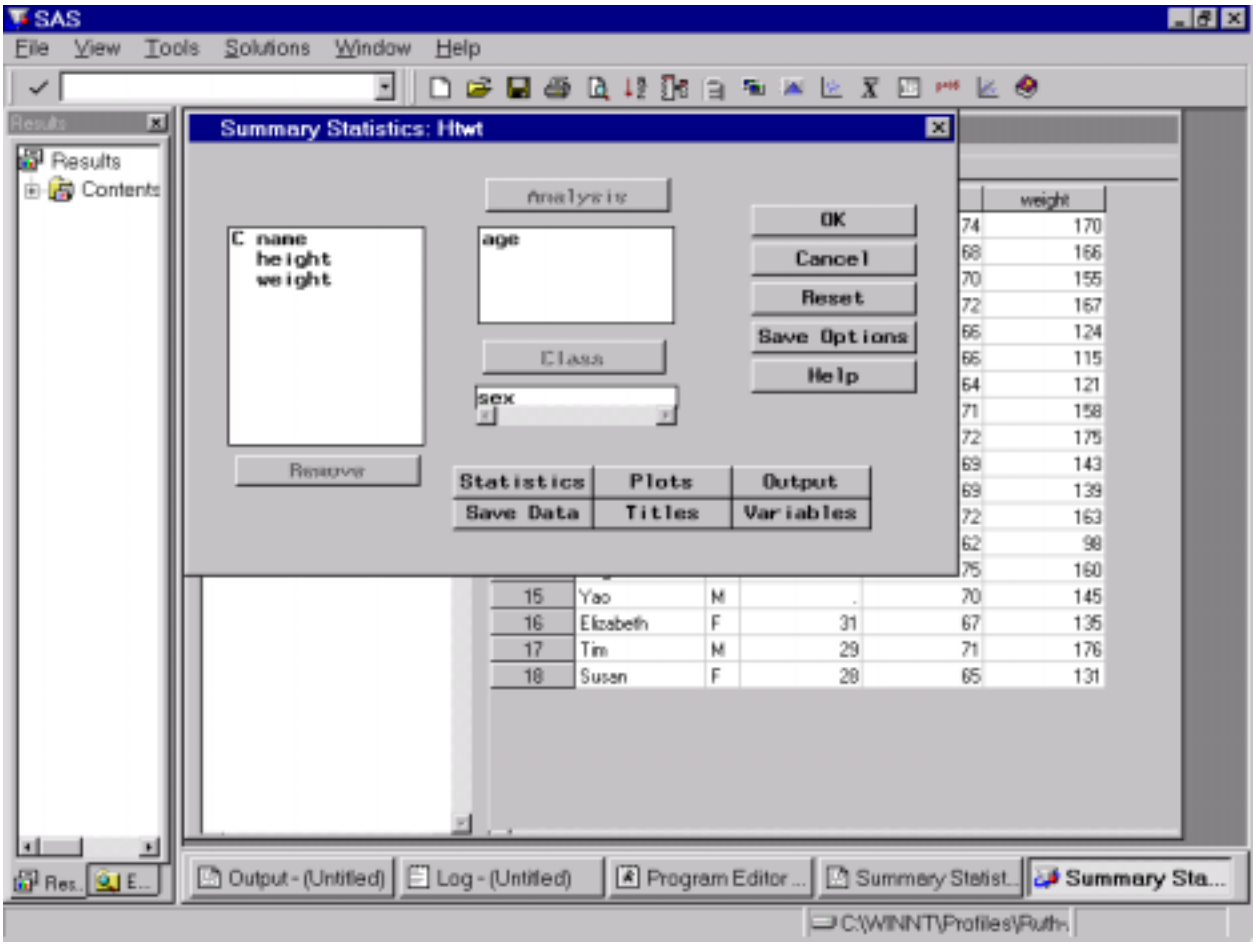
***** If starting a brand new session, access the height/weight data set in Analyst:

> From the menu at the top of the screen, select **FILE → OPEN BY SAS NAME → WORK → HTWT → OK**.

1. **Generate descriptive statistics:**
   - From the menu at the top of the screen, select **STATISTICS → DESCRIPTIVE → SUMMARY STATISTICS**
   - If there are any variables listed in the **ANALYSIS** box, remove them (highlight them and click the **REMOVE** button)

- The analysis in this example applies only to the ***age*** variable, so it should be selected as the only **ANALYSIS** variable.

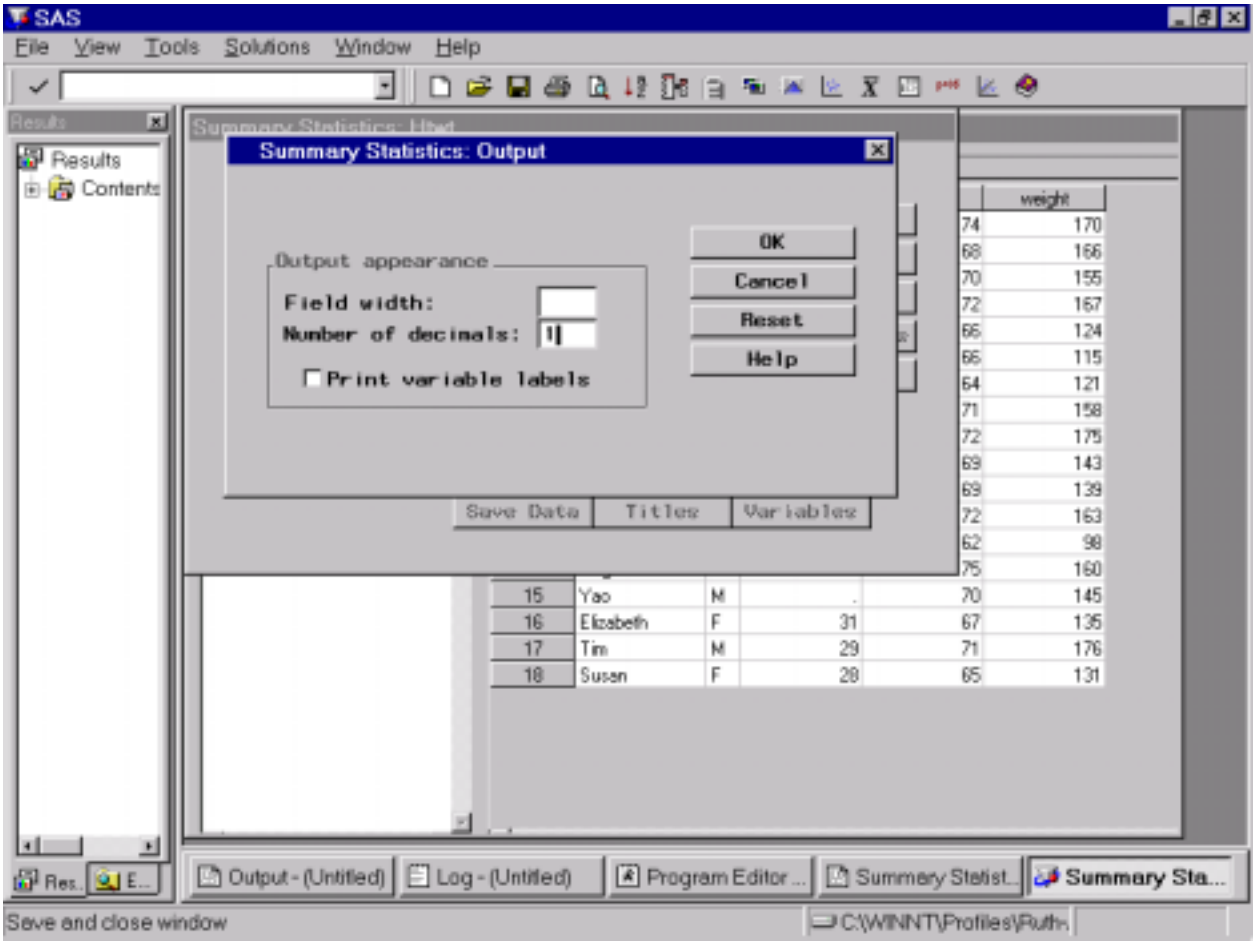
- The analysis will be separated by values of ***sex*** so this should be the chosen as the **CLASS** variable.



- To select the statistics you want to generate click on the **STATISTICS** button.

  The basic statistics are: Mean, Minimum, Maximum, Number of observations, Number of missing values and sum. Click **OK** when finished.

- Specify the number of decimal places: click on the **OUTPUT** button. Type **1** for number of decimals. Click **OK**.



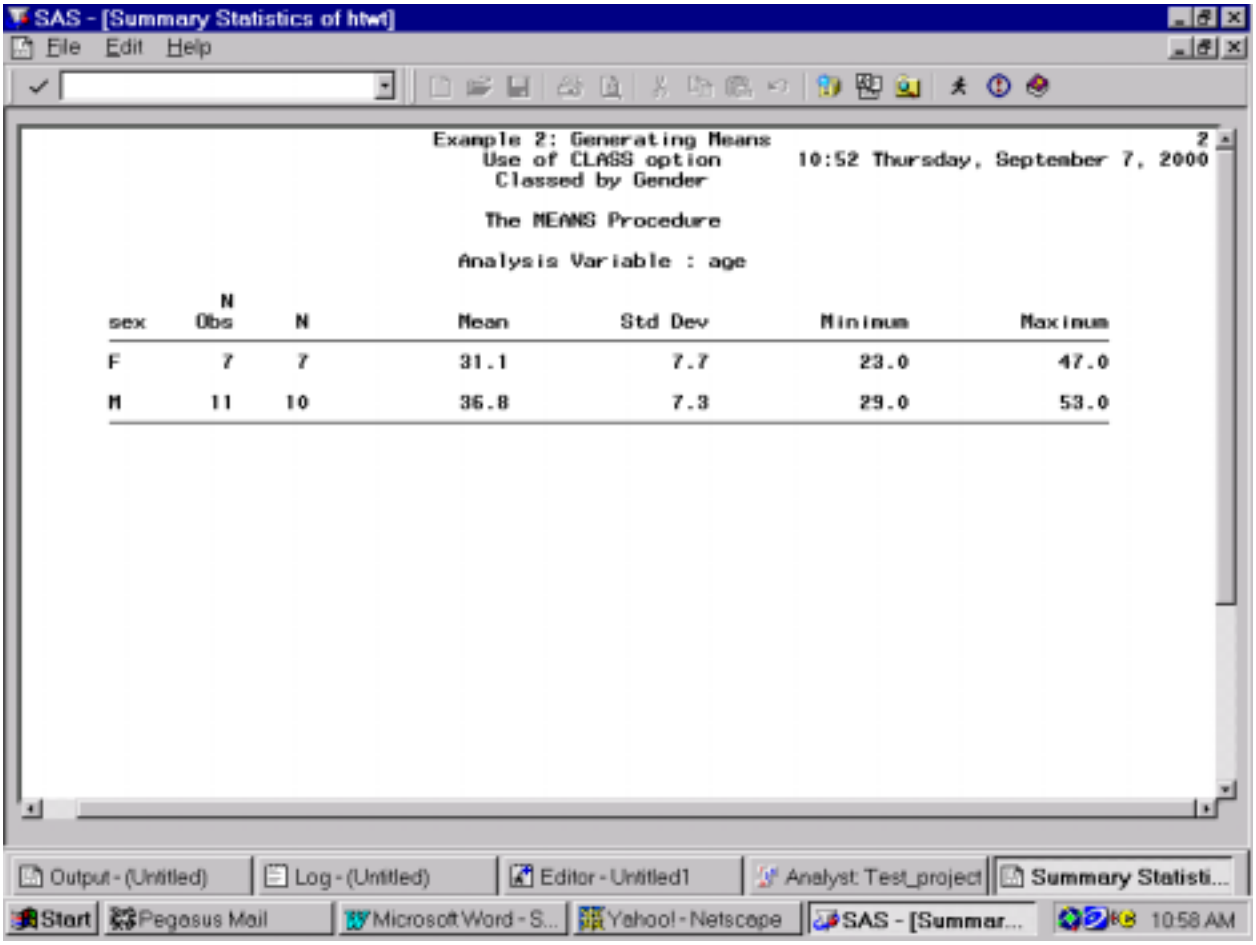
2. Add titles to the output report: click on the **TITLES** button. Type in the titles, one per line:

   Example 2: Generating Means
   Use of CLASS option
   Classed by Gender

   Click **OK**

3. Click **OK**.

Output should look like:

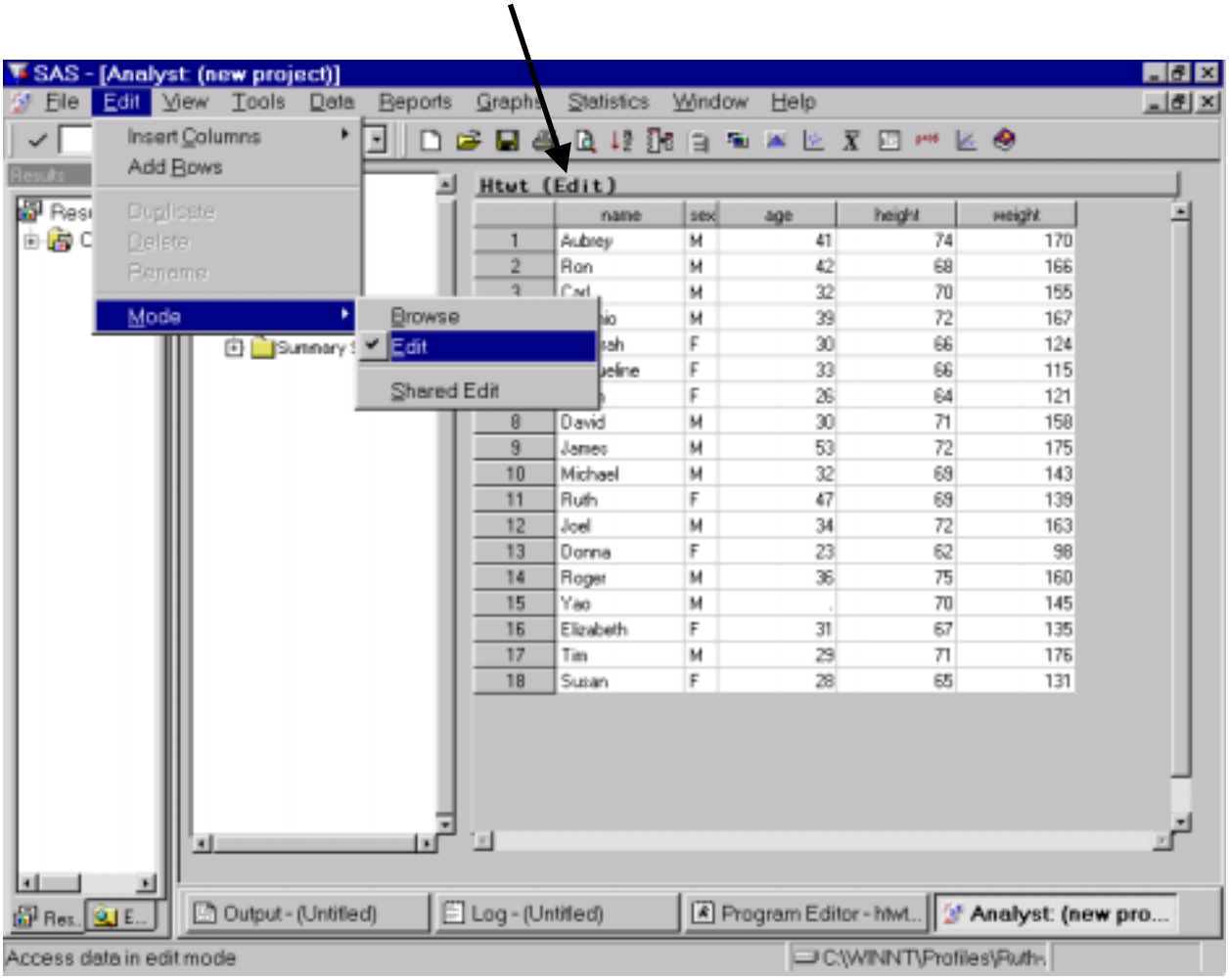

SAS - [Summary Statistics of htwt]
File Edit Help
Example 2: Generating Means
Use of CLASS option
Classed by Gender
2
10:52 Thursday, September 7, 2000
The MEANS Procedure
Analysis Variable : age
sex | N Obs | N | Mean | Std Dev | Minimum | Maximum
F | 7 | 7 | 31.1 | 7.7 | 23.0 | 47.0
M | 11 | 10 | 36.8 | 7.3 | 29.0 | 53.0
Output - (Untitled)
Log - (Untitled)
Editor - Untitled1
Analyst Test_project
Summary Statisti...
Start
Pegasus Mail
Microsoft Word - S...
Yahoo! - Netscape
SAS - [Summar...
10:58 AM

Example 3: Generating Means: BY option

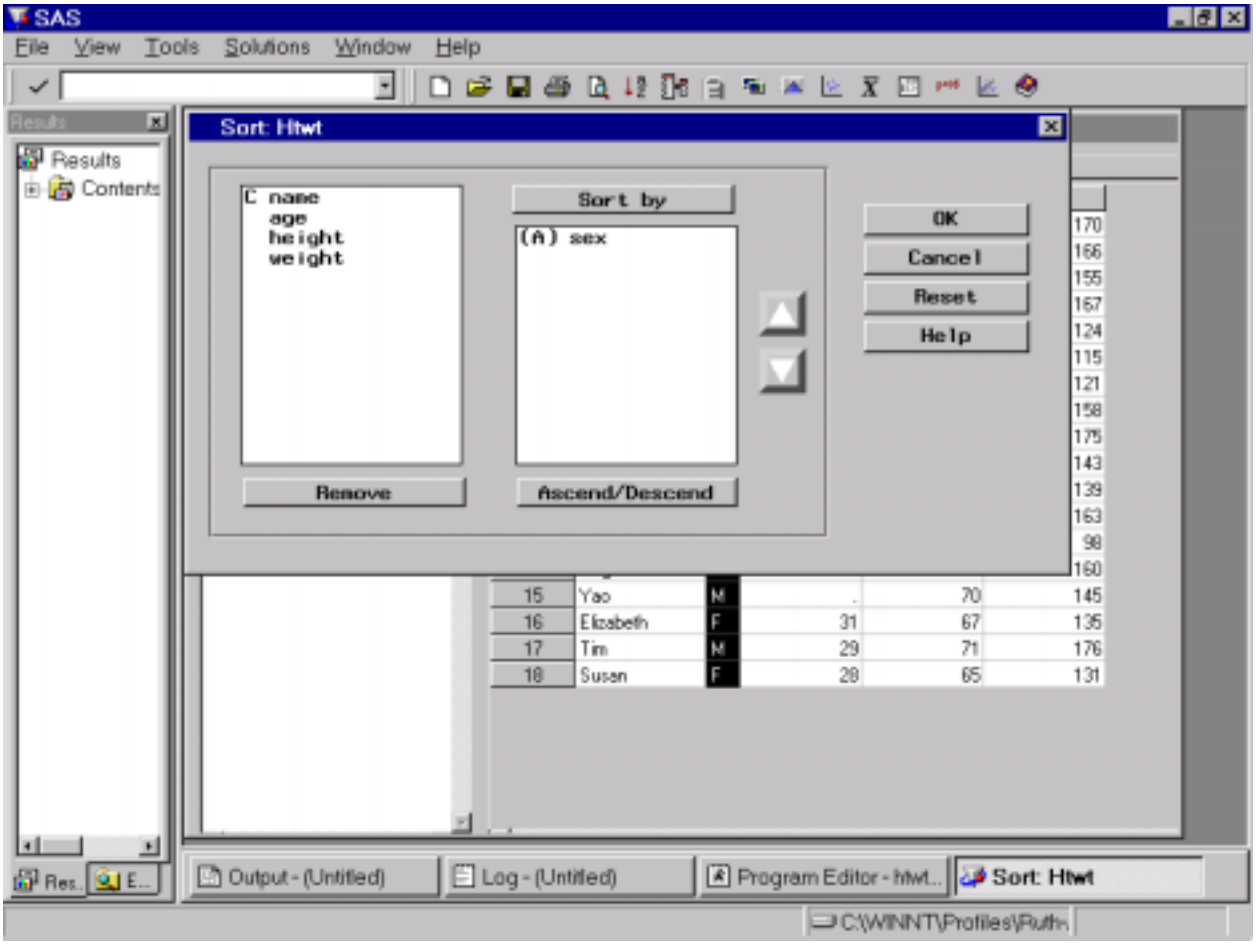
In this example, the data will first be sorted by sex.

1. In order to edit the data set, the edit mode must be changed from BROWSE to EDIT. From the menu at the top of the screen, select **EDIT → MODE → EDIT**.

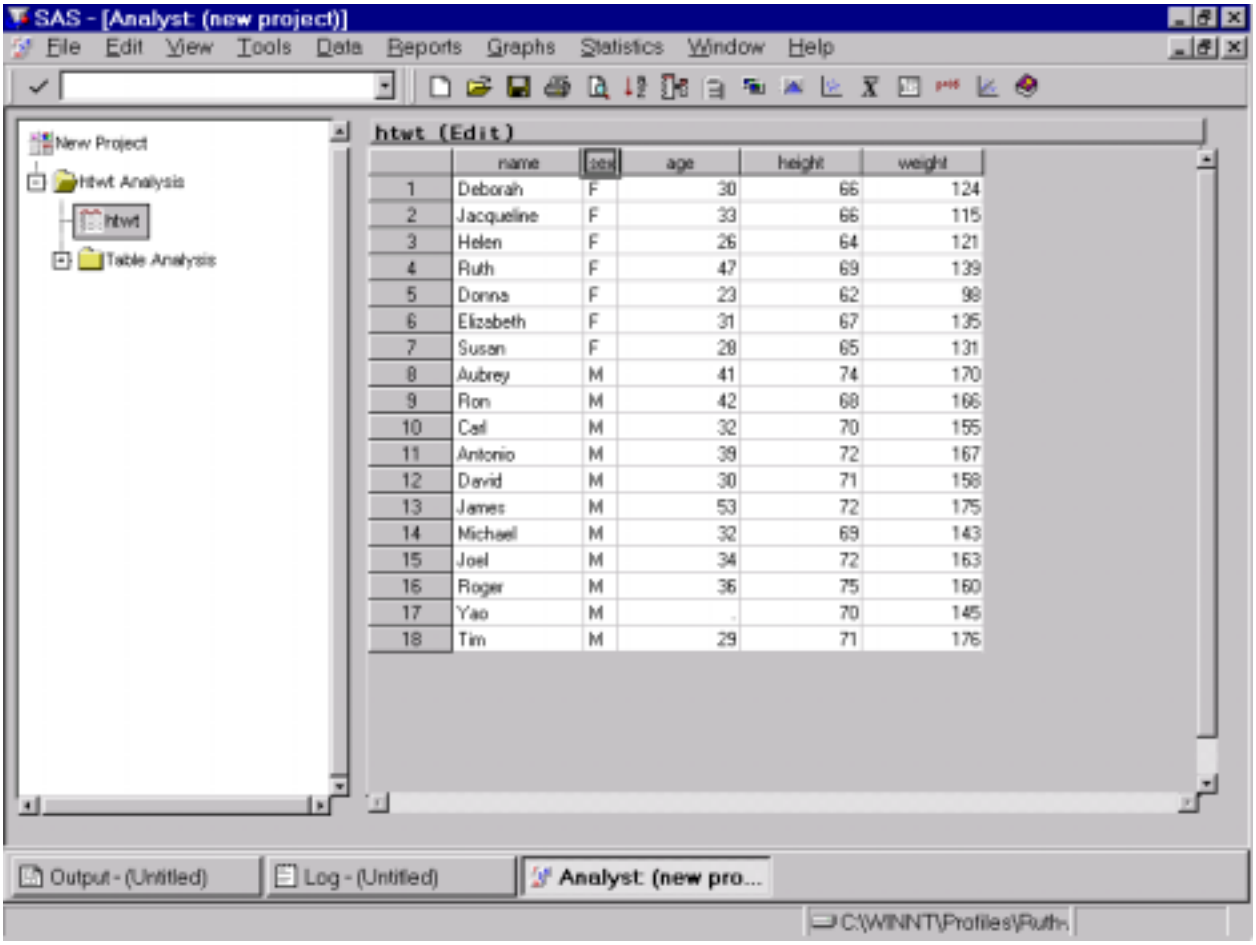
   Note that the current mode appears in parentheses beside the name of the data set.



2. To sort analyses by sex, right click on the top of the sex column and choose **SORT** from the pop-down menu. Ensure that it will be sorted in ascending order. Click **OK**:



In the list of data, all females should now appear at the beginning.

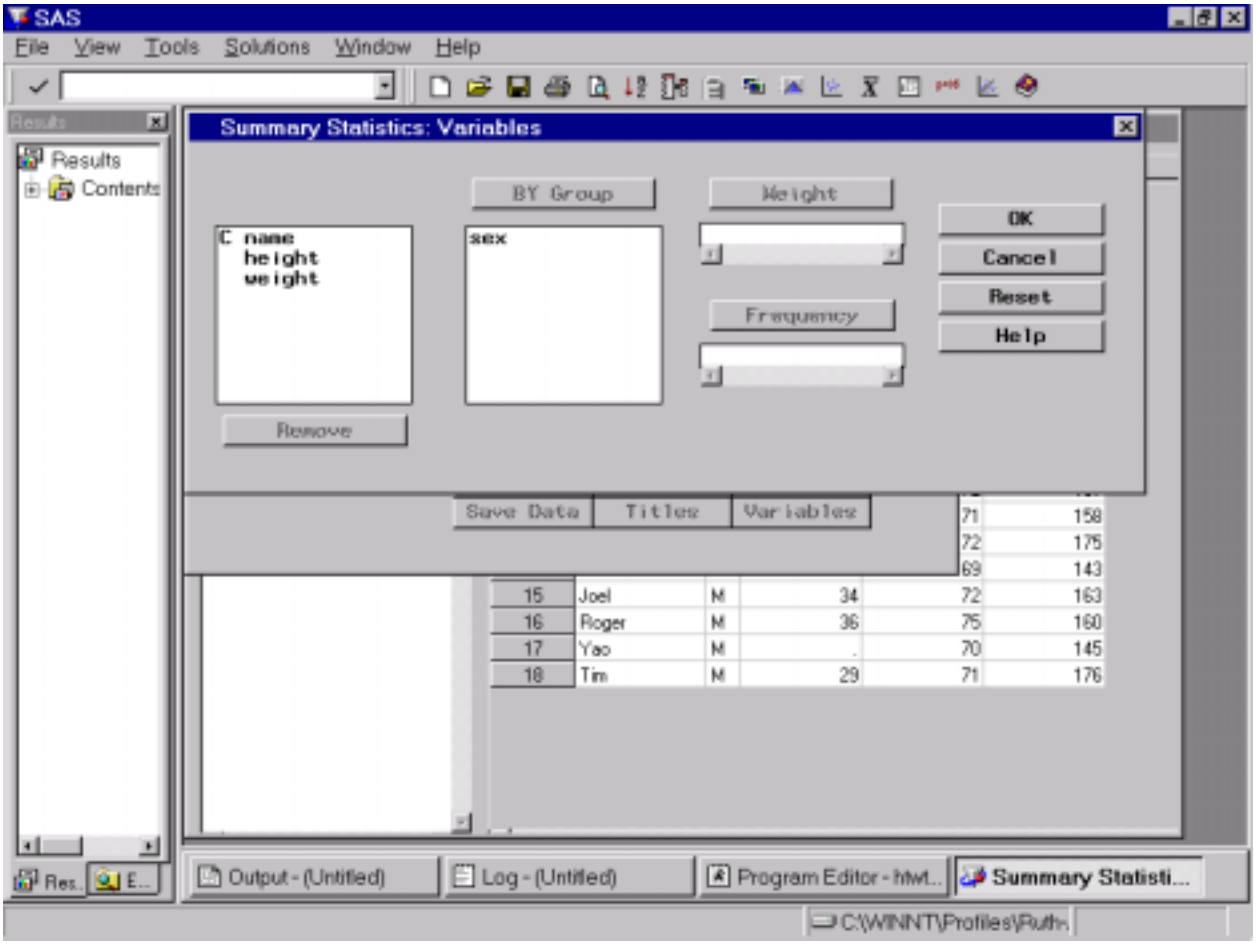


To separate output by sex:

3. From the menu at the top of the screen, select **STATISTICS → DESCRIPTIVE → SUMMARY STATISTICS**. Clear the **ANALYSIS** and **CLASS** boxes by highlighting the variable that appears in them and clicking on the **REMOVE** button

4. Select ***age*** as the **ANALYSIS** variable.

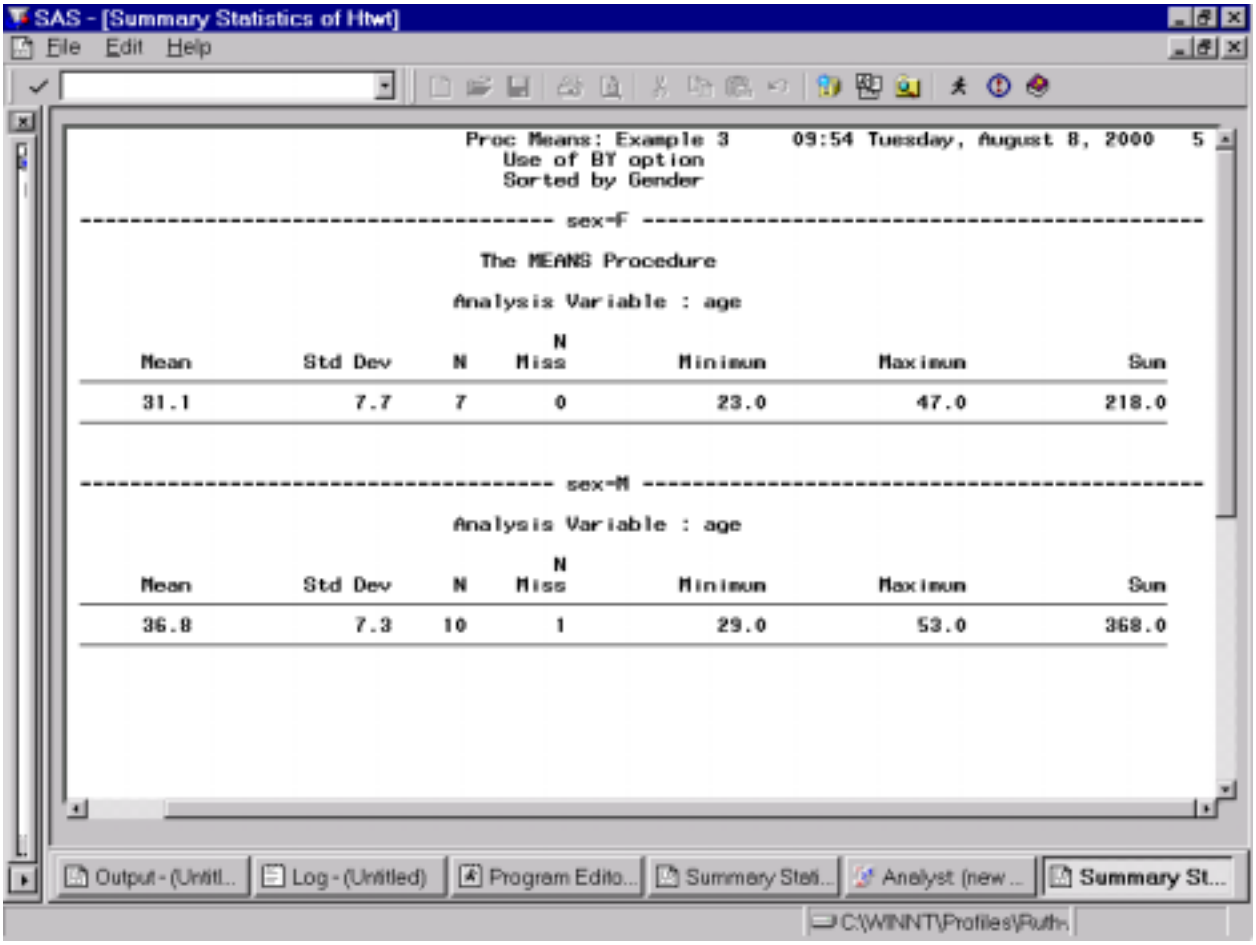
5. Click on the **VARIABLES** button. Highlight the *sex* variable and click on the **BY GROUP** button. Click **OK** when finished.



6. To change the title of the output report, click on the **TITLES** button. Up to three titles can be added - one in each line:

    1. Calculating Means: Example 3
    2. Sorted by Gender

Output will look like:



7. To generate more statistics on the *age* variable, go to the menu at the top of the screen, select **STATISTICS → DESCRIPTIVE → SUMMARY STATISTICS.** Click on the **STATISTICS** button in this window. The typical statistics chosen are:

   Mean, Standard deviation, Standard error, Variance, Minimum, Maximum, Range, Sum, Median, Number of observations, Skewness, Kurtosis, Student's t, Coefficient of variation, Corrected sum of squares, Uncorrected sum of squares

IV. Explore the Data: Creating tables

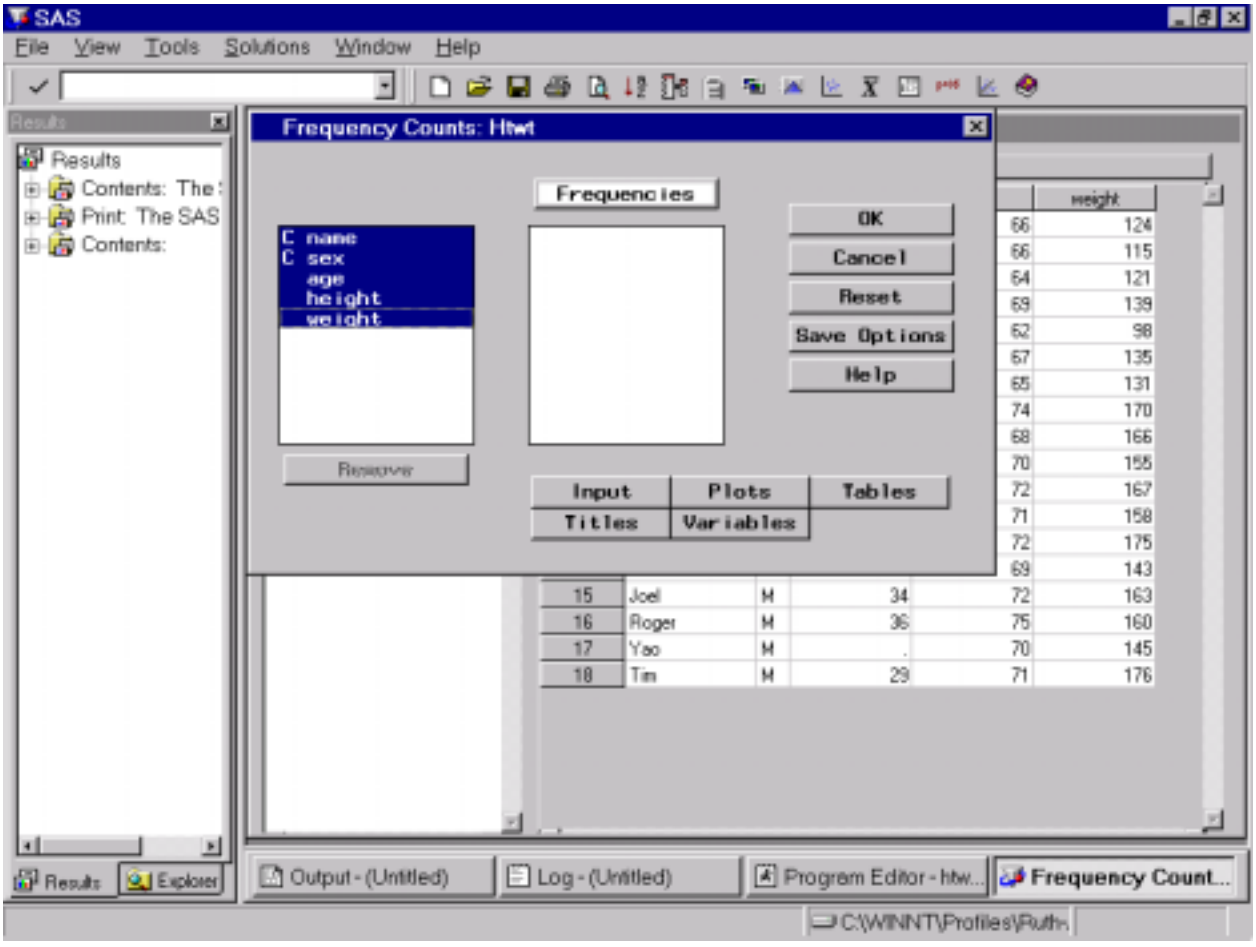
Calculating frequencies produces summary data in tabular form. Each of the following five examples show various options available with this procedure.

Example 1: Frequency Tables

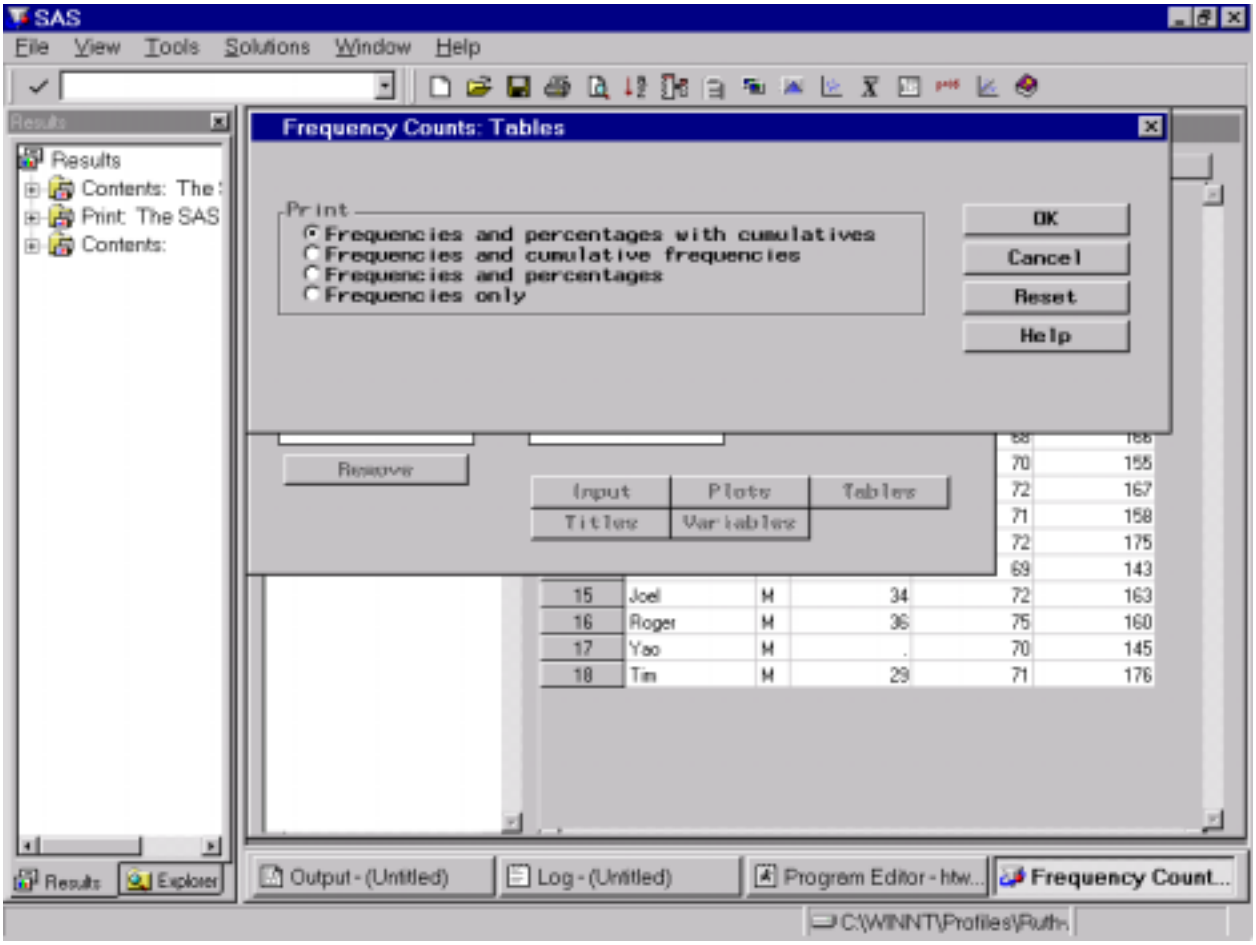
1. To generate frequency counts for each variable, from the menu at the top of the screen, select **STATISTICS → FREQUENCY COUNTS.**

2. In Box 1, highlight all variables and then click on the **FREQUENCIES** button.

   * To generate counts for only specific variables, highlight only those variables and then click on the FREQUENCIES button.



3. Click on the **TABLES** button to choose the counts you want displayed in the tables. Ensure that the first option is checked. Click **OK**.



4. To add a title to the table, click on the **TITLES** button. In the first line, type in: *Example 1: Frequency Table*. Click **OK** when finished.

   The output should look like this:

```
Example 1: Frequency Table                                  6
                                                  14:04 Wednesday, August 9, 2000

                              The FREQ Procedure

                                                    Cumulative    Cumulative
name          Frequency     Percent     Frequency      Percent
ffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffff
Antonio               1        5.56             1         5.56
Aubrey                1        5.56             2        11.11
Carl                  1        5.56             3        16.67
David                 1        5.56             4        22.22
Deborah               1        5.56             5        27.78
Donna                 1        5.56             6        33.33
Elizabeth             1        5.56             7        38.89
Helen                 1        5.56             8        44.44
Jacqueline            1        5.56             9        50.00
James                 1        5.56            10        55.56
Joel                  1        5.56            11        61.11
Michael               1        5.56            12        66.67
```

Roger	1	5.56	13	72.22
Ron	1	5.56	14	77.78
Ruth	1	5.56	15	83.33
Susan	1	5.56	16	88.89
Tim	1	5.56	17	94.44
Yao	1	5.56	18	100.00

sex	Frequency	Percent	Cumulative Frequency	Cumulative Percent
F	7	38.89	7	38.89
M	11	61.11	18	100.00

Example 1: Frequency Table 7

14:04 Wednesday, August 9, 2000

The FREQ Procedure

age	Frequency	Percent	Cumulative Frequency	Cumulative Percent
23	1	5.88	1	5.88
26	1	5.88	2	11.76
28	1	5.88	3	17.65
29	1	5.88	4	23.53
30	2	11.76	6	35.29
31	1	5.88	7	41.18
32	2	11.76	9	52.94
33	1	5.88	10	58.82
34	1	5.88	11	64.71
36	1	5.88	12	70.59
39	1	5.88	13	76.47
41	1	5.88	14	82.35
42	1	5.88	15	88.24
47	1	5.88	16	94.12
53	1	5.88	17	100.00

Frequency Missing = 1

height	Frequency	Percent	Cumulative Frequency	Cumulative Percent
62	1	5.56	1	5.56
64	1	5.56	2	11.11
65	1	5.56	3	16.67
66	2	11.11	5	27.78
67	1	5.56	6	33.33
68	1	5.56	7	38.89
69	2	11.11	9	50.00
70	2	11.11	11	61.11
71	2	11.11	13	72.22
72	3	16.67	16	88.89
74	1	5.56	17	94.44
75	1	5.56	18	100.00

Example 1: Frequency Table 8
14:04 Wednesday, August 9, 2000

The FREQ Procedure

weight	Frequency	Percent	Cumulative Frequency	Cumulative Percent
98	1	5.56	1	5.56
115	1	5.56	2	11.11
121	1	5.56	3	16.67
124	1	5.56	4	22.22
131	1	5.56	5	27.78
135	1	5.56	6	33.33
139	1	5.56	7	38.89
143	1	5.56	8	44.44
145	1	5.56	9	50.00
155	1	5.56	10	55.56
158	1	5.56	11	61.11
160	1	5.56	12	66.67
163	1	5.56	13	72.22
166	1	5.56	14	77.78
167	1	5.56	15	83.33
170	1	5.56	16	88.89
175	1	5.56	17	94.44
176	1	5.56	18	100.00

Example 2: Creating 2-way tables

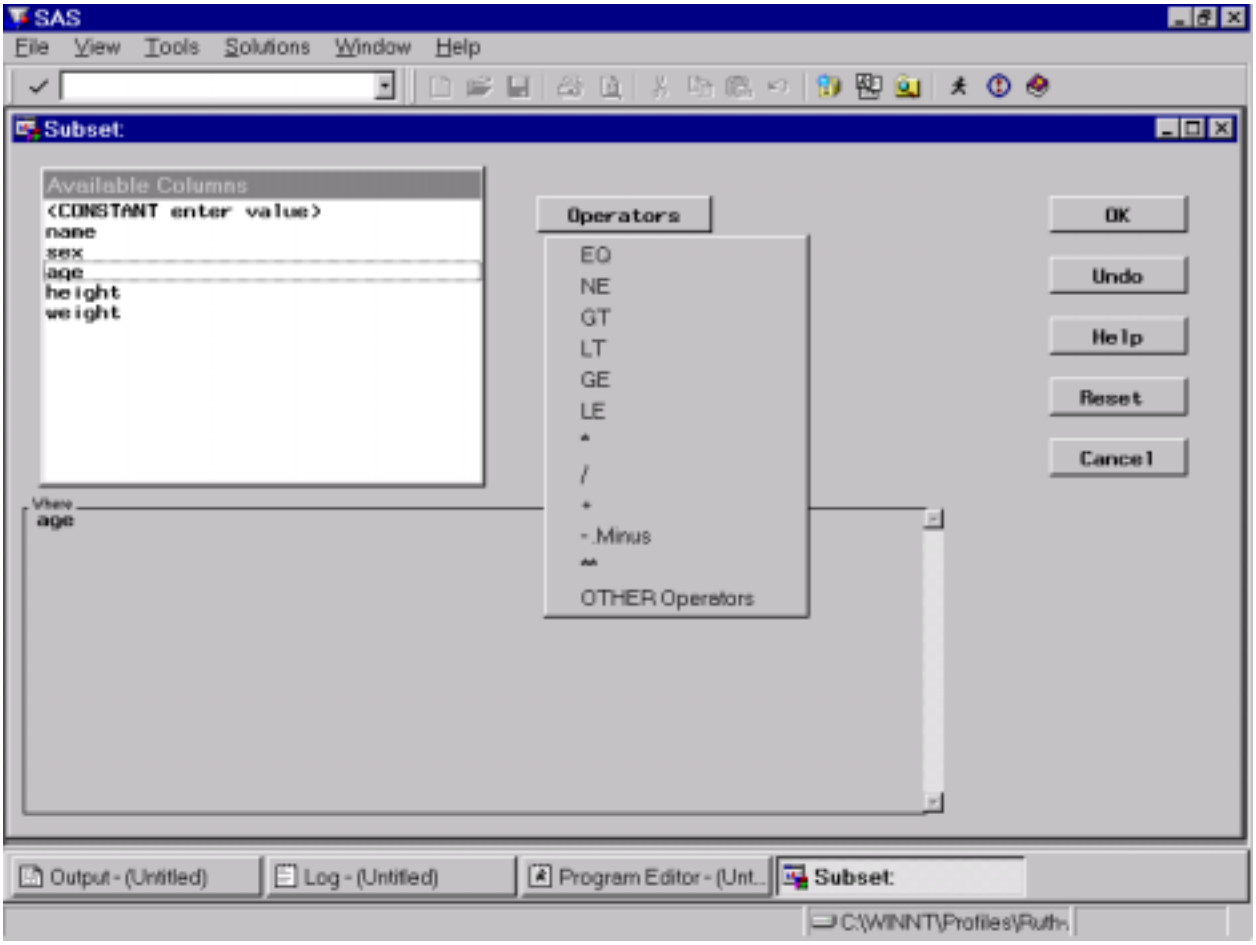
This example produces a two-way table (cross-tab) from a subset of the data.

1. Generate a cross-tab (*sex* BY *age*) only for records where *age* is between 0 and 29. From the menu at the top of the screen, select **DATA → FILTER→ SUBSET DATA**.

2. In the dialogue window that pops up, select the ***age*** variable from the **AVAILABLE COLUMNS**. It automatically appears in the **WHERE** box of the window. In this box you will write a statement identifying the conditions that determine which records will be included (i.e. you specify: $0 \leq age \geq 29$).

3. When you select *age*, a list of possible OPERATORS also appears. These operators help set the conditions for the *age* variable: EQ stands for Equal To
   NE stands for Not Equal To
   GT stands for Greater Than
   and so on…

Because one record in the data set is missing the *age* variable, you must first specify that this record is not to be included. Select **OTHER Operators** to extend the list.

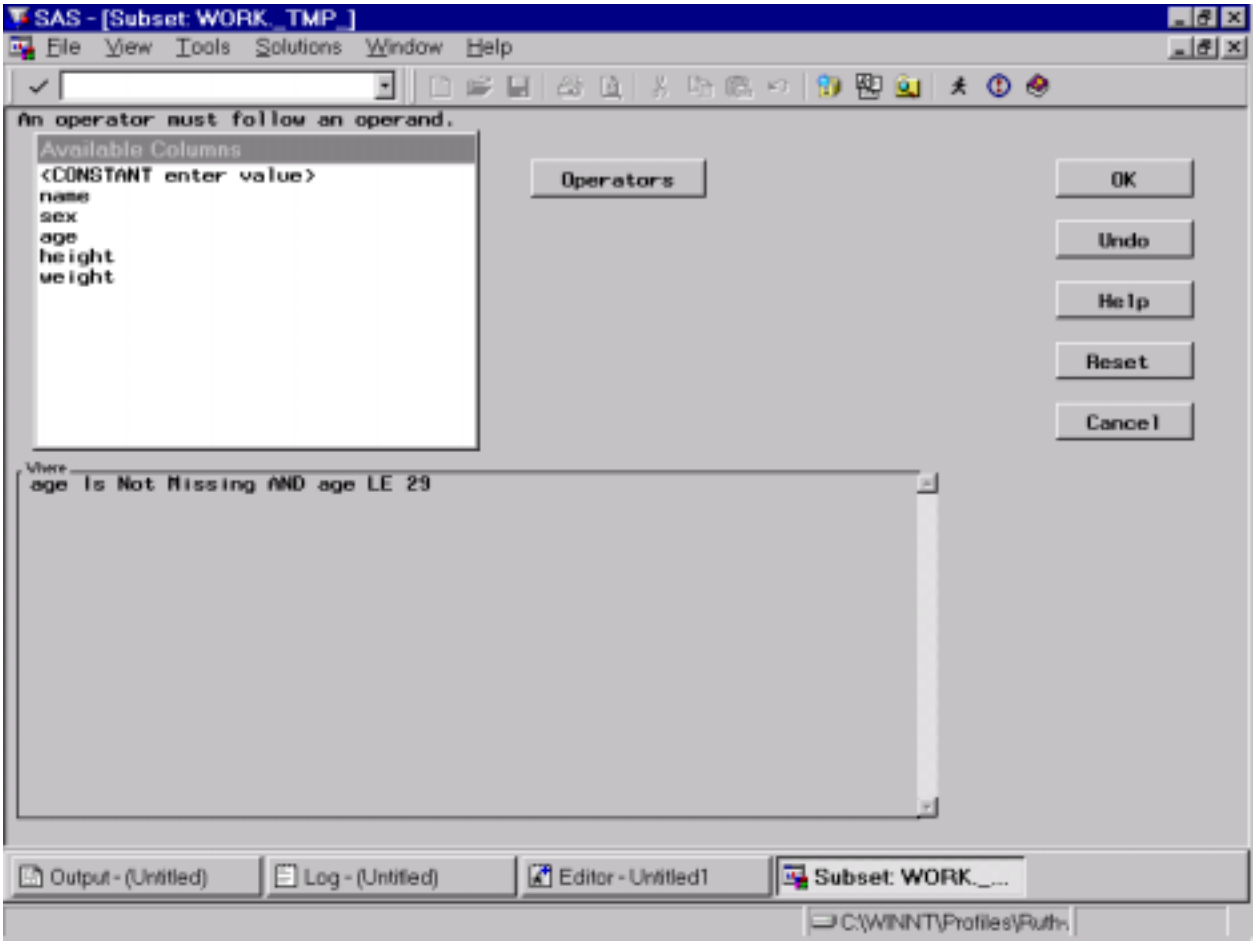


4. From this extended list, select **'IS NOT MISSING'**.

*** NOTE:** Use the **UNDO** button at the side of the window to delete/replace any incorrect selections you make.

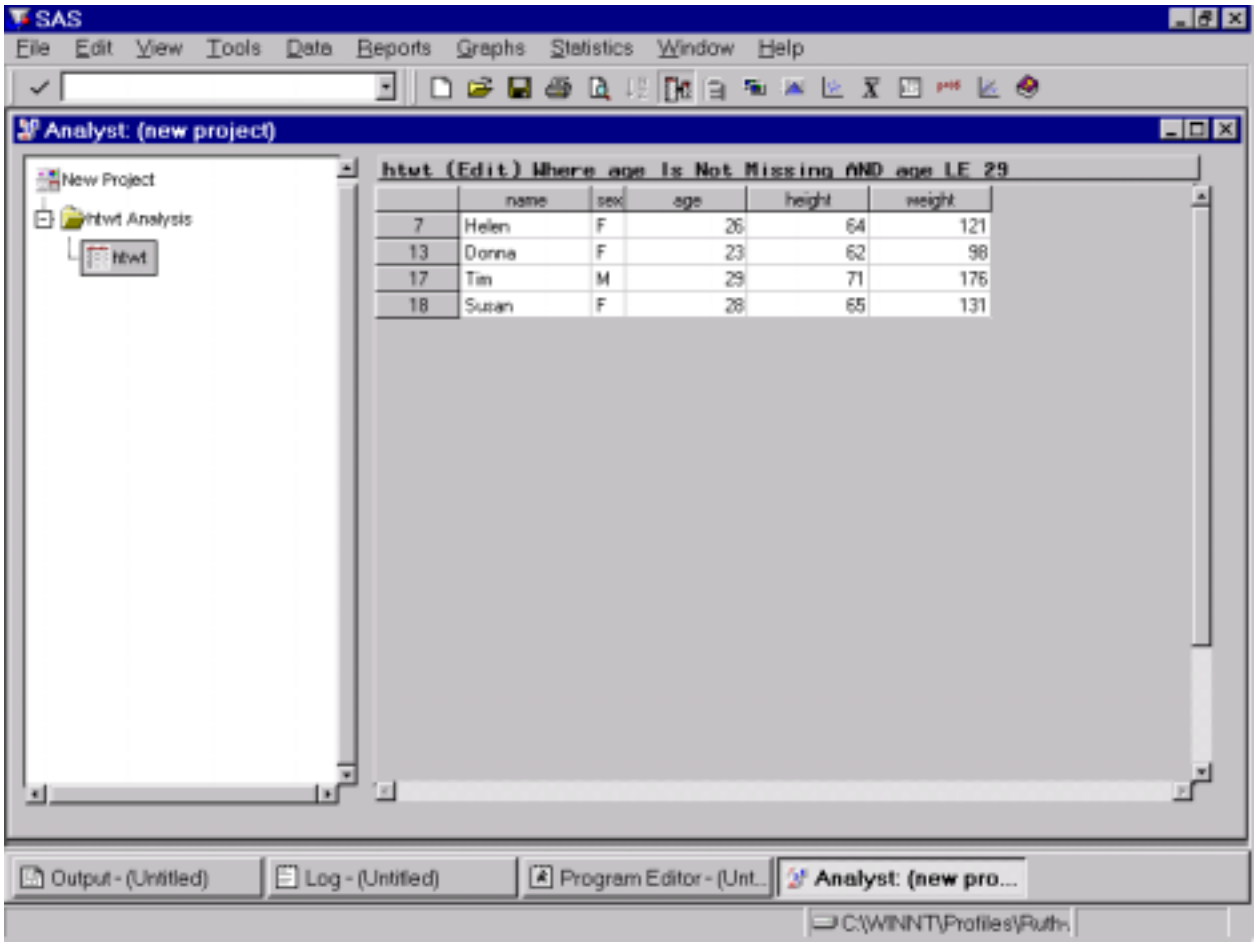
5. In order to specify that ***age*** must be less/equal to 29, you must make a compound statement. Click on the **OPERATORS** button and select **AND**.

6. Select ***AGE*** from the **AVAILABLE COLUMNS**. Select **LE** (*Less or Equal to)* from the list of **OPERATORS**.

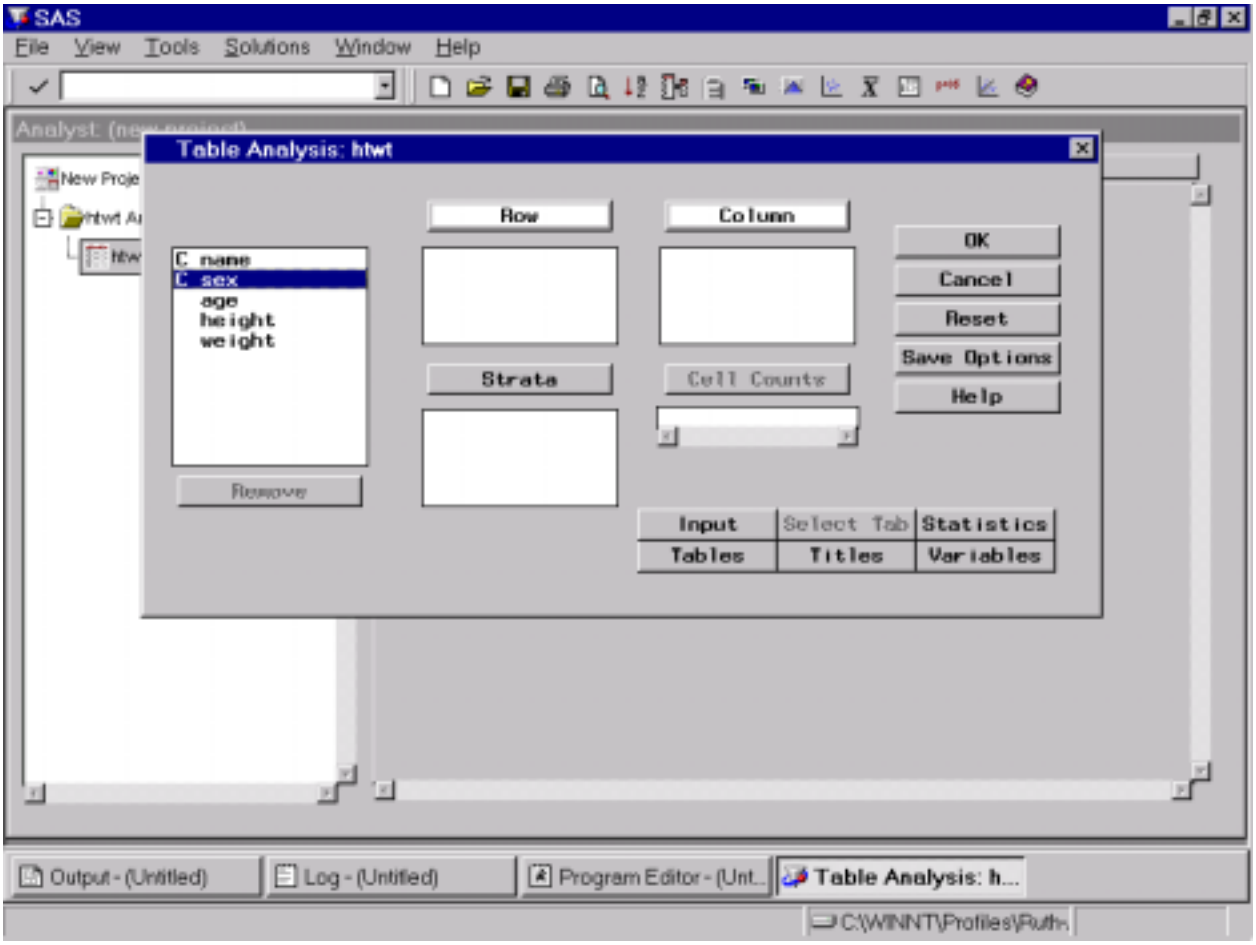


7. To specify the maximum ***age***, select '**<LOOKUP distinct values>**' from the **AVAILABLE COLUMNS.** This will open another window that lists all the values for ***age*** in the data set. Select 29. Click **OK**.

The resulting data listing should be as follows:

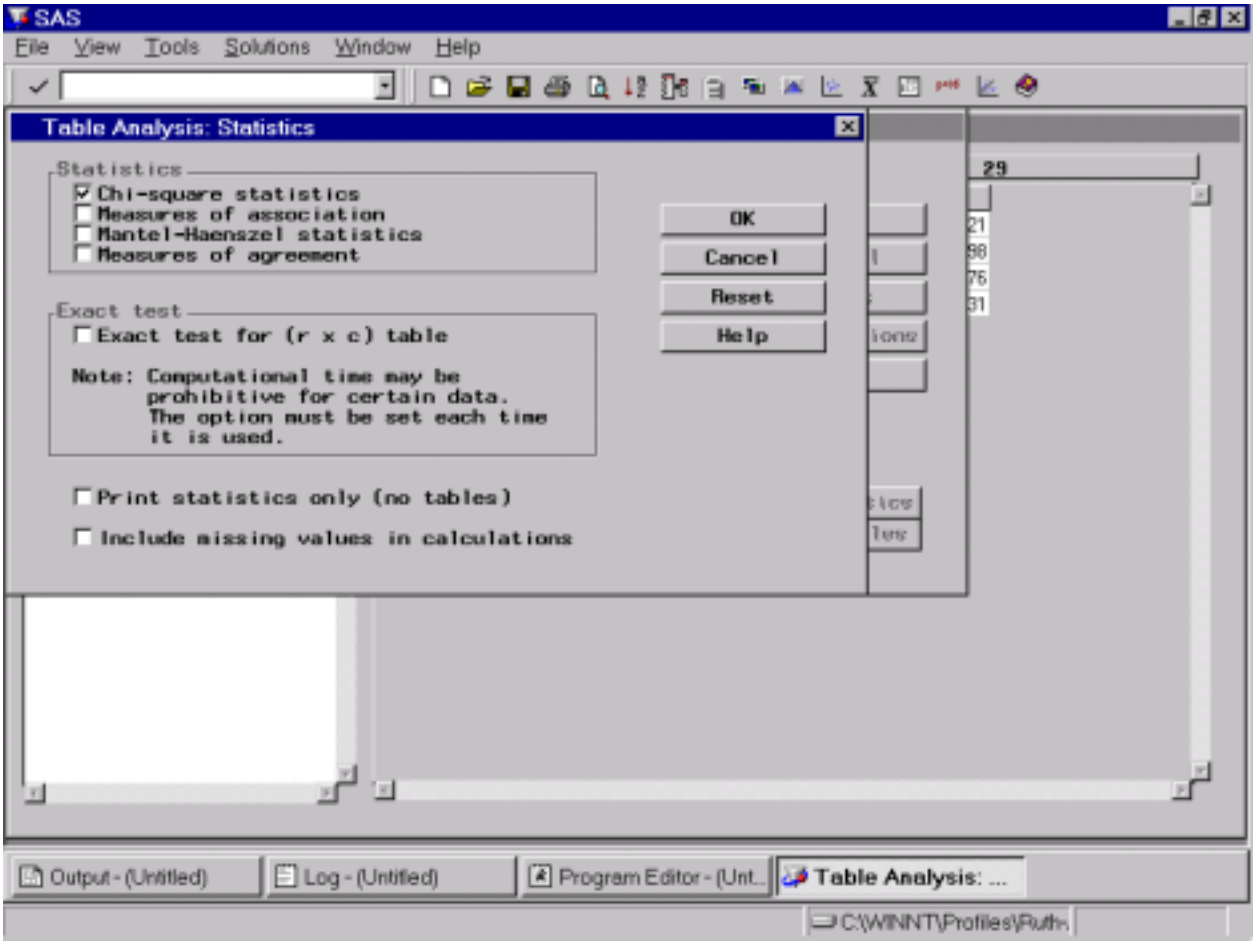


8. Now it's time to generate the two-way table. From the menu at the top of the screen, select **STATISTICS → TABLE ANALYSIS**. The window that pops up allows you to specify the layout of the table:

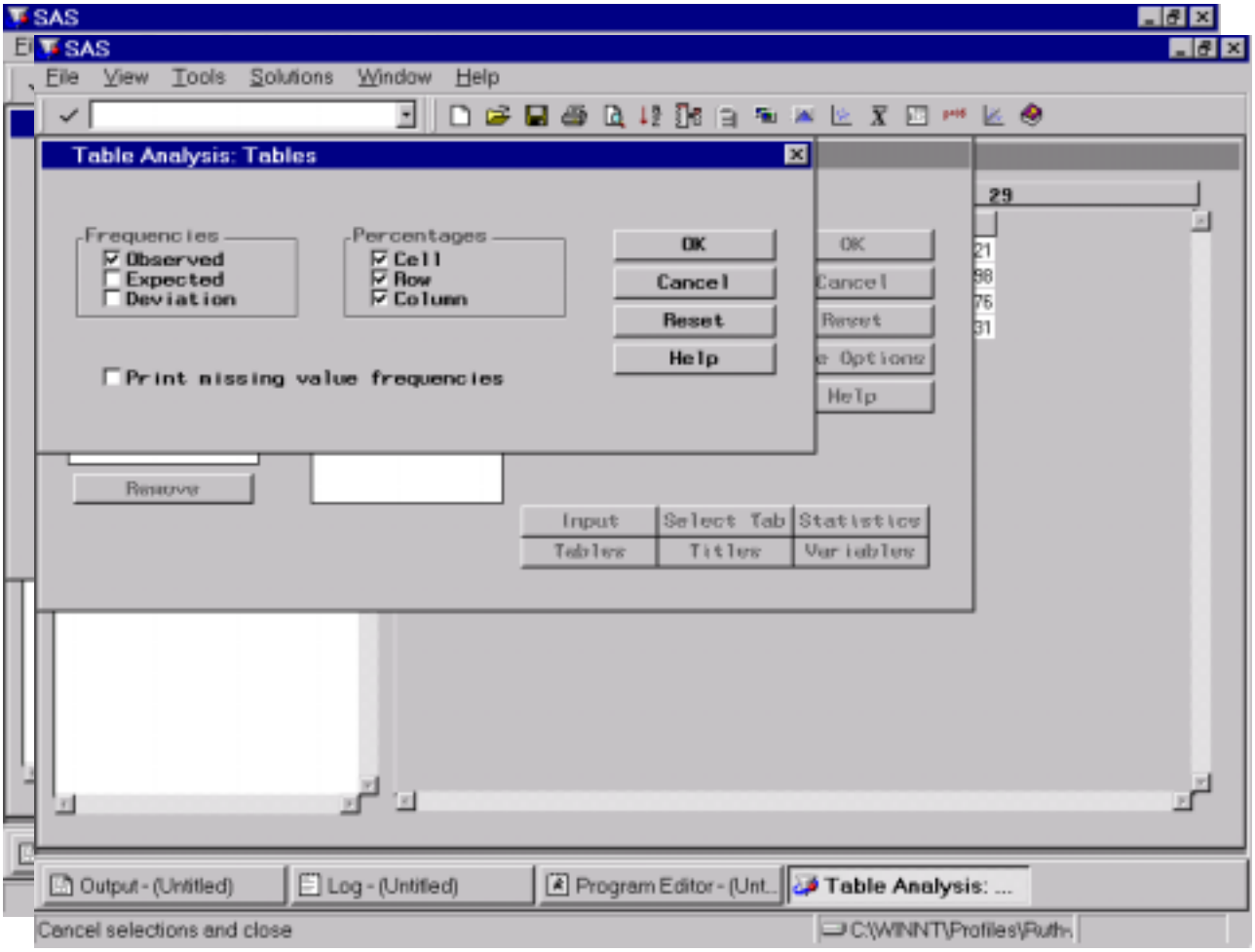


9. In this example, you want to create a table that plots ***sex*** in rows and and ***age*** in columns. Highlight ***sex*** and then click on **ROW**. Highlight ***age*** and click on **COLUMN**.

10. To choose the chi-square test, click on the **STATISTICS** button and check it off. Click **OK**.



11. To specify the frequency and percentage values you want in the table, click on the **TABLES** tab. Check off the desired values.



11. To add titles to the table, click on the **TITLES** button. Up to three titles can be added - one in each line: 1 Calculating Frequencies: Example 2
    2 Two-way Table: Sex X Age
    3 Subsetting ages 0 to 29
    Click **OK** when finished.

12. Click **OK** to run analysis.

Output should look like the following:

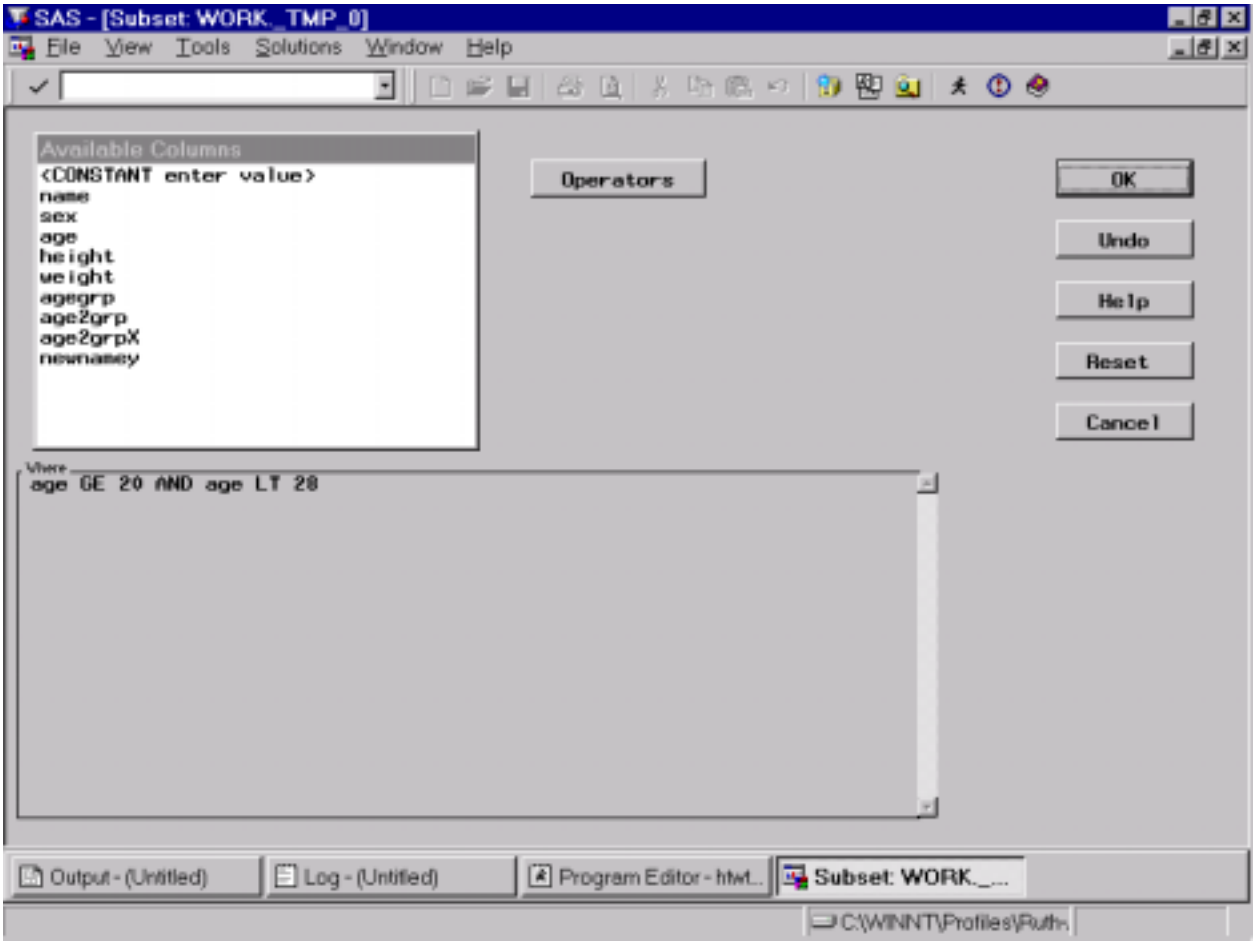
```
Calculating Frequencies: Example 2                              3
                                  Two-Way Table: Sex X Age     10:09 Thursday, August 10, 2000
                                  Subsetting ages 0 to 29
                          Where age Is Not Missing AND age LE 29

                                    The FREQ Procedure

                                   Table of sex by age

                    sex       age

                    Frequency,
                    Percent  ,
                    Row Pct  ,
                    Col Pct  ,      23,      26,      28,      29,  Total
                    ffffffffˆffffffffˆffffffffˆffffffffˆffffffffˆ
                    F        ,      1 ,      1 ,      1 ,      0 ,      3
                             ,  25.00 ,  25.00 ,  25.00 ,   0.00 ,  75.00
                             ,  33.33 ,  33.33 ,  33.33 ,   0.00 ,
                             , 100.00 , 100.00 , 100.00 ,   0.00 ,
                    ffffffffˆffffffffˆffffffffˆffffffffˆffffffffˆ
                    M        ,      0 ,      0 ,      0 ,      1 ,      1
                             ,   0.00 ,   0.00 ,   0.00 ,  25.00 ,  25.00
                             ,   0.00 ,   0.00 ,   0.00 , 100.00 ,
                             ,   0.00 ,   0.00 ,   0.00 , 100.00 ,
                    ffffffffˆffffffffˆffffffffˆffffffffˆffffffffˆ
                    Total           1        1        1        1        4
                                25.00    25.00    25.00    25.00   100.00


                              Statistics for Table of sex by age

                    Statistic                     DF       Value      Prob
                    fffffffffffffffffffffffffffffffffffffffffffffffffffff
                    Chi-Square                     3      4.0000    0.2615
                    Likelihood Ratio Chi-Square    3      4.4987    0.2124
                    Mantel-Haenszel Chi-Square     1      1.1905    0.2752
                    Phi Coefficient                       1.0000
                    Contingency Coefficient               0.7071
                    Cramer's V                            1.0000

                    WARNING: 100% of the cells have expected counts less
                             than 5. Chi-Square may not be a valid test.

                                    Sample Size = 4
```

Example 3: Three-Way Table on a Subset of data

This example creates a three-way table using three variables on a subset of the data. The first variable (***name***) represents the control variable for which separate output (cross-tabs of the other two variables - ***height***, ***weight***) is created for each of its values.

1. Filter the data and create a subset based on *age*, keeping only ages 0 to 27:
    - From the menu at the top of the screen, select **Data → Filter → Subset Data.**
    - In the dialogue window that pops up, select the ***age*** variable from the **AVAILABLE COLUMNS**.
    - Following the instructions in Example 3, 'write' a statement identifying the conditions to determine which records will be included. From the list of possible **OPERATORS** that also appears beside the **WHERE** box, select **GE** (which stands for 'greater than or equal to).
    - To specify the minimum ***age*** to be included, select **<CONSTANT enter value>** from the **AVAILABLE COLUMNS**. Type in 20 in the window that pops up and click **OK**.
    - To finish the statement of conditions, select ***age*** from **AVAILABLE COLUMNS**. Select **AND**, and then **LT** (less than) from the list of **OPERATORS**.
    - To specify the maximum ***age***, select **<CONSTANT enter value>** and type in 28. Click **OK**.
    - Click **OK** when finished

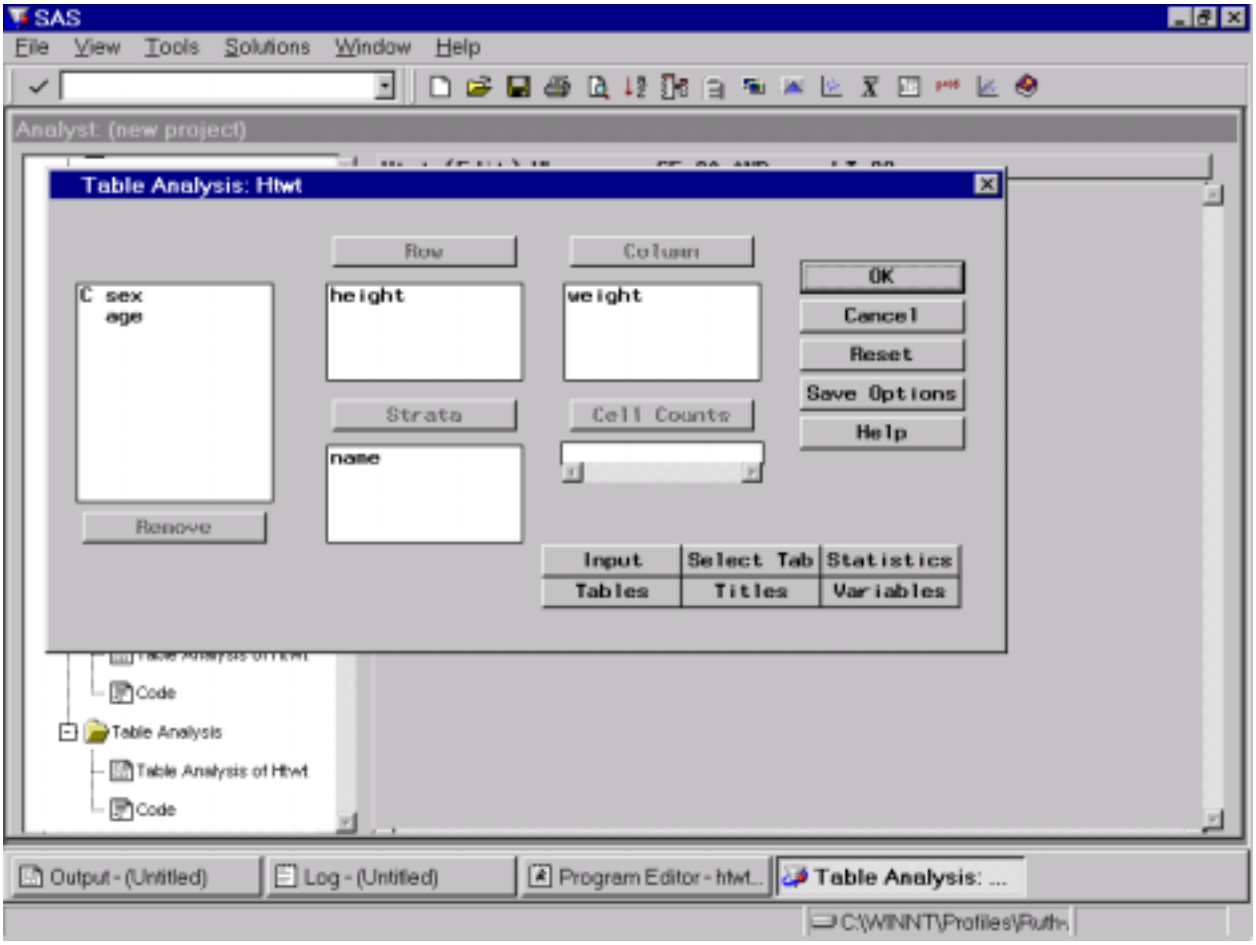
    - The data list will now show only records for Donna and Helen since they are the only ones whose ages are between 0 and 27.



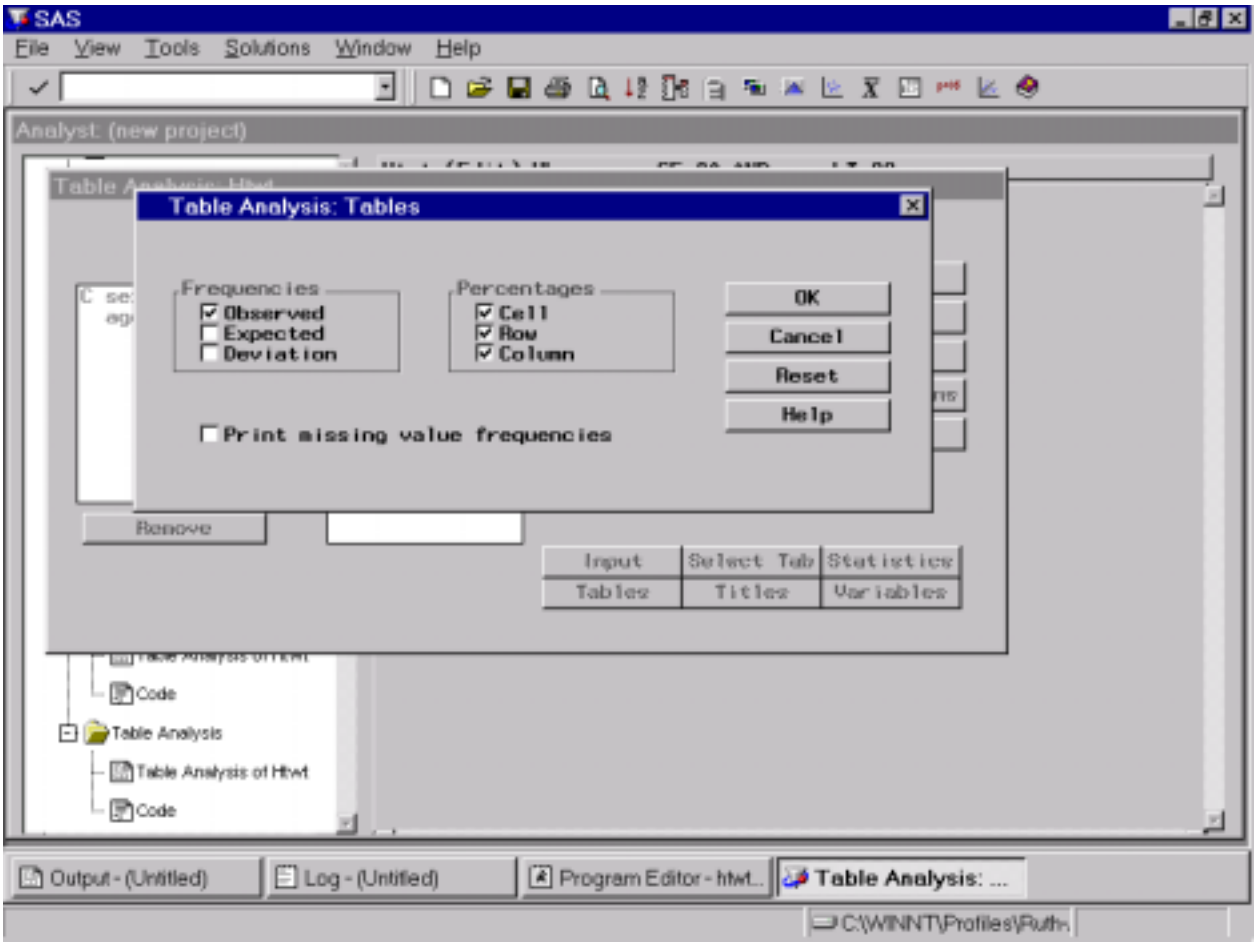
2. Create tables:

- From the menu at the top of the page, select **STATISTICS → TABLE ANALYSIS**

- In the dialogue window that pops up, you can specify the row, column and strata variables. Select the desired variable and click on the appropriate varible type: ***height*** is the **ROW** variable, ***weight*** is the **COLUMN** variable, and ***name*** is the **STRATA** variable (also determines the name of the table).



- Given the selected variables, two tables are possible. To select the preferred table, click on the **SELECT TAB** button and then highlight the three way table option: **NAME*HEIGHT*WEIGHT**. Click **OK.**
- To specify the information to be included in the table, click on the **TABLES** button. For this example, under **FREQUENCIES**, check off **observed**, and under **PERCENTAGES**, check off **cell**, **row**, and **column**. Click **OK.**


SAS
File View Tools Solutions Window Help
Analyst: (new project)
Table Analysis: Tables
Frequencies
Observed
Expected
Deviation
Percentages
Cell
Row
Column
OK
Cancel
Reset
Help
Print missing value frequencies
Remove
Input
Select Tab
Statistics
Tables
Titles
Variables
Code
Table Analysis
Table Analysis of Htwt
Code
Output - (Untitled)
Log - (Untitled)
Program Editor - htwt...
Table Analysis: ...
C:\WINNT\Profiles\Ruth

- The output should look like:

```
Where age GE 20 AND age LT 28

                        The FREQ Procedure

                    Table 1 of height by weight
                     Controlling for name=Donna

                height     weight

                Frequency,

Percent   ,
                Row Pct   ,
                Col Pct   ,      98,     121,  Total
                ffffffffˆffffffffˆffffffffˆ
                      62 ,       1 ,       0 ,       1
                         , 100.00 ,    0.00 , 100.00
                         , 100.00 ,    0.00 ,
                         , 100.00 ,       . ,
                ffffffffˆffffffffˆffffffffˆ
                      64 ,       0 ,       0 ,       0
                         ,    0.00 ,    0.00 ,    0.00
                         ,       . ,       . ,
                         ,    0.00 ,       . ,
                ffffffffˆffffffffˆffffffffˆ
                Total            1         0         1
                             100.00     0.00   100.00


                    Table 2 of height by weight
                     Controlling for name=Helen

                height     weight

                Frequency,
                Percent   ,
                Row Pct   ,
                Col Pct   ,      98,     121,  Total
                ffffffffˆffffffffˆffffffffˆ
                      62 ,       0 ,       0 ,       0
                         ,    0.00 ,    0.00 ,    0.00
                         ,       . ,       . ,
                         ,       . ,    0.00 ,
                ffffffffˆffffffffˆffffffffˆ
                      64 ,       0 ,       1 ,       1
                         ,    0.00 , 100.00 , 100.00
                         ,    0.00 , 100.00 ,
                         ,       . , 100.00 ,
                ffffffffˆffffffffˆffffffffˆ
                Total            0         1         1
                                       0.00   100.00   100.00
```

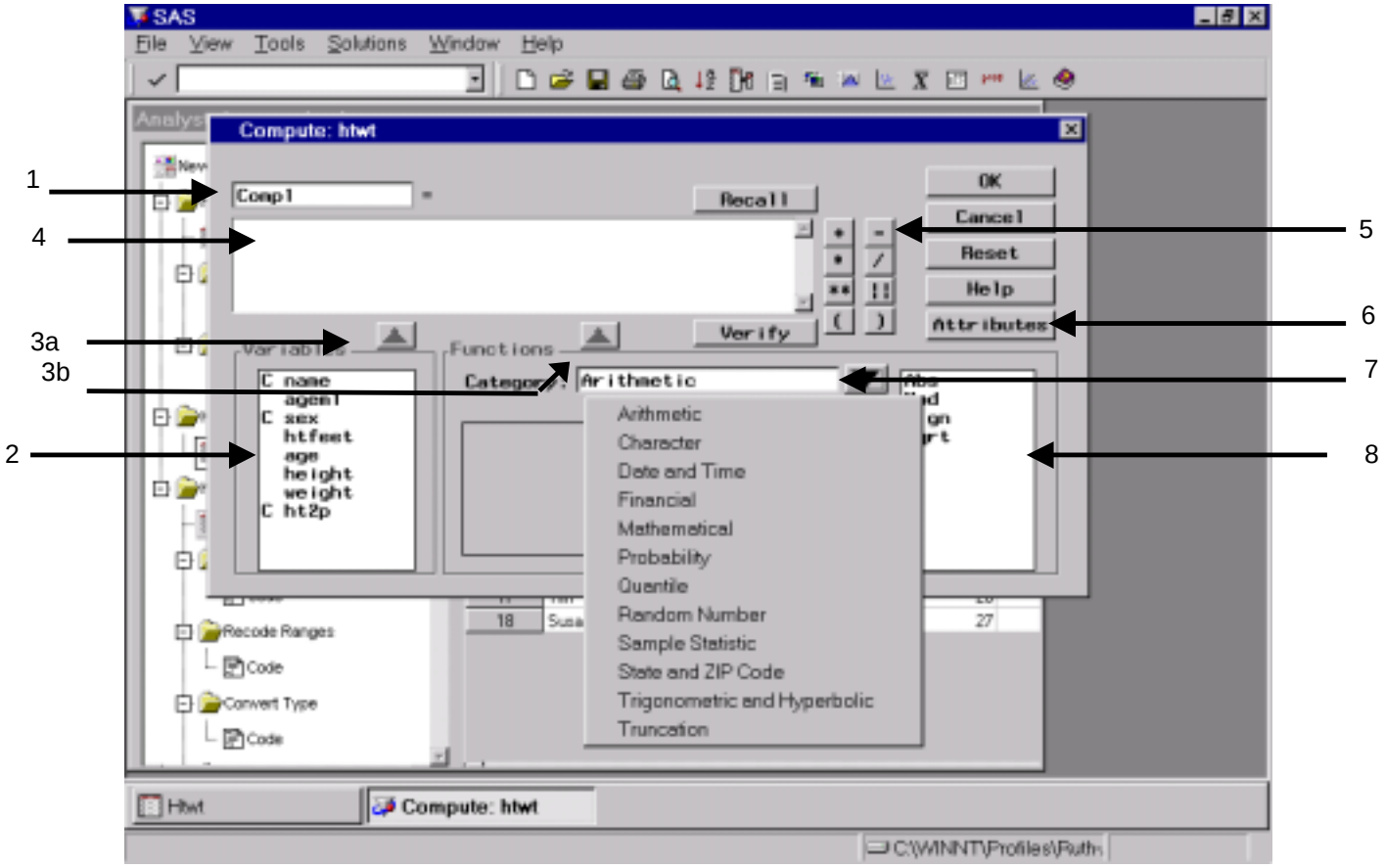
IV. Data Manipulation: Creating New Variables

A. Create new variables that incorporate computations made on existing variables.

If continuing from the previous example, take off the filter: from the menu at the top of the screen, select **DATA → FILTER → NONE.**

Ensure that the edit mode is set to EDIT: From the menu at the top of the screen, select **EDIT → MODE → EDIT.**

Each of the following computations must be made separately in the **Compute window**. Go to the menu at the top of the screen and select **DATA → Transform → Compute**.



1. New Variable Name
2. Variable List
3. a) Variable Select Button
   b) Function Select Button
4. Statement Box
5. Mathematical Operators
6. Attributes
7. Functions Categories
8. Function Sub-categories

1. **Add *'2'* to *height* for each record**.

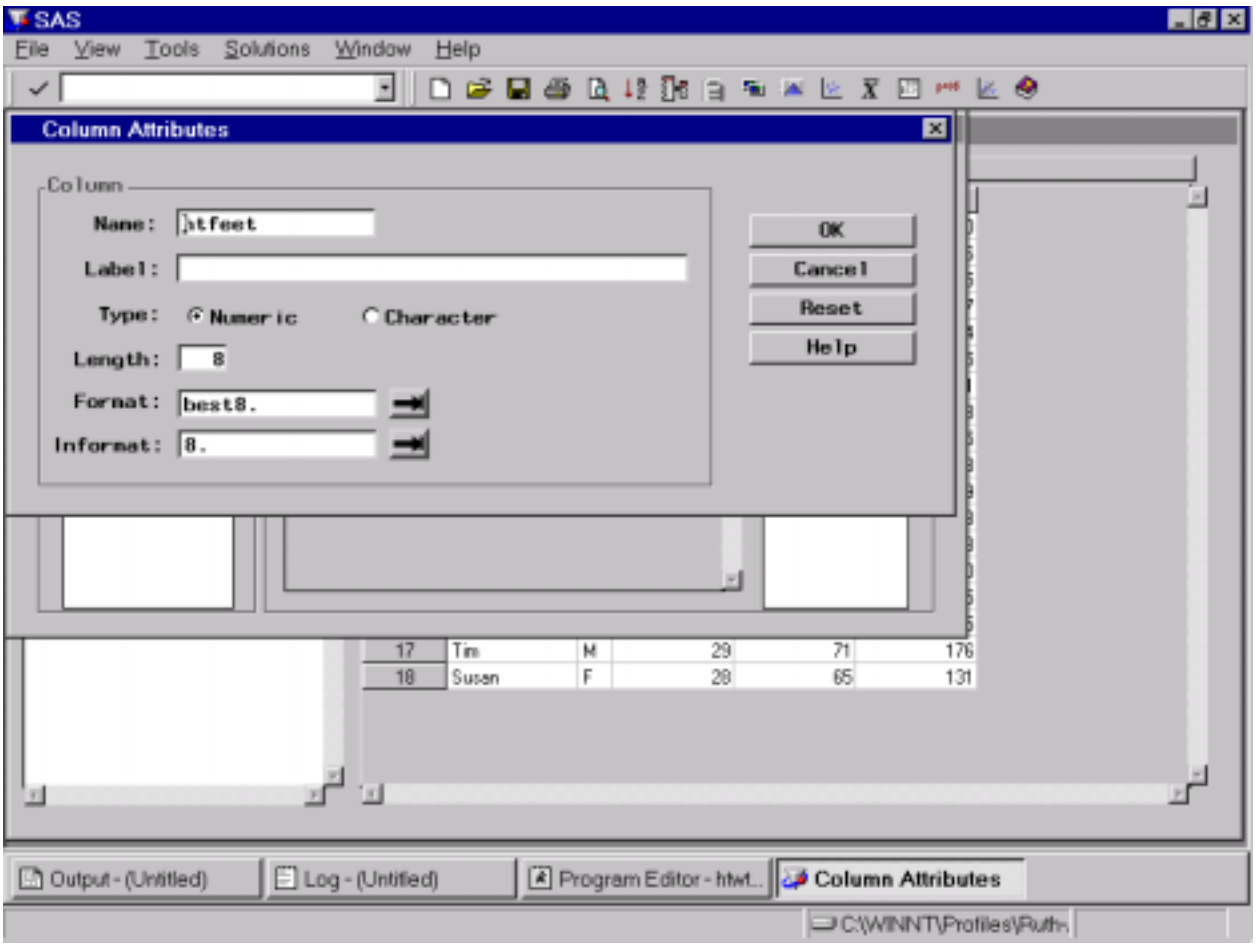
   - Select ***'height'*** from the **Variable List** (#2) and click on the **VARIABLE SELECT** button (#3a). This variable now appears in the **Statement Box** (#4).
   - Click on the **PLUS SIGN** from the selection of **Mathematical Operators** (#5).
   - Type '2' in the **Statement Box**
   - To verify that the expression is valid, click on the **VERIFY** button just below the **Statement Box**.
   - Type in the **New Variable Name** (#1): **ht2p**. Click **OK** when finished.

2. **Subtract '1' from *age* for each record.**

   - Select ***age*** from the **Variable List** and click on the **VARIABLE SELECT** button.
   - Click on the **MINUS** sign from the selection of **Math Operators**.
   - By default the column name is *Comp1*. Replace it with **agem1**
   - Click **OK** when finished.

3. **Convert *inches* to *feet* for *height*** (i.e. divide value for height by 12) **and limit to one decimal place for each record.**

   - Select 'height' from the **Variable List** and click on the **VARIABLE SELECT** button
   - Click on the **DIVIDE** sign from the selection of math operators. Type in 12 on the same line.
   - Type in the new column name: **htfeet**.
   - To limit the number of decimal places, click on the **ATTRIBUTES** button.

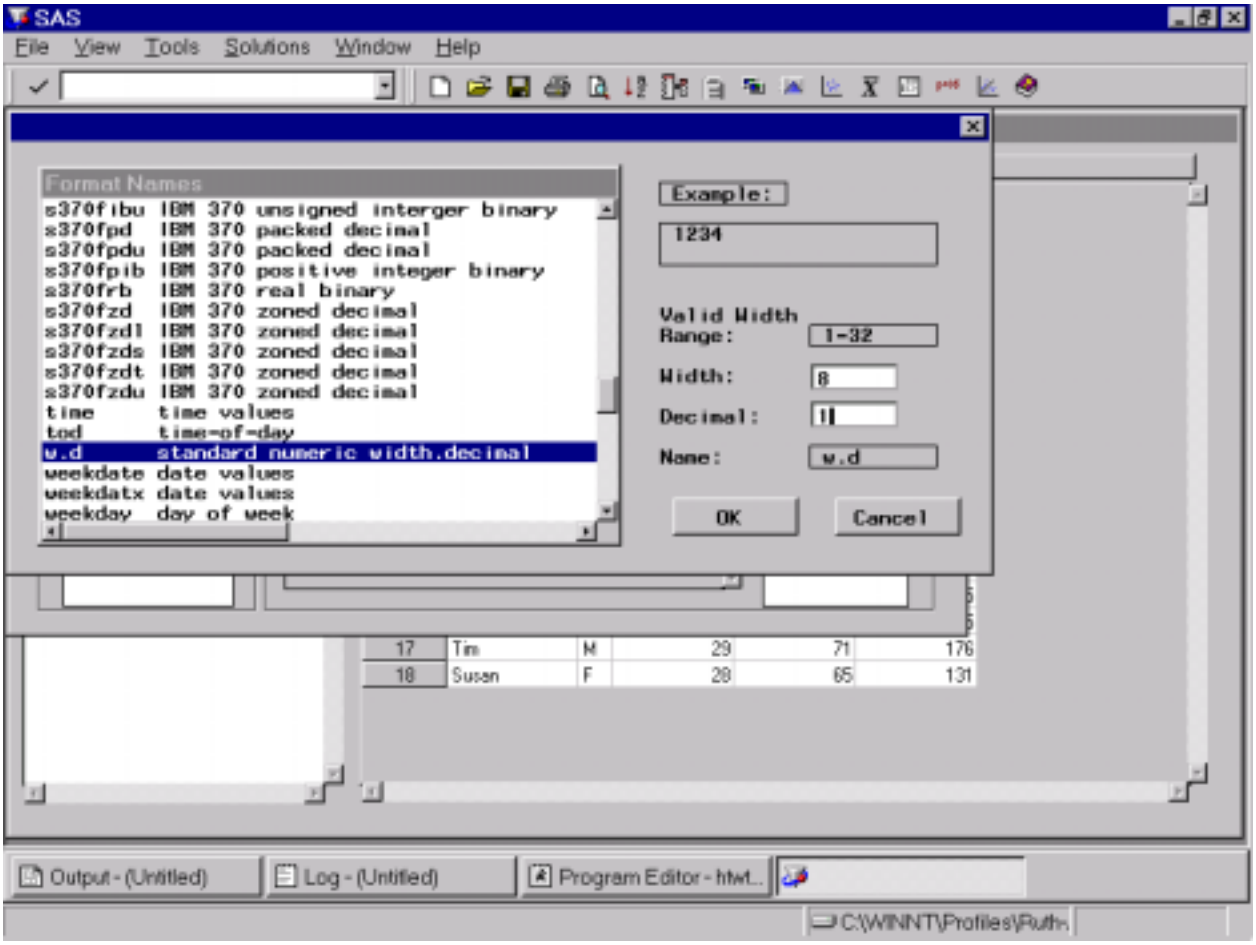


The **COLUMN ATTRIBUTES WINDOW** enables you to specify the attributes for your computed column.

- The name of the column should be **htfeet**.
- Specify whether the column type is **Numeric** or **Character**. In this example, it is **numeric.**
- The variable length in the **Length** field can remain at the default of **8**.
- The arrows next to **Format** and **Informat** allow you to select the format and informat for the column. They should be consistent with each other and with the variable length. A format controls how values are displayed, while an informat controls how values are read by the SAS System.

  Scroll down the list of **FORMATNAMES** until you come to: **w.d. standard numeric width, decimal,** and select it. This option allows you to set the number of decimal places to **1**. Click **OK** when finished.

Note: the **WIDTH** field specifies the maximum number of places before the decimal places. For this example, leave the default of 8.



- Follow the same procedure to change **INFORMAT** to **1** decimal place.

4. **Convert pounds to ounces for *weight* for each record.**

- Select ***weight*** from the **VARIABLE LIST** and click on the **VARIABLE SELECT** button.
- Click on the **multiplication** sign from the selection of math operators. Type in **16** on the same line.
- Type in the new column name: **wtounce**

- Click on **OK**.

B. Create new variables that truncate the values of existing variables.

Ensure that the edit mode has been changed from BROWSE to EDIT: From the menu at the top of the screen, select **EDIT → MODE → EDIT.**

Each of the following computations must be made separately in the **Compute window**. Go to the menu at the top of the screen and select **DATA → Transform → Compute**.

Refer to the figure in the previous example to help identify the various fields and boxes mentioned below.

1. **Truncate *name* to display only the first two values.**

   - Change column name to **name1**.
   - Because ***name*** is a character variable, change the **Functions Categories** (Box #7) to **Character.**
   - Click on the **ATTRIBUTES** button and change the **VARIABLE TYPE** to **character.** Click **OK**.
   - Scroll down the **FUNCTION SUB-CATEGORIES** and select **SUBSTR.**
   - The following expression appears in the **STATEMENT BOX**: substr(string,n<,n>)
   - Delete ***string*** and replace with ***name*** by highlighting ***name*** in the **VARIABLE LIST** and clicking on the **FUNCTION SELECT BUTTON.**
   - The first ***n*** represents the starting point for the truncation. Replace it with **1** and delete the '**<**' beside it.
   - The second ***n*** represents the number of values you want to be displayed. Replace it with **2** and delete the '**>**' beside it.
   - Click **OK**.

2. **Truncate *name* to display the first 3 values**.

   Follow the instructions from Example 1 except:
   - Change column name to **name2**
   - Replace the ***n's*** in the substring expression with **1** and **3**, respectively.

3. **Truncate *name* to display the first 3 values, starting at the second value**.

   Follow the instructions from Example 1 except:
   - Change the column name to **name3**
   - Replace the *n's* in the substring expression with 2 and 3, respectively.

C. Creating and grouping formats to create new variables

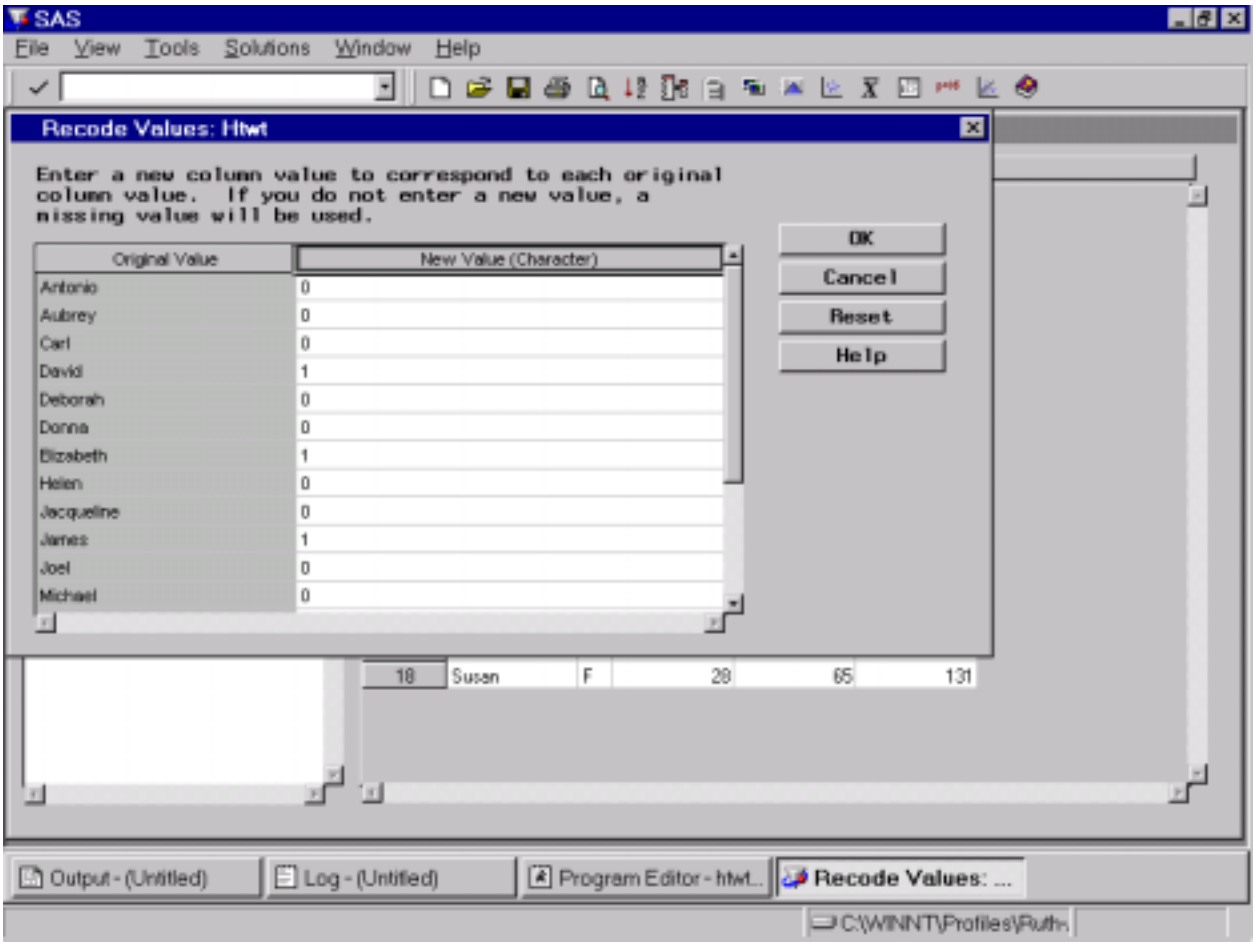
Character Variables

Character variables can be grouped in one of two ways - by recoding the values or by integrating SAS code with the point-and-click approach of Analyst.

This example will show you how to recode the *name* variable both ways. 'Elizabeth', 'David', and 'James' will be grouped as one new variable.

1. **Recode Values**

   - From the menu at the top of the screen, select **DATA → TRANSFORM → RECODE VALUES**
   - Click in the line for **COLUMN TO RECODE** for a list of variables. Select ***'name'***
   - Select **CHARACTER** for **NEW COLUMN TYPE**
   - Click **OK**
   - In the dialogue window that pops up, enter **'1'** as the new value for 'Elizabeth', 'David', and 'James' and **'o'** for the other names. Hit the **TAB** key twice to move to the next line.



   - A new column entitled *'name_recoded'* now appears in the data list.

NOTE: This method is simple and straightforward for a small data set or for binary (dichotomous) variables, but not very efficient for large data sets or for non-dichotomous variables with several value options.

2. PUT statement

SAS uses formats to read, display, and print data stored in a SAS data set.

The PUT function associates a format with an existing variable to create a new variable. It can be used with either numeric or character variables. The PUT statement can also be used to convert numeric variables into character variables.

- Create a format for the *name* variable: In the PROGRAM EDITOR of (base) SAS, type in the following code: proc format;
  Value $namef 'Elizabeth', 'David', 'James' = '1'
  Other = '0';
  Run;
  The '$' denotes character values.

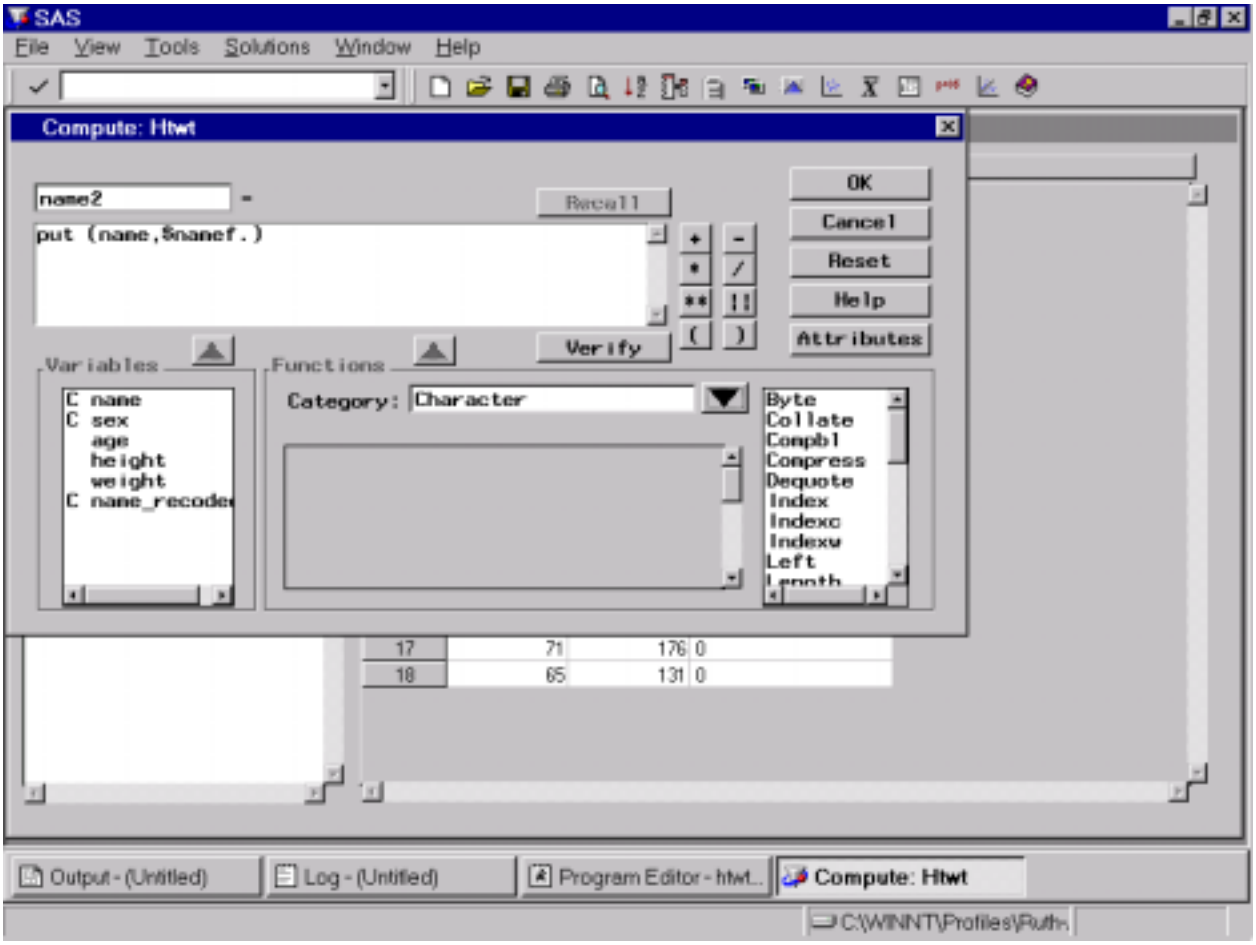
- Submit the program by clicking on the **SUBMIT** icon at the top of the screen 

  This creates a 'look up' table that converts or labels variables, and can be referred to for further analyses in the same SAS session.

- In ANALYST, from the menu at the top of the screen select **DATA → TRANSFORM → COMPUTE**.

- In the dialogue window that pops up, any valid SAS code can be entered into the statement box. Type in the following statement: put (name,$namef.)

  This statement instructs SAS to create a new variable using the format created above.



- To ensure that Analyst knows this is a character variable, click on the **ATTRIBUTES** button and check off this option in the dialogue window that pops up. Click **OK**.

- Change the name of the new variable by replacing COMP1 with **name2.** Click **OK**.

This generates a new column with values of either 0 or 1 instead of actual names.

Numeric Variables

Note: Numeric variables can also be converted to character variables to save storage space or if there are leading zeros you wish to keep.