Chapter Four

Utilities

Overview

This chapter reviews the procedures required to name and label variables, label the discrete values of variables, define missing values, attach text documents to data files, compute and recode variables, select cases for analysis, and sort data.

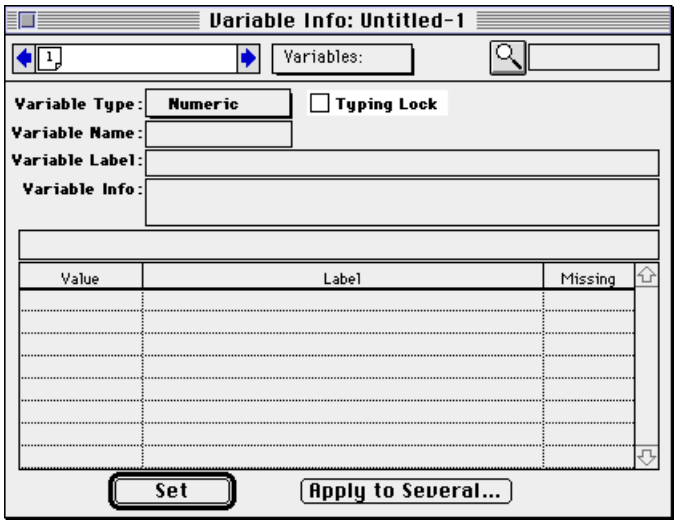
Defining the data

Information entered in the FLO•STAT data matrix can be both numerical and character (e.g., a, b, c....) . The default variable type is numeric. Any tab-delimited text can be pasted or imported into the data matrix. However, since FLO•STAT processes numeric and character data differently when calculating statistical tests, the type of data stored in each column must be correctly specified.

Select **Var Name, Labels...** from the **Utilities** menu or double click in any of the cells in the data matrix to open the Variable Info window.

Variable type

The **Variable Info** window contains the **Variable Type** menu near the top left of the window. Variable type **Numeric** is set by default. To change the variable type to character, select **Character** from this menu.



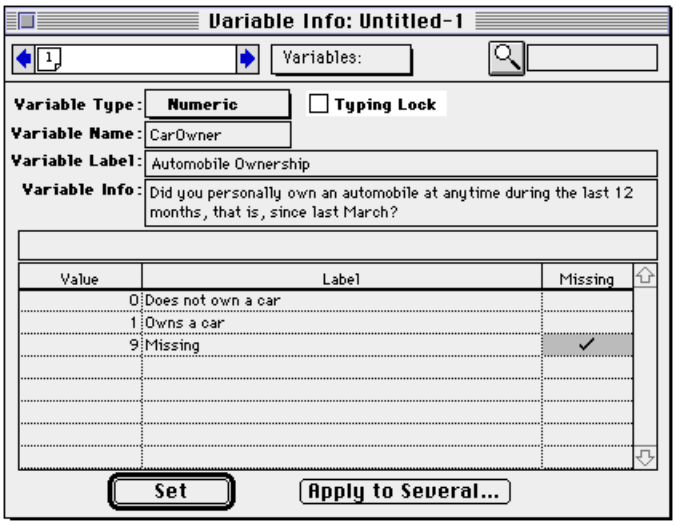
Moving about the variable info window

All of the information shown in the **Variable Info** window applies to the matrix column (i.e., variable) shown as the number shown the scroll bar in the top left of the window. The **Variable Info** window shown above is displaying the information for column 1. To select another column, click on the scroll bar, or the menu marked **Variables**, to the immediate right of the scroll bar, until the desired column is located.

To locate a particular text string anywhere in the Variable Info window, enter the string in the edit field at the top right and click the spy glass icon. For example, to locate an income variable, you might type "salary" and then click the spy glass icon. If the string, "Salary," exists anywhere among the information defining the variables information about that variable will be displayed.

Naming variables

A new variable name is entered in the **Variable Name** edit field. FLO•STAT follows the conventional rules regarding variable names and accepts a maximum of eight characters in the variable name. Furthermore, all variable names must begin with letters, may contain numbers and periods, and may not include special characters (i.e., ^, *, !, ?, +) . Further, variable names cannot be duplicated. A warning dialog will appear if these conventions are not followed.



Variable label, variable info, value labels, and missing values

The variable label for the variable named CarOwner, is entered in the edit field immediately below the variable name edit field. Variable labels can be up to 256 characters in length.

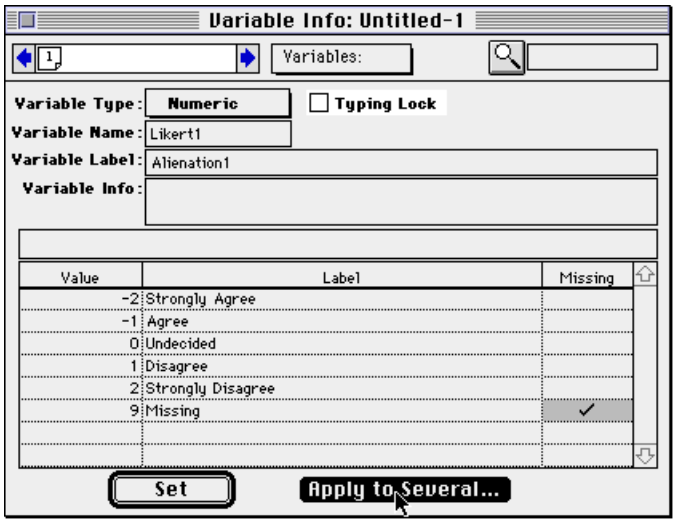
Additional information about each variable can be entered and stored in the **Variable Info** field. This field is designed to hold information about the wording of a survey question, calculations used in creating the variable, or general comments and notes about the variable. The **Variable Info** field can store several thousand characters of information on each variable.

Value labels are entered in the matrix at the bottom of the window. Labels can be 80 characters in length.

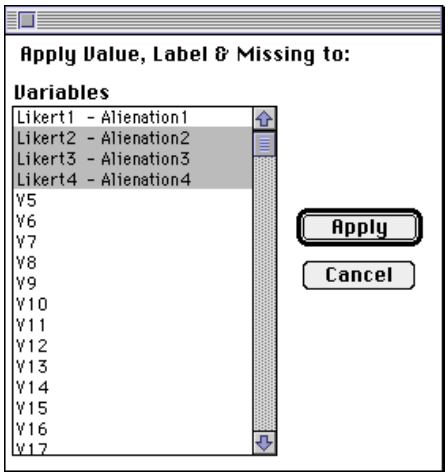
A variable's missing value(s) is set by clicking the cell under the missing column at the far right. Missing values can be turned off by clicking the same cell a second time.

Applying the same labels to more than one variable

To apply an existing set of labels to more than one variable, click the **Apply to Several...** button at the bottom of the Variable Info window.



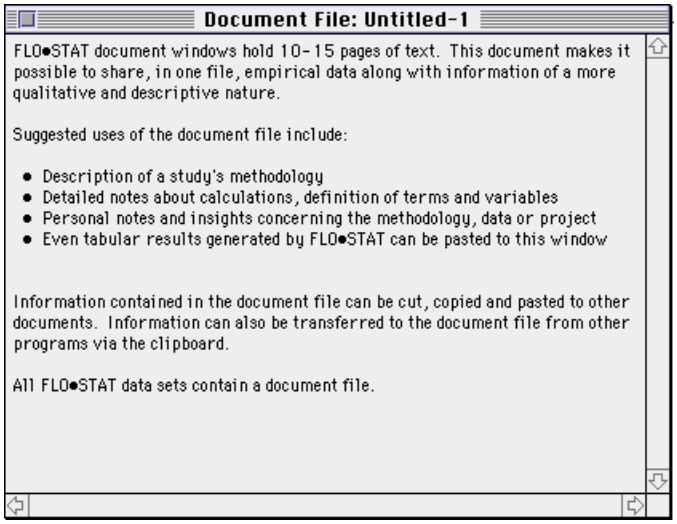
As illustrated below, select the appropriate variables from the list and click Apply.



The document file

Each FLO•STAT data set has dynamic space reserved for an accompanying text file. This file is opened by clicking on the first icon at the left of the icon bar. The file provides an area for technical documentation and project notes.

Standard Macintosh editing features involving **Cut**, **Copy** and **Paste** apply in the Document window.



Data transformation

To create new variables from either the transformation of existing variables or the mathematical combination of several variables, click the transformation icon or select **Transformation** from the **Utilities** menu.

The Transformation window contains four drop down menus, a graphic numeric keypad, an equation editor and an error message field.

The creation of new information and the transformation of existing data is completed by entering compute(s), recode(s) or if-then statements in the editor located in the center of the window.

Each statement can be typed directly or built in part from selecting items from one or all of the four drop down menus at the top of the window.

Statements entered in the transformation window remain with their respective data file, permitting you to accumulate and verify your work.

On the following pages are several examples of computing new variables.

Computing new variables

Example #1: Using four exam scores, named **Test1**, **Test2**, **Test3**, and **Test4**, compute an average test score named FGrade. The average score values should contain two digits to the right of the decimal place. (Note: if a selected variable on the right hand side of the equation contains a missing value, the computed value will be missing. Missing values appear as "•" in the data set.)

FGrade = (Test1 + Test2 + Test3 + Test4) / 4

or

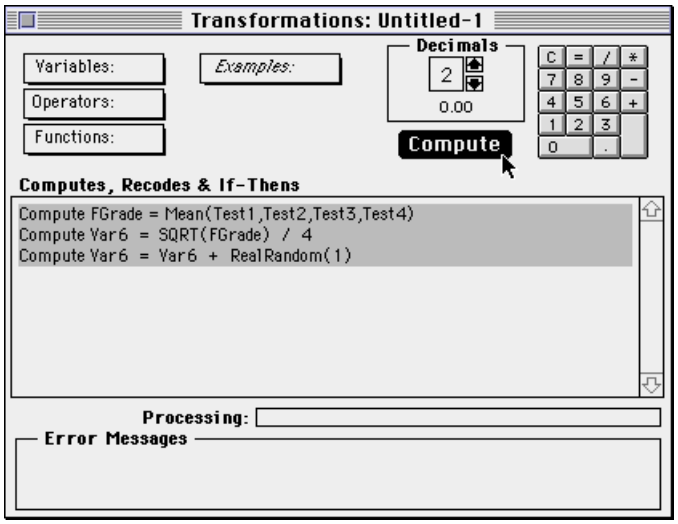
FGrade = Mean(Test1,Test2,Test3,Test4)

Example #2: Calculate the square root of variable 5 and divide by 4. Var6 should contain two digits to the right of the decimal place.

Var6 = SQRT(FGrade) / 4

Example #3: Replace Var6 with the sum of its own value added to a randomly generated real number varying between 0 and 1. Var6 should contain two digits to the right of the decimal place.

Var6 = Var6 + Random(0,1)



Enter the equation, select the statement(s), set the number of decimal places for the output values, and click the Compute button. New variables will be added as the last column in the data matrix.

Arithmetic Functions

There are a number of built-in functions for use in computing new variables. The functions and their use are outlined below.

Abs(n) Absolute value (|-var| is var)
CaseNum(n) Count each case in increments of (n)
log10(x)Base 10 logarithm (10 raised to the power of x)
ln() Natural or Naperian logarithm (base e or natural logarithm)
RealRandom(n) Uniform random real number between 0 and n
IntRandom(n) Uniform random integer number between 0 and n

Normal(m,SD)	Normal random number (distributed with a mean of m and a standard deviation of SD)
Round(x)	Round to an integer (e.g., Round(12.9) is 13)
Sqr(n)	$variable_x$ * $variable_x$ (e.g., 4 * 4 is 16)
Sqrt(n)	Square root (e.g., √9 is 3)
TRUNC(n)	Truncate to an integer (e.g., Trunc(12.9) is 12)
Sin(x)	Sine (argument in radians)
Cos(x)	Cosine (argument in radians)
Tan(x)	Tangent (argument in radians)
Arctan(x)	Arctangent

Statistical Functions

Sum(arg list)	Sum of values (Var1,Var2,Var3,Var7)
Mean(arg list)	Mean of values (Var1,Var2,Var3,Var7)
Sd(arg list)	Standard deviation of variables in the arg list
Variance(arg list)	Variance of the variables in the arg list
CFVar(arg list)	Coefficient of variation of the variables in list
Min(arg list)	Minimum value across the arg list
Max(arg list)	Maximum value across the arg list

Recoding variables

Often data are not in the form needed for a particular analysis-- you have character data which needs to be converted into numeric values , or continuous values collapsed in to a few broad categories (e.g., 18-24, 25-44, 45+).

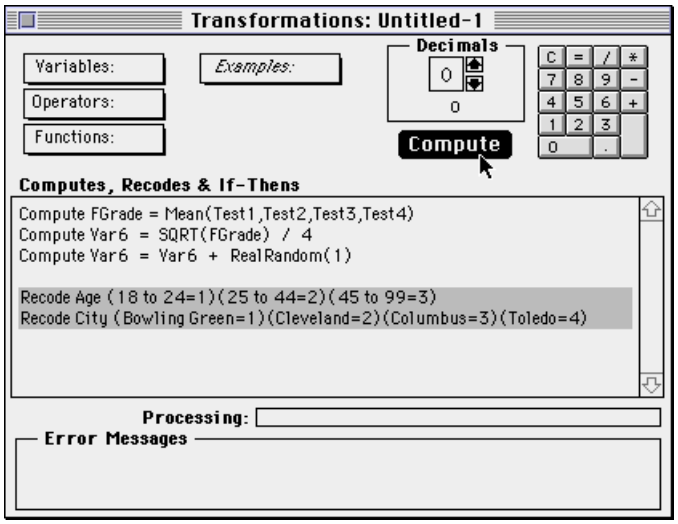
It is easy to recode existing data values using FLO•STAT.

Open the **Transformation** window by clicking on the transformation icon or selecting **Transformations...** from the **Utilities** menu.

Below are several examples using the recode utility.

Example #1: Recoding age from single year values to broader age categories.

Example #2: Recoding city name to a numeric value (i.e., a text value to a numeric value). Whenever a character variable is being changed to a numeric value or a numeric to character, the variable must be first defined as variable type character.



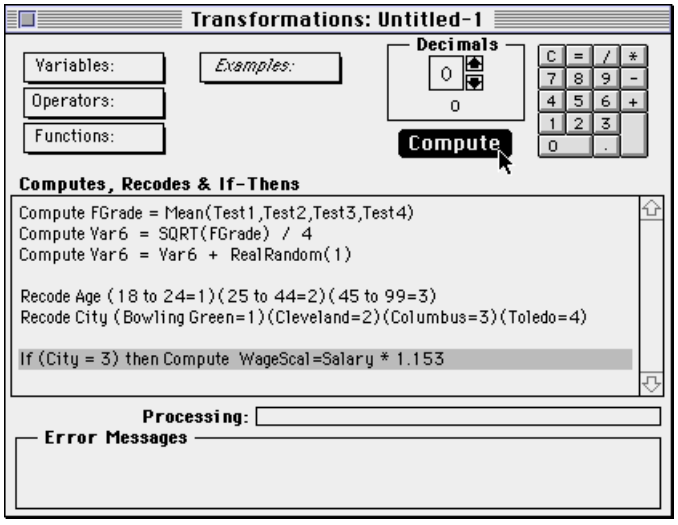
Additional examples are available in the pull down menu marked *Examples* at the top of the window.

Conditional transformations

Both Computes and Recodes can be conditional on the values of other variables.

In the examples above, both the compute and recode statements could have been conditional on whether or not the value of a particular variable was less than or equal to 50, for example.

To illustrate, assume a new variable representing wage scale is to be adjusted for the city in which each person lives. In the example below, in each case where the value for city = 3 (i.e., Columbus) wage scale will multiplied by a factor of 1.153 to more accurately reflect a higher pay differential in a state capital. In all other cities wage scale will remain unaltered.



Additional examples of conditional transformations are available in the pull down menu marked *Examples* at the top of the window.

Sorting cases

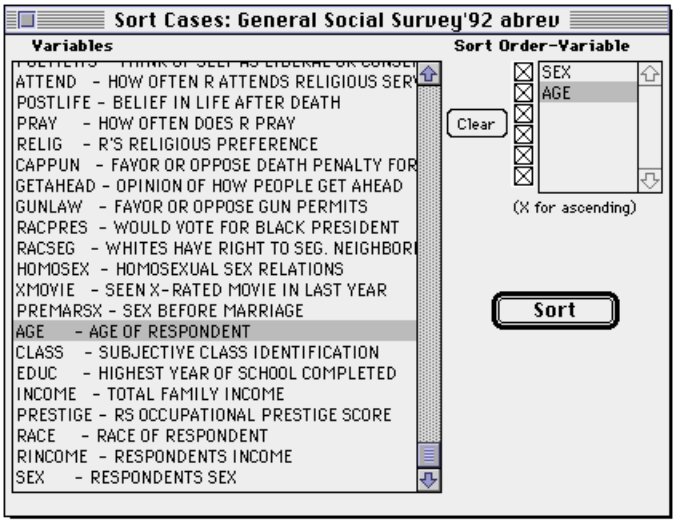
To generate properly sorted report listings or view sorted information directly from the data matrix window, the sort utility can be employed using as many as six separate sort keys.

Cases can be sorted on a particular variable in either ascending (up) or descending (down) order.

Click the sort icon or select **Sort...** from the **Utilities** menu.

The **Sort Cases** window contains a list of all variables from which up to six variables may be selected for use in sorting the data matrix.

In the example below, all cases will be sorted by age, from youngest to oldest, separately for males and females.

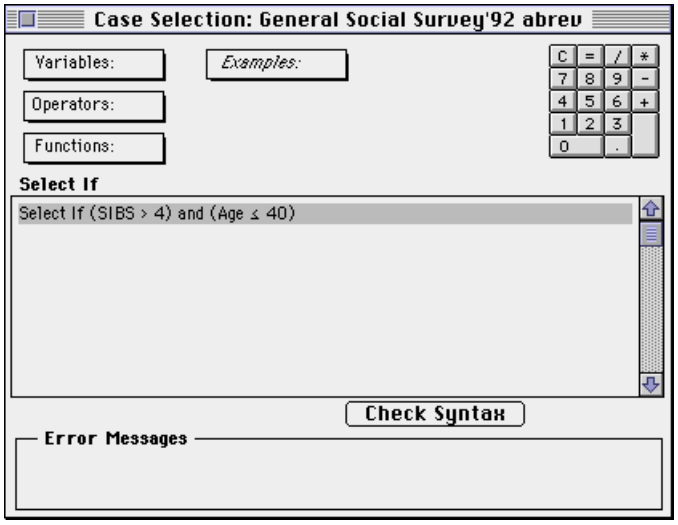


Selecting cases

Being able to select certain cases for analysis is essential in many studies. For example, from among your company's sales staff you may want to focus on only those staff members working in the western half of the country. Or, from among a list of companies, select only those which have been in business for more than 5 years, or report annual sales in excess of $100 million. Selecting certain cases on the basis of specific criteria is easy with Flo•Stat.

Click the select cases icon or choose **Select Cases...** from the **Utilities** menu.

In the example below, statistics will be generated on only those cases in which the respondent reports having four or more siblings and is less than or equal to 40 years of age.



Additional examples of case selections are available in the pull down menu marked *Examples* at the top of the window.

While it is not required, it is recommended that each statement's syntax be checked before moving on to a statistical procedure. To do so, click the **Check Syntax** button to test the statement's syntax.
The **Case Selection** window has many similarities to the Transformation window. The same four pull down menus appear along with the graphic numeric keypad and the text editor field in the center of the window. Besides having a different purpose than the transformation window, the Case Selection window has a different affect on the data matrix.

In the **Transformation** window, when the **Compute** button was clicked action was immediate - new data were created or existing data altered. The Case Selection window, on the other hand, behaves more like a switch being turned on. As long as a **Select If** statement in the **Case Selection** window is highlighted, all statistical analyses from then on are subject to the selection criteria set in the window. To turn the selection off, simply return to the window and click once on the statement.

Note the **Select If...** flag which appears below the pull down Table menu in all tabular output when **Case Selection** is in effect.