# Data manipulation in SuperMix spreadsheet

## Contents

## 2.1    Data manipulation

The SuperMix spreadsheet can be manipulated in various ways. Rows and columns can be changed directly, and simple computations or more complex built-in functions can be used in individual cells. Some of these manipulations are discussed and illustrated in the following sections, using **demo.ss3**. There are 15 cases (patients) in the data set.

The spreadsheet window for **demo.ss3** is opened as follows:

- o  Use the **Open** option on the **File** menu of the main window to load the **Open Spreadsheet** dialog box.
- o  Browse for the file **demo.ss3** in the **Examples** folder.
- o  Select the file and click on the **Open** button to open the following SuperMix spreadsheet window.

The variables include:

- o Group is a variable with 3 categories, indicating the group number of the patient (5 patients in each group).
- o Age is the age of the patient.
- o WT_kg denotes the weight of the patient in kg.
- o PFat is a measure of percentage body fat.

### 2.5.1 Basic data manipulations

It is important to note that any change of the data file will not be saved unless you use the **Save** option on the **File** menu.

### Cells

A careful examination of the data shows that the Age entry of the 3rd observation is 330.00. This is obviously a typing error. Upon further investigation, it turns out that the correct age value is 33. To correct this error, select the cell, change the value of the formula box from 330.00 to 33.00, and then click on the **Apply** button.
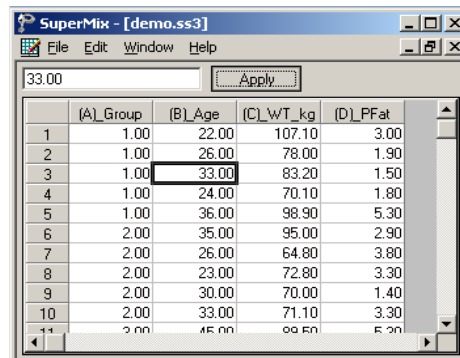


### Rows

To work with a row (or rows) of the data file, click on the row tab(s) to select the complete row(s) and then right-click on the selection to display all the options from the pop-up menu.

### Insert a row

For example, to insert another row (observation) between the first and the second rows, select the second row by clicking on the row 2 tab, right-click on the selected row to activate the menu and select the **Insert Row** option to create the window as shown below.
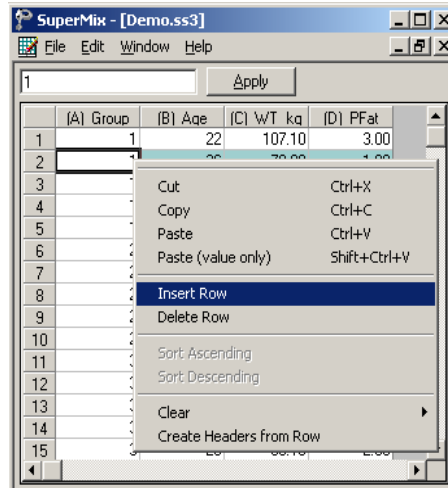
An empty row is added to the spreadsheet above the previous second row and the total sample size is changed to 16 as shown below.



**Delete a row**

To delete the empty row that was inserted, select the second row by clicking on the row 2 tab. Right click on the selected row and select the **Delete Row** option to delete the second row.

**Columns**

To work with a column (or columns) of the data file, first click on the column header(s) to select the column(s) and then right-click on one of the selected columns to see all the options listed on the pop-up menu.

**Insert a column**

To insert another variable between (A)_Group and (B)_Age, first click on the header of (B)_Age, right-click on the column to activate the menu and select the **Insert Column** option to create the window as shown below.
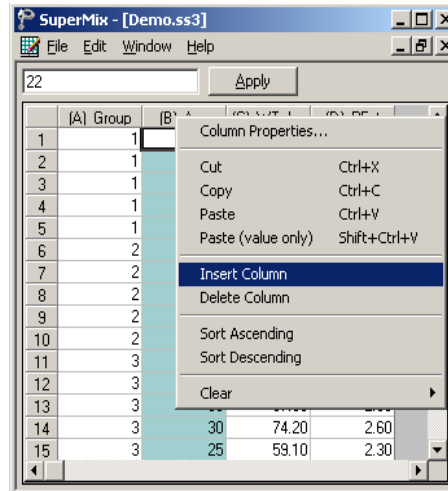


A new column (variable) is added to the spreadsheet. Change the variable name by selecting the column header, right-click and select **Column Properties** to load the dialog box as shown below.
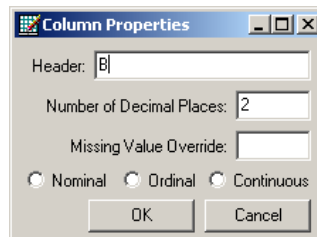


Input a variable name, such as NewVar, and then click on the **OK** button to return to the spreadsheet window as shown below.

### Delete a column

To delete the NewVar column that was inserted, select the variable NewVar by clicking on the column header. Right-click and select the **Delete Column** option to delete column B.

### 2.5.2   Simple computations

### Assigning values to a new variable

In **demo.ss3**, the variable WT_kg is a variable denoting weight in kilograms. We would like to use the variable WT_lb, that is, the corresponding weight in pounds. To create this variable, first insert a column and change the column header to WT_lb as illustrated earlier in Section 2.5.1 to generate the following spreadsheet window.



Select the column containing the variable WT_lb, input the function **2.20462*(C1)** in the formula box and click on the **Apply** button to get the new variable WT_lb as shown below. The formula applies to each row of (D)_WT_lb provided that

- o  this column is selected (highlighted)
- o  the first cell of the variable(s) in the formula, in this case C1, is referenced.



Save the changes to **demo.ss3** by clicking on the **Save** option on the **File** menu.

## 2.5.3   Built-in functions

**LN function**

In mixed-effects models, we often consider the natural log of a variable. For example, the natural log of Age in **demo.ss3** may be a more appropriate explanatory variable than the original age given in years. The variable LnAge can easily be created by using the options available in the SuperMix spreadsheet.

First, create a new column with the header LnAge. Next, select the column containing the variable LnAge, input the function **LN(B1)** in the formula box, and click on the **Apply** button. Each value of the new variable LnAge is the natural log of the corresponding values of the variable Age as shown below.

### SQRT function

To add another variable, SqrtAge, which is the square root of Age, we proceed as follows. As above, first create a new column with the header SqrtAge.

Select the column containing the variable SqrtAge, input the function **SQRT(B1)** in the formula box, and click on the **Apply** button. Each value of the new variable SqrtAge is the square root value of the corresponding value of the variable Age as shown below.



Save the changes to **demo.ss3** by clicking on the **Save** option on the **File** menu.

Table 2.16 contains a selection of the built-in functions in SuperMix. A list of values can be any of the following types:

- o (B1, B2, B3) is the selection of the first three values of the variable in column B.

- o (A1:A15) selects all the values of the variable in column A. Starting with the first and ending with the one in row 15.
- o (A1:A5, A7, A11:A15) selects the values of row 1 to row 5, row 7, and row 11 to row 15 of column A.
- o (A3:A6, B7, D12:D15) contains a list of values from more than one column. This selection includes the values of row 3 to row 6 of column A, row 7 of column B and row 12 to row 15 of column D.

**Table 2.16: Selection of SuperMix functions**

| Function | Definition |
|---|---|
| ABS(value) | Absolute value |
| AVERAGE (list of values) | Average |
| EXP(value) | Exponent base e |
| LN(value) | Natural log |
| LOG(value) | Logarithm |
| MAX(list of values) | Maximum value |
| MEDIAN(list of values) | Median |
| MIN(list of values) | Minimum value |
| MODE(list of values) | Mode |
| SQRT(value) | Square root |
| SQUARE(value) | Square |

### 2.5.4   Other useful data manipulations

The data manipulation capabilities of the SuperMix spreadsheet window can be used to create interaction terms, essentially a product of variables, for use in modeling. It can also be used to perform grand mean centering of variables.

### Absolute references

If you do not want SuperMix to adjust references when you copy a formula to a different cell, use an **absolute** reference. For example, if your formula multiplies cell A5 with cell C1 (=A5*C1) and you copy the formula to another cell, SuperMix will adjust both references. You can create an absolute reference to cell C1 by placing a **dollar sign** ($) before the parts of the reference that do not change. To create an absolute reference to cell C1, for example, add dollar signs to the formula as follows: = A5*$C$1.

### Creating an interaction term

Suppose that we want to study the possible interaction between a subject's age and weight (in pounds). This product of Age and WT_lb, is created in the SuperMix spreadsheet window as follows.

First, create a new column with header Age_WT. Then, select this column, and input the function **(B1)\*(E1)** in the formula box. Click on the **Apply** button. Each value of the new variable Age_WT is equal to the product of the corresponding values of Age and WT_lb as shown below.



## Grand mean centering

It is often useful to center a predictor variable around its grand mean. To illustrate, we grand mean center the variable PFat.



To do so, first create a new column with the header of PFat_Mea. Then, select the PFat_Mea column, input the function **(H1)–AVERAGE(H$1:H$15)** in the formula box and click on the **Apply** button. Each value of the new variable of PFat_Mea now contains the difference

between the corresponding original PFat value and the grand mean of all the PFat values. As illustrated above, the spreadsheet functions are not case sensitive.