Tutorials and worked examples for simulation, curve fitting, statistical analysis, and plotting.
http://www.simfit.org.uk

It has been explained that data for analysis by SIMFIT must be supplied in the form of a rectangular table of numbers with no missing values. Further, row and column labels can be present as long as they either have no spaces as in `Time_of_Day` or are double quoted as in `"Time of Day"`.

So, for many purposes, it is adequate merely to copy the table to the clipboard from a spreadsheet program such as Excel or Calc, and then use the [Paste] button on the SIMFIT file opening control, which simply makes a temporary file in SIMFIT format from the clipboard data. There is nothing wrong with this way of proceeding but two things must then be realized.

1. **Archiving**

   Doing it this way using copy and paste means it has to be done every time you want to repeat the process, for instance, to fit several models to the same data. Saving well-named data files with short meaningful titles makes retrospective use so much easier. In addition, it permits the gathering of files together to fit several data sets simultaneously, or plot multiple sets of coordinates, say using SIMFIT library files.

2. **Environments**

   Many analytical procedures require more than just the data table. For instance.

   (a) Setting parameter starting estimates and limits for nonlinear model fitting.

   (b) Providing initial conditions and range for numerical solution of differential equations, as well as the limits and number of points for plotting trajectories.

   (c) Defining starting clusters for K-means clustering.

   (d) Indicating variables to include or suppress in multivariate analysis.

   (e) Assigning variables to groups as in canonical correlation.

   (f) Adding row and column labels to data files to use in multivariate analysis plots.

   Evidently it would be extremely tedious to have to do this every time analysis is carried out using the clipboard to copy and paste spreadsheet data data into SIMFIT for analysis, as these additional parameters would then have to be edited interactively each time for use by the calling program.

Before proceeding any further an important point must be made about such data files.

SIMFIT data files are simple ASCII text files which means that, given such a file, it is easy to edit it retrospectively in any text editor, such as notepad, in order to add, remove, or edit any of the information in it.

However, if rows of numbers are added to the data table or removed from it, then the first integer on the second line of the file which indicates the number of rows must be corrected.

There is also another matter which may cause concern if it is not understood.

Numbers in data files prepared by SIMFIT are usually represented in scientific notation with a fixed number of significant figures. So if you input 1, 2, 3 from the clipboard it will be written to file as 1.0000000E+00, 2.0000000E+00, 3.0000000E+00 or similar. Of course calculations by SIMFIT are carried out to 64-bit precision, so in the unlikely event that you do want to input data with more significant figures, just input in CSV format.

These are the ways to create data files in the SimFIT format.

1. Paste in from the clipboard but then save the temporary file created with a new name.
2. Use a macro with your spreadsheet program. For instance `simfit6.xls` with Excel.
3. Read a spreadsheet export file into program **maksim** or paste a table in from the clipboard.
4. Create a data file using a text editor such as **notepad**, or better **notepad++**.
5. Create a data file using one of the SimFIT programs such as **makfil** for curve fitting files, or **makmat** for arbitrary data tables.

Having created a data file then any environments that need to be added can be pasted in anywhere at the end of the data table using a text editor.

For small data sets it may be convenient to create data files using the SimFIT file creating programs **makmat** and **makfil** which guarantees correctly formatted data files. Then, for simple editing to correct, add, or deleet a few values it is probably easiest to use a text editor like **notepad**. However, when it comes to serious editing of data files the SimFIT data file editing programs **editmt** and **editfl** provide many procedures that are very difficult if not impossible to perform using a text editor or spreadsheet program. So the following features of the SimFIT data preparation and editing programs should be realized before the functionalities are discussed.

The SimFIT data preparation programs may only be useful when creating a file from relatively small data sets, but do have some advantages that will be outlined. The SimFIT file editing programs read in a source file and output a target file, but the source file will never be altered.

Standard data files

SimFIT program **makmat**

With this program you can simply type in numbers into an empty grid in the usual way. However, in order to facilitate the creation of special data sets, matrices can be zeroed with selected numbers, which can be very useful where diagonals have special significance. After filling in all the cells the matrix can be edited before exit. If the file creation process is closed before all the cells are filled in then uncompleted cells are set to 1 which can only be changed by further editing.

SimFIT program **editmt**

Some of the functionality is summarized.

- Patches of the matrix can be written to file and new patches can be added from files. This is a very useful way to fuse multiple data sets that all have the same number of rows, or alternatively the same number of columns.
- Individual rows or columns can be deleted or restored which is a convenient way to swap rows or columns
- Individual rows or columns can be transformed by algebraic, probability, or trigonometric functions.
- Individual rows or columns can be set to fixed values.
- The total matrix can be edited to change selected values or for such processes as centering, scaling, or centering and scaling rows or columns. Such editing can be aborted at any stage without overwriting the current default matrix.
- On exit the title and trailer section of the data set can be edited.

Of course users must be aware of the need to proceed in an orderly and methodical fashion if these procedures are to applied sensibly with the desired mathematical results.

Curve fitting Files

There are also several special considerations with curve fitting files that must be considered briefly here, noting that there is much more detail on this subject in the SimFIT reference manual.

These have either two columns x and y, or three columns x, y, and s which have the following meanings.

X in column 1

The independent variable known with great accuracy, e.g. time, weight, concentration.

Usually x values are increasing order because of four reasons.

1. The first x_i, y_i pairs are used to obtain starting estimates for model parameters that have influence at low x.
2. The last x_i, y_i pairs are used to obtain starting estimates for model parameters that have influence at high x.
3. Numerical estimation of differential equations is best done sequentially onwards from the initial conditions to avoid unnecessary re-calculations.
4. SimFIT parses the data first time and assigns logical variables to identify groups of replicates so that the model error is only calculated for the first member of each group of replicates to avoid unnecessary re-calculations.

Y in column 2

The measured response assumed to result from random experimental error added to a deterministic effect.

S in column 3

The weights for fitting are calculated using $w_i = 1/s_i^2$.

There are five possibilities, all of them being controversial.

1. All $s_i = 1$. Constant variance is assumed.
   This is also the case when only two columns x the y are supplied. In other words, there is no such thing as unweighted regression.
2. The s_i are investigated independently and are know accurately.
   This is unquestionably the best method but is seldom used.
3. The s_i are estimated using the sample standard deviations based on replicates.
   This is only acceptable if the sample sizes are sufficiently large, definitely ≥ 5.
4. The s_i are assumed to be functions of the data i.e., y_i.
   This means that replicates will be weighted differently.
5. The s_i are assumed to be functions of the best-fit model.
   Whatever functional dependence is assumed the weights will be different for each iteration and only make sense if the fitted model is actually the correct one, the assumed functional dependence is correct, and in addition the weights only become asymptotically reliable as the regression converges to the solution point.

SimFIT program **makfil**

The user can choose to make a x, y or a x, y, s file and can choose whether to input x in increasing order or, for special use where this is not necessary, in arbitrary order. Note that x can also be input for data such as those from doubling dilution experiments as described in the information available when the program is run, but this option is only to be used when it is properly understood. If the option to make a x, y file is chosen, the output file will have a third column with $s_i = 1$.

As this program is designed to prepare data files for curve fitting you will be forced to only input x in increasing order unless this option has been suppressed, and if you choose to make a x, y, s file, you will be forced to input meaningful s_i values with $s_i > 0$.

Note that you can plot the x, y values when the data input phase has been completed, and this is a very valuable way to check that sensible data have been input. So, if outliers are seen suggesting a typing error, this editing can be done before exit.

SimFIT program **editfl**

Just as with **editmt** you specify a source file and a target file in case an undo functionality is required, and you can fuse multiple curve fittng data files together. A valuable feature is to rearrange data so that x is in nondecreasing order, and a check is provided to make sure that s_i, y_i pairs suggesting a sensible signal to noise ration have been input. If replicates have been provided these can be used to calculate weights and error bars, although some SimFIT programs can do this at run-time.

Before final exit the ability to edit the title and trailer section is provided in case environments such as

```
begin{limits} ...  end{limits}
```

need to be added or updated.

The great advantage of using programs **makfil** and **editfl** is that the extensive checks for consistency, x order, sensible signal to noise ratios for y_i, s_i pairs, and visual checking for accidental outliers during the data input phase, greatly decreases the chance of a spurious result from trying to fit badly formatted data.

Missing values

Data tables used by SimFIT for statistical analysis must have no missing values. So, if you are in the unfortunate situation of requiring such dishonesty for the greater good, then you will have to use the Excel macro called `simfit6.xls`, or some other program dedicated to cheating in (one hopes) the least objectional way.

However there are sometimes cases where analysis of a matrix with unequal length columns can proceed and where missing values do not need to be replaced by estimates. For instance, 1-way ANOVA, analysis of multiple samples for equality of variance, creating box and whisker plots, etc. Such situations can be handled using individual column vectors, specifying data samples using a library file, or choosing individual samples from your project archive. Note that now these procedures can also use incomplete matrix files which will be described separately.