Tutorials and worked examples for simulation, curve fitting, statistical analysis, and plotting.
https://simfit.uk
https://simfit.org.uk
https://simfit.silverfrost.com

Many procedures require sets of data with possibly unequal sample sizes, and the standard ways to deal with such situations in SimFIT has been to use one of the following three techniques.

1. Opening a library file to reference sample files, e.g., anova1.tfl.

2. Opening individual files like column1.tf1, column1.tf2 such as are specified in anova1.tf1.

3. Opening collections of files like column1.tf1, column1.tf2 from your vector project archive.

There is now also the possibility of using an incomplete matrix file with missing values where the columns are padded out with non-numeric character strings like blanks, X, NA, etc. However, note that this technique is not intended for situations where a value is missing because an observation was not recorded, or because an outlier is suspected, and an estimate is to be used in order to allow statistical analysis to proceed. For this the Excel macro simfit6.xls can be used. Rather it is simply designed so that a single matrix can be used for convenience whenever the numbers of numeric data are not the same in every column of a matrix.

The format for incomplete matrix files

Here are the rules.

- In order for such data to be interpreted correctly the matrix columns must be separated in an unambiguous way, for example, using commas, semicolons, or tabs.

- The separators used cannot occur elsewhere in the file except as column separators.

- Down any given column the missing values can occur at the level of any row.

- Unambiguous non-numeric character strings must be used to pad out the cells with missing values.

- Headers and trailers can be added as long as they do not contain the column separators.

- If labels are required for plotting they cannot be added to the trailer but must be added from the configuration files, or interactively, e.g. from a separate labels file.

Example 1: comma separated variables

The SimFIT test file incomplete.tf1 uses commas as separators as follows.

```
23,    29,    38,    30,    31
27,    25,    31,    27,    33
26,    33,    28,    28,    31
19,    36,    35,    22,    28
30,    32,    33,    33,    30
  ,    28,    36,    34,    24
  ,    30,      ,    34,    29
  ,    31,      ,    32,    30
```

Such a file is just a normal comma separated file such as a `*.csv` file exported from any spread sheet program. Of course this format cannot be used if continental notation is used for non-integer values, such as using 1,234 for 1.234. Cells can use scientific notation for 1.234 such as 1.234E+00 or 0.1234E+01, 1.234e0, etc.

Example 2: semicolon separated variables

The SIMFIT test file `incomplete.tf2` uses semicolons as separators as follows.

```
23;   29;   38;   30;   31
27;   25;   31;   27;   33
26;   33;   28;   28;   31
19;   36;   35;   22;   28
30;   32;   33;   33;   30
  ;   28;   36;   34;   24
  ;   30;     ;   34;   29
  ;   31;     ;   32;   30
```

Such a file would be output as a `*.csv` file from a spreadsheet program set up to use commas instead of the decimal points used in standard scientific documents.

Example 3: tab separated variables

This how test file `incomplete.tf3` will look in **Notepad**

```
23    29    38    30    31
27    25    31    27    33
26    33    28    28    31
19    36    35    22    28
30    32    33    33    30
      28    36    34    24
      30          34    29
      31          32    30
```

while here is how it would look in **Notepad++** if the option to display tabs is switched on.
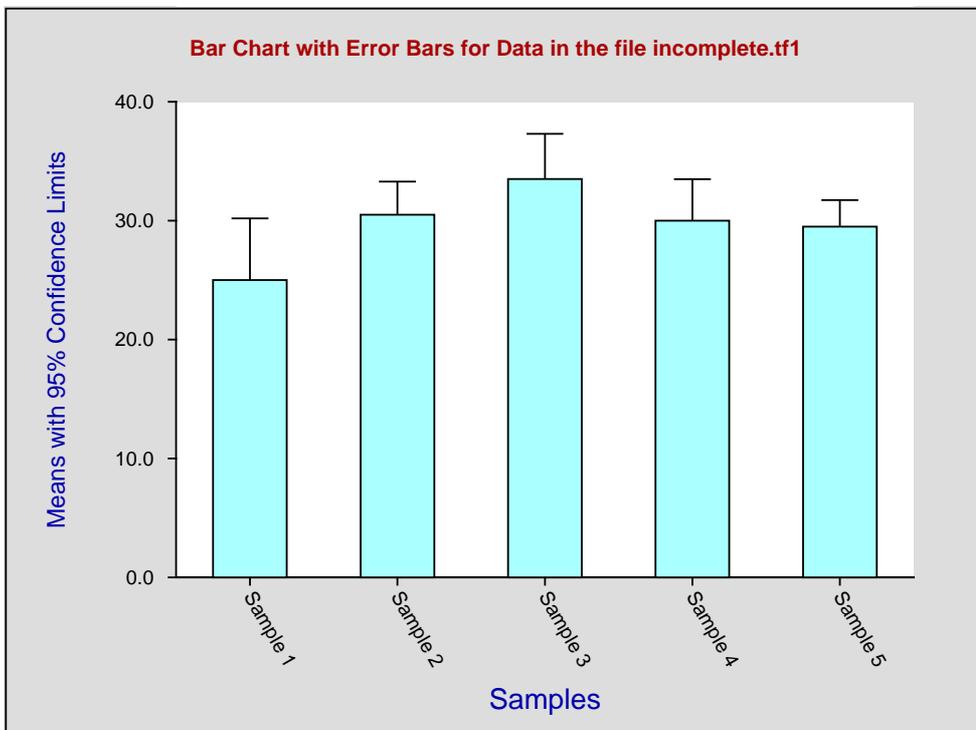
```
23->29->38->30->31
27->25->31->27->33
26->33->28->28->31
19->36->35->22->28
30->32->33->33->30
  ->28->36->34->24
  ->30->  ->34->29
  ->31->  ->32->30
```

This is usually the default format to copy to the clipboard or export to files with many spreadsheet programs, as it allows either continental or scientific notation to be used for non-integer values. However, considerable care must be taken when editing such files in a simple editor like **Notepad** which does not display tabs, and the more advanced editior **Notepad++** must be used to prevent accidental deletion of tabs, when it would no longer be recognized by SIMFIT as an incomplete matrix file. Here are some of the graph plotting styles that can be used with both incomplete and complete SIMFITdata files.
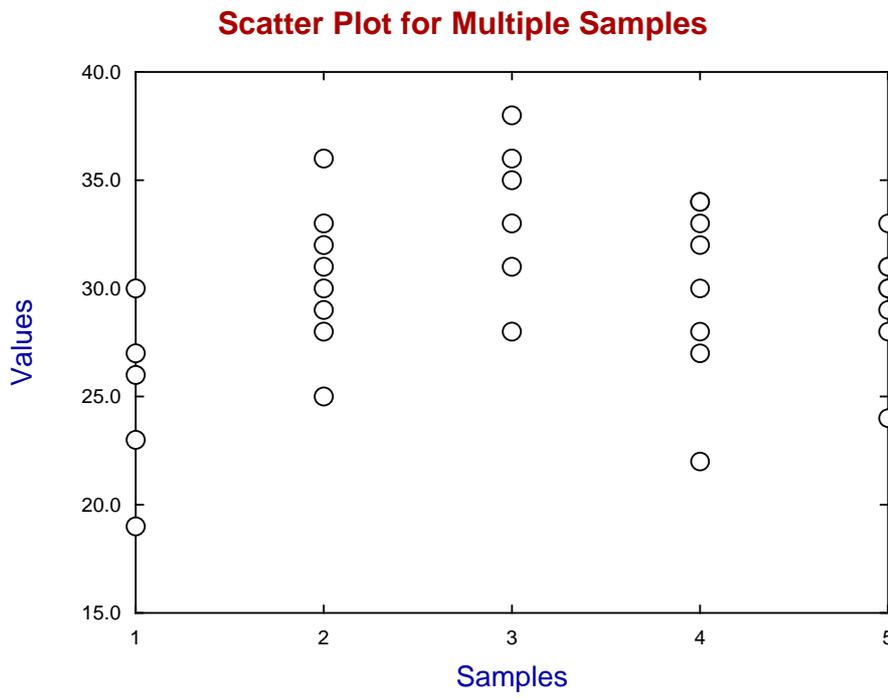
**Plot 1:** Means as symbols with error bars calculated from individual samples

Error Bars for Data in the file incomplete.tf1

Means and 95% Confidence Limits

Samples

Plot 2: A bar chart plot with added error bars calculated as for Plot 1

Bar Chart with Error Bars for Data in the file incomplete.tf1

Means with 95% Confidence Limits

Sample 1   Sample 2   Sample 3   Sample 4   Sample 5

Samples

Plot 4: Showing all the sample values as clusters

Scatter Plot for Multiple Samples

Plot 5: A box and whisker plot with added data

Box and Whisker Plot with Scattered Data
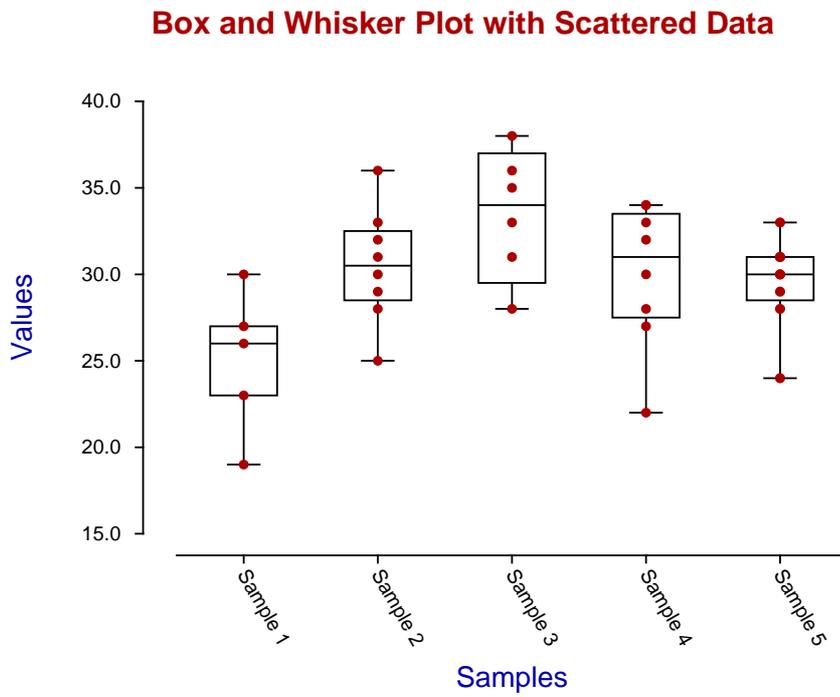
Plot 6: Connecting the percentiles by straight line segments

Range and Percentiles