



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

Each time a data set is analyzed the results are written to a file called `f$result.txt`, and the current files are renamed so that the existing `f$result.txt` becomes `f$result.001` while `f$result.001` becomes `f$result.002` and so on. These can be viewed using the [Results] option from the main SIMFIT menu.

These SIMFIT results files are formatted so that the numbers displayed only contain the number of significant figures that are meaningful in context. For instance, probabilities will usually only have four digits after the decimal point, which indicates that it does not make any sense to consider any subsequent digits for purpose of statistical testing, and in any case probability estimates will not be accurate for more than about four digits. Furthermore, as experimental data are rarely more accurate than about three or four significant figures anyway, it may be wishful thinking to ever consider more than say six. In addition, the tables are formatted using a fixed font with scientific notation to line up column entities irrespective of absolute size, and many users do not want this in a thesis or published document. Naturally, these arguments do not apply to integers.

As an example consider the following case with the title Table 1.

Table 1						
1-Way Analysis of Variance: 1 (Grand Mean 4.316E+01)						
Transformation:- x (untransformed data)						
Source	SSQ	NDOF	MSQ	F	p	
Between Groups	2.19344E+03	4	5.48316E+02	5.61546E+01	0.0000	
Residual	2.44130E+02	25	9.76520E+00	...	...	
Total	2.43757E+03	29	...	...	...	

Actually, most users would want to import such a table formatted as tabbed-text, html, xml, or  $\LaTeX$  into documents such as a report, thesis, or publication looking something like Table 2.

Table 2						
1-Way Analysis of Variance: 1 (Grand Mean 43.16)						
Transformation:- x (untransformed data)						
Source	SSQ	NDOF	MSQ	F	p	
Between Groups	2193.44	4	548.316	56.1546	0.0000	
Residual	244.130	25	9.76520			
Total	2437.57	29				

Or even, for those with artistic leanings, possibly something like Table 3.

Table 3					
1-Way Analysis of Variance: Grand Mean 43.16					
Transformation: x (untransformed data)					
Source	SSQ	NDOF	MSQ	F	p
Between Groups	2193.44	4	548.316	56.1546	0.0000
Residual	244.130	25	9.76520		
Total	2437.57	29			

This article explains the procedures required to export tables from from SIMFIT results files into forms suitable for inclusion into word processors, spreadsheet programs, website scripts, or even professional document preparation systems such as L<sup>A</sup>T<sub>E</sub>X. Also, decimal points can be replaced by commas as in continental notation if required.

## The procedure

It is important to realize that the [Results] option from the SIMFIT main menu gives access to all the currently saved results files.

- **Choosing a results file.**

Early versions of SIMFIT allowed users to name results files individually to avoid anything being lost. However, now that up to 100 results files are saved and users have the option [Results] from the main SIMFIT menus from which to view, print, save, edit, or export tables, this is no longer usually necessary. Clearly, if results are always required for retrospective use, regular back-up or saving will be necessary.

- **Extracting a table.**

From the [Extract tables] option view the file to make sure it is the one required then copy to the clipboard only the table required along with any associated header and trailer sections ... but nothing else.

- **Preliminary editing.**

Sometimes editing of the file is required to make sure that every row of the table has exactly the same number of columns. So note that, for extracting a table there can be no empty cells, and each cell must contain precisely one word. Any column titles must be edited so that they consist of one word, for instance changing `Time of Day` to `Time_of_Day`, or filling empty cells by three dots. Added underscores and sets of three contiguous dots can be removed when the final table is written to file. A pre-processing option is provided for editing before attempting to create a table.

Note that often tables have cells with added comments relating to goodness of fit or results of statistical testing, and these do not need to be underscored. There are also special tables with only two columns containing several words in some cells, and Example 1 later will make this clear.

- **Viewing the hashtag table.**

The algorithm attempts to identify cells in a table by inserting a hashtag between every column. If the algorithm succeeds there would be no need to view this hashtag table. However this option should be switched on until the process of the algorithm is understood, or if it fails and you need to see why.

The hashtag table is clarified in Example 1.

- **Headers and trailers.**

Frequently tables have header and trailer sections that are descriptive and not part of the table itself. As these can have strings of word and numbers that would confuse the algorithm checking that every row must have the same number of columns, they must be identified. This is done by using buttons on a window that allows the header and trailer lines to be highlighted. If this is not done the table creating algorithm will fail.

Selecting headers and trailers is clarified in Example 2.

- **Saving the table.**

For programs that produce Windows quality hardcopy the table should be saved as html or xml as these will import directly into word processors or spreadsheets. Tabbed-text is also available but is much less versatile than html or xml, and L<sup>A</sup>T<sub>E</sub>X is available for those up to it.

- **Fine tuning.**

Inevitably there will often be the need for dealing with details. For instance, users will sometimes want to replace `alpha` by  $\alpha$  or `chi-sqd` by  $\chi^2$  and this can be done for html, xml, and LaTeX output but

not for tabbed-text. On the other hand html, xml, and  $\LaTeX$  have reserved letters and these must be dealt with retrospectively.

For instance, consider the transformation of the following expressions which can be done in html, xml, and  $\LaTeX$  but not in tabbed text.

Original	Transformed
>=	$\geq$
=<	$\leq$
alpha	$\alpha$
beta	$\beta$
delta	$\delta$
gamma	$\gamma$
lambda	$\lambda$
infinity	$\infty$
$P(\text{chi-sqd} \geq \text{TS}) = 0.2037$	$P(\chi^2 \geq \text{TS}) = 0.2037$

Where ambiguity could arise in  $\LaTeX$  using underscores or similar special characters they will be replaced by question marks. So  $\LaTeX$  users should search for ? characters to replace for the final table.

- **Padding with zeros.**

As the [Extract tables] option will never remove significant figures a problem arises if users wish to replace numbers in scientific notation by floating point representation. In order to extend the range over which this can be done, padding zeros can be introduced as illustrated in this next table.

Scientific notation	Padding	Floating point representation
1.234E-01	0	.1234
1.234E+00	0	1.234
1.234E+01	0	12.34
1.234E+02	0	123.4
1.234E-02	2	0.01234
1.234E-01	2	0.12340
1.234E+00	2	1.23400
1.234E+01	2	12.3400
1.234E+02	2	123.400
1.234E+03	2	1234.00
1.234E+04	2	12340.0
1.234E-04	4	0.0001234
...	...	...
1.234E+06	4	1234000.0

Evidently increasing the number of padding zeros increases the range over which transformation from scientific to floating point representation can be achieved, and the default is two which allows a wide range, but four ensures that a mixture of transformed and untransformed numbers will occupy the same width in the columns of a table.

However, there is here a problem because adding padding zeros could suggest all trailing zeros are meaningful. For instance, the number 1.2341213179 stored in the computer could be written as 1.234E+00 in `SIMFIT` output because the analysis in question only justifies accuracy or meaning for up to four significant figures. However, 1.23400 could be mistaken for indicating the internal representation with twelve significant figures. So users may wish to suppress trailing zeros in such cases, noting that this could result in numbers with different widths in a column. In any case, the transformation from scientific notation to floating point can be switched off.

Sometimes, for instance with numerical analysis where more significant figures are justified than with data analysis, eight significant figures are output, and some procedures can optionally allow more. In addition special DLLs dedicated for particular routines can be supplied for this purpose.

## Example 1

This example shows the transformation of a special type of SIMFIT table containing just two columns separated by equals signs (i.e., =) and containing cells with multiple words. As long as the equals signs are perfectly lined up and there is no header or trailer section, this type of table can always be transformed. Here is the table with no header or trailer sections as extracted using the [Results] then [Extract tables] options from the main SIMFIT menu using the results file f\$result.txt following the exhaustive analysis of a vector process used to analyze data contained in the default test file normal.tfl.

Sample size	=	50
Minimum value	=	-2.20820E+00
Maximum value	=	1.61750E+00
Coefficient of skew	=	-1.66905E-02
Coefficient of kurtosis	=	-7.68395E-01
Lower Hinge (25th percentile)	=	-8.55015E-01
Median value (50th percentile)	=	-9.73615E-02
Upper Hinge (75th percentile)	=	7.85965E-01
Sample mean	=	-2.57897E-02
Sample standard deviation	=	1.00553E+00
Coefficient of variation (CV%)	=	> 100%
Standard error of the mean	=	1.42203E-01
Upper 2.5% t-value	=	2.00958E+00
Lower 95% con lim for mean	=	-3.11558E-01
Upper 95% con lim for mean	=	2.59978E-01
Sample variance	=	1.01109E+00
Lower 95% con lim for variance	=	7.05519E-01
Upper 95% con lim for variance	=	1.57006E+00
Shapiro-Wilks W statistic	=	9.62693E-01
Significance level for W	=	0.1153
Conclusion	=	Tentatively accept normality

This is the corresponding intermediate hashtag table.

Sample size	#	50
Minimum value	#	-2.20820E+00
Maximum value	#	1.61750E+00
Coefficient of skew	#	-1.66905E-02
Coefficient of kurtosis	#	-7.68395E-01
Lower Hinge (25th percentile)	#	-8.55015E-01
Median value (50th percentile)	#	-9.73615E-02
Upper Hinge (75th percentile)	#	7.85965E-01
Sample mean	#	-2.57897E-02
Sample standard deviation	#	1.00553E+00
Coefficient of variation (CV%)	#	> 100%
Standard error of the mean	#	1.42203E-01
Upper 2.5% t-value	#	2.00958E+00
Lower 95% con lim for mean	#	-3.11558E-01
Upper 95% con lim for mean	#	2.59978E-01
Sample variance	#	1.01109E+00
Lower 95% con lim for variance	#	7.05519E-01
Upper 95% con lim for variance	#	1.57006E+00
Shapiro-Wilks W statistic	#	9.62693E-01
Significance level for W	#	0.1153
Conclusion	#	Tentatively accept normality

The hashtag table is very useful for detecting the source of errors. The table making algorithm attempts to locate the position separating columns and writes a hashtag there. If every row has the same number of columns then every row will have the same number of hashtags and the algorithm has succeeded. Observing this hashtag table when the algorithm has failed will allow you identify then correct the error.

Here is the selected table as it would be written to the output file.

Sample size	50
Minimum value	-2.2082000
Maximum value	1.6175000
Coefficient of skew	-0.0166905
Coefficient of kurtosis	-0.7683950
Lower Hinge (25th percentile)	-0.8550150
Median value (50th percentile)	-0.0973615
Upper Hinge (75th percentile)	0.7859650
Sample mean	-0.0257897
Sample standard deviation	1.0055300
Coefficient of variation (CV%)	> 100%
Standard error of the mean	0.1422030
Upper 2.5% t-value	2.0095800
Lower 95% con lim for mean	-0.3115580
Upper 95% con lim for mean	0.2599780
Sample variance	1.0110900
Lower 95% con lim for variance	0.7055190
Upper 95% con lim for variance	1.5700600
Shapiro-Wilks W statistic	0.9626930
Significance level for W	0.1153
Conclusion	Tentatively accept normality

Here it is with a few minor cosmetic changes.

Exhaustive analysis of a vector	
Sample size	50
Minimum value	-2.2082000
Maximum value	1.6175000
Coefficient of skew	-0.0166905
Coefficient of kurtosis	-0.7683950
Lower Hinge (25th percentile)	-0.8550150
Median value (50th percentile)	-0.0973615
Upper Hinge (75th percentile)	0.7859650
Sample mean	-0.0257897
Sample standard deviation	1.0055300
Coefficient of variation (CV%)	> 100%
Standard error of the mean	0.1422030
Upper 2.5% t-value	2.0095800
Lower 95% confidence limit for mean	-0.3115580
Upper 95% confidence limit for mean	0.2599780
Sample variance	1.0110900
Lower 95% confidence limit for variance	0.7055190
Upper 95% confidence limit for variance	1.5700600
Shapiro-Wilks W statistic	0.9626930
Significance level for W	0.1153
Conclusion:	<i>Tentatively accept normality</i>

## Example 2

From fitting a two-exponential model to data in the test file `exfit.tf4` using program `exfit` the following results can be extracted from the results file.

For best-fit 2-exponential function					
Parameter	Value	Std.Error	Lower95%cl	Upper95%cl	p
A(1)	8.5255E-01	6.7731E-02	7.1332E-01	9.9177E-01	0.0000
A(2)	1.1765E+00	7.4779E-02	1.0227E+00	1.3302E+00	0.0000
k(1)	6.7935E+00	8.5386E-01	5.0383E+00	8.5486E+00	0.0000
k(2)	1.1121E+00	5.1128E-02	1.0070E+00	1.2172E+00	0.0000
AUC	1.1834E+00	1.4714E-02	1.1531E+00	1.2136E+00	0.0000
AUC is the area under the curve from $t = 0$ to $t = \text{infinity}$					
Initial time point (A) = 3.5983E-02					
Final time point (B) = 1.6110E+00					
Area over range (A,B) = 9.3832E-01					
Average over range (A,B) = 5.9575E-01					

Now the file has an additional head and trailer section so, if the full table is selected, it will have to be highlighted as follows in the header and trailer selection control as shown next, where the header is colored magenta and the trailer colored cyan.

For best-fit 2-exponential function					
Parameter	Value	Std.Error	Lower95%cl	Upper95%cl	p
A(1)	8.5255E-01	6.7731E-02	7.1332E-01	9.9177E-01	0.0000
A(2)	1.1765E+00	7.4779E-02	1.0227E+00	1.3302E+00	0.0000
k(1)	6.7935E+00	8.5386E-01	5.0383E+00	8.5486E+00	0.0000
k(2)	1.1121E+00	5.1128E-02	1.0070E+00	1.2172E+00	0.0000
AUC	1.1834E+00	1.4714E-02	1.1531E+00	1.2136E+00	0.0000
AUC is the area under the curve from $t = 0$ to $t = \text{infinity}$					
Initial time point (A) = 3.5983E-02					
Final time point (B) = 1.6110E+00					
Area over range (A,B) = 9.3832E-01					
Average over range (A,B) = 5.9575E-01					

Using two padding zeros this leads to the following table.

For best-fit 2-exponential function					
Parameter	Value	Std.Error	Lower95%cl	Upper95%cl	p
A(1)	0.852550	0.067731	0.713320	0.991770	0.0000
A(2)	1.176500	0.074779	1.022700	1.330200	0.0000
k(1)	6.793500	0.853860	5.038300	8.548600	0.0000
k(2)	1.112100	0.051128	1.007000	1.217200	0.0000
AUC	1.183400	0.014714	1.153100	1.213600	0.0000
AUC is the area under the curve from $t = 0$ to $t = \infty$					
Initial time point (A) = 0.035983					
Final time point (B) = 1.611					
Area over range (A,B) = 0.93832					
Average over range (A,B) = 0.59575					

However, note that, with this example, three points emerge.

1. Numbers outside the main table will also be transformed into floating point numbers.
2. Equals signs lined up the trailer will not lead directly to secondary tabulation.
3. Some special words, like infinity, will be recognized.

So, because a certain amount of fine tuning will be required, the possibilities for handcrafting are endless. Here, for example, the header is enlarged by adding a formula, while the trailer is added in the form of a footnote to the main table.

For the best-fit 2-exponential function					
$f(t) = A_1 \exp(-k_1 t) + A_2 \exp(-k_2 t)$					
Parameter	Value	Std.Error	Lower95%cl	Upper95%cl	<i>p</i>
$A_1$	0.852550	0.067731	0.713320	0.991770	0.0000
$A_2$	1.176500	0.074779	1.022700	1.330200	0.0000
$k_1$	6.793500	0.853860	5.038300	8.548600	0.0000
$k_2$	1.112100	0.051128	1.007000	1.217200	0.0000
<i>AUC</i>	1.183400	0.014714	1.153100	1.213600	0.0000

$$\text{Area under the curve } AUC = \int_0^{\infty} \hat{f}(t) dt$$

$$\text{Initial time point (A)} = 0.035983$$

$$\text{Final time point (B)} = 1.611$$

$$\text{Area over range (A,B)} = 0.93832$$

$$\text{Average over range (A,B)} = 0.59575$$

### Example 3

A special situation exists with symmetric matrices where just a lower or upper triangle is displayed, and also some other related situations. For instance, following on from the previous example we have the parameter correlation matrix expressed in the following form.

Parameter correlation matrix			
1.0000			
-0.8758	1.0000		
-0.5964	0.8996	1.0000	
-0.8480	0.9485	0.8200	1.0000

Clearly, the algorithm to count the number of columns per row in order to insert hashtags will fail because all the rows have different numbers of columns, unless editing is performed like this.

Parameter correlation matrix			
1.0000	...	...	...
-0.8758	1.0000	...	...
-0.5964	0.8996	1.0000	...
-0.8480	0.9485	0.8200	1.0000

Now transformation would be possible leading to a table such as the following.

Parameter correlation matrix			
1			
-0.8758	1		
-0.5964	0.8996	1	
-0.8480	0.9485	0.8200	1

Another example to consider is from correlation analysis which leads to  $r$  values in the strict upper triangle and significance levels in the strict lower triangle as shown next followed by the extracted table.

Pearson correlation results							
Upper triangle = $r$ , Lower = corresponding two-tail $p$ values							
.....	0.5295	0.2874	0.0662	0.1941	0.6255	-0.5876	0.3010
0.0766	.....	0.3285	-0.0219	0.7930	0.5338	-0.4230	0.3006
0.3650	0.2971	.....	-0.2833	0.2165	0.0264	0.2314	-0.0304
0.8381	1.0000	1.0000	.....	0.2787	0.2837	-0.5238	-0.1166
0.5455	0.0021	0.4992	0.3804	.....	0.2029	-0.1949	0.2144
0.0296	0.0738	0.9351	1.0000	0.5271	.....	-0.4532	0.1360
1.0000	1.0000	0.4694	1.0000	1.0000	1.0000	.....	-0.1696
0.3418	0.3424	1.0000	1.0000	0.5035	0.6735	1.0000	.....
Test for absence of any significant correlations							
H0: correlation matrix is the identity matrix							
Determinant		= 2.476E-03					
Test statistic (TS)		= 4.501E+01					
Degrees of freedom		= 28					
P(chi-sq >= TS)		= 0.0220 Reject H0 at 5% sig.level					

#### Pearson correlation results

Upper triangle =  $r$ , Lower = corresponding two-tail  $p$  values

.....	0.5295	0.2874	0.0662	0.1941	0.6255	-0.5876	0.3010
0.0766	.....	0.3285	-0.0219	0.7930	0.5338	-0.4230	0.3006
0.3650	0.2971	.....	-0.2833	0.2165	0.0264	0.2314	-0.0304
0.8381	1.0000	1.0000	.....	0.2787	-0.2837	-0.5238	-0.1166
0.5455	0.0021	0.4992	0.3804	.....	0.2029	-0.1949	0.2144
0.0296	0.0738	0.9351	1.0000	0.5271	.....	-0.4532	0.1360
1.0000	1.0000	0.4694	1.0000	1.0000	1.0000	.....	-0.1696
0.3418	0.3424	1.0000	1.0000	0.5035	0.6735	1.0000	.....

Test : for absence of any significant correlations

$H_0$  : correlation matrix is the identity matrix

Determinant = 0.002476

Test statistic(TS) = 45.01

Degrees of freedom = 28

$P(\chi^2 \geq TS) = 0.0220$  Reject  $H_0$  at 5% significance level

Here the five dots (.....) denote that the diagonal elements have no meaning and this is just a convenient way to conserve space by having one matrix instead of two. Note that the option to blank out three dots (...) used as temporary column separators does not blank out groups with less three or more than three contiguous dots.

A summary of the options available and procedure to be used comes next.

## Summary

- The first step is to select just one table from the results file.
- This can be followed by an optional pre-processing step to edit the table so that every row has exactly the same number of columns.
- Empty cells must be denoted by a three dot ellipsis (...) and cells containing multiple words must have them joined by underscores or similar.
- Three dot symbols to denote empty cells are deleted from the output file.
- The input table can have optional header and trailer sections if required, but these must be highlighted by the control to select headers and trailers.
- There is an option to transform scientific numbers into floating point format by specifying the number of padding zeros required. This option can be switched off.
- Numbers in the header and trailer will also be transformed in this way.
- If it is required, decimal points in floating point numbers can be replaced by commas.
- If transformation fails then the option for pre-processing should be switched on, and also the hashtag table should be requested. By viewing the hashtag table most errors can easily be diagnosed, then rectified by a re-run using pre-process editing.
- If tabbed-text output is selected the resulting file will have to be input into a spreadsheet program for formatting before importing into a word processing program.
- Both html and xml output can allow a certain number of further changes, like changing alpha into  $\alpha$ , or adding cell borders.
- $\LaTeX$  output will have question marks (?) inserted to replace forbidden character such as underscores which must be edited retrospectively depending the intention, e.g., linking words, or denoting subscripts.

Three further things should be emphasized.

1. Some tables have specialized features such as lined up equals signs that allow multiple words in a column and, as long as every row in the table has an equals sign in exactly the same position, this feature will be recognized.
2. Some `SIMFIT` results files output tables to the display without three dot separators (...) to create a more pleasing effect, but add them to the results files to assist the processing described in this document.
3. There are several widely used tables that can have empty cells and multi-word titles that the parsing routine will recognize and format automatically.

Finally, should you require further worked examples, you can browse the `SIMFIT` tutorials, or the document `w_examples.pdf`, where a large number of alternative display styles are demonstrated.