

SAAM II Version 2.2

Basic Tutorials

Working with Data

	Introduction	Data – 2
Part 1	Importing Data from a Spreadsheet	Data – 19
Part 2	Assigning Weights to Data - 1	Data – 25
Part 3	Assigning Weights to Data - 2	Data – 39
Part 4	Weighting and Unweighting data	Data – 48
Part 5	Data Replicates	Data – 60

This page is intentionally left blank

Working with Data

Prerequisites

The prerequisite for this tutorial is having worked through the **SAAM II** introductory tutorial, “Getting Started with **SAAM II Compartmental**.”

What you will learn in this tutorial

The purpose of this tutorial is to show you how to work with your data in your SAAM II study file. You will learn:

- Basic SAAM II **Data** window features (Introduction)
- How to import data from a spreadsheet (Part 1)
- How to use the variance model options “absolute-relative” and “data-model” (Part 2)
- How to use the different weighting schemes FSD, SD, POIS and GEN (Part 3)
- How to weight and unweight data (Part 4)
- How to work with replicated data (Part 5)

Files Required

Study Files: The study files for this tutorial are

study_0.stu
study_0_nodata.stu
study_0_all.stu
fsd_sd.stu

These files are included as part of this tutorial. The file **study_0.stu** is the same that is installed in the **SAAM II** program folder and referred to in the **SAAM II** examples and tutorials.

Data Files: The data files for this tutorial are

study_0.dat

and the Excel® files

study_0.xls
study_0_all.xls
replicate_data.xls

Introduction

This tutorial will show you how to use the different options in SAAM II to work with your data. Most of what follows in the introduction is also available in the SAAM II Online Help. The different parts of the tutorial will reinforce this information in the SAAM II Online Help and the Users Manual.

Part 1 will show you how to import data from a spreadsheet into the **Data** window, and to check to be sure the data format is correct.

Part 2 will illustrate the different variance model options. The fractional standard deviation weighting scheme will be used. The effect of using data based versus model based, and absolute versus relative weighting will be illustrated. These options are described in more detail below.

Part 3 will examine the different default weighting schemes available in SAAM II, and how the choice of a particular scheme can affect the outcome of your modeling exercise.

Part 4 will illustrate how to unweight a datum, or a series of data, as part of the modeling exercise.

Part 5 will show you how to deal with data replicates. This is the situation where, at each sample time, you may have data replicates. SAAM II deals with this using the **Sample** tool.

The simplest example of a SAAM II **Data** window is provided in study_0; the **Data** window is shown below. This is the most typical application where one has a single set of data collected at specific time points.

```
Data - study_0.dat
DATA
(FSD 0.1)
t      plasma
0.007  46780
0.042  43522
0.125  42535
0.25   40125
0.375  36221
0.5    35562
1.0    28194
2.0    19573
3.0    14403
4.0    11278
5.0    8081
6.0    6999
7.0    5653
8.0    5139
9.0    4210
END
Data Format is okay
```

The data are bounded by the DATA and END statements. The second line in the file, “(FSD 0.1)” specifies the weighting scheme that is to be applied to all data. The next line, “t plasma” contains the names of the two columns, “t” the independent variable “time”, and “plasma” the name of the data.

The SAAM II **Data** window, however, has been designed for flexibility to accommodate the different kinds of experimental data that can be analyzed using SAAM II. The salient features will be discussed in the following.

SAAM II Data format

Two kinds of data can be entered in the SAAM II **Data** window. One is the experimental data, and the other is known “constants” such as weight, height, or dose.

Experimental data can be arranged in one or more tables of up to twenty columns each. Each table must begin with a DATA statement and end with an END statement. Each of these statements must be on a separate line.

After the DATA statement, you can enter the weighting scheme if it is to be used for all experimental data. If, for example, you are using a fractional standard deviation of 10%, you would write:

```
DATA  
(FSD 0.1)
```

The specification of the weighting scheme is enclosed in parentheses. This line is not required. If you enter no specification, the default is (FSD 0.1). If you have several columns of data, you can enter the weighting scheme associated with each column.

Just before the actual data are entered, you must include a header line that includes the name of the data element for each column of data. Below is an example of a data set (modified from **study_0**) with replicates for three plasma samples:

t	plasma1 (FSD 0.1)	plasma2 (FSD 0.1)	plasma3 (FSD 0.1)
0.007	46780	43593	44002
0.042	43522	45876	43166
0.125	42535	40725	40658
0.25	40125	39613	40978
0.375	36221	38968	39048
0.5	35562	36670	36661
1	28194	27516	28115
2	19573	19881	19136
3	14403	14775	14594
4	11278	10511	10299
5	8081	8123	8553
6	6999	7064	7229
7	5653	6014	5727
8	5139	5178	4851
9	4210	4212	4152
END			

As illustrated in the **Data** window above, after the DATA statement and just before the data, you must include a header line that includes the name of the data element for the independent variable and for each column of data, usually called a data set. In the above example, these names are “t”, “plasma1”, “plasma2” and “plasma3”.

You can see (FSD 0.1) is associated with each of the three columns of data. It is not necessary that “0.1” be used for each column. There may be occasions, e.g. experiments with plasma and urine data, when you may want to assign (FSD 0.1) to the plasma data and (FSD 0.25) to the urine data.

The independent variable (the default is "t" for time) must be in the first column in each table. This situation is illustrated in the previous figure. You can change the name of the independent variable in the **Experimental Attributes** dialog box which is opened in the **Compute** menu. Below is an example where the name was changed from “t” to “time”.

In all equations where the independent variable appears, “time” must now be used. If, for example, you wanted to write the expression for a mono-exponential, you would write:

$$y(\text{time}) = A * \exp(-a * \text{time})$$

If you had specified the name of the independent variable as “time” and written the above equation

$$y(t) = A * \exp(-a * t)$$

in the **Equations** window, “t” would appear as a parameter in the **Parameters** dialog box.

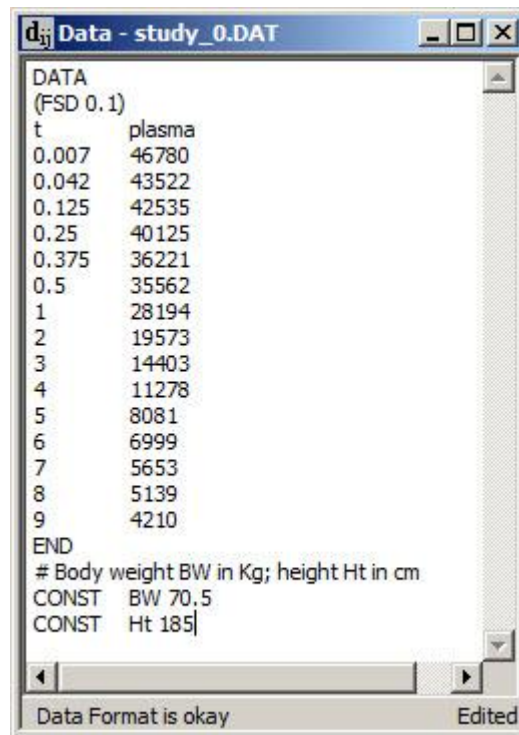
Columns must be separated by **one or more tabs** or spaces with no additional punctuation. Use “n” or a word beginning with “n” where **no value exists**. This is illustrated in the **Data** window above where there are no experimental data at time zero. You can add comments inside the **Data** window. A number sign (#) on a line causes the rest of the line to be ignored. Blank lines are also allowed.

All values for a given data set must be in the same table. For example, Table 1 might contain “t” and “plasma1”, while Table 2 contains “t”, “plasma2”, and “plasma3”. A column of “plasma1” values in the second table will be ignored.

The data shown in the **Data** window above does not need to be entered as a table with four columns. They can be entered as separate tables as shown below:

```
d:\ Data - study_0_3B.dat
DATA
t      plasma1 (FSD 0.1)
0.007  46780
0.042  43522
0.125  42535
0.25   40125
0.375  36221
0.5    35562
1      28194
2      19573
3      14403
4      11278
5      8081
6      6999
7      5653
8      5139
9      4210
END
DATA
t      plasma2 (FSD 0.1)
0.007  43593
0.042  45876
0.125  40725
0.25   39613
0.375  38968
0.5    36670
1      27516
2      19881
3      14775
4      10511
5      8123
6      7064
7      6014
8      5178
9      4212
END
DATA
t      plasma3 (FSD 0.1)
0.007  44002
0.042  43166
0.125  40658
0.25   40978
0.375  39048
0.5    36661
1      28115
2      19136
3      14594
4      10299
5      8553
6      7229
7      5727
8      4851
9      4152
END
Data Format is okay
```


Constants are entered using the CONST statement. CONST statements cannot lie between a DATA and END statement. A CONST statement must be used for each variable entered. Below is an example of two constants added to **study_0.dat**:



```
DATA
(FSD 0.1)
t      plasma
0.007  46780
0.042  43522
0.125  42535
0.25   40125
0.375  36221
0.5    35562
1      28194
2      19573
3      14403
4      11278
5      8081
6      6999
7      5653
8      5139
9      4210
END
# Body weight BW in Kg; height Ht in cm
CONST  BW 70.5
CONST  Ht 185
```

The screenshot shows a window titled "Data - study_0.DAT" with a scrollable text area containing the data shown in the code block above. At the bottom of the window, there are two status indicators: "Data Format is okay" and "Edited".

In addition to the two constants, a comment line denoted by “#” has been added.

As noted previously, the format for the **Data** window has been designed to be as flexible as possible in order to accommodate the many different kinds of experimental data that can be analyzed using SAAM II.

Conversion of data

A separate line may appear anywhere within a data table containing an operator and a numerical value. An example would be “(/60)” which indicates that each subsequent value in the column should be divided by “60” prior to use in a SAAM II computation. When a conversion line is used, each column must have an entry even if only empty brackets “()” (this signifies no conversion).

Acceptable conversion operators are “+”, “-”, “*”, and “/”. Functions cannot be used as conversion operators.

Data conversions are normally used to change the units associated with some time values or data themselves. For example, if the units of “t” (time) for the experiment are hours, but the data has been collected at time points expressed in minutes, you can easily convert minutes to hours by entering “(/60)” on top of the “t” data column.

The following **Data** window is an example where conversions of both time and data are used. This is an experiment in which radioiodinated low-density lipoprotein was injected. Plasma and 24 hour urine samples were collected, and the urine data presented as cumulative and 24 hour samples. Thus besides the column for “t”, there are three columns of data. The plasma data are in units of dpm (disintegrations per minute) per ml; the urine data are in units of dpm.

```

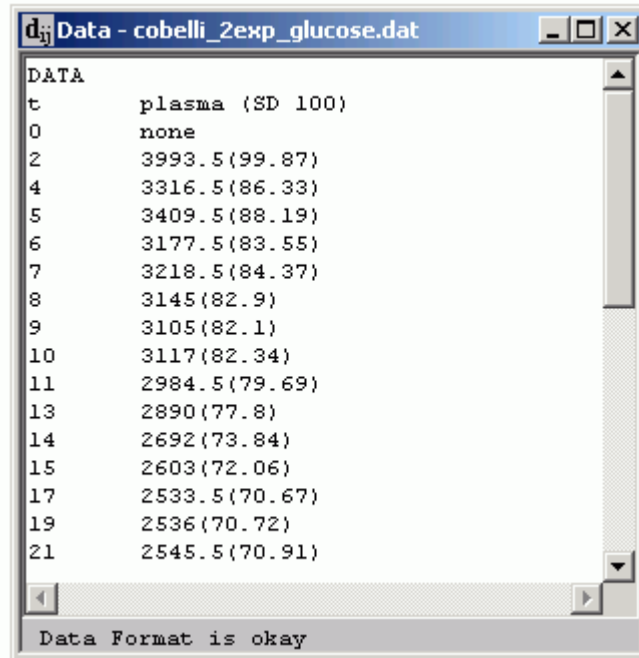
dij Data - LDL.dat
DATA
(FSD 0.1)
      t      plasma      cum_urine      ind_urine
#
# The (/24) converts the time units from hours to days.
#
      (/24)      ()      ()      ()
      0          n          n          n
      0.16      29745      n          n
      0.5       31008      n          n
      1         29114      n          n
      2         27907      n          n
      4         27593      n          n
      6         27483      n          n
      9         26610      n          n
      12        25791      n          n
#
# The "()" in all columns stops (/24) since the time units
# are now in days.
#
      ()      ()      ()      ()
      1.000    20899.    6979047.    6979047.
#
# The plasma data are in dpm/ml while the cumulative the urine
# samples are in dpm. The (*10000000) makes it easier to enter
# the data.
#
      ()      ()      (*10000000)      (*10000000)
      2.000    13646    2.424          1.671
      3.000    9714.    3.908          1.556
      4.000    6456.    5.090          1.185
#
# After a certain point, the (*10000000) is no longer needed to
# enter the individual urine samples.
#
      ()      ()      (*10000000)      ()
      5.000    4806.    5.779          8091944.
      6.000    3804.    6.752          5846717.
      7.000    3267.    6.942          3832041.
      8.000    2744.    7.401          2696780.
      9.000    2279.    7.508          1882786.
      10.000   2131.    7.698          1394947.
      11.000   1971.    7.141          985794.
      12.000   1720.    7.902          876272.
      13.000   1667.    7.626          712898.
      14.000   1622.    7.843          643487.
END
Data Format is okay

```

Comment lines have been added to the file to explain the intent of each conversion using “()”.

Assigning weights to data

The assignment of weights to each datum is done in the **Data** window. The same weighting scheme can be assigned to all data if the specification is made between the DATA statement and the entries of the data themselves. They can also be assigned to individual columns of data as was illustrated above with the columns “plasma1”, “plasma2” and “plasma3”. Finally, they can be assigned individually. Below is an example:



This is a portion of a **Data** window in which the plasma measurements are dpm of carbon atoms in glucose. The (SD 100) following “plasma” indicates that the weighting scheme is standard deviation. Associated with each datum, in parenthesis, is a value for the calculated standard deviation for that datum. This number will over ride the value “100” in (SD 100).

How are weights calculated, what is the relationship between the weights and the variance model (error in the data), and what are the options in SAAM II?

If $\sigma_{i,j}$ is the standard deviation of the i^{th} datum in the j^{th} data set, then the variance of the datum is defined

$$Var_{i,j} = \bar{v}_j \cdot \sigma_{i,j}^2 \quad (1)$$

where \bar{v}_j is a possibly unknown constant for the j^{th} data set. If \bar{v}_j is known, the weights are absolute; if it is unknown and estimated from the data, the weights are relative. When \bar{v}_j is estimated from the data, it is returned following a successful Fit in the **Statistics** window; it is called the **scaled data variance**.

A frequently used model for $\sigma_{i,j}^2$ is

$$\sigma_{i,j}^2 = A + B \cdot \left| \frac{\text{model}}{\text{data}} \right|^C \quad (2)$$

In this equation, A , B and C are non-negative numbers. The “model” or “data” in the absolute value sign means $\sigma_{i,j}^2$ can either be calculated based upon the model calculated value for the i^{th} - j^{th} datum, or based upon the datum itself. The former serves as the basis for “model” based weights and the latter for “data” based weights.

You specify absolute or relative weights and model or data based weights in the **Variance Model** pane of the **Computational Settings** dialog box. The **Variance Model** pane is shown below set in the default mode “data-relative”.



The weight assigned to each datum is the inverse of its variance:

$$w_{i,j} = \frac{1}{\text{Var}_{i,j}} = \frac{1}{\bar{v}_j \cdot \sigma_{i,j}^2} \quad (3)$$

The definition of $\sigma_{i,j}^2$ (2) in SAAM II is invoked using (GEN A B C) where, depending upon the variance model chosen,

$$(\text{GEN A B C}) = A + B \cdot \left| \frac{\text{model}}{\text{data}} \right|^C \quad (4)$$

For example, if the weights are data based and $A = 0$, $B = 1$ and $C = 2$, in the absolute weight situation when \bar{v}_j is equal to 1, the weight would equal

$$w_{i,j} = \frac{1}{\text{data}_{i,j}^2} \quad (5)$$

You can select **four** options for your weighting scheme: **FSD**, **POIS**, **SD** and **GEN**. **FSD**, **POIS** and **SD** are all special cases of (GEN A B C). **FSD** assigns a constant coefficient of variation as the variance to each datum; **POIS** assigns a Poisson distribution to each datum, and **SD** assigns a constant standard deviation and hence variance to each datum.

FSD

FSD is equivalent to (GEN 0 B 2). That is (4) becomes

$$(GEN\ 0\ B\ 2) = B \cdot \left| \frac{\text{model}}{\text{data}} \right|^2 \quad (6)$$

For the i-jth datum,

$$\sigma_{i,j} = \sqrt{\sigma_{i,j}^2} = \sqrt{B \cdot \left| \frac{\text{model}}{\text{data}} \right|^2} = \sqrt{B} \cdot \left| \frac{\text{model}}{\text{data}} \right| \quad (7)$$

Thus if (FSD \sqrt{B}) is entered as the weighting option in the data file:

$$\text{Data based absolute} \quad w_{i,j} = \frac{1}{B \cdot |\text{data}_{i,j}|^2} \quad (8a)$$

$$\text{Data based relative} \quad w_{i,j} = \frac{1}{\bar{v}_j \cdot B \cdot |\text{data}_{i,j}|^2} \quad (8b)$$

$$\text{Model based absolute} \quad w_{i,j} = \frac{1}{B \cdot |\text{model}_{i,j}|^2} \quad (8c)$$

$$\text{Model based relative} \quad w_{i,j} = \frac{1}{\bar{v}_j \cdot B \cdot |\text{model}_{i,j}|^2} \quad (8d)$$

In the above, $\text{data}_{i,j}$ is the ith datum in the jth data set; $\text{model}_{i,j}$ is the model calculated value at the time of $\text{data}_{i,j}$.

From (7), (FSD 0.1) will calculate a constant coefficient of variation (standard deviation) for each datum equal to 0.1, or 10%.

SD

SD is equivalent to (GEN A 0 0), that is (4) becomes

$$(GEN\ A\ 0\ 0) = A \quad (9)$$

For the i-jth datum,

$$\sigma_{i,j} = \sqrt{\sigma_{i,j}^2} = \sqrt{A} \quad (10)$$

Thus if (SD \sqrt{A}) is entered as the weighting option in the data file:

$$\text{Data based absolute} \quad w_{i,j} = \frac{1}{A} \quad (11a)$$

$$\text{Data based relative} \quad w_{i,j} = \frac{1}{\bar{v}_j \cdot A} \quad (11b)$$

$$\text{Model based absolute} \quad w_{i,j} = \frac{1}{A} \quad (11c)$$

$$\text{Model based relative} \quad w_{i,j} = \frac{1}{\bar{v}_j \cdot A} \quad (11d)$$

From (10), (SD 10) will assign a standard deviation of 10 to each datum. The weights, in the absolute case, will be $\frac{1}{10^2} = \frac{1}{100} = 0.01$.

POIS

FSD is equivalent to (GEN 0 B 1), that is (4) becomes

$$(GEN 0 B 1) = B \cdot \left| \frac{\text{model}}{\text{data}} \right| \quad (12)$$

For the i-jth datum,

$$\sigma_{i,j} = \sqrt{\sigma_{i,j}^2} = \sqrt{B \cdot \left| \frac{\text{model}}{\text{data}} \right|} \quad (13)$$

Thus if (POIS B) is entered as the weighting option in the data file:

$$\text{Data based absolute} \quad w_{i,j} = \frac{1}{\sqrt{B \cdot |\text{data}_{i,j}|}} \quad (14a)$$

$$\text{Data based relative} \quad w_{i,j} = \frac{1}{\sqrt{\bar{v}_j \cdot B \cdot |\text{data}_{i,j}|}} \quad (14b)$$

$$\text{Model based absolute} \quad w_{i,j} = \frac{1}{\sqrt{B \cdot |\text{model}_{i,j}|}} \quad (14c)$$

$$\text{Model based relative} \quad w_{i,j} = \frac{1}{\sqrt{\bar{v}_j \cdot B \cdot |\text{model}_{i,j}|}} \quad (14d)$$

In the above, $\text{data}_{i,j}$ is the i^{th} datum in the j^{th} data set; $\text{model}_{i,j}$ is the model calculated value at the time of $\text{data}_{i,j}$.

GEN

GEN is the general format for entering the variance, and hence weight, associated with each datum. The only restriction is that A , B and C must be non-negative. The formula is:

$$(\text{GEN } A \ B \ C) = A + B \cdot \left| \frac{\text{model}}{\text{data}} \right|^C \quad (15)$$

For the i - j^{th} datum,

$$\sigma_{i,j} = \sqrt{\sigma_{i,j}^2} = \sqrt{A + B \cdot \left| \frac{\text{model}}{\text{data}} \right|^C} \quad (16)$$

Thus if (GEN A B C) is entered as the weighting option in the data file:

$$\text{Data based absolute} \quad w_{i,j} = \frac{1}{A + B \cdot |\text{data}_{i,j}|^2} \quad (17a)$$

$$\text{Data based relative} \quad w_{i,j} = \frac{1}{\bar{v}_j \cdot (A + B \cdot |\text{data}_{i,j}|^2)} \quad (17b)$$

$$\text{Model based absolute} \quad w_{i,j} = \frac{1}{A + B \cdot |\text{model}_{i,j}|^2} \quad (17c)$$

$$\text{Model based relative} \quad w_{i,j} = \frac{1}{\bar{v}_j \cdot (A + B \cdot |\text{model}_{i,j}|^2)} \quad (17d)$$

In the above, $\text{data}_{i,j}$ is the i^{th} datum in the j^{th} data set; $\text{model}_{i,j}$ is the model calculated value at the time of $\text{data}_{i,j}$.

Weights in SAAM II

It is important to remember that, no matter what option you choose for the variance model and hence the weights, the variance is calculated using the formula

$$\text{Var}_{i,j} = \bar{v}_j \cdot (A + B \cdot \left| \frac{\text{model}}{\text{data}} \right|^2) \quad (18)$$

while, for a the i - j^{th} datum, the standard deviation is

$$\sigma_{i,j} = \sqrt{\sigma_{i,j}^2} = \sqrt{A + B \cdot \left| \frac{\text{model}}{\text{data}} \right|^2} \quad (19)$$

In SAAM II, the weight for the i - j^{th} datum is

$$w_{i,j} = \frac{1}{\text{Var}_{i,j}} = \frac{1}{\bar{v}_j \cdot \sigma_{i,j}^2} \quad (20)$$

so the weight equals the reciprocal of the square of the standard deviation only when \bar{v}_j equals 1.

Variance model – absolute versus relative

The variance is either known, \bar{v}_j equals 1, in which case the variance is the square of the standard deviation, or it is known (in terms of the model – FSD, SD, etc.) up to a proportionality constant, \bar{v}_j . If the variance model is “absolute”, then \bar{v}_j is assumed equal to 1. If the variance model is not known, \bar{v}_j is assumed unknown, and will be estimated from the data as part of the fitting process. It is called the scaled data variance in the **Statistics** window.

How can \bar{v}_j be used? It can be used as a measure of how good your estimate of the standard deviation, or variance, is. If, when you use relative weighting, the scaled data variance \bar{v}_j for the j^{th} data set is close to 1, then your estimate is very close since the variance will be close to

the square of the standard deviation. If \bar{v}_j is not close to 1, then you can modify your estimate for the standard deviation by setting:

$$\text{new_}\sigma_{i,j}^2 = \bar{v}_j \cdot \sigma_{i,j}^2 \quad (21)$$

or

$$\text{new_}\sigma_{i,j} = \sqrt{\bar{v}_j} \cdot \sigma_{i,j} \quad (22)$$

That is you can multiply your original estimate for the specification of the standard deviation by $\sqrt{\bar{v}_j}$ to produce a specification for a standard deviation that is closer to the true value (assuming the variance model “(GEN A B C)” is “correct”).

Variance model – data versus model based

In the calculation of the variance for the i-jth datum, either the actual datum or the model calculated value can be used. If in the variance model, the “data” option is chosen, then clearly from (18) the variance and hence the weight need to be calculated only once. When this option is chosen, fitting (optimization) in SAAM II is via weighted least square (WLS). On the other hand, if in the variance model, the “model” option is chosen, then again from (18) the variance and hence the weight must be recalculated at each iteration of the fitting (optimization process). When this option is chosen, fitting (optimization) in SAAM II is via extended weighted least square (ELS).

An example

Here is an example of a data table containing weighted values and conversions taken from the SAAM II Online Help system:

```
DATA
FSD 0.1)
t          smpl1 (SD 50)          smpl2
0.0              none          100.0
0.2              10000.0        95.2
1.0              9000.0 (100)    90.6 (0.05 +)
2.0              8000.0          84.1
3.0              7000.0          76.3
4.0              6000.0 (-)      68.8
6.0              5000.0          57.4 (-)
8.0              4000.0          50.0
10.0             3000.0          41.7
12.0             2500.0          34.1
END
# This is a comment
CONST concentration 0.34 # This is also a comment
DATA
(POIS 0.2)
t          smpl3
()          (*10)
0.0              none
0.3              11000.0
1.1              9200.0
()              ()
2.1              83000.0
3.2              74000.0
END
```

In this example:

1. At "t = 0.0", "smpl1" has no value and hence no standard deviation.
2. At "t = 0.2", "smpl1" has a value of "10000.0" with an "SD = 50". The "(SD 50)" in the header following "smpl1" overrides the global weight of "(FSD 0.1)" which appears after "DATA".
3. At "t = 1.0", "smpl1" has a value of "9000.0" with an "SD = 100". The "(100)" following the value overrides the "(SD 50)" set for the column.
4. At "t = 4.0", "smpl1" has a value of "6000.0", but the data point is ignored for fitting because "(-)" follows the value.
5. At "t = 0.0" and "t = 0.2", "smpl2" has an "FSD" of "0.1", due to the global assignment.
6. At "t = 1.0", "t = 2.0", "t = 3.0" and "t = 4.0", "smpl2" has an FSD of "0.05". The "(0.05 +)" assigns "0.05" to all values in the column until another entry appears.
7. At "t = 6.0", "smpl2" has a value of "57.4", but the data point is ignored for fitting.

-
8. From "t = 8.0" through "t = 12.0", "smp12" reverts to "FSD = 0.1", the global value, since the "(+)" specification at "t = 1.0" was canceled by the "(-)" at "t = 6.0".
 9. A constant named "concentration" is assigned a value of "0.34" at all data times.
 10. All values for "smp13" are assigned standard deviations using the Poisson distribution.
 11. All values of "smp13", before "t = 2.1", are multiplied by 10 prior to use. The remaining values are used as is. The parentheses "()" reset the conversion operators to null.

With this introduction to the various options for the data format in SAAM II, and the discussion of how to define the variance model for the error in your data, and hence the weight of each datum, you can proceed to those parts of the tutorial in which you are interested.

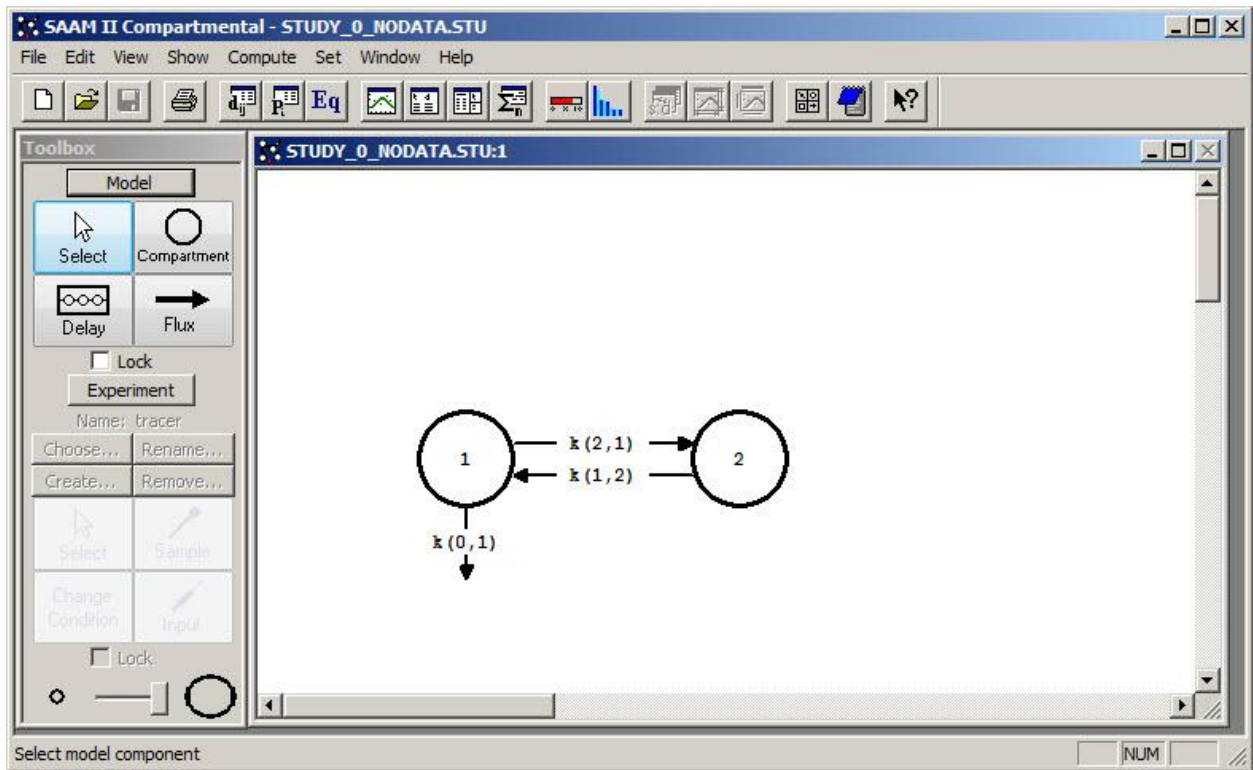
Part 1. Importing data from a spreadsheet

It is often convenient to prepare your data for a SAAM II modeling exercise first in a spreadsheet, and then import the data into the **Data** window. This tutorial will illustrate this process using an Excel® spreadsheet.

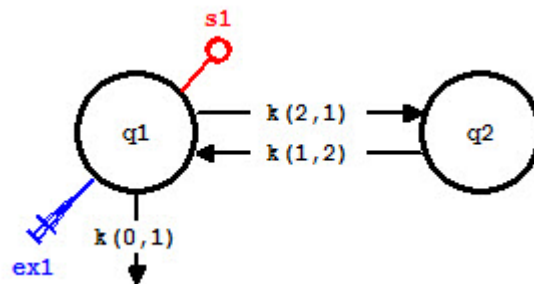
1. **Start** Excel® and open **study_0.xls**. You will have to move to the folder where you put this file. The spreadsheet will appear in part as follows:

	A	B
1	Example DATA	
2		
3	(FSD 0.1)	
4	t	plasma
5	0.007	46780
6	0.042	43522
7	0.125	42535
8	0.25	40125
9	0.375	36221
10	0.5	35562
11	1	28194
12	2	19573
13	3	14403
14	4	11278
15	5	8081
16	6	6999
17	7	5653
18	8	5139
19	9	4210
20		
21	END	


2. **Start** the **SAAM II Compartmental** application. The **SAAM II Compartmental** main window will open.
3. Open the **SAAM II Compartmental** study file **study_0_nodata**. This study file contains all information present in **study_0** except the **Data** window is empty.
 - a. The file **study_0_nodata.stu** should appear in the file list; if it does not, find the folder where you put this tutorial.
 - b. In the **File** menu, click **Open**. The **SAAM II Compartmental** main window will appear as shown below:

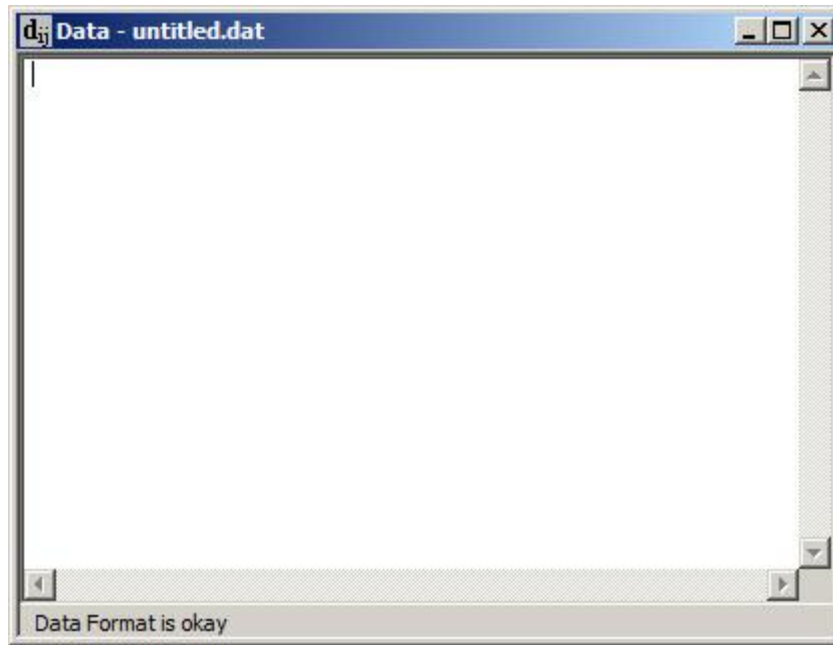


4. View the model and the experiment on the model. In the **SAAM II Toolbox**, click **Experiment**. The model of the experiment will appear on the **Drawing Canvas** as follows:



Remember the sample circle, **s1**, is not filled because there are no data in the **Data** window to associate with the sample. The experimental input **ex1** and the parameter values are all those from **study_0**. You can verify this by examining the **Exogenous Input** or **Parameters** dialog boxes.

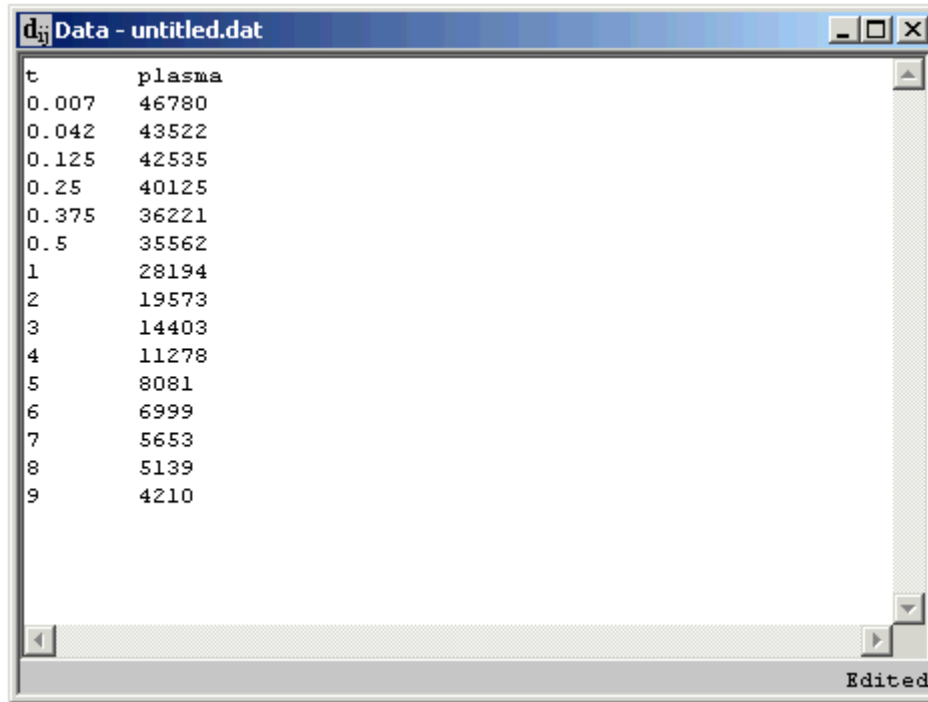
5. Import the data from **study_0.xls**.
 - a. In the **Show** menu, click **Data**, or alternatively, on the **SAAM II Toolbar** click **Data** . The **Data** window will open as follows:



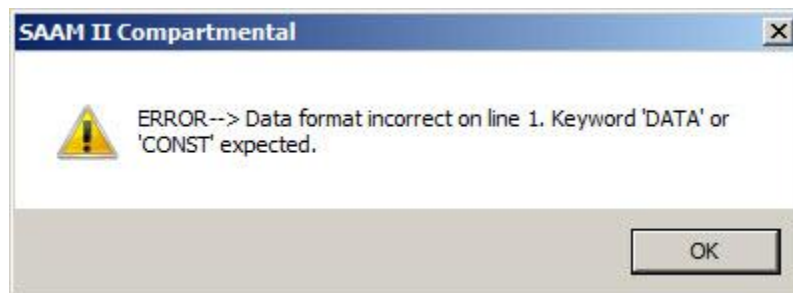
- b. Return to the Excel® spreadsheet.
- c. Select the data in the spreadsheet you wish to import into the SAAM II **Data** window. The spreadsheet will appear as follows:

	A	B
1	t	plasma
2	0.007	46780
3	0.042	43522
4	0.125	42535
5	0.25	40125
6	0.375	36221
7	0.5	35562
8	1	28194
9	2	19573
10	3	14403
11	4	11278
12	5	8081
13	6	6999
14	7	5653
15	8	5139
16	9	4210

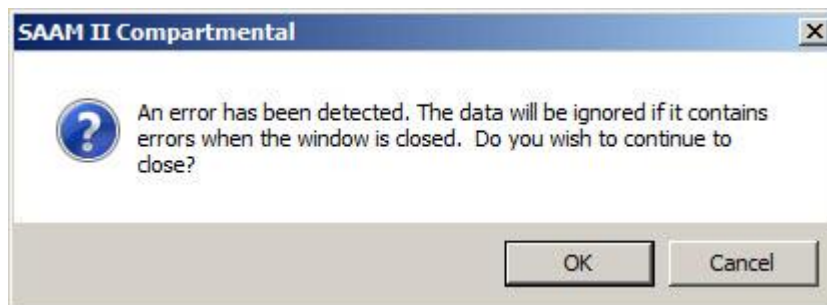
- d. Copy the data from the spreadsheet.
- e. Return to the SAAM II **Data** window.
- f. In the **Edit** menu, click **Paste**. The **Data** window will appear as follows:



g. Close the **Data** window. An **Error** message will appear as follows:



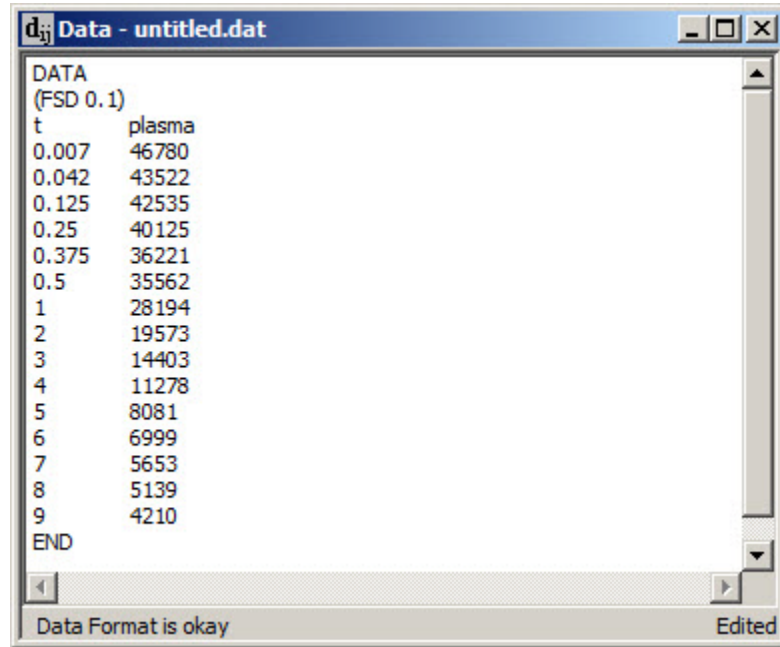
Click **OK**. The following message will appear:



The reason why these error messages appear is that the first line of the data file in the **DATA** window must be either **DATA** (or **CONST**), and if you are entering a stream of experimental data, the last line must be **END**.

h. Click **Cancel**. You will return to the **Data** window.

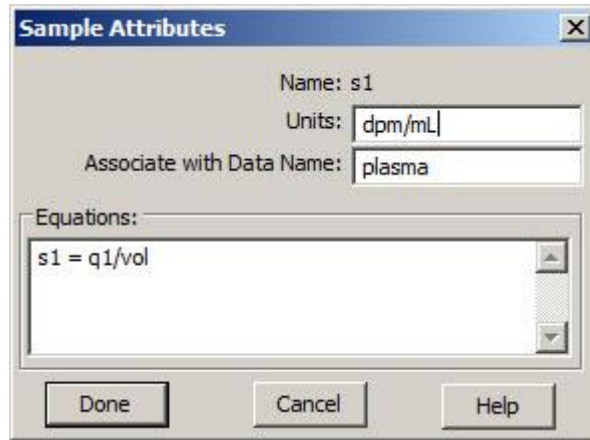
-
- i. Type “DATA” for the first line, and (FSD 0.1) for the second line. Type “END” at the end of the data stream.
 - j. In the **Edit** menu, click **Check Data Format**. Your **Data** window should appear:



If there are any errors in the **Data** window, they will be reported and you can fix them. When there are no errors, the statement “Data Format is okay” will appear as shown at the lower left of the **Data** window.

- k. Close the **Data** window.
- l. Associate the sample **s1** with the data.
 - (1) Double-click **s1** to open the **Sample Attributes** dialog box.
 - (2) Type “plasma” in the **Associate with Data Name** box.

The **Sample Attributes** dialog box will appear as follows:



(3) Close the **Sample Attributes** dialog box. You can now proceed with the modeling exercise.

Quit the **SAAM II Compartmental** application. Do not save the changes to **study_0_nodata**.

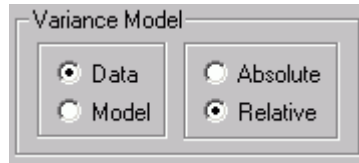


Importing data from a spreadsheet. In importing data from a spreadsheet, you can also import segments of data. This can be used, for example, to replace parts of the data in a **Data** window. When you do this, you may accidentally introduce an error. For this reason, it is always wise to check the data format using the **Check Data Format** option in the **Edit** menu. If there are errors, **SAAM II** will indicate the general nature of the first error and the line on which it occurs. You should check the data format after correcting each error until you get the “**Data Format is okay**” message.



Part 2. Effect of the variance model when using the FSD weighting option

In this tutorial, you will use **study_0** to investigate the effect of the variance model when using the FSD weighting option. The variance model is set in the **Variance Model** pane in the **Computational Settings** dialog box. The pane is shown below:

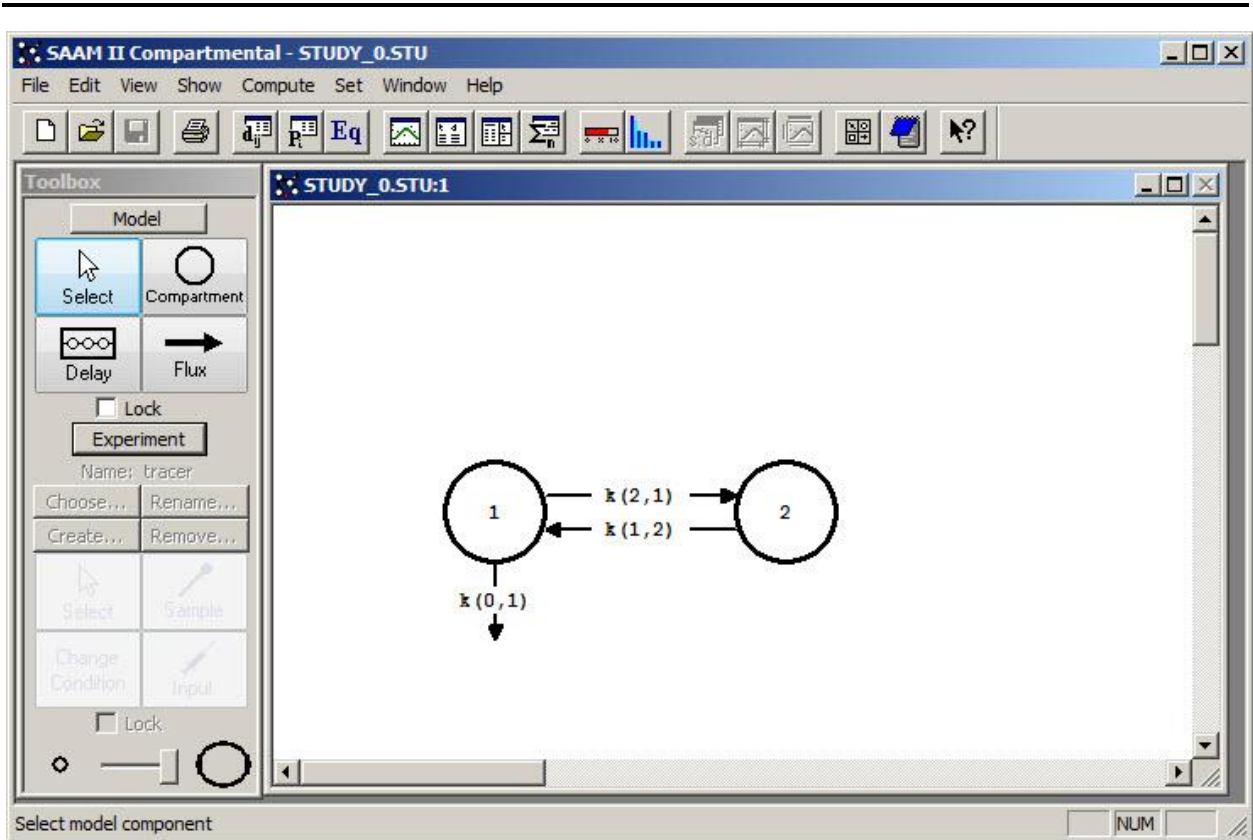


As discussed in the Introduction, the variance model can be based upon the data or model calculation, and can be absolute or relative. All options will be investigated with different values for the FSD. This will give you a much better feeling for the variance model, and what purpose the scaled data variance parameter $\bar{\nu}_j$ discussed in the Introduction serves.

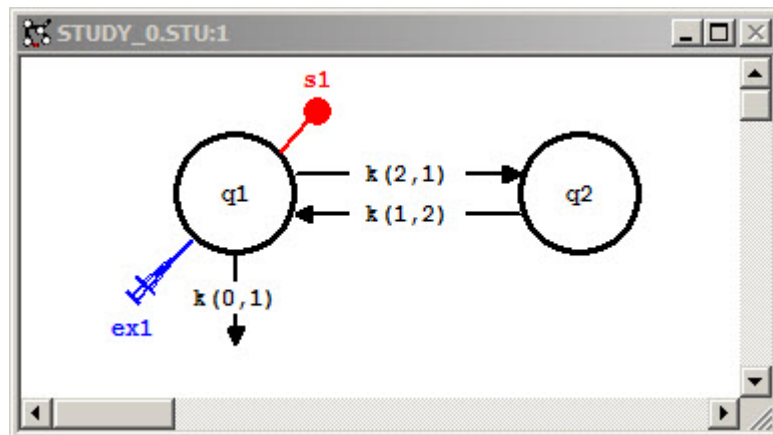
This tutorial will use the **Solutions** option in the **Compute** menu to create a baseline solution from which you will work. If you are not familiar with this option, it is recommended you work through the Using SAAM II Tutorial “Saving and Restoring Solutions”.

Since there is just one data set in this tutorial, we will refer to the scaled data variance parameter $\bar{\nu}_j$ simply as ν .

1. **Start** the **SAAM II Compartmental** application. The **SAAM II Compartmental** main window will open.
2. Open the **SAAM II Compartmental** study file **study**.
 - a. The file **study_0.stu** should appear in the file list; if it does not, find the folder where you put this tutorial.
 - b. In the **File** menu, click **Open**. The **SAAM II Compartmental** main window will appear as shown below:



3. View the model and the experiment on the model. In the **SAAM II Toolbox**, click **Experiment**. The model of the experiment will appear on the **Drawing Canvas** as follows:



To understand the effect of setting different values for FSD, and using different variance models, you will Fit **study_0** using the following:



FSD = 0.1	absolute data base absolute model base	relative data base relative model base
FSD = 0.25	absolute data base absolute model base	relative data base relative model base

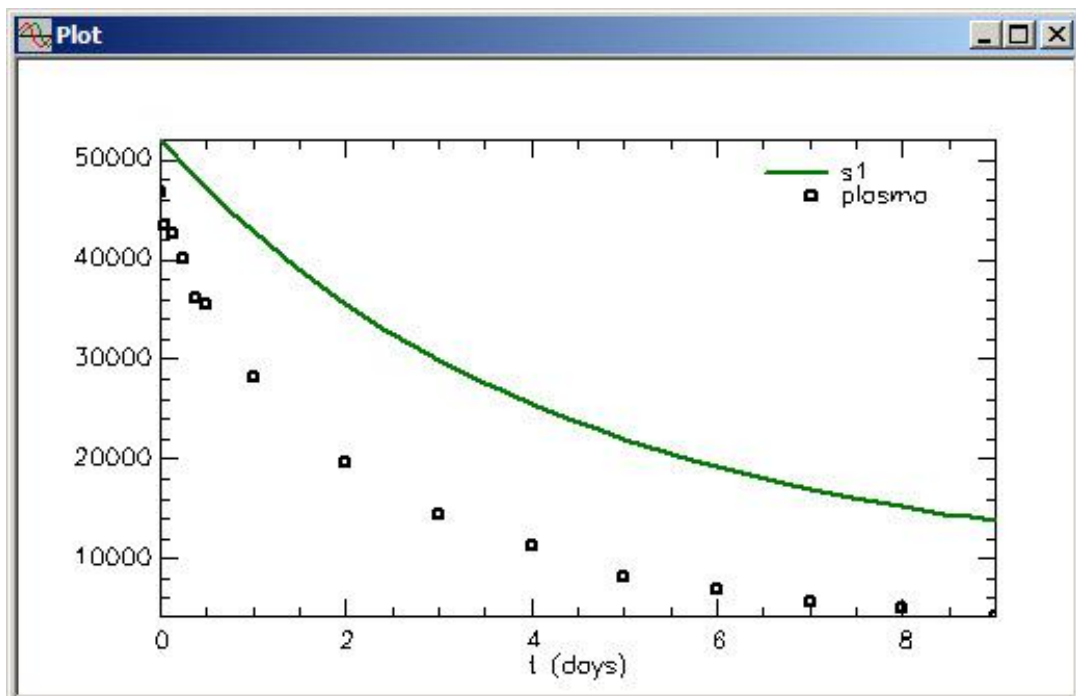
FSD = 0.5	absolute data base absolute model base	relative data base relative model base
FSD = 1.0	absolute data base absolute model base	relative data base relative model base

A worksheet is provided at the end of the tutorial for you to record your results as you proceed. Remember the default variance model in SAAM II is relative data base, and the weighting scheme in **study_0** is (FSD 0.1).

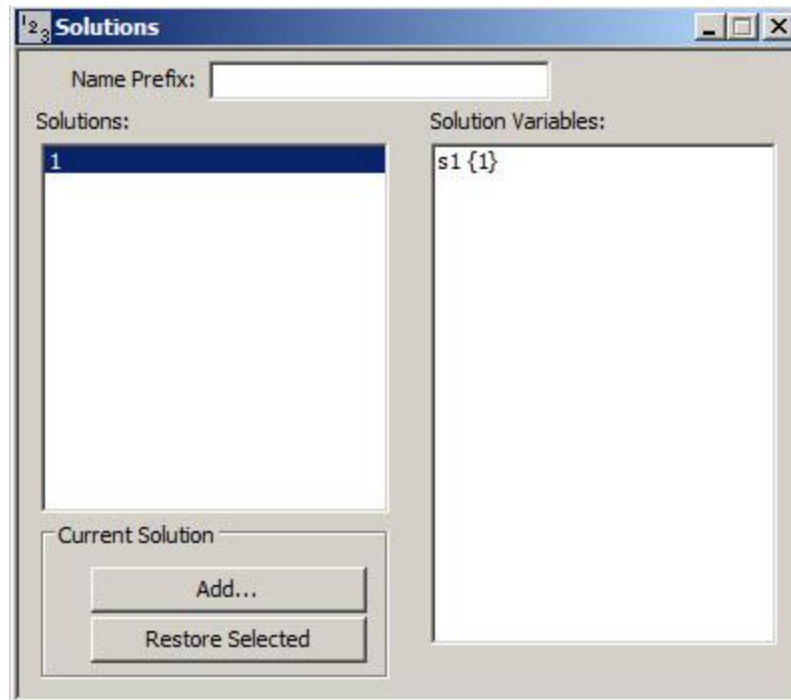
4. Create a baseline solution.

It is important in this series of Fits to always start from the same set of initial parameter estimates.

- a. On the **Compute** menu, click **Solve**, or alternatively click **Solve**  on the **SAAM II Toolbar**.
- b. Plot the data. On the **Show** menu, click **Plot**, or alternatively click **Plot**  on the **SAAM II Toolbar**.
- c. On the **Set** menu, click **Plot/Table Variables**. The **Plot and Table Variables** dialog box will open. Be sure the **List All Variables** box is not checked.
- d. Click **s1:plasma** to move this to the **Current Selection** window.
- e. Click **Done**. Your plot (in semi-log) will appear:



- f. Close the **Plot** window.
- g. On the **Compute** menu, click **Solutions**. The **Solution** dialog box will open.
- h. In the **Current Solution** box, click **Add**. The **Solutions Variables** box will open with a list of variables you can save.
- i. Click **s1**.
- j. Click **Done**.
- k. In the **Solutions** box, click **1**. The list of saved variables and parameters (**s1** only in this case) will appear in the **Solutions Variables** box. The **Solutions** dialog box will appear:



The {1} in **s1{1}** refers to solution number **1**; **s1{1}** can be accessed anywhere other model variables can be accessed. An example is the **Plot and Table Variables** dialog box when the **List All Variables** box is checked.

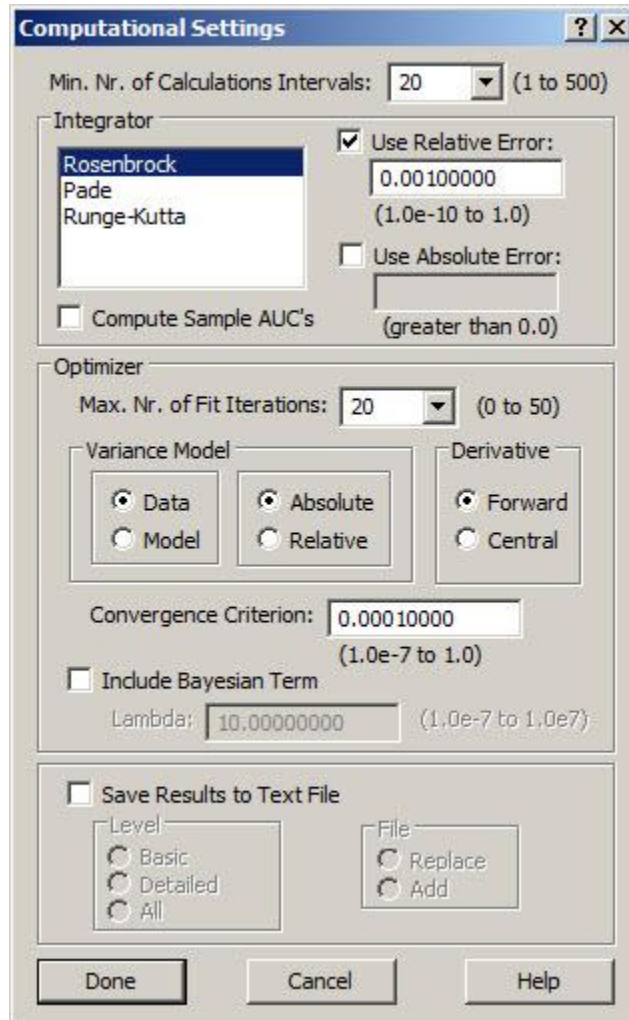
- l. Close the **Solutions** dialog box.


This solution will be your baseline solution to which you will return before resetting the variance model and Re-Fitting the model to the data.


5. Perform the first fit.

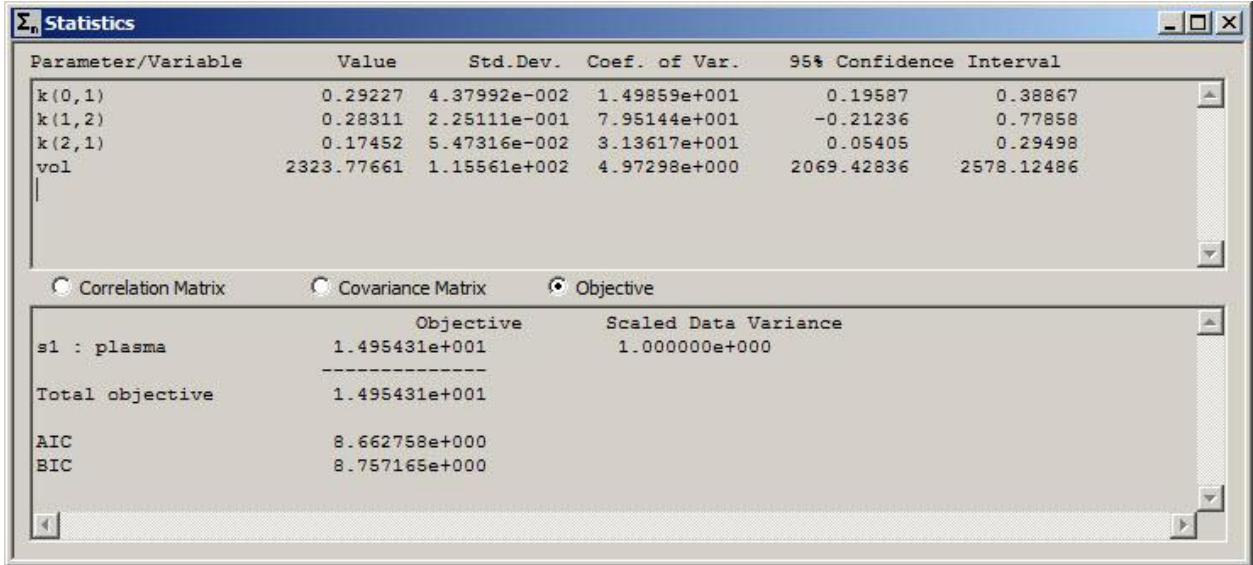
The variance model in the first Fit is $FSD = 0.1$ and absolute data. Thus the variance model must be changed from the default setting.

- a. In the **Compute** menu, click **Settings**. The **Computational Settings** dialog box will appear.
- b. In the **Variance Model** pane, change the variance model from “relative” to “absolute”. The **Computational Settings** dialog box will appear as follows:



- c. Click **Done**.
- d. Fit your model to your data. On the **Compute** menu, click **Fit**, or alternatively click **Fit**  on the **SAAM II Toolbar**.

- e. Plot the solution. You can see the fit is quite satisfactory. Leave your **Plot** window open.
- f. On the **Show** menu, click **Statistics**, or alternatively, click **Statistics**  on the **SAAM II Toolbar**. The **Statistics** window will appear as follows:



- g. Record in your worksheet the best estimate and coefficient of variation for the four parameters of your model. The value for v is 1 since the weighting scheme is absolute and has been entered. If the weighting scheme were relative, this value would not be equal to 1, and would need to be entered. This part of your worksheet should appear:

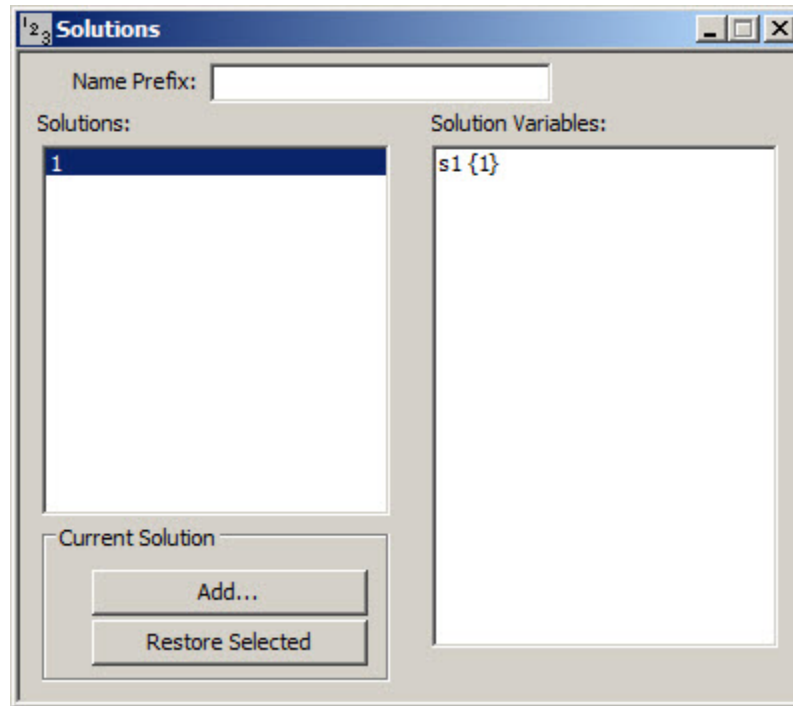
		Data Based Absolute
FSD 0.1	k(0,1)	.292(15)
	k(2,1)	.175(31)
	k(1,2)	.283(80)
	vol	2324(5)
	v	1

- h. Close the **Statistics** window.
6. Perform the second Fit.

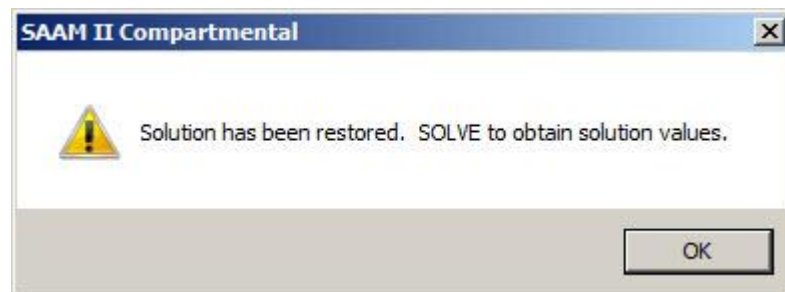
The second Fit is data base relative. Thus you need to return to the original model and be sure the variance model is correctly set.

- a. In the **Compute** menu, click **Solutions**. The **Solutions** dialog box will open.

- b. In the **Solutions** pane, click **1**. Notice the **Restore Selected** button is now active in the **Current Solution** pane as shown below.

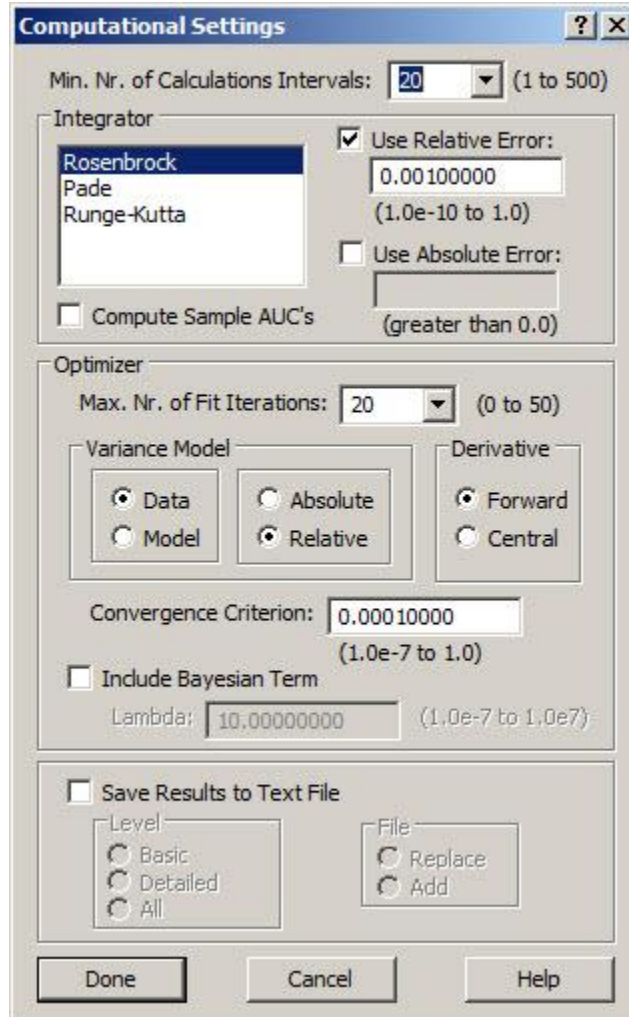


- c. Click **Restore Selected**. The following message will appear:



This means that before you proceed, you must Solve the model again to return to the original **study_0** which is the solution you saved.

- d. Click OK.
- e. Close the **Solutions** dialog box.
- f. Re-Solve the model. You can verify you have returned to the original model by viewing the plot, and checking the parameters and computational settings.
- g. Open the **Computational Settings** dialog box to reset the variance model to data relative. The **Computational Settings** dialog box should appear as follows:





Click **Done**.



Restoring solutions. When a saved solution is restored, only the parameter values associated with the solution and input information are retained. If the computational settings have been change, they will remain at their changed values or settings.



- h. Fit your model to your data. On the **Compute** menu, click **Fit**, or alternatively click **Fit**  on the **SAAM II Toolbar**.
- i. Plot the solution. You can see the fit is quite satisfactory. Leave your **Plot** window open.

- j. On the **Show** menu, click **Statistics**, or alternatively, click **Statistics**  on the **SAAM II Toolbar**. The **Statistics** window open.
- k. Record in your worksheet the best estimate and coefficient of variation for the four parameters of your model. The value for ν of has not been entered since the weighting scheme is relative. Your worksheet should appear as follows:

		Data Based Absolute	Data Based Relative
FSD 0.1	k(0,1)	.292(15)	.292(5)
	k(2,1)	.175(31)	.175(10)
	k(1,2)	.283(80)	.283(27)
	vol	2324(5)	2324(2)
	ν	1	1.11E-01

Notice that the parameter estimates do not change, but the error estimates of the parameters do change.


- l. Close the **Statistics** window.
6. Perform the third and fourth Fits.

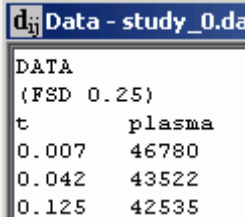
Repeat the above. After you have restored the baseline solution, you will change the variance model first to model-absolute, and then model-relative. When you have completed these Fits, your worksheet should appear as follows:

		Data Based Absolute	Data Based Relative	Model Based Absolute	Model Based Relative
FSD 0.1	k(0,1)	.292(15)	.292(5)	.294(14)	.294(5)
	k(2,1)	.175(31)	.175(10)	.176(32)	.176(11)
	k(1,2)	.283(80)	.283(27)	.295(77)	.294(26)
	vol	2324(5)	2324(2)	2341(5)	2320(2)
	ν	1	1.11E-01	1	1.12E-01

You can see the major differences in the parameter coefficients of variation between absolute and relative weights; there are only small differences in the parameter estimates between data and model based weights. More specifically, between Absolute and Relative, there are minimal differences in parameter values except for volume in the Model-Based variance models. As expected, there are significant differences in coefficients of variation. Between Data and Model, there are more noticeable differences in parameter values but the coefficients of variations are very close. There is minimal difference in Scaled Data Variance between Data and Model based variance models.

As you proceed, you should plot the best fit of the model to the data; you will see (and this is supported by the very small changes in the actual parameter values) that the Fits are essentially identical.

7. Repeat the above with $FSD = 0.25$
 - a. In the **Show** menu, click **Data**, or alternatively, on the **SAAM II Toolbar** click **Data** . The **Data** window will open.
 - b. Change the FSD from 0.1 to 0.25. The initial portion of the **Data** window will appear as follows:



DATA	
(FSD 0.25)	
t	plasma
0.007	46780
0.042	43522
0.125	42535

- c. Close the **Data** window.
 - d. Restore the baseline solution.
 - e. Fit the model to the data with data-absolute, and record the results.
 - f. Repeat with data-relative, model-absolute and model-relative.
8. Repeat Step 7 with $FSD = 0.5$.
9. Repeat Step 7 with $FSD = 1.0$

When you have finished, your spreadsheet should appear as follows:

		Data Based Absolute	Data Based Relative	Model Based Absolute	Model Based Relative
FSD 0.1	k(0,1)	.292(15)	.292(5)	.294(14)	.294(5)
	k(2,1)	.175(31)	.175(10)	.176(32)	.176(11)
	k(1,2)	.283(80)	.283(27)	.295(77)	.294(26)
	vol	2324(5)	2324(2)	2341(5)	2320(2)
	v	1	1.11E-01	1	1.12E-01
FSD 0.25	k(0,1)	.292(37)	.292(5)	.294(33)	.294(5)
	k(2,1)	.175(78)	.175(10)	.176(77)	.176(11)
	k(1,2)	.283(199)	.283(27)	.295(183)	.294(26)
	vol	2324(12)	2324(2)	2454(12)	2320(2)
	v	1	1.78E-02	1	1.79E-02
FSD 0.5	k(0,1)	.292(75)	.292(5)	.294(58)	.294(5)
	k(2,1)	.175(157)	.175(10)	.176(133)	.176(11)
	k(1,2)	.283(398)	.283(27)	.294(318)	.294(26)
	vol	2324(25)	2324(2)	2798(20)	2320(2)
	v	1	4.45E-03	1	4.48E-03
FSD 1	k(0,1)	.292(150)	.292(5)	.294(82)	.294(5)
	k(2,1)	.175(314)	.175(10)	.176(188)	.176(11)
	k(1,2)	.283(795)	.283(27)	.294(449)	.294(26)
	vol	2324(50)	2424(2)	3749(29)	2320(2)
	v	1	1.11E-03	1	1.12E-03

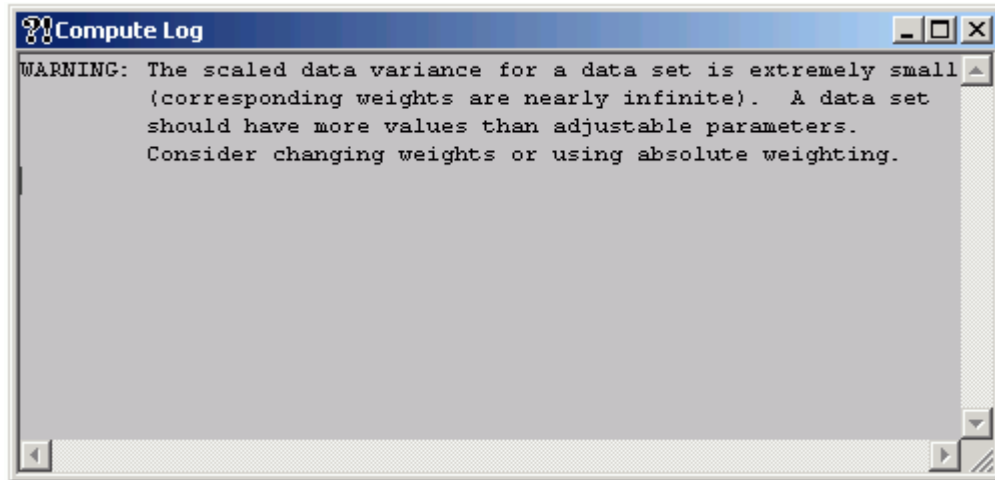
There may be some small differences; these probably will be due to the computer you are using, and the computational algorithms supported by the computer

In going through this tutorial, you have seen that:

- How you assign the variance model can have a dramatic effect on the statistical precision of the parameter estimates.
- When absolute weights are used, in both the data and model based variance models, the higher the FSD, the poorer the precision of the parameter estimates.
- When relative weights are used, in both the data and model based variance models, the parameter estimates and their precision are the same (there are small differences in the parameter estimates between data and model based weights). In this case, as the FSD increases, the scaled data variance ν decreases (indicating you are moving away from the “best” value for the FSD).
- When using the model-absolute variance model, as the FSD increases, the estimated volume increases resulting in a very poor fit.

In general, it is recommended to use relative weighting. SAAM II uses data-relative as the default because you may not have the “correct” model when you begin your model development exercise. Model-relative assumes that the model is “correct”. Once you have developed what you believe is the “correct” model, you can switch to the model-relative variance model.

Relative as opposed to absolute is also recommended since it is rare you will know exactly the value to use in the weight specification. There are times, however, when the relative option may not be appropriate. You may try to fit your model to your data, and the following warning message may appear:



In this instance, if you believe your model for your weights is correct (e.g. FSD is appropriate), then you should change the variance model to absolute, and try again.

Quit the **SAAM II Compartmental** application. Do not save the changes to **study_0**.



The variance model. You have seen in this tutorial that, when you use absolute weights, there is a large discrepancy in the precision estimates for the parameters depending upon the numerical value you use for FSD. In the above, for example, with $FSD = 0.1$, the scaled data variance is 0.111. From equation (22) in the introduction, you can “correct” this to obtain a value which, when you use data-relative, the scaled data variance will be close to 1. The value is $0.1 \cdot \sqrt{0.111} \approx 0.033$. If you change “0.1” to “0.033” in the **Data** window, and Re-Fit the model to the data, the **Statistics** window will appear as follows:

Parameter/Variable	Value	Std.Dev.	Coef. of Var.	95% Confidence Interval	
k(0,1)	0.29227	1.46155e-002	5.00071e+000	0.26010	0.32444
k(1,2)	0.28311	7.51183e-002	2.65335e+001	0.11777	0.44844
k(2,1)	0.17452	1.82636e-002	1.04652e+001	0.13432	0.21472
vol	2323.77661	3.85619e+001	1.65945e+000	2238.90219	2408.65103

	Objective	Scaled Data Variance
s1 : plasma	1.336743e+001	1.022512e+000
Total objective	1.336743e+001	
AIC	7.935987e+000	
BIC	8.053995e+000	

The statistics are good, and the scaled data variance is essentially 1.

Suppose you wanted “better” statistics in terms of parameter precision. You could use data-absolute, and set FSD = 0.01. Re-Fitting the model to the data, the **Statistics** window will appear as follows:

Parameter/Variable	Value	Std.Dev.	Coef. of Var.	95% Confidence Interval	
k(0,1)	0.29227	4.37992e-003	1.49859e+000	0.28263	0.30191
k(1,2)	0.28311	2.25111e-002	7.95144e+000	0.23356	0.33265
k(2,1)	0.17452	5.47316e-003	3.13617e+000	0.16247	0.18656
vol	2323.77661	1.15561e+001	4.97298e-001	2298.34178	2349.21144

	Objective	Scaled Data Variance
s1 : plasma	1.843326e+001	1.000000e+000
Total objective	1.843326e+001	
AIC	1.040223e+001	
BIC	1.049664e+001	

The parameter precision is better with FSD = 0.01 than FSD = 0.033. This stands to reason since it is assuming a 1% error to the data as opposed to a 3.3% error.

The point is that you can manipulate the results when you use the absolute option in the variance model. This is not wise, which is why the relative option is highly recommended.



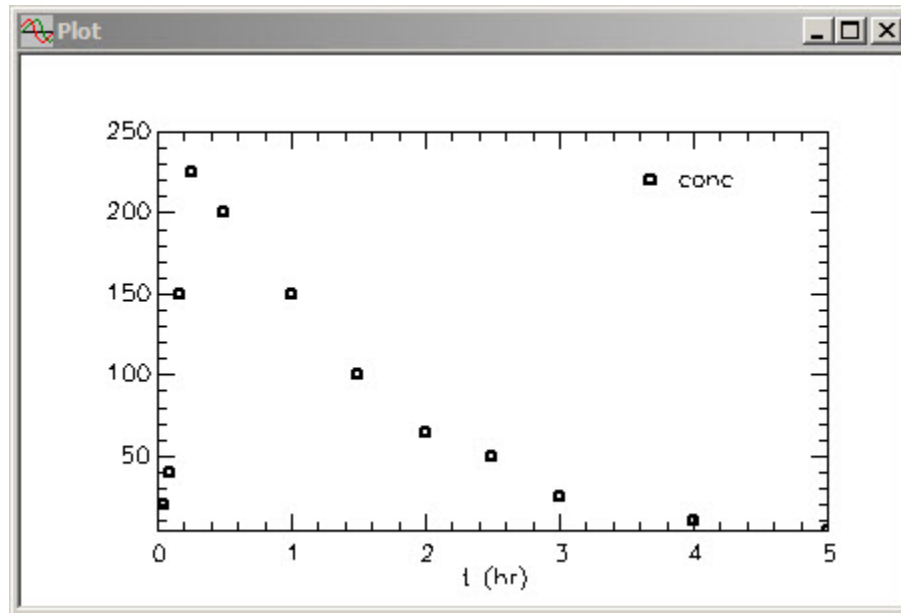
Worksheet for tutorial

You can use this worksheet to record the results of your fitting exercise:

		Data Based Absolute	Data Based Relative	Model Based Absolute	Model Based Relative
FSD 0.1	k(0,1)				
	k(2,1)				
	k(1,2)				
	vol				
	v	1		1	
FSD 0.25	k(0,1)				
	k(2,1)				
	k(1,2)				
	vol				
	v	1		1	
FSD 0.5	k(0,1)				
	k(2,1)				
	k(1,2)				
	vol				
	v	1		1	
FSD 1	k(0,1)				
	k(2,1)				
	k(1,2)				
	vol				
	v	1		1	

Part 3. Effect of using the FSD or SD weighting option

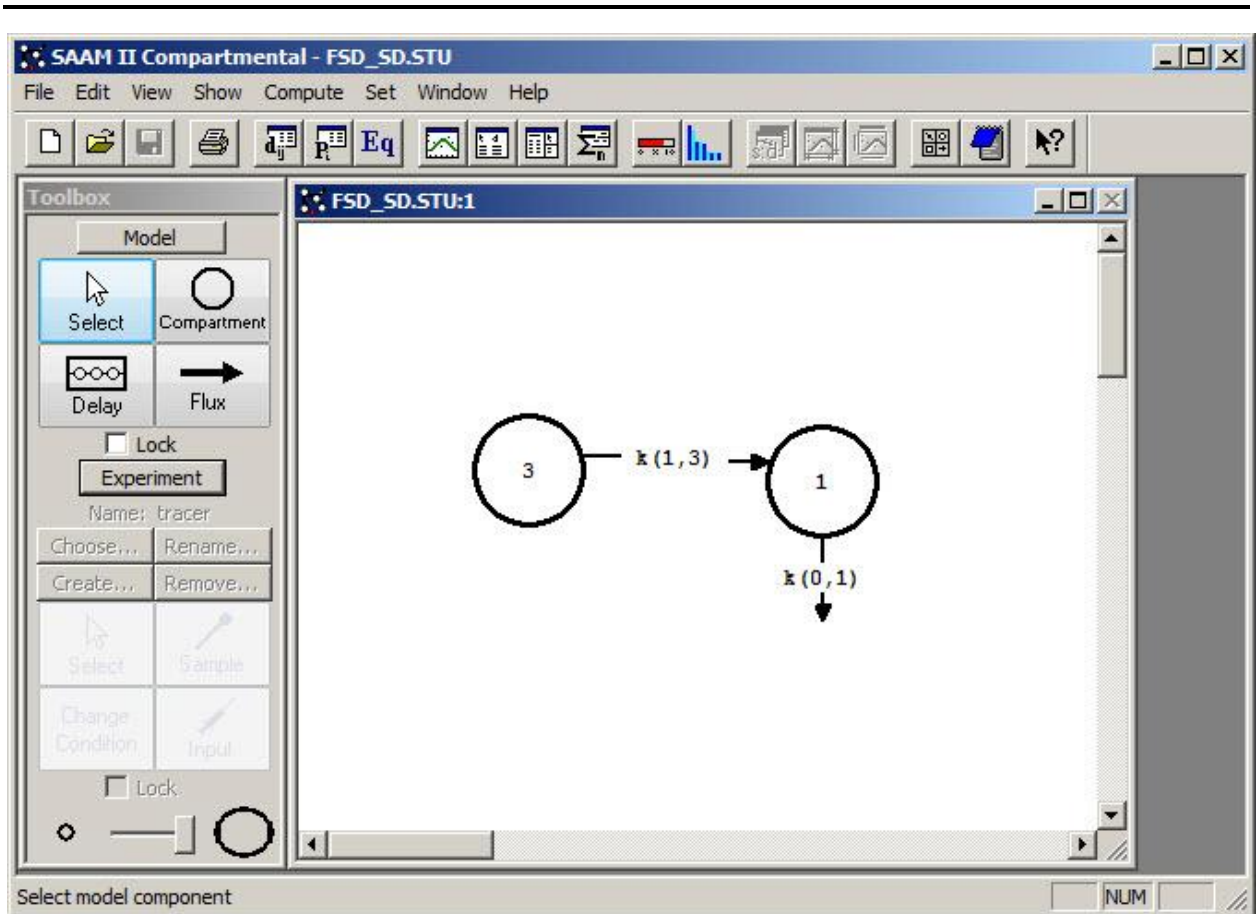
In this tutorial, you will use the study file **fsd_sd.stu** to investigate the effect of using the FSD and SD weighting options in your modeling exercise. The data to be analyzed are shown in the figure below:



These are plasma data (called “conc” for concentration) following an oral input of 23000 μ g of a drug, and plasma samples were taken for 5 hours. These data are characteristic of an experiment where the data rise and fall, and the peak is determined by a very few samples. This tutorial will illustrate the difference between using FSD and SD as the weighting option to analyze these data, and explain why there is a difference.

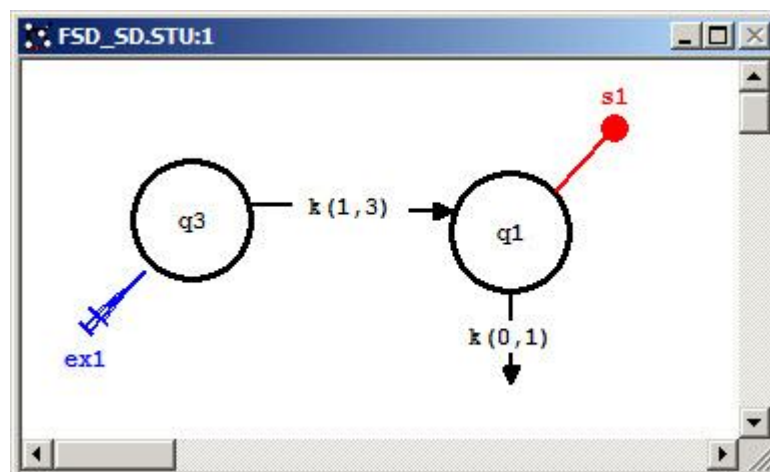
This tutorial will use the **Solutions** option in the **Compute** menu to create a baseline solution from which you will work. If you are not familiar with this option, it is recommended you work through the Using SAAM II Tutorial “Saving and Restoring Solutions”.

1. **Start** the **SAAM II Compartmental** application. The **SAAM II Compartmental** main window will open.
2. Open the **SAAM II Compartmental** study file **study**.
 - a. The file **fsd_sd.stu** should appear in the file list; if it does not, find the folder where you put this tutorial.
 - b. In the **File** menu, click **Open**. The **SAAM II Compartmental** main window will appear as shown below:



In the pharmacokinetic literature, this is known as the one-compartment model with absorption even though there are really two compartments in the model.



3. View the model and the experiment on the model. In the **SAAM II Toolbox**, click **Experiment**. The model of the experiment will appear on the **Drawing Canvas** as follows:

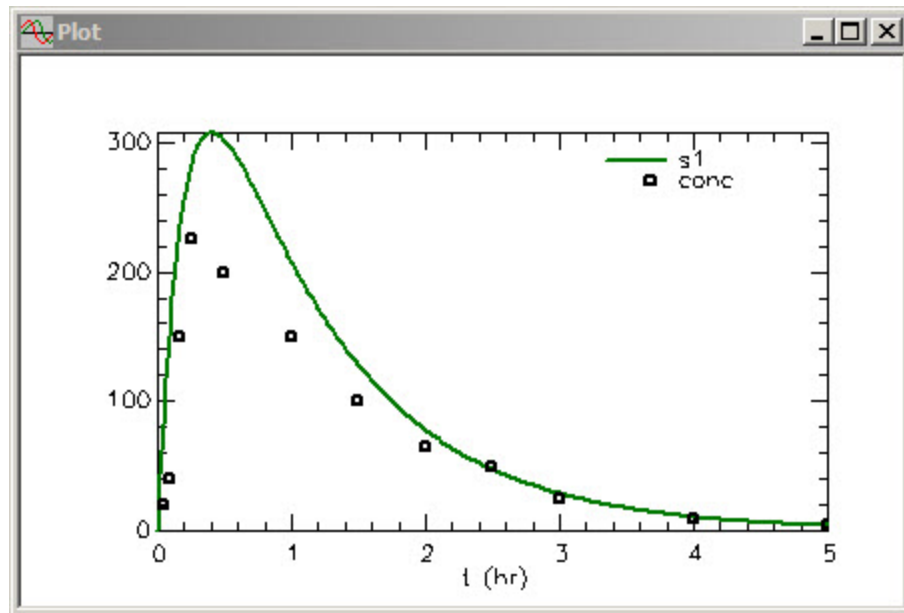


You can view information about the model and the experiment on the model by viewing the **Exogenous Input**, the **Parameters**, and the **Sample Attributes** dialog boxes.

4. Create a baseline solution.

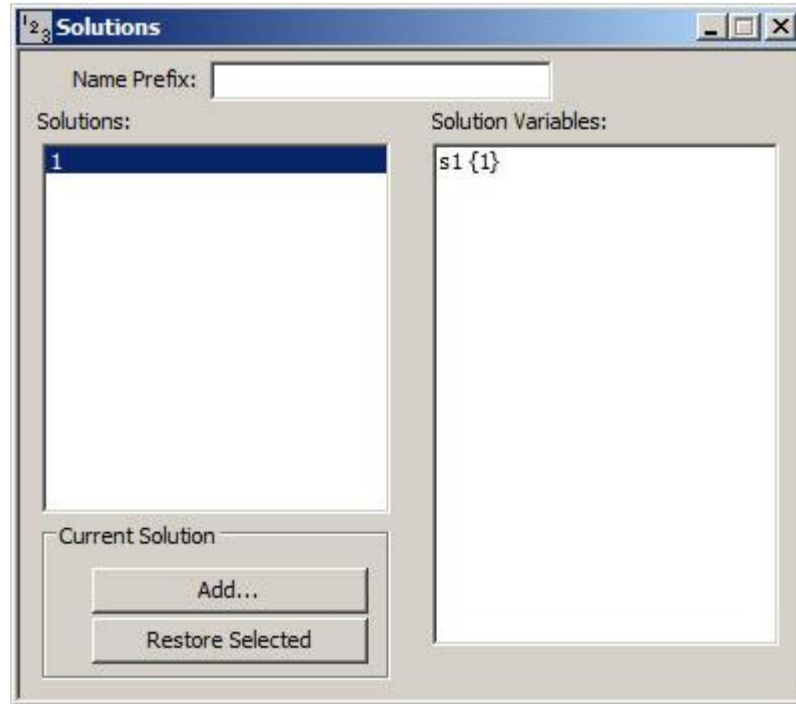
It is important in the Fits you will do to always start from the same set of initial parameter values.

- a. On the **Compute** menu, click **Solve**, or alternatively click **Solve**  on the **SAAM II Toolbar**.
- b. Plot the data. On the **Show** menu, click **Plot**, or alternatively click **Plot**  on the **SAAM II Toolbar**.
- c. On the **Set** menu, click **Plot/Table Variables**. The **Plot and Table Variables** dialog box will open. Be sure the **List All Variables** box is not checked.
- d. Click **s1:conc** to move this to the **Current Selection** window.
- e. Click **Done**. Your plot (in linear mode) will appear:



- f. Close the **Plot** window.
- g. On the **Compute** menu, click **Solutions**. The **Solution** dialog box will open.
- h. In the **Current Solution** box, click **Add**. The **Solutions Variables** box will open with a list of variables you can save.
- i. Click **s1**.

- j. Click **Done**.
- k. In the **Solutions** box, click **1**. The list of saved variables and parameters (**s1** only in this case) will appear in the **Solutions Variables** box. The **Solutions** dialog box will appear:

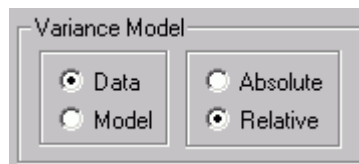


The {1} in **s1{1}** refers to solution number **1**; **s1{1}** can be accessed anywhere other model variables can be accessed. An example is the **Plot and Table Variables** dialog box when the **List All Variables** box is checked.

- l. Close the **Solutions** dialog box.


This solution will be your baseline solution to which you will return before resetting the weighting scheme and Re-Fitting the model to the data.

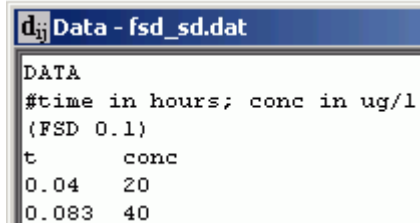
5. Perform the first fit using $FSD = 0.1$ and the data-relative variance model.
 - a. Confirm the data-relative variance model. In the **Compute** menu, click **Settings**. The **Computational Settings** dialog box will open. In this box, the **Variance Model** pane should appear as follows:



If it does not appear as shown above, change the settings so that it does appear with the data-relative option.

Close the **Computational Settings** dialog box.

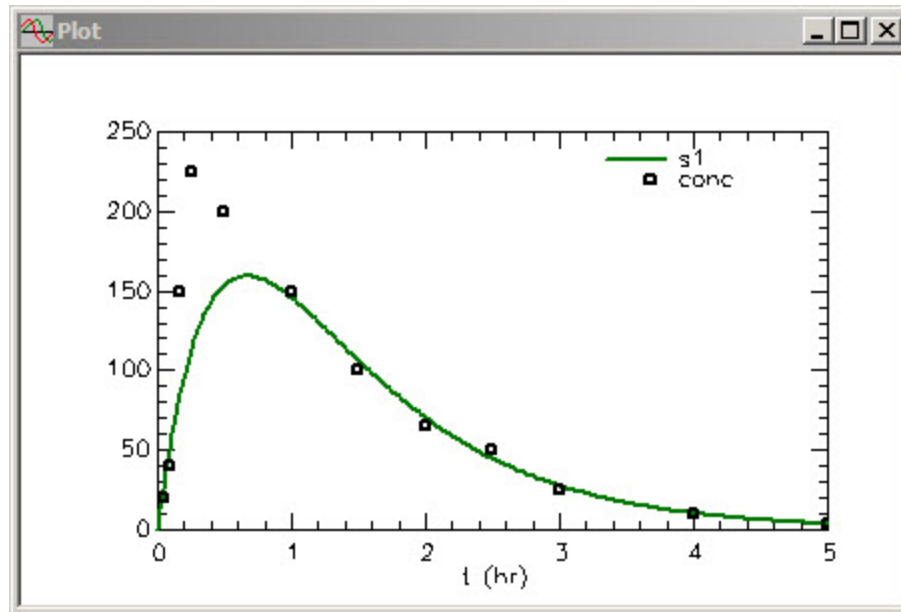
- b. Confirm the weighting scheme is (FSD 0.1). In the **Show** menu, click **Data**, or alternatively, on the **SAAM II Toolbar** click **Data** . The **Data** window will open, in part, as follows:



```
DATA
#time in hours; conc in ug/l
(FSD 0.1)
t      conc
0.04   20
0.083  40
```

The weighting scheme should be (FSD 0.1). If it is not, change the scheme so that it is, and the first part of the **Data** window appears as shown above. Close the **Data** window.

- c. Re-Solve the model (this is the baseline solution) and open the **Plot** window. The linear plot of **s1:conc** should appear.
- d. Fit the model to the data. The **Plot** window will be updated as follows (the Y Axis scale has been set from 0 to 250):

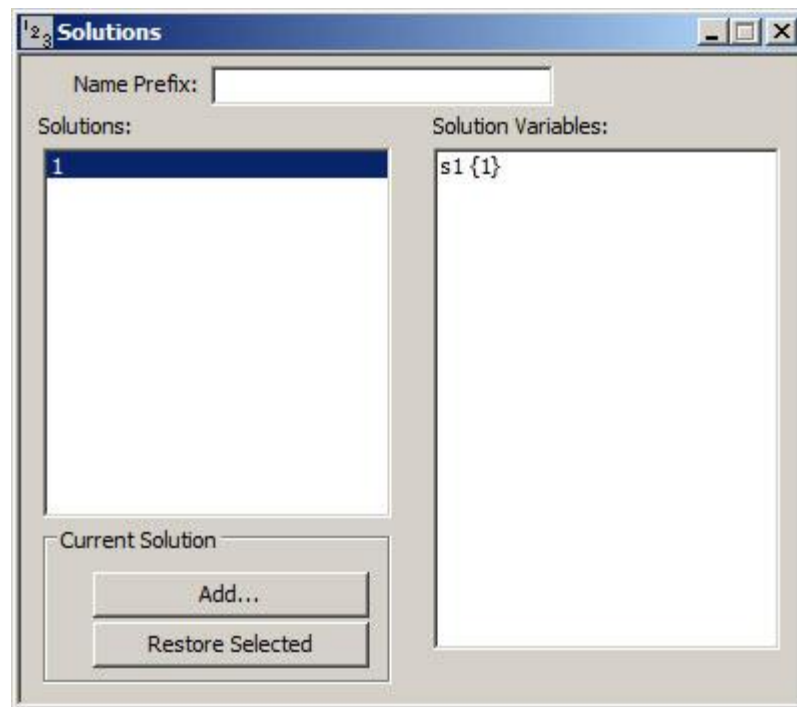


The statistics associated with the best fit are also shown. Notice that while the precision of the parameter estimates is quite reasonable, the Fit itself is not

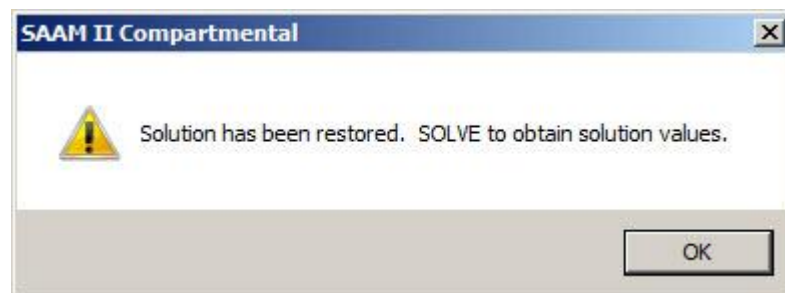
satisfactory at all because the peak of the concentration curve is not described at all! More will be said about this later in the tutorial.

- e. Close the **Plot** and **Statistics** windows.

6. Perform the first fit using $SD = 5$ and the data-relative variance model.
 - a. Restore the baseline solution.
 - (1) In the **Compute** menu, click **Solutions**. The **Solutions** dialog box will open.
 - (2) In the **Solutions** pane, click **1**. Notice as shown below, in the **Current Solution** pane, **Restore Selected** is active.



- (3) Click **Restore Selected**. The following message will appear:




This means that before you proceed, you must Re-Solve the model to return to the original **fsd_sd** which is the solution you saved.

(4) Click OK.

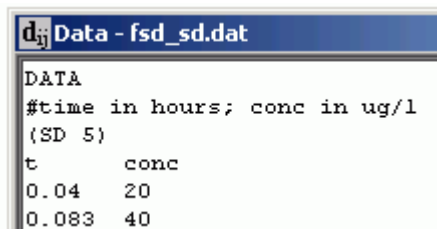
(5) Close the **Solutions** dialog box.

(6) Re-Solve the model. You can verify you have returned to the original model by viewing the plot, and checking the parameters and computational settings.

b. Change the error model to $SD = 5$.

(1) In the **Show** menu, click **Data**, or alternatively, on the **SAAM II Toolbar** click **Data** . The **Data** window will open.

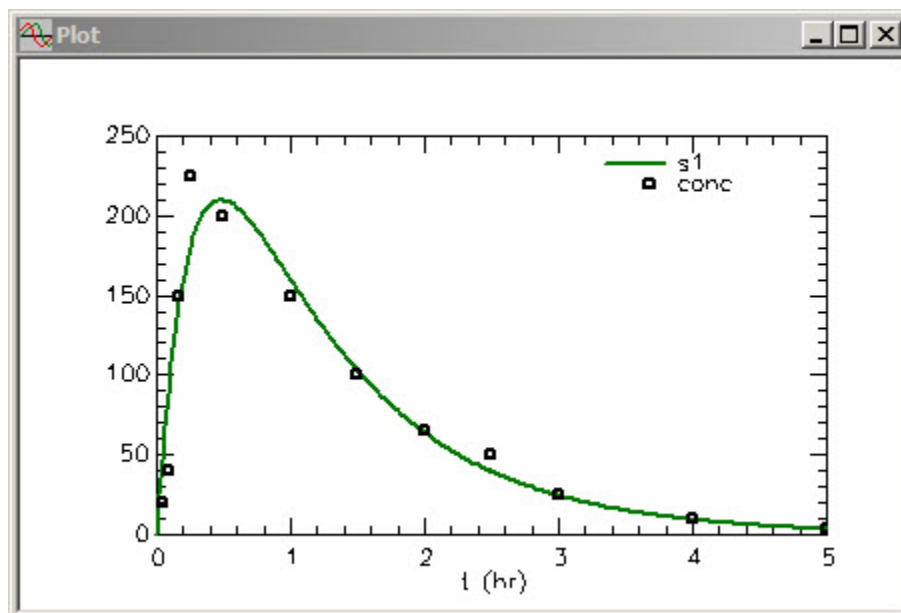
(2) Replace “(FSD 0.1)” with “(SD 5)”. The first lines of the **Data** window should appear as follows:



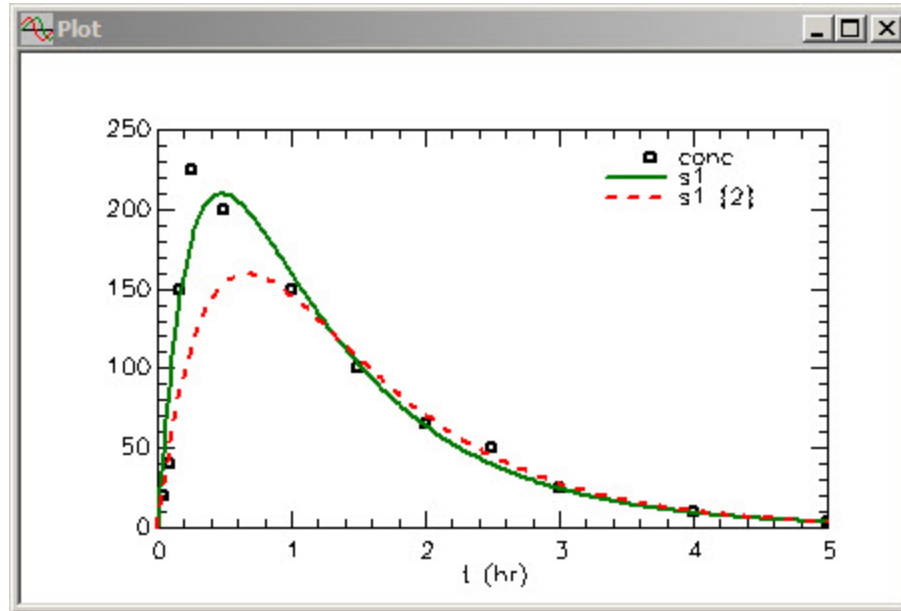
```
dij Data - fsd_sd.dat
DATA
#time in hours; conc in ug/l
(SD 5)
t      conc
0.04   20
0.083  40
```

(3) Close the **Data** window.

c. Fit the model to the data. The **Plot** and **Statistics** windows will appear as follows:



The parameter precision is reasonable, as was the case with the FSD weighting scheme. Now, however, the peak is much better described. In fact if you compare the two solutions, as shown below, you can see the differences clearly.



The solid line, **s1**, is the current solution with the SD weighting scheme while the dotted line, **s1{2}**, is the best fit with the FSD option.

d. Close the **Plot** and **Statistics** windows.

Why are there such differences between using the FSD and SD weighting options? To understand these differences, we discuss the following interrelated factors:

- Error in the data
- Variance for the error in the data
- Weight of a datum (the inverse of the variance of the error in the data).

The error in the data (variance) is specified both by the weighting scheme, i.e. (GEN A B C), FSD, SD or POIS, and the variance model, i.e. data or model, and absolute or relative. The weight assigned to a datum is the inverse of the variance.

In the case of a constant FSD, the error is a constant percentage of a datum. This means that data with small values contribute more because the weight associated with a small datum is large. In the case of FSD = 0.05 and absolute weighting, for example:

$$100(5\%) = 100 \pm 5 \quad w = \frac{1}{SD^2} = \frac{1}{25} = 0.04$$

$$1.0(5\%) = 1.0 \pm 0.05 \quad w = \frac{1}{SD^2} = \frac{1}{0.0025} = 400$$

where “w” is the weight. So you can see “1” contributes far more in terms of “importance” than “100.”

In the case of a constant standard deviation, the error is constant over all data and the weights of the data are the same. In this case, data with large values contribute more. For example, if SD = 0.5,

$$100 \pm 0.5 = 100(0.5\%) \quad w = \frac{1}{SD^2} = \frac{1}{0.25} = 4$$

$$1.0 \pm 0.5 = 1.0(50\%) \quad w = \frac{1}{SD^2} = \frac{1}{0.25} = 4$$

What do we mean by “contribute more?” Remember in fitting, a weighted residual is created:

$$wres_i = w_i \cdot (datum_i - model_i)^2$$

The weight can contribute significantly to this term as is the case in the FSD option. In the SD case where all weights are the same, large data contribute more simply because the term $(datum_i - model_i)^2$ is normally larger for large data.

With the data in this tutorial, there are only 2 – 3 data of the 12 that define the peak. In the case of the FSD, the small values outweigh these, and hence “drag down” the best fit. On the other hand, in the SD case, these 2 – 3 data do carry more weight, and tend to “help” the fit achieve a better description of the peak.

In your modeling exercise, when you have rising and falling data such as these, you may want to consider using the SD option for this reason.

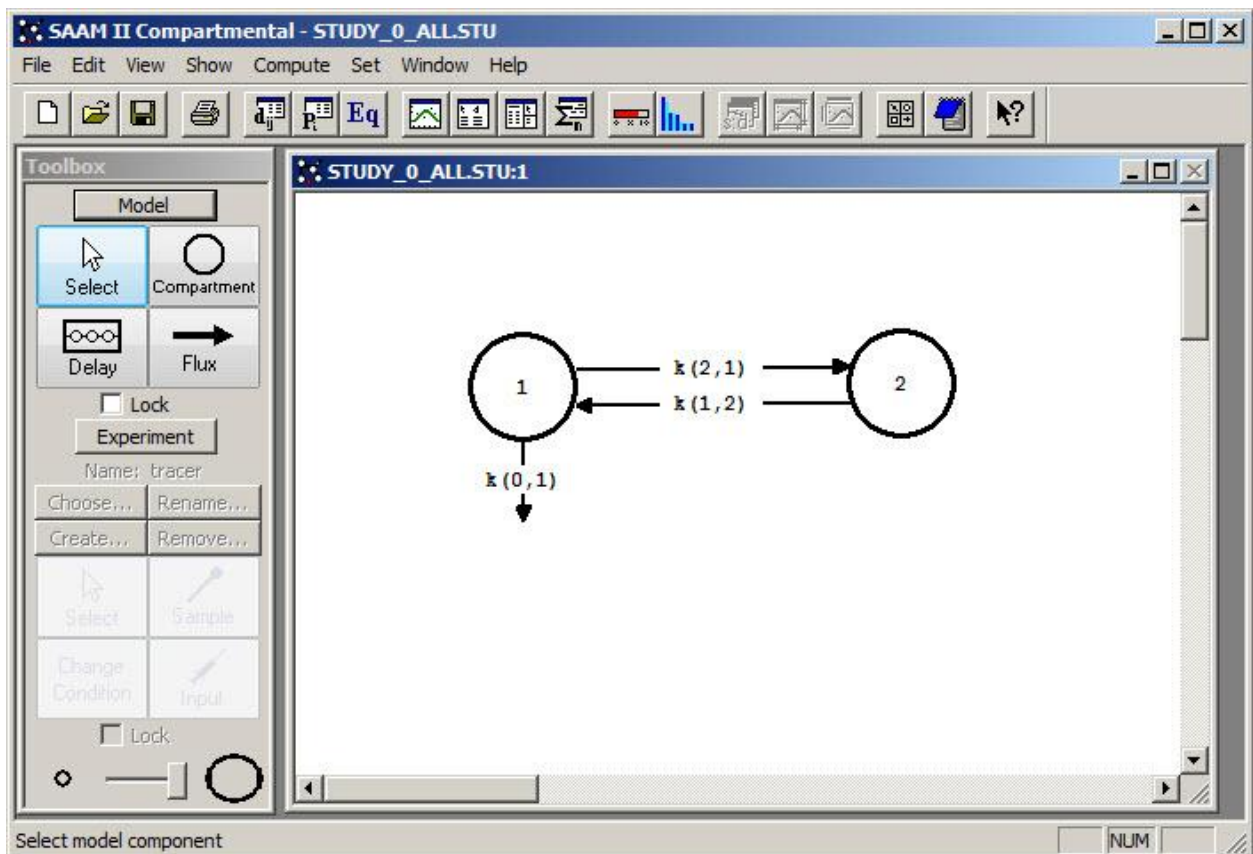
Quit the SAAM II Compartmental application. Do not save the changes to `fsd_sd`.

Part 4. Weighting and unweighting data

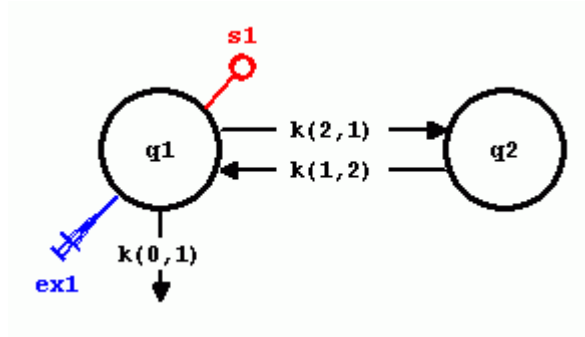
As part of your modeling exercise, there may be occasions when you feel a datum, or sequence of data, is spurious. It is possible that just one datum, an outlier, can create problems especially when you fit your model to your data.

This tutorial will explain how to unweight (and re-weight) a single datum and series of data.

1. Start the **SAAM II Compartmental** application. The **SAAM II Compartmental** main window will open.
2. Open the **SAAM II Compartmental** study file **study**.
 - a. The file **study_0_all.stu** should appear in the file list; if it does not, find the folder where you put this tutorial.
 - b. In the **File** menu, click **Open**. The **SAAM II Compartmental** main window will appear as shown below:



3. View the model and the experiment on the model. In the **SAAM II Toolbox**, click **Experiment**. The model of the experiment will appear on the **Drawing Canvas** as follows:



The model is the same as the model for **study_0**. The differences are:

- The study lasts 19 days instead of 9 (as can be seen by viewing the **Experiment Attributes** dialog box opened in the **Set** menu).
- Sample **s1** appears as an open circle because no data has been associated with the sample. There is no data in the Data window.

The experimental input and initial parameter values are the same as those in **study_0**. This can be seen by viewing the **Exogenous Input** and **Parameters** dialog boxes.


4. Copy the data into the **Data** window, and view the data.

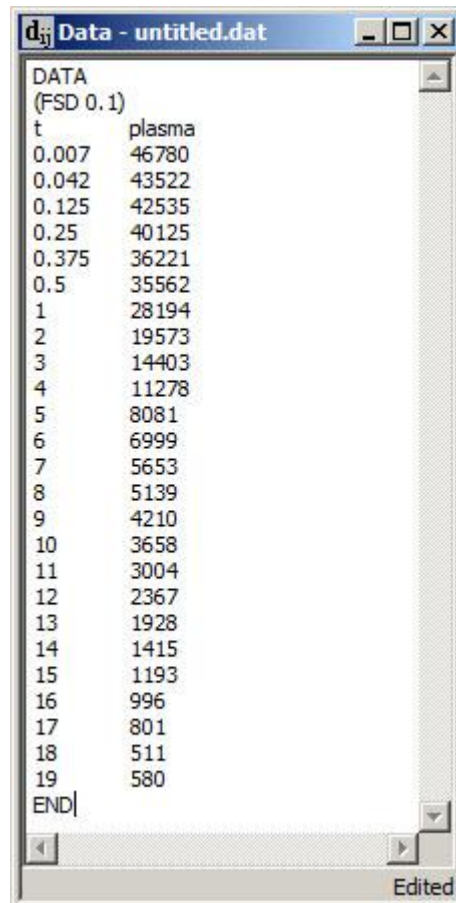
The data are in the Excel® spreadsheet **study_0_all.xls**.


- a. **Start** Excel® and open **study_0_all.xls**. You will have to move to the folder where you put this file. The spreadsheet will appear in part as follows:

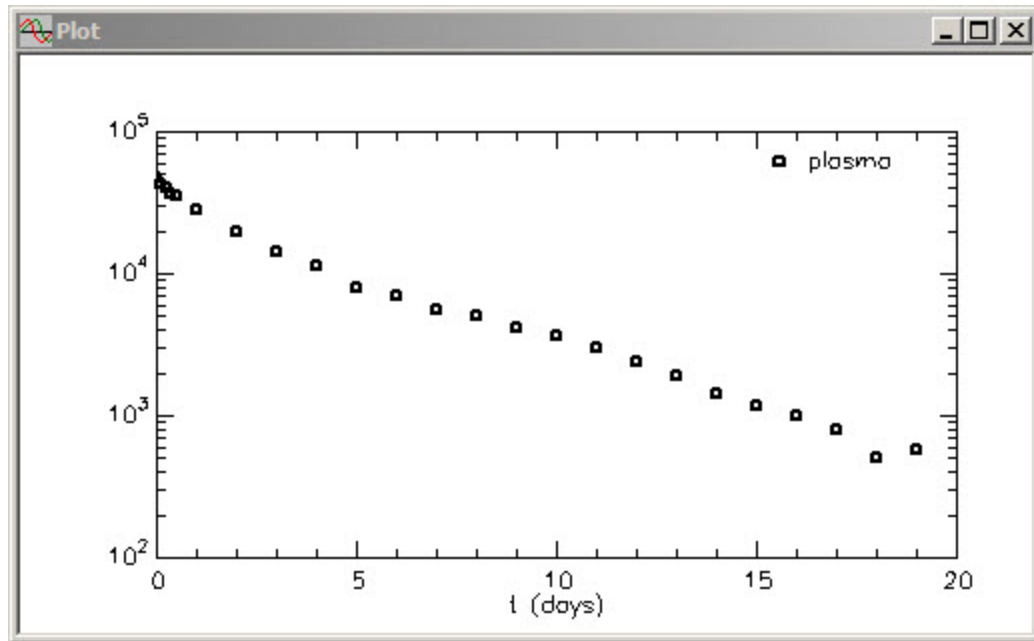
	A	B
1	DATA	
2	(FSD 0.1)	
3	t	plasma
4	0.007	46780
5	0.042	43522
6	0.125	42535
7	0.25	40125
8	0.375	36221
9	0.5	35562
10	1	28194
11	2	19573
12	3	14403
13	4	11278

- b. Select and copy the data beginning with the DATA line and ending with the END line.

- c. Return to the SAAM II study file **study_0_all**, and open the **Data** window. In the **Show** menu, click **Data**, or alternatively, on the **SAAM II Toolbar** click **Data** . The **Data** window will open.
- d. In the **Edit** menu, click **Paste**. In the **Edit** menu, click **Check Data Format**. Correct any errors. The **Data** window should appear as follows.

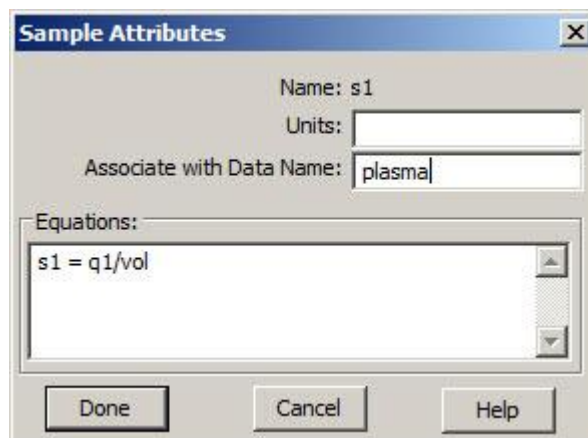


- e. Close the **Data** window.
- f. Plot the data. On the **Show** menu, click **Plot**, or alternatively click **Plot**  on the **SAAM II Toolbar**. The **Plot and Table Variables** dialog box will open. Be sure the **List All Variables** box is checked.
- g. Click **plasma** to move this to the **Current Selection** window.
- h. Click **Done**. Your plot (in semi-log mode and the X Axis maximum set equal to 20) will appear:





Setting the X Axis maximum equal to 20 allows you to visualize the final data better. You can see that the datum at day 18 may be an outlier. In addition, you can see there could be a “break” in the curve around day 9. Both will be addressed in the following.

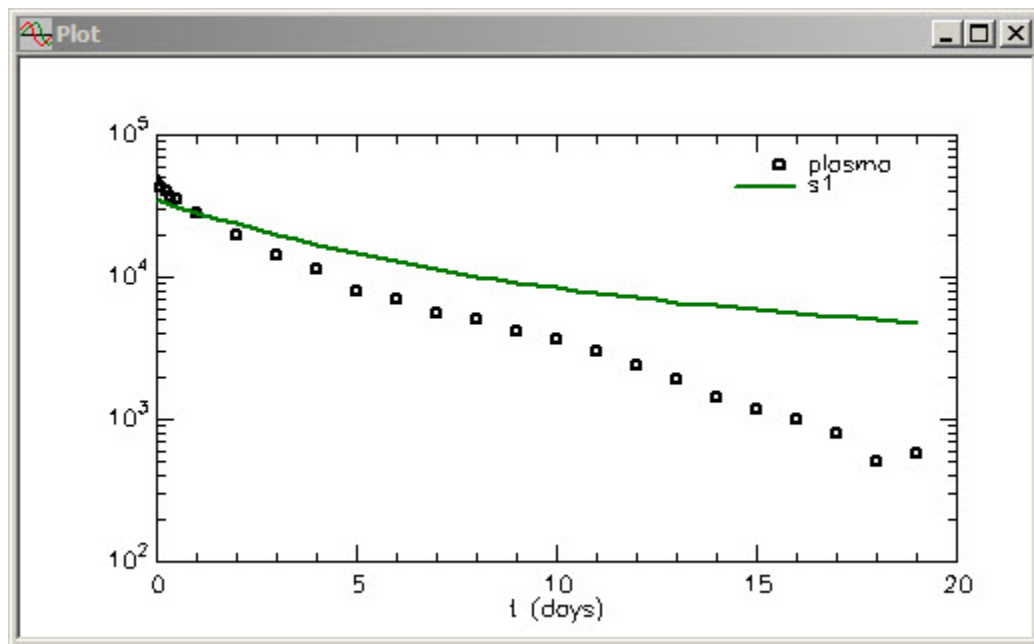
- i. Close the **Data** window.
5. Associate **s1** with the plasma data.
 - a. Double-click **s1** to open the **Sample Attributes** dialog box.
 - b. Type “plasma” in the **Associate with Data Name** box. The **Sample Attributes** dialog box will appear as follows:



- c. Click **Done**.

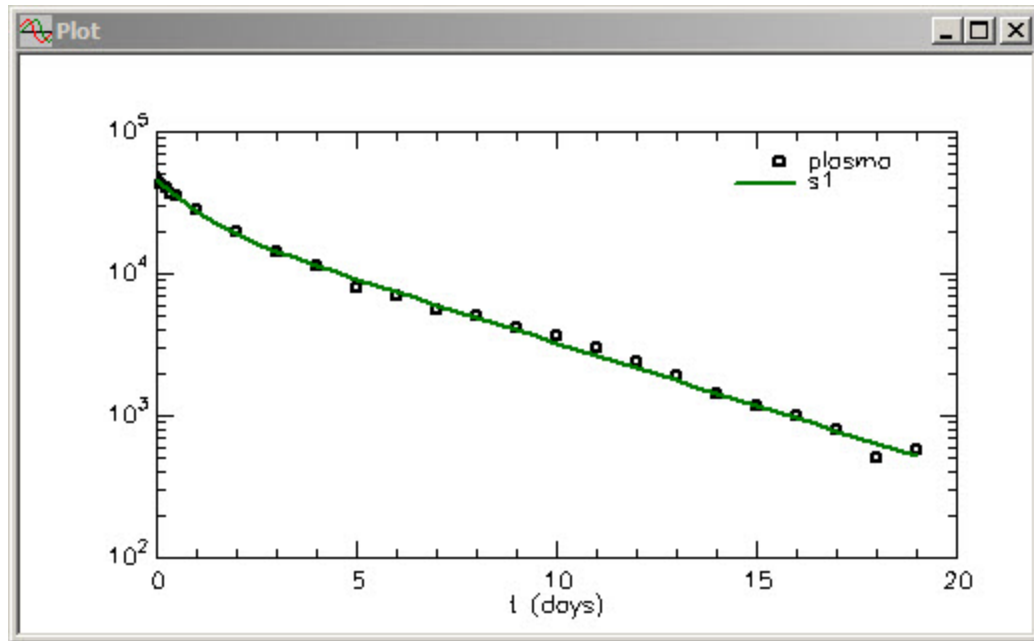
6. Fit the model to the data.

- a. On the **Compute** menu, click **Solve**, or alternatively click **Solve**  on the **SAAM II Toolbar**.
- b. Plot the solution. On the **Show** menu, click **Plot**, or alternatively click **Plot**  on the **SAAM II Toolbar**. The **Plot and Table Variables** dialog box will open. Be sure the **List All Variables** box is not checked.
- c. Click **s1:plasma** to move this to the **Current Selection** window.
- d. Click **Done**. Your plot (in semi-log with the X Axis maximum set equal to 20) will appear as follows:



Leave the **Plot** window open.

- e. Fit your model to your data. On the **Compute** menu, click **Fit**, or alternatively click **Fit**  on the **SAAM II Toolbar**. The **Plot** window will be updated as follows:



You can see the fit is reasonable except the datum at day 18 appears as an outlier, and there is clearly a “hump” in the data around day 9 or 10. The statistics associated with the Fit are shown as follows:

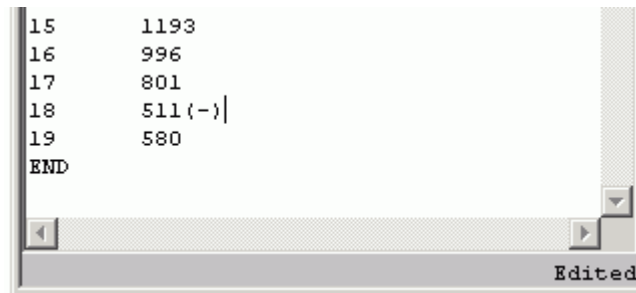
Parameter/Variable	Value	Std.Dev.	Coef. of Var.	95% Confidence Interval	
k(0,1)	0.32256	1.67863e-002	5.20411e+000	0.28765	0.35747
k(1,2)	0.67453	1.60509e-001	2.37957e+001	0.34073	1.00832
k(2,1)	0.27456	1.09778e-001	3.99835e+001	0.04626	0.50286
vol	2263.71700	1.03475e+002	4.57102e+000	2048.52937	2478.90464

	Objective	Scaled Data Variance
s1 : plasma	1.331310e+001	6.976760e-001
Total objective	1.331310e+001	
AIC	7.775486e+000	
BIC	7.897374e+000	

- f. Close the **Plot** and **Statistics** windows.

7. Unweight the datum at day 18, and Re-Fit the model to the data.
 - a. Open the **Data** window.

- b. Type “(-)” after “511” for the datum at day 18. This part of the **Data** window will appear as follows:



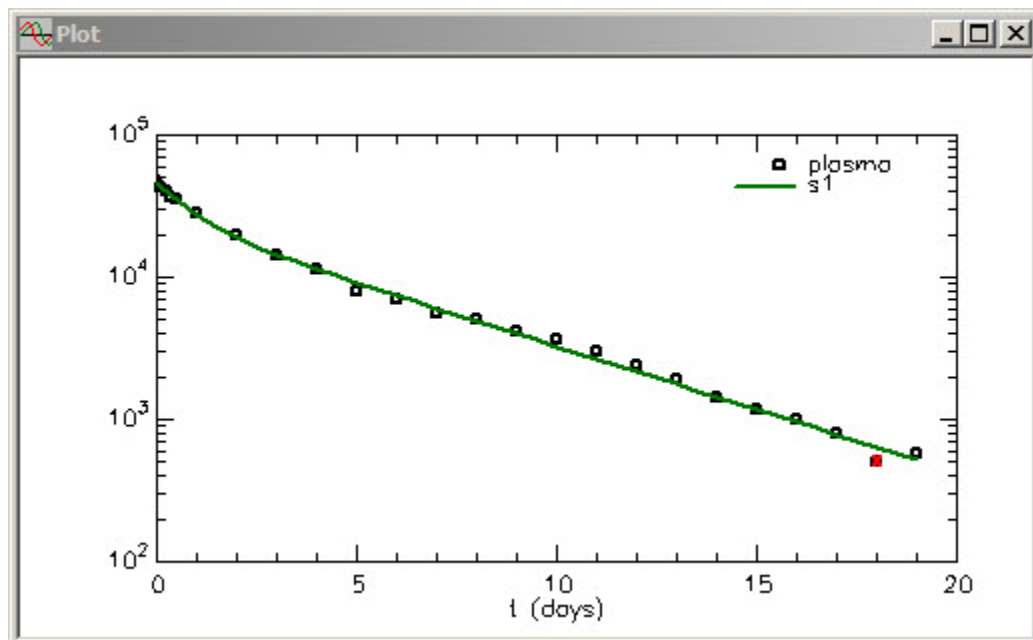
15	1193
16	996
17	801
18	511 (-)
19	580
END	



Unweighting a single datum. To unweight a single datum, simply type “(-)” after the value. The datum will appear on subsequent plots with a red “x” through it. The datum will not be taken into account during the fitting process.

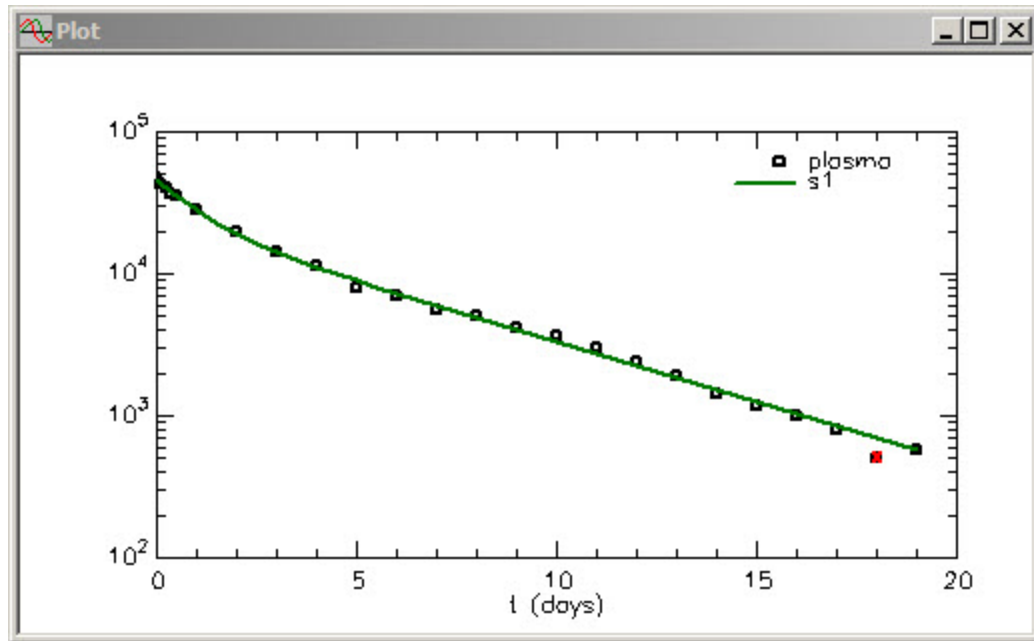


- c. Close the **Data** window.
- d. Re-Solve the model, and plot the solution. Since the parameters following the previous Fit are active, the plot will appear as follows:



Note the red “x” through the datum at day 18 to indicate it is unweighted.

- e. Re-Fit the model to the data. The plot will be updated as follows:

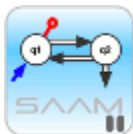


You can see there is a small difference with the final decay. The statistics for the fit are shown as follows:

Parameter/Variable	Value	Std.Dev.	Coef. of Var.	95% Confidence Interval	
k(0,1)	0.32024	1.03390e-002	3.22851e+000	0.29867	0.34181
k(1,2)	0.58339	8.25974e-002	1.41583e+001	0.41109	0.75568
k(2,1)	0.24742	5.83648e-002	2.35890e+001	0.12568	0.36917
vol	2274.02446	6.47359e+001	2.84675e+000	2138.98806	2409.06086

Objective		
s1 : plasma	Objective	Scaled Data Variance
	1.264153e+001	2.919730e-001

Total objective	1.264153e+001	
AIC	7.448036e+000	
BIC	7.570750e+000	



Statistics and unweighting a datum. You can see some improvement in the statistics just by removing the single “outlier”. In this situation, the change is not dramatic. In some situations, removing a single “outlier” can mean the difference between a successful and unsuccessful Fit.

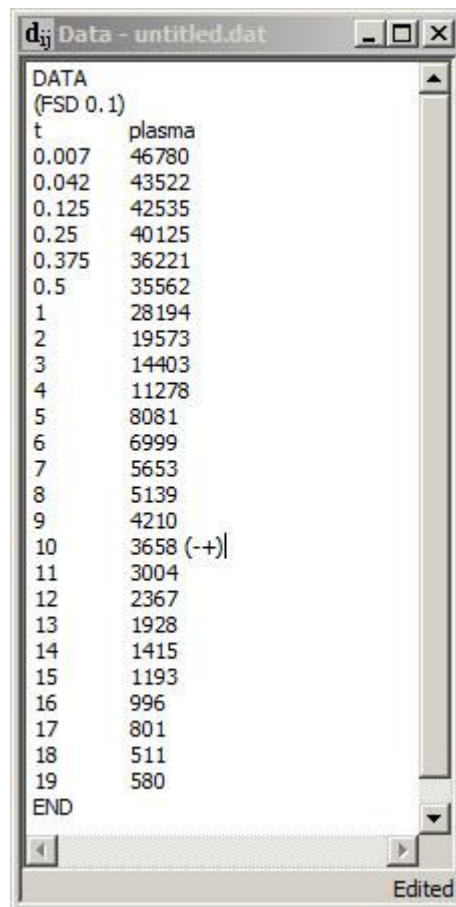


Close the **Plot** and **Statistics** windows.

- f. Open the **Data** window, and remove “(-)” from the datum at day 18. This will essentially re-weight this datum. Re-Solve the model. The best-fit parameter values from the previous fit are the parameters that are current.
8. Unweight the data from day 10 to day 19, and Re-Fit the model to the data.

There may be occasions when you may want to unweight a string of data. This could be done, for instance, if you believe there has been a perturbation that will change a model parameter starting at some point during an experiment. This part of the tutorial will show you have to unweight such a string, and the effect this can have on the model solution.

- a. Open the **Data** window.
- b. Type “(-+)” after “3658” for the datum at day 10. The **Data** window will appear as follows:



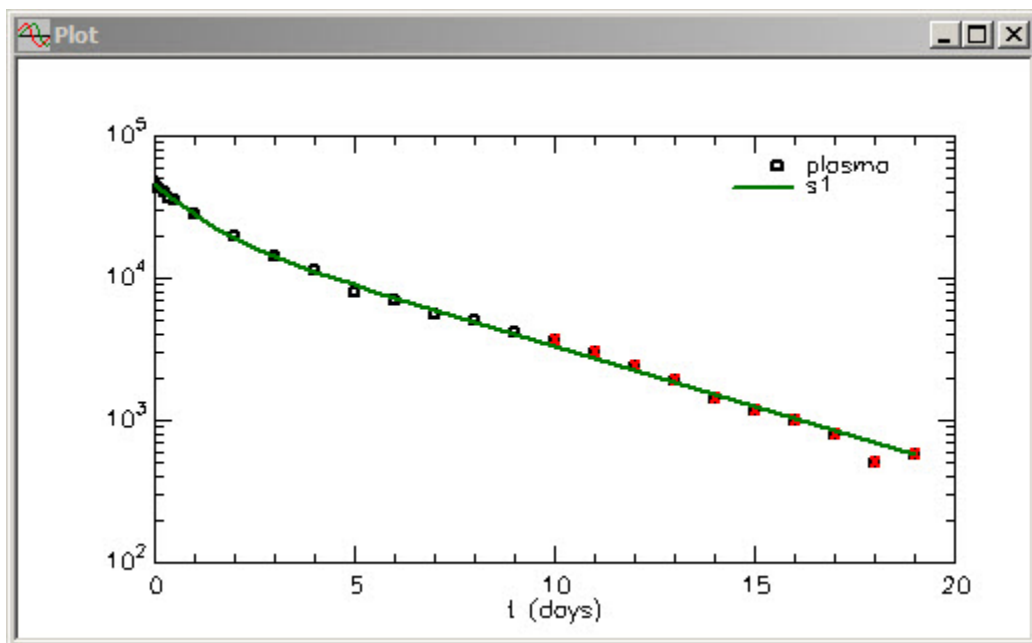


Unweighting a string of data. You can unweight a string of data by typing “(-+)” after the first datum to be unweighted. All subsequent data will be unweighted until a “()” is encountered associated with the first datum to be re-weighted; this means to weight this and subsequent data.



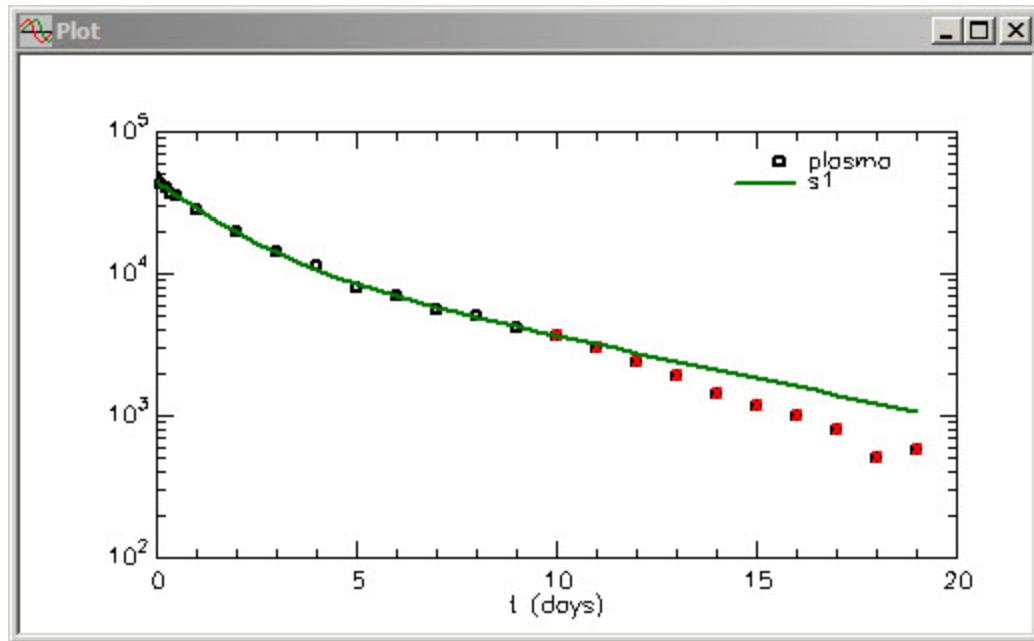
Close the **Data** window.

- c. Solve the model and view the solution. The plot will appear as follows:



Notice the red “x” through all data starting at day 10; this means these data are unweighted and will not be used when the model is fitted to the data.

- d. Re-Fit the model to the data. The plot will be updated as follows:



You can clearly see the difference between the fits. In this case, the fit of the initial decay to day 9 (which is the same as **study_0**) is quite good; the model predictions to day 19 indicate the deviation in the data from the model predicted values. This is indicative of a perturbation that has changed the value of one or more rate constants in the model.

Unweighting strings of data is a very useful modeling technique to test for such perturbations.

The statistics associated with this fit are shown as follows:

Parameter/Variable	Value	Std.Dev.	Coef. of Var.	95% Confidence Interval	
k(0,1)	0.29251	1.45634e-002	4.97884e+000	0.26045	0.32456
k(1,2)	0.28391	7.51308e-002	2.64628e+001	0.11855	0.44927
k(2,1)	0.17471	1.83399e-002	1.04975e+001	0.13434	0.21507
vol	2323.46997	3.85704e+001	1.66003e+000	2238.57695	2408.36298

	Objective	Scaled Data Variance
s1 : plasma	1.336757e+001	1.113667e-001
Total objective	1.336757e+001	
AIC	7.936055e+000	
BIC	8.054063e+000	

The numbers are obviously different from the previous fits because only data to day 9 have been fitted by the model.

- e. Close the **Plot** and **Statistics** window.

Quit the **SAAM II Compartmental** application. Do not save the changes to **study_0_all**.

Part 5. Working with data replicates.

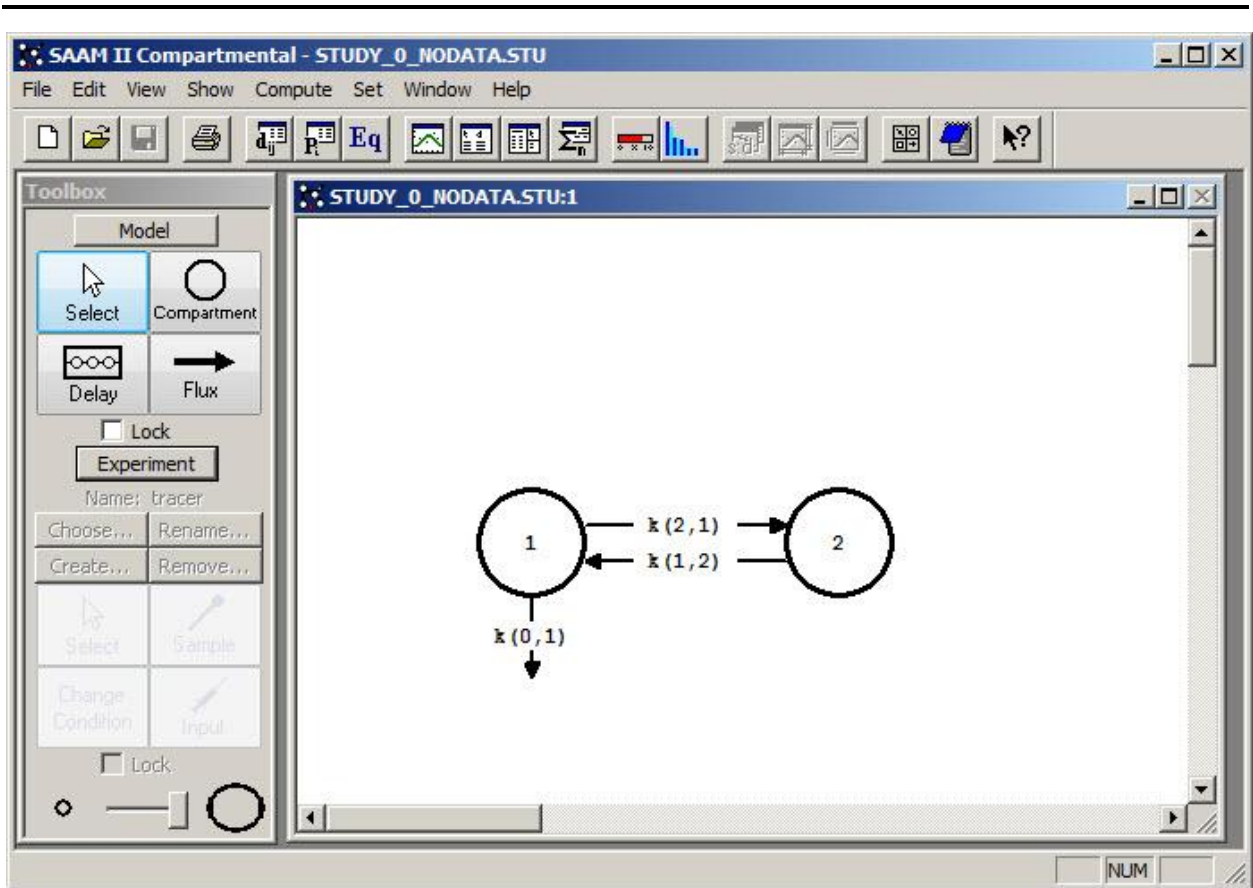
Suppose you have an experimental design where, at each sample time, you collect data in replicates. As an example, suppose the amount of material in each plasma sample is measured three times. Thus you would have plasma data in triplicates. This situation can be treated as having more than one measurement from the same sampling site of a single study or as separate studies.

This tutorial will show you have to deal with data replicates in SAAM II. There are essentially two ways this can be done. One way is to enter all data in a single table, but for the times at which there are replicates, change the time slightly of the replicates. For example, at time “1” there were three plasma measurements of 113, 105 and 120, you could enter in the table:

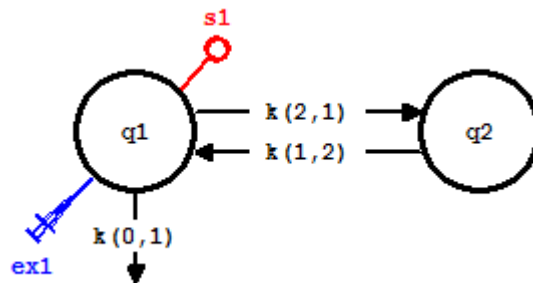
0.99	113
1.00	105
1.01	120

This can be cumbersome, and does not really reflect what was done in the laboratory. The other way of dealing with replicates will be illustrated here; this uses the Sample tool by creating a separate sample for each replicate.

1. **Start the SAAM II Compartmental** application. The **SAAM II Compartmental** main window will open.
2. Open the **SAAM II Compartmental** study file **study**.
 - a. The file **study_0_nodata.stu** should appear in the file list; if it does not, find the folder where you put this tutorial.
 - b. In the **File** menu, click **Open**. The **SAAM II Compartmental** main window will appear as shown below:



- View the model and the experiment on the model. In the **SAAM II Toolbox**, click **Experiment**. The model of the experiment will appear on the **Drawing Canvas** as follows:



The model is the same as the model for **study_0**. The difference is:

- The data have not been copied into the **Data** window so the sample **s1** appears as an open circle.

The experimental input and initial parameter values are the same as those in **study_0**. This can be seen by viewing the **Exogenous Input** and **Parameters** dialog boxes.


-
4. Copy the data into the **Data** window, and view the data.

The data are in the Excel® spreadsheet **replicate_data.xls**.


- a. **Start** Excel® and open **replicate_data.xls**. You will have to move to the folder where you put this file.
- b. Select and copy the data beginning with the DATA line and ending with the END line. The selected data will appear as follows:

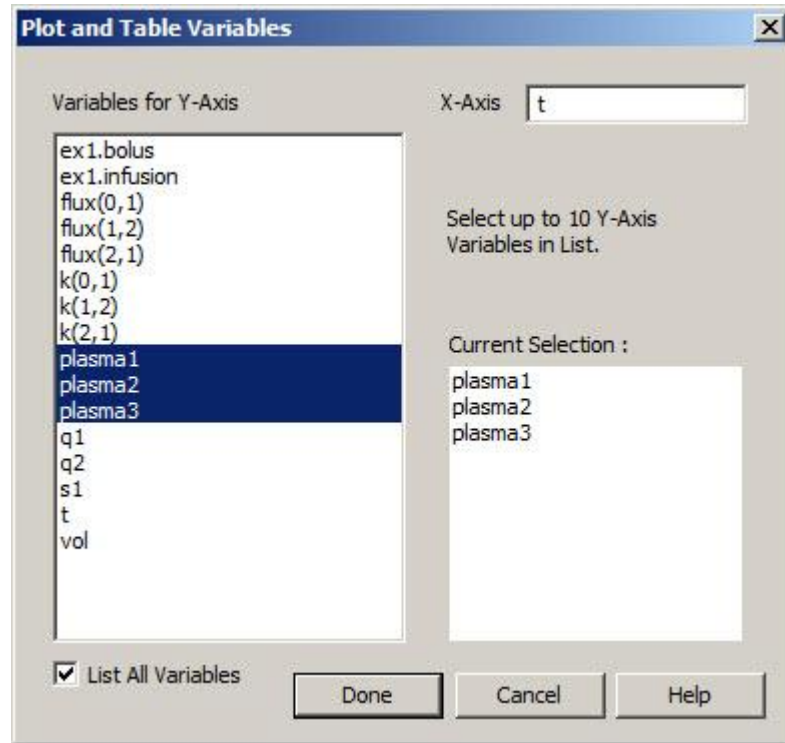
	A	B	C	D
1	DATA			
2	(FSD 0.1)			
3	t	plasma1	plasma2	plasma3
4	0.007	46780	43593	44002
5	0.042	43522	45876	43166
6	0.125	42535	40725	40658
7	0.25	40125	39613	40978
8	0.375	36221	38968	39048
9	0.5	35562	36670	36661
10	1	28194	27516	28115
11	2	19573	19881	19136
12	3	14403	14775	14594
13	4	11278	10511	10299
14	5	8081	8123	8553
15	6	6999	7064	7229
16	7	5653	6014	5727
17	8	5139	5178	4851
18	9	4210	4212	4152
19	END			

The file contains three replicates of the plasma data at each time point; these are listed in the columns “plasma1”, “plasma2” and “plasma3”.

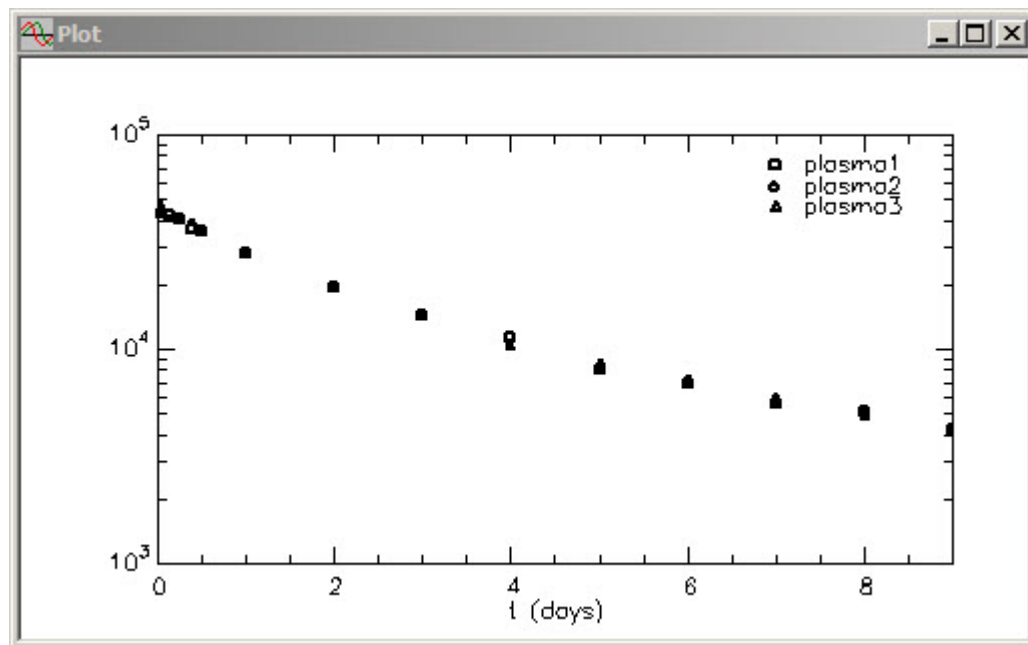
- c. Return to the SAAM II study file **study_0_nodata**, and open the **Data** window. In the **Show** menu, click **Data**, or alternatively, on the **SAAM II Toolbar** click **Data** . The **Data** window will open.
- d. In the **Edit** menu, click **Paste**. In the **Edit** menu, click **Check Data Format**. Correct any errors. The **Data** window should appear as follows.

t	plasma1	plasma2	plasma3
0.007	46780	43593	44002
0.042	43522	45876	43166
0.125	42535	40725	40658
0.25	40125	39613	40978
0.375	36221	38968	39048
0.5	35562	36670	36661
1	28194	27516	28115
2	19573	19881	19136
3	14403	14775	14594
4	11278	10511	10299
5	8081	8123	8553
6	6999	7064	7229
7	5653	6014	5727
8	5139	5178	4851
9	4210	4212	4152

- e. Close the **Data** window.
- f. Plot the data. On the **Show** menu, click **Plot**, or alternatively click **Plot**  on the **SAAM II Toolbar**. If the **Plot and Table Variables** dialog box is not open, open it by clicking **Plot/Table Variables** in the **Compute** menu. Be sure the **List All Variables** box is not checked.
- g. Click **plasma1**, **plasma2**, and **plasma3** to move these to the **Current Selection** window. The **Plot and Table Variables** dialog box will appear as follows:



h. Click **Done**. Your plot (in semi-log) will appear:

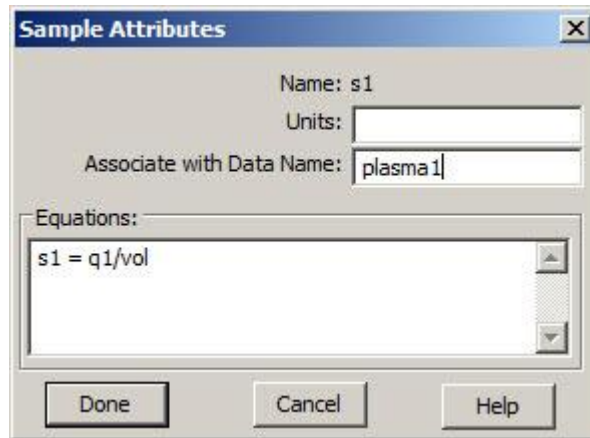


You can see each set of data is represented by a different symbol.

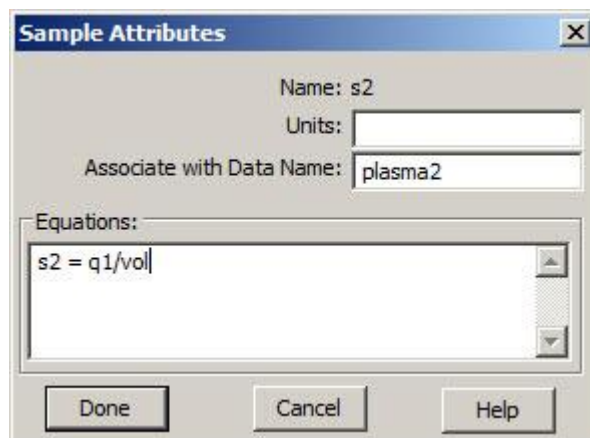
i. Close the **Data** window.

5. Associate **s1** with **plasma1** data.

- a. Double-click **s1** to open the **Sample Attributes** dialog box.
- b. Type “plasma1” in the **Associate with Data Name** box. The **Sample Attributes** dialog box will appear as follows:



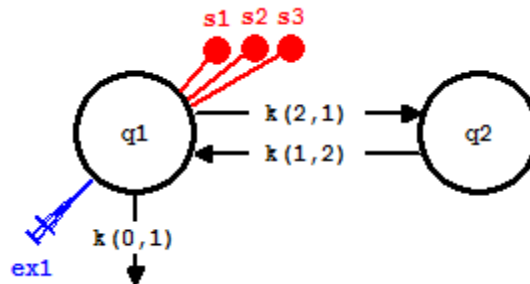
- c. Click **Done**.
6. Create samples **s2** and **s3**, and associate them with **plasma2** and **plasma3** respectively.
 - a. In the **SAAM II Toolbox**, click **Sample**.
 - b. Click Compartment **q1** and then the **Drawing Canvas**. A sample **s2** will appear associated with Compartment **q1**.
 - c. Double-click **s2** to open the **Sample Attributes** dialog box.
 - d. Type “plasma2” in the **Associate with Data Name** box.
 - e. Edit the sample equation to read “ $s2=q1/vol$ ”. The **Sample Attributes** dialog box will appear as follows:





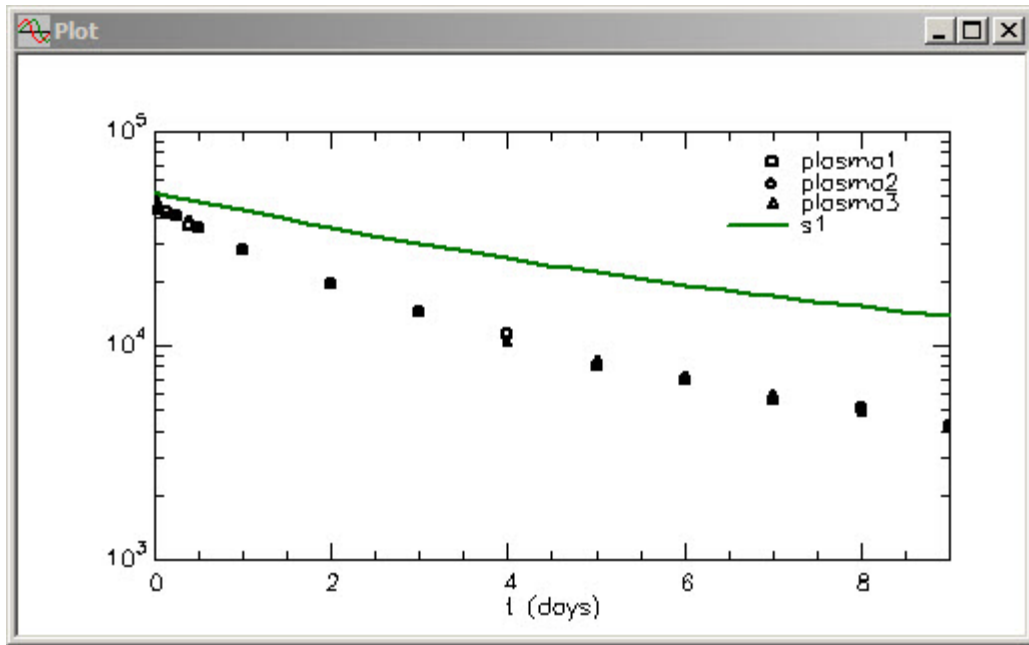
Notice the sample equation for **s2** is the same as for **s1**; this is because the model solution does not change; what changes are the data sets between the two samples.

Click **Done**.

- f. Repeat the process creating a sample **s3** and associating it with “plasma3”. Remember the sample equation will be “ $s3=q1/vol$ ”. When you have finished, your model will appear as follows.

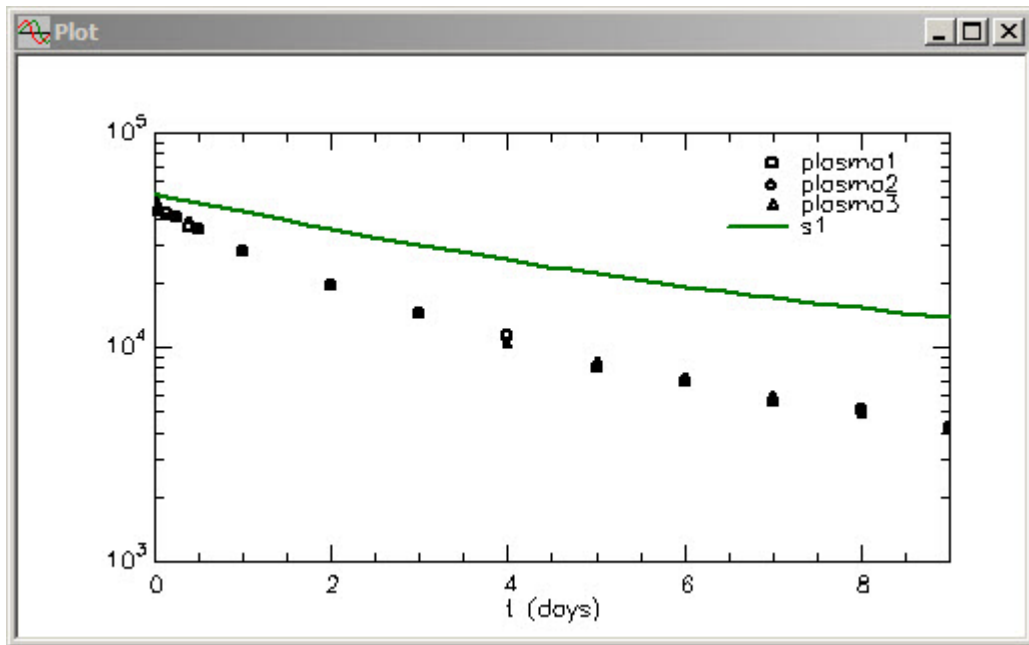


7. Fit the model to the data.
- On the **Compute** menu, click **Solve**, or alternatively click **Solve**  on the **SAAM II Toolbar**.
 - Plot the solution. On the **Show** menu, click **Plot**, or alternatively click **Plot**  on the **SAAM II Toolbar**. The **Plot and Table Variables** dialog box will open. Be sure the **List All Variables** box is checked.
 - Press the **CTRL** key, and click **s1**, **plasma1**, **plasma2**, and **plasma3** to move these to the **Current Selection** window.
 - Click **Done**. Your plot (in semi-log) will appear as follows:

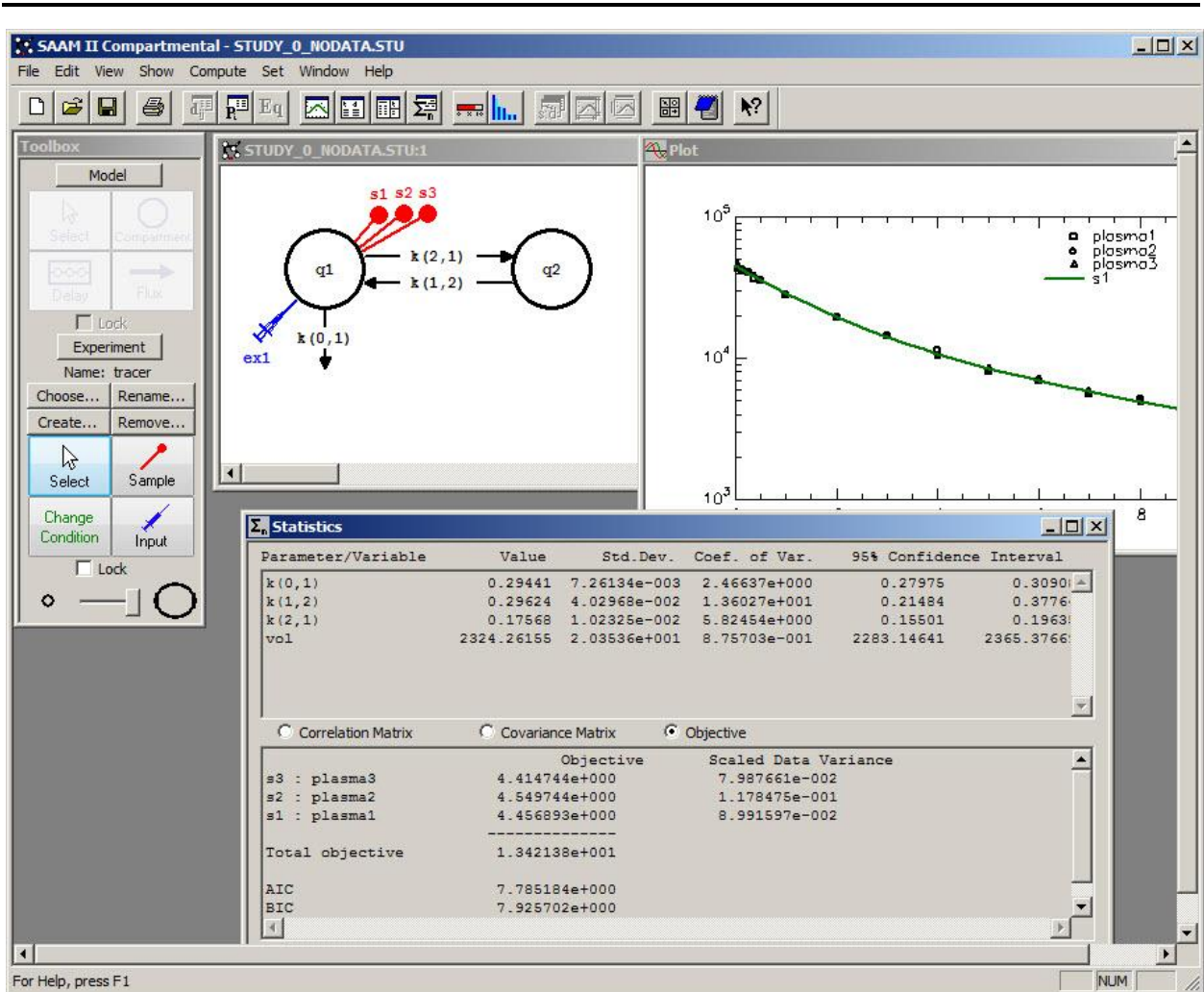


Leave the **Plot** window open.

- f. Fit your model to your data. On the **Compute** menu, click **Fit**, or alternatively click **Fit**  on the **SAAM II Toolbar**. The **Plot** window will be updated as follows:



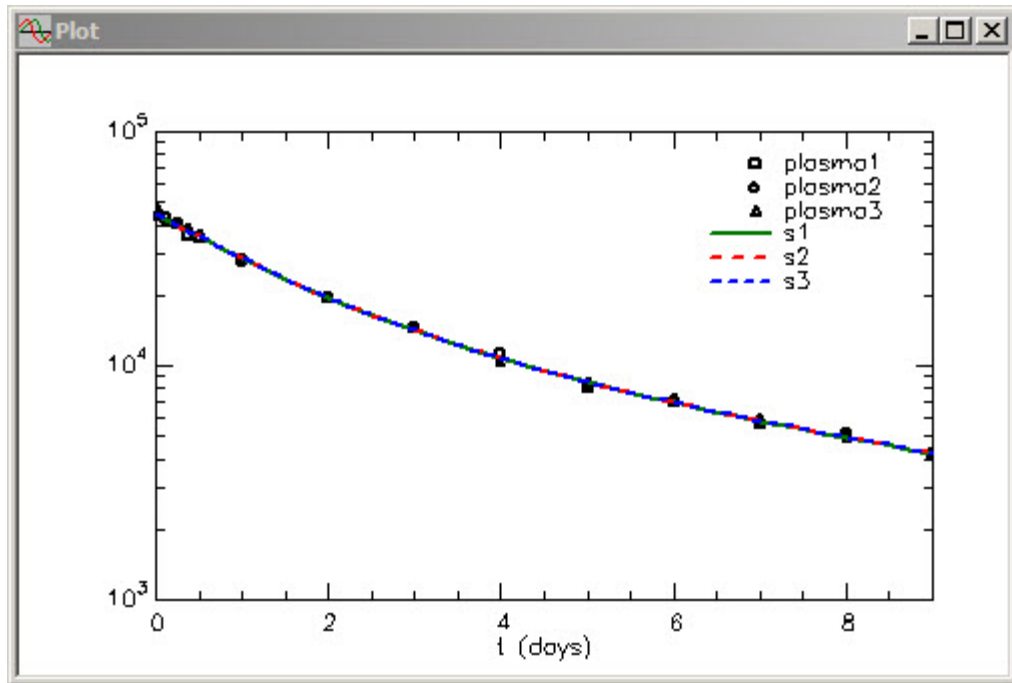
The model, **Plot** and **Statistics** windows will appear on the **Drawing Canvas** with the model as follows:



g. Close the **Plot** and **Statistics** windows.



Working with data replicates. The easiest and probably best way to work with data replicates is to create a sample for each set of data replicates. Remember, however, that the sample equation for each sample will be the same. In the **Plot and Table Variables** dialog box, when the **List All Variables** check box is not selected, you will be presented (in this example) with the options **s1:plasma1**, **s2:plasma2**, and **s3:plasma3**. If you choose this, you will get plots of **s1**, **s2** and **s3** which, of course, are all identical. The plot will appear as follows:



There is nothing wrong with this except it appears a little “busier” because of **s2** and **s3**. Thus the presentation of the results is a matter of personal preference. You can do it this way, or select the **List All Variables** check box, and plot only **s1** together with the data replicates as was done in this tutorial.



Quit the **SAAM II Compartmental** application. Do not save the changes to **study_0_nodata**.