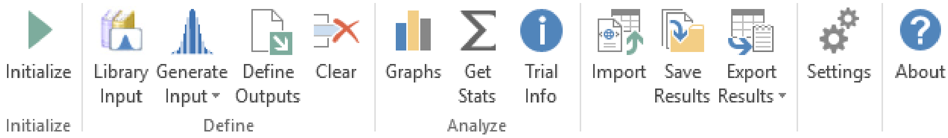


SIPmath™ Modeler Tools for Excel

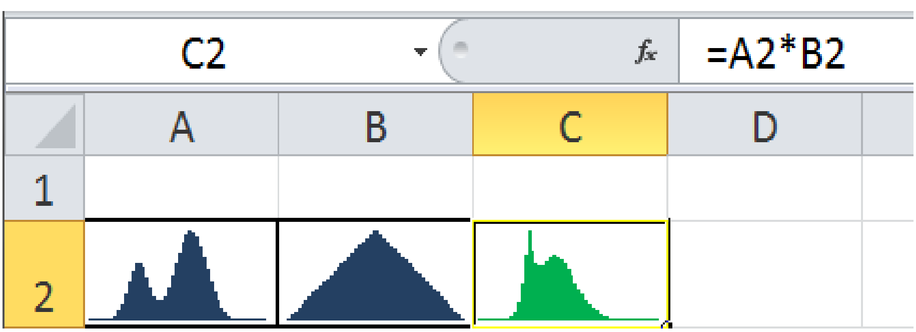
REFERENCE MANUAL

by

Sam Savage, Marc Thibault, and Dave Empey

Rev. 29 May 2017

SIPmath Modeler Tools Version 3.4.0



Compatibility: Excel 2010-2016 for Windows, Excel 2016 for Mac

Contents

1 INTRODUCTION 4

2 SIPMATH OPERATING MODES 7

3 SIPMATH MODELER TOOLS OVERVIEW 9

4 EXERCISE: MODEL IN GENERATE MODE 12

5 EXERCISE: MODEL IN SIP LIBRARY MODE 17

5.1 Initialize 17
5.2 Input Distributions 19
5.3 The Model 21
5.4 The Output Distributions 23
5.5 Presenting the results 24
5.6 Running Your model Without the tools 26

6 THE SIPMATH RIBBON DETAILS 28

6.1 Initialize 28
6.2 Library Inputs 29
6.3 Generate Input 31
6.4 Define Outputs 44
6.5 Clear 47
6.6 Graphs 48
6.7 Get Stats 52
6.8 Trial Info 52
6.9 Import 54
6.10 Save Results 54
6.11 Export Results 55
6.12 Settings 56
6.13 About 57

7 SIPMATH ARCHITECTURE 58

7.1 Overview 58
7.2 Stochastic Libraries 60
7.3 Input Cells 61

Contents

7.4 Output Cells 62
7.5 The Data Table 62
7.6 The Process 63
8 SIP LIBRARY SHEET FORMAT 64

Supporting Files:

SIPmath Modeler Tools 3.4.*.xlam

Tool Macro file, does not need to be present to run models

Assets Library 3.1.xlsx

File for exercise in section 5

SPT Metalogs for SIPmath.xlsx

Template for generating Metalog distributions

CBImport_SIP.xlam

File required for importing SIP libraries from Oracle Crystal Ball

ARImport_SIP.xlam

File required for importing SIP libraries from Palisade @RISK

1 INTRODUCTION

The SIPmath Modeler Tools for Excel (the **Tools**) facilitate the development of stochastic spreadsheet applications, making use of the native Data Table function[i]. Model inputs can be based on SIP Libraries or dynamic random number generation.

Once created, **Tools** applications do not require the **Tools** add-in to run. The **Tools** work with current Microsoft Excel versions on Windows and Mac OS. Although any of the functions of the **Tools** may be performed manually in Excel, the **Tools** significantly reduce the work needed to design and develop stochastic models.

Probability Management is a decision support framework for dealing with uncertainties much in the same way that traditional data management deals with regular spreadsheet numbers[ii],[iii].It is based on four principles:

1 **Communicating Distributions as Data**

Normally, spreadsheets allow each cell to store only a single value. Thus, when estimating uncertainties, spreadsheets show just one number, typically a single "average" future scenario. This leads to a set of systematic errors known collectively as the *Flaw of Averages*,[iv] which explains why so many projects are behind schedule, beyond budget, and below projection.

The SIPmath Modeler Tools enable Excel to model thousands of potential future scenarios simultaneously, so that we can make better decisions in the face of uncertain future prices, demand, project duration, etc.

These potential future scenarios are stored in *Stochastic Information Packets*, or SIPs, that contain hundreds or thousands of simulated or historical outcomes of an uncertainty. That is, SIPs are arrays filled with realizations and optional metadata concerning the uncertainty in question. A central goal of ProbabilityManagement.org is to develop cross-platform standards for formatting SIP Libraries – repositories of reliable and auditable probability distributions.

2 Interactive Simulation

Microsoft Excel has become so powerful that worksheets may be vectorized; formulas are applied to whole arrays instead of individual numbers. Expressions based on random number generators such as the `RAND()` formula, or pre-compiled distributions stored as SIPs, may be used in interactive simulations with an approach we call SIPmath. Excel can calculate output distributions based on tens of thousands of trials, and store them in SIPs, in the time it takes your finger to leave the <Enter> key.

SIPmath works in Microsoft Excel using only the `INDEX()` and Data Table function, so models created with the SIPmath Modeler Tools can be shared with people who do not have the Tools themselves.

3 Credibility of Uncertain Estimates

Many organizations are not comfortable with managers who express uncertainty. As a cure for this problem, the discipline of probability management suggests creating a Chief Probability Officer (CPO) role. The CPO takes responsibility for the integrity and provenance of the SIP Libraries used by an organization. This provides access to centralized, cross platform, and standardized representations of uncertainty that yield repeatable and identical results for all authorized model users.

4 Coherence of relevant distributions

When there are multiple uncertainties associated with each other, it is important to make sure they stay *coherent*. For example, consider two SIPs of 1,000 trials each, one representing the value of your house, and the other representing the payout of your fire insurance policy. If there is one chance in 1,000 that your house will burn down, then one element of the house value SIP is zero (total loss), while the other 999 take on its undamaged value. However, a SIP representing the fire insurance payout is clearly relevant to the house value SIP: it must have one positive element in which it pays up to the coverage terms, while the other 999 entries contain the negative value (your insurance premium). These two SIPs are said to be coherent if the realization of the value loss (fire) for the value SIP lines up (appears on the same trial) with the realization of payment in the insurance payout SIP. A coherent set of SIPs is also known as a Stochastic Library Unit with Relationships Preserved or SLURP.

If the house and fire insurance SIPs were not coherent, it would be like buying insurance on someone else's house. There would be an event when the house has burned down (zero value) but insurance has not paid off, or the insurance pays off but the house doesn't burn down.

2 SIPMATH OPERATING MODES

The SIPmath Modeler Tools is an Excel add-in that can be configured to run in one of the two major modes:

Generate Mode: Generate Mode is much like traditional Monte Carlo simulation, except that the computations take place interactively with up to tens of thousands of trials taking place before your finger leaves the <Enter> key. The **Tools** make use of random number generators, including the built-in Excel `RAND()` function, to derive model inputs.

In addition to generated distributions, historical data may be resampled with a formula like

```
=INDEX(Data,RANDBETWEEN(1,N))
```

where N is the number of data elements.

Generate Mode is a quick way to get a sense of a stochastic process, or to generate a SIP Library for cataloging and later use in Library mode. *Models in this mode are only auditable if the HDR generators are used instead of `RAND()` (see below).*

Generate Mode with HDR Numbers: The tools support a second random number generator that can be used in place of `RAND()`, referred to as *HDR (Hubbard Decision Research)*. Unlike `RAND()`, `HDR()` returns seeded, repeatable pseudo random values. In general terms, HDR numbers have all the characteristics of random numbers except unpredictability; a given pair of inputs will always generate the same output. The HDR algorithm is somewhat experimental, but has passed preliminary tests of randomness and is very convenient for generating interactive and auditable results, using pure Excel formulas. It takes as input, a Variable ID, and PM_Index, the simulation counter that drives the Data Table, and in this way creates a virtual SIP on the fly; it does not require macros to run, because the tools insert the generating formula into the designated input cell.

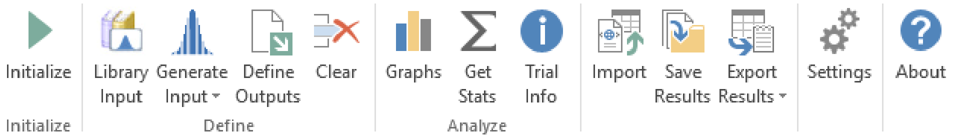
SIP Library Mode: In this mode, the inputs are drawn from a SIP Library that may be stored within the model, or be linked as an external file. This is the mode that provides the cross-platform auditability required by the discipline of probability management. From the perspective of the Excel user it adds a third dimension to the worksheet, representing thousands of potential outcomes. You may scroll through this third dimension either with the View Trials tool or by changing the value of PM_Index on the PMTable sheet.

This is fundamentally different from earlier forms of Monte Carlo simulation. SIPmath can draw its values from prepared libraries of Stochastic Information Packets (SIPs). SIPs hold arrays of pre-computed trial values created by statistical or subject matter experts, using simulation or real-world data. SIPs also maintain data provenance, an audit trail, and reproducible results, and thus are suitable for informing robust business decisions.

As a result, if various SIPmath applications are based on the same set of coherent input SIPs, the results may be networked together. Thus, the results of complex stochastic models may be aggregated to form hierarchical model structures.

3 SIPMATH MODELER TOOLS OVERVIEW

To start working with the SIPMath package, open a new workbook with the `SIPmath Modeler Tools.xlam` add-in installed. The Excel ribbon should include a *SIPmath Modeler Tools* tab. Select that tab and the Modeler Tools ribbon should appear.



The *Initialize* tool is used to identify the stochastic library containing the input SIPs in SIP Library mode, or specifies the number of trials to run in Generate Mode. It sets up the PMTable sheet on which the simulation data will be generated. This tool will also reinitialize existing models and, in Generate Mode, allow the number of trials to be changed.

Once a Library has been Initialized, the *Library Input* tool identifies input cells and links them to one or more SIPs in the Library. This tool may also be used to add a library to a model created in Generate Mode, or to add a library to a model created in Library Mode. Care must be taken when doing this, because the number of trials will be changed to the value in the PM_Trials range in the last loaded Library. This can create problems: For a given model, all the SIPs must have the same number of trials.

The *Generate Input* tool (see section 6.3) selects probability distributions and their parameters to generate one or more input cells. This tool is used primarily in Generate Mode, but may also be used to add inputs to models in Library Mode.

The *Define Outputs* tool identifies the model's output cells with names, and feeds them to the data table. The result is a SIP for each output cell so identified.

The *Clear* tool allows the modeler to remove inputs and outputs from the model. It clears the associated results from both the PMTable sheet and SIPmath Chart Data sheet.

The *Graphs* tool quickly creates histograms and cumulative probability distribution graphs for the output cells, after a simulation has been created. It can also create graphs from pre-existing arrays of data in Excel.

The *Get Stats* tool is useful for returning statistics about multiple input or output SIPs. For example, suppose cell D10 is an output cell named "Portfolio01." To find the standard deviation of Portfolio01, you may enter `=STDEV(Portfolio01)` directly into a cell, or enter `=STDEV(D10)` into the cell and then click on the Get Stats button. This is particularly useful for generating statistics for ranges of output cells. Note that for some statistical functions, such as STDEV, the cell will display an error message until the formula is converted with Get Stats.

The *Trial Info* tool can be used to step through the input SIP values one trial at a time, or to display optional metadata, such as the average or a specified percentile, if any metadata are stored in the SIP.

The *Import* tool (Windows only) allows the modeler to import SIP Library files stored in XML or CSV format, and to store them in Excel format.

The *Save Results* tool allows the modeler to convert a model's simulated data and optional metadata on the PMTable sheet into an Excel format SIP Library.

The *Settings tool* is used to set various parameters. Excel Calculations must be set to Automatic for interactive simulation. This tool sets

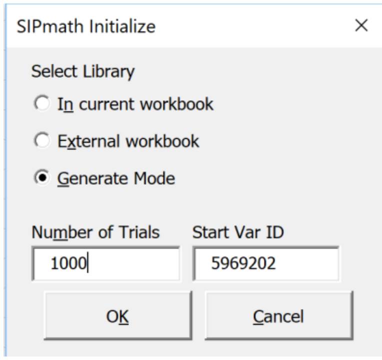
Automatic warnings and Calculation options, sets the default number of Bins for charts, and activates the Crystal Ball or @Risk interface if they are installed.

The *About* tool displays version and copyright information.

Note: It is a good idea to save your work before initializing SIPmath.

Due to the differences in versions and platforms of Excel, some action sequences may make Excel unstable; if this happens, please make sure the problem is repeatable and submit a bug report with a list of software versions and user actions to Support@ProbabilityManagement.org.

4 EXERCISE: MODEL IN GENERATE MODE



SIPmath operates in two main modes: *Generate* and *SIP Library*. If `RAND()` is used in Generate Mode the model will produce different results every time it is run. Use HDR generators for repeatable results.

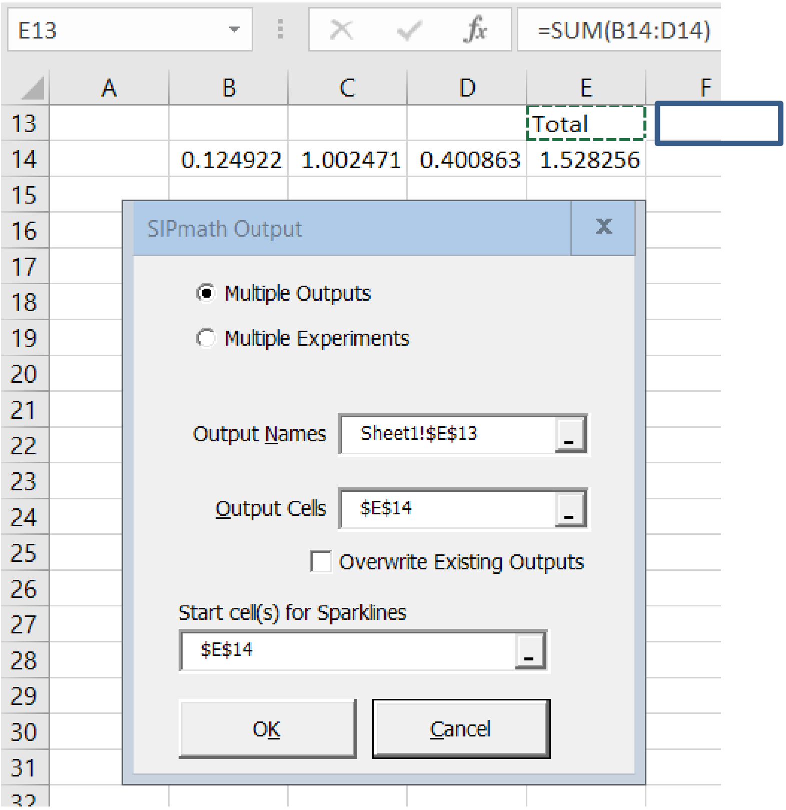
Let's say we are interested in finding the distribution of the sum of three uncertain variables. Such a distribution is easy to simulate with the SIPmath Modeler Tools in just a few steps:

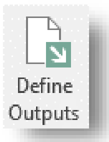
Step 1. Initialize in Generate Mode

Click the *Initialize* tool, choose *Generate Mode*, type the desired number of trials, and click OK. This will set up the PMTable and SIPmath Chart Data sheets in your workbook.

Step 2. Generate Random Inputs. Select the cells you want to use as input cells. For the purpose of this exercise, select B14:D14. Then click the *Generate Input* tool. This drops down a list of distribution choices. Select the distribution you want. For this exercise, select Uniform. This will bring up a dialog that's specific to the parameters needed to calculate your selected distribution (see section 6.3). For this exercise, select **Uniform** and **Rand**.

Step 3. Using Excel formulas create the cell you wish to output. In this case we sum the three random variables in cells B14:D14 and give the result vector the name "Total". Put *=sum(B14:D14)* in cell E14 and put the title “Total” in cell E13.

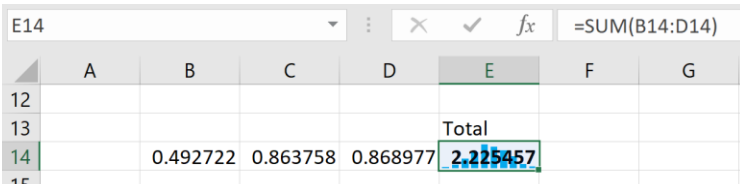




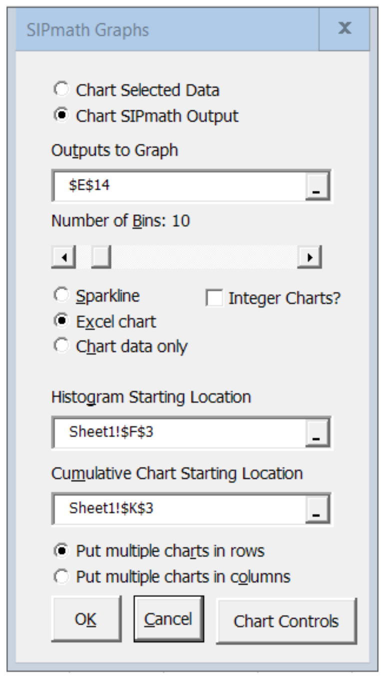
Step 4. With the cursor in output cell (E14) select *Define Outputs* on the ribbon and specify the name (cell E13). If there are multiple output cells (several random variables to study), they can be entered together in the normal Excel fashion (as continuous ranges or lists). Names should follow Excel conventions and include only letters, underscores, periods, and numbers and not overlap with built-in identifiers or cell addresses.

After you click on OK, the Tools will convert the text found in cell E13 into a range name addressing the result vector, and use cell E14 to feed consecutive trials into a data table on the PMTable sheet. It also places a

Sparkline graph in a specified cell E14 since this option was specified in the *Define Outputs* dialog.



At this point, the simulation is complete and the spreadsheet has 1,000 samples (or whatever you defined in step 1) of the sum of the three random variables, stored as a range named "Total" on the PMTable sheet. Now, try changing the parameters of the distributions defined as input (in B14:D14) or the formula in cell E14 to instantly run a new simulation.

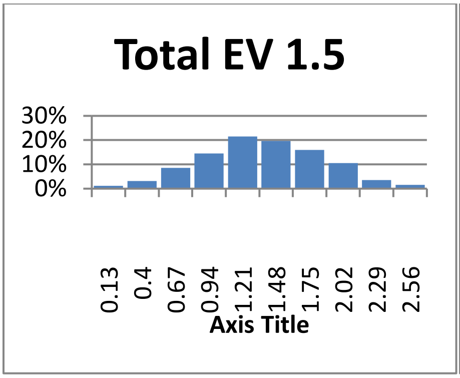


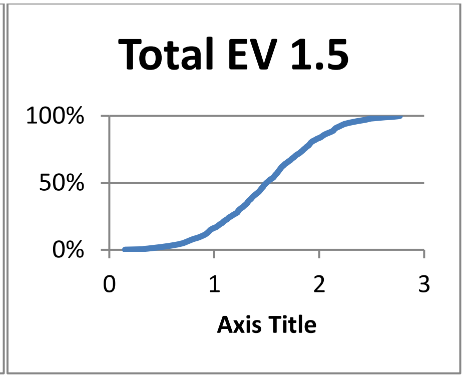
Step 5. Graphs

To create histograms or cumulative graphs as either sparklines or regular Excel graphs, select the simulation output cell (E14) and click the Graphs button.

In this example we will specify both an Excel histogram and cumulative graph in cells F3 and K3 respectively.

The number of decimal places displayed in the graph, the offset of the horizontal axis, and the cumulative curve direction may be specified in the SIPmath Chart Data worksheet.

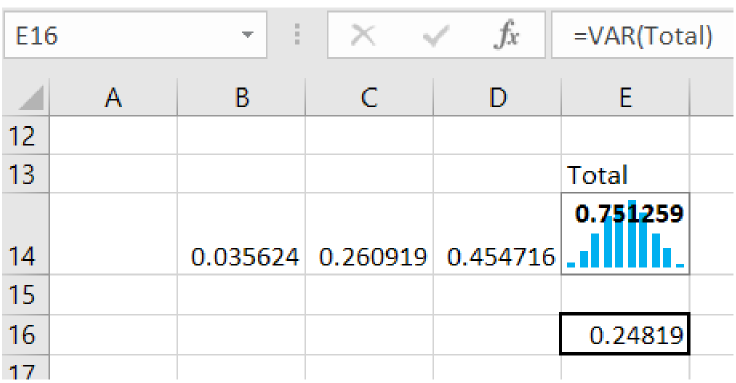




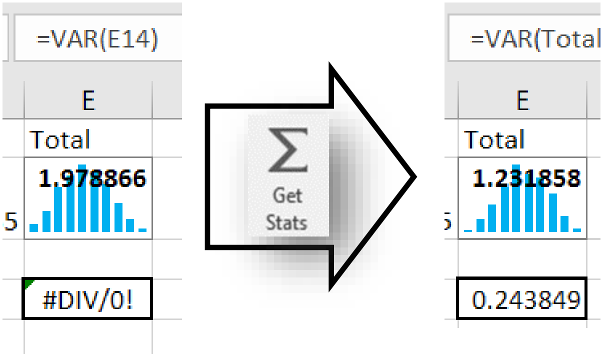
For this exercise, since we're using RAND() to generate the input, the shape will change whenever Excel recalculates.

Step 6. Statistics

We can use the built-in Excel functions like `Var()` to estimate various statistics of the resulting distribution by referring to its range name:

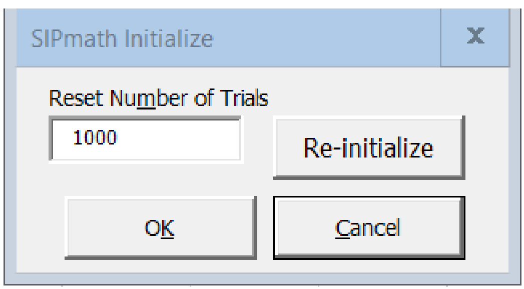


Alternately you may apply the statistical formula to the output cell address and then press the Get Stats button.



This is particularly useful when copying statistical formulas across a range of output cells.

Step7. Re-initializing



If you want to run a larger number of trials, press the *Initialize* button again, enter the new number of trials, and press *OK*. (Pressing *Re-initialize* will delete all defined inputs and outputs and open the main *Initialize* dialog.)

5 EXERCISE: MODEL IN SIP LIBRARY MODE

5.1 INITIALIZE

We will now demonstrate how to build a SIP Library Mode application using Excel and the *SIPmath Modeler Tools*. You can start with an existing model and add distributions to it or, as we describe, start with a blank workbook, and build a model with input and output distributions.

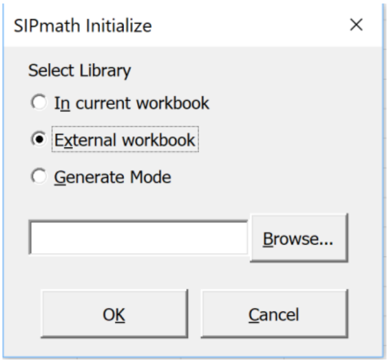
We are going to build a financial portfolio simulation so that we can compare the expected risks and returns for a variety of investment portfolio choices. We will present these as a graph plotting the mean against the variance of the return for each portfolio.

In financial circles, variance is a proxy for risk; better portfolios have higher returns and lower variance. Risk and return are trade-offs, so the best portfolio depends on our risk attitude. That's why we need a picture.

We will simulate five different portfolios and be able to compare their risk/return characteristics while making changes to their mix of assets.

Typically, we would start by developing a deterministic model, then test and debug it before adding stochastic inputs and outputs. In this case, the model is pretty simple, so we will start with the inputs and work our way to the outputs and graphical presentation.

Start by opening a blank workbook in Excel. In our example we renamed "Sheet1" to "Model". Give the file a name of our choice.



The first thing we need is a source of data about assets we can invest in. The data we want are probability distributions of the assets' annual returns. In a real-world application, these would come from a *Stochastic Library* maintained by a financial organization with the appropriate credentials. Since this is just an exercise, we will use the Assets Library 3.1.xlsx workbook that comes with the SIPmath Modeler Tools.

To get a stochastic library, use the *Initialize* tool on the ribbon. Under *Select Library*, select *External Workbook.* You are then prompted to open a stochastic library file.

Instead of linking our model to an external Library file, we have the option to use a stochastic library within the current workbook. This is used for creating self-contained models (not linked to other files) for convenient distribution as a single Excel file. To do this we may open a library, or copy a worksheet from a desired library to the application workbook, using the *Sheet Move* or *Copy* command in Excel, then choose *In current workbook* under *Select Library*.

For this exercise, click on *Browse* and pick the file named *Assets Library 3.1.xlsx* that came with the **Tools**.

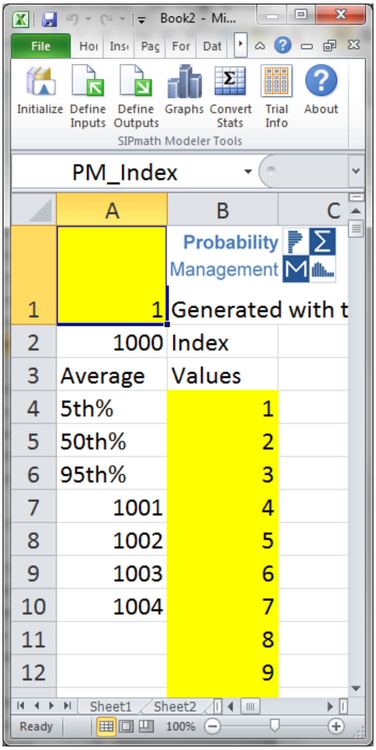
When you click OK, the *Initialize* function will perform several tasks that you need not understand at this stage, but which may be important as you perform more advanced modeling.

1 Opens the library file.

2 Creates a "PMTable" sheet.

3 Names cell A1 of the PMTable sheet PM_Index, and add links to Meta Data such as Average, 5th Percentile, etc., which are stored in the SIP Library.

4 Creates a SIPmath Chart Data sheet.



Cell A1 of the PMTable sheet will be the *Column Input Cell* of the Data Table that will run the simulation.

5.2 INPUT DISTRIBUTIONS

The next task is to set up the input variables, which get linked to the distributions from the library and PM_Index cell.

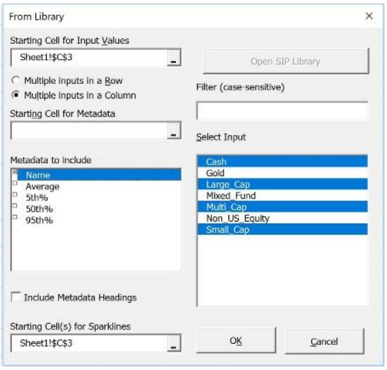
For this we will use the *Library Input* tool. We need to specify where to put the input cells for the distributions we'll be using. In this case we'll put the names (optionally) in a column starting at B3 and the values in a column starting at C3.

On the Model sheet, select cell C3 and click *Library Input* on the SIPmath ribbon. In the dialog that comes up, put the cursor in the *Starting Cell for*

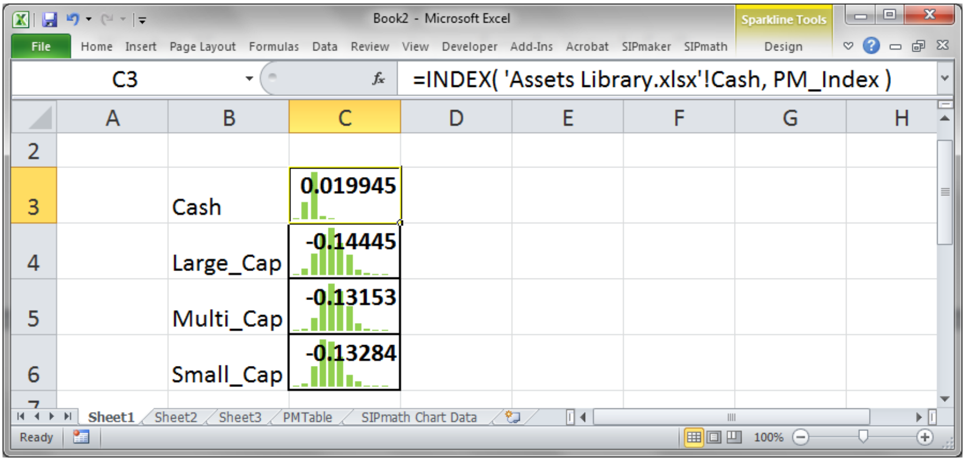
Metadata field, and select cell B3. Under *Metadata to Include*, check the *Name* box. This will put the asset names in column B.

Make sure the *Multiple inputs as Column* radio button is ticked to get the asset data laid out in a column with one row per asset. Finally, select some assets to use in the portfolios. Use the Shift and Ctrl keys in the usual ways to select multiple entries from the list.

Here we have chosen `Cash`, `Large_Cap`, `Multi_Cap` and `Small_Cap`. If we enter *Starting Cell(s) for Sparklines*, sparklines will be created for the selected inputs at this time.

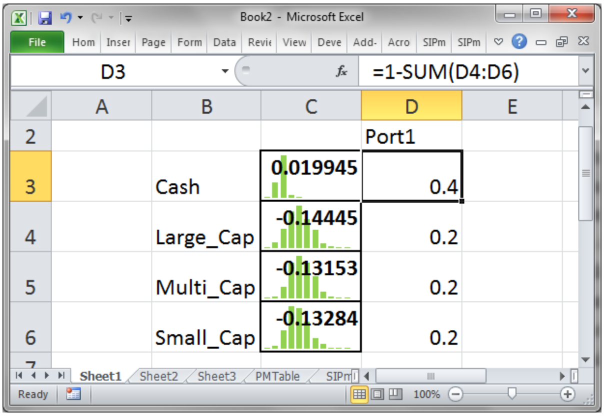


When you click OK, *Define Inputs* will put the names of the assets you chose into a column below B3 and the index formulas that access the SIPs in a column below C3. The numbers in column C are the first trial of each asset distribution. Because we opted to create Sparkline graphs for Inputs and Outputs, they are also created at this time. Note that the Sparklines could also have been placed in any other cells in the worksheet, for example those containing the names.

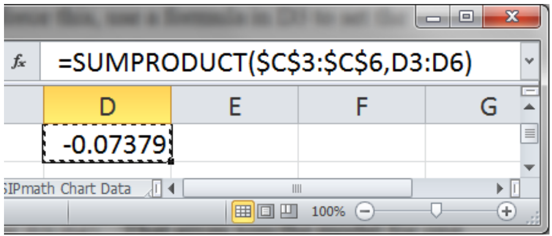


Warning If you use range names in your formulas instead of cell addresses, you'll need to have different names for the model input cells versus the corresponding input SIPs. That is, cell C3 may not be named Cash, because that name is already taken by the Cash SIP range. Consider naming the Output Variable "d_Cash" to distinguish it from the scalar "Cash" if you want both the scalar and the distribution to have range names. The "d_" can signify that the range name is a distribution.

5.3 THE MODEL



The model will consist of a few portfolios and their calculated returns. Define a portfolio in column D with its name in D2 and a weight for each asset. The weights must add up to 1.0 because you want them to be the fraction of total investment allocated to each asset. To enforce this, use a formula in D3 to set the Cash weight so that the portfolio sum is 1.0.



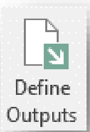
Then, to calculate the return from this portfolio, multiply each asset's return by its weight and add up all the results. Excel's `SUMPRODUCT()` function achieves this in one operation in cell D7.

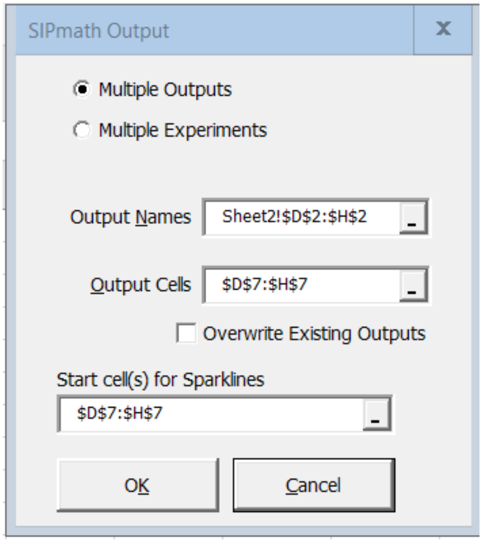
That gives us the model for one portfolio. (Note that $ appear on C3:C6 in the formula but not on D3:D6). This will ensure that when we create multiple portfolios, they will all be pointing at the same inputs. We will make four more portfolios now. Select D2 through to D7 and drag/copy them over to column H. Excel will automatically increment the name "Port1" as it copies and it will adjust the formula for each column. Now, change the portfolio weights so that each portfolio is different. You should have something like this:

		Port1	Port2	Port3	Port4	Port5
Cash	0.019945	1	0.6	0.3	0	0
Large_Cap	-0.14445	0	0.2	0.2	0.4	0
Multi_Cap	-0.13153	0	0.2	0.3	0.4	0
Small_Cap	-0.13284	0	0	0.2	0.2	1
		0.019945	-0.04323	-0.08893	-0.13696	-0.13284

The formulas in column C hook the model to the input distributions. Now you need to hook the model up to the data table and to create the interactive simulation.

5.4 The Output Distributions

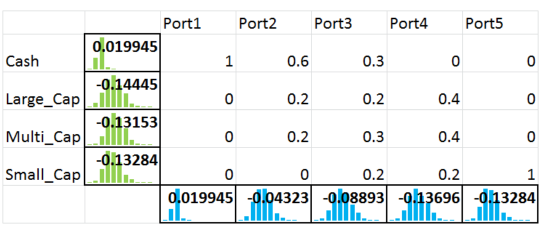




For this we'll use the *Define Outputs* tool on the SIPmath ribbon. Select the cells with the model results, D7:H7, and click *Define Outputs*. In the dialog box that comes up, set the Output Names field to D2:H2 by selecting the cells containing "Port1" to "Port5" while the cursor is in the Names field.

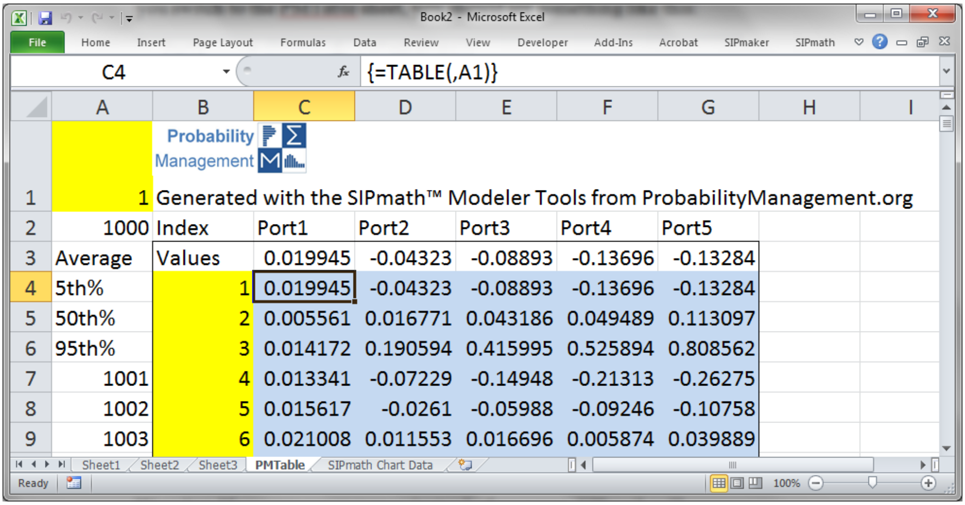
When you click OK, the simulation is set up and run, with Sparklines appearing in the output cells as specified in the *Define Outputs* dialog. At this point if you change any of the weights of the portfolios, the model will instantly run another 1,000 trials.

The *Define Outputs* command creates named ranges in the data table for each portfolio, using the output names you selected.



		Port1	Port2	Port3	Port4	Port5
Cash	0.019945	1	0.6	0.3	0	0
Large_Cap	-0.14445	0	0.2	0.2	0.4	0
Multi_Cap	-0.13153	0	0.2	0.3	0.4	0
Small_Cap	-0.13284	0	0	0.2	0.2	1
		0.019945	-0.04323	-0.08893	-0.13696	-0.13284

Each of these will be an output column holding the results of a 1,000 trial simulation. If you switch to the PMTable sheet, you should see something like this:


C4 {=TABLE(,A1)}

	A	B	C	D	E	F	G
		Probability Management					
1	1	Generated with the SIPmath™ Modeler Tools from ProbabilityManagement.org					
2	1000	Index	Port1	Port2	Port3	Port4	Port5
3	Average	Values	0.019945	-0.04323	-0.08893	-0.13696	-0.13284
4	5th%	1	0.019945	-0.04323	-0.08893	-0.13696	-0.13284
5	50th%	2	0.005561	0.016771	0.043186	0.049489	0.113097
6	95th%	3	0.014172	0.190594	0.415995	0.525894	0.808562
7	1001	4	0.013341	-0.07229	-0.14948	-0.21313	-0.26275
8	1002	5	0.015617	-0.0261	-0.05988	-0.09246	-0.10758
9	1003	6	0.021008	0.011553	0.016696	0.005874	0.039889

The Data Table contains 5000 calculated numbers

5.5 Presenting the Results

As we showed in the Generate Mode exercise, you can create histograms and cumulative graphs of the output cells with the *Graphs* function. In this exercise we will focus on calculating and comparing the variance and average of each portfolio.

We could proceed by entering the following formulas for portfolio 1.

In cell D9, put `=VAR(Port1)`

In cell D10, put `=AVERAGE(Port1)`

But that would just set up portfolio 1, and the results could not be copied to the other portfolios because the range names would not translate.

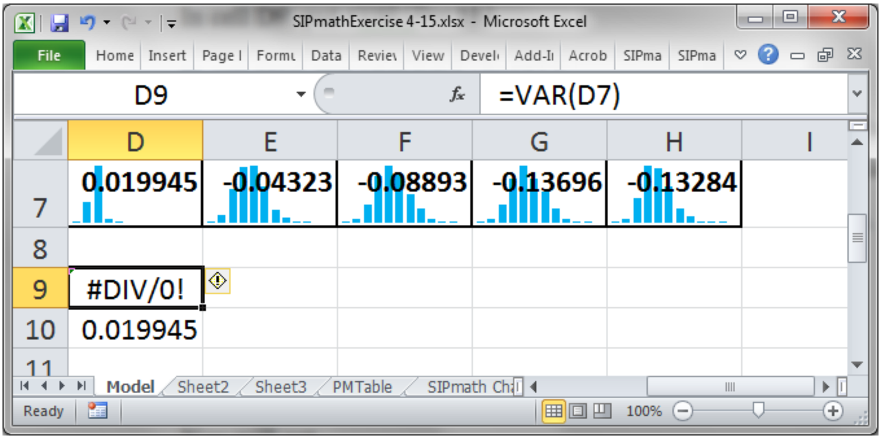
We solve this problem with the *Convert Stats* button as follows.

First:

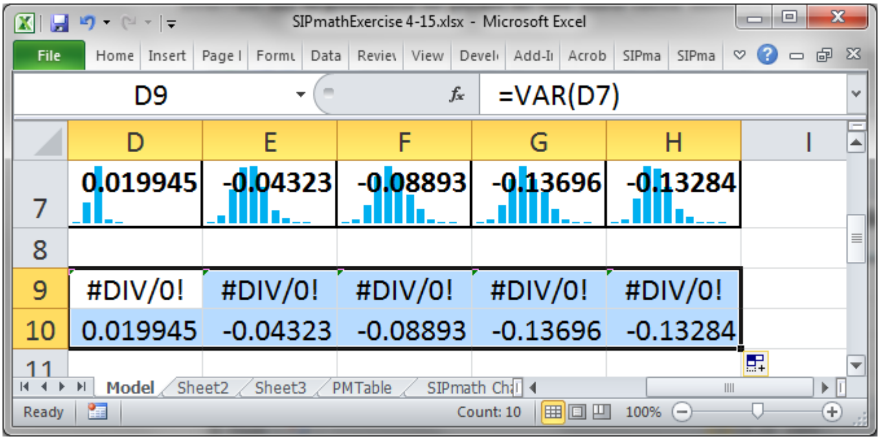
In cell D9, put `=VAR(D7)`

In cell D10, put `=AVERAGE(D7)`

You will get #DIV/0! in cell D9, but don't worry about it, as we will convert all arguments to point to the output ranges in the data table shortly.



Next: copy cells D9 and D10 to E9 through H10.



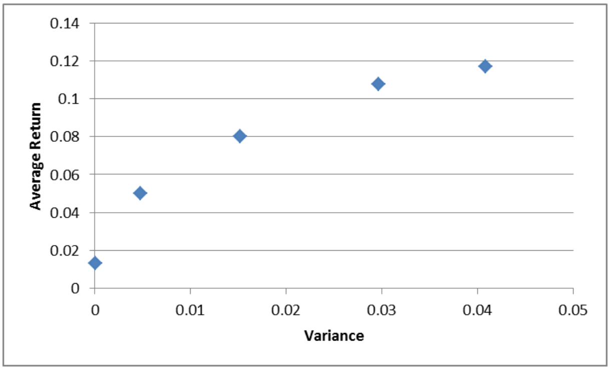
Finally, with all ten of the formulas selected, click on the Get Stats button, and they will be linked to the ranges on the PMTable sheet as shown.

NOTE: DO NOT USE =INDIRECT( ) TO SOLVE THIS PROBLEM

In theory, the `Indirect()` formula could help in copying formulas, but it seriously degrades the speed of the Data Table.

Now we will use the averages and variances in a scatter plot, so that they can be compared more easily.

Select D9:H10, the variances and averages. On the Excel ribbon, click *Insert / Charts / Scatter* and choose the style with no lines. With a little bit of decorating, you should end up with a similar result:

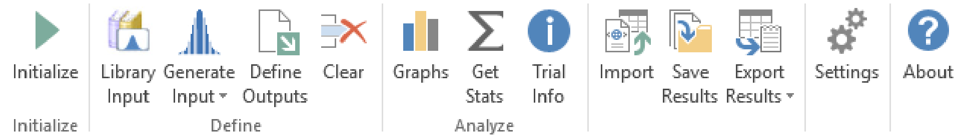


The scatterplot vertical axis is expected return, and the horizontal axis is variance. Up and to the left is better: higher return, lower risk. As you change portfolios, 'better' has a lot to do with how big a return you want and how much risk you can tolerate. Now, as you change the portfolio weights, you will see the dots move around in the graph to reflect the risk and return for those weights.

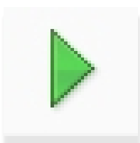
5.6 Running Your model Without the tools

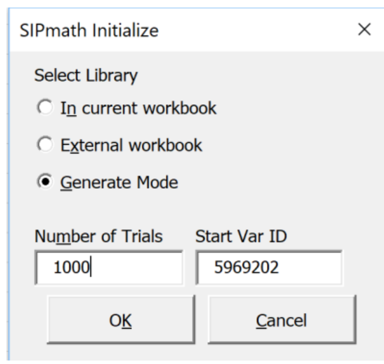
One of the important design goals for the SIPmath Modeler Tools is that the model must work without the SIPmath add-in installed. If you used a library in the current workbook (or created a model in Generate Mode), all you need to do is open your model spreadsheet. SIPmath-generated spreadsheets are easy to share.

6 THE SIPMATH RIBBON DETAILS



6.1 INITIALIZE

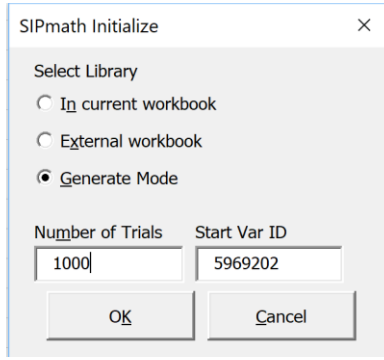




The *Initialize* tool must be used if the plan is to create a model in either SIP Library or Generate Mode.

In SIP Library mode, the *Initialize* tool dialog tells the SIPmath Modeler Tools where to find the distributions that will be used for input. They can be in the model workbook or an external workbook. To specify an external workbook, click “Browse” and use the file dialog to select the desired library file.

The model will be linked to the library file through the use of `INDEX()` formulas, the second argument of which is `PM_Index`, a cell on the PMTable sheet that does the simulation and records the iteration results. If you want to distribute a model widely, you may want to place the library in the same workbook as the model. This can be accomplished by opening an existing library, right clicking on the sheet tab, then using the “Copy” command to copy the library sheet to the desired workbook.



If Generate Mode is selected, initialization provides an input field for the number of trials to simulate. The variable `PM_Trials`, containing the number of simulation trials, is set as a defined name.

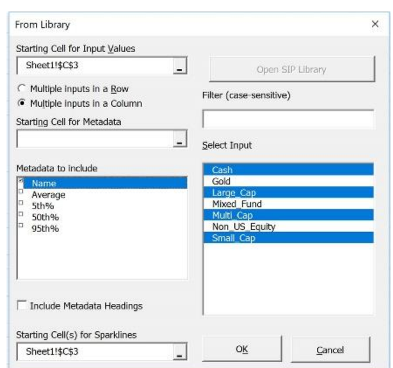
Reinitializing

Clicking the Intialize button on an existing model will give you the option to reinitialize the model or, in Generate Mode, change the number of trials. Changing the number of trials impacts the size of the PMTable and the size of the model. The number of trials can be increased or decreased.

6.2 Library Inputs

For use in the SIP Library mode only.

Use the *Library Input* tool to tell the SIPmath Modeler Tools which distributions to use from the library chosen during initialization, and which cells the model will use to get input trial values.



Use the *Select Input* dialog to select the SIPs you want linked to your model.

The Filter box can facilitate your search when you have a large number of variables. Use the shift and control keys in the usual way to select multiple SIPs. Click OK when you're done.

You may optionally specify cells into which to put the SIP names and other metadata (e.g. Average). Select the metadata you want included in the output in the `Metadata to Include` box. Usually, this will include the *Name* value to identify each SIP.

The values in the cells are from the SIPs in the library, usually the first trial. The `PM_Index` default value is 1, but you may view any value with the Trial Info button.

In the picture below, if you look at the formulas under the values, you'll see that the tool has set them to something like

```
=INDEX( Library01.xlsx!Distribution3, PM_Index )
```

Where `Library01.xlsx` is the library file name, `Distribution3` is the distribution name and `PM_Index` selects the element in the distribution's range.

Unless you cleared the Starting Cell for Sparklines field, you will see that a light green sparkline histogram has been placed in the cell with the `INDEX()` formula, and the cell has a black border drawn around it. An example is shown below.

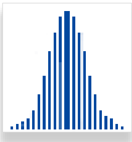
Cash	Gold	Large_Cap	Mixed_Fund	Multi_Cap	Non_US_Equity	Small_Cap
0.019945362	-0.001127084	-0.144446604	-0.144579575	-0.131532497	-0.157626654	-0.132836872

This example is generated using the "*Multiple Inputs as Row*" setting in the "*From Library*" dialog.

Using the "*Multiple Inputs as Column*" setting produces a table like this:

Cash	0.019945
Large_Cap	-0.14445
Multi_Cap	-0.13153
Small_Cap	-0.13284

6.3 GENERATE INPUT



The famous mathematician and father of the modern computer, John von Neumann, said *"Anyone who attempts to generate random numbers by deterministic means is, of course, living in a state of sin."* This did not mean that he thought it could not be done; in fact, his first computer launched the field of Monte Carlo simulation. However, generating such numbers is not straightforward and, some 70 years after the advent of Monte Carlo simulation, this is still an active research area. As described below, the SIPmath Tools provide internal random number generation, as well as the option to use SIPs of random numbers generated from any external source.

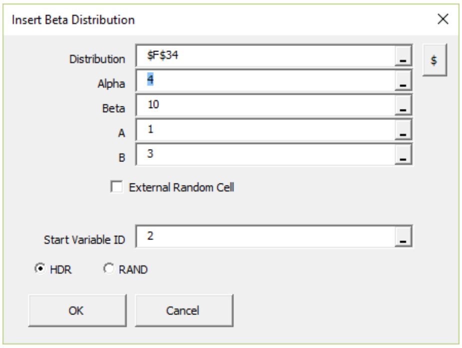
The distributions currently supported are listed below.

1. Uniform
2. Beta
3. Binomial
4. Chi-Square
5. Discrete
6. Exponential
7. F
8. Gamma
9. Lognormal
10. Metalog
11. Myerson
12. Normal
13. Poisson
14. Resample
15. Triangular
16. T
17. Weibull
18. Correlated Uniform
19. Correlated Normal

All the distribution functions in the SIPmath Modeler Tools are driven by Uniform random numbers between 0 and 1. These are referred to as U(0,1) variables. The Tools offer three options for U(0,1)s to drive your simulations.

1) The Native Excel `RAND()` function.
   This computes results the fastest, but does not provide replicable results.
2) The HDR Generator
   A Random Access Pseudo-Random Number Generator from Hubbard Decision Research (the HDR Generator). This is an evolving family of simple generators that use pure Excel formulas to provide repeatable results in an interactive simulation environment. Each instance of an HDR U(0,1) must have a unique Variable ID to avoid correlated results. The version of the HDR used in the tools has passed some basic random number tests, but should not be considered to be a "pedigreed" generator. However, this is the suggested approach for convenient everyday modeling. Because the modular nature of SIPmath, you may always go back and replace your U(0,1)s at a later date.
3) Your own favorite generator.
   By creating SIP Libraries from of your favorite generator, you may use U(0,1)s of any sort. For example, you may use any pedigreed random number generator, or even numbers generated from atmospheric noise by Random.org. Such SIPs are entered in your model through the Library Input command, which creates an index into the specified SIP.

6.3.1 SAMPLE DISTRIBUTION DIALOG:



Distribution: Enter the cell or range of cells where you want the formula for the chosen distribution to appear. A range must be either a single row or a single column.

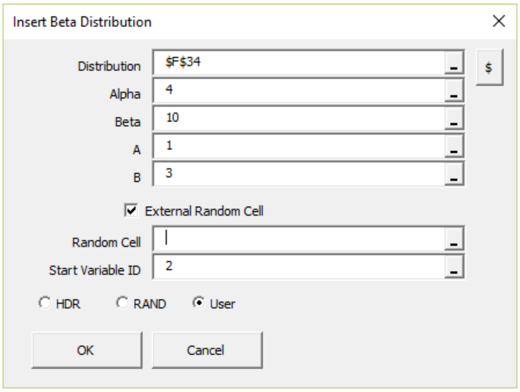
$: The $ button changes the address of an input parameter from absolute to relative. Click in the box of the parameter you want to change and then press $ to change the address. The $ button will cycle through the various combinations of holding the Row constant, Column constant or both Row and Column Constant. If you don't use the $ button, then SIPmath will ignore $s and make a best guess whether to use absolute addresses when entering the distribution formula; usually SIPmath will use relative addresses, unless the input parameter is a single cell and the distribution address is two or more cells.

Alpha, **Beta**, **A, B**: These parameters determine the pattern of results computed by the formula for the Beta distribution. They may be entered either as constant values, single cells, or a range of cells. In the latter case, the range must be the same size and shape as the Distribution range. Note that every distribution has its own set of input values.

External Random Cell: check this box if you want to either correlate your distribution with other distributions or specify your own U(0,1)s to drive the Distribution formula. The External cell (or range of cells) will be used as random inputs to the formula. For instance, a Normal distribution in cell A1 might be computed as

`=NORMINV(A2, 0, 1)`, where `A2` would be the External Random Cell containing a random uniform.

Leave the box unchecked if you want the Distribution formula to use a random value computed inside the formula (RAND or HDR only). For instance, a Normal distribution in cell A1 might be computed as `=NORMINV(RAND(),0,1)`



Checking the box shows the Random Cell field below; unchecking it will hide the Random Cell field.

Random Cell: If External Random Cell is checked, enter a cell or range of cells that will contain the random values. If this is a range, it must have the same size and shape as the Distribution range.

HDR: Select this option to generate pseudo-random values with the Hubbard Decision Research generator. This causes the sequence of pseudo-random values to be repeated with the same inputs every time a model is calculated, allowing for repeatable and auditable results. External Random Cells will be filled with the HDR formula for generating pseudo-random values.

Start Variable ID: This is required when using the HDR generator. Enter a constant, a cell, or a range of cells that has a value to identify the variable ID for the HDR generator. If you use a range, it must have the same size and shape as the Distribution range. This input can also be thought of as the “seed” for the random distribution. Having a unique

Start Variable ID or seed is important to ensure the independence of the random variables.

RAND: Select this option to generate pseudo-random values with Excel's `RAND()` function. External Random Cells will be filled with `=RAND()` formulas.

User: Select this option to use your own source of external random values. If this option is chosen, the Random Cells specified will be left unchanged, so you can put whatever formula you prefer in the Random Cells. For example, by using the Library Input command to specify data from a known generator.

OK: Press this button to fill the Distribution range with formulas to compute the chosen distribution.

Cancel: Press this button to close the dialog without doing anything.

The example above shows the dialog for the Beta distribution. The other dialogs, except for the Myerson, Poisson, Triangular, and Correlated distributions (marked by [1]) are very similar, differing only in that the input values **Alpha**, **Beta**, **A** and **B** will have different names, and not all dialogs have 4 inputs.

Distribution	Input parameters			
Uniform	<none>			
Beta	Alpha	Beta	A	B
Binomial	Number of Trials	Chance of Success		

[1] These distributions require an External Random cell to hold the U(0,1).

Distribution	Input parameters			
Chi-Square	Degrees of Freedom			
Discrete	Values	Probabilities	Cumulative Probabilities	
Exponential	Alpha			
F	Degrees Free 1	Degrees Free 2		
Gamma	Alpha	Beta		
Lognormal	50^{th} percentile	Top percentile rank	Top percentile value	
Metalog[1,2]	Low, 50^{th}, and high percentiles	Low-percentile probability	Bounds (if any)	
Myerson [1]	10^{th} percentile	50^{th} percentile	90^{th} percentile	
Normal	Mean	Standard Deviation		
Poisson [1]	Lambda			
Resample	Data to Resample			
Triangular [1]	Minimum	Most Likely	Maximum	
T	Alpha			
Weibull	Lambda	K		
Correlated Normal [1]	Mean	Covariance Matrix		
Correlated Uniform [1]	Correlation Matrix			

1: These distributions require an external random cell(s) to hold the U(0,1)

2: These are SPT Metalogs. See below for more information.

6.3.2 BRIEF DISTRIBUTION DESCRIPTIONS

The descriptions below are indications of where these distributions may be applied. We strongly recommend consulting appropriate published references before applying any of these in practice.[2]

Uniform: Every value between 0 and 1 is equally likely. Drives all other generators.

Beta: Used to model random variables limited to an interval. For example, the percentage of defective parts in a batch (between 0 and 100%).

Binomial: Gives the number of successes in the given Number of Trials with the given Chance of Success. It can give any whole number from 0 to ∞. Note that, in Excel, ∞ is about 10^{15}.

Chi-Square: The chi-square distribution with K degrees of freedom is the distribution of the sum of the squares of K independent standard normal random variables. It can give any value from –∞ to ∞.

Discrete: Gives a distribution of user-specified values which occur with user-specified probabilities. The Cumulative Probabilities field must be a blank range with the same size and shape as the Probabilities field; it will be filled in by SIPmath with the correct formulas to compute this distribution.

Exponential: Gives the time between events, when the average time is alpha. Can give any value greater than or equal to 0

F: Used to test whether two data sets have the same degree of diversity. If U and V are chi-square distributions with degrees of freedom given by

---

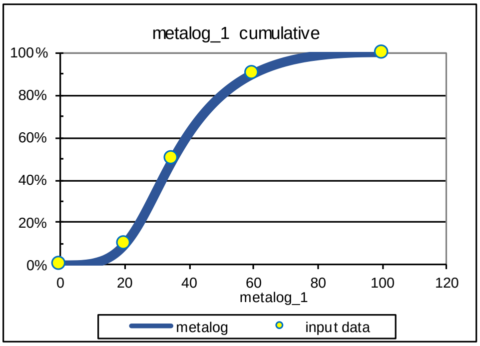
[2] These distributions require an External Random cell to hold the U(0,1).

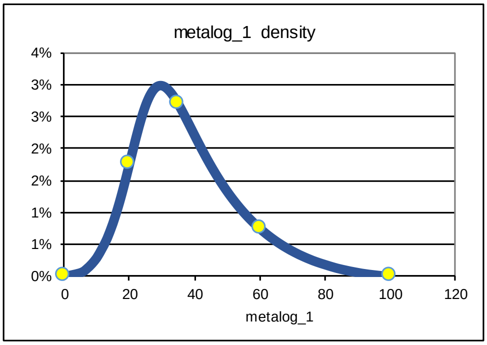
`Degrees Free 1` and `Degrees Free 2`, then F `is (U/Degrees Free 1)/(V/Degrees Free 2)`. Can give any value from 0 to ∞.

Gamma: Gives the waiting time until alpha events have happened, when the average time between events is Beta. (Thus when Alpha=1, this is the same as the exponential distribution.) Alpha and Beta are (possibly fractional) values > 0.

Lognormal: The log of this distribution is a normal distribution. The top percentile rank can be between 0.6 and 0.99. If this rank is (say) 0.75, then 75% of the generated values will fall below the top percentile value. Can give any value from 0 to ∞.

Metalog: Distribution family, invented by Tom Keelin, is comprised of unbounded, semi-bounded, and bounded distributions, each of which has nearly unlimited shape flexibility. All metalogs are quantile-parameterized, meaning that they are parameterized by (probability, quantile) points on the cumulative. Known lower and/or upper bounds are optional parameters. If no bounds are specified, the metalog is unbounded. All metalogs have closed-form equations for both cumulative and density functions. The graphs below show a bounded metalog parameterized by cumulative-points (0.1, 20), (0.5,35), and (0.9,60), along with lower and upper bounds of 0 and 100 respectively.





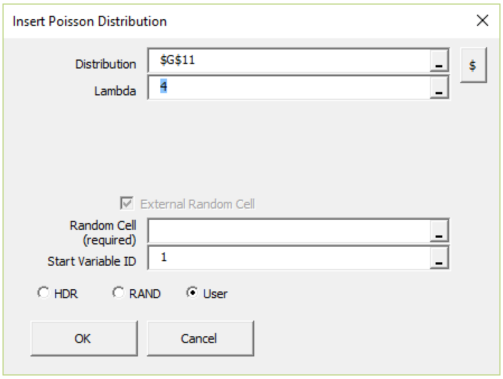
When a metalog distribution is requested, SIPmath loads a metalog template at the location specified in the SIPmath dialog box. The user may choose between vertical and horizontal formats and whether to include the above charts. With the above charts, the vertical and horizontal templates contain 152 rows/3 columns and 132 rows/11 columns respectively. Without charts, they contain 14 rows/3 columns and 4 rows/11 columns respectively. Please ensure that there are enough empty rows and columns available for the template to load without impacting any other part of the current model. Once loaded, the user may choose to cut/paste any template cells to different locations. For example, if charts are included, the bottom 117 rows contain only chart data, which the user may wish to cut/paste to separate worksheet.

The metalog implementation in SIPmath currently encompasses the SPT (symmetric percentile triplet, e.g. 10th, 50th, and 90th percentile) metalogs. For a tutorial on using SPT metalogs in SIPmath, see http://www.metalogdistributions.com/downloads/sipmath.html . For more general information, including on SPT metalogs, on general metalogs (even more flexible metalogs parameterized by an unlimited number of data points), equations, publications, and downloadable Excel workbooks, see www.metalogdistributions.com.

Myerson: Distribution, invented by Roger Myerson, is a generalization of both the Normal and Lognormal distributions, and can imitate either. If the 10th, 50th and 90th percentiles are evenly spaced (e.g. 3, 5, 7) then the distribution is a Normal distribution with the specified percentiles. If the percentiles increase geometrically (e.g. 1, 4, 16) then the distribution is a Lognormal. Other patterns will give intermediates between Normal and Lognormal.

Normal: Also called the bell shaped curve, or Gaussian distribution. It has the mean and standard deviation specified by the inputs. Can give any value from -∞ to ∞.

Poisson: The number of events (such as arrivals of customers) in a given time interval, given an average arrival rate, Lambda.



Resample: Chooses a random value (i.e. re-samples data with replacement) from a row or column of cells specified by the user.

Triangular: Often used when a distribution must be guessed at. The minimum, most likely, and maximum values are as specified in the inputs.

T: Also called Student's t-distribution. Similar to a Normal distribution, but more likely to have extreme values. The larger Alpha is, the closer this is to a Normal distribution.

Weibull: This can be interpreted as the time-to-failure of a process. If k > 1, the rate of failure increases with time. If k < 1, it decreases. K must be > 0. Lambda characterizes the average time to failure; about 63.2% of the values will be less than Lambda.

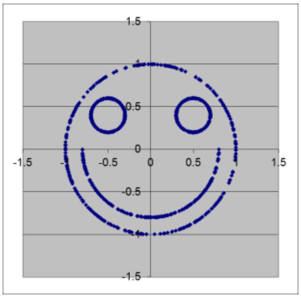
Correlated distributions: Creates sets of distributions which are correlated as specified by the user. Either uniform or normal distributions can be correlated. More detail below.

6.3.3 DISTRIBUTIONS REQUIRING EXTERNAL RANDOM CELLS

The Myerson, Poisson, Triangular, Correlated Uniform and Correlated Normal distribution formulas require the use of a Random Cell. Those dialogs have no External Random Cell checkbox, merely an External Random Cell field.

6.3.4 CORRELATED DISTRIBUTIONS

It is often desirable to generate interrelated, or statistically dependent random variables. In classical statistics this was generally described in terms of ***correlation***, which measures the degree to which the scatter plot of two variables lie along a straight line. Like many concepts from classical statistics, correlation is most applicable to normal distributions. But for many sorts of variables correlation is misleading. For example, the X and Y values of the scatter plot on the right have a correlation of zero, yet there is an obvious interrelationship. A major benefit of SLURPs is that they preserve such patterns within the data, so the interrelationships are handled automatically. This captures, a much broader set of relationships than correlation, and may be accomplished through the Resample distribution.

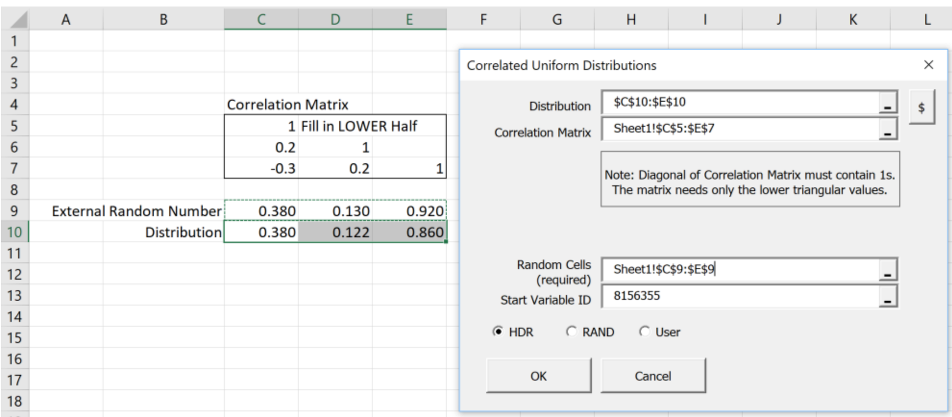


For those trained in statistics, and familiar with Pearson's and Spearman's definitions of correlation, the Tools provide two types of correlated distributions. These generate sets of Uniform or Normal pseudo-random values which have user-specified correlations. When these distributions are created, SIPmath adds a new worksheet to your model named SIPmath Cholesky Decomposition. This sheet contains intermediate formulas that are required to compute the correlated values.

6.3.4.1 Correlated Uniform Distribution

The primary reason to create correlated uniforms is to drive other distributions through external random cells. For example, one could generate Poisson and Triangular variables that were correlated to one another.

Distribution: Enter the range of cells (a row or column) where you want the correlated random values to appear.



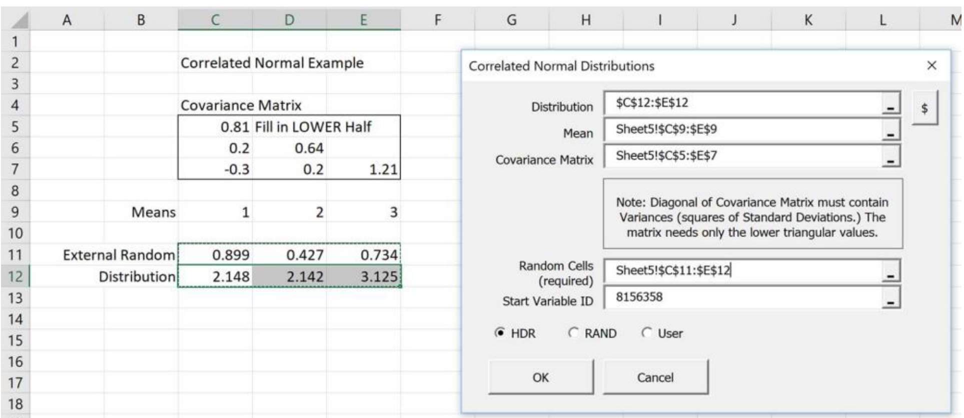
Correlation Matrix is a square range of cells containing the correlations. The main diagonal of this matrix must contain the value 1. The correlations should be put below the main diagonal; values above the diagonal will be ignored. The range must have as many cells in its sides as there are cells in the Distribution range. For instance, if the Distribution range is four cells, the Correlation matrix must be a 4x4 range.

NOTE: The correlation entered for the uniform is approximately the Spearman's Rank correlation, which is quite close to the Pearson's correlation.

External Random Cells is a required range that is parallel to the Distribution range.

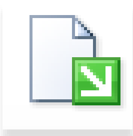
6.3.4.2 Correlated Normal Distribution

Mean: Enter the mean value of the normal distributions here, either as a single constant, or as a cell or cell range. If this is a range, it must have the same size and shape as the Distribution range.

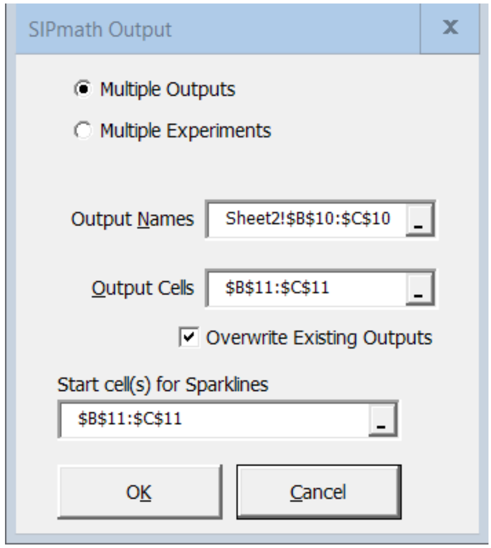


Covariance Matrix is like the correlation matrix for correlated uniform distributions, except you use covariances and variance instead of correlations. Note, the variances are the squares of the standard deviations of the normals.

6.4 Define Outputs



Define Outputs is actually two different dialogs: Multiple Outputs, which is displayed in the examples, or Multiple Experiments, for "What If", analysis. The dialogs work the same way in both SIP Library and Generate Mode, but the Experiments setting is less informative in Generate Mode.



Warning: If Multiple Experiments is used with an existing Multiple Outputs model, the current PMTable state is overwritten by the analysis output. Thus, any model calculation or graph that depends on the earlier PMTable output will have errors. You will need to save the model as a different version when using Multiple Experiments.

6.4.1 MULTIPLE OUTPUTS

Multiple Outputs runs a single set of trials through one or more output cells, each of which will correspond to a column in the PMTable sheet. This dialog lets you specify names for (a) the simulation result distributions and (b) the cells containing the formulas to be simulated.

If you specify a Start Cell(s) for Sparklines range in the dialog, then the cells in that range will have a light blue sparkline histogram and a black border.

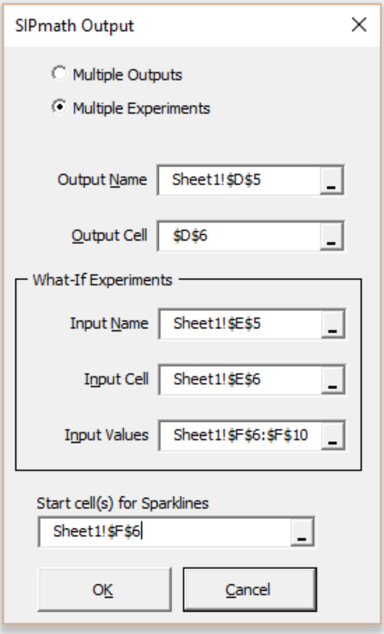
Define Outputs will, for each output cell, put a named column range in PMTable, using the associated name that you specify. These names, referring to the output distributions, can be used when creating graphs and summary statistics, using native Excel formulas anywhere in your model.

You may specify ranges of output cells. If you check “Overwrite Existing Outputs,” a new table will be created, overwriting any data currently in place. Otherwise it will append your selections to the right of the current output columns. This means you don’t have to specify all your outputs in one shot. Also, the output cells don’t have to be contiguous.

6.4.2 MULTIPLE EXPERIMENTS: WHAT-IF ANALYSIS

You get the second dialog by selecting **Multiple Experiments**. The purpose of this tool is to find out how changes in one particular value assumption impact the model results. The formal name for this is sensitivity analysis.

In this context, an ‘experiment’ is a single run of the model, and the ‘What If’ analysis consists of some number of experiments, driven by an input variable that changes value for each experiment.



The ‘Input Values’ field holds the list of values that you will use to drive the analysis. The model will be run once for each of these values. ‘Input Name’ and ‘Input Cell’ address the cell that you will use to pass successive values from the Input Values list to the model.

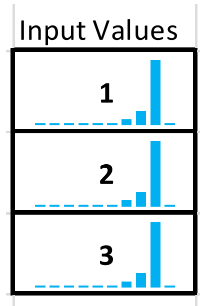
For each model run, the ‘Output Cell’ trial values will be appended to PMTable, using a naming convention that prevents collisions and makes it easy to refer to analysis results for populating reports and graphs. Each PMTable column is given a range name taken from the names assigned in the SIPmath Output dialog. That is:

`OutputName_InputName_InputValue.`

For example, if the output name was “Cost”, the input name was “Stock”, and the input stocking levels were 10,15 and 20, then the column names in the PMTable sheet would be: Cost_Stock_10, Cost_Stock_15, and Cost_Stock_20.

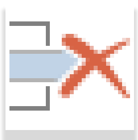
Setting up the analysis is the same as setting up for any model, except that you can define only one output, and you need to define the input list to drive the analysis.

Note that the experiment Input Values *are the values you want to test, not the same as the input distributions you specified in the* Define Inputs *dialog, which will still drive each model run. The "What-If" values will drive the multiple experiments.*



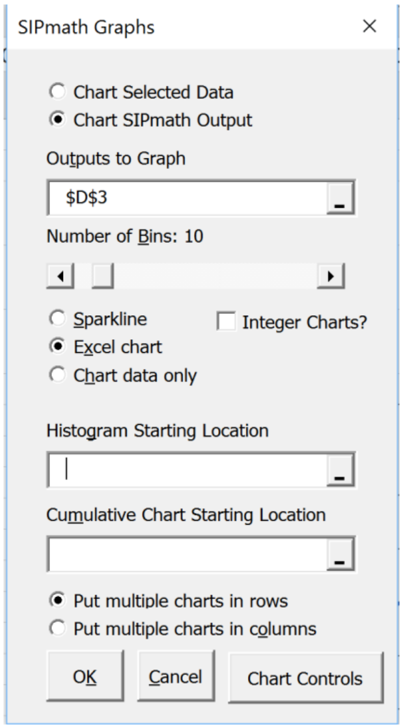
If you specify a Start Cell(s) for Sparklines range in the dialog, then the cells in that range will have a light blue sparkline histogram and a black border.

6.5 Clear

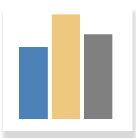


Clicking the *Clear* tool once removes input and output formulas and formatting from the selected cells. It also removes any related chart data from the SIPmath Chart Data worksheet. With an output variable selected, clicking the clear button once will remove the output variable from the PMTable worksheet. This will reduce the size of the model and speed up the calculation time. But graphs can no longer be generated. Clicking the Clear tool a second time will delete the formula. If the cell selected is a random variable but not an output then the cell contents are deleted. This can also be accomplished by using the keyboard delete key.

Clear also provides a simple way to remove the relatively slowly-recalculating formulas that compute data for the SIPmath cumulative probability (CDF) chart. If a SIPmath CDF chart for an output is selected when Clear is pressed, the chart and the formulas for computing the cumulative chart will be removed. Histogram charts and their formulas will be left alone.



6.6 Graphs



The *Graphs* button lets you create histograms and cumulative graphs from either *SIPmath Output Cell(s)* or pre-existing *Selected Data* ranges.

Specify a location to put the Histogram(s) and/or a separate location for the cumulative. If only the Histogram field has a cell reference, then the data for the CDF will not be generated. If the histogram cell reference is not present, then the histogram graph will not be generated but the data will be calculated on SIPmath Chart Data. The only way to not generate the histogram data is to not request Sparklines when defining the output variable and not specifying either a histogram and CDF graph.

For pre-existing data, select a range of values to be charted. You may either: 1) select that range before pressing the *Graphs* button, 2) type the range specification or name of the set in the Data field manually, or 3) click in the Data field and select the range to be charted.

The Number of Bins slider sets the number of histogram bars. (Cumulative (or CDF) charts are always charted with 105 points using Excel's `PERCENTILE()` function.)

If you check the Integer Charts checkbox, *Graphs* will assume that the data to be charted consists of integers (whole numbers only), such as 1, 2, 87, and −452. The number of bins will be automatically determined by *Graphs*, and the slider control will be greyed out when you check Integer Charts. The histogram will have one bar for each whole number between the smallest and largest whole number in the data, up to 100 bars. The bars will be labeled as whole numbers.

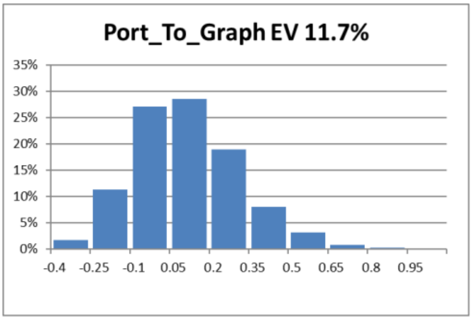
Note This option is not recommended if the range of data from smallest to largest value is more than 100.

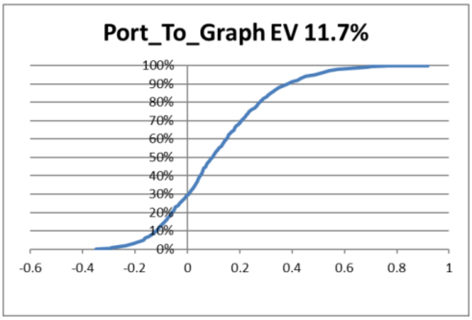
If you select Sparkline[3] or Excel Chart the dialog will expand to ask you which graphs you want and where you'd like them put. Note that sparklines are not available for CDF charts.

If you are charting more than one set of data, additional charts can be placed in a row to the right of the first chart, or in a column below the first chart. You can select which with the radio buttons at the bottom.

Pressing OK will create the chart data and Sparkline or Excel chart, if any, in the specified location. Here's what histogram and cumulative distributions might look like once created:

[3] In Excel 2010 or later.





The the series label is the output name followed by the expected value of the distribution. On the SIPmath Chart Data there is an option to change the CDF to an Exceedance chart. An Exceedance chart plots the probability of being a certain value or more, whereas the CDF graph plots the probability of being that value or less.

CDF charts use Excel's `PERCENTILE()` function, which can be slow when used with a large number of trials. If you find the CDF chart is slowing down Excel's response time, you can remove the CDF chart and the `PERCENTILE()` functions by selecting the CDF chart and pressing the SIPmath *Clear* button.

6.6.1 SIPMATH CHART DATA

	A	B
1		d_uniform1
2	Expected Value	0.5
3	Series Name	d_uniform1 EV 0.5
4	Integer chart?	FALSE
5	Loss Exceedance	FALSE
6	Number of Bins	10
7	Label position	10
8	Bin Width	0.1
9	Decimals	2
10	Scale	1
11	Use Axis Title?	FALSE
12	Axis Title	
13	Min	0
14	Bin Range	0.1
15	Frequency	0.2
16	Labels	0.3
17	Cumulative	0.4
18		0.5
19		0.6

When you initialize a model or click Ok to close the *Graphs* dialogue box under the Select Data option, a new worksheet will appear titled "SIPmath Chart Data." Here is a screenshot of example chart data.

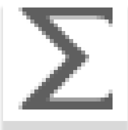
The SIPmath Chart Data will store information on every Output Variable. Each variable will occupy one column. The rows contain different types of data. The color of the data indicates the type of data. Row 1 will have the Output Variable name and Row 3 will specify the Series Name that appears in either the histogram or CDF graph.

The worksheet uses the outline function to facilitate displaying only information desired. For example, to see only the CDF data, click on the "1" outline and then expand the last set of data for the Cumulative. When a Sparkline graph, histogram or CDF graph is requested the sheet is set back to outline "2".

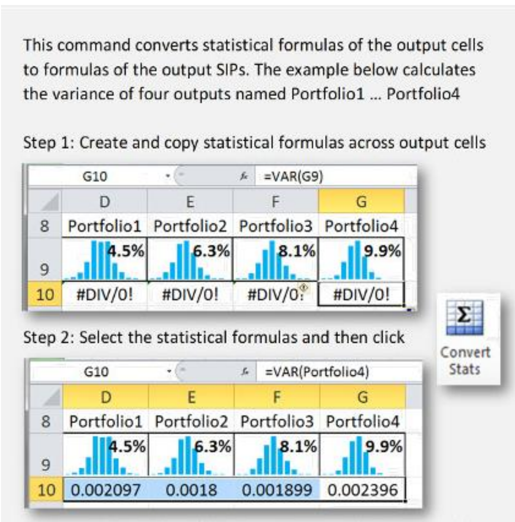
Data in this worksheet can be manually edited to change the way your output histogram on the other sheet looks. Bins can be manually edited, an offset can be applied, the number of decimal places allowed for class widths can be changed, and the histogram bar sizes (frequencies) can be altered.

Selecting "Chart Data Only" in the *Graphs* box will make the chart data table without creating the chart. You can then use this to create your own charts using Excel's *Insert Chart* tool. The ranges for the bins are suggestively named (e.g. Profit_bins, Profit_freq, assuming your data had a heading of “Profit”).

6.7 Get Stats

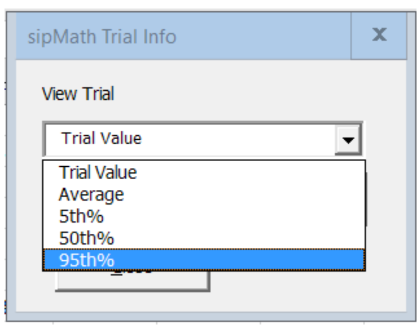


The *Get Stats* tool is designed to simplify the application of statistical formulas such as `Var()`, `Percentile()`, etc. to model outputs. Output variables are essentially SIPs (arrays of observed realizations), and must be referred to by the range names that were automatically created when they were defined (see *Define Outputs* above). If you forget how to use this button, click on any cell without a formula and click the button. This explanation will appear:



6.8 Trial Info

The *Trial Info* tool lets you step through the simulation trials one at a time.



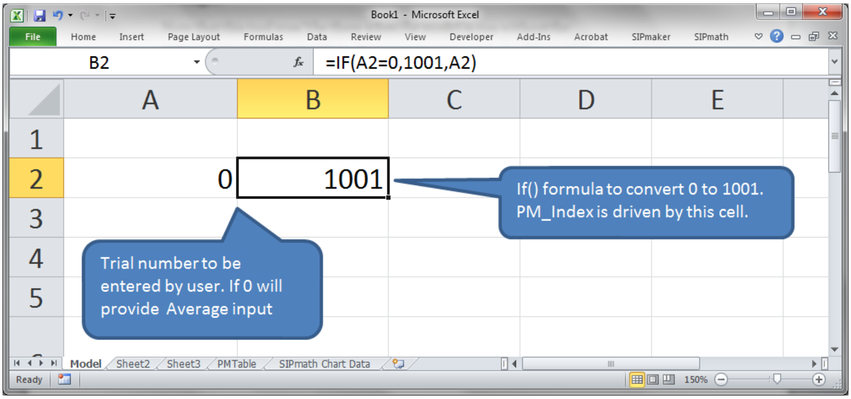
Wherever you set the trial number, you can see the corresponding trial values in the input cells and the corresponding result values in the output cells. *Trial*

Info is actually incrementing the number in the PM_Index range on the PMTable sheet; this can be done manually.

If the SIP Library contains metadata extensions to the distributions, you can also use those to instantiate model parameters. Keep in mind that the choice is only for your inputs. For example, if you choose "Average", you'll see the output values based on average inputs, NOT the average of the output values.

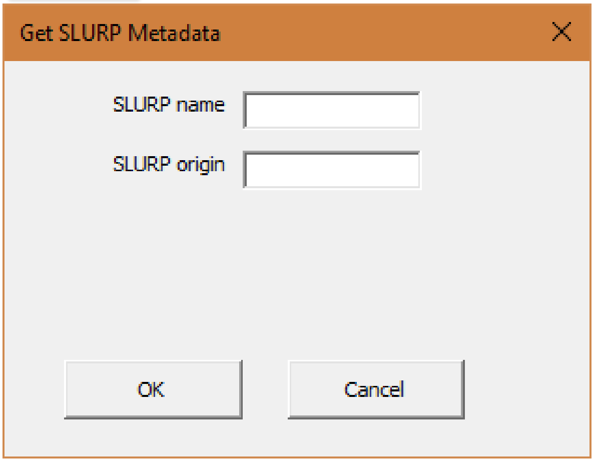
If your spreadsheet model is non-linear, you will find an interesting difference between the *output* given *average* inputs and the *average output given input distributions*. This is due to Jensen's Inequality, commonly called the Flaw of Averages.

Note that this tool won't be there when the model is run without the SIPmath Modeler Tools add-in. If you want your user to be able to step through the trials, you'll need to let them change the PM_Index range and include a control to access the metadata at the ends of the SIPs if desired. Suppose, for example, that the "Average" input is stored at location 1001. Then it is convenient to create two cells as follows to drive PM_Index.



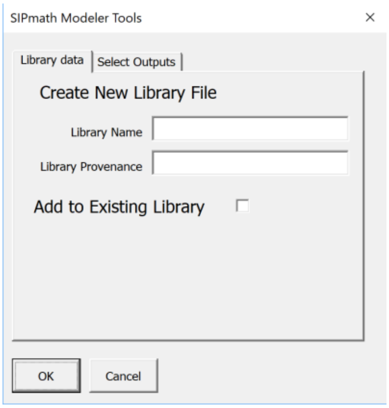
6.9 Import

The *Import* tool allows the modeler to import SIP library files in XML or CSV format. Selecting *Import* brings up the Windows *File Open* dialog to identify the file being imported. The tool creates another file (which should be a standard SIP/XML or SIP/CSV file) and unpacks the selected file into it. The resulting file's *SLURP* worksheet holds the SIP metadata and SIP data in the standard Excel SIP library layout.

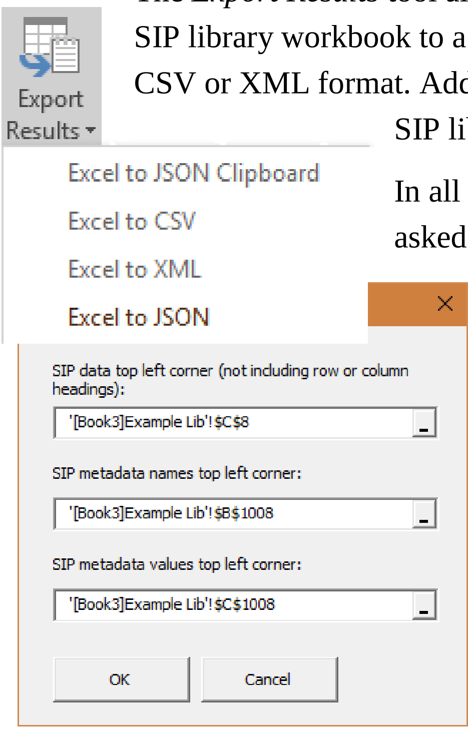


6.10 Save Results

The *Save Results* tool allows the modeler to export the model's data to an Excel SIP library workbook. Selecting *Save Results* brings up a dialog for specifying a worksheet name and selecting the SIPs to be exported. Use *Save As* to write the SIP Library file compatible with the format expected by the Library Input tool.



6.11 Export Results



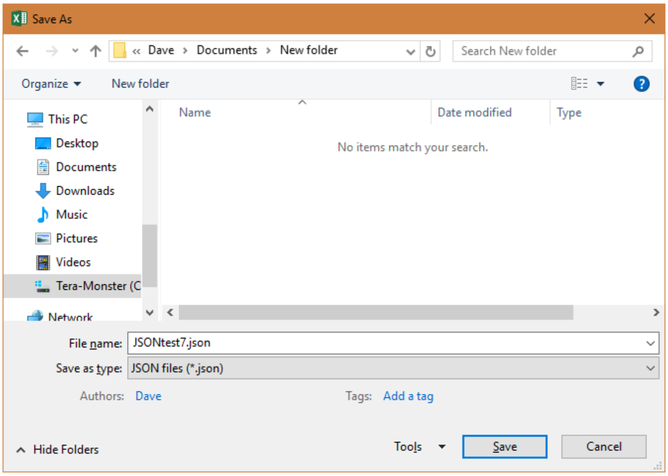
The *Export Results* tool allows the modeler to convert an Excel SIP library workbook to a SIP library file saved in either JSON, CSV or XML format. Additionally, the modeler may copy the SIP library to the clipboard in JSON format.

In all four cases, the modeler will first be asked to define the locations of some of the SIP data and metadata. SIPmath will supply its best guess as to these locations by default. These are usually correct, especially if the modeler is Exporting a library created with the *Save Results* button.

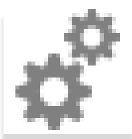
When these locations are specified, SIPmath asks for a file name, unless exporting to the clipboard.

Finally, SIPmath asks for the name and origin of the SIP Library (or SLURP).When all the information has been entered, SIPmath converts the

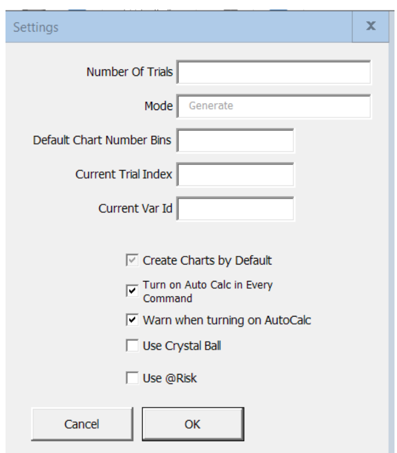
Excel library to the specified format and saves the data with the specified file name.



6.12 Settings



The *Settings* tool brings up a dialog for setting the tool's default values and behaviors. The Number of Trials is initially set to 1,000. We can either increase or decrease that number depending on our needs. The Mode is automatically set to generate but can be changed to SIPmath Library Mode as described in section 2. The Default Chart Number Bins is 10. Specifying more bins provides greater definition but might also result in a more erratic looking histogram. Current Trial Index is the current value of the variable PM_Index which is defined on the



PMTable worksheet. The Current Variable ID refers to the value when using the HDR random variable. This will change as random variables are defined.

6.13 ABOUT

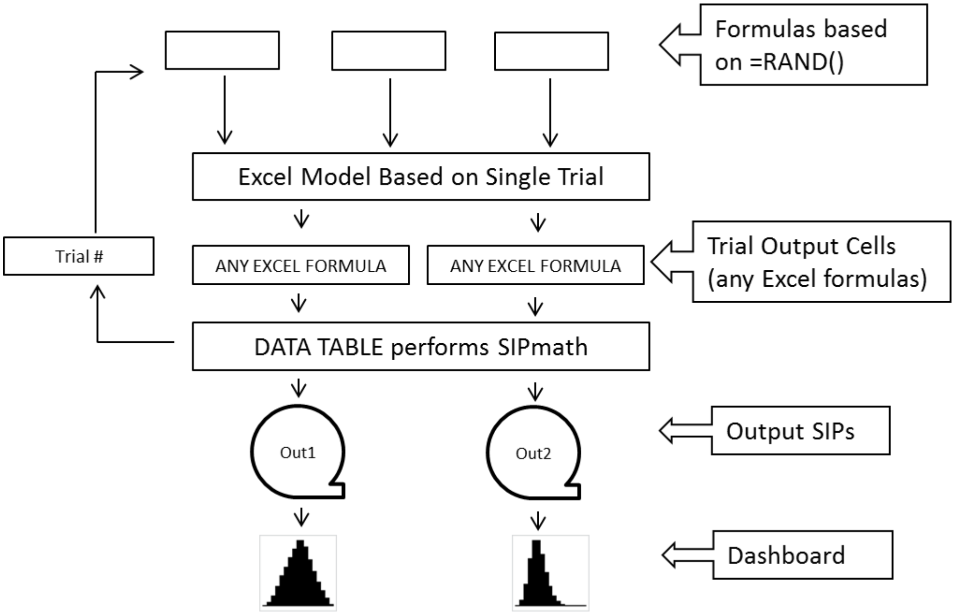
The *About* tool brings up a box with information about the Modeler Tools.

7 SIPMATH ARCHITECTURE

7.1 OVERVIEW

Generate Mode

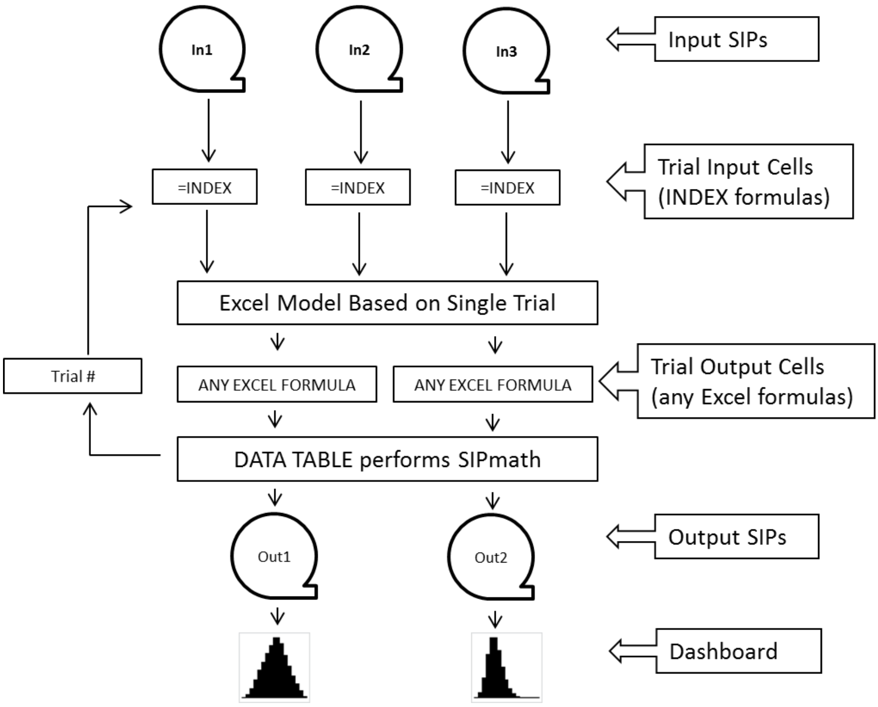
When applied in random simulation mode, no SIP library is needed, and the input expressions are based on `HDR` or `RAND()` formulas as shown here.



The Data Table function iterates the model through the trials of random inputs, calculates, and collects the results in arrays in the PMTable sheet. This is accomplished entirely with built-in commands and without using VBA. So, once created, such models may be run in native Excel without the Tools.

The SIPmath Modeler Tools is a set of macros that facilitate the development of such models by automatically creating a data table, and assisting in setting up the inputs, outputs, graphs, etc. The model calculations are formulated as if the inputs and outputs are simple numerical cells in Excel. Then the Excel data table function causes the simulation to run the model for the number of iterations stored in the `PM_Trials` variable. Data Table computations are very efficient, and for medium sized models, thousands of calculations may be performed nearly instantaneously.

SIP Library Model



In SIP Library Mode the input formulas driven by the `HDR` or `RAND()` formula are replaced with `INDEX()` formulas that step the model through the trials of prepared input SIPs, and collect the results into output SIPs.

7.2 Stochastic Libraries

When not using Generate Mode, input distributions come from *SIP Libraries* – Excel worksheets containing SIPs and SLURPs (coherent sets of SIPs). There are some specifics that must be met for the SIPmath Modeler Tools to use them:

1. The SIPs can be in rows or in columns, but each one must be defined as a contiguous named range.
2. Optionally, there can be metadata appended to a SIP's trial values. (See below for metadata range naming rules.) These are useful for observing the results of your models under specified scenarios, such as average values of the inputs. If used, there must be a range named PM_Meta, containing the names of the metadata elements, and another named range PM_Meta_Index with the displacements to the metadata values. These displacements must be common to all the SIPs in the library. For example, assuming a 1,000 trials, the two ranges might appear as:

PM_Meta_Index	PM_Meta
1001	Average
1002	Median
1003	Minimum
1004	Maximum

These names can be used by the *Trial Info* dialog to let the modeler see model instances for specific trials. The named ranges for the input distributions do not include metadata. However, An additional range, with the same name as the SIP but with “.MD” appended to the name must include the PM_Meta elements so that *Trial Info* can use the range name with an INDEX() formula to find them. The values of the metadata are not calculated automatically, and must be specified when creating the library.

Important

1. Metadata elements are not included in input SIP defined ranges.
2. SIP data and Metadata are included in a defined range given the same name as the SIP but with “.MD” appended to the name.
3. Metadata values are data, not formulas, so if a SIP trial value is changed, the SIP’s metadata must be recalculated. The SIP Standard says nothing, and the Tools do nothing, about disagreement between a SIP’s trial values and its metadata.
4. The number of metadata elements are not included in the value of variable PM_TRIALS.

7.3 Input Cells

Input cells are the interface between the input SIPs and the model.

In SIP Library mode, the interface between the model and the SIPs in the library is a pair of ranges – one for the SIP names and another for the SIP trial values as specified by the user in the *Library Input* tool. Each input distribution must have an input value cell and optionally one input name cell. The SIPmath Input tool will initialize the value cells with INDEX()

formulas pointing to the first trial in each SIP. As the data table increments PM_Index, these will evaluate to successive values from the corresponding input SIPs. The model refers to these cells to get its input data. In Generate Mode, the Index formulas are replaced with formulas using `RAND()` or HDR and the input cells do not require names.

7.4 OUTPUT CELLS

Output cells interface between the model and the data table

With the *Define Outputs* tool, the modeler will specify one or more cells that will evaluate to the results of the model calculation for each trial. These results will be written into successive rows in the data table, collectively making up the output distributions. The use of this tool is the same in both modes. However, in Generate Mode, the output SIPs will change every time a calculation is performed or the F9 Key is pressed.

7.5 THE DATA TABLE

The data table steps through the input distributions and records the output distributions.

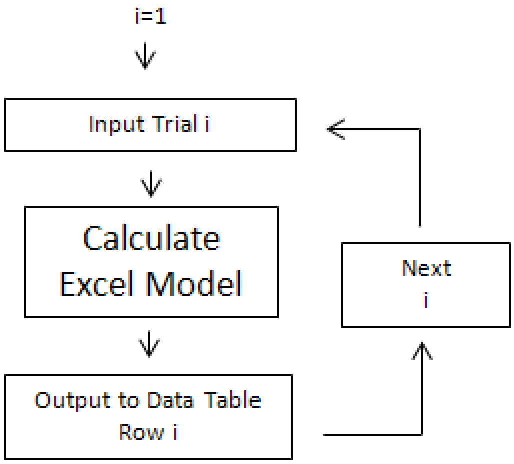
The Tools set up the data table on a worksheet called PMTable as part of the initialization. This includes defining PM_Index to be used as the trial number by the input cell formulas.

It also fills the leftmost column of the data table with a count from 1 to the total number of trials. These will be fed, one at a time, to PM_Index, which in turn will be used by the input cell formulas to select specific trials from the input distributions.

The *Define Outputs* tool completes the picture by setting up the output column or columns of the data table. It sets the formula driving the data table to a simple assignment that copies the model output cell.

7.6 The Process

A sheet recalculation event causes the data table to sequentially copy all the numbers from 1 to the number of trials into PM_Index. This results in re-evaluating the model for each of the trials in the stochastic library.



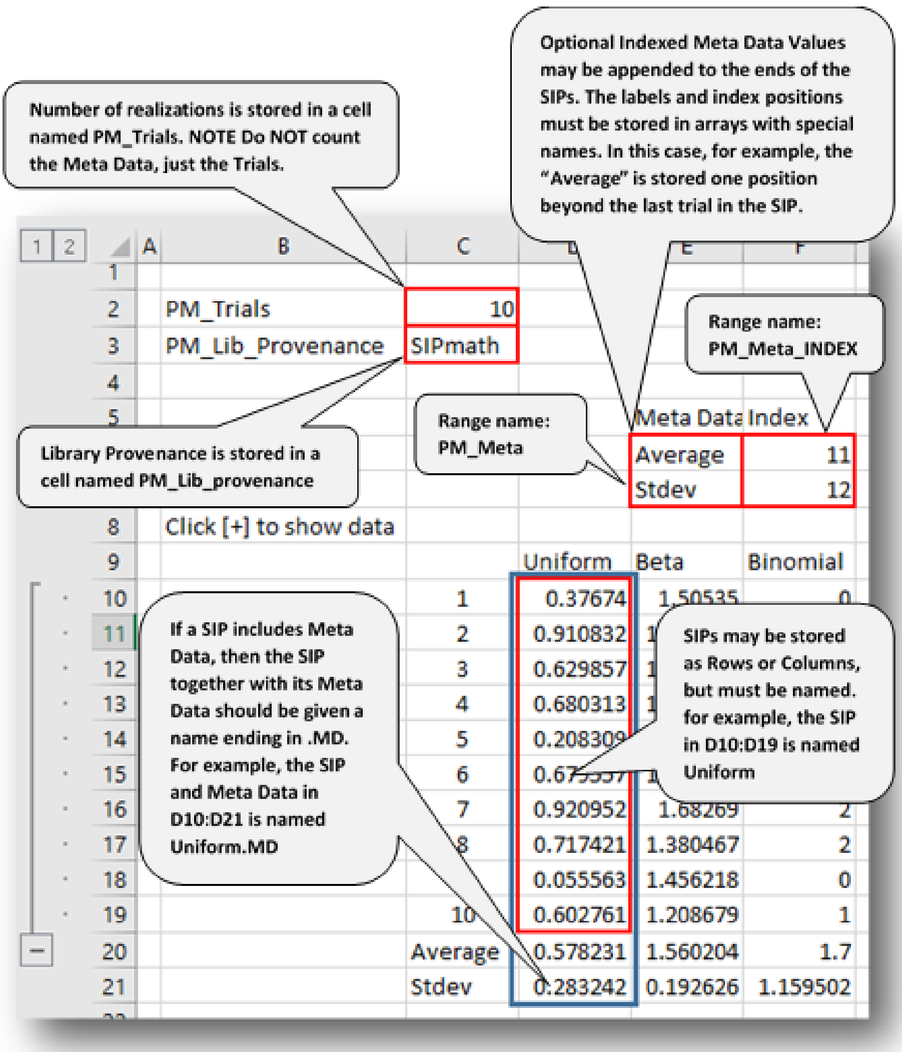
The Simulation Loop

The results appear in the corresponding rows of the output columns. Because these calculations are all internal to Excel, they are much faster than they would be if the numbers were pasted into Excel by a macro.

The process ends when the desired number of trials is reached.

8 SIP LIBRARY SHEET FORMAT

Libraries are free-format, with information conveyed in Excel variable names & range names. Other Formats include XML and CSV.



[i]Distribution Processing and the Arithmetic of Uncertainty, Sam Savage, November/December 2012, Analytics Magazine

[ii]Probability Management, Sam Savage, Stefan Scholtes and Daniel Zweidler, OR/MS Today, February 2006, Volume 33 Number 1

[iii]Probability Management 2.0, by Sam Savage and Melissa Kirmse, ORMS Today, October 2014

[iv]Sam L. Savage, The Flaw of Averages, Why we Underestimate Risk in the Face of Uncertainty, John Wiley 2009, 2012