

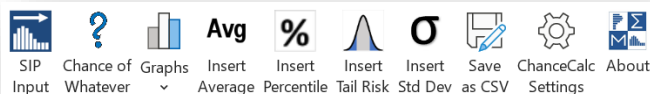
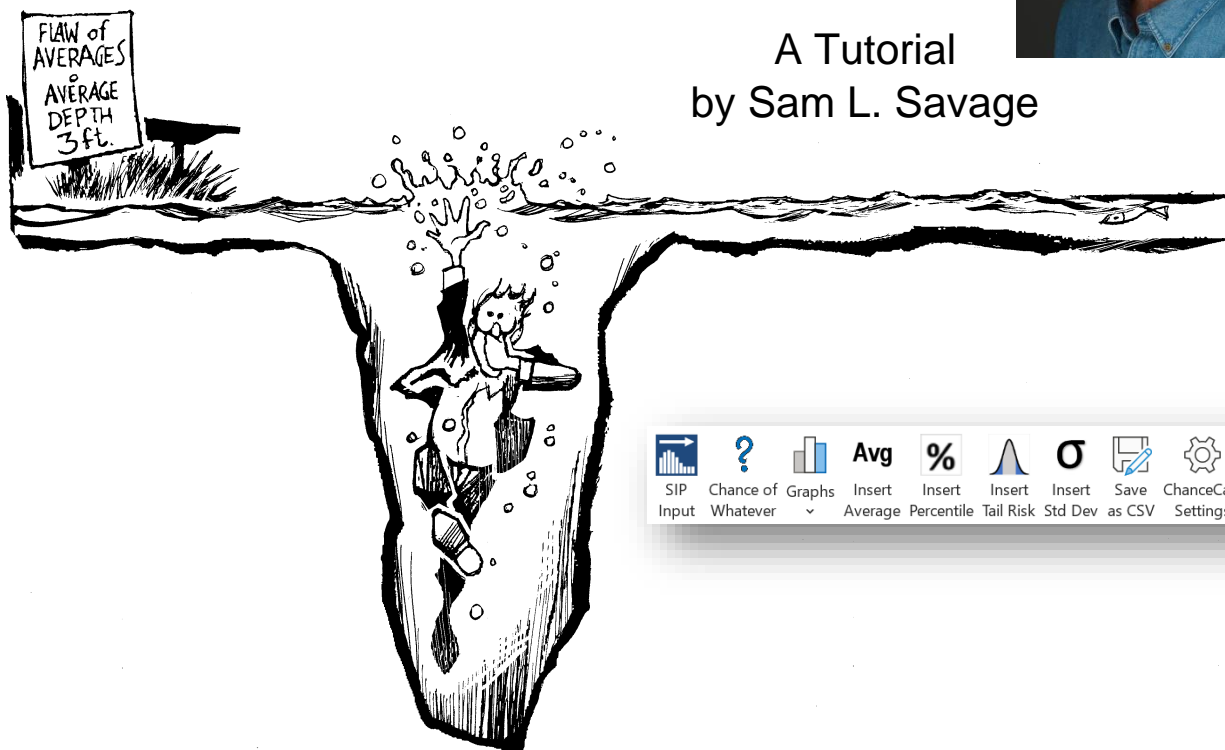


# Curing the Flaw of Averages

with



A Tutorial  
by Sam L. Savage



**ChanceCalc** is an Excel add-in for creating chance-informed dashboards in your spreadsheet based on uncertainties expressed in the SIPmath™ Standard. The models created with **ChanceCalc** are stand-alone Excel files that perform hundreds or thousands of simulation trials per second and do not require the add-in to run.

## Acknowledgments

I want to thank Doug Hubbard and Tom Keelin, without whose complementary technologies the 3.0 Standard would not exist. Nik Kalyani and Kevin McDonald encouraged and assisted us with the JSON format. Dan Fylstra of Frontline Systems Implemented the standard in Analytic Solver, providing a powerful platform that can export SIP libraries for use by **ChanceCalc**. Dave Empey has done the bulk of the computer programming that got ProbabilityManagement.org off the ground including the code for **ChanceCalc**. Daniel O'Neil and Aaron Brown assisted with the architecture and developed Python code for creating and reading the 3.0 Standard. Bridget Cash, Melissa Kirmse, Mary Claire Meijer, and Matthew Raphaelson have all influenced my approach to expressing probabilistic concepts without triggering Post-Traumatic Statistics Disorder (PTSD).

If you have not yet viewed it, we recommend watching the **ChanceCalc** [introductory video](#) before proceeding with the tutorial.

ChanceCalc 5-9.mp4

# ChanceCalc

Arithmetic of Uncertainty

Probability Management

Chance of Being	
Beyond Budget	77%
Below Projection	50%
Behind Schedule	88%

0:15 / 4:18

# Contents

- Introduction ..... 1
- Lesson 1a – Demonstrating the Flaw of Averages ..... 3
  - Beyond Budget in Operating Costs ..... 3
  - Nonlinearity ..... 3
- Lesson 1b – A Simple Simulation Based on Historical Data ..... 4
  - Models based on SIPs ..... 4
  - Chance of Exceedance ..... 5
  - Chance-Informed Stocking Decision ..... 5
- Using ChanceCalc ..... 6
  - Loading ChanceCalc ..... 6
- Lesson 2 – Loading a SIP Library from your Computer ..... 7
  - A Capital Investment Problem ..... 7
  - Loading the Library ..... 7
  - Finding the Simulated Average Profit ..... 8
  - The Chance of Whatever ..... 9
  - Percentiles..... 10
  - Creating Graphs ..... 10
  - Was 100,000 the Optimal Capacity? ..... 11
- Virtual SIPs ..... 12
- Lesson 3 – Using a Library from the Cloud..... 12
  - A Scheduling Model ..... 12
  - Chance-Informed Schedule..... 14
- Lesson 4 – A Chance-Informed Decision with Multiple Experiments ..... 15
  - Performing Multiple Experiments..... 15
  - Creating a Chance-Informed Graph ..... 17
- Lesson 5 – Failure Analysis, Multiple Identical Variables..... 18
  - Predicting the Failure of Aircraft Parts ..... 18
  - Linking Simulations to Serial Numbers ..... 19
- Chancification and the SIPmath Standard ..... 20
- SIP Libraries..... 21
- Lesson 6 – Creating a CSV Library of Profit from Lesson 2 ..... 22
- Tools for Generating SIP Libraries..... 23
- Appendix: Worksheets added by ChanceCalc ..... 24

## Introduction

Have you heard about the statistician who drowned in the river that was on average three feet deep? This is just one example of what can go wrong when an average is substituted for an uncertainty.



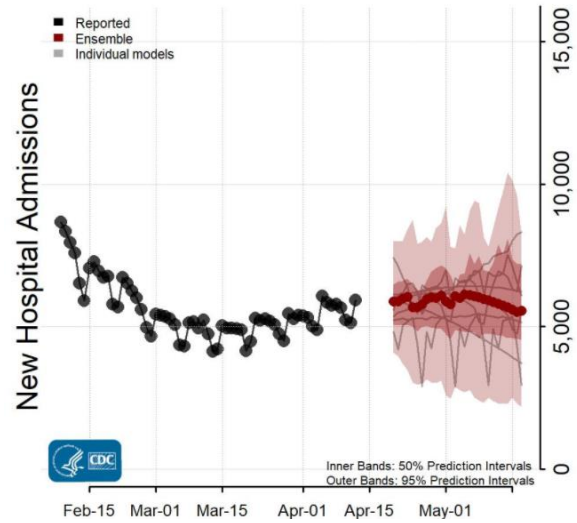
Two decades ago, I coined the term Flaw of Averages to describe a whole family of such systematic errors, which imply that plans based on average assumptions are wrong on average. It explains, for example, why so many projects are beyond budget, below projection and behind schedule. I refer to these as the Flaw of Averages Trilogy. We will meet all three below. Today, new standards from nonprofit ProbabilityManagement.org are curing this problem by conveying uncertainty as auditable data called SIPs. Instead of containing a single number, SIPs may contain thousands of different numbers representing possible outcomes of an uncertainty. For example, the SIP of rolling a single die might be depicted as shown below.

 = **6, 1, 2, 3, 6, 2, 5, 5, 4, 5, 2, ... 2, 4, 1**  
The SIP of a die roll

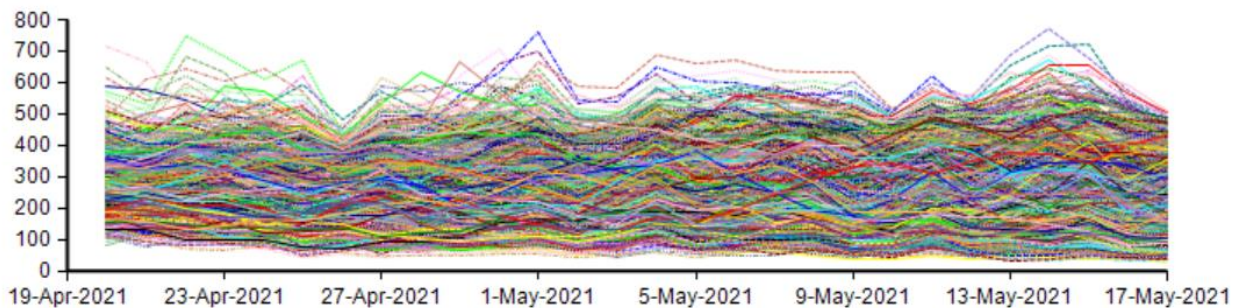
The new SIPmath™ 3.0 Standard can represent the same information in a tiny fraction of the storage in JSON format. In general, SIPs require statistical expertise to generate. However, they may be used by non-statisticians in Excel, R, Python, or virtually any other software environment. This tutorial will show you how to use SIP Libraries in Excel with a revolutionary add-in called **ChanceCalc**. You will see how “chance-informed” thinking leads to better decisions in the long run.

## How is uncertainty captured?

Nearly all forecasting methods yield information on their own accuracy, but instead of being used, this is often discarded. Probability management provides a way to get that “uncertainty” to the decision maker in an actionable form. As an important example, the CDC provides forecasts of COVID-19 hospitalizations that provide ranges of outcomes<sup>i</sup>. In the graph on the right, there is a 95% chance that future hospitalizations will lie within the light red band and a 50% chance it will lie within the darker red band. The darkest red line represents the median: that is, there is a 50/50 chance of hospitalizations above or below this level for each week of the forecast.



These colored bands are not of much use to a decision maker, so people are likely to go with the single number median for each time period, which again leads to the Flaw of Averages. But the CDC also provides data that can be used by analysts to generate SIP Libraries through a process known as Monte Carlo simulation. In effect, this displays future hospitalizations in 1,000 parallel universes, in any one of which we might find ourselves, as shown below. This results in 28 SIPs of 1,000 trials, one for each day in the forecast.



I will begin by demonstrating the systematic errors that result from using single numbers in place of uncertainties.

## Lesson 1a – Demonstrating the Flaw of Averages

- Beyond Budget in Operating Costs
- Nonlinearity

### Beyond Budget in Operating Costs

Open **Inventory.xlsx**. Here we model the quantity of a perishable drug to stock in the face of uncertain monthly demand. We only consider the operating costs associated with the inventory, and not the cost of the drug itself. Demand is uncertain, but the month over month average is five units (A2), so obviously we will stock five units (C2). However, if monthly demand is less than the amount stocked, then any remaining units must be discarded with an expiration cost of \$50 per unit as shown in the formula in cell D5.

	A	B	C	D	E	F
1	Demand		Amt Stocked			
2	5		5		Uncertainty	
3					Decision	
4			Costs	Per Unit	Total	Other Inputs
5	Expiration Cost		\$50		\$0	
6	Air Freight		\$150		\$0	
7						
8	Operating Cost				\$0	

D5: =IF(Amt\_Stocked>Demand,(Amt\_Stocked-Demand)\*Exp\_Cost,0)

If on the other hand, demand exceeds the amount stocked, the extra units must be airfreighted in at \$150 per unit as shown in cell D6.

D6: =IF(Amt\_Stocked<Demand,(Demand-Amt\_Stocked)\*F\_Cost,0)

The Operating Cost in cell D8 is simply the sum of the two costs above, at most one of which will be greater than zero.

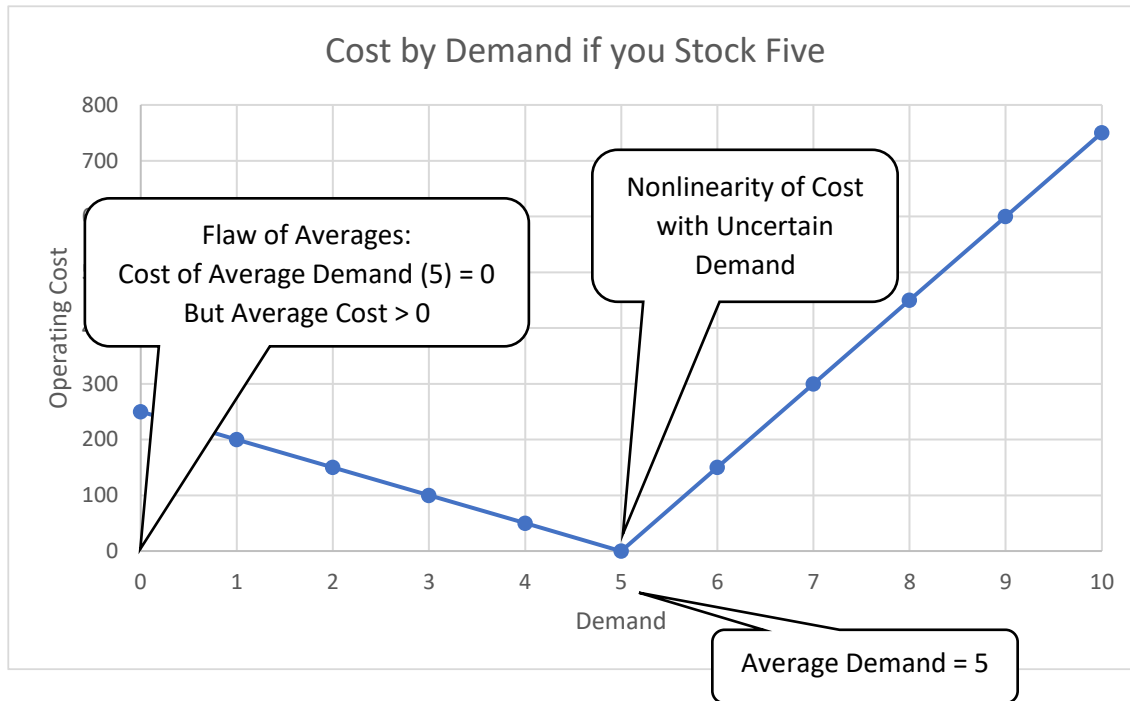
#### Experiment with the model

Change the demand in cell A2 to be both greater than and less than 5, and observe the formulas in cells D5, D6, and D8.

Suppose the boss comes in and asks what the operating cost will be next month. Many people will plug the average of five into cell A2 and estimate the cost to be zero.

### Nonlinearity

This is our first encounter with the Flaw of Averages. If demand deviates either direction from five, there will be a positive cost. This nonlinearity in the model triggers a mathematical concept called Jensen's Inequality, which results in a particular form of the Flaw of Averages. The cost of the average demand is zero, but the average cost is greater than zero.

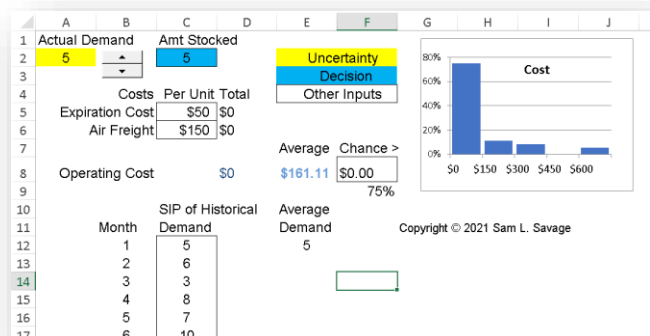


The cost is said to be nonlinear with demand because the line above is not straight but bent. Whichever way demand goes from average, cost goes up. This demonstrates why averages can yield systematically bad results, but it does not tell us how much to budget for operating costs on average. We will answer this in the next lesson.

## Lesson 1b – A Simple Simulation Based on Historical Data

- SIPmath
- Simulated Average Cost
- Chance of Exceedance
- Chance-Informed Stocking Decision

Open **Beyond Budget.xlsx**. In this file we have explicitly modeled the uncertainty in demand with 36 months of historical data in cells C12:C47. This is a simple example of a SIP, or Stochastic Information Packet, analogous to the rolls of the die depicted above. Start by scrolling through a few of the possible demands using the control in column B.



### Models based on SIPs

This file runs the entire SIP of demand through the model with each keystroke. That is, every calculation is done 36 times, or simulated, by running each of the historical demands through an INDEX formula in

A2. The 36 resulting operating costs appear in a range called Output\_1 on a sheet called PMTable. As with the hospitalizations discussed above, imagine observing the problem in 36 parallel universes to get a better idea of what will happen in the one universe in which we live. The Output\_1 range was generated automatically by **ChanceCalc** when the model was created and makes use of the Excel Data Table command.

	A	B	C	D	E
1	1				
2		Index	Output_1		
3		Values	\$0.00		
4		1	0		
5		2	150		
6		3	100		
7		4	450		
8		5	300		
9		6	750		

### Simulated Average Cost

Cell E8 of the Model page contains the average of the operating cost simulated over all 36 months of \$161.11 and is determined with the formula

E8: =AVERAGE( Output\_1 )

So, to review, if you had estimated operating cost based on the average demand of five, you would have budgeted \$0, but on average you need a budget of over \$160 per month. This example of the Flaw of Averages highlights the importance of explicitly modeling uncertainty.

### Chance of Exceedance

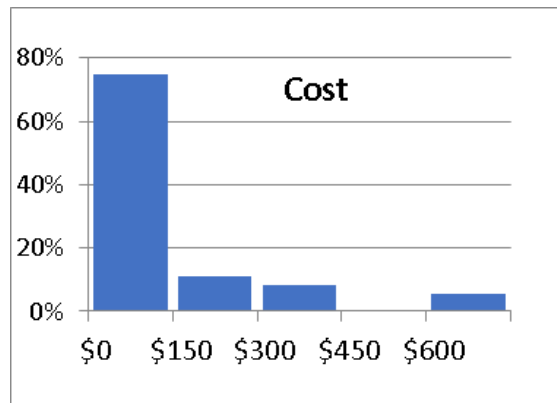
The model also displays the chance of exceeding any specified cost. Cell F9 displays the chance of exceeding the costs entered in cell F8, and uses the formula:

F9: =COUNTIF( Output\_1, ">" & F8 ) / 36

It counts all occurrences of the cost (Output\_1) which exceed the number in cell F8, and then divides by 36 to turn it in to a probability. Try entering other costs in cell F8 to determine their chance of exceedance. NOTE: You do not need to remember this formula because **ChanceCalc** enters it for you.

#### The Histogram

The histogram on the right shows the relative chances of various costs.



### Chance-Informed Stocking Decision

In making decisions under uncertainty, we usually face tradeoffs between average outcomes and adverse outcomes. Let's imagine that in this case you were not so foolish as to budget \$0 for operating costs but had budgeted \$200. Now of course you want to minimize your average cost, but you are also worried about the risk of exceeding \$200 on any month.

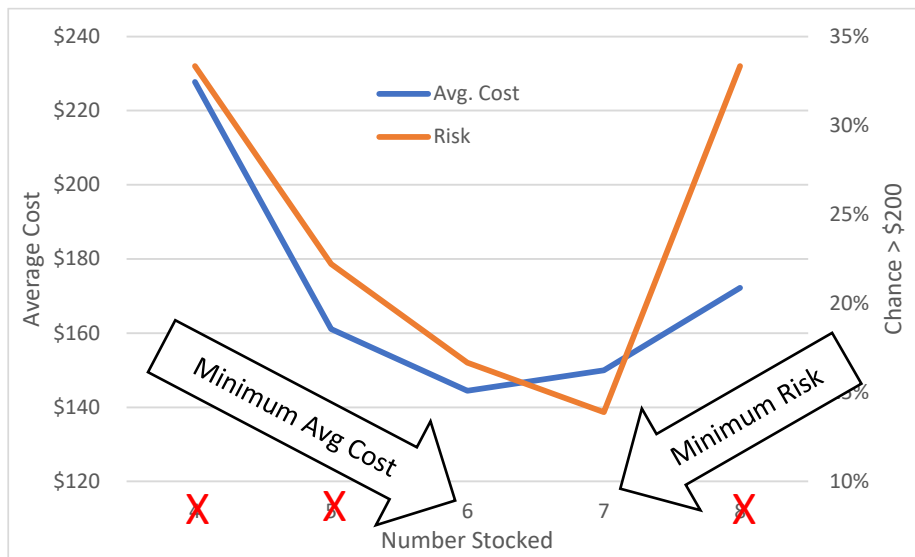




Set cell F8 to \$200 and experiment with the quantity to stock in cell C2. Observe both the average cost, and the chance that cost is greater than \$200. If you try all stocking levels between 4 and 7, you should observe the results in the table on the left. Remember that before modeling uncertainty we were simply stocking five units and expecting zero costs. The table shows that the minimum average cost occurs with six units, while the minimum risk of exceeding \$200 occurs for seven units. Risk is in the eye of the beholder, and which of these two you would choose would depend on your risk attitude. But none of the other levels, including the formerly obvious quantity of five, make sense regardless of risk attitude. In describing this situation to decision makers, I find it useful to create graphs like the one below, which will be explained further on in this tutorial.

Number Stocked	Average Cost	Chance Cost > \$200
4	\$227.78	33%
5	\$161.11	22%
6	\$144.44	17%
7	\$150.00	14%
8	\$172.22	33%

Observe the results in the table on the left. Remember that before modeling uncertainty we were simply stocking five units and expecting zero costs. The table shows that the minimum average cost occurs with six units, while the minimum risk of exceeding \$200 occurs for seven units. Risk is in the eye of the beholder, and which of these two you would choose would depend on your risk attitude. But none of the other levels, including the formerly obvious quantity of five, make sense regardless of risk attitude. In describing this situation to decision makers, I find it useful to create graphs like the one below, which will be explained further on in this tutorial.



## Using ChanceCalc

### Loading ChanceCalc




In Windows you may run the **ChanceCalc** Install file. On Mac you must open ChanceCalc.xlam and Enable Macros.

## Lesson 2 – Loading a SIP Library from your Computer

- A Capital Investment Problem
- Loading a SIP Library from your computer
- The Simulated Average
- The Chance of Whatever
- Percentiles
- Creating Graphs

### A Capital Investment Problem

 Open **Below Projection Example.xlsx**. This model involves a capital investment in manufacturing capacity, given an uncertain demand. For simplicity we will assume that cost of production is zero once the capacity has been paid for up front. It is known that average demand will be 100,000, so the current plan is to invest in capacity of that amount at \$30 per unit for a total investment of \$3,000,000. If we sell the average demand for \$40 per unit, we will gross \$4,000,000 in revenue for a profit of \$1,000,000. Does that sound familiar? This is usually accepted practice, but given that demand is uncertain, do you think we can really expect \$1,000,000 on average? Again, think of parallel universes as we did in lesson 1, but this time there will be 1,000 of them instead of 36.

	A	B	C	D
1				
2			Cost of Capacity per Unit	\$30
3			Selling Price per Unit	\$40
4				
5			Actual Demand (Uncertain)	100,000
6				
7			Decision: Capacity	100,000
8				
9			Sales	100,000
10			Revenue	\$4,000,000
11			Cost of Goods Sold	\$3,000,000
12			Profit	\$1,000,000
13				

### Explore the model

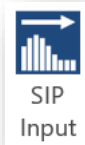
Start out by experimenting with the demand in cell D5. The nonlinearity in this model comes from the formula for Sales in cell D9.


D9: =MIN(Actual,Capacity)

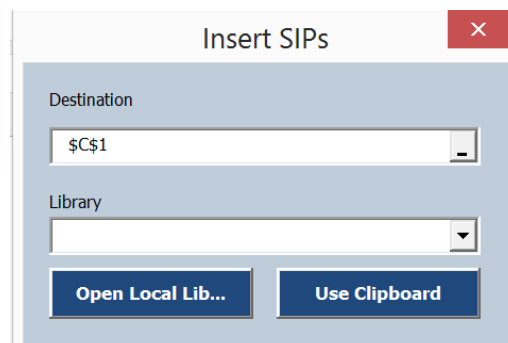
If you invest in capacity for 100,000, that's the *most* you can sell. There is no upside, so \$1,000,000 can't be the average profit, but is the maximum possible or best-case scenario. To find the simulated profit we will use **ChanceCalc**.

### Loading the Library

Imagine that your marketing department has forecast demand for the product and has emailed you a SIP of 1,000 trials, which is now stored on your computer.

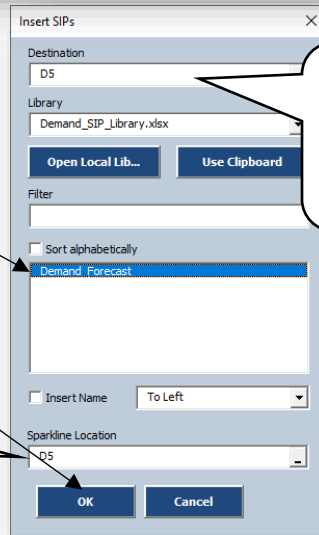
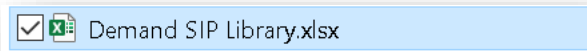


1.  Place the cursor in the Demand cell, D5, and click the **SIP Input** button.
2. Select the **Open Local Library** option.



3. In the Library window, navigate to the Demand SIP Library and open the file.

4. Select **Demand Forecast** in the lower window and click **OK**.



Make sure the Destination refers to cell D5.

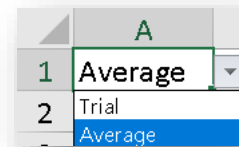
A Sparkline graph representing the SIP distribution will be placed in the cell referred to here, or you may leave the field blank for no Sparkline.

### The Result of Loading the SIP

Two or more sheets may be added to your workbook when you click OK after inserting the first SIP from a new library. This is described in the appendix, which I suggest that you examine later.

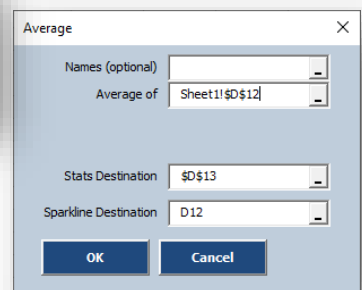
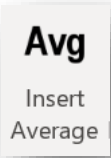
### Experiment with the model

1. Look at various trials (parallel universes) by changing cell A2. Most people in the organization are imagining \$1,000,000. In that light, trial 9 is sobering.
2. Select **Average** off the menu in cell A1 to observe the Profit given Average Demand.



### Finding the Simulated Average Profit

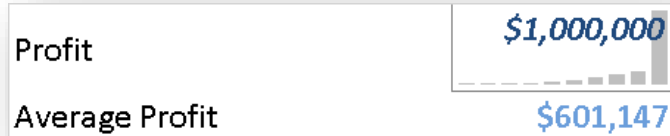
1. Type "Average Profit" in cell B13
2. Place the cursor in cell D13
3. Click the **Avg** button
4. You may provide an optional name for the formula cell (D12). Type Profit in this field, or alternatively click in cell B12. Names are important when you plan to save SIPs from one model for use in other models as discussed in a later lesson.
5. Move to the **Average of** field and select cell D12. (This should automatically fill the Sparkline Destination with D12 as well.)
6. You may leave the **Sparkline Destination** as D12, (or choose another location) then click **OK**.



Cell D12 now has a Sparkline graph histogram showing that there is a downside but no upside. I like to increase the row height for these cells so you can see the histogram more clearly. And the Simulated Average Profit is only about \$600,000, not the promised million.

Why?

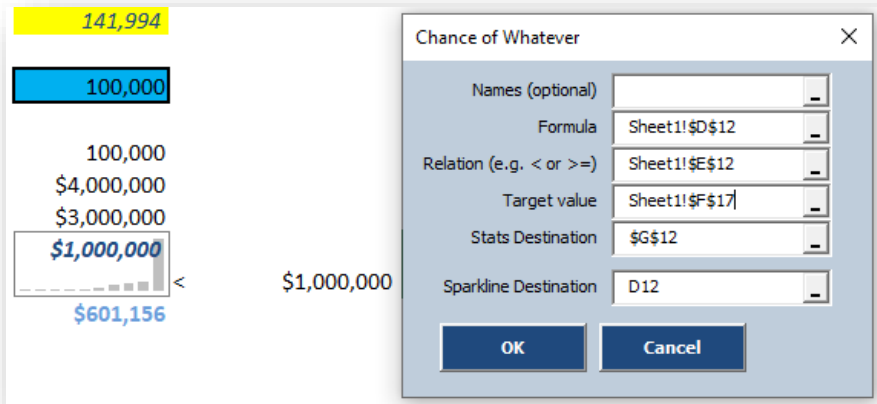
Easy. If demand is less than 100,000 you won't make your \$1 million, but if demand is greater than 100,000, then the maximum you can make is \$1,000,000, because you will have exceeded your capacity. The Flaw of Averages strikes again!



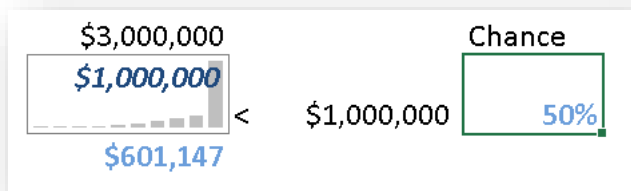
## The Chance of Whatever

Let's determine the chance that profit is less than any particular amount.

1. Type < (for less than) in cell E12, and \$1,000,000 in cell F12
2. Put the cursor in cell G12 and click the **Chance of Whatever** button
3. Click in the specified cells in each field and click **OK**




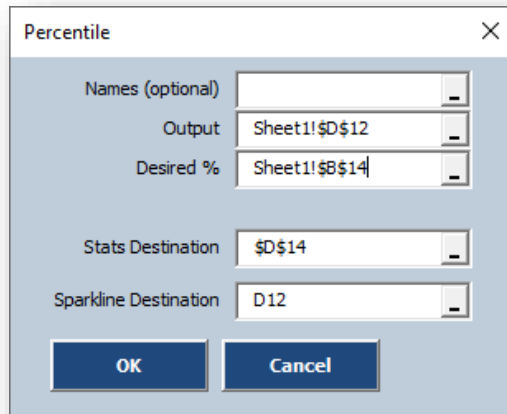
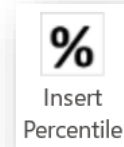
You will see that there is a 50% chance of making less than \$1 million. Next see what the chance is of losing money by typing \$0 in cell F12.



## Percentiles

The percentile is the value, or less, that Profit will take on with a given probability. We will start by finding the 10<sup>th</sup> Percentile of Profit. (Note: all statistical results provide the option to name the output).

1.  Enter 10% in cell B14
2. Type "Percentile" in cell C14
3. Put the cursor in cell D14 and click the **Insert Percentile** button
4. Click in the specified cells in each field and click **OK**

A dialog box titled "Percentile" with a close button (X) in the top right. It contains several input fields: "Names (optional)" (empty), "Output" (Sheet1!\$D\$12), "Desired %" (Sheet1!\$B\$14), "Stats Destination" (\$D\$14), and "Sparkline Destination" (D12). At the bottom are "OK" and "Cancel" buttons.


We see that there is a 10% chance of losing about \$280,000 or more! Determine the 5<sup>th</sup> percentile by changing cell B14 to 5%.

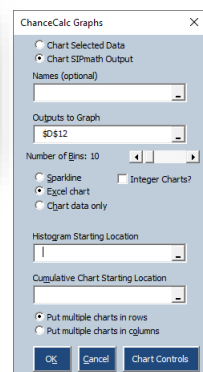
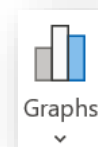
Profit	\$1,000,000
Average Profit	\$601,147
10% Percentile	-\$279,288

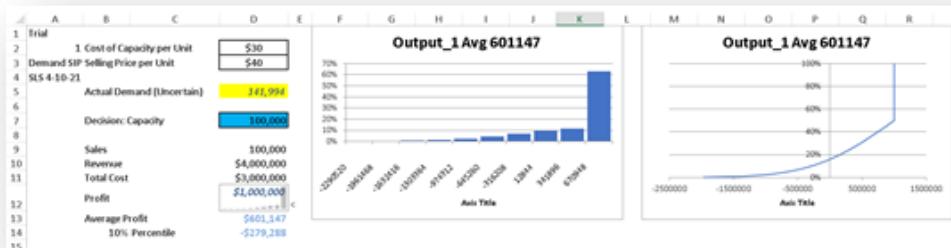
I prefer using chances rather than percentiles as I find it a more natural way for most people to think. If you enter -\$279,288 into cell F12 you will find that the chance of being less than this number is 10%, confirming what we just learned. Percentiles, however, were easier for mathematicians before computers, and are still in common use today. So, it is important to be able to work with both ways of expressing similar things.

## Creating Graphs

**ChanceCalc** supports Histograms, Cumulative, Exceedance, and Scatter plots. Here we will discuss the first two.

1.  Place the cursor in the Profit cell, D12, and click the **Graph** button
2. Choose locations for the Histogram and/or Cumulative Graphs, then click **OK**. Note that the locations do not need to be empty cells as the Graphs will be pasted over and can be moved aside. As with the Average button, you will have the option to name the formula cell.

A dialog box titled "ChanceCalc Graphs" with a close button (X) in the top right. It has radio buttons for "Chart Selected Data" and "Chart SPmath Output". Below are "Names (optional)" (empty), "Outputs to Graph" (\$D\$12), "Number of Bins: 10", "Sparkline" (checked), "Integer Charts?" (unchecked), and "Chart data only" (checked). There are also "Histogram Starting Location" and "Cumulative Chart Starting Location" fields. At the bottom are "OK", "Cancel", and "Chart Controls" buttons.

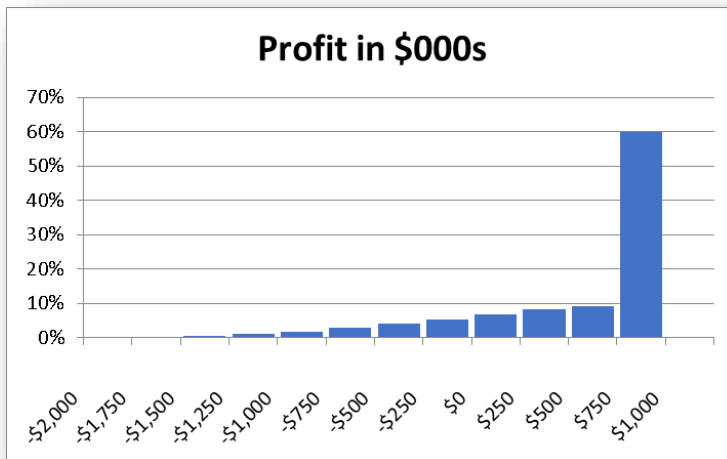


Many formatting options are available for the graphs on the SIPmath Chart Data page.

I made the following changes to get the histogram below.

1. Changed Format in row 11 to Currency
2. Removed decimals from Currency Format in row 25
3. Set minimum in row 14 to -2,000,000
4. Set bin width in row 8 to 250,000
5. Set number of bins in row 6 to 13
6. Set Scale in row 10 to 1,000
7. Set Series Name in row 3 to Profit in \$000s


	A	B
1		Output_1
2	Expected Value	601147
3	Series Name	Output_1 Avg
4	Integer chart?	FALSE
5	Loss Exceedance	FALSE
6	Number of Bins	10
7	Label position	10
8	Bin Width	329052
9	Decimals	0
10	Scale	1
11	Format Char	General
12	Use Axis Title?	General
13	Axis Title	Number
14	Min	Currency
15	Bin Range	Accounting
16	Frequency	Date
17	Labels	Time
18	Cumulative	Percentage
		Fraction
		-1305364
		-974312



### Was 100,000 the Optimal Capacity?

If we ordered more, maybe we could take advantage of that excess demand. Or maybe not. As with Lesson 1b, there are tradeoffs between profit and risk. If you set up your model correctly, you should have a Simulated Average Profit of \$601,147 and a 16% chance of losing money. The spike in the histogram above indicates that profit is capped at \$1 million because we only invested in a capacity of 100,000 units. It makes sense that if we invested more in capacity, we could take advantage of the possibility of higher demand and increase our average profit. It makes sense but it is dead wrong.

Increase capacity to 110,000. Average profit drops to \$469,651 and the chance of loss goes up to 24%, so that would be a terrible decision. Lesson 4 will provide the answers.

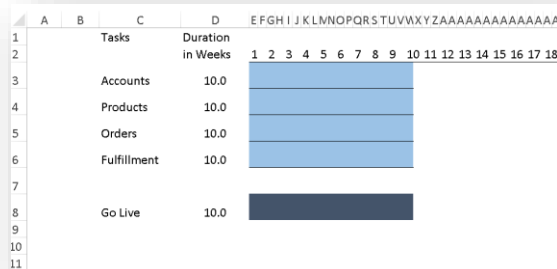
 Save this version of the file under a different name as we will use it again in Lesson 6 on creating SIP Libraries.

## Virtual SIPs

The SIPs in the earlier examples were analogous to collections of previous rolls of dice and are stored in the SIPmath 2.0 Standard<sup>ii</sup>. Virtual SIPs are more like the dice themselves. You can roll them as many times as you like, and they take up very little storage as JSON objects<sup>iii</sup>. An important difference from real dice is that they are controlled by Doug Hubbard’s HDR random number generator, which can produce repeatable, identical rolls across multiple platforms such as R, Python, or Excel. They are also based on Tom Keelin’s Metalog distribution, which can represent virtually any continuous uncertainty with a single family of formulas.


## Lesson 3 – Using a Library from the Cloud

- A Scheduling Model
- Virtual SIPs
- Why so many projects are behind schedule
- Chance Informed Schedule



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AAAAAAAAAAAAAAAA
1																											
2			Tasks	Duration	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18					
3			Accounts	10.0																							
4			Products	10.0																							
5			Orders	10.0																							
6			Fulfillment	10.0																							
7																											
8			Go Live	10.0																							
9																											
10																											
11																											

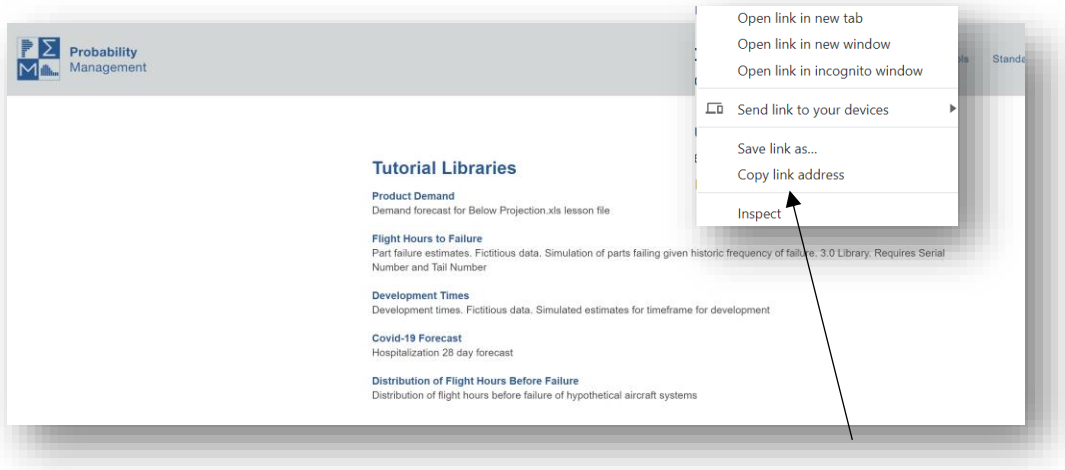
## A Scheduling Model

 Open **Behind Schedule Example.xlsx** as shown above. This file models the software development of the four major components of an online business: Customer Accounts, Product List, Orders, and Fulfillment. Each of these is being developed in parallel by separate teams, and from past data, it is known that each such development takes ten weeks on average. Of course, you don’t go live until the last one is done. So, when the boss comes in and demands to know when the project will be done, what do you say?

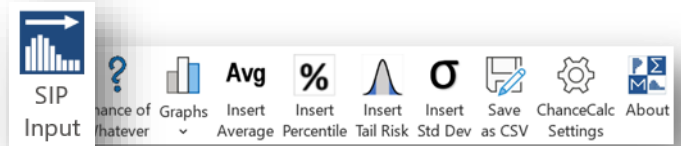
### Load the Development Time Library from the cloud

1.  Select cells D3:D6 in the model

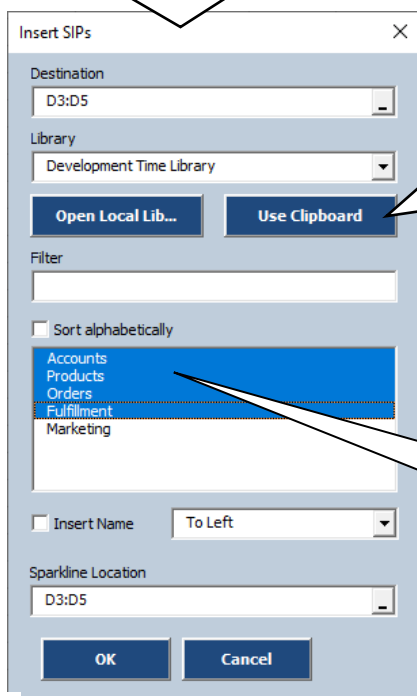
- Go to the ProbabilityManagement.org tutorial libraries page:  
<https://www.probabilitymanagement.org/tutorial-libraries>



- Right-click the third library, **Development Times**, and choose **Copy link address**.
- Back on the Excel model, click the **SIP Input** button on the far left of the Ribbon:



5. You will see this dialog box:



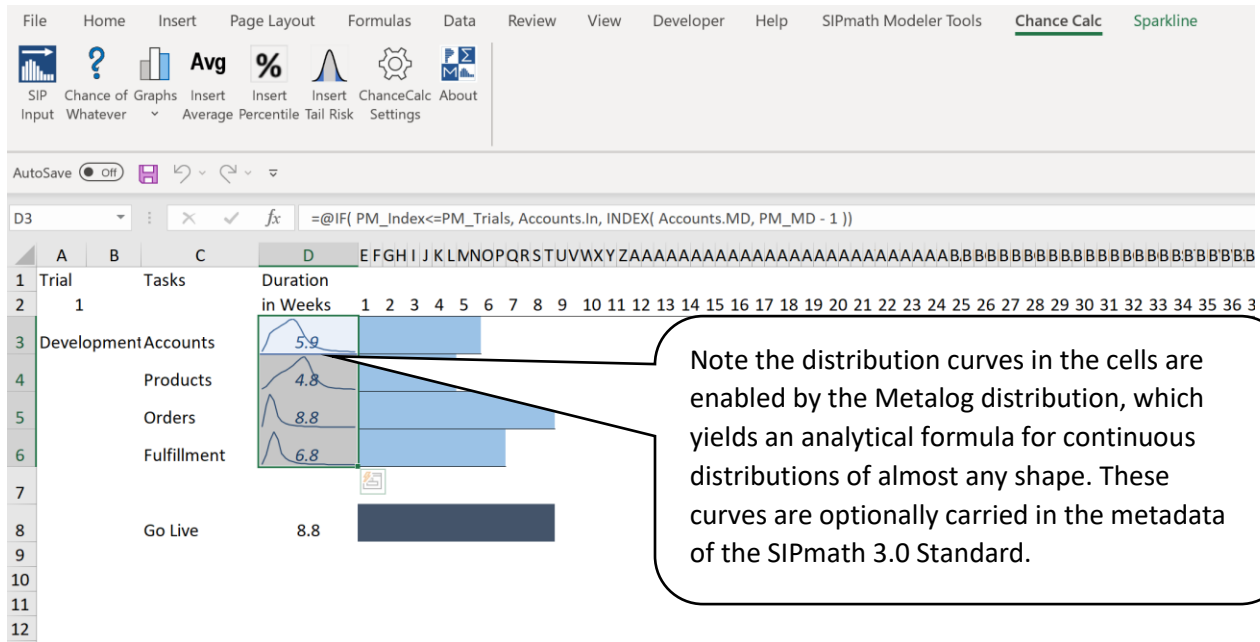
6. Click on the **Use Clipboard** button which will load the virtual library with five SIPs.

7. For this example, choose only the first four (Accounts, Products, Orders, Fulfillment) by holding the shift key and clicking on them.

- Then click **OK**



The spreadsheet will now look like this:



### Explore the model

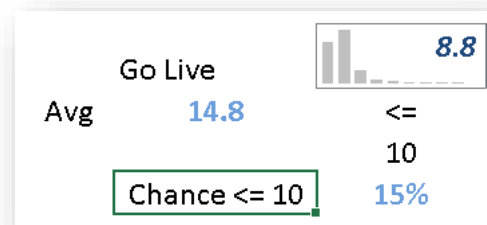
1. Run through a few trials by changing cell A2 to gauge the uncertainty. Notice that the MAX formula in D8 ensures that we go live only when the final task is done.
2. Put the cursor in cell A1 and select the mean durations of 10 weeks. What do you suppose the chance is of finishing in 10 or fewer weeks? When will we go live on average over 1,000 parallel universes? We will answer these questions next.

	A
1	Trial
2	Trial
3	min
4	mean
5	Trial1
6	Trial2
7	Trial3
8	max

### Chance-Informed Schedule

The steps are very similar to the ones in Lesson 2, so we will merely show the results here for you to check your work against. You may be shocked that the chance of finishing in 10 weeks is so low, but here is the intuition. If the four distributions of development time had been independent and symmetric, then each task would have had a 50/50 chance of finishing over or under 10 weeks. The only way to finish in 10 weeks or less would be like flipping four heads in a row on a coin, for which the chance is 1/16 or about 6%.


So, the surprise, if any, is that the chance is as high as 15%. Also note that we finish in 10 weeks given the average inputs, but that the simulated average is closer to 15 weeks, a classic case of the Flaw of Averages. Given that many contracts involve bonuses and penalties based on finish date, consistently having projects exceed your estimates by 50% does not lead to a sustainable business.



## Lesson 4 – A Chance-Informed Decision with Multiple Experiments

- Performing multiple experiments on the Below Projection example.
- Creating a Chance-Informed graph

### Performing Multiple Experiments

 Open **Below Projection Multi Experiment.xlsx** as shown below. This model, based on Lesson 2, will simultaneously analyze the average profit and chance of loss for each of eight capacity decisions.

	A	B	C	D	E	F	G	H	I	J	K	L
1												
2		Cost of Capacity per Unit		\$30								
3		Selling Price per Unit		\$40								
4												
5		Actual Demand (Uncertain)		100,000								
6												
7		Initial Capacity		50,000								
8		Capacity Increment		10,000								
9		Sales		50,000	60,000	70,000	80,000	90,000	100,000	100,000	100,000	
10		Revenue		\$2,000,000	\$2,400,000	\$2,800,000	\$3,200,000	\$3,600,000	\$4,000,000	\$4,000,000	\$4,000,000	
11		Total Cost		\$1,500,000	\$1,800,000	\$2,100,000	\$2,400,000	\$2,700,000	\$3,000,000	\$3,300,000	\$3,600,000	
12		Profit		\$500,000	\$600,000	\$700,000	\$800,000	\$900,000	\$1,000,000	\$700,000	\$400,000	
13												
14		Decision: Capacity		50,000	60,000	70,000	80,000	90,000	100,000	110,000	120,000	
15												
16		Threshold	\$0	Average	Chance < 0							
17												

### Experiment with the model

1. Inspect the formulas in rows 9 through 12 to see how they depend on the capacity in row 14.
2. Try various Actual Demands in cell D5 to see how the Profit changes in row 12.

### Load the Demand SIP


You may use the same steps as in Lesson 2 under **Loading the Library** to load the library from your computer. Alternately, right click the Product Demand library on the tutorials page, then place uncertain Demand into cell D5. The model should appear as shown on the right.

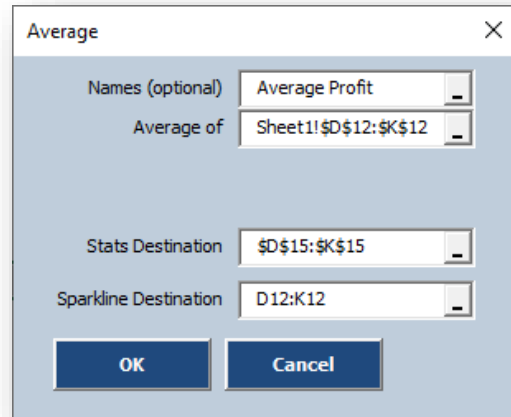
	A	B	C	D
1	Trial			
2		1 Cost of Capacity per Unit		\$30
3		Demand SIP Selling Price per Unit		\$40
4		SLS 4-10-21		
5		Actual Demand (Uncertain)		141,994

### Performing the Experiment

Find Simulated Average Profit for each capacity.


**Avg**  
Insert  
Average

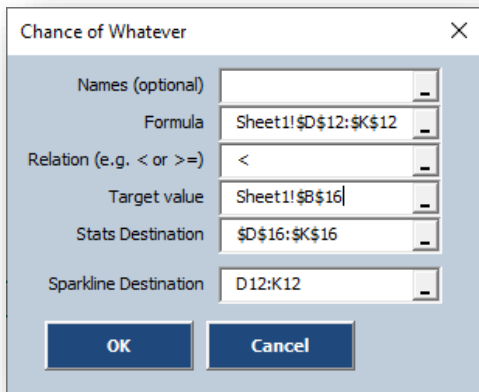
1.  Select cells D15:K15
2. Click the **Insert Average** button and select the range of Profit cells to enter in the Average of field, then click **OK**



If you did this correctly, you should see our old friend \$601,147 from Lesson 2 in cell I15.

### Find the Chance of Loss for each capacity


1.  Select cells D16:K16
2. Click the **Chance of Whatever** button and select the range of Profit cells to enter in the **Formula** field
3. Type a < sign in the **Relation** field
4. Select cell B16 in the **Target value** field
5. Click **OK**

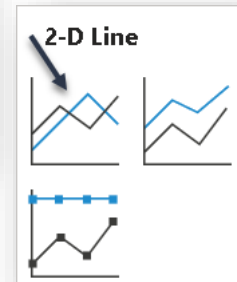


Cells C14:K16 should now contain the results below.

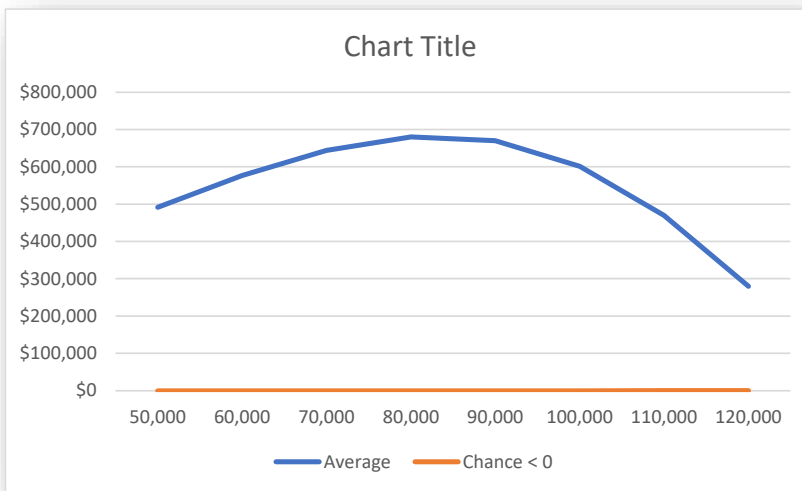
	50,000	60,000	70,000	80,000	90,000	100,000	110,000	120,000
Average	\$491,598	\$576,846	\$643,986	\$679,882	\$669,651	\$601,147	\$469,651	\$279,882
Chance < 0	1%	1%	3%	6%	10%	16%	24%	35%

## Creating a Chance-Informed Graph

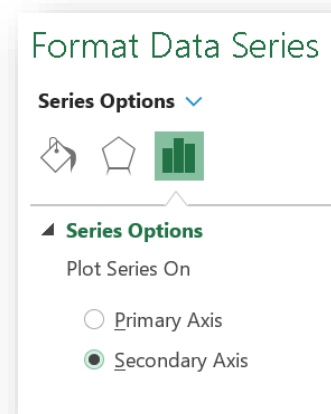
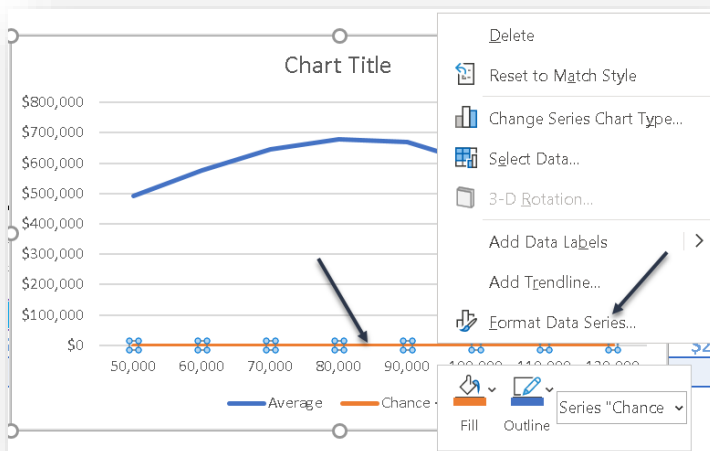
1.  Select cells C14:K16, then select the line graph from the Insert menu from the Excel ribbon.



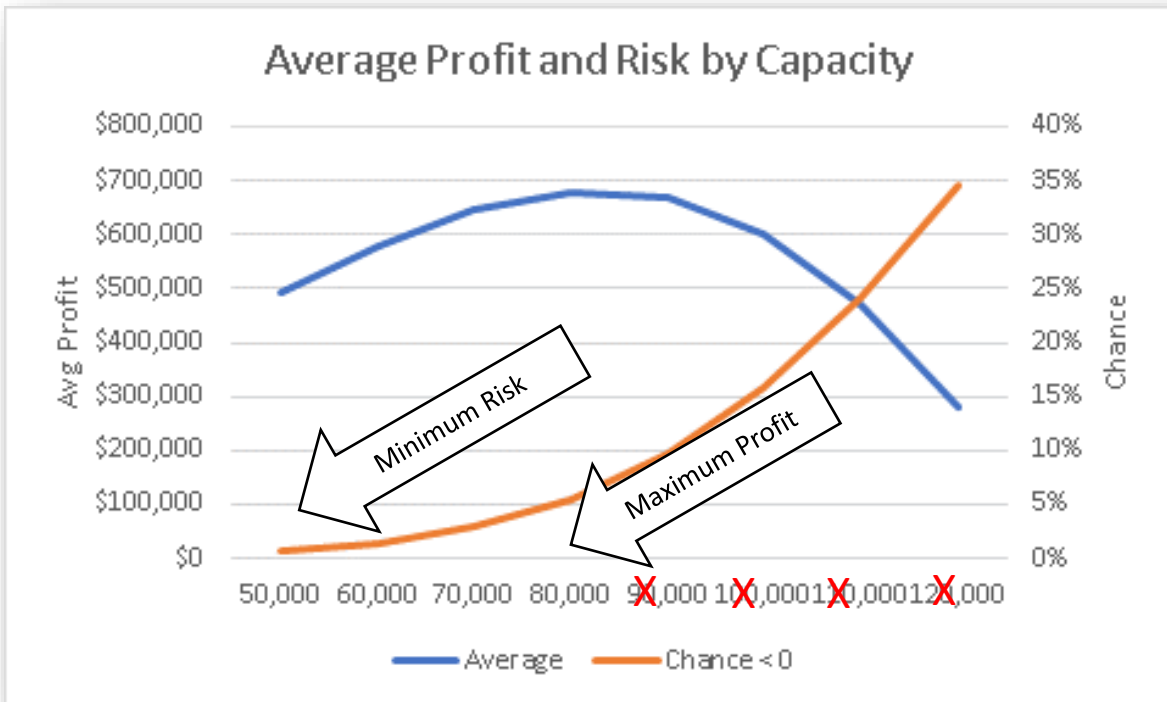
2. Your graph should look like this.



3. Because the chance of loss is so much smaller than the average profit, we will put them on separate vertical axes.
  - a. Select the "Chance" line on the graph, then **right click**.
  - b. Select **Format Data Series**.
  - c. Select **Secondary Axis** from the **Format** pane.



d. Add a title and axis labels and your graph should now look like this:



We see that a capacity of 80,000 maximizes average profit. Anything above that, including the 100,000 we started with, makes no sense because it reduces average profit while increasing risk. For capacities below 80,000 there is a rational tradeoff between profit and risk, so again the choice will be guided by the risk attitude of the decision maker.


## Lesson 5 – Failure Analysis, Multiple Identical Variables

- Predicting the Failure of Aircraft Parts
- Linking Simulations to Serial Numbers

### Predicting the Failure of Aircraft Parts

Consider an aircraft part, which on average fails and needs replacement every six flight hours. That is, it has a Mean Flight Hours Before Failure (MFHBF) of six. Now suppose there are six such parts on the plane, and if any one of them fails, the plane must be grounded and repaired. The expected time for this part to ground the plane is the minimum of six independent failure times. Replacing these with the mean (another word for average), we find that the minimum of 6,6,6,6,6 and 6 is 6 flight hours. But hopefully by now you're not going to fall for the Flaw of Averages, and you should be saying let's simulate the minimum of 6 failures.

Suppose we had a SIP of failure time for the part. That wouldn't work either. If all six parts refer to the same SIP, then they would all fail at the same instant, and although there would now be some variability, the average time to first failure would still be 6 flight hours. In the next example, we will use a Virtual SIP whose HDR seeds are tied to the serial numbers of the parts and tail number of the plane to estimate the chance that the time to the failure of the first part is at least 6 flight hours, and also what it is on average.

 Open **First Part Failure.xlsx** as shown. Note that we are only modeling the time it would take for Part 1 to ground the plane. Clearly, such models could be generalized to include much larger numbers of both identical and different parts.

## Linking Simulations to Serial Numbers

 **Load Flight Hours Library**



	A	B	C	D	E
1			Aircraft Tail Number		
2			103		
3			Time to	Serial	
4		Part type	Failure	Number	
5		Part 1	6	736451	
6		Part 1	6	736452	
7		Part 1	6	736453	
8		Part 1	6	736454	
9		Part 1	6	736455	
10		Part 1	6	736456	
11	Time to First Failure		6		
12					

1. Select cells C5:C10, then click the **SIP Input** button.
2. Copy the **Flight Hours to Failure Library** link from <https://www.probabilitymanagement.org/tutorial-libraries>
3. Click the **Use Clipboard** button, then select only **Part1**

1 Make sure the Destination refers to cells C5:C10.

2 Use Clipboard

3 IMPORTANT. Only select **Part1**, then click **OK**.

	A	B	C	D
			Aircraft Tail Number	
			103	
			Time to	Serial
		Part type	Failure	Number
		Part 1	6	736451
		Part 1	6	736452
		Part 1	6	736453
		Part 1	6	736454
		Part 1	6	736455
		Part 1	6	736456
	Time to First Failure		6	

- You will be prompted to specify a range of **Random Cells** (one for each of the six parts), and then a range of **Serial Numbers**, and the **Tail Number** of the plane, which will be used as seeds for the HDR random number generator.

4.1. You must specify a range of six blank cells, e.g. E5:E10, in which HDR random number formulas will be placed.

4.2. You must specify a range of unique Serial Numbers

4.3. You must specify a cell containing the **Tail Number**. Technically, this example would have worked with **Serial Number** alone, but multiple seeds are available. Then click **OK**.

	A	B	C	D	E
1	Trial		Aircraft Tail Number		
2		1	103		
3	Flight_Hours_to_Failure_Li	Time to	Serial		
4	SLS 4-10-21	Part type	Failure	Number	
5		Part 1	5	736451	
6		Part 1	5	736452	
7		Part 1	5	736453	
8		Part 1	5	736454	
9		Part 1	5	736455	
10		Part 1	5	736456	
11	Time to First Failure		5		
12					
13					
14					
15					

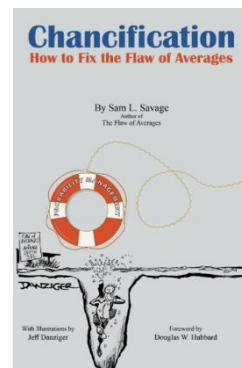
### Chance-Informed Model

11	Time to First Failure		6	>=	6	0.8%
12		Average	3.143			

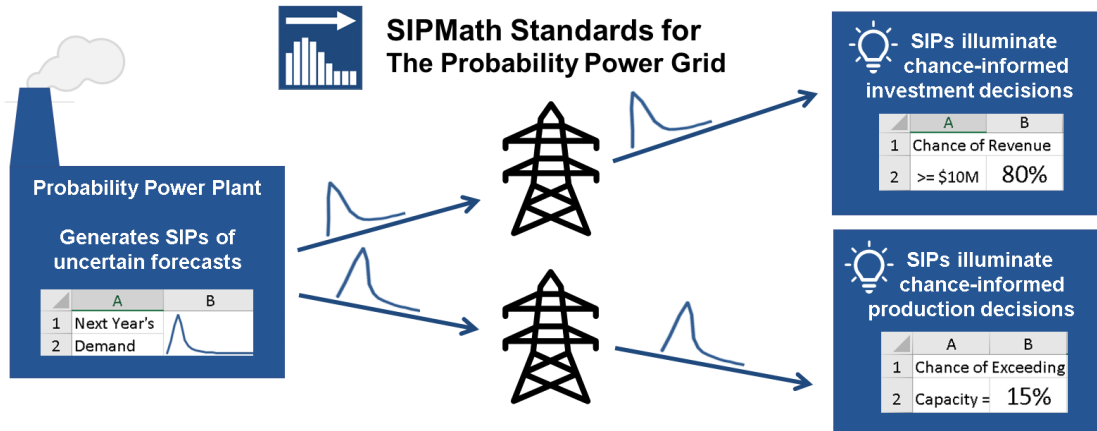
Using the **Insert Average** and **Chance of Whatever** buttons on the **ChanceCalc** ribbon demonstrated in **Lesson 2**, it is easy to demonstrate that there is virtually no chance of achieving 6 hours of flight time before failure of at least one of Part 1, and that on average the first failure will occur in just over three flight hours.

## Chancification and the SIPmath Standard

As described in my 2022 book, which is available on Kindle, Chancification is how we refer to the process of replacing calculations using numbers with calculations using probabilities, and it has some parallels with electrification. Just as electrification enables a small number of power engineers to supply electricity to millions of people for illuminating their homes, Chancification allows a small number of data scientists and statisticians to supply SIP Libraries to a wide audience of managers for illuminating chance-informed decisions.



**ChanceCalc** is the first application of its kind. Why? Because before the discipline of probability management, which established standards for communicating uncertainty as actionable data, this application would have been as useless as a lightbulb before electricity. The [SIPmath 2.0](#) and [recent SIPmath 3.0](#) Standards play roles analogous to the DC and AC current standards that ushered in the electric age.



For more I recommend visiting the [Standards Page of ProbabilityManagement.org](#)

and the article on [Chancification in OR/MS Today](#).

**SIPmath™ Standards**

Conveying Uncertainty as Data That Obey Both the Laws of Arithmetic and the Laws of Probability

May 11, 2021 in Probability Management

**Chancification: Wiring Your Organization for Probability**

By Sam L. Savage, Shayne Kavanagh, Aaron Brown

## SIP Libraries

**ChanceCalc** can read SIP Libraries of three formats: CSV (comma separated values) Files, SIPmath 2.0 XLSX files, and SIPmath 3.0 JSON files, either from your computer or from the cloud. CSV Files may be saved from **ChanceCalc** as described below. SIPmath 2.0 Standard libraries are XLSX files which may contain metadata in addition to SIPs. They may be created with the SIPmath Modeler Tools from ProbabilityManagement.org. SIPmath 3.0 libraries have a number of further advantages including vastly smaller file size. Furthermore, unlike the CSV and XLSX libraries, which have a fixed number of trials, SIPmath 3.0 libraries allow the user to specify the number of trials. SIPmath 3.0 libraries may be created with Python, R, or software such as Analytic Solver from Frontline Systems. From the users' perspective, the primary differences are described in the table below.

**Primary differences between CSV, SIPmath 2.0 and SIPmath 3.0 libraries from the users' perspective.**

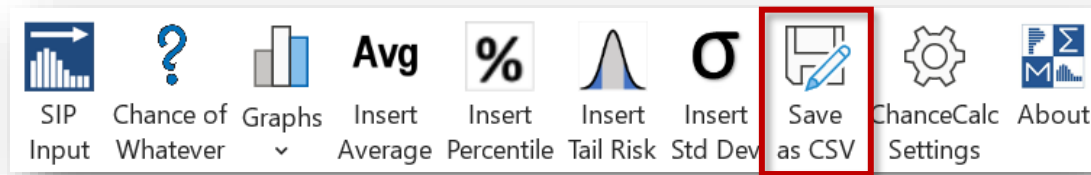
	Number of Trials	Metadata	File Type
CSV File	Fixed	NA	CSV Names in row 1




SIPmath 2.0	Fixed	Optional during creation	XLSX containing range names
SIPmath 3.0	User Specified	Optional during creation	JSON, may have .SIPmath extension

## Lesson 6 – Creating a CSV Library of Profit from Lesson 2

CSV SIP Libraries may be saved directly from **ChanceCalc** as follows.

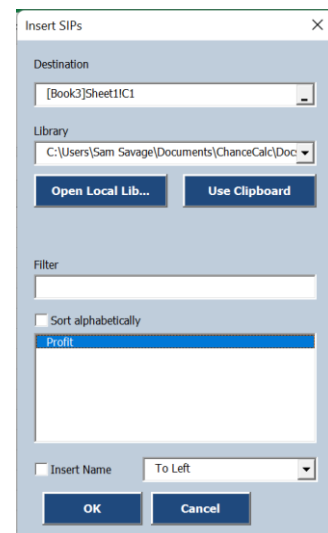


1.  Open the saved version of the Below Projection model from **Lesson 2** and click the **Save as CSV** button.
2. Navigate to a folder you can remember, give it the name of **Profit Lib.csv** and click the **Save** button.

File name:	Profit Lib.csv
Save as type:	CSV Files (*.csv)

3. To use the Library in another model, click the **SIP Input** button and select **Open Local Library**. Select **Profit Lib.csv**, select the **Profit** variable, and click **OK**. Although there will be no metadata available in cell A1, and no sparkline graph in the cell containing profit, all the functions of **ChanceCalc** remain available.

	A	B	C
1	Trial		1000000
2	Trial		



## Tools for Generating SIP Libraries

Below is a discussion of various packages that can generate SIP Libraries for **ChanceCalc**, starting with those from ProbabilityManagement.org.

### Package



SIPmath Modeler  
Tools for Excel

### Organization

ProbabilityManagement.org

The SIPmath Modeler Tools create Monte Carlo simulations using Excel's Data Table. This means that:

1. Unlike other simulation packages, you can run the models in native Excel without having the tools installed.
2. The models are interactive, running thousands of trials per keystroke.

There is both a free and Enterprise version, and both can generate their own random numbers or read and write SIPmath 2.0 SIP Libraries, which can be read by **ChanceCalc**. They are a sensible upgrade from **ChanceCalc** for those who want to move beyond using SIP Libraries generated by others to generating their own.

### Package



Analytic Solver

### Organization

Frontline Systems  
Solver.com

Frontline Systems was the first commercial software package to support SIPs and SLURPs in 2006, and their Analytic Solver was the first to both import and export SIPmath 3.0 Standard Libraries. It also features powerful optimization engines, and in particular can optimize under uncertainty. For more on Analytic Solver's support of the SIPmath Standard see the video on the right.



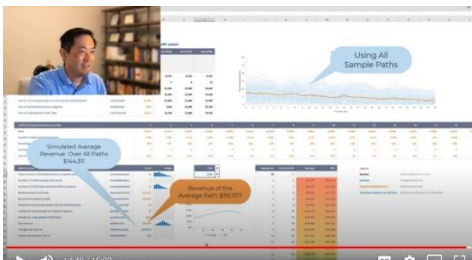
### Package



visionary modeling

### Organization

Lumina.com



Analytica is a powerful multi-dimensional modeling language, which can deliver a complete probability management solution, including simulation and optimization. It has recently been used to generate SIPmath 2.0 SIP Libraries from hospitalization forecasts from the CDC. See the video on the left.

**Package**

All the following packages can run Monte Carlo simulations and save the results as CSV files.

**Organization**

**Package**



**Organization**

Open Source

**Package**



**Organization**

Open Source

**Package**



**Organization**

Oracle Corp.  
oracle.com/applications/crystalball/

**Package**



**Organization**

Palisade Corp.  
palisade.com

## Appendix: Worksheets added by ChanceCalc

ChanceCalc models work in native Excel without the add-in. This is accomplished by adding several types of worksheets to your workbook as you create a chance-informed dashboard.

	A	B	C	D	E	F	G
1	Trial		5.91065				
2		1	4.8185				
3	2. Gaussian Copula		8.84994				
4	SLS/DGE		6.75224				
5			4.8185				
6			6.553023				

Contains an Intermediate Calculations

Contains settings for Histograms and Cumulative Graphs of simulation ..

Contains an Excel Data Table for performing simulation.

Contains Data from a SIP Library.

<sup>i</sup> <https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/hospitalizations-forecasts.html>

<sup>ii</sup> <https://www.probabilitymanagement.org/20-standard>

<sup>iii</sup> <https://www.probabilitymanagement.org/30-standard>