A SunCam online continuing education course

Engineering Methods in *Microsoft Excel*

Part 5: Simulation and Systems Modeling II

by

Kwabena Ofosu, Ph.D., P.E., PTOE

Engineering Methods in Excel
A SunCam online continuing education course

Abstract

This course is part of a series that presents *Microsoft Excel* methods that are useful for a wide range of engineering analyses and data management. This course presents simulation and systems modeling. Simulation is a technique for conducting experimentation of a system or process, virtually, on a computer. This course presents the fundamentals of the Monte Carlo simulation technique, and demonstrates, in detail, the framework for applying the technique to model a real engineering system, on an *Excel* spreadsheet. This course presents techniques for analyzing the results and how they are interpreted and applied in decision making. This course presents techniques for selecting and validating statistical distributions that are used to model the elements of the system being simulated, and how they are implemented in *Excel*. Upon completion of this course, practitioners will be able to apply the methods learned to a variety of engineering problems, and also to identify situations in their fields of specialty where the innovative application of these tools and methods will be advantageous to their output and to their work product.

List of Figures

TABLE OF CONTENTS

1. INTRODUCTION

1.1 Overview

Engineers are constantly challenged with solving a wide range of complex analytical and computational problems in their fields of specialty. These problems involve analyses methodologies and the management of data. The application of computers enables repetitive, time-consuming and often tedious calculations to be conducted rapidly, efficiently, and less prone to errors. The application of computer tools also enables the results and outputs of such engineering analyses to be readily transferred and incorporated into reports and other engineering documents. An even greater advantage, in terms of productivity and efficiency, is realized when these calculations and outputs are replicated across numerous projects. Competence in computer skills predisposes engineers to pursue and develop more creative and innovative solutions to problems than otherwise.

Microsoft Excel is widely and increasingly being used as a tool to assist engineers in conducting and replicating intricate calculations and analyses, designing complex systems, and managing large data sets. *Microsoft Excel* is an electronic spreadsheet program developed by the *Microsoft* Company, and part of the *Microsoft Office* suite of products. A spreadsheet is a grid that organizes data and calculations into columns and rows. The intersection of a column and a row is called a cell. An electronic spreadsheet enables users to store, organize, manipulate, and analyze data in the cells of the spreadsheet.

This course presents fundamental principles and engineering applications of simulation and systems modeling, and demonstrates the *Microsoft Excel* tools, methods, and strategies that can be used to simulate and model real-life engineering systems. Simulation techniques involve conducting virtual experimentation of a system or process, on a computer, using mathematical and statistical models. This course presents how to formulate, implement and solve simulation and systems modeling problems in *Microsoft Excel*.

Upon completion of this course, participants will have gained insight into applying *Excel* tools, methods, and strategies in formulating, implementing and analyzing simulation models. Participants will also be able to identify professional situations where the application of these innovative *Excel* techniques will be of great benefit and advantage, and will enable practitioners to significantly improve their productivity and the quality of their work product.

2. SIMULATION

2.1 Introduction

Simulation is a statistical analysis tool used in many fields such as the sciences, engineering, business, management and many others. Simulation techniques have been used to study a wide variety of problems and complex physical phenomena such as traffic congestion, the spread of public health epidemics, weather forecasting, the performance of financial markets, military operations, emergency response scenarios, the quality of product or duration of an industrial or manufacturing process, etc., etc. Simulation has been used to solve mathematical problems for which a direct solution is impractical.

If the processes being **simulated** involve an element of probabilistic behavior (random chance) the simulation is referred to as **Monte Carlo simulation**. Many complex and large-scale engineering problems are amenable to Monte Carlo simulation due to the fact that the simulation technique can handle large numbers of **random variables**, numerous statistical distributions, and nonlinear mathematical models.

Monte Carlo simulation is useful in situations where direct experimentation of a system or process is impractical, infeasible, time or cost prohibitive, or simply impossible. For example, it is not possible to conduct an experiment on the spread or impacts of a highly contagious disease outbreak on a large or densely populated U.S. city. Likewise, it would be impractical or time prohibitive for a light bulb manufacturer to test a large sample of a product over the product's entire design life. In Monte Carlo simulation, the experimentation is conducted rapidly and many times over on a computer using a simulation model. The incorporation of random variables in the simulation model to describe the discreet elements of the system or process enables many multiple scenarios of the system or process to be modelled, thus synthesizing model output data. The model output data set is then analyzed by appropriate statistical methods, to draw conclusions and to make decisions and recommendations about the system or process.

Thus, Monte Carlo simulation enables one to see all possible outcomes of a process and make decisions that take into account the probability (uncertainty or risk) associated with the discrete elements of the process.

2.2 Illustration of a Simulation Model

In this section a very simple example shall be used to illustrate a Monte Carlo simulation model.

Consider a small local contractor hired to prepare a site for a national company to come in and build a luxury apartment complex. The local contractor's activities, or work items, include clearing the site and setting up trailers for site offices and other facilities. The contractor breaks down the project into the following discrete tasks in sequential order, and their estimated completion times.

	Work Item	Description	Completion Time (days)
1.	Clearing and Grubbing	Clear the site of vegetation, pull tree stumps, and remove all debris.	10
2.	Furnish and install trailers	Supply trailers and all ancillary work to complete installation.	5
3.	Connect services	Hook up power, utilities, and telecommunication services	7
4.	Furnish and install interior amenities	Install bathrooms, kitchenette equipment, office furniture, security devices	4
5.	Final cleanup	Final cleanup of interior and exterior	2
	Project completion		28

The completion times are based on the local contractor's recent experience with similar projects, clients and suppliers. In this approach, the local contractor considers these times to be set and fixed to yield the overall project completion time. This is called a **deterministic** approach. Prior to any knowledge or experience in Monte Carlo methods, the deterministic approach would be the intuitive approach to handle this type of problem.

The fact that the deterministic approach considers the work items' completion times to be set and fixed is a major weakness of the approach. Practically, all of the work items' completion times are inherently **random** to varying degrees. For example, the completion of the clearing and grubbing is subject to randomness (or **uncertainty**) due to factors such as weather, equipment reliability, availability and punctuality of earth moving equipment and the operators, etc. etc. The work item to furnish and install the trailers is subject to randomness due to traffic conditions, availability or schedule of escort vehicles required to accompany the truck(s) that bring the trailer(s) to the site, as well as the randomness associated with obtaining the permits from governmental agencies to convey oversized cargo on state roads. Once the trailer(s) reaches the site, further uncertainty may be due to soil conditions that may slow down (or speed up for that matter) the process of anchoring and securing the trailer(s). There is also randomness associated with the quality of the workmanship, or lack thereof that may necessitate rework or additional inspections prior to approval. The uncertainty associated with each work item contributes to the uncertainty associated with the overall project completion time. The degree of uncertainty is called **risk**, and exposes the local contractor to potential unfavorable outcomes, particularly financial loss.

It is therefore justified to take a **probabilistic** approach to the analysis in order to incorporate the uncertainty associated with each work item's completion time. This can be done by describing each work item's completion time with an appropriate **random variable** (also called a **probability distribution**). The work items' completion times can be given at certain probabilities based on the appropriate probability distribution for that work item. Overall project completion times can then be determined at given probabilities. The local contractor can now select an acceptable level of uncertainty (or risk) and reach a more informed conclusion regarding the project completion time.

Using assumed probability distributions, the work items' completion times for this project can be described probabilistically as follows

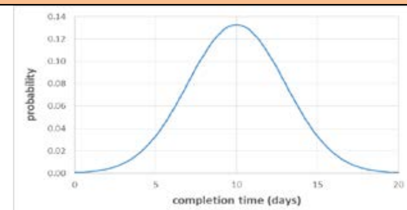
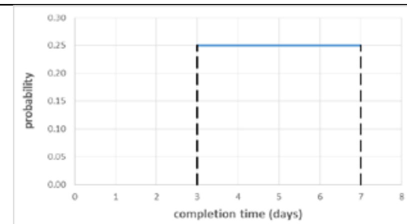
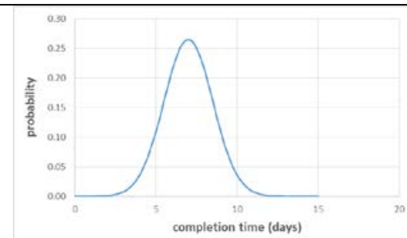
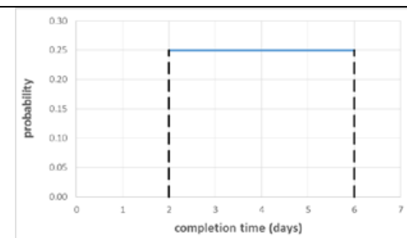
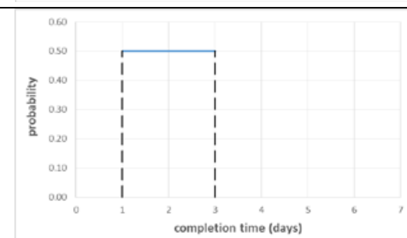
	Work Item	Probability Distribution	Completion Time (days)
1.	Clearing and Grubbing	Normal distribution Parameters: Mean (μ) = 10 Standard deviation (σ) = 3	
2.	Furnish and install trailers	Uniform distribution Parameters: α = 3 β = 7	
3.	Connect services	Normal distribution Parameters: Mean (μ) = 7 Standard deviation (σ) = 1.5	
4.	Furnish and install interior amenities	Uniform distribution Parameters: α = 2 β = 6	
5.	Final cleanup	Uniform distribution Parameters: α = 1 β = 3	
	Total (Project) Duration		$\sum(completion\ times)$

Figure 2. 1: Framework for simulation

To generate a possible scenario for a particular work item's completion time, the contractor can randomly pick a probability value and read off the completion time from the probability distribution graph (or calculate it from the graph function). The process is repeated for each work item, and the project completion time for the scenario is the aggregate of the individual completion times obtained. This process can now be replicated several times over to synthesize a data set of project completion times. The project completion time data can be analyzed to obtain the descriptive statistics, quartiles and percentiles, or other statistical measures which are used for appropriate decision making.

Although it is not a required prerequisite for this course, it will be helpful if readers take a moment to review and refresh their knowledge on random variables and statistical distributions. An earlier part of this course series provides a preparatory in-depth presentation of those topics.

3. THE SIMULATION PROCEDURE

3.1 Monte Carlo Simulation

The Monte Carlo simulation procedure can be concisely described by the following steps.

Step 1:
Take the first work item, which is a random variable (X), guess a random probability value (F). Use the inverse cumulative density function (also known as the percent point function or the quantile function) of the random variable that describes the work item (X), to determine the corresponding value (x) (an actual completion time) of the random variable (X) for that probability value (F).

Step 2:
Repeat Step 1 for each of the other work items.

Step 3:
Sum the x values (completion times) to generate the overall completion time for the current simulation (or iteration).

Step 4:
Repeat Step 1 through Step 3 $(n - 1)$ times to synthesize a dataset of completion times of size n for the n simulations conducted.

Step 5:
Compute statistical measures for the synthesized completion time dataset, such as mean, median, standard deviation, quartiles, percentiles etc.
Use the statistical measures to make conclusions, recommendations, decisions etc., based on engineering performance criteria, industry standards, regulatory requirements, business and financial considerations etc., etc.

It is pertinent to note that in practice, the number of simulations (n) needed to be conducted is not subject to hard and fast rules. It is determined by judgement or experience, based on factors such as the type and complexity of the system being simulated, as well as the complexity of the breakdown of the work items, which itself may be driven by the available data. A small-scale civil engineering model may use a few hundred or a few thousand simulations, whereas a large-scale mechanical system may need several hundred-thousand simulations to synthesize good data for appropriate decision making.

The Monte Carlo simulation procedure can be depicted as follows.

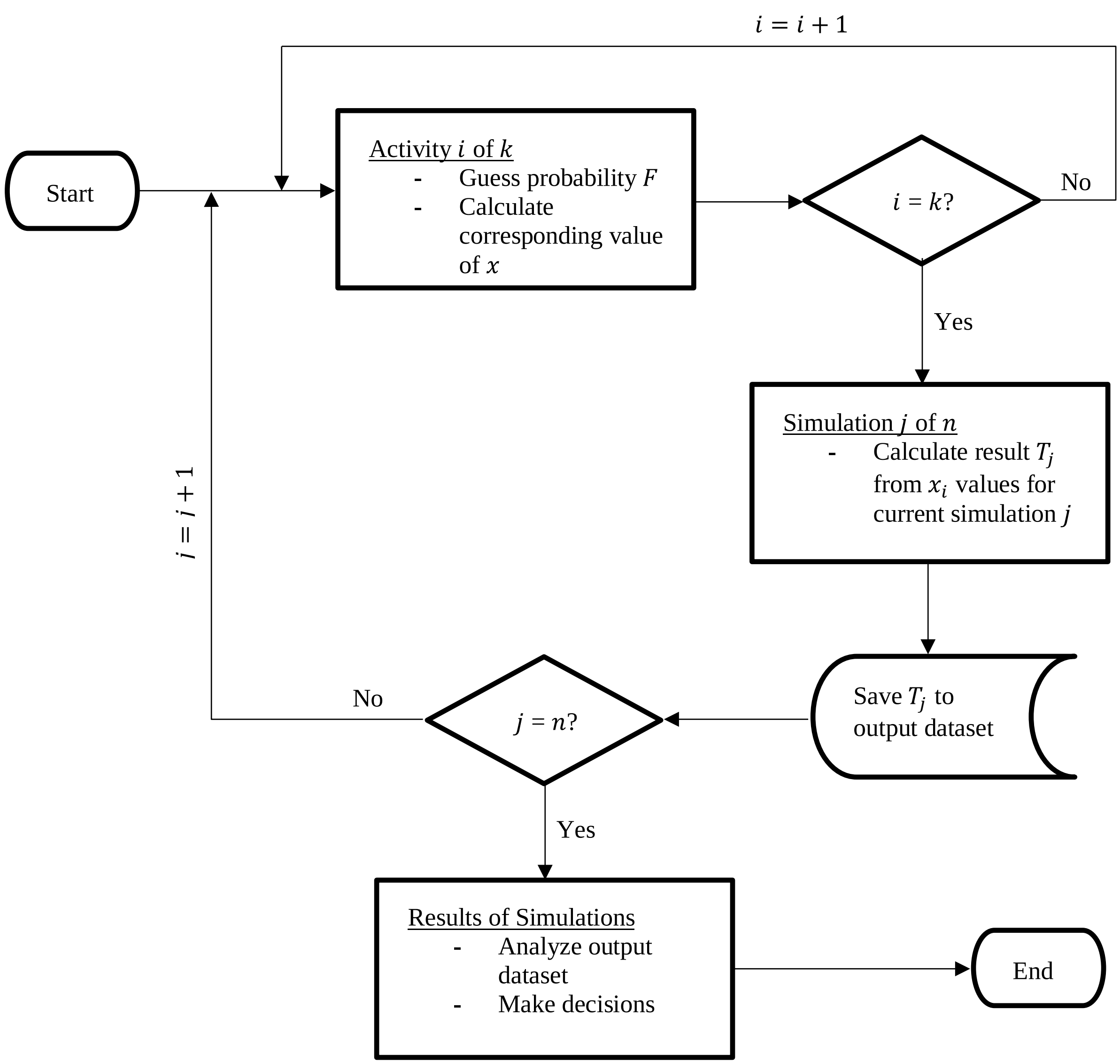


Figure 3. 1: Monte Carlo simulation procedure

Conducting the Monte Carlo procedure by hand in a tabular format will be of the following structure.

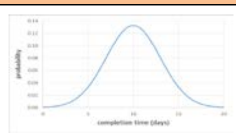
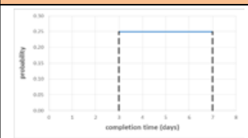
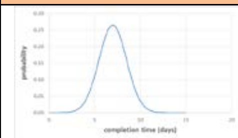
Simulation	Activity 1	Activity 2	Activity 3	…	Completion Time
				…	
	Distribution: Parameters, Quantile function	Distribution: Parameters, Quantile function	Distribution: Parameters, Quantile function	…	
1.	guess probability F_{11}	guess probability F_{12}	guess probability F_{13}	…	
	calculate duration x_{11}	calculate duration x_{12}	calculate duration x_{13}	…	$T_1 = \sum x_{1i}$
2.	guess probability F_{21}	guess probability F_{22}	guess probability F_{23}	…	
	calculate duration x_{21}	calculate duration x_{22}	calculate duration x_{23}	…	$T_2 = \sum x_{2i}$
3.	guess probability F_{31}	guess probability F_{32}	guess probability F_{33}	…	
	calculate duration x_{31}	calculate duration x_{32}	calculate duration x_{33}	…	$T_3 = \sum x_{3i}$
:	:	:	:	…	:
$n.$	guess probability F_{n1}	guess probability F_{n2}	guess probability F_{n3}	…	
	calculate duration x_{n1}	calculate duration x_{n2}	calculate duration x_{n3}	…	$T_n = \sum x_{ni}$

Figure 3. 2: Monte Carlo simulation tabulation

The synthesized completion times (T_j) can now be analyzed using appropriate statistical measures from which conclusions, recommendations, decisions, etc., can be made.

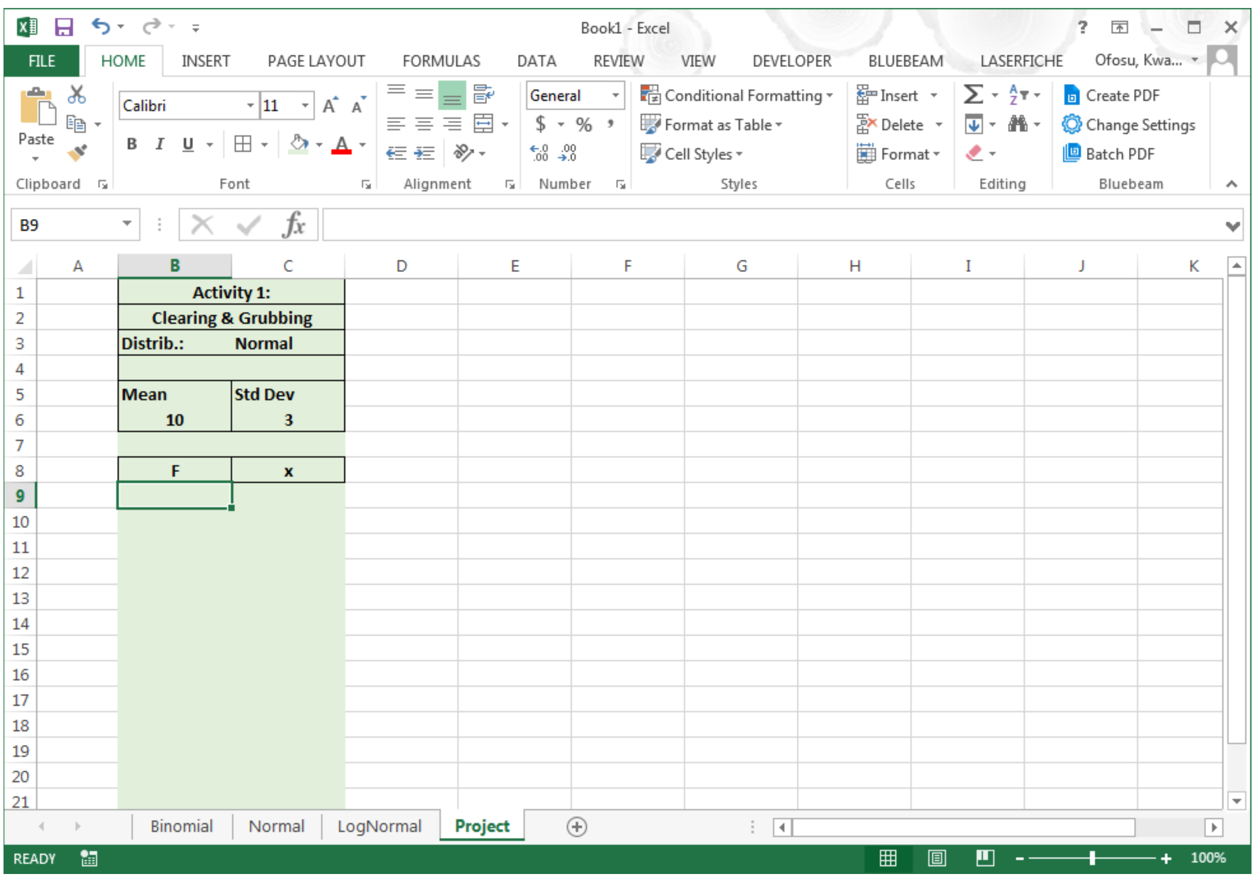
4. MONTE CARLO SIMULATION IN *EXCEL*

4.1 Implementation in *Excel*

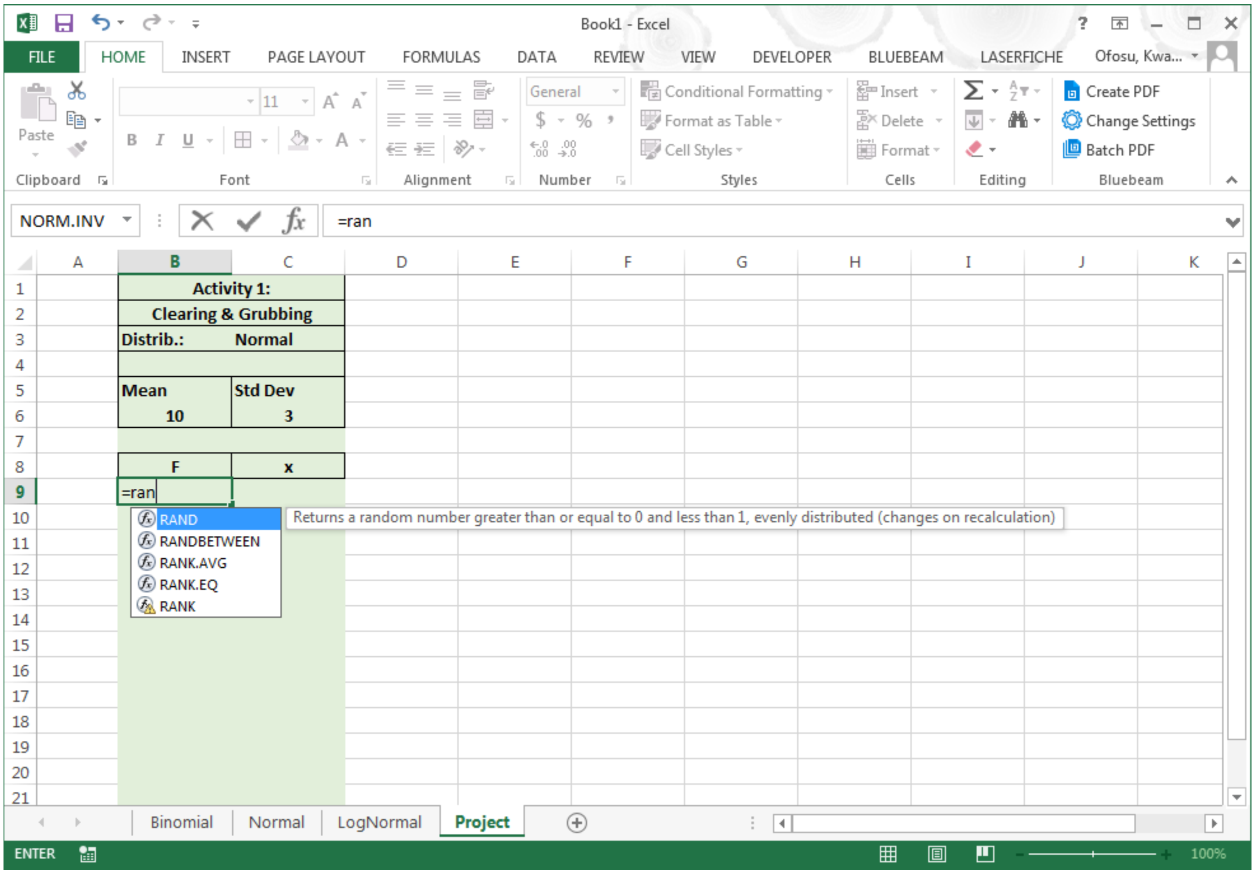
In this section, the site clearing project described in Section 2.2 shall be implemented in a Monte Carlo simulation model, in Microsoft *Excel* following the tabulated framework depicted in Figure 3.2. The details of the project activities and their descriptions and statistical parameters are shown in Figure 2.1.

Open a new session of *Excel*.

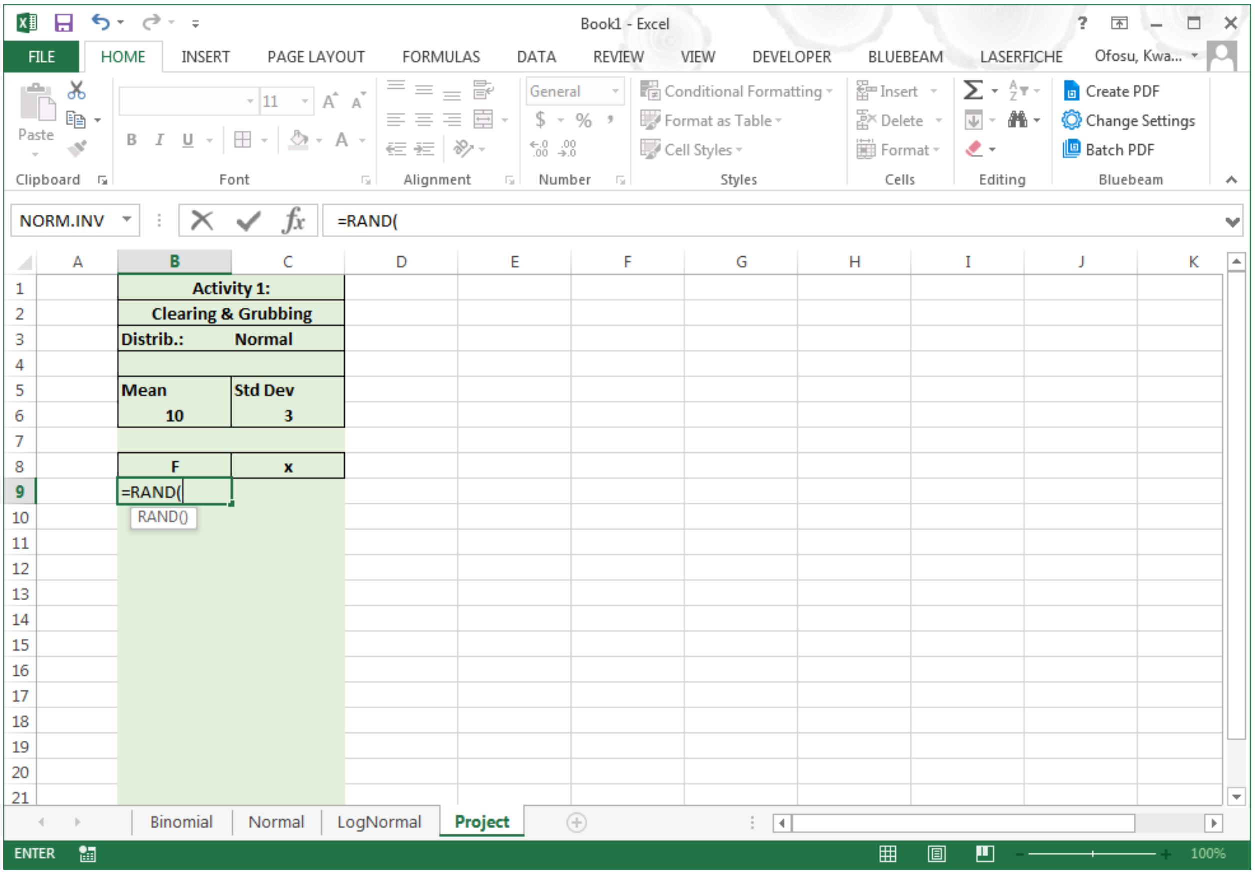
Set up the following headers for the first work item, Activity 1: Clearing and Grubbing.



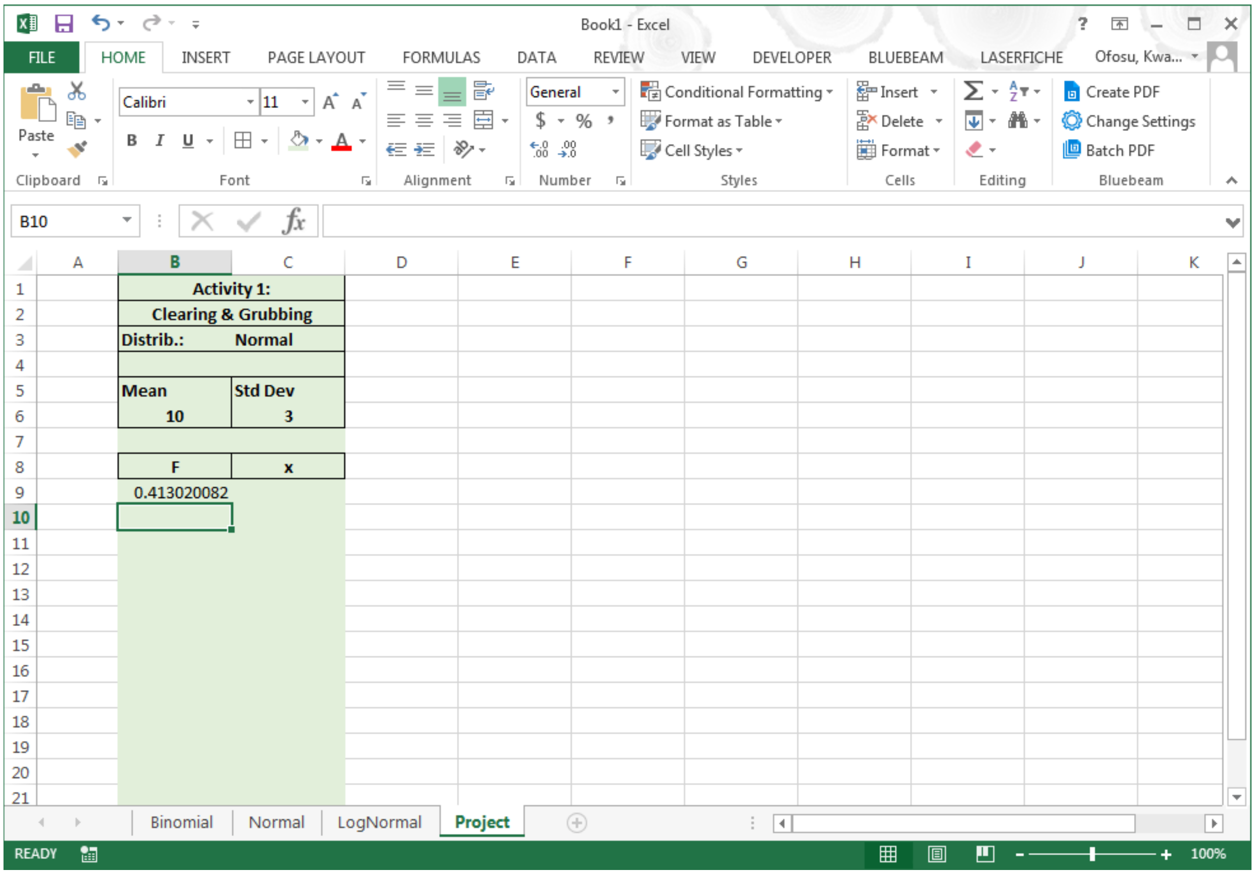
To conduct a simulation of this activity, call the *Excel* random number generator function RAND as follows.



Double click on RAND from the candidate list.



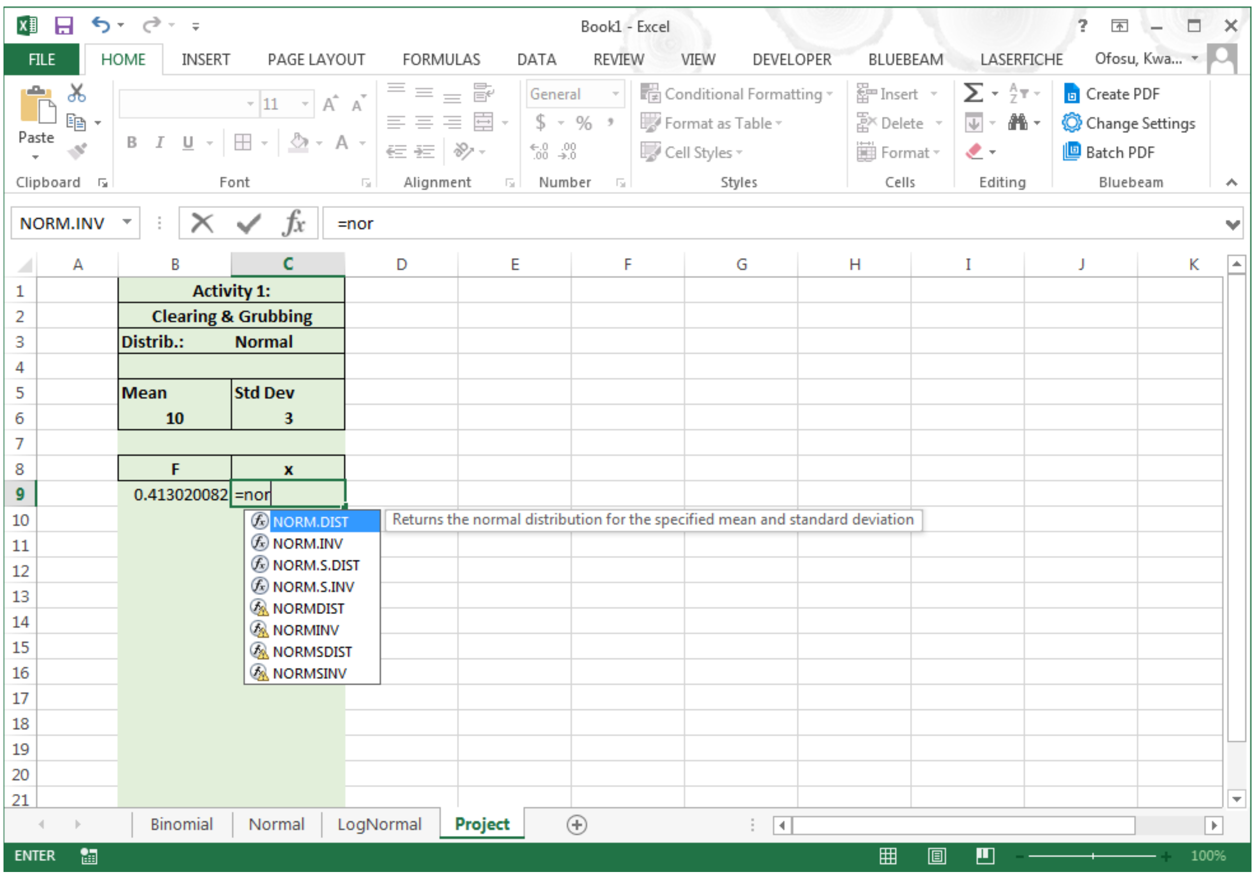
Close parenthesis.
Hit **Enter** on the keyboard.
The function is implemented in the current cell and a random value generated.



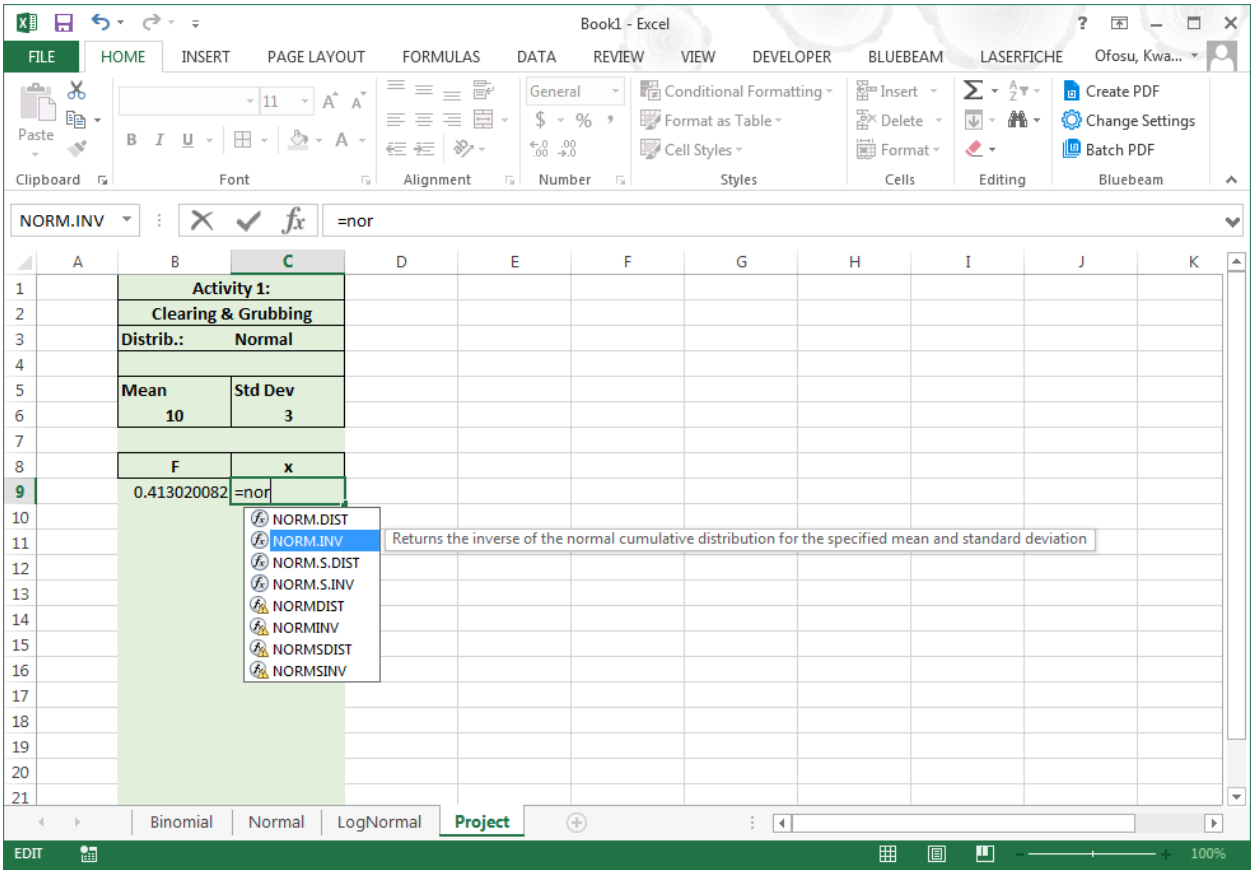
ATTENTION! Due to the random nature of the output generated by the RAND function, one may get a completely different value than shown in this demonstration. Also, one will notice the random value changes intermittently automatically whenever some other operation is conducted elsewhere on the spreadsheet. For now, both of the above are normal and not a problem. Do not be alarmed. The focus for now is on the procedure and implementation of the simulation model.

The next task is to call the inverse cumulative density function (also known as the percent point function or quantile function) of the distribution that describes the random behavior of the activity, to calculate the x value (completion time of the activity) corresponding to the random probability value F previously generated. This first activity is described by the Normal distribution.

With the cursor in the x value cell, begin typing "NORMAL" to activate a candidate list of relevant functions.

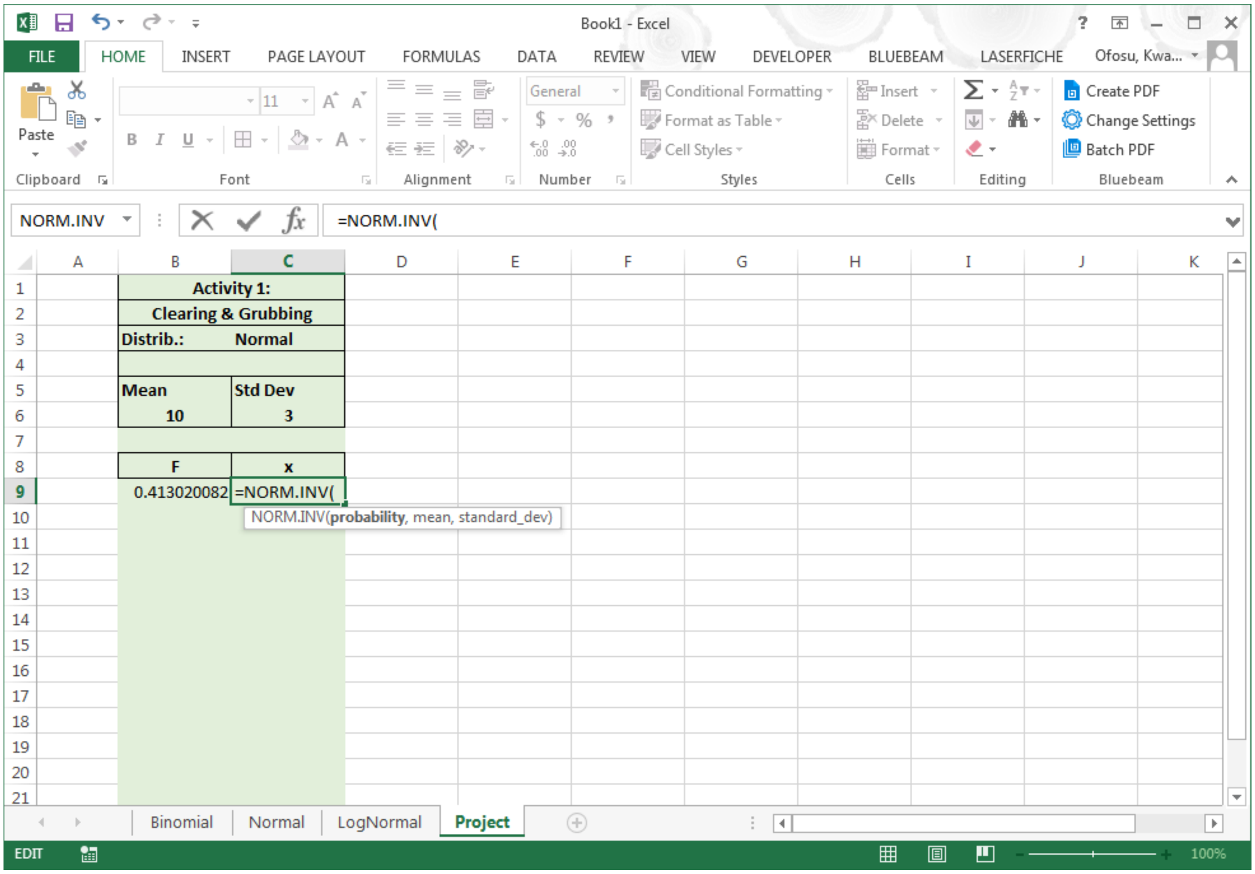


Select NORM.INV and review the tooltip.

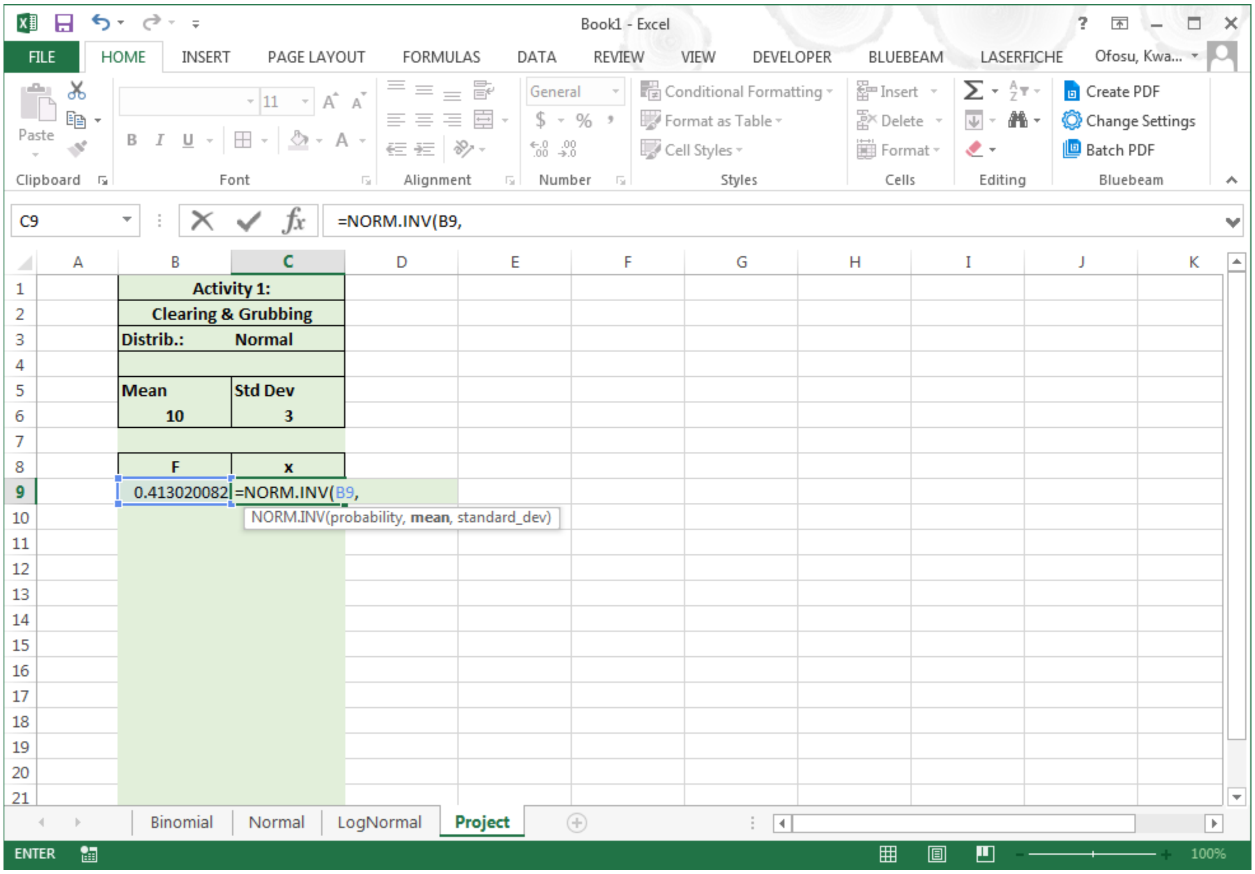


This is the function needed.

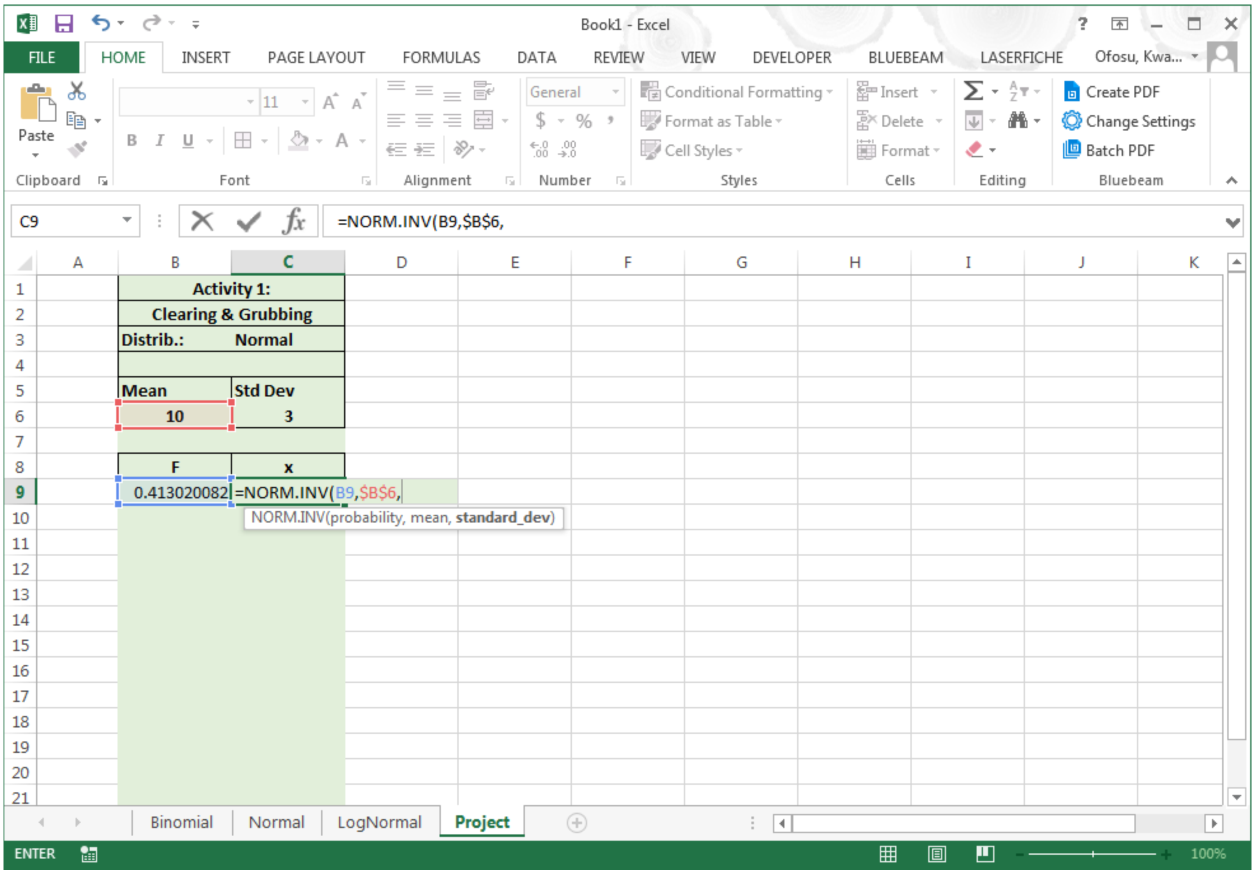
Double click on NORM.INV.



For the **probability** value, point and click on the *F* value.
Type a comma (“,”).

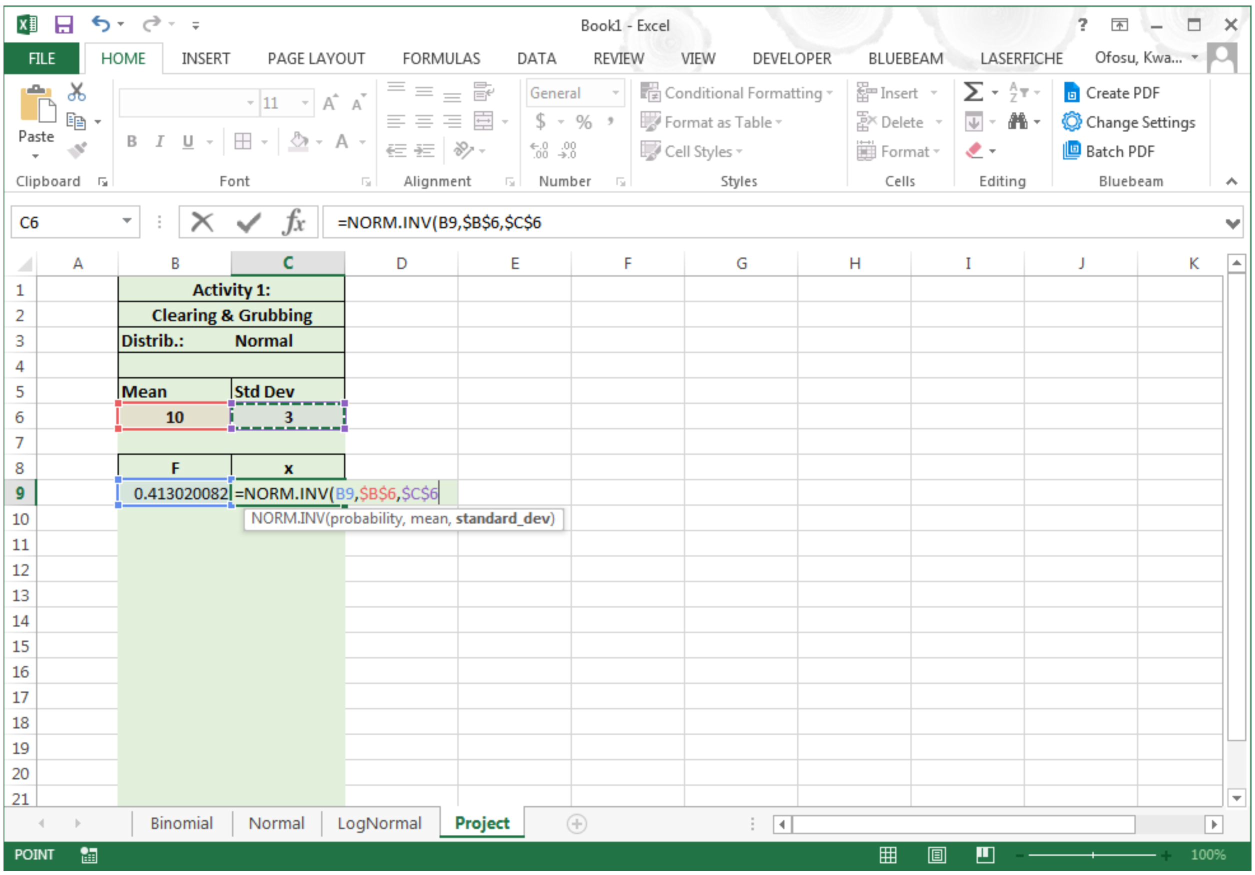


For the **mean** value, point and click on the mean value.
Hit **F4** on the keyboard.
Type a comma (“,”).

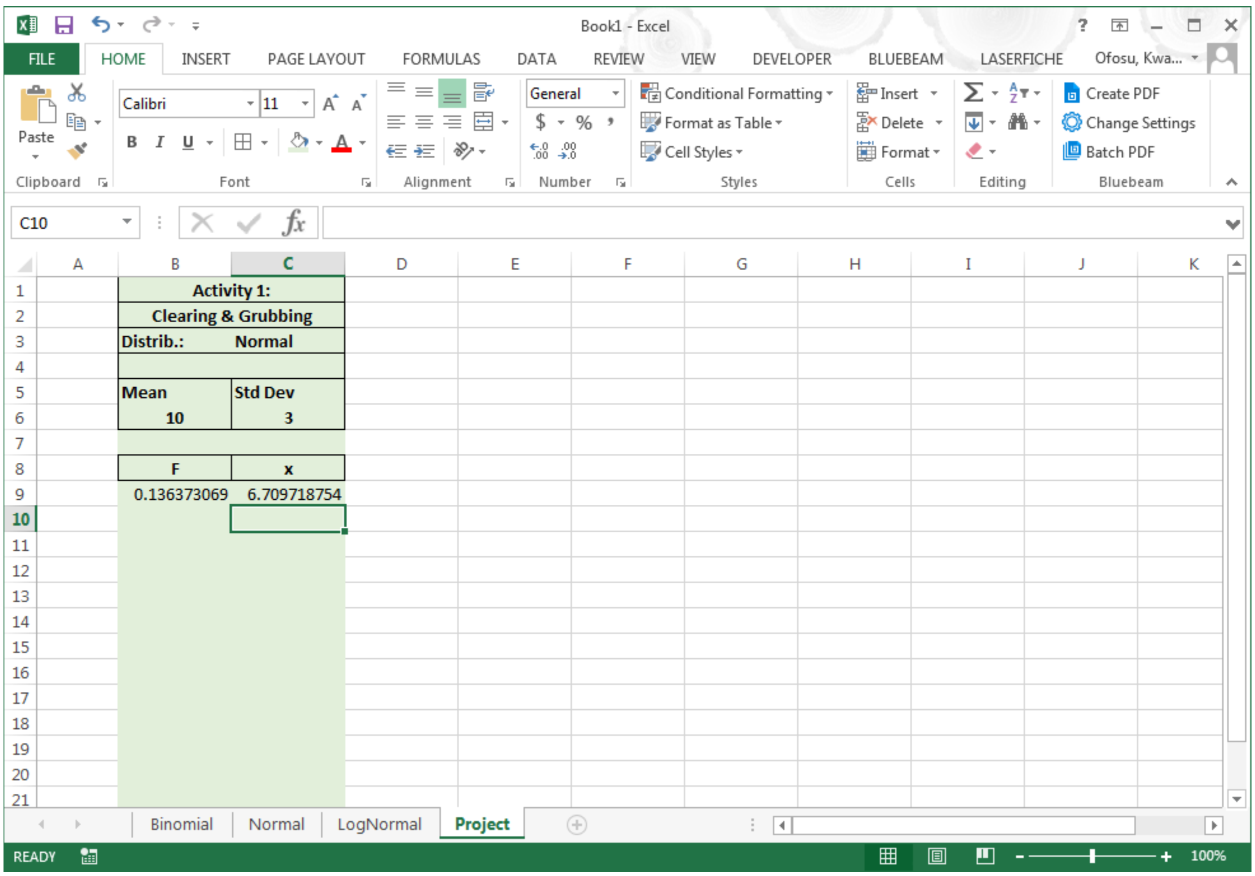


The **F4** button is used to make the cell reference to the mean value an **Absolute Reference**. This will prevent the mean value cell reference from changing, albeit inadvertently, later when the formula is copied and replicated down the table for the subsequent simulations.

For the **standard_dev** value, point and click on the standard deviation value.
Hit **F4** on the keyboard.
Type a comma (“,”).



Close parenthesis.
Hit **Enter** on the keyboard.
The x value is calculated.

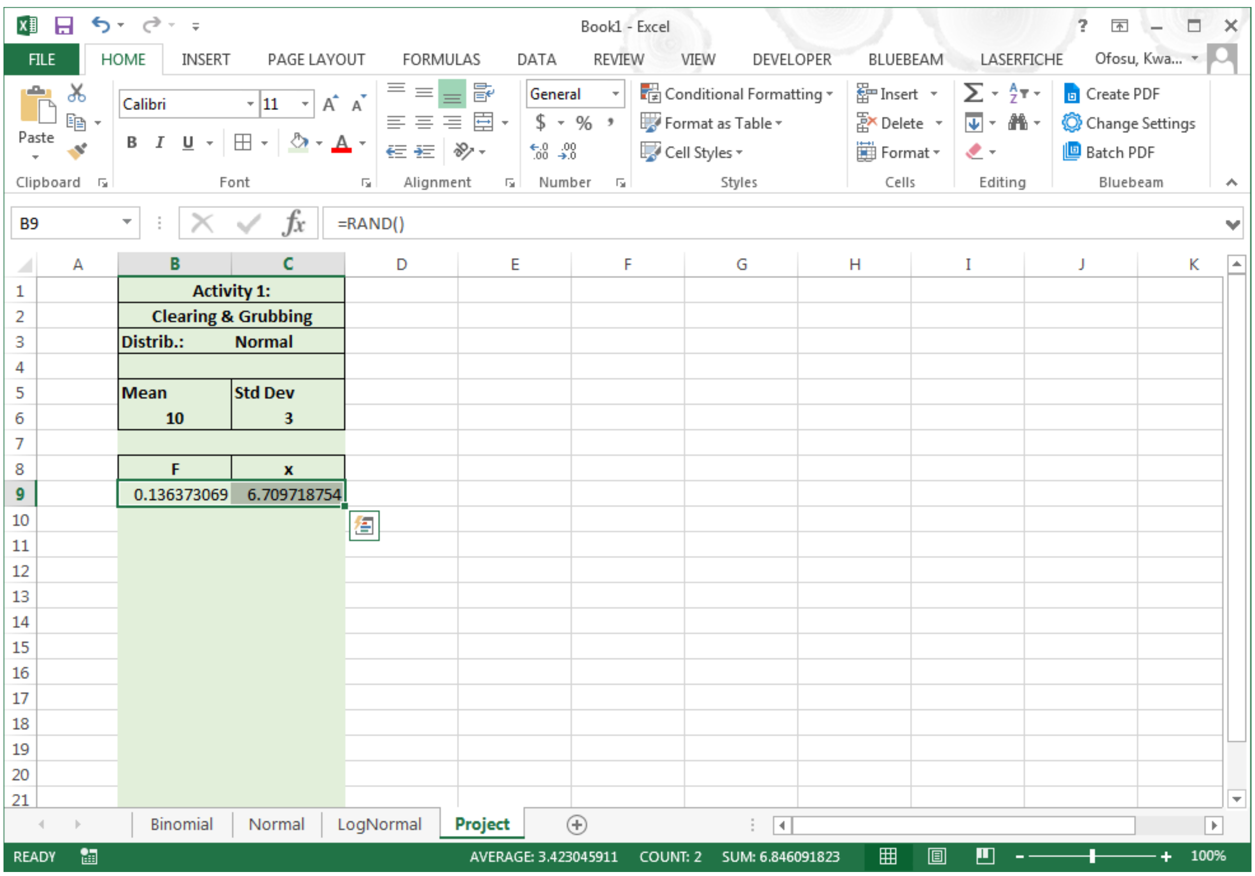


Note that the original random number has changed, but due to the cell referencing used in the NORM.INV calculation, the corresponding value (x) for the current probability value (F) is computed and displayed.

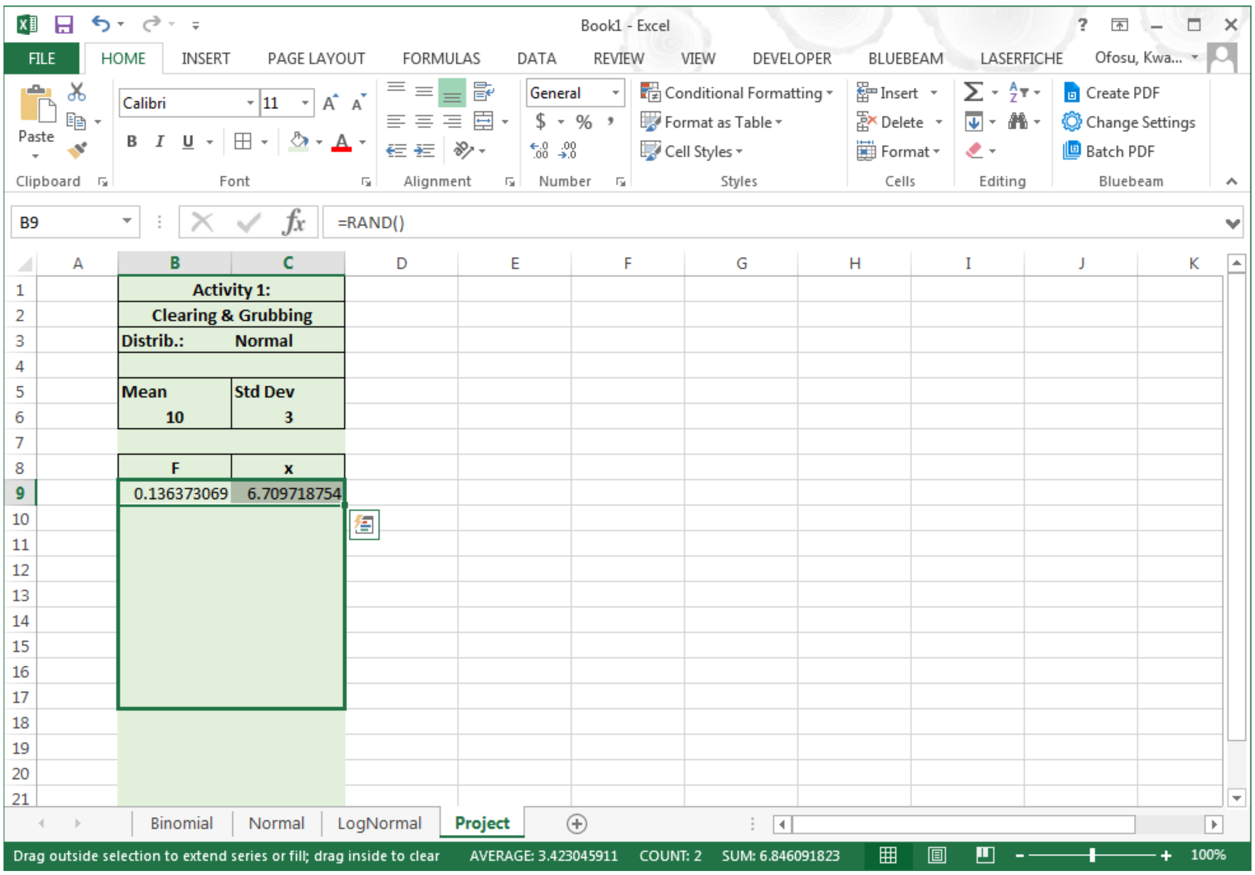
This result, stated in ordinary language, is that there is 0.136 probability that the activity will be completed in less than or equal to 6.709 days. Or one could say there is 0.136 probability that the activity's completion time will not exceed 6.709 days. (Again! The actual values you get may be completely different from those depicted here due to the random numbers being generated and passed into the formula).

One may now explore other scenarios (or simulations) of the activity completion time based on other random probability values.

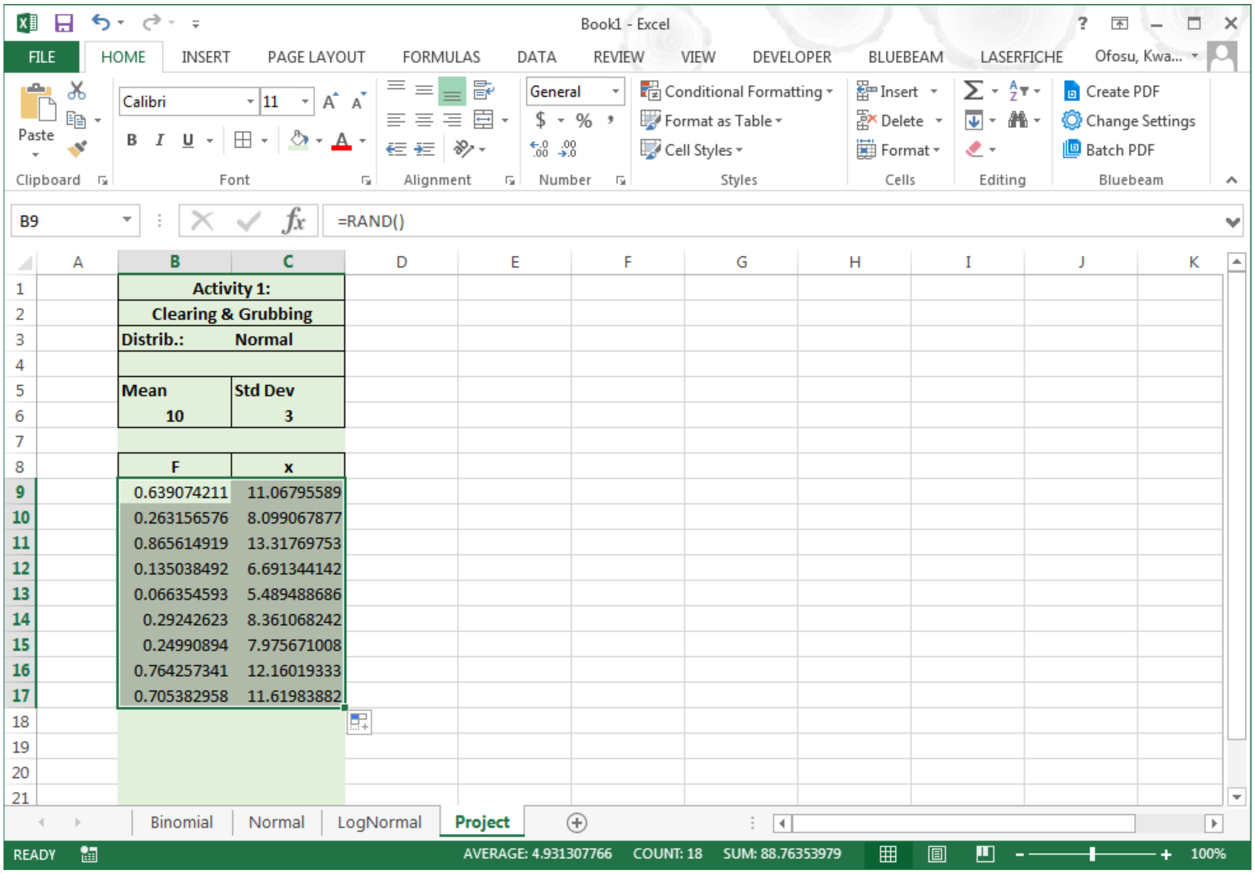
Select the F and x values previously calculated.
Hover over the bottom right border of the selection.
The cursor will change from a thick plus sign to a thinner plus sign.



With the thinner plus sign on the bottom right corner of the selection, click and hold the click down.
Drag the cursor down a few rows.



Release the click.
The original formulas/ calculations are replicated down the sheet.



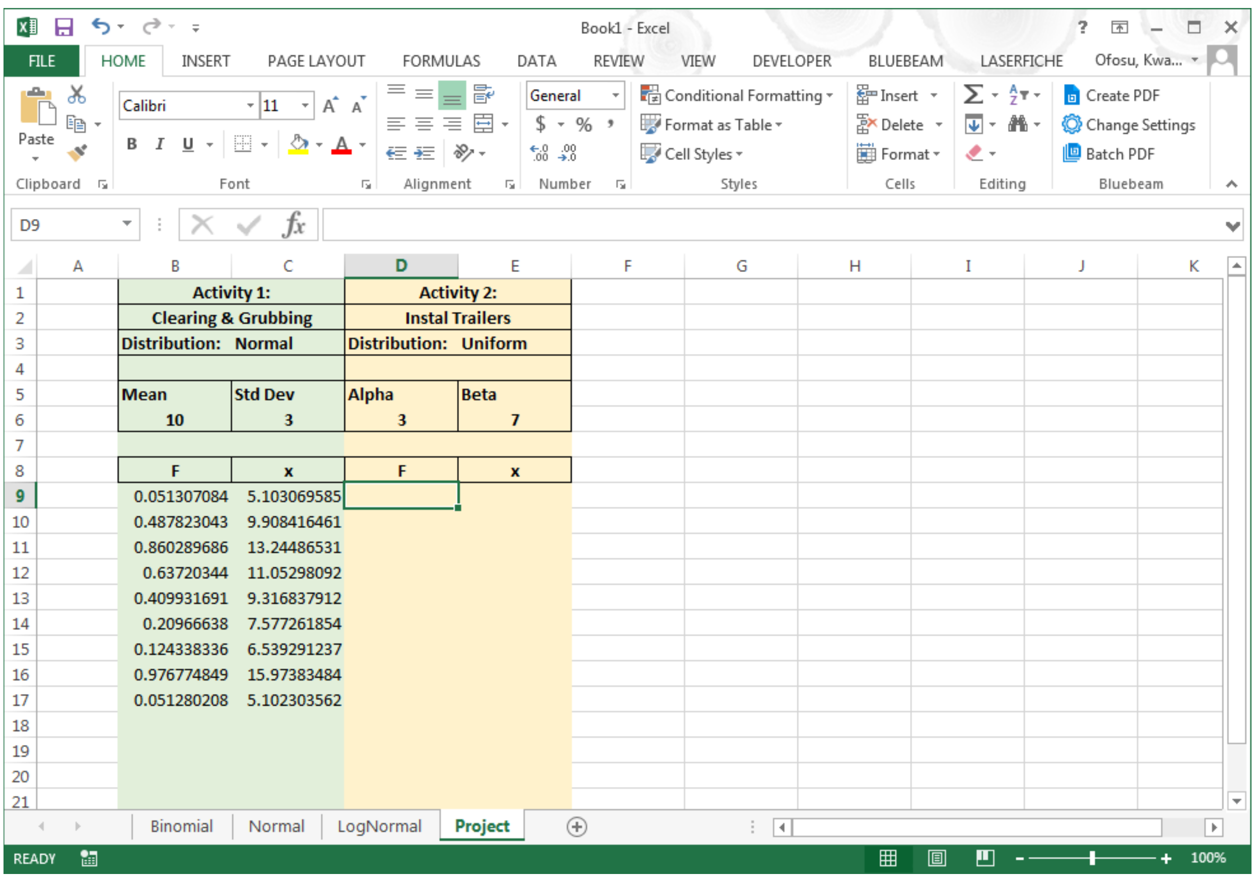
The process of selecting cells, and then clicking, holding down and dragging to replicate the formulas/ contents of the cell is the *Excel* technique called **fill handling**.

Review the probabilities (F) of the completion times (x) of the scenarios (simulations) generated. Again, one may obtain completely different numbers than depicted due to the random numbers being generated and passed into the formulas. Again, notice that the original random value has changed, again. Again, this is not a problem to worry about, the focus is on the procedure and implementation of the simulation model.

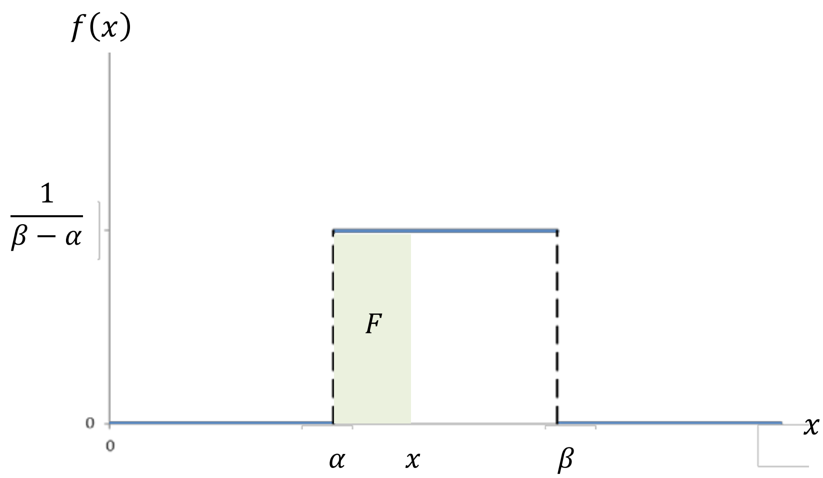
Save your work.

At this time, it is pertinent for readers to note that the random numbers generated by the *Excel* RAND function are actually not truly random. They are the best "random" numbers *Excel* can provide. The "random" numbers are being pulled from a Uniform distribution. The "randomness" of the values can be improved by changing the random seed settings of the *Excel* program. (Adjusting the random seed settings is beyond the scope of this course and participants are encouraged to research this issue on their own). Adjusting the *Excel* random seed generator will yield some improvement, but it still does not make the "random" numbers truly random. Henceforth "random" numbers generated by the *Excel* program will be better described as **pseudo-random numbers**.

Revert to the simulation spreadsheet.
Set up the headers for Activity 2: Furnish & Install Trailers.

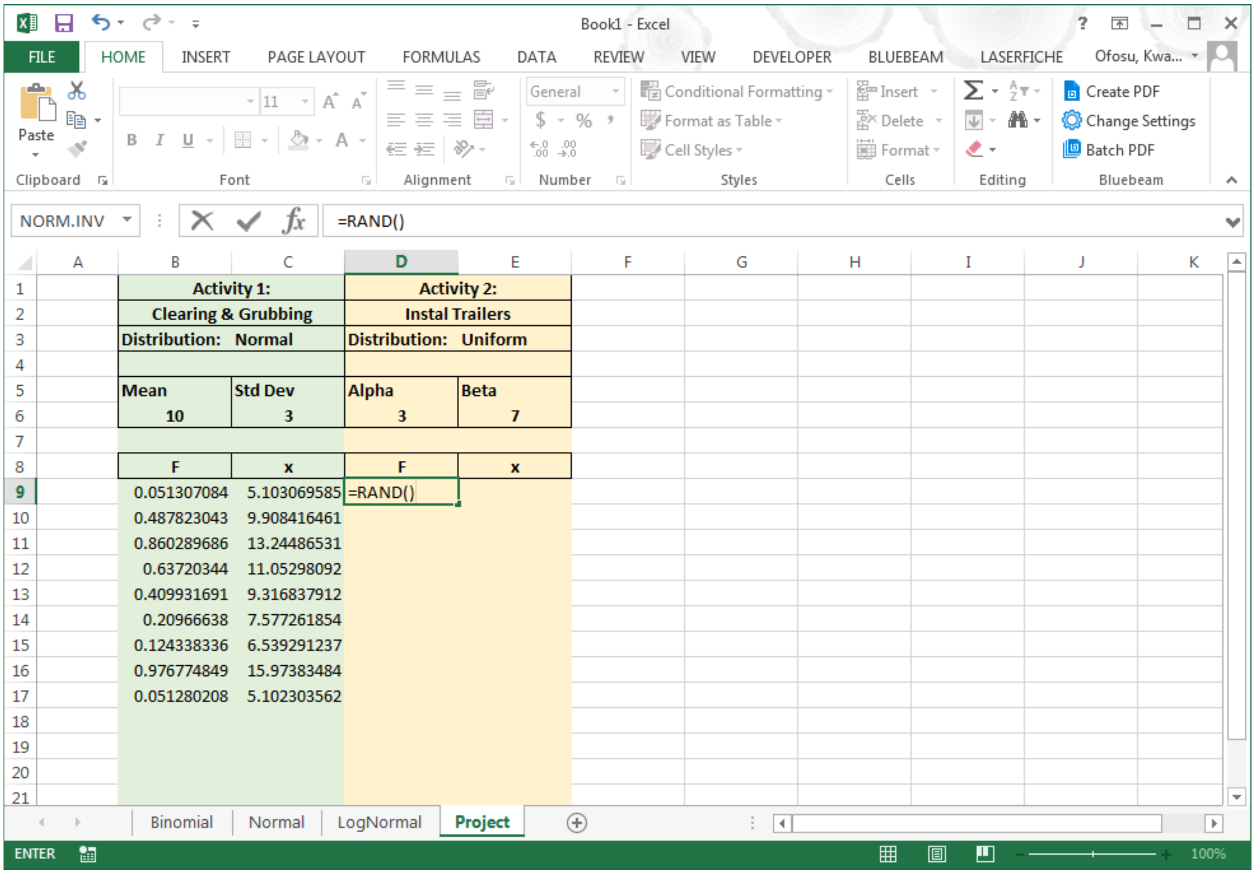


This Activity 2 completion time is described by a Uniform distribution with parameters
$\alpha = 3, \beta = 7,$
where the probability of any x in $[\alpha, \beta]$ equals $1/(\beta - \alpha)$
and the probability of any x not in $[\alpha, \beta]$ equals zero.
Thus, the density function of the Uniform distribution is a horizontal line.
The cumulative density of a value x in $[\alpha, \beta]$ is the rectangular area under the density function from α through x.



Therefore, for a pseudo-random cumulative probability value F,
$x = \alpha + F(\beta - \alpha)$

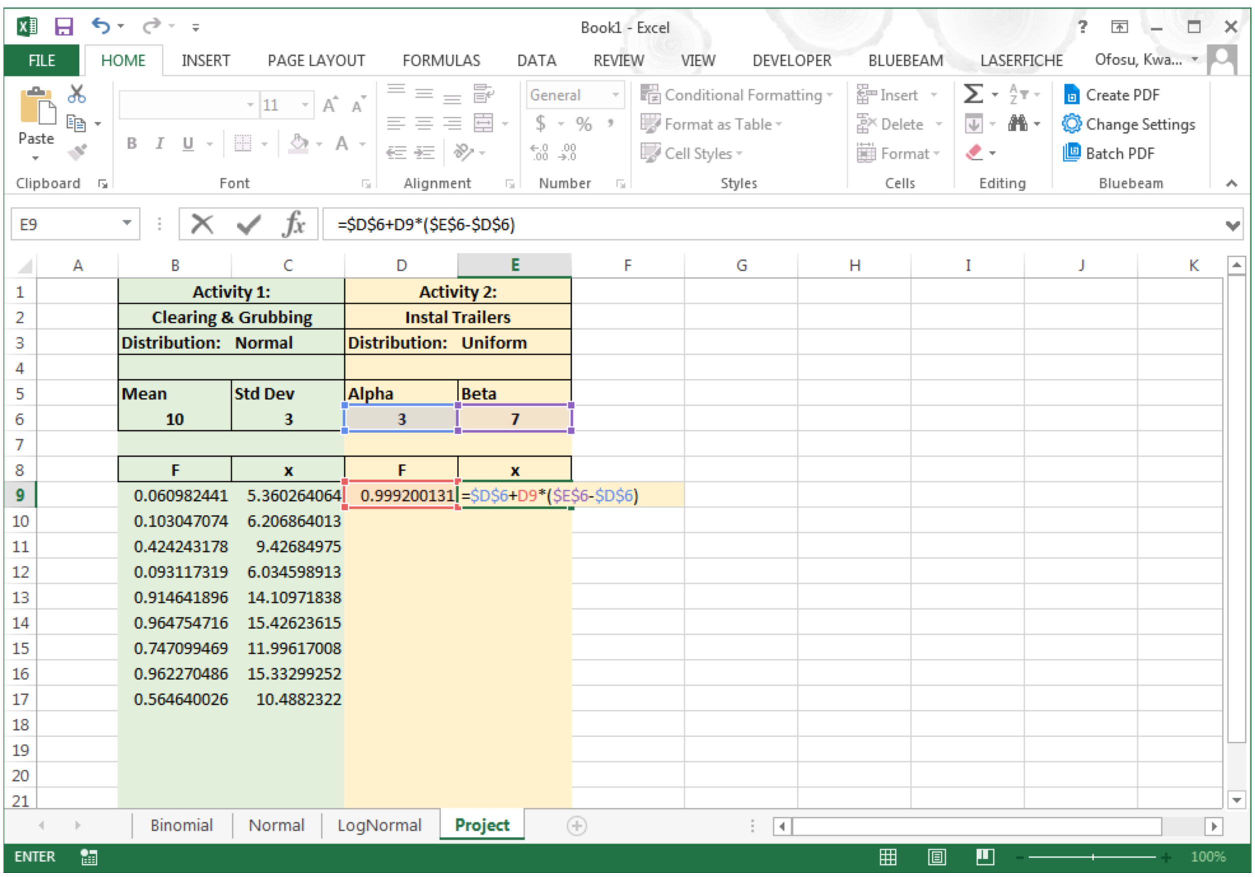
On the spreadsheet, generate a pseudo-random value F.



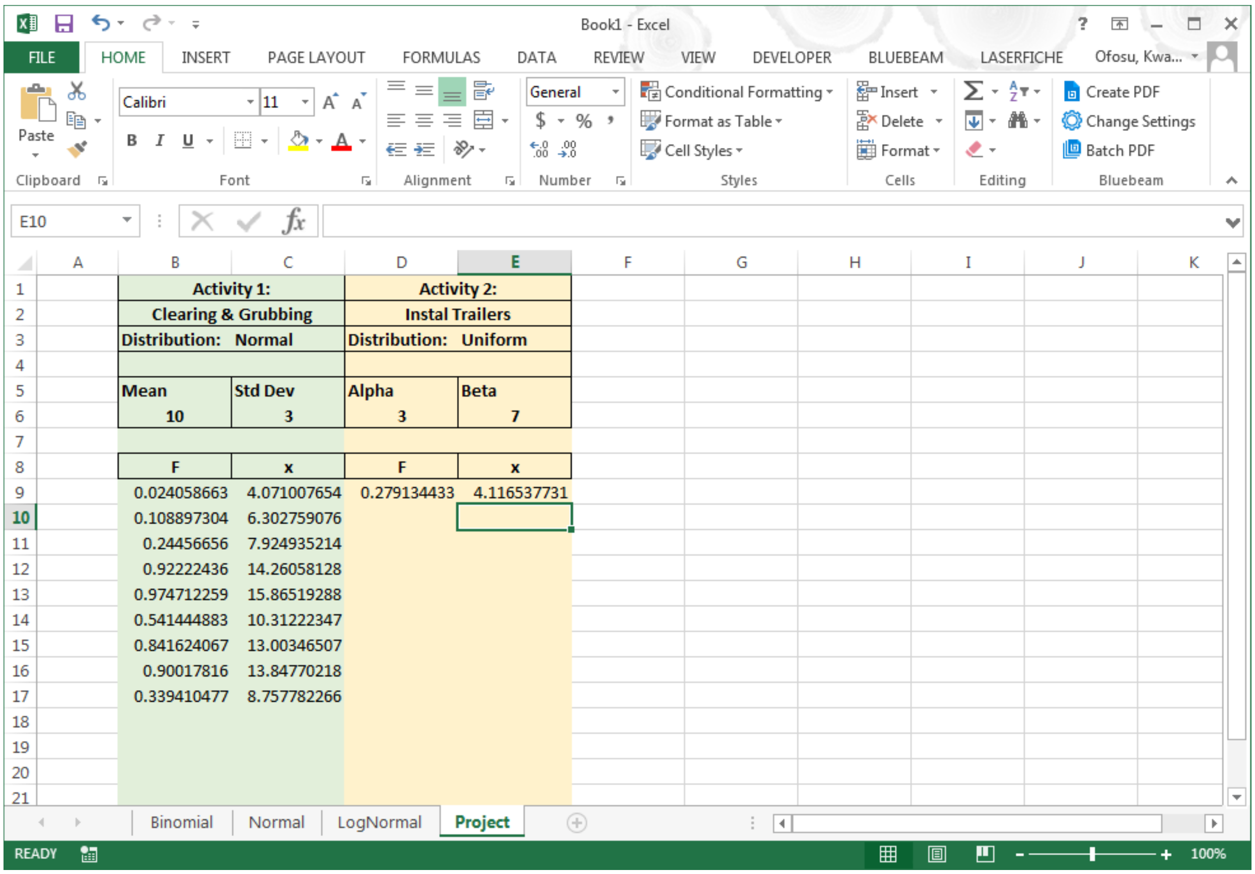
Hit **Enter** on the keyboard.

Enter the formula for the corresponding x value from the uniform distribution of the activity as follows,

Type " = "
Click on the cell holding the α value.
Hit **F4** on the keyboard to make the reference to the α value an absolute reference.
Type " + "
Click on the cell holding the F value.
Type " * "
Type " ("
Click on the cell holding the β value.
Hit **F4** on the keyboard to make the reference to the β value an absolute reference.
Type " - " (minus sign).
Click on the cell holding the α value.
Hit **F4** on the keyboard to make the reference to the α value an absolute reference.
Type ") "

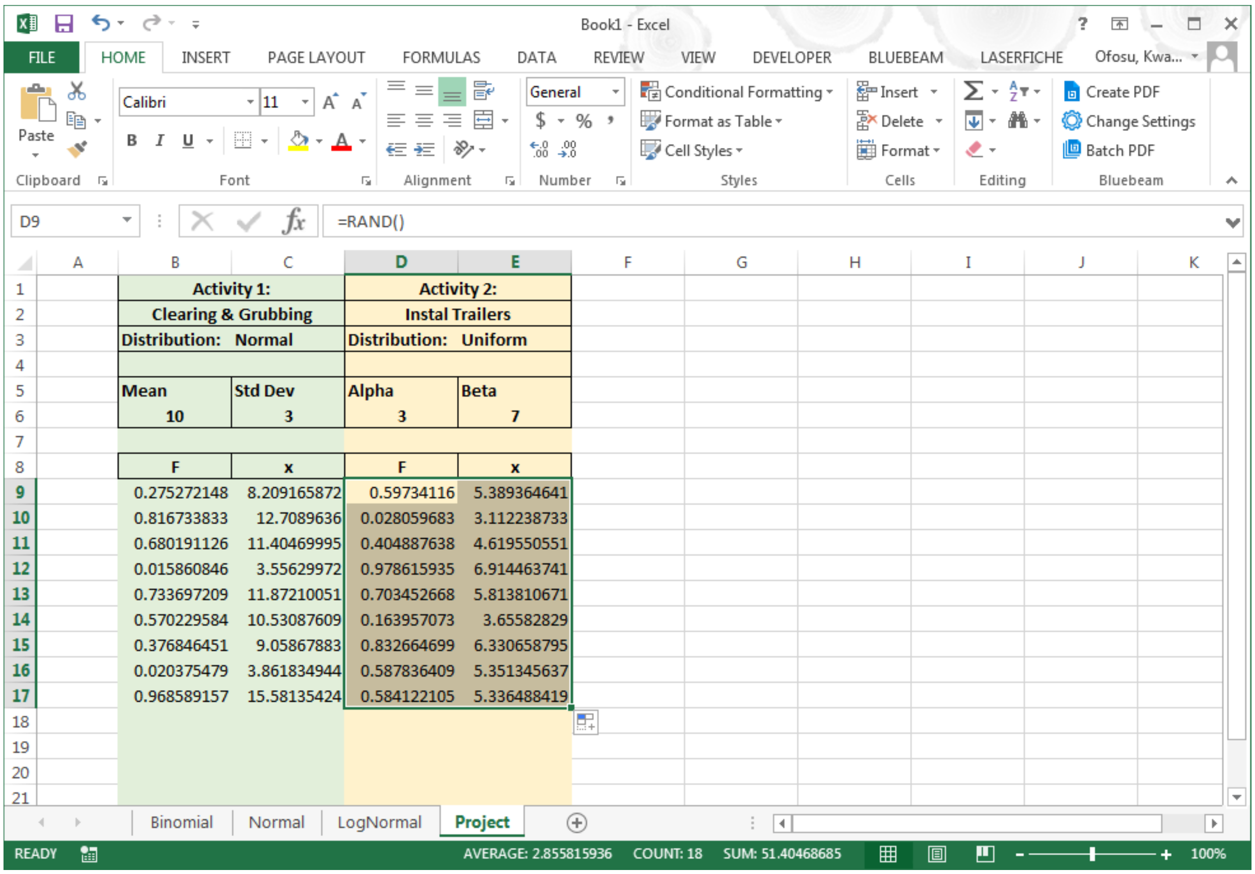


Hit **Enter** on the keyboard to implement the formula.



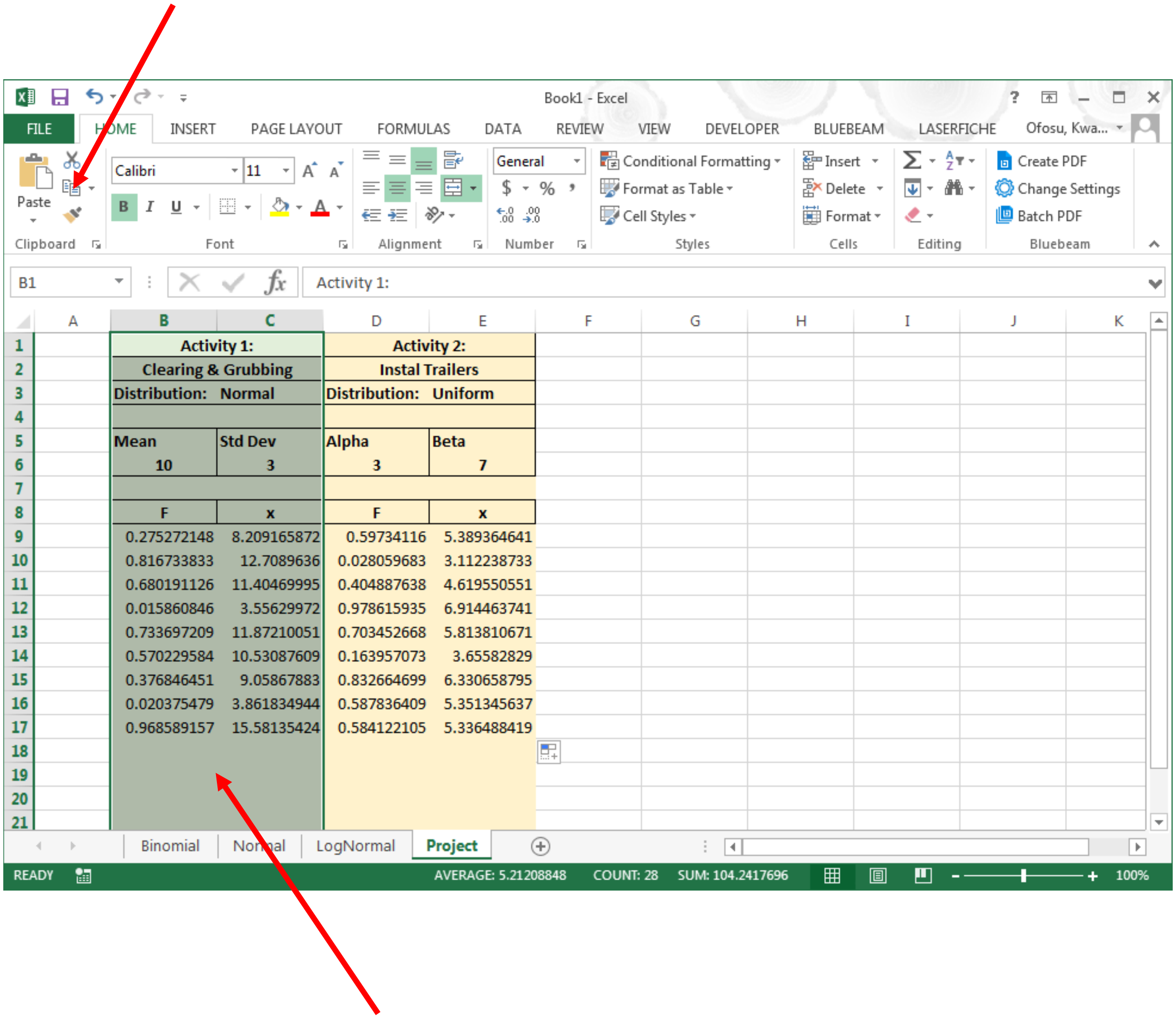
Select the F and x values
Fill handle a few rows.
Review the results.

The results appear to be normal and reasonable.

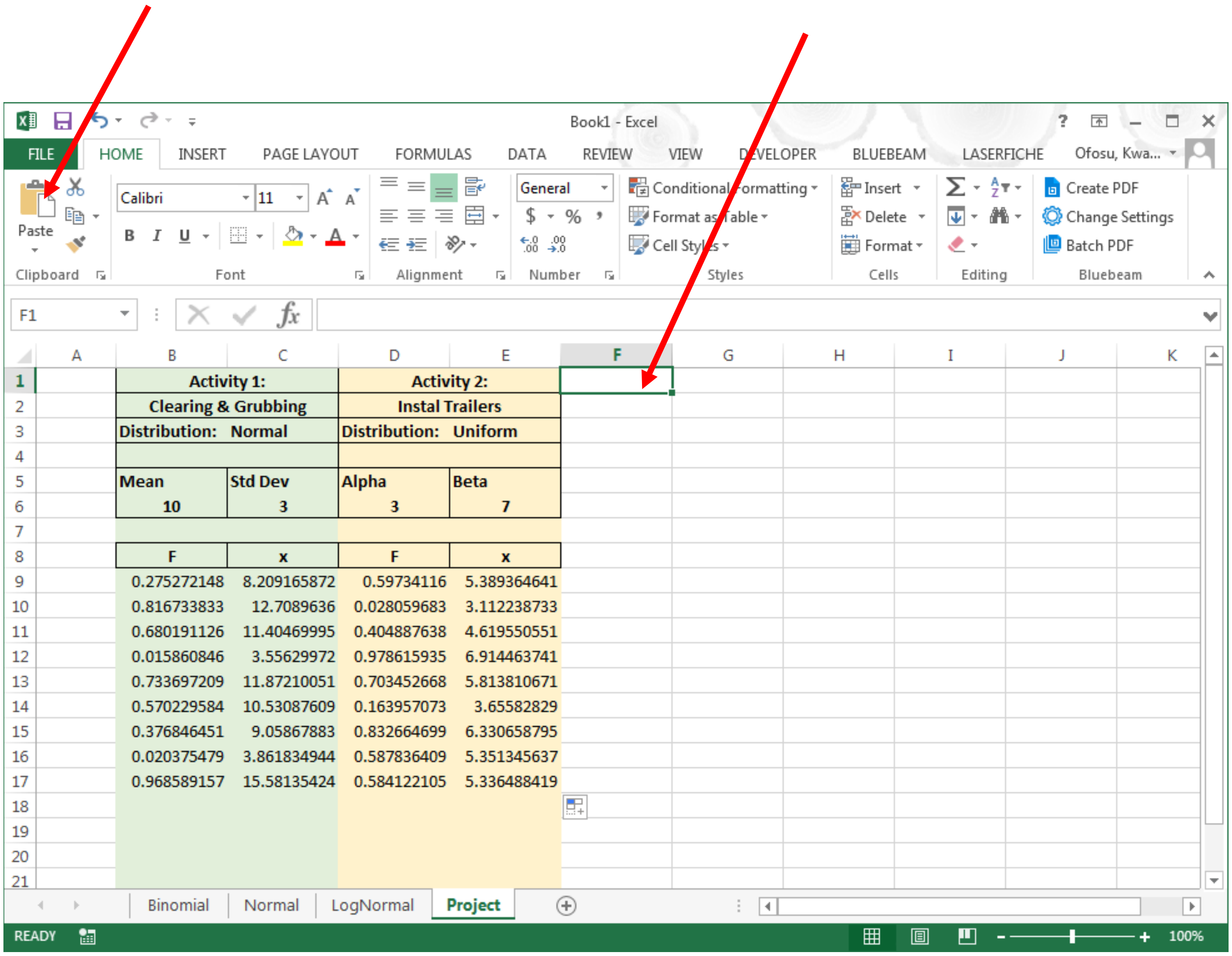


Activity 3 is described by a Normal distribution. Therefore, to save effort, we may copy over Activity 1 and update the statistical parameters and other input data.

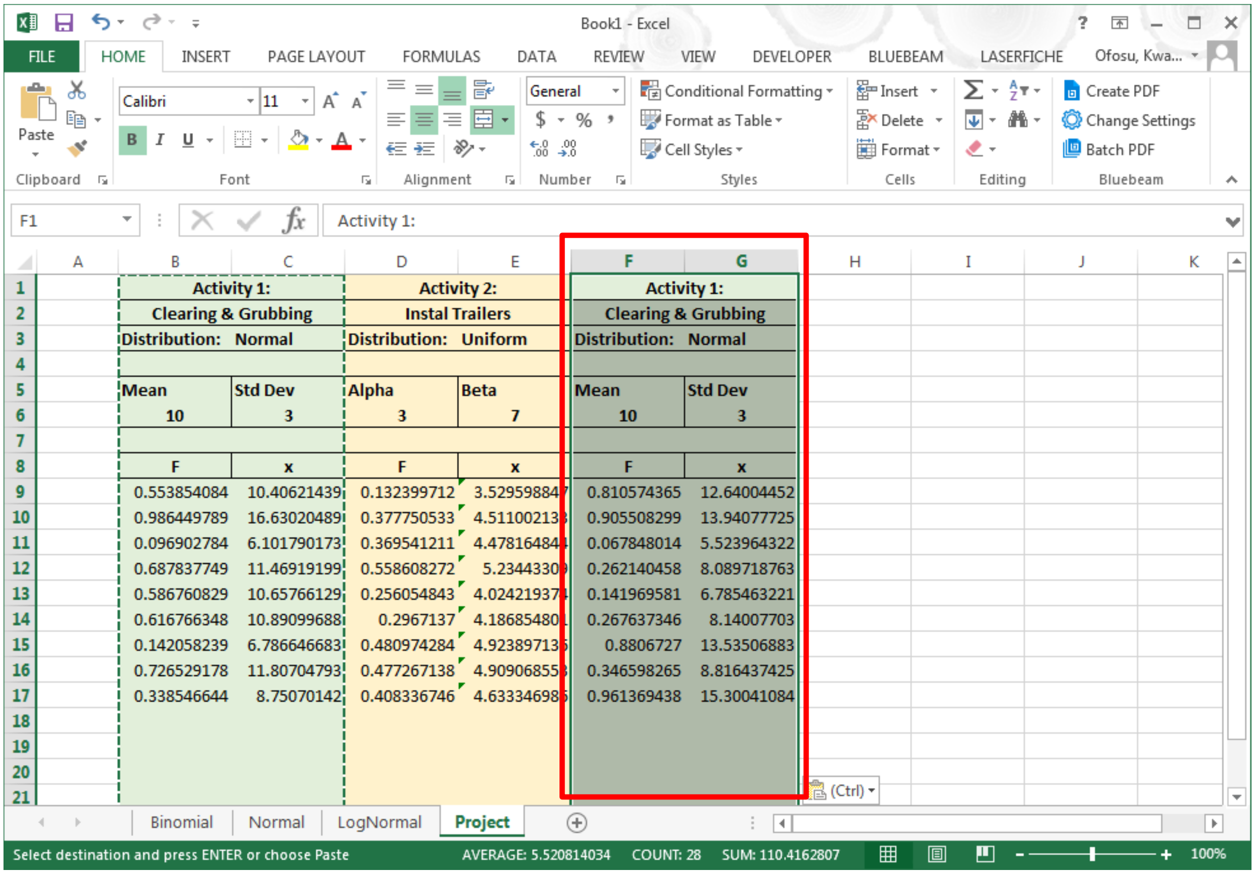
Select the Activity 1 data.
Press the Copy icon.



Click on the destination cell.
Press the Paste icon.

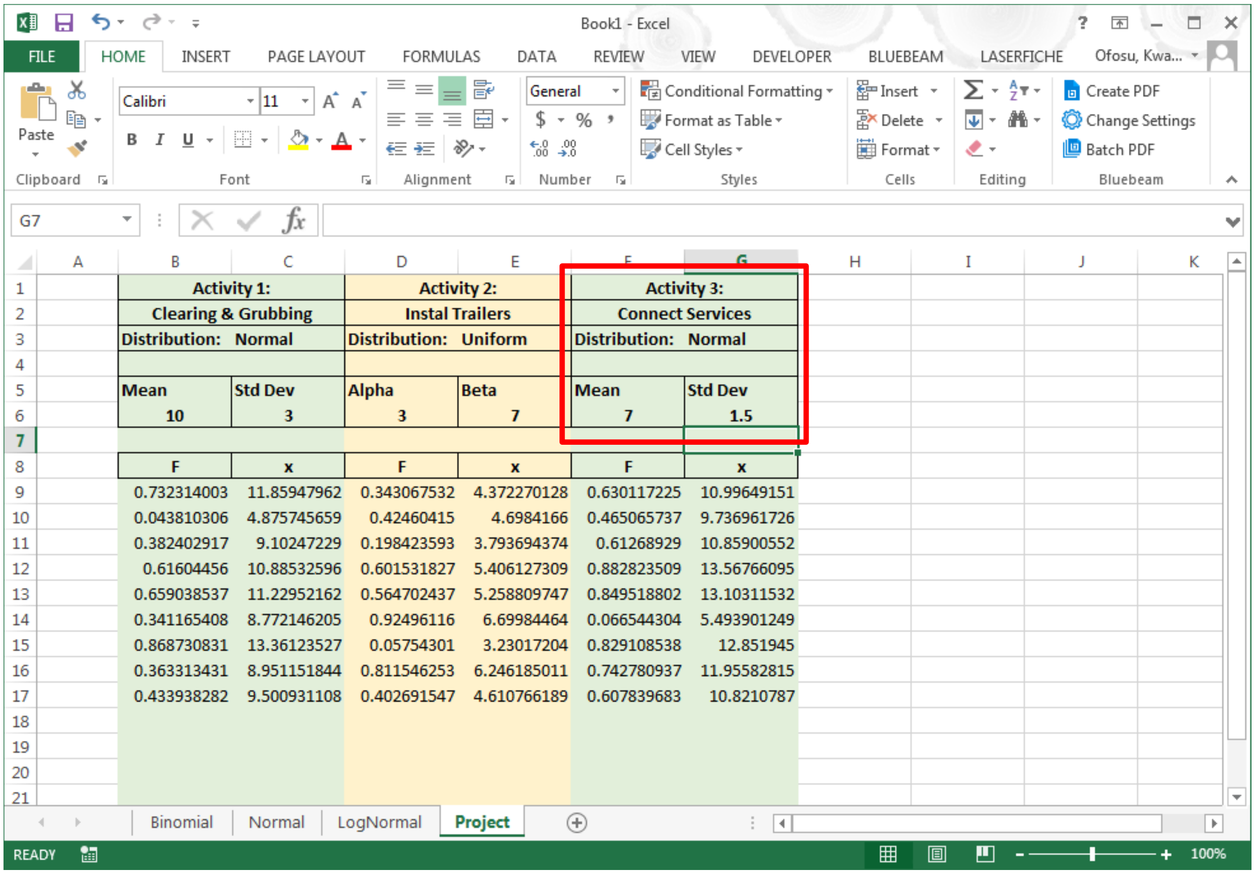


The data is copied over.



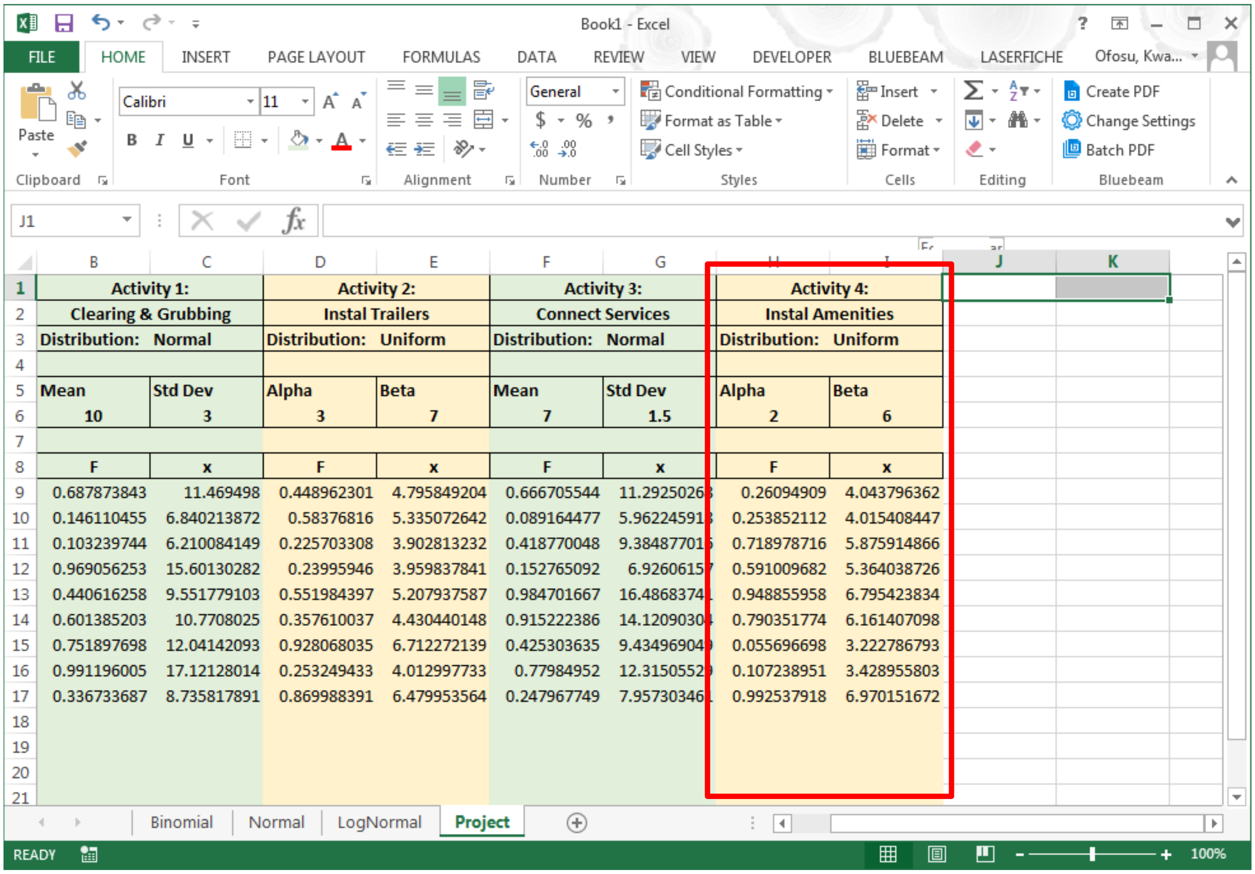
	B	C	D	E	F	G
1	Activity 1:		Activity 2:		Activity 1:	
2	Clearing & Grubbing		Instal Trailers		Clearing & Grubbing	
3	Distribution:	Normal	Distribution:	Uniform	Distribution:	Normal
4						
5	Mean	Std Dev	Alpha	Beta	Mean	Std Dev
6	10	3	3	7	10	3
7						
8	F	x	F	x	F	x
9	0.553854084	10.40621439	0.132399712	3.52959884	0.810574365	12.64004452
10	0.986449789	16.63020489	0.377750533	4.51100213	0.905508299	13.94077725
11	0.096902784	6.101790173	0.369541211	4.47816484	0.067848014	5.523964322
12	0.687837749	11.46919199	0.558608272	5.2344330	0.262140458	8.089718763
13	0.586760829	10.65766129	0.256054843	4.02421937	0.141969581	6.785463221
14	0.616766348	10.89099688	0.2967137	4.18685480	0.267637346	8.14007703
15	0.142058239	6.786646683	0.480974284	4.92389713	0.8806727	13.53506883
16	0.726529178	11.80704793	0.477267138	4.90906855	0.346598265	8.816437425
17	0.338546644	8.75070142	0.408336746	4.63334698	0.961369438	15.30041084

Update the data accordingly for Activity 3.



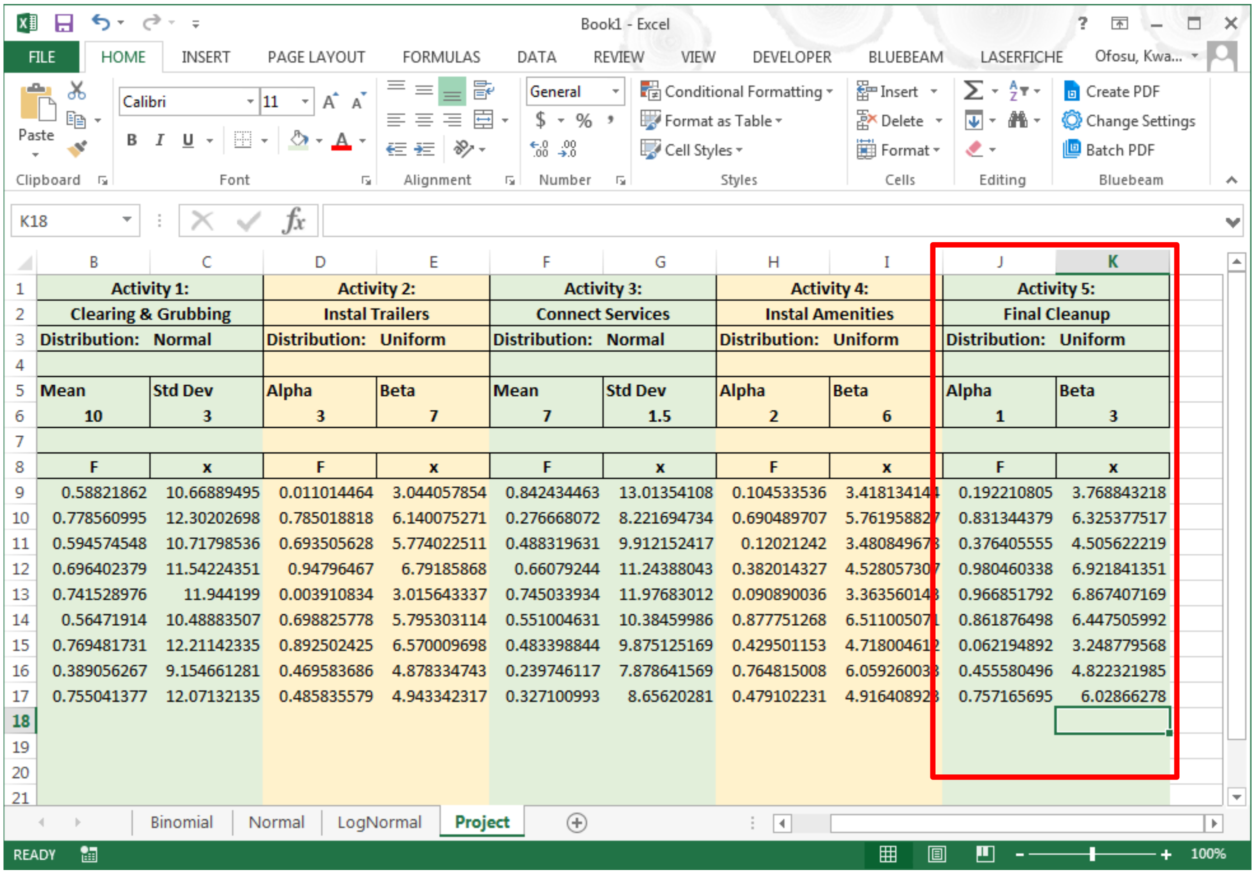
	B	C	D	E	F	G
1	**Activity 1:**		**Activity 2:**		**Activity 3:**	
2	**Clearing & Grubbing**		**Instal Trailers**		**Connect Services**	
3	**Distribution:**	**Normal**	**Distribution:**	**Uniform**	**Distribution:**	**Normal**
4						
5	**Mean**	**Std Dev**	**Alpha**	**Beta**	**Mean**	**Std Dev**
6	**10**	**3**	**3**	**7**	**7**	**1.5**
7						
8	**F**	**x**	**F**	**x**	**F**	**x**
9	0.732314003	11.85947962	0.343067532	4.372270128	0.630117225	10.99649151
10	0.043810306	4.875745659	0.42460415	4.6984166	0.465065737	9.736961726
11	0.382402917	9.10247229	0.198423593	3.793694374	0.61268929	10.85900552
12	0.61604456	10.88532596	0.601531827	5.406127309	0.882823509	13.56766095
13	0.659038537	11.22952162	0.564702437	5.258809747	0.849518802	13.10311532
14	0.341165408	8.772146205	0.92496116	6.69984464	0.066544304	5.493901249
15	0.868730831	13.36123527	0.05754301	3.23017204	0.829108538	12.851945
16	0.363313431	8.951151844	0.811546253	6.246185011	0.742780937	11.95582815
17	0.433938282	9.500931108	0.402691547	4.610766189	0.607839683	10.8210787

Activity 4 is described by a Uniform distribution. Therefore, to save effort, we may copy over Activity 2 and update the statistical parameters and other input data, as follows.



	B	C	D	E	F	G	H	I
1	Activity 1:		Activity 2:		Activity 3:		Activity 4:	
2	Clearing & Grubbing		Instal Trailers		Connect Services		Instal Amenities	
3	Distribution: Normal		Distribution: Uniform		Distribution: Normal		Distribution: Uniform	
4								
5	Mean	Std Dev	Alpha	Beta	Mean	Std Dev	Alpha	Beta
6	10	3	3	7	7	1.5	2	6
7								
8	F	x	F	x	F	x	F	x
9	0.687873843	11.469498	0.448962301	4.795849204	0.666705544	11.29250268	0.26094909	4.043796362
10	0.146110455	6.840213872	0.58376816	5.335072642	0.089164477	5.962245918	0.253852112	4.015408447
11	0.103239744	6.210084149	0.225703308	3.902813232	0.418770048	9.384877016	0.718978716	5.875914866
12	0.969056253	15.60130282	0.23995946	3.959837841	0.152765092	6.92606157	0.591009682	5.364038726
13	0.440616258	9.551779103	0.551984397	5.207937587	0.984701667	16.48683741	0.948855958	6.795423834
14	0.601385203	10.7708025	0.357610037	4.430440148	0.915222386	14.12090304	0.790351774	6.161407098
15	0.751897698	12.04142093	0.928068035	6.712272139	0.425303635	9.434969049	0.055696698	3.222786793
16	0.991196005	17.12128014	0.253249433	4.012997733	0.77984952	12.31505529	0.107238951	3.428955803
17	0.336733687	8.735817891	0.869988391	6.479953564	0.247967749	7.957303461	0.992537918	6.970151672

Activity 5 is described by a Uniform distribution. Therefore, to save effort, we may copy over Activity 4 and update the statistical parameters and other input data, as follows.

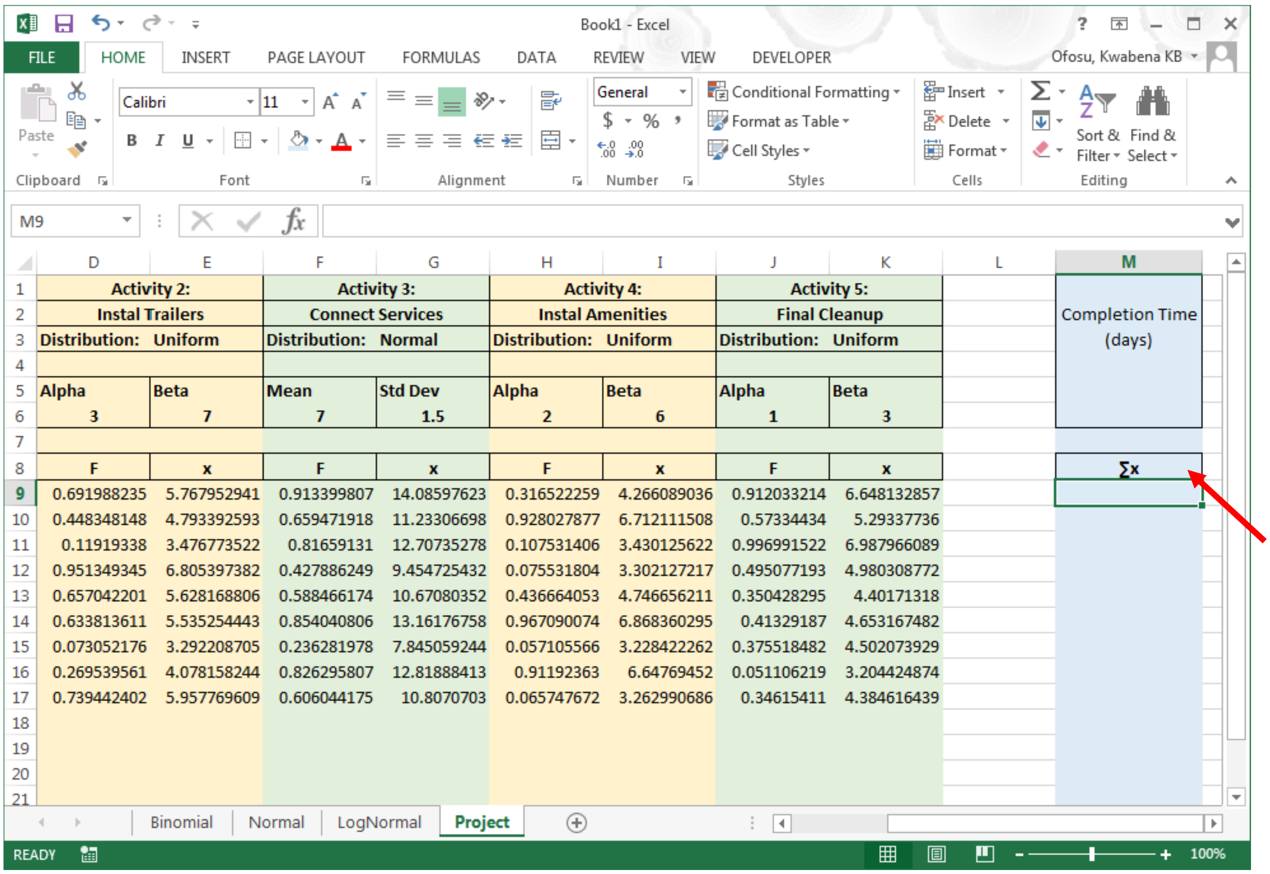


	B	C	D	E	F	G	H	I	J	K
1	Activity 1:		Activity 2:		Activity 3:		Activity 4:		Activity 5:	
2	Clearing & Grubbing		Instal Trailers		Connect Services		Instal Amenities		Final Cleanup	
3	Distribution:	Normal	Distribution:	Uniform	Distribution:	Normal	Distribution:	Uniform	Distribution:	Uniform
4										
5	Mean	Std Dev	Alpha	Beta	Mean	Std Dev	Alpha	Beta	Alpha	Beta
6	10	3	3	7	7	1.5	2	6	1	3
7										
8	F	x	F	x	F	x	F	x	F	x
9	0.58821862	10.66889495	0.011014464	3.044057854	0.842434463	13.01354108	0.104533536	3.418134144	0.192210805	3.768843218
10	0.778560995	12.30202698	0.785018818	6.140075271	0.276668072	8.221694734	0.690489707	5.761958827	0.831344379	6.325377517
11	0.594574548	10.71798536	0.693505628	5.774022511	0.488319631	9.912152417	0.12021242	3.480849678	0.376405555	4.505622219
12	0.696402379	11.54224351	0.94796467	6.79185868	0.66079244	11.24388043	0.382014327	4.528057307	0.980460338	6.921841351
13	0.741528976	11.944199	0.003910834	3.015643337	0.745033934	11.97683012	0.090890036	3.363560148	0.966851792	6.867407169
14	0.56471914	10.48883507	0.698825778	5.795303114	0.551004631	10.38459986	0.877751268	6.511005071	0.861876498	6.447505992
15	0.769481731	12.21142335	0.892502425	6.570009698	0.483398844	9.875125169	0.429501153	4.718004612	0.062194892	3.248779568
16	0.389056267	9.154661281	0.469583686	4.878334743	0.239746117	7.878641569	0.764815008	6.059260038	0.455580496	4.822321985
17	0.755041377	12.07132135	0.485835579	4.943342317	0.327100993	8.65620281	0.479102231	4.916408928	0.757165695	6.02866278
18										
19										
20										
21										

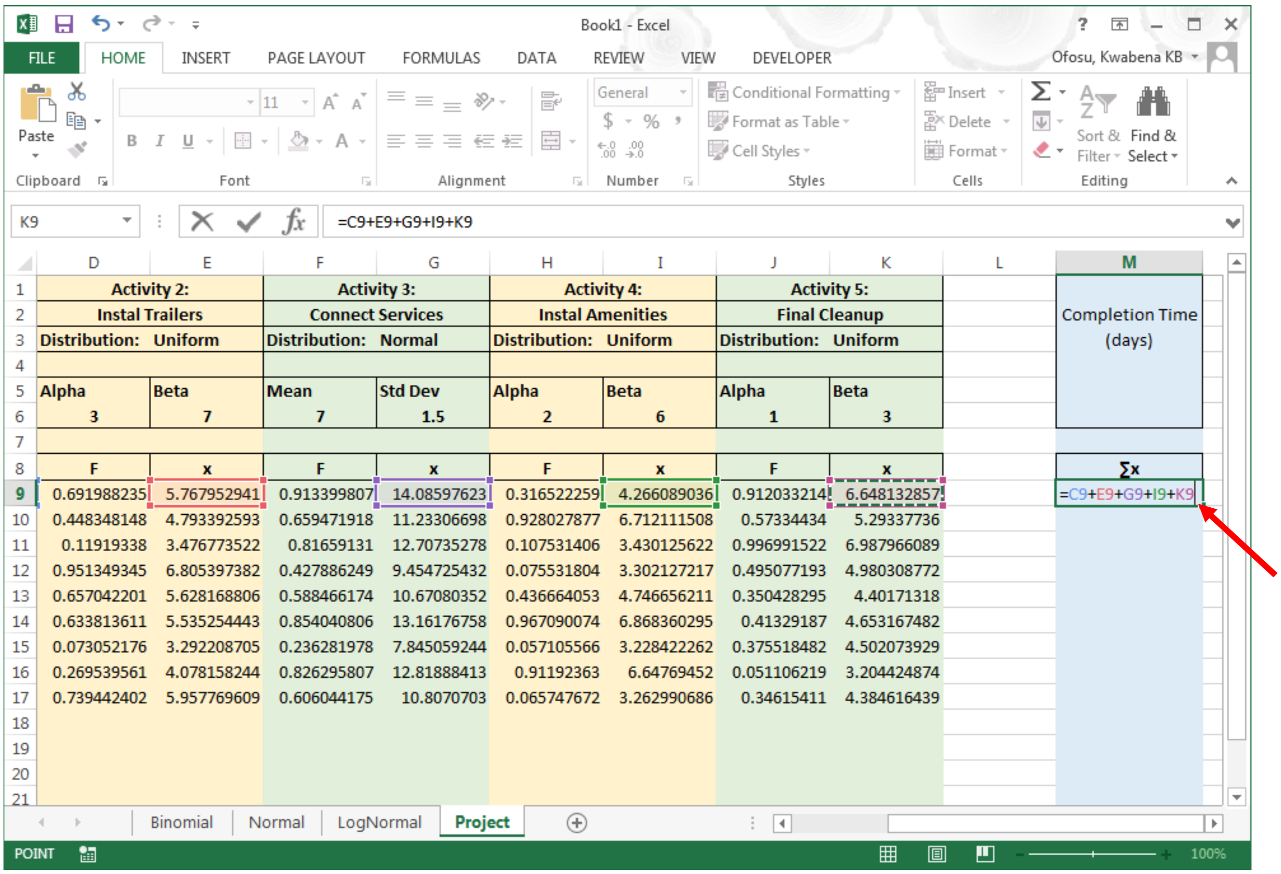
The activities' simulations are now implemented. Additional simulations may be added by fill handling down the spreadsheet as needed.

The next step is to calculate the overall project completion time for a simulation, and then replicate this calculation for the multiple simulations conducted, to synthesize a dataset of the simulated project completion times that will be analyzed statistically to reach the conclusions, recommendations and decisions of the simulation study.

Add a header for the project completion time.



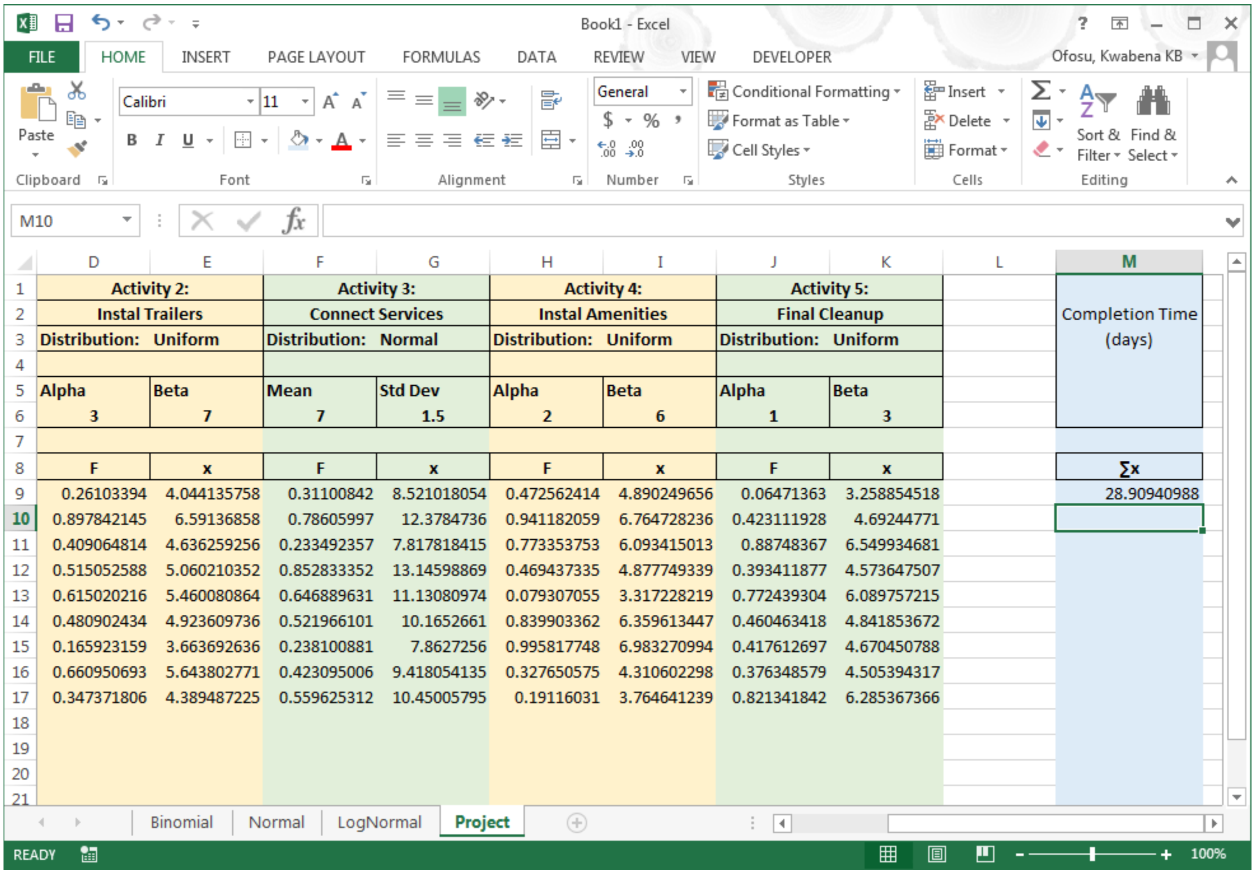
Enter an addition formula that sums the completion time (x) values of Activity 1 through Activity 5 of the first simulation.



	D	E	F	G	H	I	J	K	L	M
1	Activity 2:		Activity 3:		Activity 4:		Activity 5:			
2	Instal Trailers		Connect Services		Instal Amenities		Final Cleanup			Completion Time (days)
3	Distribution: Uniform		Distribution: Normal		Distribution: Uniform		Distribution: Uniform			
4										
5	Alpha	Beta	Mean	Std Dev	Alpha	Beta	Alpha	Beta		
6	3	7	7	1.5	2	6	1	3		
7										
8	F	x	F	x	F	x	F	x		Σx
9	0.691988235	5.767952941	0.913399807	14.08597623	0.316522259	4.266089036	0.912033214	6.648132857		=C9+E9+G9+I9+K9
10	0.448348148	4.793392593	0.659471918	11.23306698	0.928027877	6.712111508	0.57334434	5.29337736		
11	0.11919338	3.476773522	0.81659131	12.70735278	0.107531406	3.430125622	0.996991522	6.987966089		
12	0.951349345	6.805397382	0.427886249	9.454725432	0.075531804	3.302127217	0.495077193	4.980308772		
13	0.657042201	5.628168806	0.588466174	10.67080352	0.436664053	4.746656211	0.350428295	4.40171318		
14	0.633813611	5.535254443	0.854040806	13.16176758	0.967090074	6.868360295	0.41329187	4.653167482		
15	0.073052176	3.292208705	0.236281978	7.845059244	0.057105566	3.228422262	0.375518482	4.502073929		
16	0.269539561	4.078158244	0.826295807	12.81888413	0.91192363	6.64769452	0.051106219	3.204424874		
17	0.739442402	5.957769609	0.606044175	10.8070703	0.065747672	3.262990686	0.34615411	4.384616439		

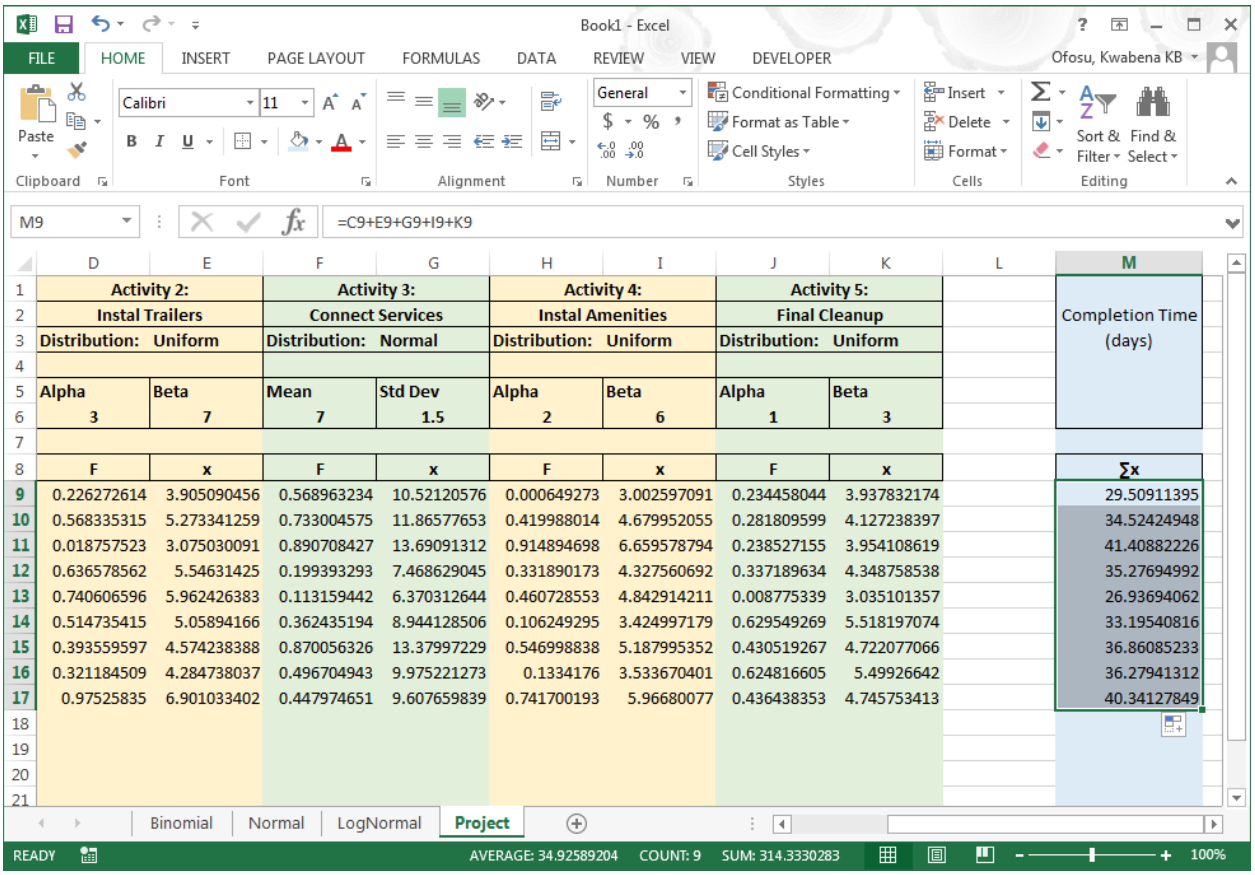
Hit **Enter** on the keyboard.

The overall project completion time for the first simulation is computed.



	D	E	F	G	H	I	J	K	L	M
1	Activity 2:		Activity 3:		Activity 4:		Activity 5:			
2	Instal Trailers		Connect Services		Instal Amenities		Final Cleanup			Completion Time (days)
3	Distribution:	Uniform	Distribution:	Normal	Distribution:	Uniform	Distribution:	Uniform		
4										
5	Alpha	Beta	Mean	Std Dev	Alpha	Beta	Alpha	Beta		
6	3	7	7	1.5	2	6	1	3		
7										
8	F	x	F	x	F	x	F	x		∑x
9	0.26103394	4.044135758	0.31100842	8.521018054	0.472562414	4.890249656	0.06471363	3.258854518		28.90940988
10	0.897842145	6.59136858	0.78605997	12.3784736	0.941182059	6.764728236	0.423111928	4.69244771		
11	0.409064814	4.636259256	0.233492357	7.817818415	0.773353753	6.093415013	0.88748367	6.549934681		
12	0.515052588	5.060210352	0.852833352	13.14598869	0.469437335	4.877749339	0.393411877	4.573647507		
13	0.615020216	5.460080864	0.646889631	11.13080974	0.079307055	3.317228219	0.772439304	6.089757215		
14	0.480902434	4.923609736	0.521966101	10.1652661	0.839903362	6.359613447	0.460463418	4.841853672		
15	0.165923159	3.663692636	0.238100881	7.8627256	0.995817748	6.983270994	0.417612697	4.670450788		
16	0.660950693	5.643802771	0.423095006	9.418054135	0.327650575	4.310602298	0.376348579	4.505394317		
17	0.347371806	4.389487225	0.559625312	10.45005795	0.19116031	3.764641239	0.821341842	6.285367366		

Select the first project completion time.
Fill handle a few rows.
Review the results.



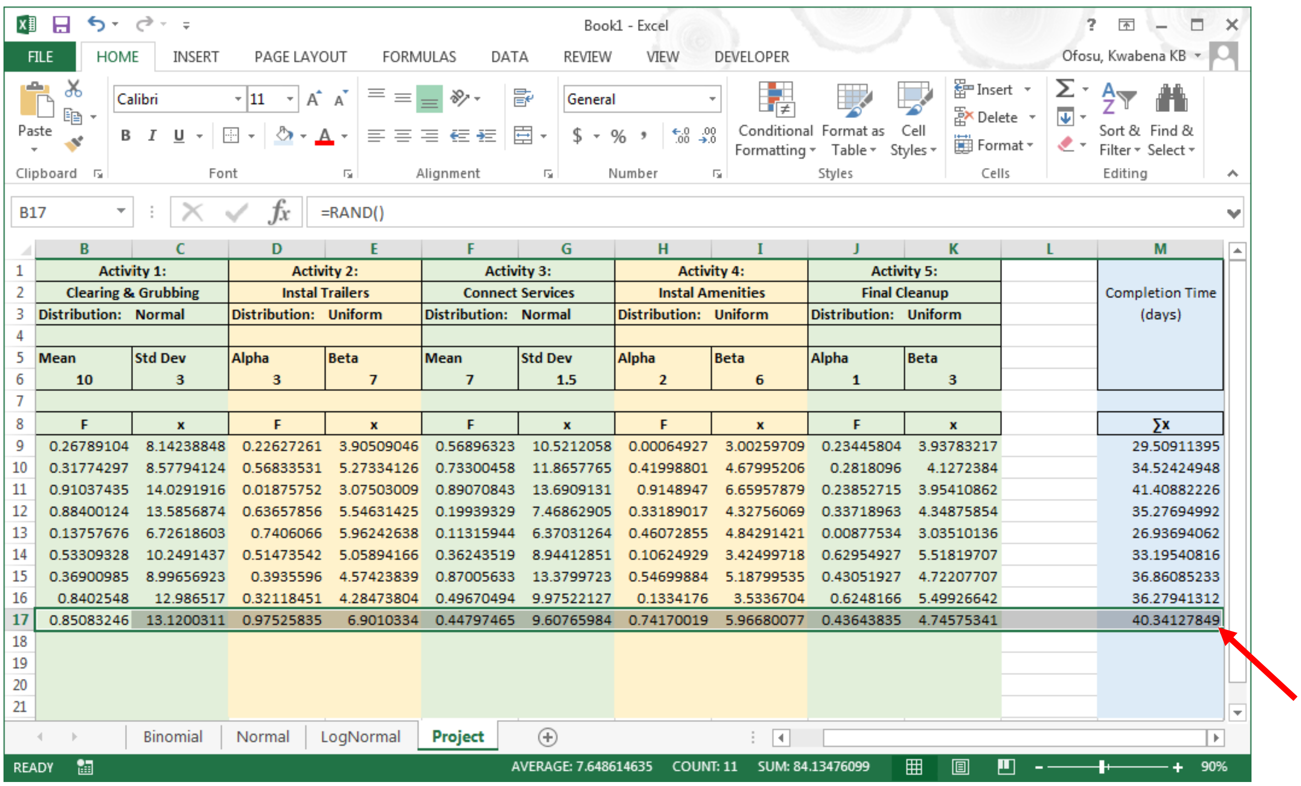
M9: =C9+E9+G9+I9+K9

	D	E	F	G	H	I	J	K	L	M
1	Activity 2:		Activity 3:		Activity 4:		Activity 5:			
2	Instal Trailers		Connect Services		Instal Amenities		Final Cleanup			Completion Time (days)
3	Distribution: Uniform		Distribution: Normal		Distribution: Uniform		Distribution: Uniform			
4										
5	Alpha	Beta	Mean	Std Dev	Alpha	Beta	Alpha	Beta		
6	3	7	7	1.5	2	6	1	3		
7										
8	F	x	F	x	F	x	F	x		∑x
9	0.226272614	3.905090456	0.568963234	10.52120576	0.000649273	3.002597091	0.234458044	3.937832174		29.50911395
10	0.568335315	5.273341259	0.733004575	11.86577653	0.419988014	4.679952055	0.281809599	4.127238397		34.52424948
11	0.018757523	3.075030091	0.890708427	13.69091312	0.914894698	6.659578794	0.238527155	3.954108619		41.40882226
12	0.636578562	5.54631425	0.199393293	7.468629045	0.331890173	4.327560692	0.337189634	4.348758538		35.27694992
13	0.740606596	5.962426383	0.113159442	6.370312644	0.460728553	4.842914211	0.008775339	3.035101357		26.93694062
14	0.514735415	5.05894166	0.362435194	8.944128506	0.106249295	3.424997179	0.629549269	5.518197074		33.19540816
15	0.393559597	4.574238388	0.870056326	13.37997229	0.546998838	5.187995352	0.430519267	4.722077066		36.86085233
16	0.321184509	4.284738037	0.496704943	9.975221273	0.1334176	3.533670401	0.624816605	5.49926642		36.27941312
17	0.97525835	6.901033402	0.447974651	9.607659839	0.741700193	5.96680077	0.436438353	4.745753413		40.34127849

AVERAGE: 34.92589204 COUNT: 9 SUM: 314.3330283

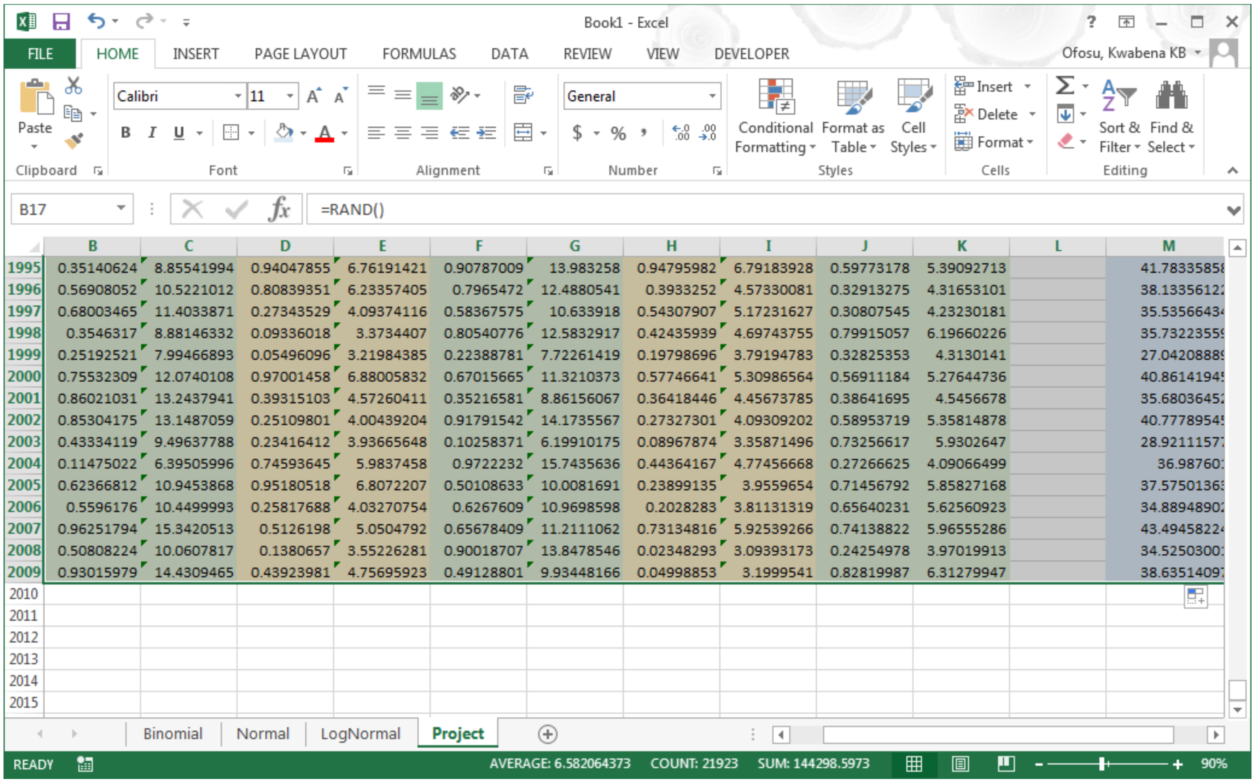
We will now like to run 2,000 simulations of this system. In other words, we shall investigate 2,000 possible scenarios of this project, taking into account the random outcomes of the project's constituent activities.

Select the entire row of the last simulation.



	B	C	D	E	F	G	H	I	J	K	L	M
1	Activity 1:		Activity 2:		Activity 3:		Activity 4:		Activity 5:			
2	Clearing & Grubbing		Instal Trailers		Connect Services		Instal Amenities		Final Cleanup			Completion Time (days)
3	Distribution:	Normal	Distribution:	Uniform	Distribution:	Normal	Distribution:	Uniform	Distribution:	Uniform		
4												
5	Mean	Std Dev	Alpha	Beta	Mean	Std Dev	Alpha	Beta	Alpha	Beta		
6	10	3	3	7	7	1.5	2	6	1	3		
7												
8	F	x	F	x	F	x	F	x	F	x		∑x
9	0.26789104	8.14238848	0.22627261	3.90509046	0.56896323	10.5212058	0.00064927	3.00259709	0.23445804	3.93783217		29.50911395
10	0.31774297	8.57794124	0.56833531	5.27334126	0.73300458	11.8657765	0.41998801	4.67995206	0.2818096	4.1272384		34.52424948
11	0.91037435	14.0291916	0.01875752	3.07503009	0.89070843	13.6909131	0.9148947	6.65957879	0.23852715	3.95410862		41.40882226
12	0.88400124	13.5856874	0.63657856	5.54631425	0.19939329	7.46862905	0.33189017	4.32756069	0.33718963	4.34875854		35.27694992
13	0.13757676	6.72618603	0.7406066	5.96242638	0.11315944	6.37031264	0.46072855	4.84291421	0.00877534	3.03510136		26.93694062
14	0.53309328	10.2491437	0.51473542	5.05894166	0.36243519	8.94412851	0.10624929	3.42499718	0.62954927	5.51819707		33.19540816
15	0.36900985	8.99656923	0.3935596	4.57423839	0.87005633	13.3799723	0.54699884	5.18799535	0.43051927	4.72207707		36.86085233
16	0.8402548	12.986517	0.32118451	4.28473804	0.49670494	9.97522127	0.1334176	3.5336704	0.6248166	5.49926642		36.27941312
17	0.85083246	13.1200311	0.97525835	6.9010334	0.44797465	9.60765984	0.74170019	5.96680077	0.43643835	4.74575341		40.34127849
18												
19												
20												
21												

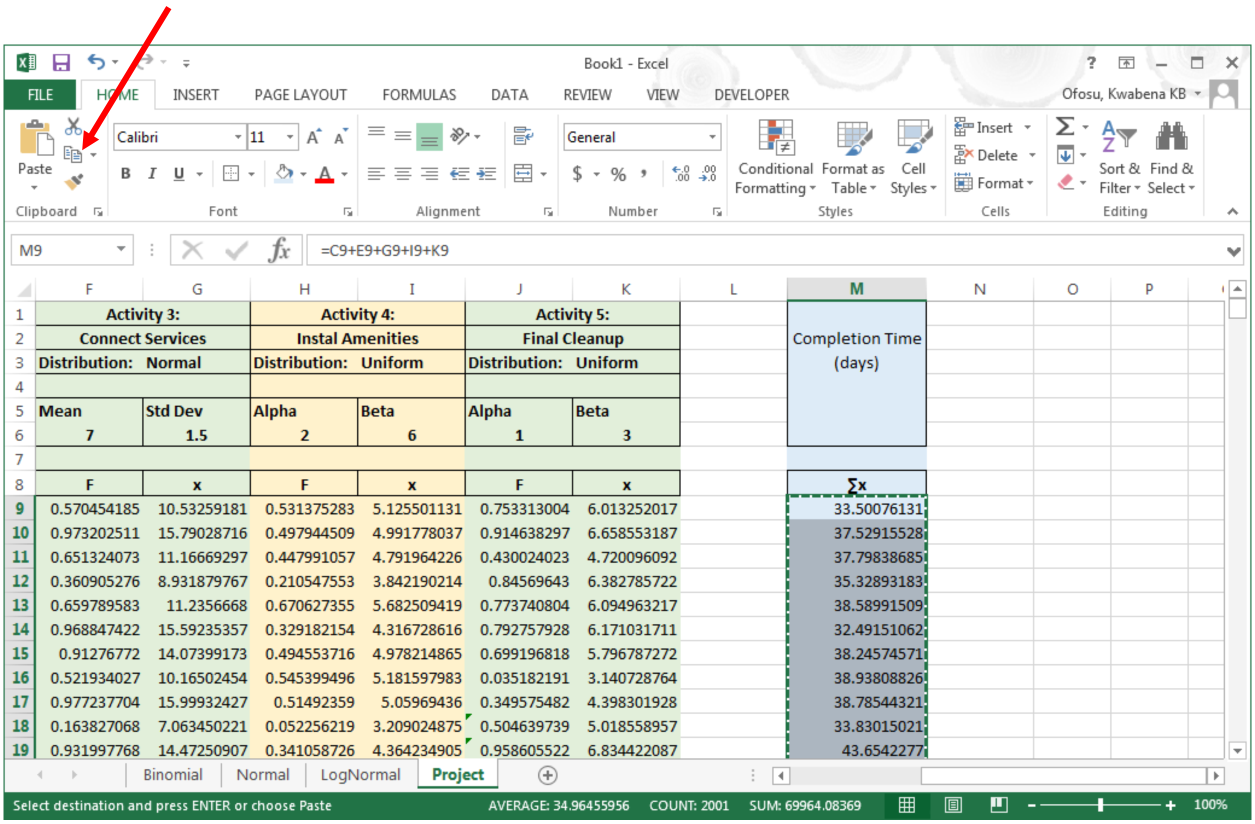
Fill handle for a total of 2,000 rows (simulations).



	B	C	D	E	F	G	H	I	J	K	L	M
1995	0.35140624	8.85541994	0.94047855	6.76191421	0.90787009	13.983258	0.94795982	6.79183928	0.59773178	5.39092713		41.7833585
1996	0.56908052	10.5221012	0.80839351	6.23357405	0.7965472	12.4880541	0.3933252	4.57330081	0.32913275	4.31653101		38.1335612
1997	0.68003465	11.4033871	0.27343529	4.09374116	0.58367575	10.633918	0.54307907	5.17231627	0.30807545	4.23230181		35.5356643
1998	0.3546317	8.88146332	0.09336018	3.3734407	0.80540776	12.5832917	0.42435939	4.69743755	0.79915057	6.19660226		35.7322355
1999	0.25192521	7.99466893	0.05496096	3.21984385	0.22388781	7.72261419	0.19798696	3.79194783	0.32825353	4.3130141		27.0420888
2000	0.75532309	12.0740108	0.97001458	6.88005832	0.67015665	11.3210373	0.57746641	5.30986564	0.56911184	5.27644736		40.8614194
2001	0.86021031	13.2437941	0.39315103	4.57260411	0.35216581	8.86156067	0.36418446	4.45673785	0.38641695	4.5456678		35.6803645
2002	0.85304175	13.1487059	0.25109801	4.00439204	0.91791542	14.1735567	0.27327301	4.09309202	0.58953719	5.35814878		40.7778954
2003	0.43334119	9.49637788	0.23416412	3.93665648	0.10258371	6.19910175	0.08967874	3.35871496	0.73256617	5.9302647		28.9211157
2004	0.11475022	6.39505996	0.74593645	5.9837458	0.9722232	15.7435636	0.44364167	4.77456668	0.27266625	4.09066499		36.98760
2005	0.62366812	10.9453868	0.95180518	6.8072207	0.50108633	10.0081691	0.23899135	3.9559654	0.71456792	5.85827168		37.5750136
2006	0.5596176	10.4499993	0.25817688	4.03270754	0.6267609	10.9698598	0.2028283	3.81131319	0.65640231	5.62560923		34.8894890
2007	0.96251794	15.3420513	0.5126198	5.0504792	0.65678409	11.2111062	0.73134816	5.92539266	0.74138822	5.96555286		43.4945822
2008	0.50808224	10.0607817	0.1380657	3.55226281	0.90018707	13.8478546	0.02348293	3.09393173	0.24254978	3.97019913		34.5250300
2009	0.93015979	14.4309465	0.43923981	4.75695923	0.49128801	9.93448166	0.04998853	3.1999541	0.82819987	6.31279947		38.6351409
2010												
2011												
2012												
2013												
2014												
2015												

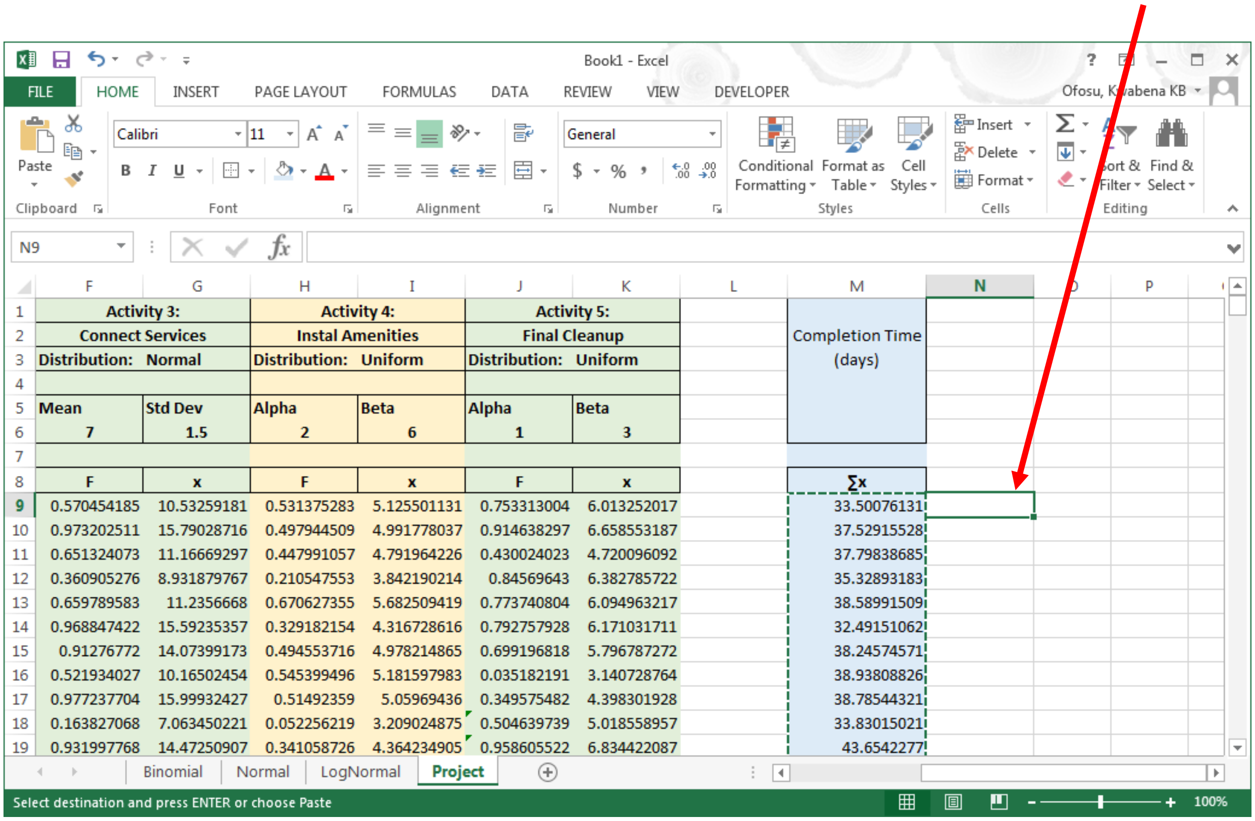
To obtain a stable dataset (that does not continuously, automatically update itself due to the random number function), select the synthesized project completion times.

Press the Copy icon.

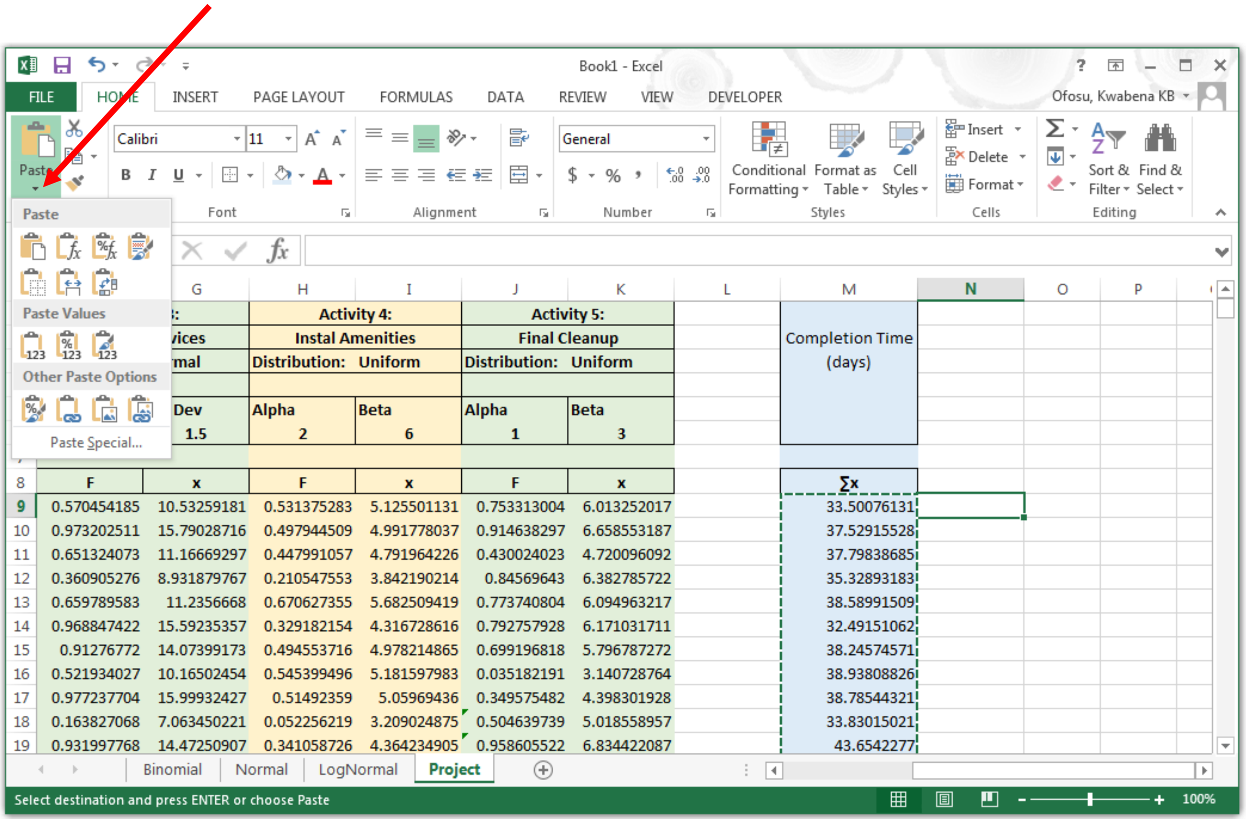


	F	G	H	I	J	K	L	M
1	Activity 3:		Activity 4:		Activity 5:			
2	Connect Services		Instal Amenities		Final Cleanup			Completion Time (days)
3	Distribution: Normal		Distribution: Uniform		Distribution: Uniform			
4								
5	Mean	Std Dev	Alpha	Beta	Alpha	Beta		
6	7	1.5	2	6	1	3		
7								
8	F	x	F	x	F	x		∑x
9	0.570454185	10.53259181	0.531375283	5.125501131	0.753313004	6.013252017		33.50076131
10	0.973202511	15.79028716	0.497944509	4.991778037	0.914638297	6.658553187		37.52915528
11	0.651324073	11.16669297	0.447991057	4.791964226	0.430024023	4.720096092		37.79838685
12	0.360905276	8.931879767	0.210547553	3.842190214	0.84569643	6.382785722		35.32893183
13	0.659789583	11.2356668	0.670627355	5.682509419	0.773740804	6.094963217		38.58991509
14	0.968847422	15.59235357	0.329182154	4.316728616	0.792757928	6.171031711		32.49151062
15	0.91276772	14.07399173	0.494553716	4.978214865	0.699196818	5.796787272		38.24574571
16	0.521934027	10.16502454	0.545399496	5.181597983	0.035182191	3.140728764		38.93808826
17	0.977237704	15.99932427	0.51492359	5.05969436	0.349575482	4.398301928		38.78544321
18	0.163827068	7.063450221	0.052256219	3.209024875	0.504639739	5.018558957		33.83015021
19	0.931997768	14.47250907	0.341058726	4.364234905	0.958605522	6.834422087		43.6542277

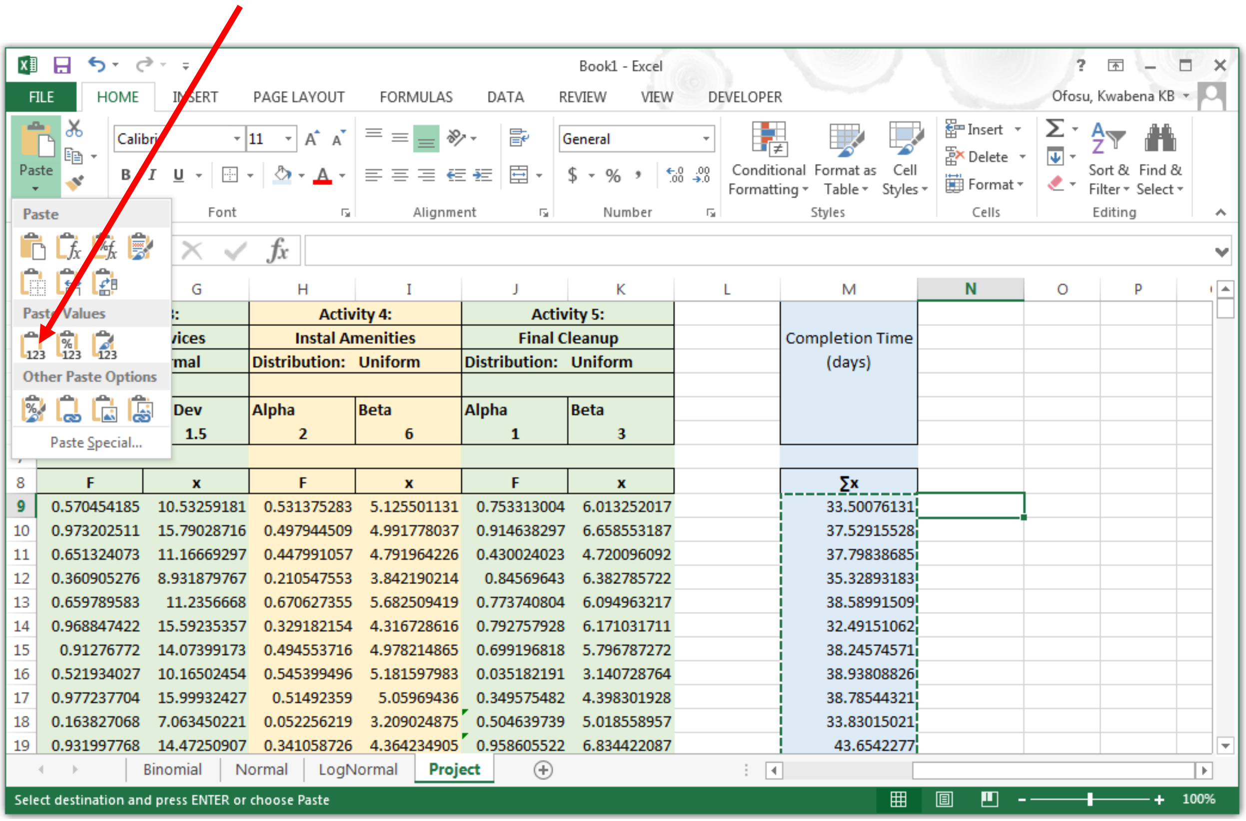
Select the destination cell.



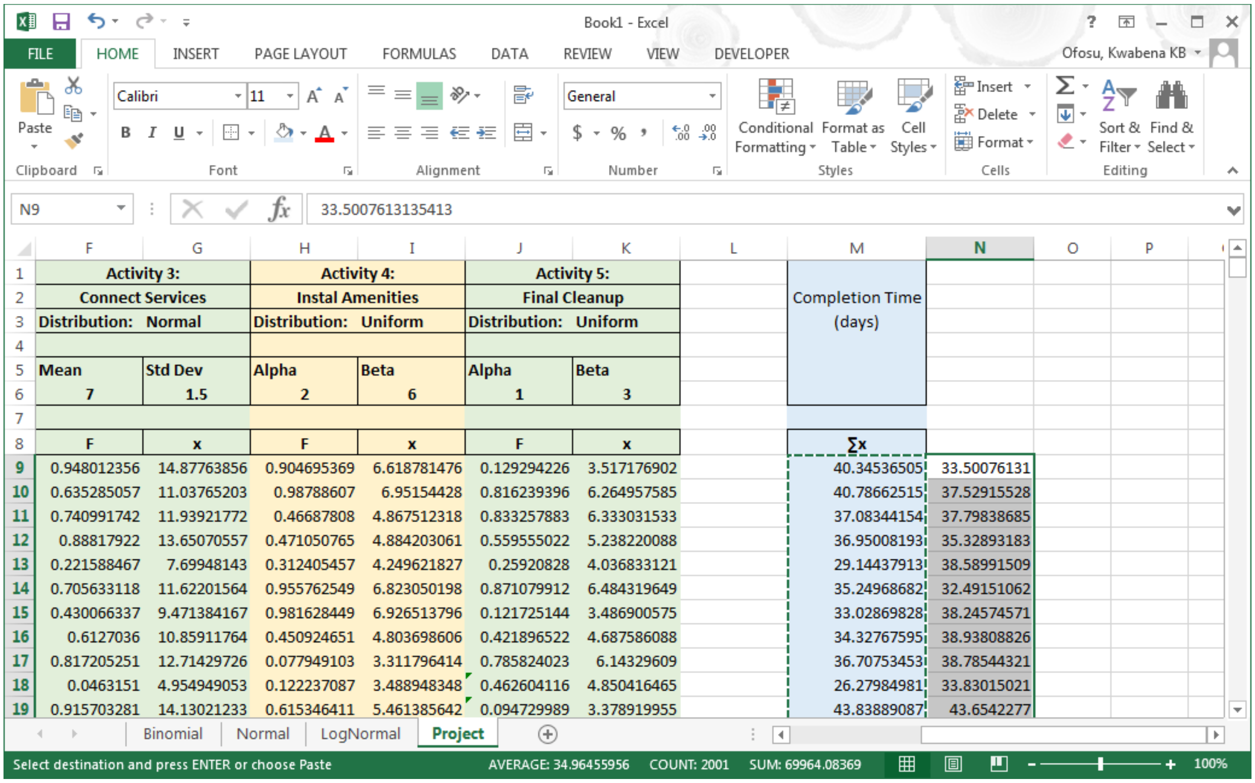
Press the Paste drop-down button.



Click on Paste Values



A stable (non-changing) dataset of completion times values is obtained.

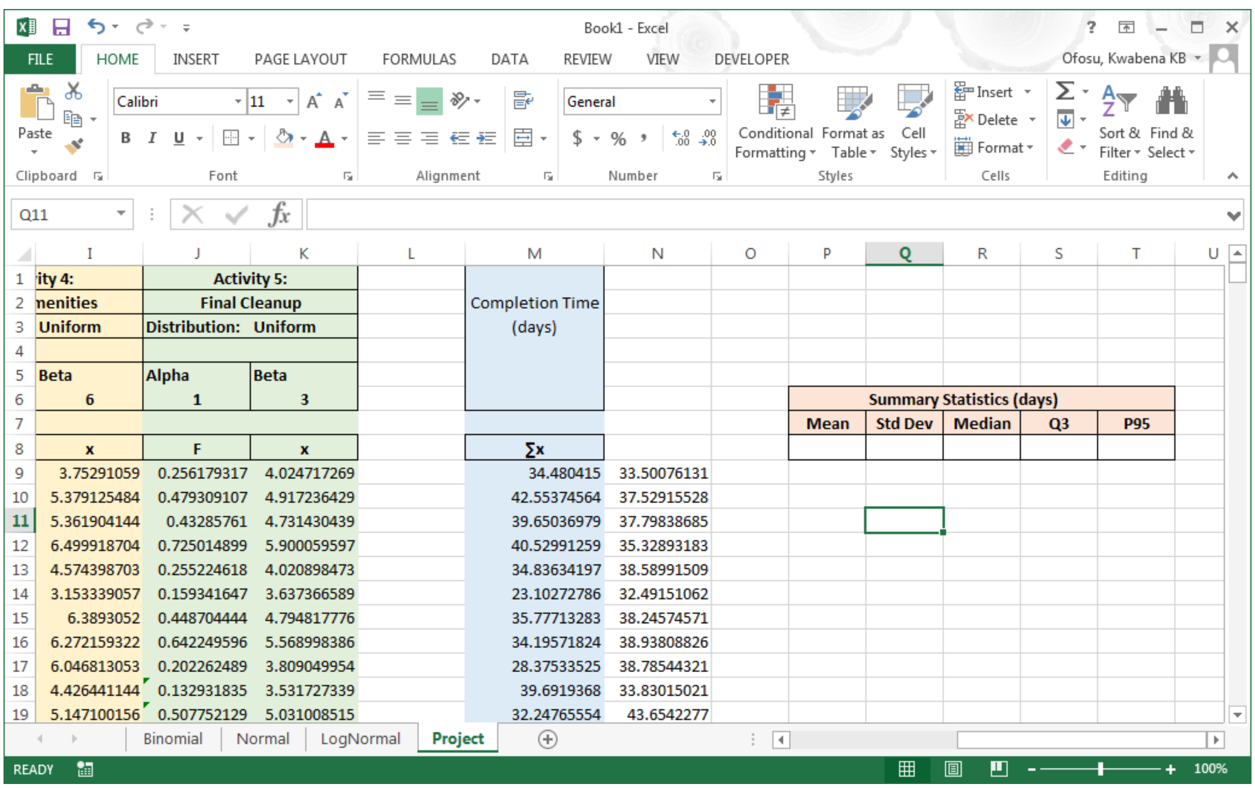


	F	G	H	I	J	K	L	M	N
1	Activity 3:		Activity 4:		Activity 5:			Completion Time (days)	
2	Connect Services		Instal Amenities		Final Cleanup				
3	Distribution: Normal		Distribution: Uniform		Distribution: Uniform				
4									
5	Mean	Std Dev	Alpha	Beta	Alpha	Beta			
6	7	1.5	2	6	1	3			
7									
8	F	x	F	x	F	x		∑x	
9	0.948012356	14.87763856	0.904695369	6.618781476	0.129294226	3.517176902		40.34536505	33.50076131
10	0.635285057	11.03765203	0.98788607	6.95154428	0.816239396	6.264957585		40.78662515	37.52915528
11	0.740991742	11.93921772	0.46687808	4.867512318	0.833257883	6.333031533		37.08344154	37.79838685
12	0.88817922	13.65070557	0.471050765	4.884203061	0.559555022	5.238220088		36.95008193	35.32893183
13	0.221588467	7.69948143	0.312405457	4.249621827	0.25920828	4.036833121		29.14437913	38.58991509
14	0.705633118	11.62201564	0.955762549	6.823050198	0.871079912	6.484319649		35.24968682	32.49151062
15	0.430066337	9.471384167	0.981628449	6.926513796	0.121725144	3.486900575		33.02869828	38.24574571
16	0.6127036	10.85911764	0.450924651	4.803698606	0.421896522	4.687586088		34.32767595	38.93808826
17	0.817205251	12.71429726	0.077949103	3.311796414	0.785824023	6.14329609		36.70753453	38.78544321
18	0.0463151	4.954949053	0.122237087	3.488948348	0.462604116	4.850416465		26.27984981	33.83015021
19	0.915703281	14.13021233	0.615346411	5.461385642	0.094729989	3.378919955		43.83889087	43.6542277

The next step is to compute statistical measures that shall serve as the performance measures for this simulation study. For the purposes of this exercise, the following statistical measures shall be computed.

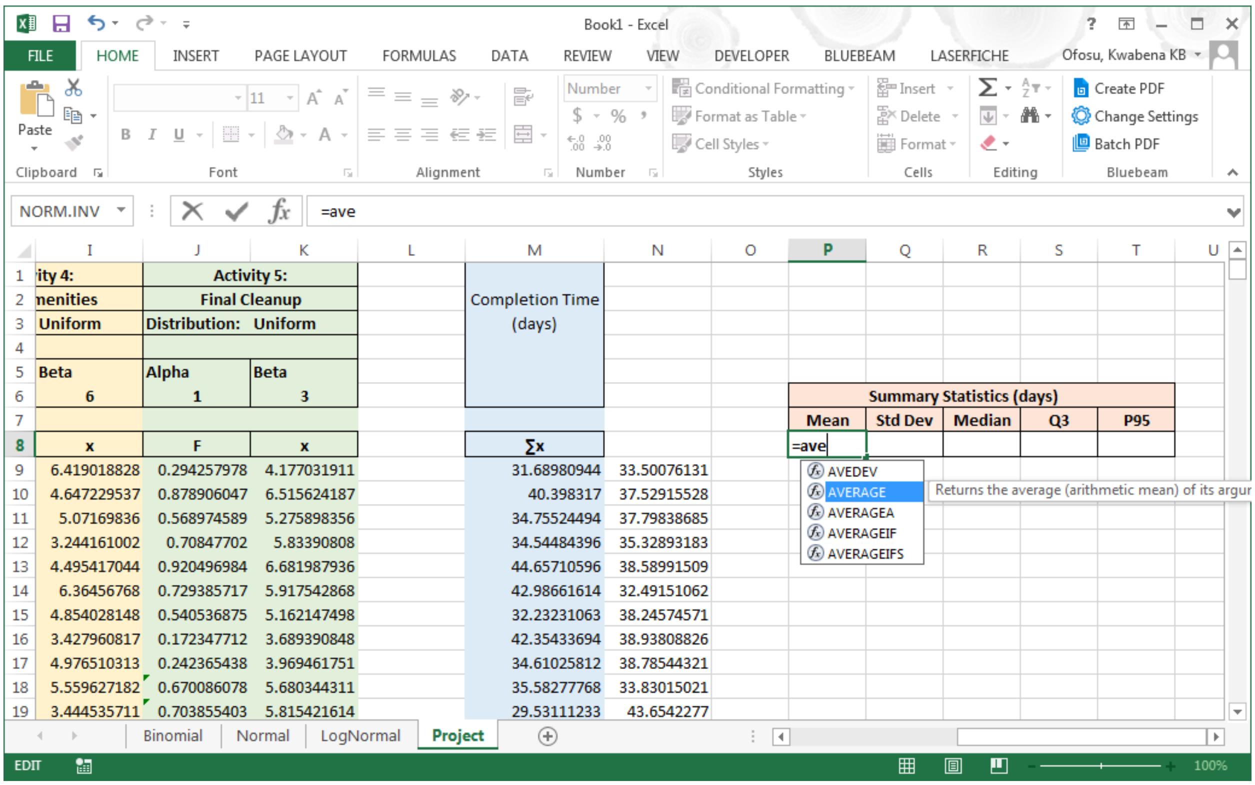
- Mean
- Standard Deviation
- Median
- 3rd Quartile
- 95% Percentile

Add a summary table to the spreadsheet as follows.

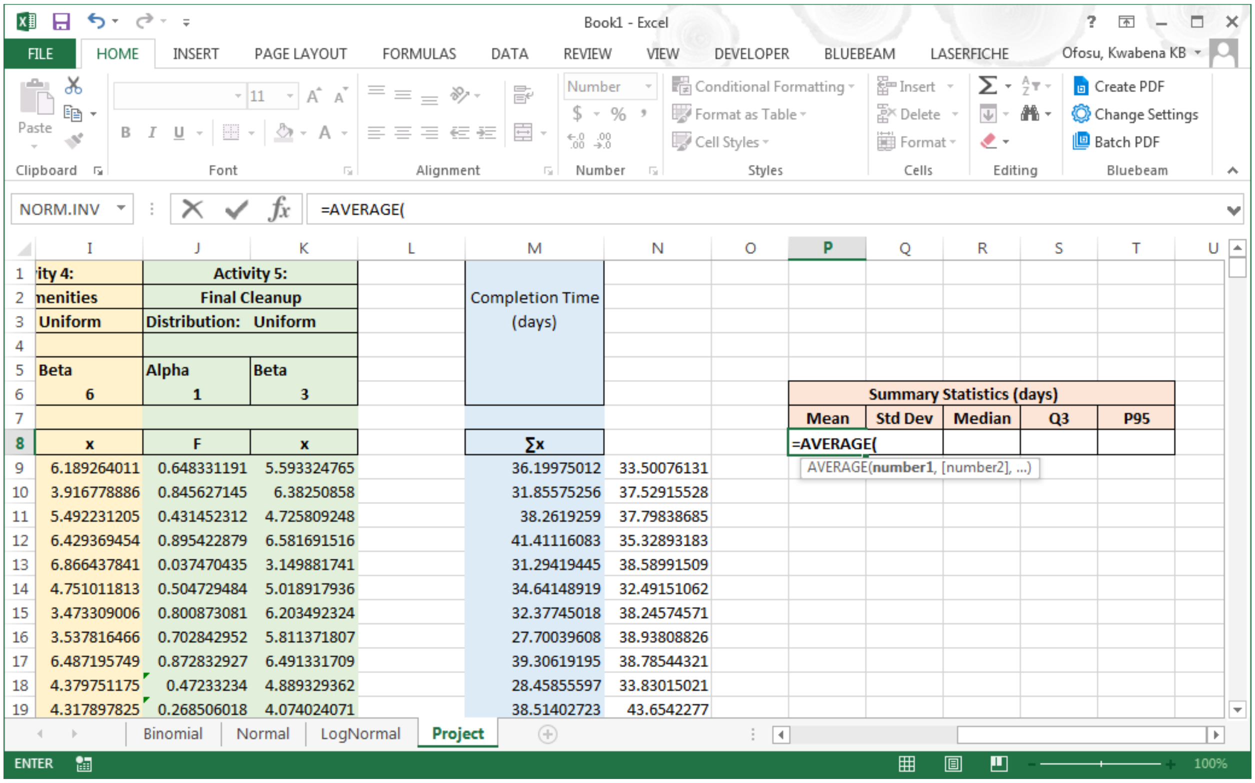


Mean (or Average)

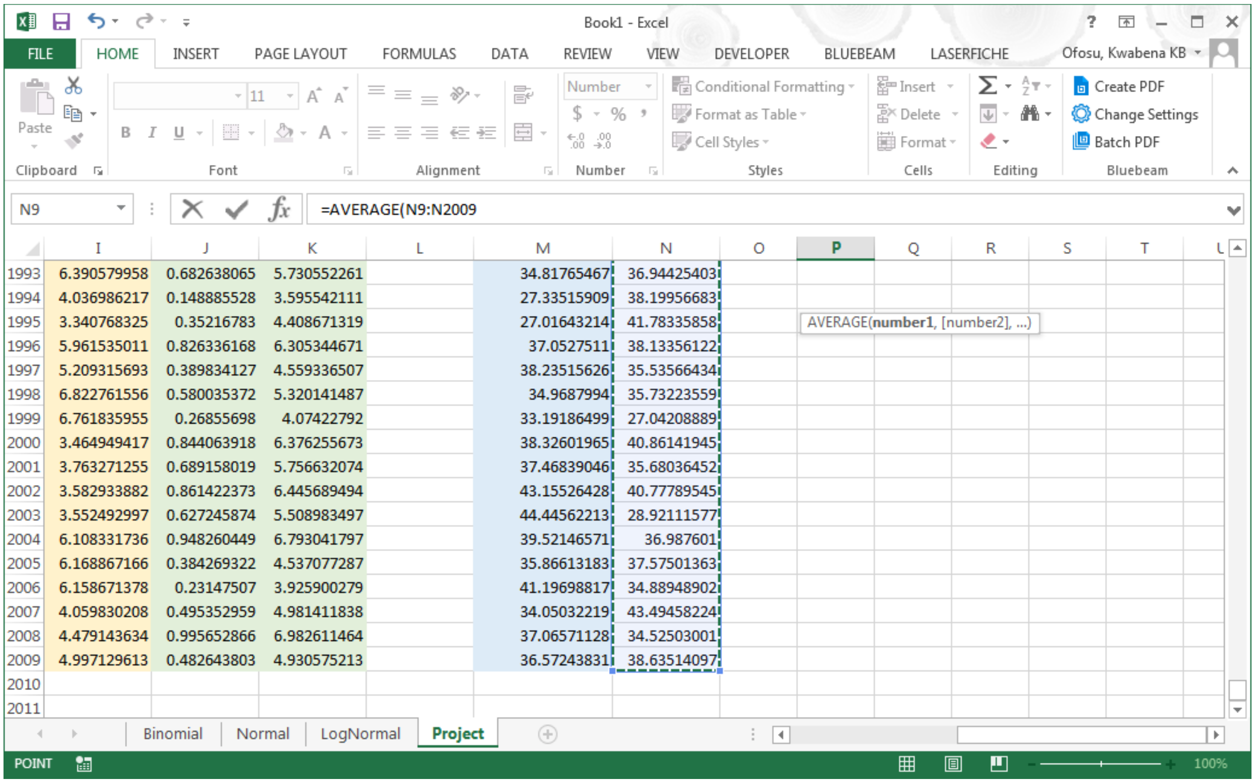
Type "=AVERAGE"



Select, and then double click on the AVERAGE function.

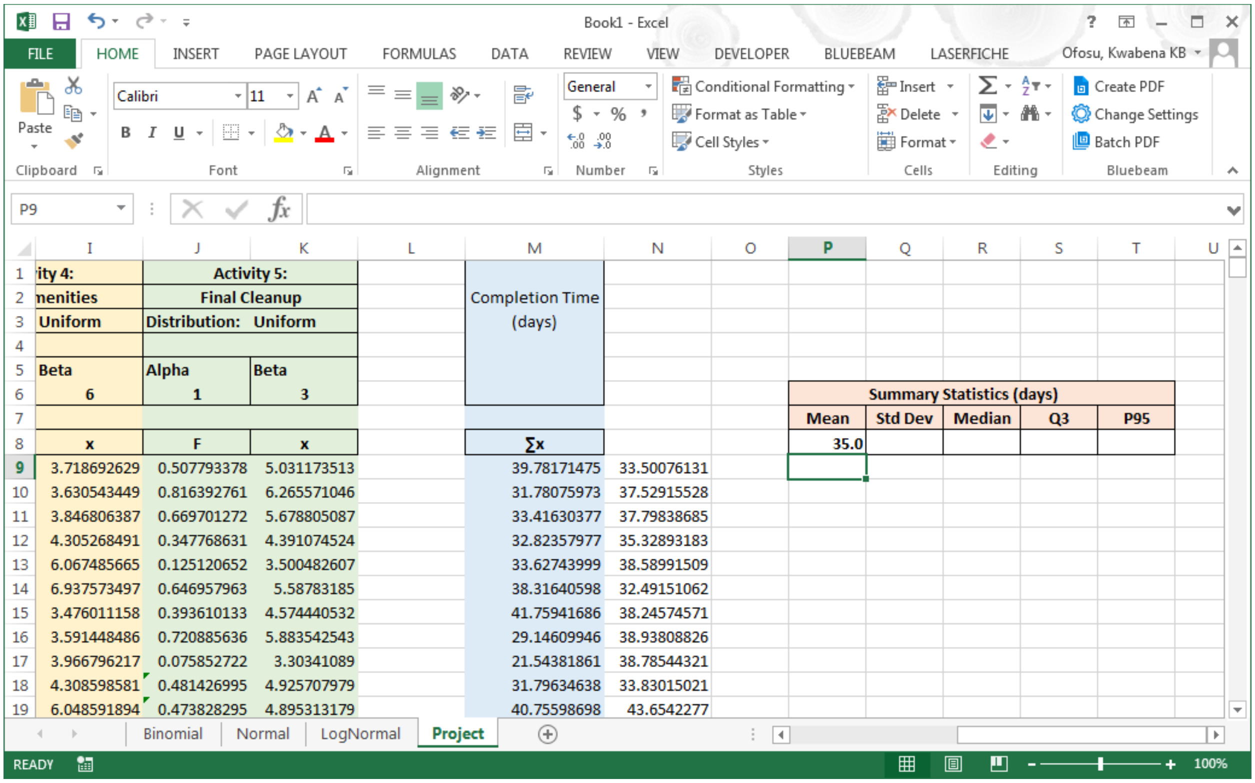


Select your “stable” completion times data.



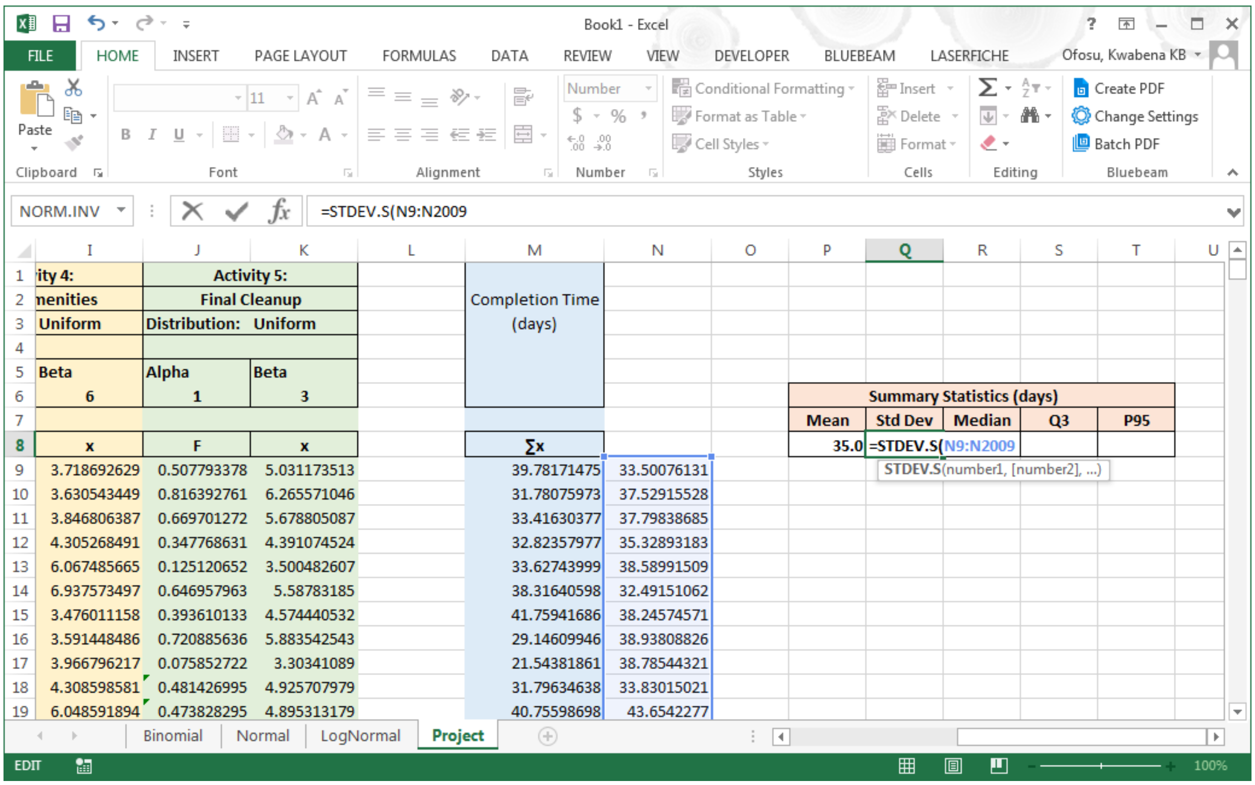
Close parenthesis.
Hit **Enter** on the keyboard.

The mean completion time is computed and displayed.

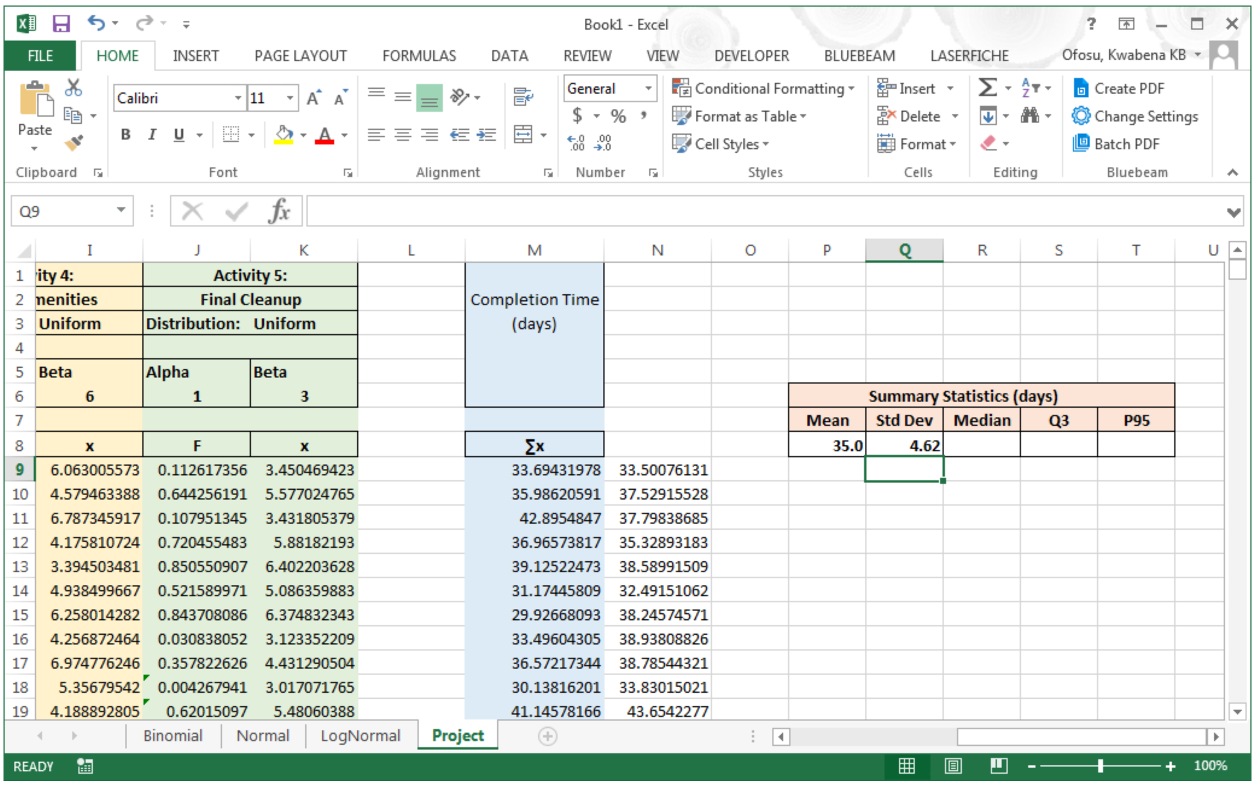


<u>Standard Deviation</u>

Type “=STANDARD”
Select, and then double click on STDEV.S
Select your “stable” completion times data.



Close parenthesis.
Hit **Enter** on the keyboard.
The standard deviation of the completion times is computed and displayed.



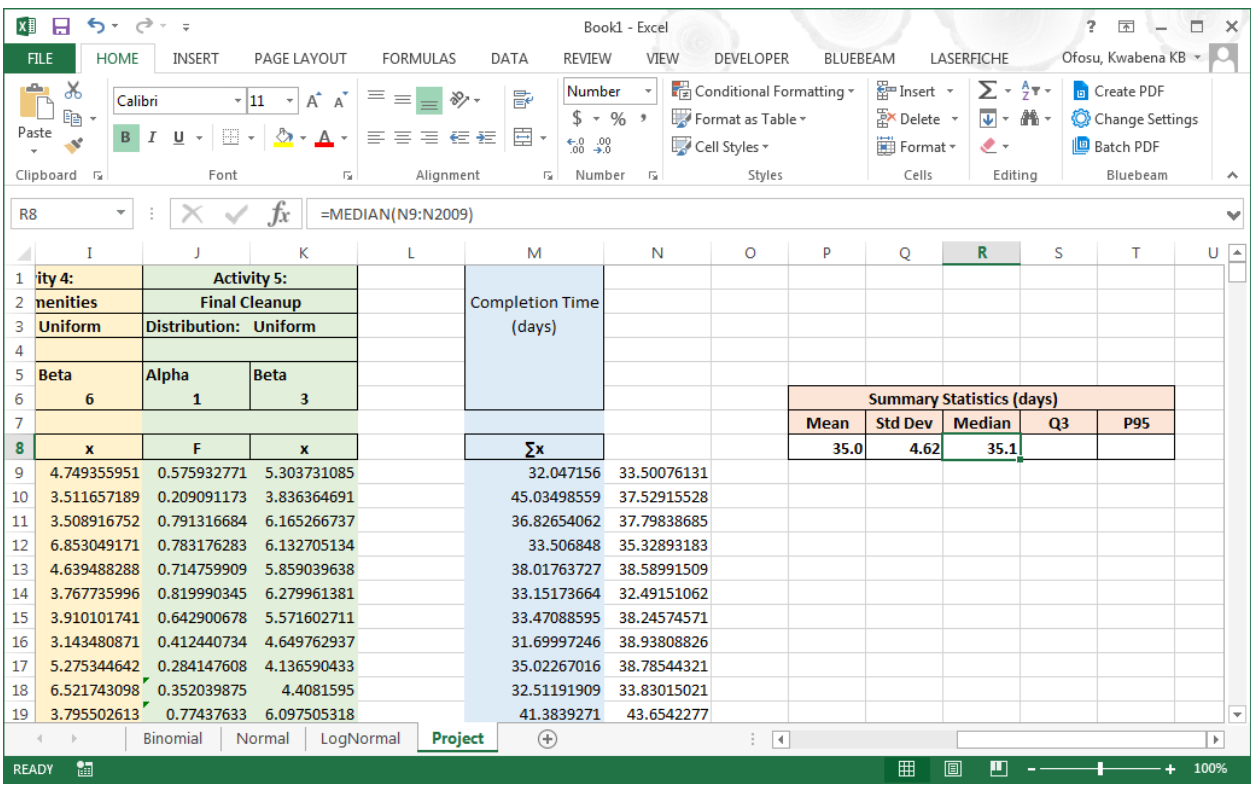
	I	J	K	L	M	N	O	P	Q	R	S	T
1	ity 4:	Activity 5:										
2	nenities	Final Cleanup			Completion Time (days)							
3	Uniform	Distribution: Uniform										
4												
5	Beta	Alpha	Beta									
6	6	1	3					Summary Statistics (days)				
7								Mean	Std Dev	Median	Q3	P95
8	x	F	x		∑x			35.0	4.62			
9	6.063005573	0.112617356	3.450469423		33.69431978	33.50076131						
10	4.579463388	0.644256191	5.577024765		35.98620591	37.52915528						
11	6.787345917	0.107951345	3.431805379		42.8954847	37.79838685						
12	4.175810724	0.720455483	5.88182193		36.96573817	35.32893183						
13	3.394503481	0.850550907	6.402203628		39.12522473	38.58991509						
14	4.938499667	0.521589971	5.086359883		31.17445809	32.49151062						
15	6.258014282	0.843708086	6.374832343		29.92668093	38.24574571						
16	4.256872464	0.030838052	3.123352209		33.49604305	38.93808826						
17	6.974776246	0.357822626	4.431290504		36.57217344	38.78544321						
18	5.35679542	0.004267941	3.017071765		30.13816201	33.83015021						
19	4.188892805	0.62015097	5.48060388		41.14578166	43.6542277						

Thus, on the "average" the completion times deviate from the mean value by 4.62 days.

Median

The median is a value such that half (or 50%) of the data values are less than it, and 50% of the data is greater than it. Thus, there is a 50% probability that the project completion time will be less than this value and a 50% probability that the project completion time will be greater than this value.

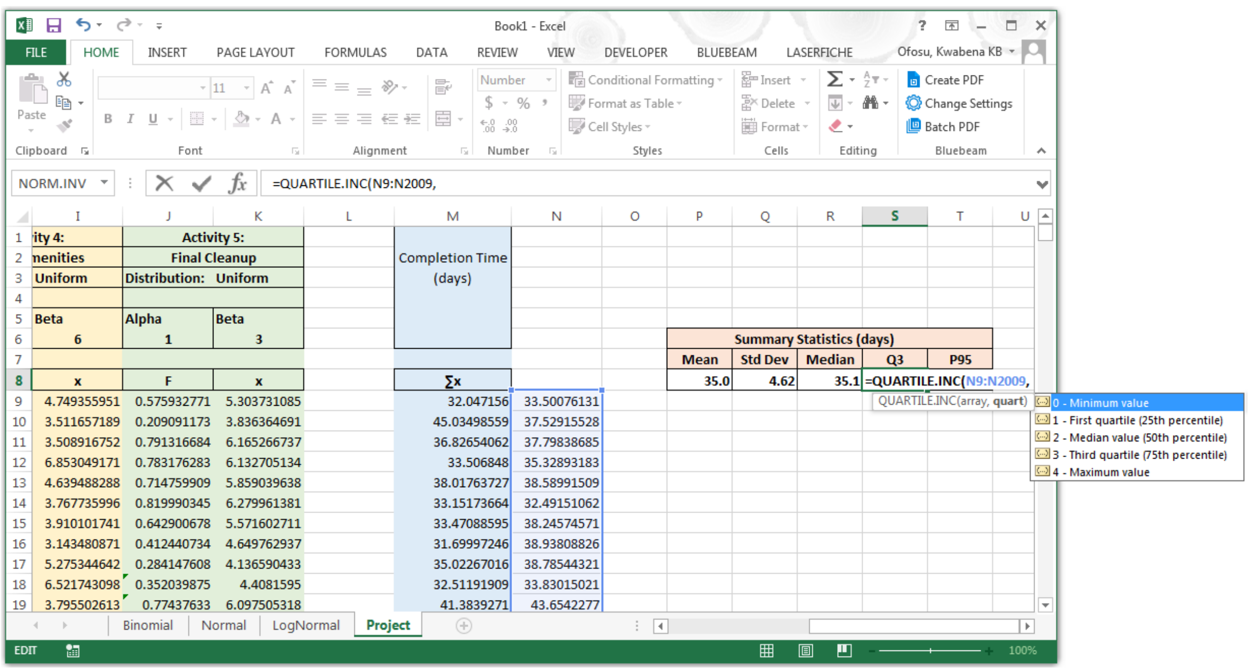
Type “=MEDIAN”
Select, and then double click on MEDIAN
Select your “stable” completion times data.
Close parenthesis.
Hit **Enter** on the keyboard.
The median completion time is computed and displayed.



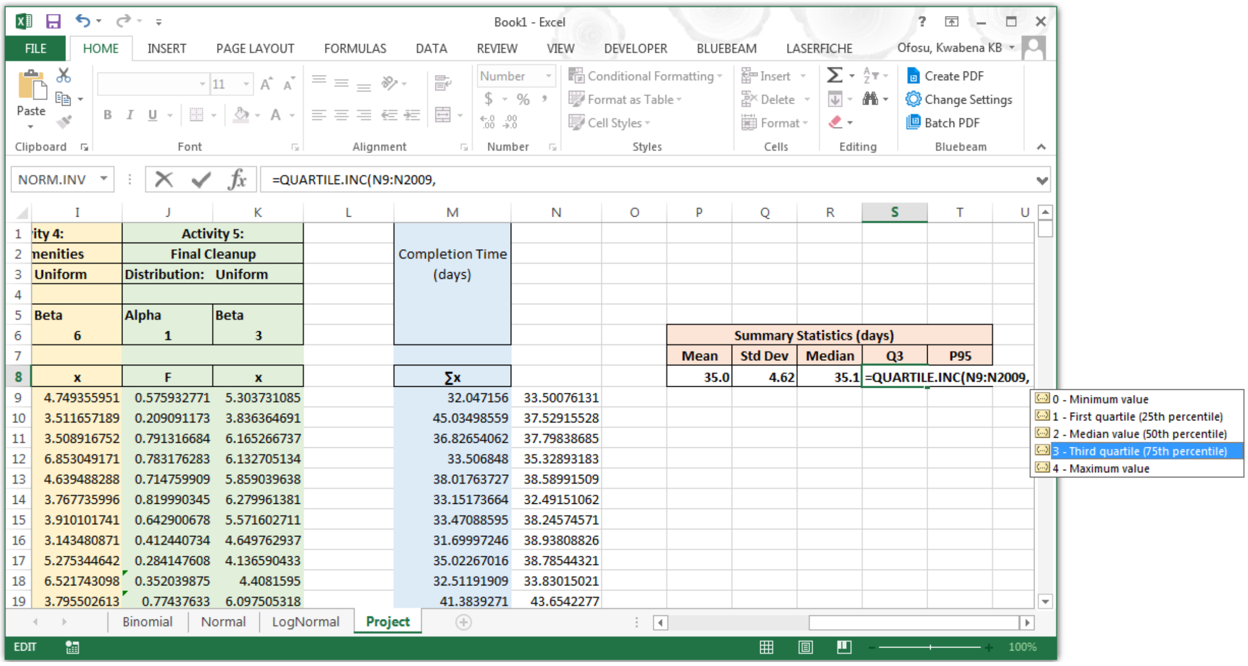
3rd Quartile (also called the Upper Quartile)

The 3rd quartile is a value such that three-quarters (or 75%) of the data values are less than it. Thus, there is a 75% probability that the project completion time will not exceed this value.

Type "=QUARTILE"
Select, and then double click on QUARTILE.INC (note that QUARTILE.EXC also computes quartiles by using a different *Excel* algorithm).
For the **array** argument, select your "stable" completion times data.
Type a comma (",")

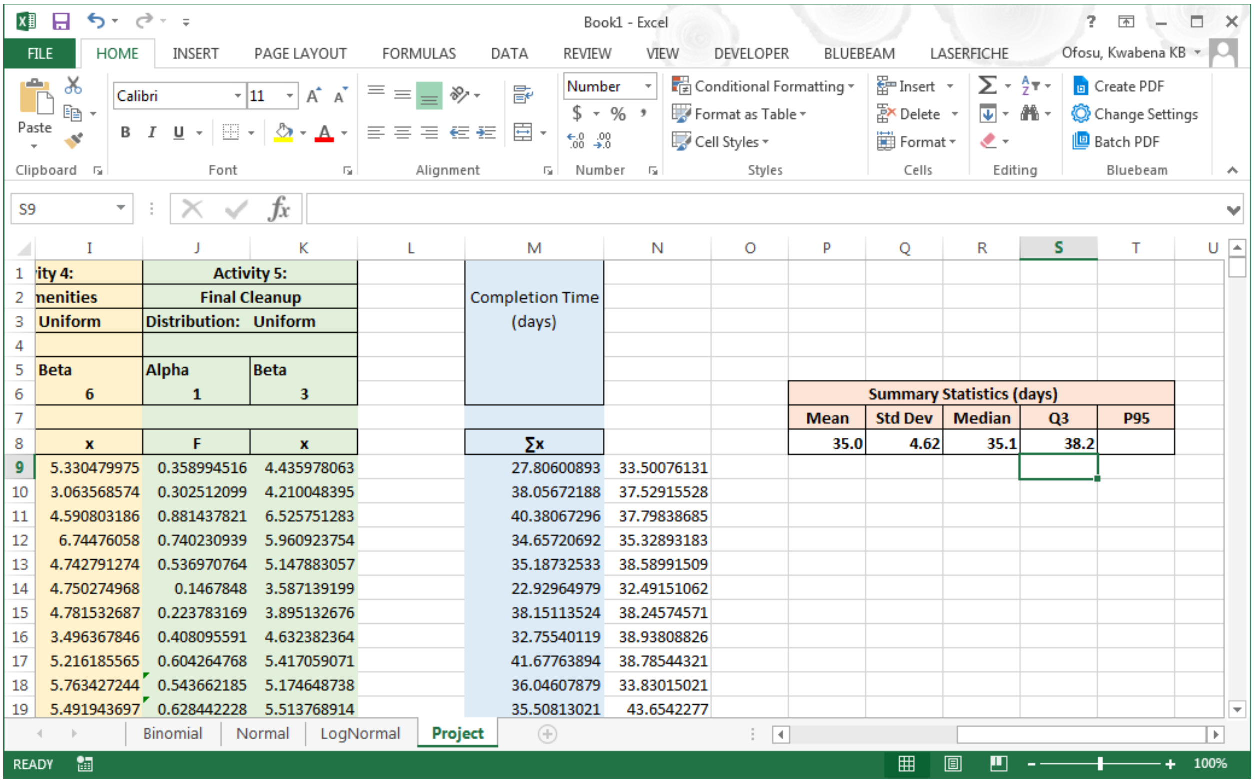


For the **quart** argument, select, and then double click "Third quartile (75th percentile)"



Close parenthesis.
Hit **Enter** on the keyboard.

The 3rd quartile completion time is computed and displayed.

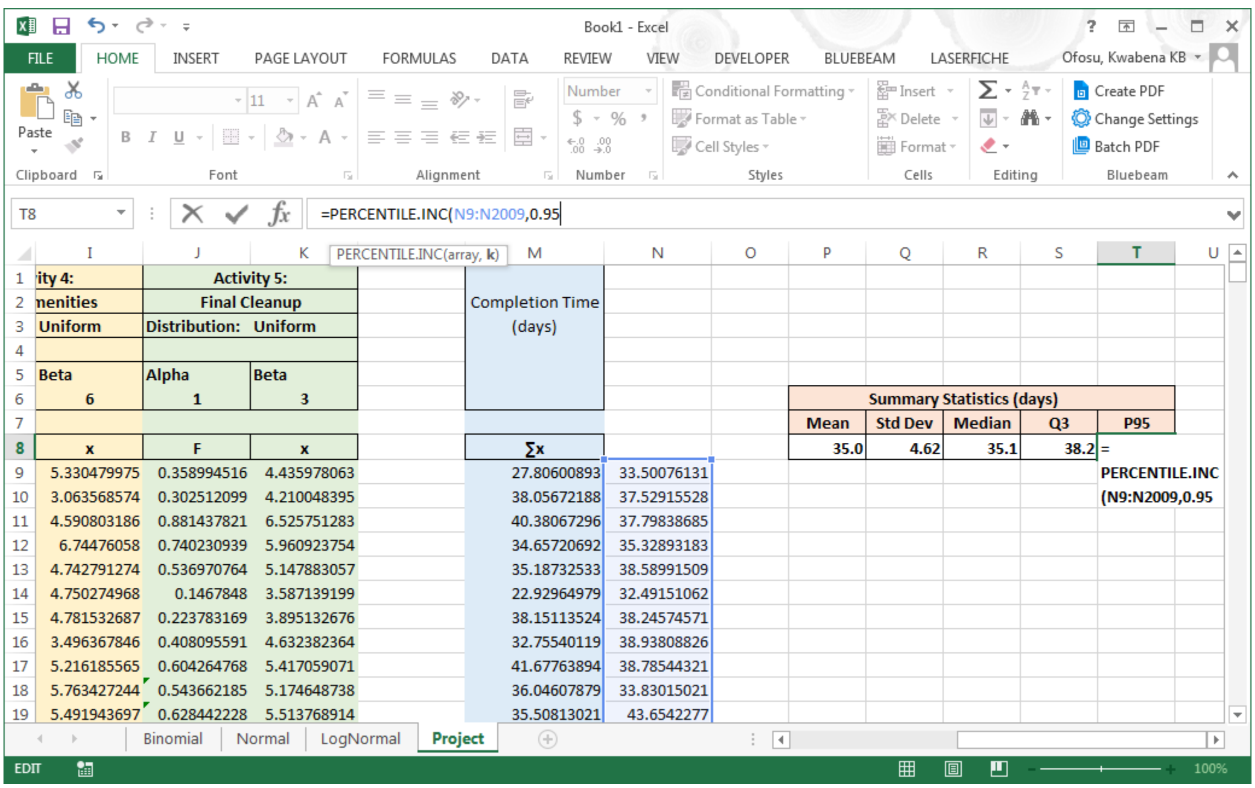


	I	J	K	L	M	N	O	P	Q	R	S	T
1	ity 4:	Activity 5:			Completion Time (days)							
2	nenities	Final Cleanup										
3	Uniform	Distribution: Uniform										
4												
5	Beta	Alpha	Beta									
6	6	1	3					Summary Statistics (days)				
7								Mean	Std Dev	Median	Q3	P95
8	x	F	x		∑x			35.0	4.62	35.1	38.2	
9	5.330479975	0.358994516	4.435978063		27.80600893	33.50076131						
10	3.063568574	0.302512099	4.210048395		38.05672188	37.52915528						
11	4.590803186	0.881437821	6.525751283		40.38067296	37.79838685						
12	6.74476058	0.740230939	5.960923754		34.65720692	35.32893183						
13	4.742791274	0.536970764	5.147883057		35.18732533	38.58991509						
14	4.750274968	0.1467848	3.587139199		22.92964979	32.49151062						
15	4.781532687	0.223783169	3.895132676		38.15113524	38.24574571						
16	3.496367846	0.408095591	4.632382364		32.75540119	38.93808826						
17	5.216185565	0.604264768	5.417059071		41.67763894	38.78544321						
18	5.763427244	0.543662185	5.174648738		36.04607879	33.83015021						
19	5.491943697	0.628442228	5.513768914		35.50813021	43.6542277						

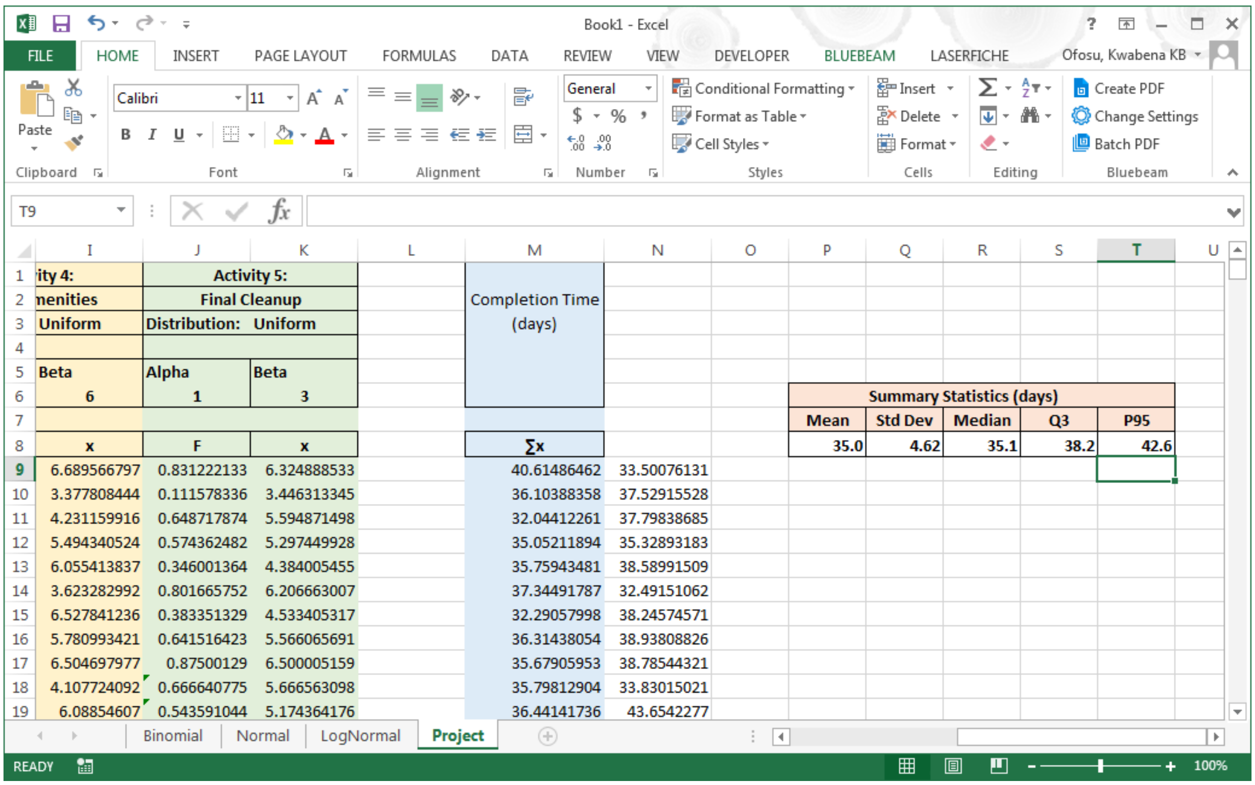
95th Percentile

The 95th percentile is a value such that 95% of the data values are less than it. Thus, there is a 95% probability that the project completion time will not exceed this value. One is therefore 95% confident that the project will be completed by this time.

Type "=PERCENTILE"
For the **array** argument, select, and then double click on PERCENTILE.INC (note that PERCENTILE.EXC also computes percentiles by using a different *Excel* algorithm).
Select your "stable" completion times data.
Type a comma (",")
For the **k** argument, type in "0.95"



Close parenthesis.
Hit **Enter** on the keyboard.
The 95th percentile completion time is computed and displayed.



	I	J	K	M	N
1	ity 4:	Activity 5:			
2	nenities	Final Cleanup		Completion Time (days)	
3	Uniform	Distribution:	Uniform		
4					
5	Beta	Alpha	Beta		
6	6	1	3		
7					
8	x	F	x	∑x	
9	6.689566797	0.831222133	6.324888533	40.61486462	33.50076131
10	3.377808444	0.111578336	3.446313345	36.10388358	37.52915528
11	4.231159916	0.648717874	5.594871498	32.04412261	37.79838685
12	5.494340524	0.574362482	5.297449928	35.05211894	35.32893183
13	6.055413837	0.346001364	4.384005455	35.75943481	38.58991509
14	3.623282992	0.801665752	6.206663007	37.34491787	32.49151062
15	6.527841236	0.383351329	4.533405317	32.29057998	38.24574571
16	5.780993421	0.641516423	5.566065691	36.31438054	38.93808826
17	6.504697977	0.87500129	6.500005159	35.67905953	38.78544321
18	4.107724092	0.666640775	5.666563098	35.79812904	33.83015021
19	6.08854607	0.543591044	5.174364176	36.44141736	43.6542277

Summary Statistics (days)				
Mean	Std Dev	Median	Q3	P95
35.0	4.62	35.1	38.2	42.6

The next step is the analysis of the results, for making conclusions, recommendations or decision-making.

Many engineering fields have adopted various statistical measures such as those demonstrated above as criteria for reaching conclusions and decision-making. For example, in the field of traffic engineering, the posted speed limit on a roadway is based on the 85th percentile speed (except where the posted speed limit is prescribed by state law, for example, on the interstate highway system in the State of Florida). Another example from traffic engineering is the length of a left-turn lane at an intersection controlled by a traffic signal. Generally, the turn lane length is based on the 95th percentile vehicular queue length obtained from a simulation model.

For the construction site example, a deterministic approach using the estimated average duration of each constituent work item, the conclusion was that the project will be completed in 28 days. However, using a probabilistic approach and modeling the uncertainty associated with each constituent activity by some known probability distribution, and conducting 2,000 simulations, the average project completion time is 35 days, and furthermore there is a 95% probability that the project will be completed in no more than approximately 43 days.

5. SELECTING DISTRIBUTIONS FOR SIMULATION

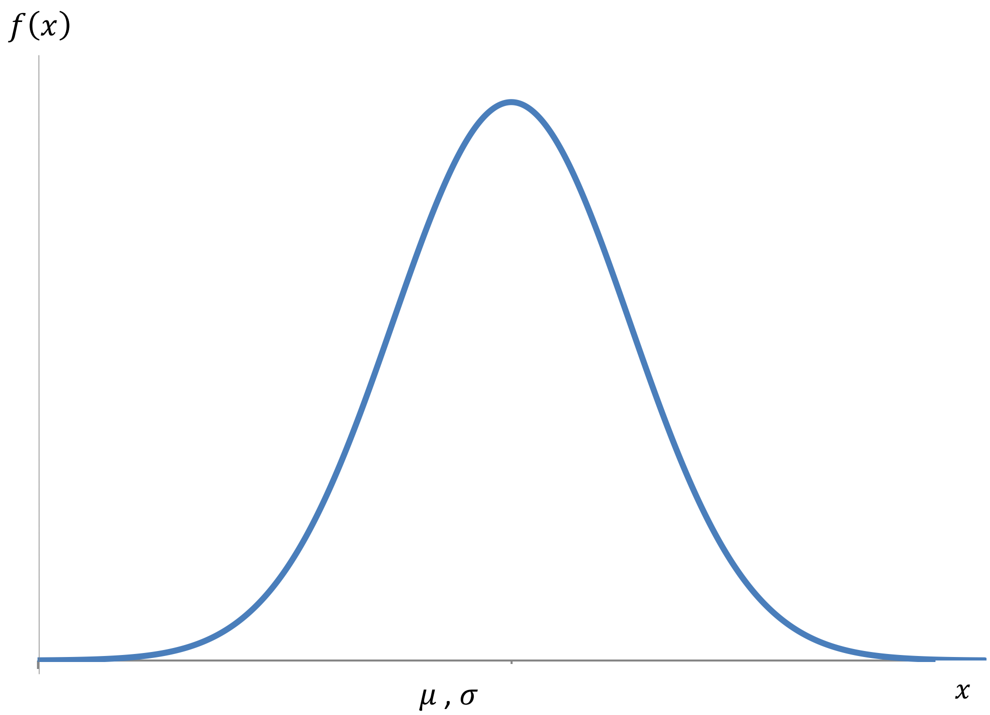
5.1 Selecting a Probability Distribution

In the construction site project simulation, each constituent activity of the project was described, or modelled, by some known statistical distribution, such as the Normal distribution, the Uniform distribution, the LogNormal distribution etc. This leads to the question as to how the distributions were selected or justified. In other words, how does one know what distribution to pick or how does one confirm or otherwise, that a given theoretical distribution is suitable to describe the activity, based on data for that activity.

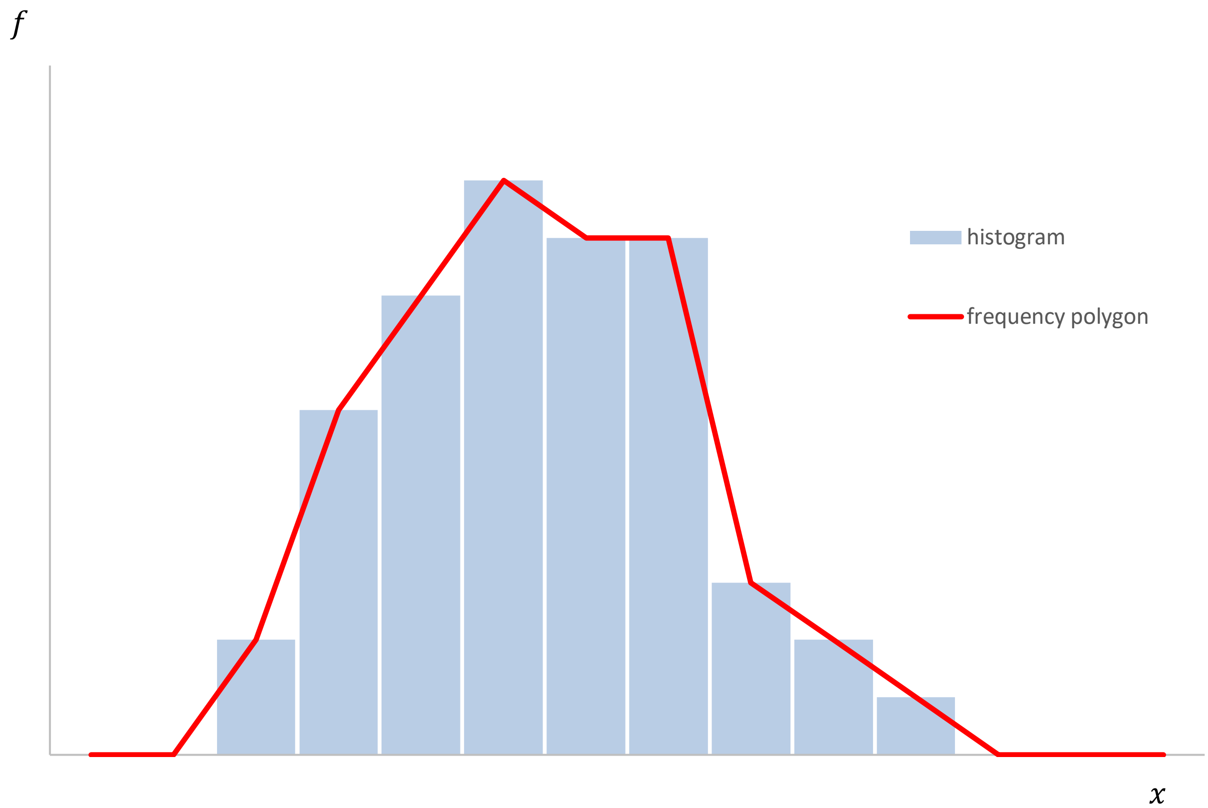
One cursory method of selecting a distribution or confirming a **distributional assumption** for an activity is to look at the shape of the histogram or frequency polygon of the data for the activity and compare it to the shape of the probability density function of the theoretical candidate distribution.

[A comprehensive presentation of histograms, frequency polygons and other frequency diagrams can be found elsewhere, including a course by this author that is available on Suncam].

For example, consider the following data for an activity, for which the simulation engineer makes an educated guess to go with the Normal distribution. Recall that the distribution function (of the data) will have a bell-shape if it is indeed described by a (theoretical) Normal distribution, such as

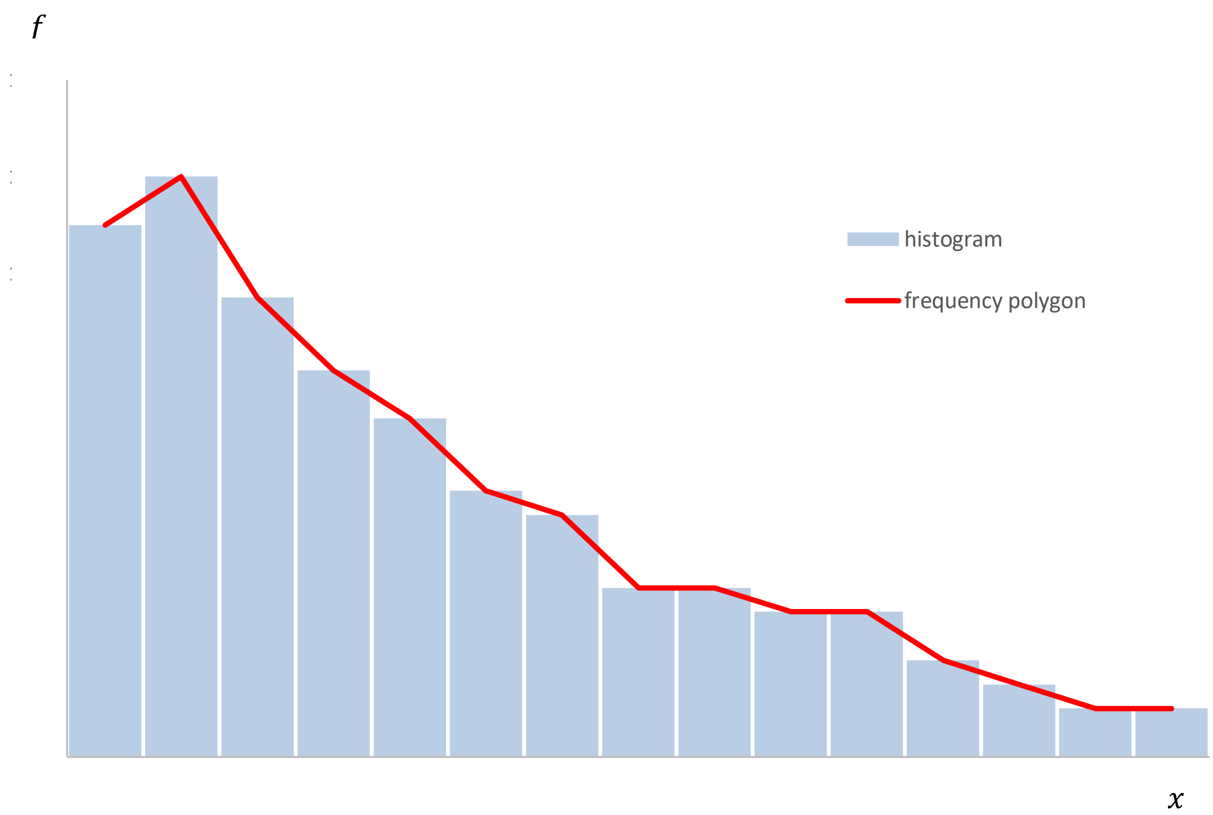

$f(x)$
μ , σ
x

The histogram (and superimposed frequency polygon) of the data turns out as follows.



The histogram (or the frequency polygon) does show a rise and fall trend of a somewhat symmetrical bell shape. Thus, the Normal distribution is a good candidate distribution to describe this data. In fact, other bell-shaped distributions may also be suitable, for example, the Logistic distribution.

In a similar analysis, the following example suggests that a LogNormal distribution may be suitable to describe the data.



In fact, a Weibull distribution or some other skewed distribution may also be a good candidate.

Inspecting the frequency diagram provides a quick and useful means of assessing the suitability of a distribution or checking a distributional assumption. However, as with graphical methods in general, it does involve an element of subjective judgement.

5.2 Probability Plot

A probability plot, also known as a **quantile plot** or **Q-Q plot**, is a graphical method for checking whether a dataset follows a given theoretical distribution or otherwise. The method involves using the cumulative probability values observed in the data to compute theoretical variates (x-values) from the quantile function of the candidate theoretical distribution. The theoretical values versus the corresponding actual (data) values are plotted. If the plot appears to be a straight line through the origin at an angle of 45 degrees to the horizontal, one can conclude that the data does indeed follow that candidate theoretical distribution. Or one may say the data comes from the candidate distribution. Generally, a dataset of at least 15 to 20 observations is required for the conclusions of the probability plot technique to be tenable.

The probability plot can be constructed by the following steps.

1. Arrange the data from smallest to largest

2. Divide the area of the distribution (the data) into $(n + 1)$ equal sub-areas of area $^{1}/_{(n+1)}$, where n is the number of observations in the data.

3. Compute the theoretical variate from the quantile function of the candidate distribution for each ordered observation, where the cumulative probability of an ordered observation i is given by

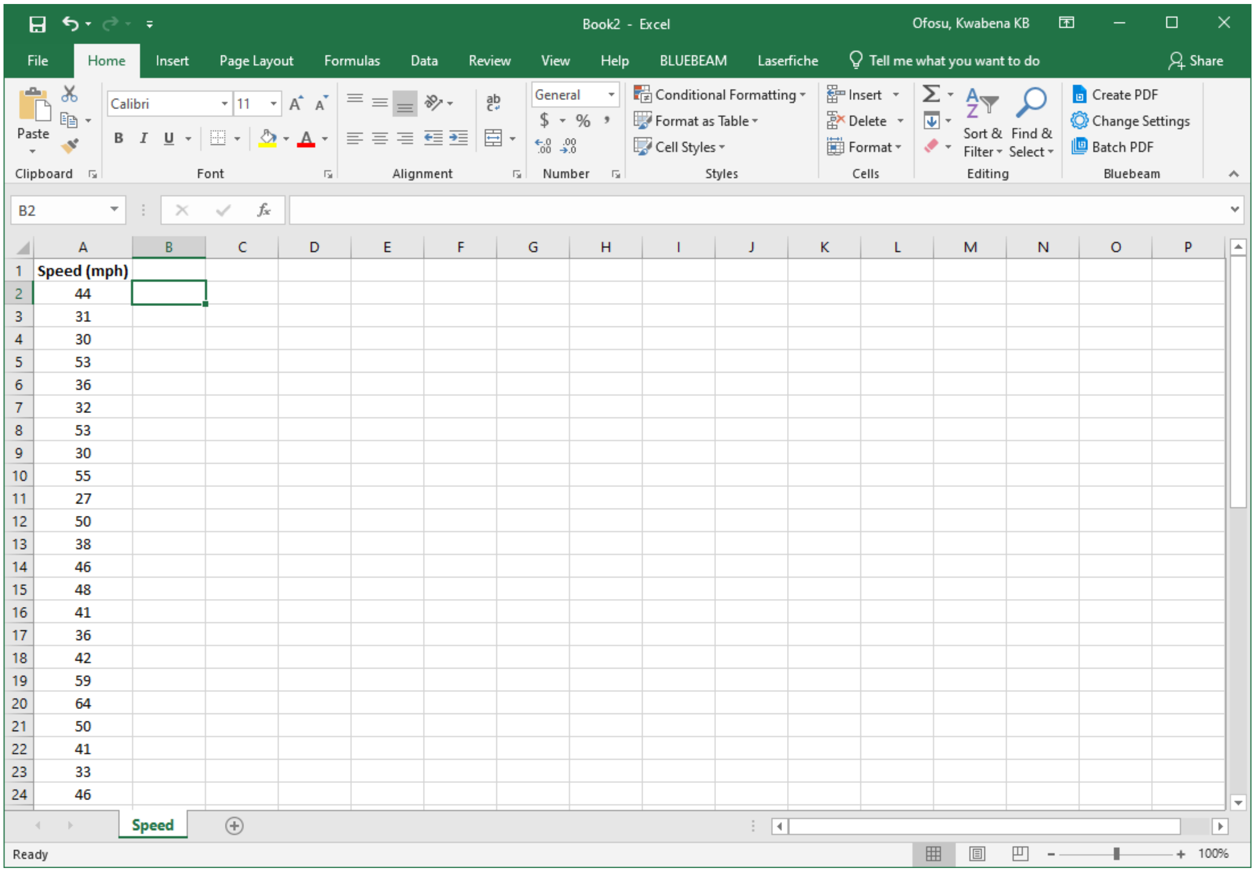
$$\sum_{k=1}^{i}\left(\frac{1}{n+1}\right)$$

4. Plot the ith ordered observation versus the ith theoretical variate.

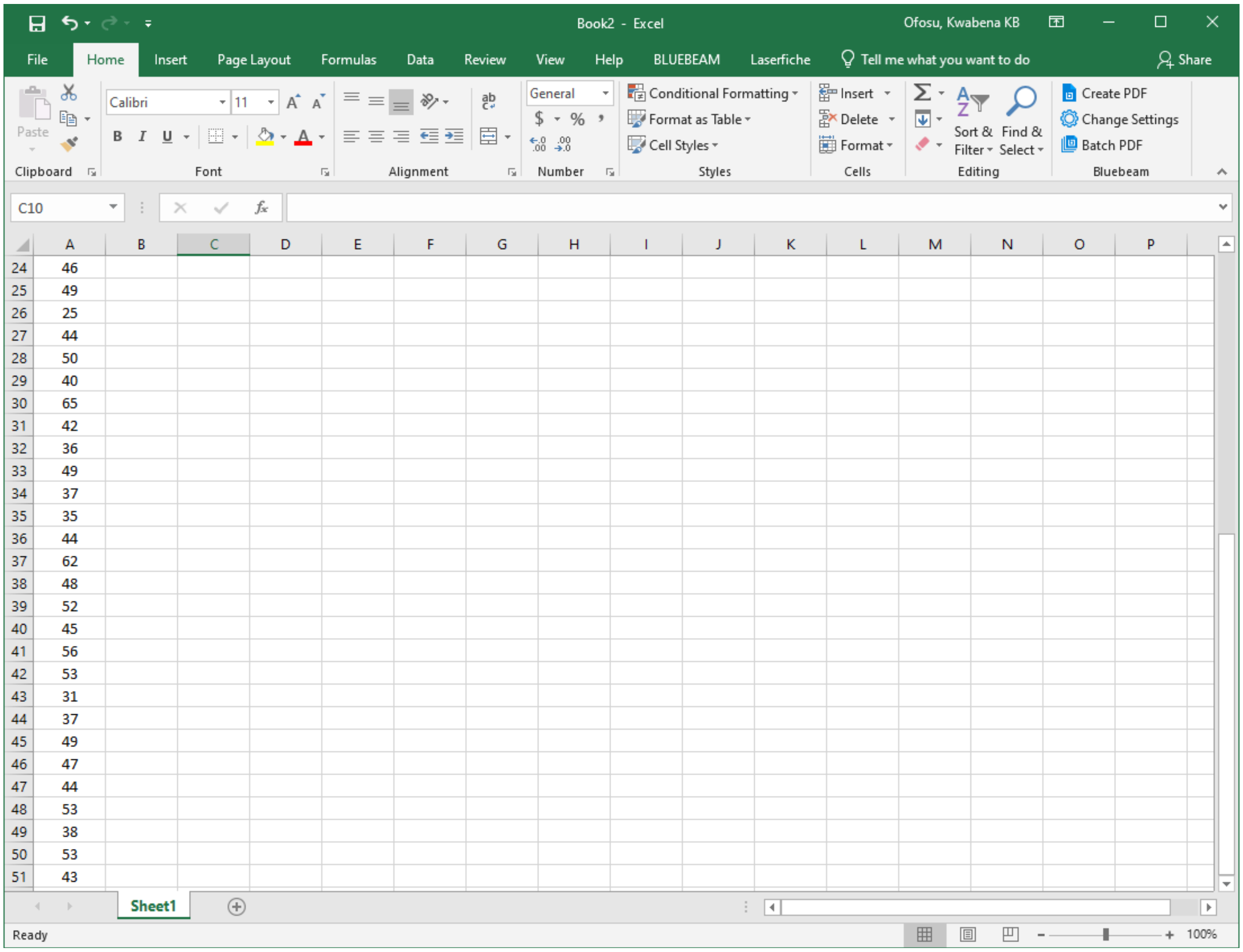
Example:

A traffic engineer constructs a probability plot from 50 vehicle speed measurements (in miles per hour) to check whether the speed data follows a Normal distribution.

The data is as follows.


Book2 - Excel
Ofosu, Kwabena KB
File
Home
Insert
Page Layout
Formulas
Data
Review
View
Help
BLUEBEAM
Laserfiche
Tell me what you want to do
Share
Paste
Calibri
11
General
Conditional Formatting
Format as Table
Cell Styles
Insert
Delete
Format
Sort & Filter
Find & Select
Create PDF
Change Settings
Batch PDF
Clipboard
Font
Alignment
Number
Styles
Cells
Editing
Bluebeam
B2
A
B
C
D
E
F
G
H
I
J
K
L
M
N
O
P
Speed (mph)
44
31
30
53
36
32
53
30
55
27
50
38
46
48
41
36
42
59
64
50
41
33
46
Speed
Ready
100%

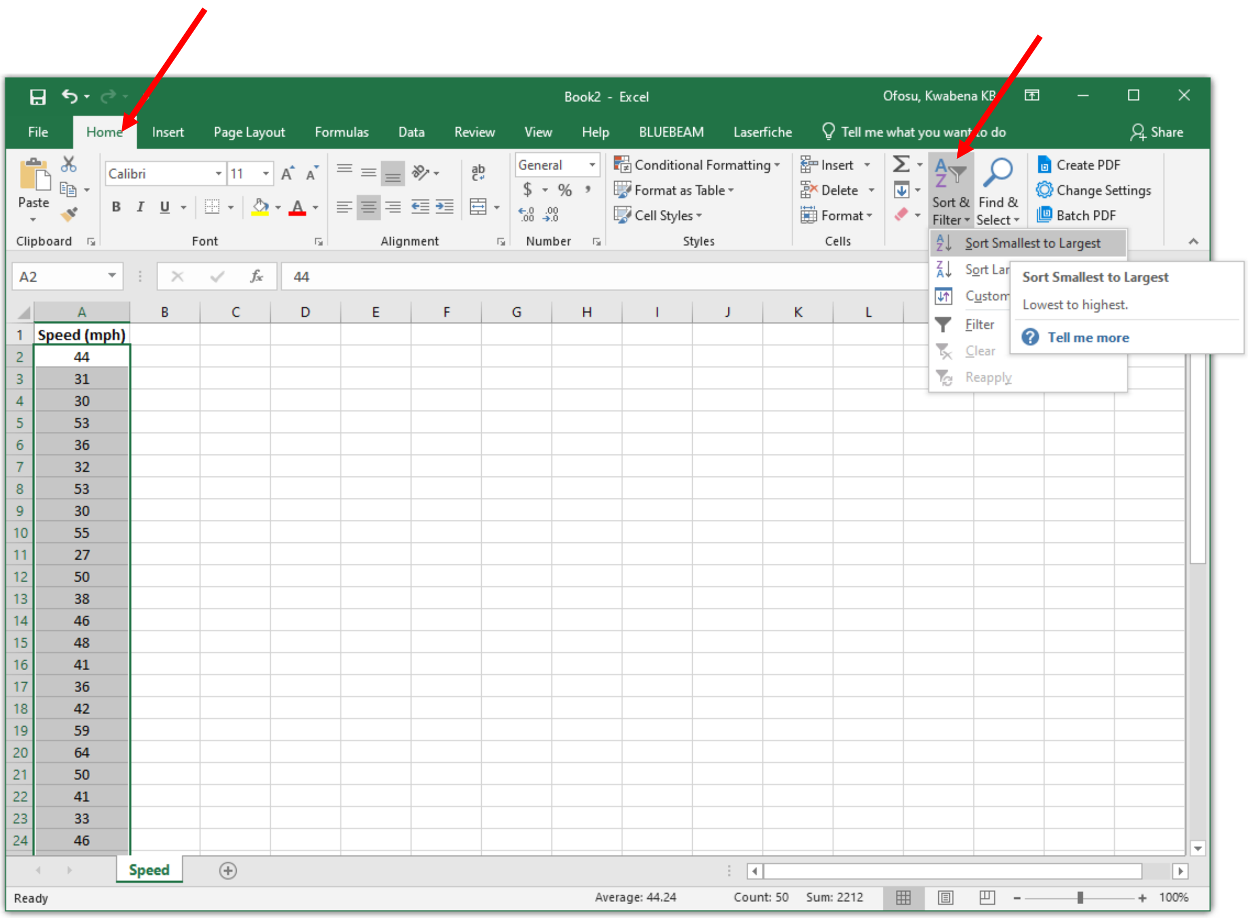
and



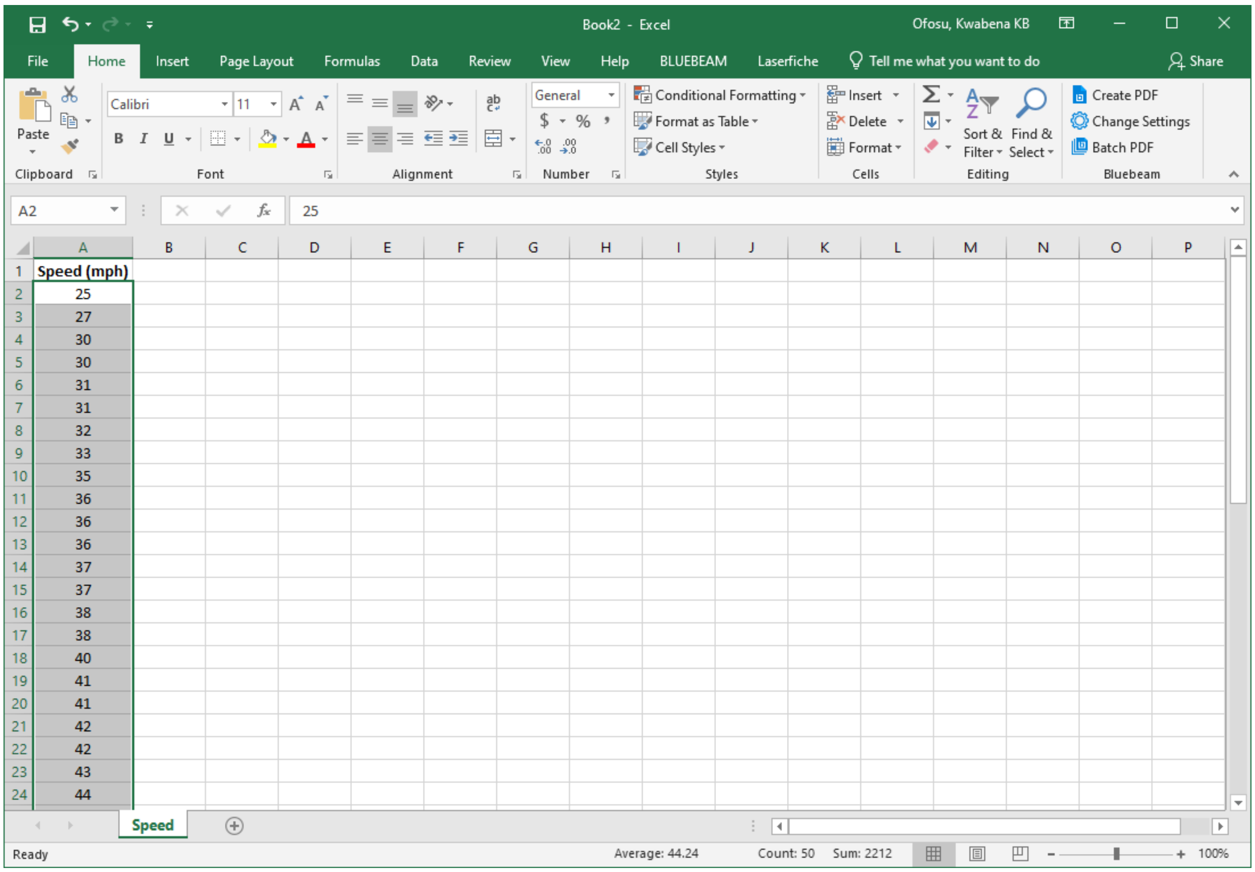
	A
24	46
25	49
26	25
27	44
28	50
29	40
30	65
31	42
32	36
33	49
34	37
35	35
36	44
37	62
38	48
39	52
40	45
41	56
42	53
43	31
44	37
45	49
46	47
47	44
48	53
49	38
50	53
51	43

Step 1:

Select the data.
Click on **Home**.
Click on **Sort & Filter**.
Click on **Sort Smallest to Largest**.



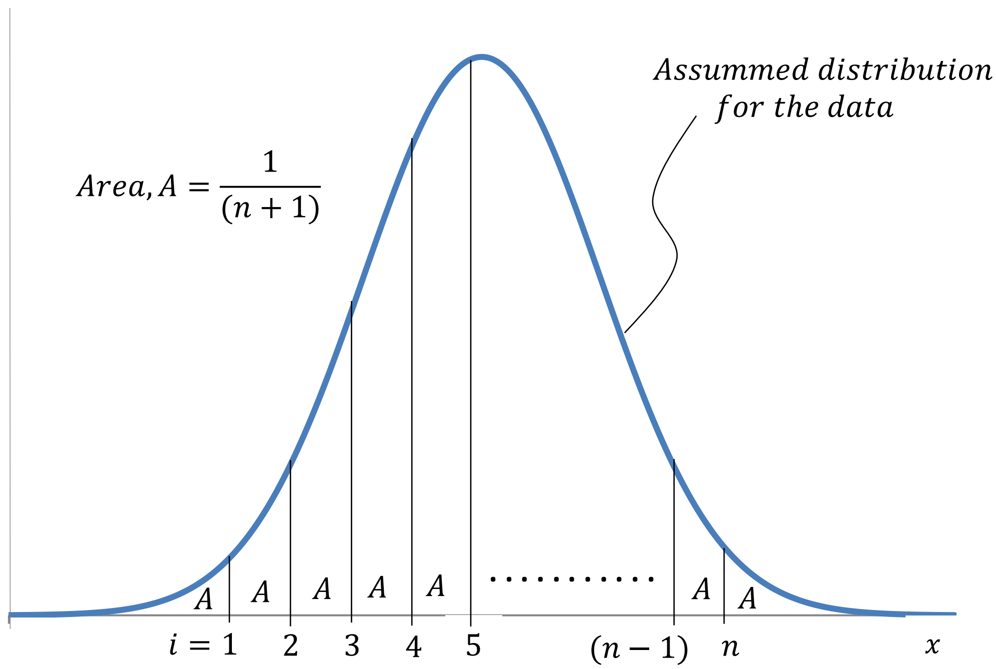
The data is rearranged in ascending order.



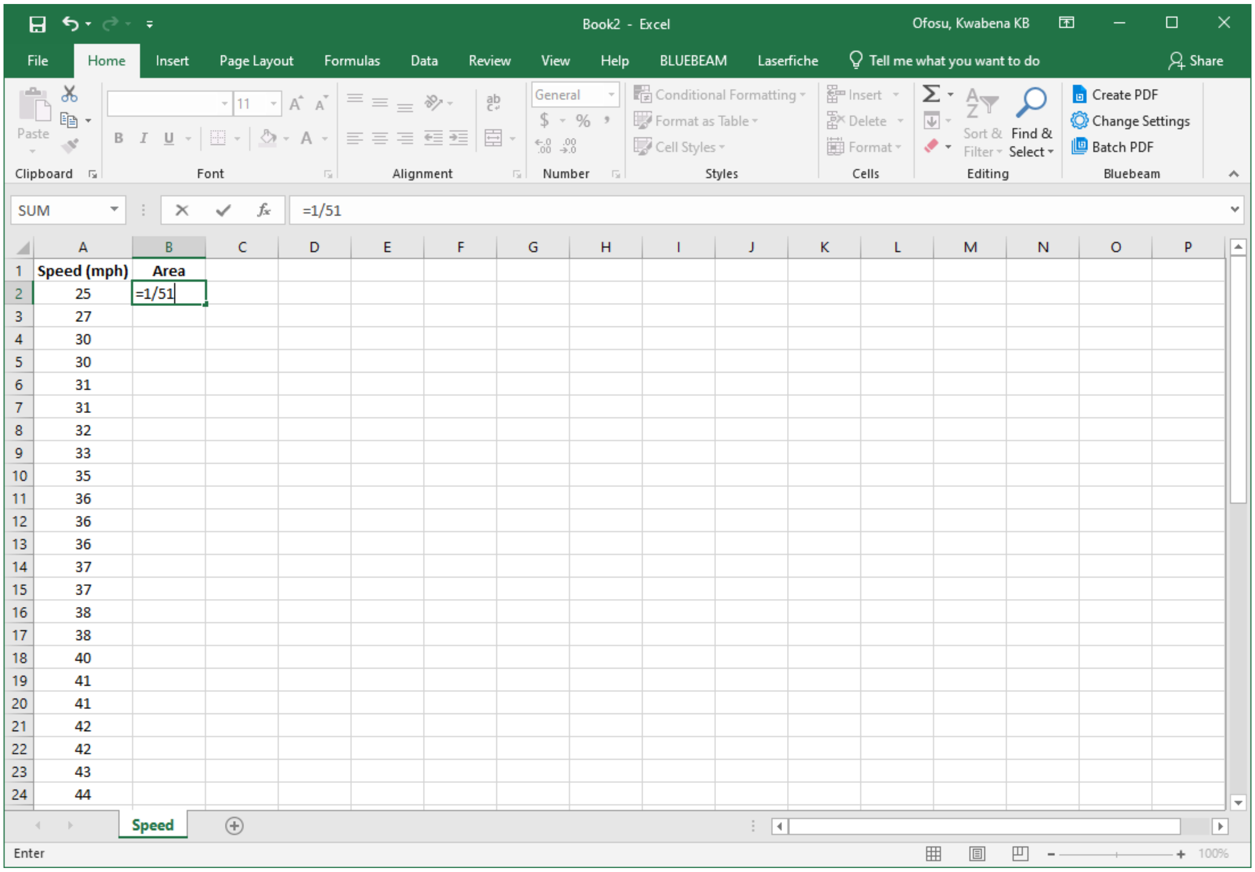
Step 2:

Divide the area under the distribution (the data) into $(50 + 1)$ equal sub-areas of area $^{1}/_{51}$ or 0.0196.

Graphically,

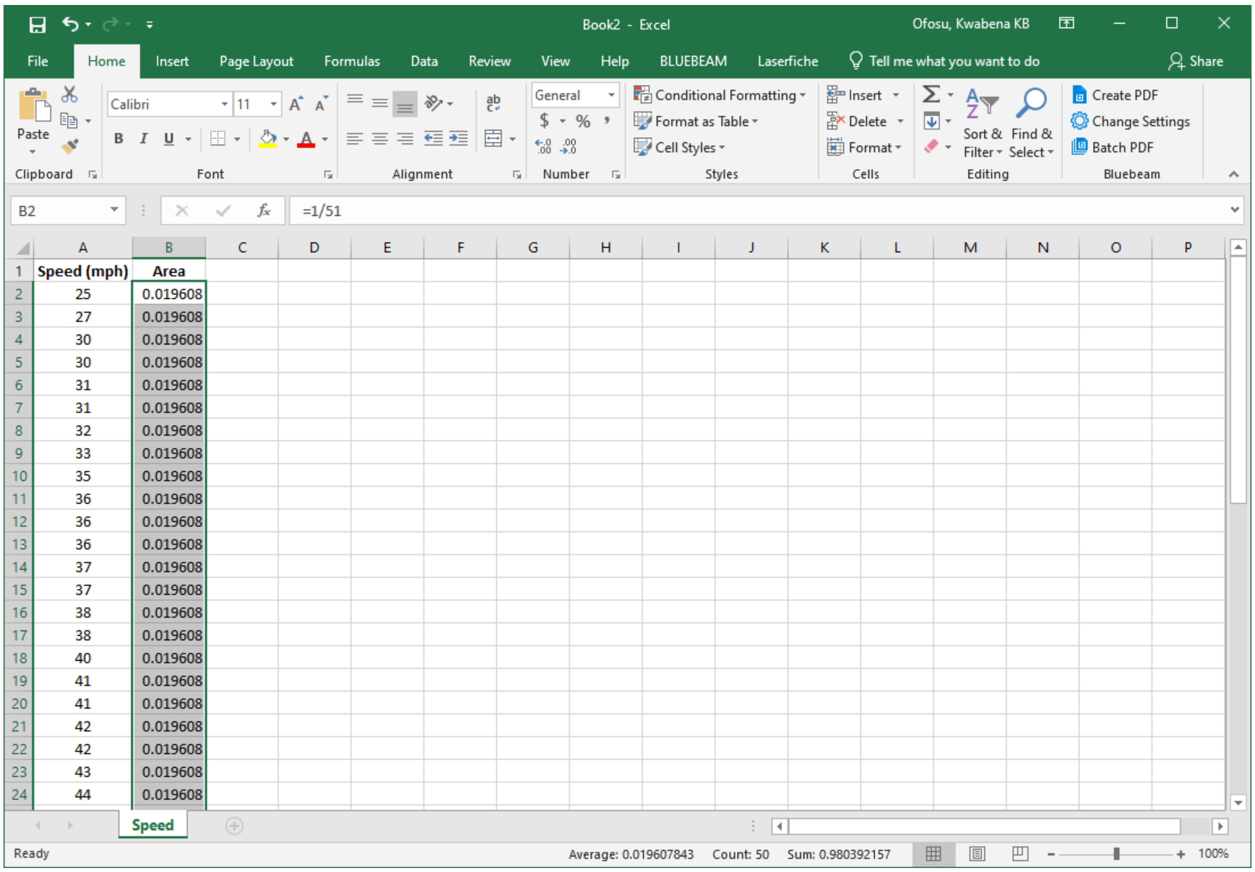


In tabular format on the spreadsheet.
Type in the formula as follows.



Hit the **Enter** key on the keyboard.

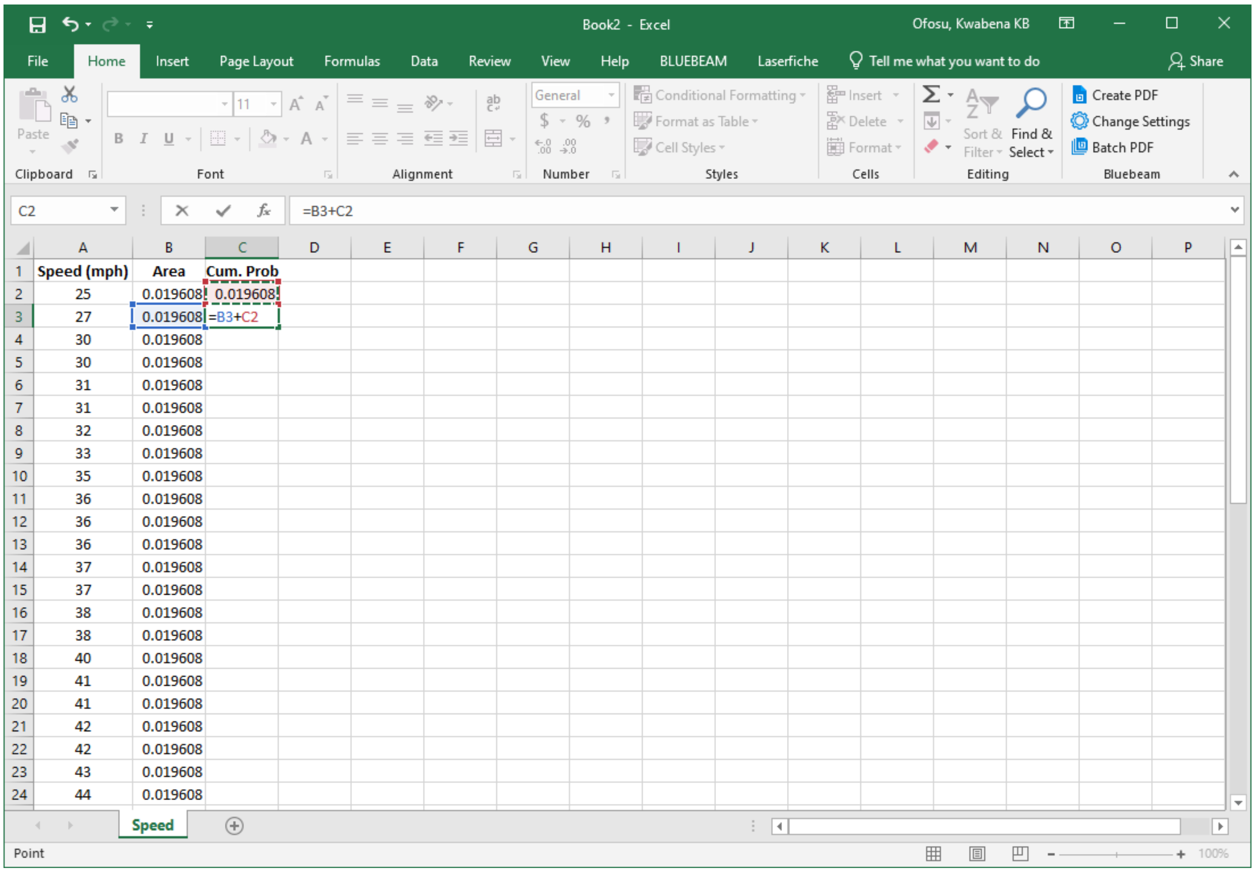
Fill handle down the entire table.



	A	B
1	Speed (mph)	Area
2	25	0.019608
3	27	0.019608
4	30	0.019608
5	30	0.019608
6	31	0.019608
7	31	0.019608
8	32	0.019608
9	33	0.019608
10	35	0.019608
11	36	0.019608
12	36	0.019608
13	36	0.019608
14	37	0.019608
15	37	0.019608
16	38	0.019608
17	38	0.019608
18	40	0.019608
19	41	0.019608
20	41	0.019608
21	42	0.019608
22	42	0.019608
23	43	0.019608
24	44	0.019608

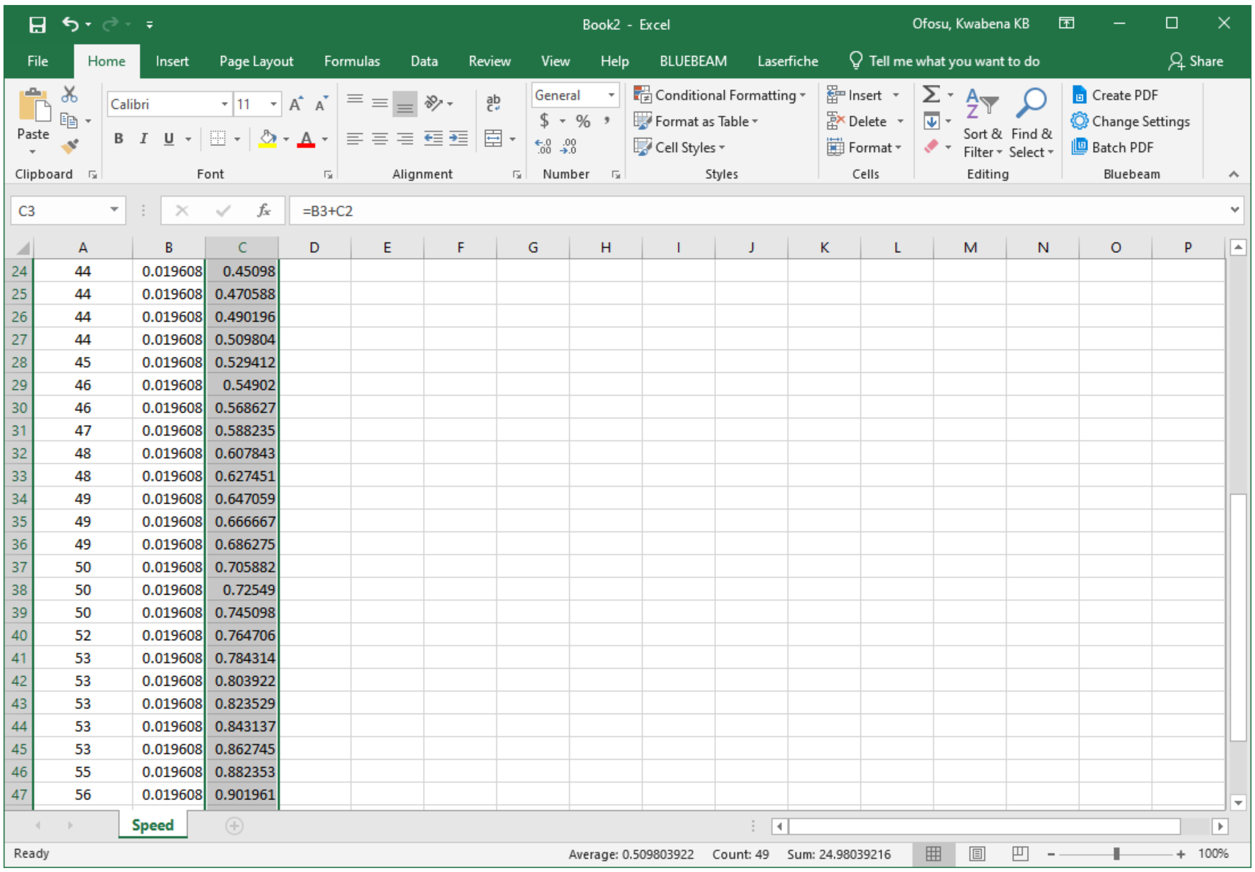
Compute the cumulative areas (cumulative probabilities) as follows.
The cumulative probability of an ordered value is the sum of the probabilities of the values up to that value. Therefore, the cumulative probability of the second value is the area for the second value plus the cumulative probability of the preceding value, the first value, and so on and so forth.

Replicate the following formula.



Hit the **Enter** key on the keyboard.

Fill handle down the entire table.



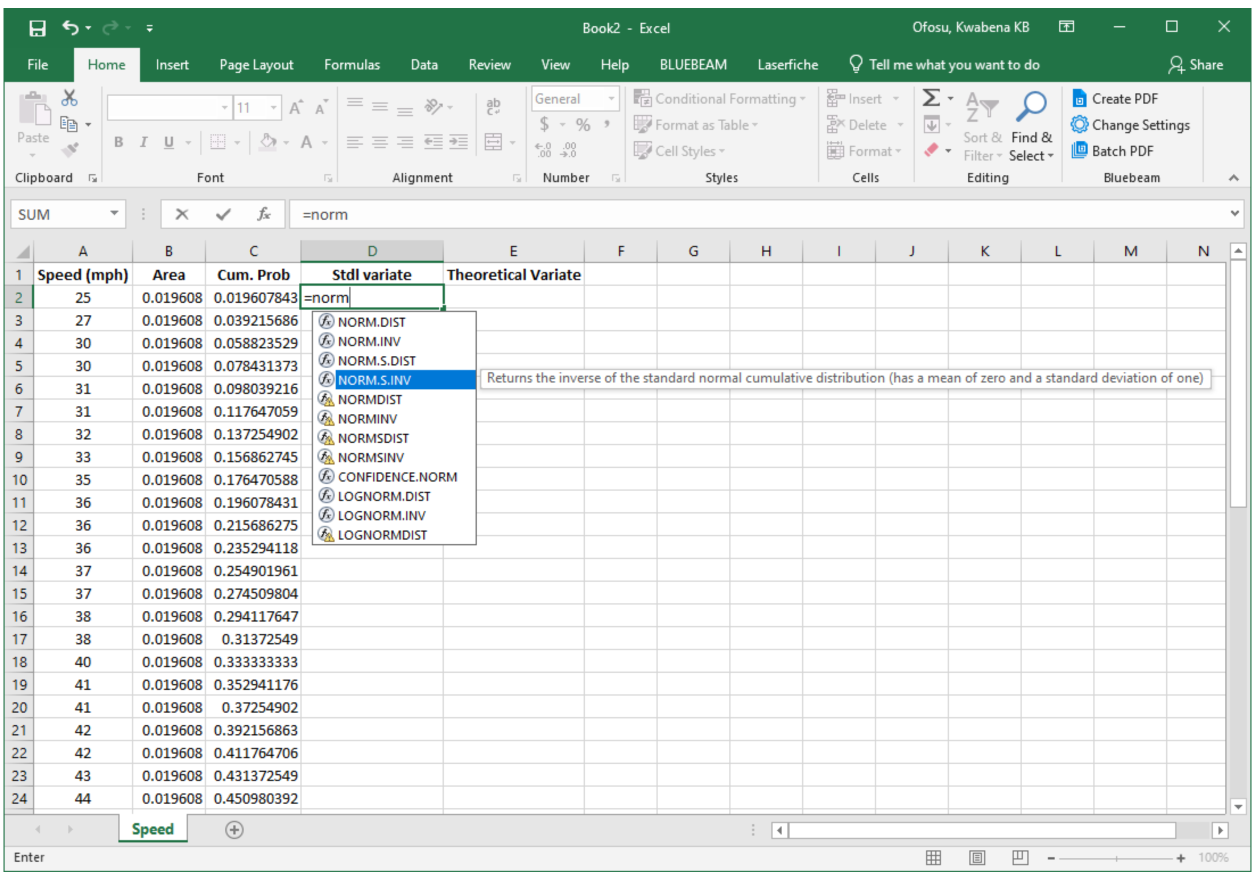
	A	B	C
24	44	0.019608	0.45098
25	44	0.019608	0.470588
26	44	0.019608	0.490196
27	44	0.019608	0.509804
28	45	0.019608	0.529412
29	46	0.019608	0.54902
30	46	0.019608	0.568627
31	47	0.019608	0.588235
32	48	0.019608	0.607843
33	48	0.019608	0.627451
34	49	0.019608	0.647059
35	49	0.019608	0.666667
36	49	0.019608	0.686275
37	50	0.019608	0.705882
38	50	0.019608	0.72549
39	50	0.019608	0.745098
40	52	0.019608	0.764706
41	53	0.019608	0.784314
42	53	0.019608	0.803922
43	53	0.019608	0.823529
44	53	0.019608	0.843137
45	53	0.019608	0.862745
46	55	0.019608	0.882353
47	56	0.019608	0.901961

These are the observed cumulative probabilities.

Step 3:

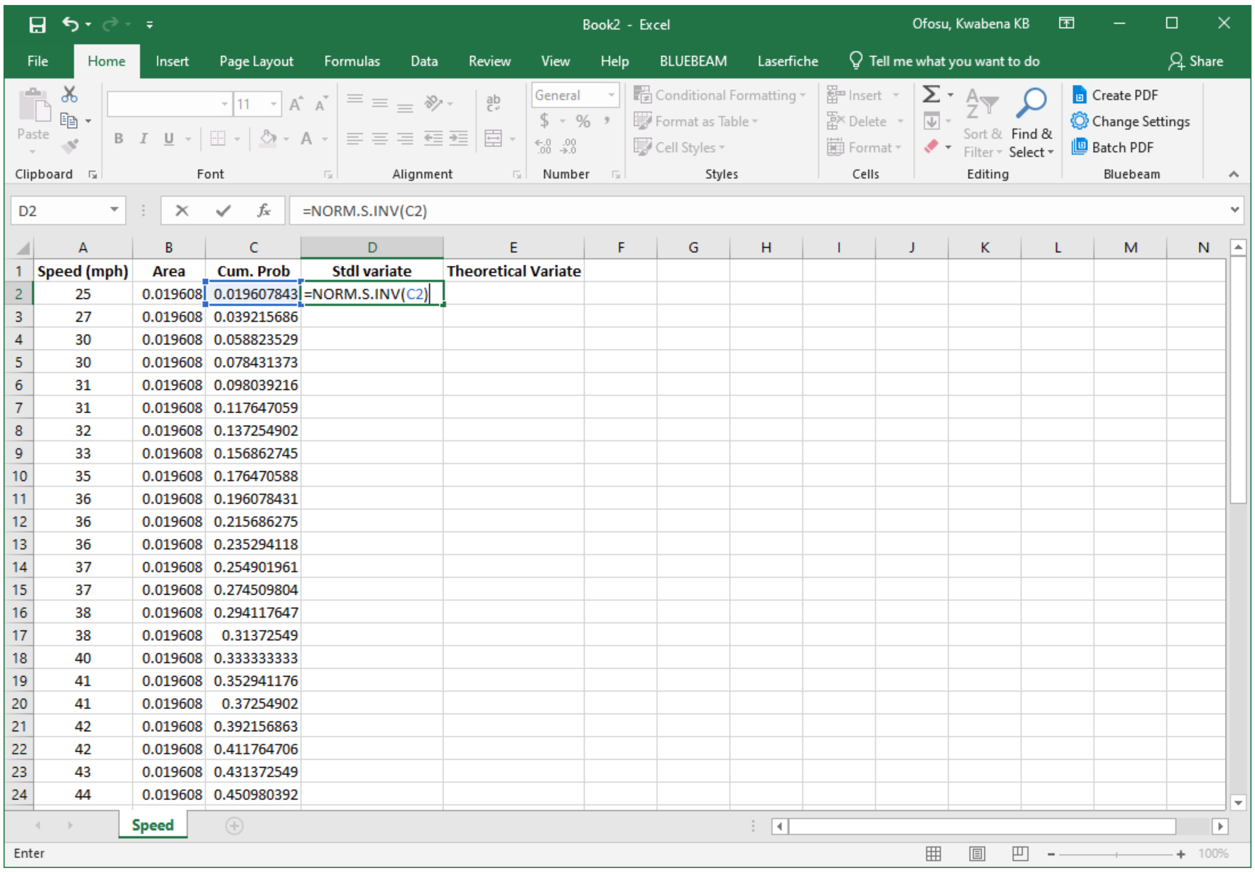
In this step the theoretical variate shall be computed from the quantile function of the candidate distribution. In this example the candidate distribution is the Normal distribution. For the special case of the Normal distribution, if this exercise were to be done by hand, one would compute the Standard Normal variate (commonly called the z-score), and then convert it to the Normal variate (x).

Compute the Standard Normal variate as follows.
Type in "NORM.S.INV" to look up and select the *Excel* Standard Normal distribution quantile function.



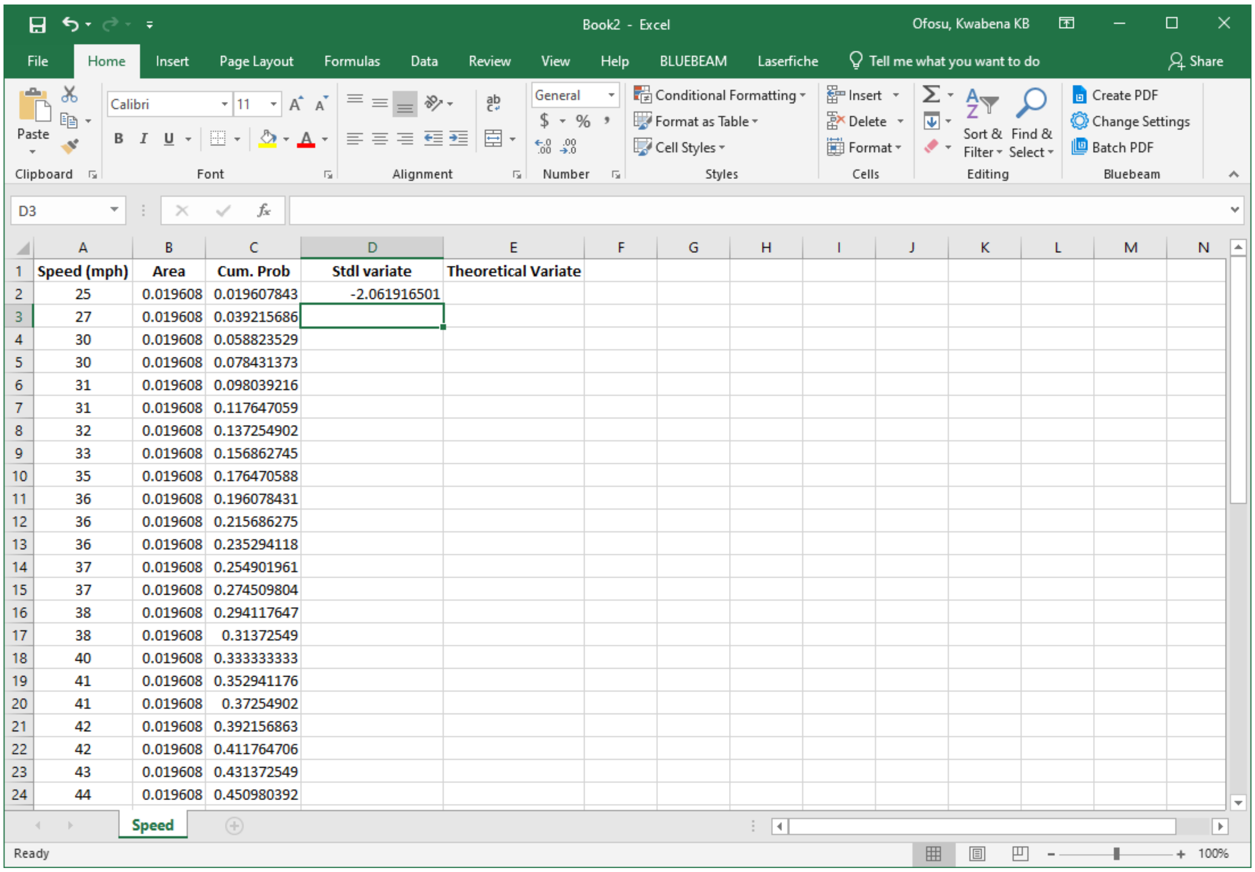
Double click on NORM.S.INV

Select the cumulative probability computed from the data.



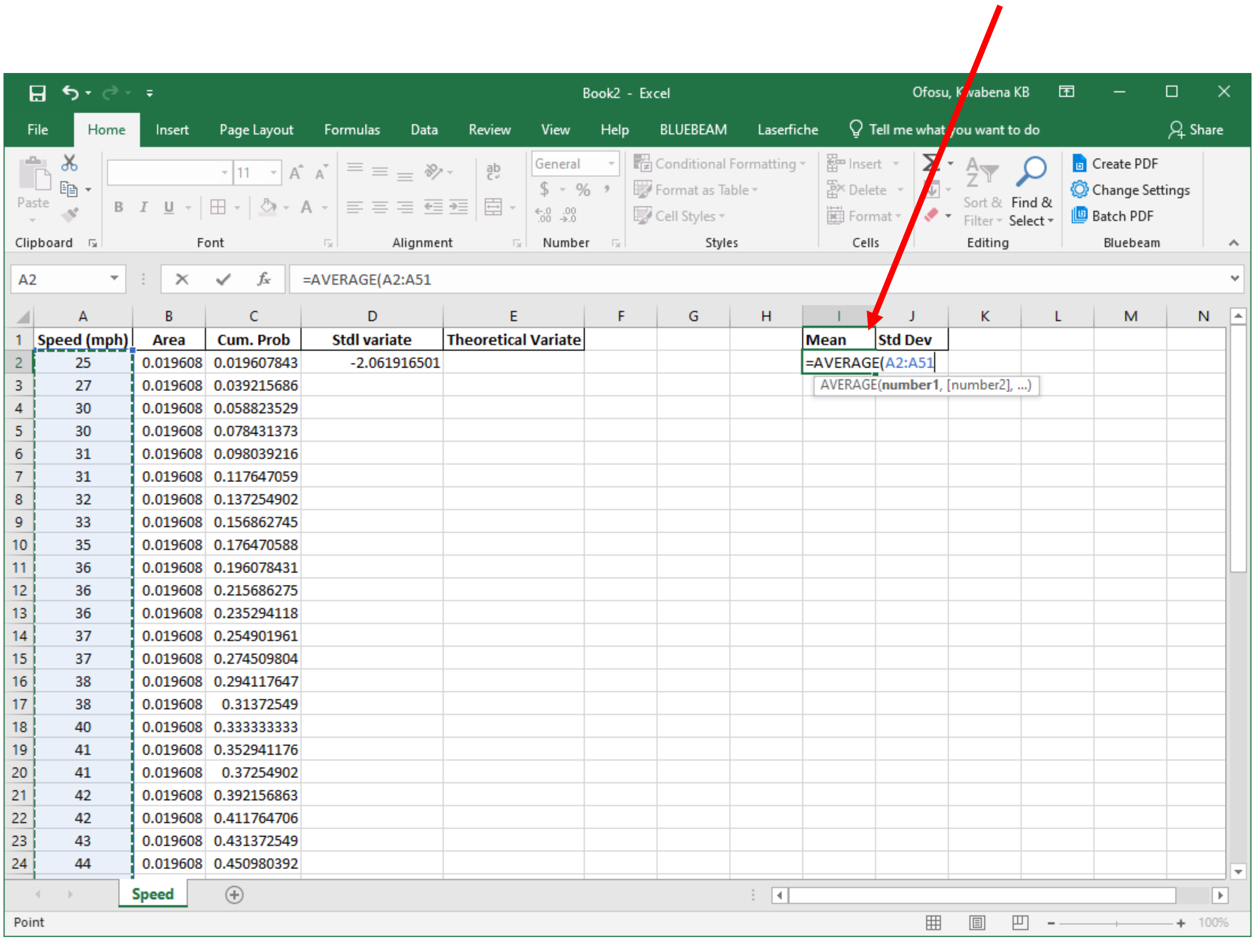
Hit the **Enter** key on the keyboard.

The Standard Normal variate (z-score) is computed.



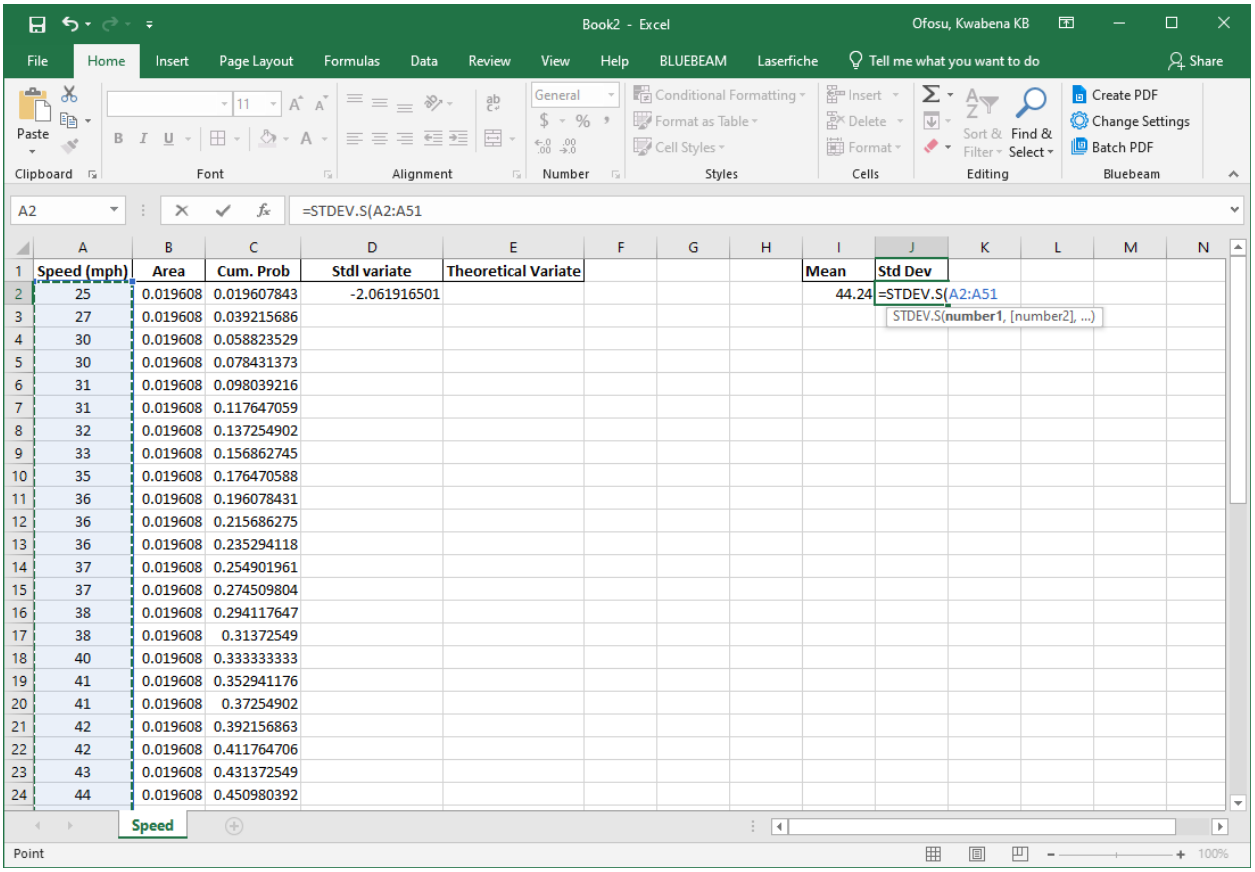
	Speed (mph)	Area	Cum. Prob	Stdl variate	Theoretical Variate
2	25	0.019608	0.019607843	-2.061916501	
3	27	0.019608	0.039215686		
4	30	0.019608	0.058823529		
5	30	0.019608	0.078431373		
6	31	0.019608	0.098039216		
7	31	0.019608	0.117647059		
8	32	0.019608	0.137254902		
9	33	0.019608	0.156862745		
10	35	0.019608	0.176470588		
11	36	0.019608	0.196078431		
12	36	0.019608	0.215686275		
13	36	0.019608	0.235294118		
14	37	0.019608	0.254901961		
15	37	0.019608	0.274509804		
16	38	0.019608	0.294117647		
17	38	0.019608	0.31372549		
18	40	0.019608	0.333333333		
19	41	0.019608	0.352941176		
20	41	0.019608	0.37254902		
21	42	0.019608	0.392156863		
22	42	0.019608	0.411764706		
23	43	0.019608	0.431372549		
24	44	0.019608	0.450980392		

Compute the mean (μ) of the data as follows.



Hit the **Enter** key on the keyboard.

Compute the standard deviation (σ) of the data as follows.



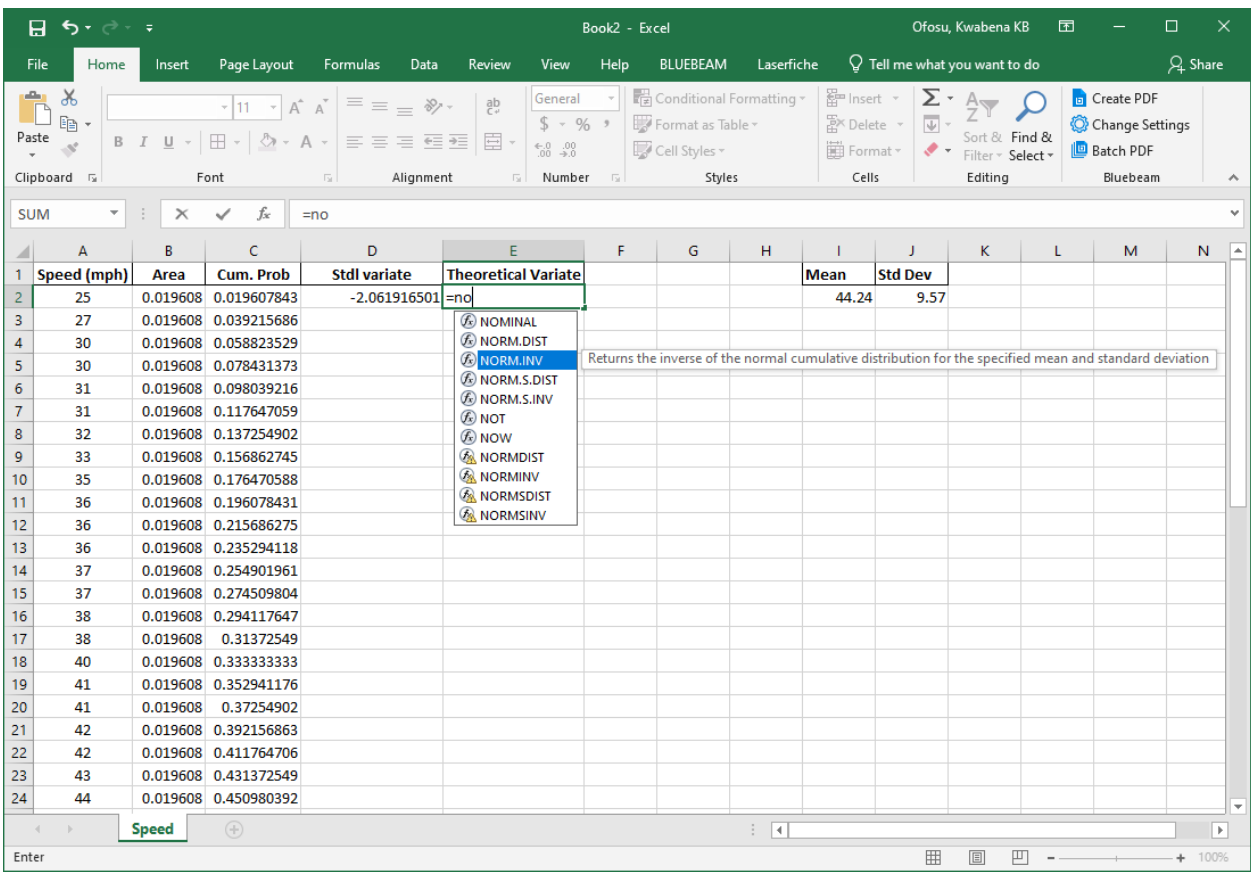
Hit the **Enter** key on the keyboard.

The theoretical variate (x) can be computed from the Standard Normal variate (z-score) by the relation

$$z = \frac{x - \mu}{\sigma}$$

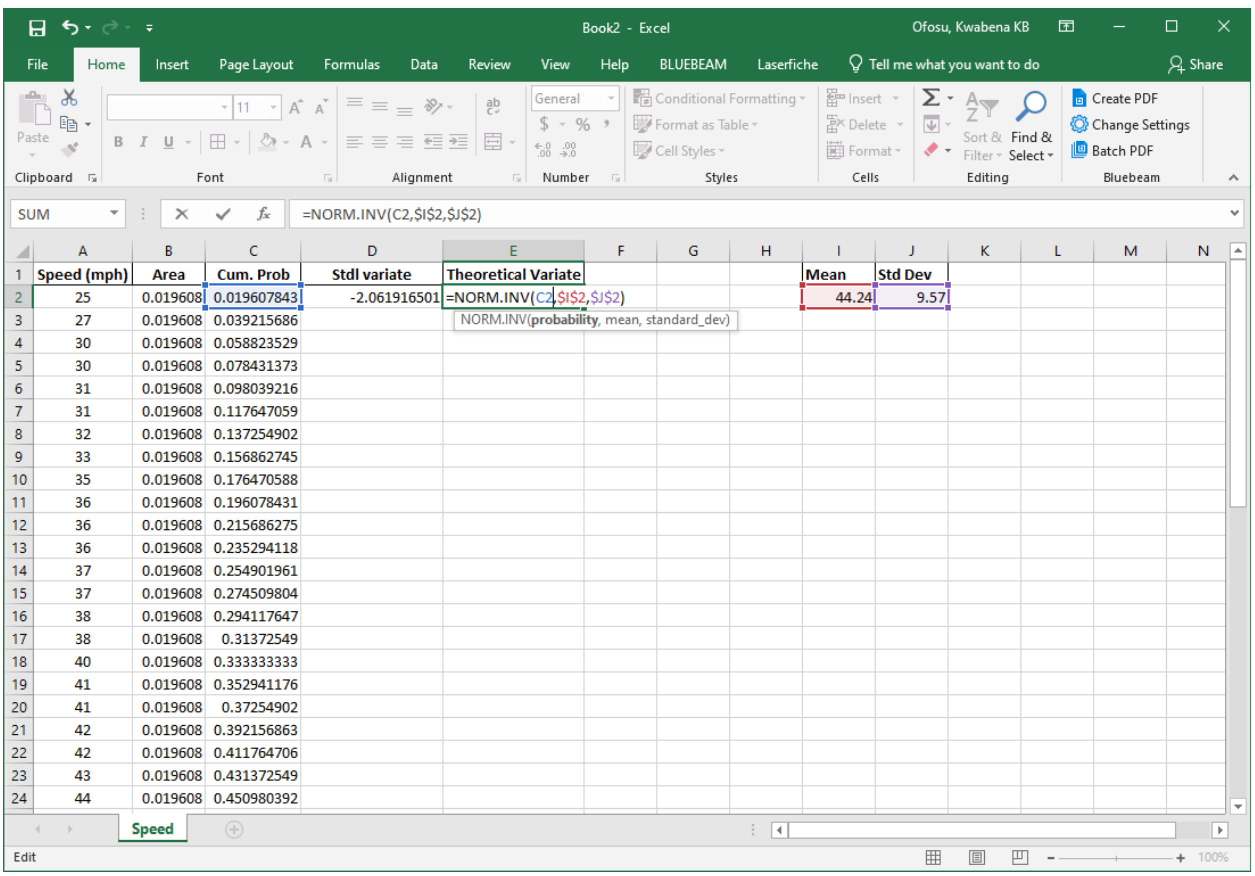
or by the *Excel* NORM.INV function.

Begin to type "NORM.INV" to pull up the *Excel* NORM.INV function



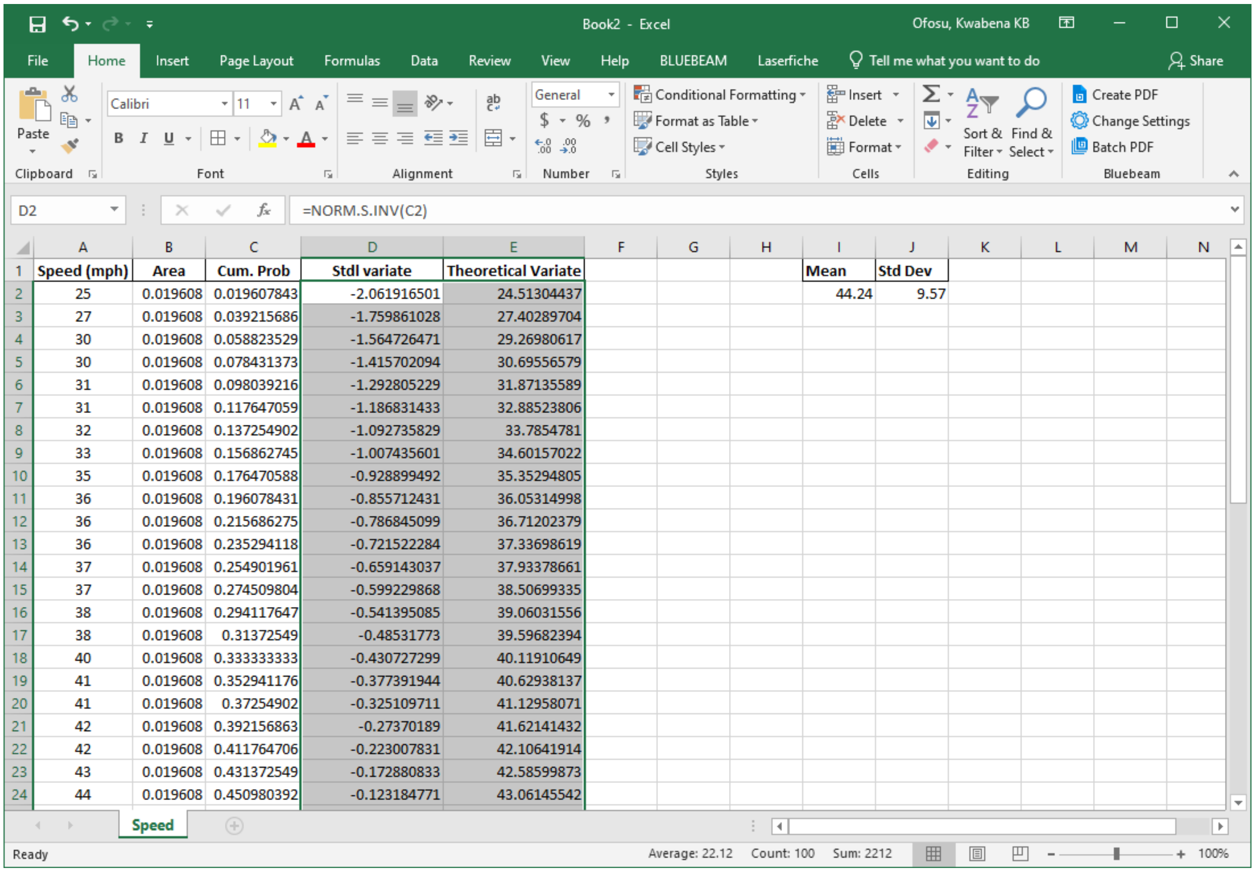
Double click on NORM.INV.

Select the cumulative **probability** value.
Type a comma (“,”).
Select the **mean** (μ) value.
Hit **F4** on the keyboard.
Type a comma (“,”).
Select the **standard_dev** (σ) value.
Hit **F4** on the keyboard.
(Note that the **F4** manipulation is needed to make those cell references absolute references which shall be needed to replicate the formula, by fill handle, correctly down the spreadsheet in the subsequent step).



Hit the **Enter** key on the keyboard.

Select the computed standard variate and the theoretical variate cells.
Fill handle down the entire table.



D2 =NORM.S.INV(C2)

	Speed (mph)	Area	Cum. Prob	Stdl variate	Theoretical Variate
2	25	0.019608	0.019607843	-2.061916501	24.51304437
3	27	0.019608	0.039215686	-1.759861028	27.40289704
4	30	0.019608	0.058823529	-1.564726471	29.26980617
5	30	0.019608	0.078431373	-1.415702094	30.69556579
6	31	0.019608	0.098039216	-1.292805229	31.87135589
7	31	0.019608	0.117647059	-1.186831433	32.88523806
8	32	0.019608	0.137254902	-1.092735829	33.7854781
9	33	0.019608	0.156862745	-1.007435601	34.60157022
10	35	0.019608	0.176470588	-0.928899492	35.35294805
11	36	0.019608	0.196078431	-0.855712431	36.05314998
12	36	0.019608	0.215686275	-0.786845099	36.71202379
13	36	0.019608	0.235294118	-0.721522284	37.33698619
14	37	0.019608	0.254901961	-0.659143037	37.93378661
15	37	0.019608	0.274509804	-0.599229868	38.50699335
16	38	0.019608	0.294117647	-0.541395085	39.06031556
17	38	0.019608	0.31372549	-0.48531773	39.59682394
18	40	0.019608	0.333333333	-0.430727299	40.11910649
19	41	0.019608	0.352941176	-0.377391944	40.62938137
20	41	0.019608	0.37254902	-0.325109711	41.12958071
21	42	0.019608	0.392156863	-0.27370189	41.62141432
22	42	0.019608	0.411764706	-0.223007831	42.10641914
23	43	0.019608	0.431372549	-0.172880833	42.58599873
24	44	0.019608	0.450980392	-0.123184771	43.06145542

Mean	Std Dev
44.24	9.57

Step 4:

Plot the ordered observed data, in this case the ordered speed data, versus the theoretical variate. This is the probability plot.

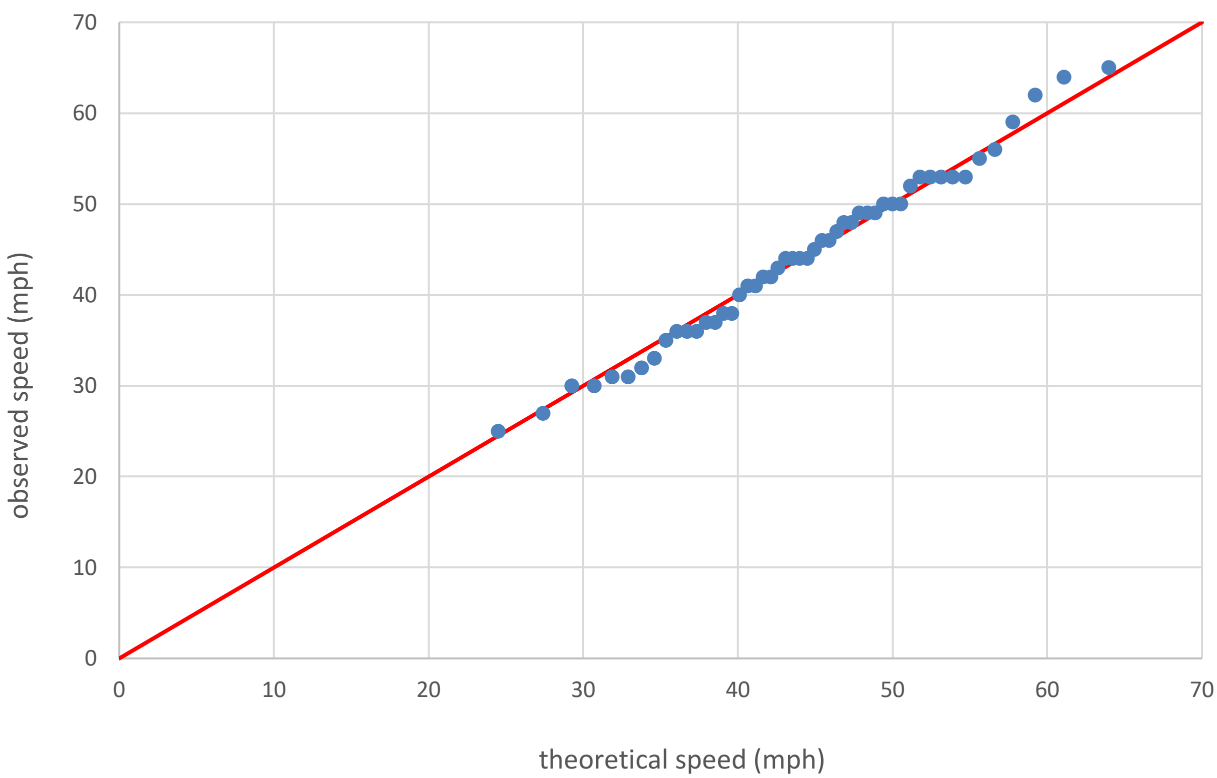


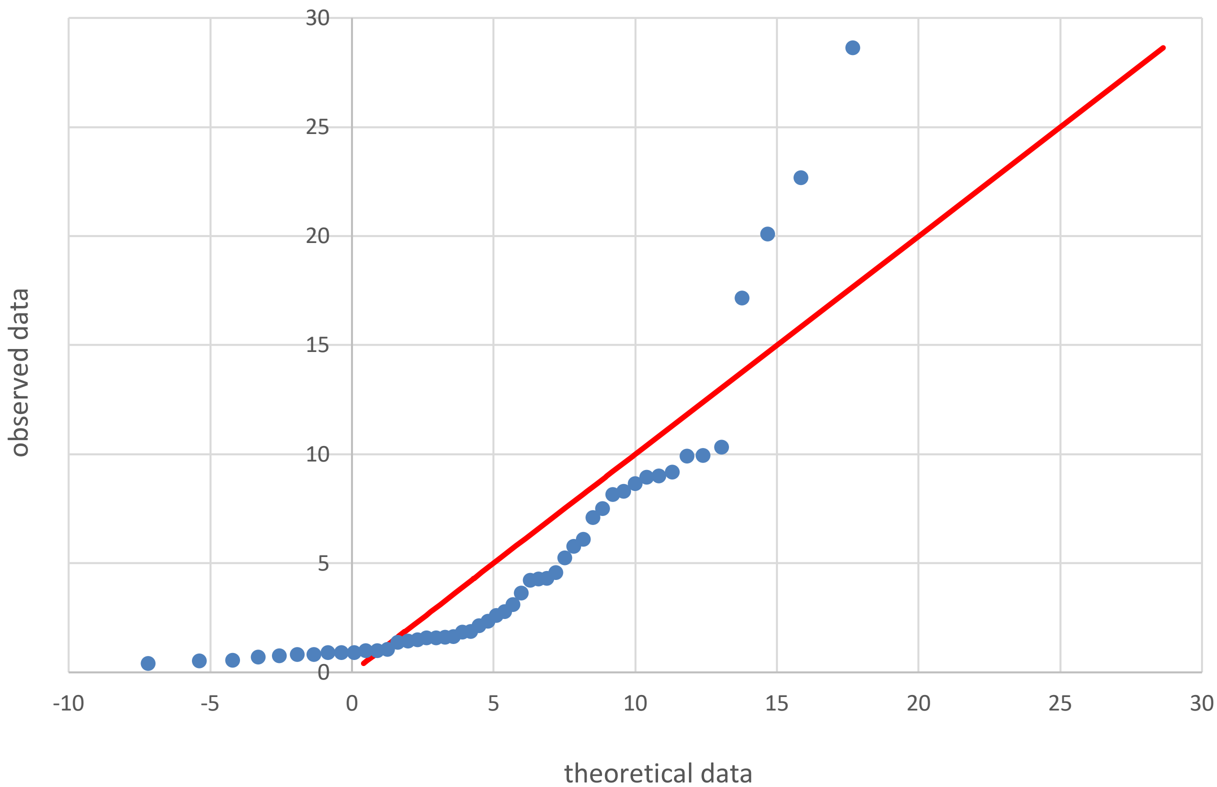
Figure 5. 1: Normal probability plot

The plotted points appear to follow a straight line through the origin at an angle of 45 degrees to the horizontal. It can therefore be concluded that the data come from the candidate distribution, in this case the Normal distribution.

The above plot is a **Normal probability plot**. The probability plot for any other distribution will be constructed in a similar manner, the only difference being using the quantile function of the candidate distribution to compute the theoretical variate values.

This concludes the probability plot example.

The following Normal probability plot illustrates an example where the plotted points do not appear to line up with a straight line through the origin at an angle of 45 degrees to the horizontal.

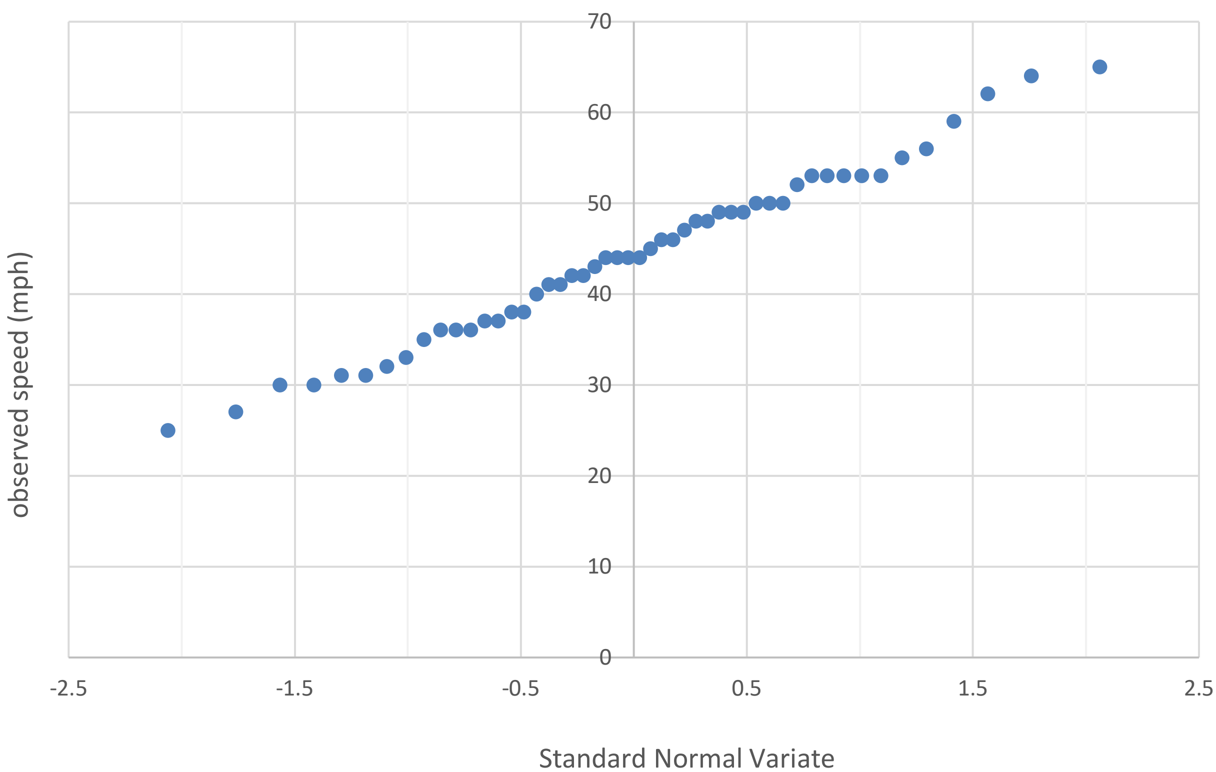


Thus, the conclusion is that the data do not come from a Normal distribution. An analyst will now have to repeat the probability plot process as many times as needed, testing other known theoretical probability distributions, until a suitable distribution is found.

It is pertinent to note that the suitability or otherwise of one theoretical distribution does not imply or infer that some other theoretical distribution(s) may be suitable or otherwise. In other words, there can be multiple suitable theoretical distributions for a data set. In such cases, it is up to the analyst to conduct the relevant testing and ultimately select the appropriate theoretical distribution based on a combination of the results, experience and expert knowledge of the system under investigation.

<u>Other Probability Plotting Techniques</u>

There are many variations of the probability plot procedure demonstrated in this chapter. One popular variation found in the literature, particularly for the Normal probability plot, is to plot the Standard Normal variate (*z*-score) on the horizontal axis. For the speed data example previously demonstrated in this chapter, the probability plot will be as follows.



The straight-line behavior of the plotted points is apparent, confirming that the data follows a Normal distribution. If a trendline is fitted to the points, it can be seen that the intercept of the trendline on the vertical axis will be equal to the mean of the data, and the slope of the trendline will be equal to the standard deviation of the data.

Another popular variation of the probability plot involves plotting the rank (j) of the ordered observed data versus the Standard Normal variate (z-score). The ranking is such that the lowest observed value in the data has a rank of 1, the next value has a rank of 2 and so on, and the highest observed value has a rank of n, where n is the number of observations in the dataset. Each Standard Normal variate is computed from a cumulative probability (p) that is approximated by the formula,

$$p = \frac{j - 0.5}{n}$$

Probability plots may also be constructed from probabilities, not cumulative probabilities. Such probability plots are often referred to as P-P plots.

5.3 Other Methods

Other methods of assessing the validity of a probability distribution for modeling a random phenomenon fall under the **Goodness-of-Fit Tests**. Goodness-of-fit tests are statistical inference tests (or **Hypothesis tests**) that result in a conclusion of "**reject**", or "**fail to reject**" the given distribution based on the data. The Goodness-of-fit tests are generally applied to verify the validity of a given theoretical distribution after the visual inspections of histograms and probability plots suggest the given distribution is suitable. The Goodness-of-fit tests available include the **Chi-Square** test, the **Kolmogorov-Smirnov** (or **K-S**) test, and the **Anderson-Darling** (or **A-D**) test.

The Chi-Square test (pronounced "kai" or "chai" or "shy" or "chee" or "she") involves comparing observed frequencies of the data with corresponding theoretical frequencies computed from the candidate theoretical distribution model. As with any hypothesis test, a test statistic is computed and a comparison is made with a critical value at a given level of confidence for the test procedure, based on which, the candidate theoretical distribution will be rejected or will fail to be rejected.

In the K-S test, the observed cumulative frequencies of the data are compared to the cumulative distribution function of the given theoretical distribution. If the discrepancy between the observed and theoretical frequencies is greater than normally expected for a given sample size, then the given theoretical distribution is rejected. If the discrepancy is less than a critical value at a given level of confidence for the test procedure, then the conclusion is to fail-to-reject the theoretical distribution.

The A-D test compares the observed cumulative frequencies of the data to the cumulative distribution function of the given theoretical distribution, but captures discrepancies between the two that are intrinsically weighted towards the tails of the distributions, unlike the other Goodness-of-fit tests which are unable to make this discrimination.

A detailed presentation of the Goodness-of-fit tests is beyond the scope of this course. Readers are strongly encouraged to review the extensive literature available on these topics on their own.

6. CONCLUSION

This course presented an overview of the Monte Carlo simulation technique, and how it can be implemented in *Excel*.

This course began with an illustrative introductory example of a simulation problem encountered by a practicing engineer. This was followed by a presentation of the fundamentals of the overall structure of the Monte Carlo simulation method. A demonstration of the implementation of the elements of a Monte Carlo simulation model on an *Excel* spreadsheet was presented. This course covered how the results of a Monte Carlo simulation can be analyzed, interpreted and applied in decision making. The final part of this course presented techniques used to select and validate statistical distributions that are used to model the constituent components or activities of the system or process that is being simulated, and how these techniques can be conducted in *Excel*.

Upon completion of this course, participants have gained skills in statistical distributions and Monte Carlo simulation and will be able to apply these skills in the simulation and modeling of real engineering systems. This course has enabled participants to identify professional situations where the innovative application of techniques learned in this course are relevant and will be of benefit to their productivity, efficiency, and the quality of their work product. Practitioners are strongly encouraged to look for situations in their domains of expertise where simulation and systems modeling are applicable and will be of benefit to their work product and to their organization.

A successful application of engineering methods in *Excel* requires a careful and meticulous approach, and can only be mastered and retained by practice and repetition. It has been my utmost pleasure presenting this topic to you. Thank you.

REFERENCES

Johnson, R. A. (2011). *Miller and Freund's Probability and Statistics for Engineers* (Eighth ed.).

May, A. D. (1990). *Traffic Flow Fundamentals.*

Meeker, W. Q., & Escobar, L. A. (1998). *Statistical Methods for Reliability Data.* John Wiley & Sons.

The National Instititute of Standards and Testing. (2013). NIST/SEMATECH e-Handbook of Statistical Methods. Retrieved November 2019, from http://www.itl.nist.gov/div898/handbook/

Images were all drawn/ prepared by Kwabena. Ofosu

