Calculating and Understanding Sampling Distributions in Excel

Authored by

Mohammed looti

November 6, 2025

Understanding how to calculate and analyze a sampling distribution is arguably one of the most fundamental concepts in modern **statistical inference**. A sampling distribution does not describe the population itself, but rather represents the probability distribution of a particular statistic--such as the mean--derived from numerous random samples taken from a single underlying population. By simulating this distribution, we gain crucial insight into the stability and variability of sample statistics.

This comprehensive tutorial offers a structured, step-by-step guide on how to simulate, analyze, and visualize sampling distributions using **Microsoft Excel**. By leveraging built-in statistical functions, we can effectively demonstrate complex theoretical principles, such as the Central Limit Theorem, through practical data simulation. This process transforms abstract concepts into tangible results that are easy to manipulate and understand.

We will walk through the entire simulation lifecycle, ensuring that you can replicate these results for any population parameters or sample sizes. Our workflow is designed to cover the following essential stages of analysis, moving from data generation to probabilistic conclusions:

Generating the raw data required for a robust sampling distribution simulation.
Calculating the empirical statistics (mean and standard deviation) for each sample.
Visually inspecting the resulting distribution using appropriate charts, such as a histogram.
Calculating and comparing the empirical parameters against theoretical expectations predicted by statistical theorems.
Determining specific probabilities related to the calculated sample means, which is a core function of inferential statistics.

Generating Simulated Data for the Sampling Distribution

To effectively demonstrate the creation of a sampling distribution, we must first simulate a large dataset representing repeated sampling. For this demonstration, we will generate **1,000 independent samples**, where the size of each sample (n) is set to **20**. We will assume the underlying population follows a normal distribution with a specified population mean (μ) of **5.3** and a population standard deviation (σ) of **9**.

We can efficiently generate this substantial volume of data in Excel by pairing the inverse cumulative distribution function with a random number generator. The function NORM.INV is ideal for this purpose, as it takes a uniform random probability (generated by the RAND() function) and converts it into a value drawn from the specified normal distribution. To begin, enter the following formula into cell A2 of your worksheet:

=NORM.INV(RAND(), 5.3, 9)

Once the formula is entered, the next step is to replicate it across the entire required range. Click on cell A2 and use the fill handle (the small plus sign in the bottom right corner) to drag this formula horizontally across 20 columns (extending to column T) and vertically down 1,000 rows (down to row 1001). This action instantly creates the simulated dataset, which should span the range A2:T1001, providing a total of 20,000 data points.

It is vital to recognize the structure of this generated data: each complete row in the range A2:T1001 represents one independent sample of size 20. Every single value within these cells is drawn from the specified normal population (Mean = 5.3, Standard Deviation = 9). The image below illustrates the result of this data generation process:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Sample 6	Sample 7	Sample 8	Sample 9	Sample 10	Sample 11	Sample 12	Sample 13	Sample 14	Sample 15	Sample 16	Sample 17	Sample 18	Sample 19	Sample 20
2	-14.3478	9.427724	6.41191	12.41153	-8.24302	6.381618	11.5354	14.50201	5.103956	11.26152	5.878546	26.79076	5.257913	22.77662	-6.071	-1.76196	18.22686	11.31325	13.93815	0.479712
3	27.16546	25.44272	0.882208	4.039182	4.181897	9.708091	11.48947	11.12963	11.78162	8.597135	7.193303	18.64945	4.876111	13.41833	19.22029	9.507522	9.241467	-3.73597	6.762841	19.90909
4	13.79445	-5.98839	0.809056	12.59813	4.707145	8.149118	-3.08282	9.212429	14.83148	-0.16035	17.50281	9.990323	10.28493	2.317118	16.60559	-2.24892	-5.34456	3.556385	1.854363	3.002711
5	-4.78071	-3.29673	12.86278	1.434419	12.67984	1.387565	-0.6056	6.770139	3.290494	12.07852	-0.36996	-13.3463	3.944265	22.6152	3.833341	-1.15977	0.814789	8.223065	13.30782	-9.23518
6	17.02983	3.375123	4.306389	0.911502	1.181418	10.45451	-10.6861	3.255356	-10.0396	-1.41185	1.717989	4.332279	8.355191	-1.47665	4.267777	1.589566	8.90931	2.660249	-6.3752	3.829714
7	-3.23865	26.52023	-1.5878	-6.26995	8.71232	-3.73426	4.818732	-5.32073	1.121144	12.11585	-3.69324	0.018888	-0.86529	8.679084	6.37755	-0.44015	6.191809	7.541581	-8.42945	5.261255
8	-2.87737	-9.55328	-8.2743	1.022951	17.4909	-3.42453	4.051349	-3.83616	6.251832	13.11334	3.644683	7.531315	8.776285	14.13575	7.215605	-9.33324	4.986041	14.52475	9.902011	15.87092
9	15.63742	-4.39394	6.285594	10.41558	10.52037	17.08537	-12.8884	-4.34923	-2.62318	16.49861	1.329042	13.54197	0.361993	10.35637	17.32712	7.771755	14.26686	8.609849	6.172321	7.860733
10	3.26881	7.722144	2.722006	1.755557	0.299	0.64552	11.40958	13.69103	-12.7907	1.22483	16.58394	4.196994	3.651579	-3.36173	15.33022	15.12278	13.10987	-5.74766	-11.3612	-1.22457
11	14.08954	-7.4207	18.07621	7.623785	-0.67097	18.88473	-9.94538	-1.88383	4.788754	16.42689	8.532977	12.70508	-11.3424	-4.72177	4.802661	0.222133	-1.63861	5.718594	-3.5927	2.335083
12	-5.63741	-7.02729	5.518323	19.37283	7.551738	5.985714	3.159656	3.889748	7.563939	27.52694	-1.89179	1.751063	2.013756	-0.42772	28.72887	10.74831	-3.17044	10.53117	7.14759	11.57939
13	15.79509	2.727566	1.487302	2.770588	15.1024	5.978842	7.69092	7.08799	4.570933	2.98918	0.595897	-11.8499	3.196037	8.761552	0.440966	11.51041	2.987297	15.49738	-1.31741	20.23163
14	5.367749	10.46109	9.583662	0.647727	8.837532	4.590372	12.11238	-0.29913	28.07044	3.09704	-4.04612	-0.39319	7.842675	3.864414	-9.83307	11.39651	1.114558	7.08748	-4.3725	-3.0126
15	8.907808	23.75535	11.17992	8.050923	6.318122	0.480674	1.686195	-0.3514	-1.69552	3.920678	-2.39118	3.419415	21.62053	1.701101	0.409567	12.63149	11.59317	-9.11533	12.85583	23.30873
16	-3.53068	-0.11266	14.20008	4.819141	9.376628	-12.1205	19.49012	3.791216	3.908823	2.319907	15.86166	8.473103	18.36289	1.892983	9.143392	0.356941	10.99721	4.457816	8.78991	1.907188
17	4.47008	17.64061	15.2252	8.729333	-12.836	-3.36272	4.663955	15.15908	-8.82142	-4.81252	16.79467	3.714213	10.31241	13.844	18.1205	-0.44859	1.707714	10.75024	13.46804	10.99492
18	-4.48259	16.44781	-4.13243	18.75596	17.05798	9.045342	0.11317	4.298675	-1.72524	4.139834	-14.3815	6.208324	16.99288	-0.50087	6.08703	1.692208	3.47329	-5.88791	8.492441	14.32321
19	28.27343	-0.11177	-0.77956	-5.10458	10.993	1.573969	-1.76437	10.06757	9.946887	6.244451	-8.23187	1.495842	21.15129	17.02533	-11.9013	-1.20163	20.53252	0.706525	19.89147	-2.46186
20	-2.57817	-1.75948	6.682566	2.331538	0.061475	3.255007	0.337898	-7.52274	2.497519	0.416612	10.04643	10.48019	23.1283	-5.11329	-1.07627	3.837494	17.58466	2.474282	-3.21888	9.963007
21	-5.64917	15.54493	0.113937	9.955225	-9.55085	-6.794	9.143788	12.43555	-3.2781	-1.5712	5.71225	0.554566	-0.04992	1.309828	-5.22846	19.35693	-1.83687	-2.26399	21.88116	19.49978
22	8.762842	-2.0507	3.104038	11.36072	1.442022	6.946865	11.49612	8.949299	14.2931	-6.16227	11.60043	1.684628	7.028243	-2.62814	4.313823	-2.43979	3.573794	7.44069	-5.16497	6.494702
23	-0.54716	-1.76507	6.236773	-2.53364	5.74823	1.340863	-4.25959	20.7382	16.13412	1.34754	8.211887	2.30144	19.84482	13.65197	-2.24629	-11.5175	16.40694	-3.43766	2.801031	6.446707
24	11.9622	-4.53673	8.319412	-3.85999	2.904457	20.70921	14.69749	11.56554	4.846614	2.225501	14.55177	12.13013	10.57308	22.07002	11.88017	17.96653	-5.34646	-8.11144	10.80995	11.36282

Calculating the Sample Means for the Distribution

The core objective of simulating a sampling distribution of the mean is to collect and analyze the means of the individual samples. Therefore, the logical next step after data generation is to calculate the sample mean for each of the 1,000 samples we just created. This entire collection of 1,000 means will constitute the empirical sampling distribution we wish to analyze.

To calculate the sample mean for the first sample (located in Row 2, spanning A2 through T2), we use Excel's built-in AVERAGE function. Input the following formula into cell U2 of your spreadsheet. This column (Column U) will serve as the storage location for our resulting sampling distribution data:

=AVERAGE(A2:T2)

Once the formula is correctly entered in U2, use the fill handle again. Instead of dragging, you can simply double-click the handle at the bottom right corner of cell U2. Since the adjacent data (A2:T1001) is already filled, Excel will automatically copy this formula down to all 1,000 rows, populating the range U2 through U1001. Column U now holds the mean value derived from every single sample generated in the previous step.

A quick inspection of the results in Column U highlights the inherent variability that exists in random sampling. For instance, you might observe that the first sample yielded a mean of 7.56, while the second sample might result in a mean of 10.97. This fluctuation is normal and illustrates how individual sample means vary around the true population mean, leading us toward the eventual analysis of the distribution's shape and spread.

P	Q	R	S	T	U	V
Sample 16	Sample 17	Sample 18	Sample 19	Sample 20	Mean	
-1.76196	18.22686	11.31325	13.93815	0.479712	7.563684	
9.507522	9.241467	-3.73597	6.762841	19.90909	10.97299	
-2.24892	-5.34456	3.556385	1.854363	3.002711	5.61955	
-1.15977	0.814789	8.223065	13.30782	-9.23518	3.5224	
1.589566	8.90931	2.660249	-6.3752	3.829714	2.309342	
-0.44015	6.191809	7.541581	-8.42945	5.261255	2.688946	
-9.33324	4.986041	14.52475	9.902011	15.87092	4.560943	
7.771755	14.26686	8.609849	6.172321	7.860733	6.989309	
15.12278	13.10987	-5.74766	-11.3612	-1.22457	3.812402	
0.222133	-1.63861	5.718594	-3.5927	2.335083	3.649505	
10.74831	-3.17044	10.53117	7.14759	11.57939	6.745719	
11.51041	2.987297	15.49738	-1.31741	20.23163	5.812735	
11.39651	1.114558	7.08748	-4.3725	-3.0126	4.605851	
12.63149	11.59317	-9.11533	12.85583	23.30873	6.914303	
0.356941	10.99721	4.457816	8.78991	1.907188	6.119259	
-0.44859	1.707714	10.75024	13.46804	10.99492	6.765686	
1.692208	3.47329	-5.88791	8.492441	14.32321	4.80088	
-1.20163	20.53252	0.706525	19.89147	-2.46186	5.817264	
3.837494	17.58466	2.474282	-3.21888	9.963007	3.591407	
19.35693	-1.83687	-2.26399	21.88116	19.49978	3.964269	
-2.43979	3.573794	7.44069	-5.16497	6.494702	4.502273	
-11.5175	16.40694	-3.43766	2.801031	6.446707	4.745179	
17.96653	-5.34646	-8.11144	10.80995	11.36282	8.336015	

Comparing Theoretical vs. Actual Sampling Distribution Parameters

With the 1,000 sample means compiled in Column U, we can now calculate the descriptive statistics for this simulated sampling distribution. These empirical statistics--the overall mean and standard deviation of the sample means--are critical for comparison against established statistical theory. This comparison serves as a powerful validation of the simulation process.

To compute the final descriptive statistics for the entire dataset in Column U, you will use simple Excel functions like AVERAGE and STDEV.S (or STDEV.P, depending on whether you treat the 1,000 samples as a population or a sample, though STDEV.S is generally safer). Use the following functions to summarize the data in column U:

V	W	X	Y	Z
	Mean	5.367869	=AVERAGE(U2:U1001)	
	SD	2.075396	=STDEV(U2:U1001)	

According to the highly important Central Limit Theorem (CLT), the expected mean of the sampling distribution must closely approximate the population mean (μ). Since our simulation used a population mean of 5.3, our calculated sampling mean of **5.367869** demonstrates a remarkably strong alignment with this theoretical expectation, providing evidence of a successful simulation.

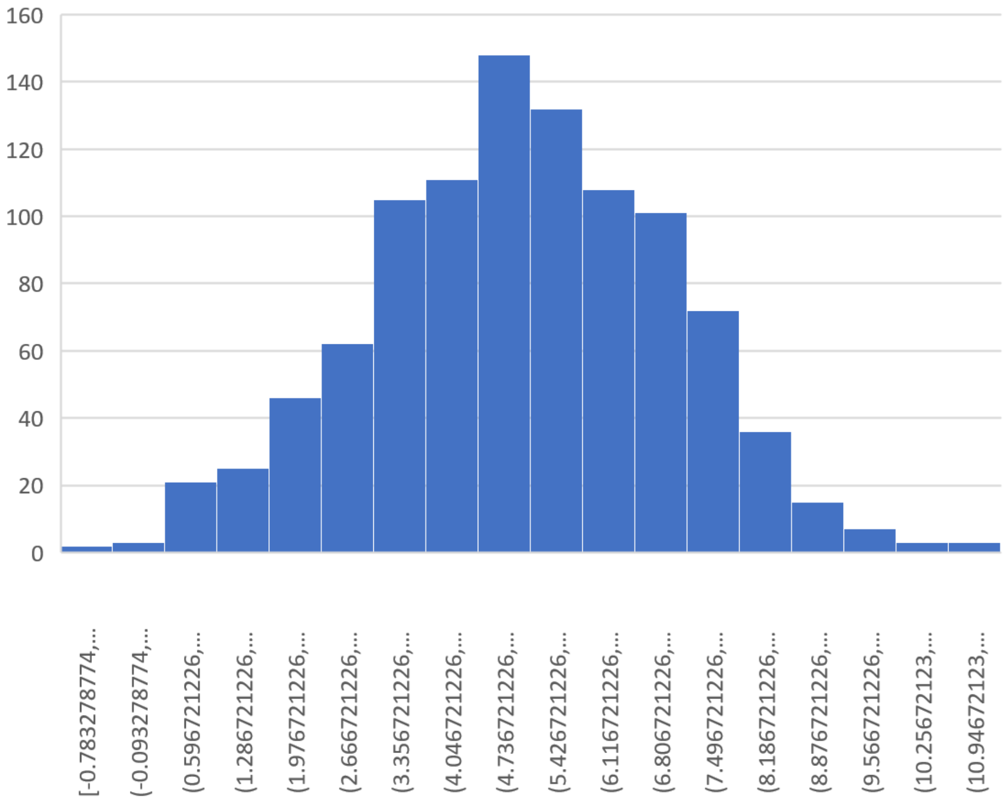
Similarly, the CLT dictates the expected spread of the sampling distribution, known as the standard error of the mean. This theoretical value is calculated by dividing the population standard deviation (σ) by the square root of the sample size (√n). In our scenario, the theoretical standard error is 9 / √20, which approximates 2.012. Our simulated standard deviation--the empirical standard error--was calculated as **2.075396**. The closeness of this value to 2.012 confirms that our simulation accurately captures the theoretical spread of sample means.

Visualizing the Sampling Distribution with a Histogram

While numerical analysis confirms the theoretical predictions, visualization provides immediate clarity regarding the shape and characteristics of the generated sampling distribution. Creating a histogram is the most effective way to display the frequency distribution of the 1,000 sample means.

To generate the chart, first select the data for the sample means located in Column U (the range U2:U1001). Next, navigate to the **Insert** tab on the Excel ribbon. Within the **Charts** section, locate and click on the **Statistics Chart** icon, then choose the **Histogram** option. Excel will automatically bin the data and generate the visualization based on the frequencies of the sample means.

The resulting histogram below provides a clear, visual confirmation of the simulation's validity:



Crucially, as predicted by the Central Limit Theorem (Link 2/5), the sampling distribution of the means exhibits a classic, symmetrical bell-shape. It is clearly centered closely around the population mean of 5.3. The histogram visually confirms that the vast majority of sample means cluster tightly near the center, demonstrating that most samples are representative of the underlying population.

However, by analyzing the tails of the distribution, we can also observe the natural variation inherent in the sampling process. The far ends of the distribution show that, occasionally, some random samples yielded relatively extreme means--for example, some greater than 10 or less than 0. This visualization clearly maps the probability of obtaining these less common outcomes.

Calculating Probabilities Based on Sample Means

The primary applied utility of the sampling distribution is to calculate the probability of obtaining a specific range of values for a sample mean, given known population parameters and the sample size. This is essential for hypothesis testing and confidence interval construction.

For instance, let us pose the question: What is the probability that a random sample of size 20 yields a sample mean (Link 2/5) less than or equal to 6? We can calculate this probability directly from our simulated data set in Column U by counting how many of our 1,000 samples satisfy this condition and then dividing by the total number of samples.

We use a combination of the **COUNTIF** and **COUNT** functions in Excel to perform this empirical calculation. Enter the following formula into a blank cell:

=COUNTIF(U2:U1001, "<=6")/COUNT(U2:U1001)

Based on our 1,000 simulations, the probability that the sample mean is less than or equal to 6 is determined to be approximately **0.638** (or 63.8%). This result provides a strong empirical estimation of the theoretical probability.

This empirical result typically aligns very closely with the value calculated using precise theoretical tools, such as an online Sampling Distribution Calculator, which applies the Z-score transformation and the standard normal distribution tables. The image below shows how a theoretical calculation yields a similar result, confirming the accuracy of our simulation:

	V	W	X	Y	Z	AA	AB	AC
		Mean	5.367869	=AVERAGE(U2:U1001)				
5		SD	2.075396	=STDEV(U2:U1001)				
2								
7		Prob mean<=6	0.638	=COUNTIF(U2:U1001, "<=6")/COUNT(U2:U1001)				
3								
5								
4								
7								
1								
9								
1								
1								
6								
3								
4								
7								
8								
8								
7								

μ (population mean)

5.3

σ (population standard deviation)

9

n (sample size)

20

X (random variable)

6

CALCULATE

$P(\bar{x} \leq 6)$: 0.63602

$P(\bar{x} \geq 6)$: 0.36398

Further Statistical Resources

To deepen your understanding of these core statistical concepts and their practical application in various fields, explore the following educational resources:

An Introduction to Sampling Distributions
Sampling Distribution Calculator
An Introduction to the Central Limit Theorem