How do you plot a cumulative distribution function (CDF) in Excel?

Authored by

stats writer

May 3, 2024

A cumulative distribution function (CDF) in Excel is a graphical representation of the cumulative probability distribution of a set of data. It shows the probability of a certain value or less occurring in a data set. To plot a CDF in Excel, first organize the data in ascending order. Then, create a new column next to the data with the formula "=RANK.AVG(cell,range,0)/COUNT(range)" to calculate the cumulative probability for each data point. Next, select the data and insert a scatter plot. Right-click on the plot and select "Change Series Chart Type" and choose "Line with Markers." Finally, add axis labels and a chart title to complete the CDF plot. This allows for a visual representation of the distribution of the data and can be useful in analyzing and understanding the behavior of the data set.

Plot a CDF in Excel

A cumulative distribution function (CDF) describes the probability that a random variable takes on a value less than or equal to some number.

We can use the following function in Excel to calculate cumulative distribution probabilities:

=NORM.DIST(x, MEAN, STANDARD_DEVIATION, TRUE)

The following example shows how to calculate and plot a CDF in Excel.

Example: Calculate & Plot CDF in Excel

First, let's create the following dataset in Excel:

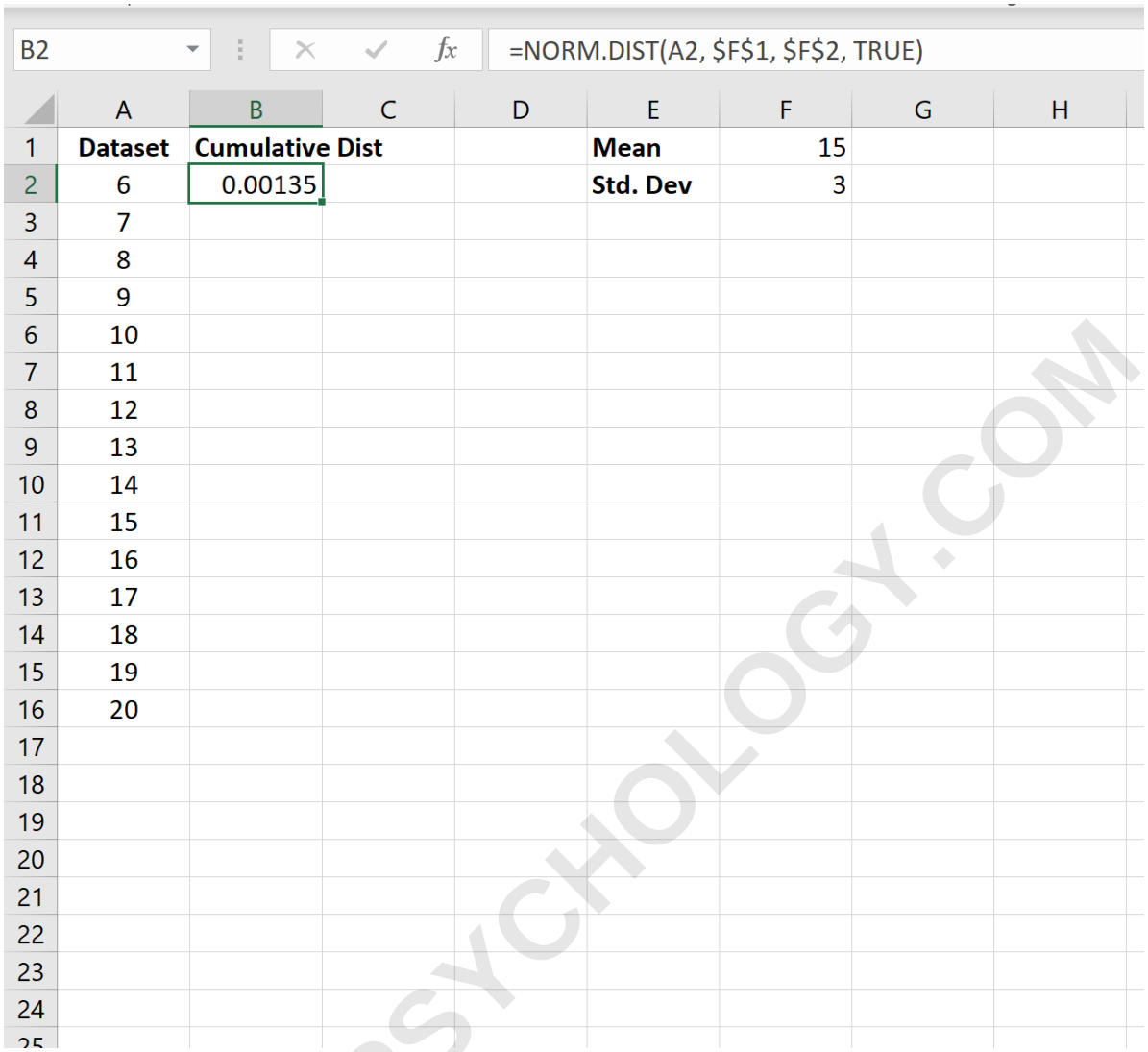
	A	B	C	D	E	F	G	H
1	**Dataset**							
2	6							
3	7							
4	8							
5	9							
6	10							
7	11							
8	12							
9	13							
10	14							
11	15							
12	16							
13	17							
14	18							
15	19							
16	20							
17								
18								
19								
20								
21								
22								
23								
24								

Next, let's specify the mean and standard deviation of the distribution:

	A	B	C	D	E	F	G
1	**Dataset**				**Mean**	15	
2	6				**Std. Dev**	3	
3	7						
4	8						
5	9						
6	10						
7	11						
8	12						
9	13						
10	14						
11	15						
12	16						
13	17						
14	18						
15	19						
16	20						
17							
18							
19							
20							
21							
22							
23							
24							
25							
26							

Next, we can calculate the cumulative distribution probability for the first value in the dataset by using the following formula:

=NORM.DIST(A2, F1, F2, TRUE)



	A	B	C	D	E	F	G	H
1	**Dataset**	**Cumulative Dist**			**Mean**	15		
2	6	0.00135			**Std. Dev**	3		
3	7							
4	8							
5	9							
6	10							
7	11							
8	12							
9	13							
10	14							
11	15							
12	16							
13	17							
14	18							
15	19							
16	20							
17								
18								
19								
20								
21								
22								
23								
24								

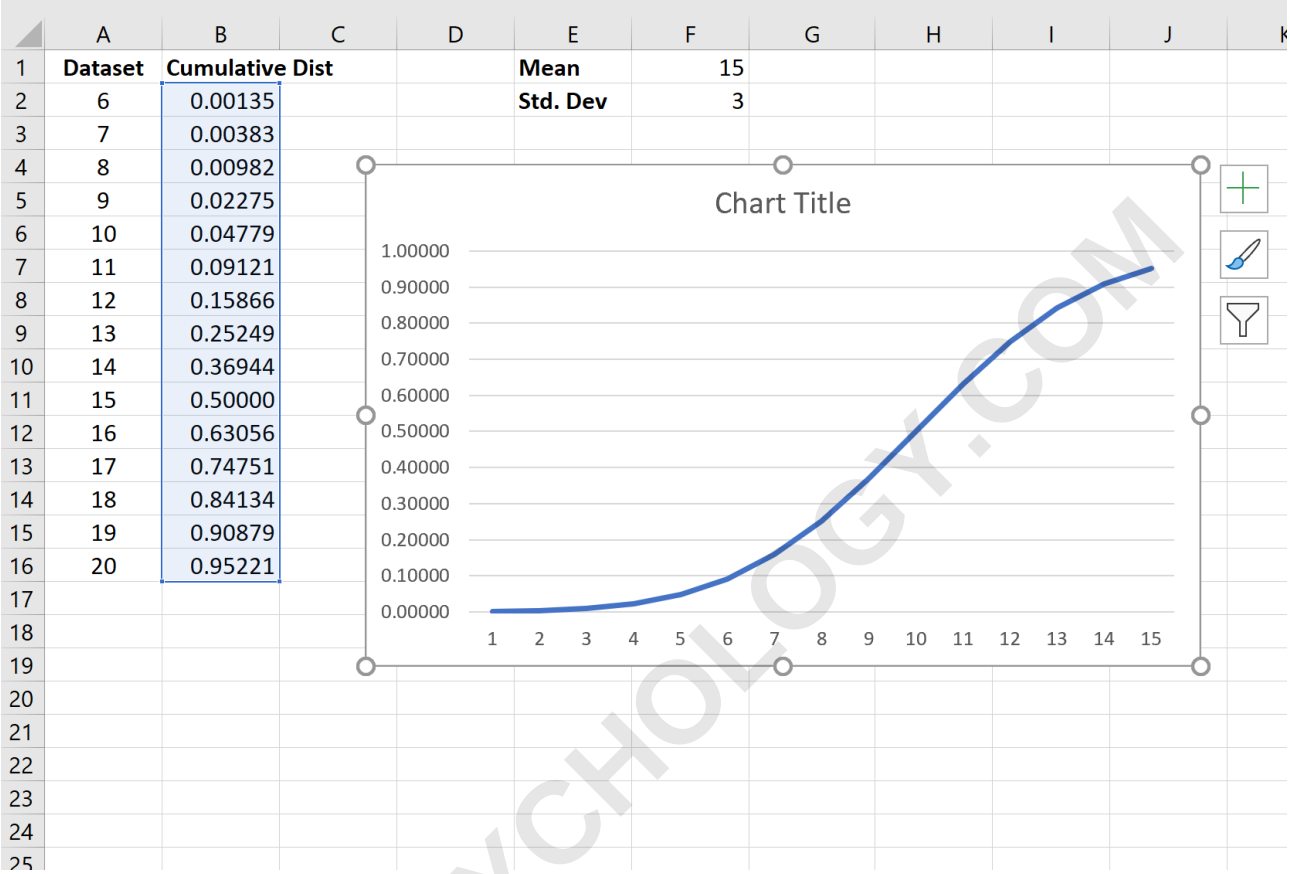
Next, we can copy and paste this formula down to every other cell in column B:

	A	B	C	D	E	F	G	H
1	**Dataset**	**Cumulative Dist**			**Mean**	15		
2	6	0.00135			**Std. Dev**	3		
3	7	0.00383						
4	8	0.00982						
5	9	0.02275						
6	10	0.04779						
7	11	0.09121						
8	12	0.15866						
9	13	0.25249						
10	14	0.36944						
11	15	0.50000						
12	16	0.63056						
13	17	0.74751						
14	18	0.84134						
15	19	0.90879						
16	20	0.95221						
17								
18								
19								
20								
21								
22								
23								
24								
25								

The CDF is now complete. The way we interpret the values is as follows:

The probability that the random variable will take on a value equal to or less than 6 is .00135.The probability that the random variable will take on a value equal to or less than 7 is .00383.The probability that the random variable will take on a value equal to or less than 8 is .00982.

And so on.



	A	B	C	D	E	F	G	H	I	J
1	**Dataset**	**Cumulative Dist**			**Mean**	15				
2	6	0.00135			**Std. Dev**	3				
3	7	0.00383								
4	8	0.00982								
5	9	0.02275								
6	10	0.04779								
7	11	0.09121								
8	12	0.15866								
9	13	0.25249								
10	14	0.36944								
11	15	0.50000								
12	16	0.63056								
13	17	0.74751								
14	18	0.84134								
15	19	0.90879								
16	20	0.95221								
17										
18										
19										
20										
21										
22										
23										
24										
25										

The values along the x-axis show the values from the dataset and the values along the y-axis show the CDF values.