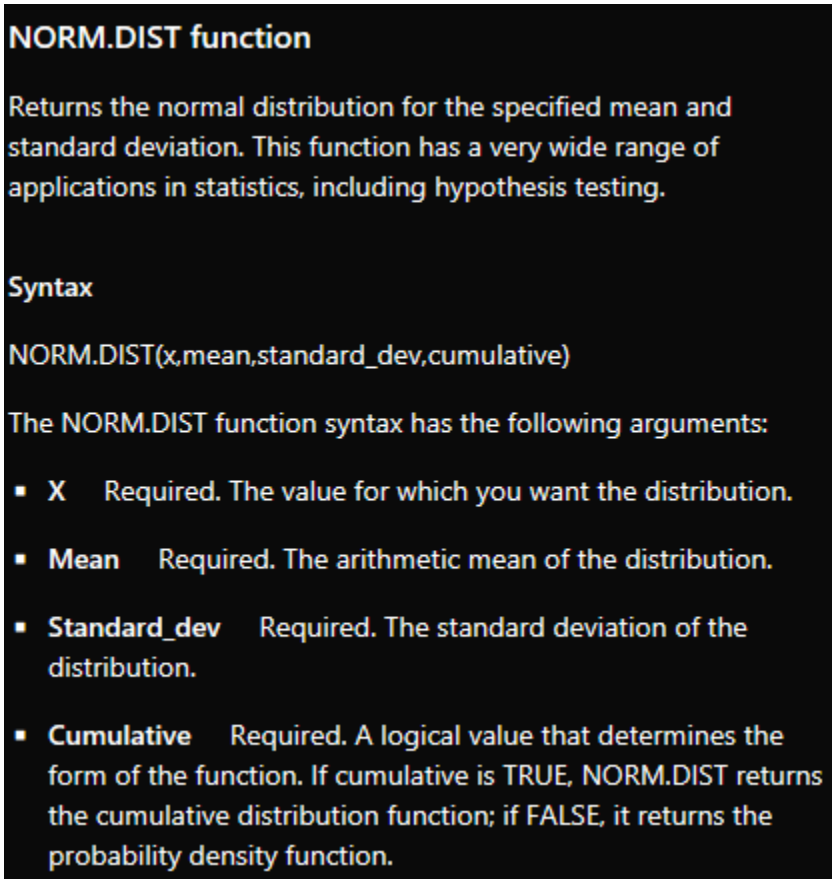
Calculation for Normal Distribution:

Theoretical:



Mean = 228.079
Variance = 397292.0595 => Standard_dev = 630.31
I chose X as : {5,10,20,40,60,80,100,200,500,600,700,900,1000,1500,2000}
Cumulative is set to TRUE as I'm calculating the CDF.

Using the formula for each value of X given in the set above, we get:

F(5) = 0.0005
F(10) = 0.364
F(20) = 0.370
F(40) = 0.382
F(60) = 0.394
F(80) = 0.407
F(100) = 0.419
F(200) = 0.482
F(500) = 0.666
F(600) = 0.722

F(700) = 0.772
F(900) = 0.856
F(1000) = 0.889
F(1500) = 0.978
F(2000) = 0.997

Practical:

F(x) = number of values in the dataset less than or equal to x / total number of samples

Using the above formula we get:

F(5) = 0.0524
F(10) = 0.1432
F(20) = 0.2912
F(40) = 0.4404
F(60) = 0.5268
F(80) = 0.5858
F(100) = 0.6288
F(200) = 0.7406
F(500) = 0.879
F(600) = 0.9002
F(700) = 0.9206
F(900) = 0.9448
F(1000) = 0.953
F(1500) = 0.9782
F(2000) = 0.9882

From the excel file plot for the normal distribution we can see that it does quite seem to fit well.

P.T.O

Calculation for Beta Distribution:

Theoretical:

BETA.DIST function

Returns the beta distribution.

The beta distribution is commonly used to study variation in the percentage of something across samples, such as the fraction of the day people spend watching television.

Syntax

BETA.DIST(x,alpha,beta,cumulative,[A],[B])

The BETA.DIST function syntax has the following arguments:

- **X** Required. The value between A and B at which to evaluate the function
- **Alpha** Required. A parameter of the distribution.
- **Beta** Required. A parameter of the distribution.
- **Cumulative** Required. A logical value that determines the form of the function. If cumulative is TRUE, BETA.DIST returns the cumulative distribution function; if FALSE, it returns the probability density function.
- **A** Optional. A lower bound to the interval of x.
- **B** Optional. An upper bound to the interval of x.

As before,
I chose X as : {5,10,20,40,60,80,100,200,500,600,700,900,1000,1500,2000}
Alpha = 0.38
Beta = 4
Cumulative = TRUE as we are calculating CDF.
A = 0.444
B = 2000
These are the upper and lower limit values to the interval of X

Using the formula for each value of X given in the set above, we get:

F(5) = 0.182
F(10) = 0.241
F(20) = 0.316
F(40) = 0.409
F(60) = 0.475
F(80) = 0.525
F(100) = 0.568
F(200) = 0.710
F(500) = 0.897
F(600) = 0.927
F(700) = 0.949
F(900) = 0.976
F(1000) = 0.984
F(1500) = 0.999
F(2000) = 1

Practical:
Practical values remain the same as before.

By comparing the values and plotting(refer excel sheet), we can see that this gives us a much better fit compared to normal distribution.

Hence, beta distribution seems like a good fit to this data compared to normal distribution.