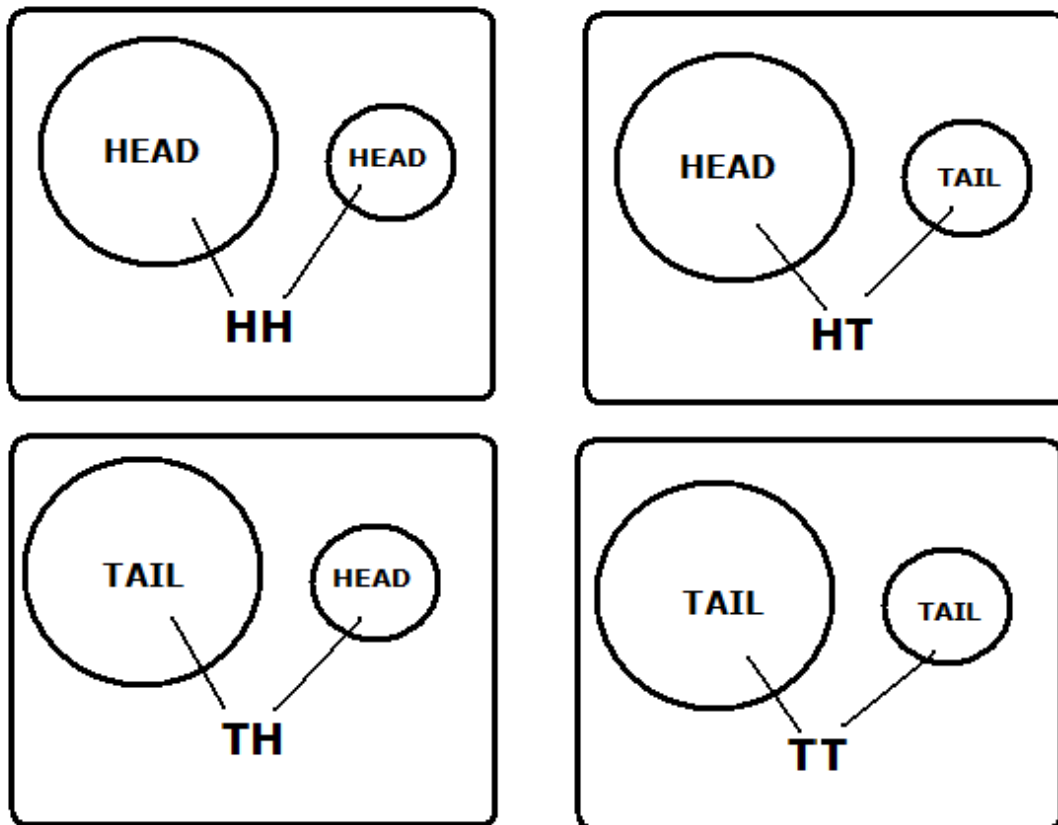


Basic Probability Calculations

Probability of an event is a number between 0 and 1 that gives us a measure of likelihood of the occurrence of that event.

Example 1: If we toss a fair coin with two faces head or tail, the probability that a head will show up is $\frac{1}{2}$ and the probability that a tail will show up is also $\frac{1}{2}$ because in a large number of tosses, a fair coin will give us head half of the times and tail half of the times.

Example 2: If we toss a fair quarter and a fair dime. Then there are **FOUR** possible outcomes as shown in the diagram below.



Since all the outcomes are equally likely, the probability that we shall have two heads is $P(HH) = \frac{1}{4}$

There is another way to approach this to note that

P(Head on the quarter) is $\frac{1}{2}$

and

P(Head on the dime) is $\frac{1}{2}$

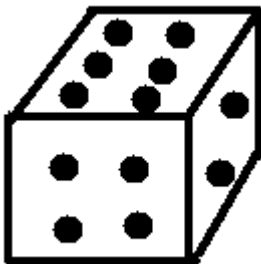
Therefore the probability of heads on both the quarter and the dime is $\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$

In general if events A and B are independent,

$$P(A \text{ and } B) = P(A) \times P(B)$$

Example 3:

Refer to a six faced die

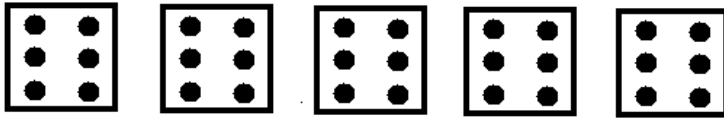


the probability that a fair six faced die will have six dots on the top face



is $\frac{1}{6}$

If we roll five fair dice, because the outcome on each die is independent of the outcome on the other dice, the probability that all five will show six dots facing up



$$\text{is } \frac{1}{6} \times \frac{1}{6} \times \frac{1}{6} \times \frac{1}{6} \times \frac{1}{6} = \frac{1}{7776} \cong 0.0001286$$

In general, when we have independent trials with only two outcomes "success" or "failure" in each trial then can apply binomial probability calculations provided that the probability of success stays the same in each trial.

Let us look at the following example, where we can apply binomial probability calculations.

Example 4:

According to a report from the ABC news available at

<http://abcnews.go.com/Health/ReproductiveHealth/story?id=3024044&page=1>

" Traditionally, the expected ratio of male to female births is believed to be 1.05. That is, for every 100 girls born, you would expect 105 boys to be born. "

If we agree with the above statement, the probability that a new born will be a girl is $\frac{100}{205} = \frac{20}{41}$

Let us apply the binomial probability calculation in the following situation:

There are 20 babies expected to be born in a community where the gender of the baby is not known. We would like to know the probability that exactly five of these 20 will be girls.

Note that we can treat this situation as of 20 independent trials

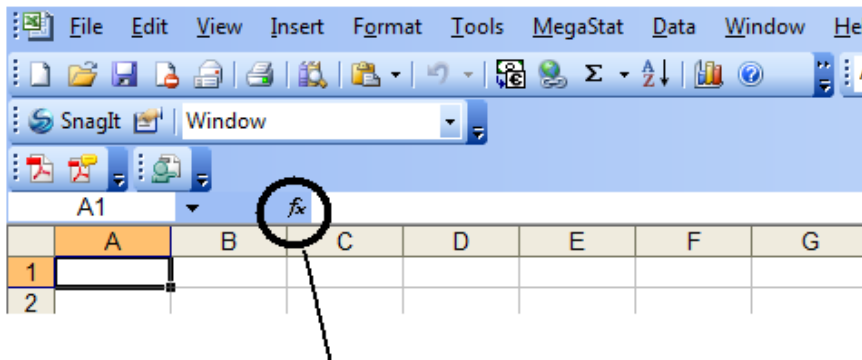
term the birth of a girl as a "success"

the probability of a "success" in any trial is $\frac{20}{41}$

We would like to find the probability of exactly 5 "successes"

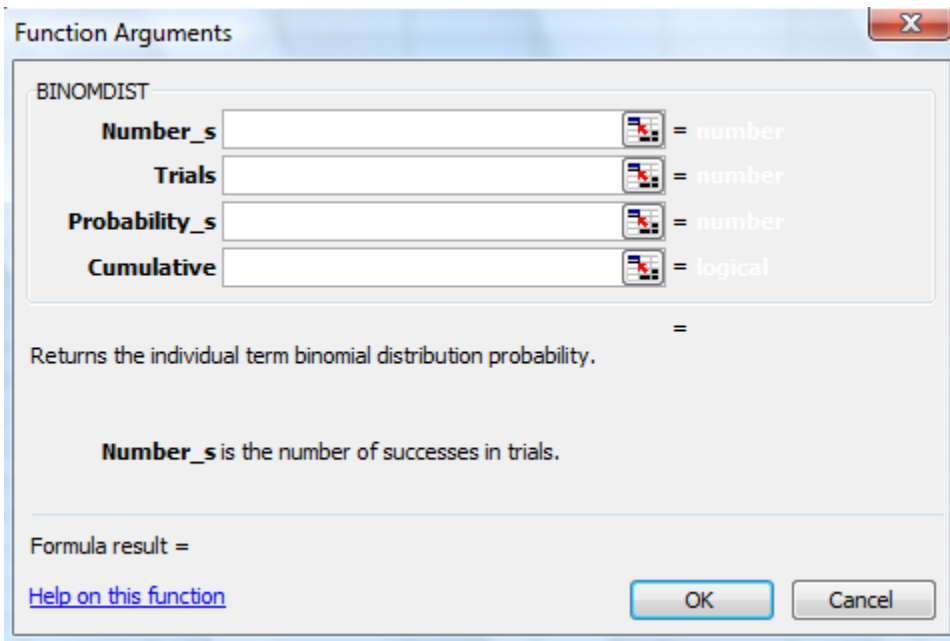
Instead of using the binomial probability formula, we may utilize the "paste function" feature in Excel by clicking on the

 symbol located in

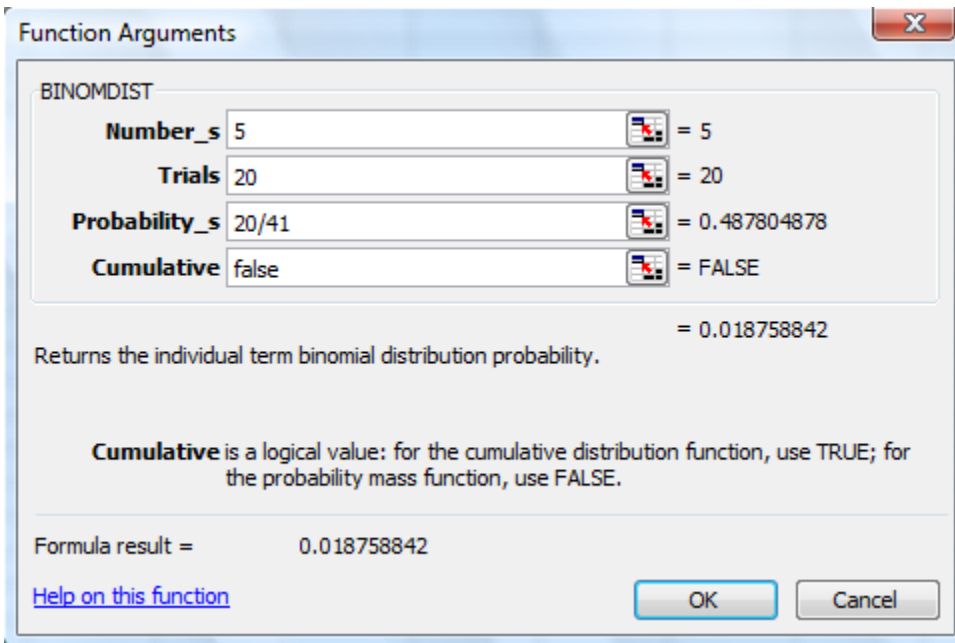


paste function

Find the statistical feature and then select the BINOMDIST click OK to see the following dialogue box

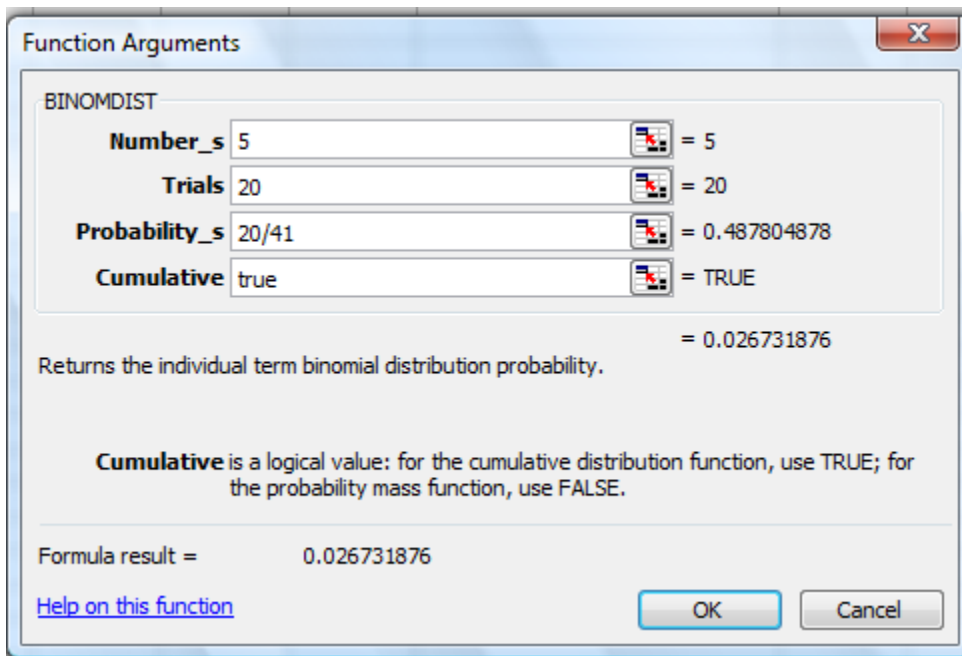


Since the probability here is not cumulative, we shall enter false for the cumulative frequency



The probability of exactly five girls in 20 births is 0.018758842

On the other hand, if we would like the probability the number of girls among these 20 new born will be 5 OR less, we are calculating cumulative probability and this value can be obtained by writing "true" in the cumulative probability box



The probability the number of girls will be five or less among these 20 new born is 0.026731876

When the number of successes for a binomial distribution becomes very large, we may use the Poisson Distribution.

The Poisson Distribution is described below

When we are looking at the # of successes over an interval (could be time interval, or a region) such that the number of successes over any two disjoint intervals are independent

and that average is appropriately proportion to the size of the interval.

Poisson probability calculations may be applied in the study of birth defects and genetic mutations. One such reference is "Functional genetic analysis of mouse chromosome 11" published in *Nature* | Volume 425 | 4 SEPTEMBER 2003 page 82

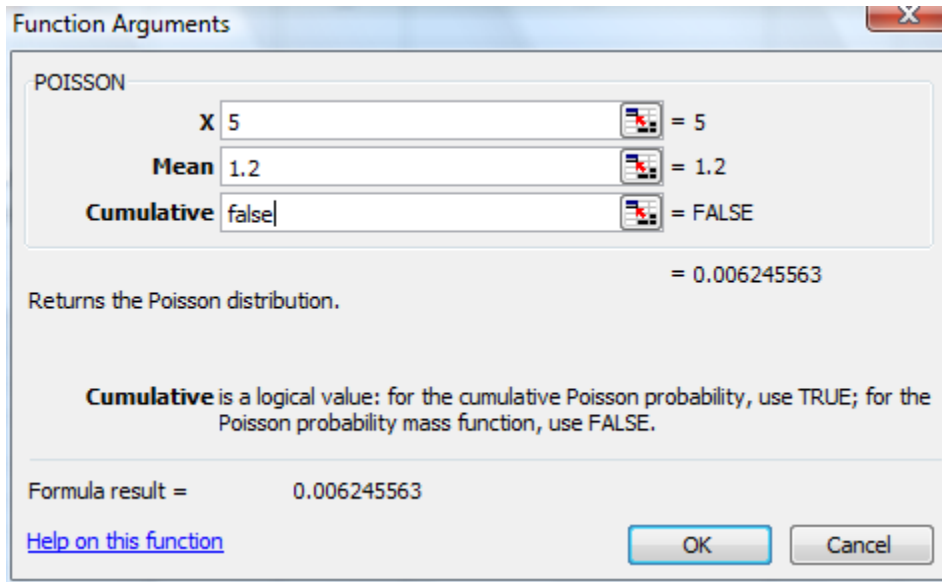
Let us understand Poisson probability calculations using the following example from Statistical Research Methods in Life Sciences by P V Rao, Thomson Publishing page 42.

Assume that the number of bacterial colonies found in an ager plate follows a Poisson Distribution and that on an average, an ager plate with an area of 100 cm^2 will contain 12 colonies. To find the probability that 5 colonies will be found in a 10-cm^2 agar plate.

Note that for a 100 cm^2 area, the average for the number of "successes" is 12

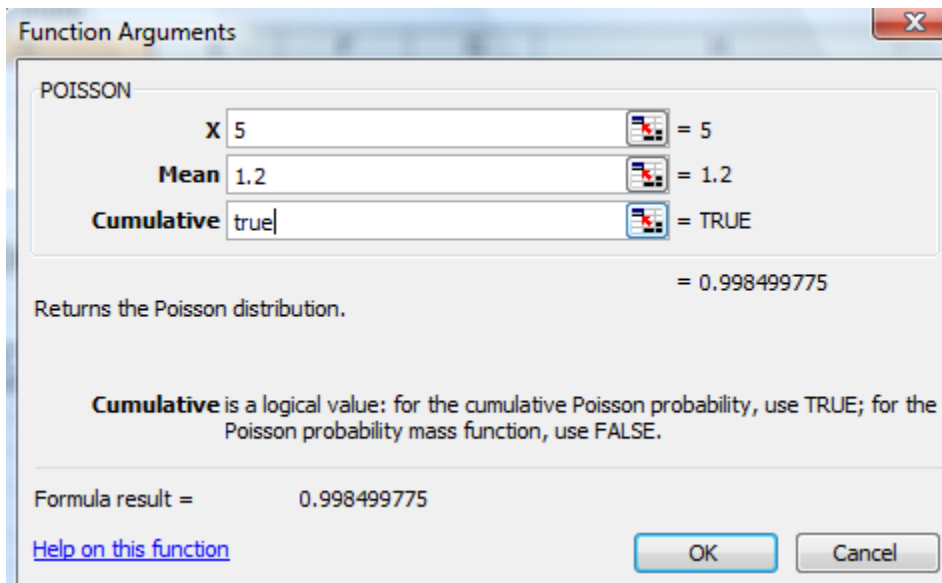
Therefore for a 10-cm^2 area, the average number of "successes" is $12 \times \frac{10}{100} = 1.2$

To find the probability of 5 (here they mean) successes in a 10-cm^2 area, we can use the Poisson option in statistical part of paste function in Excel



The probability of exactly 5 colonies is 0.006245563

If we would like the cumulative probability of 5 i.e. the probability that the number of colonies will be five or less, we can modify the above dialogue box to give us the answer



The probability that the number of colonies will be 5 or less is 0.998499775

Note the probability calculations shown above involved discrete random variables because the variable could take only whole number values such as 0, 1, 2, 3, 4,



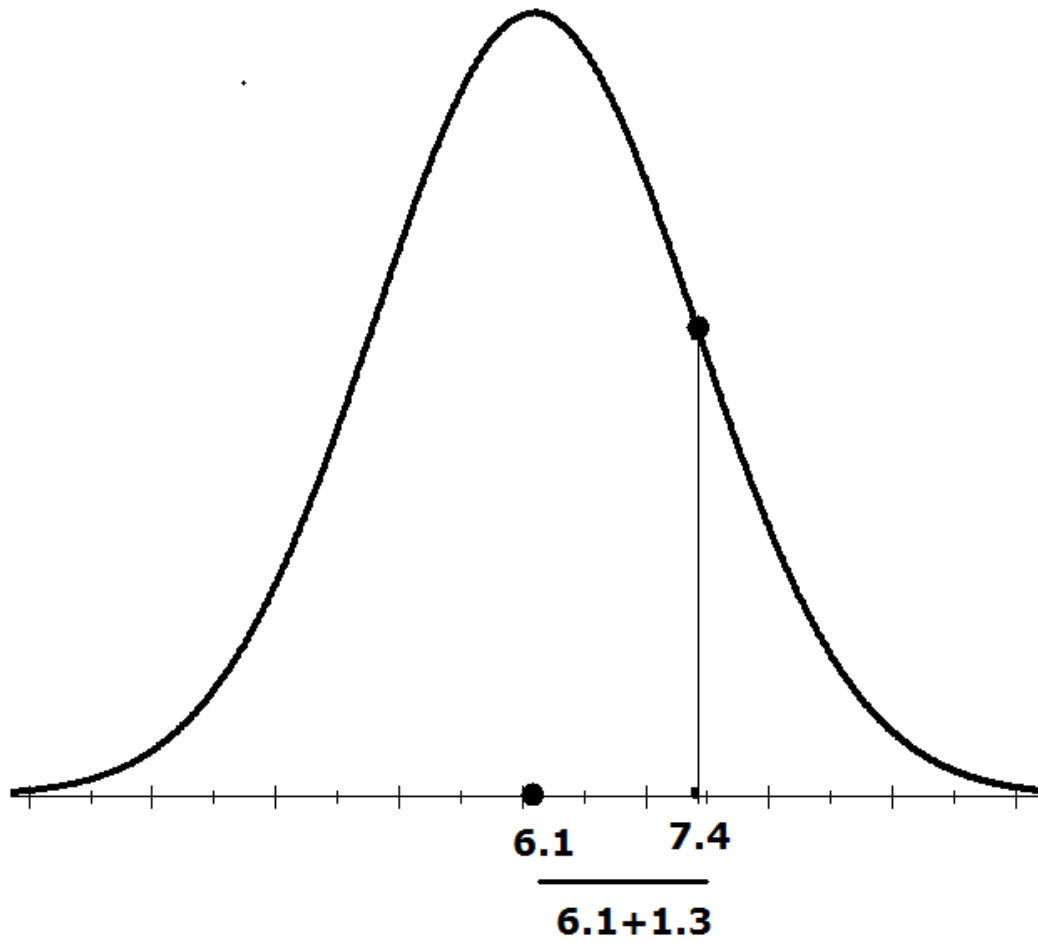
Now, we shall take up the random variables that take values on a continuous scale, such as weight, height, temperature etc.



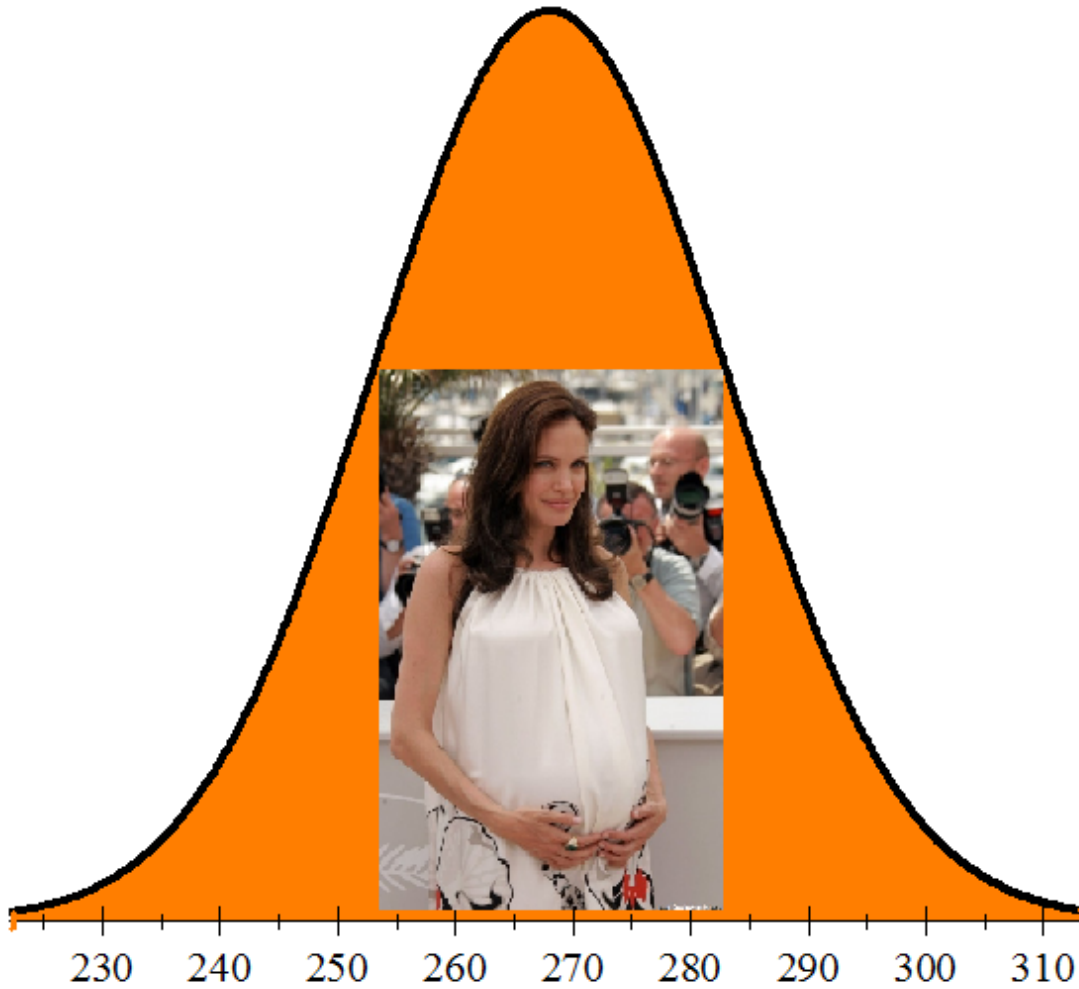
Let us take up a very commonly used continuous distribution called Normal Distribution or Gaussian Distribution. We saw a few examples in the context of standard deviations.

A graph normal distribution is a bell shaped curve with the shape determined by the parameters the mean and the standard deviation.

Example : The length of female polyhaete worms are normally distributed with mean 6.1 mm and standard deviation 1.3 mm. The following is a display of this distribution

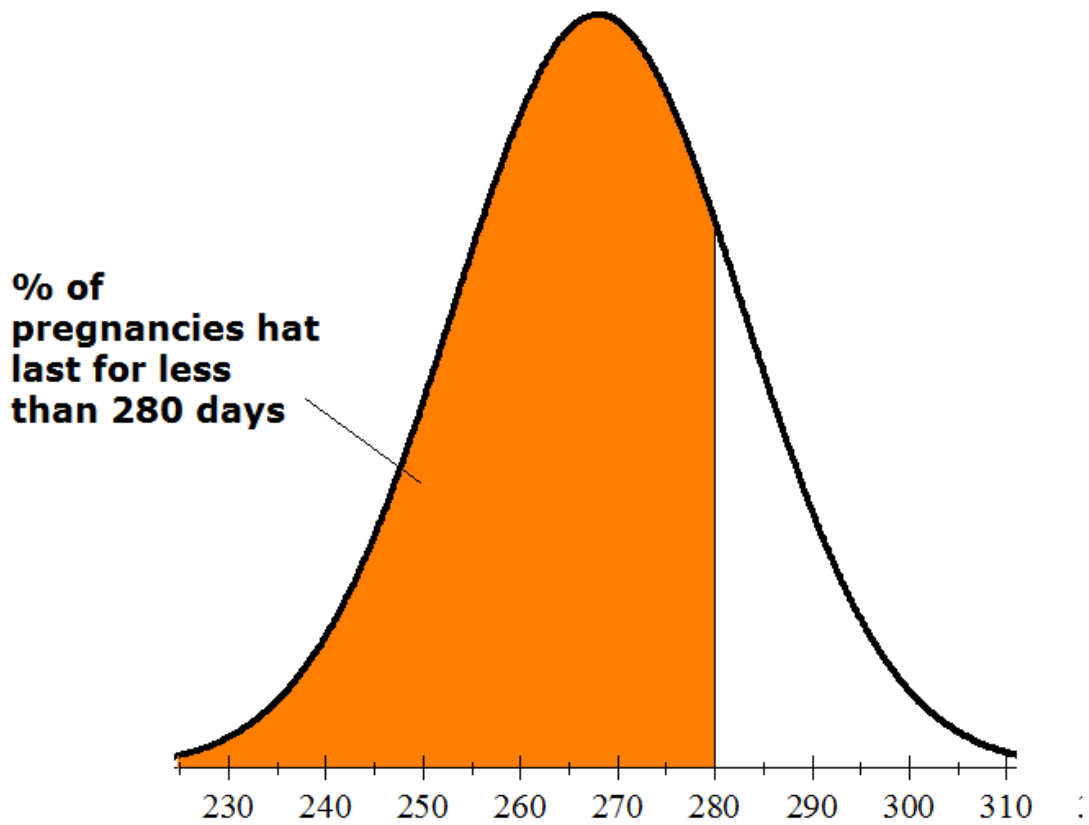


Example 2: The length of human pregnancy shows a normal distribution with a mean of 268 days and standard deviation of 15 days. The distribution is displayed below

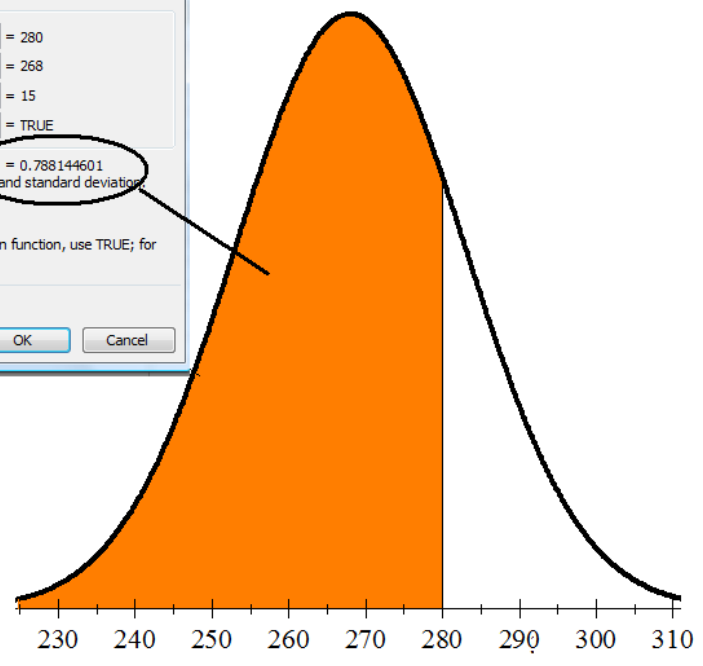
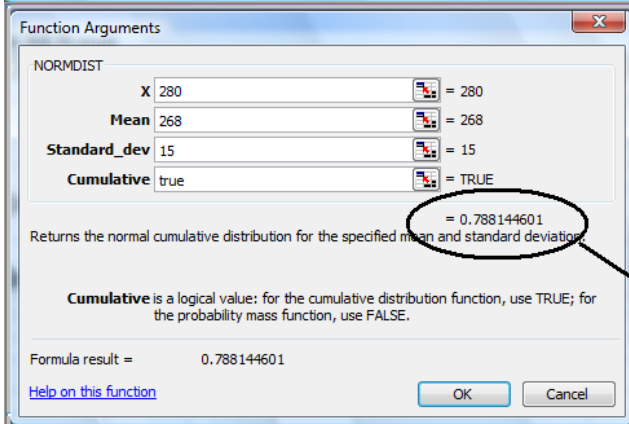
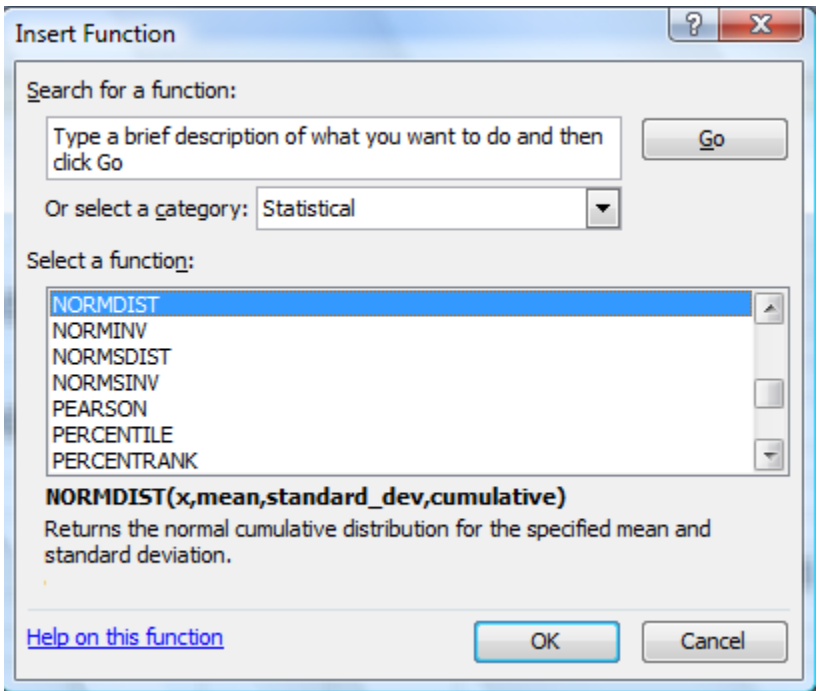


We can use the paste function feature in Excel to find the area between any two x-values to answer questions such as

What percentage of the pregnancies last for less than 280 days?

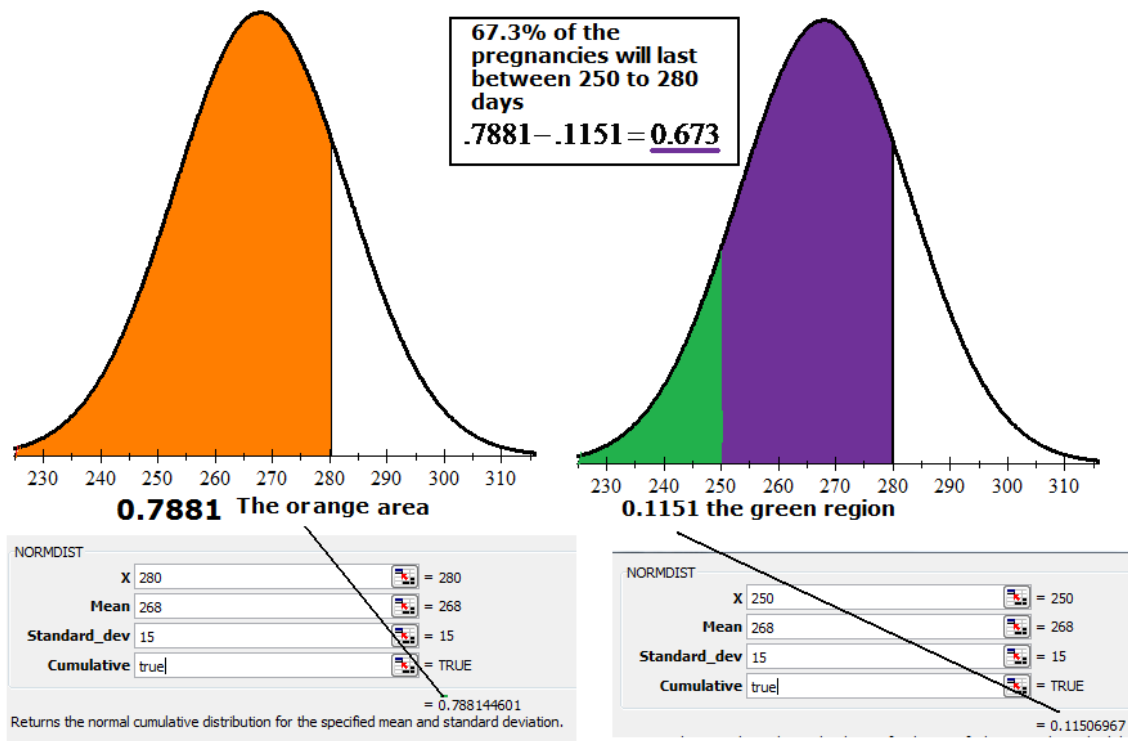


The proportion may be found by using the facility NORMDIST in paste function in Excel



Therefore 78.81% of the human pregnancies last for less than 280 days.

If we have to find the percentage of pregnancies that last between 250 and 280 days, note that



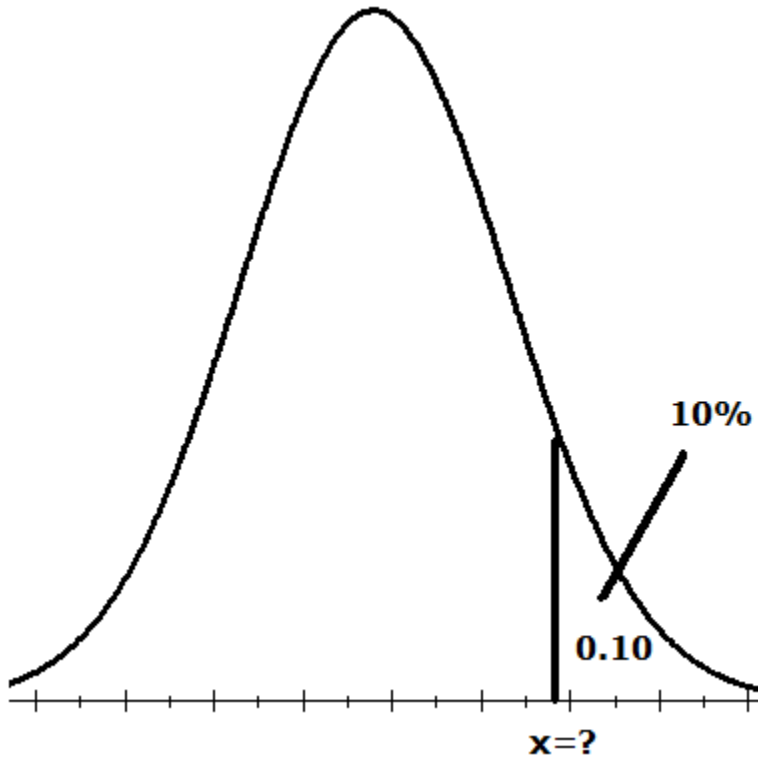
Therefore 67.3% of the pregnancies last between 250 to 280 days.

To find the percentage of pregnancies that last for more than 280 days, note that 78.81% last for less than 280 days therefore $100 - 78.81 = 21.19\%$ last for more than 280 days.

To answer the question:

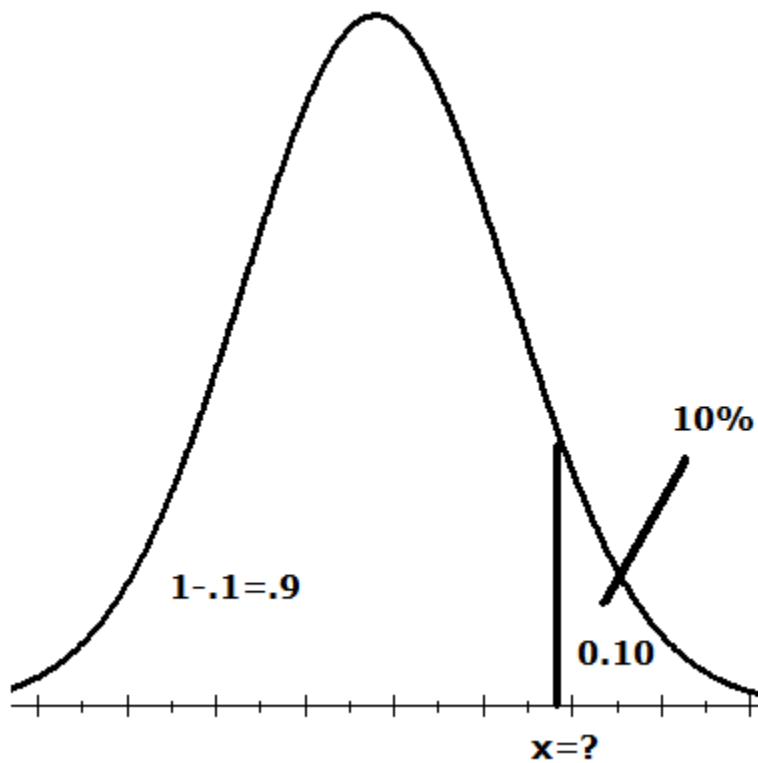
How long should a pregnancy be so that it is among top 10% of all the durations of human pregnancies?

We have to find the value of x so that the area to the right of x is 0.10.



OR

We have to find the value of x so that the area to the left of x is 0.90.



OR we have to find the 90th percentile of the normal distribution with mean 268 days and standard deviation 15 days

In Excel, the statistical feature in paste function has a choice norminv,

NORMINV			
Probability	.90		= 0.9
Mean	268		= 268
Standard_dev	15		= 15
			= 287.2232735
Returns the inverse of the normal cumulative distribution for the specified mean and standard deviation.			

Shows that 287.23 is the 90th percentile of this distribution.

