Binomial Distribution

**Notes**

* In probability theory and statistics, the binomial distribution is the discrete probability distribution of the number of successes in a sequence of n independent yes/no experiments, each of which yields success with probability p.
* The binomial distribution is frequently used to model the number of successes in a sample of size n drawn with replacement from a population of size N.
* A binomial distribution is very different from a normal distribution.
* The key difference is that a binomial distribution is discrete, not continuous. In other words, it is NOT possible to find a data value between any two data values.

**Binomial Distribution**

* The binomial distribution is used when there is a fixed number of trials, and each trial has two possible outcomes - a success or a failure (something happened or it did not).
* The fixed number of trials is n.
* The probability of success is p.
* For example, we can play a game where we roll a die 10 times and count how many times the number 6 shows up.
* In this example, the number of trials is n = 10, and the probability that a 6 will show up is p = 1/6 = 0.1667
* Symbolically, we can say X~Binomial (n=10, p=1/6), which is read as "X follows a binomial distribution with n = 10, and p = 1/6"

**The requirements for a binomial distribution:**

1. The of interest is the count of successes in n trials
2. The number of trials (or sample size), n, is fixed
3. Trials are independent, with fixed value p = P(success on a trial)
4. There are only two possible outcomes on each trial, called "success" and "failure." (This is where the "bi" prefix in "binomial" comes from.

**Rules for finding binomial probabilities in Excel**

1. If the question asks you to find the probability of exactly one number, use BINOMDIST(successes,trials,probability,FALSE).
2. If the question asks you to find the probability of up to and including a number (less than or equal to a number), use BINOMDIST(successes,trials,probability,TRUE).
3. If the question asks you to find the probability of less than a number, use BINOMDIST(successes,trials,probability,TRUE) - BINOMDIST(successes,trials,probability,FALSE).
4. If the question asks you to find the probability of at least one number (greater than or equal to a number), use (1 - BINOMDIST(successes,trials,probability,TRUE)) + BINOMDIST(successes,trials,probability,FALSE).
5. If the question asks you to find the probability of greater than a number, use 1 - BINOMDIST(successes,trials,probability,TRUE).

**Illustration**

* Say that you have five trials, with the probability of success being .25. After five outcomes, you may observe,
* 0 successes
* 1 success
* 2 successes
* 3 successes
* 4 successes
* 5 successes
* Each outcome has its own probability, which can be found using

BINOMDIST(successes,5,.25,FALSE).

* The results are:
* Pr(0 successes) = .237
* Pr(1 success) = .396
* Pr(2 successes) = .264
* Pr(3 successes) = .088
* Pr(4 successes) = .015
* Pr(5 successes) = .001
* These probabilities will sum to one (within rounding).

**The above rules work as follows:**

1. You are asked to find the probability of observing exactly 3 successes. This is equal to .088.
2. You are asked to find the probability of observing up to 3 successes. In other words, you need to report the probability of observing a number of successes less than or equal to 3.

* The CUMULATIVE argument to the BINOMDIST function adds all of the probabilities up to the specified number of successes.
* Thus, BINOMDIST(3,5,.25,TRUE) is equivalent to .237 + .396 + .264 + .088.

1. You are asked to find the probability of observing less than 3 successes.

* This can be found by using BINOMDIST(3,5,.25,TRUE) to add up .237 + .396 + .264 + .088, but then you must subtract out the .088.
* Thus, the formula is:

BINOMDIST(3,5,.25,TRUE) - BINOMDIST(3,5,.25,FALSE), which is equivalent to (.237 + .396 + .264 + .088) – (.088).

1. You are asked to find the probability of observing at least three successes.

* Because the probabilities sum to one, we can first find the probability of observing more than three successes as 1 - BINOMDIST(3,5,.25,TRUE).
* The result is (1 – (.237 + .396 + .264 + .088)), which is equivalent to .015 + .001. However, this is only the probability of observing greater than 3 successes, when we need to find greater than or equal to 3 successes. We therefore have to add the probability of exactly 3 successes back in.
* The full syntax is (1 - BINOMDIST(3,5,.25,TRUE)) + BINOMDIST(3,5,.25,FALSE). This is equivalent to (1 – (.237 + .396 + .264 + .088)) + (.088).

1. You are asked to find the probability of observing more than three successes. Because the probabilities sum to one, we can find this probability as 1 - BINOMDIST(3,5,.25,TRUE).

* Doing so performs the calculation (1 – (.237 + .396 + .264 + .088)), which is the same as .015 + .001.

**Application**

1. Warranty records show that the probability that a new car needs a warranty repair in the first 90 days is 0.05.  If a sample of 3 new cars is selected:
2. What is the probability that none needs a warranty repair?
3. What is the probability that at least one needs a warranty repair?
4. What is the probability that more than one needs a warranty repair?

The key components of this problem are:

* Sample size (trials) = 3 new cars
* Probability of “success” = 5%
* Number of “successes”:

A: Exactly 0

B: 1 or more

C: More than one

Solutions are:

* For A: BINOMDIST(0,3,0.05,FALSE)
* For B: 1-BINOMDIST(0,3,0.05,FALSE)
* For C: 1-BINOMDIST(1,3,0.05,TRUE)

1. According to a recent article in *The Wall Street Journal*, 40% of web surfers who view Internet banner ads remember them.  Suppose that a random sample of 5 web surfers is selected and asked if they remember a specific Internet banner that they previously accessed.
2. What is the probability that none of the web surfers will remember the banner ad?
3. What is the probability that exactly 1 of the web surfers will remember the banner ad?
4. What is the probability that 3 or more of the web surfers will remember the banner ad?

The key components of this problem are:

* Sample size (trials) = 5 Web surfers
* Probability of “success” = 40%
* Number of “successes”:

A: Exactly 0

B: Exactly 1

C: 3 or more

Solutions are:

* For A: BINOMDIST(0,5,0.4,FALSE)
* For B: BINOMDIST(1,5,0.4,FALSE)
* For C: (1-BINOMDIST(3,5,0.4,TRUE) + BINOMDIST(3,5,0.4,FALSE)

1. An important part of the customer service responsibilities of a telephone company relates to the speed with which troubles in residential services can be repaired.  Suppose past data indicate that the likelihood is 0.70 that troubles in residential service can be repaired on the same day.  For the first 5 troubles reported on a given day:
2. What is the probability that all 5 will be repaired on the same day?
3. What is the probability that at least 3 will be repaired on the same day?
4. What is the probability that fewer than 2 will be repaired on the same day?

The key components of this problem are:

* Sample size (trials) = 5 reported troubles
* Probability of success = 70%
* Number of successes:

A: Exactly 5

B: 3 or more

C: Fewer than two

Solutions are:

* For A: BINOMDIST(5,5,0.7,FALSE)
* For B: (1-BINOMDIST(3,5,0.7,TRUE)) + BINOMDIST(3,5,0.7,FALSE)
* For C: BINOMDIST(2,5,0.7,TRUE) - BINOMDIST(2,5,0.7,FALSE)

1. A student is taking a multiple-choice exam in which each question has 4 choices.  Assuming that she has no knowledge of the correct answers to any of the questions, she has decided on a strategy in which she will place 4 balls (marked A, B, C and D) into a box.  She randomly selects one ball for each question and replaces the ball in the box.  The marking on the ball will determine her answer to the question.  There are 5 multiple-choice questions on the exam.
2. What is the probability that she will get 5 questions correct?
3. What is the probability that she will get at least 4 questions correct?
4. What is the probability that she will get no questions correct?
5. What is the probability that she will get no more than 2 questions correct?

The key components of this problem are:

* Sample size (trials) = 5 problems
* Probability of success = 25%
* Number of successes:

A: Exactly 5

B: 4 or more

C: Exactly 0

D: Two or Fewer

Solutions are:

* For A: BINOMDIST(5,5,0.25,FALSE)
* For B: (1 - BINOMDIST(4,5,0.25,TRUE) ) + BINOMDIST(4,5,0.25,FALSE)
* For C: BINOMDIST(0,5,0.25,FALSE)
* For D: BINOMDIST(2,5,0.25,TRUE)