Understanding When to Use Binomial, Poisson, Normal, or Sampling Functions (Mixed Probability Problems)

There is no "set of rules" for knowing when to use which equation. Judgment should be used when reading over a question to try to distinguish which probability function to use.
The best way to understand which distribution to use, and when, is to PRACTICE.

However, there are hints and tips you can use to guide your decision.

1. **After you read a question, lay out all the data that is given to you in the question**

EXAMPLE 1

A survey shows that high school students' average fail rate per semester is 1 out of 5 tests, with a standard deviation of 0.041 tests.

a. If students have 12 tests this year, what is the expected number of tests that will be failed?

What is given to you?

Number of tests = n = 12
Probability of failing a test = p = $\frac{1}{5}$
Standard deviation = 0.041

2. **Once all the data is organized, try to understand what the question is asking for. This step is important because sometimes a question can give you unnecessary data, and you must determine what values to use.**

In the above example, the question is asking for the expected number of failed tests. Thus, it is asking for the EXPECTED VALUE, or in other words, the MEAN number of failed tests.
With the data provided, can we find the expected value?

Yes.
The equation for expected value (μ) is:
$E(x) = \mu = np$
Thus, $E(x) = \mu = (12)(\frac{1}{5}) = 2.4$.

Thus, the expected number (or average number) of failed tests for a high school student who has 12 tests in a year is 2.4 tests.

Notice that this question did not require us to use the standard deviation.

b. Compute the probability that during the year, a student fails at least 1 test.

In this case we are looking to compute $P(x \geq 1)$. Let's review what information we have now:

$n = 12$

Probability of failing a test = $p = \frac{1}{5}$

$\sigma = 0.041$

$E(x) = \mu = 2.4$

If you do not know which function to use, you can compare methods:

Binomial	Normal	Poisson
To find $P(x \geq 1)$ using a binomial function, you need: n, p, x n= 12 $p = \frac{1}{5}$ x = 0 Can we use binomial? Yes. All data is given.	To find $P(x \geq 1)$ using a normal function, you need: Upper, lower, σ, μ However, a normal function is only used when there is an infinite amount of possibilities of x. In this case x is limited to *n*, the number of observations. Thus, we cannot use a normal distribution function.	To find $P(x \geq 1)$ using a Poisson function, you need: x, μ x = 0 $\mu = 2.4$ Can we use Poisson? No. Although you have all the necessary info, the question requires you to consider this average in relation to a set number of observations (*n*). If we were to use Poisson, the average of 2.4 would be assumed no matter how many tests are taken by a person in a year. The Poisson function does not take into account the 12 tests; therefore, we should not use it.

As exemplified by this answer, you can have sufficient information to use a distribution, but it may not be the distribution we need to use. In this case, by looking at the importance of the *n* value, we understand which distribution function to use.

Some tips/differentiating factors:

1. Binomial and Poisson distributions can only compute the probability of **discrete** variables, e.g. number of people, number of tests
2. Normal distributions compute the probability of **continuous** variables, e.g. time, money, kilometers. This is because the curve of a normal distribution never touches the x-axis.
3. Binomial distributions require an *n* value
4. Poisson distributions measure probability over a period of time/space
5. Sampling distributions compute the probability of continuous variables in a sample using information from a general population

EXAMPLE 2:

A survey claims that students spend an average of 5 hours studying the day before a test, with a standard deviation of 60 minutes. If there are 42 students who have a test tomorrow, what is the probability that more than half of them will spend at least 3 hours studying today?

1. **After you read a question, lay out all the data that is given to you in the question**

   $n = 42$ students $\sigma = 1$ hour
   $\mu = 5$ hours

2. **Once all the data is organized, try to understand what the question is asking for.**

Since the question is asking for the probability of *half the students* studying for more than 3 hours, we know that a binomial function must be used, since # of students is a discrete variable. Since 42/2 = 21, we are looking for $P(x > 21)$.

HOWEVER, the question asks for the probability that these students will spend *at least 3 hours studying*. This means that a normal function must also be used since time is a continuous variable. Thus, we are also looking for $P(x \geq 3)$.

Which function do we use first?

NORMAL
We use the normal function first because we do not have all the information necessary for a binomial calculation – specifically, we need to find the p-value.

Thus, using DIST → NORM → Ncd our input would be:
Lower = 3 $\mu = 5$
Upper = Big # $\sigma = 1$

Our result would be P = 0.3446. This is the probability of students, in general, studying for at least 3 hours. We can now use this probability to calculate the probability of *more than 21 students* studying for at least 3 hours.

Thus, using DIST → BINM → Bcd our input would be:
$n = 42$
$x = 21$
$p = 0.3446$
So $P(x > 21) = 1 - P(x \leq 21)$
$= 1 - 0.9872$
$= 0.0128$