# Binomial Distribution

Colby Community College

### Example 1

Let us assume there is a 0.85 probability that a randomly chosen adult has heard of Twitter.

Four people are chosen at random:

Ariana ($A$),    Brittany ($B$),    Carlton ($C$),    Damian ($D$)

We want the probability exactly one of them will have heard of Twitter.

There are four combinations possible:

$P(A = \text{yes and } B = \text{no and } C = \text{no and } D = \text{no})$
$= 0.85 \cdot 0.15 \cdot 0.15 \cdot 0.15 = (0.85)^1 (0.15)^3 = 0.002869$

$P(A = \text{no and } B = \text{yes and } C = \text{no and } D = \text{no})$
$= 0.15 \cdot 0.85 \cdot 0.15 \cdot 0.15 = (0.85)^1 (0.15)^3 = 0.002869$

$P(A = \text{no and } B = \text{no and } C = \text{yes and } D = \text{no})$
$= 0.15 \cdot 0.15 \cdot 0.85 \cdot 0.15 = (0.85)^1 (0.15)^3 = 0.002869$

$P(A = \text{no and } B = \text{no and } C = \text{no and } D = \text{yes})$
$= 0.15 \cdot 0.15 \cdot 0.15 \cdot 0.85 = (0.85)^1 (0.15)^3 = 0.002869$

So, the probability exactly one has heard of Twitter is
$0.002869 + 0.002869 + 0.002869 + 0.002869 = 0.11475 = 11.475\%$

## Definition

The **binomial distribution** is used to describe the number of successes in a fixed number or trials.

## Notation

| | |
|---|---|
| $p$ | The probability of a success. |
| $q = 1 - p$ | The probability of a failure. |
| $n$ | The fixed number of trials. |
| $k$ | The number of successes. |

## Note

Example 1 is how to find a binomial distribution the hard way.

## Note

If all the scenarios are independent of each other, then we can calculate the final probability as:

$$[\text{\# of scenarios}] \cdot P(\text{single scenario})$$

## Definition

The **factorial**, for any positive integer *n*, is

$$0! = 1$$
$$1! = 1$$
$$2! = 2 \cdot 1 = 2$$
$$3! = 3 \cdot 2 \cdot 1 = 6$$
$$4! = 4 \cdot 3 \cdot 2 \cdot 1 = 24$$
$$\vdots$$
$$n! = n \cdot (n-1) \cdot (n-2) \cdot \cdots \cdot 4 \cdot 3 \cdot 2 \cdot 1$$

## Note

Factorials can be calculated iteratively. i.e.

$$(n+1)! = n! \cdot (n+1)$$

### Definition

The **binomial coefficients** gives the number of ways to choose $k$ successes in $n$ trials.:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Read "$n$ choose $k$."

### Example 2

The number of ways to choose $k = 3$ successes in $n = 4$ trials:

$$\binom{4}{3} = \frac{4!}{3!(4-3)!} = \frac{4!}{3!1!} = \frac{4 \cdot 3 \cdot 2 \cdot 1}{3 \cdot 2 \cdot 1 \cdot 1} = \frac{4 \cdot \cancel{3} \cdot \cancel{2} \cdot 1}{\cancel{3} \cdot \cancel{2} \cdot 1 \cdot 1} = 4$$

## Binomial Distribution

Suppose the probability of a single trial being a success is $p$. Then the probability of observing exactly $k$ successes in $n$ independent trials is given by

$$\binom{n}{k}p^k(1-p)^{n-k} = \frac{n!}{k!(n-k)!}p^k(1-p)^{n-k}$$

The mean, variance, and standard deviation of the number of observed successes are

$$\mu = np, \qquad \sigma^2 = np(1-p), \qquad \sigma = \sqrt{np(1-p)}$$

## Is It Binomial?

Every binomial distribution has to satisfy the following:

- The trials are independent.
- The number of trials, $n$, is fixed.
- Each trial outcome can be classified as either a *success* or *failure*.
- The probability of a success, $p$, is the same for each trial.

### Example 3

From Example 1 we have $p = 0.85$, $n = 4$, and $k = 1$.

$$
\begin{aligned}
P\,(\text{exactly one has heard of twitter}) &= \binom{n}{k} p^k (1-p)^{n-k} \\
&= \binom{4}{1} (0.85)^1 (1 - 0.85)^{4-1} \\
&= \frac{4!}{1!(4-1)!} (0.85)^1 (0.15)^3 \\
&= \frac{4 \cdot 3 \cdot 2 \cdot 1}{1 \cdot 3 \cdot 2 \cdot 1} (0.85)^1 (0.15)^3 \\
&= \frac{4 \cdot \cancel{3} \cdot \cancel{2} \cdot 1}{1 \cdot \cancel{3} \cdot \cancel{2} \cdot 1} (0.85)^1 (0.15)^3 \\
&= 4 \cdot (0.85)^1 (0.15)^3 \\
&= 0.11475
\end{aligned}
$$

### Example 4

Assume that 70% of customers won't exceed their car insurance deductible. Let's the probability that 5 of 8 randomly selected customers won't exceed their premium.

Start by identifying

$$p = 0.7, \qquad q = 1 - p = 0.3, \qquad n = 8, \qquad k = 5$$

$$\begin{aligned}
\binom{n}{k} p^k (1-p)^{n-k} &= \binom{8}{5}(0.7)^5(0.3)^{8-5} \\
&= \frac{8!}{5!(8-5)!}(0.7)^5(0.3)^3 = \frac{8!}{5!(3)!}(0.7)^5(0.3)^3 \\
&= \frac{8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 \cdot 3 \cdot 2 \cdot 1}(0.7)^5(0.3)^3 \\
&= \frac{8 \cdot 7 \cdot 6 \cdot \cancel{5} \cdot \cancel{4} \cdot \cancel{3} \cdot \cancel{2} \cdot 1}{\cancel{5} \cdot \cancel{4} \cdot \cancel{3} \cdot \cancel{2} \cdot 1 \cdot 3 \cdot 2 \cdot 1}(0.7)^5(0.3)^3 \\
&= 56 \cdot (0.7)^5(0.3)^3 \\
&= 0.254122
\end{aligned}$$

## Example 5

Assume the probability that a smoker will develop a severe lung condition in their life time is 0.3.

*If you have four friends who smoke, are the conditions for the binomial model satisfied?*

It is likely that independence is not satisfied, since they probably all know each other.

## Example 6

Suppose instead four people are randomly selected.

*Is the binomial model appropriate to find the probability that none of them will develop a severe lung condition?*

We are assuming that the four are randomly selected, yes.

$$\binom{n}{k} p^k (1-p)^{n-k} = \binom{4}{0} (0.3)^0 (1-0.3)^{4-0} = \frac{4!}{0!(4-0)!} (0.3)^0 (0.7)^4$$

$$= 1 \cdot 1 \cdot (0.7)^4 = 0.2401$$

## Example 7

Let consider finding the probability than no more than one of the four people will develop a severe lung condition.

The events that "none of them develops a severe lung condition" and "exactly one develops a severe lung condition" are mutually exclusive.

$P\,(\text{none}) + P\,(\text{exactly one})$

$$= \binom{4}{0}(0.3)^0(1-0.3)^{4-0} + \binom{4}{1}(0.3)^1(1-0.3)^{4-1}$$

$$= \frac{4!}{0!(4-0)!}(0.3)^0(0.7)^4 + \frac{4!}{1!(4-1)!}(0.3)^1(0.7)^3$$

$$= \frac{4 \cdot 3 \cdot 2 \cdot 1}{1 \cdot 4 \cdot 3 \cdot 2 \cdot 1}(0.3)^0(0.7)^4 + \frac{4 \cdot 3 \cdot 2 \cdot 1}{1 \cdot 3 \cdot 2 \cdot 1}(0.3)^1(0.7)^3$$

$$= \frac{\not4 \cdot \not3 \cdot \not2 \cdot 1}{1 \cdot \not4 \cdot \not3 \cdot \not2 \cdot 1}(0.3)^0(0.7)^4 + \frac{4 \cdot \not3 \cdot \not2 \cdot 1}{1 \cdot \not3 \cdot \not2 \cdot 1}(0.3)^1(0.7)^3$$

$$= 0.2401 + 0.4116$$

$$= 0.6517 = 65.17\%$$

## Example 8

Lets consider finding the probability that at least two of the four people will develop a severe lung condition.

The complement of "at least two will develop a severe lung condition" is "no more than one will develop a severe lung condition."

We know from Example 7 that $P(\text{no more than one}) = 0.6517$.

So,
$$P(\text{at least two}) = 1 - P(\text{no more than one})$$
$$= 1 - 0.6517 = 0.3483 = 34.83\%$$

## Example 9

*Out of seven randomly selected smokers, how many would we expect to develop a severe lung condition?*

The mean of the binomial model is
$$\mu = np = 7 \cdot 0.3 = 2.1$$
On average, we would expect 2.1 of 7 randomly chosen smokers to develop a severe lung condition.

### Example 10

$$\binom{n}{0} = \frac{n!}{0!(n-0)!} = \frac{n!}{n!} = 1$$

### Example 11

$$\binom{n}{1} = \frac{n!}{1!(n-1)!} = \frac{n!}{(n-1)!} = \frac{n \cdot (n-1)!}{(n-1)!} = n$$

### Example 12

$$\binom{n}{n-1} = \frac{n!}{(n-1)!(n-(n-1))!} = \frac{n!}{(n-1)!1!} = \frac{n \cdot (n-1)!}{(n-1)!} = n$$

### Example 13

$$\binom{n}{n} = \frac{n!}{n!(n-n)!} = \frac{n!}{n!0!} = \frac{n!}{n!} = 1$$

### Example 14

Approximately 15% of the US population smokes cigarettes.

A local government believed their community has a lower smoker rate and commissioned a survey of 400 randomly selected individuals.

The survey found that only 42 of the 400 participants smoke cigarettes.
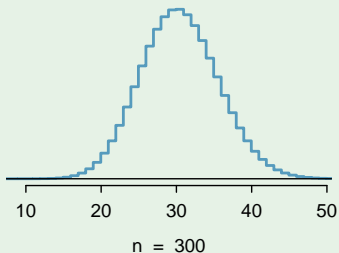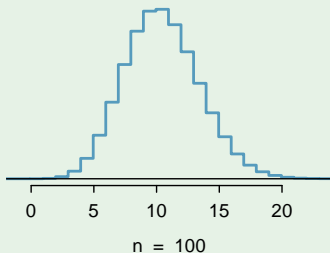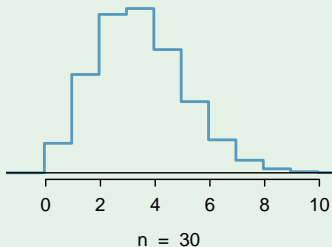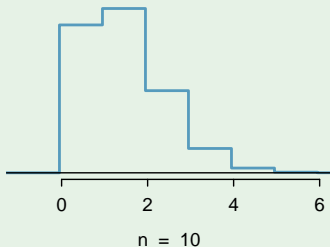
If the true proportion of smokers in the community was really 15%, what is the probability of observing 42 or fewer smokers in a sample of 400 people?

$$P(k = 0 \text{ or } k = 1 \text{ or } \cdots \text{ or } k = 42)$$
$$= P(k = 0) + P(k = 1) + \cdots + P(k = 42)$$
$$= 5.8558 \times 10^{-29} + 4.133335 \times 10^{-27} + \cdots + 0.001985$$
$$= 0.0054$$

### Note

When certain conditions are met, we can actually use the normal distribution to approximate the binomial distribution.

## Example 15



Histograms of samples from the binomial model when $p = 0.10$.

## Normal Approximation of the Binomial Distribution

The binomial distribution with probability of success *p* is approximately normal when the sample size *n* is sufficiently large so that:

$$np \geq 10 \qquad \text{and} \qquad n(1-p) \geq 10$$

The approximate normal distribution has parameters corresponding to the mean and standard deviation of the binomial distribution:

$$\mu = np \qquad\qquad \sigma = \sqrt{np(1-p)}$$

### Example 16

Let us check to see if the binomial distribution in Example 14 is approximately normal.

Recall that $p = 0.15$, $n = 400$, so we check:

$$np = 400 \cdot 0.15 = 60 \geq 10$$
$$n(1 - p) = 400 \cdot (1 - 0.15) = 400 \cdot 0.85 = 340 \geq 10$$

We may now use the normal distribution to approximate the binomial distribution for observing 42 or fewer smokers:

$$\mu = np = 60$$
$$\sigma = \sqrt{np(1 - p)} = 7.14143$$
$$z = \frac{x - \mu}{\sigma} = \frac{42 - 60}{7.14143} = -2.5205$$

Using technology gives:

$$P(z \leq -2.52) = 0.005859$$

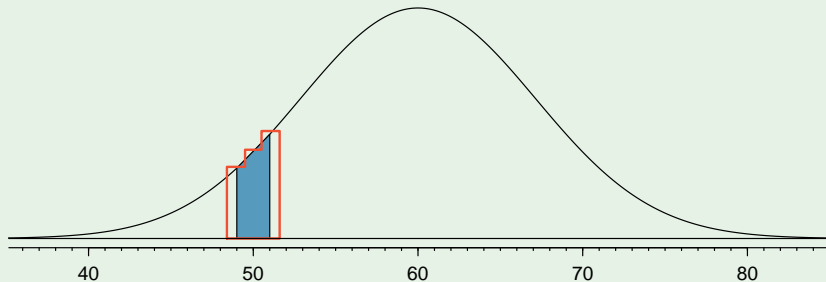This is very close to the value of 0.0054 we calculated in Example 14.

## Example 17

Suppose we want to compute the probability of observing 49, 50, or 51 smokers in 400 when $p = 0.15$.

It's tempting to apply the normal approximation here, as well. But,

Binomial: 0.0649          Normal: 0.0421



The area representing the binomial probability is outlined in red while the area representing the normal approximation is shaded in blue. Notice that the width of the normal approximation is too narrow.

### Improving the Normal Approximation

The normal approximation to the binomial distribution for intervals of values is usually improved if the cutoff values are modified slightly.

The cutoff values for the lower end of a shaded region should be reduced by 0.5.

The cutoff values for the upper end of a shaded region should be increased by 0.5.

### Example 18

For computing the probability of observing 49, 50, or 51 smokers in 400 when $p = 0.15$ we get the three values:

Binomial: 0.0649

Unmodified Normal: 0.0421

Modified Normal: 0.0633