Binomial Random Variables: Mean and Standard Deviation

Learning Objective: Fit the binomial model when appropriate, and use it to perform simple calculations.

Mean and Standard Deviation of the Binomial Random Variable

Now that we understand how to find probabilities associated with a random variable X which is binomial, using either its probability distribution formula or software, we are ready to talk about the mean and standard deviation of a binomial random variable. Let's start with an example:

Example: Blood Type B—Mean

Overall, the proportion of people with blood type B is 0.1. In other words, roughly 10% of the population has blood type B.

Suppose we sample 120 people at random. On average, how many would you expect to have blood type B?

The answer, 12, seems obvious; automatically, you'd multiply the number of people, 120, by the probability of blood type B, 0.1. This suggests the general formula for finding the mean of a binomial random variable:

Claim:

If X is binomial with parameters n and p, then

$$\mu_X = np$$

Although the formula for mean is quite intuitive, it is not at all obvious what the variance and standard deviation should be. It turns out that:

Claim:

If X is binomial with parameters n and p, then

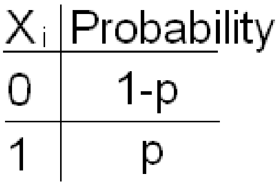
$$\sigma_X^2 = np\left(1-p\right); \sigma_X = \sqrt{np(1-p)}$$

For those who are interested, read below to see how these formulas were derived.

Many Students Wonder ...

Question: How are formulas for the mean and standard deviation of a binomial random variable derived?

Answer: For those who are interested, these formulas may be derived with the aid of the rule for variance of a sum of random variables. We think of our binomial X with n and p as the sum of n identical "Bernoulli" random variables X1 through Xn, each representing one of the n trials. Each of the Xi has just two possibilities: success (X = 1, with probability p), or failure (X = 0, with probability 1 - p), as shown in the probability distribution table below.



X_i	Probability
0	1-p
1	p

The mean of each Xi is 0(1 - p) + 1(p) = p, and the variance is

$$\sigma_i^2 = (0-p)^2(1-p) + (1-p)^2 p = p(1-p)[p+(1-p)] = p(1-p)$$

Because variance is additive, if we sum up n such random variables to form the binomial random variable X = X1 + ... + Xn, its variance will be the sum of n variances p(1 - p), or np(1 - p).

Comment

The binomial mean and variance are special cases of our general formulas for the mean and variance of any random variable.

$$\mu_X = x_1p_1 + x_2p_2 + \ldots + x_np_n = \sum_{i=1}^{n} x_ip_i$$

$$\sigma_X^2 = \left(x_1-\mu_X\right)^2 p_1 + \left(x_2-\mu_X\right)^2 p_2 + \ldots + \left(x_n-\mu_X\right)^2 p_n$$

$$= \sum_{i=1}^{n}\left(x_i-\mu_X\right)^2 p_i$$

Clearly it is much simpler to use the "shortcut" formulas

$\mu_X = np$ and $\sigma_X^2 = np\left(1-p\right); \sigma_X = \sqrt{np(1-p)}$ than it would be to calculate the mean and variance or standard deviation from scratch.

Example: Blood Type B—Standard Deviation

Suppose we sample 120 people at random. The number with blood type B should be about 12, give or take how many? In other words, what is the standard deviation of the number X who have blood type B?

Since n = 120 and p = .1,

$$\sigma_X^2 = 120\left(0.1\right)\left(1-0.1\right) = 10.8; \sigma_X = \sqrt{10.8} \approx 3.3$$

In a random sample of 120 people, we should expect there to be about 12 with blood type B, give or take about 3.3.

Scenario: Death Penalty Gallup Poll

A Gallup Poll in May 2004 estimated that roughly 70% of U.S. adults are in favor of the death penalty for a person convicted of murder. A random sample of 750 U.S. adults is chosen. Let X be the number of adults (out of 750) who favor the death penalty.

Did I Get This (1/1 point)

The distribution of X is binomial. What are the values of n and p?

Your Answer:

n=750, p=0.7

Our Answer:
Each sampled adult is a trial, so n = 750. In this case "success" (our outcome of interest) is "favoring the death penalty for convicted murderers." The Gallup Poll estimates that p = 0.7.

Resubmit Reset

Did I Get This

1/1 point (graded)
The mean and standard deviation of X are:

○ $\mu_X = 12.55$, $\sigma_X = 525$

● $\mu_X = 525$, $\sigma_X = 12.55$ ✓

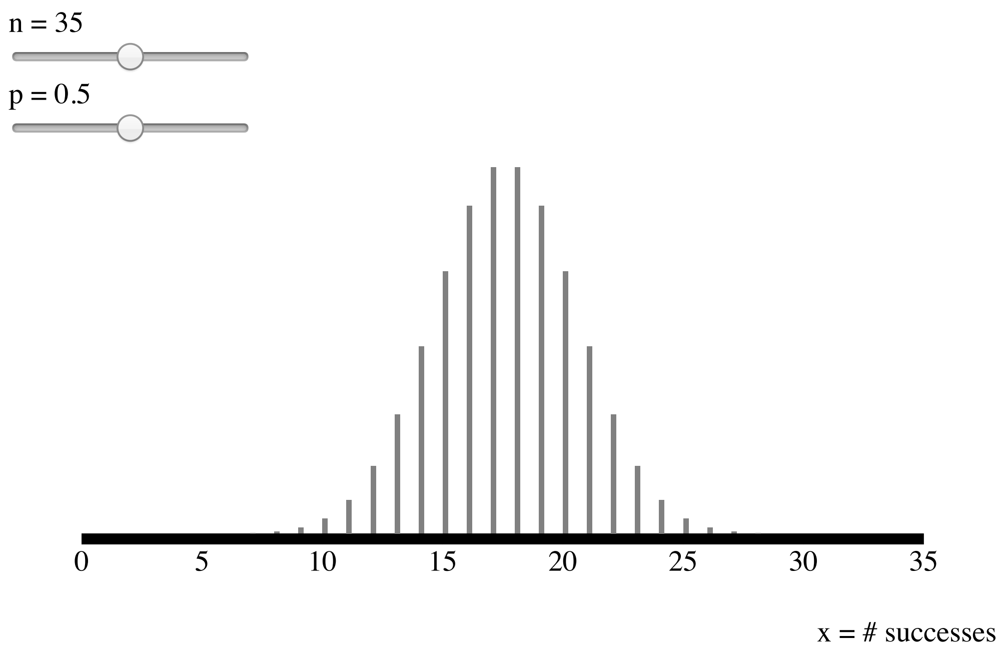
○ $\mu_X = 525$, $\sigma_X = 157.5$

Answer
Correct:
Indeed, since X is a binomial random variable, $\mu_X = np = 750(.7) = 525$. $\sigma^2_X = np(1 - p) = 750(.7)(.3) = 157.5$, and therefore $\sigma_X = \text{sqrt}(157.5) = 12.55$. (Note that "sqrt" stands for square root.)

Submit

Interactive Simulation

Before we move on to continuous random variables, let's investigate the shape of binomial distributions using the following simulation. Using the simulation below we will see that for different values of n and p, binomial distributions can be symmetric, skewed right or skewed left.



Learn By Doing (1/1 point)

The value of p should be 0.5 as the simulation launches. Move the slider for n from the far left (n = 10) to the far right (n = 60). What is the shape of the binomial distributions for p = 0.5 and the different values of n? Does the general shape change as n increases from 10 to 60?

Your Answer:

General shape doesn't change though it becomes thinner

Our Answer:

When p=0.5, the distribution is symmetric for any value of n from 10 to 60.

Resubmit **Reset**

Learn By Doing

1/1 point (graded)

Change the value of p to 0.2, and set n all the way to the left to n=10.

What is the shape of the distribution?

○ symmetric

● skewed right ✓

○ skewed left

Answer
Correct: The distribution is not symmetric. Most of it is on the left, with a long tail to the right.

Submit

Learn By Doing

1/1 point (graded)
Increase the value of n by moving the slider to the right.

As the value of n increases, what happens to the shape of the distribution?

● symmetric ✓

○ skewed right

○ skewed left

Answer
Correct:
The distribution shape becomes roughly symmetric when n is large with most of the data (high bars) in the middle and very little data (low bars) in the extremes.

Submit

Learn By Doing

1/1 point (graded)
Change the value of p to 0.8, and set n all the way to the left to n=10.

What is the shape of the distribution?

○ symmetric

○ skewed right

● skewed left ✓

Answer
Correct: The distribution is not symmetric. Most of it is on the right, with a tail to the left.

Submit

Learn By Doing

1/1 point (graded)
Increase the value of n by moving the slider to the right.

As the value of n increases, what happens to the shape of the distribution?

● symmetric ✓

○ skewed right

○ skewed left

Answer
Correct:
The distribution shape becomes roughly symmetric when n is large with most of the data (high bars) in the middle and very little data (low bars) in the extremes.

Submit