

Distribution of Sample Proportions

by Sophia Tutorial



WHAT'S COVERED

This tutorial will cover the distribution of sample proportions, which is called a sampling distribution. Our discussion breaks down as follows:

1. Sample Proportions
 - a. Mean
 - b. Standard Deviation
 - c. Shape

1. Sample Proportions

Many different situations can provide you with proportions.

➦ **EXAMPLE** Suppose that you were taking a poll during a political season, and you calculated the proportion of people that were going to vote for a particular candidate.

However, proportions like this are typically sample proportions. The only way to obtain the true population proportion, which is the parameter we're trying to estimate, is by taking a census. If you had some binomial question type-- meaning, are you going to vote for one candidate or the other--and you took a census, you would be able to know the parameter.

In most cases, you only deal with samples. You will want to figure out what the **distribution of sample proportions** actually looks like, which is the distribution of all possible sample proportions for a certain size, n .

➦ **EXAMPLE** Consider flipping a coin ten times. Obviously, you would expect 50% heads and 50% tails, however, it doesn't always work out exactly that way.

Suppose the first time you flipped ten coins, you got 6 heads, so a percentage of 60% heads.

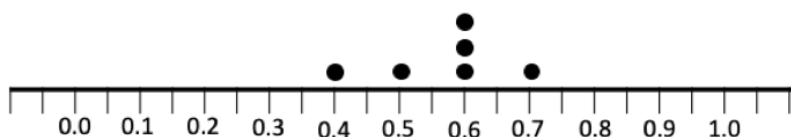


The next time you flipped ten coins, you got 70% heads. So it seems like the proportion of heads might change from trial to trial, or sample to sample rather. First time, you got 60% heads in your sample. The second time you got 70% heads in your sample. Suppose you do this a lot of times, and obtain sample

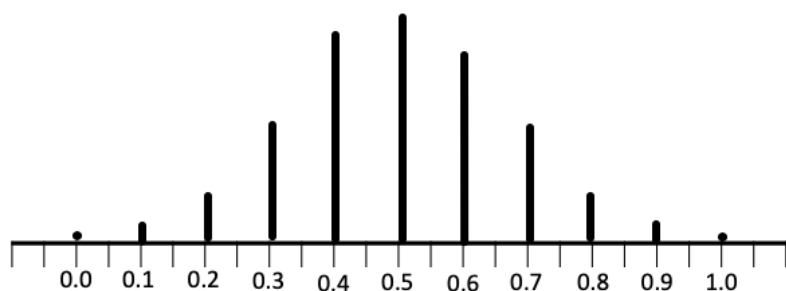
proportions of heads every time.

Sample	Sample Prop Heads
$S_1 = \text{HHHTHTTTHHT}$	$\hat{p}_1 = 0.6$
$S_2 = \text{HTHHHTHTHH}$	$\hat{p}_2 = 0.7$
$S_3 = \text{HHTHHHHTTT}$	$\hat{p}_3 = 0.6$
$S_4 = \text{TTHHTHTTHTH}$	$\hat{p}_4 = 0.4$
$S_5 = \text{TTTTHTHTHHH}$	$\hat{p}_5 = 0.5$
$S_6 = \text{HHHHHTTTTH}$	$\hat{p}_6 = 0.6$

Next, you can start to graph those sample proportions on a dot plot. Take the 0.6 and graph it, and then the 0.7, then the 0.6 again, stacking up the second dot on top of the first dot.



Repeat this process for every possible sample of size ten. Eventually, you would obtain a distribution that looks like this:



This is the distribution of the sample proportions of heads. This is what is called a sampling distribution of proportions.

TERM TO KNOW

Distribution of Sample Proportions

The distribution of all possible sample proportions for a certain size, n .

1a. Mean

For the scenario above, notice that it peaks at 0.5, exactly where you would expect. Also, notice that it sort of falls in almost a normal-looking shape off to each side. Very rarely did you get all of them being heads (a sample proportion of one) and very rarely did you get none of them being heads (a sample proportion of zero).

Notice the mean of the distribution of sample proportions, is the value of p , which is the actual probability of getting heads, which was 0.5. It centers around what the proportion, or probability, of heads is going to be for a single trial.



FORMULA

Mean of a Distribution of Sample Proportions

$$\mu_{\hat{p}} = p$$

1b. Standard Deviation

The number of successes is actually a binomial variable, meaning either you do it or you don't, and each trial is independent and all of the requirements for it being binomial are there. Since this is the case, when we graph the proportion of successes --which is the number of successes divided by the sample size, n -- the standard deviation will be the standard deviation of the binomial distribution, divided by n .

Therefore, the **standard deviation of a distribution of sample proportions** is the square root of n times p times q , divided by n . After some algebra, this simplifies to this square root of p times q over n . This is also known as the **standard error**.



FORMULA

Standard Deviation of a Distribution of Sample Proportions

$$\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}}$$



HINT

If p is the probability of success, q is the probability of failure, which is equal to $1-p$.



TERMS TO KNOW

Standard Deviation of a Distribution of Sample Proportions

A measure calculated by taking the square root of the quotient of $p(1-p)$ and n .

Standard Error

The standard deviation of the sampling distribution of sample proportions.

1c. Shape

For a distribution of sample proportions, we have discussed that the mean is equal to the probability of success and the standard deviation is the equal to the square root of p times q over n .

You're going to use the binomial numerator again to determine the shape. Since the sampling distribution of sample proportions is a binomial variable divided by a constant --that is, it's some number of successes

divided by n -- the rules for the shape of it are going to follow that of the binomial distribution.

That is, it's going to be skewed to the left when the value of p is high and the sample size is low. It's going to be skewed to the right when the probability of success is low and the sample size is low. Then, when the sample size is large, it will be approximately normal.

Again, how large is large? When n times p is at least ten and when n times q is at least ten, the distribution of sample proportions will be approximately normal, with the mean of p and the standard deviation of the square root of p times q over n .

This is going to be one of our conditions for inference if you're going to use normal calculations, which you'll want to do because they're easy to deal with. You're going to require that n times p is at least ten, and n times q is also at least ten.



BIG IDEA

A condition for inference with a distribution of sample proportions states that n times p is at least ten and n times q is at least ten.

$$n \cdot p \geq 10$$

AND

$$n \cdot q \geq 10$$



SUMMARY

You've learned about the distribution of sample proportions, the standard deviation of a distribution of sample proportions, and standard error, which is the same thing as the standard deviation of the sampling distribution. The sampling distribution of sample proportions has an approximately normal sampling distribution when the number of trials is large, referring to the shape. Its mean is the proportion of successes in the population--that's the center. In addition, the standard deviation of the sampling distribution, which is also called standard error, is the square root of the product of the probabilities of success and failure, divided by the number of trials. That's the spread.

Good luck!

Source: Adapted from Sophia tutorial by Jonathan Osters.



TERMS TO KNOW

Distribution of Sample Proportions

A distribution where each data point consists of a proportion of successes of a collected sample. For a given sample size, every possible sample proportion will be plotted in the distribution.

Standard Deviation of a Distribution of Sample Proportions

The square root of the product of the probabilities of success and failure (p and q , respectively) divided by the sample size.

Standard Error

The standard deviation of the sampling distribution of sample proportions.



FORMULAS TO KNOW

Mean of a Distribution of Sample Proportions

$$\mu_{\hat{p}} = p$$

Standard Deviation of a Distribution of Sample Proportions

$$\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}}$$