

7.4: The Central Limit Theorem for Proportions

The Central Limit Theorem tells us that the point estimate for the sample mean, \bar{x} , comes from a normal distribution of \bar{x} 's. This theoretical distribution is called the sampling distribution of \bar{x} 's. We now investigate the sampling distribution for another important parameter we wish to estimate; p from the binomial probability density function.

If the random variable is discrete, such as for categorical data, then the parameter we wish to estimate is the population proportion. This is, of course, the probability of drawing a success in any one random draw. Unlike the case just discussed for a continuous random variable where we did not know the population distribution of X 's, here we actually know the underlying probability density function for these data; it is the binomial distribution. The random variable is X = the number of successes and the parameter we wish to know is p , the probability of drawing a success which is of course the proportion of successes in the population. The question at issue is: from what distribution was the sample proportion, $p' = \frac{x}{n}$ drawn? The sample size is n and X is the number of successes found in that sample. This is a parallel question that was just answered by the Central Limit Theorem: from what distribution was the sample mean, \bar{x} , drawn? We saw that once we knew that the distribution was the Normal distribution then we were able to create confidence intervals for the population parameter, μ . We will also use this same information to test hypotheses about the population mean later. We wish now to be able to develop confidence intervals for the population parameter " p " from the binomial probability density function.

In order to find the distribution from which sample proportions come we need to develop the sampling distribution of sample proportions just as we did for sample means. So again imagine that we randomly sample say 50 people and ask them if they support the new school bond issue. From this we find a sample proportion, p' , and graph it on the axis of p 's. We do this again and again etc., etc. until we have the theoretical distribution of p 's. Some sample proportions will show high favorability toward the bond issue and others will show low favorability because random sampling will reflect the variation of views within the population. What we have done can be seen in Figure 7.4.1. The top panel is the population distributions of probabilities for each possible value of the random variable X . While we do not know what the specific distribution looks like because we do not know p , the population parameter, we do know that it must look something like this. In reality, we do not know either the mean or the standard deviation of this population distribution, the same difficulty we faced when analyzing the X 's previously.

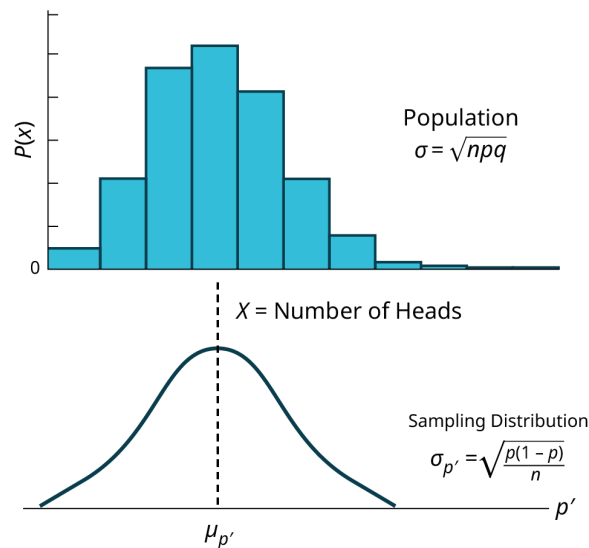


Figure 7.4.1

Figure 7.10 places the mean on the distribution of population probabilities as $\mu = np$ but of course we do not actually know the population mean because we do not know the population probability of success, p . Below the distribution of the population values is the sampling distribution of p 's. Again the Central Limit Theorem tells us that this distribution is normally distributed just like the case of the sampling distribution for \bar{x} 's. This sampling distribution also has a mean, the mean of the p' s, and a standard deviation, $\sigma_{p'}$.

Importantly, in the case of the analysis of the distribution of sample means, the Central Limit Theorem told us the expected value of the mean of the sample means in the sampling distribution, and the standard deviation of the sampling distribution. Again the Central Limit Theorem provides this information for the sampling distribution for proportions. The answers are:

1. The expected value of the mean of sampling distribution of sample proportions, $\mu_{p'}$, is the population proportion, p .
2. The standard deviation of the sampling distribution of sample proportions, $\sigma_{p'}$, is the population standard deviation divided by the square root of the sample size, n .

Both these conclusions are the same as we found for the sampling distribution for sample means. However in this case, because the mean and standard deviation of the binomial distribution both rely upon p , the formula for the standard deviation of the sampling distribution requires algebraic manipulation to be useful. We will take that up in the next chapter. The proof of these important conclusions from the Central Limit Theorem is provided below.

$$E(p') = E\left(\frac{x}{n}\right) = \left(\frac{1}{n}\right) E(x) = \left(\frac{1}{n}\right) np = p \quad (7.4.1)$$

(The expected value of X , $E(x)$, is simply the mean of the binomial distribution which we know to be np .)

$$\sigma_{p'}^2 = \text{Var}(p') = \text{Var}\left(\frac{x}{n}\right) = \frac{1}{n^2} (\text{Var}(x)) = \frac{1}{n^2} (np(1-p)) = \frac{p(1-p)}{n} \quad (7.4.2)$$

The standard deviation of the sampling distribution for proportions is thus:

$$\sigma_{p'} = \sqrt{\frac{p(1-p)}{n}} \quad (7.4.3)$$

Table 7.4.1

Parameter	Population distribution	Sample	Sampling distribution of p's
Mean	$\mu = np$	$p' = \frac{x}{n}$	p' and $E(p') = p$
Standard Deviation	$\sigma = \sqrt{npq}$		$\sigma(p) = \sqrt{\frac{p(1-p)}{n}}$

Table 7.4.1 summarizes these results and shows the relationship between the population, sample and sampling distribution. Notice the parallel between this Table and Table 7.1 for the case where the random variable is continuous and we were developing the sampling distribution for means.

Reviewing the formula for the standard deviation of the sampling distribution for proportions we see that as n increases the standard deviation decreases. This is the same observation we made for the standard deviation for the sampling distribution for means. Again, as the sample size increases, the point estimate for either μ or p is found to come from a distribution with a narrower and narrower distribution. We concluded that with a given level of probability, the range from which the point estimate comes is smaller as the sample size, n , increases. Figure 7.8 shows this result for the case of sample means. Simply substitute p' for \bar{x} and we can see the impact of the sample size on the estimate of the sample proportion.

7.4: The Central Limit Theorem for Proportions is shared under a [not declared](#) license and was authored, remixed, and/or curated by LibreTexts.