

## 5.3: Sampling Distribution of Sample Proportions

### Learning Objectives

- State the relationship between the sampling distribution of sample proportions(  $\hat{p}$  ) and a normal distribution.
- State the expected value (mean) and standard deviation of the sampling distribution of sample proportions.
- State the requirements for modeling the sampling distribution of sample proportions with a normal distribution.
- Apply the above to reasonably predict the proportion measures of various samples (all of the same size  $n$ ) from a population.

### Review and Preview

In regard to a random variable of a population, we have discussed the importance of understanding how various samples taken from the population produce different measures from each other as well as from the population's related measures. In the first section of this chapter, we saw that some statistical measures (such as samples' means, samples' variances, and samples' proportions for samples of a specific size  $n$ ) are considered unbiased since the various samples' statistics tend to crowd around the actual population's parameter. However, there are other statistical measures (such as samples' ranges, samples' standard deviation, and samples' medians) that do not behave this way, and hence are considered biased estimators.

Digging deeper in the last section, we have seen how the sample means from all possible samples are actually very predictable as a group. Under certain requirements, the sample means for samples of one specific size,  $n$ , act as a random variable themselves and that, although we can't predict what will happen with any one chosen simple random sample, the collection of all simple random samples' means form a normal distribution called a sampling distribution. Furthermore, this sampling distribution's mean value is the same as the population's mean value and the spread (standard deviation) in the sampling distribution is smaller than the standard deviation of the population. In notational form, we designated this with  $\mu_{\bar{x}} = \mu$  and  $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ .

Now we embark on a similar investigation of the sampling distribution's behavior for the proportion measure. Recall from [here](#) in Section 5.1 that we have seen an example of building a sampling distribution for a small population (our family of five) in which our "proportion" variable of interest was the proportion of the family members that wear glasses. In this small example, there were three of the five family members that wear glasses making our population's proportion measure  $p = \frac{3}{5}$ ; however, when we selected samples of size three from the population, the ten different samples produced various sample proportion measures ( $\hat{p}$ ), none of which were the same as the population's parameter measure of  $p = \frac{3}{5}$ . However, the sampling distribution of these various sample proportions--the probability distribution of sample proportions--had a distribution mean that did match the population's proportion. That is, we saw in this small example that  $\mu_{\hat{p}} = p$  even though  $p \neq \hat{p}$  for any of the actual samples of size three. So, as in the general sampling distribution of sample means, we wonder if the sampling distribution of sample proportions is just as predictable.

### The Sampling Distribution of Sample Proportions

First, we need to recognize that sample proportion measures fall into the realm of a binomial experiment with the number of trials being the sample size,  $n$ , and the probability of success,  $p$ , is the proportion of that population meeting the definition of "success" in the binomial experiment. Each time we select a member to be part of our sample, we are performing a binomial experiment. As a reminder from [Section 4.3](#), recall that the random variable,  $X$ , of a binomial experiment was the number of successes that could occur with a sample of size  $n$  taken from the population and that the possible values for the random variable were  $0, 1, 2, \dots, n$ . In general, it is possible to have a sample in which none (0) of the sample group met the "success" definition of the binomial experiment or a sample in which only 1 of the  $n$  was a "success," or that 2 of the  $n$ , ... , all the way to all  $n$  of the  $n$  were a success. We could build the binomial probability distribution from such information based on combination counts and our probability multiplication rule. We see this relationship established more formally below concerning a large population in which many, many samples of a size  $n$  are possible, or, in the case of a small finite population, with random sampling with replacement on samples of a size  $n$  being possible. Using our prior concepts of Section 4.3, we can build the binomial distribution concerning samples of size  $n = 3$  coming from a population in which the population proportion measure of success is  $p = \frac{3}{10} = 30\%$ . Hence, the population proportion measure of failure is  $q = \frac{7}{10} = 70\%$ . Using our binomial distribution approach of Section 4.3, we produce the following distribution table.

Table 5.3.1: Binomial probability distribution

Number of Successes in $n$ trials: $X$	Probability $P(x) = P(\hat{p})$
0	${}_3C_0 \cdot \left(\frac{3}{10}\right)^0 \cdot \left(\frac{7}{10}\right)^3 = 1 \cdot (1) \cdot \left(\frac{343}{1000}\right) = \frac{343}{1000} = 34.3\%$
1	${}_3C_1 \cdot \left(\frac{3}{10}\right)^1 \cdot \left(\frac{7}{10}\right)^2 = 3 \cdot \left(\frac{3}{10}\right) \cdot \left(\frac{49}{100}\right) = \frac{441}{1000} = 44.1\%$
2	${}_3C_2 \cdot \left(\frac{3}{10}\right)^2 \cdot \left(\frac{7}{10}\right)^1 = 3 \cdot \left(\frac{9}{100}\right) \cdot \left(\frac{7}{10}\right) = \frac{189}{1000} = 18.9\%$
3	${}_3C_3 \cdot \left(\frac{3}{10}\right)^3 \cdot \left(\frac{7}{10}\right)^0 = 1 \cdot \left(\frac{27}{1000}\right) \cdot (1) = \frac{27}{1000} = 2.7\%$

We also recall that the binomial probability distribution was found to have an expected value or mean of  $\mu = n \cdot p = 3 \cdot 0.30 = 0.90$  and variance of  $\sigma^2 = n \cdot p \cdot q = 3 \cdot 0.30 \cdot 0.70 = 0.63$ , and a standard deviation of  $\sigma = \sqrt{n \cdot p \cdot q} = \sqrt{0.63} \approx 0.793725$ . As the binomial distribution demonstrates and

our past work confirms, not all samples of size  $n$  taken from a population will have the same number of success  $x$ ; and the related proportion of successes  $\frac{x}{n}$  will tend to vary from sample to sample. We saw this in our small family member example mentioned in Section 5.1.

We now connect the sample proportion measures to this binomial distribution. Notice that our random variable  $X$  on the number of successes can be transformed into sample proportion measures simply by dividing by the sample size. For example, suppose we have a sample of size  $n = 3$  with  $x = 2$  successes. In that case, the sample's proportion measure of success is  $\hat{p} = \frac{2}{3} \approx 66.7\%$ . We can do this for each possible value of our random variable  $X$ , producing the following distribution table.

Table 5.3.2: Binomial probability distribution

Number of Successes in $n$ trials: $X$	Proportion of Success $\hat{p} = \frac{x}{n}$	Probability $P(x) = P(\hat{p})$
0	$\frac{0}{3} = 0.00 = 0\%$	34.3%
1	$\frac{1}{3} \approx 0.333.00 = 33.3\%$	44.1%
2	$\frac{2}{3} \approx 0.667 = 66.7\%$	18.9%
3	$\frac{3}{3} = 1.00 = 100\%$	2.7%

This  $\hat{p}$ -probability distribution is the sampling distribution, and below is a graphic of that binomial distribution and its related sampling distribution of sample proportions.

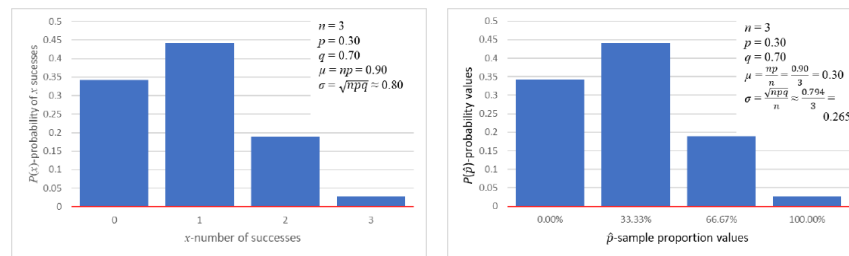


Figure 5.3.1: Binomial distribution (left) transformed into the sampling distribution of sample proportions (right)

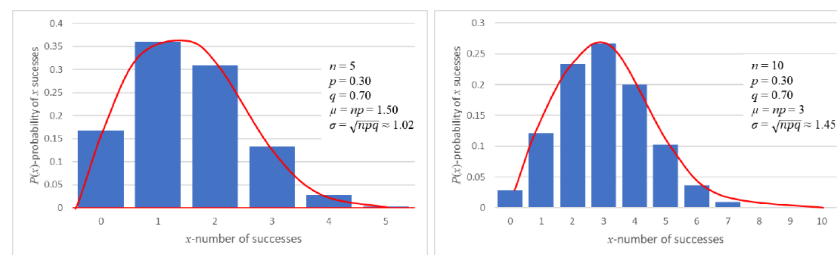
We should recognize that the only change occurring when moving from the binomial distribution to the sampling distribution is a rescaling of the horizontal axis, which results in a rescaling of the mean,  $\mu$ , and the standard deviation,  $\sigma$ , both caused by our division by the sample size  $n = 3$ .

Of course, with this sampling distribution table or with its graph, we understand what values occur for the sample proportions concerning all the various simple random samples of size  $n = 3$ . For example, we may wish to know how probable it is to find a random sample of size 3 from this population in which the sample proportion is below 50%...that is to find  $P(\hat{p} < 0.50)$ . From our distribution, we can determine that

$$P(\hat{p} < 0.50) = P(\hat{p} = 0.00) + P(\hat{p} = 0.33) \approx 34.3\% + 44.1\% = 78.4\%.$$

That is, over 75% of random samples of size 3 from this population will produce sample proportion  $\hat{p}$  measures below 50%.

The process from above is how we build the sampling distribution of sample proportions with **small** sample sizes, as the binomial distributions only have a few possible outcomes for the random variable "a number of successes." However, these binomial distribution tables get very large and cumbersome if working with large samples. For example, if dealing with samples of size  $n = 100$ , we would need to build a table with 101 rows with sample proportion measures from 0%, 1%, 2%, ..., 100%. We can quickly see how even larger but often used sample sizes such as 2000 could be difficult to work with. To find a way around this, we continue our theory building in which the population proportion is  $p = 0.30$ , and then work with various samples of size 5, 10, 25, and 50. Using the same approach as above, we can produce the binomial distribution tables (not shown), and then the graphics for each of those distribution tables, shown below in Figure 5.3.2. Notice how the binomial distributions become more bell-shaped and symmetrical as the sample size gets larger, specifically in the distributions for the largest two sample sizes of  $n = 25$  and  $n = 50$ .



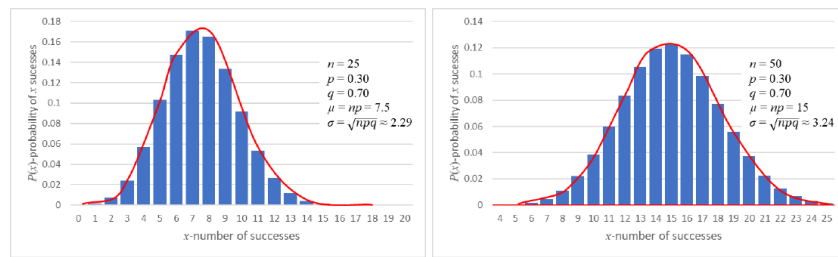


Figure 5.3.2: Binomial distributions approaching a normal distribution

This same behavior tends to occur regardless of the actual population proportion value,  $p$ , provided the sample size,  $n$ , is sufficiently large. Without delving too deeply into the underlying mathematical reasoning, the binomial distribution can be considered an approximately normal distribution in behavior provided  $n \cdot p > 5$  and  $n \cdot q > 5$ . Statisticians wishing to be even more conservative to achieve greater accuracy in the use of a normal distribution to approximate a binomial distribution will often require  $n \cdot p > 10$  and  $n \cdot q > 10$ . The above also demonstrates that as  $n$  increases, the normal distribution approximation becomes a better and better fit for the binomial distribution. We also have the mean and standard deviation of this approximating normal distribution due to our knowledge of the binomial distribution; that is, our normal distribution approximation to the binomial distribution will also have a mean of  $\mu = n \cdot p$  and standard deviation of  $\sigma = \sqrt{n \cdot p \cdot q}$ .

Now let us examine how this is related to the "sample proportion" random variable instead of the "number of success" random variable. As in Figure 5.3.1, we can adjust to "proportion" measures instead of "number of success" in the distributions by dividing each of our random variable's  $x$ -values by the sample size  $n$ . As can be seen below in Figure 5.3.3 this change to the proportion variable only causes a change in the scaling of the  $x$ -axis and measures related to that axis, but does not change the distribution probability measures nor the basic shape of the distribution.

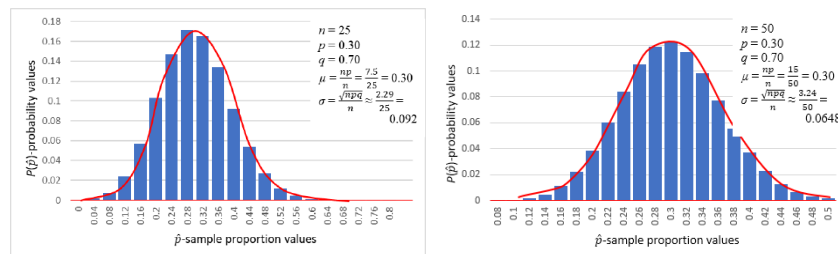


Figure 5.3.3: Binomial Distributions rescaled to Proportion Distributions

Therefore, we note that under sufficient requirements, our sampling distribution for the proportion values will be approximately a bell-shaped distribution with key measures of  $\mu_{\hat{p}} = \frac{n \cdot p}{n} = p$  and  $\sigma_{\hat{p}} = \frac{\sqrt{n \cdot p \cdot q}}{n} = \sqrt{\frac{n \cdot p \cdot q}{n^2}} = \sqrt{\frac{p \cdot q}{n}}$ . We can then use a normal probability distribution to estimate the binomial probability values over intervals; a normal probability distribution with appropriate probability density function with the same mean and standard deviation as we found above. Similar to the Central Limit Theorem (CLT) for predicting the distribution of all possible sample means in a specific situation, we have a theorem for predicting the distribution of all possible sample proportions within a particular situation.

### Sampling Distribution of Sample Proportions

Given a binomial situation within a population of interest in which the following conditions are known:

- the requirements for a binomial distribution are met with
  - the population proportion of interest (probability of success) is  $p$
  - the complement proportion (probability of failure) is  $q = 1 - p$
  - the sample size (number of finite trials) is  $n$
- the requirements of  $n \cdot p > 5$  and  $n \cdot q > 5$  are met

then the sampling distribution of all possible sample proportions can be approximated as a normal random variable with  $\mu_{\hat{p}} = p$  and  $\sigma_{\hat{p}} = \sqrt{\frac{p \cdot q}{n}}$ .

If desiring more reliable measures in using the normal distribution to approximate the distribution of sample proportions, we instead use the more conservative requirements of  $n \cdot p > 10$  and  $n \cdot q > 10$ . If the values of  $n \cdot p$  and  $n \cdot q$  do not surpass at least 5, then we do not approximate with a normal distribution but instead use the binomial probability distribution adjusted to sample proportions to model the sampling distribution.

### Applying the Theorem on Sampling Distribution of Sample Proportions

With a specific population of interest, our theorem allows us to understand which sample proportions are likely to happen and which are unlikely. This knowledge is important for understanding how we can have confidence in predicting a population's proportion from a single sample as we turn to inferential statistics in the next chapter. So, to prepare for this, we apply this theorem with the following text exercises.

? Text Exercise 5.3.1

It is believed that 21% of all U.S. female adults are over 66 inches in height. Determine the probability of selecting a simple random sample of U.S. female adults in which over 30% of the sample group is over 66 inches tall for each sample size given below. What do you notice about the probabilities as  $n$  increases?

1.  $n = 25$

**Answer**

We first note that this can be considered a "binomial" experiment in which we are defining "success" as a U.S. female adult having a height measure over 66 inches. The population's proportion is  $p = 21\% = 0.21$  and the "failure" proportion is  $q = 1 - 0.21 = 0.79$ . Instead of building a binomial distribution to answer the question, we can answer this question using our above theorem since  $n \cdot p = 25 \cdot 0.21 = 5.25$  and  $n \cdot q = 25 \cdot 0.79 = 19.75$  are both values above 5. (We also note that 5.25 is not much above 5 and right on the border of meeting the requirements of the theorem; in general, when getting close in value to the requirements, we understand that our measures are not as reliable and should not be used for highly important or costly decision making.) Based on our developed theory above, the sampling distribution of sample proportions is approximately normal with a mean  $\mu_{\hat{p}} = p = 0.21$  and a standard deviation  $\sigma_{\hat{p}} = \sqrt{\frac{p \cdot q}{n}} = \sqrt{\frac{0.21 \cdot 0.79}{25}} \approx 0.0815$ . Sketching a graphic of this normal distribution, we see the following.

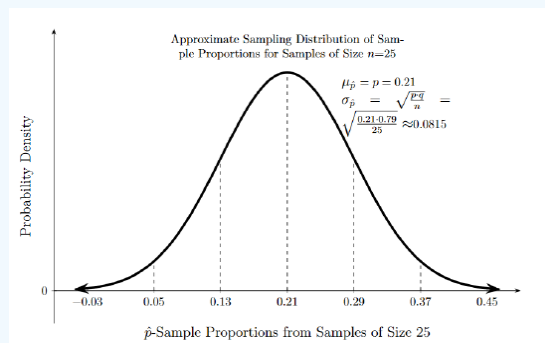


Figure 5.3.4 Approximate sampling distribution of sample proportions

We can compute the approximate probability for randomly selecting a sample of size 25 in which the proportion measure from that sample is larger than 30%; that is, we can find approximately from our normal distribution the value of  $P(\hat{p} > 30\%)$  by finding the shaded region displayed below in our related sampling distribution.

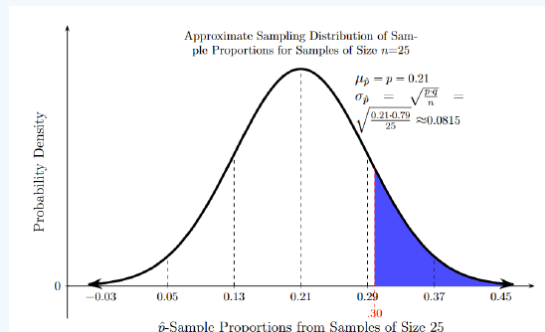


Figure 5.3.5 Approximate sampling distribution of sample proportions

Using our spreadsheet's NORM.DIST function concerning the normal distribution above, we produce:

$$P(\hat{p} > 30\%) = 1 - \text{NORM.DIST}(0.30, 0.21, \sqrt{\frac{0.21 \cdot 0.79}{25}}, 1) \approx 13.462\%$$

About 13.462% of all possible samples of size  $n = 25$  from the population of U.S. female adults will have over 30% of the women in the sample being over 66 inches tall. Stated equivalently, the probability of randomly selecting 25 U.S. female adults in which over 30% of the women are over 66 inches tall is about 13.462%. Although not highly likely, such a sample result would generally not be considered unusual.

2.  $n = 50$

**Answer**

The problem setup remains the same; we update for the sample size of 50, noting that we still meet our requirements with  $n \cdot p = 50 \cdot 0.21 = 10.5$  and  $n \cdot q = 50 \cdot 0.79 = 39.5$  both above 10. In our graphic below, we point out for emphasis the scaling change that occurred in the

horizontal axis as compared to part 1. of this text exercise above.

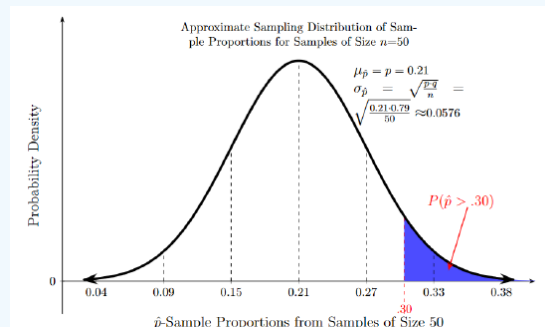


Figure 5.3.6 Approximate sampling distribution of sample proportions

$$P(\hat{p} > 30\%) = 1 - \text{NORM.DIST}(0.30, 0.21, \sqrt{\frac{0.21 \cdot 0.79}{50}}, 1) \approx 5.9092\%$$

The probability of randomly selecting 50 U.S. female adults in which over 30% of the women samples are over 66 inches tall is about 5.9092%. We note that this is a less likely outcome as compared to such in samples of size 25.

3.  $n = 100$

**Answer**

The problem setup remains the same; we update the sample size to 100 and note that we still meet our requirements with  $n \cdot p = 100 \cdot 0.21 = 21.0$  and  $n \cdot q = 100 \cdot 0.79 = 79.0$  both well above 10.

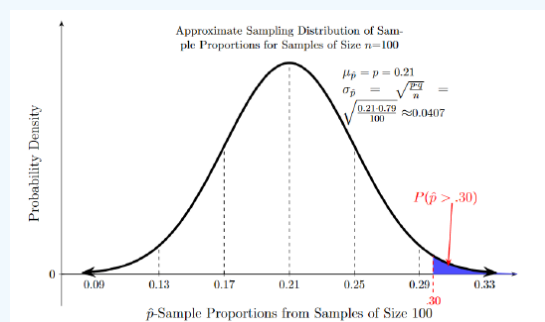


Figure 5.3.7 Approximate sampling distribution of sample proportions

$$P(\hat{p} > 30\%) = 1 - \text{NORM.DIST}(0.30, 0.21, \sqrt{\frac{0.21 \cdot 0.79}{100}}, 1) \approx 1.3565\%$$

4.  $n = 250$

**Answer**

We again update to the sample size of 250 and note that we easily meet our restrictive requirements with  $n \cdot p = 250 \cdot 0.21 = 52.5$  and  $n \cdot q = 250 \cdot 0.79 = 197.5$ .

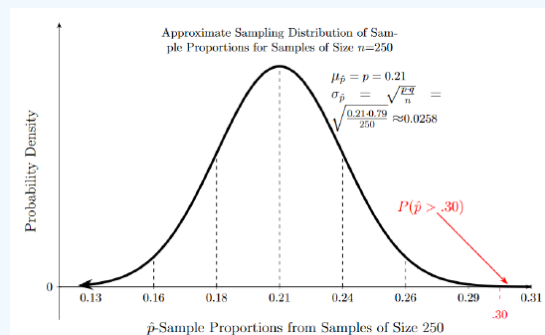


Figure 5.3.8 Approximate sampling distribution of sample proportions

$$P(\hat{p} > 30\%) = 1 - \text{NORM.DIST}(0.30, 0.21, \sqrt{\frac{0.21 \cdot 0.79}{250}}, 1) \approx 0.0238\%$$

We note in this last case that it is extremely unlikely to randomly select a sample of 250 U.S. adult women in which over 30% of the sample group are over 66 inches tall.

Finally, looking across all four exercises, we notice that as the sample size increases, the standard deviation of the sampling distribution decreases. In paying attention to the horizontal axis scale as it changes through these exercises, if the sample size,  $n$ , were to increase close to the size of the population, we would see almost a 100% chance of the various sample proportions being very very close to the population's proportion of 21%. That is, the larger  $n$  is, the various possible random sample proportions will usually be very close to the population's proportion. In the next chapter, we will develop more specific measures for the vague term "close."

### ? Text Exercise 5.3.2

In Kansas, 35% of adults over 25 years old have a bachelor's degree or higher.

1. Determine the probability of randomly selecting fifty Kansas adults over 25 years old in which less than 20% of the sample group have a bachelor's degree or higher.

#### Answer

We are imagining randomly selecting fifty individuals from the population of Kansas adults over 25 years of age and determining the sample proportion ( $\hat{p}$ ) of those selected who have a bachelor's degree or higher (hence a binomial situation). We should understand at this point that different samples will produce different  $\hat{p}$  values, and in using random sampling, we do not know which sample we will get. However, we are interested in the following probability:  $P(\hat{p} < 20\%)$ . So, we must turn to the sampling distribution of sample proportions to answer the probability question. Utilizing our developed theory and the given information, we note that  $n = 50$ ,  $p = 35\% = 0.35$ , and  $q = 65\% = 0.65$ . Since  $n \cdot p = 50 \cdot 0.35 = 17.5$  and  $n \cdot q = 50 \cdot 0.65 = 32.5$  are both larger than 5, we can reasonably approximate the sampling distribution of the various possible  $\hat{p}$  values as a normal distribution with a mean  $\mu_{\hat{p}} = p = 0.35$  and a standard deviation  $\sigma_{\hat{p}} = \sqrt{\frac{p \cdot q}{n}} = \sqrt{\frac{0.35 \cdot 0.65}{50}}$ . Sketching a graphic of this described sampling distribution, we see the following.

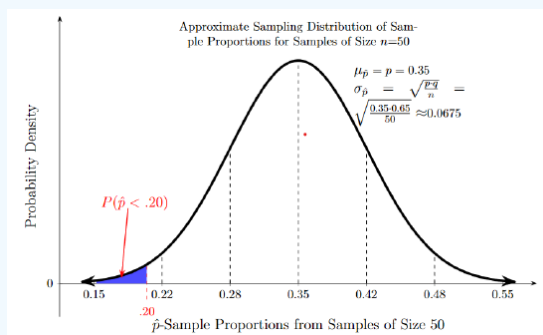


Figure 5.3.9 Approximate sampling distribution of sample proportions

Using our spreadsheet to compute the area/probability measure highlighted, we have:

$$P(\hat{p} < 0.20) = \text{NORM.DIST}(0.20, 0.35, \sqrt{\frac{0.35 \cdot 0.65}{50}}, 1) \approx 1.3083\%$$

We note that randomly selecting such a sample is not very likely, though also not impossible.

2. Determine the probability of randomly selecting eight hundred Kansas adults over 25 years old in which the sample's proportion will be within 2% of the actual population's proportion of 35%. That is, what proportion of samples of size eight hundred from this population of interest will produce a  $\hat{p}$  measure between 33% and 37%?

#### Answer

We are again working in the same basic situation context, only with a larger sample size of 800. Again, turning to our developed theory we first note that  $n \cdot p = 800 \cdot 0.35 = 280$  and  $n \cdot q = 800 \cdot 0.65 = 520$  are both larger than 5. Therefore, we can reasonably approximate the sampling distribution of the possible  $\hat{p}$  values as a normal distribution with a mean  $\mu_{\hat{p}} = p = 0.35$  and a standard deviation  $\sigma_{\hat{p}} = \sqrt{\frac{p \cdot q}{n}} = \sqrt{\frac{0.35 \cdot 0.65}{800}}$ . Sketching a graphic of this described sampling distribution, we see the following.

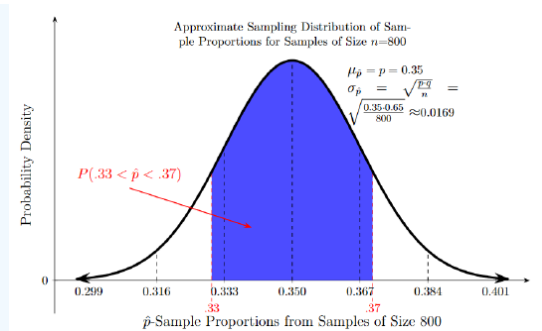


Figure 5.3.10 Approximate sampling distribution of sample proportions

We are interested in the proportion of samples in which the  $\hat{p}$  values are within 2% = 0.02 of the population's measure of 35%. Thus, we need the area/probability measure in our distribution between  $\hat{p}$  scale measures of 33% = 0.33 and 37% = 0.37. Using our technology to compute our area:

$$P(0.33 < \hat{p} < 0.37) \approx \text{NORM.DIST}(0.37, 0.35, \sqrt{\frac{0.35 \cdot 0.65}{800}}, 1) - \text{NORM.DIST}(0.33, 0.35, \sqrt{\frac{0.35 \cdot 0.65}{800}}, 1) \approx 0.882189 - 0.117811 \approx 76.4378\%$$

Around three-quarters of the samples of size 800 from the population of Kansas adults over 25 years old will produce sample proportion measures within 2 percentage points of the actual population proportion of 35%. Understanding such results gives us some confidence in using a single sample's measure from a sample of size 800 as a close approximation to what is happening in the population. We realize some samples will not meet this condition, but most will.

3. Regarding this situation involving Kansas adults over age 25, determine the probability of randomly selecting twelve Kansas adults over 25 years old in which at most 25% of the sample group have a bachelor's degree or higher. That is, determine  $P(\hat{p} \leq 0.25)$ .

#### Answer

We are again working in the same basic context as question 1. and 2. above, but we should also notice we are working with a somewhat small sample size. So checking our theory requirements, we first note that  $n = 12$ ,  $p = 35\% = 0.35$ , and  $q = 65\% = 0.65$ , and thus our important requirement measures are  $n \cdot p = 12 \cdot 0.35 = 4.2$  and  $n \cdot q = 12 \cdot 0.65 = 7.8$ . Since both are not larger than 5, we should NOT use the normal distribution for approximating the binomial probability distribution; we must go back to our discrete values table approach of the binomial probability distribution discussed in Section 4.3 of this text, with the needed adjustment to sample proportion as the random variable.

Building this table as per Section 4.3 concepts (for efficiency we use the BINOM.DIST function in a spreadsheet to produce the probability measures), and then converting to proportion measures on number of success as discussed above in this section, we produce the following distribution table:

Table 5.3.3 Binomial to sample proportion probability distribution

Number of Successes in $n$ trials: $X$	Proportion of Success $\hat{p} = \frac{x}{n}$	Probability $P(x) = P(\hat{p})$
0	$\frac{0}{12} = 0.00 = 0\%$	0.5688%
1	$\frac{1}{12} \approx 0.08333 = 8.333\%$	3.6753%
2	$\frac{2}{12} \approx 0.16667 = 16.667\%$	10.8846%
3	$\frac{3}{12} = 25.000\%$	19.5365%
4	$\frac{4}{12} \approx 33.333\%$	23.6692%
5	$\frac{5}{12} \approx 41.667\%$	20.3920%
6	$\frac{6}{12} = 50.000\%$	12.8103%
7	$\frac{7}{12} \approx 58.333\%$	5.9125%
8	$\frac{8}{12} \approx 66.667\%$	1.9898%
9	$\frac{9}{12} = 75.000\%$	0.4762%
10	$\frac{10}{12} \approx 83.333\%$	0.0769%
11	$\frac{11}{12} \approx 91.667\%$	0.00753%
12	$\frac{12}{12} = 100.000\%$	0.000338%

This table is a representation of the  $\hat{p}$ -sampling distribution. We see that the sample proportion measure of 25% for samples of size 12 occurs when the binomial random variable is  $x = 3$ . Thus, by adding all associated probability measures when  $\hat{p} \leq 0.25$ , we find

$$P(\hat{p} \leq 0.25) = P(x \leq 3) \approx 0.346653 \approx 34.67\%.$$

Thus, about 34.67% of all the various possible samples of size 12 from the population of Kansas adults over 25 years old will produce a sample proportion value (representing the proportion of those with a bachelor's degree or higher) of at most 25%.

To see why this "check of requirements" was so important, we notice that if we had instead incorrectly used a normal distribution in this situation, we would have computed

$$P(\hat{p} \leq 0.25) \approx \text{NORM.DIST}\left(0.25, 0.35, \sqrt{\frac{0.35 \cdot 0.65}{12}}, 1\right) \approx 0.2333836 \approx 23.338\%$$

which is a significantly poor approximation value to the actual probability measure of the sampling distribution for proportions computed above of 34.67%.

4. Regarding this situation involving Kansas adults over age 25, determine the interval of sample proportions  $\hat{p}$ , which captures the central 95% of all possible proportion values from samples of size eight hundred. That is, determine  $\hat{p}$ -values  $a$  and  $b$  such that  $P(a < \hat{p} < b) = 0.95$ .

#### Answer

We are again working in the same basic situation as question 2. above, so we again go to the normal distribution modeling the sampling distribution of sample proportions: a normal distribution with a mean  $\mu_{\hat{p}} = p = 0.35$  and a standard deviation  $\sigma_{\hat{p}} = \sqrt{\frac{p \cdot q}{n}} = \sqrt{\frac{0.35 \cdot 0.65}{800}}$ . However, this time we have a central area/probability region of 95% and are looking for the boundary values on the  $\hat{p}$ -axis that captures this amount of area. Sketching a graphic of this described sampling distribution, we see the following:

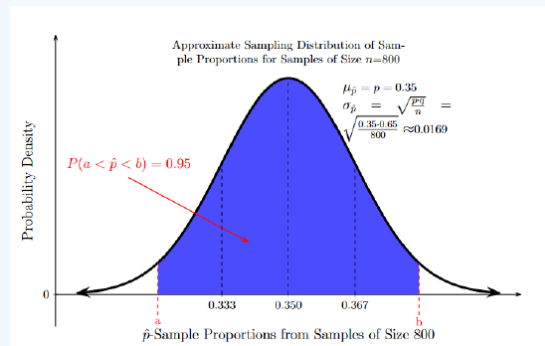


Figure 5.3.11: Approximate sampling distribution of sample proportions

Reminding ourselves that we can find horizontal axis scale values in normal distributions tied to left area measures using our spreadsheet's NORM.INV function, we compute:

$$a = \text{NORM.INV}(0.025, 0.35, \sqrt{\frac{0.35 \cdot 0.65}{800}}) \approx 0.3169 = 31.69\%,$$

$$b = \text{NORM.INV}(0.975, 0.35, \sqrt{\frac{0.35 \cdot 0.65}{800}}) \approx 0.3831 = 38.31\%,$$

Thus,  $P(0.3169 < \hat{p} < 0.3831) \approx 95\%$  or, stated in words, about 95% of all the various possible samples of size 800 from the population of Kansas adults over 25 years old will produce a sample proportion value (representing the proportion of those with a bachelor's degree or higher) between 31.69% and 38.31%.

We notice another implication from our work. Our boundary values of 31.69% and 38.31% each deviate from the population proportion value  $p = 35\%$  by  $|a - \mu_{\hat{p}}| = |b - \mu_{\hat{p}}| = |31.69\% - 35\%| = |38.31\% - 35\%| = 3.31\%$ . This tells us that about 95% of random samples in this situation will produce a sample proportion measure  $\hat{p}$  that is different from the population's proportion measure  $p$  by no more than 0.0331 = 3.31%. So most samples' proportions from samples of size 800 in this population are relatively close in value to the population's proportion, and only about 5% of samples will deviate from the population's proportion by more than 3.31%.

In summary, as long as certain requirements are met, we can often use normal distributions to analyze sampling distributions of sample proportions and understand how varied sample proportions can be within a specific binomial situation. As the Text Exercise 5.3.2.4 demonstrated, this will enable us to understand how we can infer, with some confidence, a population's proportion from a single random sample.

5.3: Sampling Distribution of Sample Proportions is shared under a [Public Domain](#) license and was authored, remixed, and/or curated by The Math Department at Fort Hays State University.

- 9.8: Sampling Distribution of  $p$  by David Lane is licensed Public Domain. Original source: <https://onlinestatbook.com>.