

6.2: Confidence Intervals for Proportions

Learning Objectives

- Recognize that a proportion of a random sample proportion is an estimation of the related proportion measure of the population
- Develop and apply the margin of error measure for using the proportion of a sample proportion to estimate the proportion of the population
- Develop and apply the confidence interval measures for using the proportion of a sample to estimate the proportion of the population
- Develop and apply sample size measures to control margin of error

Review and Preview

As stated numerous times before, an important area of inferential statistics is the ability to use a single measure from a sample to predict the related measure for the entire population (such as using the mean of a sample to predict the mean of the population or using the proportion from a sample to predict the proportion of the population.) In the previous Section 6.1, we discussed the general concepts of margin of error, confidence intervals, confidence levels, and α value; all of which are important measures of inferential statistics. We now focus on the specific situation of using a proportion measure from a random sample to predict a proportion measure from the population.

To further review, we remind ourselves of [Section 5.3](#) and the sampling distribution of sample proportions where we noted that different samples of a specific chosen size, n , produce a collection of various sample proportion measures \hat{p} . In our past investigations, most if not all of the various sample proportion measures were not the same value as the population's proportion, that is $\hat{p} \neq p$. It was also important for us to recognize that in the large collection of various \hat{p} values, that under certain restrictions, the distribution of \hat{p} values formed an approximately normal distribution. (The restrictions required that $n \cdot p > 5$ and $n \cdot q > 5$, both of which tend to be easily met if working with large sample sizes.) Furthermore, this sampling distribution's mean value will be the same as the population's proportion 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 $\mu_{\hat{p}} = p$ and $\sigma_{\hat{p}} = \sqrt{\frac{p \cdot q}{n}}$.

As one final review note, we re-examine the third part of [Text Exercise \(5.3.2\)](#). In that exercise, we found the central interval in the sampling distribution that contained 95% of possible sample proportion results. That is, we found within the given context how far away (the margin of error) from the population proportion's value 95% of the various samples' proportions would be.

Now we use these previous findings to develop a routine method for building a confidence interval in the proportion measure situation.

Sampling Distribution of Sample Proportions and Confidence Intervals

Let us begin in a specific context to help frame our work. Suppose that we are interested in predicting the proportion of the U.S. adult population which has received the latest flu vaccine. Naturally, we would not be able to ask every U.S. adult due to the population size and likely limited resources/finances to collect such data. However, it would be reasonable for us to randomly contact 1,000 such adults in the United States and determine which of those had and which had not taken the latest vaccine. Suppose 735 of those had received the vaccine; then this one collected sample had a proportion measure of $\hat{p} = \frac{735}{1000} = 0.735 = 73.5\%$. Naturally, we do not claim the population's proportion, p , is the same value, but our work with sampling distributions should convince us that we can expect the population's measure to be reasonably close to this sample's measure. This predictable sampling distribution of sample proportions allows us to consider a random sample's proportion, \hat{p} , to be a valid **point estimate** of the population's proportion; after all, the sampling distribution shows that most of the time a random sample's proportion will be "close" to the population's proportion measure. However, we need to have a measure for "close".

Due to the predictable sampling distribution of sample proportions (shown in Figure 6.2.1 below), we will determine a measure of "close" by choosing a confidence level (CL) value, such as 95%. Our measure of "close" will be a calculated margin of error measure designated as ME. Recall that the distribution below shows that most random samples (in fact the percentage given by our choice of CL) will produce \hat{p} values that fall within the ME distance of the actual population's proportion p which is at the center of the sampling distribution. As long as we choose a large CL value, we have a very good chance that our one collected random

sample's proportion will fall on the horizontal axis scale under the blue region. (Yes, it is possible that our random sample's \hat{p} will not be within this group, but the probability of such an outcome is only $\alpha = 100\% - \text{CL}$ —the α value discussed in Section 6.1. Once more, we can see that if we keep our choice of CL close to 100%, then α will be small: close to 0%.

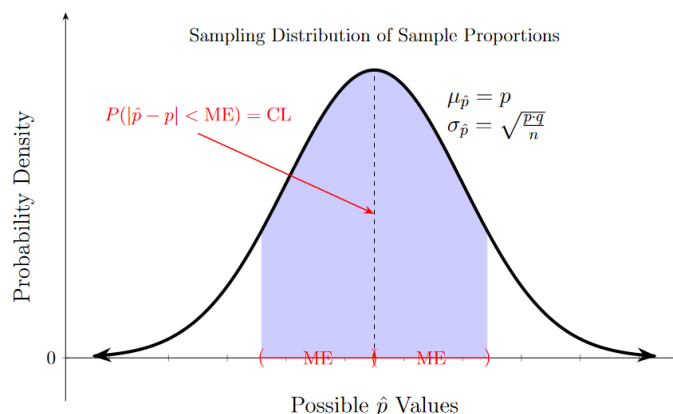
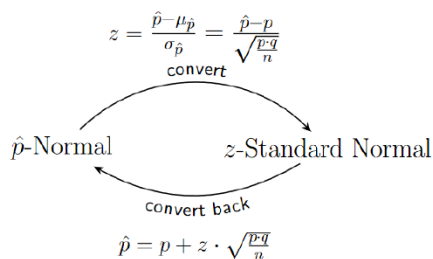
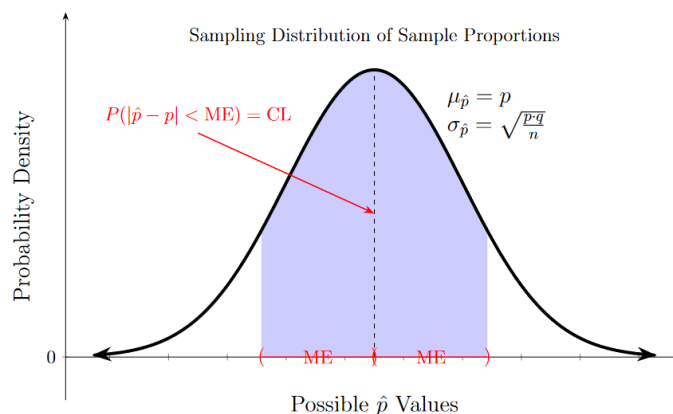


Figure 6.2.1: Sampling distribution of sample proportions

Now in relation to our given situation of estimating the proportion of all U.S. adults that have received the flu vaccine by using a sample of size $n = 1000$, we note that we do not know p and so, unlike our previous work in Chapter 5, we cannot determine the scaling of our horizontal axis in the natural scale of proportion measures. We address this in the next subsection on determination of the margin of error ME value.

Determining the Margin of Error in the Proportion Situation

We now use our powerful standardization feature on normal distributions from [Section 4.5](#); where any normal distribution can be converted in scale to the standard normal distribution. Consider the following figures illustrating the transformation process to our normal distribution of sample proportions.



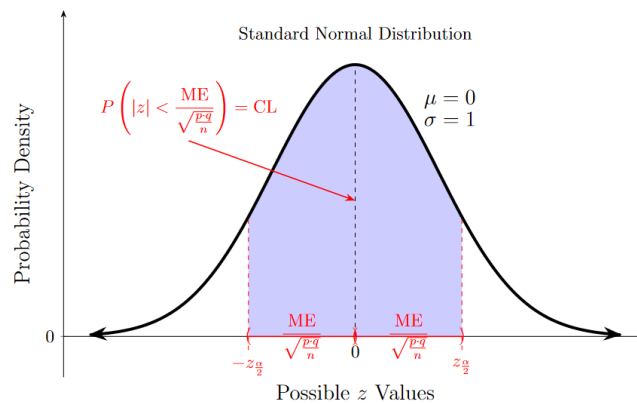


Figure 6.2.2: Sampling distribution of sample proportions transformed into the standard normal distribution

Notice that under the standard normal transformation process, the margin of error is scaled by a factor of one over the standard deviation of the sampling distribution; this occurs since we divided by the standard deviation value of $\sqrt{\frac{p \cdot q}{n}}$ in our scale transformation. Also, since we are now considering the standard normal distribution, we know the mean and standard deviation of the distribution. Therefore, using our computation technology, we can find the boundary values in the z -scale that will produce the desired confidence. These are represented by $\pm z_{\frac{\alpha}{2}}$ in the figure above. We call these points **critical z -values** or simply **critical values** in this process; these are completely determined once we have a chosen confidence level. Note that $z_{\frac{\alpha}{2}}$ represents the z -scale value where the area under the standard normal distribution to the right is $\frac{\alpha}{2}$ and $-z_{\frac{\alpha}{2}}$ represents the z -scale value where the area under the standard normal distribution to the left is $\frac{\alpha}{2}$. In converting back to the proportions' sampling distribution with $\hat{p} = p + z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}}$, or in related equivalent form of $\hat{p} - p = z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}}$, we should recognize something important. These critical z -values are telling us how many standard deviations of the sampling distribution we must differ from the mean of that distribution to capture the chosen CL percentage of sample results. That is, we have a measure of our "closeness" by $ME = \pm z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{p \cdot q}{n}}$.

We do have an issue in this computation of the margin of error since it needs the value of p , yet the value of p is unknown and what we are trying to estimate. However, since samples' proportions tend to be close in value to the population proportion, we will use our sample's proportion measure in the calculation. That is, we will find the margin of error measure of closeness by $ME \approx \pm z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}}$. For practical purposes, the use of \hat{p} and \hat{q} instead of p and q is reasonable as long as we meet the large-sample requirements of $n \cdot \hat{p} > 5$ and $n \cdot \hat{q} > 5$. It is worth noting that using $\hat{p} \cdot \hat{q} \approx p \cdot q$ is not the same as using $\hat{p} \approx p$; there is smaller error in the former than the latter. One illustration of this, which is developed further later in the section, is the fact that $p \cdot q$ is never larger than 0.25. For example, if $p = 0.4$ and we get $\hat{p} = 0.3$ then $p \cdot q = 0.24$ and $\hat{p} \cdot \hat{q} = 0.21$; even large differences between p and \hat{p} may still yield small differences between $p \cdot q$ and $\hat{p} \cdot \hat{q}$. This is worth observing so that it does not seem as if we are chasing our own tail. Our goal is to estimate p ; it would be pointless to use a formula to do so if the formula implicitly used $p \approx \hat{p}$. If we are concerned with the accuracy of our error measure, we can be more conservative and instead require $n \cdot \hat{p} > 10$ and $n \cdot \hat{q} > 10$ as discussed in Section 5.3. By using larger sample sizes, we can be more assured of our theory and hence in the validity of our measures produced by this theory. One should not use the above ideas on proportions measures if working with small sample sizes. We summarize the above in the following.

Margin of Error in Sample Proportions

Given a desire to estimate a population proportion measure p using a simple random sample's proportion \hat{p} in which the following conditions are known or reasonably believed to exist:

- the requirements for a binomial distribution are met with a sample size of n
- the requirements of $n \cdot \hat{p} > 5$ and $n \cdot \hat{q} > 5$ are met
- a confidence level of CL has been chosen and hence $\alpha = 100\% - CL$

then the margin of error in using the random sample's proportion measure is measured by

$$ME = \pm z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}}$$

where $\pm z_{\frac{\alpha}{2}}$ are the two critical values capturing the center CL% of the standard normal distribution.

If being more conservative in our approach, we may instead use requirements of $n \cdot \hat{p} > 10$ and $n \cdot \hat{q} > 10$.

With this theory in place, we now apply this to our specific context of the flu vaccine. We recall that we were interested in estimating the proportion of all U.S. adults who had taken the most recent flu vaccine. We had collected a random sample of 1,000 adults in which 735 had taken the vaccine, producing $\hat{p} = 0.735 = 73.5\%$. We note that the requirement for a binomial distribution are met with this context in relation to samples of size $n = 1,000$ and that $n \cdot \hat{p} = 1000 \cdot 73.5\% = 735 > 5$ and $n \cdot \hat{q} = 1000 \cdot 26.5\% = 265 > 5$.

Next, we do expect the actual population's proportion to be close to this 73.5% value due to our sampling distribution theory, but we need a measure of how close: a measure of the likely margin of error in the sample's result. To do so, we first must set a confidence level, say we choose CL = 95%. This means that this process will produce an interval which contains p 95% of the time. Then, to determine this margin of error, we proceed to the standard normal distribution to find the associated critical z -values tied to a central area of 95% and left/right tail areas of $\frac{\alpha}{2} = 2.5\%$, illustrated below in Figure 6.2.3.

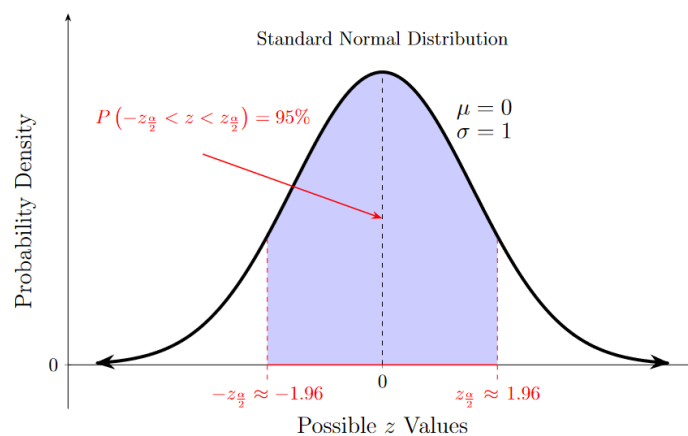


Figure 6.2.3: Standard normal distribution for a 95% confidence level

Using our approach of Section 4.6, we find these critical z -scores using our spreadsheet's NORM.S.INV function:

$$\text{left critical } z_{\frac{\alpha}{2}} = \text{NORM.S.INV}(0.025) \approx -1.95996$$

$$\text{right critical } z_{\frac{\alpha}{2}} = \text{NORM.S.INV}(0.975) \approx 1.95996$$

Of course, as seen in earlier work, since the standard normal distribution is symmetric about its mean scale value of 0, we need not actually compute both critical z -values as both will be the same sized value, just one negative and the other positive.

This now lets us determine our margin of error as tied to this chosen confidence level of 95% :

$$\begin{aligned} ME &= \pm z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}} \\ &\approx \pm 1.95996 \cdot \sqrt{\frac{0.735 \cdot 0.265}{1000}} \\ &\approx \pm 1.95996 \cdot 0.013956 \approx \pm 0.02735 = \pm 2.735\% \end{aligned}$$

Thus we have 95% confidence that our one random sample's proportion of $\hat{p} = 73.5\%$ is no more than 2.735% away from the population's actual proportion measure p . That is, in assuming our one collected sample's proportion is one of the central 95% of possible sample proportion values that can occur from samples of size 1,000, then our sample's proportion will be found on the horizontal scale somewhere below the shaded region, no more than 2.735% from the actual true population's proportion, as illustrated in Figure 6.2.4 below.

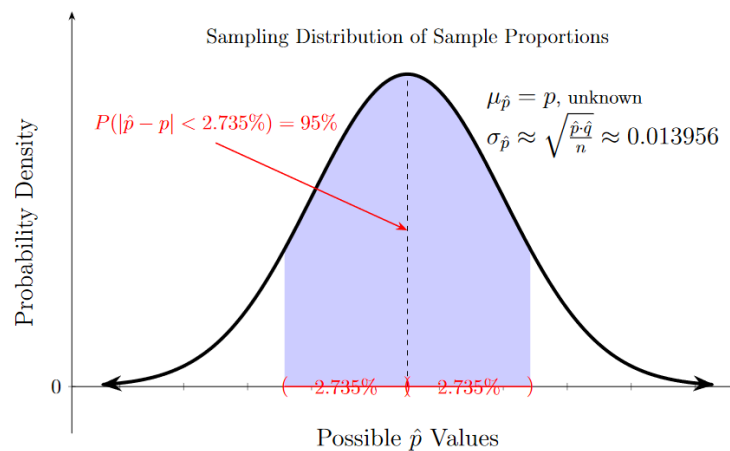


Figure 6.2.4: Illustration of margin of error in a sampling distribution

In the above computation of the margin of error measure, one should note that, although the confidence level was $CL = 95\%$ and complement alpha level was thus $\alpha = 5\%$, in determination of the critical z -values within the NORM.S.INV function, neither of these two numbers were directly used. Instead, since the spreadsheet's function requires use of only a left-area measure, we instead had to use $\frac{\alpha}{2} = 2.5\%$ and its complement measure of $1 - \frac{\alpha}{2} = 97.5\%$ within the spreadsheet function. This is a technology computational requirement that must be recalled when constructing these measures.

As a final summary of our specific example, we are able to state that we have 95% confidence that the true proportion of the U.S. adult population that had taken the most recent flu vaccine is approximately 73.5% with no more error than 2.735%. This now easily leads us to the final concept of this section, the **confidence interval** for the proportion situation.

Constructing Confidence Intervals for Proportions

Once we have a margin of error measure determined, we easily construct the confidence interval for the population proportion.

$$(\hat{p} - ME, \hat{p} + ME) = \left(\hat{p} - z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}}, \hat{p} + z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}} \right)$$

Or, instead of using algebraic interval notation, we may instead indicate the confidence interval as follows.

$$\hat{p} \pm ME = \hat{p} \pm z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}}$$

So in our flu vaccine context, we have a confidence interval of $73.5\% \pm 2.735\%$ or equivalently $(0.735 - 0.02735, 0.735 + 0.02735) = (0.707646, 0.762354) = (70.7646\%, 76.2354\%)$.

This allows us to state that we are 95% confident that the actual proportion of the U.S. adult population that had taken the most recent flu vaccine is between 70.7646% and 76.2354%, or equivalently $73.5\% \pm 2.735\%$.

Let us try a few more text exercises using the same theory but in varied contexts.

? Text Exercise 6.2.1

Use our theory on margin of error and confidence intervals established above, determine the following.

1. A state's department of education is interested in the proportion of all eighth-grade students in their state that will score at less-than-proficient in math on a national assessment. A random sample of 450 eighth-grade students from the state were given the national assessment and 345 of those students scored less-than-proficient in math. Develop an appropriate estimate from this information for the department of education, including the margin of error and related confidence interval based upon a choice of a 90% confidence level. Include a final concluding statement with the developed confidence interval.

Answer

We proceed by first developing the sample's proportion measure:

$$\hat{p} = \frac{345}{450} \approx 0.76666667 \approx 76.67\%$$

Thus, in the sample, about $\hat{p} = 76.67\%$ of the sampled eighth-grade students scored less-than-proficient on the national assessment and $\hat{q} \approx 23.33\%$ scored above less-than-proficient. We also note that we meet the basic requirements for our theory since $n \cdot \hat{p} = 345$ and $n \cdot \hat{q} = 105$ are both well above 5 and the situation is based on a random sample and a binomial experiment.

We now use this sample measure as a "best estimate" for the population's proportion, but also need to determine the possible likely margin of error in this estimate. We continue by developing $ME = \pm z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}}$. Since working with a chosen confidence level of 90%, then $\alpha = 10\%$, leading to determination of the following critical z -values.

$$\pm z_{0.05} = \pm \text{NORM.S.INV}(0.05) \approx \pm 1.64485$$

The statistical margin of error is

$$\begin{aligned} ME &= \pm z_{0.05} \cdot \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}} \\ &\approx \pm 1.64485 \cdot \sqrt{\frac{0.7667 \cdot 0.2333}{450}} \\ &\approx \pm 1.64485 \cdot 0.019938 \approx \pm 0.0328 = \pm 3.28\% \end{aligned}$$

Thus, based upon 90% confidence, we have at most a 3.28% margin of error in this sample estimate; leading to a confidence interval of $76.67\% \pm 3.28\%$, or in interval notation, $(0.7339, 0.7995) = (73.39\%, 79.95\%)$.

As a final summary, we are able to state that we have 90% confidence that the true proportion all eighth-grade students in this state that will score at less-than-proficient in math on the state assessment is approximately 76.67% with no more error than 3.28%; we are 90% confident that the true population proportion p falls somewhere between 73.39% and 79.95%.

2. A marketing researcher is interested in the proportion of European consumers who are aware of a U.S. branded product. A random sample of 375 European consumers were asked if they recognized the U.S. branded product; 75 stated they knew of the product. Develop an appropriate estimate of the proportion of all European consumers who are aware of the U.S. branded product from this information for the researcher, including the margin of error and related confidence interval based upon a choice of a 99% confidence level.

Answer

We again first develop the sample's proportion measure:

$$\hat{p} = \frac{75}{375} = 0.2000 = 20\%$$

Thus, in the sample, $\hat{p} = 20\%$ of the sampled Europeans were aware of the U.S. branded produce and $\hat{q} = 80\%$ were not aware. We also note that we meet the basic requirements for our theory since $n \cdot \hat{p} = 75$ and $n \cdot \hat{q} = 300$ are both well above 5 and the situation is based on a random sample and a binomial experiment.

We use this sample proportion as a "best estimate" for the population's proportion, but must determine the possible likely margin of error in this estimate. Since we are working with a chosen confidence level of 99%, then we have $\alpha = 1\%$, leading to determination of the following critical z -values of

$$\pm z_{0.005} = \pm \text{NORM.S.INV}(0.005) \approx \pm 2.57583$$

Thus, the statistical margin of error is

$$\begin{aligned} \text{ME} &= \pm z_{0.005} \cdot \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}} \\ &\approx \pm 2.57583 \cdot \sqrt{\frac{0.20 \cdot 0.80}{375}} \\ &\approx \pm 2.57583 \cdot 0.020656 \approx \pm 0.0532 = \pm 5.32\% \end{aligned}$$

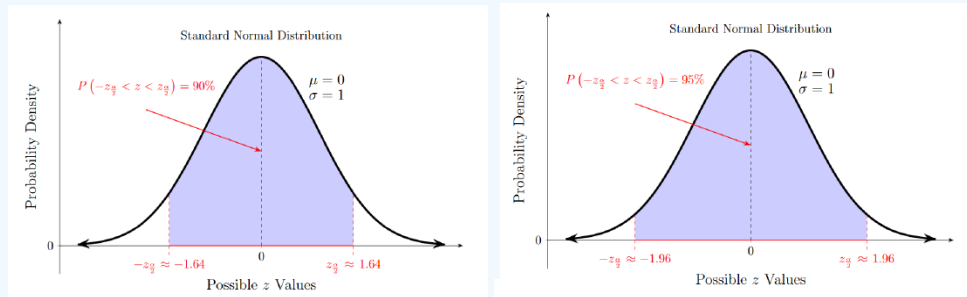
Based upon 99% confidence, we have at most a 5.32% margin of error in this sample estimate; leading to a confidence interval of $20\% \pm 5.32\%$, or in interval notation, $(0.1468, 0.2532) = (14.68\%, 25.32\%)$.

As a final summary, we are able to state that we have 99% confidence that the true proportion all Europeans that recognize the U.S. branded product is approximately 20% with no more error than 5.32%; we are 99% confident that the actual population proportion p falls somewhere between 14.68% and 25.32%.

3. What will happen to the margin of error if one increases the desired level of confidence?

Answer

Since the confidence level CL is tied to us assuming our random sample's proportion is within the central CL% of the sampling distribution or its standardization, we need only recognize what happens to our horizontal axis interval in relation to any adjustment of the level. Using a confidence level of 90 first and then of 95%, we can visually reason that increasing the confidence level increases the size of the horizontal axis interval (and hence the size of the related critical z -values) as is illustrated in the diagrams below.



Such an increase in the chosen confidence level will then cause the margin of error measure to be larger and hence the confidence interval to be wider. So, choosing to increase only the desired level of confidence (while also not changing any other option) will cause a larger margin of error. The ethical researcher will always set the confidence level before beginning the statistical analysis (not set after some statistical work just to force a smaller margin of error.) Those aware of this will also notice when research sets an unusually low confidence level, possibly in an attempt to narrow the margin of error in sample results so as to mislead consumers of the research.

4. What will happen to the margin of error if one decreases the sample size used to produce the sample proportion estimate?

Answer

Since $\text{ME} = z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$, we have that n is in the denominator. Using our number sense within simple arithmetic, we see that as n decreases, our denominator in our sampling distribution's standard deviation measure, \sqrt{n} , also decreases. When dividing by smaller and smaller numbers, the result of the quotient is larger and larger (for example, $\frac{1}{500} = 0.002$, $\frac{1}{50} = 0.02$, $\frac{1}{5} = 0.2$, and so on. As we decrease the denominator, the value of our fraction increases getting closer and closer to 1. Thus, the margin of error becomes larger as the sample size gets smaller. This should match our natural number sense that smaller samples are more likely to produce statistics which deviate more from the population parameter in comparison to larger samples.

Sample Size Determination in Confidence Intervals on Proportions

Based upon part 4 of the last text exercise group, we notice that sample size choice plays some role in controlling the magnitude of the margin of error. We can apply a bit of algebraic manipulation to develop a formula allowing us to pre-predict the sample size

needed to control the margin of error within any specific chosen level of confidence. Using our developed margin of error formula,

$$ME = \pm z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}},$$

we note that we can solve this algebraic formula for n , as illustrated below.

$$\begin{aligned} (ME)^2 &= \left(\pm z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}} \right)^2 && \text{— square both sides} \\ ME^2 &= \left(z_{\frac{\alpha}{2}} \right)^2 \cdot \frac{\hat{p} \cdot \hat{q}}{n} && \text{— simplify} \\ n \cdot ME^2 &= \left(z_{\frac{\alpha}{2}} \right)^2 \cdot \frac{\hat{p} \cdot \hat{q}}{\cancel{n}} \cdot \cancel{n} && \text{— multiply both sides by } n \\ \frac{n \cdot \cancel{ME^2}}{\cancel{ME^2}} &= \left(z_{\frac{\alpha}{2}} \right)^2 \cdot \hat{p} \cdot \hat{q} \cdot \frac{1}{ME^2} && \text{— divide both sides by } ME^2 \\ n &= \left(\frac{z_{\frac{\alpha}{2}}}{ME} \right)^2 \cdot \hat{p} \cdot \hat{q} && \text{— simplify} \end{aligned}$$

So we have developed a related formula that will tell us how large of sample we need once we have chosen a confidence level (so we can determine the critical z -value), a margin of error size, and some previous study's sample results (so we have values for \hat{p} and \hat{q} .) It would be nice to eliminate the requirements of a previous study, and we can do so if we take just a brief time to notice that the product $\hat{p} \cdot \hat{q}$ is predictable. Recall that \hat{q} is the complement of \hat{p} , so as illustrated by the table of values below.

Table 6.2.1: Products of proportion values and their complements

\hat{p}	\hat{q}	Product $\hat{p} \cdot \hat{q}$
0.00	1.00	$0.00 \cdot 1.00 = 0.00$
0.10	0.90	$0.10 \cdot 0.90 = 0.09$
0.20	0.80	$0.20 \cdot 0.80 = 0.16$
0.30	0.70	0.21
0.40	0.60	0.24
0.50	0.50	0.25
0.60	0.40	0.24
\vdots	\vdots	\vdots
1.00	0.00	0.00

We can inductively reason that the maximum product is $0.50 \cdot 0.50 = 0.25$. So a required sample size in the proportion situation can be found without a preliminary study by our developed formula given by:

$$n = \left(\frac{z_{\frac{\alpha}{2}} \cdot 0.50}{ME} \right)^2 = \left(\frac{z_{\frac{\alpha}{2}}}{ME} \right)^2 \cdot 0.25$$

The above leads to the following key findings.

Sample Size for Estimation of Population Proportions

Given a desire to estimate a population proportion measure p using a simple random sample's proportion \hat{p} in which the required conditions are to be met, then the sample size needed to meet a confidence level of CL% and margin of error of no more than ME can be found by the following computations:

$$n = \left(\frac{z_{\frac{\alpha}{2}}}{ME} \right)^2 \cdot \hat{p} \cdot \hat{q} \quad \text{— if a preliminary value of } \hat{p} \text{ is known}$$

$$n = \left(\frac{z_{\frac{\alpha}{2}}}{ME} \right)^2 \cdot 0.25 \quad \text{— if no preliminary value of } \hat{p} \text{ is known}$$

Now we apply these sample size concepts within a few exercises.

? Text Exercise 6.2.2

Using our sample size findings above, determine the following.

1. A state's department of education is interested in the proportion of all eighth-grade students in their state that will score at less-than-proficient in math on a national assessment. A researcher is interested in controlling the margin of error to no more than 1.5% while working under a 95% confidence level. A previous study from three years ago produced a sample proportion measure of $\hat{p} = 58\%$. What size sample is required for the researcher to meet the desired conditions?

Answer

We proceed by applying our developed sample size formula in which we need the margin of error to be at most $ME = 1.5\%$ and also happen to have a preliminary value of $\hat{p} = 0.58$ known. First, we must determine the critical z -scores $z_{\frac{\alpha}{2}}$ tied to the prescribed confidence level of 95%. Therefore $\alpha = 5\%$ and $\frac{\alpha}{2} = 2.5\%$.

$$z_{0.025} = \pm \text{NORM.S.INV}(0.025) \approx \pm 1.95996.$$

Hence, our sample size calculation is

$$\begin{aligned} n &= \left(\frac{z_{\frac{\alpha}{2}}}{ME} \right)^2 \cdot \hat{p} \cdot \hat{q} \\ &= \left(\frac{\pm 1.95996}{0.015} \right)^2 \cdot 0.58 \cdot 0.42 \\ &= 4159.019 \end{aligned}$$

Now, sample size must be a natural number, so we must round up any fractional-valued results. Common rounding which would often be rounding down will allow the margin of error to go slightly above the desired 1.5%, thus our need to always round up any resulting computed fractional amounts (the same is true for final interpretation of all sample size computations: we round up any fractional values.)

As a final summary, the researcher must collect a sample of at least size 4,160 in order to keep the margin of error at most 1.5% while also requiring 95% confidence. If, upon reflection, the researcher decides it is unreasonable (possibly due to cost) to collect data from such a large number of eighth-grade students in Kansas, then either the allowed margin of error must be increased or else the confidence level must be decreased in order to decrease the required sample size.

2. A marketing researcher is interested in the proportion of European consumers who are aware of a U.S. branded product. The researcher is interested in controlling the margin of error to no more than 5% while working under a 99% confidence level. No previous study has been found about this topic. What size sample is required for the researcher to meet the desired conditions?

Answer

This time we proceed by applying our developed sample size formula in which we need the margin of error to be at most $ME = 5\%$ but in which we have no preliminary value of \hat{p} known. So, again we must first determine the critical z -scores $z_{\frac{\alpha}{2}}$ tied to the prescribed confidence level of 99%. Therefore $\alpha = 1\%$ and $\frac{\alpha}{2} = 0.5\%$.

$$z_{0.005} = \pm \text{NORM.S.INV}(0.005) \approx \pm 2.57583.$$

Hence, our sample size calculation is

$$\begin{aligned}n &= \left(\frac{z_{\frac{\alpha}{2}}}{\text{ME}} \right)^2 \cdot 0.25 \\&= \left(\frac{\pm 2.57583}{0.05} \right)^2 \cdot 0.25 \\&= 663.4897\end{aligned}$$

We once again round this results to a needed sample size of 664.

As a final summary, the researcher must conservatively collect a sample of at least size 664 in order to keep the margin of error at most 5% while also requiring 99% confidence.

It is worth noting that these are the minimum sample sizes needed if the sample is obtained via a simple random process. Other methods of sampling may require larger sample sizes. It is also worth noting that all the theory discussed in this section, as well as all the examples, operates under the assumption that the sample is a simple random sample, meaning, all samples of size n are equally likely. Use of the methodology developed here on samples not obtained in this way could lead to a much higher probability of inaccuracy. Bear this in mind when reading statistical analyses.

6.2: Confidence Intervals for Proportions is shared under a [Public Domain](#) license and was authored, remixed, and/or curated by The Math Department at Fort Hays State University.

- **10.12: Proportion** by [David Lane](#) is licensed [Public Domain](#). Original source: <https://onlinestatbook.com>.
- **9.8: Sampling Distribution of p** by [David Lane](#) is licensed [Public Domain](#). Original source: <https://onlinestatbook.com>.
- **1.10: Distributions** by [David Lane](#) is licensed [Public Domain](#). Original source: <https://onlinestatbook.com>.