**The Sampling Distribution of a Sample Proportion**

* Recall that the binomial distribution has a mean .
* The sampling distribution of the sample proportion, i.e., the distribution of all possible samples of size n, is approximately normal with mean and standard deviation as long as both *n*⋅*p* ≥ 10 and *n*⋅(1 – *p*) ≥ 10.

**Recall that the general form for a confidence interval**

Point Estimate ± Margin of Error

Point Estimate ± (Critical Value)⋅(Standard Deviation of the Statistic)

Just as is used to estimate a population mean, the sample proportionis used to estimate of the population proportion *p*.

The confidence interval for a population proportion *p* requires

1. A point estimate of *p*, which is the sample proportion , pronounced “p-hat”
2. The appropriate critical value (z\*)
3. The standard error of ,

Unfortunately, we have another “Catch 22.” The computation of requires knowledge of *p*. In the absence of a known or assumed value for *p*, we estimate *p* with the point estimate . The estimated standard deviation is denoted by :

.

The estimate is known as the **standard error** of .

**Why not t\*?** When the sample mean estimates the population mean , a separate estimation of the standard deviation leads to more error, and thus, a t distribution. When the sample proportion estimates the population proportion *p*, the spread depends on *p*, not a separate parameter. The normal approximation is only slightly less accurate when replacing *p* in the standard deviation by .

**Standard Error:** When the standard deviation of a statistic is estimated from data, the result is called the **standard error** of the statistic.

* Since estimates the value of , we call the standard error of the sample mean.
* Since estimates the value of , we call the standard error of the sample proportion.

When the conditions are met, a C% confidence interval for the unknown proportion *p* is

where *z*\* is the critical value for the z-distribution, with C% of the area between –*z*\* and *z*\*.

* We call the quantity the **standard error** of the sample proportion**.**
* We call the quantity the **margin of error** of the estimation.

The conditions for this inference procedure are:

1. A random sample has been selected from the population.
2. The large counts condition: The values of *n* and  must satisfy both *n*⋅ ≥ 10 and *n*⋅(1 –) ≥ 10.

**Example 1**

From a random sample of 1411 youths from across the state of Pennsylvania, researchers found that 219 youths were obese. The point estimate for the proportion of all PA youths that are obese is . Construct and interpret a 99% confidence interval for the true proportion of obese youths in Pennsylvania.

|  |
| --- |
| **Solution**  **Choose the correct inference procedure.**  We will calculate a 99% confidence interval for a population proportion.  **Check the conditions**   * A simple random sample was selected. ✓ * The large counts condition:   + *n* = 1411(0.155) = 219 > 10 ✓   + *n*(1 –) = 1411(0.845) = 1192 > 10 ✓   **Carry out the inference procedure**  For a 99% level of confidence, *z*\* = 2.576.    **State the conclusion by interpreting the confidence interval**  We are 99% confident that the interval (0.130, 0.180) contains the true proportion of obese youths in Pennsylvania. |

In the calculation of the confidence interval in example 1, identify the value of each of the following.

1. The point estimate of the population proportion \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
2. The critical value \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
3. The standard error of the point estimate \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
4. The margin of error \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Determining a Sample Size**

We now wish to determine the minimum sample size needed to estimate a population proportion to within a given margin of error *M* and a specified level of confidence. As we did with the confidence interval for the population mean, set the expression for the margin of error equal to *M*.

Multiply both sides of the equation by . 

Divide both sides of the equation by *M*. 

Finally, solve for *n* by squaring both sides of the equation. 

As before, this formula will rarely yield a whole number. Since a sample size must be a whole number, always round your answer **up** to the nearest whole number so that the margin of error is less strict than *M*.

**Example 2**

Return to example 1. Determine the minimum sample size needed to estimate the true proportion of obese youths to within 1.5% with a 99% level of confidence. Use the estimate for *p* from example 1.

|  |
| --- |
| **Solution**  Use the formula with = 0.155, *M* = 0.015, and *z*\* = 2.576 (the critical value for a 99% confidence interval).  =  As expected, the result is not a whole number. Round this number up to 3863 youths.  **Answer:** A minimum sample of size 3863 youths is needed to estimate the true proportion of obese youths in Pennsylvania with a confidence level of 99%. |

|  |  |
| --- | --- |
|  | **Example 3**  How many M&Ms must be sampled to estimate the true proportion of red M&Ms to within 2% of the actual value with a 95% level of confidence? The point estimate for the proportion of red M&Ms is 0.22. |
| **Solution**  Use the formula with = 0.22, *M* = 0.02, and *z*\* = 1.96 (the critical value for a 95% confidence interval).  =  **Answer**  A minimum sample of size 1649 M&Ms is needed to estimate the true proportion of red M&Ms to within 0.02 of the actual proportion. | |