**A *z*-Confidence Interval for a Population Proportion**

This section presents a procedure for constructing a confidence interval for a population proportion p.

Recall the following important results from chapter 7….

* The sample proportion is an unbiased estimate of the population proportion *p*.
* The sampling distribution of the sample proportion is approximately normal with mean and standard deviation as long as the population size is at least ten times the size of the random sample ( *N* > 10*n* ) and both *n*⋅*p* > 10 and *n*⋅(1 – *p*) > 10.

Also, recall that the general form for a confidence interval is

**Point Estimate ± Margin of Error**

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

Clearly, the confidence interval for a population proportion *p* requires (1) a point estimate of *p* () , (2) the appropriate critical value (z\*), and (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 . We denote the estimate of **** is denoted by . Thus,

****.

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

**Definition: 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.

**8.7 A *z*-Confidence Interval for a Population Proportion**

|  |
| --- |
| **A *z*-Confidence Interval for a Population 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 10% condition: Check that *N* > 10*n*. 3. The large counts condition: The values of *n* and  must satisfy both *n*⋅ > 10 and *n*⋅(1 –) > 10. |

**Example 14**

From a random sample of 1411 youths from across the state of Pennsylvania, researchers found the 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. ✓ * 10% condition: It is safe to assume that there are more than 14,111 youths in PA. ✓ * 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.  = (0.130, 0.180).  **State the conclusion by interpreting the confidence interval**  We are 99% confident that the true proportion of obese youths in Pennsylvania is between 0.130 and 0.180. |

**8.7 A *z*-Confidence Interval for a Population Proportion**

**Example 15**

In the calculation of the confidence interval in example 14, 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.  



This final version of the formula will be used if a “ballpark” estimate of *p* is known. If the value of *p* is completely unknown, it can be shown (using Calculus methods) that the standard error for  is a maximum when = 0.50.

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*.

**8.7 A *z*-Confidence Interval for a Population Proportion**

**Example 17: A Sample Size Exercise**

Return to example 15. 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 15.

|  |
| --- |
| **Solution**  We will use the formula  where = 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 18: A Sample Size Exercise**  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? |
| **Solution**  We will use the formula  where *M* = 0.02 and *z*\* = 1.960.  In the absence of a preliminary estimate of , we will use . This is the value of  that maximizes the standard error of the sample proportion.  .  In a rare occurrence, the answer is a whole number. I will still “bump up” the answer to the next whole number so that the margin of error is truly within 0.02.  **Answer**  A minimum sample of size 2402 M&Ms is needed to estimate the true proportion of red M&Ms to within 0.02 of the actual proportion. | |