If you were asked to estimate the percent of Americans without health insurance, you might guess and say “about 20%” or “about 20%, give or take 4%”. Notice the second estimate is an “interval”. What would your confidence level be? Are you about 90% confident? If you needed to be more confident, say 95% confident, you might say between 20% – 5% and 20% + 5%. The “interval” is larger to allow for less error. So your answer is your estimate plus or minus an estimating error or error of estimate. For your information, the actual percent in 2012 was 15.4%; the website I used did not give a level of accuracy such as “plus or minus __ percent”.

Now let’s look at a schematic of a population and a sample; \( \rho \) represent the population proportion. This is a value we may never know due to the size of the population, but it can be estimated from a sample, as long as the sample is random and representative of the population. So, if you want to estimate the population proportion, you will need a “point estimate” of it. That would come from the sample proportion, \( \hat{p} \), called P-hat. Your estimate is also dependent on the spread of the data. What measures the spread of the data? Isn’t that the standard deviation?

Finally, you have to decide or be told how accurate your estimate needs to be. Do you want to be 90% accurate, 95% accurate, or 99% accurate? What “number” could be used to represent these levels of accuracy? It would be a “Z” value; 90% accuracy can be represented by the number 1.645; 95% by 1.96 and 99% by 2.58.

Let’s call the error of estimate “E”; E must have the following components: A measure of spread - we’ll use the standard deviation; a sample size, \( n \), and a degree of accuracy - we’ll use a z score. Do you remember how to find the standard deviation for a proportion? Using our point estimate of the proportion, \( \hat{p} \), called P-hat, the standard deviation is \( \sqrt{\frac{\hat{p} (1-\hat{p})}{n}} \). With this in mind, the error of estimate is \( E = Z \sqrt{\frac{\hat{p} (1-\hat{p})}{n}} \).

Let’s try an example. Should a traffic light have a left-turn signal in one direction of an intersection? You are familiar with the hard-rubber “counters” that are stretched across roads; let’s say that data from the counters indicated there were 117 left turns from 1587 people entering the intersection. What is P-hat? Find a 95% confidence interval for \( \rho \), the population proportion for all left turns for that intersection.

What is \( z \)? What is E? Add and subtract E from the point estimate of \( \rho \), the sample proportion \( \hat{p} = 0.0737 \). If it were rounded to three decimal places, 0.013. See the “caution” on the next page.

0.074 – 0.013 \( \leq \rho \) \( \leq \) 0.074 + 0.013 or 0.061 \( \leq \rho \) \( \leq \) 0.087 This will help Penn DOT make a decision.
Wow, I hope this can be done on the calculator! It can be! Look at the next page.

Focus – Estimating the Population Proportion

Press STAT, scroll over to TESTS and choose “A”: 1-Prop Z Int. Enter x = 117, n = 1587; choose the confidence level = 95 and “Calculate”; press ENTER. There’s your interval: $0.06087 \leq \rho \leq 0.08658$  
To find E, take the upper limit and subtract $\hat{p}$: $0.08685 - 0.07372 = 0.01286$.

CAUTION: THERE ARE RESTRICTIONS ON USING THIS METHOD!

This method cannot be used for all proportions. P-hat, $\hat{p}$ is actually a binomial. The binomial setting is a situation in which a subject has a particular characteristic or does not have it. In this example, a person entering the intersection either makes a left turn or does not make a left turn; there are only two choices and they are exclusive of each other. The “prefix “bi” means two, as you know from a previous chapter.

This process of estimating the population proportion depends on the distribution being approximately normal. The distribution will be approximately normal if $n \hat{p} (1 - \hat{p}) \geq 10$ and the sample size, n, is less than or equal to 0.05 of the entire population, N.

Is $n \hat{p} (1 - \hat{p}) \geq 10$ and $n \leq 0.05N$? If this is true, you can construct a confidence interval for $\rho$.

Example: Let’s use some raw data and do a problem. What percent of workers are participating in a retirement plan? This could be a company-sponsored plan with matching or non-matching funds, or a separate IRA plan through a bank or insurance company or a Roth IRA – an after-tax plan. In order to get this estimate, a nation-wide SRS (simple random sample) that best represents the current working force is gathered by a national news organization that has the resources to do so. Let’s say the news organization found the following to be true: 1378 out of 2394 surveyed participated in a retirement-savings plan. Can this problem be done using the method described above? What is $\hat{p}$? What is the standard deviation of the sample proportion? What is $Z$? What is the error of estimate, $E$? What is 90% confidence interval?

$\hat{p} = 1378 / 2394 = 0.5756$  
The standard deviation is 
$\sqrt{0.5756(0.4244)/2394} \approx 0.0101$  
$Z = \text{InvNorm}(0.05) = -1.645$  
$E = 1.6445 (0.0101) = 0.017$  
Interval: $0.576 - 0.017 \leq \rho \leq 0.576 + 0.017$ or $0.559 \leq \rho \leq 0.593$  
Calculator answer: $0.55899 \leq \rho \leq 0.59222$

HANDOUTS at www.hacc.edu  
Choose “Lancaster Campus” from the “pull-down” menu. Scroll to bottom and choose “Tutoring” under the “STUDENTS” heading. Choose Lancaster Tutoring” on the left side. Scroll down to “Guides for TI 83 or 84 graphing calculators” – in red letters – click on it. Scroll down to ”Handouts”. If you are unfamiliar with the calculator, choose “Calculator Handout – Basic Operations” first, or
choose the topic you are learning and, if you need to do so, print it. You can also come to the Lancaster Tutoring Center for free personal help! No appointment is needed!