13 – Confidence Intervals, Part II

There are two more types of confidence intervals you should know: those for *proportions* and for *standard deviation*.

A proportion is a percentage computed about a sample or population. For example, we know that 25.47% of Point Park students are on campus residents (975 of the 3827 Fall 2012 students). Proportions will be expressed as decimals for the formula below.

To compute of a confidence interval for population proportion, we use the formula $\hat{p} \pm z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$. The parameters in this formula are: $\hat{p}$ is the sample proportion

z is a z-score from the standard normal distribution (determined by level of certainty)

n is the number of data points in the sample

Example 1: The portion of Point Park students who live on campus is 25.47% (975 of the 3827 Fall 2012 students). Using this as a sample for the nationwide college student population, compute a 90% confidence interval for the portion of college students nationwide that live on campus.

Solution: The values of $\hat{p}$ and n are given directly in the problem: $\hat{p}$ = 0.2547 and n = 3827. We will use the formula $\hat{p} \pm z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$. The value of z needed will correspond to the 90% level of certainty prescribed by the problem. The value we seek is the z-score that corresponds to an area of 10% being *shared* by the two tails in the standard normal distribution. That is, we must look up the z-score that gives an area of 5% = .0500 to its left. That value is z = -1.645 (midway between -1.64 and -1.65).

The formula thus becomes 0.2547 ± (-1.645)$\sqrt{\frac{0.2547 \times 0.7453}{3827}}$. Computation give us the two values of 0.2431 and 0.2662. These numbers represent the lower and upper boundaries of the confidence interval and our conclusion is that we are 90% certain that the portion of students nationwide who live on campus is between 24.31% and 26.62%.

Just as it was done for the population mean, we can algebraically solve the formula $E = z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ for n. That solution is $n = \hat{p}(1-\hat{p})\left(\frac{z}{E}\right)^2$. This enables us to determine an appropriate sample size when a certain precision is required. For example, if we want to compute a 99% confidence interval for population proportion and we need that interval to be within 6.5 units of the sample proportion, we use E = 6.5 in the formula.

However, there is an issue with the formula for sample size: it involves $\hat{p}$, or the sample proportion. So, if we want to compute a confidence interval *after* finding the appropriate sample size, we need to optimize. To do that, consider the product $\hat{p}(1 - \hat{p})$ in the formula. Since $\hat{p}$ is necessarily a decimal between 0 and 1, the table to the right explores various values of that product.

$\hat{p}$	$1 - \hat{p}$	$\hat{p}(1 - \hat{p})$
0.1	*0.9*	*0.09*
0.2	*0.8*	*0.16*
0.3	*0.7*	*0.21*
0.4	*0.6*	*0.24*
0.5	*0.5*	*0.25*
.06	*0.4*	*0.24*
0.7	*0.3*	*0.21*
0.8	*0.2*	*0.16*
0.9	*0.1*	*0.09*

Using $\hat{p}$ = 0.5, or 50%, produces the largest value of this product, which will make for the largest possible (yet reasonable) value of n in our formula.

So, if we just use $\hat{p}$ = 0.5 for all cases, the formula is much easier:

$$n = \hat{p}(1-\hat{p})\left(\frac{z}{E}\right)^2 = (0.5)(0.5)\left(\frac{z}{E}\right)^2 = (0.25)\left(\frac{z}{E}\right)^2 = \frac{1}{4}\left(\frac{z}{E}\right)^2 = \left(\frac{z}{2E}\right)^2$$

This will be the formula we use for finding sample size for confidence intervals of a prescribed size: $n = \left(\frac{z}{2E}\right)^2$

Example 2: A student group wishes to compute a 90% confidence interval for the portion of drivers who are talking on their cell phones. If the confidence interval is to be no more than 0.15 wide, how many drivers should the students observe?

Solution: The parameters given in the problem should be used to identify *E* = 0.075 (error is half the width of the interval) and *z* = -1.645 (midway between *z*-values for 0.0495 and 0.0505). So, the sample size should be $n = \left(\frac{z}{2E}\right)^2 = n = \left(\frac{-1.645}{2\times 0.075}\right)^2 \approx 120.3$. So, they should observe 121 drivers (always round up).

Confidence intervals for standard deviation use a new distribution: the *Chi-square distribution*. This is a right-skewed distribution of positive values. Because it is not symmetrical, confidence intervals cannot be computed by adding and simultaneously subtracting an error term. Rather, the two bounds of the confidence interval are computed separately.

To compute of a confidence interval for population standard deviation, we use the formula $\left(\sqrt{\frac{(n-1)s^2}{\chi_r^2}}, \sqrt{\frac{(n-1)s^2}{\chi_l^2}}\right)$

The parameters in this formula are:

s is the sample standard deviation

χ_l^2 and χ_r^2 are values from the chi-square distribution (determined by level of certainty)

n is the number of data points in the sample

Looking up values in the chi-square distribution is different that looking up values in the *z* and *t* distributions. Table G, found on page 649 of the text or on the foldout, lists values of the variable within the table and area to the *right* of the values at the tops of the columns. Rows correspond to degrees of freedom, and again *df* = *n* – 1.

Example 3: When computing a 95% confidence interval for standard deviation, we must find values in the chi-square distribution that reflect 95% of the area between them. If the sample size is 30, then *df* = 29. The area to the right of χ_r^2 and the area to the left of χ_l^2 will both be 2.5%, or 0.0.25.

Since the table directly provides values with area to the right, we find 0.025 on the top border and χ_r^2 = 45.722. We must subtract 0.025 from 1 to get 0.975 for χ_l^2, and χ_l^2 = 16.047.

Example 3, continued: In a manufacturing process of markers of diameter 1.5 cm, the actual diameter of each marker varies by a small amount. A quality controller measures 30 markers. The standard deviation of the sample of 30 markers is 0.12 cm. Find a 95% confidence interval for the standard deviation of all markers produced by the manufacturing process.

Solution: The values of χ_r^2 and χ_l^2 were found in Table G above. They are and χ_r^2 = 45.722 and χ_l^2 = 16.047. Using the formula $\left(\sqrt{\frac{(n-1)s^2}{\chi_r^2}}, \sqrt{\frac{(n-1)s^2}{\chi_l^2}}\right)$, the bounds of the interval are computed separately with *n* = 30 and *s* = 0.12.

$$\sqrt{\frac{(n-1)s^2}{\chi_r^2}} = \sqrt{\frac{29\times(0.12)^2}{45.722}} \approx 0.096$$

$$\sqrt{\frac{(n-1)s^2}{\chi_l^2}} = \sqrt{\frac{29\times(0.12)^2}{16.047}} \approx 0.161$$

We can be 95% certain that the standard deviation of all markers produced is between 0.096 cm and 0.161 cm.