

## Exact Confidence Limits for a Proportion

Since

$$\hat{\mathbf{p}} \sim \mathbf{N}\left(\mathbf{p}, \frac{\mathbf{p}(1-\mathbf{p})}{\mathbf{n}}\right) \quad (1)$$

Then the confidence interval for  $\hat{\mathbf{p}}$  is:

$$\mathbf{P}\left(\hat{\mathbf{p}} - z_{\alpha/2} \sqrt{\frac{\mathbf{p}(1-\mathbf{p})}{\mathbf{n}}} < \mathbf{p} < \hat{\mathbf{p}} + z_{\alpha/2} \sqrt{\frac{\mathbf{p}(1-\mathbf{p})}{\mathbf{n}}}\right) = 1 - \alpha \quad (2)$$

Because  $\mathbf{p}$  is not known, the usual assumption is to appeal to the Central Limit Theorem and assume that:

$$\hat{\mathbf{p}} \sim \mathbf{N}\left(\mathbf{p}, \frac{\hat{\mathbf{p}}(1-\hat{\mathbf{p}})}{\mathbf{n}}\right) \quad (3)$$

And the confidence interval is:

$$\mathbf{P}\left(\hat{\mathbf{p}} - z_{\alpha/2} \sqrt{\frac{\hat{\mathbf{p}}(1-\hat{\mathbf{p}})}{\mathbf{n}}} < \mathbf{p} < \hat{\mathbf{p}} + z_{\alpha/2} \sqrt{\frac{\hat{\mathbf{p}}(1-\hat{\mathbf{p}})}{\mathbf{n}}}\right) = 1 - \alpha \quad (4)$$

However, when  $\mathbf{p}$  is small equation (4) often produces intervals that are unreasonable.

A better solution is to work with equation (2) as follows:

$$\begin{aligned} & \mathbf{P}\left(\hat{\mathbf{p}} - z_{\alpha/2} \sqrt{\frac{\mathbf{p}(1-\mathbf{p})}{\mathbf{n}}} < \mathbf{p} < \hat{\mathbf{p}} + z_{\alpha/2} \sqrt{\frac{\mathbf{p}(1-\mathbf{p})}{\mathbf{n}}}\right) = \\ & \mathbf{P}\left(-z_{\alpha/2} \sqrt{\frac{\mathbf{p}(1-\mathbf{p})}{\mathbf{n}}} < \mathbf{p} - \hat{\mathbf{p}} + z_{\alpha/2} \sqrt{\frac{\mathbf{p}(1-\mathbf{p})}{\mathbf{n}}}\right) = \mathbf{P}\left(|\mathbf{p} - \hat{\mathbf{p}}| < z_{\alpha/2} \sqrt{\frac{\mathbf{p}(1-\mathbf{p})}{\mathbf{n}}}\right) = 1 - \alpha \end{aligned}$$

Squaring both sides of the last expression we get:

$$\begin{aligned} & \mathbf{P}\left((\mathbf{p} - \hat{\mathbf{p}})^2 < z_{\alpha/2}^2 \frac{\mathbf{p}(1-\mathbf{p})}{\mathbf{n}}\right) = \mathbf{P}\left(\mathbf{p}^2 - 2\mathbf{p}\hat{\mathbf{p}} + \hat{\mathbf{p}}^2 < z_{\alpha/2}^2 \frac{\mathbf{p}(1-\mathbf{p})}{\mathbf{n}}\right) = \\ & \mathbf{P}\left(\mathbf{p}^2 \left(1 + \frac{z_{\alpha/2}^2}{\mathbf{n}}\right) + \mathbf{p} \left(-2\hat{\mathbf{p}} - \frac{z_{\alpha/2}^2}{\mathbf{n}}\right) + \hat{\mathbf{p}}^2\right) = 1 - \alpha \end{aligned} \quad (5)$$

The equation inside the last  $\mathbf{P}(\cdot)$  is a standard quadratic in  $p$ . Recall that if:

$$\mathbf{ax}^2 + \mathbf{bx} + \mathbf{c} = \mathbf{0} \text{ then } \mathbf{x} = \frac{-\mathbf{b} \pm \sqrt{\mathbf{b}^2 - 4\mathbf{ac}}}{2\mathbf{a}}$$

Here the solution is:

$$\mathbf{p} = \frac{(2\hat{\mathbf{p}} + \frac{\mathbf{z}_{\alpha/2}^2}{\mathbf{n}}) \pm \sqrt{\left(-2\hat{\mathbf{p}} - \frac{\mathbf{z}_{\alpha/2}^2}{\mathbf{n}}\right)^2 - 4\left(1 + \frac{\mathbf{z}_{\alpha/2}^2}{\mathbf{n}}\right)\hat{\mathbf{p}}^2}}{2\left(1 + \frac{\mathbf{z}_{\alpha/2}^2}{\mathbf{n}}\right)} \quad (6)$$

Note that the two solutions for  $\mathbf{p}$  in equation (6) are the confidence limits in equation (2).

Equation (6) is clearly superior to equation (4) in that the lower bound *cannot be less than zero* and the upper bound *cannot be greater than one*.