Exact Confidence Limits for a Proportion

Since

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

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

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

Because **p** is not known, the usual assumption is to appeal to the Central Limit Theorem and assume that:

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

And the confidence interval is:

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

However, when **p** is small equation (4) often produces intervals that are unreasonable.

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

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

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

Squaring both sides of the last expression we get:

$$\mathbf{P}\left((\mathbf{p} - \hat{\mathbf{p}})^2 < \mathbf{z}^2_{\alpha/2}\frac{\mathbf{p}(1-\mathbf{p})}{\mathbf{n}}\right) = \mathbf{P}\left(\mathbf{p}^2 - 2\mathbf{p}\hat{\mathbf{p}} + \hat{\mathbf{p}}^2 < \mathbf{z}^2_{\alpha/2}\frac{\mathbf{p}(1-\mathbf{p})}{\mathbf{n}}\right) =$$

$$\mathbf{P}\left(\mathbf{p}^2(1 + \frac{\mathbf{z}^2_{\alpha/2}}{\mathbf{n}}) + \mathbf{p}(-2\hat{\mathbf{p}} - \frac{\mathbf{z}^2_{\alpha/2}}{\mathbf{n}}) + \hat{\mathbf{p}}^2\right) = 1 - \alpha \tag{5}$$

The equation inside the last **P(..)** is a standard quadratic in p. Recall that if:

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

Here the solution is:

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

Note that the two solutions for **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*.