Set 24: Section 6.1.2, Confidence Intervals for Binomial proportion

We construct a confidence interval for the unknown p in the model $X \sim \text{Binomial}(n, p)$. We require $np \geq 5$ and $n(1-p) \geq 5$ so that we can use the approximation

$$X \approx \text{Normal}[np, np(1-p)].$$

Let $\hat{p} = X/n$ and $\hat{p}_{\text{obs}} = x_{\text{obs}}/n$. A $(1-\alpha)\%$ confidence interval for p is obtained via:

$$\begin{aligned}
& \mathrm{P}\left(-z_{\frac{\alpha}{2}} < \frac{X - np}{\sqrt{np(1-p)}} < z_{\frac{\alpha}{2}}\right) = 1 - \alpha \\
\Leftrightarrow \quad & \mathrm{P}\left(-z_{\frac{\alpha}{2}} < \frac{\hat{p} - p}{\sqrt{p(1-p)/n}} < z_{\frac{\alpha}{2}}\right) = 1 - \alpha \\
\Leftrightarrow \quad & \ldots \\
\Leftrightarrow \quad & \mathrm{P}\left(\hat{p} - z_{\frac{\alpha}{2}}\sqrt{p(1-p)/n} < p < \hat{p} + z_{\frac{\alpha}{2}}\sqrt{p(1-p)/n}\right) = 1 - \alpha
\end{aligned}$$

Therefore,

$$\hat{p}_{\text{obs}} \pm z_{\frac{\alpha}{2}}\sqrt{\hat{p}_{\text{obs}}(1 - \hat{p}_{\text{obs}})/n} \tag{1}$$

is an approximate $(1-\alpha)100\%$ CI for p. The CI (1) is based on two approximations:

1. **approximating the Binomial with the Normal**
2. **replacing p with $\hat{p}$**

Example: A sample of 1380 randomly selected books produced by a publishing company finds that 25 have bookbinding errors. Find a 95% confidence interval for p, the proportion of books with bookbinding errors.

The estimate for p, the proportion with bookbinding errors is $\hat{p} = 25/1380 \approx 0.018$.

Since the number which do and which don't have bookbinding errors are both greater than 5**, then** $\hat{p}$ **is approximately normal.**

$$\textbf{approx CI} = 0.018 \pm 1.96(0.004) = 0.018 \pm 0.008$$