Normal Approximation for Sample Proportions

Central Limit Theorem

Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from any distribution with a finite mean μ and finite standard deviation σ. For a sufficiently large sample size, n, the distribution of $\hat{P} = \frac{X}{n}$ is approximately normally distributed regardless of the shape of the distribution of X.

For a normal distribution, $\mathrm{N}(\mu, \sigma)$, we need the mean and standard deviation of the distribution.
If we do not know the population mean and population standard deviation, we can approximate the distribution using a sample's mean and standard deviation.

Mean

$$\mathrm{E}(\hat{P}) \approx \bar{x} = \hat{p}$$

Variance

$$\mathrm{var}(\hat{P}) \approx s^2 = \frac{\hat{p}(1-\hat{p})}{n}$$

Standard Deviation

$$\mathrm{sd}(\hat{P}) \approx s = \sqrt{\mathrm{var}(\hat{P})} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Therefore we get:

$$\hat{P} \sim \mathrm{N}\left(\hat{p}, \frac{\hat{p}(1-\hat{p})}{n}\right) \text{ approximately as } n \to \infty \qquad Z = \frac{X-\mu}{\sigma} = \frac{\hat{P}-\hat{p}}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \to \mathrm{N}(0,1) \text{ as } n \to \infty$$

Determining Probabilities with Sample Proportions using Normal Approximation

The probability that the proportion of a specific sample of size n exceeds a desired proportion, $\hat{p}$ given that the proportion of the population, p, is $\Pr(\hat{P} > \hat{p})$ where $\hat{P} \sim \mathrm{N}\left(p, \frac{p(1-p)}{n}\right)$.

Using a Transformed Normal Distribution

$$\hat{P} \sim \mathrm{N}\left(p, \frac{p(1-p)}{n}\right), \qquad \Pr(\hat{P} > \hat{p})$$

Using a Standardised Normal Distribution

$$Z = \frac{\hat{P}-p}{\sqrt{\frac{p(1-p)}{n}}} \sim \mathrm{N}(0,1), \therefore \Pr(\hat{P} > \hat{p}) = \Pr\left(Z > \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}}\right)$$

Example

If a cow gets a certain disease, there is only a 25% probability that it will recover.
If 200 cows get the disease, what is the probability that at least 30% will recover? Give an approximate answer using a normal approximation and the probability without using normal approximation.

$$p = 0.25, \qquad s^2 = \frac{0.25 \times 0.75}{200} = \frac{3}{3200} = \frac{6}{6400} \Rightarrow s = \sqrt{\frac{6}{6400}} = \frac{\sqrt{6}}{80}$$

Normal Approximation

$$\hat{P} \sim \mathrm{N}\left(0.25, \frac{3}{3200}\right), \qquad \Pr(\hat{P} \geq 0.3) = 0.0512, \qquad Z \sim \mathrm{N}(0,1), \qquad \Pr\left(Z \geq \frac{0.3-0.25}{\sqrt{6}/80}\right) = 0.0512$$

Probability using Binomial Distribution

$$\hat{P} = \frac{X}{200}, \qquad X \sim \mathrm{Bi}(200, 0.25^2), \qquad \Pr(\hat{P} \geq 0.3) = \Pr(X \geq 60) = 0.0625$$