- **Confidence intervals for proportions**

  Visit the Hyperstat site to learn about setting up confidence intervals for proportions. Be sure to only work through the 3 pages on this topic.

-

Applying the general formula for a confidence interval, the confidence interval for a proportion, π, is:

$p \pm z\, \sigma_p$

where p is the proportion in the sample, z depends on the level of confidence desired, and σ_p, the standard error of a proportion, is equal to:

$$\sigma_p = \sqrt{\frac{\pi(1-\pi)}{N}}$$

where π is the proportion in the population and N is the sample size. Since π is not known, p is used to estimate it. Therefore the estimated value of σ_p is:

$$\sqrt{\frac{p(1-p)}{N}}$$

As an example, consider a researcher wishing to estimate the proportion of X-ray machines that malfunction and produce excess radiation. A random sample of 40 machines is taken and 12 of the machines malfunction. The problem is to compute the 95% confidence interval on π, the proportion that malfunction in the population.

The value of p is 12/40 = 0.30. The estimated value of σ_p is $\sqrt{\frac{(0.3)(0.7)}{40}}$ = 0.072.

A z table can be used to determine that the z for a 95% confidence interval is 1.96. The limits of the confidence interval are therefore:

Lower limit = .30 - (1.96)(0.072) = .16
Upper limit = .30 + (1.96)(0.072) = .44.
The confidence interval is: $0.16 \le \pi \le .44$.

Correction for Continuity

Since the sampling distribution of a proportion is not a continuous distribution, a slightly more accurate answer can be arrived at by applying the correction for continuity. This is done simply by subtracting 0.5/N from the lower limit and adding 0.5/N to the upper limit. For the present example, 0.5/N = 0.5/40 = 0.01. Therefore the corrected interval is: $0.15 \le \pi \le 0.45$.

Summary of Computations

1. Compute p
2. Estimate σ_p by $\sqrt{\frac{p(1-p)}{N}}$
3. Find z for the level of confidence desired with a z table.
4. Lower limit = p - (z) (Estimated σ_p) - 0.5/N
5. Upper limit = p + (z) (Estimated σ_p) + 0.5/N
6. Lower limit $\leq \pi \leq$ Upper limit

Assumptions

1. Observations are sampled randomly and independently.
2. The adequacy of the normal approximation depends on the sample size (N) and π. Although there are no hard and fast rules, the following is a guide to needed sample size:

   If π is between 0.4 and 0.6 then an N of 10 is adequate. If π is as low as 0.2 or as high as 0.8 then N should be at least 25. For π as low as 0.1 or as high as 0.9, N should be at least 30.

   A more conservative rule of thumb that is often recommended is that $N\pi$ and $N(1 - \pi)$ should both be at least 10.

Click here for an interactive demonstration of the normal approximation to the binomial to explore the validity of these rules of thumb.