Estimating the standard deviation of a normal population

When to use this

- You want to estimate the standard deviation of some characteristic of a population that can be modeled reasonably well by a normal distribution.

Assumptions

The theory underlying the interval estimates depends on the data having a normal distribution.

A good approach is rather than thinking "Is my data normal?" think "Is the normal distribution a good model for my data?".

Point estimates

The point estimate for the population standard deviation σ is the sample standard deviation s:

$$\hat{\sigma} = s = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \overline{y})^2}{n-1}}$$

The divisor $(n-1)$ makes s^2 an unbiased estimate of σ^2. Dividing by n gives the maximum likelihood estimate. In practice, it makes very little difference unless the sample is tiny.

Interval estimates

Confidence intervals are constructed using the fact that under the normality assumptions the sample variance s^2 has a χ^2 distribution with $n-1$ degrees of freedom.

The formula for the upper and lower bounds of a $100(1-\alpha)\%$ confidence interval for σ^2 is:

$$\frac{(n-1)s^2}{\chi^2_{1-\alpha/2}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{\alpha/2}}$$

Where:

- s is the sample standard deviation

- n is the sample size
- $L = \chi^2_{\alpha/2,n-1}$ is the value x such that the proportion of a χ^2_{n-1} polulation lying to the left of x is $\alpha/2$.
- $U = \chi^2_{1-\alpha/2,n-1}$ is the value x such that the proportion of a χ^2_{n-1} polulation lying to the left of x is $1-\alpha/2$.