

Chauvenet's Criterion

Before the method is implemented, the user must be queried for a “Chauvenet's Criterion Threshold,” the probability (C) of finding a datum farther from the mean than the suspected outlier. This must be left up to the user, but should not be changed for each analysis. I think we should bury it in an “options” menu somewhere, with a default of 0.50, the most commonly accepted value.

1. Calculate the mean (μ) and standard deviation (σ) of the n data points.
2. For each datum, calculate $z = \frac{|x - \mu|}{\sigma}$.
3. For the largest z , calculate the probability of finding a point farther from the mean.

This is $P = n \cdot \sqrt{\frac{2}{\pi}} \int_z^{\infty} e^{\left(-\frac{1}{2}x^2\right)} dx$ or $P = n \cdot \text{erfc}\left(\frac{z}{\sqrt{2}}\right)$.

4. If $P < C$ then discard the data point corresponding to the high z value.
5. Recalculate the mean and standard deviation of the data points without the discarded outlier, (n becomes $n-1$), and repeat the process until no more data points may be discarded.