ERROR BARS

This document describes the algorithms for error bar computation of the mean, median and their confidence intervals for a simple random sample.

Notation

The following notation is used throughout this document unless otherwise noted:

Let $y_1 \le ... \le y_m$ be m ordered observations for the sample and $w_1,...,w_m$ be the corresponding case weights. Then

$$ww_i = \sum_{k=1}^{i} w_k = \text{cumulative sum of weights up to and including } y_i$$

and

$$W = ww_m = \sum_{k=1}^{m} w_k = \text{total sum of weights}$$

$p = \dfrac{CI}{100}$, CI is the confidence interval level $0 \le CI < 100$.

Descriptive Statistics

Mean ($\bar{y}$)

$$\bar{y} = \frac{\sum_{i=1}^{m} w_i y_i}{W}$$

Confidence Interval for the Mean

Lower bound = $\bar{y} - IDF.T(\dfrac{p+1}{2}, W-1) \cdot SE$

Upper bound = $\bar{y} + IDF.T(\dfrac{p+1}{2}, W-1) \cdot SE$

where SE is the standard error, and IDF.T is the inverse student t function documented in the COMPUTE command.

Variance (s^2)

$$s^2 = \frac{1}{W-1}\sum_{i=1}^{m} w_i (y_i - \bar{y})^2$$

Standard Deviation

$$s = \sqrt{s^2}$$

Standard Error

$$SE = \frac{s}{\sqrt{W}}$$

Median

The Aempirical method in the EXAMINE procedure is used for computation of the median. Let

$$v = \frac{W}{2}$$

and k satisfies

$$ww_k \le v < ww_{k+1}$$

Then,

$$g = v - ww_k$$

Let *m* be the estimated median, then it is defined as

$$m = \begin{cases} (y_k + y_{k+1})/2, & g = 0 \\ y_{k+1}, & g > 0 \end{cases}$$

Confidence Interval for the Median

Note: the case weights $w_1, \cdots, w_m$ must be integers for the following computation. If at least one weight is not integer, an error message is issued.

Let

$$
\begin{aligned}
b_i &= \Pr[Binomial(W,0.5) \geq i] \\
&= \sum_{j=i}^{W} \binom{W}{i} 0.5^W \\
&= IB(0.5; i, W-i)
\end{aligned}
$$

where IB is the incomplete Beta function.

Define

$$
\begin{aligned}
\gamma_i &= \Pr[i \leq Binomial(W,0.5) \leq W-i] \\
&= b_i - b_{W-i+1}
\end{aligned}, \qquad i = 0,1,\ldots, floor(W/2)
$$

and define

$\gamma_{w/2+1} = 0$, if W is even;

$\gamma_{(w+1)/2} = 0$, if W is odd.

Algorithm: Hettmansperger-Sheather Interpolation (1986)

1. Re-index all the cases to be $x_1 \leq x_2, \ldots, \leq x_W$ in which

$$
\begin{aligned}
x_1 &= x_2, \ldots &= x_{ww_1} &= y_1 \\
x_{ww_1+1} &= x_{ww_1+2} \ldots &= x_{ww_2} &= y_2 \\
&\vdots \\
x_{ww_{m-1}+1} &= x_{ww_{m-1}+2} \ldots &= x_{ww_m} &= y_m
\end{aligned}
$$

2. If W is even, compute $\gamma_0, \ldots, \gamma_{W/2}$. If W is odd, compute $\gamma_0, \ldots, \gamma_{(W+1)/2}$.
3. Choose the smallest index k such that $\gamma_{k+1} \leq p$. If k is found, go to Step 4; otherwise, stop and issue a message.
4. Compute

$$
l = \frac{\gamma_k - p}{\gamma_k - \gamma_{k+1}},
$$

and

$$
\lambda = \frac{(W-k)l}{k + (W-2k)l}.
$$

The p confidence interval is

Lower bound = $\lambda \cdot x_{k+1} + (1-\lambda) \cdot x_k$

Upper bound = $\lambda \cdot x_{W-k} + (1-\lambda) \cdot x_{W-k+1}$

References

Hettmansperger, T. P., and Sheather, S. J. 1986. Confidence Interval Based on Interpolated Order Statistics, Statistical Probability Letters, 4: 75–79.