Trimmed mean

Mean is too sensitive to extreme observations. Trimmed mean is designed to solve that problem. It involves trimming α percent observations from both ends. E.g.: If you are asked to compute 10% trimmed mean, $\alpha=0.10$

Given a bunch of observations, X_i

1. First, find n - number of observations
2. Reorder them as "order statistics" X_i from the smallest to the largest.
3. Compute $k=n\alpha$ (n times α)

If $n\alpha$ is an integer use $k=n\alpha$ and trim k observations at both ends. $R = remaining\ observations = n - 2k=n*(1-2*\alpha)$

$$T = \frac{1}{\lfloor R \rfloor} * \sum_{\lfloor k \rfloor+1}^{n-\lfloor k \rfloor} X_i$$

Note: α is ***half*** of the percent of trimming. Mean is $\alpha = 0\%$ trimmed mean. Median is $\alpha = \frac{1}{2} - \frac{1}{2*n}$

Example 1: Find 10% trimmed mean of 2, 4, 6, 7, 11, 21, 81, 90, 105, 121

$n=10, \alpha=0.10, k=n\alpha=1$, which is an integer so trim exactly one observation at each end, since $k=1.$ Thus trim off 2 and 121. we are left with $R=n-2k=10-2=8$ observations.

*10% trimmed mean = (1 / 8) * (4 + 6 + 7 + 11 + 21 + 81 + 90 + 105)*

If $n\alpha$ has a fractional part present, trimmed mean is a bit more complicated. In the above example, if we wanted 15% trimmed mean, $\alpha=0.15, n=10, k=n\alpha=1.5.$ Calculations yield k has an integer part 1, and a fractional part 0.5. $R=n-2k=10-2*1.5=10-3=7.$ Thus $R=7$ observations are retained. This R is then the correct denominator of trimmed mean.
Since integer part of k is 1, we throw out the smallest 1 observation, i.e., we throw out $Xmin=2.$ The next observation is 4. $k=1.5$, has fractional part 0.5, so we throw out only 0.5 part of 4, and keep the remainder 0.5 times 4, or 2. Similarly, at the high end we throw out $Xmax=121$, and throw out 0.5 part of the next largest (=105) and keep the rest (105*0.5)=52.5.

*15% trimmed mean = (1 / 7) * ((**0.5 *** 4) +6 + 7 + 11 + 21 + 81 + 90 + (0.5 * 105))*

Now find 27% trimmed mean. $\alpha=0.27, k=n\alpha=2.7,$ has a fractional part. $R=n-2k=10-5.4= 4.6$ observations are retained. This R is the denominator. We throw 2 numbers from each end, and also throw 0.7 parts of the third numbers (keeping their 0.3 parts):

*27% trimmed mean = (1 / 4.6) * ((0.3 * 6) + 7 + 11 + 21 + 81 + (0.3 * 90))*

General formula for trimmed mean

$$T = \frac{1}{n(1-2*\alpha)} * \left((1-r) * (X_{g+1} + X_{n-g}) + \sum_{i=g+2}^{n-g-1} X_i \right)$$

X – sorted numerical vector
n – length of X
α – proportion to trim from each end
g – **integer** part of the $n * \alpha$, or floor
r – **fraction** part of $n * \alpha$, $n * \alpha - g$