Computation of the semi-variance

by Jean-Victor Côté

December 26, 2011

An algorithm proposed by A. Björk in 1978 to the authors Chan, Golub and LeVeque from Stanford, as cited by them on page 243 of the paper cited on the page you gave an address to, should do the job nicely for you. The latter authors call it the "corrected two-pass algorithm", so you might find useful references about this on the Web. The first pass is, as in the textbook algorithm, to compute the mean, as a shift parameter for the data. Other shift parameters could be used, but for your purpose, I suggest that you use the mean, because you can use it to filter the elements to keep only the observations which are under the mean to go into the computation of the sum of squares.

Let $d_i = x_i - \bar{x}$, where x_i is observation i for all i in the set of indices I and $\bar{x}$ is the corresponding sample mean.

Let $\{n_i\}$ be the subset of $\{d_i\}$ such that $n_i < 0$ for the subset of I thus selected, called N (for negative).

Let n be the number of observations which are smaller than the mean, i.e. the number of indices in N.

Then the sum of squares S_N, restricted to the x_i smaller than $\bar{x}$, is given by the following formula (derived from formula 1.7, page 243 in [1] or formula S_{22} in [2]):

$$S_N = \sum_{i \in N} d_i^2 - \frac{1}{n} \left(\sum_{i \in N} d_i \right)^2$$

It is similar to the textbook formula, except that the textbook formula sums the x_i over all observations, in order to compute the sum of squares for the variance, whereas the formula just derived sums the differences $(x_i - \bar{x})$ over all i for which $(x_i - \bar{x}) < 0$, in order to compute the sum of squares for the semi-variance. Please note that you still have to divide by $(n-1)$ to get the semi-variance.

Both papers score that kind of formula highly for variance computation, so that you should not go wrong by using the formula I derived from it in order to compute the semi-variance.

This formula could be refined, using for instance the pairwise summation algorithm described on page 243 in [1], but such refinement and complication might not be warranted for.

References

[1] Chan, Tony F.; Golub, Gene H.; LeVeque, Randall J. (1983). Algorithms for Computing the Sample Variance: Analysis and Recommendations. The American Statistician 37, 242-247.

[2] Ling, Robert F. (1974). Comparison of Several Algorithms for Computing Sample Means and Variances. Journal of the American Statistical Association, Vol. 69, No. 348, 859-866.