The lost variance

One-to-one correspondence between central values and variances is *sine qua non* in mathematical statistics. In geostatistics, however, one-to-one correspondence between distance-weighted averages and variances is *null and void*. Following are formulae for the distance-weighted average, for the lost variance of the set, and for the lost variance of the distance-weighted average:

Distance-weighted average

$$\bar{x} = \sum_{i=1}^{n} \left(w_i \cdot x_i \right)$$

where: w_i = *ith* weighting factor
x_i = *ith* independently measured value
n = number of measured values in set

Variance of set

$$\text{var}(x) = \sum \left[n \cdot w_i \cdot \left(\bar{x} - x_i \right)^2 \right] \div \left[\left(1 \div \sum w_i^2 \right) - 1 \right]$$

where: n = number of measured values in set
w_i = ith weighting factor
$(1 \div \sum w_i^2) - 1$ = degrees of freedom

Variance of distance-weighted average

$$\text{var}(\bar{x}) = \text{var}(x) \cdot \sum w_i^2$$

where: var(x) = variance of set
$\sum w_i^2$ = sum of squared weighting factors

So who lost the variance of a ***single*** distance-weighted average? And who found the variance of a ***set*** of distance-weighted averages? Alas, the geostatocracy does not respond to questions that require rationally reasoned responses!

Matheron made up the term *geostatistics* because of his mind-boggling rationale that geologists stress *structure* and statisticians stress *randomness*! He claimed, *"A statistician who is not familiar with mining may well be discouraged before he can even get a good idea of the problem at hand"*. Matheron may well have been familiar with mining but certainly not with mathematical statistics. As a matter of fact, Matheron and his staff at the newly minted Centre de Géostatistique did not even know that each distance-weighted average has its own variance. So it is a profound mystery how the variance of a ***set*** of distance-weighted averages could possibly have turned into one of the cornerstones of Matheronian geostatistics.

Matheron and his disciples were not the only thinkers afflicted with statistical dyslexia. Krige himself did not know that distance-weighted averages had variances before they were reborn as honorific kriged estimates. David's 1977 textbook mentions the *"famous"* central limit theorem but omits the variance of the distance-weighted average. Krige, in his *Preface* to David's textbook, recalled *"some stormy receptions"* but did not reveal who objected to what and why. Philip and Watson's 1986 *Matheronian Geostatistics, Quo Vadis*, was mocked and thrashed while the new science kept kriging and smoothing along to nowhere.

Neither David nor Krige knew that weighted averages converge on the arithmetic mean and variances of weighted averages on the central limit theorem as all the weighted factors converge on ***1/n***. Area, count, density, length, mass and volume-weighted averages, unlike distance-weighted averages, do have variances. So why do the textbooks published since 1977 ignore the undeniable fact that variances of weighted averages converge on the ubiquitous central limit theorem?

Journel, an early minion of Matheron, mentioned in 1992, *"spatially distributed data should be considered a priori as dependent one to another, unless proven otherwise",* but rejected "*Fischerian"* [sic!] statistics to prove *otherwise.* Journel mentions a zero kriging variance on page 308 of *Mining Geostatistics*. Armstrong, true to form uninformed in 1989 of the unassailable statistical fact that distance-weighted averages have variances, was so troubled by the fall of kriging variances that she cautioned against oversmoothing. Of course, zero variances suggest severe oversmoothing, which Armstrong too, much to her credit, found hard to believe. Regrettably, she may still not comprehend what the difference between functional dependence and spatial dependence is all about.

Clark's hypothetical uranium data in *Practical Geostatistics* do not display spatial dependence but the author interpolates within the sample space all the same. Most distance-weighted averages in the infinite set defined by Clark's data converge on the arithmetic mean while their variances converge on the central limit theorem.