Sampling and Statistics Explained

Towards commonsensical sampling practices and scientifically sound statistical methods

J W Merks

Chapter 1 Introduction

Sampling is the practice of selecting a part of a whole and measuring the central value of the stochastic variable of interest in the part in such a manner that it is an unbiased estimate for the unknown true value of the whole, and that this estimate is precise within acceptable and affordable confidence limits. Trueness, accuracy, bias, and precision are fundamental concepts in sampling theory and practice. Statistical methods as defined in mathematical statistics are indispensable in sampling practice.

Sampling and statistics sounds simple but if it were, the Bre-X salting fraud would have been detected based on bogus grades for the first three to five boreholes, and Hecla's Grouse Creek mine would have made the predicted grade. In the Bre-X case, geostatistics converted bogus grades and barren rock into a phantom gold resource. In the Hecla case, geostatistics converted a small set of native gold stringers into a large block of high-grade ore. The question is then why the world's mining industry puts up with a variant of mathematical statistics that converted Bre-X's bogus grades and Busang's barren rock into a phantom gold resource, and caused too many reserves to shrink during mining.

Various aspects of this question will be examined at a later stage in a work about a cast of statistically challenged characters who hailed geostatistics as a new science in the early 1960s, and turned peer review into a blatantly biased, shamelessly self-serving sham. It is about mining professionals who displayed the same abysmal knowledge of mathematical statistics and disturbing lack of common sense by accepting the junk science of assuming continued mineralization between boreholes. Most of all, it will deal in detail with the chronology of events that began long before the Bre-X fraud unraveled and is bound to linger until the mining industry accepts scientifically sound statistical methods.

Generally, geostatistics is based on assuming spatial dependence between measured values in a temporally or *in situ* ordered set, interpolating by kriging, selecting the least biased subset of some infinite set of kriged estimates, smoothing its pseudo kriging variance to perfection, and rigging elementary rules of mathematical statistics with impunity. In the Bre-X case, assuming continued gold mineralization between ore zones in each line of boreholes in Busang's South East Zone was essential in converting bogus grades and barren rock into a phantom gold resource. Most Bre-X's investors lost money, many lost their life savings, and a few lost their lives. This is why all the facts behind geostatistics and its role in the Bre-X fraud must be recorded.

This text on sampling and statistics explains why geostatistics is an invalid variant of mathematical statistics, and why only mathematical statistics provides a scientifically sound basis to sampling practice and statistical analysis. The irrefutable statistical facts behind the genesis of geostatistics make a strange tale with more twists and turns than an Agatha Christie whodunit. This tale is not so much about who did what to whom and when but why the first generation of geostatistical scholars decided to replace the true variance of the ***single*** distance-weighted average with the false variance of a ***set*** of distance-weighted averages.

The distance-weighted average is the first and only weighted average whose variance vanished several years before it was reborn as an honorific *kriged estimate* to recognize the work of Professor D G Krige, the pioneering plotter of distance-weighted averages at the Witwatersrand gold reef complex in South Africa in the 1950s. Krige discovered that two or more gold assays, when determined in samples selected at positions with different coordinates in a finite sample space, define an infinite set of distance-weighted average gold assays. In those days, Krige did not know that each distance-weighted average has its own variance. By implication, he was unfamiliar with the requirement of functional independence and the concept of degrees of freedom in mathematical statistics. Otherwise, he would have known that two or more functionally dependent values such as distance-weighted averages gives exactly zero degrees of freedom. Therefore, variances of sets of distance-weighted averages cannot possibly give unbiased confidence limits for metal contents and grades of ore reserves.

It is beyond dispute that the variance of the single distance-weighted average vanished on Krige's watch at the Witwatersrand complex in the early 1950s. It is equally indisputable that this human error would not have turned into a scientific fraud if it were not for the work of Professor Dr G Matheron who found out about Krige's infinite sets of distance-weighted averages in the early 1960s. Matheron and his students did not know that one-to-one correspondence between distance-weighted averages and variances is inviolable in mathematical statistics. On the contrary, Matheron was so impressed with Krige's aptitude in augmenting small sets of gold grades in the Witwatersrand gold reef complex with infinite sets of distance-weighted average gold grades that he conferred on Krige the ubiquitous *krige* eponym. Matheron's résumé reveals that he taught probability theory and geostatistics, first at the *Nancy School of Mines* and later at the *Paris School of Mines* where he became the director of research at the *Centre de Morphologie Mathématique* in 1968, and established his *Centre de Géostatistique* in 1986.

Matheron, in his *Foreword* to Journel and Huijbregts's 1978 *Mining Geostatistics*, credits himself with coining the term *geostatistics* in the early 1960s because geologists stress structure whereas statisticians stress randomness. This strange dichotomy of structure and randomness may have inspired Matheron to prefix *geo* to *statistics* but it also underscores his tentative grasp of mathematical statistics in general, and of spatial dependence in particular. Initially, geostatistics was applied to sparse boreholes in ore deposits where Matheron and his statistically challenged following embraced the profusion of kriged boreholes with reckless abandon. In time, geostatistics was applied not only to sparse data in large sample spaces such as boreholes in ore deposits, tree species counts in forests, and shrimp counts along coastlines but also to dense data in small sample spaces such as bacteria counts in culture dishes. Nowadays, Krige's infinite sets of kriged estimates are largely obfuscated by software-driven kriging models that make Matheron's new science of geostatistics a fatally flawed but formidable tool to augment sparse data in small and large sample spaces alike.

Matheron's opinion about mathematical statistics and its practitioners is recorded in his *Foreword* to *Mining Geostatistics* in which he claims, *"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."* Perhaps ironically, Matheron did not have the foggiest idea of his own problem because one-to-one correspondence between distance-weighted averages and variances was as profound a mystery to him as it was to Krige when he discovered the infinite set of degrees-of-freedom and variance-deprived distance-weighted average gold grades at the Witwatersrand gold reef complex in South Africa in the 1950s.

Sir Ronald A Fisher was knighted in 1953 because of his work on analysis of variance. Matheron's synthesis of structure and randomness into his new science of geostatistics suggests that analysis of variances and the properties of variances were missing on the curriculum at the *Centre de Géostatistique*. Given his prejudice against mathematical statistics and its practitioners, it is not at all surprising that Fisher's work did not impact across the Channel where Matheron and his minions were transforming Krige's human error into a fundamentally flawed new science with single-minded determination.

Meanwhile, across the Atlantic, Professor Dr A G Journel had been teaching Matheron's new science for many years before he postulated in 1992 that spatially distributed data should be considered *a priori* as dependent one to another, unless proven otherwise. It appears that Journel, a Stanford professor and a dedicated disciple of Matheron, rather assumes than proves spatial dependence by applying analysis of variance. Evidently, Matheron's nascent geostatisticians were not taught how to verify spatial dependence between measured values in ordered sets, or how to derive unbiased confidence limits for metal contents and grades of ore reserves.

Fisher's F-test is the essence of analysis of variance (ANOVA). This test is routinely applied to verify whether two variances are statistically identical or differ significantly. In mineral exploration and mining, the F-test proves whether an ordered set of measured values displays a significant degree of spatial dependence, or is randomly distributed within its sample space. Paradoxically, it is Fisher's F-test that partitions uncertainties in sample spaces into Matheron's structure and randomness. A sampling variogram is a visual interpretation of Fisher's F-test, which shows where a significant degree of spatial dependence in a sample space dissipates into randomness. Sampling variograms are described in several ISO Standards such as those developed by ISO Technical Committee 69–*Applications of statistical methods*.

In retrospect, it is understandable that Krige was honored by his peers because he was the first mining professional who discovered that two or more gold grades, determined in samples selected at positions with different coordinates in a finite sample space, define an infinite set of distance-weighted average gold grades. Mining engineers and geologists are invariably short of reliable but expensive borehole grades, which explains why kriged borehole grades were embraced with statistical-dyslexia-driven exuberance. Some mining professionals expressed concern but few seemed to comprehend why kriging variances of sets of kriged boreholes are invalid measures for variability, precision, and risk as defined mathematical statistics. It is not surprising then that functional independence, degrees of freedom, and one-to-one correspondence between distance-weighted average borehole grades and variances, went missing in Matheron's new science. Yet, it is easy to prove heuristically that the variances of all weighted averages converge on the *Central Limit Theorem* as variable weighting factors converge on a constant weighting factor.

In his 1978 *Geostatistical Ore Reserve Estimation*, Professor Dr M David refers to the *"famous"* central limit theorem. Yet, he appeared unaware that variances of all weighted averages converge on that ubiquitous variance of the arithmetic mean as all weighting factors converge on ***1/n***. In *Chapter 12 Orebody Modelling* [sic] of this first textbook on geostatistics, the author mentions the infinite set of simulated values that a kriging model generates and suggests to apply conditional simulation, *"to make that infinite set smaller and get that model closer to reality."*

Figure 203 on page 286 shows sixteen (16) distance-weighted average holes in a four by four matrix, each of which is *"estimated"* (calculated!) from the same nine (9) holes. The author proclaims on the same page, *"Writing all the necessary covariances for that system of equations might be a good test to find out whether one really understands geostatistics."* In fact, counting the degrees of freedom for that system of equations is a good test to find out whether one really understands mathematical statistics. David's system of equations would not give a single degree of freedom. In contrast, a set of nine (9) holes gives $df=n-1=8$ degrees of freedom whereas the ordered set gives $df_o=2(n-1)=16$ degrees of freedom.

In *V.A. Theory of Kriging* of their 1979 *Mining Geostatistics*, Journel and Huijbregts, refer to the zero kriging variance. While it is true that distances between ordered kriged estimates in an infinite set are indeed infinitesimal, the question is why the authors even bothered to mention that bizarre zero kriging variance in their textbook. What they failed to explain is why the real variance of a ***single*** distance-weighted average was replaced with the pseudo kriging variance of a ***set*** of kriged estimates.

Pseudo kriging variances have the same dimension as real variances. However, pseudo kriging variances are as meaningless a measure for variability, precision, and risk as pseudo kriging covariances are for spatial dependence. Pioneering geostatistical scholars concocted pseudo kriging variances and pseudo kriging covariances of subsets of infinite sets of kriged estimates. Matheron's muddled thinking about synthesizing structure and randomness into geostatistics is at least partially responsible for the fact that it became an invalid variant of mathematical statistics. What is known with absolute certainty is that kriging variances and kriging covariances of least biased subsets of all sorts of infinite sets of kriged estimates became the cornerstones of geostatistics. Ruthlessly effective peer review was the key to enforcing the troubling tenets of Matheron's new science.

H G Wells (1866-1946) ranked statistical thinking as necessary for efficient citizenship as reading and writing. Geostatistical thinking is all about assuming spatial dependence between temporally or *in situ* ordered sets of measured values, interpolating by kriging, selecting the least biased subset of some infinite set of kriged estimates, smoothing its pseudo kriging variance to perfection, and rigging elementary rules of mathematical statistics with reckless abandon. Wells's opinion about statistical thinking may well have been inspired by Fisher's analysis of variance. Would Wells have ranked geostatistical thinking quite as necessary as reading and writing?

Sampling and statistics are important in mineral exploration, mining, processing, smelting and refining, in a wide range of engineering and scientific disciplines, and in the international trade of coals, mineral concentrates, ores, and scores of other materials in bulk. Some understanding of probability theory and mathematical statistics is required when applying statistical analysis to test results determined in all sorts of samples taken from all types of sampling units and sample spaces. Implementing and evaluating manual sampling protocols and designing mechanical sampling systems for solids and slurries also demand in-depth knowledge of sampling and statistics.

Generally, the objective of sampling is to estimate unknown true values of stochastic variables in sampling units or sample spaces. The concept of *unknown true value* is necessary to define the absence or presence of bias. The unknown true value should be estimated in an unbiased manner and with an acceptable and affordable degree of precision. How to test for bias and how to compute confidence intervals and ranges for estimates of those elusive unknown true values of stochastic variables in sampling units and sample spaces are key elements of metrology, the science of measurement.

Uncertainties in measurement chains or hierarchies can be partitioned into random variations and systematic errors or biases. The sum of all random variations is statistically identical to zero whereas the sum of all systematic errors or biases departs significantly from zero. The variance is the fundamental measure for the sum of all random variations in a measurement chain. The variance with its squared dimension stands in sharp contrast to the linear dimension of the sum of systematic errors.

The term e*rror*, with or without adjectives, is traceable to Gy's sampling theory. In this text, *error* will not be used. On the contrary, a compelling case will be made that Gy's *errors* be avoided altogether. Gy refers in 1953 to Visman's 1947 PhD thesis, in which this author postulates that the sampling variance for a heterogeneous sampling unit such as a mass of unwashed coal is the sum of the composition variance and the segregation variance. Visman's work is incorporated in ASTM D2234 on the sampling of coal, which is the first internationally recognized standard that includes a precision statement for ash content. In his later work, Gy become more introspective and refers mostly to his own work. Visman's work is discussed in detail to ensure that his priority in sampling theory and practice is recognized and recorded in the history of sampling and statistics.

Almost invariably, the primary sample selection stage adds most to the variance of the measurement procedure, and, thus, to the measurement variance. Typically, the sample preparation stage adds more to the measurement variance than the analytical stage. The additive property of variances in a measurement hierarchy is of critical importance when optimizing sampling protocols. This additive property is not necessarily commutative. The question of whether the difference between two variances is itself a valid variance estimate is solved by applying Fisher's F-test. In this application, too, the F-test is based on comparing the calculated F-value between the highest and lowest variances with tabulated values of F-distributions at 5% and 1% probability, which are listed as a function of the degrees of freedom for each of these variances.

A sampling unit is deemed a static stochastic system when a set of primary increments is selected from the wet mass of material contained in one or more bulk bags, trucks, wagons, or holds of a vessel. A sampling protocol conforms to the ideal equiprobable model when a properly designed sample probe penetrates the entire particle stratum such that each primary increment represents a single cell of the set into which the content of a bulk bag, truck, wagon, or hold is divided. The dimension of the sampling ratio between the mass of a primary increment and the mass it represents is a probability. Mechanical sampling probe systems are widely used to sample mineral concentrates and recycled materials with high precious metal contents.

A sampling unit is deemed a dynamic stochastic system when a set of primary increments is selected during transfer from one stationary situation to another. Stratified systematic sampling at constant mass intervals conforms closest to the ideal equiprobable sampling model. Stratified systematic sampling at constant time intervals tends to depart from the ideal model when handling rates become highly variable. Stratified systematic sampling is applied in mechanical systems for coals and ores at bulk handling terminals all over the world. The advantage of stratified random sampling is that periodic changes in the stochastic variable of interest are less likely to become synchronized with intervals of constant mass or time. Periodic changes are more likely to synchronize during loading of ships from unit trains than during discharge of ships. A problem with stratified random sampling is that, from time to time, a pair of primary increments is taken in rapid succession so that the secondary system should be designed to handle such surges in primary sample mass flow.

Interleaved sampling protocols provide precision estimates at little or no additional costs. Interleaved primary samples are obtained by partitioning a set of primary increments into a pair of subsets such that one consists of odd-numbered increments (A-primary sample in ISO jargon) and the other of even-numbered increments (B-primary sample). The variance of test results determined in test portions of test samples prepared from A- and B-primary samples is a reliable estimate for the measurement variance (the sum of the variances of the primary sample selection, sample preparation and analytical stages). Test results for duplicate test portions taken from test samples give an unbiased estimate for the analytical variance. Test results for duplicate test samples, prepared of A-primary samples, of B-primary samples, or of both primary samples, give unbiased estimates for the sum of the variances of the sample preparation and analytical stages.

A pair of interleaved primary samples gives only a single degree of freedom. Shipments of coals, concentrates and ores are often divided into four of more lots, which give no less than four degrees of freedom for a tabulated t-value of $t_{0.05;4}=2.776$. Daily shifts give 27 to 30 degrees of freedom for monthly production periods for a tabulated t-value of no less than $t_{0.05;27}=2.052$. Mechanical sampling systems can be modified to routinely select A- and B-final system samples by alternating the mass flow from the final division stage between two receptacles in time-delayed synchronization with the primary sampler. Interleaved sampling protocols generate precision estimates at the lowest possible cost, which explains why interleaved sampling is rapidly becoming the *de facto* standard in sampling practice.

Sampling theory and practice in mineral exploration and mining require the concept of the sample space. A single borehole may intersect several ore zones with sufficient mineralization to permit classification as ore grade material that can be mined and processed at a profit. When a borehole intersects mineralization, test results for *in situ* ordered core samples in each ore zone may display a significant degree of associative dependence, or *spatial dependence* in this context. Spatial dependence gives a higher degree of precision for the central value of an *in situ* ordered set. Spatial dependence also makes it possible to construct a down-hole sampling variogram and convert a linear set of measured values into a cylindrical volume of *in situ* ore, whose grade and content can be reported with a 95% confidence limits. In other words, an ore zone defines its own sample space whenever test results for a set of *in situ* ordered core samples display a significant degree of spatial dependence.

When boreholes intersect massive sulfide mineralization, test results determined in sets of *in situ* ordered core samples within ore zones are more likely to display stratification than spatial dependence. In such cases, it makes sense to partition contiguous core sections at visual strata boundaries, and measure the lengths and densities of core samples before whole core is divided into halves. It is common practice to prepare and assay one half, and wrap, seal, and store the other half for further testing. In the case of massive sulfides, the central value of a set of core samples is the length- and density-weighted average of all the test results in the mineralized zone.

Verifying spatial dependence between central values of ore zones within boreholes on a line proves or disproves in probabilistic terms the continuity of mineralization between ore zones. If the probability to encounter continued mineralization during mining were to exceed 95%, it would be justified to replace cylindrical volumes with contiguous blocks, and to compute 95% confidence intervals and ranges not only for the cumulative content of all blocks but also for the weighted average grade of the ore reserve. If the probability were less than 95%, it would still be justified to prove a reserve within a resource based on the set of boreholes that defined cylindrical volumes, and compute 95% confidence intervals and ranges for the metal content and grade of the cumulative volume and mass of the reserve within the resource.

Variances of metal contents of volumes of in situ ore, and of masses of mined ore, mill feed, concentrate and tailing are additive. The additive property of variances of contained metals can be applied in process simulation in mineral exploration, mining, processing, smelting, and refining. It is easy to set up simulation models with spreadsheet software and generate confidence limits for derived statistics.

Extraneous variances such as those associated with dividing whole core samples into halves, and with selecting and assaying test portions of test samples, may be subtracted before a down-hole sampling variogram is constructed and the sample space of a borehole is derived. The variance of selecting and assaying test portions of test samples is, in fact, the analytical variance. In geostatistics, the analytical variance turned into a *nugget effect*, a misnomer that belies its squared dimension is that of a true variance. It becomes irrelevant in massive sulfides and rather ridiculous when counting tree species in forests, shrimp along shorelines or bacteria in culture dishes.

Interleaved sampling protocols should be applied when selecting bulk samples of crushed ore from adits, drifts, pits, or trenches. Interleaved bulk samples give a reliable estimate for the measurement variance (the sum of the variances of the primary sample selection, sample preparation and analytical stages), which adds to the intrinsic variance of *in situ* ore, and may be subtracted before a sampling variogram is constructed. The variances of selecting face and wall samples in adits, drifts, pits or trenches are significantly higher than those of interleaved bulk samples. Such variances, too, are extraneous to the intrinsic variance of a stochastic variable in a sample space, and may be subtracted to obtain the least biased estimate of the intrinsic variance of a stochastic variable within is sample space. Bulk samples provide the highest possible degree of precision for the intrinsic variance of the stochastic variable of interest.

Spatial dependence between metal grades of ore zones within lines or profiles makes it possible to compute 95% confidence intervals and ranges for metal contents and grades as a measure for the risk associated with mining ore reserves. In contrast, geostatistically inflated mineral inventories make more compelling annual reports and tend to enhance performance bonuses but are bound to shrink during mining when recovered tonnages and grades are lower than predicted, and when the statute of limitations for commercial fraud may well have expired.