Sampling and Statistics Explained

Towards commonsensical sampling practices and scientifically sound statistical methods

J W Merks

Preface

Sampling and statistics have found wide application in science and engineering, and play important roles in the international trade of commodities, goods, and services. Sampling theory evolved from games of chance, mathematical probability, properties of probability distributions, and parameters of homogeneous populations with defined characteristics and compositions. The most relevant parameters in sampling theory are population means and variances of all types of homogeneous populations.

The transition from sampling theory with homogeneous populations to sampling practice where sample spaces and sampling units display variable degrees of heterogeneity require thorough knowledge of mathematical statistics. Many applications in sampling practice are based on the properties of variances. The requirement of functional independence and concept of degrees of freedom are of critical importance. The most relevant statistics in sampling practice are central values and variances of sets of independently measured values with equal or variable weights. An essential element of sampling practice is the verification of spatial dependence between independently measured values in temporally or *in situ* ordered sets in small and large sample spaces alike. Associative dependence between independently measured values of two or more stochastic variables in the same sample space may provide valuable information.

The arithmetic mean is the central value of a set of independently measured values with equal weights whereas a weighted average is the central value of a set of independently measured values with variable weights. *Independently measured* implies that neither the primary sample selection stage, nor the sample preparation stage or the analytical stage cause a significant degree of associative dependence between contiguous subsets of the whole set. *Measured value*, without the adjective *independent,* is used predominantly because contiguous subsets of test results do not often display associative dependence at some stage in a measurement chain or hierarchy.

All weighted averages converge on the arithmetic mean as variable weights converge on the same weight. Each central value is a functionally dependent *(calculated)* value of a set of measured values. Generally, a set of n measured values gives $df=n-1$ degrees of freedom, and the ordered set gives $df_o=2(n-1)$ for the first variance term. The number of degrees of freedom is a positive integer for a set of measured values with equal weights, and a positive irrational for a set of measured values with variable weights.

The effectiveness of interleaved sampling protocols derives from the fact that a pair of measured values gives a single degree of freedom. Interleaved sampling protocols give reliable precision estimates for shipments of coals, concentrates, ores, industrial minerals, and other materials in bulk when sampled where custody changes between sources and destinations. Interleaved sampling protocols are easy to implement in mechanical sampling systems and manual sampling procedures. Interleaved sampling is incorporated in a number of internationally recognized standards including those approved by ISO Technical Committee 69–*Applications of Statistical Methods*.

S*patial dependence* implies a statistically significant degree of associative dependence between *in situ* or temporally ordered sets of measured values. Analysis of variance verifies whether an ordered set of measured values displays a significant degree of spatial dependence in its sample space. Interpolation between measured values is permissible if the ordered set displays a significant degree or spatial dependence. However, if the ordered set does not display a significant degree of spatial dependence, interpolation is not permissible and extrapolation beyond the sample space is a scientific fraud. It may be necessary at times to take into account a significant degree of associative dependence between measured values of two or more stochastic variables, when determined in samples selected at positions with different coordinates in the same sample space.

The primary objective in sampling practice is to obtain an unbiased estimate for the *unknown true value* of a stochastic variable in a sample space or a sampling unit. Such an estimate should not only be unbiased but also precise within acceptable and affordable confidence limits. The concept of *unknown true value* is important when testing for the absence or presence of bias at the primary sample selection stage, the sample preparation stage, or the analytical stage in a measurement hierarchy or chain. Testing for bias at the primary sampling stage is much more complex than testing for analytical bias.

The variance is the fundamental measure for variability. The properties of variances are of critical importance in sampling practice and statistical analysis because variances are amenable to mathematical analysis. Due to its squared dimension, the variance is less intuitive a measure for variability than the standard deviation, which has the same dimension as the central value of a set of measured values. In contrast, the coefficient of variation is equal to the standard deviation as a percentage of the central value. Equally intuitive and user-friendly measures for precision are 95% confidence intervals and symmetric and asymmetric 95% confidence ranges. Confidence intervals are reported in absolute units and relative percentages whereas lower and upper limits of symmetric and asymmetric confidence ranges can only be reported in absolute units.

Testing for bias and estimating precision are key elements of sampling practice. Student's t-test is the bias test *par excellence* for paired or pooled sets of measured values. The power of the t-test to detect a bias is reported in terms of bias detection limits for the Type I statistical risk only, and for the combined Type I and Type II statistical risks. If the difference between central values of paired or pooled sets of measured values implies the presence of bias, the probable bias ranges for the observed difference can be reported with the same statistical risks. While it is true that the terms *bias* and *systematic error* are deemed synonyms, the noun *error* without the adjective *systematic* is avoided in defining statistical risks and throughout the text for reasons to be discussed in due course.

Analysis of variance (ANOVA) verifies whether variances are statistically identical or differ significantly. Fisher's F-test is the essence of ANOVA. The F-test verifies whether an observed degree of spatial dependence is statistically significant by comparing the observed F-value between the variance of the set and the first variance term of the ordered set with tabulated values of F-distributions at 5% and 1% probability and proper degrees of freedom for each of these variances. Plotting statistically significant variance terms of the ordered set against the variance of a set and the lower limits of its asymmetric 99% and 95% confidence ranges gives a sampling variogram. Sampling variograms show where spatial dependence in sample spaces dissipates into randomness.

The terms *random* and *order* are used side-by-side in table headers to show how spatial dependence impacts precision estimates for central values. The numerical value of the *variance of set* does not depend on the order in which the differences between measured values and its central value are squared and summated. In contrast, the order in which the differences between contiguous measured values are squared and summated impacts the *first variance term of ordered set*. A statistically significant degree of spatial dependence gives a higher degree of precision for the central value of a set of measured values, and does so at no additional cost. The terms *variance of set* and *first variance term of ordered set* are unambiguous when comparing observed F-values with appropriate tabulated F-values to determine whether the degree of spatial dependence is statistically significant at 95% or 99% probability.

Fisher's F-test is also applied to estimate the intrinsic variance of a stochastic variable in its sample space, and to optimize a sampling protocol by partitioning the measurement variance into components. It played a key role in unraveling the Bre-X fraud by proving that the intrinsic variance of Busang's gold was statistically identical to zero.

Various internationally recognized standards describe how to test mechanical sampling systems and manual sampling procedures for absence or presence of bias, how to apply statistical analyses to test results for bias test programs, and how to compute unbiased confidence limits for metal contents and grades of shipments of concentrates and ores. The methodology provides confidence limits for metal contents and grades of mill feed, tailing, and concentrate. It can be applied to simulate confidence limits for functionally dependent values such as recoveries in mineral processing plants.

Although the same methodology provides confidence limits for metal contents and grades of ore reserves, geostatistical reviewers deemed it at variance with their own practice. Paradoxically, geostatistics itself is at variance with mathematical statistics because it violates the requirement of functional independence and ignores the concept of degrees of freedom. These interrelated statistical inferences did not trouble those whose tenuous grasp of mathematical statistics is traceable to Matheron's teachings at his celebrated *Centre de Géostatistique*, Fontainebleau, France.

Matheron's misgivings about mathematical statistics and its practitioners are preserved for posterity in his *Foreword* to Journel and Huijbregts's 1978 *Mining Geostatistics*. Predictably, his prejudice against mathematical statistics bred endemic statistical dyslexia at his *Centre de Géostatistique*. So much so that the variance of the distance-weighted average that went missing on Krige's watch at the Witwatersrand complex in the 1950s, was still missing when Matheron's new science evolved at his *Centre de Géostatistique*. It is an irrefutable statistical fact that each weighted average had its own variance when Sir Ronald A Fisher was knighted in 1953 for his contribution to analysis of variance. When Matheronian geostatistics was hailed as a new science in the early 1960s, the distance-weighted average turned out to be the first and only weighted average that was reborn as an honorific but variance-deprived *kriged estimate*.

Incredibly, *kriging variances* and *kriging covariances* of ***sets*** of *kriged estimates* became the cornerstones of geostatistics. Incredible indeed because the pseudo variance of a ***set*** of distance-weighted averages is as meaningless a measure for variability, precision, and risk as its pseudo covariance is for spatial dependence. David's 1977 *Geostatistical Ore Reserve Estimation*, Journel and Huijbregts's 1978 *Mining Geostatistics*, Clarke's 1979 *Practical Geostatistics*, and similarly flawed works, show that the true variance of the ***single*** distance-weighted average has indeed been replaced with the false variance of a ***set*** of distance-weighted averages. Predictably, none of these early textbooks refers to the requirement of functional independence and the concept of degrees of freedom.

Two or more independently measured values, determined in samples selected at positions with different coordinates in a finite sample space, define not only an infinite set of distance-weighted averages but also a correspondingly infinite set of variances. One-to-one correspondence between distance-weighted averages and variances is inviolable in mathematical statistics but irrelevant in geostatistics where kriged estimates and kriging variances are not similarly paired.

The question is not so much who lost the variance of the distance-weighted average but who replaced the genuine variance of the distance-weighted average with the pseudo kriging variance of a set of kriged estimates. The question is simple but the geostatocracy is silent. In fact, the first generation of geostatistical scholars neither admits nor accepts that each kriged estimate has its own variance. On the contrary, the *krige* inspired eponym spawned a cult like lingo of neologisms, the most mind numbing of which are zero kriging variances and unity kriging covariances of infinite sets of kriged estimates. The rise of kriging covariances and the fall of kriging variances troubled geostatistical scholars that a caution against oversmoothing was issued. Apparently, the requirement of functional independence may be violated a little but not a lot.

The requirement of functional independence and the concept of degrees of freedom in mathematical statistics invalidate kriging variances and kriging covariances of ***sets*** of kriged estimates just as much as the laws of thermodynamics shattered the prospect of perpetual motion. A compelling case can be made that the requirement of functional independence and the concept of degrees of freedom in mathematical statistics evolved to ensure that geostatistics is doomed to genuine junk science status.

The practice of geostatistics becomes even more peculiar upon realizing that each kriging method gives it own infinite set of kriged estimates. Selecting the least biased subset of an infinite set of kriged estimates and smoothing its pseudo kriging variance to perfection is a daunting task indeed. A different kriging method would give another infinite set of kriged estimates, another least biased subset, and another perfectly smoothed pseudo kriging variance. Geostatistics is about selecting the very subset that seamlessly matches the stochastic variable of interest in the sample space under examination, and gives the least biased estimate for its unknown true value.

In geostatistics, the least biased estimate is called the *Best Linear Unbiased Estimator* for which *BLUE* is the approved acronym. Unlike those of winning *6/49* lotteries, the odds of selecting BLUEs of infinite sets of kriged estimates are immeasurable. Despite staggering odds, geostatisticians routinely select BLUEs in small and large sample spaces alike. An insidious problem is that functionally dependent BLUEs do not give degrees of freedom and cannot possibly provide unbiased confidence limits for contents and grades. A recent addition to the geostatistician's toolbox is the *Kolmogorov-Wieder Best Linear Unbiased Prediction*. It is highly improbable that the *KWBLUP* method gives unbiased confidence limits for metal contents and grades of ore reserves.

Scores of *krige*-inspired eponyms obfuscate the irrefutable fact that geostatistics derives from mathematical statistics. The supporters of this new science continue to search for a distinct terminology that underscores its uniqueness. For example, *variance of set* became *sill value*, *analytical variance* turned into *nugget effect*, and *sampling variogram* metamorphosed into *semi-variogram* or *variogram*. Although the differences between the semi-variogram and the variogram are irrelevant, they do spark peculiar squabbles among geostatistical scholars. Paradoxically, such squabbles are exercises in futility because only the sampling variogram shows where a significant degree of spatial dependence in a sample spaces dissipates into randomness.

Calculating 95% confidence limits for metal contents and grades of ore reserves would become infinitely more practical and scientifically sound if the mining industry were to accept the incontrovertible statistical fact that each distance-weighted average has its own variance. Matheron's new science would return to its old roots in mathematical statistics where all weighted averages have variances, and the distance-weighted average is no exception. Sampling variograms would display where statistically significant degrees of spatial dependence in sample spaces dissipate into randomness. Confidence limits for metal contents and grades would be routinely reported in annual reports as measures for the risks associated with mining ore reserves. Matheron's new science of geostatistics, as it stands, is beyond salvation.

Mathematical statistics, unlike geostatistics, complements sampling practice in a logical and intuitive manner. Not surprisingly, sampling and statistics are inseparable subjects in a number of internationally recognized standard methods. ISO Technical Committee 69–*Applications of statistical methods*, deals with statistics and sampling for a broad range of practical applications for the natural resource industry in general, and for mineral exploration, mining, processing, smelting, and refining in particular.