Sampling in Mineral Processing

J W Merks[1]

ABSTRACT

The objective of sampling in mineral processing is to estimate grades and contents of sampling units in an unbiased manner and with an acceptable and affordable degree of precision. Sampling units are classified as dynamic and static stochastic systems. The paper examines the most relevant topics of sampling practice and applied statistics such as how to test for bias, how to estimate precision, how to quantify associative dependence between measured values in ordered sets, how to select suitable sampling procedures and how to optimize sampling protocols. Examples are given to illustrate the application of sampling theory in practice. Extensive references to publications describing sampling procedures and guidelines for mineral processing applications are provided.

INTRODUCTION

Sampling theory and practice play an important role in mining and metallurgy. The sampling of materials in bulk is well-documented in the literature (Gy 1979; Merks 1985; Visman 1962) but concise definitions, uniform symbols and common rules remain elusive targets. Various *Technical Committees* (TCs) of the *International Organization for Standardization* (ISO) have developed guidelines on the sampling of coal (TC27), iron ore (TC102), and copper, lead and zinc concentrates (TC183). Detailed information can be found in several ISO Draft International Standards for copper, lead and zinc concentrates (see *References, ISO/DIS*), and in standard methods developed by ISO/TC69-*Applications of Statistical Methods*.

Generally, sampling is the process of selecting a part of a whole such that a measured value for the part is an unbiased estimate for the whole. In mineral processing, a *whole* is referred to as a sampling unit such as a mass of mill feed, dewatered concentrate or bullion, or a volume of cyclone overflow or tailings slurry. A sampling unit is classified as a dynamic stochastic system when sampled during transfer, and as a static stochastic system when sampled while stationary.

The wet mass of mill feed can be estimated in an unbiased manner and with an acceptable degree of precision (Merks and Merks 1992) but SAG mills, and gravity and flash concentrates, have made it difficult to obtain precision estimates for metal grades and contents of mill feed. The variances of metals contained in tailings, concentrates and thickener inventories can be used to obtain reliable precision estimates for monthly mill feed grades. It is beyond the scope of this paper to explain how Monte Carlo simulations can be applied to estimate confidence limits for the metal grade of mill feed on the basis of its wet mass, and of the metal contents and variances of tailings and concentrates (Merks 1991; 1999). The validity of this method depends critically on how slurry flows in mineral processing plants are interrogated, and how the variances of stochastic variables are estimated. On-stream data give valuable statistics that can be plotted in sampling variograms to show where orderliness in the sample space of time dissipates into randomness (Merks 1999).

Applied statistics provides scores of powerful techniques to test for bias, to estimate precision, to optimize sampling protocols, and to determine the degree of associative dependence between measured values in ordered sets (ASTM 1985; Mandel 1964; Merks 2000; Volk 1980). The statistical analyses applied in this paper are based on comparing F- and t-statistics with values tabulated as a function of degrees of freedom in the F-distributions at 5% and 1% probability, and in the t-distribution (Handbook 1968; Volk 1980).

[1] Matrix Consultants Limited, Vancouver, British Columbia

A sampling protocol that is based on dividing a set of primary increments into odd- and even-numbered subsets (see *Appendix A*), and preparing a test sample of each primary sample (see *Appendix B*), gives an unbiased estimate for the variance of the entire measurement chain. A pair of interleaving (or interpenetrating) primary samples is referred to as A- and B-samples (ISO/DIS 12743). The symbols *var(t)* and *var(spa)* refer to the total variance of a measurement procedure (the sum of the variances of the primary sample selection, preparation and analytical stages). Interleaving sampling protocols are equally effective when applied to slurry flows in mineral processing and bulk samples in mineral exploration.

Uncertainties in a measurement chain can be partitioned into randomly distributed variations (*random variations* for simplicity) and biases. The sum of all random variations is statistically identical to zero, and the sum of all biases is statistically different from zero. The variance is the fundamental measure for random variations. Analysis of variance (ANOVA) can be applied to optimize sampling protocols by partitioning the sum of the variance of the primary sample selection stage, the variance of the sample preparation stage and the variance of the analytical stage into its components. This application of Fisher's F-test, which is the essence of ANOVA, is examined in a separate section.

DEFINITIONS

Through the years, probability theory and applied statistics have developed a distinct jargon. Since statistical tests and techniques are applied in all scientific and engineering disciplines, it is unsurprising that vastly different definitions and symbols abound. Elementary concepts such as *trueness, accuracy, bias* and *precision* received a great deal of attention and scrutiny from ISO/TC69–*Applications of Statistical Methods*, and from technical committees that deal with the sampling of concentrates, coals and various types of ores. In time, ambiguous terms such as *measurement error, margin of error, sampling error, sill value, semi-variogram, nugget effect,* and scores of others, will be replaced with concise definitions, and be assigned the proper statistical symbols.

Accuracy

A generic term that implies closeness of agreement between a single measured value or the central value of a set (the arithmetic mean or a weighted average), and the unknown true value of the stochastic variable.

This definition reflects that *accuracy* is an abstract concept. By contrast, a lack of accuracy can be measured and quantified in terms of a *bias* or *systematic error*. Webster defines *accuracy* as *free from error*. Thus, unbiased measurements are accurate by definition. The term *unbiased* implies that a properly designed bias test was applied, and that a single measured value or the central value of a set is indeed an unbiased estimate for the unknown true value of the stochastic variable in the sampling unit or sample space under examination.

Bias

A statistically significant difference between a single measured value or the central value of a set, and an unbiased estimate of the unknown true value of the stochastic variable.

Testing for the absence or presence of bias is an essential part of sampling in mineral processing. Terms such as *random error*, or *error* without adjuncts or adjectives, will not be used to avoid confusion with randomly distributed variations for which the variance is the fundamental and unambiguous measure.

Testing for relative bias and estimating analytical precision are key elements of statistical quality control (SQC) and statistical grade control (SGC). Testing for analytical bias demands the use of *Certified Reference Materials* (CRMs). The presence of bias at the analytical stage is easy to detect but sometimes difficult to eliminate at affordable cost. Some sources of bias at the primary sample selection stage and the sample preparation stage are intrinsic to the applied procedure, which makes a bias difficult to detect and impossible to eliminate.

Student's t-test, the bias test *par excellence*, is described in a separate section. The t-test can be applied to paired test results obtained by employing different analytical methods to replicate test portions taken from each of a set of test samples. The test can also be applied to paired test results obtained by employing different sampling procedures to the same sampling unit, or different sample preparation procedures to the same primary or secondary sample.

The presence of analytical bias suggests that at least one of the procedures is suspect, and the absence of analytical bias implies that both procedures are most probably unbiased (ASTM 1985; Davies and Goldsmith 1972; Mandel 1964; Merks 1985; Volk 1980). The t-test should be routinely applied to assays determined at mineral processing plants and control assays reported by commercial laboratories, and to exchange assays between mines and smelters (Merks 1989).

One-way ANOVA is applied to test results determined in the same laboratory by employing the same analytical method to replicate test portions taken from each of five up to ten test samples prepared of the same sample mass under carefully controlled conditions. Tests for homogeneity should precede crosscheck programs to ensure that each participating laboratory receives a subset of test samples selected from a homogeneous set.

Two-way ANOVA is employed to test for analytical bias when three or more laboratories participate in interlaboratory crosscheck programs by applying the same analytical technique to duplicate test portions taken from each of a set of no fewer than five test samples. Logically, the complete set should pass the test for homogeneity before subsets are submitted to the participating laboratories.

Precision

A generic term that refers to the magnitude of randomly distributed variations (random variations) in the measurement procedure applied to estimate the central value of the stochastic variable of interest.

Precision, too, is an abstract concept. For example, the precision is low or poor, or the degree of precision is high or excellent, are valid but ambiguous, non-quantitative and vacuous just the same. Quantitative measures for precision such as confidence intervals in absolute values or relative percentages, and symmetric and asymmetric confidence ranges in absolute values, derive from the variance of the central value for the stochastic variable of interest.

Sample

A part of a sampling unit or a sample space selected such that a measured value for the part is an unbiased estimate for the sampling unit or the sample space.

A sample is often referred to as *a representative part* of a population or a whole but the concept of representativeness is widely abused and misused in sampling practice (Huff 1954). In reality, the measured value for a sample is an unbiased estimate for the sampling unit if, and only if, each stage of the applied measurement procedure is unbiased.

Interleaving Test Samples

A pair of test samples obtained by dividing a set of primary increments into odd- and even-numbered subsets (A- and B-primary samples), and preparing a test sample of each primary sample (A- and B-test samples).

Selecting a pair of interleaving primary samples (see *Appendices A & B*) is the most effective procedure to obtain an unbiased estimate for the variance of an entire measurement chain. Since one pair gives only a single degree of freedom (see *Degrees of Freedom*), the estimate for *var(t)*, the total variance of the measurement procedure, is extremely imprecise (see Table 3). By contrast, 28−31 pairs of interleaving samples give 27−30 degrees of freedom so that the monthly metallurgical balance is significantly more precise than single daily metallurgical balances.

SAMPLING PROTOCOLS

Generally, sampling protocols can be divided into random sampling, stratified random sampling and stratified systematic sampling. Random sampling and stratified random sampling are routinely applied to consumer products. Stratified systematic sampling is most effective for all types of materials in bulk and for slurry flows in mineral processing plants.

Stratified Systematic Sampling

This sampling protocol is commonly applied to bulk solids such as concentrates, coals and ores, preferably during transfer with a conveying system (dynamic stochastic system) but often while in storage (static stochastic system). Stratified systematic sampling is based on dividing the sampling unit into a set of elementary units (dynamic strata or static cells), and selecting a primary increment from each elementary unit. For example, dewatered or dried concentrate is divided into dynamic strata during transfer with a conveying system by selecting a set of primary increments at intervals of constant mass or time. Dewatered or dried concentrates should not be sampled mechanically during transfer because significant moisture losses will inevitably occur.

Concentrate in trucks or wagons is divided into static cells, and a primary increment is selected from the center of each cell with a properly designed probe. Mechanical probe sampling systems make it simple to select pairs of interleaving samples and implement effective risk analysis and loss control at mines and smelters. Interleaving sampling protocols give unbiased precision estimates at the lowest possible cost. It does so at no additional cost if the wet mass of a sampling unit (a lot) is increased by the factor 2.

Whenever the set of primary increments is combined into a single primary sample, the variance of the primary sample selection stage and the variance of the sample preparation stage cannot be estimated. Only the analytical variance can be measured and monitored by assaying replicate test portions of a test sample prepared of the primary sample.

On-stream analyzers interrogate slurry flows either continuously or intermittently. The large set of on-stream data generated during a shift gives a high degree of precision for the arithmetic mean. However, this central value is an unbiased estimate for the central value only if the analyzer is in a proper state of calibration. Moreover, slurry densities and metal grades in flotation circuits may exhibit associative dependence so that the density weighted average is a more reliable estimate for the metal grade of a slurry than the arithmetic mean. A sampling module, designed to take a pair of interleaving secondary samples from a primary sample flow at constant time intervals during each shift, would make it possible to implement meaningful metallurgical accounting procedures.

The degree of associative dependence between on-stream data in the sample space of time (*spatial dependence* for simplicity) impacts the variances of ordered sets, and, thus, the precision of central values. Slurry flows are usually interrogated at constant time intervals which simplifies the calculation of variances and central values (see *Appendix D*).

MEASURES OF CENTRAL TENDENCY

Measured values tend to cluster around a central value which is often referred to as the central tendency of a set. The arithmetic mean is an unbiased estimate for the central value of a set of measured values with equal weighting factors whereas the weighted average is a more reliable (less bias prone) estimate for a set of measured values with variable weighting factors. Count, density, distance, length, mass and volume weighted averages are important measures of central tendency in mining and metallurgy.

Central values are measured with finite precision because each is merely an estimate (an unbiased estimate one would hope) for that most elusive central value, the unknown true value of the stochastic variable of interest in the sampling unit or sample space under examination.

The arithmetic mean is the central value of a set of measured values with equal weighting factors. The equation for the arithmetic mean is elementary, and has its own function in spreadsheet software. The weighted average is the central value of a set of measured values with variable weighting factors:

$$\bar{x} = \sum(w1i * xi) \qquad [Eq\ 1]$$

where $\bar{x}$ = *weighted average*
xi = *ith measured value*
$w1i$ = *first weighting factor for ith measured value*

Given that $w1i=mi/\sum mi$ for the mass weighted average grade of $\bar{x}=0.0386*32.1+\ldots+0.2684*29.8=30.71\%$ (see Table 2 and *Appendix C*), and $\sum(1/n)=1$ for the arithmetic mean grade of 30.90% *(see Appendix C)*, it follows that $\sum w1i$ is also unity. The second weighting factor of $w2i= mi/\sum mi=mi/\bar{m}$ is convenient in spreadsheet templates to obtain the variance of a set of measured values (see *Measures of Variability*), and the variance of its central value (see *Variances of Central Values*).

Weighted averages play a key role in a wide range of applications. For example, the length and density weighted average is the central value of a set of measured values for core samples of variable length and density. Similarly, the distance weighted average is the central value of a set of measured values with variable coordinates in a two- or three-dimensional sample space. In geostatistics, the distance weighted average transmogrified into the ubiquitous *kriged estimate*.

Table 1 gives, in addition to the set of paired data (dry masses in tonnes and metal grades in percent), $w1i$, the first weighting factor, which is required to calculate this mass weighted average grade of 30.71%, and $w2i$, the second weighting factor, which simplifies the equations for the variance of the set, and for the variance of its mass weighted average grade.

Table 1 Weighting factors

Unit	mass in mt	grade in %	$w1i$	$w2i$
1	12.1	32.1	0.0386	0.1929
2	54.5	30.3	0.1737	0.8687
3	72.6	30.5	0.2314	1.1572
4	90.3	31.8	0.2879	1.4393
5	84.2	29.8	0.2684	1.3420

The sum of the first set of weighting factors and the arithmetic mean of the second set are both unity which implies that $\sum w1i=\bar{w}2i=1$. This relationship makes it simple to check the correctness of both weighting factors in spreadsheet templates. The weighting factors in Table 1 are used in several calculations (see *Appendix C*).

MEASURES OF VARIABILITY

The variance is the fundamental measure of variability. Variances are amenable to mathematical analysis. All measures of variability and precision derive from the variance. For example, the standard deviation is the square root of the variance, and the coefficient of variation is the standard deviation as a relative percentage. The properties of variances are the essence of probability theory and applied statistics. The additive property of the variances of volume, mass and content has scores of powerful applications in mining and metallurgy.

Since the occurrence of spatial dependence between measured values in ordered sets is of critical importance in sampling theory and practice, this matter will be examined in a separate section (see *Testing for Spatial Dependence*). Almost invariably, measured values are ordered either in time (on-stream and production data) or in space (rounds in a drift or a trench; core samples in a borehole; boreholes in a section).

The question is then whether a set of measured values displays a statistically significant degree of spatial dependence, or is randomly distributed within its sample space. For example, the set of metal grades in Table 1 is ordered in time but it does not exhibit a significant degree of spatial dependence. Therefore, the fundamental measure of variability is the variance of the

randomized set. Thus, the term *"randomized"* implies that these metal grades constitute a randomly distributed set within this sample space of time.

By contrast, the set of on-stream data given in Appendix D exhibits a highly significant degree of spatial dependence. As a result, the variance of the ordered set gives a significantly higher degree of precision for its central value than the variance of the randomized set.

The example in Table 1 is more complicated in the sense that each metal grade represents a different mass. In this case, the mass weighted average grade is a measure of central tendency, the variance of the set is a measure of variability, and the variance of its central value is a measure of precision (see *Measures of Precision*).

The sequence in which the various equations for variances are given in the following sections reflects the fact that ordered sets occur more frequently than randomized sets. The terms *"ordered"* and *"randomized"* are juxtaposed, and combined with *"equal weighting factors"* or *"variable weighting factors"* to show how to calculate the corresponding measure of variability.

Ordered; Equal Weighting Factors

The variance of an ordered set of measured values with equal weighting factors (equidistant point estimates) is:

$$varj(x) = \frac{\sum (xi+j-xi)^2}{2(n-j)} \qquad [Eq\ 2]$$

where $varj(x)$ = *jth variance term of ordered set*
$xi+j$ = *(i+j)th measured value*
xi = *ith measured value*
j = *jth spacing between measured values*
n = *number of measured values for jth variance term*
$2(n-j)$ = *degrees of freedom for jth variance term*

This variance has found many applications in science and engineering. For example, the first variance term of the ordered set of metal grades with equal weighing factors (see *Table 1*) is $var1(x)=[(30.3-32.1)^2+...+(29.8-31.8)^2]/[2*(5-1)]=1.1212\%^2$. Given that degrees of freedom for ordered sets of measured values are not universally embraced (Merks 1993, 1997), it is necessary to explore this concept in a separate section (see *Degrees of Freedom*).

Ordered; Variable Weighting Factors

The variance terms of an ordered set of measured values with variable weighting factors are computed as follows:

$$varj(x) = \frac{\sum [w2j * (xi+j-xi)^2]}{(2/\sum w1j^2)-2j} \qquad [Eq\ 3]$$

where $varj(x)$ = *jth variance term of ordered set*
$xi+j$ = *(i+j)th measured value*
xi = *ith measured value*
j = *jth spacing between measured values*
$w1j$ = *first weighting factor for jth variance term*
$w2j$ = *second weighting factor for jth variance term*
$(2/\sum w1j^2)-2$ = *degrees of freedom for jth variance term*

Variance terms or ordered sets are useful in mineral processing, smelting and refining when variables are measured at different intervals, and in mineral exploration and mining where measured values for core samples of variable length and bulk samples of variable mass are ordered in space.

Randomized; Equal Weighting Factors

The variance of a set of measured values with equal weighting factors is elementary, and has its own function in spreadsheet software. The correct variance function in spreadsheet software should be selected to obtain the *"sample variance"*, the variance of a sample of a population rather than the *"population variance"* itself. These functions are *=VAR(xi..xj)* in Excel, and *@VARS(xi..xj)* in Lotus. The denominators in both functions are the degrees of freedom for the set (see *Degrees of Freedom*). The variance of the randomized set of metal grades (see *Table 1*) is $var(x)=[(32.1-30.90)^2+\ldots+(29.8-30.90)^2]/(5-1)=0.9950\%^2$ (see *Appendix C*).

Randomized; Variable Weighting Factors

The variance of a randomized set of measured values with variable weighting factors *w1i* and *w2i* as defined in *Measures of Central Tendency* is:

$$var(x) = \frac{\sum[w2i * (\bar{x}-xi)^2]}{(1/\sum w1i^2)-1} \qquad [Eq\ 4]$$

where $\bar{x}$ = *weighted average*
xi = ith measured value
w1i = first weighting factor for ith measured value
w2i = second weighting factor for ith measured value
$(1/\sum w1i^2)-1$ = *degrees of freedom*

The weighted average must be calculated before the differences between $\bar{x}$ and *xi* can be squared and multiplied with *w2i*, the corresponding weighting factors. Therefore, several columns in a spreadsheet template are required to obtain not only the central value and the variance of the set but also the variance of the central value (see *Appendix C*).

Given that the order in which squared differences are added does not impact the numeric values of variances, it does make sense to refer to *randomized* sets. Whenever a set of measured values is ordered, either in space (core samples in a borehole; rounds of crushed ore taken from a drift or trench) or in time (on-stream and production data), testing for spatial dependence becomes an important element of statistical analysis (see *Testing for Spatial Dependence*).

Table 2 gives, in addition to the mass weighted average grade of 30.71 % for the data set in Table 1, the most common measures of variability (see also *Appendix C*).

Table 2 Basic statistics

Statistic	Symbol[1]	Symbol[2]	Value
Mass weighted average grade in %abs	$\bar{x}$	xbar	30.71
Variance of randomized set in $(\%abs)^2$	var(x)	var(x)	0.8143
Standard deviation in %abs	sd(x)	sd(x)	0.9024
Coefficient of variation in %rel	CV	CV	2.9

[1] *text* [2] *template*

Due to its squared dimension the variance is not a user-friendly measure for variability. By contrast, the coefficient of variation (the standard deviation in relative percent) makes it simple to check and compare different degrees of variability at a glance.

TESTING FOR SPATIAL DEPENDENCE

A statistically significant degree of spatial dependence gives a lower variance of the ordered set, and, thus, a higher degree of precision for its central value. Testing for spatial dependence is also an important element of statistical analysis when optimizing sampling protocols. Fisher's F-test is applied to assess whether two variances are statistically identical or differ significantly

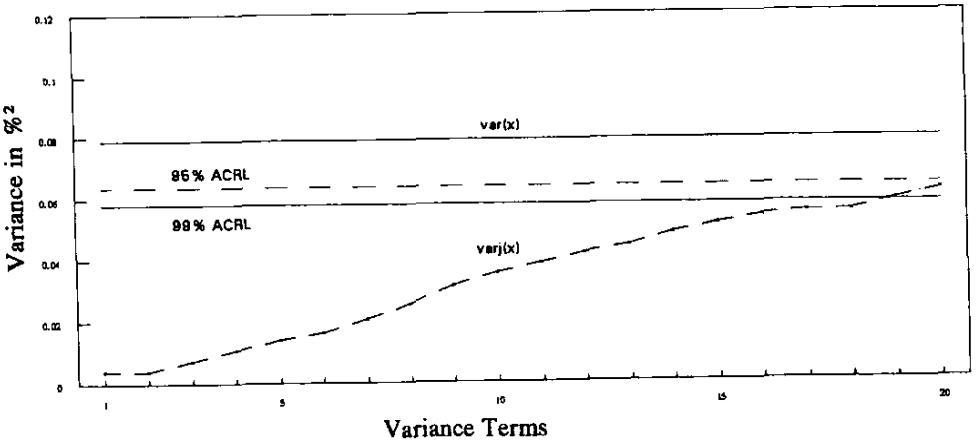


Figure 1 Sampling Variogram

by comparing the ratio between the highest variance and the lowest variance (the F-statistic or calculated F-value) with values tabulated in the F-distributions at 5% and 1% probability with the applicable degrees of freedom.

Applying the F-test to *var1(x)=1.1212%²*, the first variance term of the ordered set of metal grades in Table 1, and *var(x)=0.9950%²*, and the variance of the randomized set, gives F=1.1212/0.9950=1.13. Since this F-statistic is lower than F0.05;8;4=6.04 at 5% probability, these variances are statistically identical. By implication, the ordered set of metal grades does not exhibit a significant degree of spatial dependence. Thus, this set is classified as *"randomized"* within its sample space.

The tabulated values in the F-distribution at 5% probability rank from F0.05;1;1=161 to F0.05;∞;∞=1, and those in the F-distribution at 1% probability rank from F0.01;1;1=4,052 to F0.01;∞;∞=1 (Handbook 1986). Therefore, the F-statistic is always the ratio between the highest variance and the lowest variance (Volk 1980).

Sampling Variogram

The sampling variogram in Figure 1 is based on a set of 96 on-stream data obtained at 15 min intervals during a 24-hour shift. The variance terms of the ordered set, the variance of the randomized set, and the lower limits of the asymmetric confidence ranges at 95% and 99% probability are given in *Appendix D*. Since the F-statistic of F=0.0790/0.0038=20.79 for *var(x)=0.0790*, the variance of the randomized set, and *var2(x)=0.0038*, the second variance term of the ordered set, is higher than the tabulated value of F0.01;95;188=1.33 at 1% probability, the degree of spatial dependence at a spacing of 30 min is statistically significant.

By contrast, the F-statistic of F=0.0790/0.0618=1.28 for the variance of the randomized set, and *var20(x)=0.0618*, the 20th variance term of the ordered set, is lower than F0.05;95;154= 1.34 at 5% probability. Evidently, the ordered set of on-stream data no longer exhibits a significant degree of spatial dependence at a spacing of 5 hours.

When the variance of the randomized set (var(x)=0.0790%²), the lower limit of its asymmetric 95% confidence range *[95% ACRL=var(x)/F0.05;df;∞=0.0790/1.24=0.0581%²]*, and the lower limit of its 99% confidence range *[99% ACRL=var(x)/F0.01;df;∞=0.0790/1.36)=0.0637%²]*, are also plotted, the sampling variogram illustrates whether the degree of spatial dependence is statistically significant and where orderliness dissipates into randomness. In fact, a sampling variogram is a visual interpretation of Fisher's F-test when applied to check the degree of spatial dependence in the sampling unit or sample space under examination.

Generally, the existence of spatial dependence at spacing *j* is verified by applying the F-test to the variance of the randomized set and the *jth* variance term of the ordered set. If the

calculated F-value is higher than the tabulated F-value with the applicable degree of freedom, either at 5% or at 1% probability, then the degree of spatial dependence at the *jth* spacing is statistically significant at the corresponding probability level.

Although the first variance term determines the intrinsic variability of a dewatered sample for a shift, it does not give a variance estimate. Only a pair of interleaving samples gives an unbiased variance estimate that takes into account the second variance term of ordered on-stream data. Whenever the variance of a randomized set and the first variance term of the ordered set are statistically identical, the differences between consecutive on-stream data become random numbers that cannot be used for process control.

Higher variance terms have fewer degrees of freedom than lower terms because the last but one datum is not used for the second term, the last but two for the third, and so on. Therefore, F-statistics for small sets should be interpreted with caution. In addition, mathematical analysis ought not to be applied to differences between statistically identical variances (Merks 1993).

DEGREES OF FREEDOM

The concept of degrees of freedom in applied statistics is the corollary of the fundamental requirement of functional or mathematical independence in probability theory. The difference between *var(x)*, the variance of a sample, and σ^2, the population variance, explains why degrees of freedom are finite in applied statistics but deemed infinite in probability theory.

The differences between n measured values and the arithmetic mean of the set are $x1-\bar{x}$, ..., $xi-\bar{x}$, ..., $xn-\bar{x}$ so that the sum of n differences equals $(x1+\ldots+xi+\ldots+xn)-n\bar{x}$. By definition, the arithmetic mean of a set of n measured values is $\bar{x}=(x1+\ldots+xi+\ldots+xn)/n$. Hence, $(x1+\ldots+xi+\ldots+xn)-n\bar{x}=0$.

Logically, if $n-1$ differences are given, the missing one is determined because the sum of n differences is zero. Because a set of n measured values has $n-1$ independent differences and a single dependent difference, a randomly distributed set of n measured values has $n-1$ degrees of freedom.

By contrast, the first variance term of an ordered set has $2(n-1)$ or $(2/\sum w1j^2)-2$ degrees of freedom because all but the first and last datum are used twice which implies that each higher term has two fewer degrees of freedom than the preceding term.

One measured value does not give any information on that elusive population variance simply because $\sum(x1-\bar{x})^2/(n-1)=0/0$, which is indeterminate as it ought to be. It can be proved by induction that adding any number of functionally (or mathematically) dependent values to a set of measured (or independent) values does not add a single degree of freedom.

Degrees of freedom are positive integers for sets of measured values with equal weighting factors but become positive irrational numbers for sets of measured values with variable weighting factors.

VARIANCES OF CENTRAL VALUES

The variances of central values are pivotal statistics in sampling theory and practice because they play a critical role in bias testing of mechanical sampling systems, manual sampling procedures, sample preparation techniques and analytical methods. The variances of central values also underlie confidence intervals and ranges, bias detection limits as measures for the power or sensitivity of Student's t-test, and probable ranges as intuitive measures for the probabilistic limits within which an observed bias is expected to fall.

Variance of Arithmetic Mean

The variance of the arithmetic mean of a set of n measured values with equal weighting factors is elementary:

$$var(\bar{x}) = \sum(1/n)^2 * var(x) = var(x)/n \qquad [Eq\ 5]$$

where $var(\bar{x})$ = *variance of arithmetic mean*
$var(x)$ = *variance of set*
n = *number of measured values in set*

The intermediate term in Equation 5 reflects an important step in the derivation of this equation from the variance of a general function as defined in probability theory (Volk 1980). This simple relationship between the variance of a set of measured values with equal weighting factors and the variance of its arithmetic mean is often referred to as the *Central Limit Theorem.* Perhaps ironically because the variances of weighted averages are required in many applications in mineral exploration, mining, processing, smelting and refining.

Variance of Weighted Average

The variance of a count, density, distance, length, mass or volume weighted average of a set of measured values with variable weighting factors is:

$$var(\bar{x}) = \sum w1i^2 * var(x) \qquad [Eq\ 6]$$

where $var(\bar{x})$ = *variance of weighted average*
$var(x)$ = *variance of set*
$w1i$ = *first weighting factor for ith measured value*

The variance of the weighted average of any set of measured values with variable weighting factors converges on the variance of the arithmetic mean when all $w1i$ approach $1/n$. When formulated as the sum of two or more products of squared weighting factors and variances, the *central limit theorem* proves that the variance of the central value of two or more sets of primary samples, selected from multinomial, binomial or Poisson distributions, converges on the normal (or Gaussian) distribution.

How to determine whether or not a set of measured values exhibits a Gaussian or normal probability distribution is described in a draft standard under development by ISO/TC69. Whenever a set of measured values departs from normality, it can be partitioned into subsets such that each subset approaches a straight line segment in a log-normal plot of a numerically ordered set. The next step is to calculate the central value of the set and its variance from the arithmetic means and the variances of all subsets (Merks 1998).

VARIANCE OF CONTAINED METAL

The additive property of the variance of contained metal (the mass of contained metal or metal content) underlies various applications in mineral exploration, mining, processing and smelting (Merks 1985, 1988, 1991, 1999; Merks and Merks 1991). ISO/DIS 13543 describes how to determine the variance of the mass of metal contained in a lot. This method is based on the premise that pairs of interleaving primary samples are routinely taken from all lots. Pairs of interleaving samples also give the variances of metal contained in tailings, concentrates and thickener inventories, which can be used to calculate reliable precision estimates for metal grades of mill feed.

The mass of metal contained in a quantity of crushed ore or mineral concentrate is a function of its wet mass, moisture content and metal grade:

$$Me = Mw * MF * GF \qquad [Eq\ 7]$$

where Me = *mass of contained metal in mt*
Mw = *wet mass in mt*
MF = *moisture factor : 1 − [0.01 * %H2O] (dimensionless)*
$\%H_2O$ = *moisture content in percent*
GF = *grade factor : 0.01 * %Me (dimensionless)*
$\%Me$ = *metal grade in percent on dry basis*

The variance of contained metal is obtained by substituting in the equation for variance of a general function the squared partial derivatives for Equation 7 and the variances of these stochastic variables :

$$var(Me) = (MF*GF)^2 * var(Mw) + (Mw*GF)^2 * var(MF) + (Mw*MF)^2 * var(GF) \qquad [Eq\ 8]$$

where $var(Me)$ = *variance of contained metal in* mt^2
$var(Mw)$ = *variance of wet mass in* mt^2
$var(MF)$ = *variance of moisture factor (dimensionless)*
$var(GF)$ = *variance of grade factor (dimensionless)*

Multiplying the mass term in Equation 8 with Mw^2/Mw^2, the moisture term with MF^2/MF^2, and the grade term with GF^2/GF^2, dividing each term by Me^2, and multiplying the sum of all terms with Me^2, gives the following equation for the variance of contained metal:

$$var(Me) = Me^2 * [var(Mw)/Mw^2 + var(MF)/MF^2 + var(GF)/GF^2] \qquad [Eq\ 9]$$

Extreme care should be exercised to ensure that variables and variances are correctly entered into Equation 9. Changing from grades in percent (%Me) to grade factors (GF), from precious metal grades in g/mt to contained metal in kg, and from moisture contents (%H2O) to moisture factors (MF), demands close attention to derivatives, dimensions and decimal places. Scale calibration data can often be used to obtain the variance of wet mass (Merks and Merks 1992).

The additive property of variances also makes it simple to determine the variance of the mass of metal contained in an ore deposit, and to calculate confidence limits for its metal content and grade as a measure for the risk associated with the least precise measurement procedure in mining and metallurgy. The variances of dividing whole core samples into halves, preparing test samples of selected halves, and taking and assaying test portions of test samples, are extraneous to the sample space, and add to the variance of the stochastic variable within its sample space. Extraneous variances may be subtracted from the variances of the randomized and ordered sets before Fisher's F-test is applied and confidence limits for contents and grades of ore deposits are computed (Merks and Merks 1991, Merks 2000).

VARIANCE OF GY's SAMPLING CONSTANT

Gy's sampling theory proposes that the primary sample mass required for a specified degree of precision can be estimated *a priori*. In sampling practice, however, it can only be determined experimentally because the degree of heterogeneity of a stochastic variable within a sampling unit defies *a priori* estimation (Merks 1985, Visman 1962). The variance of the primary sample selection stage is the sum of the composition variance (the variance between particles within primary increments) and the distribution variance (the variance between primary increments within a sampling unit). It is the latter variance that defies *a priori* estimation in heterogeneous sampling units, and that causes the ordered set of on-stream data to exhibit spatial dependence (see Figure 1) and gives a higher degree of precision for the central value (see *Appendix D*).

Gy's sampling theory suggests that $\sigma^2(FE)$, his fundamental error, is a function of C, his sampling constant, and d^3, the cube of the top size of the particulate matter. Gy's sampling constant C, in turn, is a function of four factors (Gy 1979). The variance of Gy's sampling constant, too, derives from the variance of a general function (Volk 1980):

$$var(C) = C^2 * [var(c)/c^2 + var(l)/l^2 + var(f)/f^2 + var(g)/g^2] \qquad [Eq\ 10]$$

where $var(C)$ = *variance of sampling constant*
$var(c)$ = *variance of mineralogical composition factor*
$var(l)$ = *variance of liberation factor*
$var(f)$ = *variance of particle shape factor*
$var(g)$ = *variance of size range factor*

Logically, C is a constant only if each of these variances is infinitesimally small but, in the real world, variances are finite. In fact, a single pair of interleaving primary samples gives an imprecise estimate for $var(spa)$, the sum of the variances of the primary sample selection, preparation and analytical stages because it has but one degree of freedom. In sampling practice,

the question is not so much whether Gy's sampling constant is indeed a constant but how imprecise a variance estimate with a single degree of freedom really is.

Fortuitously, applied statistics gives a relationship between degrees of freedom and confidence limits for variances. Symmetric 95% confidence ranges for variances are computed from values of the χ^2-distribution at different probability levels (Handbook 1968). Table 3 gives the tabulated χ^2-values at 2.5% and 97.5% probability, and the lower limits *[95% CRL= df * var(spa)/χ²0.975;df]* and upper limits *[95% CRU=df * var(spa)/χ²0.025;df]* of the symmetric 95% confidence ranges for *var(spa)=0.10* when estimated with increasing degrees of freedom (Volk 1980).

Table 3 Symmetric 95% confidence ranges for *var(spa)=0.10*

Degrees of freedom	$\chi^2$0.975	$\chi^2$0.025	95% CRL	95% CRU
1	5.02	0.001	0.020	101.8
5	12.8	0.831	0.039	0.602
10	20.5	3.25	0.049	0.308
25	40.5	13.1	0.062	0.191
∞			0.10	0.10

Table 3 underscores the astounding precision of variances when degrees of freedom are infinite. Given that Gy's sampling constant is a function of four stochastic variables whose variances have finite degrees of freedom, it is implausible that *var(C)*, the variance of Gy's sampling constant, is infinitesimally small. The more so because the variance of the cube of the topsize is $3^2=9$ times larger than the variance of the topsize itself.

Neither is it plausible that $\sigma^2(FE)$, Gy's fundamental error, gives a meaningful *a priori* estimate for the primary sample mass required for a specified degree of precision for a heterogeneous sampling unit. After all, the distribution component of the variance of the primary sample selection stage, which is a measure for the degree of segregation or heterogeneity in a sampling unit, can only be estimated from a sampling experiment based on taking 20−30 pairs of small and large increments (ASTM D2234, Merks 1985, Visman 1962).

When this experiment is applied to a dynamic sampling unit, the distribution variance, the very statistic to be estimated, is reduced. This is the corollary of Heisenberg's uncertainty principle in sampling practice where the measurement procedure impacts the outcome. The same experiment does give an estimate of the composition variance but does so at high cost.

MEASURES OF PRECISION

The fundamental measure of precision is the variance of a central value but derived measures of precision such as confidence intervals and ranges are more intuitive and transparent than variances. For many applications, 95% confidence intervals (95% CI) and 95% confidence ranges (95% CR) are acceptable, but if the risk associated with a wrong decision is high, confidence intervals and ranges at 99% or 99.9% probability should be considered. Confidence intervals are given in absolute values and relative percentages whereas confidence ranges are given in absolute values only.

Confidence Interval

The calculation of a 95% confidence interval for the central value of a randomized set requires a tabulated value of the t-distribution at 5% probability with $df=n-1$ or $df=(1/\sum wli^2)-1$ degrees of freedom. However, if the first variance term of the ordered set is significantly lower than the variance of the randomized set, the t-values at $df=2(n-1)$ or $df=(2/\sum wli^2)-2$ degrees of freedom may be used.

Since t0.05;60=2.000 for 60 degrees of freedom, and t0.05∞=z0.05=1.96 for infinite degrees of freedom, the z-value of normal distribution is can be rounded to 2, the following equation applies to all sets:

$$95\%\ CI = sd(\bar{x}) * t0.05;df \qquad [Eq\ 11]$$

where $95\%\ CI$ = *95% confidence interval (absolute value)*
$sd(\bar{x})$ = *standard deviation of central value*
$t0.05;df$ = *tabulated t-value at 5% probability*
df = *degrees of freedom*

*Table 4 gives the mass weighted average grade of 30.71% for the set of paired dry masses and metal grades in Table 1, its 95% confidence interval of ±1.37%abs (absolute percent), and its 95% confidence interval of 95% CI=1.37*100/30.71 = ±4.5%rel (relative percent).*

Table 4 Confidence interval

Statistic	Symbol[1]	Symbol[2]	Value
Mass weighted average grade in %abs	$\bar{x}$	xbar	30.71
95% Confidence interval in %abs	95% CI	95% CI	±1.37
95% Confidence interval in %rel	95% CI	95% CI	±4.5

[1] text [2] templates

Appendix C also gives 95% CIs in absolute values (%abs in this case), and in relative percent (%rel) but without ± - symbols. Comparing the 95% CI of ±4.5%rel in Table 4 with 95% CI = ±1.05%rel for the randomized set of on-stream data (see *Appendix D*), and 95% CI = ±0.24%rel for the ordered set illustrates how a large data set and a significant degree of spatial dependence impact the precision of the central value of 5.22%. Confidence intervals at 99% and 99.9% are obtained by multiplying $sd(\bar{x})$ with t0.01;df and t0.001;df respectively.

Symmetric Confidence Range

The lower and upper limits of a symmetric 95% confidence range are obtained as follows:

$$95\%\ CRL = \bar{x} - 95\%\ CI \qquad [Eq\ 12]$$
$$95\%\ CRU = \bar{x} + 95\%\ CI \qquad [Eq\ 13]$$

where $95\%\ CRL$ = *lower limit of 95% confidence range*
$95\%\ CRU$ = *upper limit of 95% confidence range*
$\bar{x}$ = *central value of set*
$95\%\ CI$ = *95% confidence interval*

Table 5 is based on the mass weighted average grade of 30.71% for the set of paired wet masses and metal grades in Table 1, and on the derived statistics in Appendix C.

Table 5 Confidence range

Statistic	Symbol[1]	Symbol[2]	Value
Mass weighted average grade in %abs	$\bar{x}$	xbar	30.71
95% Confidence range	95% CR	95% CR	
Lower limit in %abs	95% CRL	95% CRL	29.3
Upper limit in %abs	95% CRU	95% CRU	32.1

[1] text [2] template

Confidence ranges at 95% and 99% probability are convenient control and action limits in statistical quality control (SQC) charts.

Asymmetric Confidence Range

The lower limit of an asymmetric 95% confidence range is the central value of the set minus its 90% confidence interval. Similarly, the upper limit of an asymmetric 95% confidence range is the central value of the set plus its 90% confidence interval:

$$95\%\ ACRL = \tilde{x} - 90\%\ CI \qquad \textit{[Eq 14]}$$

$$95\%\ ACRU = \tilde{x} + 90\%\ CI \qquad \textit{[Eq 15]}$$

where $95\%\ ACRL$ = *lower limit of 95% confidence range*
$95\%\ ACRU$ = *upper limit of 95% confidence range*
$\tilde{x}$ = *central value of set*
$90\%\ CI$ = *90% confidence interval*

These lower limits and upper limits are mutually exclusive. In other words, either the lower limit or the upper limit is valid. Together, however, the same limits give a symmetric 90% confidence range. For large sets, the tabulated value of $t0.10;df$ converges on $z0.10 = 1.645$.

STUDENT's t-TEST

The t-test is applied to examine whether the difference between identifiably different paired test results is due to random variations or caused by the presence of bias. Typical examples are test results for reference increments and system samples, for different laboratories, for the same laboratory but at different times or by different technicians, or for different analytical methods. In every case, the question is whether two central values differ significantly, and, thus, whether their difference indicates the presence of bias (reject null hypothesis). Alternatively, the difference between central values is statistically identical to zero (accept null hypothesis), and its numeric value merely reflects the effect of random variations in measurement procedures.

An observed bias is either significantly higher than an accepted value such as a certified value for a Certified Reference Material or the central value of a set of reference increments (a positive bias), or significantly lower than the certified or reference value (a negative bias). When a set of paired test results reported by two laboratories fails the bias test, the t-test does not reveal which laboratory is suspect but only that the difference between their central values is higher than random variations alone could explain.

Most textbooks on applied statistics give the t-distribution with tabulated values for probabilities ranging from 90% to 0.1% for one degree of freedom to infinite degrees of freedom. If the t-statistic is much lower than $t0.90;df$, it may reflect the *too-good-to-be-true* effect, which could be indicative of tampering with test results.

Since the t-statistic (the calculated t-value) is the ratio between the difference between two central values and the standard deviation of the difference, the following equations apply:

$$t = \frac{\tilde{x}1 - \tilde{x}2}{\sqrt{[var(\Delta x)/n]}} = \frac{\Delta\tilde{x}}{sd(\Delta\tilde{x})} \qquad \textit{[Eq 16]}$$

where t = *t-statistic*
$\tilde{x}1$ = *central value of first set*
$\tilde{x}2$ = *central value of second set*
$\Delta\tilde{x}$ = *difference between central values*
$var(\Delta x)$ = *variance of differences*
n = *number of pairs*
$sd(\Delta\tilde{x})$ = *standard deviation of difference*

The central limit theorem also underlies the relation between $sd(\Delta\tilde{x})$, the standard deviation of the difference between two central values, and $sd(\Delta x)$, the standard deviation of the differences between paired test results. Given that $sd(\Delta\tilde{x}) = \sqrt{[var(\Delta x)/n]} = sd(\Delta x)/\sqrt{n}$, it follows that three variables interact and determine the t-statistic and the power of the t-test.

Appendix E gives the t-statistics for a test program designed to test for bias between reference increments and final system samples. This bias test program is based on comparing 30 pairs of test results determined in reference increments removed from a stopped belt with the aid of a sampling frame, and in final system samples obtained with a multistage mechanical sampling system. Table 6 gives the basic t-statistics for the test program.

Table 6 Basic t-statistics

Statistic	Symbol[1]	Symbol[2]	Value
Central value in % : reference increments	$\bar{x}(R)$	xbar(R)	8.10
Central value in % : final system samples	$\bar{x}(s)$	xbar(s)	7.49
Difference in %abs	$\Delta\bar{x}$	dxbar	-0.61
Difference in %rel	$\Delta\bar{x}$	dxbar	-7.5
Calculated t-value	t	t	11.296
Significance			***

[1] *text* [2] *template* *** *significant at 0.1% probability*

The calculated t-value of 11.296 exceeds the tabulated value of $t0.001;29=3.674$ at 0.1% probability (Handbook 1968; Volk 1980) so that the difference of −0.61%abs or −7.5%rel implies the presence of bias. The spreadsheet template in Appendix E gives three asterisks to indicate statistical significance at 0.1% probability. Two asterisks would have been printed for statistical significance at 1%, and a single one at 5%. In addition, *ns* (not significant) would have been printed in the same cell if the t-statistic were lower than $t0.05;df$ at 5% probability.

The variance of differences is calculated from the differences between paired data. This variance is the sum of the variances of all systems and procedures used to obtain the set. Given that the variance of differences and the number of paired data determine the power of the t-test, it is possible to prove that even a small and commercially insignificant difference is a bias if the number of pairs is large enough. A preliminary bias test may be needed to estimate the number of test results necessary to prove statistical significance at a specified probability level. The t-test can also be applied to pairs of measured values with variable weighting factors such as central values for on-stream data and test results for slurry samples for the same production period, or to the exchange assays for lots of variable mass. It is beyond the scope of this paper to present a numerical example.

Bias Detection Limits

Bias detection limits (BDLs) are intuitive measures for the power or sensitivity of the t-test to detect a bias or systematic error between two central values. BDLs are defined for the Type I statistical risk only, and for the combined Type I and Type II statistical risks. A simple analogy exists between these statistical risks and the role of a fire alarm. The Type I statistical risk refers to the event that a fire occurs but the fire alarm does not sound. The Type II statistical risk refers to the event that the fire alarm sounds but no fire occurs. Finally, the combined Type I and Type II statistical risks refer to a fire and the sound of a fire alarm.

The effect of the number of pairs on the power of the t-test becomes evident upon realizing that these statistical risks are obtained by multiplying the standard deviation of the difference either with the tabulated t-value at 5% probability, or with the sum of the tabulated t-values at 5% and 10% probability. A symmetric two-sided 5% probability for the Type I risk only, and an asymmetric one-sided 5% probability for the Type II risk, are widely accepted.

Several ISO Standards on bias testing of mechanical sampling systems and manual sampling procedures specify statistical risks in the same manner and at the same probability levels. Based on this convention, the bias detection limits for the Type I risk only, and for the combined Type I and Type II risks, are defined as follows:

$$BDL(I)=sd(\Delta\bar{x}) * t0.05;df \qquad [Eq\ 17]$$

$$BDL(I\&II) = sd(\Delta\bar{x}) * [t0.05;df + t0.10;df] \quad \textit{[Eq 18]}$$

where BDL(I) = BDL for Type I risk only
BDL(I&II) = BDL for combined Type I and Type II risks
$sd(\Delta\bar{x})$ = standard deviation of difference
t0.05;df = tabulated t-value at 5% probability
t0.10;df = tabulated t-value at 10% probability
df = degrees of freedom

Table 7 gives the bias detection limits in absolute and relative percent on the basis of the differences between the central values of 8.10% for reference increments and 7.49% for final system samples (see also Table 6 and *Appendix E*).

Table 7 Bias detection limits

Statistic	Symbol	%abs	%rel
Difference	$\Delta\bar{x}$	−0.61	−7.5
Bias detection limits	BDLs		
Type I risk only	BDL(I)	±0.11	±1.4
Type I and II risks	BDL(I&II)	±0.20	±2.5

Bias detection limits are effective control and action limits for SQC charts in which precision and bias of measurement systems and procedures are monitored as a function of time. A strong case can be made that metal grades and contents of concentrate shipments should be measured and monitored to ensure that biases are detected before losses become punitive.

Probable Ranges

Probable ranges (PRs) define the limits within which an observed bias is expected to fall. Whenever a difference between two central values turns out to be statistically significant, and exceeds either the bias detection limits for the Type I statistical risk only, or the combined Type I and Type II statistical risks, the following relationships give the lower and upper limits of the corresponding probable ranges for the observed bias:

$$PRL(I) = \Delta\bar{x} - BDL(I) \quad \textit{[Eq 19]}$$
$$PRU(I) = \Delta\bar{x} + BDL(I) \quad \textit{[Eq 20]}$$
$$PRL(I\&II) = \Delta\bar{x} - BDL(I\&II) \quad \textit{[Eq 21]}$$
$$PRU(I\&II) = \Delta\bar{x} + BDL(I\&II) \quad \textit{[Eq 22]}$$

where PRL(I) = lower limit of probable range for Type I risk only
PBU(I) = upper limit of probable range for Type I risk only
PRL(I&II) = lower limit of probable range for combined Type I and II risks
PBU(I&II) = upper limit of probable range for combined Type I and II risks
$\Delta\bar{x}$ = observed bias
BDL(I) = bias detection limit for Type I risk only
BDL(I&II) = bias detection limit for Type I and Type II risks

Reporting probable ranges for an observed bias makes sense only if the difference between two central values is indeed indicative of the presence of bias, and the null hypothesis is rejected. If a difference exceeds the bias detection limit for the Type I risk, the lower and upper limits of the corresponding probable range are reported. For example, the difference of −0.61%abs between 8.10% for reference increments and 7.49% for system samples is lower than BDL(I)= −0.11% and BDL(I&II)= −0.20% (see Table 7). Therefore, the lower and upper limits of the

probable ranges are defined not only for the Type I risk but also for the combined Type I and Type II risks.

Logically, the difference of −7.5 %rel between reference increments and system samples is also lower than BDL(I) = −1.4 %rel and BDL(I&I) = −2.5 %rel (see Table 7) which implies that the probable bias ranges are defined for the Type I risk only and for the combined Type I and Type II risks. Table 8 gives the probable ranges for the observed bias.

Table 8 Probable ranges

Statistic	Symbol	%abs	%rel
Difference	$\Delta\bar{x}$	−0.61	−7.5
Probable range	PRs		
Type I risk only	PR(I)		
Lower limit	PRL(I)	−0.72	−8.9
Upper limit	PRU(I)	−0.50	−6.1
Type I and II risks	PR(I&II)		
Lower limit	PRL(I&II)	−0.81	−10.0
Upper limit	PRU(I&II)	−0.41	−5.0

If a difference between two central values is statistically identical to zero, and the lower and upper limits of the probable range for the Type I risk are not defined, the abbreviation *na* (not applicable) may be printed in the appropriate cells of the spreadsheet template (see *Appendix E*). Whenever an observed bias in moisture content or metal grade impacts the cumulative mass of metal contained in concentrate production, it would make sense to convert probable ranges into monetary units.

FISHER's F-TEST

Fisher's F-test is applied to determine whether two variances are statistically identical or differ significantly. The F-test is based on comparing the ratio between the highest variance and the lowest variance with tabulated values from the F-distributions at 5% and 1% probability and with the applicable degrees of freedom for each variance (Handbook 1968; Volk 1980). If the calculated F-value is lower than the tabulated value of $F0.05;df1;df2$ at 5% probability, then the variances are statistically identical. The probability that this statistical inference is true exceeds 95%. Conversely, the probability that this inference is false is less than 5%.

Alternatively, if the F-statistic is higher than $F0.05;df1;df2$ at 5% probability, the variances differ significantly, and the probability is less than 5% that this inference is false. Similarly, if the F-statistic is higher than $F0.01;df1;df2$ at 1% probability, the variances differ significantly but in this case, the probability is less than 1% that the inference is false.

Tabulated F-values, too, reflect that $df1$ and $df2$ are the degrees of freedom for the numerator and denominator in the F-test. The fact that $F0.05;\infty;\infty = F0.01;\infty;\infty = 1$ explains why the concept of degrees of freedom is of critical importance when analysis of variance is applied to test for spatial dependence and to optimize sampling protocols.

Optimizing Sampling Protocols

Suppose that a sampling experiment gives $var(spa) = 0.075$ for the sum of the variances of the primary sample selection, preparation and analytical stages, and $var(a) = 0.050$ for the variance of taking and assaying a test portion of a test sample. The question of whether these variance estimates are statistically identical or differ significantly can only be solved if the applied sampling protocol and the degrees of freedom for $var(spa)$ and $var(a)$ are taken into account.

For example, a pair of interleaving primary samples gives a single degree of freedom for $var(spa) = 0.0750$, and duplicate test portions taken from each of a pair of test samples give two degrees of freedom for $var(a)/2 = 0.050/2 = 0.025$, the analytical variance of the arithmetic

mean of duplicate test results. This is the reason why the analytical variance (the variance of taking and assaying a single test portion of a test sample) is divided by the factor 2. Since $F=0.075/0.025=3.00$ is lower than $F0.05;1;2=18.51$ at 5% probability, the difference of $0.075-0.025=0.050$ between *var(spa)* and *var(a)/2* is not a valid estimate for *var(sp)*, the sum of the variances of the primary sample selection and preparation stages. Therefore, no statistical significance should be attached to this difference of 0.050, nor should mathematical analysis be applied to such differences (Merks 1993).

By contrast, 20 pairs of interleaving primary samples would give 20 degrees of freedom for $var(spa)=0.075$ whereas duplicate test portions taken from each of 40 test samples would give 40 degrees of freedom for $var(a)/2=0.050/2=0.025$. In this case, the F-statistic of $F=0.075/0.025=3.00$ exceeds not only $F0.05;20;40=1.54$ at 5% probability but also $F0.01;20;40=2.37$ at 1% probability. Hence, the same difference of $0.075-(0.050/2)=0.050$ is a valid estimate for *var(sp)*, the sum of the variances of the primary sample selection and preparation stages. The probability that this statistical inference is false is less than 1%.

The latter F-test shows that *var(sp)*, the sum of *var(s)*, the variance of the primary sample selection stage, and *var(p)*, the variance of the sample preparation stage, adds most to *var(spa)*, the variance of the entire measurement chain. The most effective method to reduce *var(s)* the variance of the primary sample selection stage, is to increase the number of primary increments (Gy 1979, Merks 1985, Visman 1962).

The variance of the sample preparation stage can be estimated by preparing duplicate test samples of each of a pair of interleaving samples, and assaying duplicate test portions of each test sample. For example, 10 pairs of interleaving primary samples would generate 20 pairs of test samples and 40 pairs of test portions, and give 10 degrees of freedom for *var(spa)*, the sum of the variances of the primary sample selection, preparation and analytical stages, 20 for *var(pa)*, the sum of the variances of the sample preparation and analytical stages, and 40 for *var(a)*, the analytical variance (see *Appendices A & B*).

The variance of the sample preparation stage can be reduced to a minimum by comminuting and homogenizing dried sample masses prior to division. The key is always to find a compromise between acceptability and expediency, a task that requires some understanding of experiment design and statistical analysis of test results. Sample preparation procedures are prone to bias due to cross contamination and loss of dust, moisture or native metal while comminuting, homogenizing and dividing sample masses (Merks 1985, 1988, 1993).

SUMMARY

Sampling in mineral processing is based on scientifically sound elements of probability theory and applied statistics. The properties of variances are the quintessence of sampling theory and practice. The additive property of the variances of volume, mass and contained metal play a key role in metallurgical accounting procedures.

Combining a set of primary increments into a single primary sample does not give a variance estimate. Dividing a set of primary increments into a pair of interleaving primary samples is the most effective procedure to estimate the variance of the entire measurement chain. Interleaving sampling protocols are equally effective for slurry flows in mineral processing and bulk samples in mineral explorations. Sampling protocols can be optimized by applying analysis of variance to partition the sum of the variances of the primary sample selection, preparation and analytical stages into its components, and by examining which variance component should be reduced to improve the precision of the measurement procedure.

On-stream data almost invariably exhibit a significant degree of spatial dependence. Metal grades of contiguous sets of core samples within a borehole, or a set of adjoining rounds in a drift or trench, may also display a significant degree of spatial dependence. When plotted in a graph the variance terms of an ordered set display a sampling variogram. When the variance of the randomized set and the lower limits of its asymmetric 95% and 99% confidence ranges are also plotted, the sampling variogram shows whether the degree of spatial dependence is statistically significant and where orderliness in the sampling unit or sample space under examination has dissipated into randomness.

The computations discussed in this paper are carried out with spreadsheet software. Setting up effective spreadsheet templates is an important element of sampling practice in mining and metallurgy. A strong case can be made that sound elements of probability theory and applied statistics be implemented in all the measurement procedures commonly applied in mineral exploration, mining, processing, smelting and refining.

ACKNOWLEDGMENT

The author dedicates this paper to the memory of Len Green, formerly with Falconbridge Limited, and for many years the driving force behind the Canadian Advisory Committee to ISO Technical Committee 183 on copper, lead and zinc concentrates.

REFERENCES

ASTM Standards on Precision and Bias for Various Applications, 1985, Second Edition

ASTM D2234, 1997, Standard Practice for the Collection of a Gross Sample of Coal, Annual Book of ASTM Standards, Volume 05.05, Gaseous Fuels; Coal and Coke

Davies O L and Goldsmith P L, 1972, Statistical Methods in Research and Production, Longman Group Limited, London

Gy, P M, 1979, Sampling of Particulate Matter; Theory and Practice, Elsevier Scientific Publishing, Amsterdam

Handbook of Tables for Probability and Statistics, 1968, The Chemical Rubber Company, Cleveland, Ohio

Huff, D, 1954, How to Lie with Statistics, Penguin Books, Middlesex, England

ISO/DIS 10251, Determination of Mass Loss of Bulk Material on Drying

ISO/DIS 12743, Sampling Procedures for Determination of Metal and Moisture Content

ISO/DIS 12744, Experimental Methods for Checking the Precision of Sampling

ISO/DIS 12745, Precision and Bias for Mass Measurement Techniques

ISO/DIS 13292, Experimental Methods for Checking the Bias of Sampling

ISO/DIS 13543, Determination of Mass of Contained Metal in the Lot

Mandel, J, 1964, The Statistical Analysis of Experimental Data, Dove Publications, New York

Merks, J W, 1985, Sampling and Weighing of Bulk Solids, TransTech Publications, Clausthal-Zellerfeld

Merks, J W, 1988, Sampling and Weighing of Mineral Concentrates, Bulk Solids Handling, Volume 8, Number 2

Merks, J W, 1989, A New Approach to the Exchange of Assays, Sampling in the Non-Ferrous Metals Industry, TransTech Publications, Clausthal-Zellerfeld

Merks, J W and Merks E A T, 1991, Precision Estimates for Ore Reserves, Erzmetall, October

Merks, J W, 1991, Simulation Models for Mineral Processing Plants, CIM Bulletin, September

Merks, J W and Merks E A T, 1992, Precision and Bias for Mass Measurement Techniques, Matrix Consultants Limited, Vancouver

Merks, J W, 1993, Abuse of Statistics, CIM Bulletin, January

Merks, J W, 1994, Precision and Bias—The Keys to Total Quality Management, Proceedings of the Eighteen International Precious Metals Conference, International Precious Metal Institute, Allentown, Pennsylvania

Merks, J W, 1997, Applied Statistics in Mineral Exploration, Mining Engineering, February

Merks, J W, 1998, Private communication prepared for litigation

Merks, J W, 1999, Process Simulation with Spreadsheet Software, Minerals & Metallurgical Processing, May

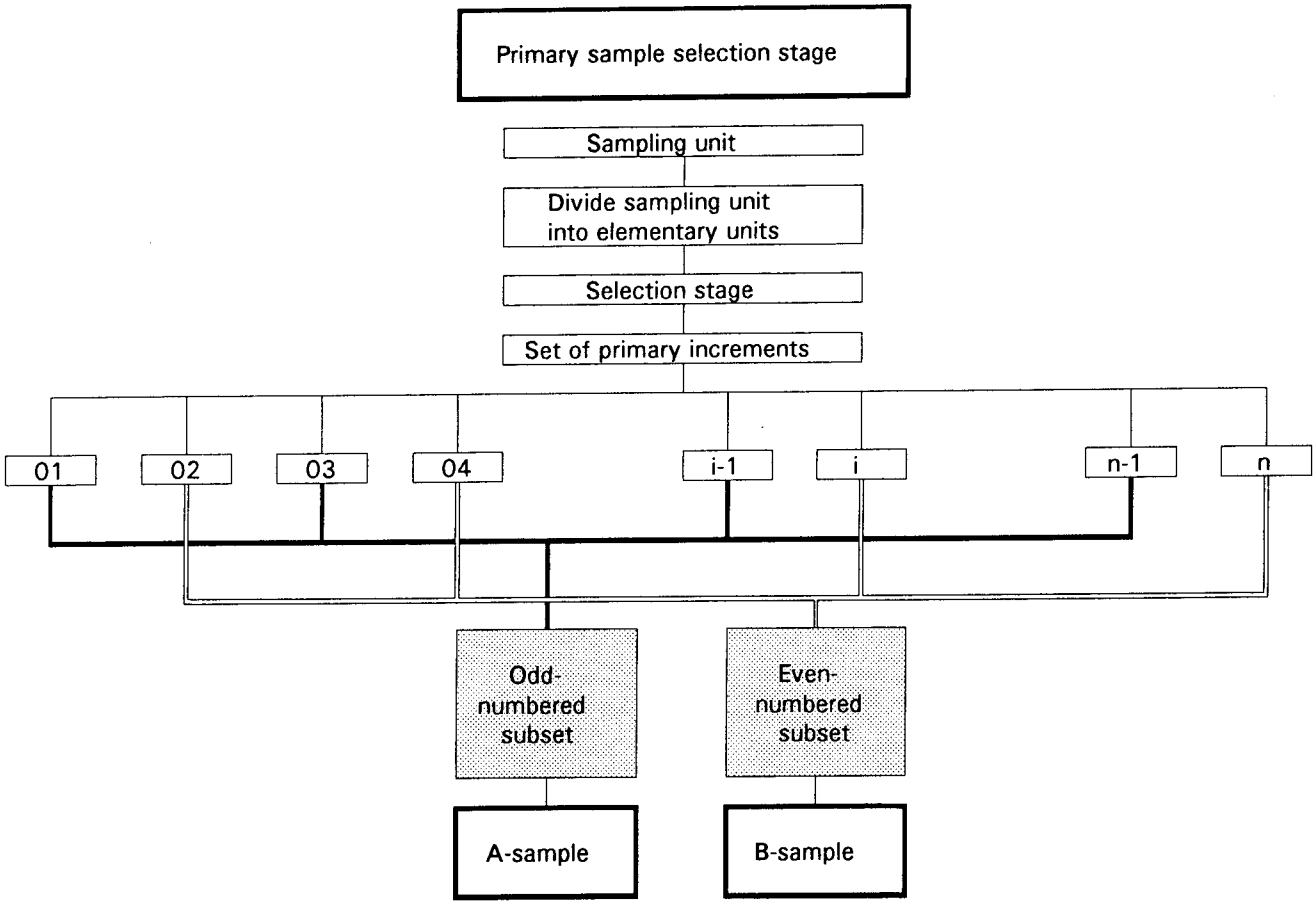
Merks, J W, 2000, Borehole Statistics with Spreadsheet Software, SME Transactions, Volume 308

Visman, J, 1962, Towards a Common Basis for the Sampling of Materials, CANMET Research Report R93, July

Volk, W, 1980, Applied Statistics for Engineers, Robert R Krieger Publishing Company, Huntingdon

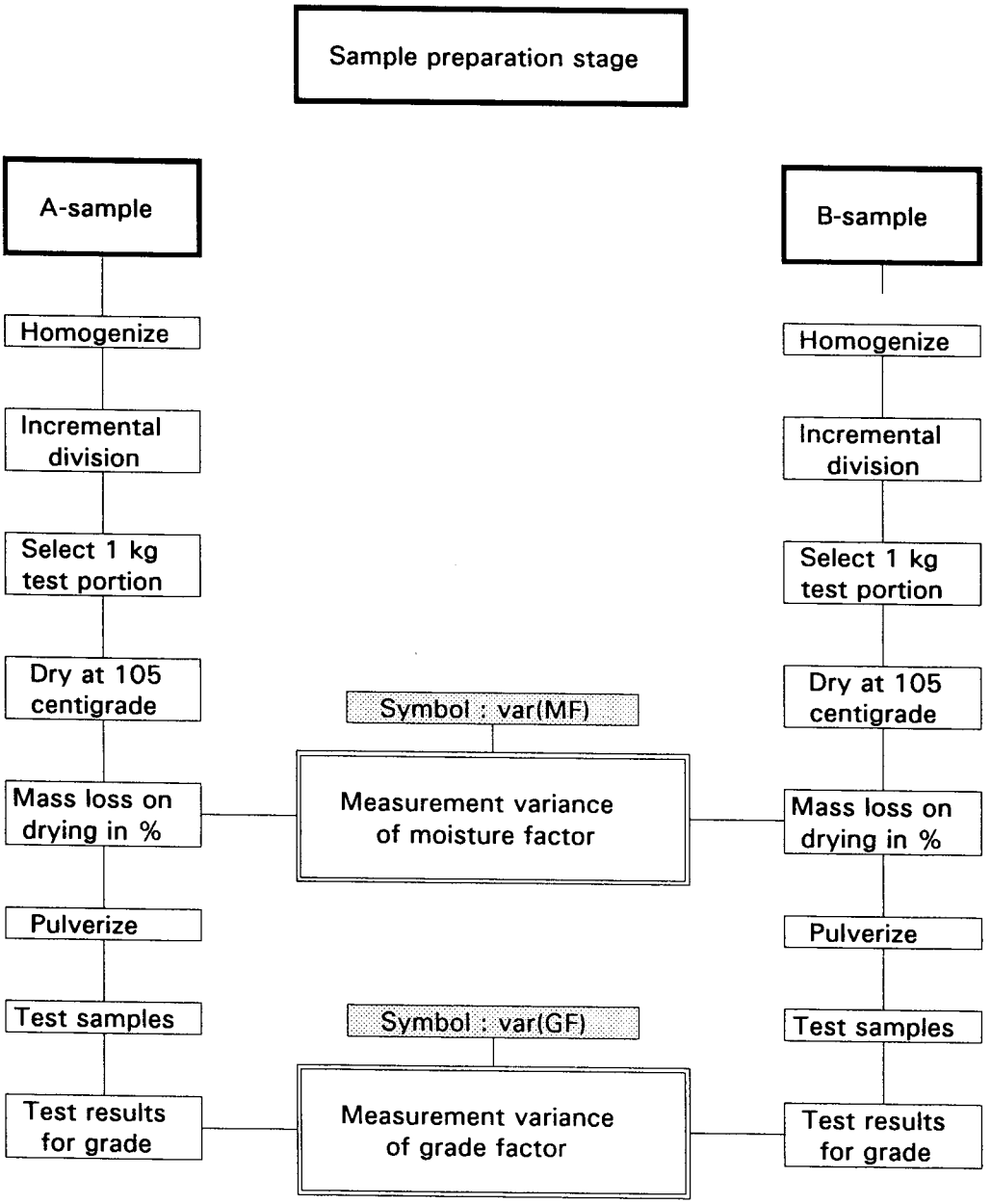
APPENDIX A

Interleaving Sampling Protocol



APPENDIX B

Interleaving Sampling Protocol



APPENDIX C

Measures for Variability and Precision

Statistic	Symbol	Value
Arithmetic mean in %	xbar	30.90
Variance in %^2	var(x)	0.9950
Standard deviation in %	sd(x)	0.9975
Coefficient of variation in %rel	CV	3.2
Number of measured values	n	5
Variance of mean in %^2	var(xbar)	0.1995
Standard deviation in %	sd(xbar)	0.4467
95% Confdence interval in %abs #	95% CI	1.24
95% Confdence interval in %rel	95% CI	4.0
95% Confidence range	95% CR	
Lower limit in %abs	95% CRL	29.7
Upper limit in %abs	95% CRU	32.1
Degrees of freedom	df	4
Tabulated t-value at 5% probability	t0.05;df	2.776

Statistic	Symbol	Value
Mass weighted average in %	xbar	30.71
Variance in %^2	var(x)	0.8143
Standard deviation in %	sd(x)	0.9024
Coefficient of variation in %rel	CV	2.9
Sum of squared weighting factors	sum(w1i^2)	0.2401
Variance of mean in %^2	var(xbar)	0.1955
Standard deviation in %	sd(xbar)	0.4422
95% Confdence interval in %abs #	95% CI	1.37
95% Confdence interval in %rel	95% CI	4.5
95% Confidence range	95% CR	
Lower limit in %abs	95% CRL	29.3
Upper limit in %abs	95% CRU	32.1
Degrees of freedom	df	3.2
Tabulated t-value at 5% probability ##	t0.05;df	3.100

\# based on 95% CI = sd(xbar)*t0.05;df

by linear interpolation

APPENDIX D

Sampling Variogram for On-stream Data

Variance terms of ordered set		Symbol	varj(x)	var(x)	95% ACRL	99% ACRL
1	1st	var1(x)	0.0038	0.0790	0.0637	0.0581
	2nd	var2(x)	0.0038	0.0790	0.0637	0.0581
	3rd	var3(x)	0.0070	0.0790	0.0637	0.0581
	4th	var4(x)	0.0106	0.0790	0.0637	0.0581
5	5th	var5(x)	0.0140	0.0790	0.0637	0.0581
	6th	var6(x)	0.0161	0.0790	0.0637	0.0581
	7th	var7(x)	0.0205	0.0790	0.0637	0.0581
	8th	var8(x)	0.0253	0.0790	0.0637	0.0581
	9th	var9(x)	0.0313	0.0790	0.0637	0.0581
10	10th	var10(x)	0.0355	0.0790	0.0637	0.0581
	11th	var11(x)	0.0385	0.0790	0.0637	0.0581
	12th	var12(x)	0.0417	0.0790	0.0637	0.0581
	13th	var13(x)	0.0440	0.0790	0.0637	0.0581
	14th	var14(x)	0.0483	0.0790	0.0637	0.0581
15	15th	var15x)	0.0511	0.0790	0.0637	0.0581
	16th	var16x)	0.0535	0.0790	0.0637	0.0581
	17th	var17(x)	0.0548	0.0790	0.0637	0.0581
	18th	var18x)	0.0550	0.0790	0.0637	0.0581
	19th	var19(x)	0.0587	0.0790	0.0637	0.0581
20	20th	var20(x)	0.0618	0.0790	0.0637	0.0581

Statistic	Symbol	Value
Degrees of freedom for :		
Randomized set	df(r)	95
Ordered set	df(o)	190
Tabulated F-value at :		
5% Probability	F0.05;95;oo	1.24
1% Probability	F0.01;95;oo	1.36

APPENDIX E

Student's t-test for Paired Test Results

Statistic	Symbol	Value
Aritmetic mean in % : reference increments	xbar(R)	8.10
Aritmetic mean in % : system samples	xbar(s)	7.49
Difference in %abs	dx	0.61
Difference in %rel	dx	8.1
Variance of differences in %^2	var(dx)	0.5728
Standard deviation in %	sd(dx)	0.7569
Coefficient of variation in %rel	CV	9.3
Number of paired test results	n	30
Variance of difference	var(dxbar)	0.0191
Standard deviation	sd(dxbar)	0.1382
Calculated t-value	t	4.390
Significance		***
Bias detection limits in %abs	BDLs	
Type I statistical risk only	BDL(I)	0.28
Type I & II statistical risks	BDL(I&II)	0.52
Bias detection limits in %rel	BDLs	
Type I statistical risk only	BDL(I)	3.6
Type I & II statistical risks	BDL(I&II)	6.6
Probable bias range in %abs		
Type I risk only : lower limit	PBL(I)	0.32
Type I risk only : upper limit	PBU(I)	0.89
Type I & II risks : lower limit	PBL(I&II)	0.09
Type I & II risks : upper limit	PBU(I&II)	1.12
Degrees of freedom	df	29
Tabulated t-values at :		
10% Probability	t0.10;df	1.701
5% Probability	t0.05;df	2.048
1% Probability	t0.01;df	2.763
0.1% Probability	t0.001;df	3.674

*** significant at 0.1% probability