Sampling and Statistics

A Synopsis of Tools and Techniques

Compiled by:

Jan W Merks
President
Matrix Consultants Limited
1357 Napier Place
Coquitlam, British Columbia
Canada V3B 7A3

Phone/Fax : (604) 941-1213
Email : merksmatrix@telus.net
Website : geostatscam.com

Section	Contents	Page

1. Terminology and Symbols

Through the years mathematical probability and applied statistics have developed a distinct jargon. Because so many scientific and engineering disciplines apply statistical tests and techniques, a common terminology with consistent symbols remains an elusive target. The International Organization for Standardization, and particularly Technical Committee 69 – *Applications of Statistical Methods*, is attempting to assist in replacing terms such as *margin of error, total error, nugget effect, sill value, semi-variogram*, and scores of others that have evolved in isolation.

This Appendix introduces statistical terms, symbols, tests and techniques required to test for bias, to estimate precision, to check for associative dependence, and to examine compatibility of measured values of stochastic variables and their variances. Table B lists the terms and symbols used to describe the basic statistics in the text of the report, and in the appendix with printed copies of spreadsheet templates.

TABLE B1

Terms and Symbol for Basic Statistics

Statistic	*Text*	*Template*
Arithmetic mean	$\bar{x}$	*xbar*
Weighted average	$\bar{x}$	*xbar*
Variance of randomly distributed set	*var(x)*	*var(x)*
Variance terms of ordered set	*varj(x)*	*varj(x)*
Standard deviation of randomly distributed set	*sd(x)*	*sd(x)*
Standard deviations for the ordered set	*sdj(x)*	*sdj(x)*
Coefficient of variation	*CV*	*CV*
Number of measured values	*n*	*n*
Variance of mean or average	$var(\bar{x})$	*var(xbar)*
Standard deviation	$sd(\bar{x})$	*sd(xbar)*
95% Confidence interval	*95% CI*	
Symmetric 95% confidence range	*95% CR*	
Lower limit	*95% CRL*	
Upper limit	*95% CRU*	
Asymmetric 95% confidence range	*95% ACR*	
either : lower limit	*95% ACRL*	
or : upper limit	*95% ACRU*	

Symbols that contain x_i, $\bar{x}(?)$ or *var(?)* are also used to define other statistics. For example, *var(t)* is the total variance of a measurement chain, *var(a)* is the variance of the analytical stage, and $\bar{x}(A)$ and $\bar{x}(B)$ are the central values for A- and B-sets (test results for odd- and even-numbered primary samples obtained by partitioning a set of primary increments into a pair of interleaving or interpenetrating subsets).

The $\bar{x}$-symbol refers to the arithmetic mean of a set of measured values with equal weighting factors, and to the weighted average of a set of measured values with variable weighting factors. The same symbols are used for the arithmetic mean and the weighted average because the context clarifies to which of these estimates of the central value the $\bar{x}$-symbol refers.

Statistical tests and techniques to estimate precision, to test for absence or presence of bias, and to examine associative dependence between measured values of stochastic variables (random variables or variates) require additional terms and symbols. Table B2 lists the terms and symbols used to describe these statistics in the text and in the templates:

TABLE B2

Terms and Symbols for Statistical Tests

Statistic	*Text*	*Template*
Difference	Δx	*dx*
Mean difference	$\Delta\bar{x}$	*dxbar*
Variance of differences	$var(\Delta x)$	*var(dx)*
Standard deviation	$sd(\Delta x)$	*sd(dx)*
Variance of mean difference	$var(\Delta\bar{x})$	*var(dxbar)*
Standard deviation	$sd(\Delta\bar{x})$	*sd(dxbar)*
Bias detection limits	*BDLs*	
Type I statistical risk only	*BDL(I)*	
Type I & II statistical risks	*BDL(I&II)*	
Probable bias ranges	*PBRs*	
Type I risk : lower limit	*PBL(I)*	
Type I risk : upper limit	*PBU(I)*	
Type I & II risks : lower limit	*PBL(I&II)*	
Type I & II risks : upper limit	*PBU(I&II)*	
Student's t-value	*t*	
Tabulated t-values	*t0.0x;n-1*	
Fisher's F-value	*F*	
Tabulated F-values	*F0.0x;n1-1;n2-1*	
Bartlett's χ^2*-value*	χ^2	*X^2*
Tabulated χ^2*-value*	$\chi^2 0.0x;df$	*X^2;0.0x;df*

The symbols under the heading *Text* are often used in applied statistics and sampling theory because they are simple to print in the text. The symbols under the heading *Template* are selected because they are easier to print and recognize in spreadsheet templates. The symbol *0.0x* refers to either 5%, 1% or 0.1% probability. The terms and symbols in Tables B1 and B2 are not only used in this appendix but also in the Matrix report of which it forms part.

2. Definitions

Basic concepts such as accuracy, precision and bias have received a great deal of attention not only from ISO/Technical Committee 69 – *Application of Statistical Methods*, but even more so from ISO Technical Committees that deal with the sampling of materials in bulk such as coals, concentrates, and various types of ores.

2.1. Accuracy

The following definition for accuracy reflects its descriptive nature in measurement technology.

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

This definition reflects that *accuracy* is an abstract concept. By contrast, a lack of accuracy can be measured and quantified in terms of a systematic error or bias. Webster defines *accuracy* as *free from error*. By implication, unbiased measurements are accurate by definition.

2.2. Bias

Testing for the absence or presence of bias is an essential part of sampling in mining and metallurgy. The term *systematic error* is merely a synonym for bias. Terms such as *random error*, or *error* without adjectives, will not be used to avoid possible confusion with *random variations*.

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 of interest.*

Testing for relative bias and estimating analytical precision are key elements of statistical quality control (SQC). Tests for analytical bias require 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. The presence of a bias at the primary sample selection and sample preparation stages is more difficult to detect, and at times impossible to eliminate.

Tests for bias are commonly based on Student's t-test for paired data. This test can be applied to paired test results determined by employing different analytical procedures to duplicate test portions taken from each of a set of test samples. It can also be applied to paired test results determined by employing different sample preparation procedures to duplicate test samples taken from each of a set of secondary or primary samples. 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.

Analysis of variance (ANOVA) is a powerful technique to test for analytical bias when three or more laboratories participate in interlaboratory crosscheck programs. One-way ANOVA is applied to sets of test results determined by the same laboratory with the same analytical method to replicate test portions taken from each of five (5) up to ten (10) test samples that are prepared from a sample mass under carefully controlled conditions. This test should precede a crosscheck program to ensure that all participating laboratories receive test samples from a homogeneous set.

Two-way ANOVA is applied when different laboratories employ the same analytical method to analyze duplicate test portions taken from each of a set of no less than five (5) test samples. Logically, such sets of test samples should pass the test for homogeneity.

2.3. Precision

The definition for precision also reflects its descriptive nature and its lack of purport as a quantitative measure.

Precision *A generic term that refers to the magnitude of random variations associated with 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.

2.4. Sample

Webster presents definitions for simple and advanced applications. Given that the term *"representative"* is widely abused and misused in sampling practice, a more succinct definition is presented.

Sample *A part selected from the whole such that a measured value for the part is an unbiased estimate for the whole.*

2.5. Sampling

Some texts refer to sampling as the process of selecting a representative part of a population. The problem with this definition in sampling practice can be avoided if the objective of sampling is explained in statistical terms.

Sampling *The process of selecting a set of primary increments from a sampling unit, combining the set of primary increments into a pair of primary samples, preparing test samples from primary samples, and analyzing test portions of test samples, such that the stochastic variable of interest can be estimated in a unbiased manner and with an acceptable and affordable degree of precision.*

Stratified systematic sampling is commonly applied to materials in bulk such as coals, concentrates and ores, preferably during transfer on a conveying system, or while in storage when necessary. Stratified systematic sampling is the process of dividing a quantity of material (the sampling unit) into a symmetric set of cells or strata (elementary units), and selecting a primary increment throughout the center of each elementary unit. The degree of associative dependence between ordered sets of test results for single primary increments (spatial dependence) impacts the variances of ordered sets, and, thus, the precision for the stochastic variable of interest.

In practical sampling applications, a set of primary increments is composited into a pair of interleaving or interpenetrating subsets in such a manner that one subset (A-sample) consists of all odd-numbered primary increments, and the other (B-sample) consists of all even-numbered primary increments. This sampling protocol gives unbiased precision estimates at the lowest possible costs. In the case of mineral concentrates, it does so at no additional cost if the mass of a lot is increased by a factor 2.

Random sampling and stratified random sampling are commonly applied to consumer products but they should not be applied to materials in bulk simply because test results for a single primary sample for a single sampling unit elude rigorous statistical analysis.

3. Measures for Central Tendency

The arithmetic mean and the weighted average are measures for the central values of probability distributions encountered in sampling practice. Each of these measures gives the expected value of the stochastic variable (the most probable estimate of its unknown true value). This statement implies that stochastic variables can only be measured with a finite degree of precision. For each measured value is an estimate, and, one would hope, an unbiased estimate for that elusive unknown true value.

For most applications the arithmetic mean of a set of measured values is the most reliable (least biased) estimate for the central value of the probability distribution. The formula for the arithmetic mean is trivial:

$$\bar{x} = \sum xi/n$$

where :

$\bar{x}$ = *arithmetic mean*
xi = *ith measured value*
n = *number of measured values*

In mining and metallurgy the weighted average is often a more reliable estimate of the expected value than the arithmetic mean is. For example, the mass weighted average is the most reliable estimate for the expected value for a set of core samples of uniform length but variable density while the length and mass weighted average is the most reliable estimate for a set of core samples of variable length and density. In geostatistics, the distance weighted average mutated into a *kriged estimate*. A set of measured values has only one arithmetic mean but two measured values in a n-dimensional sample space define an infinite set of distance weighted averages. Selecting the least biased subset of kriged estimates from the infinite set is a daunting task indeed. The more so because all kriged estimates converge on the arithmetic mean as the distances increase between their position and those of the measured values.

The general formula for the weighted average of a set of ***n*** measured values with variable weighting factors is:

$$\bar{x} = \sum^{n} wi\, xi$$

where :

$\bar{x}$ = *weighted average*
wi = *weighting factor for ith measured value*
xi = *ith measured value*
n = *number of measured values*

Because $\sum(1/n)=1$ for the arithmetic mean, the same condition should apply to the weighted average so that $\sum wi=1$. It applies to the length weighted average of a set of core samples of variable length, to the length and mass weighted average of a set of core samples of variable length and density, and to the distance weighted average of a set of measured values in a n-dimensional sample space.

4. Measures for Variability

The variance is the fundamental measure for variability and precision. Variances are amenable to mathematical analysis. The additive property of the variances of volume, mass and content in particular plays an important role in mining and metallurgy. The basic formula for the variance of a set of measured values with equal weighting factors is:

$$var(x) = \frac{\sum (x_i - \bar{x})^2}{n-1}$$

where :

$\bar{x}$	=	*arithmetic mean*
x_i	=	*ith measured value in the set*
n	=	*number of measured value in the set*
$n-1$	=	*degrees of freedom*

The variance for a set of measured values with variable weighting factors w_i such that their sum meets the requirement $\sum w_i / n = 1$ is:

$$var(x) = \frac{\sum [w_i (x_i - \bar{x})]^2}{n-1}$$

where :

$\bar{x}$	=	*weighted average*
x_i	=	*ith measured value in the set*
w_i	=	*weighting factor for ith measured value*
n	=	*number of measured values in the set*
$n-1$	=	*degrees of freedom*

Both variance formulas require that the central value be calculated before the differences are squared and added. Before the advance of computers these formulas caused some measure of uncertainty due to rounding. The next formula is equivalent to the basic formula for the arithmetic mean but more precise and much easier to use with spreadsheet software:

$$var(x) = \frac{\sum x_i^2 - (\sum x_i)^2 / n}{n-1}$$

where :

$\sum x_i^2$	=	*sum of squared measurements*
$\sum x_i$	=	*sum of measured values*
n	=	*number of measured values in the set*
$n-1$	=	*degrees of freedom*

This formula is implemented in all spreadsheet software to obtain ***var(x)***, the sample variance (the variance of a sample of the population as opposed to σ^2, the population variance). The formula for the variance of a set of measured values with variable weighting factors cannot be simplified but the calculation is elementary with spreadsheet software.

Given that the order in which the squared differences between $\bar{x}$ and x_i are summed is irrelevant, this formula gives the variance of the randomly distributed set of measured values (the randomized set). If the set is ordered, either in space (exploration data) or in time (on-stream data), testing for associative dependence in each of these sample spaces is an important element of statistical analysis.

The formula for the variance of ordered sets of measured values has found scores of applications in mineral exploration, mining and processing:

$$var_j(x) = \frac{\sum (x_{i+j} - x_i)^2}{2(n-1)}$$

where :

$var_j(x)$ = *jth variance term of ordered set*
x_{i+j} = *(i+j)th measured value*
x_i = *ith measured value*
j = *jth spacing between measured values*
n = *number of measured values for jth variance term*
$2(n-1)$ = *degrees of freedom for jth variance term*

Some textbooks on geostatistics give a homologue of this formula in which the variance is multiplied with the factor 2, and the denominator is rounded to ***n***. Nonetheless, the term ***2(n − 1)*** gives the degrees of freedom for the ordered set of ***n*** measured values of the stochastic variable of interest. The concept of degrees of freedom is fundamental in applied statistics but quite irrelevant in geostatistical practice.

Because each variance term for the ordered set is multiplied with the factor 2, the graph in which the terms are plotted against their spacing ***j*** gives a semi-variogram in geostatistics. In sampling theory and practice, however, a graph with the correct variance terms is simply a sampling variogram. If the variance of the randomized set and the lower limits of its asymmetric 95% and 99% confidence limits are plotted in the same sampling variogram, it shows whether the degree of associative dependence is significant and where it has dissipated into randomness. A sampling variogram gives a visual interpretation of analysis of variance (ANOVA) when applied to examine the degree of associative dependence in a sample space.

The requirement of functional or mathematical independence is fundamental in probability theory. Its corollary in applied statistics is the concept of degrees of freedom. Since a *kriged estimate* is a dependent variable whose value is defined by its position in relation to the set of measured values, it is a functionally dependent variable. Therefore, the variance of a set of kriged estimates is a mathematical aberration. Just the same, kriging variances and covariances became the quintessence of geostatistics.

ANOVA is one of the most powerful tools in applied statistics. Fisher's F-test is the essence of ANOVA. The test is applied to determine whether two variances are either statistically identical or differ significantly by comparing their ratio (the calculated F-value) with tabulated F-values at 5% and 1% probability levels with applicable degrees of freedom.

For example, the existence of spatial dependence at spacing ***j*** can be 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-ratio is below the tabulated F-value, then the variances are statistically identical, and their difference is merely a random number. A significant F-ratio, by contrast, implies the existence of spatial dependence. In mineral exploration, this condition implies that the probability of continued mineralization between measured values in the sample space under examination can be quantified.

Mathematical analysis ought not to be applied to differences between pairs of statistically identical variances. In geostatistics, however, they are entered into smoothing relationships to predict tonnages and grades. Given that the difference between a pair of statistically identical variances is merely a random number, it follows that entering it into any relationship is simply an abuse of applied statistics.

The next formula gives the variance of differences between identifiably different sets of paired data such as test results determined in different samples, in different laboratories, at different times in the same laboratory, or by applying different analytical techniques:

$$var(\Delta x) = \frac{\sum \Delta xi^2 - (\sum \Delta xi)^2/n}{n-1}$$

where :

$var(\Delta x)$	=	*variance of differences*
$\sum \Delta xi^2$	=	*sum of squared differences*
$\sum \Delta xi$	=	*sum of differences*
n	=	*number of paired measurements in the set*
$n-1$	=	*degrees of freedom*

The variance of differences between identifiably different sets of paired data and the number of pairs determine the sensitivity or power of Student's t-test to detect a systematic error. Hence, the variance of differences and the number of pairs are the most salient statistics when testing for bias and calculating bias detection limits and probable bias ranges.

The following formula gives the variance of a single measured value from a set of simultaneous duplicates (as opposed to staged replicates):

$$var(x) = (\pi/4)*[(\sum |xi1-xi2|)/n]^2 = (\pi/4)*|\Delta\bar{x}|^2$$

where :

$var(x)$	=	*variance of a single measurement*
$xi1$	=	*first measurement for ith pair*
$xi2$	=	*second measurement for ith pair*
$\lvert\Delta\bar{x}\rvert$	=	*mean of absolute differences*
n	=	*number of pairs*

Absolute differences between simultaneous duplicates give a measure for precision only. By contrast, relative differences between identifiably different pairs such as staged replicates, with their signs taken into account, give measures for precision and bias.

The $\pi/4$-formula is less sensitive for extreme values than the Δx^2-formula. Therefore, it provides a diagnostic tool to check whether a set of measured values exhibits a normal distribution. If the set is normally distributed, the ratio between the variance of differences divided by the factor 2, and the variance between measured values, is not expected to exceed the tabulated F-value at 5% probability with the appropriate degrees of freedom.

Due to its squared dimension the variance does not make it easy to check and compare variability and precision at a glance. The standard deviation, which is the square root of the variance and has the same dimension as the variable of interest, is a more readily understood statistic when examining variability and precision.

Derived statistics such as the coefficient of variation (CV) are an effective measure for the precision of single measurements. Confidence limits such as confidence intervals (CIs), and lower limits (CRLs) and upper limits (CRUs) of symmetric confidence ranges (CRs), are readily understood measures for the precision of central values. At times, either the lower limits (ACRLs) or the upper limits (ACRUs) of asymmetric confidence ranges (ACRs) are more appropriate than those of symmetric confidence ranges.

In science and engineering 95% confidence intervals (95% CIs) and 95% confidence ranges (95% CRs) are often used but if the risk associated with the wrong decision is high, confidence limits at 99% and 99.9% probability should be considered. Confidence limits for ore reserves and resources might be defined at lower probability levels.

5. Degrees of Freedom

Given that degrees of freedom are fundamental in applied statistics it is quite extraordinary that they are irrelevant in geostatistics. The disappearance of degrees of freedom becomes even more peculiar upon realizing that kriging, the process of calculating functionally dependent values from independently measured values, not only enhances spatial dependence but also induces it where it does not exist.

The concept of degrees of freedom in applied statistics is the corollary of the basic requirement of functional or mathematical independence in probability theory. The formula for the variance of a sample selected from a population can be used to explain the relevance of degrees of freedom. For a set of ***n*** measured values the differences between all values and the arithmetic mean of the set are:

$$x_1-\bar{x},\ x_2-\bar{x},\ \ldots,\ x_i-\bar{x},\ x_{i+1}-\bar{x},\ \ldots,\ x_{n-1}-\bar{x},\ x_n-\bar{x}$$

By implication, the sum of ***n*** differences is equal to:

$$(x_1 + x_2 + \ldots + x_i + x_{i+1} + \ldots x_{n-1} + x_n) - n\bar{x}$$

By definition, the arithmetic mean for set of ***n*** measured values is:

$$\bar{x} = (x_1 + x_2 + \ldots + x_i + x_{i+1} + \ldots + x_{n-1} + x_n)/n$$

so that:

$$(x_1 + x_2 + \ldots + x_i + x_{i+1} + \ldots x_{n-1} + x_n) - n\bar{x} = 0$$

Logically, if a set of $n-1$ differences is given, the missing one is known because the sum of ***n*** differences is zero. Therefore, a set of ***n*** measured values has $n-1$ independent differences and one *(1)* dependent difference. Hence, a set of n measured values has $n-1$ degrees of freedom.

A single measured value does not give any information on the population variance. After all, $\sum(x_1-\bar{x})^2/(n-1)=0/0$, which is indeterminate as it ought to be. It can be demonstrated by induction that adding functionally dependent (calculated) values to a set of measured values does not add a single degrees of freedom.

6. Central Limit Theorem

The central limit theorem is fundamental in applied statistics and sampling practice. It underlies many applications such as bias testing of mechanical sampling systems and manual sampling procedures, and forms the basis for calculating confidence limits for arithmetic means, bias detection limits as measures for the power or sensitivity of Student's t-test, and probable bias ranges as measures for the probabilistic ranges within which an observed bias of systematic error is expected to fall.

The formula for the variance of the arithmetic mean of ***n*** independently measured values with equal weighting factors is:

$$var(\bar{x}) = \sum^{n} (1/n)^2\, var(xi) = var(xi)/n$$

where :

$var(\bar{x})$ = *variance of arithmetic mean*
$var(xi)$ = *variance between measured values*
n = *number of measured values*

The sum formula reflects an intermediate step in the derivation of the central limit theorem from the variance of a general function as given in probability theory. The same formula can also be used to derive a homologue for the variance of the distance, mass or length weighted average of a set of ***n*** independently measured values with variable weighting factors:

$$var(\bar{x}) = \sum^{n} wi^2\, var(xi)$$

where :

$var(\bar{x})$ = *variance of weighted average*
$var(xi)$ = *variance for ith subset*
wi = *weighting factor for ith subset*
n = *number of measured values*

The formulas for the variances of the arithmetic mean and the weighted average become identical if each $w_i = 1/n$. The central limit theorem, when given as the sum of the products of squared weighting factors and variances, implies that the variance of the central values of several samples, selected from populations with different characteristics (multinomial, binomial or Poisson) converge on the Gaussian or normal probability distribution.

The central limit theorem also plays an important role in Student's t-test (see Section 8). For this application the theorem is formulated as follows:

$$var(\Delta\bar{x}) = var(\Delta xi)/n$$

where :

$var(\Delta\bar{x})$ = *variance of mean difference*
$var(\Delta xi)$ = *variance of differences*
n = *number of paired data*

7. Measures for Precision

The coefficient of variation (CV) is a useful measure to quantify the precision for a measurement procedure, or for each stage of a measurement chain. CVs can be plotted in SQC charts to monitor the precision of measurement systems and procedures as a function of time. Numerically, the CV is the standard deviation as a percentage of the arithmetic mean or the weighted average so that the following formula applies:

$$CV = \sqrt{var(x)}*100/\bar{x} = sd(x)*100/\bar{x}$$

where :
- $var(x)$ = *variance*
- $sd(x)$ = *standard deviation*
- x = *arithmetic mean or weighted average*

Given that its dimension is a percentage, the coefficient of variation makes it simple to compare at a glance the variability of a stochastic variable, or the precision of a measurement system or procedure used to obtain one or more estimates of the stochastic variable of interest.

Confidence intervals and ranges, too, are useful measures for the precision of central values. In many applications, 95% confidence intervals in absolute values and relative percentages, and symmetric 95% confidence ranges in absolute values are routinely reported. In hypothesis testing, however, asymmetric and symmetric 99% and 99.9% confidence ranges are also used. Confidence limits, either for a single measured value or for the arithmetic mean or weighted average, make intuitive and user-friendly precision estimates.

If the number of measured values used to estimate the variance is unknown but large, the 95% CI is the product of the standard deviation of the mean and the factor $z0.05 = 1.96$. This z-value from the Gaussian or normal distribution is usually rounded to 2 so that $95\%\ CI = sd(\bar{x})*z0.05 \simeq sd(\bar{x})*2$. If the number of measured values is known and small, the value tabulated in the t-distribution at $df = n-1$ degrees of freedom should be used to compute the 95% confidence interval so that $95\%\ CI = sd(\bar{x})*t0.05;n\text{-}1$.

The lower limit of the symmetric 95% confidence range is equal to the measured value minus the 95% confidence interval, and the upper limit is equal to the measured value plus the 95% confidence interval. In formula:

$$95\%\ CRL = \bar{x} - 95\%\ CI = \bar{x} - sd(\bar{x})*t0.05;n\text{-}1 \text{ or } \bar{x} - sd(\bar{x})*z0.05$$
$$95\%\ CRU = \bar{x} + 95\%\ CI = \bar{x} + sd(\bar{x})*t0.05;n\text{-}1 \text{ or } \bar{x} + sd(\bar{x})*z0.05$$

where :
- $\bar{x}$ = *arithmetic mean or weighted average*
- $sd(\bar{x})$ = *standard deviation of mean or average*
- $t0.05;n\text{-}1$ = *t-value at 5% probability*
- $z0.05$ = *z-value at 5% probability*

The lower limit of the asymmetric 95% confidence range is obtained by subtracting from the central value (the arithmetic mean or a weighted average) the 95% confidence interval as an absolute value. Similarly, the upper limit of the asymmetric 95% confidence range is obtained by adding to the central value the 95% confidence interval as an absolute value. In formula:

$$95\%\ ACRL = \bar{x} - 90\%\ CI = \bar{x} - sd(\bar{x}) * t0.10;n\text{-}1 \quad or \quad \bar{x} - sd(\bar{x}) * z0.10$$
$$95\%\ ACRU = \bar{x} + 90\%\ CI = \bar{x} + sd(\bar{x}) * t0.10;n\text{-}1 \quad or \quad \bar{x} + sd(\bar{x}) * z0.10$$

where :

$\bar{x}$	=	*central tendency*
$sd(\bar{x})$	=	*standard deviation of central tendency*
$t0.10;n\text{-}1$	=	*t-value at 10% probability*
$z0.10$	=	*z-value at 10% probability*

Lower limits and upper limits of asymmetric confidence ranges are mutually exclusive. In other words, such limits are one-sided in the sense that either the lower limit or the upper limit is valid at 95% probability. Together these lower and upper give, of course, a symmetric 90% confidence range.

Confidence limits at 95% and 99% probability are also effective control and action limits for statistical quality control (SQC) charts. More detailed information on this subject can be found in Section 14. References.

8. Student's t-test

The t-test is applied to assess whether the differences between identifiably different paired data sets (such as test results determined in different samples or in different laboratories, at different times in the same laboratory, or by applying different analytical techniques) are due to random variations or reflect the presence of a bias or systematic error. In each of those cases, the question is whether or not two means differs significantly, and, thus, that the mean difference is indicative of the presence of bias (reject null hypothesis). The alternative inference is that the means are statistically identical, and, thus, that the mean difference is statistically identical to zero (accept null hypothesis). In the latter case, the numerical difference reflects the effect of random variations in the measurement chain.

For example, if the mean difference between paired test results determined in different laboratories is statistically different from zero (reject null hypothesis), the analytical acumen of one of the laboratories is suspect. However, if the mean difference is statistically identical to zero (accept null hypothesis), it is most likely that both laboratories submitted unbiased test results.

It is possible but not highly probable that both laboratories submitted biased test results. Regrettably, the t-test does not reveal which laboratory is suspect but only that the mean difference between paired test results is suspect. Therefore, it is of critical importance that laboratories participate in interlaboratory crosscheck programs to ensure the absence of analytical bias and an acceptable degree of precision for the most critical elements.

Since Student's t-value is the ratio between the mean difference and its standard deviation, the following formulas apply:

$$t = \frac{\bar{x}1 - \bar{x}2}{sd(\Delta\bar{x})} = \frac{\Delta\bar{x}}{sd(\Delta\bar{x})}$$

where :

t	=	*calculated t-value*
$\bar{x}1$	=	*mean for first data set*
$\bar{x}2$	=	*mean for second data set*
$\Delta\bar{x}$	=	*mean difference*
$sd(\Delta\bar{x})$	=	*standard deviation of mean difference*

This formula shows that central limit theorem also underlies the relationship between $sd(\Delta\bar{x})$, the standard deviation of the mean difference, and $sd(\Delta x)$, the standard deviation of the differences between paired data. Given that $sd(\Delta\bar{x}) = \sqrt{[var(\Delta x)/n]} = sd(\Delta x)/\sqrt{n}$, it follows that three variables interact in the t-test and determine the calculated t-value.

The mean difference between paired test results is either identical to zero (accept null hypothesis), or indicative of the presence of bias between different measurement systems or procedures (reject null hypothesis). Similarly, the difference between the certified value of a Certified Reference Material (CRM) and the mean of a set of test results is either identical to zero or indicative of the presence of bias in the analytical procedure. In the latter case, the observed bias is either higher (positive) or lower (negative) than the certified value.

The variance of differences is a function of the differences between a set of paired data, and thus of the precision of all the systems and procedures used to obtain the set. Because the variance of difference and the number of paired data determine the bias detection limits, it is possible to prove that even a small and practically insignificant difference is a bias if the set is large enough. Therefore, bias detection limits (BDLs) are useful to quantify the sensitivity or power of the t- test to detect a bias, and to determine the number of test results required to prove statistical significance at some predetermined probability level.

The question whether a difference between two means is indicative of the presence of bias can be assessed by comparing the calculated t-value with tabulated t-values at 5%, 1% and 0.1% probability. Bias detection limits (BDLs) are more intuitive to examine the absence or presence of bias than is the comparison of calculated and tabulated values. Probable bias ranges (PBRs) define the probabilistic limits within which the observed bias is expect to fall. BDLs and PBRs will be addressed in the next sections.

8.1. Bias Detection Limits

Bias Detection Limits (BDLs) are measures for the power or sensitivity of Student's t-test to detect a bias or systematic error between identifiably different paired data. BDLs are defined for the Type I statistical risk only, and for the combined Type I and 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 alarm does not sound. The Type II statistical risk refers to the event that the alarm sounds but no fire occurs. The combined Type I and Type II statistical risks refer to the event that a fire occurs and the alarm sounds.

The effect of the number of paired data on the sensitivity of the Student's t-test becomes evident upon realizing that each of these statistical risks is obtained by multiplying the standard deviation of the mean difference either with the t-value at 5% probability, or with the sum of the t-values at 5% and 10% probability. Tabulated values of the t-distribution are a function of degrees of freedom, and, thus, of the number of paired data in a set.

In science and engineering, a symmetric two-sided 5% probability for the Type I statistical risk only and an asymmetric one-sided 5% probability for the Type II statistical risk are usually acceptable. The question of whether a mean difference is statistically identical to zero (accept null hypothesis), or indicative of the presence of bias (reject null hypothesis) can only be solved in probabilistic terms. In other words, the risk associated with a decision is always translates into a finite probability. Moreover, extremely low t-values are often of more concern than highly significant t-values because they may be indicative of tampering of test results. It is not surprising then that the t-test has found wide application in mineral exploration, mining, processing, smelting and refining.

Several ISO Standards on bias testing of mechanical sampling systems and manual sampling procedures specify these statistical risks in the same manner and at the same probability levels. Based on this convention the bias detection limits are defined as follows:

$$BDL(I) = sd(\Delta\bar{x}) * t0.05;n\text{-}1$$
$$BDL(I\&II) = sd(\Delta\bar{x}) * [t0.10;n\text{-}1 + t0.05;n\text{-}1]$$

where :

$BDL(I)$	=	*BDL for Type I statistical risk only*
$BDL(I\&II)$	=	*BDL for Type I and II statistical risks*
$sd(\Delta\bar{x})$	=	*standard deviation of mean difference*
$t0.10;n\text{-}1$	=	*tabulated t-value at 10% probability*
$t0.05;n\text{-}1$	=	*tabulated t-value at 5% probability*
n	=	*number of paired data*
$n-1$	=	*degrees of freedom*

Bias detection limits, for the Type I statistical risk only and for the combined Type I and II statistical risks, 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.

8.2. Probable Bias Ranges

If the mean difference between a set of paired test results is statistically significant, the bias detection limits for the Type I statistical risk and for the combined Type I and II statistical risks can be added to the observed bias and subtracted from it. The resulting probabilistic limits are referred to as Probable Bias Ranges (PBRs). In formula:

$$\begin{aligned} PBL(I) &= \Delta\bar{x} - BDL(I) \\ PBU(I) &= \Delta\bar{x} + BDL(I) \\ PBL(I\&II) &= \Delta\bar{x} - BDL(I\&II) \\ PBU(I\&II) &= \Delta\bar{x} + BDL(I\&II) \end{aligned}$$

where :

$\Delta\bar{x}$ = *observed bias*
$BDL(I)$ = *BDL for Type I risk only*
$BDL(I\&II)$ = *BDL for Type I and II risks*

$PBL(I)$ = *lower limit for Type I risk only*
$PBU(I)$ = *upper limit for Type I risk only*

$PBL(I\&II)$ = *lower limit for Type I and II risks*
$PBU(I\&II)$ = *upper limit for Type I and II risks*

Reporting probable bias ranges makes sense only if the mean difference is indeed indicative of the presence of bias (reject null hypothesis). Therefore, the difference between two means should exceed the bias detection limit for the Type I statistical risk before the corresponding probable bias range is reported. Similarly, it should exceed the bias detection limit for the Type I and Type II statistical risks before the corresponding probable bias range is reported. The abbreviation *na* (not applicable) is placed in the appropriate cell of the spreadsheet template if the observed mean difference is identical to zero (accept null hypothesis) and the probable bias range is undefined.

9. Tukey's WSD-Test

Tukey's wholly significant difference test is applied to determine whether sets of three (3) or more central values derive from the same population or from different populations. Alternatively, Tukey's test can be applied to test results for primary samples selected from the same population but under different conditions.

The test is applied to the absolute differences between all possible pairs of central values. Student's t-test can be applied to a set of three (3) means because three (3) means give only three (3) pairs. However, the number of pairs is 10 for a set of four, 15 for a set of five, and 21 for a set of six. If the t-test were applied to 21 pairs, one of the mean differences may well turn out to statistically significant simply because 1 out of 20 is expected to exceed the tabulated value of t0.05;df at 5% probability.

Tukey's wholly significant difference is computed from the weighted average variance, the number of sampling units, the number of primary samples, and *WSD/sd*($\bar{x}$), a value given in Tukey's table at the 5% probability level. If the absolute difference between a pair of means exceeds the wholly significant differences, it is justified to infer that the means differ significantly, and thus derive from statistically different sampling units. Alternatively, if the means are statistically identical, they derive from statistically identical sampling units, which, in turn, may derive from the same population.

10. Fisher's F-test

Fisher's F-test is applied to determine whether or not two variances are statistically identical. The F-test is based on comparing the ratio between the largest and smallest variance with tabulated values of the F-distribution at 5% and 1% probability. If the calculated F-value is below the tabulated value of $F0.05;n_1-1;n_2-1$, they are deemed compatible but if it exceeds the tabulated value of $F0.05;n_1-1;n_2-1$, the probability is less than 5% that the variances are compatible. Similarly, if the calculated F-value exceeds the tabulated value of $F0.01;n_1-1;n_2-1$, the probability is less than 1% that the variances are compatible. Tabulated F-values, too, reflect that the terms n_1-1 and n_2-1 are the degrees of freedom for the numerator and denominator in the calculated F-ratio.

The degree of associative dependence between an ordered set of measured values can be examined by comparing the ratio between the variance of the randomly distributed set and the first variance term of the ordered set with tabulated F-values at 5% and 1% probability with appropriate degrees of freedom. Unless the ratio between two variances is statistically significant, their difference is merely a random number to which mathematical analysis ought not to be applied. The fact that $F0.05;\infty;\infty = F0.01;\infty;\infty = 1$ explains the futility of analysis of variance without degrees of freedom.

On-stream measurements at mineral processing plants almost invariably display a significant degree of associative dependence. Metal grades of contiguous core samples within boreholes, or of adjoining rounds within drifts, may display a significant degree of associative dependence. When plotted in a sampling variogram the variance terms for an ordered set may exhibit a sampling variogram. When the variance of the randomly distributed set and the lower limits of its asymmetric 95% and 99% confidence ranges are plotted in the same sampling variogram, it will show whether the degree of associative dependence is statistically significant and where orderliness has dissipated into randomness.

11. Bartlett's χ^2-test

Bartlett's chi square test is applied to determine whether or not three (3) or more variances constitute a homogeneous set. The calculated χ^2-value is the difference between the product of the natural logarithm of the weighted average variance for all sampling units and the total number of degrees of freedom and the sum of the product of natural logarithm of the variance and the number of degrees of freedom for each sampling unit. In formula:

$$\chi^2 = \ln var(\bar{x}) * \sum (ki-1) - \sum [\ln var(xi) * (ki-1)]$$

where :

χ^2	=	*calculated chi square value*
$var(\bar{x})$	=	*weighted average variance*
$\sum ki$	=	*total number of degrees of freedom*
$var(xi)$	=	*variance of ith sampling unit*
ki	=	*measured values for ith sampling unit*

The calculated χ^2-value is compared with tabulated values of $\chi^2 0.05;df$ at 5% probability, $\chi^2 0.01;df$ at 1% probability, and $\chi^2 0.001;df$ at 0.1% probability. If the χ^2-test indicates that four or more variances do not constitute a homogeneous set, either the lowest variance or the highest variance should be rejected, and the reduced set be tested for homogeneity.

12. Correlation-Regression

Correlation implies the occurrence of associative dependence between pairs of stochastic variables, and regression quantifies this relationship. The basic measure for the degree of associative dependence is r^2, the coefficient of determination. The calculated correlation coefficient or r-value is compared with tabulated values of the r-distribution at different probability levels for the appropriate degrees of freedom to verify whether or not the degree of associative dependence is statistically significant.

A correlation coefficient of $r \simeq 0$ implies a complete lack of associative dependence, and a correlation coefficient of $r \simeq \pm 1$ an almost perfect degree of associative dependence. By comparing calculated r-values with tabulated values from the r-distribution, with appropriate degrees of freedom and at different probability levels, the level of significance can be quantified. F- and t-distributions can also be used to verify whether or not the covariances between pairs of measured values are statistically significant.

Following is the basic formula to calculate the correlation coefficient or r-value between stochastic variable x and y:

$$r = \frac{covar(x,y)}{\sqrt{[var(x)*var(y)]}}$$

where :	r	=	*correlation coefficient*
	$covar(x,y)$	=	*covariance between x and y*
	$var(x)$	=	*variance of measured x-value*
	$var(y)$	=	*variance of measured y-value*

Because $covar(x,y) = [\sum xy - \sum x \sum y/n]/(n-1)$, it follows that the covariance is simply a homologue of the formula for the variance. In fact, the term $n-1$ reflects the degrees of freedom for n pairs but they cancel in the above formula because the same term occurs twice in the denominator.

Precision is often a function of concentration and grade. For example, the question whether the correlation between the means of duplicate test results and the variance between each pair is significant can be resolved by testing whether their correlation coefficient is significant. The degree of association between grade and *in-situ* density, in massive sulphides in particular, should be taken into account when calculating confidence limits for grades and contents of ore reserves.

The covariance of a set of distance weighted averages *(kriged estimates)* is relevant in geostatistics but the covariance of a set of functionally dependent variables is a mathematical aberration in applied statistics. ANOVA ought to be applied to a covariance before significance is attached to the degree of associative dependence, and before the existence of a causal relationship is inferred. ANOVA requires F- or χ^2-distributions whose values depend on degrees of freedom. Therefore, the question of whether or not covariances are statistically significant cannot be solved if the concept of degrees of freedom is deemed irrelevant.

13. Variances of Functions

The variance is the basic measure for random variations in measured values of a stochastic variable such as a volume of *in-situ* ore, its grade or its density, or in measured values of a stochastic system such as the mass of metal contained in a volume of *in-situ* ore. Probability theory teaches that the variance of the mass of metal contained in a quantity of *in-situ* ore is a function of volume, *in-situ* density and grade, and of the variance of each of these variables. Similarly, the variance of the mass of metal contained in a quantity of mined ore or mineral concentrate is a function of mass, moisture content and metal grade, and of the variance of each of these variables.

The general formula for a function between a dependent variable $\boldsymbol{y}$ and a set of $\boldsymbol{n}$ independent variables is $y=f(x1, x2,..., xn)$. The variances of the variables $x1, x2,..., xn$ for this function are $\sigma^2(x1), \sigma^2(x2),..., \sigma^2(xn)$. In probability theory the $\boldsymbol{\sigma^2}$ symbol refers to the *population* variance. In applied statistics the *sample* variance, estimated from a finite set of measured values, replaces the unknown *population* variance. The replacement of $\boldsymbol{\sigma^2}$ with ***var(x)*** reflects the transition from probability theory to applied statistics.

This transition is not only a matter of changing a few symbols but involves scores of tables in which values of probability distributions are given as a function of degrees of freedom. Statistical tables imply that variances of finite sets of measured values do not instill the same level of confidence that those elusive population variances would.

The variances of basic functions such as a weighted average, the arithmetic mean and the metal content of a volume or a mass, can be derived from the formula for the variance of a general function as defined in probability theory. The variance of the weighted average, despite its significance in exploration, mining, mineral processing, smelting and refining, is not as well-documented in textbooks on applied statistics as is its homologue for the arithmetic mean. Just the same, the ubiquitous Central Limit Theorem, the fundamental relationship between the variance of a set of measured values with equal weighting factors and the variance of the arithmetic mean derives from the formula for the weighted average of a set of measured values with variable weighting factors. The variances of both central values are of critical importance in sampling theory and practice.

The variances of contained metals for mill feed, concentrate and tailing are used in statistical simulation models for mineral processing plants. The additive property of the variance of contained metal are used to calculate confidence limits for grades and contents of ore reserves. ISO Technical Committee 183 on copper, lead and zinc sulphides concentrates, too, has accepted the variance of contained metal as the most basic measure for the risk that mines and smelters encounter when trading copper, lead and zinc concentrates (ISO/DIS 13543 1994).

13.1. Variance of a General Function

The formula for the variance of a general function between a dependent variable and a set of independent variables finds its origin in calculus and probability theory. The variance of a general function is equal to the sum of ***n*** terms each consisting of a squared partial derivative for a variable multiplied by its variance. In formula:

$$\sigma^2(y) = \left[\frac{\partial y}{\partial x1}\right]^2 * \sigma^2(x1) + \left[\frac{\partial y}{\partial x2}\right]^2 * \sigma^2(x2) + \ldots + \left[\frac{\partial y}{\partial xn}\right]^2 * \sigma^2(xn) \qquad [1]$$

in which : $\sigma^2(y)$ = *population variance of a general function*
$\sigma^2(x1)$ = *population variance of first variable*
$\sigma^2(x2)$ = *population variance of second variable*
$\sigma^2(xn)$ = *population variance of nth variable*

The formula for the variance of a general function is often referred to as the additive property of variances. This additive property is not necessarily commutative. In fact, differences between variances are only valid variance estimates if their ratio is statistically significant. Analysis of variance, one of the most powerful tools in applied statistics, can be used to partition the sum of the variances in a measurement hierarchy into its components.

13.2. 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. ISO Draft International Standard 13543 – *Determination of mass of contained metal*, describes how to calculate the variance of the mass of metal contained in a concentrate shipment. It is simple to calculate in a similar manner the mass of metal contained in a volume of *in-situ* ore. The additive property of variances is a practical attribute to assess the risk associated with the measurement of metal grades and contents of ore reserves.

13.2.1. *In-situ* Ore

The mass of metal contained in a quantity of *in-situ* ore is a function of its volume, density and metal grade, and of its moisture if conditions demand. If its moisture content does not impact the precision for grade and content, the formula for the mass of metal contained in an elementary unit of *in-situ* ore is:

$$Me = V * ID * GF$$

in which :

Me = *contained metal in mt*
V = *volume in* m^3
ID = *in-situ density in* mt/m^3
GF = *grade factor :* $\%Me/100$
$\%Me$ = *metal grade in percent on dry basis*

Substituting these variables and their variances in the formula for the variance of the mass of contained metal gives:

$$var(Me) = (ID*GF)^2*var(V) + (V*GF)^2*var(ID) + (V*ID)^2*var(GF)$$

Multiplying the volume term with V^2/V^2, the density term with ID^2/ID^2, and the grade factor term with GF^2/GF^2, dividing each of these terms by Me^2, and multiplying the sum of all terms with Me^2, gives the following formula for the variance of the mass of metal contained in a volume of *in-situ* ore with a given grade and density:

$$var(Me) = Me^2\,[var(V)/V^2 + var(ID)/ID^2 + var(GF)/GF^2]$$

in which :

$var(Me)$ = *variance of contained metal in* mt^2
$var(V)$ = *variance of volume in* $(m^3)^2$
$var(ID)$ = *variance of density in* $(mt/m^3)^2$
$var(GF)$ = *variance of grade factor (dimensionless)*

The elementary unit is defined as a quantity of *in-situ* ore for which a single grade estimate is available. The grade squared partition technique is applied to calculate the variance of the grade of each elementary unit from the variance of metal contained in a contiguous set of elementary units.

Almost invariably, the variance of grade, and, thus, the variance of the grade term $var(GF)/GF^2$, adds most to the variance of the mass of contained metal or metal content. How to obtain unbiased variance estimates for all variables that interact in the function of interest is a key element of metrology, the science of measurement.

13.2.2. Mill Feed, Concentrate or Tailing

The mass of metal contained in a quantity of concentrate, mill feed or tailing is a function of its mass, moisture content and metal grade. The formula for the mass of contained metal is:

$$Me = Mw * MF * GF$$

in which :

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

Substituting these variables and their variances in the formula for the variance of the mass of contained metal gives:

$$var(Me) = (MF*GF)^2*var(Mw) + (Mw*GF)^2*var(MF) + (Mw*MF)^2*var(GF)$$

Multiplying the wet mass term with Mw^2/Mw^2, moisture factor term with MF^2/MF^2, and the grade factor 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 formula for the variance of metal contained in mill feed, concentrate or tailing:

$$var(Me) = Me^2\,[var(Mw)/Mw^2 + var(MF)/MF^2 + var(GF)/GF^2]$$

in which :

$var(Me)$	=	*variance of contained metal in mt²*
$var(Mw)$	=	*variance of wet mass in mt²*
$var(MF)$	=	*variance of moisture factor (dimensionless)*
$var(GF)$	=	*variance of grade factor (dimensionless)*

In the case of precious metals extreme care should be exercised to ensure that the variables and variances are correctly entered into the appropriate formula for the variance of contained metal. Changing from metal grades in percent (%Me) to dimensionless metal factors (GF), from precious metal grades in g/mt to contained metals in kg, and from moisture contents (%H_2O) to dimensionless moisture factors (MF), requires close attention to derivatives, dimensions and decimal places.

The variances of contained metal can also be used in simulation models to generate precision estimates for recoveries at mineral processing plants (Merks 1991). The properties of variances can be utilized in a multitude of applications in mining and metallurgy.

14. References

ASTM Manual on Presentation of Data and Control Chart Analysis
Manual Series MNL 7, Philadelphia, 1991

Davies, O L and Goldsmith, P L, Statistical Methods in Research and Production
Longman Group, London, 1977

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

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

Mandel, J, The statistical analysis of experimental data
Dover Publications, New York 1964

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

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

Merks, J W and Merks, E A T, Precision Estimates for Ore Reserves
Erzmetall 44, October 1991

Merks, J W, Sampling in Exploration: Theory and Practice
Sampling Practices in the Minerals Industry, November 1992

Merks, J W, Abuse of Statistics
CIM Bulletin, Volume 86, No 966, January 1993

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

Merks, J W, Process Simulation with Spreadsheet Software
Mnerals & Metallurgical Processing, Vol 16, No 2, May 1999

Merks, J W, Sampling in Mineral Processing
SME Symposium, Vancouver, BC, October20-24, 2002

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

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