CHAPTER 3
CLASSICAL STATISTICAL THEORY IN GEOSTATISTICS

3.1 Introduction

Classical statistical theory can be effectively applied in any geostatistical analysis. Recall from Chapter 1 that geostatistical data contains the measurement of an attribute at various locations within a spatial area. Classical statistics can be used to statistically analyze the measured attribute without taking into account the locations at which it was measured. Statistics such as the mean value and variance of the attribute can be estimated, as can its value distribution. The value distribution, mean value and variance of the attribute are actually very important when it comes to geostatistical analyses, as will be seen later in this dissertation.

This chapter presents some classical statistical techniques and considerations that can be successfully used in analyzing the measured attribute. The dissertation assumes that the reader has a statistical background, and thus basic classical statistical theory such as the measures of central tendency (mean, mode and median) and the measures of variability (variance and range) have not been included here. However, for the interested reader, a description of these and some other basic classical statistics and considerations have been included in Appendix B.

Note 3.1 Before continuing, it is important to clarify some concepts that were introduced in Chapter 2. Chapter 2 is concerned with geostatistics in the mining industry. As has already been mentioned, geostatistics can also be applied outside of the mining industry. It is important to realize that the concepts of support and the support effect are not unique to mining geostatistics. Regardless of the field in which the geostatistical analysis is carried out, the concepts of support and the support effect will hold.

Outside of the mining industry, a sample is simply a measurement of a non-mining attribute taken over a relatively small support, while a block value is simply a measurement that has been taken over a larger (block) support. In this dissertation, the value given to an attribute over a block is always equivalent to the average of all the point values of the attribute within that block. Thus, instead of support being defined as the volume, area or shape of a sample or block of ore, it can equally be defined as the volume, area or shape of a sample or block of something else that is unrelated to mining. Thus, regardless of whether geostatistics is being applied in mining, forestry or elsewhere, the attributed being studied can be measured over various supports, giving rise to the concepts of support and the support effect.

3.2 The Distribution of the Sample Population

(Section 3.2 is a summary of David, 1988 : 1-3).

Geostatistical studies start off with a set of sample values[†] taken at various locations within a spatial area. These samples are usually of point support, while estimates are often required of the attribute over a larger (block) support. The first step in the classical statistical analysis of the measured attribute is to construct a histogram from the available set of sample values.

3.2.1 Construction of the sample histogram

The construction of the sample histogram may appear straightforward and of little importance, but the converse is true. The sample histogram provides insight into the possible distribution of the sample population and therefore fulfils an important role in, amongst other things, simulation problems[‡]. Once the population distribution is estimated for the basic sampling units, it becomes

[†] In a mining geostatistical study these will be samples of ore (see definition in Section 2.2). In a non-mining geostatistical study these will be samples of a non-mining attribute - as has been described in Note 3.1.

[‡] In a geostatistical simulation (as in all simulations), it is required that the simulated distribution of the sample population be approximately the same as the actual distribution of the sample population.

possible to undertake the estimation of population distributions of units of different support (this is explained in Section 3.3).

The ideal situation would be for samples to be taken on a regular grid or at random locations from the spatial area. In most cases this would provide an unbiased sample histogram. Unfortunately, it is common practice in many disciplines for samples not to be taken independently. In the mining industry, for example, it is usual for the drilling of samples to be undertaken more intensively in areas of high mineral grade. This practice is to confirm the existence of these rich areas of the ore body. The histogram constructed from these sample values would definitely be biased. The high grade areas would be over represented resulting in a biased overall average and variance value.

Theoretical solutions for the construction of unbiased sample histograms are often difficult to implement and not very practical. Switzer (1977) offers one solution by suggesting that each sample be assigned a weight related to its polygon of influence. A far simpler solution, called the *declustering technique*, involves simply ignoring some of the samples in the over-sampled areas.

3.2.2 The declustering technique

The idea behind the declustering technique is to divide the spatial study area into cells of a certain size and then to randomly select one sample value from each cell. Averaging all sample values within a cell and using this value as the cell value is discouraged. Such a procedure would introduce smoothing and the problems associated with it[†]. An inherent shortcoming of the declustering technique is that although it is a practical solution to unbiased histogram construction, no unique solution will exist. The results depend on the samples that are chosen and will also differ as the sizes of the cells are varied.

[†] Additionally, since different cells contain different numbers of sample values, the average values of the cells will have different distributions.

3.3 The Distribution of the Measured Attribute over other Supports

(Section 3.3 is a summary of David, 1988 : 12).

In most geostatistical studies it is important not only to estimate the population distribution of the sample values, but also to estimate the distribution of block values of different support. Again, a number of theoretical solutions for obtaining the required block support distribution exist, many of which are tedious to implement. According to David (1988 : 12), some possible solutions are discussed by Matheron (1984). A far simpler solution, called the *affine correction model*, has been promoted by both Dagbert & David (1983) and Journel (1983).

3.3.1 The affine correction model

Suppose that samples have been taken at various locations within the spatial study area Ω and that an unbiased histogram for these samples has been created. Because of the support effect (see Section 2.4.1 and Note 3.1), the histogram associated with block values over a different support, say v, should have the same average value m as the sample histogram but a different variance $g^2(v|\Omega)$. Suppose that the samples are of point support. The blocks are then obviously of a greater support than the samples, and hence the variance $g^2(o|\Omega)$ of the sample histogram will be greater than the variance $g^2(v|\Omega)$ of the block support histogram. The block support variance $g^2(v|\Omega)$ can be estimated from the semi-variogram of the sample values (see Section 6.4 and Equations 6.21 and 6.22).

However, it is not sufficient to simply know the average value and variance of the block support histogram. To fully describe the block support histogram, the affine correction model assumes that the general shape of the sample and block histograms are the same. More strictly speaking, the assumption implies that the standardized value distributions of the samples and the blocks are identical, i.e.

$$\frac{(X-m)}{g(o\,|\,\Omega)} \sim \frac{(X_v-m)}{g(v\,|\,\Omega)}$$

where X and X_v are random variables representing the sample and block values of the attribute respectively.

This assumption means that

$$X_v \sim \frac{(X-m)g(v\,|\,\Omega)}{g(o\,|\,\Omega)} + m$$

and allows for the estimation of the cumulative distribution of block values of support v through the use of the formula

$$F_{X_v}(x_v) = F_X\left(m + \frac{(x-m)g(v\,|\,\Omega)}{g(o\,|\,\Omega)}\right).$$

3.4 The Lognormal Distribution

Thus far in Chapter 3 much has been said about estimating the statistical distribution of the measured attribute. In Chapter 2 the importance of the lognormal distribution was expressed. Geological attributes are by and large non-negative and display positively skew distributions that are often well approximated by the lognormal distribution. The lognormal distribution is widely used and referred to in this dissertation, and hence its theory is presented here. Much of the theory presented in Section 3.4 has been researched and taken from Johnson & Kotz (1970 : 112-136) and Crow & Shimizu (1988 : 2-3, 9-11, 27-32, 113-114, 122-123).

3.4.1 The two-parameter lognormal distribution

A positively skewed variable X is *lognormally distributed* with two parameters μ and σ^2 if $Y=\ln X$ is normally distributed with mean μ and variance σ^2.

In other words,

$$X \sim \mathrm{LN}(\mu,\sigma^2) \quad \text{if} \quad Y \sim \mathrm{N}(\mu,\sigma^2).$$

The probability density function of $X \sim \mathrm{LN}(\mu,\sigma^2)$ is

$$f(x)=\begin{cases}\dfrac{1}{\sqrt{2\pi}\sigma x}\exp\left\{\dfrac{-(\ln x-\mu)^2}{2\sigma^2}\right\} & x>0\\ 0 & x\le 0\end{cases}$$

The r^{th} moment of X about the origin follows directly from the moment generating function of the normal distribution and is given as

$$\mathrm{E}(X^r)=\mathrm{E}(e^{rY})=\exp\left(r\mu+\frac{1}{2}r^2\sigma^2\right). \tag{3.1}$$

Thus, from Equation 3.1 it follows that

$$\mathrm{E}(X)=\exp\left(\mu+\frac{1}{2}\sigma^2\right)=\tau \tag{3.2}$$

$$\begin{aligned}\mathrm{var}(X)&=\exp(2\mu+\sigma^2)(\exp(\sigma^2)-1)\\ &=\tau^2(\exp(\sigma^2)-1)\\ &=\tau^2(w-1)\\ &=\kappa^2\end{aligned}$$

$$\mathrm{skew}(X)=\sqrt{\beta_1}=(w-1)^{1/2}(w+2)$$

and

$$\mathrm{kurt}(X)=\beta_2=w^4+2w^3+3w^2-3$$

where $w=\exp(\sigma^2)$.

It can be seen that the variance of a lognormal distribution is not independent of its mean. Consequently, as is mentioned in Clark and Harper (2000 : 73), the variance of a lognormal distribution is not really an effective measure of spread, but rather a measure of skewness. A far better measure of spread is provided by σ^2, the variance of the logarithmic transformed variable Y (see Figure 3.1).

It is also easily shown that,

$$\text{median}(X) = \exp(\mu)$$
$$\text{mode}(X) = \exp(\mu - \sigma^2)$$

and, as is expected of a positively skew distribution,

$$\text{mode}(X) < \text{median}(X) < \text{E}(X).$$

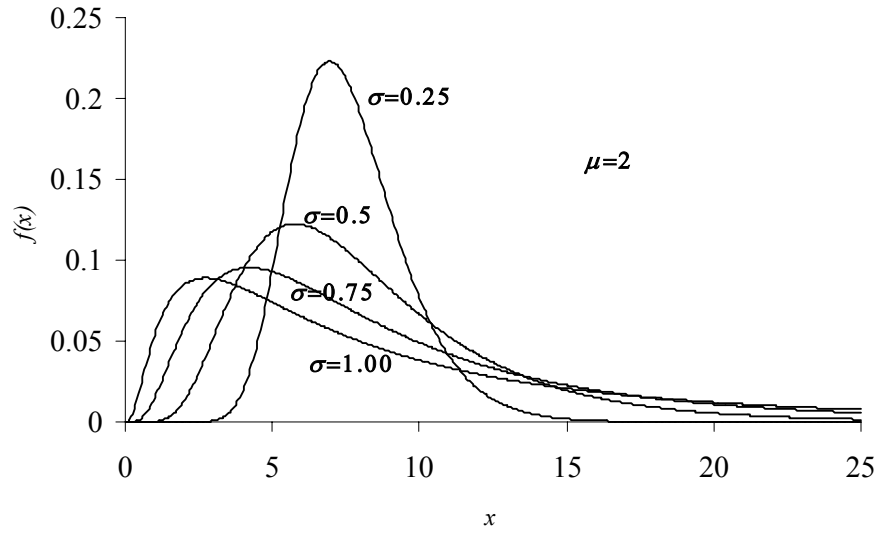

Figure 3.1 The probability density function of X~LN(2,σ^2) for σ = 0.25, 0.5, 0.75 and 1.00. This figure has been taken from Hogg & Tanis (1997 : 223) and shows the various shapes of the two-parameter lognormal distribution. It also illustrates that σ^2 is a good measure of spread of the lognormal distribution.

A number of techniques exist for determining whether a set of observations come from a two-parameter lognormal distribution. It is possible, for example, to calculate the natural logarithms of the set of observations and then to test these values for normality (see Appendix B). Alternatively, techniques involving the 'raw' observation values may be implemented, such as the use of logarithmic probability paper. The cumulative distribution of a lognormal distribution with two parameters will plot as a straight line on this graph paper (Rendu, 1981 : 5). More detail, including the estimation of the lognormal mean using logarithmic probability paper, can be found in Rendu (1981 : 5-7, 9).

Now, let $x_1, x_2, ..., x_n$ be a set of n observations taken independently from a LN(μ,σ^2) distribution. Define

$$y_i = \ln x_i$$

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

$$s_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2 .$$

Considering Equation 3.2, it would appear that the most instinctive estimator of the lognormal mean would be

$$\hat{\tau} = \exp\left(\bar{y} + \frac{1}{2}s_y^2\right) \tag{3.3}$$

This estimator is biased however, although it does produce a reasonable approximation when the number of observations n is large, say above 40 (Clark and Harper, 2000 : 73)

The uniformly minimum variance unbiased (UMVU) estimator of the mean and variance of a two-parameter lognormal distribution was calculated by Finney (1941) as

$$\hat{\tau} = e^{\bar{y}}\,{}_0F_1\left(\frac{n-1}{2}; \frac{(n-1)^2}{4n}s_y^2\right) \tag{3.4}$$

and

$$\hat{\kappa}^2 = e^{2\bar{y}}\left\{{}_0F_1\left(\frac{\mathrm{n}\text{-}1}{2}; \frac{(n-1)^2}{n}s_y^2\right) - {}_0F_1\left(\frac{n-1}{2}; \frac{(n-1)(n-2)}{2n}s_y^2\right)\right\} \tag{3.5}$$

respectively, where ${}_0F_1$ is a particular member of the generalized hypergeometric functions, i.e.

$$ {}_0F_1(\alpha, z) = \sum_{j=0}^{\infty} \frac{z^j}{(\alpha)_j \, j!} $$

$$ (\alpha)_j = \begin{cases} \alpha(\alpha+1)\ldots(\alpha+j-1) & \text{j} \geq 1 \\ 1 & \text{j} = 0 \end{cases} $$

and α and z are a complex (or real) number and complex (or real) variable respectively.

Comprehensive mathematical detail (including proofs) of these estimators can be found in Finney (1941) and Crow & Shimizu (1988 : 28-32).

A few years later, Sichel (1949) derived the same estimator for the two-parameter lognormal mean. He was unaware of Finney's work at the time (Sichel, 1966 : 107). Sichel named this estimator the t-estimator and produced tables for ease of calculation.

The t-estimator is defined as:

$$ \hat{\tau} = t = e^{\bar{y}} \gamma_n(V_y) \tag{3.6} $$

where

$$ \gamma_n(V_y) = {}_0F_1\left(\frac{n-1}{2}; \frac{(n-1)}{4} V_y\right) $$

and

$$ V_y = \frac{n-1}{n} s_y^2 $$

It is easily seen that Equations 3.4 and 3.6 are the same. However, instead of having to calculate $\gamma_n(V_y)$ by summation of an infinite series as in Equation 3.4, these values can be read off Sichel's tables. Included in this dissertation (Appendix A – Table A1) is a table of $\gamma_n(V_y)$ taken from Sichel (1966 : 117-118).

It turns out that the t-estimator "… is a much more precise estimator of the unknown true average [of a two-parameter lognormal distribution] … than the conventional arithmetic mean. A very practical advantage [of the t-estimator] is its non-sensitivity to one or two very high values…." (Sichel, 1966 : 108)

Part of the example that follows has been taken and adapted from Sichel (1966 : 108-109). In this example it is shown how to use the t-estimator and the tables presented by Sichel (1966).

Example 3.1

Consider a sample† consisting five observations that have been taken from a two-parameter lognormal distribution with a true average value τ of 698. The values of the five observations are 409, 357, 191, 269 and 5151. Suppose that τ is unknown and requires estimation. Using these five observations, the mean of the two-parameter distribution will be estimated using the

i. arithmetic mean of the five observations
ii. intuitive estimator, i.e. Equation 3.3
iii. the t-estimator

The sample† size is very small and one of the observations is an extreme tail-value from the two-parameter lognormal distribution. Statistically, these factors will certainly affect each of the three estimates.

i. $\hat{\tau} = \frac{1}{5}[409 + 357 + 191 + 269 + 5151] = 1\,275.4$

ii. By defining $y_i = \ln(x_i)$ for $i = 1, 2, \ldots, 5$

$\bar{y} = 6.257$

$s_y^2 = 1.724$

and

† Statistical definition.

$$\hat{\tau} = \exp\left[(6.257) + \frac{1}{2}(1.724)\right]$$
$$= 1\,235.152$$

iii. $\hat{\tau} = t = \exp(\bar{y})\gamma_n(V_y)$

From ii.

$$\bar{y} = 6.257$$

$$V_y = \frac{4}{5}s_y^2 = 1.379$$

Reading from Table A1 in Appendix A, it is seen that

$$\gamma_5(1.3) = 1.807$$
$$\gamma_5(1.4) = 1.884.$$

Thus, by linear interpolation

$$\gamma_5(1.379) = 1.868.$$

Therefore,

$$\hat{\tau} = \exp(6.257)\gamma_5(1.379)$$
$$= 974.43$$

All three estimates were affected by the large tail-value and the small sample† size. However, the t-estimator was not affected to the same extent as the other two estimates, and is best suited for estimating the mean of a two-parameter lognormal distribution from a small sample† size.

A situation such as in Example 3.1 may occur in mining. When a mineral deposit is initially identified, a small number of boreholes are usually drilled, say between

† Statistical definition.

5 and 20, and used to estimate the average grade and tonnage of the potential ore deposit (Sichel, 1987 : 3). It has already been mentioned that mineral grades often follow a two-parameter lognormal distribution and "as boreholes are usually sited long distances apart, we can say with some confidence that the serial correlation between borehole intersection areas is zero for all practical purposes." (Sichel, 1987 : 3). The t-estimator can thus be successfully used in such situations.

Construction of confidence intervals for the t-estimator is also possible. Sichel (1966 : 109-114) shows that the central 100(1-p)% confidence interval for the true mean of the two-parameter lognormal distribution is

$$\left[t \times \Psi_{p/2}\left(V_y, n\right), t \times \Psi_{1-p/2}\left(V_y, n\right) \right].$$

Once again, tables of $\Psi_{\frac{p}{2}}\left(V_y, n\right)$ and $\Psi_{1-\frac{p}{2}}\left(V_y, n\right)$ have been produced by Sichel for ease of calculation.

The exact mathematical details of the function $\Psi_{\frac{p}{2}}\left(V_y, n\right)$ are relatively complicated, and are presented in Sichel (1966).

Included in Appendix A is Table A2, a table of $\Psi_{0.05}\left(V_y, n\right)$ and $\Psi_{0.95}\left(V_y, n\right)$ values as taken from Sichel (1966 : 119). To illustrate the calculation of such a confidence interval, consider the next example, which is a continuation of Example 3.1.

Example 3.2 (Taken from Sichel, 1966 : 114-115)
In Example 3.1 Sichel's t-estimator gave an estimate of 974 for the mean of the two-parameter lognormal distribution. Furthermore, we had $n = 5$ and $V_y = 1.379$.

Suppose that a 90% central confidence interval for the true mean τ is required. From Table A2 it is seen that:

$$\Psi_{0.05}\left(1.3, 5\right) = 0.421$$

$$\Psi_{0.95}(1.3, 5) = 11.75$$

and

$$\Psi_{0.05}(1.4, 5) = 0.4083$$

$$\Psi_{0.95}(1.4, 5) = 13.55.$$

Therefore, by linear interpolation

$$\Psi_{0.05}(1.379, 5) = 0.411$$

$$\Psi_{0.95}(1.379, 5) = 13.17.$$

The 90% central confidence interval for τ is thus calculated as

$$\tau \in [400, 12828].$$

It can be seen that the true mean of the two-parameter lognormal distribution (viz. $\tau = 698$) does indeed fall within this interval. The construction of such confidence intervals via Sichel's t-estimator approach to the estimation of two-parameter lognormal means is definitely an advantage over the other two approaches also explored in Example 3.1.

One concern that may arise from this example is the high value of the upper limit of the confidence interval. Sichel (1966 : 115) assures that the reason for this is the purely the result of the very small sample[†] size of only five observations.

3.4.2 The three-parameter lognormal distribution

A positively skewed variable *X* is *lognormally distributed* with three parameters μ, σ^2 and β if $Y=\ln(X-\beta)$ is normally distributed with mean μ and variance σ^2.

In other words,

$$X \sim \text{LN}(\mu, \sigma^2, \beta) \quad \text{if} \quad Y \sim \text{N}(\mu, \sigma^2).$$

[†] Statistical definition.

The probability density function of X~LN(μ,σ^2,β) is

$$f(x)=\begin{cases}\dfrac{1}{\sqrt{2\pi}\sigma(x-\beta)}\exp\left\{\dfrac{-(\ln(x-\beta)-\mu)^2}{2\sigma^2}\right\} & \beta<x<\infty \\ 0 & x\le\beta\end{cases}$$

and β is often called the threshold parameter.

Note 3.2 In certain statistical and geostatistical texts the three-parameter lognormal distribution is described slightly differently. In these texts X is said to be three-parameter lognormally distributed if Y=ln(X+α) ~ N(μ,σ^2). The parameter α is then appropriately called the additive constant. It can be seen, however, that by simply replacing the constant α by $-\beta$, this description of the three-parameter lognormal distribution and the description presented in this dissertation are exactly the same.

The r^{th} moment of X about β is

$$E\left[(X-\beta)^r\right]=\exp\left(r\mu+\frac{1}{2}r^2\sigma^2\right) \qquad (3.7)$$

since the random variable (X-β) possesses a two-parameter lognormal distribution.

Thus, from Equation 3.7

$$\mathrm{E}(X)=\beta+\exp\left(\mu+\frac{1}{2}\sigma^2\right)$$

$$\mathrm{var}(X)=\exp(2\mu+\sigma^2)\{\exp(\sigma^2)-1\}$$

$$\mathrm{skew}(X)=\sqrt{\beta_1}=(w-1)^{1/2}(w+2)$$

and

$$\text{kurt}(X) = \beta_2 = w^4 + 2w^3 + 3w^2 - 3$$

where $w = \exp(\sigma^2)$.

It is also easily shown that

$$\text{median}(X) = \exp(\mu) + \beta$$
$$\text{mode}(X) = \exp(\mu - \sigma^2) + \beta\,.$$

The cumulative frequency of a three-parameter lognormal variable X will not plot as a straight line on logarithmic probability paper (Rendu, 1981 : 5). The variable (X-β), being lognormally distributed with two parameters, will however plot as a straight line. Therefore, one method for determining whether a set of n observations $v_1, v_2, \ldots, v_n$ come from a three-parameter lognormal distribution is to see if it is possible to obtain an estimate $\hat{\beta}$ such that the cumulative probability of the observations $(v_1 - \hat{\beta})$, $(v_2 - \hat{\beta})$, ..., $(v_n - \hat{\beta})$ plots as a straight line on the logarithmic probability paper.

Numerous methods have been proposed for obtaining the estimate $\hat{\beta}$. According to Clark & Harper (2000 : 77) one method is to experiment with various $\hat{\beta}$ values and choose the $\hat{\beta}$ value which produces a sample[†] skewness coefficient based on the set of observations $\ln(v_1 - \hat{\beta})$, $\ln(v_2 - \hat{\beta})$, ..., $\ln(v_n - \hat{\beta})$ closest to zero.

Rendu (1981: 7) suggests the estimate

$$\hat{\beta} = \frac{f_1 f_2 - m^2}{f_1 + f_2 - 2m}$$

where m corresponds to the median of the set of observations $v_1, v_2, \ldots, v_n$, and f_1 and f_2 are the values associated with p and 1-p cumulative frequency of the

[†] Statistical definition.

sample† distribution respectively. Rendu (1981 : 7) also offers a proof for this estimate. It is advised that for best results,

$$0.05 \leq p \leq 0.20 .$$

This estimator does however require a large sample† size. Estimation of the threshold parameter is by no means a trivial matter, and much consideration needs to be taken when calculating an estimate.

Suppose that a set of *n* observations $x_1, x_2, ..., x_n$ are indeed drawn randomly from a LN(μ,σ^2,β) distribution. Crow & Shimizu (1988 : 122-123) mentions that Cohen (1951) has demonstrated that the local maximum likelihood estimator of β can be determined by solving the following equation for $\hat{\beta}$:

$$AB - n\sum_{i=1}^{n} \frac{\ln\left(x_i - \hat{\beta}\right)}{\left(x_i - \hat{\beta}\right)} = 0$$

where

$$A = \left[\sum_{i=1}^{n} \left(x_i - \hat{\beta}\right)^{-1}\right]$$

$$B = \left[\sum_{i=1}^{n} \ln\left(x_i - \hat{\beta}\right) - \sum_{i=1}^{n} \ln^2\left(x_i - \hat{\beta}\right) + \frac{1}{n}\left\{\sum_{i=1}^{n} \ln\left(x_i - \hat{\beta}\right)\right\}^2\right]$$

and $\hat{\beta} < x_{(1)}$ where $x_{(1)}$ is the smallest sample value.

Upon estimating the threshold parameter of a three-parameter lognormal distribution from a set of *n* observations, it is possible to estimate the mean and variance of the three-parameter lognormal distribution by simply applying Equations 3.4 and 3.5 to the values $x_1 - \hat{\beta}, x_2 - \hat{\beta}, ..., x_n - \hat{\beta}$ and remembering to

† Statistical definition.

add $\hat{\beta}$ to the average value afterwards. Sichel (1966) advocated the same approach when dealing with the t-estimate. According to Sichel (1966) the t-estimate of a three-parameter lognormal distribution is defined as:

$$t = e^{\bar{y}} \gamma_n \left(V_y \right)$$

where

$$x_i \sim \text{LN}(\mu, \sigma^2, \beta) \text{ and } y_i = \ln\left(x_i - \hat{\beta} \right) \quad i=1, 2, \ldots, n.$$

and

$$\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i$$

$$V_y = \frac{1}{n} \sum_{i=1}^{n} \left(y_i - \bar{y} \right)^2 .$$

The average of the three-parameter lognormal distribution is then estimated as

$$\hat{\tau} = \text{t} + \hat{\beta} .$$

Similarly, the central 100(1-p)% confidence interval for the true mean of a three-parameter lognormal distribution is simply

$$\left[t \times \Psi_{p/2}\left(V_y, n \right) + \hat{\beta}, t \times \Psi_{1-p/2}\left(V_y, n \right) + \hat{\beta} \right].$$

3.5 Grade-Tonnage Curves:

As the name suggests, a grade-tonnage curve is a graphical technique used extensively in the mining industry. In the author's opinion, however, this graphical technique can be equally effectively used outside of mining. In particular, the author believes that grade-tonnage curves can be successfully used in the analysis of the measured attribute in any geostatistical study. Much of Section 3.5 has been researched from David (1977 : 39-48, 302-308), Rendu (1981 : 49-53) and Clark & Harper (2000 : 45-49, 80-85).

In the mining industry it is generally uneconomical to mine the entire ore deposit. An economic cut-off grade is determined below which it is not profitable to extract the mineral. Thus the mining company would obviously like to estimate the proportion of the ore deposit graded above the cut-off, as well as the average value of the grade which can be profitably mined.

Note 3.3 For conciseness, only the grade-tonnage calculations related to proportion and mean value *above* the cut-off will be considered.

Given a probability density function $f(x)$, the proportion of observations above a value x_c is

$$T(x_c) = \int_{x_c}^{\infty} f(x)dx = 1 - F(x_c) \tag{3.8}$$

and

$$\mu_c = E(X \mid X \geq x_c) = \int_{x_c}^{\infty} \frac{xf(x)}{1 - F(x_c)} dx . \tag{3.9}$$

Hence, if X is:

- normally distributed

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma} \right)^2 \right) \qquad -\infty < x < \infty$$

$$T(x_c) = 1 - \Phi\left(\frac{x_c - \mu}{\sigma} \right) \tag{3.10}$$

and

$$\mu_c = \mu + \frac{\sigma}{\left(1-\Phi\left(\frac{x_c-\mu}{\sigma}\right)\right)\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x_c-\mu}{\sigma}\right)^2\right) \qquad (3.11)$$

where $\Phi(y)$ represents the cumulative distribution function of the standard normal distribution (see Appendix B), i.e.

$$\Phi(y) = \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-t^2}{2}\right) dt \, .$$

Tables of $\Phi(y)$ are widely available, and one taken from Hogg & Tanis (1997) has been included in Appendix A (Table A3).

- lognormally distributed (two-parameter)

$$f(x) = \begin{cases} \frac{1}{\sqrt{2\pi}\sigma x} \exp\left\{\frac{-(\ln x-\mu)^2}{2\sigma^2}\right\} & x > 0 \\ 0 & x \le 0 \end{cases}$$

$$T(x_c) = 1-\Phi\left(\frac{\ln(x_c)-\mu}{\sigma}\right) \qquad (3.12)$$

and

$$\mu_c = \frac{1-\Phi\left(\frac{1}{\sigma}\ln\left(\frac{x_c}{\tau}\right)-\frac{\sigma}{2}\right)}{1-\Phi\left(\frac{1}{\sigma}\ln\left(\frac{x_c}{\tau}\right)+\frac{\sigma}{2}\right)}\tau \qquad (3.13)$$

where τ represents the mean value of the lognormal distribution.

Note 3.4 The proof of Equation 3.11 is provided in Proof 3.1 on the following page. The proofs of Equations 3.12 and 3.13, however, are for conciseness purposes not shown. The interested reader may consult David (1977 : 41-43).

Proof 3.1

$$E(X \mid X \geq x_c) = \int_{x_c}^{\infty} \frac{x}{(1-F(x_c))\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) dx$$

$$= \frac{1}{\left(1-\Phi\left(\frac{x_c-\mu}{\sigma}\right)\right)\sqrt{2\pi}\sigma} \int_{x_c}^{\infty} x \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) dx \quad \ldots(1)$$

Let $\frac{x-\mu}{\sigma} = u \;\Rightarrow\; x = \sigma u + \mu$ and $du = (1/\sigma)dx.$

$$\text{Thus, } (1) = \frac{1}{\left(1-\Phi\left(\frac{x_c-\mu}{\sigma}\right)\right)\sqrt{2\pi}\sigma} \int_{\frac{x_c-\mu}{\sigma}}^{\infty} (\sigma u + \mu)\exp(-0.5u^2)\sigma\, du$$

$$= \frac{1}{\left(1-\Phi\left(\frac{x_c-\mu}{\sigma}\right)\right)\sqrt{2\pi}} \left[\int_{\frac{x_c-\mu}{\sigma}}^{\infty} \sigma u \exp(-0.5u^2) du + \int_{\frac{x_c-\mu}{\sigma}}^{\infty} \mu \exp(-0.5u^2) du \right] \quad \ldots(2)$$

Let $-0.5u^2 = t \;\Rightarrow\; dt = -udu.$

$$\text{Thus, } (2) = \frac{1}{\left(1-\Phi\left(\frac{x_c-\mu}{\sigma}\right)\right)\sqrt{2\pi}} \left[\int_{-0.5\left(\frac{x_c-\mu}{\sigma}\right)^2}^{-\infty} -\sigma \exp(t) dt + \int_{\frac{x_c-\mu}{\sigma}}^{\infty} \mu \exp(-0.5u^2) du \right]$$

$$= \frac{1}{\left(1-\Phi\left(\frac{x_c-\mu}{\sigma}\right)\right)\sqrt{2\pi}} \left[\sigma \exp\left(-0.5\left(\frac{x_c-\mu}{\sigma}\right)^2\right) + \mu\sqrt{2\pi}\left(1-\Phi\left(\frac{x_c-\mu}{\sigma}\right)\right) \right]$$

$$= \mu + \frac{\sigma}{\left(1-\Phi\left(\frac{x_c-\mu}{\sigma}\right)\right)\sqrt{2\pi}} \exp\left(-0.5\left(\frac{x_c-\mu}{\sigma}\right)^2\right).$$

If the grade distribution of the ore deposit has been estimated (see discussions in Section 3.2 and Section 3.3), the proportion of the deposit above cut-off grade and

the average grade of this proportion can be estimated using Equations 3.8 and 3.9 respectively. As the cut-off value is altered, so the proportion and mean of values above the cut-off also change. This relationship between cut-off grade, proportion of ore deposit above cut-off grade and mean grade value of proportion above cut-off grade can be displayed graphically as a grade-tonnage curve. The reason for the name is that the mean grade above cut-off and the proportion of tonnage of the ore deposit above cut-off are plotted together for a number of possible cut-off values. Again the author states his belief that this graphical technique can be used outside of mining. To illustrate, consider the following example:

Example 3.3

An area near an industrial site has been exposed to a certain chemical pollutant. A clean-up team has divided this area into a number of smaller regions and has sampled the pollution level in some of these regions. It is found that the pollution level in the regions follow a normal distribution with mean pollution level of 120 mg/m^2 (milligrams per meter squared) and standard deviation of 20 mg/m^2. Cleaning the area of pollution is a costly procedure, and it is decided that only the regions contaminated above a determined pollution cut-off level will be treated. A 'grade-tonnage' curve, or in this case a 'pollution-region' curve is constructed to determine the proportion of the polluted area that will be treated and the average pollution level of the regions that are contaminated above the determined pollution cut-off level. The results are attained by using Equations 3.10 and 3.11, and are shown in Table 3.1 and Figure 3.2.

Table 3.1 Table showing the proportion of the polluted area above the pollution cut-off level as well as the average pollution value of these regions, for a number of possible pollution cut-off values.

Pollution cut-off level (mg/m^2)	Proportion of regions above pollution cut-off level	Average value of the regions contaminated above pollution cut-off level (mg/m^2)
120	0.500	135.958
125	0.401	139.271
130	0.309	142.822
135	0.227	146.576
140	0.159	150.503
145	0.106	154.576
150	0.067	158.774
155	0.040	163.075
160	0.023	167.464
165	0.012	171.928
170	0.006	176.455
175	0.003	181.034
180	0.001	185.659

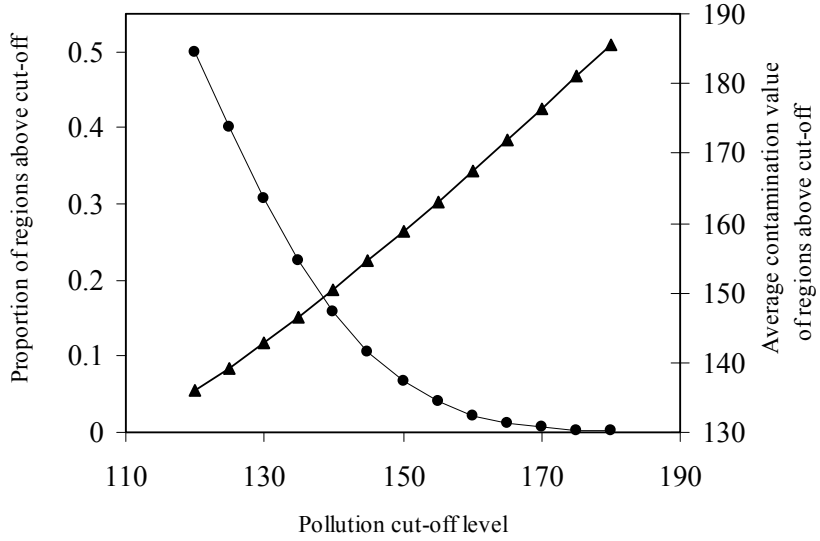

Figure 3.2 'Grade-tonnage' type curve displaying graphically the information contained in Table 3.1

It is found that pollution levels above 145 mg/m^2 are considered hazardous, and it is required that all regions above this pollution cut-off level are treated. From Table 3.1 and Figure 3.2 the clean-up team will be aware that just under 11% of the polluted area will require treatment, and that the regions being treated will on average have a contamination level of 154.576 mg/m^2.

As is pointed out in Rendu (1981 : 50), it is of utmost importance to realize that the probability density function *f(x)* will differ as the attribute is measured over different sized supports (see Section 2.4.1 and Note 3.1). The 'grade-tonnage' relationship (i.e. the relationship between proportion above cut-off and average value of this proportion) will thus also be affected by a change of support, and it is therefore crucial to identify the correct probability density function before calculating any grade-tonnage curve. As an illustration, consider the following example which has been taken and adapted from David (1977 : 46-48).

Example 3.4

Assume an iron ore deposit of 100 Mt (mega-tons) displays a normal distribution of grades and averages 54%. The standard deviation of samples is found to be 7%. It is decided that all ore graded above 60% should be sold to Company A.

According to the sample distribution and Equation 3.10, the total amount of ore that will be sold to Company A is:

$$T = 100\left[1 - \Phi\left(\frac{60-54}{7}\right)\right] = 19.5 \text{ Mt}$$

and from Equation 3.11 this ore will average:

$$\mu_c = 54 + \frac{7}{\left(1 - \Phi\left(\frac{60-54}{7}\right)\right)\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{60-54}{7}\right)^2\right) = 63.88\,\%$$

However, the ore will not be mined on blocks the size of the samples. The ore will be mined on much larger blocks which naturally possess a smaller standard deviation, say 4%.

So from Equation 3.10, the actual amount of ore that will be sold to Company A is:

$$T = 100\left[1 - \Phi\left(\frac{60-54}{4}\right)\right] = 6.7 \text{ Mt}$$

and from Equation 3.11 this ore will average:

$$\mu_c = 54 + \frac{4}{\left(1 - \Phi\left(\frac{60-54}{4}\right)\right)\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{60-54}{4}\right)^2\right) = 61.75\,\%$$

Note 3.5 It is noticed in Example 3.4 that mining the smaller support samples would have resulted in a better tonnage recovery and average grade than mining the larger mining blocks. Consequently, it would seem that mining smaller blocks would achieve higher recoveries and higher grades.

However, the mining techniques and mining equipment used by the company will have a large say on the minimum size of mining blocks. The extra cost involved in mining smaller blocks may therefore not justify the additional recoveries and higher grade that could be achieved.

Another important aspect is that the 'grade-tonnage' calculations automatically assume that it is possible to accurately distinguish between values above and below the cut-off. In general, the attribute values are unknown and need to be estimated (by, for instance, geostatistics), and it can naturally occur that values

truly below the cut-off are estimated as being above the cut-off value, and vice versa.

In conclusion, the grade-tonnage curve is a simple graphical technique that can undoubtedly add significant value to many geostatistical studies, but the following shortcomings, as pointed out by David (1977 : 308), should not be ignored.

- The grade-tonnage curve informs on what proportion above cut-off exists, but not where in space this proportion can be found.
- Grade-tonnage calculations assume that it is possible to make out the real population values when deciding if the population unit is valued above or below the cut-off.
- Often the block value distributions (i.e. larger support distributions) are not so straightforward to calculate from the sample distribution.