JOURNAL OF THE CHEMICAL METALLURGICAL & MINING SOCIETY OF SOUTH AFRICA

The Society, as a body, is not responsible for the statements and opinions advanced in any of its Publications
Reproduction from this Journal is allowed only with full acknowledgment of the source

Vol. 52 No. 6 DECEMBER 1951 Price 6/-

A STATISTICAL APPROACH TO SOME BASIC MINE VALUATION PROBLEMS ON THE WITWATERSRAND

By D. G. KRIGE, M.Sc. (Eng.) (Rand)

SUMMARY

Certain fundamental concepts in the application of statistics to mine valuation on the Witwatersrand are discussed, and general conclusions are drawn regarding the application of the lognormal curve to the frequency distribution of gold values.

An indication is given of the reliability of present valuation methods on the Rand. It is shown that the existing over- and under-valuation of blocks of ore listed as high-grade and low-grade, respectively, can be explained statistically. Suggestions are made for the elimination of such errors and for the improvement of the general standard of mine valuation by the use of statistical theory.

I. INTRODUCTION

The estimation of the tonnage and grade of payable ore in a mine and the correct policy of selective mining based on such estimates, is of vital importance to the mining engineer. It is surprising, therefore, that more attention has not been devoted on the Witwatersrand to the scientific improvement of mine valuation methods. At present these methods consist almost entirely of the application of simple arithmetic and empirical formulae guided by practical experience, and ignore the many advantages to be gained from a careful statistical analysis of the behaviour of the gold values.

The science of statistics has expanded rapidly during the last two decades and its value as a powerful and indispensable tool is now recognized not only by research workers and scientists but also, ever-increasingly, by the commercial and industrial world. This being the case it is noteworthy that in a mining field such as the Rand with its highly developed and advanced mining methods, singularly little attention has been paid to the analysis of mine valuation problems on a modern statistical basis. This omission is even more striking when cognizance is taken of the wealth of sampling data concerning the gold ore which is available and of the far-reaching decisions and deductions constantly being based on such data. Some contributions have however been made from time to time towards the application of statistics to mine valuation on the Rand (References Nos. 7, 8, 9, 11 and 12), but none of these have been generally applied in practice and a systematic practical approach based on clearly defined fundamental concepts still appears to be lacking.

The object of this paper is, therefore, to define these statistical concepts and to indicate briefly the lines along which statistical methods can be applied profitably in solving some of the existing problems and in improving the general standard of mine valuation on the Rand. For this purpose a knowledge of mathematical statistics is essential, but as this paper is primarily intended to provide the mine valuator with a practical introduction to statistical application in his field, the mathematics is confined to a short annexure, together with references from which details can be obtained.

II. DEFINITION OF CERTAIN FUNDAMENTAL CONCEPTS

The intelligent observer has no doubt often been amazed at the regularity and order behind what at first glance appears to be a chaotic variation in the attributes of an object, event or condition. The individual heights of the people forming the population of a town, for example, appear from a casual investigation to vary haphazardly, and yet when such height measurements are grouped according to the frequency of occurrence of individual sizes, a surprisingly uniform and regular trend in such frequencies will be found. Thus, intelligent observation and analysis will generally disclose a regular pattern behind the apparent chaos.

Even an experienced mine valuator on the Rand may believe that the variation between gold values along a stretch of drive, raise or stope face is haphazard. This is not the case, however, and it follows naturally that the observation of the regular pattern followed by such values, and the correct interpretation thereof, must open up new avenues of approach to the benefit of mine valuation in general.

It is as well to stress at this stage that the basic problem of mine valuation is that the actual gold content of a block of ore to be stoped is unknown and that it can only be estimated from the limited number of values available round its periphery, the orthodox estimate being based on the arithmetic mean of such a set of available values, i.e. the mean of such values is accepted as being the indicated mean value of the whole block of ore. The two main objects of a statistical approach to mine valuation are firstly, to determine the reliability of such existing methods of estimation and secondly, to develop methods which will on average, yield closer and more reliable estimates of the actual mean value of the ore from the limited available sampling information.

Before this can be done, however, the following fundamental statistical terms and their application to mine valuation on the South African gold fields have to be defined.

Population

The common concept of a 'population' is that of a large group of persons, each 'member' of the population being identified by his or her own particular attributes such as height, weight, age, wealth, etc. In the statistical sense, however, the measurements of any one attribute of the individual persons in such a group constitute a population (of measurements) and each such measurement is regarded as a member of the population.

In the case of a gold mine, the ore body can be regarded as a single ore parcel which can be subdivided into a large number of small parcels of ore, each of these smaller parcels having its own attributes, the vital one naturally being its gold content. The aim in framing the ideal policy of selective mining is to select for stoping purposes only those parcels of ore which contain sufficient gold to pay for all expenditure incurred up to and including the extraction of this gold and to leave intact all parcels with an insufficient gold content to cover such costs. In practice, except in the case of unusually wide auriferous reef bodies, this process of selection is effected in respect of reef 'parcels' which in each case occupy the entire width of the reef body (or economic band of reef), and the 'payable' and 'unpayable' parcels can consequently be depicted on the plane of the reef by the areas covered by these parcels. For practical purposes, therefore, a reef body in a particular mine can be regarded as a large 'area' of reef consisting of smaller individual reef 'areas,' each 'area' being identified in particular by the gold content of the ore parcel (or volume or tonnage of ore) it represents. The gold contents of such individual small reef 'areas' within a large reef 'area' can from a statistical angle, consequently be regarded as the members of a population.

The smallest 'area' of reef, the gold content of which is measured in practice, is that represented by the cross sectional area of the standard size channel cut in the process of sampling across the width of the reef body at a sampling section and, on average, measures approximately 6 sq in. For mine valuation purposes, therefore, the measurements of the gold contents of all the standard size (6 sq in) reef 'areas' which constitute a larger reef 'area' will be regarded as a population, and every such individual measurement will be a member of the population. The basic population is comprised of the *actual* gold contents of these 'areas' but these can in practice only be *measured* by underground sampling and hence the observed population consists of a number of *measurements* of the actual gold contents concerned.

In the practical case of a block of ore measuring, say, 200 ft by 200 ft, which has been sampled at 5 ft intervals round its periphery, the measured gold contents of the 160 standard size reef 'areas' at the corresponding number of sample sections will constitute the only known members of the population of measurements of the gold contents of the odd million standard size reef 'areas' constituting the entire block.

The gold content of any one such standard size reef 'area' (6 sq in) will be measured by the assayed gold content of the sample(s) obtained from the channel cut at the corresponding sampling section, i.e. by the (weighted average) dwt/ton of the sample(s) × tonnage of sample(s). Now, since the tonnage of the sample(s) is directly proportional to the volume of the sample(s), and the volume is in turn directly proportional to the overall sampled width (when the cross-sectional area of the channel cut for every sample is identical), it follows that the measurement of the gold content of a standard size reef 'area' is directly proportional to the average dwt/ton over the sampled width × the sampled width = total inch-dwt for the sample section concerned.

The inch-dwt of a sample section can, therefore, be accepted as a measurement of the gold content of a standard size reef 'area' (6 sq. in.) corresponding to this sampling section.

Where the reef width is relatively narrow the stoping width is determined entirely by practical mining considerations and is fairly constant. In such a case the inch-dwt value of a sampling section divided by the more or less constant factor of the stoping width so as to yield the dwt/ton value over the stoping width will also, therefore, provide a measurement of the gold content of the relevant standard size reef area.

Similarly, in the case of a wide variable reef width having a definite influence on the stoping width, but where neither of these widths appears on average to be related to the corresponding inch-dwt values (i.e. where the full range of stoping width variations is likely to be associated with every category of inch-dwt values), the dwt/ton value over the stoping width at a sampling section will on average also provide a measure of the gold content of the corresponding standard size reef area.

In the unusual case where there appears to be a definite relationship between the stoping widths and corresponding inch-dwt values at the various sample sections, the problem is more complicated and will not be considered in this paper.

From a practical point of view, therefore, the statistical use of either the inch-dwt value or the dwt/ton value over the stoping width at a sampling section can be justified and should yield the same eventual answer if a sufficient number of values is available, since the average dwt/ton value for the tonnage of ore in a block is the quotient of the average inch-dwt value and the average stoping width.

For the purpose of this paper the inch-dwt measure will be used almost invariably and a population will therefore be considered as being comprised of a number of inch-dwt values of sample sections corresponding to 'standard' size reef areas. In the case of a block of ore, for example, the population will consist of all the theoretically possible inch-dwt values which could be obtained if the block were to be extracted by a process of continuous sampling.

Similarly, the sample values obtained from a stretch of drive, raise or stope face can be considered to be equivalent to that obtained from a relatively narrow and elongated 'area' of reef containing a population of sample section values.

A case where the area concept is departed from is in the analysis of the distribution of calculated ore reserve values. In this case the population in effect comprises the indicated mean values of a number of blocks (i.e. 'areas') of ore. In order, however, to allow for the fact that in practice the areas of these blocks are usually very divergent in size, the tonnages of ore in the various value categories provide a better relative frequency measure. The population will therefore in this case consist of all the individual tons of ore in the ore reserves each at the indicated average value of the ore block of which it forms part.

Sampling from a population

In the statistical sense 'sampling' implies the selection of a limited number of members of a population, the group of selected members constituting the so-called 'sample.' To the mine valuator, 'sampling' implies the physical act of chiselling out a few pounds of reef (and waste) material for assay purposes, and 'samples' imply the separate packages of reef (and waste) material obtained in 'sampling.' It is, therefore, obvious that in the application of statistics to mine valuation a clear distinction is required between the above dual meanings of both 'sampling' and 'sample.' Bearing in mind the stated intention of this paper, the valuation interpretation of these two terms will be maintained and the corresponding statistical terms will be referred to in the following manner :—

Statistical term	*Valuation equivalent*
'Sample' ...	A set of sample values drawn from a population of sample values.
'Sampling' ...	The act of drawing a set of sample values from a population of such values.

The term sample, therefore, unless qualified, is used in the mine valuation sense and an individual sample value will be the inch-dwt value at a sampling section, i.e. a member of a population of individual sample values.

Random and systematic sampling

A considerable part of statistical theory has been built up round the basic concept of 'random' sampling (statistical sense), i.e. the concept of the drawing of a set of sample values (valuation sense) from a population of such values in a purely random and unbiassed manner. Briefly this means that for every selection of a single member every individual member of the population must have an equal chance of selection.

In practice, samples can only be taken round the periphery of an ore block and common sense has dictated the spacing of these at equal intervals. The theoretical sample values in the interior of the block, therefore, have no chance of selection, and the peripheral values are selected not at random but systematically. The exact relationship between the results of such systematic sampling and the theoretical random procedure is at present being investigated. It appears, however, from practical tests on mine A[5] that where the selection of the locations of the drives and raises bounding the blocks of ore has not been influenced in any way by sampling values previously known, the systematic perimeter samples may be accepted on average as representative of the blocks as a whole. The author is, however, fully aware of the limitations of this statement, particularly in cases where value trends are evident and where individual blocks are irregular in shape.

Frequency histogram

The statistical analysis of a population of values consists primarily of the segregation of such values into a range of selected value categories. The population is then represented graphically by plotting the limits of the range of values within each value category as abscissae and on each such range of values as base, a rectangle with area in direct proportion to the frequency of occurrence of the values in the value category concerned. The resultant step diagram is called a frequency histogram, and where the value ranges are made sufficiently small, this step diagram will in the limiting case, merge into a smooth curve called a frequency curve.

III. THE LOGNORMAL FREQUENCY DISTRIBUTION AS APPLIED TO GOLD VALUES

The shape of the lognormal curve

The fact that the gold values obtained in sampling a reef area could be represented by a frequency histogram of definite shape was known as far back as 1919[12] but it was not until recent years that the type of frequency curve which could be fitted satisfactorily to such a histogram was recognized as the lognormal frequency curve.[7, 8, 9] A typical lognormal curve is illustrated on Diagram No. 1. For this purpose 28,334 inch-dwt values representing the development results for a large section of Mine A were analyzed. The observed percentage frequency of occurrence of the values falling into the various value categories are represented by the areas of the rectangles forming the step diagram to which the lognormal curve was fitted, e.g., 12 per cent of the total number of values were found to lie in the category between 150 and 200 inch-dwt. The most distinctive features of the lognormal curve are its marked skewness to the left and the long drawn-out tail towards the higher value categories.

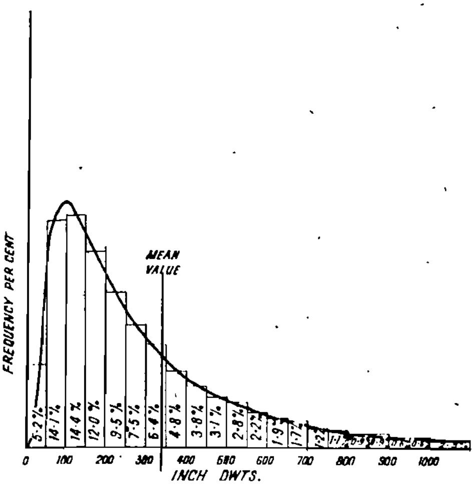

Diagram No. 1—Showing frequency distribution of 28,334 inch-dwt values on Mine A by means of a histogram and the lognormal curve fitted to these observations

As suggested by the name 'lognormal,' this curve is related to the well-known symmetrical 'normal' curve of error, and can be transformed into the latter by plotting the abscissae, i.e. the inch-dwt values on a logarithmic scale.

The application of the lognormal curve in various fields

The lognormal frequency curve is not peculiar to the distribution of gold values and has been found to be applicable in a large number of widely different fields, as the following brief list will indicate :—

The incomes of individuals in a nation.[9]

The sizes of grains in samples from sedimentary deposits.[6]

The sizes of sandgrains in samples from windblown sand.[1]

The sizes of particles of silver in a photographic emulsion.[3]

Sensitiveness of animals of same species to drugs.[3]

Numbers of plankton caught in different hauls with a net.[3]

Amounts of electricity used in medium class homes in the U.S.A.[3]

Reaction times of human beings in a word test.[3]

Diameters of particles of airborne dust in coal mines.[3]

Number of words in sentences from works of G. B. Shaw.[3]

As far as the Witwatersrand goldfield and its extensions are concerned, evidence from a number of the chain of mines stretching for more than a hundred miles from Heidelberg in the east to the West-Wits line in the west, as well as from mines in the Klerksdorp sector, indicates that it is highly probable that the lognormal curve will find application to the distributions of gold values in most, if not all, of the known economic reef horizons. An analysis of the borehole values for the Basal Reef in the Orange Free State field confirm the natural expectation that the pattern exhibited by these values is also lognormal in trend.

Characteristics of the lognormal curve

A brief list of the more important formulae based on a mathematical analysis of the lognormal distribution is contained in the annexure to this paper with the relevant references. For the stated purpose of this paper it is sufficient to mention only two basic characteristics of the lognormal distribution, i.e. the mean of the population and the relative variation (co-efficient of variation) between the values constituting the distribution. Referring to Diagram No. 1 the mean lies at the inch-dwt value corresponding to the centre of gravity of the area between the curve and the x-axis, and the relative variation is measured by the relative spread of the curve. All the other characteristics of the distribution, e.g. skewness, kurtosis (peakedness), modal value, median value and geometric mean, can be expressed in terms of either or both of the above two basic characteristics.

A lognormal distribution is, therefore, completely determined by its mean value and the relative variation between the units of the population.

Effect of a variation in the size of the reef area

The question naturally arises whether the gold values in a reef body are distributed lognormally, irrespective of the size of the reef area concerned. The selection of the boundaries of the section of Mine A, for example, in respect of which the lognormal distribution shown on Diagram No. 1 was observed, was purely arbitrary and it is, therefore, natural to expect the distributions of gold values within smaller (or larger) sections of such a mine also to be lognormal. This has been confirmed on Mine A for various sizes of reef areas down to a size smaller than that of the average ore reserve block on the mine.[5]

Now, in dealing with the incomes of the members of a human population it is only natural to expect a larger relative variation between the incomes of members of a town's population than between those of members of the population in a single suburb of that town. Similarly, it can be expected that gold values in a whole mine will be subject to a larger relative variation than those in a portion of the mine, and hence that the relative variation between gold values on any mine will tend to increase with an increase in the size of the reef area concerned. Practical experiments on Mine A have confirmed this.[5]

A conclusion arrived at as a result of these practical investigations, and which may be of considerable importance in future in improving the general standard of mine valuation, was that the relative variation between values within equal size reef areas appeared to be fairly stable and possibly constant (at least for practical purposes) in individual sections of a mine with minor changes from section to section. This aspect will be referred to again in Section VI.

The effect of a variation in the size of the ore samples

Since the decision to select samples corresponding to reef areas of 6 sq.in. is also arbitrary, the size of the sample should also not affect the typical lognormal pattern of the distribution of gold values. This is confirmed by the fact that in an ideal case (where the natural distribution of the ore values has not been upset by previous mining operations), the average values of ore blocks (pay and unpay) in a mine have also been found to be distributed lognormally, an ore reserve block being equivalent to an exceptionally large ore sample.

It is found in practice, therefore, that the inch-dwt values within a block of ore are distributed lognormally, that these values from all the blocks in the mine when considered together still form a lognormal distribution with a curve of a somewhat different shape, and that if the mean values of the individual blocks are considered, these in turn also constitute a similar type of distribution. The mathematical relationship between the relative shapes of these three types of distributions is formulated in the annexure (Formula 19) and can be applied in practice in predicting the percentage payability and average payable ore reserve value for a new mine from relatively little basic information such as borehole values and a limited number of underground development values. A detailed discussion of this aspect falls outside the scope of this paper and could possibly be dealt with at a later date.

General basic conclusions to be drawn from the lognormal distribution of gold values

It is immediately evident from the illustration of a typical lognormal curve (Diagram No. 1) that the inch-dwt values cover the entire theoretical range from zero to infinity. Also since the curve approaches the x-axis asymptotically in the range of the higher value categories, the frequency of occurrence of extremely high values is relatively small but can only become zero for infinitely large values.

In drawing a set of values at random from a population of values, the probability of drawing a value in any particular value category is measured by the relative frequency of occurrence of values in this category, e.g. referring to Diagram No. 1, it is obvious that if one value is drawn at random from the 28,334 values represented by the curve, the probability of drawing a value between, say, 150 and 200 inch-dwt is 12·0 per cent, i.e. approximately 1 in 8.

It is also evident that every lognormal distribution, no matter what its mean value may be, must comprise a mixture of values ranging theoretically from zero to infinity.* A distribution with a low average value will, therefore, always contain a proportion (even if small) of relatively high values, and *vice versa* a distribution with a high average value will contain a proportion of relatively low values. In mine valuation, therefore, *the occurrence of relatively high values (even if only occasionally) in a low-grade block of ore is quite natural,* and similarly also the occurrence of low values in a high-grade block of ore.

Considering now the practical aspect of, say, a block of ore or a stope face, in respect of which only a limited number of all the possible values is available, it is evident that the possible combinations of, say, ten sample values each, which can be drawn from the complete distribution formed by all the sample values, will be infinite and, therefore, that the probability of striking two identical sets of ten values each is slight. Further, in view of the wide range of values covered by the parent population, the striking of a set in which all ten values are identical, is virtually impossible.

A further basic conclusion to be drawn from the knowledge of the lognormal frequency distribution of gold values is that the individual sample values available in respect of a block of ore or a stope face represent only a few known values out of a virtually infinite number of values which can be obtained by repeated sampling. Where the few known sample values are distributed over the range of values in approximately the same proportions as the total number of possible sample values, the mean value of these few samples will naturally correspond closely to the true mean value of all the possible samples. In practice, however, some of the relatively few extremely high values in the parent population of values, must at one time or another be struck in taking a set of samples, and will in such an event appear to be out of accord with the rest of the sample values in the set, and will raise the average value of the set to an abnormally high figure. Such values are generally regarded as 'anomalous,' 'freak,' or the result of bad sampling, and are in practice usually 'cut' or 'adjusted' by arbitrary methods in order to yield what, at any rate, appears by intuition to be a more reliable average result. Such apparently anomalous values are, however, genuine members of the population of values along the stope face or in the block of ore, and are, therefore, in no sense truly anomalous or freak. The correct approach to the problem of estimating the true mean value of the unknown population of values (i.e. of the stope face or ore block) from the few known members of this population (i.e. from the few available sample values) is, therefore, to fill in the gaps between these known values in such a way as to result in the best estimate of the parent distribution of values, i.e. the population, without discarding or 'cutting' any one value which may appear to be anomalous. This is one of the basic aims in approaching the problems of mine valuation from a statistical angle.

It is also evident that since even adjoining sample values cannot be expected to be identical, a fact which has in a practical way often been observed from the results of check sampling in the same or an adjacent groove, *any sample value cannot be regarded as having a so-called 'area or distance of influence,' in the generally accepted sense.* An occasional high value encountered in sampling successive stope faces in a low-grade block of ore is, therefore, not

* In practice the maximum gold value possible will be that corresponding to pure gold, i.e. some 583,000 dwt/ton, or say, 29 million inch-dwt for a 50 in. stoping width.

necessarily indicative of a patch of high-grade ore (extending half way from the relevant sample section to surrounding sampling sections), and the same principle naturally applies to borehole values.

IV. THE RELIABILITY OF INDIVIDUAL FACE AND BLOCK VALUATIONS

As stressed at the outset, the individual sample values as computed from assay results have been employed throughout, and no attempt has been made to determine the bias errors which may be introduced in the actual physical acts of sampling and assaying. A useful paper on this aspect was published by Sichel[8] in 1947. Reference to the reliability of block and face values in this investigation, therefore, merely implies the reliability of accepting the mean of a limited number of sample values as an estimate of the *correct average sample value* of the relevant block or face. The criterion for measuring reliability will, consequently, be the average value of a theoretically infinite number of sample sections in the relevant block or stope face. This average value has been, and will be, referred to as the *true mean value* of the block and will only equal the *actual mean value* of the block where on average no over- or under-sampling is carried out.

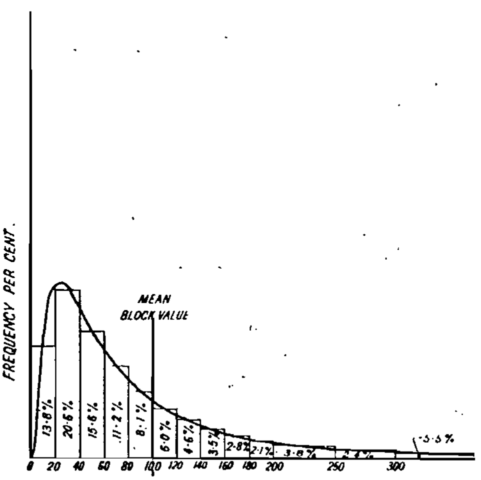

Diagram No. 2—Showing the frequency distribution of inch-dwt values within an average ore reserve block on the Witwatersrand

The frequency distributions of gold values in individual ore reserve blocks vary in characteristics from mine to mine and reef to reef and even within mines from section to section, but if the individual values are expressed as percentages of the true mean block value, Diagram No. 2 can serve as an indication of what conditions might be on average.* The curve on this diagram, therefore, approximately represents the distribution of individual gold values in an average ore reserve block on the Witwatersrand, the position being possibly somewhat different in any particular mine or section of a mine.

If now, one sample is taken at random from this average block of ore (or stope face), the probability of drawing a value falling within any specific value category will be represented by the percentage of values in this category, e.g. the chances of obtaining a value falling within the range of, say, 80 per cent to 100 per cent of the true mean value of the block, i.e. a value corresponding to a negative error of between 0 and 20 per cent will be 8·1 per cent, i.e., say, 1 in 12.

Similarly, the probability of an error not exceeding plus or minus 40 per cent of the true mean value is represented by the percentages in the categories from 60 per cent to 140 per cent of the mean value, i.e. 29·9 per cent. There will, therefore, only be a chance of about three in ten of not exceeding an error of 40 per cent. In other words, in an average of seven cases out of every ten the value obtained will be in error by more than 40 per cent of the true mean value of the block of ore. As could have been anticipated, one sample value as an estimate of the true block value will serve little practical purpose.

Consider now the case where, say, ten values are drawn at random from a parent population of values and where the arithmetic mean of such a set of ten values is accepted as an estimate of the true mean value of the population. If this process is repeated many times, it will result in a series of estimates of this true mean value and such estimates will in turn yield a distribution of estimates which can, as in the above case of individual values, be

* Parameter σ^2 of distribution taken as 1·02—see Annexure.

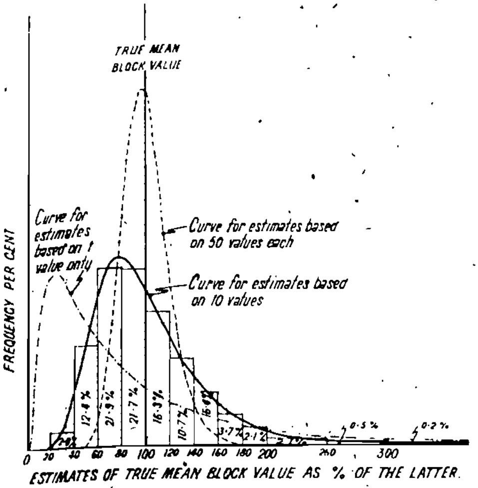

Diagram No. 3—Showing the frequency distributions of estimates of the true mean value of an average ore reserve block based on 1, 10 and 50 values per estimate respectively

represented by a smooth frequency curve. Statistical theory provides the method of determining the distribution curve of the estimates (to a first approximation in the case of a lognormal parent distribution).*

In the case of the so-called average ore reserve block considered above, the distribution of the estimates of the true mean value of the block if based on ten values in each case only, can be represented by the full curve on Diagram No. 3.† The curve for individual values (from Diagram No. 2) is shown dotted and it is immediately evident how the estimates based on sets of ten values each are clustered more closely on either side of the overall true mean block value.

From Diagram No. 3 it is evident that if a large number of sets of ten samples each are taken from this particular block of ore, the chances are that, e.g. 70·6 per cent of all the observed mean values of these sets of ten sample values each will be within the range of values from 60 per cent to 140 per cent of the true mean value of the block. There is therefore approximately a seven in ten chance of not exceeding an error of 40 per cent if the arithmetic mean of ten sample values is accepted as the true mean value of the block.

Similarly, the distribution curve for estimates based on any specific number of available sample values can be obtained, and Fig. 3 shows the curve (dotted) for sets of 50 values each. Based on a series of such curves, Table 1 was prepared to indicate the overall reliability of block valuation in the case of an average ore reserve block on the Rand.

It becomes clear from this table what 'chances' are taken by the valuator in the valuation of individual blocks and stope faces where conditions approach the

* See Annexure and Ref. 5.

† Based on random sampling theory. For systematic sampling as practised on the Witwatersrand the curve will probably be slightly different, an aspect receiving attention at present.

TABLE 1

No. of samples per set	Maximum error either side of true mean value of block				
	10 per cent	20 per cent	30 per cent	40 per cent	50 per cent
	Probability percent				
1	7	14	22	30	39
5	14	28	42	56	68
10	19	38	55	71	82
20	27	51	71	85	92
50	41	72	90	97	99
100	55	87	97	99	100
200	71	97	100	100	100
500	91	100	100	100	100

E.g., there will be a 51 per cent probability, say, an even chance, of the observed mean of 20 sample values not being in error by more than 20 per cent either side of the true mean value.

average. In order to ensure, say, that the observed mean of the sample values for a block or face will not be in error by more than 10 per cent in an average of nine cases out of every ten (i.e. a 90 per cent probability or confidence limit), a total of some 500 samples will be necessary. This is far in excess of the usual number of sample sections available for an average ore reserve block and appears nearly astronomical when compared with the number of stope sampling sections per face, usually ranging from 10 to 30.

It is therefore obvious that the customary practice of starting and stopping stope faces, or portions of stope faces, on the evidence of one, or even two or three, stope samplings, must inevitably result in the stoping of some unpay ore and in the rejection of a percentage of pay ore.

This statement will be more evident from Diagram No. 4 which is based on the same fundamental assumptions and on which the position for the average ore reserve block (or stope face) is presented in relation to the pay limit and to a probability of 95 per cent, i.e. 19 in 20.

If, therefore, a wrong decision in not more than an average of one case out of every 20 is regarded as a reasonable risk to take in block valuation, it is evident from the diagram that a block valuation based on, say, 50 sample sections will have to be at least 1·336 times the pay limit

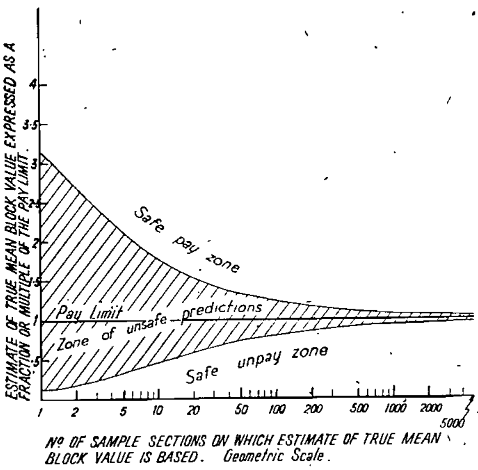

Diagram No. 4—Showing safe limits for selecting ore as payable or unpayable where 'safe' implies an incorrect classification on average in not more than one case out of 20. Based on approximately average conditions on the Witwatersrand

before the block can safely be classed as payable, or will have to be at most 0·723 times the pay limit for the block to be classed safely as unpayable. No safe classification can be made if the arithmetic mean of the 50 observed values lies anywhere between these two limits.

Where the pay limit is, say, 3 dwt/ton, a block valuation based on 50 sample sections will therefore have to be less than 2·17 dwt/ton or more than 4·01 dwt/ton for the block to be classed safely as unpayable or payable, respectively. Similarly, a face valuation based on, say, ten values, will be indecisive anywhere between 1·42 dwt/ton and 5·38 dwt/ton.

It is evident from the above analysis which it must be stressed applies only under the stated average conditions, that a reasonably reliable estimate of the value of a block of ore or of a stope face cannot be based on a limited number of values such as are generally available from a single stope sampling or even from two or three such samplings.

It seems, therefore, that the present practice of basing selective stoping policy on current stope sampling results is open to justifiable criticism and requires careful reconsideration. It appears to the author that serious thought should be given to the curtailment and possibly even the discontinuation of stope sampling and the diversion of all efforts rather towards increasing the number of available development sampling sections per ore block in order to improve at any rate the general reliability of block valuations to within safe limits. Before this general but radical suggestion can be put into effect on any specific mine, it will naturally be essential to analyze very carefully local conditions, such as—

(*a*) the average variation between values in ore blocks;

(*b*) the extent of the change (if any) in this variation from block to block;

(*c*) shoot and general value trends (if any);

(*d*) the extent to which perimeter sampling can be accepted as representative; and

(*e*) the extent to which stope sampling is essential for purposes other than selective stoping policy.

V. BIAS ERRORS INTRODUCED IN ORE RESERVE VALUATIONS DUE TO THE NATURE OF THE DISTRIBUTION OF GOLD VALUES AND THE LIMITED NUMBER OF AVAILABLE VALUES PER ORE RESERVE BLOCK

The question which now naturally arises is what the overall effect will be of valuing all the blocks of ore in a mine on a limited number of sample values per ore block. For this purpose a practical example, which is a reasonable representation of the position on Mine C, will be considered.

The true values of the payable and unpayable ore blocks in this mine are distributed lognormally in the various grade categories as shown in Columns 1A and 1B of Table 2,* e.g. 7·10 per cent of the total number of ore blocks is of a true grade lying between 2·55 dwt/ton and 2·95 dwt/ton and has an average true value of 2·75 dwt/ton. The ore blocks are in practice valued on an average of only 60 sample section values per block, and as indicated in the previous subsection such estimates of the true block values are, therefore,

* Assumptions for Table 2 :
Parameter σ_b^2 of distribution of true block means $= 0{\cdot}184$.
Overall mean value $= 4{\cdot}9$ dwt/ton.
Parameter σ_i^2 of distribution of individual values within blocks $= 1{\cdot}183$
Distributions of means of sample sets based on random theory and assumption of lognormality.
Determinations made graphically and therefore not necessarily correct to the second decimal point.

TABLE 2

					Distribution of means of 60 samples per block								
		Col. No.	1A	1B	1	2	3	4	5	6	7	8	9
	Row No.		Value category limits—dwt		2·55	2·95	3·95	4·95	5·95	6·95	7·95	8·95	
	1A	Value category limits—dwt		% freq.	12·57	6·32	22·42	19·21	13·46	9·36	6·20	3·66	6·80
	1B		Per cent freq.	Mean values	2·05	2·76	3·45	4·43	5·42	6·41	7·41	8·41	11·20
Distribution of true mean values of all blocks	1	— 2·55 —	9·55	2·12	8·26	0·48	0·81	—	—	—	—	—	—
	2	— 2·95 —	7·10	2·75	2·88	2·02	2·06	0·14	—	—	—	—	—
	3	— 3·95 —	22·03	3·46	1·43	3·41	12·67	4·03	0·49	—	—	—	—
	4	— 4·95 —	20·74	4·43	—	0·41	5·81	9·44	3·94	0·97	0·17	—	—
	5	— 5·95 —	15·35	5·42	—	—	1·00	4·45	5·45	3·15	1·01	0·23	0·06
	6	— 6·95 —	10·06	6·41	—	—	0·07	0·99	2·67	3·12	2·06	0·80	0·35
	7	— 7·95 —	6·20	7·41	—	—	—	0·15	0·77	1·55	1·77	1·12	0·84
	8	— 8·95 —	3·69	8·41	—	—	—	0·01	0·14	0·51	0·92	0·89	1·22
	9		5·28	10·88	—	—	—	—	—	0·06	0·27	0·62	4·33
	10	Mean of true values in columns		4·90	2·42	3·19	3·70	4·57	5·40	6·17	6·97	7·89	9·73
	11	Theoretical Block Plan Factor		—	% 118	% 116	% 107	% 103	% 100	% 96	% 94	% 94	% 87
	12	Actual B.P.F.—Mine C		—	—	% 119	% 106	% 108	% 105	% 101	% 98	% 93	% 81

subject to considerable fluctuation or errors. For example, valuations of the blocks of ore falling in the second grade category (i.e. the 7·10 per cent) with a true mean value of 2·75 dwt/ton, will yield the following results (Row No. 2) :—

2·88 per cent will be valued at between	nil	and	2·55	dwt/ton
2·02 ,, ,, ,,	2·55	,,	2·95	,,
2·06 ,, ,, ,,	2·95	,,	3·95	,,
0·14 ,, ,, ,,	3·95	,,	4·95	,,
7·10 per cent will be valued at an average of ...		...		2·75 dwt/ton

Similarly, the block valuations of ore falling in the other true grade categories are reflected in the other rows up to No. 9 in the table. The overall effect of all the block valuations can now be analyzed as follows :—

Taking, for example, Column No. 2 which represents all the 6·32 per cent of blocks listed from block valuations as having values between 2·55 and 2·95 dwt/ton with an overall indicated mean value of 2·76 dwt/ton. Due to the fact that these blocks were each valued on 60 sample values only, these blocks actually comprise—

0·48 per cent with true mean values between	nil	and	2·55	dwt/ton
2·02 ,, ,, ,,	2·55	and	2·95	,,
3·41 ,, ,, ,,	2·95	,,	3·95	,,
0·41 ,, ,, ,,	3·95	,,	4·95	,,
6·32 per cent with a true mean value of ...		...	...	3·19 dwt/ton

The true mean value of these blocks valued at an average of 2·76 dwt/ton, is therefore 3·19 dwt/ton representing a considerable undervaluation.

In a similar manner the true values of the blocks falling in the other 'indicated' value categories can be determined and are shown in Row No. 10. The indicated average values of the blocks within these categories, as obtained from sampling results, will however be as shown in Row No. 1B. Comparing now the values in this row with those in Row No. 10, it is obvious that *block valuation based on a limited number of samples per block will result in the general under-valuation of blocks listed in the low-grade categories and over-valuation of blocks listed in the high-grade categories.*

Now, it is a well-known fact on most mines on the Witwatersrand that when ore reserve blocks are extracted, the results from blocks valued as high-grade are on average disappointing, whereas results from blocks valued as low-grade on average exceed expectations. The best practical measure of these phenomena is provided by the Block Plan Factors* observed in the various categories, those for Mine C being listed in Row No. 12 (Table 2).

The question naturally arises why current stope sampling can be accepted as a criterion for criticising ore reserve valuations in the various categories, but the explanation can be supplied readily. Take, for example, the blocks as valued in the category 2·55 to 2·95 dwt/ton (Column 2, Table 2), and in fact, comprising a range of blocks with true values in the categories from 0 to 4·95 dwt. The current face samplings in respect of these blocks (based on an even smaller number of samples per block than in the case of block valuations) will err on either side of the true values in the range 0 to 4·95 dwt, but the mean of *all* such samplings (particularly over a whole year of operation) will provide a reasonable

* Block Plan Factor is the ratio expressed as a percentage which the gold content of the ore broken from Ore Reserves as indicated by current stope sampling results bears to the content as computed from block valuations.

estimate of the true mean value of the combination of blocks under consideration.*

The theoretical Block Plan Factors in Row No. 11, Table 2 (i.e. theoretical true values in Row No. 10 divided by indicated values in Row 1B) can now be compared with the observed Block Plan Factors (in the various value categories) for Mine C (Row No. 12), where the conditions assumed in the table are roughly approximated. The agreement is sufficiently close to indicate that the contentions on which the calculation of the theoretical Block Factors was based can provide a reasoned solution to the problem of the variation in the Block Plan Factors as observed in the various value categories.

A direct and more accurate method of arriving at the theoretical Block Plan Factors is provided by the theory of log-normal correlation and regression (Formulae Nos. 22 and 23 in the annexure).

Referring now to Diagram No. 5, the curve *ab* represents the theoretically determined trend of relationship of true block values corresponding to the observed block values for Mine C. This relationship as observed in practice from the Block Plan Factors in the various grade categories is shown by the small circles, and it is evident that the agreement between theory and

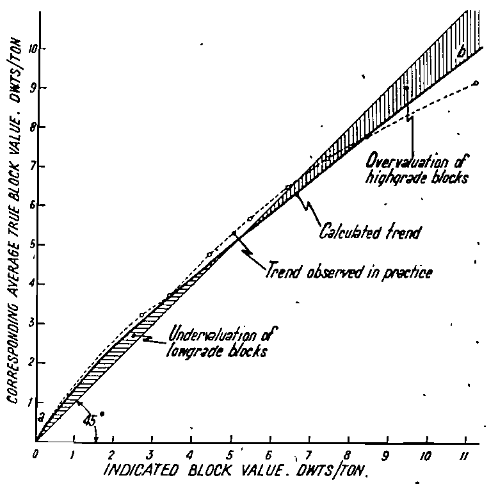

Diagram No. 5—Showing theoretical and observed trend of relationship between block valuations based on 60 values per block and the true block values for Mine C

* Provided the total number of samples taken from blocks in the category is sufficiently large.

practical results is close. This diagram clearly indicates the extent of under-valuation of blocks listed as low-grade and the over-valuation of blocks listed as high-grade.

It is obvious from the foregoing that, where the number of samples per block is limited, and no cutting of individual or block values is done, it is only natural and not due to wrong blocking policies, to expect Block Plan Factors exceeding 100 per cent in the low-grade ore categories and less than 100 per cent in the high-grade ore categories. The extent of the deviation either side will depend partly on the basic behaviour of the distributions of gold values in the mine and in individual blocks, but mainly on the average number of sample values available per block, and these deviations can only be expected to become insignificant as the number of sample values per block is increased considerably.

The average value above any pay limit.—Referring again to Table 2, and computing now the average indicated and true values above any pay limit, the interesting position reflected in Table 3 is found.

TABLE 3

Value limit = pay limit	Average indicated value above pay limit	Average true value above pay limit	Theoretical Block Plan Factor for all ore above pay limit
	Values in dwt/ton		Per cent
0	4·9	4·9	100
2·55	5·33	5·28	99·1
2·95	5·53	5·45	98·6
3·95	6·32	6·11	96·7
4·95	7·24	6·85	94·6
5·95	8·18	7·59	92·8
6·95	9·18	8·37	91·2
7·95	10·22	9·17	89·7
8·95	11·20	9·90	88·4

The overall theoretical Block Plan Factors for various pay limits are shown by the curve *AB* on Diagram No. 6 and it is clear that as the pay limit is raised the overall Block Plan Factor will decrease, i.e. the

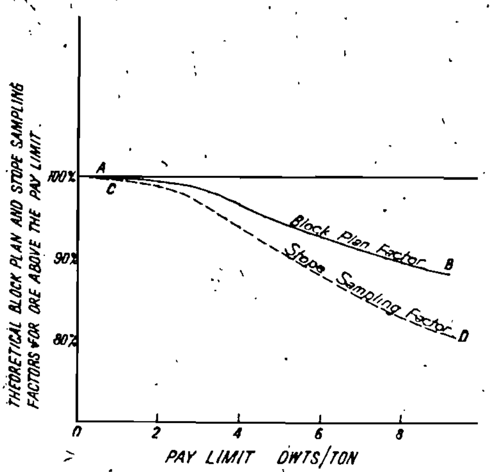

Diagram No. 6—Showing overall effect above various pay limits for Mine C

over-valuation of ore listed as payable will increase.

*The effect on the Mine Call Factor.**—The customary practice on the Rand in stoping to a pay limit is to base the policy of stopping and starting stope faces, i.e. of selecting the payable ore for stoping, largely on current stope sampling except, possibly, where an ore reserve block falls well within the payable category. In many cases portions of stope faces, which can from a mining point of view be stopped without upsetting the general layout, are stopped on the results of one or two, or at most, three unpay stope samplings, the total number of samples on which such a decision is based being as low, possibly, as 10 or 15 (e.g. a 100 ft length of face sampled at 20 ft intervals). The average number of stope samples on which any such decision is based is probably larger but will in nearly all cases be considerably smaller than the average number of samples per ore reserve block. If, therefore, a table similar to Table 2 is prepared to allow for the selection of tonnages in value categories on, say, 30 sampling sections per working face only, it is obvious that the variation of the indicated values in the rows will be larger, and that the under-valuation in the low-grade and over-valuation in the high-grade categories will be accentuated.

On this basis, a second edition of Table 3 was prepared and is reflected in Table 4.

From the above it is evident that in working to a pay limit and even where there is no physical over-sampling underground, *there will always be a consistent and natural tendency for the ore selected for stoping to be over-valued.* Further, in the case of a mine with a relatively low payability (i.e. a high pay limit in relation to the average value of all the ore in the mine) the percentage over-valuation in selecting tonnages of ore as payable on a limited number of sample values can be appreciable, and can thus account directly for a proportion of the difference between the gold called for by sampling and that accounted for by the reduction works. The effect on Mine C is shown graphically on Diagram No. 6 by the dotted line *CD*. The explanation of a Mine Call Factor of less than 100 per cent can therefore to some extent be sought in the limited number of sample values available

* Mine Call Factor is the ratio expressed as a percentage which the gold accounted for by the reduction works bears to the gold called for by current sampling results.

TABLE 4

Pay limit: dwt/ton	0	2·55	2·95	3·95	4·95	5·95	6·95	7·95	8·95
Average indicated value above pay limit: dwt/ton	4·9	5·42	5·68	6·49	7·41	8·39	9·41	10·45	11·49
Average true value above pay limit: dwt/ton	4·9	5·31	5·51	6·08	6·71	7·35	8·00	8·63	9·27
Stope sampling factor %	100	98	97	94	91	88	85	83	81

per working face and on which the selection of payable and unpayable ore is based.

Over a period the Mine Call Factor as observed on a mine can, therefore, in addition to the well-known contributing factors such as any actual physical over-sampling underground, and any gold lost in mining, also be ascribed to the invisible but inevitable over-valuation of the ore classed as payable on a limited number of values per stope face or ore block.

The major effect of this unsatisfactory over- and under-valuation of blocks in the upper and lower grade categories, as well as the undesirable overall over-valuation of all blocks above the pay limit, can of course be eliminated by applying the correctly calculated Block Factors in the various grade categories. This should also have a definite steadying effect on the Mine Call Factor on any mine, particularly as every change in the working costs of a mine and/or the price of gold, has an immediate effect on the pay limit. This in turn will have an effect on the Mine Call Factor unless the corrections, based on the suggestions above, are effected, in the grade categories. Such corrections will, however, only result in a corrected average valuation of all blocks in every grade category, and many individual blocks will, therefore, remain graded incorrectly.

A further significant observation is that, as in the case of the Block Plan Factors, the Mine Call Factors will vary from one grade category to another. The Mine Call Factor at the pay limit grade can, therefore, not be expected to be the same as that observed for the mine as a whole, and pay limit determinations based on the overall Mine Call Factor will consequently be subject to a bias error. In the case of the ore reserve pay limit, however, this bias error will be eliminated automatically if the correct Block Plan Factors for the various grade categories are applied first.

VI. IMPROVED ESTIMATES BASED ON STATISTICAL THEORY

There are various additional ways in which the unsatisfactory features present in existing mine valuation practice can be reduced to such an extent as to render the effects insignificant. With the present methods based on the straight arithmetic mean of sample values an obvious remedy is to increase where necessary the number of values on which any individual valuation of an ore block or stope face is made. In the case of a stope face, practical considerations exclude the possibility of obtaining the hundreds of sample sections required to reduce valuation errors to safe limits. In the case of ore blocks, however, and provided these are blocked out realistically so as to be of a reasonable size with as much of the periphery as possible exposed, it is quite practical to obtain the required number of sample sections. For most mines, satisfactory block valuation may necessitate at most the doubling or possibly the trebling of the development sampling programme, and this will not call for any additional samplers if, as suggested previously, the stope sampling programme, on which relatively little reliance for valuation purposes can be placed, is curtailed drastically. The policy to be advocated for any specific mine will naturally depend on local conditions as stressed in a previous section.

The application of statistical theory to our problems, however, opens up a new avenue of approach. Where the gold values follow the lognormal pattern it is possible to arrive at improved statistical estimates of the true values of ore blocks (or stope faces), i.e. calculations subject to reduced errors as compared with existing estimates. Thus the same basic sample values can be employed to greater advantage.

Two types of improved estimates have been developed and the choice between these will depend on local conditions. The first type will be dealt with in some detail, and applies where analysis of the distribution of gold values in a mine indicates that the relative variation between values in ore blocks (or stope faces) of equal size is sufficiently stable to allow it to be accepted as constant for valuation purposes throughout the relevant mine or section of that mine. Where this proves to be the case (and practical experiments carried out on Mine A[5] and recently also on Mine D, suggest that this warrants serious investigation on all mines) the improved estimation procedure is straightforward. The *geo-*

metric mean of the sample values is determined and multiplied by a factor related to the number of available values and the known relative variation between all the possible values in the ore block (or stope face), i.e. the parent population (Formulae 18 and 18*a* in annexure).

The improvement obtained by this method compared to the present method of accepting the arithmetic mean, can be measured in the most practical manner by the relative numbers of sample values required by these two methods to yield equally reliable estimates of the true value of any particular ore block. Allowing for the variation in conditions which have so far been observed for a number of Witwatersrand mines, it can be stated that for equivalent reliability the number of sample values required on the present orthodox method will be at least one and a half times, and may be as high or even exceed, three times the number required on the suggested improved method (see notes following Formula 18*a* in annexure).

Practical illustration.—The above conclusions were tested by taking a developed section of Mine B with 3,600 available development sampling values. For our purposes the true mean value of this section of the mine can be accepted as the arithmetic mean of these 3,600 values, i.e. 477 inch-dwt. The relative variation between these values was calculated and can also for practical purposes be accepted as equivalent to that between all the values of the parent population of values concerned. Based on this calculated figure the relevant factors by which the geometric mean of any set of sample values requires to be multiplied to yield the improved estimate of the population mean were determined.

Thirty sets of twenty equidistantly spaced sample values* each were now selected from these 3,600 samples, and the mean and geometric mean of every set was calculated. The geometric mean of every set was then multiplied by the required factor, i.e. 1·7514, to yield the 'improved' estimate of the true mean value (477 inch-dwt).

* These values were therefore not selected in a random fashion but systematically. For practical comparative purposes, however, this aspect can be disregarded at this stage.

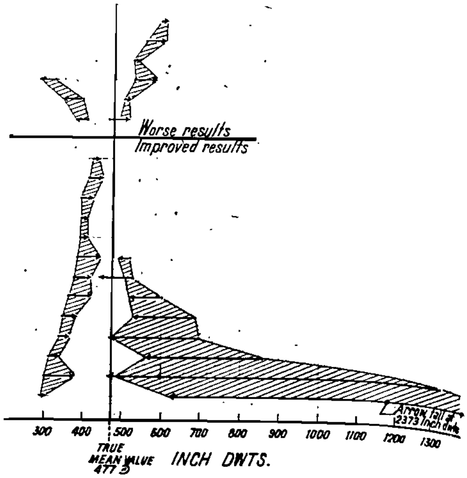

Diagram No. 7—Showing comparison between orthodox and corresponding improved estimates of true mean value of a reef area obtained from 30 sets of 20 samples each

Arrow tail = orthodox estimate
Arrow head = improved estimate

The 30 'improved' estimates were found on the whole to be clustered much more closely around the true mean value of 477 inch-dwt than the corresponding 30 straight means and 'improvements' were obtained in 21 out of the 30 cases.† The results for these 30 cases are shown graphically on Diagram No. 7, and the overall improvement obtained is evident.

The following is an example of how the need for 'cutting' a high value falls away :—

Set No. 13 :

Inch-dwt values : 35, 40, 66, 74, 81, 112, 146, 161, 202, 209, 244, 266, 290, 305, 319, 528, 1,004, 1,233, 2,341, 18,928.

Straight mean value : 1,329·2 inch-dwt.
'Improved' estimate : 466·7 inch-dwt.
True mean value : 477 inch-dwt.

Table 5 indicates the results of six sets of 100 samples each treated in the identical manner, and the overall advantage of the suggested improved method is obvious.

In the same way, one set of 600 samples gave an arithmetic mean value of 555·9 inch-dwt, and an 'improved' estimate

† Average extent of improvements in 21 cases = 189 inch-dwt. Average extent of worsening in 9 cases = 34 inch-dwt.

TABLE 5

Sample Set No.	1	2	3	4	5	6
Mean value of set—inch-dwt ...	403·3	454·2	445·9	832·6	517·2	682·4
'Improved' estimate of true mean value—inch-dwt	405·7	432·6	471·7	476·3	485·9	488·8

(True mean value = 477 inch-dwt.)

of 461·0 inch-dwt, the former being in error by +16·5 per cent and the latter by only −3·4 per cent.

Where detailed analysis on any mine shows that the relative variation between all the possible values in individual ore blocks (or stope faces) varies significantly from block to block (or face to face) the second type of improved estimation procedure can be employed. This procedure has been analyzed in detail by Sichel (1949 and 1951), and involves the use of formula (20). The practical application of this procedure will be facilitated by the tables to be published in Ref. 10. Although this procedure will yield improvements of a somewhat lesser extent than those discussed above, and is not quite so straightforward, it has other advantages and can be applied without any detailed preliminary research.

There is, therefore, at this stage no longer any excuse for not claiming the obvious advantages to be gained by the application of either of the two statistical valuation procedures referred to in this section. Without any increase in the number of samples taken regularly on a mine, these methods will result in reduced valuation errors in individual block and face valuations and will thus also reduce the overall unsatisfactory features associated with ore reserve determinations as discussed in the previous section.

CONCLUSION

This concludes the necessarily brief introduction to the general statistical procedures which can be applied profitably in mine valuation on the Witwatersrand. More specific applications in e.g. the analysis of borehole results and the correlation between gold and uranium values have been omitted from this paper for practical reasons and may possibly be dealt with at a later date. The object of this presentation will, however, have been served fully if a general and keen interest in the application of statistical procedures in the field of mine valuation has been aroused.

ACKNOWLEDGMENT

The author wishes to express his appreciation to the Mines Department for placing the necessary facilities for investigation at his disposal, to a number of colleagues who have assisted him with suggestions and advice, and to the University of the Witwatersrand for permitting this paper to be based partly on his recent Master's thesis.

REFERENCES

1. Bagnold, R. A. The physics of blown sand and desert dunes. Methnen & Co., Ltd., London (1941).

2. Finney, D. J. On the distribution of a variate whose logarithm is normally distributed. Supplement to *The Journal of the Royal Statistical Society*, Vol. VII, No. 2 (1941).

3. Gaddum, Prof. J. H. Lognormal distributions. *Nature*, Vol. 156 (1945), p. 463.

4. Juan, Pae Tsi. On the logarithmic frequency distribution and the semi-logarithmic correlation surface. *Annals of Mathematical Statistics*, Vol. IV (1933).

5. Krige, D. G. A statistical approach to some mine valuation and allied problems on the Witwatersrand. Master's Thesis, University of the Witwatersrand (1951).

6. Pettijohn, F. J. Sedimentary rocks. Harper & Bros., New York (1949).

7. Ross, F. W. J. The development and some practical applications of a statistical value distribution theory for the Witwatersrand auriferous deposits. Master's Thesis, University of the Witwatersrand (1950).

8. Sichel, H. S. An experimental and theoretical investigation of bias error in mine sampling, with special reference to Narrow Gold Reefs. *Transactions of the Institution of Mining and Metallurgy*, London (February, 1947).

9. SICHEL, H. S. Mine valuation and maximum likelihood: Master's Thesis, University of the Witwatersrand (1949).

10. SICHEL, H. S. New methods in the statistical evaluation of mine sampling data. Paper to be published shortly in the *Transactions of the Institution of Mining and Metallurgy*, London (1952).

11. TRUSCOTT, S. J. The computation of the probable value of ore reserves from assay results. *Transactions of Institution of Mining and Metallurgy*, London, Vol. 39 (1929/30).

12. WATERMEYER, G. A. Application of the theory of probability in the determination of ore reserves. *Journal of the Chemical, Metallurgical and Mining Society of South Africa*, Vol. 19 (January, 1919).

ANNEXURE OF MATHEMATICAL FORMULAE

The lognormal frequency distribution * can be represented by the formula :—

$$y = M \exp\left[-a^2 (\log_e z - b)^2\right] \qquad \ldots (1)$$

where $M = \dfrac{a}{\sqrt{\pi}\exp(b + 1/4a^2)}$ for unit area under the curve, z is the variable concerned, a and b are the two parameters, and y the ordinate corresponding to an abscissa value of z.

Substituting $\sigma^2 = \dfrac{1}{2a^2}$ and $\xi - \sigma^2 = b$, (1) reduces to :

$$\Psi(z)\,dz = \left[\sqrt{2\pi}\cdot\sigma\right]^{-1} \exp\left[\frac{\sigma^2}{2} - \xi - \frac{1}{2\sigma^2}(\log_e z - \xi + \sigma^2)^2\right] dz \qquad \ldots (2)$$

with mean* $= \theta = \exp\left(\xi + \dfrac{\sigma^2}{2}\right)$ ... (3)

geometric mean = median value † $= \exp(\xi)$ ... (4)

modal value † $= \exp(\xi - \sigma^2)$ ... (5)

height of mode † $= \left(\sqrt{2\pi}\cdot\sigma\right)^{-1} \exp\left(\dfrac{\sigma^2}{2} - \xi\right)$ ... (6)

variance ‡ $= \mu_2 = [\exp(\sigma^2) - 1]\,[\exp(2\xi + \sigma^2)]$ ... (7)

standard deviation ‡ $= \left[\exp\left(\xi + \dfrac{\sigma^2}{2}\right)\right]\left[\exp(\sigma^2) - 1\right]^{\frac{1}{2}}$ ... (8)

coefficient of variation ‡ (i.e. relative variation) $= [\exp(\sigma^2) - 1]^{\frac{1}{2}}$ ... (9)

3rd moment about the mean : ‡

$$\mu_3 = \left[\exp(\sigma^2) - 1\right]^2 \left[\exp(\sigma^2) + 2\right] \left[\exp\left(3\xi + \frac{3\sigma^2}{2}\right)\right] \qquad \ldots (10)$$

4th moment about the mean : ‡

$$\mu_4 = [\exp(\sigma^2) - 1]^2\,[\exp(4\sigma^2) + 2\exp(3\sigma^2) + 3\exp(2\sigma^2) - 3]\,[\exp(4\xi + 2\sigma^2)] \qquad \ldots (11)$$

skewness : $\beta_1 = \dfrac{{\mu_3}^2}{{\mu_2}^3} = \exp(3\sigma^2) + 3\exp(2\sigma^2) - 4$ ... (12)

* References 2, 4, 5, 7, 8, 9 and 10.

† References 4, 5, 7, 9 and 10.

‡ References 4 and 5.

and kurtosis (peakedness) :

$$\beta_2 = \frac{\mu_4}{\mu_2^{\,2}} = \exp(4\,\sigma^2) + 2\exp(3\,\sigma^2) + 3\exp(2\,\sigma^2) - 3 \quad \ldots \ldots \ldots \ldots (13)$$

For real values of σ, the lognormal curve will, therefore, always be positively skew, with its mode to the left of the mean and its long drawn-out tail towards the right hand side. The curve is also more sharply peaked than the Gaussian Normal Curve, i.e. it is leptokurtic. Values commonly encountered in dealing with the distribution of individual gold inch-dwt values on the Witwatersrand cover the following ranges :—

σ^2	2·00	1·389	1·02	0·781
β_1	564·22	108·75	40·44	20·73
β_2	3,948·62	432·92	122·04	54·91

The normalised frequency function :†

No. (2) can be normalized by substituting $x = \log_e z$ to yield :

$$f(x)\,dx = \left[\sqrt{2\pi}\cdot\sigma\right]^{-1} \exp\left[-\frac{1}{2\sigma^2}(x - \xi)^2\right] dx \quad \ldots \ldots \ldots \ldots \ldots (14)$$

which is the Gaussian Normal Frequency Distribution Law with mean ξ and standard deviation σ.

Frequency of values above value z_i : ‡

$$= \frac{1}{\sqrt{2\pi}} \int_{w_i}^{\infty} \exp\left(-\frac{w^2}{2}\right) dw \quad \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots (15)$$

where $w = \frac{1}{\sigma}(\log_e z - \xi)$

Average of all z values above value z_i : ‡

$$= \exp\left(\xi + \frac{\sigma^2}{2}\right) \frac{\int_{w_i-\sigma}^{\infty} \exp\left(-\frac{w^2}{2}\right) dw}{\int_{w_i}^{\infty} \exp\left(-\frac{w^2}{2}\right) dw} \quad \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots (16)$$

† References 2, 5, 9 and 10.

‡ References 5 and 7.

Sampling distribution of arithmetic mean : †

The sampling distribution of the mean of a sample of size n, from a lognormal parent population has variance $= \frac{\mu_2}{n}$, skewness $= \frac{\beta_1}{n}$ and kurtosis $= \frac{\beta_2 - 3}{n} + 3$.

For the same variance, this distribution is more skew and peaked than the lognormal. Preliminary practical tests † have, however, indicated that as a first and practical approximation for mine valuation purposes this distribution can be accepted as longnormal with variance $= \frac{\mu_2}{n}$.

Sampling distribution of geometric mean : †

In the case of a parent lognormal population, this sampling distribution is also lognormal with variance

$$= \left[\exp\left(\frac{\sigma^2}{n}\right) - 1\right] \left[\exp\,(2\,\xi + \sigma^2)\right] \quad \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \quad (17)$$

unbiased error being exp $(\bar{x} - \frac{\sigma^2}{2n})$

Maximum Likelihood Estimator when σ is known *a priori* : †

This estimator is also lognormally distributed with variance as defined by (17), the estimator being $=$ $t'' = \exp\left[\frac{n-1}{2n}\,(\sigma^2) + \bar{x}\right]$ ... ... (18)

where $\bar{x}$ = mean of natural logarithms of observed values in a sample of size n

i.e. t'' = geometric mean of sample values multiplied by $\exp\left[\frac{n-1}{2n}\cdot\sigma^2\right]$ ... (18a)

The relative efficiencies of t'' and the arithmetic mean are evident on comparison of their variances. For equally reliable results, and assuming the range of σ^2 values listed above (i.e. 0·781 to 2·000), the arithmetic mean will require from $1\frac{1}{2}$ to 3 times as many individual observations as the estimator t''.

The combination of lognormal subpopulations with identical σ_s values and lognormally distributed means : †

Since the product of two lognormally distributed variables corresponds to the sum of two normally distributed variables (after normalization), this product is lognormal in distribution. The parent population comprising the above combination of subpopulations is, therefore, lognormal with mean equal to the overall mean of these subpopulations and parameter σ_p^2 = parameter σ_s^2 of the subpopulations + parameter σ_m^2 of the distribution of the means of these subpopulations ... (19)

Maximum Likelihood Estimator t when σ is unknown : ‡

$$t = e^{\bar{x}}\left[1 + \tfrac{1}{2}V + \frac{n-1}{2^2.2!\,(n+1)}\,V^2 + \frac{(n-1)^2}{2^3.3!\,(n+1)\,(n+3)}\,V^3 + \cdots\right] \quad \ldots (20)$$

† Reference 5.

‡ References 2, 9 and 10.

where $\bar{x}$ = mean of natural logarithms of observed values and V = variance of natural logarithms of observed values.

For n large, (20) tends to $t' = \exp\left(\bar{x} + \frac{V}{2}\right)$ ... ... (21)

The solution of (20) can be effected by use of the tables in Reference 10; this reference also discusses the variances and efficiences of t and t'.

Lognormal correlation and regression.*

For an ideal lognormal correlation surface corresponding on a double logarithmic grid to a normal correlation surface with homoscedastic regression system and linear regression, the lognormal regression curves (or lines) will be of the type $z_i = K \exp(p \log z_j)$ (where K and p are constants and z_i and z_j the two correlated variables) and this will be a straight line only when $p = 1$, i.e. $z_i = K z_j$.

In this particular case, the corresponding curve of regression of z_j on z_i is defined by :—

$$\log_e z_j = \frac{\sigma_j^2}{\sigma_i^2} \log_e z_i + \xi_j + \frac{\sigma_j^2}{2} - \frac{\sigma_j^2}{\sigma_i^2}\left(\xi_i + \frac{\sigma_i^2}{2}\right) + \frac{\sigma_j^2}{2}\left(1 - \frac{\sigma_j^2}{\sigma_i^2}\right) \quad \ldots \ldots \ldots (22)$$

where σ_i and ξ_i = parameters of z_i distribution; and σ_j and ξ_j = parameters of z_j distribution.

In the special case when the means of the z_i and z_j distributions are identical, (22) reduces to :—

$$\log_e z_j = \frac{\sigma_j^2}{\sigma_i^2} \log_e z_i + \left(\xi_j + \sigma_j^2\right)\left(1 - \frac{\sigma_j^2}{\sigma_i^2}\right) \ldots \ldots \ldots \ldots \ldots \ldots \ldots (23)$$

Straightline graphical fit of the lognormal curve *

This can be effected on logarithmic-probability paper but is unsuitable for small samples (statistical sense) and is in any case subject to human errors.

* Reference 5.

A STATISTICAL APPROACH TO SOME BASIC MINE VALUATION PROBLEMS ON THE WITWATERSRAND

By D. G. KRIGE, M.Sc. (Eng.) (Rand)

December 1951

ERRATA.

Page 126 : Footnote :—σ should read σ^2.

Page 129: Footnote:—σ_b should read $\sigma_b{}^2$.
σ_i should read $\sigma_i{}^2$.

Page 137 : Table between formulae (13) and (14) :—σ should read σ^2.

Page 138 : Formula (17) :—Delete σ^2 in second factor and add after the formula ' the unbiased estimator being

$$\exp\left[\bar{x} - \frac{\sigma^2}{2n}\right]'$$

Line above formula (18) :—Delete (17) and substitute $\left[\exp\left(\frac{\sigma^2}{n}\right)-1\right]\left[\exp(2\xi+\sigma^2)\right]$

2nd line below formula (18a) :—σ should read σ^2.

Formula (19) :—σ_p, σ_s and σ_m should read $\sigma_p{}^2$, $\sigma_s{}^2$ and $\sigma_m{}^2$, respectively.

Page 139 : 1st line :—x should read $\bar{x}$.