A STATISTICAL EVALUATION OF MINERAL DEPOSIT SAMPLING

by

HOWARD W. BRISCOE
S.B., Massachusetts Institute of Technology
(1952)

SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
June 1954

Signature of Author ______________________________
Department of Geology and Geophysics, May 24, 1954

Certified by ______________________________
Thesis Supervisor

Accepted by ______________________________
Chairman, Departmental Committee on Graduate Students

ACKNOWLEDGMENTS

The author wishes to express his appreciation to Professor Roland D. Parks, for supervising this thesis. Acknowledgment is also made to Dr. L. H. Ahrens, and Enders A. Robinson for their kind cooperation and suggestions.

A STATISTICAL EVALUATION OF MINERAL DEPOSIT SAMPLING

by

Howard W. Briscoe

Submitted to the Department of Geology and Geophysics on May 24, 1954 in partial fulfillment of the requirements for the degree of Master of Science.

Abstract

If the procedure of mathematical statistics are to be used to improve the analysis of a set of samples of any kind, the samples being analysed must be statistically unbiased; and some information about the size or distribution of the population being sampled must be available. The geologic meaning of a statistically unbiased sample is discussed with respect to collecting and weighting the sample units. An investigation of the presently available knowledge about the distribution of geologic populations indicated that the expected log-normal distribution does not hold for all kinds of deposits, and there are indications that the shift from the log-normal distribution may be related to the effect of weathering of the deposit. A more thorough study is needed before any safe conclusions about the general distribution of elements in various geologic bodies can be reached.

Possible statistical aides in detecting zoning in geologic deposits are discussed. These procedures are based on the number or length of runs of values above and below the mean or median of an ordered sequence of samples.

The use of the "t" distribution for estimating the average assay value of a deposit from the assay values of samples taken from the deposit is described, and possible variations on the procedure are discussed.

Finally, the method of Sequential Analysis and its potential geologic applications are considered. The use of Sequential Analysis based on the distribution of the sampled population is impractical without a more dependable knowledge of the analytic form of this distribution, but the use of Sequential Analysis based on the distribution of the mean of all possible samples of a fixed size merits further investigation.

TABLE OF CONTENTS

TABLE OF CONTENTS (continued

I. INTRODUCTION

Thousands of dollars each year are spent for sampling geologic bodies, and millions of dollars are spent each year on the basis of the results of these samples. As a result, the problems of effectively sampling mineral deposits and of reliably interpreting the resulting samples have been considered imaginatively; intuitively; empirically; and even, occasionally, mathematically.

The application of imaginative, intuitive, and empirical methods is limited only by the fact that the accuracy of the results is somewhat unpredictable. On the other hand the mathematical approaches offer controlled accuracy but are limited by, first, the validity of the assumptions on which the mathematical method is based; second, the geologist's ability to understand and correctly use the mathematics involved; and, third, the extent of the calculations that must be made to use the test.

In this paper, the application of several statistical procedures to the design of efficient methods of sampling of geologic bodies will be investigated. Specifically, the statistical interpretation of samples and its effect on the procedures for cutting samples will be considered first. Next the use of statistical tests of randomness of a series of numbers will be considered as a method of testing for zoning in a deposit, and a method of obtaining statistical confidence limits on the average assay value of a geologic body from a series of samples will be described. Finally the possible value of applications of the techniques of sequential analysis to geologic sampling will be discussed.

It must be realized from the start that statistics, at best, can be used only as a valuable tool to help the geologist to answer, on a mathematical basis, a few of the many problems that confront him. The problem is still a geologic problem, and the parts of the problem to be analysed statistically must be clearly defined and must supply some information about the geology of the deposit. It would be a waste of time, for example, to spend several hours testing a deposit for possible zoning if the geology of the deposit was such that it was sure to be zoned. The statistical tests alone can have no value. The assumptions which form the basis for the tests must be geologically reasonable, and the results of the test must be such that they can be interpreted geologically.

II. THE STATISTICAL NOTION OF A SAMPLE

Geologic samples may serve several purposes; they may be used to help determine the geologic structure or age of a deposit; they may be used to investigate the ease with which certain valuable ingredients can be separated from the gangue or freed from the chemical compounds in which they are found; they may be used to determine boundaries of commercial deposits; or they may be used to estimate the grade of a deposit that is to be mined. The concept of the sample is different for each of the purposes for which a sample is to be used; however, the same sample may serve more than one purpose. In one case the sample may be considered as a collection of fossils or sand grains; in another the sample may be considered as an aggregate of chemical compounds. When the samples are to be used to outline a deposit, each sample must be considered as representing a specific part of the deposit, most commonly the volume of the deposit bounded by the surface equidistant from that sample and the surrounding samples closest to it.

The most valuable applications of statistics are associated with sampling to estimate the grade of a deposit, and the concepts and characteristics of samples taken for this purpose will be considered in some detail.

The Sample Universe and Unit Samples.

In order to develop the concept of a geologic sample which is appropriate for statistical purposes, the geologic body being sampled is divided into a large number of small parts, each of equal size and similar shape. The aggregate of all of these small parts will be referred to as the "sample universe," and each part will be called a "sample unit." A "sample of size N" will refer to any group of N individual sample units. For simplicity, this discussion will assume that the assay value of each sample unit can take one of a finite number of possible values (this assumption is valid in that it is possible to measure the concentration of any component of the sample unit to only a finite accuracy, say to .01 percent of the weight of the sample unit). If the assay value in percent of the weight of the sample unit is plotted as the abscissa and the number of sample units in the sample universe which have the given assay value is plotted as the ordinate, the resulting curve is the "frequency distribution" of the sample universe. The frequency distribution of the sample universe completely describes the amount of the interesting component in the deposit and the variation of the concentration of that component, but that it does not describe the geometry of the zoning in the deposit is obvious since no measure of distance or direction is involved. The average assay value of the universe of samples is the value of the abscissa such that equal areas lie under the parts of the curve on each side of it.

If the assay value is plotted as the abscissa, and for every assay value the percent of the number of samples in the sample universe with that assay value or less is plotted as the ordinate, the resulting curve is the "cumulative distribution" or simply "the distribution" of the sample universe. The average assay value is that value of the abscissa corresponding to an ordinate of 50 percent. Any two samples for which the ratios of corresponding ordinates of the frequency distributions are equal will have the same cumulative distribution.

The statistical concept of a sample is based on the fact that a random sample of size N will tend to have a distribution similar to the distribution of the universe being sampled. As N is increased, the distribution of the sample is increasingly likely to be similar to the distribution of the universe; and, when N includes all possible sample units, the distribution of the sample must be the same as the distribution of the universe.

The various procedures for statistical sample analysis are based on determining the probability that the sample taken has a distribution similar to the distribution of the universe. In general the sample alone is not enough information on which to base a rigid test. At least the size of the sample universe must be known before any calculations can be made to determine the probability that the sample distribution will be similar to the distribution of the universe. Usually some information about the actual distribution of the universe or of some characteristic of the sample is known, and all statistical tests that have been developed place some requirement on the minimum amount of information that must be used in that test. The information used by many tests is in the form of one or more assumptions concerning the distribution of the sample universe and may not ever enter the analytic formulas for applying the test, and it is important that the validity of the assumptions be considered before the test is applied. It is evident that a statistical test which uses as much of the available information as possible should be chosen for any analysis because the more information the test has to work with the more reliable the results of the test will be for a given sample size.

Geologic Sampling.

If geologic samples are to be considered statistically they must be unbiased; that is, they must be representative of a universe of sample units. The most obvious way of getting biased samples would be to take most of the samples from a zone that is not representative of the whole deposit, a zone of high grade ore for example. The use of improper weighting factors in computing the average assay value of the sample will also result in a biased sample because such weighting factors would cause certain samples to receive undue emphasis. A third cause of biased samples may be concentration or dilution of the samples while they are being cut.

Another form of error can be introduced into the samples by taking sample units of different sizes. Consider, for example, a deposit in which the interesting component of the rock occurs in the form of masses of one pound or less. If the size of the sample unit is one pound, it is possible to get an assay value of 100 percent; but, if the size of the sample unit is five pounds, the maximum possible assay value is 20 percent. Thus a sample composed of both sizes will result in an inconsistent sample frequency distribution and a biased sample.

The closer the sample universe comes to being a homogeneous mixture of the valuable constituent and the gangue, the smaller the error introduced by any of the errors listed above will be. In extremely erratic deposits, like many gold deposits, it may be very difficult to get a safely unbiased sample.

III. NONHOMOGENEOUS DEPOSITS

Sample Patterns and Weighting Factors.

The existence of high or low grade zones in a geologic deposit is obviously of interest to the mining company in the production stage. Such zoning must be taken into account in the sampling stage as well as the production stage, since a preponderance of samples from any one zone will result in a biased sample. The only important limitation on the pattern of cutting samples is the requirement that the pattern should be designed to give a sample that is representative of the sample universe. If the deposit being sampled is homogeneous, almost any sample pattern will result in a representative sample; but, if the deposit is zoned, the sample pattern should be designed so that the number of samples taken from each zone is proportional to the size of that zone. If the deposit is considered to be divided into cubes and sample units are taken at all corners of each cube so that the sample units are equally spaced in three dimensions, any zone which is larger than one of the cubes will be represented in about the right proportion in the sample. Unfortunately, it is not often convenient to take samples such that the sample units are equally spaced in all three dimensions, and the sample pattern may even be entirely determined by physical conditions that make it possible to cut samples from only certain accessible places in the deposit. If the bounds of the various zones have been determined by preliminary sampling or by geologic considerations, there are an unlimited number of possible sample patterns that may be used to include the correct proportion of samples from each zone.

When the sample units must be cut with no knowledge of the zoning or with only a vague idea of where the zones may be, the boundaries of the zones can often be determined from the sample; and appropriate weighting factors may then be used to correct for the zones. For example, consider a sample which indicates a zone of high grade ore which includes 40 percent of the deposit. If only 30 percent of the sample has been taken from that zone, each sample unit in the high grade zone should be multiplied by a factor of $4/3$; and each sample unit from the rest of the body should be multiplied by $6/7$ in order to compute an unbiased estimate of the average value of the deposit. Since the correction is a correction on the frequency of the occurrence of the sample unit in the sample, the unbiased estimate of the variance must be found from the formula

$$s^2 = \Sigma(x_i - \overline{x})^2 \, fw/N + 1 \tag{1}$$

where f is the frequency of sample units with value x_i, $\overline{x}$ is the unbiased estimate of the mean, and w is the weighting factor (4/3 for samples in the zone and 6/7 for those not in the zone in the example above). If the nature of the zone is very difficult from the rest of the deposit, the use of the corrective weighting factors may not result in truly unbiased samples. When the zones are so different that the use of weighting factors is invalid, it is likely that the zones will require different metallurgical methods for refining the ore, or that different methods will be used to mine the zones, and the separate zones will be treated as separate deposits.

The application of various weighting factors to geologic samples is a common practice; however, many of the weighting factors are not based on statistically sound reasoning and will actually cause the samples to be biased. Weighting functions involving powers of the frequency or assay values of the unit samples along a fixed direction are often used to correct for geometric considerations in the deposit. These weighting factors, particularly factors of the square or cube of the frequency are valid for a few special combinations of zoning geometry and sample pattern; but the general use of weighting factors based on a fixed function of the frequency is unsound and may result in serious errors. Any legitimate weighting function that is used to correct for nonhomogeneity of the deposit must be based on some knowledge of the relative size and the relative grade of the various zones and must be logically designed to cause each zone to be represented in proportion with its size in the final estimates of the parameters (average assay value, variance, etc.) of the distribution of the universe.

The units used for expressing the assay values may introduce an unsound weighting factor into the values of the sample units. Assay values should be measured in units equivalent to the percent of the volume or weight of the sample unit that is accounted for by the constituent being considered. Units of pounds-of-x per pound-of-sample or cubic-feet-of-x per cubic-foot-of-sample (where x is the constituent of the sample that is of interest), for example, would be appropriate units for the assay values. When the assay values are found by multiplying the percent of x by half the distance to the adjacent samples, the distance factor is actually a weighting factor and may place undue emphasis on certain samples which, for some reason unrelated to the assay value, have more than the average spacing between samples. Hidden weighting factors of this kind are probably the most common error in current geologic sample analysis procedure.

Detection of Zoning.

The existence of zones in many deposits can be determined from geologic information before any samples are taken, and the boundaries of many zones coincide with geologic boundaries which are also known before sampling has been started. In other cases the zones may not be recognized until preliminary or even final samples have been taken, but by the time the sampling has been finished any zones existing in the deposit except those that are smaller than the sample spacing should have been recognized or at least suspected. If zones are completely missed by the sample pattern or are arranged such that they appear as random variations in the sample values, there is no way to detect their presence except closer sampling. When zones are not expected or detected by the sampling, they will usually not be large enough to have an important affect on the average assay value; or, if they appear as random variations in the samples, the sampling pattern may give a representative sample in spite of the zones. There will, of course, be a few isolated cases where important zoning will go undetected and the sample will be misleading, but these cases will be rare if a reasonable number of sample units are taken throughout the deposit.

Exceptional cases in which reasonably large zones of slightly high or low grade ore must be detected may occur occasionally, and in such cases there may be some question about the significance of a group of sample units that are consistently above or below the average assay value by a small margin. In these rare cases statistical tests can be used on the samples along any line through the suspected zone to find the probability that the given sequence of values is a random sequence. Two such tests will be considered here.

Consider an ordered sequence of numbers each of which must come from one or the other of two mutually exclusive classes of numbers. Any group of one or more consecutive numbers from the same class will be referred to as a run. If the sequence contains r numbers from Class I and s numbers from Class II and if the sequence is a random sequence, all possible combinations of r numbers of Class I and s numbers of Class II are equally probable; and the probability of getting k runs in the sequence of $T = r + s$ numbers is

$$P(k) = 2 \frac{(r-1)!(s-1)!\, r!\, s!}{(k/2-1)!(k/2-1)!(r-k/2)!(s-k/2)!\; T!} \qquad (2)$$

if k is even, and

$$P(k) = \frac{(r-1)!(s-1)!\, r!\, s!}{\frac{k-1}{2}!\, \frac{k-3}{2}!\, T!} \left[\frac{1}{r-\frac{k+1}{2}!\, s-\frac{k-1}{2}!} + \frac{1}{r-\frac{k-1}{2}!\, s-\frac{k+1}{2}!} \right] \qquad (3)$$

if k is odd.[1]

A sequence of numbers filling all the requirements of sequence just described can be constructed from the consecutive assay values of sample units along a line through a suspected zone by computing the deviation of each value from the mean assay value for the sample units along that line and associating positive deviations with one class and negative deviations with the other class. If the line along which the samples are taken passes through a zone, all combinations of positive and negative deviations are not equally probable; but all possible sequences are equally probable if the line lies entirely within one homogeneous zone.

The probability of getting more than A runs and less than B runs in a random sequence can be found from the probability, P(k), of k runs by using the formula

$$P(A < k < B) = \sum_{k=A}^{k=B} P(k) = \sum_{k=2}^{k=B} P(k) - \sum_{k=2}^{k=A} P(k). \qquad (4)$$

A test for the randomness of a sequence of numbers can now be made by determining the minimum number, A, and the maximum number, B, of runs such that the expected number of runs from a random sequence will be between A and B in a predetermined percent of the possible sequences. The predetermined percent of the time that k will be between A and B is known as the "level of significance," and the values of A and B that result in a given level of significance are the values of A and B for which $P(A < k < B)$ is equal to the significance level. The use of such a test has been made practical by published tables of the probability that k will be between 2 and B. The probability is tabulated as a function of r, s, and B.[2] We may reject the hypothesis that the series is random if there are either too many or too few runs in the series, and in any test values of A and B should be determined so that the probability that k is less than A is equal to the probability that k is greater than B. This condition will be satisfied if

$$P(k < A) = P(k > B) = \frac{1 - P(A < k < B)}{2} = \frac{1}{2}\left[1 - \sum_{k=2}^{A=B} P(k) + \sum_{k=2}^{k=A} P(k)\right] \qquad (5)$$

An example may help to clarify the test described above. Consider the sequence of average assay values from the line of drill holes listed in table I. This sequence was taken from the test holes along the line 500 W in the lateritic iron deposit described in Appendix A. For this sequence

$r = - 5$

$s = - 5$

$k = - 7$

and the values of A and B such that

$P(A < k < B) = .95$

will be determined. Thus the level of significance will be 95 percent. From Eq. (5) the probability that k is less than A is found to be $1/2 \ 1 - P(A < k < B)$ or $P(k < A) = 1/2\ (1 - .95) = 1/2\ (.05) = .025$.

The probability that k is less than B is one minus the probability that k is greater than B, so that

$$P(k < B) = 1 - .025 = .975.$$

The required values of A and B such that

$$\sum_{k=2}^{k=B} P(k) = .975$$

and

$$\sum_{k=2}^{k=B} P(k) = .025$$

are looked up in the tables and are found to be

$$A = -2$$
$$B = -9,$$

and the sequence is found to be random since the value of k in the sequence seven.

In some cases it may be more appropriate to base the test on the probability of getting a run as long as the longest run in the sequence rather than the probability of getting the number of runs found in the sequence. A test based on the probability of getting a run of size s on either side of the median of a sequence and the tables for using the test have been derived.[3] The computations involved in setting up the tables for this test are more complex than those for the test based on the number of runs, and the formulas will not be reproduced here. On the other hand, the use of the tables is simpler because, first, the hypothesis of a random sequence can be rejected only on the basis of a run that is too long and, therefore, only one value must be looked up in the table; and, second, the test is based on runs above and below the median instead of the mean so the probabilities need be tabulated only for two variables, the number of samples in the sequence and the size of the run. The tables available for this test are rather limited.

The length of run for which the hypothesis of a random sequence should be rejected on the 95 percent significance level is tabulated as a function of the number of samples in the sequence in table II.

Applying this test to the sequence described in table I to determine whether the sequence is random on a 95 percent significance level it is found that the number of samples is ten and the longest run is a run of two which is shorter than the run of five needed to reject the hypothesis of a random sequence, and the sequence is found to be a random sequence. In each of these tests the acceptance of the hypothesis of a random sequence is associated with a lack of significant zoning along the line of samples being tested.

Clearly, any test for zoning which is based on the number or size of the runs of high and low values can detect only those zones which are large enough or have sufficient contrast to affect the size or number of the runs and will not detect many possible configurations of small zones. The test based on the number of runs will be more effective for detecting small zones with high contrast between zones because the high contrasts will place the mean safely between the two zones so that each zone will be one run and the number of runs will be small (approximately equal to the number of zones along the sequence).

Errors Resulting from the Methods of Cutting Samples.

Dilution or concentration of samples as a result of the method used to cut samples has been mentioned as a possible cause of biased sampling. Obviously cutting procedures that may introduce errors into the samples should be avoided; but, often, there is no other way to get the samples; and in these cases the assay values of the samples are usually adjusted to correct for the error. Here, again, zoning in the deposit may complicate the problem since a zone of higher or lower grade ore than the rest of the deposit is often also a zone of different physical characteristics (more friable or more coarsely crystalline for example) from the rest of the deposit. Thus, the error resulting from cutting the samples may differ in the various zones because the ability to cut a good sample depends on the tendency of the rock to crumble. As a result, weighting factors in zoned deposits may have to be designed to correct for cutting errors as well as the proportion of samples in the various zones. The determination of such weighting factors must be based on experience with the procedure being used to take samples.

Weathered zones, contact zones, and other zones that are not typical of the main body of the deposit must be considered in a similar manner to zones while taking samples. The method used to cut the sample units and the sampling pattern must be such that a preponderance of samples is not cut from unimportant zones of this kind.

IV. THE SELECTION OF A TEST FOR ESTIMATING THE AVERAGE VALUE OF A DEPOSIT

Once the size and shape of the geologic deposit have been found and weighting factors to account for zoning, cutting errors, and other possible errors have been determined, the problem becomes one of estimating the grade of the deposit; and this is primarily a statistical problem. Several tests are available for the estimation of the average assay value, and a number of factors must be considered before choosing the test to be used.

Errors and Confidence Limits.

The purpose of sampling for the grade of a deposit is primarily to determine whether the concentrations of the various significant constituents are within the limits which define a commercially valuable deposit. The analysis may result in information in one of several forms all of which satisfy the basic purpose of the sampling; but some forms supply more additional information than others. The most common tests result in either rejection or acceptance of a hypothesis that the mean of the sample universe has a given probability of lying within certain predetermined limits, or they determine the confidence limits above and below the mean of the sample within which the mean of the universe will fall with a predetermined probability. The latter result is usually preferable because it gives the limits within which the mean of the universe is expected to be, regardless of how high or low the limits are, instead of merely determining whether the mean is expected to be in a fixed interval.

The final decision made on the basis of the statistical analysis of the sample may make one of two possible errors. An "error of the first kind" is made when the average assay value of the sample universe is actually within the limits of a commercial deposit and the results of the analysis indicate that the average assay value is not within these limits. An "error of the second kind" is made when the average value of the universe is not within the commercial limits and the analysis indicates that it is within these limits. A decision to accept the hypothesis that the average assay value is within the commercial range is seldom questioned. As soon as the decision is made, mining operations are begun; and large sums of money may be put into the mining of a deposit before an error of the second kind is detected. On the other hand, when the analysis of the sample indicates that the deposit is not a commercial deposit, the deposit will probably either be subjected to more detailed sampling, or it may be set aside until the commercial limits change, but it will not often be completely discarded. Thus either kind of error may be costly and, obviously, undesirable; but an error of the second kind may be more costly than an error of the first kind unless the mining company doing the sampling depends on the success of the deposit in question for its existence.

The ideal test, then, is one which allows the probability of an error of either kind to be determined or, preferably, preset. There are, however, very few tests that allow both errors to be controlled. Even the choice of which error is to be controlled is not usually open to the investigator; although there are usually some practical reasons for preferring control over a particular error. The test chosen should, of course, be the one which leads to the smallest error of one kind when the error of the other kind is fixed. However, it is often extremely difficult or even impossible to determine the error not involved explicitly in the test.

If the error of the first kind is to be reduced to zero, we must accept the hypothesis that the deposit is of commercial value all of the time. This would make the probability of an error of the second kind rather high. Conversely, reducing the probability of an error of the second kind to zero requires that the hypothesis should never be accepted, and the probability of an error of the first kind is, then, large. Thus, the extreme reduction of one error will cause an increase in the other, and the practical use of a statistical test requires that a compromise which does not ask too much safety in either direction be found.

The statistical probability of error, significance level, or percent of confidence interval should not be confused with the tolerance limit on the assay value. For example, if the 95 percent confidence interval on the mean of a distribution is 100 + 10, it means that in 95 percent of the possible occurrences of the sample taken the sample will be from a universe with an average value between 90 and 110. The tolerance limits on the mean are 10 percent above or below the estimated mean of 100 for this case, but the level of significance of the statistical test is 95 percent. The significance level and allowable tolerance limits are not independent, and for a given sample size an attempt to get an extremely high significance level will result in wide tolerance limits. For geologic sampling a significance level greater than 95 percent may be unwise and even unrealistic in that it may be expecting an absolute guarantee from a statistical method. In the various stages of the sampling procedure, from the preliminary survey to the final production control, tolerance limits of from 10 percent to less than one percent are not unreasonable.

Frequency Distribution of Geologic Samples.

It has already been pointed out that some information besides the sample itself is needed for the statistical analysis of the sample, and a test which uses as much information as possible should be chosen for any analysis. The procedure should not use any information that is subject to any doubt, however. The most desirable situation is one in which the available information is in the form of an assumption concerning the frequency distribution of the universe from which the sample was taken. The maximum amount of information in this form would be a knowledge of an analytical expression for the frequency distribution of the universe in which the parameter being estimated (the average assay value in geologic work) appears as the only unknown parameter. Almost all of the statistical procedures that have been worked out in detail are based on the assumption that the universe has one of three convenient analytical forms in which the distribution is completely described by three parameters or less. These three distributions are the normal, the Poisson, and the binomial distributions. The binomial distribution is concerned with the distribution of sample units which may take either of two specific values when the probability of a sample unit of each value is known, and it is not a distribution that could apply to geologic samples. The Poisson distribution occurs as the limit of several forms of distributions as the samples get large subject to certain restrictions and would not be expected as the distribution of a geologic sample. The normal distribution is the only one of the three common distributions that may fit the distribution in a geologic deposit, and it is a distribution that occurs quite commonly in nature.

Before attempting to choose an efficient statistical procedure for analysing geologic samples, it is imperative that the expected nature of the distribution of the universes from which the samples will be taken should be investigated. If the distribution is found to be normal, a wealth of procedures allowing either or both of the two kinds of errors to be determined will be available for estimating the mean of the assay value of a deposit. There is reason to believe that geologic distributions will not be normal, but that they will be log-normal.[4] This means that the distribution of the log of the assay values of the sample units will be normally distributed, and that an analytic expression for the distribution is available.

To determine the distribution of the assay value of the sample units in a geologic deposit, the deposit would have to be completely mined out in units the size of the sample units that would be used for sampling. Each unit would have to be analysed separately because any combining of the samples before the chemical analysis would smooth the distribution. Such a procedure is obviously impossible, and some other method must be used to determine the desired distributions. If the distribution is to be of any practical value, it must hold for all deposits, or for a very large class of them. A distribution that is sufficiently universal to be useful should be evident in the samples from a large number of deposits.

The investigation of the distributions of elements in geologic deposits was based on 17 sample distributions from seven mining properties sampled by the United States Bureau of Mines,[5] an article on drill core sampling in the Whitwatersrand,[6] and the iron deposit described in Appendix A. Of these deposits one is in the soft, altered material in a volcanic agglomerate; three are laterites; one is a contact metamorphic deposit; one is a replacement deposit in sedimentary strata; two are sedimentary; and one is in a schist. The distributions found in these samples can be divided into two groups. The log-normal distribution found in igneous rocks by Dr. Ahrens[4] also appeared in all the samples of the contact metamorphic deposit, in all the sedimentary deposits, in the replacement deposit, in the schist deposit, and in the non-metallic constituents (SiO_2, sulphur, and phosphorous) of a laterite deposit (see figures 1 and 2). The characteristics of the second group were not clearly determined, but the distributions tended to be more symmetric than the log-normal distribution and varied from bi-modal to slightly negatively skewed distributions (see figures 6, 7, and 8). The samples whose distribution fell into the second group included the metallic constituents (iron, nickel, and chromium) of the laterites and the deposit in the altered volcanic agglomerate.

The results of this investigation have suggested several interesting directions in which further work may prove valuable. Most of the samples used were taken as part of preliminary investigations of geologic bodies and were not designed to accurately determine the details of the distribution of the deposits; and, therefore, the following discussion must be treated as mere speculation until considerably more work is done to prove or disprove the suggested theories.

Figures 1 and 2 are frequency distributions for two of the deposits with typical log-normal distributions. The cumulative distributions of the logs of these values are plotted on normal probability paper in figures 3 and 4. If the distributions are log-normal, the plots in figures 3 and 4 should be straight lines. The curve in both these lines indicates a tendency to have too many values greater than the peak value, and neither distribution is actually log-normal. One of the deposits that has a log-normal distribution (a barite deposit that replaces sedimentary beds) has been sampled by trench and drill samples, and the frequency distributions of the two kinds of samples are plotted in figure 5. The trench samples are found to have a higher percentage of samples after the peak than the hole samples. If the trenches were in a weathered zone along the surface of the deposit, the shift in the frequency distribution may be related to the weathering. It is not statistically sound to make such a general statement on the basis of one example, and a much more thorough study should be made before the suggestion can be seriously considered. If the second kind of distribution can be considered as the end result of the shift in the frequency distribution, however, the fact that this distribution seems to occur only as the result of extreme weathering would support the suggestion that the shift is due to weathering. The degree of shifting of the frequency distribution of a sample from the log-normal distribution may even provide a quantitative measure of the degree of weathering in the zone from which the sample was taken.

Once the characteristics of the frequency distributions for various kinds of deposits have been determined they may be used as a check for sample errors. If the distribution of a sample is not similar to the characteristic distribution for the type of deposit it was taken from, there may be reason to believe that the sample is not representative of the deposit; that errors have been introduced by the cutting method; or that the geologic interpretation of the deposit has not considered some important factor such as weathering. The second possibility may be most important practically, for it has been shown that some methods of cutting samples may introduce large errors into the samples.[6] In the case of the carbon lead of the Whitwatersrand the erroneous drill samples and the carefully controlled hand samples were found to have similarly shaped log-normal distributions; although the assay values involved were of different magnitudes. Thus, in that case the shape of the distribution curve of the drill samples does not indicate the error in the samples. However, in some types of deposits in which the physical characteristics of the rock change with the grade of the ore, the errors caused by concentration or dilution of the sample while it is being cut may be limited to very high or very low grade samples, and a shift in the sample distribution may be found as a result of the errors.

One fact is certain, that no simple analytic expression can be used for the frequency distribution of all forms of deposits. Complex curve fitting techniques may be used to derive a general distribution, but it would not be practical to use an expression which could only be understood by a trained mathematician. Without further research it would be unwise to base the analysis of any broad group of deposits on the log-normal distribution because the group of deposits to which it could be applied has not been clearly defined yet, and there is some question as to how close any deposits come to actually being log-normal. Probably the constituents of most unweathered igneous deposits are log-normally distributed.

Tests Based More Widely Applicable Assumptions

When the distribution of a function is not known in detail, a class of assumptions based on the distribution of the sample values of the parameter being estimated is often a convenient and valid basis for analysis. As in the case of procedures based on assumptions concerning the distribution of the universe, most of the procedures based on the sample distribution of a parameter of the distribution of the universe have been worked out for only a few cases. In particular, the normal distribution and the distribution of χ^2 are most extensively described.

Any procedures based on the sample distribution of the mean that are used in estimating the average assay value of a geologic deposit should be expected to give results that are slightly pessimistic since they would not use either the available geologic information about the deposit or some of the limited information available about the distribution of the universe being sampled. Actually, information concerning the distribution of the sample universe is one statistical form of information about the geology of the area, and attempts to learn more about the distributions of various geologic deposits are basically attempts to translate information from the language of geology to the language of statistics.

V. A VALID STATISTICAL PROCEDURE FOR ESTIMATING THE AVERAGE ASSAY VALUE OF A GEOLOGIC DEPOSIT

The Basis for the Analysis.

It can be shown that if x has a distribution with mean $\overline{X}$ and standard deviation S for which the moment-generating function exists, then the variable

$$z = (\bar{x} - \overline{X})\, N^{1/2}/S \tag{6}$$

has a distribution which approaches the standard normal distribution as r gets large.[7] The moment-generating function of the distribution does not have to be used explicitly in the procedure to be considered here; and, therefore, the method of determining it will not be described.[8]

The calculation of z requires the standard deviation, S, which may not be known exactly. When S is not known, the standard deviation of the sample, s, can be substituted; but the distribution of

$$t = (\bar{x} - \overline{X})\, N^{1/2}/s \tag{7}$$

is not normal. The distribution of t is known and tabulated as a function of the sample size, N, or the number of degrees of freedom which is N - 1.[9] The t distribution holds rigidly for normal sample universes, and is approximately correct for large samples from other distributions (a sample greater than 50 is usually considered a safe sample in any case).

The Procedure for Estimating the Assay Value of the Deposit.

The first step in any estimation of the average assay value of a deposit is, obviously, to assume that it is approximately given by the average assay value of the sample. The various statistical procedures are, essentially, methods of determining the tolerance limits around the sample mean within which the universe mean is expected to fall. In the procedure to be described here the desired statistical level of significance, L, is decided first. The values t_1 and t_2 such that t is less than t_1 for $(1 - L)/2$ percent of the time and t is greater than t_2 for $(1 - L)/2$ percent of the time are found from a table of the t distribution for a sample of size N. Solving the equation (7) for $\bar{x} - \bar{X}$,

$$\bar{x} - \bar{X} = ts/N^{1/2} \tag{8}$$

where $\bar{x}$ = sample mean

$\bar{X}$ = universe mean

s = sample standard deviation

N = sample size.

The values t_1 and t_2 are substituted in Eq. (7) to find the values, x - X, of the tolerance limits on the estimated value for the mean at a significance level L.

When calculating the values of x and s, the weighting factors being applied must be applied in the proper place. Weighting factors may apply either to the assay value or to the frequency of the assay value.

The following formulas should be used:

$$x = \frac{\sum_{i=1}^{M} x_i\, w(x_i)\, f_i\, w(f_i)}{\sum_{i=1}^{M} f_i\, w(f_i)} \tag{9}$$

and

$$s^2 = \frac{\sum_{i=1}^{M} \left\{x_i \left[w(x_i)\right] - x\right\}^2 f_i\, w(f_i)}{\left[\sum_{i=1}^{M} f_i\, w(f_i)\right] - 1} \tag{10}$$

where

M = number of possible values of the assay value

x_i = i' th assay value

f_i = frequency of i' th assay value

$w(x_i)$ = weighting factor applied to i' th assay value

$w(f_i)$ = weighting factor applied to frequency of i' th assay value

s = positive square root of s^2.

The denominator of 9 and 10 is the generalized form of the sample size and, if it is not equal to N, it should be used in place of N in all calculations involving s or $\bar{x}$.

An Example of the Use of the "t" Distribution.

To illustrate the use of the procedure described above and to investigate certain characteristics of the results, the test was used on four samples of different sizes drawn from the iron deposit described in Appendix A. The results of this analysis are not as close to the production figures for the deposit as would be desired, but this is probably because the cut-off limits of the ore body as determined by the mining company were not clear in some places, and the samples included in this analysis may not be entirely taken from the part of the deposit mined as ore.

The analysis of a sample composed of alternate drill holes along the north-south coordinates on the map in Appendix A will be described in detail. The majority of the drill cores from this deposit were divided into five foot sections for analysis; and for that reason, the sample unit used in the calculations was a five foot section of drill core. The average assay value of each drill hole was computed, and each value was weighted by the number of five foot sections in the length of the corresponding hole. The same results could have been reached by treating each five foot section separately, but the calculations would have been less easily mechanized. Table III shows the form of the calculations for this sample. For a significance level of 95 percent

$$t_1 = -1.96$$

$$t_2 = +1.96$$

then

$$(\bar{x} - \bar{X})_1 = \ 1.96\ (3.054)/24.94 = \ .239$$

$$(\bar{x} - \bar{X})_2 = -1.96\ (3.054)/24.94 = -.239,$$

and the average assay value of the deposit is estimated as $58.75 \pm .24$ to a significance level of 95 percent.

An Evaluation of the Procedure and its Variations.

The only error possible in the actual statistical procedure described above is that of accepting the tolerance limits on the mean assay value when this value is actually outside of the tolerance limits. If the sample is taken in such a way that it is a reasonably unbiased sample, the level of significance determines the probability of making this error. The geologic hypothesis which is to be accepted or rejected on the basis of the results of the analysis is usually the hypothesis that the average grade of the deposit is above the minimum commercially valuable limit. If the whole tolerance interval for the mean is above or below the commercial border value, the hypothesis is respectively accepted or rejected. The values of t_1 and t_2 were chosen such that the probability that the mean is outside the tolerance interval on the low side is equal to the probability that it is outside on the high side, and the probability of either is $(1 - L)/2$ where L is the significance level chosen. The maximum probability of errors of the first or second kind occur in that order when the upper or lower limit of the tolerance interval is on the commercial limit. In each case the probability of being out of the tolerance limit on the side that corresponds to the commercial cut-off limit is the probability of the corresponding error and is $(1 - L)/2$. Thus, if the tolerance interval on the mean is entirely above or entirely below the commercial cut-off limit, then the maximum probability of an error of the first kind is equal to the maximum probability of an error of the second kind and is $(1 - L)/2$. If the commercial cut-off limit is inside the tolerance interval on the mean assay value, the probability of an error can be found by determining the confidence level, L^*, such that the tolerance interval corresponding to L^* is entirely above or below the cut-off limit. The maximum probability of an error is then $(1 - L^*)/2$.

If there is reason to prefer a higher probability of one kind of error than of the other, the values of t_1 and t_2 can be determined so that the probability of t being below t_1 is not equal to the probability that t will be above t_2. In any case the sum of the probability of being above t_2 and the probability of being below t_1 is $1 - L$.

The possible variations in the procedure and the limits on the accuracy of the procedure can be best realized by recognizing that there are effectively three parameters in the procedure, and any two of these parameters can be fixed arbitrarily for any one analysis. The third parameter is determined when the first two are prescribed. The three parameters are the size of the sample, N; the confidence level, L; and the tolerance limits on the mean, $\bar{x} - \bar{X}$.

First, consider the variations on the procedure that can be obtained by fixing different combinations of the three parameters. In the description of the procedure above, the level of significance and the size of the sample was fixed. The resulting procedure determined a tolerance interval for the average assay value of a fixed sample. The second possibility, that of fixing N and the tolerance interval, leads to a test which, effectively, determines whether the mean lies within a given interval. These two types of test were mentioned on page 13 of this paper, and it was decided that the former was the preferred test.

The last combination leads to a rather important variation. When the tolerance interval on the average assay value and the level of significance are both fixed; the procedure that results is a form of sequential sampling procedure which provides a way of minimizing the size of the sample necessary for the test. In many geologic situations the cost of cutting samples is quite high, and a method of evaluating the accumulated sample at each step so that sampling can be stopped as soon as the required level of statistical accuracy is reached may prove to be a valuable tool. Such a procedure consists of fixing a hypothesis that the mean assay value lies within certain limits, say between x_1 and X_2, and fixing the level of the significance in the form of two probabilities, L_1 and L_2 such that

$$L_1 = 1 - L_2 \qquad (11a)$$

and

$$L_2 - L_1 = L. \qquad (11b)$$

An initial sample of size N is taken, and the sample mean is computed. The terms $\bar{x} - \bar{X}_1$ and $\bar{x} - \bar{X}_2$ are then computed and the two values of t corresponding to them for the sample size are determined. The probability of t falling in the interval between the two calculated values of t can be found from a table of the t distribution for the appropriate size of the sample taken; and, if this probability is between L_1 and L_2, the size of the sample is increased; and the procedure is repeated for the new sample size. If the probability of t being between the two calculated values is less than L_1, the sampling is over; and the hypothesis is rejected. If the probability that t is between the calculated values is greater than L_2, the sampling is stopped; and the hypothesis is accepted. It has not been proven that the probability that t is within the calculated interval will ever become less than L_1 or greater than L_2, and an attempt to get too high a significance level, L, may result in a sample sequence that never terminates.

The affect of a change in one of the parameters or a second parameter when the third is held fixed provides a basis for investigating the effect of an increase in the required accuracy of the tolerance interval or significance level. Consider, first, the case in which the level of significance is held fixed. The level of significance determines the value of t and the tolerance limits are then given by

$$\bar{x} - \bar{X} = c/N^{1/2} \qquad (c = ts = \text{constant}) \tag{12}$$

The tolerance limits are seen to be inversely proportional to the square root of the sample size for a fixed significance level. In order to halve the tolerance interval the sample size should be increased by a factor of four, for example. Of course, the relationship is not that simple because x is related to N in a random fashion, but the above relationship still indicates the order of magnitude of the sample increase as a function of the decrease in tolerance limits. The effects of the significance level on the tolerance limits and sample size are not so obvious. The effect of the significance level on the tolerance interval can best be seen from a plot of the cumulative distribution of t for the fixed sample size (figure 9). Since the tolerance interval is directly proportional to t, a plot of the tolerance interval against the significance level has the same shape as the t distribution. The important part of the curve is that part that corresponds to high significance levels. For example, to halve the probability of an error from a significance level of .95 requires that the tolerance interval be increased by a factor of about .19, assuming a sample size of 40. Thus, small changes in the required tolerance interval may cause large changes in the probability of making an error. The square root of the sample size is also proportional to t so that the probability of making an error (1 - L) is even more sensitive to changes in N than to changes in the tolerance limits for high levels of significance.

A more realistic example of the interaction of the three parameters may be found by actually calculating the tolerance limits of the mean assay value of a deposit for varying sample sizes. Estimates of the mean of the iron content of the deposit described in Appendix A were computed for four sample sizes with the following results:

Sample Size	Estimated Mean
320	$60.44 \pm .25$
622	$58.76 \pm .24$
812	$59.18 \pm .184$
1103	$58.96 \pm .176$

The last three samples are seen to be converging to a value of about 59.00 at a rate that is indicated by the tolerances, but the first value is obviously incorrect. The drill holes used in the first sample were the holes used as a preliminary sample (holes with numbers from 1 to 211 on the map in Appendix A), and they may not be a representative sample of the deposit. It may be noticed that the ratio of the size of the second to the size of the fourth samples is slightly under a factor of two, and the corresponding change in the tolerance limits is slightly under the square root of two. Thus, the experimental results are quite close to the theoretical results for a change in the sample size, and the estimated average seems to be converging to a value of about 59.

VI. SEQUENTIAL ANALYSIS

In the last twenty years, a statistical procedure known as "Sequential Analysis," designed to be used for testing hypotheses concerning one or more parameters from a distribution of a known form, has been developed.[10] The methods of Sequential Analysis have several advantages over most other procedures for the type of statistical analysis needed for geologic samples. The desired probabilities of errors of the first and second kind enter into the procedure explicitly; the procedure is, as the name implies, a sequential sampling procedure; the procedure has required as few as half the number of samples needed by other methods of analysis for the same statistical accuracy; and the amount of information that can be used in this test is quite flexible after the minimum amount of information for using the procedure is available. Obviously, there must be a limit to the statistical accuracy that can be obtained for a given amount of information or to the minimum number of samples that must be taken to reach a given level of significance. The savings in the number of samples used by the Sequential Analysis procedures may be due, in part, to the ability of the test to use a more flexible amount of information than many other tests by allowing all of the known parameters of a distribution to be used in estimating the unknown parameters.

To use the Sequential Analysis procedure, the two constants

$$A = \frac{\beta}{1 - \alpha}$$

and

$$B = \frac{1 - \beta}{\alpha}$$

where

α = the probability of an error of the first kind

β = the probability of an error of the second kind

are computed. Next, the values of the "sequential probability ratio," P_1/P_0 (where P_0 is the probability of getting the latest sample if the hypothesis is true and P_1 is the probability of getting the latest sample if the hypothesis is false), is computed for the latest sample.

If

$P_1/P_0 < A$, the hypothesis is accepted.

$A < P_1/P_0 < B$, the sample size is increased and the new P_1/P_0 is tested.

$B < P_1/P_0$, the hypothesis is rejected.

The Application of Sequential Analysis Based on the Distribution of the Sample Universe.

Unless the distribution of the universe being sampled is known in analytic form (and even in some cases when it is known) the calculation of P_0 and P_1 may be quite difficult. The calculation of the sequential probability ratio for a sample from a geologic deposit, for example, could be computed as follows:

Let M = number of possible assay values,

R_k = k'th possible assay value ($k = 1$ to M),

$P(R_k)$ = percentage of the total number of sample units in the deposit with a value R_k,

N = Sample size.

The probability of getting any given sample of N sample units

$R_k^{(1)}, R_k^{(2)}, \ldots R_k^{(N)}$ (each k may take any value from 1 to M)

is of the form

$$P^* = \left[\frac{N!}{N(R_1)!\ N(R_2)!\ \ldots\ N(R_M)!}\right] P(R_1)^{N(R_1)}\ P(R_2)^{N(R_2)} \ldots P(R_M)^{N(R_M)} \tag{13}$$

where $N(R_k)$ = number of times R_k appears in the sample.

The hypothesis to be tested is in the form of stated limits on the M values of the probabilities of the M possible assay values. The probability P_1 is, now, the sum of probabilities of the form of Eq. (13) for all possible combinations of the $P(R_k)$'s that will make the hypothesis false; and the probability P_0 is the sum of the probabilities of the form of Eq. (13) for all possible combinations of the $P(R_k)$'s for which the hypothesis is true. Thus, the calculations involved in computing the sequential probability ratio at each step of the sample would involve computing the multinomial probability stated above for all M possible values of the N percentages $P(R_k)$, a total of M^N times in all. If the assay values were carried only to the nearest .1 percent, M would be 1000; and the sample size would surely be greater than, for example, 20 and would increase at each step. Thus, on the order of 1000^{20} computations of Eq. (13) would be required at each step. Even using the fastest electronic calculators in existence this would take on the order of 10^{15} years.

These calculations could be simplified by placing restrictions on the allowable values of the $P(R_k)$'s and by making approximations in Eq. (13). However, until more dependable information is available about the distributions of geologic samples the use of Sequential Analysis procedures based on the distribution of the sample will be of no practical value.

Application of Sequential Analysis Based on the "t" Distribution.

A more practical approach to the use of the procedure of Sequential Analysis may be to use a combination of multiple and sequential sampling techniques. If the hypothesis tested is in the form of a statement that the average of the sample means for samples of size N is within a certain interval and the sample size, N, is large enough that the sample mean can be assumed to have a "t" distribution, the sequential probability ratio may be relatively easy to calculate. When the test indicates that the sample being tested should be larger, another sample of size N must be taken; and its mean is included in the computation of the next sequential probability ratio. This approach may not be of much value if many sample units are needed because each sample unit is the mean of a geologic sample of size N, and N must be large enough that the "t" distribution is valid. Thus, the number of geologic samples needed increases rapidly with each step of the analysis and final number of geologic samples may be much higher for this sequential test than for a fixed sample method such as the one described in Section V. On the other hand, this method may terminate rapidly and may prove to be extremely efficient; it certainly merits a more thorough study.

VII. SUMMARY

Geologic sampling must serve many purposes, and the procedures of mathematical statistics can be applied to only a few of them. However, when statistical procedures are applicable, they should provide reliable and efficient techniques for getting as much information as possible from as few samples as possible.

Statistical methods cannot be applied to any collection of samples without regard to their origin. If statistical concepts are to be used in analysing the samples, they must also be used in collecting and weighting them in order that the samples will be statistically representative of the deposit from which they are taken.

In order to get the most accurate and efficient results from a statistical analysis all available valid information about the distribution of the deposit being sampled should be incorporated into the analysis. The procedures discussed in this paper were based on the rather sketchy and incomplete knowledge now available concerning the nature of the distribution from which samples of geologic deposits are drawn. As our understanding of the statistical properties of mineral deposits becomes more complete, more efficient procedures of statistical analysis will become applicable to geologic problems. The procedures which seem to offer the most valuable potentialities are those that allow sequential sampling of the deposit being analysed, and the most important of these is the method known as "Sequential Analysis."

Table I

Calculations for test of randomness of samples from north to south on Line 500W of the deposit described in Appendix A.

Hole Number	Average Assay Value	$\bar{x}$ - x	Sign of x-Median
543	60.60	+ .2	-
552	62.04	-1.24	+
549	62.65	-1.85	+
468	60.13	+ .67	-
460	63.90	-3.10	+
416	61.37	- .77	+
415	59.12	+1.68	-
1B	61.76	- .96	+
13	60.17	+ .63	-
538	56.29	+4.51	-

Table II

Tables for the test of randomness of a series at a significance level of .95 based on the length of the longest run in the series.

Size of Sample	Minimum Length of Run for an Indication of a Non-Random Series.
10	5
20	7
30	8
40	9
50	10

Table III

Calculations for estimating the average assay value of the deposit in Appendix A from a sample of composed alternate drill holes in the north-south lines.

Hole Number	Average Value of Hole (x)	Number of Sample Units in Hole (f)	$(x - \bar{x})^2$
26	62.42	20	13.396
555	49.91	20	78.323
453	61.25	30	6.200
452	60.40	20	2.69
546	58.29	25	.221
451	60.57	34	3.276
419	61.51	15	7.563
538	56.29	29	6.101
1B	61.76	35	9.000
11	61.79	17	9.181
4	63.52	6	22.658
519	59.99	33	1.513
6	59.17	37	.168
409	57.96	40	.640
903	56.68	31	4.326
413	57.22	13	2.372
493	55.57	16	10.176
533	53.58	21	26.832
461	63.00	15	17.978
463	62.26	14	12.250
460	63.90	1	26.420
549	62.65	12	15.132
543	60.60	10	3.386
537	56.91	18	3.423

Table III (continued)

Hole Number	Average Value of Hole (x)	Number of Sample Units in Hole (f)	$(x - \bar{x})^2$
547	56.03	4	7.453
558	62.02	5	10.628
467	61.71	8	8.703
544	61.65	3	8.352
534	54.71	19	16.403
900	58.53	22	.053
531	54.28	19	16.403
505	55.92	10	8.066
4	59.02	10	.068
3	60.54	10	3.168

$$\bar{x} = \frac{\Sigma fx}{\Sigma f} = 58.759$$

$$s^2 = \frac{\Sigma(x - \bar{x})^2 f}{\Sigma f} = 9.325$$

$$N^{1/2} = (\Sigma f)^{1/2} = 24.940$$

Estimated average assay value $= \bar{x} \pm 1.96\,(s/N^{1/2}) = 58.759 \pm .240$.

FIGURES

Figure 1.

Distribution of copper in a contact metamorphic deposit.

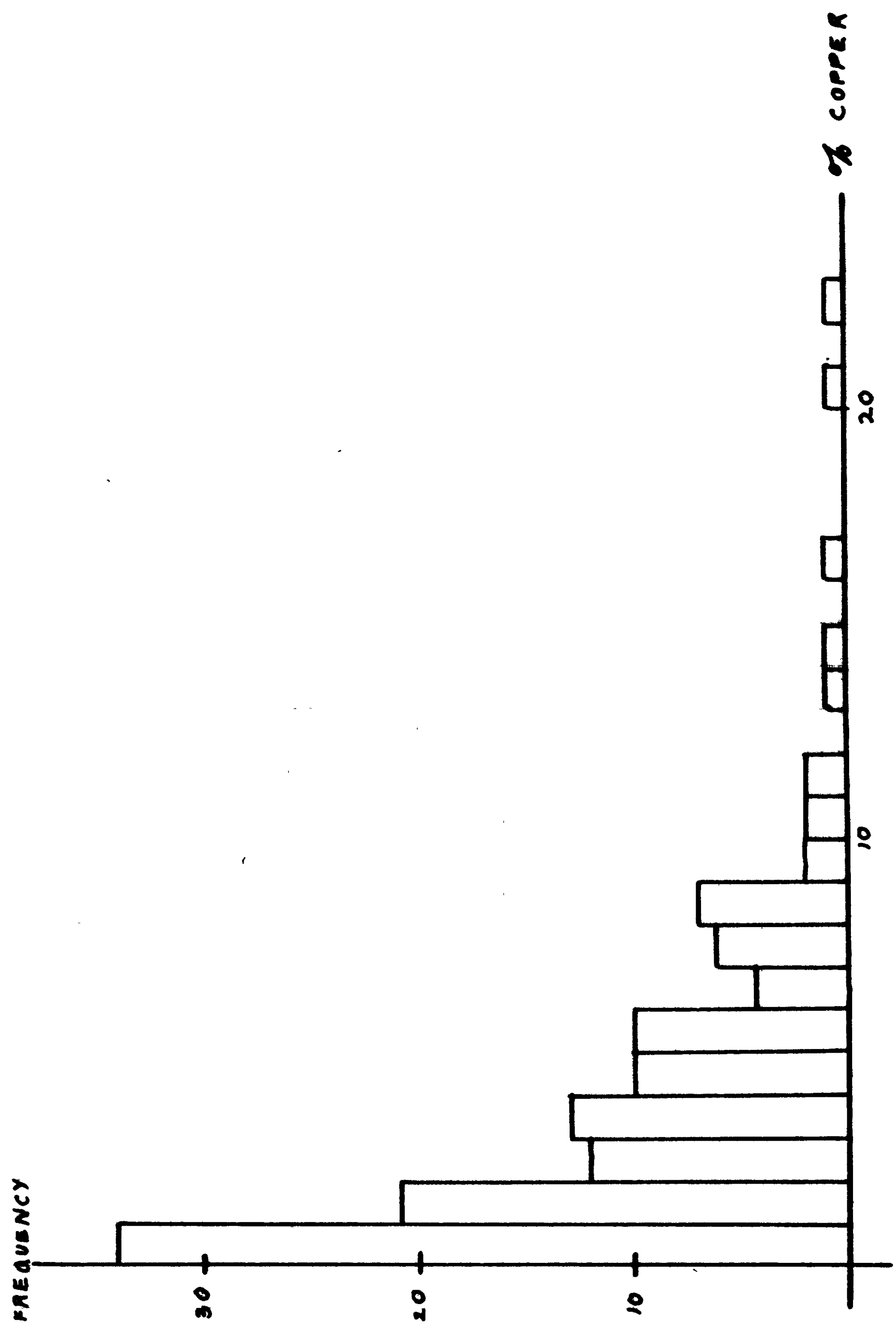
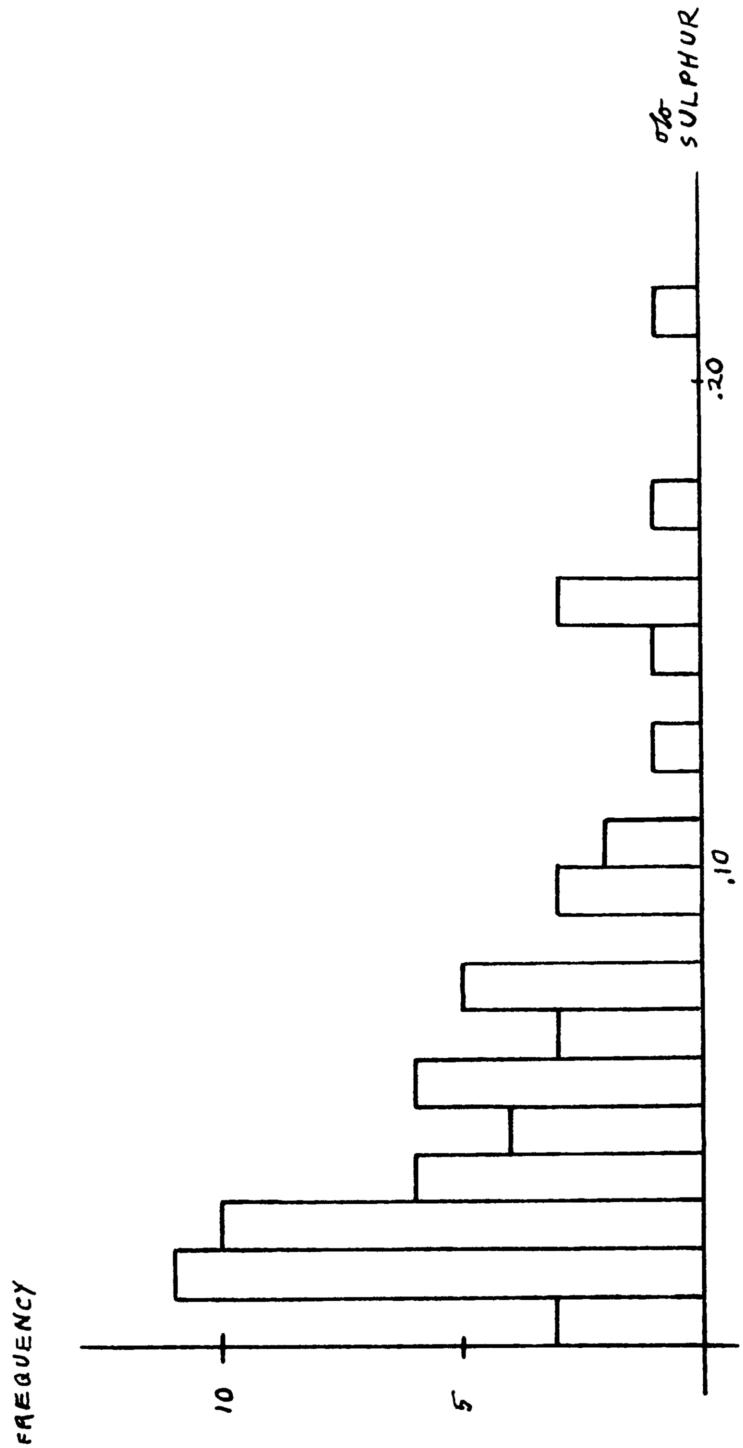


Figure 2.

Distribution of sulphur in a lateritic deposit.



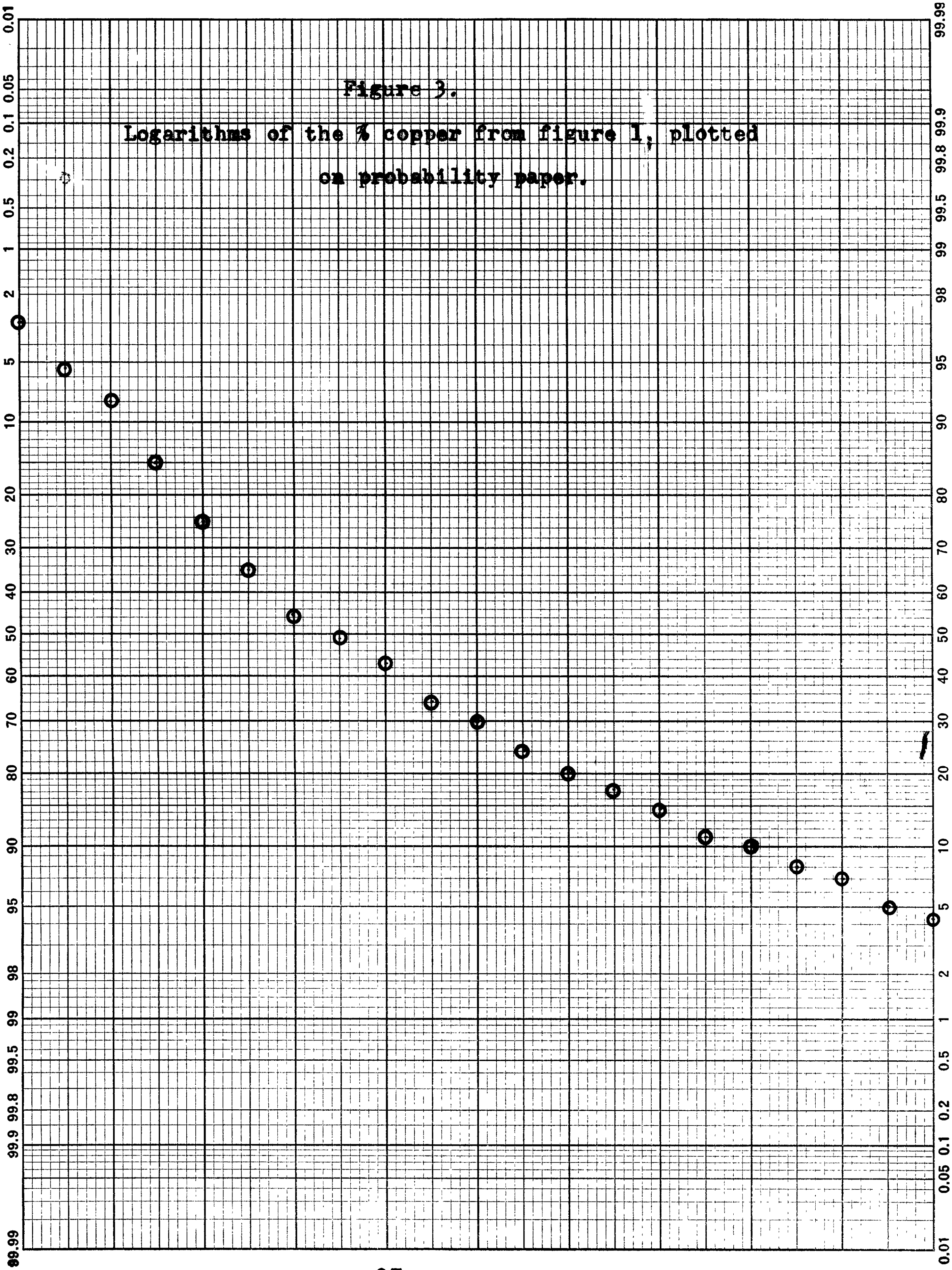


Figure 3.

Logarithms of the % copper from figure 1, plotted on probability paper.

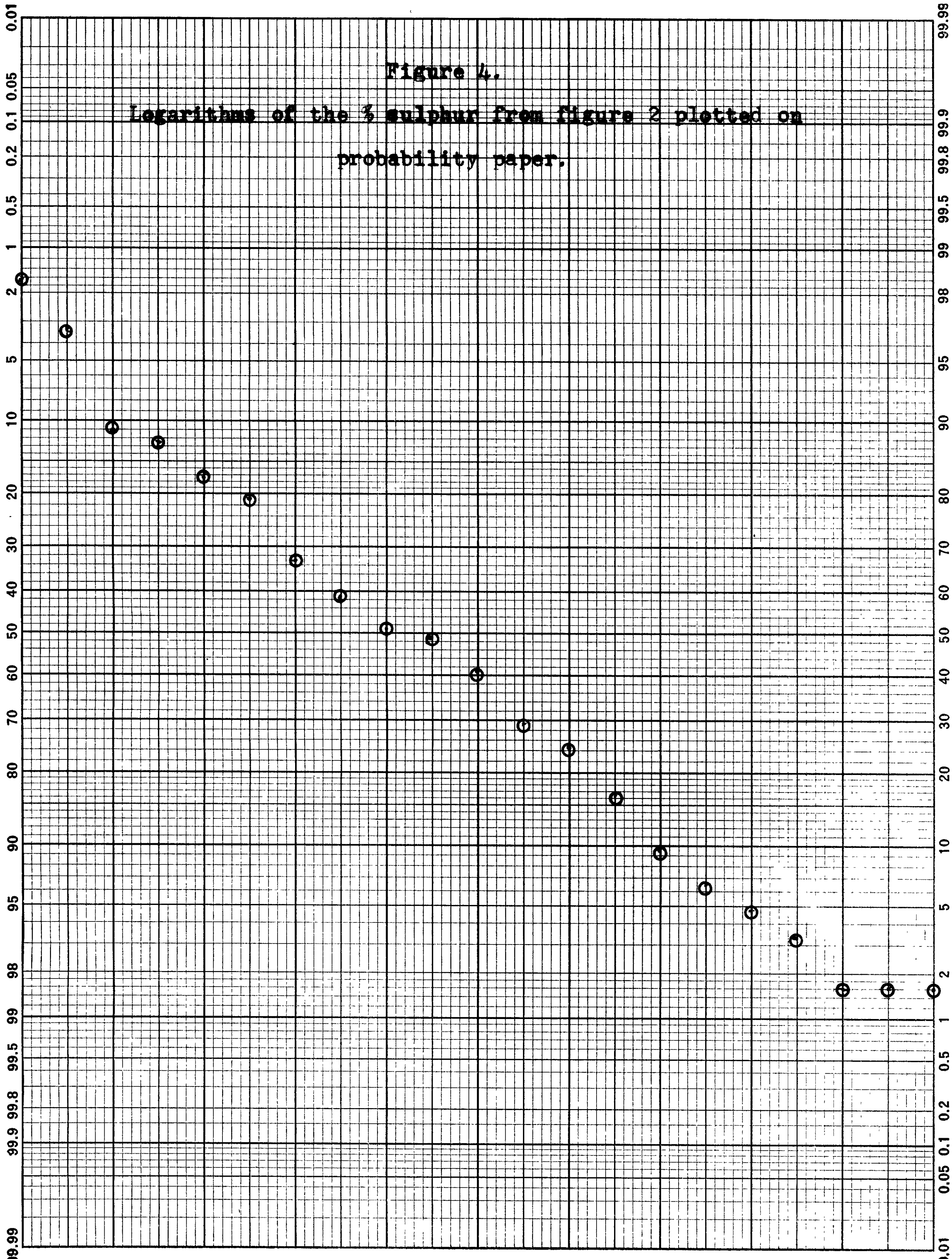
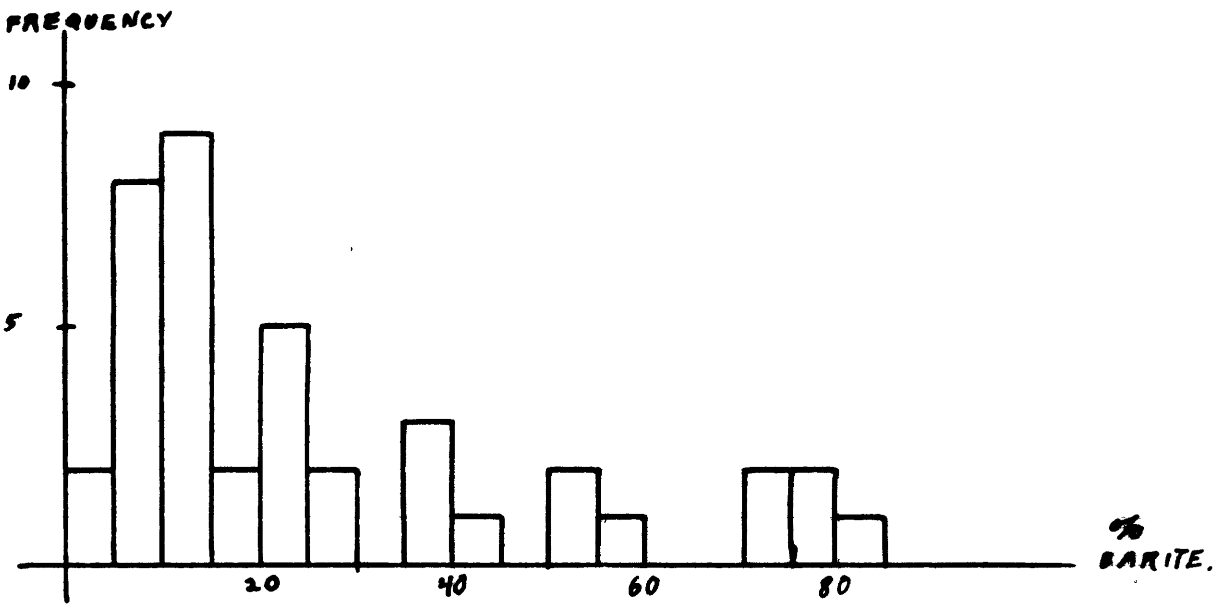


Figure 4.

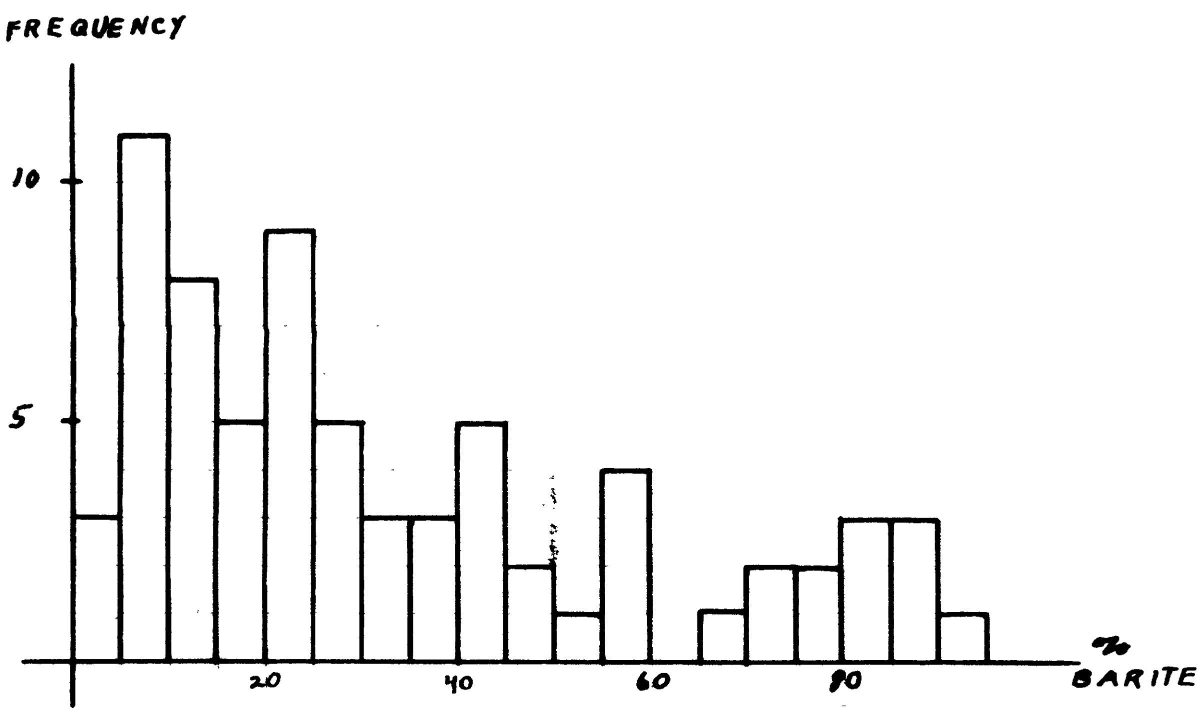
Logarithms of the % sulphur from figure 2 plotted on probability paper.

Figure 5.

Distributions of trench and drill samples from a replacement deposit of barite.



Drill Samples



Trench Samples

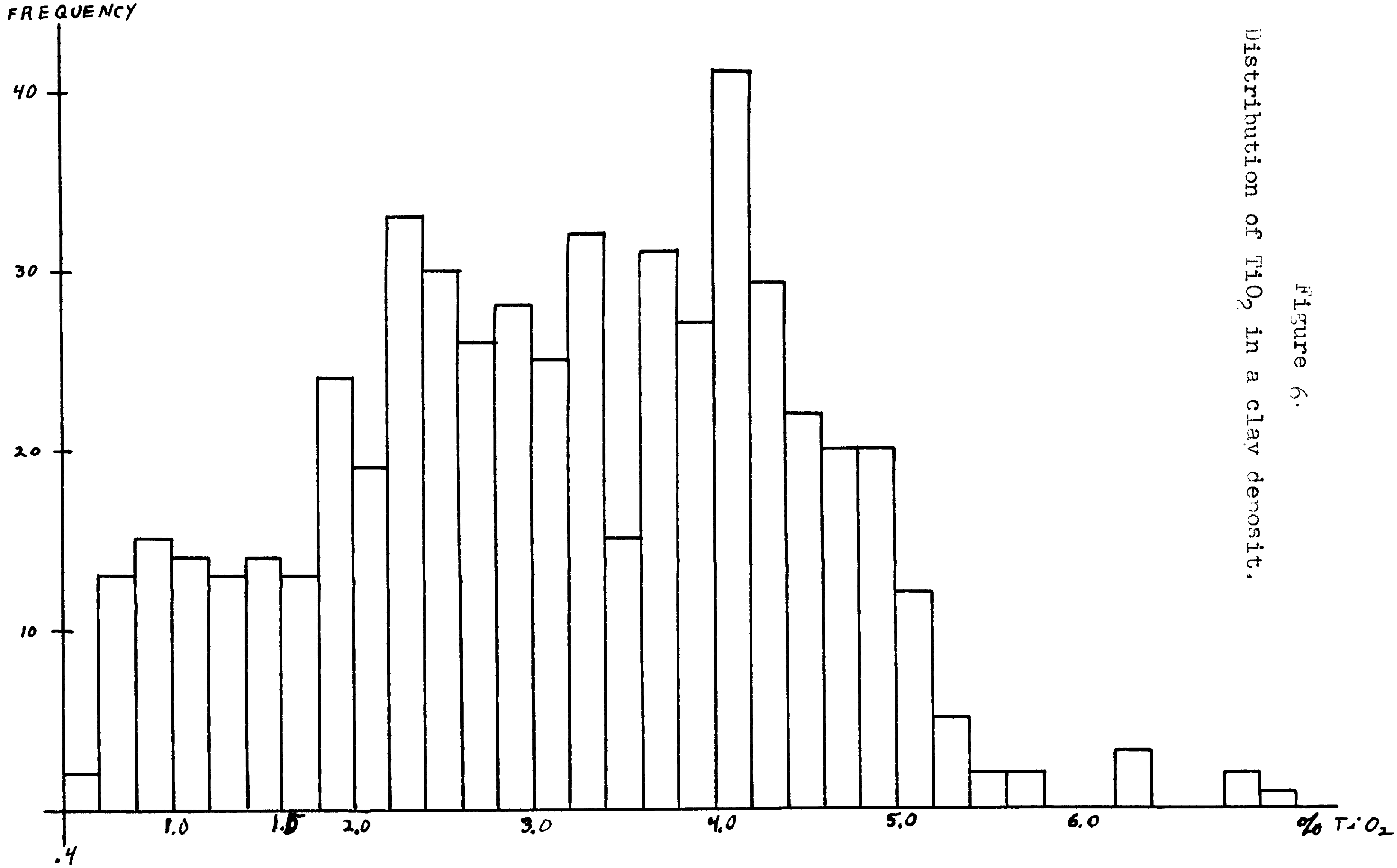
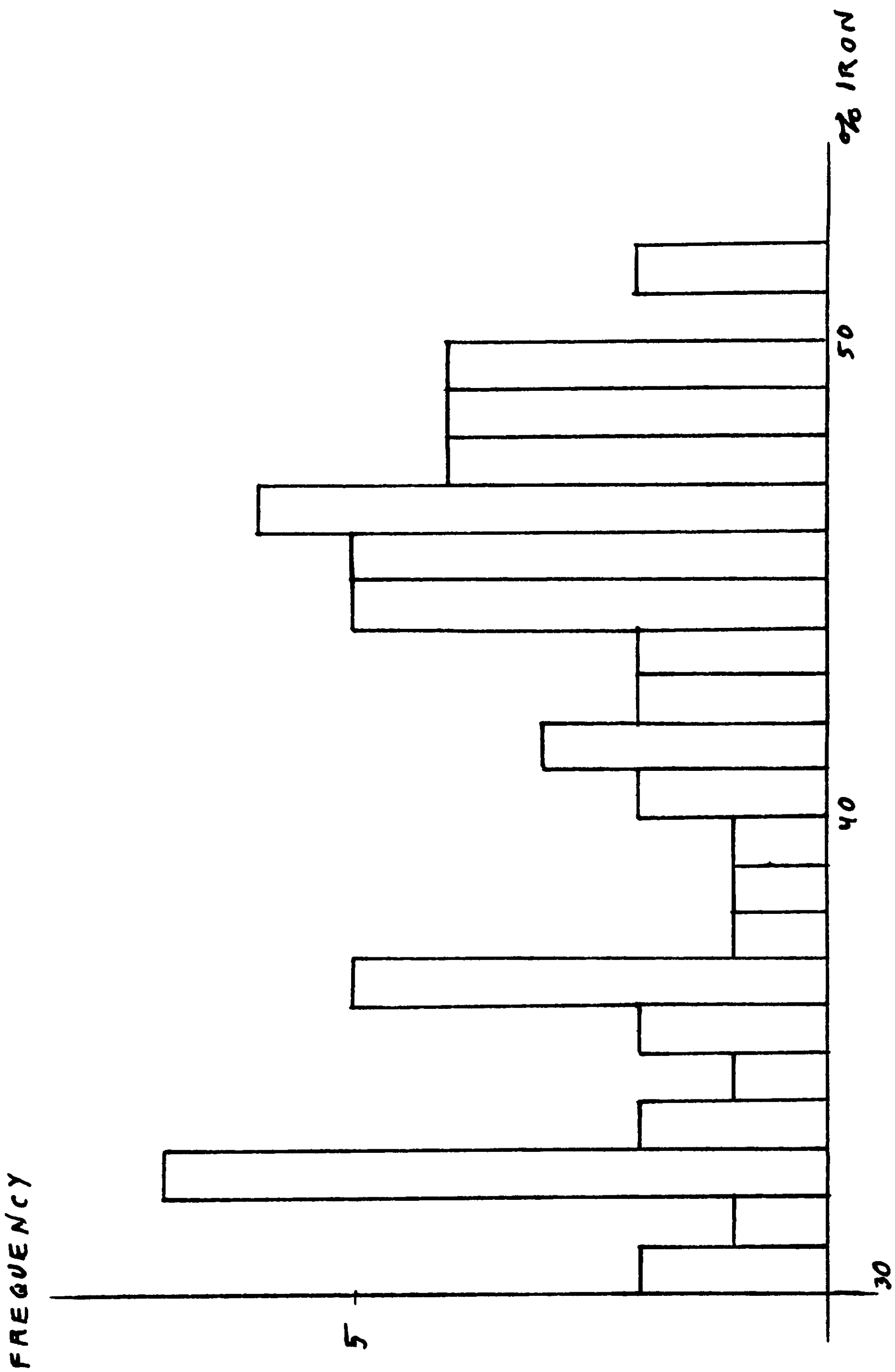


Figure 6.
Distribution of TiO_2 in a clay deposit.

Figure 7.

Distribution of iron in a lateritic deposit of iron and nickel.



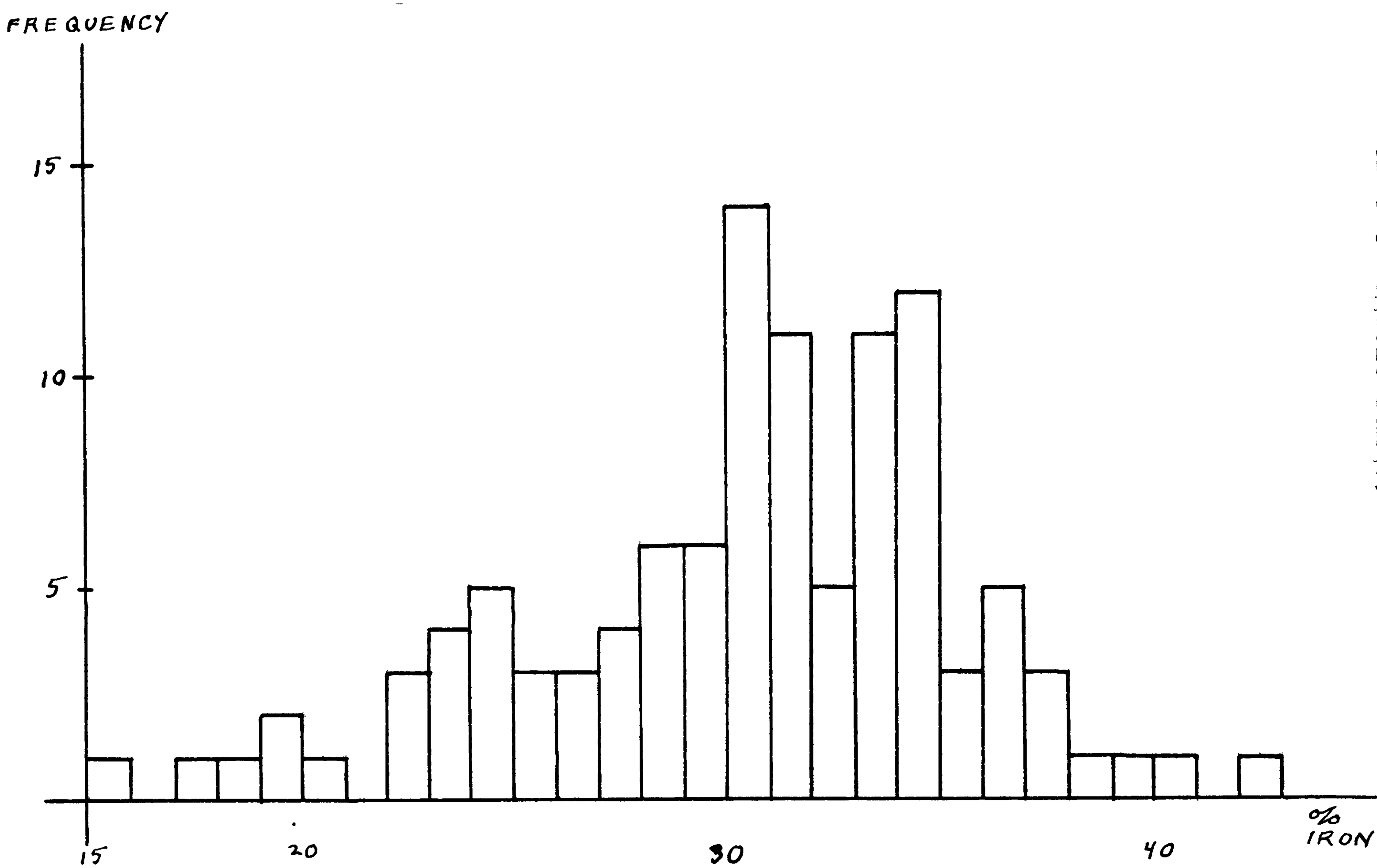
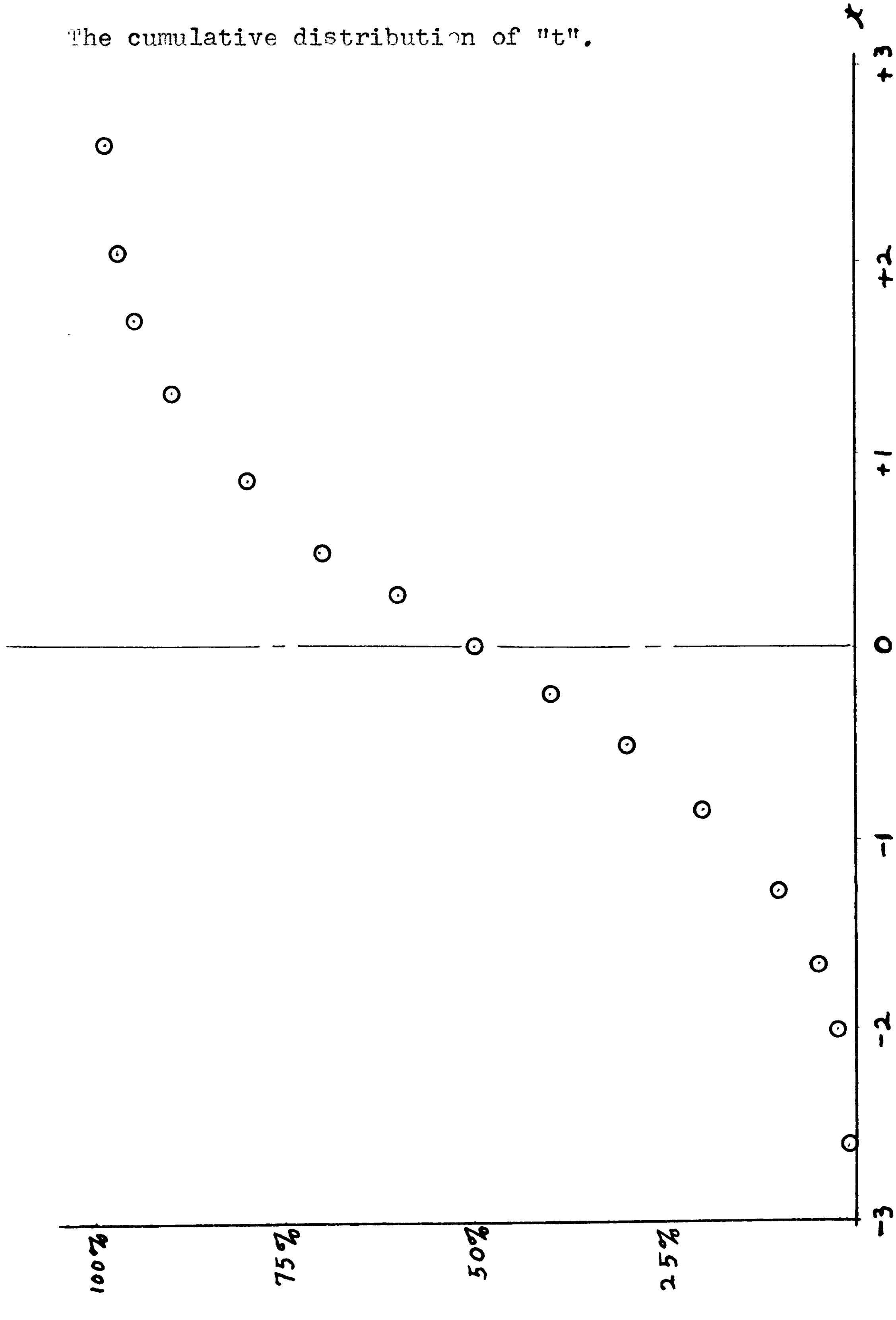


Figure 8.
Distribution of iron in a laterite in the western part of the Gogebic range.

Figure 9.

The cumulative distribution of "t".



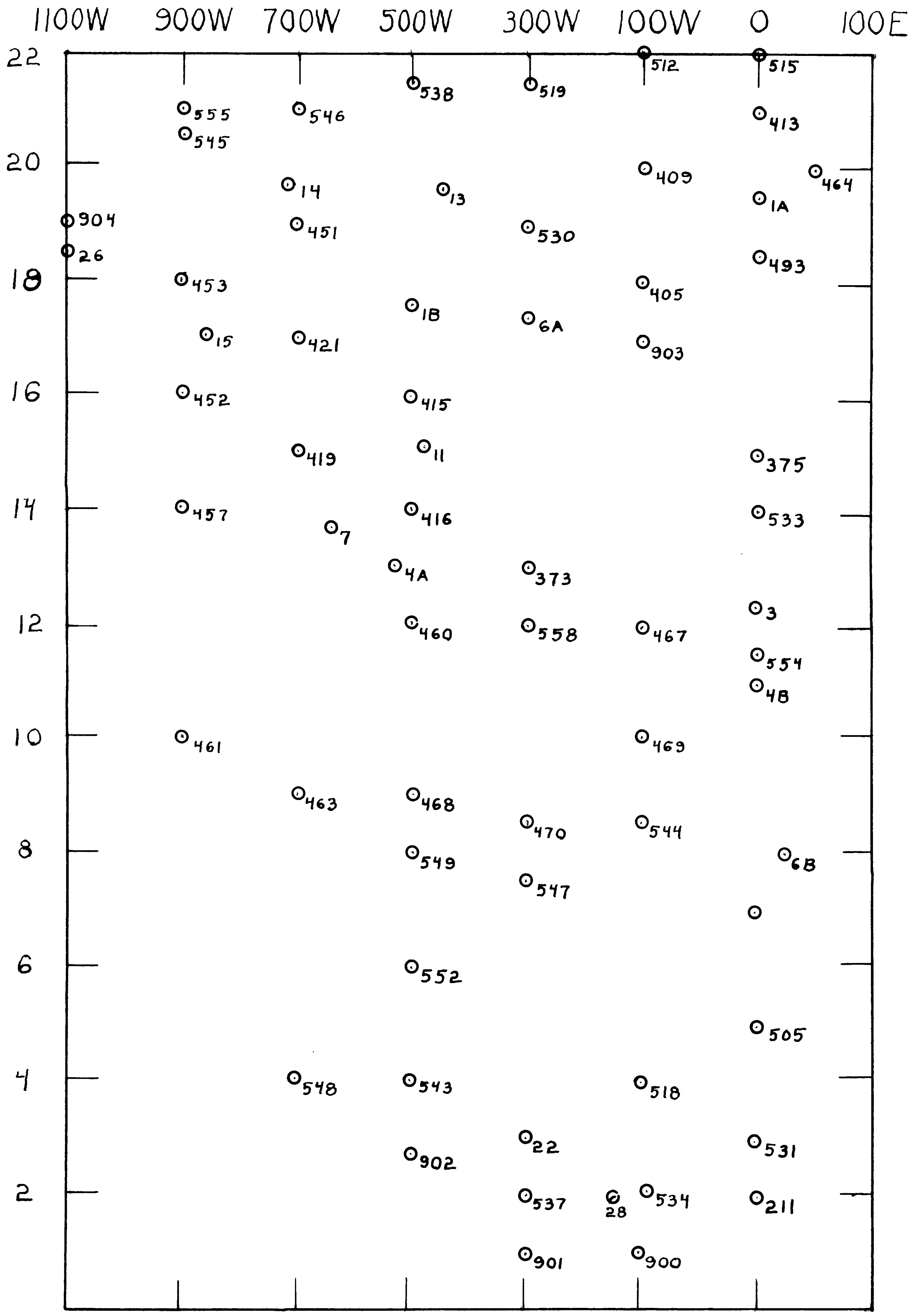


Figure 10

Sample pattern for the iron deposit used as an example in the text.

Figure 11.

Frequency distribution (histogram) for the iron deposit used as the standard example in the text.

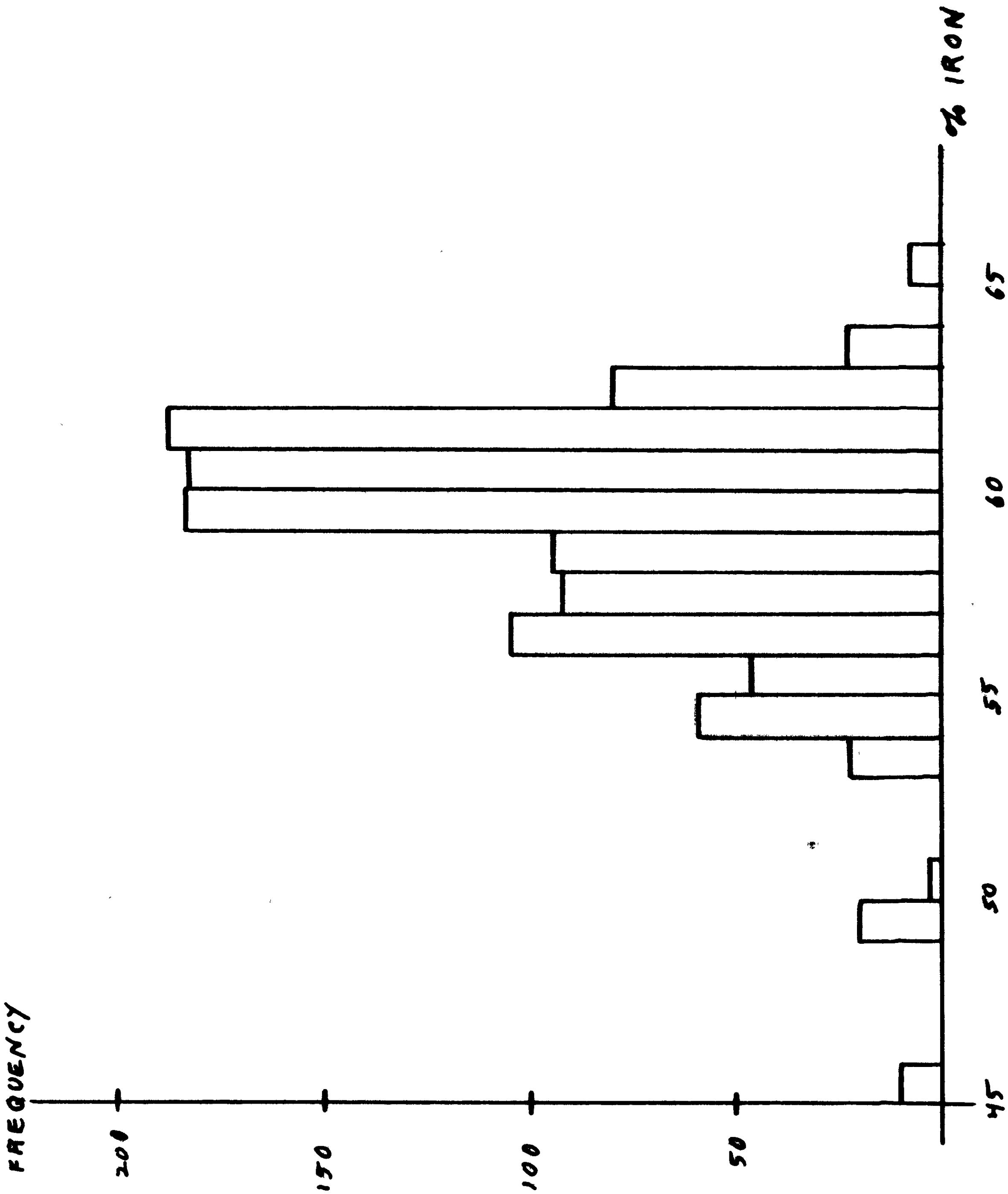
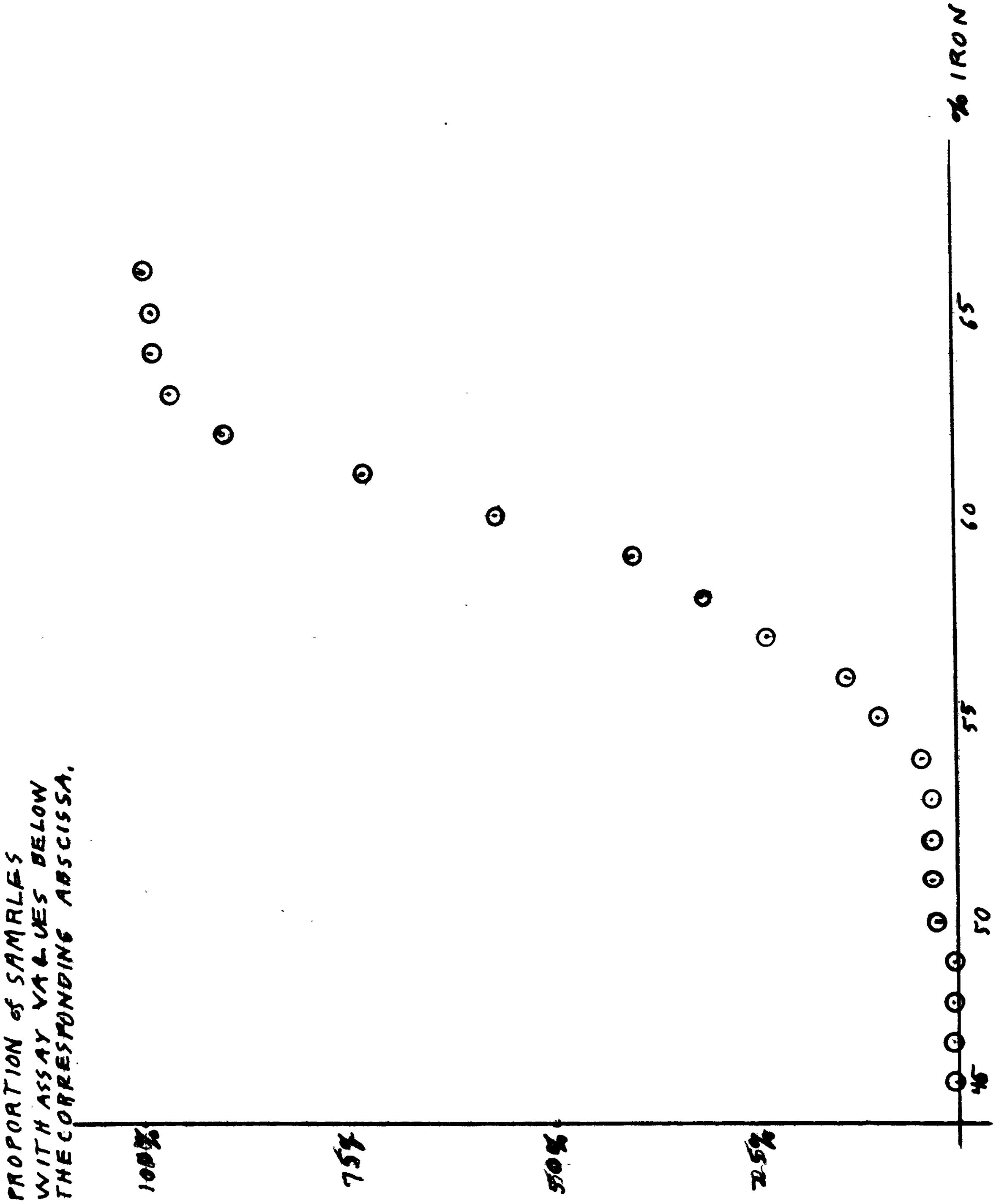


Figure 12.

Cumulative distribution of the iron deposit used as a standard example in the text.



APPENDIX A

All examples used in this paper have been taken from one lateritic iron deposit. Table IV describes the average iron content and the length of each of the sample holes in the ore. Since this paper is not concerned with determining cut-off limits for mining the deposit, only the holes in the ore are described. Figure 10 is a map of the drill hole pattern, and Figures 11 and 12 are, respectively, the frequency distribution and cumulative distribution of the samples.

Table IV

Assay Values from an Iron Deposit

Hole Number	Average Assay Values in Percent	Number of Five Foot Sample Units in Hole
1A	60.15	16
1B	61.76	35
3	60.54	10
4A	63.52	6
4B	59.02	10
6A	59.17	37
6B	59.31	9
7	65.18	8
11	61.79	17
13	60.17	40
14	61.14	30
15	62.23	23
22	60.03	16
26	62.42	20
28	60.41	22
211	54.04	21
373	61.20	3
375	59.85	15
405	57.16	38
409	57.96	40
413	57.22	13
415	59.12	22
416	61.37	2
419	61.51	15
421	61.60	24
451	60.57	34
452	60.40	20
453	61.25	30
457	61.32	16
460	63.90	1
461	63.00	15
463	62.26	14
464	59.29	13
467	61.71	8
468	60.13	5
469	59.23	8
470	50.48	3
493	55.57	16
505	55.92	10
512	59.39	24
514	56.82	10

Table IV (continued)

Hole Number	Average Assay Values in Percent	Number of Five Foot Sample Units in Hole
515	55.88	10
518	59.13	12
519	59.99	33
530	58.90	39
531	54.28	19
533	53.58	21
534	54.71	19
537	56.91	18
538	56.29	29
543	60.60	10
544	61.65	3
545	56.62	13
546	58.29	25
547	56.03	4
548	61.69	5
549	62.65	12
552	62.04	6
554	58.98	9
555	49.91	20
558	62.02	5
900	58.53	21
901	55.69	10
902	60.06	9
903	56.68	31
904	45.77	10

The Normal and Lognormal Distributions.

The frequency distribution of a normally distributed variate, x, is defined as

$$Y(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

where

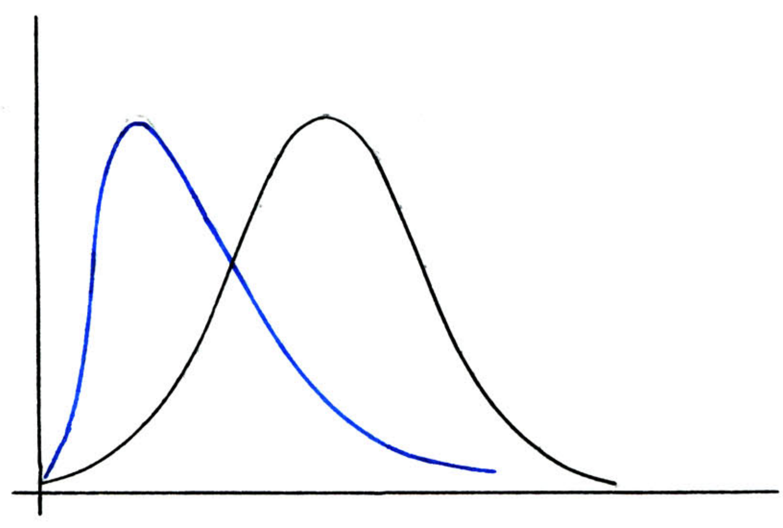
σ = the standard deviation

μ = the mean of the distribution

Y = the frequency of the value, x.

This curve is always symmetric, but its shape will vary for different values of σ . The black curve in the figure below is a normal frequency distribution.

The lognormal distribution is a distribution in which the log of the variate is normally distributed, and is always positively skewed. The blue curve in the figure below is a lognormal frequency distribution.



References

1. Hoel, Paul G., Introduction to Mathematical Statistics, John Wiley and Sons, New York, pp. 180-181.

2. Swed, F. and Eisenhart, C., "Tables for Testing Randomness of Grouping in a Sequence of Alternatives," Annals of Mathematical Statistics, Vol. XIV (1943), pp. 66-87.

3. Mosteller, F., "Note on Application of Runs to Quality Control Charts," Annals of Mathematical Statistics, Vol. XII (1941), pp. 228-232.

4. Ahrens, L. H., "The Lognormal Distribution of the Elements," Geochimica et Cosmochimica Acta, Vol. V (1954), pp. 49-73.

5. United States Bureau of Mines Reports of Investigations Numbers 3900, 3971, 4155, 4189, 4214, 4349, and 4503.

6. DeKock, W. P., "The Carbon Leader of the Far West Rand With Special Reference to the Proportion of the Gold Content of This Horizon Recovered in Drilling," Transactions of the Geological Society of South Africa, Vol. L1 (1949), pp. 213-247.

7. Hoel, Paul G., Introduction to Mathematical Statistics, John Wiley and Sons, New York, pg. 68.

8. Ibid., pp. 35-36.

9. Dixon, Wilfred J. and Massey, Frank J., Introduction to Statistical Analysis, McGraw-Hill Book Company, Inc., New York, pp. 97-101, 108.

10. Wald, A., Sequential Analysis, John Wiley and Sons, New York.

Bibliography

1. Ahrens, L. H., "The Lognormal Distribution of the Elements," Geochimica et Cosmochimica Acta, Vol. V (1954), pp. 49-73.

2. DeKock, W. P., "The Carbon Leader of the Far West Rand With Special Reference to the Proportion of the Gold Content of This Horizon Recovered in Drilling," Transactions of the Geological Society of South Africa, Vol. L I (1949), pp. 213-247.

3. Dixon, Wilfred J. and Massey, Frank J., Introduction to Statistical Analysis, McGraw-Hill Book Company, Inc., New York, 1951.

4. Feller, William, An Introduction to Probability Theory and its Applications, Vol. I, John Wiley and Sons, Inc., New York, 1950.

5. Hoel, Paul G., Introduction to Mathematical Statistics, John Wiley and Sons, Inc., New York, 1947.

6. Mosteller, F., "Note on Application of Runs to Quality Control," Annals of Mathematical Statistics, Vol. XII (1941), pp. 228-232.

7. Parks, Roland D., Examination and Valuation of Mineral Property, Addison-Wesley Press, Inc., 1949, 3rd edition.

8. Statistical Research Group, Sequential Analysis, Columbia University Press, 1945.

9. Swanson, C. O., "Probability in Estimating the Grade of Gold Deposits," Transactions of the Canadian Institute of Mining and Metallurgy, Vol. 48 (1945), pp. 323-335.

10. Swed, F. and Eisenhart, C., "Tables for Testing Randomness of Grouping in a Sequence of Alternatives," Annals of Mathematical Statistics, Vol. XIV (1943), pp. 66-87.

11. United States Bureau of Mines Reports of Investigations Numbers 3900, 3971, 4155, 4189, 4214, 4349, and 4503.

12. Wald, A., Sequential Analysis, John Wiley and Sons, Inc., New York.