DANIEL F. MERRIAM, Editor

SAMPLING A GEOLOGICAL POPULATION

By

JOHN C. GRIFFITHS
Pennsylvania State University

and

CHARLES W. ONDRICK
Kansas Geological Survey

in cooperation with the
American Association of Petroleum Geologists
Tulsa, Oklahoma

COMPUTER CONTRIBUTION 30
State Geological Survey
The University of Kansas, Lawrence
1968

Editor's Remarks

The following remarks are the author's preface to our "Workshop on experiment in sampling." The workshop is the fifth in a series of colloquia on Computer Applications in the Earth Sciences. The colloquium is sponsored by the Geological Survey, U.S. Geological Survey, Water Resource Branch (Kansas District), and University Extension. We would like to thank all people concerned for help in preparation of the meeting.

It is recognized generally that effective teaching of a science subject is achieved by having students perform experiments which illustrate the principles and laws upon which the subject is based. In the geosciences the subject is the earth and most experiments with earth materials rely heavily on sampling. This report outlines a series of experiments in sampling common materials of the earth's crust; the sampling and measurements are performed in the field, a feature which caters to the emphasis on field work that characterizes many geological investigations.

Sampling of a gravel pit has the "simple" objective of measuring the size and shape of some of the pebbles it contains; it is not difficult to alter the location either in the field or from the field to the laboratory. Again, it is possible also to change the variate from measurement of pebble size to counts of pebbles and counting may be performed in the field or laboratory. Thus, these sampling experiments include a wide range of potential studies--the procedure is invariant to change of scale over orders of magnitude of fractions of a millimeter to kilometers.

The original experiments commenced in 1951 with measurement of size of quartz grains from the Homewood Quartzite (Pennsylvanian of Pennsylvania) and studies of the Montoursville gravel, Montoursville, Pennsylvania, a fluvioglacial terrace gravel in the valley of the Susquehanna River near Williamsport, Pennsylvania, commenced in 1956 (with kind permission of the Lycoming Silica Sand Company, Montoursville, Pennsylvania). From 1956 the experiments have alternated each year between the Montoursville gravel and screes of Tuscarora Quartzite. The two rock types represent an approximation to the extremes of the sampling model (Griffiths, in Milner, 1962; Griffiths, 1967; and this report, Fig. 2), and recommendations resulting from the work cover a wide range of geological populations.

The account was composed as a rough draft syllabus for a graduate course in The Pennsylvania State University and was later completed in a syllabus for an American Geological Institute short course prior to the Annual Meeting of the Geological Society of America in San Francisco in November, 1966. The results of measuring size and shape of white felsite pebbles in the Elliott Plant gravel pit of the Pacific Cement and Aggregates Co. near Livermore, California are similar to those of the white quartzite and red sandstone pebbles from the Montoursville gravel, Pennsylvania.

The sampling experiments therefore are applicable to a wide range of materials and serve to illustrate the use of a number of statistical procedures such as the construction of frequency distributions, calculation of moment measures, tests for randomness, quality control graphs, analysis of variance, and regression and correlation analysis. In addition the recording of data on Port-a-Punch Cards in the field, data processing utilizing an electronic computer, and construction and use of different algorithms incorporated in the systems program also are exemplified. The experiments thus may be used to introduce the student to the use of statistics and computers in geological problems.

Objectives of the experiment are not by any means trivial because the variation in the outcomes of different sampling arrangements illustrate the problems involved in estimation of means and variances. The procedure is of general interest in many scientific fields and of considerable importance in many industrial fields such as in the estimation of reserves. The use of teams of operators serves to illustrate the effect of differences among investigators in performing experiments and the design requirements that must be met to ensure that differences among operators do not vitiate the finding of differences among materials.

The experiments completed to date yield general recommendations for sampling geological populations and are of value in research both in suggesting efficient sampling designs and in formulating an objective means of specifying the structure or pattern of variation of geological materials.

Success of the experiments has depended mainly on cooperation of students who have carried out the data gathering and analysis as part of their graduate course work; graduate assistants have helped with the experiments, and in particular, Dr. E.C. Dahlberg was invaluable in the Short Course experiment in San Francisco in 1966. Dr. Ondrick is responsible for compiling the separate steps into a single systems program which increases the efficiency of the analysis and represents an example of the use of an algorithm for problem solving in the geosciences.

The experiment that will be performed within this outline will be concerned with the measurement of size and shape of white (carbonate) pebbles. List and Clark Construction Co. allowed use of the Holliday glacial gravel pit near Zarah, Kansas.

John C. Griffiths and Charles W. Ondrick, December 1968

An up-to-date list of computer and related publications can be obtained by writing the Editor, COMPUTER CONTRIBUTION Series, Kansas Geological Survey, The University of Kansas, Lawrence, Kansas 66044.

SAMPLING A GEOLOGICAL POPULATION

by

John C. Griffiths and Charles W. Ondrick

INTRODUCTION

Statement of Problem

When the outcome of an experiment is important it possesses some tangible value. For example, the Bureau of Census was created to sample the (human) population of the United States for information required by government. Although this population is a sizeable unit, census taking is evidently a worthwhile aid to governmental decisions. Many private polls also supply--at a price--information on characteristics of the American people. The important principle here is that information is of value, and when it is, people will pay the necessary price. The population of the United States consists of only about 200 million individuals; the geologist is faced with the problem of sampling units of the earth's crust, each of which includes a larger population of elements. If geological information is valuable, it is worthwhile investing the necessary effort to ensure that is is adequately reliable for the purpose for which it is to be used.

No experiment is better than the constraints included in its sampling arrangement, therefore it is necessary to be deliberate in planning an experiment, and in particular, in planning its sampling design to fit the objective for which the experiment is performed (Steinmetz, 1962). In our experiment, a geological population is characterized by describing certain properties of its individual elements. The experiment illustrates different results that arise from various sampling arrangements. Essentially the objective is simple, that is to determine the mean and dispersion around the mean of some specified characteristics of the population. It is assumed that these two estimators are valuable items of information and both are <u>necessary</u> and <u>sufficient</u> in describing the specific <u>characteristics of the</u> population (Griffiths, 1961, 1964).

A number of constraints are imposed on the experiment by these apparently simple requirements. In order to fulfill the constraints, the procedure becomes elaborate (Fig. 1). All sampling experiments to determine "best estimators" of population parameters are among the most difficult and expensive experiments to complete successfully (Stephan and McCarthy, 1963). In order to achieve the objectives within specified limits it is necessary to be specific about each step in the procedure, and the steps are designed so that the data fulfill the constraints (Fig. 1).

Defining the Problem

The objectives entail determining <u>statistical estimators</u> of the required <u>population parameters</u>. Thus from a population and its <u>elements</u>, a <u>sample</u> of size "n" of the elements is <u>taken</u>. The <u>problem</u> of estimation of parameters of a detrital sediment is outlined by Griffiths (in press). Notation is noted in Table 1. It is essential that as the sample size "n" increases towards the population size "N", sample estimators ($\overline{X}$, $\hat{\sigma}^2$) converge on their respective population parameters (μ, σ^2).

Table 1.–Relationship between statistical estimators and population parameters.

Characteristic	Statistical Estimator	Population Parameter
Mean	$\overline{X}$	μ
Variance	$\hat{\sigma}^2$	σ^2
Sample Size	n	N

The statistical estimators are "<u>best estimators</u>" of their respective parameters if they <u>are consistent</u>, unbiased, efficient and sufficient (Fisher, 1948, p. 10). To obtain statistical estimators which fulfill the requirements, it is necessary to commence with random samples from known frequency distributions. Design of the experiment therefore is based essentially on the constraints. In effect, we have decided answers to our questions and worked backwards to the experimental design which will yield these answers. Most of the analysis is concerned with ensuring that the requirements are fulfilled adequately.

The second step is to select the characteristic parameters to estimate. Some parameters are defined as <u>fundamental</u> properties of populations (Griffiths, 1961, 1967) and the analysis may be extended simply to the remaining properties and to <u>derived</u> properties. The fundamental properties of a <u>population</u> are a function of composition, size, shape, orientation, and packing. Data are measurements or counts. <u>Measurements</u> arise if the variate exhibits a <u>continuous range of variation</u>, such as the properties of size and shape. <u>Counts</u> arise if the variate differs in discrete steps. <u>Experimental</u> determination of the proportions of elements of different types, or composition, generally

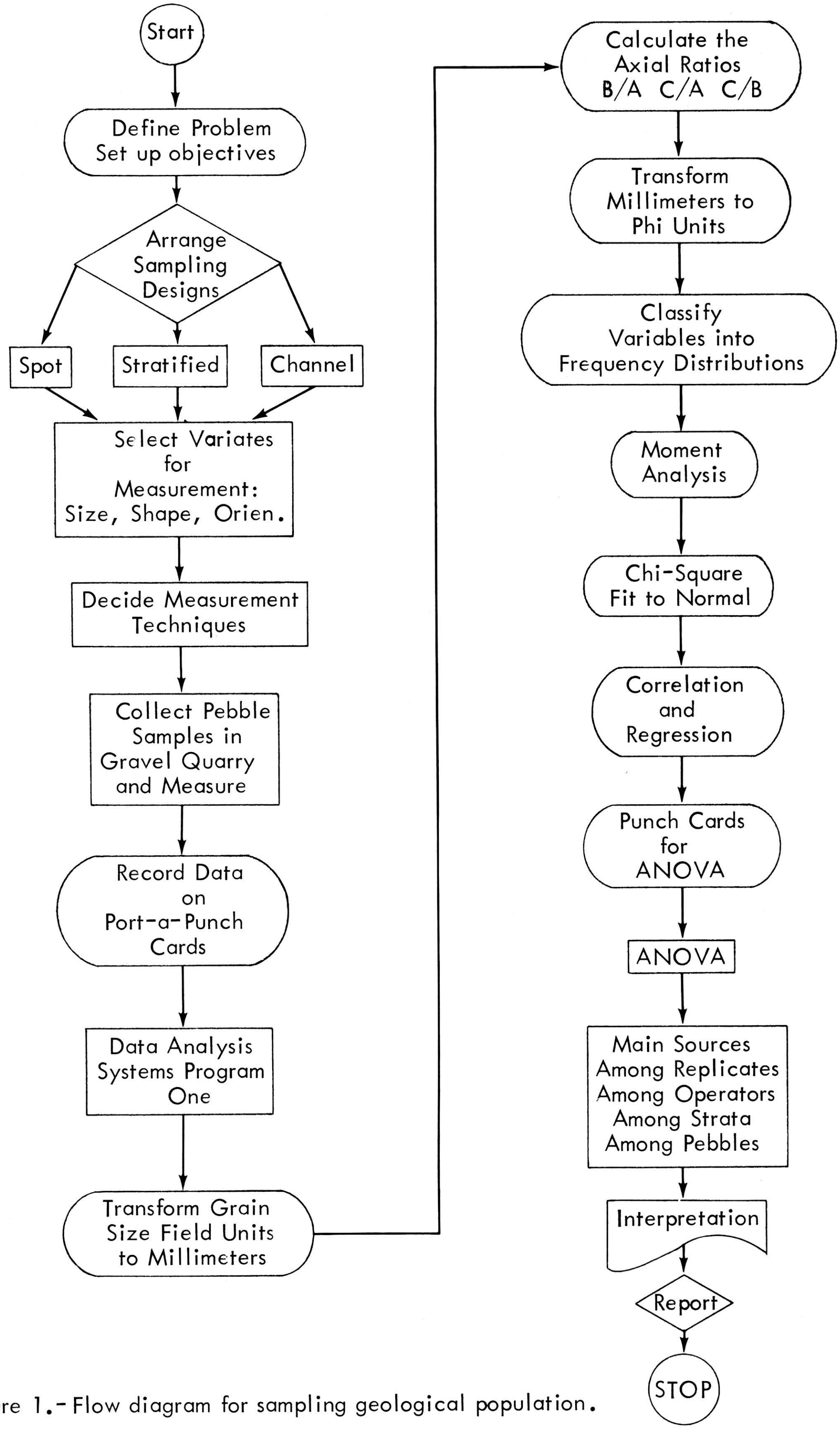


Figure 1.- Flow diagram for sampling geological population.

yields discrete data as does packing (Kahn, 1956; Miller and Kahn, 1962).

An important distinction between the two types of data arises essentially from their different frequency distributions. Continuous variates, if measured appropriately, exhibit normal frequency distributions. Count data results generally in binomial and Poisson distributions. These are called constant probability models. By suitable definition of the sampling design and measurement procedure, it is possible usually to achieve a constant probability model. Again, knowing the result, that is the constant probability model, it is possible to use tests of random sampling which permit adjustment of the sampling design so that a constant probability model is appropriate (Griffiths and Drew, 1966).

Now this is a somewhat elaborate and rigorously prescribed procedure to ensure outcome of the experiments are of the desired form. It may well be questioned whether a less elaborate procedure would not be adequate in a geological sense. Fortunately it is possible to use all the constraints as an interpretive base to yield conclusions of geological interest and there appear to be no other procedures which yield such unequivocal answers to geological questions. The desirable features ("best estimators" as described) are obtained by arranging an equivalence between the geological and statistical questions.

If for example, a variate is continuous and if arrangement of different values of the variate in the population is random, then random samples are simple to achieve and variate values will generate a normal distribution after suitable transformation. Geologically, this is equivalent to random sampling a population which possesses a massive structure, that is there is no discernible systematic arrangement of variation in this variate in the population.

On the other hand, suppose that the population is stratified or layered, then it is necessary to be more circumspect about the sampling. If it is unnecessary to describe the layering, channel samples are adequate to achieve the required estimators. Channel samples, if they are random, will be self-weighting and will yield unbiased best estimators. To ensure that the channel samples are unbiased, and self-weighting, each channel sample must cross the entire population. This can be achieved only if the entire population is accessible to the sampling procedure.

If it is necessary to define the layering as well, additional constraints are imposed on the sampling arrangement. It is necessary to ensure that the samples are confined to single layers, for example, sedimentation unit sampling is required (Otto, 1938). If samples transgress the layers the estimators will be biased and certainly will not be "best estimators." In fact the frequency distributions usually will, in this instance, not yield constant probability models.

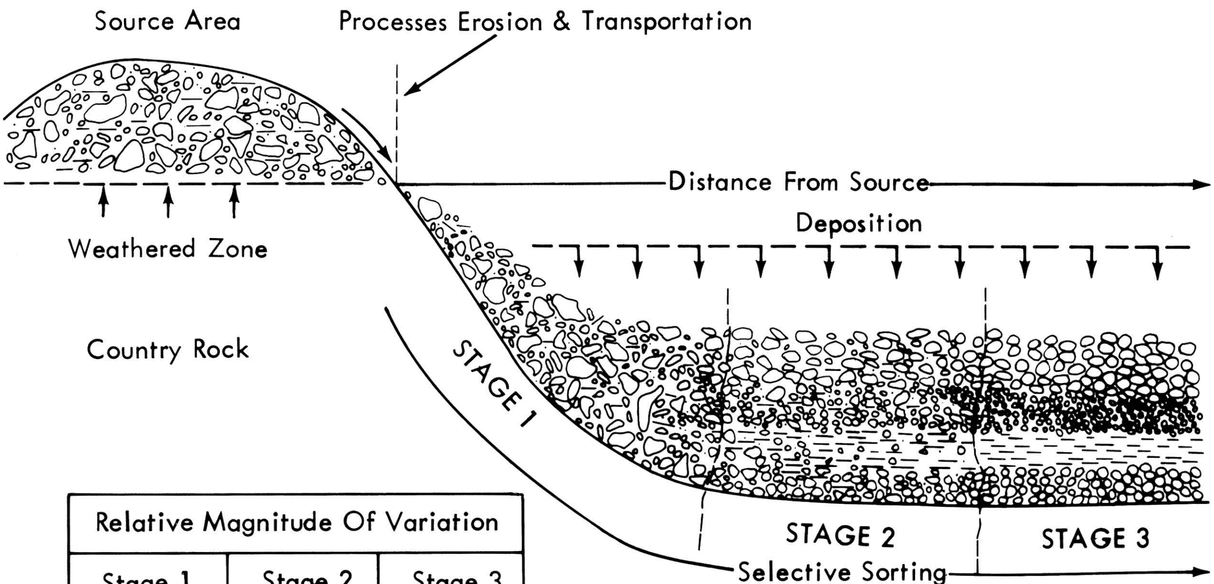
On this basis, by suitable sampling design, it is possible to determine the pattern of variation in the population, or in other words, to define the structure of the population. A sampling model for detrital sediments using this approach has been described by Griffiths (in Milner, 1962; Fig. 2). Tests of the model are described by Steinmetz (1962), Wood and Griffiths (1963), Dalhlberg (1964) and Ondrick (1968).

It is possible by ensuring fulfillment of the constraints to define structure of the variate in the population and achieve best estimates of the population parameters. In contrast, without this level of attainment there is no guarantee of achieving either objective. If the information resulting from the experiment is valuable then the estimators must be reliable, and in general, best estimators are essential. In order to ensure best estimators, fulfillment of the constraints is essential.

Suppose the object of interest in the investigation (target population in the terminology of Cochran, Mosteller and Tukey, 1954; Krumbein, 1960; Krumbein and Graybill, 1965) is a gravel pit, although it could be any population, and the variate is length of the long axis, suitably defined, of white pebbles. The population, therefore, is defined as all long axes of white pebbles in the gravel pit. In this example attention is confined to the exposed face of the pit so that all elements of the population are equally accessible. The different types of target populations which a geologist considers have been defined by Griffiths (in Milner, 1962; 1967) as follows:

1. The hypothetical population - that population of elements which was formed by geological processes operating in a specified geographical area and through a specified time interval. The whole population may no longer exist because of erosion or nondeposition, thus it is hypothetical and defined on the basis of geological information.
2. The existent population - that population of elements which now exists in some circumscribed unit volume of the earth's crust. This is the population which may be defined rigorously and yields the most statistical and geological information.

Frequently the geologist complains that the existent population is not accessible equally and so he cannot sample it. Such a complaint presupposes a judgement of value; what the geologist is actually saying is that it will cost too much to make this existent population accessible to random sample. In other words, he has decided that the result is not worth the investment! He compromises and collects data which costs little, and of course, is also worth little. Until we resolve this compromise and show that geological information, properly collected, is worth the investment necessary to collect it, our level of predictability will suffer. Indeed, the information collected on the basis of the compromise will generally be obtained cheaply and of little value. In this experiment the exposed face of the gravel is the available, and by definition, represents the existent population. It is emphasized that all conclusions, however, are limited to the face of the gravel pit!



		Relative Magnitude Of Variation		
		Stage 1	Stage 2	Stage 3
Source Of Variation	Between Samples	Small	Moderate	Large
	Within Samples	Large	Moderate	Small

Figure 2.-Source area, process, stage and variation among and within samples of different sedimentary deposits; basic model for defining sampling patterns.

3. The available population - that population of elements which is available readily and usually sampled. If it is equivalent to the existent population, then the information will be adequate. If, however, this available population is either not sampled randomly or not representative of the existent population, the estimators will be biased. The bias cannot be measured and the conclusions and interpretation will apply solely to the sampled population. In this instance it may be difficult to relate to the existent geological population.

It seems worthwhile emphasizing this aspect of sampling and investigation because few scientists, except perhaps in the behavioral sciences, have accepted the available in place of the existent population. Results of the Lanarkshire Milk experiment (Student, 1931; Pearson and Wishart, 1942), the presidential polls of 1937 and 1948 and the Kinsey report (Cochran, Mosteller, and Tukey, 1954), have been criticized severely because the conclusions, while correctly applied to the available population, were extrapolated incorrectly to a different target population (Stephan and McCarthy, 1963, p. 235-272). Indeed if the experimental result is used in decision making, and most geological investigations are thus used, then the outcome should be reliable so that the decisions are rationally sound. Geological information is considered low order because of its poor predictability, which is a function of reliability of the information. To improve reliability and predictability it is necessary to invest the effort and money required to make the data worthwhile. It is necessary to ensure that the conclusions and interpretations actually apply to the existent population.

In summary then it has been decided to obtain random samples from homogeneous populations, and the investigation is concerned with the existent population. These are stringent requirements and the result will suffice not only to yield unbiased estimators of population parameters but by suitable arrangement will yield information on the pattern of variation or structure of the population. Random samples of homogeneous populations will yield frequency distributions that are constant probability models.

SAMPLING AND MEASUREMENT PROCEDURE

The procedure actually consists of two steps, the first dealing with details of sampling and the second with process of measurement. Both steps are in effect sampling procedures and are equally important in ensuring fulfillment of the requirements.

Size, Shape and Arrangement of Samples

First assume that the pattern of variation in the population is unknown but an attempt will be made to determine the structure. Suppose the population is homogeneous, then no systematic pattern of variation exists. In geological terms this is a massive population as far as the specific variate is concerned. In such an instance a simple sampling procedure will suffice to yield random samples and unbiased estimators. Each sample will yield a mean and a variance which will be "best estimators" of the population parameters. Differences between the sample estimators will not be statistically (or geologically) significant.

On the other hand if the population possesses a structure or pattern of variation then it may be considered a collection of homogeneous subpopulations. It may be necessary, then, to random sample these subpopulations. Thus if the samples cross all subpopulations with appropriate weights, they will be random samples of the populations and will yield unbiased estimators. These are channel samples. Each sample yields a mean and variance that are unbiased estimators of population parameters and they will not be statistically (or geologically) different. They will, however, contain no information on the pattern of variation or structure.

To obtain information on the structure it is necessary to sample within the homogeneous subpopulations. Each sample will yield a mean and a variance. These will not be necessarily "best estimators" unless appropriately weighted in the proportions in which they occur within the population. If the population is composed of layers, wherein each layer is a homogeneous subpopulation, a channel sample will cross all layers and if self-weighting, it contains the appropriate contribution from each subpopulation.

When samples are taken within the layers the combined set of samples will yield "best estimators" of the population parameters only if the layers are representative of the entire population. By ensuring that the estimators are similar to those from channel sampling the same population, the appropriate weights are determined. Similarity is defined as "no statistically significant difference" between the means and variances from the two sampling designs. In effect this answers the question of how many different layers occur in the population.

Suppose that the population represents one of the stages in Figure 2. Samples are collected small enough to fall within a layer and large enough to embrace the largest element in the population, and the sample sites are arranged in a rectangular grid. The sample statistics are grouped into horizontal rows and vertical columns. If the mean of the row means equals the mean of the column means, the population is massive, that is representative of stage 1. If the mean of the row means differs significantly, in a statistical sense, from the mean of the column means then the population has a structure which may resemble stage 2 or 3. In other words this sampling arrangement enables us to differentiate the layered from the massive population.

In practice the layers may be present but may not be horizontal. Because the samples cross several layers, their arrangement does not fulfill the requirements outlined previously, and we will generally conclude wrongly that the population is a stage 1, or homogeneous type. It is necessary therefore to group the sample estimators in different ways to determine whether a structural arrangement is present or not. Generally it is necessary to specify a structure and test the results of sampling against the model. Otherwise an infinite number of possibilities result. The orthogonal (rectangular) grid may be rotated and different types of groupings examined by means of a computer; examples of the procedure applied to determining reservoir heterogeneity are described by Bennion (1965) and Bennion and Griffiths (1966), and in determining a gradient in a single specimen of a detrital sediment by Dahlberg (1964, 1965) and Dahlberg and Griffiths (1967).

Several requirements outlined here suffice to define size, shape and arrangement of samples in a population. It is advisable to take small samples where possible and sample sizes of 4 or 5 elements are usually an advantage.

In designing experiments usually more than one operator is employed and then it is necessary to remove the effect of operator variation (Griffiths and Rosenfeld, 1954). The use of several operators has a number of advantages. Firstly, it is possible to reduce the work of each operator and yet collect a relatively large number of observations. Secondly, by using several operators, the results may be generalized for all operators and final recommendations are therefore independent of operator differences. For generalized conclusions the multioperator sampling design is a requirement.

Operators differ, therefore we can arrange the experiment so that operator differences are independent of, that is orthogonal to, the source of variation we wish to isolate. The grid arrangement of sampling sites and isolation of operator differences both have a similar role--to introduce independence or orthogonality, into the analysis. It is necessary to be able to test for the presence of independence or orthogonality. The term interaction or discrepance is a measure of degree of nonindependence and the discrepance is the source of variation we wish to reduce. If the interaction between rows and columns or between operators and rows and columns is present, differences in the former instance between subpopulations cannot be isolated. In the latter instance, possible differences among subpopulations cannot be isolated from those due to differences among operators.

By envisaging possible results of the experiment, we have defined sample size, shape and arrangement and also an experimental design. Any variate (X) will contain variation from each source, for example from differences among channels (or layers) and from among operators. We wish to test the interaction between the two sources which gives a third source of variation. To test this interaction we need a fourth item, usually called replication, which leads to the experimental structure

$$X_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + [\gamma_{k(ij)} + \epsilon_{ijk}].$$

Variation in a variate represented by X_{ijk} receives contributions from differences among subpopulations α_i, differences among operators β_j, interaction (or lack of independence) of the two sources of variation if present $(\alpha\beta_{ij})$, repeated measurements within each subpopulation $(\gamma_{k(ij)})$ where (ij) represents within each α_i and β_j, and from all other unassigned sources (ϵ_{ijk}). In this design we are unable to isolate ϵ_{ijk} from $\gamma_{k(ij)}$, but it is well to be aware of its presence.

In practice, j operators will sample i subpopulations by observing a value for the variate (X) on k pebbles from each subpopulation. Small sets of operators are advantageous and we generally use 4 or 5; therefore j = 1 to q, where q = 4 or 5. Examples of designs used in sampling the Montoursville gravel in Montoursville, Pennsylvania are represented in Figures 3, 4, and 5 for stratified, channel and spot sampling respectively. These designs are utilized here.

In the stratified sampling plan, there are q operators, where q = 6 in Figure 3 or in terms of the equation above j = 1...q, where q = 4 or 5. There are "l" subpopulations or layers where l = 8 in the figure or in terms of the above equation, i = 1...l. In the experiments we will change this item for different teams using l = 10, 15 and 25 to determine what effect changing the number of layers has on the estimators. Observations were performed on "p" pebbles by each operator within each layer, or in terms of the equation above, k = 1...p. Again we will change the number of pebbles so that the total number of measurements is reasonably similar in each case.

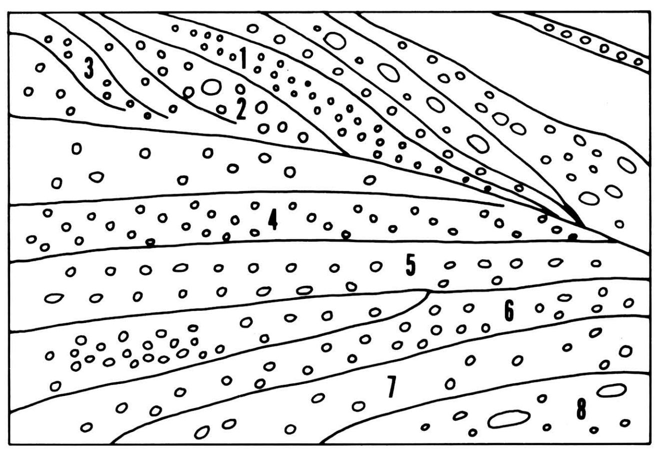
Channel and spot sampling procedures follow the same experimental structure. Spot sampling is included in the experiment because it has been advocated as an appropriate sampling procedure. From our experiments, however, it appears to be woefully inefficient. A spot sample is arranged so that it crosses some but not all layers (about equivalent to

Plan 1. Stratified Random Sampling

No. of items	Source of Variation	Degrees of Freedom	Expected Mean Square
q = 6	Operators	(q-1) = 5	$\sigma_e^2 + p\sigma_{ql}^2 + lp\sigma_q^2$
l = 8	Layers	(l-1) = 7	$\sigma_e^2 + p\sigma_{ql}^2 + qp\sigma_l^2$
	Operators by Layers	(q-1)(l-1) = 35	$\sigma_e^2 + p\sigma_{ql}^2$
p = 7	Pebbles Within Operators & Layers	ql(p-1) = 288	σ_e^2
qlp=336	Total	(qlp-1) = 335	σ_{total}^2

Population = Size of Outcrop

Operators 1 - 6



Layers 1 - 8

Pebbles	Number
Per Layer	42
Per Operator	46
TOTAL	336

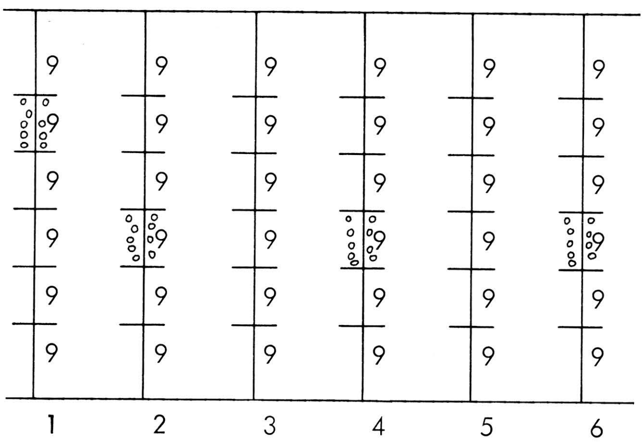
Figure 3.-Plan 1 stratified random sampling of white quartzite pebbles in Montoursville gravel Pennsylvania (1958-1959).

Plan 2. Channel Sampling

No. of items	Source of Variation	Degrees of Freedom	Expected Mean Square
q = 6	Operators	(q-1) = 5	$\sigma_e^2 + p\sigma_{cq}^2 + cp\sigma_q^2$
c = 6	Channels	(c-1) = 5	$\sigma_e^2 + p\sigma_{cq}^2 + pq\sigma_c^2$
	Operators by Channels	(q-1)(c-1) = 25	$\sigma_e^2 + p\sigma_{cq}^2$
p = 9	Pebbles Within Operators & Channels	cq(p-1) = 288	σ_e^2
qcp = 324	Total	(qcp-1) = 323	σ_{total}^2

Population = Size of Outcrop

Operators 7 - 12



Channels 1 - 6

Pebbles	Number
Per Channel	54
Per Operator	54
TOTAL	324

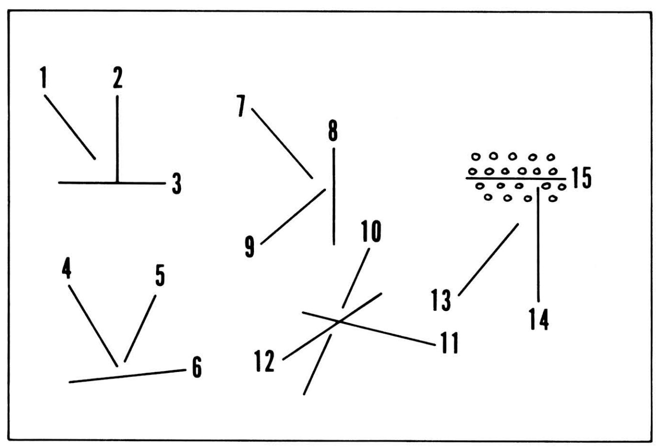
Figure 4.-Plan 2 channel sampling of white quartzite pebbles in Montoursville gravel, Pennsylvania (1958-1959).

Plan 3. Spot Sampling

No. of items	Source of Variation	Degrees of Freedom	Expected Mean Square
q = 5	Operators	(q-1) = 4	$\sigma_e^2 + p\sigma_{qs}^2 + sp\sigma_q^2$
s = 15	Spots	(s-1) = 14	$\sigma_e^2 + p\sigma_{qs}^2 + qp\sigma_s^2$
	Operators by Spots	(q-1)(s-1) = 56	$\sigma_e^2 + p\sigma_{qs}^2$
p = 4	Pebbles Within Operators & Spots	qs(p-1) = 225	σ_e^2
qsp = 300	Total	(qsp-1) = 299	σ_{total}^2

Population = Size of Outcrop

Operators 13 - 17



Spots 1 - 15

Pebbles	Number
Per Spot	20
Per Operator	60
TOTAL	300

Figure 5.-Plan 3 spot sampling of white quartzite pebbles in Montoursville gravel, Pennsylvania (1958-1959).

Table 2. Sampling designs and distribution of samples for measurement.

Sampling Design	Design Number	Replicate Number	Number of Channels (c) Layers (l) Spots (s)	Pebbles per c, l, s	TOTAL (n)	(q) Number of Operators	Pebbles Measured per Operator	(p) Pebbles Measured per Operator per c, l, s	Remarks
	1	1	8	36	288	4	72	9	Each operator samples two channels
	2	2	8	36	288	4	72	9	
Channel	3	1	6	50	300	5	60	10	Each operator samples one channel and as a team samples an additional channel
	4	2	6	50	300	5	60	10	
	5	1	5	75	375	5	75	15	
	6	2	5	75	375	5	75	15	
TOTAL					1926	28			
c = 8,6,5,; q = 4,5,; p = 9,10,15; n = qcp = 288,300,375									
	7	1	10	30	300	5	60	6	
	8	2	10	30	300	5	60	6	
Stratified (Layers)	9	1	15	20	300	5	60	4	
	10	2	15	20	300	5	60	4	
	11	1	25	10	250	5	50	2	
	12	2	25	10	250	5	50	2	
TOTAL					1700	30			
l = 10,15,25; q = 5,5,; p = 6,4,2; n = qlp = 300,300,250									
	13	1	10	35	350	5	70	7	
	14	2	10	35	350	5	70	7	
Spot	15	1	15	25	375	5	75	5	
	16	2	15	25	375	5	75	5	
	17	1	25	15	375	5	75	3	
	18	2	25	15	375	5	75	3	
TOTAL					2200	30			
s = 10,15,25; q = 5,5,5; p = 7,5,3; n = qsp = 350,375,375									
GRAND TOTAL					5826	88			

taking a handful, bagful, or small shovelful from a beach or dune sand). As a result it is a cross between the channel and stratified sample with all the disadvantages of both and with few of the advantages of either. Estimators based on this type of sample usually are biased, the frequency distributions are rarely constant probability models and convergence with increasing sample size or number is slow. The spot sampling design is included in the experiment to show, by example, what happens if the sampling arrangement is not prescribed carefully in size, shape and arrangement.

Figures 3,4, and 5 include the experimental design, the sampling procedure and the form of analysis of variance table which will be used to analyze the data. The figures illustrate that the sampling design and experimental design are interrelated closely and once the sampling arrangement is decided, the form of the statistical analysis is fixed.

Sample-Gathering Procedure

The procedure for selecting the samples will differ depending on the sampling plan, and the operational procedure will consist of the measurement of pebbles.

Selection of pebbles for measurement is common to all sampling procedures. Select the first (white) pebble at random and then take the next contiguous four white pebbles as a sample. The selection of the first pebble usually introduces some bias, that is the largest or smallest pebble commonly is selected. By taking the next four contiguous pebbles this bias is reduced, assuming that pebbles of different sizes are arranged randomly within the sampling unit.

For channel sampling place the pebbles from each channel in a heap and randomize them across operators. Position in the channel is not important and to ensure a random sample each operator selects pebbles from each heap, i.e. channel (Table 2).

In stratified sampling an operator selects a pebble (say a white one) at random from one layer. The next four contiguous white pebbles are taken within the layer and observations are made on the five (Table 2).

Spot sampling consists of a 3-foot span taken successively at increments of approximately 22.5° across the outcrop and an operator selects pebbles along the span. There will be (s) spot samples per operator consisting of (p) pebbles, that is, each operator measures (s) spots x (p) pebbles = (sp) pebbles for the spot sampling plan. Because there are (q) operators this leads to a total of qsp (white) pebbles (Table 2).

Channel Sampling

Select five, six or eight channels, depending on the design assigned, across the (existent) population. Each operator samples one channel by collecting 36, 50, or 75 pebbles respectively from the entire length of the channel. In the experimental design where eight channels are selected for sampling, each operator samples two channels, that is four operators form the team. The operators forming teams to measure the pebbles in the sampling designs composed of five and six channels, select at random from the top tenth or fifteenth of the channel a white pebble. They then collect the next four white pebbles, this comprises a set of five pebbles. Take the next tenth or fifteenth and repeat the performance and so forth to complete the channel. The operators forming the team to measure pebbles in the sampling design composed of eight channels follow the above procedure, however, collect a white pebble and the next three white pebbles. In each channel there are (Z x B) pebbles, where Z is the number of sets per channel and B is the number of pebbles collected within a set. In the present experiment, therefore, there are either 9 x 4 = 36, 10 x 5 = 50 or 15 x 5 = 75 pebbles per channel. Assemble each of the 36, 50, and 75 pebbles into individual groups.

Each of the five operators (four operators in the example of eight channels) measures nine, ten or fifteen pebbles chosen at random from each group (channel). Thus, each operator measures 72, 60 or 75 pebbles; because one sampling design has six channels, each operator samples a single channel and the sixth channel is sampled by each member of the team. Each operator collects from the sixth channel two sets of 5 pebbles or 2 x 5 x 5 = 50 pebbles. Total number of pebbles measured for this design is 6 x 10 x 5 = 300 (see Table 2).

Stratified Sampling

A total of "l" layers (l = 10, 15, 25) are selected and each of five operators selects a single pebble and "r" contiguous pebbles (r = 6, 4, 2 respectively) within the same layer. The operators assemble the pebbles for each layer and sampling design and place them in a heap for measurement. In the sampling designs composed of 10, 15, 25 strata, each operator of each team measures 6, 4 and 2 pebbles respectively from each heap or layer.

Spot Sampling

Select a vertical line 3 feet long and choose a white pebble at random within this span of 3 feet. Select the next five or six white pebbles. Repeat this sampling within each spot.

Select the next line "f" feet away and at an angle of 22.5° clockwise from vertical and repeat the sampling; proceed with this sampling procedure for 10, 15, 25 spots depending upon the experimental design (Table 2). For channel and stratified sampling designs the operators assemble the pebbles for each spot in a heap for measurement. In sampling designs composed of 10, 15 and 25 spots each operator of each team and sampling design measures 7, 5 and 3 pebbles respectively for each heap or spot.

Experimental Procedure and Data Recording

Columns 2 through 18 on Port-a-Punch cards are reserved for identification (Fig. 6). In columns 2-6 record the replicate number (Table 2), this number is assigned based on the experimental design to which the operator has been designated. The operator numbers (columns 8-10) range from 1 to 5 within a particular experimental design and are assigned prior to the experiment. Each member of the experiment may be numerically identified based on operator number and experimental design of which he is a member. The number of the channel, strata or spot and the pebble measured within these are recorded in columns 12-14 and 16-18 respectively.

In the present experiment for the measurement of size and shape of pebbles the length, breadth and thickness is measured in inches and eighths of an inch by means of a steel tape marked off in 16ths. The size data are recorded on the Port-a-Punch card in columns 24-30 inclusive for the intermediate b axis, in columns 34-40 inclusive for the long a axis and in columns 44-50 inclusive for the short c axis (Fig. 6).

Selection of the axes must be standardized and the following procedure is advocated. Allow the pebble to attain its stable position on a flat surface. It will expose usually its maximum projection area. Select the shortest diameter across the maximum projection of the pebble and complete the tangent rectangle around the image of the pebble (Fig. 7). This will yield the intermediate b axis and the long axis respectively; turn the pebble until its image yields the minimum projection area, approximately perpendicular to the maximum, and again choose the shortest diameter; this yields the c axis. The a, b, and c axes are at right angles to each other. The procedure has been used on clay particles (in electron micrographs), sand particles, as individual grains and in thin section (Smith, 1966) and on pebbles (Griffiths and Rosenfeld, 1950, Griffiths and Smith, 1964, Griffiths, 1959). It is a standardized procedure, not necessarily the best, but one which has been tested on a large number of different populations. It will lead to three axes in which $a > b > c$.

DATA ANALYSIS

Data may consist of numbers with a continuous range of variation such as measurement of pebble axes or of numbers in discrete steps, for example counts of pebbles of various colors. If sampled correctly both continuous and discrete data follow constant probability models. Continuous variates exhibit normal distributions and discrete data result generally in binomial or Poisson distributions. The computer scientist recognizes the difference between data by referring to continuous data as floating point and to discrete data as fixed point. Although only continuous data are used in the experiment both types are discussed. Two computer programs are required for data analysis, SYSTEMS PROGRAM ONE (Fig. 1 and appendix) and ANOVA (Analysis of Variance, Fig. 1).

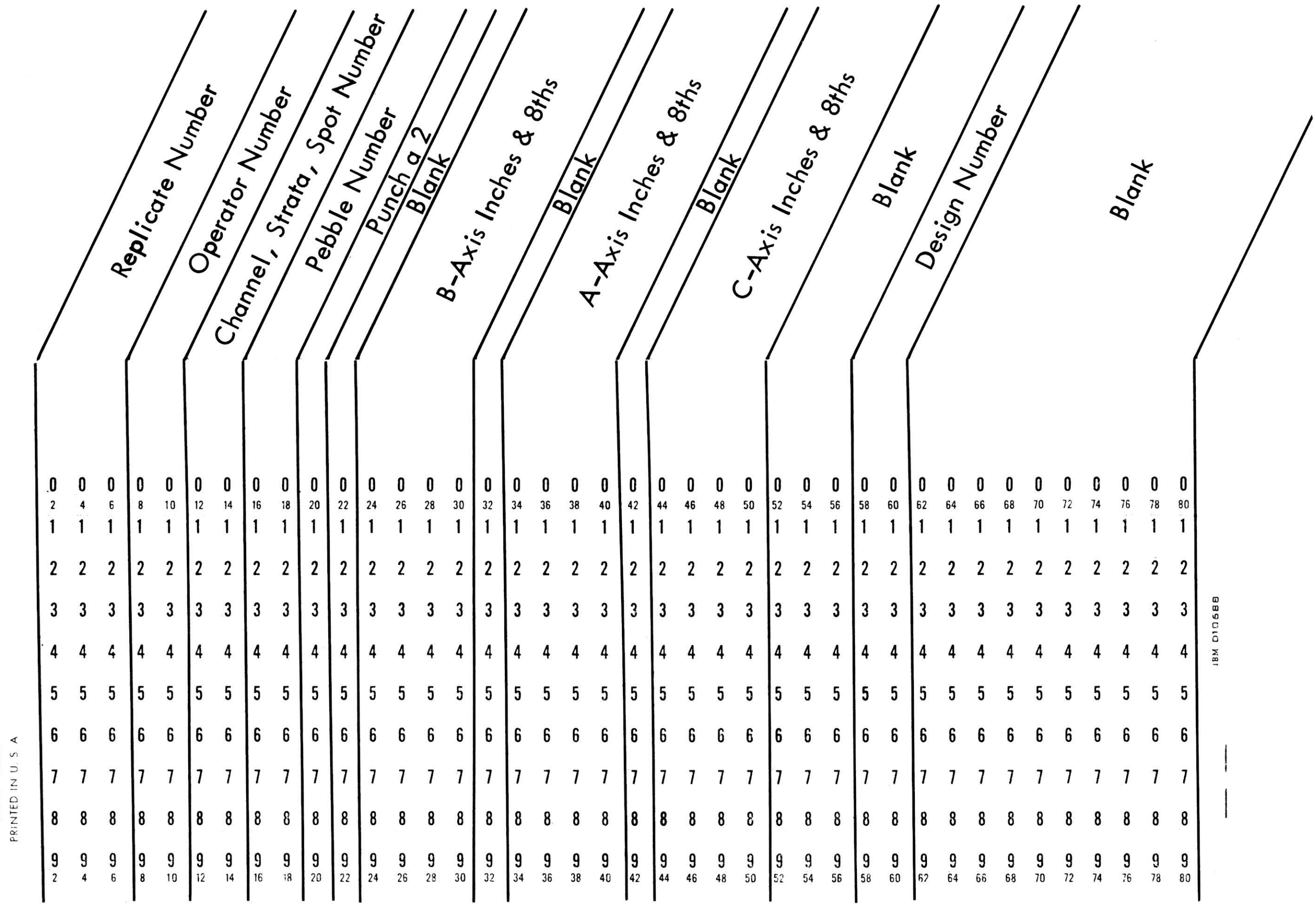


Figure 6.- Sample Port-a-Punch data card. All entries are right justified and no decimals are punched.

Data for this investigation consist of the triaxial measurements of pebbles recorded with a Port-a-Punch in the field. Theoretically Port-a-Punch cards can be "read in" a computer in the same manner as conventional Hollerith cards. In practice, however, high-speed readers usually mutilate the Port-a-Punch cards. Therefore prior to data analysis Port-a-Punch cards are passed through low-speed computers or card reproducers and converted to Hollerith cards.

The Hollerith cards are listed and checked for recording errors such as double punches, field shifts or errors causing a print check. Verifying avoids serious problems arising with high-speed computing equipment.

SYSTEMS PROGRAM ONE (Univariate Analysis; Pebble Size)

The verified grain-size data are ready for analysis. SYSTEMS PROGRAM ONE converts field units (inches and eighths) to millimeters, calculates the axial ratios in millimeters for b/a, c/a, c/b and transforms millimeters to phi (ϕ) units. The logarithmic transformation is used because frequency distri-

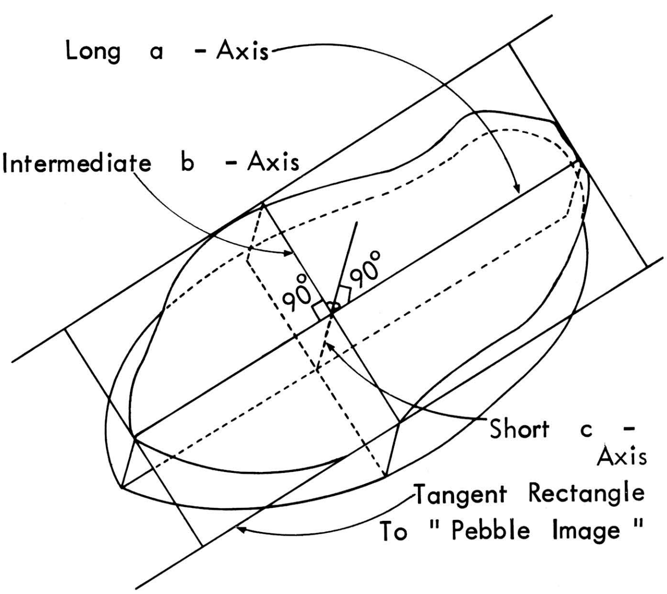


Figure 7.- Selection and relationship of pebble axes.

Table 3.-Frequency and cumulative frequency percent of a, b, c, axes in phi units for channel and stratified sampling plans applied to Montoursville gravel 1960-61.

CHANNEL

	A-AXIS φ			B-AXIS φ			C-AXIS φ		
Class Limits	Freq.	Cum. Freq.	Cum. Freq. %	Freq.	Cum. Freq.	Cum. Freq. %	Freq.	Cum. Freq.	Cum. Freq. %
-8.00 to -7.50	1	1	0.19	0	0	0.00	0	0	0.00
-7.50 to -7.00	16	17	3.23	3	3	0.57	0	0	0.00
-7.00 to -6.50	39	56	10.67	20	23	4.38	2	2	0.38
-6.50 to -6.00	71	127	24.19	41	64	12.19	23	25	4.76
-6.00 to -5.50	96	223	42.48	77	141	26.86	41	66	12.57
-5.50 to -5.00	126	349	66.48	103	244	46.48	75	141	26.86
-5.00 to -4.50	101	450	85.73	128	372	70.87	123	264	50.29
-4.50 to -4.00	53	503	95.82	91	463	88.20	109	373	71.06
-4.00 to -3.50	11	514	97.92	38	501	95.44	80	453	86.30
-3.50 to -3.00	8	522	99.44	14	515	98.11	41	494	94.11
-3.00 to -2.50	3	525	100.01	8	523	99.63	25	519	98.87
-2.50 to -2.00	0	525	100.01	2	525	100.01	3	522	99.44
-2.00 to -1.50	0	525	100.01	0	525	100.01	3	525	100.01
TOTAL	525	525	100.01	525	525	100.01	525	525	100.01

STRATIFIED

	A-AXIS φ			B-AXIS φ			C-AXIS φ		
Class Limits	Freq.	Cum. Freq.	Cum. Freq. %	Freq.	Cum. Freq.	Cum. Freq. %	Freq.	Cum. Freq.	Cum. Freq. %
-8.00 to -7.50	0	0	0.00	0	0	0.00	0	0	0.00
-7.50 to -7.00	7	7	1.56	0	0	0.00	0	0	0.00
-7.00 to -6.50	31	38	8.44	12	12	2.67	0	0	0.00
-6.50 to -6.00	56	94	20.89	30	42	9.33	10	10	2.22
-6.00 to -5.50	71	165	36.67	70	112	24.89	39	49	10.89
-5.50 to -5.00	105	270	60.00	72	184	40.89	57	106	23.56
-5.00 to -4.50	99	369	82.00	121	305	67.78	92	198	44.00
-4.50 to -4.00	60	429	95.33	92	397	88.22	105	303	67.33
-4.00 to -3.50	14	443	98.44	34	431	95.78	74	377	83.78
-3.50 to -3.00	5	448	99.55	15	446	99.11	50	427	94.89
-3.00 to -2.50	1	449	99.78	3	449	99.78	18	445	94.89
-2.50 to -2.00	1	450	99.99	1	450	99.99	3	448	99.55
-2.00 to -1.50	0	450	99.99	0	450	99.99	2	450	99.99
TOTAL	450	450	100.00	450	450	100.00	450	450	100.00

butions in terms of raw measurements are highly skewed and the means are related to the variances. The transformation tends to normalize, or at least increase symmetry of the frequency distribution and make the means independent of variances. Theoretical recommendations for the log-normal frequency distributions are discussed in Griffiths (1967, p. 271).

The systems program calculates for each sampling design the frequency distribution of each of the six variables (a, b, c axes phi, b/a, c/a, c/b mm). For pebble axes the data are classified into quarter (0.25) phi classes, for example -1.00 to -1.25, etc., and for the axial ratios a class interval of 0.05 is desirable. Selection of small class intervals (0.25 and 0.05) permits "folding" of classes if the frequency distribution appears ragged or choppy, that is frequency classes may be added together to increase the class interval to 0.50 and 0.10respectively and to smooth the distributions. Table 3 is an example of a 0.50 class interval for the Montoursville gravel data. Oper-

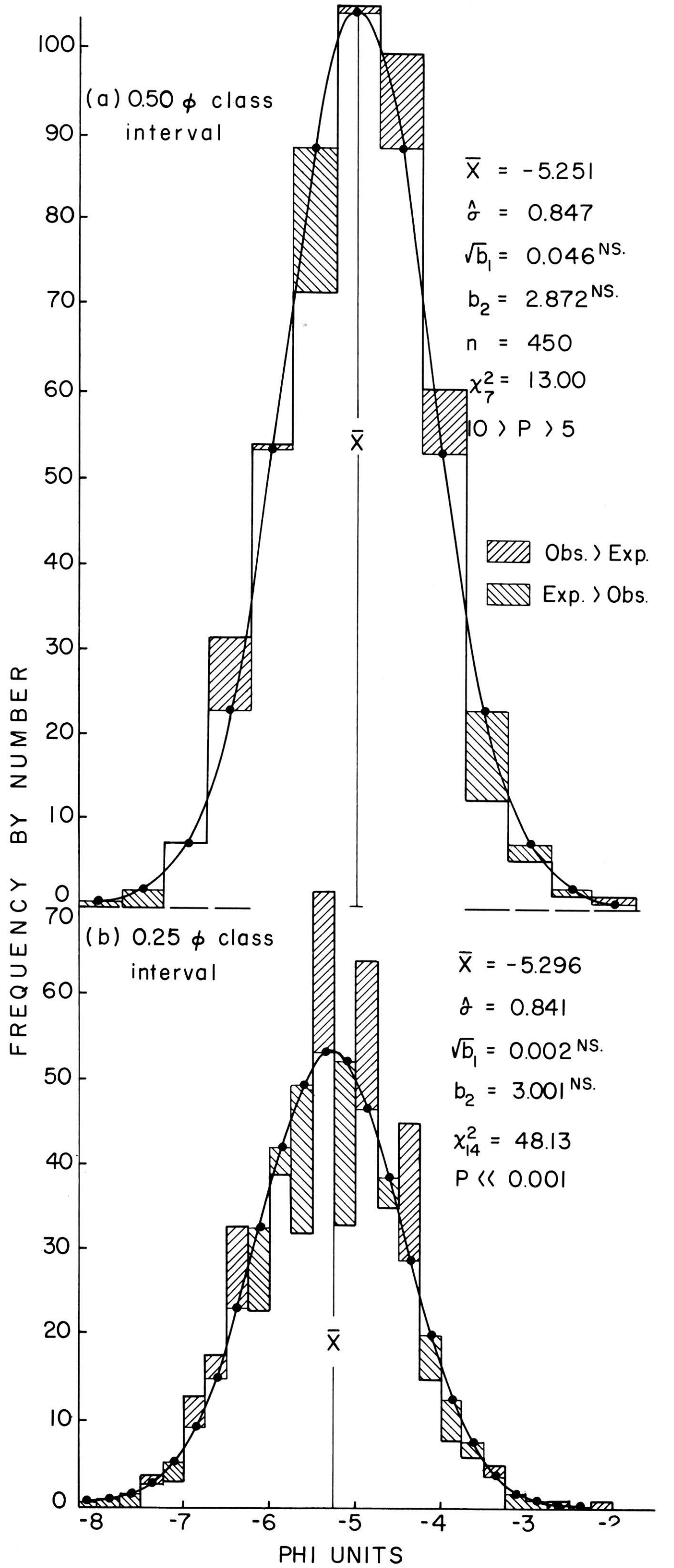


Figure 8.-Frequency histograms of a phi axis, stratified sampling plan, Montoursville gravel, 1960-61, in 0.25 and 0.50 classes.

ators will draw frequency histograms and cumulative probability curves of the data (see Fig. 8 and 9 for examples from Table 3). Frequency histograms should be drawn with class intervals of both 0.25 and 0.50 units for comparison.

From earlier discussion about results of the experiment, it was decided to obtain "best estimators" from known frequency distributions; the frequency distributions are constant probability models. In other words it is expected that frequency distributions will approximate specific types of models.

Measurement data may be calibrated against phi or log-normal models; count data may be calibrated against the binomial model for common constituents and the Poisson model for rare constituents. Sampling conditions that give rise to these models require random sampling from homogeneous populations. Also sampling procedures are arranged to fulfill the requirements as far as possible without specific knowledge of the arrangement of variation in the population. It is necessary to test for the fulfillment of the requirements and where not fulfilled to adjust the sampling arrangements until they are fulfilled. This is the only guarantee that the statistics will be "best estimators."

Consider for example the binomial model, it is required that the outcome of an event (experiment) be classifiable into two states, success or failure, white pebbles or nonwhite pebbles, etc. Probability of the occurrence of a white pebble then is constant from event to event. The fact that a white pebble has occurred is independent of the preceding or following event for all events.

The condition of sampling leads to a constant probability binomial model. If the conditions are fulfilled and a binomial distribution is generated then distribution of white pebbles in the population is a random event. Thus the occurrence of a white pebble is homogeneous throughout the population.

The mean of a binomial distribution is a sufficient estimator and as the sample size increases the binomial converges on the normal. Similar features characterize the Poisson model.

The entire analysis of the data is, in effect, an attempt to find why the sampling departs from randomness or why the observed frequency distributions do not match the required constant probability models. Two series of tests are concerned specifically with the form of the frequency distributions. The first, moment analysis, describes the shape of the curve. The expected moment values for constant probability models are known (Table 4).

Statistical estimators may be calculated from the observed data by the method of moments and compared with equivalent parametric values; the moment values are calculated by SYSTEMS PROGRAM ONE. (An example of the calculation of moment statistics by desk calculator is given in Table 5 for one set of the Montoursville gravel data.) The sampling distribution of statistical estimators of asymmetry ($\sqrt{b_1}$) and

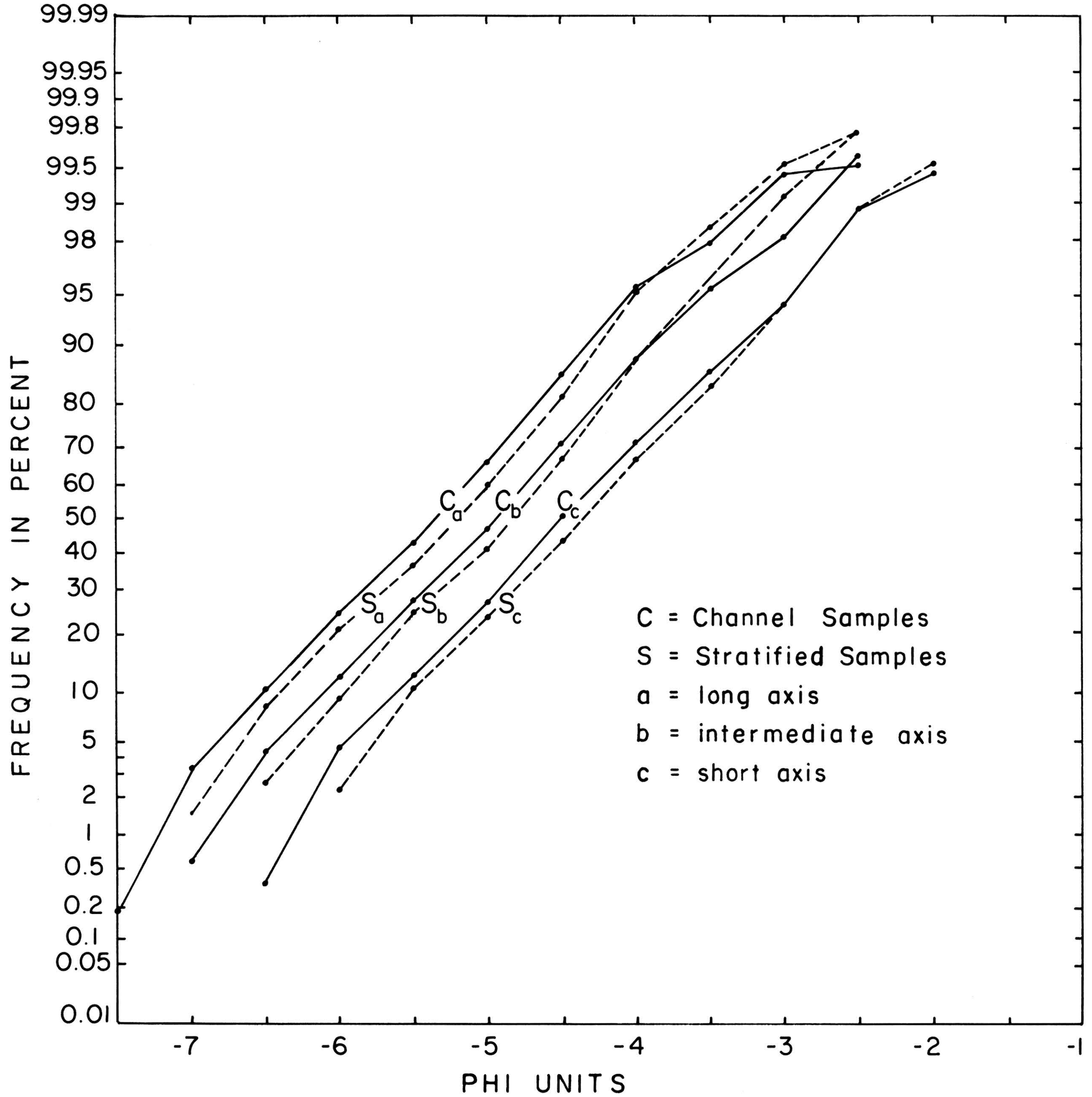


Figure 9.-Cumulative probability frequency curves of phi axis of white quartzite pebbles from Montoursville gravel; stratified and channel sampling design, 1960-61; 0.50 class interval.

Table 4.-Expected values of parameters for constant probability models.

Parameter	Mean	Variance	Standard Deviation	Asymmetry	Peakedness
Normal	μ	σ^2	σ	$\sqrt{\beta_1} = 0.0$	$\beta_2 = 3.0$
Binomial	np	npq	$\sqrt{npq}$	$\beta_1 = \frac{(q-p)^2}{npq}$	$\beta_2 = 3 + \frac{(1-6pq)}{npq}$
Poisson	np	np	$\sqrt{np}$	$\beta_1 = 1/np$	$\beta_2 = 3 + 1/np$

where n = sample size, p = probability of event occurring and q = 1 - p.

Table 5. Data sheet for calculation of first four moments of frequency distribution. Montoursville gravel 1960-61; Stratified Plan 1.

Sample Number: 2 — Computations by: J.C.G.: Aϕ

Definitions: X = mid-point of each class; $\overline{X}'$ = assumed mean = -5.75ϕ; $d = \frac{X-\overline{X}'}{c}$

$\overline{X}$ = arithmetic mean; c = class interval = 0.50ϕ; f = frequency

Class Limits	X	f	d	df	d^2f	d^3f	d^4f	$(d-1)^4f$	$(d-1)^4$
-7.50 to -7.00	-7.25	7	-4	-28	112	-448	1792	4375	625
-7.00 to -6.50	-6.75	31	-3	-93	279	-887	2511	7936	256
-6.50 to -6.00	-6.25	56	-2	-112	224	-448	896	4536	81
-6.00 to -5.50	-5.75	71	-1	-71	71	-71	71	1136	16
-5.50 to -5.00	-5.25	105	0	0	0	0	0	105	1
-5.00 to -4.50	-4.75	99	1	99	99	99	99	0	0
-4.50 to -4.00	-4.25	60	2	120	240	480	960	60	1
-4.00 to -3.50	-3.75	14	3	42	126	378	1134	224	16
-3.50 to -3.00	-3.25	5	4	20	80	320	1280	405	81
-3.00 to -2.50	-2.75	1	5	5	25	125	625	256	256
-2.50 to -2.00	-2.25	1	6	6	36	216	1296	625	625
TOTAL		450		-12	1292	-186	10,664	19,658	
		Σf		Σdf	Σd^2f	Σd^3f	Σd^4f	$\Sigma(d-1)^4f$	

Moment Calculation

CHARLIER CHECK ON TOTALS: $\Sigma(d-1)^4f = \Sigma d^4f - 4\Sigma d^3f + 6\Sigma d^2f - 4df + \Sigma f$

$19,658 = 10,664 + 744 + 7752 + 48 + 450$

ARITHMETIC MEANS OF COLUMNS: $n_1^2 = 0.000711$; $n_1^3 = 0.000019$; $n_1^4 = 0.000001$

$n_1 = \Sigma df/\Sigma f = 0.026667$; $n_2 = \Sigma d^2f/\Sigma f = 2.87111$; $n_3 = \Sigma d^3f/\Sigma f = -0.413333$;
$n_4 = \Sigma d^4f/\Sigma f = 23.697778$

ARITHMETIC MEAN OF DISTRIBUTION: $\overline{X} = \overline{X}' + cn_1 = -5.25 - (0.5 \times 0.026667) = -5.251$

VARIANCE: $m_2 = c^2(n_2 - n_1^2) = 0.25\ (2.871111 - 0.000711) = 0.717600 = \hat{\sigma}^2$

STANDARD DEVIATION: $\hat{\sigma} = \sqrt{\hat{\sigma}^2} = 0.847113$

THIRD MOMENT: $m_3 = c^3(n_3 - 3n_2n_1 + 2n_1^3) = 0.125(-0.413333 + 0.229692 - 0.037926) = 0.027696$

FOURTH MOMENT: $m_4 = c^4(n_4 - 4n_3n_1 + 6n_1^2n_2 - 3n_1^4) = 0.0625(23.697778 - 0.044089 + 0.012248$
$- 0.000003) = 1.479121$

α_3 or $\sqrt{b_1} = m_3/\hat{\sigma}^3 = 0.04556$; $\hat{\sigma}^3 = 0.607888$ — α_4 or $b_2 = m_4/\hat{\sigma}^4 = 2.872$; $\hat{\sigma}^4 = 0.514950$

SKEWNESS: $Sk = \alpha_3/2 = 0.02278$ — KURTOSIS: $K = b_2 - 3 = -0.1276$

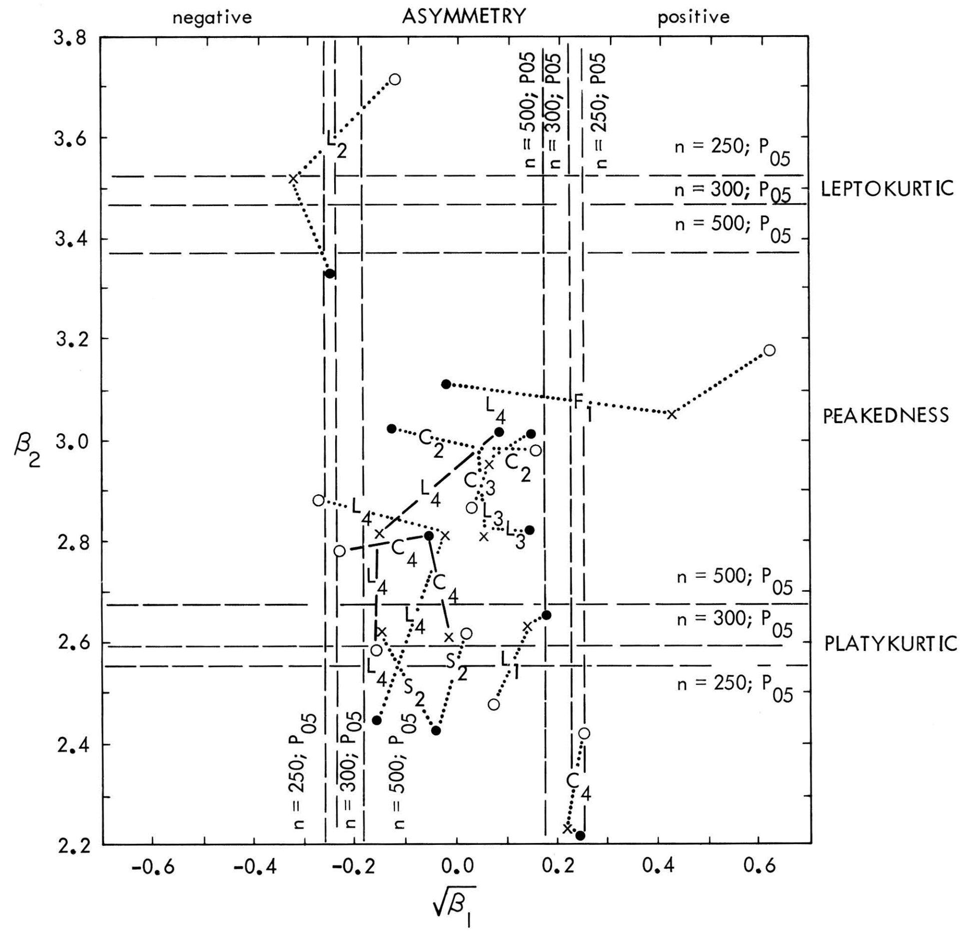
$\overline{X} = -5.251$; $\hat{\sigma}^2 = 0.718$; $\hat{\sigma} = 0.847$; $\sqrt{b_1} = 0.046$; $b_2 = 2.872$

peakedness (b_2) around their corresponding parametric values ($\sqrt{\beta_1}$ and β_2) have been tabulated for the normal distribution by Pearson and Hartley (1954, p. 183-184). A graph may be drawn showing variation in asymmetry and peakedness due to inadequacies of the sampling arrangement, that is due to nonrandomness (Fig. 10).

As the sample size "n" increases for the binomial and Poisson, the parametric values of both approach those of the normal model. The value of "n" appears in the denominator and for samples of 200 or more, departures from normality are small. With a sample of size n = 200, for example, asymmetry and peakedness in a binomial distribution changes as shown in Table 6. Departures from values expected in a normal distribution are small. This is an example of convergence following the central limit law. The normal model therefore is a basis for comparison in most of our experimental investigations.

Results from successive sampling of the Montoursville gravel by various sampling designs is shown in Figure 10. Channel sampling (C) generally yields estimators which fall within the limits of sampling

MONTOURSVILLE GRAVEL



Sampling Pattern		Year	57-58 1	58-59 2	60-61 3	64-65 4
			Sample Numbers = n			
Channel	C		---	324	525	240
Stratified	L		319	336	450	240
Spot	S		---	300	---	---
Field	F		104	---	---	---

Axis: a=o; b=x; c=● Color: White pebbles ...●... Red pebbles ---●---

Figure 10.-Graphical representation of test for nonrandomness by means of moment measures for Montoursville gravel (1957-1965).

Table 6.-Values of asymmetry (β_1) and peakedness (β_2) in a binomial with sample size n = 200.

p	q	pq	$(q-p)^2$	npq	β_1	6pq	1-6pq	(1-6pq)/(npq)	β_2
0.5	0.5	0.25	0.00	50	0.0000	1.50	-0.50	-0.001	2.999
0.4	0.6	0.24	0.04	48	0.0008	1.44	-0.44	-0.009	2.991
0.3	0.7	0.21	0.16	42	0.0040	1.26	-0.26	-0.006	2.994
0.2	0.8	0.16	0.36	32	0.0110	0.96	+0.04	+0.001	3.001
0.1	0.9	0.09	0.64	18	0.0356	0.54	+0.46	+0.026	3.026

error and are truly unbiased estimators. Estimators from the stratified sampling (L) range widely. By adjustments in number of layers and pebbles per layer, it is possible to cause convergence on unbiased estimators of the moment measures [see L_3, the 60-61 data and L_4 (red), the 64-65 data in Figure 10 and Tables 7 and 8]. If stratified sampling plans yield estimators of the third and fourth moments that do not differ significantly from those of the channel plan, the mean and variance of the stratified plan are unbiased estimators. The correct weighting of pebbles per layer and number of layers has then been achieved, that is the patterned variation or structure of the population now is explicit.

Spot sampling (S) also yields biased estimators (Fig. 10, and Table 7, 8). Although it is possible to adjust the number of spots and pebbles per spot to achieve unbiased estimators, results do not yield information of comparable geological value to those from the stratified plan. Spots are not related to subpopulations in any obvious manner, therefore the correct weighting achieves the same level of information as channel samples but at much greater expense. Convergence of spot sampling toward unbiased estimators is slow; total sample sizes of $n > 15,000$ may be necessary (Griffiths, 1967, p. 96).

The data set labelled (F) in Figure 10 represents stratified sampling and measurement in the field by a single operator. The sample size (n = 104) is too small to yield satisfactory results. Note that a axes yield unbiased estimators but both b and c axes depart widely from unbiasedness.

A second test procedure, the chi-square test,

Table 7. Sampling arrangements for Montoursville gravel; 1957-1965.

Sampling	Year Number	Operators	Layers or Channels	Pebbles	Total	Remarks	Normal
Channel (C)	58-59 (2)	6	6	9	324	C_2	OK
	60-61 (3)	7	5	15	525	C_3 3 sets of 5 Pebbles	OK
	64-65 (4)	6	10	4	240	C_4	Biased
	64-65 (4)	6	10	4	240	C_4 Red	OK (?)
Stratified (L)	57-58 (1)	4	8	10	319	L_1 1 missing value	Biased
	58-59 (2)	6	8	7	336	L_2	Biased
	60-61 (3)	6	15	5	450	L_3	OK
	64-65 (4)	6	20	2	240	L_4	Biased
	64-65 (4)	6	20	2	240	L_4 Red	OK (?)
Spot (S)	58-59 (2)	5	15	4	300	S_2	Biased
Field (F)	57-58 (1)	1	random (?)		104	F_1 talus slope	Biased

Table 8.-Summary statistics for a, b, c, axes of pebbles from Montoursville gravel (0.50 phi class interval).

Axis	Pebble Color	WHITE		WHITE			WHITE		WHITE		RED	
	Year	1957-58		1958-59			1960-61		1964-65		1964-65	
	Statistic	Stratified (Laboratory)	Stratified Single Operator	Channel	Stratified	Spot	Channel	Stratified	Channel	Stratified	Channel	Stratified
A	$\bar{X}\phi$	-4.314	-5.303	-5.265	-4.867	-5.705	-5.394	-5.317	-5.814	-5.725	-5.604	-5.620
	$\hat{\sigma}^2$	0.769	0.928	0.590	0.725	0.651	0.707	0.659	0.705	0.538	0.612	0.456
	$\hat{\sigma}\phi$	0.877	0.963	0.768	0.852	0.807	0.841	0.812	0.839	0.734	0.782	0.657
	$\sqrt{b_1}$	0.096^{NS}	0.617^{**}	0.143^{NS}	-0.102^{NS}	0.018^{NS}	0.035^{NS}	0.043^{NS}	0.246^{NS}	-0.275^{NS}	-0.215^{NS}	-0.161^{NS}
	b_2	2.498^{*}	3.176^{NS}	2.989^{NS}	3.721^{*}	2.607^{NS}	2.860^{NS}	2.997^{NS}	2.407^{*}	2.883^{NS}	2.792^{NS}	2.587^{NS}
	χ^2	8.04^{NS}	6.11^{NS}	4.23^{NS}	8.90^{NS}	5.80^{NS}	13.39^{NS}	13.00^{NS}	7.151^{NS}	7.125^{NS}	4.259^{NS}	7.393^{NS}
	d.f.	7	6	6	7	6	8	7	5	5	6	5
	P	50-30	50-50	70-50	30-20	50-30	10-05	10-05	30-20	30-20	70-50	20-10
B	$\bar{X}\phi$	-3.877	-4.832	-4.850	-4.481	-5.303	-4.997	-4.925	-5.397	-5.325	-5.112	-5.106
	$\hat{\sigma}^2$	0.767	0.943	0.617	0.711	0.657	0.753	0.653	0.724	0.521	0.595	0.461
	$\hat{\sigma}\phi$	0.876	0.971	0.785	0.843	0.811	0.868	0.808	0.849	0.722	0.771	0.679
	$\sqrt{b_1}$	0.128^{NS}	0.431^{*}	0.063^{NS}	-0.320^{**}	-0.145^{NS}	0.055^{NS}	0.005^{NS}	0.207^{NS}	-0.002^{NS}	-0.003^{NS}	-0.141^{NS}
	b_2	2.624^{*}	3.064^{NS}	2.996^{NS}	3.515^{*}	2.617^{*}	2.980^{NS}	2.825^{NS}	2.225^{**}	2.832^{NS}	2.621^{NS}	2.804^{NS}
	χ^2	12.76^{NS}	8.60^{NS}	5.82^{NS}	13.36^{NS}	6.19^{NS}	13.39^{NS}	13.38^{*}	13.181^{*}	2.367^{NS}	8.364^{NS}	2.211^{NS}
	d.f.	7	7	7	7	6	8	6	5	5	6	5
	P	10-05	20-10	70-50	10-05	50-30	10-05	05-02	05-02	80-70	30-20	90-80
C	$\bar{X}\phi$	-3.363	-4.221	-4.276	-3.981	-4.719	-4.493	-4.413	-4.897	-4.831	-4.200	-4.195
	$\hat{\sigma}^2$	0.790	1.172	0.690	0.769	0.595	0.769	0.723	0.724	0.544	0.762	0.697
	$\hat{\sigma}\phi$	0.889	1.083	0.831	0.877	0.771	0.877	0.850	0.851	0.737	0.872	0.834
	$\sqrt{b_1}$	0.182^{NS}	-0.062^{NS}	-0.134^{NS}	-0.252^{*}	-0.034^{NS}	-0.153^{NS}	0.134^{NS}	0.238^{NS}	-0.146^{NS}	-0.049^{NS}	0.091^{NS}
	b_2	2.644^{NS}	3.110^{NS}	3.025^{NS}	3.329^{NS}	2.435^{**}	3.006^{NS}	2.820^{NS}	2.223^{**}	2.467^{**}	2.828^{NS}	3.044^{NS}
	χ^2	17.49^{**}	18.45^{**}	7.04^{NS}	8.20^{NS}	16.25^{**}	12.35^{NS}	6.62^{NS}	18.161^{**}	12.487^{*}	27.933^{***}	9.011^{NS}
	d.f.	7	6	7	6	6	8	7	6	5	8	6
	P	02-01	01-001	50-30	30-20	02-01	20-10	50-30	01-001	05-02	<0.001	20-10
	n	319	104	324	336	300	525	450	240	240	240	240

Table 9.-The chi square test for normality, goodness of fit of normal and observed distribution. Montoursville gravel 1960-61; a -phi; statified sampling 0.50 class interval.

$\bar{X} = -5.251$ $\hat{\sigma} = 0.847$

Class Limits	1 (x) $X - \bar{X}$	2 $\frac{(x)}{\hat{\sigma}}$	3 Area ϕ (t)	4 Diff.	5 (E) n x Diff.	6 Observed	7 O-E	8 $(O-E)^2$	9 $\frac{(O-E)^2}{E}$
-8.5 to -8.0	-2.749	-3.246	.999415	.000585	0.26	0			
-8.0 to -7.5	-2.249	-2.655	.996034	.003381	1.52	0			
-7.5 to -7.0	-1.749	-2.065	.980537	.015497	6.97 } 8.75	7	-1.75	3.06	0.35
-7.0 to -6.5	-1.249	-1.475	.929891	.050646	22.79	31	+8.21	67.40	2.96
-6.5 to -6.0	-0.749	-0.884	.811649	.118242	52.21	56	+2.79	7.78	0.15
-6.0 to -5.5	-0.249	-0.294	.615620	.196029	88.21	71	-17.21	296.18	3.36
-5.5 to -5.0	0.251	0.296	.616384	.232004	104.40	105	+1.60	2.56	0.02
-5.0 to -4.5	0.751	0.887	.812458	.196704	88.23	99	+10.77	115.99	1.31
-4.5 to -4.0	1.251	1.477	.930160	.117702	52.97	60	+7.03	49.42	0.93
-4.0 to -3.5	1.751	2.067	.980632	.050472	22.71	14	-8.71	75.86	3.34
-3.5 to -3.0	2.251	2.658	.996069	.015437	6.95	5	-1.95	3.80	0.55
-3.0 to -2.5	2.751	3.248	.999419	.003350	1.51 } 1.77	1 } 2	+0.23	0.05	0.03
-2.5 to -2.0	3.251	3.838	.999419	.000581	0.26	1			
TOTAL				1.000000	449.99	450	+1.01	622.10	$13.00 = X^2$

n = 450
D.F. = No. of classes - 3 = 10 - 3 = 7
1 D.F. lost for mean
1 D.F. lost for $\hat{\sigma}$ $P_{.10} = 12.017$
1 D.F. lost for total

$P = .10 > P > .05$
$P_{.05} = 14.067$

The items in column 3 are found by subtracting from 0.5000 the figures found for each $-(X - \bar{X})/\hat{\sigma}$ and adding to 0.5000 the figures for each $+(X - \bar{X})/\hat{\sigma}$ from table of normal integral.
Capital X equals lower class limit; small x equals $(X - \bar{X})$.

is independent of shape of the curve. It is concerned therefore with a different type of departure from randomness than that dealt with by the moment measures. In this procedure, class-by-class comparison of frequencies of an observed distribution with those expected from a specified theoretical distribution leads to clearer pinpointing of the excesses and deficiencies due to nonrandom sampling. SYSTEMS PROGRAM ONE calculates the individual chi-square class-by-class contributions for each of the variables measured per sampling design.

The theoretical model for measurement data is the (log) normal distribution. By equating derived mean and standard deviation from moment measures to parametric values, the expected frequencies of a normal distribution may be calculated from tables (Pearson and Hartley, 1954, p. 104-110). In Figure 8(a) the ϕ class interval is 0.25, the observed mean (X) and standard deviation ($\hat{\sigma}$) equals -5.251 and 0.847 respectively, therefore the expected frequencies for a normal distribution with mean (μ) = -5.251 and standard deviation (σ) = 0.847 is calculated in Table 9. The resulting normal curve is shown in Figure 8. In this figure excesses (EXP. > OBS.) are hachured on the right and deficiencies (EXP < OBS) are hachured on the left. The calculated value of the chi-square criterion is (χ^2) = 48.13; degrees of freedom are 3 less than the number of classes or (d.f.) = c - 3 = 17 - 3 = 14. The loss of three degrees of freedom arises from equating areas under the curve (the zero moment), the means and standard deviations. Expressing the test as a <u>null hypothesis</u> H_0:

> These data are random samples from a normal model with area, mean and standard deviation given.

It would be expected therefore that if the null hypothesis were true, a chi-square value with 14 degrees of freedom would exceed 23.685 about five times in every 100 such samples and a chi-square value of 36.123 would occur about one time in 1000 (Arkin and Colton, 1950, Table 14).

If the null hypothesis is accepted, this sample represents an event that occurs much less than one time in 1000; so we decide to reject the hypothesis and now may question any and all of the assumptions underlying the test. Because we expect the normal model, the item under suspicion is the randomness of the sampling. It is, of course, reasonable to challenge the model, but it is necessary then to substitute another and test against this new model. What is the alternative model and how is one to be chosen?

If we use a 0.50 ϕ class interval, the chi-square value becomes (χ^2) = 13.00 with 7 degrees of freedom (Table 9) and leads to a probability statement of $.10 > P > .05$. In other words, if the null hypothesis is accepted, this is an event which would arise by chance in random sampling a normal distribution with the given area, mean and standard deviation between five and ten times in every 100. This is not an unusual event.

Evidently it is not the model which is at fault but the measurement procedure (a sampling procedure)! The results suggest that a 0.25 phi unit class interval is too small and therefore grouping errors are too large at this level to yield a normal distribution. Class-by-class frequency fluctuations with a 0.25 phi interval are greater than attributed to chance. There is evidently some bias introduced into the measurements causing a nonrandom fluctuation which is reduced by smoothing to a 0.50 phi unit class interval.

It should be noted that in both instances the moment measures are not statistically significant, illustrating that the moments and chi-square testing do not necessarily yield the same results. They estimate different types of departure from normality. We have found generally the sensitivity to be in the smaller class intervals. Fluctuations, or errors, in the measurements are too sensitive using the 0.25 phi unit as a basis for classification into a frequency distribution.

Fisher (1948, p. 51) states that errors of grouping are $(1/\sqrt{12n})c$, where n = sample size and c = class interval. This value should be less than 1/10 of the standard error or less than $1/10\,\hat{\sigma}_{\bar{X}}$. We may write this as an inequality

$$(1/\sqrt{12n})\,c > 1/10\,\hat{\sigma}_{\bar{X}} = \hat{\sigma}/10\sqrt{n}\ ,$$

where n = 450, c = 0.25 and 0.50, and $\hat{\sigma}$ = 0.841 and 0.847 respectively, then,

$$0.25/\sqrt{12 \times 450} \not> 0.841/10\sqrt{450}$$
$$0.50/\sqrt{12 \times 450} \not> 0.847/10\sqrt{450}\ .$$

Cancelling the common 450 from both denominators

$$0.25/\sqrt{12} \not> 0.841/10$$
$$0.50/\sqrt{12} \not> 0.847/10$$
$$0.25 \times 0.3015 \not> 0.0841$$
$$0.50 \times 0.3015 \not> 0.0847$$

or

$$0.0754 \not> 0.0841$$
$$0.1507 \not> 0.0847\ .$$

It seems we are losing information by using the 0.50 phi unit class interval but if we use the 0.25 phi interval we introduce a sampling bias.

Results of testing frequency distributions from different sampling arrangements is summarized in Table 8. The objective is to obtain nonsignificance in both the moment measures and the chi-square testing. Any departure is an indication of nonrandomness in the sampling procedure and requires an adjustment in the sampling arrangement.

There is one disturbing result that is common and concerns the erratic behavior of the c axis. Where a and b axes are nonsignificant the c axis seemingly changes independently. This is important on two counts. First, it means that different sampling arrangements are required to obtain "best estimators" for the a, b, and c axes. This is not feasible and suggests we may be studying the wrong variables. Secondly, it warns us to expect different relationships throughout the analysis for the three axes. In particular, if a is compared with b the results may differ from a compared with c. In other words, shape may change independently of size. In subsequent experiments aimed at isolating this feature we have generally found that information on shape is contained in variation among the a and c axes (Griffiths and Smith, 1964; Hulbe, 1957; Griffiths, 1959). In boulders from the Tuscarora scree (Griffiths, 1959) the relationship implies that the jointing which controls a and b axes plays a role different from the bed thickness which controls the c axis. The relationship among the axes seemingly persists throughout the life-history of individual boulders and pebbles of quartzite.

The theoretical model for count data is the binomial distribution for common pebble types and the Poisson for rare pebble types. The same procedure outlined for measurement data above is followed in analysis of count data. Frequency distributions are analysed for their statistical estimators based on the moments. The observed estimators may be compared with those expected, that is calculated from the formulae in Table 4. We expect convergence towards a normal model with increasing "n" therefore departures from the expected distribution model may be used in the same way as in the case of measurement data (i.e. as a measure of the nonrandomness of the samples), except that generally the statistical testing is not exact.

The chi-square goodness of fit test also may be applied. Expected frequencies for a binomial model with set size 5 may be calculated using the National Bureau of Standards Table (1950). Similarly, the expected frequencies for rare elements may be calculated by means of the Poisson distribution with Molina's (1942) tables. In both instances the areas under the observed and theoretical curves are made equal and the model parameter $\mu = np$ also are equated. There is, therefore, a loss of 2 degrees of freedom in the chi-square comparison for a binomial and Poisson model but otherwise the test proceeds along similar lines.

SYSTEMS PROGRAM ONE (Bivariate Analysis; Pebble Shape)

The concept of shape is independent of size.

Table 10. Summary of representation of axial-ratio shape (a vs b axes) of pebbles measured directly.

AXIAL-RATIO SHAPE

Define a = long axis, then log a = X.
Define b = intermediate axis, then log b = Y.

Shape is expressed in	Millimeters	Logarithms
	$b = ka^m$	$Y = k' = mX$
Case 1: If m = 1 (Shape is constant over size)	$b = ka$	$Y = k' + 1X$
Case 2: If m > 1, say m = 2 (Shape changes; elongation increases with increase in size)	$b = ka^2$	$Y = k' + 2X$
Case 3: If m < 1, say m = 1/2 (Shape changes; circularity increases with increasing size)	$b = ka^{1/2} = k\sqrt{a}$	$Y = k' + 1/2X$

We may now set up models in the form of hypotheses for testing as follows:

Hypothesis: H_0; shape in phi units	$Y = \beta X + \alpha$		
For constant circular shape	$\beta = 1$	and	$\alpha = 0$
For constant shape	$\beta = 1$	and	α = positive value
For circularity increasing with size	$\beta < 1$		
For elongation increasing with size	$\beta > 1$		

The test statistic is $(\beta - m)/\hat{\sigma}_m$, and it follows Student's "t" with degrees of freedom n - 2.

There are spheres and ellipsoids of all sizes and the character of a sphere or an ellipsoid is maintained independently of the size of the body. Nevertheless, it is difficult to operationally measure size and shape separately. Usual measurement procedures result in a variate (P) in which both size (s) and shape (sh) are confounded, or symbolically

$$P = f(s, sh).$$

It is feasible to design an experiment in which, by subsequent mathematical analysis, the variate called size may be separated from the variate called shape. The preceding univariate analysis represents a study of pebble size in which any of the three axes may be chosen to represent the variate with different results for different axes (Hulbe, 1957). One measure of shape that maintains the effect of size may be performed through bivariate analysis in which variations in pairs of axes are considered simultaneously. This may be extended to multivariate analysis in which variations in all three axes are considered together. Bivariate analysis will be used in the following example.

If any two axes are plotted in a scatter diagram in terms of direct measurement scales such as millimeters, inches or microns, the resulting graph shows that the spread of variation increases with increasing size in both dimensions. The variation thus shows heteroscedasticity (Griffiths, 1959, 1967). The use of an axial ratio, such as b/a, is frequently advocated to avoid this heteroscedastic result but the variation in both axes is confounded in a single figure and therefore variation in value of the ratios are difficult to evaluate. Alternately if any pair of axes is plotted in a scatter diagram in terms of logarithmic scales (or phi units) the scatter is homoscedastic, or the variance tends to be homogeneous (Griffiths, 1959, 1967). This is most desirable for bivariate statistical analysis and the variation in the two axes may be considered separately. Both correlation and regression analysis of each of the fifteen paris of variables [n variables = 6; variable pairs = n x (n-1)/2] are calculated as part of SYSTEMS PROGRAM ONE for each sampling design in this experiment.

The model used is the straight line, $Y = \alpha + \beta X$. It may be argued that the points need not follow a straight line and indeed for individual pebbles the scatter around a line may be large. For means of sets of pebbles, however, it may be shown empirically and checked by statistical testing that the straight line is indeed an adequate representation. Theoretically if all pebbles were spheres any pair of axes would plot as a straight line bisecting the scatter diagram (assuming that the scales used on both axes are in the same units). Similarly, if the relationship between pairs of axes are constant for the size range observed, the points representing pairs of axes would fall on a straight line parallel to, but not necessarily coincident with, the bisectrix. If, on the other hand, "sphericity increases with increasing size" all points may fall on a line that converges on the bisectrix as the size increases, and, of course, vice versa for sphericity increasing with decreasing size (Table 10).

As a study of shape, therefore, in terms of pairs of axes in phi (logarithmic) units the two parameters of the linear model, β measuring the slope of the line and α measuring the (Y) intercept, or distance from zero along the (Y) axis, adequately characterizes the change of shape with size. Because the values of the parameters for the bisectrix are $\beta = 1$ and $\alpha = 0$, it is possible to test the departures of observed sets of points from the parametric values. The observed sets of points may be analyzed by simple linear regression (Krumbein and Graybill, 1965, p. 221) to yield statistical estimators $\hat{\beta}$ of β and $\hat{\alpha}$ of α and the observed line is symbolized as

$$Y = \hat{\alpha} + \hat{\beta}X + \epsilon ,$$

where ϵ represents unaccounted variation or error. The model representation and appropriate statistical test are summarized in Table 10.

A third parameter that characterizes the relationship among the points in a bivariate scatter diagram is the correlation coefficient, ρ. This is a measure of how closely the points cluster about the line; for $\rho = 1$ the points are on the line and for $\rho = 0$ the points do not cluster about the line. The statistical estimator of ρ is r, the sample correlation coefficient.

For our purpose, because we used normality as a model for each of the three variates, the bivariate normal model is adequate for the three pairs of axes. Because we defined the axes as $a > b > c$ in millimeters, however, the results in phi units (a negative log transform) lead to $a<b<c$. In both instances this results in a cut-off in variation of the axes and because $b < a$ only half the scatter diagram will be used. The relationship induces some degree of association among the variate pairs and so there exists some degree of pseudocorrelation in the association of pairs of axes. C.M. Smith of the Pennsylvania State University Computation Center analyzed, by Monte Carlo sampling on a computer, the effect of this ordering of the axes and found that a maximum association of 0.66 may be induced. In statistical testing of our observed correlation coefficient, r, therefore this value is used as a comparative basis. If $r < 0.66$ no association is present among the variates. As will be seen from the results, the values of r which occur, are nearly all greater than 0.90 so that there is no doubt about the association if means of sets of axes are used to represent the shape changes. This is not necessarily true of the association among axes for individuals, particularly for a versus c axes (Griffiths and others, 1955).

The parametric correlation coefficient, ρ, or its square, ρ^2, is a valuable measure of linear association among the pairs of axes and the sample estimators, r and r^2, are included in the studies. Because ρ is dimensionless this may be considered a measure of "shape" independent of "size". It may be extended by partial correlation to three axes with interesting results (Griffiths, 1959).

Some results of this type of bivariate analysis applied to data from the sampling experiments performed on white quartzite pebbles and red sandstone pebbles from the Montoursville gravel are illustrated in Table 11 and Figures 11, 12, and 13. The data comprise the means of the set of pebbles in each sampling unit, layers, channels and spots and in Figure 13 grand means of each experiment are plotted.

The straight line appears to be an adequate representation because the points are close to a straight line ($r > 0.9$ in all but one instance and greater than 0.95 in all but three instances). Where r is below 0.95 for the b versus c and a versus c axes for the grand means (Table 11, last two rows), the value is lowered by the presence of two sampling experiments on red pebbles. If we confine attention to white quartzite pebbles only, one out of fifteen r values falls below 0.95. This is possibly an accident of sampling which we expect to occur under the null hypothesis one in every twenty times.

The next step is to test whether the lines show any evidence of departure from parallelism with the bisectrix. We may test the observed $\hat{\beta}$ against the parametric value of $\beta = 1$. Results of the test are summarized in the last column of Table 11 and only one out of eighteen is significantly different. Evidently the shape of pebbles is constant across the size range studied. This is a feature we have found in quartz grains measured directly (Hulbe, 1957), in thin section (Griffiths and others, 1955) and boulders (Griffiths, 1959). It should be emphasized that the conclusion applies to means of sets of pebbles and grains not necessarily to individual pebbles or grains.

In general we have found that a and b axes are related and the degree of relationship increases from that of individual pebbles to means and means of means etc. The a and c axes may be independent if the data treated represent individuals but again, as in the present instance, the means of relatively small sets tend to associate and the association grows with increasing sample size.

The shape of quartzite pebbles tends to be ellipsoidal and constant through a wide range of size. There is no evidence of increasing sphericity with increasing or decreasing grain size. We have too little evidence to generalize about the shapes of pebbles of other compositions but the two sets of red pebbles in Figure 13 are of a different shape from that of the white pebbles. Both are ellipsoidal and the relationship among the axes differs in white quartzite from red sandstone.

Ideally if the red sandstone pebbles occur on straight lines, this offers a powerful tool for the study of shapes of grains and pebbles. We could test first to determine if red (or other types of pebbles) occur

Table 11. Relationship between pairs of axes in Montoursville gravel, Montoursville, Pennsylvania.

Year	Sampling Design	Axial Pairs	Regression Equation	Correlation Coefficient	%Coefficient Determination	Test (1-β)
58-59	Stratified	a vs b	Y = 1.028X-0.518	0.997	99.4	NS
		b vs c	Y = 0.972X-0.393	0.997	99.5	NS
		a vs c	Y = 0.997X-0.886	0.992	98.5	NS
	Channel	a vs b	Y = 1.106X-0.972	0.988	97.7	NS
		b vs c	Y = 0.963X-0.402	0.942	88.8	NS
		a vs c	Y = 1.096X-1.503	0.959	91.9	NS
	Spot	a vs b	Y = 0.936X-0.041	0.994	98.9	NS
		b vs c	Y = 0.835X-0.290	0.989	97.8	NS
		a vs c	Y = 0.783X-0.248	0.985	97.0	*
60-61	Channel	a vs b	Y = 1.010X-0.452	0.992	98.3	NS
		b vs c	Y = 0.780X+0.597	0.971	94.3	NS
		a vs c	Y = 0.803X+0.159	0.982	96.5	NS
	Stratified	a vs b	Y = 0.997X-0.376	0.997	99.3	NS
		b vs c	Y = 0.993X-0.467	0.987	97.4	NS
		a vs c	Y = 0.993X-0.867	0.986	97.3	NS
All (57-65)	Various	a vs b	Y = 0.987X-0.361	0.995	99.0	NS
		b vs c	Y = 0.922X-0.220	0.934	87.1	NS
		a vs c	Y = 0.880X-0.385	0.897	80.4	NS

NS = not significantly different from $\beta = 1.0$

* = significantly different from $\beta = 1.0$ at the 5 percent level

on a line parallel to the bisectrix which would imply consistency in shape through the size range investigated. This would be a test of the slope of the line against a parametric value of $\beta = 1$; then we could test $\hat{\beta}$ for the red pebbles against that of the white. If these are similar the $\hat{\alpha}$'s should be tested as a measure of difference in ellipsoidicity.

Such an analysis is a rigorous test procedure and may be made sensitive by increasing sample size. It would represent a comprehensive determination of shape changes of boulders, pebbles and grains. Here again experimental design and balanced sampling is essential to obtain unbiased estimators and reproducible results.

Analysis of Variance (ANOVA)

The succeeding data analysis by means of analysis of variance, quality control graphs etc. are further tests of nonrandomness of sampling against different arrangements of sources of variation. Thus the entire analysis amounts to setting up certain constant probability models which are presumed to represent distribution of variation in the population; then to interpret the deviations which are statistically, and therefore geologically, significant as implications about the pattern of variation.

Once the requirements are fulfilled and the samples are shown to be random samples from homo-

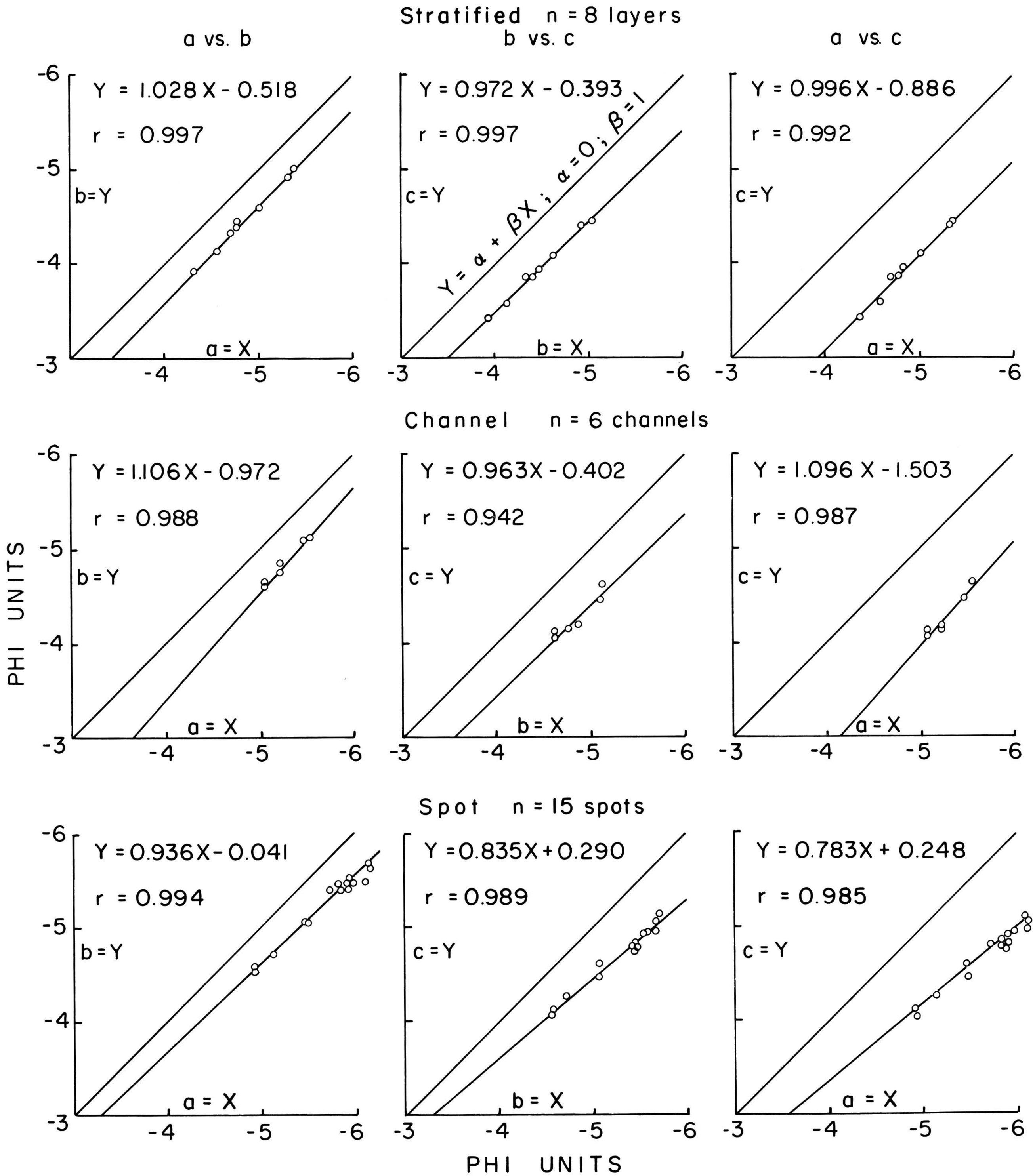


Figure 11.-Relationship between axes of white quartzite pebbles in Montoursville gravel, Montoursville, Pennsylvania; 1958-59 data; means of sampling units.

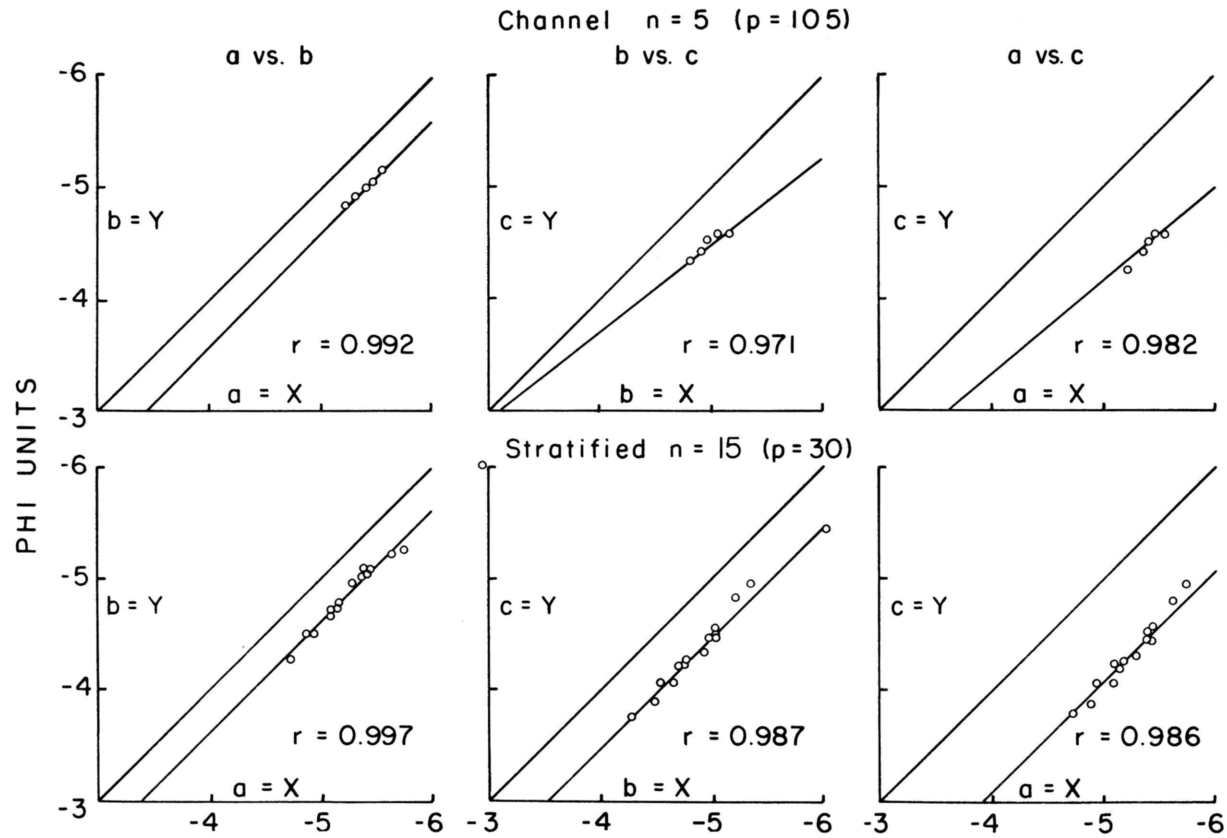


Figure 12.-Relationship between axes of white quartzite pebbles in Montoursville gravel, Montoursville, Pennsylvania; 1960-61 data; means of sampling units. "p" represents number of pebbles per channel or layer.

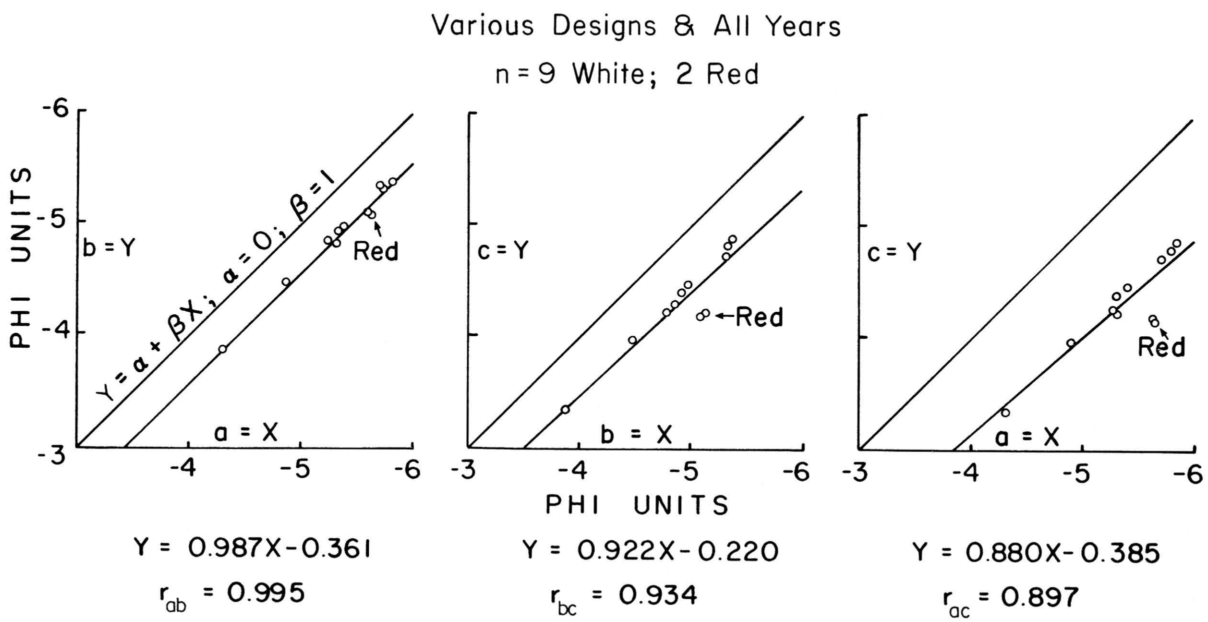


Figure 13.-Relationship between axes of white quartzite and red sandstone pebbles, Montoursville gravel, Montoursville, Pennsylvania; 1957-1965 and various designs (channel, stratified, spot); grand means.

geneous populations, which yield constant probability models, the successful sampling arrangement leads to a definition of what constitutes random sampling from the populations. The pattern of variation, therefore, in the population is defined and the structure, equated to the pattern of variation, also is defined.

Depending on the required results (Fig. 2), the entire population may be homogeneous (defined by means of channel sampling) and the characteristics represented by the mean and variance (or standard deviation) are sufficient statistics. On the other hand if the population is composed of homogeneous subpopulations, for example layers, then stratified sampling is required to yield the characteristics of the layers (again the means and variances are sufficient). The only criterion for deciding if the stratified sampling has been successful is by comparison and agreement with the results of channel sampling that have been shown to be random samples of the population.

These features arise from the process of formation of the population and its subunits. Thus varying current velocity and changing its range from time to time results in layers. Each period is characterized by constant conditions and yields a homogeneous population (sedimentation unit of Otto, 1938). The mean and variance are sufficient statistics for the description of the subpopulations.

This is why the data analysis emphasizes tests for nonrandomness. The advantage of the present approach is that it is consistent internally and yields criteria to decide when the objective has been achieved. Without this type of analysis different investigations by different investigators may, and generally do, yield different results and there is no method of deciding which is "right."

The next step in the analysis is to examine the effects of operators and to determine if the results from different operators may be combined. If the results from different operators are compatible, we may examine the realtionships between the different sources of variation arising from "natural causes." This variation is within and among channels, layers and spots, and we decide which is the most efficient sampling arrangement.

Each sampling experiment leads to a similar structure; thus each observation contains variation from at least six separate sources which may be identified, and using the present design (Fig. 3, 4, 5), five of these may be isolated. This is expressed symbolically as

$$X_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + [\gamma_{k(ij)} + \epsilon_{ijk}],$$

where X represents the variate; X_{ijk} a single observation of that variate; μ is the population mean, an unknown fixed constant; α_i is a contribution from operators around μ; and β_j is a contribution from different sampling units around μ, such as channels or layers or spots. Operators do differ and in the absence of criteria for deciding which one is "right" we would like them to differ consistently; α_i measures the degree of consistent difference, whereas $\alpha\beta_{ij}$ estimates whether the operators are inconsistent from sampling unit to sampling unit. We expect $\alpha\beta_{ij} = 0$, but, if it is significant, differences among operators cannot be separated from differences among sampling units.

There are differences also among pebbles (γ_k) within each operator and sampling unit (ij). There are many other possible sources of variation of unknown importance called "errors" which are designated ϵ_{ijk}. The "errors" source will be included in differences among pebbles in the present design. It could be isolated by each operator repeating the measurement on each axis on each pebble but this would double the effort in gathering the data.

The form of the analysis of variance table for each sampling design is illustrated in Figures 3, 4, 5 for stratified, channel and spot sampling respectively. We expect <u>channel means not to differ significantly</u>, that is the <u>variation from differences among channel</u> means should not significantly exceed differences among pebble measurements and error combined (because we expect $\alpha\beta_{ij}$ also to be zero). On the other hand we expect layer means to <u>differ significantly</u>, that is to show variation <u>significantly larger than</u> variation among pebbles. If this source is not significantly larger than the pebble variation we cannot detect layers (in terms of this variate). Ideally, differences among spot means should not exceed differences among pebbles within spots. As an example, the results for the a-ϕ axis of white quartzite pebbles from the Montoursville gravel, using channel and stratified sampling designs, are illustrated in Table 12.

<u>Channel Sampling</u>

First note that the total variation or sum of squares is 370.997 (Table 12). Variation among individual pebbles, set means plus "error" is 315.858 or some 85 percent of the total variation. This is typical and indicates that we need a large number of degrees of freedom at this level to control the variation if we are to find significant differences from any other source.

The pebbles are grouped into sets and if we include this item it changes the structure of the experiment to

$$X_{ijkl} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \delta_{k(ij)} + [\gamma_{l(ijk)} + \epsilon_{ijkl}].$$

Then variation among set means may be compared with that for "error" (including pebble to pebble variation) as an F ratio. $F_{\alpha, \nu_1, \nu_2} = 1.328/0.531 =$

Table 12.-Analysis of variance for long a axis of white quartzite pebbles from Montoursville gravel (1960-61); channel sampling and stratified sampling.

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Squares	F Ratio
		Channel Sampling		
Operator	6	10.031	1.672	1.28[NS]
Channels	4	6.800	1.700	1.046[NS]
Operator by Channel	24	39.022	1.626	2.52** 1.22[NS]
Sets	70	92.953	1.328	2.50**
Pebbles plus "Error"	420	222.905	0.530726	
TOTAL	524	370.997	0.708	
Pooled	490	315.858	0.644608	
		Stratified Sampling		
Operator	5	1.957	0.391	0.806[NS]
Layers	14	74.973	5.355	5.355**
Operator by Layer	70	45.018	0.643	1.32[NS]
Pebbles plus "Error"	360	174.561	0.485	
TOTAL	449	296.509	0.660	
Pooled	430	219.579	0.511	

2.50 which with $\nu_1 = 70$ and $\nu_2 = 420$ degrees of freedom is significant beyond the 1 percent (α percent) level (Arkin and Colton, 1950, Table 13). Evidently the sets are more variable than the pebbles plus error.

Compared with variation among set means, the discrepance is not significantly larger, nor are the differences among channel means or operator means. On this basis the deposit apparently is homogeneous and the channel means ($\bar{X}$) are unbiased estimators of the population mean (μ).

If, however, the grouping into sets is ignored by pooling sets with "error" then using this as a basis for comparison the discrepance (operator by channel) is significantly larger. That is the F ratio = 1.626/0.645 = 2.52 which with 24 and 490 degrees of freedom is significantly larger than the tabulated value of 1.84 for 24 and 400 degrees at the 1 percent level (Arkin and Colton, 1950, p. 120).

The differences among channel means and operator means are not significantly larger than the discrepance, however, if the operators are inconsistent across the channels then this tells little. It does suggest that the experiment is out of control. It is essential that the discrepance shall be nonsignificant, $\alpha\beta_{ij} = 0$, or else no firm conclusions may be drawn about operator or channel means. Obviously it is necessary to randomize operators across sets and to remove this effect if the experiment is to be in control.

It is not possible on the basis of this experiment alone to isolate the reason for this unfortunate result of significant operator by channel interaction. Compared with channel sampling experiments that achieved their goal, this experiment has several features which should be avoided. First, from experience, it appears that 4 to 6 operators are ideal and less than 4 or more than 6 introduces variation out of proportion to the degrees of freedom involved. This coincides with the findings of team experiments in psychology and sociology. Secondly there are too many pebbles and too few channels to balance the design. Seemingly, there is a delicate balance in the arrangements just as there is in stratified sampling and too few or too many channels or pebbles per channel are equally critical. One may deduce also that the large variation among pebble measurements is not stabilized by using set means where a set is 5 pebbles but at least 15 pebbles per set are needed to achieve stability.

Finally, it is necessary to randomize operators across sets to remove an interaction that may arise at this level. Fortunately, the results of other experiments using channel sampling with different weighting indicates that the difficulties may be overcome (Table 7).

The relatively wide range of variation in numbers of sampling units and numbers within sampling units in the experiment should show how critical this balance is in all types of sampling design.

Stratified Sampling

Here again the largest source of variation is at the pebble plus "error" level. This accounts for some 59 percent of the total variation in contrast to the channel experiment with 85 percent. As would be expected, variation within layers is usually less than within channels. This may be tested as an F ratio, using the largest variance in the numerator as follows

$$F_{\alpha,\nu_1,\nu_2} = F_{\alpha,490,360} = 0.645/0.485 = 1.336,$$

which is significant at the 1 percent level.

If the "error" source of variation is tested against discrepance (operator by layer) the F ratio is $F_{\alpha,\nu_1,\nu_2} = F_{.05,\ 70,\ 360} = 0.643/0.485 = 1.326$, which is not significant. Formally this implies that error plus discrepance is no greater than error alone so that the inconsistency of operators across layer means is effectively zero i.e. $\alpha\beta_{ij} = 0$ as required.

Then, using either error or discrepance as a conservative test (larger value of mean square and less degrees of freedom), the differences among layer means are highly significant, $F_{\alpha,\nu_1,\nu_2} = F_{\alpha,14,\ 70} = 5.355/0.643 = 8.327$. This most assuredly brings out the presence of layering.

On the other hand the contribution of operator mean differences are not significant if tested against discrepance or error, $F_{\alpha,\nu_1,\nu_2} = F_{\alpha,5,\ 360} =$

Table 13.-Analyses of variance of A, B and C axes in phi units for quartzite pebbles from Montoursville gravel using three sampling arrangements (1958-59 data) (computer results).

Plan	Number of Items	Source of Variation	Degrees of Freedom	Aϕ Sum of Squares	Aϕ Mean Square	Aϕ Variance Ratio (F)	Bϕ Sum of Squares	Bϕ Mean Square	Bϕ Variance Ratio (F)	Cϕ Sum of Squares	Cϕ Mean Square	Cϕ Variance Ratio (F)
Channel	6	Operators	5	22.1441	4.429	2.625^{*}	28.6303	5.726	3.523^{*}	14.9544	2.991	1.739^{NS}
	6	Channels	5	11.4381	2.288	1.361^{NS}	14.1738	2.835	1.744^{NS}	40.4359	8.087	4.702^{**}
	36	Ops x Chan.	25	42.0143	1.681	4.549^{**}	40.6350	1.625	4.490^{**}	43.0035	1.720	4.475^{**}
	9	Error	288	106.4040	0.369		104.2672	0.362		110.6967	0.384	
	324	Total	323	182.0005	0.563		187.7063	0.581		209.0905	0.647	
Stratified	6	Operators	5	26.1774	5.235	10.515^{**}	28.7704	5.754	8.832^{**}	28.5687	5.714	7.873^{**}
	8	Layers	7	38.0239	5.432	10.910^{**}	39.7595	5.680	8.718^{**}	33.4626	4.780	6.587^{**}
	48	Ops x Layer	35	23.6372	0.675	1.418^{NS}	22.8021	0.651	1.478^{NS}	25.4006	0.726	1.419^{NS}
	7	Error	288	117.1852	0.476		126.9335	0.441		147.2900	0.511	
	336	Total	335	225.0237	0.671		218.2655	0.652		234.7219	0.701	
Spot	5	Operators	4	6.6435	1.661	3.500^{**}	6.6649	1.666	3.338^{**}	8.5192	2.130	4.470^{**}
	15	Spots	14	51.6285	3.688	7.700^{**}	45.4327	3.245	6.502^{**}	32.6069	2.329	4.888^{**}
	75	Ops x Spots	56	22.2182	0.397	< 1	18.8088	0.336	< 1	19.8853	0.355	< 1
	4	Error	225	106.7890	0.475		112.3019	0.499		107.2133	0.477	
	300	Total	299	187.2795	0.626		183.2083	0.613		168.2247	0.563	

0.391/0.485 = < 1. Because operator discrepance is effectively zero this suggests that differences among operator means are independent of differences among layer means, a desirable result.

The total variance from the two experiments may also be compared as an F ratio where

$$F = 0.707/0.695 = 1.073 \text{ with } \nu_1 = 523 \text{ and } \nu_2 = 449,$$

and these are not significantly different. From this it should be clear that by careful attention to experimental design and sampling arrangement it is possible to control interfering sources of variation and isolate those of real interest. By the same token if this is not done, any differences which emerge are difficult to assign without ambiguity. Comparison of the two experiments shows that whereas the total variation is similar, the arrangement or apportionment of the variation is not the same. Thus it is possible to determine the presence or absence of a structural arrangement.

Comparison with Other Experimental Results

A more extensive series of results, using the same sampling designs plus spot sampling and all three axes, is displayed in Table 13. These may be taken to typify the general results except that, based on these experiments, somewhat better control was achieved in the 1960-61 investigation.

Generally operator means are significantly different suggesting that experimental design is required to ensure that discrepancies among different operators do not vitiate the achievement of the objective. It suggests also that results of experiments performed by single operators may yield different results because the operators differ, and without testing for consistency among operators, it is difficult to assign differences unequivocally to "natural causes."

Channel means generally do not differ although the c axes in Table 13 do show significant differences. This emphasizes that the c axis does not behave like the a and b axes.

Operator by channel interaction (discrepance) is a common occurrence and careful balance in the design is required to remove it. From the successive experiments it is our opinion that this interaction is not solely an operator effect but is a function of the number of channels and number of pebbles per channel. Usually more channels and less pebbles per channel are required to induce $\alpha\beta_{ij} = 0$ as a stable result.

The discrepance among operator by layer means is generally nonsignificant whereas both operator and layer means (main effects) are significantly different for all three axes. Similarly differences among operator by spot means is not significantly larger than "error" in the experiments. The variations in the discrepance are illustrated in Figure 14. If operators are different but consistent (discrepance not significant), the lines on a graph such as that in Figure 14 are parallel. If discrepance is significant, there are usually many crossovers among the lines. In some instances discrepance becomes significant if the lines diverge without crossing. It is difficult to detect differences among the sampling designs in Figure 14, but the layers and spots possess no significant discrep-

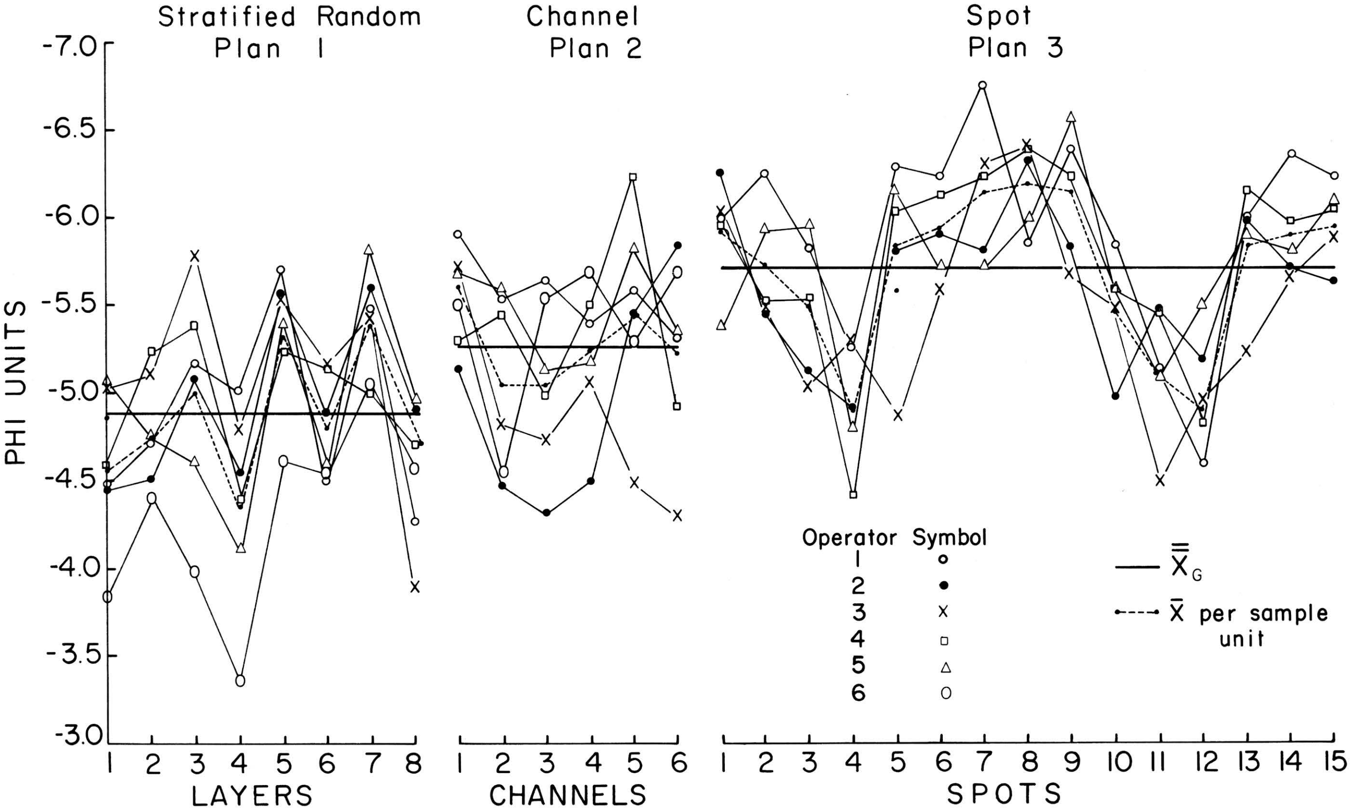


Figure 14.-Operator by sampling unit interaction in different sampling designs (Montoursville gravel, Montoursville, Pennsylvania, Aϕ-axis, 1958-59 data).

pance. The channels by operators, however, are significantly greater than their appropriate error term.

Again, the differences among channel means, and layer means, for the experiments of 1958-59 and 1960-61 are illustrated in Figure 15. in the form of a quality control graph. The spot means for the 1958-59 experiment is included also in the figure. Solid lines represent the grand means of each sampling design. The dashed lines represent confidence belts at ±3 standard deviations from the means. If the means for the sampling units, plotted as points, are random samples from homogeneous populations only 3 in 1000 should occur outside the limits.

In both channel samplings the sample means occur within the limits implying each channel mean is a random sample from a homogeneous population. Some means for layers and spots occur outside the limits. This is expected for the layers because their means should be significantly different. If this happens for spots, however, it implies that we need either more spots or more pebbles per spot or both.

It is clear also that in the 1958-59 data all three sampling designs yield significantly different grand means. If we assume the channel samples to be in control then the other estimators are biased in opposite directions.

In 1960-61 both channel and stratified sampling yield grand means that are not significantly different; both estimators therefore are adequate. Note that in both series of experiments the spread for the layer means is larger than that for the channel means. As would be expected layer means differ more than channel means.

Spot samples also show large variation and would require an increase in the number of spots and pebbles within spots to reduce the variation. This type of sampling seemingly is less efficient than channel sampling and less informative than stratified sampling.

The above experiments indicate that to obtain "best estimators" requires carefully controlled sampling. It is appropriate also to observe that if the experiment is not properly arranged, the sources of variation contribute their quota to the total variation. It is not possible then to decide what source is contributing most and thus geological conclusions are open to equivocation (and controversy).

In our experiment a modified version of the Pennsylvania State University Computation Center library program (AOV) analysis of variance factorial design is used for the analysis of variance of the pebble measurement data. There are four main sources of variation; among replicates, among operators, among channels, layers or spots and among pebbles.

CONCLUSIONS

On the basis of the sampling experiments it has been found that to achieve satisfactorily unbiased or (more demanding) "best estimators" of the population means and variances of the gravels, it is necessary to use a combination of channel and stratified sampling. The former is used to obtain unbiased estimators of the population mean and variance. If the latter yields estimators that do not differ significantly from the estimators based on the channel sampling, then we may conclude that the estimates of the mean and variance based on stratified sampling also are unbiased and "best estimators." The channel sampling is, therefore, used as a calibration standard.

The stratified sampling establishes the presence of layers, that is defines the structural arrangement. It yields estimates of the number of different layers (the weighting which yields population estimators that are unbiased) and the means and variances of the layers yield the characteristics of each subpopulation.

Studies of shape by bivariate analysis supplies information on the relationship among axes and thus indicates whether shape is constant or changes with changes in size. The regression values and associated correlation coefficients suffice to define shape and its relationship to size.

The total number of measurements (= number of pebbles) required to stabilize the estimates is about 300 or more pebbles for the Montoursville gravel; 200 is too few and 400 is about optimum.

The balance of number of pebbles per channel or per layer and number of layers (or channels) is critical. Generally five channels with 70 pebbles in each is adequate (for Montoursville gravel). Similarly 15-20 layers of 30-20 pebbles in each represents

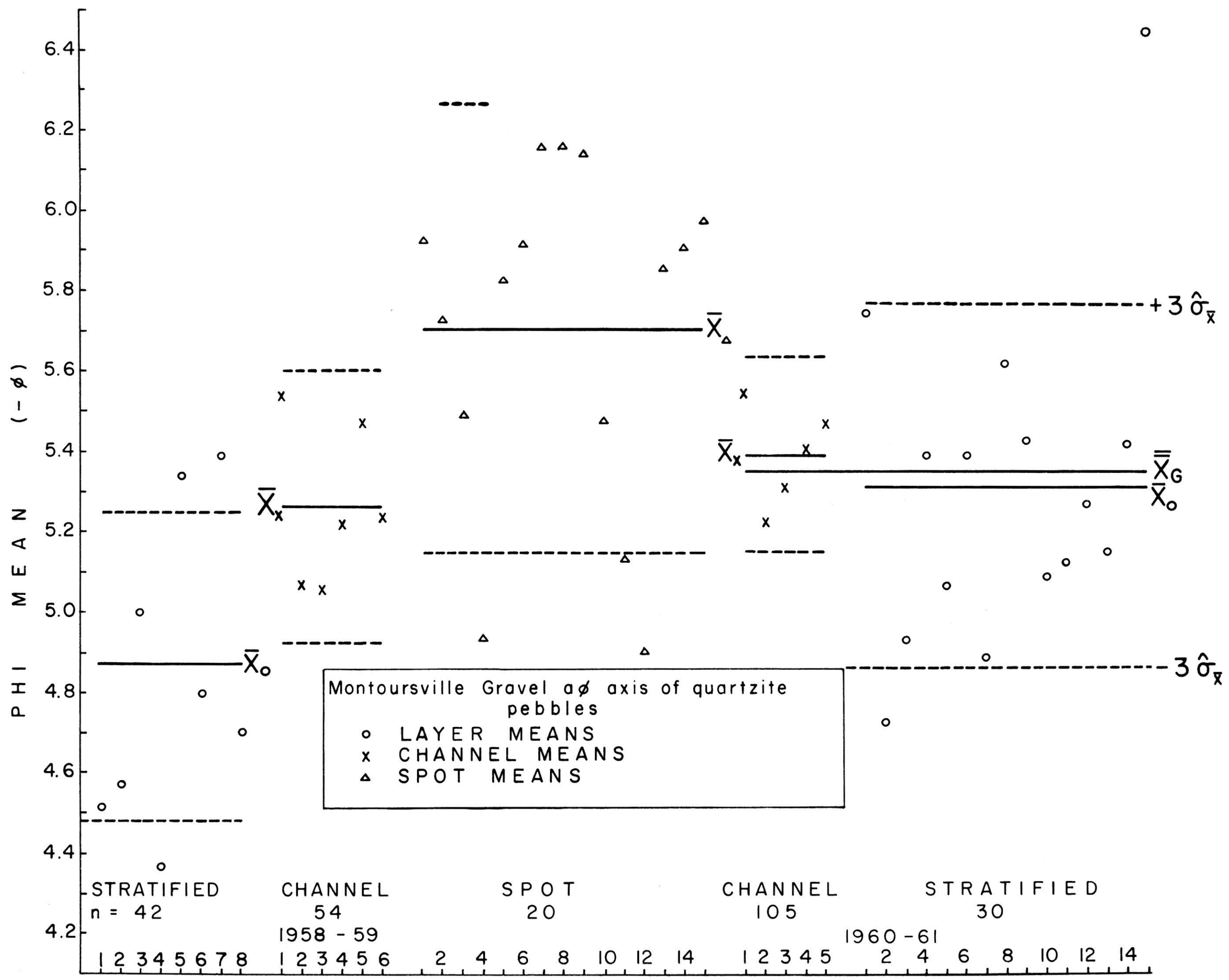


Figure 15.-Quality control graph to illustrate variation of means of sampling units around their grand means. Montoursville gravel, Montoursville, Pennsylvania, aϕ-axis, (1958-1961 data; means of sampling units).

the limiting numbers for stratified sampling. In other words, a 400-450 total sample size is required, proportioned according to the number of layers by number of pebbles per layer.

Spot sampling seemingly requires too many spots and pebbles to compete with the channel sampling and does not supply information on structure thus fails to compete with the stratified sampling design.

DISCUSSION

Reviewing achievements that result from extensive experiments indicates they are esoteric. The results include best estimators for the population mean and variance and similar estimators for the layer means and variance. From the results the most efficient sampling design may be adopted to find subsequent estimators for similar gravel deposits. From experiments on the Homewood Quartzite and on Tuscarora scree the recommendations seemingly are of general use in sampling deposits.

In the absence of knowledge on the existence of layering it is advisable to assume layering is present. From the results of the investigation then it is possible to decide whether the population is homogeneous (massive) or structured (for example layered).

The results by themselves are not informative geologically. It is possible, however, by examining other gravel deposits, to compare and contrast estimators and to establish whether there are any differences with changes in geographical direction or with stratigraphic direction, that is changes with time. This remains at the descriptive empirical level with perhaps some implications that may be applied in engineering. To see what the results really indicate in a geological sense, it is necessary to decide why size and shape are measured in the first place.

The conventional model for the petrogenesis of detrital sediments commences with a source material (or source area). This initial detritus is subjected to various processes such as weathering, erosion, transportation, deposition and diagenesis ultimately to produce a detrital sedimentary rock (Fig. 2). We may generalize this model as (Griffiths, 1966)

Source Area ⟶ Process ⟶ Product (Stage)*

The estimators determined in the sampling experiments are presumed to characterize certain aspects of the processes in this model. Thus, given a medium which is eroding, transporting and depositing detritus, it will follow in general a relationship between velocity of current and size of material of the form of Stokes Law or the Impact Law, where

$$V = Cr^2, \quad \text{or } V^2 = Cr, \text{ respectively.}$$

The relationships are log linear, that is

$$\log V = \log C + 2 \log r, \quad \text{or } 2 \log V = \log C + \log r,$$

which indicates a one-to-one mapping between log velocity and log size. This, of course, supplies some theoretical basis for using the log tansform (or the phi transformation).

On this basis, any conclusions regarding size of materials also refers to the current velocities characteristic of the carrying medium. This is why we measure size and shape to obtain information of geological interest. It is possible with this background to translate the "structure" of the material as defined in the sampling experiments into a description of the behavior of the carrying medium.

For example, suppose the medium fluctuates in velocity around some constant (average) velocity, the size range will show analogous variation. This will result in a homogeneous (massive) deposit. On the other hand, if the current velocity fluctuates in discontinuous steps through time where for relatively short periods of time it fluctuates around a low (average) velocity and later at a higher velocity and so on, the process will tend to develop an analogue in size variation in the deposit; it will be layered. A similar model may be constructed for graded bedding wherein the average changes continuously with time, and of course, two or more of these may be superimposed (Dahlberg and Griffiths, 1967).

In all instances the only evidence we possess concerning the behavior of the process is represented in the fluctuation in the size and shape of specific constituents. This is the motivation behind the sampling experiments. It is critical to the result that we know the "structure" to describe the fluctuations in velocity and therefore to describe aspects of the paleogeography. Any mistake in the experimental analysis will result in a mistake in the inferred paleogeography.

It is necessary to realize that indirect measuring techniques (for example sieving) which treat all the constituents as if they were similar in behavior will confuse the interpretation. Similarly, attempts to measure "size" which really measure something else as well, will confuse the interpretation (Griffiths, 1961, 1967). Again if white quartzite pebbles yield information on process and all red sandstone pebbles are equivalent hydraulically they will not add new information on process.

If after adjusting the variation in size and shape of white quartzite pebbles and red sandstone pebbles for their hydraulic equivalence, there remains a residual variation, it may reflect the effects of other factors than process, for example change in source material. Because relative proportions of different pebble types (or mineral types that is mineral composition) are affected by the processes, it is necessary to remove the process effect on the variables to "see" the effect of change in factors that preceded

*The Northwestern school (see Krumbein, W.C. and Graybill, F.A., 1965; and Whitten, 1964) use a process-response model for this representation but it seems to us to require a starting material as well to complete the model.

the carrying processes.

In fact the only way to "see" the effect of changes in source material is to first remove the effect of interfering sources of variation such as that induced by process (Griffiths, 1966). A mistake in estimating the effects of processes, and this is equivalent to a mistake in estimating the sizes and shapes of materials, will confuse the entire petrogenetic story.

The sampling experiments set up the most effective method of measuring the effect of processes in the petrogenesis of detrital sediments. The results are critical in deciding the petrogenesis of the material as well as in determining its industrial usefulness.

The continued controversy about petrogenetic episodes is likely to be in no small part due to the lack of reproducibility of experimental investigations which are not adequately designed to achieve their objective. The sampling experiments and their accompanying experimental designs exemplify what is required if the experimental results are to be reproducible and meaningful. Much literature in this field is based on inadequate sampling, poorly designed experiments and lack of control so that, for more effective results, some improvement in carrying out the experiments is necessary.

REFERENCES

Arkin, H., and Colton, R.R., 1950, Tables for statisticians: Barnes & Noble, Inc., New York, 152 p.

Bennion, D.W., 1965, A stochastic model for predicting reservoir rock properties: Unpubl. doctoral dissertation, The Pennsylvania State Univ., 117 p.

Bennion, D.W., and Griffiths, J.C., 1966, A stochastic model for predicting variations in reservoir rock properties: Jour. Soc. Petrol. Engr.: v. 6, p.9-16.

Cochran, W.G., Mosteller, F., and Tukey, J.W., 1954, Statistical problems of the Kinsey Report: Am. Stat. Assoc. Publ., Washington, D.C., p. 309-331.

Dahlberg, E.C., 1964, Statistical analysis of petrographic variability in a graded bed: Unpubl. master's thesis, The Pennsylvania State Univ., 156 p.

Dahlberg, E.C., 1965, Defining a gradient in a sample of sedimentary rock (abst.): Am. Assoc. Petroleum Geologists Bull., v. 49, no. 3, p. 337

Dahlberg, E.C., and Griffiths, J.C., 1967, Defining gradients in a sample of sedimentary rock: Sedimentology, v. 8, no. 4, p. 291-309.

Fisher, R.A., 1948, Statistical methods for research workers: Oliver and Boyd, Edinburgh (10th ed.), 354 p.

Griffiths, J.C., 1959, Size and shape of rock-fragments, Tuscarora scree, Fishing Creek, Lamar, central Pennsylvania: Jour. Sed, Pet., v. 29, no. 3, p. 391-401.

Griffiths, J.C., 1961, Measurement of the properties of sediments: Jour. Geology, v. 69, no. 5, p. 487-498.

Griffiths, J.C., 1966, A genetic model for the interpretive petrology of detrital sediments: Jour. Geology, v. 74, no. 5, pt. 2, p. 655-672.

Griffiths, J.C., 1967, Scientific method in analysis of sediments: McGraw-Hill Book Co., New York, 508 p.

Griffiths, J.C., in press, Sedimentary parameters, in Encyclopedia of Earth Sciences: Rheinhold Publ. Corp., New York.

Griffiths, J.C., and Rosenfeld, M.A., 1950, Progress in measurement of grain orientation of Bradford sand: Bull. Mineral Ind. Expt. Sta., The Pennsylvania State Univ., v. 56, p. 202-236.

Griffiths, J.C., and Rosenfeld, M.A., 1954, Operator variation in experimental research: Jour. Geology, v. 62, no. 1, p. 74-91.

Griffiths, J.C., Cochran, J.A., Hutta, J.J., and Steinmetz, R., 1955, Petrographical investigation of Salt Wash sediments: U.S. Atomic Energy Comm. RME 3122, Ann. Rept. April 1, 1954 to April 1, 1955, 63 p.

Griffiths, J.C. and Smith, C.M., 1964, Relationship between volume and axes of some quartzite pebbles from the Olean Conglomerate, Rock City, N.Y.: Am. Jour. Sci., v. 262, no. 4, p. 497-512.

Griffiths, J.C., and Drew, L.J., 1966, Grid spacing and success ratios in exploration for natural resources: The Pennsylvania State Univ., Min. Ind. Exp. Sta., Spec. Pub. no. 2-65, v. 1, p. Q1-Q24.

Hulbe, C.W.H., 1957, An investigation of the size and shape of quartz grains, Pedro Beach, California: Unpubl. master's thesis, The Pennsylvania State Univ., 69 p.

Kahn, J.S., 1956, The analysis and distribution of the properties of packing in sand-size sediments, 1. On the measurement of packing in sandstones; 2. The distribution of the packing measurement and an example of packing analysis: Jour. Geology, v. 64, no. 4, p. 385-395; no. 6, p. 578-606.

Krumbein, W.C., 1960, The geological population as a framework for analyzing numerical data in geology: Liverpool and Manchester Geol. Jour., v. 2, p. 341-368.

Krumbein, W.C., and Graybill, F.A., 1965, An introduction to statistical models in geology: McGraw-Hill Book Co., New York, 475 p.

Miller, R.L., and Kahn, J.S., 1962, Statistical analysis in the geological sciences: John Wiley & Sons, New York, 483 p.

Milner, H.B., 1962, Sedimentary petrography (4th ed.): George Allen and Unwin, Ltd., London, v. 1, 643 p.

Molina, E.C., 1942, Poisson's exponential binomial limit: D. Van Nostrand Co., Inc., Princeton, New Jersey, 47 p.

National Bureau of Standards, 1950, Tables of the binomial probability distribution: National Bur. of Standards, Appl. Math. Ser. No. 6, 387 p.

Ondrick, C.W., 1968, Petrography and geochemistry of the Rensselaer Graywacke, Troy, New York: Unpubl. doctoral dissertation, The Pennsylvania State Univ., 218 p.

Otto, G.H., 1938, The sedimentation unit and its use in field sampling: Jour. Geology, v. 46, no. 4, p. 569-582.

Pearson, E.S., and Hartley, H.O., 1954, Biometrika tables for statisticians, v. 1: Cambridge Univ. Press, New York, 238 p.

Pearson, E.S., and Wishart, H.D., 1942, 'Students' collected papers: Cambridge Univ. Press, New York, 224 p.

Smith, R.E., 1966, Grain size measurement in thin section and in grain mount: Jour. Sed. Pet., v. 36, no. 3, p. 841-843.

Steinmetz, R., 1962, Sampling and size distribution of quartzose pebbles from three New Jersey gravels: Jour. Geology, v. 70, no. 1, p. 56-73.

Stephan, F.F., and McCarthy, P.J., 1963, Sampling opinions; an analysis of survey procedure: John Wiley & Sons, New York, 451 p.

Student, 1931, The Lanarkshire milk experiment: Biometrika, v. 23, p. 398-406.

Whitten, E.H.T., 1964, Process-response models in geology: Geol. Soc. America Bull., v. 75, no. 5, p. 455-464.

Wood, G.V., and Griffiths, J.C., 1963, Modal analysis of three quartzites and their textural implications: Jour. Geology, v. 71, no. 4, p. 405-421.

APPENDIX

Description of SYSTEMS PROGRAM ONE

The following program is a composite of five 1620 computer programs linked together, revised and adapted to the IBM System/360 Model 67. The program is designed to accept grain-size data (a, b and c axes) from either Port-a-Punch or Hollerith cards, convert field units to phi units for a, b and c axes and obtain the ratios of b/a, c/b, c/a in millimeters. The six variables are classified into frequency distributions, moment measures calculated, and expected normal derived for each observed distribution based on the observed mean and standard deivations. A regression analysis for each of fifteen paris of variables also is given.

The program includes the following options:

1. Accepts Port-a-punch (40 punched column cards) or Hollerith cards.
2. Punches the six variables on cards for AOV or other analysis
3. Bypasses the log (phi) transformation and ratio determination and accepts any six variables for the above analysis.

Card 1 Contains the number of individual sets of data (I5).

Card 2 Contains five parameters, in order of appearance on the card they are: multiplication factor for field units to millimeters (F10.5), sum of frequencies for all six variables (F10.5), number of variables (6) (I4), reading option (see READING OPTION) (F5.0), punching option (see PUNCHING OPTION) (F5.0). FORMAT for Card 2 therefore is (2F10.5, I4, 2F5.0).

Card 3 If the reading option is 5 or 8 the card contains the format to which the data will be read in, otherwise the card immediately after Card 2 is considered as Card 4.

Card 4 plus 5 cards Contains the lower class limits, class interval and number of classes desired from the lower class limit (≤40). FORMAT of Card 4 therefore is (2F7.3, I4). The card is replicated for each of the six variables.

Data Cards

If the reading option punched on parameter Card 2 was 5 or 8, data cards will be conventional Hollerith cards. A format card submitted with the run by the user (input parameter Card 3) will state the format in which the data are punched. Alternatively, if the reading option selected on Card 2 is "BLANK", data will be read in on Port-a-Punch cards according to a fixed format. A description of that format follows:

Columns	
2-6	Replicate number (see Table 2)
8-10	Operator number
12-14	Number of channel, strata or spot
16-18	Pebble number

(Actually any series of 4 sequence numbers will suffice)

20	Punch a 2 in this column as a check on alignment of fields on each card.
24-30	b-axis (inches and eighths)
34-40	a-axis (inches and eighths)
44-50	c-axis (inches and eighths)
58-60	Design number (see Table 2)

All entries to the Port-a-Punch cards must be right justified.

Last Data Card contains a 999999 across the entire card.

Cards-Data + 1 through 6 Contains titles of each of the six variables in order of appearance on data cards (columns 1 through 36).

Repeat cards 2 through (Data + 1 through 6) based on the number on Card 1.

Reading Option

On Card 2 the fourth parameter controls the reading of the cards. Leave columns 28-29 blank to accept Port-a-Punch cards for a, b and c axes (phi conversion); punch a 5. to accept Hollerith cards for a, b and c axes (phi conversion); punch an 8. to bypass log transformation and accept any six variables for analysis. The format must be of the form (4IX, 6FX.X).

Punching Option

On Card 2 the fifth parameter controls whether or not the converted data are punched. Leave blank if no punched cards are desired; punch a 3. in columns 33-34 if punched output is desired. Data are punched in the Format (4I3,6F10.5).

```
C     $$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$
C     ********************SYSTEMS PROGRAM ONE***************************
C     $$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$
C     THE FOLLOWING PROGRAM IS A COMPOSITE OF FIVE 1620 PROGRAMS LINKED
C     TOGETHER,REVISED AND ADAPTED TO THE 360/67 BY CHARLES W. ONDRICK
C     FEB. 1968. IT IS DESIGNED TO ACCEPT GRAIN SIZE DATA(A,B,AND C AXES
C     FROM EITHER PORTA-PUNCH OR HOLLARITH CARDS-CONVERT FIELD UNITS TO
C     PHI UNITS FOR A,B,AND C AXES AND OBTAIN THE RATIOS OF B/A,C/B,C/A
C     IN MILLIMETERS. THESE SIX VARIABLES ARE THEN CLASSIFIED INTO FREQU
C     ENCY DISTRIBUTIONS,THE MOMENT MEASURES CALCULATED AND THE EXPECTED
C     NORMAL DERIVED FOR EACH OBSERVED DISTRIBUTION BASED ON THE OBSER-
C     VED MEAN AND STANDARD DEVIATION. A REGRESSION ANALYSIS FOR EACH OF
C     15 PAIRS OF VARIABLES IS ALSO GIVEN.
C     A NUMBER 2 MUST APPEAR IN COLUMN 20 FOR EACH DATA CARD(PORT-A-
C     PUNCH CARDS ONLY)
      DIMENSION X(6),FREQ(6,40),CLINT(6),XLCL(6),NOCL(6)
      DIMENSION T(10),NCLAS(6),UCL(40),CLMP(40)
      DIMENSION KTOP(6),LOW(6)
      DIMENSION FD(40),FD2(40),FD3(40),FD4(40),FD5(40),D(40)
      DIMENSION FREQEX(40),CONTB(40),BZI(40),AREA(40),DIFF(40)
      DIMENSION FMT(20)
      READ 5,KIDD
    5 FORMAT (I5)
      DO 390 LLD=1,KIDD
      PRINT 365
      PRINT 10
   10 FORMAT (1H ,21X,78HUNITS OF MEASUREMENT TO MILLIMETERS TO PHI-FREQ
     1UENCIES PLUS SUMMARY STATISTICS//)
      PRINT 15
   15 FORMAT (31X,7HKANSAS 57HGEOLOGICAL SURVEY SHORT COURSE ON SAMPLING
     1 DECEMBER 1968 //)
      READ 20,P,Q,NVAR,FIVE,THREE
   20 FORMAT (2F10.5,I4,2F5.0)
      IF (FIVE-5.0)  35,25,25
   25 READ 30,FMT
   30 FORMAT (20A4)
   35 PRINT 40,NVAR
   40 FORMAT (1H ,43X,21HNUMBER OF VARIABLES =,I3)
      PRINT 45
   45 FORMAT (1H ,//30X,14HIDENTIFICATION,2X,5HA-PHI,5X,5HB-PHI,5X,5HC-P
     1HI,5X,5H B/A ,5X,5H C/B ,5X,5H C/A )
      DO 55 I=1,NVAR
      READ 50,XLCL(I),CLINT(I),NOCL(I)
   50 FORMAT (2F7.3,I4)
   55 CONTINUE
C     ZERO OUT FREQ ARRAY
      DO 60 I=1,NVAR
      JOKE=NOCL(I)
      DO 60 J=1,JOKE
      FREQ(I,J)=0.0
   60 CONTINUE
      ZAA=0.0
      ZAB=0.0
      ZAC=0.0
      ZAD=0.0
      ZAE=0.0
      ZAF=0.0
      ZAAB=0.0
      ZAAC=0.0
      ZAAD=0.0
```

```
      ZAAE=0.0
      ZAAF=0.0
      ZABC=0.0
      ZABD=0.0
      ZABE=0.0
      ZABF=0.0
      ZACD=0.0
      ZACE=0.0
      ZACF=0.0
      ZADE=0.0
      ZADF=0.0
      ZAEF=0.0
      COAA=0.0
      COBB=0.0
      COCC=0.0
      CODD=0.0
      COEE=0.0
      COFF=0.0
C     READ PORTA PUNCH CARDS
   65 IF (FIVE-5.0)  70,85,100
   70 CONTINUE
      READ 75,IX2,IX4,IX6,IX8,IX10,IX12,IX14,IX16,IX18,IX20,IX24,IX26,IX
     128,IX30,IX34,IX36,IX38,IX40,IX44,IX46,IX48,IX50
   75 FORMAT (1XI1,1XI1,1XI1,1XI1,1XI1,1XI1,1XI1,1XI1,1XI1,1XI1,3XI1,1XI
     11,1XI1,1XI1,3XI1,1XI1,1XI1,1XI1,3XI1,1XI1,1XI1,1XI1)
      IF (IX20-2)  150,80,150
   80 IB=IX24*100+IX26*10+IX28
      XB=IB
      X30=IX30
      B=(XB+X30/8.0)*P
      IA=IX34*100+IX36*10+IX38
      XA=IA
      X40=IX40
      A=(XA+X40/8.0)*P
      IC=IX44*100+IX46*10+IX48
      XC=IC
      X50=IX50
      C=(XC+X50/8.0)*P
      I=IX2*100+IX4*10+IX6
      J=IX8*10+IX10
      K=IX12*10+IX14
      L=IX16*10+IX18
      GO TO 95
   85 READ FMT,I,J,K,L,XAA,XAB,XAC
      IF (XAA-9999.0)  90,150,150
   90 A=XAA*P
      B=XAB*P
      C=XAC*P
   95 AC=(1.0/ALOG(2.0))
      X(1)=-((ALOG(A))*AC)
      X(2)=-((ALOG(B))*AC)
      X(3)=-((ALOG(C))*AC)
      X(4)=B/A
      X(5)=C/B
      X(6)=C/A
      GO TO 105
  100 READ FMT,I,J,K,L,    X(1),X(2),X(3),X(4),X(5),X(6)
      IF (X(1)-9999.0)  105,150,150
  105 PRINT 110,I,J,K,L,X(1),X(2),X(3),X(4),X(5),X(6)
  110 FORMAT (1H ,26X,4I3,6F10.5)
      IF (THREE-3.0)  125,115,125
```

```
  115 PUNCH 120,I,J,K,L,X(1),X(2),X(3),X(4),X(5),X(6)
  120 FORMAT (4I3,6F10.5)
  125 ZAA=ZAA+X(1)
      ZAB=ZAB+X(2)
      ZAC=ZAC+X(3)
      ZAD=ZAD+X(4)
      ZAE=ZAE+X(5)
      ZAF=ZAF+X(6)
      ZAAB=ZAAB+X(1)*X(2)
      ZAAC=ZAAC+X(1)*X(3)
      ZAAD=ZAAD+X(1)*X(4)
      ZAAE=ZAAE+X(1)*X(5)
      ZAAF=ZAAF+X(1)*X(6)
      ZABC=ZABC+X(2)*X(3)
      ZABD=ZABD+X(2)*X(4)
      ZABE=ZABE+X(2)*X(5)
      ZABF=ZABF+X(2)*X(6)
      ZACD=ZACD+X(3)*X(4)
      ZACE=ZACE+X(3)*X(5)
      ZACF=ZACF+X(3)*X(6)
      ZADE=ZADE+X(4)*X(5)
      ZADF=ZADF+X(4)*X(6)
      ZAEF=ZAEF+X(5)*X(6)
      COAA=COAA+X(1)*X(1)
      COBB=COBB+X(2)*X(2)
      COCC=COCC+X(3)*X(3)
      CODD=CODD+X(4)*X(4)
      COEE=COEE+X(5)*X(5)
      COFF=COFF+X(6)*X(6)
C     BUILD FREQUENCY DISTRIBUTIONS
      DO 145 I=1,NVAR
      JOKE=NOCL(I)
      DO 140 J=1,JOKE
      Z=J
      QP=XLCL(I)+Z*CLINT(I)
      IF (X(I)-QP)  135,135,140
  135 FREQ(I,J)=FREQ(I,J)+1.
      GO TO 145
  140 CONTINUE
  145 CONTINUE
      GO TO 65
  150 CONTINUE
C     ELIMINATION OF EMPTY CLASSES ON BOTH ENDS
C     LOWER END
      PRINT 365
      DO 370 L=1,NVAR
      READ 155,T(1),T(2),T(3),T(4),T(5),T(6),T(7),T(8),T(9)
  155 FORMAT (9A4)
      NCLAS(L)=0
      JOKE=NOCL(L)
      DO 160 J=1,JOKE
      IF (FREQ(L,J))  160,160,165
  160 CONTINUE
  165 LOW(L)=J
      Z=LOW(L)
C     NEW LOWEST BOUND
      XLCL(L)=XLCL(L)+Z*CLINT(L)-CLINT(L)
C     UPPER END
      DO 170 J=1,JOKE
      K=(JOKE-J)+1
      IF (FREQ(L,K))  170,170,175
```

```
  170 CONTINUE
  175 KTOP(L)=K
      NCLAS(L)=(KTOP(L)-LOW(L))+1
C     COLLECT NEW CLASSES, LIMITS AND FREQUENCIES
      KOP=KTOP(L)
      LOP=LOW(L)
C    PUNCH OUT UPPER CLASS LIMITS AND OBSERVED FREQUENCIES FOR NVAR VARI
      PRINT 180,T(1),T(2),T(3),T(4),T(5),T(6),T(7),T(8),T(9),CLINT(L),NC
     1LAS(L)
  180 FORMAT (1H ,26X,9A4,5X,6HINT = ,F7.3,2X,I4,8H CLASSES/)
      PRINT 185
  185 FORMAT (1H ,32X,12HCLASS NUMBER,5X,11HUPPER LIMIT,5X,8HMIDPOINT,5X
     1,19HFREQUENCY(OBSERVED))
      Z=0.0
      DO 195 J=LOP,KOP
      Z=Z+1.
      UCL(J)=XLCL(L)+Z*CLINT(L)
      CLMP(J)=UCL(J)-(CLINT(L))/2.
      PRINT 190,Z,UCL(J),CLMP(J),FREQ(L,J)
  190 FORMAT (1H ,35X,F5.0,8X,F10.4,5X,F10.4,9X,F10.0)
  195 CONTINUE
      CONTINUE
C     *SUMSTATPHI* PROGRAM TO DERIVE MOMENTS FOR FREQUENCY DISTRIBUTIONS
C     DERIVED BY *FREQPHI*
      N=NCLAS(L)
      C=CLINT(L)
C     SELECT CLASS WITH MAXIMUM FREQUENCY
      YMAX=FREQ(L,LOP)
      KK=LOP
      KLOP=LOP+1
      ILOP=KLOP+N
      DO 210 K=KLOP,ILOP,1
      IF (YMAX-FREQ(L,K))  205,210,210
  205 YMAX=FREQ(L,K)
      KK=K-LOP+1
  210 CONTINUE
C     PRINT MAX. FREQ., CLASS NO., VALUE OF X AT THIS CLASS
      PRINT 215
  215 FORMAT (1H ,/40X,17HMAXIMUM FREQUENCY,5X,5HCLASS,5X,8HMIDPOINT)
      IJOL=KK+LOP-1
      PRINT 220,YMAX,KK,CLMP(IJOL)
  220 FORMAT (1H ,42X,F7.0,13X,I4,5X,F8.3)
C     ASSIGN D SCALE
      DO 225 I=1,ILOP
      D(I)=(CLMP(I)-CLMP(IJOL))/C
  225 CONTINUE
      SF=0.0
      SD=0.0
      SD2=0.0
      SD3=0.0
      SD4=0.0
      SD5=0.0
      DO 230 I=1,ILOP
      FD(I)=FREQ(L,I)*D(I)
      SD=SD+FD(I)
      FD2(I)=FREQ(L,I)*(D(I)**2)
      SD2=SD2+FD2(I)
      FD3(I)=FREQ(L,I)*(D(I)**3)
      SD3=SD3+FD3(I)
      FD4(I)=FREQ(L,I)*(D(I)**4)
      SD4=SD4+FD4(I)
```

```
      FD5(I)=FREQ(L,I)*((D(I)-1.0)**4)
      SD5=SD5+FD5(I)
C     CALCULATE TOTAL FREQUENCY, SF
      SF=SF+FREQ(L,I)
  230 CONTINUE
C     CALCULATE GRAM-CHARLIER CHECK AND TEST
      GCK=SD4-(4.0*SD3)+(6.0*SD2)-(4.0*SD)+SF
      IF (GCK-SD5)  235,250,235
  235 PRINT 240
  240 FORMAT (1H ,/48X,10HGRAM CHECK,3X,10H     SUMF4)
      PRINT 245,GCK,SD5
  245 FORMAT (1H ,48X,F10.6,5X,F10.6)
  250 GO TO 255
C     CALCULATE MOMENTS AROUND ASSUMED MEAN
  255 AMOM1=SD/SF
      AMOM2=SD2/SF
      AMOM3=SD3/SF
      AMOM4=SD4/SF
C     CALCULATE MOMENTS AROUND TRUE MEAN
      TMOM1=(AMOM1)*C
      TMOM2=(C**2)*(AMOM2-((AMOM1)**2))
      TMOM3=(C**3)*(AMOM3-(3.0*AMOM2*AMOM1)+(2.0*(AMOM1**3)))
      PT4A=(AMOM4-(4.0*AMOM3*AMOM1)+(6.0*(AMOM1**2)*AMOM2))
      PT4B=PT4A-(3.0*(AMOM1**4))
      TMOM4=(C**4)*(PT4B)
C     CALCULATE SUMMARY STATISTICS
      AVEX=CLMP(IJOL)+TMOM1
      VAR=TMOM2
      STDV=SQRT(VAR)
      RTB1=(TMOM3/((SQRT(VAR))**3))
      B2=(TMOM4/(TMOM2**2))
C     PRINT SUMMARY STATISTICS
      PRINT 260
  260 FORMAT (1H ,/40X,7HAVERAGE,3X,8HVARIANCE,3X,18HSTANDARD DEVIATION)
     1
      PRINT 265,AVEX,VAR,STDV
  265 FORMAT (1H ,39X,F8.3,1X,F11.5,4X,F8.4)
      PRINT 270
  270 FORMAT (1H ,/38X,7HROOT B1,8X,2HB2,8X,18HSUM OF FREQUENCIES)
      PRINT 275,RTB1,B2,SF
  275 FORMAT (1H ,37X,F8.3,5X,F8.3,10X,F8.0//)
      PRINT 280
  280 FORMAT (1H ,20X,80H**************************************************
     1******************************//)
      PRINT 285
  285 FORMAT (1H ,50X,30HFIT TO THE NORMAL DISTRIBUTION//)
      A1=0.09979268
      A2=0.04432014
      A3=0.00969920
      A4=-0.00009862
      A5=0.00058155
      SOBFQ=0.0
      IIM=NCLAS(L)+1
      NDF=NCLAS(L)-3
      DO 290 J=LOP,KOP
      SOBFQ=SOBFQ+FREQ(L,J)
      BZI(J)=(UCL(J)-AVEX)/STDV
  290 CONTINUE
      SDIFF=0.0
      SEXPF=0.0
      CHISQ=0.0
```

```
      IJK=0
      IKO=LOP-1
      AREA(IKO)=0.0
      DO 335 I=LOP,KOP
      IIV=I-1
      IF (BZI(I))  295,300,300
  295 AVT=-BZI(I)
      GO TO 305
  300 AVT=BZI(I)
  305 AREA(I)=1.0+AVT*(A1+AVT*(A2+AVT*(A3+AVT*(A4+AVT*A5))))
      AREA(I)=0.5/(AREA(I)**8)
      IF (BZI(I))  310,315,315
  310 DIFF(I)=AREA(I)-AREA(IIV)
      GO TO 330
  315 IJK=IJK+1
      IF (IJK-1)  320,320,325
  320 DIFF(I)=1.0-AREA(I)-AREA(IIV)
      GO TO 330
  325 DIFF(I)=AREA(IIV)-AREA(I)
  330 SDIFF=SDIFF+DIFF(I)
      FREQEX(I)=DIFF(I)*SF
      SEXPF=SEXPF+FREQEX(I)
      CONTB(I)=((FREQ(L,I)-FREQEX(I))**2)/FREQEX(I)
  335 CHISQ=CHISQ+CONTB(I)
      PRINT 340
  340 FORMAT (1H ,30X,11HUPPER LIMIT,5X,19HFREQUENCY(OBSERVED),5X,19HFRE
     1QUENCY(EXPECTED),5X,10HCHI SQUARE//)
      DO 350 I=LOP,KOP
      PRINT 345,UCL(I),FREQ(L,I),FREQEX(I),CONTB(I)
  345 FORMAT (1H ,30X,F10.4,5X,F10.0,14X,F13.3,11X,F10.3/)
  350 CONTINUE
      PRINT 355,SOBFQ,SEXPF,CHISQ
  355 FORMAT (1H ,25X,5HTOTAL,15X,F10.0,14X,F13.3,11X,F10.3//)
      PRINT 360,NDF
  360 FORMAT (1H ,25X,I5,19H DEGREES OF FREEDOM)
      PRINT 365
  365 FORMAT (1H1)
  370 CONTINUE
      BABY=(COAA-(ZAA**2/Q))
      CHOOK=(COBB-(ZAB**2/Q))
      HELP=(COCC-(ZAC**2/Q))
      TRAMP=(CODD-(ZAD**2/Q))
      STICK=(COEE-(ZAE**2/Q))
      BEER=(COFF-(ZAF**2/Q))
      CREEP=(ZAAB-((ZAA*ZAB)/Q))
      GRIFF=(ZAAC-((ZAA*ZAC)/Q))
      THORN=(ZAAD-((ZAA*ZAD)/Q))
      ARNY=(ZAAE-((ZAA*ZAE)/Q))
      SUHR=(ZAAF-((ZAA*ZAF)/Q))
      GENE=(ZABC-((ZAB*ZAC)/Q))
      GUESS=(ZABD-((ZAB*ZAD)/Q))
      WHAT=(ZABE-((ZAB*ZAE)/Q))
      YOYO=(ZABF-((ZAB*ZAF)/Q))
      BLOAK=(ZACD-((ZAC*ZAD)/Q))
      CROAK=(ZACE-((ZAC*ZAE)/Q))
      PRIDE=(ZACF-((ZAC*ZAF)/Q))
      ACNE=(ZADE-((ZAD*ZAE)/Q))
      DOOR=(ZADF-((ZAD*ZAF)/Q))
      SAND=(ZAEF-((ZAE*ZAF)/Q))
      ASK=CREEP/BABY
      BAT=GRIFF/BABY
```

```
      CAT=THORN/BABY                                                    SYSP1835
      DOZ=ARNY/BABY                                                     SYSP1840
      EAT=SUHR/BABY                                                     SYSP1845
      FAT=GENE/CHOOK                                                    SYSP1850
      GAG=GUESS/CHOOK                                                   SYSP1855
      HAT=WHAT/CHOOK                                                    SYSP1860
      PAT=YOYO/CHOOK                                                    SYSP1865
      QAT=BLOAK/HELP                                                    SYSP1870
      RAT=CROAK/HELP                                                    SYSP1875
      SAT=PRIDE/HELP                                                    SYSP1880
      TAT=ACNE/TRAMP                                                    SYSP1885
      UAT=DOOR/TRAMP                                                    SYSP1890
      VAT=SAND/STICK                                                    SYSP1895
      FUNNY=(ZAB-(ASK*ZAA))/Q                                           SYSP1900
      BUNNY=(ZAC-(BAT*ZAA))/Q                                           SYSP1905
      GUNNY=(ZAD-(CAT*ZAA))/Q                                           SYSP1910
      HUNNY=(ZAE-(DOZ*ZAA))/Q                                           SYSP1915
      PUNNY=(ZAF-(EAT*ZAA))/Q                                           SYSP1920
      QUNNY=(ZAC-(FAT*ZAB))/Q                                           SYSP1925
      RUNNY=(ZAD-(GAG*ZAB))/Q                                           SYSP1930
      SUNNY=(ZAE-(HAT*ZAB))/Q                                           SYSP1935
      TUNNY=(ZAF-(PAT*ZAB))/Q                                           SYSP1940
      VUNNY=(ZAD-(QAT*ZAC))/Q                                           SYSP1945
      WUNNY=(ZAE-(RAT*ZAC))/Q                                           SYSP1950
      XUNNY=(ZAF-(SAT*ZAC))/Q                                           SYSP1955
      YUNNY=(ZAE-(TAT*ZAD))/Q                                           SYSP1960
      ZUNNY=(ZAF-(UAT*ZAD))/Q                                           SYSP1965
      OUNNY=(ZAF-(VAT*ZAE))/Q                                           SYSP1970
      AVEX1=ZAA/Q                                                       SYSP1975
      AVEX2=ZAB/Q                                                       SYSP1980
      AVEX3=ZAC/Q                                                       SYSP1985
      AVEX4=ZAD/Q                                                       SYSP1990
      AVEX5=ZAE/Q                                                       SYSP1995
      AVEX6=ZAF/Q                                                       SYSP2000
      VARA=BABY/(Q-1.0)                                                 SYSP2005
      VARB=CHOOK/(Q-1.0)                                                SYSP2010
      VARC=HELP/(Q-1.0)                                                 SYSP2015
      VARD=TRAMP/(Q-1.0)                                                SYSP2020
      VARE=STICK/(Q-1.0)                                                SYSP2025
      VARF=BEER/(Q-1.0)                                                 SYSP2030
      STDVA=SQRT(VARA)                                                  SYSP2035
      STDVB=SQRT(VARB)                                                  SYSP2040
      STDVC=SQRT(VARC)                                                  SYSP2045
      STDVD=SQRT(VARD)                                                  SYSP2050
      STDVE=SQRT(VARE)                                                  SYSP2055
      STDVF=SQRT(VARF)                                                  SYSP2060
      R21=(CREEP**2)/(BABY*CHOOK)                                       SYSP2065
      R22=(GRIFF**2)/(BABY*HELP)                                        SYSP2070
      R23=(THORN**2)/(BABY*TRAMP)                                       SYSP2075
      R24=(ARNY**2)/(BABY*STICK)                                        SYSP2080
      R25=(SUHR**2)/(BABY*BEER)                                         SYSP2085
      R26=(GENE**2)/(CHOOK*HELP)                                        SYSP2090
      R27=(GUESS**2)/(CHOOK*TRAMP)                                      SYSP2095
      R28=(WHAT**2)/(CHOOK*STICK)                                       SYSP2100
      R29=(YOYO**2)/(CHOOK*BEER)                                        SYSP2105
      R210=(BLOAK**2)/(HELP*TRAMP)                                      SYSP2110
      R211=(CROAK**2)/(HELP*STICK)                                      SYSP2115
      R212=(PRIDE**2)/(HELP*BEER)                                       SYSP2120
      R213=(ACNE**2)/(TRAMP*STICK)                                      SYSP2125
      R214=(DOOR**2)/(TRAMP*BEER)                                       SYSP2130
      R215=(SAND**2)/(STICK*BEER)                                       SYSP2135
```

```
      R1=SQRT(R21)                                                      SYSP2140
      R2=SQRT(R22)                                                      SYSP2145
      R3=SQRT(R23)                                                      SYSP2150
      R4=SQRT(R24)                                                      SYSP2155
      R5=SQRT(R25)                                                      SYSP2160
      R6=SQRT(R26)                                                      SYSP2165
      R7=SQRT(R27)                                                      SYSP2170
      R8=SQRT(R28)                                                      SYSP2175
      R9=SQRT(R29)                                                      SYSP2180
      R10=SQRT(R210)                                                    SYSP2185
      R11=SQRT(R211)                                                    SYSP2190
      R12=SQRT(R212)                                                    SYSP2195
      R13=SQRT(R213)                                                    SYSP2200
      R14=SQRT(R214)                                                    SYSP2205
      R15=SQRT(R215)                                                    SYSP2210
      PRINT 375                                                         SYSP2215
  375 FORMAT (1H ,///60X,19HREGRESSION ANALYSIS//)                      SYSP2220
      PRINT 380                                                         SYSP2225
  380 FORMAT (1H ,1X,8HIDENTIF.,2X,12HSUMSQ DEV(X),2X,12HSUMSQ DEV(Y),2XSYSP2230
     1,11HSUMPR DEVXY,2X,8HAVERAGEX,2X,8HAVERAGEY,2X,7HSTANDVX,2X,7HSTANSYSP2235
     1DVY,3X,5HSLOPE,2X,9HINTERCEPT,2X,12HCOEFF DETERM,2X,8HCORRELTN)   SYSP2240
      IVAN=1                                                            SYSP2245
      MARG=2                                                            SYSP2250
      PRINT 385,IVAN,MARG,BABY,CHOOK,CREEP,AVEX1,AVEX2,STDVA,STDVB,ASK,FSYSP2255
     1UNNY,R21,R1                                                       SYSP2260
      MARG=3                                                            SYSP2265
      PRINT 385,IVAN,MARG,BABY,HELP,GRIFF,AVEX1,AVEX3,STDVA,STDVC,BAT,BUSYSP2270
     1NNY,R22,R2                                                        SYSP2275
      MARG=4                                                            SYSP2280
      PRINT 385,IVAN,MARG,BABY,TRAMP,THORN,AVEX1,AVEX4,STDVA,STDVD,CAT,GSYSP2285
     1UNNY,R23,R3                                                       SYSP2290
      MARG=5                                                            SYSP2295
      PRINT 385,IVAN,MARG,BABY,STICK,ARNY,AVEX1,AVEX5,STDVA,STDVE,DOZ,HUSYSP2300
     1NNY,R24,R4                                                        SYSP2305
      MARG=6                                                            SYSP2310
      PRINT 385,IVAN,MARG,BABY,BEER,SUHR,AVEX1,AVEX6,STDVA,STDVF,EAT,PUNSYSP2315
     1NY,R25,R5                                                         SYSP2320
      IVAN=2                                                            SYSP2325
      MARG=3                                                            SYSP2330
      PRINT 385,IVAN,MARG,CHOOK,HELP,GENE,AVEX2,AVEX3,STDVB,STDVC,FAT,QUSYSP2335
     1NNY,R26,R6                                                        SYSP2340
      MARG=4                                                            SYSP2345
      PRINT 385,IVAN,MARG,CHOOK,TRAMP,GUESS,AVEX2,AVEX4,STDVB,STDVD,GAG,SYSP2350
     1RUNNY,R27,R7                                                      SYSP2355
      MARG=5                                                            SYSP2360
      PRINT 385,IVAN,MARG,CHOOK,STICK,WHAT,AVEX2,AVEX5,STDVB,STDVE,HAT,SSYSP2365
     1UNNY,R28,R8                                                       SYSP2370
      MARG=6                                                            SYSP2375
      PRINT 385,IVAN,MARG,CHOOK,BEER,YOYO,AVEX2,AVEX6,STDVB,STDVF,PAT,TUSYSP2380
     1NNY,R29,R9                                                        SYSP2385
      IVAN=3                                                            SYSP2390
      MARG=4                                                            SYSP2395
      PRINT 385,IVAN,MARG,HELP,TRAMP,BLOAK,AVEX3,AVEX4,STDVC,STDVD,QAT,VSYSP2400
     1UNNY,R210,R10                                                     SYSP2405
      MARG=5                                                            SYSP2410
      PRINT 385,IVAN,MARG,HELP,STICK,CROAK,AVEX3,AVEX5,STDVC,STDVE,RAT,WSYSP2415
     1UNNY,R211,R11                                                     SYSP2420
      MARG=6                                                            SYSP2425
      PRINT 385,IVAN,MARG,HELP,BEER,PRIDE,AVEX3,AVEX6,STDVC,STDVF,SAT,XUSYSP2430
     1NNY,R212,R12                                                      SYSP2435
      IVAN=4                                                            SYSP2440
```

```
      MARG=5                                                            SYSP2445
      PRINT 385,IVAN,MARG,TRAMP,STICK,ACNE,AVEX4,AVEX5,STDVD,STDVE,TAT,YSYSP2450
     1UNNY,R213,R13                                                     SYSP2455
      MARG=6                                                            SYSP2460
      PRINT 385,IVAN,MARG,TRAMP,BEER,DOOR,AVEX4,AVEX6,STDVD,STDVF,UAT,ZUSYSP2465
     1NNY,R214,R14                                                      SYSP2470
      IVAN=5                                                            SYSP2475
      MARG=6                                                            SYSP2480
      PRINT 385,IVAN,MARG,STICK,BEER,SAND,AVEX5,AVEX6,STDVE,STDVF,VAT,OUSYSP2485
     1NNY,R215,R15                                                      SYSP2490
  385 FORMAT (1H ,/,I2,4H VS.,I2,2X,F12.6,2X,F12.6,2X,F11.5,3X,F9.4,1X,FSYSP2495
     19.4,1X,F8.4,1X,F8.4,1X,F7.3,2X,F9.4,4X,F7.5,5X,F6.5)              SYSP2500
  390 CONTINUE                                                          SYSP2505
C     RETURN TO NEXT ELEMENT OR STOP IF LAST RUN                        SYSP2510
      STOP                                                              SYSP2515
      END                                                               SYSP2520
```

```
    1
  25.4       50.0        6        3.
-13.00      .50    40
-13.00      .50    40
-13.00      .50    40
  0.0       .09    40
  0.0       .09    40
  0.0       .09    40
   1 0   1   1   1 2     1 0 4     1 6 0       9 1
   1 0   1   1   2 2     1 5 0     1 9 1       4 7
   1 0   1   2   1 2     1 8 4     2 1 4       9 0
   1 0   1   2   2 2     1 3 1     1 9 5       6 3
   1 0   1   3   1 2     1 0 4     2 1 0       3 4
   1 0   1   3   2 2     1 1 7     2 2 2       6 3
   1 0   1   4   1 2     1 3 3     2 2 7       9 6
   1 0   1   4   2 2     1 1 4     1 8 1       5 5
   1 0   1   5   1 2       8 1     1 9 4       4 6
   1 0   1   5   2 2     1 5 0     1 9 2       7 6
   1 0   2   1   1 2     2 7 4     3 2 7     1 1 1
   1 0   2   1   2 2     1 4 3     2 2 3       7 7
   1 0   2   2   1 2     1 2 1     1 2 4       7 4
   1 0   2   2   2 2       3 7       4 4       3 6
   1 0   2   3   1 2       4 7       8 6       4 2
   1 0   2   3   2 2       9 2     1 1 3       6 3
   1 0   2   4   1 2     1 3 4     1 7 4       4 6
   1 0   2   4   2 2       9 0     1 1 5       4 6
   1 0   2   5   1 2     1 0 1     1 8 5       7 6
   1 0   2   5   2 2       8 7     2 0 0       4 2
   1 0   3   1   1 2     1 0 4     1 5 0       4 7
   1 0   3   1   2 2       9 2     1 2 7       4 2
   1 0   3   2   1 2       9 3     1 3 0       5 5
   1 0   3   2   2 2     1 0 3     1 2 6       3 4
   1 0   3   3   1 2       9 1     3 6 1       6 2
   1 0   3   3   2 2       8 7     1 6 4       8 5
   1 0   3   4   1 2     1 2 5     2 4 3     1 0 7
   1 0   3   4   2 2     1 3 4     1 8 2       5 0
   1 0   3   5   1 2       5 0     1 5 5       2 0
   1 0   3   5   2 2       7 6     1 5 3       6 5
   1 0   4   1   1 2     1 0 1     1 6 3       4 7
   1 0   4   1   2 2       5 6     1 5 4       4 1
   1 0   4   2   1 2       6 1       8 2       4 2
   1 0   4   2   2 2       7 5       9 0       6 7
   1 0   4   3   1 2     1 3 5     3 0 1       4 4
   1 0   4   3   2 2       9 1     1 3 0       5 0
   1 0   4   4   1 2     1 0 0     1 4 1       3 2
   1 0   4   4   2 2       6 7     1 7 0       5 3
   1 0   4   5   1 2       5 7     1 0 4       5 1
   1 0   4   5   2 2       3 4       7 3       2 5
   1 0   5   1   1 2       4 1     1 2 6       2 4
   1 0   5   1   2 2       6 0     1 2 5       5 4
   1 0   5   2   1 2       5 4     1 2 2       5 1
   1 0   5   2   2 2       7 7       9 2       5 6
   1 0   5   3   1 2       7 0       9 0       4 7
   1 0   5   3   2 2       7 2       9 5       6 0
   1 0   5   4   1 2     5 1 4     5 6 7     1 1 4
   1 0   5   4   2 2     1 8 3     3 4 7       7 6
   1 0   5   5   1 2     1 1 1     1 2 1       7 3
   1 0   5   5   2 2       9 1     1 5 6       6 7
                   9
A-AXIS PHI TEST BEAR MEADOWS
B-AXIS PHI TEST BEAR MEADOWS
C-AXIS PHI TEST BEAR MEADOWS
B/A TEST BEAR MEADOWS
C/B TEST BEAR MEADOWS
C/A TEST BEAR MEADOWS
```

UNITS OF MEASUREMENT TO MILLIMETERS TO PHI-FREQUENCIES PLUS SUMMARY STATISTICS

KANSAS GEOLOGICAL SURVEY SHORT COURSE ON SAMPLING DECEMBER 1968

NUMBER OF VARIABLES = 6

IDENTIFICATION				A-PHI	B-PHI	C-PHI	B/A	C/B	C/A
10	1	1	1	-8.66675	-8.05907	-7.85658	0.65625	0.86905	0.57031
10	1	1	2	-8.92414	-8.57364	-6.95216	0.78431	0.32500	0.25490
10	1	2	1	-9.09302	-8.87621	-7.83668	0.86046	0.48649	0.41860
10	1	2	2	-8.96137	-8.38100	-7.33918	0.66879	0.48571	0.32484
10	1	3	1	-9.05907	-8.05907	-6.47411	0.50000	0.33333	0.16667
10	1	3	2	-9.14249	-8.23661	-7.33918	0.53371	0.53684	0.28652
10	1	4	1	-9.18245	-8.40822	-7.95216	0.58470	0.72897	0.42623
10	1	4	2	-8.84666	-8.19032	-7.15861	0.63448	0.48913	0.31034
10	1	5	1	-8.95216	-7.68912	-6.91468	0.41667	0.58462	0.24359
10	1	5	2	-8.93354	-8.57364	-7.62095	0.77922	0.51667	0.40260
10	2	1	1	-9.70567	-9.44811	-8.14249	0.83650	0.40455	0.33840
10	2	1	2	-9.15057	-8.51224	-7.64403	0.64246	0.54783	0.35196
10	2	2	1	-8.31061	-8.26667	-7.57364	0.97000	0.61856	0.60000
10	2	2	2	-6.83668	-6.62095	-6.57364	0.86111	0.96774	0.83333
10	2	3	1	-7.79604	-6.95216	-6.75422	0.55714	0.87179	0.48571
10	2	3	2	-8.17455	-7.87621	-7.33918	0.81319	0.68919	0.56044
10	2	4	1	-8.79604	-8.42164	-6.91468	0.77143	0.35185	0.27143
10	2	4	2	-8.20591	-7.83668	-6.91468	0.77419	0.52778	0.40860
10	2	5	1	-8.88592	-8.00660	-7.62095	0.54362	0.76543	0.41611
10	2	5	2	-8.98868	-7.81650	-6.75422	0.44375	0.47887	0.21250
10	3	1	1	-8.57364	-8.05907	-6.95216	0.70000	0.46429	0.32500
10	3	1	2	-8.35325	-7.87621	-6.75422	0.71845	0.45946	0.33010
10	3	2	1	-8.36719	-7.89557	-7.15861	0.72115	0.60000	0.43269
10	3	2	2	-8.33918	-8.04179	-6.47411	0.81373	0.33735	0.27451
10	3	3	1	-9.84168	-7.85658	-7.31061	0.25260	0.68493	0.17301
10	3	3	2	-8.71115	-7.81650	-7.77528	0.53788	0.97183	0.52273
10	3	4	1	-9.27408	-8.32496	-8.10970	0.51795	0.86139	0.44615
10	3	4	2	-8.85658	-8.42164	-6.98868	0.73973	0.37037	0.27397
10	3	5	1	-8.63254	-6.98868	-5.66675	0.32000	0.40000	0.12800
10	3	5	2	-8.60927	-7.62095	-7.39467	0.50406	0.85484	0.43089
10	4	1	1	-8.70018	-8.00660	-6.95216	0.61832	0.48148	0.29771
10	4	1	2	-8.62095	-7.19032	-6.71115	0.37097	0.71739	0.26613
10	4	2	1	-7.71115	-7.28146	-6.75422	0.74242	0.69388	0.51515
10	4	2	2	-7.83668	-7.59749	-7.44811	0.84722	0.90164	0.76389
10	4	3	1	-9.57964	-8.43494	-6.83668	0.45228	0.33028	0.14938
10	4	3	2	-8.36719	-7.85658	-6.98868	0.70192	0.54795	0.38462
10	4	4	1	-8.48693	-7.98868	-6.36719	0.70796	0.32500	0.23009
10	4	4	2	-8.75422	-7.44811	-7.09302	0.40441	0.78182	0.31618
10	4	5	1	-8.05907	-7.22134	-7.02431	0.55952	0.87234	0.48810
10	4	5	2	-7.54940	-6.47411	-6.05907	0.47458	0.75000	0.35593
10	5	1	1	-8.33918	-6.71115	-5.98868	0.32353	0.60606	0.19608
10	5	1	2	-8.32496	-7.25172	-7.12619	0.47525	0.91667	0.43564
10	5	2	1	-8.28146	-7.12619	-7.02431	0.44898	0.93182	0.41837
10	5	2	2	-7.87621	-7.64403	-7.19032	0.85135	0.73016	0.62162
10	5	3	1	-7.83668	-7.47411	-6.95216	0.77778	0.69643	0.54167
10	5	3	2	-7.93354	-7.52474	-7.25172	0.75325	0.82759	0.62338
10	5	4	1	-10.49648	-10.35325	-8.19032	0.90549	0.22330	0.20220
10	5	4	2	-9.79087	-8.86643	-7.62095	0.52688	0.42177	0.22222
10	5	5	1	-8.26667	-8.14249	-7.54940	0.91753	0.66292	0.60825
10	5	5	2	-8.64403	-7.85658	-7.44811	0.57937	0.75342	0.43651

A-AXIS PHI TEST BEAR MEADOWS INT = 0.500 8 CLASSES

CLASS NUMBER	UPPER LIMIT	MIDPOINT	FREQUENCY(OBSERVED)
1.	-10.0000	-10.2500	1.
2.	-9.5000	-9.7500	4.
3.	-9.0000	-9.2500	6.
4.	-8.5000	-8.7500	18.
5.	-8.0000	-8.2500	13.
6.	-7.5000	-7.7500	7.
7.	-7.0000	-7.2500	0.
8.	-6.5000	-6.7500	1.

MAXIMUM FREQUENCY	CLASS	MIDPOINT
18.	4	-8.750

AVERAGE	VARIANCE	STANDARD DEVIATION
-8.610	0.42040	0.6484

ROOT B1	B2	SUM OF FREQUENCIES
-0.034	3.510	50.

****

FIT TO THE NORMAL DISTRIBUTION

UPPER LIMIT	FREQUENCY(OBSERVED)	FREQUENCY(EXPECTED)	CHI SQUARE
-10.0000	1.	0.801	0.050
-9.5000	4.	3.446	0.089
-9.0000	6.	9.441	1.254
-8.5000	18.	14.680	0.751
-8.0000	13.	12.962	0.000
-7.5000	7.	6.497	0.039
-7.0000	0.	1.848	1.848
-6.5000	1.	0.296	1.669
TOTAL	50.	49.971	5.700

5 DEGREES OF FREEDOM

B-AXIS PHI TEST BEAR MEADOWS   INT = 0.500   9 CLASSES

CLASS NUMBER	UPPER LIMIT	MIDPOINT	FREQUENCY(OBSERVED)
1.	-10.0000	-10.2500	1.
2.	-9.5000	-9.7500	0.
3.	-9.0000	-9.2500	1.
4.	-8.5000	-8.7500	5.
5.	-8.0000	-8.2500	16.
6.	-7.5000	-7.7500	15.
7.	-7.0000	-7.2500	7.
8.	-6.5000	-6.7500	4.
9.	-6.0000	-6.2500	1.

MAXIMUM FREQUENCY	CLASS	MIDPOINT
16.	5	-8.250

AVERAGE	VARIANCE	STANDARD DEVIATION
-7.910	0.48440	0.6960

ROOT B1	B2	SUM OF FREQUENCIES
-0.351	4.439	50.

****

FIT TO THE NORMAL DISTRIBUTION

UPPER LIMIT	FREQUENCY(OBSERVED)	FREQUENCY(EXPECTED)	CHI SQUARE
-10.0000	1.	0.067	12.965
-9.5000	0.	0.491	0.491
-9.0000	1.	2.375	0.796
-8.5000	5.	6.981	0.562
-8.0000	16.	12.513	0.972
-7.5000	15.	13.677	0.128
-7.0000	7.	9.119	0.492
-6.5000	4.	3.707	0.023
-6.0000	1.	0.918	0.007
TOTAL	50.	49.848	16.437

6 DEGREES OF FREEDOM

C-AXIS PHI TEST BEAR MEADOWS   INT = 0.500   6 CLASSES

CLASS NUMBER	UPPER LIMIT	MIDPOINT	FREQUENCY(OBSERVED)
1.	-8.0000	-8.2500	3.
2.	-7.5000	-7.7500	10.
3.	-7.0000	-7.2500	15.
4.	-6.5000	-6.7500	16.
5.	-6.0000	-6.2500	4.
6.	-5.5000	-5.7500	2.

MAXIMUM FREQUENCY	CLASS	MIDPOINT
16.	4	-6.750

AVERAGE	VARIANCE	STANDARD DEVIATION
-7.110	0.34040	0.5834

ROOT B1	B2	SUM OF FREQUENCIES
0.122	2.770	50.

****

FIT TO THE NORMAL DISTRIBUTION

UPPER LIMIT	FREQUENCY(OBSERVED)	FREQUENCY(EXPECTED)	CHI SQUARE
-8.0000	3.	3.179	0.010
-7.5000	10.	9.417	0.036
-7.0000	15.	16.142	0.081
-6.5000	16.	13.867	0.328
-6.0000	4.	5.967	0.648
-5.5000	2.	1.283	0.401
TOTAL	50.	49.855	1.504

3 DEGREES OF FREEDOM

B/A TEST BEAR MEADOWS INT = 0.090 9 CLASSES

CLASS NUMBER	UPPER LIMIT	MIDPOINT	FREQUENCY(OBSERVED)
1.	0.2700	0.2250	1.
2.	0.3600	0.3150	2.
3.	0.4500	0.4050	5.
4.	0.5400	0.4950	9.
5.	0.6300	0.5850	6.
6.	0.7200	0.6750	8.
7.	0.8100	0.7650	9.
8.	0.9000	0.8550	7.
9.	0.9900	0.9450	3.

MAXIMUM FREQUENCY	CLASS	MIDPOINT
9.	4	0.495

AVERAGE	VARIANCE	STANDARD DEVIATION
0.639	0.03240	0.1800

ROOT B1	B2	SUM OF FREQUENCIES
-0.192	2.175	50.

****

FIT TO THE NORMAL DISTRIBUTION

UPPER LIMIT	FREQUENCY(OBSERVED)	FREQUENCY(EXPECTED)	CHI SQUARE
0.2700	1.	1.009	0.000
0.3600	2.	2.020	0.000
0.4500	5.	4.314	0.109
0.5400	9.	7.216	0.441
0.6300	6.	9.445	1.256
0.7200	8.	9.679	0.291
0.8100	9.	7.766	0.196
0.9000	7.	4.876	0.926
0.9900	3.	2.398	0.151
TOTAL	50.	48.721	3.371

6 DEGREES OF FREEDOM

C/B TEST BEAR MEADOWS   INT = 0.090   9 CLASSES

CLASS NUMBER	UPPER LIMIT	MIDPOINT	FREQUENCY(OBSERVED)
1.	0.2700	0.2250	1.
2.	0.3600	0.3150	6.
3.	0.4500	0.4050	4.
4.	0.5400	0.4950	10.
5.	0.6300	0.5850	6.
6.	0.7200	0.6750	6.
7.	0.8100	0.7650	6.
8.	0.9000	0.8550	6.
9.	0.9900	0.9450	5.

MAXIMUM FREQUENCY	CLASS	MIDPOINT
10.	4	0.495

AVERAGE	VARIANCE	STANDARD DEVIATION
0.614	0.04129	0.2032

ROOT B1	B2	SUM OF FREQUENCIES
0.038	1.923	50.

****

FIT TO THE NORMAL DISTRIBUTION

	UPPER LIMIT	FREQUENCY(OBSERVED)	FREQUENCY(EXPECTED)	CHI SQUARE
	0.2700	1.	2.267	0.708
	0.3600	6.	3.025	2.926
	0.4500	4.	5.213	0.282
	0.5400	10.	7.408	0.907
	0.6300	6.	8.677	0.826
	0.7200	6.	8.380	0.676
	0.8100	6.	6.675	0.068
	0.9000	6.	4.381	0.598
	0.9900	5.	2.372	2.911
TOTAL		50.	48.397	9.902

6 DEGREES OF FREEDOM

C/A TEST BEAR MEADOWS   INT = 0.090   9 CLASSES

CLASS NUMBER	UPPER LIMIT	MIDPOINT	FREQUENCY(OBSERVED)
1.	0.1800	0.1350	4.
2.	0.2700	0.2250	8.
3.	0.3600	0.3150	13.
4.	0.4500	0.4050	12.
5.	0.5400	0.4950	4.
6.	0.6300	0.5850	7.
7.	0.7200	0.6750	0.
8.	0.8100	0.7650	1.
9.	0.9000	0.8550	1.

MAXIMUM FREQUENCY	CLASS	MIDPOINT
13.	3	0.315

AVERAGE	VARIANCE	STANDARD DEVIATION
0.380	0.02431	0.1559

ROOT B1	B2	SUM OF FREQUENCIES
0.735	3.563	50.

****

FIT TO THE NORMAL DISTRIBUTION

UPPER LIMIT	FREQUENCY(OBSERVED)	FREQUENCY(EXPECTED)	CHI SQUARE
0.1800	4.	5.002	0.201
0.2700	8.	7.031	0.134
0.3600	13.	10.441	0.627
0.4500	12.	11.212	0.055
0.5400	4.	8.709	2.546
0.6300	7.	4.891	0.910
0.7200	0.	1.987	1.987
0.8100	1.	0.583	0.299
0.9000	1.	0.123	6.233
TOTAL	50.	49.978	12.991

6 DEGREES OF FREEDOM

REGRESSION ANALYSIS

IDENTIF.	SUMSQ DEV(X)	SUMSQ DEV(Y)	SUMPR DEVXY	AVERAGEX	AVERAGEY	STANDVX	STANDVY	SLOPE	INTERCEPT	COEFF DETERM	CORRELTN
1 VS. 2	20.595947	23.626953	17.36060	-8.6325	-7.9231	0.6483	0.6944	0.843	-0.6467	0.61935	.78699
1 VS. 3	20.595947	14.562744	8.34717	-8.6325	-7.1367	0.6483	0.5452	0.405	-3.6381	0.23230	.48198
1 VS. 4	20.595947	1.506409	1.21851	-8.6325	0.6379	0.6483	0.1753	0.059	1.1486	0.04786	.21876
1 VS. 5	20.595947	2.017731	3.41626	-8.6325	0.6151	0.6483	0.2029	0.166	2.0470	0.28084	.52994
1 VS. 6	20.595947	1.212892	3.30801	-8.6325	0.3859	0.6483	0.1573	0.161	1.7724	0.43806	.66186
2 VS. 3	23.626953	14.562744	12.52979	-7.9231	-7.1367	0.6944	0.5452	0.530	-2.9350	0.45629	.67549
2 VS. 4	23.626953	1.506409	-2.50938	-7.9231	0.6379	0.6944	0.1753	-0.106	-0.2036	0.17692	.42062
2 VS. 5	23.626953	2.017731	4.15788	-7.9231	0.6151	0.6944	0.2029	0.176	2.0094	0.36264	.60219
2 VS. 6	23.626953	1.212892	1.50914	-7.9231	0.3859	0.6944	0.1573	0.064	0.8919	0.07947	.28191
3 VS. 4	14.562744	1.506409	-1.60268	-7.1367	0.6379	0.5452	0.1753	-0.110	-0.1475	0.11709	.34218
3 VS. 5	14.562744	2.017731	-0.90485	-7.1367	0.6151	0.5452	0.2029	-0.062	0.1717	0.02786	.16693
3 VS. 6	14.562744	1.212892	-1.27335	-7.1367	0.3859	0.5452	0.1573	-0.087	-0.2382	0.09180	.30298
4 VS. 5	1.506409	2.017731	-0.32661	0.6379	0.6151	0.1753	0.2029	-0.217	0.7534	0.03510	.18734
4 VS. 6	1.506409	1.212892	0.71209	0.6379	0.3859	0.1753	0.1573	0.473	0.0843	0.27753	.52681
5 VS. 6	2.017731	1.212892	1.11101	0.6151	0.3859	0.2029	0.1573	0.551	0.0472	0.50437	.71019

KANSAS GEOLOGICAL SURVEY COMPUTER PROGRAM
THE UNIVERSITY OF KANSAS, LAWRENCE

PROGRAM ABSTRACT

Title (If subroutine state in title):

SYSTEMS PROGRAM ONE

Date: 1 December 1968

Author, organization: John C. Griffiths, The Pennsylvania State University and
Charles W. Ondrick, Kansas Geological Survey

Direct inquiries to: Authors, or

Name: Daniel F. Merriam Address: Kansas Geological Survey
Univ. of Kansas, Lawrence, Kans. 66044

Purpose/description: Computes frequencies, summary statistics, expected normal distribution, chi-square and linear regression-correlation for six variables. An option is provided to accept either Port-a-Punch or standard Hollerith cards.

Mathematical method:

Restrictions, range: The number of sets of data processable is limited only by the time and/or records available to the user on a particular system.

Computer manufacturer: IBM or GE Model: 360/67 or 625

Programming language: FORTRAN IV

Memory required: 10 K Approximate running time:

Special peripheral equipment required: none

Remarks (special compilers or operating systems, required word lengths, number of successful runs, other machine versions, additional information useful for operation or modification of program)

Port-a-Punch cards should be repunched in peripheral card reproducers to avoid card jam on high speed (>300 cards/minute) card readers.

COMPUTER CONTRIBUTIONS

Kansas Geological Survey

1. Mathematical simulation of marine sedimentation with IBM 7090/7094 computers, by J.W. Harbaugh, 1966 . . . . . $1.00
2. A generalized two-dimensional regression procedure, by J.R. Dempsey, 1966 . . . . . $0.50
3. FORTRAN IV and MAP program for computation and plotting of trend surfaces for degrees 1 through 6, by Mont O'Leary, R.H. Lippert, and O.T. Spitz, 1966 . . . . . $0.75
4. FORTRAN II program for multivariate discriminant analysis using an IBM 1620 computer, by J.C. Davis and R.J. Sampson, 1966 . . . . . $0.50
5. FORTRAN IV program using double Fourier series for surface fitting of irregularly spaced data, by W.R. James, 1966 . . . . . $0.75
6. FORTRAN IV program for estimation of cladistic relationships using the IBM 7040, by R.L. Bartcher, 1966 . . . . . $1.00
7. Computer applications in the earth sciences: Colloquium on classification procedures, edited by D.F. Merriam, 1966 . . . . . $1.00
8. Prediction of the performance of a solution gas drive reservoir by Muskat's Equation, by Apolonio Baca, 1967. . . . . $1.00
9. FORTRAN IV program for mathematical simulation of marine sedimentation with IBM 7040 or 7094 computers, by J.W. Harbaugh and W.J. Wahlstedt, 1967 . . . . . $1.00
10. Three-dimensional response surface program in FORTRAN II for the IBM 1620 computer, by R.J. Sampson and J.C. Davis, 1967 . . . . . $0.75
11. FORTRAN IV program for vector trend analyses of directional data, by W.T. Fox, 1967 . . $1.00
12. Computer applications in the earth sciences: Colloquium on trend analysis, edited by D.F. Merriam and N.C. Cocke, 1967 . . . . . $1.00
13. FORTRAN IV computer programs for Markov chain experiments in geology, by W.C. Krumbein, 1967 . . . . . $1.00
14. FORTRAN IV programs to determine surface roughness in topography for the CDC 3400 computer, by R.D. Hobson, 1967 . . . . . $1.00
15. FORTRAN II program for progressive linear fit of surfaces on a quadratic base using an IBM 1620 computer, by A.J. Cole, C. Jordan, and D.F. Merriam, 1967 . . . . . $1.00
16. FORTRAN IV program for the GE 625 to compute the power spectrum of geological surfaces, by J.E. Esler and F.W. Preston, 1967. . . . . $0.75
17. FORTRAN IV program for Q-mode cluster analysis of nonquantitative data using IBM 7090/7094 computers, by G.F. Bonham-Carter, 1967 . . . . . $1.00
18. Computer applications in the earth sciences: Colloquium on time-series analysis, D.F. Merriam, editor, 1967 . . . . . $1.00
19. FORTRAN II time-trend package for the IBM 1620 computer, by J.C. Davis and R.J. Sampson, 1967 . . . . . $1.00
20. Computer programs for multivariate analysis in geology, D.F. Merriam, editor, 1968 . . . $1.00
21. FORTRAN IV program for computation and display of principal components, by W.J. Wahlstedt and J.C. Davis, 1968 . . . . . $1.00
22. Computer applications in the earth sciences: Colloquium on simulation, D.F. Merriam and N.C. Cocke, editors, 1968 . . . . . $1.00
23. Computer programs for automatic contouring, by D.B. McIntyre, D.D. Pollard, and R. Smith, 1968 . . . . . $1.50
24. Mathematical model and FORTRAN IV program for computer simulation of deltaic sedimentation, by G.F. Bonham-Carter and A.J. Sutherland, 1968 . . . . . $1.00
25. FORTRAN IV CDC 6400 computer program for analysis of subsurface fold geometry, by E.H.T. Whitten, 1968 . . . . . $1.00
26. FORTRAN IV computer program for simulation of transgression and regression with continuous-time Markov models, by W.C. Krumbein, 1968 . . . . . $1.00
27. Stepwise regression and nonpolynomial models in trend analysis, by A.T. Miesch and J.J. Connor, 1968. . . . . $1.00
28. KWIKR8 a FORTRAN IV program for multiple regression and geologic trend analysis, by J.E. Esler, P.F. Smith, and J.C.Davis, 1968 . . . . . $1.00
29. FORTRAN IV program for harmonic trend analysis using double Fourier series and regularly gridded data for the GE 625 computer, by J.W. Harbaugh and M.J. Sackin, 1968 . . . $1.00
30. Sampling a geological population (workshop on experiment in sampling), by J.C. Griffiths and C.W. Ondrick, 1968 . . . . . $1.00