ClibPDF - www.fastio.com

ClibPDF - www.fastio.com

TECHNICAL REPORT STANDARD TITLE PAGE

1. REPORT NO.	2. GOVERNMENT ACCESSION NO.	3. RECIPIENT'S CATALOG NO.
4. TITLE AND SUBTITLE Development of a Correlation Program		5. REPORT DATE June 1975
		6. PERFORMING ORGANIZATION CODE 19107-762503-631153
7. AUTHOR(S) Benson, Paul E.		8. PERFORMING ORGANIZATION REPORT NO. CA-DOT-TL-1153-5-75-06
9. PERFORMING ORGANIZATION NAME AND ADDRESS Transportation Laboratory 5900 Folsom Boulevard Sacramento, California 95819		10. WORK UNIT NO.
		11. CONTRACT OR GRANT NO. F-4-20
12. SPONSORING AGENCY NAME AND ADDRESS California Department of Transportation Division of Construction and Research Sacramento, California 95807		13. TYPE OF REPORT & PERIOD COVERED Interim Report
		14. SPONSORING AGENCY CODE

15. SUPPLEMENTARY NOTES

16. ABSTRACT

A statewide correlation program aimed at improving test method reproducibility is discussed. Statistical methods such as Analysis of Variance, Regression Analysis, Principal Components Analysis and Nonparametric Analysis are used for developing an integrated program of interlaboratory testing. Four computer programs using these methods are presented. Guidelines for conducting a correlation program are given.

17. KEY WORDS Correlation, interlaboratory testing, reproducibility, statistical methods, analysis of variance, regression analysis, guidelines, research & development, computer applications, statistics.	18. DISTRIBUTION STATEMENT Unlimited

19. SECURITY CLASSIF. (OF THIS REPORT) Unclassified	20. SECURITY CLASSIF. (OF THIS PAGE) Unclassified	21. NO. OF PAGES 50	22. PRICE

MBC 1242 (ORIG. 9-72)

75-06

ClibPDF - www.fastio.com

ClibPDF - www.fastio.com

STATE OF CALIFORNIA
DEPARTMENT OF TRANSPORTATION
DIVISION OF CONSTRUCTION AND RESEARCH
TRANSPORTATION LABORATORY

June 1975

FHWA No. F-4-20
TL No. 631153

Mr. R. J. Datel
Chief Engineer

Dear Sir:

I have approved and now submit for your information this interim research project report titled:

DEVELOPMENT OF
A CORRELATION PROGRAM

Study made by . . . . . General Services Branch

Under the Supervision of . . . . . W. H. Ames, P.E. and
C. R. Sundquist, P.E.

Principal Investigator . . . . . P. E. Benson

Report Prepared by . . . . . P. E. Benson

Very truly yours,

George A. Hill

GEORGE A. HILL
Chief,
Office of Transportation Laboratory

Attachment

ClibPDF - www.fastio.com

ClibPDF - www.fastio.com

TABLE OF CONTENTS

	Page
INTRODUCTION	1
FINDINGS AND CONCLUSIONS	2
IMPLEMENTATION	2
THE DESIGN	3
THE PROGRAMS	6
"Scat"	7
"Aggva"	18
"Interlab"	24
"Ranklab"	36
REFERENCES	42
GLOSSARY	
APPENDIX	A-1

ClibPDF - www.fastio.com

ClibPDF - www.fastio.com

ACKNOWLEDGEMENTS

The author wishes to express his appreciation to Messrs. Carl Sundquist, Wallace Ames, and Richard Trimble for the opportunity to explore this interesting topic in detail and for their helpful review of this report. Another whose comments, suggestions and leadership in the field of interlaboratory correlation contributed significantly to this report was Dr. John Mandel, Statistical Consultant for the National Bureau of Standards. Also, special thanks go to Mr. Dilford Onodera for his assistance in debugging the computer programs.

This project was conducted in cooperation with the Federal Highway Administration under Agreement No. F-4-20. The contents of this report reflect the views of the Transportation Laboratory which is responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the State of California or the Federal Highway Administration. This report does not constitute a standard, specification, or regulation.

ClibPDF - www.fastio.com

ClibPDF - www.fastio.com

INTRODUCTION

Correlating materials test methods has long been a part of the Department of Transportation's policy of self-inspection. The responsibility for carrying out this policy rests with those who design and maintain test methods since they are best prepared to study their accuracy, precision and effectiveness. Hence, the Transportation Laboratory, while itself engaging in several nationwide correlation programs (AASHTO Materials Reference Laboratory and the NBS Cement and Concrete Reference Laboratory), implemented correlation work between district laboratories within the California Department of Transportation. Participation by the district laboratories in the past, however, has not been uniform and the results have yielded limited conclusions.

The reason for this lack of a more effective correlation program was threefold. First, no integrated correlation program design had been proposed. Second, effective analytical tools were not readily available. Third, no formal system had been provided to motivate more comprehensive district participation.

This report is addressed to the first two of these deficiencies. By combining statistical techniques with a flexible time-share computer system, an interlaboratory correlation design which is both versatile and practical has been developed and is presented in the first section of the report. The second section briefly explains the theory and operation of four computer programs written to analyze the data generated by the design. Those intending to use these programs are encouraged to study all sections carefully.

The capabilities of this correlation program package have been proven by a one year pilot study. In fact, all the examples shown are derived from actual test results obtained during the pilot study. A second report treats these results in more detail(2).

ClibPDF - www.fastio.com

Copies of the programs may be obtained by writing to:

California Department of Transportation
Transportation Laboratory
5900 Folsom Boulevard
Sacramento, California 95819

FINDINGS AND CONCLUSIONS

As a direct result of these and related investigations, it has been determined that:

1. Testing bias due to different operators and laboratories can be isolated;

2. Performance of individual laboratories can be monitored closely with a minimum of testing; and

3. The relative magnitude of general error types such as operator errors, random errors or equipment errors can often be determined.

It is the conclusion of the author that many tangible and useful results regarding test precision can be obtained with a reasonable expenditure of manpower by implementing a correlation program. These results will not only provide a measure of test precision, but will also indicate areas where test precision can be improved. The decision to make improvements will, of course, depend largely on the magnitude of the testing errors in comparison with the overall variation in test results.

IMPLEMENTATION

A permanent correlation program similar to the one described herein has been implemented by the California Department of

ClibPDF - www.fastio.com

Transportation. The program incorporates some of the basic concepts mentioned in this report, but is based on a more simplified design. While this design limits the ability to isolate systematic sources of error such as operator or equipment, it allows more laboratories and test methods to be studied at a reasonable cost. California is offering the program at cost to city and county laboratories throughout the state. It is hoped that the results will:

1) Establish on-going measures of test precision for the California community of testing laboratories;

2) Identify problem areas that may exist;

3) Generate aggressive interest in improving test precision;

4) Promote uniformity in the application of materials test methods; and

5) Generate a more competitive spirit for assessing the relative validity or signifiance of given test results in case of legal actions.

THE DESIGN

The success of a correlation program depends, to a large degree, upon the flexibility and feasibility of its design. More specifically the design should:

1. Provide the amount and kinds of data needed to test with reasonable accuracy the pertinent hypotheses;

2. Cover the full range of material normally encountered by each test method;

ClibPDF - www.fastio.com

3. Use manageable sample sizes; and

4. Spread testing workloads over a reasonable period of time.

It is apparent from these criteria that a balance between the ideal amount of data required and the practical constraints of logistics must be achieved. To do this, studies of prior routine testing data were made under this project which yielded approximations of the sensitivity of proposed analytical techniques as functions of sample size. Practical considerations of test scheduling, sample preparation and costs were then used in selecting an optimum sample size. Once this size was determined, the details of scheduling and procedure were worked out in accordance with the criteria listed above. A description of the design follows.

Two large samples representing a common material type such as AC or PCC aggregate were prepared by the Transportation Laboratory every three months. These samples were obtained from different sources, but had similar properties.

Each sample was split down into 64 sub-samples. Each of these sub-samples consisted of enough material to run the full set of tests required. Forty-eight of the sub-samples were randomly assigned tags denoting the laboratory to which they were to be sent. The remaining 16 sub-samples were kept as a contingency. Thus, each of the 12 participating laboratories was randomly assigned 4 sub-samples from each sample (see Figure 1). This could be expanded for larger samples if more laboratories were to be included in the correlation program.

At the beginning of each three-month interval each of the participating laboratories received two sets of four sub-samples each. When time permitted, they each chose two operators and set aside

ClibPDF - www.fastio.com

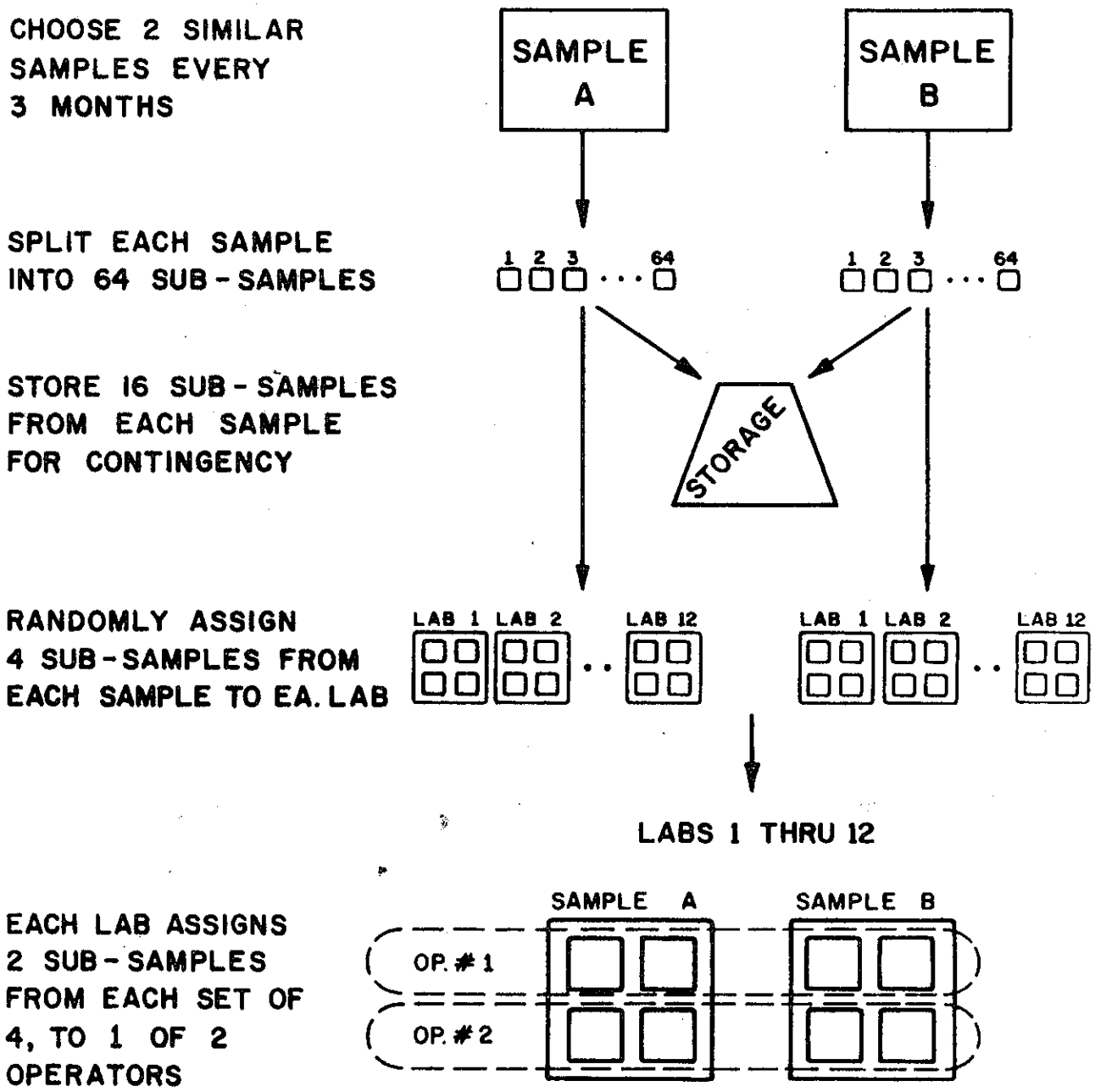


Figure 1 SAMPLE DISTRIBUTION

ClibPDF - www.fastio.com

one set of equipment. On the day or days that the tests were to be run, each operator was given two sub-samples from each sample. The operators ran all the indicated tests on each individual sub-sample following their usual procedure. Any deviations from the published test method were noted. Each operator used the same set of equipment for the four sub-samples he tested. Ideally, the two operators ran all their tests on the same day. Their results were recorded on a specially designed form and submitted to the Transportation Laboratory.

When designing the form to be used for recording results, several important things were noted. First, the order and placement of results on the form should provide for direct keypunching of results for computer submittal. Second, it is often important to encourage greater numerical accuracy in reporting results than is normally required. For instance, sieve analysis results should be recorded to the nearest 0.1% rather than the nearest 1% called for by the test method. If this is not done, the sensitivity of the analysis can be affected. Third, an area to record calculations and interim results should be provided. Also, laboratories should be required to submit their original worksheets. Often, important details are omitted from copies. Fourth, operators should be identified, and fifth, remarks encouraged and provided for.

THE PROGRAMS

The correlation programs that were developed are called "Scat", "Aggva", "Interlab", and "Ranklab". Each will now be discussed in detail.

ClibPDF - www.fastio.com

"Scat"

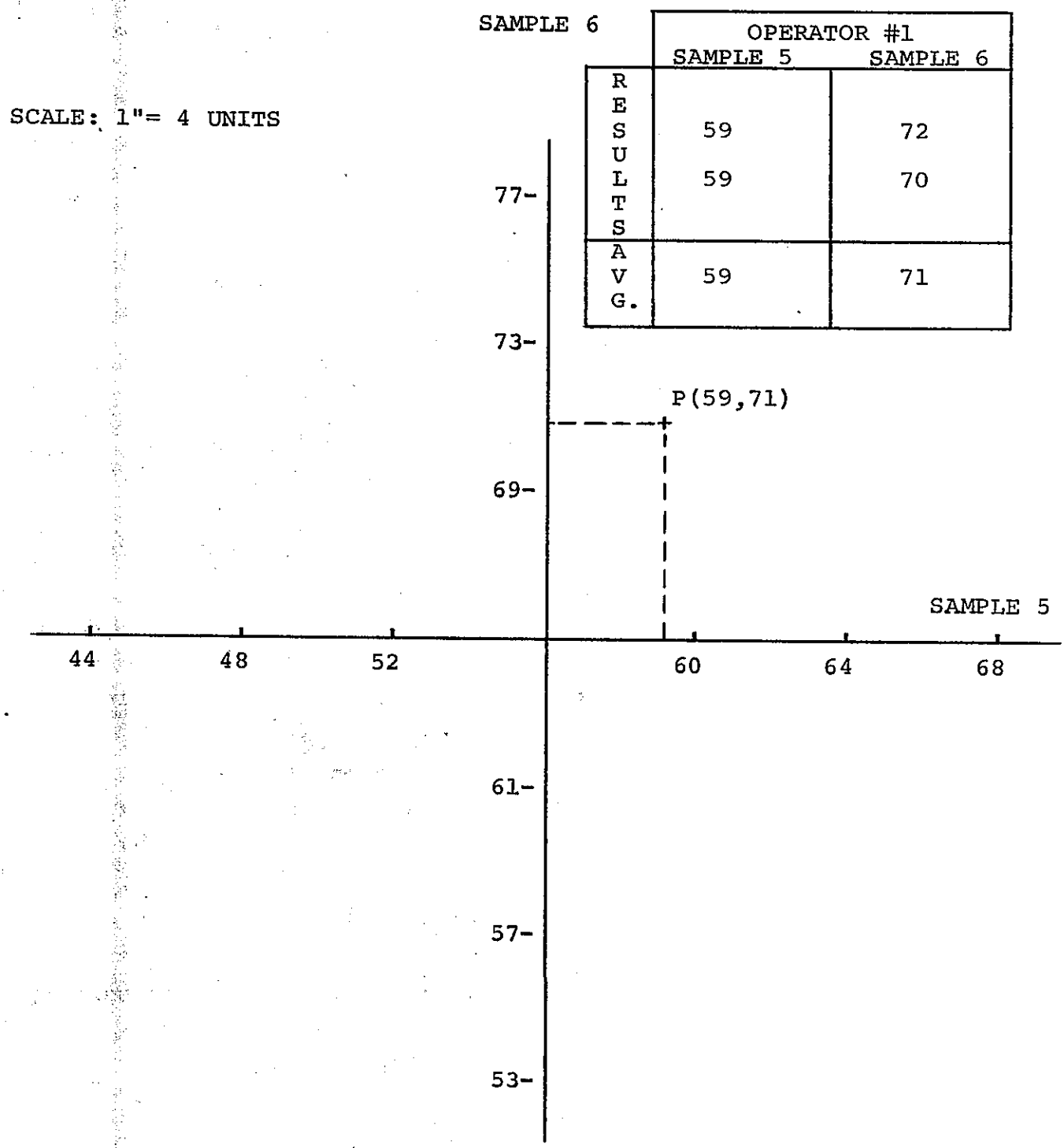
The correlation design previously discussed suggests that two samples having similar properties be sent to each participating laboratory. This would constitute one cycle in a continuous correlation program. These two samples are perfectly suited to a graphical display of the dispersion of results called the scatter diagram(15).

The main advantage of the scatter diagram is its simplicity. Many qualitative conclusions can be drawn from the diagram if one simple principle is remembered: Systematic error, if present and measurable, will be consistent for both samples. More illustratively stated, any operator-equipment combination obtaining abnormally high results from one sample is expected to obtain high results from all similar samples. With this in mind, the scatter diagram can be defined as it pertains to this specific design.

The scatter diagram is simply a cartesian coordinate representation of a group of data that can be divided into two distinct sub-groups. In this case these two sub-groups are the two samples for which each operator-equipment combination performs replicate tests. To construct the diagram, the average of an operator's two results on the first sample is assigned as his x-coordinate. His y-coordinate is assigned in the same way, but is based on his results for the second sample (see Figure 2). A set of axes covering the range of results for all operators on each sample are drawn with the origin at the point representing the grand means of the two samples for all operators. These axes divide the "scattered" data points into four separate quadrants as labeled in Figure 3.

If the test method being studied had completely flawless precision and the sub-samples are all identical, all points would

ClibPDF - www.fastio.com



	OPERATOR #1 SAMPLE 5	OPERATOR #1 SAMPLE 6
RESULTS	59 59	72 70
AVG.	59	71

Figure 2 CONSTRUCTION OF SCATTER DIAGRAM

ClibPDF - www.fastio.com

S C A T T E R D I A G R A M

EXAMPLE

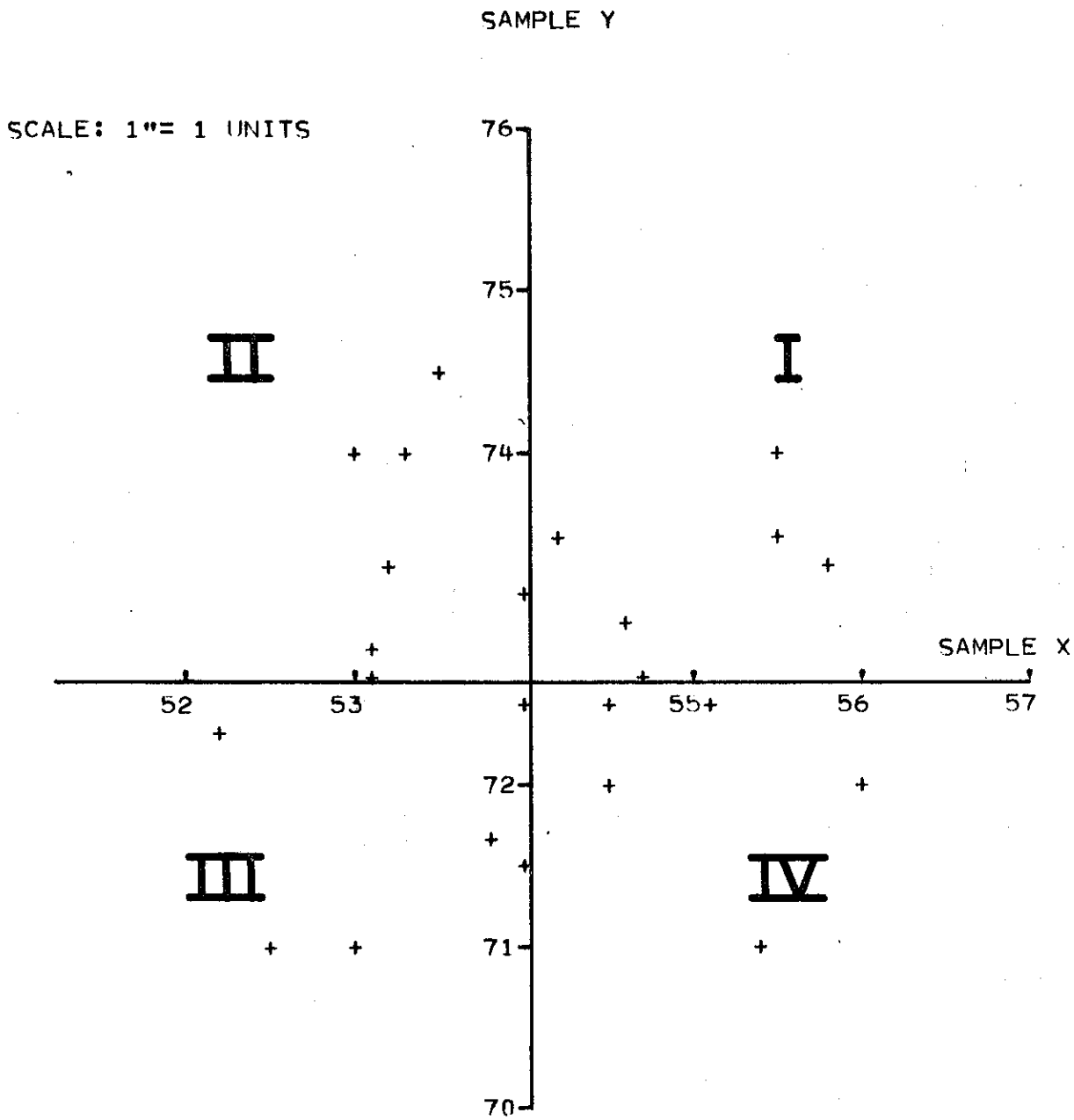


Figure 3 QUADRANTS AND TYPICAL SCATTER FROM TEST EXHIBITING RANDOM ERROR ONLY

ClibPDF - www.fastio.com

be expected to coincide at the origin. Barring this implausible configuration, however, the random error inherent in the test will tend to cause the points to form a circular cluster about the origin as shown in Figure 3. The radius of this cluster is an indication of the precision of the test. If, however, a systematic or consistent error in the results is present and measurable, then quadrants I and III will contain significantly more points than quadrants II and IV (see Figure 4). This follows from the basic principle already mentioned: Operators obtaining consistently high or low results on one sample usually obtain high or low results, respectively, on other similar samples. Returning to the ideal case, if all scattering effects could be eliminated, the scatter diagram of a test method exhibiting systematic error would have all its points lying on a line intersecting the origin and quadrants I and III. This line will be termed the major axis.

The "Scat" program provides a two page output for each test analyzed. The first page consists of the scatter diagram itself along with several pertinent statistics (see Figure 5). The second page contains an overall analysis and laboratory performance summary (see Table 1).

For our specific design, which includes 12 participating laboratories, there are 24 separate data points plotted on the first page. Letters of the alphabet represent the 12 different laboratories. Since there are two operators participating from each laboratory, each letter appears twice. The scale, automatically set by the computer, is given in the upper left corner of the diagram. The axes are labeled for the sample they represent. The statistics given for each sample on the first page are: the mean (the overall average result for that sample), the range (the difference between the extreme high

ClibPDF - www.fastio.com

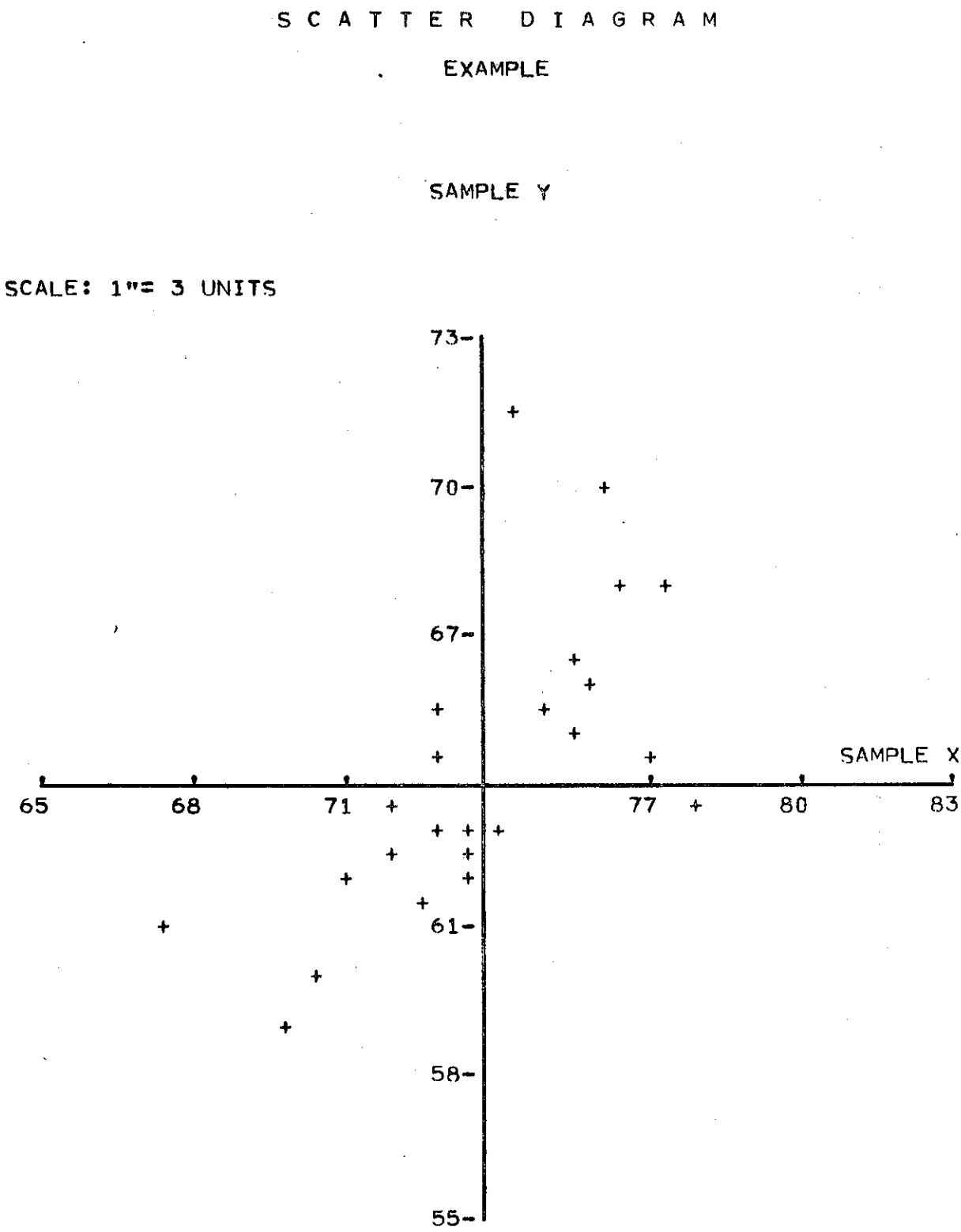


Figure 4 TYPICAL SCATTER FROM TEST EXHIBITING SIGNIFICANT SYSTEMATIC ERROR

ClibPDF - www.fastio.com

TRANSLAB CORRELATION PROGRAM

TEST METHOD: FINE DURABILITY MATERIAL: PCC
DATE: 08/14/74

	SAMPLE 1	SAMPLE 2
MEAN	73.8	64.2
RANGE	10.5	12.5
STANDARD DEVIATION	2.55	3.06

PHI (RADIANS) = .917

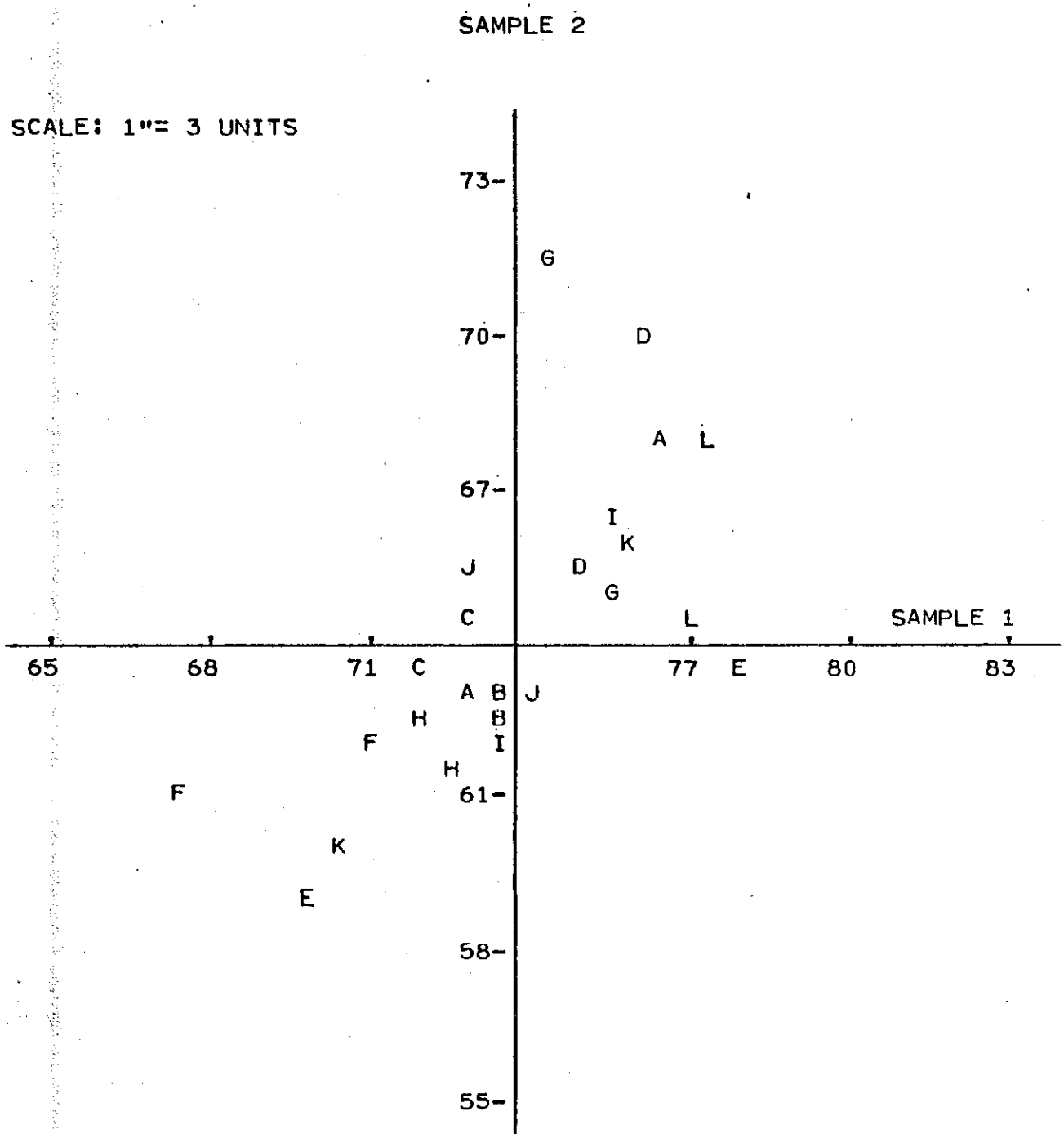


Figure 5 FIRST PAGE OF "SCAT" OUTPUT

ClibPDF - www.fastio.com

TRANSLAB CORRELATION PROGRAM
LABORATORY PERFORMANCE SUMMARY

TEST METHOD: FINE DURABILITY MATERIAL: PCC
DATE: 01/03/75

OVERALL RESULTS

MAJOR AXIS		MINOR AXIS			COMPONENTS OF VARIANCE	
VARIANCE	SIGMA	VARIANCE	SIGMA	R-RATIO	RANDOM	SYSTEMATIC
13.38	3.66	2.49	1.58	2.32	2.49	5.45

P E R F O R M A N C E S U M M A R Y

BETWEEN LAB

LAB	DISTANCE FROM ORIGIN
C	.8
J	1.3
B	1.3
A	1.5
J	1.5
D	1.8
B	1.8
G	1.9
C	1.9
I	2.3
H	2.5
K	2.8
I	2.8
H	3.0
L	3.2
90% LIMITS - - - - - - - - - -	
F	3.6
95% LIMITS - - - - - - - - - -	
E	4.3
A	4.6
99% LIMITS - - - - - - - - - -	
L	5.3
K	5.4
E	6.5
D	6.5
F	7.1
G	7.3

WITHIN LAB

LAB	DISTANCE BETWEEN OPERATORS
B	.5
H	1.1
C	1.4
J	2.7
L	3.5
F	3.6
- - - - - - - - - - 90% LIMITS	
I	4.9
D	4.9
- - - - - - - - - - 95% LIMITS	
A	6.1
G	6.6
- - - - - - - - - - 99% LIMITS	
K	8.1
E	9.2

TABLE I SECOND PAGE OF "SCAT" OUTPUT

ClibPDF - www.fastio.com

and low values for a given sample), the standard deviation (a measure of the dispersion of the results), and Phi (a parameter that will be discussed later).

The second page of "Scat" output analyzes the amount of systematic error exhibited by the test being studied and summarizes between and within laboratory performance (see Table 1). This is done by means of a bivariate (variability in two dimensions) analytical technique called principal components analysis(4). Isolation of the systematic and random components of variance is made using the first special case of the additive model discussed in 10. This model assumes that the magnitude of random and systematic errors is the same for both samples of the scatter diagram. In the event that this assumption is not valid, reference 10 should be consulted for a more appropriate model.

The probability distribution of a bivariate normal population, such as the scatter diagram, can be represented by a series of concentric ellipses. In order to evaluate systematic error and laboratory performance, the "Scat" program first calculates the unit variance ellipse. This is an ellipse whose major and minor axes are coincident with the eigenvectors of the scatter diagram's variance-covariance matrix. The orientation of these axes is given on the first page of output by Phi (Ø), the angle between the X-axis and the major axis of the ellipse. The magnitudes of the two eigenvectors are given as the variances under "Overall Results" on the second page of output. From this information an ellipse circumscribing any desired confidence zone for the data can be plotted on the scatter diagram using a plotting program developed for this study called "Eliplot". Figure 6 shows the elliptical 90% confidence limits (approximated by using the circular normal probability density function) for the example already shown in Figure 5. The orientation of the axes as defined by Ø is also illustrated.

ClibPDF - www.fastio.com

S C A T T E R D I A G R A M

EXAMPLE

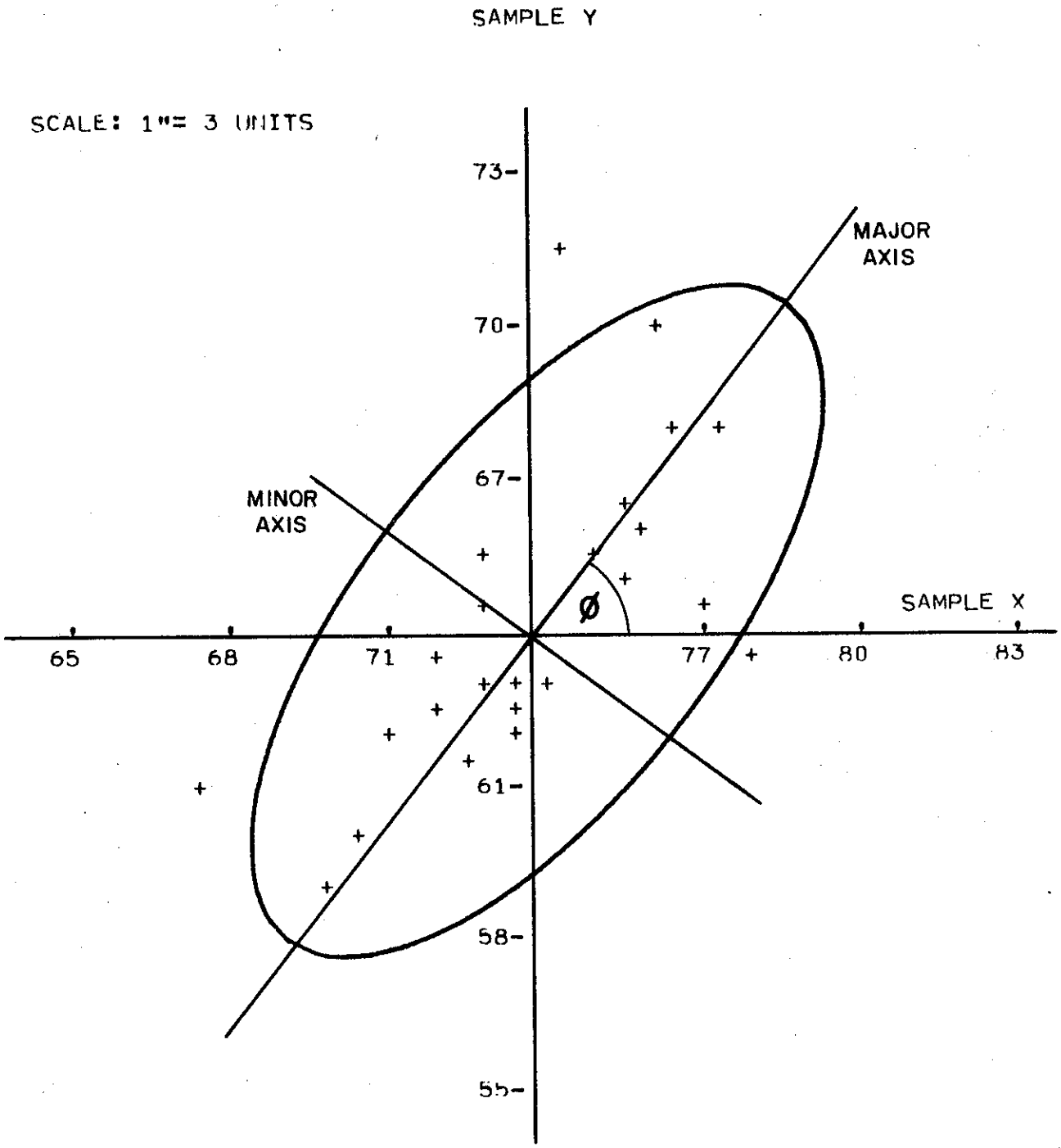


Figure 6 ELLIPTICAL 90% CONFIDENCE LIMITS

ClibPDF - www.fastio.com

As was mentioned earlier, random error will tend to cause a circular cluster of points centered around the origin. Therefore, the confidence limits of a test method whose error is predominantly random will tend to have a circular shape (i.e. the length of the major and minor axes will tend to be equal). Any significant additional length of the major axis over the minor axis may be in part attributable to systematic error. It is useful then if one considers two univariate distributions: 1) variations in the direction of the minor axis caused by random errors only, and 2) variations in the direction of the major axis caused by both random and systematic errors. A sphericity test (R-Ratio, 10) is used to test the hypothesis that two distributions are identical (i.e. that systematic error is immeasurable). Also, the random and systematic components of variance (for averages of two) are computed using the additive model already mentioned (see Table 1, "Overall Results").

"Scat's" evaluation of laboratory performance is meant to isolate results which deviate significantly from the average results for all laboratories. Performance is judged by the characteristic random error of the test method. Figure 7 shows two 90% confidence zones, one circular and one elliptical. The elliptical zone is based on a combination of random and systematic error while the circular zone is based on random error only. Three points fall outside the ellipse and 9 points fall outside the circle. "Scat" evaluates laboratory performance based on the circular zone. The additional 6 points that fall outside the circular zone are considered outliers because of their apparent systematic errors.

The "Between Lab Summary" shown in Table 1 ranks results in order of ascending distances from the origin with the 90, 95,

ClibPDF - www.fastio.com

S C A T T E R D I A G R A M

EXAMPLE

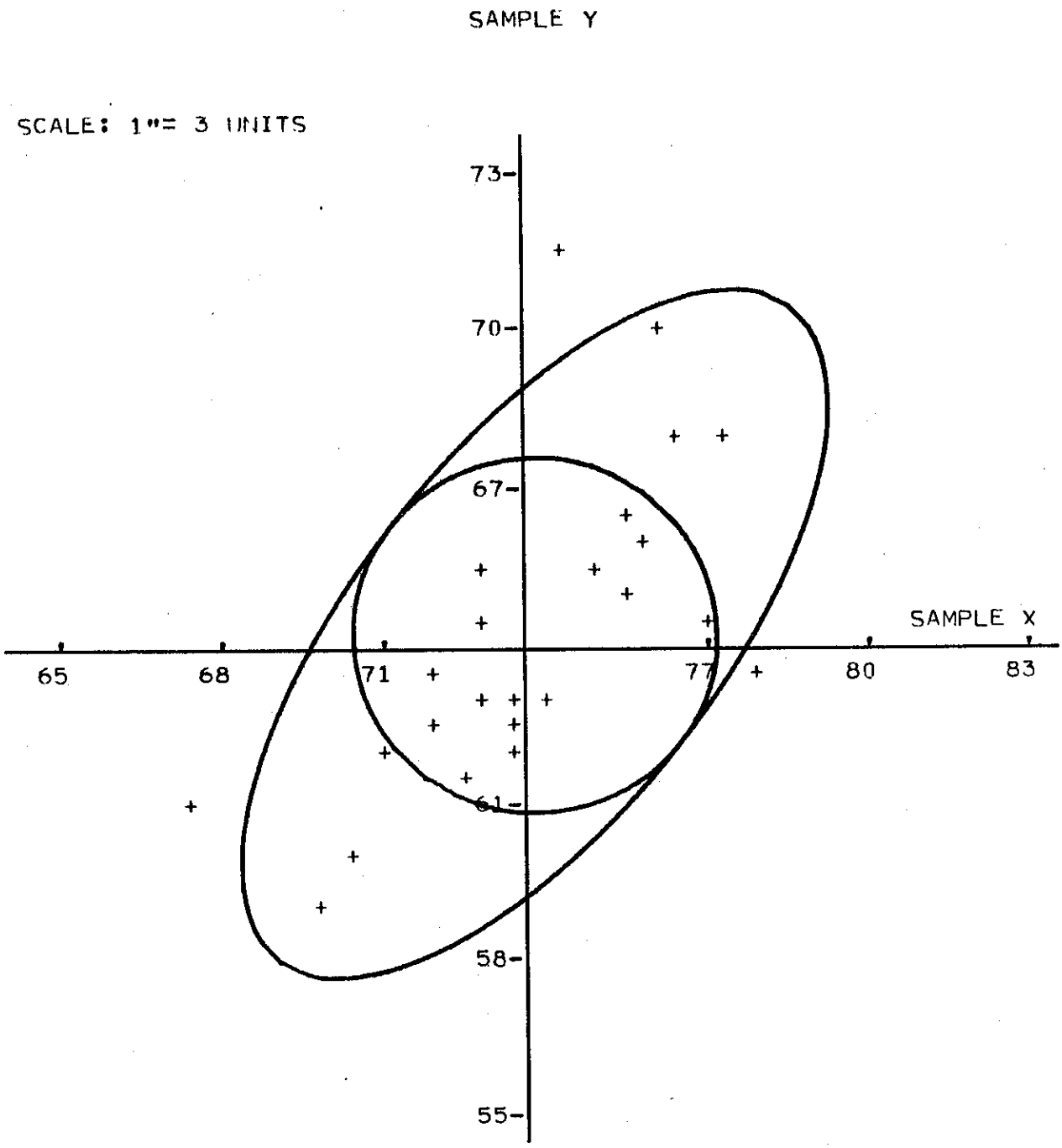


Figure 7 ELLIPTICAL (Random + Systematic Error) AND CIRCULAR (Random Error only) 90% CONFIDENCE ZONES

ClibPDF - www.fastio.com

and 99% confidence limits being superimposed (see Figure 8). The "Within Lab Summary" also shown in Table 1 evaluates the ability of a laboratory to reproduce its results from one operator to the next. The distance between each pair of points for a laboratory is used as a relative measure of this reproducibility. Confidence limits for the analysis are based on the distribution of distances between randomly chosen pairs of points for the test method assuming random error exclusively.

Both the "Within Lab Summary" and "Between Lab Summary" are more qualitative in nature than quantitative. Laboratories that fall outside the confidence limits should check their procedures. All other laboratories, regardless of their ranking, should be considered equally capable and unbiased in their procedures.

If a scatter diagram has an insignificant R-Ratio indicating little or no systematic error, it does not mean that precision cannot or should not be improved. That decision requires quantitative knowledge of all sources of variation - materials, sampling and testing - in addition to manpower, cost-effectiveness and implementation considerations. If the decision is to improve test precision, then tightening tolerance on equipment and procedure or requiring more tests per test result can significantly reduce random errors.

"Aggva"

The data generated by the proposed correlation program can also be analyzed by a technique known as Analysis of Variance (ANOVA, 6) using the "Aggva" program. This analytical treatment yields a more quantitative measure of precision than the previously discussed scatter diagrams. The results can easily be compared

ClibPDF - www.fastio.com

TRANSLAB CORRELATION PROGRAM

TEST METHOD: FINE DURABILITY MATERIAL: PCC
DATE: 08/08/74

	SAMPLE 1	SAMPLE 2
MEAN	73.8	64.2
RANGE	10.5	12.5
STANDARD DEVIATION	2.55	3.06

PHI (RADIANS) = .917

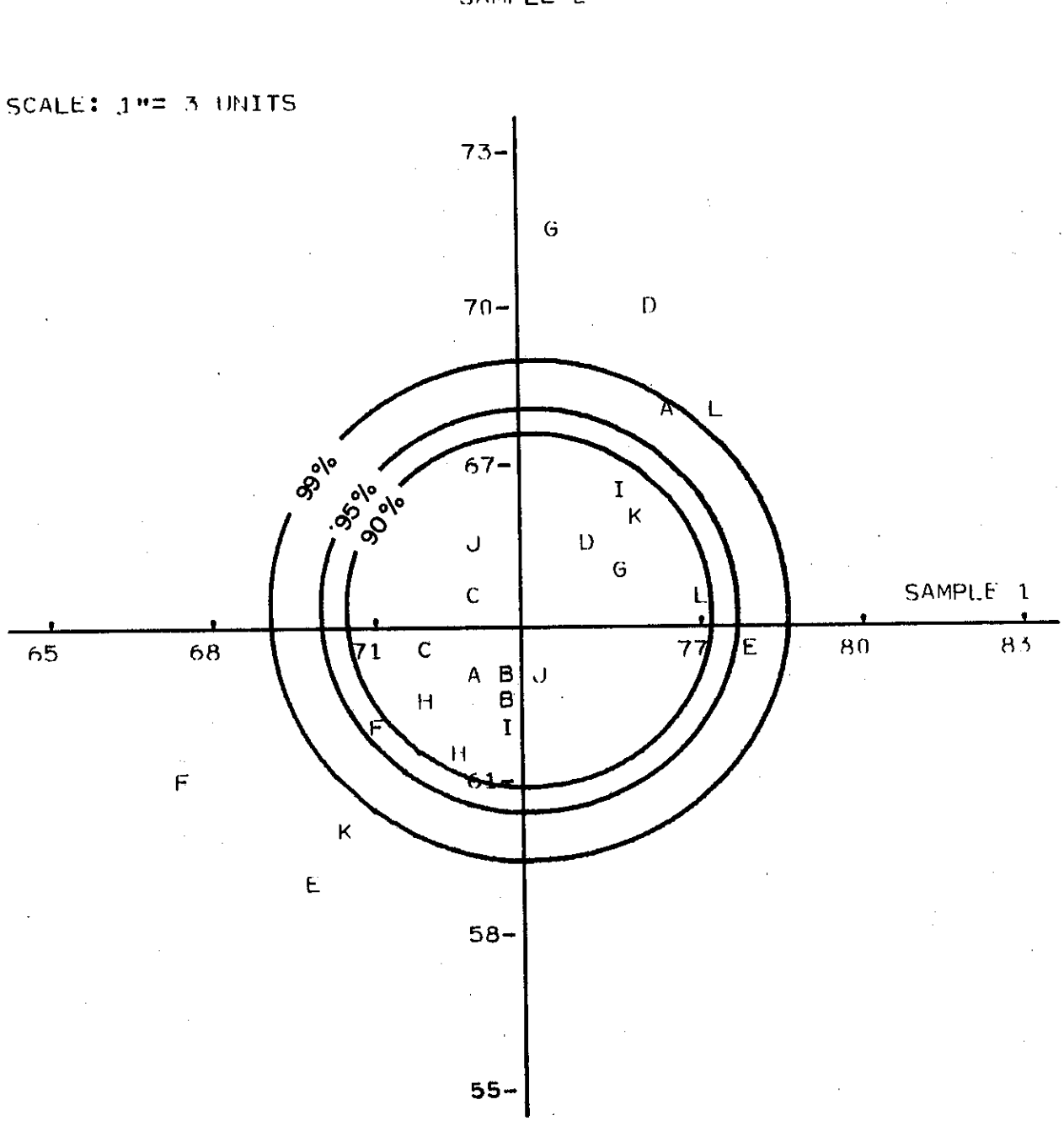


Figure 8 90,95 & 99% CONFIDENCE ZONES BASED ON RANDOM ERROR

ClibPDF - www.fastio.com

to the other contributing sources of variation (materials and sampling) to determine whether improvement in the test precision would significantly reduce the variability of the actual test result.

The ANOVA technique is primarily used to study the effect of different treatments on a well defined operation. The number of treatments and their relationship to one another define a model on which the ANOVA is based. In our case, there are three treatments: operators, samples and laboratories. The operator treatment is termed a nested treatment(6). Since there are two different operators in each laboratory, it is considered nested within the laboratory treatment. Also, since different operators may be used over the period of time samples are sent out, it may often be considered nested within the sample treatment. Hence, two models must be provided for in the program (see the Appendix for model equations).

ANOVA measures the variances of the treatment means and estimates the experimental error based on replications. These are compared for significance using an F-Ratio Test in a fashion determined by the Expected Mean Square (EMS) Table(6). The EMS Table is, in turn, formulated from the model equation (see Appendix for EMS Tables). The components of variance contributed by each treatment and each interaction between treatments can be estimated by also using the EMS Table. These components can be recombined to give an estimate of the variance due to any combination of treatments.

Table 2 is a sample of "Aggva" output. The title of the test method being analyzed and any other pertinent facts are read from the data file and printed on the third line. Then the coding and level or number of each treatment are summarized. The level for laboratories, for example, is 12 because there are 12 participating laboratories. On the other hand, the level

ClibPDF - www.fastio.com

TABLE 2

ANALYSIS OF AGGREGATE TEST METHODS

THREE FACTORIAL PROGRAM --- ANALYSIS OF VARIANCE
TEST METHOD: LA RATTLER (500 REV)

SOURCE	CODE	LEVEL
OPERATOR	O	2
SAMPLE	S	4
LAB	L	12

GRAND MEAN = 1.556E+01

SOURCE	SS	DF	M-SQUARE	F-RATIO
O	34.50	48	.72	.66
S	661.04	3	220.35	306.60
L	460.12	11	41.83	14.10
SL	97.90	33	2.97	4.13
ERROR	104.93	96	1.09	
TOTAL	1358.49	191	7.11	

COMPONENTS OF VARIANCE:

OPERATOR = 0.00
SAMPLE = 4.58
LAB = 2.43
SL = 5.62E-01
ERROR = 1.09

SINGLE-OPERATOR STANDARD DEVIATION = 1.05

MULTILABORATORY STANDARD DEVIATION = 1.88

REPEATABILITY: 3.0 % LOST (BASED ON 96 DEGREES OF FREEDOM)
REPRODUCIBILITY: 5.3 % LOST (BASED ON 96 DEGREES OF FREEDOM)

ClibPDF - www.fastio.com

for operators is only 2, although there are certainly more than 2 operators participating. This is explained by the fact that the operator treatment is nested, and that while there are more than 2 operators in the program, there are only 2 in each laboratory.

The grand mean for the data is listed next in scientific notation (E+01 = 10^1). Then the ANOVA Table giving the sums of squares, degrees of freedom and mean squares (variances) for all the sources of variation in the model equation is printed. Also, the F-Ratios for treatments and interactions are computed. To test these ratios for significance, an F-Ratio Table should be consulted using the appropriate degrees of freedom. For example, consider the F-Ratio for the laboratory treatment in Table 3. This ratio equals 14.1. Since operators are nested within districts and samples, the EMS Table on page A-3 is used to determine which mean square is used to test the significance of the laboratory mean square. It can be seen from the EMS Table that the interaction mean square "SL" has all the expected components of variance that the laboratory mean square has, except for the laboratory component. Hence, if the laboratory component is truly zero (i.e., the laboratory treatment is insignificant), the ratio of the laboratory mean square to the interaction mean square "SL" is <u>expected</u> to be one. Due to random fluctuations it may be greater than or less than one. However, if it exceeds the table value assigned to a ratio whose degrees of freedom are 11 in the numerator and 33 in the denominator, it may be considered more than a random fluctuation. In other words, we may accept the hypothesis that at least one laboratory is consistently getting different results. In the example this hypothesis is accepted (with 95% confidence of being correct) since the tabular value is 2.1.

The next block of values printed by the computer are the components of variance. They are computed by equating each mean

ClibPDF - www.fastio.com

square to the appropriate EMS. If the value computed is negative, a zero is printed.

The single-operator and multilaboratory standard deviations are computed next. The former is equal to the square root of the error component of variance, while the latter equals the square root of the sum of the error, operator and laboratory components of variance. From these values, the repeatability and reproducibility as defined by ASTM Designation C670-71T are calculated(1).

"Interlab"

If the results from "Scat" continually indicate the presence of a laboratory bias, and the testing variations involved, as measured by "Aggva", appear significantly large, then an effort should be made to pinpoint the sources or types of error existing in the test method. This can be done by designing a special experiment for the test method in question. However, before going to this expense, full use of the data generated from the correlation program should be made.

An approach proposed by Dr. John Mandel (Statistician, National Bureau of Standards) has proved capable of isolating general error types without requiring additional data(8 and 9). It also provides information on calibration deficiencies. This useful analytical method has been adopted as the basis for the program entitled "Interlab".

Dr. Mandel's technique is based on a linear model and combines both ANOVA and regression principles. It requires replicate results from a series of laboratories on samples drawn from the range of material types normally encountered. The model itself relies on the assumption that test precision is a constant over the range of material types. If this assumption is not satisfied, a data transformation is used. In addition, the "Aggva" output should be rerun using the same transformation.

ClibPDF - www.fastio.com

A basic understanding of the linear model is essential for those wishing to interpret the results of "Interlab". The reader should study the notation shown on Table 3, and then apply it to the linear model for laboratory No. 1 illustrated in Figure 9.

The line representing laboratory No. 1's performance is found by linearly regressing the laboratory's sample averages against the overall sample averages. If laboratory No. 1 is an "ideal" laboratory, this line will coincide with the dashed 45° line also shown on the graph; however, this is rarely the case. Usually, the line will vary from the ideal not only in position, but also in slope. These two parameters, along with a measure of the variability of the individual points about the line, form the basis for distinguishing the relative amounts of four generalized error types. The four types are categorized into two groups: Within laboratory errors and between laboratory errors.

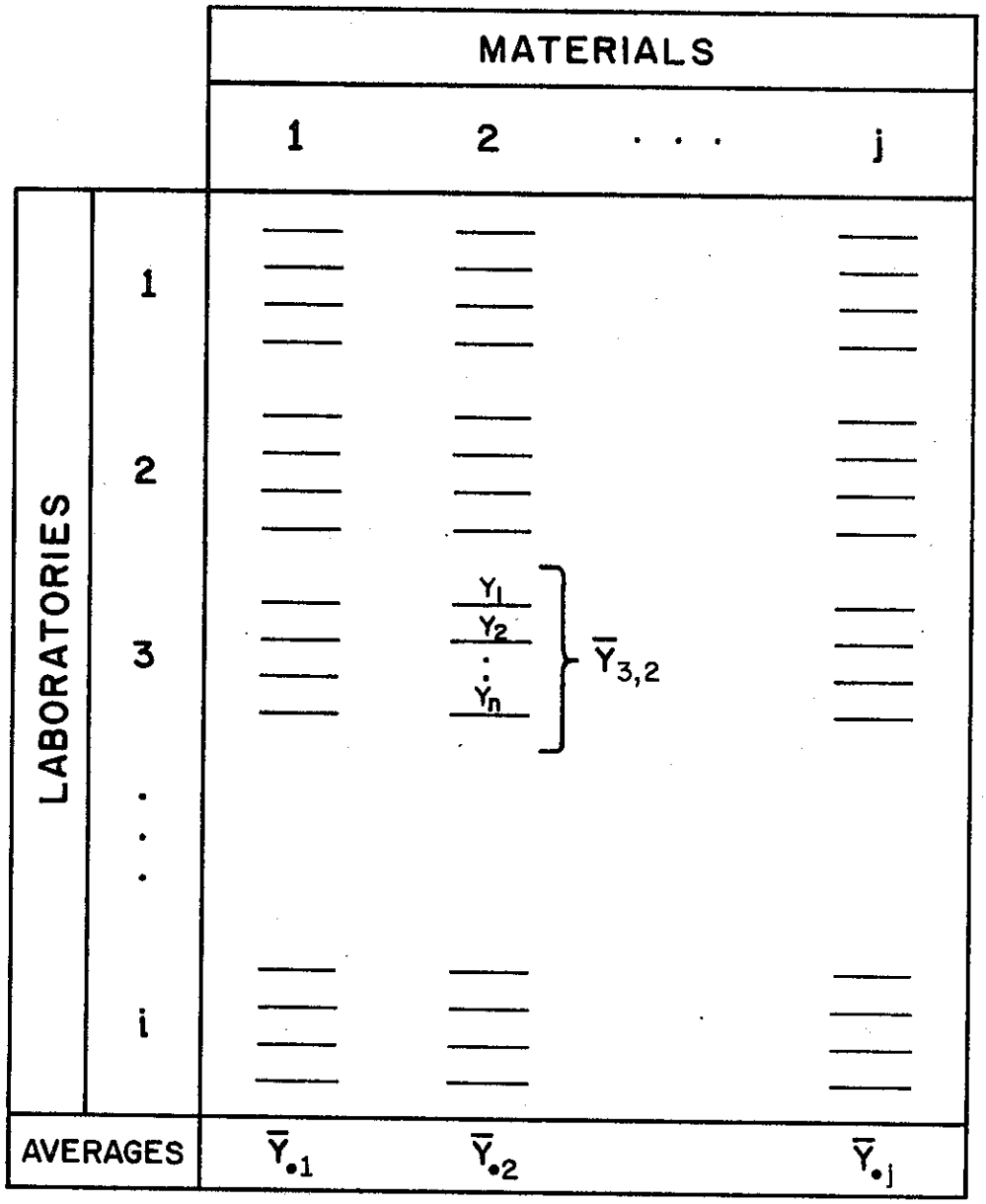
The first type of within laboratory error will be termed epsilon variability. This is a measure of the average variability of individual test results taken under similar conditions (i.e. same equipment and sample). It represents random testing errors that can be improved by either tightening tolerances for the method and equipment involved, or by increasing the number of replications averaged for a test result.

The second type of within laboratory error is called lambda variability. Here, the "goodness of fit" of the sample averages used to determine the regression line is measured. A good fit (low lambda variability) indicates that the laboratories are consistent in their performance of the test method throughout its range of operation. A poor fit (high lambda variability)

ClibPDF - www.fastio.com

TABLE 3

DESIGN SCHEMATIC



ClibPDF - www.fastio.com

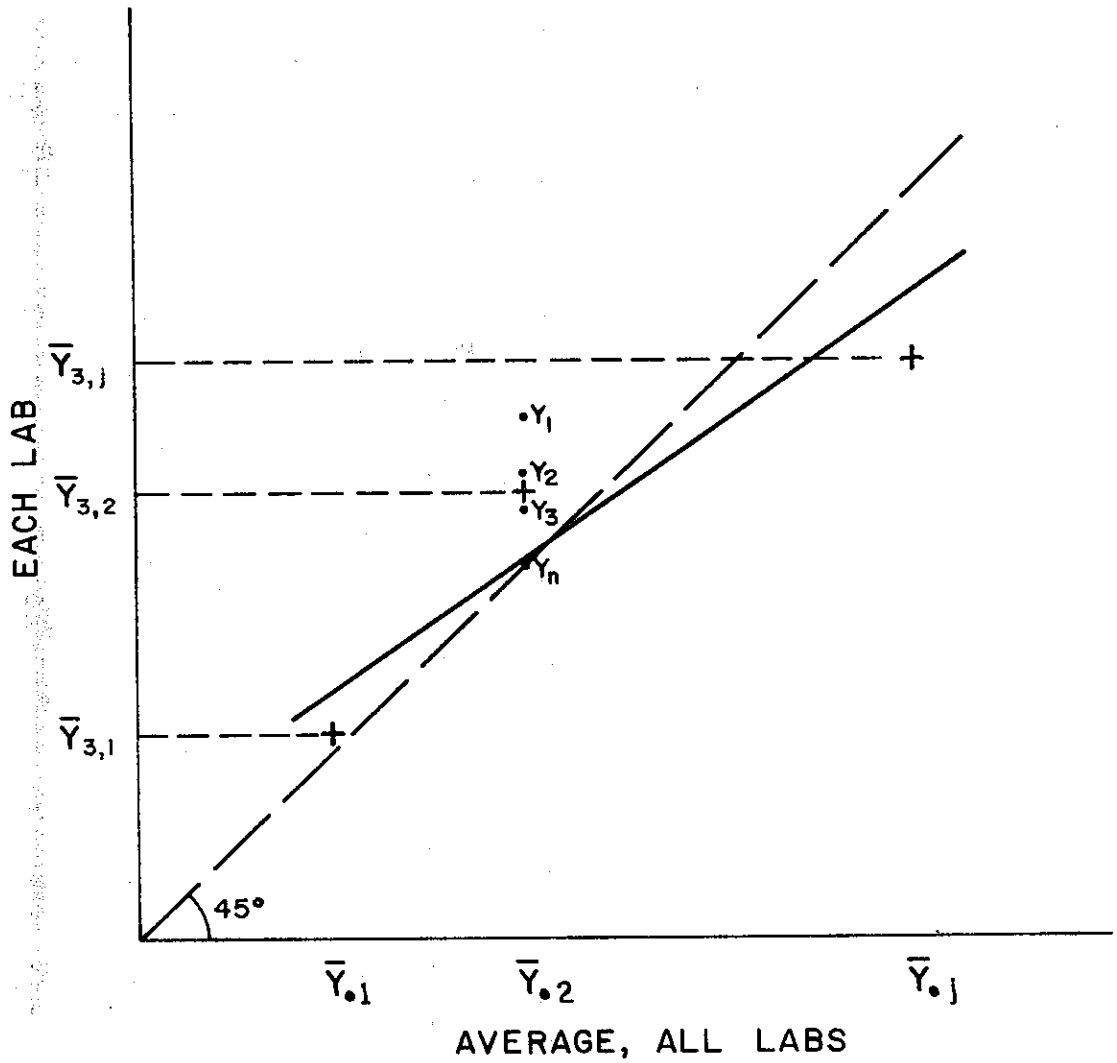


Fig. 9 CODED RESULTS FOR LAB NO. 3

ClibPDF - www.fastio.com

indicates the presence of an error occurring between the ranges of samples tested. The most frequent cause of this type of error is the changing of apparatus for different sample ranges. For instance, a significant lambda variability would be expected for asphalt viscosity tests since different diameter viscometer tubes are used to test asphalts with different viscosities.

Illustrations of the interrelationship between epsilon and lambda error types are shown in Figures 10 and 11. Also note the changes in relative magnitude of the errors with respect to the sample averages. Figure 10 is an example of high epsilon variability for a single laboratory, occurring simultaneously with low lambda variability. A situation of this type may be improved by averaging several tests for each test result, or by reducing some of the random errors from the test method (e.g. specifying stricter rules on sample preparation or replacing manual techniques with machine operations). Figure 11 shows the opposite case: Low epsilon variability occurring with high lambda variability. This type of configuration is illustrative of an operation that is in good control within each sample, but out of control from sample to sample. A typical solution for test methods in which equipment is changed from sample to sample would be the tightening of manufacturing tolerances for the apparatus used. However, for test methods in which the equipment is not changed for different sample ranges, the problem would most likely be caused by poor calibration procedures which could then be corrected.

The second group of errors measured by Mandel's technique are those occurring between laboratories. Of these there are again two types: Mu variability and delta variability. The nature of these errors is somewhat different from their within laboratory counterparts. Mu variability measures a systematic or biased error occurring between laboratories (similar to

ClibPDF - www.fastio.com

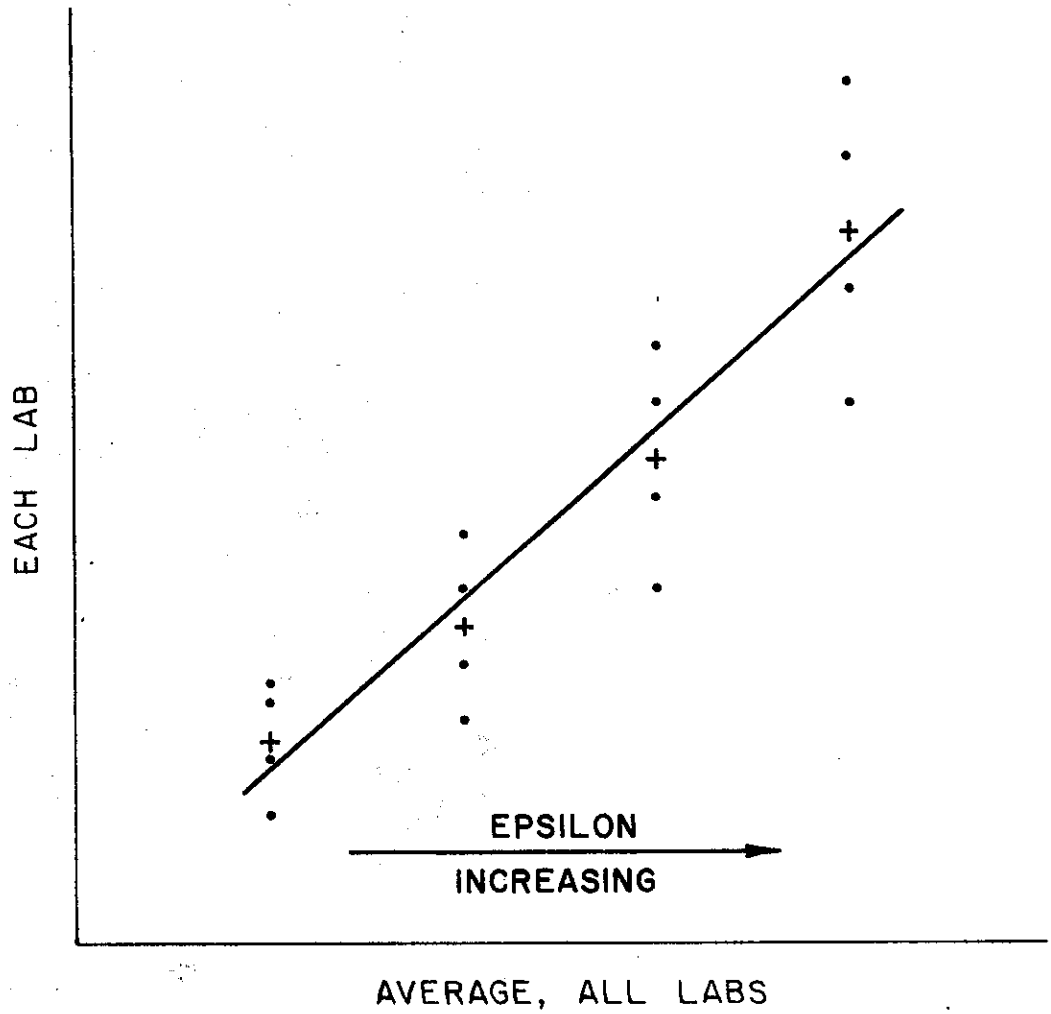


Fig.10 HIGH EPSILON - LOW LAMBDA

ClibPDF - www.fastio.com

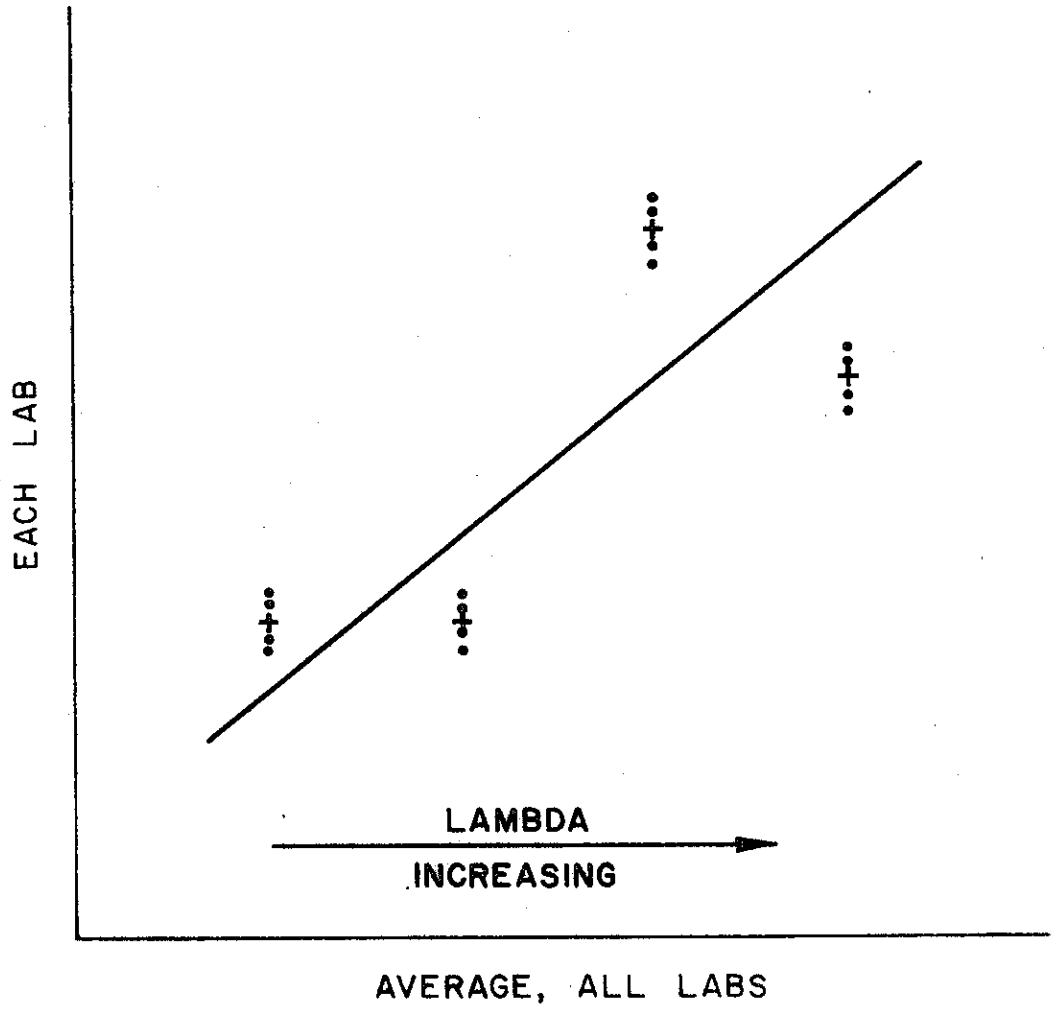


Fig. 11 LOW EPSILON - HIGH LAMBDA

ClibPDF - www.fastio.com

what the scatter diagram analysis measures). For instance, if some laboratories consistently get high results while others get low results, a large mu variability is measured. The common reasons for this type of error are inadequate calibration between laboratories and/or differences in testing procedure. If, however, some laboratories get high results on low range material and low results on high material or vice versa, a large delta variability is measured. This is indicative of an error resulting from adjustments in procedure or equipment over the range of operations. For example, an instrument is used to test a certain material property. Each of several laboratories have their own instrument, but all have been calibrated on a single standard sample. A dial adjustment is used to bring the instrument into equilibrium for the material being tested. This adjustment is the potential source of delta variability. While all laboratories may correlate well on samples within the range of the standard sample, the dial adjustment may give some laboratories high results and others low for ranges of material different than the standard sample. The solution to this problem can be relatively simple: Use at least two standard samples chosen to cover the full range of materials tested.

Figures 12 and 13 show the difference between mu and delta variabilities in terms of a graphical representation. They also point out changes in the magnitude of the errors with respect to the sample averages.

A sample of "Interlab" output is given in Tables 4 and 5. At the top of the first page of output the results of a regression analysis run between test precision and sample average are shown. The slope of the regression equation should be equal or nearly equal to zero in order to satisfy the assumption that test precision is a constant over the range of materials tested.

ClibPDF - www.fastio.com

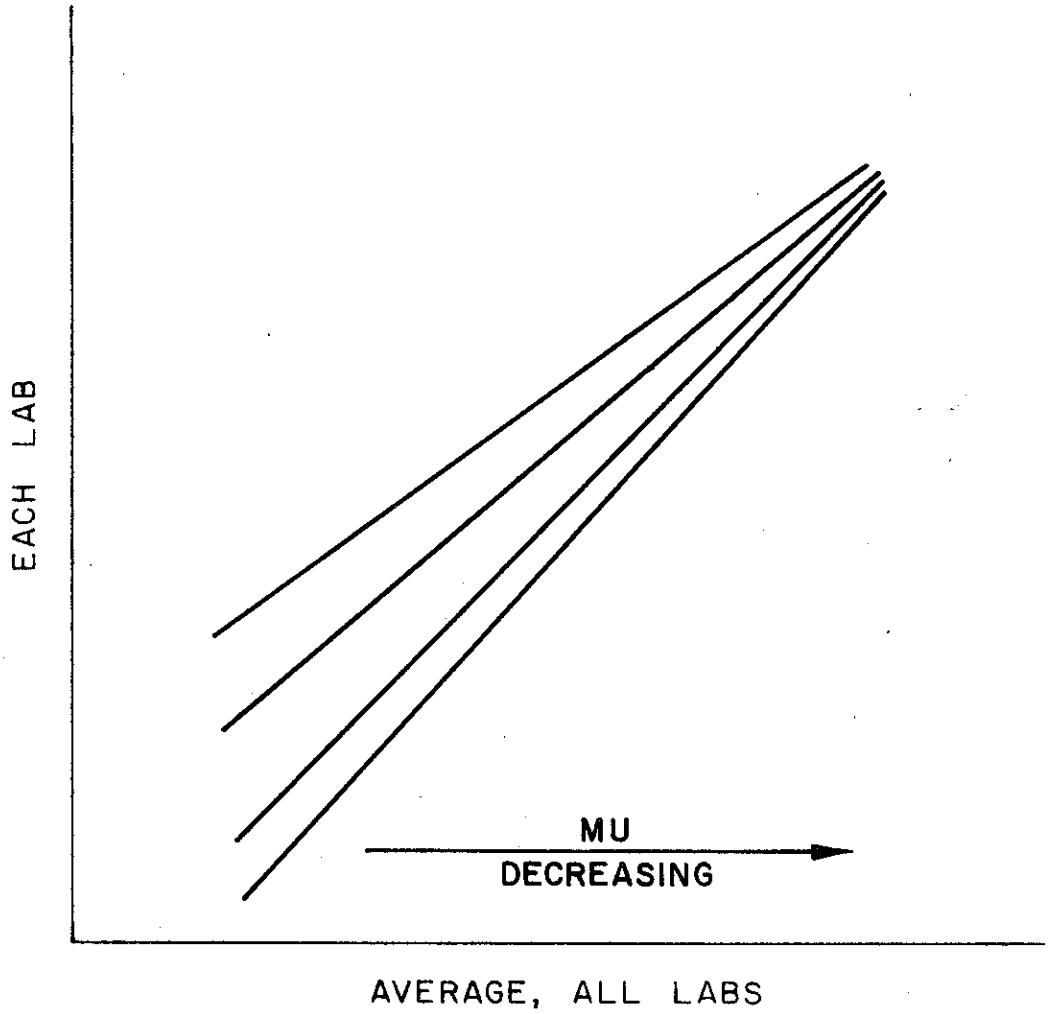


Fig. 12 HIGH MU - LOW DELTA

ClibPDF - www.fastio.com

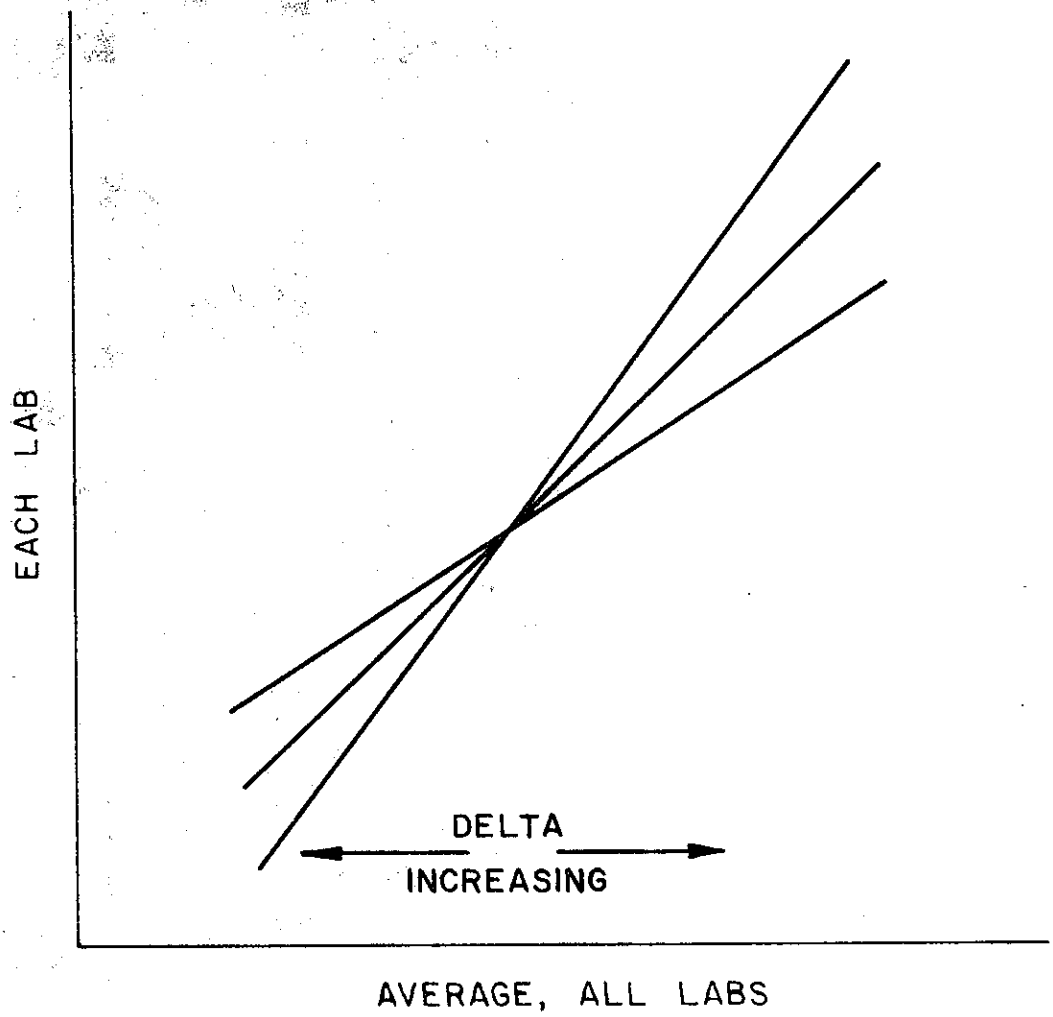


Fig.13 LOW MU - HIGH DELTA

ClibPDF - www.fastio.com

TABLE 4

ANALYSIS OF INTERLABORATORY RESULTS

% CRUSHED PARTICLES (RET. #4)

SUMMARY OF RESULTS:

REGRESSION EQUATION-STANDARD DEVIATION VS. TEST PARAMETER:
Y = 18.449 + (-.165)X INDEX OF DETER. = .714 F-RATIO = 4.98

LABORATORY	SAMPLE 5	6	7	8
A	61.37	44.35	73.30	91.08
	14.06	12.44	4.90	2.21
B	93.55	64.13	86.43	96.30
	.17	7.03	1.32	.70
C	86.45	62.23	75.62	94.18
	3.62	9.17	5.99	1.48
D	82.33	57.75	88.75	95.00
	13.04	16.65	1.50	.82
E	90.85	60.32	81.10	93.98
	2.37	3.59	2.54	1.39
F	74.00	43.75	81.15	93.83
	3.92	4.99	4.25	.91
G	95.00	74.00	86.75	94.00
	.82	1.41	2.36	2.45
H	95.00	69.75	90.75	96.75
	4.24	8.73	6.70	2.63
I	89.50	35.25	79.58	92.98
	2.08	2.50	1.45	1.17
J	88.50	62.00	74.75	88.00
	1.73	9.42	5.62	2.45
K	99.50	87.00	95.05	98.18
	.58	2.83	2.03	1.11
L	78.50	59.25	75.43	91.65
	8.23	7.09	4.08	2.66
AVERAGE	86.21	59.98	82.39	93.83*
	6.43	8.34	4.01	1.82**

* - SAMPLE MEAN
**- STANDARD DEVIATION

ClibPDF - www.fastio.com

TABLE 5

ANALYSIS OF INTERLABORATORY RESULTS

% CRUSHED PARTICLES (RET. #4)

REGRESSION SUMMARY:

LABORATORY	SLOPE	Y-AVG	MSE	PLOTTING POINTS X	PLOTTING POINTS Y	X	Y
A	1.205	67.525	131.248	64.00	47.51	97.00	87.29
B	.993	85.100	6.798	64.00	68.61	97.00	101.39
C	.917	79.619	23.322	64.00	64.39	97.00	94.66
D	1.081	80.956	31.274	64.00	63.02	97.00	98.67
E	1.029	81.562	10.399	64.00	64.47	97.00	98.44
F	1.416	73.181	45.881	64.00	49.68	97.00	96.40
G	.638	87.437	12.432	64.00	76.84	97.00	97.90
H	.845	88.062	7.191	64.00	74.04	97.00	101.92
I	1.807	74.325	32.865	64.00	44.33	97.00	103.95
J	.812	78.312	31.706	64.00	64.84	97.00	91.62
K	.366	94.931	5.077	64.00	88.86	97.00	100.93
L	.891	76.206	15.471	64.00	61.42	97.00	90.81
AVERAGE	1.000	80.602	29.472				

ANALYSIS OF VARIANCE:

SOURCE	DEGREES OF FREEDOM	SUMS OF SQUARES	MEAN SQUARES	F-RATIO
LABORATORIES	11	2479.99	225.45	4.68
SAMPLES	3	7616.75	2538.92	
INTERACTION	33	1590.80	48.21	

AVERAGE	TOTAL VARIANCE	SOURCES OF VARIABILITY IN PERCENT: WITHIN LABORATORY EPSILON	WITHIN LABORATORY LAMBDA	BETWEEN LABORATORY MU	BETWEEN LABORATORY DELTA
86.21	85.73	38	25	36	1
59.98	214.43	15	10	70	5
82.39	96.82	34	22	44	0
93.83	71.47	46	30	18	6
56.00	244.74	13	9	72	6
60.00	214.30	15	10	70	5
64.00	186.72	17	11	67	4
68.00	162.01	20	13	64	2
72.00	140.16	23	15	60	1
76.00	121.18	27	18	55	0
80.00	105.07	31	20	49	0
84.00	91.83	35	23	41	0
88.00	81.45	40	26	32	2
92.00	73.94	44	29	23	4

ClibPDF - www.fastio.com

If the slope differs significantly from zero, the index of determination is greater than 0.5 and the F-Ratio for regression greater than 4, a logarithmic transformation of the form $Z = K \ln 10(A+By)+C$* should be made.

The remainder of the first page is devoted to a summary of results. For each laboratory-sample combination the mean and standard deviation of the four test results (two replicates by two operators) are listed. At the bottom of the summary, the overall average and pooled standard deviation for each sample is given. These are the numbers from which Mandel's linear model is constructed.

The second page of "Interlab" output gives the results of the analysis in three blocks. The first block is a summary of the regression analyses run for each laboratory. Plotting points are given for each line so that the linear model may easily be graphed. Also, erratic results from any laboratory may be spotted by studying the mean square error (MSE) column. The larger this value the poorer the fit of the test results to a line. If one or several of the laboratories have MSE's much larger than the rest, the analyst may wish to exclude their results from the study.

Next, an analysis of variance is given. This is somewhat redundant since "Interlab" will only be run on data which has already exhibited a significant laboratory bias when run on "Aggva". However, it is included so as to make "Interlab" output independent from the other analyses.

The last block lists the relative amounts (in percent) of the four error types for each sample tested. A line separates these values from an incremental summary over the range of

*A = y-intercept, B = slope, K and C are arbitrary constants chosen for convenience.

ClibPDF - www.fastio.com

material studied. If a data transformation has been used an additional two columns list the average and standard deviation retransformed to the original scale (see Tables 6 and 7).

It is important that the analyst remember this simple procedure when examining results from "Interlab". First, study the relative magnitudes of the sources of variability throughout the range of materials analyzed. Then develop a credible hypothesis to account for the distribution of these sources of variability. This hypothesis should be verified by conducting a specifically designed experiment. If the hypothesis is then accepted, implement the most economically appropriate solution. The continuing correlation program will soon show if the implemented solution is effective.

"Ranklab"

Evaluating the performance of individual laboratories is an important part of the correlation program since systematic errors often come from discrepant laboratories. A means for conveniently identifying these laboratories is needed. "Scat" can only fulfill this need for two samples at a time, while "Interlab's" laboratory results are too complex to be interpreted readily. Hence, a nonparametric analysis was chosen to meet the need (16). This formed the basis for the program entitled "Ranklab".

Nonparametric analyses offer simplicity while sacrificing very little power, provided large enough amounts of data are used. In the case of "Ranklab", nonparametrics are solely concerned with the ranks of laboratory results within each sample rather than the results themselves. These ranks are analyzed according to the probabilities of nonparametrics. The data used by "Ranklab" is essentially the same as the data summarized on the first page

ClibPDF - www.fastio.com

TABLE 6

ANALYSIS OF INTERLABORATORY RESULTS

% CRUSHED PARTICLES (RET. #4)

SUMMARY OF RESULTS:

REGRESSION EQUATION-STANDARD DEVIATION VS. TEST PARAMETER:
Y = 4.289 + (.041)X INDEX OF DETER. = .239 F-RATIO = .63
TRANSFORMATION EQUATION: Z=100*LOG10(18.449-(.165*Y))-15

LABORATORY	SAMPLE 5	6	7	8
A	75.68	89.08	65.03	38.23
	12.64	8.19	5.77	4.79
B	32.90	74.23	47.17	25.78
	.41	6.45	2.25	1.95
C	46.85	75.72	62.18	31.27
	5.92	8.11	6.88	3.77
D	51.09	78.41	42.97	29.27
	16.51	13.86	2.81	2.11
E	38.67	77.84	55.37	31.78
	5.18	3.12	3.63	3.44
F	64.34	89.95	55.10	32.20
	4.45	3.14	5.90	2.18
G	29.27	64.49	46.50	31.50
	2.11	1.60	4.23	6.13
H	28.36	68.45	37.51	24.04
	10.29	8.84	13.41	7.61
I	41.46	95.13	57.55	34.19
	4.07	1.41	1.94	2.65
J	43.42	75.90	63.26	44.26
	3.23	8.25	6.70	4.42
K	15.75	46.01	28.95	20.11
	2.04	4.78	5.17	3.57
L	58.00	78.54	62.63	36.93
	10.87	5.59	5.04	5.48
AVERAGE	43.82	76.15	52.02	31.63*
	8.03	7.02	6.06	4.35**

* - SAMPLE MEAN
** - STANDARD DEVIATION

ClibPDF - www.fastio.com

TABLE 7

ANALYSIS OF INTERLABORATORY RESULTS

% CRUSHED PARTICLES (RET. #4)

REGRESSION SUMMARY:

LABORATORY	SLOPE	Y-AVG	MSE	PLOTTING POINTS X	PLOTTING POINTS Y	X	Y
A	.989	67.005	195.387	41.00	57.21	61.00	76.99
B	1.129	45.020	13.412	41.00	33.84	61.00	56.42
C	.987	54.007	40.433	41.00	44.24	61.00	63.97
D	1.039	50.437	78.021	41.00	40.15	61.00	60.92
E	1.075	50.912	18.843	41.00	40.26	61.00	61.77
F	1.184	60.400	122.653	41.00	48.67	61.00	72.36
G	.816	42.940	48.253	41.00	34.86	61.00	51.18
H	1.043	39.590	28.571	41.00	29.26	61.00	50.12
I	1.427	57.083	31.060	41.00	42.95	61.00	71.49
J	.779	56.707	54.250	41.00	48.99	61.00	64.57
K	.655	27.704	44.858	41.00	21.22	61.00	34.32
L	.877	59.025	36.904	41.00	50.34	61.00	67.83
AVERAGE	1.000	50.903	59.387				

ANALYSIS OF VARIANCE:

SOURCE	DEGREES OF FREEDOM	SUMS OF SQUARES	MEAN SQUARES	F-RATIO
LABORATORIES	11	5045.41	458.67	8.44
SAMPLES	3	12721.55	4240.52	
INTERACTION	33	1793.12	54.34	

TRANSFORMED SCALE AVERAGE	TOTAL VARIANCE	SOURCES OF VARIABILITY IN PERCENT: WITHIN LABORATORY EPSILON	WITHIN LABORATORY LAMBDA	BETWEEN LABORATORY MU	BETWEEN LABORATORY DELTA	NORMAL SCALE AVER.	NORMAL SCALE STD. DEV.
43.82	179.96	24	27	49	0	88.33	7.25
76.15	235.33	18	21	61	0	62.38	17.46
52.02	192.75	22	25	53	0	83.45	9.07
31.63	162.56	26	30	44	0	94.08	5.21
28.00	157.74	27	31	42	0	95.50	4.72
33.00	164.42	26	30	45	0	93.51	5.40
38.00	171.42	25	28	47	0	91.28	6.19
43.00	178.74	24	27	49	0	88.77	7.09
48.00	186.38	23	26	51	0	85.96	8.13
53.00	194.33	22	25	53	0	82.80	9.31
58.00	202.61	21	24	55	0	79.26	10.57
63.00	211.21	20	23	57	0	75.29	12.[illegible]
68.00	220.12	19	22	59	0	70.84	14.0[illegible]
73.00	229.36	18	21	60	0	65.84	16.03

ClibPDF - www.fastio.com

of "Interlab" (see Table 5). For each sample, the laboratory means and standard deviations are ranked from lowest to highest by the program. A rank of 1.0 indicates the lowest result for a particular sample by a particular laboratory. Conversely, a rank of 12.0 (provided 12 laboratories are being considered) indicates the highest result. When results are identical the ranks involved are averaged and this average rank assigned to the tied laboratories. This explains the occurrence of ranks incremented by 0.5. Study of Tables 4 and 8 should provide a more thorough understanding of the procedure.

Once the ranks are assigned for each sample, a rank sum is calculated for each laboratory. Results are then recorded by laboratory from lowest to highest rank sum. Simultaneously, the columnar arrangement of results is reordered to read from the lowest sample average to the highest. This completed rearrangement is the form in which results appear on "Ranklab's" output.

The ranking summary and analysis is presented in two separate blocks (see Table 8). The first block ranks the mean or average results obtained by laboratories while the second ranks the within laboratory standard deviations summarized by "Interlab". The rank sum corresponding to each laboratory is shown, in addition to the Kendall Rank Correlation Coefficient(13). This coefficient is a measure of increases or decreases in rank standing corresponding to increases in the sample average. Values approaching 1.0 or -1.0 are indicative of laboratories requiring tighter calibration over the full operational range of the test.

Both the rank sum and Kendall Coefficient are marked by an asterisk if significant at the 95% level. This means there is only a 5% chance that the marked value is caused by random fluctuations.

ClibPDF - www.fastio.com

TABLE 8

LABORATORY RANKING SUMMARY - NONPARAMETRIC ANALYSIS

% CRUSHED PARTICLES (RET. #4)

MEAN RANKS:

LAB CODE	SAMPLE 6	7	5	8	RANK SUM	KENDALL RANK CORRELATION COEFFICIENT
A	3.0	1.0	1.0	2.0	7.0 *	-.17
L	5.0	3.0	3.0	3.0	14.0	-.55
J	7.0	2.0	6.0	1.0	16.0	-.67
F	2.0	7.0	2.0	5.0	16.0	.17
I	1.0	5.0	7.0	4.0	17.0	.33
C	8.0	4.0	5.0	8.0	25.0	.17
E	6.0	6.0	8.0	6.0	26.0	.18
D	4.0	10.0	4.0	9.0	27.0	.17
B	9.0	8.0	9.0	10.0	36.0	.50
G	11.0	9.0	10.5	7.0	37.5	-.67
H	10.0	11.0	10.5	11.0	42.5	.50
K	12.0	12.0	12.0	12.0	48.0 *	.00
				MEAN SCORE	26.0	

SIGMA RANKS:

LAB CODE	SAMPLE 6	7	5	8	RANK SUM	KENDALL RANK CORRELATION COEFFICIENT
B	6.0	1.0	1.0	1.0	9.0	-.50
K	3.0	4.0	2.0	4.0	13.0	.18
I	2.0	2.0	5.0	5.0	14.0	.67
G	1.0	5.0	3.0	9.5	18.5	.67
E	4.0	6.0	6.0	6.0	22.0	.50
F	5.0	8.0	8.0	3.0	24.0	-.17
D	12.0	3.0	11.0	2.0	28.0	-.73
J	10.0	10.0	4.0	9.5	33.5	-.50
C	9.0	11.0	7.0	7.0	34.0	-.50
L	7.0	7.0	10.0	12.0	36.0	.83
A	11.0	9.0	12.0	8.0	40.0	-.33
H	8.0	12.0	9.0	11.0	40.0	.58
				MEAN SCORE	26.0	

* - EXCEEDS CRITICAL LIMITS AT APPROXIMATE 95% LEVEL.

ClibPDF - www.fastio.com

"Ranklab" is equipped with coding and laboratory deletion options. These, combined with its easily understood ranking format, make it an ideal accompaniment to "Scat" for reporting results to participating laboratories.

ClibPDF - www.fastio.com

REFERENCES

1. ASTM-C670-71T, "Preparing Precision Statements for Test Methods for Construction Materials".

2. Benson, Paul E., "Precision of Selected Aggregate Test Methods," California Department of Transportation, Report #CA-DOT-TL-1153-2-75-05, 1974.

3. Crandall, J. R. and Blaine, R. L., "Statistical Evaluation of Interlaboratory Cement Tests," Proceedings, Am. Soc. Testing Mats., Vol. 59, p. 1129, 1959.

4. Davis, John C., Statistics and Data Analysis in Geology, New York, John Wiley & Sons, Inc., 1973.

5. Duncan, A. J., Quality Control and Industrial Statistics, 3rd Ed., Homewood: R. D. Irwin, Inc., 1965.

6. Hicks, C. R., Fundamental Concepts in the Design of Experiments, New York: Holt, Rinehart and Winston, 1964.

7. Lashof, T. W., "Ranking Laboratories and Evaluating Methods of Measurement in Round-Robin Tests," Materials Research and Standards Vol. 4, No. 8, Aug. 1964, pp. 397-407.

8. Mandel, John, "The Measuring Process," Technometrics, Aug., 1959, pp. 251-267.

9. Mandel, John and Lashof, T. W., "The Interlaboratory Evaluation of Testing Methods," ASTM Bulletin, No. 239, July 1959, pp. 53-61.

ClibPDF - www.fastio.com

10. Mandel, John and Lashof, T. W., "Interpretation and Generalization of Youden's Two-Sample Diagram," Journal of Quality Technology, Vol 6, No. 1, January, 1974, pp. 22-36.

11. Neville, A. M. and Kennedy, J. B., Basic Statistical Method for Engineers and Scientists, Scranton: International Textbook Company, 1964.

12. Owen, D. B., Handbook of Statistical Tables, Palo Alto: Addison-Wesley, Inc., 1962.

13. Siegel, Sidney, Nonparametric Statistics for the Behavioral Sciences, New York: McGraw-Hill, 1956.

14. Thompson, W. A. and Willke, T. A., "On an Extreme Rank Sum Test for Outliers," Biometrika, Vol. 52, Nos. 3 and 4, December 1963, p. 375.

15. Youden, W. J., "Graphical Diagnosis of Interlaboratory Test Results," Industrial Quality Control, Vol. XV, p. 24, May, 1959.

16. Youden, W. J., "Ranking Laboratories by Round Robin Tests," Materials Research and Standards, Vol. 3, No. 1, pp. 9-13, 1963.

ClibPDF - www.fastio.com

GLOSSARY

ANALYSIS OF VARIANCE (ANOVA): A numerical technique which computes the variances of a measurable process under the influence of different treatments, and assesses the relative importance of these treatments.

BIVARIATE DISTRIBUTION: A distribution in which each observational unit is characterized by two variables.

CORRELATION PROGRAM: A specific plan mutually benefiting all participants in which the goal is to encourage and aid participants to strive for uniform application of a process (in this case particular test methods).

COVARIANCE: A measure of the joint variation of two variables about their common mean.

DEGREES OF FREEDOM: The number of independent values that can be assigned to a set of observations.

EIGENVECTOR: A mathematical parameter associated with matrix algebra. As used in this report, eigenvectors correspond with the principal axes of an ellipse whose shape is characteristic of the shape of the confidence limits of a bivariate distribution.

EXPECTED MEAN SQUARE (EMS): The long range average value of a mean square in algebraic form.

F-RATIO: A statistic consisting of the ratio of two variances: $F = S_1^2/S_2^2$. May be used to determine whether both variances come from the same population.

ClibPDF - www.fastio.com

INTERACTION: A type of variation attributed to the situation where the effect of one treatment changes for different levels of some other treatment.

MEAN SQUARE ERROR (MSE): The 'Average' cell variance of an ANOVA. The term error is used because this is a measurement of the error or "noise" inherent in the particular ANOVA.

MEAN SQUARE: An unbiased estimate of a population variance. Determined by dividing a sum of squares by its degrees of freedom.

MODEL EQUATION: An algebraic representation of each value in an ANOVA matrix.

NESTED TREATMENT: A treatment whose levels are chosen within the levels of another treatment is said to be nested within that treatment.

NONPARAMETRIC STATISTICS: A branch of statistics which makes no hypotheses about the value of parameters in a statistical density function.

POWER OF A TEST: The probability $1-\beta$, where β is the probability of accepting a hypothesis when it is false.

RANDOM ERROR: An error that does not follow a consistent, definable pattern.

RANGE: The difference between the extreme high and low values of a sample.

REPEATABILITY: A measure of the greatest difference between two results that could be expected to be exceeded in only 1 out of 20 cases when properly conducted repetitive determinations are made on identical portions of material by a competent operator using one set of equipment.

ClibPDF - www.fastio.com

REPLICATE TESTS: Ideally, two or more tests conducted under identical conditions on identical material by the same operator.

REPRODUCIBILITY: A measure of the greatest difference between two results that could be expected to be exceeded in only 1 out of 20 cases when properly conducted determinations are made by two different operators in different laboratories on identical portions of material.

SCIENTIFIC NOTATION: A shorthand form of writing very large or very small numbers using powers of ten ($0.00019=1.9\times10^{-4}=1.9E-04$). Useful in standardizing the output format of a computer program that will be used on a wide range of numbers.

STANDARD DEVIATION: The square root of the variance, also known as the root mean square deviation. This is often a more practical statistic than the variance since it is in the same units as the variate and is directly related to several important physical characteristics of the normal distribution.

SYSTEMATIC ERROR: An error that follows a consistent, definable pattern.

UNIVARIATE DISTRIBUTION: A distribution in which each observational unit is characterized by a single variable.

VARIANCE: A statistical measure of data dispersion defined as the mean of the squares of the deviation of each data point from the arithmetic mean.

VARIANCE-COVARIANCE MATRIX: A matrix which mathematically represents the "variance" of a multivariate (in this case, bivariate) normal distribution.

ClibPDF - www.fastio.com

APPENDIX

ANALYSIS OF VARIANCE:

Source	Code
Operator	O
Sample	S
Laboratory	L

I. 1st Model: Operators nested in laboratories.

A. Model Equation

$$X_{ijkl} = \mu = O_{i(k)} + S_j + L_k + SL_{jk} + OS_{ij(k)} + e_{l(ijk)}$$

For i = 1 to I Operators/Lab
j = 1 to J Samples
k = 1 to K Labs
l = 1 to L Replicates,

and where μ equals the grand mean.

B. Expected Mean Squares*

Source	EMS
Operator	$V_e + (L)V_{os} + (JL)V_o$
Sample	$V_e + (L)V_{os} + (IKL)V_s$
OS	$V_e + (L)V_{os}$
Laboratory	$V_e + (L)V_{os} + (IL)V_{sl} + (JL)V_o + (IJL)V_l$
SL	$V_e + (L)V_{os} + (IL)V_{sl}$
Error	V_e

*V Symbolizes Variance.

ClibPDF - www.fastio.com

C. F-Ratios

$$F_o = \frac{MS_o}{MS_{os}}$$

$$F_s = \frac{MS_s}{MS_{os}}$$

$$F_l = \frac{MS_l}{MS_{sl}+MS_o-MS_{os}} \quad \text{(Psuedo F-Test)}$$

$$F_{os} = \frac{MS_{sl}}{MS_e}$$

$$F_{sl} = \frac{MS_{sl}}{MS_{os}}$$

D. Components of Variance

$$V_o = \frac{MS_o-MS_{os}}{JL}$$

$$V_s = \frac{MS_s-MS_{os}}{IKL}$$

$$V_l = \frac{MS_l+MS_{os}-MS_o-MS_{sl}}{IJL}$$

$$V_{os} = \frac{MS_{os}-MS_e}{L}$$

$$V_{sl} = \frac{MS_{sl}-MS_{os}}{IL}$$

ClibPDF - www.fastio.com

II. 2nd Model: Operators nested in laboratories and samples.

A. Model Equation

$$X_{ijkl} = \mu + O_{i(jk)} + S_j + L_k + SL_{jk} + e_{l(ijk)}$$

For i = 1 to I Operators/Lab
j = 1 to J Samples
k = 1 to K Labs
l = 1 to L Replicates,

and where μ equals the grand mean.

B. Expected Mean Squares

Source	EMS
Operator	$V_e + (L)V_o$
Sample	$V_e + (L)V_o + (IKL)V_s$
Laboratory	$V_e + (L)V_o + (IL)V_{sl} + (IJL)V_l$
SL	$V_e + (L)V_o + (IL)V_{sl}$
Error	V_e

C. F-Ratios

$$F_o = \frac{MS_o}{MS_e}$$

$$F_s = \frac{MS_s}{MS_o}$$

$$F_l = \frac{MS_l}{MS_{sl}}$$

$$F_{sl} = \frac{MS_{sl}}{MS_o}$$

ClibPDF - www.fastio.com

D. Components of Variance

$$V_o = \frac{MS_o - MS_e}{L}$$

$$V_s = \frac{MS_s - MS_o}{IKL}$$

$$V_1 = \frac{MS_1 - MS_{s1}}{IJL}$$

$$V_{s1} = \frac{MS_{s1} - MS_o}{IL}$$

ClibPDF - www.fastio.com

ClibPDF - www.fastio.com

ClibPDF - www.fastio.com