# Progress Report

PINPOINTING SUSPECT TRIPLICATE UNCONFINED COMPRESSIVE STRENGTH VALUES IN A SERIES OF SOIL-ADDITIVE STRENGTH DETERMINATIONS

by

H. T. David, D. T. Davidson and C. A. O'Flaherty

Project HR-1 of the Iowa Highway Research Board
Project 283-S of the Iowa Engineering Experiment Station, Iowa State University, Ames, Iowa

Presented at the Sixty-Second Annual Meeting of the American Society for Testing Materials, Atlantic City, New Jersey, June 26-30, 1961.

Reproduction in whole or in part is permitted for any purpose of the United States Government.

IOWA ENGINEERING EXPERIMENT STATION

IOWA STATE UNIVERSITY of Science and Technology / Ames, Iowa

# PINPOINTING SUSPECT TRIPLICATE UNCONFINED COMPRESSIVE STRENGTH VALUES IN A SERIES OF SOIL-ADDITIVE STRENGTH DETERMINATIONS[1]

By
H. T. David,[2] D. T. Davidson[2] and C. A. O'Flaherty[2]

In recent years, various types of organic and inorganic materials have been investigated for use as soil stabilizing agents in the construction of highways and airports. Since the properties and environmental conditions of soils vary so greatly from place to place, a stabilizing agent that is suitable for one type of soil may not be satisfactory for another. As a result, it is often desirable to evaluate several stabilizing agents under varying treatment conditions before deciding on a specific one to be used with a given soil. In addition many research programs have been initiated which investigate the effects of these stabilizing agents upon soils.

The unconfined compressive strength test is probably the most commonly used test in such soil stabilization investigations. The general procedure is, for one given test condition, to prepare and test several specimens, after which the average of the several strength values is reported. Three specimens per test condition are commonly used. Because of the many variables involved, the total number of specimens which may have to be tested may range from the hundreds to the thousands, depending upon the size and scope of the investigation.

Since such large numbers of specimens are involved, it is likely that some unconfined compressive strength results will be obtained that are, seemingly, not what they should be. The question then arises whether these unusual observations are the result of expected normal experimental variation, or whether they are due to an experimental or material abberration and should therefore be discarded. In cases where three specimens are prepared per test condition, a commonly used solution to this question is to discard any single measurement which deviates by more than ten percent from the average of all three measurements, as prescribed in ASTM Method of Test for Compressive Strength of Hydraulic Cement Mortars (C109-58)[3]. In the event of such a disqualifying deviation, the average of the remaining two strength values is then reported.

It is felt that this blanket-type disqualifying percentage should be reappraised from a statistical point of view, since it is very possible that entirely valid triplicate unconfined compressive strength values may attain this percentage simply by virtue of expected statistical fluctuation. Thus many values may be unjustly disqualified. Since unjustly disqualified strength values carry information which is as valid as that carried by their supposedly more reliable neighbors, uncritical adherence to such a blanket-type disqualifying percentage causes needless loss of information. In addition, bias is introduced when any strength observation is wrongfully discarded.

In summary, this paper deals with triplicate unconfined compressive strength testing, and outlines a procedure which attempts to control the rate of wrongful disqualifications by replacing the commonly used "blanket" disqualifying percentage by a percentage tailored to the specific investigation at hand. In addition, a method is given for examining the series as a whole for reliability, homogeneity and normality.

Proposed Disqualification Test

Step 1: The statistical theory of the present approach requires the existence and the estimation of a constant coefficient of variation--abbreviated CV--for the entire series of observations. The CV of any observation equals the dispersion to which that observation is subject divided by the true value that the observation is supposed to estimate. It should be a constant for all the observations of a single investigation.

A simple nomographic procedure has been devised for establishing and estimating this constant CV.

Procedure for establishing and estimating the CV:

1(a). For each set of triplicate unconfined compressive strength values, compute the ratio, r, of the range, R, of the three values to the average, $\bar{X}$, of the three values. The range is defined as the difference between the largest value and the smallest value of the three. Thus

$$r = \frac{R}{\bar{X}} = \frac{X_{max} - X_{min}}{(X_1 + X_2 + X_3)/3}$$

1(b). Arrange all the r values so obtained in ascending order of magnitude. This can easily be done by plotting them on ordinary graph paper.

1(c). Choose approximately thirty well spaced r values. For each selected r value, find the number of other r values less than it and express this number as a percentage of the total number of r values.

1(d). Plot each percentage against its corresponding r value on the nomograph, using scale A for the r values and scale B for the percentages.

1(e). Fit the thirty points so obtained with a straight line - hereafter called the CV line - passing through the origin. If the points lie reasonably close to the straight line, then constancy of the CV is established and the proposed test is applicable. (Questions of objective fit and closeness criteria are touched upon in this discussion).

Outliers, if present, will tend to unduly enlarge r. This will cause the r pattery to form an arched rather than straight line. In such cases, the points furthest from the origin should be excluded from the straight line fit. A technical though perhaps impractical refinement here is to eliminate far points until the remaining replotted points form a satisfactory straight line.

The CV itself is estimated by the value on scale A at which the CV line attains a height of 24 on scale B.

It might be noted that prior workers in this general area have worked with the assumption of constant CV (1)[4]. In addition, a considerable number of experimental sets of data have been examined for constancy of the CV at the Iowa Engineering Experiment Station, and it has been found to hold in every case.

Step 2. Upon the establishment and estimation of the constant CV, it is now possible to test for possible incorrect unconfined compressive strength values. The procedure is as follows:

Procedure for disqualification of extreme strength values:

2a. For each set of triplicate values compute the ratio, U, of the largest value ($X_{max}$) - the average value ($\bar{X}$) to the average value ($\bar{X}$). Thus

$$U = \frac{X_{max} - \bar{X}}{\bar{X}} \qquad (2)$$

2b. For each set of triplicate values, compute the ratio, V, of the average value ($\bar{X}$) - smallest value ($X_{min}$) to the average value ($\bar{X}$). Thus

$$V = \frac{\bar{X} - X_{min}}{\bar{X}} \qquad (3)$$

2c. Enter scale D at the total number of triplicate sets. Through this point draw a horizontal line until it intersects the CV line through the origin. Read on scale A the value t of the abscissa of this intersection point.

2d. t is the critical value for both U and V. Any triplicate whose U exceeds t should have its $X_{max}$ discarded; similarly, any triplicate whose V exceeds t should have its $X_{min}$ discarded. In other words the t value, when expressed in percentage form, is the disqualifying percentage for the investigation at hand.

It must be realized that, although the suggested procedure controls the rate of wrongful disqualifications, it cannot reduce this rate to zero. It is therefore possible that valid observations may be disqualified. Similarly, a certain number of outliers will not be detected.

Wrongful disqualifications can occur either when all three members of the triplicate set are subject only to normal experimental variation or possibly because the two remaining values are, in fact, the illegitimate ones. The investigator seeking additional controls for errors of this type may wish to cross-check the disqualifications suggested by the present procedure against the disqualifications suggested by the magnitude of the corresponding residuals from fitted regression functions (2). This cross check is not further discussed in this paper.

Where, however, the cross-check is not used, it is recommended that if one observation is disqualified, the middle observation of the original three then be reported. If it should happen that both U and V are extreme for one triplicate set, the entire triplicate set should then be discarded.

Step 3. In some cases it may be of interest to check on the reliability of the investigation as a whole. This may be necessary for many reasons, such as suspected unreliability of the operator, non-normality, or inhomogeneity of the material under test.

Criterion for the reliability of the entire investigation:

3(a). Arrange all the U values in ascending order of magnitude. This is most easily done by plotting them on ordinary graph paper.

3(b). Select approximately thirty well spaced U values. For each selected U value, find the number of other U values that are less than the selected U value and express this number as a percentage of the total number of U values.

3(c). Using the nomograph, plot on scale E each percentage obtained in 3b against its corresponding U value on scale A.

3(d). Fit the points so obtained by a straight line - hereafter called the U line - through the origin.

3(e). Similarly, do 3(a), 3(b) and 3(c) and 3(d) for V so as to obtain a V line.

The extent of non-coincidence of the three lines obtained in 1(d), 3(d) and 3(e), and the extent to which the three sets of points fail to be fitted by the CV line, indeed the actual shape of the sets themselves, will provide clues concerning series-wide unreliability, inhomogeneity and non-normality. For example, inhomogeneity, in the sense of more than one underlying coefficient of variation, will cause the three sets to form similar "S" shaped curves, arching first downward then upward, the first arch typically being the more pronounced. This effect is similar to that arising under "inadvertent plot splitting" in half-normal plot analyses (3), and is due to similar causes. Again, certain types of operator fabrication will manifest themselves in distinctive patterns. For example, fabricating a triplicate from a single determination by adding and subtracting fixed proportions of the single determination will cause a vertical discontinuity to appear in all three plots. On the other hand, fabricating a triplicate from a pair of determinations by interpolation will cause a configuration similar to but typically less extreme than that arising under inhomogeneity.

Should serious series-wide non-normality be uncovered, the clash of non-normal data with normal theory should, as a rule, be resolvable in favor of the theory. In other words, non-normality of data often will have an identifiable and removable cause.

## Examples

The proposed technique is now applied to two series of triplicate determinations. The first example involves 134 triplicate sets of unconfined compressive strength determinations of soil-calcium lignosulfonate-aluminum sulfate specimens (4). The second example involves 152 triplicate sets of unconfined compressive strength determinations of soil-lime-sodium silicate specimens (5).

As shown in Fig. 1, the estimated CV for the first example is 0.048, and the critical t is 0.114, corresponding to a disqualifying percentage of 11.4. None of the 134 triplets were disqualified by this criterion. As shown in Fig. 2, the CV line and V line coincide, with the U-points and V-points falling close to this joint line. All indications therefore point to the fact that this investigator was in thorough control of his experiment.

The estimated CV for the second example is approximately 0.074, indicating a degree of experimental precision lower than that of the first example. This lower precision probably does not represent an operator effect, but is probably due to the well known rapid jell-forming ability of sodium silicate. Low precision does not by itself constitute evidence of experimental inefficiency but, as is likely in the present case, can be the result of inherent material properties. The critical t-value for this example is approximately 0.182, corresponding to a disqualifying percentage of 18.2. As regards the reliability check carried out in Fig. 2, the CV line, U line and V line are seen not to coincide. Moreover, the U points and V points do not lie close to their respective lines. The tendency to downward curvature exhibited by both the U points and V points suggests the possibility of inhomogeneity of experimental material.

It is important to note that the critical percentage of 11.4 for the first experimental series is near the blanket percentage of 10%, which, parenthetically, is exceeded by 3 triplicate sets of this series. This 10% is also exceeded by 38 triplicates of the second series. Use of the critical percentage "tailor-made" to inherent experimental variability thus leads to a reduction in the number of disqualifications in the case of both experimental series. These are, namely, zero versus 3 for example No. 1 and 18 versus 38 for example No. 2.

Note that the two types of nomographic computations shown in Figs. 1 and 2 can be performed on a single nomograph. A sample of such a nomograph, called "Outlier Paper" is given in Fig. 3.

## Discussion

It is planned to discuss the details underlying the proposed procedure in a separate technical publication. However it seems appropriate to give a brief theoretical discussion here.

The Outlier Paper of Fig. 3 is based upon the following facts.

(A) The ratio, $\frac{r}{CV}$, has approximately the distribution of the range of three unit normal deviates, and $\frac{U}{CV}$ and $\frac{V}{CV}$ have approximately the distribution of the largest minus the average of three unit normal deviates. Verifying computations indicate that these approximations are sufficiently exact as long as the coefficient of variation is less than 0.15. Scales A and B represent inverse probability transformations corresponding to the above two functions of unit normal variables. The linearizing property of inverse probability transformations has been exploited before (3).

(B) In view of the above, the cumulative distribution functions for r, U and V are straight lines through the origin and have slope of $\frac{1}{CV}$ when plotted on the outlier paper. This enables the CV line, which is in fact the estimated cumulative distribution of r, to yield critical values for U and V i.e. to be used as if it were in fact the cumulative distribution function of U and V.

It is important to note that, ideally, the construction of the CV line should be based on a statistic that is as insensitive as possible to outliers, whereas the disqualifying percentage derived from this CV line should be applied to statistics that are as sensitive as possible to outliers. Triplicate observations lend themselves only partially to these objectives if, as is assumed in this paper, both large and small outliers are involved. In view of this, the plot of the partially sensitive r values may show some downward curvature. In such cases, as has already been recommended, the CV line should be fitted on the basis of the r points less likely to be contaminated by the outliers, i.e. the r points closer to the origin.

In cases where it is known that only large outliers are present, an ideal insensitive statistic is the ratio of the difference to the mean of the middle and smallest observation.

With quadruple observations, almost complete insensitivity and sensitivity can be achieved even when both large and small outliers are present. A suitable insensitive statistic is the ratio of the difference to the average of the middle two observations, and a suitable sensitive statistic is the largest minus the average of the middle two, divided by the average of the middle two observations.

(C) The method of obtaining the disqualifying percentage is based upon the "multiple-comparison" point of view that experimental series not containing outliers, regardless of their length should suffer no disqualification with probability $\frac{1}{2}$. It is realized that other points of view regarding the question of risk will lead to different D scales.

It is of interest to note the manner in which the critical disqualifying values for U and V depend upon the total number of triplicate sets and also upon the constant coefficient of variation. When the number of triplicate sets increases, the critical t value increases, which means that the critical U and V values also increase. This follows from the present point of view regarding risk and may be explained by the fact that, since a greater number of triplicates are involved, natural experimental variation is expected to produce greater numbers of extreme U and V values. The critical t value also increases with increasing CV. This is a reflection of the fact that the data are expected to be more erratic whenever the natural experimental error, of which the constant CV is a measure, is large.

Further theoretical considerations revolve about the manner of fitting the CV line and the manner of assessing the goodness-of-fit of the r, U and V points to this line. As a rule, an eye-fit will be adequate for the CV line, as other more sophisticated methods probably will not provide sufficiently greater accuracy to compensate for their greater computational complexities. A measure of goodness-of-fit is provided by the maximum vertical deviation, in units of percentage, of the thirty points from the straight line. This deviation may be approximately judged in terms of the known distribution of the maximum vertical discrepancy between a population CDF and its corresponding sample CDF (6). However this distribution theory should be taken only as a rough guide since (a) only thirty points of the sample CDF have been plotted, (b) the CDF to which this sample CDF is being compared is a fitted rather than a true CDF and (c) whatever outliers are present are actually contributing to the discrepancy between the two CDF's; alternatively, if one attempts to eliminate outliers by the refinement given in 1(e), maximum vertical deviations will arise that are considerably smaller than those expected according to the standard distribution theory.

Acknowledgments

The authors thank H. O. Hartley and J. J. Moder for helpful contributions to this investigation. Appreciation is also due to R. Altenberger and M. Grimes for high-speed computational assistance.

The subject matter of this report was obtained as part of the research being carried on at the Engineering Experiment Station of the Iowa State University, under contract with the Iowa Highway Research Board of the Iowa State Highway Commission.

Appreciation is due to the Statistical Laboratory of Iowa State University for making available professional and high-speed computer time.

Footnotes: David, Davidson and O'Flaherty

1. For a similar treatment of this subject, reference is made to ASTM Tentative Recommended Practice for Dealing with Outlying Observations to be issued later this year under E 000 - 61 T. (Number will be assigned following its adoption by Committee E-11 on Quality Control of Materials.)
2. Associate Professor of Statistics, Professor of Civil (Soil) Engineering, and Assistant Professor of Engineering Graphics and Civil Engineering Graduate Student, respectively, Iowa State University of Science and Technology, Ames, Iowa.
3. ASTM Method of Test for Compressive Strength of Hydraulic Cement Mortars (Using 2-in. Cube Specimens) (c 109 - 58). 1958 Book of ASTM Standards, Part 4, p. 130.
4. The boldface numbers refer to the list of references appended to this paper.
4. When the ultimate interest is in estimating a mean, another point of view is offered by F. J. Anscombe, "Rejection of Outliers," <u>Technometrics</u>, Vol. 2, pp. 123 ff. (1960).

REFERENCES

(1) Waddell, J. J., "Quality Control in the Construction Industry," Industrial Quality Control, Vol. XVII, No. 7 (1961).

(2) Deming, W. E., "Statistical Adjustment of Data, " John Wiley and Sons, Inc., New York, N. Y., pp. 168-171 (1943).

(3) Daniel, C., "Use of Half-normal Plots in Interpreting Factorial Two-level Experiments", Technometrics, Vol. 1 (1959), pp. 333-335.

(4) Demirel, T., "Use of Calcium Lignosulfonate with Aluminum Sulfate for Soil Stabilization," unpublished M. S. Thesis, Iowa State University of Science and Technology Library, Ames, Iowa (1959).

(5) Ruff, C. G., "Sodium Silicate-Lime Stabilization," unpublished M. S. Thesis, Iowa State University of Science and Technology Library, Ames, Iowa (1960).

(6) Miller, L. H., "Table of Percentage Points of Kolmogorov Statistics," Journal of the American Statistical Association, Vol. 51 (1956), pp. 111-21.

(7) Pearson, E. S., and Harley, H. O., "Biometrika Tables for Statisticians," Cambridge University Press, Cambridge, Vol. I (1956), Table 23.

(8) Youden, W. J., "Sets of Three Measurements," Scientific Monthly, Vol. 77 (1953), pp. 143-147.

(9) Youden, W. J., "The Fallacy of the Best Two Out of Three," Natl. Bur. Standards (U.S.), Technical News Bulletin, Vol. 33 (1949), p. 77.

REFERENCES, Con't.

(10) Nair, K. R., "The Distribution of the Extreme Deviate from the Sample Mean and its Studentized Form," Biometrika, Vol. 35 (1948).

(11) Dixon, W. J., "Analysis of Extreme Values," Annuals of Mathematical Statistics, Vol. 21 (1950) pp. 488-506.

(12) Kruskal, W. H., "Some Remarks on Wild Observations," Technometrics, Vol. 2 (1960), pp. 1-3.

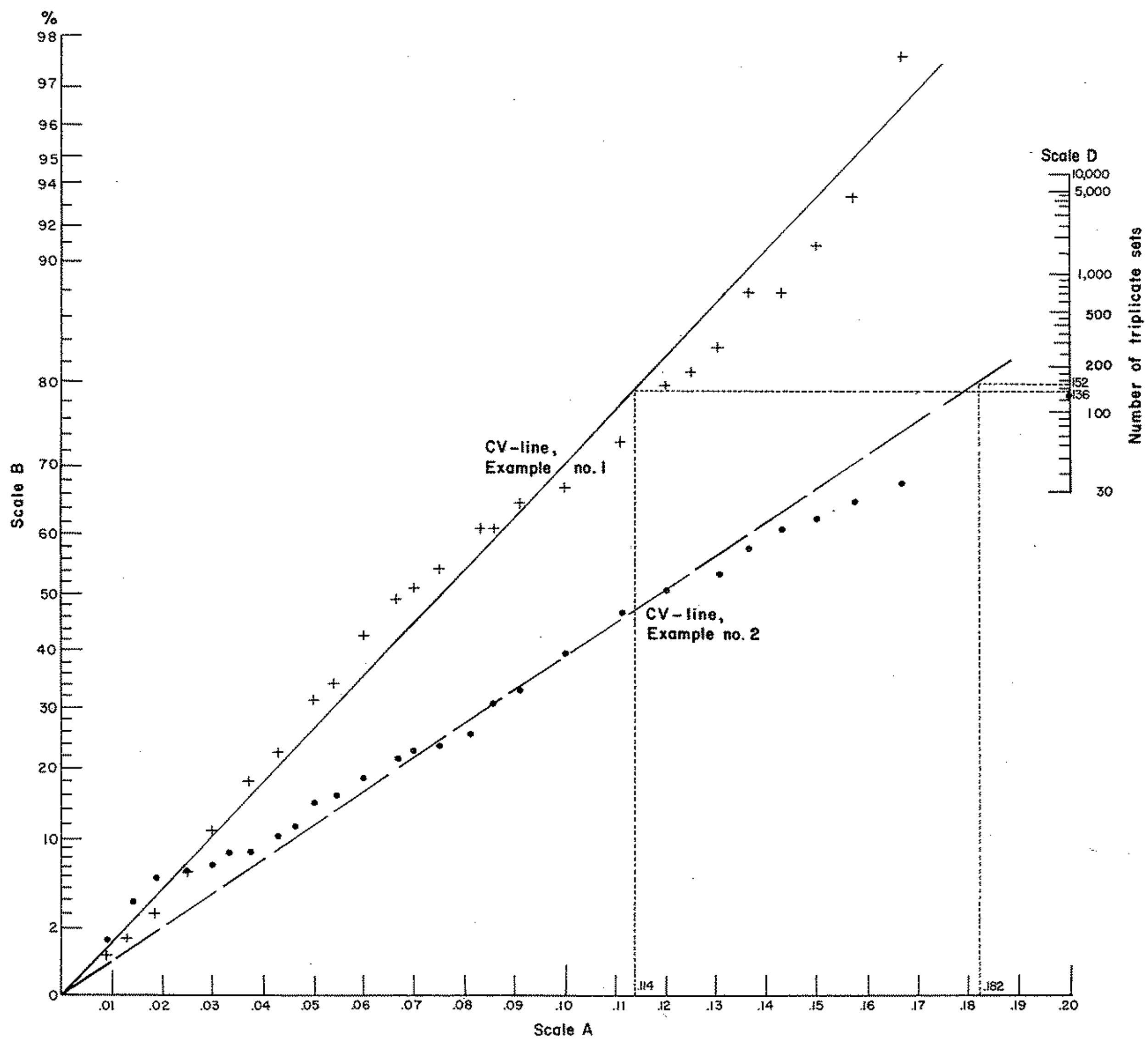


Fig. 1.—Nomographic computation of disqualifying critical values

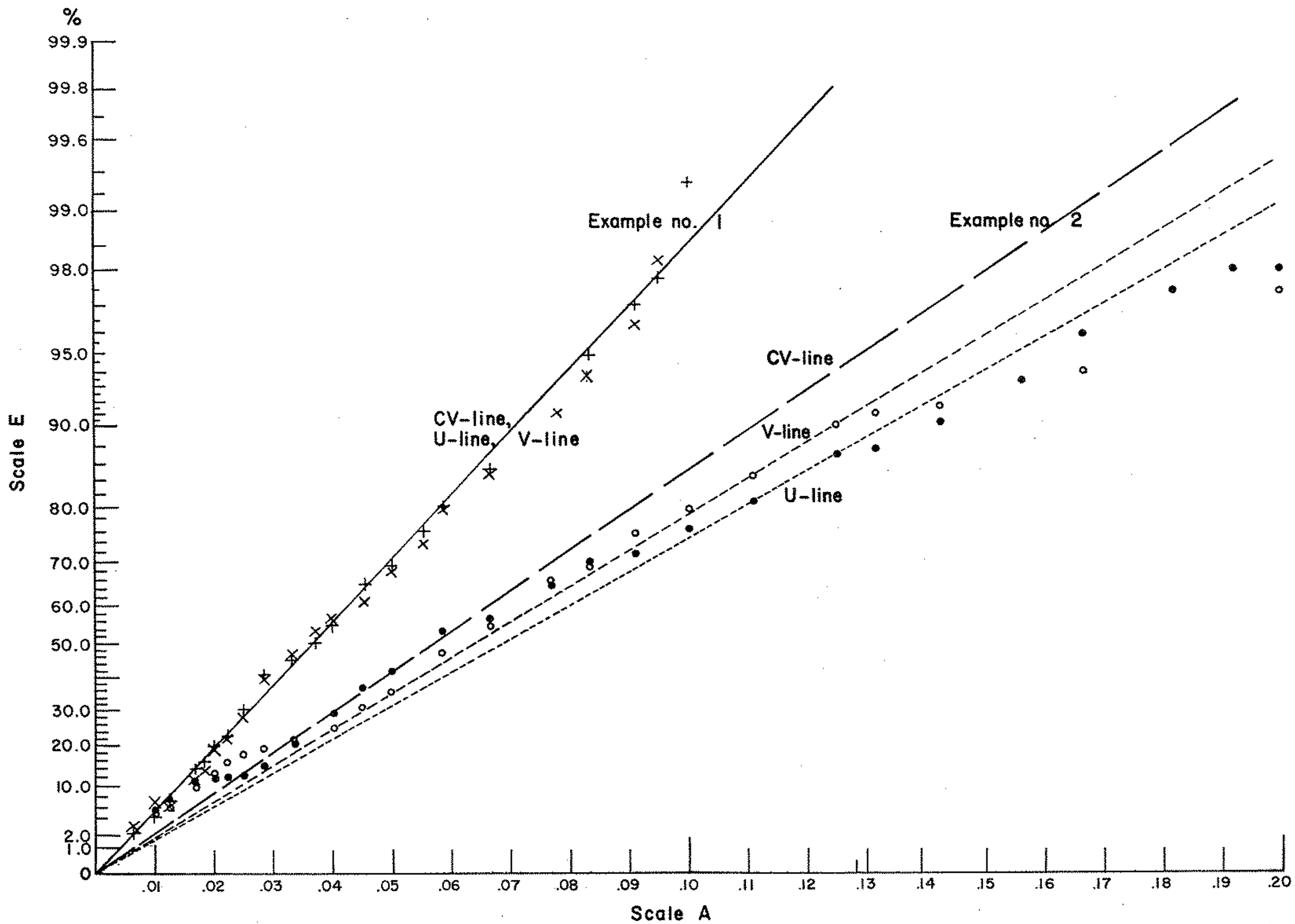


Fig. 2.— Nomographic assessment of series reliability

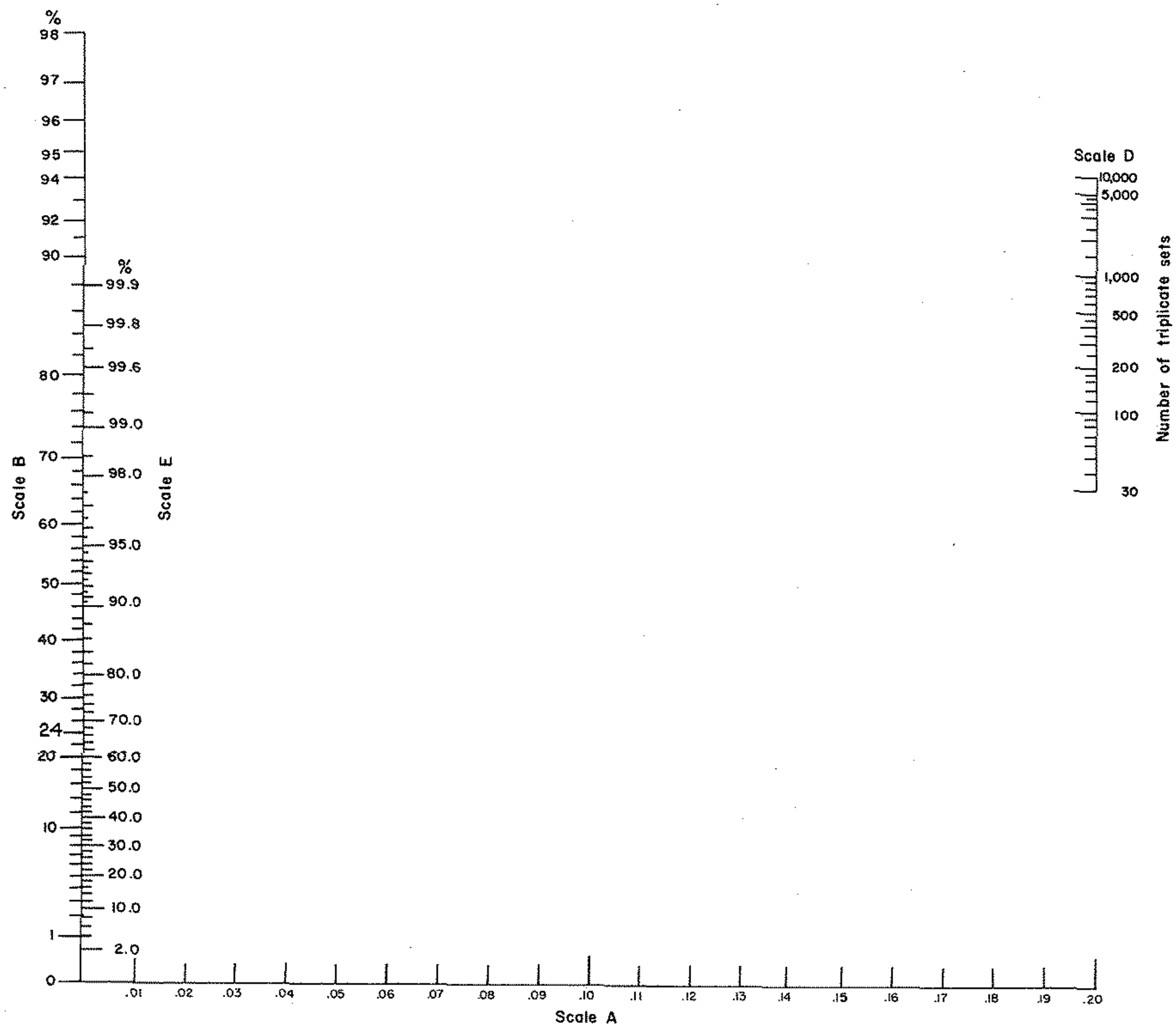


Fig. 3. — Outlier paper