A PRELIMINARY EVALUATION OF CATEGORICAL FIELD OBSERVATIONS FOR REGIONAL STREAM SEDIMENT SAMPLES

(82 F, K)

By **P. Matysek, W. K. Fletcher, A. J. Sinclair, and A. Bentzen**
Department of Geological Sciences, University of British Columbia

INTRODUCTION

We are in the process of examining regional stream geochemical data obtained during a joint Federal-Provincial Uranium Reconnaissance Program of NTS map-areas 82 F and K released in 1978. The purposes of our study are to evaluate the usefulness of individual variables coded in the course of these surveys, to develop a statistical procedure for extracting useful information from the data, and to utilize the data base as an effective means of defining problems of geological interest that warrant further investigation. Our work is developmental in nature and is confined to data for NTS map-area 82F.

Initial work (Sinclair and Fletcher, 1980) considered only numeric (concentration) data and emphasized a systematic approach to its evaluation using standard statistical procedures. However, in addition to the quantitative data the Uranium Reconnaissance Program files contain field observations describing the drainage sediments and collection site for each sample. For example, in addition to rock type, coded comments are recorded to describe the presence of contamination, the nature of the bank material, water and sediment colour, texture, the presence of organics or precipitates, and such physiographic features as landscape maturity, drainage pattern, and stream class. Although similar data are often collected as part of drainage surveys, they are so seldom utilized in any systematic fashion that a recent review (Meyer, *et al.*, 1979) suggested limiting field observation to those of proven significance. Our ongoing studies attempt to establish which field observations are significant and to assess their influence on metal content of the drainage sediments. Two complementary procedures are being utilized:

(1) Contingency tables and the chi-square test, and
(2) Duncan's multiple range test.

CONTINGENCY TABLES

Two-way contingency tables represent an ordered arrangement of counts of 'intersection' of pairs of variables. A simple example is illustrated in Table 1 for rock type versus stream velocity for 1 318 stream sediment samples from NTS map-area 82F. The count of 245 at the intersection of the GRNT-35 and MOD columns means that 245 sediments from the area are characterized by both these features. With such a table one can quickly, if subjectively, evaluate the distribution of one variable in relation to another. For example, in Table 1 we can look at the distribution of samples in velocity categories for individual rock types (that is, distribution of counts along rows) and compare the distribution for one rock type with the distribution for another. Conversely the data can be viewed in terms of the distribution of counts along each column. That is, dominant stream velocity categories may differ over each rock type and different distribution patterns may occur in different columns.

It is apparent that contingency tables permit rapid qualitative evaluation of paired variables although the importance of this feature is not particularly obvious in viewing Table 1. However, the usefulness of a computer-generated two-way contingency table for 40 or more variables, as is the case with stream sediment samples collected during the Uranium Reconnaissance Program, is apparent. Two-way tables for such a large number of variables are difficult and impractical to obtain manually. They can, however, be produced with ease on the computer and a program in FORTRAN designed specifically for dealing with publically available magnetic tapes of regional geochemical data for the Uranium Reconnaissance Program has been developed. Because the table is symmetrical about the main diagonal, only half is printed by the program. Row and column totals are also output. For 182 categories of about 15 variables, 45 pages of computer output are necessary to generate an entire contingency table.

In reality such a two-way table consists of a series of self-contained subsets of which Table 1 is one example. Using a chi-square test these subsets can be examined rigorously to see whether one variable is statistically dependent on the second variable.

As an application of the chi-square test, consider the data of Table 1. The chi-square test requires that there are no zero values and that no more than 20 per cent of the values are less than 5. Rows and columns must be grouped or eliminated to meet these conditions. In the case of Table 1 a possible grouping leads to the arrangement in Table 2.

In conducting a chi-square test, it is assumed that summations of rows and columns represent a best estimate of independence of the two variables. Consequently, these summations are used to estimate expected values for each intersection according to the formula

$$E_{ij} = \frac{r_i c_j}{N}$$

where is the sum of row i and c_j is the sum of column j.

The values determined represent an expected value assuming that the variables are independent. Differences between observed and calculated values $(O_{ij} - E_{ij})$ are examined by the chi-square test to evaluate whether or not they could result from random sampling:

$$X^2_{calc.} = \sum_{ij} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

If differences are small $X^2_{calc.}$ is small, and the two variables are said to be independent. If $X^2_{calc.}$ is greater than some critical value, $X^2_{(d.f.,\ L)}$, that depends on the number of degress of freedom and the level of significance of the test, then one variable is said to be dependent on the other.

For the example of Table 2, $X^2_{calc.}$ is 32.63 which is much greater than the critical value of 18.31 obtained from tables such as those given in Krumbein and Graybill (1965) for 10 degrees of freedom and a test level of 0.05. Consequently, we can say with assurance that stream velocity depends on rock type; more specifically, some rock types are characterized by high stream velocities and others by low velocities. Because higher stream velocities generally mean steeper gradients we see the physiographic information contained within such tests. Within the test area, for example, siltstone (SLTE) is generally resistant and forms steep slopes in contrast to gneiss (GNSS) which is recessive and more prevalent in valleys.

A similar test can be done between any two variables. For example, we can compare sediment colour against rock type, bank type versus stream velocity, and so on.

DUNCAN'S MULTIPLE RANGE TEST

METHOD

Duncan's multiple range test (Duncan, 1955, 1957) provides a method of testing if differences among a group of means are significant. It has previously been applied to regional geochemical data by Miesch (1976) and Doyle and Fletcher (1979). The test assumes that the means $m_1, m_2 \ldots\ldots m_n$ are independently drawn from 'n' normal populations having true means of $\mu_1, \mu_2 \ldots\ldots \mu_n$ respectively. However, as previously reported (Sinclair and Fletcher, 1980) much of the metal concentration data from the Uranium Reconnaissance Program is log-normally distributed and it is often multimodal. Consequently, before the significance of field parameters for a background population could be evaluated using Duncan's multiple range test, it was necessary to log-transform the data and eliminate anomalous results. For this report the procedures involved and results will be illustrated with respect to streams draining granites (GRNT-35) in map-area 82F.

The first step is to partition the log-probability plot, using the method of Sinclair (1976), for each element into low (probably background) and high (probably anomalous) populations. Anomalous samples are then rejected leaving only background samples for classification into groups based on the field observations. For example, considering copper in sediments associated with GRNT-35, 12 samples (from a total of 485) are rejected as anomalous; if sediment colour is the field observation of interest the remaining 473 sediments can then be divided into seven (red, white, black, yellow, green, grey, and pink) groups and log means and standard deviations calculated. However, the only colours recorded with reasonable frequency are red (n = 341), white (105), and black (23) with corresponding means of 9, 9, and 15 parts per million (ppm). The significance of the differences among these means are then calculated (0.05 confidence level) using Duncan's multiple range test. Results of the test establish that concentrations of copper in white and red sediments are indistinguishable but those in black sediments are significantly greater. The results are conveniently presented as Venn diagrams in which overlapping or shared circles indicate groups whose means are not significantly different (Fig. 52; figures at end of text, pages 155 to 158).

RESULTS

Data for 11 of the elements reported were subdivided into groups according to their classification with respect to four sediment characteristics (fines, sand, organic content, and colour) and six environmental parameters (physiography, water flow, stream class, drainage pattern, bank type, and contamination). The significance of differences of means among groups were then tested and presented as Venn diagrams as shown on Figures 52 to 54. Results can be summarized as follows:

(1) Bulk composition — fines (absent; minor <33 per cent; medium 33 to 67 per cent; major >67 per cent, Fig. 52). Except for tungsten, variation in the content of fines in the sediment has no apparent influence on metal concentrations.

(2) Bulk composition — sand (absent; minor <33 per cent; medium 33 to 67 per cent; major >67 per cent). Although lead, zinc, uranium, manganese, and mercury concentrations in sediments estimated to contain medium amounts of sand are lower than those in other sediments, it is only for lead that this group forms a statistically independent population. Similarly, sediments in which sand is absent or a minor component contain relatively high concentrations of zinc, uranium, manganese, and mercury without their constituting a significantly different statistical group. Sand content has no apparent influence on concentrations of copper, nickel, molybdenum, iron, and cobalt.

(3) Bulk composition – organic (absent; minor <33 per cent; medium 33 to 67 per cent; major >67 per cent, Fig. 53). Sediments can be divided according to the presence or absence of organic matter; those with no organic matter contain significantly lower concentrations of zinc, lead, uranium, manganese, and mercury. Copper, molybdenum, iron, and cobalt do not appear to be affected by variations in organic content. The difference between the presence of minor or medium quantities of organic matter does not seem to be important for any of the elements.

(4) Sediment colour (red, white, black, yellow, green, grey, and pink, Fig. 54). As might be anticipated, red sediments contain higher average concentrations of iron and manganese than white or black sediments; uranium shows the same pattern. Between the red and white sediments the difference in means for these elements is also significant. In contrast, the greater average concentrations of copper, nickel, lead, and mercury form a statistically significant group associated with black sediments. For lead and mercury, red sediments comprise a statistically independent group of intermediate concentrations.

(5) Physiography (plateau, hilly undulating, mountainous mature, mountainous youthful). Zinc, lead, manganese, copper, and mercury have their lowest concentrations in the youthful category which, for the first three of these elements, forms a statistically significant group, Conversely, molybdenum (and tungsten) have their maximum values in this category. Nickel, cobalt, and iron concentrations show no relationship to the physiographic classification.

(6) Stream class (permanent, secondary, tertiary, quaternary). Zinc, molybdenum, uranium, tungsten, iron, and mercury show no significant differences in concentration related to stream class. Concentrations of copper are significantly different in each of the three classes, highest concentrations being associated with secondary drainages. Secondary drainage also comprises a significantly distinct group of high concentrations for cobalt and nickel. In contrast, maximum manganese content is associated with quaternary streams.

(7) Water flow (zero, slow, moderate, fast, torrential). Molybdenum, uranium, tungsten, iron, and manganese concentrations cannot be subdivided on the basis of flow velocities. For the remaining elements (except cobalt) there is a tendency for maximum values to be associated with slow flow rates. There is, however, considerable overlap between the groups.

(8) Drainage pattern (poorly defined, dendritic, herringbone, rectangular, discontinuous, basinal, other). Only copper and lead form significantly distinct groups with their lowest concentrations in areas with poor and herringbone drainages respectively. Lowest concentrations of zinc, manganese, and mercury are also found in areas of poor drainage but concentrations are not significantly different between those with dendritic or herringbone patterns.

(9) Bank type (undefined, alluvial, colluvial, glacial till, glacial outwash, bare rock, talus, organic predominant). With seven categories sample size is often small. Nevertheless, relatively high nickel, molybdenum, and cobalt values are significantly associated with talus slopes. Results for uranium and iron show a similar pattern but they are not significantly different to all other groups.

(10) Contamination (none, possible, probable, definite, mining, agriculture, forestry, domestic). Of the six categories of contamination only the association of low concentrations of copper and cobalt with domestic contamination is significantly different to concentrations in all other categories.

DISCUSSION

Despite their very qualitative, subjective character it is apparent that the field observations can be related to variations in the trace element content of sediments associated with a single rock unit, in this case GRNT-35. Under these circumstances it is of obvious interest to:

(1) Consider if the relationships observed are consistent with factors known to influence trace element behaviour, and
(2) To establish the interactions between the field parameters with a view to eliminating those that are either redundant or appear to have little influence on trace element concentrations.

Considerably more work is required on both topics, however, a summary (Fig. 55) does indicate those field parameters which influence the greatest number of elements. From this we note that sediment colour, physiography, and bank type significantly influence concentrations of 8 out of the 11 elements whereas, rather surprisingly, content of fines only influences zinc concentrations. Conversely, the susceptibility of an element to influence by the field parameters decreases from lead, which is significantly affected by eight factors, in the order manganese and zinc (7), copper, nickel, and uranium (6), cobalt and mercury (5), tungsten (4), and iron and molybdenum (2).

Clearly the composition of a particular sediment reflects the interaction and relative strengths of many factors which, by reinforcing or counteracting each other, impart a low, average, or high metal content. At present we do not know why many of the factors, especially those related to physiography, produce the results observed. However, it is encouraging to note that the results are in accord with some well-known controls on the behaviour of trace elements in sediments. For example, scavenging of metals by organic matter and hydrous oxide precipitates probably cause the associations between zinc, lead, uranium, manganese, and organics and mercury and of zinc, copper, lead, uranium, manganese, mercury, and iron with red sediments.

Finally, the need for caution in interpreting interacting factors should be emphasized. For example, relatively low background concentrations of copper are apparently associated with streams receiving domestic contaminants. Although this may reflect good housekeeping it seems more likely that it means that townsites were developed in valleys, alongside quaternary drainages which have low copper contents. Higher in the mountains secondary and tertiary drainages have relatively higher copper contents.

CONCLUSIONS

As far as we are aware, this is the first systematic attempt to evaluate the significance of field observations in relation to background variations in metal contents of drainage sediments. We conclude:

(1) Despite their subject character field observations can be related to significant variations in metal content of drainage sediments associated with a single rock unit.
(2) Two-way contingency tables are a useful means for rapidly identifying those paired categorical variables for which enough samples exist for statistical analysis. For practical use, however, a computerized system of generating such tables is essential.
(3) Subsets from large two-way contingency tables for regional stream sediment samples can be tested rigorously for dependence or independence using a chi-square test. An example

from NTS 82F map-area indicated preferential occurrence of certain rock types in certain physiographic environments as indicated by stream velocity.

(4) Duncan's multiple range test, used in conjunction with probability plots, enables the significance of field observations to be systematically related to variations in background metal content of sediments. This provides a basis for studying the interactions of environmental factors and determining which are most relevant to geochemical exploration programs.

ACKNOWLEDGMENTS

This work has been supported financially by a grant from the British Columbia Ministry of Energy, Mines and Petroleum Resources to W. K. Fletcher and A. J. Sinclair.

REFERENCES

Doyle, P. J. and Fletcher, W. K. (1979): Regional Geochemical Mapping in Areas Lacking Surface Drainages: Cu, Fe, Mn and Zn Content of Overburden and Soil in South Central Saskatchewan, *Cdn. Jour. Earth Sci.*, Vol. 16, pp. 1086-1093.

Duncan, D. B. (1955): Multiple Range and Multiple F Test, *Biometrics*, Vol. 11, pp. 1-42.

.... (1957): Multiple Range Tests for Correlated and Heterosiedostic Means, *Biometrics*, Vol. 13, pp. 164-176.

Krumbein, W. C. and Graybill, F. A. (1965): An Introduction to Statistical Models in Geology, *McGraw-Hill Book Company Inc.*, New York, N.Y., 475 pp.

Meyer, W. T., Theobald, P. K., and Bloom, H. (1979): Stream Sediment Geochemistry, *in* Geophysics and Geochemistry in the Search for Metallic Ores, P. J. Hood, editor, *Geol. Surv., Canada*, Econ. Geol., Rept., No. 31, pp. 411-434.

Miesch, A. T. (1976): Geochemical Survey of Missouri – Methods of Sampling, Laboratory Analysis, and Statistical Reduction of Data, *U.S.G.S.*, Prof. Paper 954-A, 39 pp.

Sinclair, A. J. (1976): Applications of Probability Graphs in Mineral Exploration, *Assoc. Explor. Geochem.*, Spec. Vol. 4, 95 pp.

Sinclair, A. J. and Fletcher, W. K. (1980): Evaluation Procedure for Geochemical Data Uranium Reconnaissance Program, *B.C. Ministry of Energy, Mines & Pet. Res.*, Geological Fieldwork, 1979, Paper 1980-1, pp. 130-141.

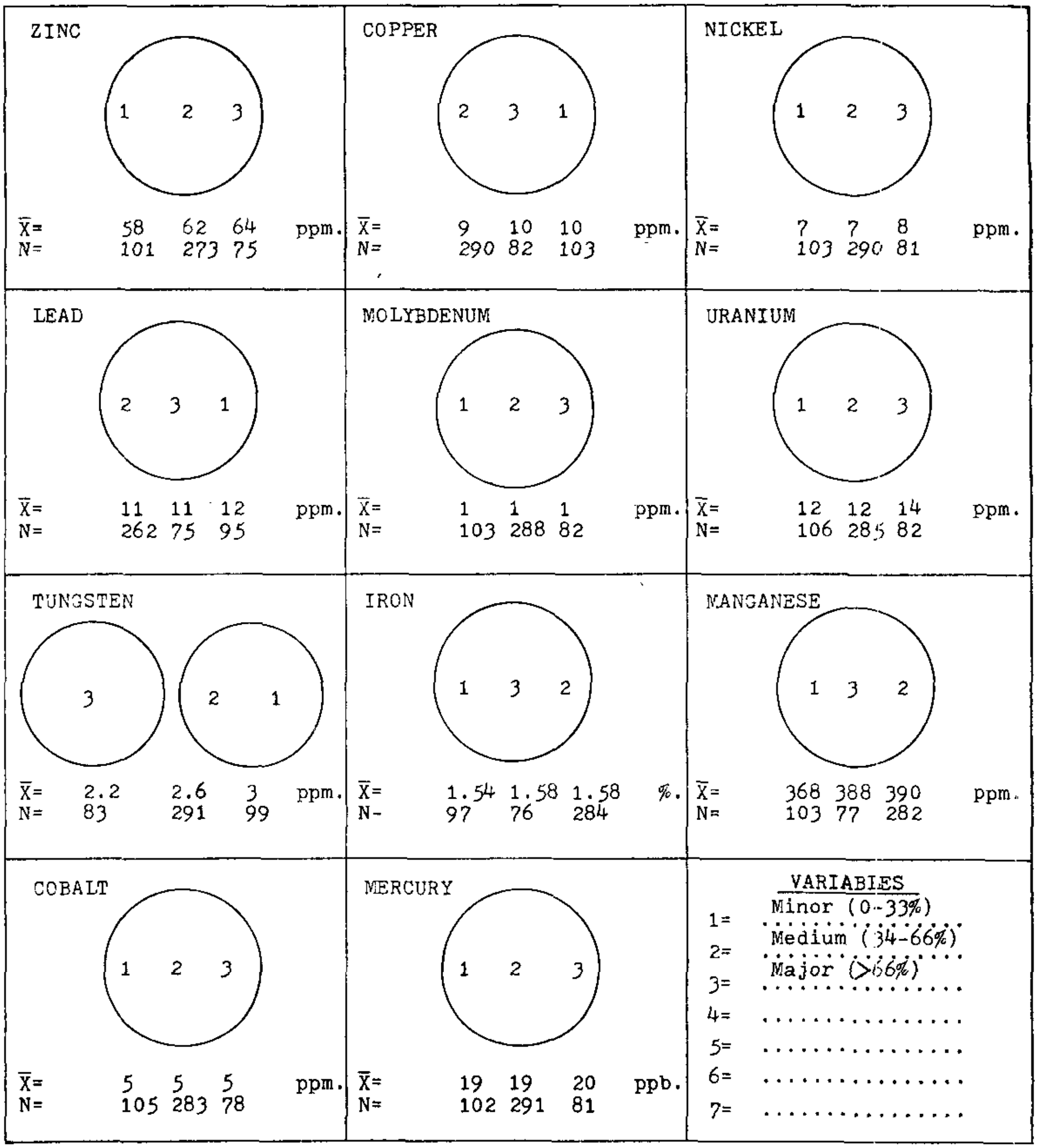

Figure 52. Duncan's multiple range test for the influence of per centage of fines on metal content of stream sediments associated with granites (GRNT-35), map-area 82F. Common or overlapping circles indicate that group means are not significantly different at the 0.05 confidence level.

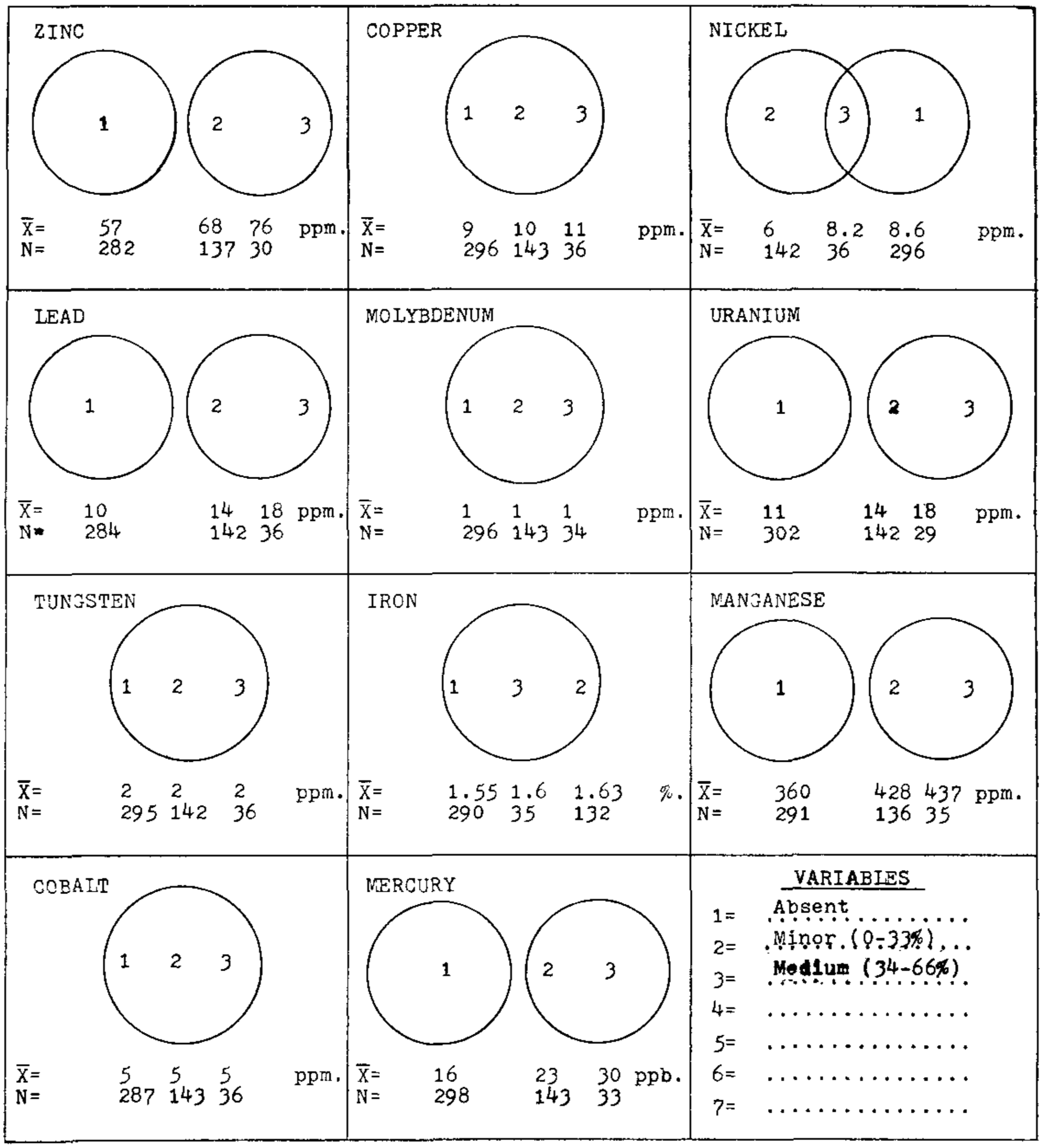

Figure 53. Duncan's multiple range test for the influence of organic matter content on metal content of stream sediments associated with granites (GRNT-35), map-area 82F. Common or overlapping circles indicate that group means are not significantly different at the 0.05 confidence level.

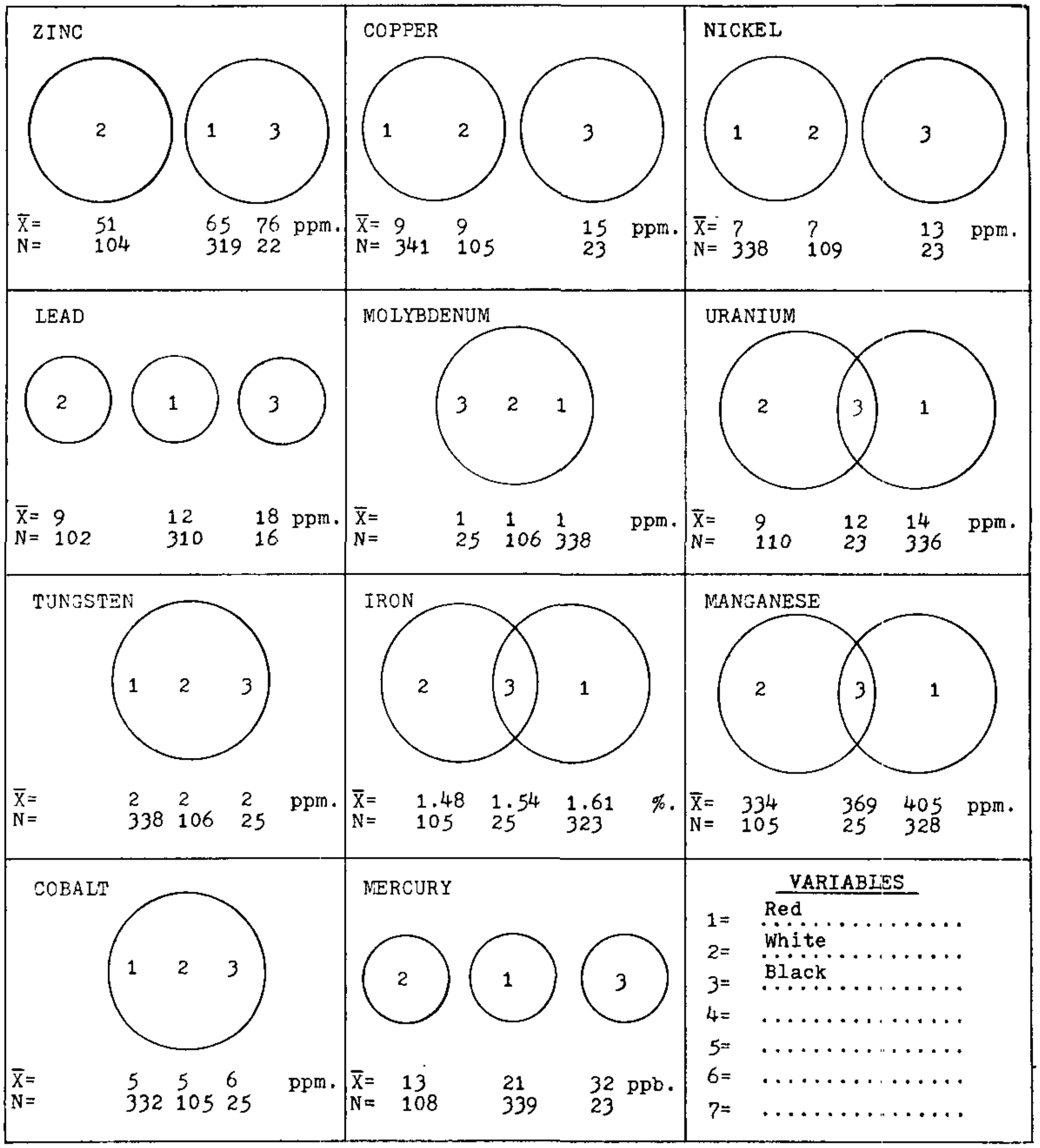

Figure 54. Duncan's multiple range test for the influence of sediment colour on metal content of stream sediments associated with granites (GRNT-35), map-area 82F. Common or overlapping circles indicate that group means are not significantly different at the 0.05 confidence level.

FACTORS AFFECTING ELEMENTAL CONCENTATIONS

	Zn	Cu	Ni	Pb	Mo	U	W	Fe	Mn	Co	Hg	Total
Bulk comp. (Fines)							X					(1)
Bulk comp. (Sands)	X			X		X	X		X		X	(6)
Bulk comp. (Orgs.)	X		X	X		X			X		X	(6)
Sediment Color	X	X	X	X		X		X	X		X	(8)
Physio-graphy	X	X		X	X	X	X		X		X	(8)
Waterflow Rate	X	X	X	X						X	X	(6)
Stream Class		X	X	X					X	X		(5)
Drain Pattern	X	X		X		X	X		X	X		(7)
Bank Type	X		X	X	X	X		X	X	X		(8)
Contamin-ation		X	X							X		(3)
Total	(7)	(6)	(6)	(8)	(2)	(6)	(4)	(2)	(7)	(5)	(5)	

Figure 55. Chart summarizing the relationships between metal concentrations and field observations for stream sediments associated with granites (GRNT-35), map-area 82F; X indicates the presence of a significant interaction on the basis of the Duncan's-multiple-range-test data.