CASE STUDY ON APPLICATION OF QUALITATIVE DATA ANALYSIS TECHNIQUES TO AN URANIUM MINERALIZATION

H.G.PEREIRA
Centro de Valorização de Recursos Minerais
Technical University of Lisbon
Instituto Superior Técnico, Av. Rovisco Pais
1096 LISBOA CODEX

ABSTRACT. Several data analysis techniques were applied to qualitative attributes in order to visualize their connections and calculate a discriminant function to classify samples into two groups. Recoding numerical values into intervals, exploratory data analysis techniques produce different views of the relationship pattern among variables and discriminant analysis quantify the capability of each variable to differenciate two groups of samples.

THE PROBLEM

During the late stages of an exploration program for Uranium, a set of 50 samples was collected in the Central-Northern region of Portugal (Pereira et al., 1984). On each sample, the following in situ attributes were observed:

. lithology
. redox condition
. metamorphism
. presence of a fault
. expressed mineralization
. presence of quartz

Also, the U_3O_8 grade and depth were measured. Furthermore, a leaching test on standard operational conditions was conducted for each sample, and the recovery was calculated.

Using this data set as an input, a battery of data analysis techniques was applied, in order to get a proper insight into the relationships among variables, and identify targets for further investigation.

The main particularity of this data set is the heterogeneity of variables - it is believed that numerical and nominal data must be, jointly, taken into account by the processing methodology and, therefore, a problem of statistical coherence arises. Hence, numerical values (depth, U_3O_8 grade and recovery) were recoded into intervals and qualitative data analysis techniques were applied to the data set as a whole, taking

C. F. Chung et al. (eds.), *Quantitative Analysis of Mineral and Energy Resources*, 617–624.

globally the available information.

Moreover, it is worth noting that there is a "special" measure - recovery - which should be treated as an "external variable" to be related to in situ attributes, and account for economic favorability.

METHODOLOGY

Bearing in mind the previously mentioned particularities of the problem, a method of combining several data analysis techniques was devised, in order to process this type of data.

The method includes two main steps: an EXPLORATORY DATA ANALYSIS to describe variables and visualize their connections and a DISCRIMINANT ANALYSIS to derive an objective criterium for classification of samples into one of two groups, related to "high" and "low" recovery.

Concerning the techniques of the first step (Exploratory Data Analysis), it was decided to use Q-Analysis (Atkin, 1974, Griffiths, 1983) in order to produce a connectivy pattern among variables, where the most common links are visualized; also, Characteristic and Correspondence Analysis were tried in this step, to obtain graphic diagrams depicting the projection of variables onto the principal factorial axes. Characteristic Analysis (Botbol, 1971) was applied in its simplest formulation, calculating and incidence-coincidence matrix (Geoffroy et al, 1972) and finding its eigenvalues/eigenvectors.

Correspondence Analysis (Benzécri, 1973; Greenacre, 1984) was used to check stability of results to the recoding procedure. Indeed, Correspondence Analysis is based on a special similarity matrix, denoted "Burt Matrix", which can be seen as the analogue of the incidence-coincidence table, in the case where all states of variables are considered.

In the second step, the objective is to derive a discriminant criterium for qualitative data. As the Classical Fisher linear function does not apply, the correspondence analysis technique was modified in order to produce only one factorial axis, where all categories of "explicative" variables are projected, according to their capability to differenciate two groups of samples. The discriminant power of this axis was assessed through the proportion of misclassified samples in the original training set. Also, a discriminant weight was calculated for each variable and an equation was derived to assign any new sample (defined by a vector of "explicative" attributes) to one of the groups, without conducting further leaching tests.

RECODING DATA

As the data set contains nominal and numeric data, it is necessary to transform variables such as depth, U_3O_8 grade and recovery into intervals, in order to assure statistical homogeneity. The interval limits were selected by a trial and error procedure, controling results through geological and metallurgical interpretation.

After several attempts, it was decided to summarize data in a

50 x 9 boolean matrix, each column of which representing the state of a dichotomous variable (presence/absence of a certain attribute or assignment to a certain interval). The recoded variables are shown in Table 1.

TABLE 1 - Recoded variables

VARIABLE	BOOLEAN CODE 1	BOOLEAN CODE 0
LITHOLOGY	SHALE	OTHER
REDOX CONDITION	REDUCED	OXIDIZED
METAMORPHISM	ABSENCE	PRESENCE
FAULT	PRESENCE	ABSENCE
EXPRESSED MINERALIZATION	PRESENCE	ABSENCE
QUARTZ	PRESENCE	ABSENCE
U_3O_8 GRADE	> .2%	< .2%
DEPTH	> 36.7m	< 36.7m
RECOVERY	> 85%	< 85%

RESULTS

Exploratory Data Analysis

The first step of the methodology was applied to the data set, recoded according to Table 1. Denoting T the matrix of recoded data, the incidence coincidence matrix C was calculated by $C = T'T$ (where T' is the transposed of T). C. contains the frequences of matching for all pairs of positive categories and can be depicted in a valuated graph, the vertices of which are variables and the arcs represent connections. The length of each arc is simply given by the absolute frequence of matches. Using a thereshold of 10 links, the graph obtained is sketched in Fig. 1.

It is worth noting, in Fig. 1, the central position occupied by the vertex "recovery > 85%", which links to all pertinent variables (presence of quartz and fault were disregarded applying the threshold).

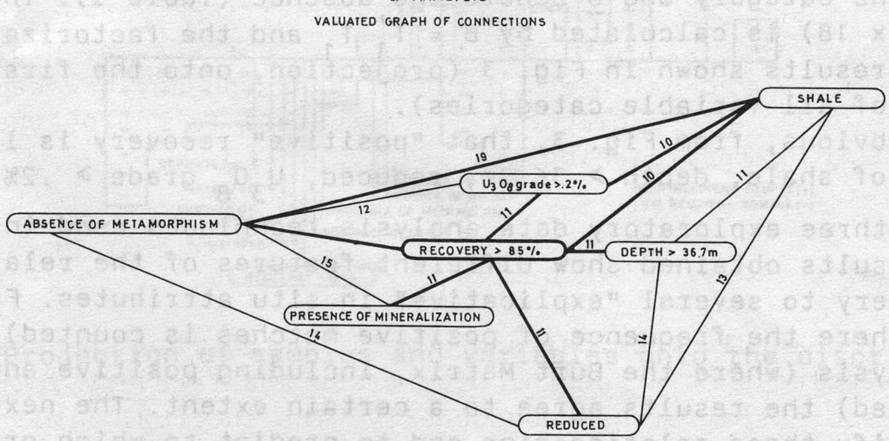

Fig. 1 - Valuated Graph of connections for threshold of 10 links.

If the eigenvalues and eigenvectors of matrix C are calculated, it is possible to project each variable onto the principal axes (eigenvectors for greater eigenvalues). These projections are shown if Fig. 2, using the first and second factors, which account for 70% of the data set variability.

Fig. 2 depicts, in a different way, the same pattern of relationships among variables. In particular, recovery is related with the pertinent "explicative" variables - shale, absence of methamorphism, depth, reducted, U_3O_8 grade, and presence of mineralization.

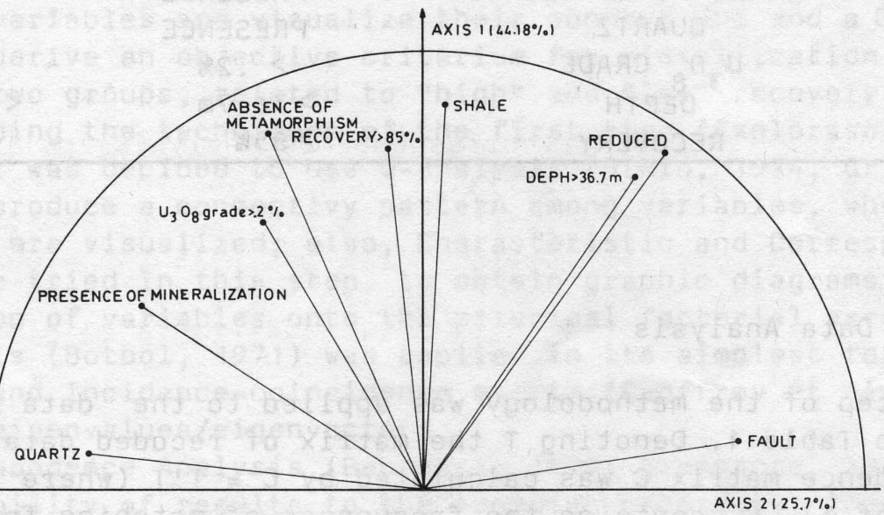

Fig. 2 - Projection of variables onto 1st and 2nd factors using incidence-coincidence matrix.

The correspondence analysis technique takes as input the Burt Matrix calculated from a new table of logical codes (T_1). In T_1, each row is a sample and there are two columns for each variable - Values 1 denote the presence of the category and 0 denote the absence (Table 1). The Burt Matrix B (18 x 18) is calculated by $B = T_1' T_1$ and the factorization of B produces the results shown in Fig. 3 (projection, onto the first and second axis, of all variable categories).

It is obvious, from Fig. 3, that "positive" recovery is linked to the presence of shale, depth > 36.7m, reduced, U_3O_8 grade > .2%.

In the three exploratory data analysis techniques used in this case study, the results obtained show different features of the relationship linking recovery to several "explicative" in situ attributes. From Q-analysis (where the frequence of positive matches is counted) to correspondence analysis (where the Burt Matrix, including positive and negative matches is used) the results agree to a certain extent. The next step aims to quantify these relationships and to predict to which group (low and high recovery) a new sample is most likely to belong.

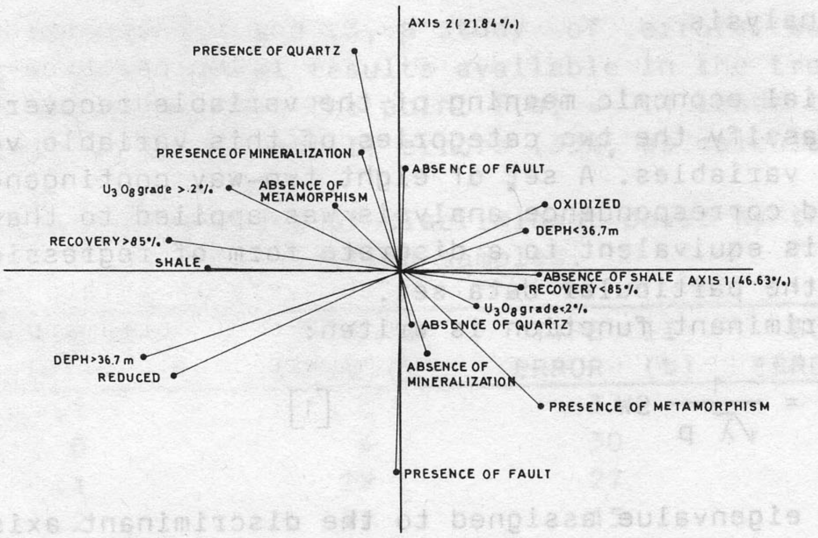

Fig. 3 - Output of correspondence analysis applied to Burt Matrix.

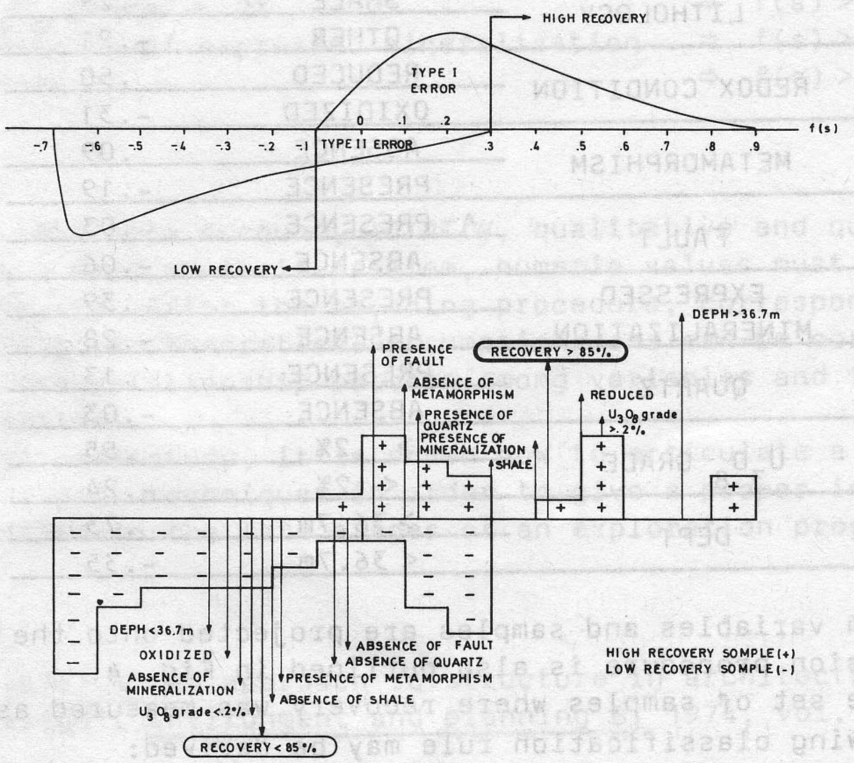

Fig. 4 - Projection of samples and variables onto the discriminant axis

Discriminant Analysis

Given the special economic meaning of the variable recovery, it was decided to cross classify the two categories of this variable versus the states of explicative variables. A set of eight two-way contingency tables were constructed and correspondence analysis was applied to that input matrix. This approach is equivalent to a discrete form of regression, adapted to the nature of the particular data set.

The discriminant function is writen:

$$f(s) = \frac{1}{\sqrt{\lambda}\; p} SW \qquad [1]$$

where λ is the eigenvalue assigned to the discriminant axis ($\lambda = .09122$)
p is the number of variables (8).
S is the boolean vector of the sample.
W is the vector of weights for each category of variable, given by their projection onto the discriminant axis (TABLE 2).

TABLE 2 - Weights of variable categories in the discriminant function

VARIABLE	CATEGORY	WEIGHT
LITHOLOGY	SHALE	.29
	OTHER	-.21
REDOX CONDITION	REDUCED	.50
	OXIDIZED	-.31
METAMORPHISM	ABSENCE	.09
	PRESENCE	-.19
FAULT	PRESENCE	.03
	ABSENCE	-.06
EXPRESSED MINERALIZATION	PRESENCE	.39
	ABSENCE	-.28
QUARTZ	PRESENCE	.13
	ABSENCE	-.03
U_3O_8 GRADE	> .2%	.55
	< .2%	-.24
DEPT	> 36.7m	.73
	< 36.7m	-.35

In Fig. 4 variables and samples are projected onto the discriminant axis. The decision procedure is also outlined in Fig. 4.

Using the set of samples where recovery was measured as a training set, the following classification rule may be derived:

IF $f(s) > .3$ THEN THE SAMPLE BELONGS TO HIGH RECOVERY GROUP
IF $f(s) < -.1$ THEN THE SAMPLE BELONGS TO LOW RECOVERY GROUP

For the region between -.1 and .3, a study of errors was performed according to the experimental results available in the trainning set, and the conclusion was drawn that the point f(s) = -.1 leads to a minimum sum of misclassification type I and II errors (33%, as calculated in TABLE 3).

TABLE 3 - Assessment of discriminant power by the proportion of misclassified samples.

DISCRIMINANT LIMIT	TYPE I ERROR (%)	TYPE II ERROR (%)	SUM OF ERRORS (%)
-.1	0	33	33
0	6	30	36
.1	29	27	56
.2	47	15	62
.3	59	0	59

So, assuming a risk of 33% of misclassification, a simple and practical decision rule may be derived:

Any sample, taken at a depth greater than 36.7m, has 2 chances in 3 to provide a "high recovery" product if one of the following conditions is met:

Reduced	⇒ f(s) > .22
U_3O_8 grade > .2%	⇒ f(s) > .2
Presence of expressed mineralization	⇒ f(s) > .07
Shale	⇒ f(s) > -.09

CONCLUSIONS

In order to take into account, *jointly*, qualitative and quantitative information in a data analysis problem, numeric values must be recoded into ordinal intervals. After the recoding procedure, correspondence analysis satisfies its basic theoretical assumptions and can be correctly applied to describe the relationship pattern among variables and to discriminate groups of samples.

In this case study, it is shown how to articulate a set of qualitative data analysis techniques in order to give a proper insight into a decision problem in the late stages of an exploration program for Uranium.

REFERENCES

ATKIN, R. (1974) - 'An approach to structure in architectural and urban design'. Environment and planning B, 1974, vol. 1, pp. 51-67.

BENZECRI, J.P. (1973) - L'analyse des Données, Dunod, Paris, vol I and II

BOTBOL, J. (1971) - 'An application of characteristic analysis to mineral exploration', CIM Special vol 12, pp. 92-99.

GEOFFROY, J. and WIGNALL, T. (1972) - 'A statistical study of geological characteristics of porphyry-copper-molybdenum deposits in the Cordilleran Belt - Application to the rating of porphyry prospects'. Ec. Geol. vol. 67, 1972, pp. 656-668.

GREENACRE, M. (1984) - Theory and applications of correspondence analysis, Academic Press.

GRIFFITHS, J. (1983) - 'Geological Similarity by Q-Analysis', Math. Geol vol 15, nº 1, pp 85-108.

PEREIRA, H.G., SOUSA, A.J., FIUZA, A., COSTA, L. (1984) - 'Improving of interpretation of uranium leaching tests by inclusion of geological information', in 18th APCOM, IMM, pp 441-447.