International Journal of Science and Engineering Applications
Volume 11-Issue 06, 72 - 75, 2022, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1106.1001

DCA Method in Mineral Exploration, Example: Predict the Location of New Samples

Hamed Nazerian
University of Catania
Italy

Bahareh Hedayat
Shahrood University of Technology
Iran

Behnam Kakavand

Abstract: As the name implies, Discriminant Correspondence Analysis (DCA) is a format of Discriminant analysis (DA) and correspondence analysis (CA). Like differential analysis, the goal is to categorize observations into predefined groups and, like Correspondence analysis, to use nominal variables. The main idea of the DCA is to represent each set of observations and perform a simple Correspondence analysis on the variables (a matrix). The main observations are complementary elements, and each observation is attributed to the closest group. A comparison between the predictions and predictions of classification leads to evaluating the Discriminant correspondence. This information can be used for a similar case to classify new observations, and the validation of estimates can also be examined using cross-validation techniques such as Jack Knife or Bootstrap. For example, samples were taken from different regions. After scoring the parameters and training in this analysis, a new sample was entered, and a group (region) was determined compared to the previous samples.

Keywords: Discriminant Correspondence Analysis, DCA, CA, DA, Mineral Sampling.

1. INTRODUCTION

In mining engineering, many statistical methods have been adapted from other disciplines. These methods can be used in mining engineering with a simple idea [1-15]. As the name of this method suggests, discriminant analysis is a format of Discriminant analysis and correspondence analysis. Like differential analysis, the goal is to categorize observations into predefined groups and, like Correspondence analysis, to use nominal variables [16-18].

The main idea of the DCA is to represent each set of observations and perform a simple Correspondence analysis on the variables (a matrix). The main observations are complementary elements, and each observation is attributed to the closest group. A comparison between the predictions and predictions of classification leads to evaluating the Discriminant correspondence. This information can be used for a similar case to classify new observations, and the validation of estimates can also be examined using cross-validation techniques such as Jack Knife or Bootstrap [19-35].

2. DISCUSSION

For example, the type of mineral usually depends on its origin. For example, we sampled 12 samples from 3 different regions (4 samples from each region) and asked one person (unaware of their origin) to rate the samples on a 5-point scale.

The scores were then converted to binary code in an index matrix (which can be used in Correspondence analysis). For example, a score of 2 became the binary value of "0 1 0". The data are given in Table 1.

2.1 Subject description

We have K groups; in each group, Ik is observed, and I represent the sum of the observations. To simplify, we assume that our observations are rows, and our variables are columns that contain the J variable, which we call the J × I matrix. The index matrix of K × I is called Y. The value of 1 indicates the belonging of the row that represents the group, and 0 indicates the non-belonging to the group in the columns. The J × K matrix, also named N, is the group matrix, which holds the number of variables for each group. For example, we found that:

$$\mathbf{N} = \mathbf{Y}^{\mathrm{T}}\mathbf{X} = \begin{bmatrix} 3 & 1 & 0 & 0 & 1 & 3 & 0 & 2 & 2 & 2 & 2 & 0 & 1 & 1 & 1 & 1 \\ 1 & 2 & 1 & 1 & 2 & 1 & 2 & 1 & 1 & 0 & 1 & 3 & 1 & 1 & 1 & 1 \\ 0 & 1 & 3 & 3 & 1 & 0 & 1 & 1 & 2 & 3 & 1 & 0 & 1 & 1 & 1 & 1 \end{bmatrix} \quad (1)$$

2.2 Method

Doing CA in the group N matrix produces two points, one for groups (set F) and one for variables (set G). In general, these factor scores have general values so that their variance is equal to the specific values attributed to the factors. The above table is named N, and the first step in the analysis is to calculate the probability matrix Z = N-1N.

Table 1: Information of 3 regions: 12 samples from 3 different regions in 5 descriptive points. A value of 1 indicates that the sample is variable. The W sample is unknown as a complementary observation.

		1			2			3			4			5			
Sample		1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	4
1		1	0	0	0	0	1	0	1	0	1	0	0	1	0	0	0
2	Area 1	0	1	0	0	0	1	0	0	1	0	1	0	0	0	1	0
3		1	0	0	0	1	0	0	1	0	1	0	0	0	1	0	0
4		1	0	0	0	0	1	0	0	1	0	1	0	0	0	0	1
Σ	Sum	3	1	0	0	1	3	0	2	2	2	2	0	1	1	1	1
5		1	0	0	0	1	0	1	0	0	0	0	1	0	0	1	0
6	Area 2	0	1	0	1	0	0	1	0	0	0	0	1	0	1	0	0
7		0	0	1	0	1	0	0	1	0	0	1	0	1	0	0	0
8		0	1	0	0	0	1	0	0	1	0	0	1	0	0	0	1
Σ	Sum	1	2	1	1	2	1	2	1	1	0	1	3	1	1	1	1
9		0	0	1	1	0	0	0	0	1	1	0	0	1	0	0	0
10		0	1	0	1	0	0	0	0	1	1	0	0	0	1	0	0
11	Area 3	0	0	1	0	1	0	0	1	0	0	1	0	0	0	0	1
12		0	0	1	1	0	0	1	0	0	1	0	0	0	0	1	0
Σ	Sum	0	1	3	3	1	0	1	1	2	3	1	0	1	1	1	1
W?	?	1	0	0	0	0	1	0	1	0	0	1	0	1	0	0	0

r is the sum of the rows Z and the parameter C is the sum of the placed columns, Dc and Dr are the diagonals C and r.

International Journal of Science and Engineering Applications
Volume 11-Issue 06, 72 - 75, 2022, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1106.1001

The scoring factor is obtained by breaking down the following single value:

$$D_r^{-1/2}(Z-rc^T)D_c^{-1/2}=P\Delta Q^T \quad (2)$$

Delta (Δ) is a diagonal matrix with single values, and 2Δ = Λ is a matrix of unique values.

The row and (respectively) column of the score invoice is obtained from the following formula:

$$G=D_c^{-1/2}Q\Delta \quad \& \quad F=D_r^{-1/2}P\Delta \quad (3)$$

The square of the distance from the row and column to the relevant body centers is calculated as follows:

$$d_c=\mathrm{diag}\{GG^T\} \quad \& \quad d_r=\mathrm{diag}\{FF^T\} \quad (4)$$

The square cosine between row i and factor l and column j and factor l is created as follows:

$$o_{i,\ell}=\frac{f_{i,\ell}^2}{d_{r,i}^2} \quad , \quad o_{j,\ell}=\frac{g_{j,\ell}^2}{d_{c,j}^2} \quad (5)$$

With $d_{r,I}{}^2$ and $d_{c,j}{}^2$the element i from dr and the element j from d_c are created, respectively. The cosine of squares helps determine the location of important factors to observe. The share of row i to factor l and column j to factor l is obtained as follows:

$$t_{i,\ell}=\frac{f_{i,\ell}^2}{\lambda_\ell} \quad \& \quad t_{j,\ell}=\frac{g_{j,\ell}^2}{\lambda_\ell} \quad (6)$$

Coexistence (or contribution) helps identify the location of essential observations for a factor.

Complementary or rational elements that can represent factors are called transfer formulas.

Specifically, $i^T{}_{sup}$ as an illustrative row and j_{sup} as an illustrative column. The $f_{sup}$and g_{sup} specifications are obtained through the following:

$$\mathbf{f}_{sup}=\left(\mathbf{i}_{sup}^T\mathbf{1}\right)^{-1}\mathbf{i}_{sup}^T\mathbf{G}\boldsymbol{\Delta}^{-1} \quad \& \quad \mathbf{g}_{sup}=\left(\mathbf{j}_{sup}^T\mathbf{1}\right)^{-1}\mathbf{j}_{sup}^T\mathbf{F}\boldsymbol{\Delta}^{-1} \quad (7)$$

Note that the scalar rules $^{-1})i^T{}_{sup}$ (and $^{-1})j^T{}_{sup}1$ (are used to ensure the sum of the elements i_{sup} or j_{sup} are equal to one. If this condition is met, there is no need for this law.

After the group's analyses were performed, the main observations were stored as complementary elements and factor scores in a matrix called F_{sup}. To calculate these scores, the first-row profile matrix is calculated:

$$R=(\mathrm{diag}\{X1\})^{-1}X \quad (8)$$

And apply Equation 7 below:

$$F_{sup}=RG\Delta^{-1} \quad (9)$$

The Euclidean distance between the observations and the groups calculated from the score factor is equal to the distance ϰ2 between their row profiles.

The K × I distance matrix between observations and groups is calculated as follows:

$$D=S_{sup}1^T+1S^T-2F_{sup}F^T \quad (10)$$

with

$$S=\mathrm{diag}\{FF^T\} \quad \text{و} \quad S_{sup}=\mathrm{diag}\{F_{sup}F^T{}_{sup}\} \quad (11)$$

Each observation will be assigned to the nearest group.

2.2.1. Model evaluation

The resolution quality can be considered a model with fixed effects or a random model. For the fixed effect model, the correct classifications are compared with the answers obtained from Equation 10. The fixed-effect model evaluates the classification quality in the sample used to construct the model, and the stochastic model evaluates the classification quality based on the new observations. Typically, this step is done using cross-validation techniques such as Jack Knife or Bootstrap.

Table 2: Factor scores, cosine squares, and auxiliary variables (set J). Negative scores of the correspondence part are also shown in italics.

			1			2			3			4			5			
			1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	4
Axis	λ	%	Factor Scores															
1	.251	55	.93	−.05	−.88	−.88	−.05	.93	−.51	.33	.04	−.14	.33	−.20	0	0	0	0
2	.201	44	−.04	.35	−.31	−.31	.35	−.04	.64	−.13	−.28	−.74	−.13	1.40	0	0	0	0
Axis			Squared Cosines															
1			.998	.021	.892	.892	.021	.998	.384	.864	.021	.035	.864	.021	0	0	0	0
2			.002	.979	.108	.108	.979	.002	.616	.137	.979	.965	.137	.979	0	0	0	0
Axis			Contributions															
1			.231	.001	.207	.207	.001	.231	.051	.029	.001	.007	.029	.008	0	0	0	0
2			.0006	.0405	.0313	.0313	.0405	.0006	.1019	.0056	.0324	.2235	.0056	.4860	0	0	0	0

Table 3: Invoice scores, cosine squares, region alignment, and complement row for unknown sample region (W). Negative scores of the correspondence part are also shown in italics.

				Area 1 Samples					Area 2 Samples					Area 3 Samples					
Axis	λ	%	Area 1	1	2	3	4	Area 2	1	2	3	4	Area 3	1	2	3	4	W?
									Factor Scores									
1	.251	55	0.66	0.82	0.50	0.43	0.89	−0.10	0.07	−0.66	−0.11	0.29	−0.56	−0.74	−0.41	−0.11	−0.96	1.01
2	.201	44	−0.23	−0.42	−0.05	−0.25	−0.22	0.63	1.05	0.93	−0.10	0.64	−0.39	−0.73	−0.43	−0.10	−0.32	−0.15
									Squared Cosines									
1			.89	.79	.99	.75	.94	.03	.00	.33	.56	.17	.67	.51	.47	.56	.90	.98
2			.11	.21	.01	.25	.06	.97	1.00	.67	.44	.83	.33	.49	.53	.44	.10	.02
									Contributions									
1			.58	.	.	.	.	.01	.	.	.	.	.41	.	.	.	.	.
2			.09	.	.	.	.	.65	.	.	.	.	.26	.	.	.	.	.

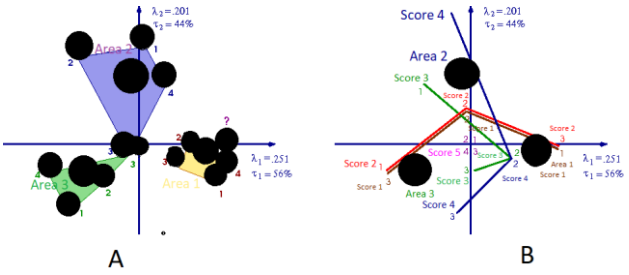

Figure 1: DCA analysis shown in two dimensions. (A) Set I: Rows (samples), sample projected as complementary elements, sample? The sample is unknown. (B) Set J: Column (number of points). Sample categories are also included for ease of interpretation. Both shapes have the same scale (some points shifted slightly for readability). (Tables 2 and 3).

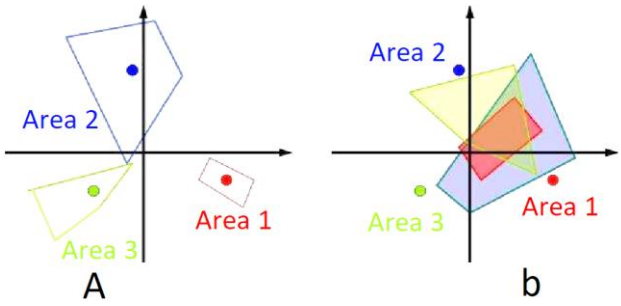

Figure 2: DCA analysis. 2D display. (A) Fixed effect model. Three areas and indicators for samples. (B) Stochastic effect model. Samples of the chipped jack behind the fixed effect solution are shown. The index shows that the stochastic effect classification is more varied and transferred.

3. CONCLUSION

Place Tables 2 and 3 show the analysis results, and Figure 1 shows them. The quality of the fixed effect of the model was evaluated with the following configuration matrix:

$$\begin{bmatrix} 4 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 1 & 4 \end{bmatrix}$$

In this matrix, rows are assigned to predicted groups, and columns are natural groups. For example, out of 5 points assigned to Zone 3 (Group 3), one is from Zone 2 (Group 2), and four samples are from Zone 3. The overall quality can be calculated from the diameter of the matrix. Here we find that (4 + 3 + 4) 11 samples of R12 are correctly classified.

Jack Knife's estimation method was used to evaluate the generalization capacity of the analysis to the new sample (for example, this relates to random effect analysis).

Each sample was placed outside the sample set, a DCA was performed on the remaining samples (11), and the sample was assigned to the nearest group. This method gives us the following configuration matrix:

$$\begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}$$

As expected, the performance of the random effect model was lower than the fixed-effect model, and only 6 (2 + 2 + 2) samples out of 12 samples were correctly identified. The difference between the fixed effect model and the stochastic effect is shown in Figure 2 when the data is jacked up (using a multidimensional metric scaling). The quality of the model can be evaluated by drawing the polygon of each category. For the fixed effect model, the centers of gravity are polygons of the categories, indicating that DCA is a least-squares estimation technique. The random model did not function properly due to the more significant variance (it has larger covered polygon areas) and rotated around the area (the polygon was not based on the center of gradation but the center of gravity). Codes of this method were also made available for further research.

4. REFERENCES

1. Khayer, K., et al., *Permeability Estimation from Stoneley Waves in Carbonate Reservoirs.* Geological Bulletin of Turkey, 2022. **65**: p. 42.

2. Khosravi, V., et al., *Hybrid Fuzzy-Analytic Hierarchy Process (AHP) Model for Porphyry Copper Prospecting in Simorgh Area, Eastern Lut Block of Iran.* Mining, 2022. **2**(1): p. 1-12.

3. Nazerian, H., et al., *Design of an Artificial Neural Network (BPNN) to Predict the Content of Silicon Oxide (SiO2) based on the Values of the Rock Main Oxides: Glass Factory Feed Case Study.* International Journal of Science and Engineering Applications (IJSEA), 2022. **2**(11): p. 41-44.

4. Shirazy, A., et al., *K-Means Clustering and General Regression Neural Network Methods for Copper Mineralization probability in Chahar-Farsakh, Iran.* Türkiye Jeoloji Bülteni, 2022. **65**(1): p. 79-92.

5. Hedayat, B., et al., *Feasibility of Simultaneous Application of Fuzzy Neural Network and TOPSIS Integrated Method in Potential Mapping of Lead and Zinc Mineralization in Isfahan-Khomein Metallogeny Zone.* Open Journal of Geology, 2022. **12**(3): p. 215-233.

6. Ahmadi, M.E., et al., *Assessment of the Influence of Sulfuric Acid/Hydrogen Peroxide Mixture on Organic Sulfur Reduction of High Sulfur Coals and Their Chemical Composition.* Open Journal of Geology, 2022. **12**(3): p. 199-214.

7. Aali, A.A., et al., *Geophysical Study to Identify Iron Mineralization Anomalies Using Terrestrial Magnetometry in the Chak-Chak Exploration Area, Iran.* Türkiye Jeoloji Bülteni, 2022. **65**(2): p. 159-170.

8. Nazerian, H., et al., *Design of an Artificial Neural Network (BPNN) to Predict the Content of Silicon Oxide (SiO2) based on the Values of the Rock Main Oxides: Glass Factory Feed Case Study.* International Journal of Science and Engineering Applications (IJSEA), 2022. **2**: p. 41-44.

9. Khayer, K., et al., *Permeability Estimation from Stoneley Waves in Carbonate Reservoirs.* Türkiye Jeoloji Bülteni, 2022. **65**(1): p. 1-8.

10. Adel, S., Z. Mansour, and H. Ardeshir, *Geochemical behavior investigation based on k-means and artificial neural network prediction for titanium and zinc, Kivi region, Iran.* Известия Томского политехнического университета. Инжиниринг георесурсов, 2021. **332**(3): p. 113-125.

11. Shirazy, A., A. Shirazi, and H. Nazerian, *Application of Remote Sensing in Earth Sciences–A Review.* International Journal of Science and Engineering Applications, 2021. **10**(5): p. 45-51.

12. Khayer, K., et al., *Determination of Archie's Tortuosity Factor from Stoneley Waves in Carbonate Reservoirs.* International Journal of Science and Engineering Applications (IJSEA), 2021. **10**: p. 107-110.

13. Shirazy, A., et al., *Geophysical study: Estimation of deposit depth using gravimetric data and Euler method (Jalalabad iron mine, kerman province of IRAN).* Open Journal of Geology, 2021. **11**(8): p. 340-355.

14. Shirazy, A., et al., *Investigation of Geochemical Sections in Exploratory Boreholes of Mesgaran Copper Deposit in Iran.* International Journal for Research in Applied Science and Engineering Technology (IJRASET), 2021. **9**(8): p. 2364-2368.

15. Nazerian, H., et al., *Predict the Amount of Cu Using the Four Ca, Al, P, S Elements by Multiple Linear Regression Method.* International Journal for Research in Applied Science and Engineering Technology (IJRASET), 2021. **9**: p. 1088-1092.

16. Shirazi, A., A. Hezarkhani, and A.B. Pour, *Fusion of Lineament Factor (LF) Map Analysis and Multifractal Technique for Massive Sulfide* Copper *Exploration: The Sahlabad Area, East Iran.* Minerals, 2022. **12**(5): p. 549.

17. Williams, L.J., et al., *A tutorial on multiblock discriminant correspondence analysis (MUDICA): a new method for analyzing discourse data from clinical populations.* 2010.

18. Abdi, H., *Discriminant correspondence analysis*. 2007, Sage Thousand Oaks, CA. p. 1-10.

19. Shirazy, A., et al., *Investigation of Magneto-/Radio-Metric Behavior in Order to Identify an Estimator Model Using K-Means Clustering and Artificial Neural Network (ANN)(Iron Ore Deposit, Yazd, IRAN).* Minerals, 2021. **11**(12): p. 1304.

20. Shirazi, A., et al., *Geochemical and Behavioral Modeling of Phosphorus and Sulfur as Deleterious Elements of Iron Ore to Be Used in Geometallurgical Studies, Sheytoor Iron Ore, Iran.* Open Journal of Geology, 2021. **11**(11): p. 596-620.

21. Shirazi, A. and A. Shirazy, *Introducing Geotourism Attractions in Toroud Village, Semnan Province, IRAN*. International Journal of Science and Engineering Applications, 2020. **9**(16): p. 79-86.

22. Khakmardan, S., et al., *Evaluation of Chromite Recovery from Shaking Table Tailings by Magnetic Separation Method.* Open Journal of Geology, 2020. **10**(12): p. 1153-1163.

23. Doodran, R.J., et al., *Minimalization of Ash from Iranian Gilsonite by Froth Flotation.* Journal of Minerals and Materials Characterization and Engineering, 2020. **9**(1): p. 1-13.

24. Shirazy, A., et al., *Cementation exponent estimate in carbonate reservoirs: A new method*. Global Journal of Computer Sciences: Theory and Research, 2020. **10**(2): p. 66-72.

25. Shirazy, A., A. Shirazi, and A. Hezarkhani, *Behavioral Analysis of Geochemical Elements in Mineral Exploration:- Methodology and Case Study*. 2020: LAP LAMBERT Academic Publishing.

26. Shirazy, A., et al., *Geochemical and geostatistical studies for estimating gold grade in tarq prospect area by k-means clustering method*. Open Journal of Geology, 2019. **9**(6): p. 306-326.

27. Shirazi, A., A. Hezarkhani, and A. Shirazy, *Exploration Geochemistry Data-Application for Cu Anomaly Separation Based On Classical and Modern Statistical Methods in South Khorasan, Iran.* International Journal of Science and Engineering Applications (IJSEA), 2018. **7**(4): p. 39-44.

28. Shirazi, A., A. Hezarkhani, and A. Shirazy, *Remote Sensing Studies for Mapping of Iron Oxide Regions, South of Kerman, IRAN*. International Journal of Science and Engineering Applications (IJSEA), 2018. **7**(4): p. 45-51.

29. Shirazi, A., et al., *Geostatistics studies and geochemical modeling based on core data, sheytoor iron deposit, Iran.* Journal of Geological Resource and Engineering, 2018. **6**: p. 124-133.

30. Alahgholi, S., A. Shirazy, and A. Shirazi, *Geostatistical studies and anomalous elements detection, Bardaskan Area, Iran.* Open Journal of Geology, 2018. **8**(7): p. 697-710.

31. Khakmardan, S., et al., *Copper oxide ore leaching ability and cementation behavior, mesgaran deposit in Iran.* Open Journal of Geology, 2018. **8**(09): p. 841.

32. Shirazi, A., A. Shirazy, and J. Karami, *Remote sensing to identify copper alterations and promising regions, Sarbishe, South Khorasan, Iran.* International Journal of Geology and Earth Sciences, 2018. **4**(2): p. 36-52.

33. Shirazy, A., et al., *Exploratory Remote Sensing Studies to Determine the Mineralization Zones around the Zarshuran Gold Mine.* International Journal of Science and Engineering Applications, 2018. **7**(9): p. 274-279.

34. Shirazi, A., et al., *Introducing a software for innovative neuro-fuzzy clustering method named NFCMR.* Global Journal of Computer Sciences: theory and research, 2018. **8**(2): p. 62-69.

35. Shirazy, A., A. Shirazi, and A. Hezarkhani, *Predicting gold grade in Tarq 1: 100000 geochemical map using the behavior of gold, Arsenic and Antimony by K-means method.* Journal of Mineral Resources Engineering, 2018. **2**(4): p. 11-23.