Creative Economy
2025, VOL. 9, NO. 2, 179-182
DOI: 10.47297/wspceWSP2516-251931.20250902

Based on the application of the correspondence analysis method in archaeology

Manli Zhang, Guangli Xu, Rui Chen, Zhenhuan Kang

Taishan University, Taian 271000

ABSTRACT

Seeing the world with two eyes, the world in our eyes determines our understanding, thinking and mastery of the world. Sometimes, the complicated relationship of the world should not be concluded only by feeling, but also combined with new scientific and technological means to help us better clarify the relationship and understand the connotation. The corresponding analysis method is a multivariate statistical method developed based on the basis of factor analysis. It can intuitively represent the relationship between samples and variables, and intuitively explain the correlation relationship without the complex mathematical operation process such as factor selection and rotation of factor axis in the factor analysis. With the help of an archaeological data, the application of the analysis method can be developed from the line outline, column outline and corresponding analysis map.

KEYWORDS

corresponding analysis; archaeological data; line outline; column outline

1 Introduction

The corresponding analysis[1] is also called the corresponding analysis was proposed by the French Benzecri in 1970, initially popular in France and Japan, and then popular to the United States. The main body of the corresponding analysis is the contingency table obtained from the data collection, whose ultimate purpose is to represent the relationship between rows and columns in a low-dimensional graph. The biggest characteristic is that it can present many samples and many variables on an icon at the same time. Corresponding analysis is widely used in the fields of geological research, archaeological analysis, market segmentation and product positioning. It can directly show the connection between the data that are not directly visible to the naked eye with visually acceptable positioning maps.

2 The corresponding analysis method

The body of the corresponding analysis is the contingency n_{ij}table $i$as shown $j$in Table 1: in the table represents the frequency of the row and column categories. The corresponding frequency matrix can be calculated $p_{ij}=\frac{n_{ij}}{n}$,$i=1,2\cdots,p$,$j=1,2\cdots,q$;$p_{i\cdot}=\sum_{j=1}^{q}p_{ij}=\frac{n_{i\cdot}}{n}$;$p_{\cdot j}=\sum_{i=1}^{p}p_{ij}=\frac{n_{j}}{n}$;in Table 1, using the following formula: then the corresponding matrix is the frequency matrix as shown in Table 2.

Table 1 p x q Contingency Table

row \ column	1	2	⋯	q	Subtotal
1	n_{11}	n_{12}	⋯	n_{1q}	$n_{1\cdot}$
2	n_{21}	n_{22}	⋯	n_{2q}	$n_{2\cdot}$
⋮	⋮	⋮		⋮	⋮
p	n_{p1}	n_{p2}	⋯	n_{pq}	$n_{p\cdot}$
Subtotal	$n_{\cdot 1}$	$n_{\cdot 2}$	⋯	$n_{\cdot q}$	n

Table 2 Correspondence Matrix

row \ column	1	2	⋯	q	Subtotal
1	p_{11}	p_{12}	⋯	p_{1q}	$p_{1\cdot}$
2	p_{21}	p_{22}	⋯	p_{2q}	$p_{2\cdot}$
⋮	⋮	⋮		⋮	⋮
p	p_{p1}	p_{p2}	⋯	p_{pq}	$p_{p\cdot}$
Subtotal	$p_{\cdot 1}$	$p_{\cdot 2}$	⋯	$p_{\cdot q}$	1

Using the data in the corresponding matrix we can use the formula to calculate $\boldsymbol{r}_i=\left(\frac{p_{i1}}{p_{i\cdot}},\frac{p_{i2}}{p_{i\cdot}},\cdots,\frac{p_{iq}}{p_{i\cdot}}\right)=\left(\frac{n_{i1}}{n_{i\cdot}},\frac{n_{i2}}{n_{i\cdot}},\cdots,\frac{n_{iq}}{n_{i\cdot}}\right)$; $\boldsymbol{r}_i 1=1$, $i=1,2,\cdots,p$the row profile of any row; using the formula

$$\boldsymbol{c}_j=\left(\frac{p_{1j}}{p_{\cdot j}},\frac{p_{2j}}{p_{\cdot j}},\cdots,\frac{p_{pj}}{p_{\cdot j}}\right)'=\left(\frac{n_{1j}}{n_{\cdot j}},\frac{n_{2j}}{n_{\cdot j}},\cdots,\frac{n_{pj}}{n_{\cdot j}}\right)';1'\boldsymbol{c}_j=1,\ j=1,2,\cdots,q$$

Calculate the column outline of any column, here the line outline and the column outline are actually a conditional distribution, namely the frequency distribution, from the row profile of each row and the column outline of each column can get the corresponding row profile matrix and the column profile matrix:

$$\boldsymbol{R}=\boldsymbol{D}_r^{-1}\boldsymbol{P}=\begin{pmatrix}\boldsymbol{r}_1'\\ \boldsymbol{r}_2'\\ \mathrm{M}\\ \boldsymbol{r}_p'\end{pmatrix}=\begin{pmatrix}\frac{p_{11}}{p_{1\cdot}} & \frac{p_{12}}{p_{1\cdot}} & \mathrm{L} & \frac{p_{1q}}{p_{1\cdot}}\\ \frac{p_{21}}{p_{2\cdot}} & \frac{p_{22}}{p_{2\cdot}} & \mathrm{L} & \frac{p_{2q}}{p_{2\cdot}}\\ \mathrm{M} & \mathrm{M} & & \mathrm{M}\\ \frac{p_{p1}}{p_{p\cdot}} & \frac{p_{p2}}{p_{p\cdot}} & \mathrm{L} & \frac{p_{pq}}{p_{p\cdot}}\end{pmatrix}\boldsymbol{C}=\boldsymbol{P}\boldsymbol{D}_c^{-1}=(\boldsymbol{c}_1,\boldsymbol{c}_2,\mathrm{L},\boldsymbol{c}_q)=\begin{pmatrix}\frac{p_{11}}{p_1} & \frac{p_{12}}{p_2} & \mathrm{L} & \frac{p_{1q}}{p_q}\\ \frac{p_{21}}{p_1} & \frac{p_{22}}{p_2} & \mathrm{L} & \frac{p_{2q}}{p_q}\\ \mathrm{M} & \mathrm{M} & & \mathrm{M}\\ \frac{p_{p1}}{p_1} & \frac{p_{p2}}{p_2} & \mathrm{L} & \frac{p_{pq}}{p_q}\end{pmatrix}$$

Line profile matrix and column profile matrix can be illustrated in R language to show the corresponding row outline, row edge frequency and the data of column outline and column edge frequency.

In the corresponding analysis of the independence of the test is necessary, we usually assume that the line contains variables and column contains variables are related, if there is no relationship we do not need to corresponding χ^2analysis, so the independence of the $\chi^2=n\sum_{i=1}^{p}\sum_{j=1}^{q}\frac{\left(p_{ij}-p_{i\cdot}p_{j}\right)^2}{p_{i\cdot}p_{\cdot j}}$line $\chi^2\geq\chi_\alpha^2\left[(p-1)(q-1)\right]$and the column test is particularly important, we use a distribution of test statistics, specific statistics for, then reject the original hypothesis of independence.

Note that the value $\chi^2$is related. Here the larger *n*the value, the more *n*information the sample contains, the more likely it is to deviate from the original hypothesisχ^2. Therefore*n*, we often remove the above distribution and define a new variable

I$\frac{\chi^2}{n}=\sum_{i=1}^{p}\sum_{j=1}^{q}\frac{\left(p_{ij}-p_{i\cdot}p_{j}\right)^2}{p_{i\cdot}p_{j}}$t is called the total inertia, which is used to measure the correlation between rows and column variables. Total inertia $\sum_{i=1}^{p}p_{i\cdot}\sum_{j=1}^{q}\frac{\left(p_{ij}/p_{i\cdot}-p_{j}\right)^2}{p_j}=\sum_{i=1}^{p}p_{i\cdot}\left(\boldsymbol{r}_i-\boldsymbol{c}\right)'\boldsymbol{D}_c^{-1}\left(\boldsymbol{r}_i-\boldsymbol{c}\right)$can also be expressed in the form of row profiles and column profiles that we mentioned earlier. When it is expressed in a line outline;

When it is expressed with a column $\sum_{j=1}^{q}p_j\sum_{i=1}^{p}\frac{\left(p_{ij}/p_{\cdot j}-p_{i\cdot}\right)^2}{p_{i\cdot}}=\sum_{j=1}^{q}p_j\left(\boldsymbol{c}_j-\boldsymbol{r}\right)'\boldsymbol{D}_r^{-1}\left(\boldsymbol{c}_j-\boldsymbol{r}\right)$outline$\left(\boldsymbol{r}_i-\boldsymbol{c}\right)'\boldsymbol{D}_c^{-1}\left(\boldsymbol{r}_i-\boldsymbol{c}\right)=\sum_{j=1}^{q}\frac{\left(p_{ij}/p_{i\cdot}-p_{j}\right)^2}{p_j}$*ir*_*i*C; where it is *j*called the *c*_*j*chir-square distance $\left(\boldsymbol{c}_j-\boldsymbol{r}\right)'\boldsymbol{D}_r^{-1}\left(\boldsymbol{c}_j-\boldsymbol{r}\right)=\sum_{i=1}^{p}\frac{\left(p_{ij}/p_{j}-p_{i\cdot}\right)^2}{p_{i\cdot}}$of the row contour to the center of the row contour, it is the weighted square Eucoid distance; also the chi-square distance from the row contour to the center of the column contour. The total inertia expressed by the above formula can be regarded as the weighted average of the chi-square distance from the row contour to its center, but also as the weighted average of the chi-square distance from the column contour to its center, that is, it is the total variation between the row contour and the column contour. The greater the difference between the profiles of the ranks, the stronger the correlation between the ranks, and vice versa.

Previously mentioned centralized line outline and column outline can be marked in the low-dimensional coordinate system, the amount of information on the coordinates is generally different, usually our main information is focused on the front dimension, like select the principal component principle, the principal components of the information is different, the corresponding relationship is a low dimensional coordinates, the most common is to two dimensions, in the form of plane map show this is the birth of the corresponding analysis diagram. In the corresponding analysis diagram, the farther away from the origin, and the closer to the origin, the actual frequency of the category combination is higher

than the expected frequency in the case of two variables.

3 Example application of the corresponding analysis method

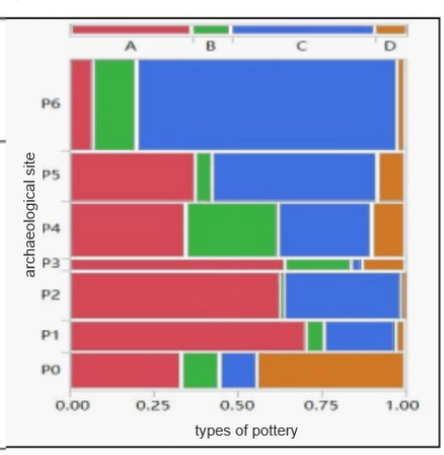
Table 3 Archaeological Data

archaeological site \ types of pottery	A	B	C	D	Subtotal
P0	30	10	10	39	89
P1	53	4	16	2	75
P2	73	1	41	1	116
P3	20	6	1	4	31
P4	46	36	37	13	132
P5	45	6	59	10	120
P6	16	28	169	5	218
Subtotal	283	91	333	74	781

We first calculate the travel profile data and the line contour graph as shown in Table 4 and Figure 1; the mosaic graph shows the line contour and the row edge frequency, the width of the four small squares: A, B, C, D shows the row contour and the small squares from P0 to P67 of the archaeological sites. From Figure 1, we can see that the profiles of P1 and P2 are similar; P4 and P5 are similar; but the profiles of P0 and P6 are the least similar, which can be judged directly by the line outline.

Table 4 Line profile data

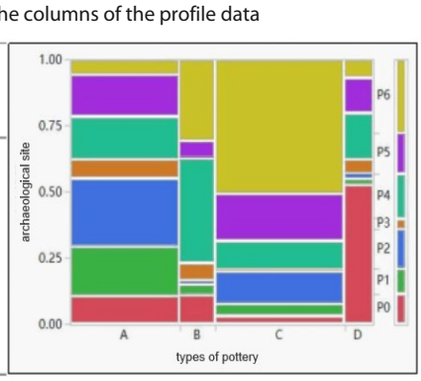
archaeological site \ types of pottery	A	B	C	D
P0	0.337	0.112	0.112	0.438
P1	0.707	0.053	0.213	0.027
P2	0.629	0.009	0.353	0.009
P3	0.645	0.194	0.032	0.129
P4	0.348	0.273	0.280	0.099
P5	0.375	0.050	0.492	0.083
P6	0.073	0.128	0.775	0.023



The same calculated column outline data and column outline diagram are shown in Table 5 and Figure 2; the mosaic diagram shows the calculated data of the column outline and column edge frequency; Figure 2 shows the height of the small square of 7 archaeological sites from P0 to P6, and the width of the four small squares A, B, C, and D shows the column edge frequency. From Figure 2, we can also see that the profiles of the four ceramics are not very similar, which is also an intuitive display of the column outline diagram.

Table 5 shows the columns of the profile data

archaeological site \ types of pottery	A	B	C	D
P0	0.106	0.110	0.030	0.527
P1	0.187	0.044	0.048	0.027
P2	0.258	0.011	0.123	0.014
P3	0.071	0.066	0.003	0.054
P4	0.163	0.396	0.111	0.176
P5	0.159	0.066	0.177	0.135
P6	0.057	0.308	0.507	0.068



We calculated the main inertia and the contribution rate as shown in Table 6. From Table 6, we conclude that the value of the total inertia is large, that is, the difference between the rows and the column profiles is large, which also reveals the strong correlation between the rows and columns.

Table 6 Main inertia and contribution rate values

dimension	1	2	3	total value
principal moments	0.2836	0.1701	0.0588	0.5125
Contribution rate	0.5534	0.3319	0.1147	1.000
Cumulative contribution rate	0.5534	0.8853	1.0000	

Sometimes it is not clear to observe only through the row outline and column outline, but it is more intuitive to analyze through the corresponding analysis chart. We show the corresponding analysis chart through R language in Figure 1.

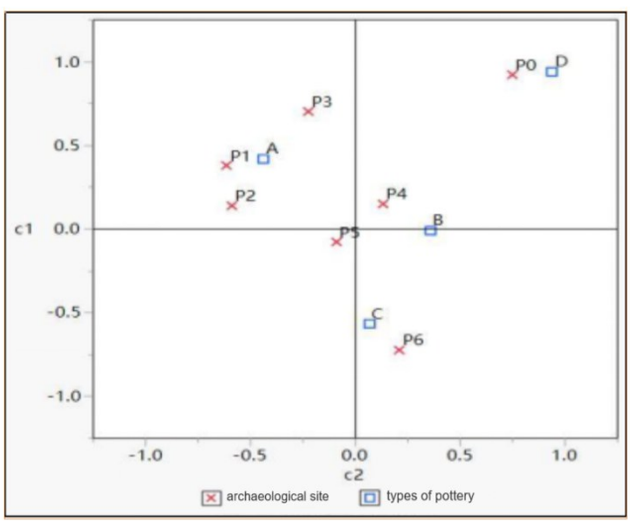


Figure 1 Corresponds to the analysis

We judge the degree of similarity by the secondary distance of the points in the figure. From Figure 3 we can see that the distance between P1 and P2 is relatively close, which means that their contours are relatively similar; similarly, P4 and P5 are relatively close, so their contours are relatively similar; and P0 and P6 are the most distant, thus the most similar. According to the figures, the list points between each other of the four pottery sites are not very close, thus indicating that the profiles of the four pottery sites are not very similar. However, we can also find that C and B are relatively less different; but C and D have the farthest distance and the largest difference. Further, we can also find that P0 and D are very close together, indicating that they have a close connection, that is, D pottery was unearthed in large numbers in P0 archaeological sites. The same relationship between P0 and D, P1 and A, P2 and A, P3 and A, P4 and B, and P6 and C. From this perspective, the benefits of the corresponding analysis graph are intuitive.

4 Summary

This paper mainly explains the principle and the application of the corresponding analysis. This paper expounds the basic principle of the corresponding analysis in the archaeology.

About the Author

Manli Zhang (1985.09--), Mujer, Nacionalidad:Han, Tai'an* City, Shandong Province.Estudiantes de posgrado,Research direction:Applied statistics.

References

[1] Wang Xuemin. Applied a multivariate statistical analysis [M]. Shanghai: Shanghai University of Finance and Economics Press, 2021.

[2] He Xiaoqun, Ma Xuejun. Multivariate statistical analysis of the —is based on the R language [M]. Beijing: China Renmin University Press, 2021.