G. Antille

Descriptive Analysis of Matrix-Valued Time-Series

In this article we present a technique of data analysis applied to three-dimensional tables as, for instance, matrix-valued time-series. The main goal of the method is to describe the evolution of the statistical units with respect to time in a space summarizing the set of matrices. Moreover, our technique points out similar statistical units provided by a classification of their trajectories.

1. Introduction

Large data sets are common in most sciences. In particular, factor analysis is widely used to study a set of observations of p variables measured on n statistical units. Nevertheless, few methods [Kroonenberg (1983)], [Escoufier (1985)], [Casin (1995), (1996)] are developed to analyse repeated observations of $n \times p$ matrices. Without loss of generality we will consider repetitions of these matrices only over the time.

As a global approach we conduct a three-steps analysis of this type of data:

1. Analysing each data matrix to get an idea of the data structure at time t, $t \in \{1,\dots,T\}$;
2. Constructing and analysing the $T \times p$ matrix whose rows contain the means or medians of each data matrix to get an idea of the global evolution of the process under study;

and, finally,

3. Finding a common space to describe the evolution of statistical units and relationships between variables with respect to time.

In the article we have mainly developed the third point with techniques based on Principal Component Analysis (PCA). PCA is also a way to perform the above mentioned first and second steps.

The concern of Section 2 will be the construction of the common space, on which the data sets will be projected. The choice of that space is based on an optimality criterion applied to measures of dispersion. In Section 3 we present a descriptive method of analysis of matrix valued time series which consists in projections, with respect to time, of statistical units or variables on principal directions of the common space. We call these projections *trajectories,* and we classify them to exhibit similarities. In Section 4 we apply our method to a $26 \times 8 \times 26$ matrix: 26 years of observations of 8 variables of rates of mortality in 26 Swiss cantons.

We define a data set as a three-dimensional matrix denoted by $\mathbf{K} = (x_{ijt})_{i=1,\dots,n;\, j=1,\dots,p\,;\, t=1,\dots,T}$, where x_{ijt} represents the value of the j^{th} variable for the i^{th} statistical unit at time t.

Another way to define such a data set is given by $\mathbf{K} = \{X_t | t = 1,\dots,T\}$, where X_t is the $n \times p$ matrix of observations at time t.

The first step in the analysis of our data sets is to perform PCA on the T matrices X_t to explore the structure of each matrix in order to point out important changes in the structure of the data. In this article we do not discuss this type of question.

The second step consists in applying PCA to $\bar{X}_u$, the $T \times p$ matrix of the means over statistical units, which t^{th} row equals $(\bar{x}_{.jt})_{j=1,\dots,p}$. The goal of this PCA is to summarize and exhibit the global evolution of the T matrices in subspaces generated by principal axes of $\bar{X}_u$.

In the third step, PCA is conducted on $\bar{X} = (\bar{x}_{ij.})$, the $n \times p$ mean matrix over the time, to define directions of projection of statistical units or variables to study their evolution with respect to t. We remark that if $T = 1$, there is no common space to define, and usual PCA provides exactly what we search for. Let us recall that the optimum properties of PCA are direct consequences of the properties of the Rayleigh quotient $r_V(\nu) = \dfrac{\nu' V \nu}{\nu' \nu}$, where V is the variance-covariance matrix related to X.

Hence the concern of this article is to extend the PCA approach for the cases when $T > 1$.

In this context the following question arises: how to define V, when T matrices of size $n \times p$ have to be analysed simultaneously, and the common space will be defined through a criterion based on a ratio similar to $r_V(\nu)$.

2. General Framework

Given a data set $\mathbf{K}$ there exists at least four ways to define the matrix V with respect to $\mathbf{X}$, a matrix constructed by means of the set of matrices X_t.

Let us consider $1' = (1 \quad 1 \quad \dots \quad 1)$, the identity matrix I, and $X_t^c = (I - 11'/n)X_t$. To simplify the notation let $X_t = X_t^c$ and $V_t = (X_t^c)' X_t^c$.

1. $X = (X_1 \quad X_2 \quad \dots \quad X_T)', \qquad W_1 = \sum V_t;$

2. $X = (X_1 \quad X_2 \quad \dots \quad X_T), \qquad W_2 = \begin{pmatrix} X_1'X_1 & X_1'X_2 & \dots & X_1'X_T \\ X_2'X_1 & & & \vdots \\ \vdots & & & \vdots \\ X_T'X_1 & X_T'X_2 & \dots & X_T'X_T \end{pmatrix};$

3. $X = \begin{pmatrix} X_1 & 0 & \dots & 0 \\ 0 & X_2 & & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & X_T \end{pmatrix}, \quad W_3 = \begin{pmatrix} X_1'X_1 & 0 & \dots & 0 \\ 0 & X_2'X_2 & & 0 \\ \vdots & & & \vdots \\ 0 & 0 & \dots & X_T'X_T \end{pmatrix};$

4. $X = \sum c_t X_t, \qquad W_4 = \sum k_{ij} X_i' X_j, \; k_{ij} = c_i c_j.$

Given V, let $\Gamma = \{\nu_1, \nu_2, \dots, \nu_r\}$ be a set of r orthogonal vectors and let us define $\varphi(\Gamma, V) = \sum r_V(\nu_k)$ as the sum of the squared lengths of the projections of the rows of $\mathbf{X}$ on ν_k, $k = 1, 2, \dots, r$. In cases 1 and 4 $\nu_k \in \Re^p$ and φ defines a global measure of dispersion of the data set captured by the space generated by the vectors of Γ.

At any time t, $\varphi(\Gamma, V_t)$ is the dispersion of the data matrix X_t captured by the space generated by Γ. As

$$\varphi(\Gamma, W_1) = \sum_t \varphi(\Gamma, V_t),$$

$\frac{1}{T}\varphi(\Gamma, W_1)$ measures the mean dispersion of the data set captured by the space generated by the set Γ.

Cases 2 and 3 are not relevant to our problem as the dimension of the projection directions do not match the original data.

In Case 4 we have

$$W_4 = \sum_t c_t^2 V_t + \sum_{i \neq j} c_i c_j X_i' X_j,$$

and we can write

$$\varphi(\Gamma, W_4) = \sum_t c_t^2 \varphi(\Gamma, V_t) + \sum_{i \neq j} c_i c_j \varphi(\Gamma, X_i' X_j),$$

which shows that the dispersion captured by Γ is composed of a within each year part and a between every couple of years part.

Moreover, if

$$c_t = \frac{1}{T},$$

then

$$\varphi(\Gamma, W_4) = \frac{1}{T^2}\left(\varphi(\Gamma, W_1) + \sum_{i \neq j} \varphi(\Gamma, X_i' X_j)\right).$$

As in the standard PCA we can consider a dual approach based on the study of the columns instead of the rows of the data matrices, and we define for $\Lambda = \{u_1, u_2, \ldots, u_s\}$, a set of orthogonal vectors, $\psi(\Lambda, V') = \sum r_{V'}(u_k)$.

In that framework we should solve the following optimisation problem (OP):

$$\max_{\Gamma} \varphi(\Gamma, V) \text{ and } \max_{\Lambda} \psi(\Lambda, V')$$

to get the directions of projections of the statistical units and the variables on their respective optimal subspace.

In fact, essentially for computational reasons we consider a slightly different optimisation problem and we solve OP sequentially. We start to solve OP with Γ and Λ containing one vector. Then we solve OP, with Γ and Λ still containing only one element, under the constraint of orthogonality of that solution to the previous one and so on.

It is well known that such solutions are given by the singular value decomposition of the above given $\mathbf{X}$ matrix. The obtained eigenvectors generate the common spaces Γ_c and Λ_c.

Classical methods as generalized PCA, generalised canonical analysis, or STATIS method fit in that general framework [Antille (2001)].

3. Trajectories

In the previous section we proposed a way to obtain a common space to represent the data set.

In the common space Γ_c, at any time t, the statistical units have their respective positions given by $X_t D$, where D is the $p \times r$ matrix of eigenvectors of $\mathbf{X}'\mathbf{X}$. Similarly, at any time t, $X_t' G$ gives the coordinates of the variables in Λ_c, where G is the $n \times r$ matrix of eigenvectors of $\mathbf{X}\mathbf{X}'$; and where r, with $r \leq rank(\mathbf{X})$, is equal to the number of eigenvectors we select to describe the data.

A k-dimensional trajectory of the i^{th} statistical unit with respect to time is defined by $p_i = [p_{i1}; p_{i2}; \ldots; p_{iT}]$, where $p_{it} = X_t^{(i)} D_{(k)}$, $X_t^{(i)}$ is the i^{th} row of X_t and $D_{(k)}$ is a $p \times k$ sub-matrix of D; the coordinates of the t^{th} vertex of that trajectory are given by p_{it}. As k can be seen as a degree of freedom left to the analyst there are as many as 2^k trajectories; some of them being interesting for graphical purposes, others for clustering statistical units.

Among the one-dimensional or two-dimensional trajectories only those corresponding to the largest singular values have statistical interpretation. The one-dimensional trajectories should be plotted versus the time, and graphical comparisons are quite easy (see Figures 4 and 5). The two-dimensional trajectories have to be plotted in the plane generated by two chosen principal axes (see Figure 3)

In order to compare trajectories we propose two distinct points of view, a location and an evolution one.

In the location approach, comparisons are based on distances between vertices of trajectories i and j, defined by

$$d_p^L(i,j) = \|p_i - p_j\|_p = \sum_t \|p_{it} - p_{jt}\|_p,$$

$\|\ \|_p$ being the L_p — *norm*. In this case two trajectories are equal if they match exactly.

In the evolution approach, proximities are based on distances between e_i^{tq}, lags of order q for the trajectory i, and e_j^{tq} — lags for j, where

$$e_l^{tq} = p_{lt} - p_{lt-q},\ t = q+1,\ldots,T,\ 1 \le q \le T-1,$$

with too large values of q being meaningless. In this case two trajectories are similar if they are linked by a translation.

As it can be easily seen, graphical presentations of trajectories are often useless as there are too many overlappings or simply too many trajectories on the plot. Clustering trajectories provide a way to detect similar statistical units with respect to the defined principal axes.

4. Application

As an illustration of our descriptive approach to analysing matrix-valued time-series we study the evolution of 8 mortality causes in the 26 Swiss cantons during 26 years. As there exist important geographical, economic, and cultural differences between cantons we expect to point out these differences by analysing our set of data, a $26 \times 8 \times 26$ matrix. Moreover, sizes of the population are also very different, so we have to consider the percentages of death due to infection, tumour, diabetes, hart disease, respiration, accident, suicide, and others for each canton. To construct the common space we choose the mean matrix over the time and perform a PCA on that matrix after standardization. In that case the Kaiser criterion implies that only the first three components are interesting as they capture 74.11% of the total dispersion. Eigenvectors, correlation with the axes, and contribution of the variables to the construction of the axes are given in Table 1. As computation was performed on a standardized matrix, correlations are equal to the coordinates of the variables on the corresponding axis. Figure 1 contains the representation of the variables in the common space of dimension two. As it can be seen, the first axis opposes heart disease to infection and other causes of death, and the second axis opposes accident to tumour and diabetes.

Table 1

Eigenvectors, correlation, and contribution of the variables

	CP1			CP2			CP3		
Infection	–.45	–.81	19.9	.13	.16	1.64	.03	.03	.06
Tumour	–.34	–.61	11.27	–.49	–.60	24.37	.25	.26	5.97
Diabetes	.30	.54	8.72	–.37	–.45	13.57	.44	.47	19.00
Heart disease	.52	.94	26.71	.14	.17	2.05	–.16	–.18	2.72
Respiration	–.26	–.47	6.82	–.25	–.30	6.18	–.19	–.20	3.50
Accident	–.14	–.25	1.85	.70	.84	47.63	.15	.16	2.14
Suicide	–.10	–.17	.93	–.17	–.21	2.98	–.80	–.87	65.24
Others	–.49	–.88	23.8	.13	.15	1.58	.12	.13	1.37

The locations of the cantons (the list of abbreviations is provided in the appendix to the paper) on the first PCA plane provide information on the similarities of causes of death; for instance, on the first axis we observe that GE and VS have the highest rate of mortality due to infection as pointed out on Figure 1.

The main interest of this descriptive method is due to the possibility of following graphically the evolution of cantons with respect to time and making comparisons. The weakness of the method is a large amount of information, which could be included in such projections. Drawing all the trajectories does not make sense as, usually, they will overlap, and the figure will be almost covered by lines. Drawing trajectories for a few cantons, which seems similar on the compromised

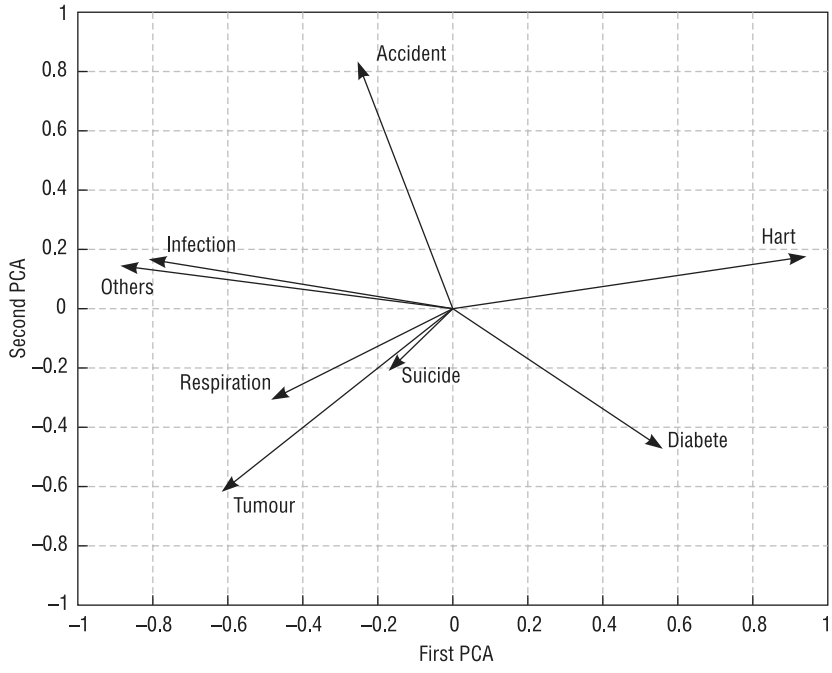

Figure 1. Plot of the variables on the first PCA plane

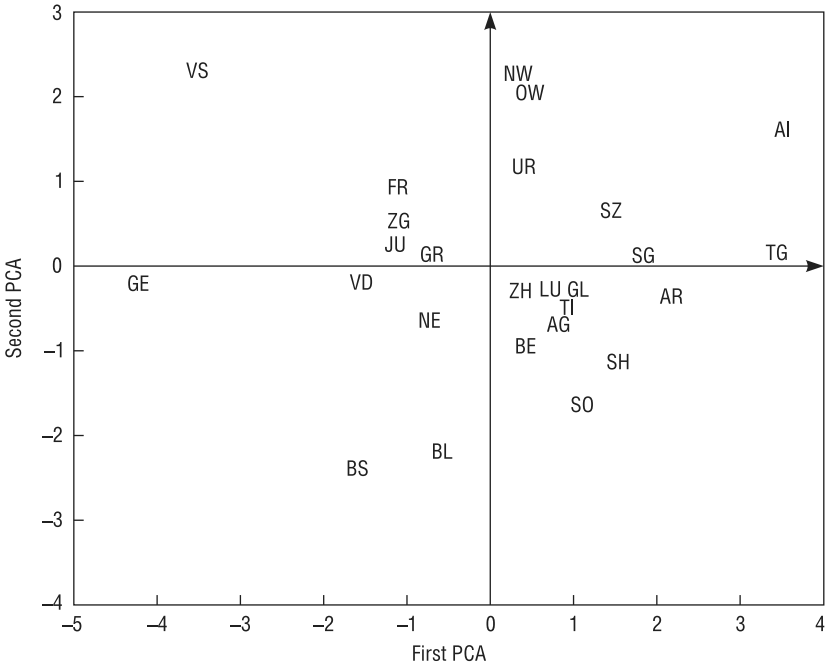

Figure 2. Plot of the cantons on the first PCA plane

space, allows exhibiting differences as shown on Figure 3 for GE and VS on the first PCA plane. Drawing one-dimensional trajectories, as for examples first PCA versus time or second PCA versus time, is sometimes more informative (as shown on Figure 4 and Figure 5, again for GE and VS.) Differences are obvious.

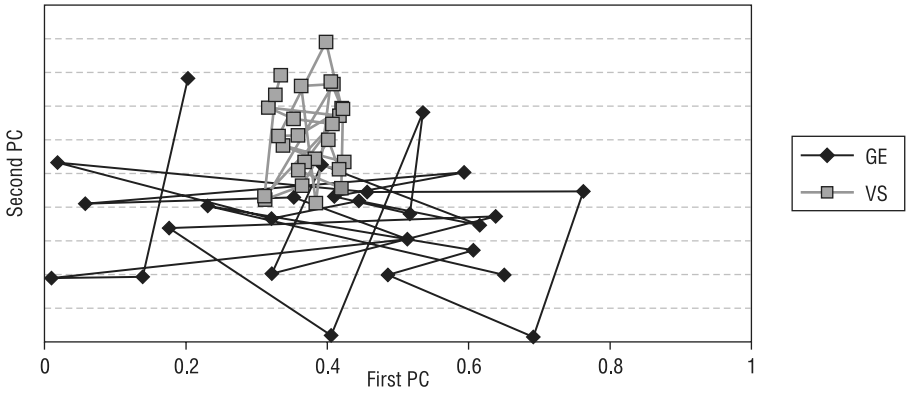

Figure 3. Trajectories of GE and VS on the first PC plane

These plots provide information to compare cantons or, more generally, to compare statistical units.

Plots of variables versus time allow observing the evolution of causes of death as shown on Figure 6, where we see the increase of diabetes. The trajectory of the infections changed sharply around 1985, when the AIDS started to be counted as an infection.

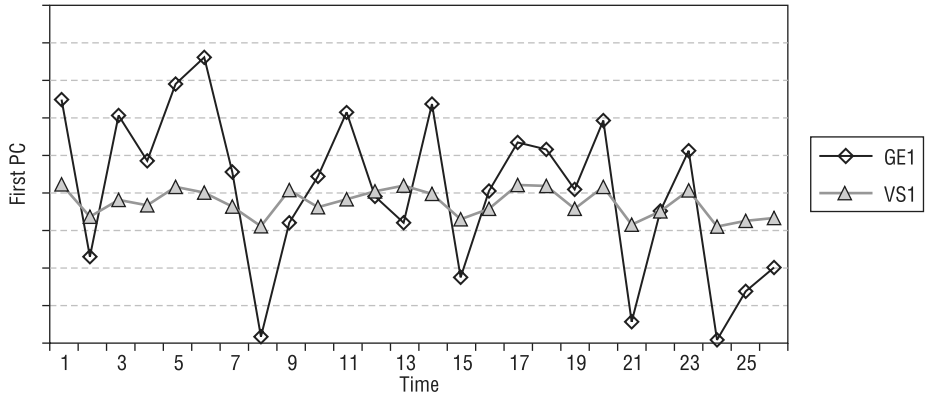

Figure 4. Trajectories of GE and VS, first PC versus time

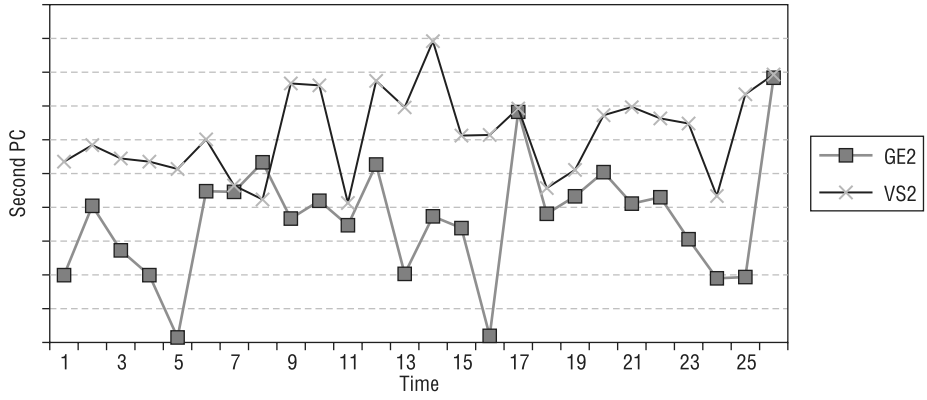

Figure 5. Trajectories of GE and VS, second PC versus time

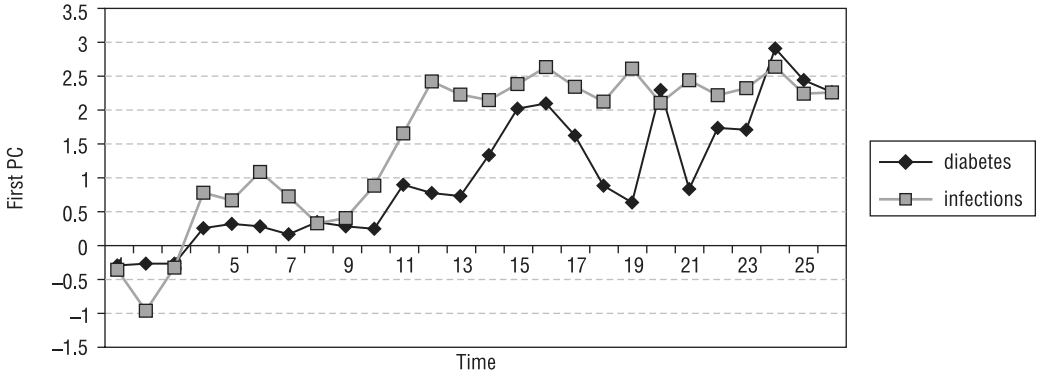

Figure 6. Diabetes and infections, first PC versus time

For general comparison we suggest using classification of trajectories. Figure 7 presents a classification tree of the cantons performed on Euclidean distances between two-dimensional trajectories of the cantons with respect to the first PCA plane. As it can be seen, the structure provided by the classification is close to the latent structure shown on Figure 2. This fact is explained by the repartition of causes of death, which was almost stable during the observation period as the structure reflected by time is similar to the one given by the common space.

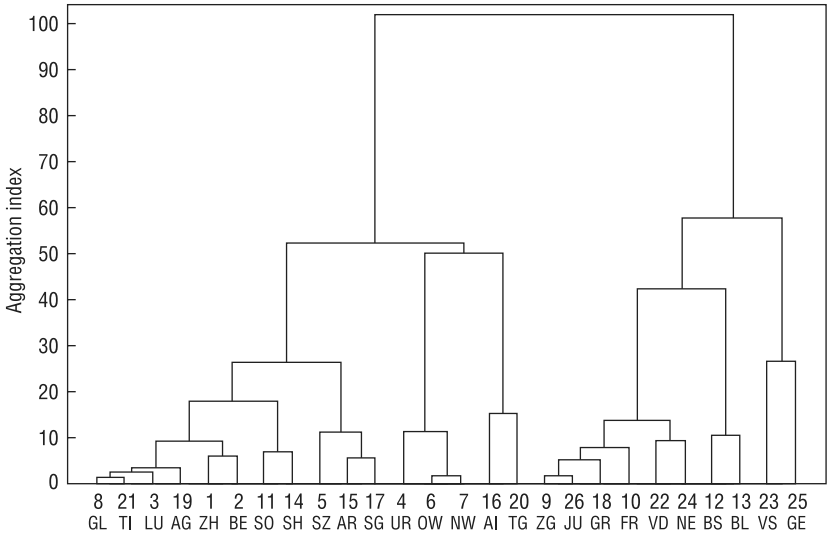

Figure 7. Classification tree of cantons based on two-dimensional trajectories

5. Conclusion

In this article a descriptive method of analysing three-dimensional matrices is presented. Principal component analysis of a matrix, summarizing the data, provides directions of projections of statistical units for construction of their trajectories with respect to time. Plots, clustering methods, and classification trees of trajectories allow comparison of the evolution of the units. For the variables a dual approach can be performed, and comparisons with respect to time are possible.

Appendix

The following table contains the list of abbreviations of the Swiss cantons.

Table 2

1. Zurich	ZH	8. Glarus	GL	15. Appenzell R	AR	22. Vaud	VD
2. Berne	BE	9. Zug	ZG	16. Appenzell I	AI	23. Valais	VS
3. Lucerne	LU	10. Fribourg	FR	17. St Gallen	SG	24. Neuchâtel	NE
4. Uri	UR	11. Solothurn	SO	18. Grisons	GR	25. Geneva	GE
5. Schwytz	SZ	12. Basel-City	BS	19. Aargau	AG	26. Jura	JU
6. Obwalden	OB	13. Basel-L	BL	20. Thurgau	TG		
7. Nidwalden	NI	14. Schaffhausen	SH	21. Ticino	TI		

References

Antille G. Analyse de la composante chronologique dans les tableaux croisés // *Revue suisse d'économie et de statistique.* 2001.

Casin Ph. L'analyse discriminante des tableaux évolutifs //*Revue de Statistique Appliquée XLIII.* 1995.

Casin Ph. L'analyse en composantes principales généralisée //*Revue de Statistique Appliquée XLIV.* 1996.

Escoufier Y. Objectifs et procédures de l'analyse conjointe de plusieurs tableaux de données // *Statistique et Analyse de données.* 1985. № 10.

Kroonenberg P. Three mode principal component analysis. Theory and applications. Leiden, reprint. 1983.

Ж. Антиль

Дескриптивный анализ временных рядов с матрицами в качестве значений

В настоящей статье мы представляем метод анализа данных, применяемый для трехмерных таблиц, например, для временных рядов с матрицами в качестве значений. Основная задача данного метода — описание эволюции статистических единиц относительно времени в пространстве, суммирующем множество матриц. Более того, эта техника позволяет выявить похожие статистические единицы, обнаруженные на основе классификации их траекторий.

Краткое изложение

Крупные массивы данных часто встречаются в научных исследованиях. В частности, факторный анализ широко применяется для изучения множества наблюдений p переменных, измеренных для n статистических единиц. Тем не менее, среди существующих [Kronenberg (1983)], [Escoufier (1985)], [Casin (1995), (1996)] крайне мало методов для анализа матриц размерности $T\, n \times p$.

Для этого типа данных может быть проведен трехступенчатый анализ.

- Анализ каждой матрицы данных для получения точечных представлений об их структуре.
- Конструирование и анализ матрицы $T \times p$, чьи строки содержат средние или медианные значения каждой матрицы данных для получения представления о глобальной эволюции изучаемого процесса.
- Нахождение общего пространства для описания эволюции статистических единиц и взаимоотношений между переменными относительно $t \in T$.

Каждый из этих шагов основан на анализе главных компонент (*АГК*), выполненном для различных матриц. Общеизвестно, что оптимальные свойства *АГК* являются непосредственным следствием свойств отношения Релея (Rayleigh) $r_V(\mathbf{v}) = \dfrac{\mathbf{v}'V\mathbf{v}}{\mathbf{v}'\mathbf{v}}$, где V — ковариационная матрица этого множества данных. В данном контексте возникает следующий вопрос: как определить V, когда матрицы T размером $n \times p$ должны быть проанализированы одновременно? Тогда общее пространство будет определено через критерий, основанный на отношении, сходном с $r_V(\mathbf{v})$.

В статье представлено четыре разных способа описания множеств матриц. Таким образом, мы имеем четыре различные ковариационные матрицы V, используемые для определения четырех глобальных мер дисперсии $\varphi(\Gamma, V) = \sum r_V(\mathbf{v}_k)$ множества данных, охваченных пространством, порожденным множеством ортогональных векторов $\Gamma = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$. Как

и в случае стандартного *АГК*, при использовании двойственного подхода мы получаем $\psi(\Lambda, V^t) = \sum r_{V^t}(u_k)$ — глобальную меру дисперсии в двойственном пространстве.

При такой постановке задачи, для получения направлений проекций статистических единиц и переменных на их соответствующее оптимальное подпространство, мы должны решить следующую оптимизационную задачу: $\max_{\Gamma} \varphi(\Gamma, V)$ и $\max_{\Lambda} \psi(\Lambda, V^t)$. Общеизвестно, что решения этой задачи получаются спектральным разложением V и V^t. Эти собственные векторы генерируют общие пространства. Классические методы, такие как обобщенные *АГК*, канонический анализ или метод *СТАТИС* соответствуют этой постановке задачи [Antille (2001)].

Дескриптивный метод анализа временных рядов *с матрицами в качестве значений*, обсуждаемый в настоящей статье, состоит в проецировании относительно времени статистических единиц или переменных на основные направления общего пространства. Мы называем эти проекции *траекториями*. Размерность траектории равна размерности подпространства, на которое производится проекция. Среди одномерных или двумерных траекторий могут быть статистически интерпретированы только траектории, соответствующие максимальным собственным значениям. Одномерные траектории должны быть нанесены на график против T, что позволяет достаточно просто производить графические сравнения (рис. 4 и 5). Двумерные траектории наносятся на плоскость, генерируемую двумя главными осями (рис. 3).

Для сравнения траекторий, на основании классификации, мы предлагаем использовать две разные точки зрения — расположения и эволюции. С точки зрения расположения, сравнения основываются на расстояниях между вершинами двух или более траекторий. В этом случае две траектории равны, если они полностью совпадают. В подходе, основанном на эволюции, сравнения строятся на расстояниях между временными интервалами одного и того же порядка q для двух или более траекторий. В этом случае две траектории подобны, если связаны перемещением. Кластеры траекторий дают возможность выявлять сходные статистические единицы относительно уже определенных главных осей.

В качестве иллюстрации вышеизложенного изучается уровень смертности по 8 причинам в 26 швейцарских кантонах в течение 26 лет. Для конструирования общего пространства выбрана матрица средних значений за этот период времени и проведена *АГК* этой матрицы после стандартизации. На рис. 1 воспроизведены переменные в общем пространстве размерности 2. Как показано на рисунке, первая ось противопоставляет сердечно-сосудистые и инфекционные заболевания другим причинам смерти, а на второй оси — диабет — самоубийствам. Положение кантонов на плоскости первых двух главных компонент дает информацию о сходствах в причинах смертности. Но основной интерес данного дескриптивного метода заключается в получении графического инструмента для представления эволюции кантонов во времени и возможности сравнивать их (рис. 4 и 5). Отображение «переменных против времени» позволяет наблюдать эволюцию причин смертности (рис. 6). Для глобального сравнения рекомендуется использовать классификацию траекторий (рис. 7). Структура, представленная этой классификацией, близка скрытой (латентной), показанной на рис. 2. Этот факт объясняется тем, что распределение причин смертности оставалось практически неизменным в течение всего периода наблюдения.