Projet STATIS

Feuille de route

Introduction

On se propose d'étendre l'ACP à l'analyse factorielle d'un multi-tableau.

Il existe plusieurs façons de le faire, STATIS en est une.

Il existe aussi plusieurs types de multi-tableaux. Nous en distinguerons deux:

Tableau à trois entrées (situation 1)

Il s'agit d'un tableau à trois indices, par exemple un tableau décrivant n individus à l'aide de p variables à T dates. Les indices courants seront notés $1 \leq i \leq n$, $1 \leq j \leq p$ et $1 \leq t \leq T$.

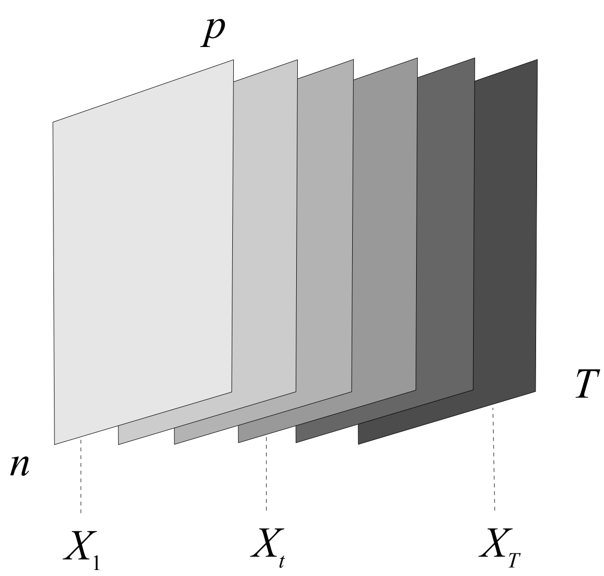

Figure 1. Tableau à trois entrées

Un tel tableau peut être décomposé en T tableaux juxtaposés de dimensions *identiques* (n, p) . Cette identité de dimensions entre les T tableaux homologues permet une extension de l'ACP dans laquelle on considère chaque tableau comme une "variable" décrivant $n \times p$ "individus" (i, j) .

Nous formaliserons d'abord cette extension de l'ACP, appelée STATIS1.

Tableau avec partition thématique (situation 2).

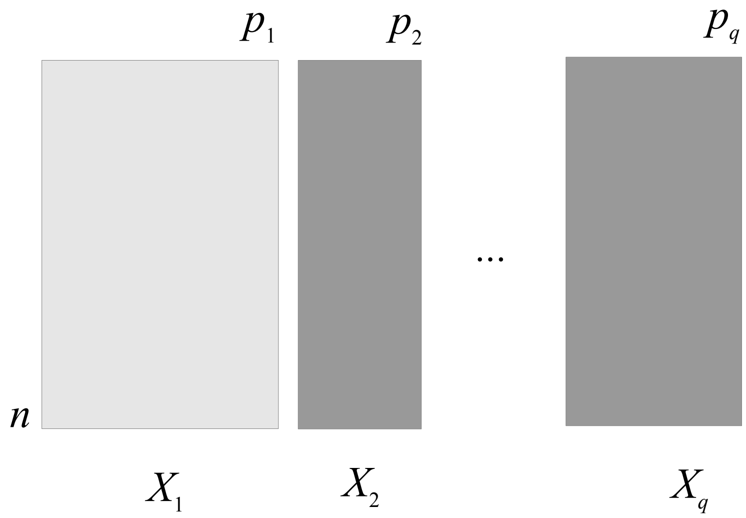

Figure 2. Tableau à partitionnement thématique

Un tel tableau est constitué de q tableaux décrivant les n individus à l'aide de groupes de variables différents, chaque groupe appartenant conceptuellement à un thème précis. Chacun de ces tableaux $X_m, 1 \leq m \leq q$ possède p_m variables (colonnes).

Le fait que ces tableaux ne possèdent qu'une dimension commune, celle des individus, ne permet pas d'utiliser directement STATIS1. Nous devrons donc adapter cette dernière à la situation 2.

I - Situation 1

1. Préliminaires

0. Cherchez (ou fabriquez) un tableau à trois entrées décrivant n individus avec p variables numériques à T dates. N.B. Il doit s'en trouver sur le site de l'INSEE, mais ce n'est pas forcément la meilleure source; j'en ai demandé à une autre source, au cas où.

Vous pourrez centrer-réduire les colonnes de vos tableaux si ce sont des indicateurs numériques, mais ce n'est qu'une option parmi d'autres, parce qu'en définitive, plusieurs pré-traitements seront possibles selon la nature du tableau comme de la variante de STATIS envisagée.

1.1. Produit scalaire

La matrice diagonale des poids des n individus est W (par défaut, $W = \frac{1}{n} I_n$).

On considérera également une matrice diagonale des poids des p colonnes : $C = \frac{1}{p} I_p$ par défaut.

1. Le produit scalaire entre deux matrices A et B de taille (n, p) est :

$$[A|B] = tr(CA'WB)$$

La norme d'une matrice A correspondant à ce produit scalaire sera notée $[\![A]\!]$.

a) Écrivez ce produit scalaire sous forme $tr(\tilde{A}'\tilde{B})$ (produit scalaire de Frobénius) en explicitant la transformation $Z \to \tilde{Z}, \ \forall Z(n,p)$.

b) Montrez qu'il s'agit bien d'un produit scalaire.

c) Explicitez ce produit scalaire sous forme de double somme.

d) Programmez ce produit scalaire en R, ainsi que la norme d'une matrice.

1.2. Coefficient RV

2. On définit le coefficient RV d'Escoufier entre deux matrices A et B de taille (n,p) par :

$$R(A,B) = \frac{[A|B]}{[\![A]\!][\![B]\!]}$$

a) Interprétez RV en termes géométriques.

b) Programmez RV en R, ainsi qu'une fonction donnant la matrice Γ des coefficients RV entre T tableaux $X_t(n,p)$.

2. Programme de STATIS1

2.1. Programme

a) Interprétez en termes d'inertie le programme suivant :

$$\max_{\|u\|^2=1} [\![\sum_{t=1}^{T} u_t \frac{X_t}{[\![X_t]\!]}]\!]^2$$

b) Résolvez le programme.

c) Écrivez le programme R fournissant les vecteurs u solutions des équations du premier ordre. Montrez que ces vecteurs u forment une base I -orthonormée.

2.2. Équivalence avec l'ACP d'un tableau juxtaposé "dépliant" le tableau cubique

a) Montrez que les vecteurs u obtenus sont les vecteurs propres unitaires de l'ACP d'un tableau dont les X_t sont les "variables", et dont on précisera les "individus" ainsi que leurs poids.

b) En déduire que les composantes principales $F^1, \ldots, F^k, \ldots$ sont orthogonales au sens du produit scalaire convenable.

c) Donnez la formule des composantes principales correspondant aux vecteurs u .
Programmez la fonction calculant la $k^{\text{ième}}$ composante F^k ainsi que la composante normée f^k correspondante.

Programmez la fonction donnant la représentation graphique des individus en plan principal (k,l) .

d) Calculez le cosinus entre X_t et F^k .

Programmez la fonction donnant la représentation graphique des tableaux en plan principal (k,l) .

2.3. Quelles ACP d'autres "dépliages" du tableau cubique ?

a) Quels sont les autres dépliages possibles du tableau cubique initial?

b) Quelle(s) ACP des tableaux dépliés pouvez-vous envisager? Pour chacune: quels sont ses "individus" et ses "variables"? Quel centrage paraît opportun? Quelle réduction? Quels éléments sont utiles à projeter en supplémentaire?

c) Quels sont les avantages et les inconvénients de chaque ACP d'un tableau "déplié"?

Finalement, le "dépliage" opéré par STATIS1 était-il le plus opportun?

II - Situation 2