

Analyse en composantes principales et analyse factorielle: Les bleuets et les grenouilles

Voir texte de Tarkonen (notes de cours)

- Tarkonen propose l'analogie suivante:
 - En analyse en composantes principales, on a cueilli des bleuets et on pense que la cueillette a été parfaite, il n'y a que des bleuets. On peut verser directement dans la tarte aux bleuets.
 - En analyse factorielle, il y a quelques grenouilles et quelques branches qui se sont glissées parmi les bleuets. Il faut nettoyer avant de verser dans la pâte à tarte.

4

Les "types" d'analyse factorielle

L'analyse en composantes principales

- Elle met l'accent sur ce qui est propre à l'ensemble de la variance. On décompose toute la variance de l'ensemble des variables.
- On cherche à expliquer le maximum de variance.
- Il existe toujours une solution parce qu'il s'agit d'une méthode "numérique". On ne fait pas d'estimation, on décompose.
- La solution est unique, il en existe une seule.
- La solution est fortement dépendante du premier facteur (les autres lui sont orthogonaux).

5

Les types d'analyse factorielle

L'analyse factorielle dite "théorique" ou "en facteurs communs" ou "des psychologues"

- Elle est basée sur un modèle théorique.
- Elle décompose uniquement la variance commune à au moins deux variables.
- Elle cherche à reproduire la matrice de corrélation originale (et non pas les données individuelles). C'est une méthode d'estimation fonctionnant par itération visant à minimiser une fonction.
- Il y a de multiples solutions.
- Il y a plusieurs types d'extraction et plusieurs méthodes de rotation.

6

Types de variables

- Les variables doivent être de type métrique (quantitatives, continues).
- La pratique dans l'école anglo-saxonne est de traiter les variables ordinales (avec des échelles de type Likert) comme des variables quantitatives.
- **On ne peut pas introduire des variables nominales dans l'analyse** MAIS on peut dichotomiser des variables nominales en plusieurs variables de type 0 et 1 pour les utiliser dans l'analyse.

13

Méthodes d'extraction

- Composantes principales: décompose de façon simple l'ensemble de la variance.
- Factorielle (liste non exhaustive):
 - ▶ **Maximum de vraisemblance**: maximise la généralisabilité à la population (Test de χ^2).
 - Très sensible à la normalité des distributions.
 - ▶ **Moindres carrés non-pondérés**: minimise les résidus.
 - Appropriée en particulier quand les distributions des variables entrées dans l'analyse ne sont pas normales.
 - ▶ **Alpha**: maximise l'inférence psychométrique (la capacité de créer des échelles).
 - Moins utilisée mais néanmoins intéressante.

14

Méthodes de rotation

- Aucune rotation:
 - ▶ Avec l'analyse en composantes principales: donne un aperçu de la composition du premier facteur commun.
- Orthogonale: postule que les facteurs sont indépendants les uns des autres.
- Oblique: postule qu'il y a corrélation entre les facteurs (ce qui est habituellement le cas en sciences sociales)
 - ▶ Mais on ne veut pas avoir des facteurs trop corrélés entre eux.

15

Y-a-t-il une solution factorielle appropriée?

Divers indices permettent de le déterminer...

- Le KMO (Kaiser-Meyer-Olkin) indique jusqu'à quel point la solution est appropriée (similaire au Alpha de Cronbach, minimum de 0,70).
- Le déterminant de la matrice: doit être le plus petit possible mais non égal à zéro (parce que cela voudrait dire qu'une variable est une copie ou une combinaison parfaite de plusieurs variables déjà dans l'analyse).
- Le test de Bartlett vérifie que l'hypothèse d'indépendance entre les variables peut être rejetée.

16

Est-ce que chacune des variables doit être gardée dans l'analyse?

- La qualité de la représentation de chaque variable devrait être plus grande que 0,20.
 - L'information qu'il faut regarder est celle disponible dans "statistiques initiales" avant extraction de l'analyse factorielle.
- Chaque variable devrait avoir une saturation plus grande que 0,30
 - Avec au moins un facteur...
 - Mais avec un SEUL facteur (sinon variable complexe).

17

Que cherche-t-on?

Une structure simple

- On cherche à obtenir une structure simple où toutes les variables sont liées à **un facteur et à un seul facteur**.
- L'obtention d'une telle structure est l'indice d'une bonne validité convergente et discriminante, soit :
 - Les variables liées au même facteur sont fortement corrélées entre elles.
 - Et elles sont peu corrélées aux autres facteurs.
- Pour ce faire, on utilise un processus itératif où les variables qui ne répondent pas aux critères d'appartenance à la solution factorielle sont tour à tour retirées jusqu'à l'obtention de la structure désirée.

18

Un exemple....

- Voir sorties informatiques.
