L'analyse factorielle

Cours SOL6210 Analyse quantitative avancée

© Claire Durand, 2024
Professeur titulaire,
Département de sociologie,
Université de Montréal

1

Un peu de contexte

Les facteurs de 36

- 2 x 18
- 2 x 2 x 9
- 4 x 9
- 4 x 3 x 3
- 2 x 2 x 3 x3
- On est dans le même univers... On décompose en facteurs de sorte de pouvoir retrouver le nombre original.

2

Un peu de contexte

Les hypothèses de base (voir Héraux et Novi, 1974)

- Hypothèse générale: Il existe de grandes dimensions. Le positionnement des répondants sur ces dimensions explique leurs réponses similaires à un certain nombre de questions mesurant ces dimensions.
- H1 Il existe des facteurs rendant compte de ces dimensions.
- H2 Il existe une fonction décrivant les valeurs de x (les réponses aux questions) en fonction des saturations sur les facteurs.
- H3 Les facteurs sont...ou ne sont pas... indépendants entre eux (orthogonalité).
- H4 La distribution des valeurs des indicateurs (les réponses aux questions) suit une loi normale (mais certaines méthodes d'extraction tiennent compte de la non normalité).

3

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.

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).

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.

Les “types” d’analyse factorielle

L’analyse factorielle des correspondances

- Il s’agit d’une “classe à part” que nous verrons dans un prochain cours. Elle est fondée sur les mêmes principes que l’analyse en composantes principales.
- Elle sert surtout pour analyser les variables nominales.
- Elle permet de mettre sur le même plan indicateurs et objets (cas, répondants).

7

Les “types” d’analyse factorielle

Exploratoire ou confirmatoire

- On fait habituellement des analyses exploratoires, du moins lorsqu’on utilise les grands progiciels d’analyse disponibles.
- On peut faire des analyses confirmatoires en utilisant les modèles d’équations structurelles (que nous verrons au dernier cours).
  - Il s’agit alors de vérifier des hypothèses quant à des relations prédéterminées entre l’ensemble des mesures, d’une part, et des facteurs latents, d’autre part.

8

Pourquoi faire une analyse factorielle?

- Pour comprendre la structure des relations entre les divers indicateurs.
- Pour être en mesure de réduire le nombre de variables, extraire de grandes dimensions qui résument la majeure partie de l’information pertinente.
  - → On cherche à expliquer le plus de variance possible avec le moins de facteurs possibles.
- Pour nous guider dans la création de nouvelles variables indépendantes entre elles, ce qui sera utile dans des analyses de régression subséquentes.

9

Information recherchée

- Combien de facteurs doit-on retenir?
- Les facteurs sont-ils indépendants ou corrélés entre eux?
- Quelles variables sont liées à quel facteur? Lesquelles garder, lesquelles enlever?
- Quelle proportion de la variance la solution factorielle explique-t-elle?
- **Comment nommer les facteurs obtenus?** Que représentent-ils?

10

Comment faire?

- Tout ou presque est permis en autant que l'on demeure "modeste", c'est-à-dire
  - Garder à l'esprit que l'on est dans l'exploration et non la confirmation (lorsque c'est le cas).
  - Garder à l'esprit que l'échantillon **devrait** être un échantillon aléatoire de la population. Si ce n'est pas le cas, rester d'autant plus modeste quant à la généralisation.

11

Comment faire?

- Choisir les variables.
  - Attention, on souhaite un minimum de 10 cas par variable. Si cela n'est pas possible:
    - Possibilité de faire une analyse de classification sur les variables.
    - Possibilité de faire des analyses sur des sous-ensembles de variables dans un premier temps.
- Choisir le mode d'extraction approprié.
- Choisir le mode de rotation approprié.
- Demander l'ensemble de l'information nécessaire (**voir notes de cours**).

12

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.