

Les Méthodes Exploratoires

* Il existe un grand nombre de méthodes destinées à traiter des données. Ces méthodes se divisent en 2 groupes.

→ Lorsque'il n'est pas possible d'envisager, parmi un ensemble de variables, des relations de causalité (une ou plusieurs variables explicatives et une ou plusieurs variables expliquées), le chercheur fait appel aux méthodes exploratoires.

* Les méthodes exploratoires cherchent à identifier pourquoi et comment les variables sont liées les unes aux autres, ou encore si il existe des sujets (entreprises, personnes, boissons) qui peuvent être regroupés en raison de leur proximité.

Pour mettre en évidence des structures préexistantes dans les données.

Elles peuvent être utilisées avec des données qualitatives ou quantitatives

Tableau: Panorama des Méthodes Exploratoires

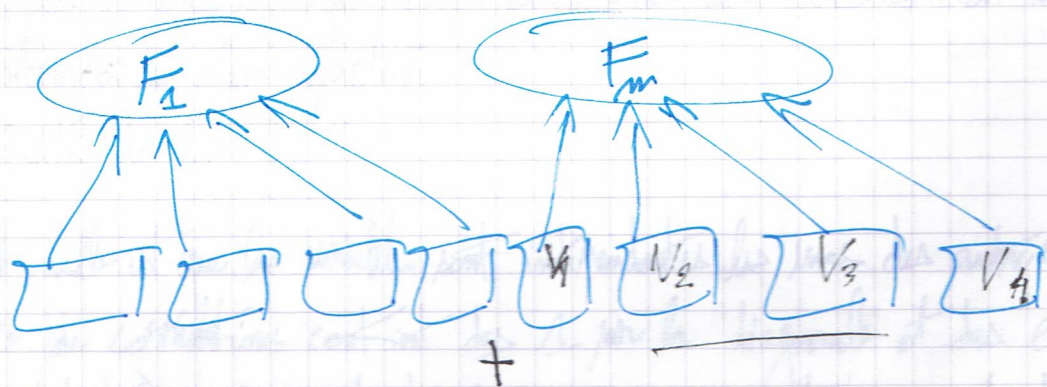
	Intervalle ou rapport	Ordinal	Nominal
Variables	Analyse Factorielle en Composantes Principales	Analyse des Similarités et des Préférences	Analyse Factorielle des Correspondances
Sujets	Analyse typologique		

Analyse Factorielle en Composantes Principales (ACP):

L'ACP est une technique explicative dont l'objet est de chercher, à partir d'un ensemble de K variables mesurées sur une échelle d'intervalle ou de rapport, une logique, une structure sous-jacente... Si une telle structure existe, l'identifier permet de simplifier l'information contenue dans une base de données, en substituant aux K variables initiales un ~~plus~~ groupe plus petit de m facteurs.

Toute simplification entraîne une perte d'information. Ainsi l'information contenue dans l'ensemble de K variables, est plus détaillée... Le passage à m facteurs par une représentation moins précise de l'état des observations, dont la variance résiduelle est inférieure à celle de la variance initiale.

* L'objectif du chercheur est de trouver le meilleur compromis entre les bénéfices de la simplification et les inconvénients de la perte d'information due au passage de K variables à m facteurs (F).



m facteurs ($m < K$)
~~Nombre de variables~~ Variance
 résiduelle



K variables initiales
 Nombre de variance

Figure 0: Objectif de l'analyse factorielle en composantes principales.

* Une autre utilisation courante de l'ACP est la construction d'échelles.

Démarche 0:

(A) Vérification des conditions d'application 0

La construction de facteurs synthétiques se fait sur la base d'un lien observé entre les

→ plus les individus auront répondu de façon liée à deux questions, par exemple, plus la probabilité que les deux variables correspondantes contribuent à la formation d'un même axe est grande. Les corrélations entre les variables initiales, évaluant les liens elles, fournissent donc la matière première de l'analyse. Des variables fortement corrélées, c'est-à-dire dont le coefficient de corrélation tend à 1 en valeur absolue, ont donc des chances d'être regroupées au sein d'un même axe.

→ À l'inverse, des variables explicatives, c'est-à-dire dont la corrélation tend vers 0, ne peuvent contribuer à un même facteur.

Le test de sphéricité de Bartlett

L'hypothèse nulle est que les variables sont indépendantes les unes des autres, c'est-à-dire que leur matrice de corrélations contient des 1 sur la diagonale et des 0 partout ailleurs (matrice identité). La significativité associée à ce test permet de conclure si la situation réelle est proche de cette situation théorique (significativité $> \alpha$, [la factorisation n'est pas pertinente]).

L'indice K_1

L'indice de Kaiser-Meyer-Olkin s'intéresse, quant à lui, aux corrélations partielles (c'est-à-dire spécifiques à chaque paire de variables et en contrôlant les liens avec toutes les autres) qu'il compare aux corrélations. Jugé inacceptable s'il est inférieur à 0,5, il est d'autant meilleur qu'il est proche de 1.

② Détermination du nombre de facteurs :

Le nombre m de facteurs qu'il est possible de retenir est compris entre 1 et le nombre de variables - 1 puisque l'objectif est d'obtenir un nombre de facteurs plus petit que le nombre de variables. Les approches les plus communément utilisées s'appuient sur la variance expliquée par les facteurs.

Le critère de Kaiser

La valeur propre d'un facteur est le rapport entre la variance qu'il restitue et la prop. $1/k$. Si ce rapport est inférieur à 1, cela signifie que le facteur restitue moins de variance que la variance initiale, et c'est un mauvais résultat pour un facteur synthétiser plusieurs variables. Kaiser (1960) propose de ne retenir que les facteurs dont la valeur propre excède 1.

Le critère de Cattell

Le chercheur, partant du ~~premier~~ premier facteur, doit estimer ce qu'il gagne de variance en ajoutant un deuxième facteur, puis un troisième etc. L'augmentation du nombre de facteurs nuait à la simplicité de la solution factorielle, jusqu'où cela va. Cette question peut être résolue graphiquement, en représentant les valeurs propres décroissantes sous une forme semblable à celle de la figure en de

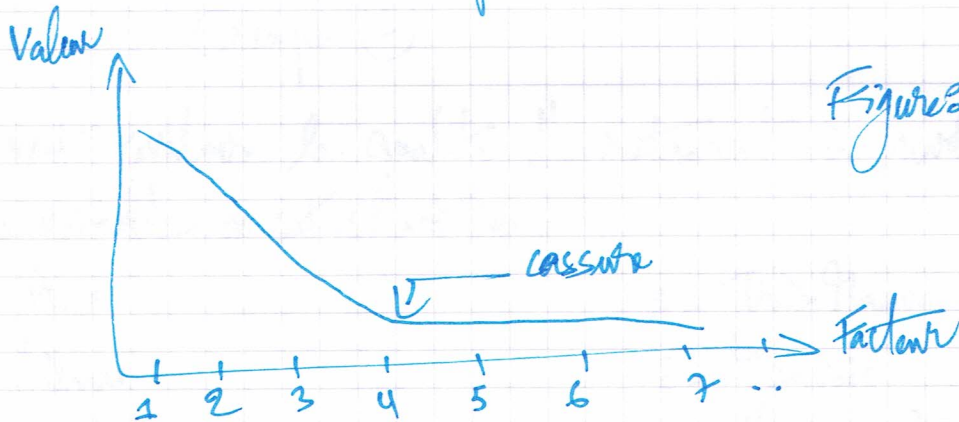


Figure 2 Graphique des Valeurs Propres et le Test de Cattell.

Les derniers facteurs n'ont plus que très peu de variance propre et leurs valeurs propres tendent à s'homogénéiser, donnant une forme à peu près horizontale à la courbe \Rightarrow La cassure intervient au quatrième facteur, indiquant qu'il convient de retenir les trois premiers selon ce critère.

La variance marginale restituée par chaque facteur supplémentaire.