1. Qu'est-ce que l'analyse factorielle? A quoi çà sert?

On utilise le terme générique d'analyse factorielle pour parler de deux types d'analyse ayant de nombreux liens de parenté mais légèrement différentes: l'analyse en composantes principales et l'analyse factorielle proprement dite. Avant d'examiner les différences entre ces deux types d'analyse, il est pertinent de voir les points communs à la famille des analyses factorielles.

L'analyse factorielle cherche à **réduire un nombre important d'informations** (prenant la forme de valeurs sur des variables) **à quelques grandes dimensions**. Comme dans toute analyse statistique, on tente donc d'expliquer la plus forte proportion de la variance (de la covariance dans le cas de l'analyse factorielle) par un nombre aussi restreint que possible de variables (appelées ici composantes ou facteurs). On utilise le terme de variables latentes pour parler de ces variables qui existent au plan conceptuel seul et qui ne sont pas mesurées.

Exemple:

- De façon à mesurer la satisfaction des gens face à leur travail, j'ai d'abord déterminé que celle-ci portait sur trois grandes dimensions: la qualité des relations interpersonnelles, la nature même du travail et les aspects extrinsèques (salaire, horaire,...).

- Pour chacune des dimensions, j'ai posé quatre (4) questions du type "Etes-vous très satisfait, assez satisfait, peu satisfait ou pas du tout satisfait a) de la qualité de vos relations avec vos collègues... b) de la qualité de vos relations avec vos supérieurs... c) de la qualité de vos relations avec vos subordonnés d) de la qualité générale des relations interpersonnelles à votre travail...

- En agissant ainsi, je **suppose qu'une dimension générale de satisfaction face au climat des relations interpersonnelles existe** et que le positionnement des individus face à cette dimension "explique", "prédit" leur positionnement sur chacune des "variables mesurées".

- **Si cette hypothèse est vraie,** les personnes auront tendance à répondre de la même manière aux quatre questions portant sur cette dimension **et** leurs réponses à ces questions seront plus correlées entre elles qu'avec les autres variables pour lesquelles on demande leur degré de satisfaction.

- Cette perspective suppose aussi que l'on conçoit que les variables mesurées constituent un échantillon de l'ensemble des variables aptes à mesurer le concept choisi.

L'analyse factorielle tente de donner un sommaire des patrons de corrélations entre les variables. Elle tente de décomposer les patrons de corrélations pour les expliquer par un nombre restreint de dimensions. Elle est souvent utilisée comme méthode d'analyse exploratoire en vue de créer des échelles.

Les équations:

Contrairement à l'entendement intuitif, il faut comprendre que ce sont les réponses aux variables mesurées qui dépendent des scores aux facteurs et non pas l'inverse. Ainsi, l'analyse pose que la réponse des individus à diverses questions portant, par exemple, sur leur satisfaction face aux relations avec leurs collègues ou avec leurs supérieurs dépend de deux éléments: d'une part, la satisfaction générale face aux relations interpersonnelles et d'autre part, un élément résiduel qui comprend l'erreur de mesure et un aspect unique propre à la satisfaction spécifique qui est mesurée.

Les équations s'écrivent ainsi:

S'il y a un seul facteur et trois variables:

$x_1 = b_1F+U_1$
$x_2 = b_2F+U_2$
$x_3 = b_3F+U_3$

où: x_1 est la valeur pour la variable 1
b_1 est la valeur du coefficient de régression et donne l'importance de l'influence de F sur x_1
F est la valeur théorique du facteur F
et U_1 est la valeur de l'élément résiduel comprenant le facteur unique à la variable 1.

Similairement, s'il y avait n facteurs, l'équation pour chaque variable s'écrirait:

$x_i= b_{i1}F_1 + b_{i2}F_2+...b_{in}F_n + U_i$

Ainsi, dans le cas par exemple d'un questionnaire de sondage ou d'un test, on peut dire que la réponse que donne un individu à une question posée est conçue comme une combinaison linéaire 1) de la réponse théorique qu'il donnerait à une variable globale que l'on ne peut pas mesurer directement et 2) d'autres facteurs reliés entre autres à la question spécifique qui est posée.

Type de questions auxquelles l'analyse factorielle permet de répondre (Tabachnik et Fidell, 1989: 601)

- Combien de facteurs sont nécessaires pour donner une représentation juste et parcimonieuse des données?

- Quelle est la nature de ces facteurs, comment peut-on les interpréter?

- Quelle proportion de la variance des données peut être expliquée par un certain nombre de dimensions (facteurs) majeures?

- Jusqu'à quel point la solution factorielle est conforme à la théorie que je voulais vérifier?

- La structure factorielle est-elle la même pour divers groupes?

- (Quel score auraient obtenu les sujets si on avait pu mesurer les facteurs?)

Analyse en composantes principales et analyse factorielle:

- L'analyse en composantes principales (ACP) cherche une solution **à l'ensemble de la variance des variables mesurées**. De plus, elle cherche une solution où les composantes sont orthogonales (c'est-à-dire indépendantes entre elles)[1]. Quelque soit la matrice de corrélations, il y a toujours une solution en ACP. L'ACP maximise la variance expliquée.

- L'analyse factorielle (A.F.) cherche une solution à **la covariance entre les variables mesurées**. Elle tente d'expliquer seulement la variance qui est commune à au moins deux variables et présume que chaque variable possède aussi une variance unique représentant son apport propre. Les divers modes d'extraction visent à maximiser une bonne reproduction de la matrice de corrélations originale.

[1] Quoiqu'il soit possible de faire des rotations orthogonales ou obliques en ACP, cette utilisation ne respecte pas les bases mêmes de l'ACP, à savoir une solution unique et des composantes indépendantes entre elles qui expliquent chacune une proportion décroissante de la variance.

Analyse exploratoire et analyse confirmatoire:

L'analyse habituellement effectuée par les logiciels courants (SPSS, BMDP, SAS) est une analyse de type exploratoire puisqu'elle ne permet pas de déterminer à l'avance quelles variables devraient être liées à quels facteurs. Lorsque la solution factorielle proposée par le logiciel (la solution statistique) confirme nos hypothèses de départ, c'est bon signe. Lorsque ce n'est pas le cas, ceci n'infirme pas nécessairement nos hypothèses, ceci parce qu'une multitude de solutions sont possibles pour chaque analyse et que le logiciel ne peut en proposer qu'une seule, celle qui est la plus appropriée statistiquement. Une autre solution, plus conforme à nos hypothèses, peut être presque aussi bonne que la solution proposée et on ne peut pas le vérifier.

L'analyse factorielle confirmatoire permet de déterminer non seulement le nombre de facteurs mais aussi l'appartenance de chaque variable à un ou plusieurs facteurs. Ce type d'analyse doit être utilisé avec précaution, lorsque l'on est vraiment à l'étape finale de la confirmation d'un modèle. Elle nécessite l'utilisation de logiciels permettant de faire des analyses par équations structurales (EQS et LISREL, maintenant AMOS en module de SPSS version 7.0). LISREL VII était disponible comme module de SPSS (versions antérieures à la version 7.0); toutefois, la version VIII, une version Windows, est nettement plus conviviale.

2. Considérations théoriques et pratiques (*Tabachnik et Fidell, 1989: 603-605)*

- Pour qu'une variable soit intégrée dans l'analyse, sa distribution doit montrer une certaine variance i.e. elle doit discriminer les positions des individus.

- Idéalement, on cherche une ***structure simple***, c'est-à-dire une solution où chaque variable détermine fortement un et un seul facteur.

- Lorsqu'une variable est corrélée à plus d'un facteur, on dit que c'est une ***variable complexe***; on peut dire que la signification des réponses à cette variable s'interprète selon plusieurs dimensions.

- La structure factorielle peut être différente pour différentes populations. Il faut faire attention à ne pas regrouper pour l'analyse des populations trop différentes.

- Pour qu'une structure factorielle soit stable, elle doit avoir été vérifiée sur un minimum de cas. **La règle veut qu'il y ait <u>un minimum</u> de 5 cas par variable.** Lorsque cette règle n'est pas respectée, plusieurs problèmes peuvent survenir dont celui de la "matrice malade" (ill-conditioned matrix) ou le fait qu'une deuxième analyse avec une population différente donne des regroupements très différents. Il y a donc des problèmes de stabilité, de fidélité, de la solution factorielle.

- Les variables utilisées pour l'analyse devraient **se distribuer normalement**. Toutefois, lorsqu'on utilise l'analyse factorielle uniquement comme outil exploratoire, il est possible de "transgresser" cette règle. Il faut alors utiliser une procédure d'extraction (Moindres carrés non pondérés ou ULS en anglais) qui tient compte du fait que la distribution des variables n'est pas normale. Si le but de l'analyse est l'inférence, le postulat de normalité est plus important et certaines transformations normalisant la distribution peuvent être effectuées.

- La relation entre les paires de variables est présumée **linéaire**.

- On devrait idéalement repérer et éliminer les cas ayant des patrons de réponses "anormaux" (**Cas aberrants**)

- **La matrice de corrélation ne peut pas être singulière pour ce qui est de l'AF pure.** Ceci signifie que les variables ne peuvent pas être à ce point corrélées entre elles qu'une variable constitue une combinaison linéaire d'une ou plusieurs autres variables; il y a alors redondance, c'est-à-dire que la même information est inscrite à deux reprises. Mathématiquement, les produits de matrices nécessaires à l'estimation ne peuvent être effectués dans une telle situation.

- **La matrice de corrélation doit contenir un patron, une solution factorielle.** Certains ensembles de variables doivent être corrélés entre eux, suffisamment pour qu'on puisse dire qu'ils constituent un facteur. La solution factorielle doit aussi expliquer une proportion suffisamment intéressante de la variance pour que la réduction à un nombre restreint de facteurs ne se fasse pas au prix d'une perte importante d'information.

- **Toutes les variables doivent faire partie de la solution c'est-à-dire être corrélées minimalement avec une ou plusieurs variables, sinon elles constituent des cas aberrants et doivent par conséquent être retirées de l'analyse.**

3. Les types d'extraction d'une solution factorielle.

A) l'extraction de type ACP pour analyse en composantes principales (ou PC pour principal components en anglais)
Ce type d'extraction produit nécessairement une solution et la solution produite est unique. Il s'agit d'un solution maximisant la variance expliquée par les facteurs.

B) l'extraction pour l'analyse factorielle

Il y a plusieurs méthodes d'extraction. Il faut souligner que lorsque la solution factorielle est stable, les diverses méthodes donnent des résultats similaires, la plupart du temps identiques.
Les méthodes les plus utilisées sont:

- ML pour maximum de vraisemblance (ou maximum likelihood en anglais) : maximise la probabilité que la matrice de corrélation reflète une distribution dans la population. Cette méthode produit aussi un test de χ^2 de rapport de vraisemblance qui indique si la solution factorielle est plausible. **La probabilité de ce test doit être supérieure à .05, c'est-à-dire que l'on ne doit pas rejeter l'hypothèse nulle qui veut que le modèle soit compatible avec les données.**

La méthode est toutefois sensible aux déviations à la normalité des distributions: on rencontre fréquemment des problèmes lorsqu'on utilise cette méthode avec des échelles ordinales de type très, assez, peu, pas du tout.

- ULS pour moindres carrés non pondérés (ou unweighted least square en anglais) : minimise les résidus. Cette méthode est privilégiée lorsque les échelles de mesure sont ordinales ou que la distribution des variables n'est pas normale. Cette situation se présente fréquemment en sciences sociales, particulièrement lorsque l'on mesure des attitudes.

alpha-maximisation : Cette méthode est très peu utilisée et très peu connue. Elle s'avère pertinente lorsque le but de l'analyse est de créer des échelles puisqu'elle tente de maximiser l'homogénéité à l'intérieur de chaque facteur et donc, la fidélité.

4. Les types de rotation

La rotation est le processus mathématique qui permet de faciliter l'interprétation des facteurs en maximisant les saturations les plus fortes et en minimisant les plus faibles de sorte que chaque facteur apparaisse déterminé par un ensemble restreint et unique de variables. Ce processus est effectué par rotation, repositionnement des axes.

Deux types de rotations:

rotation orthogonale: On utilise cette rotation avec l'ACP[2] et avec l'analyse factorielle (AF) lorsque l'on croit qu'il est possible de déterminer des facteurs qui soient indépendants les uns des autres. Une solution orthogonale **est toujours préférable parce qu'une telle solution indique que chaque facteur apporte une information unique, non partagée par un autre facteur. Toutefois, ce type de solution est rarement possible en sciences sociales puisque habituellement, il existe des liens conceptuels entre les facteurs**. Il existe trois méthodes pour produire une rotation orthogonale; la plus fréquemment utilisée est **VARIMAX.**

rotation oblique: La rotation oblique, utilisée surtout avec l'A.F. puisqu'elle est conceptuellemnet plus appropriée dans ce cas, permet qu'il y ait corrélation entre les facteurs. Comme elle correspond habituellement mieux à la réalité, elle est fréquemment utilisée en sciences sociales. La méthode utilisée est **OBLIMIN.**

[2] voir note de bas de page 1 page 3

5. Les étapes de l'analyse factorielle de type exploratoire

a) déterminer l'ensemble des variables qui seront analysées conjointement.

b) idéalement, examiner cet ensemble de façon conceptuelle et déterminer la solution qui apparaîtrait plausible au plan du nombre de facteurs et du regroupement des variables.

c) - Effectuer une analyse en composantes principales avec rotation orthogonale (varimax) en laissant la procédure définir le nombre de facteurs par défaut;
- effectuer en même temps une analyse factorielle (lorsque c'est le but final de l'analyse) avec une rotation orthogonale et une rotation oblique (oblimin). **Le nombre de facteurs par défaut est déterminé par un critère, celui d'une valeur propre plus grande que 1.0. Le Graphique des valeurs propres donne aussi des indications quant au nombre de facteurs appropriés.**

d) Examiner cette analyse pour déterminer les éléments suivants:

- comparer la ou les solutions proposées avec l'hypothèse de regroupement faite au départ.

- Pour chacune des variables, décider du maintien dans les analyses subséquentes à partir des critères suivants:

1) voir si la qualité de la représentation ("communality" - A.F. statistiques initiales) est suffisamment bonne pour le maintien dans l'analyse.

2) voir si les variables appartiennent à un seul facteur ou à plusieurs. Une trop grande complexité d'une variable justifierait son retrait.

3) examiner parallèlement la pertinence des regroupements et la pertinence théorique de maintenir ou de retirer une variable plutôt qu'une autre.

4) examiner les divers indices de pertinence de la solution factorielle

e) refaire l'analyse de façon itérative jusqu'à arriver à une solution simple satisfaisante.

Le test d'une bonne analyse factorielle réside, en fin de compte, dans la signification des résultats. C'est le chercheur qui "décode" la signification conceptuelle de chaque facteur. Il faut donc pouvoir nommer chaque facteur.

Deux problèmes se posent:

1) Le critère de la justesse de l'analyse est en partie subjectif (Est-ce que le regroupement fait du sens?). Il faut faire particulièrement attention à la tendance qu'ont certains chercheurs à donner aux facteurs des noms qui font du sens et qui impressionnent mais qui ne reflètent pas ce qui a été mesuré.

2) Il y a une infinité de solutions possibles après rotation; il n'y a donc pas une seule "bonne" solution. Il est alors difficile de décréter que la solution présentée est "la solution". Il faut la présenter comme une solution plausible qui n'est pas contredite par les données.

6) Outils de diagnostic de la solution factorielle :

Si le principal critère d'un bonne solution factorielle demeure sa justesse au plan théorique, au plan du sens, ***il demeure que plusieurs outils statistiques nous guident dans la recherche de la meilleure solution possible.*** Voici une brève présentation des principaux outils diagnostiques utilisés. L'exemple présenté permettra d'en comprendre l'utilisation de façon plus poussée.

A) adéquation de la solution globale

a) Le déterminant de la matrice:

Un déterminant égal a zéro signifie qu'au moins une variable est une combinaison linéaire parfaite d'une ou de plusieurs autres variables. Il y a donc une variable qui ne rajoute aucune information nouvelle au-delà de celle fournie par les autres variables. Dans ce cas l'analyse ne peut procéder pour des raisons mathématiques (Il est impossible d'inverser la matrice). *Notons que nous recherchons un **déterminant très petit**, ce qui constitue un bon indice de l'existence de patrons de corrélations entre les variables, **mais non égal à zéro**.*

On obtient le déterminant
- en indiquant DéTERMINANT dans la fenêtre CARACTéRISTIQUES (SPSS Windows)[3].
- en indiquant DET dans la sous-procédure /PRINT (SPSS syntaxe)

[3] La mention SPSS Windows réfère à l'item ANALYSE FACTORIELLE du menu déroulant ANALYSE-FACTORISATION La mention SPSS syntaxe réfère à la version UNIX ou PC ainsi qu 'à la syntaxe que l'on peut éditer soi-même dans la version Windows.

b) La mesure de Kaiser-Meyer-Olkin

Plus communément appelé le KMO, la mesure de Kaiser-Meyer-Olkin est un indice d'adéquation de la solution factorielle. Il indique jusqu'à quel point l'ensemble de variables retenu est un ensemble cohérent et permet de constituer une ou des mesures adéquates de concepts. Un KMO élevé indique qu'il existe une solution factorielle statistiquement acceptable qui représente les relations entre les variables.

Une valeur de KMO de moins de .5 est inacceptable
.5 est misérable
.6 est médiocre
.7 est moyenne
.8 est méritoire
.9 est merveilleuse (ref: SPSS professional statistics)

Le KMO reflète le rapport entre d'une part les corrélations entre les variables et d'autre part, les corrélations partielles, celles-ci reflétant l'unicité de l'apport de chaque variable.

On obtient le KMO
- en indiquant Indice KMO et test de Bartlett dans la fenêtre CARACTéRISTIQUES (SPSS Windows).
- en indiquant KMO dans la sous-procédure /PRINT (SPSS syntaxe)

c) Le test de sphéricité de Bartlett:

Ce test vérifie l'hypothèse nulle selon laquelle toutes les corrélations seraient égales à zéro. On doit donc tenter de rejeter l'hypothèse nulle i.e. que le test doit être significatif (la probabilité d'obtenir la valeur du test doit être plus petite que .05). Toutefois le test est très sensible au nombre de cas; il est presque toujours significatif lorsque le nombre de cas est grand. Ses résultats sont donc intéressants presque uniquement lorsqu'il y a moins de 5 cas par variable.

On obtient le test de sphéricité automatiquement
- avec l'indication Indice KMO et test de Bartlett dans la fenêtre CARACTéRISTIQUES (SPSS Windows).
- avec l'indication KMO dans la sous-procédure /PRINT (SPSS syntaxe)

d) Le test du coude de Cattell

Le Graphique des valeurs propres donne une représentation graphique des informations sur les valeurs propres de chaque facteur présentées dans le tableau des statistiques initiales. Dans cette représentation, il faut rechercher le point (parfois les points) de cassure qui représente le nombre de facteurs au-delà duquel l'information ajoutée est peu pertinente. *Plus la courbe est accentuée, plus il apparaît qu'un petit nombre de facteurs explique la majeure partie de la variance. A partir du moment où la courbe devient presque une ligne droite horizontale, il apparaît que les facteurs subséquents apportent peu de nouvelles informations.*

Note : Les valeurs propres représentent la variance expliquée par chaque facteur. Elles sont constituées de la somme des poids factoriels au carré de toutes les variables pour un facteur déterminé.

On obtient cette représentation
- en indiquant Graphique des valeurs propres dans l'item "afficher" de la fenêtre EXTRACTION (SPSS Windows).
- en indiquant EIGEN dans la sous-procédure /PLOT (SPSS syntaxe)

e) La matrice reconstituée et les résidus

L'analyse en composantes principales, tout comme l'analyse factorielle, constitue une décomposition de la matrice des corrélations entre les variables.

Ainsi, si l'on effectue une analyse en composantes principales et que l'on demande autant de facteurs qu'il y a de variables dans l'analyse, la matrice de corrélation reconstituée sera identique à la matrice de corrélation initiale.

Ce n'est pas le cas pour l'analyse factorielle puisque celle-ci tente d'expliquer non pas la variance totale mais uniquement la covariance entre les variables. Donc, même avec autant de facteurs que de variables, **la matrice qui est créée lorsque l'on tente l'opération inverse,** c'est-à-dire de reproduire les corrélations d'origine à partir des informations extraites des facteurs suite à l'analyse, **ne reproduira pas les corrélations originales à la perfection, il restera des résidus. Il restera d'autant plus de résidus que l'on garde seulement une partie des facteurs.**

Donc, plus la solution factorielle est bonne, plus la matrice reconstituée s'approche de la matrice de corrélation initiale et moins les résidus sont importants. L'indication d'une proportion faible de résidus plus grands que .05 signifie que la solution est appropriée. Ces résidus devraient quand même être examinés pour voir s'il n'y a pas des cas aberrants.

On obtient la matrice reproduite et la matrice des résidus
- en demandant RECONSTITUéE dans la fenêtre CARACTéRISTIQUES (SPSS Windows).
- en indiquant REPR dans la sous-procédure /PRINT (SPSS syntaxe)

f) La structure obtenue

La structure obtenue, c'est-à-dire le tableau des corrélations entre les variables et les facteurs (Matrice des composantes en rotation orthogonale et Matrice des types en rotation oblique), doit être ***simple***, **ce qui veut dire que chaque variable doit avoir une corrélation plus grande que .3 avec au moins un facteur et avec un seul facteur.**

Ces matrices sont imprimées automatiquement
- en indiquant Structure après rotation dans la fenêtre ROTATION (SPSS Windows). Pour avoir ces mêmes tableaux avant rotation, on indique Structure factorielle sans rotation dans la fenêtre EXTRACTION.
- en indiquant DEFAULT dans la sous-procédure /PRINT (SPSS syntaxe)

B) Pertinence de garder une variable dans la solution.

a) Les statistiques descriptives des variables

Les variables, par définition, doivent montrer une certaine variation du positionnement des individus quant à ce qui est mesuré. En ce sens, un écart-type important et une moyenne qui se rapproche du milieu de l'échelle de mesure (exemple: moyenne de 2.5 pour une échelle à 4 catégories) sont de bons indices que la variable apporte une information susceptible d'aider à différencier les individus.

Ces statistiques sont produites
- en indiquant Caractéristiques univariées dans la fenêtre CARACTéRISTIQUES (SPSS Windows).
- en indiquant UNIVARIATE dans la sous-procédure /PRINT (SPSS syntaxe)

b) La qualité de la représentation de chaque variable avec la solution factorielle initiale

On doit examiner les statistiques de qualité de la représentation pour l'analyse factorielle proprement dite (et non l'ACP) et ceci avant l'extraction d'un nombre restreint de facteurs. La qualité de la représentation indique alors l'appartenance de chaque variable à la covariance de l'ensemble des variables. ***C'est la variance de chaque variable qui peut être expliquée par l'ensemble des autres variables.*** On considère que la qualité de la représentation doit être minimalement de .20 pour justifier le maintien de la variable dans l'analyse.

Par ailleurs, quoique ce soit un indice différent, il pourrait arriver qu'une variable ait une bonne qualité de la représentation avec la solution initiale mais non avec la solution après extraction d'un nombre restreint de facteurs. Ceci se refléterait probablement par le fait que cette variable ne se regrouperait pas avec les autres dans la solution factorielle; elle déterminerait entièrement un facteur.

NOTE: La somme des poids factoriels au carré pour une variable donnée égale la qualité de la représentation de cette variable.

On obtient ces informations
- en indiquant Structure initiale dans la fenêtre CARACTéRISTIQUES (SPSS Windows).
- via DEFAULT dans la sous-procédure /PRINT (SPSS syntaxe)

c) La simplicité ou la complexité de chaque variable dans la solution factorielle finale

Une variable est dite complexe lorsqu'elle est corrélée substantiellement (saturation factorielle plus grande que 0,30 à plus d'un facteur). On peut dire que les réponses à cette variable reflètent plus d'un concept. Ainsi, il pourrait arriver que la satisfaction face aux avantages sociaux soit corrélée à deux facteurs, un portant sur la rémunération (associé au salaire) et l'autre portant sur la sécurité d'emploi. Cela entraîne un problème lorsque l'on veut créer des échelles : Avec quelle échelle devrait-on regrouper cette variable, celle portant sur la rémunération ou celle portant sur la sécurité?

Il y a plusieurs manières de traiter ce problème en fonction de l'importance théorique de la variable, du choix quant au nombre de facteurs, de la rotation (orthogonale ou oblique) privilégiée. Dans le cas où d'autres variables amènent une information similaire, on peut retirer la variable considérée comme complexe. Il arrive qu'on la maintienne dans l'analyse; lorsque l'on veut créer les échelles, on décide de son appartenance à une échelle plutôt qu'une autre à partir des informations fournies par les analyses de fidélité et à partir de considérations théoriques.

Ces informations sont tirées des matrices factorielles (Matrice des composantes ou Matrice des types) que l'on obtient
- en indiquant Structure après rotation dans la fenêtre ROTATION (SPSS Windows).
- en indiquant DEFAULT dans la sous-procédure /PRINT (SPSS syntaxe)

d) Un cas spécial: le Heywood case

On parle de cas Heywood lorsque, dû aux relations entre les variables, il y a un problème dans les calculs et la qualité de la représentation devient plus grande que 1.0. On repère un cas Heywood case
- lorsque la qualité de la représentation d'une variable est notée comme étant .9998 ou .9999.
- lorsque la saturation factorielle de la variable est plus grande que 1.0. (SPSS émet un avertissement disant que la qualité de la représentation d'une ou de plusieurs variables est plus grande que 1.0).

Le cas Heywood est dû au fait que la variable est trop fortement corrélée à une ou plusieurs autres variables. Dans ce cas, il faut décider soit de retirer la variable des analyses subséquentes soit de retirer une autre variable avec laquelle elle est fortement corrélée.

En résumé, voici les commandes pour l'analyse factorielle avec Spss -Windows:
→ Allez dans ANALYSE,
→ choisir FACTORISATION - ANALYSE FACTORIELLE
Dans le tableau principal de l'analyse factorielle

a) Choisir les VARIABLES que l'on veut analyser

b) Dans CARACTÉRISTIQUES :
- STATISTIQUES → CARACTÉRISTIQUES UNIVARIÉES
→ STRUCTURE INITIALE
- MATRICE DE CORRÉLATIONS
→ COEFFICIENTS
→ DÉTERMINANT
→ INDICE KMO ET TEST DE BARTLETT
→ MATRICE DES CORRÉLATION RECONSTITUÉE

c) Dans EXTRACTION :
- METHODE[4] →COMPOSANTES PRINCIPALES (PC)
→MOINDRES CARRÉS NON PONDÉRÉS (ULS)
→ALPHA-MAXIMISATION
→...
- EXTRAIRE →VALEURS PROPRES SUPÉRIEURES À 1.0
→NOMBRE DE FACTEURS =
→MAXIMUM DES ITÉRATIONS POUR CONVERGER (pour augmenter le nombre d'itérations nécessaires à la convergence).
- AFFICHER →STRUCTURE FACTORIELLE SANS ROTATION
→GRAPHIQUE DES VALEURS PROPRES

d) Dans ROTATION:
- MÉTHODE →VARIMAX
→OBLIMIN DIRECTE
- DISPLAY →STRUCTURE APRÈS ROTATION
→CARTE(S) FACTORIELLE(S) (par défaut: 3-D pour 3 premiers facteurs; si on veut un graphique des facteurs deux par deux, il faut éditer la syntaxe en rajoutant une parenthèse (n1 n2) après l'indication PLOT)

e) Dans OPTIONS:
- VALEURS MANQUANTES
→ EXCLURE TOUTE OBSERVATION INCOMPLÈTE OU EXCLURE SEULEMENT LES COMPOSANTES NON VALIDES OU REMPLACER PAR LA MOYENNE
- AFFICHAGE DES PROJECTIONS
→CLASSEMENT DES VARIABLES PAR TAILLE
→SUPPRIMER LES VALEURS ABSOLUES INFÉRIEURES À (.30)

[4] Nota bene : On ne peut pas choisir plus d'une méthode à la fois. C'est la même chose pour les rotations. Il est habituellement préférable d'éditer le fichier de syntaxe de façon à pouvoir faire plusieurs analyses à la fois, sinon il faut éditer le fichier Résultats pour ne pas faire imprimer plusieurs fois les mêmes informations.

7) Commandes pour l'analyse factorielle avec SPSS

En résumé, une commande d'analyse factorielle dans SPSS qui donnerait l'ensemble des informations nécessaires présentées plus haut aurait la forme suivante:

COMMANDE, SOUS-COMMANDE	SIGNIFICATION
FACTOR /VAR=NATURE3 TO RECON8 CARRIER1 STABIL2 PERFECT9 SECUR13 SALAIR10 TO HORAIR12 ** Dans Windows, sélectionner les variables*	FACTOR: demande une analyse factorielle /VAR= liste des variables qui seront analysées
/PRINT DEFAULT UNIVARIATE CORRELATION REPR DET KMO **Dans Windows, choisir Structure initiale, Caractéristiques univariées, Coefficients, Reconstituée, Déterminant et Indice KMO et test de Bartlett dans CARACTÉRISTIQUES*	/PRINT indique les informations qui devront être imprimées, dans ce cas-ci, les statistiques par défaut, les statistiques univariées pour chaque variable, la matrice de corrélation originale, la matrice reproduite, le déterminant, le KMO et le test de sphéricité.
/CRITERIA FACTORS (4) ** Dans Windows, cette option est disponible dans la fenêtre EXTRACTION*	/CRITERIA permet de spécifier, *lorsque nécessaire,* des critères tels le nombre de facteurs et le nombre d'itérations; dans ce cas-ci, on demande 4 facteurs.
/PLOT EIGEN *ROTATION (1,2) ** Dans Windows, on demande le Graphique des valeurs propres (eq. de PLOT EIGEN) dans la fenêtre EXTRACTION et Carte factorielle dans la fenêtre ROTATION (eq. de ROTATION). Par défaut , cette dernière indication donne un graphique en 3 dimensions des 3 premiers facteurs, le cas échéant. Pour obtenir des graphiques en deux dimensions , il faut éditer la syntaxe.*	/PLOT permet de demander des graphiques des valeurs de eigen ou des facteurs; dans ce cas-ci on demande le graphique des valeurs de eigen et celui des deux premiers facteurs
/FORMAT SORT BLANK (.3) ** Dans Windows, la fenêtre OPTIONS donne l'Affichage des projections. On clique Classement des variables par taille et Supprimer les valeurs absolues inférieures à __. On change la valeur par défaut (.10) à .30.*	/FORMAT contrôle l'apparence; dans ce cas-ci SORT permet que les variables apparaissent dans les matrices factorielles en fonction de leur importance et selon les facteurs et BLANK (.3) permet que les saturations factorielles inférieures à .3 n'apparaissent pas dans l'impression ce qui facilite la lecture.

/EXTRACTION PC /ROTATION VARIMAX /EXTRACTION ULS /ROTATION VARIMAX /ROTATION OBLIMIN. **Dans SPSS Windows, il faut, pour chaque extraction ou rotation, refaire la commande d'Analyse facorielle au complet (i.e. choisir le mode d'extraction approprié dans EXTRACTION et la rotation désirée dans ROTATION). Il est nettement préférable d'éditer la commande pour demander les extractions et rotations pour un ensemble de variables déterminé en une seule commande. Attention, c'est dans la fenêtre EXTRACTION que l'on demande le Graphique des valeurs propres (Scree plot) ainsi que la solution factorielle sans rotation et que l'on définit les critères (nombre de facteurs et/ou d'itérations) le cas échéant. Dans ROTATION, on demande la Structure après rotation et les cartes factorielles (loading plots).*	EXTRACTION spécifie le type d'extraction et ROTATION le type de rotation. Dans ce cas-ci, on demande une première analyse avec une extraction PC (en composantes principales) avec une rotation orthogonale et une deuxième analyse avec extraction ULS (moindres carrés non pondérés) comprenant une rotation orthogonale et une rotation oblique.

8) Qu'est ce que l'analyse de fidélité?

Blalock (1968):

"Les sociologues théoriciens utilisent souvent des concepts qui sont formulés à un assez haut niveau d'abstraction. Ce sont des concepts relativement différents des variables utilisées qui sont le lot des sociologues empiriques... Le problème du lien entre la théorie et la recherche peut donc être vu comme une question d'erreur de mesure".

La mesure peut être vue comme le "processus permettant de lier les concepts abstraits aux indicateurs empiriques" (Carmines et Zeller, 1979).

Deux concepts clés: La fidélité et la validité

Fidélité: *Consistance dans la mesure :* Jusqu'à quel point plusieurs mesures prises avec le même instrument donneront les mêmes résultats dans les mêmes circonstances.

Exemple: Je fais passer un questionnaire portant sur l'idéologie deux fois aux mêmes personnes à deux mois d'intervalle et j'obtiens des résultats différents entre les deux passations. Est-ce que l'idéologie d'une personne peut changer si vite ou si c'est l'instrument qui n'est pas fiable?

La fidélité demeure au plan empirique: elle dit si en soi l'instrument est un bon instrument.

Validité: *Jusqu'à quel point l'instrument mesure ce qu'il est supposé mesurer.*

Exemple: Si j'utilise une série de questions censées mesurer les préférences idéologiques et que je me rends compte que j'ai en fait mesuré l'identification à un parti politique, ma mesure est une mesure non valide de l'idéologie.

La validité concerne la relation entre la théorie et les concepts qui lui sont reliés d'une part et la mesure d'autre part: elle est concernée par l'adéquation de la traduction du concept en mesure.

Plusieurs types de validité:

validité reliée au critère*:** relation entre l'instrument et ce à quoi il devrait théoriquement être relié. On parle de *validité* ***prédictive quand le critère est mesuré après et de *validité* ***concurrente*** quand le critère est mesuré en même temps.

validité de contenu: relation entre le /les concepts à mesurer et l'instrument utilisé. Un instrument de mesure de l'aliénation mesurera le sentiment d'absence de pouvoir, d'absence de normes, d'isolation sociale, etc. Ceci implique que le domaine et les concepts doivent être bien définis.

validité de construit: relation entre l'instrument et d'autres instruments supposé mesurer des concepts reliés.

validité convergente et discriminante*:* voir analyse factorielle: Jusqu'à quel point chaque indicateur constitue une mesure d'un et d'un seul concept.

La fidélité:

→ théorie classique des tests:

→ le score observé (qui peut-être par exemple la réponse d'une personne à une question) est une combinaison linéaire d'une partie représentant le score vrai et d'une partie d'erreur aléatoire.

X=t+e
où "t" représente le score vrai
et "e" représente l'erreur aléatoire

→ la corrélation entre deux scores observés constitue un estimé de la fidélité des mesures → Dès qu'on a plus d'une mesure d'un même concept, on peut estimer la fidélité ρ ; on peut concevoir l'erreur moyenne de mesure comme la réciproque de la fidélité, "1-ρ ". Dans la théorie classique des tests, ceci est vrai si les scores vrais et la variance d'erreur sont identiques (i.e. les mesures sont parallèles).

Diverses manières de mesurer la fidélité:

fidélité test-retest: corrélation entre la mesure prise à un temps 1 et la mesure prise à un temps 2: fidélité dans le temps.

fidélité "split-half" entre deux sous-ensembles: Jusqu'à quel point deux sous-ensembles des items constituent deux mesures fidèles du même concept.

fidélité entre différentes formes: Jusqu'à quel point deux ensembles différents d'items peuvent mesurer le même concept.

consistance interne: Jusqu'à quel point chacun des items constitue une mesure équivalente d'un même concept.

On peut mesurer la consistance interne en utilisant le

→ *Alpha de Cronbach:*

$$\alpha = \frac{k * \bar{cov}/\bar{var}}{1 + (k-1)\bar{cov}/\bar{var}}$$

où "k" est le nombre d'items
cōv est la covariance moyenne entre les items
et vār est la variance moyenne des items

Si les items sont standardisés de façon à avoir la même variance, la formule se modifie comme suit:

$$\alpha = \frac{k * \bar{\rho}}{[1 + \bar{\rho}(k-1)]}$$

où "k" représente le nombre d'items dans l'échelle
et "ρ̄" est la corrélation moyenne

Ce coefficient Alpha peut être considéré comme la moyenne des coefficients alpha que l'on obtiendrait pour toutes les combinaisons possibles de deux sous-ensembles des items mesurant un même concept. Il peut aussi être vu comme l'estimé de la corrélation que l'on obtiendrait entre un test et une forme alternative du même test comprenant le même nombre d'items.

Le coefficient alpha est la borne inférieure de la fidélité réelle i.e la fidélité réelle ne peut pas être inférieure à la valeur du alpha et elle est égale à cette valeur lorsque les items sont parallèles i.e. les scores vrais ont la même moyenne et la variance d'erreur est la même.

Remarquer que la valeur de alpha augmente avec le nombre d'items, mais ce à la condition que la corrélation moyenne inter-item ne soit pas diminuée avec l'ajout de nouveaux items (i.e. toutes choses égales par ailleurs). L'amélioration du alpha devient marginale au-delà d'un certain nombre d'items (environ 6-7).

9) L'analyse de fidélité : aspects pratiques et outils diagnostiques

Plusieurs outils sont disponibles pour évaluer la fidélité d'un ensemble de variables. La procédure POSITIONNEMENT-ANALYSE DE FIABILITé de SPSS permet d'examiner les informations pertinentes:

Matrice de corrélation: Tout comme avec l'analyse factorielle, cette matrice permet de voir jusqu'à quel point les items sont corrélés entre eux et quels items sont plus fortement corrélés. S'il s'avérait que deux concepts sont mesurés plutôt qu'un seul, les corrélations pourraient nous permettre de repérer cette possibilité.

→ Dans SPSS Windows, on indique Corrélations dans COHéRENCE INTER-ITEMS
→ Dans SPSS syntaxe, on indique la sous-commande CORR dans /STATISTICS

Statistiques univariées:
Pour chacun des items, on peut obtenir la moyenne et l'écart type, ce qui permet de voir si les statistiques descriptives sont similaires pour les divers items,

→ Dans SPSS Windows, on indique Item dans CARACTéRISTIQUES
→ Dans SPSS syntaxe, on indique la sous-commande DESCRIPTIVE dans /STATISTICS

Les statistiques d'échelle: indiquent quelle serait la moyenne, la variance et l'écart-type de l'échelle si on additionnait les réponses à chacun des items. Donne une idée des propriétés futures de l'échelle à créer → la variance sera-t-elle suffisante?

→ Dans SPSS Windows, on indique Echelle dans CARACTéRISTIQUES
→ Dans SPSS syntaxe, on indique la sous-commande SCALE dans /STATISTICS

Le sommaire des statistiques d'items: les moyennes: Donnent les indications sur les différences de moyennes entre les items i.e minimum, maximum, moyenne, ratio maximum/minimum. Si ces différences sont trop importantes, on pourrait penser que chaque item ne mesure pas le concept de façon équivalente.

→ Dans SPSS Windows, on indique Moyennes dans PRINCIPALES STATISTIQUES
→ Dans SPSS syntaxe, on indique la sous-commande MEANS dans /SUMMARY

Le sommaire des statistiques d'items: les variances: Donnent les indications sur les différences de variances entre les items. Même interprétation que pour les moyennes.

→ Dans SPSS Windows, on indique Variances dans PRINCIPALES STATISTIQUES
→ Dans SPSS syntaxe, on indique la sous-commande VARIANCES dans /SUMMARY

Les statistiques d'items: les corrélations inter-items: Donnent les indications **très importantes** sur les différences de corrélations entre les items. Des corrélations très faibles ou négatives devraient être repérées dans la matrice de corrélation et la pertinence de garder des items montrant une corrélation négative avec un ou plusieurs items évaluée. Si cette situation existe, soit que plus d'un concept est mesuré, soit que certains items mesurent mal le concept ou que l'échelle aurait due être inversée pour cet item (Par exemple, dans le cas où des items sont formulés négativement et d'autres positivement).

→ Dans SPSS Windows, on indique Corrélations dans PRINCIPALES STATISTIQUES
→ Dans SPSS syntaxe, on indique la sous-commande CORR dans /SUMMARY

Les statistiques de la relation entre chaque item et le total (échelle):

Moyenne, variance qui résulterait si on enlevait un item : La moyenne et la variance de l'échelle ne devraient pas être fortement modifiés par le retrait d'un item sinon, ceci signifie que l'item en question ne contribue pas de la même manière à la mesure du concept que les autres items.

Corrélation corrigée: Il s'agit de la corrélation entre l'item et les autres items de l'échelle (i.e l'échelle moins l'item en question). Cette corrélation devrait être minimalement de .30 (comme dans le cas de la corrélation entre chaque item et le facteur dans l'analyse factorielle).

Corrélation multiple au carré: Il s'agit de la variance de l'item expliquée par les autres items. Plus elle est élevée, plus l'item est une mesure commune du concept. Il s'agit de l'équivalent de la "communauté" dans les statistiques initiales en analyse factorielle.

alpha si l'item était enlevé: Il est très important de regarder cette information. Si un item n'appartient pas à l'échelle, alors la fidélité telle que mesurée par le alpha **serait supérieure si l'item était enlevé**. Donc, lorsque le "alpha if item deleted" est supérieur au alpha standardisé, ceci signifie que l'item en question est une mesure qui détériore la fidélité de l'échelle.

Toutes les statistiques précédentes sont obtenues
→ Dans SPSS Windows, on indique Echelle sans l'item dans CARACTéRISTIQUES
→ Dans SPSS syntaxe, on indique la sous-commande SCALE dans /STATISTICS

ALPHA ET ALPHA STANDARDISÉ: Ces deux mesures sont similaires lorsque les moyennes et les variances des items diffèrent peu. L'évaluation d'un bon alpha est similaire à celle du KMO: On recherche une valeur supérieure à .70, une valeur supérieure à .90 étant magnifique.

10) L'analyse de fidélité avec SPSS

En résumé, voici les commandes pour l'analyse de fidélité avec Spss -Windows:
→ Allez dans ANALYSE,
→choisir POSITIONNEMENT- ANALYSE DE FIABILITÉ

Dans le tableau principal de l'analyse de fidélité
a) Choisir les VARIABLES (items) que l'on veut analyser

b) Cliquer sur LISTER LES ÉTIQUETTES D'ITEM

c) Dans MODÈLE : - choisir le type d'analyse → ALPHA
- autres choix (split-half, Guttman, parallel, strict parallel)

d) Dans STATISTIQUES : - CARACTÉRISTIQUES POUR
→ ITEM
→ ÉCHELLE
→ ÉCHELLE SANS L'ITEM
- PRINCIPALES STATISTIQUES
→ MOYENNES
→ VARIANCES
→ CORRÉLATIONS
→ COVARIANCES
- COHÉRENCE INTER-ITEM
→ CORRÉLATIONS
→ COVARIANCES
- TABLEAU ANOVA
→ AUCUN
→ (TEST F - teste qu'il n'y a pas de différence significative entre les différentes mesures de l'échelle)

Autres informations disponibles : - T carré de Hotelling
- Test d'additivité de Tukey

En résumé, la syntaxe:

```
RELIABILITY
  /VARIABLES=q65c q65d q65e q65f q65g q65h
  /FORMAT=LABELS
  /SCALE(ALPHA)=ALL
  /MODEL=ALPHA
  /STATISTICS=DESCRIPTIVE SCALE CORR
  /SUMMARY=TOTAL MEANS VARIANCE CORR .
```

11) Notes additionnelles

a) Relation entre l'analyse factorielle, l'analyse en composantes principales et l'analyse de fidélité.

Le but premier de ces analyses est d'en arriver à regrouper ensemble les items qui mesurent le même concept de façon à ce qu'une addition des réponses à un ensemble d'items constitue une nouvelle mesure, composite, d'un concept. Par exemple, si on additionne les réponses de chaque répondant à chacun des items mesurant la satisfaction envers un aspect extrinsèque de son travail, on obtiendra pour chaque répondant une mesure de satisfaction extrinsèque.

L'analyse en composantes principales (ACP) décompose la matrice de corrélation en tenant compte de l'ensemble de la variance des items. Elle en extrait un certain nombre de facteurs indépendants. *Le but de l'analyse en composantes principales est d'expliquer le plus de variance possible avec un nombre de composantes le plus restreint possible.* Après extraction, une part seulement de la variance totale est expliquée. Le mode d'extraction et de rotation permet de déterminer les sous-ensembles d'items qui sont plus fortement corrélés entre eux et qui peuvent donc constituer des mesures d'un nombre restreint de concepts.

L'analyse factorielle (AF) fait la même chose que l'ACP mais tient compte uniquement de la variance commune à l'ensemble des items. Elle extrait des facteurs qui peuvent être indépendants ou corrélés entre eux. *Son but est de reproduire le plus fidèlement possible la matrice de corrélation.* Comme l'ACP, elle permet de déterminer des sous-ensembles plus fortement corrélés entre eux.

Comme l'ACP et l'AF ne retiennent qu'une partie de la variance totale dans la solution finale, les résultats de l'analyse de fidélité peuvent contredire en partie ceux de l'analyse factorielle ou de l'ACP. On peut expliquer cette situation par le fait que la variance commune d'un item est bien reliée à celle d'un autre item mais que leurs variances spécifiques sont peu ou pas du tout corrélées ou même en corrélation négative. Comme l'analyse de fidélité considère l'ensemble de la variance, cette situation peut faire qu'un item bien identifié à un facteur en AF se révèle un mauvais contributeur à l'échelle.

b) Attention aux valeurs manquantes:

Comme il est possible que certains répondants n'aient pas répondu à tous les items d'un ensemble donné, le nombre de cas valides peut varier d'une analyse à l'autre selon que l'on retire ou que l'on ajoute un item. **Ceci peut modifier les résultats**. Il existe dans certaines procédures des moyens d'estimer les valeurs manquantes, entre autres en remplaçant la valeur manquante par la moyenne du groupe. **Comme règle générale, les valeurs manquantes ne sont pas estimées: les cas qui n'ont pas répondu à toutes les questions n'apparaissent pas dans l'analyse**. Nous n'avons pas le temps d'aborder toute cette problématique durant le cours (Il existe dans BMDP, par exemple, 15 manières différentes d'estimer les valeurs manquantes, chacune ayant son biais propre) mais il demeure qu'il faut examiner les résultats et la modification du nombre de cas valides selon les analyses.

Bibliographie

Tabachnik,Barbara G.; Fidell,Linda S. (1996): Using multivariate statistics. Harper and Row, New York. 509 pages.

Norusis,M. J. (1993): Spss professional statistics. SPSS, Chicago.

Norusis,M. J. (1993): Spss advanced statistics. SPSS, Chicago.

Analyse factorielle et fidélité:

Carmines,Edward G.; Zeller,Richard A. (1979): Reliabiblity and validity assessment. (Quantitative applications in the social sciences, 17.) Sage, Beversly Hills. 71 pages.

Kim,Jae-On; Mueller,Charles (1978): Introduction to factor analysis. What it is and how to do it. (Quantitative applications in the social sciences.) Sage, Beverly Hills.

Kim,Jae-On; Mueller,Charles (1978): Factor analysis, statistical methods and practical issues. (Quantitative applications in the social sciences.) Sage, Beverly Hills.

Factor analysis and Principal components analysis, what's the difference?

by Lauri Tarkkonen

Factor analysis is based on the model

x = Bf+e, and cov(x)=Acov(f)A'+cov(e), one can make some further assumptions say cov(f)=I and cov(e)=D (diagonal), so the covariance is simplified to cov(x) = BB'+D = S.

Principal components are by definition:

u = A'x, such as V(u(i)) = a(i)'x is maximum a'a = 1 and the u(i) are orthogonal. Then cov(u) = A'cov(x)A, A'A and AA' are I, thus cov(x) = S = AA' . The factor analysis is a statistical model and the principal componentsare just a linear transformation of the original variables. Anyway, in factor analysis S-D = BB' and in principal components S = AA'.The difference in B and A is that you remove the D, the variances of the errors in FA but not in PCA. If there is no errors (it is easy to show that the same applies if all error variances are equal) the two methods will give you the same results.

If there is significant differences in the communalities of the variables, the two methods differ. So what is the proper one. If you do not assume measurement errors then use PCA; if you think there are measurement errors use FA.

I would like to put it this way:

> You all know the thing called the blueberry pie. First you take some dough, make a bottom of it, go to the forrest and pick some blueberries. If you believe in PCA, you put everything in your basket on top of the how. If you believe that there is something in your basket like leaves, needles, frogs that do not belong to the blueberry pie, you blow strongly while you pour your berries to the pie. Most things that do not belong to the pie fly away, you might loose some berries as well, but the taste is perhaps more of blueberries. If you did good job in picking your berries, there is no difference.

So why is this so difficult? Why do we always have this discussion?

First: Tradition.

In the beginning, there was T.W Anderson and his Introduction to Multivariate Analysis (-58). It started with PCA and told us that in the appendix (was it F) there is a funny thing called the Factor analysis. All the statisticians thought that PCA is the proper method and FA is some magic developed by some psychologists and you should not take it seriously.

Remember FA had this rotational indeterminacy.

Then: Lawley come up with the Maximum Likelihood solution. Now the statisticians had to accept FA as a bona fide statistical model. If there was Maximum Likelihood estimates for something, it must be the real thing. They realized that there was no theoretical base but the simplicity and the lack of indeterminacy speaking for the PCA.

More confusion:

Because both solutions were derived by the same solution of the eigenvalue problem, the spectral decomposition of a symmetric matrix, both analysis was performed with the same computer program. You just told it: do you have the communality estimation or not. Some programmers, like the maker of SYSTAT did not even understand the difference. Because the default value varied, so the naive users kept on getting whatever the programmers had to think as the main method.

Still more confusion:

The calculation of the factor solution has been two stage. Earlier the first stage was calculated by the 'principal axes' method. Some people do not see the difference with 'principal axes' and 'principal components'. They might have factors but they claim they have principal components.

Rotation

If you bother to stick to the definition of 'principal components' you will not rotate them, because if you rotate, the maximum variance part is not true anymore.

More tradition

In some areas the use of multivariate methods was started during the time when the statistician felt that PCA is the method and FA is some trick, there they swear still by PCA, because all the articles have PCA.

GREED

Some like PCA better because it gives sometimes higher loadings, but for some reason they still want to remove leaves and frogs from their blueberry pie.

Hope this helps.

- Lauri Tarkkonen

Lauri Tarkkonen / email: lauri.tarkkonen@helsinki.fi Tel:+358 0 666108
Korkeavuorenkatu 2 b B 11, 00140, Helsinki, Finland FAX +358 0 1913379