



Esprit et enjeux de l'analyse factorielle exploratoire

André Achim ^a  

^aDépartement de Psychologie, Université du Québec à Montréal

Abstract ■ L'analyse factorielle exploratoire est présentée en considérant les facteurs comme des sources communes de variance exprimées dans les données et responsables des corrélations observées. Les variables ayant généralement chacune une part de variance propre, il s'y trouve ainsi plus de sources d'information qu'il n'y a de variables observées. Alors que l'analyse en composantes principales inclut nécessairement une part des variances propres (souvent de l'erreur de mesure) dans chaque composante, l'analyse factorielle exploratoire vise à expurger les variances propres pour n'expliquer que les corrélations, c'est-à-dire l'information partagée provenant de facteurs communs. Il existe maintenant une méthode de modélisation pour identifier par un test formel combien de dimensions sont nécessaires et suffisantes pour rendre compte des corrélations dans les données. L'interprétation factorielle de ces dimensions demande généralement des rotations. Les enjeux de ces rotations et de l'obtention de scores des individus sur les facteurs retenus sont également discutés. Enfin, on propose une méthode simple pour l'analyse factorielle de mesures répétées. ■ Exploratory factor analysis is presented by viewing factors as sources of variance expressed in the data that are completely responsible for the correlations observed among the variables. Since each variable also expresses some unique variance besides that from common factors, the data express more sources of information than there are variables. While principal component analysis necessarily includes some of the unique variances (often measurement error) in each component, exploratory common factor analysis concentrates on explaining the correlations, which entirely depend on shared information from common factors, ignoring as much as possible the unique sources of variance. There is now a modelling method to identify by formal testing how many dimensions are necessary and sufficient to account for the data correlations. The factorial interpretation of these dimensions generally requires rotations. The issues involved in rotations and in obtaining factor scores for the retained factor solution are also discussed. Finally, a simple method is proposed to factor analyse repeated measurements.

Keywords ■ Analyse factorielle, nombre de facteurs, rotations, saturations nulles, facteurs doublets, mesures répétées.

 achim.andre@uqam.ca

 [10.20982/tqmp.16.4.p213](https://doi.org/10.20982/tqmp.16.4.p213)

Acting Editor ■ Denis Cousineau (Université d'Ottawa)

Reviewers

■ Louis Laurencelle (Université du Québec à Trois-Rivières)

■ et un évaluateur anonyme.

An automatic translation in English of this text is available in the supplementary material on the journal's web site.

Introduction

L'analyse factorielle exploratoire (AFE) représente une famille de méthodes quantitatives par lesquelles on cherche à comprendre un ensemble de variables observées en termes de dimensions plus abstraites qu'on appelle facteurs. Pour explorer, par exemple, les attitudes des personnes face à un thème social donné, on pourrait obtenir des données sociodémographiques et des réponses à divers items sur des échelles de Likert. L'AFE pourrait alors contribuer à faire émerger de ces données un petit

nombre de dimensions, qu'on appellera variables latentes ou facteurs. Les facteurs sont conçus comme des dispositions abstraites qu'on ne peut souvent mesurer qu'indirectement mais qui toutefois s'expriment dans diverses variables mesurées faisant que celles-ci sont corrélées entre elles. Pour bien utiliser l'AFE, il est utile d'en comprendre l'esprit et le mécanisme, ce que propose ce texte qu'on pourrait considérer comme une théorie de l'AFE.

Nombre d'utilisateurs de l'AFE la voient d'abord comme une entreprise pour déterminer quelles variables *vont ensemble*, soit comment on peut les regrouper pour, en



quelque sorte, mieux les comprendre et en faire un résumé conceptuel. On cherche ainsi à établir une relation partant des variables vers des facteurs utiles. Cette optique reflète explicitement l'esprit et le modèle de l'analyse en composantes principales (ACP), une technique apparentée. Pour l'ACP, il s'agit d'approximer au mieux les données par un nombre réduit de composantes. Les composantes perdent d'ailleurs cette désignation, prenant celle de facteurs quand, par une opération dite de rotation, on tente de créer du sens en associant à chaque variable latente des sous-groupes de variables observées. On peut qualifier cette approche de fonctionnelle, les facteurs étant appréciés pour leur utilité à condenser l'information des variables en un petit ensemble de concepts.

Souvent, cependant, les données visent à mesurer quelque chose de plus abstrait mais de réel, qu'on présume exister chez les personnes mesurées et qui les caractérise. L'acquisition des données est ainsi vue comme une prise de mesures concrètes qui visent à quantifier du réel abstrait. Il n'est pas nécessaire qu'on sache déjà quelles seraient ces dimensions abstraites; on peut alors espérer que ce soient les données qui nous révèlent ce qui cause leurs corrélations. Mais on garde l'attitude que les facteurs à identifier sont réels, pas seulement fonctionnels. On considère aussi que les valeurs des individus sur ces dimensions abstraites ne sont pas nécessairement fixes et peuvent changer, par exemple, avec le temps et l'expérience.

De plus, souvent les mesures qu'on peut prendre reflètent simultanément plus d'une dimension abstraite. Prenons en exemple des épreuves d'évaluation de l'aptitude à l'attention partagée qui, typiquement, demandent aux participants de retenir des consignes multiples concernant différents types de stimuli. La capacité de la mémoire de travail à garder rapidement accessibles ces consignes pourra influencer les mesures, induisant une dimension (cognitive) supplémentaire à s'exprimer dans les mesures visant l'attention partagée. Ces deux dimensions sont supposées préexister chez les participants avant qu'on les soumette au test, chacun ayant son propre niveau d'aptitude en attention partagée et sa propre capacité de mémoire de travail.

Ainsi, telle que présentée ici, l'AFE s'inscrit dans une approche qu'on peut qualifier de réaliste, par rapport à l'approche utilitaire, en ce sens qu'elle suppose la réalité de dimensions abstraites qui déterminent, au moins en partie, diverses mesures qu'on peut prendre et sont donc imputables des corrélations entre ces mesures. On pourrait désigner « axes naturels » ces dimensions abstraites exprimées dans un ensemble de données. L'AFE est ainsi à la recherche de ces axes naturels, tout comme en EEG de l'épilepsie on pourrait être à la recherche des foyers

d'activité cérébrale réellement à l'origine des crises d'un patient. Il n'est toutefois pas nécessaire d'avoir une idée préalable de ce que devraient être ces axes ou sources. En fait, l'AFE est particulièrement appropriée pour aider à identifier les axes, ou variables abstraites, exprimés dans les corrélations observées.

Quoi qu'on pense du principe de réifier les facteurs ou variables latentes mis à jour par l'AFE et quel que soit son champ d'étude particulier où cette attitude pourrait être plus difficile à afficher, on devrait trouver ci-dessous matière à raffiner sa compréhension de l'AFE. Cette approche conceptuelle de l'AFE centrée sur le flux d'information des axes naturels vers les mesures a déjà le mérite d'avoir mené à l'élaboration d'une méthode objective pour déterminer, par tests statistiques, combien de facteurs communs sont exprimés dans les variables, et celui d'avoir dicté un modèle opérationnel simple pour appliquer l'AFE à des mesures répétées.

On commence, ci-dessous, par décrire les étapes générales de l'AFE, tant pour le néophyte que pour assurer un langage commun. On touche ensuite à l'importance de l'examen préliminaire des données, particulièrement pour viser à les rendre conformes au modèle selon lequel les effets de divers facteurs sur les variables sont additifs. Suivent une discussion générale de la décomposition de variables en facteurs, puis un exemple fictif de données populationnelles où les variables observables proviennent de l'addition pondérée (ici par des poids arbitraires) des valeurs des participants sur les facteurs. La discussion de cette structure de données fait voir que les corrélations attendues entre les variables ne dépendent que des facteurs communs. On discute ensuite de l'ACP, démontrant que toutes ses composantes comportent nécessairement une part d'erreur de mesure. Cela mène à décrire le principe de la factorisation en axes principaux; celle-ci ne vise qu'à expliquer les corrélations entre variables, ignorant autant que possible la variance non partagée exprimée dans les variables individuelles. On discute ensuite comment estimer la part de variance partagée par les variables. Ayant ainsi une vision globale de l'AFE, on aborde ensuite des aspects pratiques, à commencer par la détermination du nombre de facteurs requis pour expliquer les corrélations. Un espace à autant de dimensions que de facteurs est alors donné dans lequel se trouvent les facteurs recherchés. La discussion de cette recherche prend la forme de principes généraux, débouchant sur la considération de rotations orthogonales puis obliques. On discute ensuite de l'interprétation des facteurs fixés pour délimiter l'espace factoriel en termes de contribution des facteurs aux variables, incluant comment on peut juger que certaines contributions puissent être effectivement nulles dans la population, même si elles ne le sont pas dans l'échantillon. On dis-



cute ensuite des scores factoriels, d'abord de l'incertitude inhérente au fait qu'ils sont estimés à partir de variables contenant une part d'erreur de mesure, puis du mérite relatif de différentes options pour leur estimation. Le texte se termine sur trois aspects utiles à connaître mais pas encore discutés, soit premièrement le rapport de l'indice KMO avec des facteurs dits doublets (ne s'exprimant que dans une paire de variables) et l'indétermination de paramètres qui en découle, ensuite de l'analyse à propos d'analyser les covariances sous certaines conditions, et enfin d'une approche simple pour traiter de mesures répétées.

L'AFE dans ses grandes lignes

Quel que soit le point de vue envisagé, fonctionnel ou réaliste, l'AFE vise à cadastrer la complexité du phénomène étudié, à commencer par combien de dimensions abstraites communes sont requises pour rendre compte des corrélations entre les variables observées. Elle vise ensuite à suggérer des interprétations de ces facteurs. Elle peut aussi estimer les valeurs des participants sur chacune des variables latentes identifiées. L'AFE comporte donc trois phases. La première consiste à identifier ce qu'on appelle l'espace factoriel, ce qui inclut la détermination du nombre de dimensions dans lesquelles les variables observées partagent de l'information, de même que la délimitation de cet espace rendant compte des corrélations. La part d'information partagée par les variables observées ainsi que les facteurs mêmes qui fournissent cette information partagée devraient tenir dans l'espace factoriel identifié. À cette étape, pour deux dimensions par exemple, c'est le *plan* des facteurs qui sera identifié. Les axes initialement utilisés pour délimiter ce plan n'obéissent toutefois qu'à des contraintes mathématiques qui sont le plus souvent sans rapport direct avec les facteurs à l'œuvre dans la réalité.

Situés dans ce plan (ici, à 2 dimensions), les facteurs à identifier peuvent être vus comme des réarrangements des axes initiaux, réarrangements qu'on décrit alors comme un processus de rotation dans le plan depuis les axes initiaux vers les facteurs recherchés. Ceci constitue la seconde étape, celle de l'interprétation du plan ou, plus généralement, de l'espace factoriel, quel que soit son nombre de dimensions et qu'on appelle plus exactement *hyperplan*. La rotation des axes initiaux consiste à rechercher, dans le même hyperplan, de nouvelles directions qui représentent des facteurs interprétables et qui définissent *comment les facteurs s'expriment dans les variables*. Le modèle de l'AFE décrit en effet un flux d'information des facteurs vers les variables. C'est le partage d'information provenant des facteurs et exprimée dans les variables qui

explique les corrélations observées entre les variables. Une troisième étape suit optionnellement : elle consiste à estimer les valeurs que, selon la solution retenue, chaque répondant obtiendrait sur les différents facteurs, valeurs qu'on appelle communément scores factoriels.

On peut aussi utiliser l'AFE uniquement pour réduire un grand nombre de mesures observées en un plus petit nombre de valeurs dérivées, les scores factoriels, épurant les données en vue de tests statistiques ultérieurs. Les facteurs fonctionnels ainsi dérivés seront alors vus comme un moyen commode de résumer les relations entre les variables observées, sans véhiculer expressément de signification pour les grandeurs ainsi produites. La rotation est alors facultative. Cette pratique pourrait, par exemple, précéder une analyse discriminante où des différences entre groupes viendraient proposer une hypothèse d'explication parcimonieuse et crédible pour certains des facteurs réels impliqués.

Le mot *analyse* dans l'expression AFE fait référence à la décomposition d'un phénomène en ses parties constituantes. Il ne faut pas y voir la notion de test statistique, trop souvent et faussement considérée comme sens premier de l'expression *analyse de variance*. On verra que des tests statistiques peuvent servir l'AFE, en particulier pour la détermination du nombre de facteurs nécessaires, mais le but général de l'AFE est surtout d'obtenir un éclairage sur la structure interne des données.

Pour bien camper la nature de l'AFE, on parlera ici de *recettes de fabrication* des variables observées à partir des facteurs. Les facteurs réfèrent conceptuellement à des dimensions réelles mais qu'on ne peut observer directement, dimensions sur chacune desquelles les répondants ont des valeurs numériques (sur des échelles arbitraires, typiquement exprimées en cotes standardisées z^1) et qui s'expriment dans des variables observables via des recettes de fabrication : tant du premier facteur, plus tant du second facteur, etc. Tout se passe comme si chaque répondant était caractérisé par un ensemble de valeurs sur les facteurs impliqués et que celles-ci se projetaient de façon appropriée sur des instruments de mesure pour y produire l'ensemble correspondant de valeurs observées.

Examen préalable des données et modèle additif

Ce sont les instruments de mesure qui fixent les grandeurs quantifiées des observations et le point d'origine des échelles, qui aussi captent ou occasionnent les erreurs de mesure et qui parfois déforment l'échelle de telle sorte que deux différences de même valeur numérique ne représentent pas des différences de même importance. Penseriez-vous, par exemple, qu'une augmentation de sa-

1. En fait, les valeurs des répondants sur les facteurs pourraient n'être que positives. Elles deviennent centrées sur 0 parce que l'analyse porte sur les corrélations ou les covariances. L'AFE, telle que pratiquée, ne dit rien sur la moyenne des scores factoriels.



laire de 2 000\$ produise la même émotion chez qui gagne 20 000\$ par année que chez qui en gagne 200 000? Si on met les salaires bruts, sans transformation, en relations linéaires avec d'autres variables, on assume implicitement qu'une même différence correspond à des effets identiques partout le long de l'échelle. Il importe donc de bien inspecter les variables qui seront soumises à l'AFE, notamment pour *défaire les distorsions* et corriger les anomalies que peut avoir introduites l'opération de mesure.

Implicite dans l'exemple de différence de 2.000\$ donné ci-dessus est la notion que l'AFE suppose que les recettes de fabrication sont strictement additives. Pour le moment, en effet, le modèle mathématique de toutes les techniques d'AFE n'admet que des effets additifs des facteurs, tout comme c'est le cas pour les autres techniques statistiques multivariées.

Comme des combinaisons additives d'effets multiples tendent à donner des distributions normales, la symétrie des distributions observées peut servir de guide sur le besoin de transformation des mesures d'origine. Si une variable reflète un seul facteur sous-jacent et que sa distribution observée n'est pas normale, le problème qui se pose pour la correction de son échelle, pour la juste estimation de ses corrélations avec les autres variables qui reflètent le même facteur et pour son éventuel comportement additif est réduit à un éventuel besoin de transformation. Par exemple, si l'instrument de mesure (y) reflète le carré de la contribution (x) du facteur (normalement distribuée et non négative avant l'exclusion des moyennes) plus une éventuelle origine arbitraire dépendant de l'instrument de mesure, p. ex. $y = x^2 + 10$, alors la distribution observée de la variable y est asymétrique et étirée vers la droite. On peut alors prendre la racine carrée de la valeur observée après avoir utilement corrigé son origine, soit ici $x = \sqrt{y - 10}$, restituant ainsi une variable symétriquement répartie. Le passage vers une distribution raisonnablement normale est un bon guide dans le choix de la correction d'origine et de transformation d'échelle à appliquer.

Toutefois, la situation peut être plus complexe pour une variable de distribution non symétrique qui reflète l'effet de plus d'un facteur. Le cas est particulièrement compliqué si la nature doit combiner *des transformations* de valeurs des répondants sur deux (ou plus) facteurs normalement distribués (un exemple suit), ce qui résulte en une variable observée qui n'est pas normalement distribuée. Alors une transformation de la variable observée qui défait la transformation d'un des facteurs risque de déformer celle d'un autre de ses facteurs contributifs. Ainsi, une mesure qui proviendrait de la somme d'un facteur normalement distribué et du carré d'un autre facteur également normalement distribué aurait une distri-

bution observée asymétrique, mais la racine carrée qui corrigerait la contribution du second facteur déformerait en même temps celle du premier facteur (produisant alors une asymétrie négativement pour la contribution de ce facteur à la variable). Il n'y a généralement pas de transformation de variables observées qui corrige pour des transformations non linéaires (exposant, racine, logarithme, etc.) des contributions des facteurs *avant* leur combinaison additive, voire pour leur combinaison autre qu'additive.

Rappelons-le donc : le modèle de l'AFE ne peut correctement rendre compte que de combinaisons *additives* des scores des répondants sur les facteurs. Avant de soumettre les observations à l'AFE, toute structuration non linéaire du résultat de telles combinaisons additives (potentiellement des transformations introduites par l'instrument de mesure) doit être neutralisée, au risque de ne pouvoir bien retrouver les facteurs sous-jacents. L'examen préliminaire des données et leur transformation vers des distributions normales (ou à tout le moins symétriques) augmente les chances de refléter correctement un tel modèle additif.

Prenons l'exemple de questions d'examen qui reflèteraient dans diverses proportions la facilité (F) à comprendre la matière et l'ardeur (A) à étudier. L'AFE ne permet que des modèles où la note à toute question donnée est produite par la somme $aF + bA$, a et b étant des coefficients que l'AFE devrait pouvoir retrouver pour chacune des questions. Le modèle mathématique ne soutient pas la possibilité que la note reflète par exemple le produit $A^a \times F^b$. Dans certains cas particuliers, remplacer les mesures observées par leur logarithme pourrait transformer un modèle intrinsèquement multiplicatif en modèle additif, p. ex. $\log(A^a \times F^b)$ devenant $a \cdot \log(A) + b \cdot \log(F)$, mais la situation doit être très particulière pour que cela fonctionne. Si une mesure reflétait à la fois de la multiplication et de l'addition, la transformation logarithmique ne conviendrait plus.

Analyse factorielle

L'ACP est trop souvent utilisée comme une technique d'AFE. L'ACP transforme p variables corrélées en p facteurs indépendants appelés composantes principales. En ce sens, elle n'offre ni économie de données ni compréhension. Toutefois, comme les composantes principales sont successivement identifiées telles que chacune explique le plus de variance résiduelle possible – la variance non encore expliquée par les composantes précédentes –, on peut se croire justifié de ne retenir que les k premières composantes, où k est inférieur à p et est adéquatement choisi. Ces composantes peuvent alors être traitées comme des facteurs, et des opérations ultérieures, les rotations, peuvent être appliquées en cherchant à produire des facteurs plus interprétables.



En psychologie, les situations sont nombreuses où on veut quantifier des dispositions abstraites à partir de mesures observables, par exemple des tests ou encore des questionnaires basés sur des scores gradués de type Likert. Chaque mesure étant sujette à une bonne part de variance propre, attribuable à la spécificité de la mesure et à l'erreur aléatoire qui s'y attache, il est typique de prévoir de nombreuses mesures plus ou moins redondantes qui, collectivement, ne reflètent qu'un tout petit nombre de facteurs. Le modèle de l'AFE stipule explicitement que chaque variable observée contient une part de variance qui lui est propre; on l'identifiera comme résidu de variance, au-delà de la variance qui sera reproduite par les recettes de fabrication des variables à partir des facteurs communs. Il s'agit donc d'expliquer le partage de l'information dans divers groupes de variables. Cela consistera souvent à rendre compte des *corrélations* entre les variables, tandis que parfois ce seront plutôt leurs *covariances* qu'on cherchera à éclairer (voir plus loin).

L'AFE s'exprime principalement par la technique de factorisation en axes principaux (FAP) et parfois par d'autres techniques, telle celle de maximum de vraisemblance, techniques dont nous ne parlerons ici que peu. Dans tous les cas, il y a effort d'identifier la part de variance propre (i.e., unique, non partagée) de *chacune* des variables pour pouvoir l'ignorer. Considérer, dans l'ACP, que cette variance propre correspond aux dernières composantes principales qu'on laisse tomber relève d'une incompréhension profonde. Comme on le démontrera sous peu, l'ACP incorpore nécessairement une part de la variance unique des variables dans chacune des composantes principales, surestimant en cela la part de variance due à chaque facteur commun. C'est pour cela que l'ACP *tronquée* (c.-à-d. comportant moins de facteurs que le nombre de variables) n'est pas un équivalent adéquat de l'AFE. *En sciences humaines, c'est très généralement l'AFE et non l'ACP qu'il faut utiliser.*

Beaucoup de publications décrivant l'utilisation de l'AFE incluent des prescriptions pour la constitution de l'ensemble de variables sur lesquelles on l'appliquera. On propose alors la notion de *structure simple* d'après laquelle quelques variables reflètent chaque facteur et surtout chaque variable ne relève autant que possible que d'un seul des facteurs retenus. Ces prescriptions sont utiles quand on veut développer un instrument pour quantifier un ensemble de dimensions déjà identifiées conceptuellement. L'analyse factorielle confirmatoire, un cas particulier de modélisation par équations structurelles développée des années après l'AFE, est en principe bien plus adaptée pour l'examen de tels facteurs.

Les situations dont on parle ici portent sur le problème plus général où l'on a un ensemble de variables dont

on s'attend qu'elles expriment un ensemble de facteurs communs, mais pas nécessairement selon une structure simple. On pourrait avoir, par exemple, des résultats de questions d'examen qui dépendent tous de mêmes dispositions (i.e., facteurs à identifier, tels la facilité à comprendre une matière et l'ardeur à l'étudier) dans des proportions qui peuvent varier d'un item à l'autre, sans qu'aucun item ne reflète qu'un seul facteur. Toute collection de mesures susceptible de refléter des sources communes d'information encore peu identifiées tant en nombre qu'en nature se prête à l'AFE pour aider à en défricher les concepts attendus.

Modèle mathématique de l'AFE

Pour bien saisir certaines propriétés de l'AFE, considérons des données simulées à partir de deux variables internes indépendantes, les facteurs F1 et F2, dont les valeurs sont assignées au hasard à partir de distributions de moyennes nulles et de variances unitaires. Que les distributions suivent une loi normale n'importe pas particulièrement ici. Fabriquons quatre variables à partir de recettes de fabrication arbitraires quelconques. Disons donc que le processus qu'utilise la nature pour exprimer les variables *A*, *B*, *C* et *D* à partir de F1, de F2 et d'erreurs de mesure est celui donné au Tableau 1, chaque recette étant accompagnée, à droite, par la variance qu'elle génère dans la population lorsque F1 et F2, sans corrélation entre eux, ont une variance unitaire.

F1 et F2 sont appelés *facteurs communs*; ils s'expriment chacun dans plus d'une variable. Il n'est pas nécessaire qu'une des variables, ici *C*, ne reflète qu'un seul des facteurs communs. Il est toutefois fort préférable que chaque facteur commun s'exprime dans au moins trois variables: on reviendra là-dessus.

Bien qu'on parlera de l'espace des variables et de celui des facteurs, il est utile de pouvoir relier ces deux sous-espaces à l'espace initial des répondants. Chaque répondant, étant indépendant des autres, peut être représenté comme un axe d'un espace à autant de dimensions (*N*) qu'on a de répondants. Toute variable à analyser peut alors être vue comme un vecteur, soit une flèche partant de l'origine jusqu'à un point dans l'espace des répondants dont la coordonnée est constituée du score de chaque répondant (i.e., chaque axe) sur cette variable.

Pour étudier les covariances, on sépare ces vecteurs en deux parties orthogonales, qui sont la projection de la variable sur l'hyper-diagonale (direction des coordonnées identiques sur tous les répondants) et son complément pour reproduire exactement la variable. En soustrayant ainsi la moyenne de chaque variable et donc ne gardant que ces compléments, les variables n'ont plus rien dans la direction de l'hyper-diagonale; deux variables



Tableau 1 ■ Recettes de fabrication et variances attendues de quatre variables à partir de deux facteurs et de quatre termes d'erreur

	Variance	
	(F1 et F2) + erreur	= totale
$A = 10 \times F1 + 8 \times F2 + 5 \times N(0; 1) + 42$	164 + 25	= 189
$B = 200 \times F1 + 25 \times F2 + 50 \times N(0; 1) + 200$	40625 + 2500	= 43125
$C = 0,8 \times F2 + 0,2 \times N(0; 1) + 5$	0,64 + 0,04	= 0,68
$D = 38 \times F1 + 2 \times F2 + 10 \times N(0; 1) + 30$	1588 + 100	= 1688

Note. $N(0; 1)$ indique une distribution aléatoire normale de moyenne 0 et variance 1.

indépendantes sont alors attendues à 90° l'une de l'autre, engendrant une somme nulle des produits de leurs coordonnées respectives sur l'ensemble des axes (répondants). L'exclusion de l'hyper-diagonale correspond au degré de liberté perdu dans l'espace des répondants, les variables tenant maintenant dans seulement $N-1$ dimensions, avec plus rien sur l'hyper-diagonale.

Pour étudier les variables sans égard à leurs unités de mesure initiales, on peut les « normaliser », soit ramener à 1 la longueur (mathématiquement appelée norme) de chaque variable (après exclusion de l'hyper-diagonale); cela équivaut à leur donner une variance (carré de la longueur) de 1. L'espace délimité par p variables occupe seulement p dimensions à l'intérieur de l'espace des répondants (ou moins, si une des variables peut être parfaitement reproduite par combinaison d'autres variables) hormis l'hyper-diagonale. L'analyse factorielle cherche à réduire l'espace de p variables en un espace encore plus compact de k dimensions, délimité par autant de facteurs. Dans ce sous-espace, les variables n'auront plus leur pleine longueur (hyper-diagonale exclue) : une part de leur variance, celle non partagée avec d'autres variables, leur manquera.

Pour les données de population, chacune des variables données au Tableau 1 peut être standardisée (moyenne 0, variance 1) en divisant ses diverses pondérations par la racine carrée de la variance totale (la somme des carrés des pondérations des diverses sources de variance), et en laissant tomber la constante. On obtient alors le Tableau 2, où les entrées donnent les poids attribués aux sources de variance unitaire en entête, lesquelles incluent le terme d'erreur unique de chaque variable. Dans cette forme, quand les variables internes (facteurs + terme d'erreur) ont chacune une variance de 1, la variance attendue de chaque variable observée est la somme des carrés de ses poids et, ce qui est plus pertinent encore, la corrélation attendue entre

deux variables est la somme des produits de leurs poids respectifs.

Alors que chaque ligne définit une somme de carrés (SC) de 1, pour un total de 4 découlant de nos quatre variables, la SC de ces paramètres faite par colonne montre la propriété que les facteurs communs, lesquels contribuent de la variance sur plus d'une variable, finissent par cumuler passablement de variance tandis que les sources d'erreur (ou plus généralement les sources de variance unique) ne contribuent qu'à une seule variable, et ce à seule fin dans notre modèle de compléter sa variance à 1. De plus, répétons-le, ce qui est particulièrement important pour l'AFE, seuls les poids sur les facteurs communs contribuent aux corrélations attendues entre les variables.

Comme le montre le Tableau 2, pour engendrer ces quatre variables, on s'est servi de 6 distributions indépendantes de variance unitaire. Les variables observées s'inscrivent donc comme un sous-espace à quatre dimensions (parce qu'il n'y a que quatre variables observées) d'un espace de variables latentes (internes) à six dimensions. En plus de noter que la somme des variances est de 4, ce qui est le nombre de variables standardisées² observées, on remarquera surtout que les deux premiers axes de l'espace initial à six dimensions contribuent un total de $2,3121 + 1,3796 = 3,6917$ de variance, alors que le 0,3083 de variance qui reste est cumulé sur quatre coordonnées en dehors de ces deux axes (dans les six dimensions).

Étant donné les recettes de fabrication du Tableau 2, la matrice de corrélations (somme des produits sur les six coordonnées standardisées³) des variables dans la population est telle que donnée au Tableau 3.

2. Les expressions normalisées et standardisées traduisent des concepts très voisins. Une variable observée transformée en cotes z est standardisée. La longueur du vecteur qui la représente est la racine carrée du carré moyen des déviations de sa moyenne, longueur qui implique une division par la racine de $N-1$. On comprend mieux les enjeux quand on pose que les directions des variables sont « normalisées », c'est-à-dire de longueur 1, sans égard à la taille d'un échantillon particulier.

3. Par exemple, la corrélation entre les variables A et B s'obtient, à partir du Tableau 2, par $0,7274 \times 0,9631 + 0,5819 \times 0,1204 + 0 = 0,7706$, les entrées vides du Tableau 2 étant en fait des zéros.



Tableau 2 ■ Pondérations des cotes z des variables du Tableau 1

Var	F1	F2	errA	errB	errC	errD
A	0,7274	0,5819	0,3637			
B	0,9631	0,1204		0,2408		
C	0,9701				0,2425	
D	0,9249	-0,2921				0,2434
SC	2,3121	1,3796	0,1323	0,0580	0,0588	0,0592
						total : 4

Note. Les entrées vides sont des 0. SC = somme des carrés.

Tableau 3 ■ Corrélations populationnelles entre les quatre variables

	A	B	C	D
A	1,0000	0,7706	0,5645	0,5028
B	0,7706	1,0000	0,1168	0,8556
C	0,5645	0,1168	1,0000	-0,2834
D	0,5028	0,8556	-0,2834	1,0000

Analyse en composantes principales.

Avant de considérer la matrice de corrélations d'un échantillon, voyons d'abord l'ACP de la matrice de corrélations de la population. Cela fera comprendre pourquoi l'ACP n'est pas un choix approprié pour identifier l'espace des facteurs communs. Rappelons ici que la première phase de la FAP et de l'ACP tronquée ne cherche qu'à identifier l'espace des facteurs communs, pas nécessairement ces facteurs eux-mêmes. Il faut donc d'abord déterminer combien il y a de facteurs communs présents pour ensuite identifier ce qu'on appellera un système d'axes dans lequel s'inscriront ces facteurs.

Les procédures à appliquer ultérieurement pour développer des interprétations vraisemblables des facteurs communs seront discutées plus loin. Le système d'axes (directions mutuellement orthogonales) dans lequel s'inscrivent les facteurs lors de leur extraction est arbitraire à la fois par l'orthogonalité imposée aux facteurs et par sa géométrie (directions spécifiques identifiées pour circonscrire l'espace retenu). Ce n'est qu'ensuite qu'il faut trouver un nouveau système de directions, pas nécessairement orthogonales, dans l'espace des axes initiaux, directions qui correspondront à ce que pourraient être les facteurs. Valider la réalité d'une hypothèse sur ce que pourraient être les véritables facteurs communs peut demander perspicacité et créativité. Il est prudent de retenir que toute solution que produit une application de l'AFE, même avec une rotation apparemment appropriée, n'est qu'une suggestion, une approximation possible de ce que peut être la réalité sous-jacente. Bien qu'on veuille les fac-

teurs comme des entités abstraites réelles, on doit donc se garder de réifier d'emblée une solution factorielle. Le jugement et l'ouverture à des solutions autres demeurent toujours de mise.

Signalons ici que quatre variables non parfaitement redondantes occupent un espace à autant de dimensions, espace dans lequel le coefficient de corrélation entre deux variables correspond au cosinus de l'angle qui les sépare. L'ACP d'une matrice de corrélations identifie successivement des axes (directions) dans l'espace occupé par les variables tels que chacun explique le plus de variance possible et tel que chaque nouvel axe soit orthogonal à chacun des axes précédemment identifiés. La variance d'une variable expliquée par une direction est le carré de la longueur de la projection de la variable sur la direction (la variable étant alors décomposée en deux parties orthogonales, une partie dans la direction en question et un résidu; c'est le théorème de Pythagore qui s'applique ici).

Pour la matrice des corrélations théoriques (celles dans la population, données au Tableau 3) l'ACP trouve d'abord, dans l'espace des quatre variables, une direction qui, à elle seule, peut expliquer une somme de carrés partielle aussi élevée que 2,4658, soit 61,64% du total de 4. Elle identifie ensuite une seconde direction, perpendiculaire à la première, qui explique une autre part de la somme des carrés totale, soit 1,3743 (34,36% du total), pour un cumulatif de 3,8401, soit 96,0% de la variance totale.

On note ici que ce total dépasse par environ 0,15 celui de 3,6917 qu'on retrouve dans le plan de F1 et F2 de l'espace original à six dimensions. Pour nos deux dimensions sur quatre variables, ce 0,15 est environ la moitié

4. La part de variation propre d'une variable, celle qu'elle ne partage avec aucune autre variable de l'ensemble considéré, est souvent dénommée « variance unique » ou « unicité ». Dans la théorie des tests classique, cette unicité se décompose en une part de variance vraie spécifique (c.-à-d. associée à une grandeur réelle qui ne se retrouve pas dans l'autre (ou aucune autre) variable corrélée, plus une part d'erreur aléatoire, laquelle est par



Tableau 4 ■ Contributions des composantes principales aux variables.

	CP1	CP2	CP3	CP4
A	0,8899	0,3879	0,2400	0,0012
B	0,9668	-0,1518	-0,1277	-0,1612
C	0,2614	0,9504	-0,1535	0,0695
D	0,8191	-0,5455	-0,0611	0,1668
SC	2,4658	1,3743	0,1012	0,0587

(deux parts sur quatre) de la variance totale contribuée par les quatre termes de variance propre⁴. Cela vient de ce que la variance contribuée par ces quatre termes d'erreur, indépendants entre eux, est partiellement répartie dans les deux dimensions de F1 et F2 : *c'est seulement en six dimensions et plus que les quatre dernières dimensions sont perpendiculaires aux deux premières*. Même si la variance totale des quatre variables dépend de six sources indépendantes, les quatre variables disponibles ne délimitent qu'un espace à quatre dimensions, superposant les contributions des deux facteurs à celles des quatre termes d'erreur.

Après avoir identifié deux dimensions qui expliquent déjà 96,0% (= 3,84/4) de la variance totale, une nouvelle dimension perpendiculaire aux deux premières peut être identifiée capable de contribuer jusqu'à 0,1012 à la somme des carrés des quatre variables. Il reste une dernière dimension, perpendiculaire aux trois déjà identifiées, qui explique le 0,0587 encore manquant pour reproduire parfaitement la somme des carrés des quatre variables.

Les coordonnées des quatre variables standardisées et leurs sommes de carrés sur les quatre composantes principales sont présentées au Tableau 4. On peut concevoir que CP1 à CP4 sont quatre sources hypothétiques indépendantes et de variance unitaire, avec les inscriptions du tableau présentant la pondération qu'on doit donner à chaque source pour reproduire exactement chacune des variables. Ces poids sont souvent appelés *saturations* (on parle de saturation des variables sur les facteurs), mais le vocable *pondération* semble mieux véhiculer l'idée que, selon le modèle, le transfert d'information se fait des facteurs vers les variables observées, en donnant des poids plus ou moins élevés aux scores qu'auraient les répondants sur ces facteurs présumés. Si, ici, on ignorait CP3 et CP4, on ne reproduirait qu'approximativement les quatre variables observées, sans toutefois reproduire exactement leur variance fiable étant donné que CP1 et CP2 incluent forcément une part des termes d'erreurs des quatre variables.

Alors que les sommes de carrés de chacune des quatre variables sur les quatre composantes principales sont

toutes 1, celles sur seulement les deux premières composantes principales (CP1 et CP2) sont A : 0,9424, B : 0,9577, C : 0,9716 et D : 0,9684 (dont le total donne 3,8401). Ces nombres sont, pour chaque variable, la proportion de sa variance expliquée par le plan des deux composantes. Lorsqu'on ne garde que certaines composantes, ici deux, ces nombres représentent ce qu'on appelle les *communalités* respectives des variables.

Les contributions successives des quatre composantes à la somme des carrés totale, soit 2,4658, 1,3743, 0,1012 et 0,0587, sont désignées *valeurs propres* (eigenvalues) et les directions (axes) qui leur sont respectivement associées sont désignées *vecteurs propres* (eigenvectors) ou composantes principales.

L'appendice A explique l'essence de la méthode par laquelle les directions successives de variance maximale, sujettes à l'orthogonalité avec les directions déjà identifiées, peuvent être obtenues.

Factorisation en axes principaux

La technique de l'ACP peut aussi s'appliquer à des matrices (de covariances ou de corrélations) dont la diagonale principale est réduite, et c'est cela qui constitue la FAP. Bien que « ACP » fasse habituellement référence à la décomposition d'une matrice non-réduite, on emploie aussi parfois le terme pour décrire la technique d'extraction des axes à partir d'une matrice réduite. Le contexte dicte s'il s'agit du nom d'une analyse (ACP au sens propre) ou simplement de la technique de décomposition utilisée en FAP.

Pour introduire la FAP, revenons d'abord au Tableau 2. Lorsque des variables sont ainsi exprimées en coordonnées sur des axes orthogonaux et ont chacune une somme de carrés de 1, alors leur coefficient de corrélation est simplement la somme des produits des coordonnées sur l'ensemble des axes. À partir de cette propriété, on comprend que les termes de variance propre, parce qu'ils ne s'expriment chacun que sur une seule variable, ne contribuent pas à la somme des produits des coordonnées de la paire de variables. Au niveau populationnel, les corrélations n'expriment que le partage d'information sur

définition du « bruit » pur, sans corrélation systématique avec quoi que ce soit. Le premier type d'unicité contribue à la corrélation test-retest, mais non le second.



des facteurs communs.

Si on éliminait les contributions de variance propre, on réduirait de fait les six dimensions qui ont présidé à la fabrication des valeurs observées aux seules dimensions de F1 et F2. Les sommes de produits de chaque paire de variables sur ces deux facteurs demeureraient les mêmes, tandis que la somme des carrés de chaque variable serait réduite de sa variance propre. Autrement dit, si on part de la matrice de corrélation mais que, aux 1 de la diagonale principale, on substitue des valeurs de variance qui correspondent à la somme des carrés des pondérations des variables sur les seuls facteurs communs, on reproduit de fait non seulement les sommes de carrés mais aussi les sommes de produits dues aux facteurs communs. Une ACP (au sens technique) d'une telle matrice de corrélations populationnelles réduite devrait alors pouvoir reproduire exactement cette matrice à partir de ses seules (ici deux) premières composantes principales. Si les corrélations de la matrice ont pu être engendrées par seulement deux sources de variance, deux composantes principales récupéreront l'espace commun de ces deux sources de variance. La variance totale à expliquer sera la somme des valeurs mises en diagonale de la matrice, une somme appelée *trace* de la matrice.

La FAP recherche donc les facteurs communs en s'intéressant à expliquer (c.-à-d. à reproduire) seulement les corrélations entre les variables, plutôt que leurs variances complètes. Pour cela, elle applique une ACP à une matrice dont la diagonale estime, pour chaque variable, la variance attribuable aux seuls facteurs communs. Plutôt que les corrélations, ce pourrait être les covariances qu'on cherche à expliquer, l'idée commune étant qu'on veut en défalquer les contributions des sources de variance unique. D'autres techniques d'AFE, dont GLS (moindres carrés généralisés), visent le même résultat (expliquer les corrélations) par des moyens autres que l'estimation des variances dues aux facteurs communs suivie d'une ACP. Nous n'en parlerons pas ici.

Rappelons que ce qu'on appelle communauté d'une variable est la part de sa variance qu'on considère expliquée par son association aux facteurs communs, part dérivée de ses corrélations avec les autres variables qui s'y rattachent. Dans le cas de l'ACP d'une pleine matrice de corrélations où cette « explication » repose sur un nombre suffisant de *composantes principales*, on sait que cela surestime la communauté réelle, soit la variance due aux seuls facteurs communs, puisque la variance unique de l'ensemble des variables est répartie en autant de dimensions qu'on a de variables.

La FAP expurge cette contamination et s'intéresse seulement à la variance provenant des facteurs communs. Supposons donc, dans un premier temps, que les

communautés réelles (plutôt qu'estimées) des variables sont connues, ce qui implique qu'on connaisse la part de variance unique de chaque variable, soit les carrés des pondérations des termes d'erreur des cotes z des différentes variables données au Tableau 2. La communauté réelle d'une variable est alors la différence entre 1, sa variance totale, et le carré de la pondération de son terme de variance unique. Se référant aux recettes de fabrication initiales, ces communautés sont exactement la variance de chaque variable attribuable à F1 et F2 au Tableau 1, divisée par sa variance totale, soit respectivement $164/189=0,8677$, $40625/43125=0,9420$, $0,64/0,68=0,9412$ et $1588/1688=0,9408$. La variance totale à expliquer devient donc la somme de ces quatre valeurs, soit 3,6917. Si donc on remplace les 1 dans la diagonale de la matrice de corrélations par ces communautés réelles (lesquelles, dans un cas concret, nous seraient inconnues), cela nous donne la matrice réduite de corrélations du Tableau 5. La décomposition ACP de cette matrice est donnée au Tableau 6.

La première composante principale (C1) de cette matrice réduite présente une valeur propre de 2,3842 (plus que le 2,2321 fourni par F1), la seconde (C2) de 1,3075 (moins que le 1,3796 fourni par F2), pour un total de 3,6917, soit exactement la variance totale contribuable aux quatre variables par F1 et F2. Ceci n'est pas une coïncidence. On est en fait à une rotation près de F1 et F2. Autrement dit, la relation entre d'une part C1 et C2 de la solution trouvée ici, et d'autre part F1 et F2 de la solution réelle qu'on aimerait trouver est uniquement celle d'une rotation. F1 et F2, les *vrais* facteurs, sont deux directions orthogonales dans le plan identifié par C1 et C2.

Les directions de C1 et C2 elles-mêmes ne sont pas les facteurs recherchés. C1 est plutôt *la* direction dans le plan de F1 et F2 qui explique le plus de variance possible des quatre variables hors de leurs variances uniques. C2 prend en compte ce qu'il reste à exprimer du plan de F1 et F2. Les deux dernières composantes principales de la matrice réduite du Tableau 5 ont leurs valeurs propres nulles.

Ainsi, si on pouvait identifier correctement les communautés réelles des variables et si les corrélations entre les variables étaient estimées sans erreur, on retrouverait exactement l'espace des facteurs communs dépeint dans l'illustration ci-après. Cet espace serait toutefois identifié non pas par les facteurs sous-jacents F1 et F2 eux-mêmes, mais par des axes C1 et C2 successivement identifiés, répondant chacun à un critère de variance maximale expliquée, tenant compte des axes précédemment fixés.

La Figure 1 représente, en bleu, des portions des variables A , B , C , D dans le plan de F1 et F2, ainsi que, en noir gras, des composantes C1 et C2 extraites de leur matrice réduite des coefficients de corrélations. Autant les



FIGURE 1 ■ Positions de la part expliquée des variables *A*, *B*, *C*, et *D* dans l'espace délimité par les véritables facteurs *F1* et *F2*, et par les composantes principales *C1* et *C2* de leur matrice réduite de corrélations.

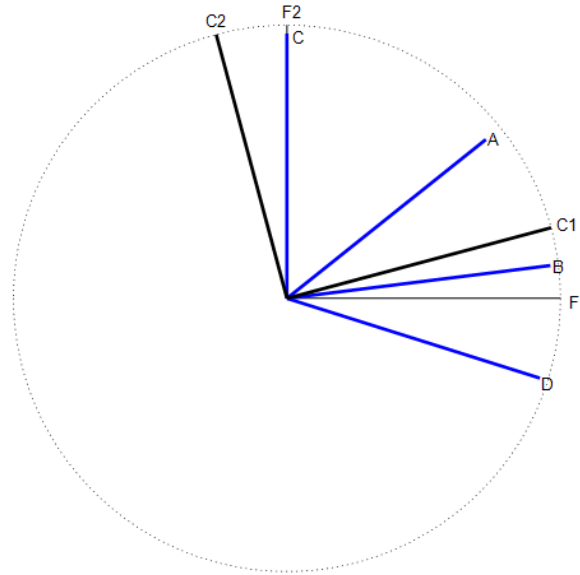


Tableau 5 ■ Matrice réduite des corrélations populationnelles

	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>
<i>A</i>	0,8677	0,7706	0,5645	0,5028
<i>B</i>	0,7706	0,9420	0,1168	0,8556
<i>C</i>	0,5645	0,1168	0,9412	-0,2834
<i>D</i>	0,5028	0,8556	-0,2834	0,9408

facteurs *F1* et *F2*, en noir fin, que les composantes *C1* et *C2* ont une longueur unitaire (i.e., touchent le cercle pointillé de rayon 1) étant orthogonaux entre eux; ce sont deux systèmes d'axes possibles pour décrire le plan commun des variables. Ces dernières ont une longueur correspondant à la racine carrée de leur communauté (variance due à *F1* et *F2*).

Estimation des communautés

En FAP, les communautés doivent être estimées, puis insérées à la place des 1 de la diagonale de la matrice de corrélations. Les dimensions sont ensuite extraites par ACP de la matrice de corrélations ainsi réduite. Si les communautés étaient correctement identifiées et les corrélations précisément celles de la population, cela identifierait exactement l'espace des facteurs communs, expliquant toute la variance de la matrice réduite, tel que décrit ci-dessus en relation avec les Tableaux 5 et 6.

Ce qui varie entre les diverses méthodes de FAP, c'est essentiellement la façon d'estimer les communautés. Une première approche, la plus classique et la plus simple, passe par l'équivalent de la régression multiple de chacune des variables sur toutes les autres. On utilise alors, comme communauté estimée d'une variable, la proportion de sa variance qui est expliquée par les autres variables, soit le R^2 de sa régression multiple. En pratique, on obtient ces valeurs de R^2 simultanément pour toutes les variables par une procédure simple qui implique l'inversion de la matrice de corrélations.⁵

Deux autres approches apparentées entre elles utilisent une méthode itérative pour estimer les communautés. Disons d'abord que lorsqu'une matrice de corrélations est trop fortement réduite, sa décomposition en composantes principales produit au moins une valeur propre négative, symptôme d'un manque de variance dans la diagonale de la matrice.

5. En langage de Matlab, on a : $R_{reduit} = R - \text{diag}(1 / \text{diag}(\text{inv}(R)))$, où R est la pleine matrice de corrélations.



Tableau 6 ■ Décomposition en composantes principales de la matrice réduite du Tableau 5

	C1	C2	C3	C4
A	0,8532	0,3739	0	0
B	0,9614	-0,1328	0	0
C	0,2510	0,9371	0	0
D	0,8179	-0,5214	0	0
SC	2,3842	1,3075	0	0

Une méthode pour produire une estimation améliorée des communautés utilise les R^2 comme estimations initiales puis les réévalue à partir des pondérations de toutes les composantes à valeurs propres positives, réitérant jusqu'à ce que la solution se stabilise. Si, dans un cycle donné d'itération (pour un assez grand nombre de variables observées), on a cinq valeurs propres positives, la somme des carrés des pondérations des cinq premières composantes sur une variable devient la nouvelle estimation de sa communauté. Si, au cours d'une autre itération, la matrice réduite n'a que quatre valeurs propres positives, alors les communautés sont réévaluées à partir des quatre premières composantes principales de la matrice réduite courante. Les itérations s'arrêtent quand un résultat de communautés estimées devient essentiellement le même que celui du cycle antérieur, soit lorsque la plus grande modification de communauté depuis le cycle précédent est inférieure en valeur absolue à un seuil arbitraire, comme 0,001, ou bien qu'un nombre maximum d'itérations, p. ex. 25, est atteint. Cette méthode ne demande pas d'avoir déjà estimé le nombre de dimensions exprimées dans les données.

Une troisième méthode d'estimation des communautés, celle utilisée entre autres par SPSS et par la fonction `fa()` du `package psych` du système de programmation R, demande une hypothèse quant au nombre de dimensions exprimées dans les données, disons k , qui doit être moindre que le nombre de variables. Elle consiste à estimer itérativement les communautés à partir des k premières composantes principales de la matrice précédente, jusqu'à convergence comme ci-dessus. L'estimation initiale est obtenue par les k premières composantes principales de la pleine matrice de corrélations, ce qui évite d'avoir moins que k valeurs propres positives dans les itérations successives. Parce que cette méthode utilise plus d'information, nommément l'hypothèse sur le nombre de dimensions, on doit s'attendre à ce qu'elle produise des estimations plus précises des communautés lorsque le nombre de dimensions supposé est correct. Cela s'avère effectivement dans les simulations, et seule cette seconde méthode basée sur les valeurs propres est considérée ici dorénavant, sauf stipulation contraire.

En pratique, la FAP dotée de la méthode itérative donne

occasionnellement des solutions aberrantes au sens où la recette de fabrication proposée expliquerait plus que 100% de la variance d'au moins une variable. Cet effet est dû à une dérive du processus d'itérations pour estimer quelles communautés mettre en diagonale de la matrice de corrélations. Ces situations, désignées « cas de Heywood », signifient qu'une solution interprétable n'a pas été trouvée. Nous reviendrons là-dessus en regard de la discussion de l'indice KMO et de la désirabilité que chaque facteur s'exprime dans au moins trois variables observées.

Nombre de dimensions

La problématique du nombre de facteurs à retenir est celle d'identifier combien de sources d'information sont exprimées dans les corrélations des variables. Il s'agit de découvrir ou de juger combien de dimensions extraites de la matrice de corrélations réduite sont appropriées pour expliquer les corrélations observées. Tel que montré dans l'appendice A, ces dimensions sont identifiées successivement en vertu d'hypothèses provisoires successives, à savoir qu'on suppose implicitement chaque fois qu'une seule dimension suffit à rendre compte de toute la variance encore à expliquer. L'acceptation d'une deuxième composante vient donc réfuter l'hypothèse sous laquelle la première composante a pu être identifiée. Ce premier axe, de même que le second forcé de lui être perpendiculaire ne sont pas crédibles comme facteurs.

L'espace factoriel, soit l'hyper-plan des facteurs retenus, est toutefois généralement bien identifié, bien que les axes initiaux qui le délimitent ne correspondent généralement pas aux véritables facteurs recherchés. La fonction des rotations sera alors de proposer un nouveau système d'axes (facteurs orthogonaux) ou de facteurs partiellement corrélés (facteurs obliques) qui sous-tendent l'espace déjà identifié. Il faut donc d'abord avoir décidé du nombre de dimensions requises, ce qui requiert en même temps de circonscrire l'espace factoriel par les k premières composantes principales de la matrice réduite de corrélations (ou de covariances).

On revoit ci-dessous diverses méthodes utilisées pour aider à déterminer combien de dimensions retenir, pour en arriver à décrire la méthode la plus fiable, basée sur la comparaison statistique des données avec leur meilleur



modèle à k facteurs, pour des valeurs croissantes de k . Les méthodes qui précèdent la section sous-titrée « NEST : Next Eigenvalue Sufficiency Test », sont décrites à motif de culture générale, permettant de comprendre pourquoi on ne veut pas les utiliser sachant qu'il y a maintenant beaucoup plus fiable et parfaitement justifié.

Reconnaissons d'abord que l'AFE basée sur la méthode du maximum de vraisemblance offre déjà un test statistique pour déterminer si k dimensions sont suffisantes, ces k dimensions étant ajustées pour rendre les corrélations observées le plus probables possible (étant donné le sous-espace à k dimensions en voie d'identification). Le test statistique appliqué dans cette approche dépend de ce que les corrélations traitées émanent d'un ensemble de variables multi-normales. Il s'agit alors de vérifier si une seule dimension est suffisante, sinon si deux dimensions conviennent, et ainsi de suite jusqu'à ce qu'on ait suffisamment de facteurs pour que les corrélations observées puissent être considérées comme essentiellement celles produites par le modèle à k facteurs, affectées seulement de petits écarts attribuables à l'erreur d'échantillonnage. On obtient donc ici un espace sur lequel la somme des produits des coordonnées de toute paire de variables reproduit de près leur corrélation observée. Par parcimonie, on veut que k soit aussi petit que possible, tout en permettant que le modèle à k facteurs ne soit pas statistiquement rejeté par les corrélations observées. Le modèle à k facteurs constitue donc, pour le test statistique associé, l'hypothèse nulle quant au partage d'information entre les variables. Cette approche de vraisemblance maximale est hélas souvent affectée de cas de Heywood; ceux-ci sont toutefois signalés à l'utilisateur, qui sait alors que le modèle factoriel résultant n'était pas fiable.

Pour d'autres techniques d'AFE et surtout pour l'ACP, on a longtemps recouru à des critères heuristiques dont les logiques respectives dépendent de conditions souvent non réalisées. Parmi les plus facilement disponibles, on retrouve le critère de Kaiser, qui consiste à retenir toutes les dimensions qui affichent une valeur propre supérieure à 1 et seulement celles-là. Cela s'applique aux valeurs propres de la pleine matrice de corrélations, c'est-à-dire à celles de l'ACP (dans laquelle la diagonale principale est occupée par des 1). Son principe est qu'on ne retienne que des variables latentes qui rendent compte de plus de variance que n'en possède une variable individuelle quelconque. Ce principe ne correspond pas à ce qu'on veut faire quand on cherche à identifier tous les facteurs communs et eux seuls, excluant la variance unique qu'on pourrait par ailleurs étudier séparément s'il y avait lieu.

Une autre approche facilement disponible est celle du test visuel des éboulis (Scree test), montrant en graphique les valeurs propres décroissantes associées aux

composantes principales successivement extraites. Ce test suppose que les communautés réelles des variables sont sensiblement les mêmes pour toutes les variables. Il en résulterait que, jusqu'à ce qu'on ait identifié toutes les composantes principales nécessaires, les valeurs propres consécutives décroissent plutôt rapidement mais que, lorsqu'il ne reste plus que des unicités (ou erreurs de mesure) à peu près égales entre elles, l'algorithme de l'ACP choisit comme composante suivante la dimension *accidentellement* la plus grande parmi celles restantes, un critère d'opportunité qui dépend en grande partie d'un hasard d'échantillonnage. Ainsi, quand on choisit la plus grande dimension dans ce qui serait, dans la population, des dimensions équivalentes de bruit, on produit une décroissance *lente et graduelle* des valeurs propres. Le test des éboulis, appliqué d'ordinaire visuellement, exclut donc cette dernière portion des valeurs propres qui affiche une décroissance essentiellement régulière à partir d'un certain rang.

Ce n'est toutefois pas généralement le cas que les parts de variance unique soient toutes à peu près égales. Des variables psychométriques, par exemple, peuvent avoir des corrélations test-retest (dites *fidélités*) passablement plus faibles que celles d'autres variables mesurées presque sans erreur. Une façon d'utiliser le test des éboulis plus lucidement serait d'apprécier le graphe des valeurs propres au regard des communautés des variables estimées après chaque extraction d'une nouvelle composante principale. Si, par exemple, les données auxquelles on applique l'ACP incluent une variable orpheline (ne dépendant d'aucun facteur commun et donc n'exprimant dans l'échantillon que des corrélations accidentelles très faibles avec d'autres variables), cette variable gardera une communauté estimée presque nulle (presque toute sa variance restant à expliquer) jusqu'à ce que l'extraction exprime cette variable isolée comme prochaine source de variance maximale qu'on peut extraire. Il en résulte une composante avec une valeur propre proche de l'unité qui peut fort bien ne pas être en continuité avec les valeurs propres suivantes. De telles variables orphelines, quand on les a identifiées, devraient être exclues de l'analyse factorielle, quitte à les réutiliser séparément s'il y a lieu.

Une autre méthode, longtemps préconisée quoique moins généralement disponible, est nommée analyse parallèle (HORN, 1965). Cette méthode est en quelque sorte une sophistication du critère de Kaiser. Au lieu de fixer la valeur critique uniformément à 1, quel que soit le rang (à partir de 1) de la valeur propre, on fixe une valeur critique différente à chaque rang, basée sur des simulations comportant autant de *répondants* et de variables que dans les données analysées, où ces *répondants* sont fictifs et les variables générées sont aléatoires et statistiquement



indépendantes, chacune étant normalement distribuée.

Si l'échantillon avait une taille infinie, l'analyse parallèle engendrerait uniquement des corrélations parfaitement nulles. Les valeurs propres des matrices de corrélations seraient alors toutes égales à l'unité, la valeur critère de Kaiser étant ainsi partout reproduite. Avec des échantillons de taille finie toutefois, le hasard d'échantillonnage de variables toutes indépendantes produit des corrélations habituellement non nulles.⁶ Bien que toutes les dimensions aient la même variance attendue de 1, ce hasard d'échantillonnage pour une taille finie de cas fait en sorte qu'on y trouvera des dimensions offrant des valeurs propres supérieures à l'unité, et d'autres inférieures. On a, en fait, la même situation que pour le test des éboulis après avoir extrait les sources de variance partagées, sauf qu'ici, par construction, on est assuré que la prémisses de variances uniques identiques pour toutes les variables est absolument respectée dans les simulations. Les valeurs propres critiques (ou valeurs de coupure pour chaque rang des composantes principales extraites) sont obtenues par la moyenne (ou la médiane) d'un certain nombre de simulations indépendantes. Ce nombre de simulations importe d'ailleurs peu, sauf lorsqu'une valeur propre observée est très proche de la valeur de coupure et alors, il faudrait des simulations plus nombreuses pour augmenter la précision du seuil établi. On trouve invariablement qu'environ la première moitié des valeurs propres critiques identifiées par l'analyse parallèle est au-dessus de l'unité, un résultat statistiquement attendu. Si on appliquait le critère de Kaiser, plutôt que ces seuils, à des simulations sans aucun facteur partagé, on fixerait en général le nombre de dimensions apparemment requises non pas à zéro (car il n'y a aucun facteur commun) mais à environ la moitié du nombre de variables.

Pendant assez longtemps, on a considéré l'analyse parallèle comme la méthode de choix pour identifier le nombre de dimensions à retenir en AFE (ou en ACP). Toutefois, la justification logique de cette approche est lacunaire, sauf pour la situation où il n'y aurait en jeu aucun facteur commun. Quand il y a véritablement de l'information partagée par les variables, les données pourraient refléter, par exemple, trois composantes telles que les deux premières expliquent beaucoup plus de variance que la troisième et ainsi nettement plus de variance que ce que prédit l'analyse parallèle pour les deux premières dimensions. Cette grande part de variance expliquée par les deux premières dimensions peut donc avoir pour effet d'abaisser la variance restante au-dessous de ce que laissent prévoir les deux premières composantes principales des simulations de l'analyse parallèle. Cela aboutit alors au rejet d'une pos-

sible troisième composante, même en présence d'information partagée encore à expliquer. En ce sens, on peut s'attendre à ce que l'analyse parallèle sous-estime le nombre de dimensions communes exprimées par les données.

Une compréhension inadéquate de l'analyse parallèle a toutefois amené divers auteurs à constater plutôt le contraire, soit une tendance à surestimer le nombre de dimensions, ce qui a conduit à la suggestion de remplacer, à chaque rang, la moyenne des valeurs propres des données simulées par leur 95^e percentile. Cela découlait de ce que, dans toutes les conditions étudiées où l'analyse parallèle appliquée à l'ACP montrait une tendance à surestimer le nombre de dimensions effectivement utilisées pour engendrer les données de test, on n'avait pas pris conscience de la présence de variables orphelines. Ces variables n'étaient pas considérées par les auteurs des études comme devant être comptées parmi les dimensions à conserver. Avec autant de facteur que la moitié du nombre de variables, la composante correspondant à la variable orpheline, de valeur propre proche de l'unité, tend ainsi à dépasser le critère de l'analyse parallèle qui passe sous l'unité au rang correspondant à la moitié du nombre de variables.

Étonnamment, ce n'est que plutôt récemment (publications de 2012) que des approches de modélisation ont été développées pour les techniques d'AFE (incluant l'ACP tronquée) basées sur les moindres carrés, à l'instar de celle utilisant le principe du maximum de vraisemblance. Ce qui manquait n'était pourtant qu'un test pour déterminer si k dimensions *adéquatement identifiées* suffisent comme modèle de l'espace des facteurs communs. Alors que l'analyse parallèle n'implante que le seul modèle nul (où $k = 0$), l'approche de modélisation prescrit que, quelle que soit la valeur de k considérée comme nombre correct de facteurs, on démontre que les valeurs propres des données observées sont typiques de celles émanant du modèle à k dimension(s).

Dans la même année 2012, deux publications indépendantes ont décrit, pour l'ACP ou pour la FAP, des approches basées sur le principe de moindres carrés plutôt que sur la vraisemblance maximale, comportant modélisation pour k facteurs et test statistique afférent. D'une part, GREEN, LEVY, THOMPSON, LU et LO (2012) ont proposé d'utiliser comme modèle d'espace à k dimensions (pour des grandeurs successives de k) les k premières composantes principales soit de la pleine matrice de corrélations, pour l'ACP, soit d'une matrice de corrélations réduite, pour la FAP. Dans les deux cas, la variance de chaque variable générée pour les simulations est complétée à l'unité par une source indépendante de variance unique. Le test que k dimensions sont suffisantes

6. L'erreur-type d'une corrélation de Pearson émanant de deux variables aléatoires indépendantes est égale à $1/\sqrt{n-1}$, quelles que soient les densités (normales ou non) des variables impliquées (cf. ALLAIRE & LAURENCELLE, 2014).



est alors donné par le rang centile de la $k + 1^{\text{ième}}$ valeur propre des données parmi celles correspondantes issues de données simulées produites par le modèle. Les valeurs propres comparées sont issues de la même technique, ACP ou FAP, que celle ayant fourni le modèle à k dimensions. S'il manque même une seule dimension au modèle, les données observées auront leur $k + 1^{\text{ième}}$ valeur propre nettement plus grande que celle de la majorité des simulations à ce rang. Un rang de la $k + 1^{\text{ième}}$ valeur propre des données parmi les 5% plus grandes, i.e. au-delà du 95^e centile de toutes les $k + 1^{\text{ième}}$ valeurs propres disponibles, justifie donc de rejeter l'hypothèse nulle que k composantes soient suffisantes.

Avant d'aller plus loin, revoyons ce qui se passe quand on utilise comme modèle les k premières composantes principales de la pleine matrice de corrélations. On a vu au Tableau 4 que la somme des valeurs propres des deux premières composantes principales dépasse la variance effectivement due aux deux facteurs communs. La raison en est que les quatre sources de variance unique, étant indépendantes, occupent nécessairement les quatre dimensions délimitées par les quatre variables. Il est naïf de croire que les deux premières composantes principales ne contiendraient que la variance due aux deux facteurs communs alors que les deux dimensions restantes rassembleraient tout le bruit. Ainsi, pour le nombre correct k de dimensions, prendre les k premières composantes principales comme modèle produira plus de variance, sur les k premières composantes principales des données simulées, que n'en ont produite, dans les données réelles, les k facteurs à identifier. Il en résulte que la variance restante, à expliquer par les dimensions ultérieures (les valeurs propres de rang $k + 1$ et suivantes), sera moindre dans les données simulées qu'elle ne l'est dans les données observées. Cela crée, pour la $k + 1^{\text{ième}}$ valeur propre des données observées, une nette tendance à être plus grande que celles des nombreuses simulations, d'où un biais à surestimer le nombre de dimensions requises.

La même année 2012, RUSCIO et ROCHE (2012) publient une approche par modélisation qui n'utilise que des facteurs issus de l'AFE comme modèle des données. Leur test est basé sur la comparaison d'ensemble des valeurs propres (issues de l'ACP) des données observées à celles issues de chacun des jeux de données simulées. La statistique sur laquelle se base leur test est, pour chaque jeu de données simulées, la moyenne quadratique des différences de leurs valeurs propres avec celles des données (pensez *variance d'erreur*). La technique a été optimisée pour 500 jeux de données simulés pour chaque hypothèse quant au nombre de dimensions présentes. Tant que l'ensemble

de 500 différences quadratiques moyennes pour $k + 1$ dimensions est significativement différent (moindre) de celui des 500 différences quadratiques moyennes issues du modèle à k dimensions, k est considéré comme insuffisant et augmenté de 1. Un test de rangs est utilisé pour comparer deux ensembles de 500 différences quadratiques moyennes pour des modèles qui diffèrent par une dimension. Pour réduire la tendance de cette procédure à surestimer le nombre de facteurs, les auteurs ont préféré fixer le seuil alpha de leur test de rang à 0,3 (en mode bidirectionnel).

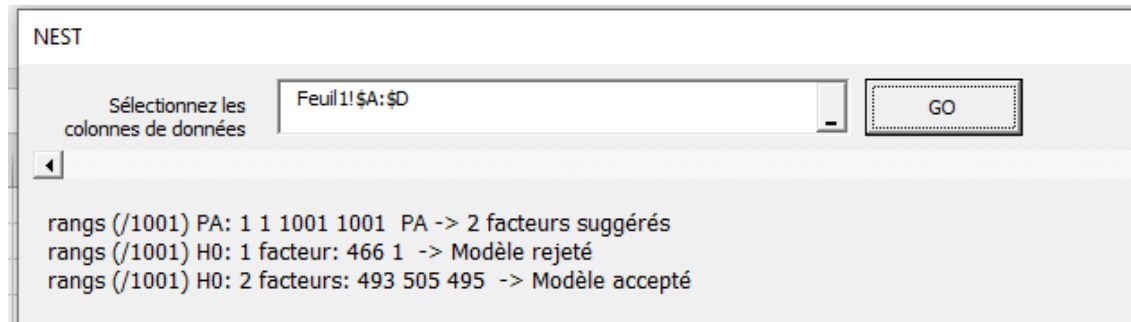
Bien que la version de modélisation de RUSCIO et ROCHE (2012) utilise la FAP, on note deux différences pertinentes entre les deux approches, outre le critère statistique utilisé pour fixer le nombre de dimensions. La première est que Ruscio et Roche appliquent leur test aux valeurs propres issues de la pleine matrice de corrélations alors que Green et collègues utilisent celles de la matrice réduite de corrélations (après que le modèle ait été produit par la FAP). La seconde différence est plus subtile et porte sur la façon de réduire la matrice de corrélations. Green et collègues la réduisent simplement en substituant les valeurs de R^2 , facilement calculables,⁷ dans la diagonale de la matrice de corrélations. Ceci s'avère beaucoup plus rapide que l'estimation itérative pour le calcul des valeurs propres issues des simulations avec lesquelles sont comparées les valeurs propres de la matrice réduite des valeurs observées. De leur côté, Ruscio et Roche utilisent l'approche itérative qui tient compte du nombre de dimensions faisant l'objet de l'hypothèse courante quant à ce nombre. Puisqu'ils comparent ensuite les données réelles à celles simulées quant aux valeurs propres issues de l'ACP et non de la FAP, la réduction de matrice n'est effectuée que pour produire le modèle générateur pour chaque valeur de k .

NEST : Next Eigenvalue Sufficiency Test. Dans une variété de structures factorielles testées, ACHIM (2017) a comparé entre elles huit méthodes de fixation du nombre de dimensions requises. Ces méthodes incluaient l'analyse parallèle, l'approche par vraisemblance maximale, celle de RUSCIO et ROCHE (2012) et quatre versions dérivées de l'approche de GREEN et collègues. (2012) où le test que k dimensions sont suffisantes provient du rang de la $k + 1^{\text{ième}}$ composante principale des données observées parmi celles correspondantes issues de données engendrées par un modèle à k composantes. Ces quatre variantes portent collectivement le nom de « Next Eigenvalue Sufficiency Test » (NEST). Elles incluent la version originale de Green et collègues, avec réduction des matrices de corrélations par les R^2 , autant pour établir de modèle que pour l'analyse des données

7. Tel que mentionné plus haut, il s'agit des coefficients de détermination de la régression de chaque variable à partir de toutes les autres de la matrice de données, le calcul se faisant tel qu'indiqué en note 5.



FIGURE 2 ■ Interface de NEST.xlsx donnant le résultat de l'analyse de 200 cas engendrés selon le modèle du Tableau 1.



simulées. Une variante utilise aussi l'évaluation du modèle via les R^2 mais met en rang les valeurs propres issues des pleines matrices de corrélations autant des données observées que de celles des données simulées. Les deux autres variantes utilisent les mêmes mises en rangs de valeurs propres issues soit de matrices réduites (par les R^2), soit de matrices pleines, mais elles estiment le modèle par la méthode itérative tenant compte du nombre de dimensions dont on cherche à vérifier la suffisance. Globalement, les quatre variantes NEST surpassent très nettement les autres méthodes, même celle de vraisemblance maximale. Le meilleur taux de succès est obtenu par la variante qui utilise la méthode itérative pour estimer les communautés en vue d'établir le modèle des données, mais qui met en rangs les valeurs propres issues des pleines matrices de corrélations. Alors que les autres variantes ont occasionnellement donné, sous certaines conditions liées à la plus petite valeur propre à conserver, plus que leur taux nominal de 5% de faux rejets du nombre correct de dimensions, la version gagnante n'a pas dépassé le taux alpha nominal quel qu'était le niveau de la plus petite valeur propre à ne pas ignorer. Autrement dit, son taux de surestimation du nombre de facteurs ne dépassait jamais le seuil d'erreur du test.

C'est à la variante gagnante mentionnée ci-dessus qu'est désormais attribué le nom de méthode NEST. Pour la rendre facilement accessible, un lien est fourni dans l'article (ACHIM, 2017) vers du code approprié en MATLAB^{md} et en langage R de même que vers un fichier Excel (NEST.xlsx) implantant NEST en macros cachées (supportées sous Windows mais dans peu de versions sur MacIntosh). La Figure 2 montre l'interface de NEST.xlsx appliquée aux données engendrées par le modèle du Tableau 1 pour $N = 200$.

La première ligne de sortie teste un modèle nul et est donc équivalente à l'analyse parallèle (PA dans la figure). Elle donne le rang des valeurs propres des données parmi les 1001 valeurs propres de même rang disponibles, dont

1000 viennent de données engendrées avec le modèle à aucun facteur, avec autant de tels rangs qu'il y a de variables, ici 4. Si le rang obtenu par la première valeur propre ne se situait pas à 50 ou moins (signifiant $p < .05$), il faudrait conclure qu'aucun facteur n'est nécessaire. La procédure se poursuit en testant l'hypothèse que 1 facteur serait suffisant (« H0 : 1 facteur » dans la figure). On y voit que la première valeur propre des données arrive ici au rang 466 parmi les 1001 premières valeurs propres disponibles, tel que prévu du fait que l'ensemble de simulations a déjà inclus le (ou, éventuellement les facteurs, pour $k > 1$) précédemment accepté(s). On voit surtout que ce modèle n'épuise pas l'information partagée par les variables puisque la deuxième valeur propre des données (donc ici celle au rang $k + 1 = 2$) occupe le plus haut rang (=1) parmi les 1001 telles valeurs propres, ce qui signifie $p < 0,001$ (en fait $p = 1/1001$). Enfin, puisque le modèle à un facteur a été rejeté, la troisième ligne teste le modèle à deux facteurs et indique que celui-ci est tout à fait acceptable, les valeurs propres des données jusqu'à la $k + 1$ ème (ici la 3e) inclusivement étant quelconques parmi toutes celles disponibles. Ce modèle est donc accepté.

Plus récemment, une macro pour SPSS a été développée pour l'application de NEST. On peut le retirer par le lien NEST <https://www.tqmp.org/RegularArticles/vol16-4/p213/>. Dans un fichier de syntaxe, on commence par activer la macro par : « INSERT [...chemin.../]NEST.sps. », en n'oubliant pas le point final. Ensuite, on appelle simplement NEST suivi de la liste des variables séparées par des espaces, où une suite de variables consécutives dans le tableau de données SPSS peut être spécifiée par ses extrêmes séparés par « TO », p. ex. « VAR1 VAR3 TO VAR7. », encore en n'oubliant pas de terminer la ligne par un point. Il n'y a plus qu'à faire exécuter ces deux lignes. La sortie contient d'abord des informations textuelles sur la procédure et comment y faire référence. La suite, pour les données avec $N = 200$,



est montrée dans l'Encadré 1. S'il y a lieu, les cas de Heywood rencontrés en estimant les communautés pour un modèle à k facteurs sont rapportés. Le '1' à la ligne qui suit l'avertissement de cas de Heywood complète en fait la phrase précédente (disant ici que le cas de Heywood fut observé pour le modèle à 1 facteur), la procédure de programmation SPSS ne permettant pas de l'accoler directement au texte qui précède. Après le signalement d'un cas de Heywood, les deux séries de communautés sont rapportées; la première donne des valeurs où une communauté dépasse l'unité, la deuxième les communautés de l'itération précédente utilisées pour le modèle. On voit ici que c'était la variable B qui était en cause. Rapporter les cas de Heywood vise à attirer l'attention sur un problème potentiel, mais on n'a généralement pas à s'en inquiéter, surtout si le cas est apparu en élaborant un modèle autre que celui qu'on va choisir de retenir.

La prochaine section de sortie de NEST.sps donne le test d'un nombre croissant de facteurs comme hypothèse nulle. L'entrée critique pour chacune des hypothèses est celle dont le rang est exprimé en négatif (simplement pour le marquer). Contrairement à l'implantation de NEST pour Excel, les tests d'hypothèses consécutives ne s'arrêtent que lorsque la probabilité critique est supérieure à 0,25. Cela est destiné à montrer les cas où une hypothèse est acceptée malgré une probabilité pas beaucoup supérieure à 0,05, ce qui pourrait représenter un manque de puissance à détecter la présence d'un facteur de plus. L'utilisateur peut alors décider d'examiner aussi la légitimité éventuelle du facteur surnuméraire. Des travaux autres que très préliminaires restent toutefois à faire pour confirmer que les dimensions non détectées selon le critère 0,05 engendrent des rangs plus proches du 50 critique que du 50 attendu sous H_0 . La dernière ligne de la sortie donne le nombre de dimensions à retenir selon un seuil d'erreur alpha de 0,05. Encore ici, le nombre apparaît sur une ligne séparée du texte qu'il complète.

Ainsi, la procédure NEST disponible sous plusieurs formes permet d'estimer le nombre de dimensions à retenir, plus précisément le nombre de facteurs qui semblent nécessaires et suffisants pour rendre compte des corrélations observées entre les variables. Notons toutefois qu'une erreur de type 1 (faux positif) ferait retenir une dimension de trop et qu'une erreur de type 2 (faux négatif) ferait manquer une dimension pourtant présente dans les données. Ainsi, si NEST permet d'identifier que k dimensions semblent, d'une part, *nécessaires*, puisque l'hypothèse de $k - 1$ a été rejetée et, d'autre part, *suffisantes*, puisque cette hypothèse de k n'a pas été rejetée, il faut rester ouvert à des solutions ayant un nombre différent de dimensions. Rappelons toutefois que les faux rejets par NEST sont systématiquement moindres que le taux nomi-

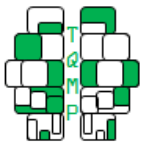
nal de 5%. Si l'hypothèse de k facteurs a été rejetée avec un rang de 1/1001, l'erreur de type 1 paraîtra improbable, bien que la possibilité demeure. Une erreur possible de type 2 pourrait être suggérée par un rang de la $k + 1^{\text{ième}}$ valeur propre observée pas très nettement au-dessus de 50/1001, ou du moins pas nettement proche du rang médian. Dans les deux applications ci-dessus, NEST.xlsx et NEST.sps, le rang 1 (pour la deuxième valeur propre) menant au rejet de l'hypothèse d'un seul facteur, de même que les rangs 495 (Excel) ou 462 (SPSS) pour la troisième valeur propre menant à accepter que deux facteurs suffisent laissent peu à redouter une erreur de type 1 ou de type 2. Toutefois, tout simple d'utilisation et tout utile que soit NEST, il ne faut jamais renoncer ni à réfléchir ni à rester vigilant quant à la solution qu'on s'appête à accepter.

Interprétation de l'espace factoriel

La fixation du nombre de dimensions à retenir et l'extraction par la FAP des axes définissant ces dimensions constituent les deux premières étapes de l'AFE. Il reste à interpréter cet espace factoriel, puis éventuellement, à estimer les scores factoriels selon la solution retenue. Concernant l'interprétation de l'espace factoriel, le Tableau 7 permet de comparer les communautés de la solution de FAP (de SPSS) pour l'échantillon de taille $N = 200$, avec celles du modèle générateur et celles de l'ACP. Dans SPSS, les communautés sont données sous l'étiquette « Qualité de représentation ». Le tableau 7 confirme en particulier que l'ACP surestime systématiquement les communautés.

Avant d'aller plus avant, récapitulons un peu. On a des variables corrélées entre elles et on conceptualise que l'information partagée par ces variables vient de ce que celles-ci expriment la variance de facteurs communs. On peut, en particulier par la procédure NEST, identifier combien de facteurs communs sont requis, au minimum, pour rendre compte du patron de corrélations. L'espace de ces facteurs est généralement bien approximé par celui des k premières composantes principales de la matrice réduite de corrélations. Toutefois la méthode d'extraction reposant sur des critères purement mathématiques, les axes initiaux servant à délimiter l'espace des facteurs réels ne correspondent généralement pas à ces facteurs. L'exception principale à cette réalité est triviale : si les données ne reflètent qu'un seul facteur, la question de rotation ne se pose pas et la dimension identifiée approxime déjà directement cet unique facteur.

Coordonnées orthogonales ou obliques. Une fois l'espace factoriel identifié, il reste donc à le décrire comme un système de coordonnées significatives. On réservera ici le terme *axe* pour un ensemble de directions mutuellement orthogonales. Une rotation orthogonale est en fait la substitution d'un nouveau système d'axes à l'ancien, sans



Encadré 1 ■ Sortie partielle de la procédure NEST . sps pour les données avec $N = 200$.

```
Run MATRIX procedure:
Heywood case observed. Communalities returned from (next) to (following) output
  lines in preparing model with nb of factor(s)=
  1
T (COMM)
  ,733215013  1,019110721  ,063310714  ,547031784
T (COMM)
  ,7525626178  ,9898799292  ,0639111524  ,5590836330
Ranks (/1001) of data eigenvalues among those of surrogate data. Critical ranks
  given as negative
H0:0 fct   -1    1  1001  1001
H0:1 fct   570   -1  1001  1001
H0:2 fct   527   493 -462   408
NEST indicates that the necessary and apparently sufficient number of factors is
  2
----- END MATRIX -----
```

Tableau 7 ■ Contributions des deux axes principaux, C1 et C2, avant rotation aux quatre variables pour l'échantillon de 200 cas, accompagnées de leurs communautés selon trois points de vue.

	Axe C1	Axe C2	Communauté FAP	Communauté population	Communauté ACP
A	0,8879	0,3049	0,8814	0,8677	0,9424
B	0,9658	-0,1830	0,9663	0,9421	0,9577
C	0,3639	0,8956	0,9346	0,9411	0,9716
D	0,8152	-0,5151	0,9300	0,9408	0,9685
SC	2,5182	1,1940	3,7123	3,6916	3,8401

déplacer l'origine. Ce nouveau système est le point d'arrivée d'une rotation rigide des anciens axes optimisant un certain critère. Quand on considèrera la possibilité de facteurs corrélés, on parlera simplement de *directions*, un terme plus général. La rotation n'est alors plus contrainte de préserver les angles droits qui caractérisent les axes.

Lorsqu'un espace est délimité par des axes, le théorème de Pythagore s'applique : le carré de la longueur de la partie d'une variable dans cet espace (sa variance expliquée par l'hyperplan) est la somme des carrés de ses saturations sur les divers axes. Ceci reste vrai quel que soit le système d'axes choisi et n'est donc pas affecté par une rotation orthogonale. De plus, si on suppose que les variables ne partagent d'information que dans cet espace, la corrélation entre deux variables est tout simplement la somme des produits de leurs coordonnées sur l'ensemble des axes, quel que soit le choix d'axes.

Il n'en va pas de même pour des directions corrélées, c.-à-d. mutuellement obliques dans l'espace des répondants. Lorsque les parties expliquées des variables sont plutôt présentées comme des sommes pondérées de directions

corrélées, tout ceci se complique pour les variables qui ne sont pas parfaitement alignées sur un unique facteur et dont on voudrait calculer la corrélation avec d'autres variables à partir des contributions des divers facteurs à chacune. Il n'est pas nécessaire d'entrer dans ces considérations ici, surtout qu'on sait que les relations entre les variables dans l'espace factoriel retenu, tant la longueur des variables que les angles qui les séparent, ne changent pas. La rotation consiste simplement à coller une nouvelle grille d'interprétation sur laquelle lire des coordonnées des variables, lesquelles restent évidemment immobiles. Ces coordonnées ne sont pas la projection de la variable sur le facteur, qui serait le point d'où part une perpendiculaire au facteur passant par la variable; elles sont la longueur qu'on doit parcourir dans la direction de chaque facteur pour atteindre la variable. Cela correspond à la recette de fabrication de la variable selon cette solution factorielle.

Ainsi, par exemple, un plan peut être redéfini par un système de deux axes, lesquels peuvent être une rotation des deux composantes principales qui l'ont initialement



défini. Des facteurs corrélés ne forment pas un système d'axes. On peut toutefois décrire le même plan à partir des deux directions corrélées correspondant à une hypothèse sur les facteurs à identifier. La notion de recette de fabrication continue à s'appliquer : en termes graphiques, chaque variable est encore la somme de deux directions, pondérées par leurs saturations respectives. Les directions elles-mêmes sont généralement exprimées comme des vecteurs de longueur unitaire. Ainsi, si les coordonnées de la partie reproduite d'une variable sur deux facteurs corrélés, désignons-les A et B , sont 0,4 et 0,7 alors la variable est décrite comme $0,4A + 0,7B$, ce qui veut dire que pour s'y rendre à partir de l'origine, on fait un pas de longueur 0,4 dans la direction A ainsi qu'un autre de longueur 0,7 dans la direction B . La longueur depuis l'origine au point ainsi atteint sera plus grande que la racine carrée de la somme des carrés de 0,4 et de 0,7 si l'angle entre les facteurs est aigu (pour une corrélation positive) ou moins que cela s'il est obtus (pour une corrélation négative des facteurs⁸).

Structure simple. Les rotations ont pour but de mieux localiser et identifier les entités abstraites, les facteurs, qui ont présidé à la fabrication de la part de variance des différentes variables due à des facteurs communs. Que cherche-t-on au juste? La réponse est que cela dépend de la situation. Souvent, il est raisonnable de penser que la plupart des variables n'expriment qu'un seul des facteurs impliqués. En analyse factorielle confirmatoire, ces variables sont dites des variables indicatrices. S'il y a des corrélations entre de telles variables indicatrices associées à des facteurs distincts, cela est alors attribué à une corrélation entre les facteurs eux-mêmes. Dans ces cas, en AFE, la rotation cherchera à identifier des directions dans l'espace factoriel qui passent près de chacune des variables reflétant un même facteur, levant la contrainte d'orthogonalité pour ces directions.

Structure pas toujours simple. Il se peut aussi que plusieurs des variables présentes reflètent simultanément les mêmes sources, plus ou moins indépendantes, d'information. C'est vraisemblablement le cas pour des questions d'examen qui reflètent toutes, mais dans des proportions variées, à la fois de la facilité à comprendre la matière et de l'ardeur déployée à l'étudier. On ne s'attend alors pas à ce qu'aucune de ces questions reflète seulement l'une ou seulement l'autre de ces deux sources de succès à l'examen. Pour fixer correctement les directions des facteurs, il faudrait probablement définir des points d'ancrage, c'est-à-dire des questions dont on prédéfinit ex cathedra la contribution de chacune des sources de succès, ou encore

établir a priori une mise en ordre attendu des items sur chacune des dimensions. En général, la fixation d'axes significatifs pour de telles décompositions factorielles s'annonce difficile, sinon discutable. Les méthodes heuristiques décrites ci-après peuvent toutefois aider à atteindre une interprétation crédible. On retiendra ici qu'une solution se conformant au modèle d'une structure simple n'est pas toujours à attendre.

On cherche en général à décrire l'espace factoriel par des directions qui correspondent à des variables internes ou facteurs vraisemblables. Ces facteurs sous-tendent le même espace que celui initialement identifié par les composantes principales de la matrice de corrélations réduite. Il s'agit en fait de chercher à identifier ce qu'on a appelé plus tôt les axes naturels, soit les véritables variables internes (et les recettes de fabrication auxquelles elles contribuent) qui auraient présidé à forger les grandeurs associées aux variables observées dans l'échantillon. Comme on l'a déjà expliqué, les nouvelles directions choisies pour décrire les variables dans l'espace identifié ne changeront ni les communautés (parts de variance expliquées par les facteurs) ni les corrélations reproduites. Elles changeront toutefois les relations entre les facteurs et les variables, soit, en particulier, les recettes présumées par lesquelles la *nature* convertit des grandeurs non observables sur les facteurs en mesures observables sur les variables. Ces directions retenues déterminent en retour les valeurs qu'on peut attribuer aux participants sur les facteurs présumés.

La recherche de directions plus significatives à considérer comme facteurs n'a de solution unique correcte que dans des approches dites de décomposition trimodale, parfois appelées analyse factorielle parallèle (à ne pas confondre avec l'analyse parallèle de Horn), où les mêmes facteurs s'expriment avec les mêmes recettes de fabrication dans au moins deux groupes, mais avec des profils différents de variance des facteurs dans les groupes. Ces analyses simultanées de plus d'une matrice de corrélations sont toutefois peu courantes et surtout peu accessibles à la majorité des chercheurs.

Retenons pour le moment que l'analyse d'une seule matrice de corrélations (ou de covariances) n'apporte pas suffisamment de contraintes pour que toute solution mathématiquement unique soit nécessairement le reflet correct de la réalité sous-jacente. Les contraintes ajoutées pour rendre une solution mathématiquement unique n'ont souvent rien à voir avec le phénomène étudié. C'est notamment le cas pour l'identification initiale de l'espace factoriel, dans laquelle la contrainte est d'extraire successivement le maximum de variance laissée inexpliquée par les

8. Pour cet exemple, le lecteur pourra vérifier que, sous un angle aigu (corrélation positive) de 60° dont le cosinus est 0,5, le carré de la longueur du vecteur de la variable est $0,4^2 + 0,7^2 + 2 \times 0,4 \times 0,7 \times \cos(60) = 0,93$, et de $= 0,37$ pour un angle obtus de 120° (corrélation négative; cosinus $= -0,5$), le cas orthogonal (90°) donnant 0,65.



composantes ou facteurs précédemment extraits. Parfois, les contraintes ajoutées pour proposer les directions des facteurs dans l'espace factoriel offrent des chances raisonnables que l'interprétation factorielle corresponde d'assez près à la réalité. C'est le cas de diverses procédures de rotation qui ont été développées pour l'AFE. Toujours, l'acceptabilité conceptuelle d'une solution proposée par un algorithme de rotation doit demeurer un critère pour son acceptation.

Parmi les contraintes qu'on peut inclure dans une procédure de rotation, il y a celle que les facteurs soient indépendants les uns des autres, c.-à-d. orthogonaux, sans corrélation entre eux. De toutes façons, si l'espace factoriel contient effectivement k dimensions, il existe nécessairement au moins k sources indépendantes de variance impliquées. C'est notamment le cas dans les simulations numériques, où il faut au moins deux sources de valeurs indépendantes pour engendrer deux variables imparfaitement corrélées. Cette nécessité d'autant de sources de variance que de dimensions/facteurs pourrait déjà justifier une contrainte d'orthogonalité des facteurs. Toutefois, la compréhension optimale du partage d'information entre les variables ne se situe pas ordinairement à ce niveau.

Parfois, en effet, la compréhension pourra être plus grande, ou plus directe, si on admet des facteurs qui peuvent être corrélés entre eux. Mais pour que deux facteurs soient à la fois distincts et corrélés, il est nécessaire qu'ils soient eux-mêmes répondants à des recettes de fabrication qui leur font partager de l'information initialement indépendante. Autrement dit, si on développe une solution où les facteurs sont corrélés, il doit aussi exister une méta-solution factorielle en proto-facteurs orthogonaux pour expliquer les corrélations admises entre les facteurs. Le niveau de solution qui contribue le mieux à la compréhension des relations entre les variables observées devrait dépendre de l'état courant des connaissances dans le domaine.

En principe, si les données s'interprètent mieux sous une hypothèse de facteurs corrélés, on voudra identifier cette hypothèse de facteurs corrélés. Souvent, on ne pourra que spéculer sur les sources indépendantes de variation sous-tendant ces facteurs. Il est toutefois plutôt rare que des solutions en facteurs corrélés soient commentées quant à la source ou l'explication de leurs corrélations. En exemple, si on trouvait qu'une corrélation négative entre la facilité à comprendre la matière et l'ardeur à étudier aide la compréhension de leur expression dans les items d'un examen, on pourrait spéculer que l'ardeur à étudier se soit davantage développée chez ceux dotés de moins de facilité à comprendre la matière mais motivés à réussir, sans quoi ils ne seraient pas inscrits dans la population d'où provient l'échantillon mesuré. Les lecteurs devraient

pouvoir se référer à leur propre expérience là-dessus, la plupart d'entre eux ayant plutôt dû faire du rattrapage quant à leur ardeur au travail, leurs capacités intellectuelles ayant assuré leur réussite dans la première partie de leur cursus d'études !

Bien qu'on ne puisse être assuré que la solution factorielle de conformera toujours à une structure simple, on rencontrera de nombreuses situations où la plupart des variables devraient chacune refléter principalement un seul facteur. Les heuristiques de rotation devraient s'avérer utiles et efficaces. Une heuristique est une procédure de recherche de solution donnant généralement, mais pas nécessairement, le bon résultat. Dans le cas des rotations, les heuristiques visent à identifier un nouvel ensemble de directions qui ont une chance raisonnable de bien correspondre aux axes naturels des variables (les facteurs réels qui s'expriment dans les variables et causent leurs corrélations).

Rotations orthogonales

On appelle rotation orthogonale une rotation qui exprime l'espace initialement délimité par des composantes principales (donc orthogonales en vertu de la méthode de calcul) par un nouvel ensemble d'axes orthogonaux (pivotant autour de la même origine) : on peut voir cela comme une rotation rigide des axes deux par deux dans l'espace original. Comme on l'a déjà exprimé, toute rotation, orthogonale ou oblique, ne change rien aux relations objectives des variables entre elles ; on fait simplement le choix d'une nouvelle grille de référence pour lire leurs coordonnées. Imaginez, par exemple, un objet rigide (un système d'axes) pointant dans les trois directions de la pièce où vous vous situez, par exemple les arêtes d'une petite boîte. Par rapport au point de rencontre de ces trois directions, on peut lire des coordonnées, par exemple, celles du bout de votre nez. Faites tourner cet objet sans changer le point de rencontre de ses trois directions. Votre nez a maintenant de nouvelles coordonnées sur ces nouvelles directions sans que votre nez ni rien d'autre n'aient bougé dans la pièce (excluant le bras qui a fait tourner le système de référence, mais une rotation mathématique se passe du bras !).

Il y a moins de paramètres à ajuster pour une telle rotation dite rigide que pour un changement de système de référence permettant aussi de changer les angles entre les directions sur lesquelles on pourrait lire les coordonnées des objets. On peut commencer par obtenir une approximation orthogonale d'une solution dont on accepterait toutefois que les facteurs soient corrélés. Il se peut même que la solution orthogonale soit déjà en elle-même satisfaisante. Une solution orthogonale est intéressante d'une part parce qu'elle ne pose pas explicitement la question d'une méta-compréhension de la corrélation des facteurs (c.-à-d.



une explicitation des corrélations entre ceux-ci) et d'autre part parce que la variance des variables est alors comprise comme la somme de sources indépendantes de variances, sans correction par des termes d'interaction entre les facteurs.

La rotation orthogonale qui est généralement la plus satisfaisante est la rotation Varimax. Elle identifie de nouveaux axes orthogonaux du sous-espace retenu (axes qu'on appellera facteurs et non plus composantes principales) tels que chaque facteur contribue soit beaucoup, soit peu à chacune des variables. Cela singularise des facteurs qui s'exprimeraient le plus purement possible dans un petit nombre de variables (si on ne garde que les pondérations les plus élevées). Cette qualité, que les contributions des facteurs aux variables aient leur carré s'approchant soit de 1, soit de 0, avec le moins possible de valeurs intermédiaires, correspond à un critère objectif de VARIance MAXimale des (carrés des) pondérations, d'où le nom de la rotation. Ces variances à maximiser sont calculées pour chaque facteur sur l'ensemble des variables, puis additionnées sur les facteurs, comme si les variables étaient des répondants auxquels on cherche à donner le profil le plus différencié possible. Si on maximisait plutôt la variance de chaque variable sur l'ensemble des facteurs, on appliquerait une rotation *Quartimax* (qui est souvent moins satisfaisante). Le résultat de la rotation Varimax sur les données de l'échantillon de 200 cas est rapporté au Tableau 8.

En comparant avec le Tableau 7, on voit que les communautés FAP au Tableau 8 sont restées les mêmes, mais que la variance est répartie différemment entre les deux axes. On peut aussi remarquer que, malgré que la rotation soit orthogonale, la somme des produits des contributions des deux facteurs à l'ensemble des variables ne donne plus zéro, comme c'était le cas avant la rotation. Elle est ici de 0,132. Autrement dit, si on pouvait considérer les variables comme indépendantes (mais elles ne le sont pas, par nature, en AFE), cela signifierait que les nouveaux axes ne seraient pas orthogonaux dans un espace dont les variables seraient les axes. La rotation est plutôt orthogonale en ce sens qu'elle garde les facteurs orthogonaux *dans l'espace initial des répondants* (qui sont théoriquement indépendants les uns des autres). Seules les composantes principales (de matrices pleines ou réduites) ont une somme nulle de produits des saturations de chaque paire de facteurs sur l'ensemble des variables. Ainsi, seules les composantes principales sont orthogonales à la fois dans l'espace des répondants et dans celui des variables. L'intérêt que les facteurs demeurent orthogonaux dans l'espace des répondants, si leur interprétation est acceptable, c'est que les scores des répondants seront indépendants (corrélations nulles) d'un facteur à l'autre.

Le Tableau 9 rappelle les véritables pondérations par

lesquelles les données ont été engendrées, excluant les unités complétant de variance à 1. Malgré qu'on note que la solution Varimax donne à V2 un poids négatif de -0,2225 pour la variable B alors que F2, le véritable facteur, y contribue avec un poids positif de 0,1204, il demeure assez remarquable qu'on arrive si près des véritables *recettes de fabrication*, considérant que ce modèle n'avait pas ce qu'on appelle une structure simple. Rappelons que la structure simple est définie par divers critères, mais suppose surtout qu'un facteur contribue soit beaucoup, soit très peu à une variable et qu'une variable dépend soit beaucoup soit très peu d'un facteur. La variable A ne se comportait pas comme cela, recevant des parts de variance comparables des deux facteurs. Autrement, dit, dans notre exemple avec des recettes de fabrication des variables passablement arbitraires, il n'y avait pas lieu de s'attendre à retrouver une solution de rotation qui corresponde à la recette, celle-ci ne se qualifiant pas de structure simple (sauf partiellement par le fait que la variable C ne reçoit d'information que du facteur F2).

On se souviendra que les rotations sont des heuristiques. Elles appliquent des principes qui ont une chance raisonnable de bien marcher, mais cela ne garantit pas de représenter correctement ce que la nature a fait quand les mesures ont été prises.

Rotations obliques.

Pour discuter de l'opportunité de recourir à une rotation oblique, il est utile de considérer les pondérations faibles dans une solution de rotation orthogonale. SPSS, comme d'autres logiciels, offre l'option de masquer les pondérations de valeur absolue inférieure à un certain seuil. Faire usage de cette option n'est pas judicieux. Pour qu'il soit logique de considérer les pondérations faibles comme nulles dans la population, celles-ci devraient au moins avoir des polarités aléatoires dans la solution factorielle. La présence d'un patron discernable dans les pondérations faibles sur une paire de facteurs est signe de l'utilité d'une rotation oblique, ainsi qu'on peut en juger dans la situation caricaturale de la Figure 3, où deux facteurs effectivement corrélés s'expriment chacun dans quatre variables.

Dans cette illustration, la rotation Varimax donne les deux axes orthogonaux tracés en noir. Chaque variable a une coordonnée élevée sur un axe et une coordonnée faible sur l'autre. Ici, pour des variables issues de deux facteurs sous-jacents corrélés positivement (l'angle d'écartement étant aigu), non seulement toutes les pondérations fortes (sur leur facteur dominant), mais aussi toutes les pondérations faibles (sur l'autre facteur) sont de polarité positive dans la solution orthogonale. On pourrait, dans d'autres cas, avoir les pondérations fortes d'une pola-



Tableau 8 ■ Résultat de la rotation Varimax vers les axes V1 et V2 de la solution à deux facteurs pour l'échantillon avec $N = 200$.

	V1	V2	(communauté)
A	0,6893	0,7374	(0,8814)
B	0,9575	-0,2225	(0,9663)
C	-0,0286	0,9663	(0,9346)
D	0,9538	-0,1422	(0,9300)
SC	2,3024	1,4098	(3,7123)

Tableau 9 ■ Rappel des contributions réelles des facteurs aux variables (Tableau 2) dans la population.

Var	F1	F2
A	0,7274	0,5819
B	0,9631	0,1204
C	0,	0,9701
D	0,9249	-0,2921

rité et faibles de la polarité opposée, indiquant une structure simple pour des facteurs corrélés négativement.

Après une rotation orthogonale, il est donc utile d'inspecter même les saturations faibles, une paire de facteurs à la fois. Si deux facteurs s'expriment sur les mêmes variables formant deux sous-ensembles, chacun affectant assez fortement un sous-ensemble et faiblement l'autre, on pourrait être tenté d'ignorer les saturations faibles pour n'associer chacun des facteurs qu'avec son sous-ensemble à fortes saturations. Faisant cela, on considérerait implicitement les saturations faibles comme nulles dans la population. Si c'était le cas, toutefois, la polarité de ces saturations faibles dans l'échantillon devrait être aléatoire. Présumer ceci sera toutefois peu raisonnable si on peut constater que les saturations faibles des deux sous-ensemble de variables sont très majoritairement de même signe. Avant d'ignorer et de traiter comme nulles dans la population ces faibles saturations dans l'interprétation des facteurs, on peut envisager qu'une solution plus simple et plus crédible soit possible en admettant des facteurs obliques. Par exemple, dans la Figure 3, nettement caricaturale, on a quatre variables proches de l'axe vertical, mais toutes avec un poids positif sur l'axe horizontal. On a aussi l'autre groupe, proche de l'axe horizontal, dont tous les poids sur l'autre axe sont aussi positifs. En permettant aux deux facteurs d'être corrélés, on annulerait les poids faibles de chacun des facteurs sur les variables exprimant l'autre facteur.

Dans des cas réels, les variables ne seraient pas aussi bien alignées sur leurs directions obliques, mais si (presque) toutes les saturations faibles des variables fortement influencées par l'autre facteur de la paire sont de même signe, une rotation oblique rendra ces saturations non seulement plus petites en valeurs absolues, mais aussi,

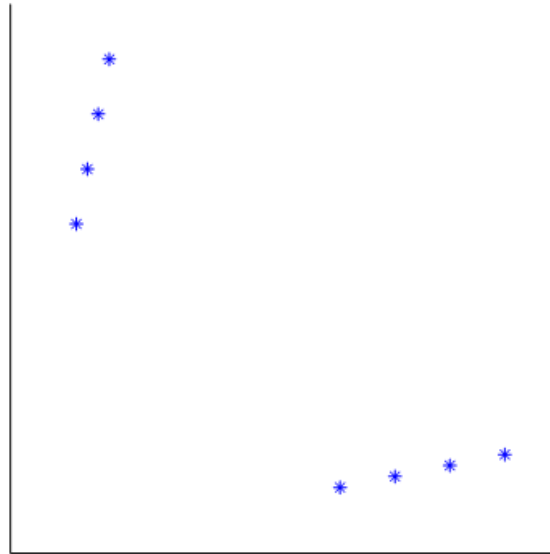
vraisemblablement, de polarités plus aléatoires. Il deviendra alors raisonnable de considérer ces pondérations faibles pour les facteurs obliques comme des corrélations aléatoires (entre facteur et variable) qui seraient vraisemblablement nulles dans la population. Nous reviendrons plus loin sur les questions d'interprétation, notamment celle de décréter comme nulles dans la population des pondérations plutôt faibles.

Ainsi, souvent pour bien comprendre la structure des données, on voudra opérer à un niveau conceptuel qui admet des corrélations entre les facteurs : on utilisera alors une rotation oblique. Il y a plusieurs algorithmes de rotations obliques disponibles, quoique SPSS n'en offre que deux, désignés Oblimin direct et Promax. Certains auteurs proposent de considérer les résultats de diverses rotations obliques ou de choisir pour elles des paramètres de contrôle autres que ceux proposés par défaut, afin d'être alerté de diverses possibilités d'interprétation si leurs résultats diffèrent, c'est-à-dire si leurs critères respectifs de structure simple, mettant l'accent sur des aspects différents, amènent des propositions d'interprétation qualitativement distinctes. On propose aussi parfois de commander d'emblée une solution de rotation oblique, puis d'examiner les corrélations entre facteurs afin de décider si celles-ci semblent ou non compatibles avec des corrélations qui soient nulles dans la population. La première recommandation pourrait s'avérer fructueuse si des solutions nettement distinctes en émanaient. Mais commander d'abord une rotation Varimax permet de justifier, par l'examen des saturations faibles, qu'une rotation orthogonale n'est pas suffisante.

Lorsqu'une rotation oblique a été commandée, quatre matrices sont habituellement rapportées par SPSS. L'une d'elles est désignée « matrice des facteurs » ; c'est la solu-



FIGURE 3 ■ Représentation caricaturale d'une solution de rotation orthogonale qui appelle une rotation oblique.



tion factorielle avant la rotation. Une autre est désignée « matrice de forme » (en anglais, pattern matrix) et donne les saturations des variables sur les facteurs; c'est là la recette de fabrication de la partie commune des variables observées (reproduisant leurs communautés) selon la solution factorielle finalisée par la rotation. La communauté d'une variable est plus facilement obtenue à partir de l'extraction orthogonale primaire par ACP de la matrice réduite de corrélations, puisque c'est exactement la même communauté qu'avant toute rotation. Ce que la rotation change, c'est le profil hypothétique de l'influence des sources d'information des facteurs maintenant corrélés sur chaque variable, autrement dit, les recettes de fabrication des variables à partir des facteurs qu'on choisit de retenir.

Les deux autres matrices rapportées sont la « matrice de structure » et la « matrice de corrélation des facteurs ». L'interprétation de cette dernière va de soi. La matrice de structure, quant à elle, contient les corrélations entre les pleines variables (pas seulement leur partie reproduite) d'une part et les facteurs d'autre part dans l'espace des répondants. Pour une rotation orthogonale, ces corrélations sont identiques aux saturations. Il n'en est pas de même pour des facteurs corrélés : après une rotation oblique, les corrélations diffèrent alors des saturations. On explique ci-après ce que représentent ces corrélations. Dans une conception utilitaire de l'AFE, elles peuvent servir à interpréter les facteurs, même si une corrélation peut

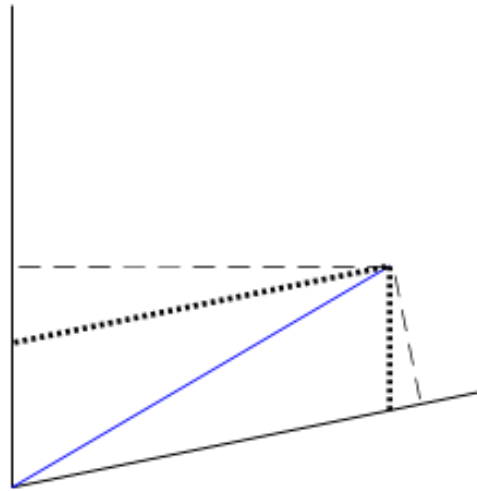
avoir une polarité opposée à celle de la saturation correspondante. Dans notre approche attribuant une réalité aux facteurs, ce sont les saturations, données dans la matrice de forme, qui devraient guider l'interprétation des facteurs.

En termes de représentation graphique, la Figure 4 montre une variable (en bleu) dans l'espace de deux facteurs (lignes pleines noires, de longueur unitaire) corrélés positivement ($r=0,2$). Les lignes *pointillées*, respectivement parallèles à chacun des facteurs, rejoignent l'autre facteur à une distance de l'origine correspondant à la saturation de la variable sur le facteur. Cela illustre la *recette de fabrication* (donnée dans la matrice de forme) au sens où, selon ce modèle, la nature aurait construit la variable en prenant ces portions des (scores des répondants sur chacun de ces) facteurs, les mettant bout à bout, pour produire la partie significative (hors unicité) de la variable.

Les lignes *hachurées*, quant à elles, sont perpendiculaires à chaque facteur, les touchant à une distance de l'origine correspondant à la corrélation du facteur avec la variable, telle que donnée par la matrice de structure. Les longueurs des facteurs de l'origine jusqu'à ces points d'intersection sont ainsi les projections de la variable sur les facteurs. Ces corrélations ne sont pas le cosinus de l'angle illustré dans la figure, parce que la ligne bleue de la figure ne représente que la partie expliquée de la variable. La variable elle-même, normalisée avec toute sa variance, a implicitement son unicité perpendiculaire au plan de la figure, laquelle ne participe pas à la longueur



FIGURE 4 ■ Saturations (en pointillé) de deux facteurs sur une variable et corrélations (en hachuré) des facteurs avec la variable. L'unicité de la variable est perpendiculaire à ce plan mais est prise en compte dans la corrélation.



de la projection sur les facteurs. L'angle entre la variable complète (variance partagée et unicité) est plus grand que celui représenté dans le seul plan de la figure.

Pour l'interprétation des facteurs, la matrice de structure risque d'induire en erreur, en particulier si deux facteurs sont négativement corrélés, tel qu'illustré dans la figure 5. Les lignes en pointillé montrent encore comment on peut rejoindre l'extrémité de la partie expliquée de la variable en se déplaçant uniquement parallèlement aux facteurs. On peut aller d'abord le long du facteur horizontal, en allant plus loin que le bout de la variable, parce que l'autre facteur nous fera régresser horizontalement quand on ira vers le haut. On va ensuite dans la direction de l'autre facteur pour une longueur indiquée par le point d'intersection de la ligne pointillée sur ce facteur. L'ordre dans lequel on effectue ces deux pas n'importe pas. On a ici utilisé les saturations, soit les valeurs données dans la matrice de forme. On voit que chacun des facteurs contribue positivement à la variable. Si la variable avait presque toute sa variance expliquée par les deux facteurs corrélés négativement, on pourrait facilement avoir une valeur plus grande que l'unité comme saturation sur le facteur qui contribue le plus à la variable. Cela ne serait pas une anomalie, l'autre facteur ramenant l'extrémité de la variable à une coordonnée moindre que l'unité sur ce premier facteur.

Les lignes hachurées de la Figure 4 sont, encore en Figure 5, la projection de la variable sur la direction de cha-

acun des facteurs. Elles indiquent une corrélation négative de la variable avec le facteur présenté obliquement. La corrélation négative de la variable avec un des facteurs est donc ici trompeuse. Pour l'interprétation des facteurs, c'est donc la recette de fabrication, soit la matrice de forme, qu'il importe de prendre en considération pour comprendre que le facteur oblique s'exprime dans la variable avec une saturation positive.

Une autre matrice est parfois rapportée (si requise), dont le nom inclut le mot *coefficient*. Ces coefficients peuvent être vus comme des poids à appliquer aux cotes z des variables pour approximer les scores sur les facteurs. Il est également risqué de se servir de ces coefficients pour interpréter les facteurs, étant donné que les variables observées sont corrélées. Le problème illustré à la Figure 5 pour approximer une variable à partir de facteurs corrélés se retrouve à l'identique quand on cherche à approximer un facteur à partir de variables corrélées pour produire les scores factoriels. Cela revient à combiner bout à bout des longueurs appropriées dans la direction de chaque variable pour atteindre chaque facteur. La combinaison des variables pour atteindre le facteur peut, elle aussi, être trompeuse quant aux polarités de la contribution du facteur à la variable. La Figure 6 illustre cela en caricature pour une solution orthogonale (lignes noires), dans laquelle deux variables (exprimées en bleu) se retrouvent dans le même quadrant. Pour se rendre à l'extrémité de la ligne représentant le facteur horizontal partant de l'ori-

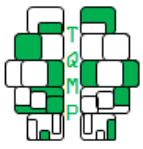
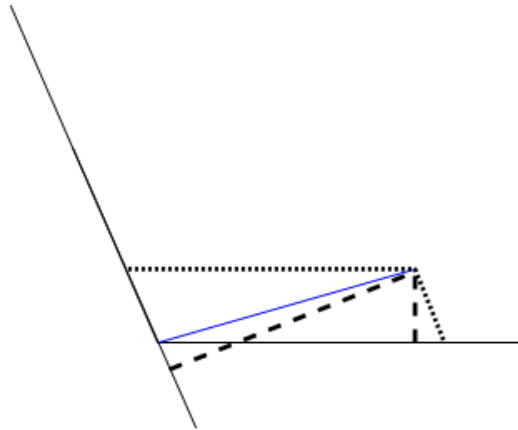


FIGURE 5 ■ Comme la Figure 4, mais pour des facteurs corrélés négativement pour montrer que le signe de la corrélation peut ne pas correspondre au signe de la contribution du facteur à la variable.



gine, le trajet de la ligne pointillée suit la variable proche de ce facteur mais doit la prolonger pour que la direction de l'autre variable permette d'atteindre le but. Le trajet pointillé revient, en quelque sorte, sur ses pas en donnant un poids négatif à l'autre variable, même si cette dernière est, elle aussi, influencée positivement par ce facteur horizontal.

En résumé, pour les rotations orthogonales, la matrice factorielle est à la fois la matrice de forme et celle de structure. Par contre, pour les rotations obliques, la matrice de forme donne les contributions des facteurs aux variables et sert à interpréter les facteurs sans les risques de fausses pistes que sont susceptibles de suggérer les corrélations facteurs-variables ou les coefficients des scores factoriels.

Interprétation des facteurs

Une fois fixés dans leur espace factoriel, les facteurs devraient être interprétés comme des sources d'information qui s'expriment dans les variables. L'expression *recette de fabrication* convient bien, particulièrement pour des variables qui reflètent simultanément plus d'un facteur. Le score observé pour un répondant sur une variable est alors vu comme la somme de son score sur chacun des facteurs, chacun pondéré par la saturation de la variable sur le facteur, laquelle est donnée dans la matrice de forme dans le cas de facteurs corrélés. On peut penser ici à l'exemple des questions d'examen où on ne s'attend pas particulièrement à une structure simple. Ce n'est qu'après avoir identifié en tant que facteurs les deux sources d'information (i. e., après avoir décidé de la solution à retenir) qu'on pourra

estimer pour chaque participant ses valeurs sur ces deux concepts théoriques.

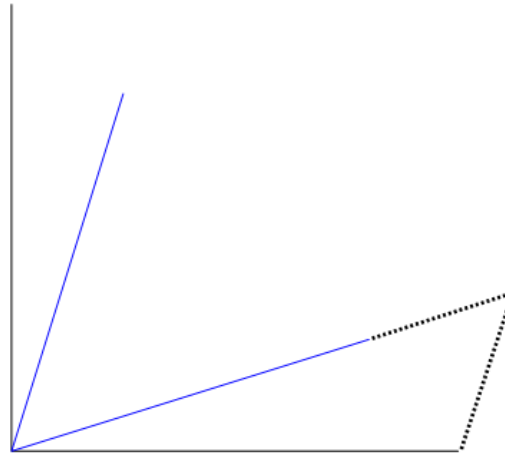
Comme déjà mentionné, si on crée des tâches ou des items de questionnaires dont chacun n'est destiné à ne refléter qu'un concept théorique, alors on s'attend à une structure simple et on peut espérer que l'AFE reflète assez facilement les relations des items aux facteurs. Toutefois, c'est l'analyse factorielle confirmatoire qui serait plus appropriée pour cela, vu que la structure des items est déjà l'objet d'une hypothèse précise qu'on souhaite valider. Une analyse factorielle confirmatoire bien menée pourra faire découvrir que, par exemple, pour rendre le modèle statistiquement acceptable, un item conceptuellement associé à un facteur est aussi influencé, dans une moindre mesure, par un autre facteur.

Souvent, toutefois, il n'est pas acquis d'avance que les données soient engendrées sous une relation de structure simple entre les variables et les facteurs sous-jacents, surtout si on n'a pas conçu les items avec en tête une structure factorielle claire. On devrait juger du mérite d'une solution factorielle par le degré auquel elle se prête à une interprétation qui a du sens. C'est pourquoi il est tout à fait légitime de comparer diverses solutions tant avec des rotations différentes qu'avec des nombres différents de facteurs.

Pour interpréter chaque facteur, la question qu'on se pose à partir des pondérations (obtenues de la matrice factorielle ou de celle de forme si la rotation était oblique) est la suivante : Quel est le concept qui s'exprimerait avec des pondérations non triviales dans tel sous-ensemble des



FIGURE 6 ■ Illustration de l'approximation d'un score factoriel à partir de deux variables dans le même quadrant.



variables? Se rappeler qu'une pondération à signe négatif indique qu'un score positif élevé sur le facteur contribue à un score négatif élevé⁹ sur la variable. Notez que dans le cas d'un facteur exprimé avec une prédominance de pondérations négatives, on peut inverser simultanément le signe de toutes ses pondérations, une opération qui revient simplement à inverser la direction du facteur dans l'espace factoriel global.

Si une rotation oblique produit des facteurs fortement corrélés, ces corrélations pourraient aussi entrer dans l'appréciation qu'on se fait de l'interprétation de la solution : d'où vient ce grand partage d'information? Les véritables facteurs qu'on voudrait identifier ne proviendraient-ils pas plutôt de deux sources d'information, l'une subsumant et générant ces facteurs corrélés, l'autre les distinguant? Par exemple, une rotation pourrait produire deux facteurs corrélés qu'on serait porté à identifier respectivement comme propension à fuir et propension à affronter les situations conflictuelles. Ces deux facteurs très corrélés pourraient avoir en commun de refléter la sensibilité aux situations conflictuelles vécues et se différencier par le style de réaction aux conflits, sensibilité et style étant deux dimensions passablement indépendantes. Ainsi, selon cette dernière interprétation orthogonale, plus un participant vit gravement les situations conflictuelles, plus son score sera élevé sur un des deux facteurs de l'interprétation oblique, lequel étant déterminé en fonction de son score de style.

Ainsi, selon que le participant est porté à fuir ou à affronter les situations conflictuelles, son score sera plus élevé en valeur absolue sur l'un ou sur l'autre des facteurs corrélés. On a ainsi deux interprétations possibles, l'une avec des facteurs obliques, l'autre avec des facteurs orthogonaux. Laquelle préférera-t-on? Celle de deux facteurs indépendants, de sensibilité aux conflits et de style de réponse, fait déjà l'économie d'avoir à expliquer la corrélation des facteurs en faisant de l'explication potentielle les facteurs mêmes. Si les données étaient compatibles avec l'hypothèse de seulement deux dimensions, il se pourrait même que la solution avant rotation soit la meilleure. Le premier facteur extrait exprimerait déjà ce qu'il y a de commun entre les deux facteurs suggérés par la rotation oblique, tandis que le second facteur extrait avant rotation expliquerait leur différence.

Le chercheur consciencieux s'efforce de comprendre et expliquer le réel tel qu'il se manifeste par les données observées, et non à simplement produire un compte rendu quelconque qui leur correspondrait. Pour cela, il est primordial de rester curieux et d'exercer son esprit critique, de réfléchir à la lumière des connaissances et concepts qu'on a déjà acquis tant sur le sujet traité que sur la mécanique de l'analyse factorielle. Il convient aussi de rester ouvert à plus d'une interprétation compatible avec les données.

On peut par ailleurs confronter une solution factorielle à d'autres données. La question de l'estimation de scores

9. Il est utile ici de reconnaître que cette idée de scores négatifs sur un facteur signifie seulement une valeur dans la partie inférieure de la distribution. Les scores sur un facteur (bien que pas ceux qu'on calcule) pourraient fort bien, de fait, être tous positifs. L'AFC opère en effet en dehors de l'hyper-diagonale, non seulement pour les variables mais aussi pour les facteurs.



des participants sur les facteurs d'une solution est discutée plus loin. Considérons ici que tels scores ont été obtenus. Considérons par ailleurs que les participants proviennent de divers groupes qui diffèrent sur une ou plusieurs des dimensions exprimées dans les données. En ce cas, l'analyse discriminante peut se montrer féconde pour confirmer ou infirmer le résultat d'une rotation. Cette analyse détermine s'il existe des différences entre groupes dans l'espace des variables qui lui sont soumises, lesquelles peuvent être des scores factoriels. Si les groupes diffèrent selon un test multivarié, une dimension est extraite des données pour rendre compte du maximum de variance qu'il y a entre les groupes, puis, pourvu qu'il y ait plus de deux groupes, un nouveau test est appliqué sur les résidus pour déterminer s'il reste des différences de groupes à expliquer.

Pour confronter une solution de rotation en AFE, on peut soumettre les scores factoriels à l'analyse discriminante. Pour peu que les groupes diffèrent, l'analyse discriminante identifie la direction dans l'espace factoriel sur laquelle les groupes diffèrent le plus. S'il devait y avoir deux dimensions sur lesquelles diffèrent les groupes, on se retrouverait avec une indétermination à une rotation près de ces deux facteurs, mais peut-être la disposition des moyennes de groupes (appelées barycentres ou *centroïdes*) dans le plan des deux dimensions identifiées par l'analyse discriminante indiquerait-elle les nouvelles directions (facteurs) qui décrivent ce plan avec le plus de sens. Si toutefois une seule dimension suffit à rendre compte des différences entre les groupes, alors, selon un principe de parcimonie, cette dimension a d'emblée une grande crédibilité et devrait correspondre à un des facteurs de la solution factorielle recherchée. Si on a utilisé les scores factoriels après rotation et que l'analyse discriminante associe une combinaison de deux facteurs à son unique fonction discriminante retenue, cela discrédite fortement la solution de rotation. Pour continuer à la défendre, il faudrait pouvoir soutenir que c'est par pure coïncidence que les différences unidimensionnelles entre groupes soient strictement proportionnelles sur deux facteurs. La Figure 7 illustre un tel cas.

Ici, trois moyennes de groupes (*) forment pratiquement une ligne dans l'espace de deux facteurs X et Y après rotation. Les variables elles-mêmes auxquelles contribuent X et Y ne sont pas représentées, mais les coordonnées des moyennes sur X et Y sont marquées le long des axes respectifs. Selon la solution de rotation des facteurs, il faudrait penser que les groupes diffèrent à la fois sur X et sur Y, mais avec cette étrange coïncidence que les rapports $x_1 : x_2 : x_3$ soient pratiquement les mêmes que les rapports $y_1 : y_2 : y_3$. Penser que la rotation n'était pas optimale et que les groupes ne diffèrent en fait que sur une seule dimension donne une explication plus parcimonieuse. Avec des

moyens plus sophistiqués, toutefois, on peut utiliser des différences de groupe (ou de temps) pour fixer les facteurs sans indétermination de rotation (voir la note 11).

Pondérations négligeables

L'interprétation des facteurs dans une solution proposée se fait à partir de ses pondérations non négligeables sur les diverses variables. Il convient pour cela de définir ce qu'est une pondération non négligeable. Pour une solution orthogonale, les pondérations sont les coefficients de corrélation entre la variable et chacun des facteurs. Pour une solution oblique, la matrice de structure donne elle aussi les corrélations entre les variables et les facteurs, mais celles-ci sont fortement influencées par les corrélations entre les facteurs eux-mêmes. Une variable v qui serait directement alignée sur le facteur F1 aurait une pondération nulle (et donc manifestement négligeable) sur tout autre facteur, bien qu'elle soit corrélée avec eux en vertu des corrélations de ces autres facteurs avec F1. Ce sont donc les pondérations données par la matrice de forme qui sont à juger si elles sont négligeables ou pas dans l'interprétation d'un facteur.

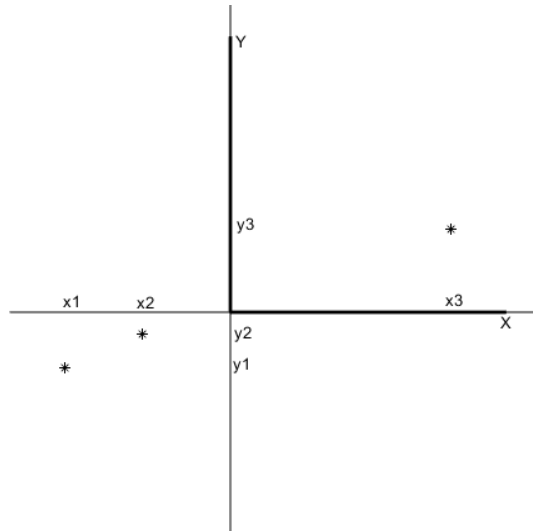
Ici, la question se divise en deux décisions. D'une part, devrait-on utiliser ou non telle variable dans l'interprétation d'un facteur donné? Après avoir jugé négligeable la saturation de la variable sur le facteur *pour ce qui est de lui attribuer une signification*, il reste à décider s'il est raisonnable de considérer comme nulle la contribution du facteur à la variable selon le modèle retenu. Un facteur peut en effet influencer une variable suffisamment pour que cela se détecte, mais trop peu pour que la variable contribue utilement à définir l'identité du facteur.

On peut donc commencer par se demander si une pondération donnée est suffisamment élevée pour qu'on s'en serve pour identifier ce que serait le concept correspondant à un facteur. Même si certains auteurs le préconisent, il n'est pas sage d'utiliser à cette fin une valeur limite prédéterminée : cela doit dépendre plutôt de l'ensemble des pondérations de la solution. Une pondération de 0,35 pourra être utile si presque toutes les pondérations sont sous 0,5, mais elle deviendra négligeable si, par exemple, plusieurs pondérations excèdent 0,6. Ce sont les variables avec des saturations dominantes sur le facteur qui vont surtout servir à son interprétation.

La question associée est de décider quelles pondérations plutôt faibles pourraient bien être nulles dans la population, en supposant que la solution trouvée soit essentiellement correcte. En effet, une corrélation nulle dans la population entre une variable et un facteur ne le sera pas exactement dans un échantillon. Ici aussi, il n'est pas très avisé de se fier à une valeur de coupure fixe, même si l'on suggère parfois que, par exemple, une



FIGURE 7 ■ Illustration d'une fonction discriminante sur les scores de deux facteurs.



pondération inférieure à 0,3 en valeur absolue peut être ignorée.

Des simulations exploratoires inédites ont montré, autant pour des facteurs orthogonaux que corrélés, que les pondérations nulles dans la population se comportent sensiblement comme des coefficients de corrélation dans des échantillons, avec des valeurs critiques qui dépendent de la taille de l'échantillon. Ainsi, avec une centaine de répondants pour lesquels le coefficient de corrélation critique au seuil de probabilité de 0,05 est d'environ 0,20 ($\approx 1,96/\sqrt{100}$), il est raisonnable d'accepter qu'une pondération de 0,15 dans une solution de FAP puisse être nulle dans la population. Mais pour le triple de répondants, la même pondération de 0,15 aura une probabilité estimée d'environ 0,009 selon un test bilatéral. Il devient alors plutôt gênant de prétendre qu'elle soit nulle dans la population. L'interprétation du facteur, si c'est la solution qu'on accepte, devrait alors considérer qu'il puisse conceptuellement contribuer au moins minimalement à la variable, pas assez pour que la variable serve à identifier le construit correspondant au facteur, mais trop pour qu'on considère qu'il soit sans effet sur cette variable.

Indétermination des scores factoriels

Une fois les relations entre les facteurs et les variables fixées par une rotation appropriée résultant en facteurs interprétables, on veut souvent estimer les scores des répondants sur ces facteurs. Même avant de considérer qu'on pouvait avoir de tels scores pour valider une rotation, par exemple par une analyse discriminante, on avait

déjà fait allusion aux scores factoriels en mentionnant l'existence d'une matrice de coefficients liant les variables aux facteurs. On présentait alors ces coefficients comme les poids à donner aux variables normalisées (cotes z) pour approximer chacun des facteurs. Produire ces coefficients revient donc à identifier des combinaisons linéaires des variables observées pour retrouver au mieux les scores des participants sur les facteurs identifiés.

Après avoir identifié les facteurs comme des directions dans un sous-espace de l'espace des variables, lui-même un sous-espace de celui des répondants, on souhaite pouvoir lire directement la coordonnée d'un facteur donné sur l'axe de chaque répondant pour savoir quel score chacun obtient sur le facteur. Il y a toutefois une complication liée à ce que la réduction de la matrice de corrélations pour en soustraire l'unicité des variables ne nous dit pas quelle est, pour chaque répondant, la valeur de cette unicité (souvent réductible à une erreur de mesure). La situation reste toutefois conceptuellement simple pour peu qu'on accepte qu'on ne peut pas retrouver exactement, sans contamination ou part d'erreur, ces scores. Cela vient de ce qu'on ne pourra les estimer que par des combinaisons linéaires des variables, lesquelles contiennent chacune une part de variance d'erreur impossible à enlever, même par corrélation multiple, vu que les prédicteurs contiennent aussi de l'erreur.

Considérons un exemple simple illustrant ce qui peut être fait. Un facteur s'est exprimé dans seulement deux variables, et la solution a retrouvé les deux saturations correspondantes. En principe, telle solution n'est pas



déterminée (voir la discussion plus loin sur l'indice KMO et le besoin d'au moins trois variables par facteur), mais acceptons que les saturations sont connues, ce qui nous importe ici étant d'illustrer la problématique de l'approximation des scores factoriels.

Supposons donc la recette de fabrication donnée au Tableau 10, où les colonnes sont des axes et les rangées des variables, parmi lesquelles on a inclus le facteur lui-même (qu'on voudra approximer par les variables A et B) :

Les entrées de ce tableau sont les coordonnées de F , A et B dans un système d'axes orthogonaux désignés F , err_A et err_B . Leurs sommes de carrés sont toutes 1 et leurs sommes des produits par paire sur les trois axes sont les corrélations. La corrélation AB est ainsi $0,8 \times 0,75 = 0,6$; AF est $1 \times 0,8$ et BF $1 \times 0,75$. Selon notre convention, les saturations auraient été identifiées correctement comme les lignes 2 et 3 de la première colonne du Tableau 10. On peut compléter ces saturations sur le seul facteur affectant A et B en ajoutant les unicités. Souvenons-nous que A et B sont connus pour chaque répondant et qu'on cherche des coefficients à leur appliquer pour calculer F .

Ce qu'on va faire maintenant s'appelle une orthogonalisation de Gramm-Schmidt pour les trois « variables » A , B et F dans cet ordre. Cette procédure consiste à fixer la première variable comme premier axe, puis à représenter chaque variable suivante avec une partie sur chacun des axes antérieurs, de façon à conserver, sur ces nouveaux axes, les mêmes sommes de produits que sur les axes d'origine, puis de compléter la somme des carrés de chaque variable par la valeur appropriée sur un nouvel axe. Autrement dit, chaque nouvel axe est associé à la part de variance de chaque variable pas déjà prise en compte par les axes précédents.

On trouve les coordonnées ci-dessus, illustrées en Figure 8 pour le plan des deux premiers axes définis par les variables A et B :

La valeur de $0,3375$ fut obtenue comme solution en X de $0,6 \times 0,8 + 0,8 \times X = 0,75$ (qui est la corrélation entre B et F). Le $0,4961$ est la racine carrée de $0,2461$ ($= 1 - 0,8^2 - 0,3375^2$). Dans ce nouveau système d'axes, chaque variable, incluant F , a la somme des carrés de ses coordonnées à 1. De plus, les trois corrélations, AB , AF et BF , sont les mêmes qu'à l'origine.

On peut se rendre à n'importe quel point du plan de A et B par une combinaison additive de ces deux variables, mais on ne peut atteindre aucun point en dehors de leur plan. Les intersections des lignes pointillées avec les deux variables indiquent la longueur qu'on doit donner à chacune pour se rendre à l'extrémité de la partie du facteur dans ce plan. Cela correspond tout à fait à la régression multiple du facteur sur les variables. Les pondérations des variables donnent ici une variance de $0,8^2 + 0,3375^2 =$

$0,7539$, soit le R^2 de la régression multiple. On ne peut donc reproduire de F , stipulé de variance unitaire et sans erreur de mesure, que sa partie en bleu dans l'illustration de la Figure 8, qui correspond ici à environ 75% de sa variance. Cela veut dire que 24,6% de la variance du facteur ne peut être reproduite à partir des variables. Le 75,4% représente même un gain par rapport à A ou B en solo comme estimatrices de F , sur lesquelles la source d'information que représente le facteur aurait respectivement $0,8^2 = 64\%$ ou $0,75^2 = 56,25\%$ de sa variance reproduite.

La troisième coordonnée de F , $0,4961$, pourrait être peinte, dans une troisième dimension, autant derrière que devant le plan de A et B . Le facteur F entier ne se situe donc pas entièrement dans le plan de A et B , mais situer F dans ce plan est la meilleure (en fait la seule) estimation qu'on peut en faire à partir des variables. La part de F qui manque n'est pas accessible. La limite conceptuelle de seulement deux positions possibles pour reconstituer l'entièreté de F , devant ou derrière le plan de A et B , est ici un artifice lié à ne figurer la situation qu'avec tout au plus trois dimensions. Dans l'espace de N répondants où les p variables occupent p dimensions, après exclusion aussi de l'hyper-diagonale, on aurait $N - p - 1$ dimensions restantes où peut se trouver la vraie position de F tout entier, sans moyen de connaître cette vraie position. La partie en bleu dans l'illustration est donc la partie de F qu'on peut approximer au mieux à partir des valeurs observées sur les variables. L'échelle choisie pour les scores factoriels étant arbitraire, ce qu'il importe d'approximer est la direction de cette partie reproductible du facteur.

Alors que, dans de nombreuses techniques, l'erreur d'estimation peut être arbitrairement réduite en augmentant la taille de l'échantillon, pour les scores factoriels c'est d'abord le nombre de variables dans lesquelles s'exprime un facteur qu'il faudrait augmenter pour mieux approximer les scores sur ce facteur. Par exemple, 8 variables influencées par un facteur définissent un espace à 8 dimensions dont on peut atteindre n'importe quel point et ces 8 dimensions épuiseront une grande part de la variance exacte du facteur, mais pas toute la variance : il restera un résidu de sa variance dans une 9e dimension. De seulement augmenter la taille de l'échantillon permet de rendre les estimations de corrélations plus proches de ce qu'elles sont dans la population; cela en retour réduit l'erreur du modèle (erreur d'orientation du facteur estimé dans l'espace des répondants par rapport à celle dans la population) mais, pour un modèle donné, augmenter seulement le nombre de répondants ne réduit pas la part d'erreur de mesure associée à chaque variable et contribue donc très peu à augmenter la précision des scores associés aux facteurs du modèle retenu.



Tableau 10 ■ Modèle simple à un facteur exprimé dans deux variables pour discuter de l'indétermination des scores factoriels.

	<i>F</i>	<i>err_A</i>	<i>err_B</i>
<i>F</i>	1	0	0
<i>A</i>	0,8	0,6	0
<i>B</i>	0,75	0	0,6144

Tableau 11 ■ Résultat de la rotation de Gramm-Schmidt pour exprimer le facteur *F* dans le plan des variables *A* et *B*.

	axe		
	a1	a2	a3
<i>A</i>	1	0	0
<i>B</i>	0,6	0,8	0
<i>F</i>	0,8	0,3375	0,4961

Estimation des scores factoriels

La procédure appliquée pour confectionner les scores factoriels est partiellement arbitraire en raison de la redondance entre les variables. Les poids relatifs appliqués aux cotes *z* des différentes variables définissent la direction de la partie reproduite du facteur dans l'espace des variables. Les poids absolus définissent la variance de ces scores, leur échelle. La moyenne des scores sera toutefois nécessairement nulle sur l'ensemble des répondants puisque la moyenne de toute somme pondérée de cotes *z* est nécessairement nulle.

Ayant vu que la partie des facteurs à reproduire est bien définie à partir des relations entre les facteurs et les variables, données par la matrice de forme, on peut se demander : Pourquoi alors SPSS, par exemple, propose-t-il trois méthodes (Régression, Bartlett et Anderson-Rubin) pour estimer les scores ? En quoi ces méthodes diffèrent-elles ? Quels sont les enjeux du choix d'une méthode ? Laquelle est préférable ?

Toutes ces méthodes peuvent être qualifiées de solutions de moindres carrés généralisés. Cela veut dire que, explicitement ou implicitement dans leurs formules de calcul, les variables sont chacune associée à une unicité (ici considérée comme erreur d'estimation) de même grandeur. Cela fait que, si une variable coïncidait parfaitement avec un facteur de telle sorte qu'elle ne soit affectée d'aucune erreur de mesure, alors le score sur le facteur en question serait parfaitement connu à partir de cette seule variable. En effet, pour donner à l'unicité (nulle) de cette variable la même longueur qu'aux unicités des autres variables, il faudrait rendre infinie la taille de la variable, ce qui rendrait dérisoires (nulles) les contributions des autres variables à reproduire le facteur. Dans le cas plus général, les variables qui ont le moins d'unicité sont celles dont les longueurs relatives en moindres carrés généralisés sont les

plus longues pour reproduire la direction des scores factoriels. Pour une variable de grande communauté (faible unicité) dépendant de plus d'un facteur, d'autres variables sont mises à contribution, par des coefficients appropriés, pour annuler la contribution des facteurs autres que celui qui est la cible d'une estimation.

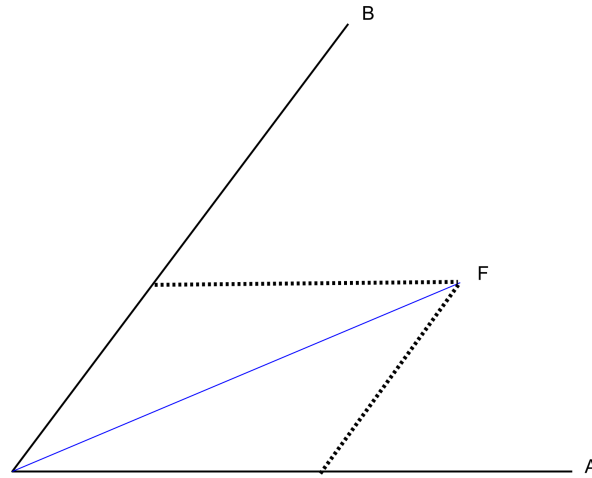
Pour les rotations orthogonales, la seule différence à attendre des trois méthodes proposées est la variance des scores, car les scores obtenus par chacune de ces méthodes ont une corrélation parfaite de 1 entre eux. Pour des rotations obliques, les directions des scores diffèrent un peu, quoique les corrélations entre les scores obtenus par des méthodes différentes demeurent très élevées.

Seule la méthode d'Anderson-Rubin donne une variance de 1 aux scores de chaque facteur, quelle que soit la rotation. Elle impose aussi des corrélations nulles entre scores des divers facteurs. Ceci pose question, puisque la notion de facteurs corrélés se comprend généralement comme ceux-ci étant corrélés dans l'espace initial des participants, c'est-à-dire reflétant la corrélation des scores des participants sur ces facteurs. À moins de concevoir une autre interprétation de la corrélation entre facteurs, la méthode d'Anderson-Rubin n'est intéressante que si on a une raison impérative et justifiée de souhaiter des scores sans corrélation même pour des facteurs que le modèle accepté dit corrélés. Cette méthode n'estime donc pas véritablement les scores pour les facteurs de la solution retenue, mais des scores modifiés afin que leurs corrélations deviennent nulles : l'interprétation de tests appliqués à partir de ces scores devient alors ... problématique. Bien entendu, il serait absurde et futile d'utiliser les scores obtenus par la méthode d'Anderson-Rubin pour rechercher des méta-facteurs susceptibles d'expliquer les corrélations entre facteurs.

Les deux autres méthodes donnent des scores qui reproduisent de près les corrélations admises entre les fac-



FIGURE 8 ■ Approximation (des scores) d'un facteur par les deux variables dans lesquelles il s'exprime.



teurs. La méthode dite de Régression donne des scores d'une variance égale au R^2 de la régression. C'est là une caractéristique importante qui mérite d'être rapportée dans toute publication faisant usage des scores factoriels. Cela donne une idée de la fiabilité de la mesure et donc de sa marge d'erreur. La méthode de Bartlett donne des variances plus grandes que l'unité et qui sont très proches de $1/R^2$. Il est donc plus simple et tout aussi satisfaisant d'utiliser la méthode de Régression.

Tests de sphéricité et indice KMO

Il est d'usage, en début d'AFE dans SPSS, de commander le test de sphéricité de Bartlett et l'indice KMO d'adéquation de l'échantillonnage. Tant qu'on y est, autant demander aussi le déterminant de la matrice de corrélations et demander à voir la matrice anti-image. On doit généralement aussi choisir l'analyse en axes principaux plutôt que l'analyse en composantes principales et indiquer si on veut analyser la matrice de corrélation ou celle de covariances. Toutes ces options méritent explicitation et discussion.

Le déterminant de la matrice de corrélations, s'il est très petit (disons $< 0,001$), indique un problème de multicolinéarité dans les données, à savoir que les p variables ne forment pas véritablement un espace avec autant que p dimensions. Ce pourrait être dû à une variable qui est la somme d'autres variables aussi incluses dans l'analyse ou aux mêmes données incluses sous deux formes (par ex., poids en livres et en kilogrammes). Pour éviter des problèmes s'apparentant à une division par zéro, on voudra identifier l'anomalie et la résoudre avant de continuer

l'analyse.

Le test de sphéricité de Bartlett est en fait un test que les données ne sont pas que des variables toutes indépendantes les unes des autres. Pour toute analyse sérieuse, basée sur des variables dont on s'attend qu'elles nous informent sur leurs facteurs communs, l'hypothèse nulle que toutes les corrélations ne sont qu'accidentelles sera rejetée par le test de sphéricité. Mais la procédure NEST, en déterminant le nombre de dimensions nécessaires et suffisantes pour rendre compte des corrélations, fournit déjà la réponse à cette question. Le test de Bartlett est alors simplement redondant avec le premier test effectué par NEST, même si, dans SPSS, il est automatiquement commandé en même temps que le KMO.

L'indice de qualité d'échantillonnage des variables KMO (Kaiser-Meyer-Olkin) est toutefois utile. Il estime dans quelle mesure les dimensions présentes dans les données sont appuyées sur plus de deux variables. L'indice minimal recommandé est 0,6, mais on peut tout de même procéder s'il est inférieur.

Pour comprendre l'utilité relative du KMO, il faut d'abord savoir qu'un facteur doit s'exprimer dans au moins trois variables pour qu'on puisse fixer ses pondérations. Quand un facteur ne détermine de corrélation que sur une paire de variables, on le dit « facteur doublet ». L'AFE doit donner aux variables de ce facteur un poids sur chacune d'elle alors qu'il n'y a qu'une corrélation à représenter : une corrélation de 0,64, par exemple, est obtenue tout aussi bien par une contribution du facteur doublet de 0,8 à chacune des deux variables que par une contribution de 1 à l'une et de 0,64 à l'autre,

10. Numériquement, des poids de 1,02 et de 0,6275 feraient également l'affaire, hormis qu'en un tel cas la variable avec un poids supérieur à l'unité



ou vice versa, ainsi que par toute solution intermédiaire ou même au-delà.¹⁰ Avec trois variables affectées par un facteur, on a trois saturations à établir et trois corrélations pour contraindre le choix de ces saturations, pour quatre variables, on a six contraintes pour quatre saturations, etc. Il se peut aussi que les variables affectées par un facteur doublet le soient aussi par un autre facteur, de sorte que le facteur doublet n'explique qu'une partie de leur corrélation.

Le KMO tente de signaler les situations où les paramètres à trouver pour les facteurs ne sont pas suffisamment contraints (moins de trois variables par facteurs, ou au moins trois mais avec de grandes unicités). Cela est fait en divisant la somme des carrés des corrélations par la même somme plus celle des carrés des corrélations partielles. Pour les facteurs exprimés dans seulement deux variables, il n'y a pas d'autres variables portant l'information, de sorte que les corrélations partielles sont presque aussi grandes que les corrélations entières. Pour la paire de variables, cela double presque sa contribution au dénominateur, tirant le KMO vers la valeur 0,5. Comme toutes les variables contiennent de l'unicité, les corrélations partielles tombent rarement à 0 même quand d'autres variables expriment l'information provenant des facteurs qui rendent deux variables corrélées. À cause de cela, il n'y a pas de valeur stricte que le KMO doit dépasser pour assurer que les paramètres trouvés seront fiables. On suggère de s'interroger sérieusement si le KMO tombe sous 0,6 alors qu'on sera très confiant s'il déborde même 0,7.

Manifestement, un facteur s'exprimant dans une seule paire de variables aura d'autant moins d'effet sur le KMO général qu'il y a plus de variables dépendant d'autres facteurs dans l'analyse. La diagonale de la matrice anti-image des corrélations donne, en négatif, l'indice KMO des variables individuelles (la somme de leurs corrélations au carré sur celle de leurs corrélations au carré plus leurs corrélations partielles au carré) : cela peut servir à identifier la paire de variables problématiques, ce qui reste difficile si les variables affectées pas un facteur doublet sont aussi affectées par un autre facteur.

L'AFE mène parfois au signalement d'un cas de Heywood. Cela est fait par une note, au bas d'un tableau vide intitulé *Matrice des facteurs*, qui dit « La qualité de représentation d'une variable a dépassé 1. L'extraction a été interrompue ». Rappelons que la qualité de représentation est la communauté. Il y a alors lieu de suspecter la présence d'un facteur doublet, lequel informerait deux variables dont le KMO individuel est proche de 0,5, pour peu que toute la corrélation de ces deux variables soit due à ce seul facteur doublet. On devrait alors

considérer de retirer ces variables de l'analyse. Toutefois, si on tient à obtenir une solution factorielle qui sauvegarde les variables concernées, on peut plafonner le nombre d'itérations permises pour que les itérations s'arrêtent avant que la communauté d'une variable n'atteigne 1. Il faut toutefois se rappeler que, dans une telle solution, les pondérations du facteur doublet sur ses variables n'ont de contrainte que de reproduire la corrélation entre les deux variables. Qu'une des variables ait sa communauté à 1 (i.e., toute sa variance expliquée par le facteur) ne signifie alors pas que cette variable coïncide parfaitement avec le facteur. De plus, les scores factoriels sur un facteur doublet sont très peu fiables. Pour représenter le facteur, mieux vaut alors utiliser la moyenne des cotes z des deux variables, comme si la solution leur avait donné des saturations égales à la racine carrée de leur corrélation.

Quand préférer analyser les covariances

La principale raison pour analyser par défaut la matrice de corrélations plutôt que celle de covariances est que les unités de mesure des données brutes sont typiquement hétérogènes. Que l'âge des participants soit donné en jours ou en années, leurs poids en grammes ou en kilogrammes et leur durée de sommeil en minutes ou en heures ne change pas la nature de l'information portée mais affecte considérablement la variance des variables qui portent l'information. Pour leur donner des échelles commensurables, il convient de les analyser en uniformisant leurs échelles, notamment par leurs cotes z équivalentes. C'est ce que fait implicitement la matrice des corrélations.

Mais il arrive que toutes les variables soient dans la même métrique. Pour bien illustrer la situation, prenons l'exemple de variables mesurées sur une échelle de Likert, nonobstant les réserves qu'on devrait avoir touchant l'AFE de telles variables ordinales. Supposons qu'un item laisse les répondants plutôt indifférents de sorte que toutes les réponses sont limitées à deux niveaux adjacents au milieu de l'échelle de 1 à 7 (p. ex. les scores 3 et 4) indiquant de faibles tendances, et supposons qu'un autre item polarise les participants de sorte que les réponses ne soient qu'à une ou l'autre des extrémités de l'échelle (p. ex. seulement des 1 et des 7). Pour une ACP, il apparaît évident que les transformations en cotes z sous-tendant les corrélations donneraient autant d'importance (autant de variance) à chacun de ces deux items. Pour l'AFE, qui ne cherche à rendre compte que de l'information partagée par les variables, excluant leurs unicités, on peut s'attendre à des partages plus nets d'information impliquant la variable avec des scores extrêmes que pour celle avec des scores regroupés. L'analyse des corrélations entraîne une inutile perte d'information lorsque toutes les variables

aurait plus que 100% de sa variance expliquée par le facteur : la détermination itérative s'arrête lorsque 1 est atteint et un cas de Heywood est signalé.



sont d'emblée sur une même échelle. Il pourrait ainsi être avantageux du point de vue de la précision des estimations, même pour l'AFE, qu'on analyse la matrice des covariances plutôt que celle des corrélations. Ainsi, l'analyse des covariances devrait mieux identifier que celle des corrélations les facteurs qui affectent fortement les réponses, accordant plus de poids aux items qui ont de plus grandes covariances (issues de leurs plus grandes variances).

Le problème avec l'AFE d'items sur des échelles de Likert, c'est que le modèle de contributions additives des facteurs a peu de chance de bien s'appliquer. Il semblerait que pour ces items l'utilisation des corrélations polychoriques rende les AFE plus fiables. C'est là un domaine que nous n'avons pas exploré. Notre *attitude* face à ces échelles a jusqu'ici été de transformer les valeurs pour rendre si possible leur distribution symétrique, sous le pari que cela rende le modèle additif plus vraisemblable. Avec certaines variables transformées et d'autres pas ou toutes transformées mais pas toutes de la même manière, c'est la matrice de corrélations qu'on analyse en vertu des variances qui ne reflètent plus l'importance discriminative des variables.

AFE de mesures répétées

Un problème rarement discuté est comment il convient d'effectuer une AFE sur des mesures répétées. Par exemple, des chercheurs peuvent appliquer une batterie de mesures et obtenir des protocoles de données avant, durant ou après un traitement. Veut-on effectuer une AFE séparée à chaque temps de mesure, puis chercher des rotations qui alignent les facteurs des différents temps de mesure avant d'obtenir leurs scores respectifs? La plupart des modèles d'AFE qui ont été proposés pour des mesures répétées admettent des modifications des saturations des facteurs d'un temps de mesure à l'autre. CORBALLIS et TRAUB (1970), par exemple, ont proposé une méthode s'appliquant à seulement deux temps de mesure et basée sur des contraintes d'orthogonalité de chaque facteur avec tous les autres du même temps et des autres temps, où tant les saturations d'un facteur que les scores factoriels peuvent changer d'un temps de mesure à un autre, même si on espère que ce ne soit pas le cas pour les premières. DASSA (1982) propose d'intégrer des fonctions temporelles à appliquer aux scores factoriels des répondants pour guider l'analyse mais sa méthode demande de fixer d'avance la structure simple des facteurs. Plus récemment, c'est l'analyse factorielle confirmatoire qui a été mise de l'avant

pour modéliser l'évolution temporelle de facteurs.

Se concentrant sur la notion de *recettes de fabrication* à identifier, on peut admettre que de très nombreuses situations devraient donner lieu à des changements sur certains facteurs sans que leur relation avec les variables dans lesquelles ils s'expriment ne change. Admettons toutefois qu'il n'est pas universellement assuré que les facteurs s'expriment avec des saturations constantes dans les variables, surtout si celles-ci doivent être mesurées avec des instruments différents aux différents temps de mesure, comme pour des mesures de capacités cognitives prises de dix ans en dix ans à partir de cinq ans.

C'est toutefois souvent le cas qu'on puisse considérer que les facteurs conservent dans le temps leur intensité d'expression dans les variables, les changements se concrétisant dans les scores des participants sur certains facteurs. Face à un simple changement de variance des scores sur un facteur d'un temps de mesure à un autre, des AFE séparées pour les différents temps devraient donner des saturations distinctes pour les mêmes facteurs. Cela complique inutilement la compréhension du changement temporel, alors que des recettes de fabrication communes peuvent être obtenues. Avec des variances factorielles variant dans le temps, on n'a ici rien de fondamentalement différent de l'AFE de participants appartenant à des groupes ayant des moyennes différentes sur certains facteurs, bien que l'AFE ne prenne pas en compte l'appartenance au groupe. C'est alors une analyse ultérieure utilisant les scores factoriels qui révèle les différences de groupes.

La compréhension de l'AFE développée ci-dessus dicte une première solution plausible et simple¹¹ pour l'AFE de données longitudinales qu'on peut appliquer soi-même dans SPSS. Les saturations des facteurs seront en quelque sorte moyennées sur les temps de mesure, de façon à leur appliquer la même recette de fabrication¹²; des analyses ultérieures pourront vérifier tant des changements de moyenne que de variance sur les scores factoriels. Cette solution a vraisemblablement déjà été décrite, tellement elle découle naturellement du modèle factoriel, mais nous n'en connaissons pas de référence.

Voici comment se présente la situation : on s'attend à des facteurs dotés de recettes fixes pour leur expression dans les variables : ce qui changerait d'un temps de mesure à l'autre, ce sont les valeurs des répondants sur certains facteurs. Étant donné cette perspective, on considérera d'abord les différents temps de mesure comme s'ils

11. Un traitement plus avancé, impliquant de la décomposition trimodale, est décrit dans le manuscrit : ACHIM, A. (en préparation) *Exploiting the hyper-diagonal in factor analysis using PARAFAC models*. On y décrit l'exploitation de différences de moyennes entre groupes ou temps de mesure pour identifier certains facteurs sans incertitude de rotation.

12. La variance de 1 donnée implicitement aux facteurs dans le modèle de l'AFE ne vise qu'à fixer l'échelle des saturations. Une fixation commune pour tous les temps de mesure permet d'avoir la même recette de fabrication et la même règle pour engendrer les scores factoriels. La procédure proposée ici est ainsi très analogue à la fixation d'une moyenne commune de tous les participants pour ensuite comparer leurs moyennes de groupes.



émanaient de nouvelles vagues de répondants. Pour t temps de mesure de v variables auprès de N répondants, on effectuera l'AFE sur les données mises sous la forme de $t \times N$ rangées pour les v variables. Toutefois, comme il ne s'agit plus de répondants indépendants, si on utilise NEST pour déterminer le nombre de facteurs à retenir, ce sont plutôt les moyennes des t répétitions qu'il faudra lui soumettre, vu que le nombre correct de répondants indépendants est crucial pour NEST. On effectue donc l'AFE sur les $t \times N$ rangées de données en demandant d'extraire par FAP le nombre k de facteurs suggéré par NEST. Une fois satisfait du résultat d'une rotation appropriée, on fait produire les k scores factoriels pour chacune des $t \times N$ rangées, dont on sait que les t ensembles de N rangées successives correspondent à nos différents temps de mesure. On reporte ces sous-ensembles comme autant de groupes de k colonnes correspondant aux différents temps de mesure. On pourrait alors soumettre ces scores factoriels à une ANOVA à mesures répétées ou à modèle mixte s'il y a des groupes. Si on préfère une MANOVA, l'étape de rotation n'importe que pour une éventuelle interprétation des résultats; des différences de temps ou des interactions groupe par temps donneront les mêmes effets multivariés, qu'on ait fait une rotation ou pas.

La procédure de passer par $t \times N$ rangées fait que les facteurs dont les valeurs sont modifiées au cours du traitement sont des sources de variance plus importantes que les facteurs non affectés, ce qui devrait faciliter l'identification de ces facteurs. Les valeurs des répondants sur les variables reflétant des facteurs dont la moyenne change auront une variance accrue sur les $t \times N$ rangées par rapport aux variables reflétant des facteurs non affectés au cours du temps. Ainsi, les facteurs affectés par un traitement, par exemple, ont de bonnes chances de se retrouver parmi ceux expliquant le plus de variance.

Conclusion

Le but du présent texte était de faciliter une compréhension interne de l'AFE et de sa dynamique. Cela aura permis de comprendre pourquoi la FAP est préférable à l'ACP et d'être conscient que ces techniques sont basées sur des modèles additifs de l'effet des facteurs sur les variables, modèles que compromettent notamment de trop fortes asymétries des données. On devrait retenir qu'il n'est plus nécessaire de dépendre de simples heuristiques pour estimer le nombre de dimensions requises, et on devrait pouvoir juger de l'à-propos d'une rotation orthogonale versus une rotation oblique (tout en gardant alors à l'esprit la question importante

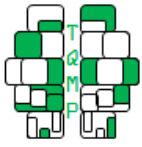
du partage d'information entre les facteurs). On devrait, de plus, retenir des éléments de réflexion pour gérer les éventuels cas de Heywood signalés, comprendre qu'on doit se méfier de critères fixes sur les pondérations (ou saturations) à prendre en compte pour l'interprétation des facteurs ou qu'on peut considérer comme nulles dans la population, préférer des scores factoriels qui respectent les corrélations entre les facteurs, plutôt que ceux produits par la méthode Anderson-Rubin, considérer la variance de ces scores factoriels comme un indice de la part d'incertitude qui les affecte encore, et pouvoir aborder intelligemment des mesures prises à deux ou plusieurs moments dans le temps.

Note de l'auteur

Remerciements à Louis Laurencelle pour ses encouragements et ses nombreux commentaires constructifs sur diverses versions de ce texte.

Références

- ACHIM, A. (2017). Testing the number of required dimensions in exploratory factor analysis. *The Quantitative Methods for Psychology*, 13(1), 64-74. doi :10.20982/tqmp.13.1.p064
- ALLAIRE, D. & LAURENCELLE, L. (2014). The sampling variance of $r(X,Y)$ for unrelated X and Y : a simple algebraic demonstration. *Communications in Statistics*, 33, 2389-2392.
- CORBALLIS, M. C. & TRAUB, R. E. (1970). Longitudinal factor analysis. *Psychometrika*, 35, 79-98.
- DASSA, C. (1982). L'analyse des données axée sur les développements éducatifs. *Revue des sciences de l'éducation*, 8, 415-429. Récupérée à partir de <https://doi.org/10.7202/900382ar>
- GREEN, S. B., LEVY, R., THOMPSON, M. S., LU, M. & LO, W.-J. (2012). A proposed solution to the problem with using completely random data to assess the number of factors with parallel analysis. *Educational and Psychological Measurement*, 72, 357-374. doi :10.1177/0013164411422252
- HORN, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30, 179-185. doi :10.1007/BF02289447
- RUSCIO, J. & ROCHE, B. (2012). Determining the number of factors to retain in an exploratory factor analysis using comparison data of a known factorial structure. *Psychological Assessment*, 24, 282-292. doi :10.1037/a0025697



Appendix A : Extraction des composantes principales par la méthode de puissance (ou « power method »).

La méthode de puissance est une méthode itérative qui s'amorce en posant une solution approximative, même lointaine, de l'éventuelle bonne solution, et qu'on améliore par des itérations successives conditionnées sur les données. L'algorithme décrit ci-dessous passe par une alternance entre, d'une part, estimer les scores factoriels des répondants selon l'estimation courante des saturations sur la composante recherchée et, d'autre part, ré-estimer les saturations (i.e. la composante) à partir des données et de l'estimation courante des scores sur cette composante. On finit par regrouper de tels cycles en une seule opération où on passe d'une estimation des saturations à la suivante sans estimer les scores. Le rôle intermédiaire des scores factoriels est implicite dans la procédure mais facilite sa compréhension. À chaque étape, même avant l'unification des deux phases en une seule opération par cycle, on donne à l'estimation obtenue (des saturations ou des scores) une somme de carrés de 1 et on utilise un paramètre scalaire (a ci-dessous) pour conserver l'information d'échelle. L'approximation initiale de la composante importe peu, pourvu qu'elle ne définisse pas un vecteur orthogonal à l'espace des données. S'il arrivait que l'approximation initiale corresponde parfaitement à une composante principale de plus petite valeur propre, la convergence des itérations serait immédiate et on aurait juste extrait cette composante avant son tour.

Le modèle des données est le suivant. Pour n répondants, p variables observées et k facteurs sous-jacents, on a les valeurs $S(n, k)$ des répondants sur les facteurs, les recettes $F(p, k)$ de fabrication des p variables à partir des k facteurs, et les données $X(n, p)$, telles que $X(n, p) = S(n, k)F'(k, p) + aleat(n, p)$, où $aleat$ est une source d'unicité non corrélée entre les répondants et entre les variables, typiquement de l'erreur de mesure (quoique cela pourrait aussi exprimer de l'information fiable non partagée). $F'(k, p)$ est la transposée de $F(p, k)$.

Supposons $k = 1$ et posons $F(p, 1) = a\underline{F}(p, 1)$ où le souligné indique un vecteur standardisé (somme de carrés = 1). La constante a s'obtient par la racine carrée de $F'F$. Si on connaît \underline{F} , comment peut-on évaluer S et a ? En utilisant \approx plutôt que $=$ pour rappeler qu'il manque le terme d'erreur dont la procédure de projection utilisée minimise la somme des carrés, on a :

$$X(n, p) \approx S(n, 1)F'(1, p) = S(n, 1)a\underline{F}'(1, p) \quad (1)$$

en multipliant chaque côté de l'égalité à droite par \underline{F} , on obtient :

$$\begin{aligned} X(n, p)\underline{F}(p, 1) &\approx S(n, 1)a\underline{F}'(1, p)\underline{F}(p, 1) \\ &\approx S(n, 1)a = aS(n, 1) \end{aligned} \quad (2)$$

On peut ainsi facilement déterminer a , qui doit être l'écart-type du résultat de $X(n, p)\underline{F}(p, 1)$ (i.e du vecteur " $aS(n, 1)$ " obtenu) si on veut assurer que S , le vecteur des scores sur le facteur, ait un écart-type de 1 tel que posé dans le modèle.

Si on connaît plutôt S , comment peut-on évaluer \underline{F} et a ? On a comme en (1) :

$$X(n, p) \approx S(n, 1)a\underline{F}'(1, p) \quad (3)$$

En multipliant cette fois chaque côté de l'égalité à gauche par S' , on obtient

$$\begin{aligned} S'(1, n)X(n, p) &\approx S'(1, n)S(n, 1)a\underline{F}'(1, p) \\ &\approx (n - 1)a\underline{F}'(1, p) \end{aligned} \quad (4)$$

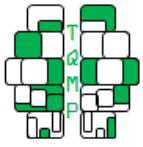
où $S'(1, n)S(n, 1)$ est la somme des carrés des scores, qu'on devrait diviser par $(n - 1)$ pour avoir la variance (et ensuite l'écart type) de S , laquelle doit être 1.

Il sera commode de transposer le dernier résultat et de diviser par $(n - 1)$, soit :

$$a\underline{F}(p, 1) \approx (1/(n - 1))X'(p, n)S(n, 1) \quad (5)$$

Considérons maintenant une série d'approximations de \underline{F} , désignées f_i , commençant avec f_0 , une hypothèse quelconque de ce que serait \underline{F} . À chaque approximation f_i correspondront une amplitude a_i et une évaluation des scores s_i . Par la méthode pour évaluer a et S à partir de \underline{F} , on a donc, selon l'équation (2) :

$$a_i s_i(n, 1) \approx X(n, p)f_i(p, 1) \quad (6)$$



De là, on peut réévaluer F par le prochain vecteur f désigné f_{i+1} ; on réévaluera en fait a_{i+1} et f_{i+1} sans sortir a_i de l'expression précédente (c'est pourquoi le terme de gauche commence par a_i). À partir de l'équation (5), on trouve :

$$\begin{aligned} aF(p, 1) &\approx (1/(n - 1))X'(p, n)S(n, 1) \\ a_i(a_{i+1}f_{i+1}(p, 1)) &\approx (1/(n - 1))X'(p, n)(a_i s_i(n, 1)) \\ a_i a_{i+1} f_{i+1}(p, 1) &\approx (1/(n - 1))X'(p, n,)X(n, p)f_i(p, 1) \end{aligned} \tag{7}$$

Or, $(1/(n - 1))X'(p, n)X(n, p) = C(p, p)$ la matrice de covariance.

Lorsque f_{i+1} est devenu essentiellement égal à f_i , alors a_{i+1} est aussi essentiellement égal à a_i et f_i est la meilleure estimation (au sens des moindres carrés des résidus) qu'on peut faire de F . On a alors

$$a^2 f(p, 1) \approx C(p, p)f(p, 1) \tag{8}$$

En soustrayant $a^2 f f'$ de C , on peut recommencer la procédure et extraire la prochaine composante principale, qui minimisera aussi le carré des résidus et sera sans corrélation avec les composantes précédentes parce que C n'aura plus de variance dans les directions de ces composantes (grâce à la soustraction de $a^2 f f'$).

Si la matrice $C(p, p)$ analysée est une matrice de corrélations réduite où la réduction est excessive, on arrivera à une composante pour laquelle l'évaluation des a_i et des f_i successifs convergera à ceci près que la polarité de f_i ou celle de a_i alternera à chaque itération. Si on s'assure que ce soit celle des a_i qui alterne, alors le produit des deux derniers a_i sera négatif. Cela correspond à une valeur propre négative. Tous les facteurs suivants auront aussi des valeurs propres négatives qui reflètent une carence de variance plutôt que de la variance supplémentaire encore à expliquer.

Open practices

📄 The *Open Material* badge was earned because supplementary material(s) are available on [the journal's web site](#).

Citation

ACHIM, A. (2020). Esprit et enjeux de l'analyse factorielle exploratoire. *The Quantitative Methods for Psychology*, 16(4), 213-247. doi :[10.20982/tqmp.16.4.p213](https://doi.org/10.20982/tqmp.16.4.p213)

Copyright © 2020, ACHIM. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Received: 10/12/2019 ~ Accepted: 25/02/2020