

L'analyse factorielle et son utilisation dans l'examen des phénomènes sociocomportementaux : quelques clarifications méthodologiques

Eric Jean^a

RÉSUMÉ. Parmi les diverses méthodes d'analyse des données, l'analyse factorielle a reçu une attention particulière de la part des chercheurs de plusieurs disciplines scientifiques. L'avènement des logiciels statistiques conviviaux permettant de procéder aisément à certaines analyses, par exemple, la validation de construit d'un instrument psychométrique, peut en partie expliquer cette popularité. Il s'avère toutefois que cette convivialité peut dans certains cas favoriser une utilisation inadéquate de l'analyse factorielle et que le choix des méthodes utilisées a peu évolué, malgré l'amélioration croissante des outils disponibles. L'objectif du présent article, qui ne prétend pas livrer une analyse exhaustive des considérations techniques relatives à ce type d'analyse, consiste à apporter quelques précisions permettant de prendre des décisions éclairées lorsque le choix d'une analyse factorielle est envisagé par un chercheur ou un étudiant-chercheur. Afin de guider la réflexion, une question précise est posée : est-il possible de faire une rotation oblique lors d'une analyse en composante principale? La réponse à cette question nécessite d'abord de distinguer certains concepts liés à l'analyse factorielle pour ensuite proposer quelques éléments de réponse.

ABSTRACT. *Among the different methods of data analysis, factor analysis has been the subject of special attention by researchers in several scientific disciplines. The advent of user-friendly statistical software can partly explain this popularity. It turns out, however, that this user-friendliness can in some cases lead to improper use of factor analysis and that practices have not changed much, despite the increasing improvement of computer technology. The purpose of this article is to provide some insight that can help researchers choose appropriate methods based on their research objectives. To guide the reflection, a question is asked: is it possible to choose an oblique rotation following a principal component analysis? The answer to this question requires some clarification that may allow a better understanding of the factor analysis.*

L'analyse factorielle, dans son sens générique, réfère à un ensemble d'analyses multivariées qui permettent de réduire un jeu de données en identifiant des facteurs. Cela permet d'identifier un plus petit nombre de variables sous-jacentes à partir d'un plus grand nombre de variables mesurées. L'analyse factorielle est également utilisée par les chercheurs afin d'examiner la structure interne (les dimensions) d'un instrument de mesure, notamment en psychologie, en sciences humaines et sociales et en sciences de la gestion. L'analyse factorielle est alors utile à la vérification de la validité de construit d'un instrument, c'est-

à-dire s'assurer que l'instrument évalue véritablement la ou les dimensions qu'il est censé mesurer. De nombreux logiciels statistiques permettent aujourd'hui la réduction de données d'une façon rapide et simplifiée tout en proposant de nombreux choix, autant sur le plan des méthodes d'extraction des données que sur celui de la rotation des facteurs, et ce, afin de produire une solution factorielle simple et facile à interpréter. Il s'avère toutefois que la convivialité offerte par les logiciels disponibles peut dans certains cas favoriser une utilisation inadéquate de l'analyse factorielle (Bourque, Poulin, Cleaver, Noël-

^a Professeur, Ph. D., CRIA, Université du Québec à Chicoutimi

Gaudreault et Kalubi, 2006) et que le choix des techniques utilisées a peu évolué au cours des dernières années (Howard, 2016). Comme les options sont nombreuses et qu'il est facile de sélectionner une technique ne convenant pas à l'objectif de recherche, il convient de distinguer deux méthodes d'analyse apparentées, mais convenant à des usages distincts : l'analyse en composante principale et l'analyse factorielle proprement dite. Le questionnement de départ à la base du cheminement présentée est le suivant : peut-on utiliser une rotation oblique lors d'une analyse en composante principale? Une réflexion en six étapes est proposée afin de bien comprendre ce dont il est question.

1) AFE et AFC. D'abord, il faut faire une distinction entre l'analyse factorielle de type exploratoire (AFE) faisant l'objet du présent article et l'analyse factorielle de type confirmatoire (AFC). Logiquement, l'AFE précède l'AFC, car elle permet d'abord de tester un modèle et de proposer une solution factorielle cohérente avec les bases théoriques et les données observées. Dans l'AFE, les facteurs sont dérivés des résultats statistiques (p. ex., avec IBM SPSS) et ne peuvent donc être identifiés qu'après l'analyse factorielle, sans nécessairement savoir préalablement combien de facteurs existent réellement ou quelles variables se regrouperont précisément dans chacun des construits. En ce sens, l'AFE impose aux chercheurs beaucoup plus de contraintes techniques que l'AFC (Pedhazur et Schmelkin, 1991). Lors d'une AFC, le chercheur doit préciser le nombre de facteurs et l'appartenance des variables aux facteurs avant d'effectuer les calculs (p. ex., avec le logiciel d'équations structurelles Lisrel). Ainsi, au lieu de nous indiquer le nombre de facteurs et leur poids factoriel (« loadings ») respectif comme avec l'AFE, l'AFC nous indique statiquement la qualité du modèle testé et nécessite donc *a priori* un modèle théorique de mesure (Hair, 2010).

Si l'objectif est, par exemple, d'examiner les qualités psychométriques d'une version en ligne d'un instrument de mesure initialement développé en format papier crayon, l'AFE pourrait être utilisée afin de vérifier si les facteurs émergents des données (construits) ainsi que l'appartenance des variables à ces facteurs reflètent les résultats obtenus par l'instrument original (Jean, 2015). Si des

données scientifiques existent sur la version traduite utilisée, l'AFE pourrait ainsi permettre de confronter les résultats obtenus à des études existantes.

Afin d'effectuer une AFE, les chercheurs peuvent, en fonction de l'objectif poursuivi, avoir recours à l'analyse en composantes principales et à l'analyse factorielle.

2) ACP et AF. Il est également important de faire une distinction claire entre l'analyse en composante principale (ACP) et l'analyse factorielle (AF) comme méthodes d'extraction des données. L'objectif de l'ACP est de pouvoir extraire le plus de variance observée possible (variance totale) avec un nombre réduit de composantes en produisant une solution mathématique unique. Toutefois, si les quelques premières composantes identifiées n'expliquent pas au moins 50 % de la variance, il y aurait peu de gains à l'utiliser (Pedhazur et Schmelkin, 1991). Quant à l'AF, elle cherche également à extraire une solution, mais l'objectif principal est d'identifier des facteurs, et non des composantes, facteurs qui reflètent ce que les variables partagent comme variance commune (Tabachnick et Fidell, 2013)¹. Dans leur article, Park, Dailey et Lemus (2002) présentent deux exemples d'utilisation de l'ACP et de l'AF. Par exemple, si un chercheur vise à regrouper les variables « revenu », « éducation » et « niveau de dette » afin de déterminer le statut socioéconomique, l'ACP pourrait être appropriée. Toutefois, si l'objectif est d'examiner l'agressivité par le regroupement des variables « tendance à perdre son sang-froid », « utilisation d'insultes » et « recours à des commentaires offensants », l'AF pourrait être envisagée. En effet, l'AF vise à faire émerger des construits latents à partir des variables observées et au besoin produire des scores comme si ces facteurs pouvaient être mesurés directement. Attention, l'utilisation de l'AF nécessite une base théorique sous-jacente et des attentes quant aux construits pouvant émerger (Pedhazur et Schmelkin, 1991). L'ACP est donc à privilégier pour la réduction du nombre de variables ou comme étape préalable à l'AF afin de révéler le nombre de facteurs possibles et leur nature potentielle. Quant à l'AF, elle est à privilégier lorsque des construits sont théoriquement envisagés dans le but de produire des scores sur les variables regroupées (Tabachnick et Fidell,

2013). Il est important de rappeler que l'ACP produit des « composantes » et l'AF des « facteurs ». Cette distinction est importante, car même si les termes sont similaires, les deux techniques sont différentes (Preacher et MacCallum, 2003). Les objectifs de recherche guideront le choix de l'une ou l'autre de ces deux techniques d'analyse.

Pour reprendre l'exemple cité au point 1, comme l'objectif du chercheur va bien au-delà d'unique-ment « extraire » le plus de variance observée possible dans les données (variance totale) avec un nombre réduit de composantes, il pourrait être préférable d'utiliser l'AF, particulièrement si la production de scores est envisagée, ce qui est généralement le cas en psychologie pour le développement d'instruments de mesure.

3) Extraction. Maintenant, il est nécessaire de différencier « extraction » et « rotation » lors d'une analyse factorielle exploratoire. L'extraction consiste essentiellement à dégager un certain nombre de facteurs à partir de la matrice de corrélation, mais ces facteurs pourraient ne pas être interprétables. L'interprétation des facteurs est donc un élément crucial à toute forme d'analyse factorielle exploratoire (Tabachnick et Fidell, 2013). Cela est particulièrement vrai en ce qui concerne l'AF. En effet, comme l'AF vise justement à refléter ce que les variables partagent entre elles (variance commune), le fait de pouvoir nommer ces facteurs devient primordial. Alors que l'ACP a pour principal objectif de réduire les nombreuses variables en quelques facteurs expliquant le maximum de variance en ordre descendant, l'AF vise principalement à identifier des construits. D'ailleurs, Preacher et MacCallum (2003) recommandent fortement de ne pas utiliser l'ACP à moins d'être spécifiquement intéressé à la réduction de données. En effet, comme les deux types d'analyse ne partagent pas un même calcul de la variance (notamment en ce qui a trait à l'erreur de mesure) utiliser malencontreusement l'ACP comme une AF rend l'interprétation des composantes problématique (Preacher et MacCallum, 2003).

4) Rotation. Quant à la rotation, c'est sur le plan de l'interprétation que celle-ci prend tout son

sens. Elle permet en effet d'améliorer, ou de simplifier, l'interprétation des facteurs extraits. Deux types de rotation sont possibles : orthogonale et oblique. La rotation orthogonale oblige les facteurs à être non corrélés (orthogonaux, ou indépendants) alors que la rotation oblique n'a pas cette restriction et autorise la corrélation entre les facteurs. Dans leur article, Preacher et McCallum (2003) précisent que de façon générale, si le chercheur ne connaît pas la nature des liens entre les facteurs, il n'y a aucune raison de présumer qu'ils sont indépendants. Il est ainsi toujours préférable de présumer la non-indépendance dans ce genre de situation. Dans leur ouvrage, Tabachnick et Fidell (2013, p. 642) affirment qu'une rotation orthogonale facilite l'interprétation, la description et la publication des résultats, mais insistent sur le fait qu'elle force la réalité, à moins que le chercheur soit convaincu de l'indépendance des facteurs. Pour leur part, Pedhazur et Schmelkin (1991, p. 615) vont plus loin et ajoutent que les solutions orthogonales (rotation), malgré leur simplicité, sont dans la plupart des cas des représentations naïves et irréelles des phénomènes sociocomportementaux. La question cruciale est donc celle-ci : les éléments du construit multidimensionnel proposé sont-ils corrélés entre eux ? Si la réponse à cette question n'est pas explicite, il est plutôt recommandé d'effectuer les deux types de rotation et de considérer la rotation orthogonale seulement si la corrélation entre les facteurs est négligeable (Pedhazur et Schmelkin, 1991). Dans le cas contraire, une rotation oblique fournit une bien meilleure information. Il est important de mentionner que les facteurs ne seront pas nécessairement corrélés en utilisant une rotation oblique. En effet, Tabachnick et Fidell (2013) rapportent que souvent, en fait, les facteurs ne sont pas corrélés et alors le chercheur présente la rotation orthogonale, plus simple.

5) Méthodes d'extraction. Lorsque le choix du chercheur s'arrête sur une analyse factorielle (AF) plutôt que sur une analyse en composante principale (ACP), plusieurs méthodes d'extraction sont alors offertes par les logiciels statistiques spécialisés. Par exemple, celles offertes par IBM SPSS sont présentées au tableau 1.

<ul style="list-style-type: none"> ▪ Analyse des composantes principales (Principal components) <ul style="list-style-type: none"> ○ Cette option apparaît au menu, mais correspond en fait à l'ACP ▪ Factorisation en axes principaux (Principal axis factoring) ▪ Alpha de Cronbach (Alpha factoring) ▪ Factorisation en projection (Image factoring) ▪ Méthode des moindres carrés non pondérés (Unweighted least squares factoring) ▪ Méthode des moindres carrés généralisés (Generalized (weighted) least squares factoring) ▪ Méthode du maximum de vraisemblance. (Maximum likelihood factor extraction)

Tableau 1 – Exemples de méthodes d'extraction en analyse factorielle (IBM SPSS)

Quoique chacune de ces méthodes d'extraction pour l'AF possède ses particularités, le choix de l'une ou l'autre peut devenir un exercice futile, car l'impact de ces différences est peu documenté et en pratique les résultats peuvent être similaires peu importe la méthode utilisée, sous certaines conditions (Tabachnick et Fidell, 2013). En effet, ceci est particulièrement vrai lorsque le nombre de variables est grand et qu'il existe de fortes corrélations entre certaines d'entre elles, lorsque le nombre de facteurs est le même entre les méthodes et que ce nombre est décidé avec soin, et que la qualité des représentations (« communalités ») des variables est similaire. Bref, lorsqu'une solution factorielle est stable, elle le demeure, peu importe la méthode d'extraction (ou de rotation) utilisée. De plus, les différences qui ressortent entre les méthodes après

l'extraction ont tendance à disparaître à la suite d'une rotation (Tabachnick et Fidell, 2013). Les méthodes d'AF largement utilisées par les chercheurs sont la « factorisation en axes principaux » et le « maximum de vraisemblance » (McNeish, 2017).

6) Méthode de rotation. Un dernier point, la rotation. Ici encore, plusieurs options s'offrent aux chercheurs et elles présentent chacune leurs avantages et leurs inconvénients. Selon le logiciel statistique utilisé, ces choix peuvent varier. En ce qui concerne IBM SPSS, les options offertes sont présentées au tableau 2, et ce, autant pour l'ACP que pour l'AF.

Orthogonales	Obliques
<ul style="list-style-type: none"> ▪ Varimax ▪ Quartimax ▪ Equamax 	<ul style="list-style-type: none"> ▪ Oblimin ▪ Promax

Tableau 2 – Exemples de méthodes de rotation en analyse factorielle (IBM SPSS)

Le choix d'une méthode de rotation dépend du type d'analyse effectuée. Par exemple, la méthode Varimax s'est historiquement avérée adaptée comme approche analytique afin d'obtenir des facteurs orthogonaux, c'est-à-dire qui ne sont pas corrélés entre eux (Hair, 2010). Malgré la popularité de la rotation Varimax et comme les outils statistiques sont beaucoup plus performants et accessibles que par le passé, Preacher et McCallum (2003) recommandent d'éviter l'utilisation d'une rotation ortho-

gonale de façon systématique en toutes circonstances et propose plutôt l'utilisation d'une rotation oblique, car les facteurs sont rarement non corrélés dans les études empiriques. L'une des meilleures méthodes semble être Oblimin (Tabachnick et Fidell, 2013) et de ce fait, celle-ci est fréquemment utilisée par les chercheurs.

Enfin, pour Pedhazur et Schmelkin (1991) la rotation de composantes (ici l'ACP) de même que le fait d'y associer une signification ne font pas de sens en soi, même si cela est possible, car la rotation a pour objectif d'améliorer l'interprétation des variables regroupées alors que l'ACP vise essentiellement la réduction de données. Pourtant, certains ouvrages de référence consultés (Hair, 2010; Tabachnick et Fidell, 2013) ne semblent pas faire explicitement mention de cette mise en garde sur la rotation en ACP.

Conclusion

En conclusion, pour répondre à la question de départ, oui il est possible d'effectuer une rotation oblique lors d'une extraction en composantes principales. Néanmoins, dans la mesure où l'ACP vise principalement la réduction de données et non

l'identification de construits latents, son utilisation doit être pleinement justifiée. Ce qui pouvait sembler questionable, au départ, était plutôt l'idée d'effectuer une rotation lors d'une ACP (peu importe le type de rotation). Toutefois, comme cette réserve n'est soulevée que par quelques auteurs et que l'utilisation de rotations en ACP est largement répandue, son utilisation demeure recommandée tout en gardant à l'esprit que l'ACP et l'AF poursuivent des objectifs distincts. Il faut également retenir que la rotation oblique est préférable dans le domaine de la psychologie et des sciences sociales en général, car il existe très souvent des relations entre les concepts étudiés. Compte tenu de ses avantages, il est possible de trouver des défenseurs de la rotation orthogonale, mais il semble que les écrits convergent vers une interprétation plus restrictive de son utilisation.

NOTE

- 1 Notez que les auteurs Tabachnick et Fidell précisent dans le texte qu'il existe une distinction entre facteur et composante, mais utilisent par la suite uniquement la notion de facteurs afin de minimiser la confusion.

RÉFÉRENCES

- Barbara, P., Sara, C. et Janine, C. (2013). Work-family and work-life pressures in Australia: advancing gender equality in "good times"? *International Journal of Sociology and Social Policy*, 33(9/10), 594-612. <http://dx.doi.org/10.1108/IJSSP-11-2012-0100>.
- Bourque, J., Poulin, N., Cleaver, A. F., Noël-Gaudreault, M. et Kalubi, J.-C. (2006). Évaluation de l'utilisation et de la présentation des résultats d'analyses factorielles et d'analyses en composantes principales en éducation. *Revue des sciences de l'éducation*, 32(2), 325-344.
- Hair, J. F. (2010). *Multivariate data analysis* (7^e éd.). Upper Saddle River, N.J.: Prentice Hall.
- Howard, M. C. (2016). A review of exploratory factor analysis decisions and overview of current practices: what we are doing and how can we improve? *International Journal of Human-Computer Interaction*, 32(1), 51-62. <http://dx.doi.org/10.1080/10447318.2015.1087664>.
- Jean, E. (2015). Les enjeux liés à la collecte de données en ligne : le cas d'une recherche auprès de gestionnaires. *La Revue des Sciences de Gestion*, 2(272), 13-21.
- McNeish, D. (2017). Exploratory factor analysis with small samples and missing data. *Journal of Personality Assessment*, 99(6), 637-652. <http://dx.doi.org/10.1080/00223891.2016.1252382>.
- Park, H. S., Dailey, R. et Lemus, D. (2002). The use of exploratory factor analysis and principal components analysis in communication research. *Human Communication Research*, 28(4), 562-577. <http://dx.doi.org/10.1111/j.1468-2958.2002.tb00824.x>.
- Pedhazur, E. J. et Schmelkin, L. P. (1991). *Measurement, design, and analysis: An integrated approach* (édition de l'étudiant). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Preacher, K. J. et MacCallum, R. C. (2003). Repairing Tom Swift's electric factor analysis machine. *Understanding Statistics*, 2(1), 13-43.
- Tabachnick, B. G. et Fidell, L. S. (2013). *Using multivariate statistics* (6^e éd.). Boston, MA: Pearson Education.