Méthodes psychométriques en qualité de vie

Christophe Lalanne
EA 7334 REMES
Unité de Méthodologie des critères d'évaluation
Université Paris-Diderot, Sorbonne Paris-Cité

30 ANS
ÉCOLE D'ÉTÉ DE SANTÉ PUBLIQUE ET D'ÉPIDÉMIOLOGIE DE BICÊTRE
Institut National de la Santé et de la Recherche Médicale - Université Paris-Sud
Faculté de Médecine Paris-Sud - Ecole de Santé Publique Paris-Sud

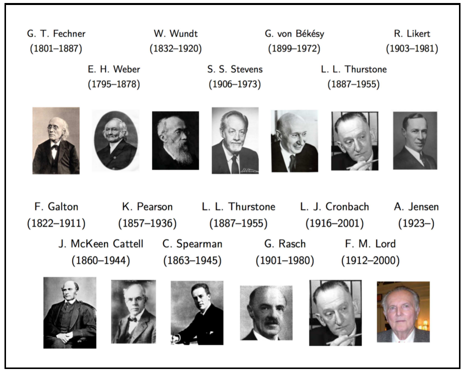
Analyses factorielles

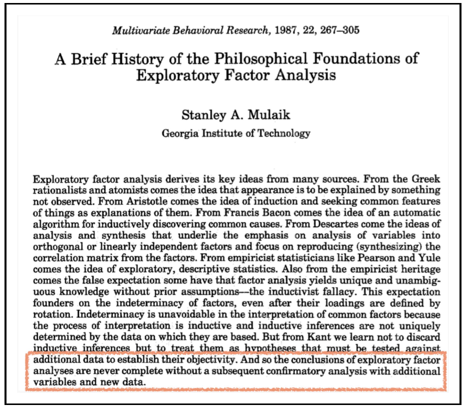
- Analyse en composantes principales et analyse factorielle
- Analyse factorielle exploratoire
- Analyse factorielle confirmatoire

It is rather surprising that systematic studies of human abilities were not undertaken until the second half of the last century. . . An accurate method was available for measuring the circumference of the earth 2,000 years before the first systematic measures of human ability were developed. [1]

1. J Nunnally et I Bernstein. *Psychometric Theory*. 3rd. McGraw-Hill, 1994.

Avant Jan de Leeuw & Bengt Muthén





Multivariate Behavioral Research, 1987, 22, 267–305

A Brief History of the Philosophical Foundations of Exploratory Factor Analysis

Stanley A. Mulaik

Georgia Institute of Technology

Exploratory factor analysis derives its key ideas from many sources. From the Greek rationalists and atomists comes the idea that appearance is to be explained by something not observed. From Aristotle comes the idea of induction and seeking common features of things as explanations of them. From Francis Bacon comes the idea of an automatic algorithm for inductively discovering common causes. From Descartes come the ideas of analysis and synthesis that underlie the emphasis on analysis of variables into orthogonal or linearly independent factors and focus on reproducing (synthesizing) the correlation matrix from the factors. From empiricist statisticians like Pearson and Yule comes the idea of exploratory, descriptive statistics. Also from the empiricist heritage comes the false expectation some have that factor analysis yields unique and unambiguous knowledge without prior assumptions—the inductivist fallacy. This expectation founders on the indeterminacy of factors, even after their loadings are defined by rotation. Indeterminacy is unavoidable in the interpretation of common factors because the process of interpretation is inductive and inductive inferences are not uniquely determined by the data on which they are based. But from Kant we learn not to discard inductive inferences but to treat them as hypotheses that must be tested against additional data to establish their objectivity. And so the conclusions of exploratory factor analyses are never complete without a subsequent confirmatory analysis with additional variables and new data.

ACP et AF

▷ 01a-scores.pdf

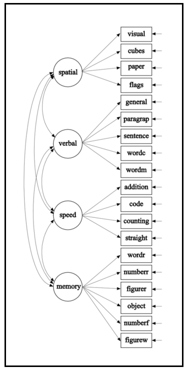
Les composantes C_i $(i = 1, \ldots, p)$ de l'analyse en composantes principales (ACP) sont construites comme de simples combinaisons linéaires des p variables d'origine : $C_i = \sum_{j=1}^{p} w_{ij} x_j$.

Dans le cadre de l'analyse factorielle, on considère au contraire des combinaisons linéaires de facteurs[2] :

$$x_i \approx \sum_{j=1}^{k} w_{ij} F_j.$$

2. W Revelle. *An introduction to psychometric theory with applications in R*. http ://www.personality-project.org/r/book/. 2016, chap. 6.

Modèle de Holzinger & Swineford



Comparaison ACP *versus* AF

```
library(psych)
principal(HS[,c("visual","cubes", paper")], nfactors = 3,
          rotate = "none")

Standardized loadings (pattern matrix) based upon correlation matrix
       PC1   PC2   PC3 h2       u2 com
visual 0.77 -0.41  0.48  1  0.0e+00 2.3
cubes  0.70  0.71  0.10  1 -4.4e-16 2.0 ❶
paper  0.80 -0.22 -0.56  1 -6.7e-16 2.0

                       PC1  PC2  PC3
SS loadings           1.72 0.72 0.56
Proportion Var        0.57 0.24 0.19
```

```
fa(HS[,c("visual", "cubes", "paper")], nfactors = 1)

Standardized loadings (pattern matrix) based upon correlation matrix
       MR1   h2   u2 com
visual 0.62 0.39 0.61   1
cubes  0.48 0.23 0.77   1 ❶
paper  0.71 0.50 0.50   1

               MR1
SS loadings    1.12
Proportion Var 0.37
```

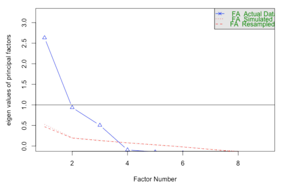

Sélection de modèle

- Sélection des variables à inclure : analyse d'items ou hypothèses *a priori*
- Sélection du nombre de facteurs : méthode exploratoire, hypothèses *a priori*, analyse parallèle [3]
- Type de rotation : en fonction des hypothèses théoriques
- Méthode d'estimation (OLS, ML, WLS et PA)
- Matrice de corrélation (Pearson, tétra- ou polychorique)
- Nombre de sujets nécessaires [4]

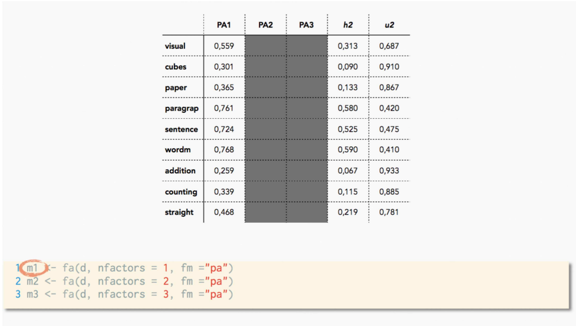
3. Lloyd G. Humphreys et Richard G. Montanelli. « An investigation of the parallel analysis criterion for determining the number of common factors ». In : *Multivariate Behavioral Research* 10 (1975), p. 193–205.
4. Rouquette A et Falissard B. « Sample size requirements for the internal validation of psychiatric scales ». In : *International Journal of Methods in Psychiatric Research* 20.4 (2011), p. 235–249.

Analyse parallèle

```
d <- HS[,7:15]
describe(d)
fa.parallel(d, fm = "pa", fa = "fa", main = "")
```



Solution factorielle à 1 facteur



	PA1	PA2	PA3	h2	u2
visual	0,559			0,313	0,687
cubes	0,301			0,090	0,910
paper	0,365			0,133	0,867
paragrap	0,761			0,580	0,420
sentence	0,724			0,525	0,475
wordm	0,768			0,590	0,410
addition	0,259			0,067	0,933
counting	0,339			0,115	0,885
straight	0,468			0,219	0,781

```
m1 <- fa(d, nfactors = 1, fm ="pa")
m2 <- fa(d, nfactors = 2, fm ="pa")
m3 <- fa(d, nfactors = 3, fm ="pa")
```

Solution factorielle à 2 facteurs

	PA1	PA2	PA3	*h2*	*u2*
visual	0,275	0,430		0,341	0,659
cubes	0,134	0,244		0,100	0,900
paper	0,060	0,449		0,223	0,777
paragrap	0,851	0,005		0,727	0,273
sentence	0,854	-0,038		0,709	0,291
wordm	0,828	0,033		0,705	0,295
addition	-0,034	0,434		0,180	0,820
counting	-0,083	0,642		0,383	0,617
straight	0,007	0,734		0,542	0,458

```
m1 <- fa(d, nfactors = 1, fm ="pa")
m2 <- fa(d, nfactors = 2, fm ="pa")
m3 <- fa(d, nfactors = 3, fm ="pa")
```


Solution factorielle à 3 facteurs

	PA1	PA2	PA3	*h2*	*u2*
visual	0,196	0,591	0,032	0,477	0,523
cubes	0,043	0,510	-0,121	0,256	0,744
paper	-0,062	0,685	0,020	0,453	0,547
paragrap	0,846	0,016	0,007	0,728	0,272
sentence	0,885	-0,065	0,007	0,753	0,247
wordm	0,805	0,080	-0,013	0,692	0,308
addition	0,045	-0,154	0,732	0,512	0,488
counting	-0,034	0,121	0,691	0,524	0,476
straight	0,032	0,380	0,458	0,461	0,539

```
m1 <- fa(d, nfactors = 1, fm ="pa")
m2 <- fa(d, nfactors = 2, fm ="pa")
m3 <- fa(d, nfactors = 3, fm ="pa")
```


Analyse exploratoire et confirmatoire (CFA) [5]

La CFA revient à imposer une structure particulière, c.a.d. à contraindre certains paramètres du modèle, et à tester l'adéquation du modèle avec les données.

	PA1	PA2	PA3	*h2*	*u2*
visual		0,591		0,477	0,523
cubes	0	0,510		0,256	0,744
paper		0,685	0	0,453	0,547
paragrap	0,846			0,728	0,272
sentence	0,885			0,753	0,247
wordm	0,805	0		0,692	0,308
addition			0,732	0,512	0,488
counting	0		0,691	0,524	0,476
straight			0,458	0,461	0,539

5. DL Jackson, JA Gillaspy et R Purc-Stephenson. « Reporting practices in confirmatory factor analysis : An overview and some recommendations ». In : *Psychological Methods* 14.1 (2009), p. 6–23 ; L.T. Hu et P.M. Bentler. « Cutoff criteria for fit indexes in covariance structure analysis : Conventional criteria versus new alternatives ». In : *Structural Equa8on Modeling* 6 (1999), p. 1–55.

Utilisation de `lavaan`

Le package `lavaan` [6] dispose de 4 commandes essentielles : `lavaan`, `cfa`, `sem`, `growth`.

Les commandes inclut entre autres des procédures d'estimation par intervalles (bootstrap), de simulation et de transfert de données/modèles avec Mplus. [7]

`http://lavaan.ugent.be`

6. Y Rosseel. « lavaan, An R Package for Structural Equation Modeling ». In : *Journal of Statistical SoftwareJournal of Statistical Software* 48.2 (2012), p. 1–36.
7. AA Beaujean. « Factor Analysis Using R ». In : *Practical Assessment, Research & Evaluation* 18.4 (2013), p. 1–11 ; AA Beaujean. *Latent Variable Modeling Using R, A Step-by-Step Guide*. New York : Routledge, 2014.

Modèle en traits corrélés

```
library(lavaan)
library(semPlot)
m <- 'Visual =~ visual + cubes + paper                      ❶
      Verbal =~ paragrap + sentence + wordm
      Speed  =~ addition + counting + straight'
r <- cfa(m, data = d)                                       ❷
summary(r, fit.measures = TRUE)                             ❸
semPaths(r, whatLabels = "est")                             ❹
```

```
> summary(r, fit.measures = TRUE)
lavaan (0.5-20) converged normally after  35 iterations

  Number of observations                           301
```



```
  Estimator                                         ML
  Minimum Function Test Statistic               85.306
  Degrees of freedom                                24
  P-value (Chi-square)                           0.000

Model test baseline model:

  Minimum Function Test Statistic              918.852
  Degrees of freedom                                36
  P-value                                        0.000

User model versus baseline model:

  Comparative Fit Index (CFI)                    0.931        ❶
  Tucker-Lewis Index (TLI)                       0.896

Loglikelihood and Information Criteria:
```



```
  Loglikelihood user model (H0)              -3737.745
  Loglikelihood unrestricted model (H1)      -3695.092

  Number of free parameters                         21
  Akaike (AIC)                                7517.490
  Bayesian (BIC)                              7595.339
  Sample-size adjusted Bayesian (BIC)         7528.739

Root Mean Square Error of Approximation:

  RMSEA                                          0.092        ❷
  90 Percent Confidence Interval          0.071  0.114
  P-value RMSEA <= 0.05                          0.001

Standardized Root Mean Square Residual:

  SRMR                                           0.065        ❸
```



```
Parameter Estimates:

  Information                                 Expected
  Standard Errors                             Standard

Latent Variables:
                   Estimate  Std.Err  Z-value  P(>|z|)
  Visual =~
    visual            1.000                                   ❹
    cubes             0.554    0.100    5.554    0.000
    paper             0.729    0.109    6.685    0.000
  Verbal =~
    paragrap          1.000
    sentence          1.113    0.065   17.014    0.000
    wordm             0.926    0.055   16.703    0.000
  Speed =~
    addition          1.000
```

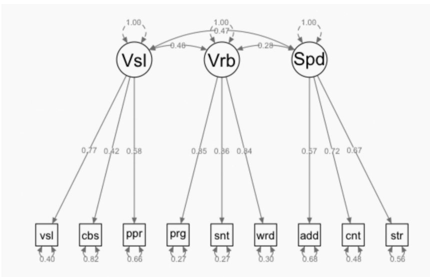
```
    counting          1.180    0.165    7.152    0.000
    straight          1.082    0.151    7.155    0.000

Covariances:
                   Estimate  Std.Err  Z-value  P(>|z|)
  Visual ~~
    Verbal            0.408    0.074    5.552    0.000
    Speed             0.262    0.056    4.660    0.000
  Verbal ~~
    Speed             0.173    0.049    3.518    0.000

Variances:
                   Estimate  Std.Err  Z-value  P(>|z|)
    visual            0.549    0.114    4.833    0.000
    cubes             1.134    0.102   11.146    0.000
    paper             0.844    0.091    9.317    0.000
    paragrap          0.371    0.048    7.779    0.000
    sentence          0.446    0.058    7.642    0.000
```

❺


```
    wordm             0.356    0.043    8.277    0.000
    addition          0.799    0.081    9.823    0.000
    counting          0.488    0.074    6.573    0.000
    straight          0.566    0.071    8.003    0.000
    Visual            0.809    0.145    5.564    0.000
    Verbal            0.979    0.112    8.737    0.000
    Speed             0.384    0.086    4.451    0.000
```

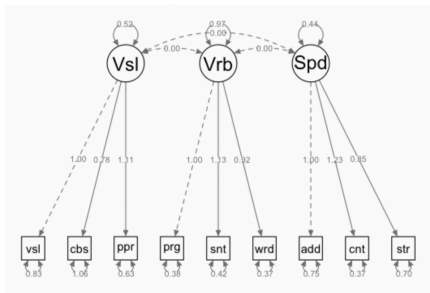

Modèle en traits corrélés (paramétrisation alternative)

```
r <- cfa(m, data = d, std.lv = TRUE)
```



Modèle en traits orthogonaux

```
r <- cfa(m, data = d, orthogonal = TRUE)
```



Modèles de mesure en analyse factorielle

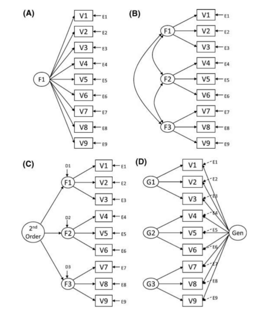


FIGURE 1.—Model A, a unidimensional model; Model B, a correlated traits model; Model C, a second-order model; and Model D, a bifactor model. F = factor; E = error; D = disturbance; V = measured variable.

Modèles alternatifs pour les données HS

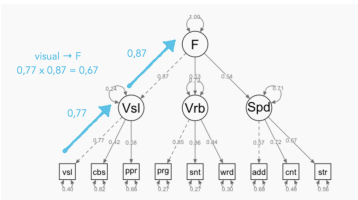
— **Modèle de second ordre** : modèle de mesure placé directement au niveau de la corrélation entre les facteurs spécifiques : les facteurs sont corrélés car ils « partagent une cause commune ». L'effet facteur primaire est appelé effet indirect.

— **Modèle bifactoriel** : tous les items sont associés à un même facteur général, et ce modèle inclut des facteurs spécifiques orthogonaux, appelés facteurs communs, qui résument la variance non expliquée par le facteur général.[8]

8. S.P. Reise, T.M. Moore et M.G. Haviland. « Bifactor Models and Rotations : Exploring the Extent to Which Multidimensional Data Yield Univocal Scale Scores ». In : *Journal of Personality Assessment* 92.6 (2010), p. 544–559.

Modèle de second ordre

```
m3 <- 'Visual =~ visual + cubes + paper
       Verbal =~ paragrap + sentence + wordm
       Speed  =~ addition + counting + straight
       F =~ Visual + Verbal + Speed'
```



Applications

1. Refaire l'analyse CFA séparément dans les deux groupes définis par la variable `school` et comparer les charges factorielles entre les deux échantillons.
2. À partir des données `HADS.RData`,[9] réaliser une analyse d'items et vérifier la dimensionnalité de l'échelle.
3. Estimer les paramètres d'un modèle CFA en traits corrélés et non corrélés.

9. F. Bartolucci, S. Bacci et M. Gnaldi. *Statistical Analysis of Questionnaires. A unified approach based on R and Stata*. CRC Press, Taylor & Francis Group, 2016.

Fichier de données et scripts R disponibles à l'adresse suivante :
`https://bitbucket.org/chlalanne/eespe11`

– Typeset with FoilTEX (version 2), Revision f4328f7