

Méthodes psychométriques en qualité de vie

Christophe Lalanne

EA 7334 REMES

Unité de Méthodologie des critères d'évaluation

Université Paris-Diderot, Sorbonne Paris-Cité





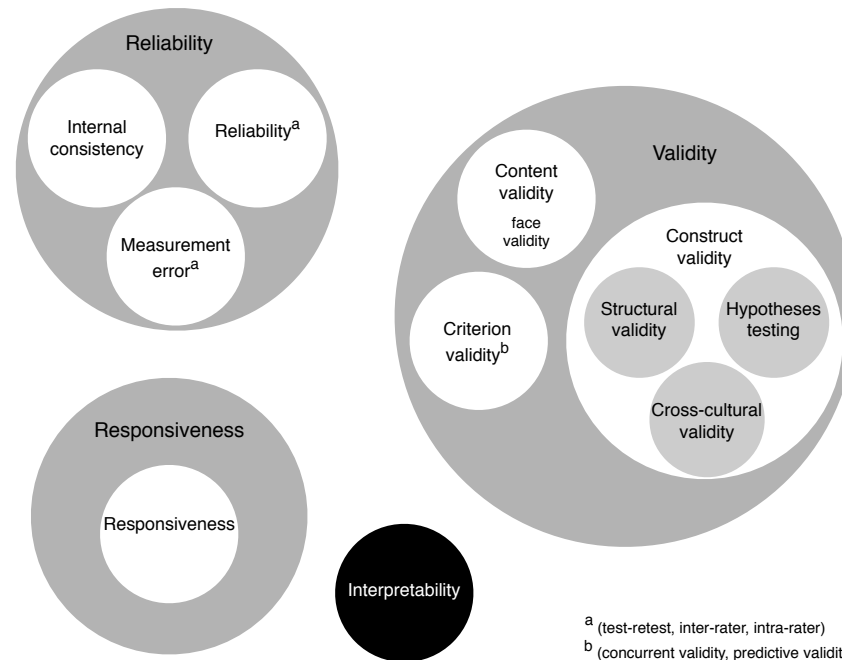
Théorie classique des tests

- Définition des mesures subjectives en santé, intérêt et enjeu.
- Approche psychométrique de la mesure.
- Propriétés d'un bon instrument de mesure (validité, fidélité de mesure).
- Modèles de mesure, modèles à variables latentes.

Validation of instruments is the process of determining whether there are grounds for believing that the instrument measures what it is intended to measure, and that it is useful for its intended purpose.¹

1. PM FAYERS et D MACHIN. *Quality of life. The assessment, analysis and interpretation of patient-reported outcomes*. Wiley, 2000.

Taxonomie COSMIN²



2. LB MOKKINK et al. « The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes ». In : *Journal of Clinical Epidemiology* 63 (2010), p. 737–745.

Instruments et mesures

Un instrument est utilisé pour mettre en relation ou associer **quelque chose observé dans le monde réel** à **quelque chose de mesurable dans un certain cadre conceptuel**.

On parlera de **variable manifeste** (p.ex., réponse à la question "vous sentez-vous déprimé le matin en vous réveillant") et de **variable latente** (p.ex., "état dépressif" du patient).

On peut voir le processus de mesure comme une tâche consistant à assigner des nombres à des catégories.³

3. SS STEVENS. « On the theory of scales of measurement ». In : *Science* 103 (1946), p. 677–680 ; P DE BOECK, M WILSON et GS ACTON. « A Conceptual and Psychometric Framework for Distinguishing Categories and Dimensions ». In : *Psychological Review* 112.1 (2005), p. 129–158.

Construction d'un instrument de mesure⁴

- A **Construct map** features a coherent and substantive definition for the content of the construct which is composed of an underlying continuum (for ordering respondents and/or items responses).
- **Items design** deals with the standardized construction of items that are supposed to stimulate responses, assimilable to observations about the construct.
- The **Outcome space** is the set of well-defined categories, finite and exhaustive, ordered, context-specific, and research-based.
- A **Measurement model** is needed in order to relate the scored outcomes from the items design and the outcome space back to the construct that was the original inspiration of the items.

4. M WILSON. *Constructing measures. An item response modeling approach*. Taylor & Francis Group, 2005.

Modèles à variables latentes

La nature des variables manifestes et latentes permet généralement de guider le choix d'un modèle statistique.⁵

		Variables manifestes	
		Métrique	Catégorielle
Variables latentes	Métrique	Analyse factorielle	Analyse en traits latents
	Catégorielle	Analyse en profils latents	Analyse en classe latente

5. DJ BARTHOLOMEW et M KNOTT. *Latent Variable Models and Factor Analysis : A Unified Approach*. 3rd. Wiley, 2011; S RABE-HESKETH et A SKRONDAL. « Classical latent variable models for medical research ». In : *Statistical methods in medical research* (2008).

Des items aux échelles de mesure

Un « item » constitue l'unité de mesure fondamentale et il peut être considéré comme une évaluation critériée d'un construit théorique (« construct ») psychologique (fluence verbale, habileté cognitive, qualité de vie, traits de personnalité).

L'aggrégation des réponses fournies par plusieurs individus à plusieurs items permet de construire une échelle de mesure, souvent confondue avec la dimension (construct) qu'elle est supposée refléter.

En supposant que cette échelle soit bel et bien unidimensionnelle, il est alors possible d'assigner un score numérique aux patterns de réponse individuels et ainsi positionner l'individu sur un trait latent.

Illustration

MOS 36-item version courte (1 = « all the time », . . . , 6 = « none of the time »)

		1	2	3	4	5	6
23.	Did you feel full of pep?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
24.	Have you been a very nervous person?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
25.	Have you felt so down in the dumps that nothing could cheer you up?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
26.	Have you felt calm and peaceful?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
27.	Did you have a lot of energy?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
28.	Have you felt downhearted and blue?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
29.	Did you feel worn out?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
30.	Have you been a happy person?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
31.	Did you feel tired?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Type d'items

- **item dichotomique** (variable binaire) : deux options de réponse ou plusieurs options de réponse mais une seule réponse correcte (p.ex., QCM) ;
- **item polytomique** (variable catégorielle ordinale) : options ou catégories de réponse naturellement ordonnées (p.ex., échelle de Likert)
- score numérique (standardisé ou non) : application d'un algorithme pour dériver un score de résumé
- choix préférentiels : choix ordonnés de réponses parmi un ensemble d'options de réponse

Exemple d'échelles multidimensionnelles

- Le questionnaire **HADS** est composé de deux dimensions (2×7 items) permettant d'explorer les symptômes anxio-dépressifs.⁶
- Le **NEOPI-R** est un inventaire de personnalité (240 ou 60 items) basé sur le modèle en 5 facteurs (extraversion, agreeableness, conscientiousness, neuroticism, and openness; 6 facettes chacun).⁷
- Le questionnaire **MOS-HIV** évalue la qualité de vie spécifique du VIH à l'aide de 35 items couvrant 10 dimensions suivantes.⁸

6. AS ZIGMOND et RP SNAITH. « The hospital anxiety and depression scale ». In : *Acta Psychiatry Scandinavica* 67.6 (1983), p. 361–370.

7. RR McCRAE et OP JOHN. « An introduction to the five-factor model and its applications ». In : *Journal of Personality* 60.2 (1992), p. 175–215.

8. AW WU et al. « Applications of the Medical Outcomes Study health-related quality of life measures in HIV/AIDS ». In : *Quality of Life Research* 6.6 (1997), p. 531–554.

Propriétés attendues (en terme de mesure)

- **unidimensionalité** : l'échelle de mesure mesure un seul construct qui est supposé expliquer les réponses d'un individu aux items selon sa position sur le trait latent ;
- **indépendance locale** : si le trait latent est maintenu à une valeur constante, les réponses aux items sont indépendantes ;
- **invariance de mesure** : les réponses individuelles ne dépendent pas de caractéristiques intrinsèques aux individus (p.ex., sexe, origine ethnique).

Les deux premières propriétés sont indispensables pour pouvoir attribuer un score à l'individu (et aux items) sur le trait latent, tandis que la troisième propriété permet des comparaisons inter-groupes valides.

Les différentes facettes de la validité

Plusieurs définitions ont été proposées sans grand consensus.⁹

- **validité de contenu** : adéquation des domaines ou dimensions couvertes par les items ;
- **validité de critère** : les échelles et donc les scores construits à partir d'elles ont des associations empiriques avec des critères externes (gold standard ou instruments mesurant des constructs similaires) ;
- **validité de construit** (convergente et discriminante/divergente) : relation cohérentes des items et des sous-échelles entre elles.

9. B FALISSARD. *Mesurer la subjectivité en santé. Perspective méthodologique et statistique*. Masson, 2008 ; BD ZUMBO. « Validity : Foundational Issues and Statistical Methodology ». In : *Handbook of Statistics, Vol. 26 : Psychometrics*. Sous la dir. de CR RAO et S SINHARAY. Elsevier Science B.V. : The Netherlands, 2007, p. 45–79.

Théorie classique des tests (TCT)

Encore appelée « théorie du score vrai », cette approche permet de formaliser la construction d'un score permettant de caractériser un individu par rapport à une échelle de mesure et de quantifier les sources potentielles de variabilité ou d'erreur.

Le **score** qui est assigné à un individu à partir de ses réponses à un instrument de mesure standardisé (à un moment donné et dans des conditions spécifiques) est ainsi envisagé comme une combinaison du **score vrai** et d'une **erreur de mesure**.

Modèle de mesure

Pour un individu i évalué sur une seule occasion, son score x_i peut être exprimé comme suit :

$$x_i = \tau_i + \varepsilon_i \quad (i = 1, \dots, n), \quad \varepsilon_i \sim \mathcal{N}(0; \sigma_e^2),$$

d'où l'on en déduit naturellement que $\mathbb{E}(x) = \tau$.

Si l'on suppose que T et E sont indépendants, on a également

$$\mathbb{V}(X) = \mathbb{V}(T) + \mathbb{V}(E).$$

Comment construire X

Une façon simple de dériver un score résumant l'ensemble des réponses d'un individu à un instrument comportant plusieurs items consiste à agréger (par sommation ou moyennage) l'ensemble des scores numériques associés à chaque item.

L'aggrégation peut être simple (non pondérée) mais il est également possible de considérer une combinaison linéaire des scores aux items. Les poids de chaque score peuvent être simples (coefficients entiers de pondération) ou plus complexes (charges factorielles).

Que trouve-t-on dans ε

L'erreur de mesure peut se décomposer en deux sources distinctes :

1. erreur systématique (ou biais), à éliminer au maximum dans la mesure du possible ;
2. erreur aléatoire, propre à la variabilité intrinsèque de l'instrument de mesure.

Borsboom discute extensivement l'intérêt et les limitations d'une telle approche pour la construction de scores, notamment la définition circulaire de la TCT.¹⁰

10. D BORSBOOM. « The attack of the psychometricians ». In : *Psychometrika* 71.3 (2006), p. 425–440 ;
D BORSBOOM. *Measuring the mind : Conceptual issues in contemporary psychometrics*. Cambridge : Cambridge University Press, 2005.

Illustration

301 enfants de deux écoles auxquels on a administré 26 tests permettant d'évaluer les compétences suivantes : spatiales, verbales, vitesse de raisonnement, mémoire, mathématiques.

Holzinger, K. J. and Swineford, F. A. *A study in factor analysis : The stability of a bi-factor solution*. Supplementary Education Monographs, 48. University of Chicago, 1939.

Illustration

	id ↕	sex ↕	ageyr ↕	agemo ↕	school ↕	grade ↕	visual ↕	cubes ↕	paper ↕	paragrap ↕	sentence ↕	wordm ↕	addition ↕	counting ↕	straight ↕
1	1	1	13	1	Pasteur	7	3.333	7.75	0.375	2.333	5.75	1.286	3.39	5.75	6.36
2	2	2	13	7	Pasteur	7	5.333	5.25	2.125	1.667	3.00	1.286	3.78	6.25	7.92
3	3	2	13	1	Pasteur	7	4.500	5.25	1.875	1.000	1.75	0.429	3.26	3.90	4.42
4	4	1	13	2	Pasteur	7	5.333	7.75	3.000	2.667	4.50	2.429	3.00	5.30	4.86
5	5	2	12	2	Pasteur	7	4.833	4.75	0.875	2.667	4.00	2.571	3.70	6.30	5.92
6	6	2	14	1	Pasteur	7	5.333	5.00	2.250	1.000	3.00	0.857	4.35	6.65	7.50
7	7	1	12	1	Pasteur	7	2.833	6.00	1.000	3.333	6.00	2.857	4.70	6.20	4.86
8	8	2	12	2	Pasteur	7	5.667	6.25	1.875	3.667	4.25	1.286	3.39	5.15	3.67
9	9	2	13	0	Pasteur	7	4.500	5.75	1.500	2.667	5.75	2.714	4.52	4.65	7.36
10	11	2	12	5	Pasteur	7	3.500	5.25	0.750	2.667	5.00	2.571	4.13	4.55	4.36
11	12	1	12	2	Pasteur	7	3.667	5.75	2.000	2.000	3.50	1.571	3.74	5.70	4.31
12	13	1	12	11	Pasteur	7	5.833	6.00	2.875	2.667	4.50	2.714	3.70	5.15	4.14
13	14	2	12	7	Pasteur	7	5.667	4.50	4.125	2.667	4.00	2.286	5.87	5.20	5.86
14	15	2	12	8	Pasteur	7	6.000	5.50	1.750	4.667	4.00	1.571	5.13	4.70	4.44
15	16	1	12	6	Pasteur	7	5.833	5.75	3.625	5.000	5.50	3.000	4.00	4.35	5.86
16	17	2	12	1	Pasteur	7	4.667	4.75	2.375	2.667	4.25	0.714	4.09	3.80	5.14
17	18	2	14	11	Pasteur	7	4.333	4.75	1.500	2.000	4.00	1.286	3.70	6.65	5.25

Illustration

```
data(HolzingerSwineford1939, package="lavaan")  
HS <- HolzingerSwineford1939  
names(HS)[7:15] <- c("visual", "cubes", "paper",  
                    "paragrap", "sentence", "wordm",  
                    "addition", "counting", "straight")
```

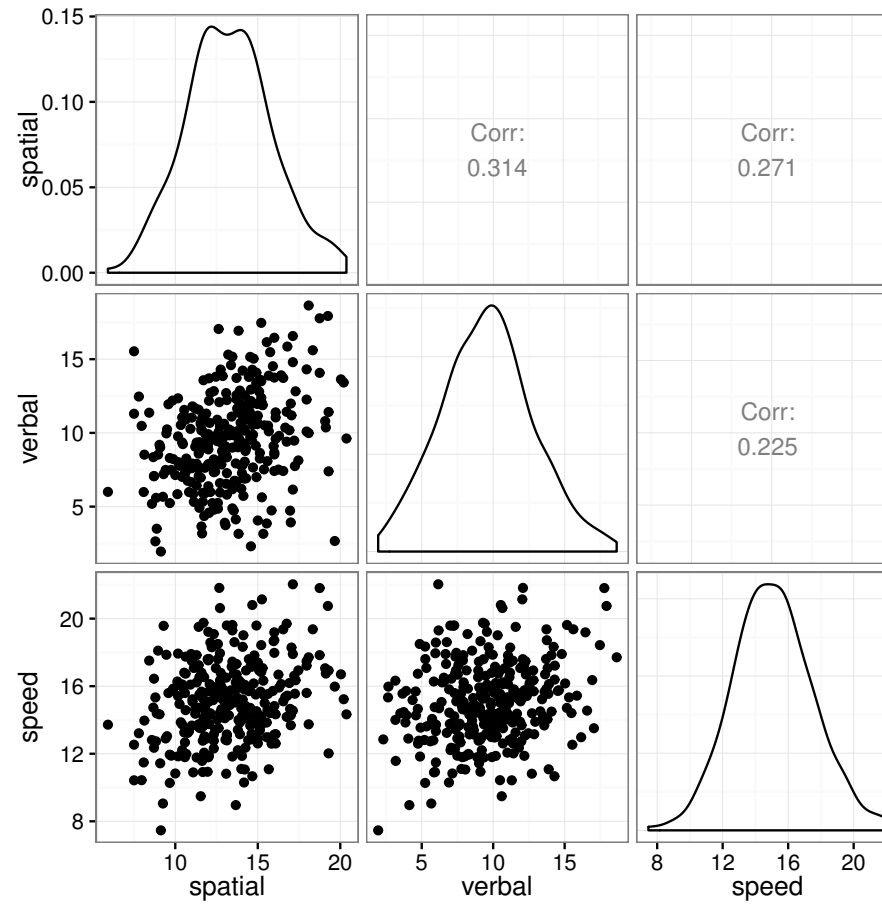
①
②
③

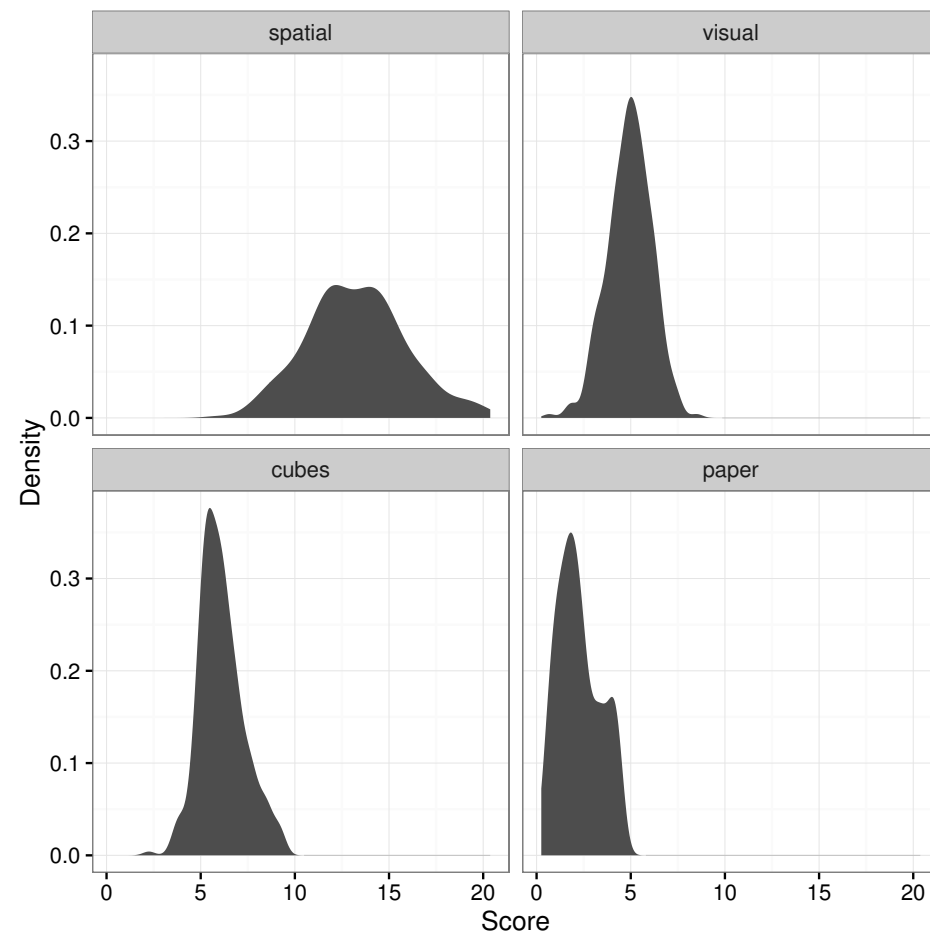
visual + cubes + paper = score (composite) spatial

Illustration

Calcul des scores composites :

```
HS$spatial <- rowSums(HS[,c("visual", "cubes", "paper")])
HS$verbal <- rowSums(HS[,c("paragrap", "sentence", "wordm")])
HS$speed <- rowSums(HS[,c("addition", "counting", "straight")])
colMeans(HS[,c("spatial", "verbal", "speed")])
```





Approche par ACP

Les composantes C_i ($i = 1, \dots, p$) de l'analyse en composantes principales (ACP) sont construites comme de simples combinaisons linéaires des p variables d'origine :

$$C_i = \sum_{j=1}^p w_{ij} x_j.$$

Les charges w_{ij} représentent la contribution de chaque variable dans la composante C_i . Les valeurs propres représentent la part de variance expliquée par chaque composante, et les vecteurs propres leur direction dans l'espace.

L'ACP constitue une approche préliminaire pour vérifier l'unidimensionnalité d'une échelle de mesure.

Application pour l'échelle spatiale

```
library(FactoMineR)
pca <- PCA(HS[,c("visual", "cubes", "paper")],
           scale.unit = TRUE, graph = FALSE)

> pca$eig
      eigenvalue percentage of variance cumulative percentage of variance
comp 1  1.7221050             57.40350             57.40350 ❶
comp 2  0.7227525             24.09175             81.49525
comp 3  0.5551425             18.50475            100.00000

> pca$var
$coord
      Dim.1      Dim.2      Dim.3
visual 0.7742336 -0.4082765 0.4836038 ❷
cubes  0.6951525  0.7115009 0.1026133
```

```
paper 0.7996439 -0.2232247 -0.5574409
```

```
$cor
```

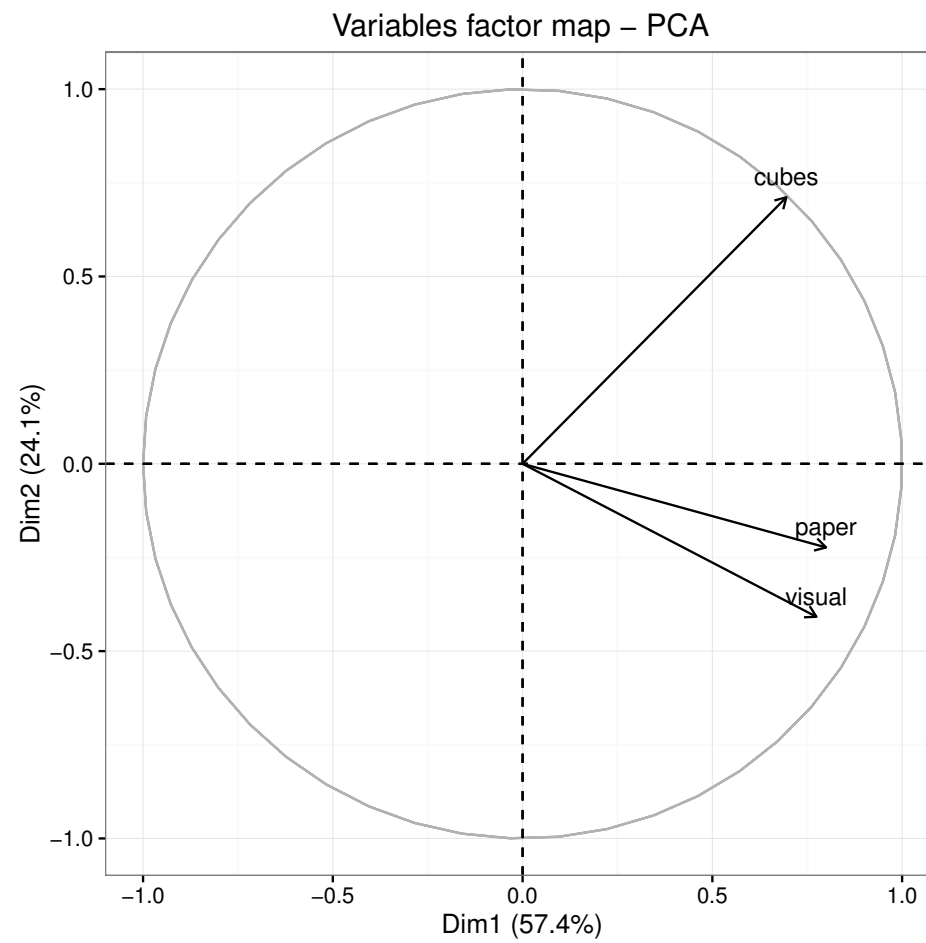
	Dim.1	Dim.2	Dim.3	
visual	0.7742336	-0.4082765	0.4836038	③
cubes	0.6951525	0.7115009	0.1026133	
paper	0.7996439	-0.2232247	-0.5574409	

```
$cos2
```

	Dim.1	Dim.2	Dim.3	
visual	0.5994377	0.16668969	0.2338726	④
cubes	0.4832369	0.50623355	0.0105295	
paper	0.6394303	0.04982925	0.3107404	

Packages R pour la visualisation des résultats : ggbiplot, factoextra.

Voir aussi : <http://gastonsanchez.com/how-to/2012/06/17/PCA-in-R/>.



Scores individuels

visual + cubes + paper = score (composite) spatial

$w_{11}\text{visual} + w_{12}\text{cubes} + w_{13}\text{paper} = \text{PC1} \approx \text{score (factoriel) spatial}$

```
HS$PC1 <- pca$ind$coord[,1]  
cor(HS[,c("spatial", "PC1")])
```

L'ACP peut être vue comme une méthode de réduction de dimension, de résumé graphique d'une matrice de corrélation, voire d'inférence multivariée (sous certaines hypothèses distributionnelles), mais dans tous les cas cette approche suppose que les variables X_j sont mesurées sans erreur.

Lorsque le nombre d'items est grand, l'ACP constitue une bonne approximation de l'analyse factorielle.

Application

1. Calculer la matrice de corrélation de Pearson et réaliser une ACP sur les données HolzingerSwineford1939 en considérant l'ensemble des variables (`visual`, . . . , `straight`).
2. Calculer les scores factoriels pour chaque dimension définie *a priori* (3 variables/dimension) et calculer leur corrélation avec les scores totaux non pondérés.
3. Calculer les corrélations entre les sous-scores totaux et le score total calculé en ignorant les dimensions (9 variables).
4. Comparer les sous-scores totaux et factoriels entre les deux groupes de participants définis par la variable `school`.

Fichier de données et scripts R disponibles à l'adresse suivante :
<https://bitbucket.org/chlallanne/eespe11>

– Typeset with Foil_{TEX} (version 2), Revision e967a78