# SEMINR

## PSYCHOMETRIE ET ANALYSE FACTORIELLE

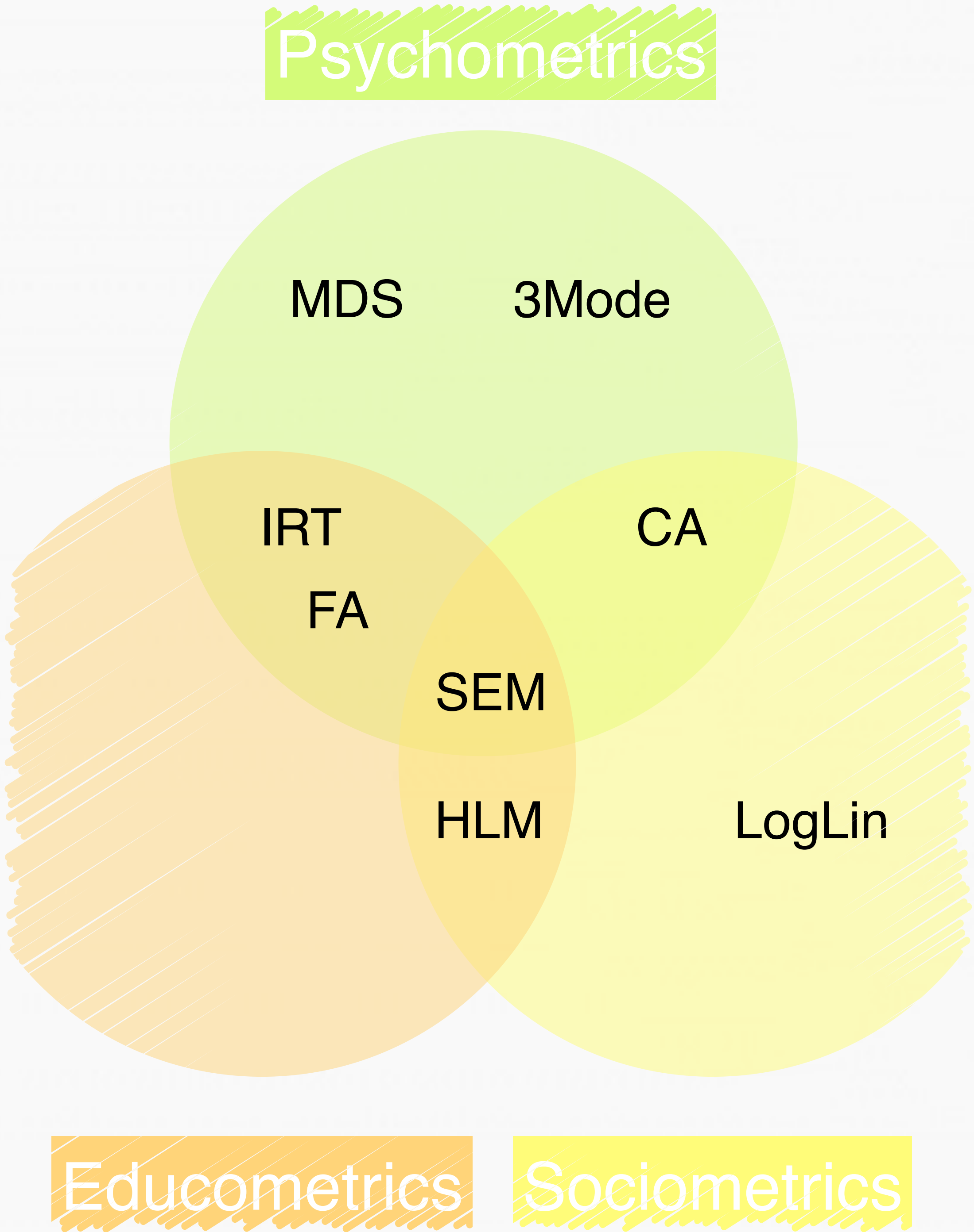
CHRISTOPHE LALANNE

www.aliquote.org

“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.” — Nunnally & Bernstein (1994)

# PSYCHOMETRIE

# FOOMETRICS




JAN DE LEEUW, USER! 2006

If Foo is a science then Foo often has both an area Foometrics and an area Mathematical Foo. Mathematical Foo applies mathematical modeling to the Foo subject area, while Foometrics develops and studies data analysis techniques for empirical data collected in Foo. Each of the social and behavioural sciences has a form of Foometrics, although they may not all use a name in this family.

*Multivariate Behavioral Research*, 1987, 22, 267–305

# A Brief History of the Philosophical Foundations of Exploratory Factor Analysis

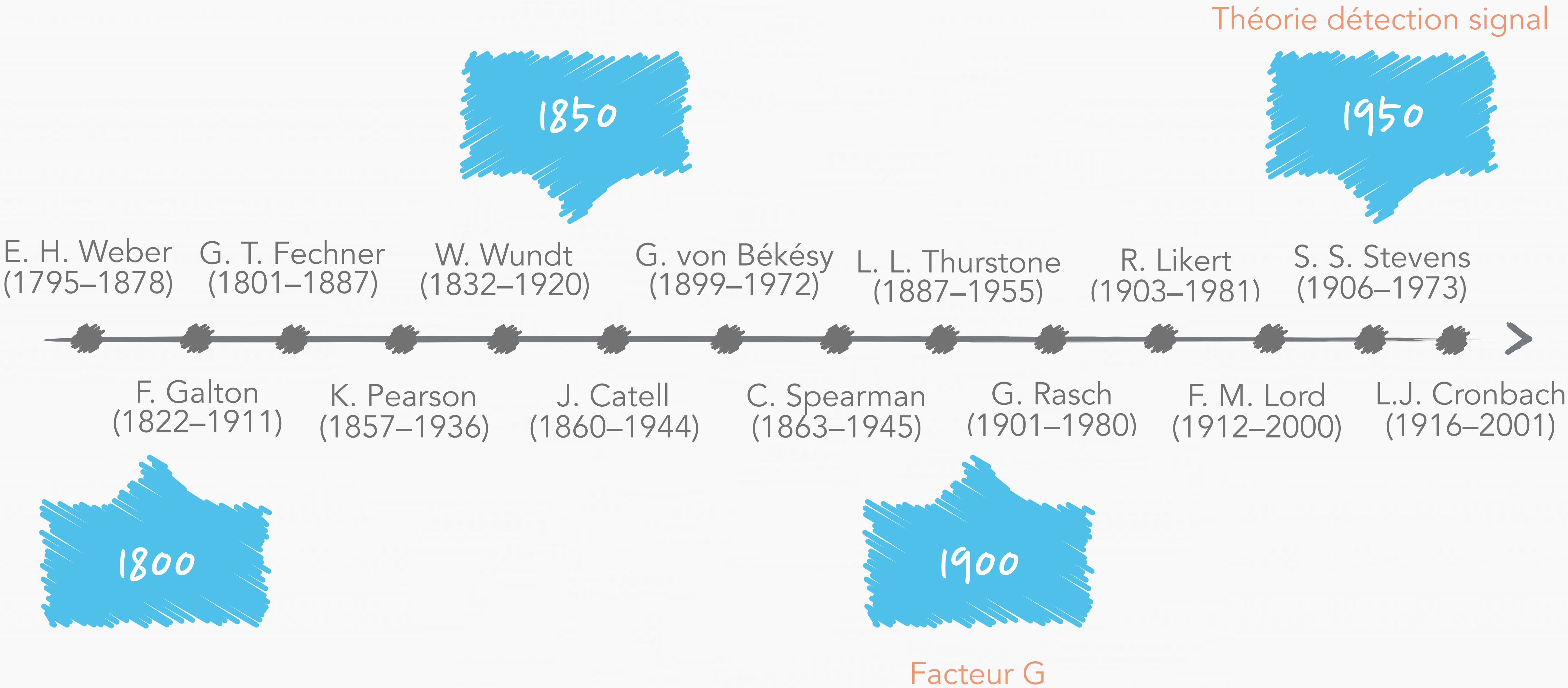
Stanley A. Mulaik

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.

# UN PEU D'HISTOIRE

(avant Jan de Leeuw et Bengt Muthen)




## APPROCHES PSYCHOPHYSIQUE ET PSYCHOMETRIQUE

# MESURE ET INSTRUMENT

- Un instrument (de mesure) est utilisé pour mettre en relation, ou « associer », quelque chose observé dans le monde réel à quelque chose conceptualisé dans un modèle théorique.
- Dans le premier cas on parle de variable manifeste, et dans le second de variable latente (ou facteur).
- Le processus de mesure peut être assimilé à une tâche consistant à assigner des réponses individuelles à des catégories ou à des nombres (Stevens, 1946 ; De Boeck et al., 2005), dans le but de

  (…) provide a reasonable and consistent way to summarize the responses that people make to express their achievements, attitudes, or personal points of view through instruments such as attitude scales, achievement tests, questionnaires, surveys, and psychological scales. (Wilson, 2005)

# MODELES A VARIABLES LATENTES

| Variable latente \ Variable manifeste | **Métrique** | **Catégorielle** |
|---|---|---|
| **Métrique** | Analyse factorielle | Analyse en traits latents |
| **Catégorielle** | Analyse en profiles latents | Analyse en classes latentes |

(Bartholomew & Knott, 2011 ; Rabe-Hesketh & Skrondal, 2008)

# DES ITEMS AUX ECHELLES

- L'agrégation des réponses à différents items aboutit à une échelle de mesure (parfois confondue avec la dimension qu'elle est censée refléter).
- En supposant qu'il n'y a qu'un seul « construct » (Borsboom, 2006), il est possible d'assigner des scores aux patterns de réponses observées ou aux scores totaux (théorie de réponse à l'item) afin d'inférer la position de chaque individu sur le trait latent (et plus généralement d'ordonner les individus sur ce même trait latent).
- Deux propriétés attendues : unidimensionnalité et indépendance locale.

  L'instrument doit mesurer un « construct » unique qui rend compte de la réponse d'un individu à un item spécifique étant donné sa position sur le trait latent (« habileté »), indépendamment d'autre facteurs.

  Si la variable latente est maintenue à un même niveau, les réponses à n'importe quel item doivent être indépendantes.

# EXEMPLES

## HADS

Deux échelles (2 x 7 items, 21 points max.) pour l'anxiété et la dépression (Zigmond & Snaith, 1983).

## NEOPI-R

Inventaire de personnalité (240/60 items) décrivant le modèle à 5 facteurs selon 6 facettes (extraversion, agreeableness, conscientiousness, neuroticism, openness) (McCrae & John, 1992)

## MOS-HIV

Questionnaire de qualité de vie spécifique du VIH (35 items) : general health perception, pain, physical functioning, role functioning, … (Wu et al., 1997)

## PROMIS

http://www.nihpromis.org

Banque d'items (adultes/enfants) pour plusieurs pathologies.

# MODELES DE MESURE

- Théorie classique des tests : suppose l'existence d'un score « vrai » défini comme la réponse attendue pour un individu à l'issue de l'administration infinie du même test (Borsboom, 2006).

  score observé = score vrai + erreur

  Principaux champs d'application : étude de fidélité de mesure et analyse d'items. Le score est généralement la somme des réponses de l'individu aux différents items et constitue une statistique de rang.

- Modèle(s) de réponse à l'item : modèle statistique selon lequel la probabilité qu'un individu choisisse une modalité de réponse dépend de paramètres propres aux items (« difficulté ») et aux individus (« habileté »). La difficulté de l'item (ou les seuils de transition dans le cas des items polytomiques) représente la position sur le trait latent qui maximise la discrimination entre les individus.

  Principaux champs d'application : banque d'items (Reeve et al., 2007), testing adaptatif (Rebollo et al., 2010), invariance de mesure (Teresi, 2006).

# ILLUSTRATION

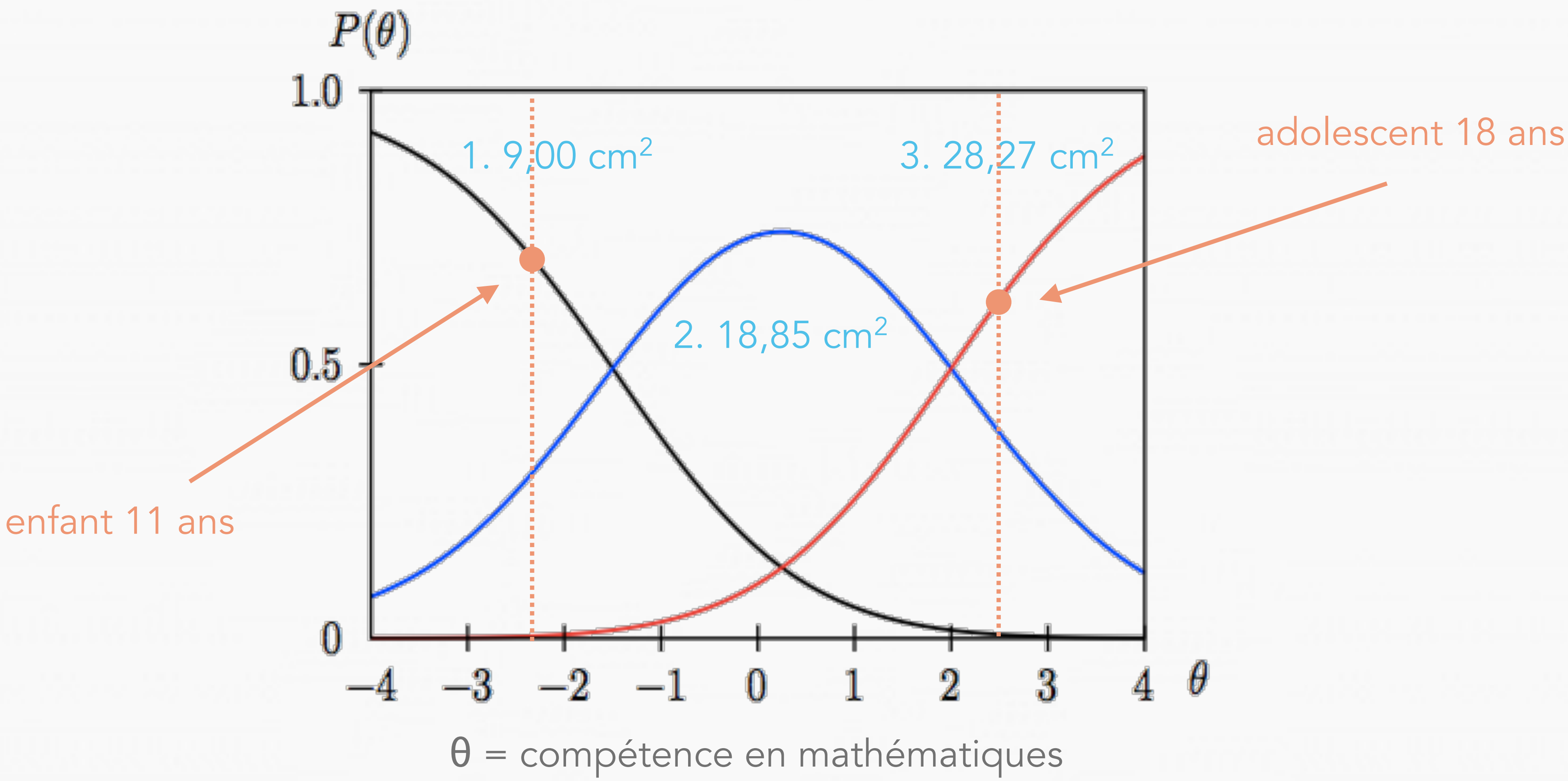
Partchev (2004)

Quelle est la surface d’un cercle ayant un rayon de 3 cm ?

1. 9,00 $cm^2$
2. 18,85 $cm^2$
3. 28,27 $cm^2$

1. réponse probablement la plus naïve
2. réponse erronée mais supposant un certain degré de connaissance (surface confondue avec circonférence)
3. réponse correcte

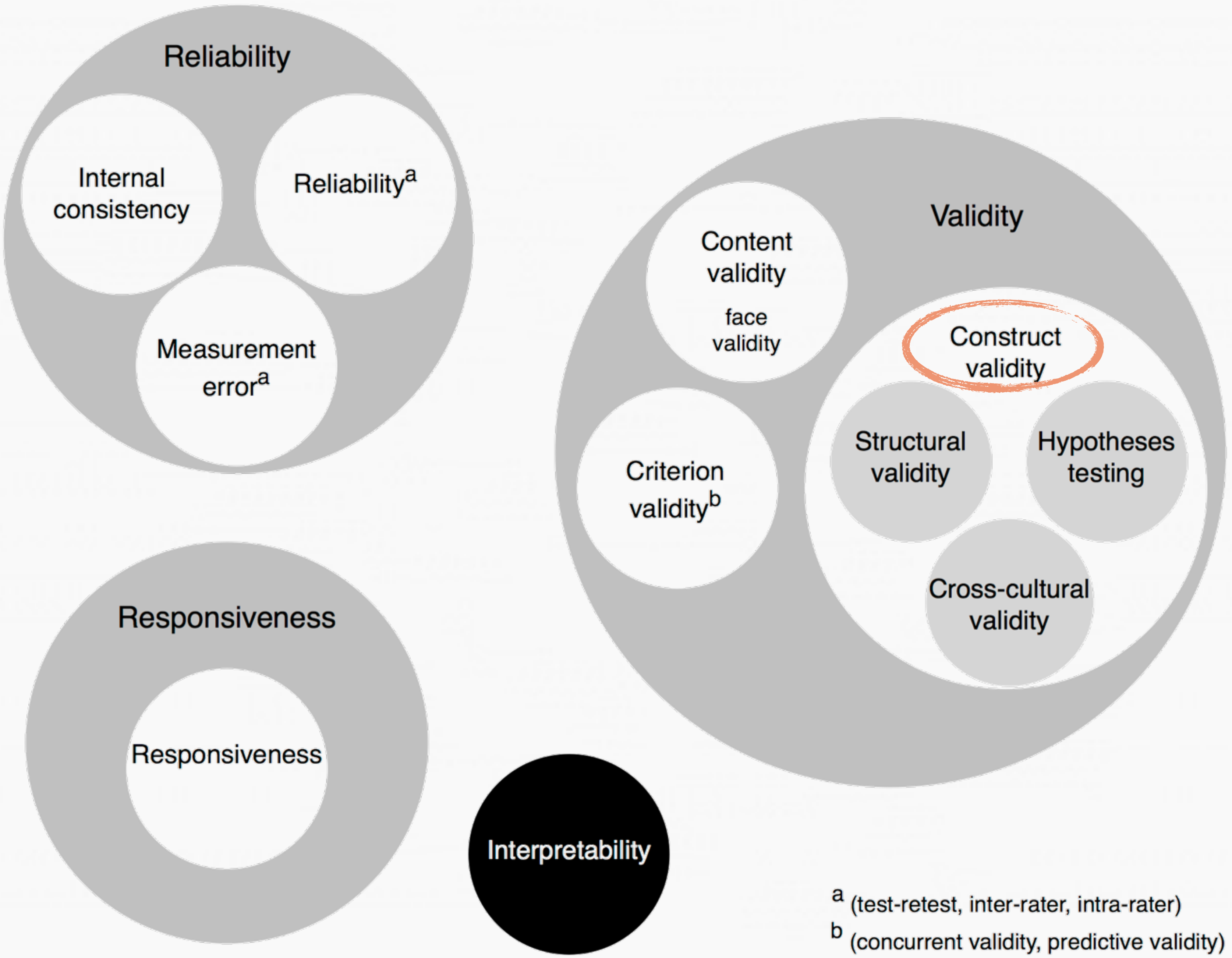
# ILLUSTRATION



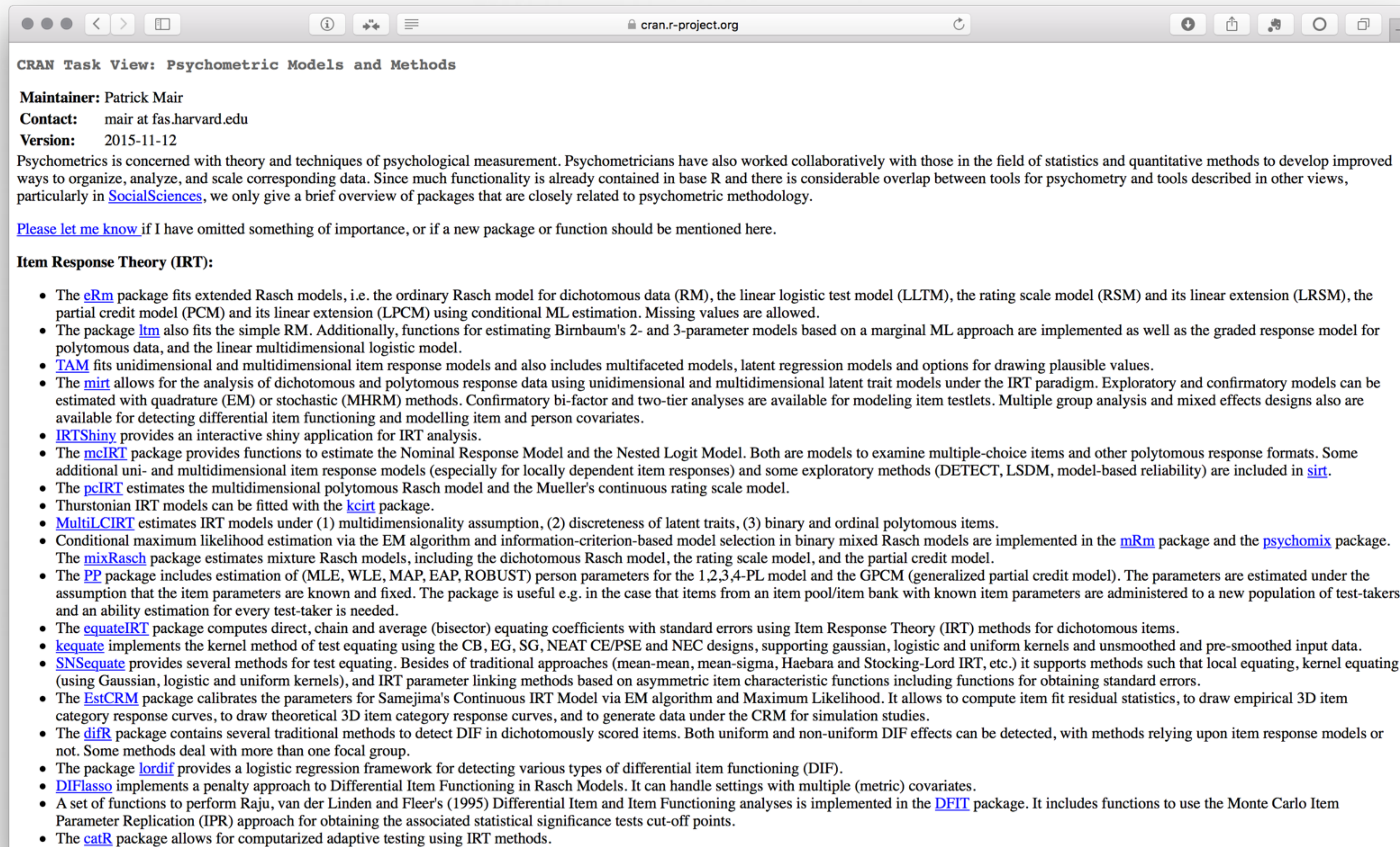

Modèle probabiliste (Partial Credit Model)

# COSMIN TAXONOMY

Mokkink et al. (2010)



# PACKAGES R



cran.r-project.org

CRAN Task View: Psychometric Models and Methods

**Maintainer:** Patrick Mair
**Contact:** mair at fas.harvard.edu
**Version:** 2015-11-12

Psychometrics is concerned with theory and techniques of psychological measurement. Psychometricians have also worked collaboratively with those in the field of statistics and quantitative methods to develop improved ways to organize, analyze, and scale corresponding data. Since much functionality is already contained in base R and there is considerable overlap between tools for psychometry and tools described in other views, particularly in SocialSciences, we only give a brief overview of packages that are closely related to psychometric methodology.

Please let me know if I have omitted something of importance, or if a new package or function should be mentioned here.

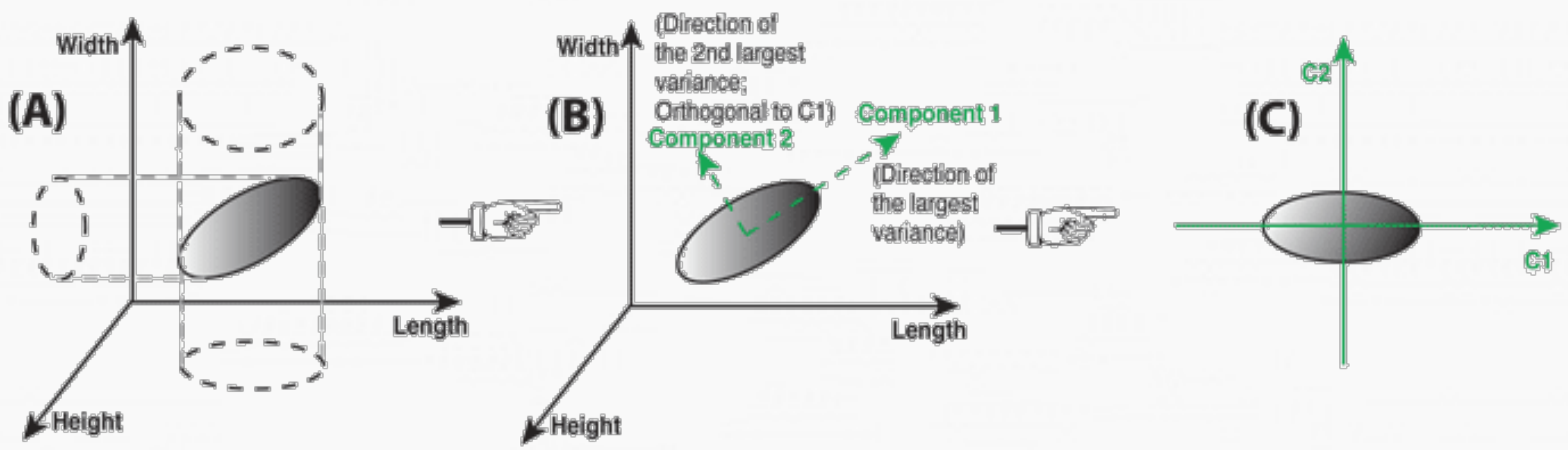
**Item Response Theory (IRT):**

- The eRm package fits extended Rasch models, i.e. the ordinary Rasch model for dichotomous data (RM), the linear logistic test model (LLTM), the rating scale model (RSM) and its linear extension (LRSM), the partial credit model (PCM) and its linear extension (LPCM) using conditional ML estimation. Missing values are allowed.
- The package ltm also fits the simple RM. Additionally, functions for estimating Birnbaum's 2- and 3-parameter models based on a marginal ML approach are implemented as well as the graded response model for polytomous data, and the linear multidimensional logistic model.
- TAM fits unidimensional and multidimensional item response models and also includes multifaceted models, latent regression models and options for drawing plausible values.
- The mirt allows for the analysis of dichotomous and polytomous response data using unidimensional and multidimensional latent trait models under the IRT paradigm. Exploratory and confirmatory models can be estimated with quadrature (EM) or stochastic (MHRM) methods. Confirmatory bi-factor and two-tier analyses are available for modeling item testlets. Multiple group analysis and mixed effects designs also are available for detecting differential item functioning and modelling item and person covariates.
- IRTShiny provides an interactive shiny application for IRT analysis.
- The mcIRT package provides functions to estimate the Nominal Response Model and the Nested Logit Model. Both are models to examine multiple-choice items and other polytomous response formats. Some additional uni- and multidimensional item response models (especially for locally dependent item responses) and some exploratory methods (DETECT, LSDM, model-based reliability) are included in sirt.
- The pcIRT estimates the multidimensional polytomous Rasch model and the Mueller's continuous rating scale model.
- Thurstonian IRT models can be fitted with the kcirt package.
- MultiLCIRT estimates IRT models under (1) multidimensionality assumption, (2) discreteness of latent traits, (3) binary and ordinal polytomous items.
- Conditional maximum likelihood estimation via the EM algorithm and information-criterion-based model selection in binary mixed Rasch models are implemented in the mRm package and the psychomix package. The mixRasch package estimates mixture Rasch models, including the dichotomous Rasch model, the rating scale model, and the partial credit model.
- The PP package includes estimation of (MLE, WLE, MAP, EAP, ROBUST) person parameters for the 1,2,3,4-PL model and the GPCM (generalized partial credit model). The parameters are estimated under the assumption that the item parameters are known and fixed. The package is useful e.g. in the case that items from an item pool/item bank with known item parameters are administered to a new population of test-takers and an ability estimation for every test-taker is needed.
- The equateIRT package computes direct, chain and average (bisector) equating coefficients with standard errors using Item Response Theory (IRT) methods for dichotomous items.
- kequate implements the kernel method of test equating using the CB, EG, SG, NEAT CE/PSE and NEC designs, supporting gaussian, logistic and uniform kernels and unsmoothed and pre-smoothed input data.
- SNSequate provides several methods for test equating. Besides of traditional approaches (mean-mean, mean-sigma, Haebara and Stocking-Lord IRT, etc.) it supports methods such that local equating, kernel equating (using Gaussian, logistic and uniform kernels), and IRT parameter linking methods based on asymmetric item characteristic functions including functions for obtaining standard errors.
- The EstCRM package calibrates the parameters for Samejima's Continuous IRT Model via EM algorithm and Maximum Likelihood. It allows to compute item fit residual statistics, to draw empirical 3D item category response curves, to draw theoretical 3D item category response curves, and to generate data under the CRM for simulation studies.
- The difR package contains several traditional methods to detect DIF in dichotomously scored items. Both uniform and non-uniform DIF effects can be detected, with methods relying upon item response models or not. Some methods deal with more than one focal group.
- The package lordif provides a logistic regression framework for detecting various types of differential item functioning (DIF).
- DIFlasso implements a penalty approach to Differential Item Functioning in Rasch Models. It can handle settings with multiple (metric) covariates.
- A set of functions to perform Raju, van der Linden and Fleer's (1995) Differential Item and Item Functioning analyses is implemented in the DFIT package. It includes functions to use the Monte Carlo Item Parameter Replication (IPR) approach for obtaining the associated statistical significance tests cut-off points.
- The catR package allows for computarized adaptive testing using IRT methods.

psych, lavaan, sem, OpenMX, semPLS, semPlot, semTools, …

AcP

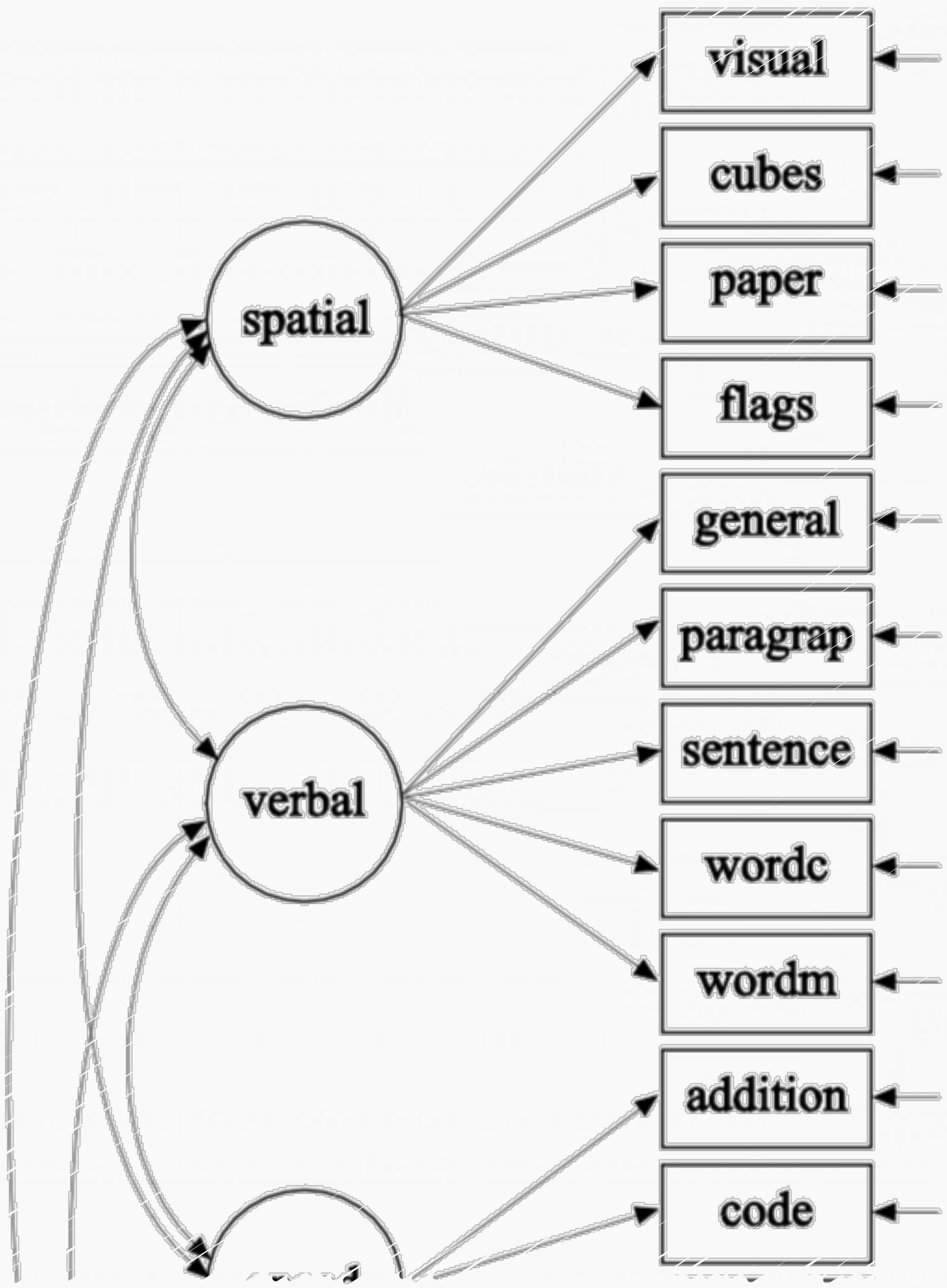
# L'ACP EN 6 MOTS




TROUVER LES DIRECTIONS DE VARIANCE MAXIMALE

http://mengnote.blogspot.fr/2013/05/an-intuitive-explanation-of-pca.html

# DONNEES D'ILLUSTRATION




## HOLZINGER & SWINEFORD

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.

```
lavaan::HolzingerSwineford1939
MBESS::HS.data
```

# LOAD

tests spatiaux

| | 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 |

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

# Score spatial

```
Console ~/Desktop/SEMinR/
> HS$spatial <- apply(HS[,c("visual", "cubes", "paper")], 1, sum)
> cor(HS[,c("spatial", "visual", "cubes", "paper")])
        spatial visual cubes paper
spatial   1.000  0.767 0.726 0.778
visual    0.767  1.000 0.297 0.441
cubes     0.726  0.297 1.000 0.340
paper     0.778  0.441 0.340 1.000
> apply(HS[,c("visual", "cubes", "paper")], 2, range)
     visual cubes paper
[1,]  0.667  2.25  0.25
[2,]  8.500  9.25  4.50
>
```

apply(data, margin, function)

score « spatial » = somme non pondérée des 3 subtests (de variance et d'étendue différentes)

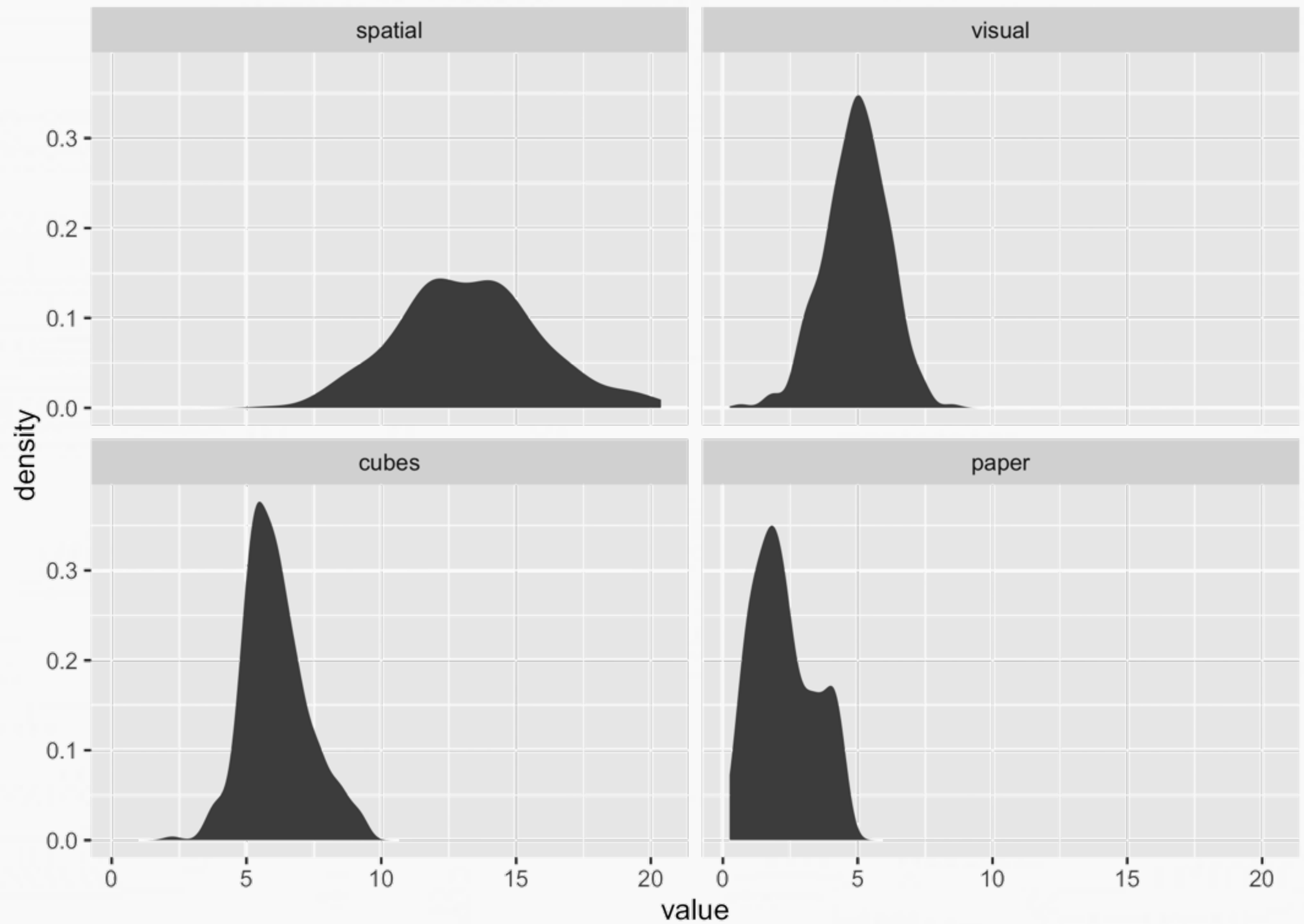
# Score spatial

Console ~/Desktop/SEMinR/

```
> d <- HS[,c("spatial", "visual", "cubes", "paper")]
> head(d)
  spatial visual cubes paper
1    11.5   3.33  7.75 0.375
2    12.7   5.33  5.25 2.125
3    11.6   4.50  5.25 1.875
4    16.1   5.33  7.75 3.000
5    10.5   4.83  4.75 0.875
6    12.6   5.33  5.00 2.250
> library(reshape2)
> d <- melt(HS[,c("id", "spatial", "visual", "cubes", "paper")])
Using id as id variables
> head(d)
  id variable value
1  1  spatial  11.5
2  2  spatial  12.7
```

melt (wide → long) ⇆ cast (long → wide)

# Score spatial




```r
p <- ggplot(data = d, aes(x = value))
p + geom_density(colour = "transparent", fill = "grey30") + facet_wrap(~ variable, nrow=2)
```

# PRINCIPE DE L'ACP

| visual | cubes | paper |
|---|---|---|
| $w_{11}$ x 3.333 | $w_{12}$ x 7.75 | $w_{13}$ x 0.375 |
| 5.333 | 5.25 | 2.125 |
| 4.500 | 5.25 | 1.875 |
| 5.333 | 7.75 | 3.000 |
| 4.833 | 4.75 | 0.875 |
| 5.333 | 5.00 | 2.250 |
| 2.833 | 6.00 | 1.000 |
| 5.667 | 6.25 | 1.875 |
| 4.500 | 5.75 | 1.500 |

## CALCUL DES COMPOSANTES

$$C_i = \sum w_{ij} x_j$$

Les composantes sont construites comme de simples combinaisons linéaires des variables d'origine.

Les charges ($w_{ij}$) représentent la contribution de chaque variable dans la composante $C_i$ (i=1,…, 3).

Les valeurs propres représentent la part de variance expliquée par chaque composante (et les vecteurs propres leur direction dans l'espace).

# FACTOMINER::PCA

Console ~/Desktop/SEMinR/

```
> pca$eig
       eigenvalue percentage of variance cumulative percentage of variance
comp 1      1.722                   57.4                              57.4
comp 2      0.723                   24.1                              81.5
comp 3      0.555                   18.5                             100.0
> pca$var$coord
       Dim.1  Dim.2  Dim.3
visual 0.774 -0.408  0.484
cubes  0.695  0.712  0.103
paper  0.800 -0.223 -0.557
> head(pca$ind$coord)
   Dim.1  Dim.2   Dim.3
1 -1.074  2.280  0.5450
2 -0.244 -0.731  0.2063
3 -0.801 -0.330 -0.0921
4  1.355  0.845 -0.0806
```

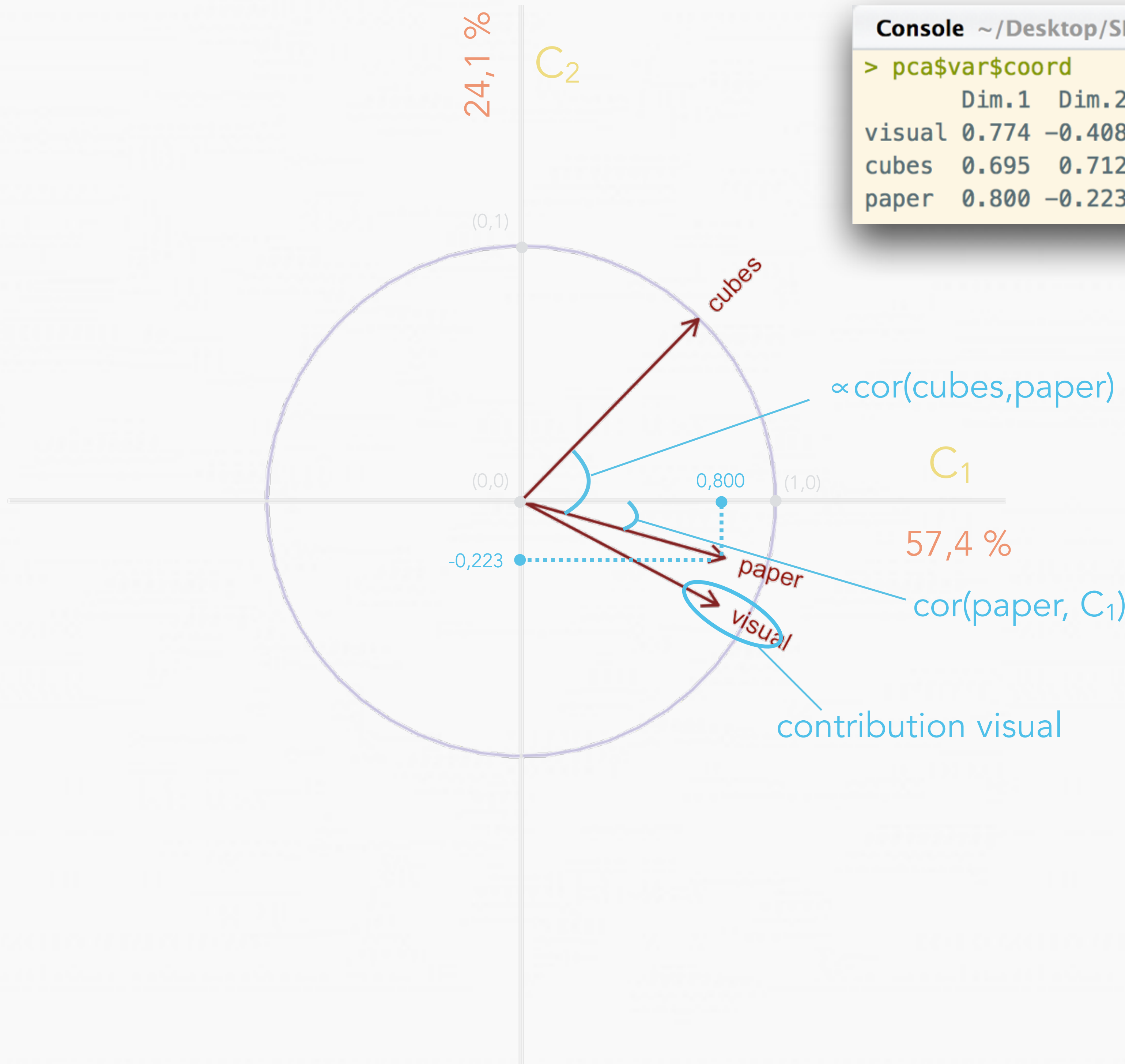
1.722 / (1.722+0.723+0.555)

$w_{ij}$

$C_i$

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

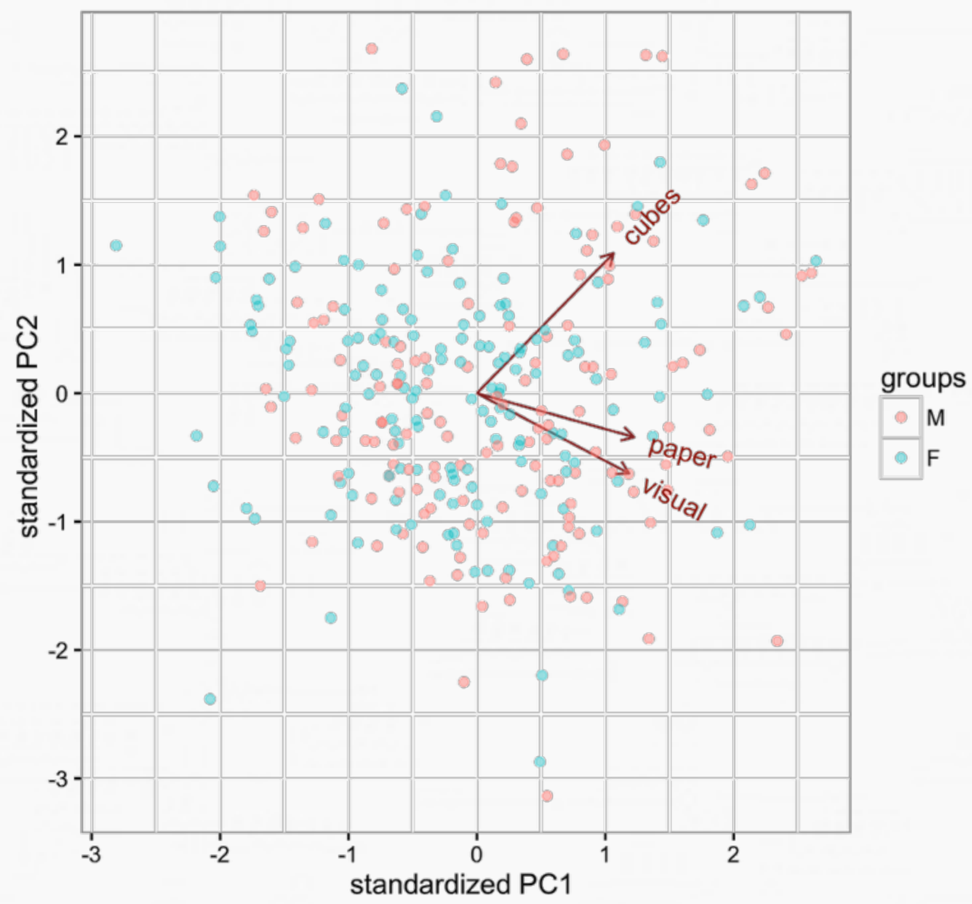
# CERCLE DES CORRELATIONS




Console ~/Desktop/SEMinR/

```
> pca$var$coord
        Dim.1  Dim.2  Dim.3
visual  0.774 -0.408  0.484
cubes   0.695  0.712  0.103
paper   0.800 -0.223 -0.557
```

# BIPLOT

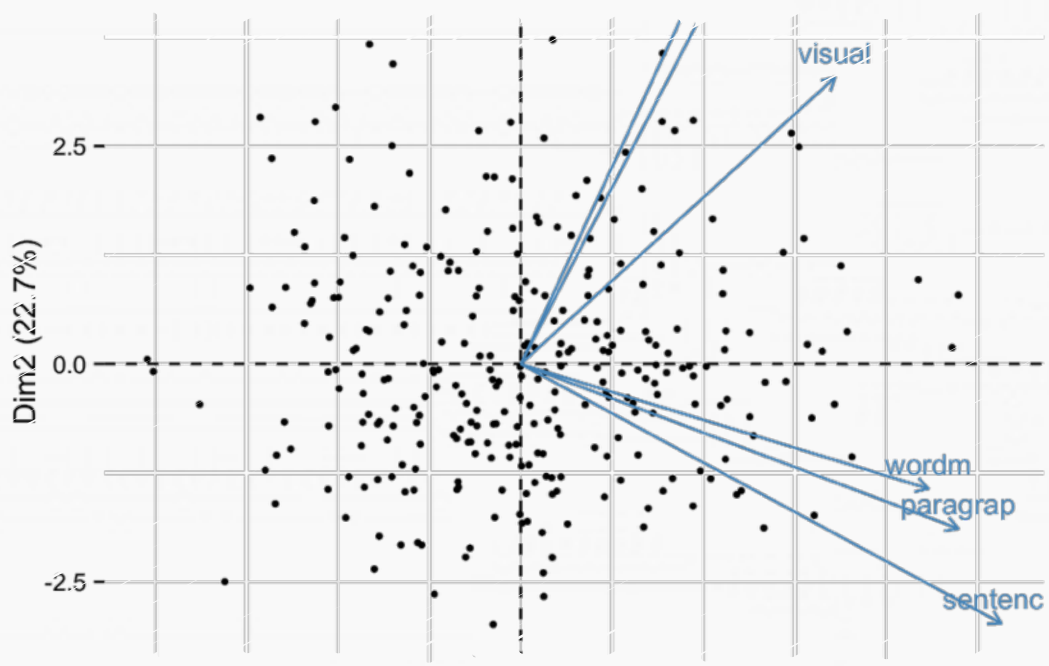



```
library(ggbiplot)
ggbiplot(pca, groups = HS$sex, varname.size = 4, alpha = .5) + theme_bw()
```

`devtools::install_github("vqv/ggbiplot")`

EFA

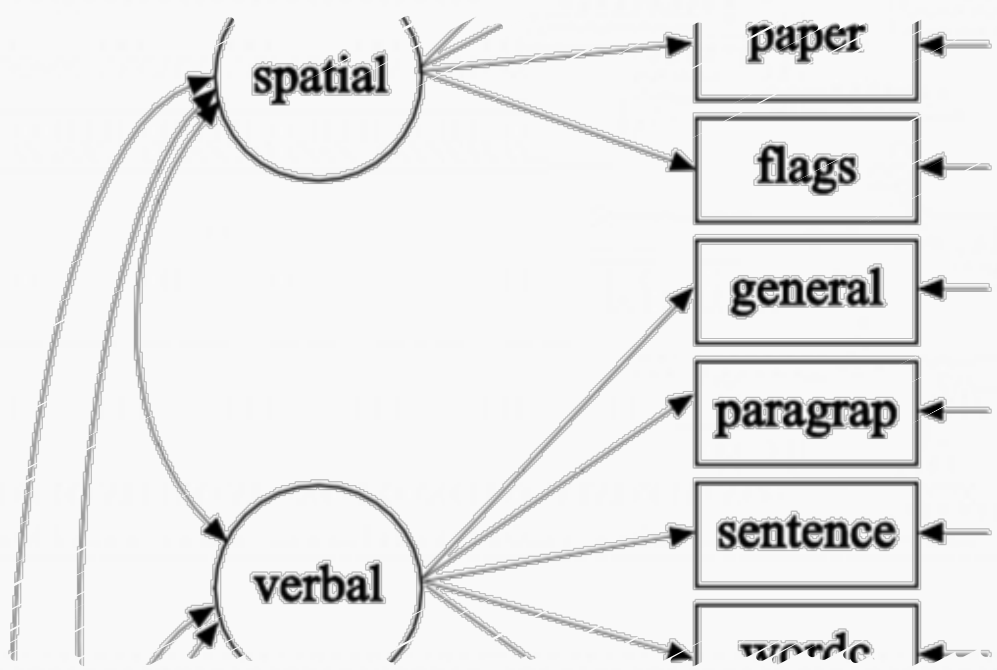
# ACP VERSUS EFA




## ACP

L'ACP peut être vue comme une méthode de réduction de dimension, de résumé graphique d'une matrice de corrélation, voire d'inférence multivariée supposant l'absence d'erreur de mesure.

$$C_i = \sum w_{ij} x_j$$




## EFA

L'EFA vise à explorer la relation entre variables manifestes en supposant que leur correlation reflète l'existence d'une ou plusieurs variables latentes (facteurs) fixées.

$$x_i \approx \sum w_{ij} F_j$$

# LE PACKAGE PSYCH

Console ~/Desktop/SEMinR/

```
> library(psych)
> principal(HS[,c("visual", "cubes", "paper")], nfactors = 3, rotate = "none")
Principal Components Analysis
Call: principal(r = 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
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
Cumulative Var        0.57 0.81 1.00
Proportion Explained  0.57 0.24 0.19
Cumulative Proportion 0.57 0.81 1.00
```

MODELE POUR LES VARIABLES MANIFESTES

# PSYCH::FA

Console ~/Desktop/SEMinR/

```
> fa(HS[,c("visual", "cubes", "paper")], nfactors = 1)
Factor Analysis using method =  minres
Call: fa(r = 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
paper  0.71 0.50 0.50   1

                MR1
SS loadings    1.12
Proportion Var 0.37

Mean item complexity =  1
Test of the hypothesis that 1 factor is sufficient.

The degrees of freedom for the null model are  3  and the objective function was  0.37 w
ith Chi Square of  110
```

uniquenesses $1-\sum w^2_{ij}$

communality $\sum w^2_{ij}$

minres
wls
gls
pa
ml
minchi

MODELE POUR LES VARIABLES LATENTES

# HELP(FA)

## Exploratory Factor analysis using MinRes (minimum residual) as well as EFA by Principal Axis, Weighted Least Squares or Maximum Likelihood

### Description

Among the many ways to do latent variable exploratory factor analysis (EFA), one of the better is to use Ordinary Least Squares (OLS) to find the minimum residual (minres) solution. This produces solutions very similar to maximum likelihood even for badly behaved matrices. A variation on minres is to do weighted least squares (WLS). Perhaps the most conventional technique is principal axes (PAF). An eigen value decomposition of a correlation matrix is done and then the communalities for each variable are estimated by the first n factors. These communalities are entered onto the diagonal and the procedure is repeated until the sum(diag(r)) does not vary. Yet another estimate procedure is maximum likelihood. For well behaved matrices, maximum likelihood factor analysis (either in the fa or in the factanal function) is probably preferred. Bootstrapped confidence intervals of the loadings and interfactor correlations are found by fa with n.iter > 1.

### Usage

```
fa(r,nfactors=1,n.obs = NA,n.iter=1, rotate="oblimin", scores="regression",
residuals=FALSE, SMC=TRUE, covar=FALSE,missing=FALSE,impute="median",
min.err = 0.001,  max.iter = 50,symmetric=TRUE, warnings=TRUE, fm="minres",
 alpha=.1,p=.05,oblique.scores=FALSE,np.obs,use="pairwise",cor="cor",...)
```

http://www.personality-project.org/r/book/ (Chapitre 6)

# CHOIX D'UN MODELE

- Choix des variables à inclure dans le modèle ? Analyse d'items ou hypothèses *a priori*.
- Choix du nombre de facteurs ? Méthode exploratoire (ACP, corrélation) ou hypothèses *a priori* ; comparaison *a posteriori* des différentes solutions (1, 2, 3, . facteurs)…
- Choix du type de rotation ? Dépend la plupart du temps du modèle postulé.
- Choix de la méthode d'estimation ? Maximum de vraisemblance suppose une distribution normale, peu adapté au cas des petits échantillons ; WLS et PA classiquement utilisés.
- Quelle matrice de corrélation analyser ? Corrélations de Pearson dans le cas des variables continues ou corrélations tétra- ou polychoriques dans le cas des items dicho- ou polytomiques.
- En pratique, le point essentiel reste le choix du nombre de facteurs et des variables à inclure dans le modèle.

# APPLICATION

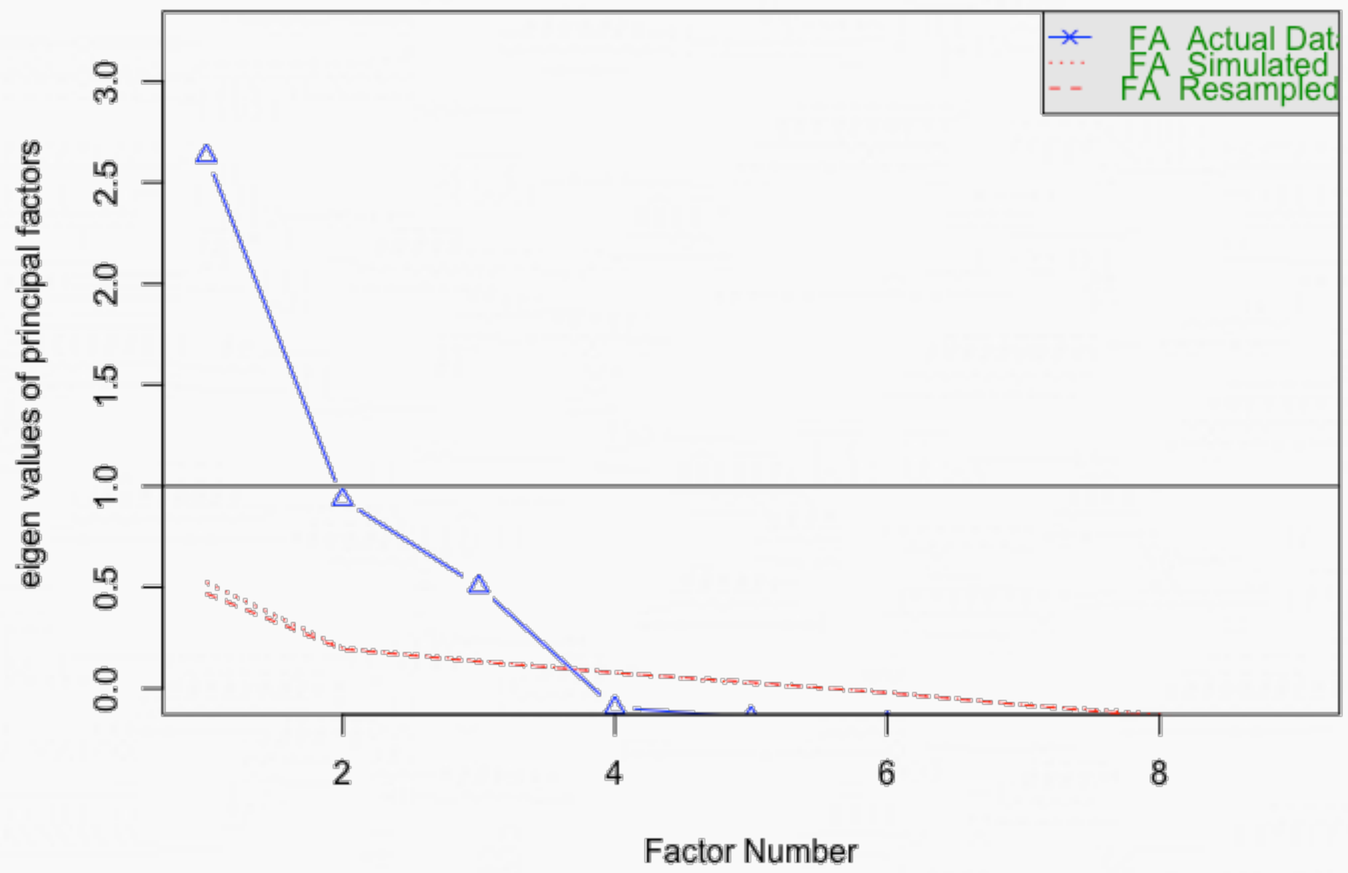
Console ~/Desktop/SEMinR/

```
> d <- HS[,7:15]
> names(d)
[1] "visual"   "cubes"    "paper"    "paragrap" "sentence" "wordm"    "addition" "counting"
[9] "straight"
> describe(d)
         vars   n mean   sd median trimmed  mad  min   max range  skew kurtosis   se
visual      1 301 4.94 1.17   5.00    4.96 1.24 0.67  8.50  7.83 -0.25     0.31 0.07
cubes       2 301 6.09 1.18   6.00    6.02 1.11 2.25  9.25  7.00  0.47     0.33 0.07
paper       3 301 2.25 1.13   2.12    2.20 1.30 0.25  4.50  4.25  0.38    -0.91 0.07
paragrap    4 301 3.06 1.16   3.00    3.02 0.99 0.00  6.33  6.33  0.27     0.08 0.07
sentence    5 301 4.34 1.29   4.50    4.40 1.48 1.00  7.00  6.00 -0.35    -0.55 0.07
wordm       6 301 2.19 1.10   2.00    2.09 1.06 0.14  6.14  6.00  0.86     0.82 0.06
addition    7 301 4.19 1.09   4.09    4.16 1.10 1.30  7.43  6.13  0.25    -0.31 0.06
counting    8 301 5.53 1.01   5.50    5.49 0.96 3.05 10.00  6.95  0.53     1.17 0.06
straight    9 301 5.37 1.01   5.42    5.37 0.99 2.78  9.25  6.47  0.20     0.29 0.06
> 
```

```
d <- HS[,7:15]
names(d)

describe(d)
```

# APPLICATION




```
fa.parallel(d, fm = "pa", fa = "fa", main = "")
```

# Solution 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 |

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

# Solution 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 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")
```

# References


1. Nunnally, J.C. and Bernstein, I.H. *Psychometric Theory*. New York: McGraw-Hill, 1994.
2. Mulaik, S.A. A Brief History of the Philosophical Foundations of Exploratory Factor Analysis. *Multivariate Behavioural Research*, 22: 267–305, 1987.
3. Stevens, S. On the Theory of Scales of Measurement. *Science*, 103: 677–680, 1946.
4. De Boeck, P., Wilson, M., and Acton, G.S. A Conceptual and Psychometric Framework for Distinguishing Categories and Dimensions. *Psychological Review*, 112(1): 129-158, 2005.
5. Wilson, M. *Constructing Measures. An Item Response Modeling Approach*. Taylor & Francis Group, 2005.
6. Bartholomew, D. and Knott, M. *Latent Variable Models and Factor Analysis: A Unified Approach*. Wiley, 2011.
7. Rabe-Hesketh, S. and Skrondal, A. Classical Latent Variable Models for Medical Research. *Statistical methods in medical research*, 17: 5–32, 2008.
8. Borsboom, D. The Attack of the Psychometricians. *Psychometrika*, 71(3): 425–440, 2006.
9. Zigmond, A. and Snaith, R. The Hospital Anxiety and Depression Scale. *Acta Psychiatry Scandinavia*, 67(6): 361–370, 1983.
10. McCrae, R. and John, O. An Introduction to the Five-Factor Model and Its Applications. *Journal of Personality*, 60(2): 175–215, 1992.

# References

11. Wu, A., Hays, R., Kelly, S., Malitz, F. and Bozzette, S. Applications of the Medical Outcomes Study Health-Related Quality of Life Measures in HIV/AIDS. *Quality of Life Research*, 6(6): 531-554, 1997.

12. Reeve, B., Hays, R., Bjorner, J., Cook, K. and Crane, P. et al. Psychometric Evaluation and Calibration of Health-Related Quality of Life Item Banks. *Medical Care*, 45(5): S22–S31, 2007.

13. Rebollo, P., Castejón, I., Cuervo, J., Villa, G. and García-Cueto, E. Validation of a Computer-Adaptive Test to Evaluate Generic Health-Related Quality of Life. *Health and Quality of Life Outcomes*, 8: 147, 2010.

14. Teresi, J. Different Approaches to Differential Item Functioning in Health Applications: Advantages, Disadvantages and Some Neglected Topics. *Medical Care*, 44, S152–S170, 2006.

15. Partchev, I. *A visual guide to item response theory*. Friedrich-Schiller-Universität Jena, 2004.

16. Mokkink, L., Terwee, C., Patrick, D., Alonso, J. and Stratford, P. et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measure- ment properties for health-related patient-reported outcomes. *Journal of Clinical Epidemiology*, 63: 737–745, 2010.