Exemples d'analyse en composantes principales

1.1.1 Mini-exemple

Ci-dessous, un tableau de notes attribuées à 9 sujets dans 5 matières.

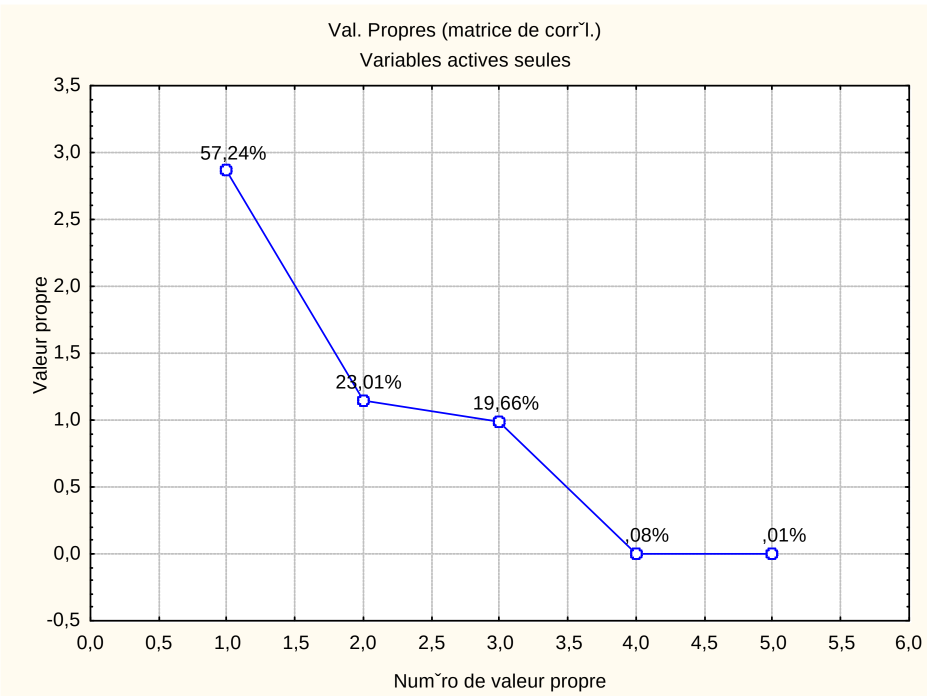
Sujet	Math	Sciences	Français	Latin	Musique
Jean	6	6	5	5,5	8
Aline	8	8	8	8	9
Annie	6	7	11	9,5	11
Monique	14,5	14,5	15,5	15	8
Didier	14	14	12	12	10
André	11	10	5,5	7	13
Pierre	5,5	7	14	11,5	10
Brigitte	13	12,5	8,5	9,5	12
Evelyne	9	9,5	12,5	12	18

L'ACP étudie les lignes et les colonnes de la matrice centrée-réduite :

Sujet	Math	Sciences	Français	Latin	Musique
Jean	-1,0865	-1,2817	-1,5037	-1,6252	-1,0190
Aline	-0,4939	-0,6130	-0,6399	-0,7223	-0,6794
Annie	-1,0865	-0,9474	0,2239	-0,1806	0,0000
Monique	1,4322	1,5604	1,5197	1,8058	-1,0190
Didier	1,2840	1,3932	0,5119	0,7223	-0,3397
André	0,3951	0,0557	-1,3597	-1,0835	0,6794
Pierre	-1,2347	-0,9474	1,0878	0,5417	-0,3397
Brigitte	0,9877	0,8916	-0,4959	-0,1806	0,3397
Evelyne	-0,1975	-0,1115	0,6559	0,7223	2,3778

1.2 Valeurs propres et inerties

	Val. propr	Variance (%)	Variance cumul (%)
1	2,8618	57,24	57,24
2	1,1507	23,01	80,25
3	0,9831	19,66	99,91
4	0,0039	0,08	99,99
5	0,0004	0,01	100,00



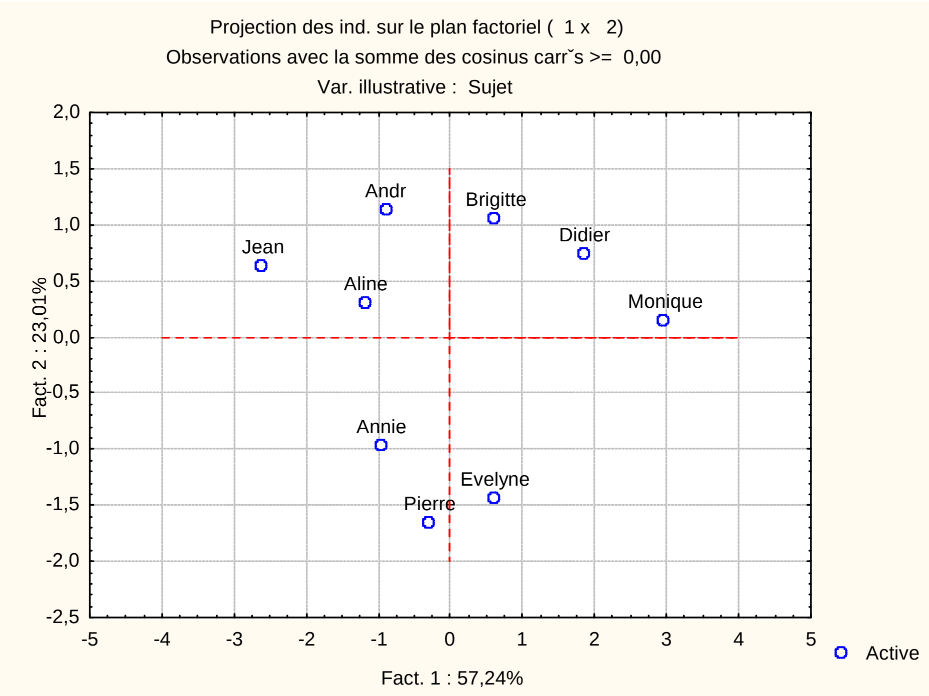
La variation totale (100%) est répartie selon 5 valeurs propres. D'où l'idée de ne garder que les valeurs propres (et directions propres) qui représentent au moins 20% de variation. Dans le cas d'une ACP normée, cela revient à conserver les valeurs propres supérieures à 1.
Variante : on observe une brusque décroissance des valeurs propres entre la 3è et la 4è valeur propre.
Au final, on décide de ne garder que trois valeurs propres.

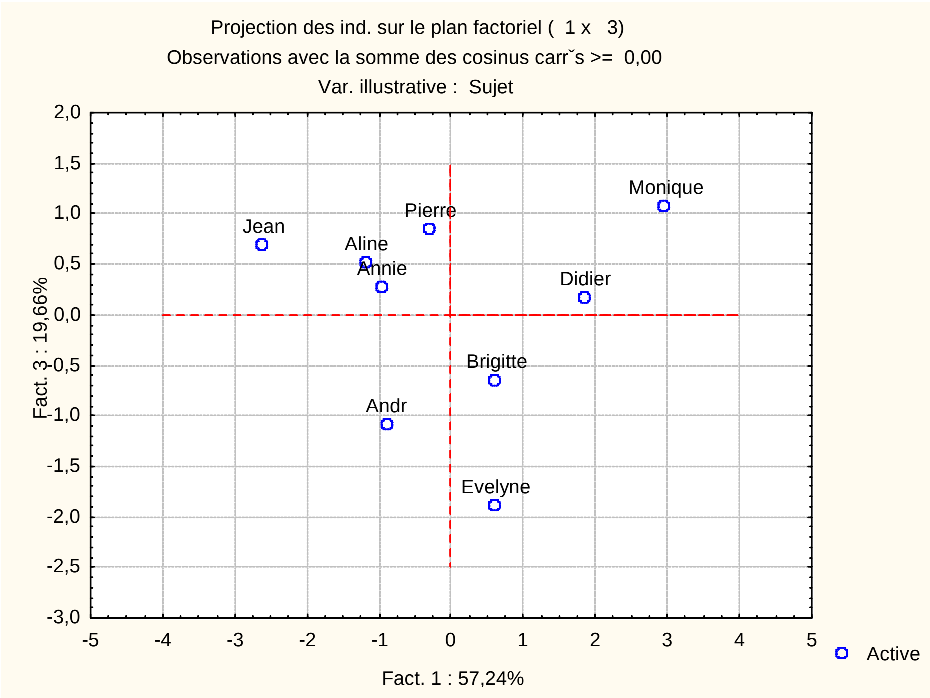
1.3 Résultats relatifs aux individus

1.3.1 Scores des individus

Les scores des individus sont les valeurs des composantes principales sur les individus :

	Fact. 1	Fact. 2	Fact. 3
Jean	-2,7857	0,6765	0,7368
Aline	-1,2625	0,3303	0,5549
Annie	-1,0167	-1,0198	0,2881
Monique	3,1222	0,1659	1,1442
Didier	1,9551	0,7879	0,1892
André	-0,9477	1,2014	-1,1401
Pierre	-0,3250	-1,7548	0,9095
Brigitte	0,6374	1,1298	-0,6919
Evelyne	0,6231	-1,5173	-1,9909





Contributions des individus

La contribution relative d'un individu i à la formation de la composante principale α est l'inertie relative de cet individu sur l 'axe factoriel k. Elle est définie par :

$$CTR_\alpha(i) = \frac{(Score\ de\ i\ sur\ l'axe\ \alpha)^2}{n\ \lambda_\alpha}$$

Par exemple : $CTR_1(Jean) = \dfrac{(-2,7857)^2}{9 \times 2,8618} = 0,3013$

Contributions des individus exprimées en pourcentages

Sujet	Fact. 1	Fact. 2	Fact. 3
Jean	30,13	4,42	6,14
Aline	6,19	1,05	3,48
Annie	4,01	10,04	0,94
Monique	37,85	0,27	14,80
Didier	14,84	5,99	0,40
André	3,49	13,94	14,69
Pierre	0,41	29,73	9,35
Brigitte	1,58	12,33	5,41
Evelyne	1,51	22,23	44,79

Qualités de la représentation des individus

La qualité de la représentation d'un individu i par la composante principale α est définie par :

$$QLT_\alpha(i) = \frac{(Score\ de\ i\ \ sur\ \ l'axe\ \alpha)^2}{\sum_l (Score\ de\ i\ sur\ l'axe\ \ l)^2}$$

Par exemple :

$$QLT_1(Jean) = \frac{(-2,7857)^2}{2,7857^2 + 0,6765^2 + ... + 0,0332^2} = \frac{(-2,7857)^2}{1,0865^2 + 1,2817^2 + 1,5037^2 + 1,6252^2 + 1,0190^2} = 0,8855$$

Cosinus carré :

Sujet	Fact. 1	Fact. 2	Fact. 3
Jean	0,8855	0,0522	0,0619
Aline	0,7920	0,0542	0,1530
Annie	0,4784	0,4813	0,0384
Monique	0,8786	0,0025	0,1180
Didier	0,8515	0,1383	0,0080
André	0,2465	0,3962	0,3568
Pierre	0,0263	0,7671	0,2061
Brigitte	0,1877	0,5898	0,2211
Evelyne	0,0583	0,3458	0,5954

Les qualités de représentation sont additives. Par exemple, la qualité de représentation d'un individu i par le plan 1-2 est donnée par :

$$QLT_{1,2}(i) = \frac{(Score\ de\ i\ sur\ l'axe 1)^2 + (Score\ de\ i\ sur\ l'axe\ 2)^2}{\sum_l (Score\ de\ i\ sur\ l'axe\ l)^2}$$

Pour le sujet 1 (Jean), la qualité de représentation par le plan factoriel 1-2 est : 0,8855+0,0522=0,9377.

1.4 Résultats relatifs aux variables

Saturation des variables

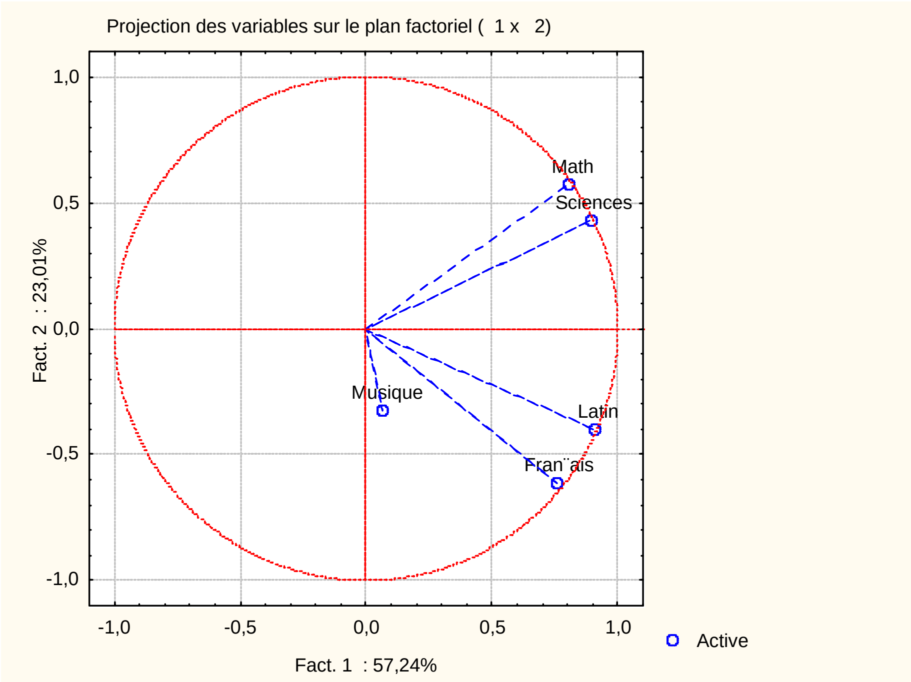
Les saturations des variables sont les coordonnées factorielles des variables. Elles sont égales au coefficients de corrélation entre les variables (centrées réduites) de départ et les scores des individus : $\phi_{j\alpha} = \rho(j, \psi_\alpha)$

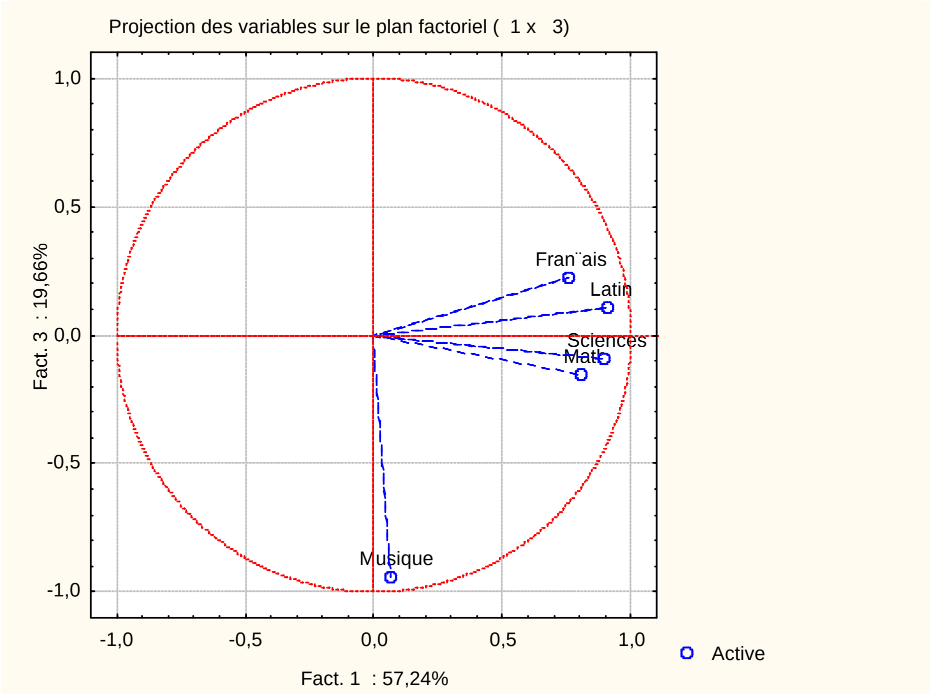
N.B. Les variables de départ sont centrées réduites, les scoress sont centrées, et de variances égales aux valeurs propres correspondantes. On peut donc retrouver les saturations à l'aide d'un calcul tel que :

$$\phi_{jean,1} = \frac{(-1{,}0865)(-2{,}7857) + (-0{,}4939)(-1{,}2625) + (-1{,}0865)(-1{,}0168) + (1{,}4322)(3{,}1222)}{9\sqrt{2{,}8618}} +$$
$$\frac{(1{,}2840)(1{,}9551) + (0{,}3951)(-0{,}9478) + (-1{,}2347)(-0{,}3250) + (0{,}9877)(0{,}6373) + (-0{,}1975)(0{,}6231)}{9\sqrt{2{,}8618}}$$

Coord. factorielles des variables :

	Fact. 1	Fact. 2	Fact. 3
Math	0,8059	0,5714	-0,1534
Sciences	0,8970	0,4308	-0,0929
Français	0,7581	-0,6110	0,2257
Latin	0,9103	-0,3975	0,1084
Musique	0,0667	-0,3275	-0,9425





Contributions des variables

Les contributions des variables à la formation des composantes principales sont définies de la même façon que celles des individus. Par exemple : $CTR_1(Math) = \dfrac{0{,}8059^2}{2{,}8618} = 0{,}2269$

Contributions des variables

	Fact. 1	Fact. 2	Fact. 3
Math	0,2269	0,2837	0,0239
Sciences	0,2812	0,1613	0,0088
Français	0,2008	0,3245	0,0518
Latin	0,2895	0,1373	0,0120
Musique	0,0016	0,0932	0,9035

Qualités de la représentation des variables

La qualité de la représentation d'une variable par une composante principale est définie de la même façon que pour les individus :

$$QLT_{\alpha}(j)=\frac{(Saturation\,de\ j\ sur\ l'\,axe\,\alpha)^2}{\sum_l (Saturation\,de\ j\,sur\ \ l'\,axe\,l)^2}=(Saturation\,de\ j\,sur\ l'\,axe\,\alpha)^2$$

Mais, comme les variables *j* sont normées, la qualité est simplement le carré de la saturation de la variable par rapport à la composante principale.

Comme dans le cas des individus, les qualités des représentations d'une variable selon les composantes principales s'additionnent. Le tableau ci-dessous donne les qualités de représentation selon la première composante principale, selon le plan des deux premières composantes et dans l'espace défini par les trois premières composantes.

Communautés des variables

	Avec 1 facteur	Avec 2 facteurs	Avec 3 facteurs
Math	0,6495	0,9759	0,9995
Sciences	0,8046	0,9902	0,9988
Français	0,5747	0,9481	0,9990
Latin	0,8286	0,9866	0,9983
Musique	0,0044	0,1117	1,0000

1.5 Interprétation conjointe des plans factoriels des individus et des variables

On utilise la formule : $\phi_{j\alpha}=\rho(j,\psi_{\alpha})$