

### 2.2.3. Illustration de la logique de calcul.

Si on note  $X_1, \dots, X_j, \dots, X_p$  les différentes variables (Statistiques, Mathématiques, Comptabilité et Gestion financière –  $p=4$ ) initiales plus ou moins corrélées entre elles, l'ACP va rechercher de nouvelles variables  $Y_1, \dots, Y_j, \dots, Y_p$  que l'on appelle **composantes principales** qui elles ne seront pas corrélées entre elles (chaque composante principale apportera donc une information différente des autres) et qui devront **résumer au mieux les données de départ**. Ces composantes principales sont en fait des indices synthétiques c'est-à-dire des combinaisons linéaires des variables initiales ( $X_1, \dots, X_j, \dots, X_p$ ).

Avant d'aller plus avant, il est nécessaire de définir ce que représente une combinaison linéaire de variables. Pour cela, imaginons un individu décrit par deux variables. Un exemple simple de combinaison linéaire serait tout simplement le résultat du calcul de sa moyenne :  $\bar{x} = \frac{x_1 + x_2}{2} = \frac{1}{2}x_1 + \frac{1}{2}x_2$ . On parle de combinaison car toutes les variables apparaissent dans le calcul et elle est linéaire car les variables ne sont pas multipliées par elles-mêmes ou élevées à une puissance quelconque.

Sachant que les individus sont décrits dans un espace à  $p$  variables (dimensions), comme nous l'avons précédemment précisé, le principe de l'ACP consiste à trouver des espaces de dimensions « plus petites » afin d'y voir au mieux les individus. Pour introduire la logique du calcul du premier axe, nous allons nous appuyer sur un exemple comptant trois individus décrits par deux variables (notes)  $X_1$  et  $X_2$  :

$$\underset{(3,2)}{X} = \begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ x_{31} & x_{32} \end{pmatrix} = \begin{pmatrix} 5 & 9 \\ 11 & 8 \\ 17 & 13 \end{pmatrix}$$

Figure n°1 : données brutes.

La moyenne de  $X_1$  vaut 11.00 et sa variance 24.00 tandis que la moyenne et la variance de  $X_2$  sont respectivement égales à 10.00 et 4.67. Si on ne réduit pas les deux variables, ce qui sera le cas ici, l'inertie totale est donc égale à  $I=24.00+4.67=28.67$ .

En revanche, et comme nous l'avons vu précédemment, nous commençons toujours dans une ACP par centrer nos données. Cette opération nous conduit à une nouvelle matrice de données dont les variances (et les covariances) restent identiques :

$$\underset{(3,2)}{X} = \begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ x_{31} & x_{32} \end{pmatrix} = \begin{pmatrix} -6 & -1 \\ 0 & -2 \\ 6 & 3 \end{pmatrix}$$

Figure n°2 : données centrées.

On va à partir de ces données créer une première variable,  $Y_1$ , combinaison linéaire des deux précédentes. Soient  $c_1$  et  $c_2$  les coefficients de la combinaison linéaire,  $Y_1$  devient égal à  $Y_1 = c_1X_1 + c_2X_2$ .

Si on pose  $c_1 = c_2 = \frac{1}{2}$ , on obtient  $Y_1 = \begin{pmatrix} -3.5 \\ -1.0 \\ 4.5 \end{pmatrix}$  c'est-à-dire que  $Y_1$  est un vecteur composé des moyennes

des notes centrées obtenues par chaque étudiant. Bien sûr, si l'on n'avait pas centré nos données, on aurait

obtenu  $Y_1 = \begin{pmatrix} 7.0 \\ 9.5 \\ 15.0 \end{pmatrix}$ .

Quoi qu'il en soit, remplacer deux notes par une seule (la moyenne des notes) relève bien du principe de l'ACP qui cherche à diminuer le nombre de variables (on passe de deux notes à une moyenne). Le problème consiste maintenant à savoir si le choix de la moyenne (c'est-à-dire dans le cadre de notre exemple utiliser une combinaison linéaire dont les coefficients sont  $c_1 = c_2 = \frac{1}{2}$ ) est le meilleur choix possible. Pour cela il nous faut définir un critère. L'un des objectifs de l'ACP est de voir au mieux les différences entre les individus. On verra d'autant mieux les différences entre les individus qu'ils seront, toutes choses égales par ailleurs, d'autant plus dispersés autour de leur centre de gravité.

En fonction de ce que nous avons vu précédemment, on sait que la dispersion autour du centre de gravité se mesure au travers de l'inertie qui se trouve également être égale à la somme des variances. Partant de là, un indicateur immédiat permettant de mesurer la dispersion des individus autour de leur moyenne se trouve être la variance. En clair, cela revient à calculer l'inertie des individus sur cette première nouvelle variable.

Ainsi avec  $c_1 = c_2 = \frac{1}{2}$ , on obtient  $Y_1 = \begin{pmatrix} -3.5 \\ -1.0 \\ 4.5 \end{pmatrix}$ ,  $\bar{Y}_1 = 0$  (la moyenne de variables centrées est

nulle) et  $Var(Y_1) = 11.17$ . En choisissant  $c_1 = \frac{1}{2}$  et  $c_2 = -\frac{1}{2}$ , on obtient  $Y_1 = \begin{pmatrix} -2.5 \\ 1.0 \\ 1.5 \end{pmatrix}$ ,  $\bar{Y}_1 = 0$  et

$Var(Y_1) = 3.17$ . Ces deux cas sont résumés graphiquement ci-dessous :

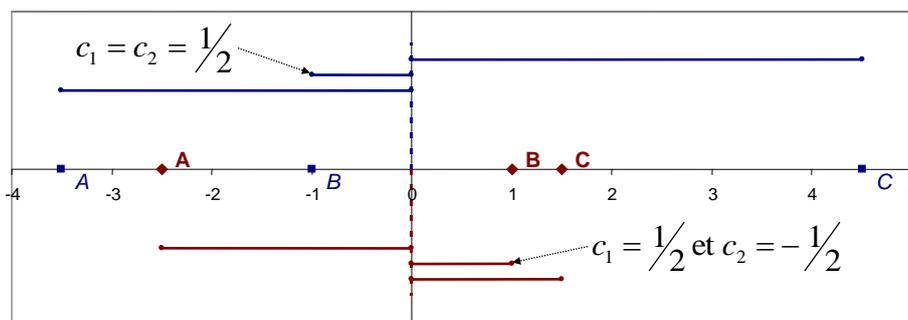


Figure n°3 :  $c_1 = c_2 = \frac{1}{2}$  vs.  $c_1 = \frac{1}{2}$  et  $c_2 = -\frac{1}{2}$ .

On constate que les différences entre individus sont mieux mises en valeur dans le premier cas que dans le deuxième. Maintenant, on peut bien sûr envisager une infinité de valeurs pour  $c_1$  et  $c_2$ . Par

exemple,  $c_1 = 20$  et  $c_2 = -20$  donnent  $Y_1 = \begin{pmatrix} -100 \\ 40 \\ 60 \end{pmatrix}$ ,  $\bar{Y}_1 = 0$  et  $Var(Y_1) = 5066.67$ . Ce choix de valeurs

pour  $c_1$  et  $c_2$  (comparé avec  $c_1 = c_2 = \frac{1}{2}$ ) est résumé sur le graphique ci-dessous :

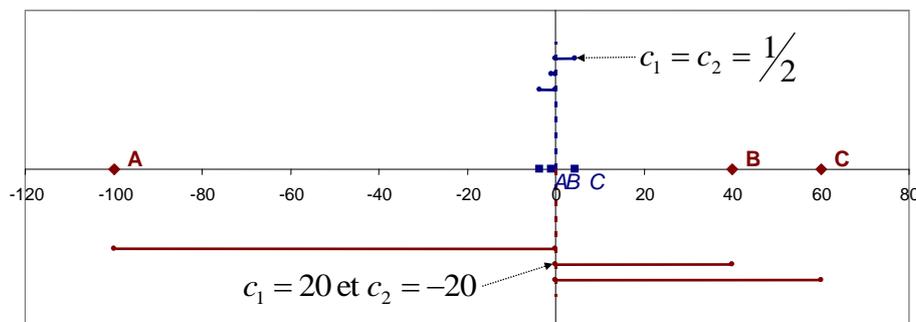


Figure n°4 :  $c_1 = c_2 = \frac{1}{2}$  vs  $c_1 = 20$  et  $c_2 = -20$ .

En comparant les trois cas (attention au changement d'échelle entre les deux graphiques), il est évident que la dispersion des individus est d'autant plus forte que l'un des coefficients tend vers plus l'infini et l'autre vers moins l'infini. Ainsi, en posant  $c_1 = +\infty$  et  $c_2 = -\infty$  (ou le contraire) on obtient  $Var(Y_1) = +\infty$ . Ce résultat, trivial, est évidemment sans intérêt.

Ce constat fait, il va falloir imposer une contrainte aux coefficients de la combinaison linéaire. Par exemple, on peut imposer que leur somme soit égale à une valeur quelconque comme par exemple  $c_1 + c_2 = 0$ . Ce choix n'est manifestement pas pertinent car  $c_1 (= +\infty) + c_2 (= -\infty) = 0$  ce qui ne résout pas notre problème. En fait, la solution retenue consiste à imposer que la somme des coefficients élevés au carré soit nulle (dans notre exemple,  $c_1^2 + c_2^2 = 1$ ). Dans les deux tableaux ci-dessous, on a pris quelques couples de valeurs pour  $c_1$  et  $c_2$  (les cinq premiers cas ne respectent pas la contrainte  $c_1^2 + c_2^2 = 1$ ) :

Cas	$c_1$	$c_2$	$Y_1$	$Var(Y_1)$
1	0.50	0.50	(-3.5, -1.0, 4.5)	11.17
2	0.50	-0.50	(-2.5, 1.0, 1.5)	3.17
3	1.58	-1.58	(-7.9, 3.2, 4.7)	31.43
4	10.00	-10.00	(-50, 20, 30)	1266.67
5	20.00	-20.00	(-100, 40, 60)	5066.67

Cas	$c_1$	$c_2$	$Y_1$	$Var(Y_1)$
6	0.71	0.71	(-4.9, -1.4, 6.4)	22.33
7	0.71	-0.71	(-3.5, 1.4, 2.1)	6.33
8	0.94	-0.34	(-5.3, 0.7, 4.6)	16.68
9	0.94	0.34	(-6.0, -0.7, 6.7)	26.88

Tableau n°1 : évolution de  $Var(Y_1)$  en fonction de  $c_1$  et  $c_2$ .

En fait, à partir des données numériques que nous avons retenues, il apparaît que la combinaison  $(c_1 = 0.94, c_2 = 0.34)$  est celle qui permet de rendre maximale la variance des individus sur  $Y_1$  tout en respectant la contrainte  $c_1^2 + c_2^2 = 1$ .

Sachant que l'inertie totale du nuage de points est égale à 28.67 et que la variance (inertie) des individus sur  $Y_1$  (nouvelle variable ou composante principale) vaut 26.88, on peut donc en déduire que  $Y_1$  restitue  $26.88/28.67=94\%$  de l'inertie initiale du nuage.

Dans l'exemple précédent, nous avons obtenu les valeurs numériques de  $c_1$  et  $c_2$  en résolvant un programme mathématique de maximisation d'une variance sous contrainte. De façon plus formelle, le **premier axe principal** est en fait une droite (définie dans  $\mathbb{R}^p$ ) passant par le centre d'inertie du nuage de points centré ou centré-réduit et maximisant son inertie sur cette droite. Sans reprendre la démonstration mathématique conduisant à ce résultat<sup>1</sup>, on montre que :

- 1) si on ne fait que centrer les données, cette droite est engendrée par le **vecteur propre de la matrice des variances-covariances  $C$  associé à la plus grande valeur propre** (maximisation de l'inertie),
- 2) si on centre et on réduit les données, cette droite est engendrée par le **vecteur propre de la matrice des corrélations  $R$  associé à la plus grande valeur propre** (maximisation de l'inertie).

---

<sup>1</sup> Le lecteur intéressé par une telle démonstration pourra se rapporter aux ouvrages de G. Saporta cités dans la bibliographie.