

BASES D'ALGÈBRE LINÉAIRE POUR LA MODÉLISATION EN SANTÉ

Code couleur : vert pour les exemples, bleu pour les infos supplémentaires (hors diapo)

I. POURQUOI MODÉLISER EN SANTÉ ?

Modéliser permet de reproduire informatiquement une situation dans le but de tester des scénarios, des configurations .

Dans le contexte de **Big Data** (données massives) c'est-à-dire de données de différentes natures, hétérogènes ... il est possible de les structurer dans des grands tableaux (**matrices**) et d'effectuer des opérations mathématiques sur ces matrices (évolution, distances entre profils...).

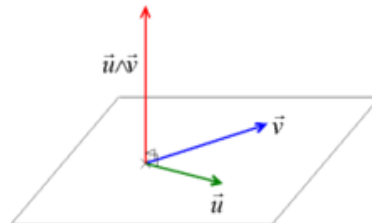
*Par exemple, estimer l'effet d'une augmentation du prix des cigarettes sur le système de santé (maladies, coût pour la société...), on est typiquement dans un **problème multifactoriel** et de Big data (il faut un grand nombre d'individus fumeurs ou non-fumeurs pour évaluer cet impact)*

Le calcul matriciel est précieux dans le domaine des **statistiques multivariées** (ACP, AFC...).

II. DÉFINITION DE L'ALGÈBRE LINÉAIRE

L'algèbre linéaire est un domaine des mathématiques qui étudie les transformations linéaires et les espaces vectoriels.

Un espace vectoriel est une **structure stable** par addition de vecteurs et par multiplication par un scalaire. Autrement dit, on peut ajouter deux éléments d'un tel espace, ou les multiplier par un nombre, le résultat appartiendra encore à l'espace de départ.



Le tutorat est gratuit. Toute vente ou reproduction est interdite.

Les transformations linéaires et les espaces vectoriels sont des outils pour l'**analyse multivariée de données** dans un contexte de BIG DATA.

A. BASES DU CALCUL MATRICIEL

Une **matrice** est un tableau de nombres à **n lignes** et **p colonnes** : A (n, p) c'est par exemple la donnée de n individus mesurés selon p variables (p≥1).

Point tut' : Une matrice A (n,p) est un tableau à np éléments.

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,p} \\ a_{2,1} & a_{2,2} & \dots & a_{2,p} \\ \dots & \dots & \dots & \dots \\ a_{n,1} & a_{n,2} & \dots & a_{n,p} \end{pmatrix} \text{ avec } (a_{i,j}) \text{ coefficients de la matrice}$$

$$1 \leq i \leq n$$

$$1 \leq j \leq p$$

- Si p = 1, on parle de matrice univariée (=matrice colonne). $A = \begin{pmatrix} 7 \\ 6 \end{pmatrix}$
- Si p ≥ 2, on parle de matrice multivariée. Ex : $B = \begin{pmatrix} 5 & 9 \\ 2 & 11 \\ 33 & 5 \end{pmatrix}$
- Si n=p, on parle de matrice carrée d'ordre n, cette matrice a autant de lignes (n) que de colonnes (p=n). Ex : $C = \begin{pmatrix} 8 & 14 \\ 4 & 10 \end{pmatrix}$ est une matrice carrée d'ordre 2.

Pour calculer le **produit** de deux matrices A et B, il faut que le nombre de **lignes** de la **deuxième** matrice soit égale au nombre de **colonnes** de la **première** matrice. Ainsi $A(n,p) \cdot B(p, m) = C(n, m)$

Tut'Astuce : Il n'est pas toujours facile de faire le produit de deux matrices, pour s'aider on peut placer les matrices de la façon suivante :

<div style="background-color: #4a7ebb; width: 20px; height: 20px; display: inline-block; margin-right: 5px;"></div> $\begin{bmatrix} 9 \\ 2 \end{bmatrix}$	Le but étant d'avoir un carré en haut à gauche comme dans le cas bleu. Si on n'obtient pas de carré alors on sait qu'on ne peut pas calculer le produit (cas rouge).	<div style="background-color: #e74c3c; width: 20px; height: 20px; display: inline-block; margin-right: 5px;"></div> $\begin{bmatrix} 2 & 3 \\ 1 & -1 \end{bmatrix}$
$\begin{bmatrix} 2 & 3 \\ 1 & -1 \end{bmatrix}$		$\begin{bmatrix} 9 \\ 2 \end{bmatrix}$

Les **puissances** d'une matrice n'existent que pour des **matrices carrées**, en effet pour effectuer un produit de matrices il faut que le nombre de lignes de la deuxième matrice soit égale au nombre de colonnes de la première matrice, or dans ce cas $B=A$ et donc $n=p$.

Ex : On peut calculer $A = \begin{pmatrix} 2 & 4 \\ 6 & 7 \end{pmatrix}^2$ mais pas $D = \begin{pmatrix} 43 \\ 57 \end{pmatrix}^2$

B. TRANSPOSÉE, NILPOTENTE, INVERSE ET DÉTERMINANT

La **transposée** d'une matrice revient à présenter l'information qui est en colonnes en lignes. Ainsi la transposée de $A(n,p)$, notée tA est une matrice à p lignes et n colonnes. Autrement dit, on s'intéresse à p variables déclinées selon n individus. tA s'obtient en échangeant les lignes et les colonnes de A.

Ex : Si $A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}$, alors on a ${}^tA = \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix}$

Point tut' : La transposée d'une matrice existe toujours.

Lorsque l'on effectue le **produit ${}^tA.A$** , on obtient une **matrice carrée d'ordre p**, qui est une matrice **symétrique** (par rapport à la diagonale qui part du coefficient en haut à gauche de la matrice).

Ex : $\begin{pmatrix} 17 & 22 & 27 \\ 22 & 29 & 36 \\ 27 & 36 & 45 \end{pmatrix}$ est symétrique (de chaque côté de cette diagonale)

Point tut' : Soit A une **matrice carrée** d'ordre n.

- 1) A est dite symétrique si et seulement si ${}^tA = A$
- 2) A est dite asymétrique si et seulement si ${}^tA = -A$

Exemples :

$\begin{pmatrix} 1 & 3 \\ 3 & 2 \end{pmatrix}$ est symétrique. $\begin{pmatrix} 0 & -7 \\ 7 & 0 \end{pmatrix}$ est anti-symétrique.

Une matrice est dite **nilpotente** d'ordre n lorsque $A^n=0$ et $A^{n-1} \neq 0$. Une matrice est nilpotente lorsqu'il existe une puissance pour laquelle cette matrice est égale à la matrice nulle.

Ex : $\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}^2 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$

L'**inverse** d'une matrice n'existe que pour des **matrices carrées** sous conditions (**detA ≠ 0**), il est défini comme A^{-1} tel que **$AA^{-1}=I$** où I est la matrice identité c'est-à-dire avec tous les coefficients diagonaux égaux à 1 et tous les autres coefficients nuls. Le déterminant sera utile pour calculer l'inverse.

Ex : $I = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ est la matrice identité d'ordre 3

Point tut' : La transposée d'une matrice inversible est toujours inversible.

Le **déterminant** d'une matrice (det A) est donné par la formule suivante :

- $\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc$

Ex : $\det \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = 1 \times 4 - 2 \times 3 = -2$

- $\det \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} = a \times \det \begin{pmatrix} e & f \\ h & i \end{pmatrix} - b \times \det \begin{pmatrix} d & f \\ g & i \end{pmatrix} + c \times \det \begin{pmatrix} d & e \\ g & h \end{pmatrix}$

Ex :

$$\begin{aligned} \det \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} &= 1 \times \det \begin{pmatrix} 5 & 6 \\ 8 & 9 \end{pmatrix} - 2 \times \det \begin{pmatrix} 4 & 6 \\ 7 & 9 \end{pmatrix} + 3 \times \det \begin{pmatrix} 4 & 5 \\ 7 & 8 \end{pmatrix} \\ &= 1 \times (5 \times 9 - 6 \times 8) - 2 \times (4 \times 9 - 6 \times 7) + 3 \times (4 \times 8 - 5 \times 7) \\ &= -3 + 12 - 9 = 0 \end{aligned}$$

Pour la matrice 3x3, pas la peine de retenir la formule, retenez visuellement.

Soit $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, une matrice carrée d'ordre 2 :

Si $\det(A) \neq 0$, alors $A^{-1} = \frac{1}{ad-bc} \times \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$

C. MATRICES DIAGONALES (supplément)

Une matrice carrée est dite **diagonale** lorsque ses coefficients en dehors de la diagonale principale sont nuls. Les coefficients de la diagonale peuvent être ou ne pas être nuls.

Toute puissance d'une matrice diagonale est une matrice diagonale dont les coefficients diagonaux sont les puissances des coefficients diagonaux de la matrice de départ.

Soit $A = \begin{pmatrix} 8 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$. $A^n = \begin{pmatrix} 8^n & 0 & 0 \\ 0 & 2^n & 0 \\ 0 & 0 & 1 \end{pmatrix}$; $A^2 = \begin{pmatrix} 64 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 1 \end{pmatrix}$

Dans le cas d'une matrice diagonale inversible, les éléments diagonaux de l'inverse sont les inverses des éléments de la matrice. Sachant qu'une matrice diagonale est inversible si ses coefficients diagonaux sont non nuls.

Soit $A = \begin{pmatrix} 8 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$. $A^{-1} = \begin{pmatrix} 1/8 & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$.

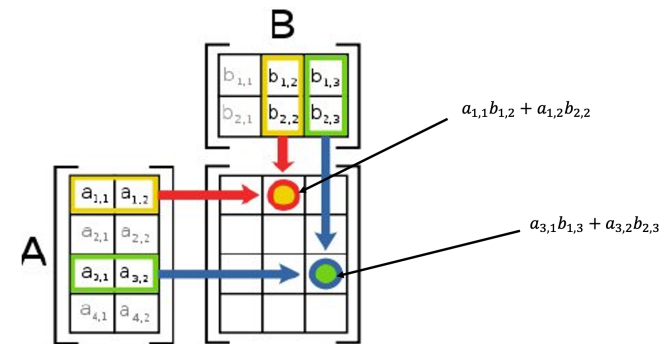
$B = \begin{pmatrix} 5 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ n'est pas inversible

Toujours pour une matrice diagonale inversible : $(A^n)^{-1} = (A^{-1})^n$, soit : « l'inverse de la puissance est égal à la puissance de l'inverse ».

Soit $A = \begin{pmatrix} 8 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$, $(A^n)^{-1} = \begin{pmatrix} 1/8^n & 0 & 0 \\ 0 & 1/2^n & 0 \\ 0 & 0 & 1 \end{pmatrix}$.

Toute matrice diagonale commute avec n'importe quelle de ses puissances, et leur inverse. De manière plus générale, toute matrice diagonale commute avec toute autre matrice diagonale.

D. CALCUL MATRICIEL, PRODUIT DE MATRICES



Soit $A = \begin{pmatrix} -1 & 2 & 1 \\ 3 & 4 & 5 \end{pmatrix}$ et $B = \begin{pmatrix} 1 & -4 & 2 \\ 3 & 0 & 1 \\ 6 & 5 & 7 \end{pmatrix}$, on a $AB = C = \begin{pmatrix} 11 & 9 & 7 \\ 45 & 13 & 45 \end{pmatrix}$. On remarque qu'on peut calculer AB mais pas BA. Pour trouver AB, on a calculé les coefficients un à un :

$c_{1,1} = (-1) \times 1 + 2 \times 3 + 1 \times 6 = 11$

$c_{1,2} = (-1) \times (-4) + 2 \times 0 + 1 \times 5 = 9$

$c_{1,3} = (-1) \times 2 + 2 \times 1 + 1 \times 7 = 7$

$$c_{2,1} = 3 \times 1 + 4 \times 3 + 5 \times 6 = 45$$

$$c_{2,2} = 3 \times (-4) + 4 \times 0 + 5 \times 5 = 13$$

$$c_{2,3} = 3 \times 2 + 4 \times 1 + 5 \times 7 = 45$$

Il peut être pratique de poser les matrices au brouillon comme ceci pour éviter de se tromper dans les lignes et colonnes :

$$\begin{pmatrix} -1 & 2 & 1 \\ 3 & 4 & 5 \end{pmatrix} \begin{pmatrix} 1 & -4 & 2 \\ 3 & 0 & 1 \\ 6 & 5 & 7 \\ 11 & 9 & 7 \\ 45 & 13 & 45 \end{pmatrix}$$

Dans le cas général, le produit de matrice AB est différent de BA. Il se peut même que l'un des produits n'existe pas. Dans l'exemple précédent, le produit BA n'est pas possible car le nombre de colonnes de B est différent du nombre de lignes de A.

⇒ Si **AB = BA** on dit que les **matrices commutent +++**

Aussi, pour toute matrice carrée, si A^{-1} existe, alors A^{-1} et A commutent toujours.

On peut avoir un produit de matrices nul sans que l'une des matrices soit nulle.

Exemple : Pour $A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ et $B = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$, on a $AB = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$.

On appelle une **matrice « creuse »**, une matrice qui contient beaucoup de 0. Ainsi, le produit de deux matrices creuses peut donner une matrice nulle.

$$\text{Ex : } \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 2 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 \\ 10 & 2 & 9 \\ 2 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

III. Calcul matriciel pour LES ANALYSES MULTIVARIÉES (analyses factorielles) – modélisation en santé

Le terme **d'analyses factorielles** désigne un ensemble de techniques d'ajustement linéaire dont le but est de résumer l'essentiel de l'information contenue dans des gros tableaux de données (plusieurs dizaines, centaines d'individus observés selon un grand nombre de variables [supérieures à 100 par exemple]).

Le procédé consiste à passer d'un espace de grandes dimensions à un espace plus petit (factorisation du tableau de données) avec une **perte d'information minimale et contrôlée**.

2 techniques principales d'analyse factorielle :

- L'analyse en **composantes principales** ou ACP : la plus ancienne des deux (1933) mais véritablement développée avec l'informatique, employée dans le cadre de variables quantitatives, homogènes ou pas.
- L'analyse **factorielle des correspondances** ou AFC (années 70), pour l'étude de tableaux de contingence (données qualitatives)

D'un point de vue méthodologique, le fonctionnement des deux méthodes, est le même, aussi on ne décrira ici que l'ACP.

A. ACP - INTÉRÊT ET ANALYSE DE DONNÉES

L'intérêt de l'ACP est de :

- Extraire le **maximum d'informations** sous une forme simple et cohérente à partir d'un ensemble très important de données (description synthétique)

- Permettre de mettre en évidence
 - Des **interrelations entre variables** (redondance)
 - Des **ressemblances ou oppositions** entre individus
- Les résultats se présentent sous forme de combinaisons linéaires de variables différenciant les individus statistiques

L'ACP s'applique uniquement sur des variables quantitatives qui peuvent être exprimées :

- Soit dans une même unité (exemple : le % de cas de Covid 19 dans les passages aux urgences)
- Soit dans des unités différentes (exemple : la mort infantile, le revenu par habitant ...)

Le tableau de données est constitué de n individus statistiques (unités spatiales, individus ...) caractérisés par p variables quantitatives. Ce tableau de données D constitue une **matrice d'informations (n,p)**.

- Chaque ligne = vecteur ligne décrit un individu selon p variables
- Chaque colonne = vecteur colonne décrit un indicateur selon n individus.

Exemple :

	Taille (m)	Masse (kg)	IMC	Tour de tête (cm)	Âge (années)	Moyenne au bac	...
Patient 1	1.60	50	19,531	55	18	17	...
Patient 2	1,82	78	23,548	60	23	14	...
...

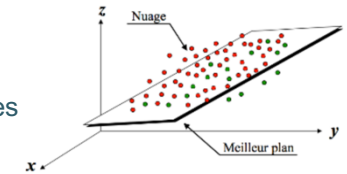
Ainsi, le but de l'ACP est de prendre l'information contenue dans ce tableau de dimension importante et la représenter sous forme simplifiée.

B. ACP - MÉTHODE

L'ACP consiste à réduire la taille du nuage de points multi-dimensionnels en un nuage de points en 3-4 dimensions.

Point tut' : Lorsqu'on analyse 1 variable, on a besoin que d'une dimension pour la représenter : c'est la représentation axiale. Pour analyser 2 variables, on passe à 2 dimensions (sur un graph, on peut par exemple représenter la masse sur l'axe des abscisses et la taille sur l'axe des ordonnées). Quand on ajoute une 3^{ème} variable, on passe à 3 dimensions (cf image), on ajoute un nouvel axe et ainsi de suite.

Au-delà de 3 variables, on va utiliser l'ACP qui permettra de condenser les informations de variables différentes dans les premiers axes



Pour cela on fait une projection selon des axes (axes factoriels ou facteurs). Ces axes sont des combinaisons linéaires de variables .

$$F_i = A_1 X_1 + A_2 X_2 + \dots + A_p X_p \text{ avec la plupart des } A_i = 0$$

Les **coefficients A_i** permettent de mesurer l'intensité de la relation de chaque variable avec le facteur considéré (de max à nul selon les composantes), les coefficients A_i changent d'un facteur à l'autre. Les coefficients étant les données que l'on trouve dans les matrices. Les coefficients A_i servent comme dans une régression multiple, à exprimer la part de chaque variable dans la constitution de l'axe.

Les facteurs sont **hiérarchisés** : l'axe 1 compte le maximum d'informations, c'est l'axe de plus grande dispersion du nuage de points, mais il laisse de côté des résidus. C'est le 2^{ème} axe qui prend en compte le maximum d'informations résiduelles et ainsi de suite pour les axes suivants.

Par construction, tous les axes (facteurs) sont non corrélés, ils forment donc des angles droits deux à deux. Lorsque les facteurs sont corrélés ils forment des angles aigus ou obtus.

Aussi, un plan est formé par deux vecteurs directeurs ou deux axes de même origine.

C. ACP – CALCUL DES AXES FACTORIELS

L'idée est de transformer la **matrice d'information** en une **matrice de projection** des individus statistiques sur les axes.

Pour cela, une étape importante est de **centrer-réduire les données** :

Centrer-réduire les données permet de gommer les effets taille. Si les données sont assez **homogènes**, on pratique un simple **centrage**. Si les données sont assez **hétérogènes**, le **centrer-réduire** est obligatoire (ramène la moyenne à 0 et l'écart-type à 1) ce qui donne une ACP normée.

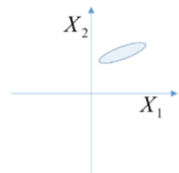
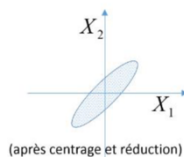


Fig. 7 ACP générale



(après centrage et réduction)
Fig. 9 ACP normée

L'ACP normée consiste en :

- Variables centrées réduites
- Projections orthogonales
- Méthode des moindres carrés

Point tut' :

Les axes (=facteurs) sont hiérarchisés : le premier axe contient le maximum d'informations, c'est l'axe sur lequel le nuage se déforme le moins possible en projection. On cherche ensuite le second axe sur lequel le nuage se déforme le moins **après** le premier axe (en projection), tout étant orthogonal (perpendiculaire) au premier axe. On réitère jusqu'à l'obtention de p axes.

Le meilleur axe (premier axe) est celui où la dispersion du nuage de points est maximale, c'est-à-dire où l'**inertie** du nuage est maximale. On va voir la notion de « matrice inertie ».

Détermination des axes factoriels :

La matrice de données D (individus variables) est une matrice à n lignes et p colonnes. À cette matrice on associe une matrice **D'**, la **matrice transposée de D**.

En faisant le produit de D'.D on obtient une matrice carrée, symétrique d'ordre p appelée **matrice d'inertie**, notée **T**.

Toute l'ACP repose sur du calcul matriciel. Les axes sont définis par des **vecteurs propres** et **valeurs propres** :

- Un **vecteur propre** est un vecteur tel que **T.V=μV** où **μ** est une **valeur propre**.

Exemple : Soit $T = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$.

Pour trouver la valeur propre μ , on résout le système $T.V=\mu V$.

$$T = \mu I$$

$$\text{On a : } T - \mu I = 0, \text{ donc } T - \mu I = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} - \mu \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1-\mu & 2 \\ 2 & 1-\mu \end{pmatrix} = 0$$

En effet, quand on divise par V des 2 côtés de l'égalité, comme c'est du calcul matriciel on ne peut pas écrire $T = \mu$, on est obligé de garder une forme matricielle à gauche comme à droite, donc on multiplie par I, ici la matrice identité d'ordre 2.

Pour trouver une solution à ce type d'équation, on cherche à ce que le déterminant de la matrice $T - \mu I$ soit égal à 0. Ainsi :

$$\text{Det}(T - \mu I) = (1-\mu)^2 - 4 = (-1-\mu)(3-\mu) = 0$$

La résolution de cette équation donne $\mu = -1$ et $\mu = 3$.

On résout alors les systèmes $T.V = -V$ et $T.V = 3V$ (le prof ne détaille que le 2^{ème} système).

$$\begin{aligned}
 T.V = 3V &\Leftrightarrow \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} * V = 3V \\
 &\Leftrightarrow \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} * \begin{pmatrix} V_1 \\ V_2 \end{pmatrix} = 3 \begin{pmatrix} V_1 \\ V_2 \end{pmatrix} \\
 &\Leftrightarrow \begin{pmatrix} V_1 + 2V_2 \\ 2V_1 + V_2 \end{pmatrix} = \begin{pmatrix} 3V_1 \\ 3V_2 \end{pmatrix} \\
 &\Leftrightarrow \begin{cases} V_1 + 2V_2 = 3V_1 \\ 2V_1 + V_2 = 3V_2 \end{cases} \\
 &\Leftrightarrow V_1 = V_2 = \begin{pmatrix} 1 \\ \frac{1}{\sqrt{2}} \\ 1 \\ \frac{1}{\sqrt{2}} \end{pmatrix} \text{ car les vecteurs sont normés}
 \end{aligned}$$

De même, l'autre équation mène à $V_3 = V_4 = \begin{pmatrix} -1 \\ \frac{1}{\sqrt{2}} \\ 1 \\ \frac{1}{\sqrt{2}} \end{pmatrix}$

Remarquons que les deux vecteurs obtenus sont bien orthogonaux (leur produit scalaire est nul : les axes sont donc bien décorrélés).

D. ACP – part d'explication des axes factoriels

La **part d'explication** d'un axe est donné par la formule suivante :

$$\mu_i \% = \frac{\mu_i}{\sum \mu_i} * 100$$

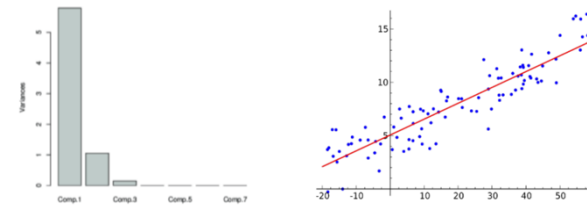
Dans le cas à n dimensions, deux situations peuvent se produire :

- Soit l'**histogramme des valeurs propres** est assez droit, le nuage de points est plutôt arrondi sans axe d'allongement véritablement marqué : on peut en déduire que les interrelations entre les variables sont sans doute faibles et qu'il ne se dégage pas de combinaisons simples de l'ensemble des données.

- Soit l'**histogramme des valeurs propres** est assez concentré, les valeurs propres sont très différenciées et on perçoit l'existence de deux axes d'allongement très marqués, on peut s'attendre alors à ce qu'il ressorte une structure de différenciation forte (profils marqués)

Point tut' : La valeur propre permet de caractériser l'importance de cet axe par rapport aux valeurs de notre nuage de points.

Exemple 1 : Dans un plan 2D, ce nuage de points peut être caractérisé de manière très significative par la droite rouge, qui est alors un facteur (=axe factoriel). On peut supposer que ce facteur aura alors une valeur propre μ et une part d'explication (implication dans la forme du nuage) importante ! On pourrait alors avoir un histogramme assez concentré avec un axe d'allongement au niveau de μ_1 très marqué, comme si contre :



Exemple 2 : Dans un plan 2D, sur ce type de nuage de points, on n'observe aucun axe d'allongement marqué, les valeurs sont dispersées. L'histogramme des valeurs propres est alors assez droit, aucune ne prend le dessus sur les autres. On peut en déduire que les interrelations entre les variables sont faibles.

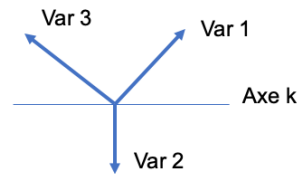


E. ACP – Interprétation d'une analyse factorielle

Plusieurs données doivent être analysées :

- Les **coordonnées** sur les axes factoriels : elles donnent la **position** des individus par rapport aux axes factoriels. On peut alors mettre en évidence des **oppositions entre groupe d'individus** par rapport aux combinaisons de variables définies par les axes.
- **La qualité de représentation** des individus sur les axes factoriels : deux points distincts peuvent avoir la même projection sur l'axe factoriel, mais l'un sera mieux représenté que l'autre (angle plus petit).

Exemple : les variables 1 et 3 sont bien corrélées à l'axe k, elles forment un angle aigu avec l'axe k (bonne saturation sur l'axe), elles sont bien représentées. Tandis que la variable 2 est orthogonale à l'axe k (saturation nulle), elle n'a pas une bonne qualité de représentation.



- **La contribution des individus** dans la formation de l'axe : les individus contribuent plus ou moins à déterminer la direction des différents axes d'allongement du nuage. Elle est mesurée par la part de l'individu dans la variance.
- **La part de l'individu dans l'inertie totale du nuage** = INR : elle est proportionnelle à sa distance au centre de gravité (G). Elle donne une idée de la spécificité de l'individu par rapport à la moyenne.

Ce cours peut sembler complexe au premier abord.

Concentrez-vous surtout sur la partie des matrices qui pour l'instant est la seule partie de ce cours à être tombée à l'examen et qui fait l'objet des qcms de fin de diapo du prof. Essayez de comprendre la partie ACP, ce sont des questions simples qui tomberont, pas la peine de retenir les détails. Tout ce qui est écrit en bleu sont des aides à la compréhension ajoutées par vos vieux de Biostat et moi.

Faites des qcms, ce sont souvent les mêmes types de questions qui reviennent, pour cela annatus + DMs. Faites les qcms du diapo du prof lorsqu'il sortira, un DM des qcms du prof sera prévu avec une correction plus détaillée que sur ses diapos. De plus, des DMs matrices seront postés afin de bien vous entraîner.

Enfin, à partir de la semaine prochaine (10/10), lors des permanences Discord, des exercices de matrices avec une correction détaillée étape par étape seront réalisés. La SDA n'a sûrement pas permis à tout le monde de comprendre comment fonctionnaient les matrices. J'ai été un peu pris par le temps désolé !

La Biostat n'est pas forcément la matière la plus accueillante. Pour autant cette année elle comptera vraiment pour votre classement. N'hésitez pas à regarder les annales, les qcms qui tombent ne sont pas aussi complexes que ce que l'on peut trouver dans les cours.

Bonnes révisions !

Fin