Introduction à la description statistiques multidimensionnelles

Présenté par : M. HAMMAD

La statistique classique étudie un nombre restient de caractéres mesurés sur un petit nombre d'individus, en revanche, l'analyse des données multidimensionnelles permet de traiter un ensemble de variables observées sur un ensemble d'individus, la notion de données multidimensionnelles se référant surtout au nombre important de variables et non au nombre d'individus concernés.

0.1 Échantillon d'une variable

Définition 0.1.1.
Soit I un ensemble de "n" individus décrit par une variable X.
L'ensemble $E = \{x_1, x_2, x_3, \cdots, x_n\}$ des n valeurs x_i, est échantillon de la variable quantitative X, à chaque individu i est associé un poids $p_i > 0$, tel que $\sum_{i \in I} p_i = 1$.

Remarque 0.1.1.
Il y a un cas particulier où les individus sont de même poids, $p_i = \frac{1}{n}$.

Réprésentation de l'échantillon E

On peut représenter E de deux manières différentes :
- l'ensemble E peut être considéré comme un point ou vecteur de $\mathbb{R}^n$:

$$X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}.$$

- l'ensemble E de peut être considéré comme sous-ensemble de $\mathbb{R}$, il est représenté graphiquement par les "n" valeurs.

Remarque 0.1.2. .

1. *La représentation de l'ensemble E dépend de l'ordre des valeurs x_i.*
2. *l'échantillon E sera noté X quel que soit le mode de représentation choisi.*

0.1.1 Indices statistiques sur E

La moyenne

Elle est définie par :$\bar{x} = \sum_{i \in I} p_i x_i$.

- Géométriquement, sur la droite $\mathbb{R}$, est le centre de gravité noté g des points x_i munis des poids p_i.

La variance et l'écart-type

Elles sont définie par :

- La variance : $Var(X) = \sum_{i=1}^{n} p_i (x_i - \bar{x})^2$.
- L'écart-type : $\sigma_x = \sqrt{Var(X)}$.

Remarque 0.1.3.

1. $Var(X) = 0 \Rightarrow$ *l'ensemble E est constitué de n valeurs égales.*
2. *L'écart type intervient en probabilité pour comparer des variables ou des distributions entre elles.*

0.1.2 Interprétation de la moyenne et la variance dans $\mathbb{R}$

Notre but c'est de trouver une valeur t qui est proche à toutes les valeurs $x_1, x_2, \cdots, x_n$ pour réduire l'échantillon E.
D'où la démarche suivante :

- Choisir une distance d dans $\mathbb{R}^n$.

- Chercher le point T dans $\mathbb{R}^n$ qui minimise $d(T, X)$.

Soit T un vecteur tel que :

$$T = \begin{pmatrix} t \\ t \\ \vdots \\ t \end{pmatrix} = t * \mathbf{1}, \text{ où } \mathbf{1} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$$

et $d(x,T)$ est la distance euclidienne entre x et T définie par :

$$\begin{aligned} d^2(x,T) &= \sum_{i=1}^{n} p_i(x_i-t)^2 \\ &\Rightarrow \quad d(x,T) = \sqrt{\sum_{i=1}^{n} p_i(x_i-t)^2}, \end{aligned}$$

on veut chercher t qui minimise $d(x,T)$, soit, $\dfrac{\partial}{\partial t} d(x,T) = 0$,
$\Rightarrow -2(\sum_{i=1}^{n} p_i x_i - \sum_{i=1}^{n} p_i t) = 0$, avec $\sum_{i=1}^{n} p_i = 1$,
$\Rightarrow t = \sum_{i=1}^{n} p_i x_i = \bar{x}$.

Conclusion 0.1.1.

$\bar{x}$ est le point le plus proche du nuage, i.e, la valeur qui résume "le mieux" l'échantillon E.

0.2 La Métrique des poids dans $\mathbb{R}^n$

Définition 0.2.1.

Si les données ont été recueillies à la suite d'un tirage aléatoire à probabilités égales, les individus ont tous même importance $\frac{1}{n}$, dans le calcul des caractéristiques de l'échantillon. Il n'en est pas toujours ainsi et il est utile pour certaines applications de travailler avec des poids p_i éventuellement différents d'un individu à l'autre (échantillons redressés ; données regroupées $\cdots$). Ces poid, qui sont des nombres positifs de somme 1 comparables à des fréquences, sont regroupées dans une matrice diagonale D_p de taille n : nomée matrice des poids

$$D_p = \begin{pmatrix} p_1 & 0 & \cdots & 0 \\ 0 & p_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & p_n \end{pmatrix}.$$

Dans le cas le plus usuel de poids égaux : $D_p = \frac{1}{n}.I_n$, avec I_n la matrice d'identité d'ordre n.

0.2.1 Calcul de la moyenne d'une variable par la Métrique D_p

On a défini déja la moyenne par : $\bar{x} = \sum_{i=1}^{n} p_i x_i$,

soit $\mathbf{1} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$ et $X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \in E$,

on a, le produit

$$X'.D_p = \begin{pmatrix} x_1, x_2, \cdots, x_n \end{pmatrix} . \begin{pmatrix} p_1 & 0 & \cdots & 0 \\ 0 & p_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & p_n \end{pmatrix} = \begin{pmatrix} p_1x_1, p_2x_2, \cdots, p_nx_n \end{pmatrix}$$

et par la suite

$$\bar{x} = X'.D_p.\mathbf{1} = \begin{pmatrix} p_1x_1, p_2x_2, \cdots, p_nx_n \end{pmatrix} . \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} = \sum_{i=1}^{n} p_i x_i \Longrightarrow \bar{x} = D_p(X, \mathbf{1}).$$

0.2.2 Calcul de la variance d'une variable par la Métrique D_p

On sait que la variance est égale à : $Var(X) = \sum_{i=1}^{n} p_i (x_i - \bar{x})^2$, soit $Z = X - \bar{x}.\mathbf{1}$ (Z est un vecteur centré)

$$Z = \begin{pmatrix} x_1 - \bar{x} \\ x_2 - \bar{x} \\ \vdots \\ x_n - \bar{x} \end{pmatrix} = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{pmatrix},$$

$$
\begin{aligned}
Var(X) &= p_1(x_1-\bar{x})^2 + p_2(x_2-\bar{x})^2 + \cdots + p_n(x_n-\bar{x})^2 \\
&= \begin{pmatrix} x_1-\bar{x}, x_2-\bar{x}, \cdots, x_n-\bar{x} \end{pmatrix} . \begin{pmatrix} p_1(x_1-\bar{x}) \\ p_2(x_2-\bar{x}) \\ \vdots \\ p_n(x_n-\bar{x}) \end{pmatrix} \\
&= \begin{pmatrix} x_1-\bar{x}, x_2-\bar{x}, \cdots, x_n-\bar{x} \end{pmatrix} . \begin{pmatrix} p_1 & 0 & \cdots & 0 \\ 0 & p_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & p_n \end{pmatrix} . \begin{pmatrix} x_1-\bar{x} \\ x_2-\bar{x} \\ \vdots \\ x_n-\bar{x} \end{pmatrix} \\
&= Z' D_p Z \\
&= \sum_{i=1}^{n} p_i Z_i^2 .
\end{aligned}
$$

0.3 Échantillon d'un couple de variables (X, Y)

Soit un ensemble I de n individus décrits par deux variables quantitatives X et Y. Un échantillon E du couple (X, Y) est donc l'ensemble des valeurs : $E = \{(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)\}$. à chaque individu est associé un poids $p_i > 0$ tel que $\sum_{i \in I} p_i = 1$.

-Les individus ont souvent le même poids, $p_i = \frac{1}{n}$.

À cet échantillon E on peut associer deux échantillons d'une seule variable :

-Un échantillon E_x de la varible X : $E_x = \{ \ x_i, i \in I \ \}$

- Un échantillon E_y de la variable Y : $E_y = \{ \ y_i, i \in I \ \}$

L'étude des échantillons E_x et E_y ne suffit pas pour étudier l'échantillon E; il faut mettre en évidence les liens ou l'absence de liens entre les deux variables X et Y.

Réprésentation graphique de l'échantillon E dans $\mathbb{R}^2$

0.3.1 Indices statistiques sur E pour deux variables

	Moyenne	Variance	Covariance	Coefficion de corrélation
X	$\bar{x} = \sum_{i=1}^{n} p_i x_i$	$\sum_{i=1}^{n} p_i (x_i - \bar{x})^2$	$Cov(X,Y) = \sum_{i=1}^{n} p_i (x_i - \bar{x})(y_i - \bar{y})$	$\dfrac{Cov(X,Y)}{\sigma_x \cdot \sigma_y}$
Y	$\bar{y} = \sum_{i=1}^{n} p_i y_i$	$\sum_{i=1}^{n} p_i (y_i - \bar{y})^2$		

Matrice des variances-covariances V

La matrice des variances-covariances V associée à l'échantillon E est symetrique à la quelle :

$$V = \begin{pmatrix} Var(X) & Cov(X,Y) \\ Cov(Y,X) & Var(Y) \end{pmatrix}.$$

Matrice de corrélation R

$$R = \begin{pmatrix} \dfrac{Var(X)}{\sigma_x^2} = 1 & \dfrac{Cov(X,Y)}{\sigma_x \cdot \sigma_y} \\ \dfrac{Cov(Y,X)}{\sigma_x \cdot \sigma_y} & \dfrac{Var(Y)}{\sigma_y^2} = 1 \end{pmatrix} = \begin{pmatrix} 1 & \rho_{X,Y} \\ \rho_{Y,X} & 1 \end{pmatrix},$$

où $\rho_{X,Y} = \dfrac{Cov(X,Y)}{\sigma_x \cdot \sigma_y}$ est le coefficient de correlation entre x et y.

La métrique $D_{\frac{1}{\sigma}}$ dans $\mathbb{R}^2$

On note $D_{\frac{1}{\sigma}}$ la matrice diagonale des inverses des écart-types :

$$D_{\frac{1}{\sigma}} = \begin{pmatrix} \dfrac{1}{\sigma_x} & 0 \\ 0 & \dfrac{1}{\sigma_y} \end{pmatrix}.$$

0.3.2 Relation entre V, R et $D_{\frac{1}{\sigma}}$

on peut trouver la matrice de corrélation R par la relation suivante :

$$R = D_{\frac{1}{\sigma}}.V.D_{\frac{1}{\sigma}}$$

Preuve 0.3.1.
on a :

$$D_{\frac{1}{\sigma}}.V = \begin{pmatrix} \dfrac{1}{\sigma_x} & 0 \\ 0 & \dfrac{1}{\sigma_y} \end{pmatrix} . \begin{pmatrix} Var(X) & Cov(X,Y) \\ Cov(Y,X) & Var(Y) \end{pmatrix} = \begin{pmatrix} \dfrac{Var(X)}{\sigma_x} & \dfrac{Cov(X,Y)}{\sigma_x \cdot \sigma_y} \\ \dfrac{Cov(Y,X)}{\sigma_x \cdot \sigma_y} & \dfrac{Var(Y)}{\sigma_y} \end{pmatrix}$$

et par la suite

$$\begin{pmatrix} \frac{Var(X)}{\sigma_x} & \frac{Cov(X,Y)}{\sigma_x \cdot \sigma_y} \\ \\ \frac{Cov(Y,X)}{\sigma_x \cdot \sigma_y} & \frac{Var(Y)}{\sigma_y} \end{pmatrix} \cdot \begin{pmatrix} \frac{1}{\sigma_x} & 0 \\ \\ 0 & \frac{1}{\sigma_y} \end{pmatrix} = \begin{pmatrix} \frac{Var(X)}{\sigma_x^2} & \frac{Cov(X,Y)}{\sigma_x \cdot \sigma_y} \\ \\ \frac{Cov(Y,X)}{\sigma_x \cdot \sigma_y} & \frac{Var(Y)}{\sigma_y^2} \end{pmatrix}$$

$$\Rightarrow R = D_{\frac{1}{\sigma}}.V.D_{\frac{1}{\sigma}}.$$

0.3.3 La relation entre la covariance et la Métrique D_p

On peut écrire la covariance en fonction de la métrique D_p :

$$\begin{aligned} Cov(X,Y) &= \sum_{i=1}^{n} p_i(x_i - \bar{x})(y_i - \bar{y}) \\ &= (x - \bar{x})' D_p (y - \bar{y}) \\ &= D_p(x - \bar{x}, y - \bar{y}). \end{aligned}$$

La relation matricielle entre le Coefficient de corrélation et la Métrique D_p

On a :

$$\begin{aligned} \rho_{X,Y} &= \frac{Cov(X,Y)}{\sigma_x \cdot \sigma_y} \\ &= \frac{D_p(x - \bar{x}\mathbf{1}, y - \bar{y}\mathbf{1})}{\|x - \bar{x}\mathbf{1}\|_{Dp} . \|y - \bar{y}\mathbf{1}\|_{Dp}}. \end{aligned}$$

D_p-**angle** Représente le cosinus d'angle fromé par $(x - \bar{x})$ et $(y - \bar{y})$.

Cas particulier x et y sont centrés

$\bar{x} = 0, \bar{y} = 0 \Rightarrow g = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ tel que g est le centre de gravité,

ceci entraîne que :

	Variance	Covariance	Coefficient de corrélation linéarie
X	$D_p(x,x) = \|x\|_{D_p}^2$	$D_p(x,y)$	$\frac{D_p(x,y)}{\|x\|_{Dp}\|y\|_{Dp}}$
Y	$D_p(y,y) = \|y\|_{D_p}^2$		

Si $x \underset{D_p}{\perp} y = D_p(x,y) = 0 \Leftrightarrow < x, y >_{D_p} = 0,$ on dit que x et y ne sont pas corrélées.

0.4 Échantillon de p variables

I est l'échantillon de variable X^j d'écrit par les n individus $E_j = \{x_{j1}, x_{j2}, \cdots, x_{jn}\}$ tel que $1 \leq j \leq p$ et K l'échantillon d'individu X_i d'écrit par les p variables $E_i = \{x_{1i}, x_{2i}, \cdots, x_{pi}\}$ tel que $1 \leq i \leq n$.

0.4.1 Représentation des données

Dans notre cas les données sont quantitatives, leurs représentation sous forme d'un tableau à p lignes (caractéres) et n colonnes (individus), $p \leq n$.

$$
\begin{array}{cc}
 & \begin{array}{ccccc} X_1 & X_2 & \cdots & \cdots & X_n \\ \downarrow & \downarrow & & & \downarrow \end{array} \\
X = & \begin{pmatrix} x_{11} & x_{12} & \cdots & \cdots & x_{1n} \\ x_{21} & \ddots & & & x_{2n} \\ \vdots & & x_{ij} & & \vdots \\ \vdots & & & \ddots & \vdots \\ x_{p1} & x_{p2} & \cdots & \cdots & x_{pn} \end{pmatrix} \begin{array}{l} \longleftarrow X^1 \\ \longleftarrow X^2 \\ \quad\vdots \\ \quad\vdots \\ \longleftarrow X^p \end{array}
\end{array}
$$

Remarque 0.4.1.

$$\begin{cases} X_i : \textit{le } i^{ime}\textit{individu} \\ X^j : \textit{la } j^{ime}\textit{variable} \end{cases}$$

x_{ji} : *valeur prise par le* i^{ime} *individu sur la* j^{ime} *variable.*

Espace des variables ou caractéres

Une ligne j caracterise une variable X^j est une liste de n valeur numeriques, c'est un vecteur de l'espace $\mathbb{R}^n$ de dimension n :

$$X^j = \begin{pmatrix} x_{j1} \\ x_{j2} \\ \vdots \\ x_{jn} \end{pmatrix}.$$

Espace des individus ou observations

La colonne i caracterise l'individu i et chaque individu X_i est un point dans un espace $\mathbb{R}^p$ de dimension p,

$$X_i = \begin{pmatrix} x_{1i} \\ x_{2i} \\ \vdots \\ x_{pi} \end{pmatrix}$$

Les individus sont munis de poids p_i tel que $p_i > 0, \sum_{i=1}^{n} p_i = 1$, eventuellement différents d'un individu à l'autre.

Soit D_p la matrice diagonale des poids. Le point moyen g, ou centre de gravité résume dans l'espace $\mathbb{R}^p$ le nuage E; la j^{ime} composante de ce point est le nombre :

$$\bar{x}^j = \sum_{i=1}^{n} p_i x_{ji} \ j = \overline{1,p}.$$

Si on désigne par $\mathbb{1}$ le vecteur (colonne) dont tous les éléments sont égaux à 1, on obtient :

$$g = X'.D_p.\mathbb{1}.$$

La Matrice des variances-covariances des p variables V :

$$V = \begin{pmatrix} Var(X^1) & Cov(X^1,X^2) & \cdots & Cov(X^1,X^p) \\ Cov(X^1,X^2) & Var(X^2) & \ddots & \vdots \\ \vdots & \ddots & \ddots & Cov(X^{p-1},X^p) \\ Cov(X^1,X^p) & \cdots & Cov(X^{p-1},X^p) & Var(X^p) \end{pmatrix}$$

V est une matrice carée d'ordre p, symétrique définе positive.

Remarque 0.4.2.
Si la matrice X est centrée alors on peut définir la matrice V de variance covariance par la métrique D_p :

$$V = X \cdot D_p \cdot X'.$$

Si non, on le centre par g on le note $X_c = (x_{ji} - \bar{x}^j)_{\substack{1 \le i \le n \\ 1 \le j \le p}}$ c'est à dire V vas devenir : $V = X_c \cdot D_p \cdot X_c'$.

De même l'ensemble des coefficients de corrélation est regroupé dans la matrice de corrélation R dont les termes diagonaux vaut 1 puisque $\rho(X^j, X^j) = 1$.

$$R = \begin{pmatrix} 1 & \rho(X^1, X^2) & \cdots & \rho(X^1, X^p) \\ \rho(X^1, X^2) & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \rho(X^p - 1, X^p) \\ \rho(X^1, X^p) & \cdots & \rho(X^p - 1, X^p) & 1 \end{pmatrix}.$$

Si on note $D_{\frac{1}{\sigma}}$ la Métrique dans $\mathbb{R}^n$ des inverses des écartes-types

$$D_{\frac{1}{\sigma}} = \begin{pmatrix} \frac{1}{\sigma_{x^1}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sigma_{x^2}} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \frac{1}{\sigma_{x^p}} \end{pmatrix}.$$

On a la relation matricielle entre R et $D_{\frac{1}{\sigma}}$: $R = D_{\frac{1}{\sigma}} \cdot V \cdot D_{\frac{1}{\sigma}}$, définie dans (0.3.2).

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

- Déterminer la matrice des poids D_p.
- Centré la matrice X.
- Calculer la matrice de variance covariance V.
- Déterminer la matrice $D_{\frac{1}{\sigma}}$ et la matrice R des coefficients de corrélation.

```
## On représante le tableau des données par la matrice X:##
>
A=matrix(c(6,8,6,14.5,14,11,5.5,13,9,6,8,7,14.5,14,10,7,12.5,9.5,5,8,11,15.5,
12,5.5,14,8.5,12.5,5.5,8,9.5,15,12,7,11.5,9.5,12,8,
9,11,8,10,13,10,12,18),nrow=9,ncol=5,byrow=FALSE)
> A
      [,1] [,2] [,3] [,4] [,5]
 [1,]  6.0  6.0  5.0  5.5    8
 [2,]  8.0  8.0  8.0  8.0    9
 [3,]  6.0  7.0 11.0  9.5   11
 [4,] 14.5 14.5 15.5 15.0    8
 [5,] 14.0 14.0 12.0 12.0   10
 [6,] 11.0 10.0  5.5  7.0   13
 [7,]  5.5  7.0 14.0 11.5   10
 [8,] 13.0 12.5  8.5  9.5   12
 [9,]  9.0  9.5 12.5 12.0   18
>
>X=t(A)
> X
     [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9]
[1,]  6.0    8  6.0 14.5   14 11.0  5.5 13.0  9.0
[2,]  6.0    8  7.0 14.5   14 10.0  7.0 12.5  9.5
[3,]  5.0    8 11.0 15.5   12  5.5 14.0  8.5 12.5
[4,]  5.5    8  9.5 15.0   12  7.0 11.5  9.5 12.0
[5,]  8.0    9 11.0  8.0   10 13.0 10.0 12.0 18.0
>
>
>##Soit Dp la matrice des poids ##
>
>Dp=1/9*diag(9)
>
>Dp
>
>##c le vecteur moyennes de variables##
>
>c= apply(X,1,mean)
>
> c
[1]  9.666667  9.833333 10.222222 10.000000 11.000000
>
>
```

```
##Xc la matrice centré de X ##
>
>Xc=sweep(X,1,c,FUN="-")
>
>Xc
>
>## La matrice des variances-covariances V ##
>
>V=Xc%*%Dp%*%t(Xc)
> V
          [,1]       [,2]       [,3]      [,4]       [,5]
[1,] 11.3888889 9.91666667  2.6574074 4.5833333 0.11111111
[2,]  9.9166667 8.94444444  4.1203704 5.2500000 0.05555556
[3,]  2.6574074 4.12037037 12.0617284 9.1944444 0.38888889
[4,]  4.5833333 5.25000000  9.1944444 7.6666667 0.72222222
[5,]  0.1111111 0.05555556  0.3888889 0.7222222 8.66666667
>
>
>sigma=sqrt(diag(V))
> sigma
[1] 3.374743 2.990726 3.473000 2.768875 2.943920
>
>##Ds: la métrique des inverses des écarts types##
>Ds=1/(sigma)*diag(5)
>
> Ds
         [,1]      [,2]      [,3]      [,4]      [,5]
[1,] 0.2963189 0.0000000 0.0000000 0.0000000 0.0000000
[2,] 0.0000000 0.3343669 0.0000000 0.0000000 0.0000000
[3,] 0.0000000 0.0000000 0.2879355 0.0000000 0.0000000
[4,] 0.0000000 0.0000000 0.0000000 0.3611576 0.0000000
[5,] 0.0000000 0.0000000 0.0000000 0.0000000 0.3396831
>
>## La matrice de corrélation R##
>R=Ds%*%V%*%Ds
>
> R
          [,1]        [,2]       [,3]       [,4]        [,5]
[1,] 1.00000000 0.982535729 0.22673193 0.49049826 0.011183835
[2,] 0.98253573 1.000000000 0.39669324 0.63398550 0.006309933
[3,] 0.22673193 0.396693238 1.00000000 0.95613107 0.038035989
[4,] 0.49049826 0.633985503 0.95613107 1.00000000 0.088601589
[5,] 0.01118384 0.006309933 0.03803599 0.08860159 1.000000000
```

Espace métrique des individus

Définition 0.4.1.

En physique, la distance entre deux points de l'espace se calcul facilement par la formule de pythagore, car les dimensions sont de même nature : ce sont des longueurs avec la même unité.

$$d^2 = (x_1^k - x_2^k)^2 + (x_1^j - x_2^j)^2$$

Il n'en pas de même en statistique où chaque dimension correspond à un caractère qui s'exprime avec son unité particulière : comment calculer la distance entre deux individus décrits par les trois caractères : âge, salaire, nombre d'enfant ?. La formule de Pythagore est alors aussi arbitraire qu'une autre. Si on veut donner des importances différentes à chaque caractère, on pourrait choisir une formule de type :

$$d^2 = (x_1^1 - x_2^1)^2 + (x_1^2 - x_2^2)^2 + \cdots + (x_1^p - x_2^p)^2,$$

ce qui revient à multiplier par $\sqrt{m_j}$ *chaque caractère, De plus la formule de Pythagore n'est valable que si les axes sont orthogonaux, mais en statistique c'est par pure convention que l'on représente les caractères sur des axes orthogonaux : on aurait pu prendre des axes obliques* θ :

La formule donnant la distiance fait alors intervenir en plus des carrés des différences de coordonnées les produits des différences :

$$d^2 = (x_1^k - x_2^k)^2 + (x_1^j - x_2^j)^2 - 2(x_1^k - x_2^k)(x_1^j - x_2^j)\cos\theta.$$

Ainsi,la distance d entre deux individus peut être définie de manière générale par :

$$d^2(t_1 - t_2) = \sum_{k=1}^{p}\sum_{j=1}^{p} m_{kj}(x_1^k - x_2^k)(x_1^j - x_2^j).$$

Soit en notant M *la matrice d'élements* m_{kj} :

$$d^2(t_1 - t_2) = (t_1 - t_2)'M(t_1 - t_2),$$

M *peut être n'importe quelle matrice symétrique définie positive. La formule de Pythagore revient à choisir pour* M *la matrice unité* $\mathbb{I}$.

Ceci revient à définir le produit scalaire de deux vecteurs t_1 *et* t_2 *dans l'espace des individus par :*

$$< t_1, t_2 >_M = t_1' M t_2,$$

on dit que l'on a muni l'espace des individus d'une structure euclidienne, la matrice M *s' appelle alors la métrique de l'espace. Le produit scalaire de* t_1 *par lui-même est noté* $\|t_1\|_M^2$ *et* $\|t_1\|_M^2$, *qui est analogue de la longueur du vecteur* t_1, *s'appelle la M-norme de* t_1.

0.5 Schéma de dualité

Définition 0.5.1.

Soit E un espace d'individu qui est isomorphe à $\mathbb{R}^p$ de base canomique $\{e_j/j = 1, \cdots, p\}$ et E^* son dual de base $\{e_j^*/j = 1, \cdots, p\}$ tel que $x_{ji} =< e_j^*, x_i >$ est la j^{me} composante de x_i par rapport à e_j.

De même pour F : un espace des variables qui est isomorphe à $\mathbb{R}^n$ de base canomique $\{f_i/i = 1, \cdots, n\}$ et son dual F^* de base $\{f_i^*/i = 1, \cdots, n\}$ telque $x_{ji} =< f_i^*, x^j >$ est la i^{me} composante de x^j par rapport à f_i.

Rappel

On a vu dans la définition (**??**), (**??**) que, si on a un produit scalaire f dans un espace euclidien de dimension n qui est associée à M et qui est définie par :

$$\begin{aligned} f_x : E &\longrightarrow \mathbb{R} \qquad\qquad (f_x \in E^*) \\ y &\longrightarrow f_x(y). \end{aligned}$$

Alors la matrice associée à φ :

$$\begin{aligned} \varphi : E &\longrightarrow E^* \qquad\qquad \text{n'est autre que}\, M \\ y &\longrightarrow \varphi(x) = f_x. \end{aligned}$$

Donc si on pose E l'espace des individus muni d'une métrique M et F l'espace des variables muni d'une métrique N alors les métriques M et N sont des matrices associées aux applications suivantes :

$$\begin{array}{rcl} M:\, E & \xrightarrow{M} & E^* \\ (x_i, x_{i'}) & \longrightarrow & M(x_i, x_i') \end{array}$$

et

$$\begin{array}{rcl} N:\, E & \xrightarrow{N} & F^* \\ (x^j, x^{j'}) & \longrightarrow & N(x^j, x^{j'}) \end{array}$$

0.5.1 Construction des applications X et X'

– L'application X :

$$\begin{aligned} X : F^* &\longrightarrow E \\ f_i^* &\longrightarrow X(f_i^*) = x_i. \end{aligned}$$

Cette application X est associée à la matrice X qui est le tableau des données.

– L'application X' :

$$\begin{aligned} X' : E^* &\longrightarrow F \\ f_j^* &\longrightarrow X'(e_j^*) = x^j. \end{aligned}$$

Cette application X' est associée à la matrice X qui est le transposé de tableau des données.

0.5.2 La métrique W dans F^*

On vas construire la métrique W tel qu'on garde les mêmes distances, c'est-à-dire :

$$\|f_i^* - f_i'^*\|_W^2 = \|x_i - x_i'\|_M^2,$$

on a :
$W(f_i^*, f_{i'}^*) = M(x_i, x_{i'}) \;/\; X(f_i^*) = x_i$
$\qquad = M(X(f_i^*), X(f_{i'}^*))$
$\qquad =< MXf_i^*, Xf_{i'}^* >$
$\qquad = X'MX(f_i^*, f_{i'}^*)$
d'où $W = X'MX$.

Propriété 0.5.1.

$W(\cdot,\cdot)$ est une forme bilinèaire sur F^* symétrique semi définie positife.
$W(f_i^*, f_{i'}^*) = 0 \Rightarrow M(x_i, x_{i'}) = 0$
$\qquad x_i = 0$
$\qquad X(f_i^*) = 0$
$\qquad KerX = \{0\}$
$\qquad X\, est\, injective,$
si X est injective alors W est définie.
W est un produit scalaire qui concerve les normes de manière générale W est un semi produit scalaire vérifiant $W(f_i^*, f_{i'}^*) = M(x_i, x_{i'})$ $\forall i, i'$.

0.5.3 La métrique V dans E^*

On fait les mêmes démarches pour construire la métrique V telque on grade les mêmes distances :
$V(e_j^*, e_{j'}^*) = N(x^j, x^{j'}) \;/\; X'(e_j^*) = x^j$
$\qquad = N(X'(e_j^*), X'(e_{j'}^*))$
$\qquad =< NX'e_j^*, X'e_{j'}^* >$
$\qquad = XNX'(e_j^*, e_{j'}^*)$
d'ou $V = XMX'$.

Propriété 0.5.2.

$V(\cdot,\cdot)$ est une forme bilinèaire sur E^* symétrique semi définie positife.
$V(e_j^*, f_{j'}^*) = 0 \Rightarrow N(x^j, x^{j'}) = 0$
$\qquad x^j = 0$
$\qquad X'(e_j^*) = 0$
$\qquad KerX' = \{0\}$
$\qquad X'\, est\, injective.$
Si X' est injective alors V est définie.

V est un produit scalaire qui concerve les normes de manière général V est un semi produit scalaire vérifiant $V(e^*_j, e^*_{j'}) = N(x^j, x^{j}{'})\forall j, j'$.

On résume ces propriétés graphiqement dans la figure (2.3)

0.6 L'inertie et moment d'inertie

0.6.1 L'inertie d'un point

Définition 0.6.1.

On applle inertie I_a de nuage η des individus sur E qui a pour métrique M par rapport à un point a :

$$I_a = \sum_{i=1}^{n} p_i \|x_i - a\|^2_M.$$

I_a : mesure la proximitée de η par rapport à $a \in E$.

0.6.2 L'inertie total I_g d'un point

Soit $\eta = \{(x_i, p_i)\}$ le nuage de points. On note $\mu = \{(y_i, p_i)\}$ le nuage centré, où on a ramenté le centre de gravité à l'origine du repère.

Définition 0.6.2.

On applle inertie totale de nuage des individus et on note I_g, la moyenne pondérée des carrés de toutes les distances entre les individus :

$$I_g = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} p_i p_j \|x_i - x_j\|^2_M.$$

Cas particulier

$$I_g = \sum_{i=1}^{n} p_i \|x_i - g\|^2_M I_g = \sum_{i=1}^{n} p_i \|y_i\|^2_M.$$

C'est-à-dire l'inertie correspond à la moyenne des carrés des distances des points au centre de gravité.
L'inertie mesure la dispersion des points individus autour du centre de gravité g, elle est parfois appelée variance du nuage. L'inertie du nuage η est évidemment égale à l'inertie du nuage centré μ.

Proposition 0.6.1.

$$I_a = I_g + \|g - a\|^2_M.$$

Preuve 0.6.1.

$$
\begin{aligned}
I_a &= \sum_{i=1}^{n} p_i \|x_i - a\|_M^2 \\
&= \sum_{i=1}^{n} p_i \|(x_i - g) + (g - a)\|_M^2 \\
&= \sum_{i=1}^{n} p_i [\|x_i - g\|_M^2 + \|g - a\|_M^2 + 2M(x_i - g, g - a)] \\
&= \sum_{i=1}^{n} p_i \|x_i - g\|_M^2 + \sum_{i=1}^{n} p_i \|g - a\|_M^2 + 2\sum_{i=1}^{n} p_i M(x_i - g, g - a) \;\; telque \sum_{i=1}^{n} p_i = 1 \\
&= \sum_{i=1}^{n} p_i \|x_i - g\|_M^2 + \|g - a\|_M^2 + 2M\left(\sum_{i=1}^{n} p_i x_i - g, g - a\right), M(0, g - a) = 0 \\
&= I_g + \|g - a\|_M^2.
\end{aligned}
$$

Remarque 0.6.1.

On vois que la proximitée est minimale si $a = g$, c'est-à-dire :

$$
\begin{aligned}
I_a = I_g &\Longleftrightarrow \|g - a\|_M^2 = 0 \\
&\Longleftrightarrow g = a,
\end{aligned}
$$

donc g est le point de E le plus proche de η au sens de l'inertie :

$$I_g = \sum_{i=1}^{n} p_i \|x_i - g\|_M^2 = \sum_{i=1}^{n} p_i \|y_i\|_M^2.$$

0.6.3 Expression matricielle de l'inertie I_g

Dans la suite de chapitre, on supposera que le nuage est centré.

Propriété 0.6.1.

$$tr\left(\sum_{i=1}^{n} p_i x_i x_i' M\right) = tr(VM).$$

Preuve 0.6.2.

$I_g = \sum_{i=1}^{n} \|x_i\|_M^2$, *car le nuage est centré*

on a, $\|x_i\|_M^2 = M(x_i, x_i) = x_i' M x_i$, $tr(x_i' M x_i) = x_i' M x_i$ *car c'est un scalaire,*
$x_i' M x_i = x_i' x_i M$ *alors* $I_g = \sum_{i=1}^{n} \|x_i\|_M^2 = tr\left(\sum_{i=1}^{n} p_i x_i' x_i M\right) = tr\left(\sum_{i=1}^{n} p_i x_i x_i' M\right)$

On sait que $\sum_{i=1}^{n} p_i x_i x_i'$ *est la matrice de variance covariance V.*

D'où le résultat : $tr\left(\sum_{i=1}^{n} p_i x_i x_i' M\right) = tr(VM)$.

Remarque 0.6.2.

VM c'est un endomorphisme de E qui est défini par :

$VM : E \xrightarrow{M} E^* \xrightarrow{V} E$ *(car V et M sont des applications linéaires)*

0.6.4 Moment d'inertie de η par rapport un sous espace vectoriel

Définition 0.6.3.

Soit W un espace vectoriel de E et $W^{\perp}$ le supplimentaire M -orthogonal de W dans E tel que : $E = W \oplus W^{\perp}, W \cap W^{\perp} = \{0_E\}$ c'est-à-dire :

$\forall x_i \in E,\ \exists!\, \alpha_i \in W,\ \exists!\, \beta_i \in W^{\perp} :$ *telque* $x_i = \alpha_i + \beta_i.$

Proximité entre x_i et W

Soit $\|\beta_i\|_M = \|x_i - \alpha_i\|_M$, car $\beta_i = x_i - \alpha_i$.

Pour calculer la proximité entre le nuage η et W il suffit de calculer la proximité de chaque x_i dans E et sa projection α_i alors : $I_W = \sum_{i=1}^{n} p_i \|x_i - \alpha_i\|_M^2$.

Proposition 0.6.2.

$$I_g = I_{W^{\perp}} + I_W.$$

Preuve 0.6.3.

$$\begin{aligned}
I_g &= \sum_{i=1}^{n} p_i \|x_i\|_M^2, \ \textit{on remplace } x_i \textit{ par } \alpha_i + \beta_i \\
&= \sum_{i=1}^{n} p_i \|\alpha_i + \beta_i\|_M^2 \\
&= \sum_{i=1}^{n} p_i M(\alpha_i + \beta_i, \alpha_i + \beta_i) \\
&= \sum_{i=1}^{n} p_i M(\alpha_i, \alpha_i) + \sum_{i=1}^{n} p_i M(\beta_i, \beta_i) + 2\sum_{i=1}^{n} p_i \underbrace{M(\alpha_i, \beta_i)}_{=0} \\
&= \sum_{i=1}^{n} p_i \|\alpha_i\|_M^2 + \sum_{i=1}^{n} p_i \|\beta_i\|_M^2, \alpha_i = x_i - \beta_i \ \textit{et} \ \beta_i = x_i - \alpha_i \\
&= \sum_{i=1}^{n} p_i \|x_i - \beta_i\|_M^2 + \sum_{i=1}^{n} p_i \|x_i - \alpha_i\|_M^2 \\
&= I_{W^{\perp}} + I_W.
\end{aligned}$$

0.6.5 Moment d'inertie de η par rapport à un sous espace affine

Définition 0.6.4.

Soit W_1 un sous espace affine de E parallèle à un sous espace vectoriel W de E c'est-à-dire :

$$W_1 = \{y \in E, \exists x \in W \textit{ tel que } y = x + a\} = W + a, a \in E.$$

Remarque 0.6.3.

- $\forall x_i \in E, \exists! \alpha_i \in W, \exists! \beta_i \in W^{\perp}$ *tel que* $x_i = \alpha_i + \beta_i$.
- $W^{\perp} \cap W_1 = \{a\}$.

Poximité entre x_i et W_1

Soit :

$$I_{W_1} = \sum_{i=1}^{n} p_i \|\beta_i - a\|_M^2,$$

on a $x_i = \alpha_i + \beta_i = (a - \alpha_i) + (\beta_i - a)$

$$\textit{donc}\, I_{W_1} = \sum_{i=1}^{n} p_i \|x_i - (\alpha_i + a)\|_M^2$$

Proposition 0.6.3.

$I_{W_1} = I_W + \|a\|_M^2$.

Preuve 0.6.4.

$$I_{W_1} = \sum_{i=1}^{n} p_i \|\beta_i - a\|_M^2$$

$$= \sum_{i=1}^{n} p_i \|x_i - (a + \alpha_i)\|_M^2$$

$$= \sum_{i=1}^{n} p_i M(\beta_i - a, \beta_i - a)$$

$$= \sum_{i=1}^{n} p_i M(\beta_i, \beta_i) + 2\sum_{i=1}^{n} p_i M(\beta_i, a) + \sum_{i=1}^{n} p_i M(a, a)$$

$$= \sum_{i=1}^{n} p_i M(\beta_i, \beta_i) + \sum_{i=1}^{n} p_i M(a, a), M(\beta_i, a) = 0 \ car \, \beta_i \perp a$$

$$= \sum_{i=1}^{n} p_i \|x_i - \alpha\|_M^2 + \|a\|_M^2$$

$$= I_W + \|a\|_M^2.$$

Remarque 0.6.4.

- *si* $\eta \in W \Rightarrow I_W = 0.$
- $I_{W_1} \geq I_W.$

Remarque 0.6.5.

Le centre de gravité g est contenu dans le sous espace affine optimal car est un espace vectoriel.

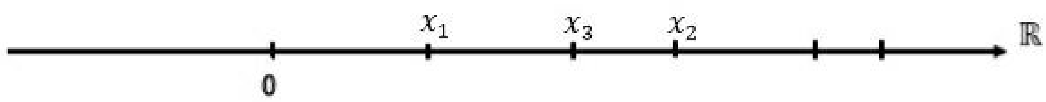

FIGURE 1 – Représentation de E dans $\mathbb{R}$

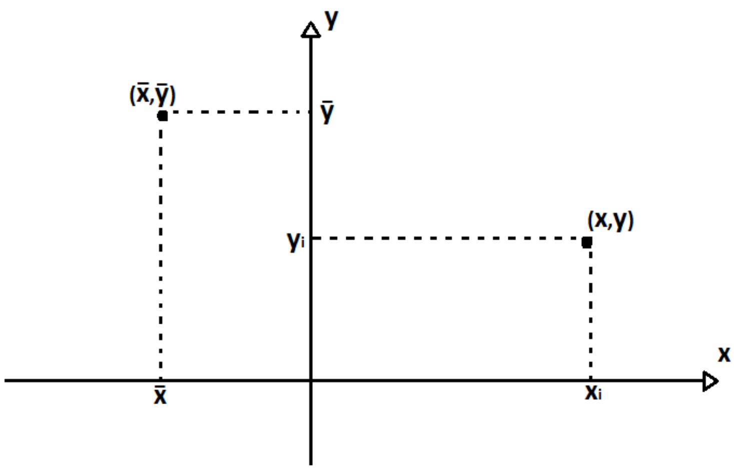

FIGURE 2 – Représentation de E dans $\mathbb{R}^2$

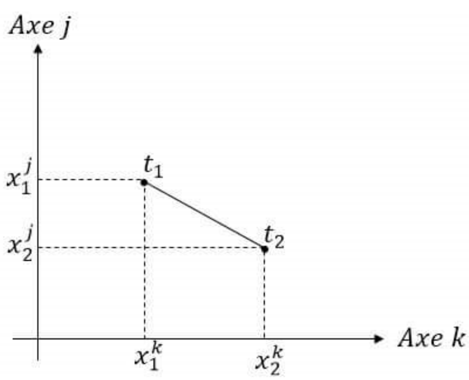

FIGURE 3 – Distance entre individus

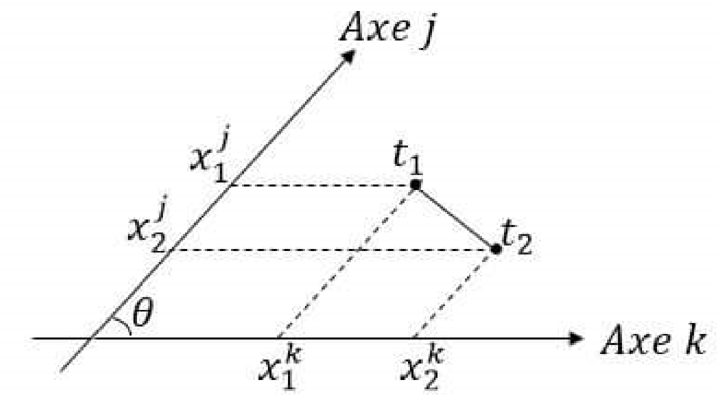

FIGURE 4 – Représentation graphique : cas axes sont obliques

$$\begin{array}{ccc} F^* & \xrightarrow{X} & E \approx \mathbb{R}^p \\ N \uparrow \downarrow W & & V \uparrow \downarrow M \\ \mathbb{R}^n \approx F & \xleftarrow{X'} & E^* \end{array}$$

FIGURE 5 – Construction de schéma de dualité

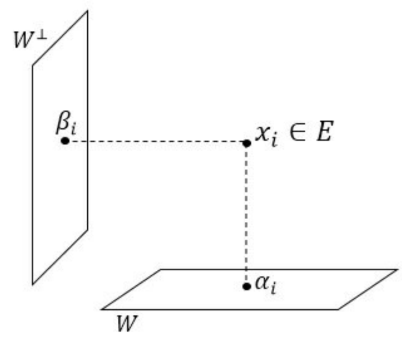

FIGURE 6 – Proximité graphique des individus par raport à W

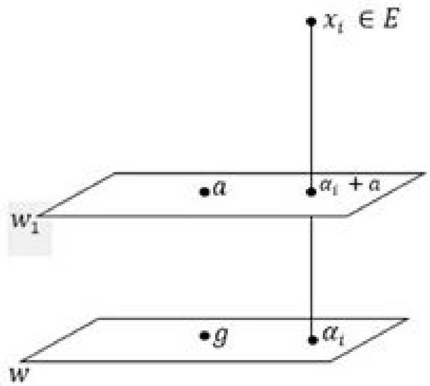

FIGURE 7 – Représentation affine