

# Analyse des données

**Jaouad Madkour**

[jmadkour@uae.ac.ma](mailto:jmadkour@uae.ac.ma)

Faculté des Sciences Juridiques, Économiques et Sociales de Tanger  
Département des Sciences économiques et Gestion

2 avril 2026



[jmadkour.org](https://jmadkour.org)





## PROBLÉMATIQUE

- Il est impossible de représenter un objet 3D (ou plus) sur un plan 2D sans le déformer.
- Toute déformation d'un objet entraîne une perte d'information.
- Comment déformer un objet en perdant le moins d'information possible ?

## EN ÉCONOMIE ET GESTION

- Les **individus** sont caractérisés par de nombreuses **variables**.
- Les individus sont des objets **multidimensionnels**.
- Avec la statistique uni-variée, on peut juste décrire une seule variable !
- Avec la statistique bi-variée, on peut juste décrire la corrélation linéaire entre deux variables !
- Comment décrire des individus "multi-variés" avec un ensemble réduit de variables ?

## EXEMPLE

Lors d'une étude statistique sur les clients d'une société, les informations collectées sur chaque individu ont été enregistrées dans la base de données suivante :

ID	Age	Profession	Résidence	Fréquence	Montant	Satisfaction	Fidèle
1	28	Étudiant	Tanger	4	120	4	Non
2	45	Cadre	Assilah	2	320	5	Oui
3	36	Artisan	Tétouan	4	240	3	Non
4	62	Retraité	Tanger	3	150	4	Oui
5	24	Employé	Larache	1	110	2	Non
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

- **Problème** : Les individus sont multidimensionnels, ils sont représentés par 7 variables.
- **Question** : Comment établir le profile des clients avec deux "variables" <sup>a</sup> uniquement ?
- **Réponse** : Appliquer les outils de l'analyse des données.

---

a. Il s'agit de variables virtuelles appelées **facteurs**.

## OBJECTIFS DU COURS

- Revoir les éléments de la statistique descriptive uni-variée (moyenne, variance, écart-type...).
- Approfondir les concepts de la statistique descriptive bi-variée (covariance, corrélation...).
- Traduire géométriquement des concepts clés de l'algèbre linéaire (vecteur, matrice...).
- Étudier l'Analyse en Composantes Principales, l'Analyse Factorielle des Correspondances...
- Appliquer l'analyse des données dans le domaine de la gestion des entreprises.

## PRÉREQUIS

- Essentiellement, la statistique descriptive uni- et bi-variée.
- Accessoirement, l'algèbre linéaire.

## RÉFÉRENCES BIBLIOGRAPHIQUES



## SOMMAIRE

CHAPITRE 1 : APPROFONDISSEMENTS DE LA STATISTIQUE DESCRIPTIVE

CHAPITRE 2 : GÉOMÉTRIE DE L'ALGÈBRE LINÉAIRE

CHAPITRE 3 : ANALYSE EN COMPOSANTES PRINCIPALES

**1. Approfondissements de la statistique descriptive**

2. Géométrie de l'algèbre linéaire

3. Analyse en Composantes Principales

## CHAPITRE 1

### APPROFONDISSEMENTS DE LA STATISTIQUE DESCRIPTIVE

SECTION 1.1 : DESCRIPTION D'UNE SEULE VARIABLE STATISTIQUE

SECTION 1.2 : LIAISON ENTRE DEUX VARIABLES STATISTIQUES

SECTION 1.3 : VERS LE CAS MULTIVARIÉ

## SECTION 1.1

### DESCRIPTION D'UNE SEULE VARIABLE STATISTIQUE

- DESCRIPTION D'UNE VARIABLE QUANTITATIVE
- DESCRIPTION D'UNE VARIABLE QUALITATIVE

## TERMINOLOGIE

- Les individus sont étudiés selon un caractère donné  $X$  (âge, profession, lieu de résidence, etc.)
- Ce caractère est appelé **variable statistique** et peut prendre plusieurs **modalités**  $x_1, x_2, \dots, x_k$
- Un nombre  $n_i$  d'individus possèdent la modalité  $x_i$ . Ainsi,  $n_i$  est l'**effectif** de  $x_i$ .
- $N$  est l'**effectif total** de tous les individus : i.e.  $N = n_1 + n_2 + \dots + n_k$ .
- Une proportion  $f_i = \frac{n_i}{N}$  d'individus possèdent la modalité  $x_i$ . Ainsi,  $f_i$  est la **fréquence** de  $x_i$ .

## APPLICATION

$x_i$	20	21	22	23	24	$\Sigma$
$n_i$	10	20	40	20	10	100
$f_i$	0.1	0.2	0.4	0.2	0.1	1

## DESCRIPTION D'UNE VARIABLE QUANTITATIVE

### VARIABLE QUANTITATIVE DISCRÈTE

$x_i$	$x_1$	$x_2$	$\dots$	$x_k$	$\Sigma$
$n_i$	$n_1$	$n_2$	$\dots$	$n_k$	$N$

### VARIABLE QUANTITATIVE CONTINUE

$[x_i, x_{i+1}[$	$[x_1, x_2[$	$[x_2, x_3[$	$\dots$	$[x_k, x_{k+1}[$	$\Sigma$
$n_i$	$n_1$	$n_2$	$\dots$	$n_k$	$N$

- $[x_i, x_{i+1}[$  : la classe  $i$  de la variable statistique  $X$  ;
- $c_i = (x_i + x_{i+1})/2$  : le centre de la classe  $[x_i, x_{i+1}[$ .

## DESCRIPTION NUMÉRIQUE D'UNE VARIABLE QUANTITATIVE

- MOYENNE ARITHMÉTIQUE
- VARIANCE ET ÉCART-TYPE
- FRÉQUENCE

## MOYENNE ARITHMÉTIQUE

**Variable discrète**

$$\bar{x} = \frac{1}{N} \sum_{i=1}^k n_i x_i$$

**Variable continue**

$$\bar{x} = \frac{1}{N} \sum_{i=1}^k n_i c_i$$

## VARIANCE ET ÉCART-TYPE

**Variable discrète**

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^k n_i (x_i - \bar{x})^2 = \overline{x^2} - (\bar{x})^2$$

$$\sigma = \sqrt{\sigma^2}$$

**Variable continue**

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^k n_i (c_i - \bar{x})^2 = \overline{c^2} - (\bar{x})^2$$

$$\sigma = \sqrt{\sigma^2}$$

## FRÉQUENCE

**Variable discrète**

$$f_i = \frac{n_i}{N}$$

**Variable continue**

$$f_i = \frac{n_i}{N}$$

## APPLICATION

### Variable discrète

$x_i$	20	21	22	23	24	$\Sigma$
$n_i$	10	20	40	20	10	100
$f_i$	0.1	0.2	0.4	0.2	0.1	1

$$\bar{x} = \frac{1}{N} \sum_{i=1}^5 n_i x_i = 22$$

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^5 n_i (x_i - \bar{x})^2 = 1.2$$

$$\sigma = \sqrt{\sigma^2} \approx 1.1$$

### Variable continue

$x_i$	[20,21[	[21,22[	[22,23[	[23,24]	$\Sigma$
$c_i$	20.5	21.5	22.5	23.5	-
$n_i$	10	30	40	20	100
$f_i$	0.1	0.3	0.4	0.2	1

$$\bar{x} = \frac{1}{N} \sum_{i=1}^4 n_i c_i = 22.2$$

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^4 n_i (c_i - \bar{x})^2 = 0.81$$

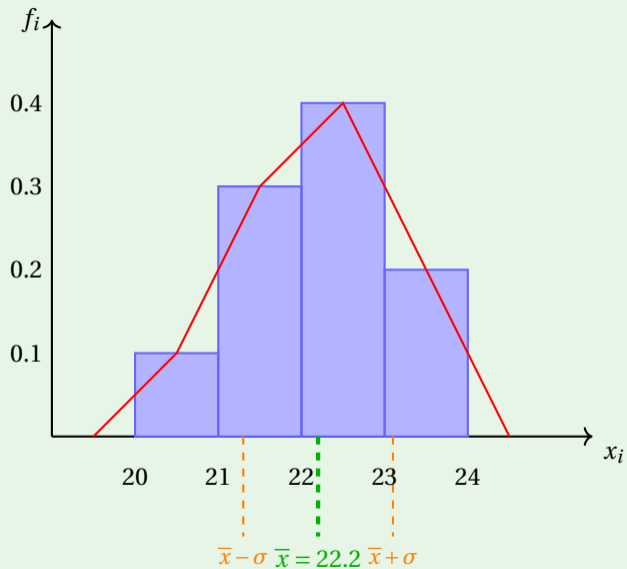
$$\sigma = \sqrt{\sigma^2} = 0.9$$

## RÉSUMÉ GRAPHIQUE D'UNE VARIABLE QUANTITATIVE

- DIAGRAMME EN BÂTONS
- HISTOGRAMME
- POLYGONE DES FRÉQUENCES



## HISTOGRAMME - POLYGONE DES FRÉQUENCES



## DESCRIPTION D'UNE VARIABLE QUALITATIVE

### VARIABLE QUALITATIVE NOMINALE

Catégories	Catégorie 1	Catégorie 2	...	Catégorie $k$	$\Sigma$
$n_i$	$n_1$	$n_2$	...	$n_k$	$N$

### VARIABLE QUALITATIVE ORDINALE

Catégories	1 <sup>ère</sup> catégorie	2 <sup>ème</sup> catégorie	...	$k^{\text{ème}}$ catégorie	$\Sigma$
$n_i$	$n_1$	$n_2$	...	$n_k$	$N$

## DESCRIPTION NUMÉRIQUE D'UNE VARIABLE QUALITATIVE

- MOYENNE ARITHMÉTIQUE
- VARIANCE ET ÉCART-TYPE
- FRÉQUENCE

## APPLICATION

Mentions	Passable	Assez bien	Bien	Très bien	Excellent	$\Sigma$
$n_i$	10	20	40	20	10	100
$f_i$	0.1	0.2	0.4	0.2	0.1	1

Spécialités	Économie	Finance	Marketing	Logistique	Comptabilité	$\Sigma$
$n_i$	10	20	40	20	10	100
$f_i$	0.1	0.2	0.4	0.2	0.1	1

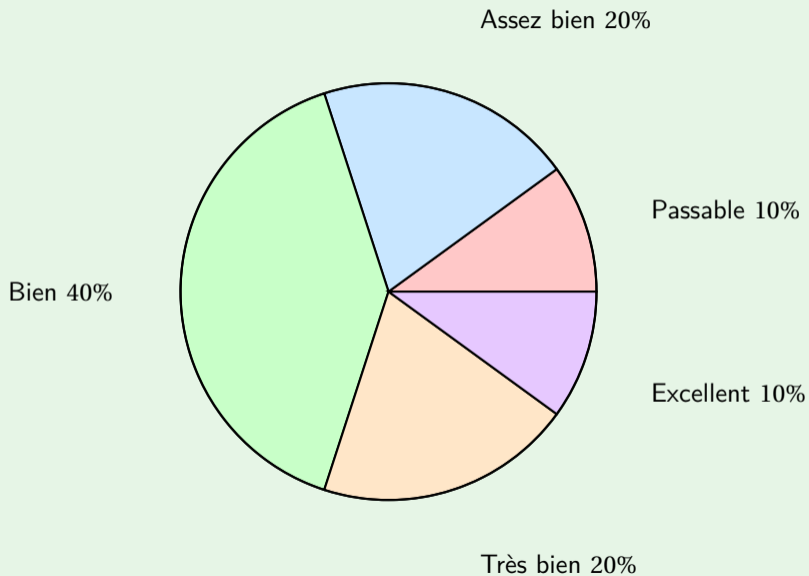
## RÉSUMÉ GRAPHIQUE D'UNE VARIABLE QUALITATIVE

- DIAGRAMME EN BARRES
- DIAGRAMME EN SECTEURS

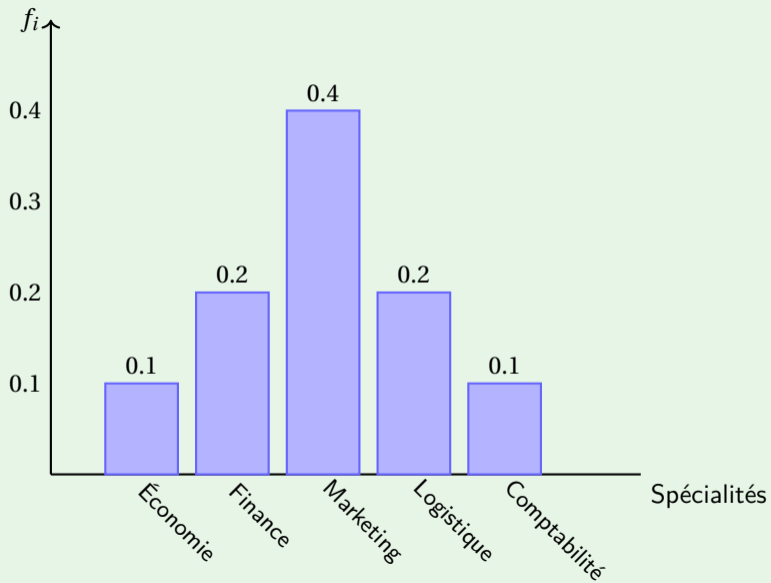
## DIAGRAMME EN BARRES



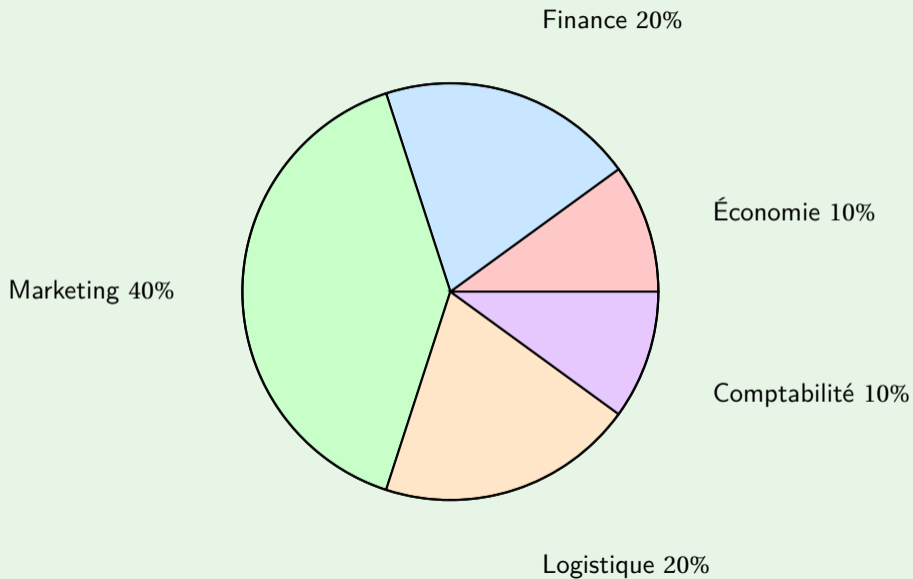
## DIAGRAMME EN SECTEURS



## DIAGRAMME EN BARRES



## DIAGRAMME EN SECTEURS



## SECTION 1.2

### LIAISON ENTRE DEUX VARIABLES STATISTIQUES

- DISTRIBUTIONS MARGINALES
- DISTRIBUTIONS CONDITIONNELLES
- DISTRIBUTION JOINTE

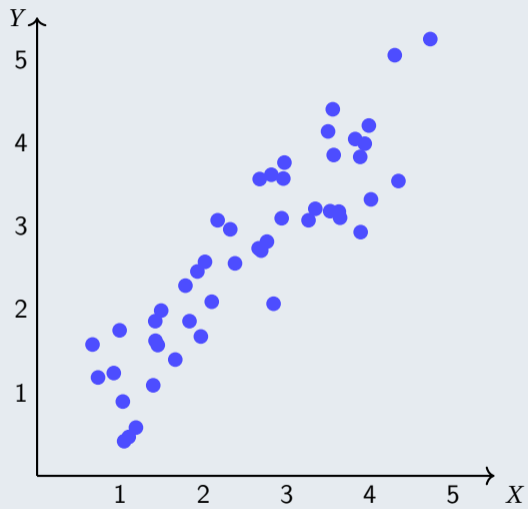
## TABLEAU DE CONTINGENCE

$(x_i, y_j)$	$y_1$	$y_2$	$\dots$	$y_l$	$\Sigma$
$x_1$	$n_{11}$	$n_{12}$	$\dots$	$n_{1l}$	$n_{1.}$
$x_2$	$n_{21}$	$n_{22}$	$\dots$	$n_{2l}$	$n_{2.}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
$x_k$	$n_{k1}$	$n_{k2}$	$\dots$	$n_{kl}$	$n_{k.}$
$\Sigma$	$n_{.1}$	$n_{.2}$	$\dots$	$n_{.l}$	$n_{..}$

## PARFOIS

$(x_i, y_i)$	$(x_1, y_1)$	$(x_2, y_2)$	$\dots$	$(x_k, y_k)$	$\Sigma$
$n_i$	$n_1$	$n_2$	$\dots$	$n_k$	$n$

## NUAGE DE POINTS



## DISTRIBUTION MARGINALE DE X

$(x_i, y_j)$	$y_1$	$y_2$	$\dots$	$y_l$	$\Sigma$
$x_1$	$n_{11}$	$n_{12}$	$\dots$	$n_{1l}$	$n_{1.}$
$x_2$	$n_{21}$	$n_{22}$	$\dots$	$n_{2l}$	$n_{2.}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
$x_k$	$n_{k1}$	$n_{k2}$	$\dots$	$n_{kl}$	$n_{k.}$
$\Sigma$	$n_{.1}$	$n_{.2}$	$\dots$	$n_{.l}$	$n_{..}$

$x_i$	$x_1$	$x_2$	$\dots$	$x_k$	$\Sigma$
$n_{i.}$	$n_{1.}$	$n_{2.}$	$\dots$	$n_{k.}$	$n_{..}$

## MOYENNE MARGINALE ET VARIANCE MARGINALE DE X

$$\bar{\bar{x}} = \frac{1}{n_{..}} \sum_{i=1}^k n_{i.} x_i$$

$$\sigma_x^2 = \frac{1}{n_{..}} \sum_{i=1}^k n_{i.} (x_i - \bar{\bar{x}})^2 = \overline{x^2} - (\bar{\bar{x}})^2$$

## DISTRIBUTION MARGINALE DE Y

$(x_i, y_j)$	$y_1$	$y_2$	$\dots$	$y_l$	$\Sigma$
$x_1$	$n_{11}$	$n_{12}$	$\dots$	$n_{1l}$	$n_{1.}$
$x_2$	$n_{21}$	$n_{22}$	$\dots$	$n_{2l}$	$n_{2.}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
$x_k$	$n_{k1}$	$n_{k2}$	$\dots$	$n_{kl}$	$n_{k.}$
$\Sigma$	$n_{.1}$	$n_{.2}$	$\dots$	$n_{.l}$	$n_{..}$

$y_j$	$y_1$	$y_2$	$\dots$	$y_l$	$\Sigma$
$n_{.j}$	$n_{.1}$	$n_{.2}$	$\dots$	$n_{.l}$	$n_{..}$

## MOYENNE MARGINALE ET VARIANCE MARGINALE DE Y

$$\bar{y} = \frac{1}{n_{..}} \sum_{j=1}^l n_{.j} y_j$$

$$\sigma_y^2 = \frac{1}{n_{..}} \sum_{j=1}^l n_{.j} (y_j - \bar{y})^2 = \overline{y^2} - (\bar{y})^2$$

## DISTRIBUTION CONDITIONNELLE DE $X|(Y = y_2)$

$(x_i, y_j)$	$y_1$	$y_2$	$\dots$	$y_l$	$\Sigma$
$x_1$	$n_{11}$	$n_{12}$	$\dots$	$n_{1l}$	$n_{1.}$
$x_2$	$n_{21}$	$n_{22}$	$\dots$	$n_{2l}$	$n_{2.}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
$x_k$	$n_{k1}$	$n_{k2}$	$\dots$	$n_{kl}$	$n_{k.}$
$\Sigma$	$n_{.1}$	$n_{.2}$	$\dots$	$n_{.l}$	$n_{..}$

$x_i$	$x_1$	$x_2$	$\dots$	$x_k$	$\Sigma$
$n_{i2}$	$n_{12}$	$n_{22}$	$\dots$	$n_{k2}$	$n_{.2}$

## MOYENNE CONDITIONNELLE ET VARIANCE CONDITIONNELLE DE $X$

$$\bar{x}_2 = \frac{1}{n_{.2}} \sum_{i=1}^k n_{i2} x_i$$

$$\sigma_{x,2}^2 = \frac{1}{n_{.2}} \sum_{i=1}^k n_{i2} (x_i - \bar{x}_2)^2$$

## DISTRIBUTION CONDITIONNELLE DE $Y|(X = x_2)$

$(x_i, y_j)$	$y_1$	$y_2$	$\dots$	$y_l$	$\Sigma$
$x_1$	$n_{11}$	$n_{12}$	$\dots$	$n_{1l}$	$n_{1.}$
$x_2$	$n_{21}$	$n_{22}$	$\dots$	$n_{2l}$	$n_{2.}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
$x_k$	$n_{k1}$	$n_{k2}$	$\dots$	$n_{kl}$	$n_{k.}$
$\Sigma$	$n_{.1}$	$n_{.2}$	$\dots$	$n_{.l}$	$n_{..}$

$y_j$	$y_1$	$y_2$	$\dots$	$y_l$	$\Sigma$
$n_{2j}$	$n_{21}$	$n_{22}$	$\dots$	$n_{2l}$	$n_{2.}$

## MOYENNE CONDITIONNELLE ET VARIANCE CONDITIONNELLE DE $Y$

$$\bar{y}_2 = \frac{1}{n_{2.}} \sum_{j=1}^l n_{2j} y_j$$

$$\sigma_{y,2}^2 = \frac{1}{n_{2.}} \sum_{j=1}^l n_{2j} (y_j - \bar{y}_2)^2$$

## DISTRIBUTION JOINTE DU COUPLE (X, Y)

$(x_i, y_j)$	$y_1$	$y_2$	$\dots$	$y_l$	$\Sigma$
$x_1$	$n_{11}$	$n_{12}$	$\dots$	$n_{1l}$	$n_{1.}$
$x_2$	$n_{21}$	$n_{22}$	$\dots$	$n_{2l}$	$n_{2.}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
$x_k$	$n_{k1}$	$n_{k2}$	$\dots$	$n_{kl}$	$n_{k.}$
$\Sigma$	$n_{.1}$	$n_{.2}$	$\dots$	$n_{.l}$	$n_{..}$

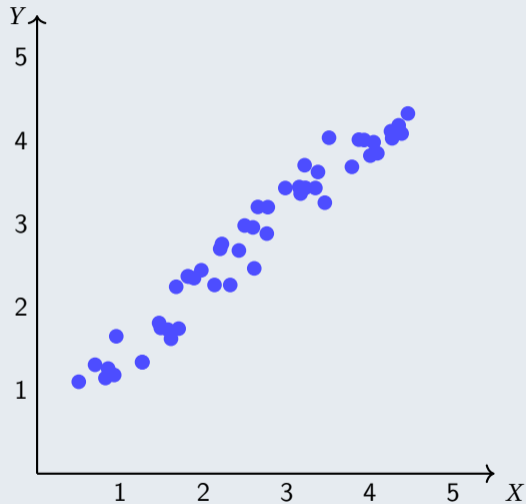
## MOYENNE JOINTE ET COVARIANCE DU COUPLE (X, Y)

$$\overline{xy} = \frac{1}{n_{..}} \sum_{i=1}^k \sum_{j=1}^l n_{ij} x_i y_j \quad \text{cov}(X, Y) = \frac{1}{n_{..}} \sum_{i=1}^k \sum_{j=1}^l n_{ij} (x_i - \bar{x}) (y_j - \bar{y}) = \overline{xy} - \bar{x} \cdot \bar{y} \in \mathbb{R}$$

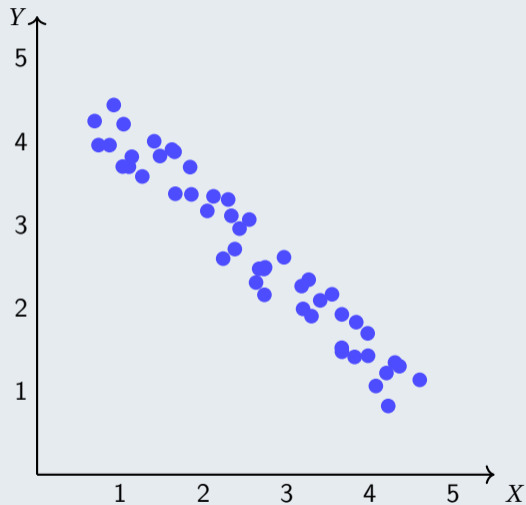
## COEFFICIENT DE CORRÉLATION LINÉAIRE ENTRE X ET Y

$$\rho(X, Y) = \frac{\text{cov}(X, Y)}{\sigma_x \sigma_y} \quad -1 \leq \rho(X, Y) \leq 1$$

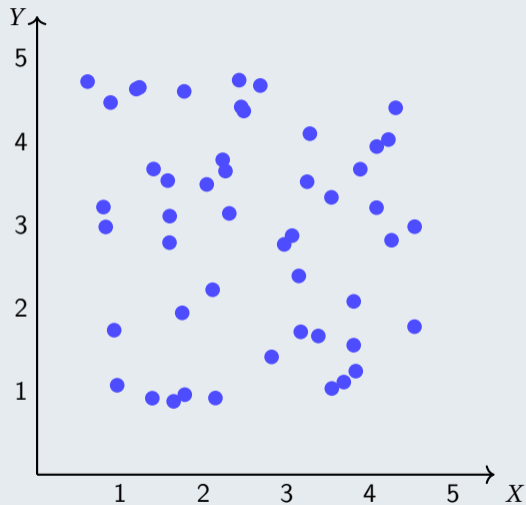
## COVARIANCE POSITIVE (CORRÉLATION POSITIVE)



## COVARIANCE NÉGATIVE (CORRÉLATION NÉGATIVE)



## COVARIANCE NULLE (CORRÉLATION NULLE)



## MATRICE DES VARIANCES-COVARIANCES

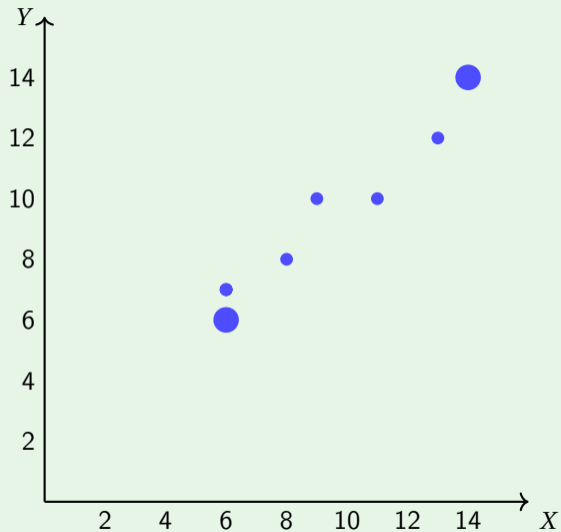
$cov(X, Y)$	$X$	$Y$
$X$	$cov(X, X) = \sigma_x^2$	$cov(X, Y)$
$Y$	$cov(X, Y)$	$cov(Y, Y) = \sigma_y^2$

## MATRICE DES CORRÉLATIONS

$\rho(X, Y)$	$X$	$Y$
$X$	$\rho(X, X) = 1$	$\rho(X, Y)$
$Y$	$\rho(X, Y)$	$\rho(Y, Y) = 1$

## APPLICATION

Élève	Math	Phys
A	6	6
B	8	8
C	6	6
D	14	14
E	14	14
F	11	10
G	6	7
H	13	12
I	9	10



## APPLICATION

Élève	Math	Phys
A	6	6
B	8	8
C	6	6
D	14	14
E	14	14
F	11	10
G	6	7
H	13	12
I	9	10

$\Rightarrow$

(X, Y)	6	7	8	10	12	14	$\Sigma$
6	2	1	0	0	0	0	3
8	0	0	1	0	0	0	1
9	0	0	0	1	0	0	1
11	0	0	0	1	0	0	1
13	0	0	0	0	1	0	1
14	0	0	0	0	0	2	2
$\Sigma$	2	1	1	2	1	2	9

X : Notes de mathématiques

Y : Notes de physique

## DISTRIBUTION MARGINALE DE X

(X, Y)	6	7	8	10	12	14	Σ
6	2	1	0	0	0	0	3
8	0	0	1	0	0	0	1
9	0	0	0	1	0	0	1
11	0	0	0	1	0	0	1
13	0	0	0	0	1	0	1
14	0	0	0	0	0	2	2
Σ	2	1	1	2	1	2	9

$x_i$	6	8	9	11	13	14	Σ
$n_{i.}$	3	1	1	1	1	2	9

## MOYENNE MARGINALE ET VARIANCE MARGINALE DE X

$$\bar{\bar{x}} = \frac{1}{n_{..}} \sum_{i=1}^6 n_{i.} x_i = 9.67$$

$$\sigma_x^2 = \frac{1}{n_{..}} \sum_{i=1}^6 n_{i.} (x_i - \bar{\bar{x}})^2 = \overline{x^2} - (\bar{\bar{x}})^2 = 10.48$$

## DISTRIBUTION MARGINALE DE Y

(X, Y)	6	7	8	10	12	14	$\Sigma$
6	2	1	0	0	0	0	3
8	0	0	1	0	0	0	1
9	0	0	0	1	0	0	1
11	0	0	0	1	0	0	1
13	0	0	0	0	1	0	1
14	0	0	0	0	0	2	2
$\Sigma$	2	1	1	2	1	2	9

$y_j$	6	7	8	10	12	14	$\Sigma$
$n_{.j}$	2	1	1	2	1	2	9

## MOYENNE MARGINALE ET VARIANCE MARGINALE DE Y

$$\bar{y} = \frac{1}{n_{..}} \sum_{j=1}^6 n_{.j} y_j = 9.67$$

$$\sigma_y^2 = \frac{1}{n_{..}} \sum_{j=1}^6 n_{.j} (y_j - \bar{y})^2 = \overline{y^2} - (\bar{y})^2 = 8.89$$

## DISTRIBUTION CONDITIONNELLE DE $X|Y = 8$

$(X, Y)$	6	7	8	10	12	14	$\Sigma$
6	2	1	0	0	0	0	3
8	0	0	1	0	0	0	1
9	0	0	0	1	0	0	1
11	0	0	0	1	0	0	1
13	0	0	0	0	1	0	1
14	0	0	0	0	0	2	2
$\Sigma$	2	1	1	2	1	2	9

$x_i$	6	8	9	11	13	14	$\Sigma$
$n_{i3}$	0	1	0	0	0	0	1

## MOYENNE CONDITIONNELLE ET VARIANCE CONDITIONNELLE DE $X|Y = 8$

$$\bar{x}_3 = \frac{1}{n_{.3}} \sum_{i=1}^6 n_{i3} x_i = 8$$

$$\sigma_{x,3}^2 = \frac{1}{n_{.3}} \sum_{i=1}^6 n_{i3} (x_i - \bar{x}_3)^2 = 0$$

## DISTRIBUTION CONDITIONNELLE DE $Y|(X = 13)$

$(X, Y)$	6	7	8	10	12	14	$\Sigma$
6	2	1	0	0	0	0	3
8	0	0	1	0	0	0	1
9	0	0	0	1	0	0	1
11	0	0	0	1	0	0	1
13	0	0	0	0	1	0	1
14	0	0	0	0	0	2	2
$\Sigma$	2	1	1	2	1	2	9

$y_j$	6	7	8	10	12	14	$\Sigma$
$n_{5j}$	0	0	0	0	1	0	1

## MOYENNE CONDITIONNELLE ET VARIANCE CONDITIONNELLE DE $Y|X = 13$

$$\bar{y}_5 = \frac{1}{n_5} \sum_{j=1}^6 n_{5j} y_j = 12$$

$$\sigma_{y,5}^2 = \frac{1}{n_5} \sum_{j=1}^6 n_{5j} (y_j - \bar{y}_5)^2 = 0$$

## DISTRIBUTION JOINTE DU COUPLE (X, Y)

(X, Y)	6	7	8	10	12	14	$\Sigma$
6	2	1	0	0	0	0	3
8	0	0	1	0	0	0	1
9	0	0	0	1	0	0	1
11	0	0	0	1	0	0	1
13	0	0	0	0	1	0	1
14	0	0	0	0	0	2	2
$\Sigma$	2	1	1	2	1	2	9

## MOYENNE JOINTE ET COVARIANCE DU COUPLE (X, Y)

$$\overline{xy} = \frac{1}{n_{..}} \sum_{i=1}^6 \sum_{j=1}^6 n_{ij} x_i y_j = 102.89 \quad \text{cov}(X, Y) = \frac{1}{n_{..}} \sum_{i=1}^6 \sum_{j=1}^6 n_{ij} (x_i - \bar{x})(y_j - \bar{y}) = \overline{xy} - \bar{x} \cdot \bar{y} = 9.44$$

Une covariance positive signifie que les variables  $X$  (notes des mathématiques) et  $Y$  (notes de physique) varient dans le même sens.

## COEFFICIENT DE CORRÉLATION LINÉAIRE ENTRE $X$ ET $Y$

$$\rho(X, Y) = \frac{\text{cov}(X, Y)}{\sigma_x \sigma_y} = 0.98$$

$$-1 \leq \rho(X, Y) \leq 1$$

Un coefficient de corrélation linéaire assez proche de 1 signifie que les variables  $X$  (notes des mathématiques) et  $Y$  (notes de physique) entretiennent une très forte relation positive.

## MATRICE DES VARIANCES-COVARIANCES

$\text{cov}(X, Y)$	$X$	$Y$
$X$	10.48	9.44
$Y$	9.44	8.89

## MATRICE DES CORRÉLATIONS

$\rho(X, Y)$	$X$	$Y$
$X$	1	0.98
$Y$	0.98	1

## SECTION 1.3

### VERS LE CAS MULTIVARIÉ

- MATRICE DES VARIANCES-COVARIANCES
- MATRICE DES CORRÉLATIONS
- TABLEAU DE NUAGES

- On suppose que les individus sont décrits par  $p$  variables  $X_1, X_2, \dots, X_p$ .
- L'objectif du cours est d'exposer les techniques de la statistique descriptive multidimensionnelle.
- Il est possible de débiter une exploration de données multidimensionnelles en adaptant simplement les méthodes déjà étudiées.

## MATRICE DES VARIANCES-COVARIANCES

$cov(X_i, X_j)$	$X_1$	$X_2$	$\dots$	$X_p$
$X_1$	$cov(X_1, X_1) = \sigma_{x_1}^2$	$cov(X_1, X_2)$	$\dots$	$cov(X_1, X_p)$
$X_2$	$cov(X_2, X_1)$	$cov(X_2, X_2) = \sigma_{x_2}^2$	$\dots$	$cov(X_2, X_p)$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
$X_p$	$cov(X_p, X_1)$	$cov(X_p, X_2)$	$\dots$	$cov(X_p, X_p) = \sigma_{x_p}^2$

## MATRICE DES CORRÉLATIONS

$\rho(X_i, X_j)$	$X_1$	$X_2$	$\dots$	$X_p$
$X_1$	$\rho(X_1, X_1) = 1$	$\rho(X_1, X_2)$	$\dots$	$\rho(X_1, X_p)$
$X_2$	$\rho(X_2, X_1)$	$\rho(X_2, X_2) = 1$	$\dots$	$\rho(X_2, X_p)$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
$X_p$	$\rho(X_p, X_1)$	$\rho(X_p, X_2)$	$\dots$	$\rho(X_p, X_p) = 1$

## NOTES SUR 20 DE 9 ÉLÈVES DANS 4 MATIÈRES

Élève	Mathématiques	Physique	Français	Anglais
A	6	6	5	6
B	8	8	8	8
C	6	6	11	10
D	14	14	15	15
E	14	14	12	13
F	11	10	6	7
G	6	7	14	12
H	13	12	9	10
I	9	10	12	12

## ANALYSE À UNE DIMENSION

Matière	Moyenne	Variance	Écart-type
Mathématiques	9.67	10.44	3.23
Physique	9.67	8.89	2.98
Français	10.22	10.62	3.26
Anglais	10.33	7.78	2.79

- En moyenne, les meilleures notes sont obtenues en Anglais (10.33) et en Français (10.22),
- Les notes obtenues en Français sont les plus dispersées (plus grand écart-type 3.26),
- Les notes obtenues en Anglais sont les plus homogènes (plus petit écart-type 2.79).

## ANALYSE À DEUX DIMENSIONS : MATRICE DES VARIANCES-COVARIANCES

	Mathématiques	Physique	Français	Anglais
Mathématiques	10.44	10.44	0.67	4.17
Physique	10.44	8.89	3.67	4.86
Français	0.67	3.67	10.62	8.69
Anglais	4.17	4.86	8.69	7.78

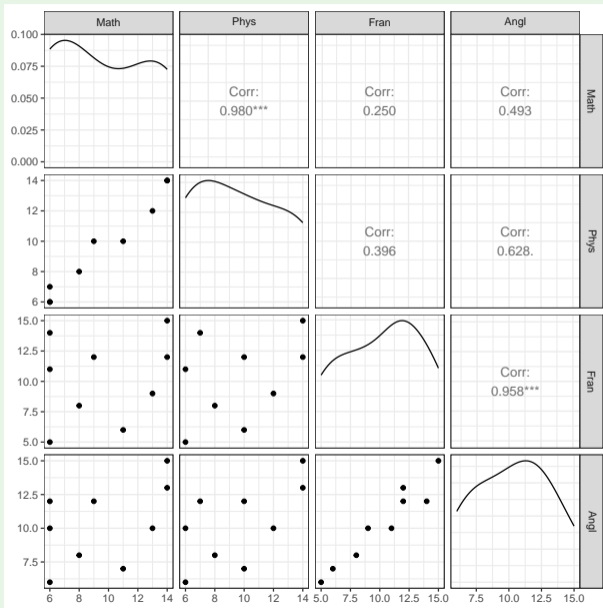
- Les éléments diagonaux de la matrice des variances-covariances sont les variances,
- Les éléments hors diagonaux sont des covariances,
- Les covariances étant toutes positives, on conclut que les notes évoluent dans le même sens.

## ANALYSE À DEUX DIMENSIONS : MATRICE DES CORRÉLATIONS

	Mathématiques	Physique	Français	Anglais
Mathématiques	1	0.98	0.25	0.49
Physique	0.98	1	0.40	0.63
Français	0.25	0.40	1	0.96
Anglais	0.49	0.63	0.96	1

- La corrélation entre les matières scientifiques est très forte (0.98),
- La corrélation entre les matières littéraires est très forte (0.96),
- La corrélation entre les mathématiques et le français est faible (0.25),
- La corrélation entre la physique et l'anglais est moyenne (0.63).

# TABLEAU DE NUAGES



1. Approfondissements de la statistique descriptive

**2. Géométrie de l'algèbre linéaire**

3. Analyse en Composantes Principales

## CHAPITRE 2

### GÉOMÉTRIE DE L'ALGÈBRE LINÉAIRE

SECTION 2.1 : MATRICE, VECTEUR ET SCALAIRE

SECTION 2.2 : MULTIPLICATION PAR UN SCALAIRE, PRODUITS SCALAIRE ET MATRICIEL

SECTION 2.3 : INTERPRÉTATION GÉOMÉTRIQUE

SECTION 2.4 : VALEURS ET VECTEURS PROPRES D'UNE MATRICE CARRÉE

## SECTION 2.1

## MATRICE, VECTEUR ET SCALAIRE

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n1} & \cdots & a_{nm} \end{pmatrix}$$

- $A$  est une **matrice** de taille  $n \times m$  (elle contient  $n$  **lignes** et  $m$  **colonnes**), on la note  $A_{n \times m}$ .
- Si  $n = m$  alors la matrice  $A$  est **carrée**, sinon elle est **rectangulaire**.
- Si  $n = 1$  ou  $m = 1$  alors la matrice  $A$  est un **vecteur**, noté  $\mathbf{x}$ .
- Si  $n = m = 1$  alors la matrice  $A$  est un **scalaire** (un nombre), noté  $\alpha$ .

$$A = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 3 & 1 \\ 1 & 2 & 3 \end{pmatrix}$$

$$\mathbf{x} = \begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix}$$

$$\alpha = 2$$

## SECTION 2.2

### MULTIPLICATION PAR UN SCALAIRE, PRODUITS SCALAIRE ET MATRICIEL

#### MULTIPLICATION D'UNE MATRICE PAR UN SCALAIRE

La multiplication d'une matrice  $A$  par un scalaire  $\alpha$  est donnée par :

$$\alpha A = \begin{pmatrix} \alpha a_{11} & \alpha a_{12} & \cdots & \alpha a_{1m} \\ \alpha a_{21} & \alpha a_{22} & \cdots & \alpha a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha a_{n1} & \alpha a_{n1} & \cdots & \alpha a_{nm} \end{pmatrix}$$

- Chacun des éléments  $a_{ij}$  de la matrice  $A$  est multiplié par le scalaire  $\alpha$ .

$$\alpha = 2$$

$$A = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 3 & 1 \\ 1 & 2 & 3 \end{pmatrix}$$

$$2A = \begin{pmatrix} 2 \times 2 & 2 \times 1 & 2 \times 0 \\ 2 \times 0 & 2 \times 3 & 2 \times 1 \\ 2 \times 1 & 2 \times 2 & 2 \times 3 \end{pmatrix}$$

## PRODUIT SCALAIRE

Le produit scalaire des vecteurs  $\mathbf{x} = (x_1 \ x_2 \ \cdots \ x_n)$  et  $\mathbf{y} = (y_1 \ y_2 \ \cdots \ y_n)$ , noté  $\mathbf{xy}$ , est donné par :

$$\mathbf{xy} = x_1y_1 + x_2y_2 + \cdots + x_ny_n$$

- Les vecteurs  $\mathbf{x}$  et  $\mathbf{y}$  contiennent le même nombre d'éléments.
- Les vecteurs  $\mathbf{x}$  et  $\mathbf{y}$  n'ont pas nécessairement les mêmes dimensions.

$$(3 \ 2 \ 1)(2 \ 1 \ 0) = 3 \times 2 + 2 \times 1 + 1 \times 0 = 8$$

$$\begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix} \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} = 3 \times 2 + 2 \times 1 + 1 \times 0 = 8$$

$$(3 \ 2 \ 1) \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} = 3 \times 2 + 2 \times 1 + 1 \times 0 = 8$$

## PRODUIT MATRICIEL

Le produit matriciel des matrices  $A$  et  $B$  est une matrice  $AB$  telle que :

$$\begin{matrix} A & \times & B & = & AB \\ n \times m & & m \times p & & n \times p \end{matrix}$$

- Le nombre de colonnes de la matrice  $A$  est égal au nombre de lignes de la matrice  $B$ .
- $AB$  a le même nombre de lignes que  $A$  et le même nombre de colonnes  $B$ .
- Chaque élément de  $AB$  est le produit scalaire de la ligne et de la colonne correspondantes dans  $A$  et  $B$  respectivement.

## EXEMPLE

$$\begin{pmatrix} 2 & 1 & 0 \\ 0 & 3 & 1 \\ 1 & 2 & 3 \end{pmatrix} \times \begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 3 & 1 \\ 1 & 2 & 3 \end{pmatrix} \begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \times 3 + 1 \times 2 + 0 \times 1 \\ 0 \times 3 + 3 \times 2 + 1 \times 1 \\ 1 \times 3 + 2 \times 2 + 3 \times 1 \end{pmatrix} = \begin{pmatrix} 8 \\ 7 \\ 10 \end{pmatrix}$$

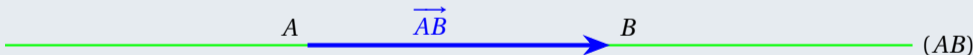
## SECTION 2.3

### INTERPRÉTATION GÉOMÉTRIQUE

#### VECTEUR GÉOMÉTRIQUE

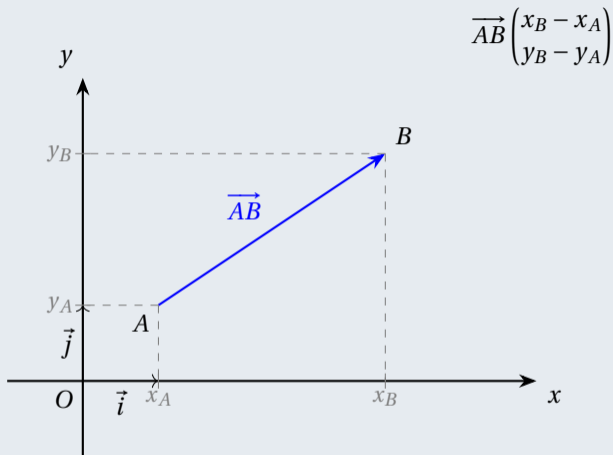
Un **vecteur géométrique**  $\overrightarrow{AB}$  est un segment de droite orienté, représenté par une flèche, défini par trois caractéristiques fondamentales :

- Une direction : la droite support  $(AB)$  ou ses parallèles ;
- Un sens : de l'origine  $A$  vers l'extrémité  $B$  ;
- Une norme : sa longueur  $AB$ .



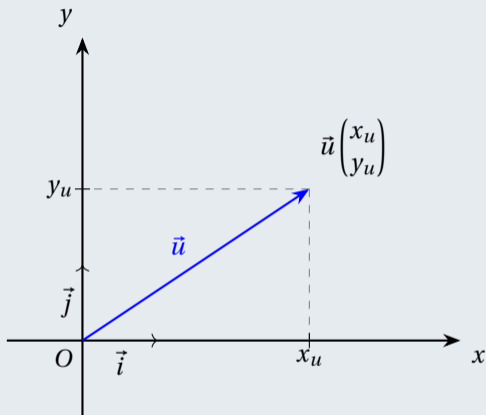
Si l'une des trois caractéristiques d'un vecteur  $\overrightarrow{AB}$  est modifiée, le vecteur qui en résulte n'est plus le même que le vecteur  $\overrightarrow{AB}$ .

## COORDONNÉES D'UN VECTEUR DANS UN PLAN



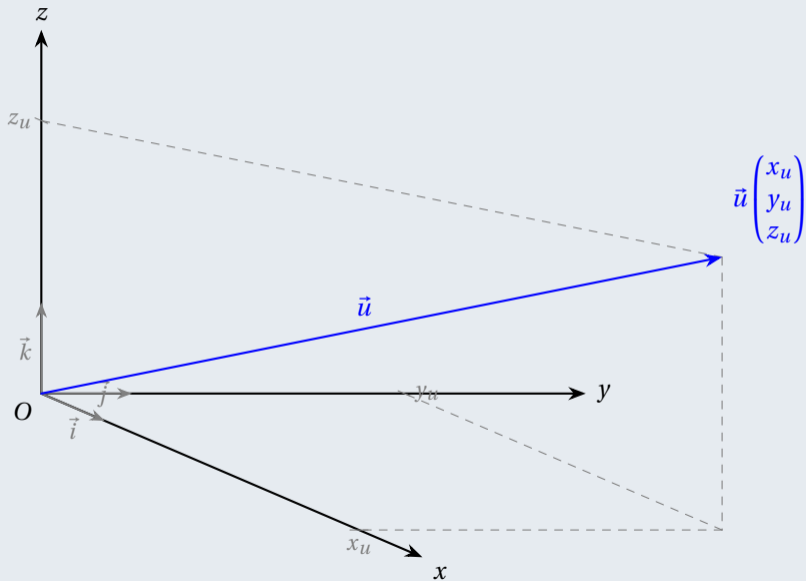
Dans un repère cartésien, le vecteur algébrique  $(x_B - x_A, y_B - y_A)$  représente les coordonnées du vecteur géométrique  $\vec{AB}$ .

## COORDONNÉES D'UN VECTEUR DANS UN PLAN



Dans un repère cartésien, le vecteur algébrique  $(x_u, y_u)$  représente les coordonnées du vecteur géométrique  $\vec{u}$  d'origine  $O$ .

## COORDONNÉES D'UN VECTEUR DANS UN ESPACE 3D



## CONTEXTE DE L'ENTREPRISE

Une société technologique souhaite lancer un nouveau produit sur le marché. Pour assurer le succès de ce lancement, la direction marketing doit décider de la répartition de son budget de communication entre trois canaux principaux.

## DONNÉES

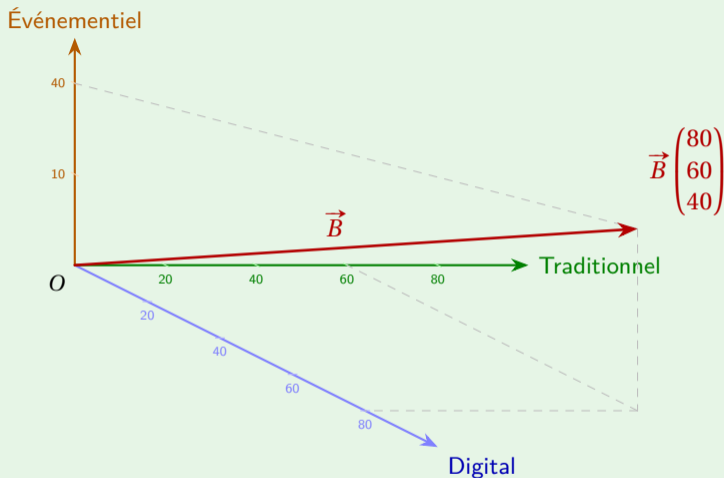
Le budget total alloué est de **180000**. La stratégie retenue se décompose comme suit :

- **Marketing Digital** : 80000 (Réseaux sociaux, SEA, SEO)
- **Marketing Traditionnel** : 60000 (Presse, Radio, Affichage)
- **Événementiel** : 40000 (Salons, Conférences, Lancements)

## TRAVAIL DEMANDÉ

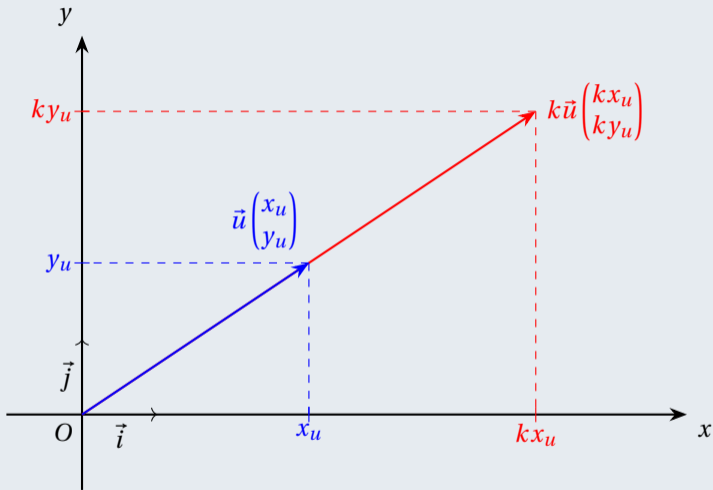
1. Modéliser cette stratégie par un vecteur  $\vec{B}$  dans un repère orthonormé  $(O, \vec{i}, \vec{j}, \vec{k})$  où chaque axe représente un canal.
2. Calculer la norme  $\|\vec{B}\|$  du vecteur. Que représente-t-elle économiquement ?

## APPLICATION : VECTEUR DE BUDGET MARKETING ( $\times 10000$ )



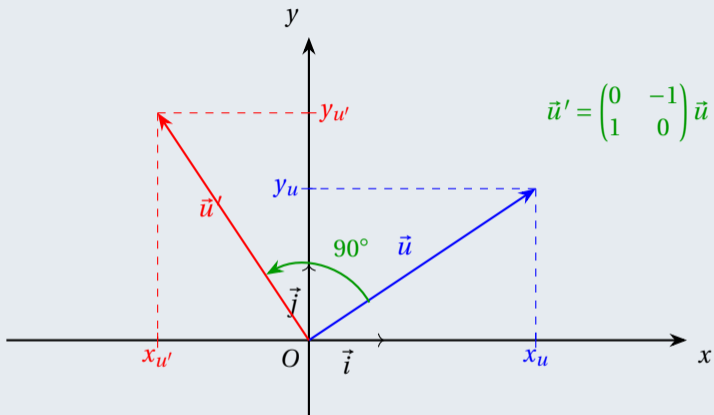
$\|\vec{B}\| = \sqrt{80^2 + 60^2 + 40^2} \approx 107.7$ . Elle représente l'intensité de l'effort marketing global.

## MULTIPLICATION D'UN VECTEUR PAR UN SCALAIRE



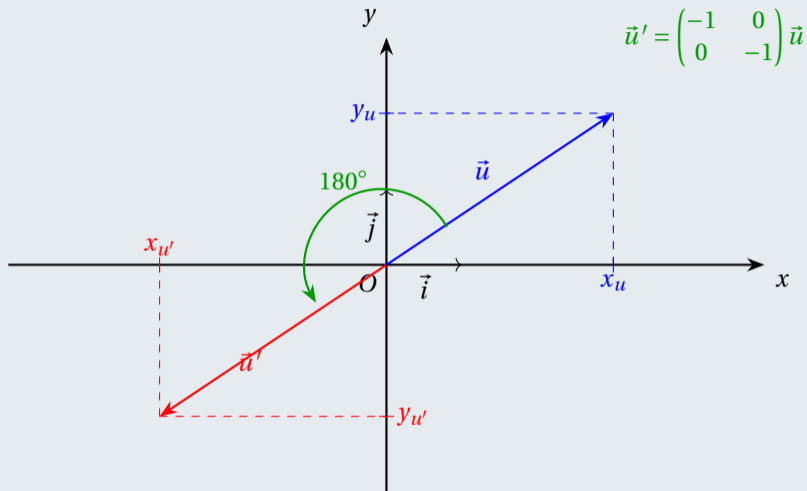
- $|k| > 1$  :  $k\vec{u}$  est plus long que  $\vec{u}$ .
- $|k| < 1$  :  $k\vec{u}$  est plus court que  $\vec{u}$ .
- $k > 0$  :  $k\vec{u}$  et  $\vec{u}$  sont orientés dans le même sens.
- $k < 0$  :  $k\vec{u}$  et  $\vec{u}$  sont orientés en sens inverses.

## ROTATION À 90° d'un vecteur



$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$  est la matrice de rotation à 90°.

## ROTATION À 180° d'un vecteur



$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$  est la matrice de rotation à 180°.

## SECTION 2.4

### VALEURS ET VECTEURS PROPRES D'UNE MATRICE CARRÉE

#### VALEUR ET VECTEUR PROPRES D'UNE MATRICE CARRÉE

Soit  $A$  une matrice carrée d'ordre  $n$ ,  $\lambda$  un scalaire et  $v$  un vecteur non nul de taille  $n$ .

$v$  est le **vecteur propre** de la matrice  $A$  associé à la **valeur propre**  $\lambda$  si et seulement si  $\lambda$  et  $v$  vérifient l'équation  $Av = \lambda v$ .

L'équation précédente s'écrit également  $(A - \lambda I)v = 0$ , où  $I$  est la matrice identité d'ordre  $n$ . Pour qu'il existe une solution non nulle à cette équation, il faut que  $\det(A - \lambda I) = 0$ . C'est l'**équation caractéristique** dont les racines sont les valeurs propres de la matrice  $A$ .

## EXEMPLE

$$A = \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}$$

$$v_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \rightarrow Av_1 = \begin{pmatrix} 2 \\ 0 \end{pmatrix} = 2v_1 \rightarrow \lambda_1 = 2$$

$$v_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \rightarrow Av_2 = \begin{pmatrix} 0 \\ 3 \end{pmatrix} = 3v_2 \rightarrow \lambda_2 = 3$$

1. Approfondissements de la statistique descriptive

2. Géométrie de l'algèbre linéaire

**3. Analyse en Composantes Principales**