Statistique, probabilités et algèbre linéaire

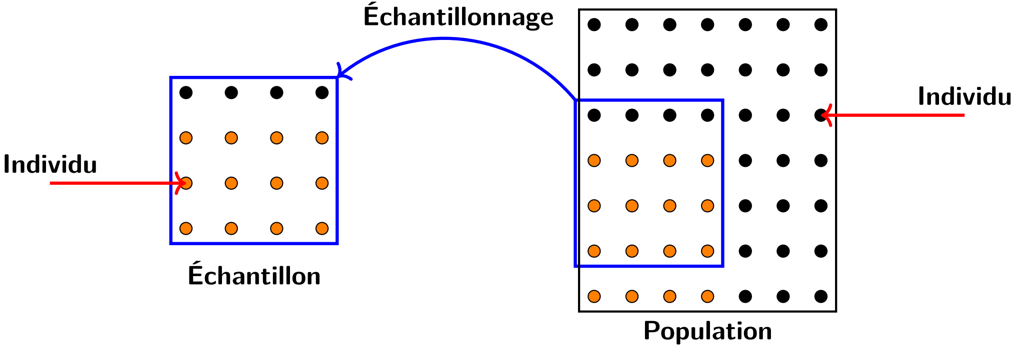
Rappels de l'essentiel pour l'analyse des données, l'estimation et l'économétrie

Jaouad Madkour
`jmadkour@uae.ac.ma`

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

6 mars 2026

jmadkour.org


Échantillonnage
Individu
Individu
Échantillon
Population

Statistique descriptive

- Les caractéristiques des individus sont observables et la population est de taille finie ;
- **et** les données statistiques sont disponibles et à moindre coût.
- $\Rightarrow$ Décrire toute la population avec les outils de la statistique descriptive.

Statistique inférentielle ($\approx$ échantillonnage et estimation)

- Les caractéristiques des individus sont observables **mais** la population est de taille infinie ;
- **ou** la population est de taille finie mais les données ne sont pas disponibles et sont coûteuses.
- $\Rightarrow$ Estimer les paramètres de la population sur un échantillon avec la statistique inférentielle.

Probabilités

- Les caractéristiques des individus sont inobservables.
- On peut juste émettre des hypothèses ;
- $\Rightarrow$ Calculer les probabilités des hypothèses.

Étude statistique sur les étudiants au Maroc selon leurs âges

Population : Tous les étudiants

- La population des étudiants au Maroc est de taille infinie ;
- On prélève un échantillon et on applique les outils de la **statistique inférentielle**.

Population : Étudiants de la FSJES de Tanger

- La population des étudiants de la FSJES de Tanger est de taille finie ;
- Les données ne sont pas disponibles et elles sont coûteuses.
- On prélève un échantillon et on applique les outils de la **statistique inférentielle**.

Application

- Échantillon : 100 étudiants $\Rightarrow$ Moyenne d'échantillon : $\overline{x} = 22$ ans.
- L'âge moyen des étudiants au Maroc μ est **estimé** à 22 ans.
- Il y a une probabilité de 95% que l'âge moyen appartienne à l'**intervalle de confiance** $[20, 24]$.

Population : Étudiants de la FSJES de Tanger

- La population des étudiants de la FSJES de Tanger est de taille finie ;
- Les données sont disponibles et à moindre coût.
- Décrire toute la population à l'aide des outils de la **statistique descriptive**.

Application

- Population : 1500 étudiants ⇒ Moyenne de la population : $\mu = 21$ ans
- L'âge moyen des étudiants de la FSJES Tanger μ est **égal** à 21 ans. Il est connu avec certitude.

Population : Les étudiants au Maroc l'année prochaine

- Les étudiants de l'année prochaine ne sont pas observables ;
- On émet des hypothèses et on calcule des probabilités.

Application

A partir des statistiques des années précédentes, on établit la distribution des probabilités des âges :

Age x	20	21	22	23	24
$P(x)$	0.1	0.2	0.4	0.2	0.1

Sommaire

Chapitre 1

Rappels de statistique descriptive

Section 1.1 : Description d'une variable statistique

Section 1.2 : Liaison entre deux variables statistiques

Section 1.1

Description d'une 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** $x_1, x_2, \cdots, x_k$
- 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. $N = n_1 + n_2 + \cdots + n_k$.
- Une proportion $f_i = \dfrac{n_i}{N}$ d'individus possèdent la modalité x_i. Ainsi, f_i est la **fréquence** de x_i.

Application

x_i	20	21	22	23	24	Σ
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	$\cdots$	x_k	Σ
n_i	n_1	n_2	$\cdots$	n_k	N

Variable quantitative continue

$[x_i, x_{i+1}[$	$[x_1, x_2[$	$[x_2, x_3[$	$\cdots$	$[x_k, x_{k+1}[$	Σ
n_i	n_1	n_2	$\cdots$	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	Variable continue
$\overline{x} = \frac{1}{N}\sum_{i=1}^{k} n_i x_i$	$\overline{x} = \frac{1}{N}\sum_{i=1}^{k} n_i c_i$

Variance et écart-type

Variable discrète	Variable continue
$\sigma^2 = \frac{1}{N}\sum_{i=1}^{k} n_i \left(x_i - \overline{x}\right)^2 = \overline{x^2} - \left(\overline{x}\right)^2$	$\sigma^2 = \frac{1}{N}\sum_{i=1}^{k} n_i \left(c_i - \overline{x}\right)^2 = \overline{c^2} - \left(\overline{x}\right)^2$
$\sigma = \sqrt{\sigma^2}$	$\sigma = \sqrt{\sigma^2}$

Fréquence

Variable discrète	Variable continue
$f_i = \frac{n_i}{N}$	$f_i = \frac{n_i}{N}$

Application

Variable discrète

x_i	20	21	22	23	24	Σ
n_i	10	20	40	20	10	100
f_i	0.1	0.2	0.4	0.2	0.1	1

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

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

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

Variable continue

x_i	$[20,21[$	$[21,22[$	$[22,23[$	$[23,24]$	Σ
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

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

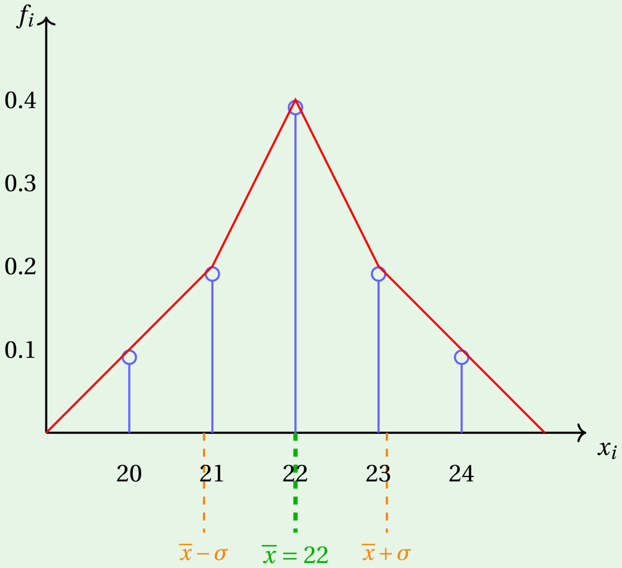
$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{4} n_i \left(c_i - \overline{x}\right)^2 = 0.81$$

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

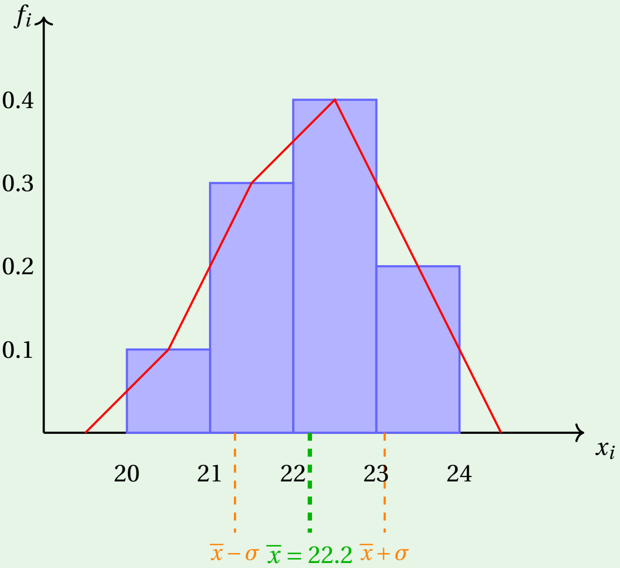
Résumé graphique d'une variable quantitative

- Diagramme en bâtons
- Histogramme
- Polygone des fréquences

Diagramme en bâtons - Polygone des fréquences



Histogramme - Polygone des fréquences



Description d'une variable qualitative

Variable qualitative nominale

Catégories	Catégorie 1	Catégorie 2	$\cdots$	Catégorie k	Σ
n_i	n_1	n_2	$\cdots$	n_k	N

Variable qualitative ordinale

Catégories	1ère catégorie	2ème catégorie	$\cdots$	$k$ème catégorie	Σ
n_i	n_1	n_2	$\cdots$	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	Σ
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	Σ
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

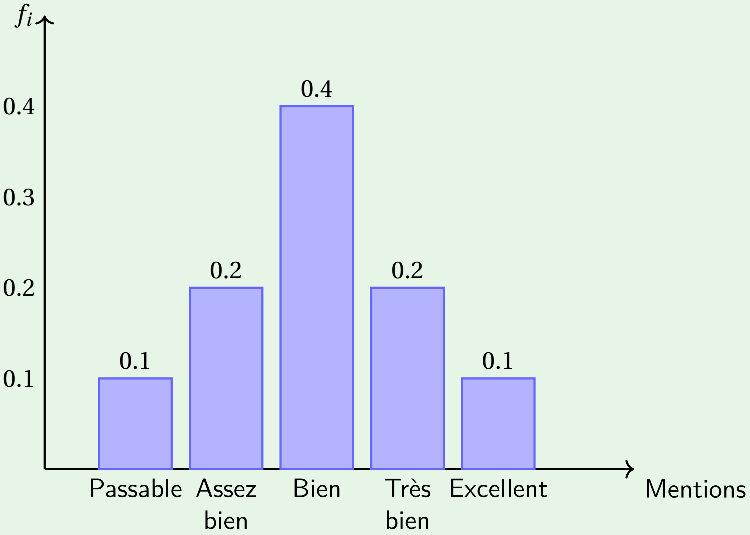


Diagramme en secteurs

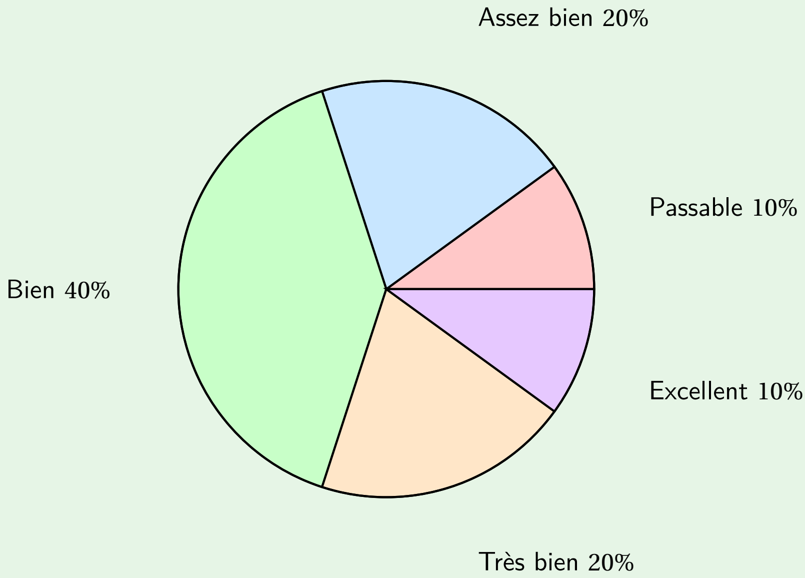


Diagramme en barres

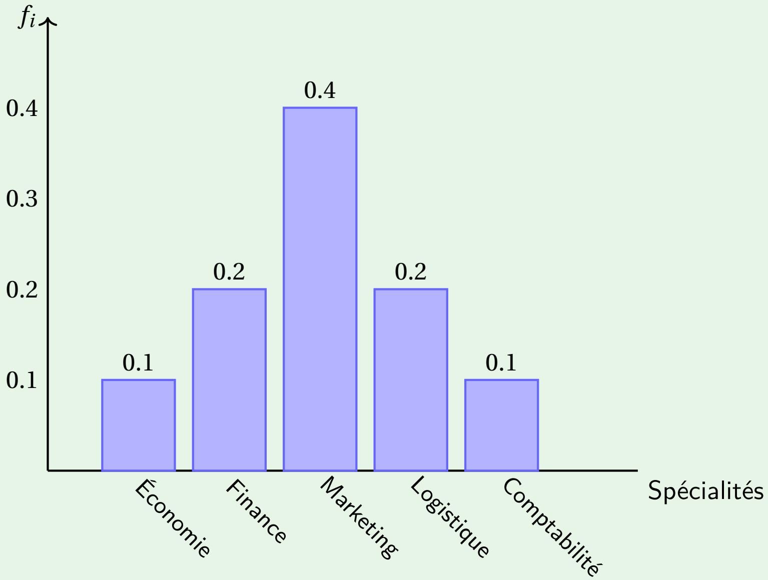
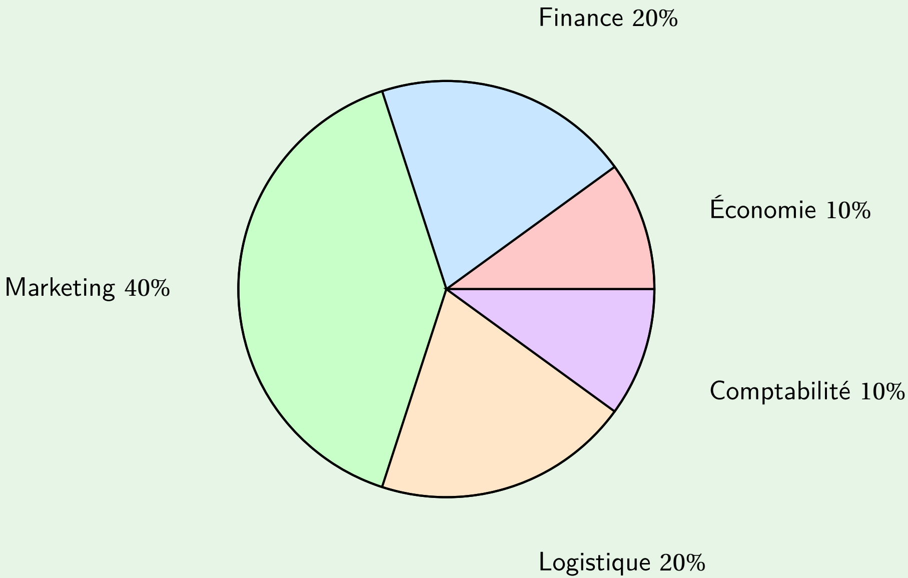


Diagramme en secteurs



Section 1.2

Liaison entre deux variables statistiques

- Distributions marginales
- Distributions conditionnelles
- Distribution jointe

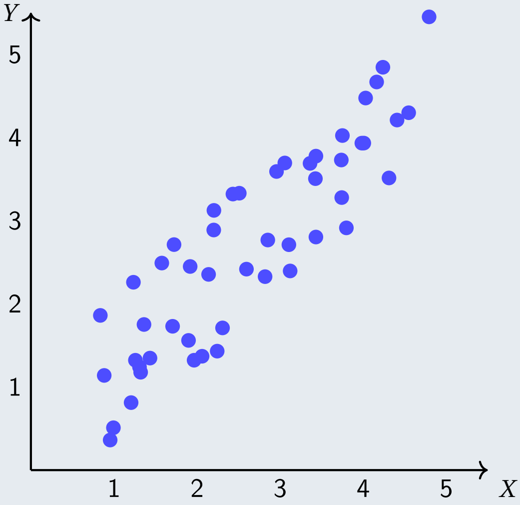
Tableau de contingence

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

Parfois

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

Nuage de points



Distribution marginale de X

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

x_i	x_1	x_2	$\cdots$	x_k	Σ
$n_{i.}$	$n_{1.}$	$n_{2.}$	$\cdots$	$n_{k.}$	$n_{..}$

Moyenne marginale et variance marginale de X

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

Distribution marginale de Y

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

y_j	y_1	y_2	$\cdots$	y_l	Σ
$n_{.j}$	$n_{.1}$	$n_{.2}$	$\cdots$	$n_{.l}$	$n_{..}$

Moyenne marginale et variance marginale de Y

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

Distribution conditionnelle de $X|(Y = y_2)$

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

x_i	x_1	x_2	$\cdots$	x_k	Σ
n_{i2}	n_{12}	n_{22}	$\cdots$	n_{k2}	$n_{.2}$

Moyenne conditionnelle et variance conditionnelle de X

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

$$\sigma^2_{x,2} = \frac{1}{n_{.2}} \sum_{i=1}^{k} n_{i2} \left(x_i - \overline{x}_2\right)^2$$

Distribution conditionnelle de $Y|(X = x_2)$

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

y_j	y_1	y_2	$\cdots$	y_l	Σ
n_{2j}	n_{21}	n_{22}	$\cdots$	n_{2l}	$n_{2.}$

Moyenne conditionnelle et variance conditionnelle de Y

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

$$\sigma^2_{y,2} = \frac{1}{n_{2.}} \sum_{j=1}^{l} n_{2j} \left(y_j - \overline{y}_2\right)^2$$

Distribution jointe du couple (X, Y)

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

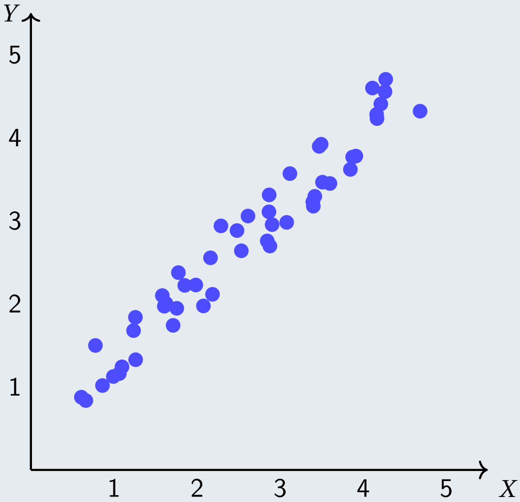
Moyenne jointe et covariance du couple (X, Y)

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

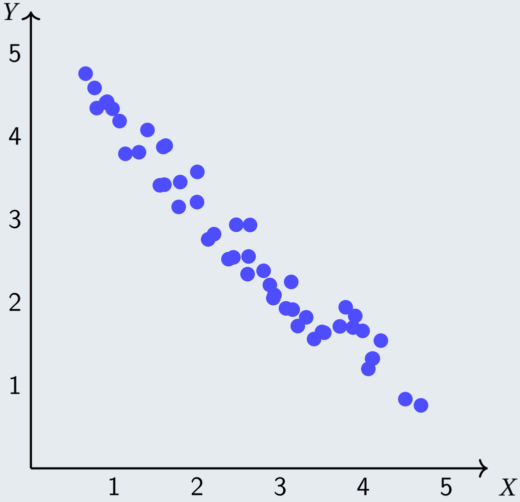
Coefficient de corrélation linéaire entre X et Y

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

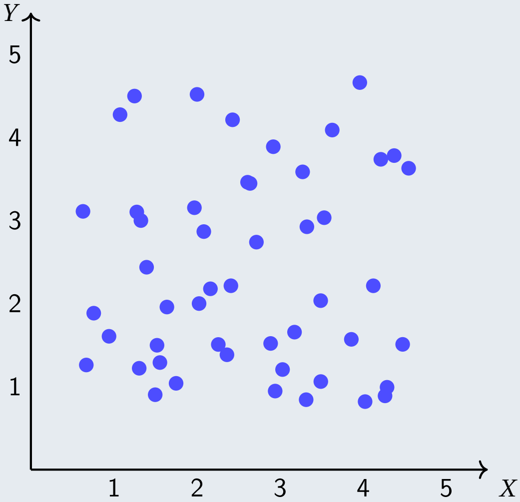
Covariance positive (corrélation positive)



Covariance négative



Covariance nulle



Matrice des variances-covariances

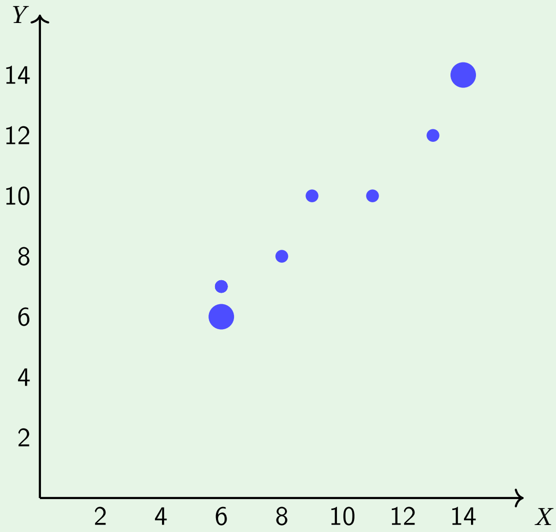
$cov(X,Y)$	X	Y
X	$cov(X,X)=\sigma_x^2$	$cov(X,Y)$
Y	$cov(X,Y)$	$cov(Y,Y)=\sigma_y^2$

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

$\Longrightarrow$

(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 : 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

$$\overline{\overline{x}} = \frac{1}{n_{..}} \sum_{i=1}^{6} n_{i.} x_i = 9.67 \qquad \sigma_x^2 = \frac{1}{n_{..}} \sum_{i=1}^{6} n_{i.} \left(x_i - \overline{\overline{x}}\right)^2 = \overline{\overline{x^2}} - \left(\overline{\overline{x}}\right)^2 = 10.48$$

Distribution marginale de Y

(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

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

Moyenne marginale et variance marginale de Y

$$\overline{\overline{y}} = \frac{1}{n_{..}} \sum_{j=1}^{6} n_{.j} y_j = 9.67 \qquad \sigma_y^2 = \frac{1}{n_{..}} \sum_{j=1}^{6} n_{.j} \left(y_j - \overline{\overline{y}}\right)^2 = \overline{\overline{y^2}} - \left(\overline{\overline{y}}\right)^2 = 8.89$$

Distribution conditionnelle de $X|(Y=8)$

(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_{i3}	0	1	0	0	0	0	1

Moyenne conditionnelle et variance conditionnelle de $X|Y=8$

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

$$\sigma^2_{x,3} = \frac{1}{n_{.3}} \sum_{i=1}^{6} n_{i3} \left(x_i - \overline{x}_3\right)^2 = 0$$

Distribution conditionnelle de $Y|(X=13)$

(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

y_j	6	7	8	10	12	14	Σ
n_{5j}	0	0	0	0	1	0	1

Moyenne conditionnelle et variance conditionnelle de $Y|X=13$

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

$$\sigma^2_{y,5} = \frac{1}{n_{5.}} \sum_{j=1}^{6} n_{5j} \left(y_j - \overline{y}_5\right)^2 = 0$$

Distribution jointe du couple (X, Y)

(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

Moyenne jointe et covariance du couple (X, Y)

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

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

$$\rho(X,Y) = \frac{cov(X,Y)}{\sigma_x \sigma_y} = 0.98 \qquad\qquad -1 \leq \rho(X,Y) \leq 1$$

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

$cov(X,Y)$	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

Sommaire

Chapitre 2

Rappels de probabilités

Section 2.1 : Distribution d'une variable aléatoire

Section 2.2 : Liaison entre deux variables aléatoires

Terminologie

- **Expérience aléatoire :** expérience dont on ne peut prévoir le résultat avec certitude.
- **Événement :** résultat d'une expérience aléatoire
- **Ensemble fondamental Ω :** ensemble des résultats possibles d'une expérience aléatoire.
- **Événement élémentaire $\omega \in \Omega$:** chaque élément de l'ensemble fondamental Ω.
- **Variable aléatoire X :** sa valeur est déterminée par le résultat ω d'une expérience aléatoire.
- **Variable aléatoire discrète :** si $X(\Omega)$ est dénombrable.
- **Variable aléatoire continue :** si $X(\Omega)$ est non dénombrable de $\mathbb{R}$.

Exemple

- **Expérience aléatoire :** Jeu de pile ou face.
- **Ensemble fondamental :** $\Omega = \{\text{pile}, \text{face}\}$
- **Événements élémentaires :** $\omega_1 = \text{pile}$, $\omega_2 = \text{face}$.
- **Variable aléatoire X** : "gagner 5 dirhams si pile, perdre 3 dirhams si face".
- On a $X(\text{pile}) = 5$ et $X(\text{face}) = -3$.
- $X(\Omega) = \{5, -3\}$ est dénombrable. X est une <u>variable aléatoire discrète</u>.

Section 2.1

Distribution d'une variable aléatoire

Variable aléatoire discrète

Fonction de **probabilités** $p(x)$:

$$0 \leq p(x) \leq 1 \qquad \sum_{i=1}^{k} p(x_i) = 1$$

x_i	x_1	x_2	$\cdots$	x_k	Σ
$p(x_i)$	$p(x_1)$	$p(x_2)$	$\cdots$	$p(x_k)$	1

Variable aléatoire continue

Fonction de **densité** $f(x)$:

$$f(x) \geq 0 \qquad \int_{-\infty}^{+\infty} f(x)\,\mathrm{d}x = 1$$

Moments d'une variable aléatoire

- Espérance mathématique
- Variance et écart-type

Variable statistique vs Variable aléatoire

Illustration : Dé équilibré

Variable statistique

x_i	1	2	3	4	5	6	Σ
f_i	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	1

$$\begin{aligned}\overline{x} &= \frac{1}{N}\sum_{i=1}^{k} n_i x_i \\ &= \sum_{i=1}^{k} \frac{n_i}{N} x_i \\ &= \sum_{i=1}^{k} f_i x_i\end{aligned}$$

Variable aléatoire

x_i	1	2	3	4	5	6	Σ
$p(x_i)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	1

$$E(X) = \sum_{i=1}^{k} p(x_i) x_i$$

Moments d'une variable aléatoire

Espérance mathématique

Variable discrète	**Variable continue**
$E(X) = \sum_{i=1}^{k} x_i \, p(x_i)$	$E(X) = \int_{-\infty}^{+\infty} x f(x) \, \mathrm{d}x$

Variance et écart-type

Variable discrète	**Variable continue**
$\sigma^2 = E\left[(X - E(X))^2\right] = E\left(X^2\right) - E^2(X)$	$\sigma^2 = E\left[(X - E(X))^2\right] = E\left(X^2\right) - E^2(X)$
$\sigma = \sqrt{\sigma^2}$	$\sigma = \sqrt{\sigma^2}$

Lois de probabilités

Variable discrète	**Variable continue**
• Loi uniforme discrète : $X \sim \mathscr{U}\{x_1, x_2, \cdots, x_n\}$	• Loi uniforme continue : $X \sim \mathscr{U}[a, b]$
$p(x) = 1/n$	$f(x) = \dfrac{1}{b-a}$ pour $a \leq x \leq b$
$E(X) = \dfrac{n+1}{2}$	$E(X) = \dfrac{a+b}{2}$
$V(X) = \dfrac{n^2-1}{12}$	$V(X) = \dfrac{(b-a)^2}{12}$

Lois de probabilités

Variable discrète

- Loi binomiale : $X \sim \mathscr{B}(n,p)$

$$p(x) = \mathscr{C}_n^x p^x (1-p)^{n-x}$$
$$E(X) = np$$
$$V(X) = np(1-p)$$

Variable continue

- Loi normale : $X \sim \mathscr{N}(\mu, \sigma^2)$

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$
$$E(X) = \mu$$
$$V(X) = \sigma^2$$

Lois de probabilités

Variable discrète

- Loi Poisson : $X \sim \mathscr{P}(\mu)$

$$p(x) = \frac{e^{-\mu}\mu^x}{x!}$$
$$E(X) = \mu$$
$$V(X) = \mu$$

Variable continue

- Loi exponentielle : $X \sim \mathscr{E}(\lambda)$

$$f(x) = \lambda e^{-\lambda x} \quad \text{pour } x \geq 0$$
$$E(X) = \frac{1}{\lambda}$$
$$V(X) = \frac{1}{\lambda^2}$$

Application

Variable discrète

x_i	20	21	22	23	24	Σ
$p(x_i)$	0.1	0.2	0.4	0.2	0.1	1

$$E(X) = \sum_{i=1}^{5} x_i\, p(x_i) = 22$$

$$\sigma^2 = E\left[(X - E(X))^2\right] = E\left(X^2\right) - E^2(X) = 1.2$$

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

Variable continue

$$\boxed{f(x) = \frac{1}{0.9\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-22}{0.9}\right)^2\right]}$$

$$E(X) = \int_{-\infty}^{+\infty} x\, f(x)\, \mathrm{d}x = 22$$

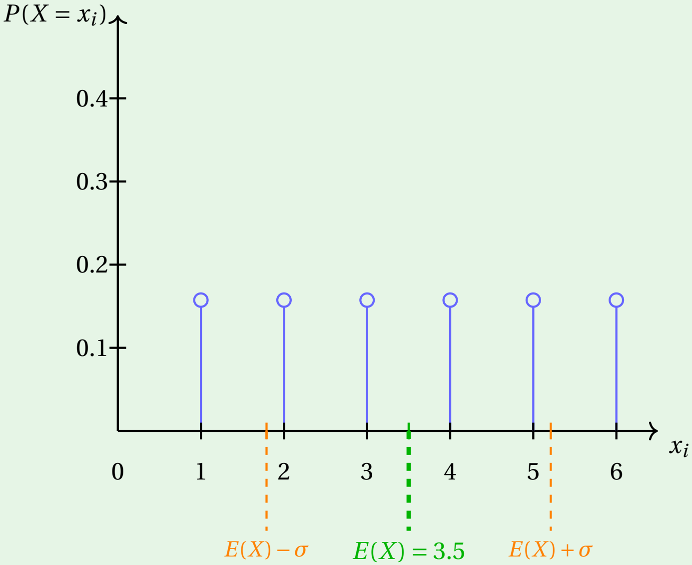
$$\sigma^2 = E\left[(X - E(X))^2\right] = E\left(X^2\right) - E^2(X) = 0.81$$

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

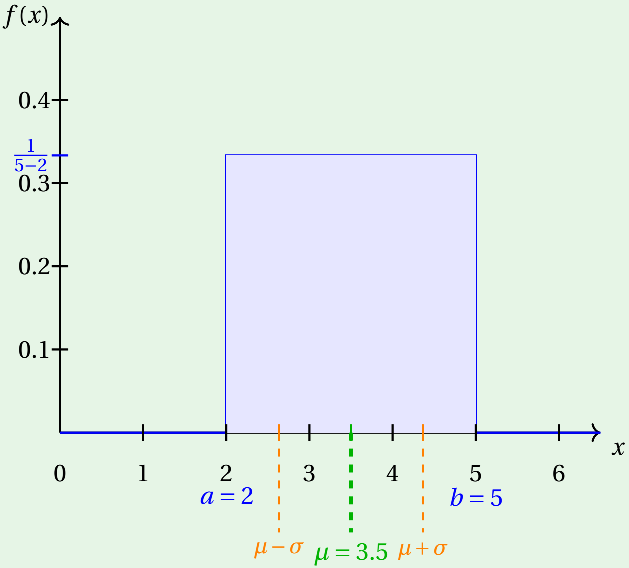
Résumé graphique d'une variable quantitative

- Diagramme en bâtons
- Histogramme
- Polygone des fréquences

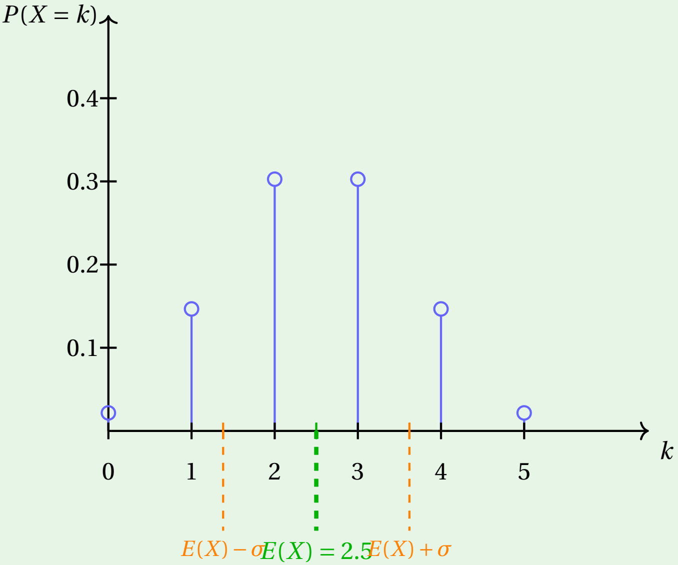
Probabilité loi uniforme discrète



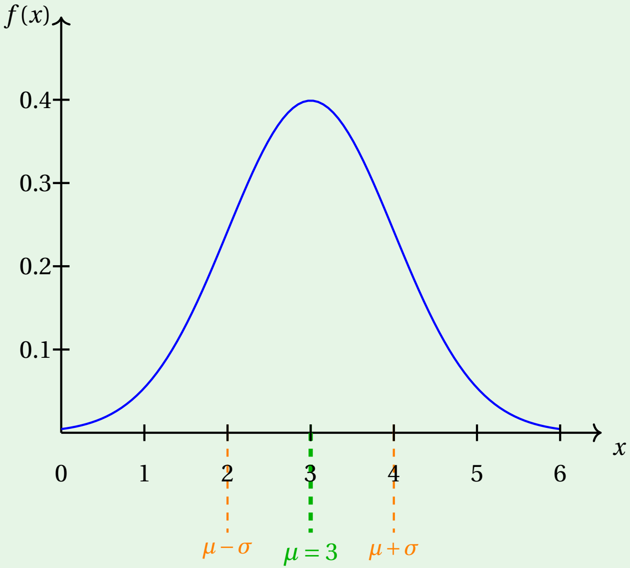
Probabilité loi uniforme continue



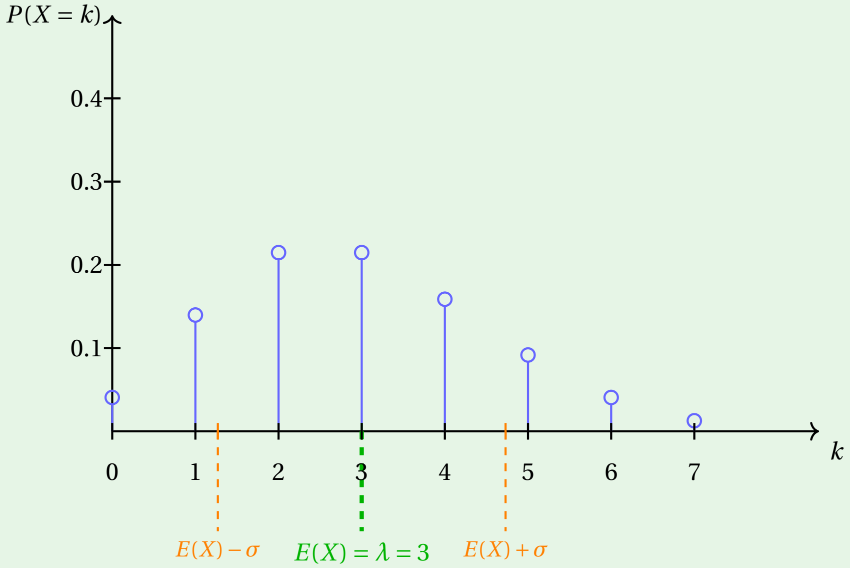
Probabilité loi uniforme binomiale



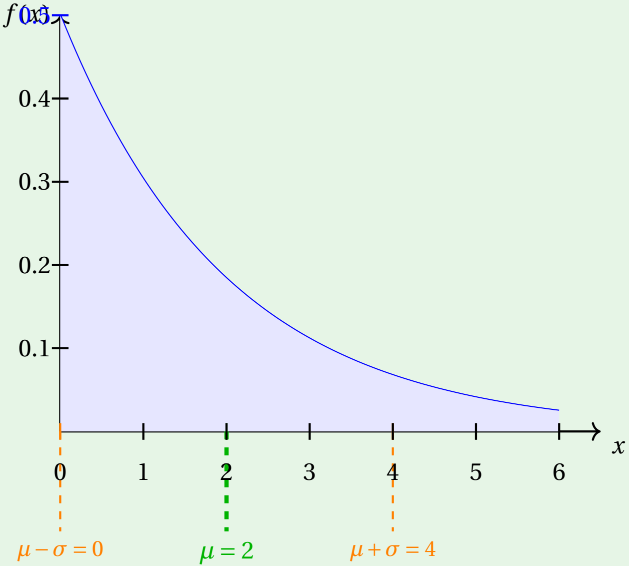
Densité de la loi normale



Densité loi Poisson



EXPONETIELLE



Section 2.2

Liaison entre deux variables aléatoires

- Distributions marginales
- Distributions conditionnelles
- Distribution jointe

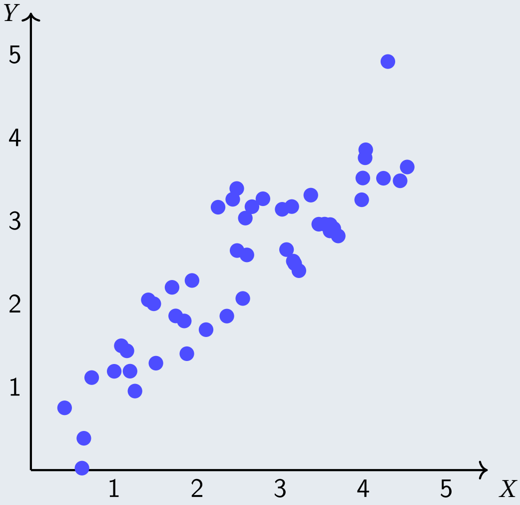
Tableau de contingence

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

Parfois

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

Nuage de points



Distribution marginale de X

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

x_i	x_1	x_2	$\cdots$	x_k	Σ
$n_{i.}$	$n_{1.}$	$n_{2.}$	$\cdots$	$n_{k.}$	$n_{..}$

Moyenne marginale et variance marginale de X

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

Distribution marginale de Y

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

y_j	y_1	y_2	$\cdots$	y_l	Σ
$n_{.j}$	$n_{.1}$	$n_{.2}$	$\cdots$	$n_{.l}$	$n_{..}$

Moyenne marginale et variance marginale de Y

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

Distribution conditionnelle de $X|(Y = y_2)$

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

x_i	x_1	x_2	$\cdots$	x_k	Σ
$n_{i.}$	$n_{1.}$	$n_{2.}$	$\cdots$	$n_{k.}$	$n_{..}$

Moyenne conditionnelle et variance conditionnelle de X

$$\overline{x}_2 = \frac{1}{n_{.2}} \sum_{i=1}^{k} n_{i2} x_i \qquad \sigma^2_{x,2} = \frac{1}{n_{.2}} \sum_{i=1}^{k} n_{i2} \left(x_i - \overline{x}_2\right)^2$$

Distribution conditionnelle de $Y|(X = x_2)$

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

y_j	y_1	y_2	$\cdots$	y_l	Σ
$n_{.j}$	$n_{.1}$	$n_{.2}$	$\cdots$	$n_{.l}$	$n_{..}$

Moyenne conditionnelle et variance conditionnelle de Y

$$\overline{y}_2 = \frac{1}{n_{2.}} \sum_{j=1}^{l} n_{2j} y_j \qquad \sigma^2_{y,2} = \frac{1}{n_{2.}} \sum_{j=1}^{l} n_{2j} \left(y_j - \overline{y}_2\right)^2$$

Distribution jointe du couple (X, Y)

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

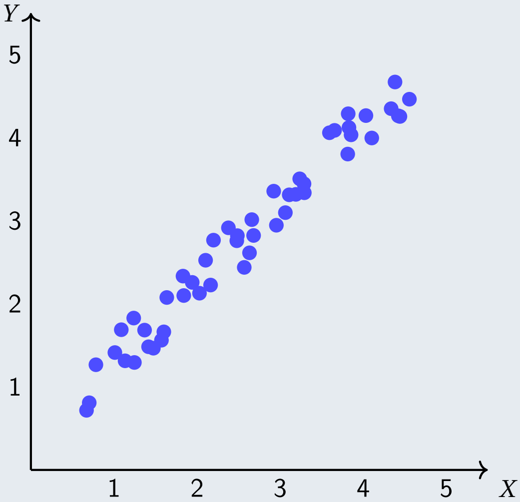
Moyenne jointe et covariance du couple (X, Y)

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

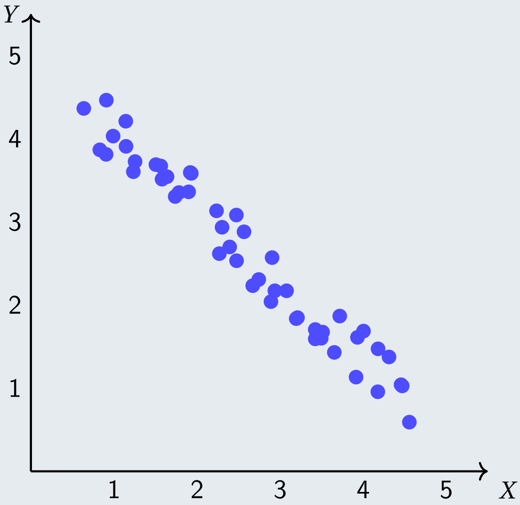
Coefficient de corrélation linéaire entre X et Y

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

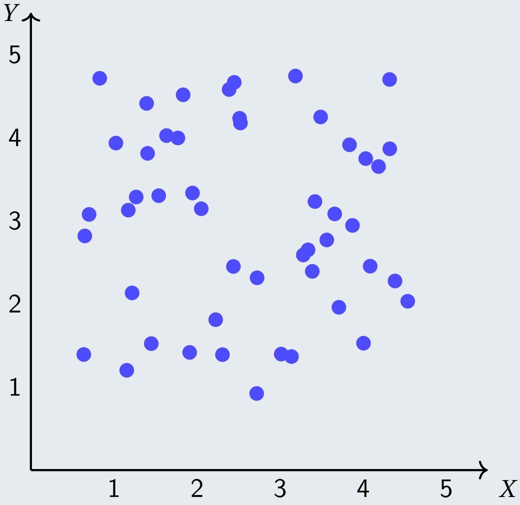
Covariance positive (corrélation positive)



Covariance négative



Covariance nulle



Matrice des variances-covariances

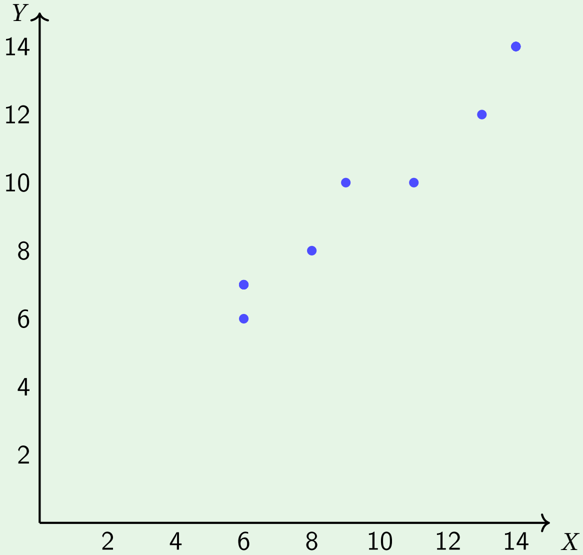
$cov(X,Y)$	X	Y
X	$cov(X,X) = \sigma_x^2$	$cov(X,Y)$
Y	$cov(X,Y)$	$cov(Y,Y) = \sigma_y^2$

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	7
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	7
D	14	14
E	14	14
F	11	10
G	6	7
H	13	12
I	9	10

$\Longrightarrow$

(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 : 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

$$\overline{\overline{x}} = \frac{1}{n_{..}} \sum_{i=1}^{6} n_{i.} x_i = 9.67 \qquad \sigma_x^2 = \frac{1}{n_{..}} \sum_{i=1}^{6} n_{i.} \left(x_i - \overline{\overline{x}}\right)^2 = \overline{\overline{x^2}} - \left(\overline{\overline{x}}\right)^2 = 10.48$$

Distribution marginale de Y

(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

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

Moyenne marginale et variance marginale de Y

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

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

Distribution conditionnelle de $X|(Y=8)$

(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
$n_{i.}$	0	1	0	0	0	0	1

Moyenne conditionnelle et variance conditionnelle de $X|Y=8$

$$\overline{x}_2 = \frac{1}{n_{.2}} \sum_{i=1}^{6} n_{i2} x_i = 8$$

$$\sigma_{x,2}^2 = \frac{1}{n_{.2}} \sum_{i=1}^{6} n_{i2} \left(x_i - \overline{x}_2\right)^2 = 0$$

Distribution conditionnelle de $Y|(X=13)$

(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

y_j	6	7	8	10	12	14	Σ
$n_{.j}$	0	0	0	0	1	0	1

Moyenne conditionnelle et variance conditionnelle de $Y|X=13$

$$\overline{y}_2 = \frac{1}{n_{2.}} \sum_{j=1}^{6} n_{2j} y_j = 12 \qquad \sigma^2_{y,2} = \frac{1}{n_{2.}} \sum_{j=1}^{6} n_{2j} \left(y_j - \overline{y}_2\right)^2 = 0$$

Distribution jointe du couple (X, Y)

(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

Moyenne jointe et covariance du couple (X, Y)

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

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

$$\rho(X,Y) = \frac{cov(X,Y)}{\sigma_x \sigma_y} = 0.98 \qquad\qquad -1 \leq \rho(X,Y) \leq 1$$

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

$cov(X,Y)$	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

Sommaire

Chapitre 3

Rappels d'algèbre linéaire

Section 3.1

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 dénote $\underset{n\times m}{A}$.
- 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**, dénoté $\mathbf{x}$.
- Si $n = m = 1$ alors la matrice A est un **scalaire** (un nombre), dénoté α.

$$A = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 3 & 1 \\ 1 & 2 & 3 \end{pmatrix} \qquad \mathbf{x} = \begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix} \qquad \alpha = 2$$

Multiplication d'une matrice par un scalaire

La multiplication d'une matrice A par un scalaire α 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 α.

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

Produit scalaire

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

$$\mathbf{xy} = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n$$

- 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 :

$$\underset{n\times m}{A} \times \underset{m\times p}{B} = \underset{n\times p}{AB}$$

- 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

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

Section 3.3

Interprétation géométrique

Vecteur géométrique

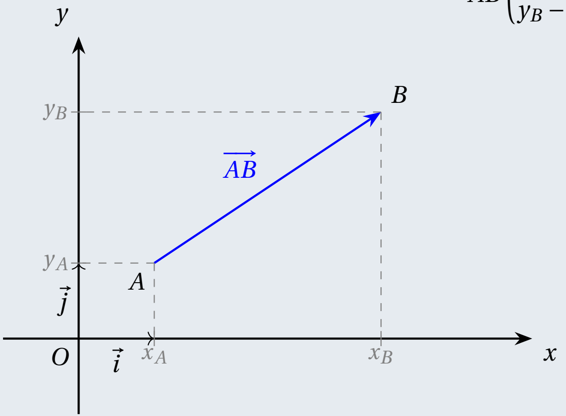
Un **vecteur géométrique** $\overrightarrow{AB}$ est un segment de droite orienté, représenté par une flèche, défini par trois caractéristiques fondamentales :

- 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 $\overrightarrow{AB}$.

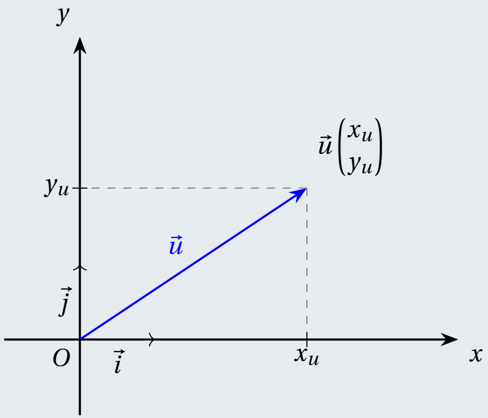
Coordonnées d'un vecteur dans un plan

$$\overrightarrow{AB}\begin{pmatrix} x_B - x_A \\ y_B - y_A \end{pmatrix}$$



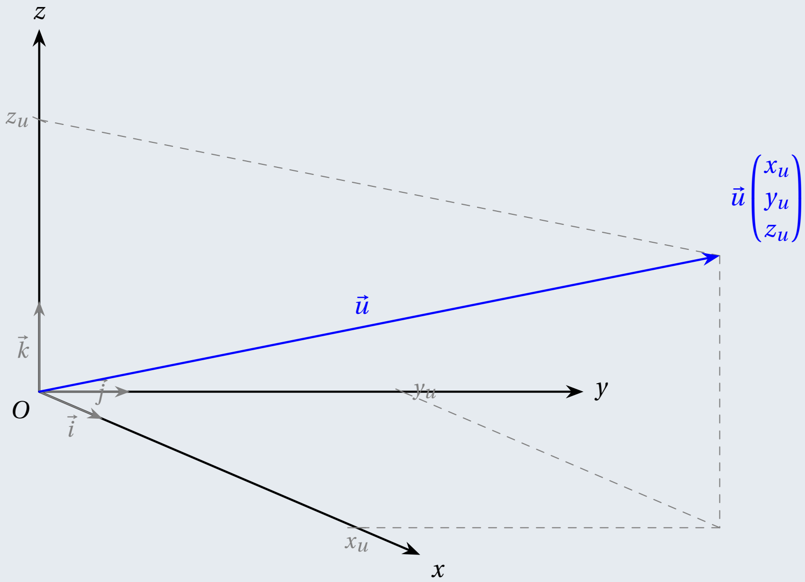
Dans un repère cartésien, le vecteur algébrique $(x_B - x_A, y_B - y_A)$ représente les coordonnées du vecteur géométrique $\overrightarrow{AB}$.

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

La société *InnovTech* 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

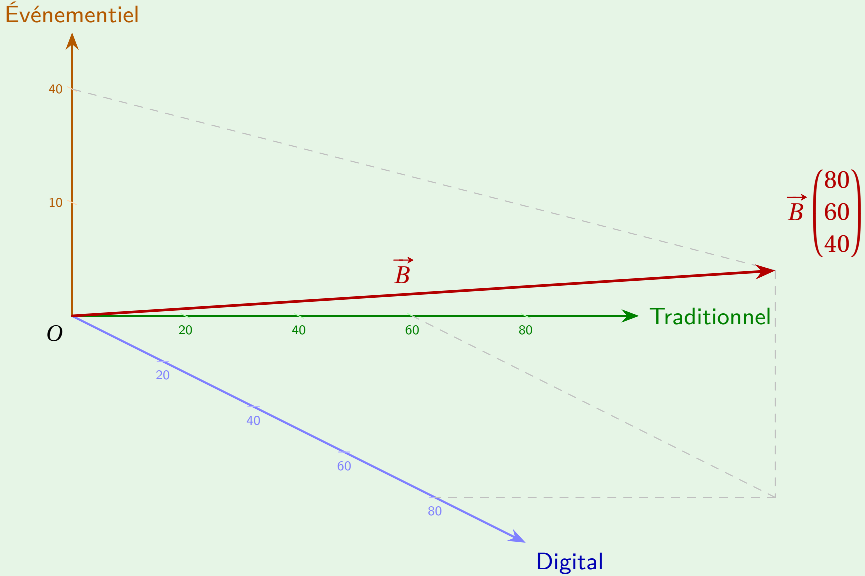
Le budget total alloué est de **180000**. La stratégie retenue se décompose comme suit :

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

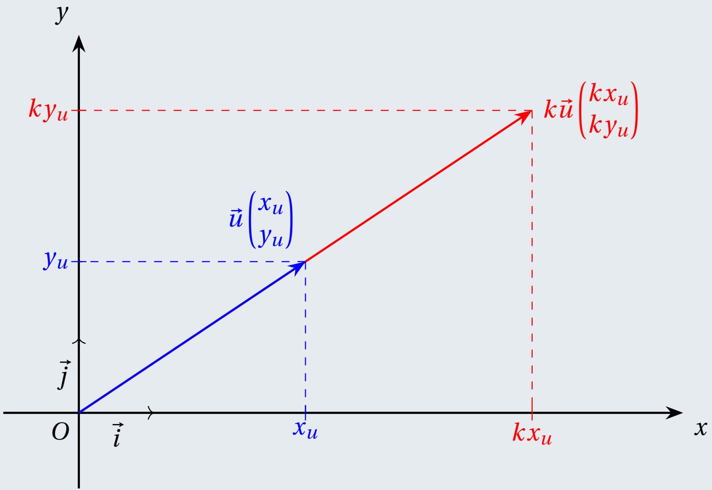
Travail demandé

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 ?

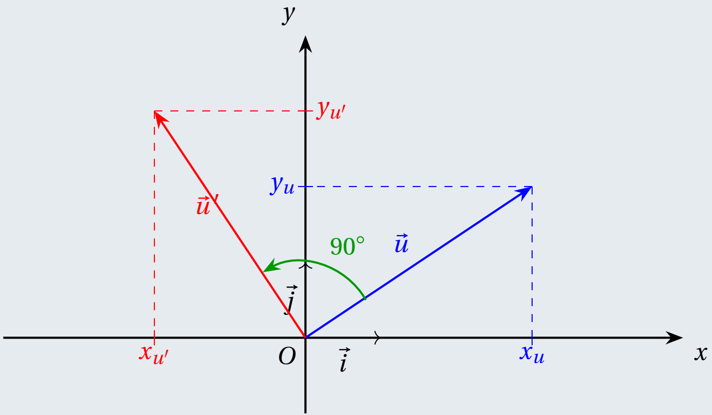
Application : Vecteur de Budget Marketing (en k€)



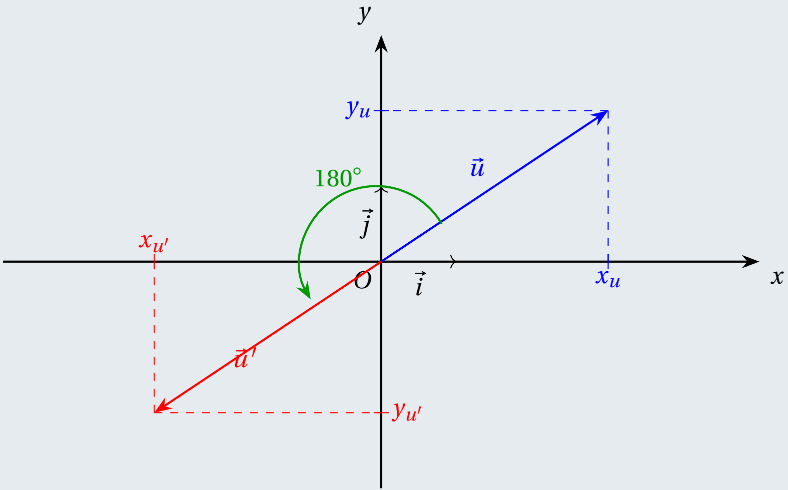
Multiplication d'un vecteur par un scalaire



Rotation d'un vecteur 90°



Rotation d'un vecteur 180°



Sommaire

Chapitre 4

Initiation au langage R

Section 4.1

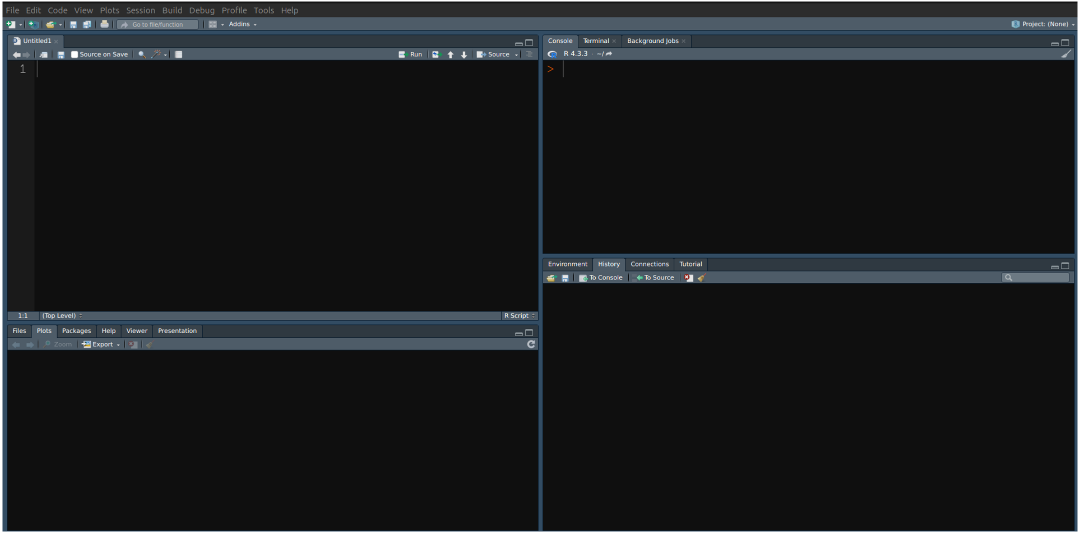
Présentation et installation

Qu'est-ce que R ?

- Langage de programmation dédié aux **statistiques** et à la **visualisation**.
- Open Source et gratuit.
- Communauté immense (CRAN).

Installation

- Télécharger **R** et **RStudio** sur `https://posit.co/download/rstudio-desktop`
- Installer **R** en premier.
- Installer l'interface graphique **RStudio** ensuite.


File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
Project: (None)
Untitled1
Source on Save
Run
Source
1
1:1
(Top Level)
R Script
Console
Terminal
Background Jobs
R 4.3.3 · ~/
>
Environment
History
Connections
Tutorial
To Console
To Source
Files
Plots
Packages
Help
Viewer
Presentation
Zoom
Export

Section 4.2

Variable, Vecteur et Data frame

Variables et Affectation

Pour affecter une valeur à une variable, on utilise la flèche <- (ou =).

```r
# Création de variables
x <- 10
y <- 5
z <- x + y

# Affichage
print(z)
z  # Taper le nom de la variable suffit aussi
```

Attention !

R est sensible à la casse : une majuscule est différente d'une minuscule (`Variable` ≠ `variable`).

Les Vecteurs

La structure de base en **R** est le **vecteur** (une liste de valeurs du même type). On utilise la fonction `c()` (concaténer).

```r
# Créer un vecteur de notes
notes <- c(12, 15, 9, 18, 14)

# Opérations sur tout le vecteur
notes + 2         # Ajoute 2 à chaque note
notes * 0.5       # Multiplie chaque note par 0.5
sum(notes)        # Somme totale
length(notes)     # Nombre d'éléments
```

Les Data Frames (Tableaux de données)

Pour l'analyse de données, on utilise des tableaux (lignes = observations, colonnes = variables).

```r
# Création manuelle d'un data frame
etudiants <- data.frame(
  nom = c("Ali", "Karim", "Fatima"),
  age = c(20, 22, 21),
  note = c(15, 12, 18)
)

# Voir les données
View(etudiants) # Ouvre une fenêtre type Excel
head(etudiants) # Affiche les premières lignes
```

Section 4.3

Calculatrice et statistique descriptive

Utiliser R comme une calculatrice

R peut faire des opérations mathématiques de base directement.

```r
# Opérations arithmétiques
2 + 2
10 / 5
3^2       # Puissance
sqrt(16) # Racine carrée
log(10)  # Logarithme naturel
```

Le symbole # permet d'écrire des commentaires (ignorés par **R**).

Statistiques Descriptives (Base R)

R possède des fonctions natives.

```r
notes <- c(12, 15, 9, 18, 14, 12, 10)

mean(notes)   # Moyenne
median(notes) # Médiane
sd(notes)     # Écart-type
min(notes)    # Minimum
max(notes)    # Maximum

# Résumé complet en une commande
summary(notes)
```

Utilisation de Packages

La force de **R** réside dans ses extensions (packages).

1. **Installation** (une seule fois) : Télécharge le code.
2. **Chargement** (à chaque session) : Rend les fonctions disponibles.

```r
# Installation du package 'dplyr' (manipulation de données)
install.packages("dplyr")

# Chargement du package
library(dplyr)

# Exemple d'utilisation (filtrage)
filter(etudiants, note > 14)
```

Statistiques avec un Package (Exemple)

Prenons l'exemple du package psych pour des statistiques avancées.

```r
install.packages("psych")
library(psych)

# Statistiques descriptives détaillées
describe(notes)

# Cela donne : mean, sd, median, trimmed, mad, min, max, skew, kurtosis...
```

Graphiques

Graphiques de Base (Base R)

R permet de tracer des graphiques très rapidement.

```r
# Données
x <- 1:10
y <- x^2

# Nuage de points
plot(x, y, main="Mon premier graphique",
     xlab="Axe X", ylab="Axe Y", col="blue")

# Histogramme
hist(notes, col="lightgreen", border="darkgreen")
```

Boxplot (Boîte de Tukey)

Utile pour voir la dispersion et les valeurs aberrantes.

```r
# Comparaison de deux groupes
groupe_A <- c(10, 12, 14, 15, 16)
groupe_B <- c(8, 9, 20, 22, 25)

boxplot(groupe_A, groupe_B,
        names=c("Groupe A", "Groupe B"),
        col=c("orange", "purple"),
        main="Comparaison des groupes")
```

Introduction à ggplot2 (Le standard)

Pour aller plus loin, le package ggplot2 est la référence.

```r
library(ggplot2)

ggplot(data = etudiants, aes(x = age, y = note)) +
  geom_point(size = 3, color = "red") +
  theme_minimal()
```

À retenir

- Base R pour de l'exploration rapide.
- ggplot2 pour des graphiques plus sophistiqués.

Section 4.6

Importer et Exporter des données

Pour travailler sur vos propres données, il faut savoir lire et écrire des fichiers.

```r
# 1. Vérifier le répertoire de travail
getwd()
# Pour le changer (attention aux slashes / )
# setwd("C:/Users/Nom/Documents/Cours_R")

# 2. Importer un fichier CSV dans un data frame
donnees <- read.csv("mon_fichier.csv", header = TRUE, sep = ";")

# 3. Vérifier l'import
head(donnees)    # Aperçu des premières lignes
str(donnees)     # Structure du data frame

# 4. Exporter un data frame vers un CSV
write.csv(donnees, "resultats_export.csv", row.names = FALSE)
```