

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

UNIVERSITE MOHAMED KHIDER, BISKRA

FACULTE des SCIENCES EXACTES et des SCIENCES de la NATURE et de la VIE

DEPARTEMENT DE MATHEMATIQUES



Mémoire présenté par

Machana Selsabil

En vue de l'obtention du Diplôme de

MASTER en Mathématiques

Option : **Statistique**

Titre

Mesures d'association

Membres du Comité d'Examen

Dr. Benameur Sana	UMKB	Présidente
Pr. MERAGHNI Djamel	UMKB	Encadreur
Dr. Ouanoughi Yasmina	UMKB	Examinatrice

Juin 2022

Dédicace

Je dédie cet humble travail,

A mon cher père Bachir.

*A la fontaine qui ne se lasse pas de donner, et qui a tissé mon bonheur avec les fils de
son coeur, ma chère mère Sabah.*

A celles dont l'amour coule dans mes veines et mon coeur par leur souvenir, mes soeurs :

Radja et ses deux fils Rawan et Mehیار, Chahrazed et Chaima.

A mes frères Rabah et Kaiss.

A mes oncles Ibrahim et Abdesslam.

A mon oncle maternel Ali.

Avec une dédicace spéciale,

*A la mémoire de mon oncle Ali que Allah lui fasse miséricorde
et le place dans son paradis.*

A tous les membres des familles Machana et Ouani.

*A tous mes collègues qui m'ont accompagnée dans mon parcours académique et m'ont
aidée à développer ce travail.*

Merci à toutes et à tous sans exception.

Selsabil Machana

REMERCIEMENTS

Au nom de Allah, le Très Miséricordieux.

Prière et paix soient sur le plus honorable des prophètes et messagers.

Je remercie Allah Le Tout-Puissant de m'avoir conciliée dans cet humble travail.

8J'adresse mes plus pofonds remerciements et tout mon respect au superviseur

Prof. Djamel MERAGHNI.

Ses précieux conseils et encouragements m'ont aidée énormément à réaliser ce mémoire.

Je remercie également les membres du jury Mesdames S. Benameur et Y. Ouanoughi

pour le temps qu'elles ont consacré pour juger ce travail.

Je n'oublie pas non plus de remercier tous les enseignants du département de Mathématiques qui m'ont enseignée et aidée dans tous mon parcours académique.

Je dois également rendre hommage à mon père qui m'a aidée et subvenu à tous mes besoins pour accomplir mon travail.

Je demande à Allah Le Tout-Puissant de le récompenser.

Que la paix, la miséricorde et les bénédictions de Allah soient sur vous tous.

Selsabil Machana

Table des matières

Dédicace	i
Remerciements	ii
Table des matières	iii
Liste des figures	v
Liste des tableaux	vi
Introduction	1
1 Couples aléatoires	3
1.1 Définition	3
1.1.1 Variable aléatoire	3
1.1.2 Couple aléatoire	4
1.2 Loi de probabilité	5
1.2.1 Fonction de répartition	5
1.2.2 Masse ou densité de probabilité	5
1.2.3 Lois marginales	6
1.2.4 Lois conditionnelles	7
1.2.5 Indépendance	8
1.3 Paramètres statistiques	9

1.3.1	Espérance	9
1.3.2	Matrice de covariance	9
1.3.3	Matrice de corrélation	10
1.4	Exemples	12
1.4.1	Couple discret	12
1.4.2	Couple continu	13
2	Coefficients de corrélation	18
2.1	Nuage de points	18
2.2	Coefficient de corrélation de Pearson	19
2.2.1	Définition et interprétation	19
2.2.2	Coefficient de corrélation empirique	21
2.3	Coefficient de corrélation de Spearman	22
2.4	Coefficient de corrélation de Kendall	25
2.4.1	Concept de concordance	25
2.4.2	Tau de Kendall	26
2.4.3	Fonction R de calcul des coefficients de corrélation	27
	Conclusion	29
	Annexe A : Codes R	31
	Annexe B : Abréviations et Notations	33

Table des figures

2.1	Liaison linéaire positive (gauche), absence de liaison (centre) et liaison linéaire négative (droite).	19
2.2	Représentation de l'argent hebdomadaire moyen attribué à 10 en fonction de leur âge.	20
2.3	Notes attribuées par 2 juges à 10 sportifs	25

Liste des tableaux

1.1	Distribution de probabilité d'un couple aléatoire avec ses marginales	12
1.2	Distribution conditionnelle de X sachant Y	12
1.3	Distribution conditionnelle de Y sachant X	12
2.1	Age et argent de poche hebdomadaire de 10 enfants	19
2.2	Paramètres statistiques des mesures de taille et de poids de 17 individus . .	22
2.3	Notes attribuées par deux juges à 10 sportifs lors d'une compétition	24
2.4	Résultats relatifs au calcul du coefficient de corrélation de Spearman	24
2.5	Exemple de calcul du corrélation de Kendal	27

Introduction

La statistique est une science mathématique qui étudie diverses méthodes de description, d'estimation et d'inférence sur des phénomènes aléatoires. Elle s'applique dans des domaines très variés de la vie socio-économique. Parmi les axes de la statistique, il y a l'étude des liaisons (ou effets) qui peuvent exister entre deux ou plusieurs sujets. Ces liaisons sont évaluées par ce qu'on appelle mesures d'association ou coefficients de corrélation. A titre d'exemple, en épidémiologie, on cherche à savoir si l'exposition à un facteur de risque conduit à l'apparition de symptômes de la maladie. En sociologie, on peut s'intéresser à l'effet de la profession des parents sur le choix de la branche d'études universitaires de leurs enfants.

Dans l'analyse statistique des données multivariées, la valeur du coefficient de corrélation indique le degré de la relation entre deux ou plusieurs variables. La valeur calculée indique si la relation est directe (valeur positive) ou inverse (valeur négative). Un coefficient égal à zéro n'implique pas nécessairement l'indépendance des deux variables car d'autres types de coefficients sont possibles. Parfois, le coefficient de corrélation est examiné en le comparant avec une norme relative au sujet d'étude. La valeur de ce coefficient est comprise entre -1 et 1 .

Dans ce mémoire, j'ai essayé de mettre en avant le concept d'association en m'intéressant au rôle fondamental du coefficient de corrélation, à la finalité de son utilisation et à la manière de l'appliquer. Il existe plusieurs types de coefficients de corrélation dont les plus connus sont celui de Pearson, le rho de Spearman et le tau de Kendall. Le choix du

coefficient à utiliser dépend du type de données ou de variables étudiées ou analysées. Ce mémoire se compose de deux chapitres :

- Chapitre 1 : consacré aux définitions et caractéristiques fondamentales des couples aléatoires.
- Chapitre 2 : traite de l'explication et de la simplification du concept des différents coefficients de corrélation et de leurs implications, des méthodes de leur calcul, de leurs équations mathématiques ou des lois de leur calcul, et les utilise avec des exemples illustratifs simplifiés en présentant le graphique appelé la diffusion forme de deux variables liées à une relation (direct ou inverse).

Enfin, je tiens à souligner que le contenu de ce mémoire s'inspire essentiellement des références [5], [4] et [7], et que les études de simulation, avec calculs et graphes, sont réalisées en utilisant le logiciel de traitement statistique R [3].

Chapitre 1

Couples aléatoires

Dans ce chapitre, on généralise les définitions et propriétés relatives à une variable aléatoire (v.a) réelle au cas où on dispose de deux v.a. On parle alors de couple aléatoire réel ou de vecteur aléatoire de dimension 2.

1.1 Définition

Soit un (Ω, \mathcal{F}, P) un espace de probabilités.

1.1.1 Variable aléatoire

Une v.a réelle est une application mesurable X définie sur l'ensemble fondamental Ω à valeurs dans \mathbb{R} :

$$\begin{aligned} X : \Omega &\rightarrow \mathbb{R} \\ \omega &\mapsto X(\omega) \end{aligned}$$

Selon l'ensemble des valeurs $X(\Omega)$, on distingue deux types de v.a.

1. v.a discrète : concerne les sujet de dénombrement. Les v.a discrètes usuelles sont Bernoulli (succès-échec), binomiale (nombre de succès lors d'un certain nombre

d'épreuves indépendantes), Poisson (nombre d'arrivées pendant une certaine période),...

2. v.a continue : traite les problèmes de mesure. Les plus connues sont les variables de Gauss (normale), de Pearson (Khi-deux), de Student, de Fisher ainsi que la v.a exponentielle.

Exemple 1.1.1

1. Nombre d'étudiants absents pendant un mois : Poisson.
2. Temps d'attente dans un certain service administratif : exponentielle.

1.1.2 Couple aléatoire

Un couple aléatoire réel (X, Y) est une application mesurable de Ω dans \mathbb{R}^2 . Il s'écrit sous la forme d'un vecteur colonne $(X, Y)^t$ dont les composantes X et Y sont des v.a réelles définies sur l'ensemble Ω .

La notion de couple est nécessaire dans l'étude du comportement simultané des deux v.a X et Y .

Remarque 1.1.1

1. Cette définition se généralise au cas où l'ensemble d'arrivée est \mathbb{R}^n , avec $n \geq 2$. On parle alors de vecteur aléatoire réel de dimension n : $(X_1, X_2, \dots, X_n)^t$.
2. Lorsque les v.a X et Y sont toutes les deux discrètes, on dira que le couple (X, Y) est discret. Si les deux v.a X et Y sont continues, on dira que (X, Y) est continu.

1. Etude de la relation entre le nombre b'absences d'un étudiant et le nombre de modules acquis.
2. Etude de l'influence de la taille d'une personne sur son poids.

1.2 Loi de probabilité

Soit (X, Y) un couple aléatoire. Sa loi de probabilité est déterminée par la fonction de répartition. Dans le cas continu, cette loi peut aussi être obtenue par la fonction de densité et dans le cas discret on utilise les probabilités individuelles.

1.2.1 Fonction de répartition

Définition 1.2.1 (fonction de répartition conjointe) *On appelle fonction de répartition conjointe de (X, Y) , qu'on note F , la fonction définie sur \mathbb{R}^2 par*

$$F(x, y) := P(X \leq x, Y \leq y).$$

Propriété 1.2.1 *La fonction de répartition d'un couple aléatoire (X, Y) jouit des propriétés suivantes :*

- $\forall (x, y) \in \mathbb{R}^2, 0 \leq F(x, y) \leq 1$.
- F est croissante par rapport à chacune des variables x et y .
- $\lim_{x, y \rightarrow +\infty} F(x, y) = 1$ et $\lim_{x, y \rightarrow -\infty} F(x, y) = 0$.
- F est continue à droite en tout point (x_0, y_0) de \mathbb{R}^2 : $\lim_{(x, y) \xrightarrow{\geq} (x_0, y_0)} F(x, y) = F(x_0, y_0)$.

1.2.2 Masse ou densité de probabilité

Cas discret

Dans ce cas, le couple (X, Y) prend des valeurs $\{(x_i, y_j), 1 \leq i, j \leq n\}$, où $n \in \mathbb{N}^*$. La masse de probabilité conjointe, que l'on notera par P , donne les probabilités associées à ces couples de valeurs. Donc, pour tout couple (x_i, y_j) , on a

$$p_{ij} = P(x_i, y_j) := P(X = x_i, Y = y_j).$$

Elle vérifie les deux conditions suivantes :

- $p_{ij} \geq 0, i, j = 1, \dots, n.$
- $\sum_{i=1}^n \sum_{j=1}^n p_{ij} = 1.$

Cas continu

Dans ce cas, la fonction de répartition F est dérivable par rapport aux deux variables. La densité de probabilité conjointe, notée par f , est alors définie sur \mathbb{R}^2 par

$$f(x, y) := \frac{\partial^2 F(x, y)}{\partial x \partial y}.$$

Elle vérifie les deux conditions suivantes :

- $f(x, y) \geq 0, \forall (x, y) \in \mathbb{R}^2.$
- $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XY}(x, y) dx dy = 1.$

Remarque 1.2.1 *La fonction de répartition du couple (X, Y) se calcule par*

$$F(x, y) = \begin{cases} \sum_{u \leq x} \sum_{v \leq y} P(u, v) & \text{si } (X, Y) \text{ est discret,} \\ \int_{-\infty}^x \int_{-\infty}^y f(u, v) du dv & \text{si } (X, Y) \text{ est continu.} \end{cases}$$

1.2.3 Lois marginales

Soit (X, Y) un couple aléatoire de fonction de répartition F . Quand on s'intéresse à un évènement relatif à l'une des deux v.a X ou Y indépendamment de l'autre, on définit ce que l'on appelle lois marginales.

Définition 1.2.2 (fonctions de répartition marginales) *Les fonction de répartition marginales de X et Y sont respectivement définie sur \mathbb{R} par*

$$F_X(x) := P(X \leq x, Y \in \mathbb{R}) = \lim_{y \rightarrow +\infty} F(x, y),$$

et

$$F_Y(y) := P(X \in \mathbb{R}, Y \leq y) = \lim_{x \rightarrow +\infty} F(x, y).$$

Les masses et densités marginales sont alors égales à

$$P_X(x_i) = \sum_{j=1}^n P(x_i, y_j) \text{ et } P_Y(y_j) = \sum_{i=1}^n P(x_i, y_j),$$

et

$$f_X(x) = \int_{-\infty}^{+\infty} f(x, y) dy \text{ et } f_Y(y) = \int_{-\infty}^{+\infty} f(x, y) dx.$$

1.2.4 Lois conditionnelles

On s'intéresse maintenant à la loi de probabilité de l'une des deux v.a sachant que l'autre est connue.

Cas discret

La loi de X sachant Y est définie par

$$p_{i|j} = P(X = x_i | Y = y_j) := \frac{P(X = x_i, Y = y_j)}{P(Y = y_j)}, \quad i = 1, \dots, n.$$

De même, la loi de Y sachant X est définie par

$$p_{j|i} = P(Y = y_j | X = x_i) := \frac{P(X = x_i, Y = y_j)}{P(X = x_i)}, \quad j = 1, \dots, n.$$

Remarque 1.2.2 On vérifie que ce sont bien des lois de probabilité. En effet,

1. Sur $\Omega_X = \{x_i, i = 1, \dots, n\}$, on a, pour j fixé,

$$\sum_{i=1}^n P(X = x_i | Y = y_j) = \frac{1}{P(Y = y_j)} \sum_{i=1}^n P(X = x_i, Y = y_j) = 1.$$

2. Sur $\Omega_Y = \{y_j, j = 1, \dots, n\}$, on a, pour i fixé,

$$\sum_{j=1}^n P(Y = y_j | X = x_i) = \frac{1}{P(X = x_i)} \sum_{j=1}^n P(X = x_i, Y = y_j) = 1.$$

Cas continu

Les lois de X sachant Y et de Y sachant X sont définies par les densités conditionnelles respectives

$$f_{X|Y}(x) := \frac{f(x, y)}{f_Y(y)} \text{ et } f_{Y|X}(y) := \frac{f(x, y)}{f_X(x)}.$$

Remarque 1.2.3 On vérifie que ce sont bien des densités de probabilité. En effet, on a

$$\int_{-\infty}^{\infty} f_{X|Y}(x) dx = \frac{1}{f_Y(y)} \int_{-\infty}^{\infty} f(x, y) dx = 1.$$

De même pour l'autre densité conditionnelle.

1.2.5 Indépendance

Définition 1.2.3 Deux v.a X et Y sont dites indépendantes si tout événement relatif à X est indépendant de tout événement défini par Y .

Remarque 1.2.4 L'indépendance se traduit par l'égalité entre la loi conditionnelle et la loi marginale, c-à-d

$$p_{i|j} = P(X = x_i) \text{ ou } p_{j|i} = P(Y = y_j),$$

ou

$$f_{X|Y}(x) = f_X(x) \text{ ou } f_{Y|X}(y) = f_Y(y).$$

En utilisant les définitions des lois conditionnelles, on obtient la condition qui permet de définir l'indépendance entre deux v.a X et Y :

$$p_{ij} = P(X = x_i)P(Y = y_j) \text{ ou } f(x, y) = f_X(x)f_Y(y).$$

1.3 Paramètres statistiques

Soit $V = (X, Y)^t$ un couple aléatoire.

1.3.1 Espérance

On appelle espérance ou moyenne de V , $E(V)$ en abrégé, le vecteur des moyennes des deux v.a X et Y . On la note généralement par μ :

$$\mu = E(V) := \begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix},$$

où, pour une v.a Z , on a

$$\mu_Z = E(Z) := \begin{cases} \sum_i z_i P(Z = z_i) & \text{si } Z \text{ est discrète,} \\ \int z f_Z(z) dz & \text{si } Z \text{ est continue.} \end{cases}$$

Remarque 1.3.1

1. *L'espérance est linéaire* : $E(U + \lambda V) = E(U) + \lambda E(V)$, $\lambda \in \mathbb{R}$.
2. *Si X et Y sont indépendantes, alors*

$$E(XY) = E(X)E(Y).$$

1.3.2 Matrice de covariance

La matrice de covariance de V , $Cov(V)$ en abrégé, est une matrice carrée symétrique d'ordre 2, généralement notée par Σ :

$$\Sigma = Cov(V) := \begin{bmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_Y^2 \end{bmatrix}, \quad (1.1)$$

où

$$\sigma_{XY} = \text{cov}(X, Y) := E\{(X - E(X))(Y - E(Y))\} = E(XY) - E(X)E(Y),$$

représente la covariance entre les deux v.a X et Y , et

$$\sigma_Z^2 = \sigma_{ZZ},$$

représente la variance d'une v.a Z (sa racine carrée σ_Z est appelée écart type de Z).

Remarque 1.3.2

1. Σ est une matrice définie positive.
2. $\Sigma = E(VV') - \mu\mu'$.
3. Si X et Y sont indépendantes, alors Σ est une matrice diagonale. En effet, on a

$$\sigma_{XY} = E(XY) - E(X)E(Y) = E(X)E(Y) - E(X)E(Y) = 0.$$

La réciproque n'est pas toujours vraie.

1.3.3 Matrice de corrélation

La matrice de corrélation de V , $Cor(V)$ en abrégé, est une matrice carrée symétrique d'ordre 2, définie par

$$Cor(V) := \begin{bmatrix} 1 & \theta \\ \theta & 1 \end{bmatrix}, \quad (1.2)$$

où

$$\theta = \text{cor}(X, Y) := \frac{\sigma_{XY}}{\sigma_X \sigma_Y}, \quad (1.3)$$

représente le coefficient de corrélation linéaire des deux v.a X et Y .

Remarque 1.3.3

1. $Cor(V)$ est une matrice définie positive.
2. Si X et Y sont indépendantes, alors $Cor(V)$ est égale à la matrice identité. La réciproque n'est pas toujours vraie.
3. En termes du coefficient de corrélation, on a

$$\Sigma = \begin{bmatrix} \sigma_X^2 & \theta\sigma_X\sigma_Y \\ \theta\sigma_X\sigma_Y & \sigma_Y^2 \end{bmatrix}. \quad (1.4)$$

Remarque 1.3.4 On peut aussi définir les paramètres statistiques par rapport aux lois conditionnelles.

Définition 1.3.1 (espérance conditionnelle) On appelle espérance de Y sachant que $X = x$ et on note $E(Y | X = x)$ la quantité

$$E(Y | X = x) := \begin{cases} \sum_{j \in J} y_j P(Y = y_j | X = x) = \sum_{j \in J} y_j P_{j|i}, & \text{cas discret} \\ \int_{\mathbb{R}} y f_{Y|X}(x, y) dy, & \text{cas continu} \end{cases}.$$

Définition 1.3.2 (variance conditionnelle) On appelle variance de Y sachant que $X = x$ et on note $V(Y | X = x)$ la quantité

$$V(Y | X = x) := E \{ [Y - E(Y | X = x)]^2 | X = x \}.$$

Pour plus de détails sur l'espérance et la variance conditionnelles, voir [7, pages 71-75].

$Y \backslash X$	-2	0	2	marge
-1	0.1	0.2	0.1	0.4
2	0.2	0.2	0.2	0.6
marge	0.3	0.4	0.3	1

TAB. 1.1 – Distribution de probabilité d’un couple aléatoire avec ses marginales

	-2	0	2
$X Y = -1$	1/4	1/2	1/4
$X Y = 2$	1/3	1/3	1/3

TAB. 1.2 – Distribution conditionnelle de X sachant Y

1.4 Exemples

1.4.1 Couple discret

Cet exemple se trouve dans [5, page 109]. Soit $V = (X, Y)$ un couple aléatoire dont la distribution est donnée par le tableau 1.1.

Lois conditionnelles

Les Lois conditionnelles sont résumées dans les tableaux 1.2 et 1.3.

Paramètres statistiques

D’après la définition de l’espérance d’un couple aléatoire, on a

$$\mu = E(V) = \begin{pmatrix} 0 \\ 0.8 \end{pmatrix}.$$

On a

$$Var(X) = 2.4, \quad Var(Y) = 3.44 \text{ et } Cov(X, Y) = 0.$$

Alors ,d’après (1.1) et (1.2), les matrices de covariance et corrélation de V sont respecti-

	-1	2
$Y X = -2$	1/3	2/3
$Y X = 0$	1/2	1/2
$Y X = 2$	1/3	2/3

TAB. 1.3 – Distribution conditionnelle de Y sachant X

vement

$$\Sigma = \begin{bmatrix} 2.4 & 0 \\ 0 & 3.44 \end{bmatrix} \text{ et } \text{Cor}(V) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Remarque 1.4.1 Bien que $\text{Cov}(X, Y) = \text{Cor}(X, Y) = 0$, les deux v.a ne sont pas indépendantes (voir les lois marginales et conditionnelles). On a ici un contre exemple à la remarque 3 de (1.3.2)(ou la remarque 2 de (1.3.3)).

1.4.2 Couple continu

Exemple 1 : loi normale bidimensionnelle $V \sim \mathcal{N}_2(\mu, \Sigma)$ ([7, page 90])

$$f(x, y) = \frac{1}{2\pi\sqrt{\det \Sigma}} \exp \left\{ -\frac{1}{2} (v - \mu)^t \Sigma^{-1} (v - \mu) \right\}$$

En utilisant la forme (1.4), on a $\det \Sigma = \sigma_X^2 \sigma_Y^2 (1 - \theta^2)$ et

$$\Sigma^{-1} = \frac{1}{\sigma_X^2 \sigma_Y^2 (1 - \theta^2)} \begin{bmatrix} \sigma_Y^2 & -\theta \sigma_X \sigma_Y \\ -\theta \sigma_X \sigma_Y & \sigma_X^2 \end{bmatrix}.$$

La densité de probabilité s'écrit alors

$$f(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\theta^2}} \times \exp \left\{ -\frac{1}{2((1-\theta^2))} \left[\left(\frac{x-\mu_X}{\sigma_X} \right)^2 - 2\theta \frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} + \left(\frac{y-\mu_Y}{\sigma_Y} \right)^2 \right] \right\}.$$

Si les v.a X et Y sont centrées et réduites, alors $V \sim \mathcal{N}_2(0, I_2)$ et la densitéo conjointe devient

$$f(x, y) = \frac{1}{2\pi} \exp \left\{ -\frac{1}{2} [x^2 + y^2] \right\}.$$

Dans ce cas particulier, les densités marginales sont

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\right\} \text{ et } f_Y(y) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{y^2}{2}\right\},$$

et les densités conditionnelles sont

$$f_{X|Y}(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\right\} \text{ et } f_{Y|X}(y) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{y^2}{2}\right\}.$$

On constate que $f_{X|Y}(x) = f_X(x)$ et $f_{Y|X}(y) = f_Y(y)$ sur \mathbb{R} . Ceci entraîne que

$$f(x, y) = f_X(x)f_Y(y)$$

et par conséquent les deux v.a X et Y sont indépendantes. D'où le résultat de la Proposition 1.4.1.

Puisque X et Y sont de même distribution $\mathcal{N}(0, 1)$, alors l'espérance et les matrices de covariance et de corrélation de V sont

$$E(V) = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \text{ et } Cov(V) = Cor(V) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Proposition 1.4.1 *Soit (X, Y) un couple aléatoire Gaussien. Alors les deux v.a X et Y sont indépendantes si et seulement si elles sont non corrélées.*

Exemple 2 : cet exemple est extrait du cours [2]. On suppose que la densité de probabilité du couple V est définie sur \mathbb{R}^2 par

$$f(x, y) = \begin{cases} 2 \exp\{-(x+y)\} & \text{si } 0 \leq x \leq y \\ 0 & \text{sinon} \end{cases}$$

La densité marginale de X est définie, pour $x \geq 0$, par

$$\begin{aligned} f_X(x) &= \int_x^\infty f(x, y) dy = 2 \exp(-x) \int_x^\infty \exp(-y) dy \\ &= -2 \exp(-x) \exp(-y) \Big|_x^\infty = 2 \exp(-2x), \end{aligned}$$

c-à-d

$$f_X(x) = 2 \exp(-2x) \mathbf{1}_{[0, +\infty[}(x).$$

Par le même raisonnement, on obtient la densité marginale de Y qui est définie pour $y \geq 0$ par

$$\begin{aligned} f_Y(y) &= \int_0^y f(x, y) dx = 2 \exp(-y) \int_0^y \exp(-x) dx \\ &= 2 \exp(-y) (-\exp(-x) \Big|_0^y) = 2 \exp(-y) (1 - \exp(-y)), \end{aligned}$$

c-à-d

$$f_Y(y) = 2 \exp(-y) (1 - \exp(-y)) \mathbf{1}_{[0, +\infty[}(y).$$

La densité conditionnelle de Y (sachant $X = x \geq 0$) est

$$f_{Y|X}(y) = \frac{2 \exp(-(x+y)) \mathbf{1}_{\{0 \leq x \leq y\}}}{2 \exp(-2x)} = \exp(x-y) \mathbf{1}_{\{0 \leq x \leq y\}}.$$

De la même façon, on trouve la densité conditionnelle de X (sachant $Y = y \geq 0$) est

$$f_{X|Y}(x) = \frac{2 \exp(-(x+y)) \mathbf{1}_{\{0 \leq x \leq y\}}}{2 \exp(-y) (1 - \exp(-y))} = \frac{\exp(-x)}{(1 - \exp(-y))} \mathbf{1}_{\{0 \leq x \leq y\}}.$$

Remarque 1.4.2

- On constate que $f_X(x)$ est la densité de la loi exponentielle de paramètre 2, c-à-d la v.a $X \sim \mathcal{E}(2)$.

- *Le produit des densités marginales est*

$$f_X(x)f_Y(y) = 4 \exp(-(2x + y))(1 - \exp(-y))\mathbf{1}_{[0,+\infty[^2}(x, y).$$

Il est différent de la densité conjointe $f(x, y)$, alors les v.a X et Y ne sont pas indépendantes.

Espérance de V : elle est égale à

$$E(V) = \begin{pmatrix} E(X) \\ E(Y) \end{pmatrix} = \begin{pmatrix} \frac{1}{2} \\ \frac{3}{2} \end{pmatrix}.$$

En effet, puisque $X \sim \mathcal{E}(2)$ alors $E(X) = 1/2$. Pour la v.a Y , on a

$$\begin{aligned} E(Y) &= \int_{\mathbb{R}} y f_Y(y) dy = 2 \int_0^{\infty} y (\exp(-y) - \exp(-2y)) dy \\ &= 2 \left[\int_0^{\infty} y \exp(-y) dy - \int_0^{\infty} y \exp(-2y) dy \right] =: 2 [I_1 - I_2]. \end{aligned}$$

Les valeurs

$$I_1 = 1 \text{ et } I_2 = \frac{1}{4},$$

sont obtenues par des intégrations par parties. D'où $E(Y) = 3/2$.

Matrice de covariance de V

On a $X \sim \mathcal{E}(2) \longrightarrow \text{Var}(X) = 1/2^2 = 1/4$; $\text{Var}(Y) = 2$ et $\text{Cov}(X, Y) = 1/4$. D'où

$$\text{Cov}(V) = \begin{bmatrix} \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & 2 \end{bmatrix}.$$

Matrice de corrélation de V

D'après (1.3) on a $\theta = \frac{1}{4} / \left(\frac{1}{2} \times \sqrt{2}\right) = \sqrt{2}/4$. Donc de 1.2, on a

$$\text{Cor}(V) = \begin{bmatrix} 1 & \frac{\sqrt{2}}{4} \\ \frac{\sqrt{2}}{4} & 1 \end{bmatrix}.$$

Chapitre 2

Coefficients de corrélation

Ce chapitre est consacré aux mesures d'association les plus populaires (Pearson, Spearman, Kendall). Une attention particulière est faite au coefficient de corrélation linéaire de Pearson car c'est le plus important. Il est, entre autres, utilisé en régression linéaire lors de l'ajustement d'une variable par rapport à une autre par une relation affine.

2.1 Nuage de points

Une distribution statistique bidimensionnelle (X, Y) peut être représentée graphiquement sous la forme d'un nuage de points dans le plan. Chaque point a pour abscisse la valeur de la v.a X et pour ordonnée la valeur de la v.a Y . L'allure globale du nuage permet de caractériser la nature de la liaison qui peut exister entre les deux v.a. Les formes les plus courantes sont représentées par la figure 2.1. La liaison linéaire positive signifie que les deux v.a X et Y évoluent dans le même sens alors que la liaison linéaire négative est synonyme d'une évolution en sens inverses. L'absence de liaison veut dire que la valeur de X ne donne aucune indication sur celle de Y et inversement.

Exemple 2.1.1 ([1, page 335]) *On s'intéresse à l'argent de poche donné à 10 jeunes dont l'âge est compris entre 11 et 16 ans. On a mesuré pour chaque individu de l'échan-*

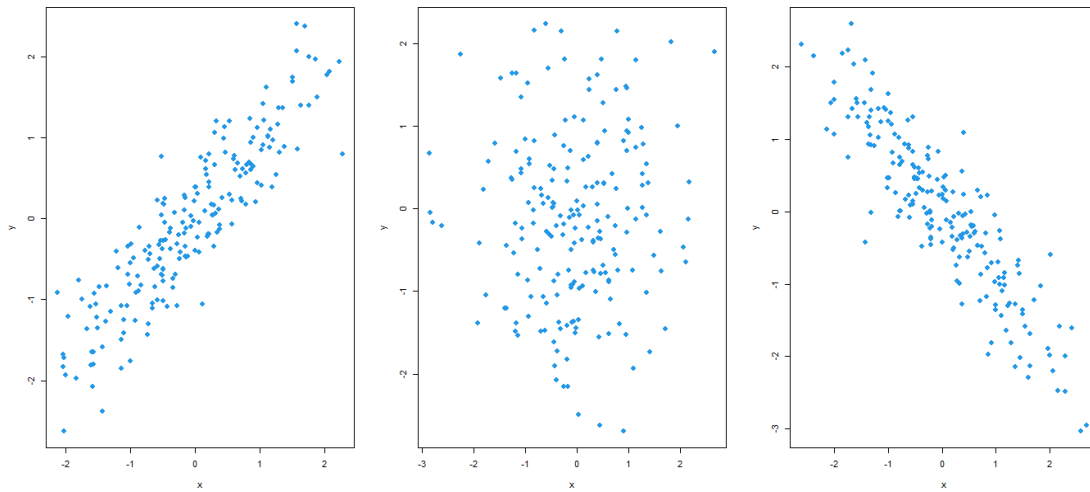


FIG. 2.1 – Liaison linéaire positive (gauche), absence de liaison (centre) et liaison linéaire négative (droite).

tillon, l'âge (x) et le montant hebdomadaire moyen (y) (sur une année). La série observée est donnée dans le tableau 2.1 et représentée graphiquement par la figure 2.2.

x	12	12	15	14	16	14	12	13	11	11
y	4.1	3.4	11.3	10.2	11.5	7.2	6	7.8	3.5	3

TAB. 2.1 – Age et argent de poche hebdomadaire de 10 enfants

Sur la figure 2.2, on voit que l'argent hebdomadaire moyen donné aux enfants a tendance à varier dans le même sens que leur âge. Il s'agit donc d'une liaison linéaire positive entre les deux variables.

2.2 Coefficient de corrélation de Pearson

2.2.1 Définition et interprétation

Définition 2.2.1 (Pearson) *Le coefficient de corrélation linéaire de Pearson entre deux v.a X et Y est une normalisation de leur covariance par le produit de leur écarts-types. Il est défini par la formule (1.3).*

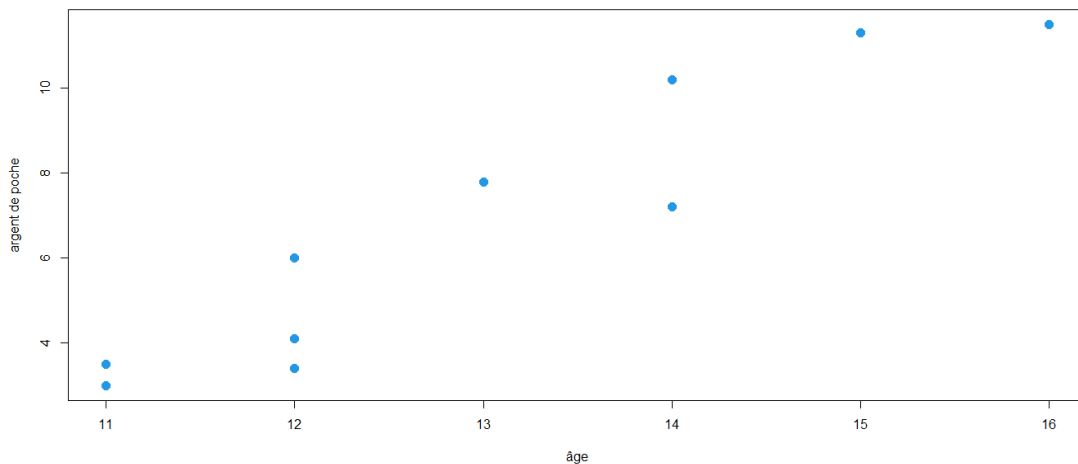


FIG. 2.2 – Représentation de l’argent hebdomadaire moyen attribué à 10 en fonction de leur âge.

Propriété 2.2.1 *La corrélation possède les propriétés suivantes :*

1. On a $-1 \leq \theta \leq 1$ ou, en d’autres termes, $|\theta| \leq 1$.
2. Le coefficient θ est indépendant des unités de mesure.
3. La corrélation d’une variable avec elle même est $Cor(X, X) = 1$.
4. Si X et Y sont indépendantes, alors $\theta = 0$. La réciproque est fautive, sauf lorsque le couple (X, Y) suit une loi normale bivariée.

Interprétation

Selon la valeur du coefficient θ , on a les remarques suivantes :

- Si θ est proche de 1, il existe une forte liaison linéaire positive entre X et Y (voir le graphe de gauche de la figure 2.1).
- Si θ est proche de 0, il n’y a pas de liaison linéaire entre X et Y (voir le graphe du centre de la figure 2.1).
- Si θ est proche de -1 , il existe une forte liaison linéaire négative entre X et Y (voir le graphe de droite de la figure 2.1).

Remarque 2.2.1 *Le signe de r indique le sens de la liaison tandis que sa valeur absolue indique l'intensité de la liaison.*

2.2.2 Coefficient de corrélation empirique

Définition 2.2.2 *Le coefficient de corrélation linéaire empirique est défini, à partir d'un échantillon $\{(X_1, Y_1), \dots, (X_n, Y_n)\}$, de taille $n \geq 1$, du couple (X, Y) , par*

$$R := \frac{S_{XY}}{S_X S_Y},$$

où

$$S_{XY} := \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}),$$

représente la covariance empirique et

$$S_X^2 := \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \text{ et } S_Y^2 := \frac{1}{n} \sum_{i=1}^n (Y_i - \bar{Y})^2,$$

sont les variances empiriques respectives de X et Y , avec

$$\bar{X} := \frac{1}{n} \sum_{i=1}^n X_i \text{ et } \bar{Y} := \frac{1}{n} \sum_{i=1}^n Y_i,$$

étant les moyennes empiriques de X et Y respectivement.

Remarque 2.2.2

1. Les v.a R , S_{XY} , S_X^2 et S_Y^2 sont les estimateurs naturels des paramètres θ , σ_{XY} , σ_X^2 et σ_Y^2 respectivement.
2. R est un estimateur asymptotiquement non biaisé et convergent de θ . On a

$$E(R) = \theta - \frac{\theta(1-\theta^2)}{2n} \text{ et } \text{Var}(R) = \frac{(1-\theta^2)^2}{n}.$$

3. Le coefficient de corrélation empirique R vérifient les mêmes propriétés que son équivalent théorique θ , avec la même interprétation.

4. La valeur observée de la v.a R est

$$r := \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}.$$

Exemple 2.2.1 Dans l'exemple 2.1.1, on a

	moyenne	variance	covariance
X_i	13	2.88	5.33
Y_I	6.8	11.09	

TAB. 2.2 – Paramètres statistiques des mesures de taille et de poids de 17 individus

La valeur du coefficient de corrélation est donc 0.94. Ceci indique qu'il y a une liaison positive très forte entre l'argent qu'on donne aux enfants et leur âge.

2.3 Coefficient de corrélation de Spearman

Ce coefficient, aussi appelé rho de Spearman, a été développé par Charles Spearman en 1904. Il permet d'analyser les liaisons entre non pas les observations mais leurs rangs. Il est égal au coefficient de corrélation de Pearson appliqué sur les rangs. Dans ce qui suit, on va se concentrer sur la version empirique de ce coefficient. Soit $\{(X_i, Y_i)\}_{1 \leq i \leq n}$ un échantillon, de taille $n \geq 1$, d'un couple aléatoire (X, Y) . On note les rangs observés de X et de Y par

$$R_i = \text{rang}(X_i) \text{ et } S_i = \text{rang}(Y_i), \quad i = 1, \dots, n.$$

Définition 2.3.1 (rho de spearman) *Le coefficient de corrélation de Spearman est donné*

(comme v.a) par

$$\rho := \text{Cor}(R, S) = \frac{\sum_{i=1}^n (R_i - \bar{R})(S_i - \bar{S})}{\sqrt{\sum_{i=1}^n (R_i - \bar{R})^2} \sqrt{\sum_{i=1}^n (S_i - \bar{S})^2}}.$$

Sa valeur observée est égale à

$$\rho = \frac{\sum_{i=1}^n (r_i - \bar{r})(s_i - \bar{s})}{\sqrt{\sum_{i=1}^n (r_i - \bar{r})^2} \sqrt{\sum_{i=1}^n (s_i - \bar{s})^2}}.$$

Le fait que les rangs soient des permutations de l'ensemble $\{1, \dots, n\}$ simplifie les calculs de ce coefficient. On a le résultat ci-dessous.

Proposition 2.3.1 La valeur du rho de Speraman est égale à

$$\rho = 1 - \frac{6}{n(n^2 - 1)} \sum_{i=1}^n d_i^2,$$

où $d_i := r_i - s_i$, $i = 1, \dots, n$, représente la différence des rangs d'un même objet selon les deux classements.

Preuve. En l'absence d'ex-aequo, on a

$$\bar{r} = \bar{s} = \frac{n+1}{2} \text{ et } S_r^2 = S_s^2 = \frac{n^2 - 1}{12}.$$

D'où

$$\rho = \frac{12}{n^2 - 1} \left(\frac{1}{n} \sum_{i=1}^n r_i s_i - \left(\frac{n+1}{2} \right)^2 \right). \quad (2.1)$$

On a

$$\sum_{i=1}^n (r_i - s_i)^2 = \sum_{i=1}^n r_i^2 + \sum_{i=1}^n s_i^2 - 2 \sum_{i=1}^n r_i s_i.$$

Or la somme des carrés des n premiers entiers naturels est

$$\sum_{i=1}^n r_i^2 = \sum_{i=1}^n s_i^2 = \frac{n(n+1)(2n+1)}{6},$$

d'où on obtient

$$\sum_{i=1}^n r_i s_i = -\frac{1}{2} \sum_{i=1}^n d_i^2 + \frac{n(n+1)(2n+1)}{6}.$$

En remplaçant cette somme dans (2.1), on obtient le résultat de la proposition. ■

Remarque 2.3.1

1. On a $-1 \leq \rho \leq 1$.
2. Si $\rho = 1$, alors les deux classements sont identiques.
Si $\rho = -1$, alors les deux classements sont inverses l'un de l'autre.
Si $\rho = 0$, alors il n'y a pas de relation entre les deux classements.

Exemple 2.3.1 ([1, page 348]) Deux juges attribuent des points à 10 sportifs au cours d'une compétition. Les notes présentées dans le tableau 2.3.

Sportif	A	B	C	D	E	F	G	H	I	J
note du 1 ^{er} juge (x)	8.5	7.6	9.1	9.5	8.4	6.9	9.2	7.8	8.6	8.2
note du 2 ^e juge (y)	7.9	7.4	9.1	9.3	8.4	7.5	9.0	7.2	8.2	8.1

TAB. 2.3 – Notes attribuées par deux juges à 10 sportifs lors d'une compétition

A chaque note attribuée par un juge, on associe son rang et on résume les calculs dans le tableau 2.4.

Sportif	A	B	C	D	E	F	G	H	I	J
r_i	5	2	8	10	6	1	9	3	7	4
s_i	4	2	9	10	7	3	8	1	6	5
d_i	-1	0	-1	0	-1	-2	-1	2	-1	-1
d_i^2	1	0	1	0	1	4	1	4	1	1

TAB. 2.4 – Résultats relatifs au calcul du coefficient de corrélation de Spearman

On trouve $\sum_{i=1}^6 d_i^2 = 14$ et la valeur de ρ est donc égale à 0.92. Ceci indique que les avis des deux juges très proches comme le montre la figure 2.3.

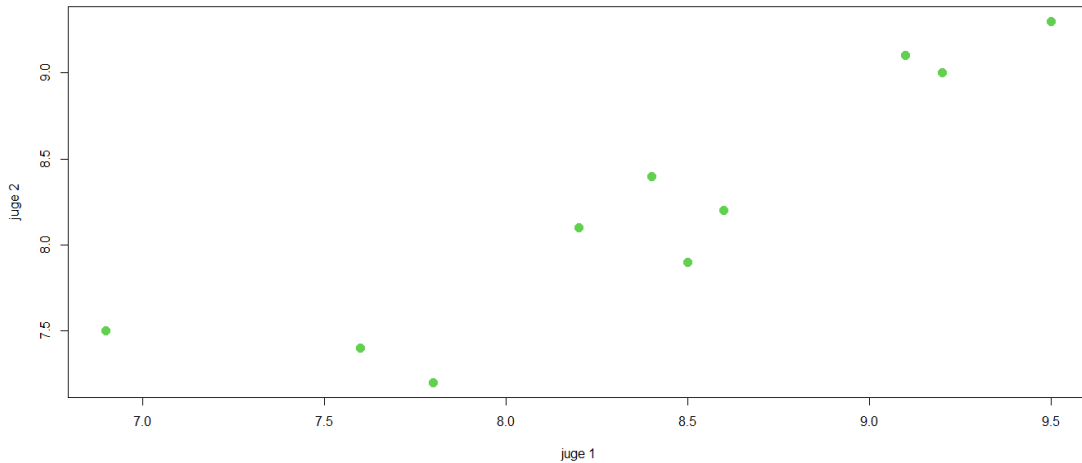


FIG. 2.3 – Notes attribuées par 2 juges à 10 sportifs

2.4 Coefficient de corrélation de Kendall

Le tau de Kendall sert à mesurer la liaison non linéaire entre deux v.a. Il donne une mesure de la corrélation entre les rangs des observations. On peut l'exprimer de deux manières différentes, soit en fonction des observations ou en fonction de la concordance.

2.4.1 Concept de concordance

Soit $\{(X_1, Y_1), \dots, (X_n, Y_n)\}$ un échantillon, de taille $n \geq 1$, d'un couple aléatoire (X, Y) .

Il existe $C_n^2 = n(n-1)/2$ paires d'observations distinctes (x_i, y_i) et (x_j, y_j) . Ces couples sont dite concordants ou discordants selon :

1. **Concordants** : $(x_i - x_j)(y_i - y_j) > 0$ c-à-d $(x_i < x_j \text{ et } y_i < y_j)$ ou $(x_i > x_j \text{ et } y_i > y_j)$.
2. **Discordants** : $(x_i - x_j)(y_i - y_j) < 0$ c-à-d $(x_i < x_j \text{ et } y_i > y_j)$ ou $(x_i > x_j \text{ et } y_i < y_j)$.

Définition 2.4.1 (fonction de concordance) *La fonction de concordance est la différence entre la probabilité de concordance et celle de discordance entre deux couples (X_1, Y_1)*

et (X_2, Y_2) . Elle est donnée par

$$Q := P[(X_1 - X_2)(Y_1 - Y_2) > 0] - P[(X_1 - X_2)(Y_1 - Y_2) < 0].$$

Remarque 2.4.1 Lorsque les deux v.a X et Y sont continues, alors

$$Q = 2P[(X_1 - X_2)(Y_1 - Y_2) > 0] - 1.$$

En effet, on a

$$P[(X_1 - X_2)(Y_1 - Y_2) < 0] = 1 - P[(X_1 - X_2)(Y_1 - Y_2) > 0].$$

2.4.2 Tau de Kendall

Définition 2.4.2 (tau de Kendall) Le tau de Kendall, noté τ , est défini par

$$\tau := P[(X_1 - X_2)(Y_1 - Y_2) > 0] - P[(X_1 - X_2)(Y_1 - Y_2) < 0].$$

Définition 2.4.3 (tau de Kendall empirique) L'équivalent empirique du tau de Kendall est

$$\tau = \frac{n_c - n_d}{C_n^2} = 2 \frac{n_c - n_d}{n(n-1)},$$

où n_c et n_d désignent les nombres de paires concordantes et discordantes respectivement.

Remarque 2.4.2 1. On a $-1 \leq \tau \leq 1$.

2. Si $\tau = 1$, alors toutes les paires d'observations sont concordantes.

Si $\tau = -1$, alors toutes les paires d'observations sont discordantes.

Si $\tau = 0$, alors il n'y a pas de relation entre les deux classements.

Remarque 2.4.3 Il y a une relation entre le rho de Spearman et le tau de Kendall (voir [6, 59])

$$-1 \leq 3\tau - 2\rho \leq 1.$$

Lorsque n est assez grand, et τ , ρ pas trop proches de 1 :

$$\rho = \frac{3}{2}\tau.$$

Si le couple (X, Y) suit une loi normale bivariée, alors

$$\tau = \frac{2}{\pi} \arcsin \rho.$$

Exemple 2.4.1 ([1, page 350]) *On réordonne les individus de façon à présenter les notes du premier juge selon un ordre croissant, comme le montre le tableau 2.5.*

Individus	F	B	H	J	A	E	I	G	C	D
r_i	1	2	3	4	5	6	7	8	9	10
s_i	3	2	1	5	4	7	6	9	8	10
concor.	7	7	7	5	5	3	3	1	1	0
disconcor.	2	1	0	1	0	1	0	1	0	0

TAB. 2.5 – Exemple de calcul du corrélation de Kendal

On a $n_c = 39$ et $n_d = 6$. Il s'ensuit que le taux de Kendall vaut 0.73.

2.4.3 Fonction R de calcul des coefficients de corrélation

La fonction `cor()` de **R** peut être utilisée pour calculer le coefficient de corrélation entre deux variables, X et Y .

Un format simplifié de cette fonction est :

$$\text{cor}(x,y,\text{method}=\text{c}(\text{"pearson"}, \text{"spearman"}, \text{"kendall"}))$$

Exemple

```
x<-c(44.4,45.9,41.9,53.3,44.7,44.1,50.7,45.2,60.1)
```

```
y<-c(2.6,3.1,2.5,5.0,3.6,4.0,5.2,2.8,3.8)
```

```
cor(x,y,method="pearson")
```

```
[1] 0.5711816cor(x,y,method="spearman")
```

```
[1] 0.6
```

```
cor(x,y,method="kendall")
```

```
[1] 0.4444444
```

Conclusion

La statistique est une science vaste et complète avec de multiples domaines d'application. Parmi ces derniers, il y a l'analyse de corrélation qui signifie l'étude de relations entre deux ou plusieurs variables. Son objectif principal est de déterminer la direction et l'intensité de la relation qui varie de zéro (pas de corrélation) à une corrélation complète (négative ou positive). La mesure de la corrélation est nécessaire dans tous les secteurs de l'activité humaine pour confirmer ou infirmer les liens qui peuvent exister entre différents phénomènes.

Dans ce mémoire, on s'est intéressé aux définitions et propriétés fondamentales des coefficients de corrélation les plus populaires, à savoir le coefficient de corrélation linéaire de Pearson, le rho de Spearman et le tau de Kendall. Des exemples d'application ont été traités par le logiciel d'analyse statistique R.

Il est à noter qu'il existe d'autres mesures d'association mais qui ne sont pas aussi largement utilisées. Parmi ces dernières, on peut citer le phi de Pearson, le V de Cramer,...

Bibliographie

- [1] Dreesbeke, J.J (1988) . Eléments de Statistique. Office des Publications Universitaires, Alger.
- [2] Hafayed, M. (2021) . Probabilités approfondies. Cours de 1^{ère} année master. Université de Biskra.
- [3] Ihaka, R. and Gentleman, R. (1996). *R* : A Language for Data Analysis and Graphics. Journal of Computational and Graphical Statistics 5, 299-314.
- [4] Lejeune, M. (2010). Statistique, la théorie et ses applications. Springer.
- [5] Lecoutre, J.P. (2012). Statistique et probabilités. Dunod, Paris.
- [6] Rakotomalala, R. (2017). Analyse de corrélation. Etude de dépendances-Variables quantitatives. Université Lumière Lyon 2.
- [7] Saporta, G. (2011). Probabilités, analyse des données et statistique. Technip, Paris

Annexe A : Codes R



1. Code de la figure 2.1

```
library(mvtnorm)

sigma1<- matrix(c(1,0.9,0.9,1), ncol=2)

v1<-rmvnorm(n=200, mean=c(0,0), sigma=sigma1)

sigma2<- matrix(c(1,0,0,1), ncol=2)

v2<-rmvnorm(n=200, mean=c(0,0), sigma=sigma2)

sigma3<- matrix(c(1,-0.9,-0.9,1), ncol=2)

v3<-rmvnorm(n=200, mean=c(0,0), sigma=sigma3)

x<-v1[,1];y<-v1[,2]

op <- par(mfrow=c(1,3))

plot(v1, col=4, pch=19, cex=1.5, xlab="x", ylab="y")

plot(v2, col=4, pch=19, cex=1.5, xlab="x", ylab="y")

plot(v3, col=4, pch=19, cex=1.5, xlab="x", ylab="y")

par(op)
```

2. Code de la figure 2.2

```
x=c(12,12,15,14,16,14,12,13,11,11)
y=c(4.1,3.4,11.3,10.2,11.5,7.2,6,7.8,3.5,3)
plot(x,y, col=4, pch=19, cex=1.5, xlab="âge", ylab="argent de poche")
```

3. Code de la figure 2.3

```
x=c(8.5,7.6,9.1,9.5,8.4,6.9,9.2,7.8,8.6,8.2)
y=c(7.9,7.4,9.1,9.3,8.4,7.5,9,7.2,8.2,8.1)
plot(x,y, col=3, pch=19, cex=1.5, xlab="juge 1", ylab="juge 2")
```

Annexe B : Abréviations et Notations

Les différentes abréviations et notations utilisées tout au long de ce mémoire sont expliquées ci-dessous.

Notation	Significations
c-à-d	c'est-à-dire
$Cor(X, Y)$	corrélation entre X et Y
$Cov(X, Y)$	covariance entre X et Y
$F(x, y)$	fonction de répartition conjointe
$f(x, y)$	densité de probabilité conjointe
$F_X(x), F_Y(y)$	fonctions de repartition marginales
$f_{X Y}, f_{Y X}$	densités conditionnelles
f_{x-y}, f_{y-x}	densité conditionnelle.
n_c, n_d	nombres de paires concordantes et discordantes.
P_{ij}	masse de probabilité.
$P_X(x_i), P_Y(y_i)$	masse de probabilité marginale.
P_{i-j}, P_{j-i}	loi conditionnelle, cas discret.
Q	fonction de concodance.

R	coefficient de corrélation empirique
R_i	rang de X_i .
ρ	rho de Spearman.
S_i	rang de Y_i .
S_{XY}	covariance empirique entre X et Y
S_X^2, S_Y^2	variances empiriques
τ	tau de Kendall.
μ	moyenne du couple.
μ_Z	moyenne de Z
σ_Z^2	variance de Z .
Σ	matrice de covariance.
θ	coefficient de corrélation.
\bar{X}, \bar{Y}	moyenne empirique
v.a	variable aléatoire
$:=$	égalité par définition
1	fonction indicatrice.