Note de cours de statistique descriptive

Niveau : L1 Mathématiques et Informatiqoue

rédigée par

KPATA Akon A. Bérenger P.

Enseignant-Chercheur à l'UFR SFA

Université Nangui Abrogoua

Côte d'Ivoire

16 janvier 2024

Table des matières

1 Vocabulaire de la statistique descriptive **3**

1.1 Généralités 3

1.2 Description d'une population statistique 4

1.3 Effectifs et fréquences 7

2 Séries statistiques à une variable **9**

2.1 Tableaux statistiques 9

2.2 Paramètres de position ou paramètres de tendance centrale 11

2.2.1 Moyenne arithmétique 11

2.2.2 Moyenne géométrique 12

2.2.3 Moyenne harmonique 12

2.2.4 moyenne quadratique 13

2.2.5 Mode 13

2.2.6 Médiane 15

2.2.7 Les quantiles 18

2.3 Paramètres de dispersion 19

2.3.1 Etendue 19

2.3.2 Ecart interquantile 19

2.3.3 Varaiance et écart-type 20

2.3.4 Coefficient de variation 20

2.3.5 Ecarts absolus moyens 21

2.4 Changement de variable affine et variable centrée réduite 21

2.4.1 Changement de variable affine 21

2.4.2 Variable centrée réduite 21

2.5 Moments d'une série statistique 22
2.5.1 Moments non centrés 22
2.5.2 Moments centrés 22
2.6 Paramètres de forme 23
2.6.1 Asymétrie 23
2.6.2 Aplatissement 24
2.7 Représentations graphiques 26
2.7.1 Cas des variables discrètes 26
2.7.2 Cas des variables continues 27
2.7.3 Cas des variables qualitatives 29

3 Séries statistiques à deux variables **31**
3.1 Généralités 31
3.2 Présentation des tableaux statistiques 31
3.2.1 Tableau de données ponctuelles 31
3.2.2 Tableau à double entrée ou tableau de contingence 32
3.3 Représentation graphique 32
3.4 Distributions asssociées et indépendance statistique 33
3.4.1 Distributions marginales ou séries marginales 33
3.4.2 Fréquences 34
3.4.3 Indépendance statistique 35
3.4.4 Distributions conditionnelles 35
3.5 Paramètres d'une série statistique à deux variables 36
3.5.1 Cas d'un tableau de données ponctuelles 36
3.5.2 Cas d'un tableau de contingence 38
3.6 Ajustement 39
3.6.1 Régression-corrélation 39
3.6.2 Ajustement affine 40
3.6.3 Ajustement affine par la méthode des moindres carrées . . . 40

Bibliographie **42**

Chapitre 1

Vocabulaire de la statistique descriptive

1.1 Généralités

Définition 1.1.1. ***(Les statistiques)***
Le mot " statistiques ", au pluriel, désigne l'ensemble des données chiffrées qui regroupent toutes les observations faites sur des faits relatifs à un phénomène qui concerne un groupe d'individus ou d'objets. Ces données sont essentiellement tirées des récensements de la population, des déclarations du registre d'état civil ou d'enquêtes appropriées et sont groupeés sous forme de tableaux, de graphiques ou d'indicateurs statistiques. On trouve des statistiques qui concernent la démographie, l'emploi, la santé, l'industrie, les transports, le commerce intérieur, le commerce extérieur, les indices de prix, la finance... etc.

Définition 1.1.2. ***(La statistique)***
Le mot " statistique ", au singulier, désigne la discipline scientifique constituée par l'ensemble des procédés, des techniques ou des méthodes d'analyse visant, d'une part, à fournir au moyen d'un nombre limité de caractéristiques une ***description*** *simple et la plus complète possible d'une population envisagée sous l'angle d'un caractère donné. D'autre part, la statistique permet d'****interpréter*** *les caractéristiques ainsi déterminées afin de tirer des conclusions concernant la population étudiée et de prendre des décisions.*

La statistique, en tant que méthode d'analyse des données quantitatives et qualitatives comportent deux niveaux :

• La **statistique descriptive**, qui consiste en la collecte et la présentation des données, ainsi que leur première analyse. Le but est de présenter d'une manière compréhensible et utilisable l' information fournie par les données.

• **La statistique inférentielle ou statistique mathématique**, qui cherche à trouver les caractéristiques de la population mère à partir des observations faites sur un échantillon. Elle s'appuie sur la statistique descriptive et fait appel au calcul des probabilités.

Une opération statistique se déroule en général en quatre étapes que sont :

1. La collecte des données. Cette étape peut se réaliser de deux manières à savoir le récensement ou le sondage.
2. La représentation et l'organisation statistique des données. Il s'agit de faire des représentations graphiques et de calculer les indicateurs statistiques
3. La modélisation.
4. L'interprétation des résultats.

1.2 Description d'une population statistique

Définition 1.2.1.

1. ***Population***

   C'est l'ensemble sur lequel porte l'étude statistique.

   Le terme "population" s'applique à des ensembles de toute nature : les êtres humains, les objets, les animaux, les plantes....

2. ***Echantillon***

   C'est un sous-ensemble ou une partie de la population statistique.

3. ***Individu ou unité statistique***

   On appelle individu ou unité statistique tout élément de la population.

4. ***Taille de la population***

   C'est le nombre d'individus qui constituent la population.

5. ***Caractère***

   *C'est l'aspect particulier des individus auxquels on s'intéresse. C'est aussi le **critère statistique** d'une population donnée.*

   *Le caractère est aussi appelé **variable**.*

On distingue deux grands types de caractères que sont : les caractères qualitatifs et les carctères quantitatifs.

a- Caractère qualitatif

Le caractère qualitatif est lié à une observation ne faisant pas l'objet d'une mesure ou d'un calcul. Il est donc non mesurable.

Exemple 1.2.1. *Considérons le personnel administratif de l'UFR SFA et intéressons nous à la situation matrimoniale (marié(e), veuf ou veuve, divorcé(e), célibataire) de chacun de ses membres.*
Etant donné que la situation matrimoniale n'est pas mesurable, elle est un caractère qualitatif ou une variable qualitative.

Les différentes rubriques d'un caractère qualitatif sont appelées les **modalités** de ce caractère. Ainsi, dans l'exemple précédent, marié(e), veuf ou veuve, divorcé(e) et célibataire constituent les différentes modalités du caractère "situation matrimoniale".

Notation 1.2.1. *Dans toute la suite, les modalités d'un caractère qualitatif donné seront notées m_i.*

Les modalités d'un caractère qualitatif sont :
-<u>exhaustives</u> : à chaque individu doit correspondre une modalité du caractère ;
-<u>mutuellement incompatibles</u> : chaque individu doit pouvoir être classé dans une seule modalité.
Ainsi, chaque individu de la population statistique doit pouvoir être classé dans une et une seule modalité.

Les modalités d'un caractère qualitaif peuvent être ordinales ou nominales.
Les modalités sont dites **ordinales** si elles peuvent être classées ou hiérarchisées.

Exemple 1.2.2. *La mention obtenue au bac , le niveau d'études scolaires.*

Les modalités sont dites **nominales** si elles ne peuvent pas être classées ou hierarchisées.

Exemple 1.2.3. *: Les modalités de la situation matrimoniale.*

b- Caractère quantitatif

Il est mesurable. On peut effectuer des calculs numériques sur ce caractère.

Exemple 1.2.4. *L'âge, le poids, la taille, le nombre d'enfants.*

Remarque 1.2.1.

Les valeurs prises par un caractère quantitatif sont parfois appelées modalités.

Les caractères quantitatifs peuvent être classés en variables discrètes ou en variables continues.

b-1) Variables discrètes

Les variables discrètes prennent des valeurs isolées.

L'ensemble des valeurs prises par une variable discrète est fini ou dénombrable.

Exemple 1.2.5. *Le nombre d'enfants par menage.*

Notation 1.2.2. *Dans toute la suite, nous désignerons par x_i ou y_i les valeurs prises par les variables discrètes.*

b-2) variables continues

Les variables continues peuvent prendre toutes les valeurs d'un intervalle contenu dans $\mathbb{R}$.

D'une manière générale, ces intervalles sont notés $Cl_i = [e_i, e_{i+1}[$ et sont appelés **classes**.

- e_i est la **borne inférieure** de la classe Cl_i,
- e_{i+1} est la **borne supérieure** de la classe Cl_i,
- le nombre $e_{i+1} - e_i$ est l'**amplitude** de la classe Cl_i,
- $c_i = \dfrac{e_{i+1} + e_i}{2}$ est le **centre** de la classe Cl_i.

Remarque 1.2.2. *Les classes d'une variable continue sont toujours deux à deux disjointes.*

1.3 Effectifs et fréquences

Définition 1.3.1.

1. ***Effectifs***

   • *Le nombre d'individus pour lequel le carctère prend soit la modalité x_i ou m_i soit toutes ses valeurs dans la classe Cl_i est appelé* ***effectif*** *de cette modalité et sera noté n_i. L'effectif n_i est aussi appelée fréquence absolue.*

   • *La taille de la population est l'****effectif total*** *et sera notée N.*

   N est donc la somme des effectifs de toutes les modalités. En d'autres termes, s'il y a r modalités,

   $$N = n_1 + n_2 + ... + n_r = \sum_{i}^{r} n_i.$$

   $\sum_{i}^{r} n_i$ se lit somme pour i allant de 1 à r des n_i.

2. ***Fréquence***

   On appelle fréquence ou fréquence relative de la modalité x_i ou m_i ou de la classe Cl_i le nombre noté f_i défini par

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

3. ***Pourcentage***

   On appelle pourcentage de la modalité x_i ou m_i ou de la classe Cl_i le nombre noté p_i défini par

   $$p_i = 100 f_i.$$

4. ***Effectif cumulé croissant***

   C'est une notion qui n'a de sens que si le caractère étudié est quantitatif ou qualitaif ordinal.

   L'effectif cumulé croissant associé à la modalité x_i ou m_i ou à la classe $Cl_i = [e_i,\ e_{i+1}[$ est le nombre d'individus dont la modalité du caractère est strictement inférieure à x_i ou à m_i ou à e_{i+1}.

   L'effectif cumulé croissant sera noté N_i.

5. ***Effectif cumulé décroissant***

   On appelle effectif cumulé décroissant associé à la modalité x_i ou m_i ou à la classe

$Cl_i = [e_i,\, e_{i+1}[$ le nombre $N - N_i$.

C'est donc le nombre d'individus dont la modalité du caractère est supérieure ou égale à x_i ou à m_i ou à e_{i+1}.

6. ***Fréquence cumulée croissante***

   On appelle fréquence cumulée croissante de la modalité x_i ou m_i ou de la classe Cl_i le nombre noté F_i défini par

$$F_i = \frac{N_i}{N}.$$

7. ***Fréquence cumulée décroissante***

   On appelle fréquence cumulée décroissante de la modalité x_i ou m_i ou de la classe Cl_i le nombre noté $\widetilde{F_i}$ défini par

$$\widetilde{F_i} = \frac{N - N_i}{N}.$$

8. ***Pourcentage cumulé croissant***

   On appelle fréquence cumulée croissante de la modalité x_i ou m_i ou de la classe Cl_i le nombre noté P_i défini par

$$P_i = 100F_i.$$

9. ***Pourcentage cumulé décroissant***

   On appelle fréquence cumulée décroissante de la modalité x_i ou m_i ou de la classe Cl_i le nombre noté $\widetilde{F_i}$ défini par

$$\widetilde{P_i} = 100\tilde{F_i}.$$

10. ***Fonction de répartition***

    On appelle fonction de répartition, la fonction notée F qui, à chaque nombre réel x associe la fréquence des individus ayant pour modalité une valeur strictement inférieur à x.

Remarque 1.3.1. *Pour un caractère ayant r modalités, on a les relations suivantes :*

- $0 \leq f_i \leq 1 \quad \forall 1 \leq i \leq r.$
- $\displaystyle\sum_{i=1}^{r} f_i = 1.$
- $\displaystyle\sum_{i=1}^{r} p_i = 100.$

Chapitre 2

Séries statistiques à une variable

2.1 Tableaux statistiques

On appelle série statistique l'ensemble des couples $(x_i,\ n_i)$ ou $(m_i,\ n_i)$ ou $([e_i,\ e_{i+1}[,\ n_i)$. Pour pouvoir être exploitées, les séries statistiques sont représentées dans un tableau dit tableau statistique.

Exemple 2.1.1. *La situation matrimoniale des enseignants du groupe scolaire "les anges" se présentent comme suit :*

$$\begin{matrix} M & M & D & C & C & M & C & C & C & M \\ C & M & V & M & V & D & C & C & C & M \end{matrix}$$

où C, M, D et V désignent célibatauire, marié(e), divorcé(e) ou veuf(ve) respectivement.

1. *Préciser les éléments suivants :*
   - population statistique étudiée
   - taille de la population
   - caractère étudié et sa nature
2. *Quelles sont les modalités du caractère étudié*
3. *Présenter le tableau statistique de cette série en y faisant figurer le dépouillement.*
4. *Calculer les fréquences.*

Exemple 2.1.2. *On considère l'ensemble des notes obtenues, lors d'un test noté sur* 20, *par* 50 *candidats.*

10	8	3	12	13	9	12	9	12	11
11	11	8	5	13	14	14	6	12	16
11	7	8	10	13	9	13	9	7	13
7	11	10	10	2	15	12	10	1	14
11	19	9	4	10	8	9	6	7	14

1. *Quelle est la population statistique ?*
2. *Quel est le caractère étudié et sa nature ?*
3. *Dépouiller ces données et présenter les résultats dans un tableau statistique. (On prendra les classes suivantes :* $[0,5[$, $[5,\ 7[$, $[7,9[$, $[9,\ 11[$, $[11,13[$, $[13,\ 15[$ *et* $[15,\ 20[$.*)*
4. *Déterminer l'amplitude et le centre de chaque classe.*
5. *Calculer les fréquences et les fréquences cumulées croissantes.*
6. *Quelle est la proportion des candidats ayant une note inférieure à* 9 *?*
7. *Quelle est la proportion des candidats ayant une note supérieure ou égale à* 13 *?*
8. *Quelle est la proportion des étudiants ayant une note comprise entre* 5 *et* 20 *?*

Exemple 2.1.3. *On a relevé le nombre d'enfants de* 40 *menages ayant au moins un enfant. On a obtenu le résultat suivant :*

Nombre d'enfants	Nombre de menages
1	5
2	10
3	11
4	9
5	5

1. *Déterminer la population étudiée, le caractère étudié et sa nature.*
2. *Dresser le tableau statistique de cette série en y faisant figurer les fréquences et les fréquences cumulées décroissantes.*

2.2 Paramètres de position ou paramètres de tendance centrale

2.2.1 Moyenne arithmétique

Soit X une variable statistique discrète dont les modalités sont x_1, x_2, ..., x_r d'effectifs respectifs n_1, n_2,..., n_r.

Définition 2.2.1. *On apelle* ***moyenne arithmétique*** *ou* ***moyenne*** *de cette série statistique le nombre*

$$\begin{aligned}\overline{X} &= \frac{n_1x_1 + n_2x_2 + ... + n_rx_r}{N} \\ &= \frac{1}{N}\sum_{i=1}^{r} n_ix_i \\ &= \sum_{i=1}^{r} f_ix_i.\end{aligned}$$

Remarque 2.2.1. *Lorsque la variable X est continue, on remplace chaque x_i par le centre c_i de chaque classe $[e_i,\ e_{i+1}[$.*

Proposition 2.2.1. *Soient $\{x_i\}_{i\in I}$ une série statistique et $\{y_i\}_{i\in I}$ la série définie par*

$$\forall i \in I, \quad y_i = ax_i + b$$

où a et b sont deux réels quelconques. Alors

$$\overline{Y} = a\overline{X} + b.$$

Remarque 2.2.2. *La multiplication par a correspond à un changement d'échelle (d'unité) et l'addition de b à un changement d'origine.*
La propriété précédente permet de simplifier le calcul de la moyenne arithmétique.

Proposition 2.2.2. *Soient p séries statistiques de même nature : $\{x_{1i}\}$ d'effectif total n_1, $\{x_{2i}\}$ d'effectif total n_2, ... et $\{x_{pi}\}$ d'effectif total n_p.*
Alors la moyenne arithmétique des p séries est la moyenne arithmétique des moyennes

arithmétiques de chacune des séries, pondérées par leurs effectifs respectifs :

$$\overline{X} = \frac{\displaystyle\sum_{i=1}^{p} n_i \overline{X_i}}{\displaystyle\sum_{i=1}^{p} n_i}.$$

2.2.2 Moyenne géométrique

Soit X une variable statistique discrète dont les modalités x_1, x_2, ..., x_r sont **positives** d'effectifs respectifs n_1, n_2,..., n_r.

Définition 2.2.2. *On appelle **moyenne géométrique** de cette série statistique le nombre noté G_X défini par*

$$G_X = \left(\prod_{i=1}^{r} x_i^{n_i}\right)^{\frac{1}{N}} = \prod_{i=1}^{r} x_i^{f_i}.$$

Remarque 2.2.3. *Lorsque la variable X est continue, on remplace chaque x_i par le centre c_i de chaque classe $[e_i,\ e_{i+1}[$.*

Remarque 2.2.4. *Dans le cas où toutes les modalités x_i sont strictement positives, le calcul de G peut s'effectuer grâce à la relation*

$$\ln G_X = \frac{\displaystyle\sum_{i=1}^{r} n_i \ln x_i}{N}.$$

Proposition 2.2.3. *Soient $X = \{x_i\}_{i\in I}$ et $Y = \{y_i\}_{i\in I}$ deux séries statistiques positives de même nature ; $Z = \{z_i\}_{i\in I}$ et $T = \{t_i\}_{i\in I}$ les séries définies par :*

$$\forall i \in I, \quad z_i = x_i y_i \quad \textit{et} \quad t_i = \frac{x_i}{y_i} \quad (y_i \neq 0).$$

Alors

$$G_Z = G_X G_Y \quad \textit{et} \quad G_T = \frac{G_X}{G_Y}.$$

2.2.3 Moyenne harmonique

Soit X une variable statistique discrète dont les modalités x_1, x_2, ..., x_r sont **strictement positives** d'effectifs respectifs n_1, n_2,..., n_r.

Définition 2.2.3. *La **moyenne harmonique** de cette série statistique est égale à l'inverse de la moyenne arithmétique des inverses des valeurs observées. On la note généralement H_X. Ainsi :*

$$H_X = \frac{N}{\displaystyle\sum_{i=1}^{r} \frac{n_i}{x_i}} = \frac{1}{\displaystyle\sum_{i=1}^{r} \frac{f_i}{x_i}}.$$

Remarque 2.2.5. *Lorsque la variable X est continue, on remplace chaque x_i par le centre c_i de chaque classe $[e_i,\ e_{i+1}[$.*

2.2.4 moyenne quadratique

Soit X une variable statistique discrète de modalités x_1, x_2, ..., x_r d'effectifs respectifs n_1, n_2,..., n_r.

Définition 2.2.4. *La **moyenne quadratique** de cette série statistique est la racine carrée de la moyenne arithmétique des carrés des valeurs observées. On la note généralement Q_X. Ainsi :*

$$Q_X = \sqrt{\frac{1}{N} \sum_{i=1}^{r} n_i x_i^2} = \sqrt{\sum_{i=1}^{r} f_i x_i^2}.$$

Remarque 2.2.6. *Lorsque la variable X est continue, on remplace chaque x_i par le centre c_i de chaque classe $[e_i,\ e_{i+1}[$.*

Remarque 2.2.7. *Soit une série statistique X pour laquelle les quatres moyennes définies ci-dessus existent. On a alors*

$$H_X \leq G_X < \overline{X} \leq Q_X.$$

2.2.5 Mode

Définition 2.2.5. *Le mode est noté Mo.*

• *Dans le cas d'une variable qualitative ou d'une variable quantitative discrète, on appelle mode une modalité ayant le plus grand effectif.*

• *Dans le cas d'une variable continue dont les classses sont d'amplitudes égales, on appelle **classe modale**, une classse qui a l'effectif le plus grand.*

Désignons par $[e_i,\ e_{i+1}[$ *une classe modale et notons* n_i *l'effectif de cette classe. L'effectif de la classe qui précède la classe modale est* n_{i-1} *et l'effectif de la classe qui suit la classe modale est* n_{i+1}.
La détermination du mode à partir de la classe modale est donnée par la formule

$$Mo = e_i + a_i\left(\frac{r_1}{r_1 + r_2}\right) = \frac{e_i r_2 + e_{i+1} r_1}{r_1 + r_2}$$

où

$r_1 = n_i - n_{i-1}$ $\qquad$ $r_2 = n_i - n_{i+1}$ $\qquad$ $a_i = e_{i+1} - e_i =$ *amplitude de la classe modale.*

• *Dans le cas d'une variable continue dont les classses sont d'amplitudes inégales (mais finies), on commmence par calculer l'effectif corrigé de chaque classe .*
On appelle effectif corrigé ou densité de la classe $[e_i,\ e_{i+1}[$, *le nombre*

$$d_i = \frac{n_i}{a_i}$$

où $a_i = e_{i+1} - e_i$.
On appellle alors classe modale, toute classe ayant le plus grand effectif corrigé.
Désignons par $[e_i,\ e_{i+1}[$ *une classe modale et notons* d_i *l'effectif corrigé de cette classe. L'effectif de la classe qui précède la classe modale est* d_{i-1} *et l'effectif corrigé de la classe qui suit la classe modale est* d_{i+1}.
La détermination du mode à partir de la classe modale est donnée par la formule

$$Mo = e_i + a_i\left(\frac{\widetilde{r}_1}{\widetilde{r}_1 + \widetilde{r}_2}\right) = \frac{e_i \widetilde{r}_2 + e_{i+1} \widetilde{r}_1}{\widetilde{r}_1 + \widetilde{r}_2}$$

où

$\widetilde{r}_1 = d_i - d_{i-1}$ $\qquad$ $\widetilde{r}_2 = d_i - d_{i+1}$ $\qquad$ $a_i = e_{i+1} - e_i =$ *amplitude de la classe modale.*

Remarque 2.2.8. *Une série statistique peut présenter plusieurs modes. On dit alors qu'elle est plurimodale.*

Exemple 2.2.1. *Déterminer le mode de la série statistique suivante et interpréter le résultat obtenu.*

$$14 \quad 16 \quad 12 \quad 9 \quad 11 \quad 18 \quad 7 \quad 8 \quad 9 \quad 16 \quad 7 \quad 9 \quad 18.$$

Exemple 2.2.2. *On considère le tableau statistique suivant :*

Classes	*Effectifs*
$[0,\ 5[$	4
$[5,\ 7[$	3
$[7,\ 9[$	8
$[9,\ 11[$	12
$[11,\ 13[$	11
$[13,\ 15[$	9
$[15,\ 20[$	3

Déterminer le mode de cette série statistique et interpréter le résultat obtenu.

2.2.6 Médiane

La médiane d'une série statistique est la valeur qui partage cette dernière (préalablement classée selon les valeurs croissantes prises par la variable) en deux séries de même effectif. Elle est notée Me.
Dans la première série on trouve les valeurs strictement inférieures à la médiane et dans la deuxième série on trouve les valeurs strictement supérieures à la médiane.
La détermination de la médiane dépend du type de données. On distingue quatre cas :
- la série est non groupée d'effectif total impair et aucune valeur n'est répétée
- la série est non groupée d'effectif total pair et aucune valeur n'est répétée
- la série est discrète et groupée par valeurs
-la série est continue et groupée par classe.

a) Calcul de la médiane : effectif impair où aucune valeur n'est répétée.
C'est le cas idéal, celui qui permet de mieux comprendre la notion de médiane.

Exemple 2.2.3. *Déterminer la médiane de la série suivante :*

$$8 \quad 5 \quad 3 \quad 4 \quad 12 \quad 2 \quad 24.$$

• *Classsement par ordre croissant :*

$$2 \quad 3 \quad 4 \quad 5 \quad 8 \quad 12 \quad 24.$$

• *Détermination du rang de la médiane.*
Le rang de la médiane est $\frac{N+1}{2}$ *où* N *est l'effectif total.*
Ici $\frac{N+1}{2} = \frac{7+1}{2} = 4$. *La médiane se trouve à la* 4^e *position. Donc* $Me = 5$.

b) Calcul de la médiane : effectif pair où aucune valeur n'est répétée.

C'est le cas idéal, celui qui permet de mieux comprendre la notion de médiane.

Exemple 2.2.4. *Déterminer la médiane de la série suivante :*

$$3 \quad 1 \quad 9 \quad 10 \quad 2 \quad 4 \quad 12 \quad 7.$$

• *Classer la série par ordre croisssant de valeurs*

$$1 \quad 2 \quad 3 \quad 4 \quad 7 \quad 9 \quad 10 \quad 12.$$

• *Déterminer le rang de la médiane.*
Celui-ci est donné par $\frac{N}{2}$. *Mais le rang de la médiane ne nous donne pas la médiane. Il indique que la médiane est comprise entre la* $\frac{N}{2}$ *ième valeur et la* $\frac{N}{2}+1$ *ième valeur. Ces valeurs constituent l'intervalle médian.*
La médiane est égale au centre de cet interval.
Ici $\frac{N}{2} = \frac{8}{2} = 4$. *L'intervalle médian est* $[4,\ 7]$. *Donc* $Me = \frac{4+7}{2} = 5,5$.

c) Calcul de la médiane : valeurs discrètes et groupées.

x_i	n_i	f_i	$F_i = FCC$	$N_i = Ecc$
2	2	$0,067$		
8	3	$0,1$		
9	4	$0,133$		
10	4	$0,133$		
11	5	$0,167$		
12	3	$0,1$		
13	6	$0,2$		
15	1	$0,033$		
18	2	$0,067$		

Remarque 2.2.9. *Dans le cas d'une variable discrète ayant un effectif N pair, il n'est pas toujourss possible de déterminer avec exactitude la médiane. De plus, le rang de la médiane n'indique pas nécessairement la médiane.*

d) Calcul de la médiane : valeurs discrètes et groupées.

Classes	Effectifs	f_i	$F_i = FCC$	$N_i = ECC$
$[150,\ 160[$	15			
$[160,\ 180[$	10			
$[180,\ 200[$	24			
$[200,\ 250[$	26			
$[250,\ 300[$	10			

La valeur 43 ne correspond pas à une valeur des ECC. Nous allons utiliser la **méthode d'interpolation linéaire**. Pour cela, remarquons que $43 \in [25,\ 49]$.

$$\begin{array}{ccc} 25 & 43 & 49 \\ \longleftarrow & \!\!\!\!\text{---------}\!\!\!\! & \longrightarrow \\ \longleftarrow & \!\!\!\!\text{---------}\!\!\!\! & \longrightarrow \\ 180 & Me & 200 \end{array}$$

On choisit un sens de lecture. Le notre est le suivant : $\leftarrow$.

$$49 - 25 \quad \longleftrightarrow \quad 200 - 180$$

$$43 - 25 \quad \longleftrightarrow \quad Me - 180$$

$$24 \longleftrightarrow 20$$
$$18 \longleftrightarrow Me - 180$$

On fait ensuite le produit en croix.

$$24(Me - 180) = 18 \times 20$$

On obtient alors

$$Me = 195$$

Exercice : Completer le tableau ci-dessus et calculer la médiane à l'aide des fréquences cumulées croissantes.

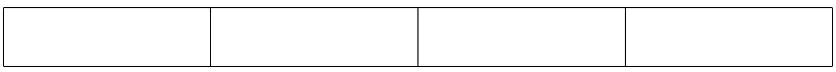
Exemple 2.2.5. *Si* $\dfrac{N+1}{2}$ *est une valeur des ECC alors celle-ci se trouve entre deux classes de type* $[e_i,\ e_{i+1}[$ *et* $[e_{i+1},\ e_{i+2}[$. *Par suite,* $F(e_{i+1}) = \dfrac{N+1}{2}$ *et* $M_e = e_{i+1}$.

2.2.7 Les quantiles

Les **quantiles** sont des indicateurs de position. Le quantile d'ordre α $(0 \leq \alpha \leq 1)$ noté U_α est la valeur du caractère telle qu'une proportion α des individus aient pour modalité une valeur strictement inférieur à U_α.
Si F désigne la fonction de répartition ou la fonction "des fréquences cumulées croissantes" alors $F(U_\alpha) = \alpha$.
Dans la pratique, on utilise les quantiles d'ordre $\frac{1}{4}$, $\frac{1}{2}$ et $\frac{3}{4}$. Ces quantiles portent aussi le nom de **quartiles** et les notations suivantes sont utilisées pour les désigner.
$Q_1 = U_{\frac{1}{4}} = U_{0,25}$ c'est le premier quartile ;
$Q_2 = U_{\frac{1}{2}} = U_{0,50}$ c'est le deuxième quartile ;
$Q_3 = U_{\frac{3}{4}} = U_{0,75}$ c'est le troisième quartile.

Remarque 2.2.10.

- *Les trois* (3) *quartiles partagent la série statistique en quatre* (4) *séries de même taille.*



- *Le deuxième quartile c'est-à-dire Q_2 est la médiane. Autrement dit, $Q_2 = Me$.*
- *Les quantiles se déterminent de la même manière que la médiane.*

De manière similaire aux quartiles, on définit les déciles notés D_1, D_2, D_3, D_4, D_5, D_6, D_7, D_8, D_9 en posant

$$D_i = U_{\frac{i}{10}} \quad \text{pour} \quad i \in \{1,\ 2,\ 3,\ 4,\ 5,\ 6,\ 7,\ 8,\ 9\}.$$

On définit également les 99 centiles par

$$D_i = U_{\frac{i}{10}} \quad \text{pour} \quad i \in \{1,\ 2,\ ...,\ 99\}.$$

2.3 Paramètres de dispersion

Ce sont des nombres qui mesurent la dispersion des valeurs observées autour d'un paramètre de position notamment autour de la moyenne arithmétique, la médiane et le mode. Ils permettent de comparer les séries statistiques.

2.3.1 Etendue

L'étendue est la différence entre la plus grande et la plus petite des valeurs observées.

2.3.2 Ecart interquantile

On appelle :
• intervalle interquartile l'intervalle $[Q_1,\ Q_3]$ qui contient 50% des valeurs observées.
La valeur $Q_3 - Q_1$ est l'écart interquartile
• écart interdécile le nombre $D_9 - D_1$
• écart intercentile le nombre $U_{\frac{99}{100}} - U_{\frac{1}{100}}$.

2.3.3 Varaiance et écart-type

Toutes les définitions de cette sous-section seront données pour une distribution quantitative discrète $\{(x_i,\ n_i)\,/\,1 \leq i \leq r\}$. Lorsque la variable est continue, on remplace x_i par les centres c_i de chaque classe.

Définition 2.3.1. *On appelle **variance** d'une variable X le nombre réel positif ou nul noté $Var(X)$ et défini par :*

$$\begin{aligned} Var(X) &= \frac{1}{N}\sum_{i=1}^{r} n_i(x_i - \overline{X})^2 \\ &= \sum_{i=1}^{r} f_i(x_i - \overline{X})^2. \end{aligned}$$

Dans la pratique, pour calculer la variance, on utilise le théorème de Koenigs :

$$Var(X) = \left(\frac{1}{N}\sum_{i=1}^{r} n_i x_i^2\right) - \overline{X}^2.$$

Remarque 2.3.1. *L'unité de la variance est l'unité de la variable X au carré.*

Définition 2.3.2. *L'**écart-type** noté σ_X est défini par*

$$\sigma_X = \sqrt{Var(X)}.$$

Remarque 2.3.2. *L'écart-type s'exprime dans la même unité que les valeurs observées et mesure la dispersion autour de la moyenne. Ainsi, plus l'écart-type est grand, plus la dispersion autour de la moyenne est grande.*

2.3.4 Coefficient de variation

Définition 2.3.3. *le coefficient de variation noté CV est défini par $CV = \dfrac{\sigma}{\overline{X}}$.*

Remarque 2.3.3. *Le coefficient de variation est un nombre sans dimension, indépendant des unités choisies.*
Le coefficient de variation permet de comparer des séries statistiques exprimées dans des unités différentes.

2.3.5 Ecarts absolus moyens

a) Ecart absolu moyen par rapport à la moyenne

C'est la moyenne des écarts à la moyenne. On le note $e_{\overline{X}}$ et on a :

$$e_{\overline{X}} = \frac{1}{N}\sum_{i=1}^{r} n_i|x_i - \overline{X}|.$$

b) Ecart absolu moyen par rapport à la médiane

C'est la moyenne des écarts à la médiane. On le note e_{Me} et on a

$$e_{Me} = \frac{1}{N}\sum_{i=1}^{r} n_i|x_i - Me|.$$

Remarque 2.3.4. *Pour toute distribution statistique on a*

$$e_{Me} \leq e_{\overline{X}} \leq \sigma_X.$$

2.4 Changement de variable affine et variable centrée réduite

2.4.1 Changement de variable affine

Soient a et b deux nombres réels. On considère une variable quantitative X et on lui fait subir une application affine qui la transforme en une variable Y définie par $Y = aX + b$. Alors :

- les observations y_i sont données par $y_i = ax_i + b$.
- la moyenne de Y est : $\overline{Y} = a\overline{X} + b$
- la variance de Y est : $Var(Y) = a^2Var(X)$
- l'écart-type de Y est $\sigma_Y = |a|\sigma_X$.

2.4.2 Variable centrée réduite

Onn considère une variable X de moyenne $\overline{X}$, de variance $Var(X)$ et d'écart-type σ_X.

On définit une nouvelle variable Y en posant :

$$Y = \frac{X - \overline{X}}{\sigma_X}.$$

Cette variable est appelée variable centrée réduite associée à X.

En effet, elle est :

-centrée : $\overline{Y} = \dfrac{\overline{X} - \overline{X}}{\sigma_X} = 0$

- réduite : $\sigma_Y = |\dfrac{1}{\sigma_X}| \times \sigma_X = \dfrac{1}{\sigma_X} \times \sigma_X = 1.$

Remarque 2.4.1. *La variable centrée réduite Y est sans unité.*

2.5 Moments d'une série statistique

2.5.1 Moments non centrés

Définition 2.5.1. *Le moment non centré d'ordre k ($k \in \mathbb{N}$) d'une variable X est le nombre $m_k(X)$ défini par*

$$m_k(X) = \frac{1}{N}\sum_{i=1}^{r} n_i x_i^k = \sum_{i=1}^{r} f_i x_i^k$$

Remarque 2.5.1. *Nous avons :*

- $m_0(X) = 1$
- $m_1(X) = \overline{X}$
- $\sqrt{m_2(X)}$ *est la **moyenne quadratique** de X.*
- $Var(X) = m_2(X) - (m_1(X))^2$.

2.5.2 Moments centrés

Définition 2.5.2. *Le moment centré d'ordre k ($k \in \mathbb{N}$) d'une variable X est le nombre $\mu_k(X)$ défini par*

$$\mu_k(X) = \frac{1}{N}\sum_{i=1}^{r} n_i (x_i - \overline{X})^k = \sum_{i=1}^{r} f_i (x_i - \overline{X})^k.$$

Remarque 2.5.2.

- $\mu_0(X) = 1$
- $\mu_1(X) = 0$

- $\mu_2(X) = Var(X)$.
- *Soit* a *et* b *deux réels quelconques, et les séries statistiques :* $X = \{x_i\}_{i\in I}$, $Y = \{y_i\}_{i\in I}$ *telles que pour tout* i *élément de* I, $y_i = ax_i + b$. *Alors pour tout* $k \in \mathbb{N}$,

$$\mu_k(Y) = a^k \mu_k(X).$$

2.6 Paramètres de forme

Les polygones des fréquences nous livrent une représentation approximative de la distribution réelle des fréquences. Pour avoir une idée satisfaisante et plus précise sur la forme de la distribution, il est recommandé de calculer les indicateurs de forme. On distingue les indicateurs d'asymétrie et les indicateurs d'aplatissement. Ces indicateurs sont sans unités de mesure. Ils sont indépendants d'un changement d'échelle et / ou d'origine.

2.6.1 Asymétrie

Une distribution est dite asymétrique si les observations se répartissent dans la même proportion de part et d'autre des trois valeurs centrales (mode, médiane et moyenne).

Les mesures d'asymétrie permettent de quantifier le dégré de déviation de la forme de distribution par rapport à une distribution symétrique.

a) Coefficient d'asymétrie de Fisher

Le coefficient d'asymétrie de Fisher est noté γ_1 et est défini par :

$$\gamma_1 = \frac{\text{moment centré d'ordre 3}}{(\text{écart-type})^3} = \frac{\mu_3(X)}{\sigma_X^3}$$

La distribution est dite symétrique dans le cas où $\gamma_1 = 0$.

La distribution est dite étalée à gauche dans le cas où $\gamma_1 < 0$.

La distribution est dite étalée à droite dans le cas où $\gamma_1 > 0$.

a) Coefficient d'asymétrie de Yule

Le coefficient d'asymétrie de Yule est noté C_Y et est basé sur les quartiles. Il est défini par :

$$C_Y = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{Q_3 - Q_1}.$$

La distribution est dite symétrique dans le cas où $C_Y = 0$.

La distribution est dite étalée à gauche dans le cas où $C_Y < 0$.

La distribution est dite étalée à droite dans le cas où $C_Y > 0$.

c) Coefficient d'asymétrie de Pearson

Le coefficient d'asymétrie de Yule est noté C_P et est basé sur la moyenne, le mode et l'écart-type. Il est défini par :

$$C_p = \frac{\overline{X} - Mo}{\sigma_X}.$$

La distribution est dite symétrique dans le cas où $C_p = 0$.

La distribution est dite étalée à gauche dans le cas où $C_P < 0$.

La distribution est dite étalée à droite dans le cas où $C_P > 0$.

2.6.2 Aplatissement

Une distribution est d'autant plus "plate" que la dispersion des observations autour des valeurs centrales est forte.

a) Coefficient d'aplatissement de Pearson

Il est noté β et est défini par

$$\beta = \frac{\mu_4(X)}{\sigma_X^4} = \frac{\text{moment centré d'ordre } 4}{(\text{écart-type})^4}.$$

La distribution est dite normale dans le cas où $\beta = 3$.

La distribution est dite hyponormale ou platykurtyque (plus aplatie que la distribution normale) dans le cas où $\beta < 3$.

La distribution est dite hypernormale ou leptokurtique (moins aplatie que la distribution normale) dans le cas où $\beta > 3$.

b) Coefficient d'aplatissement de Fisher

Il est noté γ_2 et est défini par

$$\gamma_2 = \frac{\mu_4(X)}{\sigma_X^4} - 3 = \beta - 3.$$

La distribution est dite normale dans le cas où $\gamma_2 = 0$.
La distribution est dite hyponormale (plus aplatie que la normale) dans le cas où $\gamma_2 < 0$.
La distribution est dite hypernormale (moins aplatie que la normale) dans le cas où $\gamma_2 > 0$.

2.7 Représentations graphiques

Pour toute représentation graphique, il est impératif de préciser l'échelle, le nom des axes et le nom du graphique et éventuellement la source des informations.

2.7.1 Cas des variables discrètes

a) Diagrammes en bâtons

Pour représenter une série statistique discrète, on porte sur l'axe des abscisses les valeurs discrètes et sur l'axe des ordonnées les effectifs n_i ou les fréquences f_i associé(e)s à la variable. On trace ensuite des **bâtons** verticaux dont la longueur est proportionnelle aux effectifs ou aux fréquences. On obtient ainsi le diagramme en bâtons.

Remarque 2.7.1. *Le mode de la série statistique discrète est la modalité x_i qui a le bâton le plus élevé.*

Nous appelons polygone statistique ou diagramme polygonal, la ligne obtenue en joignant les sommets des bâtons.

b) Courbe de la fonction de répartition

C'est la représentation graphique de la fonction de répartition F.

Rappelons que si $\{(x_i,\, n_i)\,/\,1 \leq i \leq r\}$ est une série statistique à une variable, alors

$$F(x) = \begin{cases} 0 & \text{si} \quad x \leq x_1 \\ F_{i+1} & \text{si} \quad x_i < x \leq x_{i+1} \\ 1 & \text{si} \quad x_r \leq x \end{cases}$$

où F_{i+1} est la fréquence cumulée croissante associée à la valeur x_{i+1}.

2.7.2 Cas des variables continues

a) Histogramme

Un **histogramme** est un ensemble de rectangles contigus. Chaque rectangle ayant pour base la classe et une aire proportionnelle à l'effectif ou à la fréquence de la classe.

- Dans le cas où les classes ont la même amplitude, il suffit que chaque rectangle ait une hauteur proportionnelle à l'effectif de chaque classe.

Nous traçons le polygone des fréquences en joignant les milieux des segments su-

périeurs de chaque rectangle (en ajoutant éventuellement deux classes de même amplitude et d'effectif nul de chaque côté de l'histogramme).
Le polygone des fréquences ou des effectifs a une surface égale à la surface de l'histogramme.

- Dans le cas où les classes n'ont pas la même amplitude, on construit des rectangles dont la hauteurs est proportionnelle à l'effectif corrigé.
Lorsque les classes sont d'amplitudes inégales, on peut également tracer le polygone des effectifs mais en prenant soin que la surface de celui-ci soit égale à la surface de l'histogramme.

Remarque 2.7.2. *Dans le cas d'une variable continue, une classe modale est celle qui a le rectangle le plus haut.*

b) Coube de la fonction de répartition

Rappelons que les individus étant regroupés en classes, la fréquence cumulée associée à la classe $Cli = [e_i,\, e_{i+1}[$ est la proportion d'individus dont la valeur du

caractère est strictement inférieure à e_{i+1}.
La courbe de la fonction de répartition F est affine sur chaque intervalle $[e_i,\ e_{i+1}[$.

2.7.3 Cas des variables qualitatives

Les représentations graphiques des caractères qualitatifs sont très nombreuses et sont fonctions des différentes modalités du caractère. Nous choisissons d'en présenter deux.

a) Diagramme à bandes ou diagramme en tuyaux d'orgue

En abscisse, sont disposées les différentes modalités de façon arbitraire auxquelles on associe des bandes (ou des rectangles) espacées entre elles, de largeur constante dont les hauteurs (en ordonnées) sont proportionnelles à l'effectif ou à la fréquence de chaque modalité. Les bases des bandes doivent être égales et équidistantes.

b) Diagramme à secteurs

L'effectif total est représenté par un disque (ou un demi-disque).

Chaque modalité est représentée par un secteur circulaire dont la surface est proportionnelle à l'effectif correspondant.

Dans le cas d'un disque, l'effectif total ou la fréquence 1 correspond à l'angle 360.

Dans le cas d'un demi-disque, l'effectif total ou la fréquence 1 correspond à l'angle 180.

Chapitre 3

Séries statistiques à deux variables

3.1 Généralités

Soient deux variables X et Y définies sur une même population d'effectif total N. La distribution ou la série statistique à deux variables relative au couple (X, Y) est définie par la donnée de :

- p valeurs possibles $x_1, x_2, ..., x_p$ de X ;
- q valeurs possibles $y_1, y_2, ..., y_q$ de Y
- des effectifs n_{ij} associés à chaque couple (x_i, y_j) correspondant aux observations ($X = x_i$ et $Y = y_j$).

Remarque 3.1.1. *Si X est une variable aléatoire continue, x_i est remplacé par c_i (le centre de la i^{eme} classe). Il en est de même pour Y.*

3.2 Présentation des tableaux statistiques

3.2.1 Tableau de données ponctuelles

Le tableau de données ponctuelles est utilisé lorsque à chaque x_i correspond une et une seule valeur y_i. On dit alors que les données ne sont pas groupées.

Le tableau de données ponctuelles se présente comme suit :

X	x_1	x_2	...	x_N
Y	y_1	y_2	...	y_N

3.2.2 Tableau à double entrée ou tableau de contingence

Il se présente sous la forme suivante :

$X \setminus Y$	y_1	y_2	...	y_j	...	y_q
x_1	n_{11}	n_{12}	...	n_{1j}	...	n_{1q}
x_2	n_{21}	n_{22}	...	n_{2j}	...	n_{2q}
...						
x_i	n_{i1}	n_{i2}	...	n_{ij}	...	n_{iq}
...						
x_p	n_{p1}	n_{p2}	...	n_{pj}	...	n_{pq}

A l'intersection de la i^{eme} ligne et de la j^{eme} colonne, on reporte l'effectif n_{ij} correspondant à l'observation $x = x_i$ et $Y = y_j$.

3.3 Représentation graphique

Dans un repère cartésien, on place les points de coordonées $(x_i,\, y_j)$ en indiquant, éventuellement, entre parenthèses l'effectif n_{ij}. On construit ainsi un nuage de points pondérés.

3.4 Distributions asssociées et indépendance statistique

3.4.1 Distributions marginales ou séries marginales

Etant donné un tableau de contingence ou un tableau à double entrée, on peut déduire les séries séparées de X et de Y. Ces séries sont dites marginales.
En indiquant par un point une sommation effectuée suivant l'indice i ou j, on note :
$n_{i\bullet}$ le total des effectifs de la i^{eme} ligne :

$$n_{i\bullet} = \sum_{j=1}^{q} n_{ij} = n_{i1} + n_{i2} + \ldots + n_{iq}$$

$n_{\bullet j}$ le total des effectifs de la j^{eme} colonne :

$$n_{\bullet j} = \sum_{i=1}^{p} = n_{1j} + n_{2j} + \ldots + n_{pj}$$

N l'effectif total :

$$\begin{aligned} N &= n_{\bullet\bullet} = \sum_{i=1}^{p} n_{i\bullet} = \sum_{j=1}^{q} n_{\bullet j} \\ &= \sum_{i=1}^{p} \sum_{j=1}^{q} n_{ij}. \end{aligned}$$

La distribution marginale de X ou la série marginale de X est l'ensemble des couples $(x_i,\, n_{i\bullet})$. Elle est donnée par le tableau suivant :

x_i	$n_{i\bullet}$
x_1	$n_{1\bullet}$
x_2	$n_{2\bullet}$
...	...
x_p	$n_{p\bullet}$

La distribution marginale de Y ou la série marginale de Y est l'ensemble des couples $(y_j,\, n_{\bullet j})$. Elle est donnée par le tableau suivant :

y_j	$n_{\bullet j}$
y_1	$n_{\bullet 1}$
y_2	$n_{\bullet 2}$
...	...
y_q	$n_{\bullet q}$

3.4.2 Fréquences

• La fréquence de l'observation $X = x_i$ et $Y = y_j$ est notée f_{ij} et est définie par

$$f_{ij} = \frac{n_{ij}}{N}.$$

• La fréquence marginale de l'observation $X = x_i$ est notée $f_{i\bullet}$ et est définie par

$$f_{i\bullet} = \frac{n_{i\bullet}}{N} = \sum_{j=1}^{q} f_{ij}.$$

• La fréquence marginale de l'observation $Y = y_j$ est notée $f_{\bullet j}$ et est définie par

$$f_{\bullet j} = \frac{n_{\bullet j}}{N} = \sum_{i=1}^{p} f_{ij}$$

Remarque 3.4.1.

$$\sum_{i=1}^{p} f_{i\bullet} = \sum_{j=1}^{q} f_{\bullet j} = 1 = f_{\bullet\bullet}.$$

Exemple : On considère la série statistique double donnée par le tableau suivant :

$X \setminus Y$	0	2
1	2	4
2	1	2
4	3	5

1. Déterminer les séries marginales de X et de Y.
2. En déduire l'effectif total N.

3. Calculer les fréquences marginales
4. Représenter graphiquement cette série statistique.

Remarque 3.4.2. *La connaissance des distributions marginales de X et de Y ne suffit pas en général pour déterminer la distribution du couple $(X,\ Y)$. Cela n'est possible que si X et Y sont indépendantes.*

3.4.3 Indépendance statistique

Deux variables X et Y sont indépendantes si et seulement si pour tout $i \in \{1,\ 2,\ ,...,\ p\}$ et pour tout $j \in \{1,\ ...,\ q\}$,

$$f_{ij} = f_{i\bullet} \times f_{\bullet j} \quad \text{où} \quad n_{ij} = \frac{n_{i\bullet \times n_{\bullet j}}}{N}.$$

Remarque 3.4.3. *L'indépendance statistique de X et de Y correspond au fait que les lignes (ou les colonnes) du tableau à double entrée sont proportionnelles.*

3.4.4 Distributions conditionnelles

• Pour i fixé et j variant de 1 à q, l'ensemble des couples $(y_j,\ n_{ij})$ est la distribution conditionnelle de Y sachant $X = x_i$.
• Pour j fixé et i variant de 1 à p, l'ensemble des couples $(x_i,\ n_{ij})$ est la distribution conditionnelle de X sachant $Y = y_j$.
• La fréquence conditionnelle de $Y = y_j$ sachant $X = x_i$ se note f_{y_j/x_i} et est définie par

$$f_{y_j/x_i} = \frac{n_{ij}}{n_{i\bullet}}.$$

• La fréquence conditionnelle de $X = x_i$ sachant $Y = y_i$ se note f_{x_i/y_j} et est définie par

$$f_{x_i/y_j} = \frac{n_{ij}}{n_{\bullet j}}.$$

Remarque 3.4.4. *1. On a les égalités suivantes :*

$$\sum_{j=1}^{q} f_{y_j/x_i} = 1, \qquad \sum_{i=1}^{p} f_{x_i/y_j} = 1$$

et

$$f_{ij} = f_{y_j/x_i} \times f_{i\bullet} = f_{x_i/y_j} \times f_{\bullet j}.$$

2. *Les variables X et Y sont indépendantes si et seulement si pour tout i et pour tout j*

$$f_{y_j/x_i} = f_{\bullet j} \quad ou \quad f_{x_i/y_j} = f_{i\bullet}.$$

3. *La connaissance de la distribution marginale de X (respectivement de Y) et des distributions conditionnelles de Y (respectivement de X) sachant $X = x_i$ (respectivement sachant $Y = y_j$) suffit pour déterminer la distribution du couple (X, Y).*

Exemple : On considère la statistique double définie par le tableau de contingence suivant :

$X \ \backslash \ Y$	y_1	y_2	y_3
x_1	128	81	17
x_2	64	22	88

Déterminer : $n_{\bullet\bullet}$, n_{12}, n_{21}, $n_{2\bullet}$, $n_{\bullet 2}$ et les fréquences suivantes :
f_{23}, $f_{\bullet 1}$, f_{y_2/x_1} , f_{y_1/x_2}, f_{x_2/y_1} et f_{x_2/y_2}.

3.5 Paramètres d'une série statistique à deux variables

Dans cette section, on suppose que X et Y sont des variables quantitatives.

3.5.1 Cas d'un tableau de données ponctuelles

a) Moyenne

$$\overline{X} = \frac{1}{N}\sum_{i=1}^{N} x_i \qquad \text{et} \qquad \overline{Y} = \frac{1}{N}\sum_{i=1}^{N} y_i$$

b) Variance et écart-type

$$Var(X) = \left(\frac{1}{N} \sum_{i=1}^{N} x_i^2 \right) - \overline{X}^2$$

et

$$Var(Y) = \left(\frac{1}{N} \sum_{i=1}^{N} y_i^2 \right) - \overline{Y}^2.$$

Donc

$$\sigma_X = \sqrt{Var(X)} \qquad \text{et} \qquad \sigma_Y = \sqrt{Var(Y)}$$

c) Covariance

La covariance de $(X,\ Y)$ est notée $Cov(X,\ Y)$ et est définie par

$$\begin{aligned} Cov(X,\ Y) &= \frac{1}{N} \sum_{i=1}^{N} (\overline{X} - x_i)(\overline{Y} - y_i) \\ &= \left(\frac{1}{N} \sum_{i=1}^{N} x_i y_i \right) - \overline{X}\,\overline{Y} \end{aligned}$$

La covariance est un indicateur du sens de la variation simultannée de X et de Y. Ainsi, si X et Y évoluent dans un même sens alors $Cov(X,\ Y) > 0$.

Remarque 3.5.1. *1.* $Cov(X,\ X) = Var(X)$.

2. $Cov(X,\ Y) = Cov(Y,\ X)$.

d) Coefficient de corrélation linéaire

On appelle coefficient de corrélation linéaire entre X et Y le nombre noté ρ défini par

$$\rho = \frac{Cov(X,\ Y)}{\sqrt{Var(X)Var(Y)}} = \frac{Cov(X,\ Y)}{\sigma_X\,\sigma_Y}.$$

Remarque 3.5.2. *Quel que soit le couple de variables* $(X,\ Y)$*,*

$$|\rho| \leq 1 \quad \textit{c'est-à-dire} \quad -1 \leq \rho \leq 1.$$

Afin de calculer les différents paramètres, on adopte la disposition pratique suivante :

	x_i	y_i	x_i^2	y_i^2	x_iy_i
	x_1	y_1	x_1^2	y_1^2	x_1y_1
	...	...	...	...	...
	x_N	y_N	x_N^2	y_N^2	x_Ny_N
Total					

3.5.2 Cas d'un tableau de contingence

a) Moyenne marginale

$$\overline{X} = \frac{1}{N}\sum_{i=1}^{p} n_{i\bullet}x_i \qquad \text{et} \qquad \overline{Y} = \frac{1}{N}\sum_{j=1}^{q} n_{\bullet j}y_j.$$

b) Variance et écart-type

$$Var(X) = \left(\frac{1}{N}\sum_{i=1}^{p} n_{i\bullet}x_i^2\right) - \overline{X}^2$$

$$Var(Y) = \left(\frac{1}{N}\sum_{j=1}^{q} n_{\bullet j}y_j^2\right) - \overline{Y}^2$$

$$\sigma_X = \sqrt{Var(X)} \quad \text{et} \quad \sigma_Y = \sqrt{Var(Y)}.$$

c) Covariance et coefficient de corrélation linéaire

$$\begin{aligned} Cov(X, Y) &= \frac{1}{N}\sum_{i=1}^{p}\sum_{j=1}^{q} n_{ij}(\overline{X} - x_i)(\overline{Y} - y_j) \\ &= \left(\frac{1}{N}\sum_{i=1}^{p}\sum_{j=1}^{q} n_{ij}\, x_i\, y_j\right) - \overline{X}\,\overline{Y} \end{aligned}$$

et

$$\rho = \frac{Cov(X, Y)}{\sigma_X\, \sigma_Y}.$$

Afin de calculer les différents paramètres, on adopte la disposition pratique suivante :

$X \setminus Y$	y_1	...	y_q	$n_{i\bullet}$	$n_{i\bullet}\, x_i$	$n_{i\bullet}\, x_i^2$	$\sum_{j=1}^{q} n_{ij} y_j$	$x_i \sum_{j=1}^{q} n_{ij} y_j$
x_1								
...								
x_p								
$n_{\bullet j}$				$N =$				$S =$
$n_{\bullet j} y_j$								
$n_{\bullet j} y_j^2$								
$\sum_{i=1}^{p} n_{ij} x_i$								
$y_j \sum_{i=1}^{p} n_{ij} x_i$				$S =$				

où

$$S = \sum_{i=1}^{p} \sum_{j=1}^{q} n_{ij}\, x_i\, y_j.$$

Ainsi,

$$Cov(X, Y) = \frac{1}{N} S - \overline{X}\,\overline{Y}.$$

3.6 Ajustement

3.6.1 Régression-corrélation

Soit (X, Y) une série statistique double. On s'intéresse ici à une liaison éventuelle entre les variables X et Y.

La régression porte sur une expression de cette liaison sous forme d'une fonction mathématique.

Lorsqu'on cherche à exprimer Y en fonction de X ($Y = f(X)$), on effectue une régression de Y en X. Si l'on cherche à exprimer X en fonction de Y ($X = g(Y)$), on effectue une régression de X en Y.

En général, les courbes de régression représentent des fonctions complexes et difficiles à exploiter. On préfère donc choisir en tenant compte du nuage de points une famille de fonctions simples (fonctions polynômiales le plus souvent) de laquelle on détermine l'élément dont la représentation graphique est le plus proche possible du nuage de points.

La corrélation renseigne sur l'intensité de la liaison entre les variables X et Y. Ainsi, si X et Y sont indépendantes c'est-à-dire s'il n'y a aucune liaison entre X et Y ont dit que X et Y sont non correlés. Et dans ce cas, $Cov(X,\ Y) = 0$.

Remarque 3.6.1. *$Cov(X,\ Y) = 0$ n'entraîne pas que X et Y sont indépendantes.*

3.6.2 Ajustement affine

Dans le cas où le nuage de points présente une forme allongée et linéaire, on choisit la famille des fonctions affines définies par $f(x) = ax + b$, où a et b sont des nombres réels à déterminer. On peut être ramené à ce cas après une transformation des données par exemple en posant $U = ln(X)$ ou / et $V = ln(Y)$.

3.6.3 Ajustement affine par la méthode des moindres carrées

a) Droite d'ajustement de Y en X ou droite de régression de Y en X.

Elle est notée $(D_{Y/X})$ et son équation s'écrit $y = ax + b$ où

$$a = \frac{Cov(X,\ Y)}{Var(X)} \quad \text{et} \quad b = \overline{Y} - a\overline{X}$$

car $(D_{Y/X})$ passe par le point moyen G de coordonnées $(\overline{X},\ \overline{Y})$ encore appelé centre de gravité du nuage de points.

b) Droite d'ajustement de X en Y ou droite de régression de X en Y.

Elle est notée $(D_{X/Y})$ et son équation est $x = \widetilde{a}y + \widetilde{b}$, où

$$\widetilde{a} = \frac{Cov(X,\ Y)}{Var(Y)} \quad \text{et} \quad \widetilde{b} = \overline{X} - \widetilde{a}\overline{Y}.$$

Remarque 3.6.2. 1. *L'ajustement linéaire par la méthode des moindres carrées est considéré comme justifié si le coefficient de corrélation linéaire ρ vérifie*

$$|\rho| \geq \frac{\sqrt{3}}{2} \simeq 0,867.$$

Dans ce cas, on dit qu'il y a une forte corrélation linéaire entre X et Y.

2. *Si* $0,7 \leq |\rho| < \frac{\sqrt{3}}{2}$*, on dit qu'il y a une corrélation linéaire moins forte entre* X *et* Y.
3. *Si* $0 < |\rho| < 0,7$*, il y a une faible corrélation linéaire entre* X *et* Y.
4. *Si* $\rho = 0$ *alors il n'existe pas de corrélation linéaire entre* X *et* Y.
5. *Si* $|\rho| = 1$ *alors la droite de régression de* Y *en* X *et la droite de régression de* X *en* Y *sont confondues.*
6. $a\,\widetilde{a} = \rho^2$.

Bibliographie

[1] **G. CHAUVAT et J-P. RÉAU**, *Statistiques descriptives*, Armand Colin/ HER, Paris, (2001).

[2] **F. MAZEROLE**, *Statistique descriptive*, Gualino éditeur, EIA-Paris, (2006).

[3] **H. MZALI**, *Statistique et calcul de probabilité (cours)*, Ecole Nationale d'Administration (République Tunisienne), (2013).

[4] **D. SCHLACTHER**, *De l'analyse à la prévision*, Ellipses , (1994).

ECUE 1 : Séries statistiques à une variable

Exercice 1 Les caractères suivants sont-ils qualitatifs ou quantitatifs ?

a) sexe
b) situation matrimoniale
c) taille
d) couleurs des yeux
e) âge
f) salaire mensuel
g) tension artérielle
h) quartier habité.

Exercice 2 Un nouveau concessionnaire de voitures neuves a enregistré au cours de ses 40 premières semaines d'activité le nombre d'automobiles qu'il a vendues par semaine. Il a obtenu les résultats suivants :

$$\begin{array}{cccccccccccccc} 0 & 1 & 1 & 0 & 3 & 1 & 2 & 2 & 1 & 5 & 5 & 2 & 2 & 2 \\ 2 & 2 & 4 & 4 & 4 & 4 & 4 & 4 & 6 & 3 & 3 & 3 & 3 & \\ 6 & 4 & 4 & 4 & 4 & 5 & 3 & 3 & 3 & 5 & 5 & 5 & 6 & \end{array}$$

1. Déterminer les éléments suivants : population statistique, individu statistique, taille de la population statistique, caractère étudié et sa nature, nombre de modalités, les modalités.
2. Présenter le tableau statistique en y faisant figurer le dépouillement.
3. Représenter graphioquement la série statistique.

Exercice 3 L'étude statistique d'une population a permis de regrouper les individus par classes d'égale amplitude dont les centres sont les suivants :

$$52 \quad 60 \quad 68 \quad 76 \quad 84 \quad 92$$

1. Quelle est l'amplitude de chaque classe ?
2. Calculer la limite inférieure et la limite supérieure de chaque classe.

Exercice 4 Compléter le tableau ci-contre, sachant que la première et la dernière modalité ont des effectifs égaux.

Modalités	Fréquences
A	...
B	$0,12$
C	$0,34$
D	0.27
E	0.15
F	...

Exercice 5 Compltéter le tableau suivant :

Classes	Centres	Fréquences	FCC
			0
$[10;\ 20[$	...	$0,08$	
$[20;\ 30[$	...	$0,21$	
$[30;\ 40[$	...		0,55
$[40;\ 60[$	...		0,86
$[60;\ 80[$	...		

(FCC désigne les fréquences cumulées croissantes.)

Exercice 6 On considère l'ensemble des notes obtenues, lors d'un examen noté sur 20, par 50 étudiants.

10	8	3	12	13	9	12	9	12	11
11	11	8	5	13	14	14	6	12	16
7	11	10	10	2	15	12	10	1	14
11	7	8	10	13	9	13	9	7	13
11	19	9	4	10	8	9	6	7	14

1. Quel est la population étudiée ?
2. Déterminez le caractère étudié et précisez sa nature.
3. Dépouiller ces données et présenter les résultats dans un tableau. (On prendra les classes suivantes : $[0;\ 5[$, $[5;\ 7[$,$[7;\ 9[$, $[9;\ 11[$, $[11;\ 13[$, $[13;\ 15[$, $[15;\ 20[$).
4. Calculer les fréquences cumulées croissantes.
5. Quelle est la proportion des étudiants ayant une note strictement inférieur à 9 ?
6. Quelle est la proportion des étudiants ayant une note supérieure ou égale à 13 ?
7. Quelle est la proportion des étudiants ayant une note comprise entre 5 et 20 ?

ECUE 1 : Séries statistiques à une variable

Exercice 1 Déterminer les quartiles de chacune des séries statistiques suivantes

1) 2 13 17 22 1 3 9 14 12 20 16 15 11 6 5.

2) 8 13 11 7 1 3 9 15 12 20 16 18 6 5.

Exercice 2 Dans une bibliothèque, l'ensemble des abonnnés a été reparti suivant le nombre d'ouvrage empruntés durant un mois :

x_i	0	1	2	3	4	5	6	7
n_i	18	39	57	64	42	33	21	14

Déterminer le mode, la moyenne, la médiane, le coefficient d'asymétrie de Fisher et le coefficient d'aplatissement de Pearson de cette série.
Interpréter les résultats obtenus.

Exercice 3 On dispose d'une variable X dont la distribution statistique est la suivante :

x_i	x_1	x_2	9	11
n_i	2	1	1	1

x_1 et x_2 étant inconnues.
Sachant que la moyenne arithmétique de X est égale à $6,2$ et que la variance de X est égale à $10,56$, déterminer x_1 et x_2.

Exercice 4 Le tableau suivant donne la répartion d'une population par tranches d'âge :

Classes	$[0,\ 10[$	$[10,\ 20[$	$[20,\ 30[$	$[30,\ 40[$	$[40,\ 50[$	$[50,\ 60[$	$[60,\ 70[$	$[70,\ 80[$
Nombre	18	44	68	54	42	36	16	10

1. Quelle est l'étendue de cette série ?
2. Déterminez la classe modale de cette série et calculer le mode.
3. Calculer l'âge moyen de cette population.
4. Calculer les quartiles de cette série statistique et interpréter les résultats obtenus.
5. Calculer le coefficient d'asymétrie de Yule de cette série et interpréter le résultat obtenu.
6. Calculer l'écart interdécile
7. Calculer l'écart-type et interpréter le résultat obtenu.

Exercice 5 On donne les deux distributions ci-dessous concernent la répartition des superficies des exploitations agricoles dans deux régions :

Superficies (en hectare)	pourcentage des exploitations de la région A	pourcentage des exploitations de la région B
1 à 5	33	20
5 à 10	20	16
10 à 20	24	22
20 à 50	20	30
50 à 200	3	12

Comparer les dispersions des superficies de ces deux régions à l'aide des écart-types et des coéfficients de variations. Que peut-on conclure ?

Exercice 6 Le tableau suivant donne la répartition des ménages d'une population suivant le nombre de véhicules automobiles possédés.

Véhicules automobiles	0	1	2	3	4
Nombre de menages	488	1872	884	186	18

1. Donner la représentation graphique de cette série.
2. Déssiner le diagramme cumulatif.

Exercice 7 Le tableau suivant donne les résultats d'une enquête sur le poids (en kg) des individus d'une population :

48	72	54	80	58	70	69	58	57	60
85	94	78	81	64	49	54	57	57	62
63	69	72	71	82	87	64	65	73	58
61	67	49	52	60	66	69	89	84	82
73	70	72	58	64	51	65	77	79	80
59	57	81	78	76	79	68	67	53	59

1. Classer ces résultats par classes d'amplitude 5 avec comme première classe $[45,\ 50[$.
2. Calculer les fréquences et les fréquences cumulées croissantes.
3. Construire l'histogramme et le polygone des fréquences.
4. Construire la courbe des fréquences cumulées croissantes.
5. Déterminer graphiquement la proportion des individus pesant moins de 75 kg et comparer ce résultat avec la valeur calculée (question 2.)

Exercice 8

Dans cet exercice, les résultats de tous les calculs effectués seront donnés sans être arrondis.
On donne la distribution suivante pour laquelle α et β sont des inconnues.

Salaires en miliers de francs	$[20;\ 30[$	$[30;\ 40[$	$[40;\ \alpha[$	$[\alpha;\ 70[$	$[70;\ 100[$	$[100;\ \beta[$
Nombres d'ouvriers	100	140	125	200	180	55

1. Sachant que l'étendue de cette série statistique est de 130 000 francs, déterminer β.
2. Calculer les fréquences et les fréquences cumulées croissantes.
3. Sachant que le deuxième quartile de la série statistique est 56 800 francs, calculer α.
4. Sachant que α=52 500 francs et β=120 000 francs, déterminer le mode de la série statistique.

Exercice 9 On considère une population de 1000 individus répartis en fonction de leur âge :

Age	$[0,\ 10[$	$[10,\ 15[$	$[15,\ 20[$	$[20,\ 30[$	$[30,\ 40[$	$[40,\ 60[$	$[60,\ 80[$
Nombre	120	100	140	200	180	160	100

1. Déterminer la classe modale de cette série statistique et calculer le mode.
2. Construire un histogramme représentant cette population.
3. Tracer le polygone des effectifs.

ECUE 2 : Séries statistiques à deux variables

Exercice 1 On considère la statistique double définie par le tableau de contingence suivant :

$X \ \backslash\ Y$	y_1	y_2	y_3
x_1	128	81	17
x_2	64	22	88

1. Déterminer N (l'effectif totoal), $n_{1\,2}$ $n_{2\,2}$ $n_{2\bullet}$ $f_{2\,3}$ $f_{1\bullet}$ f_{y_2/x_1} f_{x_2/y_2}.
2. Donner la distribution conditionnelle de X sachant que $Y = y_2$.

Exercice 2 Considérons la distribution statistique du couple $(X,\ Y)$ définie par le tableau suivant :

$X \ \backslash\ Y$	1	2	3	4
-1	10	42	80	17
0	21	7	12	4
1	32	8	6	11

Dresser le tableau correspondant des fréquences des fréquences $f_{i\,j}$ et des fréquences marginales $f_{i\,\cdot}$ et $f_{\cdot\,j}$.

Exercice 3 Compléter les tableaux suivants, sachant que dans chaque cas, X et Y sont indépendants.

1)

$X \ \backslash\ Y$	0	20	30	50	Distribution marginale de X
1					$0,45$
2					$0,55$
Distribution marginale de Y	$0,1$	0.3	0.4	0.2	1

2)

$X \ \backslash\ Y$	100	200	300
10	8	12	2
20		24	
30			5

Exercice 4 On considère la distribution statistique donné par le tableau de contingence suivant :

$X \ \backslash\ Y$	3	5	7
7	1	3	1
1	2	2	2

1. Caluler la moyenne de X et la moyenne de Y.
2. Calculer l' écart-type de X et l'écart-type de Y.
3. Calculer Cov(X, Y)
4. Les variables X et Y sont-elles indépendantes ?
5. Représenter graphiquement cette série statistique.

Exercice 5 A l'oral du Baccalauréat série A, chaque candidat est interrogé en première langue où il obtient la note X et en seconde langue où il obtient la note Y (notes sur 20).
Les résultats obtenus par 100 candidats sont donnés dans le tableau ci-dessous.

$X \setminus Y$	$[0;\ 4[$	$[4;\ 8[$	$[8;\ 12[$	$[12;\ 16[$	$[16;\ 20[$
$[0;\ 4[$	2	5	2		
$[4;\ 8[$	1	12	10	3	
$[8;\ 12[$		3	28	12	1
$[12;\ 16[$		1	5	10	2
$[16;\ 20[$				1	2

1. Déterminer, dans deux tableaux différents, les distributions marginales en X et en Y.
2. Déterminer la distribution conditionnelle de X sachant $Y \in [4;\ 8[$.
3. Calculer la fréquence conditionnelle dans chacun des cas suivants :
   a) $X \in [8;\ 12[$ sachant $Y \in [0;\ 4[$;
   b) $Y \in [12;\ 16[$ sachant $X \in [4;\ 8[$.
   NB : Les résultats seront donnés sous forme de fraction irréductible.

Exercice 6

Le nuage de points d'une série statistique double $(X,\ Y)$ a été ajusté par ses deux droites de regression. Ces deux droites ont pour equation dans le même repère :

$$y = -3x + 7 \quad \text{et} \quad y = -\frac{10}{3}x + \frac{23}{3}.$$

1. Déterminer les coordonnées du point moyen G de ce nuage de points.
2. Calculer le coefficient de corrélation linéaire et interpreter le résultat obtenu.
3. Sachant que la somme de la variance de X et la variance de Y est égale à 22, calculer :
   a) la variance de X,
   b) la variance de Y,
   c) la covariance de X et Y.

Exercice 7 Soit la répartition par taille et par poids de 40 animaux d'une race donnée.

X taille en cm $\setminus$ Y : Poids en kg	$[3,\ 9[$	$[9,\ 13[$	$[13,\ 19[$
De 60 à 100	15	1	0
De 100 à 120	2	10	1
De 120 à 140	1	2	1
De 140 à 180	0	0	17

1. Calculer le coefficient de correlation linéaire entre X et Y. Que peut-on en déduire ?
2. Déterminer l'équation de la droite de regression de Y en X et celle de X en Y.

Exercice 8

Dans cet exercice, les résultats seront éventuellement arrondis au dixième près.

On cobsidère la distribution donnée par le tableau à double entrée suivant :

$X \diagdown Y$	7	9	16
3	2	3	7
4	12	18	42
5	6	9	21
8	4	6	14

1. Représenter graphiquement cette série statistique
2. Déterminer le coefficient de correlation linéaire de $(X,\ Y)$.

Année : 2014-2015

Université

NANGUI ABROGOUA

EXAMEN DE STATISTIQUE DESCRIPTIVE
ECUE 1-STATAISTIQUE A UNE VARIABLE
Session 1 – Epreuve A

L1 Maths-informatique

Durée 01h 30mn

Unité de Formation et de Recherche des Sciences Fondamentales et Appliquées

---

Ce sujet comporte 4 pages

EXERCICE I :

Pour chaque question, une seule réponse est exacte. On ne demande pas de justifier.
Chaque réponse juste rapporte 1 point.
Une absence de réponse vaut 0 point.
Il sera retiré 1 point par réponse fausse.

1. Dans une série statistique, l'écart absolu moyen par rapport à la moyenne est plus petit que l'écart absolu moyen par rapport à la médiane.

A. Vrai
B. Faux

2. Soient X et Y deux variables statistiques telles que Y=2-X. Alors X et Y ont le même écart-type.

A. Vrai
B. Faux

3. Le coefficient de variation d'une série statistique est un paramètre de dispersion.

A. Vrai
B. Faux

4. La variance d'une variable statistique s'exprime dans la même unité que la moyenne arithmétique de cette même variable.

A. Vrai
B. Faux

5. Quand les classes d'une série statistique sont d'amplitude inégale, il faut obligatoirement corriger les fréquences pour calculer les quartiles.

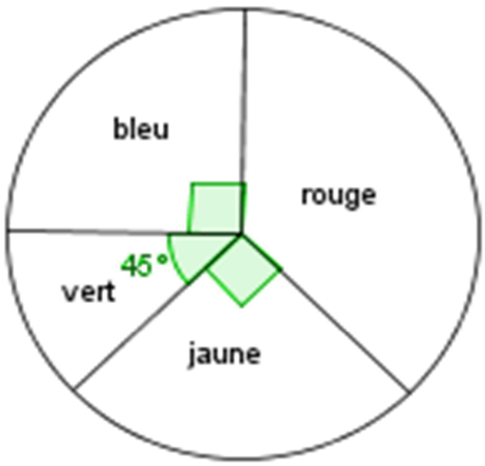
A. Vrai

B. Faux

EXERCICE II

Pour chaque question, il peut y avoir plusieurs réponses exactes. On ne demande pas de justifier.
Chaque réponse juste rapporte 1 point.
Une absence de réponse vaut 0 point.
Il sera retiré 0.5 point par réponse fausse.

6. On considère la série statistique représenté par le diagramme suivant :



A. La fréquence de la réponse « bleu » est 24%

B. La fréquence de la réponse « rouge » est 37,5%

C. La fréquence de la réponse « jaune » est 25%

D. La fréquence de la réponse « vert » est 12,5%

7. La classe modale d'une variable quantitative continue est :

A. la classe ayant la plus grande amplitude

B. la classe ayant le plus grand effectif corrigé

C. la classe ayant la plus grande fréquence

D. la classe ayant la plus grande densité

EXERCICE III

Pour chaque question, une seule réponse est exacte. On ne demande pas de justifier.
Chaque réponse juste rapporte 2,5 points.
Une absence de réponse vaut 0 point.
Il sera retiré 2 points par réponse fausse.

Une étude consacrée au budget hebdomadaire d'appels téléphoniques auprès de ménages du quartier « AKWABA » a donné les résultats suivants :

Budget (en francs CFA)	Fréquences cumulées croissantes
	0
[800 ; 1000[	
	0,08
[1000 1400[	
	0,18
[1400 ; 1600[	
	0,34
[1600 ; y [	
	0,64
[y ; 2400[	
	0.73
[2400 ; x [	
	1

On suppose que l'étendue de cette série est égale à 3200 francs CFA.

8. La borne manquante x est (en francs CFA)

A. 5500
B. 4000
C. 4500
D. 3000

9. Sachant que le budget moyen est égal à 1995 francs CFA, la borne manquante y est

A. 1800
B. 1700
C. 1600

D. 1900

10. Sachant que le budget médian est égal à 1920 francs CFA, la borne manquante y est

A. 2200
B. 2350
C. 2000
D. 2250

11. On suppose maintenant que y=2000 francs CFA.

Sachant que $\sum n_i c_i^2 = 4741200000$ (où c_i désigne le centre de chaque classe i) et Var(X)=604044, l'effectif de la classe [1600 ; y[est

A. 170
B. 300
C. 100
D. 270

Année : 2014-2015

Université

NANGUI ABROGOUA

EXAMEN DE STATISTIQUE DESCRIPTIVE
ECUE 2- STATAISTIQUE A DEUX VARIABLES
Session 1-Epreuve B

L1 Maths-informatique

Durée 01h 30mn

Unité de Formation et de Recherche des Sciences Fondamentales et Appliquées

---

Ce sujet comporte 3 pages

Pour chaque question, une seule réponse est exacte. On ne demande pas de justifier.
Chaque réponse juste rapporte 1,5 point.
Une absence de réponse vaut 0 point.
Il sera retiré 1,5 point par réponse fausse.

Uniquement pour les questions 1, 2, 3, 4, 5, 6 et 7, on considère la série statistique double donnée par le tableau suivant :

X \ Y	6	11	16
80	15	1	0
110	2	10	1
130	1	2	1
160	0	0	17

1. La fréquence marginale de l'observation Y=11 est 0,26

A. Vrai
B. Faux

2. La fréquence marginale de l'observation X=130 est 0,08.

A. Vrai
B. Faux

3. La fréquence conditionnelle de l'observation Y=6 sachant X=110 est 0,12.

A. Vrai
B. Faux

4. Le point moyen G de cette série statistique a pour coordonnées

A. (11 ; 119)
B. (119 ; 11,1)
C. (11,1 ; 119)
D. (119 ; 11)

5. La variance de X est 18,49

A. Vrai
B. Faux

6. La covariance du couple (X, Y) est

A. 18,49
B. 239
C. 1089
D. 129,1

7. Il y a une forte corrélation linéaire entre les variables X et Y.

A. Vrai
B. Faux

8. La covariance comme la variance est un nombre toujours positif.

A. Vrai
B. Faux

9. La connaissance des distributions marginales de X et de Y est suffisante pour déterminer la distribution du couple (X, Y).

A. Vrai
B. Faux

10. Si les deux droites d'ajustement (X en Y et Y en X) ont leur coefficient directeur négatif, alors le coefficient de corrélation linéaire de (X, Y) est

A. positif
B. négatif

11. Si les droites de régression de X en Y et de Yen X sont confondues alors le coefficient de corrélation linéaire vaut 1.

A. Vrai
B. Faux

12. Soit X et Y deux variables statistiques définies sur une même population. X et Y sont statistiquement indépendantes si et seulement si Cov (X, Y)=0. :

A. Vrai
B. Faux

13. Deux caractères X et Y, définis sur une même population d'effectif n=1000, sont statistiquement indépendants. Sachant que $n_{4.}$ =400 et $n_{.5}$=320. On peut affirmer que n_{45} est égal à

A. 80
B. 128
C. 720
D. 250

14. Si le coefficient de corrélation de (X, Y) est -0,6, alors le coefficient de corrélation linéaire de (Y, X) est

A. 0,6
B. -0,3
C. -0,6
D. 0,3

Université Nangui Abrogoua **Année Universitaire 2014/2015**
UFR SFA

A

Épreuve de la deuxième session de l'examen de

L'UE DE STATISTIQUE DESCRIPTIVE
ECUE 1 : SERIES STATISTIQUES UNIVARIÉES
Licence 1 : Maths - Info
Durée : 01 h 30 mn

Consigne : *Pour chacune des questions indiquer, sans justification, la bonne réponse ou les bonnes réponses sur la grille réponse.*

Barème : Pour les questions à choix direct, une réponse exacte rapporte $0,5$ point. Une réponse fausse enlève $0,5$ point. L'absence de réponse ne rapporte ni n'enlève aucun point.
Pour les questions à choix multiples, une réponse exacte rapporte $1,5$ point. Une réponse fausse enlève 1 point. L'absence de réponse ne rapporte ni n'enlève aucun point.

Exercice 1 : Questions à choix direct

Question 1 Le mode n'est pas défini pour les variables qualitatives nominales.

$\boxed{A}$ Vrai $\boxed{B}$ Faux

Question 2 Le moment d'ordre 1 est un paramètre de position.

$\boxed{A}$ Vrai $\boxed{B}$ Faux

Question 3 Le coefficient de variation est défini par le rapport : moyenne divisée par écart-type.

$\boxed{A}$ Faux $\boxed{B}$ Vrai

Question 4 Il est possible de déterminer la médiane pour les variables qualitatives ordinales.

$\boxed{A}$ Faux $\boxed{B}$ Vrai

Exercice 2 : Questions à choix multiple

Question 5 On a réalisé une étude sur un échantillon de 10 personnes. Voici les résultats obtenus

Nombre de médicaments	0	1	2	3	4
Nombre de personnes	2	4	0	3	1

$\boxed{A}$ La moyenne est égale à $1,5$

$\boxed{B}$ L'étendue de cette série est 5

$\boxed{C}$ La médiane est égale à 4

$\boxed{D}$ Le mode est égale à 1

Question 6 Le caractère quantitatif discret X admet la tableau de distribution suivant :

Valeurs	1	2	3	4	5
fréquence (en pourcentage)	$10,5$	$22,3$	$30,4$	$23,6$	$13,2$

La fréquence cumulée croissante pour $X = 3$ est

$\boxed{A}$ $67,2$ $\boxed{B}$ $30,4$ $\boxed{C}$ $63,2$ $\boxed{D}$ $32,8$

Question 7 Le tableau suivant rapporte la répartition des niveaux d'étude des patients d'un hôpital.

Niveau d'étude	Inférieur au Bac	Bac ou Bac +1	Bac + 2 ou Licence	Master et plus
Fréquence (en pourcentage)	30	15	25	30

$\boxed{A}$ La distribution de la variable niveau d'étude est unimodale.

$\boxed{B}$ Le niveau d'étude est une variable qualitative nominale.

$\boxed{C}$ Le niveau d'étude est une variable qualitative ordinale.

$\boxed{D}$ Dans un diagramme circulaire, le niveau d'étude "Bac + 2 ou Licence" serait représenté par un secteur d'angle 70 dégré

Question 8 On désigne par : $e_{\overline{X}}$ l'écart absolu moyen par rapport à la moyenne , e_{M_e} l'écart absolu moyen par rapport à la médiane et par σ l'écart-type d'une variable X. Alors on a

A $e_{\overline{X}} \leq e_{M_e} \leq \sigma$ B $e_{\overline{X}} \leq \sigma \leq e_{M_e}$ C $\sigma \leq e_{M_e} \leq e_{\overline{X}}$ D $e_{M_e} \leq e_{\overline{X}} \leq \sigma$

Question 9 On donne la série statistique suivante :

14 16 12 9 11 17 7 8 9 16 7 9 18

La médiane est égale à

A 12 B 18 C 11 D Autre réponse

Question 10 Dans une étude statistique, on a déterminé le nombre d'entorses chez 40 cascadeurs : 2 personnes présentent 0 entorse ; 4 personnes présentent 1 entorse ; 7 en présentent 2 ; 8 en présentent 3 ; 10 en présentent 4 ; 6 en présentent 5 et le reste des cascadeurs en présentent 6.

A L'entorse est l'unité satatistique

B 50% des cascaduers ont moins de 3 entorses

C L'amplitude de cette série est 6

D La variance est $2,4875$

Exercice 3 : Questions à choix multiple

On a relevé dans le tableau suivant les buts marqués par Didier Drogba lors de 50 matchs disputés dans l'équipe Chelsea entre 2010 et 2012.

Nombre de buts par match	0	1	2	3
Fréquence (en pourcentage)	48	38	12	2

Question 11 L'effectif cumulé des matchs durant lesquels Didier Drogba a marqué au plus un but est

A 48 B 14 C 43 D 86

Question 12 Le 3ème quartile de cette série est

A $37,5$ B 75 C 0 D 1

Question 13 La moyenne de cette série est

A $0,5$ B 1 C $0,68$ D $1,36$

Question 14 Le caractère étudié est

A le nombre de buts par match

B tous les matchs joués par Didier Drogba

C le nombre de matchs disputé par Chelsea

D le pourcentage de matchs

Question 15 Dans cette série, l'effectif correspondant à la valeur 3 est

A 1 B 2 C 50 D 100

Université Nangui Abrogoua **Année Universitaire 2014/2015**
UFR SFA

D

Épreuve de la deuxième session de l'examen de

L'UE DE STATISTIQUE DESCRIPTIVE
ECUE 2 : SERIES STATISTIQUES BIVARIÉES
Licence 1 : Maths - Info
Durée : 01 h 30 mn

Consigne : *Pour chacune des questions indiquer, sans justification, la bonne réponse sur la grille réponse.*

Barème : Pour les questions à choix direct, une réponse exacte rapporte 0,5 point. Une réponse fausse enlève 0,5 point. L'absence de réponse ne rapporte ni n'enlève aucun point.
Pour les questions à choix multiples, une réponse exacte rapporte 1,5 point. Une réponse fausse enlève 1 point. L'absence de réponse ne rapporte ni n'enlève aucun point.

Exercice 1 : Questions à choix direct

Question 1 Si le coefficient de corrélation linéaire vaut -1, alors les droites de régression de X en Y et de Y en X sont confondues.

[A] Vrai [B] Faux

Question 2 Si X et Y sont deux variables statistiquement indépendantes, alors les distributions marginales de X et de Y suffisent pour déterminer la distribution du couple (X, Y).

[A] Vrai [B] Faux

Question 3 La covariance de (X, Y), lorsqu'elle est non nulle, est un indicateur du sens de variation simultanéé des variables X et Y.

[A] Faux [B] Vrai

Question 4 Si $Cov(X, Y) = 0$ alors X et Y sont linéairement indépendantes.

[A] Vrai [B] Faux

Exercice 2 : Questions à choix multiple

Soit la série statistique double définie par le tableau suivant :

i	1	2	3	4	5
x_i	10	12	18	24	36
y_i	50	60	70	90	130

Question 5 La variance de Y est

[A] autre réponse [B] 488 [C] 800 [D] 560

Question 6 La covariance de (X, Y) est

[A] 264 [B] 182 [C] autre réponse [D] 356

Question 7 La droite d'ajustement de Y en X a pour équation

[A] $y = 3x - 20$ [B] $y = -3x + 20$ [C] $y = 3x + 20$ [D] autre réponse

Question 8 La corrélation entre (X, Y) est

[A] faible [B] très faible [C] très forte [D] forte

Exercice 3 : Questions à choix multiple

La population d'un quartier d'Abidjan est décrite par l'âge (X) et le nombre de numéros de téléphone possédés (Y).

$X \setminus Y$	0	1	2	3
[15 , 25[	8	8	5	2
[25 , 35[	7	20	22	20
[35 , 50[	11	34	40	37
[50 , 65[	15	37	39	35
[65 , 95[	9	31	33	12

Question 9 Le nombre d'habitants de ce quartier est

$\boxed{A}$ 425 $\boxed{B}$ 450 $\boxed{C}$ autre réponse $\boxed{D}$ 400

Question 10 La fréquence marginale $f_{2\bullet}$ est

$\boxed{A}$ $0,17$ $\boxed{B}$ $0,18$ $\boxed{C}$ autre réponse $\boxed{D}$ $0,15$

Question 11 La fréquence marginale $f_{\bullet 3}$ (arrondie au millième près) est

$\boxed{A}$ $0,250$ $\boxed{B}$ $0,327$ $\boxed{C}$ autre réponse $\boxed{D}$ 0.249

Question 12 La formule permettant de calculer $f_{2|1}$ est

$\boxed{A}$ $\frac{n_{1\,2}}{n_{1\bullet}}$ $\boxed{B}$ $\frac{n_{1\,2}}{n_{2\bullet}}$ $\boxed{C}$ $\frac{n_{2\,1}}{n_{1\bullet}}$ $\boxed{D}$ $\frac{n_{1\,2}}{n_{\bullet 1}}$

Question 13 L'âge moyen de cette population est

$\boxed{A}$ $51,1$ $\boxed{B}$ autre réponse $\boxed{C}$ $51,3$ $\boxed{D}$ $51,2$

Question 14 La fréquence conditionnelle $X \in [15\,,\,25[$ sachant $Y = 0$ est

$\boxed{A}$ autre réponse $\boxed{B}$ 0.35 $\boxed{C}$ $0,16$ $\boxed{D}$ $0,3$

Question 15 La fréquence $f_{3\,2}$ est égale à

$\boxed{A}$ autre réponse $\boxed{B}$ $0,08$ $\boxed{C}$ $0,06$ $\boxed{D}$ $0,09$

Question 16 La moyenne marginale de Y (arrondie au dixième près) est

$\boxed{A}$ $1,9$ $\boxed{B}$ autre réponse $\boxed{C}$ $1,8$ $\boxed{D}$ $1,7$

Université Nangui Abrogoua
UFR SFA

Année Universitaire 2015/2016

B

Épreuve de la première session de l'examen de

L'UE DE STATISTIQUE DESCRIPTIVE
ECUE 1 : SERIES STATISTIQUES A UNE VARIABLE
Licence 1 : Maths - Info
Durée : 01 h 30 mn

Consigne : *Le sujet comporte trois (3) pages et contient vingt (20) questions. Pour chacune des questions indiquer, sans justification, la bonne réponse sur la grille réponse.*

Barème : Une réponse exacte rapporte 1 point. Une réponse fausse enlève 1 point. L'absence de réponse ne rapporte ni n'enlève aucun point.

Exercice 1 : Questions à choix multiple

Soit la série statistique définie par le tableau suivant :

Classe d'âges (en années)	[30, 40[	[40, 50[	[50, 60[	[60, 70[	[70, 80[	[80, 90[	[90, 100[
Effectifs	10	10	50	60	40	25	5

Question 1 Le pourcentage des individus qui ont plus de 60 ans est

A autre réponse B 80% C 65% D 35%

Question 2 Le nombre des individus qui ont au moins 80 ans est

A autre réponse B 195 C 170 D 25

Question 3 Le mode de cette série statistique (au dixième près) est

A 63,3 B 65,3 C autre réponse D 60,5

Question 4 Le premier décile de cette série est

A 40 B 55 C 50 D autre réponse

Question 5 La moyenne arithmétique de cette série est

A $68,25$ B autre réponse C $75,15$ D $53,35$

Exercice 2 : Questions à choix multiple

Question 6 Une distribution statistique a un moment centré d'ordre 3 négatif. Sa représentation graphique est :

A symétrique B étalée vers la droite C étalée vers la gauche

Question 7 Mot qui désigne les différents états que peut prendre une valeur observée sur un individu :

A variable B modalité C caractère D observation

Question 8 La fonction de répartition d'une variable X est la fonction F définie par $F(x)$ égale à

A la proportion d'individus dont la valeur du caractère est strictement inférieur à $\overline{X}$
B la proportion d'individus dont la valeur du caractère est strictement supérieur à x
C la proportion d'individus dont la valeur du caractère est strictement supérieur à $\overline{X}$
D la proportion d'individus dont la valeur du caractère est strictement inférieur à x

Question 9 L'âge de la mère au moment de la naissance a été recueilli sur un échantillon de 100 femmes. Les mesures obtenues sont données dans le tableau ci-dessous :

Age de la femme (en années)	< 20	$20-24$	$25-29$	$30-34$	$35-39$	≥ 40
Effectifs	1	11	33	33	17	5

A La distribution de l'âge de la mère à la naissance est bimodale.
B Toutes les classes données ont la même amplitude.
C L'étendue de cette série statistique est 4.
D Impossible de calculer la moyenne à partir des seules données fournies.

Question 10 L'étude d'un caractère quantitatif X fournit les résultats suivants :

$$m_2(X) = \frac{1}{N}\sum_i n_i x_i^2 = 1985 \quad \text{et} \quad m_1(X) = \frac{1}{N}\sum_i n_i x_i = 42.$$

Alors l'écart-type de X vaut

A autre réponse B 221 C $44,1$ D $14,9$

Question 11 Mot ou expression qui désigne la moyenne arithmétique des carrés des écarts à la moyenne :

A écart absolu moyen par rapport à la moyenne B variance C dispersion D écart-type

Question 12 On étudie le nombre d'enfants à charge dans un échantillon de 10 adultes. Voici le tableau observé.

Nombre d'enfants	0	1	2	3	4	5
Nombre d'adultes	2	1	2	3	1	1

. A La médiane est égale à 5 B La moyenne du nombre d'enfants est $2,3$
C L'étendue est 4 D Le mode de cette série est 5

Question 13 Dans une série statistique discrète, le mot qui désigne la valeur du caractère le plus fréquent :

A médiane B amplitude C mode D fréquence

Question 14 Si la représentation graphique d'une distribution est étalée vers la droite, on a :

A $\overline{X} < Me < Mo$ B $Mo < Me < \overline{X}$ C $Mo < \overline{X} < Me$

B

Exercice 3 : Questions à choix direct

Question 15 L'intervalle interquartile correspond à l'intervalle entre le 1^{er} et le 4^{e} quartile.

A Faux   B Vrai

Question 16 Le coefficient de variation est un indicateur de position pertinent pour comparer la variabilité des phénomêmes mesurés avec différentes unités..

A Vrai   B Faux

Question 17 L'étendue d'une distribution statistique correspond à l'intervalle entre la plus petite valeur observée et la plus grande valeur observée.

A Faux   B Vrai

Question 18 Les variables qualitatives peuvent être discrètes ou continues.

A Faux   B Vrai

Question 19 Le coefficient d'asymétrie de Yule est donnée par $C_Y = \dfrac{(Q_2 - Q_1) + (Q_2 - Q_3)}{Q_1 - Q3}$.

A Vrai   B Faux

Question 20 Un histogramme est un ensemble de rectangles contigus dont chaque rectangle, associé à une classe, a une hauteur proportionnelle à l'effectif de cette classe.

A Vrai   B Faux

B

Université Nangui Abrogoua **Année Universitaire 2015/2016**
UFR SFA

Épreuve de la deuxième session de l'examen de

L'UE DE STATISTIQUE DESCRIPTIVE
ECUE 1 : SERIES STATISTIQUES A UNE VARIABLE
Licence 1 : Maths - Info
Durée : 01 h 30 mn

Consigne : *Le sujet comporte trois (3) pages et contient vingt (20) questions. Pour chacune des questions indiquer, sans justification, la bonne réponse sur la grille réponse.*

Barème : Une réponse exacte rapporte 1 point. Une réponse fausse enlève 1 point. L'absence de réponse ne rapporte ni n'enlève aucun point.

Exercice 1 : Questions à choix multiple

Question 1 Le mode de la série statistique définie par le tableau suivant

Classe d'âges (en années)	[30, 40[	[40, 50[	[50, 65[	[65, 70[	[70, 80[	[80, 90[	[90, 100[
Effectifs	10	10	60	60	40	25	5

est

A [65, 70[B 67, 5 C 65 D [50, 65[

Question 2 On désigne par : $e_{\overline{X}}$ l'écart absolu moyen par rapport à la moyenne , e_{M_e} l'écart absolu moyen par rapport à la médiane et par σ l'écart-type d'une variable X. Alors on a

A $e_{M_e} \leq e_{\overline{X}} \leq \sigma$ B $e_{\overline{X}} \leq e_{M_e} \leq \sigma$ C $e_{\overline{X}} \leq \sigma \leq e_{M_e}$ D $\sigma \leq e_{M_e} \leq e_{\overline{X}}$

Question 3 On dispose d'une variable X dont la distribution statistique est la suivante :

x_i	a	b	9	11
n_i	2	1	1	1

où a et b sont inconnues.
Sachant que $\overline{X} = 6,2$ et $Var(X) = 10,56$ on a

A $a = \dfrac{7}{3}$ B $a = 3$ C $b = \dfrac{13}{3}$ D $a = 5$

Question 4 On considère une variable X de moyenne $\overline{X}$ et d'écart-type σ. On définit une nouvelle variable Y en posant

$$Y = \frac{X - \overline{X}}{\sigma}.$$

On affirme alors que Y est une variable

A centrée et réduite B centrée et non réduite C non centrée et réduite
D non centrée et non réduite

Question 5 L'étude d'un caractère quantitatif X fournit les résultats suivants :

$$\sum_i n_i x_i^2 = 29775 \qquad \overline{X} = 42 \quad \text{et} \quad Var(X) = 221.$$

Alors la taille de la population étudiée est

[A] 14 [B] 25 [C] 20 [D] autre réponse

Exercice 2 : Questions à choix multiple

On a relevé dans le tableau suivant les buts marqués par Didier Drogba lors de 50 matchs disputés dans l'équipe Chelsea entre 2010 et 2012.

Nombre de buts par match	0	1	2	3
Fréquence (en pourcentage)	48	38	12	2

Question 6 Le neuvième décile de cette série est

[A] autre réponse [B] 2 [C] 1 [D] 75

Question 7 Dans cette série, l'effectif correspondant à la valeur 2 est

[A] 3 [B] 12 [C] 6 [D] autre réponse

Question 8 La moyenne de cette série est

[A] $1,36$ [B] $0,5$ [C] $0,68$ [D] 1

Question 9 L'effectif cumulé des matchs durant lesquels Didier Drogba a marqué au moins deux buts est

[A] 43 [B] autre réponse [C] 14 [D] 7

Question 10 Le caractère étudié est

[A] le nombre de buts par match
[B] le nombre de matchs disputé par Chelsea
[C] tous les matchs joués par Didier Drogba
[D] le pourcentage de matchs

Exercice 3 : Questions à choix direct

Question 11 L'intervalle entre le 1^{er} et le 3^e quartile est l'intervalle interquantile.

[A] Vrai [B] Faux

Question 12 Le moment non centré d'ordre 1 est un paramètre de position.

[A] Faux [B] Vrai

Question 13 Le coefficient de variation CV défini par $CV = \frac{\overline{X}}{\sigma}$ est un indicateur de dispersion sans unité.

$\boxed{A}$ Vrai $\boxed{B}$ Faux

Question 14 Une variable qualitative peut être représentée par un diagramme à bandes.

$\boxed{A}$ vrai $\boxed{B}$ Faux

Question 15 Pour toute variable quantitative, la moyenne arithmétique est toujours inférieure ou égale à la moyenne quadratique.

$\boxed{A}$ Faux $\boxed{B}$ Vrai

Question 16 L'étendue d'une distribution statistique correspond à la différence entre la plus petite valeur observée et la plus grande valeur observée.

$\boxed{A}$ Faux $\boxed{B}$ Vrai

Question 17 Dans toute série statistique dont la variable est quantitative, le cinquième décile est la médiane.

$\boxed{A}$ Vrai $\boxed{B}$ Faux

Question 18 La variance d'une variable statistique s'exprime dans la même unité que la moyenne arithmétique de cette même variable.

$\boxed{A}$ Faux $\boxed{B}$ Vrai

Question 19 La classe modale d'une variable quantitative continue est celle qui a le plus grand effectif corrigé.

$\boxed{A}$ Faux $\boxed{B}$ Vrai

Question 20 Un histogramme est un ensemble de rectangles continus dont chaque rectangle, associé à une classe, a une hauteur proportionnelle à l'effectif de cette classe.

$\boxed{A}$ Vrai $\boxed{B}$ Faux

Université Nangui Abrogoua | **Année Universitaire 2015/2016**
UFR SFA

Épreuve de la deuxième session de l'examen de

L'UE DE STATISTIQUE DESCRIPTIVE
ECUE 2 : SERIES STATISTIQUES BIVARIÉES Licence 1 : Maths - Info
Durée : 01 h 30 mn

Consigne : *Le sujet comporte trois (3) pages et contient vingt (20) questions. Pour chacune des questions indiquer, sans justification, la bonne réponse sur la grille réponse.*

Barême : Une réponse exacte rapporte 1 point. Une réponse fausse enlève 1 point. L'absence de réponse ne rapporte ni n'enlève aucun point.

Exercice 1 : Questions à choix direct

Question 1 Si le coefficient de corrélation linéaire vaut -1, alors les droites de régression de X en Y et de Y en X sont confondues.

$\boxed{\text{A}}$ Vrai $\boxed{\text{B}}$ Faux

Question 2 Si X et Y sont deux variables statistiquement indépendantes, alors les distributions marginales de X et de Y suffisent pour déterminer la distribution du couple (X, Y).

$\boxed{\text{A}}$ Vrai $\boxed{\text{B}}$ Faux

Question 3 Si les deux droites d'ajustement (X en Y et Y en X) ont leurs coefficients directeurs négatifs, alors le coefficient de corrélation linéaire est

$\boxed{\text{A}}$ positif $\boxed{\text{B}}$ negatif

Question 4 Le mot qui renseigne sur l'intensité de la liaison entre deux variables X et Y est

$\boxed{\text{A}}$ regression $\boxed{\text{B}}$ corrélation

Question 5 La covariance de (X, Y), lorsqu'elle est non nulle, est un indicateur du sens de variation simultanéé des variables X et Y.

$\boxed{\text{A}}$ Faux $\boxed{\text{B}}$ Vrai

Question 6 Si $Cov(X, Y) = 0$ alors X et Y sont linéairement indépendantes.

$\boxed{\text{A}}$ Faux $\boxed{\text{B}}$ Vrai

Exercice 2 : Questions à choix multiple

La population d'un quartier d'Abidjan est décrite par l'âge (X) et le nombre de numéros de téléphone possédés (Y).

$X \setminus Y$	0	1	2	3
[15 , 25[	8	8	5	2
[25 , 35[	7	20	22	20
[35 , 50[	11	34	40	37
[50 , 65[	15	37	39	35
[65 , 95[	9	31	33	12

Question 7 La moyenne marginale de Y (arrondie au dixième près) est

A $1,8$ B $1,7$ C $1,9$ D autre réponse

Question 8 La fréquence f_{23} (arrondie au centième près) est égale à

A $0,08$ B autre réponse C $0,05$ D $0,07$

Question 9 La fréquence conditionnelle $X \in [15\,,\,25[$ sachant $Y = 2$ (arrodie au millième près) est

A $0,035$ B $0,036$ C $0,038$ D autre réponse

Question 10 La formule permettant de calculer $f_{2|1}$ est

A $\frac{n_{12}}{n_{\bullet 1}}$ B $\frac{n_{12}}{n_{1\bullet}}$ C $\frac{n_{12}}{n_{2\bullet}}$ D $\frac{n_{21}}{n_{1\bullet}}$

Question 11 Le nombre d'habitants de ce quartier est

A 450 B autre réponse C 400 D 425

Question 12 L'âge moyen de cette population est

A autre réponse B $51,3$ C $51,2$ D $51,1$

Question 13 La fréquence marginale $f_{\bullet 2}$ (arrondie au dixième près) est

A $0,1$ B $0,3$ C autre réponse D $0,2$

Question 14 La fréquence marginale $f_{3\bullet}$ (arrondie au dixième près) est

A $0,2$ B $0,4$ C autre réponse D $0,3$

Exercice 3 : Questions à choix multiple

Soit la série statistique double définie par le tableau suivant :

i	1	2	3	4	5
x_i	10	12	18	24	36
y_i	50	60	70	90	130

Question 15 La variance de Y est

A 560 B 488 C 800 D autre réponse

Question 16 La corrélation entre (X, Y) est

$\boxed{\text{A}}$ très forte $\boxed{\text{B}}$ faible $\boxed{\text{C}}$ très faible

Question 17 La covariance de (X, Y) est

$\boxed{\text{A}}$ 356 $\boxed{\text{B}}$ autre réponse $\boxed{\text{C}}$ 182 $\boxed{\text{D}}$ 264

Question 18 La série statistique étudiée peut être représentée par un diagramme en bâtons.

$\boxed{\text{A}}$ Faux $\boxed{\text{B}}$ Vrai

Question 19 La droite d'ajustement de Y en X a pour équation

$\boxed{\text{A}}$ $y = 3x + 20$ $\boxed{\text{B}}$ $y = -3x + 20$ $\boxed{\text{C}}$ autre réponse $\boxed{\text{D}}$ $y = 3x - 20$

Question 20 Le tableau statistique ci-dessus est un tableau de contingence.

$\boxed{\text{A}}$ Vrai $\boxed{\text{B}}$ Faux

Université Nangui Abrogoua **Année universitaire** 2022-2023
UFR SFA

A

Épreuve de la deuxième session de l'examen de

L'UE DE STATISTIQUE DESCRIPTIVE
ECUE 1 : SERIES STATISTIQUES UNIVARIÉES
Licence 1 : Maths - Info
Durée : 01 h 00 mn

Consigne : *Le sujet est composé de dix (10) questions réparties en trois (3) exercices indépendants. Pour chacune des questions indiquer, sans justification, la seule bonne réponse sur la grille réponse.*

Barème : Une réponse exacte rapporte 2 points, une réponse fausse enlève 1 point et l'absence de réponse ne rapporte ni n'enlève aucun point.

Exercice 1

On considère le tableau statistique suivant :

x_i	80	a	b	160
n_i	16	13	c	17

On suppose que $\overline{X} = 119$, $Var(X) = 1089$ et que la taille de la population étudiée est 50. Les valeurs a, b et c sont inconnues et on désire les déterminer.

Question 1 Le coefficient de variation de X (arrondi au centième près) est

A 1,26 B 0,28 C 3,61 D Autre réponse

Question 2 La valeur de a est

A Autre réponse B 110 C 90 D 120

Question 3 La valeur de b est

A 145 B Autre réponse C 135 D 150

Exercice 2

Une étude statistique a donné les résultats suivants :

Classe	[30, 40[	[40, 50[	[50, 60[	[60, 70[	[70, 80[	[80, 90[	[90, 100[
Effectif	10	10	50	60	40	25	5

Question 4 La médiane de cette série est

A 60 B 65 C Autre réponse D 70

Question 5 Le mode de la série (arrondi au dixième près) est

A 63,3 B Autre réponse C 65 D 65,3

Question 6 Le premier décile de cette série est

A Autre réponse B 50 C 60 D 40

Question 7 La moyenne arithmétique de la série est

A $65,25$ B $60,25$ C $75,25$ D Autre réponse

Exercice 3

Question 8 Le coefficient d'aplatissement de Yule est donné par la formule

$$C_Y = \frac{(Q_2 - Q_1) + (Q_2 - Q_3)}{Q_1 - Q_3}$$

A Vrai B Faux

Question 9 L'écart inter-décile est la différence entre le dixième décile et le premier décile.

A Vrai B Faux

Question 10 L'étendue d'une variable qualitative est la différence entre la plus grande valeur observée et la plus petite valeur observée.

A Faux B Vrai

A

Université Nangui Abrogoua **Année universitaire** 2022-2023
UFR SFA

C

Épreuve de la deuxième session de l'examen de

L'UE DE STATISTIQUE DESCRIPTIVE
ECUE 1 : SERIES STATISTIQUES UNIVARIÉES
Licence 1 : Maths - Info
Durée : 01 h 00 mn

Consigne : *Le sujet est composé de dix (10) questions réparties en trois (3) exercices indépendants. Pour chacune des questions indiquer, sans justification, la seule bonne réponse sur la grille réponse.*

Barème : Une réponse exacte rapporte 2 points, une réponse fausse enlève 1 point et l'absence de réponse ne rapporte ni n'enlève aucun point.

Exercice 1

On considère le tableau statistique suivant :

x_i	80	a	b	160
n_i	16	13	c	17

On suppose que $\overline{X} = 119$, $Var(X) = 1089$ et que la taille de la population étudiée est 50. Les valeurs a, b et c sont inconnues et on désire les déterminer.

Question 1 Le coefficient de variation de X (arrondi au centième près) est

A Autre réponse B 0,28 C 1,26 D 3,61

Question 2 La valeur de a est

A 120 B 110 C 90 D Autre réponse

Question 3 La valeur de b est

A 135 B 145 C 150 D Autre réponse

Exercice 2

Une étude statistique a donné les résultats suivants :

Classe	[30, 40[	[40, 50[	[50, 60[	[60, 70[	[70, 80[	[80, 90[	[90, 100[
Effectif	10	10	50	60	40	25	5

Question 4 Le mode de la série (arrondi au dixième près) est

A 63,3 B Autre réponse C 65,3 D 65

Question 5 Le premier décile de cette série est

A 40 B 60 C Autre réponse D 50

Question 6 La moyenne arithmétique de la série est

A $75,25$ B Autre réponse C $60,25$ D $65,25$

Question 7 La médiane de cette série est

A 70 B 60 C 65 D Autre réponse

Exercice 3

Question 8 Le coefficient d'aplatissement de Yule est donné par la formule

$$C_Y = \frac{(Q_2 - Q_1) + (Q_2 - Q_3)}{Q_1 - Q_3}$$

$\boxed{\text{A}}$ Faux $\boxed{\text{B}}$ Vrai

Question 9 L'écart inter-décile est la différence entre le dixième décile et le premier décile.

$\boxed{\text{A}}$ Vrai $\boxed{\text{B}}$ Faux

Question 10 L'étendue d'une variable qualitative est la différence entre la plus grande valeur observée et la plus petite valeur observée.

$\boxed{\text{A}}$ Vrai $\boxed{\text{B}}$ Faux

C

Université Nangui Abrogoua **Année universitaire :** 2022-2023

UFR SFA

B

Épreuve de la deuxième session de l'examen de

L'UE DE STATISTIQUE DESCRIPTIVE
ECUE 1 : SERIES STATISTIQUES UNIVARIÉES
Licence 1 : Maths - Info
Durée : 01 h 00 mn

Consigne : *Le sujet est composé de dix (10) questions réparties en trois (3) exercices indépendants. Pour chacune des questions indiquer, sans justification, la seule bonne réponse sur la grille réponse.*

Barème : Une réponse exacte rapporte 2 points, une réponse fausse enlève 1 point et l'absence de réponse ne rapporte ni n'enlève aucun point.

Exercice 1

On considère le tableau statistique suivant :

x_i	80	a	b	160
n_i	16	13	c	17

On suppose que $\overline{X} = 119$, $Var(X) = 1089$ et que la taille de la population étudiée est 50. Les valeurs a, b et c sont inconnues et on désire les déterminer.

Question 1 Le coefficient de variation de X (arrondi au centième près) est

A 1,26 B 0,28 C 3,61 D Autre réponse

Question 2 La valeur de a est

A Autre réponse B 110 C 90 D 120

Question 3 La valeur de b est

A 145 B Autre réponse C 135 D 150

Exercice 2

Question 4 Le coefficient d'aplatissement de Yule est donné par la formule

$$C_Y = \frac{(Q_2 - Q_1) + (Q_2 - Q_3)}{Q_1 - Q_3}$$

A Faux B Vrai

Question 5 L'écart inter-décile est la différence entre le dixième décile et le premier décile.

A Vrai B Faux

Question 6 L'étendue d'une variable qualitative est la différence entre la plus grande valeur observée et la plus petite valeur observée.

A Faux B Vrai

Exercice 3

Une étude statistique a donné les résultats suivants :

Classe	[30, 40[	[40, 50[	[50, 60[	[60, 70[	[70, 80[	[80, 90[	[90, 100[
Effectif	10	10	50	60	40	25	5

Question 7 La moyenne arithmétique de la série est

A $75,25$ B Autre réponse C $60,25$ D $65,25$

Question 8 Le mode de la série (arrondi au dixième près) est

A 65 B 65,3 C 63,3 D Autre réponse

Question 9 Le premier décile de cette série est

A 50 B Autre réponse C 60 D 40

Question 10 La médiane de cette série est

A Autre réponse B 65 C 70 D 60

B

Université Nangui Abrogoua **Année universitaire** 2022-2023
UFR SFA

D

Épreuve de la deuxième session de l'examen de

L'UE DE STATISTIQUE DESCRIPTIVE
ECUE 1 : SERIES STATISTIQUES UNIVARIÉES
Licence 1 : Maths - Info
Durée : 01 h 00 mn

Consigne : *Le sujet est composé de dix (10) questions réparties en trois (3) exercices indépendants. Pour chacune des questions indiquer, sans justification, la seule bonne réponse sur la grille réponse.*

Barème : Une réponse exacte rapporte 2 points, une réponse fausse enlève 1 point et l'absence de réponse ne rapporte ni n'enlève aucun point.

Exercice 1

On considère le tableau statistique suivant :

x_i	80	a	b	160
n_i	16	13	c	17

On suppose que $\overline{X} = 119$, $Var(X) = 1089$ et que la taille de la population étudiée est 50. Les valeurs a, b et c sont inconnues et on désire les déterminer.

Question 1 Le coefficient de variation de X (arrondi au centième près) est

A Autre réponse B 0,28 C 1,26 D 3,61

Question 2 La valeur de a est

A 120 B 110 C 90 D Autre réponse

Question 3 La valeur de b est

A 135 B 145 C 150 D Autre réponse

Exercice 2

Question 4 Le coefficient d'aplatissement de Yule est donné par la formule

$$C_Y = \frac{(Q_2 - Q_1) + (Q_2 - Q_3)}{Q_1 - Q_3}$$

A Vrai B Faux

Question 5 L'écart inter-décile est la différence entre le dixième décile et le premier décile.

A Faux B Vrai

Question 6 L'étendue d'une variable qualitative est la différence entre la plus grande valeur observée et la plus petite valeur observée.

A Vrai B Faux

Exercice 3

Une étude statistique a donné les résultats suivants :

Classe	[30, 40[	[40, 50[	[50, 60[	[60, 70[	[70, 80[	[80, 90[	[90, 100[
Effectif	10	10	50	60	40	25	5

Question 7 Le premier décile de cette série est

A 40 B 60 C Autre réponse D 50

Question 8 La médiane de cette série est

A 65 B 60 C Autre réponse D 70

Question 9 La moyenne arithmétique de la série est

A $65,25$ B Autre réponse C $75,25$ D $60,25$

Question 10 Le mode de la série (arrondi au dixième près) est

A 63,3 B 65,3 C Autre réponse D 65

D

Université Nangui Abrogoua **Année universitaire 2022/2023**
UFR SFA

A

Épreuve de la deuxième session de l'examen de

L'UE DE STATISTIQUE DESCRIPTIVE
ECUE 2 : SERIES STATISTIQUES BIVARIÉES Licence 1 : Maths - Info
Durée : 01 h 00 mn

Consigne : *Le sujet est composé de dix (10) questions réparties en deux (2) exercices indépendants. Pour chacune des questions indiquer, sans justification, la seule bonne réponse sur la grille réponse.*

Barème : Une réponse exacte rapporte 2 points. Une réponse fausse enlève 1 point. L'absence de réponse ne rapporte ni n'enlève aucun point.

Exercice 1

Soit la série statistique double donnée par le tableau de contingence suivant :

$X \setminus Y$	$[0,\ 3[$	$[3,\ 9[$
$[0,\ 4[$	2	4
$[4,\ 12[$	8	3

Question 1 La variance conditionnelle de X sachant $Y \in [0,\ 3[$ est égale à

A $7,32$ B $5,76$ C $10,25$ D autre réponse

Question 2 La formule permettant de calculer $f_{1|2}$ est

A $\frac{n_{12}}{n_{\bullet 1}}$ B $\frac{n_{12}}{n_{\bullet 2}}$ C $\frac{n_{12}}{n_{2\bullet}}$ D $\frac{n_{21}}{n_{1\bullet}}$

Question 3 La moyenne marginale de X (arrondie au centième près) est égale à

A 9,09 B 5,88 C autre réponse D 3,35

Question 4 La fréquence conditionnelle de $X \in [0,\ 4[$ sachant $Y \in [3,\ 9[$ (arrondie au centième près) est

A $0,57$ B $0,67$ C $0,43$ D autre réponse

Question 5 La moyenne conditionnelle de Y sachant $X \in [4,\ 12[$ est égale à

A 1,76 B 2,73 C autre réponse D 5

Exercice 2

On considère la distibution statistique double suivante :

$X \setminus Y$	7	1
3	1	2
5	3	2
7	1	2

Question 6 Les variables X et Y sont statistiquement indépendantes.

$\boxed{\text{A}}$ Faux $\boxed{\text{B}}$ Vrai

Question 7 La moyenne marginale de X est

$\boxed{\text{A}}$ $3,73$ $\boxed{\text{B}}$ 4 $\boxed{\text{C}}$ autre réponse $\boxed{\text{D}}$ 5

Question 8 L'écart-type de Y (arrondi au millième près) est égal à

$\boxed{\text{A}}$ $1,477$ $\boxed{\text{B}}$ $3,967$ $\boxed{\text{C}}$ $2,987$ $\boxed{\text{D}}$ autre réponse

Question 9 Les variables X et Y sont linéairement indépendantes.

$\boxed{\text{A}}$ Vrai $\boxed{\text{B}}$ Faux

Question 10 La covariance du couple (X, Y) (arrondie au centième près) est égale à

$\boxed{\text{A}}$ 0 $\boxed{\text{B}}$ $0,02$ $\boxed{\text{C}}$ autre réponse $\boxed{\text{D}}$ $0,01$

$\boxed{\textbf{A}}$

Université Nangui Abrogoua **Année universitaire 2022/2023**
UFR SFA

C

Épreuve de la deuxième session de l'examen de

L'UE DE STATISTIQUE DESCRIPTIVE
ECUE 2 : SERIES STATISTIQUES BIVARIÉES Licence 1 : Maths - Info
Durée : 01 h 00 mn

Consigne : *Le sujet est composé de dix (10) questions réparties en deux (2) exercices indépendants. Pour chacune des questions indiquer, sans justification, la seule bonne réponse sur la grille réponse.*

Barème : Une réponse exacte rapporte 2 points. Une réponse fausse enlève 1 point. L'absence de réponse ne rapporte ni n'enlève aucun point.

Exercice 1

Soit la série statistique double donnée par le tableau de contingence suivant :

$X \setminus Y$	$[0,\ 3[$	$[3,\ 9[$
$[0,\ 4[$	2	4
$[4,\ 12[$	8	3

Question 1 La fréquence conditionnelle de $X \in [0,\ 4[$ sachant $Y \in [3,\ 9[$ (arrondie au centième près) est

A $0,67$ B autre réponse C $0,43$ D $0,57$

Question 2 La variance conditionnelle de X sachant $Y \in [0,\ 3[$ est égale à

A $7,32$ B $10,25$ C $5,76$ D autre réponse

Question 3 La moyenne conditionnelle de Y sachant $X \in [4,\ 12[$ est égale à

A 2,73 B 1,76 C 5 D autre réponse

Question 4 La formule permettant de calculer $f_{1|2}$ est

A $\frac{n_{12}}{n_{2\bullet}}$ B $\frac{n_{12}}{n_{\bullet 1}}$ C $\frac{n_{12}}{n_{\bullet 2}}$ D $\frac{n_{21}}{n_{1\bullet}}$

Question 5 La moyenne marginale de X (arrondie au centième près) est égale à

A 5,88 B 9,09 C autre réponse D 3,35

Exercice 2

On considère la distibution statistique double suivante :

$X \setminus Y$	7	1
3	1	2
5	3	2
7	1	2

Question 6 La moyenne marginale de X est

$\boxed{A}$ 4 $\boxed{B}$ $3,73$ $\boxed{C}$ 5 $\boxed{D}$ autre réponse

Question 7 Les variables X et Y sont statistiquement indépendantes.

$\boxed{A}$ Faux $\boxed{B}$ Vrai

Question 8 L'écart-type de Y (arrondi au millième près) est égal à

$\boxed{A}$ $3,967$ $\boxed{B}$ $1,477$ $\boxed{C}$ $2,987$ $\boxed{D}$ autre réponse

Question 9 La covariance du couple (X, Y) (arrondie au centième près) est égale à

$\boxed{A}$ $0,01$ $\boxed{B}$ 0 $\boxed{C}$ $0,02$ $\boxed{D}$ autre réponse

Question 10 Les variables X et Y sont linéairement indépendantes.

$\boxed{A}$ Vrai $\boxed{B}$ Faux

C

B

Épreuve de la deuxième session de l'examen de

L'UE DE STATISTIQUE DESCRIPTIVE ECUE 2 : SERIES STATISTIQUES BIVARIÉES Licence 1 : Maths - Info
Durée : 01 h 00 mn

Consigne : *Le sujet est composé de dix (10) questions réparties en deux (2) exercices indépendants. Pour chacune des questions indiquer, sans justification, la seule bonne réponse sur la grille réponse.*

Barème : Une réponse exacte rapporte 2 points. Une réponse fausse enlève 1 point. L'absence de réponse ne rapporte ni n'enlève aucun point.

Exercice 1

On considère la distibution statistique double suivante :

$X \setminus Y$	7	1
3	1	2
5	3	2
7	1	2

Question 1 La moyenne marginale de X est

A 5 B 4 C $3,73$ D autre réponse

Question 2 L'écart-type de Y (arrondi au millième près) est égal à

A $2,987$ B autre réponse C $3,967$ D $1,477$

Question 3 Les variables X et Y sont statistiquement indépendantes.

A Vrai B Faux

Question 4 La covariance du couple (X, Y) (arrondie au centième près) est égale à

A autre réponse B $0,01$ C 0 D $0,02$

Question 5 Les variables X et Y sont linéairement indépendantes.

A Faux B Vrai

Exercice 2

Soit la série statistique double donnée par le tableau de contingence suivant :

$X \setminus Y$	$[0,\ 3[$	$[3,\ 9[$
$[0,\ 4[$	2	4
$[4,\ 12[$	8	3

Question 6 La fréquence conditionnelle de $X \in [0,\, 4[$ sachant $Y \in [3,\, 9[$ (arrondie au centième près) est

$\boxed{\text{A}}$ $0,43$ $\boxed{\text{B}}$ $0,57$ $\boxed{\text{C}}$ $0,67$ $\boxed{\text{D}}$ autre réponse

Question 7 La moyenne conditionnelle de Y sachant $X \in [4,\, 12[$ est égale à

$\boxed{\text{A}}$ 2,73 $\boxed{\text{B}}$ 1,76 $\boxed{\text{C}}$ autre réponse $\boxed{\text{D}}$ 5

Question 8 La formule permettant de calculer $f_{1|2}$ est

$\boxed{\text{A}}$ $\frac{n_{12}}{n_{\bullet 2}}$ $\boxed{\text{B}}$ $\frac{n_{12}}{n_{\bullet 1}}$ $\boxed{\text{C}}$ $\frac{n_{21}}{n_{1\bullet}}$ $\boxed{\text{D}}$ $\frac{n_{12}}{n_{2\bullet}}$

Question 9 La variance conditionnelle de X sachant $Y \in [0,\, 3[$ est égale à

$\boxed{\text{A}}$ $7,32$ $\boxed{\text{B}}$ $5,76$ $\boxed{\text{C}}$ $10,25$ $\boxed{\text{D}}$ autre réponse

Question 10 La moyenne marginale de X (arrondie au centième près) est égale à

$\boxed{\text{A}}$ 3,35 $\boxed{\text{B}}$ 9,09 $\boxed{\text{C}}$ autre réponse $\boxed{\text{D}}$ 5,88

B

Université Nangui Abrogoua **Année universitaire 2022/2023**
UFR SFA

D

Épreuve de la deuxième session de l'examen de

L'UE DE STATISTIQUE DESCRIPTIVE
ECUE 2 : SERIES STATISTIQUES BIVARIÉES Licence 1 : Maths - Info
Durée : 01 h 00 mn

Consigne : *Le sujet est composé de dix (10) questions réparties en deux (2) exercices indépendants. Pour chacune des questions indiquer, sans justification, la seule bonne réponse sur la grille réponse.*

Barème : Une réponse exacte rapporte 2 points. Une réponse fausse enlève 1 point. L'absence de réponse ne rapporte ni n'enlève aucun point.

Exercice 1

On considère la distibution statistique double suivante :

$X \setminus Y$	7	1
3	1	2
5	3	2
7	1	2

Question 1 La covariance du couple (X, Y) (arrondie au centième près) est égale à

A $0,02$ B autre réponse C 0 D $0,01$

Question 2 La moyenne marginale de X est

A 5 B $3,73$ C 4 D autre réponse

Question 3 Les variables X et Y sont linéairement indépendantes.

A Vrai B Faux

Question 4 L'écart-type de Y (arrondi au millième près) est égal à

A $2,987$ B autre réponse C $1,477$ D $3,967$

Question 5 Les variables X et Y sont statistiquement indépendantes.

A Faux B Vrai

Exercice 2

Soit la série statistique double donnée par le tableau de contingence suivant :

$X \setminus Y$	$[0,\ 3[$	$[3,\ 9[$
$[0,\ 4[$	2	4
$[4,\ 12[$	8	3

Question 6 La moyenne marginale de X (arrondie au centième près) est égale à

A 9,09 B 3,35 C 5,88 D autre réponse

Question 7 La variance conditionnelle de X sachant $Y \in [0, 3[$ est égale à

A $5,76$ B $7,32$ C autre réponse D $10,25$

Question 8 La fréquence conditionnelle de $X \in [0, 4[$ sachant $Y \in [3, 9[$ (arrondie au centième près) est

A autre réponse B $0,67$ C $0,43$ D $0,57$

Question 9 La moyenne conditionnelle de Y sachant $X \in [4, 12[$ est égale à

A 2,73 B 5 C 1,76 D autre réponse

Question 10 La formule permettant de calculer $f_{1|2}$ est

A $\frac{n_{12}}{n_{2\bullet}}$ B $\frac{n_{12}}{n_{\bullet 1}}$ C $\frac{n_{12}}{n_{\bullet 2}}$ D $\frac{n_{21}}{n_{1\bullet}}$

D

Université Nangui Abrogoua **Année universitaire** 2020-2021
UFR SFA

A

Épreuve de la deuxième session de l'examen de

L'UE DE STATISTIQUE DESCRIPTIVE
ECUE 1 : SERIES STATISTIQUES UNIVARIÉES
Licence 1 : Maths - Info
Durée : 01 h 00 mn

Consigne : *Le sujet est composé de dix (10) questions réparties en trois (3) exercices indépendants. Pour chacune des questions indiquer, sans justification, la seule bonne réponse sur la grille réponse.*

Barème : Une réponse exacte rapporte 2 points, une réponse fausse enlève 1 point et l'absence de réponse ne rapporte ni n'enlève aucun point.

Exercice 1

On considère le tableau statistique suivant :

x_i	80	a	b	160
n_i	16	13	c	17

On suppose que $\overline{X} = 119$, $Var(X) = 1089$ et que la taille de la population étudiée est 50. Les valeurs a, b et c sont inconnues et on désire les déterminer.

Question 1 Le coefficient de variation de X (arrondi au centième près) est

A 1,26 B 0,28 C 3,61 D Autre réponse

Question 2 La valeur de a est

A Autre réponse B 110 C 90 D 120

Question 3 La valeur de b est

A 145 B Autre réponse C 135 D 150

Exercice 2

Une étude statistique a donné les résultats suivants :

Classe	[30, 40[	[40, 50[	[50, 60[	[60, 70[	[70, 80[	[80, 90[	[90, 100[
Effectif	10	10	50	60	40	25	5

Question 4 La médiane de cette série est

A 60 B 65 C Autre réponse D 70

Question 5 Le mode de la série (arrondi au dixième près) est

A 63,3 B Autre réponse C 65 D 65,3

Question 6 Le premier décile de cette série est

A Autre réponse B 50 C 60 D 40

Question 7 La moyenne arithmétique de la série est

A $65,25$ B $60,25$ C $75,25$ D Autre réponse

Exercice 3

Question 8 Le coefficient d'aplatissement de Yule est donné par la formule

$$C_Y = \frac{(Q_2 - Q_1) + (Q_2 - Q_3)}{Q_1 - Q_3}$$

$\boxed{A}$ Vrai $\boxed{B}$ Faux

Question 9 L'écart inter-décile est la différence entre le dixième décile et le premier décile.

$\boxed{A}$ Vrai $\boxed{B}$ Faux

Question 10 L'étendue d'une variable qualitative est la différence entre la plus grande valeur observée et la plus petite valeur observée.

$\boxed{A}$ Faux $\boxed{B}$ Vrai

A

Université Nangui Abrogoua
UFR SFA

Année universitaire 2020-2021

C

Épreuve de la deuxième session de l'examen de

L'UE DE STATISTIQUE DESCRIPTIVE
ECUE 1 : SERIES STATISTIQUES UNIVARIÉES
Licence 1 : Maths - Info
Durée : 01 h 00 mn

Consigne : *Le sujet est composé de dix (10) questions réparties en trois (3) exercices indépendants. Pour chacune des questions indiquer, sans justification, la seule bonne réponse sur la grille réponse.*

Barème : Une réponse exacte rapporte 2 points, une réponse fausse enlève 1 point et l'absence de réponse ne rapporte ni n'enlève aucun point.

Exercice 1

On considère le tableau statistique suivant :

x_i	80	a	b	160
n_i	16	13	c	17

On suppose que $\overline{X} = 119$, $Var(X) = 1089$ et que la taille de la population étudiée est 50. Les valeurs a, b et c sont inconnues et on désire les déterminer.

Question 1 Le coefficient de variation de X (arrondi au centième près) est

A Autre réponse B 0,28 C 1,26 D 3,61

Question 2 La valeur de a est

A 120 B 110 C 90 D Autre réponse

Question 3 La valeur de b est

A 135 B 145 C 150 D Autre réponse

Exercice 2

Une étude statistique a donné les résultats suivants :

Classe	[30, 40[	[40, 50[	[50, 60[	[60, 70[	[70, 80[	[80, 90[	[90, 100[
Effectif	10	10	50	60	40	25	5

Question 4 Le mode de la série (arrondi au dixième près) est

A 63,3 B Autre réponse C 65,3 D 65

Question 5 Le premier décile de cette série est

A 40 B 60 C Autre réponse D 50

Question 6 La moyenne arithmétique de la série est

A $75,25$ B Autre réponse C $60,25$ D $65,25$

Question 7 La médiane de cette série est

A 70 B 60 C 65 D Autre réponse

Exercice 3

Question 8 Le coefficient d'aplatissement de Yule est donné par la formule

$$C_Y = \frac{(Q_2 - Q_1) + (Q_2 - Q_3)}{Q_1 - Q_3}$$

A Faux B Vrai

Question 9 L'écart inter-décile est la différence entre le dixième décile et le premier décile.

A Vrai B Faux

Question 10 L'étendue d'une variable qualitative est la différence entre la plus grande valeur observée et la plus petite valeur observée.

A Vrai B Faux

C

Université Nangui Abrogoua **Année universitaire :** 2020-2021
UFR SFA

B

Épreuve de la deuxième session de l'examen de

L'UE DE STATISTIQUE DESCRIPTIVE
ECUE 1 : SERIES STATISTIQUES UNIVARIÉES
Licence 1 : Maths - Info
Durée : 01 h 00 mn

Consigne : *Le sujet est composé de dix (10) questions réparties en trois (3) exercices indépendants. Pour chacune des questions indiquer, sans justification, la seule bonne réponse sur la grille réponse.*

Barème : Une réponse exacte rapporte 2 points, une réponse fausse enlève 1 point et l'absence de réponse ne rapporte ni n'enlève aucun point.

Exercice 1

On considère le tableau statistique suivant :

x_i	80	a	b	160
n_i	16	13	c	17

On suppose que $\overline{X} = 119$, $Var(X) = 1089$ et que la taille de la population étudiée est 50. Les valeurs a, b et c sont inconnues et on désire les déterminer.

Question 1 Le coefficient de variation de X (arrondi au centième près) est

[A] 1,26 [B] 0,28 [C] 3,61 [D] Autre réponse

Question 2 La valeur de a est

[A] Autre réponse [B] 110 [C] 90 [D] 120

Question 3 La valeur de b est

[A] 145 [B] Autre réponse [C] 135 [D] 150

Exercice 2

Question 4 Le coefficient d'aplatissement de Yule est donné par la formule

$$C_Y = \frac{(Q_2 - Q_1) + (Q_2 - Q_3)}{Q_1 - Q_3}$$

[A] Faux [B] Vrai

Question 5 L'écart inter-décile est la différence entre le dixième décile et le premier décile.

[A] Vrai [B] Faux

Question 6 L'étendue d'une variable qualitative est la différence entre la plus grande valeur observée et la plus petite valeur observée.

[A] Faux [B] Vrai

Exercice 3

Une étude statistique a donné les résultats suivants :

Classe	[30, 40[	[40, 50[	[50, 60[	[60, 70[	[70, 80[	[80, 90[	[90, 100[
Effectif	10	10	50	60	40	25	5

Question 7 La moyenne arithmétique de la série est

A $75,25$ B Autre réponse C $60,25$ D $65,25$

Question 8 Le mode de la série (arrondi au dixième près) est

A 65 B 65,3 C 63,3 D Autre réponse

Question 9 Le premier décile de cette série est

A 50 B Autre réponse C 60 D 40

Question 10 La médiane de cette série est

A Autre réponse B 65 C 70 D 60

B

Université Nangui Abrogoua **Année universitaire** 2020-2021
UFR SFA

D

Épreuve de la deuxième session de l'examen de

L'UE DE STATISTIQUE DESCRIPTIVE
ECUE 1 : SERIES STATISTIQUES UNIVARIÉES
Licence 1 : Maths - Info
Durée : 01 h 00 mn

Consigne : *Le sujet est composé de dix (10) questions réparties en trois (3) exercices indépendants. Pour chacune des questions indiquer, sans justification, la seule bonne réponse sur la grille réponse.*

Barème : Une réponse exacte rapporte 2 points, une réponse fausse enlève 1 point et l'absence de réponse ne rapporte ni n'enlève aucun point.

Exercice 1

On considère le tableau statistique suivant :

x_i	80	a	b	160
n_i	16	13	c	17

On suppose que $\overline{X} = 119$, $Var(X) = 1089$ et que la taille de la population étudiée est 50. Les valeurs a, b et c sont inconnues et on désire les déterminer.

Question 1 Le coefficient de variation de X (arrondi au centième près) est

A Autre réponse B 0,28 C 1,26 D 3,61

Question 2 La valeur de a est

A 120 B 110 C 90 D Autre réponse

Question 3 La valeur de b est

A 135 B 145 C 150 D Autre réponse

Exercice 2

Question 4 Le coefficient d'aplatissement de Yule est donné par la formule

$$C_Y = \frac{(Q_2 - Q_1) + (Q_2 - Q_3)}{Q_1 - Q_3}$$

A Vrai B Faux

Question 5 L'écart inter-décile est la différence entre le dixième décile et le premier décile.

A Faux B Vrai

Question 6 L'étendue d'une variable qualitative est la différence entre la plus grande valeur observée et la plus petite valeur observée.

A Vrai B Faux

Exercice 3

Une étude statistique a donné les résultats suivants :

Classe	[30, 40[	[40, 50[	[50, 60[	[60, 70[	[70, 80[	[80, 90[	[90, 100[
Effectif	10	10	50	60	40	25	5

Question 7 Le premier décile de cette série est

$\boxed{A}$ 40 $\boxed{B}$ 60 $\boxed{C}$ Autre réponse $\boxed{D}$ 50

Question 8 La médiane de cette série est

$\boxed{A}$ 65 $\boxed{B}$ 60 $\boxed{C}$ Autre réponse $\boxed{D}$ 70

Question 9 La moyenne arithmétique de la série est

$\boxed{A}$ $65,25$ $\boxed{B}$ Autre réponse $\boxed{C}$ $75,25$ $\boxed{D}$ $60,25$

Question 10 Le mode de la série (arrondi au dixième près) est

$\boxed{A}$ 63,3 $\boxed{B}$ 65,3 $\boxed{C}$ Autre réponse $\boxed{D}$ 65

D

Université Nangui Abrogoua **Année universitaire** 2020-2021
UFR SFA

A

Épreuve de la deuxième session de l'examen de

L'UE DE STATISTIQUE DESCRIPTIVE
ECUE 2 : SERIES STATISTIQUES BIVARIÉES Licence 1 : Maths - Info
Durée : 01 h 00 mn

Consigne : *Le sujet est composé de dix (10) questions réparties en deux (2) exercices indépendants. Pour chacune des questions indiquer, sans justification, la seule bonne réponse sur la grille réponse.*

Barème : Une réponse exacte rapporte 2 points. Une réponse fausse enlève 1 point. L'absence de réponse ne rapporte ni n'enlève aucun point.

Exercice 1

Soit la série statistique double donnée par le tableau de contingence suivant :

$X \setminus Y$	$[0,\ 3[$	$[3,\ 9[$
$[0,\ 4[$	2	4
$[4,\ 12[$	8	3

Question 1 La variance conditionnelle de X sachant $Y \in [0,\ 3[$ est égale à

A $7,32$ B $5,76$ C $10,25$ D autre réponse

Question 2 La formule permettant de calculer $f_{1|2}$ est

A $\frac{n_{12}}{n_{\bullet 1}}$ B $\frac{n_{12}}{n_{\bullet 2}}$ C $\frac{n_{12}}{n_{2\bullet}}$ D $\frac{n_{21}}{n_{1\bullet}}$

Question 3 La moyenne marginale de X (arrondie au centième près) est égale à

A 9,09 B 5,88 C autre réponse D 3,35

Question 4 La fréquence conditionnelle de $X \in [0,\ 4[$ sachant $Y \in [3,\ 9[$ (arrondie au centième près) est

A $0,57$ B $0,67$ C $0,43$ D autre réponse

Question 5 La moyenne conditionnelle de Y sachant $X \in [4,\ 12[$ est égale à

A 1,76 B 2,73 C autre réponse D 5

Exercice 2

On considère la distibution statistique double suivante :

$X \setminus Y$	7	1
3	1	2
5	3	2
7	1	2

Question 6 Les variables X et Y sont statistiquement indépendantes.

$\boxed{\text{A}}$ Faux $\boxed{\text{B}}$ Vrai

Question 7 La moyenne marginale de X est

$\boxed{\text{A}}$ $3,73$ $\boxed{\text{B}}$ 4 $\boxed{\text{C}}$ autre réponse $\boxed{\text{D}}$ 5

Question 8 L'écart-type de Y (arrondi au millième près) est égal à

$\boxed{\text{A}}$ $1,477$ $\boxed{\text{B}}$ $3,967$ $\boxed{\text{C}}$ $2,987$ $\boxed{\text{D}}$ autre réponse

Question 9 Les variables X et Y sont linéairement indépendantes.

$\boxed{\text{A}}$ Vrai $\boxed{\text{B}}$ Faux

Question 10 La covariance du couple (X, Y) (arrondie au centième près) est égale à

$\boxed{\text{A}}$ 0 $\boxed{\text{B}}$ $0,02$ $\boxed{\text{C}}$ autre réponse $\boxed{\text{D}}$ $0,01$

$\boxed{\textbf{A}}$

Université Nangui Abrogoua **Année universitaire 2020-2021**
UFR SFA

C

Épreuve de la deuxième session de l'examen de

L'UE DE STATISTIQUE DESCRIPTIVE
ECUE 2 : SERIES STATISTIQUES BIVARIÉES Licence 1 : Maths - Info
Durée : 01 h 00 mn

Consigne : *Le sujet est composé de dix (10) questions réparties en deux (2) exercices indépendants. Pour chacune des questions indiquer, sans justification, la seule bonne réponse sur la grille réponse.*

Barème : Une réponse exacte rapporte 2 points. Une réponse fausse enlève 1 point. L'absence de réponse ne rapporte ni n'enlève aucun point.

Exercice 1

Soit la série statistique double donnée par le tableau de contingence suivant :

$X \setminus Y$	$[0,\ 3[$	$[3,\ 9[$
$[0,\ 4[$	2	4
$[4,\ 12[$	8	3

Question 1 La fréquence conditionnelle de $X \in [0, 4[$ sachant $Y \in [3, 9[$ (arrondie au centième près) est

A $0,67$ B autre réponse C $0,43$ D $0,57$

Question 2 La variance conditionnelle de X sachant $Y \in [0, 3[$ est égale à

A $7,32$ B $10,25$ C $5,76$ D autre réponse

Question 3 La moyenne conditionnelle de Y sachant $X \in [4, 12[$ est égale à

A 2,73 B 1,76 C 5 D autre réponse

Question 4 La formule permettant de calculer $f_{1|2}$ est

A $\frac{n_{12}}{n_{2\bullet}}$ B $\frac{n_{12}}{n_{\bullet 1}}$ C $\frac{n_{12}}{n_{\bullet 2}}$ D $\frac{n_{21}}{n_{1\bullet}}$

Question 5 La moyenne marginale de X (arrondie au centième près) est égale à

A 5,88 B 9,09 C autre réponse D 3,35

Exercice 2

On considère la distibution statistique double suivante :

$X \setminus Y$	7	1
3	1	2
5	3	2
7	1	2

Question 6 La moyenne marginale de X est

A 4 B $3,73$ C 5 D autre réponse

Question 7 Les variables X et Y sont statistiquement indépendantes.

A Faux B Vrai

Question 8 L'écart-type de Y (arrondi au millième près) est égal à

A $3,967$ B $1,477$ C $2,987$ D autre réponse

Question 9 La covariance du couple (X, Y) (arrondie au centième près) est égale à

A $0,01$ B 0 C $0,02$ D autre réponse

Question 10 Les variables X et Y sont linéairement indépendantes.

A Vrai B Faux

C

Université Nangui Abrogoua **Année universitaire** 2020-2021
UFR SFA

B

Épreuve de la deuxième session de l'examen de

L'UE DE STATISTIQUE DESCRIPTIVE
ECUE 2 : SERIES STATISTIQUES BIVARIÉES Licence 1 : Maths - Info
Durée : 01 h 00 mn

Consigne : *Le sujet est composé de dix (10) questions réparties en deux (2) exercices indépendants. Pour chacune des questions indiquer, sans justification, la seule bonne réponse sur la grille réponse.*

Barème : Une réponse exacte rapporte 2 points. Une réponse fausse enlève 1 point. L'absence de réponse ne rapporte ni n'enlève aucun point.

Exercice 1

On considère la distibution statistique double suivante :

$X \setminus Y$	7	1
3	1	2
5	3	2
7	1	2

Question 1 La moyenne marginale de X est

$\boxed{A}$ 5 $\boxed{B}$ 4 $\boxed{C}$ 3,73 $\boxed{D}$ autre réponse

Question 2 L'écart-type de Y (arrondi au millième près) est égal à

$\boxed{A}$ 2,987 $\boxed{B}$ autre réponse $\boxed{C}$ 3,967 $\boxed{D}$ 1,477

Question 3 Les variables X et Y sont statistiquement indépendantes.

$\boxed{A}$ Vrai $\boxed{B}$ Faux

Question 4 La covariance du couple (X, Y) (arrondie au centième près) est égale à

$\boxed{A}$ autre réponse $\boxed{B}$ 0,01 $\boxed{C}$ 0 $\boxed{D}$ 0,02

Question 5 Les variables X et Y sont linéairement indépendantes.

$\boxed{A}$ Faux $\boxed{B}$ Vrai

Exercice 2

Soit la série statistique double donnée par le tableau de contingence suivant :

$X \setminus Y$	$[0,\ 3[$	$[3,\ 9[$
$[0,\ 4[$	2	4
$[4,\ 12[$	8	3

Question 6 La fréquence conditionnelle de $X \in [0, 4[$ sachant $Y \in [3, 9[$ (arrondie au centième près) est

A $0,43$ B $0,57$ C $0,67$ D autre réponse

Question 7 La moyenne conditionnelle de Y sachant $X \in [4, 12[$ est égale à

A 2,73 B 1,76 C autre réponse D 5

Question 8 La formule permettant de calculer $f_{1|2}$ est

A $\frac{n_{12}}{n_{\bullet 2}}$ B $\frac{n_{12}}{n_{\bullet 1}}$ C $\frac{n_{21}}{n_{1\bullet}}$ D $\frac{n_{12}}{n_{2\bullet}}$

Question 9 La variance conditionnelle de X sachant $Y \in [0, 3[$ est égale à

A $7,32$ B $5,76$ C $10,25$ D autre réponse

Question 10 La moyenne marginale de X (arrondie au centième près) est égale à

A 3,35 B 9,09 C autre réponse D 5,88

B

Université Nangui Abrogoua
UFR SFA

Année universitaire 2020-2021

D

Épreuve de la deuxième session de l'examen de

L'UE DE STATISTIQUE DESCRIPTIVE
ECUE 2 : SERIES STATISTIQUES BIVARIÉES Licence 1 : Maths - Info
Durée : 01 h 00 mn

Consigne : *Le sujet est composé de dix (10) questions réparties en deux (2) exercices indépendants. Pour chacune des questions indiquer, sans justification, la seule bonne réponse sur la grille réponse.*

Barème : Une réponse exacte rapporte 2 points. Une réponse fausse enlève 1 point. L'absence de réponse ne rapporte ni n'enlève aucun point.

Exercice 1

On considère la distibution statistique double suivante :

$X \setminus Y$	7	1
3	1	2
5	3	2
7	1	2

Question 1 La covariance du couple (X, Y) (arrondie au centième près) est égale à

A $0,02$ B autre réponse C 0 D $0,01$

Question 2 La moyenne marginale de X est

A 5 B $3,73$ C 4 D autre réponse

Question 3 Les variables X et Y sont linéairement indépendantes.

A Vrai B Faux

Question 4 L'écart-type de Y (arrondi au millième près) est égal à

A $2,987$ B autre réponse C $1,477$ D $3,967$

Question 5 Les variables X et Y sont statistiquement indépendantes.

A Faux B Vrai

Exercice 2

Soit la série statistique double donnée par le tableau de contingence suivant :

$X \setminus Y$	$[0,\ 3[$	$[3,\ 9[$
$[0,\ 4[$	2	4
$[4,\ 12[$	8	3

Question 6 La moyenne marginale de X (arrondie au centième près) est égale à

A 9,09 B 3,35 C 5,88 D autre réponse

Question 7 La variance conditionnelle de X sachant $Y \in [0,\ 3[$ est égale à

[A] $5,76$ [B] $7,32$ [C] autre réponse [D] $10,25$

Question 8 La fréquence conditionnelle de $X \in [0,\ 4[$ sachant $Y \in [3,\ 9[$ (arrondie au centième près) est

[A] autre réponse [B] $0,67$ [C] $0,43$ [D] $0,57$

Question 9 La moyenne conditionnelle de Y sachant $X \in [4,\ 12[$ est égale à

[A] 2,73 [B] 5 [C] 1,76 [D] autre réponse

Question 10 La formule permettant de calculer $f_{1|2}$ est

[A] $\frac{n_{1\,2}}{n_{2\bullet}}$ [B] $\frac{n_{1\,2}}{n_{\bullet 1}}$ [C] $\frac{n_{1\,2}}{n_{\bullet 2}}$ [D] $\frac{n_{2\,1}}{n_{1\bullet}}$

D