

Biostatistique : Série 1

Exercice1

Dire, dans chacune des séries suivantes, s'il s'agit de caractères qualitatifs ou quantitatifs en précisant quelques modalités ou réponses possibles :

- 1) Taille, âge, sexe, taux de cholestérol, couleur des yeux, niveau d'instruction, type de logement, situation familiale, groupe sanguin, nombre de frères et sœurs, lieu de résidence.
- 2) Poids, nombre d'enfants, nature du bac, passe-temps favori, taux de glycémie, lieu de naissance, superficie d'un jardin, nombre de pièces de l'habitation, nombre de cigarettes/jour

Exercice2

Soit les variables suivantes

1. La couleur des cheveux notée " brun ", " blond ", " noir " et " autre "
2. Les modes d'entrée à l'hôpital notés " en provenance du domicile ", " mutation "
3. L'autonomie d'une personne pour l'habillage notée " s'habille seul ", " s'habille sous la surveillance d'un soignant ", " a besoin d'une aide partielle ", " a besoin d'une aide totale "
4. Le taux d'enzyme mesuré en unité /ml
5. La masse d'une tumeur mesurée en grammes
6. Le nombre de comprimés avalés par jour par les personnes de plus de 60 ans
7. L'heure du levé du soleil
8. L'intensité d'une cuti notée "+", "++", "+++"

Répondre par vrai (V) ou faux (F)

- a) 1,2,5 sont des variables qualitatives
- b) 7,8 sont des variables qualitatives ordinales
- c) 3,4,5 sont des variables quantitatives continues
- d) 5 est une variable quantitative discontinue
- e) 5,8 sont des variables quantitatives discontinues

Exercice3

Un sondage a été effectué à la sortie d'un centre de collecte de sang où 50 donneurs ont été interrogés sur le type de leur groupe sanguin.

Les réponses obtenues ont été :

A A A B O B AB O B O A O O AB A A A O B B AB O O O O
A A B A O AB A O O A AB A B O B A B O A A B O A A O

- 1) Quel est le caractère étudié ? Préciser sa nature
- 2) Quelles sont les modalités ?
- 3) Présenter les résultats sous forme de tableau.
- 4) Quel est le pourcentage de personnes du groupe O ? du groupe AB ?

Exercice 4

On a réalisé une étude sur un caractère quantitatif On a obtenu la distribution groupée suivante

classes	<130	[130-150[[150-170[.....
n_i	1	3	2
f_i	0,1	0,3	0,2
n_{icum}	1	4	6	
f_{icum}	0,1	0,4	0,6	

Répondre par vrai (V) ou faux (F)

1. On ne connaît pas l'effectif de l'échantillon

- L'effectif de l'échantillon est égal à 6
- Le centre de classe de la classe [130- 150[est 140
- Dans l'échantillon, il n'y a aucun individu qui a une valeur > 170

Exercice 05

50 femmes de moins de 25 ans sont soumises à une enquête. Pour chacune d'elles, on note le nombre d'enfants nés vivants. Les résultats obtenus sont :

1 3 0 3 2 2 1 2 3 2 1 1 0 2 1 2 0 3 1 1 2 2 3 2 1 3 2 3 2 0 1 2 4 1 1 2 0 4 2 4 0 2 1 1 0 2 0 1 2 3

- Quel est le caractère étudié ? Donner sa nature.
- Donner la distribution de fréquences.
- Tracer le diagramme en bâtons et le diagramme cumulé de cette série.

Exercice 6

Le taux de glucose sanguin (glycémie) déterminé chez 32 patients est donné ci-dessous en g/l
0,85 0,87 0,90 0,93 0,94 0,94 0,95 0,97 0,97 0,98 0,98 0,99 1,00 1,01 1,03 1,03 1,03 1,04 1,06
1,07 1,08 1,08 1,10 1,10 1,11 1,13 1,14 1,14 1,15 1,17 1,19 1,20

- Quel est le caractère étudié ? Donner sa nature.
- Répartir cette série en classes d'égale amplitude.
- Etablir le tableau de distribution de fréquences.
- Tracer l'histogramme et la courbe cumulative

Exercice 7

Etude de parasitisme des châtaignes. On s'intéresse au nombre de parasites par fruit.

Nbre de parasites	0	1	2	3	4	5	6	7
Effectifs	82	55	30	16	9	5	2	1

- Quel est le caractère étudié ? Donner sa nature.
- Compléter le tableau en donnant les fréquences, les effectifs cumulés et les fréquences cumulées.
- Quel est le pourcentage de fruits ayant plus d'un parasite ?
- Tracer le diagramme en bâtons, le polygone de fréquences ainsi que le diagramme cumulé.

Exercice 8

1000 électeurs ont voté pour 4 candidats : A, B, C, et D. A l'issue du scrutin on a obtenu les résultats suivants : 321 ont voté pour A, 48 pour B, 82 pour C et 549 pour D.

- Quel est le caractère étudié ? Donner sa nature.
- Donner une représentation graphique de cette série.

Exercice 9 Une étude portant sur une certaine maladie M selon l'âge a été réalisée dans une population. Les observations réalisées pendant une année sont données par le tableau suivant

Age	[20,30[[30,40[[40,50[[50,60[[60,70[Total
Nbre de malade	5	10	30	75	180	300

En complétant le tableau ci-dessus, calculer

- l'âge médian des malades
- l'âge moyen ainsi que la variance

Exercice 10

Dans une classe de 100 étudiants, une interrogation écrite notée de 0 à 5 par nombres entiers a donné les résultats suivants :

notes	0	1	2	3	4	5
Effectifs	5	20	50	85	95	100
cumulés						

1. Quel est le caractère étudié ? Donner sa nature.
2. Tracer le diagramme en bâtons et le polygone de fréquences.
3. Calculer et représenter graphiquement le mode et les différents quartiles.
4. Calculer la moyenne, la variance et l'écart-type.

Exercice 11

Calculer la moyenne, la variance et les différents quartiles de l'exercice 6

1. Pour la série brute (avant le regroupement en classes)
2. Pour la distribution de fréquences (après le regroupement en classes)

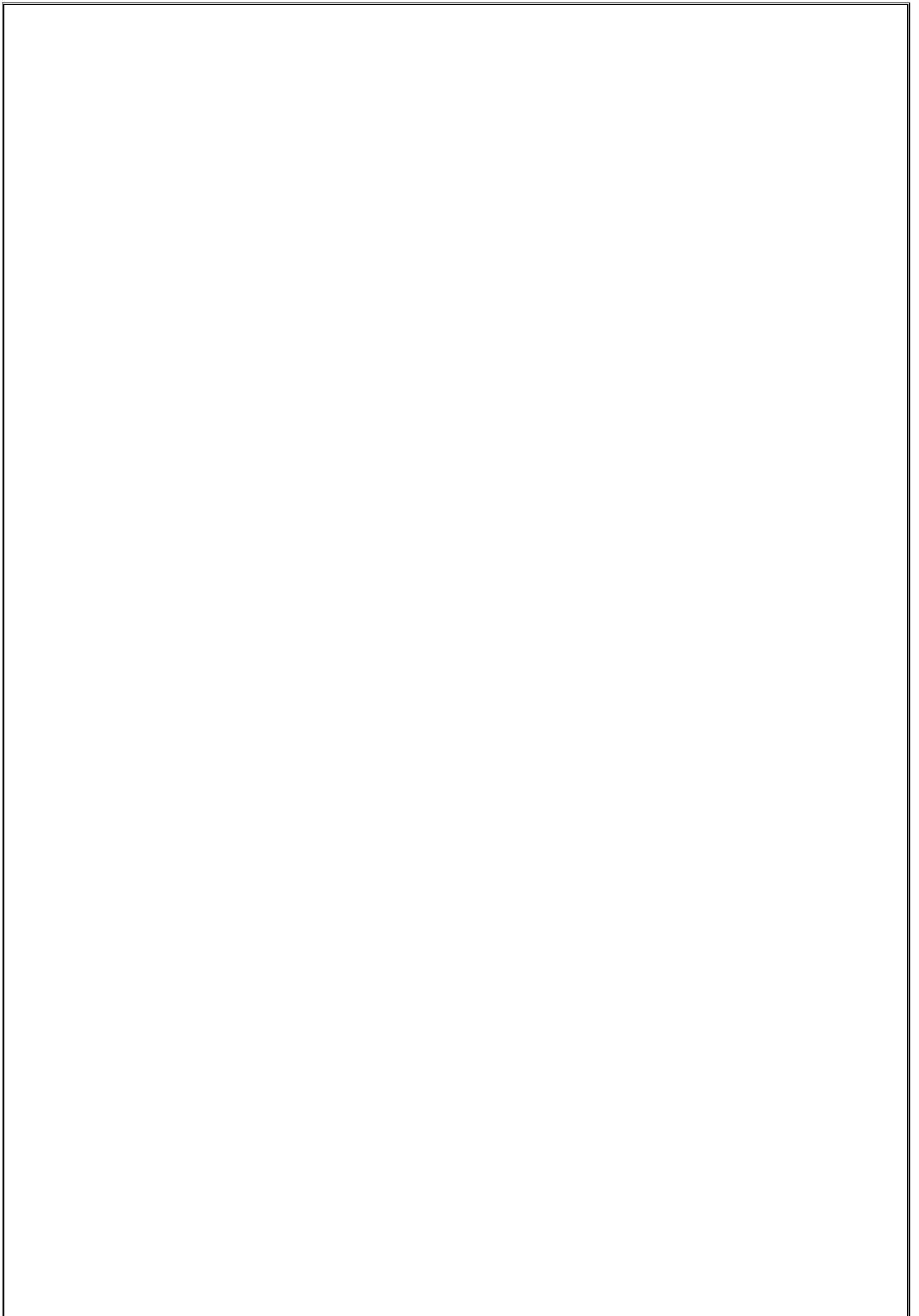
Commenter les résultats.

Exercice 12

Les données suivantes donnent le taux de glycémie à jeun d'un groupe de 36 diabétiques, (glycémie déterminée à 0,01 g/l près) :

1.52 1.25 1.38 1.18 1.49 1.63 1.25 1.54 1.71 1.42 1.50 1.63 1.40 1.65 1.59 1.28
1.64 1.42 1.58 1.48 1.44 1.42 1.61 1.47 1.61 1.44 1.53 1.36 1.52 1.35 1.50 1.32
1.37 1.28 1.57 1.41

1. Ordonner la série statistique par valeurs croissantes, et la répartir en classes d'égale amplitude, en commençant par la classe **[1.18 ;1.26[**
2. Etablir la distribution des fréquences.
3. Calculer le 1er décile et l'intervalle interquartile
4. Calculer la moyenne, la variance et l'écart-type. (Faire un changement de variable.)



Biostatistique : Série 2

Exercice 1. L'indice de masse corporelle (IMC) est défini par : $IMC = \frac{\text{masse en kg}}{(\text{taille en m})^2}$

Il permet de mesurer la corpulence d'un adulte. L'Organisation Mondiale de la Santé (OMS) a défini les critères suivants :

- Maigreur (16,5 à 18,5)
- Normal (de 18,5 à 25)
- Risque de surpoids (de 25 à 30)
- Obésité modérée (de 30 à 35)
- Obésité sévère (35 à 40).

En deçà de 16,5 (dénutrition) et au-delà de 40 (obésité massive), les risques de mortalité sont élevés.

1. On donne ci-contre le poids et la taille d'un échantillon de 13 personnes.

X : poids en kg	70	65	95	58	42	75	45	89	77	83	62	48	59
Y : taille en m	1.68	1.85	1.56	1.61	1.5	1.68	1.65	1.65	1.64	1.75	1.48	1.48	1.74

1. Représenter graphiquement le nuage statistique et calculer la covariance $cov(X, Y)$ et le coefficient de corrélation r_{XY} .
2. Déterminer la droite de régression de y en x et la représenter.
3. Conclure.

Exercice 2.

La série statistique suivante donne le nombre de stomates aérifères au m^2 d'une feuille (variable y) en fonction du nombre de jours d'exposition au soleil (variable x).

Nombre de jours x	2	4	8	10	24	40	52
Nombre de stomates y	6	11	15	20	39	62	85

1. Représenter graphiquement ce nuage de points

2. Calculer les moyennes \bar{x} et \bar{y} .
3. Calculer les écarts types s_x et s_y , la covariance $cov(x, y)$ et le coefficient de corrélation r_{XY} .
4. Déterminer la droite de régression de y en x et la représenter. Conclure.
5. Combien de stomates prévoyez-vous à 30 jours ?

Exercice 3.

Dans le cadre de travaux de recherche sur la durée de la saison de végétation en montagne, des stations météorologiques sont installées à différentes altitudes. La température moyenne (variable y en degrés Celsius) ainsi que l'altitude (variable x en mètres) de chaque station sont données dans le tableau ci-dessous :

Altitude x	1040	1230	1500	1600	1740	1950	2200	2530	2800
Température y	7,4	6	4,5	3,8	2,9	1,9	1	-1,2	-1,5

1. Mêmes questions 1 à 4 que dans l'exercice précédent.
2. Quelle température moyenne prévoyez-vous à 1300 m? à 3000 m?

Exercice 4.

Lors d'une étude sur la dynamique des populations de la tenthrède du pin (*Diprion fruterarum*), la capacité de reproduction (nombre y d'ovocytes par cocon) des individus de cet insecte a été étudiée en fonction de la longueur du cocon (x en mm).

1. Calculer les distributions marginales de x et y .
2. Calculer la moyenne et l'écart type de x et y .
3. Calculer la covariance et le coefficient de corrélation puis déterminer l'équation de la droite de régression de y en x .

$x \backslash y$	[20 ; 40[[40 ; 60[[60 ; 80[[80 ; 100[
[7 ; 7,5[8	2		
[7,5 ; 8[11	15	1	1
[8 ; 8,5[22	8	0
[8,5 ; 9[1	9	10	0
[9 ; 9,5[12	19	11
[9,5 ; 10[10	20

Exercice 5.

On a traité 80 parcelles de 0,4 hectares avec différentes quantités d'un engrais azoté pour la culture d'orge. Le tableau suivant donne la répartition de ces parcelles en fonction de la quantité d'engrais (x en kg) et du rendement obtenu (y en quintaux/hectare).

1. Calculer les distributions marginales de x et y .
2. Calculer la moyenne et l'écart type de x et y .
3. Calculer la covariance et le coefficient de corrélation puis déterminer l'équation de la droite de régression de y en x .

$y \backslash x$	50	60	70	80	90	100
[15;25[1					
[25;35[3	1				
[35;45[1	3				
[45;55[2	7		1	
[55;65[4	8	8	6	
[65;75[5	12	7	2
[75;85[2	6	1	

Biostatistique : Série 3

Exercice 1 :

Soit U une v. a de la normale N (0,1)

1. Calculer $P(U < 2.4)$; $P(U \geq 1,23)$; $P(U > 3,2)$; $P(-1,96 < U < 1,96)$
2. Déterminer u_1, u_2, u_3 , tels que $P(U < u_1) = 0,950$; $P(U > u_2) = 0,25$; $P(U < u_3) = 0,015$

Exercice 2 :

Soit X une v. a de loi normale de paramètres $m = 32$ et $\sigma^2 = 49$

1. Calculer $P(X > 32)$; $P(X \geq 32)$; $P(X < 32 + \sigma)$; $P(32 < X < 32 + \sigma)$
2. Déterminer x_1, x_2, x_3, x_4 tels que : $P(X < x_1) = 0.791$; $P(X > x_2) = 0,123$; $P(X < x_3) = 0,015$; $P(|X - 32| < x_4) = 0,20$

Exercice 3 :

Dans une population humaine, la mesure de la quantité d'urée dans le sang en (mg/100ml) a donné une moyenne de 27mg/100ml et une variance de 25 (mg/100ml)². On suppose que cette variable est de loi normale. On prend un individu au hasard

1. Quelle est la probabilité que la quantité d'urée soit inférieure à 24,5 ?
2. Déterminer la limite x_0 , appelée seuil pathologique, tel qu'il ait 2,5% d'individus dont la quantité d'urée est supérieure à cette valeur x_0 .

Exercice 4 :

On considère un échantillon de 300 étudiants dans lequel l'âge, X d'un individu est supposé être une v. a de loi $N(m, \sigma^2)$ avec $m=21$ ans.

1. Déterminer σ si l'on sait que $P(X > 22) = 30,85\%$
2. Quelle est la probabilité que l'âge d'un étudiant soit compris entre 18 et 23 ans ?
3. Quel est le nombre d'étudiants dont l'âge dépasse 19 ans ?
4. Quel est l'âge minimal des 15,87% étudiants les plus âgés ?

Exercice 5 :

On veut connaître la population P d'individus d'une population très grande, résistant à un certain antibiotique. Sur un échantillon de 1000 individus, on a observé que le nombre de moyen d'individus résistant à l'antibiotique est 62.5.

Soit X le nombre d'individus résistant à l'antibiotique parmi les 1000 observés.

1. Donner la loi de la variable a X et donner une estimation de la population P.
2. Par quelle loi peut-on approximer la loi de X ? calculer alors $P(X > 30)$ et $P(30 < X < 40)$.