

Statistique descriptive : Répétition N°2

EX1

Lors d'une campagne de prévention sur l'ostéoporose, trois variables ont été observées chez 20 femmes: l'âge (années), le poids (kg), le score Sahara (sans unité) et le facteur de risque "a déjà eu des fractures sans choc violent" (1=oui, 0=non). Le score Sahara détermine si la densité osseuse est normale (> -1) ou si elle est pathologique (≤ -1).

Age	Poids	Sahara	Fracture
59	92	0.1	0
74	71	-2	1
65	72	1.7	0
71	57	0.8	0
65	60	-1.2	1
51	63	-1	1
54	102	0.2	0
71	59	1	0
63	104	0.1	1
70	55	-0.6	0
72	80	-1.6	1
62	92	-0.7	0
65	68	-1.8	0
69	100	-1.5	0
52	62	0.1	0
66	68	-0.4	0
74	54	-1.2	1
73	66	-1.6	1
68	71	-0.9	1
66	69	-1.1	0

- Quelle est le type de chacune de ces variables ?
- Pour les variables « Age » et « Poids », déterminer le mode, la médiane, la moyenne, la variance et l'écart-type.
- Mesurer le lien entre ces deux variables à l'aide du coefficient de Pearson.
- Construire un histogramme de la variable « Sahara » (répartir préalablement les données en 8 classes d'amplitude constante). Quelle est la classe modale ?
- Calculer, séparément pour les femmes à risque et les femmes sans risque (de fracture), la moyenne du score Sahara. Que peut-on conclure ?

EX2

Le tableau de contingence ci-dessous décrit la répartition d'une population constituée de 535 ménages selon les deux variables suivantes : X représente le nombre de pièces de l'habitation et Y correspond au nombre d'enfants du ménage.

Tableau de contingence pour la série double « nombre de pièces – Nombre d'enfants ».

Valeurs de X	Valeurs de Y				
	0	1	2	3	4
1	7	3	2	1	0
2	24	32	21	2	1
3	16	35	54	26	4
4	9	28	74	55	12
5	4	12	46	13	12
6	2	6	16	11	7

- Que représente les fréquences n_{23} et n_{54} ? Calculer les fréquences relatives correspondantes.
- Déterminer les distributions marginales des variables X et Y. Calculer les moyennes, médianes et variances marginales ainsi que les modes marginaux.

- c) Déterminer la distribution conditionnelle de la variable Y sachant que le nombre de pièces du logement est égal à 4. Calculer la moyenne et la variance de cette distribution conditionnelle.

EX3

Le responsable des ressources humaines d'une entreprise s'intéresse au bien-être de ses employés. A cet effet, il interroge ceux-ci et souhaite notamment, savoir si les dépenses mensuelles pour les loisirs sont corrélées avec les revenus des travailleurs. Il recueille les données suivantes (exprimées en une certaine unité monétaire) :

X	Y	X	Y	X	Y
752	85	492	81	679	80
855	83	569	81	902	226
871	162	462	80	918	260
735	79	907	243	828	82
610	81	643	84	875	186
582	83	862	84	809	77
921	281	524	82	894	223

- a) Représenter graphiquement le nuage de points correspondant aux variables X et Y relatives, respectivement, aux revenus mensuels nets et aux dépenses mensuelles en loisirs.
 b) Déterminer la valeur du coefficient de corrélation et interpréter cette valeur en tenant compte du contexte.

EX4

Le tableau ci-dessous reprend les résultats d'un sondage sur les intentions de vote en Grande-Bretagne entre deux élections de 1974 :

	Octobre		
Février	Conservateurs	Libéraux	Travailleurs
Conservateurs	170	20	3
Libéraux	22	70	28
Travailleurs	6	12	227

- a) Quel est le pourcentage d'individus qui n'ont pas changé d'intention de vote ?
 b) En termes de pourcentage, quel parti a perdu le plus d'intentions de vote entre les deux périodes de sondage ? Quel parti en a gagné le plus ?
 c) Dresser l'évolution des tendances de vote entre les deux périodes de sondage.