

Exercice n°1 Préciser le type des variables statistiques suivantes :

Série du baccalauréat ☐ durée du trajet domicile-université ☐ mode de transport ☐
 nombre de frères et soeurs ☐ note en français au baccalauréat ☐ catégorie socio-professionnelle du père ☐

Exercice n°2 On a relevé pour vingt individus, certaines caractéristiques biologiques et sociologiques. Le tableau ci-dessous résume les résultats obtenus :

Individu	Groupe sanguin	Âge	Nombre d'enfants
1	B	38	3
2	AB	42	4
3	A	16	0
4	A	51	2
5	O	28	1
6	A	34	1
7	O	21	1
8	A	54	4
9	AB	35	0
10	A	22	3
11	A	12	0
12	B	40	2
13	O	41	1
14	O	31	2
15	A	25	1
16	AB	64	6
17	A	60	1
18	O	55	0
19	B	49	2
20	A	42	2

1. Quel est l'effectif total ?
2. Pour chacune des variables statistiques, indiquer sa nature.
3. Déterminer les effectifs des variables "groupe sanguin" et "nombre d'enfants".
4. On décide de regrouper les données individuelles de la variable "âge" selon le découpage suivant :

$[0, 18[$, $[18, 25[$, $[25, 35[$, $[35, 49[$, $[49, 55[$ et $[55, 70[$.

Déterminer la distribution des effectifs de la variable "âge" associée à ce découpage en classes.

5. Déterminer les fréquences de la variable "groupe sanguin". Représenter graphiquement cette variable. Quel est le mode de cette variable ?

Exercice n°3 Dans le but d'étudier la reproduction de la chouette effraie (*Tyto alba*) en Côte d'Or, on a compté les proies rapportées au nid par les adultes pour des jeunes âgés de moins de 21 jours. Les résultats sont reportés dans le tableau ci-dessous :

Type de proies	Campagnol	Mulot	Musaraigne	Souris	Grenouille
Effectif	24	x	7	4	3
Fréquence	0.48	0.24	0.14	y	z

Malheureusement, trois cases du tableau ont été effacées : retrouver les valeurs manquantes. Quel est la taille de l'échantillon (effectif total) ?

Exercice n°4 En 1990, lors du recensement effectué par l'INSEE, on a étudié la répartition des ménages français selon leur type. Les résultats furent les suivants (en milliers) :

Type de ménage	H. seul	F. seule	H. + enfant(s)	F. + enfant(s)	Couples	Autres
Effectif	2171	3673	225	1325	13702	443

1. Quelle est la population étudiée ?
2. Quel est le type de la variable "type du ménage" ?
3. Quel était le nombre de ménages en 1990 ?
4. Calculer la proportion d'adultes vivant seuls en 1990.
5. Quel est le mode de cette variable ?
6. Lors d'une enquête en 1999, on a de nouveau recensé la répartition des français selon leur type. Les résultats furent les suivants (en milliers) :

Type de ménage	H. seul	F. seule	H. + enfant(s)	F. + enfant(s)	Couples	Autres
Effectif	2942	4411	259	1444	14603	489

- (a) Quel était le nombre de ménages en 1999 ?
- (b) Calculer la proportion d'adultes vivant seul en 1999.
- (c) Représenter sur un même graphique les résultats observés en 1990 et 1999.

Exercice n°5 On a recensé en 1993 et 1998 la répartition du personnel enseignant du supérieur suivant leur statut. Les résultats furent les suivants :

Statut	Professeurs	Maîtres de conférences	ATER	Autres
Effectif en 1993	17000	30700	6600	11200
Effectif en 1998	16600	31400	5500	25200

1. Quel est le type de la variable "statut" ?
2. Quel était le nombre d'enseignants du supérieur en 1993 et 1998 ?
3. Calculer les fréquences de la variable "statut" en 1993, puis en 1998.
4. Quel est l'intérêt de travailler avec les fréquences plutôt qu'avec les effectifs ?
5. Représenter sur le même graphique les fréquences de la variable "statut" en 1993 et 1998.

Exercice n°6 Un sondage intitulé "les français et la vie sans télévision" a été réalisé les 5 et 6 septembre 1997 auprès de 1000 français de 18 ans et plus. Les réponses obtenues à la question "Vous arrive-t-il de regretter d'avoir la télévision ?" ont permis d'établir le tableau ci-dessous :

Réponse	Fréquence
Souvent	7,00%
De temps en temps	23,00%
Rarement	10,00%
Jamais	56,00%
N'a pas la télévision	4,00%

1. Quel est le type de la variable "réponse" ? Peut-elle être considérée comme une variable ordinale ?
2. Représenter graphiquement la distribution des effectifs de la variable "réponse".
3. Quel est le mode de la variable "réponse" ?

Exercice n°7 Cinquante exemplaires d'un emballage spécial sont soumis à des essais de résistance. Pour chacun, on a relevé le nombre de chocs nécessaires pour obtenir la rupture de l'emballage. Les résultats sont les suivants :

2 2 3 5 2 1 4 2 3 5 3 2 3 3 4 1 2 4 2 2 4 2 3 2 3
3 2 2 4 2 1 4 2 3 2 2 3 1 3 3 2 3 2 2 3 4 3 2 3 2

1. Préciser la population et la variable étudiée. Quelle est la nature de la variable étudiée ?
2. Regrouper les données dans un tableau.
3. Proposer deux méthodes pour calculer la moyenne. En choisir une et effectuer le calcul.
4. Même question pour l'écart-type.
5. Calculer les fréquences et les fréquences cumulées. Tracer les graphiques correspondants.
6. Après ce test, un nouveau procédé de fabrication a été mis au point. Les résultats des tests de résistance sur 8 emballages sont les suivants :

1 3 5 3 4 3 2 3

Le nouvel emballage vous semble-t-il plus performant ?

Exercice n°8 On observe les données suivantes :

Nb. d'enfants	0	1	2	3	4	5	6	7	8
Nb. de ménages	45	71	49	24	20	18	9	3	2

1. Préciser la population étudiée : les enfants ou les ménages ? Préciser la variable étudiée. Quelle est la nature de la variable étudiée ?
2. Représenter la série par un graphique.
3. Calculer la moyenne et l'écart-type.
4. Calculer les fréquences cumulées. Quelle est la proportion de ménages ayant moins de 6 enfants (au sens large) ?
5. Indiquer le mode.

Exercice n°9 Dans une entreprise, on étudie la distribution des salaires des employés femmes et hommes. Le salaire moyen des hommes est 1200€ et celui des femmes 1150€.

1. Dans quel cas, le salaire moyen de l'ensemble des employés est-il égal à la moyenne des salaires moyens hommes/femmes ?
2. On sait que dans l'entreprise, il y a 65% d'hommes. Quelle est le salaire moyen de l'ensemble des employés ?

Exercice n°10

En 1997, l'Etat du Québec a recensé la répartition des logements selon le nombre de pièces de l'habitation principale des ménages québécois. L'étude a consisté à répartir les ménages selon les classes suivantes :

- A : Logement de strictement moins de quatre pièces
- B : Logement de quatre pièces exactement
- C : Logement de cinq ou six pièces
- D : Logement de sept pièces et plus

Les résultats de cette étude sont les suivants :

Type de logement	A	B	C	D
Effectif	821178	392607	641358	1141857

1. Préciser la population et la variable étudiées. Quel est le type de cette variable ?
2. Quel est l'effectif total ?
3. Représenter graphiquement la distribution.
4. Déterminer le mode de la variable étudiée.
5. Calculer le nombre de ménages habitant dans un logement ayant strictement moins de cinq pièces.
6. Est-ce qu'il y a du sens à calculer les fréquences cumulées ? Justifier votre réponse.

Exercice n°11 On a relevé les nombres d'allumettes contenues respectivement dans 20 boîtes lors d'un contrôle de fabrication dans une usine. Les résultats sont les suivants :

40 42 32 38 40 48 30 38 36 40
34 40 34 40 38 40 42 44 36 42

1. Préciser la population et la variable étudiée. Quelle est la nature de la variable étudiée ?
2. Calculer la moyenne et l'écart-type.
3. Regrouper les données dans des classes d'amplitude de 4 allumettes.
4. Tracer l'histogramme de la série des données regroupées.
5. Calculer la moyenne et l'écart-type. Commenter.

Exercice n°12 Le tableau suivant indique la durée de chômage (exprimée en mois) pour un échantillon de $n = 14845$ chômeurs inscrits à l'ANPE en 1986.

Classe	Amplitude	Effectif	Fréquence	Eff. cumulé	Fréq. cumulée	Fréq. cum. décroissante
[0; 1[2221				
[1; 2[2338				
[2; 3[1587				
[3; 5[1989				
[5; 7[1402				
[7; 9[950				
[9; 12[1035				
[12; 15[786				
[15; 18[803				
[18; 30[1734				
Total						

1. Préciser la population et la variable étudiée. Quelle est la nature de la variable étudiée ?
2. Compléter le tableau.
3. Quelle est le pourcentage de durées ne dépassant pas 18 mois ?
4. Calculer la médiane, puis la situer sur le graphique précédent.
5. Représenter le polygone des fréquences cumulées (ou fonction des fréquences cumulées) et le polygone des fréquences décroissantes sur un même graphique : que représente le point d'intersection des deux polygones ?
6. Calculer la moyenne.
7. Préciser la classe modale.

Exercice n°13 En cours, on a vu une formule d'approximation linéaire pour le calcul de la médiane pour des données continues regroupées :

$$Me = a_j + (a_{j+1} - a_j) \frac{0.5 - F(a_j)}{F(a_{j+1}) - F(a_j)} .$$

Donner une formule analogue pour le quantile d'ordre $\alpha \in]0; 1[$.

Application numérique : calculer les différents quartiles pour les données de l'exercice précédent.

Exercice n°14 Une machine remplit automatiquement des paquets de tabac. On prélève un échantillon de la production et, après pesée, on obtient les résultats suivants (poids exprimées en grammes) :

Poids	< 38	< 39	< 39.5	< 40	< 40.5	< 41	< 41.5	< 42	< 42.5	< 43	< 44
Eff. cum.	0	3	8	18	31	51	69	84	95	99	100

1. Retrouver le tableau des effectifs.
2. Tracer l'histogramme de la distribution.
3. Tracer la fonction des fréquences cumulées.
4. Calculer la moyenne et l'écart-type.
5. Regrouper les données à l'aide de classe d'amplitude 2 grammes. Répondre aux mêmes questions que précédemment et commenter.

Exercice n°15 L'utilisation en vase clos d'un insecticide a donné les résultats suivants sur la durée de vie des insectes (exprimée en secondes) :

Durée de vie	$[0; 4[$	$[4; 8[$	$[8; 12[$	$[12; x_4[$	$[x_4; 22[$	$[22; 30[$	$[30; 42[$
Fréquence	6%	p_2	p_3	17%	14%	11%	3%

Certaines données sont illisibles et ont été remplacées par x_4 , p_2 et p_3 .

1. Déterminer p_2 et p_3 en sachant que le quatrième décile est égal à 9.5 secondes.
2. Déterminer x_4 sachant que la durée de vie moyenne des insectes soumis à l'insecticide est égale à 13 secondes.
3. Tracer l'histogramme de la distribution.
4. Calculer l'écart-type.

Exercice n°16 Les données ci-dessous représentent les durées (exprimées en minutes) de 400 consultations d'un site web :

Durée de consultation	$[0; 8[$	$[8; 16[$	$[16; 24[$	$[24; t[$	$[t; 44[$	$[44; 60[$	$[60; 84[$
Effectif	24	a	b	68	56	44	12

Certaines données sont illisibles et ont été remplacées par t , a et b .

1. Déterminer a et b en sachant que le quatrième décile est égal à 19 minutes.
2. Déterminer t sachant que la durée moyenne de consultation est égale à 26 minutes.
3. Tracer l'histogramme de la distribution.
4. Situer sur le graphique précédent la médiane et la calculer. Calculer les autres quartiles.
5. Calculer la moyenne.
6. Représenter les différentes caractéristiques numériques à l'aide d'un box-plot.

Exercice n°17 Soit $\{(x_j, n_j), j = 1, \dots, k, \sum_{j=1}^k n_j = n\}$ une série regroupée d'une variable quantitative discrète de moyenne \bar{x} et d'écart-type σ_x . Soit k un réel positif et l'intervalle $I_k = [\bar{x} - k\sigma_x; \bar{x} + k\sigma_x]$. Montrer que la proportion des observations à l'extérieur de I_k est inférieure à $\frac{1}{k^2}$. En déduire que la proportion des observations comprises dans I_2 est au moins égale à trois-quarts.

[Indication : on pourra décomposer la variance en une somme de deux termes à l'aide de I_k et de son complémentaire...]