0Q3 - Exercices statistiques descriptives

1. Déterminer si les variables suivantes sont quantitatives (discrètes ou continues) ou qualitatives (nominales ou ordinales).

   a) Le nombre de cours réussis au Collège.

   b) La taille d'un étudiant.

   c) Le sport préféré d'un enfant.

   d) Le temps passé devant la télé par jour.

   e) Le degré de satisfaction d'un étudiant face à son rendement scolaire.

   f) Le nombre d'enfants dans une famille.

2. Voici les résultats présentant la répartition d'un échantillon de 65 enseignants selon leurs années d'expérience.

2	2	3	3	4	4	4	5	5	5
5	5	6	6	7	7	7	7	7	7
7	8	8	8	8	8	10	10	10	10
12	12	13	14	14	14	14	15	15	15
16	16	17	17	17	17	17	18	19	20
20	21	22	22	22	23	23	24	24	27
27	28	32	33	33					

   a) Quelle est la population ?

      Quel est l'échantillon?

      Quelle est l'unité statistique?

   b) Quelle est la variable étudiée ?

   c) De quelle nature est-elle ?

   d) Déterminer le mode.

   e) Déterminer la médiane. Interpréter cette médiane.

   f) Déterminer la moyenne.

   g) Calculer l'étendue des données.

   h) Calculer la valeur de l'écart type.

   i) Calculer la valeur du coefficient de variation

3. Les salaires annuels offerts à 12 nouveaux gradués en comptabilité sont résumés dans le tableau suivant :

Salaires annuels de 12 nouveaux gradués en comptabilité			
17 500	18 400	19 200	18 000
16 750	18 500	19 000	18 200
20 000	18 800	17 950	18 740

Quel est le salaire moyen et médian ? Comment qualifieriez-vous la distribution de ces données (symétrique ou non)?

4. Voici les résultats en mathématique de 3 étudiants provenant de 3 groupes différents. On connaît la note de l'étudiant, la moyenne du groupe et l'écart type du groupe.

Nom étudiant	**Note**	**Moyenne**	**Écart type**
Océane	95	82	11
Marie-Soleil	89	79	7,2
Pierre	91	80	8

On doit choisir un seul de ces trois étudiants et, pour ce faire, on doit déterminer celui ou celle qui s'est le plus démarqué dans son groupe. Qui devrait-on choisir ? Pourquoi?

5. Un employé d'un traversier a noté le nombre d'occupants par automobile transportée. Voici les données recueillies pour le mois de juin.

Nb. occup	f	f(%)
1	201	
2	102	
3	99	
4	46	
5	21	
6	13	

a) Quel est le type de la variable présentée.

b) Compléter la 3^{e} colonne de ce tableau.

c) Quel type de graphique serait approprié pour représenter ces résultats ?

6. Le responsable des achats au collège doit faire un achat en gros d'ampoules électriques de 60 watts. Deux choix s'offrent à lui : des ampoules provenant d'un lot dont la moyenne de durée de vie est de 1000 heures avec un écart-type de 28 heures, et celles provenant d'un autre lot dont la moyenne de durée de vie est de 1300 heures avec un écart-type de 60 heures. S'il désire acheter les ampoules provenant du lot le plus homogène, quel lot choisira-t-il? Expliquer l'avantage que lui apportera le choix du lot le plus homogène.

7. Voici la distribution des notes obtenues par un groupe d'étudiants à test de statistiques.

Note (%)	f
20 – 30	4
30 - 40	12
40 - 50	12
50 - 60	7
60 - 70	46
70 – 80	20
80 – 90	3
90 – 100	1

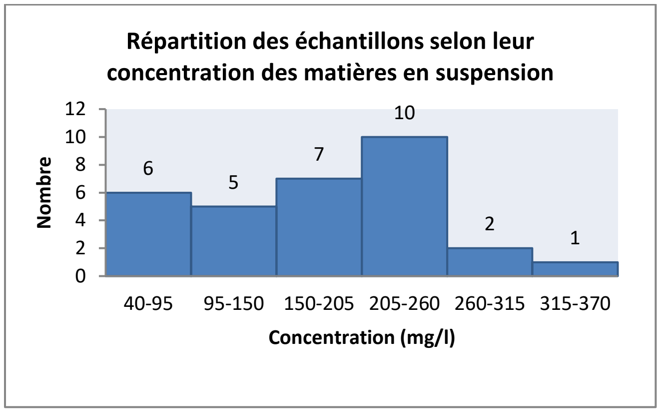
a) Le nombre de classes pour la présentation de ces données, est-il convenable. Justifier votre réponse…

b) À partir des fréquences relatives, tracer l'histogramme correspondant à ce tableau

c) Calculer une approximation (formule d'interpolation linéaire) de la note « médiane » de ce test.

d) À partir des données de ce tableau, calculer la moyenne, l'écart-type et le coefficient de variation de la variable test.

8. Voici un histogramme présentant la concentration des matières en suspension décantables d'une eau usée domestique.



a) Quelle est la taille de cet échantillon?

b) Donner une approximation de l'étendue des données de cet échantillon.

c) Donner une approximation de la moyenne.

d) Donner une approximation de l'écart-type.

9.

Déterminer les mesures de tendance centrale : mode, médiane et moyenne, des distributions suivantes.

a)

Valeurs x_i	Effectifs n_i
3	5
4	8
5	15
6	19
7	14
8	10
9	4
Total	75

b)

Valeurs x_i	Effectifs n_i
2	4
4	7
5	12
7	23
8	35
9	8
Total	89

c)

Classes $[x_i; x_{i+1}[$	Effectifs n_i
[59; 69[	5
[69; 79[	13
[79; 89[	15
[89; 99[	28
[99; 109[	22
[109; 119[	14
[119; 129[	8
Total	105

d)

Classes $[x_i; x_{i+1}[$	Effectifs n_i
[24; 34[	9
[34; 44[	19
[44; 54[	27
[54; 64[	35
[64; 74[	42
[74; 84[	24
[84; 94[	15
Total	171

10.

7. Calculer la moyenne, la variance, l'écart-type et le coefficient de variation des distributions à partir de leur tableau de distribution.

a)

Valeurs x_i	Effectifs n_i
2	9
3	17
4	26
5	32
6	24
7	18
8	9
Total	135

b)

Valeurs x_i	Effectifs n_i
2	18
3	21
4	12
5	5
6	10
7	18
8	20
Total	104

c)

Classes $[x_i; x_{i+1}[$	Effectifs n_i
[24; 28[	4
[28; 32[	9
[32; 36[	42
[36; 40[	68
[40; 44[	56
[44; 48[	18
[48; 52[	3
Total	200

d)

Classes $[x_i; x_{i+1}[$	Effectifs n_i
[24; 28[	39
[28; 32[	42
[32; 36[	12
[36; 40[	5
[40; 44[	14
[44; 48[	36
[48; 52[	52
Total	200

e) Les variables des distributions en **a** et **b** prennent les mêmes valeurs et ont la même étendue. Quelles sont les mesures qui les différencient?

f) Les classes des distributions en **a** et **b** sontles mêmes et elles ont la même étendue. Quelles sont les mesures qui les différencient?

Solutions

1. a) Quantitative discrète b) Quantitative continue c) Qualitative nominale
d) Quantitative continue e) Qualitative ordinale f) Quantitative discrète

2. a) Population : **L'ensemble** des enseignants Échantillon : 65 enseignants Unité statistique : 1 enseignant
b) Nombre d'années d'expérience
c) Quantitative continue

d) Mo = 7 années d'expérience
e) Md = 13 années d'expérience ➔ Il y a <u>au moins</u> 50% des enseignants qui ont <u>13 ans d'expérience ou moins</u>.
f) $\bar{x} \approx 13{,}6$ années d'expérience
g) E = 33 – 2 = 31 années d'expérience
h) s ≈ 8,2 années d'expérience
i) $CV = \frac{s}{\bar{x}} \times 100\% = \frac{8{,}1}{14} \times 100\% \approx 58\% > 15\%$ ➔ Donc les données sont très dispersées autour de la moyenne. La distribution est hétérogène.

3. Moyenne= 18420$ et Md = $\frac{18\ 400 + 18\ 500}{2} = 18\ 450$ $. Distribution plutôt symétrique.

4. On devrait choisir Marie-Soleil car elle a une cote z d'environ 1,39, tandis que Pierre a une cote z d'environ 1,38 et Océane a une cote z d'environ 1,18.

5. a)

X	f	f(%)
1	201	41.7
2	102	21.2
3	99	20.5
4	46	9.5
5	21	4.4
6	13	2.7
Total	482	

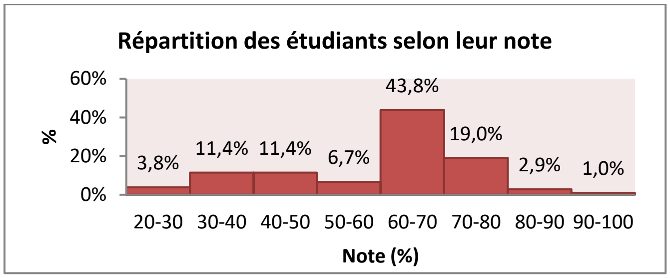
b) Quantitative discrète c) Diagramme à bâtons (ou à barres) verticaux

6.
Il devrait choisir le 1e lot pcq CV inférieur…

7.

a) oui

b)



c) 63.8%

d) Moyenne=59,9% ; écart-type=15,13% ; CV=25,3%

8.

a) n=31
b) E=330 mg/l
c) Moyenne=177,5
d) Écart-type=75,14

9.

a) On ajoute des colonnes au tableau pour calculer les effectifs cumulés et les produits de chaque valeur de la variable par son effectif.

Valeurs x_i	Effectifs n_i	Effectifs cumulés	Produits $x_i n_i$
3	5	5	15
4	8	13	32
5	15	28	75
6	19	47	114
7	14	61	98
8	10	71	80
9	4	75	36
Total	75		450

Dans la colonne des effectifs, on repère facilement que le mode de cette distribution est 6, c'est la valeur de la variable qui a le plus grand effectif.

L'effectif total est 75, en divisant par 2, cela donne 37,5. Dans la colonne des effectifs cumulés, on constate que l'effectif cumulé pour la valeur 5 est 28, la médiane est donc plus grande que 5. L'effectif cumulé pour la valeur 6 est 47. Cet effectif cumulé est supérieur à 37,5. La médiane de la distribution est donc 6.

La moyenne est le total de la colonne des produits $x_i n_i$ divisé par le total de la colonne des effectifs, soit :

$$\mu = \frac{450}{75} = 6.$$

b) On ajoute des colonnes au tableau pour calculer les effectifs cumulés et les produits de chaque valeur de la variable par son effectif.

Valeurs x_i	Effectifs n_i	Effectifs cumulés	Produits $x_i n_i$
2	4	4	8
4	7	11	28
5	12	23	60
7	23	46	161
8	35	81	280
9	8	89	72
Total	89		609

Dans la colonne des effectifs, on repère facilement que le mode de cette distribution est 8, c'est la valeur de la variable qui a le plus grand effectif.

L'effectif total est 89, en divisant par 2, cela donne 44,5. Dans la colonne des effectifs cumulés, on constate que l'effectif cumulé pour la valeur 5 est 23, la médiane est donc plus grande que 5. L'effectif cumulé pour la valeur 7 est 46. Cet effectif cumulé est supérieur à 44,5. La médiane de la distribution est donc 7.

La moyenne est le total de la colonne des produits $x_i n_i$ divisé par le total de la colonne des effectifs, soit :

$$\mu = \frac{609}{89} = 6{,}842\ \dots$$

c) On ajoute des colonnes au tableau pour calculer les effectifs cumulés, inscrire les milieux des classes et les produits de chaque milieu de classe par l'effectif de cette classe.

Classes $[x_i; x_{i+1}[$	Effectifs n_i	Effectifs cumulés	Milieux m_i	Produits $m_i n_i$
[59; 69[	5	5	64	320
[69; 79[	13	18	74	962
[79; 89[	15	33	84	1 260
[89; 99[	28	61	94	2 632
[99; 109[	22	83	104	2 288
[109; 119[	14	97	114	1 596
[119; 129[	8	105	124	992
Total	105			10 050

Dans la colonne des effectifs, on repère facilement que la classe modale de cette distribution est [89; 99[, c'est la classe modale qui a le plus grand effectif.

L'effectif total est 105, en divisant par 2, cela donne 52,5. Dans la colonne des effectifs cumulés, on constate que l'effectif cumulé pour la classe [79; 89[est 33, la médiane est donc plus grande que 89. Elle est dans la classe [89; 99[. Pour la calculer, on doit déterminer la proportion des effectifs de la classe à gauche de la médiane. Puisque 52,5 – 33 = 19,5, la proportion des effectifs de la classe [89; 99[à gauche de la médiane doit être de 19,5/28. L'amplitude de la classe est 10, on doit donc avoir :

$$Me = 89 + \frac{19{,}5}{33} \times 10 = 94{,}909 \ldots$$

La moyenne est le total de la colonne des produits $m_i n_i$ divisé par le total de la colonne des effectifs, soit :

$$\boxed{\mu = \frac{10\ 050}{105} = 95{,}714 \ldots}$$

d) On ajoute des colonnes au tableau pour calculer les effectifs cumulés, inscrire les milieux des classes et les produits de chaque milieu de classe par l'effectif de cette classe.

Classes $[x_i; x_{i+1}[$	Effectifs n_i	Effectifs cumulés	Milieux m_i	Produits $m_i n_i$
[24; 34[	9	9	29	261
[34; 44[	19	28	39	741
[44; 54[	27	55	49	1 323
[54; 64]	35	90	59	2 065
[64; 74[	42	132	69	2 898
[74; 84[	24	156	79	1 896
[84; 94[	15	171	89	1 335
Total	171			10 519

Dans la colonne des effectifs, on repère facilement que la classe modale de cette distribution est [64; 74[, c'est la classe modale qui a le plus grand effectif.

L'effectif total est 171, en divisant par 2, cela donne 85,5. Dans la colonne des effectifs cumulés, on constate que l'effectif cumulé pour la classe [44; 54[est 55, la médiane est donc plus grande que 54. Elle est dans la classe [54; 64[. Pour la calculer, on doit déterminer la proportion des effectifs de la classe à gauche de la médiane. Puisque 85,5 – 55 = 30,5, la proportion des effectifs de la classe [89; 99[à gauche de la médiane doit être de 30,5/35. L'amplitude de la classe est 10, on doit donc avoir :

$$Me = 54 + \frac{30{,}5}{35} \times 10 = 62{,}714 \ldots$$

La moyenne est le total de la colonne des produits $m_i n_i$ divisé par le total de la colonne des effectifs, soit :

$$\mu = \frac{10\ 519}{171} = 61{,}514 \ldots$$

10.

a) On ajoute une colonne au tableau de distribution pour calculer la somme des produits $n_i x_i$. On en tire la moyenne :

$$\mu = \frac{\sum_{i=1}^{k} n_i x_i}{n} = \frac{675}{135} = 5.$$

Valeurs x_i	Effectifs n_i	Produits $n_i x_i$	Produits $n_i (x_i - \mu)^2$
2	9	18	81
3	17	51	68
4	26	104	26
5	32	160	0
6	24	144	24
7	18	126	72
8	9	72	81
Total	135	675	352
	n	$\sum_{i=1}^{k} n_i x_i$	$\sum_{i=1}^{k} n_i (x_i - \mu)^2$

On ajoute une colonne supplémentaire pour calculer, pour chaque valeur de la variable, les produits de l'effectif par le carré de la différence à la moyenne μ. On utilise ces résultats pour compléter les calculs.

La variance est $\sigma^2 = \frac{\sum_{i=1}^{k} n_i (x_i - \mu)^2}{n} = \frac{352}{135} = 2,6074$, l'écart-type est $\sigma = \sqrt{\sigma^2} = \sqrt{2,6074} = 1,614 \ldots$

Le coefficient de variation est $CV = \frac{\sigma}{\mu} \times 100 = \frac{1,614 \ldots}{5} = 32,2 \ldots$

b) On ajoute une colonne au tableau de distribution pour calculer la somme des produits $n_i x_i$. On en tire la moyenne :

$$\mu = \frac{\sum_{i=1}^{k} n_i x_i}{n} = \frac{518}{104} = 4,98.$$

Valeurs x_i	Effectifs n_i	Produits $n_i x_i$	Produits $n_i(x_i-\mu)^2$
2	18	36	159,930
3	21	63	82,392
4	12	48	11,543
5	5	25	0,002
6	10	60	10,388
7	18	126	73,391
8	20	160	182,315
Total	104	518	519,9615
	n	$\sum_{i=1}^{k} n_i x_i$	$\sum_{i=1}^{k} n_i(x_i-\mu)^2$

On ajoute une colonne supplémentaire pour calculer, pour chaque valeur de la variable, les produits de l'effectif par le carré de la différence à la moyenne μ. On utilise ces résultats pour compléter les calculs.

La variance est $\sigma^2 = \frac{\sum_{i=1}^{k} n_i(x_i-\mu)^2}{n} = \frac{519,9615}{104} = 4,9996\ ...$

L'écart-type est $\sigma = \sqrt{\sigma^2} = \sqrt{4,9996\ ...} = 2,235\ ...$

Le coefficient de variation est $CV = \frac{\sigma}{\mu} \times 100 = \frac{2,235\ ...}{4,98} \times 100 = 44,8\ ...$

c) On ajoute des colonnes au tableau de distribution pour calculer les fréquences relatives f_i, les milieux de classe m_i, la somme des produits $f_i m_i$, qui est la moyenne, et la somme des produits $f_i(m_i - \mu)^2$, qui est la variance.

Classes $[x_i; x_{i+1}[$	Effectifs n_i	Fréquences relatives f_i	Milieux m_i	Produits $f_i m_i$	Produits $f_i(m_i-\mu)^2$
[24; 28[	4	0,020	26	0,52	3,165
[28; 32[	9	0,045	30	1,35	3,313
[32; 36[	42	0,210	34	7,14	4,405
[36; 40[	68	0,340	38	12,92	0,114
[40; 44[	56	0,280	42	11,76	3,275
[44; 48[	18	0,090	46	4,14	4,955
[48; 52[	3	0,015	50	0,75	1,956
Total	200	1,000		38,58	21,1836
	n			μ	σ^2

La moyenne est : $\mu = \sum_{i=1}^{k} f_i m_i = 38,58.$

La variance est $\sigma^2 = \sum_{i=1}^{k} f_i(m_i-\mu)^2 = 21,1836$, l'écart-type est $\sigma = \sqrt{\sigma^2} = \sqrt{21,1836} = 4,602\ ...$

Le coefficient de variation est $CV = \frac{\sigma}{\mu} \times 100 = \frac{4,602\ ...}{38,58} \times 100 = 11,9\ ...$

d) On ajoute des colonnes au tableau de distribution pour calculer les fréquences relatives f_i, les milieux de classe m_i, la somme des produits $f_i m_i$, qui est la moyenne, et la somme des produits $f_i(m_i - \mu)^2$, qui est la variance.

Classes $[x_i; x_{i+1}[$	Effectifs n_i	Fréquences relatives f_i	Milieux m_i	Produits $f_i m_i$	Produits $f_i(m_i - \mu)^2$
[24; 28[	39	0,195	26	5,07	30,860
[28; 32[	42	0,210	30	6,30	15,459
[32; 36[	12	0,060	34	2,04	1,259
[36; 40[	5	0,025	38	0,95	0,008
[40; 44[	14	0,070	42	2,94	0,819
[44; 48[	36	0,180	46	8,28	9,910
[48; 52[	52	0,260	50	13,00	33,908
Total	200	1,000		38,58	92,2236
	n			μ	σ^2

La moyenne est : $\mu = \sum_{i=1}^{k} f_i m_i = 38{,}58$.

La variance est $\sigma^2 = \sum_{i=1}^{k} f_i (m_i - \mu)^2 = 92{,}2236$, l'écart-type est $\sigma = \sqrt{\sigma^2} = \sqrt{92{,}2236} = 9{,}603\ ...$

Le coefficient de variation est $CV = \frac{\sigma}{\mu} \times 100 = \frac{9{,}603\ ...}{38{,}58} \times 100 = 24{,}8\ ...$

e) Dans ces deux distributions, les valeurs que peuvent prendre les variables sont les mêmes et l'étendue est la même. Les moyennes sont très légèrement distinctes, cependant les écart-types et les coefficients de variation indiquent que les données sont moins dispersées dans la distribution a.

f) Dans ces deux distributions, les valeurs que peuvent prendre les variables sont les mêmes et l'étendue est la même et la moyenne est la même. Cependant, les écart-types et les coefficients de variation indiquent que la distribution c est une plus homogène.