

Exercices de Biostatistique STA109
Fiche TD1: Statistique descriptive univariée

Exercice 1

Spécifier pour chacune de ces variables si elle est qualitative, quantitative, continue, discrète, ordinale ou nominale. Proposer des modalités adéquates pour chacune de ces variables.

Nombre d'enfants dans une famille => variable **quantitative discrète**

Couleur des yeux => variable **qualitative nominale**

modalités: marrons, verts, bleus, autre

Catégorie socioprofessionnelle => variable **qualitative nominale**

modalités: agriculteur, artisan, commerçants, cadres, professions libérales, profession intermédiaire, employés, ouvriers, autre

Département de naissance => variable **qualitative nominale**

Niveau de scolarité => variable **qualitative ordinale**

Modalités: primaire, secondaire, études supérieures, autre

Revenu => variable **quantitative continue**

Poids => variable **quantitative continue**

Sexe => variable **qualitative nominale**

modalités: féminin, masculin

Age => variable **quantitative continue**

Langue maternelle => variable **qualitative nominale**

modalités: français, anglais, espagnol, autre

Type de voiture => variable **qualitative ordinale**

modalités: petite, familiale, sportive, luxe, autre

Taille => variable **quantitative continue**

Nombre de grains de beauté sur la peau => variable **quantitative discrète**

Taille de grains de beauté => variable **quantitative continue**

Type de tumeur => variable **qualitative ordinale**

modalités: bénigne, maligne

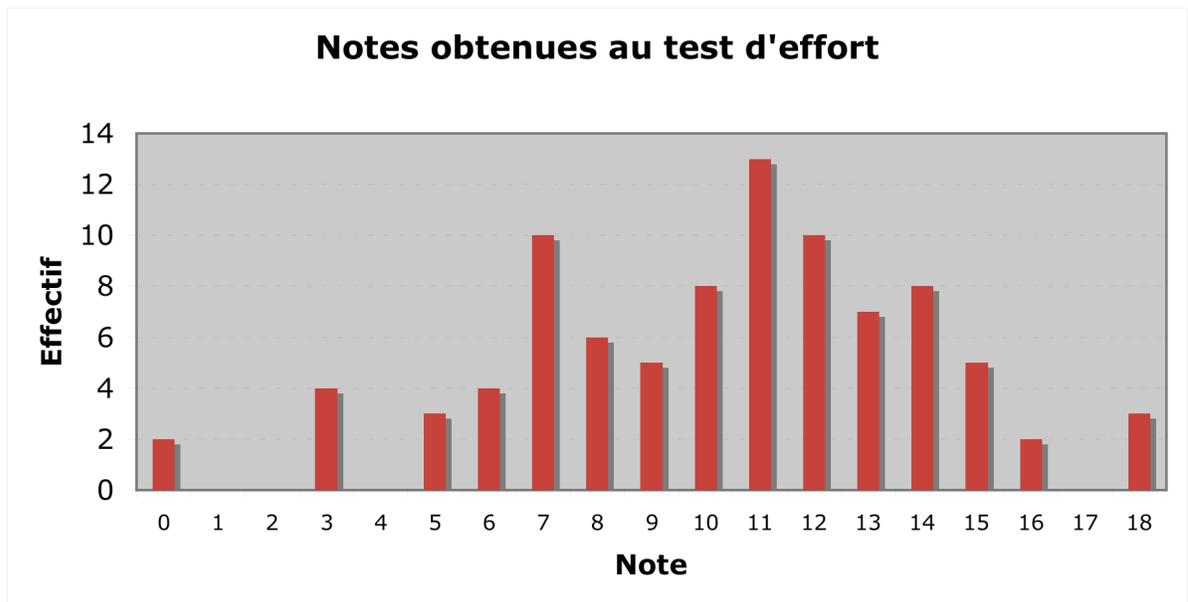
Spécifier pour chacune de ces variables si elle est qualitative, quantitative, continue, discrète, ordinale ou nominale. Proposer des modalités adéquates pour chacune de ces variables.

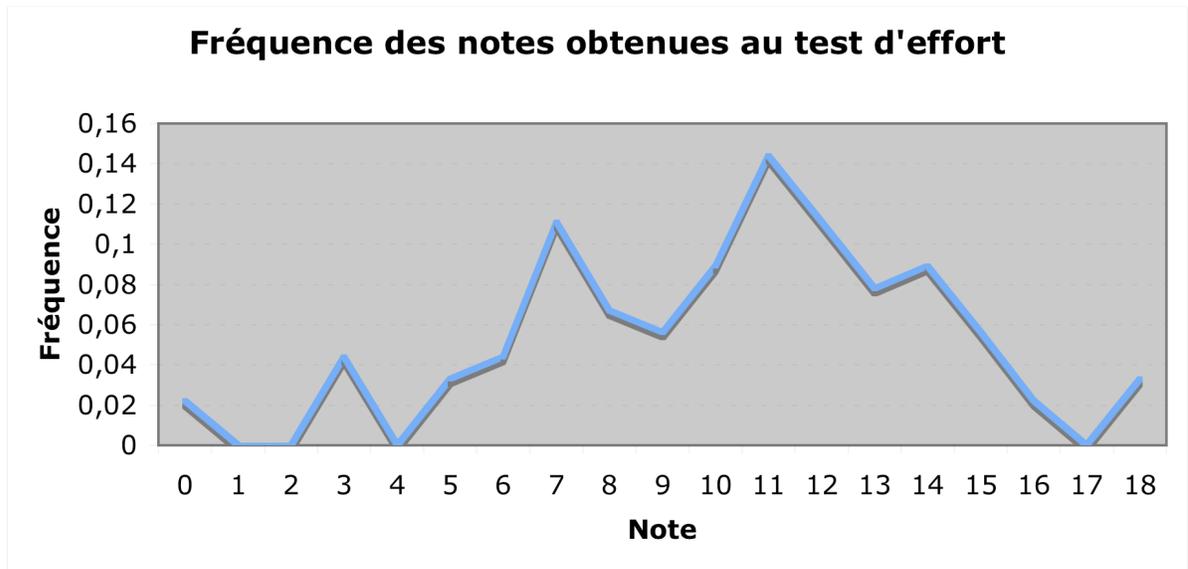
Exercice 2

Les notes obtenues par 90 candidats à la suite d'un test d'effort sont résumées dans le tableau suivant ou n_i désigne le nombre de candidats :

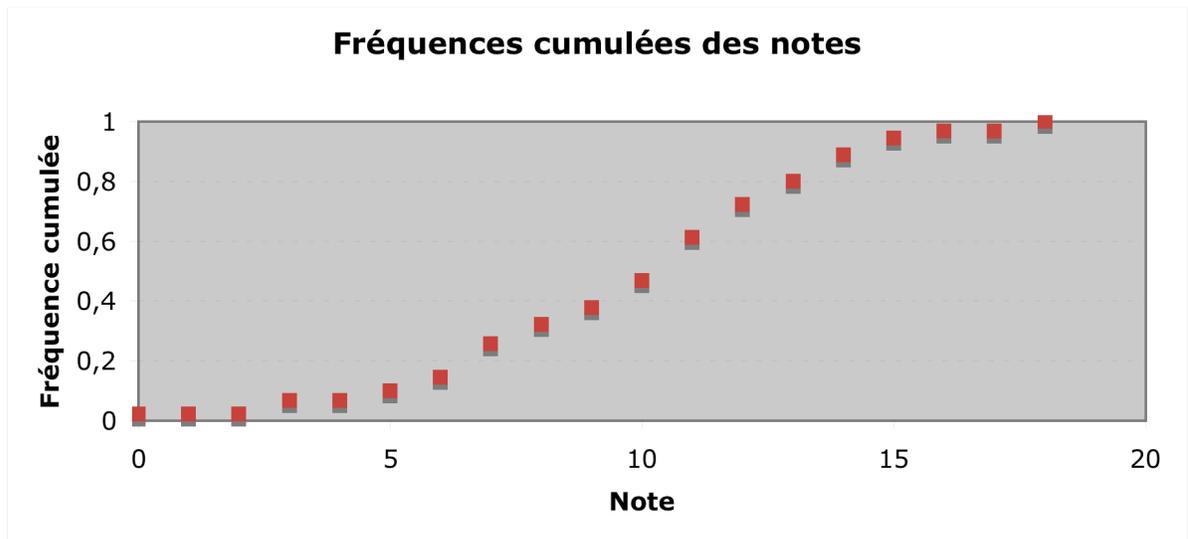
Note	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
n_i	2	0	0	4	0	3	4	10	6	5	8	13	10	7	8	5	2	0	3
N_i	2	2	2	6	6	9	13	23	29	34	42	55	65	72	80	85	87	87	90
$f_i \%$	2.2	0	0	4.4	0	3.3	4.4	11.1	6.7	5.6	8.9	14.4	11.1	7.8	8.9	5.6	2.2	0	3.3
$F_i \%$	2.2	2.2	2.2	6.7	6.7	10	14.4	25.6	32.2	37.8	46.7	61.1	72.2	80	88.9	94.4	96.7	96.7	100

1) Tracer le diagramme en bâtons et le polygone de fréquences.





2) Tracer la courbe de fréquences cumulées.



3) Déterminer le mode et la moyenne de cette distribution.

=> Mode = 11, Moyenne = 10,2

4) Déterminer la médiane, le 1er et le 3ème quartiles, graphiquement et par le calcul.

=> d'après le tableau des fréquences cumulées, Médiane=11, Q1=7, Q3=13

Par le calcul:

Médiane: n étant pair on détermine un intervalle médian $[X_{(n/2)}; X_{(n/2+1)}]=[X_{45}; X_{46}]$

D'après le tableau des fréquences cumulées, cela correspond à la valeur **11**.

Premier quartile:

$$n\alpha = 90 \cdot 0,25 = 22,5$$

Comme ce n'est pas un entier on prend $[90 \cdot 0,25 + 1] = 23$

D'après le tableau des fréquences cumulées: $Q_1 = 7$

Troisième quartile:

$$n\alpha = 90 \cdot 0,75 = 67,5$$

Comme ce n'est pas un entier on prend $[90 \cdot 0,75 + 1] = 68$

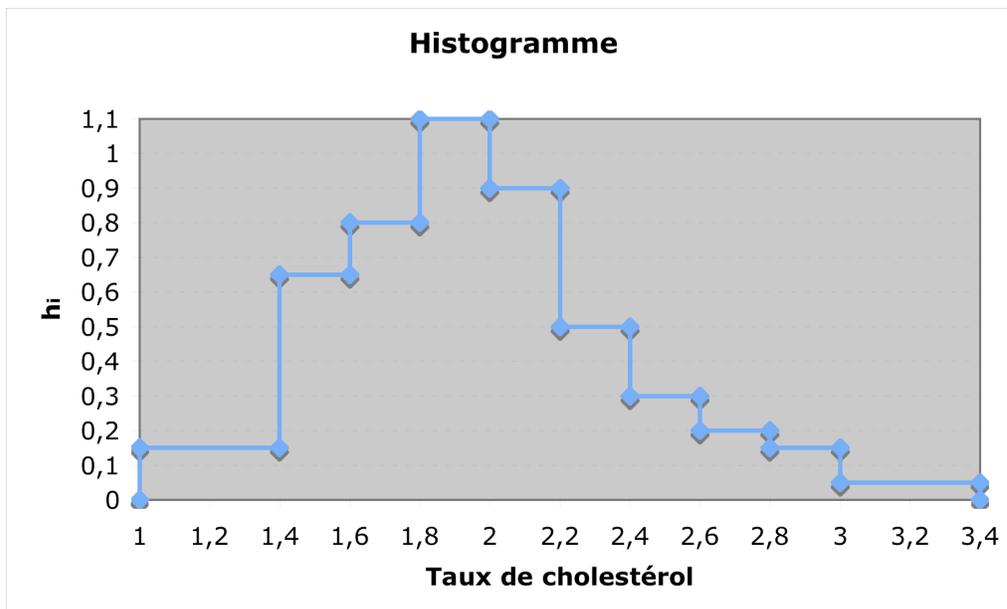
D'après le tableau des fréquences cumulées: $Q_3 = 13$

Exercice 3

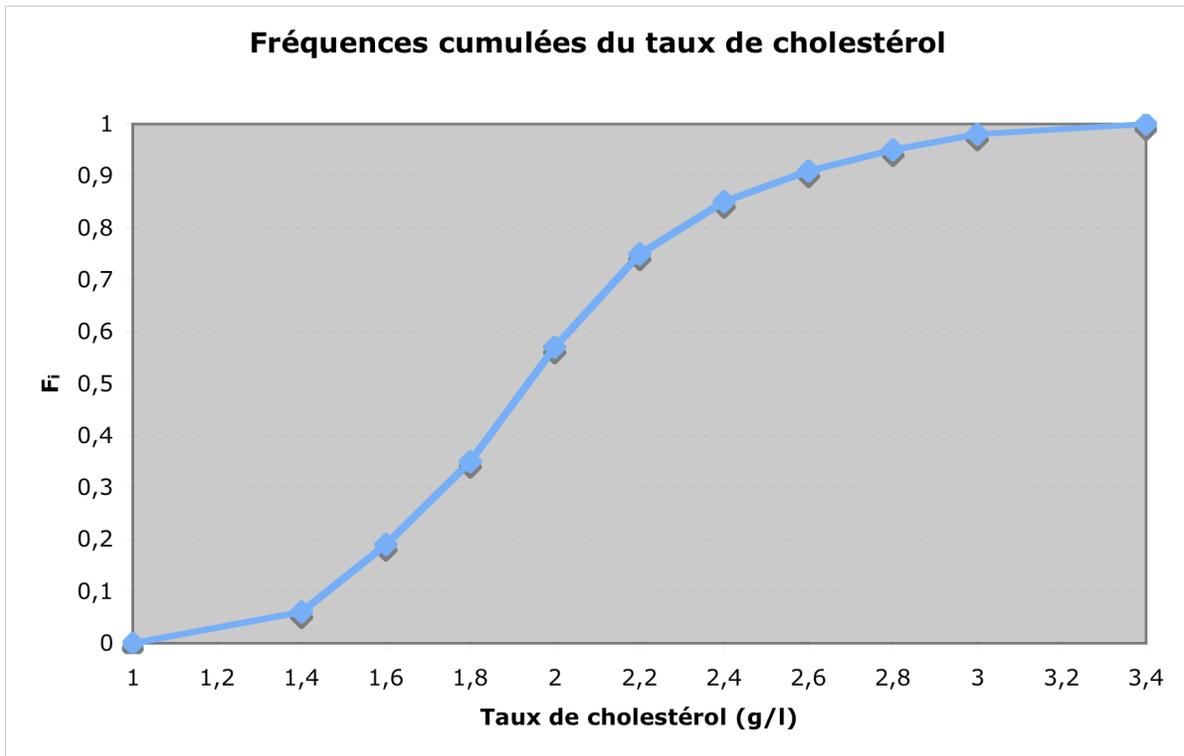
L'étude du taux de cholestérol sur un échantillon de 100 personnes a donné les résultats suivants:

Taux de cholestérol	Centre de la classe	effectif	Effectif cumulé	fréquence	Densité de fréquence	Fréquence cumulée
[1.0; 1.4[1.2	6	6	0.06	0.15	0.06
[1.4; 1.6[1.5	13	19	0.13	0.65	0.19
[1.6; 1.8[1.7	16	35	0.16	0.80	0.35
[1.8; 2.0[1.9	22	57	0.22	1.10	0.57
[2.0; 2.2[2.1	18	75	0.18	0.90	0.75
[2.2; 2.4[2.3	10	85	0.10	0.50	0.85
[2.4; 2.6[2.5	6	91	0.06	0.30	0.91
[2.6; 2.8[2.7	4	95	0.04	0.20	0.95
[2.8; 3.0[2.9	3	98	0.03	0.15	0.98
[3.0; 3.4[3.2	2	100	0.02	0.05	1

1) Tracer l'histogramme.



2) Tracer la courbe de fréquences cumulées



3) Déterminer le mode et la moyenne de cette distribution.

Pour le calcul du mode on applique la formule:

$$Mo = L + ai * (\Delta i / (\Delta i + \Delta s))$$

La classe modale est celle qui a la plus grande densité de fréquence soit: [1,8 ; 2[

$$L = 1,8$$

$$ai = 0,2$$

$$\Delta i = 22 - 16 = 6$$

$$\Delta s = 22 - 18 = 4$$

$$Mo = 1,8 + 0,2 * (6 / (6 + 4)) = \mathbf{1,92 \text{ g/l}}$$

Pour le calcul de la moyenne on utilise le centre des classes:

$$Moy = 1/100 * (6 * 1,2 + 13 * 1,5 + 16 * 1,7 + 22 * 1,9 + 18 * 2,1 + 10 * 2,3 + 6 * 2,5 + 4 * 2,7 + 3 * 2,9 + 2 * 3,2)$$

$$Moy = \mathbf{1,97 \text{ g/l}}$$

4) Déterminer la médiane, le 1er et le 3ème quartiles, graphiquement et par le calcul.

Formule de calcul de la médiane:

$$Mé = bi + (bi+1 - bi) * ((n/2) - Ni-1) / ni$$

La classe médiane est: [1,8 ; 2[, donc:

$$Mé = 1,8 + 0,2 * ((100/2) - 35) / 22 = \mathbf{1,936 \text{ g/l}}$$

Calcul du premier quartile:

La classe où se trouve le premier quartile est: [1,6 ; 1,8[, donc:

$$Q_1 = 1,6 + 0,2 * ((100/4)-19)/16 = \mathbf{1,675 \text{ g/l}}$$

Calcul du troisième quartile:

La classe où se trouve le troisième quartile est: [2 ; 2,2[, donc:

$$Q_3 = 2 + 0,2 * ((100*3/4)-57)/18 = \mathbf{2,2 \text{ g/l}}$$

Exercice 4

Les pesées de 50 nouveaux-nés dans une maternité ont permis d'établir le tableau suivant :

Poids des nouveaux-nés	Centre de la classe	effectif	Effectif cumulé	fréquence	Densité de fréquence	Fréquence cumulée
[2.0; 2.4[2.2	6	6	0.125	0.3125	0.125
[2.4; 2.8[2.6	10	16	0.2083	0.5208	0.3333
[2.8; 3.4[3.1	18	34	0.375	0.625	0.7083
[3.4; 3.8[3.6	10	44	0.2083	0.5208	0.9167
[3.8; 4.2[4	4	48	0.0833	0.2083	1

On suppose que dans chaque classe, les poids sont uniformément répartis entre les nouveaux-nés correspondants.

1) Déterminer le mode de cette distribution.

La classe modale est la classe [2.8; 3.4[qui a la plus grande densité de fréquence.

Pour calculer le mode, on utilise la formule avec les densités de fréquence

$$Mo = L + ai * (\Delta i / (\Delta i + \Delta s))$$

$$L = 2,8$$

$$ai = 0,6$$

$$\Delta i = 0,625 - 0,5208 = 0,1042$$

$$\Delta s = 0,625 - 0,5208 = 0,1042$$

$$Mo = 2,8 + 0,6 * (0,1042 / (0,1042 + 0,1042)) = \mathbf{3,1 \text{ kg}}$$

2) Calculer les premier, second et troisième quartiles.

Formule de calcul de la médiane:

$$Mé = b_i + (b_{i+1} - b_i) * ((n/2) - N_{i-1}) / n_i$$

La classe médiane est: [2,8 ; 3,4[, donc:

$$Mé = 2,8 + 0,6 * ((48/2) - 16) / 18 = \mathbf{3,07 \text{ kg}}$$

Calcul du premier quartile:

La classe où se trouve le premier quartile est: [2,4 ; 2,8[, donc:

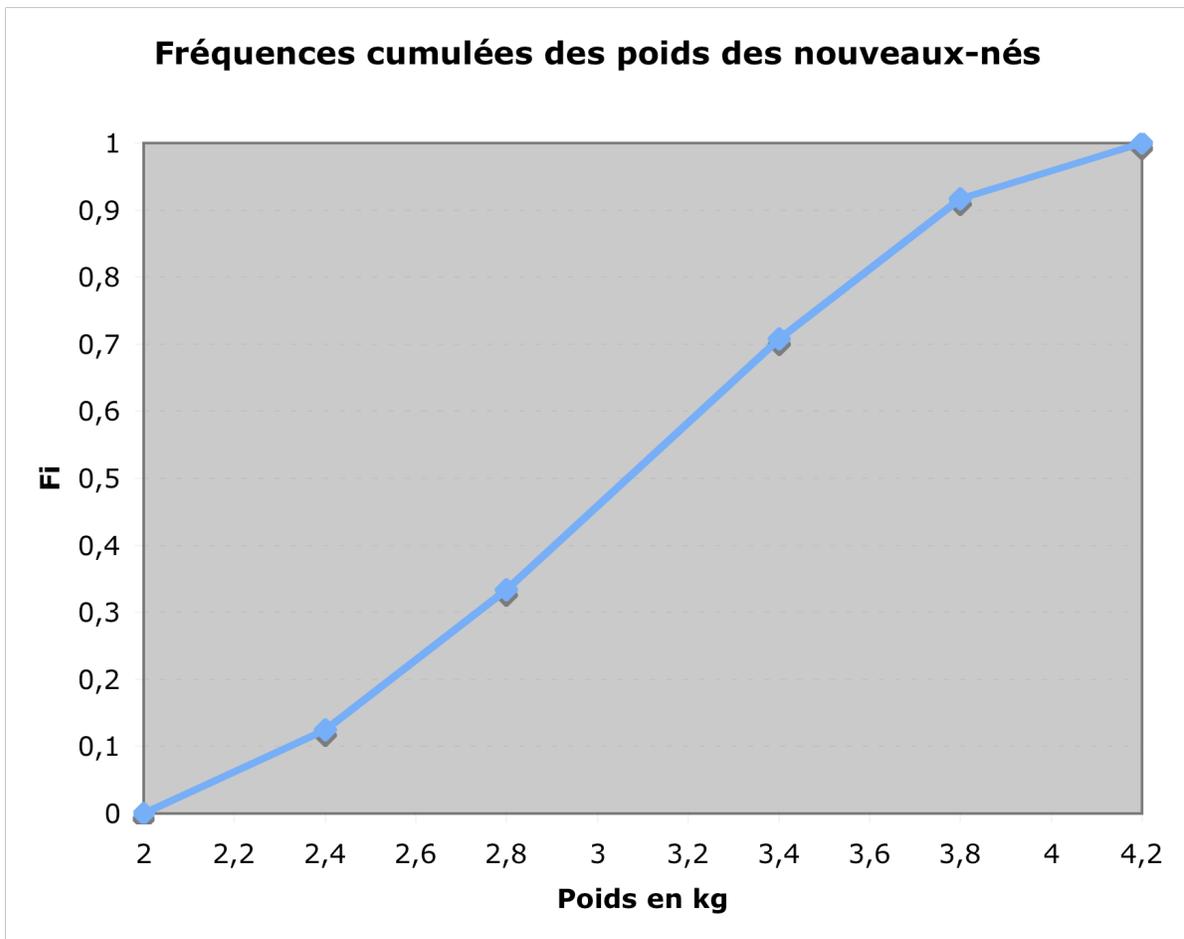
$$Q_1 = 2,4 + 0,4 * ((48/4)-6)/10 = \mathbf{2,64 \text{ kg}}$$

Calcul du troisième quartile:

La classe où se trouve le troisième quartile est: [3,4 ; 3,8[, donc:

$$Q_3 = 3,4 + 0,4 * ((48*3/4)-34)/10 = \mathbf{3,48 \text{ kg}}$$

3) Retrouver ces résultats à partir du diagramme des fréquences cumulées.



4) Déterminer le quatrième centile.

La classe où se trouve le quatrième centile (0,04) est: [2 ; 2,4[, donc:

$$C_4 = 2 + 0,4 * ((48*0,04)-0)/6 = \mathbf{2,13 \text{ kg}}$$