Cours de Statistiques - Données Continues

1. Définitions

Données continues : Ce sont des données qui peuvent prendre une infinité de valeurs dans un intervalle donné (par exemple, les poids, les tailles, les durées).

Une **classe** regroupe des données dans un intervalle $[a; b[$.

La **fréquence** (f_i) est calculée par :

$$f_i = \frac{n_i}{N}$$

2. Organisation des données dans un tableau

Pour analyser un ensemble de données continues, nous utilisons un tableau regroupant les classes, les centres des classes ($c_i = \frac{a_i+b_i}{2}$), les effectifs (n_i), et les fréquences.

44% des élèves ont eu entre 10 et 30

Exemple de tableau : résultats des examens sur 50

Classe $[a_i; b_i[$	c_i (x_i)	n_i	f_i	f_{ic}	$c_i \times n_i$ (x_i)	$n_i \times (c_i - \bar{x})^2$
$i=1$ $[a_1, b_1[$ [10 ; 20[	15	16	0,16	0,16	240	$16 \cdot (15-30,6)^2 = 3893,76$
$i=2$ $[a_2, b_2]$ [20 ; 30[	25	28	0,28	0,44	700	878,08
$i=3$ [30 ; 40[	35	40	0,4	0,84	1400	774,4
$i=4$ [40 ; 50[	45	16	0,16	1	720	3317,76
Total	-	100		-	3060	8864

100 $= n, N$; $\sum n_i$

3060 $= \sum c_i . n_i$

8864 : $\sum n_i (c_i - \bar{x})^2$

$$\bar{x} = \frac{\sum c_i n_i}{n} = \frac{3060}{100} = 30,60 \text{ (estimation)}$$

4. Questions à résoudre

1. Complétez le tableau.
2. Calculez la moyenne $\bar{x}$:

$$\bar{x} = \frac{\sum c_i \times n_i}{\sum n_i}$$

3. Calculez l'écart-type :

$$\sigma = \sqrt{\frac{\sum n_i \times (c_i - \bar{x})^2}{\sum n_i}} = \sqrt{\frac{8864}{100}}$$

$$= \sqrt{88,64} \approx 9,4$$

4. Calculez l'étendue :

$$\text{Étendue} = \text{Valeur max} - \text{Valeur min}$$

$$= 50 - 10 = 40$$

5. Interpréter les éléments suivants du tableau ou des valeurs calculées

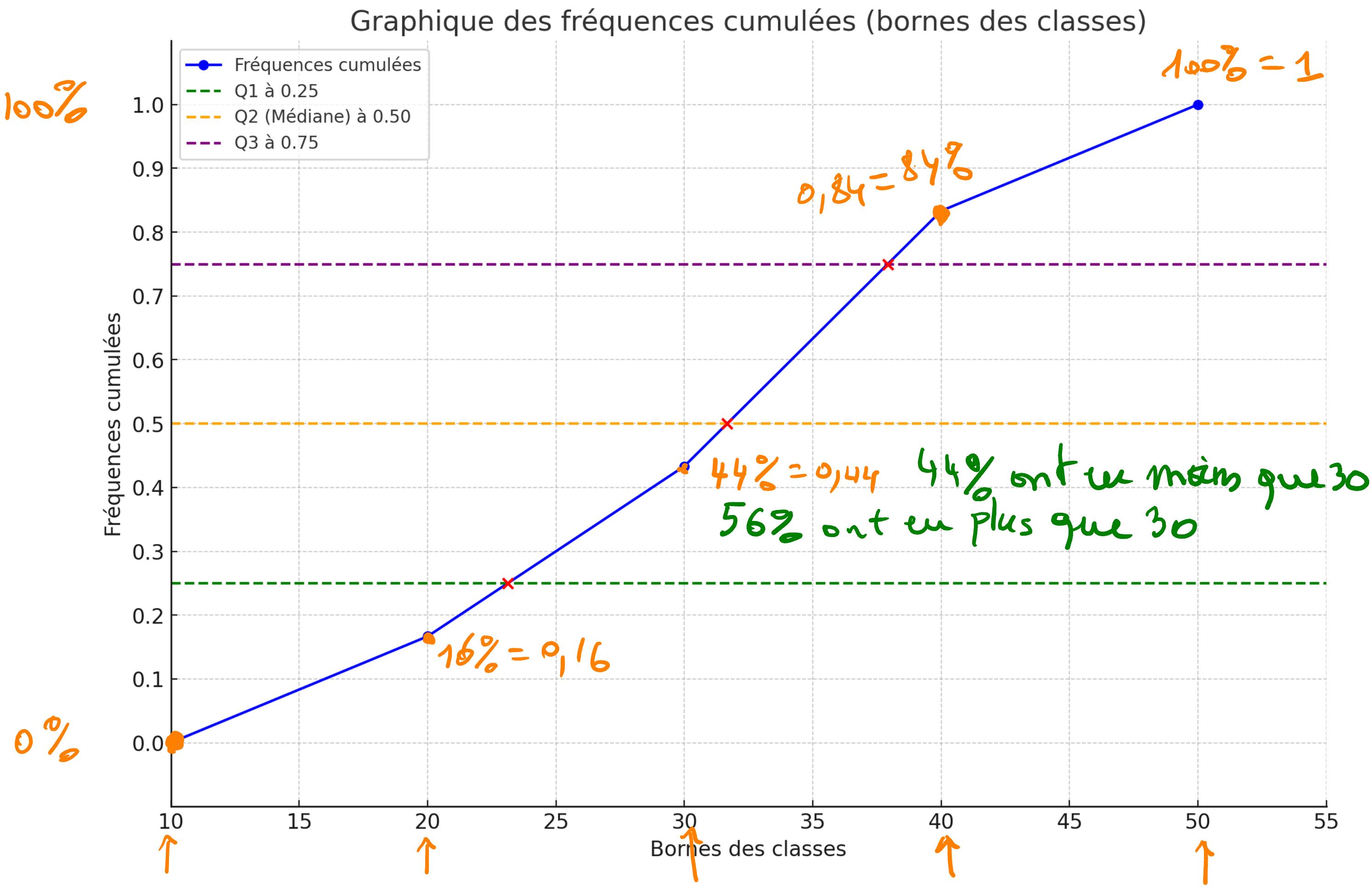
étendue (40) : toutes les notes sont dans un intervalle de 40 points (entre 10 et 50)

σ (9,4) : l'écart type : • c'est un écart "moyen" entre les notes et la moyenne

→ • (mesure de la dispersion des notes $\underbrace{\text{p/r}}_{\text{par rapport}}$ à la moyenne)

$n_3 = 40$: le nombre de notes entre 30 et 40

5. Diagramme des fréquences cumulées (variable continue)



6. Exercice : analyse statistiquement de la même façon les données suivantes :

Taille des individus dans une population de mammifères

Classe $[a_i; b_i[$	c_i	n_i	f_i	f_{ic}	$c_i \times n_i$	$n_i \times (c_i - \bar{x})^2$
[30 ; 50[	40	50	0,25	0,250	2000	$50 \times (40 - 65,5)^2 = 32512,5$
[50 ; 70[	60	75	0,375	0,625	4500	$75 \times (60 - 65,5)^2 = 2268,75$
[70 ; 90[	80	45	0,225	0,85	3600	$45 \times (80 - 65,5)^2 = 9461,25$
[90 ; 110[	100	30	0,15	1	3000	$30 \times (100 - 65,5)^2 = 35707,5$
Total	-	200	1	-	13100	79950

Σn_i $\quad$ $\Sigma c_i \times n_i$

$$\bar{x} = \frac{13100}{200} = 65,5$$

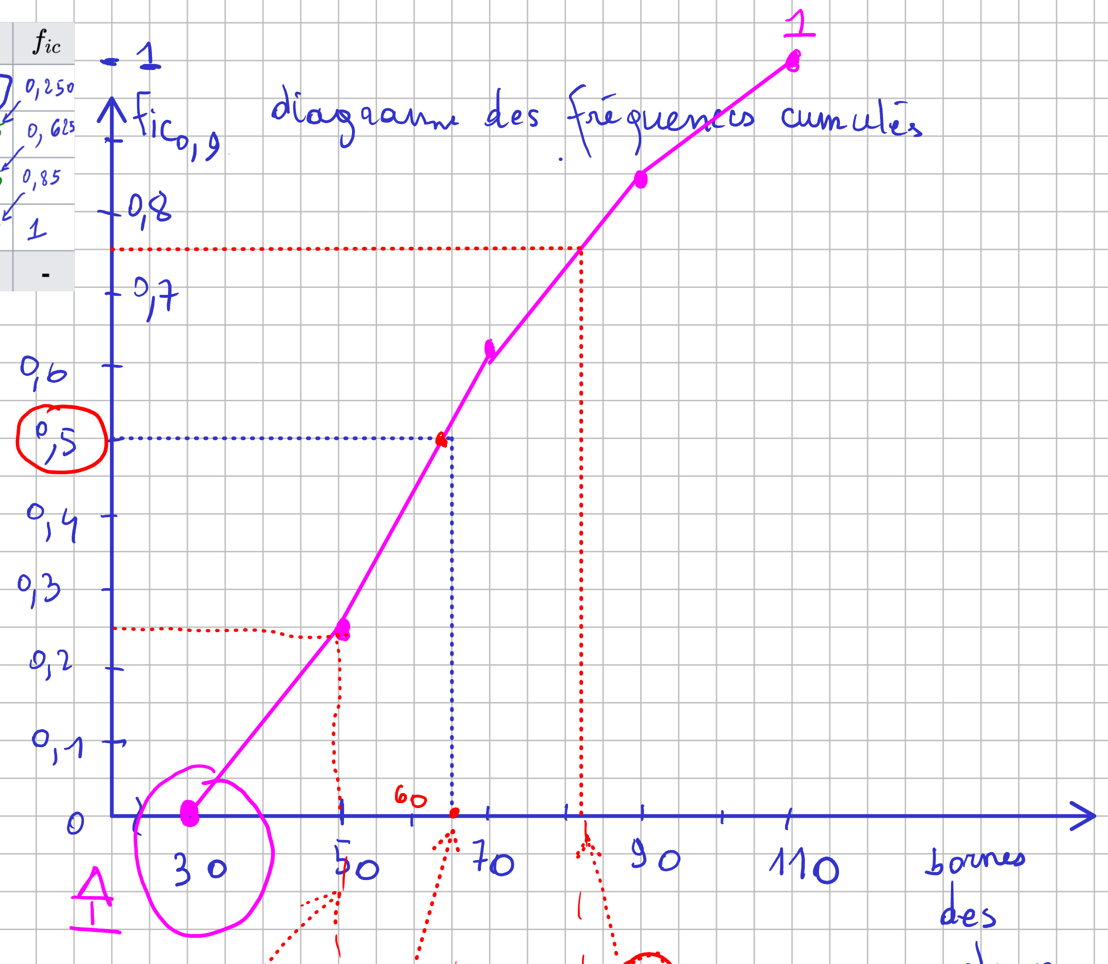
$$\frac{\Sigma n_i (c_i - \bar{x})^2}{\Sigma n_i}$$

$$\text{Variance} = \sigma^2 = \frac{79950}{200} = 399,75$$

$$\text{écart type } \sigma = \sqrt{\sigma^2} = \sqrt{399,75} \cong 20.$$

Classe $[a_i;b_i[$	c_i	n_i	f_i	f_{ic}
[30 ; 50[	40	50	0,25	0,250
[50 ; 70[	60	75	0,375	0,625
[70 ; 90[	80	45	0,225	0,85
[90 ; 110[	100	30	0,15	1
Total	-	200	1	-

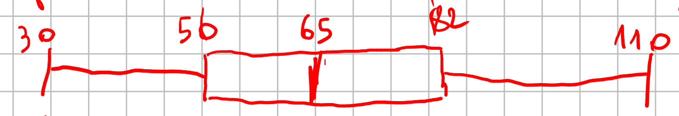
à 0



Q_1 1er quartile
25% des données sont inférieures à 50

médiane $Q_2 \simeq 65$

$Q_3 \simeq 82$ = 3ème quartile
75% des données sont inférieures à Q_3



Boîte à moustaches

écart interquartile $= Q_3 - Q_1$
$\simeq 82 - 50 = 32$

50% des données sont entre 50 et 82