

TD 5 : STATISTIQUES DESCRIPTIVES : PARAMÈTRES DE POSITION

MOYENNE ET MÉDIANE

Ex 1. Le tableau suivant rapporte les résultats d'une enquête recensant le nombre d'enfants par foyer dans un petit village de province :

| | | | | | | | |
|---------------|----|----|----|----|----|---|-------|
| Nb. d'enfants | 0 | 1 | 2 | 3 | 4 | 5 | Total |
| Effectif | 52 | 45 | 41 | 20 | 15 | 5 | 178 |

Calculer la moyenne et la médiane.

Solution. a. La moyenne est de :

$$\frac{1}{178} (52 \times 0 + 45 \times 1 + 41 \times 2 + 20 \times 3 + 15 \times 4 + 5 \times 5) = \frac{272}{178} \simeq 1,53$$

enfants par foyer. De plus, la moitié de l'effectif vaut $178/2 = 89$. On remarque que $52 < 89$ et que $52 + 41 > 89$, donc la médiane est de 1 enfant par foyer.

Ex 2. L'âge des abonnés du magazine "Tattoo & Piercing " est résumé dans le tableau suivant :

| | | | | |
|-----------|---------|---------|---------|---------|
| Âge | [10,20[| [20,30[| [30,50[| [50,90[|
| Fréquence | 27 % | 28 % | 30 % | 15% |

Calculer une valeur approchée de la moyenne. Que pouvez-vous dire de la médiane?

Solution. a. Pour approcher la moyenne, on va considérer pour simplifier que toute la population d'une classe est réduite à la valeur centrale de cette classe :

| | | | | |
|-----------|------|------|------|-----|
| Âge | 15 | 25 | 40 | 70 |
| Fréquence | 27 % | 28 % | 30 % | 15% |

L'âge moyen vaut alors :

$$0,27 \times 15 + 0,28 \times 25 + 0,30 \times 40 + 0,15 \times 70 = 33,55$$

soit 33,55 ans. Au passage, dans un calcul de moyenne (et seulement dans un calcul de moyenne!), l'hypothèse simplificatrice ci-dessus est équivalente à l'hypothèse plus sophistiquée selon laquelle la population d'une classe est uniformément répartie dans cette classe.

On a $27\% < 50\%$ et $27+28\% = 55\% > 50\%$, donc la médiane appartient à la classe $[20,30[$. Comme 55% est proche de 50%, la médiane est vraisemblablement plus proche de 30 ans. Pour en calculer une valeur plus précise (toujours sous l'hypothèse que la population est uniformément répartie dans chaque classe), il faut faire des calculs plus sophistiqués, ou s'appuyer sur un diagramme des fréquences cumulées comme dans l'exercice suivant. Ce calcul plus précis, non demandé, donnerait une médiane d'environ 28,2 ans.

FRACTILES : QUARTILES, DÉCILES, CENTILES

Ex 3. Le tableau ci-dessous résume les résultats à un examen de Droit.

| | | | | | | | | | |
|---------------------|----|----|-----|-----|-----|-----|-----|-----|------|
| Note | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
| Fréquence | 2% | 2% | 6% | 10% | 25% | 15% | 17% | 15% | 8% |
| Fréquences cumulées | 2% | 4% | 10% | 20% | 45% | 60% | 77% | 92% | 100% |

- Quelle est la nature de la variable statistique "Note" ? Quel type de graphique choisiriez-vous pour la représenter ?
- Quelle est la note la plus basse ? La plus haute ?
- Calculer la médiane, les quartiles, le premier décile et le dernier décile.

Solution. a. Tant qu'il n'y a pas de demi-points (ou autres fractions), la note est une variable quantitative et discrète (et prenant un nombre assez faible de valeurs). Un diagramme en bâton serait ici tout indiqué.

b. La note minimale est de 6, la note maximale de 14.

c. Les fréquences cumulées dépassent 50% pour une note de 11. La médiane est donc de 11.

Les fréquences cumulées dépassent 25% pour une note de 10. Le premier quartile est donc de 10.

Le second quartile (correspondant à un seuil de 75%) est de 12.

Le premier décile (seuil de 10%) pose problème : on a exactement 10% de la population en-dessous de 8, et 90% au-dessus de 9. N'importe quelle valeur entre 8 et 9 convient ; la convention la plus courante consiste à prendre leur moyenne, soit 8,5. Le dernier décile, lui, est à 13.

Ex 4. Une hôtel lance une enquête sur sa clientèle. Les données concernant l'âge sont les suivantes :

| Âge | [18,35[| [35,50[| [50,65[| [65,83] |
|---------------------|---------|---------|---------|---------|
| Fréquence | 8% | 16% | 36% | 40% |
| Fréquences cumulées | | | | |

a. Compléter le tableau ci-dessus.

b. Tracer (en grand !) le diagramme des fréquences cumulées.

c. En vous aidant de ce diagramme : quelle proportion de clients a entre 20 et 40 ans ?

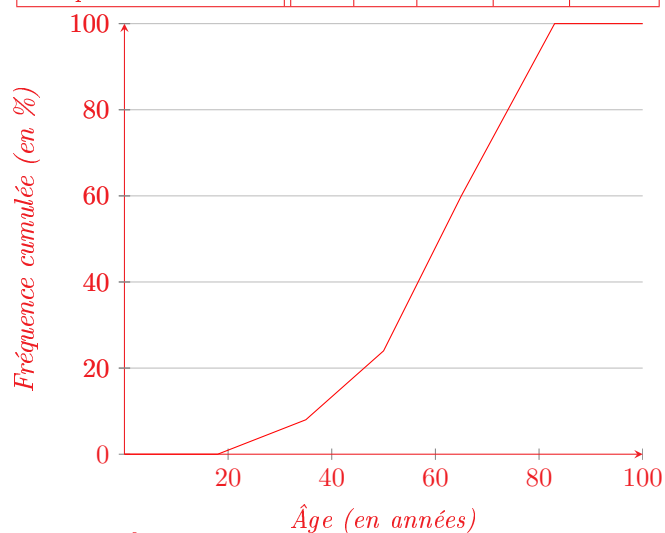
d. En vous aidant de ce diagramme : combien valent la médiane, les quartiles, le premier décile et le dernier décile ?

e. Tracer la boîte à moustaches correspondant à ce jeu de données.

Solution. a.

| Âge | 18 | 35 | 50 | 65 | 83 |
|---------------------|----|----|-----|-----|------|
| Fréquences cumulées | 0% | 8% | 24% | 60% | 100% |

b.



c. En lisant ce diagramme : la fréquence cumulée correspondant à 20 ans est d'environ 1%, et celle correspondant à 40 ans est d'environ 13%. Autrement dit, 1% des clients ont moins de 20 ans, et 13% des clients moins de 40 ans. En faisant la différence, on trouve qu'environ 12% des clients ont entre 20 et 40 ans (un calcul plus précis donnerait 12,4%).

d. La médiane est l'abscisse (âge) correspondant à une fréquence cumulée de 50%. Elle est d'environ 60 ans (un calcul précis donnerait 60,83 ans).

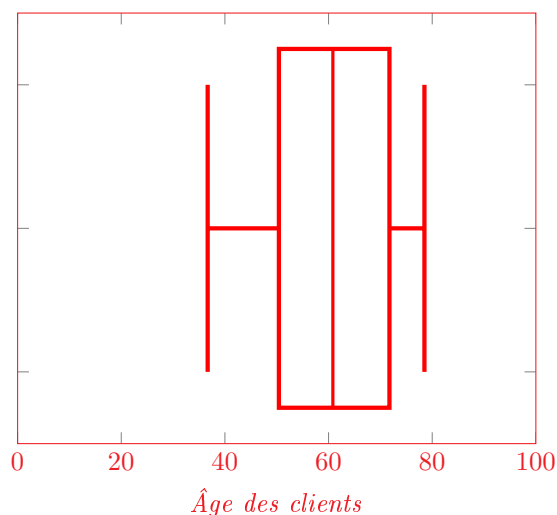
Le premier quartile est l'abscisse correspondant à une fréquence cumulée de 25%. Elle est d'environ 50 ans (un calcul précis donnerait 50,42 ans).

Le troisième quartile est l'abscisse correspondant à une fréquence cumulée de 75%. Elle est d'environ 70 ans (un calcul précis donnerait 71,75 ans).

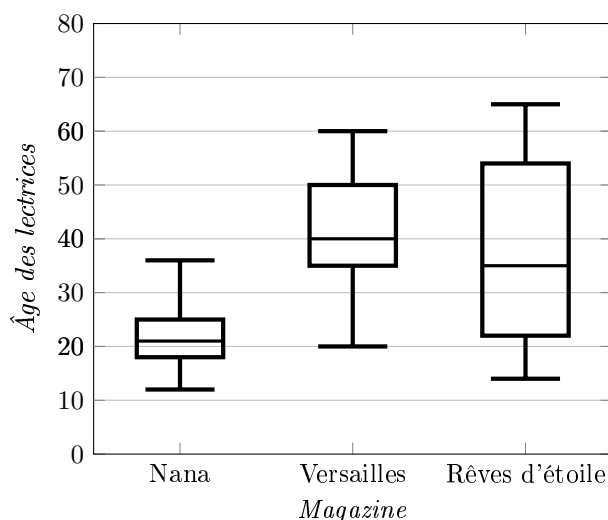
Le premier décile est l'abscisse correspondant à une fréquence cumulée de 10%. Elle est d'environ 37 ans (un calcul précis donnerait 36,68 ans).

Le dernier décile est l'abscisse correspondant à une fréquence cumulée de 90%. Elle est d'environ 80 ans (un calcul précis donnerait 78,5 ans).

e. En utilisant les données de la question précédente (médiane, quartiles, premier et dernier décile) :



Ex 5. Trois magasins de mode féminine (“Nana”, “Versailles” et “Rêves d’étoile”) ont réalisé une enquête sur l’âge de leurs lectrices. Les résultats sont résumés par le graphique suivant.



- a. Quels paramètres statistiques peut-on lire sur le graphique ?
- b. Compléter les phrases suivantes :
 - (1) Parmi les trois magasins, possède la clientèle la plus jeune.
 - (2) Parmi les trois magasins, possède la clientèle la plus hétérogène.
 - (3) 90% des lectrices de Nana ont moins de ans.
 - (4) % des lectrices de “Versailles” ont entre 35 et 50 ans.
 - (5) % des lectrices de “Versailles” ont plus de 35 ans.
 - (6) % des lectrices de “Versailles” ont entre 40 et 60 ans.
 - (7) Chez Nana, il y a autant de lectrices ayant entre 21 et 25 ans que de lectrices ayant entre et ans.

Solution. a. Pour chacun des trois magasins, on peut lire la médiane, les quartiles, ainsi que le premier et le dernier décile de l’âge du lectorat.

b.

- (1) Parmi les trois magasins, Nana possède la clientèle la plus jeune (car sa boîte est la plus basse).
- (2) Parmi les trois magasins, Rêve d’étoile possède la clientèle la plus hétérogène (car sa boîte est la plus large : l’écart entre les quartiles est particulièrement élevé, donc l’âge du lectorat est très étalé).

- (3) 90% des lectrices de Nana ont moins de 36 ans (extrémité supérieure de la moustache).
- (4) 50 % des lectrices de “Versailles” ont entre 35 et 50 ans (35 ans est le premier quartile, 50 ans le troisième quartile. L’intervalle entre les deux contient $75-25 = 50\%$ de la population).
- (5) 75 % des lectrices de “Versailles” ont plus de 35 ans (35 ans est le premier quartile. 25% de la population est en-dessous de ce seuil, donc 75% au-dessus).
- (6) 40 % des lectrices de “Versailles” ont entre 40 et 60 ans (40 ans est la médiane, 60 ans le dernier décile. L’intervalle entre les deux contient $90-50 = 40\%$ de la population).
- (7) Chez Nana, il y a autant de lectrices ayant entre 21 et 25 ans que de lectrices ayant entre 18 et 21 ans (il y a 25% du lectorat entre 21 et 25 ans, ce qui correspond aussi à l’écart entre le premier quartile et la médiane, donc entre 18 et 21 ans. Il y a aussi autant de lectrice en-dessous de 18 ans, et autant de lectrices au-dessus de 36 ans).

QUELQUE EXERCICE EN PLUS

Ex 6. On reprend les données de l’exercice précédent. On s’intéresse à l’âge des clientes du magazine “Versailles”. Compléter, à l’aide du diagramme en boîte ci-dessus, le tableau des fréquences (**non** cumulées) :

| | | | | | | |
|-----------|--------|---------|---------|---------|---------|----------|
| Âge | [0,20[| [20,35[| [35,40[| [40,50[| [50,60[| [60,100[|
| Fréquence | | | | | | |

Solution. a. Pour cela, on va passer par un tableau des fréquences cumulées :

| | | | | | | | |
|-----------|----|-----|-----|-----|-----|-----|------|
| Âge | 0 | 20 | 35 | 40 | 50 | 60 | 100 |
| Fréquence | 0% | 10% | 25% | 50% | 75% | 90% | 100% |

On fait l’hypothèse qu’aucune lectrice n’a plus de 100 ans. Toutes ces valeurs se lisent sur la boîte à moustaches (par exemple, la fréquence cumulée correspondant à 60 ans et de 75% car 60 ans est le troisième quartile). Ensuite, en faisant les différences entre fréquences cumulées successives, on retrouve les fréquences de chaque classe (et on peut même construire l’histogramme correspondant !) :

| | | | | | | |
|-----------|--------|---------|---------|---------|---------|----------|
| Âge | [0,20[| [20,35[| [35,40[| [40,50[| [50,60[| [60,100[|
| Fréquence | 10% | 15% | 25% | 25% | 15% | 10% |

Ex 7. Le premier ministre de Syldavie se félicite des résultats de sa politique sur le pouvoir d’achat. En effet, pendant son mandat, le salaire moyen annuel (indexé par l’inflation) des employés en CDI est passé de 28 800 à 29 600 euros et celui des employés en CDD de 25 100 à 26 200 euros. Un journaliste fait remarquer que, dans la même période, le pourcentage d’employés en CDD est passé de 5% à 35%.

Comment le salaire moyen des employés syldaves a-t-il évolué ?

Solution. a. Au début, 95% des employés ont un CDI et 5% un CDD. Le salaire moyen est donc de :

$$0,95 \times 28\,800 + 0,05 \times 25\,100 = 28\,615 \text{ euros.}$$

À la fin de son mandat, le salaire moyen s’élève à :

$$0,65 \times 29\,600 + 0,35 \times 26\,200 = 28\,410 \text{ euros.}$$

Ainsi, même si le salaire moyen dans chaque catégorie a augmenté, une partie de la population est passée dans la catégorie la plus pauvre, de telle sorte que le salaire moyen global a en fait diminué !