UE 22A

3

Statistiques

Un peu d'histoire

La science statistique semble exister dès la naissance des premières structures sociales. D'ailleurs, les premiers textes écrits retrouvés étaient des recensements du bétail, des informations sur son cours et des contrats divers.
On attribue souvent l'introduction du terme « statistique » au professeur **Achenwall**, qui aurait, en 1746, créé le mot Statistik, dérivé de la notion Staatskunde.
En revanche, l'étymologie du mot nous donne la définition suivante : « étude méthodique par des procédés numériques des faits sociaux qui définissent un état. ». Le nom désigne ensuite (1862) l'objet des statistiques : « ensemble de données numériques concernant une même catégorie de faits ».
Aujourd'hui, les plus gros consommateurs de statistiques sont les assureurs (risques d'accidents, de maladie des assurés), les médecins (épidémiologie), les démographes (populations et leur dynamique), les économistes (emploi, conjoncture économique), les météorologues...

Ce qu'il faut savoir

1. Définitions et vocabulaire des statistiques

Vocabulaire : la **population** est l'ensemble des individus sur lesquels porte l'étude statistique. Par exemple : une classe de CE2, les hommes, les habitants de la Réunion…
Le **caractère** ou **variable** d'une série statistique est une propriété étudiée sur chaque individu :

- lorsque le caractère ne prend que des valeurs ou **modalités** numériques, il est **quantitatif** :
  - **discret** s'il ne peut prendre que des valeurs isolées (notes, âge…) ;
  - **continu** dans le cas contraire (poids, taille…). Dans ce cas on effectue un regroupement des valeurs par **classes** ;
- sinon, on dit qu'il est **qualitatif** (couleur des yeux, sport pratiqué…) : les modalités ne sont pas des nombres.

À chaque valeur (classe) est associée un **effectif** n : c'est le nombre d'individus associés à cette valeur.
Faire des **statistiques**, c'est recueillir, organiser, synthétiser, représenter et exploiter des données dans un but de comparer, de prévoir, de constater…

DÉFINITION : Fréquence

On considère une série statistique à caractère quantitatif, dont les p valeurs sont données par : $x_1, x_2, \ldots, x_p$ d'effectifs associés $n_1, n_2, \ldots, n_p$ avec $n_1 + n_2 + ... + n_p = N$.
Pour chaque valeur on peut calculer une **fréquence** f_i grâce à la formule $\frac{n_i}{N}$: c'est un nombre compris entre 0 et 1.

L'ensemble des fréquences de toutes les valeurs du caractère s'appelle la distribution des fréquences.

Exemple Voici les notes sur 20 obtenues à une évaluation dans une classe de 30 élèves :
Série A - 2 – 3 – 3 – 4 – 5 – 6 – 6 – 7 – 7 – 7 – 8 – 8 – 8 – 8 – 8 – 9 – 9 – 9 – 9 – 9 – 9 – 10 – 10 – 11 – 11 – 11 – 13 – 13 – 15 – 16.
On obtient le tableau suivant avec des fréquences en pourcentages :

Notes	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Eff.	0	1	2	1	1	2	3	5	6	2	3	0	2	0	1	1	0	0	0
Fréq.	0	3	7	3	3	7	10	17	20	7	10	0	7	0	3	3	0	0	0

Remarque : on peut vérifier que la somme des fréquences est égale à 1 ou à 100 si on les exprime en pourcentages.

On peut également faire un regroupement par classes, ce qui rend l'étude moins précise, mais qui permet d'avoir une vision plus globale.

Exemple Pour la **série A**, on peut regrouper les données par classes d'amplitude 5 points :

Notes	[0 ; 5 [	[5 ; 10 [	[10 ; 15 [	[15 ; 20 [
Effectif	4	17	7	2
Fréquence	0,13	0,57	0,23	0,07

2. Caractéristiques de position

A. Le mode et la classe modale

DÉFINITION : Mode

Le **mode** d'une série statistique est la valeur du caractère qui a le plus grand effectif.
Si le caractère est un caractère quantitatif continu, on regroupe ses valeurs en classes.
La classe qui a le plus grand effectif est appelée **classe modale**.

REMARQUE : il peut y avoir plusieurs modes et plusieurs classes modales.

Exemple

- Dans la **série A**, le mode est égal à 9 (effectif 6) ;
- si on regroupe par classes d'amplitude 5 points, la classe modale est [5 ; 10 [(effectif 17).

B. Moyenne arithmétique

DÉFINITION : Moyenne

Soit la série statistique à caractère quantitatif dont les p valeurs sont données par x_1, x_2, ..., x_p d'effectifs n_1, n_2, ..., n_p avec $n_1 + n_2 + ... + n_p = N$.
La **moyenne arithmétique pondérée** de cette série est le nombre :

$$\overline{x} = \frac{n_1x_1 + n_2x_2 + ... + n_px_p}{n_1 + n_2 + ... + n_p} = \frac{1}{N}\sum_{i=1}^{p} n_ix_i.$$

Par la suite, on utilisera le terme **moyenne** pour désigner la moyenne arithmétique.

REMARQUE :

- Dans le cas où tous les n_i valent 1, la moyenne de la série statistique est le quotient de la somme de toutes les valeurs par l'effectif total : $\overline{x} = \dfrac{x_1 + x_2 + ... + x_n}{n}$.
- Lorsque la série est regroupée en classes, on calcule la moyenne en prenant pour valeur x_i le centre de chaque classe obtenu en faisant la moyenne des deux extrémités de la classe.

Exemple Pour la **série A**, on obtient :

- une moyenne $\overline{m} = \dfrac{1 \times 2 + 2 \times 3 + \cdots + 1 \times 16}{30} = \dfrac{254}{30} \approx 8,47$;
- si on regroupe par classes d'amplitude 5 points, une estimation de la moyenne est :
$\overline{m} = \dfrac{4 \times 2,5 + 17 \times 7,5 + 7 \times 12,5 + 2 \times 17,5}{30} = \dfrac{260}{30} \approx 8,67.$

La calculatrice permet de déterminer les éléments statistiques en allant dans le mode « stats » : 

Il suffit d'entrer les notes et les effectifs correspondants dans les listes 1 et 2, puis de choisir « 1-Var Stats » dans le menu statistique.

Ce qu'il faut savoir

PROPRIÉTÉ : Somme et multiplication

- Si on ajoute (respectivement soustrait) un même nombre k à toutes les valeurs d'une série, alors la moyenne de cette série se trouve augmentée (respectivement diminuée) de k.
- Si on multiplie (respectivement divise) par un même nombre (non nul pour la division) k toutes les valeurs d'une série, alors la moyenne de cette série se trouve multipliée (respectivement divisée) par k.

Exemple On considère la **série A** :

- si on ajoute $1,5$ points à chaque note de l'évaluation, la moyenne de classe devient : $\overline{m} = 8,47 + 1,5 = 9,97$.
- si on augmente chaque note de $10\,\%$, cela revient à multiplier chaque note par $1,1$: $\overline{m} = 8,47 \times 1,1 = 9,32$.

C. Médiane

DÉFINITION : Médiane

Soit la série statistique ordonnée dont les n valeurs sont $x_1 \leqslant x_2 \leqslant \cdots \leqslant x_n$. La **médiane** est un nombre M qui permet de diviser cette série en deux sous-groupes de même effectif.

- Si n est **impair**, M est la valeur de cette série dont le rang est $\dfrac{n+1}{2}$, notée $x_{\frac{n+1}{2}}$.
- Si n est **pair**, M appartient à l'intervalle fermé formé par les deux nombres situés « au milieu » de la série, à savoir $x_{\frac{n}{2}}$ et $x_{\frac{n}{2}+1}$.

Exemple

- La médiane de la série « 2 – 5 – 6 – 8 – 9 – 9 – 10 » est 8.
- La médiane de la série « 2 – 5 – 6 – 8 – 9 – 9 » est n'importe quel nombre situé entre 6 et 8 inclus.

Remarque : lorsque l'on a une grande série statistique et que l'on souhaite déterminer la médiane, on peut faire le tableau des effectif cumulés croissants (ECC).

Exemple Pour la **série A**, d'effectif 30, la médiane est un nombre situé entre la 15ème et la 16ème valeur qui sont 8 et 9, on peut prendre par exemple $M = 8,5$.

Notes	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Eff.	0	1	2	1	1	2	3	5	6	2	3	0	2	0	1	1	0	0	0
ECC	0	1	3	4	5	7	10	15	21	23	26	26	28	28	29	30	30	30	30

la 15ème valeur est 8 — la 16ème valeur est 9

D. Quartiles

DÉFINITION : Quartiles

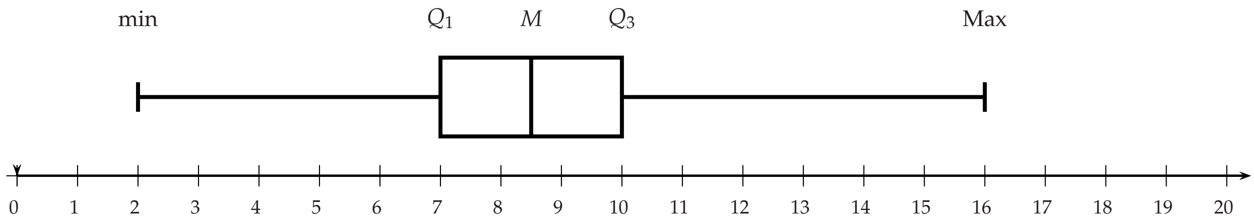
On appelle **quartiles** un triplet de réels $(Q_1; Q_2; Q_3)$ qui sépare la série en quatre groupes de même effectif. Le **premier** [respectivement le **troisième**] quartile est la plus petite valeur de la série telle que au moins 25 % [respectivement 75 %] des valeurs de la série lui soient inférieures ou égales. Q_2 correspond à la médiane de la série.

Exemple Pour la **série A**, $30 \times 0,25 = 7,5$ et $30 \times 0,75 = 22,5$ donc :

- Q_1 correspond à la $8^{\text{ème}}$ valeur : $Q_1 = 7$;
- $Q_2 = M = 8,5$;
- Q_3 correspond à la $23^{\text{ème}}$ valeur : $Q_3 = 10$.

REMARQUE : pour Q_1, on a $30 \times 0,25 = 7,5$, mais si on avait choisi la $7^{\text{ème}}$ valeur, le quota des 25 % n'aurait pas été atteint. Dans le cas d'une valeur décimale non entière, on prend systématiquement le nombre entier suivant.

On peut représenter ces données grâce à une « boîte à moustaches », ou **diagramme de Tuckey** (de son inventeur *John Tukey*, en 1977). Pour cela, il nous faut les valeurs suivantes : le minimum et le maximum ; les trois quartiles.



La boîte à moustaches résume quelques caractéristiques de position du caractère étudié et est utilisé principalement pour comparer un même caractère dans deux ou plusieurs populations de tailles différentes.

3. Caractéristiques de dispersion

DÉFINITION : Étendue

On appelle **étendue** d'une série discrète le réel égal à la différence entre la plus grande valeur et la plus petite valeur de la série.

Son principal mérite est d'exister, et de fournir une information sur la dispersion.

Exemple L'étendue de la **série A** est de $E = 16 - 2 = 14$.

DÉFINITION : Intervalle et écart inter-quartile

On appelle **intervalle inter-quartiles** l'intervalle $[\, Q_1 \,;\, Q_3 \,]$.
L'amplitude de cet intervalle est appelée **écart inter-quartiles**.

Exemple Dans la **série A**, l'intervalle inter-quartiles est $[\,7\,;\,10\,]$ dont l'écart vaut $10 - 7 = 3$.

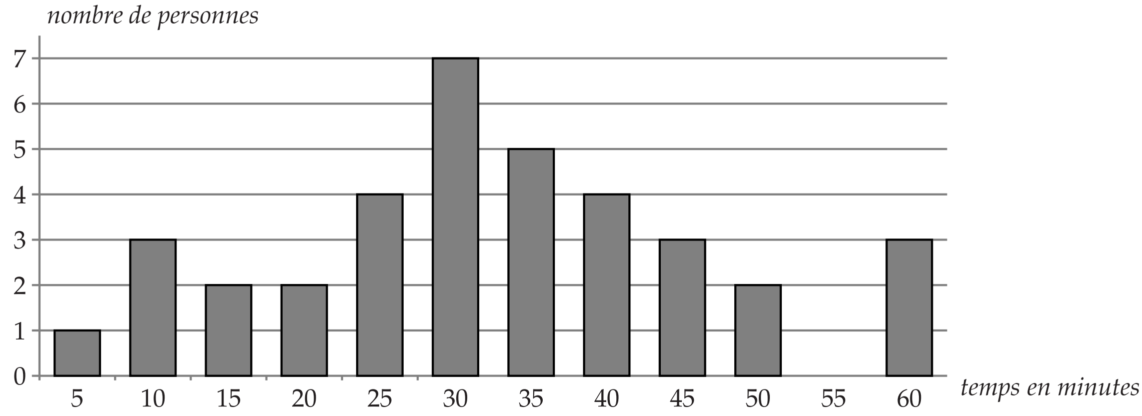
Pour s'entraîner

Maîtriser les bases avec MathenPoche

Classe	N°	Thème	Dans le cours
5e	D2	Statistiques	1.
4e	D2	Statistiques	2.
3e	N9	Statistiques et probabilités	2. et 3.
2nde	SP1	Statistiques descriptives	1., 2. et 3.

1 Quelle chance !

Le digramme en bâtons ci-dessous représente le temps de trajet journalier en minutes de 36 personnes travaillant dans l'entreprise kadubol.



On donnera les résultats numériques à 10^{-1} près sauf indication contraire.

1) Quels sont le caractère et la population étudiés ?

2) a) Construire le tableau d'effectifs et de fréquences récapitulant toutes ces valeurs.

b) Calculer l'étendue et préciser le mode de cette série.

c) Déterminer la moyenne, la médiane et les quartiles.

3) a) Construire un deuxième tableau en regroupant les effectifs par classes d'amplitude 15 minutes.

b) Calculer à nouveau la moyenne en utilisant la répartition par classes. Le résultat obtenu est-il le même que lors du calcul précédent ? Pourquoi ? Est-il plus fiable ?

2 Étude statistique d'un échantillon de notes

Le tableau suivant résume les résultats obtenus par une classe lors d'une évaluation.

Notes	3	5	6	7	8	9	10	11	12	13	14	17	18
Eff.	1	2	1	3	3	5	6	4	2	1	2	2	1

1) Combien y a-t-il d'élèves dans cette classe ? Déterminer l'étendue de la série de notes.

2) Quel est le pourcentage d'élèves ayant obtenu une note inférieure ou égale à 8 ?

3) Déterminer la moyenne, la médiane et les quartiles de la série de notes.

4) Cette évaluation était le quatrième de la période. Toutes les évaluations ont le même coefficient et jusqu'alors Bastien avait 9 de moyenne ; après ce devoir, il a 9,5 de moyenne. Quelle note a-t-il obtenue à ce devoir ?

3 Famille nombreuse

J'ai 7 enfants :

- l'âge modal est 5 ans ;
- Marie-Capucine a précisément l'âge médian, 7 ans ;
- les jumeaux ont l'âge moyen, 8 ans.

Mais quel âge a donc mon aîné ?

4 CRPE 2005 Grenoble

Voici la liste partielle des notes sur 20 obtenues par Luc et Julie aux six devoirs de mathématiques du premier trimestre :

Devoirs	1	2	3	4	5	6	Moy.
Notes de Luc	12	5	18	11	19		
Notes de Julie	20	15	4	9	x	y	12,5

1) **a)** Calculez la moyenne de Luc, si la note obtenue au sixième devoir est égale à la moyenne des cinq premiers.
b) Une meilleure note au devoir 6 aurait-elle permis à Luc d'obtenir une moyenne de 15 ?

2) La note y obtenue par Julie au devoir 6 a augmenté de 25 % par rapport à la note x qu'elle a obtenue au devoir 5.
a) Exprimez y en fonction de x.
b) Calculez x et y.

5 CRPE 2014 G1

Le cross du collège a eu lieu. 200 élèves de troisième ont franchi la ligne d'arrivée. Voici les indicateurs des performances réalisées en minutes.

Minimum	Premier quartile	Médiane	Troisième quartile	Moyenne	Étendue
12,5	14,8	15,7	16,3	15,4	4,2

Répondre aux questions suivantes en justifiant.

1) Quelle est la performance en minutes du dernier arrivé ?
2) Quelle est la somme des 200 performances en minutes ?
3) Ariane est arrivée treizième. Donner l'encadrement le plus précis possible de sa performance en minutes.
4) L'affirmation suivante est-elle vraie ?
Affirmation : Plus de 50 % des élèves ont mis un temps supérieur au temps moyen.

6 CRPE 2015 G2

Une petite entreprise emploie 7 personnes, dont 3 femmes.
Voici quelques informations sur le salaire mensuel des personnels :

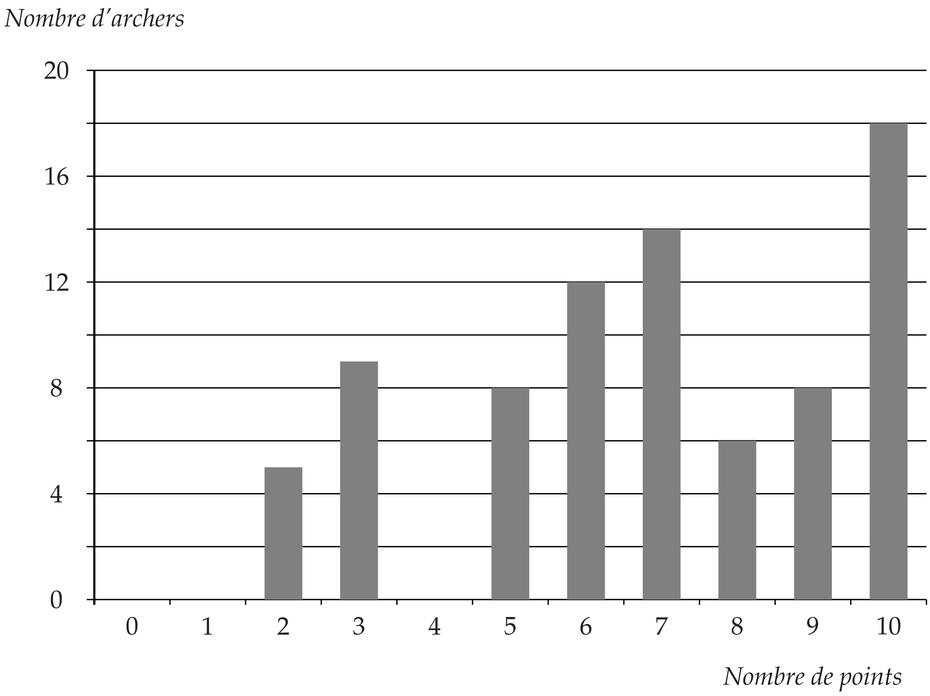
- Salaires des hommes : 1 250 € ; 1 400 € ; 1 600 € ; 3 200 €.
- Salaires des femmes : salaire médian : 1 875 € ; salaire moyen : 1 700 € ; étendue des salaires : 1 000 €.

Le patron de l'entreprise veut embaucher une femme supplémentaire pour respecter la parité.
Calculer le salaire qu'il doit verser à cette nouvelle recrue pour que les salaires moyens des hommes et des femmes soient égaux.

Vu au CRPE

7 CRPE 2016 G2

Quatre-vingts archers d'un club de tir à l'arc A ont participé à un championnat. Le nombre de points obtenus par chaque archer du club est donné par le diagramme ci-dessous.



1) Répondre à l'aide du diagramme précédent aux questions suivantes.
   a) Combien d'archers ont gagné exactement six points lors de ce championnat ?
   b) Combien d'archers ont gagné trois points ou plus lors de ce championnat ?
   c) Quel est le score médian des archers du club A ?
2) Le club de tir à l'arc voisin B a aussi participé à ce championnat. Voici quelques données relatives aux résultats des archers de ce club :
   - Le score moyen des archers lors du championnat est 7 points.
   - Le score moyen des dix meilleurs archers lors du championnat est 9,9 points.

   a) Comparer les résultats des deux clubs selon leurs scores moyens.
   b) Comparer les résultats des deux clubs selon les scores de leurs dix meilleurs archers.

8 CRPE 2017 G2

On considère une série statistique de moyenne égale à 5. On complète la série en ajoutant 5 comme valeur supplémentaire.
Affirmation : la moyenne de la série ne change pas.

9 CRPE 2017 G2

Ce tableau présente la hauteur, en millimètre, des précipitations journalières au cours du mois d'avril 2016, sur l'aéroport Roland Garros de l'île de La Réunion.

Hauteur des précipitations (mm)	0	0,3	1,3	1,7	2,5	7	13	21	28	42
Nombre de jours	4	6	4	4	3	3	2	1	2	1

1) Calculer la valeur moyenne des précipitations journalières au cours du mois d'avril 2016, arrondie au dixième de millimètre.
2) Déterminer la valeur médiane de ces précipitations journalières. Interpréter ce résultat par une phrase.
3) Quelle est l'étendue de cette série ?
4) Déterminer le nombre de jours où la hauteur des précipitations est supérieure ou égale à 13 mm, puis exprimer ce nombre en pourcentage par rapport au nombre de jours dans le mois.
5) Sachant qu'une des pistes de décollage de l'aéroport Roland Garros est rectangulaire et mesure 3 200 m de long et 50 m de large, calculer, en mètre cube, puis en litre, le volume de pluie tombé sur cette piste au cours du mois d'avril 2016.

10 D'après CRPE blanc 2017 M2/modules

Le tableau ci-dessous donne la répartition des 800 chefs d'exploitations agricoles d'une région selon leur âge.

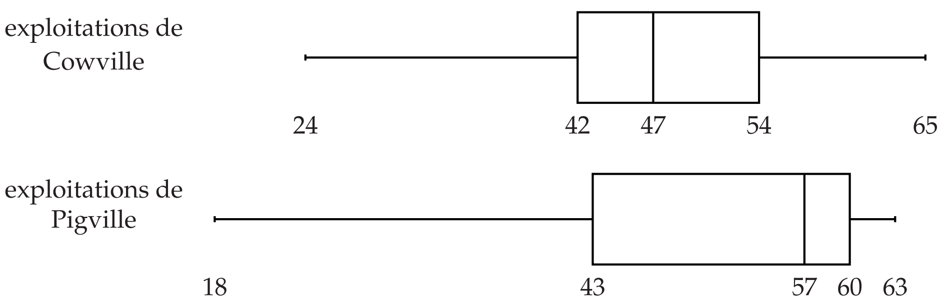
Tranche d'âge	[15 ; 25 [	[25 ; 35 [	[35 ; 45 [	[45 ; 55 [	[55 ; 65 [	total
Effectif	11	84	148	260	297	800

1) a) Expliquer pourquoi l'âge médian des chefs d'exploitation agricole est nécessairement entre 45 et 55 ans.
   b) Pour déterminer l'âge médian, la répartition des âges dans la classe [45 ; 55 [est donnée par le tableau suivant :

âge	45	46	47	48	49	50	51	52	53	54
effectif	18	21	24	31	30	31	30	27	28	20

   Déterminer l'âge médian.

2) Le diagramme en boîte des âges des chefs d'exploitation de Cowville et celui des chefs d'exploitation de Pigville sont représentés ci-dessous.



Un journaliste a écrit : « Dans leur ensemble les chefs d'exploitation de Cowville sont plus jeunes que les chefs d'exploitation de moins de Pigville. » Commenter cette affirmation en utilisant ces diagrammes en boîtes.