

S20C. Autour de la GESTION DE DONNEES

Mise en route¹

A. Alimentation

1. Le pourcentage de protéines contenu dans le muesli peut se calculer directement $\frac{1,8}{30} \times 100 = 6\%$.

De même le pourcentage de lipides est de $\frac{5,4}{30} \times 100 = 18\%$.

L'utilisation du tableau de données permet de globaliser les informations de l'énoncé, mais il n'est en rien obligatoire. Le calcul peut alors se faire en utilisant des procédures de proportionnalité.

Protéines (g)	1,8	p
Muesli (g)	30	100
Lipides (en g)	5,4	q

$$p = \frac{1,8 \times 100}{30} = 6g \text{ et } q = \frac{5,4 \times 100}{30} = 18g$$

Les 100g de muesli contiennent 6g de protéines soit 6%, et 18g de lipides soit 18%.

2. La masse de glucides peut se calculer directement puisqu'on connaît le pourcentage :

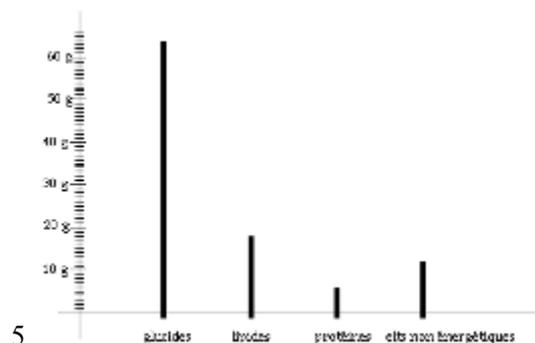
$$30 \times 63,5\% = 30 \times \frac{63,5}{100} = 19,05g \quad \text{30g de muesli contiennent 19,05g de glucides.}$$

3. La valeur énergétique (Kcal) est proportionnelle à la valeur énergétique (KJ).

valeur énergétique (KJ).	1840	552
valeur énergétique (Kcal)	440	e

$$\text{D'où } e = \frac{440 \times 552}{1840} = 132Kcal$$

4. 100g de muesli contiennent donc 6% soit 6g de protéines, 63,5% soit 63,5g de glucides et 18% soit 18g de lipides. La lecture du diagramme circulaire montre que la masse des éléments non énergétiques contenus dans 100g de muesli est égal à $100 - (63,5 + 6 + 18) = 12,5g$.



5.

¹A. Montpellier 97 – B. Source PISA 2003
Parimaths.com

B. Bonbons colorés

Il y a $6+5+3+3+2+4+2+5=30$ bonbons dans le sachet, dont 6 bonbons rouge.

La probabilité que « Kevin prenne un bonbon rouge » se calcule en divisant le nombre tirages favorables

(rouge) par le nombre de tirages possibles, soit : $\frac{6}{30} = \frac{1}{5} = 20\%$

C. Croissance²

1. $170,6 - (2,3 \times 18) = 129,2 \text{ cm}$

2. De 11 ans à 13 ans, les jeunes filles sont, en moyenne, plus grandes que les jeunes hommes du même âge.

3. En moyenne, la croissance des filles est plus lente après 12 ans. On le voit sur le graphique car entre 10 ans et 12 ans, donc sur un intervalle de 2 ans la taille augmente d'environ 17 cm (157-140), alors que de 12 ans à 14 ans elle n'augmente que de 8 cm (165-157), puis de 14 ans à 16 ans de 3 cm (168-165).

D. Diagramme en boîte

Les extrémités des segments représentent les valeurs extrêmes de la série, c'est-à-dire les températures minimale et maximale pour chaque ville : le diagramme A représente Saint Etienne avec la plus petite valeur minimale, le diagramme B représente Nice avec la plus grande valeur maximale, le diagramme C est celui de Brest.

Les côtés verticaux des rectangles représentent à gauche le premier quartile, à droite le troisième quartile, entre les deux la médiane³, et partagent chaque série en 4 groupes de trois températures. On peut rapidement comparer les trois médianes et les trois étendues, ce qui donne déjà une idée de **la position et de la dispersion** respectives de ces trois séries. On peut ainsi comparer⁴ :

- *La position des cinq nombres Min, Q1, Med, Q3, Max : pour Nice Q3 est supérieur au Max des 2 autres...ce qui traduit des températures plus chaudes et plus fréquentes que dans les deux autres villes.*
- *La position de chacune des tranches des intervalles interquartiles : le 1^{er} intervalle interquartile [Q1;Med] de Nice est au niveau du 2^{ème} de Brest [Med;Q3]...ce qui montre que les températures de Nice sont globalement plus élevées que celles de Brest tout au long de l'année*
- *la longueur des intervalles interquartiles (Q3-Q1) et des autres (Max-Q3 et Q1-Min)...montre par exemple que Saint Etienne présente des températures beaucoup plus étalées que Brest*

« Regardez de près et gardez bien en mémoire le fait que les abscisses sont des températures.

L'interprétation de ce graphique est une situation intéressante et très riche bien que le nombre de données soit faible ».

E. Evaluation

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

Devoirs	n°1	n°2	n°3	n°4	n°5	n°6	Moyenne
Notes de Luc	12	5	18	11	19		

² C. D'après Amiens2004 – D. sierra.univ-lyon1.fr/irem - E. Grenoble 2005

³ Lire les remarques du cours fiche Gestion de données

⁴ D'après sierra.univ-lyon1.fr/irem

Notes de Julie	20	15	4	9	x	y	12,5
----------------	----	----	---	---	-----	-----	------

1. La moyenne de Luc ne change pas si la note obtenue au sixième devoir est égale à la moyenne des cinq premiers.

$$\frac{12+5+18+11+19}{5}=13$$

$$\frac{12+5+18+11+19+13}{6}=13$$

2. Luc a déjà totalisé 65 points. Luc ne pourra pas avoir une moyenne de 15. En effet :

- soit on regarde avec la note maximale 20, le total est alors de 85 points soit une moyenne de 14,2 environ
- soit on raisonne sur le nombre total de points, plutôt que sur la moyenne : $15 \times 6 = 90$. Luc en a déjà 65, il faudrait une note de 25.

3. $y = x + 25\% \times x = 1,25x$

Julie a totalisé 75 points ($12,5 \times 6 = 75$), pour une somme de $20 + 15 + 4 + 9 + x + 1,25x = 48 + 2,25x$

$$48 + 2,25x = 75 \quad x = \frac{27}{2,25} = 12 \text{ et } y = 15$$

Probabilités

F. a. Un sac contient quatre jetons numérotés de 1 à 4. . On tire un premier jeton dont on note le numéro, on le remet dans le sac et on tire un deuxième jeton. **Vu qu'il y a remise du jeton après le premier tirage**, il y a 4 possibilités à chaque tirage, d'où 16 tirages possibles donnant ce tableau des résultats :

Tirages/résultat	1	2	3	4
1	1	2	3	4
2	2	4	6	8
3	3	6	9	12
4	4	8	12	16

Il y a donc un cas favorable à un résultat égal à 1 ou 9 ou 16, soit une probabilité de $\frac{1}{16}$

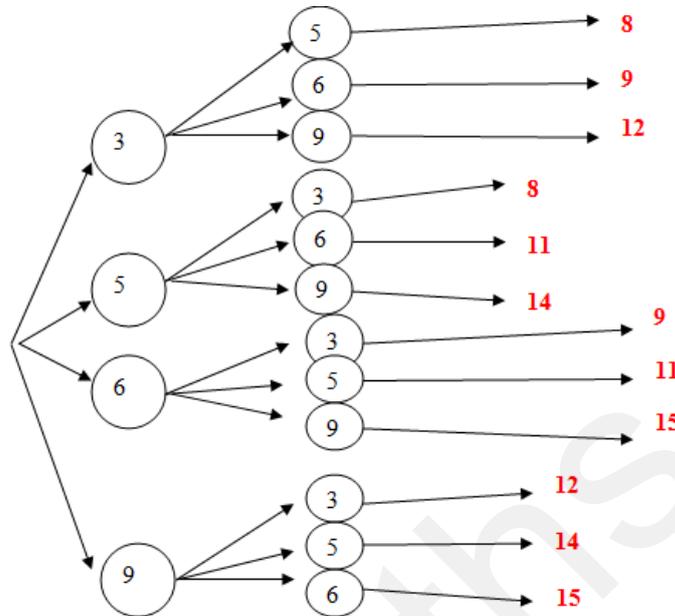
Il y a 2 cas favorables à un résultat égal à 2 ou 3 ou 6 ou 8 ou 12, soit une probabilité de $\frac{2}{16}$ ou $\frac{1}{8}$

Il y a 3 cas favorables à un résultat égal à 4, soit une probabilité de $\frac{3}{16}$

b. Il y a 3 possibilités d'obtenir un nombre strictement supérieur à 9, donc 13 de ne pas obtenir cette réponse, soit une probabilité de $\frac{13}{16}$.

G. On dispose de quatre brins de paille qui mesure 3cm, 5cm, 6cm, 9cm. On tire au hasard un premier brin, puis sans remettre ce brin, on en tire un deuxième. On met bout à bout ces deux brins et on mesure la longueur du brin obtenu.

a. Un arbre peut permettre de déterminer toutes les longueurs possibles. **Compte tenu du fait qu'il n'y a pas remise du brin tiré lors du tirage suivant**, le nombre de tirages possibles est égal à $4 \times 3 = 12$. Les différentes longueurs obtenues sont : 8cm (deux fois), 9cm (deux fois), 11cm (deux fois), 12cm (deux fois), 14cm (deux fois), 15 (deux fois).



b. La probabilité d'obtenir une longueur de 11cm est donc de $\frac{2}{12}$ soit $\frac{1}{6}$.

c. La probabilité d'obtenir une longueur inférieure ou égale à 11cm (c'est-à-dire 8cm, 9cm, 11cm) est de $\frac{6}{12}$ soit $\frac{1}{2}$.

d. Pour trouver la probabilité d'obtenir une longueur différente de 15cm (**évènement e**), il est plus rapide de chercher la probabilité de l'**évènement contraire** (une longueur égale à 15) qui est de $\frac{2}{12}$ puis de

conclure avec $P_e = 1 - \frac{2}{12} = \frac{10}{12}$ soit $\frac{5}{6}$.

Séries Statistiques

H. a. Série 1.

Population étudiée : 600 employés d'une entreprise.

Caractère qualitatif : les moyens de transport utilisés.

Série 2.

Population étudiée : 90 candidats à un concours de la DDE.

Caractère quantitatif discret : notes obtenues.

Série 3.

Population étudiée : 200 personnes

Caractère quantitatif continu : durée passée chaque jour devant le téléviseur.

Ici la répartition se fait par classe, selon des intervalles regroupant différentes durées. Les amplitudes sont irrégulières (1h, 1h, 2h, 2h, 4h). Ainsi le premier intervalle regroupe les personnes regardant la télévision moins d'une heure ; les réponses sont globalisées, que ce soit 10 minutes ou 3/4 heure. De même, le dernier intervalle regroupe celles regardant leur télévision entre 6h et 10h. Ce type de répartition permet d'avoir une vision plus globale et plus représentative de la situation étudiée.

b. Pour la série 1, fréquence d'utilisation du vélo : $f_{\text{vélo}} = \frac{18}{600} = 0,03$

Pour la série 2, fréquence de la note 13 : $f_{13} = \frac{6}{90} \approx 0,07..$

Total des effectifs pour une note inférieure ou égale à 10 : 37 fréquence: $f_{n \leq 10} = \frac{37}{90} \approx 0,41..$

Pour la série 3, fréquence d'un temps passé compris entre 4 et 6 heures par jour : $f_{4 \leq t < 6} = \frac{50}{200} = 0,25$

c. Pour la série 2, note moyenne :

$$\frac{5 \times 5 + 6 \times 9 + 7 \times 10 + 8 \times 5 + 9 \times 2 + 10 \times 6 + 11 \times 11 + 12 \times 11 + 13 \times 6 + 14 \times 5 + 15 \times 8 + 16 \times 4 + 17 \times 7 + 18 \times 0 + 19 \times 1}{90} = \frac{990}{90} = 11$$

La note médiane est celle qui partage la série en deux groupes de même effectif, soit 45 personnes : il y a 42 personnes qui obtiennent une note supérieure à 11 et 37 personnes qui obtiennent une note inférieure à 11 et 11 personnes qui obtiennent la note de 11. La note médiane est donc 11.

Pour la série 3, temps moyen passé devant le téléviseur :

$$t_{\text{moyen}} = \frac{0,5 \times 10 + 1,5 \times 20 + 3 \times 100 + 5 \times 50 + 8 \times 20}{200} = 3,725h \text{ soit } 3h43 \text{ min } 30s$$

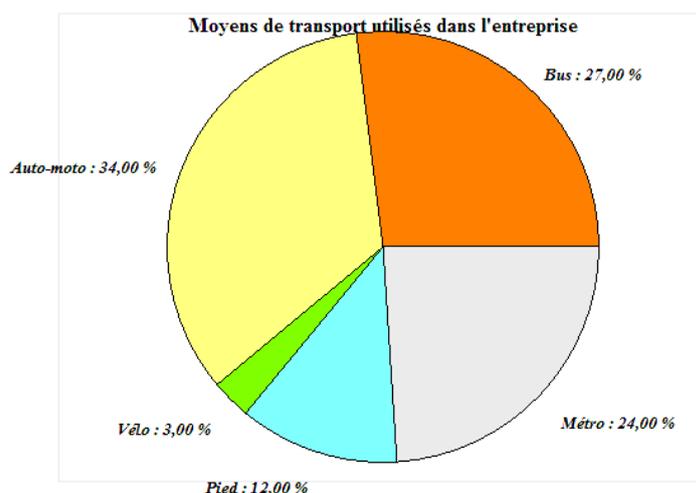
d. Représentations graphiques

Pour réaliser le diagramme circulaire, il faut calculer la mesure de chaque secteur, celle-ci étant

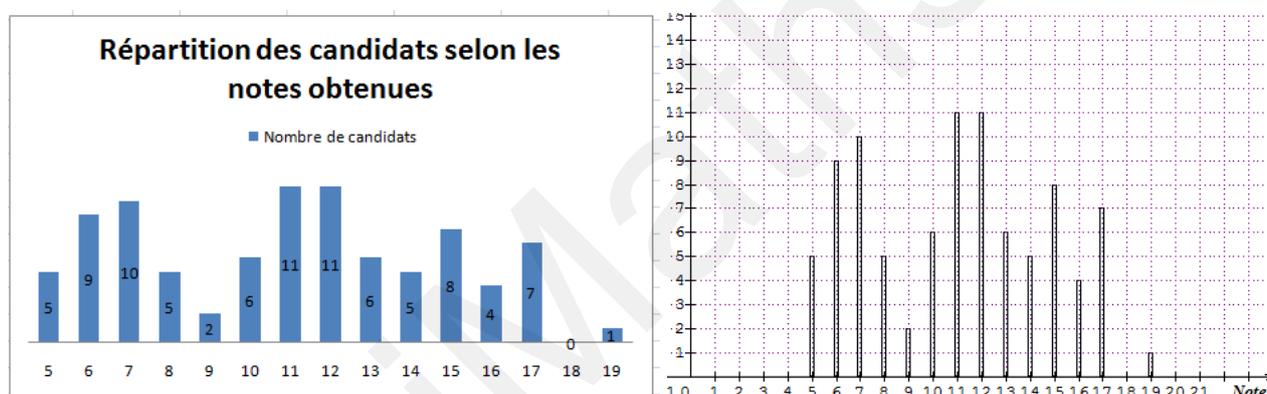
proportionnelle à l'effectif de chaque classe : $\frac{\text{effectif} \times 360}{\text{effectif total}}$

Transport	Bus	Auto-moto	Vélo	À pied	Métro
Effectif	162	204	18	72	144
Angle du secteur	$\frac{162 \times 360}{600} \approx 97,2^\circ$	$\frac{204 \times 360}{600} \approx 122,4^\circ$	$\frac{18 \times 360}{600} \approx 10,8^\circ$	$\frac{72 \times 360}{600} \approx 43,2^\circ$	$\frac{144 \times 360}{600} \approx 86,4^\circ$

Par sécurité, penser à vérifier votre total d'effectifs (600) et votre total de mesure d'angle (360°)

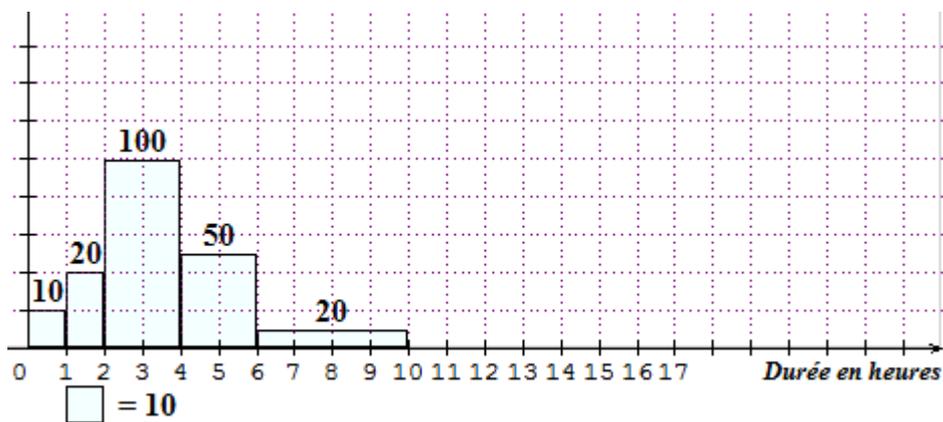


. Pour réaliser ce diagramme en barres (à gauche), en bâtons (à droite), l'axe horizontal est gradué selon les notes obtenues. On trace ensuite soit des rectangles de même largeur soit simplement des segments, dont la hauteur est proportionnelle à l'effectif. Il faut donc choisir une unité selon un axe vertical qui correspond au nombre de candidats.



. Pour réaliser cet histogramme, on gradue l'axe des abscisses selon le nombre d'heures passées, et l'axe des ordonnées selon le nombre de personnes. On peut noter que les deux premiers intervalles ont une amplitude de 1h, les deux suivants une amplitude de 2h, le dernier une amplitude de 4h. La hauteur des rectangles n'est donc pas proportionnelle à l'effectif, mais l'aire de chaque rectangle l'est. Ainsi si 1 carreau représente 10 personnes passant 1 heure devant la télé, il faudra 2 carreaux pour 20 personnes passant aussi 1 heure, répartis sur un rectangle de largeur 1, de hauteur 2. Par contre pour représenter les 100 personnes passant 2 heures, il faudra 10 carreaux répartis sur un rectangle de largeur 2, soit 5 carreaux en hauteur. De même pour les 50 personnes passant 2 heures devant leur télé, les 5 carreaux sont répartis sur un rectangle de largeur de 2, soit 2,5 carreaux en hauteur. Enfin les 20 personnes passant 4 heures sont représentées par 2 carreaux répartis sur un rectangle de largeur 4 et donc de hauteur 0,5.

Prudence donc, c'est bien l'aire du rectangle qui est proportionnelle à l'effectif.



Pour s'exercer⁵

Exercice 1

1. La population totale est égale à la somme des trois sous populations, soit 149400 milliers d'habitants. Pour trouver les fractions irréductibles, on peut décomposer numérateur et dénominateur en utilisant les critères de divisibilité.

$$\frac{31125}{149400} = \frac{5 \times 6225}{100 \times 1494} = \frac{5 \times 5 \times 5 \times 249}{100 \times 2 \times 747} = \frac{5 \times 5 \times 5 \times 249}{100 \times 2 \times 3 \times 249} = \frac{5}{24} \approx 20,8\%$$

On peut aussi utiliser la décomposition en facteurs premiers qui optimise la décomposition de chaque terme, avant la simplification : $31125 = 3 \times 5^3 \times 83$ et $149400 = 2^3 \times 3^2 \times 5^2 \times 83$

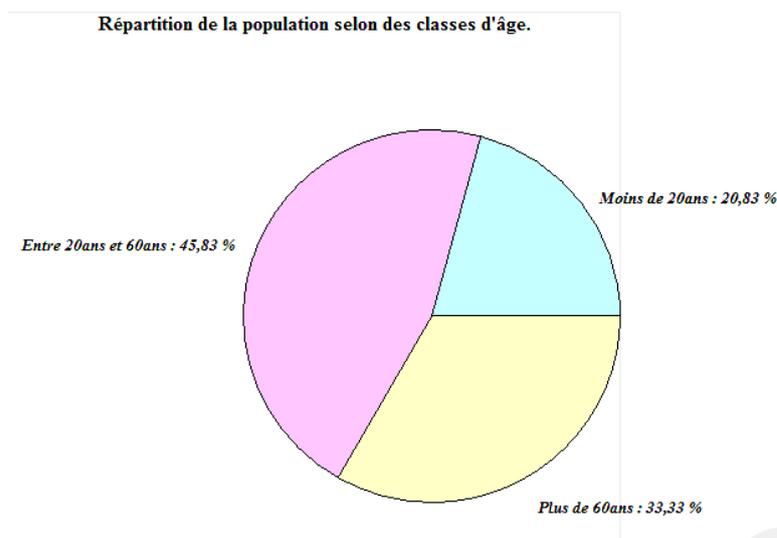
On trouve de même $\frac{68475}{149400} = \frac{11}{24} \approx 45,8\%$ et $\frac{49800}{149400} = \frac{1}{3} \approx 33,3\%$

2. Il y a proportionnalité entre la mesure des secteurs angulaires et la population représentée que l'on peut exprimer en milliers d'habitants, en fraction de la population totale (valeur exacte) ou en pourcentage (valeur approchée)

	Moins de 20 ans	Entre 20 et 60 ans	Plus de 60 ans	Total
Population (fraction de P)	$\frac{5}{24} \approx 20,8\%$	$\frac{11}{24} \approx 45,8\%$	$\frac{1}{3} \approx 33,3\%$	$\frac{24}{24} = 100\%$
Secteur	$360 \times \frac{5}{24} = 75^\circ$	$360 \times \frac{11}{24} = 165^\circ$	$360 \times \frac{1}{3} = 120^\circ$	360°

Pour construire ce camembert, on peut tracer un cercle et placer les sommets de l'hexagone ; on obtient alors un secteur de 120° . On peut obtenir le secteur de 75° en le décomposant en $60^\circ + 15^\circ$. Le secteur de 15° s'obtient en traçant deux fois les bissectrices des secteurs obtenus....

⁵ 1. Axiale, Hatier 2004 / 7. Montpellier 1997



Exercice 2

On peut calculer le nombre de salariés par catégorie, ainsi que le nombre total : l'entreprise E_1 comporte $250 + 165 + 30 = 445$ salariés, l'entreprise E_2 comporte $200 + 150 + 50 = 400$ salariés.

	Entreprise E_1			Entreprise E_2		
	$[1;1,5[$	$[1,5;3[$	$[3;8[$	$[1;1,5[$	$[1,5;3[$	$[3;8[$
Nombre d'ouvriers	250	150	0	200	100	0
Nombre de cadres	0	15	30	0	50	50
Nombre de salariés	250	165	30	200	150	50

Le PDG de l'entreprise E_2 dit à son collègue de l'entreprise E_1 : « *Mes salariés sont mieux payés que les vôtres* ». « *Faux* », répond ce dernier « *puisque mes ouvriers sont mieux payés que les vôtres et mes cadres aussi* ».

1. En supposant que tous les individus d'une classe ont pour valeur le centre de cette classe, et en exprimant les salaires moyens en milliers d'euros :

a. Salaires moyens m_1 et m_2 **des salariés** de ces deux entreprises

$$m_1 = \frac{1,25 \times 250 + 2,25 \times 165 + 5,5 \times 30}{445} \approx 1,9073$$

$$m_2 = \frac{1,25 \times 200 + 2,25 \times 150 + 5,5 \times 50}{400} = 2,15625$$

b. Salaires moyens o_1 et o_2 **des ouvriers** de ces deux entreprises :

$$o_1 = \frac{1,25 \times 250 + 2,25 \times 150 + 5,5 \times 0}{400} = 1,625$$

$$o_2 = \frac{1,25 \times 200 + 2,25 \times 100 + 5,5 \times 0}{300} \approx 1,583$$

c. Salaires moyens c_1 et c_2 **des cadres** de ces deux entreprises

$$c_1 = \frac{1,25 \times 0 + 2,25 \times 15 + 5,5 \times 30}{45} \approx 4,4$$

$$c_2 = \frac{1,25 \times 0 + 2,25 \times 50 + 5,5 \times 50}{100} = 3,875$$

2. On constate que $m_1 < m_2$. Les salariés de l'entreprise E_2 sont donc en moyenne mieux payés que ceux de l'entreprise E_1 , comme l'affirme le PDG de cette entreprise. On peut aussi constater que $o_1 > o_2$ et $c_1 > c_2$. Il est donc aussi vrai que le salaire moyen des ouvriers de l'entreprise E_1 est supérieur à

celui des ouvriers de l'entreprise E_2 , et il en est de même pour les cadres comme l'affirme aussi le PDG de cette entreprise. On voit donc ici les limites d'interprétation des moyennes et la nécessité d'étudier d'autres paramètres des séries quand on veut les comparer.

Exercice 3

Dans une entreprise où les salariés sont classés en trois catégories - ouvriers, employés, cadres – on fait une étude sur le salaire mensuel brut. Le salaire moyen des ouvriers est de 1480€, celui des employés 1620 €, et celui des cadres 3450 €.

$$1. \quad s_{moyen} = \frac{1480 + 1620 + 3450}{3} \approx 2183$$

Il est peu probable que ce salaire soit le salaire moyen des salariés car cela supposerait que le nombre d'ouvriers, le nombre d'employés et le nombre de cadres soient égaux...

2. Pour calculer le « salaire moyen » dans l'entreprise, on va « pondérer » les montants des salaires de chaque catégorie en le multipliant par l'effectif correspondant, les ajouter puis diviser par l'effectif total égal $38 + 24 + 6 = 68$.

$$s_{moyen} = \frac{1480 \times 38 + 1620 \times 24 + 3450 \times 6}{68} \approx 1703$$

Exercice 4

Lucie et Marc participent à une compétition de tir à l'arc.

$$1. \quad m_{sept\ tours} = \frac{18 + 28 + 12 + 29 + 26 + 19 + 22}{7} = \frac{154}{7} = 22$$

En ajoutant un score de 22 au troisième tour, le total devient égal à 176 et la moyenne reste donc 22.

On peut noter plus généralement que, si l'on ajoute la valeur de la moyenne au total S_{n-1} des valeurs, la

moyenne reste inchangée. $S_{n-1} = (n-1) \times m \quad S_n = (n-1)m + m = n \times m \quad \frac{S_n}{n} = m$

2. Pour obtenir une moyenne supérieure ou égale à 23, il faudrait que Marc totalise au moins $23 \times 8 = 184$ points. Il a déjà totalisé 154 points en sept tours ; il lui manque donc 30 points, ce qui est réalisable seulement par une réussite maximale des trois flèches de cette volée.

3. $x = y + \frac{40}{100}y = 1,4y$. Le total des points de Lucie est de 184 points. D'où :

$$1,4y + y + 29 + 12 + 26 + 27 + 17 + 25 = 23 \times 8 = 184 \quad 2,4y + 136 = 184 \quad y = \frac{184 - 136}{2,4} = 20 \quad \text{et } x = 1,4 \times 20 = 28$$