

laths .

# S20. Autour de la GESTION DE DONNEES Probabilités, Statistiques

### Mise en route<sup>1</sup>

#### A. Alimentation

L'étiquette d'un paquet de céréales affiche: « 30g de muesli croustillant dans 100g de lait donnent un excellent petit déjeuner énergétique. Les 30g de muesli assurent un quart des apports quotidiens recommandés en vitamines du groupe B. Les 100g de lait entier apportent protéines, calcium, phosphore et vitamines A et D. ». Le diagramme circulaire présente quelques informations nutritionnelles pour 100g de Muesli, et il est précisé que la valeur énergétique est de 1840 Kilojoules, pour 440 Kilocalories. Cependant certains renseignements ont disparu de cette étiquette...

## Informations nutritionnelles moyenne pour 100g de Muesli Autres éléments



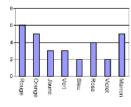
	30g de Muesli	%	
Protéines	1,8g		
Glucides	?	63,5%	
Lipides	5,4g		
Eléments non			
énergétiques			
Valeur énergétique	552KJ / ?Kcal		
en KJ/ en Kcal			

- 1. Quelle est le pourcentage de protéines et de lipides contenus dans 100 grammes de muesli?
- 2. Quelle est la masse de glucides contenue dans 30 grammes de muesli?
- 3. Quelle est la valeur énergétique de 30 grammes de muesli en kilocalories (kcal), sachant qu'elle est de 552Kjoule ?
- 4. Quelle est la masse des éléments non énergétiques contenue dans 100 grammes de muesli ?
- 5. La répartition des composants du muesli est donnée par le diagramme circulaire. En donner une représentation par un diagramme en bâtons.

#### B. Bonbons colorés

La mère de Kevin lui permet de prendre un bonbon dans un sachet opaque. Kevin ne voit donc pas les bonbons. Le nombre de bonbons de chaque couleur contenus dans le sachet est illustré par le graphique ci-contre. Parmi ces réponses, quelle est la probabilité que Kevin prenne un bonbon rouge ?

A 10 % B 20 % C 25 % D 50 %?



<sup>&</sup>lt;sup>1</sup> A. d'après Montpellier 97 – B. Source PISA 2003

\_

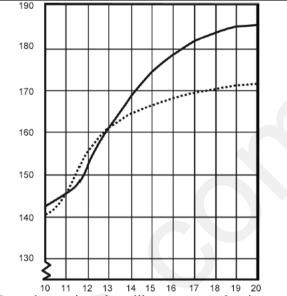
### C. Croissance<sup>2</sup>

La taille moyenne des jeunes hommes et des jeunes femmes aux Pays-Bas en 1998 est représentée par le graphique ci-contre où les âges sont représentés en abscisse et les tailles en ordonnée.

**Question 1**: Depuis 1980, la taille moyenne des jeunes filles de 20 ans a augmenté de 2,3 cm, pour atteindre 170,6 cm. Quelle était la taille moyenne des jeunes filles de 20 ans en 1980 ?

**Question 2** : D'après ce graphique, pendant quelle période de leur vie les jeunes filles sont-elles, en moyenne, plus grandes que les jeunes hommes du même âge ?

**Question 3**: Expliquez en quoi le graphique montre qu'en moyenne, la croissance des filles est plus lente après 12 ans.



En trait continu, la taille moyenne des jeunes hommes en fonction de leur âge, en trait pointillé, la taille moyenne des jeunes femmes.

# D. Diagramme en boîte

Comparaison de relevés de températures mensuelles moyennes, en degrés C, dans trois villes de France, sur une année. Que voit-on sur les trois diagrammes en boîtes superposées de l'exemple ci-dessous ?

	Saint-Etienne	Nice	Brest
Janvier	2,4	7,9	6,3
Février	3,8	8,9	6,2
Mars	5,9	10,6	7,4
Avril	8,8	13,2	8,8
Mai	12,9	16,5	11,5
Juin	16,9	20	14,1
Juillet	19,1	23	16
Août	18,3	23	16
Septembre	15,5	20,3	14,7
Octobre	11,3	16,5	12,2
Novembre	6,1	11,7	8,9
Décembre	3	8,6	7,5

# E. Evaluation

Voici la liste des notes sur 20 obtenues par 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

<sup>&</sup>lt;sup>2</sup> C. D'après Amiens2004 – D. *sierra.univ-lyon1.fr/irem* - E. Grenoble 2005 Parimaths.com CRPE 2010-2011

Notes de Luc	12	5	18	11	19		
Notes de Julie	20	15	4	9	х	У	12,5

- 1. Calculer la moyenne de Luc si la note obtenue au sixième devoir est égale à la moyenne des cinq premiers.
- 2. Une meilleure note au devoir n°6 aurait-elle permis à Luc d'obtenir une moyenne de 15 ?
- 3. La note y obtenue par Julie au devoir n°6 a augmenté de 25% par rapport à la note x qu'elle a obtenue au devoir n°5. Exprimer y en fonction de x. Calculer x et y.

### Probabilités

- **F.** 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. On multiplie les nombres obtenus.
  - a. A l'aide d'un tableau à double entrée calculer la probabilité de chaque résultat possible.
  - b. Calculer la probabilité de ne pas obtenir un résultat strictement supérieur à 9.
- G. On dispose de quatre brins de paille qui mesurent 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. A l'aide d'un arbre déterminer toutes les longueurs possibles.
  - b. Calculer la probabilité d'obtenir une longueur de 11cm.
  - c. Calculer la probabilité d'obtenir une longueur inférieure ou égale à 11 cm.
  - d. Calculer la probabilité d'obtenir une longueur différente de 15 cm.

### Séries Statistiques

H. Voici trois séries de données.

Série 1. Voici la répartition des 600 employés d'une entreprise suivant les moyens de transport.

Transport	Bus	Auto-moto	Vélo	À pied	Métro	TOTAL
Effectif	162	204	18	72	144	600

Série 2. 90 candidats se présentent à un concours départemental organisé par la DDE.

Les notes obtenues sont données dans le tableau suivant.

Note	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Effectif	5	9	10	5	2	6	11	11	6	5	8	4	7	0	1

Série 3. Une enquête sur 200 personnes porte sur la durée passée chaque jour devant le téléviseur.

Durée en heures	[0;1[	[1;2[	[2;4[	[4;6[	[6;10[	Total
Effectif	10	20	100	50	20	200

- **a.** Préciser, pour chaque série, la population étudiée et le caractère étudié. Justifiez le type de répartition des effectifs choisi pour la série 3.
- **b.** Quelle est, pour la série 1, la fréquence d'utilisation du vélo.

Quelle est, pour la série 2, la fréquence de la note 13 ? Quelle est la fréquence d'une note inférieure ou égale à 10 ?

Quelle est, pour la série 3, la fréquence d'un temps passé devant le téléviseur compris entre 4 et 6 heures par jour ?

**c.** Quelle est, pour la série 2, la note moyenne? la note médiane?

Pour la série 3, quel est le temps moyen passé devant le téléviseur ?

- d. Représenter graphiquement :
- . Par un diagramme circulaire, la répartition des 600 employés d'une entreprise suivant les moyens de transport utilisé.
- . Par un diagramme en bâtons, les notes obtenues par 90 candidats se présentant à un concours départemental organisé par la DDE.
- . Par un histogramme, la durée passée chaque jour devant le téléviseur par les 200 personnes interrogées.

# Pour s'exercer<sup>3</sup>

#### **Exercice 1**

Le tableau ci-dessous donne la répartition de la population d'un pays par groupe d'âge. Le diagramme circulaire représente les données du tableau.

Population en mil	liers d'habitants	entre 20 et 60 ans
Moins de 20 ans	31 125	
Entre 20 et 60 ans	68 475	
Plus de 60 ans	49 800	plus de 60 ans

- 1. Exprimez la part de chaque catégorie sous la forme de fractions irréductibles, puis sous la forme de pourcentages.
- 2. Calculez la mesure de chaque secteur angulaire permettant de réaliser le diagramme circulaire correspondant à ces données.
- 3. Faire la construction du « camembert » en utilisant uniquement règle et compas. (On pourra éventuellement s'aider de la construction préalable d'angles de mesure 60°; 15°; 45°).

### **Exercice 2**

Parimaths.com CRPE 2010-2011/2013 CMJ

<sup>3 1.</sup> Amiens 2002 - 2.3. Axiale, Hatier 2004 - 4. 2009Groupe4

Dans deux entreprises  $E_1$  et  $E_2$ , les salariés sont répartis en deux catégories : ouvriers et cadres. Les deux tableaux suivants donnent la répartition des salaires mensuels (en milliers d'euros) des employés de ces deux entreprises.

		Entreprise E	1	Entreprise E <sub>2</sub>			
Salaire mensuel	[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	

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 :
  - **a.** calculer les salaires moyens  $m_1$  et  $m_2$  des salariés de ces deux entreprises.
  - **b.** calculer les salaires moyens  $o_1$  et  $o_2$  des ouvriers de ces deux entreprises.
  - c. calculer les salaires moyens  $c_1$  et  $c_2$  des cadres de ces deux entreprises.
- 2. Arbitrer le désaccord entre les deux PDG

#### 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. Pour calculer le salaire moyen des salariés de cette entreprise, Alain fait la somme des trois moyennes et divise le résultat par 3. Faire le calcul, et expliquer pourquoi il est peu probable que le résultat obtenu soit pertinent.
- 2. Proposer une autre méthode de calcul du « salaire moyen » si les effectifs de chaque catégorie sont respectivement 38, 24 et 6.

### **Exercice 4**

Lucie et Marc participent à une compétition de tir à l'arc. Voici les scores partiels des 8 tours de volées de 3 flèches. (10 est le score parfait pour une flèche).

	Tour 1	Tour 2	Tour 3	Tour 4	Tour 5	Tour 6	Tour 7	Tour 8	Moyenne
Lucie	x	У	29	12	26	27	17	25	23
Marc	18	28		12	29	26	19	22	

1. Calculer la moyenne des scores de Marc si le score obtenu au tour 3 est égal à la moyenne des scores des 7 tours déjà notés dans le tableau.

- 2. Une performance meilleure au tour 3 lui aurait-elle permis d'obtenir une moyenne supérieure ou égale à celle de Lucie ? Justifier.
- 3. Le score *x* obtenu par Lucie au premier tour est supérieur de 40% au score *y* qu'elle a obtenu au second tour. Après avoir exprimé *x* en fonction de *y*, calculer *x et y*.

### A retenir

## Vocabulaire des séries statistiques

La population est l'ensemble faisant l'objet d'une étude statistique. L'effectif de la population est le nombre d'individus qui la composent.

Le caractère est la propriété étudiée sur chaque individu. Ces caractères peuvent être qualitatifs (couleur, forme,... qui ne peuvent pas être chiffrés) ou quantitatifs (taille, poids, salaire....chiffrables).

Les caractères quantitatifs peuvent être **discrets**, c'est-à-dire qu ils ne prennent qu'un certain nombre de valeurs (nombre d'élèves, âge...) ou **continus** c'est dire qu' ils prennent toutes les valeurs d'un intervalle, ou alors beaucoup trop de valeurs pour les considérer toutes (tailles d'individus en cm, salaires au centime près, .....). On regroupe alors ces valeurs par intervalles appelés **classes**.

L'étendue de la série est l'écart entre les deux valeurs extrêmes de la série. Le mode est l'une des valeurs de la série ayant le plus grand effectif.

# Fréquences, moyennes, médiane, quartile

La fréquence d'une partie de la population est le quotient de deux effectifs :

$$fréquence de la partie = \frac{effectif de la partie}{effectif de la population}$$

On peut l'exprimer sous la forme d'un nombre compris entre 0 et 1 ou sous la forme d'un pourcentage.

La moyenne d'un caractère quantitatif, simple ou pondérée par des coefficients, se calcule selon deux modes :

$$moyenne = \frac{somme \ des \ valeurs \ du \ caractère \ de \ chacun \ des \ individus}{effectif \ de \ la \ population}$$
 ou 
$$moyenne \ (pondérée) = \frac{somme \ des \ produits"effectif \times valeur"}{effectif \ de \ la \ population}$$

La médiane<sup>4</sup> est la valeur de la série qui partage la population en deux ensembles telles que la moitié de la population prend une valeur inférieure ou égale à la médiane, l'autre moitié une valeur supérieure ou égale. Lorsque le nombre de valeurs est impair, il suffit de ranger toutes les valeurs par ordre croissant, la médiane est alors la valeur centrale. Lorsque le nombre de valeurs est pair, il suffit de ranger toutes les valeurs par

CMJ

<sup>4</sup> Source Document du GEPS de mathématiques – 22/12/ 2000. Définition de la médiane adoptée dans le programme de seconde. Parimaths.com CRPE 2010-2011/2013

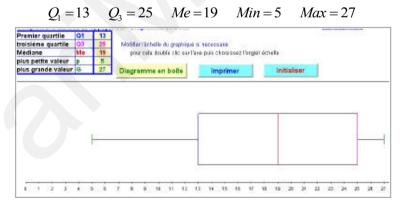
ordre croissant, la médiane est alors la demi- somme des deux valeurs centrales. Si les données sont regroupées en classe, on parle alors de classe médiane.

Plusieurs définitions peuvent être proposées selon le cadre de travail<sup>5</sup> et l'on peut remarquer que lorsque la série comporte des æquo, par exemple 1 2 2 2 2 3 5, certaine définition de la médiane comme « le » nombre *m* tel qu'exactement 50% des termes de la série sont inférieurs à *m* et exactement 50% supérieurs à *m* rend la réponse ambiguë.

D'autres paramètres de dispersion, **les quantiles**, permettent une étude plus précise des séries, en particulier les quartiles, les déciles<sup>6</sup>, les centiles....Etant donné une série de valeurs ordonnée dans l'ordre croissant, le premier quartile est le plus petit élément  $Q_1$  des valeurs de la série tel qu'au moins 25% des données sont inférieures ou égales à ce nombre $Q_1$ . Le troisième quartile est le plus petit élément  $Q_3$ , tel qu'au moins 75% des données sont inférieures ou égales à  $Q_3$ . Les calculs de la médiane et du second quartile donnent en général des résultats très proches. On appelle intervalle interquartile l'intervalle  $Q_1$ ,  $Q_3$  et écart interquartile la différence  $Q_3$  and  $Q_1$ . De même on peut définir le premier décile comme le plus petit élément d des valeurs des termes de la série, ordonnées par ordre croissant, tel qu'au moins 10% des données soient inférieures ou égales à d, et le neuvième décile comme le plus petit élément d des valeurs des termes de la série, ordonnées par ordre croissant, tel qu'au moins 90% des données soient inférieures ou égales à d.

Les diagrammes en boites, appelés aussi boite à moustaches, résument graphiquement une série selon l'idée qu'au lieu de partager l'ensemble des valeurs possibles en segments égaux, on les partage en segments (quartile, déciles, centiles) qui contiennent une proportion prédéterminée des valeurs de la série. Les diagrammes en boîtes permettent de visualiser certains phénomènes et notamment de comparer plusieurs répartitions de valeurs.

Détermination à l'aide du tableur de la médiane, premier et troisième quartile, et construction du diagramme en boîte<sup>7</sup>



Représentations graphiques<sup>8</sup>

CRPE 2010-2011/2013

CMJ

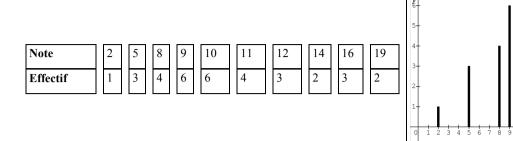
<sup>&</sup>lt;sup>5</sup> Statistiques, Mathématiques, logiciels, média

<sup>&</sup>lt;sup>6</sup> Voir fiche méthode Gestion de données

IREM -Lille 1-David Caille

<sup>&</sup>lt;sup>8</sup> Voir fiche méthode Gestion de données Parimaths.com

a. Diagramme en bâton : Il est utilisé pour représenter un caractère discret. On représente les valeurs du caractère en abscisse et leurs effectifs en ordonnée. Chaque bâton a une hauteur proportionnelle à l'effectif représenté



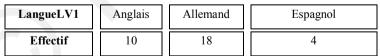
b. Histogramme : l'histogramme est utilisé pour représenter un caractère continu.

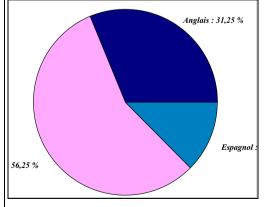
Ce sont des rectangles dont les largeurs sont celles d'une classe et dont les aires sont proportionnelles aux effectifs des classes. Si les classes sont de même amplitude, les hauteurs des rectangles sont proportionnelles à l'effectif représenté, MAIS si les classes sont d'amplitudes différentes, c'est l'aire du rectangle qui est proportionnelle à l'effectif.

Taille cm	[115; 135[	[135 ; 145[	[145; 155[	[155; 165[	[165 ; 185[	[185 ; 195[	15
Nombre d'élèves	5	15	12	10	7	2	

c. Diagramme circulaire : Il est utilisé pour représenter un caractère qualitatif ou quantitatif discret. Le disque est partagé en plusieurs secteurs, l'angle de chaque secteur étant proportionnel à l'effectif.

angle du secteur (en deg rés) = 
$$\frac{effectif}{effectif} \frac{de}{total} \times 360$$





### Probabilités

Une expérience est **aléatoire** si elle conduit à des résultats possibles qu'on est capable de nommer, mais on ne sait pas lequel de ces résultats va se produire quand on réalise l'expérience. A partir de cette expérience on définit des **évènements** qui sont des ensembles de résultats. La probabilité d'un évènement peut se définir intuitivement comme « la chance » qu'un évènement se produise. Quand les évènements sont « **équiprobables** », la **probabilité** p d'un évènement est égale au quotient :

$$p = \frac{nombre \ de \ cas \ favorables \ \grave{a} \ l'\acute{e}v\grave{e}nement}{nombre \ de \ cas \ possibles}$$

La probabilité d'un évènement est donc toujours comprise entre 0 et 1. La somme des probabilités de tous les évènements d'une expérience aléatoire est égale à 1. Si p est la probabilité d'un évènement, alors 1-p est la probabilité de l'évènement contraire.

Si on répète une expérience aléatoire un très grand nombre de fois, la fréquence de n'importe quel évènement se stabilise autour de la probabilité de l'évènement.

Ainsi, si on lance un dé dont les faces sont numérotées de 1 à 6, « *Obtenir un nombre pair* » est un évènement qui correspond à l'ensemble des résultats 2, 4, 6. La probabilité de cet évènement est égale à  $\frac{3}{6}$  puisqu'il y a trois cas favorables sur six cas possibles, soit  $\frac{1}{2}$ . La probabilité d'obtenir un AS de Cœur dans un jeu de 32 cartes est de  $\frac{1}{32}$ , celle d'obtenir un AS est de  $\frac{4}{32}$  soit  $\frac{1}{8}$ . Si l'on veut connaître la probabilité de « ne pas obtenir d'AS », on pourra considérer que cet évènement est

l'évènement contraire du précédent évènement, donc sa probabilité est de  $1 - \frac{1}{8} = \frac{7}{8}$ .