Lycée Paul Sabatier, Carcassonne — Classe de $2^{nde}$

# Chapitre 9 : Statistiques

C. Aupérin

2008-2009

"Télécharger c'est tuer l'industrie, tuons les tous" THURSTON MOORE

Dernière modification : 22 mai 2009

# Table des matières

# Cours : Statistiques

## 1 Vocabulaire

### 1.1 Population

*<u>Travail de l'élève</u> :* Feuille à trous

Au cours d'une enquête portant sur les bébés nés en 2008, on s'intéresse à la taille arrondie au cm, la couleur de leurs yeux et le temps quotidien de sommeil par intervalle d'une heure.

Les premières études étant démographiques, on en a gardé le vocabulaire.

L'ensemble sur lequel porte une étude est appelée **population**.
Un élément de cet ensemble est un **individu**.
On observe alors des propriétés, appelées **caractères**, sur les individus de cette population.
L'**effectif total** est le nombre d'individus de la population.

*Exemple :* Les bébés nés en 2008 représentent la population. On étudie trois caractères : la taille, la couleur de leurs yeux et le temps quotidien de sommeil.

### 1.2 Caractères

Lorsque le caractère étudié prend des valeurs numériques, le caractère est dit **quantitatif**. Il peut alors être :

- **Discret** lorsqu'il ne prend que des valeurs isolées
- **Continue** lorsqu'il peut prendre toutes les valeurs d'un intervalle, appelé **classe**.

Si le caractère n'est pas **quantitatif**, on dit qu'il est **qualitatif**. On appelle ses valeurs des **modalités**.

*Exemple :* La taille est un caractère quantitatif discret, un exemple de valeurs est 55 cm. La couleur des yeux est un caractère quantitatif, un exemple de modalités est bleus. Le temps de sommeil est quantitatif continue, un exemple de classe est $[21; 22[$.

L'effectif d'une valeur d'un caractère est le nombre d'individus ayant cette caleur.

La fréquence d'une valeur est la proportion d'individus possédant cette valeur.

$$\text{Fréquence de la valeur} = \frac{\text{Effectif de la valeur}}{\text{Effectif total}}$$

**Remarque :** La somme des fréquences est toujours égale à 1.

Dans le cas quantitatif, l'**effectif cumulé croissant** d'une valeur (ou d'une classe) est le nombre d'individus ayant une valeur (ou une classe) inférieur ou égale à celle considérée. On lui associe une **fréquence cumulée croissante**.

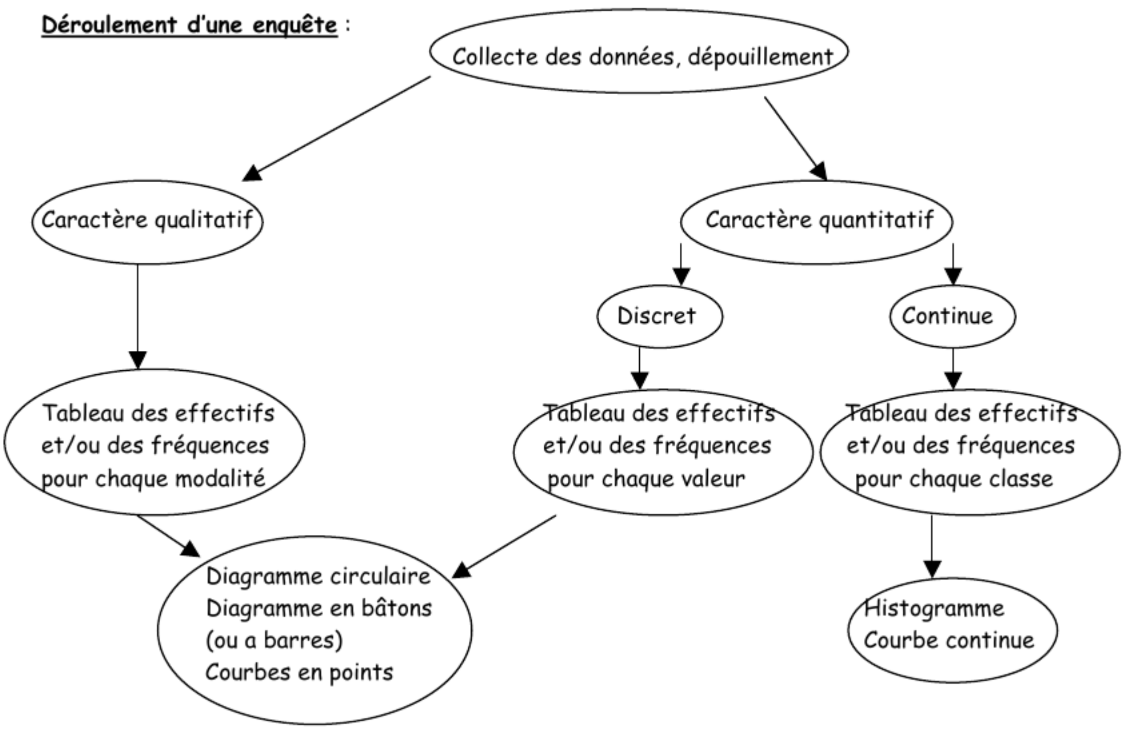
La série statistique est l'ensemble des données collectées, souvent présentées dans un tableau, ou sous forme graphique.

*Exemple :* Compléter les tableaux suivants

| Taille en cm | 44 | 45 | 46 | 47 | 48 | 49 | 50 | 51 | 52 | 53 | 54 | 55 | 56 | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Effectif | 13 | 2 | 0 | 15 | 24 | 33 | 57 | 46 | 32 | 40 | 29 | 6 | 3 | |
| Fréquence à $10^{-2}$ près | | | | | | | | | | | | | | |
| Effectif cumulé croissant | | | | | | | | | | | | | | |
| Fréquence cumulée croissante | | | | | | | | | | | | | | |

| Temps de sommeil en heures | $[18;19[$ | $[19;20[$ | $[20;21[$ | $[21;22[$ | $[22;23[$ | Total |
|---|---|---|---|---|---|---|
| Effectif | 47 | 33 | 74 | 72 | 54 | |
| Fréquence à $10^{-2}$ près | | | | | | |
| Effectif cumulé croissant | | | | | | |
| Fréquence cumulée croissante | | | | | | |

**Déroulement d'une enquête** :



# 2 Paramètres d'une série quantitative

## 2.1 Étendue et mode

L'**étendue** de la série statistique pour un caractère quantitatif est égale à la valeur absolue de la différence entre les valeurs extrêmes prises par le caractère.

Le **mode** (ou **classe modale** dans le cas continu quand les classes sont de même amplitude) est la valeur du caractère ayant le plus grand effectif.

*Exemple :* L'étendue vaut $|56 - 44| = 12$ et le mode vaut 50 pour la taille.
L'étendue du temps de sommeil est $|23 - 18| = 5$ et la classe modale est $[20; 21[$.

## 2.2 Moyenne

**Définition 1.** On considère une série statistique quantitative sur une population de $N$ individus. On note $x_1$, $x_2$, …, $x_p$ les valeurs prises par le caractère et $n_1$, $n_2$, …, $n_p$ les effectifs associés (on a toujours $p \leq N$).
La moyenne est le nombre $\bar{x} = \dfrac{n_1x_1 + n_2x_2 + \cdots + n_px_p}{N} = \dfrac{\Sigma_{i=1}^{p} n_ix_i}{N}$

**Remarque :** Dans le cas continu, on prend pour les $x_i$ le centre de chaque classe.

*Exemple :* Pour la taille $\bar{x} = 50.6$. Pour le temps de sommeil : $\bar{y} \simeq 19h19$

**Propriété 1.** Linéarité :
- Si une série statistique de valeurs $x_1$, $x_2$, …, $x_p$ a pour moyenne $\bar{x}$, la série de valeurs $ax_1$, $ax_2$, …, $ax_p$ a pour moyenne $\overline{ax}$
- Si deux séries statistiques de valeurs $x_1$, $x_2$, …, $x_p$ et $y_1$, $y_2$, …, $y_p$ ont pour moyennes $\bar{x}$ et $\bar{y}$ alors la série de valeurs $x_1 + y_1$, $x_2 + y_2$, …, $x_p + y_p$ a pour moyenne $\bar{x} + \bar{y}$
- Si une série statistique de valeurs $x_1$, $x_2$, …, $x_p$ a pour moyenne $\bar{x}$, la série de valeurs $ax_1 + b$, $ax_2 + b$, …, $ax_p + b$ a pour moyenne $a\bar{x} + b$

Conséquence : La moyenne ne change pas si on remplace les effectifs par les fréquences exactes $f_1$, $f_2$, …, $f_p$ : $\bar{x} = f_1x_1 + f_2x_2 + \cdots + f_px_p = \Sigma_{i=1}^{p} f_ix_i$

*Exemples :* Que devient la moyenne de 10 nombres lorsque :
- Tous augmentent de 5 ?
- Un seul augmente de 5 ?
- Tous son multipliés par 5 ?

*Exemples :*
- Lors d'un devoir, un professeur a noté sur 30. Il ramène toutes ses notes sur 20 en multipliant par $\dfrac{2}{3}$. Que devient la moyenne ?

- A un autre devoir, tous les élèves ont réussi la question bonus sur 2 points. Que devient la moyenne ?
- Lors de deux interrogations écrites sur 10, le professeur décide de calculer les moyennes des interros séparement, puis de les ajouter. cela revient-il au même que d'ajouter pour chaque élève ses notes, puis de faire la moyenne des notes sur 20 ?
- Lors de vos calculs de moyenne, je vous ai dit d'ajouter chacune de vos notes sur 10 et de compter coefficient 1 la nouvelle note obtenue. Si je ramène chacune de vos notes sur 10 à une note sur 20 en la multipliant par 2, quel coefficient dois-je prendre en compte pour chaque interro pour retrouver la même moyenne ?

**Propriété 2.** Si on répartit une série statistique en deux sous-groupes disjoints d'effectifs $N_1$ et $N_2$ et de moyennes respectives $m_1$ et $m_2$, alors la moyenne $\bar{x} = \dfrac{N_1 m_1 + N_2 m_2}{N_1 + N_2}$

**Remarque :** On peut généraliser cette propriété à $n$ sous-groupes disjoints.

*Exemple :* Lors d'un devoir commun, les 209 (35 élèves) ont obtenus 12 de moyenne, les 206 (34 élèves) ont obtenus 11,1 de moyenne et les 210 (31 Ìèves) ont obtenu 12.4 de moyenne. Calculer la moyenne des élèves à ce devoir.

**Définition 2.** On considère une série statistique quantitative sur une population de $N$ individus. On note $x_1, x_2, \ldots, x_p$ les valeurs prises par le caractère et $n_1, n_2, \ldots, n_p$ les effectifs associés (on a toujours $p \leq N$). Parfois, on enlève de l'étude les valeurs trop peu représentées (les considérant comme des « erreurs »). La moyenne élaguée est alors la moyenne sur les individus restants.

*Exemple :* Dans l'exemple de la première partie, faire la moyenne élaguée de la taille des bébés, en considérant que les tailles dont l'effectif est inférieur à 3 comme trop peu représentative.

**Exercice 2.1.** On effectue les pesées de 40 judokas avant une compétition. Leur poids moyen est de 72 kg. On se rend compte que la balance qui a été utilisée est mal réglée et qu'elle indique 500 g de moins que le poids réel. Quel est le poids moyen réel des 40 judokas ?

**Exercice 2.2.** Après six contrôles, un élève obtient 12 de moyenne, puis il obtient 15 au septième contrôle. Tous les contrôles ont le même coefficient. Quelle est sa nouvelle moyenne ?

**Exercice 2.3.** On doit déterminer la moyenne de 560 nombres. À la calculatrice, on trouve 115 comme moyenne. Mais on s'aperçoit que l'on a oublié d' « entrer »l'un des nombres, à savoir 171. Quelle est la moyenne des 561 nombres ?

**Exercice 2.4.** Une épreuve d'examen est constituée de deux parties indépendantes A et B. Un professeur corrige la partie A et un autre la partie B. La note totale à cette épreuve s'obtient en ajoutant la note obtenue à la partie A et la note obtenue à la partie B. Pour la partie A, la moyenne des notes est égale à 9 et pour la partie B à 7. Quelle est la moyenne des notes à cette épreuve ?

**Exercice 2.5.** Dans deux entreprises A et B, les moyennes des salaires masculins et féminins sont données par le tableau suivant :

| Salaire moyens en euros | A | B |
|---|---|---|
| Hommes | 1400 | 1500 |
| Femmes | 1000 | 1100 |

La répartition hommes/ femmes dans les deux entreprises est donnée par le tableau suivant :

| Répartition | A | B |
|---|---|---|
| Hommes | 50% | 20% |
| Femmes | 50% | 80% |

Pour chaque catégorie (hommes ou femmes), l'entreprise B paye mieux et pourtant ...
Calculer pour chaque entreprise la moyenne des salaires pour l'ensemble des employés.
Quels commentaires pouvez-vous faire ?
*Commentaire* : Cette influence de la pondération est appelée « effet de structure »

**Exercice 2.6.** La moyenne de 5 notes d'un élève est de 12. Les quatre pemières sont 13, 10, 8 et 15. Quelle est la cinquième ?

**Exercice 2.7.** Un candidat à un examen a passé les quatre premières épreuves suivantes : Les Mathématiques coefficient 3, le Français coefficient 3, l'Histoire-Géographie coefficient 2 et les Langues coefficient 1.
Sa moyenne est de 9.7. Il lui reste à passer l'épreuce déducatioin physique coefficient 2. Quelle note minimale doit-il obtenir pour que sa moyenne finale soit supérieure à 10 ?

## 2.3 Médiane

**Définition 3.** La médiane (notée $Me$) est la valeur du caractère qui partage la population en deux partie de même effectif.
Autrement dit, 50% des individus ont un caractère de valeur inférieur ou égale à la médiane et 50% des individus ont un caractère de valeur supérieur ou égale à la médiane.

**Déterminer la médiane dans le cas discret :**

– On range les valeurs du caractères par ordre croissant.

- On calcule les effectifs cumulés croissants *(ou fréquences cumulées croissantes)*
- Si l'effetif $N$ est impair, la médiane est la valeur du $\dfrac{N+1}{2}$ individu (il y a autant de personnes avant lui que après lui).
- Si $N$ est pair, la médiane est la moyenne des valeurs des $\dfrac{N}{2}$ et $\dfrac{N}{2}+1$ individus.

*Exemple :*

| Pointure des filles de $6^{ème}$A | 31 | 32 | 33 | 34 | 35 | Total |
|---|---|---|---|---|---|---|
| Effectif | 2 | 13 | 10 | 4 | 2 | |
| Effectif cumulé croissant | | | | | | |

Le $16^{ème}$ individu chausse du 33, donc la valeur médiane est 33.

| Pointure des filles de $6^{ème}$B | 31 | 32 | 33 | 34 | 35 | Total |
|---|---|---|---|---|---|---|
| Effectif | 2 | 10 | 10 | 4 | 2 | |
| Effectif cumulé croissant | | | | | | |

Le $14^{ème}$ individu chausse du 33, le $15^{ème}$ individu chausse du 33, donc la valeur médiane est 33.

| Pointure des filles de $6^{ème}$C | 31 | 32 | 33 | 34 | 35 | Total |
|---|---|---|---|---|---|---|
| Effectif | 2 | 13 | 10 | 4 | 1 | |
| Effectif cumulé croissant | | | | | | |

Le $15^{ème}$ individu chausse du 32, le $16^{ème}$ individu chausse du 33, donc la valeur médiane est 32.5.

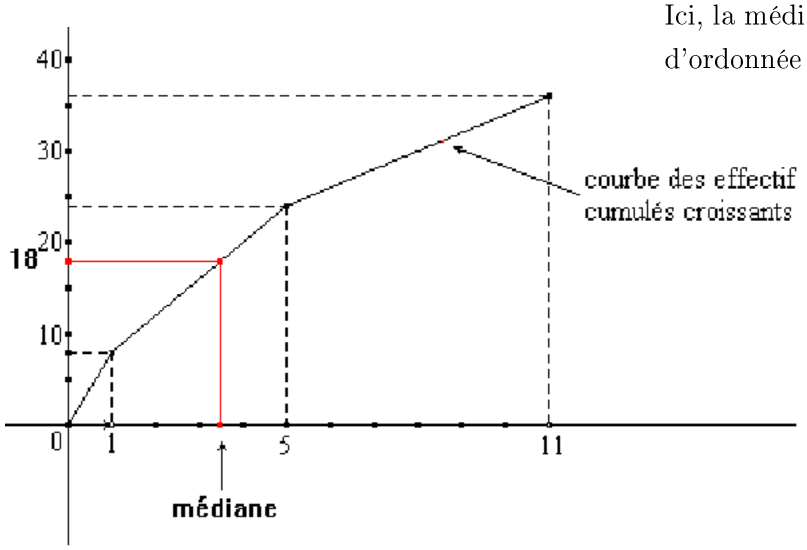
**Exercice 2.8.** Trouver la valeur médiane dans l'étude de la taille des bébés.

**Déterminer la médiane dans le cas continu :** Soit on trouve la classe médiane sur le même principe que ci-dessus, soit on trouve une valeur médiane comme ci-après. *Sur un exemple* : On étudie la distance domicile-lycée des élèves d'une classe.

- On range les classes dans « ordre croissant »
- On construit le tableau des effectifs cumulés croissants *(ou celui des fréquences cumuluées croissantes)*

| Distance (en km) | $[0;1[$ | $[1;5[$ | $[5;11[$ | Total |
|---|---|---|---|---|
| Effectifs cumulés | 8 | 24 | 36 | |

- On place dans un repère orthogonal le point $(0;0)$ puis les points $(1;8)$, $(5;24)$, $(11;36)$.
- On admet que la répartition dans chaque classe est uniforme. Ainsi on joint ces points par des segments. *On dit que la courbe obtenue est représentative d'une fonction affine par morceaux.*
- Si l'effetif $N$ est impair, la médiane est l'abscisse du point de la courbe d'ordonnée $\dfrac{N+1}{2}$.
- Si $N$ est pair, la médiane l'abscisse du point de la courbe d'ordonnée $\dfrac{N}{2}$ et $\dfrac{N}{2}+1$ individus.




Ici, la médiane est l'abscisse du point de la courbe d'ordonnée 18.5. On trouve $Me \simeq 3.5$ km.

**Remarque :** La médiane ne peut pas se retrouver à partir de médianes de sous-groupes.

**Exercice 2.9.** Déterminer la classe modale, un encadrement de l'étendue, une valeur approchée de la moyenne ainsi que la classe à laquelle appartient la médiane pour les séries suivantes :

| Valeur | $[1;3[$ | $[3;5[$ | $[5;7[$ | $[7;9[$ |
|---|---|---|---|---|
| Effectif | 5 | 10 | 15 | 2 |

| Valeur | $[0;50[$ | $[50;100[$ | $[100;200[$ | $[200;500[$ |
|---|---|---|---|---|
| Effectif | 10 | 50 | 0 | 30 |

**Exercice 2.10.** Dans un pays où la taxe d'habitation est proportionnelle à la superficie de l'habitation, la répartition des fréquences est la suivante :

| Superficie en m$^2$ | Pourcentage |
|---|---|
| $[10;40[$ | 7% |
| $[40;70[$ | 12% |
| $[70;100[$ | 27% |
| $[100;120[$ | 32% |
| $[120;140[$ | 16% |
| $[140;170[$ | 6% |

On suppose que dans chacune des classes, la répartition des superficies est régulière.

1. Un député propose d'exonérer la moitié des personnes : celles dont les habitations ont les superficies les plus faibles.
   Une personne dont l'appartement mesure 80 m$^2$ serait-elle exonérée ? Une personne dont l'appartement mesure 110 m$^2$ serait-elle exonérée ?
2. Un autre député propose d'exonérer le quart des personnes : celles dont les habitations ont les superficies les plus faibles.
   Une personne dont l'appartement mesure 80 m$^2$ serait-elle exonérée ?

**Exercice 2.11.** Le tableau suivant indique la population (en millions d'habitants) et la densité de population (en hab/km$^2$) des pays du Proche-Orient.

| Pays | Population | Densité |
|---|---|---|
| Arabie Saoudite | 20.9 | 9.7 |
| Bahreïn | 0.7 | 700 |
| Emirats Arabes Unis | 2.8 | 33.3 |
| Egypte | 66.9 | 66.8 |
| Iran | 66.2 | 40.1 |
| Irak | 22.5 | 51.8 |
| Israël | 6.1 | 290.4 |

| Pays | Population | Densité |
|---|---|---|
| Jordanie | 4.7 | 47.9 |
| Koweït | 2.1 | 116.6 |
| Liban | 4.1 | 410 |
| Oman | 2.5 | 11.7 |
| Qatar | 0.5 | 45.4 |
| Syrie | 16 | 86.4 |
| Yemen | 16.4 | 31 |

1. On considère la série statistique des populations.
   - Calculer la moyenne, la médiane et l'étendue de cette série.
   - Quel(s) est (sont) le(s) pays dont la population est la plus voisine de la moyenne ?
   - Quel(s) est (sont) le(s) pays dont la population est la plus voisine de la médiane ?
2. On considère la série des densités. Répondre pour cette série à la question 1).
3. Pour la série des populations, le Bahrein et Quatar d'une part, l'Égypte et l'Iran d'autre part, ont des valeurs exceptionnelles. Calculer :
   - La moyenne de la série des populations, élaguée de ces quatre valeurs ;
   - L'étendue de cette nouvelle série ;
   - La médiane de cette nouvelle série.
4. - Pour la série des densités, calculer la moyenne élaguée de la densité de Bahrein, ainsi que la médiane de cette nouvelle série.
   - Calculer la différence entre moyenne élaguée et la moyenne initiale, puis entre cette médiane et la médiane initiale.
   - Quels commentaires vous inspirent ces résultats ?
5. Déterminer la population totale du Proche-Orient.
6. Calculer la superficie de chaque pays. Déterminer alors la superficie totale des pays du Proche-Orient.
7. En déduire la densité des pays du Proche-Orient.

## 2.4 Regroupement par classes (facultatif)

# 3 Repésentation graphique

## 3.1 Diagramme en bâton

On l'utilise pour représenter graphiquement une série statistique dont le caractère est discret (qualitatif). On représente sur l'axe des abscisses les différentes valeurs du caractère dans l'ordre croissant (aléatoirement pour un caractère qualitatif) et sur l'axe des ordonnées les effectifs.

## 3.2 Histogramme

On l'utilise pour représenter graphiquement une série statistique dont le caractère est continue. Il s'agit d'un diagramme en colonnes dans lequel l'aire de chaque colonne est proportionnellen à l'effectif de la classe.

## 3.3 Diagramme circulaire

On l'utilise pour reprśenter graphiquement une série statistique dont le caractère est discret ou qualitatif. L'angle d'ouverture de chaque secteur est proportionnel à l'effectif :

$$\text{Angle} = \frac{\text{Effectif de la valeur}}{\text{Effectif total}} \times 360 = \text{Fréquence} \times 360$$

**Exercice 3.1.** Représenter avec le graphique approprié les caractéres étudiés sur les bébés nés en 2008.

# 4 Echantillonnage et simulation

## 4.1 Définitions

**Définition 4.** Une **expérience aléatoire** est une expérience dont on peu d'écrire l'ensemble des résultats possibles à priori, mais dont on ne peut prédire le résultat avant qu'elle ne se réalise.

**Définition 5.** Un résultat possible d'une expérience est appelé **issue**.

**Définition 6.** On appelle **échantillon** d'une expérience aléatoire un ensemble d'issues obtenues par la répétition de cette expérience. Le nombre d'issues est la **taille** de l'échantillon.

## 4.2 Exemple : Lancer d'un dé

### 4.2.1 L'expérience

Act

1. Effectuer trois séries de 15 lancers d'un dé à 6 faces et consigner les résultats dans les tableaux ci-dessous. Série 1 :

| Issues | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| Effectifs | | | | | | |
| Fréquences | | | | | | |

Série 2 :

| Issues | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| Effectifs | | | | | | |
| Fréquences | | | | | | |

Série 3 :

| Issues | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| Effectifs | | | | | | |
| Fréquences | | | | | | |

2. En utilisant 3 couleurs différentes pour chaque série, représenter les fréquences sur un même diagramme en bâtons.
3. Noter toutes les remarques que vous inspirent les résultats de cette expérience.

   Correspondent-ils à ce que l'on pouvait prévoir ?
4. Rassembler ensuite tous les résultats dans le tableau ci-dessous.

| Issues | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| Effectifs | | | | | | |
| Fréquences | | | | | | |

5. Comparer les fréquences de ce tableau avec celles des précédents
6. Regrouper maintenant les résultats obtenus séparément par chacun des élèves dans le tableau ci-dessous :

| Issues | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| Effectifs | | | | | | |
| Fréquences | | | | | | |

7. Que constatez-vous ?

Conclusion : On constate que les fréquences varient selon les expériences et le nombre d'essais, on parle de **fluctuation d'échantillonnage**.
De plus, on observe que la fréquence de sortie de chacun des chiffres est très voisine du nombre $\frac{1}{6}$ quand on effectue un grand nombre de lancers.

**Remarques :**

– Le grand nombre de lancers n'a pu être obtenu qu'avec la participatioin de tous les élèves. Un seul aurait dû y passer beaucoup de temps, d'où la nécessité d'un moyen de simulation.

- La simulation remplace l'expérience et permet d'étudier des séries statistiques comportant un très grand nombre de valeurs.
- Pour simuler une expérience aléatoire, on peut utiliser des tableurs ou des calculatrices.

### 4.2.2 Simulation sur tableur

Les statistiques nous permettent d'étudier les données récolter lors d'enquêtes. Il s'agit d'études *à postériori*.

Cependant, certaines expériences sont très longues à réaliser, comme celle que nous avons faites tous ensemble. On préfère, lorsque c'est possible, les simuler, dans notre cas sur un tableur.

*Travail de l'élève :* TP Info : Simulation d'un lancer de dé

### 4.2.3 Simulation sur calculatrice

# Les Annexes

## Vocabulaire

# 1 Vocabulaire

## 1.1 Population

On considère cet exemple dans toute cette partie du cours :

**Au cours d'une enquête portant sur les bébés nés en 2008, on s'intéresse à la taille arrondie au cm, la couleur de leurs yeux et le temps quotidien de sommeil par intervalle d'une heure.**

*Les premières études étant démographiques, on en a gardé le vocabulaire.*

L'ensemble sur lequel porte une étude est appelée . . . . . . . . . . . . . . . . . .
Un élément de cet ensemble est un . . . . . . . . . . . . . . . . . .
On observe alors des propriétés, appelées . . . . . . . . . . . . . . . . . ., sur les individus de cette population.
L'. . . . . . . . . . . . . . . . . . . . . . . est le nombre d'individus de la population.

*Exemple :* Donner la population et les caractères étudiés pour notre enquête.

## 1.2 Caractères

Lorsque le caractère étudié prend des valeurs numériques, le caractère est dit . . . . . . . . . . . . . . . . . . Il peut alors être :

- . . . . . . . . . . . . . . . . . lorsqu'il ne prend que des valeurs isolées
- . . . . . . . . . . . . . . . . . lorsqu'il peut prendre toutes les valeurs d'un intervalle, appelé . . . . . . . . . . . .

Si le caractère n'est pas . . . . . . . . . . . . . . . . . ., on dit qu'il est . . . . . . . . . . . . . . . . . . On appelle ses valeurs des . . . . . . . . . . . . . . . . . .

*Exemple :* Préciser la nature de chacun des caractères étudiés.

L'. . . . . . . . . . . . . . . . . . d'une valeur d'un caractère est le nombre d'individus ayant cette caleur.

La . . . . . . . . . . . d'une valeur est la proportion d'individus possédant cette valeur.

$$\ldots\ldots\ldots\ldots\ldots\ldots = \frac{\ldots\ldots\ldots\ldots\ldots\ldots}{\ldots\ldots\ldots\ldots\ldots\ldots}$$

**Remarques :**

- La somme des fréquences est toujours égale à . . . .
- On assimile parfois la fréquence à un pourcentage. Dans ce cas, la somme des fréquence vaut toujours . . .

Dans le cas quantitatif, l'. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . d'une valeur (ou d'une classe) est le nombre d'individus ayant une valeur (ou une classe) inférieur ou égale à celle considérée. On lui associe une . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .

La . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . est l'ensemble des données collectées, souvent présentées dans un tableau, et/ou sous forme graphique.

*Exemple :* Compléter les tableaux suivants

| Taille en cm | 44 | 45 | 46 | 47 | 48 | 49 | 50 | 51 | 52 | 53 | 54 | 55 | 56 | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Effectif | 13 | 2 | 0 | 15 | 24 | 33 | 57 | 46 | 32 | 40 | 29 | 6 | 3 | |
| Fréquence à $10^{-2}$ près | | | | | | | | | | | | | | |
| Effectif cumulé croissant | | | | | | | | | | | | | | |
| Fréquence cumulée croissante | | | | | | | | | | | | | | |

| Temps de sommeil en heures | $[18;19[$ | $[19;20[$ | $[20;21[$ | $[21;22[$ | $[22;23[$ | Total |
|---|---|---|---|---|---|---|
| Effectif | 47 | 33 | 74 | 72 | 54 | |
| Fréquence à $10^{-2}$ près | | | | | | |
| Effectif cumulé croissant | | | | | | |
| Fréquence cumulée croissante | | | | | | |

# 2 Paramètres d'une série quantitative

## 2.1 Étendue et mode

L'.................. de la série statistique pour un caractère quantitatif est égale à la valeur absolue de la différence entre les valeurs extrêmes prises par le caractère.

Le .................. (ou .................. dans le cas continu quand les classes sont de même amplitude) est la valeur du caractère ayant le plus grand effectif.

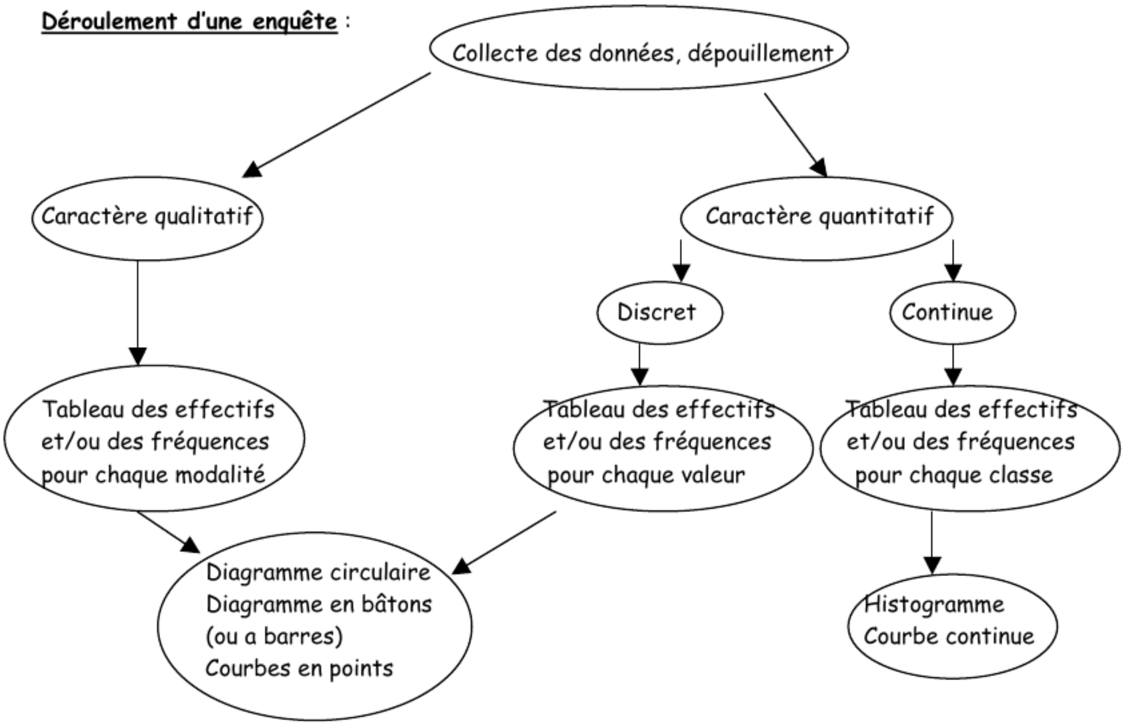
*Exemple :* Donner l'étendue et le mode relatifs au caractère taille puis l'étendue et la classe modale relatifs au temps de sommeil.

## 2.2 Moyenne

**Définition 1.** On considère une série statistique quantitative sur une population de $N$ individus. On note $x_1$, $x_2$, ..., $x_p$ les valeurs prises par le caractère et $n_1$, $n_2$, ..., $n_p$ les effectifs associés (on a toujours $p \leq N$).
La moyenne est le nombre $\bar{x} = \frac{\cdots\cdots\cdots}{\cdots\cdots\cdots} = \frac{\cdots\cdots\cdots}{\cdots\cdots\cdots}$

**Remarque :** Dans le cas continu, on prend pour les $x_i$ le ......................................

*Exemple :* Donner la moyenne de la taille et celle du temps de sommeil.




*Exemple :*

| Pointure des filles de $6^{\text{ème}}$A | 31 | 32 | 33 | 34 | 35 | Total |
|---|---|---|---|---|---|---|
| Effectif | 2 | 13 | 10 | 4 | 2 | |
| Effectif cumulé croissant | | | | | | |

Le $\ldots\ldots^{\text{ème}}$ individu chausse du ......, donc la valeur médiane est .......

| Pointure des filles de $6^{\text{ème}}$B | 31 | 32 | 33 | 34 | 35 | Total |
|---|---|---|---|---|---|---|
| Effectif | 2 | 10 | 10 | 4 | 2 | |
| Effectif cumulé croissant | | | | | | |

Le $\ldots\ldots^{\text{ème}}$ individu chausse du ......, le $\ldots\ldots^{\text{ème}}$ individu chausse du ......, donc la valeur médiane est .......

| Pointure des filles de $6^{\text{ème}}$C | 31 | 32 | 33 | 34 | 35 | Total |
|---|---|---|---|---|---|---|
| Effectif | 2 | 13 | 10 | 4 | 1 | |
| Effectif cumulé croissant | | | | | | |

Le $\ldots\ldots^{\text{ème}}$ individu chausse du ......, le $\ldots\ldots^{\text{ème}}$ individu chausse du ......, donc la valeur médiane est .......

## Exercices

**Exercice 2.1.** On effectue les pesées de 40 judokas avant une compétition. Leur poids moyen est de 72 kg. Mais la balance indique 500 g de moins que le poids réel. Quel est le poids moyen réel des judokas ?

**Exercice 2.2.** Après six contrôles, un élève obtient 12 de moyenne, puis il obtient 15 au septième contrôle. Tous les contrôles ont le même coefficient. Quelle est sa nouvelle moyenne ?

**Exercice 2.3.** On a déterminé la moyenne de 560 nombres à la calculatrice, on a trouvé 115. Mais on s'aperçoit que l'on a oublié l'un des nombres, à savoir 171. Quelle est la moyenne des 561 nombres ?

**Exercice 2.4.** Une épreuve d'examen est constituée de deux parties indépendantes A et B. Un professeur corrige la partie A et un autre la partie B. La note totale à cette épreuve s'obtient en ajoutant la note obtenue à la partie A et la note obtenue à la partie B. Pour la partie A, la moyenne des notes est égale à 9 et pour la partie B à 7. Quelle est la moyenne des notes à cette épreuve ?

**Exercice 2.5.** Dans deux entreprises A et B, les moyennes des salaires masculins et féminins ainsi que la répartition hommes/ femmes sont données par les tableaux suivants :

| Salaire moyens en euros | A | B |
|---|---|---|
| Hommes | 1400 | 1500 |
| Femmes | 1000 | 1100 |

| Répartition | A | B |
|---|---|---|
| Hommes | 50% | 20% |
| Femmes | 50% | 80% |

Pour chaque catégorie (hommes ou femmes), l'entreprise B paye mieux et pourtant ...
Calculer pour chaque entreprise la moyenne des salaires pour l'ensemble des employés.
Quels commentaires pouvez-vous faire ?

**Exercice 2.6.** La moyenne de 5 notes d'un élève est de 12. Les quatre pemières sont 13, 10, 8 et 15. Quelle est la cinquième ?

**Exercice 2.7.** Un candidat a passé les quatre premières épreuves suivantes : les Mathématiques coefficient 3, le Français coefficient 3, l'Histoire-Géographie coefficient 2 et les Langues coefficient 1. Sa moyenne est de 9.7. Il lui reste à passer l'épreuve d'éducation physique coefficient 2.
Quelle note minimale doit-il obtenir pour que sa moyenne finale soit supérieure à 10 ?

**Exercice 2.8.** Trouver la valeur médiane dans l'étude de la taille des bébés.

**Exercice 2.9.**

Dans un pays où la taxe d'habitation est proportionnelle à la superficie de l'habitation, on a la répartition des fréquences ci-contre.
On suppose que dans chacune des classes, la répartition des superficies est régulière.

| Superficie en $m^2$ | Pourcentage |
|---|---|
| $[10; 40[$ | 7% |
| $[40; 70[$ | 12% |
| $[70; 100[$ | 27% |
| $[100; 120[$ | 32% |
| $[120; 140[$ | 16% |
| $[140; 170[$ | 6% |

1. Un député propose d'exonérer la moitié des personnes : celles dont les habitations ont les superficies les plus faibles. Une personne dont l'appartement mesure 80 $m^2$ serait-elle exonérée ? Une personne dont l'appartement mesure 110 $m^2$ serait-elle exonérée ?
2. Un autre député propose d'exonérer le quart des personnes : celles dont les habitations ont les superficies les plus faibles. Une personne dont l'appartement mesure 80 $m^2$ serait-elle exonérée ?

**<u>Exercice</u> 2.10.** Le tableau suivant indique la population (en millions d'habitants) et la densité de population (en hab/$km^2$) des pays du Proche-Orient.

| Pays | Population | Densité |
|---|---|---|
| Arabie Saoudite | 20.9 | 9.7 |
| Bahreïn | 0.7 | 700 |
| Emirats Arabes Unis | 2.8 | 33.3 |
| Egypte | 66.9 | 66.8 |
| Iran | 66.2 | 40.1 |
| Irak | 22.5 | 51.8 |
| Israël | 6.1 | 290.4 |

| Pays | Population | Densité |
|---|---|---|
| Jordanie | 4.7 | 47.9 |
| Koweït | 2.1 | 116.6 |
| Liban | 4.1 | 410 |
| Oman | 2.5 | 11.7 |
| Qatar | 0.5 | 45.4 |
| Syrie | 16 | 86.4 |
| Yemen | 16.4 | 31 |

1. On considère la série statistique des populations.
   - Calculer la moyenne, la médiane et l'étendue de cette série.
   - Quel(s) est (sont) le(s) pays dont la population est la plus voisine de la moyenne ?
   - Quel(s) est (sont) le(s) pays dont la population est la plus voisine de la médiane ?
2. On considère la série des densités. Répondre pour cette série à la question 1).
3. Pour la série des populations, le Bahrein et Quatar d'une part, l'Égypte et l'Iran d'autre part, ont des valeurs exceptionnelles. Calculer :
   - La moyenne de la série des populations, élaguée de ces quatre valeurs ;
   - L'étendue de cette nouvelle série ;
   - La médiane de cette nouvelle série.
4. - Pour la série des densités, calculer la moyenne élaguée de la densité de Bahrein, ainsi que la médiane de cette nouvelle série.
   - Calculer la différence entre moyenne élaguée et la moyenne initiale, puis entre cette médiane et la médiane initiale.
   - Quels commentaires vous inspirent ces résultats ?
5. Déterminer la population totale du Proche-Orient.
6. Calculer la superficie de chaque pays. Déterminer alors la superficie totale des pays du Proche-Orient.
7. En déduire la densité des pays du Proche-Orient.

## Plus on est de fous ...

1. Effectuer trois séries de 15 lancers d'un dé à 6 faces et consigner les résultats dans les tableaux ci-dessous. Série 1 :

| Issues | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| Effectifs | | | | | | |
| Fréquences | | | | | | |

Série 2 :

| Issues | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| Effectifs | | | | | | |
| Fréquences | | | | | | |

Série 3 :

| Issues | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| Effectifs | | | | | | |
| Fréquences | | | | | | |

2. En utilisant 3 couleurs différentes pour chaque série, représenter les fréquences sur un même diagramme en bâtons.
3. Noter toutes les remarques que vous inspirent les résultats de cette expérience.

   Correspondent-ils à ce que l'on pouvait prévoir ?
4. Rassembler ensuite tous les résultats dans le tableau ci-dessous.

| Issues | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| Effectifs | | | | | | |
| Fréquences | | | | | | |

5. Comparer les fréquences de ce tableau avec celles des précédents
6. Regrouper maintenant les résultats obtenus séparément par chacun des élèves dans le tableau ci-dessous :

| Issues | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| Effectifs | | | | | | |
| Fréquences | | | | | | |

7. Que constatez-vous ?

# Rudiments sur les tableurs

Un tableur est un logiciel informatique capable de manipuler des feuilles de calcul. Par exemple, Excel est un tableur. On peut s'en servir comme d'une calculatrice, mais on peut également effectuer des tâches variées et plus complexes, telles que la gestion de bases de données, la production de graphiques (grâce à l'asitant graphique), diverses analyses et simulations statistiques, comme nous allons le voir.

Voici quelques bases pour utiliser un tableur :

- Une case s'appelle une ***cellule***, que l'on désigne par sa lettre de colonne puis son numéro de ligne.
  *Exemple :* La première cellule en haut à gauche s'appelle $A1$.

- Pour **désigner des cellules**, on a plusieurs méthodes.
  *Exemple :* « $A1; F4$ »désigne seulement les cellules $A1$ et $F4$ (donc deux cellules)
  « $A1 : F4$ »désigne les cellules entre $A1$ et $F4$, soit 24 cellules.

- Un tableur peut **effectuer des opérations** sur les cellules. Pour cela, il faut *toujours* lui préciser qu'il y a un calcul à effectuer en commençant la cellule où l'on veut faire le calcul par « $=$ ».
  *Exemple :* Dans $B2$, on note « $= 6 + 4$ »et il s'affiche 10 dans $B2$
  Si on oublie le « $=$ », il sera simplement écrit « $6 + 4$ ».
  Si on écrit « $= A4 + C2$ », il s'affichera la somme des cellules $A4$ et $C2$.

- Un tableur possède des **fonctions prédéfinies** (disponibles et décrites dans l'icône $f(x)$).
  *Exemples :* « $= SOMME(A1; B2; A4)$ »fait la somme des trois cellules écrites.
  « $= MOYENNE(A1 : A4)$ »fait la moyenne des cellules entre $A1$ et $A4$.
  « $= ALEA()$ »renvoit un nombre au hasard de $[0; 1[$.
  « $= ENT(x)$ »Donne la partie entière du nombre $x$
  « $= SI($ test ; valeur1 ; valeur2 $)$ »renvoit la valeur1 si le test est vrai, et valeur2 sinon.
  « $= SI(A10 = 4; 5; 7)$ »renvoit 5 si la cellule $A10$ vaut 4, 7 sinon.
  « $= NB.SI($plage de données ; critères $)$ »compte le nombre de cellules de la plage valant le critère.
  « $= NB.SI(A2 : J11; 2)$ »renvoit le nombre de cellules de la plage ayant la valeur 2.

- Un tableur est capable de copier automatiquement des formules.
  *Exemple :*
  - **En incrémentant des indices :** Si on rentre en $A6$ « $= MOYENNE(A1 : A4)$ »et que l'on copie cette cellule en $B6$ (on clique sur le coin en bas à gauche de la cellule et on la glisse en $B6$), alors le tableur affiche en $B6$ la moyenne des cellules $B1$ à $B4$.
  - **Sans incrémenter d'indices** : parfois, on ne veut pas que la formule change avec les cellules. Il faut alors le préciser au tableur en notant un dollars devant ce que l'on ne souhaite pas changer (les lettres et / ou les numéros de ligne). Ainsi si l'on rentre en $A7$ « $= MOYENNE(\$A1 : \$A4)$ »et que l'on copie cette cellule en $B7$ , le tableur affiche la même formule qu'en $A7$. Par contre, si on la copie en $B8$, il affichera la moyenne des cellules comprises entre $A2$ et $A6$

# Exemples : Lancer d'un dé

On veut simuler un échantillon de 100 lancers de dés puis étudier les résultats obtenus.

L'issue d'un lancer est simulée par le choix ........................
On doit donc trouver un moyen pour que le tableur fasse ce choix tout seul et ...... fois.
Ensuite, on s'intéressera aux fréquences d'apparition de chacune des issues possibles. On devra donc récuperer les effectifs d'apparition de chaque issue. Puis, on calculera les fréquences d'apparition.

Pour être clair, on présentera les résultats dans des tableaux sur la même page et sur un graphique, comme présenté au tableau.

1. Grâce aux fonctions prédéfinies du tableur données en exemple la page précédente, proposer une formule pour simuler un lancé de dé.
2. Créer le tableau des 100 simulations.
3. Créer le tableau d'analyse de données.
4. Représenter graphiquement les fréqunces de chacune des issues possibles. Pour cela :
   (a) Sélectionner ces colonnes
   (b) Dans l'assistant graphique, choisir histogramme et suivre l'assitant
5. (a) Les fréquences de sortie de chaque numéro sont-elles égales ?
   (b) Réaliser de nouvelles simulations. Les résultats varient-ils ?
   (c) Quel est le terme approprié pour décrire ce phénomène ?

## Simulation sur des échantillons plus grands

Changer de feuille de calcul. Refaire le même type d'exercice, mais pour une simulatioin de 1000 lancers de dés, puis 5000.

1. Comparer les fréquences obtenues pour les trois échantillons.
2. Quelle remarque pouvez-vous faire ?
3. Vers quelle valeur semblent se stabiliser les fréquences ?
4. Est-ce en accord avec l'intuition sur les chances de sorite d'un numéro ?

## Simulation de la somme de deux dés

Une expérience consiste à lancer deux dés et à observer la somme des faces obtenues. Simuler 1000 lancers de deux dés avec un tableur. Étudier la distribution des fr{equences obtenus par des méthodes analogue à celles de la partie précédentes.

# Devoir Surveillé 8

**Exercice 3.1.** (5 points)
Lors du troisième trimestre, un élève a une moyenne de 11 après les quatre premiers devoirs. Tous les devoirs du trimestre ont le même coefficient.

1. Cet élève obtient 14 au cinquième devoir. Quelle est sa nouvelle moyenne ?
2. L'élève effectue alors un sixième devoir qui lui permet d'obtenir une moyenne de 12. Quelle est cette sixième note ?
3. La moyenne des six devoirs au troisième trimestre est donc de 12. On sait qu'au premier trimestre, il y avait eu quatre devoirs et que la moyenne de l'élève était de 10.5. On sait aussi qu'au deuxième trimestre il y avait eu cinq devoirs et que la moyenne de l'élève était de 13.
   (a) Quelle est la moyenne des devoirs sur l'année ?
   (b) Est-ce la même que la moyenne des moyennes des trois trimestres ?
4. On décide d'augmenter chaque note du troisième trimestre de 1.
   (a) De combien augment la moyenne des devoirs au troisième trimestre ?
   (b) De combien augmente la moyenne des devoirs sur l'année ?

**Exercice 3.2.** (6 points)
La capacité vitale est le volume d'air maximal pouvant être mobilisé en une seule inspiration.
Sur un échantillon de 17 personnes, on a mesuré la capacité vitale en litres. Voila la liste des résultats :

4.15 – 4.48 – 5.24 – 4.8 – 4.95 – 4.05 – 4.3 4.7 – 5.51

4.58 – 4.12 – 5.7 – 4.85 – 5.05 – 4.65 – 4.7 – 4.28

1. Déterminer l'étendue et la moyenne de cette série. Arrondir la moyenne au centilitre près.
   *Pour la moyenne, on utilisera la calculatrice sans explication*
2. En expliquant la méthode utilisée, déterminer la médiane de cette série.
3. On décide de regrouper les valeurs de la série par classes.
   Compléter le tableau suivant :

| Capacité vitale (en L) | $[4;4.5[$ | $[4.5;5[$ | $[5;5.5[$ | $[5.5;6[$ | Total |
|---|---|---|---|---|---|
| Effectifs | | | | | |
| Effectifs cumulés croissants | | | | | |

4. A l'aide de cette répartition par classes, déterminer la moyenne des valeurs.
5. On admet que dans chaque classe, la répartition est uniforme. Tracer alors la courbe des effectifs cumulés croissants (échelle : 2 cm pour 0.5 L en abscisse, 2 cm pour 5 personnes en ordonnées, on commencera la graduation de l'axe des abscisses à 4.)
6. En déduire graphiquement la médiane de ces valeurs.

**<u>Exercice</u> 3.3.** (6 points)

On lance deux dés équilibrés à six faces et on fait le produit des résultats obtenus.

Si le résultat est un nombre pair, on perd 2 €. Par contre, si le résultat est impair, on gagne 5 €

Le but de l'exercice est d'étudier si ce jeu est intéressant pour le joueur.

1. Laquelle de ces formules permet de simuler le produit des résultats de deux dés ?

   $= ENT(24*ALEA()+1)$ $\qquad$ $= ENT(36*ALEA()+1)$ $\qquad$ $= ENT(6*ALE()+1)*ENT(6*ALEA()+1)$

2. Une simulation de 100 parties avec un tableur a donné les résultats suivants :

| Gain du joueur | $-2$ € (pair) | 5 € (impair) | Total |
|---|---|---|---|
| Effectifs | 70 | 30 | 100 |

   Calculer avec le tableau ci-dessus le gain moyen du joueur.

   D'après cette simulation, le jeu est-il favorable au joueur ?

3. Compléter le tableau suivant, donnant tous les produits possibles en lançant deux dés à six faces :

| | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 1 | | | | | | |
| 2 | | | | | | |
| 3 | | | | | | |
| 4 | | | | | | |
| 5 | | | | | | |
| 6 | | | | | | |

4. En déduire la distribution des fréquences théoriques pour les résultats pairs et impairs :

| Gain du joueur | $-2$ € (pair) | 5 € (impair) | Total |
|---|---|---|---|
| Effectifs | | | 1 |

   Calculer le gain moyen du joueur à parit des fréquences théoriques.

   D'après la théorie, le jeu est-il favorable au joueur ?

5. Quel phénomène permet d'expliquer la différence entre la conclusion obtenue après la simulation et la conclusion obtenue après l'étude théorique ?

**<u>Exercice</u> 3.4.** Questions de réflexion (3 points)

On expliquera les réponses.

1. Si on lance six fois un dé, on obtient systématiquemnt une fois le 1, une fois le 2, une fois le 3, une fois le 4, une fois le 5 et une fois le 6 ?
2. En jouant à « Pile ou Face », on a déjà eu 5 fois de suite le côté Pile. Au sixième coup, on aura plutôt Pile, Face, ou autant de chances pour Pile que pour Face ?
3. Lequel de ces événements est le plus probable ? 9 fois Pile sur 10 lancers de pièce, 90 fois Pile sur 100 lancers de pièce,ou 900 fois Pile sur 1000 lancers de pièce ?