

### Exercices de TD de statistiques

#### Rappels

Les exercices 1 à 4 sont des rappels portant sur la prise en main de la calculatrice. Ils sont corrigés en ligne.

#### Exercice 1 : Fractions

Si le résultat d'un calcul est un nombre à virgule, vous pouvez soit demander à la calculatrice de l'afficher sous forme décimale (nombre à virgule) soit sous forme de fraction (avec la touche  $\boxed{F\leftrightarrow D}$  pour une CASIO, ou la fonction  $\blacktriangleright$  **Frac** du menu  $\boxed{MATH}$  pour une TI).

- Grâce à cette fonctionnalité de la calculatrice, simplifiez chacune des fractions suivantes :  $\frac{76-34}{21}$ ,  $\frac{-77}{44}$ ,  $\frac{14}{15-7}$ ,  $\frac{-24}{-18}$ ,  $\frac{-10-81}{39 \times 3}$ ,  $\frac{-10}{35}$ .
- Classez-les dans l'ordre croissant ("du plus petit au plus grand").

#### Exercice 2 : Utilisation de la calculatrice

- La capture d'écran suivante indique un calcul effectué sur une calculatrice.



Peut-on en conclure

- que le carré de  $-1$  est  $-1$  ?
- que le carré de  $1$  est  $-1$  ?
- que  $-1$  est l'opposé du carré de  $1$  ?

De même, pour chacune des questions ci-dessous, une capture d'écran montre un calcul effectué sur la calculatrice, et trois réponses sont proposées qui interprètent (de manière correcte ou erronée) ce calcul. À chaque fois, une seule des trois réponses proposées est correcte, à vous de déterminer laquelle.

2.	$1 \div (2023 \times 2024)$ $2.442269630E-7$	(a) $\frac{1}{2023 \times 2024} \simeq 2,44$ (b) $\frac{1}{2023 \times 2024} \simeq 24422696,3$ (c) $\frac{1}{2023 \times 2024} \simeq 0,000000244$
3.	$30 - 33 \times X$ $X \times 2024$ $-3$ $-6072$	(a) $30 - 33 \times 2024 = -6072$ (b) $3 \times 2024 = -6072$ (c) $(30 - 33) \times 2024 = -6072$
4.	$-10 \times 45$ $Ans \div$ $-450$ $202500$	(a) $-10 \times 45^2 = 202500$ (b) $-450^2 = 202500$ (c) $(-10 \times 45)^2 = 202500$

#### Exercice 3 : Puissances de dix

Calculez (de préférence "de tête") chacune des expressions suivantes :

$26 \div 100$	$0,26 \times 10^2$	$0,189\,794\,312 \times 1\,000\,000$
$18\,687 \times 10^{-3}$	$150\,961 \times 10^{-2}$	$15\,650 \times 10^{-2}$
$1\,472,1 \times 100\,000\,000$	$1\,536 \times 10^{-9}$	$118\,510,05 \times 10^9$

#### Exercice 4 : Règles de calcul

- On considère les expressions suivantes :
  - $2 \times 8 + (9 + (5 + 3) \times 4 \times 4)$
  - $9 + 2 \times 8 + (4^2 \times 5 + (3))$
  - $(9 + (2 \times 8)) + 4^2(5 + 3)$
  - $9 \times 8 + 2 \times 8 + 4^2(5 + 3)$
  - $9 + 2 \times 8 + 4^2 \times 5 + 3$
  - $(9 + 2) \times 8 + (5 + 3) \times 4^2$
 Calculez la valeur de chacune de ces expressions, et observez que certaines sont égales.
- De même, calculez les expressions suivantes lorsque les variables prennent les valeurs indiquées, et observez que certaines d'entre-elles sont égales.

(a) Les expressions

- $-9 - 8x + 6$
  - $8x + 6 - 9x$
  - $(6 - 9) - 8x$
  - $6 - 3 \times \frac{9-8}{3}x$
  - $6 - (9 - 8)x$
- lorsque  $x = 6$ .

(b) Les expressions

- $\frac{a}{4 \times 8 \div b}$
  - $\frac{1}{4} \times \frac{a}{b \times 8}$
  - $a \div 8 \times b \div 4$
  - $b \frac{a}{4 \times 8}$
  - $\frac{a \div b}{4 \times 8}$
  - $\frac{b \times a}{4 \times 8}$
- lorsque  $a = 2$  et  $b = -6$ .

(c) Lorsque  $x = -6$ ,  $y = -9$  et  $z = -6$  :

- $y \frac{(8-x)}{(7-z)}$
- $y \left( \frac{8}{7} - \frac{x}{z} \right)$
- $\frac{y \times 8}{7} - \frac{y \times x}{z}$
- $(y \times 8 - x \times y) \div (7 - z)$

(d) Lorsque  $x_1 = -3$  et  $x_2 = 2$  :

- $x_2 + 3 - 5 - 3 + 2 \times x_1 + 2 \times 6$
- $x_2 - 5 - 1(2 - x_1 - 6)$
- $x_2 - 5 + 2 \times x_1 + 2 \times 6$
- $x_2 - 5 - 2(-x_1 - 6)$
- $x_2 - 5 + 2 - x_1 - 6 - (2 - x_1 - 6)$
- $x_2 - 5 - 0(2 - x_1 - 6)$
- $x_2 - 5 - (2 - x_1 - 6)$

(e) Lorsque  $a_1 = -1$ ,  $a_2 = 4$ ,  $x_1 = 8$  et  $x_2 = -2$  :

- $(x_1 - (a_1 + a_2 - x_2))$
- $(x_2 - a_1) - (a_2 - x_1)$
- $x_1 - a_1 + a_2 - x_2$
- $a_2 - a_1 + x_1 - x_2$

## Compléments

### Exercice 5 : Précision des calculs

Anaëlle, Christopher, Marie et Justin étudient en première année de psychologie à l'Université de Bourgogne.

On leur demande de calculer l'écart-type de la taille de 3 hommes, et au vu des formules du formulaire, cela revient à calculer des nombres notés " $m(X)$ ", " $m(X^2)$ ", " $Var(X)$ " et " $s(X)$ ", donnés par les formules suivantes :

$$m(X) = \frac{1,79+1,67+1,74}{3}$$

$$m(X^2) = \frac{1,79^2+1,67^2+1,74^2}{3}$$

$$Var(X) = m(X^2) - (m(X))^2$$

$$s(X) = \sqrt{Var(X)}$$

Le nombre " $s(X)$ " issu de ce calcul est l'écart-type, exprimé (dans le cas présent) en mètre, et il est demandé de déterminer cet écart-type à 1 cm près.

Ci-dessous se trouve ce que chacun.e a écrit sur sa copie ainsi que les calculs effectués sur sa calculatrice.

#### Copie d'Anaëlle

$$m(X) = \frac{1,79+1,67+1,74}{3} \simeq 1,73$$

$$m(X^2) = \frac{1,79^2+1,67^2+1,74^2}{3} \simeq 3,01$$

$$Var(X) \simeq 3,01 - 1,73^2 \simeq 0,02$$

$$s(X) \simeq \sqrt{0,02} \simeq 0,14 \text{ mètre.}$$

Donc l'écart-type est d'environ 14 cm.

#### Calculatrice d'Anaëlle

$(1.79+1.67+1.74) \div 3$	1.733333333
$(1.79^2+1.67^2+1.74^2) \div 3$	3.006866667
$3.01 - 1.73^2$	0.0171
$\sqrt{0.02}$	0.1414213562

#### Copie de Christopher

$$m(X) = \frac{1,79+1,67+1,74}{3} = \frac{5,2}{3}$$

$$m(X^2) = \frac{1,79^2+1,67^2+1,74^2}{3} = \frac{9,0206}{3}$$

$$Var(X) \simeq \frac{9,0206}{3} - \left(\frac{5,2}{3}\right)^2 \simeq 0,0024$$

$$s(X) = \sqrt{Var} \simeq 0,05 \text{ mètre.}$$

Donc l'écart-type est d'environ 5 cm.

#### Calculatrice de Christopher

$1.79+1.67+1.74$	5.2
$1.79^2+1.67^2+1.74^2$	9.0206
$9.0206 \div 3 - (5.2 \div 3)^2$	0.002422222222
$\sqrt{Ans}$	0.04921607687

**Copie de Marie**

$$m(X) = \frac{1,79+1,67+1,74}{3} \simeq 1,73$$

$$m(X^2) = \frac{1,79^2+1,67^2+1,74^2}{3} \simeq 3,01$$

$$Var(X) = m(X^2) - (m(X))^2 \simeq 0,0024$$

$$s(X) = \sqrt{Var(X)} \simeq 0,05 \text{ mètre.}$$

Donc l'écart-type est d'environ 5 cm.

**Calculatrice de Marie**

```
(1.79+1.67+1.74)÷3=M
1.733333333
(1.79²+1.67²+1.74²)÷3=M
3.006866667
M-M²=U
0.002422222222
√(U)
0.04921607687
```

**Copie de Justin**

$$s(X) = \sqrt{\frac{1,79^2+1,67^2+1,74^2}{3} - \left(\frac{1,79+1,67+1,74}{3}\right)^2} \simeq 0,05 \text{ mètre.}$$

**Calculatrice de Justin**

```
√((1.79²+1.67²+1.74²)÷3
-((1.79+1.67+1.74)÷3)²)
0.04921607687
```

Donc l'écart-type est d'environ 5 cm.

- Écrivez avec le symbole  $\sum$  la somme suivante :  $\frac{4^2}{4+6} + \frac{5^2}{5+6} + \frac{6^2}{6+6} + \frac{7^2}{7+6} + \frac{8^2}{8+6} + \frac{9^2}{9+6} + \frac{10^2}{10+6}$ .
- Calculez  $\sum_{i=1}^6 n_i (x_i)^2$  lorsque  $x_1 = 6, x_2 = 7, x_3 = 8, x_4 = 9, x_5 = 10, x_6 = 11, n_1 = 5, n_2 = 2, n_3 = 10, n_4 = 6, n_5 = 10$  et  $n_6 = 4$  (sans utiliser de liste sur la calculatrice).

- Parmi ces différentes réponses, lesquelles vous semblent correctes ? Expliquez pourquoi l'un.e des étudiant.e.s a obtenu une réponse erronée.
- On demande ensuite de calculer l'écart-type de la taille d'un échantillon de 5 femmes. Comme l'échantillon n'est plus le même, il faut cette fois-ci faire les calculs suivants :

$$m(X) = \frac{1,54+1,55+1,65+1,62+1,64}{5}$$

$$m(X^2) = \frac{1,54^2+1,55^2+1,65^2+1,62^2+1,64^2}{5}$$

$$Var(X) = m(X^2) - (m(X))^2$$

$$s(X) = \sqrt{Var(X)}$$

Rédigez ce calcul sans commettre d'erreurs d'arrondis (vous pourrez reprendre la rédaction d'un.e des étudiant.e.s ayant trouvé la bonne réponse).

**Exercice 6 : Utilisation du symbole  $\sum$**

- Calculez, sans utiliser la calculatrice, chacune des sommes suivantes :  $\sum_{n=1}^4 n, \sum_{n=-4}^4 n, \sum_{i=-3}^3 i^2$ .

**Chapitre 1 : Statistiques descriptives univariées**

**Exercice 7 : Types de variables**

- Mme Leroy, travaillant dans un bureau d'études statistiques, a recueilli les données suivantes dans le cadre de son travail :
  - le nombre de rendez-vous médicaux que différents patients ont pris au cours de l'année 2023
  - la popularité des différents footballeurs de l'équipe de France
  - le temps hebdomadaire passé devant la télévision par des enfants
  - les prénoms des enfants nés en 2023.

Dans chacun de ces cas, indiquez quelle est la population, quelle est la variable étudiée, et quelle est la nature de cette variable.

2. De même, précisez quel est le type de variable dans les situations suivantes :

- (a) la sensibilité aux questions écologiques des habitants d'un même quartier
- (b) le nombre de personnes des ménages français
- (c) le temps mis par des rats pour sortir d'un labyrinthe
- (d) les noms des rues de Dijon
- (e) le taux de testostérone parmi les patients souffrant de troubles cognitifs.

**Exercice 8 : Taille des ménages**

On note  $T$  la variable statistique indiquant le nombre de personnes d'un ménage. Au sein d'une certaine ville, on obtient les données suivantes :

Nombre de personnes dans le ménage	1	2	3	4	5	6
Effectifs : nombre de ménages	1001	1337	1505	1149	532	152

- 1. Quelles sont les modalités de la variable  $T$  ?
- 2. Calculez les fréquences et les fréquences cumulées.
- 3. Calculez la proportion de ménages d'au moins 3 personnes.
- 4. Déterminez  $\mathbb{P}_r[T < 4]$ .

**Exercice 9 : Moyens de transport**

1. Utilisation des transports en commun :

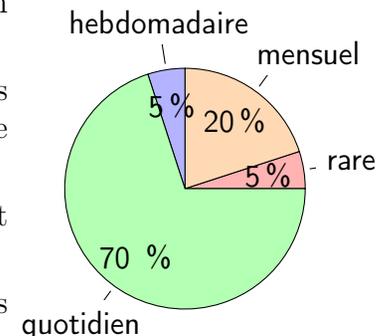
(a) Un étudiant en sociologie interroge 40 personnes de son entourage sur leur utilisation des transports en commun. Il réalise le graphique ci-dessous après les avoir regroupés en quatre groupes :

**rare** : ceux qui utilisent les transports en commun moins d'une fois par mois

**mensuel** : au moins une fois par mois, mais moins d'une fois par semaine

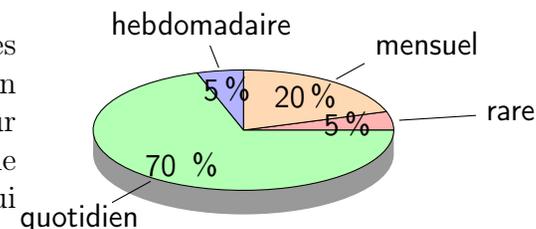
**hebdomadaire** : entre une et cinq fois par semaine

**quotidien** : au moins 5 fois par semaine



- i. Déterminez les effectifs des différentes modalités.
- ii. Quelle est, au sein de cet échantillon, la proportion d'individus qui utilisent les transports en commun moins d'une fois par semaine ?

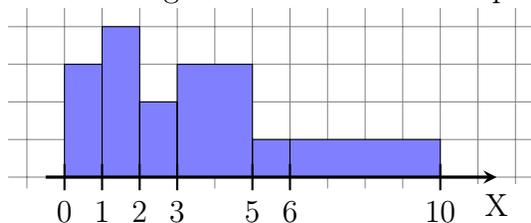
(b) Lorsqu'il entre les données dans un tableur, l'ordinateur réalise un graphique ressemblant à celui ci-contre. Que penser d'un tel graphique ?



2. Utilisation de la voiture :

Pour le même échantillon de 40 personnes, l'étudiant désigne par  $X$  le temps hebdomadaire passé dans une voiture (exprimé en

heures). Il trace un histogramme sur une feuille quadrillée :



(a) Déterminez les fréquences des différentes classes.

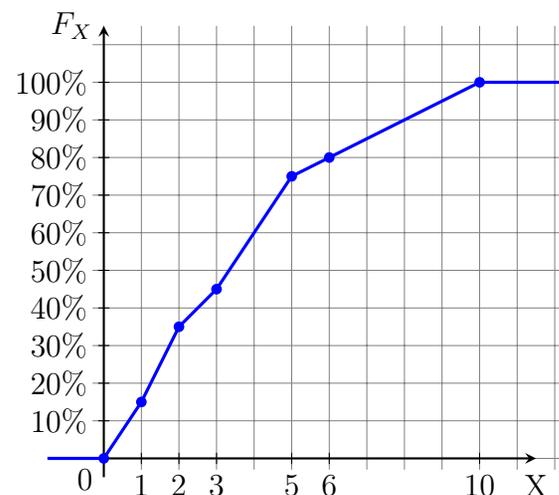
$X$	$[0; 1[$	$[1; 2[$	$[2; 3[$	$[3; 5[$	$[5; 6[$	$[6; 10[$
Fréquence						

(b) Comparez les fréquences de la première et de la dernière classes, ainsi que les hauteurs des rectangles correspondants, puis commentez.

### 3. Polygone des fréquences cumulées :

Enfin, l'étudiant décide de représenter (toujours sur son papier quadrillé) le polygone des fréquences cumulées des données de la question 2 : Le polygone est reproduit ci-après (après l'énoncé des questions)

- (a) Déterminez graphiquement (et de manière approchée) la médiane et les quartiles de  $X$ .
- (b) Déterminez graphiquement (et de manière approchée) les proportions  $\mathbb{P}_r[X \leq 4]$ ,  $\mathbb{P}_r[X \geq 7]$  et  $\mathbb{P}_r[4 \leq X \leq 7]$ .
- (c) On considère l'individu qui passe le moins de temps en voiture parmi les 10% d'individus qui passent le plus de temps en voiture. Combien de temps passe-t-il environ par semaine dans sa voiture ?



### Exercice 10 : Pays fondateurs de la zone euro

On considère le régime politique et la population (exprimée en million d'habitants, noté M.hab.) des états fondateurs de la zone euro :

Allemagne	Autriche	Belgique	Espagne	Finlande
République	République	Monarchie	Monarchie	République
82,2 M.hab.	8,9 M.hab.	11,4 M.hab.	47,4 M.hab.	5,5 M.hab.
France	Irlande	Italie	Luxembourg	Pays-Bas
République	République	République	Monarchie	Monarchie
67,8 M.hab.	5,1 M.hab.	60,4 M.hab.	0,6 M.hab.	17,3 M.hab.
Portugal				
République				
10,3 M.hab.				

1. Quelle est la proportion de monarchies parmi ces états ?
2. Parmi l'ensemble des habitants de ces états, quelle proportion vit dans une république ?
3. Parmi les états fondateurs de la zone euro, quelle est la population moyenne des républiques et celle des monarchies ? Calculez aussi les médianes et écarts type.

**Exercice 11 : Développement psychomoteur**

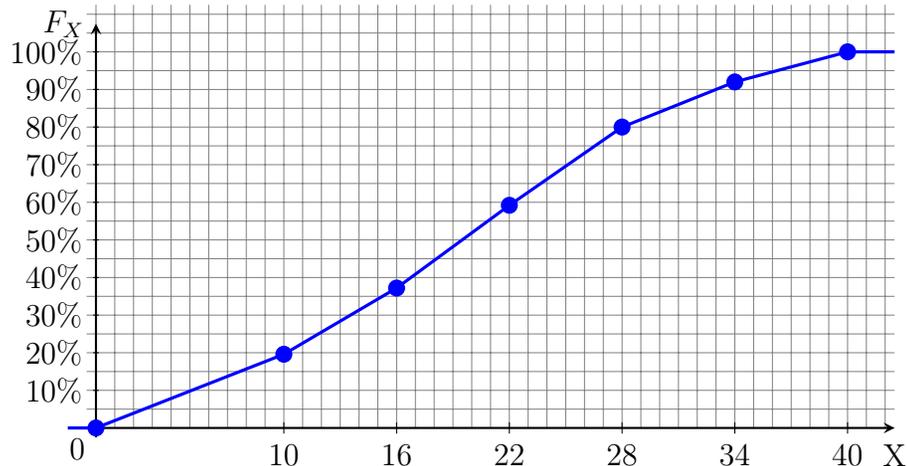
Les données ci-dessous décrivent, pour un groupe de 200 bébés âgés de deux ans, la variable statistique  $X$  représentant le “score de développement psychomoteur” (SDP) :

SDP	[55; 70[	[70; 85[	[85; 100[	[100; 115[	[115; 130[	[130; 145[
Effectif	10	44	78	49	15	4

1. Calculez la médiane et les quartiles de  $X$ .
2. Calculez la moyenne et l'écart-type de  $X$ .

**Exercice 12 : Méthodes d'apprentissage**

On teste une méthode d'apprentissage par “dessins commentés” sur un échantillon de 250 enfants. On désigne par  $X$  la note (entre 0 et 40) obtenue par les enfants à l'issue de l'apprentissage. On représente le polygone des fréquences cumulées de ces notes sur un papier quadrillé :



1. Déterminez graphiquement la médiane et les quartiles de la variable statistique  $X$ .
2. (a) Donnez une approximation de  $\mathbb{P}_r[X < 25]$ .  
 (b) Quelle est environ la proportion d'enfants dont la note est supérieure à 18 ?  
 (c) Combien vaut environ  $\mathbb{P}_r[20 \leq X < 30]$  ?

3. Plus précisément, les notes recueillies sont données ci-dessous :

X	[0; 10[	[10; 16[	[16; 22[	[22; 28[	[28; 34[	[34; 40[
Effectif	49	44	55	52	30	20

- (a) Calculez la médiane et les quartiles approchés de la variable statistique  $X$  (cette fois-ci sans lecture graphique).
- (b) Calculez la note moyenne  $m = m(X)$  et l'écart-type  $s = s(X)$  de l'échantillon.
- (c) Donnez une approximation de la proportion des enfants dont la note est comprise entre  $m - s$  et  $m + s$ .

**Exercice 13 : Âge et performances mémorielles**

On considère un échantillon de 16 personnes de 35 à 84 ans, auxquelles on attribue une note indiquant leurs performances mémorielles. On note leur âge  $X$  et leur note de performances mémorielles  $Y$ .

Les données mesurées sont les suivantes :

X	75	59	57	50	68	38	53	43	40	67	35	36	50	43	73	84
Y	27	41	58	51	22	45	22	31	38	42	68	82	38	65	35	27

Calculez la moyenne et l'écart-type de l'âge  $X$  et des performances mémorielles  $Y$  au sein de l'échantillon.

**Exercice 14 : Types de variables**

1. Mr Sauvage, travaillant dans un bureau d'études statistiques, a recueilli les données suivantes dans le cadre de son travail :
  - (a) la couleur des voitures stationnées dans le campus
  - (b) le nombre d'enfants de chaque femme dijonnaise
  - (c) le degré de satisfaction des usagers des transports en commun
  - (d) le poids des appareils photos vendus en 2023.

Dans chacun de ces cas, indiquez quelle est la population, quelle est la variable étudiée, et quelle est la nature de cette variable.

2. De même, précisez quel est le type de variable dans les situations suivantes :

- (a) les langues officielles des différents pays européens
- (b) la pluie tombée à Dijon au cours des différents mois de l'année
- (c) le nombre de pièces des appartements en location à Dijon
- (d) les numéros de sécurité sociale des enseignants de l'Université de Bourgogne
- (e) le nombre moyen d'enfants par femme dans chaque pays européen
- (f) les marques des téléphones vendus en 2023
- (g) le nombre de pages des livres d'une bibliothèque
- (h) les numéros de téléphones figurant dans les pages blanches dijonnaises
- (i) le nombre de crises d'épilepsie de différents patients au cours du mois d'octobre 2023
- (j) la douleur ressentie par différents patients atteints de la même pathologie.

**Exercice 15 :** Calculez la moyenne et l'écart-type des données de l'Exercice 8.

### Exercice 16 : Questionnaire de logique

Un chercheur établit un questionnaire visant à évaluer les compétences logiques de jeunes adolescents. Il fait remplir ce questionnaire à 12 sujets, et pour chacun d'eux il note dans un tableau le symbole « ✓ » si la réponse donnée est correcte, « 0 » si la réponse donnée est fautive, et laisse la case vide si le sujet n'a pas répondu. Il obtient le tableau suivant :

Individu	A	B	C	D	E	F	G	H	I	J	K	L
Question 1	✓	0	✓	✓		✓			0		✓	0
Question 2			0					✓	0	0	0	0
Question 3	✓	✓	✓			0	✓	0	0	✓	✓	✓
Question 4			✓	✓	0			0	0	0		
Question 5		0	✓			0	✓	0	✓	0	✓	0
Question 6	✓	✓	✓			✓	✓	0		0	0	0
Question 7	✓	0	0		✓	✓	0	0	0	0	0	✓
Question 8	0	0			✓	✓		0	0		✓	
Question 9	✓	✓	✓	0	0		✓	0	✓	✓		0
Question 10		✓	✓		✓		✓	✓		✓	0	

1. Quel a été le taux de réponse à la question 7 ?
2. Quelle proportion de sujets a répondu correctement à la question 4 ?
3. Parmi les individus ayant répondu à la question 8, quelle proportion a donné la bonne réponse ?
4. Le chercheur décide de résumer par une note les réponses données au questionnaire. Il attribue un point par bonne réponse (et zéro point par réponse erronée ou manquante). Calculez les notes (sur 10) attribuée à chaque adolescent interrogé.
5. Calculez la moyenne, l'écart-type et la médiane de ces notes.

**Exercice 17 : Regroupement de données en classes**

Voici les notes obtenues par un groupe d'élèves lors d'un contrôle noté sur 10 points :

6 ; 3 ; 2,5 ; 9 ; 5 ; 2 ; 7,5 ; 4 ; 9,5 ; 8 ; 6 ; 10 ; 4,5 ; 4 ; 9 ; 1 ; 5 ; 6 ; 6 ; 5 ; 7,5 ; 8 ; 6 ; 8 ; 8 ; 3 ; 5 ; 9 ; 9,5 ; 6 ; 7 ; 6 ; 4 ; 4 ; 6,5 et 8

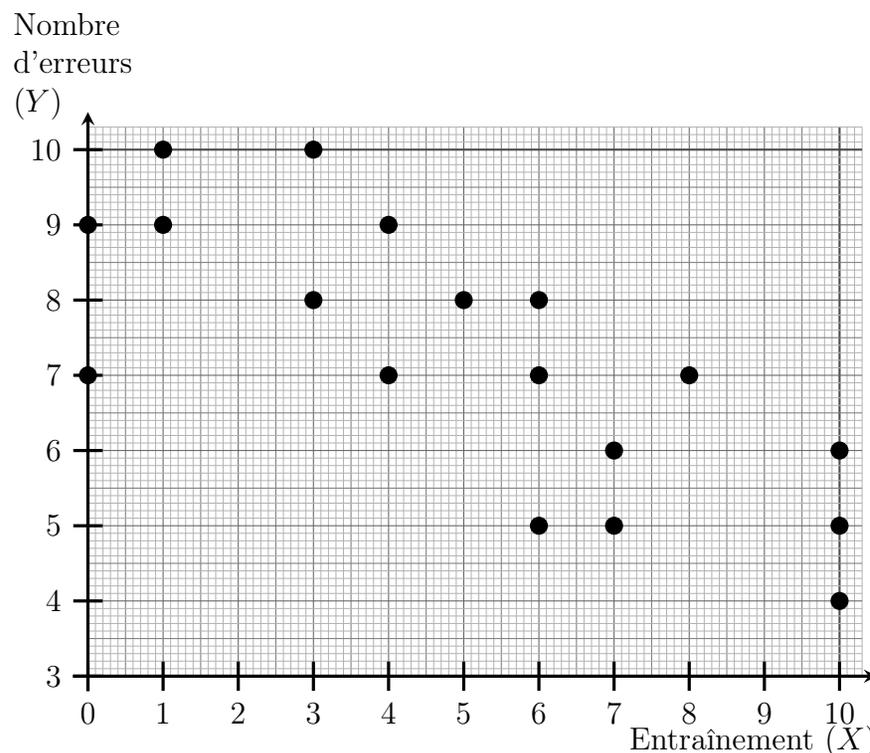
1. Faites un tri des données en indiquant l'effectif de chaque note.
2. Quelle est la taille de l'échantillon ?
3. Donner la moyenne, l'écart-type et la médiane.
4. Transformez les données en les rangeant en classes d'amplitude 1,5. Calculez les fréquences et les fréquences cumulées des classes.
5. Calculez la moyenne, l'écart-type et la médiane des notes regroupées en classes. Que remarque-t-on ?

**Chapitre 2 : Statistiques descriptives bivariées**

**Exercice 18 : Entraînement à un exercice de logique**

Pour étudier l'impact de l'entraînement sur la réussite à un test de logique, on mesure les performances d'enfants qui ont déjà effectué un certain nombre de fois un exercice similaire.

Plus précisément, pour un échantillon de 18 enfants, on considère le nombre  $Y$  d'erreurs commises à un test de logique, tandis que le nombre d'exercices similaires qu'ils ont déjà effectués auparavant est noté  $X$ . On regroupe ces résultats sous la forme d'un nuage de points (effectué sur un papier "millimétré") :



1. Extrayez de ce nuage de points les valeurs de  $X$  et  $Y$  pour chaque individu :

Entraînement $X$																	
Nombre d'erreurs $Y$																	

2. Calculez le coefficient de corrélation des rangs (de Spearman) et interprétez la valeur de ce coefficient.
3. Calculez les moyennes et écart-types des variables  $X$  et  $Y$ , puis leur coefficient de corrélation linéaire (de Pearson). Interprétez la valeur de ce coefficient.
4. Si un enfant a déjà fait 7 exercices de logique, alors combien estimeriez-vous qu'il fera d'erreurs si on lui fait à nouveau passer

un test similaire ?

5. Un enfant a commis 6 erreurs. Combien de fois estimeriez-vous qu'il avait déjà fait un test similaire pour s'entraîner ?

### Exercice 19 : Stress et temps de réponse

Un chercheur s'intéresse au temps de réponse de rats à des stimuli visuels. Il a constaté que dans les conditions d'élevage des rats dont il dispose, certains rats deviennent très stressés alors que la plupart restent beaucoup moins stressés. Il décide de mesurer d'une part (à l'aide d'indicateurs hormonaux) ce stress noté  $X$ , et d'autre part le temps de réponse aux stimuli (noté  $Y$ , et exprimé en ms) d'un échantillon de rats, obtenant les résultats suivants :

Sujet	A	B	C	D	E	F	G	H	I	J
niveau de stress $X$	2,6	7,99	2,91	8,16	3,25	8,6	2,85	2,54	3,22	3,1
temps de réponse $Y$	495	701	501	659	490	713	503	506	495	494

- Déterminez le coefficient de corrélation des rangs (de Spearman) pour les variables  $X$  et  $Y$ . Ces variables sont-elles très corrélées ?
- Déterminez le coefficient de corrélation linéaire des variables  $X$  et  $Y$ . Ces variables sont-elles très corrélées ?
- Suite à une perte d'échantillon, on ne connaît plus le niveau de stress d'un rat, dont on a mesuré qu'il met 705 ms à réagir aux stimuli visuels. À quel niveau de stress s'attend-on pour ce rat ?
- Tracez le nuage de points des variables  $X$  et  $Y$ .
- Commentez vos réponses aux questions 1, 2 et 3 en vous appuyant sur le nuage de points.

### Exercice 20 : Revenus et espérance de vie

On étudie le revenu par habitant  $X$  (en milliers de "dollars internationaux" par an) et l'espérance de vie  $Y$  (en années) dans 11 pays différents :

pays	revenu/hab. $X$	espérance de vie $Y$
Algérie	15,2	75,6
Burundi	0,8	59,6
Croatie	24,1	78
Djibouti	3,6	63,5
Estonie	31,5	77,6
Finlande	44	81,1
Guinée	2	59
Honduras	5,5	74,6
Indonésie	12,4	69,1
Jamaïque	9,2	76,2
Koweït	69,7	74,7

- Déterminez les médianes du revenu  $X$  et de l'espérance de vie  $Y$ .
- Déterminez le coefficient de corrélation linéaire de  $X$  et  $Y$ .
- Déterminez le coefficient de corrélation des rangs (de Spearman) de  $X$  et  $Y$ .
- Commentez les résultats obtenus après avoir tracé le nuage de points de  $X$  et  $Y$ .
- On ajoute à ces données réelles un pays imaginaire nommé "Utopia", où le revenu par habitant est de 1130 et l'espérance de vie est de 1800. Calculez les moyennes et médianes qu'on obtiendrait alors, ainsi que les coefficients de corrélation. Enfin, commentez les résultats obtenus.

**Exercice 21 : Données sur un groupe d'étudiants**

On demande à un groupe d'étudiants leur nombre de frères/soeurs, leur humeur du jour, leur taille, et la moyenne de leurs notes obtenues au baccalauréat. On obtient les données suivantes :

Laetitia de bonne humeur 3 frères/soeurs 1m61 bac : 13,3/20	Isabelle de relativement bonne humeur 2 frères/soeurs 1m66 bac : 19,52/20	Aurélien de très bonne humeur 1 frère/soeur 1m67 bac : 14,99/20
---	---	---

Jocelyne de mauvaise humeur 2 frères/soeurs 1m77 bac : 13,64/20	Sylvain de bonne humeur 0 frère/soeur 1m77 bac : 11,65/20	David de bonne humeur 1 frère/soeur 1m87 bac : 13,33/20
---	---	---

Chantal de bonne humeur 2 frères/soeurs 1m66 bac : 15,52/20	Théo de bonne humeur 3 frères/soeurs 1m81 bac : 11,48/20	Pascale de mauvaise humeur 1 frère/soeur 1m58 bac : 12,3/20
---	--	---

Ces données sont-elles appariées? Dès lors, est-il possible de calculer les coefficients de corrélations entre certaines de ces variables?

**Exercice 22 : Thérapie pour soulager l'angoisse**

Pour mettre en évidence l'efficacité d'une thérapie visant à réduire l'angoisse de personnes victimes d'agressions, des psychologues ont observé 15 sujets avant et après la thérapie en affectant à chaque sujet un score (plus le score est élevé, plus fort est le niveau d'angoisse). Les données sont les suivantes.

Avant	30	33	17	21	36	21	32	27	25	28	23	37	39	29	22
Après	25	16	13	15	32	16	36	22	6	38	13	32	31	30	8

1. Calculez le coefficient de corrélation linéaire et le coefficient de corrélation des rangs (de Spearman) entre les deux phases.
2. Parmi les deux droites de régression, y en a-t-il une qui a plus d'intérêt du point de vue thérapeutique? Déterminez son équation.

**Exercice 23 : Extraversion**

Dans une expérience, 10 sujets ont rempli un questionnaire de personnalité d'Eysenck comprenant, entre autres, une échelle de sociabilité (X) et une échelle d'impulsivité (Y). On a obtenu les résultats suivants :

sociabilité X	7,6	6,4	10,1	6,6	6,4	7	6,1	7,7	2,9	7,1
impulsivité Y	3,7	6,4	3,2	4,3	4	5,7	3,4	5,3	1,9	0,5

On souhaite savoir si ces deux échelles mesurent plusieurs manifestations d'un même trait de caractère : l'extraversion. Si tel était le cas, on pourrait alors additionner les scores obtenus aux deux échelles pour former une échelle unique d'extraversion.

1. Calculez le coefficient de corrélation linéaire des deux échelles X et Y.
2. Serait-il donc pertinent de résumer ces deux échelles par une échelle unique d'extraversion?

**Exercice 24 : Reconnaissance de formes**

Pour étudier l'entraînement d'enfants passant un test de reconnaissance de formes, on mesure les performances d'enfants qui ont déjà effectué un certain nombre de fois un exercice similaire.

Pour un échantillon de 14 enfants, on mesure le temps de réponse Y des enfants, tandis que le nombre d'exercices similaires qu'ils ont déjà effectués auparavant est noté X. On obtient les résultats suivants :

Entraînement X	0	0	1	1	2	3	4	5	6	6	7	7	9	10
Temps de réponse Y	3,6	4,3	3,3	3,6	3,7	2,7	2,7	2,2	2	2,1	1,8	2,3	1,4	1,2

1. Calculez le coefficient de corrélation des rangs, et interprétez la valeur de ce coefficient.

- Calculez les moyennes et écart-types des variables  $X$  et  $Y$ , puis leur coefficient de corrélation linéaire. Interprétez la valeur de ce coefficient.
- Tracez le nuage de points correspondant à ces données. Ce graphique vous conforte-t-il quant aux interprétations données aux questions précédentes ?
- Si un enfant a déjà fait 8 exercices de reconnaissance de formes, alors à combien estimeriez-vous son temps de réponse lors d'un prochain test ?
- Un enfant a mis 2,4 secondes pour répondre au test. Combien de fois estimeriez-vous qu'il avait déjà fait un test similaire pour s'entraîner ?

**Exercice 25 : Résolution de problèmes mathématiques**

On a mis au point deux méthodes d'apprentissage pour la résolution de problèmes mathématiques. La première méthode ( $\mathcal{M}\mathcal{V}$ ) est uniquement verbale, tandis que la seconde méthode ( $\mathcal{M}\mathcal{E}$ ) est écrite. Ces méthodes sont testées sur un même groupe d'enfants. Le tableau suivant représente les notes  $X$  et  $Y$  obtenues par 17 élèves à deux épreuves relatives à ces deux méthodes d'apprentissages.

$X : \mathcal{M}\mathcal{V}$	7	10	5,5	6,5	9,5	8	5,5	4	6	3,5	6	7	6,5	9,5	7	8	1,5
$Y : \mathcal{M}\mathcal{E}$	9	10	5	9	10	8,5	4,5	6,5	6	4,5	5	5,5	5	9,5	9	10	2,5

- Dessinez le nuage statistique de ces variables.
- Calculez le coefficient de corrélation des rangs des deux variables  $X$  et  $Y$ .
- Calculez les moyennes, les écart-types et le coefficient de corrélation linéaire des deux variables  $X$  et  $Y$ .
- On souhaite désormais estimer à quelle note s'attendre avec la méthode écrite pour un enfant ayant obtenu une note  $x = 9$  avec la méthode verbale :
  - Quelle droite de régression peut-on utiliser pour répondre à cette question ? Donner son équation.

- Si un enfant a obtenu une note  $x = 9$ , déterminer la note  $y$  à laquelle on s'attendrait pour cet enfant.

**Chapitre 3 : probabilités**

**Exercice 26 : Situations d'emploi**

Un petit immeuble dijonnais compte 5 habitants, dont la situation d'emploi est la suivante :

Nom	Mr Fournier	Mr Pascal	Mme Hardy	Mme Leroy	Mme Moreau
Situation	retraite	retraite	chômage	emploi	emploi

On choisit au hasard 3 habitants parmi ces 5 habitants de l'immeuble.

- Listez tous les choix possibles de trois habitants, et conclure que chaque cas a une probabilité de 10%.
- Quelle est la probabilité d'avoir
  - exactement 2 femmes dont exactement une a un emploi ?
  - strictement moins de 2 hommes ?
  - deux personnes en situation d'emploi ?
- On note  $X$  le nombre de femmes parmi les trois personnes choisies au hasard. Déterminez la loi de  $X$  (c'est-à-dire, calculez la probabilité  $\mathbb{P}[X = k]$  pour chaque possibilité de  $k$ ).

**Exercice 27 : Factorielles & coefficients binomiaux**

- Simplifiez les nombres suivants :  $\frac{21!}{19!}$ ,  $\frac{15!}{13! \times 2!}$ .
- Écrivez à l'aide de deux factorielles l'expression  $7 \times 8 \times 9 \times 10 \times 11$ .
- Calculez les nombres suivants :  $\binom{15}{2}$ ,  $\binom{12}{4}$ ,  $\binom{12}{8}$ ,  $\binom{14}{0}$ ,  $\binom{10}{1}$ .

**Exercice 28 : Croyance religieuse**

1. Sébastien et Chantal sont deux enfants nés cette année à Dijon. Compte tenu de l'environnement familial dans lequel ils vont grandir, on considère que chacun d'eux a 60% de chances de devenir athée (et donc 40% de chances de développer au contraire une croyance religieuse). Quelle est la probabilité

- (a) Qu'ils développent tous deux une croyance religieuse.
- (b) Qu'un seul d'entre eux devienne athée.
- (c) Qu'ils deviennent tous les deux athées.

Argumentez le fait que ces probabilités correspondent à une loi binomiale, dont vous donnerez les paramètres.

2. Au sein d'une famille de 5 personnes, on constate que 2 personnes ont une croyance religieuse (les 3 autres personnes sont donc athées). On choisit au hasard 2 personnes au sein de cette famille, et on note  $X$  le nombre de personnes athées parmi ces deux personnes choisies au hasard. Calculez la loi de  $X$ .

3. Au sein d'une ville de 2000 personnes, on constate que 40% (c'est-à-dire 800 personnes) ont une croyance religieuse. On choisit au hasard 2 personnes au sein de cette ville, et on note  $X$  le nombre de personnes athées parmi ces deux personnes choisies au hasard. Calculez la loi de  $X$ .

4. Comparez entre elles les lois obtenues dans ces trois situations.

**Exercice 29 : Ressources humaines**

Dans une grande entreprise, la direction demande aux personnes exerçant des responsabilités de noter les employés placés sous leur responsabilité. La direction leur impose de donner la note  $A$  (la meilleure note) à 20% des employés, la note  $B$  à 30% des employés et la note  $C$  (la moins bonne note) à 50% des employés.

On choisit, au sein de l'entreprise, 22 employés au hasard, et on suppose que les effectifs de l'entreprise sont assez grands pour que cet échantillon puisse être considéré comme un tirage "avec remise".

1. Calculez la probabilité que l'échantillon contienne exactement 4 employés notés  $C$ .
2. Quelle est la probabilité que l'échantillon contienne au moins 4 individus notés  $C$  ?
3. Quelle est la probabilité que l'échantillon contienne au moins 13 individus notés  $C$  ou  $B$  ?
4. Calculez la probabilité que l'échantillon contienne moins que 8 personnes notées  $A$ .
5. Quel est le nombre moyen d'individus notés  $B$  au sein d'un échantillon aléatoire de 22 employés ?

**Exercice 30 : Loi binomiale**

On considère une variable aléatoire  $X$  suivant une loi binomiale  $\mathcal{B}(n; p)$ , dont on sait que la moyenne est  $m(X) = 4$  et la variance est  $\text{Var}(X) = 3$ .

- 1) Calculez  $n$  et  $p$ .
- 2) Calculez les probabilités  $\mathbb{P}[9 \leq X \leq 11]$  et  $\mathbb{P}[X > 13]$ .
- 3) Trouvez toutes les valeurs entières de  $k$  telles que  $\mathbb{P}[X \geq k] \leq 0,05$ .

**Exercice 31 : Situation d'emploi dans un groupe de trois amis.**

On considère un groupe de trois amis : Alice, Bernard et Cécile. Parmi eux, Alice et Bernard exercent un emploi alors que Cécile est sans emploi.

1. On choisit au hasard le nom d'une de ces trois personnes, puis à nouveau le nom d'une de ces trois personnes au hasard (ce peut être la même personne – ou pas).
  - (a) Listez les neuf possibilités pour ces deux noms choisis au hasard.
  - (b) Parmi ces possibilités, combien comptent uniquement des personnes en situation d'emploi ?
  - (c) Quelle est la proportion, parmi ces possibilités, qui comptent une personne sans emploi, et une qui exerce un emploi ?

- (d) Pour un tel choix aléatoire de deux noms, on note  $X$  le nombre de personnes choisies exerçant un emploi. Quelle est la loi de variable  $X$  ?
- (e) Retrouvez le résultat de la question c), en utilisant cette loi.
- (f) Déterminez de même  $\mathbb{P}[X = 0]$  et  $\mathbb{P}[X = 2]$ .
2. On convient d'une autre manière de choisir deux noms : on choisit un premier nom au hasard, puis on choisit le second nom parmi les deux restant. Reprenez, avec cette nouvelle façon de choisir les deux noms, les questions 1a, 1b et 1c.

### Exercice 32 : Collection de dessins

Mr Blanc est pédiatre et utilise fréquemment de petites illustrations pour stimuler les enfants. Il a acheté à cet effet 5 cartes illustrées qu'il stocke dans un tiroir.

- Il mélange les cartes, et constate qu'elles sont dans un certain ordre à l'issue du mélange. Quelle était la probabilité qu'à l'issue du mélange, elles se retrouvent précisément dans cet ordre là ?
- S'il remélange les cartes. Calculez les probabilités des événements suivants :
  - la carte qui était en haut du tas soit à nouveau en haut du tas ?
  - les deux cartes qui étaient en haut du tas soient à nouveau en haut du tas, et dans le même ordre ?
  - les deux cartes qui étaient en haut du tas soient à nouveau en haut du tas, mais dans l'ordre inverse ?
  - les deux cartes qui étaient en haut du tas soient à nouveau en haut du tas, dans n'importe quel ordre ?
  - l'ensemble des cartes soient exactement dans le même ordre que lors du premier mélange ?
- Répondez aux mêmes questions en supposant qu'il y ait cette fois-ci 29 cartes.

- Peut-on conclure des probabilités calculées que certains événements ont particulièrement peu de chances de se produire ?

### Exercice 33 : Thérapie contre la dépression

Une psychologue a développé une nouvelle thérapie contre la dépression, et elle affirme que cette thérapie permet la rémission de 65% des patients. En interrogeant un échantillon aléatoire de 22 patients, on constate 6 rémissions.

- S'il y avait 65% de rémission parmi l'ensemble des patients, quelle loi suivrait le nombre de rémissions au sein d'un échantillon de 22 patients choisis au hasard avec remise ? Quel serait le nombre moyen de rémissions au sein d'un tel échantillon ?
- Que pensez-vous de l'hypothèse selon laquelle le tirage soit "avec remise" ?
- Sous cette même hypothèse, quelle serait la probabilité d'avoir au maximum 6 rémissions ?
- Conclure : vous semble-t-il vraisemblable qu'il y ait, comme l'affirme cette psychologue, 65% de rémissions parmi l'ensemble des patients ?

### Exercice 34 : Nombre de filles et de garçons

Dans une classe, il y a 10 filles et 19 garçons. On choisit au hasard 5 élèves distincts dans cette classe.

- Quel est le nombre de choix possibles ?
- Quel est le nombre de choix ne comportant que des garçons ?
- Quelle est la proportion de choix ne comportant que des filles ?
- Quelle est la probabilité, en choisissant ainsi un échantillon de 5 élèves au hasard, que cet échantillon comporte 4 filles et 1 garçon ?
- Quelle est donc la probabilité d'avoir choisi au plus un garçon ?

## Chapitre 4 : Loi Normale

**Exercice 35 : Loi normale centrée réduite**

On désigne par  $Z$  une variable aléatoire suivant la loi normale centrée réduite (c'est à dire de densité  $f(x) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2})$ ). On note  $F$  sa fonction de répartition, c'est à dire qu'on pose  $F(z) = \mathbb{P}[Z \leq z]$ . On trouvera la table des valeurs de  $F$  à la page 11 du formulaire.

- Pour cette question, on considère les nombres  $z_1 = 1,04$ ,  $z_2 = 1,83$  et  $z_3 = 0,144$ .
  - Déterminez  $F(z_1)$ ,  $F(z_2)$  et  $F(z_3)$  de deux façons différentes : d'une part, en utilisant la table du formulaire et, d'autre part, en utilisant les fonctions dédiées de votre calculatrice.
  - Calculez de deux manières différentes les valeurs suivantes :  $\mathbb{P}[z_1 < Z \leq z_2]$ ,  $\mathbb{P}[-z_2 < Z < -z_1]$ ,  $\mathbb{P}[z_1 \leq Z < 5]$ ,  $\mathbb{P}[-z_2 \leq Z \leq z_1]$ ,  $\mathbb{P}[-z_1 < Z < z_1]$ ,  $\mathbb{P}[Z > z_2]$  et  $\mathbb{P}[Z < z_1]$ . (On utilisera, d'une part, les valeurs collectées dans la question (a) et, d'autre part, les fonctions dédiées de la calculatrice.)
- Déterminez, à l'aide du formulaire, des valeurs approchées des nombres  $z_4$ ,  $z_5$ , et  $z_6$  tels que  $F(z_4) = 0,7967$ ,  $F(z_5) = 0,0793$  et  $F(z_6) = 0,7247$ .
  - Déterminez de deux manières différentes les valeurs approchées des nombres  $z_7$ ,  $z_8$ , et  $z_9$  tels que  $\mathbb{P}[Z \leq z_7] = 0,99$ ,  $\mathbb{P}[Z > z_8] = 0,7611$  et  $\mathbb{P}[Z < z_9] = 0,9162$ . (On utilisera, d'une part, la table du formulaire et, d'autre part, les fonctions dédiées de la calculatrice.)

**Exercice 36 : Images mentales**

Dans une expérience sur les images mentales, on demande à un échantillon de 349 enfants d'apprendre une liste de 40 mots en leur montrant, pour chaque mot, un dessin illustrant ce mot afin de faciliter leur apprentissage. On désigne par  $X$  le nombre de mots dont se souvient chaque enfant le jour suivant, et on regroupe par classes les données :

Nb de mots $X$	[10; 15[	[15; 20[	[20; 25[	[25; 30[	[30; 35[
Effectif	30	144	143	29	3

- Calculez la moyenne  $m$  et l'écart-type  $s$  de la variable statistique  $X$ .
- Quels seraient les effectifs théoriques si  $X$  suivait la loi normale  $\mathcal{N}(m; s)$  ?

**Exercice 37 : Désirabilité sociale**

On étudie les scores (notés  $X$ ) obtenus par des enfants sur une échelle de désirabilité sociale (D.S.). Une étude a montré que pour cette échelle, les scores ont, sur l'ensemble des enfants de CM2, une moyenne  $\mu \simeq 9,98$  et un écart-type  $\sigma \simeq 2,59$ .

Dans la suite, on effectue des calculs en supposant que  $X$  suit la loi normale  $\mathcal{N}(9,98; 2,59)$ .

- Calculez la proportion théorique des scores supérieurs à 10,64.
- Trouvez une valeur approchée du nombre  $a$  tel que  $\mathbb{P}[X < a] = 0,95$ .
- Déterminez les quartiles de la variable  $X$ .
- Quel est le score minimal des 20 % de ceux qui ont les scores D.S. les plus élevés ?
- Quel est le score maximal des 35 % de ceux qui ont les scores D.S. les moins élevés ?

**Exercice 38 : Taux d'échec à un concours**

Lors d'un concours le pourcentage d'échecs est de 15%. On choisit au hasard (avec remise) un échantillon de  $n$  étudiants et on désigne par  $S_n$  le nombre d'échecs au sein de l'échantillon.

1. On prend  $n = 17$ . Calculez  $p_k = \mathbb{P}[S_{17} \leq k]$  pour  $k = 0, 1, 2$  et  $3$ .
2. Pour  $n = 100$ , on souhaite calculer  $\mathbb{P}[10 \leq S_{100} \leq 20]$  :
  - a) Calculez  $\mathbb{P}[10 \leq S_{100} \leq 20]$  avec la calculatrice.
  - b) Justifiez qu'on peut approcher la loi de  $S_{100}$  par une loi normale. Calculez  $\mathbb{P}[10 \leq S_{100} \leq 20]$  de manière approchée. (On fera le calcul avec et sans correction de continuité, pour juger de l'intérêt de cette correction.)
3. On considère désormais un échantillon de taille  $n = 27\,000$ .
  - (a) Par quelle loi normale peut-on approcher la variable  $S_{27\,000}$  ? Calculez ainsi (une valeur approchée de) la probabilité  $\mathbb{P}[3\,947 \leq S_{27\,000} \leq 4\,162]$ .
  - (b) Par quelle loi normale peut-on approcher la proportion  $P_{27\,000} = \frac{S_{27\,000}}{27\,000}$  ? Calculez ainsi (une valeur approchée de) la probabilité  $\mathbb{P}[0,1462 \leq P_{27\,000} \leq 0,1541]$ .
  - (c) Trouvez une valeur approchée de  $a$  telle que  $\mathbb{P}[S_{27\,000} < a] = 0,03$ .

**Exercice 39 : Indice de masse corporelle dans la population française**

L'indice de masse corporelle (IMC) mesure la corpulence d'un individu. Il s'agit d'un élément de diagnostic de dénutrition ou d'obésité, dont l'interprétation est la suivante :

<b>IMC &lt; 16,5</b> dénutrition ou famine	<b>25 ≤ IMC &lt; 30</b> surpoids
<b>16,5 ≤ IMC &lt; 18,5</b> maigreur	<b>30 ≤ IMC &lt; 35</b> obésité modérée
<b>18,5 ≤ IMC &lt; 25</b> corpulence normale	<b>35 ≤ IMC &lt; 40</b> obésité sévère
	<b>IMC ≥ 40</b> obésité morbide

Dans cet exercice, on notera  $X$  la variable statistique rendant compte de l'IMC au sein de la population française. On ne sait pas si l'hypothèse suivante est vraie :

$$(\mathcal{H}) \quad \text{« La variable } X \text{ suit la loi normale } \mathcal{N}(25; 5) \text{. »}$$

Or, une enquête épidémiologique conduite par l'INSERM a donné la répartition suivante de l'IMC au sein d'un échantillon de 25 714 personnes, supposé représentatif de la population française :

IMC	] -∞ ; 18,5[	[18,5 ; 25[	[25 ; 30[	[30 ; 35[	[35 ; 40[	[40 ; +∞[
fréquence (en %)	3,5	49,2	32,3	10,7	3,1	1,2

1. Déterminez la probabilité (« *fréquence théorique* ») qu'aurait chaque classe sous l'hypothèse de loi normale ( $\mathcal{H}$ ).
2. Justifiez que, sous l'hypothèse ( $\mathcal{H}$ ), la proportion *théorique* d'obésité (c'est à dire  $\mathbb{P}[X \geq 30]$ , en regroupant les obésités modérées, sévères et morbides) serait d'environ 15,9%.
3. Quelle est, en fait, la proportion des sujets obèses observée dans l'échantillon épidémiologique de l'INSERM ?
4. On cherche désormais à comparer les résultats des questions 2 et 3 : pour cela, on va évaluer si la proportion  $\mathbb{P}_r[X \geq 30]$  observée dans l'échantillon de l'INSERM s'accorde avec la probabilité  $\mathbb{P}[X \geq 30]$  obtenue sous l'hypothèse ( $\mathcal{H}$ ).

- (a) Si on suppose qu'un-e français-e choisi-e au hasard a une probabilité de 15,9% d'être en situation d'obésité, quelle est la loi du nombre de personnes en situation d'obésité dans un échantillon de 25 714 français? Le fait que l'échantillon soit *avec* ou *sans remise* a-t-il la moindre importance?
- (b) Cette loi peut-elle être approximée par une loi normale, et si oui laquelle?
- (c) On considère toujours un échantillon aléatoire de 25 714 français-es, mais cette fois-ci on considère comme variable aléatoire la **proportion** de personnes, au sein de l'échantillon, qui sont en situation d'obésité. Sa loi peut-elle aussi être approximée par une loi normale, et si oui laquelle?
- (d) En utilisant cette approximation par une loi normale, calculer la probabilité qu'un échantillon aléatoire de 25 714 français-es contienne au maximum 15% de personnes en situation d'obésité.
- (e) En utilisant le résultat de la question (d) et en faisant appel à l'étude épidémiologique de l'INSERM, que pouvez-vous en conclure au sujet de l'hypothèse ( $\mathcal{H}$ )?

#### Exercice 40 : Loi normale avec moyenne et écart type

1. Déterminez les probabilités des événements suivants pour une variable aléatoire  $Z$  suivant la loi normale centrée réduite :
  - (a)  $Z$  est supérieure ou égale à 0,6 ;
  - (b)  $Z$  est comprise entre  $-1$  et  $1$  ;
  - (c)  $Z$  est comprise entre  $-2$  et  $2$  ;
  - (d)  $Z$  est comprise entre  $-3$  et  $3$ .

2. Déterminez les probabilités des événements suivants pour une variable aléatoire  $X$  suivant une loi normale de moyenne  $\mu = 108$  et d'écart-type  $\sigma = 11$  (c'est-à-dire  $X \sim \mathcal{N}(108; 11)$ ) :
  - (a)  $X$  est supérieure ou égale à  $\mu + 0,6\sigma$  ;
  - (b)  $X$  est comprise entre  $\mu - \sigma$  et  $\mu + \sigma$  ;
  - (c)  $X$  est comprise entre  $\mu - 2\sigma$  et  $\mu + 2\sigma$  ;
  - (d)  $X$  est comprise entre  $\mu - 3\sigma$  et  $\mu + 3\sigma$ .
3. Reprenez la question 2 pour une variable aléatoire  $X$  suivant la loi normale  $\mathcal{N}(13,9; 2,8)$ . Qu'observez-vous?

#### Exercice 41 : Diagnostic de la dépression

On évalue les niveaux de dépression au moyen d'un questionnaire appelé le « test Inventaire Multiphasique de la Personnalité du Minnesota ». Compte tenu du grand nombre de questions posées dans ce test, on considère que le score obtenu par un patient choisi au hasard suit une loi normale. De plus, ce test est normalisé pour que la moyenne soit  $\mu = 50$  et que l'écart-type soit  $\sigma = 10$ .

Si l'on considère qu'une note supérieure à 70 traduit un état pathologique, combien s'attend-on à trouver de personnes dépressives pathologiquement dépressives sur un ensemble de 10000 personnes?

#### Exercice 42 : Épreuve graphique d'organisation perceptive

La proportion des enfants qui réussissent parfaitement une épreuve graphique d'organisation perceptive est de 70%. On choisit au hasard (avec remise) un échantillon de  $n$  enfants et on note  $X$  la variable aléatoire égale au nombre d'enfants dans l'échantillon qui réussissent l'épreuve.

1. Lorsque  $n = 25$ , précisez la loi de probabilité de  $X$ , puis calculez  $\mathbb{P}[16 \leq X \leq 18]$ .
2. On considère désormais que  $n = 250$ .
  - (a)

- (b) Justifiez qu'on peut approcher la loi de  $X$  par une loi normale, préciser laquelle, et indiquer s'il est nécessaire de faire une correction de continuité.
- (c) Déduisez de la question précédente des valeurs approchées de  $\mathbb{P}[X \geq 165]$  et  $\mathbb{P}[160 \leq X \leq 180]$ .

**Exercice 43 : Pile ou face**

Vous jouez à "pile ou face" avec une pièce de monnaie équilibrée. Avec quelle probabilité la fréquence de "pile" est-elle comprise entre 0,48 et 0,52 si vous lancez la pièce

- 100 fois.
- 200 fois.
- 1 000 fois.
- 10 000 fois.

**Exercice 44 : Potentiels d'action d'un neurone**

On considère que le nombre de potentiels d'action (PA) émis en une minute par un neurone au repos suit une loi normale de moyenne 120 et d'écart-type 15.

- Quelle est la probabilité pour qu'un neurone émette plus de 129 PA en 1 min s'il n'est pas activé ?
- Quelle est la probabilité pour qu'un neurone émette moins de 90 PA en 1 min s'il n'est pas activé ?
- Quelle est la probabilité pour qu'un neurone émette entre 105 et 147 PA en 1 min s'il n'est pas activé ?
- On enregistre l'activité d'un neurone durant une minute, et on observe que ce neurone a émis plus de 190 PA. Selon vous, est-il plus vraisemblable que ce neurone était activé ou au repos lors de l'enregistrement ? Justifiez votre réponse.

**Chapitre 5 : Estimation****Exercice 45 : Pile ou face**

Vous jouez à « pile » ou « face » avec une pièce de monnaie. Vous la lancez 1400 fois et obtenez 757 fois pile.

- Estimez, avec la confiance 96%, la probabilité qu'a cette pièce de tomber sur pile quand on la lance.
- Conclure : pouvez-vous affirmer (avec la confiance 96%) que cette pièce soit biaisée ?

**Exercice 46 : Efficacité d'un traitement**

Un fabricant de médicaments affirme qu'en 5 jours, 55% des malades qui sont traités par l'un de ses produits guérissent de leur maladie.

Pour vérifier cette affirmation du fabricant, on administre ce produit à 500 malades. On constate alors que, au sein de cet échantillon de 500 patients traités, il y en a 353 qui guérissent en 5 jours.

- Donnez une estimation de la proportion réelle  $p$  de guérisons (pour l'ensemble des patients que l'on traiterait avec ce médicament) avec une confiance de 95%.
- Que peut-on conclure quant à l'affirmation du fabricant ?
- On veut réduire la marge de l'estimation à 1% tout en assurant une confiance de 98%. Quelle doit alors être la taille minimale de l'échantillon à considérer ?

**Exercice 47 : Satisfaction de clients**

Afin de répondre au mieux au désir de sa clientèle, une entreprise réalise un sondage pour connaître l'avis des consommateurs sur un produit qu'elle fabrique. Sur 1245 personnes interrogées, 545 se sont déclarées satisfaites. L'entreprise essaye alors d'améliorer la qualité de son produit, puis elle réalise un deuxième sondage un an après : sur 1023 personnes interrogées, il y en a alors 473 qui sont satisfaites.

1. On désigne par  $p_1$  et  $p_2$  les proportions de satisfaits sur l'ensemble de la clientèle de l'entreprise, avant et après l'amélioration apportée. Donnez une estimation de  $p_1$  et  $p_2$  par intervalle de confiance en prenant comme confiance  $c = 0,9$ . Peut-on dire que la proportion de satisfaits pour l'ensemble de la clientèle a augmenté ?
2. Lors du deuxième sondage, quelle aurait dû être la taille de l'échantillon pour avoir une estimation de  $p_2$  à 1% près avec une confiance de 0,99 ?

#### Exercice 48 : Résistance à la persuasion

On émet l'hypothèse que la résistance à la persuasion passe par une réaction active des sujets qui développent intérieurement des contre-arguments. Pour tester cette hypothèse on considère l'opinion qu'ont les étudiants sur la coopération entre étudiants et enseignants pour établir les programmes. Cette opinion est mesurée par un questionnaire pour lequel, sur l'ensemble des étudiants, la moyenne est 13,7 et l'écart type est 2,4.

1. On soumet dans un premier temps un échantillon de 90 étudiants à un argumentaire persuasif contre cette coopération, puis on mesure leur opinion à l'aide du même questionnaire. On obtient les notes suivantes :

Classe	[6 ; 8[	[8 ; 10[	[10 ; 12[	[12 ; 14[	[14 ; 16[	[16 ; 18[
Effectif	2	6	21	35	20	6

- a) Calculer la moyenne expérimentale  $m_e(X)$  et l'écart type expérimental  $s_e(X)$  de l'échantillon.
- b) Estimer pour l'ensemble des étudiants la note moyenne après avoir entendu un tel argumentaire (on l'appelle  $\mu(X)$ ). On déterminera un intervalle de confiance, avec la confiance  $c = 0,95$ .

- c) Peut-on dire avec un risque d'erreur de 5% qu'en moyenne, la note (traduisant l'opinion en faveur d'une coopération entre étudiants et enseignants) a diminué après l'argumentaire ? Justifiez votre réponse.
  - d) Quelle devrait être la taille minimale de l'échantillon à prélever pour estimer la moyenne  $\mu(X)$  à 0,5 points près avec une confiance de 0,99 ?
2. On considère désormais un échantillon de 26 étudiants à qui on demande, pendant qu'ils écoutent l'argumentaire, de réaliser des « opérations » arithmétiques (réciter mentalement la table de multiplication par 8). La moyenne et l'écart type expérimentaux sont alors  $m_e = 9,24$  et  $s_e = 4,5$ .
    - a) Avec une confiance de 0,95 déterminer la note moyenne de l'ensemble des étudiants s'ils écoutaient l'argumentaire en récitant des tables de multiplication. (*On pourra supposer que les notes suivent une loi normale.*)
    - b) Comparer cette moyenne à celle qu'ils auraient sans réciter les tables de multiplication (en étant donc en mesure de réfléchir à des contre-arguments). Ces résultats sont-ils compatibles avec l'hypothèse théorique que l'on souhaitait tester ?
    - c) Donner une estimation de l'écart type des notes, dans le deuxième cas (en récitant les tables de multiplication) avec un risque d'erreur de 5%.

**Exercice 49 : Motivation lors d'un test**

Une psychologue qui étudie la prise de décisions a demandé à 31 enfants de résoudre le plus grand nombre possible de problèmes en 30 minutes. Elle a expliqué à un groupe de 15 enfants qu'elle voulait tester leur aptitude innée à résoudre des problèmes, et aux autres enfants (16 enfants) qu'il ne s'agissait que d'une tâche destinée à les occuper. Les résultats obtenus sont les suivants :

Aptitude innée	23	23	29	34	20	41	29	26	25	20	39	31	31	29	27	
Occupation	23	14	25	18	28	11	17	12	15	22	22	19	11	21	19	22

On suppose que dans les deux cas (que l'on ait prétendu que l'on teste les enfants ou bien qu'on veut juste les occuper), la variable aléatoire qui donne le score à ce test suit une loi normale.

1. Calculer les moyennes et les écarts types des deux échantillons.
2. Donner une estimation par intervalle de confiance des moyennes et des écarts types associés aux deux conditions d'expérience. *On fera cette estimation avec une confiance de 95%.*
3. Que peut-on conclure au vu de ces résultats ?

**Exercice 50 : Taux de cholestérol**

On étudie les taux de cholestérol dans le sang chez les femmes et chez les hommes de plus de 45 ans. On suppose qu'ils suivent des lois normales.

Dans un échantillon de 17 femmes de plus de 45 ans on a mesuré un taux moyen de 177,3 mg/dL avec un écart type de 38,3 mg/dL.

Dans un échantillon de 22 hommes de plus de 45 ans on a mesuré un taux moyen de 193,8 mg/dL avec un écart type de 34,2 mg/dL.

1. Donner des estimations des écarts types des deux populations. *On demande de déterminer un intervalle de confiance, avec la confiance  $c = 99\%$ .*
2. En utilisant les estimations par intervalle de confiance à 99%, peut-on conclure que les femmes de plus de 45 ans ont en moyenne moins de cholestérol que les hommes de plus de 45 ans ?

**Exercice 51 : Épreuve de dictée**

Le tableau suivant représente les résultats obtenus par un échantillon de 12 enfants de CM1 dans deux épreuves de dictée préparée : une liste de 15 mots (épreuve  $\mathcal{E}_1$ ) et un texte (épreuve  $\mathcal{E}_2$ ).

Élève	A	B	C	D	E	F	G	H	I	J	K	L
Épreuve 1	15,7	14	18,5	14,5	12,9	15	15,1	17,3	14,5	16	19,1	15
Épreuve 2	14,7	12	13,3	12	11,3	13,6	10,1	11,7	11,7	13,5	15	12

1. Peut-on affirmer, avec un risque d'erreur de 4%, qu'une des deux épreuves donne, pour l'ensemble des enfants de CM1, une note moyenne plus élevée que l'autre épreuve ?
2. Donner des estimations des écarts types.

**Exercice 52 :**

La France compte environ 1% ou 2% de personnes boulimiques.

Un.e psychologue étudie le taux de chômage de ces patients atteints de boulimie, afin de déterminer s'il diffère du taux de chômage de l'ensemble de la population.

1. En France, le taux de chômage est de 7,8% (selon les données de l'INSEE).

On choisit au hasard un échantillon de 30000 français. Quelle est la probabilité d'avoir, au sein de cet échantillon, entre 7,6% et 8% de chômeurs ?

*Justifiez bien les éventuelles approximations que vous seriez amené à faire*

2. Sur un échantillon de 30000 patients atteints de boulimie, le psychologue a constaté que 8,107% d'entre eux étaient au chômage.

- (a) Avec la confiance  $c = 90\%$ , déterminez un intervalle de confiance pour le taux de chômage des personnes boulimiques.

Peut-on en déduire (avec la confiance 90%) que le taux de chômage des personnes boulimiques est différent de celui de l'ensemble de la population ?

- (b) Avec la confiance  $c = 98\%$ , estimez la proportion de chômeurs au sein de l'ensemble des personnes atteintes de boulimie.

Peut-on en déduire (avec le risque d'erreur  $\alpha = 2\%$ ) que le taux de chômage des personnes boulimiques est différent de celui de l'ensemble de la population ?

- (c) Pour pouvoir estimer à 0,05% près le taux de chômage des personnes boulimiques avec une confiance de 98%, quelle taille d'échantillon faudrait-il considérer ?

2. La chercheuse pense que, bien au delà de la question du daltonisme, les filles sont bien plus douées que les garçons pour détecter d'infimes nuances de couleurs. Elle teste cette hypothèse sur des enfants de 6 ans, auxquels elle fait passer un test standardisé de perception des couleurs. Sur un échantillon de 16 filles, elle obtient les données suivantes :

Score	13	9	13	15	13	10	11	15	10	19	14	13	12	12	11	14
-------	----	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----

Estimez, avec la confiance 99%, le score moyen des filles à ce test.

3. Sur un échantillon de 43 garçons auxquels la chercheuse fait passer le test, le score moyen est de 10,51 avec un écart-type de 4,54.

Estimez, avec la confiance 99%, le score moyen des garçons à ce test.

Peut-on affirmer, avec la confiance 99%, que – comme s'y attendait la chercheuse – les filles ont en moyenne une meilleure perception des couleurs que les garçons ?

4. Estimez de même les écart-types des garçons et des filles, avec la confiance 99%. Peut-on conclure qu'ils diffèrent selon le sexe ?

5. Quelle taille d'échantillon aurait-il fallu considérer pour estimer le score moyen des filles à 0,5 point près, avec une confiance de 99% ?

### Exercice 53 : Perception des couleurs

Une chercheuse, Mme Launay, se demande si la perception des couleurs diffère en fonction du sexe des individus.

1. On étudie tout d'abord la prévalence du Daltonisme.

On considère d'une part un échantillon de 3 000 femmes, parmi lesquelles 11 sont daltoniennes, et d'autre part un échantillon de 3 000 hommes, parmi lesquels 265 sont daltoniens.

Peut-on déduire, avec la confiance  $c = 95\%$ , que la prévalence du daltonisme soit plus importante chez l'homme que chez la femme ?

**Remarque :** Une grande partie des données de ces exercices sont fictives, et visent simplement à illustrer les outils mathématiques introduits en cours.