

### 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  $F\leftrightarrow D$ ) pour les casios, ou la fonction  $\blacktriangleright$ Frac du menu  $\boxed{\text{MATH}}$  des TI).

- Grâce à cette fonctionnalité de la calculatrice, simplifiez chacune des fractions suivantes :  $\frac{22 \times 6}{6+115}$ ,  $\frac{-70}{65}$ ,  $\frac{42}{-6}$ ,  $\frac{-4-22}{20-2}$ ,  $\frac{27-55}{14 \times 6}$ ,  $\frac{-104}{63-258}$ .

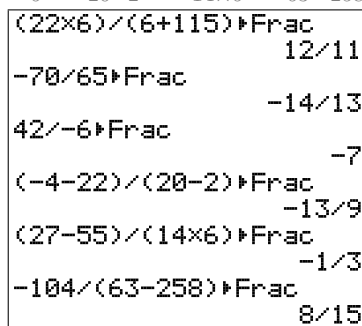
Comme en témoigne la capture d'écran de calculatrice ci-contre, on obtient :

$$\frac{22 \times 6}{6+115} = \frac{12}{11}, \quad \frac{-70}{65} = -\frac{14}{13},$$

$$\frac{42}{-6} = -7, \quad \frac{-4-22}{20-2} = -\frac{13}{9},$$

$$\frac{27-55}{14 \times 6} = -\frac{1}{3}, \quad \frac{-104}{63-258} = \frac{8}{15}.$$

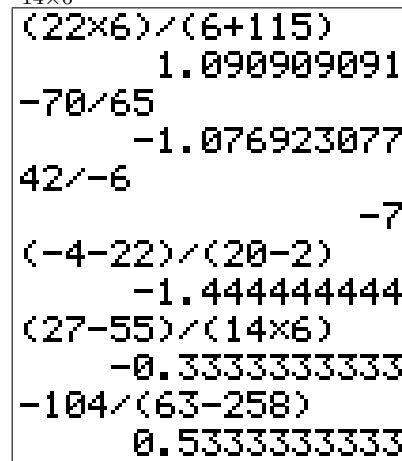
**Remarque :** La capture d'écran correspond à une TI. Sur une casio, si l'on veut par exemple calculer  $\frac{-70}{65}$ , on tape  $(-70)\div 65$  puis si le résultat affiché est un chiffre à virgule on appuie sur  $F\leftrightarrow D$  pour en faire une fraction. Au lieu d'afficher  $-\frac{14}{13}$ , certaines calculatrices affichent alors  $-1\frac{1}{13}$  (ces calculatrices considèrent que cela signifie  $-(1 + \frac{1}{13})$ ). Pour obtenir la réponse sous la bonne forme sur ces calculatrices, il faut alors appuyer sur  $\boxed{\text{SHIFT}}$  puis  $F\leftrightarrow D$  (sur la touche il est en effet écrit " $a + \frac{b}{c} \leftrightarrow \frac{d}{c}$ ").



- Classez les dans l'ordre croissant ("du plus petit au plus grand"). Cette fois-ci on utilise les résultats sous formes de "chiffres à virgules". On obtient (voir capture d'écran ci-dessous) :

$$\frac{22 \times 6}{6+115} \simeq 1,091, \quad \frac{-70}{65} \simeq -1,077, \quad \frac{42}{-6} = -7, \quad \frac{-4-22}{20-2} \simeq -1,444,$$

$$\frac{27-55}{14 \times 6} \simeq -0,333, \quad \frac{-104}{63-258} \simeq 0,5333.$$



L'ordre croissant de ces nombres est :  $-7 < -1,444 < -1,077 < -0,333 < 0,5333 < 1,091$

Donc la réponse est :  $\frac{42}{-6} < \frac{-4-22}{20-2} < \frac{-70}{65} < \frac{27-55}{14 \times 6} < \frac{-104}{63-258} < \frac{22 \times 6}{6+115}$

#### 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  $-1$  ?


(c) que  $-1$  est l'opposé du carré de  $1$  ?

La réponse (a) est fautive : le carré de  $-1$  est  $(-1)^2$ , et vaut  $1$ .


La réponse (b) est aussi fautive : le carré de  $1$  est égal à  $1$ .

La réponse (c) est quant à elle correcte : "l'opposé du carré de  $1$ " est la quantité qui s'écrit  $-(1^2)$  ou plus simplement  $-1^2$ . C'est cette quantité qui est calculée par la calculette et vaut  $-1$ .


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.  (a)  $\frac{1}{2021 \times 2022} \simeq 0,000000245$   
 (b)  $\frac{1}{2021 \times 2022} \simeq 2,45$   
 (c)  $\frac{1}{2021 \times 2022} \simeq 24471046,1$

La bonne réponse est la réponse (a). En effet **2.447104610E-7** désigne la quantité  $2,447104610 \times 10^{-7}$ , qui vaut  $0,0000002447104610$ .

3.  (a)  $-9 \times 21^2 = 35721$   
 (b)  $-189^2 = 35721$   
 (c)  $(-9 \times 21)^2 = 35721$

La bonne réponse est la réponse (c). En effet dans **X²**, **X** vaut  $-9 \times 21$ , (calculé à la ligne précédente), de sorte qu'on a calculé  $(-9 \times 21)^2$ .

4.  (a)  $4 - 28 \times 2022 = -48528$   
 (b)  $24 \times 2022 = -48528$   
 (c)  $(4 - 28) \times 2022 = -48528$

La bonne réponse est la réponse (c). En effet dans **Ans\*2022**,

**Ans** vaut  $4 - 28$ , (calculé à la ligne précédente), de sorte qu'on a calculé  $(4 - 28) \times 2022$ .

### Exercice 3 : Puissances de dix

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

$41 \div 100$	$0,41 \times 10^2$	$136 \times 10^6$
$0,110\ 403\ 069 \times 100\ 000\ 000$	$1\ 845,023\ 8 \times 100$	$1\ 445,97 \times 10^7$
$139 \times 10^{-3}$	$11,656\ 315 \div 1\ 000\ 000$	$112 \div 100\ 000\ 000$

- $41 \div 100 = 0,41$  (que l'on peut aussi écrire  $4,1 \times 10^{-1}$ )
- $0,41 \times 10^2 = 41$  (que l'on peut aussi écrire  $4,1 \times 10^1$ )
- $136 \times 10^6 = 136000000$  (que l'on peut aussi écrire  $1,36 \times 10^8$ )
- $0,110\ 403\ 069 \times 100\ 000\ 000 = 11040307$  (que l'on peut aussi écrire  $1,1040307 \times 10^7$ )
- $1\ 845,023\ 8 \times 100 = 184502,38$  (que l'on peut aussi écrire  $1,8450238 \times 10^5$ )
- $1\ 445,97 \times 10^7 = 14459700000$  (que l'on peut aussi écrire  $1,44597 \times 10^{10}$ )
- $139 \times 10^{-3} = 0,139$  (que l'on peut aussi écrire  $1,39 \times 10^{-1}$ )
- $11,656\ 315 \div 1\ 000\ 000 = 0,000011656315$  (que l'on peut aussi écrire  $1,1656315 \times 10^{-5}$ )
- $112 \div 100\ 000\ 000 = 0,00000112$  (que l'on peut aussi écrire  $1,12 \times 10^{-6}$ )

### Exercice 4 : Règles de calcul

1. On considère les expressions suivantes :

- $2 \frac{7}{5 \times 6}$
- $\frac{1}{5} \times \frac{7}{2 \times 6}$
- $\frac{7 \div 2}{5 \times 6}$
- $\frac{2 \times 7}{5 \times 6}$
- $\frac{7}{5 \times 6 \div 2}$
- $7 \div 6 \times 2 \div 5$

Calculer la valeur de chacune de ces expressions, et constater que certaines sont égales.

Comme en témoigne la capture d'écran ci-contre, on a les valeurs ci-dessous :

- $2 \frac{7}{5 \times 6} \simeq 0,4667$
- $\frac{1}{5} \times \frac{7}{2 \times 6} \simeq 0,1167$
- $\frac{7 \div 2}{5 \times 6} \simeq 0,1167$
- $\frac{2 \times 7}{5 \times 6} \simeq 0,4667$
- $\frac{7}{5 \times 6 \div 2} \simeq 0,4667$
- $7 \div 6 \times 2 \div 5 \simeq 0,4667$

2*7/(5*6)	0.4666666667
1/5*7/(2*6)	0.1166666667
7/2/(5*6)	0.1166666667
2*7/(5*6)	0.4666666667
7/(5*6/2)	0.4666666667
7/6*2/5	0.4666666667

On remarque alors d'une part que les expressions  $\frac{7}{5 \times 6 \div 2}$ ,  $\frac{2 \times 7}{5 \times 6}$ ,  $7 \div 6 \times 2 \div 5$  et  $2 \frac{7}{5 \times 6}$  sont égales, et d'autre part que les expressions  $\frac{7 \div 2}{5 \times 6}$  et  $\frac{1}{5} \times \frac{7}{2 \times 6}$  sont égales.

**Remarque :** L'égalité de ces expressions se comprend à partir des règles de parenthésage et de signe expliquées dans les "Rappels" au début des notes de cours.

2. De même, pour chacune des expressions suivantes, calculer les valeurs lorsque les variables prennent les valeurs indiquées, et remarquer que certaines d'entre elles sont égales.

(a) Les expressions suivantes

- $2 - x + 5 \times 3 + 5 \times 8$
- $2 - x - 1(5 - 3 - 8)$
- $2 - x - 0(5 - 3 - 8)$
- $2 + 6 - x - 6 + 5 \times 3 + 5 \times 8$
- $2 - x - (5 - 3 - 8)$
- $2 - x - 5(-3 - 8)$
- $2 - x + 5 - 3 - 8 - (5 - 3 - 8)$

lorsque  $x = -1$ .

Comme en témoigne la capture d'écran ci-contre, on a les valeurs ci-dessous :

- $2 - x + 5 \times 3 + 5 \times 8 = 58$
- $2 - x - 1(5 - 3 - 8) = 9$
- $2 - x - 0(5 - 3 - 8) = 3$
- $2 + 6 - x - 6 + 5 \times 3 + 5 \times 8 = 58$
- $2 - x - (5 - 3 - 8) = 9$
- $2 - x - 5(-3 - 8) = 58$
- $2 - x + 5 - 3 - 8 - (5 - 3 - 8) = 3$

-1*x	-1
2-x+5*3+5*8	58
2-x-1*(5-3-8)	9
2-x-0*(5-3-8)	3
2+6-x-6+5*3+5*8	58
2-x-(5-3-8)	9
2-x-5*(-3-8)	58
2-x+5-3-8-(5-3-8)	3

Ainsi, lorsque  $x = -1$ , on obtient d'une part que les expressions  $2 - x + 5 \times 3 + 5 \times 8$ ,  $2 + 6 - x - 6 + 5 \times 3 + 5 \times 8$  et  $2 - x - 5(-3 - 8)$  sont égales, d'autre part que les expressions  $2 - x - 1(5 - 3 - 8)$  et  $2 - x - (5 - 3 - 8)$  sont égales, et enfin que les expressions  $2 - x - 0(5 - 3 - 8)$  et  $2 - x + 5 - 3 - 8 - (5 - 3 - 8)$  sont égales.

(b) Les expressions

- $2 - b \times \frac{4-6}{b}a$
  - $(2 - 4) - 6a$
  - $6a + 2 - 4a$
  - $-4 - 6a + 2$
  - $2 - (4 - 6)a$
- où  $a = 2$  et  $b = -3$ .

Comme en témoigne la capture d'écran ci-contre, on a les valeurs ci-dessous :

- $2 - b \times \frac{4-6}{b}a = 6$
- $-4 - 6a + 2 = -14$
- $(2 - 4) - 6a = -14$
- $2 - (4 - 6)a = 6$
- $6a + 2 - 4a = 6$

2*A	2
-3*B	-3
2-B*(4-6)/B*A	6
-4-6*A+2	-14
(2-4)-6*A	-14
2-(4-6)*A	6
6*A+2-4*A	6

Ainsi, pour ces valeurs de  $a$  et  $b$ , on obtient d'une part que les expressions  $2 - b \times \frac{4-6}{b}a$ ,  $2 - (4 - 6)a$  et  $6a + 2 - 4a$  sont égales, d'autre part que les expressions  $-4 - 6a + 2$  et  $(2 - 4) - 6a$  sont égales.

(c) Lorsque  $x = -2$ ,  $y = -9$  et  $z = 4$  :

- $z \times x + (6 + (y + 8) \times 7 \times 7)$
- $(6 + z) \times x + (y + 8) \times 7^2$
- $6 + z \times x + (7^2 \times y + (8))$
- $6 \times x + z \times x + 7^2(y + 8)$
- $6 + z \times x + 7^2 \times y + 8$
- $(6 + (z \times x)) + 7^2(y + 8)$

Comme en témoigne la capture d'écran ci-contre, on a les valeurs ci-dessous :

- $z \times x + (6 + (y + 8) \times 7 \times 7) = -51$
- $(6 + z) \times x + (y + 8) \times 7^2 = -69$
- $6 + z \times x + (7^2 \times y + (8)) = -435$
- $6 \times x + z \times x + 7^2(y + 8) = -69$
- $6 + z \times x + 7^2 \times y + 8 = -435$
- $(6 + (z \times x)) + 7^2(y + 8) = -51$

-2*X	-2
-9+Y	-9
4+Z	4
Z*X+(6+(Y+8)*7*7)	-51
(6+Z)*X+(Y+8)*7^2	-69
6+Z*X+(7^2*Y+(8))	-435
6*X+Z*X+7^2*(Y+8)	-69
6+Z*X+7^2*Y+8	-435
(6+(Z*X))+7^2*(Y+8)	-51

Ainsi, pour ces valeurs de  $x$ ,  $y$  et  $z$ , on obtient d'une part que les expressions  $z \times x + (6 + (y + 8) \times 7 \times 7)$  et  $(6 + (z \times x)) + 7^2(y + 8)$  sont égales, d'autre part que les expressions  $(6 + z) \times x + (y + 8) \times 7^2$  et  $6 \times x + z \times x + 7^2(y + 8)$  sont égales, et enfin que les expressions  $6 + z \times x + (7^2 \times y + (8))$  et  $6 + z \times x + 7^2 \times y + 8$  sont égales.

(d) Lorsque  $x_1 = 8$  et  $x_2 = -1$  :

- $\frac{x_2 \times 3}{8} - \frac{x_2 \times x_1}{9}$
- $(x_2 \times 3 - x_1 \times x_2) \div (8 - 9)$

•  $x_2 \frac{(3-x_1)}{(8-9)}$

•  $x_2 \left(\frac{3}{8} - \frac{x_1}{9}\right)$

Comme en témoigne la capture d'écran ci-contre, on a les valeurs ci-dessous :

- $\frac{x_2 \times 3}{8} - \frac{x_2 \times x_1}{9} \simeq 0,5139$
- $x_2 \frac{(3-x_1)}{(8-9)} = -5$
- $(x_2 \times 3 - x_1 \times x_2) \div (8 - 9) = -5$
- $x_2 \left(\frac{3}{8} - \frac{x_1}{9}\right) \simeq 0,5139$

8*X	8
-1*Y	-1
Y*3+8-Y*X+9	0.5138888889
Y*(3-X)/(8-9)	-5
(Y*3-X*Y)/(8-9)	-5
Y*(3+8-X+9)	0.5138888889

**Remarque :** Comme la plupart des calculatrices n'utilisent pas les noms  $x_1$  et  $x_2$ , on a désigné  $x_1$  par  $X$  et  $x_2$  par  $Y$ .

Ainsi, pour ces valeurs de  $x_1$  et  $x_2$ , on obtient d'une part que les expressions  $x_2 \frac{(3-x_1)}{(8-9)}$  et  $(x_2 \times 3 - x_1 \times x_2) \div (8 - 9)$  sont égales, d'autre part que les expressions  $\frac{x_2 \times 3}{8} - \frac{x_2 \times x_1}{9}$  et  $x_2 \left(\frac{3}{8} - \frac{x_1}{9}\right)$  sont égales.

(e) Lorsque  $a_1 = -3$ ,  $a_2 = 8$ ,  $x_1 = 1$  et  $x_2 = 5$  :

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

-3+A	-3
8+B	8
1*X	1
5+Y	5
(X-(Y+A-B))	7
X-Y+A-B	-15
(B-Y)-(A-X)	7
A-Y+X-B	-15

Comme en témoigne la capture d'écran ci-contre, on a les valeurs ci-dessous :

- $(x_1 - (x_2 + a_1 - a_2)) = 7$
- $x_1 - x_2 + a_1 - a_2 = -15$
- $(a_2 - x_2) - (a_1 - x_1) = 7$
- $a_1 - x_2 + x_1 - a_2 = -15$

**Remarque :** Comme la plupart des calculatrices n'utilisent pas les noms  $a_1$ ,  $a_2$ ,  $x_1$  et  $x_2$ , on a désigné  $a_1$  par  $A$ ,  $a_2$  par  $B$ ,  $x_1$  par  $X$  et  $x_2$  par  $Y$ .

Ainsi, pour ces valeurs de  $a_1$ ,  $a_2$ ,  $x_1$  et  $x_2$ , on obtient d'une part que les expressions  $(x_1 - (x_2 + a_1 - a_2))$  et  $(a_2 - x_2) - (a_1 - x_1)$  sont égales, d'autre part que les expressions  $x_1 - x_2 + a_1 - a_2$  et  $a_1 - x_2 + x_1 - a_2$  sont égales.

### Compléments

#### Exercice 5 : Précision des calculs

Imane, Marine, Anaëlle et Alexis 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 " $m(X)$ ", " $m(X^2)$ ", " $Var$ " et " $s$ ", donnés par les formules suivantes :

$$m(X) = \frac{1,84+1,87+1,73}{3}$$

$$m(X^2) = \frac{1,84^2+1,87^2+1,73^2}{3}$$

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

$$s = \sqrt{Var}$$

Le nombre " $s$ " issu de ce calcul est l'écart-type, exprimé (dans le cas présent) en mètre. On leur demandait de calculer 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'Imane**

$$m(X) = \frac{1,84+1,87+1,73}{3} \simeq 1,81$$

$$m(X^2) = \frac{1,84^2+1,87^2+1,73^2}{3} \simeq 3,29$$

$$Var \simeq 3,29 - 1,81^2 \simeq 0,01$$

$$s \simeq \sqrt{0,01} \simeq 0,10 \text{ mètre.}$$

Donc l'écart-type est d'environ 10 cm.

**Calculatrice d'Imane**

<code>(1.84+1.87+1.73)÷3</code>	1.813333333
<code>(1.84²+1.87²+1.73²)÷3</code>	3.2918
<code>3.29-1.81²</code>	0.0139
<code>√(0.01)</code>	0.1

**Copie de Marine**

$$m(X) = \frac{1,84+1,87+1,73}{3} = \frac{5,44}{3}$$

$$m(X^2) = \frac{1,84^2+1,87^2+1,73^2}{3} = \frac{9,8754}{3}$$

$$Var \simeq \frac{9,8754}{3} - \left(\frac{5,44}{3}\right)^2 \simeq 0,0036$$

$$s = \sqrt{Var} \simeq 0,06 \text{ mètre.}$$

Donc l'écart-type est d'environ 6 cm.

**Calculatrice de Marine**

<code>1.84+1.87+1.73</code>	5.44
<code>1.84²+1.87²+1.73²</code>	9.8754
<code>9.8754÷3-(5.44÷3)²</code>	0.00362222222
<code>√(Ans)</code>	0.06018490028

**Copie d’Anaëlle**

$$m(X) = \frac{1,84+1,87+1,73}{3} \simeq 1,81$$

$$m(X^2) = \frac{1,84^2+1,87^2+1,73^2}{3} \simeq 3,29$$

$$Var = m(X^2) - (m(X))^2 \simeq 0,0036$$

$$s = \sqrt{Var} \simeq 0,06 \text{ mètre.}$$

Donc l’écart-type est d’environ 6 cm.

**Calculatrice d’Anaëlle**

```
(1.84+1.87+1.73)+3=M
1.813333333
(1.84^2+1.87^2+1.73^2)+3=M
3.2918
N-M^2+U
0.003622222222
√(U)
0.06018490028
```

**Copie d’Alexis**

$$s = \sqrt{\frac{1,84^2+1,87^2+1,73^2}{3} - \left(\frac{1,84+1,87+1,73}{3}\right)^2} \simeq 0,06 \text{ mètre.}$$

Donc l’écart-type est d’environ 6 cm.

**Calculatrice d’Alexis**

```
√((1.84^2+1.87^2+1.73^2)+3
-((1.84+1.87+1.73)+3)^2)
0.06018490028
```

1. Parmi ces différentes réponses, laquelle/lesquelles vous semble(nt) correcte(s) ?  
Expliquez pourquoi l’un·e des étudiant·e·s a obtenue une réponse erronée ?

2. 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 calculer  

$$m(X) = \frac{1,61+1,59+1,75+1,65+1,56}{5}$$

$$m(X^2) = \frac{1,61^2+1,59^2+1,75^2+1,65^2+1,56^2}{5}$$

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

$$s = \sqrt{Var}$$

Rédigez ce calcul d’une façon qui empêche des erreurs arrondis (vous pourrez reprendre la rédaction d’un·e des étudiant·e·s ayant trouvé la bonne réponse).

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

1. Calculer, sans utiliser la calculatrice, chacune des sommes suivantes :  $\sum_{i=1}^5 i$ ,  $\sum_{i=-5}^5 i$ ,  $\sum_{k=-3}^3 k^2$ .

2. Écrivez avec le symbole  $\sum$  la somme suivante  $\frac{6^2}{6+8} + \frac{7^2}{7+8} + \frac{8^2}{8+8} + \frac{9^2}{9+8} + \frac{10^2}{10+8} + \frac{11^2}{11+8}$ .
3. Calculer  $\sum_{i=1}^6 n_i (x_i)^2$  lorsque  $x_1 = 3, x_2 = 4, x_3 = 5, x_4 = 6, x_5 = 7, x_6 = 8, n_1 = 3, n_2 = 7, n_3 = 6, n_4 = 4, n_5 = 4$  et  $n_6 = 3$  (sans utiliser de liste sur la calculatrice).

**Chapitre 1 : Statistiques descriptives univariées**

**Exercice 7 : Type de variables**

1. Mme Garnier recueille des données sur
  - (a) la popularité des différents footballeurs de l’équipe de France
  - (b) le temps mis par des rats pour sortir d’un labyrinthe
  - (c) les noms des rues de Dijon
  - (d) le nombre de pièces des appartements en location à Dijon

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

2. De même quel est le type de variable lorsqu'on étudie :
- (a) les numéros de téléphones figurant dans les pages blanches dijonnaises
  - (b) le degré de satisfaction des usagers des transports en commun
  - (c) la sensibilité aux questions écologiques des habitants d'un même quartier
  - (d) le taux de testostérone parmi les patients souffrant de troubles cognitifs
  - (e) les numéros de sécurité sociale des enseignants de l'université de Bourgogne

**Exercice 8 : Taille des ménages**

On appelle  $T$  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 du ménage	1	2	3	4	5	6
Effectifs : nombre de ménages	1853	1122	1925	979	324	123

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

**Exercice 9 : Moyens de transport**

1. **Utilisation des transports en commun**

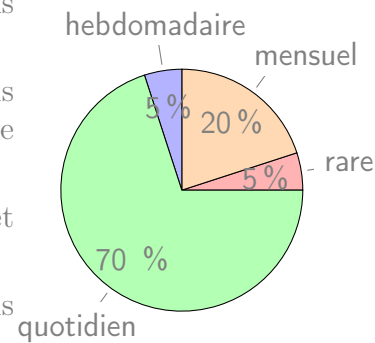
- (a) Un étudiant en sociologie interroge 20 personnes de son entourage sur leur utilisation des transports en commun. Il réalise le graphique ci-dessous après les avoir regroupés en quatre groupes :

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

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

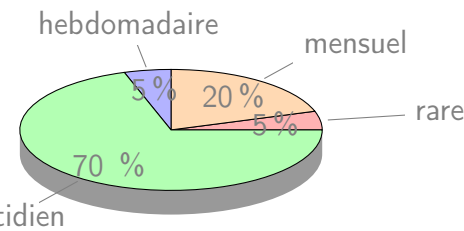
**hebdomadaire** : entre une et cinq fois par semaine

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



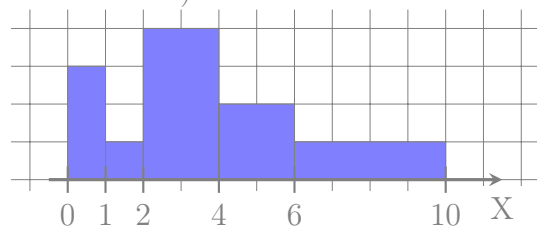
- i. Déterminer les effectifs des différentes modalités
- ii. Quelle est, au sein de cet échantillon, la proportion d'individus qui utilisent les transports en commun moins d'une fois par semaine ?

- (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, il désigne par  $X$  le le temps hebdomadaire passé dans une voiture (exprimé en heures). Il trace un histogramme sur une feuille quadrillée (et constate avec surprise que ses rectangles tombent exactement sur

le quadrillage de la feuille) :



(a) Déterminer les fréquences des différentes classes

$X$	$[0; 1[$	$[1; 2[$	$[2; 4[$	$[4; 6[$	$[6; 10[$
Fréquence					

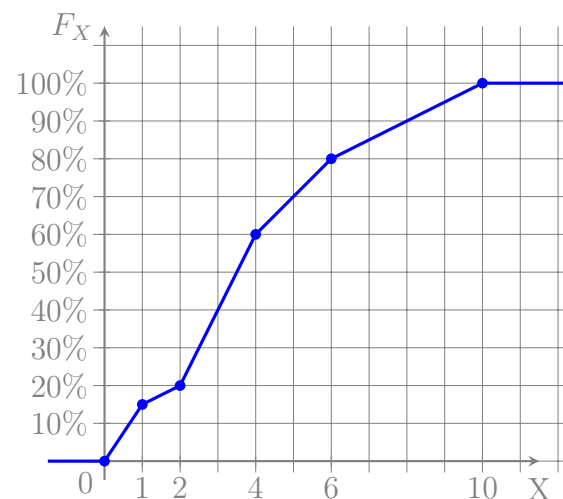
(b) Commentez : comparer les fréquences de la première et de la dernière classe, ainsi que la hauteur des rectangles correspondants.

3. **Polygone des fréquences cumulées** Enfin, il décide de représenter (toujours sur son papier quadrillé) le polygone des fréquences cumulées des données de la question 2. Celui-ci est reproduit à la fin de l'énoncé de cet exercice.

(a) Déterminer graphiquement la médiane et les quartiles approchés de  $X$ .

(b) Déterminer graphiquement les proportions  $\mathbb{P}_r[X \leq 5]$ ,  $\mathbb{P}_r[X \geq 7]$  et  $\mathbb{P}_r[5 \leq X \leq 7]$  (de manière approchée).

(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 ? Calculer aussi les médianes et écarts type.



**Exercice 11 : Développement psychomoteur**

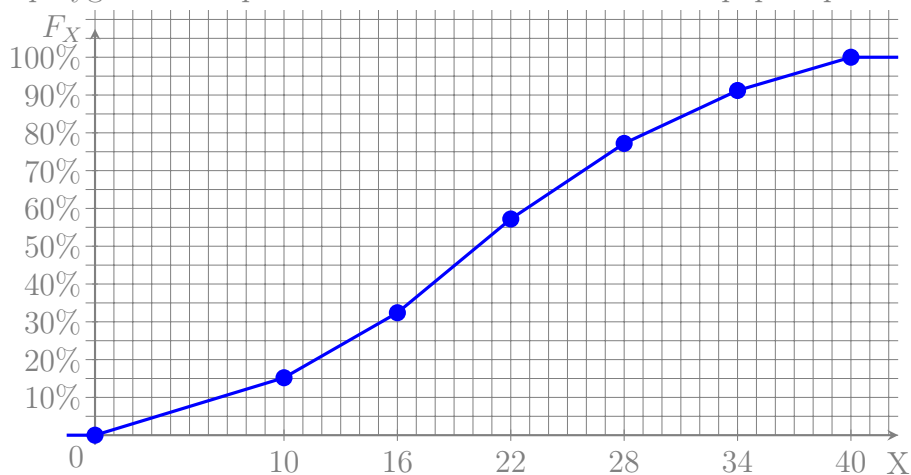
Les données ci-dessous décrivent, pour un ensemble de 200 bébés âgés de deux ans, la variable statistique  $X$  représentant le score de développement psychomoteur (SDP) :

SDP	[52 ; 62[	[62 ; 72[	[72 ; 82[	[82 ; 92[	[92 ; 102[	[102 ; 112[	[112 ; 122[	[122 ; 132[
Effectif	7	5	18	46	50	41	28	5

- Calculer la médiane et les quartiles de  $X$ .
- Calculer 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 appelle  $X$  la note obtenue par les enfants à l'issue de l'apprentissage (notée entre 0 et 40). On représente le polygone de fréquences cumulées de ces notes sur un papier quadrillé :



- Déterminer graphiquement la médiane et les quartiles de la variable statistique  $X$ .
- Donner une approximation de  $\mathbb{P}_r[X < 25]$ .
  - Quelle est environ la proportion d'enfants dont la note est supérieure à 18 ?
  - Combien vaut environ  $\mathbb{P}_r[20 \leq X < 30]$  ?

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

X	[0 ; 10[	[10 ; 16[	[16 ; 22[	[22 ; 28[	[28 ; 34[	[34 ; 40[
Effectif	38	43	62	50	35	22

- Calculer la médiane et les quartiles approchés de la variable statistique  $X$  (cette fois ci sans lecture graphique).
- Calculer la note moyenne  $m$  et l'écart type  $s$  de l'échantillon.
- Donner 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 20 personnes de 35 à 88 ans, auxquels on attribue une note indiquant leurs performances mémorielles. On note leur âge  $X$  et leur performances mémorielles  $Y$ .

Les données mesurées sont les suivantes :

X	74	42	39	62	57	43	51	43	37	71	60	70	88	35	68	51	40	54	60	36
Y	31	29	73	40	21	38	61	79	21	23	56	37	20	80	46	48	56	38	31	45

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

**Exercice 14 : Type de variables**

- Mr Delattre recueille des données sur
  - le temps hebdomadaire passé devant la télévision par des enfants
  - les langues officielles des différents pays européens
  - le nombre de pages des livres d'une bibliothèque
  - la douleur ressentie par différents patients atteints de la même pathologie

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

Cas	Population	Variable	Nature
(a)	des enfants	le temps hebdomadaire passé devant la télévision	quantitative continue
(b)	l'ensemble des pays européens	langue officielle	qualitative nominale
(c)	l'ensemble des livres de cette bibliothèque	le nombre de pages	quantitative discrète
(d)	différents patients atteints de la même pathologie	la douleur ressentie	qualitative ordinale

2. De même quel est le type de variable lorsqu'on étudie :
  - (a) le nombre de rendez-vous médicaux que différents patients ont pris au cours de l'année 2021  
quantitative discrète
  - (b) le nombre de crises d'épilepsie de différents patients au cours du mois d'octobre 2021  
quantitative discrète
  - (c) le nombre moyen d'enfants par femme dans chaque pays européen  
quantitative continue
  - (d) le poids des appareils photos vendus en 2021  
quantitative continue
  - (e) le nombre d'enfants de chaque femme dijonnaise  
quantitative discrète
  - (f) la couleur des voitures stationnées dans le campus  
qualitative nominale
  - (g) les marques des téléphones vendus en 2021  
qualitative nominale

- (h) le nombre de personnes des ménages Français  
quantitative discrète
- (i) les prénoms des enfants nés en 2021  
qualitative nominale
- (j) la pluie tombée à Dijon au cours des différents mois de l'année  
quantitative continue

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

$$\text{moyenne : } m(T) = \frac{\sum_i x_i n_i}{n} = \frac{1 \times 1853 + 2 \times 1122 + 3 \times 1925 + \dots + 6 \times 123}{6326} = \frac{16\,146}{6\,326} \simeq 2,55$$

$$m(T^2) = \frac{\sum_i x_i^2 n_i}{n} = \frac{1^2 \times 1853 + 2^2 \times 1122 + 3^2 \times 1925 + \dots + 6^2 \times 123}{6326} = \frac{51\,858}{6\,326}$$

$$\text{Var}(T) = m(T^2) - m(T)^2 = \frac{51\,858}{6\,326} - \left(\frac{16\,146}{6\,326}\right)^2 \simeq 1,68$$

$$\text{Écart-type : } s(T) = \sqrt{\text{Var}(T)} \simeq 1,3$$

**Exercice 16 : Questionnaire**

Un chercheur établit un questionnaire visant à évaluer les compétences logiques de jeunes adolescents. Il fait remplir ce questionnaire à 15 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 fausse, et a laissé 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	M	N	O
Question 1	0	0	0		✓	0	✓	✓		0	0	0	✓		
Question 2	0		0	✓		✓	✓	✓		0	0		✓	✓	
Question 3				0	0	✓		0	✓	0	✓	0	✓	✓	0
Question 4	✓		✓		✓	0			0		✓	0	0	✓	✓
Question 5		✓	✓	✓			✓	0		0		0	0	✓	
Question 6		✓	0	✓	✓	0	✓	✓	✓	✓	0	0		✓	0
Question 7	0	✓	0		✓		✓		✓	0	0	✓		0	
Question 8		0	0	✓			✓	✓	✓		0	✓	0	✓	✓
Question 9	✓	✓			✓		0	✓	✓	0		0	✓	0	0
Question 10	0	0		✓	✓	0	✓			0	✓	✓	✓	✓	0

1. Quelle a été le taux de réponse à la question 4 ?  
 En tout, 10 personnes (parmi les 15 interrogés) ont répondu à la question 4. Le taux de réponse à cette question est donc  $\frac{10}{15} \simeq 0,6667$ .

2. Quelle proportion de sujets a répondu correctement à la question 5 ?  
 Comme 5 personnes ont répondu correctement à la question 5, cette proportion est  $\frac{5}{15} \simeq 0,3333$ .

3. Parmi les individus ayant répondu à la question 8, quelle proportion à donné la bonne réponse ?  
 Parmi les 11 personnes qui ont répondu à la question 8, il y en a 7 qui ont répondu correctement, soit une proportion de  $\frac{7}{11} \simeq 0,6364$ .

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). Calculer les notes (sur 10) attribuée à chaque jeune interrogé.

Individu	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
Note	2	4	2	5	6	2	7	5	5	1	3	3	5	7	2

5. Calculer la moyenne, l'écart type et la médiane de ces notes.  
 moyenne :  $m(X) = \frac{\sum x_i}{n} = \frac{2+4+2+\dots+2}{15} = \frac{59}{15} \simeq 3,93$   
 $m(X^2) = \frac{\sum x_i^2}{n} = \frac{2^2+4^2+2^2+\dots+2^2}{15} = \frac{285}{15}$   
 $Var(X) = m(X^2) - m(X)^2 = \frac{285}{15} - \left(\frac{59}{15}\right)^2 \simeq 3,53$   
 Écart-type :  $s(X) = \sqrt{Var(X)} \simeq 1,88$   
 La médiane est la valeur numéro  $\frac{15+1}{2} = 8$  (en ordonnant par ordre croissant). C'est donc 4.

**Exercice 17 : Regroupement en classes**

Voici les notes obtenues par un groupe d'élèves lors d'un contrôle sur 10 points :  
 1 ; 7 ; 7 ; 3 ; 8,5 ; 6 ; 3,5 ; 6,5 ; 6 ; 8 ; 8 ; 6,5 ; 7 ; 6,5 ; 6 ; 7,5 ; 6 ; 5 ; 6 ; 9 ; 6 ; 5 ; 4 ; 9 ; 5 ; 3,5 ; 4 ; 6 ; 9 ; 6,5 ; 8 ; 7,5 ; 4,5 ; 5,5 ; 4 ; 5 ; 6 et 5

1. Faire un tri des données en indiquant l'effectif de chaque note.

modalité	1	3	3,5	4	4,5	5	5,5	6	6,5	7	7,5	8	8,5	9
effectif	1	1	2	3	1	5	1	8	4	3	2	3	1	3

2. Quelle est la taille de l'échantillon ?  
 La taille de l'échantillon est  $n = 38$ .

3. Donner la moyenne, l'écart type et la médiane.  
 moyenne :  $m(X) = \frac{\sum_i x_i n_i}{n} = \frac{1 \times 1 + 3 \times 1 + 3,5 \times 2 + \dots + 9 \times 3}{38} = \frac{227,5}{38} \simeq 5,99$   
 $m(X^2) = \frac{\sum_i x_i^2 n_i}{n} = \frac{1^2 \times 1 + 3^2 \times 1 + 3,5^2 \times 2 + \dots + 9^2 \times 3}{38} = \frac{1481,75}{38}$   
 $Var(X) = m(X^2) - m(X)^2 = \frac{1481,75}{38} - \left(\frac{227,5}{38}\right)^2 \simeq 3,151$   
 Écart-type :  $s(X) = \sqrt{Var(X)} \simeq 1,78$   
 La médiane est la valeur numéro  $\frac{38+1}{2} = 19,5$  (en ordonnant par ordre croissant). C'est donc 6.

4. Transformer les données en les rangeant en classes d'amplitude 1,5. Calculer les fréquences et les fréquences cumulées des classes.

classes	[0 ; 1,5[	[1,5 ; 3[	[3 ; 4,5[	[4,5 ; 6[	[6 ; 7,5[	[7,5 ; 9[	[9 ; 10,5[
effectifs	1	0	6	7	15	6	3
fréquences	0,026	0	0,158	0,184	0,395	0,158	0,079
fréquences cumulées	0,026	0,026	0,184	0,368	0,763	0,921	1

5. Calculer la moyenne, l'écart type et la médiane des notes regroupées en classes. Que remarque-t-on ?  
 moyenne :  
 $m(X) = \frac{\sum c_i n_i}{n} = \frac{0,75 \times 1 + 2,25 \times 0 + 3,75 \times 6 + \dots + 9,75 \times 3}{38} = \frac{240}{38} \simeq 6,32$   
 $m(X^2) = \frac{\sum c_i^2 n_i}{n} = \frac{0,75^2 \times 1 + 2,25^2 \times 0 + 3,75^2 \times 6 + \dots + 9,75^2 \times 3}{38} = \frac{1654,875}{38}$   
 $Var(X) = m(X^2) - m(X)^2 = \frac{1654,875}{38} - \left(\frac{240}{38}\right)^2 \simeq 3,66$

Écart-type :  $s(X) = \sqrt{Var(X)}$   
 $\simeq 1,91$

Classe de la médiane :  $[6; 7,5[$

Méd  $\simeq a_i + \frac{a_{i+1}-a_i}{F_X(a_{i+1})-F_X(a_i)}(0,5 - F_X(a_i)) \simeq 6 + \frac{7,5-6}{0,763-0,368}(0,5 - 0,368) \simeq 6,501$

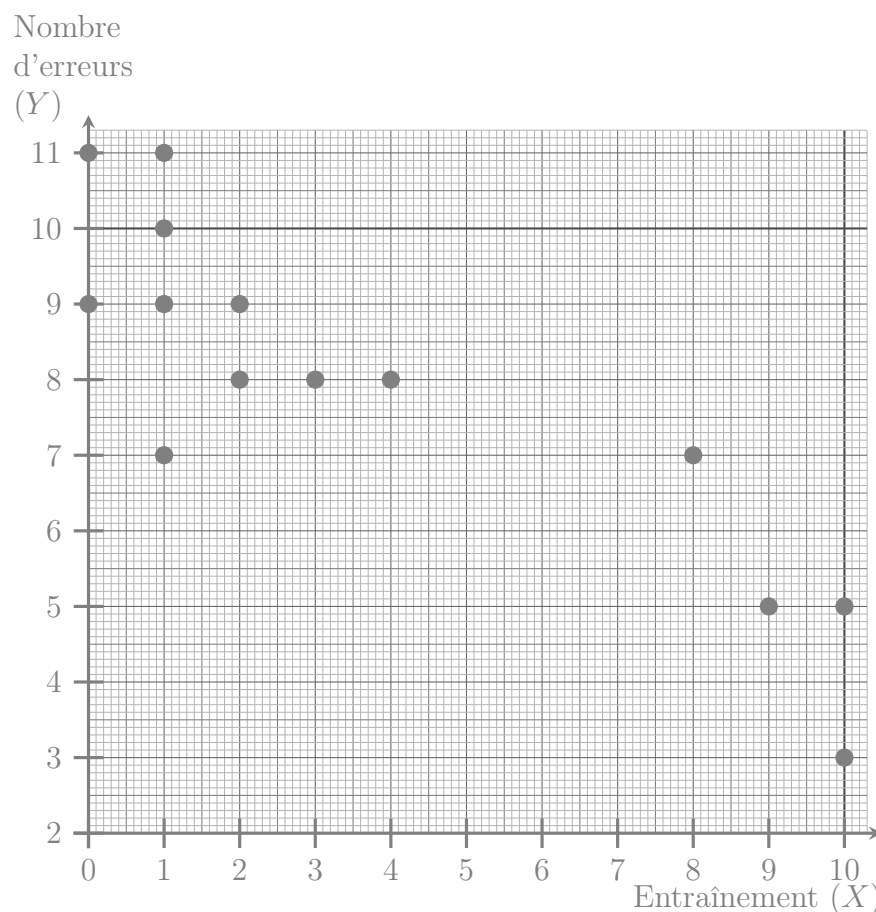
Ce regroupement en classe donne lieu à des erreurs de l'ordre de 0,25 points (dans le cas de la moyenne), ce qui est de l'ordre de 4% de la note.

### Chapitre 2 : Statistiques descriptives bivariées

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

Pour étudier l'entraînement d'enfants passant un test de logique, 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 considère le nombre  $Y$  d'erreur 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 du nuage de points suivant (effectué sur un papier "millimétré") :



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

Entraînement $X$														
Nombre d'erreurs $Y$														

2. Calculer le coefficient de corrélation des rangs de Spearman. Interpréter la valeur de ce coefficient.

3. Calculer les moyennes et écarts type des variables  $X$  et  $Y$ , puis leur coefficient de corrélation linéaire. Interpréter la valeur de ce coefficient.
4. Si un enfant a déjà fait 2 exercices de logique, alors combien estimeriez qu'il fera d'erreurs si on lui fait à nouveau passer un test similaire ?
5. Un enfant a commis 9 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 $X$	8,19	8,39	2,5	7,33	3,52	2,87	3,14	3,56	2,96	3,14
temps de réponse $Y$	701	671	538	716	479	495	509	503	530	505

1. Déterminer le coefficient de corrélation des rangs de Spearman. Ces variables sont-elles très corrélées ?
2. Déterminer le coefficient de corrélation linéaire de ces deux variables. Ces variables sont-elles très corrélées ?
3. Suite à une perte d'échantillon, on ne connaît plus le niveau de stress d'un rat, dont on a mesuré qu'il met 687ms à réagir aux stimuli visuels. Déterminer environ son niveau de stress.
4. Tracer le nuage de point des variables  $X$  et  $Y$ .
5. Commenter vos réponses aux questions précédentes 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  $Y$  dans 11 pays différents :

pays	revenu/hab. $X$	espérance de vie $Y$
Algérie	15,2	75,6
Brunei	76,7	77,7
Chili	24,6	80,5
Djibouti	3,6	63,5
Estonie	31,5	77,6
Finlande	44	81,1
Guinée-Bissau	1,8	58,9
Honduras	5,5	74,6
Indonésie	12,4	69,1
Jamaïque	9,2	76,2
Kiribati	2	66,3

1. Déterminer les médianes du revenu  $X$  et de l'espérance de vie  $Y$ .
2. Déterminer le coefficient de corrélation linéaire.
3. Déterminer le coefficient de corrélation des rangs (de Spearman).
4. Commenter (on pourra tracer le nuage de points).
5. On ajoute à ces données un pays supplémentaire (dont l'existence est sujette à caution) nommé "Utopia", où le revenu par habitant est de 1 550 et l'espérance de vie est de 1 130. Calculer les moyennes et les médianes que l'on obtiendrait alors, ainsi que les coefficients de corrélation. Commenter.

**Exercice 21 : Groupe d'étudiants**

On demande à un groupe d'étudiants leur nombre de frères et soeurs, leur humeur, leur taille, et la note qu'ils ont eue au baccalauréat. On obtient les données suivantes :

Hélène	Nicole	Sylvie
de mauvaise humeur 0 frère/soeur 1m69 bac : 14,12/20	de bonne humeur 1 frère/soeur 1m66 bac : 12,93/20	de bonne humeur 2 frères/soeurs 1m71 bac : 12,22/20

Jacques	Gilles	Françoise
de bonne humeur 0 frère/soeur 1m84 bac : 13,95/20	de très bonne humeur 1 frère/soeur 1m86 bac : 11,78/20	de très bonne humeur 1 frère/soeur 1m65 bac : 10,09/20

Florian	Marie	Angelique
de bonne humeur 3 frères/soeurs 1m76 bac : 12,37/20	de relativement bonne humeur 0 frère/soeur 1m58 bac : 10,74/20	de bonne humeur 2 frères/soeurs 1m76 bac : 14,2/20

Calculer le coefficient de corrélation linéaire entre la taille et la note obtenue au bac, pour les étudiants de cet échantillon.

**Exercice 22 : Thérapie pour réduire l'anxiété**

Pour mettre en évidence l'efficacité d'une thérapie visant à réduire l'anxiété de personnes victimes d'agressions, nous avons 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'anxiété). Les données sont les suivantes.

Avant	40	55	25	39	25	34	32	21	36	33	37	36	33	11	26
Après	26	48	7	43	20	12	23	13	26	14	16	24	25	6	13

1. Calculer les coefficients de corrélation linéaire et de Spearman entre les deux phases.
2. Parmi les deux droites de régression, y en a-t-il une qui a plus d'intérêt ? Déterminer son équation.

**Exercice 23 : Extraversion**

11 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 les résultats suivants.

sociabilité X	3,7	5,7	4,4	7,6	10,8	9,3	9,7	11,3	6	4,9	5,5
impulsivité Y	4,9	2,1	2,1	4,1	4,9	3,3	1	4,3	1,9	3,8	3,5

On souhaite savoir si ces deux échelles mesurent plusieurs manifestations d'un même caractère : l'extraversion. Si c'est le cas, on pourra additionner les scores obtenus aux deux échelles pour former une échelle unique d'extraversion.

1. Calculer le coefficient de corrélation linéaire des deux échelles X et Y.

$$\begin{aligned} \text{moyenne : } m(X) &= \frac{\sum x_i}{n} = \frac{3,7+5,7+4,4+\dots+5,5}{11} = \frac{78,9}{11} \simeq 7,17 \\ m(X^2) &= \frac{\sum x_i^2}{n} = \frac{3,7^2+5,7^2+4,4^2+\dots+5,5^2}{11} = \frac{638,47}{11} \\ \text{Var}(X) &= m(X^2) - m(X)^2 = \frac{638,47}{11} - \left(\frac{78,9}{11}\right)^2 \simeq 6,59 \\ \text{Écart-type : } s(X) &= \sqrt{\text{Var}(X)} \simeq 2,57 \end{aligned}$$

$$\begin{aligned} \text{moyenne : } m(Y) &= \frac{\sum x_i}{n} = \frac{4,9+2,1+2,1+\dots+3,5}{11} = \frac{35,9}{11} \simeq 3,26 \\ m(Y^2) &= \frac{\sum x_i^2}{n} = \frac{4,9^2+2,1^2+2,1^2+\dots+3,5^2}{11} = \frac{134,33}{11} \\ \text{Var}(Y) &= m(Y^2) - m(Y)^2 = \frac{134,33}{11} - \left(\frac{35,9}{11}\right)^2 \simeq 1,56 \\ \text{Écart-type : } s(Y) &= \sqrt{\text{Var}(Y)} \simeq 1,25 \end{aligned}$$

$$\begin{aligned} m(XY) &= \frac{\sum x_i y_i}{n} = \frac{3,7 \times 4,9 + 5,7 \times 2,1 + \dots + 5,5 \times 3,5}{11} = \frac{261,67}{11} \simeq 23,788 \\ \text{Cov}(X,Y) &= m(XY) - m(X)m(Y) = \frac{261,67}{11} - \frac{78,9}{11} \frac{35,9}{11} \simeq 0,379 \\ r(X,Y) &= \frac{\text{Cov}(X,Y)}{s(X)s(Y)} = \frac{0,379}{2,57 \times 1,25} \simeq 0,118 \end{aligned}$$

2. Serait-il pertinent de résumer ces deux échelles par une échelle d'extraversion ?

On voit donc que les deux variables ne sont pas très liées. Les résumer par leur somme fait perdre de l'information. Il semblerait plus pertinent de garder deux échelles distinctes (sociabilité et impulsivité) plutôt que de les fusionner en une seule.

**Exercice 24 : Reconnaissance de forme**

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

Pour un échantillon de 13 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	2	3	3	3	4	4	5	6	8	8	9	10
Temps de réponse $Y$	4	3,5	3,3	3,3	3,7	2,6	3,3	2,4	2,3	1,4	1,4	1,5	1,1

1. Calculer le coefficient de corrélation des rangs de Spearman. Interpréter la valeur de ce coefficient.

sujet	1	2	3	4	5	6	7	8	9	10	11	12	13
$X$	0	2	3	3	3	4	4	5	6	8	8	9	10
$Y$	4	3,5	3,3	3,3	3,7	2,6	3,3	2,4	2,3	1,4	1,4	1,5	1,1
rang $X'$	1	2	4	4	4	6,5	6,5	8	9	10,5	10,5	12	13
rang $Y'$	13	11	9	9	12	7	9	6	5	2,5	2,5	4	1
$(X' - Y')^2$	144	81	25	25	64	0,25	6,25	4	16	64	64	64	144

le coefficient de corrélation des rangs de Spearman est donc

$$1 - \left( 6 \times \frac{144+81+25+25+64+\dots+144}{13(13^2-1)} \right) \simeq -0,927$$

Cela montre un fort lien entre les deux variables, où  $Y$  diminue quand  $X$  augmente.

2. Calculer les moyennes et écarts type des variables  $X$  et  $Y$ , puis leur coefficient de corrélation linéaire. Interpréter la valeur de ce coefficient.

$$\text{moyenne : } m(X) = \frac{\sum x_i}{n} = \frac{0+2+3+\dots+10}{13} = \frac{65}{13} = 5$$

$$m(X^2) = \frac{\sum x_i^2}{n} = \frac{0^2+2^2+3^2+\dots+10^2}{13} = \frac{433}{13}$$

$$Var(X) = m(X^2) - m(X)^2 = \frac{433}{13} - \left(\frac{65}{13}\right)^2 \simeq 8,31$$

$$\text{Écart-type : } s(X) = \sqrt{Var(X)} \simeq 2,88$$

$$\text{moyenne : } m(Y) = \frac{\sum x_i}{n} = \frac{4+3,5+3,3+\dots+1,1}{13} = \frac{33,8}{13} = 2,6$$

$$m(Y^2) = \frac{\sum x_i^2}{n} = \frac{4^2+3,5^2+3,3^2+\dots+1,1^2}{13} = \frac{99,8}{13}$$

$$Var(Y) = m(Y^2) - m(Y)^2 = \frac{99,8}{13} - \left(\frac{33,8}{13}\right)^2 \simeq 0,917$$

$$\text{Écart-type : } s(Y) = \sqrt{Var(Y)} \simeq 0,958$$

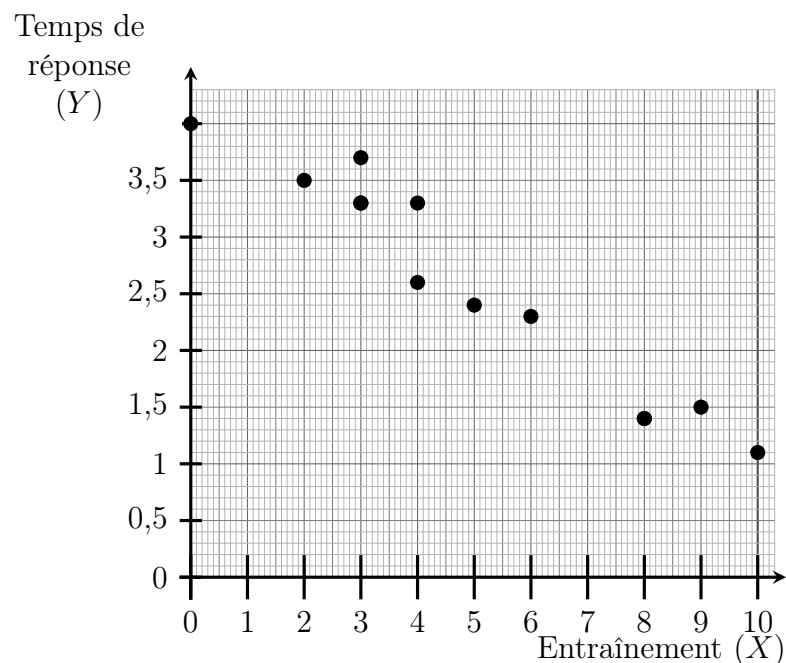
$$m(XY) = \frac{\sum x_i y_i}{n} = \frac{0 \times 4 + 2 \times 3,5 + \dots + 10 \times 1,1}{13} = \frac{134,2}{13} \simeq 10,323$$

$$Cov(X,Y) = m(XY) - m(X)m(Y) = \frac{134,2}{13} - \frac{65}{13} \frac{33,8}{13} \simeq -2,677$$

$$r(X,Y) = \frac{Cov(X,Y)}{s(X)s(Y)} = \frac{-2,677}{2,88 \times 0,958} \simeq -0,97$$

Cela montre un fort lien linéaire entre les deux variables, où  $Y$  diminue quand  $X$  augmente.

3. Tracer le nuage de points correspondant à ces données. Vous conforte t-il dans les interprétation données aux questions précédentes ?



On remarque que dans l'ensemble, les points sont assez proches d'une droite, comme attendu à partir de leurs coefficients de corrélation.

4. Si un enfant a déjà fait 1 exercices de reconnaissance de forme, alors à combien estimeriez vous son temps de réponse si on lui fait à nouveau passer ce test ?

Pour répondre à cette question, on utilise la droite  $D_{Y|X}$  (qui fait sens car il y a une forte corrélation linéaire) :

on pose  $a = \frac{Cov(X,Y)}{Var(X)} \simeq \frac{-2,677}{8,31} \simeq -0,32$  et  $b = m(Y) - a m(X) \simeq 2,6 - (-0,322) \times 5 \simeq 4,21$

D'où l'équation de la droite  $D_{Y|X} : Y = -0,32 X + 4,21$

Donc pour  $x = 1$ , on s'attend à  $y = -0,32 + 4,21 = 3,89$  .

5. Un enfant a mis 3,2 secondes pour répondre au test. Combien de fois estimeriez vous qu'il avait déjà fait un test similaire pour s'entraîner ?

Pour répondre à cette question, on utilise la droite  $D_{X|Y}$  :

on pose  $a' = \frac{Cov(X,Y)}{Var(Y)} \simeq \frac{-2,68}{0,917} \simeq -2,92$  et  $b' = m(X) - a' m(Y) \simeq 5 - (-2,92) \times 2,6 \simeq 12,59$

D'où l'équation de la droite  $D_{X|Y} : X = -2,92 Y + 12,59$

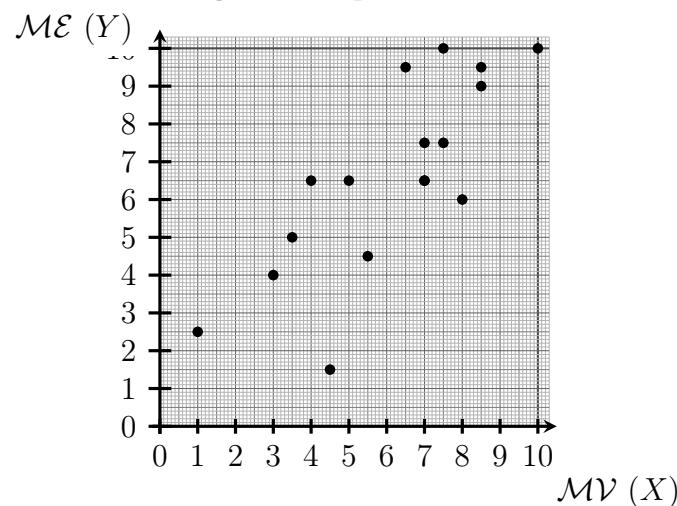
Donc pour  $y = 3,2$ , on s'attend à  $x = -2,92 \times 3,2 + 12,59 = 3,246$

### 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, la seconde ( $\mathcal{M}\mathcal{E}$ ) 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 apprentissages.

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

1. Dessiner le nuage statistique de ces variables.





2. Calculer le coefficient de corrélation des rangs de Spearman des deux variables  $X$  et  $Y$ .

sujet	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
$X$	4,5	8	10	3	7	5	7	4	7,5	3,5	6,5	5,5	8,5	7,5	1	7	8,5
$Y$	1,5	6	10	4	6,5	6,5	6,5	6,5	10	5	9,5	4,5	9,5	7,5	2,5	7,5	9
rang $X'$	5	14	17	2	10	6	10	4	12,5	3	8	7	15,5	12,5	1	10	15,5
rang $Y'$	1	6	16,5	3	8,5	8,5	8,5	8,5	16,5	5	14,5	4	14,5	11,5	2	11,5	13
$(X' - Y')^2$	16	64	0,25	1	2,25	6,25	2,25	20,25	16	4	42,25	9	1	1	1	2,25	6,25

le coefficient de corrélation des rangs de Spearman est donc

$$1 - \left( 6 \times \frac{16+64+0,25+1+2,25+\dots+6,25}{17(17^2-1)} \right) \simeq 0,761$$

3. Calculer les moyennes, les écarts types et le coefficient de corrélation linéaire des deux variables  $X$  et  $Y$ .

moyenne :  $m(X) = \frac{\sum x_i}{n} = \frac{4,5+8+10+\dots+8,5}{17} = \frac{104}{17} \simeq 6,12$

$m(X^2) = \frac{\sum x_i^2}{n} = \frac{4,5^2+8^2+10^2+\dots+8,5^2}{17} = \frac{724}{17}$

$Var(X) = m(X^2) - m(X)^2 = \frac{724}{17} - \left(\frac{104}{17}\right)^2 \simeq 5,16$

Écart-type :  $s(X) = \sqrt{Var(X)} \simeq 2,27$

moyenne :  $m(Y) = \frac{\sum x_i}{n} = \frac{1,5+6+10+\dots+9}{17} = \frac{112,5}{17} \simeq 6,62$

$m(Y^2) = \frac{\sum x_i^2}{n} = \frac{1,5^2+6^2+10^2+\dots+9^2}{17} = \frac{848,75}{17}$

$Var(Y) = m(Y^2) - m(Y)^2 = \frac{848,75}{17} - \left(\frac{112,5}{17}\right)^2 \simeq 6,13$

Écart-type :  $s(Y) = \sqrt{Var(Y)} \simeq 2,48$

$m(XY) = \frac{\sum x_i y_i}{n} = \frac{4,5 \times 1,5 + 8 \times 6 + \dots + 8,5 \times 9}{17} = \frac{763,75}{17} \simeq 44,926$

$Cov(X,Y) = m(XY) - m(X)m(Y) = \frac{763,75}{17} - \frac{104}{17} \frac{112,5}{17} \simeq 4,442$

$r(X,Y) = \frac{Cov(X,Y)}{s(X)s(Y)} = \frac{4,442}{2,27 \times 2,48} \simeq 0,789$

4. 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 :

(a) Quelle droite de régression peut-on utiliser pour répondre à cette question ? Donner son équation.

La droite qui est pertinente pour cette question est  $D_{Y|X}$ .

on pose  $a = \frac{Cov(X,Y)}{Var(X)} \simeq \frac{4,442}{5,16} \simeq 0,8609$  et  $b = m(Y) - a m(X) \simeq 6,618 - 0,8609 \times 6,118 \simeq 1,351$

D'où l'équation de la droite  $D_{Y|X} : Y = 0,8609 X + 1,351$

(b) Si un enfant a obtenu une note  $x = 9$ , donner une estimation de sa note  $y$ .

Donc pour  $x = 9$ , on s'attend à  $y = 0,8609 \times 9 + 1,351 \simeq 9,099$

### Chapitre 3 : probabilités

#### Exercice 26 : Emploi

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

Nom	Mr Girard	Mr Francois	Mme Jourdain	Mme Lucas	Mr Huet
Situation	emploi	chômage	emploi	retraite	retraite

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

- Lister 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 hommes dont un a un emploi ?
  - moins de 2 femmes ?
  - deux personnes en situation d'emploi ?
- On note  $X$  le nombre d'hommes parmi les trois personnes choisies au hasard.  
Déterminer la loi de  $X$  (c'est à dire, calculer chaque probabilité  $\mathbb{P}[X = \dots]$ ).

#### Exercice 27 : Factorielle, coefficients binomiaux

- Simplifier les quantités  $\frac{21!}{19!}$ ,  $\frac{16!}{12! \times 4!}$ .
- Écrire à l'aide de deux factorielles l'expression ci-contre :  $8 \times 9 \times 10$ .
- Calculer les nombres suivants :  $\binom{16}{4}$ ,  $\binom{15}{6}$ ,  $\binom{15}{9}$ ,  $\binom{8}{0}$ ,  $\binom{13}{1}$ .

**Exercice 28 : Croyance religieuse**

- Olivier et Brigitte 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
  - Qu'ils développent tous deux une croyance religieuse.
  - Qu'un seul d'entre eux devienne athée.
  - Qu'ils deviennent tous les deux athées.

Argumenter que ces probabilités correspondent à une loi binomiale, dont vous donnerez les paramètres.

- 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.  
Calculer la loi de  $X$ .
- 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.  
Calculer la loi de  $X$ .
- Comparer 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, et 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 parmi les employés de l'entreprise, un échantillon de 20 employés au hasard "avec remise".

- Calculer la probabilité que l'échantillon contienne exactement 3 employés notés  $A$ .
- Quelle est la probabilité que l'échantillon contienne au moins 3 individus notés  $A$  ?
- Quelle probabilité y a-t-il que l'échantillon contienne au moins 18 individus notés  $A$  ou  $C$  ?
- Calculer la probabilité que l'échantillon contienne moins que 5 personnes notées  $B$ .
- Quel est le nombre moyen d'individus notés  $C$  au sein d'un échantillon aléatoire de 20 employés ?

**Exercice 30 : Loi binomiale**

On considère une loi binomiale  $X \sim \mathcal{B}(n, p)$  de moyenne  $m(X) = 5,2$  et de variance  $\text{Var}(X) = 3,12$ .

- Calculer  $n$  et  $p$ .

$$\begin{cases} np = m(X) = 5,2 \\ np(1-p) = \text{Var}(X) = 3,12 \end{cases}$$

$$\text{d'où } 1-p = \frac{3,12}{5,2} = 0,6 \text{ donc } p = 0,4 \text{ et } n = \frac{5,2}{0,4} = 13.$$

2) Calculer les probabilités  $\mathbb{P}[6 \leq X \leq 8]$  et  $\mathbb{P}[X > 10]$ .

$$\begin{aligned} \bullet \mathbb{P}[6 \leq X \leq 8] &= \mathbb{P}[X = 6] + \mathbb{P}[X = 7] + \mathbb{P}[X = 8] \\ &= \binom{13}{6} \times 0,4^6 (1 - 0,4)^7 + \binom{13}{7} \times 0,4^7 (1 - 0,4)^6 + \binom{13}{8} \times 0,4^8 (1 - 0,4)^5 \\ &\simeq 0,1968 + 0,1312 + 0,0656 \\ &= 0,3936 \end{aligned}$$

**Remarque :** Avec la calculatrice, on obtient directement  $\mathbb{P}[6 \leq X \leq 8] = \mathbb{P}[X \leq 8] - \mathbb{P}[X \leq 5] \simeq 0,9679 - 0,5744 = 0,3935$

$$\begin{aligned} \bullet \mathbb{P}[X > 10] &= \mathbb{P}[X = 11] + \mathbb{P}[X = 12] + \mathbb{P}[X = 13] \\ &= \binom{13}{11} \times 0,4^{11} (1 - 0,4)^2 + \binom{13}{12} \times 0,4^{12} (1 - 0,4) + \binom{13}{13} \times 0,4^{13} \\ &\simeq 0,001178 + 0,000131 + 7,0 \times 10^{-06} \\ &= 0,001316 \end{aligned}$$

**Remarque :** Avec la calculatrice, on obtient directement  $\mathbb{P}[X > 10] = 1 - \mathbb{P}[X \leq 10] \simeq 1 - 0,998685 = 0,001315$

3) Trouver toutes les valeurs de  $k$  telles que  $\mathbb{P}[X \geq k] \leq 0,05$ .

$$\mathbb{P}[X \geq 13] \simeq 6,711 \times 10^{-06}$$

$$\mathbb{P}[X \geq 12] \simeq 0,0001376$$

$$\mathbb{P}[X \geq 11] \simeq 0,001315$$

$$\mathbb{P}[X \geq 10] \simeq 0,007793$$

$$\mathbb{P}[X \geq 9] \simeq 0,03208$$

$$\mathbb{P}[X \geq 8] \simeq 0,09767$$

Donc les valeurs de  $k$  qui conviennent sont lorsque  $k > 8$ .

**Remarque :** Un tel calcul sera utile en deuxième année pour savoir dans quels cas on peut (ou pas) conclure comme dans l'exercice 33.

### Exercice 31 : Troubles de l'humeur et emploi

On considère trois personnes souffrant de troubles de l'humeur : Alice, Bernard et Cécile. Parmi eux, Alice et Bernard exercent un emploi alors que Cécile est sans emploi.

- 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).

- Lister les neuf possibilités pour ces deux noms choisis au hasard.

En notant par exemple "AB" pour "Alice puis Bernard", les neuf possibilités sont :

AA, AB, AC, BA, BB, BC, CA, CB et CC.

- Parmi ces possibilités, combien comptent uniquement des personnes en situation d'emploi ?

Il y en a 4 : AA, AB, BA et BB.

- Quelle est la proportion, parmi ces possibilités, qui comptent une personne sans emploi, et une qui exerce un emploi ?

Cela correspond aux 4 possibilités AC, BC, CA et CB.

Donc à la proportion  $\frac{4}{9} \simeq 0,4444$ .

- Pour un tel choix aléatoire de deux noms, on note  $X$  le nombre de noms choisis qui correspondent à une personne exerçant un emploi. Quelle est la loi de variable  $X$  ? On réalise deux fois de suite l'expérience "choisir un des trois noms et constater si c'est une personne qui exerce un emploi", de manière indépendante. À chaque fois il y a deux chances sur trois de choisir quelqu'un qui exerce un emploi, donc  $X$  suit la loi  $\mathcal{B}(2, \frac{2}{3})$ .

- Retrouver le résultat de la question c), en utilisant cette loi.

$$\mathbb{P}[X = 1] = \binom{2}{1} (2/3)^1 (1/3)^1 \simeq 0,444.$$

- Déterminer de même  $\mathbb{P}[X = 0]$  et  $\mathbb{P}[X = 2]$ .

$$\bullet \mathbb{P}[X = 0] = \binom{2}{0} (2/3)^0 (1/3)^2 \simeq 0,111.$$

$$\bullet \mathbb{P}[X = 2] = \binom{2}{2} (2/3)^2 (1/3)^0 \simeq 0,444.$$

- On décide d'une autre façon de choisir deux noms : on choisit un premier nom au hasard, puis on choisit le second nom en demandant qu'ils soit différent du premier. Reprendre, avec cette nouvelle façon de choisir les deux noms, les questions 1a, 1b et 1c. On prendra garde qu'en question 1a le nombre de cas ne sera plus égal à neuf.

- (a) AB, AC, BA, BC, CA et CB.
- (b) Il y en a 2 : AB et BA.
- (c) Cela correspond aux 4 possibilités AC, BC, CA et CB.  
Donc à la proportion  $\frac{4}{6} \simeq 0,6667$ .

### Exercice 32 : Collection de dessins

Mme Pasquier est pédiatre et utilise fréquemment de petites illustrations pour stimuler les enfants. Elle a acheté à cet effet 6 cartes illustrées qu'elle stocke dans un tiroir.

1. Elle mélange les cartes, et constate dans quelle ordre elles se retrouvent à l'issue du mélange. Quelle était la probabilité qu'à l'issue du mélange, elles se retrouvent précisément dans cet ordre ?  
Il y a  $6! = 6 \times 5 \times \dots \times 1 = 720$  résultats possibles à l'issue du mélange. La probabilité est donc  $\frac{1}{720} \simeq 0,001389$ .
2. Si elle remélange les cartes, quelle est la probabilité qu'après le mélange
  - (a) la carte qui était en haut du tas soit à nouveau en haut du tas ?  
Il y a  $5! = 5 \times 4 \times 3 \times 2 = 120$  mélanges qui conservent la première carte (car cela correspond à ne mélanger que les 5 autres cartes). La probabilité est donc  $\frac{120}{720} \simeq 0,1667$ .
  - (b) les deux cartes qui étaient en haut du tas soient à nouveau en haut du tas, et dans le même ordre ?  
Il y a  $4! = 4 \times 3 \times 2 = 24$  mélanges qui conservent les deux premières cartes (car cela correspond à ne mélanger que les 4 autres cartes). La probabilité est donc  $\frac{24}{720} \simeq 0,0333$ .
  - (c) les deux cartes qui étaient en haut du tas soient à nouveau en haut du tas, mais dans l'ordre inverse ?  
Il y a autant de cas que dans la question précédente : en effet si on en dressait la liste, il suffirait de prendre la même liste et échanger à chaque fois les deux premières cartes. En conséquence, on a encore la probabilité  $\frac{24}{720} \simeq 0,0333$ .

- (d) les deux cartes qui étaient en haut du tas soient à nouveau en haut du tas, dans n'importe quel ordre ?  
Il suffit d'additionner les probabilités des deux dernières questions. On obtient  $\frac{24}{720} + \frac{24}{720} \simeq 0,06667$ .
- (e) l'ensemble des cartes soient exactement dans le même ordre qu'avant ?  
 $\frac{1}{720} \simeq 0,001389$ .

3. Répondre aux mêmes questions en supposant qu'il y ait cette fois-ci 31 cartes.

1.  $\frac{1}{31!} = \frac{1}{31 \times 30 \times \dots \times 1} = 1,216125041553518 \times 10^{-34}$
2. (a)  $\frac{30!}{31!} = \frac{30 \times 29 \times \dots \times 1}{31 \times 30 \times \dots \times 1} = \frac{1}{31} \simeq 0,03226$   
 (b)  $\frac{29!}{31!} = \frac{29 \times 28 \times \dots \times 1}{31 \times 30 \times \dots \times 1} = \frac{1}{31 \times 30} \simeq 0,0010753$   
 (c)  $\frac{1}{31 \times 30} \simeq 0,0010753$   
 (d)  $\frac{1}{31 \times 30} + \frac{1}{31 \times 30} \simeq 0,0021505$   
 (e)  $\frac{1}{31!} = \frac{1}{31 \times 30 \times \dots \times 1} = 1,216125041553518 \times 10^{-34}$

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

Les probabilités les plus faibles sont celles des questions 1 et 2e (probabilité déjà très faible pour 6 cartes, et quasi-impossible pour 31 cartes).

Dans la question 1 on ne peut rien en déduire : chaque résultat à l'issue du mélange a une très faible probabilité. Le résultat sur lequel on tombe sera forcément un parmi des milliards (de milliards), il avait extrêmement peu de chances de tomber sur celui-là en particulier, mais il faut tomber sur l'un de ces résultats.

En revanche, dans la question 2e, on peut vraiment déduire quelque chose : quand on remélange un paquet de cartes il est très improbable (pour 6 cartes) et même quasiment impossible (pour 31 cartes) de reproduire le même ordre qu'initialement.

**Exercice 33 : Thérapie contre la dépression**

Un psychologue a développé une nouvelle thérapie contre la dépression, et il affirme que cette thérapie permet la rémission de 70% des patients.

En interrogeant un échantillon de 22 patients, on constate 6 rémissions.

- 1) S'il y avait 70% 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?

Si l'on désigne par  $X$  le nombre de rémissions dans un échantillon aléatoire de 22 patients, et si l'on suppose qu'il y ait 70% de rémission parmi l'ensemble des patients, alors  $X$  suit la loi  $\mathcal{B}(22; 0,7)$ .

Le nombre moyen de rémissions au sein d'un tel échantillon serait alors  $m(X) = 22 \times 0,7 = 15,4$ .

- 2) Sous cette hypothèse, quelle serait la probabilité d'avoir au maximum 6 rémissions?

$$\begin{aligned} \mathbb{P}[X \leq 6] &= \mathbb{P}[X = 0] + \mathbb{P}[X = 1] + \dots + \mathbb{P}[X = 6] \\ &= \binom{22}{0} (1 - 0,7)^{22} + \dots + \binom{22}{6} \times 0,7^6 (1 - 0,7)^{16} \\ &\simeq 0 + 0 + 0 + 6,0 \times 10^{-08} + 6,8 \times 10^{-07} + 5,72 \times 10^{-06} + 3,779 \times 10^{-05} \\ &= 4,425 \times 10^{-05} \end{aligned}$$

- 3) Conclure : vous semble-t-il vraisemblable qu'il y ait, comme l'affirme ce psychologue, 70% de rémissions parmi l'ensemble des patients?

On a obtenu que s'il y avait bien 70% de rémission parmi l'ensemble des patients (comme l'affirme ce psychologue), alors il y aurait très peu de chances (environ 0,004% de chances) d'avoir un échantillon où il n'y ait pas plus de 6 rémissions. Le fait que nous soyons tombé sur un tel échantillon donne une forte présomption que le taux de rémissions n'est en fait pas de 70% (à moins par exemple que notre échantillon soit biaisé à cause de la façon dont il a été établi).

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

Dans une classe, il y a 18 filles et 14 garçons. On choisit au hasard 5 élèves distincts.

1. Quel est le nombre de choix possibles?

On choisit 5 personnes parmi l'ensemble des  $18 + 14 = 32$  élèves. Le nombre de choix possibles est donc

$$\begin{aligned} \binom{32}{5} &= \frac{32!}{5! \times 27!} = \frac{32 \times 31 \times \dots \times 1}{(5 \times 4 \times 3 \times 2) \times (27 \times 26 \times \dots \times 1)} \\ &= (32 \times 31 \times 30 \times 29 \times 28) \times \frac{1}{5 \times 4 \times 3 \times 2} = \frac{32 \times 31 \times 5 \times 6 \times 29 \times 28}{5 \times 4 \times 3 \times 2} \\ &= \frac{32 \times 31 \times 6 \times 29 \times 28}{4 \times 3 \times 2} = \frac{4 \times 8 \times 31 \times 6 \times 29 \times 28}{4 \times 3 \times 2} \\ &= \frac{8 \times 31 \times 6 \times 29 \times 28}{3 \times 2} = \frac{8 \times 31 \times 3 \times 2 \times 29 \times 28}{3 \times 2} \\ &= 8 \times 31 \times 29 \times 28 = 201376 \end{aligned}$$

2. Quel est le nombre de choix ne comportant que des garçons?

Ces choix correspondent à choisir 5 élèves parmi les 14 garçons. Le nombre de choix correspondant est donc

$$\begin{aligned} \binom{14}{5} &= \frac{14!}{5! \times 9!} = \frac{14 \times 13 \times \dots \times 1}{(5 \times 4 \times 3 \times 2) \times (9 \times 8 \times \dots \times 1)} \\ &= \frac{14 \times 13 \times 12 \times 11 \times 10 \times 9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1}{5 \times 4 \times 3 \times 2 \times 9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1} \\ &= \frac{14 \times 13 \times 12 \times 11}{4 \times 3} = \frac{14 \times 13 \times 4 \times 3 \times 11}{4 \times 3} \\ &= 14 \times 13 \times 11 = 2002 \end{aligned}$$

3. Quelle est la proportion de choix ne comportant que des filles?

Ces choix correspondent à choisir 5 élèves parmi les 18 filles. Le nombre de choix correspondant est donc

$$\begin{aligned} \binom{18}{5} &= \frac{18!}{5! \times 13!} = \frac{18 \times 17 \times \dots \times 1}{(5 \times 4 \times 3 \times 2) \times (13 \times 12 \times \dots \times 1)} \\ &= (18 \times 17 \times 16 \times 15 \times 14) \times \frac{1}{5 \times 4 \times 3 \times 2} = \frac{18 \times 17 \times 16 \times 5 \times 3 \times 14}{5 \times 4 \times 3 \times 2} \\ &= \frac{18 \times 17 \times 16 \times 14}{4 \times 3} = \frac{18 \times 17 \times 4 \times 14}{2} \\ &= \frac{18 \times 17 \times 4 \times 14}{2} = \frac{2 \times 9 \times 17 \times 4 \times 14}{2} \\ &= 9 \times 17 \times 4 \times 14 = 8568 \end{aligned}$$

Cela correspond donc à la probabilité  $\frac{8568}{201376} \simeq 0,04255$ .

4. Quelle est la probabilité, en choisissant ainsi un échantillon de 5 élèves au hasard, que cet échantillon comporte 4 filles et 1 garçon?

Les cas correspondant sont formés d'une part d'un groupe de 4 filles (pour lequel il y a  $\binom{18}{4}$  choix possibles), et pour chaque choix de ce groupe de filles, on peut y ajouter n'importe lequel des gar-

çons (pour lequel il y a  $\binom{14}{1}$  choix possibles). En conséquence, le nombre de cas qui correspondent est :

$$\begin{aligned} \binom{18}{4} \times \binom{14}{1} &= \frac{18!}{4! \times 14!} \times \frac{14!}{1! \times 13!} \\ &= \frac{18 \times 17 \times \dots \times 1}{(4 \times 3 \times 2) \times (14 \times 13 \times \dots \times 1)} \times \frac{14 \times 13 \times \dots \times 1}{13 \times 12 \times \dots \times 1} \\ &= (18 \times 17 \times 16 \times 15) \times \frac{1}{4 \times 3 \times 2} \times 14 \\ &= \frac{18 \times 17 \times 4 \times 4 \times 15}{4 \times 3 \times 2} \times 14 \\ &= \frac{18 \times 17 \times 4 \times 15}{3 \times 2} \times 14 \\ &= \frac{3 \times 6 \times 17 \times 4 \times 15}{3 \times 2} \times 14 \\ &= \frac{6 \times 17 \times 4 \times 15}{2} \times 14 \\ &= \frac{2 \times 3 \times 17 \times 4 \times 15}{2} \times 14 \\ &= 3 \times 17 \times 4 \times 15 \times 14 \\ &= 3060 \times 14 \\ &= 42840 \end{aligned}$$

D'où la probabilité  $\frac{42840}{201376} \simeq 0,2127$ .

5. Quel est le nombre de choix comportant au plus un garçon ?  
 On additionne, d'une part le nombre de choix comportant 0 garçons (calculé en question 3) et d'autre part le nombre de choix comportant exactement 1 garçon (calculé en question 4). On obtient donc  $8568 + 42840 = 51408$ .

### Chapitre 4 : Loi Normale

#### Exercice 35 : Loi normale centrée réduite

On désigne par  $Z$  la loi normale centrée réduite (de densité  $f(x) = \frac{1}{\sqrt{2\pi}} \exp(\frac{-x^2}{2})$ ). Pour  $z > 0$ , on pose

$$F(z) = \mathbb{P}[Z \leq z] = \int_{-\infty}^z f(x) dx.$$

1. Pour cette question, on pose  $z_1 = 1,57$ ;  $z_2 = 1,92$  et  $z_3 = 0,557$ .

- (a) Déterminer  $F(z_1)$ ,  $F(z_2)$  et  $F(z_3)$ . Chacune de ces probabilités peut être obtenue à partir de la table du formulaire, ou directement en utilisant les fonctions spécifiques de la calculatrice. Assurez-vous de savoir utiliser chacune de ces deux méthodes.
- (b) Calculer :  $\mathbb{P}[z_1 < Z \leq z_2]$ ,  $\mathbb{P}[-z_2 < Z < -z_1]$ ,  $\mathbb{P}[z_1 \leq Z < 5]$ ,  $\mathbb{P}[-z_2 \leq Z \leq z_1]$ ,  $\mathbb{P}[-z_1 < Z < z_1]$ ,  $\mathbb{P}[Z > z_2]$  et  $\mathbb{P}[Z < z_1]$ . On peut calculer ces probabilités à partir des valeurs obtenues dans la question a) ou en utilisant directement les fonctions spécifiques de la calculatrice. Assurez-vous de savoir utiliser chacune de ces deux méthodes.
2. (a) Déterminer des valeurs approchées de  $z_4$ ,  $z_5$ , et  $z_6$  tels que  $F(z_4) = 0,6331$ ;  $F(z_5) = 0,0401$  et  $F(z_6) = 0,8953$ .
- (b) Déterminer des valeurs approchées de  $z_7$ ,  $z_8$ , et  $z_9$  tels que  $\mathbb{P}[Z \leq z_7] = 0,96$ ;  $\mathbb{P}[Z > z_8] = 0,8577$ ;  $\mathbb{P}[Z < z_9] = 0,5636$ . On peut le faire à partir de la table du formulaire, ou bien obtenir une réponse directe de la calculatrice. Assurez-vous de savoir utiliser chacune de ces deux méthodes.

#### Exercice 36 : Images mentales

Dans une expérience sur les images mentales, on demande à un échantillon de 248 enfants d'apprendre une liste de 40 mots en leur montrant, pour chaque mot, un dessin de l'objet correspondant au mot. On désigne par  $X$  le nombre dont se souvient chaque enfant un jour après. On obtient :

Nb de mots $X$	[10; 15[	[15; 20[	[20; 25[	[25; 30[	[30; 35[	[35; 40[
Effectif	11	34	82	106	8	7

1. Calculer la moyenne  $m$  et l'écart-type  $s$  de la variable statistique  $X$ .
2. 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  $\mu \simeq 9,73$  et un écart-type  $\sigma \simeq 2,67$ .

Dans la suite, on effectue des calculs en supposant que  $X$  suit la loi normale  $\mathcal{N}(9,73; 2,67)$ .

1. Calculer la proportion théorique des scores supérieurs à 10,76.
2. Trouver la valeur de  $a$  telle que  $\mathbb{P}[X < a] \simeq 0,95$ .
3. Trouver les quartiles de la variable  $X$ .
4. Quel est le score minimal des 20 % de ceux qui ont les scores D.S. les plus élevés ?
5. 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 40%. 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  $n = 17$ . Calculer  $p_k = \mathbb{P}[S_{17} \leq k]$  pour  $k = 0; 1; 2; 3$ .
2. Pour  $n = 150$ , on souhaite calculer  $\mathbb{P}[54 \leq S_{150} \leq 70]$  :
  - a) Calculer  $\mathbb{P}[54 \leq S_{150} \leq 70]$  avec la calculette.
  - b) Argumenter que l'on peut approcher la loi de  $S_{150}$  par une loi normale, et calculer  $\mathbb{P}[54 \leq S_{150} \leq 70]$  de manière approchée. On vérifiera que la précision de l'approximation est bien meilleure si on fait la *correction de continuité*.
3. On considère désormais un échantillon de taille  $n = 15\,000$ .
  - (a) Par quelle loi normale peut on approcher la variable  $S_{15000}$  ? Calculer ainsi (une valeur approchée de) la probabilité  $\mathbb{P}[5\,891 \leq S_{15000} \leq 6\,103]$ .

- (b) Par quelle loi normale peut on approcher la proportion  $P_{15000} = \frac{S_{15000}}{15\,000}$  ? Calculer ainsi (une valeur approchée de) la probabilité  $\mathbb{P}[0,3927 \leq P_{15000} \leq 0,4069]$ .
- (c) Trouver une valeur de  $a$  telle que  $\mathbb{P}[S_{15000} < a] \simeq 0,03$ .

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

L'indice de masse corporelle mesure la corpulence d'un individu. Dans cet exercice on le notera  $X$ . Il s'agit d'un élément de diagnostic de dénutrition ou d'obésité, dont l'interprétation est la suivante :

$X < 16,5$ dénutrition ou famine	$25 \leq X < 30$ surpoids
$16,5 \leq X < 18,5$ maigreur	$30 \leq X < 35$ obésité modérée
$18,5 \leq X < 25$ corpulence normale	$35 \leq X < 40$ obésité sévère
	$X \geq 40$ obésité morbide

Une enquête épidémiologique conduite en 2012 par l'INSERM, KANTAR HEALTH et ROCHE, donne la répartition suivante de l'IMC au sein d'un échantillon de 25 714 personnes, supposé représentatif de la population française :

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

1. Déterminer la probabilité (« *fréquence théorique* ») qu'aurait chaque classe si  $X$  suivait la loi normale  $\mathcal{N}(25; 5)$ .  
Si  $X$  suit la loi  $\mathcal{N}(25; 5)$ , on obtient les probabilités suivantes :
  - $\mathbb{P}[X < 18,5] = \mathbb{P}\left[\frac{X-25}{5} < \frac{18,5-25}{5}\right] = \mathbb{P}[Z < -1,3] = F(-1,3) = 0,0968$
  - $\mathbb{P}[18,5 \leq X < 25] = \mathbb{P}\left[\frac{18,5-25}{5} \leq \frac{X-25}{5} < \frac{25-25}{5}\right] = \mathbb{P}[-1,3 \leq Z < 0] = F(0) - F(-1,3) = 0,4032$
  - $\mathbb{P}[25 \leq X < 30] = \mathbb{P}\left[\frac{25-25}{5} \leq \frac{X-25}{5} < \frac{30-25}{5}\right] = \mathbb{P}[0 \leq Z < 1] = F(1) - F(0) = 0,3413$

- $\mathbb{P}[30 \leq X < 35] = \mathbb{P}\left[\frac{30-25}{5} \leq \frac{X-25}{5} < \frac{35-25}{5}\right]$   
 $= \mathbb{P}[1 \leq Z < 2] = F(2) - F(1)$   
 $= 0,1359$
- $\mathbb{P}[35 \leq X < 40] = \mathbb{P}\left[\frac{35-25}{5} \leq \frac{X-25}{5} < \frac{40-25}{5}\right]$   
 $= \mathbb{P}[2 \leq Z < 3] = F(3) - F(2)$   
 $= 0,0215$
- $\mathbb{P}[X \geq 40] = \mathbb{P}\left[\frac{X-25}{5} \geq \frac{40-25}{5}\right]$   
 $= \mathbb{P}[Z \geq 3] = 1 - F(3)$   
 $\simeq 0,0013$

D'où les fréquences théoriques suivantes :

IMC	$[-\infty; 18,5[$	$[18,5; 25[$	$[25; 30[$	$[30; 35[$	$[35; 40[$	$[40; +\infty[$
fréquence théorique	0,0968	0,4032	0,3413	0,1359	0,0215	0,0013

2. En particulier, si  $X$  suit la loi normale  $\mathcal{N}(25; 5)$ , quelle est la proportion (*théorique*) d'obésité (c'est à dire  $\mathbb{P}[X \geq 30]$ , en regroupant les obésités modérées, sévères et morbides) ?  
 En particulier la proportion théorique d'obésité serait la somme des fréquences théoriques des trois dernières classes, c'est à dire  $\mathbb{P}[X \geq 30] \simeq 0,1359 + 0,0215 + 0,0013 \simeq 0,159$ .
3. Au contraire, quelle est dans l'échantillon de cette étude épidémiologique, la proportion d'obésité (c'est à dire  $\mathbb{P}_r[X \geq 30]$ ) ?  
 La proportion d'obésité est la somme des fréquences des trois dernières classes, c'est à dire  $10,7 + 3,1 + 1,2 \simeq 15\%$ .
4. On cherche désormais à comparer les résultats des questions 2 et 3 : on souhaite déterminer si la proportion  $\mathbb{P}_r[X \geq 30]$  est compatible avec la probabilité  $\mathbb{P}[X \geq 30]$  obtenue sous l'hypothèse d'une loi normale.
  - (a) Si on suppose qu'une personne choisie 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 ?  
 On choisit un échantillon aléatoire de taille 25 714 français

parmi une population constituée de tous les français adultes (environ 50 000 000). La taille de cette population est plus de 10 fois supérieure à celle de l'échantillon ( $50\,000\,000 > 10 \times 25\,714$ ) donc le fait d'être "avec" ou "sans remise" n'a aucun impact. On peut donc utiliser la loi binomiale  $\mathcal{B}(25714; 0,159)$ .

- (b) Cette loi peut-elle être approximée par une loi normale, et si oui laquelle ?  
 $np(1 - p) = 25\,714 \times 0,159(1 - 0,159) \simeq 3\,438,45 > 1\,000$ , donc on peut approcher  $\mathcal{B}(25714; 0,159) \approx \mathcal{N}(25\,714 \times 0,159; \sqrt{25\,714 \times 0,159(1 - 0,159)}) \simeq \mathcal{N}(4088,526; 58,638)$   
 et il n'y a pas besoin de faire de correction de continuité.
- (c) La loi de la proportion de français en situation d'obésité peut-elle aussi être approximée par une loi normale, et si oui laquelle ?  
 $np(1 - p) = 25\,714 \times 0,159(1 - 0,159) \simeq 3\,438,45 > 1\,000$ , donc on peut approcher la proportion par  $\mathcal{N}(0,159; \sqrt{\frac{0,159(1-0,159)}{25\,714}}) \simeq \mathcal{N}(0,159; 0,00228)$ , et il n'y a pas besoin de faire de correction de continuité.
- (d) En utilisant cette approximation par une loi normale, calculer la probabilité qu'un échantillon aléatoire de 25 714 personnes contienne au maximum 15% de personnes en situation d'obésité.  
 $\mathbb{P}[P \leq 0,15] \simeq 3,951 \times 10^{-05}$
- (e) Comparer avec l'échantillon de l'étude épidémiologique mentionnée en début d'exercice.  
 Conclure qu'en 2012, l'IMC des français ne suivait pas la loi de probabilité  $\mathcal{N}(25; 5)$ .  
 Donc si l'IMC des français avait vraiment suivi la loi  $\mathcal{N}(25; 5)$ , alors il n'y aurait eu aucune chance qu'un échantillon aléatoire de 25714 français ait 15% de personnes en



situation d'obésité ou moins. Or l'échantillon aléatoire de cette étude épidémiologique a justement 15% de personnes en situation d'obésité (ce qui, comme on vient de le dire, n'aurait pu arriver si l'IMC des français avait vraiment suivi la loi  $\mathcal{N}(25; 5)$ ). Donc l'IMC des français ne suivait pas la loi  $\mathcal{N}(25; 5)$ .

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

1. Pour une loi normale centrée réduite, quelle est la probabilité qu'une observation soit :

(a) supérieure ou égale à  $-0,5$  ?

$$\mathbb{P}[Z \geq -0,5] = 1 - F(-0,5) = 0,6915$$

(b) comprise entre  $-1$  et  $1$  ?

$$\mathbb{P}[-1 \leq Z \leq 1] = F(1) - F(-1) = 0,6826$$

(c) comprise entre  $-2$  et  $2$  ?

$$\mathbb{P}[-2 \leq Z \leq 2] = F(2) - F(-2) = 0,9544$$

(d) comprise entre  $-3$  et  $3$  ?

$$\mathbb{P}[-3 \leq Z \leq 3] = F(3) - F(-3) = 0,9974$$

2. Pour une loi normale de moyenne  $\mu = 123$  et d'écart-type  $\sigma = 11$ , (c'est-à-dire  $X \sim \mathcal{N}(123; 11)$ ), quelle est la probabilité qu'une observation soit :

(a) supérieure ou égale à  $\mu - 0,5\sigma$  ?

On note tout d'abord que  $\mu - 0,5\sigma = 123 - 0,5 \times 11 = 117,5$ .

$$\begin{aligned} \mathbb{P}[X \geq 117,5] &= \mathbb{P}\left[\frac{X-123}{11} \geq \frac{117,5-123}{11}\right] \\ &= \mathbb{P}[Z \geq -0,5] = 1 - F(-0,5) \\ &= 0,6915 \end{aligned}$$

(b) comprise entre  $\mu - \sigma$  et  $\mu + \sigma$  ?

On note tout d'abord que  $\mu - \sigma = 123 - 11 = 112$  et  $\mu + \sigma = 123 + 11 = 134$ .

$$\begin{aligned} \mathbb{P}[112 \leq X \leq 134] &= \mathbb{P}\left[\frac{112-123}{11} \leq \frac{X-123}{11} \leq \frac{134-123}{11}\right] \\ &= \mathbb{P}[-1 \leq Z \leq 1] = F(1) - F(-1) \\ &= 0,6826 \end{aligned}$$

(c) comprise entre  $\mu - 2\sigma$  et  $\mu + 2\sigma$  ?

On note tout d'abord que  $\mu - 2\sigma = 123 - 2 \times 11 = 101$  et  $\mu + 2\sigma = 123 + 2 \times 11 = 145$ .

$$\begin{aligned} \mathbb{P}[101 \leq X \leq 145] &= \mathbb{P}\left[\frac{101-123}{11} \leq \frac{X-123}{11} \leq \frac{145-123}{11}\right] \\ &= \mathbb{P}[-2 \leq Z \leq 2] = F(2) - F(-2) \\ &= 0,9544 \end{aligned}$$

(d) comprise entre  $\mu - 3\sigma$  et  $\mu + 3\sigma$  ?

On note tout d'abord que  $\mu - 3\sigma = 123 - 3 \times 11 = 90$  et  $\mu + 3\sigma = 123 + 3 \times 11 = 156$ .

$$\begin{aligned} \mathbb{P}[90 \leq X \leq 156] &= \mathbb{P}\left[\frac{90-123}{11} \leq \frac{X-123}{11} \leq \frac{156-123}{11}\right] \\ &= \mathbb{P}[-3 \leq Z \leq 3] = F(3) - F(-3) \\ &= 0,9974 \end{aligned}$$

3. Reprendre les questions précédentes avec la loi normale  $\mathcal{N}(15,34; 2,75)$ . Que remarque-t-on ?

(a) On note tout d'abord que  $\mu - 0,5\sigma = 15,34 - 0,5 \times 2,75 = 13,965$ .

$$\begin{aligned} \mathbb{P}[X \geq 13,965] &= \mathbb{P}\left[\frac{X-15,34}{2,75} \geq \frac{13,965-15,34}{2,75}\right] \\ &= \mathbb{P}[Z \geq -0,5] = 1 - F(-0,5) \\ &= 0,6915 \end{aligned}$$

(b) On note tout d'abord que  $\mu - \sigma = 15,34 - 2,75 = 12,59$  et  $\mu + \sigma = 15,34 + 2,75 = 18,09$ .

$$\begin{aligned} \mathbb{P}[12,59 \leq X \leq 18,09] &= \mathbb{P}\left[\frac{12,59-15,34}{2,75} \leq \frac{X-15,34}{2,75} \leq \frac{18,09-15,34}{2,75}\right] \\ &= \mathbb{P}[-1 \leq Z \leq 1] = F(1) - F(-1) \\ &= 0,6826 \end{aligned}$$

(c) On note tout d'abord que  $\mu - 2\sigma = 15,34 - 2 \times 2,75 = 9,84$

et  $\mu + 2\sigma = 15,34 + 2 \times 2,75 = 20,84$ .

$$\begin{aligned}\mathbb{P}[9,84 \leq X \leq 20,84] &= \mathbb{P}\left[\frac{9,84-15,34}{2,75} \leq \frac{X-15,34}{2,75} \leq \frac{20,84-15,34}{2,75}\right] \\ &= \mathbb{P}[-2 \leq Z \leq 2] = F(2) - F(-2) \\ &= 0,9544\end{aligned}$$

(d) On note tout d'abord que  $\mu - 3\sigma = 15,34 - 3 \times 2,75 = 7,09$  et  $\mu + 3\sigma = 15,34 + 3 \times 2,75 = 23,59$ .

$$\begin{aligned}\mathbb{P}[7,09 \leq X \leq 23,59] &= \mathbb{P}\left[\frac{7,09-15,34}{2,75} \leq \frac{X-15,34}{2,75} \leq \frac{23,59-15,34}{2,75}\right] \\ &= \mathbb{P}[-3 \leq Z \leq 3] = F(3) - F(-3) \\ &= 0,9974\end{aligned}$$

On remarque que ces résultats sont indépendants de  $\sigma$  et  $\mu$ .

#### Exercice 41 :

On évalue les niveaux de dépression au moyen du test Inventaire Multiphasique de la Personnalité du Minnesota (MMPI). Au vu du nombre important de questions du MMPI, on considère que quand on fait passer ce test à un individu choisi au hasard, son score suit une loi normale. De plus le MMPI est normalisé pour avoir une moyenne  $\mu = 50$  et un écart-type  $\sigma = 10$ .

Si l'on considère qu'une note supérieure à 65 requière un traitement contre la dépression, combien de personnes nécessitant un traitement s'attend-on à trouver sur un ensemble de 20000 personnes ?

On note  $X$  la note au MMPI, et on a  $X \sim \mathcal{N}(50; 10)$ .

$$\begin{aligned}\mathbb{P}[X \geq 65] &= \mathbb{P}\left[\frac{X-50}{10} \geq \frac{65-50}{10}\right] \\ &= \mathbb{P}[Z \geq 1,5] = 1 - F(1,5) \\ &= 0,0668\end{aligned}$$

Donc parmi un échantillon de 20 000 personnes, on s'attend à trouver environ  $20\,000 \times 0,0668 = 1\,336$  personnes nécessitant un traitement.

#### Exercice 42 :

La proportion des enfants qui réussissent parfaitement une épreuve graphique d'organisation perceptive est de 65%. On choisit au hasard (avec remise) un échantillon de  $n$  individus et on note  $X$  la variable aléatoire égale au nombre d'enfants dans l'échantillon qui réussissent l'épreuve.

1. On prend  $n = 21$ . Préciser la loi de probabilité de  $X$  et calculer la probabilité :  $\mathbb{P}[13 \leq X \leq 16]$ .

$$X \sim \mathcal{B}(21; 0,65).$$

$$\mathbb{P}[13 \leq S \leq 16] = \mathbb{P}[S \leq 16] - \mathbb{P}[S \leq 12] \simeq 0,9076 - 0,2941 = 0,6135$$

2. On prend  $n = 210$ .

(a) On peut approcher la loi de  $X$  par une loi normale. Justifier, préciser laquelle, et indiquer s'il est nécessaire de faire une correction de continuité.

$$X \sim \mathcal{B}(210; 0,65).$$

$$\begin{aligned}np(1-p) &= 210 \times 0,65(1-0,65) = \\ 47,775 &> 10, \text{ donc on peut approcher} \\ \mathcal{B}(210; 0,65) &\approx \mathcal{N}(210 \times 0,65; \sqrt{210 \times 0,65(1-0,65)}) \\ &\simeq \mathcal{N}(136,5; 6,912)\end{aligned}$$

mais il est nécessaire de faire de correction de continuité.

(b) Calculer les probabilités  $\mathbb{P}[X \geq 129]$  et  $\mathbb{P}[130 \leq X \leq 160]$ .

$$\mathbb{P}[X \geq 129] = \mathbb{P}[X \geq 128,5] \simeq 0,876$$

$$\mathbb{P}[130 \leq X \leq 160] = \mathbb{P}[129,5 \leq X \leq 160,5] \simeq 0,844$$

**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 :

1. 100 fois.

Si on note  $X$  le nombre de pile, on a  $X \sim \mathcal{B}(100; 0,5)$ .

$np(1-p) = 100 \times 0,5 (1-0,5) = 25 > 10$ , donc on peut approcher  $\mathcal{B}(100; 0,5) \approx \mathcal{N}(100 \times 0,5; \sqrt{100 \times 0,5 (1-0,5)})$ ,  
 $= \mathcal{N}(50; 5)$

mais il est nécessaire de faire de correction de continuité.

Avoir une proportion de pile entre 0,48 et 0,52 signifierait avoir  $48 \leq X \leq 52$ .

La probabilité demandée vaut donc (avec la correction de continuité) :

$$\mathbb{P}[48 \leq X \leq 52] = \mathbb{P}[47,5 \leq X \leq 52,5] \simeq 0,383$$

2. 1 000 fois.

on a  $X \sim \mathcal{B}(1000; 0,5)$ .

$np(1-p) = 1000 \times 0,5 (1-0,5) = 250 > 10$ , donc on peut approcher  $\mathcal{B}(1000; 0,5) \approx \mathcal{N}(1000 \times 0,5; \sqrt{1000 \times 0,5 (1-0,5)})$ ,  
 $\simeq \mathcal{N}(500; 15,811)$

mais il est nécessaire de faire de correction de continuité.

Avoir une proportion de pile entre 0,48 et 0,52 signifierait avoir  $480 \leq X \leq 520$ .

La probabilité demandée vaut donc (avec la correction de continuité) :

$$\mathbb{P}[480 \leq X \leq 520] = \mathbb{P}[479,5 \leq X \leq 520,5] \simeq 0,805$$

3. 2 000 fois.

on a  $X \sim \mathcal{B}(2000; 0,5)$ .

$np(1-p) = 2000 \times 0,5 (1-0,5) = 500 > 10$ , donc on peut approcher  $\mathcal{B}(2000; 0,5) \approx \mathcal{N}(2000 \times 0,5; \sqrt{2000 \times 0,5 (1-0,5)})$ ,  
 $\simeq \mathcal{N}(1000; 22,361)$

mais il est nécessaire de faire de correction de continuité.

Avoir une proportion de pile entre 0,48 et 0,52 signifierait avoir  $960 \leq X \leq 1040$ .

La probabilité demandée vaut donc (avec la correction de continuité) :

$$\mathbb{P}[960 \leq X \leq 1040] = \mathbb{P}[959,5 \leq X \leq 1040,5] \simeq 0,9299$$

4. 5 000 fois.

On considère désormais la proportion  $p = \frac{X}{n}$ .

$np(1-p) = 5000 \times 0,5 (1-0,5) = 1250 > 1000$ , donc on peut approcher la proportion par  $\mathcal{N}(0,5; \sqrt{\frac{0,5(1-0,5)}{5000}}) \simeq \mathcal{N}(0,5; 0,007071)$ , et il n'y a pas besoin de faire de correction de continuité.

La probabilité demandée vaut donc :

$$\mathbb{P}[0,48 \leq P \leq 0,52] \simeq 0,99532$$

#### Exercice 44 :

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 80 et d'écart-type 10.

1. Quelle est la probabilité pour qu'un neurone émette plus de 95 PA en 1 mn s'il n'est pas activé ?

On note  $X$  le nombre de PA émis en une minute. Si le neurone n'est pas activé on a  $X \sim \mathcal{N}(80; 10)$ .

$$\mathbb{P}[X > 95] \simeq 0,06681 \text{ (d'après la calculatrice).}$$

2. Quelle est la probabilité pour qu'un neurone émette moins de 60 PA en 1 mn s'il n'est pas activé ?

$$\mathbb{P}[X < 60] \simeq 0,02275 \text{ (d'après la calculatrice).}$$

3. Quelle est la probabilité pour qu'un neurone émette entre 70 et 85 PA en 1 mn s'il n'est pas activé ?

$$\mathbb{P}[70 \leq X < 85] \simeq 0,5328 \text{ (d'après la calculatrice).}$$

4. Un neurone émet plus de 120 PA en une minute. Selon vous, est-il plus vraisemblable que ce neurone était activé ou au repos lors de l'enregistrement ? Justifiez votre réponse.

$$\mathbb{P}[X \geq 120] \simeq 3,167 \times 10^{-05} \text{ (d'après la calculatrice).}$$

Donc si le neurone avait été au repos il aurait été très improbable d'avoir au moins 120 PA en une minute. Donc comme on observe un neurone émettant 120 PA par minute, il est plus vraisemblable qu'il soit actif.

## Chapitre 5 : Estimation

### Exercice 45 : Pile ou face

Vous jouez à « pile » ou « face » avec une pièce de monnaie. Vous la lancez 800 fois et obtenez 443 fois pile.

1. Estimez, avec la confiance 90%, la probabilité qu'a cette pièce de tomber sur pile quand on la lance.
2. Conclusion : pouvez-vous affirmer (avec la confiance 90%) 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 l'administre à 800 malades, et on constate que parmi cet échantillon de 800 patient traités, il y en a 585 qui guérissent en 5 jours.

1. Donner une estimation de la vraie proportion  $p$  de guérisons (pour l'ensemble des patients que l'on traiterait avec ce médicament) avec une confiance de 95%.
2. Que peut-on conclure par rapport à l'affirmation du fabricant ?
3. On veut réduire la marge de l'estimation à 1% tout en atteignant 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 992 personnes interrogées, 524 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 758 personnes interrogées, il y en a alors 434 qui sont satisfaites.

1. On désigne par  $p_1$  et  $p_2$  les proportions de satisfaits pour l'ensemble des clients avant et après l'amélioration apportée. Donner une estimation de  $p_1$  et  $p_2$  par intervalle de confiance en prenant comme confiance  $c = 0,99$ . Peut-on dire que la proportion de satisfaits pour l'ensemble de la clientèle a augmenté ?
2. Lors du deuxième sondage, quelle aurait du 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 pense 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 10,3 et l'écart type est 1,6.

1. On soumet dans un premier temps un échantillon de 70 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 ; 7[	[7 ; 8[	[8 ; 9[	[9 ; 10[	[10 ; 11[	[11 ; 12[	[12 ; 13[
Effectif	3	6	10	17	18	13	3

- 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 27 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  $m_e = 6,78$  et  $s_e = 4,1$ .
- 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**

Un expérimentateur qui étudie la prise de décisions a demandé à 30 enfants de résoudre le plus grand nombre possible de problèmes en 30 minutes. Il a expliqué à un groupe de 14 enfants qu'il 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	18	30	23	24	22	24	24	25	23	29	31	31	22	33	
Occupation	5	8	8	4	8	7	5	10	9	5	6	8	7	12	10

On suppose que les résultats varient normalement.

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 hommes et les femmes de plus de 50 ans. On suppose qu'ils suivent des lois normales.

Dans un échantillon de 16 femmes de plus de 50 ans on a mesuré un taux moyen de 173,9 mg/dL avec un écart type de 54,7 mg/dL.

Dans un échantillon de 19 hommes de plus de 50 ans on a mesuré un taux moyen de 194 mg/dL avec un écart type de 38,3 mg/dL.

1. Donner des estimations des écarts types des deux populations. *On demande de déterminer un intervalle de confiance, avec la confiance  $c = 96\%$ .*

**femmes** On lit dans la table du  $\chi^2$  à 15 degrés de liberté la valeur  $x_1 = 5,985$  correspondant à  $p = 0,02$  et la valeur  $x_2 = 28,26$  correspondant à  $q = 0,02$

On obtient alors  $s_e\sqrt{\frac{n}{x_2}} \simeq 54,7\sqrt{\frac{16}{28,26}} \simeq 41,2$  et  $s_e\sqrt{\frac{n}{x_1}} \simeq 54,7\sqrt{\frac{16}{5,985}} \simeq 89,4$

On estime donc que  $\sigma$  est dans l'intervalle  $[41,2; 89,4]$  avec la confiance  $c = 0,96$ .

**hommes** On lit dans la table du  $\chi^2$  à 18 degrés de liberté la valeur  $x_1 = 7,906$  correspondant à  $p = 0,02$  et la valeur  $x_2 = 32,35$  correspondant à  $q = 0,02$

On obtient alors  $s_e\sqrt{\frac{n}{x_2}} \simeq 38,3\sqrt{\frac{19}{32,35}} \simeq 29,4$  et  $s_e\sqrt{\frac{n}{x_1}} \simeq 38,3\sqrt{\frac{19}{7,906}} \simeq 59,4$

On estime donc que  $\sigma$  est dans l'intervalle  $[29,4; 59,4]$  avec la confiance  $c = 0,96$ .

2. En utilisant les estimations par intervalle de confiance à 96%, peut-on conclure que les femmes de plus de 50 ans ont en moyenne moins de cholestérol que les hommes de plus de 50 ans ?

On commence par estimer les moyennes :

**femmes** Comme  $n = 16 \leq 30$ , on cherche  $t_\alpha$  à partir de la table inverse de Student avec  $p = \frac{\alpha}{2} = 0,02$  et  $n - 1 = 15$  degrés de liberté (ddl)

On lit  $t_\alpha \simeq 2,2485$  d'où  $a_\alpha = t_\alpha \frac{se}{\sqrt{n-1}} \simeq 31,757$ .

On estime donc que  $\mu$  est dans l'intervalle  $[173,9 - 31,757; 173,9 + 31,757] \simeq [142,14; 205,66]$  avec la confiance  $c = 0,96$

**hommes** Comme  $n = 19 \leq 30$ , on cherche  $t_\alpha$  à partir de la table inverse de Student avec  $p = \frac{\alpha}{2} = 0,02$  et  $n - 1 = 18$  degrés de liberté (ddl)

On lit  $t_\alpha \simeq 2,2137$  d'où  $a_\alpha = t_\alpha \frac{se}{\sqrt{n-1}} \simeq 19,984$ .

On estime donc que  $\mu$  est dans l'intervalle  $[194 - 19,984; 194 + 19,984] \simeq [174,02; 213,98]$  avec la confiance  $c = 0,96$

Les intervalles se chevauchent donc on ne peut pas conclure que les femmes de plus de 50 ans aient en moyenne plus de cholestérol que les hommes de plus de 50 ans.

### Exercice 51 : Épreuve de dictée

Le tableau suivant représente les résultats obtenus par un échantillon de 14 enfants de CE2 dans deux épreuves de dictée préparée : une liste de 10 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	M	N
Épreuve 1	8,3	8,2	9,3	9,9	9,5	9,6	8	9,1	8,8	8,3	8,8	7,6	8,8	9,4
Épreuve 2	6,1	6,2	7,1	7,2	6,5	7	6,2	6,3	6,1	6,4	6,3	6,2	6,6	6,8

1. Peut-on affirmer, avec un risque d'erreur de 10%, qu'une des deux épreuves donne, pour des enfants de CE2, une note moyenne plus élevée que l'autre épreuve ?

- (a) On commence par calculer les moyennes et écart types expérimentaux. En notant  $X$  les résultats de l'épreuve 1 et  $Y$  les résultats de l'épreuve 2, on obtient

$$m(X) = 8,83 \qquad s(X) = 0,653.$$

$$m(Y) = 6,5 \qquad s(Y) = 0,366.$$

- (b) On estime ensuite les moyennes avec la confiance 95% :

**Épreuve 1** Comme  $n = 14 \leq 30$ , on cherche  $t_\alpha$  à partir de la table inverse de Student avec  $p = \frac{\alpha}{2} = 0,05$  et  $n - 1 = 13$  degrés de liberté (ddl)

On lit  $t_\alpha \simeq 1,7709$  d'où  $a_\alpha = t_\alpha \frac{se}{\sqrt{n-1}} \simeq 0,32073$ .

On estime donc que  $\mu(X)$  est dans l'intervalle  $[8,83 - 0,32073; 8,83 + 0,32073] \simeq [8,51; 9,15]$  avec la confiance  $c = 0,9$

**Épreuve 2** Comme  $n = 14 \leq 30$ , on cherche  $t_\alpha$  à partir de la table inverse de Student avec  $p = \frac{\alpha}{2} = 0,05$  et  $n - 1 = 13$  degrés de liberté (ddl)

On lit  $t_\alpha \simeq 1,7709$  d'où  $a_\alpha = t_\alpha \frac{se}{\sqrt{n-1}} \simeq 0,17976$ .

On estime donc que  $\mu(Y)$  est dans l'intervalle  $[6,5 - 0,17976; 6,5 + 0,17976] \simeq [6,32; 6,68]$  avec la confiance  $c = 0,9$

- (c) : Ces intervalles indiquent que  $\mu(X) > \mu(Y)$  (en effet  $\mu(X) > 8,5$  et  $\mu(Y) < 7$ ). Donc on peut dire (avec la confiance 95%) que les résultats moyens diffèrent significativement d'une épreuve à l'autre, et plus précisément que les résultats sont meilleurs pour la liste de mots que pour le texte.

**Remarque :** Comme l'échantillon est apparié, on aurait aussi pu calculer les différences entre les deux notes pour chaque individu, puis estimer la moyenne de la différence (afin de déterminer si elle est positive). Dans le cas présent la conclusion aurait été inchangée.

2. Donner des estimations des écarts types.

**Épreuve 1** On lit dans la table du  $\chi^2$  à 13 degrés de liberté la valeur  $x_1 = 5,009$  correspondant à  $p = 0,025$  et la valeur  $x_2 = 24,74$  correspondant à  $q = 0,025$

On obtient alors  $s_e \sqrt{\frac{n}{x_2}} \simeq 0,653 \sqrt{\frac{14}{24,74}} \simeq 0,491$  et  $s_e \sqrt{\frac{n}{x_1}} \simeq 0,653 \sqrt{\frac{14}{5,009}} \simeq 1,09$

On estime donc que  $\sigma(X)$  est dans l'intervalle  $[0,491; 1,09]$  avec la confiance  $c = 0,95$ .

**Épreuve 2** On lit dans la table du  $\chi^2$  à 13 degrés de liberté la valeur  $x_1 = 5,009$  correspondant à  $p = 0,025$  et la valeur  $x_2 = 24,74$  correspondant à  $q = 0,025$

On obtient alors  $s_e \sqrt{\frac{n}{x_2}} \simeq 0,366 \sqrt{\frac{14}{24,74}} \simeq 0,275$  et  $s_e \sqrt{\frac{n}{x_1}} \simeq 0,366 \sqrt{\frac{14}{5,009}} \simeq 0,612$

On estime donc que  $\sigma(Y)$  est dans l'intervalle  $[0,275; 0,612]$  avec la confiance  $c = 0,95$ .

### Exercice 52 :

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

Un.e psychologue étudie le taux de chômage de ces patients atteint 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 7,7% et 7,9% de chômeurs ?

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

- L'échantillon de 30000 personnes est choisi parmi une population de dizaines de millions de français adultes, donc  $\frac{n}{N} < \frac{1}{10}$  et on peut utiliser la loi binomiale : si on note  $X$  le nombre chômeurs d'un tel échantillon aléatoire, alors  $X \sim \mathcal{B}(30000; 0,078)$ .

- $np(1-p) = 30000 \times 0,078(1-0,078) = 2157,48 > 1000$ , donc on peut approcher la proportion par  $\mathcal{N}(0,078; \sqrt{\frac{0,078(1-0,078)}{30000}}) \simeq \mathcal{N}(0,078; 0,001548)$ , et il n'y a pas besoin de faire de correction de continuité.

- On désigne par  $P$  cette proportion et on obtient :  $\mathbb{P}[0,077 \leq P \leq 0,079] \simeq 0,482$

2. Sur un échantillon de 30000 patients atteints de boulimie, le psychologue a constaté que 7,51% d'entre eux étaient au chômage.

- (a) Avec la confiance  $c = 90\%$ , déterminer 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 du reste de la population ?

Remarques : ces 7,51% de l'échantillon correspondent à  $30000 \times 0,0751 = 2253$  personnes.

On a  $np(1-p) = 30000 \times \frac{2253}{30000} \left(1 - \frac{2253}{30000}\right) \simeq 2083,8 > 10$ , donc on peut utiliser la procédure du formulaire pour estimer la proportion  $p$ .

On a  $F(1,645) \simeq 0,95$  d'où  $z_\alpha \simeq 1,645$

d'où  $a_\alpha = z_\alpha \sqrt{\frac{pe(1-pe)}{n}} \simeq 0,0025$ .

On estime donc que  $p$  est dans l'intervalle  $[0,0726; 0,0776]$  avec la confiance  $c = 0,9$ .

Toutes les valeurs de cet intervalle sont inférieures à 7,8%, donc on peut affirmer, avec la confiance 90%, que le taux de chômage des personnes boulimiques est plus faible que parmi le reste de la population.

- (b) Avec la confiance  $c = 98\%$ , estimer 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 du reste de la population ?

On a  $np(1-p) = 30\,000 \times \frac{2\,253}{30\,000} \left(1 - \frac{2\,253}{30\,000}\right) \simeq 2\,083,8 > 10$ , donc on peut utiliser la procédure du formulaire pour estimer la proportion  $p$ .

On a  $F(2,326) \simeq 0,99$  d'où  $z_\alpha \simeq 2,326$

d'où  $a_\alpha = z_\alpha \sqrt{\frac{pe(1-pe)}{n}} \simeq 0,00354$ .

On estime donc que  $p$  est dans l'intervalle  $[0,07156; 0,07864]$  avec la confiance  $c = 0,98$ .

Cette fois-ci, la valeur  $7,8\%$  se situe à l'intérieur de l'intervalle. Donc on est incapable de déterminer, avec la confiance  $98\%$ , si les personnes boulimiques sont plus ou moins touchées que les autres personnes par le chômage.

**Remarque :** La seule différence entre les questions a et b est le niveau de confiance. Pour cet exemple, on ne peut pas conclure avec une confiance  $98\%$ , alors qu'on peut conclure avec une confiance  $90\%$ .

- (c) Pour pouvoir estimer à  $0,01\%$  près le taux de chômage des personnes boulimiques avec une confiance  $98\%$ , quelle taille d'échantillon faudrait-il considérer ?

Pour avoir une précision  $0,0001$  avec la confiance  $0,98$ , il faut  $n > z_\alpha^2 \frac{pe(1-pe)}{h^2} \simeq 2,326^2 \frac{0,0751(1-0,0751)}{0,0001^2} \simeq 37579771$

**Exercice 53 :** Un chercheur, Mr Cousin, 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  $3\,500$  femmes, parmi lesquelles  $11$  sont daltoniennes, et d'autre part un échantillon de  $3\,500$  hommes, parmi lesquels  $274$  sont daltoniens.

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

- (a) On détermine tout d'abord des intervalles de confiance pour la proportion de daltoniens parmi les femmes et les hommes :

• **parmi les femmes :**

On a  $np(1-p) = 3\,500 \times \frac{11}{3\,500} \left(1 - \frac{11}{3\,500}\right) \simeq 10,965 > 10$ , donc on peut utiliser la procédure du formulaire pour estimer la proportion  $p$ .

On a  $F(1,96) \simeq 0,975$  d'où  $z_\alpha \simeq 1,96$

d'où  $a_\alpha = z_\alpha \sqrt{\frac{pe(1-pe)}{n}} \simeq 0,001854$ .

On estime donc que  $p$  est dans l'intervalle  $[0,001289; 0,004997]$  avec la confiance  $c = 0,95$ .

• **parmi les hommes :**

On a  $np(1-p) = 3\,500 \times \frac{274}{3\,500} \left(1 - \frac{274}{3\,500}\right) \simeq 252,55 > 10$ , donc on peut utiliser la procédure du formulaire pour estimer la proportion  $p$ .

On a  $F(1,96) \simeq 0,975$  d'où  $z_\alpha \simeq 1,96$

d'où  $a_\alpha = z_\alpha \sqrt{\frac{pe(1-pe)}{n}} \simeq 0,0089$ .

On estime donc que  $p$  est dans l'intervalle  $[0,06939; 0,08719]$  avec la confiance  $c = 0,95$ .

- (b) On remarque que ces intervalles ne se chevauchent pas et indiquent une plus grande proportion chez les hommes. On peut donc affirmer, avec la confiance  $c = 95\%$ , que la prévalence du daltonisme est plus importante chez l'homme que chez la femme.

2. Le chercheur 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. Il teste cette hypothèse sur des enfants



de 7 ans, auxquels il fait passer un test standardisé de perception des couleurs. Sur un échantillon de 16 filles, il obtient les données suivantes :

Score	16	13	14	13	9	17	17	20	9	10	16	10	15	15	11	17
-------	----	----	----	----	---	----	----	----	---	----	----	----	----	----	----	----

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

- On détermine tout d’abord la moyenne et l’écart-type au sein de cet échantillon :

$$\text{moyenne : } m(X) = \frac{\sum x_i}{n} = \frac{16+13+14+\dots+17}{16} = \frac{222}{16} \simeq 13,88$$

$$m(X^2) = \frac{\sum x_i^2}{n} = \frac{16^2+13^2+14^2+\dots+17^2}{16} = \frac{3246}{16}$$

$$Var(X) = m(X^2) - m(X)^2 = \frac{3246}{16} - \left(\frac{222}{16}\right)^2 \simeq 10,36$$

$$\text{Écart-type : } s(X) = \sqrt{Var(X)} \simeq 3,22$$

- Comme  $n = 16 \leq 30$ , on cherche  $t_\alpha$  à partir de la table inverse de Student avec  $p = \frac{\alpha}{2} = 0,005$  et  $n - 1 = 15$  degrés de liberté (ddl)

$$\text{On lit } t_\alpha \simeq 2,9467 \text{ d'où } a_\alpha = t_\alpha \frac{s_e}{\sqrt{n-1}} \simeq 2,4499.$$

On estime donc que  $\mu$  est dans l’intervalle  $[13,88 - 2,4499; 13,88 + 2,4499] \simeq [11,43; 16,33]$  avec la confiance  $c = 0,99$

- Sur un échantillon de 39 garçons auxquels le chercheur fait passer le test, le score moyen est de 11,13 avec un écart-type de 4,92.

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

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

- Comme  $n = 39 > 30$ , on cherche  $z_\alpha$  tel que  $F(z_\alpha) = \frac{0,99+1}{2} = 0,995$

$$\text{On a } F(2,576) \simeq 0,995 \text{ d'où } z_\alpha \simeq 2,576 \text{ d'où } a_\alpha = z_\alpha \frac{s_e}{\sqrt{n-1}} \simeq 2,056$$

On estime donc que  $\mu$  est dans l’intervalle  $[11,13 - 2,056; 11,13 + 2,056] \simeq [9,07; 13,19]$ , avec la confiance  $c = 0,99$ .

- Les intervalles se chevauchent, donc ces échantillons ne permettent pas d’affirmer, avec la confiance 99%, que les filles aient en moyenne une meilleure perception des couleurs que les garçons.

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

(a) **Estimation de l’écart-type des filles**

On lit dans la table du  $\chi^2$  à 15 degrés de liberté la valeur  $x_1 = 4,601$  correspondant à  $p = 0,005$  et la valeur  $x_2 = 32,8$  correspondant à  $q = 0,005$

$$\text{On obtient alors } s_e \sqrt{\frac{n}{x_2}} \simeq 3,22 \sqrt{\frac{16}{32,8}} \simeq 2,25 \text{ et } s_e \sqrt{\frac{n}{x_1}} \simeq$$

$$3,22 \sqrt{\frac{16}{4,601}} \simeq 6$$

On estime donc que  $\sigma$  est dans l’intervalle  $[2,25; 6]$  avec la confiance  $c = 0,99$ .

(b) **Estimation de l’écart-type des garçons**

On lit dans la table du  $\chi^2$  à 38 degrés de liberté la valeur  $x_1 = 19,29$  correspondant à  $p = 0,005$  et la valeur  $x_2 = 64,18$  correspondant à  $q = 0,005$

$$\text{On obtient alors } s_e \sqrt{\frac{n}{x_2}} \simeq 4,92 \sqrt{\frac{39}{64,18}} \simeq 3,84 \text{ et } s_e \sqrt{\frac{n}{x_1}} \simeq$$

$$4,92 \sqrt{\frac{39}{19,29}} \simeq 7$$

On estime donc que  $\sigma$  est dans l’intervalle  $[3,84; 7]$  avec la confiance  $c = 0,99$ .

- (c) Les intervalles se chevauchent, donc on ne peut pas affirmer, avec la confiance 99%, que les écart-types diffèrent.

- Quelle taille d’échantillon aurait-il fallu considérer pour estimer le score moyen des filles à 0,2 point près, avec une confiance 99% ?

Pour avoir une précision 0,2 avec la confiance 0,99, il faut  $n > z_\alpha^2 \frac{(s_e)^2}{h^2} \simeq 2,576^2 \frac{3,22^2}{0,2^2} \simeq 1720$

**Remarque :** Une grande partie des données de ces exercices

sont fictives, et visent simplement à illustrer les outils mathématiques introduits en cours.