

### Exercices de TD de statistiques

#### Rappels

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

#### Exercice 1 : Fractions

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

- Grâce à cette fonctionnalité de la calculatrice, simplifiez chacune des fractions suivantes :  $\frac{76-34}{21}$ ,  $\frac{-77}{44}$ ,  $\frac{14}{15-7}$ ,  $\frac{-24}{-18}$ ,  $\frac{-10-81}{39\times 3}$ ,  $\frac{-10}{35}$ .

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

$$\frac{76-34}{21} = 2, \quad \frac{-77}{44} = -\frac{7}{4},$$

$$\frac{14}{15-7} = \frac{7}{4}, \quad \frac{-24}{-18} = \frac{4}{3},$$

$$\frac{-10-81}{39\times 3} = -\frac{7}{9}, \quad \frac{-10}{35} = -\frac{2}{7}.$$

$(76-34)\blacktriangleright$ Frac	2
$-77\blacktriangleright$ Frac	$-7/4$
$14\blacktriangleright(15-7)\blacktriangleright$ Frac	$7/4$
$-24\blacktriangleright-18\blacktriangleright$ Frac	$4/3$
$(-10-81)\blacktriangleright(39\times 3)\blacktriangleright$ Frac	$-7/9$
$-10\blacktriangleright 35\blacktriangleright$ Frac	$-2/7$

**Remarque :** La capture d'écran correspond à une TI. Sur une CASIO, si on veut par exemple calculer  $\frac{-77}{44}$ , on tape  $(-77)\div 44$  puis si le résultat affiché est un chiffre à virgule on appuie sur  $\boxed{F\leftrightarrow D}$  pour en faire une fraction. Au lieu d'afficher  $-\frac{7}{4}$ , certaines calculatrices affichent alors  $-1\frac{3}{4}$  (ces calculatrices considèrent que cela signifie  $-(1 + \frac{3}{4})$ ). Pour obtenir la réponse sous la bonne forme sur ces calculatrices, il faut alors appuyer sur  $\boxed{SHIFT}$  puis  $\boxed{F\leftrightarrow D}$  (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) :

$$\frac{76-34}{21} = 2, \quad \frac{-77}{44} = -1,75, \quad \frac{14}{15-7} = 1,75, \quad \frac{-24}{-18} \simeq 1,333,$$

$$\frac{-10-81}{39\times 3} \simeq -0,778, \quad \frac{-10}{35} \simeq -0,286.$$

$(76-34)\blacktriangleright 21$	2
$-77\blacktriangleright 44$	-1.75
$14\blacktriangleright(15-7)$	1.75
$-24\blacktriangleright-18$	1.333333333
$(-10-81)\blacktriangleright(39\times 3)$	-0.777777778
$-10\blacktriangleright 35$	-0.2857142857

L'ordre croissant de ces nombres est :  $-1,75 < -0,778 < -0,286 < 1,333 < 1,75 < 2$   
 Donc la réponse est :  $\frac{-77}{44} < \frac{-10-81}{39\times 3} < \frac{-10}{35} < \frac{-24}{-18} < \frac{14}{15-7} < \frac{76-34}{21}$

#### Exercice 2 : Utilisation de la calculatrice

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

$-1^2$	-1
--------	----

Peut-on en conclure

- que le carré de  $-1$  est  $-1$  ?
- que le carré de  $1$  est  $-1$  ?

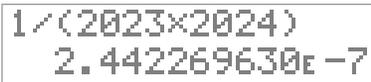
(c) que  $-1$  est l'opposé du carré de  $1$  ?

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

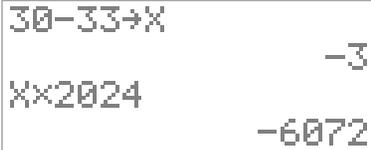
La réponse (b) est aussi fautive : le carré de  $1$  est égal à  $1$ .

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 calculatrice 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)  $\frac{1}{2023 \times 2024} \simeq 2,44$   
 (b)  $\frac{1}{2023 \times 2024} \simeq 24422696,3$   
 (c)  $\frac{1}{2023 \times 2024} \simeq 0,000000244$

La bonne réponse est la réponse (c). En effet **2.442269630E-7** désigne la quantité  $2,442269630 \times 10^{-7}$ , qui vaut  $0,0000002442269630$ .

3.  (a)  $30 - 33 \times 2024 = -6072$   
 (b)  $3 \times 2024 = -6072$   
 (c)  $(30 - 33) \times 2024 = -6072$

La bonne réponse est la réponse (c). En effet dans **X\*2024**, **X** vaut  $30 - 33$  (calculé à la ligne précédente), de sorte qu'on a calculé  $(30 - 33) \times 2024$ .

4.  (a)  $-10 \times 45^2 = 202500$   
 (b)  $-450^2 = 202500$   
 (c)  $(-10 \times 45)^2 = 202500$

La bonne réponse est la réponse (c). En effet dans **Ans^2**, **Ans**

vaut  $-10 \times 45$  (calculé à la ligne précédente), de sorte qu'on a calculé  $(-10 \times 45)^2$ .

### Exercice 3 : Puissances de dix

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

- $26 \div 100$                        $0,26 \times 10^2$                        $0,189\,794\,312 \times 1\,000\,000$
- $18\,687 \times 10^{-3}$                        $150\,961 \times 10^{-2}$                        $15\,650 \times 10^{-2}$
- $1\,472,1 \times 100\,000\,000$                        $1\,536 \times 10^{-9}$                        $118\,510,05 \times 10^9$
- $26 \div 100 = 0,26$                       (que l'on peut aussi écrire  $2,6 \times 10^{-1}$ )
- $0,26 \times 10^2 = 26$                       (que l'on peut aussi écrire  $2,6 \times 10^1$ )
- $0,189\,794\,312 \times 1\,000\,000 = 189\,794,312$   
 (que l'on peut aussi écrire  $1,89794312 \times 10^5$ )
- $18\,687 \times 10^{-3} = 18,687$   
 (que l'on peut aussi écrire  $1,8687 \times 10^1$ )
- $150\,961 \times 10^{-2} = 1509,61$   
 (que l'on peut aussi écrire  $1,50961 \times 10^3$ )
- $15\,650 \times 10^{-2} = 156,5$                       (que l'on peut aussi écrire  $1,565 \times 10^2$ )
- $1\,472,1 \times 100\,000\,000 = 147\,210\,000\,000$   
 (que l'on peut aussi écrire  $1,4721 \times 10^{11}$ )
- $1\,536 \times 10^{-9} = 0,000001536$   
 (que l'on peut aussi écrire  $1,536 \times 10^{-6}$ )
- $118\,510,05 \times 10^9 = 118\,510\,050\,000\,000$   
 (que l'on peut aussi écrire  $1,1851005 \times 10^{14}$ )

### Exercice 4 : Règles de calcul

1. On considère les expressions suivantes :

- $2 \times 8 + (9 + (5 + 3) \times 4 \times 4)$
- $9 + 2 \times 8 + (4^2 \times 5 + (3))$
- $(9 + (2 \times 8)) + 4^2(5 + 3)$
- $9 \times 8 + 2 \times 8 + 4^2(5 + 3)$
- $9 + 2 \times 8 + 4^2 \times 5 + 3$
- $(9 + 2) \times 8 + (5 + 3) \times 4^2$

Calculez la valeur de chacune de ces expressions, et observez que certaines sont égales.

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

$2*8+(9+(5+3)*4*4)$	153
$9+2*8+(4^2*5+(3))$	108
$(9+(2*8))+4^2*(5+3)$	153
$9*8+2*8+4^2*(5+3)$	216
$9+2*8+4^2*5+3$	108
$(9+2)*8+(5+3)*4^2$	216

- $2 \times 8 + (9 + (5 + 3) \times 4 \times 4) = 153$
- $9 + 2 \times 8 + (4^2 \times 5 + (3)) = 108$
- $(9 + (2 \times 8)) + 4^2(5 + 3) = 153$
- $9 \times 8 + 2 \times 8 + 4^2(5 + 3) = 216$
- $9 + 2 \times 8 + 4^2 \times 5 + 3 = 108$
- $(9 + 2) \times 8 + (5 + 3) \times 4^2 = 216$

On remarque alors d'une part que les expressions  $(9 + (2 \times 8)) + 4^2(5 + 3)$  et  $2 \times 8 + (9 + (5 + 3) \times 4 \times 4)$  sont égales, et d'autre part que les expressions  $(9 + 2) \times 8 + (5 + 3) \times 4^2$  et  $9 \times 8 + 2 \times 8 + 4^2(5 + 3)$  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, calculez les expressions suivantes lorsque les variables prennent les valeurs indiquées, et observez que certaines d'entre-elles sont égales.

(a) Les expressions

•  $-9 - 8x + 6$       •  $8x + 6 - 9x$       •  $(6 - 9) - 8x$

•  $6 - 3 \times \frac{9-8}{3}x$       •  $6 - (9 - 8)x$

lorsque  $x = 6$ .

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

- $-9 - 8x + 6 = -51$
- $6 - 3 \times \frac{9-8}{3}x = 0$
- $8x + 6 - 9x = 0$
- $6 - (9 - 8)x = 0$
- $(6 - 9) - 8x = -51$

$6-x$	6
$-9-8*x+6$	-51
$6-3*(9-8)/3*x$	0
$8*x+6-9*x$	0
$6-(9-8)*x$	0
$(6-9)-8*x$	-51

Ainsi, lorsque  $x = 6$ , on obtient d'une part que les expressions  $6 - 3 \times \frac{9-8}{3}x$ ,  $8x + 6 - 9x$  et  $6 - (9 - 8)x$  sont égales, d'autre part que les expressions  $-9 - 8x + 6$  et  $(6 - 9) - 8x$  sont égales.

(b) Les expressions

- $\frac{a}{4 \times 8 \div b}$
- $a \div 8 \times b \div 4$
- $\frac{a \div b}{4 \times 8}$
- $\frac{1}{4} \times \frac{a}{b \times 8}$
- $b \frac{a}{4 \times 8}$
- $\frac{b \times a}{4 \times 8}$

lorsque  $a = 2$  et  $b = -6$ .

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

- $\frac{a}{4 \times 8 \div b} = -0,375$
- $a \div 8 \times b \div 4 = -0,375$
- $\frac{a \div b}{4 \times 8} \simeq -0,01$
- $\frac{1}{4} \times \frac{a}{b \times 8} \simeq -0,01$
- $b \frac{a}{4 \times 8} = -0,375$
- $\frac{b \times a}{4 \times 8} = -0,375$

$2+A$	2
$-6 \div B$	-6
$A \div (4*8 \div B)$	-0.375
$A \div 8 * B \div 4$	-0.375
$A \div B \div (4*8)$	-0.01041666667
$1 \div 4 * A \div (B * 8)$	-0.01041666667
$B * A \div (4*8)$	-0.375
$B * A \div (4*8)$	-0.375

Ainsi, pour ces valeurs de  $a$  et  $b$ , on obtient d'une part que les expressions  $\frac{a}{4 \times 8 \div b}$ ,  $a \div 8 \times b \div 4$ ,  $b \frac{a}{4 \times 8}$  et  $\frac{b \times a}{4 \times 8}$  sont égales, d'autre part que les expressions  $\frac{a \div b}{4 \times 8}$  et  $\frac{1}{4} \times \frac{a}{b \times 8}$  sont égales.

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

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

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

- $y \frac{(8-x)}{(7-z)} \simeq -9,692$
- $\frac{y \times 8}{7} - \frac{y \times x}{z} \simeq -1,286$
- $y \left( \frac{8}{7} - \frac{x}{z} \right) \simeq -1,286$
- $(y \times 8 - x \times y) \div (7 - z) \simeq -9,692$

-6+X	-6
-9+Y	-9
-6+Z	-6
Y*(8-X)/(7-Z)	-9.692307692
Y*8/7-Y*X/Z	-1.285714286
Y*(8/7-X/Z)	-1.285714286
(Y*8-X*Y)/(7-Z)	-9.692307692

Ainsi, pour ces valeurs de  $x$ ,  $y$  et  $z$ , on obtient d'une part que les expressions  $y \frac{(8-x)}{(7-z)}$  et  $(y \times 8 - x \times y) \div (7 - z)$  sont égales, d'autre part que les expressions  $\frac{y \times 8}{7} - \frac{y \times x}{z}$  et  $y \left( \frac{8}{7} - \frac{x}{z} \right)$  sont égales.

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

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

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

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

-3+X	-3
2+Y	2
Y+3-5-3+2*X+2*6	3
Y-5-1*(2-X-6)	-2
Y-5+2*X+2*6	3
Y-5-2*(-X-6)	3
Y-5+2-X-6-(2-X-6)	-3
Y-5-0*(2-X-6)	-3
Y-5-(2-X-6)	-2

**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  $x_2 + 3 - 5 - 3 + 2 \times x_1 + 2 \times 6$ ,  $x_2 - 5 + 2 \times x_1 + 2 \times 6$  et  $x_2 - 5 - 2(-x_1 - 6)$  sont égales, d'autre part que les expressions  $x_2 - 5 - 1(2 - x_1 - 6)$  et  $x_2 - 5 - (2 - x_1 - 6)$  sont égales, et enfin que les expressions  $x_2 - 5 + 2 - x_1 - 6 - (2 - x_1 - 6)$  et  $x_2 - 5 - 0(2 - x_1 - 6)$  sont égales.

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

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

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

- $(x_1 - (a_1 + a_2 - x_2)) = 3$
- $x_1 - a_1 + a_2 - x_2 = 15$
- $(x_2 - a_1) - (a_2 - x_1) = 3$
- $a_2 - a_1 + x_1 - x_2 = 15$

-1→A	-1
4→B	4
8→X	8
-2→Y	-2
(X-(A+B-Y))	3
X-A+B-Y	15
(Y-A)-(B-X)	3
B-A+X-Y	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  $(x_1 - (a_1 + a_2 - x_2))$  et  $(x_2 - a_1) - (a_2 - x_1)$  sont égales, d'autre part que les expressions  $x_1 - a_1 + a_2 - x_2$  et  $a_2 - a_1 + x_1 - x_2$  sont égales.

### Compléments

#### Exercice 5 : Précision des calculs

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

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

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

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

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

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

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

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

**Copie d'Anaëlle**

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

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

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

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

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

**Calculatrice d'Anaëlle**

(1.79+1.67+1.74)÷3	
1.733333333	
(1.79²+1.67²+1.74²)÷3	
3.006866667	
3.01-1.73²	
0.0171	
√(0.02)	
0.1414213562	

**Copie de Christopher**

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

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

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

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

Donc l'écart-type est d'environ 5 cm.

**Calculatrice de Christopher**

1.79+1.67+1.74	
5.2	
1.79²+1.67²+1.74²	
9.0206	
9.0206÷3-(5.2÷3)²	
0.002422222222	
√(Ans)	
0.04921607687	

**Copie de Marie**

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

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

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

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

Donc l'écart-type est d'environ 5 cm.

**Calculatrice de Marie**

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

**Copie de Justin**

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

Donc l'écart-type est d'environ 5 cm.

**Calculatrice de Justin**

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

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

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

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

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

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

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

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

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

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

**Chapitre 1 : Statistiques descriptives univariées**

**Exercice 7 : Types de variables**

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

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

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

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

**Exercice 8 : Taille des ménages**

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

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

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

**Exercice 9 : Moyens de transport**

1. Utilisation des transports en commun :

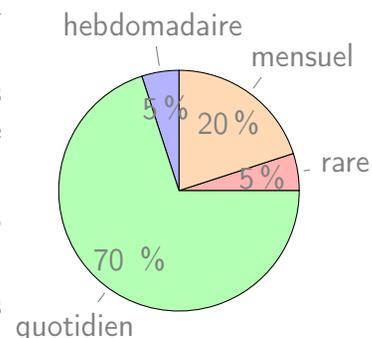
(a) Un étudiant en sociologie interroge 40 personnes de son entourage sur leur utilisation des transports en commun. Il réalise le graphique ci-dessous après les avoir regroupés en quatre groupes :

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

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

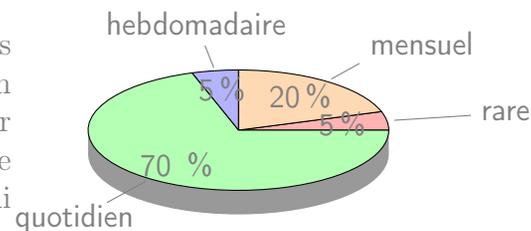
**hebdomadaire** : entre une et cinq fois par semaine

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



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

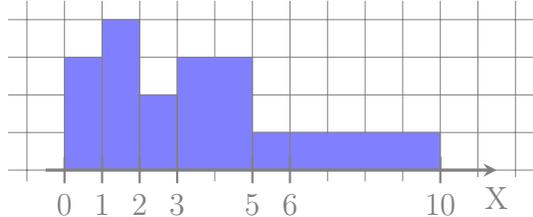
(b) Lorsqu'il entre les données dans un tableur, l'ordinateur réalise un graphique ressemblant à celui ci-contre. Que penser d'un tel graphique ?



2. Utilisation de la voiture :

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

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



(a) Déterminez les fréquences des différentes classes.

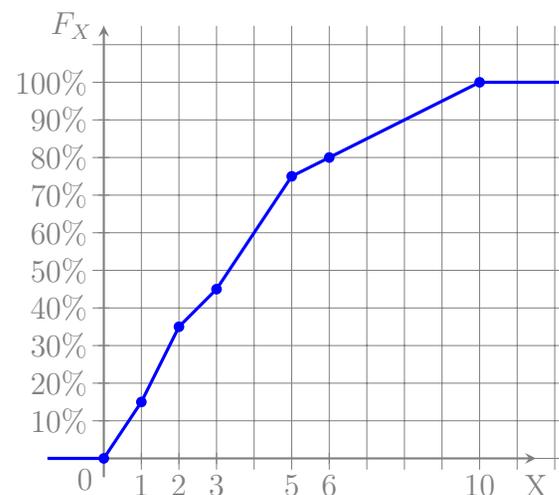
$X$	$[0; 1[$	$[1; 2[$	$[2; 3[$	$[3; 5[$	$[5; 6[$	$[6; 10[$
Fréquence						

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

**3. Polygone des fréquences cumulées :**

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

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



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

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

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

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

**Exercice 11 : Développement psychomoteur**

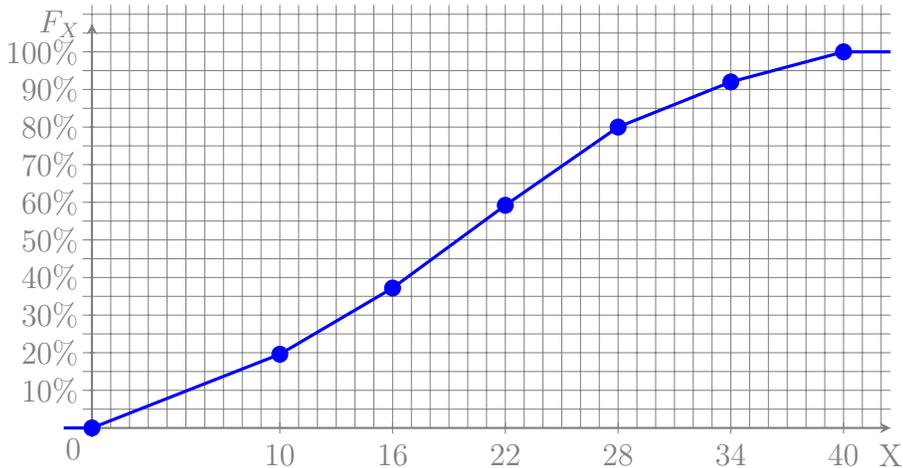
Les données ci-dessous décrivent, pour un groupe de 200 bébés âgés de deux ans, la variable statistique  $X$  représentant le “score de développement psychomoteur” (SDP) :

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

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

**Exercice 12 : Méthodes d'apprentissage**

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



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

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

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

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

**Exercice 13 : Âge et performances mémorielles**

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

Les données mesurées sont les suivantes :

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

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

**Exercice 14 : Types de variables**

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

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

Cas	Population	Variable	Nature
(a)	l'ensemble des voitures stationnées dans le campus	la couleur	qualitative nominale
(b)	les femmes dijonnaises	le nombre d'enfants	quantitative discrète
(c)	l'ensemble des usagers des transports en commun	le degré de satisfaction	qualitative ordinale
(d)	l'ensemble des appareils photos vendus en 2023	le poids	quantitative continue

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

- (a) les langues officielles des différents pays européens  
qualitative nominale
- (b) la pluie tombée à Dijon au cours des différents mois de l'année  
quantitative continue
- (c) le nombre de pièces des appartements en location à Dijon  
quantitative discrète
- (d) les numéros de sécurité sociale des enseignants de l'Université de Bourgogne  
qualitative nominale
- (e) le nombre moyen d'enfants par femme dans chaque pays européen  
quantitative continue
- (f) les marques des téléphones vendus en 2023  
qualitative nominale
- (g) le nombre de pages des livres d'une bibliothèque  
quantitative discrète

- (h) les numéros de téléphones figurant dans les pages blanches dijonnaises  
qualitative nominale
- (i) le nombre de crises d'épilepsie de différents patients au cours du mois d'octobre 2023  
quantitative discrète
- (j) la douleur ressentie par différents patients atteints de la même pathologie  
qualitative ordinale.

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

$$\text{moyenne : } m(T) = \frac{\sum_i x_i n_i}{n} = \frac{1 \times 1001 + 2 \times 1337 + 3 \times 1505 + \dots + 6 \times 152}{5676} = \frac{16\,358}{5\,676} \simeq 2,88$$

$$m(T^2) = \frac{\sum_i x_i^2 n_i}{n} = \frac{1^2 \times 1001 + 2^2 \times 1337 + 3^2 \times 1505 + \dots + 6^2 \times 152}{5676} = \frac{57\,050}{5\,676}$$

$$\text{Var}(T) = m(T^2) - m(T)^2 = \frac{57\,050}{5\,676} - \left(\frac{16\,358}{5\,676}\right)^2 \simeq 1,75$$

$$\text{Écart-type : } s(T) = \sqrt{\text{Var}(T)} \simeq 1,32$$

**Exercice 16 : Questionnaire de logique**

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

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

- Quel a été le taux de réponse à la question 7 ?  
En tout, 11 personnes (parmi les 12 interrogées) ont répondu à la question 7. Le taux de réponse à cette question est donc  $\frac{11}{12} \simeq 0,9167$ .
- Quelle proportion de sujets a répondu correctement à la question 4 ?  
Comme 2 personnes ont répondu correctement à la question 4, cette proportion est  $\frac{2}{12} \simeq 0,1667$ .
- Parmi les individus ayant répondu à la question 8, quelle proportion a donné la bonne réponse ?  
Parmi les 7 personnes qui ont répondu à la question 8, il y en a 3 qui ont répondu correctement, soit une proportion de  $\frac{3}{7} \simeq 0,4286$ .
- Le chercheur décide de résumer par une note les réponses données au questionnaire. Il attribue un point par bonne réponse (et zéro

point par réponse erronée ou manquante). Calculez les notes (sur 10) attribuée à chaque adolescent interrogé.

Individu	A	B	C	D	E	F	G	H	I	J	K	L
Note	5	4	7	2	3	4	5	2	2	3	4	2

- Calculez la moyenne, l’écart-type et la médiane de ces notes.

moyenne :  $m(X) = \frac{\sum x_i}{n} = \frac{5+4+7+\dots+2}{12} = \frac{43}{12} \simeq 3,58$

$m(X^2) = \frac{\sum x_i^2}{n} = \frac{5^2+4^2+7^2+\dots+2^2}{12} = \frac{181}{12}$

$Var(X) = m(X^2) - m(X)^2 = \frac{181}{12} - \left(\frac{43}{12}\right)^2 \simeq 2,24$

Écart-type :  $s(X) = \sqrt{Var(X)} \simeq 1,5$

La médiane est la valeur numéro  $\frac{12+1}{2} = 6,5$  (en ordonnant par ordre croissant), ou plutôt le milieu entre les valeurs numéro 6 et 7. C’est donc le milieu entre 3 et 4 c’est-à-dire  $\frac{3+4}{2} = 3,5$ .

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

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

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

- Faites un tri des données en indiquant l’effectif de chaque note.

modalité	1	2	2,5	3	4	4,5	5	6	6,5	7	7,5	8	9	9,5	10
effectif	1	1	1	2	4	1	4	7	1	1	2	5	3	2	1

- Quelle est la taille de l’échantillon ?  
La taille de l’échantillon est  $n = 36$ .
- Donner la moyenne, l’écart-type et la médiane.  
moyenne :  $m(X) = \frac{\sum_i x_i n_i}{n} = \frac{1 \times 1 + 2 \times 1 + 2,5 \times 1 + \dots + 10 \times 1}{36} = \frac{218,5}{36} \simeq 6,07$   
 $m(X^2) = \frac{\sum_i x_i^2 n_i}{n} = \frac{1^2 \times 1 + 2^2 \times 1 + 2,5^2 \times 1 + \dots + 10^2 \times 1}{36} = \frac{1512,75}{36}$   
 $Var(X) = m(X^2) - m(X)^2 = \frac{1512,75}{36} - \left(\frac{218,5}{36}\right)^2 \simeq 5,18$   
Écart-type :  $s(X) = \sqrt{Var(X)} \simeq 2,28$   
La médiane est la valeur numéro  $\frac{36+1}{2} = 18,5$  (en ordonnant par ordre croissant). C’est donc 6.

4. Transformez les données en les rangeant en classes d'amplitude 1,5. Calculez 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	2	6	5	9	7	6
fréquences	0,028	0,056	0,167	0,139	0,25	0,194	0,167
fréquences cumulées	0,028	0,084	0,251	0,39	0,64	0,834	1,001

5. Calculez la moyenne, l'écart-type et la médiane des notes regroupées en classes. Que remarque-t-on ?

moyenne :

$$m(X) = \frac{\sum c_i n_i}{\sum n_i} = \frac{0,75 \times 1 + 2,25 \times 2 + 3,75 \times 6 + \dots + 9,75 \times 6}{36} = \frac{231}{36} \simeq 6,42$$

$$m(X^2) = \frac{\sum c_i^2 n_i}{\sum n_i} = \frac{0,75^2 \times 1 + 2,25^2 \times 2 + 3,75^2 \times 6 + \dots + 9,75^2 \times 6}{36} = \frac{1689,75}{36}$$

$$Var(X) = m(X^2) - m(X)^2 = \frac{1689,75}{36} - \left(\frac{231}{36}\right)^2 \simeq 5,76$$

$$\text{Écart-type} : s(X) = \sqrt{Var(X)} \simeq 2,4$$

Classe de la médiane : [6; 7,5[

$$\text{Méd} \simeq a_i + \frac{a_{i+1} - a_i}{F_X(a_{i+1}) - F_X(a_i)} (0,5 - F_X(a_i)) \simeq 6 + \frac{7,5 - 6}{0,64 - 0,39} (0,5 - 0,39) \simeq 6,66$$

Ce regroupement des données en classes 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.

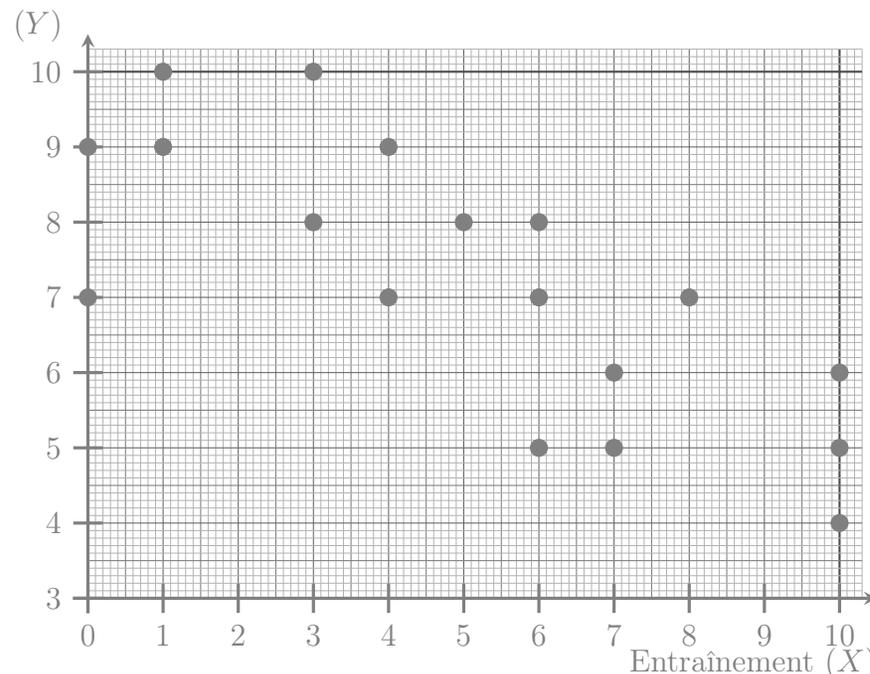
### Chapitre 2 : Statistiques descriptives bivariées

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

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

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

Nombre d'erreurs



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

Entraînement $X$																				
Nombre d'erreurs $Y$																				

2. Calculez le coefficient de corrélation des rangs (de Spearman) et interprétez la valeur de ce coefficient.

- Calculez les moyennes et écart-types des variables  $X$  et  $Y$ , puis leur coefficient de corrélation linéaire (de Pearson). Interprétez la valeur de ce coefficient.
- Si un enfant a déjà fait 7 exercices de logique, alors combien estimeriez-vous qu'il fera d'erreurs si on lui fait à nouveau passer un test similaire ?
- Un enfant a commis 6 erreurs. Combien de fois estimeriez-vous qu'il avait déjà fait un test similaire pour s'entraîner ?

**Exercice 19 : Stress et temps de réponse**

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

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

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

**Exercice 20 : Revenus et espérance de vie**

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

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

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

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

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

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

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

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

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

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

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

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

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

**Exercice 23 : Extraversion**

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

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

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

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

$$\text{moyenne : } m(X) = \frac{\sum x_i}{n} = \frac{7,6+6,4+10,1+\dots+7,1}{10} = \frac{67,9}{10} = 6,79$$

$$m(X^2) = \frac{\sum x_i^2}{n} = \frac{7,6^2+6,4^2+10,1^2+\dots+7,1^2}{10} = \frac{489,57}{10}$$

$$\text{Var}(X) = m(X^2) - m(X)^2 = \frac{489,57}{10} - \left(\frac{67,9}{10}\right)^2 \simeq 2,85$$

$$\text{Écart-type : } s(X) = \sqrt{\text{Var}(X)} \simeq 1,69$$

$$\text{moyenne : } m(Y) = \frac{\sum x_i}{n} = \frac{3,7+6,4+3,2+\dots+0,5}{10} = \frac{38,4}{10} = 3,84$$

$$m(Y^2) = \frac{\sum x_i^2}{n} = \frac{3,7^2+6,4^2+3,2^2+\dots+0,5^2}{10} = \frac{175,38}{10}$$

$$\text{Var}(Y) = m(Y^2) - m(Y)^2 = \frac{175,38}{10} - \left(\frac{38,4}{10}\right)^2 \simeq 2,79$$

$$\text{Écart-type : } s(Y) = \sqrt{\text{Var}(Y)} \simeq 1,67$$

$$m(XY) = \frac{\sum x_i y_i}{n} = \frac{7,6 \times 3,7 + 6,4 \times 6,4 + \dots + 7,1 \times 0,5}{10} = \frac{265,89}{10} = 26,589$$

$$\text{Cov}(X,Y) = m(XY) - m(X)m(Y) = \frac{265,89}{10} - \frac{67,9}{10} \frac{38,4}{10} = 0,5154$$

$$r(X,Y) = \frac{\text{Cov}(X,Y)}{s(X)s(Y)} = \frac{0,5154}{1,69 \times 1,67} \simeq 0,1826$$

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

Il semble donc que les deux variables  $X$  et  $Y$  ne sont pas liées. Les résumer par leur somme ferait 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 formes**

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

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

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

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

sujet	1	2	3	4	5	6	7	8	9	10	11	12	13	14
$X$	0	0	1	1	2	3	4	5	6	6	7	7	9	10
$Y$	3,6	4,3	3,3	3,6	3,7	2,7	2,7	2,2	2	2,1	1,8	2,3	1,4	1,2
rang $x'$	1,5	1,5	3,5	3,5	5	6	7	8	9,5	9,5	11,5	11,5	13	14
rang $y'$	11,5	14	10	11,5	13	8,5	8,5	6	4	5	3	7	2	1
$(x' - y')^2$	100	156,25	42,25	64	64	6,25	2,25	4	30,25	20,25	72,25	20,25	121	169

le coefficient de corrélation des rangs de Spearman est donc

$$1 - \left(6 \times \frac{100+156,25+42,25+64+64+\dots+169}{14(14^2-1)}\right) \simeq -0,916$$

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

2. Calculez les moyennes et écart-types des variables  $X$  et  $Y$ , puis leur coefficient de corrélation linéaire. Interprétez la valeur de ce coefficient.

$$\text{moyenne : } m(X) = \frac{\sum x_i}{n} = \frac{0+0+1+\dots+10}{14} = \frac{61}{14} \simeq 4,36$$

$$m(X^2) = \frac{\sum x_i^2}{n} = \frac{0^2+0^2+1^2+\dots+10^2}{14} = \frac{407}{14}$$

$$\text{Var}(X) = m(X^2) - m(X)^2 = \frac{407}{14} - \left(\frac{61}{14}\right)^2 \simeq 10,09$$

$$\text{Écart-type : } s(X) = \sqrt{\text{Var}(X)} \simeq 3,18$$

$$\text{moyenne : } m(Y) = \frac{\sum x_i}{n} = \frac{3,6+4,3+3,3+\dots+1,2}{14} = \frac{36,9}{14} \simeq 2,64$$

$$m(Y^2) = \frac{\sum x_i^2}{n} = \frac{3,6^2+4,3^2+3,3^2+\dots+1,2^2}{14} = \frac{108,75}{14}$$

$$\text{Var}(Y) = m(Y^2) - m(Y)^2 = \frac{108,75}{14} - \left(\frac{36,9}{14}\right)^2 \simeq 0,821$$

$$\text{Écart-type : } s(Y) = \sqrt{\text{Var}(Y)} \simeq 0,906$$

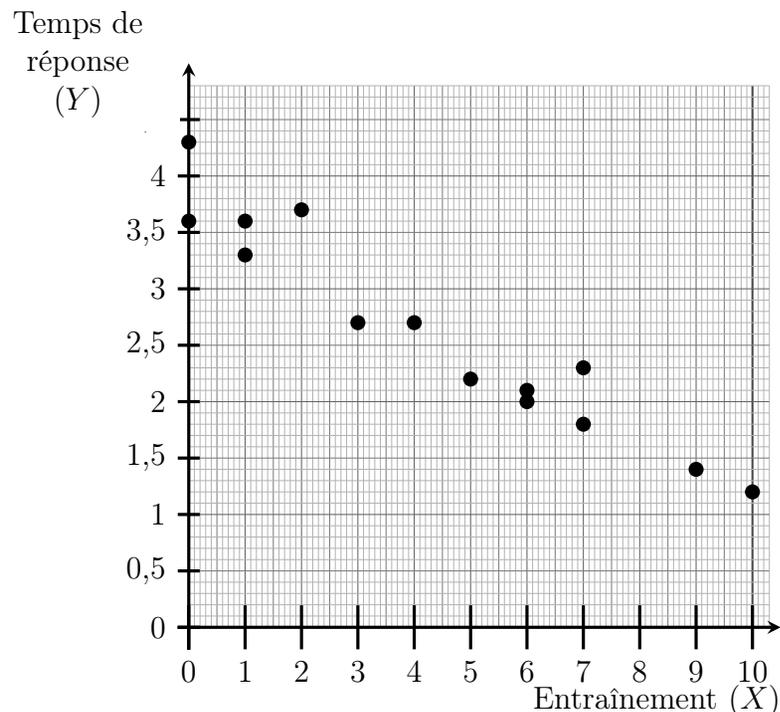
$$m(XY) = \frac{\sum x_i y_i}{n} = \frac{0 \times 3,6 + 0 \times 4,3 + \dots + 10 \times 1,2}{14} = \frac{122,1}{14} \simeq 8,721$$

$$\text{Cov}(X,Y) = m(XY) - m(X)m(Y) = \frac{122,1}{14} - \frac{61}{14} \frac{36,9}{14} \simeq -2,763$$

$$r(X,Y) = \frac{\text{Cov}(X,Y)}{s(X)s(Y)} = \frac{-2,763}{3,18 \times 0,906} \simeq -0,959$$

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

3. Tracez le nuage de points correspondant à ces données. Ce graphique vous conforte-t-il quant aux interprétations données aux questions précédentes ?



On remarque que dans l'ensemble, les points sont assez proches d'une droite, comme attendu du calcul des coefficients de corrélation.

4. Si un enfant a déjà fait 8 exercices de reconnaissance de formes, alors à combien estimeriez-vous son temps de réponse lors d'un prochain test ?

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

on pose  $a = \frac{Cov(X,Y)}{Var(X)} \simeq \frac{-2,763}{10,09} \simeq -0,27$  et  $b = m(Y) - a m(X) \simeq 2,636 - (-0,274) \times 4,357 \simeq 3,83$

D'où l'équation de la droite  $D_{Y|X} : Y = -0,27 X + 3,83$

Donc pour  $x = 8$ , on s'attend à  $y = -0,27 \times 8 + 3,83 = 1,67$  ..

5. Un enfant a mis 2,4 secondes pour répondre au test. Combien

de fois estimeriez-vous qu'il avait déjà fait un test similaire pour s'entraîner ?

Pour répondre à cette question, on utilise la droite de régression

$D_{X|Y}$  :

on pose  $a' = \frac{Cov(X,Y)}{Var(Y)} \simeq \frac{-2,76}{0,821} \simeq -3,36$  et  $b' = m(X) - a' m(Y) \simeq 4,36 - (-3,36) \times 2,64 \simeq 13,23$

D'où l'équation de la droite  $D_{X|Y} : X = -3,36 Y + 13,23$

Donc pour  $y = 2,4$ , on s'attend à  $x = -3,36 \times 2,4 + 13,23 = 5,166$

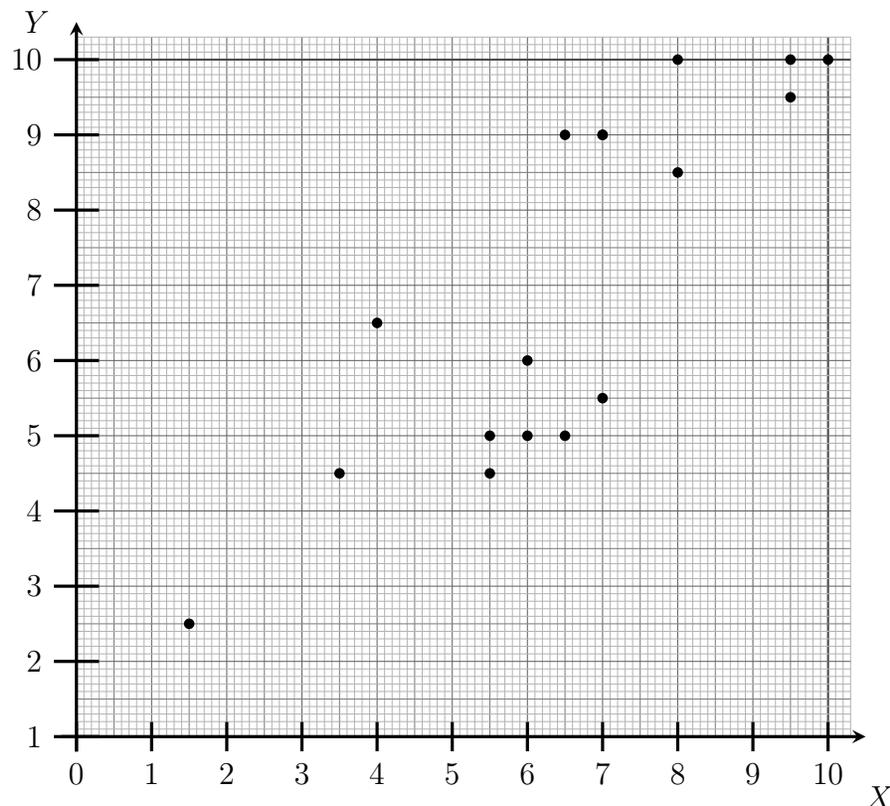
..

**Exercice 25 : Résolution de problèmes mathématiques**

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

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

1. Dessinez le nuage statistique de ces variables.



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

sujet	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
$X$	7	10	5,5	6,5	9,5	8	5,5	4	6	3,5	6	7	6,5	9,5	7	8	1,5
$Y$	9	10	5	9	10	8,5	4,5	6,5	6	4,5	5	5,5	5	9,5	9	10	2,5
rang $x'$	11	17	4,5	8,5	15,5	13,5	4,5	3	6,5	2	6,5	11	8,5	15,5	11	13,5	1
rang $y'$	12	16	5	12	16	10	2,5	9	8	2,5	5	7	5	14	12	16	1
$(x' - y')^2$	1	1	0,25	12,25	0,25	12,25	4	36	2,25	0,25	2,25	16	12,25	2,25	1	6,25	0

le coefficient de corrélation des rangs de Spearman est donc

$$1 - \left( 6 \times \frac{1+1+0,25+12,25+0,25+\dots+0}{17(17^2-1)} \right) \simeq 0,8658.$$

3. Calculez les moyennes, les écart-types et le coefficient de corrélation linéaire des deux variables  $X$  et  $Y$ .

$$\text{moyenne : } m(X) = \frac{\sum x_i}{n} = \frac{7+10+5,5+\dots+1,5}{17} = \frac{111}{17} \simeq 6,53$$

$$m(X^2) = \frac{\sum x_i^2}{n} = \frac{7^2+10^2+5,5^2+\dots+1,5^2}{17} = \frac{803}{17}$$

$$\text{Var}(X) = m(X^2) - m(X)^2 = \frac{803}{17} - \left(\frac{111}{17}\right)^2 \simeq 4,602$$

$$\text{Écart-type : } s(X) = \sqrt{\text{Var}(X)} \simeq 2,15$$

$$\text{moyenne : } m(Y) = \frac{\sum x_i}{n} = \frac{9+10+5+\dots+2,5}{17} = \frac{119,5}{17} \simeq 7,03$$

$$m(Y^2) = \frac{\sum x_i^2}{n} = \frac{9^2+10^2+5^2+\dots+2,5^2}{17} = \frac{935,75}{17}$$

$$\text{Var}(Y) = m(Y^2) - m(Y)^2 = \frac{935,75}{17} - \left(\frac{119,5}{17}\right)^2 \simeq 5,63$$

$$\text{Écart-type : } s(Y) = \sqrt{\text{Var}(Y)} \simeq 2,37$$

$$m(XY) = \frac{\sum x_i y_i}{n} = \frac{7 \times 9 + 10 \times 10 + \dots + 1,5 \times 2,5}{17} = \frac{852,5}{17} \simeq 50,147$$

$$\text{Cov}(X,Y) = m(XY) - m(X)m(Y) = \frac{852,5}{17} - \frac{111}{17} \frac{119,5}{17} \simeq 4,249$$

$$r(X,Y) = \frac{\text{Cov}(X,Y)}{s(X)s(Y)} = \frac{4,249}{2,15 \times 2,37} \simeq 0,8339$$

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  $a = \frac{\text{Cov}(X,Y)}{\text{Var}(X)} \simeq \frac{4,249}{4,602} \simeq 0,9233$  et  $b = m(Y) - a m(X) \simeq 7,029 - 0,9233 \times 6,529 \simeq 1,001$

D'où l'équation de la droite  $D_{Y|X} : Y = 0,9233 X + 1,001$

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

Donc pour  $x = 9$ , on s'attend à  $y = 0,9233 \times 9 + 1,001 \simeq 9,311$

### Chapitre 3 : probabilités

#### Exercice 26 : Situations d'emploi

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

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

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

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

### Exercice 27 : Factorielles & coefficients binomiaux

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

### Exercice 28 : Croissance religieuse

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

Argumentez le fait 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. Calculez 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. Calculez la loi de  $X$ .
- Comparez entre elles les lois obtenues dans ces trois situations.

### Exercice 29 : Ressources humaines

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

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

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

**Exercice 30 : Loi binomiale**

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

1) Calculez  $n$  et  $p$ .

$$\begin{cases} np = m(X) = 4 \\ np(1 - p) = \text{Var}(X) = 3 \end{cases}$$

d'où  $1 - p = \frac{3}{4} = 0,75$  donc  $p = 0,25$  et  $n = \frac{4}{0,25} = 16$ .

2) Calculez les probabilités  $\mathbb{P}[9 \leq X \leq 11]$  et  $\mathbb{P}[X > 13]$ .

$$\begin{aligned} \bullet \mathbb{P}[9 \leq X \leq 11] &= \mathbb{P}[X = 9] + \mathbb{P}[X = 10] + \mathbb{P}[X = 11] \\ &= \binom{16}{9} \times 0,25^9 (1 - 0,25)^7 + \binom{16}{10} \times 0,25^{10} (1 - 0,25)^6 + \binom{16}{11} \times 0,25^{11} (1 - 0,25)^5 \\ &\simeq 0,005\,825 + 0,001\,359 + 0,000\,247 \\ &= 0,007\,431 \end{aligned}$$

**Remarque :** Avec la calculatrice, on obtient directement  $\mathbb{P}[9 \leq X \leq 11] = \mathbb{P}[X \leq 11] - \mathbb{P}[X \leq 8] \simeq 0,999\,962 - 0,992\,53 = 0,007\,432$

$$\begin{aligned} \bullet \mathbb{P}[X > 13] &= \mathbb{P}[X = 14] + \mathbb{P}[X = 15] + \mathbb{P}[X = 16] \\ &= \binom{16}{14} \times 0,25^{14} (1 - 0,25)^2 + \binom{16}{15} \times 0,25^{15} (1 - 0,25) + \binom{16}{16} \times 0,25^{16} \\ &\simeq 2,515 \times 10^{-07} + 1,12 \times 10^{-08} + 2,0 \times 10^{-10} \\ &= 2,629 \times 10^{-07} \end{aligned}$$

**Remarque :** Avec la calculatrice, on obtient directement  $\mathbb{P}[X > 13] = 1 - \mathbb{P}[X \leq 13] \simeq 1 - 0,999\,999\,737\,1 \simeq 2,629 \times 10^{-07}$

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

On calcule successivement :

$$\mathbb{P}[X \geq 16] \simeq 2,328 \times 10^{-10}$$

$$\mathbb{P}[X \geq 15] \simeq 1,141 \times 10^{-08}$$

$$\mathbb{P}[X \geq 14] \simeq 2,629 \times 10^{-07}$$

$$\mathbb{P}[X \geq 13] \simeq 3,783 \times 10^{-06}$$

$$\mathbb{P}[X \geq 12] \simeq 3,811 \times 10^{-05}$$

$$\mathbb{P}[X \geq 11] \simeq 0,0002852$$

$$\mathbb{P}[X \geq 10] \simeq 0,001644$$

$$\mathbb{P}[X \geq 9] \simeq 0,00747$$

$$\mathbb{P}[X \geq 8] \simeq 0,02713$$

$$\mathbb{P}[X \geq 7] \simeq 0,07956$$

Donc les valeurs de  $k$  qui conviennent sont lorsque  $k > 7$ .

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

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

1. On choisit au hasard le nom d'une de ces trois personnes, puis à nouveau le nom d'une de ces trois personnes au hasard (ce peut être la même personne – ou pas).

(a) Listez les neuf possibilités pour ces deux noms choisis au hasard.

En notant par exemple "AB" pour "Alice puis Bernard", les neuf possibilités sont : AA, AB, AC, BA, BB, BC, CA, CB et CC.

(b) Parmi ces possibilités, combien comptent uniquement des personne en situation d'emploi ?

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

- (c) 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  $\frac{4}{9} \simeq 0,4444$ .
- (d) Pour un tel choix aléatoire de deux noms, on note  $X$  le nombre de personnes choisies exerçant un emploi. Quelle est la loi de variable  $X$  ? 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  $\mathcal{B}(2; \frac{2}{3})$ .
- (e) Retrouvez le résultat de la question c), en utilisant cette loi.  

$$\mathbb{P}[X = 1] = \binom{2}{1} (2/3)^1 (1/3)^1 \simeq 0,444.$$
- (f) Déterminez de même  $\mathbb{P}[X = 0]$  et  $\mathbb{P}[X = 2]$ .
- $\mathbb{P}[X = 0] = \binom{2}{0} (2/3)^0 (1/3)^2 \simeq 0,111.$
  - $\mathbb{P}[X = 2] = \binom{2}{2} (2/3)^2 (1/3)^0 \simeq 0,444.$
2. On convient d'une autre manière de choisir deux noms : on choisit un premier nom au hasard, puis on choisit le second nom parmi les deux restant. Reprenez, avec cette nouvelle façon de choisir les deux noms, les questions 1a, 1b et 1c.
- (a) On a cette fois-ci 6 possibilités : 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  $\frac{4}{6} \simeq 0,6667$ .
1. Il mélange les cartes, et constate qu'elle sont dans un certain ordre à l'issue du mélange. Quelle était la probabilité qu'à l'issue du mélange, elles se retrouvent précisément dans cet ordre là ? Il y a  $(5)! = 5 \times 4 \times 3 \times 2 = 120$  résultats possibles à l'issue du mélange. La probabilité est donc  $\frac{1}{120} \simeq 0,008333$ .
2. S'il remélange les cartes. Calculez les probabilités des événements suivants :
- (a) la carte qui était en haut du tas soit à nouveau en haut du tas ?  
 Il y a  $(4)! = 4 \times 3 \times 2 = 24$  mélanges qui conservent la première carte (car cela correspond à ne mélanger que les 4 autres cartes). La probabilité est donc  $\frac{24}{120} = 0,2$ .
- (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  $(3)! = 3 \times 2 \times 1 = 6$  mélanges qui conservent les deux premières cartes (car cela correspond à ne mélanger que les 3 autres cartes). La probabilité est donc  $\frac{6}{120} = 0,05$ .
- (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  $\frac{6}{120} = 0,05$ .
- (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  $\frac{6}{120} + \frac{6}{120} = 0,1$ .
- (e) l'ensemble des cartes soient exactement dans le même ordre que lors du premier mélange ?  
 $\frac{1}{120} \simeq 0,008333$ .
3. Répondez aux mêmes questions en supposant qu'il y ait cette fois-ci 29 cartes.

### Exercice 32 : Collection de dessins

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

1.  $\frac{1}{(29)!} = \frac{1}{(29) \times (28) \times \dots \times (1)} = 1,130\,996\,288\,644\,771\,5 \times 10^{-31}$
2. (a)  $\frac{(28)!}{(29)!} = \frac{(28) \times (27) \times \dots \times (1)}{(29) \times (28) \times \dots \times (1)} = \frac{1}{29} \simeq 0,034\,48$
- (b)  $\frac{(27)!}{(29)!} = \frac{(27) \times (26) \times \dots \times (1)}{(29) \times (28) \times \dots \times (1)} = \frac{1}{(29) \times (28)} \simeq 0,001\,231\,5$
- (c)  $\frac{1}{29 \times 28} \simeq 0,001\,231\,5$
- (d)  $\frac{1}{29 \times 28} + \frac{1}{29 \times 28} \simeq 0,002\,463\,1$
- (e)  $\frac{1}{(29)!} = \frac{1}{(29) \times (28) \times \dots \times (1)} = 1,130\,996\,288\,644\,771\,5 \times 10^{-31}$

4. Peut-on conclure des probabilités calculées que certains évènements 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 5 cartes, et quasi-impossible pour 29 cartes).

Dans la question 1 on ne peut rien en déduire : chaque résultat à l'issue du mélange a une très faible probabilité, et il est complètement artificiel de choisir de calculer la probabilité du tirage qui a eu lieu. Quoi qu'il arrive le résultat sur lequel on tombe sera forcément un parmi des milliards (de milliards), donc un résultat qui avait extrêmement peu de chances de tomber, mais il faut bien 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 5 cartes) et même quasiment impossible (pour 29 cartes) de reproduire le même ordre qu'initialement.

Un point clé qui explique que le calcul a un sens dans la question 2e alors qu'il ne signifie essentiellement rien dans la question 1 est que dans la question 2e, "le même ordre qu'initialement" correspond à un cas précis identifiable avant de faire le mélange, alors que dans la question 1, on choisit de quoi on calcule la probabilité après avoir vu ce qu'on a obtenu comme mélange.

### Exercice 33 : Thérapie contre la dépression

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

- 1) S'il y avait 65% de rémission parmi l'ensemble des patients, quelle loi suivrait le nombre de rémissions au sein d'un échantillon de 22 patients choisis au hasard avec remise ? Quel serait le nombre moyen de rémissions au sein d'un tel échantillon ?

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 65% de rémission parmi l'ensemble des patients, alors  $X$  suit la loi  $\mathcal{B}(22; 0,65)$ .

Le nombre moyen de rémissions au sein d'un tel échantillon serait alors  $m(X) = 22 \times 0,65 = 14,3$ .

- 2) Que pensez-vous de l'hypothèse selon laquelle le tirage soit "avec remise" ?

L'énoncé ne l'indique pas, mais on peut penser que déjà rien qu'en France, il y a plusieurs millions de personnes qui souffrent de dépression. En comparaison l'échantillon de 22 personnes est beaucoup plus petit (plus de dix fois plus petit), donc d'après le cours, les probabilités sont celles d'un tirage avec remise.

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

On aurait alors

$$\begin{aligned} \mathbb{P}[X \leq 6] &= \mathbb{P}[X = 0] + \mathbb{P}[X = 1] + \dots + \mathbb{P}[X = 6] \\ &= \binom{22}{0} (1 - 0,65)^{22} + \dots + \binom{22}{6} \times 0,65^6 (1 - 0,65)^{16} \\ &\simeq 0 + 0 + 1,0 \times 10^{-07} + 9,0 \times 10^{-07} + 8,1 \times 10^{-06} + 5,42 \times 10^{-05} + 0,000\,285\,4 \\ &= 0,000\,348\,7 \end{aligned}$$

- 4) Conclusion : vous semble-t-il vraisemblable qu'il y ait, comme l'affirme cette psychologue, 65% de rémissions parmi l'ensemble des patients ?

On a obtenu que s'il y avait bien 65% de rémission parmi l'ensemble

des patients (comme l'affirme ce psychologue), alors il y aurait très peu de chances (environ 0,03% 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 65% (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 10 filles et 19 garçons. On choisit au hasard 5 élèves distincts dans cette classe.

1. Quel est le nombre de choix possibles ?

On choisit 5 personnes parmi l'ensemble des  $10 + 19 = 29$  élèves. Le nombre de choix possibles est donc

$$\begin{aligned} \binom{29}{5} &= \frac{(29)!}{((5)! \times ((24)!)} \\ &= \frac{(29) \times (28) \times \dots \times (1)}{(5 \times 4 \times 3 \times 2) \times ((24) \times (23) \times \dots \times (1))} \\ &= ((29) \times (28) \times (27) \times (26) \times (25)) \times \frac{1}{5 \times 4 \times 3 \times 2} \\ &= \frac{(29) \times (28) \times (27) \times (26) \times (5) \times (5)}{(5) \times (4) \times (3) \times (2)} \\ &= \frac{(29) \times (28) \times (27) \times (26) \times (5)}{(4) \times (3) \times (2)} \\ &= \frac{(29) \times (4) \times (7) \times (27) \times (26) \times (5)}{(4) \times (3) \times (2)} \\ &= \frac{(29) \times (7) \times (27) \times (26) \times (5)}{(3) \times (2)} \\ &= \frac{(29) \times (7) \times (3) \times (9) \times (26) \times (5)}{(3) \times (2)} \\ &= \frac{(29) \times (7) \times (9) \times (26) \times (5)}{2} \\ &= (29) \times (7) \times (9) \times (13) \times (5) \\ &= 118755 \end{aligned}$$

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

Ces choix correspondent à choisir 5 élèves parmi les 19 garçons. Le nombre de choix correspondant est donc

$$\begin{aligned} \binom{19}{5} &= \frac{(19)!}{((5)! \times ((14)!)} \\ &= \frac{(19) \times (18) \times \dots \times (1)}{(5 \times 4 \times 3 \times 2) \times ((14) \times (13) \times \dots \times (1))} \\ &= ((19) \times (18) \times (17) \times (16) \times (15)) \times \frac{1}{5 \times 4 \times 3 \times 2} \\ &= \frac{(19) \times (18) \times (17) \times (16) \times (5) \times (3)}{(5) \times (4) \times (3) \times (2)} \\ &= \frac{(19) \times (18) \times (17) \times (16)}{(4) \times (2)} \\ &= \frac{(19) \times (18) \times (17) \times (4) \times (4)}{(4) \times (2)} \\ &= \frac{(19) \times (18) \times (17) \times (4)}{2} \\ &= (19) \times (2) \times (9) \times (17) \times (4) \\ &= (19) \times (9) \times (17) \times (4) \\ &= 11628 \end{aligned}$$

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

Ces choix correspondent à choisir 5 élèves parmi les 10 filles. Le nombre de choix correspondant est donc

$$\begin{aligned} \binom{10}{5} &= \frac{(10)!}{((5)! \times ((5)!)} \\ &= \frac{(10) \times (9) \times \dots \times (1)}{(5 \times 4 \times 3 \times 2) \times (5 \times 4 \times 3 \times 2)} \\ &= ((10) \times (9) \times (8) \times (7) \times (6)) \times \frac{1}{5 \times 4 \times 3 \times 2} \\ &= \frac{(5) \times (2) \times (9) \times (8) \times (7) \times (6)}{(9) \times (8) \times (7) \times (6)} \\ &= \frac{(4) \times (3)}{(9) \times (4) \times (2) \times (7) \times (6)} \\ &= \frac{(4) \times (3)}{(9) \times (2) \times (7) \times (6)} \\ &= \frac{(3) \times (3) \times (2) \times (7) \times (6)}{3} \\ &= (3) \times (2) \times (7) \times (6) \\ &= 252 \end{aligned}$$

Cela correspond donc à la probabilité  $\frac{252}{118755} \simeq 0,002122$ .

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 correspondants sont formés d'une part d'un groupe de 4 filles (pour lequel il y a  $\binom{10}{4}$  choix possibles), et pour chaque choix de ce groupe de filles, on peut y ajouter n'importe lequel des garçons (pour lequel il y a  $\binom{19}{1}$  choix possibles). En conséquence, le nombre de cas qui correspondent est

$$\begin{aligned} \binom{10}{4} \times \binom{19}{1} &= \frac{(10)!}{((4)!) \times ((6)!) } \times \frac{(19)!}{((1)!) \times ((18)!) } \\ &= \frac{(10) \times (9) \times \dots \times (1)}{(4 \times 3 \times 2) \times ((6) \times (5) \times \dots \times (1))} \times \frac{(19) \times (18) \times \dots \times (1)}{(18) \times (17) \times \dots \times (1)} \\ &= ((10) \times (9) \times (8) \times (7)) \times \frac{1}{4 \times 3 \times 2} \times 19 \\ &= \frac{(10) \times (9) \times (4) \times (2) \times (7)}{(4) \times (3) \times (2)} \times 19 \\ &= \frac{(10) \times (9) \times (7)}{(4) \times (3) \times (2)} \times 19 \\ &= \frac{(10) \times (3) \times (3) \times (7)}{3} \times 19 \\ &= (10) \times (3) \times (7) \times 19 \\ &= 210 \times 19 \\ &= 3990 \end{aligned}$$

D'où la probabilité  $\frac{3990}{118755} \simeq 0,0336$ .

5. Quelle est donc la probabilité d'avoir choisi au plus un garçon ?  
 On additionne, d'une part le nombre de choix comportant 0 garçons (calculé à la question 3) et d'autre part le nombre de choix comportant exactement 1 garçon (calculé à la question 4) : on obtient  $252 + 3990 = 4242$ . La probabilité d'avoir choisi au plus un garçon est donc  $\frac{4242}{118755} \simeq 0,03572$ .

### Chapitre 4 : Loi Normale

#### Exercice 35 : Loi normale centrée réduite

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

- Pour cette question, on considère les nombres  $z_1 = 1,04$ ,  $z_2 = 1,83$  et  $z_3 = 0,144$ .

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

#### Exercice 36 : Images mentales

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

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

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

**Exercice 37 : Désirabilité sociale**

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

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

1. Calculez la proportion théorique des scores supérieurs à 10,64.
2. Trouvez une valeur approchée du nombre  $a$  tel que  $\mathbb{P}[X < a] = 0,95$ .
3. Déterminez 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 15%. On choisit au hasard (avec remise) un échantillon de  $n$  étudiants et on désigne par  $S_n$  le nombre d'échecs au sein de l'échantillon.

1. On prend  $n = 17$ . Calculez  $p_k = \mathbb{P}[S_{17} \leq k]$  pour  $k = 0, 1, 2$  et  $3$ .

On note tout d'abord que  $S_{17} \sim \mathcal{B}(17; 0,15)$ .

$$\mathbb{P}[S_{17} = 0] = \binom{17}{0} (1 - 0,15)^{17} \simeq 0,063 11$$

$$\begin{aligned} \mathbb{P}[0 \leq S_{17} \leq 1] &= \mathbb{P}[S_{17} = 0] + \mathbb{P}[S_{17} = 1] \\ &= \binom{17}{0} (1 - 0,15)^{17} + \binom{17}{1} \times 0,15 (1 - 0,15)^{16} \\ &\simeq 0,063 1 + 0,189 3 \\ &= 0,252 4 \end{aligned}$$

$$\begin{aligned} \mathbb{P}[0 \leq S_{17} \leq 2] &= \mathbb{P}[S_{17} = 0] + \mathbb{P}[S_{17} = 1] + \mathbb{P}[S_{17} = 2] \\ &= \binom{17}{0} (1 - 0,15)^{17} + \binom{17}{1} \times 0,15 (1 - 0,15)^{16} + \binom{17}{2} \times 0,15^2 (1 - 0,15)^{15} \\ &\simeq 0,063 1 + 0,189 3 + 0,267 3 \\ &= 0,519 7 \end{aligned}$$

$$\begin{aligned} \mathbb{P}[0 \leq S_{17} \leq 3] &= \mathbb{P}[S_{17} = 0] + \mathbb{P}[S_{17} = 1] + \mathbb{P}[S_{17} = 2] + \mathbb{P}[S_{17} = 3] \\ &= \binom{17}{0} (1 - 0,15)^{17} + \dots + \binom{17}{3} \times 0,15^3 (1 - 0,15)^{14} \\ &\simeq 0,063 1 + 0,189 3 + 0,267 3 + 0,235 9 \\ &= 0,755 6 \end{aligned}$$

2. Pour  $n = 100$ , on souhaite calculer  $\mathbb{P}[10 \leq S_{100} \leq 20]$  :

- a) Calculez  $\mathbb{P}[10 \leq S_{100} \leq 20]$  avec la calculatrice.

On a  $S_{100} \sim \mathcal{B}(100; 0,15)$ , Ainsi, la calculatrice donne

$$\mathbb{P}[S_{100} \leq 20] \simeq 0,9337 \text{ et } \mathbb{P}[S_{100} \leq 9] \simeq 0,0551.$$

$$\text{Donc } \mathbb{P}[10 \leq S_{100} \leq 20] = \mathbb{P}[S_{100} \leq 20] - \mathbb{P}[S_{100} \leq 9] \simeq 0,933 7 - 0,055 1 = 0,878 6.$$

- b) Justifiez qu'on peut approcher la loi de  $S_{100}$  par une loi normale. Calculez  $\mathbb{P}[10 \leq S_{100} \leq 20]$  de manière approchée. (On fera le calcul avec et sans correction de continuité, pour juger de l'intérêt de cette correction.)

$$\begin{aligned} \text{On observe que } np(1 - p) &= 100 \times 0,15 (1 - 0,15) = 12,75 > 10, \text{ donc on peut approcher} \\ \mathcal{B}(100 ; 0,15) &\approx \mathcal{N}(100 \times 0,15 ; \sqrt{100 \times 0,15 (1 - 0,15)}) , \\ &\simeq \mathcal{N}(15 ; 3,571) \end{aligned}$$

mais il est nécessaire de faire de correction de continuité.

$$\mathbb{P}[10 \leq S_{100} \leq 20] = \mathbb{P}[9,5 \leq S_{100} \leq 20,5] \simeq 0,876$$

Sans la correction de continuité, on aurait obtenu

$$\mathbb{P}[10 \leq S_{100} \leq 20] \simeq 0,838 5, \text{ ce qui est bien moins proche du}$$

résultat obtenu pour la loi binomiale avec la calculatrice.

3. On considère désormais un échantillon de taille  $n = 27\,000$ .

(a) Par quelle loi normale peut-on approcher la variable  $S_{27\,000}$  ? Calculez ainsi (une valeur approchée de) la probabilité  $\mathbb{P}[3\,947 \leq S_{27\,000} \leq 4\,162]$ .

On observe que  $np(1-p) = 27\,000 \times 0,15(1-0,15) = 3\,442,5 > 1\,000$ , donc on peut approcher  $\mathcal{B}(27\,000; 0,15) \approx \mathcal{N}(27\,000 \times 0,15; \sqrt{27\,000 \times 0,15(1-0,15)}) \simeq \mathcal{N}(4\,050; 58,673)$

, et il n’y a pas besoin de faire de correction de continuité .

$$\mathbb{P}[3\,947 \leq S_{27\,000} \leq 4\,162] \simeq 0,9323$$

(b) Par quelle loi normale peut-on approcher la proportion  $P_{27\,000} = \frac{S_{27\,000}}{27\,000}$  ? Calculez ainsi (une valeur approchée de) la probabilité  $\mathbb{P}[0,1462 \leq P_{27\,000} \leq 0,1541]$ .

On observe que  $np(1-p) = 27\,000 \times 0,15(1-0,15) = 3\,442,5 > 1\,000$ , donc on peut approcher la proportion par  $\mathcal{N}(0,15; \sqrt{\frac{0,15(1-0,15)}{27\,000}}) \simeq \mathcal{N}(0,15; 0,002173)$ , et il n’y a pas besoin de faire de correction de continuité .

$$\mathbb{P}[0,1462 \leq P_{27\,000} \leq 0,1541] \simeq 0,9302$$

(c) Trouvez une valeur approchée de  $a$  telle que  $\mathbb{P}[S_{27\,000} < a] = 0,03$ .

On utilise la même approximation par une loi normale que dans la question précédente.

Pour résoudre l’équation  $\mathbb{P}[S_{27\,000} \leq a] \simeq 0,03$ , on note tout d’abord que

$$\mathbb{P}[S_{27\,000} \leq a] = \mathbb{P}[Z \leq \frac{a-4\,050}{58,673}] = F\left(\frac{a-4\,050}{58,673}\right)$$

d’où l’équation  $F\left(\frac{a-4\,050}{58,673}\right) = 0,03$ , c’est à dire  $F(z) = 0,03$ , où on note  $z = \frac{a-4\,050}{58,673}$ .

Comme  $F(z) \simeq 0,03 < 0,5$ , on sait que  $z < 0$ .

On commence donc par résoudre l’équation  $F(|z|) \simeq 0,97$  :  $F(1,88) \simeq 0,9699$  et  $F(1,89) \simeq 0,9706$ .

Donc  $1,88 \leq |z| \leq 1,89$ , que l’on peut par exemple arrondir

à  $|z| \simeq 1,88$ .

D’où  $z \simeq -1,88$ .

D’où  $\frac{a-4\,050}{58,673} = z = (-1,88)$ , qui donne  $a \simeq z \times 58,673 + 4\,050 \simeq -1,88 \times 58,673 + 4\,050 \simeq 3\,939,7$ .

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

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

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

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

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

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

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

1. Déterminez la probabilité (« fréquence théorique ») qu’aurait chaque classe sous l’hypothèse de loi normale ( $\mathcal{H}$ ).

Si  $X$  suivait la loi  $\mathcal{N}(25; 5)$ , on obtiendrait 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. Justifiez que, sous l'hypothèse ( $\mathcal{H}$ ), la proportion *théorique* d'obésité (c'est à dire  $\mathbb{P}[X \geq 30]$ , en regroupant les obésités modérées, sévères et morbides) serait d'environ 15,9%.

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. Quelle est, en fait, la proportion des sujets obèses observée dans l'échantillon épidémiologique de l'INSERM ?

La proportion d'obésité observée dans l'échantillon de l'INSERM 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 : pour cela, on va évaluer si la proportion  $\mathbb{P}_r[X \geq 30]$  observée dans l'échantillon de l'INSERM s'accorde avec la probabilité  $\mathbb{P}[X \geq 30]$

obtenue sous l'hypothèse ( $\mathcal{H}$ ).

(a) Si on suppose qu'un.e français.e choisi.e au hasard a une probabilité de 15,9% d'être en situation d'obésité, quelle est la loi du nombre de personnes en situation d'obésité dans un échantillon de 25 714 français ? Le fait que l'échantillon soit *avec* ou *sans remise* a-t-il la moindre importance ?

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 ?

$$np(1 - p) = 25\,714 \times 0,159(1 - 0,159) \simeq 3\,438,45 > 1\,000, \text{ donc on peut approcher } \mathcal{B}(25714; 0,159) \approx \mathcal{N}(25\,714 \times 0,159; \sqrt{25\,714 \times 0,159(1 - 0,159)}) \simeq \mathcal{N}(4088,526; 58,638)$$

, et il n'y a pas besoin de faire de correction de continuité .

(c) On considère toujours un échantillon aléatoire de 25 714 français-es, mais cette fois-ci on considère comme variable aléatoire la **proportion** de personnes, au sein de l'échantillon, qui sont en situation d'obésité. Sa loi peut-elle aussi être approximée par une loi normale, et si oui laquelle ?

$$np(1 - p) = 25\,714 \times 0,159(1 - 0,159) \simeq 3\,438,45 > 1\,000, \text{ donc on peut approcher la proportion par } \mathcal{N}(0,159; \sqrt{\frac{0,159(1-0,159)}{25\,714}}) \simeq \mathcal{N}(0,159; 0,00228), \text{ et il n'y a pas besoin de faire de correction de continuité .}$$

(d) En utilisant cette approximation par une loi normale, calculer la probabilité qu'un échantillon aléatoire de 25 714 français-es

contienne au maximum 15% de personnes en situation d'obésité.

Comme cette proportion  $P$  suit la loi normale trouvée dans la question (c), on doit avoir  $\mathbb{P}[P \leq 0,15] \simeq 3,951 \times 10^{-05}$

- (e) En utilisant le résultat de la question (d) et en faisant appel à l'étude épidémiologique de l'INSERM, que pouvez-vous en conclure au sujet de l'hypothèse ( $\mathcal{H}$ ) ?

Si  $X$  avait vraiment suivi la loi  $\mathcal{N}(25; 5)$ , alors il n'y aurait eu quasiment aucune chance qu'un échantillon aléatoire de 25714 français ait au plus 15% de personnes en situation d'obésité. Or l'échantillon aléatoire de l'étude épidémiologique de l'INSERM a donné 15% de personnes en situation d'obésité. Donc, en toute vraisemblance, l'IMC des français ne suit pas la loi  $\mathcal{N}(25; 5)$ .

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

1. Déterminez les probabilités des événements suivants pour une variable aléatoire  $Z$  suivant la loi normale centrée réduite :

- (a)  $Z$  est supérieure ou égale à 0,6 ;  
 $\mathbb{P}[Z \geq 0,6] = 1 - F(0,6) = 0,2743$
- (b)  $Z$  est comprise entre  $-1$  et  $1$  ;  
 $\mathbb{P}[-1 \leq Z \leq 1] = F(1) - F(-1) = 0,6826$
- (c)  $Z$  est comprise entre  $-2$  et  $2$  ;  
 $\mathbb{P}[-2 \leq Z \leq 2] = F(2) - F(-2) = 0,9544$
- (d)  $Z$  est comprise entre  $-3$  et  $3$ .  
 $\mathbb{P}[-3 \leq Z \leq 3] = F(3) - F(-3) = 0,9974$

2. Déterminez les probabilités des événements suivants pour une variable aléatoire  $X$  suivant une loi normale de moyenne  $\mu = 108$  et d'écart-type  $\sigma = 11$  (c'est-à-dire  $X \sim \mathcal{N}(108; 11)$ ) :

- (a)  $X$  est supérieure ou égale à  $\mu + 0,6\sigma$  ;  
 On note tout d'abord que  $\mu + 0,6\sigma = 108 + 0,6 \times 11 = 114,6$ .  
 $\mathbb{P}[X \geq 114,6] = \mathbb{P}\left[\frac{X-108}{11} \geq \frac{114,6-108}{11}\right]$   
 $\simeq \mathbb{P}[Z \geq 0,6] = 1 - F(0,6)$   
 $= 0,2743$

- (b)  $X$  est comprise entre  $\mu - \sigma$  et  $\mu + \sigma$  ;  
 On note tout d'abord que  $\mu - \sigma = 108 - 11 = 97$  et  $\mu + \sigma = 108 + 11 = 119$ .  
 $\mathbb{P}[97 \leq X \leq 119] = \mathbb{P}\left[\frac{97-108}{11} \leq \frac{X-108}{11} \leq \frac{119-108}{11}\right]$   
 $= \mathbb{P}[-1 \leq Z \leq 1] = F(1) - F(-1)$   
 $= 0,6826$

- (c)  $X$  est comprise entre  $\mu - 2\sigma$  et  $\mu + 2\sigma$  ;  
 On note tout d'abord que  $\mu - 2\sigma = 108 - 2 \times 11 = 86$  et  $\mu + 2\sigma = 108 + 2 \times 11 = 130$ .  
 $\mathbb{P}[86 \leq X \leq 130] = \mathbb{P}\left[\frac{86-108}{11} \leq \frac{X-108}{11} \leq \frac{130-108}{11}\right]$   
 $= \mathbb{P}[-2 \leq Z \leq 2] = F(2) - F(-2)$   
 $= 0,9544$

- (d)  $X$  est comprise entre  $\mu - 3\sigma$  et  $\mu + 3\sigma$ .  
 On note tout d'abord que  $\mu - 3\sigma = 108 - 3 \times 11 = 75$  et  $\mu + 3\sigma = 108 + 3 \times 11 = 141$ .  
 $\mathbb{P}[75 \leq X \leq 141] = \mathbb{P}\left[\frac{75-108}{11} \leq \frac{X-108}{11} \leq \frac{141-108}{11}\right]$   
 $= \mathbb{P}[-3 \leq Z \leq 3] = F(3) - F(-3)$   
 $= 0,9974$

3. Reprenez la question 2 pour une variable aléatoire  $X$  suivant la loi normale  $\mathcal{N}(13,9; 2,8)$ . Qu'observez-vous ?

- (a) On note tout d'abord que  $\mu + 0,6\sigma = 13,9 + 0,6 \times 2,8 = 15,58$ .  
 $\mathbb{P}[X \geq 15,58] = \mathbb{P}\left[\frac{X-13,9}{2,8} \geq \frac{15,58-13,9}{2,8}\right]$   
 $\simeq \mathbb{P}[Z \geq 0,6] = 1 - F(0,6)$   
 $= 0,2743$

- (b) On note tout d'abord que  $\mu - \sigma = 13,9 - 2,8 = 11,1$  et  $\mu + \sigma = 13,9 + 2,8 = 16,7$ .

$$\begin{aligned}\mathbb{P}[11,1 \leq X \leq 16,7] &= \mathbb{P}\left[\frac{11,1-13,9}{2,8} \leq \frac{X-13,9}{2,8} \leq \frac{16,7-13,9}{2,8}\right] \\ &\simeq \mathbb{P}[-1 \leq Z \leq 1] = F(1) - F(-1) \\ &= 0,6826\end{aligned}$$

- (c) On note tout d'abord que  $\mu - 2\sigma = 13,9 - 2 \times 2,8 = 8,3$  et  $\mu + 2\sigma = 13,9 + 2 \times 2,8 = 19,5$ .

$$\begin{aligned}\mathbb{P}[8,3 \leq X \leq 19,5] &= \mathbb{P}\left[\frac{8,3-13,9}{2,8} \leq \frac{X-13,9}{2,8} \leq \frac{19,5-13,9}{2,8}\right] \\ &= \mathbb{P}[-2 \leq Z \leq 2] = F(2) - F(-2) \\ &= 0,9544\end{aligned}$$

- (d) On note tout d'abord que  $\mu - 3\sigma = 13,9 - 3 \times 2,8 = 5,5$  et  $\mu + 3\sigma = 13,9 + 3 \times 2,8 = 22,3$ .

$$\begin{aligned}\mathbb{P}[5,5 \leq X \leq 22,3] &= \mathbb{P}\left[\frac{5,5-13,9}{2,8} \leq \frac{X-13,9}{2,8} \leq \frac{22,3-13,9}{2,8}\right] \\ &= \mathbb{P}[-3 \leq Z \leq 3] = F(3) - F(-3) \\ &= 0,9974\end{aligned}$$

On observe au travers de nos réponses aux questions 1, 2 et 3 que les résultats sont indépendants de  $\sigma$  et  $\mu$ . Cela apparaît d'ailleurs de manière évidente si ces calculs ont été effectués en se ramenant à chaque fois à la loi normale centrée réduite (afin d'utiliser la table du formulaire), plutôt qu'en utilisant la fonction dédiée de la calculatrice.

### Exercice 41 : Diagnostic de la dépression

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

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

On note  $X$  le score obtenu à ce test, et on a  $X \sim \mathcal{N}(50; 10)$ .

$$\begin{aligned}\mathbb{P}[X \geq 70] &= \mathbb{P}\left[\frac{X-50}{10} \geq \frac{70-50}{10}\right] \\ &= \mathbb{P}[Z \geq 2] = 1 - F(2) \\ &= 0,0228\end{aligned}$$

Donc parmi un échantillon de 10 000 personnes, on s'attend à trouver environ  $10\,000 \times 0,0228 = 228$  personnes pathologiquement dépressives.

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

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

- Lorsque  $n = 25$ , précisez la loi de probabilité de  $X$ , puis calculez  $\mathbb{P}[16 \leq X \leq 18]$ .

On a  $X \sim \mathcal{B}(25; 0,7)$ .

Donc on trouve  $\mathbb{P}[16 \leq X \leq 18] = \mathbb{P}[X \leq 18] - \mathbb{P}[X \leq 15] \simeq 0,6593 - 0,1894 = 0,4699$  en utilisant la calculatrice.

- On considère désormais que  $n = 250$ .

(a)

- (b) Justifiez qu'on peut approcher la loi de  $X$  par une loi normale, préciser laquelle, et indiquer s'il est nécessaire de faire une correction de continuité.

$$X \sim \mathcal{B}(250; 0,7).$$

$$np(1-p) = 250 \times 0,7(1-0,7) = 52,5 > 10, \text{ donc on peut approcher } \mathcal{B}(250; 0,7) \approx \mathcal{N}(250 \times 0,7; \sqrt{250 \times 0,7(1-0,7)}), \\ \simeq \mathcal{N}(175; 7,246)$$

mais il est nécessaire de faire de correction de continuité.

- (c) Déduisez de la question précédente des valeurs approchées de  $\mathbb{P}[X \geq 165]$  et  $\mathbb{P}[160 \leq X \leq 180]$ .

Avec cette approximation par une loi normale et en faisant la correction de continuité, on obtient :

$$\mathbb{P}[X \geq 165] = \mathbb{P}[X \geq 164,5] \simeq 0,9263$$

$$\mathbb{P}[160 \leq X \leq 180] = \mathbb{P}[159,5 \leq X \leq 180,5] \simeq 0,76$$

### 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  $X \sim \mathcal{B}(100; 0,5)$ .

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

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. 200 fois.

on a  $X \sim \mathcal{B}(200; 0,5)$ .

$$\text{On observe que } np(1-p) = 200 \times 0,5(1-0,5) = 50 > 10, \text{ donc on peut approcher } \mathcal{B}(200; 0,5) \approx \mathcal{N}(200 \times 0,5; \sqrt{200 \times 0,5(1-0,5)}), \\ \simeq \mathcal{N}(100; 7,071)$$

il est nécessaire de faire de correction de continuité.

Avoir une proportion de pile entre 0,48 et 0,52 signifierait avoir  $96 \leq X \leq 104$ . La probabilité demandée vaut donc (avec la correction de continuité) :

$$\mathbb{P}[96 \leq X \leq 104] = \mathbb{P}[95,5 \leq X \leq 104,5] \simeq 0,475$$

3. 1 000 fois.

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

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

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) :

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

4. 10 000 fois.

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

On observe que  $np(1-p) = 10\,000 \times 0,5(1-0,5) = 2\,500 > 1\,000$ , donc on peut approcher la proportion par  $\mathcal{N}(0,5; \sqrt{\frac{0,5(1-0,5)}{10\,000}}) = \mathcal{N}(0,5; 0,005)$ , et il n'y a pas besoin de faire de correction de continuité. La probabilité demandée vaut donc :

$$\mathbb{P}[0,48 \leq P \leq 0,52] \simeq 0,999\,936\,7$$

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

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

1. Quelle est la probabilité pour qu'un neurone émette plus de 129 PA en 1 min 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  $X \sim \mathcal{N}(120; 15)$ . Donc on a  $\mathbb{P}[X > 129] \simeq 0,2743$  (d'après la calculatrice).

2. Quelle est la probabilité pour qu'un neurone émette moins de 90 PA en 1 min s'il n'est pas activé ?

On a  $\mathbb{P}[X < 90] \simeq 0,02275$  (d'après la calculatrice).

3. Quelle est la probabilité pour qu'un neurone émette entre 105 et 147 PA en 1 min s'il n'est pas activé ?

On a  $\mathbb{P}[105 \leq X < 147] \simeq 0,8054$  (d'après la calculatrice).

4. On enregistre l'activité d'un neurone durant une minute, et on observe que ce neurone a émis plus de 190 PA. Selon vous, est-il plus vraisemblable que ce neurone était activé ou au repos lors de l'enregistrement ? Justifiez votre réponse.

On a  $\mathbb{P}[X \geq 190] \simeq 1,531 \times 10^{-06}$  (d'après la calculatrice).

Si ce neurone avait été au repos, il aurait donc été très improbable d'observer plus de 190 PA durant l'enregistrement. Ainsi, très vraisemblablement, le neurone était actif durant l'enregistrement.

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

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

1. Estimez, avec la confiance 96%, la probabilité qu'a cette pièce de tomber sur pile quand on la lance.

On a  $n p_e(1-p_e) = 1400 \times \frac{757}{1400} \left(1 - \frac{757}{1400}\right) \simeq 347,679 > 10$ , donc on peut utiliser la procédure du formulaire pour estimer la proportion  $p$ .

On a  $F(2,054) \simeq 0,98$  d'où  $z_\alpha \simeq 2,054$

d'où  $a_\alpha = z_\alpha \sqrt{\frac{p_e(1-p_e)}{n}} \simeq 0,027$  et  $p_e - a_\alpha \simeq \frac{757}{1400} - 0,027 \simeq 0,5137$  et  $p_e + a_\alpha \simeq \frac{757}{1400} + 0,027 \simeq 0,5677$ .

On estime donc que  $p$  est dans l'intervalle  $[0,514; 0,568]$  avec la confiance  $c = 0,96$ .

2. Conclure : pouvez-vous affirmer (avec la confiance 96%) que cette pièce soit biaisée ?

On peut effectivement affirmer, avec la confiance 96%, que la pièce est biaisée (on a vu qu'elle a au moins 51,4% de chances de faire pile).

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

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

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

1. Donnez une estimation de la proportion réelle  $p$  de guérisons (pour l'ensemble des patients que l'on traiterait avec ce médicament) avec une confiance de 95%.

On a  $np_e(1-p_e) = 500 \times \frac{353}{500} \left(1 - \frac{353}{500}\right) = 103,782 > 10$ , 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  $a_\alpha = z_\alpha \sqrt{\frac{p_e(1-p_e)}{n}} \simeq 0,0399$  et  $p_e - a_\alpha \simeq \frac{353}{500} - 0,0399 = 0,6661$  et  $p_e + a_\alpha \simeq \frac{353}{500} + 0,0399 = 0,7459$ .

On estime donc que  $p$  est dans l'intervalle  $[0,6661; 0,7459]$  avec la confiance  $c = 0,95$ .

2. Que peut-on conclure quant à l'affirmation du fabricant ?

Donc, avec la confiance 95%, on peut affirmer que le taux de guérison est bien supérieur à ce qu'affirme le fabricant : il est d'au moins 66%.

3. On veut réduire la marge de l'estimation à 1% tout en assurant une confiance de 98%. Quelle doit alors être la taille minimale de l'échantillon à considérer ?

Pour avoir une précision 0,01 avec la confiance 0,98, il faut  $n > z_\alpha^2 \frac{p_e(1-p_e)}{h^2} \simeq 2,326^2 \frac{0,706(1-0,706)}{0,01^2} \simeq 11229$

#### Exercice 47 : Satisfaction de clients

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

1. On désigne par  $p_1$  et  $p_2$  les proportions de satisfaits sur l'ensemble de la clientèle de l'entreprise, avant et après l'amélioration apportée. Donnez une estimation de  $p_1$  et  $p_2$  par intervalle de confiance en prenant comme confiance  $c = 0,9$ . Peut-on dire que la proportion de satisfaits pour l'ensemble de la clientèle a augmenté ?

**Estimation de  $p_1$  :** On a  $np_e(1-p_e) = 1245 \times \frac{545}{1245} \left(1 - \frac{545}{1245}\right) \simeq 306,426 > 10$ , donc on peut utiliser la procédure du formulaire pour estimer la proportion  $p_1$ .

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

d'où  $a_\alpha = z_\alpha \sqrt{\frac{p_e(1-p_e)}{n}} \simeq 0,0231$  et  $p_e - a_\alpha \simeq \frac{545}{1245} - 0,0231 \simeq 0,4147$  et  $p_e + a_\alpha \simeq \frac{545}{1245} + 0,0231 \simeq 0,4609$ .

On estime donc que  $p_1$  est dans l'intervalle  $[0,4147; 0,4609]$  avec la confiance  $c = 0,9$ .

**Estimation de  $p_2$  :** On a  $np_e(1-p_e) = 1023 \times \frac{473}{1023} \left(1 - \frac{473}{1023}\right) \simeq 254,301 > 10$ , donc on peut utiliser la procédure du formulaire pour estimer la proportion  $p_2$ .

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

d'où  $a_\alpha = z_\alpha \sqrt{\frac{p_e(1-p_e)}{n}} \simeq 0,0256$  et  $p_e - a_\alpha \simeq \frac{473}{1023} - 0,0256 \simeq 0,4368$  et  $p_e + a_\alpha \simeq \frac{473}{1023} + 0,0256 \simeq 0,488$ .

On estime donc que  $p_2$  est dans l'intervalle  $[0,4368; 0,488]$  avec la confiance  $c = 0,9$ .

**Conclusion :** Comme les intervalles de confiance pour  $p_1$  et  $p_2$  se chevauchent, on ne peut décider laquelle de ces deux proportions réelles est plus grande que l'autre. Il n'y a donc pas de conclusion à tirer pour l'entreprise au terme de ces deux sondages. (Par exemple, il serait tout à fait possible d'avoir  $p_1 = 0,46$  et  $p_2 = 0,44$  donc  $p_1 > p_2$ , mais il serait possible aussi d'avoir  $p_1 = 0,44$  et  $p_2 = 0,46$  donc  $p_1 < p_2$ .)

2. Lors du deuxième sondage, quelle aurait dû être la taille de l'échantillon pour avoir une estimation de  $p_2$  à 1% près avec une

confiance de 0,99 ?

Pour avoir une précision 0,01 avec la confiance 0,99, il faut  $n > z_{\alpha}^2 \frac{p_e(1-p_e)}{h^2} \simeq 2,576^2 \frac{0,462(1-0,462)}{0,01^2} \simeq 16493$

**Exercice 48 : Résistance à la persuasion**

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

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

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

- a) Calculer la moyenne expérimentale  $m_e(X)$  et l’écart type expérimental  $s_e(X)$  de l’échantillon.

moyenne :  $m_e(X) = \frac{\sum c_i n_i}{n} = \frac{7 \times 2 + 9 \times 6 + 11 \times 21 + \dots + 17 \times 6}{90} = \frac{1156}{90} \simeq 12,844$

$m_e(X^2) = \frac{\sum c_i^2 n_i}{n} = \frac{7^2 \times 2 + 9^2 \times 6 + 11^2 \times 21 + \dots + 17^2 \times 6}{90} = \frac{15274}{90}$

$Var_e(X) = m_e(X^2) - m_e(X)^2 = \frac{15274}{90} - \left(\frac{1156}{90}\right)^2 \simeq 4,731$

Écart-type :  $s_e(X) = \sqrt{Var_e(X)} \simeq 2,175$

- 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  $c = 0,95$ .

Comme  $n = 90 > 30$ , on cherche  $z_{\alpha}$  tel que  $F(z_{\alpha}) = \frac{0,95+1}{2} = 0,975$

On a  $F(1,96) \simeq 0,975$  d’où  $z_{\alpha} \simeq 1,96$  d’où  $a_{\alpha} = z_{\alpha} \frac{s_e}{\sqrt{n-1}} \simeq 0,4519$

On estime donc que  $\mu(X)$  est dans l’intervalle  $[12,844 - 0,4519 ; 12,844 + 0,4519] \simeq [12,39 ; 13,3]$ , 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.

Toutes les valeurs de cet intervalle de confiance sont plus petites que 13,7, donc on peut déduire avec la confiance 95% que la note moyenne a baissé.

- 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 ?

Pour avoir une précision 0,5 avec la confiance 0,99, il faut  $n > z_{\alpha}^2 \frac{(s_e)^2}{h^2} \simeq 2,576^2 \frac{2,175^2}{0,5^2} \simeq 126$

2. On considère désormais un échantillon de 26 étudiants à qui on demande, pendant qu’ils écoutent l’argumentaire, de réaliser des « opérations » arithmétiques (réciter mentalement la table de multiplication par 8). La moyenne et l’écart type expérimentaux sont alors  $m_e = 9,24$  et  $s_e = 4,5$ .

- a) Avec une confiance de 0,95 déterminer la note moyenne de l’ensemble des étudiants s’ils écoutaient l’argumentaire en récitant des tables de multiplication. (On pourra supposer que les notes suivent une loi normale.)

Comme  $n = 26 \leq 30$ , on cherche  $t_\alpha$  à partir de la table inverse de Student avec  $p = \frac{\alpha}{2} = 0,025$  et  $n - 1 = 25$  degrés de liberté (ddl)

On lit  $t_\alpha \simeq 2,0595$  d'où  $a_\alpha = t_\alpha \frac{s_e}{\sqrt{n-1}} \simeq 1,8535$ .

On estime donc que  $\mu(Y)$  est dans l'intervalle  $[9,24 - 1,8535; 9,24 + 1,8535] \simeq [7,39; 11,09]$  avec la confiance  $c = 0,95$

- 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 ?

Au vu des intervalles que l'on a calculés (et qui ne se chevauchent pas), on peut affirmer avec la confiance 95% que  $\mu(Y) < \mu(X)$ , c'est à dire qu'en moyenne, l'argumentation a plus diminué la note (donc été plus efficace) lorsque les sujets devaient réciter les tables de multiplication.

C'est aussi ce que prédisait l'hypothèse théorique, les observations sont parfaitement compatibles avec celle-ci.

- 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%.

On lit dans la table du  $\chi^2$  à 25 degrés de liberté la valeur  $x_1 = 13,12$  correspondant à  $q = 0,025$  et la valeur  $x_2 = 40,65$  correspondant à  $p = 0,025$

On obtient alors  $s_e \sqrt{\frac{n}{x_2}} \simeq 4,5 \sqrt{\frac{26}{40,65}} \simeq 3,6$  et  $s_e \sqrt{\frac{n}{x_1}} \simeq 4,5 \sqrt{\frac{26}{13,12}} \simeq 6,33$

On estime donc que  $\sigma(Y)$  est dans l'intervalle  $[3,6; 6,33]$  avec la confiance  $c = 0,95$ .

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

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

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

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

- Calculer les moyennes et les écarts types des deux échantillons.  
On note  $X$  les notes obtenues en affirmant qu'on teste leurs aptitudes innées, et  $Y$  les notes si on affirme qu'on veut les occuper.  
 $m(X) \simeq 28,467$   $s(X) \simeq 5,976$ .  
 $m(Y) \simeq 18,688$   $s(Y) \simeq 4,908$ .
- 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%.

- a) Si on leur dit qu'on teste une **aptitude innée** :
- Comme  $n = 15 \leq 30$ , on cherche  $t_\alpha$  à partir de la table inverse de Student avec  $p = \frac{\alpha}{2} = 0,025$  et  $n - 1 = 14$  degrés de liberté (ddl)
- On lit  $t_\alpha \simeq 2,1448$  d'où  $a_\alpha = t_\alpha \frac{s_e}{\sqrt{n-1}} \simeq 3,4256$ .
- On estime donc que  $\mu(X)$  est dans l'intervalle  $[28,467 - 3,4256; 28,467 + 3,4256] \simeq [25,04; 31,89]$  avec la confiance  $c = 0,95$

On lit dans la table du  $\chi^2$  à 14 degrés de liberté la valeur

$x_1 = 5,629$  correspondant à  $q = 0,025$  et la valeur  $x_2 = 26,12$  correspondant à  $p = 0,025$

On obtient alors  $s_e \sqrt{\frac{n}{x_2}} \simeq 5,976 \sqrt{\frac{15}{26,12}} \simeq 4,53$  et  $s_e \sqrt{\frac{n}{x_1}} \simeq 5,976 \sqrt{\frac{15}{5,629}} \simeq 9,76$

On estime donc que  $\sigma(X)$  est dans l'intervalle  $[4,53; 9,76]$  avec la confiance  $c = 0,95$ .

b) Si on leur dit juste qu'on les **occupe** :

Comme  $n = 16 \leq 30$ , on cherche  $t_\alpha$  à partir de la table inverse de Student avec  $p = \frac{\alpha}{2} = 0,025$  et  $n - 1 = 15$  degrés de liberté (ddl)

On lit  $t_\alpha \simeq 2,1314$  d'où  $a_\alpha = t_\alpha \frac{s_e}{\sqrt{n-1}} \simeq 2,701$ .

On estime donc que  $\mu(Y)$  est dans l'intervalle  $[18,688 - 2,701; 18,688 + 2,701] \simeq [15,99; 21,39]$  avec la confiance  $c = 0,95$

On lit dans la table du  $\chi^2$  à 15 degrés de liberté la valeur  $x_1 = 6,262$  correspondant à  $q = 0,025$  et la valeur  $x_2 = 27,49$  correspondant à  $p = 0,025$

On obtient alors  $s_e \sqrt{\frac{n}{x_2}} \simeq 4,908 \sqrt{\frac{16}{27,49}} \simeq 3,74$  et  $s_e \sqrt{\frac{n}{x_1}} \simeq 4,908 \sqrt{\frac{16}{6,262}} \simeq 7,85$

On estime donc que  $\sigma(X)$  est dans l'intervalle  $[3,74; 7,85]$  avec la confiance  $c = 0,95$ .

3. Que peut-on conclure au vu de ces résultats ?

Ces intervalles montrent que  $\mu(X) > \mu(Y)$  (en effet  $\mu(X) > 25$  alors que  $\mu(Y) < 22$ ). Cela signifie qu'avec la confiance 95%, on peut affirmer qu'en moyenne les enfants y arrivent mieux si on leur dit qu'on veut les tester que si on leur dit qu'on veut les occuper.

### Exercice 50 : Taux de cholestérol

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

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

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

1. Donner des estimations des écarts types des deux populations. *On demande de déterminer un intervalle de confiance, avec la confiance  $c = 99\%$ .*

**Chez les femmes** On lit dans la table du  $\chi^2$  à 16 degrés de liberté la valeur  $x_1 = 5,142$  correspondant à  $q = 0,005$  et la valeur  $x_2 = 34,27$  correspondant à  $p = 0,005$

On obtient alors  $s_e \sqrt{\frac{n}{x_2}} \simeq 38,3 \sqrt{\frac{17}{34,27}} \simeq 27$  et  $s_e \sqrt{\frac{n}{x_1}} \simeq 38,3 \sqrt{\frac{17}{5,142}} \simeq 69,6$

On estime donc que  $\sigma$  est dans l'intervalle  $[27; 69,6]$  avec la confiance  $c = 0,99$ .

**Chez les hommes** On lit dans la table du  $\chi^2$  à 21 degrés de liberté la valeur  $x_1 = 8,034$  correspondant à  $q = 0,005$  et la valeur  $x_2 = 41,4$  correspondant à  $p = 0,005$

On obtient alors  $s_e \sqrt{\frac{n}{x_2}} \simeq 34,2 \sqrt{\frac{22}{41,4}} \simeq 24,9$  et  $s_e \sqrt{\frac{n}{x_1}} \simeq 34,2 \sqrt{\frac{22}{8,034}} \simeq 56,6$

On estime donc que  $\sigma$  est dans l'intervalle  $[24,9; 56,6]$  avec la confiance  $c = 0,99$ .

2. En utilisant les estimations par intervalle de confiance à 99%, peut-on conclure que les femmes de plus de 45 ans ont en moyenne moins de cholestérol que les hommes de plus de 45 ans ?

On commence par estimer les moyennes :

**Chez les femmes** Comme  $n = 17 \leq 30$ , on cherche  $t_\alpha$  à partir de la table inverse de Student avec  $p = \frac{\alpha}{2} = 0,005$  et  $n - 1 = 16$  degrés de liberté (ddl)

On lit  $t_\alpha \simeq 2,9208$  d'où  $a_\alpha = t_\alpha \frac{s_e}{\sqrt{n-1}} \simeq 27,967$ .

On estime donc que  $\mu$  est dans l'intervalle  $[177,3 - 27,967; 177,3 + 27,967] \simeq [149,33; 205,27]$  avec la confiance  $c = 0,99$

**Chez les hommes** Comme  $n = 22 \leq 30$ , on cherche  $t_\alpha$  à partir de la table inverse de Student avec  $p = \frac{\alpha}{2} = 0,005$  et  $n - 1 = 21$  degrés de liberté (ddl)

On lit  $t_\alpha \simeq 2,8314$  d'où  $a_\alpha = t_\alpha \frac{s_e}{\sqrt{n-1}} \simeq 21,131$ .

On estime donc que  $\mu$  est dans l'intervalle  $[193,8 - 21,131; 193,8 + 21,131] \simeq [172,67; 214,93]$  avec la confiance  $c = 0,99$

Les intervalles se chevauchent, donc on ne peut pas conclure (avec la confiance  $c = 99\%$ ) que les femmes de plus de 45 ans aient en moyenne plus de cholestérol que les hommes de plus de 45 ans.

### Exercice 51 : Épreuve de dictée

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

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

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

- (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

$$\begin{aligned} m(X) &= 15,63 & s(X) &= 1,75. \\ m(Y) &= 12,58 & s(Y) &= 1,39. \end{aligned}$$

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

**Épreuve 1** Comme  $n = 12 \leq 30$ , on cherche  $t_\alpha$  à partir de la table inverse de Student avec  $p = \frac{\alpha}{2} = 0,02$  et  $n - 1 = 11$  degrés de liberté (ddl)

On lit  $t_\alpha \simeq 2,3281$  d'où  $a_\alpha = t_\alpha \frac{s_e}{\sqrt{n-1}} \simeq 1,2284$ .

On estime donc que  $\mu(X)$  est dans l'intervalle  $[15,63 - 1,2284; 15,63 + 1,2284] \simeq [14,4; 16,86]$  avec la confiance  $c = 0,96$

**Épreuve 2** Comme  $n = 12 \leq 30$ , on cherche  $t_\alpha$  à partir de la table inverse de Student avec  $p = \frac{\alpha}{2} = 0,02$  et  $n - 1 = 11$  degrés de liberté (ddl)

On lit  $t_\alpha \simeq 2,3281$  d'où  $a_\alpha = t_\alpha \frac{s_e}{\sqrt{n-1}} \simeq 0,97571$ .

On estime donc que  $\mu(Y)$  est dans l'intervalle  $[12,58 - 0,97571; 12,58 + 0,97571] \simeq [11,6; 13,56]$  avec la confiance  $c = 0,96$

- (c) : Ces intervalles indiquent que  $\mu(X) > \mu(Y)$  (en effet  $\mu(X) > 14$  et  $\mu(Y) < 14$ ). 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 11 degrés de liberté la valeur  $x_1 = 3,816$  correspondant à  $q = 0,025$  et la valeur  $x_2 = 21,92$  correspondant à  $p = 0,025$

On obtient alors  $s_e \sqrt{\frac{n}{x_2}} \simeq 1,75 \sqrt{\frac{12}{21,92}} \simeq 1,29$  et  $s_e \sqrt{\frac{n}{x_1}} \simeq$

$$1,75 \sqrt{\frac{12}{3,816}} \simeq 3,1$$

On estime donc que  $\sigma(X)$  est dans l'intervalle  $[1,29; 3,1]$  avec la confiance  $c = 0,95$ .

**Épreuve 2** On lit dans la table du  $\chi^2$  à 11 degrés de liberté la valeur  $x_1 = 3,816$  correspondant à  $q = 0,025$  et la valeur  $x_2 = 21,92$  correspondant à  $p = 0,025$

On obtient alors  $s_e \sqrt{\frac{n}{x_2}} \simeq 1,39 \sqrt{\frac{12}{21,92}} \simeq 1,03$  et  $s_e \sqrt{\frac{n}{x_1}} \simeq$

$$1,39 \sqrt{\frac{12}{3,816}} \simeq 2,46$$

On estime donc que  $\sigma(Y)$  est dans l'intervalle  $[1,03; 2,46]$  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 7,6% et 8% 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 :  $\mathbb{P}[0,076 \leq P \leq 0,08] \simeq 0,804$

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

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

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

Remarques : ces 8,107% de l'échantillon correspondent à  $30000 \times 0,08107 \simeq 2432$  personnes.

On a  $np_e(1-p_e) = 30000 \times \frac{2432}{30000} \left(1 - \frac{2432}{30000}\right) \simeq 2234,846 > 10$ , 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  $a_\alpha = z_\alpha \sqrt{\frac{p_e(1-p_e)}{n}} \simeq 0,00259$  et  $p_e - a_\alpha \simeq \frac{2432}{30000} - 0,00259 \simeq 0,07848$  et  $p_e + a_\alpha \simeq \frac{2432}{30000} + 0,00259 \simeq 0,08366$ .

On estime donc que  $p$  est dans l'intervalle  $[0,07848; 0,08366]$  avec la confiance  $c = 0,9$ .

Toutes les valeurs de cet intervalle sont supérieures à 7,8%, donc on peut affirmer, avec la confiance 90%, que le taux de

chomage des personnes boulimiques est plus élevé que parmi le reste de la population.

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

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

On a  $n p_e(1 - p_e) = 30\,000 \times \frac{2\,432}{30\,000} \left(1 - \frac{2\,432}{30\,000}\right) \simeq 2\,234,846 > 10$ , 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  $a_\alpha = z_\alpha \sqrt{\frac{p_e(1-p_e)}{n}} \simeq 0,00367$  et  $p_e - a_\alpha \simeq \frac{2\,432}{30\,000} - 0,003\,67 \simeq 0,077\,4$  et  $p_e + a_\alpha \simeq \frac{2\,432}{30\,000} + 0,003\,67 \simeq 0,084\,74$ .

On estime donc que  $p$  est dans l'intervalle  $[0,0774; 0,08474]$  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  $0,05\%$  près le taux de chômage des personnes boulimiques avec une confiance de  $98\%$ , quelle taille d'échantillon faudrait-il considérer ?

Pour avoir une précision  $0,0005$  avec la confiance  $0,98$ , il faut  $n > z_\alpha^2 \frac{p_e(1-p_e)}{h^2} \simeq 2,326^2 \frac{0,08107(1-0,08107)}{0,0005^2} \simeq 1612211$

### Exercice 53 : Perception des couleurs

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

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

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

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

- (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  $n p_e(1 - p_e) = 3\,000 \times \frac{11}{3\,000} \left(1 - \frac{11}{3\,000}\right) \simeq 10,96 > 10$ , 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  $a_\alpha = z_\alpha \sqrt{\frac{p_e(1-p_e)}{n}} \simeq 0,00216$  et  $p_e - a_\alpha \simeq \frac{11}{3\,000} - 0,002\,16 \simeq 0,001\,507$  et  $p_e + a_\alpha \simeq \frac{11}{3\,000} + 0,002\,16 \simeq 0,005\,827$ .

On estime donc que  $p$  est dans l'intervalle  $[0,00151; 0,00583]$  avec la confiance  $c = 0,95$ .

• **parmi les hommes :**

On a  $n p_e(1 - p_e) = 3\,000 \times \frac{265}{3\,000} \left(1 - \frac{265}{3\,000}\right) \simeq 241,592 > 10$ , 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  $a_\alpha = z_\alpha \sqrt{\frac{p_e(1-p_e)}{n}} \simeq 0,01015$  et  $p_e - a_\alpha \simeq \frac{265}{3\,000} - 0,010\,15 \simeq 0,078\,18$  et  $p_e + a_\alpha \simeq \frac{265}{3\,000} + 0,010\,15 \simeq 0,098\,48$ .

On estime donc que  $p$  est dans l'intervalle  $[0,07818; 0,09848]$  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. La chercheuse pense que, bien au delà de la question du daltonisme, les filles sont bien plus douées que les garçons pour détecter d'infimes nuances de couleurs. Elle teste cette hypothèse sur des enfants de 6 ans, auxquels elle fait passer un test standardisé de perception des couleurs. Sur un échantillon de 16 filles, elle obtient les données suivantes :

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

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

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

$$\text{moyenne : } m(X) = \frac{\sum x_i}{n} = \frac{13+9+13+\dots+14}{16} = \frac{204}{16} = 12,75$$

$$m(X^2) = \frac{\sum x_i^2}{n} = \frac{13^2+9^2+13^2+\dots+14^2}{16} = \frac{2690}{16}$$

$$\text{Var}(X) = m(X^2) - m(X)^2 = \frac{2690}{16} - \left(\frac{204}{16}\right)^2 \simeq 5,56$$

$$\text{Écart-type : } s(X) = \sqrt{\text{Var}(X)} \simeq 2,36$$

- 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)

$$\text{On lit } t_\alpha \simeq 2,9467 \text{ d'où } a_\alpha = t_\alpha \frac{s_e}{\sqrt{n-1}} \simeq 1,7956.$$

On estime donc que  $\mu$  est dans l'intervalle  $[12,75 - 1,7956; 12,75 + 1,7956] \simeq [10,95; 14,55]$  avec la confiance  $c = 0,99$

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

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

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

- Comme  $n = 43 > 30$ , on cherche  $z_\alpha$  tel que  $F(z_\alpha) = \frac{0,99+1}{2} = 0,995$

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

1,805

On estime donc que  $\mu$  est dans l'intervalle  $[10,51 - 1,805; 10,51 + 1,805] \simeq [8,71; 12,32]$ , 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.

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

(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  $q = 0,005$  et la valeur  $x_2 = 32,8$  correspondant à  $p = 0,005$

$$\text{On obtient alors } s_e \sqrt{\frac{n}{x_2}} \simeq 2,36 \sqrt{\frac{16}{32,8}} \simeq 1,65 \text{ et } s_e \sqrt{\frac{n}{x_1}} \simeq$$

$$2,36 \sqrt{\frac{16}{4,601}} \simeq 4,4$$

On estime donc que  $\sigma$  est dans l'intervalle  $[1,65; 4,4]$  avec la confiance  $c = 0,99$ .

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

On lit dans la table du  $\chi^2$  à 42 degrés de liberté la valeur  $x_1 = 22,14$  correspondant à  $q = 0,005$  et la valeur  $x_2 = 69,34$  correspondant à  $p = 0,005$

$$\text{On obtient alors } s_e \sqrt{\frac{n}{x_2}} \simeq 4,54 \sqrt{\frac{43}{69,34}} \simeq 3,58 \text{ et } s_e \sqrt{\frac{n}{x_1}} \simeq$$

$$4,54 \sqrt{\frac{43}{22,14}} \simeq 6,33$$

On estime donc que  $\sigma$  est dans l'intervalle  $[3,58; 6,33]$  avec la confiance  $c = 0,99$ .

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

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

99% ?

Pour avoir une précision 0,5 avec la confiance 0,99, il faut  $n >$

$$z_{\alpha}^2 \frac{(s_e)^2}{h^2} \simeq 2,576^2 \frac{2,36^2}{0,5^2} \simeq 148$$

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