

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 les casios, ou la fonction \blacktriangleright Frac du menu \boxed{MATH} des TI).

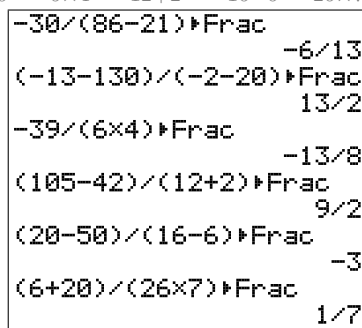
- Grâce à cette fonctionnalité de la calculatrice, simplifiez chacune des fractions suivantes : $\frac{-30}{86-21}$, $\frac{-13-130}{-2-20}$, $\frac{-39}{6\times 4}$, $\frac{105-42}{12+2}$, $\frac{20-50}{16-6}$, $\frac{6+20}{26\times 7}$.

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

$$\frac{-30}{86-21} = -\frac{6}{13}, \quad \frac{-13-130}{-2-20} = \frac{13}{2},$$

$$\frac{-39}{6\times 4} = -\frac{13}{8}, \quad \frac{105-42}{12+2} = \frac{9}{2},$$

$$\frac{20-50}{16-6} = -3, \quad \frac{6+20}{26\times 7} = \frac{1}{7}.$$

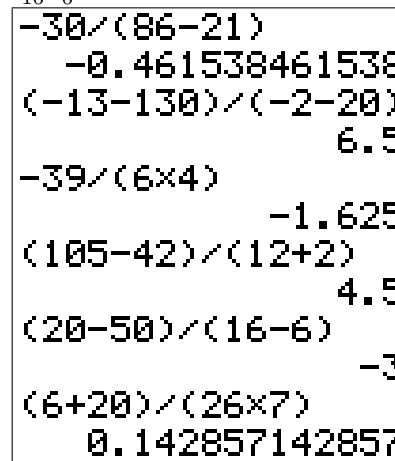


Remarque : La capture d'écran correspond à une TI. Sur une casio, si l'on veut par exemple calculer $\frac{-39}{6\times 4}$, on tappe $(-39)\div(6*4)$ puis si le résultat affiché est un chiffre à virgule on appuie sur $\boxed{F\leftrightarrow D}$ pour en faire une fraction. Au lieu d'afficher $-\frac{13}{8}$, certaines calculatrices affichent alors $-1\frac{5}{8}$ (ce qui signifie $-(1 + \frac{5}{8})$). 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) :

$$\frac{-30}{86-21} \simeq -0,462, \quad \frac{-13-130}{-2-20} = 6,5, \quad \frac{-39}{6\times 4} = -1,625, \quad \frac{105-42}{12+2} = 4,5,$$

$$\frac{20-50}{16-6} = -3, \quad \frac{6+20}{26\times 7} \simeq 0,143.$$



L'ordre croissant de ces nombres est : $-3 < -1,625 < -0,462 < 0,143 < 4,5 < 6,5$

Donc la réponse est :

$$\frac{20-50}{16-6} < \frac{-39}{6\times 4} < \frac{-30}{86-21} < \frac{6+20}{26\times 7} < \frac{105-42}{12+2} < \frac{-13-130}{-2-20}$$

Exercice 2 : Utilisation de la calculatrice

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



Peut-on en conclure

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


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

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

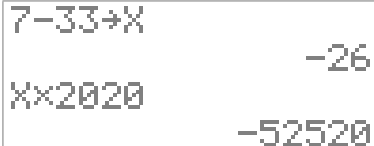
La réponse (b) est aussi fausse : le carré de 1 est égal à 1 .

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) $\frac{1}{2019 \times 2020} \simeq 24519539,62$
 (b) $\frac{1}{2019 \times 2020} \simeq 0,000000245$
 (c) $\frac{1}{2019 \times 2020} \simeq 2,45$

La bonne réponse est la réponse (b). En effet **2.451953962E-7** désigne la quantité $2,451953962 \times 10^{-7}$, qui vaut $0,0000002451953962$.

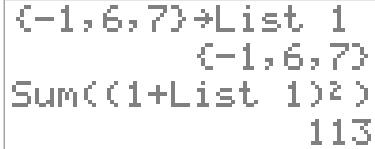
3.  (a) $7 - 33 \times 2020 = -52520$
 (b) $26 \times 2020 = -52520$
 (c) $(7 - 33) \times 2020 = -52520$

La bonne réponse est la réponse (c). En effet dans **X*2020**, **X** vaut $7 - 33$, (calculé à la ligne précédente), de sorte qu'on a calculé $(7 - 33) \times 2020$.

4.  (a) $-14 \times 5^2 = 4900$
 (b) $-70^2 = 4900$
 (c) $(-14 \times 5)^2 = 4900$

La bonne réponse est la réponse (c). En effet dans **Ans^2**, **Ans** vaut -14×5 , (calculé à la ligne précédente), de sorte qu'on a

calculé $(-14 \times 5)^2$.

5.  (a) $1 + (-1 + 6 + 7)^2 = 113$
 (b) $1 - 1^2 + 6^2 + 7^2 = 113$
 (c) $(1 - 1)^2 + (1 + 6)^2 + (1 + 7)^2 = 113$

La bonne réponse est la réponse (c). En effet **(1+List 1)^2** désigne la liste dont les éléments sont $(1 - 1)^2$, $(1 + 6)^2$ et $(1 + 7)^2$. En conséquence, leur somme **Sum((1+List 1)^2)** désigne $(1 - 1)^2 + (1 + 6)^2 + (1 + 7)^2$,

Exercice 3 : Puissances de dix

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

$$63 \div 100 \qquad 0,63 \times 10^2 \qquad 0,10316 \times 10^4$$

$$1,6786207 \times 10^9 \qquad 4 \times 1\,000\,000\,000 \qquad 1\,081,903 \times 10^6$$

$$107 \times 1\,000\,000\,000 \qquad 156\,313 \div 10\,000\,000 \qquad 0,154004 \times 10^{-7}$$

- $63 \div 100 = 0,63$ (que l'on peut aussi écrire $6,3 \times 10^{-1}$)
- $0,63 \times 10^2 = 63$ (que l'on peut aussi écrire $6,3 \times 10^1$)
- $0,10316 \times 10^4 = 1031,6$
(que l'on peut aussi écrire $1,0316 \times 10^3$)
- $1,6786207 \times 10^9 = 1678620700$
(que l'on peut aussi écrire $1,6786207 \times 10^9$)
- $4 \times 1\,000\,000\,000 = 4000000000$
(que l'on peut aussi écrire 4×10^9)
- $1\,081,903 \times 10^6 = 1081903000$
(que l'on peut aussi écrire $1,081903 \times 10^9$)
- $107 \times 1\,000\,000\,000 = 107000000000$
(que l'on peut aussi écrire $1,07 \times 10^{11}$)
- $156\,313 \div 10\,000\,000 = 0,0156313$
(que l'on peut aussi écrire $1,56313 \times 10^{-2}$)
- $0,154004 \times 10^{-7} = 0,0000000154004$
(que l'on peut aussi écrire $1,54004 \times 10^{-8}$)

Exercice 4 : Règles de calcul

Remarque : dans cet exercice, la justification des égalités n'est pas exigible en examens. Toutefois c'est une exercice très pertinent pour vous assurer que vous être à l'aise avec le type de calculs qui apparaîtront au cours du semestre.

1. On considère les expressions suivantes :

- $9 + 8 \times 4 + (5^2 \times 6 + (3))$
- $(9 + (8 \times 4)) + 5^2(6 + 3)$
- $(9 + 8) \times 4 + (6 + 3) \times 5^2$
- $9 + 8 \times 4 + 5^2 \times 6 + 3$
- $8 \times 4 + (9 + (6 + 3) \times 5 \times 5)$
- $9 \times 4 + 8 \times 4 + 5^2(6 + 3)$

(a) Calculer la valeur de chacune de ces expressions.

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

- $9 + 8 \times 4 + (5^2 \times 6 + (3)) = 194$
- $(9 + (8 \times 4)) + 5^2(6 + 3) = 266$
- $(9 + 8) \times 4 + (6 + 3) \times 5^2 = 293$
- $9 + 8 \times 4 + 5^2 \times 6 + 3 = 194$
- $8 \times 4 + (9 + (6 + 3) \times 5 \times 5) = 266$
- $9 \times 4 + 8 \times 4 + 5^2(6 + 3) = 293$

$9+8*4+(5^2*6+(3))$	194
$(9+(8*4))+5^2*(6+3)$	266
$(9+8)*4+(6+3)*5^2$	293
$9+8*4+5^2*6+3$	194
$8*4+(9+(6+3)*5*5)$	266
$9*4+8*4+5^2*(6+3)$	293

(b) Vous avez sans doute trouvé que certaines expressions étaient égales. Justifiez pourquoi elles sont égales.

La correction de cette question est reportée à la page suivante pour avoir la place de détailler. (La rédaction indique délibérément beaucoup d'étapes, chacun saurait certainement se passer d'une grande partie d'entre elles.)

- i. Les expressions $9 + 8 \times 4 + (5^2 \times 6 + (3))$ et $9 + 8 \times 4 + 5^2 \times 6 + 3$ sont égales entre elles (on a en effet trouvé qu'elles valent toutes 194). On justifie ci-dessous pourquoi elles sont égales :

$$9 + 8 \times 4 + (5^2 \times 6 + (3)) = 9 + 8 \times 4 + 5^2 \times 6 + 3 \quad (\text{En supprimant les parenthèses inutiles})$$

- ii. Les expressions $(9 + (8 \times 4)) + 5^2(6 + 3)$ et $8 \times 4 + (9 + (6 + 3) \times 5 \times 5)$ sont égales entre elles (on a en effet trouvé qu'elles valent toutes 266). On justifie ci-dessous pourquoi elles sont égales :

- $8 \times 4 + (9 + (6 + 3) \times 5 \times 5) = 8 \times 4 + 9 + (6 + 3) \times 5 \times 5$ (En supprimant les parenthèses inutiles)
- $(9 + (8 \times 4)) + 5^2(6 + 3) = 9 + 8 \times 4 + 5^2 \times (6 + 3)$ (En supprimant les parenthèses inutiles)
- $= 8 \times 4 + 9 + 5^2 \times (6 + 3)$ (En réordonnant les termes)
- $= 8 \times 4 + 9 + 5 \times 5 \times (6 + 3)$ (En réécrivant le carré)
- $= 8 \times 4 + 9 + (6 + 3) \times 5 \times 5$ (En réordonnant les facteurs)

$$\text{Donc } 8 \times 4 + (9 + (6 + 3) \times 5 \times 5) = (9 + (8 \times 4)) + 5^2(6 + 3).$$

- iii. Les expressions $(9 + 8) \times 4 + (6 + 3) \times 5^2$ et $9 \times 4 + 8 \times 4 + 5^2(6 + 3)$ sont égales entre elles (on a en effet trouvé qu'elles valent toutes 293). On justifie ci-dessous pourquoi elles sont égales :

$$(9 + 8) \times 4 + (6 + 3) \times 5^2 = 9 \times 4 + 8 \times 4 + (6 + 3) \times 5^2 \quad (\text{En développant le produit})$$

$$= 9 \times 4 + 8 \times 4 + 5^2 \times (6 + 3) \quad (\text{En réordonnant les facteurs})$$

$$\text{Donc } (9 + 8) \times 4 + (6 + 3) \times 5^2 = 9 \times 4 + 8 \times 4 + 5^2(6 + 3).$$

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

(a) Les expressions suivantes

$$\bullet \frac{4 \times 3}{x} - \frac{4 \times 9}{6}$$

$$\bullet 4 \left(\frac{3}{x} - \frac{9}{6} \right)$$

lorsque $x = -6$.

$$\bullet (4 \times 3 - 9 \times 4) \div (x - 6)$$

$$\bullet 4 \frac{(3-9)}{(x-6)}$$

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

- $\frac{4 \times 3}{x} - \frac{4 \times 9}{6} = -8$
- $4 \left(\frac{3}{x} - \frac{9}{6} \right) = -8$
- $(4 \times 3 - 9 \times 4) \div (x - 6) = 2$
- $4 \frac{(3-9)}{(x-6)} = 2$

$-6 \div x$	
-6	
$4 * 3 \div x - 4 * 9 \div 6$	-6
-8	
$4 * (3 \div x - 9 \div 6)$	-8
-8	
$(4 * 3 - 9 * 4) \div (x - 6)$	2
2	
$4 * (3 - 9) \div (x - 6)$	2
2	

i. Les expressions $\frac{4 \times 3}{x} - \frac{4 \times 9}{6}$ et $4 \left(\frac{3}{x} - \frac{9}{6} \right)$ sont égales entre elles (on a en effet trouvé qu'elles valent toutes -8). On justifie ci-dessous pourquoi elles sont égales :

$$4 \left(\frac{3}{x} - \frac{9}{6} \right) = 4 \times \frac{3}{x} - 4 \times \frac{9}{6} \quad (\text{En développant le produit})$$

ii. Les expressions $(4 \times 3 - 9 \times 4) \div (x - 6)$ et $4 \frac{(3-9)}{(x-6)}$ sont égales entre elles (on a en effet trouvé qu'elles valent toutes 2). On justifie ci-dessous pourquoi elles sont égales :

- $4 \frac{(3-9)}{(x-6)} = \frac{4 \times 3}{x-6} - \frac{4 \times 9}{x-6}$ (En développant le produit)

- $(4 \times 3 - 9 \times 4) \div (x - 6) = 4 \times 3 \div (x - 6) - 9 \times 4 \div (x - 6)$ (En développant le produit)
- $= 4 \times 3 \div (x - 6) - 4 \times 9 \div (x - 6)$ (En réordonnant les facteurs)
- $= \frac{4 \times 3}{x-6} - \frac{4 \times 9}{x-6}$ (En réécrivant sous forme de fraction)

Donc $4 \frac{(3-9)}{(x-6)} = (4 \times 3 - 9 \times 4) \div (x - 6)$.

(b) Les expressions

- $a - 6 + 3 \times 2 + 3 \times 4$
- $a - 6 - 0(3 - 2 - 4)$
- $a - 6 + 3 - 2 - 4 - (3 - 2 - 4)$
- $a - 6 - 1(3 - 2 - 4)$
- $a + b - 6 - b + 3 \times 2 + 3 \times 4$
- $a - 6 - (3 - 2 - 4)$
- $a - 6 - 3(-2 - 4)$

où $a = 4$ et $b = -2$.

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

- $a - 6 + 3 \times 2 + 3 \times 4 = 16$
- $a - 6 - 0(3 - 2 - 4) = -2$
- $a - 6 + 3 - 2 - 4 - (3 - 2 - 4) = -2$
- $a - 6 - 1(3 - 2 - 4) = 1$
- $a + b - 6 - b + 3 \times 2 + 3 \times 4 = 16$
- $a - 6 - (3 - 2 - 4) = 1$
- $a - 6 - 3(-2 - 4) = 16$

4→A	4
-2→B	-2
A-6+3*2+3*4	16
A-6-0*(3-2-4)	-2
A-6+3-2-4-(3-2-4)	-2
A-6-1*(3-2-4)	1
A+B-6-B+3*2+3*4	16
A-6-(3-2-4)	1
A-6-3*(-2-4)	16

i. Les expressions $a - 6 - 0(3 - 2 - 4)$ et $a - 6 + 3 - 2 - 4 - (3 - 2 - 4)$ sont égales entre elles (on a en effet trouvé qu'elles valent toutes -2). On justifie ci-dessous pourquoi elles sont égales :

- $a - 6 + 3 - 2 - 4 - (3 - 2 - 4) = a - 6 + 3 - 2 - 4 - 3 + 2 + 4$ (En réécrivant les signes)
 $= a - 6$ (En enlevant des termes qui s'annulent)
- $a - 6 - 0(3 - 2 - 4) = a - 6$ (En réécrivant la multiplication par 0)

Donc $a - 6 + 3 - 2 - 4 - (3 - 2 - 4) = a - 6 - 0(3 - 2 - 4)$.

ii. Les expressions $a - 6 - 1(3 - 2 - 4)$ et $a - 6 - (3 - 2 - 4)$ sont égales entre elles (on a en effet trouvé qu'elles valent toutes 1). On justifie ci-dessous pourquoi elles sont égales :

$$a - 6 - 1(3 - 2 - 4) = a - 6 - (3 - 2 - 4) \quad (\text{En réécrivant la multiplication par 1})$$

iii. Les expressions $a - 6 + 3 \times 2 + 3 \times 4$, $a + b - 6 - b + 3 \times 2 + 3 \times 4$ et $a - 6 - 3(-2 - 4)$ sont égales entre elles (on a en effet trouvé qu'elles valent toutes 16). On justifie ci-dessous pourquoi elles sont égales en montrant que $a + b - 6 - b + 3 \times 2 + 3 \times 4$ et $a - 6 - 3(-2 - 4)$ sont égaux à $a - 6 + 3 \times 2 + 3 \times 4$.

a) $a + b - 6 - b + 3 \times 2 + 3 \times 4 = a - 6 + 3 \times 2 + 3 \times 4$ (En enlevant des termes qui s'annulent)

b) $a - 6 - 3(-2 - 4) = a - 6 - (-3 \times 2 - 3 \times 4)$ (En développant le produit)
 $= a - 6 + 3 \times 2 + 3 \times 4$ (En réécrivant les signes)

Donc $a - 6 - 3(-2 - 4) = a - 6 + 3 \times 2 + 3 \times 4$.

(c) Lorsque $x = -4$, $y = 1$ et $z = -8$:

• $2 - x + y - z$

• $(2 - (x + y - z))$

$$\bullet y - x + 2 - z$$

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

- $2 - x + y - z = 15$
- $y - x + 2 - z = 15$
- $(2 - (x + y - z)) = -3$
- $(z - x) - (y - 2) = -3$

-4+X	-4
1+Y	1
-8+Z	-8
2-X+Y-Z	15
Y-X+2-Z	15
(2-(X+Y-Z))	-3
(Z-X)-(Y-2)	-3

$$\bullet (z - x) - (y - 2)$$

i. Les expressions $(2 - (x + y - z))$ et $(z - x) - (y - 2)$ sont égales entre elles (on a en effet trouvé qu'elles valent toutes -3). On justifie ci-dessous pourquoi elles sont égales :

- $(z - x) - (y - 2) = z - x - (y - 2)$ (En supprimant les parenthèses inutiles)
 $= z - x - y + 2$ (En réécrivant les signes)
- $(2 - (x + y - z)) = 2 - (x + y - z)$ (En supprimant les parenthèses inutiles)
 $= 2 - x - y + z$ (En réécrivant les signes)
 $= z - x - y + 2$ (En réordonnant les termes)

Donc $(z - x) - (y - 2) = (2 - (x + y - z))$.

ii. Les expressions $2 - x + y - z$ et $y - x + 2 - z$ sont égales entre elles (on a en effet trouvé qu'elles valent toutes 15). On justifie ci-dessous pourquoi elles sont égales :

$$2 - x + y - z = y - x + 2 - z \quad (\text{En réordonnant les termes})$$

(d) Lorsque $x_1 = -5$ et $x_2 = -3$:

- $7 - x_2 \times \frac{9-5}{x_2} x_1$
- $(7 - 9) - 5x_1$

- $-9 - 5x_1 + 7$
- $7 - (9 - 5)x_1$

$$\bullet 5x_1 + 7 - 9x_1$$

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

- $7 - x_2 \times \frac{9-5}{x_2} x_1 = 27$
- $(7 - 9) - 5x_1 = 23$
- $-9 - 5x_1 + 7 = 23$
- $7 - (9 - 5)x_1 = 27$
- $5x_1 + 7 - 9x_1 = 27$

-5*x	-5
-3+y	-3
7-y*(9-5)+y*x	27
(7-9)-5*x	23
-9-5*x+7	23
7-(9-5)*x	27
5*x+7-9*x	27

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 .

i. Les expressions $(7 - 9) - 5x_1$ et $-9 - 5x_1 + 7$ sont égales entre elles (on a en effet trouvé qu'elles valent toutes 23). On justifie ci-dessous pourquoi elles sont égales :

$$\begin{aligned} (7 - 9) - 5x_1 &= 7 - 9 - 5 \times x_1 && \text{(En supprimant les parenthèses inutiles)} \\ &= -9 - 5 \times x_1 + 7 && \text{(En réordonnant les termes)} \end{aligned}$$

Donc $(7 - 9) - 5x_1 = -9 - 5x_1 + 7$.

ii. Les expressions $7 - x_2 \times \frac{9-5}{x_2} x_1$, $7 - (9 - 5)x_1$ et $5x_1 + 7 - 9x_1$ sont égales entre elles (on a en effet trouvé qu'elles valent toutes 27). On justifie ci-dessous pourquoi elles sont égales en montrant que $7 - (9 - 5)x_1$ et $5x_1 + 7 - 9x_1$ sont égaux à $7 - x_2 \times \frac{9-5}{x_2} x_1$.

a) $7 - x_2 \times \frac{9-5}{x_2} x_1 = 7 - (9 - 5) \times x_1$ (En simplifiant la division)

b)
$$\begin{aligned} 7 - x_2 \times \frac{9-5}{x_2} x_1 &= 7 - \left(\frac{x_2 \times 9}{x_2} \times x_1 - \frac{x_2 \times 5}{x_2} \times x_1 \right) && \text{(En développant le produit)} \\ &= 7 - \frac{x_2 \times 9}{x_2} \times x_1 + \frac{x_2 \times 5}{x_2} \times x_1 && \text{(En réécrivant les signes)} \\ &= \frac{x_2 \times 5}{x_2} \times x_1 + 7 - \frac{x_2 \times 9}{x_2} \times x_1 && \text{(En réordonnant les termes)} \\ &= 5 \times x_1 + 7 - 9 \times x_1 && \text{(En simplifiant la division)} \end{aligned}$$

Donc $7 - x_2 \times \frac{9-5}{x_2} x_1 = 5x_1 + 7 - 9x_1$.

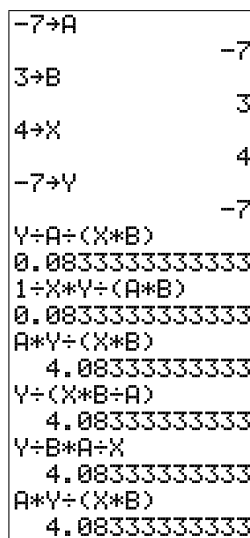
(e) Lorsque $a_1 = -7$, $a_2 = 3$, $x_1 = 4$ et $x_2 = -7$:

- $\frac{x_2 \div a_1}{x_1 \times a_2}$
- $\frac{1}{x_1} \times \frac{x_2}{a_1 \times a_2}$
- $\frac{a_1 \times x_2}{x_1 \times a_2}$

- $\frac{x_2}{x_1 \times a_2 \div a_1}$
- $x_2 \div a_2 \times a_1 \div x_1$
- $a_1 \frac{x_2}{x_1 \times a_2}$

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

- $\frac{x_2 \div a_1}{x_1 \times a_2} \simeq 0,083$
- $\frac{1}{x_1} \times \frac{x_2}{a_1 \times a_2} \simeq 0,083$
- $\frac{a_1 \times x_2}{x_1 \times a_2} \simeq 4,083$
- $\frac{x_2}{x_1 \times a_2 \div a_1} \simeq 4,083$
- $x_2 \div a_2 \times a_1 \div x_1 \simeq 4,083$
- $a_1 \frac{x_2}{x_1 \times a_2} \simeq 4,083$



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 .

i. Les expressions $\frac{x_2 \div a_1}{x_1 \times a_2}$ et $\frac{1}{x_1} \times \frac{x_2}{a_1 \times a_2}$ sont égales entre elles (on a en effet trouvé qu'elles valent toutes 0,083). On justifie ci-dessous pourquoi elles sont égales :

$$\begin{aligned} \frac{x_2 \div a_1}{x_1 \times a_2} &= \frac{1}{x_1} \times \frac{x_2 \div a_1}{a_2} \quad (\text{En réordonnant les facteurs}) \\ &= \frac{1}{x_1} \times \frac{x_2}{a_1 \times a_2} \quad (\text{En réécrivant sous forme de fraction}) \end{aligned}$$

Donc $\frac{x_2 \div a_1}{x_1 \times a_2} = \frac{1}{x_1} \times \frac{x_2}{a_1 \times a_2}$.

ii. Les expressions $\frac{a_1 \times x_2}{x_1 \times a_2}$, $\frac{x_2}{x_1 \times a_2 \div a_1}$, $x_2 \div a_2 \times a_1 \div x_1$ et $a_1 \frac{x_2}{x_1 \times a_2}$ sont égales entre elles (on a en effet trouvé qu'elles valent toutes 4,083). On justifie ci-dessous pourquoi elles sont égales en montrant que $\frac{x_2}{x_1 \times a_2 \div a_1}$, $x_2 \div a_2 \times a_1 \div x_1$ et $a_1 \frac{x_2}{x_1 \times a_2}$ sont égaux à $\frac{a_1 \times x_2}{x_1 \times a_2}$.

a) • $\frac{x_2}{x_1 \times a_2 \div a_1} = \frac{x_2}{x_1 \times a_2} \times a_1$ (En réécrivant la fraction)

• $\frac{a_1 \times x_2}{x_1 \times a_2} = \frac{x_2}{x_1 \times a_2} \times a_1$ (En réordonnant les facteurs)

Donc $\frac{x_2}{x_1 \times a_2 \div a_1} = \frac{a_1 \times x_2}{x_1 \times a_2}$.

b) • $x_2 \div a_2 \times a_1 \div x_1 = \frac{x_2}{a_2} \times \frac{a_1}{x_1}$ (En réécrivant sous forme de fraction)

• $\frac{a_1 \times x_2}{x_1 \times a_2} = \frac{x_2}{a_2} \times \frac{a_1}{x_1}$ (En réordonnant les facteurs)

Donc $x_2 \div a_2 \times a_1 \div x_1 = \frac{a_1 \times x_2}{x_1 \times a_2}$.

c) L'égalité $a_1 \frac{x_2}{x_1 \times a_2} = \frac{a_1 \times x_2}{x_1 \times a_2}$ est évidente

Compléments

Exercice 5 : Précision des calculs

Amélie, Gaétan, Nora, Assia et Pauline étudient en première année de psychologie à l’université de Bourgogne.

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

$$m(X) = \frac{1,88+1,89+1,79}{3}$$

$$m(X^2) = \frac{1,88^2+1,89^2+1,79^2}{3}$$

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

$$s = \sqrt{Var}$$

Le nombre “ s ” issu de ce calcul est l’écart-type, exprimé (dans le cas présent) en mètre. On leur demandait de calculer cet écart type à 1 cm près.

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

Copie d’Amélie

$$m(X) = \frac{1,88+1,89+1,79}{3} \simeq 1,85$$

$$m(X^2) = \frac{1,88^2+1,89^2+1,79^2}{3} \simeq 3,44$$

$$Var \simeq 3,44 - 1,85^2 \simeq 0,02$$

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

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

Calculatrice d’Amélie

```
(1.88+1.89+1.79)/3
1.853333333333
(1.88^2+1.89^2+1.79^2)/3
3.436866666667
3.44-1.85^2
0.0175
√(0.02)
0.141421356237
```

Copie de Gaétan

$$m(X) = \frac{1,88+1,89+1,79}{3} = \frac{5,56}{3}$$

$$m(X^2) = \frac{1,88^2+1,89^2+1,79^2}{3} = \frac{10,3106}{3}$$

$$Var \simeq \frac{10,3106}{3} - \left(\frac{5,56}{3}\right)^2 \simeq 0,0020$$

$$s = \sqrt{Var} \simeq 0,04 \text{ mètre.}$$

Donc l’écart-type est d’environ 4 cm.

Calculatrice de Gaétan

```
1.88+1.89+1.79
5.56
1.88^2+1.89^2+1.79^2
10.3106
10.3106/3-(5.56/3)^2
0.00202222222222
√(Ans)
0.0449691252108
```

Copie de Nora

$$m(X) = \frac{1,88+1,89+1,79}{3} \simeq 1,85$$

$$m(X^2) = \frac{1,88^2+1,89^2+1,79^2}{3} \simeq 3,44$$

$$Var = m(X^2) - (m(X))^2 \simeq 0,0020$$

$$s = \sqrt{Var} \simeq 0,04 \text{ mètre.}$$

Donc l’écart-type est d’environ 4 cm.

Calculatrice de Nora

```
(1.88+1.89+1.79)/3*M
1.853333333333
(1.88^2+1.89^2+1.79^2)/3+M
3.436866666667
M-M^2+U
0.00202222222222
√(U)
0.0449691252108
```

Copie d’Assia

$$m(X) = \frac{1,88+1,89+1,79}{3} \simeq 1,85$$

$$m(X^2) = \frac{1,88^2+1,89^2+1,79^2}{3} \simeq 3,44$$

$$Var = m(X^2) - (m(X))^2 \simeq 0,0020$$

$$s = \sqrt{Var} \simeq 0,04 \text{ mètre.}$$

Donc l’écart-type est d’environ 4 cm.

Calculatrice d’Assia

```
(1.88,1.89,1.79)+List 1
(1.88,1.89,1.79)
Sum (List 1)/3+M
1.853333333333
Sum ((List 1)^2)/3+M
3.436866666667
M-M^2+U
0.00202222222222
√(U)
0.0449691252108
```

Copie de Pauline

$$s = \sqrt{\frac{1,88^2+1,89^2+1,79^2}{3} - \left(\frac{1,88+1,89+1,79}{3}\right)^2}$$

$$\simeq 0,04 \text{ mètre.}$$

Donc l’écart-type est d’environ 4 cm.

Calculatrice de Pauline

```
√((1.88^2+1.89^2+1.79^2)/3-
((1.88+1.89+1.79)/3)^2)
0.0449691252108
```

1. Parmi ces différentes réponses, laquelle/lesquelles vous semble(nt) correcte(s) ?
Expliquez pourquoi l’un-e des étudiant-e-s a obtenu(e) une réponse erronée ?
2. On demande ensuite de calculer l’écart type de la taille d’un échantillon de 5 femmes. Comme l’échantillon n’est plus le même, il faut cette fois-ci calculer

$$m(X) = \frac{1,66+1,67+1,69+1,62+1,63}{5}$$

$$m(X^2) = \frac{1,66^2+1,67^2+1,69^2+1,62^2+1,63^2}{5}$$

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

$$s = \sqrt{Var}$$
 Rédigez ce calcul de trois façon différentes, correspondant aux

méthodes utilisées par chacun·e des étudiant·e·s ayant trouvé la bonne réponse à la première question.

Exercice 6 : Utilisation du symbole \sum

1. Calculer, sans utiliser la calculatrice, chacune des sommes suivantes :

$$\sum_{k=1}^5 k, \sum_{k=-5}^5 k, \sum_{i=-2}^2 i^2.$$

2. Écrivez avec le symbole \sum la somme suivante $\frac{6^2}{6+3} + \frac{7^2}{7+3} + \frac{8^2}{8+3} + \frac{9^2}{9+3} + \frac{10^2}{10+3} + \frac{11^2}{11+3} + \frac{12^2}{12+3} + \frac{13^2}{13+3}$.

3. On pose $x_1 = 3$, $x_2 = 4$, $x_3 = 5$, $x_4 = 6$, $x_5 = 7$, $x_6 = 8$, $n_1 = 3$, $n_2 = 10$, $n_3 = 10$, $n_4 = 9$, $n_5 = 5$ et $n_6 = 6$.

(a) Calculer $\sum_{i=1}^6 n_i (x_i)^2$ sans utiliser de liste sur la calculatrice.

(b) Calculer à nouveau $\sum_{i=1}^6 n_i (x_i)^2$, en utilisant cette fois-ci des listes sur la calculatrice.

Chapitre 1 : Statistiques descriptives univariées

Exercice 7 : Type de variables

1. Mme Collet recueille des données sur
- les langues officielles des différents pays européens
 - la sensibilité aux questions écologiques des habitants d'un même quartier
 - le nombre de personnes des ménages Français
 - le taux de testostérone parmi les patients souffrant de troubles cognitifs

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

2. De même quel est le type de variable lorsqu'on étudie :

- le temps hebdomadaire passé devant la télévision par des enfants
- le poids des appareils photos vendus en 2019
- les numéros de téléphones figurant dans les pages blanches dijonnaises
- la couleur des voitures stationnées dans le campus
- le nombre de pages des livres d'une bibliothèque

Exercice 8 : Taille des ménages On appelle T la variable statistique indiquant le nombre de personnes d'un ménage. Au sein d'une certaine ville, on obtient les données suivantes :

Nombre de personnes du ménage	1	2	3	4	5	6
Effectifs : nombre de ménages	1853	1122	1925	979	324	123

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

Exercice 9 : Moyens de transport

- Utilisation des transports en commun

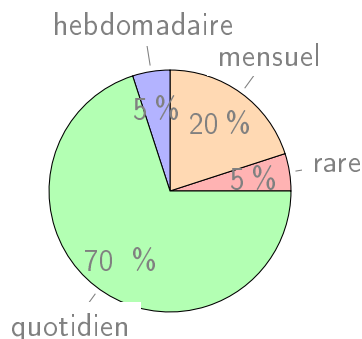
(a) Un étudiant en sociologie interroge 20 personnes de son entourage sur leur utilisation des transports en commun. Il réalise le graphique ci-dessous après les avoir regroupés en quatre groupes :

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

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

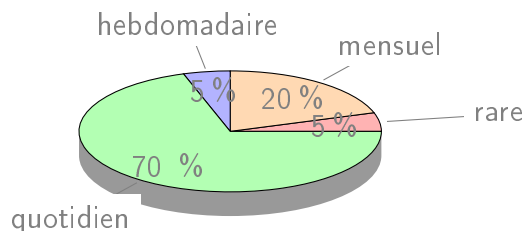
hebdomadaire : entre une et cinq fois par semaine

quotidien : au moins 5 fois par semaine

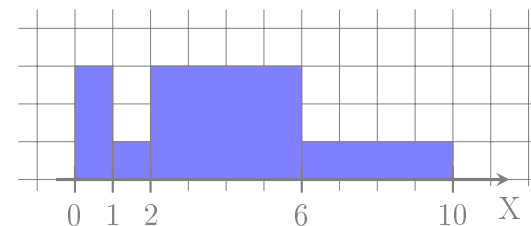


- i. Déterminer les effectifs des différentes modalités
- ii. Quelle est, au sein de cet échantillon, la proportion d'individus qui utilisent les transports en commun moins d'une fois par semaine ?

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



2. **Utilisation de la voiture** Pour le même échantillon, il désigne par X le le temps hebdomadaire passé dans une voiture (exprimé en heures). Il trace un histogramme sur une feuille quadrillée (et constate avec surprise que ses rectangles tombent exactement sur le quadrillage de la feuille) :



(a) Déterminer les fréquences des différentes classes

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

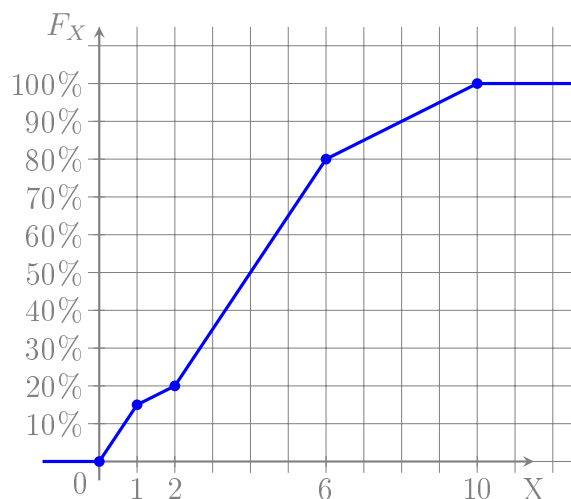
(b) Commentez : comparer les fréquences de la première et de la dernière classe, ainsi que la hauteur des rectangles correspondants.

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

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

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

(c) On considère l'individu qui passe le moins de temps en voiture parmi les 10% d'individus qui passent le plus de temps en voiture. Combien de temps passe t-il environ par semaine dans sa voiture ?



Exercice 10 : Pays fondateurs de la zone euro

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

Allemagne	Autriche	Belgique	Espagne	Finlande
République 82,2 M.hab.	République 8,7 M.hab.	Monarchie 11,3 M.hab.	Monarchie 46,4 M.hab.	République 5,5 M.hab.
France	Irlande	Italie	Luxembourg	Pays-Bas
République 67,4 M.hab.	République 4,6 M.hab.	République 60,8 M.hab.	Monarchie 0,6 M.hab.	Monarchie 17,0 M.hab.
Portugal				
République 10,4 M.hab.				

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

Exercice 11 : Développement psychomoteur

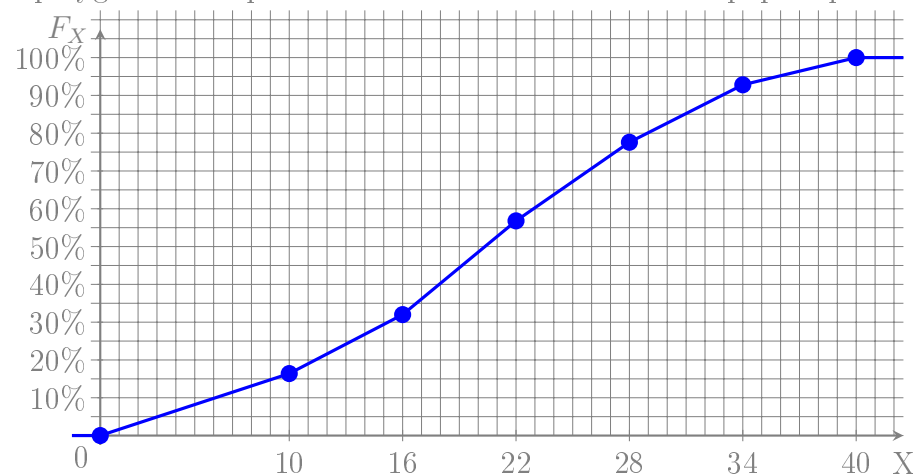
Les données ci-dessous décrivent, pour un ensemble de 160 bébés âgés de deux ans, la variable statistique X représentant le score de développement psychomoteur (SDP) :

SDP	[50 ; 65[[65 ; 80[[80 ; 95[[95 ; 110[[110 ; 125[[125 ; 140[
Effectif	5	30	55	48	18	4

1. Calculer la médiane et les quartiles de X .
2. Calculer la moyenne et l’écart type de X .

Exercice 12 : Méthodes d’apprentissage

On teste une méthode « d’apprentissage par dessins commentés » sur un échantillon de 250 enfants. On appelle X la note obtenue par les enfants à l’issue de l’apprentissage (notée entre 0 et 40). On représente le polygone de fréquences cumulées de ces notes sur un papier quadrillé :



1. Déterminer graphiquement la médiane et les quartiles de la variable statistique X .
2. (a) Donner 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	41	39	62	52	38	18

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

Exercice 13 : Calculer la moyenne et l'écart type des données de l'exercice 8.

Exercice 14 : Type de variables

- Mr Leduc recueille des données sur
 - le nombre de pièces des appartements en location à Dijon
 - le nombre moyen d'enfants par femme dans chaque pays européen
 - la douleur ressentie par différents patients atteints de la même pathologie
 - les prénoms des enfants nés en 2019

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

Cas	Population	Variable	Nature
(a)	l'ensemble des appartements en location à Dijon	le nombre de pièces	quantitative discrète
(b)	les pays européens	le nombre moyen d'enfants par femme	quantitative continue
(c)	différents patients atteints de la même pathologie	la douleur ressentie	qualitative ordinale
(d)	l'ensemble des enfants nés en 2019	le prénom	qualitative nominale

- De même quel est le type de variable lorsqu'on étudie :
 - les numéros de sécurité sociale des enseignants de l'université de Bourgogne
qualitative nominale
 - le temps mis par des rats pour sortir d'un labyrinthe
quantitative continue
 - la pluie tombée à Dijon au cours des différents mois de l'année
quantitative continue
 - le nombre de rendez-vous médicaux que différents patients ont pris au cours de l'année 2019
quantitative discrète
 - la popularité des différents footballeurs de l'équipe de France
qualitative ordinale
 - le nombre de crises d'épilepsie de différents patients au cours du mois d'octobre 2019
quantitative discrète
 - les noms des rues de Dijon
qualitative nominale

- (h) le nombre d'enfants de chaque femme dijonnaise
quantitative discrète
- (i) le degré de satisfaction des usagers des transports en commun
qualitative ordinale
- (j) les marques des téléphones vendus en 2019
qualitative nominale

Exercice 15 : Questionnaire

Un chercheur établit un questionnaire visant à évaluer les compétences logiques de jeunes adolescents. Il fait remplir ce questionnaire à 14 sujets, et pour chacun d'eux il note dans un tableau le symbole « ✓ » si la réponse donnée est correcte, « 0 » si la réponse donnée est fausse, et a laissé la case vide si le sujet n'a pas répondu. Il obtient le tableau suivant :

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

1. Quelle a été le taux de réponse à la question 10 ?
En tout, 5 personnes (parmi les 14 interrogés) ont répondu à la question 10. Le taux de réponse à cette question est donc $\frac{5}{14} \approx 0,357$.
2. Quelle proportion de sujets a répondu correctement à la question 3 ?
Comme 8 personnes ont répondu correctement à la question 3, cette proportion est $\frac{8}{14} \approx 0,571$.

3. Parmi les individus ayant répondu à la question 8, quelle proportion a donné la bonne réponse ?
Parmi les 10 personnes qui ont répondu à la question 8, il y en a 5 qui ont répondu correctement, soit une proportion de $\frac{5}{10} = 0,5$.
4. Le chercheur décide de résumer par une note les réponses données au questionnaire. Il attribue un point par bonne réponse (et zéro point par réponse erronée ou manquante). Calculer les notes (sur 10) attribuée à chaque jeune interrogé.

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

5. Calculer la moyenne, l'écart type et la médiane de ces notes.

moyenne : $m(X) = \frac{\sum x_i}{n} = \frac{5+2+\dots+7}{14} = \frac{60}{14} \approx 4,29$

$m(X^2) = \frac{\sum x_i^2}{n} = \frac{5^2+2^2+\dots+7^2}{14} = \frac{310}{14}$

$Var(X) = m(X^2) - m(X)^2 = \frac{310}{14} - \left(\frac{60}{14}\right)^2 \approx 3,78$

Écart-type : $s(X) = \sqrt{Var(X)} \approx 1,94$

La médiane est la valeur numéro $\frac{14+1}{2} = 7,5$ (en ordonnant par ordre croissant), ou plutôt le milieu entre les valeurs numéro 7 et 8. C'est donc le milieu entre 4 et 5 c'est-à-dire $\frac{4+5}{2} = 4,5$.

Exercice 16 : Regroupement en classes

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

3 ; 4,5 ; 7,5 ; 6,5 ; 7,5 ; 8 ; 7 ; 2,5 ; 5,5 ; 5,5 ; 0 ; 10 ; 7,5 ; 4 ; 5,5 ; 4,5 ; 3,5 ; 1,5 ; 2 ; 7,5 ; 5 ; 0,5 ; 9 ; 0,5 ; 7 ; 4,5 ; 5 ; 4 ; 1 ; 7 ; 8,5 ; 6,5 ; 3,5 ; 8 ; 6 ; 2,5 ; 7 ; 7,5 ; 3 ; 8,5 ; 5 ; 6 ; 5 ; 7,5 ; 3,5 ; 5,5

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

modalité	0	0,5	1	1,5	2	2,5	3	3,5	4	4,5	5	5,5	6	6,5	7	7,5	8	8,5	9	10
effectif	1	2	1	1	1	2	2	3	2	3	4	4	2	2	4	6	2	2	1	1

2. Quelle est la taille de l'échantillon ?
La taille de l'échantillon est $n = 46$.

3. Donner la moyenne, l'écart type et la médiane.

$$\text{moyenne : } m(X) = \frac{\sum_i x_i n_i}{n} = \frac{0 \times 1 + 0,5 \times 2 + 1 \times 1 + \dots + 10 \times 1}{46} = \frac{240,5}{46} \simeq 5,23$$

$$m(X^2) = \frac{\sum_i x_i^2 n_i}{n} = \frac{0^2 \times 1 + 0,5^2 \times 2 + 1^2 \times 1 + \dots + 10^2 \times 1}{46} = \frac{1532,25}{46}$$

$$\text{Var}(X) = m(X^2) - m(X)^2 = \frac{1532,25}{46} - \left(\frac{240,5}{46}\right)^2 \simeq 5,975$$

$$\text{Écart-type : } s(X) = \sqrt{\text{Var}(X)} \simeq 2,44$$

La médiane est la valeur numéro $\frac{46+1}{2} = 23,5$ (en ordonnant par ordre croissant). C'est donc 5,5.

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

classes	[0; 1,5[[1,5; 3[[3; 4,5[[4,5; 6[[6; 7,5[[7,5; 9[[9; 10,5[
effectifs	4	4	7	11	8	10	2
fréquences	0,087	0,087	0,152	0,239	0,174	0,217	0,043
fréquences cumulées	0,087	0,174	0,326	0,565	0,739	0,956	0,999

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

moyenne :

$$\begin{aligned} m(X) &= \frac{\sum_i c_i n_i}{n} \\ &= \frac{0,75 \times 4 + 2,25 \times 4 + 3,75 \times 7 + 5,25 \times 11 + 6,75 \times 8 + 8,25 \times 10 + 9,75 \times 2}{46} \\ &= \frac{252}{46} \\ &\simeq 5,48 \end{aligned}$$

$$\begin{aligned} m(X^2) &= \frac{\sum_i c_i^2 n_i}{n} \\ &= \frac{0,75^2 \times 4 + 2,25^2 \times 4 + 3,75^2 \times 7 + 5,25^2 \times 11 + 6,75^2 \times 8 + 8,25^2 \times 10 + 9,75^2 \times 2}{46} \\ &= \frac{1659,375}{46} \end{aligned}$$

$$\begin{aligned} \text{Var}(X) &= m(X^2) - m(X)^2 \\ &= \frac{1659,375}{46} - \left(\frac{252}{46}\right)^2 \\ &\simeq 6,06 \end{aligned}$$

$$\begin{aligned} \text{Écart-type : } s(X) &= \sqrt{\text{Var}(X)} \\ &\simeq 2,46 \end{aligned}$$

Classe de la médiane : [4,5; 6[

$$\text{Méd} \simeq a_i + \frac{a_{i+1} - a_i}{F_X(a_{i+1}) - F_X(a_i)} (0,5 - F_X(a_i)) \simeq 4,5 + \frac{6 - 4,5}{0,565 - 0,326} (0,5 - 0,326) \simeq 5,59$$

Ce regroupement en classe donne lieu à des erreurs de l'ordre de

0,25 points (dans le cas de la moyenne), ce qui est de l'ordre de 5% de la note.

Exercice 17 : Estime de soi

Voici les scores d'estime de soi évaluée à l'aide d'une version réduite de l'échelle de Coopersmith obtenus par un groupe d'adolescents et un groupe de jeunes adultes :

Adolescents	21	24	32	34	36	24	24	28	38	29	27	24	33	30	29	26
Adultes	41	39	40	46	38	30	37	36	36	48	34	34	30			

Calculer les moyennes et les écarts types des deux groupes.

• Adolescents :

$$\text{moyenne : } m_{\text{ado}}(X) = \frac{\sum x_i}{n} = \frac{21+24+\dots+26}{16} = \frac{459}{16} \simeq 28,69$$

$$m_{\text{ado}}(X^2) = \frac{\sum x_i^2}{n} = \frac{21^2+24^2+\dots+26^2}{16} = \frac{13525}{16}$$

$$\text{Var}_{\text{ado}}(X) = m_{\text{ado}}(X^2) - m_{\text{ado}}(X)^2 = \frac{13525}{16} - \left(\frac{459}{16}\right)^2 \simeq 22,34$$

$$\text{Écart-type : } s_{\text{ado}}(X) = \sqrt{\text{Var}_{\text{ado}}(X)} \simeq 4,73$$

• Adultes :

$$\text{moyenne : } m_{\text{adu}}(X) = \frac{\sum x_i}{n} = \frac{41+39+\dots+30}{13} = \frac{489}{13} \simeq 37,62$$

$$m_{\text{adu}}(X^2) = \frac{\sum x_i^2}{n} = \frac{41^2+39^2+\dots+30^2}{13} = \frac{18739}{13}$$

$$\text{Var}_{\text{adu}}(X) = m_{\text{adu}}(X^2) - m_{\text{adu}}(X)^2 = \frac{18739}{13} - \left(\frac{489}{13}\right)^2 \simeq 26,54$$

$$\text{Écart-type : } s_{\text{adu}}(X) = \sqrt{\text{Var}_{\text{adu}}(X)} \simeq 5,15$$

Exercice 18 : Âge et performances mémorielles

On considère un échantillon de 15 personnes de 37 à 86 ans, auxquels on attribue une note indiquant leurs performances mémorielles. On note leur âge X et leur performances mémorielles Y .

Les données mesurées sont les suivantes :

X	74	86	49	72	74	53	50	45	43	37	86	56	66	51	48
Y	34	21	27	21	23	39	37	48	59	47	22	51	34	25	21

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

• Âge X :

moyenne : $m(X) = \frac{\sum x_i}{n} = \frac{74+86+\dots+48}{15} = \frac{890}{15} \simeq 59,33$ ans

$m(X^2) = \frac{\sum x_i^2}{n} = \frac{74^2+86^2+\dots+48^2}{15} = \frac{56278}{15}$

$Var(X) = m(X^2) - m(X)^2 = \frac{56278}{15} - \left(\frac{890}{15}\right)^2 \simeq 231,42$

Écart-type : $s(X) = \sqrt{Var(X)} \simeq 15,21$ ans

• Performances mémorielles Y :

moyenne : $m(Y) = \frac{\sum x_i}{n} = \frac{34+21+\dots+21}{15} = \frac{509}{15} \simeq 33,93$

$m(Y^2) = \frac{\sum x_i^2}{n} = \frac{34^2+21^2+\dots+21^2}{15} = \frac{19487}{15}$

$Var(Y) = m(Y^2) - m(Y)^2 = \frac{19487}{15} - \left(\frac{509}{15}\right)^2 \simeq 147,66$

Écart-type : $s(Y) = \sqrt{Var(Y)} \simeq 12,15$

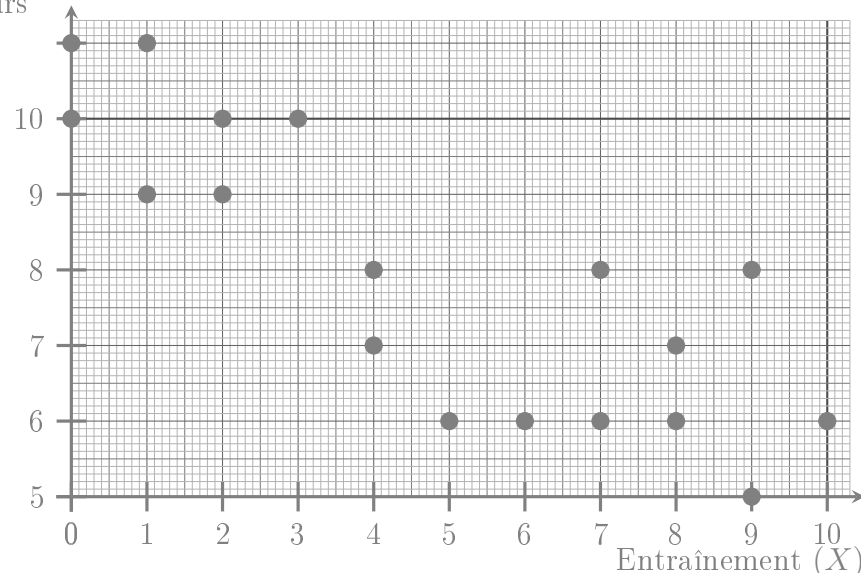
Chapitre 2 : Statistiques descriptives bivariées

Exercice 19 : Entraînement à un exercice de logique

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

Pour un échantillon de 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 du nuage de points suivant (effectué sur un papier "millimétré") :

Nombre d'erreurs (Y)



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

Entraînement X																	
Nombre d'erreurs Y																	

2. Calculer le coefficient de corrélation des rangs de Spearman. Interpréter la valeur de ce coefficient.
3. Calculer les moyennes et écarts type des variables X et Y , puis leur coefficient de corrélation linéaire. Interpréter la valeur de ce coefficient.
4. Si un enfant a déjà fait 8 exercices de logique, alors combien estimeriez qu'il fera d'erreurs si on lui fait à nouveau passer un test similaire ?
5. Un enfant a commis 8 erreurs. Combien de fois estimeriez vous qu'il avait déjà fait un test similaire pour s'entraîner ?

Exercice 20 : 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	K
niveau de stress X	7,82	3,12	3,12	2,86	3,25	3,29	3,29	8,05	2,83	8,05	2,75
temps de réponse Y	698	508	520	526	524	465	485	695	505	668	515

- Déterminer le coefficient de corrélation des rangs de Spearman. Ces variables sont-elles très corrélées ?
- Déterminer le coefficient de corrélation linéaire de ces deux variables. 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 682ms à réagir aux stimuli visuels. Déterminer environ son niveau de stress.
- Tracer le nuage de point des variables X et Y .
- Commenter vos réponses aux questions précédentes en vous appuyant sur le nuage de points.

Exercice 21 : Revenus et espérance de vie

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

pays	revenu/hab. X	espérance de vie Y
Afghanistan	2	61
Belize	8	74
Cambodge	3	64
Danemark	46	81
États-unis	56	79
Fidji	8	70
Gambie	2	59
Honduras	4	74
Israël	33	82
Jordanie	12	75
Kirghizistan	3	71

- Déterminer le coefficient de corrélation linéaire.
- Déterminer le coefficient de corrélation des rangs (de Spearman).
- Commenter (on pourra tracer le nuage de points).

Exercice 22 : Extraversion

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

$$\sum_{i=1}^{120} x_i = 845, \quad \sum_{i=1}^{120} x_i^2 = 6636, \quad \sum_{i=1}^{120} y_i = 448$$

$$\sum_{i=1}^{120} y_i^2 = 2076, \quad \sum_{i=1}^{120} x_i y_i = 3254.$$

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

1. Calculer les moyennes, les écarts types et le coefficient de corrélation des deux échelles X et Y .
2. Serait-il pertinent de résumer ces deux échelles par une échelle d'extraversion ?

Exercice 23 : Reconnaissance de forme

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

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

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

sujet	1	2	3	4	5	6	7	8	9	10	11	12	13	14
X	0	1	1	2	2	4	5	7	7	8	8	8	10	10
Y	3,7	1,9	3,7	3	4	2,6	2,8	1,9	2	1,5	1,7	2,7	1,1	1,4
rang X'	1	2,5	2,5	4,5	4,5	6	7	8,5	8,5	11	11	11	13,5	13,5
rang Y'	12,5	5,5	12,5	11	14	8	10	5,5	7	3	4	9	1	2
$(X' - Y')^2$	132,25	9	100	42,25	90,25	4	9	9	2,25	64	49	4	156,25	132,25

le coefficient de corrélation des rangs de Spearman est donc

$$1 - \left(6 \frac{132,25+9+100+42,25+90,25+\dots+132,25}{14(14^2-1)} \right) \simeq -0,766$$

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

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

moyenne : $m(X) = \frac{\sum x_i}{n} = \frac{0+1+\dots+10}{14} = \frac{73}{14} \simeq 5,21$

$m(X^2) = \frac{\sum x_i^2}{n} = \frac{0^2+1^2+\dots+10^2}{14} = \frac{541}{14}$

$Var(X) = m(X^2) - m(X)^2 = \frac{541}{14} - \left(\frac{73}{14}\right)^2 \simeq 11,45$

Écart-type : $s(X) = \sqrt{Var(X)} \simeq 3,38$

moyenne : $m(Y) = \frac{\sum y_i}{n} = \frac{3,7+1,9+\dots+1,4}{14} = \frac{34}{14} \simeq 2,43$

$m(Y^2) = \frac{\sum y_i^2}{n} = \frac{3,7^2+1,9^2+\dots+1,4^2}{14} = \frac{93,8}{14}$

$Var(Y) = m(Y^2) - m(Y)^2 = \frac{93,8}{14} - \left(\frac{34}{14}\right)^2 \simeq 0,802$

Écart-type : $s(Y) = \sqrt{Var(Y)} \simeq 0,9$

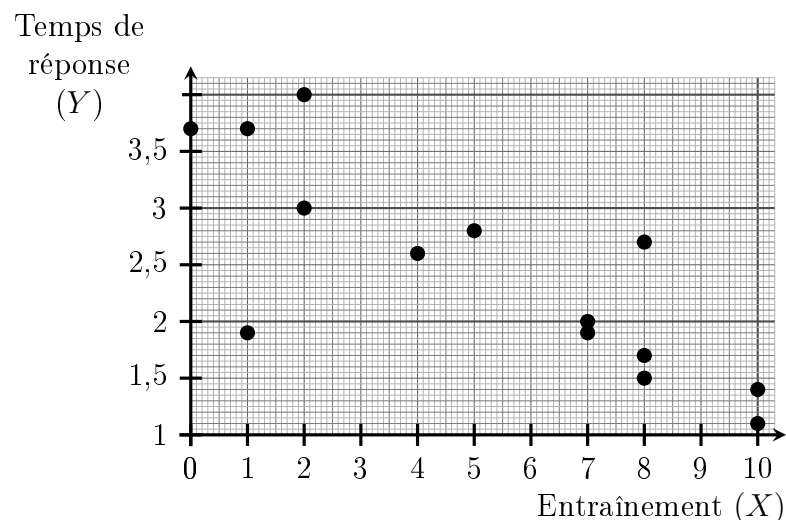
$m(XY) = \frac{\sum x_i y_i}{n} = \frac{0 \times 3,7 + 1,9 + \dots + 10 \times 1,4}{14} = \frac{143,5}{14} = 10,25$

$Cov(X,Y) = m(XY) - m(X)m(Y) = \frac{143,5}{14} - \frac{73}{14} \frac{34,0}{14} \simeq -2,413$

$r(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} = \frac{-2,413}{\sqrt{11,454 \times 0,802}} \simeq -0,796$

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

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



On remarque qu'à l'exception d'un point (le sujet 2), les points sont assez proches d'une droite, comme attendu à partir de leurs coefficients de corrélation.

4. Si un enfant a déjà fait 6 exercices de reconnaissance de forme, alors à combien estimeriez vous son temps de réponse si on lui fait à nouveau passer ce test ?

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

on pose $a = \frac{Cov(X,Y)}{Var(X)} \simeq \frac{-2,413}{11,45} \simeq -0,21$ et $b = m(Y) - a m(X) \simeq 2,429 - (-0,211) \times 5,214 \simeq 3,53$

D'où l'équation de la droite $D_{Y|X} : Y = -0,21 X + 3,53$

Donc pour $x = 6$, on s'attend à $y = -0,21 \times 6 + 3,53 = 2,27$.

5. Un enfant a mis 1,6 secondes pour répondre au test. Combien de fois estimeriez vous qu'il avait déjà fait un test similaire pour s'entraîner ?

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

on pose $a' = \frac{Cov(X,Y)}{Var(Y)} \simeq \frac{-2,41}{0,8} \simeq -3,01$ et $b' = m(X) - a' m(Y) \simeq 5,21 - (-3,01) \times 2,43 \simeq 12,52$

D'où l'équation de la droite $D_{X|Y} : X = -3,01 Y + 12,52$

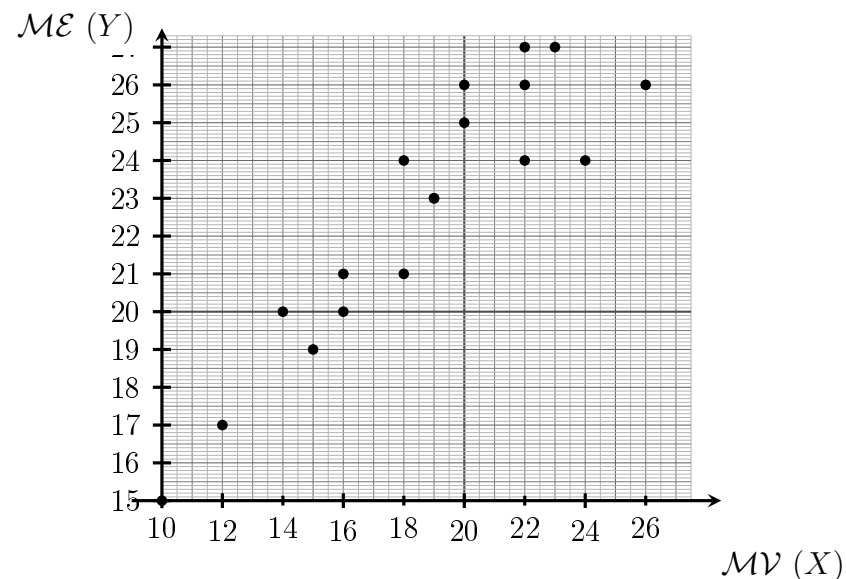
Donc pour $y = 1,6$, on s'attend à $x = -3,01 \times 1,6 + 12,52 = 7,704$

Exercice 24 : Résolution de problèmes mathématiques

On a mis au point deux méthodes d'apprentissage pour la résolution de problèmes mathématiques. La première méthode ($\mathcal{M}\mathcal{V}$) est uniquement verbale, la seconde ($\mathcal{M}\mathcal{E}$) est écrite. Ces méthodes sont testées sur des groupes d'enfants. Le tableau suivant représente les notes X et Y obtenues par ces 18 élèves à deux épreuves relatives à ces apprentissages.

$X : \mathcal{M}\mathcal{V}$	12	19	22	20	20	10	18	16	22	24	15	18	22	23	26	19	16	14
$Y : \mathcal{M}\mathcal{E}$	17	23	24	25	26	15	21	20	27	24	19	24	26	27	26	23	21	20

1. Dessiner le nuage statistique de ces variables.



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

sujet	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
X	12	19	22	20	20	10	18	16	22	24	15	18	22	23	26	19	16	14
Y	17	23	24	25	26	15	21	20	27	24	19	24	26	27	26	23	21	20
rang X'	2	9,5	14	11,5	11,5	1	7,5	5,5	14	17	4	7,5	14	16	18	9,5	5,5	3
rang Y'	2	8,5	11	13	15	1	6,5	4,5	17,5	11	3	11	15	17,5	15	8,5	6,5	4,5
$(X' - Y')^2$	0	1	9	2,25	12,25	0	1	1	12,25	36	1	12,25	1	2,25	9	1	1	2,25

le coefficient de corrélation des rangs de Spearman est donc

$$1 - \left(6 \frac{0+1+9+2,25+12,25+\dots+2,25}{18(18^2-1)} \right) \simeq 0,892$$

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

moyenne : $m(X) = \frac{\sum x_i}{n} = \frac{12+19+\dots+14}{18} = \frac{336}{18} \simeq 18,67$

$m(X^2) = \frac{\sum x_i^2}{n} = \frac{12^2+19^2+\dots+14^2}{18} = \frac{6580}{18}$

$Var(X) = m(X^2) - m(X)^2 = \frac{6580}{18} - \left(\frac{336}{18}\right)^2 \simeq 17,11$

Écart-type : $s(X) = \sqrt{Var(X)} \simeq 4,14$

moyenne : $m(Y) = \frac{\sum x_i}{n} = \frac{17+23+\dots+20}{18} = \frac{408}{18} \simeq 22,67$

$m(Y^2) = \frac{\sum x_i^2}{n} = \frac{17^2+23^2+\dots+20^2}{18} = \frac{9454}{18}$

$Var(Y) = m(Y^2) - m(Y)^2 = \frac{9454}{18} - \left(\frac{408}{18}\right)^2 \simeq 11,44$

Écart-type : $s(Y) = \sqrt{Var(Y)} \simeq 3,38$

$m(XY) = \frac{\sum x_i y_i}{n} = \frac{12 \times 17 + 19 \times 23 + \dots + 14 \times 20}{18} = \frac{7846}{18} \simeq 435,889$

$Cov(X,Y) = m(XY) - m(X)m(Y) = \frac{7846}{18} - \frac{336}{18} \frac{408}{18} \simeq 12,778$

$r(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} = \frac{12,778}{\sqrt{17,111 \times 11,444}} \simeq 0,913$

4. On souhaite désormais estimer à quelle note s'attendre avec la méthode écrite pour un enfant ayant obtenu une note $x = 22,5$ avec la méthode verbale :

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

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

on pose $a = \frac{Cov(X,Y)}{Var(X)} \simeq \frac{12,778}{17,11} \simeq 0,747$ et $b = m(Y) -$

$a m(X) \simeq 22,667 - 0,747 \times 18,667 \simeq 8,723$

D'où l'équation de la droite $D_{Y|X} : Y = 0,747 X + 8,723$

- (b) Si un enfant a obtenu une note $x = 22,5$, donner une estimation de sa note y .

Donc pour $x = 22,5$, on s'attend à $y = 0,747 \times 22,5 + 8,723 \simeq 25,531$.

Exercice 25 : Groupe d'étudiants

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

Guillaume
de bonne humeur
0 frère/soeur
1m82
bac : 10,33/20

Genevieve
de bonne humeur
1 frère/soeur
1m74
bac : 14,83/20

Alain
de mauvaise humeur
1 frère/soeur
1m71
bac : 12,93/20

Justine
de très bonne humeur
1 frère/soeur
1m69
bac : 11,94/20

Xavier
de bonne humeur
2 frères/soeurs
1m94
bac : 16,04/20

Serge
de relativement bonne humeur
0 frère/soeur
1m85
bac : 12,61/20

Gérard
de bonne humeur
0 frère/soeur
1m83
bac : 12,49/20

Marie
de mauvaise humeur
1 frère/soeur
1m72
bac : 14,12/20

Laurent
de bonne humeur
0 frère/soeur
1m86
bac : 11,36/20

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

Pour ces calculs on note X pour la taille des étudiants, et Y pour leur note au bac.

$$\text{moyenne : } m(X) = \frac{\sum x_i}{n} = \frac{1,82+1,74+\dots+1,86}{9} = \frac{16,16}{9} \simeq 1,8$$

$$m(X^2) = \frac{\sum x_i^2}{n} = \frac{1,82^2+1,74^2+\dots+1,86^2}{9} = \frac{29,0732}{9}$$

$$Var(X) = m(X^2) - m(X)^2 = \frac{29,0732}{9} - \left(\frac{16,16}{9}\right)^2 \simeq 0,006$$

$$\text{Écart-type : } s(X) = \sqrt{Var(X)} \simeq 0,08$$

$$\text{moyenne : } m(Y) = \frac{\sum x_i}{n} = \frac{10,33+14,83+\dots+11,36}{9} = \frac{116,65}{9} \simeq 12,96$$

$$m(Y^2) = \frac{\sum x_i^2}{n} = \frac{10,33^2+14,83^2+\dots+11,36^2}{9} = \frac{1537,1041}{9}$$

$$Var(Y) = m(Y^2) - m(Y)^2 = \frac{1537,1041}{9} - \left(\frac{116,65}{9}\right)^2 \simeq 2,8$$

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

$$m(XY) = \frac{\sum x_i y_i}{n} = \frac{1,82 \times 10,33 + 1,74 \times 14,83 + \dots + 1,86 \times 11,36}{9} = \frac{209,6125}{9} \simeq 23,29$$

$$Cov(X,Y) = m(XY) - m(X)m(Y) = \frac{209,6125}{9} - \frac{16,16}{9} \frac{116,65}{9} \simeq 0,018$$

$$r(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} = \frac{0,018}{\sqrt{0,006 \times 2,799}} \simeq 0,139$$

Exercice 26 : Thérapie pour réduire l'anxiété

Pour mettre en évidence l'efficacité d'une thérapie visant à réduire l'anxiété de personnes victimes d'agressions, nous avons observé 18 sujets avant et après la thérapie en affectant à chaque sujet un score (plus le score est élevé, plus fort est le niveau d'anxiété). Les données sont les suivantes.

Avant	39	29	35	15	27	27	23	28	17	43	27	29	25	29	29	24	33	16
Après	25	9	35	5	16	9	11	14	15	41	17	7	21	5	11	20	18	0

- Calculer les coefficients de corrélation linéaire et de Spearman entre les deux phases.

Calcul du coefficient de corrélation linéaire

moyenne : $m(X) = \frac{\sum x_i}{n} = \frac{39+29+\dots+16}{18} = \frac{495}{18} = 27,5$

$m(X^2) = \frac{\sum x_i^2}{n} = \frac{39^2+29^2+\dots+16^2}{18} = \frac{14519}{18}$

$Var(X) = m(X^2) - m(X)^2 = \frac{14519}{18} - \left(\frac{495}{18}\right)^2 \simeq 50,36$

Écart-type : $s(X) = \sqrt{Var(X)} \simeq 7,1$

moyenne : $m(Y) = \frac{\sum x_i}{n} = \frac{25+9+\dots+0}{18} = \frac{279}{18} = 15,5$

$m(Y^2) = \frac{\sum x_i^2}{n} = \frac{25^2+9^2+\dots+0^2}{18} = \frac{6165}{18}$

$Var(Y) = m(Y^2) - m(Y)^2 = \frac{6165}{18} - \left(\frac{279}{18}\right)^2 = 102,25$

Écart-type : $s(Y) = \sqrt{Var(Y)} \simeq 10,11$

$m(XY) = \frac{\sum x_i y_i}{n} = \frac{39 \times 25 + 29 \times 9 + \dots + 16 \times 0}{18} = \frac{8599}{18} \simeq 477,722$

$Cov(X,Y) = m(XY) - m(X)m(Y) = \frac{8599}{18} - \frac{495}{18} \frac{279}{18} \simeq 51,472$

$r(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} = \frac{51,472}{\sqrt{50,361 \times 102,25}} \simeq 0,717$

Coefficient de spearman

sujet	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
X	39	29	35	15	27	27	23	28	17	43	27	29	25	29	29	24	33	16
Y	25	9	35	5	16	9	11	14	15	41	17	7	21	5	11	20	18	0
rang X'	17	12,5	16	1	8	8	4	10	3	18	8	12,5	6	12,5	12,5	5	15	2
rang Y'	16	5,5	17	2,5	11	5,5	7,5	9	10	18	12	4	15	2,5	7,5	14	13	1
(X' - Y') ²	1	49	1	2,25	9	6,25	12,25	1	49	0	16	72,25	81	100	25	81	4	1

le coefficient de corrélation des rangs de Spearman est donc

$1 - \left(6 \frac{1+49+1+2,25+9+\dots+1}{18(18^2-1)}\right) \simeq 0,473$

- Parmi les deux droites de régression, y en a-t-il une qui a plus d'intérêt ? Déterminer son équation.

Celle qui a le plus de sens exprime l'angoisse après ("Y") en fonction de l'angoisse avant ("X"). Elle permet d'estimer, pour un patient donné, quel bénéfice attendre de la thérapie.

on pose $a = \frac{Cov(X,Y)}{Var(X)} \simeq \frac{51,472}{50,36} \simeq 1,022$ et $b = m(Y) - am(X) \simeq 15,5 - 1,022 \times 27,5 \simeq -12,605$

D'où l'équation de la droite $D_{Y|X} : Y = 1,022 X - 12,605$

Chapitre 3 : probabilités

Exercice 27 : Emploi

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

Nom	Mr Rossignol	Mme Hubert	Mr Bonnin	Mme Lemaire	Mme Charles
Situation	retraite	emploi	emploi	chômage	retraite

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

- Lister tous les choix possibles de trois habitants, et conclure que chaque cas a une probabilité de 10%.
- Quelle est la probabilité d'avoir
 - exactement 2 femmes dont une a un emploi ?
 - moins de 2 hommes ?
 - deux personnes en situation d'emploi ?
- On note X le nombre de femmes parmi les trois personnes choisies au hasard.

Déterminer la loi de X (c'est à dire, calculer chaque probabilité $\mathbb{P}[X = \dots]$).

Exercice 28 : Factorielle, coefficients binomiaux

- Simplifier les quantités $\frac{18!}{16!}$, $\frac{30!}{27! \times 3!}$.
- Écrire à l'aide de deux factorielles l'expression ci-contre : $5 \times 6 \times 7 \times 8$.
- Calculer les nombres suivants : $\binom{7}{3}$, $\binom{12}{4}$, $\binom{12}{8}$, $\binom{9}{0}$, $\binom{11}{1}$.

Exercice 29 : Croyance religieuse

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

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

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

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

Calculer la loi de X .

- Au sein d'une ville de 1000 personnes, on constate que 40% (c'est à dire 400 personnes) ont une croyance religieuse. On choisit au hasard 2 personnes au sein de cette ville, et on note X le nombre de personnes athées parmi ces deux personnes choisies au hasard.

Calculer la loi de X .

- Comparer entre elles les lois obtenues dans ces trois situations.

Exercice 30 : Loi binomiale

On considère une loi binomiale $X \sim \mathcal{B}(n, p)$ de moyenne $m(X) = 7,2$ et de variance $\text{Var}(X) = 3,96$.

- Calculer n et p .
- Calculer les probabilités $\mathbb{P}[4 \leq X \leq 7]$ et $\mathbb{P}[X > 13]$.
- Trouver toutes les valeurs de k telles que $\mathbb{P}[X \geq k] \leq 0,05$.

Exercice 31 : Troubles de l'humeur et emploi

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

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

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

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

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

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

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

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

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

Donc à la proportion $\frac{4}{9} \simeq 0,444$.

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

- Retrouver le résultat de la question 2, en utilisant cette loi.

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

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

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

- $\mathbb{P}[X = 2] = \binom{2}{2} (2/3)^2 (1/3)^0 \simeq 0,444$.
2. On décide d'une autre façon de choisir deux noms : on choisit un premier nom au hasard, puis on choisit le second nom en demandant qu'ils soit différent du premier. Reprendre, avec cette nouvelle façon de choisir les deux noms, les questions 1a, 1b et 1c. On prendra garde qu'en question 1a le nombre de cas ne sera plus égal à neuf.
- (a) AB, AC, BA, BC, CA et CB.
- (b) Il y en a 2 : AB et BA.
- (c) Cela correspond aux 4 possibilités AC, BC, CA et CB. Donc à la proportion $\frac{4}{6} \simeq 0,667$.

Exercice 32 : Collection de dessins

Mr Morin 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. Il mélange les cartes, et constate dans quelle ordre elles se retrouvent à l'issue du mélange. Quelle était la probabilité qu'à l'issue du mélange, elles se retrouvent précisément dans cet ordre ? Il y a $5! = 5 \times 4 \times 3 \times 2 = 120$ résultats possibles à l'issue du mélange. La probabilité est donc $\frac{1}{120} \simeq 0,008$.
2. S'il remélange les cartes, quelle est la probabilité qu'après le mélange
- (a) la carte qui était en haut du tas soit à nouveau en haut du tas ?
- Il y a $4! = 4 \times 3 \times 2 = 24$ 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 qu'avant ?
- $\frac{1}{120} \simeq 0,008$.
3. Répondre aux mêmes questions en supposant qu'il y ait cette fois-ci 32 cartes.
1. $\frac{1}{32!} = \frac{1}{32 \times 31 \times \dots \times 1} \simeq 3,8 \times 10^{-36}$
2. (a) $\frac{31!}{32!} = \frac{31 \times 30 \times \dots \times 1}{32 \times 31 \times \dots \times 1} = \frac{1}{32} \simeq 0,031$
- (b) $\frac{30!}{32!} = \frac{30 \times 29 \times \dots \times 1}{32 \times 31 \times \dots \times 1} = \frac{1}{32 \times 31} \simeq 0,001$
- (c) $\frac{1}{32 \times 31} \simeq 0,001$
- (d) $\frac{1}{32 \times 31} + \frac{1}{32 \times 31} \simeq 0,002$
- (e) $\frac{1}{32!} = \frac{1}{32 \times 31 \times \dots \times 1} \simeq 3,8 \times 10^{-36}$
4. Peut-on conclure des probabilités calculées que certains événements ont particulièrement peu de chances de se produire ?
- Les probabilités les plus faibles sont celles des questions 1 et 2e

(probabilité déjà très faible pour 5 cartes, et quasi-impossible pour 32 cartes).

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

En revanche, dans la question 2e, on peut vraiment déduire quelque-chose : quand on remélange un paquet de cartes il est très improbable (pour 5 cartes) et même quasiment impossible (pour 32 cartes) de reproduire le même ordre qu'initialement.

Exercice 33 : Ressources humaines

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

On choisit parmi les employés de l'entreprise, un échantillon de 25 employés au hasard "avec remise".

1. Calculer la probabilité que l'échantillon contienne exactement 3 employés notés A .

Notons X le nombre de personnes notées A dans l'échantillon. Alors $X \sim \mathcal{B}(25; 0,2)$, car on a choisi 25 employés (avec remise) parmi une population dont 20% ont la note A .

$$\begin{aligned}\mathbb{P}[X = 3] &= \binom{25}{3} (0,2)^3 (1 - 0,2)^{25-3} \\ &\simeq 0,1358\end{aligned}$$

2. Quelle est la probabilité que l'échantillon contienne au moins 3 individus notés A ?

$$\begin{aligned}\mathbb{P}[X \geq 3] &= 1 - \mathbb{P}[0 \leq X \leq 2] = 1 - \mathbb{P}[X = 0] - \mathbb{P}[X = 1] - \mathbb{P}[X = 2] \\ &= 1 - \binom{25}{0} (0,2)^0 (1 - 0,2)^{25} - \dots - \binom{25}{2} (0,2)^2 (1 - 0,2)^{25-2} \\ &\simeq 1 - 0,004 - 0,024 - 0,071 \simeq 0,9\end{aligned}$$

3. Quelle probabilité y a-t-il que l'échantillon contienne au moins 12 individus notés A ou B ?

On procède de même en notant cette fois ci Y le nombre d'individus notés A ou B dans l'échantillon. On a $Y \sim \mathcal{B}(25; 0,5)$.

$$\begin{aligned}\mathbb{P}[Y \geq 12] &= 1 - \mathbb{P}[0 \leq Y \leq 11] = 1 - \mathbb{P}[Y = 0] - \mathbb{P}[Y = 1] - \mathbb{P}[Y = 2] - \dots - \mathbb{P}[Y = 11] \\ &= 1 - \binom{25}{0} (0,5)^0 (1 - 0,5)^{25} - \dots - \binom{25}{11} (0,5)^{11} (1 - 0,5)^{25-11} \\ &\simeq 1 - 0 - 0 - 0 - 0 - 0 - 0 - 0,002 - \dots - 0,133 \simeq 0,65\end{aligned}$$

Remarque : le moyen le plus rapide d'obtenir le résultat est de calculer $\mathbb{P}[Y \leq 11] \simeq 0,345$ avec la calculette (voir explications dans le formulaire), puis de déduire $\mathbb{P}[Y \geq 12] = 1 - \mathbb{P}[Y \leq 11] \simeq 0,655$.

4. Calculer la probabilité que l'échantillon contienne moins que 8 personnes notées C .

On procède de même en notant cette fois ci Z le nombre d'individus notés C dans l'échantillon. On a $Z \sim \mathcal{B}(25; 0,5)$.

$$\begin{aligned}\mathbb{P}[Z < 8] &= \mathbb{P}[Z = 0] + \mathbb{P}[Z = 1] + \mathbb{P}[Z = 2] + \dots + \mathbb{P}[Z = 7] \\ &= \binom{25}{0} (0,5)^0 (1 - 0,5)^{25} + \dots + \binom{25}{7} (0,5)^7 (1 - 0,5)^{25-7} \\ &\simeq 0 + 0 + 0 + 0 + 0 + 0,002 + 0,005 + 0,014 \simeq 0,02\end{aligned}$$

5. Quel est le nombre moyen d'individus notés B au sein d'un échantillon aléatoire de 25 employés ?

Si on note W le nombre de personnes notées B , alors $W \sim \mathcal{B}(25; 0,3)$, donc $m(W) = 25 \times 0,3 = 7,5$.

Exercice 34 : Thérapie contre la dépression

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

En interrogeant un échantillon de 20 patients, on constate 8 rémissions.

- 1) S'il y avait 70% de rémission parmi l'ensemble des patients, quelle loi suivrait le nombre de rémissions au sein d'un échantillon de 20 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 20 patients, et si l'on suppose qu'il y ait 70% de rémission parmi l'ensemble des patients, alors X suit la loi $\mathcal{B}(20; 0,7)$.

Le nombre moyen de rémissions au sein d'un tel échantillon serait alors $m(X) = 20 \times 0,7 = 14$.

- 2) Sous cette hypothèse, quelle serait la probabilité d'avoir au maximum 8 rémissions ?

$$\begin{aligned} \mathbb{P}[X \leq 8] &= \mathbb{P}[X = 0] + \mathbb{P}[X = 1] + \mathbb{P}[X = 2] + \dots + \mathbb{P}[X = 8] \\ &= \binom{20}{0} (0,7)^0 (1 - 0,7)^{20} + \dots + \binom{20}{8} (0,7)^8 (1 - 0,7)^{20-8} \\ &\simeq 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0,001 + 0,004 \simeq 0,01 \end{aligned}$$

Remarque : Encore une fois, la calculette permet de le calculer directement (cf formulaire).

- 3) Conclure : vous semble-t-il vraisemblable qu'il y ait, comme l'affirme ce psychologue, 70% de rémissions parmi l'ensemble des patients ? On a obtenu que s'il y avait bien 70% de rémission parmi l'ensemble des patients (comme l'affirme ce psychologue), alors il y aurait très peu de chances (environ 1% de chances) d'avoir un échantillon où il n'y ait pas plus de 8 rémissions. Le fait que nous soyons tombé sur un tel échantillon donne une forte présomption que le taux de rémissions n'est en fait pas de 70% (à moins par exemple que notre échantillon soit biaisé à cause de la façon dont il a été établi).

Exercice 35 : Nombre de filles et de garçons

Dans une classe, il y a 16 filles et 12 garçons. On choisit au hasard 4 élèves distincts.

1. Quel est le nombre de choix possibles ?

On choisit 4 personnes parmi l'ensemble des $16 + 12 = 28$ élèves. Le nombre de choix possibles est donc

$$\begin{aligned} \binom{28}{4} &= \frac{28!}{4! \times 24!} = \frac{28 \times 27 \times \dots \times 1}{(4 \times 3 \times 2) \times (24 \times 23 \times \dots \times 1)} \\ &= (28 \times 27 \times 26 \times 25) \times \frac{1}{4 \times 3 \times 2} = \frac{4 \times 7 \times 27 \times 26 \times 25}{4 \times 3 \times 2} \\ &= \frac{7 \times 27 \times 26 \times 25}{3 \times 2} = \frac{7 \times 3 \times 9 \times 26 \times 25}{3 \times 2} \\ &= \frac{7 \times 9 \times 26 \times 25}{2} = \frac{7 \times 9 \times 2 \times 13 \times 25}{2} \\ &= 7 \times 9 \times 13 \times 25 = 20475 \end{aligned}$$

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

Ces choix correspondent à choisir 4 élèves parmi les 12 garçons. Le nombre de choix correspondant est donc

$$\begin{aligned} \binom{12}{4} &= \frac{12!}{4! \times 8!} = \frac{12 \times 11 \times \dots \times 1}{(4 \times 3 \times 2) \times (8 \times 7 \times \dots \times 1)} = (12 \times 11 \times 10 \times 9) \times \frac{1}{4 \times 3 \times 2} \\ &= \frac{4 \times 3 \times 11 \times 10 \times 9}{4 \times 3 \times 2} = \frac{11 \times 10 \times 9}{2} = \frac{11 \times 2 \times 5 \times 9}{2} \\ &= 11 \times 5 \times 9 = 495 \end{aligned}$$

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

Ces choix correspondent à choisir 4 élèves parmi les 16 filles. Le nombre de choix correspondant est donc

$$\begin{aligned} \binom{16}{4} &= \frac{16!}{4! \times 12!} = \frac{16 \times 15 \times \dots \times 1}{(4 \times 3 \times 2) \times (12 \times 11 \times \dots \times 1)} \\ &= (16 \times 15 \times 14 \times 13) \times \frac{1}{4 \times 3 \times 2} = \frac{4 \times 4 \times 15 \times 14 \times 13}{4 \times 3 \times 2} \\ &= \frac{4 \times 15 \times 14 \times 13}{3 \times 2} = \frac{4 \times 3 \times 5 \times 14 \times 13}{3 \times 2} \\ &= \frac{4 \times 5 \times 14 \times 13}{2} = \frac{2 \times 2 \times 5 \times 14 \times 13}{2} \\ &= 2 \times 5 \times 14 \times 13 = 1820 \end{aligned}$$

Cela correspond donc à la probabilité $\frac{1820}{20475} \simeq 0,089$.

4. Quelle est la probabilité, en choisissant ainsi un échantillon de 4 élèves au hasard, que cet échantillon comporte 3 filles et 1 garçon ?

Les cas correspondant sont formés d'une part d'un groupe de 3 filles (pour lequel il y a $\binom{16}{3}$ 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 $\binom{14}{1}$ choix possibles). En conséquence, le nombre de cas qui correspondent est :

$$\begin{aligned}
 \binom{16}{3} \times \binom{12}{1} &= \frac{16!}{3! \times 13!} \times \frac{12!}{1! \times 11!} \\
 &= \frac{16 \times 15 \times \dots \times 1}{(3 \times 2 \times 1) \times (13 \times 12 \times \dots \times 1)} \times \frac{12 \times 11 \times \dots \times 1}{11 \times 10 \times \dots \times 1} \\
 &= (16 \times 15 \times 14) \times \frac{1}{3 \times 2 \times 1} \times 12 \\
 &= \frac{16 \times 3 \times 5 \times 14}{3 \times 2} \times 12 \\
 &= \frac{16 \times 5 \times 14}{2} \times 12 \\
 &= \frac{2 \times 8 \times 5 \times 14}{2} \times 12 \\
 &= 8 \times 5 \times 14 \times 12 \\
 &= 560 \times 12 \\
 &= 6720
 \end{aligned}$$

D'où la probabilité $\frac{6720}{20475} \simeq 0,328$.

5. Quel est le nombre de choix comportant au plus un garçon ?
 On additionne, d'une part le nombre de choix comportant 0 garçons (calculé en question 3) et d'autre part le nombre de choix comportant exactement 1 garçon (calculé en question 4). On obtient donc $1\,820 + 6\,720 = 8\,540$.

Chapitre 4 : Loi Normale

Exercice 36 : Lecture de la table et utilisation de la calculatrice

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

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

1. Pour cette question, on pose $z_1 = 1,15$; $z_2 = 1,37$; $z_3 = 1,116$ et $z_4 = 1,227$.
- (a) Déterminer $F(z_1)$, $F(z_2)$, $F(z_3)$ et $F(z_4)$. Chacune de ces probabilités sera calculée à partir de la table du formulaire, puis obtenue dans un second temps en utilisant les fonctions spécifiques de la calculatrice.

- (b) Calculer : $\mathbb{P}[z_1 < Z < z_2]$; $\mathbb{P}[-z_2 < Z < -z_1]$; $\mathbb{P}[z_1 < Z < 5]$; $\mathbb{P}[-z_2 < Z < z_1]$; $\mathbb{P}[-z_1 < Z < z_1]$; $\mathbb{P}[Z > z_2]$; $\mathbb{P}[Z < z_1]$; $\mathbb{P}[Z < -z_1]$. Calculer ces probabilités à partir des valeurs obtenues dans la question a) puis vérifier le résultat en utilisant directement les fonctions spécifiques de la calculatrice.
2. (a) Déterminer des valeurs approchées de z_1 , z_2 , et z_3 tels que $F(z_1) = 0,591$; $F(z_2) = 0,93$ et $F(z_3) = 0,37$.
- (b) Déterminer des valeurs approchées de z_1 , z_2 , et z_3 tels que $\mathbb{P}[Z \leq z_1] = 0,98$; $\mathbb{P}[Z > z_2] = 0,91$; $\mathbb{P}[Z < z_3] = 0,65$.

Exercice 37 : Loi normale avec moyenne et écart type

1. Pour une loi normale centrée réduite, quelle est la probabilité qu'une observation soit :
- (a) supérieure ou égale à 1,5 ? (c) comprise entre -2 et 2 ?
 (b) comprise entre -1 et 1 ? (d) comprise entre -3 et 3 ?
2. Pour une loi normale de moyenne $\mu = 123$ et d'écart-type $\sigma = 9$, (c'est-à-dire $X \rightsquigarrow \mathcal{N}(123; 9)$), quelle est la probabilité qu'une observation soit :
- (a) supérieure ou égale à $\mu + 1,5\sigma$?
 (b) comprise entre $\mu - \sigma$ et $\mu + \sigma$?
 (c) comprise entre $\mu - 2\sigma$ et $\mu + 2\sigma$?
 (d) comprise entre $\mu - 3\sigma$ et $\mu + 3\sigma$?
3. Reprendre les questions précédentes avec la loi normale $\mathcal{N}(12,53; 1,34)$. Que remarque-t-on ?

Exercice 38 : Images mentales

Dans une expérience sur les images mentales, on demande à un échantillon de 279 enfants d'apprendre une liste de 40 mots en leur montrant, pour chaque mot, un dessin de l'objet correspondant au mot. Soit X la variable statistique égale au nombre de mots mémorisés par chaque enfant un jour après. Les effectifs observés des classes suivantes $] 5 ; 10 [$, $] 10 ; 15 [$, $] 15 ; 20 [$, $] 20 ; 25 [$ et $] 25 ; 30 [$ sont respectivement 3, 37, 113, 114 et 12.

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

Exercice 39 : 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 suisses, une moyenne $\mu \simeq 10,16$ et un écart-type $\sigma \simeq 4,32$.

Dans la suite, on effectue des calculs en supposant que X suit la loi normale $\mathcal{N}(10,16; 4,32)$.

1. Calculer la proportion théorique des scores supérieurs à 13.

$$\begin{aligned} \mathbb{P}[X > 13] &= \mathbb{P}\left[\frac{13-10,16}{4,32} < \frac{X-10,16}{4,32}\right] \simeq \mathbb{P}[0,6574 < Z] \\ &\simeq 1 - F(0,6574) \simeq 1 - 0,745 \simeq 0,255 \end{aligned}$$

2. Trouver la valeur de a telle que $\mathbb{P}[X < a] \simeq 0,95$.

Pour résoudre l'équation $\mathbb{P}[X \leq a] \simeq 0,95$, on note tout d'abord que

$$\mathbb{P}[X \leq a] = \mathbb{P}\left[Z \leq \frac{a-10,16}{4,32}\right] = F\left(\frac{a-10,16}{4,32}\right)$$

d'où l'équation $F\left(\frac{a-10,16}{4,32}\right) = 0,95$, c'est à dire $F(z) = 0,95$, où on note $z = \frac{a-10,16}{4,32}$.

$$F(1,64) \simeq 0,9495 \text{ et } F(1,65) \simeq 0,9505.$$

Donc $1,64 \leq z \leq 1,65$, que l'on peut par exemple arrondir à $z \simeq 1,64$.

D'où $\frac{a-10,16}{4,32} = z = 1,64$, qui donne $a \simeq z \times 4,32 + 10,16 \simeq z \times 4,32 + 10,16 \simeq 17,2$.

3. Trouver les quartiles de la variable X .

- Pour résoudre l'équation $\mathbb{P}[X \leq Q_1] \simeq 0,25$, on note tout d'abord que

$$\mathbb{P}[X \leq Q_1] = \mathbb{P}\left[Z \leq \frac{Q_1-10,16}{4,32}\right] = F\left(\frac{Q_1-10,16}{4,32}\right)$$

d'où l'équation $F\left(\frac{Q_1-10,16}{4,32}\right) = 0,25$, c'est à dire $F(z) = 0,25$, où on note $z = \frac{Q_1-10,16}{4,32}$.

Comme $F(z) \simeq 0,25 < 0,5$, on sait que $z < 0$.

On commence donc par résoudre l'équation $F(|z|) \simeq 0,75$:

$$F(0,67) \simeq 0,7486 \text{ et } F(0,68) \simeq 0,7517.$$

Donc $0,67 \leq |z| \leq 0,68$, que l'on peut par exemple arrondir à $|z| \simeq 0,67$.

D'où $z \simeq -0,67$.

D'où $\frac{Q_1-10,16}{4,32} = z = (-0,67)$, qui donne $Q_1 \simeq z \times 4,32 + 10,16 \simeq z \times 4,32 + 10,16 \simeq 7,3$.

- Pour résoudre l'équation $\mathbb{P}[X \leq Q_1] \simeq 0,75$, on note tout d'abord que

$$\mathbb{P}[X \leq Q_1] = \mathbb{P}\left[Z \leq \frac{Q_1-10,16}{4,32}\right] = F\left(\frac{Q_1-10,16}{4,32}\right)$$

d'où l'équation $F\left(\frac{Q_1-10,16}{4,32}\right) = 0,75$, c'est à dire $F(z) = 0,75$,
où on note $z = \frac{Q_1-10,16}{4,32}$.

$F(0,67) \simeq 0,7486$ et $F(0,68) \simeq 0,7517$.

Donc $0,67 \leq z \leq 0,68$, que l'on peut par exemple arrondir à $z \simeq 0,67$.

D'où $\frac{Q_1-10,16}{4,32} = z = 0,67$, qui donne $Q_1 \simeq z \times 4,32 + 10,16 \simeq z \times 4,32 + 10,16 \simeq 13,1$.

4. Quel est le score minimal des 20 % de ceux qui ont les scores D.S. les plus élevés ?

Pour résoudre l'équation $\mathbb{P}[X \geq m] \simeq 0,2$, on note tout d'abord que

$$\mathbb{P}[X \geq m] = \mathbb{P}\left[Z \geq \frac{m-10,16}{4,32}\right] = 1 - F\left(\frac{m-10,16}{4,32}\right)$$

d'où l'équation $1 - F\left(\frac{m-10,16}{4,32}\right) = 0,2$, c'est à dire $F(z) = 0,8$, où on note $z = \frac{m-10,16}{4,32}$.

$F(0,84) \simeq 0,7995$ et $F(0,85) \simeq 0,8023$.

Donc $0,84 \leq z \leq 0,85$, que l'on peut par exemple arrondir à $z \simeq 0,84$.

D'où $\frac{m-10,16}{4,32} = z = 0,84$, qui donne $m \simeq z \times 4,32 + 10,16 \simeq z \times 4,32 + 10,16 \simeq 13,8$.

Donc le score minimal des 20% qui ont le score D.S. le plus élevés est 13,8.

5. Quel est le score maximal des 35 % de ceux qui ont les scores D.S. les moins élevés ?

Pour résoudre l'équation $\mathbb{P}[X \leq M] \simeq 0,35$, on note tout d'abord que

$$\mathbb{P}[X \leq M] = \mathbb{P}\left[Z \leq \frac{M-10,16}{4,32}\right] = F\left(\frac{M-10,16}{4,32}\right)$$

d'où l'équation $F\left(\frac{M-10,16}{4,32}\right) = 0,35$, c'est à dire $F(z) = 0,35$, où on note $z = \frac{M-10,16}{4,32}$.

Comme $F(z) \simeq 0,35 < 0,5$, on sait que $z < 0$.

On commence donc par résoudre l'équation $F(|z|) \simeq 0,65$:

$F(0,38) \simeq 0,6480$ et $F(0,39) \simeq 0,6517$.

Donc $0,38 \leq |z| \leq 0,39$, que l'on peut par exemple arrondir à $|z| \simeq 0,39$.

D'où $z \simeq -0,39$.

D'où $\frac{M-10,16}{4,32} = z = (-0,39)$, qui donne $M \simeq z \times 4,32 + 10,16 \simeq z \times 4,32 + 10,16 \simeq 8,48$.

Donc le score maximal des 35% qui ont le score D.S. le moins élevés est 8,48.

Exercice 40 :

Lors d'un concours le pourcentage d'échecs est de 25%. 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.

- On prend $n = 18$. Calculer $p_k = \mathbb{P}[S_n \leq k]$ pour $k = 0; 1; 2; 3$.
- Pour $n = 150$, on souhaite calculer $\mathbb{P}[30 \leq S_n \leq 42]$:
 - Calculer $\mathbb{P}[30 \leq S_n \leq 42]$ avec la calculette.
 - Argumenter que l'on peut approcher la loi de S_n par une loi normale, et calculer $\mathbb{P}[30 \leq S_n \leq 42]$ de manière approchée.
On vérifiera que la précision de l'approximation est bien meilleure si on fait la *correction de continuité*.
- Pour $n = 200$, déterminer de même $\mathbb{P}[44 \leq S_n \leq 75]$ (on pourra utiliser l'approximation par une loi normale en faisant une correction de continuité).
- Pour $n = 150$, trouver une valeur de a telle que $\mathbb{P}[S_{150} < a] \simeq 0,03$ (pour simplifier les calculs, on pourra omettre la correction de continuité dans cette question).

Exercice 41 : Indice de masse corporelle dans la population française

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

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

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

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

1. Déterminer la probabilité (« *fréquence théorique* ») qu'aurait chaque classe si X suivait la loi normale $\mathcal{N}(25; 5)$.

Si X suit la loi $\mathcal{N}(25; 5)$, on obtient les probabilités suivantes :

- $\mathbb{P}[X \leq 18,5] = \mathbb{P}[\frac{X-25}{5} < \frac{18,5-25}{5}] \simeq \mathbb{P}[Z < -1,3]$
 $\simeq F(-1,3) \simeq 1 - 0,9032$
 $\simeq 0,0968$
- $\mathbb{P}[18,5 \leq X \leq 25] = \mathbb{P}[\frac{18,5-25}{5} < \frac{X-25}{5} < \frac{25-25}{5}] = \mathbb{P}[-1,3 < Z < 0,0]$
 $= F(0) - F(-1,3) \simeq 0,5000 - (1 - 0,9032)$
 $\simeq 0,5 - 0,0968 \simeq 0,4032$
- $\mathbb{P}[25 \leq X \leq 30] = \mathbb{P}[\frac{25-25}{5} < \frac{X-25}{5} < \frac{30-25}{5}] = \mathbb{P}[0,0 < Z < 1,0]$
 $= F(1) - F(0) \simeq 0,8413 - 0,5$
 $\simeq 0,3413$
- $\mathbb{P}[30 \leq X \leq 35] = \mathbb{P}[\frac{30-25}{5} < \frac{X-25}{5} < \frac{35-25}{5}] = \mathbb{P}[1,0 < Z < 2,0]$
 $= F(2) - F(1) \simeq 0,9772 - 0,8413$
 $\simeq 0,1359$

- $\mathbb{P}[35 \leq X \leq 40] = \mathbb{P}[\frac{35-25}{5} < \frac{X-25}{5} < \frac{40-25}{5}] = \mathbb{P}[2,0 < Z < 3,0]$
 $= F(3) - F(2) \simeq 0,9987 - 0,9772$
 $\simeq 0,0215$
- $\mathbb{P}[X \geq 40] = \mathbb{P}[\frac{40-25}{5} < \frac{X-25}{5}] \simeq \mathbb{P}[3,0 < Z]$
 $\simeq 1 - F(3) \simeq 1 - 0,9987$
 $\simeq 0,0013$

D'où les fréquences théoriques suivantes :

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

2. En particulier, si X suit la loi normale $\mathcal{N}(25; 5)$, quelle est la proportion (*théorique*) d'obésité sévère ou morbide (c'est à dire $\mathbb{P}[X \geq 35]$) ?

En particulier la proportion théorique d'obésité sévère ou morbide serait la somme des fréquences théoriques des deux dernières classes, c'est à dire $\mathbb{P}[X \geq 35] \simeq 0,0215 + 0,0013 \simeq 0,023$.

3. Au contraire, quelle est dans l'échantillon de cette étude épidémiologique, la proportion d'obésité sévère ou morbide (c'est à dire $\mathbb{P}_r[X \geq 35]$) ?

La proportion d'obésité sévère ou morbide est la somme des fréquences des deux dernières classes, c'est à dire $3,1 + 1,2 \simeq 4,3\%$.

4. On cherche désormais à comparer les résultats des questions 2 et 3 : on souhaite déterminer si la proportion $\mathbb{P}_r[X \geq 35]$ est compatible avec la probabilité $\mathbb{P}[X \geq 35]$ obtenue sous l'hypothèse d'une loi normale.

- (a) Si on suppose qu'une personne choisie au hasard a une probabilité de 2,3% d'être en obésité sévère ou morbide, quelle est la loi du nombre de personnes en obésité sévère ou morbide 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 ($50\,000\,000 >$

10×25714) donc le fait d'être "avec" ou "sans remise" n'a aucun impact. On peut donc utiliser la loi binomiale $\mathcal{B}(25714; 0,023)$.

- (b) Cette loi peut-elle être approximée par une loi normale, et si oui laquelle ?

On a $n = 25714 > 30$, et $np \simeq 591 > 5$, et $n(1-p) \simeq 25120 > 5$ donc on peut approximer *la loi binomiale* par $\mathcal{N}(591; 24)$ où $24 \simeq \sqrt{np(1-p)}$

- (c) En utilisant cette approximation par une loi normale, calculer la probabilité qu'un échantillon aléatoire de 25 714 personnes contienne au moins 4,3% de personnes en obésité sévère ou morbide.

On note X le nombre de personnes en obésité sévère ou morbide dans un échantillon aléatoire de 25714 français. Une proportion de 4,3% de personnes correspondrait à $25714 \times 0,043 \simeq 1105,7$ personnes, donc on calcule $\mathbb{P}[X \geq 1106]$:

$$\begin{aligned} \mathbb{P}[X \geq 1106] &= \mathbb{P}[X \geq 1105,5] = \mathbb{P}\left[\frac{1105,5-591,42}{24,038} < \frac{X-591,42}{24,038}\right] \\ &\simeq \mathbb{P}[21,3861 < Z] \simeq 1 - F(21,3861) \\ &\simeq 1 - 1 \simeq 0 \end{aligned}$$

- (d) Comparer avec l'échantillon de l'étude épidémiologique mentionnée en début d'exercice.

Conclure qu'en 2012, l'IMC des français ne suivait pas la loi de probabilité $\mathcal{N}(25; 5)$.

Donc si l'IMC des français avait vraiment suivi la loi $\mathcal{N}(25; 5)$, alors il n'y aurait eu aucune chance qu'un échantillon aléatoire de 25714 français ait au moins 4,3% de personnes en obésité sévère ou morbide. L'échantillon aléatoire de cette étude épidémiologique a justement 4,3% de personnes en obésité sévère ou morbide (ce qui, comme on vient de le dire, n'aurait pu arriver si l'IMC des français avait vraiment suivi la loi $\mathcal{N}(25; 5)$). Donc l'IMC des français ne suivait pas la loi $\mathcal{N}(25; 5)$.

Exercice 42 :

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

Si l'on considère qu'une note supérieure à 70 traduit un état pathologique, combien s'attend-on à trouver de personnes pathologiquement dépressives sur un ensemble de 50 000 personnes ?

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

$$\begin{aligned} \mathbb{P}[X \geq 70] &= \mathbb{P}\left[\frac{70-50}{10} < \frac{X-50}{10}\right] \simeq \mathbb{P}[2,0 < Z] \\ &\simeq 1 - F(2) \simeq 1 - 0,9772 \simeq 0,0228 \end{aligned}$$

Donc parmi un échantillon de 50000 personnes, on s'attend à trouver environ $50000 \times 0,0228 = 1140$ personnes pathologiquement dépressives.

Exercice 43 :

La proportion des enfants qui réussissent parfaitement une épreuve graphique d'organisation perceptive est de 60%. On choisit au hasard (avec remise) un échantillon de n individus et on note X la variable aléatoire égale au nombre d'enfants dans l'échantillon qui réussissent l'épreuve.

1. On prend $n = 28$. Préciser la loi de probabilité de X et calculer la probabilité : $\mathbb{P}[15 \leq X \leq 18]$.

$$X \sim \mathcal{B}(28; 0,6).$$

$$\mathbb{P}[15 \leq X \leq 18] = \mathbb{P}[X \leq 18] - \mathbb{P}[X \leq 14] \simeq 0,741 - 0,187 \simeq 0,554$$

2. On prend $n = 280$.

- (a) On peut approcher la loi de X par une loi normale. Justifier et préciser laquelle. Dans la suite, si vous utilisez cette approximation, il faudra faire une correction de continuité.

$$X \sim \mathcal{B}(280; 0,6).$$

On a $n = 280 > 30$, et $np = 168 > 5$, et $n(1 - p) = 112 > 5$ donc on peut approximer X par $\mathcal{N}(168; 8,2)$ où $8,2 \simeq \sqrt{np(1 - p)}$

- (b) Calculer les probabilités $\mathbb{P}[X \geq 180]$ et $\mathbb{P}[150 \leq X \leq 185]$.

$$\begin{aligned} \mathbb{P}[X \geq 180] &= \mathbb{P}[X \geq 179,5] = \mathbb{P}\left[\frac{179,5-168}{8,2} < \frac{X-168}{8,2}\right] \\ &\simeq \mathbb{P}[1,4 < Z] \simeq 1 - F(1,4) \\ &\simeq 1 - 0,9192 \simeq 0,08 \end{aligned}$$

$$\begin{aligned} \mathbb{P}[150 \leq X \leq 185] &= \mathbb{P}[149,5 \leq X \leq 185,5] \\ &= \mathbb{P}\left[\frac{149,5-168}{8,2} < \frac{X-168}{8,2} < \frac{185,5-168}{8,2}\right] \\ &\simeq \mathbb{P}[-2,26 < Z < 2,13] \\ &\simeq F(2,13) - F(-2,26) \\ &\simeq 0,9834 - (1 - 0,9881) \\ &\simeq 0,9834 - 0,0119 \\ &\simeq 0,97 \end{aligned}$$

Exercice 44 :

On considère que le nombre de potentiels d'action (PA) émis en une minute par un neurone au repos suit une loi normale de moyenne 100 et d'écart-type 10.

1. Quelle est la probabilité pour qu'un neurone émette plus de 110 PA en 1 mn s'il n'est pas activé ?

On note X le nombre de PA émis en une minute. Si le neurone n'est pas activé on a $X \sim \mathcal{N}(100; 10)$.

$$\begin{aligned} \mathbb{P}[X \geq 110] &= \mathbb{P}\left[\frac{110-100}{10} < \frac{X-100}{10}\right] \simeq \mathbb{P}[1,0 < Z] \\ &\simeq 1 - F(1) \simeq 1 - 0,8413 \simeq 0,16 \end{aligned}$$

2. Quelle est la probabilité pour qu'un neurone émette moins de 80 PA en 1 mn s'il n'est pas activé ?

$$\begin{aligned} \mathbb{P}[X \leq 80] &= \mathbb{P}\left[\frac{X-100}{10} < \frac{80-100}{10}\right] \simeq \mathbb{P}[Z < -2,0] \\ &\simeq F(-2) \simeq 1 - 0,9772 \simeq 0,02 \end{aligned}$$

3. Quelle est la probabilité pour qu'un neurone émette entre 90 et 110 PA en 1 mn s'il n'est pas activé ?

$$\begin{aligned} \mathbb{P}[90 \leq X \leq 110] &= \mathbb{P}\left[\frac{90-100}{10} < \frac{X-100}{10} < \frac{110-100}{10}\right] \\ &= \mathbb{P}[-1,0 < Z < 1,0] \\ &= F(1) - F(-1) \\ &\simeq 0,8413 - (1 - 0,8413) \\ &\simeq 0,8413 - 0,1587 \\ &\simeq 0,68 \end{aligned}$$

4. Un neurone émet plus de 150 PA en une minute. Selon vous, est-il plus vraisemblable que ce neurone était activé ou au repos lors de l'enregistrement ? Justifiez votre réponse.

$$\begin{aligned} \mathbb{P}[X \geq 150] &= \mathbb{P}\left[\frac{150-100}{10} < \frac{X-100}{10}\right] \simeq \mathbb{P}[5,0 < Z] \\ &\simeq 1 - F(5) \simeq 1 - 1 \simeq 0 \end{aligned}$$

Donc si le neurone avait été au repos il aurait été très improbable d'avoir au moins 150PA en une minute. Donc comme on observe un neurone émettant 150 PA par minute, il est plus vraisemblable qu'il soit actif.

Chapitre 5 : Estimation

Exercice 45 : Pile ou face

1. 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 :

i) 100 fois. ii) 1000 fois. iii) 2000 fois. iv) 10000 fois.

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

On a $n = 100 > 30$, et $np = 50 > 5$, et $n(1-p) = 50 > 5$ donc on peut approximer X par $\mathcal{N}(50; 5)$ où $5 = \sqrt{np(1-p)}$

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) :

$$\begin{aligned} \mathbb{P}[48 \leq X \leq 52] &= \mathbb{P}[47,5 \leq X \leq 52,5] \\ &= \mathbb{P}\left[\frac{47,5-50}{5} < \frac{X-50}{5} < \frac{52,5-50}{5}\right] \\ &= \mathbb{P}[-0,5 < Z < 0,5] \\ &= F(0,5) - F(-0,5) \\ &\simeq 0,6915 - (1 - 0,6915) \\ &\simeq 0,6915 - 0,3085 \\ &\simeq 0,38 \end{aligned}$$

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

On a $n = 1000 > 30$, et $np = 500 > 5$, et $n(1-p) = 500 > 5$ donc on peut approximer X par $\mathcal{N}(500; 15,8)$ où $15,8 \simeq \sqrt{np(1-p)}$

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) :

$$\begin{aligned} \mathbb{P}[480 \leq X \leq 520] &= \mathbb{P}[479,5 \leq X \leq 520,5] \\ &= \mathbb{P}\left[\frac{479,5-500}{15,8} < \frac{X-500}{15,8} < \frac{520,5-500}{15,8}\right] \\ &\simeq \mathbb{P}[-1,3 < Z < 1,3] \\ &\simeq F(1,3) - F(-1,3) \\ &\simeq 0,9032 - (1 - 0,9032) \\ &\simeq 0,9032 - 0,0968 \\ &\simeq 0,81 \end{aligned}$$

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

On a $n = 2000 > 30$, et $np = 1000 > 5$, et $n(1-p) = 1000 > 5$ donc on peut approximer X par $\mathcal{N}(1000; 22,4)$ où $22,4 \simeq \sqrt{np(1-p)}$

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

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

$$\begin{aligned} \mathbb{P}[960 \leq X \leq 1040] &= \mathbb{P}[959,5 \leq X \leq 1040,5] \\ &= \mathbb{P}\left[\frac{959,5-1000}{22,4} < \frac{X-1000}{22,4} < \frac{1040,5-1000}{22,4}\right] \\ &\simeq \mathbb{P}[-1,81 < Z < 1,81] \\ &\simeq F(1,81) - F(-1,81) \\ &\simeq 0,9649 - (1 - 0,9649) \\ &\simeq 0,9649 - 0,0351 \\ &\simeq 0,93 \end{aligned}$$

iv) On a $X \sim \mathcal{B}(10000; 0,5)$.

On a $n = 10000 > 30$, et $np = 5000 > 5$, et $n(1-p) = 5000 > 5$ donc on peut approximer X par $\mathcal{N}(5000; 50)$ où $50 = \sqrt{np(1-p)}$

Avoir une proportion de pile entre 0,48 et 0,52 signifierait avoir $4800 \leq X \leq 5200$.

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

$$\begin{aligned} \mathbb{P}[4800 \leq X \leq 5200] &= \mathbb{P}[4799,5 \leq X \leq 5200,5] \\ &= \mathbb{P}\left[\frac{4799,5-5000}{50} < \frac{X-5000}{50} < \frac{5200,5-5000}{50}\right] \\ &= \mathbb{P}[-4,01 < Z < 4,01] \\ &= F(4,01) - F(-4,01) \\ &\simeq 1,0000 - (1 - 1) \\ &\simeq 1 \end{aligned}$$

2. Après avoir lancé une pièce 1200 fois et obtenu 652 fois pile, vous vous demandez si la pièce était vraiment équilibrée. Avec une confiance de 95% donner une estimation de la probabilité p à chaque lancer d'obtenir pile. Que doit-on conclure ?

On a $n = 1200 > 30$, et $p_e = \frac{652}{1200} \simeq 0,543$, d'où $np_e = 652 > 5$, et $n(1-p_e) = 548 > 5$ 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 $a_\alpha = z_\alpha \sqrt{\frac{p_e(1-p_e)}{n}} \simeq 0,028$.

On estime donc que p est dans l'intervalle $[0,515; 0,571]$ avec la confiance $c = 0,95$.

On peut donc affirmer, avec la confiance 95%, que la pièce est légèrement biaisée (elle a au moins 51% de chances de faire pile).

1. Calculer la probabilité d'observer une proportion P_n comprise entre 0,38 et 0,42. *Vous pourrez utiliser pour cela l'approximation d'une loi binomiale par une loi normale*

On réalise un échantillonnage, où la taille de l'échantillon est $n = 500$. Si on suppose que le nombre total de malades (parmi lesquels on a choisi un échantillon de 500 malades) est d'au moins 5000, alors on peut utiliser une loi binomiale (il n'y a pas de différence entre "avec remise" et "sans remise").

En notant X le nombre de patients qui ont guéri au sein de l'échantillon, on a $X \sim \mathcal{B}(500; 0,4)$.

On a $n = 500 > 30$, et $np = 200 > 5$, et $n(1-p) = 300 > 5$ donc on peut approximer X par $\mathcal{N}(200; 11)$ où $11 \simeq \sqrt{np(1-p)}$.

Ainsi,

$$\begin{aligned} \mathbb{P}[0,38 \leq P_n \leq 0,42] &= \mathbb{P}[190 \leq X \leq 210] \\ &= \mathbb{P}[189,5 \leq X \leq 210,5] \\ &= \mathbb{P}\left[\frac{189,5-200}{11} < \frac{X-200}{11} < \frac{210,5-200}{11}\right] \\ &\simeq \mathbb{P}[-0,95 < Z < 0,95] \\ &\simeq F(0,95) - F(-0,95) \\ &\simeq 0,8289 - (1 - 0,8289) \\ &\simeq 0,8289 - 0,1711 \\ &\simeq 0,66 \end{aligned}$$

2. Nous avons observé un échantillon de 500 malades ayant suivi le traitement et constaté 346 guérisons.

- (a) Donner une estimation de la vraie proportion p de guérisons (parmi l'ensemble des patients) avec une confiance de 95%.

On a $n = 500 > 30$, et $p_e = \frac{346}{500} = 0,692$, d'où $np_e = 346 > 5$, et $n(1-p_e) = 154 > 5$ 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 $a_\alpha = z_\alpha \sqrt{\frac{p_e(1-p_e)}{n}} \simeq 0,0405$.

On estime donc que p est dans l'intervalle $[0,6515; 0,7325]$ avec la confiance $c = 0,95$.

Exercice 46 : Efficacité d'un traitement

Un fabricant de médicaments affirme qu'en 10 jours, 40% des malades qui sont traités par l'un de ses produits guérissent de leur maladie.

Supposons que l'affirmation du fabricant soit exacte. On choisit au hasard $n = 500$ malades ayant absorbé le médicament, et on désigne par P_n la proportion des personnes qui ont guéri parmi l'échantillon aléatoire de 500 patients.

- (b) Que peut-on conclure par rapport à l'affirmation du fabricant ?

Donc avec la confiance 95%, on peut affirmer que le taux est bien supérieur à ce qu'affirme le fabricant : il est d'au moins 65%.

- (c) Si on veut réduire la marge de l'estimation à 1% avec une confiance de 98%, quelle doit être la taille minimale de l'échantillon à considérer ?

Pour avoir une précision 0,01 avec la confiance 0,98, il faut $n > z_{\alpha}^2 \frac{p_e(1-p_e)}{h^2} \simeq 2,326^2 \frac{0,692(1-0,692)}{0,01^2} \simeq 11531$

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 1040 personnes interrogées, 612 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 970 personnes interrogées, il y en a alors 620 qui sont satisfaites.

- On désigne par p_1 et p_2 les proportions de satisfaits pour l'ensemble des clients avant et après l'amélioration apportée. Donner une estimation de p_1 et p_2 par intervalle de confiance en prenant comme confiance $c = 0,98$. Peut-on dire que la proportion de satisfaits pour l'ensemble de la clientèle a augmenté ?
- Lors du deuxième sondage, quelle aurait dû être la taille de l'échantillon pour avoir une estimation de p_2 à 1% près avec une confiance de 0,99 ?

Exercice 48 : Résistance à la persuasion

On pense que la résistance à la persuasion passe par une réaction active des sujets qui développent intérieurement des contre-arguments. Pour tester cette hypothèse on considère l'opinion qu'ont les étudiants sur la coopération entre étudiants et enseignants pour établir les programmes. Cette opinion est mesurée par un questionnaire pour lequel, sur l'ensemble des étudiants, la moyenne est 12,3 et l'écart type est 2.

- On soumet dans un premier temps un échantillon de 70 étudiants à un argumentaire persuasif contre cette coopération, puis on mesure leur opinion à l'aide du même questionnaire. On obtient les notes suivantes :

classe	[6 ; 8[[8 ; 10[[10 ; 12[[12 ; 14[[14 ; 16[[16 ; 18[
effectif	8	19	13	15	9	6

- Calculer la moyenne expérimentale $m_e(X)$ et l'écart type expérimental $s_e(X)$ de l'échantillon.
 - 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 $c = 0,95$.
 - 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.
 - 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 ?
- On considère désormais un échantillon de 28 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 $m_e = 7,26$ et $s_e = 4,8$.
 - Avec une confiance de 0,95 déterminer la note moyenne de tous les étudiants s'ils écoutaient l'argumentaire en récitant

des tables de multiplication. (On pourra supposer que les notes suivent une loi normale.)

- b) Comparer cette moyenne à celles qu'ils auraient sans réciter les tables de multiplication (en étant donc en mesure de réfléchir à des contre-arguments). Ces résultats sont-ils compatibles avec l'hypothèse théorique que l'on souhaitait tester ?
- c) Donner une estimation de l'écart type des notes, dans le deuxième cas (en récitant les tables de multiplication) avec un risque d'erreur de 5%.

Exercice 49 : Motivation lors d'un test

Un expérimentateur qui étudie la prise de décisions a demandé à 30 enfants de résoudre le plus grand nombre possible de problèmes en 30 minutes. Il a expliqué à un groupe de 16 enfants qu'il voulait tester leur aptitude innée à résoudre des problèmes, et à un autre groupe (14 enfants) qu'il ne s'agissait que d'une tâche destinée à les occuper. Les résultats obtenus sont les suivants :

Aptitude innée	33	18	27	21	27	22	26	25	33	19	29	30	14	20	32	27
Occupation	12	15	24	9	21	12	17	14	24	9	20	15	18	16		

On suppose que les résultats varient normalement.

1. 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.
- $$m(X) = 25,188 \qquad s(X) = 5,514.$$
- $$m(Y) = 16,143 \qquad s(Y) = 4,703.$$
2. Donner une estimation par intervalle de confiance des moyennes et des écarts types associés aux deux conditions d'expérience. On fera cette estimation avec une confiance de 95%.

- a) Si on leur dit qu'on teste une **aptitude innée** :
Comme $n = 16 \leq 30$, on cherche t_α 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 $a_\alpha = t_\alpha \frac{s_e}{\sqrt{n-1}} \simeq 3,0323$.
On estime donc que $\mu(X)$ est dans l'intervalle $[22,17; 28,23]$ avec la confiance $c = 0,95$

On lit dans la table du χ^2 à 15 degrés de liberté la valeur $x_1 = 6,262$ correspondant à $p = 0,025$ et la valeur $x_2 = 27,48$ correspondant à $q = 0,025$

On obtient alors $s_e \sqrt{\frac{n}{x_2}} \simeq 5,51 \sqrt{\frac{16}{27,48}} \simeq 4,2$ et $s_e \sqrt{\frac{n}{x_1}} \simeq 5,51 \sqrt{\frac{16}{6,262}} \simeq 8,81$

On estime donc que $\sigma(X)$ est dans l'intervalle $[4,2; 8,81]$ avec la confiance $c = 0,95$.

- b) Si on leur dit juste qu'on les **occupe** :
Comme $n = 14 \leq 30$, on cherche t_α à partir de la table inverse de Student avec $p = \frac{\alpha}{2} = 0,025$ et $n - 1 = 13$ degrés de liberté (ddl)
On lit $t_\alpha \simeq 2,1604$ d'où $a_\alpha = t_\alpha \frac{s_e}{\sqrt{n-1}} \simeq 2,8162$.
On estime donc que $\mu(Y)$ est dans l'intervalle $[13,28; 18,92]$ avec la confiance $c = 0,95$

On lit dans la table du χ^2 à 13 degrés de liberté la valeur $x_1 = 5,008$ correspondant à $p = 0,025$ et la valeur $x_2 = 24,73$ correspondant à $q = 0,025$

On obtient alors $s_e \sqrt{\frac{n}{x_2}} \simeq 4,7 \sqrt{\frac{14}{24,73}} \simeq 3,54$ et $s_e \sqrt{\frac{n}{x_1}} \simeq 4,7 \sqrt{\frac{14}{5,008}} \simeq 7,86$

On estime donc que $\sigma(X)$ est dans l'intervalle $[3,54; 7,86]$ 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 $\mu(X) > 22$ alors que $\mu(Y) < 19$). 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 hommes et les femmes de plus de 50 ans. On suppose qu'ils suivent des lois normales.

Dans un échantillon de 18 hommes de plus de 50 ans on a mesuré un taux moyen de 192,4mg/dL avec un écart type de 45,2mg/dL.

Dans un échantillon de 22 femmes de plus de 50 ans on a mesuré un taux moyen de 185,7mg/dL avec un écart type de 36,4mg/dL.

1. Donner des estimations des écarts types des deux populations.

On choisit par exemple de faire des estimations avec la confiance 95% :

femmes On lit dans la table du χ^2 à 21 degrés de liberté la valeur $x_1 = 10,28$ correspondant à $p = 0,025$ et la valeur $x_2 = 35,47$ correspondant à $q = 0,025$

On obtient alors $s_e \sqrt{\frac{n}{x_2}} \simeq 36,4 \sqrt{\frac{22}{35,47}} \simeq 28,7$ et $s_e \sqrt{\frac{n}{x_1}} \simeq$

$36,4 \sqrt{\frac{22}{10,28}} \simeq 53,2$

On estime donc que σ est dans l'intervalle $[28,7; 53,2]$ avec la confiance $c = 0,95$.

hommes On lit dans la table du χ^2 à 17 degrés de liberté la valeur $x_1 = 7,564$ correspondant à $p = 0,025$ et la valeur $x_2 = 30,19$ correspondant à $q = 0,025$

On obtient alors $s_e \sqrt{\frac{n}{x_2}} \simeq 45,2 \sqrt{\frac{18}{30,19}} \simeq 34,9$ et $s_e \sqrt{\frac{n}{x_1}} \simeq$

$45,2 \sqrt{\frac{18}{7,564}} \simeq 69,7$

On estime donc que σ est dans l'intervalle $[34,9; 69,7]$ avec la confiance $c = 0,95$.

2. En utilisant les estimations par intervalle de confiance à 95%, peut-on conclure que les femmes de plus de 50 ans ont en moyenne moins de cholestérol que les hommes de plus de 50 ans ?

On commence par estimer les moyennes :

femmes Comme $n = 22 \leq 30$, on cherche t_α à partir de la table inverse de Student avec $p = \frac{\alpha}{2} = 0,025$ et $n - 1 = 21$ degrés de liberté (ddl)

On lit $t_\alpha \simeq 2,0796$ d'où $a_\alpha = t_\alpha \frac{s_e}{\sqrt{n-1}} \simeq 16,519$.

On estime donc que μ est dans l'intervalle $[169,2; 202,2]$ avec la confiance $c = 0,95$

hommes Comme $n = 18 \leq 30$, on cherche t_α à partir de la table inverse de Student avec $p = \frac{\alpha}{2} = 0,025$ et $n - 1 = 17$ degrés de liberté (ddl)

On lit $t_\alpha \simeq 2,1098$ d'où $a_\alpha = t_\alpha \frac{s_e}{\sqrt{n-1}} \simeq 23,129$.

On estime donc que μ est dans l'intervalle $[169,3; 215,5]$ avec la confiance $c = 0,95$

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

Exercice 51 : Épreuve de dictée

Le tableau suivant représente les résultats obtenus par un échantillon de 15 enfants de CE2 dans deux épreuves de dictée préparée : une liste de 20 mots (épreuve \mathcal{E}_1) et un texte (épreuve \mathcal{E}_2).

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

1. Peut-on affirmer, avec un risque d'erreur de 5%, qu'une des deux épreuves donne, pour des enfants de CE2, une note moyenne plus élevée que l'autre épreuve ?

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

$$\begin{aligned} m(X) &= 15,933 & s(X) &= 1,974. \\ m(Y) &= 12,067 & s(Y) &= 1,389. \end{aligned}$$

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

Épreuve 1 Comme $n = 15 \leq 30$, on cherche t_α 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 $a_\alpha = t_\alpha \frac{s_e}{\sqrt{n-1}} \simeq 1,1315$.

On estime donc que $\mu(X)$ est dans l'intervalle $[14,8; 17,06]$ avec la confiance $c = 0,95$

Épreuve 2 Comme $n = 15 \leq 30$, on cherche t_α 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 $a_\alpha = t_\alpha \frac{s_e}{\sqrt{n-1}} \simeq 0,79621$.

On estime donc que $\mu(Y)$ est dans l'intervalle $[11,27; 12,86]$ avec la confiance $c = 0,95$

- (c) : Ces intervalles indiquent que $\mu(X) > \mu(Y)$ (en effet $\mu(X) > 14$ et $\mu(Y) < 13$). 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 χ^2 à 14 degrés de liberté la valeur $x_1 = 5,628$ correspondant à $p = 0,025$ et la valeur $x_2 = 26,11$ correspondant à $q = 0,025$

On obtient alors $s_e \sqrt{\frac{n}{x_2}} \simeq 1,974 \sqrt{\frac{15}{26,11}} \simeq 1,5$ et $s_e \sqrt{\frac{n}{x_1}} \simeq 1,974 \sqrt{\frac{15}{5,628}} \simeq 3,22$

On estime donc que $\sigma(X)$ est dans l'intervalle $[1,5; 3,22]$ avec la confiance $c = 0,95$.

Épreuve 2 On lit dans la table du χ^2 à 14 degrés de liberté la valeur $x_1 = 5,628$ correspondant à $p = 0,025$ et la valeur $x_2 = 26,11$ correspondant à $q = 0,025$

On obtient alors $s_e \sqrt{\frac{n}{x_2}} \simeq 1,389 \sqrt{\frac{15}{26,11}} \simeq 1,05$ et $s_e \sqrt{\frac{n}{x_1}} \simeq 1,389 \sqrt{\frac{15}{5,628}} \simeq 2,27$

On estime donc que $\sigma(Y)$ est dans l'intervalle $[1,05; 2,27]$ avec la confiance $c = 0,95$.

Exercice 52 :

Un psychologue étudie le taux de chômage de 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 10% (selon les données de l'INSEE).

On choisit au hasard un échantillon de 117 français. Quelle est la probabilité d'avoir, au sein de cet échantillon, entre 9% et 11% de chômeurs ?

Justifiez bien les éventuelles approximations que vous seriez amené à faire

- L'échantillon de 117 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 $X \sim \mathcal{B}(117; 0,1)$.

• On a $n = 117 > 30$, et $np \simeq 11,7 > 5$, et $n(1-p) \simeq 105 > 5$ donc on peut approximer X par $\mathcal{N}(11,7; 3,24)$ où $3,24 \simeq \sqrt{np(1-p)}$

• Une proportion entre 0,09 et 0,11 signifie que X est entre $107 \times 0,09 \simeq 9,63$ et $107 \times 0,11 = 11,77$. (Donc entre 10 et 11 comme X est entier.) On calcule donc :

$$\begin{aligned} \mathbb{P}[10 \leq X \leq 11] &= \mathbb{P}[9,5 \leq X \leq 11,5] \\ &= \mathbb{P}\left[\frac{9,5-11,7}{3,24} < \frac{X-11,7}{3,24} < \frac{11,5-11,7}{3,24}\right] \\ &\simeq \mathbb{P}[-0,68 < Z < -0,06] \\ &\simeq F(-0,06) - F(-0,68) \\ &\simeq (1 - 0,5239) - (1 - 0,7517) \\ &\simeq 0,4761 - 0,2483 \\ &\simeq 0,23 \end{aligned}$$

2. Sur un échantillon de 117 patients atteints de boulimie, le psychologue a constaté que 5,98% d'entre eux étaient au chômage.

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

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

On a $n = 117 > 30$, et $p_e = \frac{7}{117} \simeq 0,06$, d'où $np_e = 7 > 5$, et $n(1-p_e) = 110 > 5$ 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 $a_\alpha = z_\alpha \sqrt{\frac{p_e(1-p_e)}{n}} \simeq 0,0361$.

On estime donc que p est dans l'intervalle $[0,0237; 0,0959]$ avec la confiance $c = 0,9$.

Toutes les valeurs de cet intervalle sont inférieures à 10%, dont on peut affirmer, avec la confiance 90%, que le taux de chômage des personnes boulimiques est plus faible que parmi le reste de la population.

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

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

On a $n = 117 > 30$, et $p_e = \frac{7}{117} \simeq 0,06$, d'où $np_e = 7 > 5$, et $n(1-p_e) = 110 > 5$ 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 $a_\alpha = z_\alpha \sqrt{\frac{p_e(1-p_e)}{n}} \simeq 0,051$.

On estime donc que p est dans l'intervalle $[0,0088; 0,1108]$ avec la confiance $c = 0,98$.

Cette fois-ci, la valeur 10% 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 1% près le taux de chômage des personnes boulimiques avec une confiance 98%, quelle taille d'échantillon faudrait il considérer ?

Pour avoir une précision 0,01 avec la confiance 0,98, il faut $n > z_\alpha^2 \frac{p_e(1-p_e)}{h^2} \simeq 2,326^2 \frac{0,0598(1-0,0598)}{0,01^2} \simeq 3041$

Exercice 53 : Une chercheuse, Mme Lefebvre, 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 1000 femmes, parmi lesquelles 7 sont daltoniennes, et d’autre part un échantillon de 1000 hommes, parmi lesquels 94 sont daltoniens.

Avec la confiance $c = 99\%$, peut on déduire que la prévalence du daltonisme soit plus importante chez l’homme que chez la femme ?

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

• **parmi les femmes :**

On a $n = 1000 > 30$, et $p_e = \frac{7}{1000} = 0,007$, d’où $np_e = 7 > 5$, et $n(1 - p_e) = 993 > 5$ donc on peut utiliser la procédure du formulaire pour estimer la proportion p .
On a $F(2,576) \simeq 0,995$ d’où $z_\alpha \simeq 2,576$

d’où $a_\alpha = z_\alpha \sqrt{\frac{p_e(1-p_e)}{n}} \simeq 0,00679$.

On estime donc que p est dans l’intervalle $[0,00021; 0,01379]$ avec la confiance $c = 0,99$.

• **parmi les hommes :**

On a $n = 1000 > 30$, et $p_e = \frac{94}{1000} = 0,094$, d’où $np_e = 94 > 5$, et $n(1 - p_e) = 906 > 5$ donc on peut utiliser la procédure du formulaire pour estimer la proportion p .

On a $F(2,576) \simeq 0,995$ d’où $z_\alpha \simeq 2,576$

d’où $a_\alpha = z_\alpha \sqrt{\frac{p_e(1-p_e)}{n}} \simeq 0,02377$.

On estime donc que p est dans l’intervalle $[0,07023; 0,11777]$ avec la confiance $c = 0,99$.

(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 $c = 99\%$, 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 7 ans, auxquels elle fait passer un test standardisé de perception des couleurs. Sur un échantillon de 15 filles, elle obtient les données suivantes :

Score	7	6	8	0	6	6	5	2	7	7	6	4	7	6	7
-------	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

Estimer, avec la confiance 98%, le score moyen des filles à ce test.

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

$$\text{moyenne : } m(X) = \frac{\sum x_i}{n} = \frac{7+6+\dots+7}{15} = \frac{84}{15} = 5,6$$

$$m(X^2) = \frac{\sum x_i^2}{n} = \frac{7^2+6^2+\dots+7^2}{15} = \frac{534}{15}$$

$$\text{Var}(X) = m(X^2) - m(X)^2 = \frac{534}{15} - \left(\frac{84}{15}\right)^2 = 4,24$$

$$\text{Écart-type : } s(X) = \sqrt{\text{Var}(X)} \simeq 2,06$$

• Comme $n = 15 \leq 30$, on cherche t_α à partir de la table inverse de Student avec $p = \frac{\alpha}{2} = 0,01$ et $n - 1 = 14$ degrés de liberté (ddl)

On lit $t_\alpha \simeq 2,6245$ d’où $a_\alpha = t_\alpha \frac{s_e}{\sqrt{n-1}} \simeq 1,4449$.

On estime donc que μ est dans l’intervalle $[4,155; 7,045]$ avec la confiance $c = 0,98$

3. Sur un échantillon de 40 garçon auxquels la chercheuse fait passer le test, le score moyen est de 4,27 avec un écart-type de 1,73.

Estimer, avec la confiance 98%, le score moyen des garçons à ce test.

Peut on affirmer, avec la confiance 98%, que – comme s’y attendait la chercheuse – les filles ont en moyenne une meilleure perception des couleurs que les garçons ?

• Comme $n = 40 > 30$, on cherche z_α tel que $F(z_\alpha) = \frac{0,98+1}{2} = 0,99$

On a $F(2,326) \simeq 0,99$ d’où $z_\alpha \simeq 2,326$ d’où $a_\alpha = z_\alpha \frac{s_e}{\sqrt{n-1}} \simeq 0,6444$

On estime donc que μ est dans l’intervalle $[4,27 - 0,6444; 4,27 + 0,6444] \simeq [3,63; 4,91]$, avec la confiance $c = 0,98$.

- Les intervalles se chevauchent, donc ces échantillons ne permettent pas d'affirmer, avec la confiance 98%, 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 98%. Peut-on conclure qu'ils diffèrent selon le sexe ?

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

On lit dans la table du χ^2 à 14 degrés de liberté la valeur $x_1 = 4,66$ correspondant à $p = 0,01$ et la valeur $x_2 = 29,14$ correspondant à $q = 0,01$

On obtient alors $s_e \sqrt{\frac{n}{x_2}} \simeq 2,06 \sqrt{\frac{15}{29,14}} \simeq 1,48$ et $s_e \sqrt{\frac{n}{x_1}} \simeq 2,06 \sqrt{\frac{15}{4,66}} \simeq 3,7$

On estime donc que σ est dans l'intervalle $[1,48; 3,7]$ avec la confiance $c = 0,98$.

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

On lit dans la table du χ^2 à 39 degrés de liberté la valeur $x_1 = 21,42$ correspondant à $p = 0,01$ et la valeur $x_2 = 62,42$ correspondant à $q = 0,01$

On obtient alors $s_e \sqrt{\frac{n}{x_2}} \simeq 1,73 \sqrt{\frac{40}{62,42}} \simeq 1,38$ et $s_e \sqrt{\frac{n}{x_1}} \simeq 1,73 \sqrt{\frac{40}{21,42}} \simeq 2,36$

On estime donc que σ est dans l'intervalle $[1,38; 2,36]$ avec la confiance $c = 0,98$.

- (c) Les intervalles se chevauchent, on ne peut donc pas conclure, avec la confiance 98%, quant à une différence entre l'écart-type des filles et celui des garçons.

5. Quelle taille d'échantillon aurait-il fallu considérer pour estimer le score moyen des filles à 0,25 point près, avec une confiance 98% ? Pour avoir une précision 0,25 avec la confiance 0,98, il faut $n > z_\alpha^2 \frac{(s_e)^2}{h^2} \simeq 2,326^2 \frac{2,06^2}{0,25^2} \simeq 367$