Exercices de TD de statistiques

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

corrigé en ligne

Si le résultat d'un calcul est un nombre à virgule, vous pouvez soit demander à la calculatrice de l'afficher sous forme décimale (nombre à virgule) soit sous forme de fraction (avec la touche F↔D pour une Casio, ou la fonction ▸Frac du menu MATH pour une TI).

1. Grâce à cette fonctionnalité de la calculatrice, simplifiez chacune des fractions suivantes : $\frac{-18}{-23-1}$, $\frac{-165}{10-54}$, $\frac{60}{10\times 3}$, $\frac{-18}{16}$, $\frac{26+51}{46-130}$, $\frac{-26\times 2}{26\times 3}$.

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

```
-18/(-23-1)▸Frac
                  3/4
-165/(10-54)▸Frac
                 15/4
60/(10×3)▸Frac
                    2
-18/16▸Frac
                 -9/8
(26+51)/(46-130)▸Frac
               -11/12
(-26×2)/(26×3)▸Frac
                 -2/3
```

   $\frac{-18}{-23-1}=\frac{3}{4}$, $\frac{-165}{10-54}=\frac{15}{4}$,

   $\frac{60}{10\times 3}=2$, $\frac{-18}{16}=-\frac{9}{8}$,

   $\frac{26+51}{46-130}=-\frac{11}{12}$, $\frac{-26\times 2}{26\times 3}=-\frac{2}{3}$.

   Remarque : La capture d'écran correspond à une TI. Sur une Casio, si on veut par exemple calculer $\frac{-18}{16}$, on tape `(-18)÷16` puis si le résultat affiché est un chiffre à virgule on appuie sur F↔D pour en faire une fraction. Au lieu d'afficher $\frac{-9}{8}$, certaines calculatrices affichent alors $-1\frac{1}{8}$ (ces calculatrices considèrent que cela signifie $-\left(1+\frac{1}{8}\right)$). Pour obtenir la réponse sous la bonne forme sur ces calculatrices, il faut alors appuyer sur SHIFT puis F↔D (sur la touche il est en effet écrit "$a+\frac{b}{c}\leftrightarrow\frac{d}{c}$").

2. Classez-les dans l'ordre croissant ("du plus petit au plus grand").

   Cette fois-ci on utilise le résultats sous formes de chiffres à virgules. On obtient (voir capture d'écran ci-dessous) :

   $\frac{-18}{-23-1}=0{,}75$, $\frac{-165}{10-54}=3{,}75$, $\frac{60}{10\times 3}=2$, $\frac{-18}{16}=-1{,}125$,

   $\frac{26+51}{46-130}\simeq -0{,}917$, $\frac{-26\times 2}{26\times 3}\simeq -0{,}667$.

```
-18/(-23-1)
                 0.75
-165/(10-54)
                 3.75
60/(10×3)
                    2
-18/16
               -1.125
(26+51)/(46-130)
        -0.9166666667
(-26×2)/(26×3)
        -0.6666666667
```

   L'ordre croissant de ces nombres est :
   $-1{,}125 < -0{,}917 < -0{,}667 < 0{,}75 < 2 < 3{,}75$

   Donc la réponse est :
   $\frac{-18}{16} < \frac{26+51}{46-130} < \frac{-26\times 2}{26\times 3} < \frac{-18}{-23-1} < \frac{60}{10\times 3} < \frac{-165}{10-54}$

Exercice 2 : Utilisation de la calculatrice

corrigé en ligne

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

```
-1²
                   -1
```

   Peut-on en conclure

   (a) que le carré de -1 est -1 ?

   (b) 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.
```
1/(2025×2026)
  2.437448965E-7
```
(a) $\frac{1}{2025\times2026} \simeq 2{,}44$
(b) $\frac{1}{2025\times2026} \simeq 24374489{,}66$
(c) $\frac{1}{2025\times2026} \simeq 0{,}000000244$

La bonne réponse est la réponse (c). En effet `2.437448965E-7` désigne la quantité $2{,}437448965 \times 10^{-7}$, qui vaut 0,0000002437448965.

3.
```
9-24→X
            -15
X×2026
         -30390
```
(a) $9 - 24 \times 2026 = -30390$
(b) $15 \times 2026 = -30390$
(c) $(9 - 24) \times 2026 = -30390$

La bonne réponse est la réponse (c). En effet dans `X×2026`, `X` vaut $9-24$ (calculé à la ligne précédente), de sorte qu'on a calculé $(9-24) \times 2\,026$.

4.
```
-6×27
           -162
Ans²
          26244
```
(a) $-6 \times 27^2 = 26244$
(b) $-162^2 = 26244$
(c) $(-6 \times 27)^2 = 26244$

La bonne réponse est la réponse (c). En effet dans `Ans²`, `Ans` vaut -6×27 (calculé à la ligne précédente), de sorte qu'on a calculé $(-6 \times 27)^2$.

Exercice 3 : Puissances de dix

corrigé en ligne

Calculez (de préférence “de tête”) chacune des expressions suivantes :

$34 \div 100$	$0{,}34 \times 10^2$	$0{,}151\,962 \div 1\,000$
$1{,}185 \times 10^{-4}$	$1{,}342\,261\,1 \times 10^{-7}$	$12{,}838\,6 \times 100$
$123\,468{,}972 \div 100\,000\,000$	$1{,}228\,1 \times 1\,000\,000$	144×10^7

- $34 \div 100 = 0{,}34$ (que l'on peut aussi écrire $3{,}4 \times 10^{-1}$)
- $0{,}34 \times 10^2 = 34$ (que l'on peut aussi écrire $3{,}4 \times 10^1$)
- $0{,}151\,962 \div 1\,000 = 0{,}000151962$ (que l'on peut aussi écrire $1{,}51962 \times 10^{-4}$)
- $1{,}185 \times 10^{-4} = 0{,}0001185$ (que l'on peut aussi écrire $1{,}185 \times 10^{-4}$)
- $1{,}342\,261\,1 \times 10^{-7} = 0{,}00000013422611$ (que l'on peut aussi écrire $1{,}3422611 \times 10^{-7}$)
- $12{,}838\,6 \times 100 = 1283{,}86$ (que l'on peut aussi écrire $1{,}28386 \times 10^3$)
- $123\,468{,}972 \div 100\,000\,000 = 0{,}00123468972$ (que l'on peut aussi écrire $1{,}23468972 \times 10^{-3}$)
- $1{,}228\,1 \times 1\,000\,000 = 1228100$ (que l'on peut aussi écrire $1{,}2281 \times 10^6$)
- $144 \times 10^7 = 1440000000$ (que l'on peut aussi écrire $1{,}44 \times 10^9$)

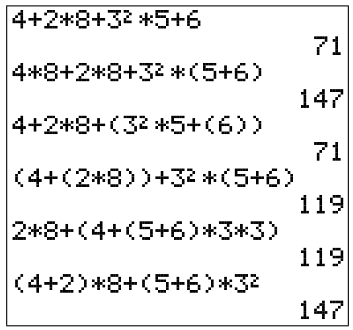
Exercice 4 : Règles de calcul

corrigé en ligne

1. On considère les expressions suivantes :
   - $4+2\times8+3^2\times5+6$
   - $4\times8+2\times8+3^2(5+6)$
   - $4+2\times8+(3^2\times5+(6))$
   - $(4+(2\times8))+3^2(5+6)$
   - $2\times8+(4+(5+6)\times3\times3)$
   - $(4+2)\times8+(5+6)\times3^2$

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

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

   - $4+2\times8+3^2\times5+6=71$
   - $4\times8+2\times8+3^2(5+6)=147$
   - $4+2\times8+(3^2\times5+(6))=71$
   - $(4+(2\times8))+3^2(5+6)=119$
   - $2\times8+(4+(5+6)\times3\times3)=119$
   - $(4+2)\times8+(5+6)\times3^2=147$

   On remarque alors d'une part que les expressions $(4+(2\times8))+3^2(5+6)$ et $2\times8+(4+(5+6)\times3\times3)$ sont égales, et d'autre part que les expressions $(4+2)\times8+(5+6)\times3^2$ et $4\times8+2\times8+3^2(5+6)$ sont égales.

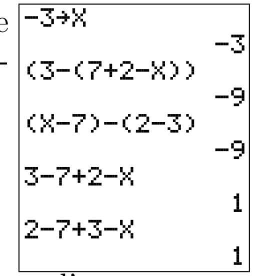
   Remarque : l'égalité de ces expressions se comprend à partir des règles de parenthésage et de signe expliquées dans les "Rappels" au début des notes de cours.

2. De même, calculez les expressions suivantes lorsque les variables prennent les valeurs indiquées, et observez que certaines d'entre-elles sont égales.

   (a) Les expressions
   - $(3-(7+2-x))$
   - $3-7+2-x$
   - $(x-7)-(2-3)$
   - $2-7+3-x$

   lorsque $x=-3$.

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

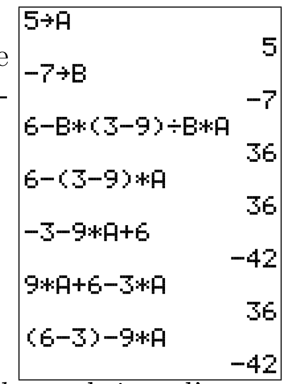
   - $(3-(7+2-x))=-9$
   - $(x-7)-(2-3)=-9$
   - $3-7+2-x=1$
   - $2-7+3-x=1$

   Ainsi, lorsque $x=-3$, on obtient d'une part que les expressions $(3-(7+2-x))$ et $(x-7)-(2-3)$ sont égales, d'autre part que les expressions $3-7+2-x$ et $2-7+3-x$ sont égales.

   (b) Les expressions
   - $6-b\times\frac{3-9}{b}a$
   - $-3-9a+6$
   - $(6-3)-9a$
   - $6-(3-9)a$
   - $9a+6-3a$

   lorsque $a=5$ et $b=-7$.

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

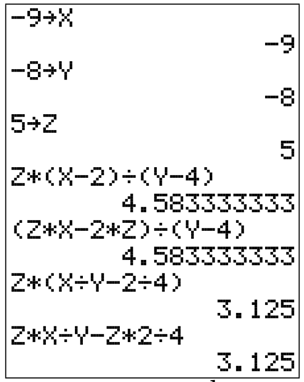
   - $6-b\times\frac{3-9}{b}a=36$
   - $6-(3-9)a=36$
   - $-3-9a+6=-42$
   - $9a+6-3a=36$
   - $(6-3)-9a=-42$

   Ainsi, pour ces valeurs de a et b, on obtient d'une part que les expressions $6-b\times\frac{3-9}{b}a$, $6-(3-9)a$ et $9a+6-3a$ sont égales, d'autre part que les expressions $-3-9a+6$ et $(6-3)-9a$ sont égales.

   (c) Lorsque $x=-9$, $y=-8$ et $z=5$:
   - $z\frac{(x-2)}{(y-4)}$
   - $z\left(\frac{x}{y}-\frac{2}{4}\right)$
   - $(z\times x-2\times z)\div(y-4)$
   - $\frac{z\times x}{y}-\frac{z\times2}{4}$

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

- $z\frac{(x-2)}{(y-4)} \simeq 4{,}583$
- $(z \times x - 2 \times z) \div (y-4) \simeq 4{,}583$
- $z\left(\frac{x}{y} - \frac{2}{4}\right) = 3{,}125$
- $\frac{z\times x}{y} - \frac{z\times 2}{4} = 3{,}125$



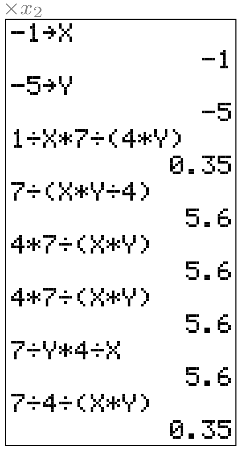
Ainsi, pour ces valeurs de x, y et z, on obtient d'une part que les expressions $z\frac{(x-2)}{(y-4)}$ et $(z \times x - 2 \times z) \div (y-4)$ sont égales, d'autre part que les expressions $z\left(\frac{x}{y} - \frac{2}{4}\right)$ et $\frac{z\times x}{y} - \frac{z\times 2}{4}$ sont égales.

(d) Lorsque $x_1 = -1$ et $x_2 = -5$:

- $\frac{1}{x_1} \times \frac{7}{4\times x_2}$
- $\frac{7}{x_1\times x_2 \div 4}$
- $\frac{4\times 7}{x_1 \times x_2}$
- $4\frac{7}{x_1\times x_2}$
- $7 \div x_2 \times 4 \div x_1$
- $\frac{7\div 4}{x_1\times x_2}$

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

- $\frac{1}{x_1} \times \frac{7}{4\times x_2} = 0{,}35$
- $\frac{7}{x_1\times x_2 \div 4} = 5{,}6$
- $\frac{4\times 7}{x_1 \times x_2} = 5{,}6$
- $4\frac{7}{x_1\times x_2} = 5{,}6$
- $7 \div x_2 \times 4 \div x_1 = 5{,}6$
- $\frac{7\div 4}{x_1\times x_2} = 0{,}35$



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

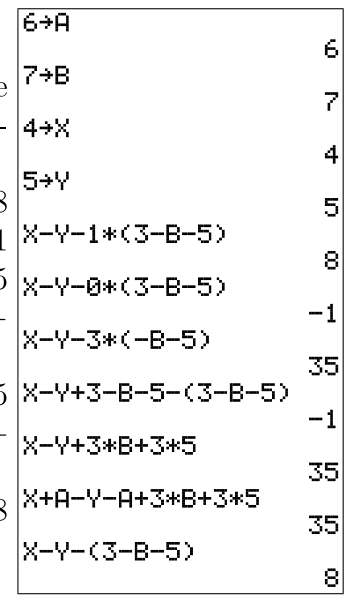
Ainsi, pour ces valeurs de x_1 et x_2, on obtient d'une part que les expressions $\frac{7}{x_1\times x_2\div 4}$, $\frac{4\times 7}{x_1\times x_2}$, $4\frac{7}{x_1\times x_2}$ et $7 \div x_2 \times 4 \div x_1$ sont égales, d'autre part que les expressions $\frac{1}{x_1} \times \frac{7}{4\times x_2}$ et $\frac{7\div 4}{x_1\times x_2}$ sont égales.

(e) Lorsque $a_1 = 6$, $a_2 = 7$, $x_1 = 4$ et $x_2 = 5$:

- $x_1 - x_2 - 1(3 - a_2 - 5)$
- $x_1 - x_2 - 0(3 - a_2 - 5)$
- $x_1 - x_2 - 3(-a_2 - 5)$
- $x_1 - x_2 + 3 - a_2 - 5 - (3 - a_2 - 5)$
- $x_1 - x_2 + 3 \times a_2 + 3 \times 5$
- $x_1 + a_1 - x_2 - a_1 + 3 \times a_2 + 3 \times 5$
- $x_1 - x_2 - (3 - a_2 - 5)$

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

- $x_1 - x_2 - 1(3 - a_2 - 5) = 8$
- $x_1 - x_2 - 0(3 - a_2 - 5) = -1$
- $x_1 - x_2 - 3(-a_2 - 5) = 35$
- $x_1 - x_2 + 3 - a_2 - 5 - (3 - a_2 - 5) = -1$
- $x_1 - x_2 + 3 \times a_2 + 3 \times 5 = 35$
- $x_1 + a_1 - x_2 - a_1 + 3 \times a_2 + 3 \times 5 = 35$
- $x_1 - x_2 - (3 - a_2 - 5) = 8$



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

Ainsi, pour ces valeurs de a_1, a_2, x_1 et x_2, on obtient d'une part que les expressions $x_1 - x_2 - 3(-a_2 - 5)$, $x_1 - x_2 + 3 \times a_2 + 3 \times 5$ et $x_1 + a_1 - x_2 - a_1 + 3 \times a_2 + 3 \times 5$ sont égales, d'autre part que les expressions $x_1 - x_2 - 1(3 - a_2 - 5)$ et $x_1 - x_2 - (3 - a_2 - 5)$ sont égales, et enfin que les expressions $x_1 - x_2 - 0(3 - a_2 - 5)$ et $x_1 - x_2 + 3 - a_2 - 5 - (3 - a_2 - 5)$ sont égales.

Compléments

Exercice 5 : Précision des calculs

Hafsa, Thomas, Imane et Yélèna étudient en première année de psychologie à l'Université de Bourgogne.

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

$m(X) = \frac{1{,}88+1{,}73+1{,}86}{3}$

$m(X^2) = \frac{1{,}88^2+1{,}73^2+1{,}86^2}{3}$

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

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

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

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

Copie de Hafsa

$m(X) = \frac{1{,}88+1{,}73+1{,}86}{3} \simeq 1{,}82$

$m(X^2) = \frac{1{,}88^2+1{,}73^2+1{,}86^2}{3} \simeq 3{,}33$

$Var(X) \simeq 3{,}33 - 1{,}82^2 \simeq 0{,}02$

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

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

Calculatrice de Hafsa

```
(1.88+1.73+1.86)÷3
                1.823333333
(1.88²+1.73²+1.86²)÷3
                3.328966667
3.33-1.82²
                     0.0176
√(0.02)
               0.1414213562
```

Copie de Thomas

$m(X) = \frac{1{,}88+1{,}73+1{,}86}{3} = \frac{5{,}47}{3}$

$m(X^2) = \frac{1{,}88^2+1{,}73^2+1{,}86^2}{3} = \frac{9{,}986\,9}{3}$

$Var(X) \simeq \frac{9{,}986\,9}{3} - \left(\frac{5{,}47}{3}\right)^2 \simeq 0{,}0044$

$s(X) = \sqrt{Var} \simeq 0{,}07$ mètre.

Donc l'écart-type est d'environ 7 cm.

Calculatrice de Thomas

```
1.88+1.73+1.86
                       5.47
1.88²+1.73²+1.86²
                     9.9869
9.9869÷3-(5.47÷3)²
             0.004422222222
√(Ans)
              0.06649979114
```

Copie d'Imane

$m(X) = \frac{1{,}88+1{,}73+1{,}86}{3} \simeq 1{,}82$
$m(X^2) = \frac{1{,}88^2+1{,}73^2+1{,}86^2}{3} \simeq 3{,}33$
$Var(X) = m(X^2) - (m(X))^2 \simeq 0{,}0044$
$s(X) = \sqrt{Var} \simeq 0{,}07$ mètre.
Donc l'écart-type est d'environ 7 cm.

Calculatrice d'Imane

```
(1.88+1.73+1.86)÷3→M
                1.823333333
(1.88²+1.73²+1.86²)÷3→N
                3.328966667
N-M²→V
             0.004422222222
√(V)
              0.06649979114
```

Copie d'Yélèna

$s(X) = \sqrt{\frac{1{,}88^2+1{,}73^2+1{,}86^2}{3} - \left(\frac{1{,}88+1{,}73+1{,}86}{3}\right)^2}$
$\simeq 0{,}07$ mètre.
Donc l'écart-type est d'environ 7 cm.

Calculatrice d'Yélèna

```
√((1.88²+1.73²+1.86²)÷3
-((1.88+1.73+1.86)÷3)²)
              0.06649979114
```

1. Parmi ces différentes réponses, lesquelles vous semblent correctes ? Expliquez pourquoi l'un·e des étudiant·e·s a obtenu une réponse erronée.
2. On demande ensuite de calculer l'écart-type de la taille d'un échantillon de 7 femmes. Comme l'échantillon n'est plus le même, il faut cette fois-ci faire les calculs suivants :
   $m(X) = \frac{1{,}57+1{,}58+1{,}72+1{,}69+1{,}66+1{,}71+1{,}65}{7}$
   $m(X^2) = \frac{1{,}57^2+1{,}58^2+1{,}72^2+1{,}69^2+1{,}66^2+1{,}71^2+1{,}65^2}{7}$
   $Var(X) = m(X^2) - (m(X))^2$
   $s(X) = \sqrt{Var(X)}$
   Rédigez ce calcul sans commettre d'erreurs d'arrondis (vous pourrez reprendre la rédaction d'un·e des étudiant·e·s ayant trouvé la bonne réponse).

Exercice 6 : Utilisation du symbole $\sum$

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

Exercice 7 : Précision des calculs

corrigé en ligne

Erwan, Sophie, Anaïs et Élisa étudient en première année de psychologie à l'Université de Bourgogne.

On leur demande de calculer l'écart-type de l'âge auquel 30 enfants ont perdu leur première *dent de laie*, et au vu des formules du formulaire, cela revient à calculer des nombres notés "$m(X)$", "$m(X^2)$", "$Var(X)$" et "$s(X)$", donnés par les formules suivantes :
$m(X) = \frac{12\times5+14\times6+4\times7}{30}$
$m(X^2) = \frac{12\times5^2+14\times6^2+4\times7^2}{30}$
$Var(X) = m(X^2) - (m(X))^2$
$s(X) = \sqrt{Var(X)}$
Le nombre "$s(X)$" issu de ce calcul est l'écart-type, exprimé (dans le cas présent) en année·s.

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

Copie d'Erwan

$m(X) = \frac{12\times5+14\times6+4\times7}{30} \simeq 5{,}73$
$m(X^2) = \frac{12\times5^2+14\times6^2+4\times7^2}{30} \simeq 33{,}33$
$Var(X) \simeq 33{,}33 - 5{,}73^2 \simeq 0{,}50$
$s(X) \simeq \sqrt{0{,}5} \simeq 0{,}707$ an.
Donc l'écart-type est d'environ 0,707 an, c'est à dire $365 \times 0{,}707 \simeq 258$ jours.

Calculatrice d'Erwan

```
(12×5+14×6+4×7)÷30
                5.733333333
(12×5²+14×6²+4×7²)÷30
                33.33333333
33.33-5.73²
                     0.4971
√(0.5)
               0.7071067812
365*0.707
                    258.055
```

Copie de Sophie

$m(X) = \frac{12\times5+14\times6+4\times7}{30} = \frac{172}{30}$
$m(X^2) = \frac{12\times5^2+14\times6^2+4\times7^2}{30} = \frac{1\,000}{30}$
$Var(X) \simeq \frac{1\,000}{30} - \left(\frac{172}{30}\right)^2 \simeq 0{,}4622$
$s(X) = \sqrt{Var} \simeq 0{,}680$ an.
Donc l'écart-type est d'environ 0,68 an, c'est à dire 248 jours.

Calculatrice de Sophie

```
12×5+14×6+4×7
                    172
12×5²+14×6²+4×7²
                   1000
1000÷30-(172÷30)²
           0.4622222222
√(Ans)
           0.6798692685
365*Ans
             248.152283
```

Copie d'Anaïs

$m(X) = \frac{12\times5+14\times6+4\times7}{30} \simeq 5{,}73$
$m(X^2) = \frac{12\times5^2+14\times6^2+4\times7^2}{30} \simeq 33{,}33$
$Var(X) = m(X^2) - (m(X))^2 \simeq 0{,}4622$
$s(X) = \sqrt{Var} \simeq 0{,}680$ an.
Donc l'écart-type est d'environ 0,68 an, c'est à dire 248 jours.

Calculatrice d'Anaïs

```
(12×5+14×6+4×7)÷30→M
            5.733333333
(12×5²+14×6²+4×7²)÷30→N
            33.33333333
N-M²→V
           0.4622222222
√(V)→S
           0.6798692685
365*S
             248.152283
```

Copie d'Élisa

$s(X) = \sqrt{\frac{12\times5^2+14\times6^2+4\times7^2}{30} - \left(\frac{12\times5+14\times6+4\times7}{30}\right)^2}$
$\simeq 0{,}680$ an.
Donc l'écart-type est d'environ 0,68 an, c'est à dire 248 jours.

Calculatrice d'Élisa

```
√((12×5²+14×6²+4×7²)÷30
-((12×5+14×6+4×7)÷30)²)
           0.6798692685
365*Ans
             248.152283
```

1. Parmi ces différentes réponses, lesquelles vous semblent correctes ? Expliquez pourquoi l'un·e des étudiant·e·s a obtenu une réponse erronée.
   Les étudiant·e·s ont toutes et tous entré la bonne formule sur la calculatrice, mais les méthodes employées sont différentes d'un.e étudiant.e à l'autre :
   - Erwan a recopié le résultat de chaque calcul sur sa copie, en arrondissant. Puis, pour la suite des calculs, il a entré dans la calculatrice les valeurs arrondies écrites sur sa copie. Cela introduit des erreurs aboutissant à un résultat final qui est trop approximatif.
   - Pour les calculs intermédiaires de $m(X)$ et $m(X^2)$, Sophie n'a calculé sur sa calculatrice que le numérateur. Elle a conservé sur sa copie une expression sous forme de fraction, sans effectuer d'arrondi. Elle a ensuite pu réutiliser ces expressions dans les calculs suivants sans introduire d'erreur liée à des arrondis approximatifs. Pour le calcul de $Var(X)$, elle a recopié sur sa copie une valeur arrondie, mais ne l'a pas utilisée sur la calculatrice (grâce à l'utilisation de la touche **Ans**). Ainsi elle n'a pas introduit d'erreur liée à des arrondis approximatifs.
   - Anaïs a écrit sur sa copie des valeurs approximatives (arrondies) de $m(X)$, $m(X^2)$ et $Var(X)$. Mais contrairement à Erwan, Anaïs n'a pas réutilisé ces valeurs arrondies pour effectuer les calculs suivants : elle a utilisé des valeurs stockées dans la mémoire de la calculatrice sans les arrondir (elle a stocké $m(X)$ sous le nom "M", $m(X^2)$ sous le nom "N" et $Var(X)$ sous le nom "V"). Ainsi elle n'a pas introduit d'erreur liée à des arrondis approximatifs.
   - Élisa a décidé de n'écrire aucun calcul intermédiaire et de calculer directement $s(X)$ à partir des formules donnant $m(X)$, $m(X^2)$ et $Var(X)$. De cette manière, elle n'a pas introduit d'erreur liée à des arrondis approximatifs.

   C'est pourquoi Sophie, Anaïs, et Élisa obtiennent la même réponse, qui est correcte, alors que Erwan obtient une réponse erronée du fait de l'utilisation de valeurs arrondies lors des calculs intermédiaires.

2. On demande ensuite de calculer l'écart-type de l'âge auquel 90 autres enfants ont perdu leur première *dent de laie*. Comme l'échantillon n'est plus le même, il faut cette fois-ci faire les calculs suivants :

$m(X) = \frac{5\times4+36\times5+35\times6+14\times7}{90}$
$m(X^2) = \frac{5\times4^2+36\times5^2+35\times6^2+14\times7^2}{90}$
$Var(X) = m(X^2) - (m(X))^2$
$s(X) = \sqrt{Var(X)}$

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

Voici des exemples de rédactions inspirées de ces étudiant·e·s :

Avec la méthode de Sophie

$m(X) = \frac{5\times4+36\times5+35\times6+14\times7}{90} = \frac{508}{90}$
$m(X^2) = \frac{5\times4^2+36\times5^2+35\times6^2+14\times7^2}{90} = \frac{2\,926}{90}$
$Var(X) \simeq \frac{2\,926}{90} - \left(\frac{508}{90}\right)^2 \simeq 0{,}6514$
$s(X) = \sqrt{Var} \simeq 0{,}807$ an.
Donc l'écart-type est d'environ 0,807 an, c'est à dire 295 jours.

Calculs correspondants

```
5×4+36×5+35×6+14×7
                              508
5×4²+36×5²+35×6²+14×7²
                             2926
2926÷90-(508÷90)²
                     0.6513580247
√(Ans)
                     0.8070675465
365*Ans
                      294.5796545
```

Avec la méthode d'Anaïs

$m(X) = \frac{5\times4+36\times5+35\times6+14\times7}{90} \simeq 5{,}64$
$m(X^2) = \frac{5\times4^2+36\times5^2+35\times6^2+14\times7^2}{90} \simeq 32{,}51$
$Var(X) = m(X^2) - (m(X))^2 \simeq 0{,}6514$
$s(X) = \sqrt{Var} \simeq 0{,}807$ an.
Donc l'écart-type est d'environ 0,807 an, c'est à dire 295 jours.

Calculs correspondants

```
(5×4+36×5+35×6+14×7)÷90→M
                      5.644444444
(5×4²+36×5²+35×6²+14×7²)÷90→N
                      32.51111111
N-M²→V
                     0.6513580247
√(V)→S
                     0.8070675465
365*S
                      294.5796545
```

Avec la méthode d'Élisa

$s(X) = \sqrt{\frac{5\times4^2+36\times5^2+35\times6^2+14\times7^2}{90} - \left(\frac{5\times4+36\times5+35\times6+14\times7}{90}\right)^2}$
$\simeq 0{,}807$ an.
Donc l'écart-type est d'environ 0,807 an, c'est à dire 295 jours.

Calculs correspondants

```
√((5×4²+36×5²+35×6²+14×7²)÷90
-((5×4+36×5+35×6+14×7)÷90)²)
                     0.8070675465
365*Ans
                      294.5796545
```

Exercice 8 : Utilisation du symbole $\sum$ corrigé en ligne

1. Calculez, sans utiliser la calculatrice, chacune des sommes suivantes : $\sum_{n=1}^{2}(-n)$, $\sum_{n=-2}^{2}(-n)$, $\sum_{k=-2}^{2}(-k)^2$.

   (a) $\sum_{n=1}^{2}(-n) = -1 - 2 = -3.$

   (b) $\sum_{n=-2}^{2}(-n) = -(-2) - (-1) - 0 - 1 - 2\,.$
   $= 2 + 1 - 1 - 2 = 0$

(c) $\sum_{k=-2}^{2}(-k)^2 = (2)^2 + (1)^2 + (0)^2 + (-1)^2 + (-2)^2$.
$= 10$

2. Écrivez avec le symbole $\sum$ la somme suivante : $\frac{6-3}{6^2} + \frac{7-3}{7^2} + \frac{8-3}{8^2} + \frac{9-3}{9^2} + \frac{10-3}{10^2} + \frac{11-3}{11^2}$.
$$\sum_{i=6}^{11} \frac{i-3}{i^2}$$

3. Calculez $\sum_{i=1}^{5}(x_i - y_i)^2$ lorsque $x_1 = 5$, $x_2 = 6$, $x_3 = 7$, $x_4 = 8$, $x_5 = 9$, $y_1 = 1$, $y_2 = 4$, $y_3 = 9$, $y_4 = 4$ et $y_5 = 5$ (sans utiliser de liste sur la calculatrice).
$$\sum_{i=1}^{5}(x_i - y_i)^2$$
$$= (x_1 - y_1)^2 + (x_2 - y_2)^2 + (x_3 - y_3)^2 + (x_4 - y_4)^2 + (x_5 - y_5)^2$$
$$= (5-1)^2 + (6-4)^2 + (7-9)^2 + (8-4)^2 + (9-5)^2 = 56$$

Chapitre 1: Statistiques descriptives univariées

Exercice 9 : Types de variables

1. Mme Meyer, travaillant dans un bureau d'études statistiques, a recueilli les données suivantes dans le cadre de son travail :
   (a) la douleur ressentie par différents patients atteints de la même pathologie
   (b) les marques des téléphones vendus en 2025
   (c) le nombre de personnes des ménages français
   (d) la pluie tombée à Dijon au cours des différents mois de l'année.

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

2. De même, précisez quel est le type de variable dans les situations suivantes :
   (a) la couleur des voitures stationnées dans le campus
   (b) le temps mis par des rats pour sortir d'un labyrinthe
   (c) les numéros de téléphones figurant dans les pages blanches dijonnaises
   (d) le taux de testostérone parmi les patients souffrant de troubles cognitifs
   (e) les prénoms des enfants nés en 2025.

Exercice 10 : Taille des ménages

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

Nombre de personnes dans le ménage	1	2	3	4	5	6
Effectifs : nombre de ménages	1013	1276	1531	1153	588	145

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

Exercice 11 : Âge et performances mémorielles

On considère un échantillon de 15 personnes de 38 à 85 ans, auxquelles on attribue une note indiquant leurs performances mémorielles. On note leur âge X et leur note de performances mémorielles Y.

Les données mesurées sont les suivantes :

X	47	68	55	45	39	38	53	55	85	53	48	40	71	79	64
Y	42	33	40	28	77	63	20	50	23	30	59	55	44	32	21

Déterminer la nature de chacune des variables X et Y.

Exercice 12 : Moyens de transport

1. **Utilisation des transports en commun :**

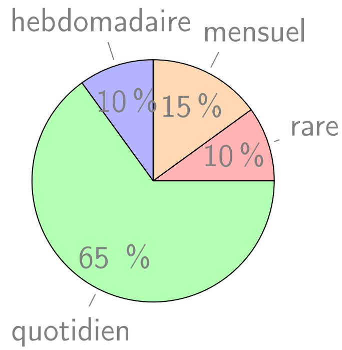
   (a) Un étudiant en sociologie interroge 40 personnes de son entourage sur leur utilisation des transports en commun. Il réalise le graphique ci-dessous après les avoir regroupés en quatre groupes :

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

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

   hebdomadaire : entre une et cinq fois par semaine

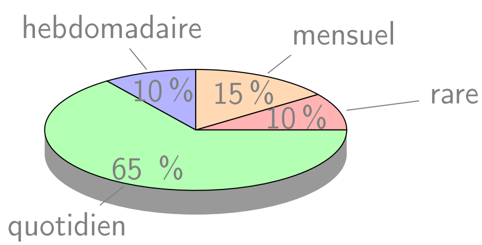
   quotidien : au moins 5 fois par semaine



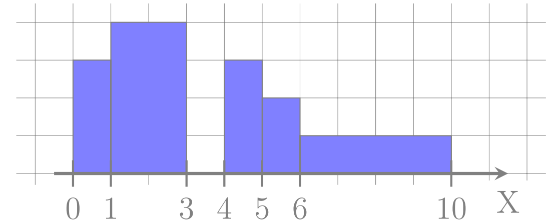
   i. Déterminez les effectifs des différentes modalités.

   ii. Quelle est, au sein de cet échantillon, la proportion d'individus qui utilisent les transports en commun moins d'une fois par semaine ?

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



2. **Utilisation de la voiture :**
   Pour le même échantillon de 40 personnes, l'étudiant désigne par X le temps hebdomadaire passé dans une voiture (exprimé en heures). Il trace un histogramme sur une feuille quadrillée :



   (a) Déterminez les fréquences des différentes classes.

X	$[0\,;1[$	$[1\,;3[$	$[3\,;4[$	$[4\,;5[$	$[5\,;6[$	$[6\,;10[$
Fréquence						

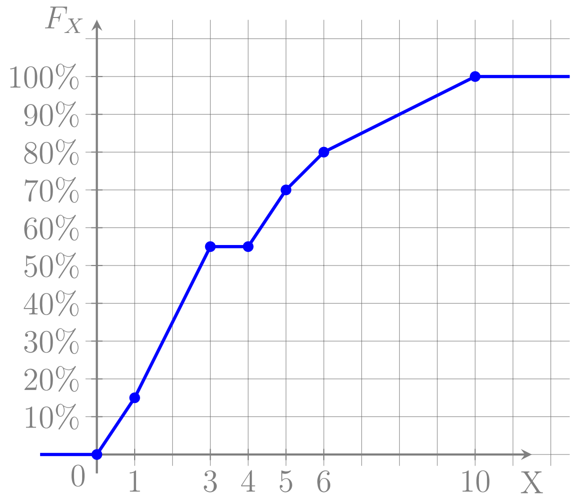
   (b) Comparez les fréquences de la première et de la dernière classes, ainsi que les hauteurs des rectangles correspondants, puis commentez.

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

   (a) Déterminez graphiquement (et de manière approchée) la médiane et les quartiles de X.

   (b) Déterminez graphiquement (et de manière approchée) les proportions $\mathbb{P}_r[X \leqslant 2]$, $\mathbb{P}_r[X \geqslant 9]$ et $\mathbb{P}_r[2 \leqslant X \leqslant 9]$.

   (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 13 : Pays fondateurs de la zone euro

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

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

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

Exercice 14 : Développement psychomoteur

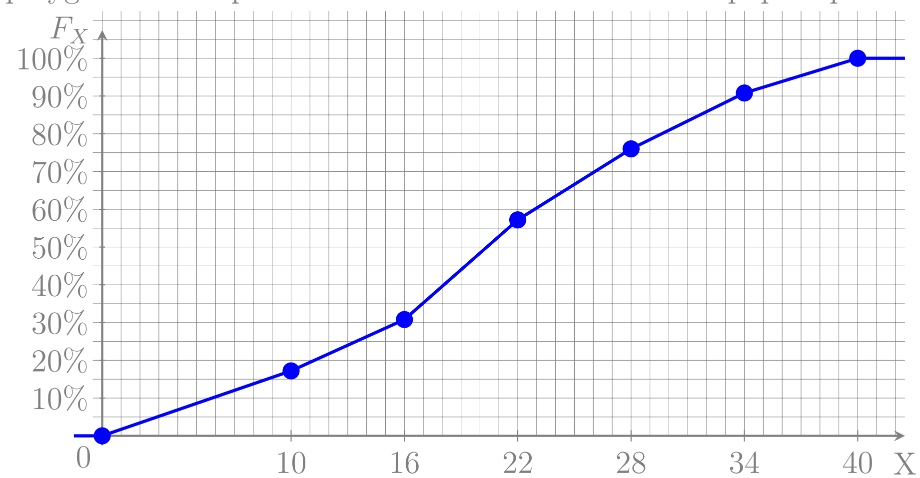
Les données ci-dessous décrivent, pour un groupe de 220 bébés âgés de deux ans, la variable statistique X représentant le "score de développement psychomoteur" (SDP) :

SDP	[50 ; 60[	[60 ; 70[	[70 ; 80[	[80 ; 90[	[90 ; 100[	[100 ; 110[	[110 ; 120[	[120 ; 130[
Effectif	1	8	31	49	53	46	22	10

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

Exercice 15 : Méthodes d'apprentissage

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



1. Déterminez graphiquement la médiane et les quartiles de la variable statistique X.
2. (a) Donnez une approximation de $\mathbb{P}_r[X < 25]$.

   (b) Quelle est environ la proportion d'enfants dont la note est supérieure à 18 ?
   (c) Combien vaut environ $\mathbb{P}_r[20 \leqslant X < 30]$?

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

X	[0 ; 10[	[10 ; 16[	[16 ; 22[	[22 ; 28[	[28 ; 34[	[34 ; 40[
Effectif	43	34	66	47	37	23

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

Exercice 16 : Moyenne et écart type

Calculez la moyenne et l'écart-type des données de l'Exercice 10.

Exercice 17 : Types de variables

corrigé en ligne

1. Mr Lecoq, travaillant dans un bureau d'études statistiques, a recueilli les données suivantes dans le cadre de son travail :
   (a) la popularité des différents footballeurs de l'équipe de France
   (b) le nombre de pages des livres d'une bibliothèque
   (c) le poids des appareils photos vendus en 2025
   (d) les noms des rues de Dijon.

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

Cas	Population	Variable	Nature
(a)	les différents footballeurs de l'équipe de France	la popularité	qualitative ordinale
(b)	l'ensemble des livres de cette bibliothèque	le nombre de pages	quantitative discrète
(c)	l'ensemble des appareils photos vendus en 2025	le poids	quantitative continue
(d)	l'ensemble des rues de Dijon	le nom	qualitative nominale

2. De même, précisez quel est le type de variable dans les situations suivantes :
   (a) les numéros de sécurité sociale des enseignants de l'Université de Bourgogne
   qualitative nominale
   (b) le nombre de crises d'épilepsie de différents patients au cours du mois d'octobre 2025
   quantitative discrète
   (c) la sensibilité aux questions écologiques des habitants d'un même quartier
   qualitative ordinale
   (d) le temps hebdomadaire passé devant la télévision par des enfants
   quantitative continue
   (e) le nombre moyen d'enfants par femme dans chaque pays européen
   quantitative continue
   (f) le nombre de rendez-vous médicaux que différents patients ont pris au cours de l'année 2025
   quantitative discrète
   (g) le nombre d'enfants de chaque femme dijonnaise
   quantitative discrète

(h) les langues officielles des différents pays européens
qualitative nominale

(i) le degré de satisfaction des usagers des transports en commun
qualitative ordinale

(j) le nombre de pièces des appartements en location à Dijon
quantitative discrète.

Exercice 18 : Questionnaire de logique

corrigé en ligne

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

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

1. Quel a été le taux de réponse à la question 4 ?
   En tout, 9 personnes (parmi les 12 interrogées) ont répondu à la question 4. Le taux de réponse à cette question est donc $\frac{9}{12} = 0{,}75$.
2. Quelle proportion de sujets a répondu correctement à la question 6 ?
   Comme 6 personnes ont répondu correctement à la question 6, cette proportion est $\frac{6}{12} = 0{,}5$.
3. Parmi les individus ayant répondu à la question 10, quelle proportion a donné la bonne réponse ?
   Parmi les 11 personnes qui ont répondu à la question 10, il y en a 6 qui ont répondu correctement, soit une proportion de $\frac{6}{11} \simeq 0{,}545\,5$.
4. Le/la chercheur/chercheurse décide de résumer par une note les réponses données au questionnaire. Il/elle attribue un point par bonne réponse (et zéro point par réponse erronée ou manquante). Calculez les notes (sur 10) attribuée à chaque adolescent interrogé.

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

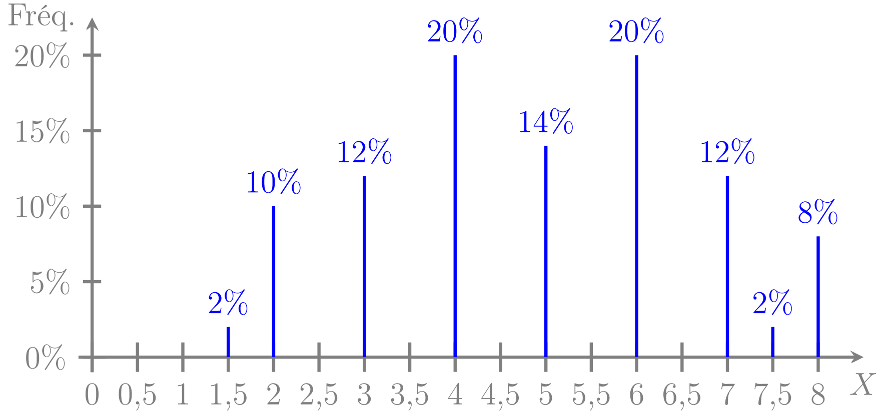
5. Calculez la moyenne, l'écart-type et la médiane de ces notes.
   moyenne : $m(X) = \frac{\sum x_i}{n} = \frac{6+2+5+\cdots+6}{12} = \frac{44}{12} \simeq 3{,}67$
   $m(X^2) = \frac{\sum x_i^2}{n} = \frac{6^2+2^2+5^2+\cdots+6^2}{12} = \frac{194}{12}$
   $Var(X) = m(X^2) - m(X)^2 = \frac{194}{12} - \left(\frac{44}{12}\right)^2 \simeq 2{,}72$
   Écart-type : $s(X) = \sqrt{Var(X)} \simeq 1{,}65$
   La médiane est la valeur numéro $\frac{12+1}{2} = 6{,}5$ (en ordonnant par ordre croissant), ou plutôt le milieu entre les valeurs numéro 6 et 7. C'est donc le milieu entre 3 et 4 c'est-à-dire $\frac{3+4}{2} = 3{,}5$.

Exercice 19 : Perception de la douleur

corrigé en ligne

Remarque : Exercice issu du contrôle commun de l'année 2019-2020.

Un groupe de 50 volontaires est soumis à des stimuli douloureux. On leur demande d'indiquer la douleur qu'ils ressentent, sur une échelle allant de 0 à 10, et on obtient les données représentées ci-dessous :



1. Déterminer les effectifs des différentes modalités et indiquez les dans le tableau ci-dessous.
   On obtient les effectifs en multipliant les fréquences par l'effectif total.
   Par exemple, pour $X = 2$, l'effectif est $\frac{10}{100} \times 50 = 5$

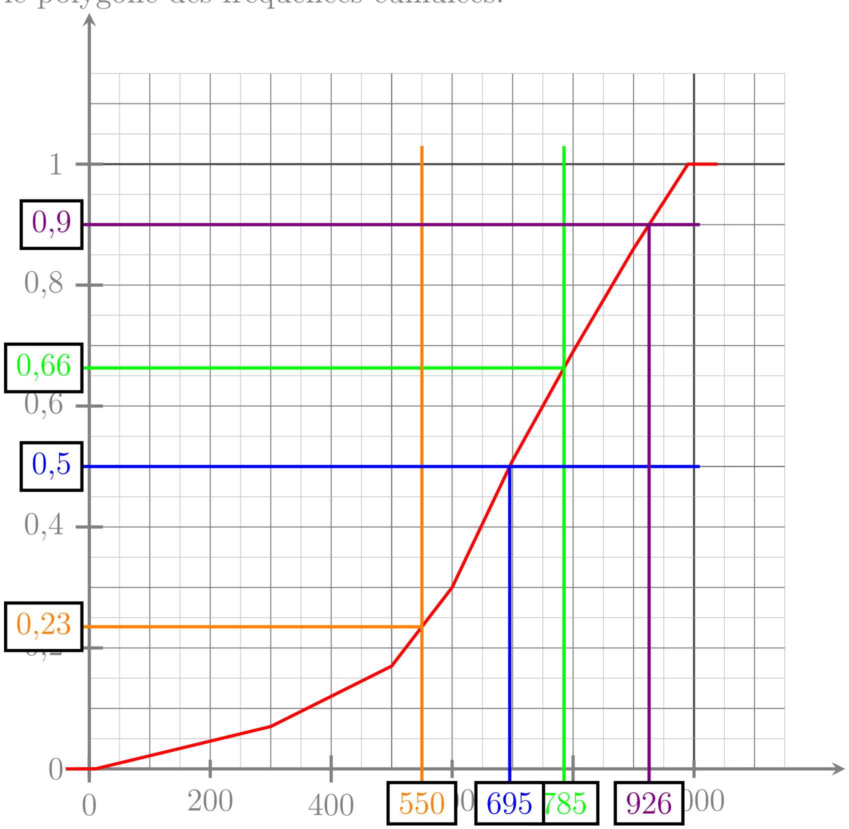
Modalités	1,5	2	3	4	5	6	7	7,5	8
Effectifs	1	5	6	10	7	10	6	1	4

2. Combien d'individus ont indiqué une douleur au moins égale à 3 ?
   Ils sont $6 + 10 + 7 + 10 + 6 + 1 + 4 = 44$ (obtenus en additionnant les effectifs correspondants aux douleurs 3, 4, 5, 6, 7, 7,5 et 8).
3. Calculer la moyenne et l'écart type de la douleur au sein de cet échantillon.
   moyenne : $m(X) = \frac{\sum_i x_i n_i}{n} = \frac{1{,}5\times1+2{,}0\times5+3{,}0\times6+\cdots+8{,}0\times4}{50} = \frac{246}{50} = 4{,}92$
   $m(X^2) = \frac{\sum_i x_i^2 n_i}{n} = \frac{1{,}5^2\times1+2^2\times5+3^2\times6+\cdots+8{,}0^2\times4}{50} = \frac{1\,377{,}5}{50}$
   $Var(X) = m(X^2) - m(X)^2 = \frac{1\,377{,}5}{50} - \left(\frac{246}{50}\right)^2 \simeq 3{,}34$
   Écart-type : $s(X) = \sqrt{Var(X)} \simeq 1{,}83$

Exercice 20 : Niveau d'anglais en France

corrigé en ligne

On considère le niveau en anglais des français nés en 1977, en se basant sur un test très répandu : le TOEIC®, qui attribue à chaque individu une note entre 10 et 990. Parmi les nombreuses études effectuées sur le niveau des Français, l'une d'entre elle indique des données pour 100 français nés en 1977. À partir des données de cette étude, on réalise le graphique ci-contre (réalisé sur un papier quadrillé), qui est le polygone des fréquences cumulées.



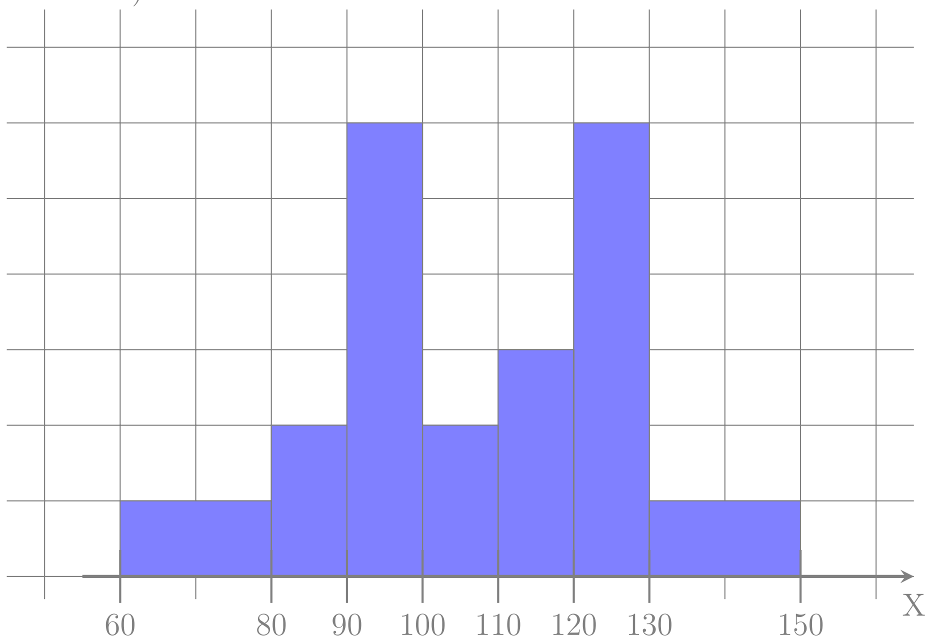
1. Quel est le score médian au sein de cet échantillon ?
   On lit sur le graphique (ajouté en bleu) le score 695 pour la fréquence cumulée 0,5.
   La médiane est donc environ 695.
2. On considère qu'un niveau avancé (correspondant à au moins B2 dans le "Cadre européen commun de référence pour les langues") est atteint lorsque l'on a au moins la note 785 au TOEIC®.
   Quelle est, au sein de cet échantillon, la proportion d'individus qui ont un niveau avancé en anglais ?
   On lit sur le graphique (ajouté en vert) la fréquence 0,66 pour le score 785. Donc (en utilisant la lettre Y pour le score au TOEIC) $\mathbb{P}_r[Y \leqslant 785] \simeq 0{,}66$, d'où l'on déduit que $\mathbb{P}_r[Y \geqslant 785] \simeq 1-0{,}66 = 0{,}34$.
   La proportion demandée vaut donc environ 0,34.
3. Le niveau B1 est un niveau où l'on commence à être indépendant et correspond aux scores entre 550 et 785 points.
   Déterminer quel est, au sein de cet échantillon, la proportion d'individus qui ont le niveau B1.
   On vient de voir que $\mathbb{P}_r[Y \leqslant 785] \simeq 0{,}66$. De plus on lit sur le graphique (ajouté en orange) que $\mathbb{P}_r[Y \leqslant 550] \simeq 0{,}23$.
   La proportion demandée est donc $\mathbb{P}_r[550 \leqslant Y \leqslant 785] \simeq 0{,}66 - 0{,}23 = 0{,}43$.
4. Combien d'individus, parmi ceux de l'échantillon, ont un niveau inférieur à B1 ?
   On a déjà vu que $\mathbb{P}_r[Y \leqslant 550] \simeq 0{,}23$. Cette proportion correspond donc environ à $100 \times 0{,}23 = 23$ individus
5. On souhaite caractériser les 10% d'individus ayant les meilleures notes. Compléter la phrase ci-dessous pour la rendre correcte :
   On lit sur la graphique (en violet) le score 926 pour la fréquence cumulée 0,9 (c'est à dire que 90% de l'échantillon a moins que ce score, tandis que 10% a plus que ce score). Ainsi,
   Dans cet échantillon, les 10% d'individus ayant la meilleure note sont ceux qui ont au moins la note 926 .

Exercice 21 : Personnalité des cadres

corrigé en ligne

On fait passer un test de personnalité à un échantillon de 69 cadres. Ce test attribue à chaque sujet un score X qui est d'autant plus élevé que l'individu a une personnalité autoritaire (le but de cette analyse serait de déterminer si les cadres sont plus autoritaires que le reste de la population).

On résume les données collectées par l'histogramme ci-dessous (où ils se trouve que les rectangles tombent exactement sur le quadrillage de la feuille).



1. Indiquez quelle est la variable statistique et quelle est sa nature.
   La variable statistique est le score X. Elle est quantitative (l'énoncé ne permet pas de dire précisemment si elle est continue ou discrète).
2. À partir de cet histogramme, déterminer les fréquences des différents intervalles.
   Pour chaque intervalle, on obtient la fréquence en divisant la surface du rectangle correspondant par la surface totale. On obtient :

Score X	[60 ; 80[	[80 ; 90[	[90 ; 100[	[100 ; 110[	[110 ; 120[	[120 ; 130[	[130 ; 150[
fréquence	0,087	0,087	0,261	0,087	0,130	0,261	0,087

3. En déduire quel est environ le score médian au sein de cet échantillon.
   On calcule les fréquences cumulées puis on déduit la médiane :

Score X	[60 ; 80[	[80 ; 90[	[90 ; 100[	[100 ; 110[	[110 ; 120[	[120 ; 130[	[130 ; 150[
Fréquences cumulées	0,087	0,174	0,435	0,522	0,652	0,913	1,000

Classe de la médiane : [100 ; 110[

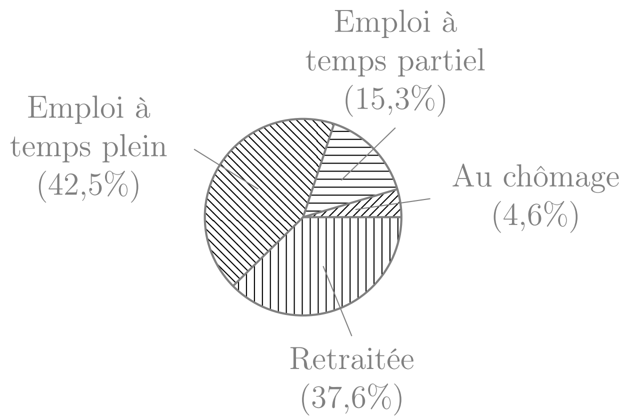
Méd $\simeq a_i + \frac{a_{i+1}-a_i}{F_X(a_{i+1})-F_X(a_i)}(0{,}5 - F_X(a_i)) \simeq 100 + \frac{110-100}{0{,}522-0{,}435}(0{,}5 - 0{,}435) \simeq 107{,}471$

Exercice 22 : Situation d'emploi corrigé en ligne

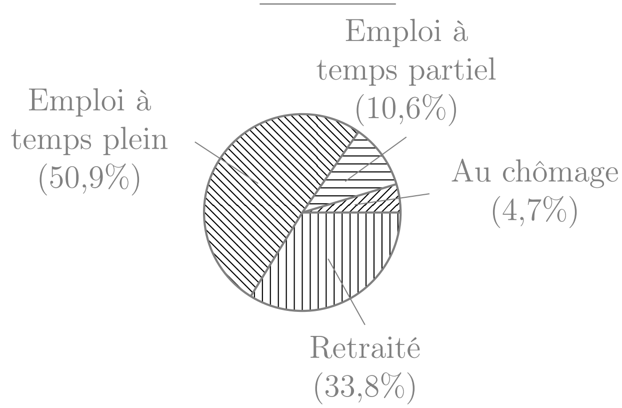
Remarque : Exercice issu du contrôle commun de l'année 2019-2024, qui portait sur les inégalités de genre en France.

On regroupe les personnes actives (en situation d'emploi ou au chômage) et les retraité·e·s (au sein de la population française), obtenant ainsi un total de 23 946 000 femmes et 23 627 000 hommes, qui se répartissent de la manière suivante (données de 2021/2022) :

Femmes



Hommes



1. Pourquoi choisir de représenter ces données sous forme de "camemberts", et pas sous la forme d'un "diagramme en bâton" ou d'un "histogramme" ?
   La variable est qualitative, et les autres types de diagrammes (diagramme en bâton, histogramme, polygone des fréquences cumulées) ne sont possibles que pour des variables quantitatives.

2. Pour les femmes Françaises, calculer les effectifs de chacune des quatre modalités (chômage / emploi à temps partiel / emploi à plein temps / retraitée).
   On calcule les effectifs en multipliant chaque pourcentage par le nombre total de femmes considérées (c'est à dire 23 946 000). Par exemple le nombre de femmes au chômage est environ $23\,946\,000 \times \frac{4{,}6}{100} = 1\,101\,516$. On obtient de même :

Modalités	Emploi à temps plein	Emploi à temps partiel	Retraité	Au chômage
Fréquence	42,5%	15,3%	37,6%	4,6%
Effectif	$\simeq 10\,177\,000$	$\simeq 3\,664\,000$	$\simeq 9\,004\,000$	$\simeq 1\,102\,000$

3. Au sein des femmes qui exercent un emploi, quelle proportion exerce à temps partiel ?
   On a obtenu qu'environ $10\,177\,000 + 3\,664\,000 = 13\,841\,000$ femmes exercent un emploi, parmi lesquelles environ 3 664 000 l'exercent à temps partiel. La proportion demandée est donc environ $\frac{3\,664\,000}{13\,841\,000} \simeq 0{,}265$.

4. Pour les hommes français, calculer les effectifs de chacune des quatre modalités (chômage / emploi à temps partiel / emploi à plein temps / retraitée). Indiquer ensuite quelle est, parmi les hommes qui exercent un emploi, la proportion qui exercent à temps partiel.
   On calcule les effectifs en multipliant chaque pourcentage par le nombre total d'hommes considérés (c'est à dire 23 627 000) :

Modalités	Emploi à temps plein	Emploi à temps partiel	Retraitée	Au chômage
Fréquence	50,9%	10,6%	33,8%	4,7%
Effectif	$\simeq$ 12 026 000	$\simeq$ 2 504 000	$\simeq$ 7 986 000	$\simeq$ 1 110 000

On a obtenu qu'environ 12 026 000 + 2 504 000 = 14 530 000 hommes exercent un emploi, parmi lesquelles environ 2 504 000 l'exercent à temps partiel. Donc parmi les hommes qui exercent un emploi, la proportion de temps partiels est environ $\frac{2\,504\,000}{14\,530\,000} \simeq 0{,}172$.

Exercice 23 : Regroupement en classes

corrigé en ligne

Voici les notes obtenues par un groupe d'élèves lors d'un contrôle noté sur 10 points :
7,5 / 6 / 4 / 7,5 / 4,5 / 6,5 / 6,5 / 2,5 / 6,5 / 7 / 5 / 3 / 3 / 9 / 6 / 6,5 / 4,5 / 2 / 8 / 5 / 3 / 4 / 8 / 7,5 / 8 / 7,5 / 4 / 6 / 8,5 / 8,5 / 4 / 7 / 2 / 6 / 5 / 3 / 4,5 / 6 / 9 / 5 / 8 et 4,5

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

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

2. Quelle est la taille de l'échantillon ?
   La taille de l'échantillon est $n = 42$.
3. Donner la moyenne, l'écart-type et la médiane.
   moyenne : $m(X) = \frac{\sum_i x_i\, n_i}{n} = \frac{2\times2+2{,}5\times1+3\times4+\cdots+9\times2}{42} = \frac{239{,}5}{42} \simeq 5{,}7$
   $m(X^2) = \frac{\sum_i x_i^2\, n_i}{n} = \frac{2^2\times2+2{,}5^2\times1+3^2\times4+\cdots+9^2\times2}{42} = \frac{1\,529{,}75}{42}$
   $Var(X) = m(X^2) - m(X)^2 = \frac{1\,529{,}75}{42} - \left(\frac{239{,}5}{42}\right)^2 \simeq 3{,}91$
   Écart-type : $s(X) = \sqrt{Var(X)} \simeq 1{,}98$
   La médiane est la valeur numéro $\frac{42+1}{2} = 21{,}5$ (en ordonnant par ordre croissant). C'est donc 6.
4. Transformez les données en les rangeant en classes d'amplitude 1,5. Calculez les fréquences et les fréquences cumulées des classes.

classes	[0 ; 1,5[	[1,5 ; 3[	[3 ; 4,5[	[4,5 ; 6[	[6 ; 7,5[	[7,5 ; 9[	[9 ; 10,5[
effectifs	0	3	8	8	11	10	2
fréquences	0	0,071	0,19	0,19	0,262	0,238	0,048
fréquences cumulées	0	0,071	0,261	0,451	0,713	0,951	0,999

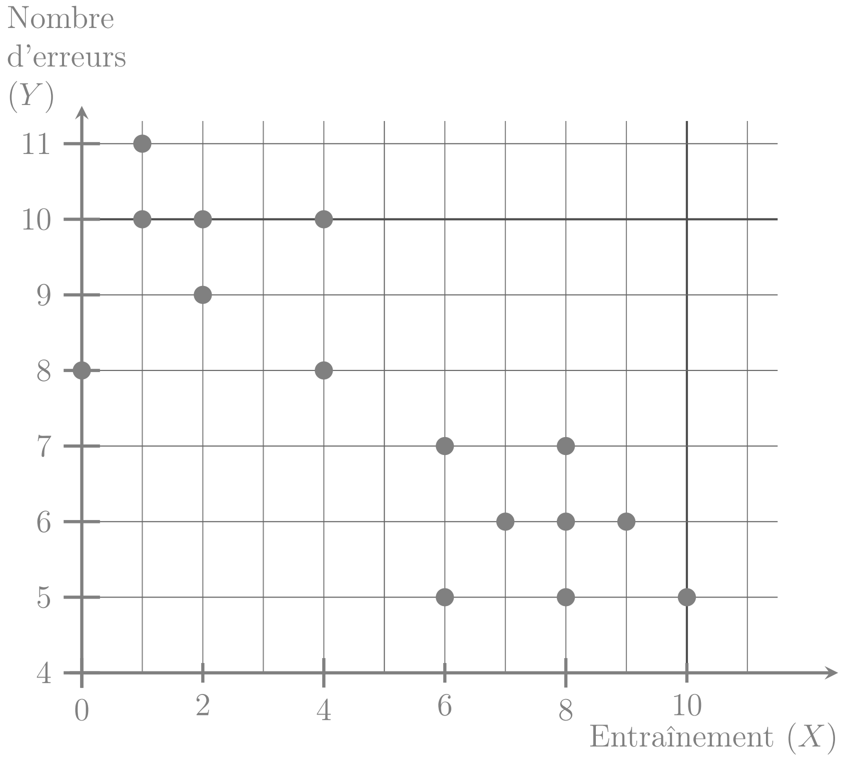
5. Calculez la moyenne, l'écart-type et la médiane des notes regroupées en classes. Que remarque-t-on ?
   moyenne :
   $$\begin{aligned} m(X) &= \frac{\sum c_i\, n_i}{p} \\ &= \frac{0{,}75\times0+2{,}25\times3+3{,}75\times8+\cdots+9{,}75\times2}{42} \\ &= \frac{255}{42} \\ &\simeq 6{,}07 \end{aligned}$$
   $$\begin{aligned} m(X^2) &= \frac{\sum c_i^2\, n_i}{n} \\ &= \frac{0{,}75^2\times0+2{,}25^2\times3+3{,}75^2\times8+\cdots+9{,}75^2\times2}{42} \\ &= \frac{1\,720{,}125}{42} \end{aligned}$$
   $$\begin{aligned} Var(X) &= m(X^2) - m(X)^2 \\ &= \frac{1\,720{,}125}{42} - \left(\frac{255}{42}\right)^2 \\ &\simeq 4{,}09 \end{aligned}$$
   Écart-type : $s(X) = \sqrt{Var(X)} \simeq 2{,}02$
   Classe de la médiane : [6 ; 7,5[
   Méd $\simeq a_i + \frac{a_{i+1}-a_i}{F_X(a_{i+1})-F_X(a_i)}(0{,}5 - F_X(a_i)) \simeq 6 + \frac{7{,}5-6}{0{,}713-0{,}451}(0{,}5 - 0{,}451) \simeq 6{,}281$
   Ce regroupement des données en classes donne lieu à des erreurs de l'ordre de 0,25 points (dans le cas de la moyenne), ce qui est de l'ordre de 4% de la note.

Chapitre 2: Statistiques descriptives bivariées

Exercice 24 : Entraînement à un exercice de logique

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

Plus précisément, pour un échantillon de 15 enfants, on considère le nombre Y d'erreurs commises à un test de logique, tandis que le nombre d'exercices similaires qu'ils ont déjà effectués auparavant est noté X. On regroupe ces résultats sous la forme d'un nuage de points (effectué sur un papier quadrillé) :



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

Entraînement X															
Nombre d'erreurs Y															

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

Exercice 25 : Stress et temps de réponse

On s'intéresse au temps de réponse de rats à des stimuli visuels. On constate que dans certaines conditions d'élevage, un parte des rats deviennent très stressés alors que la plupart restent beaucoup moins stressés. On 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	2,84	7,95	2,85	7,25	3,18	2,8	8,22	2,65	2,88	2,69	3,09
temps de réponse Y	493	694	499	704	457	467	697	528	478	488	507

1. Déterminez le coefficient de corrélation des rangs (de Spearman) pour les variables X et Y. Ces variables sont-elles très corrélées ?
2. Déterminez le coefficient de corrélation linéaire des variables X et Y. Ces variables sont-elles très corrélées ?
3. Suite à une perte d'échantillon, on ne connait plus le niveau de stress d'un rat, dont on a mesuré qu'il met 698 ms à réagir aux stimuli visuels. À quel niveau de stress s'attend-on pour ce rat ?
4. Tracez le nuage de points des variables X et Y.
5. Commentez vos réponses aux questions 1, 2 et 3 en vous appuyant sur le nuage de points.

Exercice 26 : Revenus et espérance de vie

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

pays	revenu/hab. X	espérance de vie Y
Arménie	9,1	74,8
Bénin	2,2	60
Costa Rica	17,1	79,6
Djibouti	3,6	63,5
Estonie	31,5	77,6
Finlande	44	81,1
Guinée-Bissau	1,8	58,9
Honduras	5,5	74,6
Iran	20	75,5
Jordanie	12,5	74,1
Koweït	69,7	74,7

1. Déterminez les médianes du revenu X et de l'espérance de vie Y.
2. Déterminez le coefficient de corrélation linéaire de X et Y.
3. Déterminez le coefficient de corrélation des rangs (de Spearman) de X et Y.
4. Commentez les résultats obtenus après avoir tracé le nuage de points de X et Y.
5. On ajoute à ces données réelles un pays imaginaire nommé “Utopia”, où le revenu par habitant est de 1 990 et l'espérance de vie est de 1 420. Calculez les moyennes et médianes qu'on obtiendrait alors, ainsi que les coefficients de corrélation. Enfin, commentez les résultats obtenus.

Exercice 27 : Données sur un groupe d'étudiants

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

Brigitte	Jean-Pierre	Chloe
de mauvaise humeur	de bonne humeur	de relativement bonne humeur
4 frères/soeurs	0 frère/soeur	0 frère/soeur
1m59	1m88	1m64
bac : 11,45/20	bac : 13,37/20	bac : 12,34/20

Magali	Myriam	David
de très bonne humeur	de bonne humeur	de bonne humeur
3 frères/soeurs	3 frères/soeurs	2 frères/soeurs
1m65	1m69	1m80
bac : 12,68/20	bac : 12,44/20	bac : 10,81/20

Sylvie	Bernard	Karine
de bonne humeur	de bonne humeur	de très bonne humeur
3 frères/soeurs	2 frères/soeurs	2 frères/soeurs
1m59	1m89	1m59
bac : 14,36/20	bac : 13,15/20	bac : 11,51/20

Ces données sont-elles appariées ? Dès lors, est-il possible de calculer les coefficients de correlations entre certaines de ces variables ? *On ne demande pas de calculer les coefficients, mais juste d'inquer quel·s coefficients pourraient être calculés.*

Exercice 28 : Thérapie pour soulager l'angoisse

Pour mettre en évidence l'efficacité d'une thérapie visant à réduire l'angoisse de personnes victimes d'agressions, des psychologues ont observé 16 sujets avant et après la thérapie en affectant à chaque sujet un score (plus le score est élevé, plus fort est le niveau d'angoisse). Les données sont les suivantes.

Avant	28	42	42	15	40	31	27	21	36	25	26	34	23	27	30	30
Après	7	41	25	4	44	32	16	15	31	18	9	31	6	17	10	13

On trouve que le coefficient de correlation linéaire vaut 0,8116.

Est-il raisonnable de calculer des droites de regression à partir de ces données ? Si c'est le cas, y a-t-il une des deux droites de régression, qui a plus d'intérêt du point de vue thérapeutique ?

Exercice 29 : Extraversion

corrigé en ligne

Dans une expérience, 11 sujets ont rempli un questionnaire de personnalité d'Eysenck comprenant, entre autres, une échelle de sociabilité (X) et une échelle d'impulsivité (Y). On a obtenu les résultats suivants :

sociabilité X	6,1	5,4	7,6	5,5	4,9	5,7	8,8	2,7	4,6	6,8	7,3
impulsivité Y	4,7	0,8	3,2	4,3	2,9	3	2,8	2	4,1	5,6	2,2

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

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

   moyenne : $m(X) = \frac{\sum x_i}{n} = \frac{6{,}1+5{,}4+7{,}6+\cdots+7{,}3}{11} = \frac{65{,}4}{11} \simeq 5{,}95$

   $m(X^2) = \frac{\sum x_i^2}{n} = \frac{6{,}1^2+5{,}4^2+7{,}6^2+\cdots+7{,}3^2}{11} = \frac{416{,}3}{11}$

   $Var(X) = m(X^2) - m(X)^2 = \frac{416{,}3}{11} - \left(\frac{65{,}4}{11}\right)^2 \simeq 2{,}5$

   Écart-type : $s(X) = \sqrt{Var(X)} \simeq 1{,}58$

   moyenne : $m(Y) = \frac{\sum x_i}{n} = \frac{4{,}7+0{,}8+3{,}2+\cdots+2{,}2}{11} = \frac{35{,}6}{11} \simeq 3{,}24$

   $m(Y^2) = \frac{\sum x_i^2}{n} = \frac{4{,}7^2+0{,}8^2+3{,}2^2+\cdots+2{,}2^2}{11} = \frac{133{,}72}{11}$

   $Var(Y) = m(Y^2) - m(Y)^2 = \frac{133{,}72}{11} - \left(\frac{35{,}6}{11}\right)^2 \simeq 1{,}68$

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

   $m(X\,Y) = \frac{\sum x_i\,y_i}{n} = \frac{6{,}1\times4{,}7+5{,}4\times0{,}8+\ldots+7{,}3\times2{,}2}{11} = \frac{215{,}31}{11} \simeq 19{,}574$

   $Cov(X,Y) = m(X\,Y) - m(X)\,m(Y) = \frac{215{,}31}{11} - \frac{65{,}4}{11}\frac{35{,}6}{11} \simeq 0{,}332$

   $r(X,Y) = \frac{Cov(X,Y)}{s(X)\,s(Y)} = \frac{0{,}332}{1{,}58\times1{,}3} \simeq 0{,}161\,6$

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

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

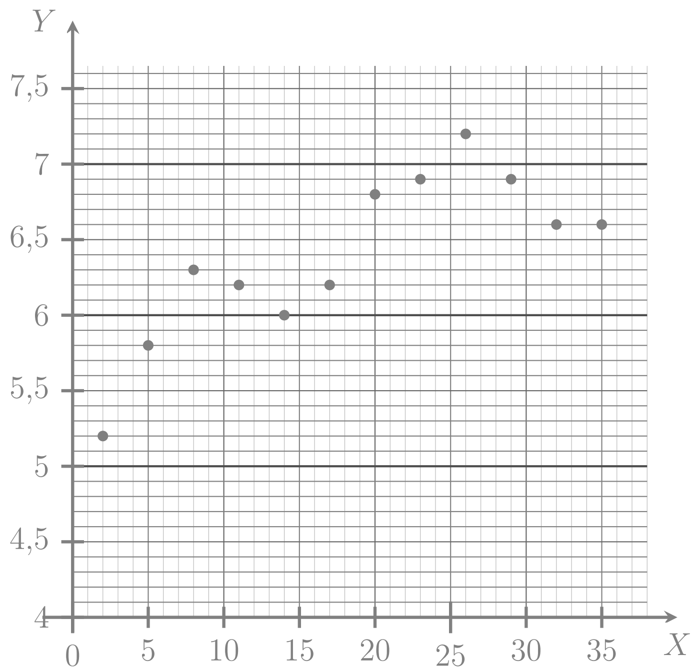
Exercice 30 : Sport et mathématiques

corrigé en ligne

Remarque : Exercice issu du CT de 2021-2022.

Mr Henry est chercheur en médecine. Il cherche à savoir si la pratique du sport, lorsqu'elle met en jeu une plus forte implication des élèves et des activités physiques cognitivement stimulantes, permet d'améliorer les performance dans d'autres disciplines comme les mathématiques. Pour cela, il considère 12 classes de CM1, où un certain nombre d'heures parmi les cours de sports de l'année sont repensées pour les rendre plus stimulantes (ce nombre d'heure est noté X). Il fait ensuite passer le même test de mathématiques à toutes ces classes, et regarde la moyenne obtenue par chaque classe (notée Y).

Il récolte ainsi des données, qu'il synthétise par le nuage de points ci-dessous :



1. Déterminer les nombres d'heures X et les moyennes en maths (Y) de chacune de ces 12 classes de CM1.

Nombre d'heures (X)	2	5	8	11	14	17	20	23	26	29	32	35
Moyenne en maths (Y)	5,2	5,8	6,3	6,2	6	6,2	6,8	6,9	7,2	6,9	6,6	6,6

2. Parmi le sous échantillon formé des classes où moins de 18 heures de sports ont été repensées pour les rendre plus stimulantes, quels sont la moyenne et l'écart type de Y ?
   Dans ce sous-échantillon, les notes Y sont : 5,2, 5,8, 6,3, 6,2, 6 et 6,2. On obtient alors $m(Y) \simeq 5{,}95$ et $s(Y) \simeq 0{,}3731$
3. Parmi le sous échantillon formé des classes où plus de 18 heures de sports ont été repensées pour les rendre plus stimulantes, quels sont la moyenne et l'écart type de Y ?
   Dans ce sous-échantillon, les notes Y sont : 6,8, 6,9, 7,2, 6,9, 6,6 et 6,6. On obtient alors $m(Y) \simeq 6{,}8333$ et $s(Y) \simeq 0{,}2055$

Exercice 31 : Reconnaissance de formes

corrigé en ligne

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

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

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

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

sujet	1	2	3	4	5	6	7	8	9	10	11	12	13	14
X	0	1	1	1	2	2	2	2	4	5	6	7	9	10
Y	3,4	3,4	3,5	4	2,9	3,2	3,5	3,5	3	2,5	2,7	2	1,2	1,3
rang x'	1	3	3	3	6,5	6,5	6,5	6,5	9	10	11	12	13	14
rang y'	9,5	9,5	12	14	6	8	12	12	7	4	5	3	1	2
$(x'-y')^2$	72,25	42,25	81	121	0,25	2,25	30,25	30,25	4	36	36	81	144	144

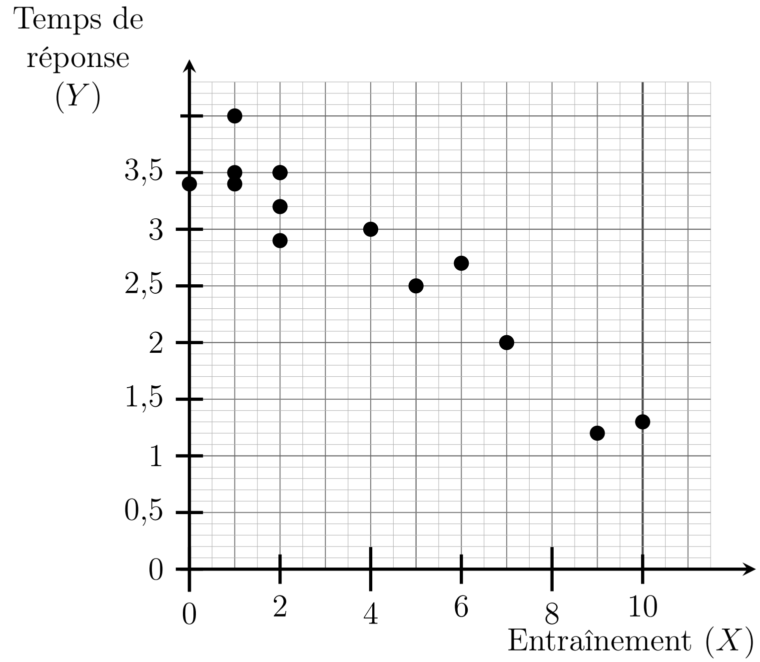
le coefficient de corrélation des rangs de Spearman est donc
$1-\left(6\times\frac{72{,}25+42{,}25+81+121+0{,}25+\cdots+144}{14\left(14^2-1\right)}\right)\simeq -0{,}812$
Cela montre un fort lien entre les deux variables, où Y diminue quand X augmente..

2. Calculez les moyennes et écart-types des variables X et Y, puis leur coefficient de corrélation linéaire. Interprétez la valeur de ce coefficient.
   moyenne : $m(X)=\frac{\sum x_i}{n}=\frac{0+1+1+\cdots+10}{14}=\frac{52}{14}\simeq 3{,}71$
   $m(X^2)=\frac{\sum x_i^2}{n}=\frac{0^2+1^2+1^2+\cdots+10^2}{14}=\frac{326}{14}$
   $Var(X)=m(X^2)-m(X)^2=\frac{326}{14}-\left(\frac{52}{14}\right)^2\simeq 9{,}49$
   Écart-type : $s(X)=\sqrt{Var(X)}\simeq 3{,}08$

   moyenne : $m(Y)=\frac{\sum x_i}{n}=\frac{3{,}4+3{,}4+3{,}5+\cdots+1{,}3}{14}=\frac{40{,}1}{14}\simeq 2{,}86$
   $m(Y^2)=\frac{\sum x_i^2}{n}=\frac{3{,}4^2+3{,}4^2+3{,}5^2+\cdots+1{,}3^2}{14}=\frac{124{,}19}{14}$
   $Var(Y)=m(Y^2)-m(Y)^2=\frac{124{,}19}{14}-\left(\frac{40{,}1}{14}\right)^2\simeq 0{,}666\,6$
   Écart-type : $s(Y)=\sqrt{Var(Y)}\simeq 0{,}816$

   $m(X\,Y)=\frac{\sum x_i\,y_i}{n}=\frac{0\times 3{,}4+3{,}4+\ldots+10\times 1{,}3}{14}=\frac{115{,}6}{14}\simeq 8{,}257$
   $Cov(X,Y)=m(X\,Y)-m(X)\,m(Y)=\frac{115{,}6}{14}-\frac{52}{14}\frac{40{,}1}{14}\simeq -2{,}382$
   $r(X,Y)=\frac{Cov(X,Y)}{s(X)\,s(Y)}=\frac{-2{,}382}{3{,}08\times 0{,}816}\simeq -0{,}948$
   Cela montre un fort lien linéaire entre les deux variables, où Y diminue quand X augmente.

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



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

4. Si un enfant a déjà fait 3 exercices de reconnaissance de formes, alors à combien estimeriez-vous son temps de réponse lors d'un prochain test ?
   Pour répondre à cette question, on utilise la droite de régression $D_{Y|X}$ (qui fait sens car il y a une forte corrélation linéaire) :
   on pose $a=\frac{Cov(X,Y)}{Var(X)}\simeq\frac{-2{,}382}{9{,}49}\simeq -0{,}25$ et $b=m(Y)-a\,m(X)\simeq 2{,}864-(-0{,}251)\times 3{,}714\simeq 3{,}8$
   D'où l'équation de la droite $D_{Y|X}$: $Y=-0{,}25\,X+3{,}8$
   Donc pour $x=3$, on s'attend à $y=-0{,}25\times 3+3{,}8=3{,}05$ ..

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 de régression $D_{X|Y}$:
on pose $a' = \frac{Cov(X,Y)}{Var(Y)} \simeq \frac{-2{,}38}{0{,}667} \simeq -3{,}57$ et $b' = m(X) - a'\,m(Y) \simeq 3{,}71 - (-3{,}57) \times 2{,}86 \simeq 13{,}92$
D'où l'équation de la droite $D_{X|Y}$: $X = -3{,}57\,Y + 13{,}92$
Donc pour $y = 1{,}6$, on s'attend à $x = -3{,}57 \times 1{,}6 + 13{,}92 = 8{,}208$
..

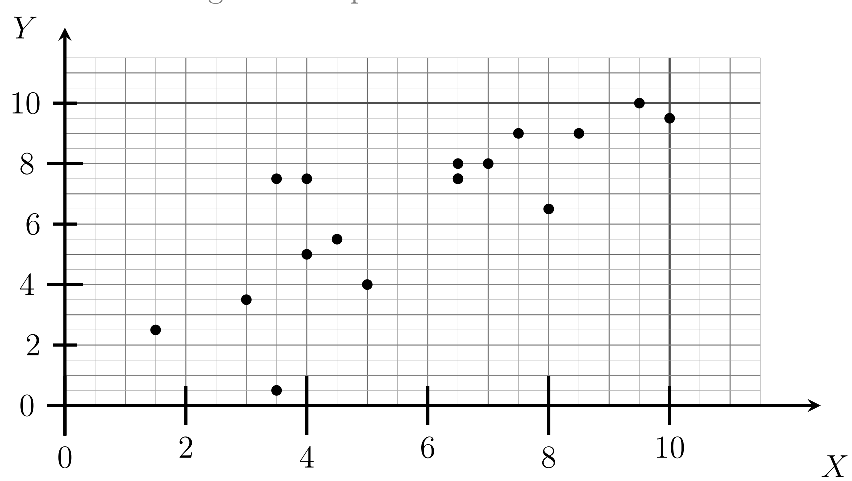
Exercice 32 : Méthodes d'apprentissage

corrigé en ligne

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

$X : \mathcal{MV}$	10	5	4,5	6,5	4	7,5	6,5	9,5	3,5	3,5	8,5	6,5	3	8	4	1,5	7
$Y : \mathcal{ME}$	9,5	4	5,5	7,5	7,5	9	8	10	0,5	7,5	9	7,5	3,5	6,5	5	2,5	8

1. Dessinez le nuage statistique de ces variables.



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

sujet	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
X	10	5	4,5	6,5	4	7,5	6,5	9,5	3,5	3,5	8,5	6,5	3	8	4	1,5	7
Y	9,5	4	5,5	7,5	7,5	9	8	10	0,5	7,5	9	7,5	3,5	6,5	5	2,5	8
rang x'	17	8	7	10	5,5	13	10	16	3,5	3,5	15	10	2	14	5,5	1	12
rang y'	16	4	6	9,5	9,5	14,5	12,5	17	1	9,5	14,5	9,5	3	7	5	2	12,5
$(x'-y')^2$	1	16	1	0,25	16	2,25	6,25	1	6,25	36	0,25	0,25	1	49	0,25	1	0,25

le coefficient de corrélation des rangs de Spearman est donc
$1 - \left(6 \times \frac{1+16+1+0{,}25+16+\cdots+0{,}25}{17\left(17^2-1\right)}\right) \simeq 0{,}830\,9.$

3. Calculez les moyennes, les écart-types et le coefficient de corrélation linéaire des deux variables X et Y.
moyenne : $m(X) = \frac{\sum x_i}{n} = \frac{10+5+4{,}5+\cdots+7}{17} = \frac{99}{17} \simeq 5{,}82$
$m(X^2) = \frac{\sum x_i^2}{n} = \frac{10^2+5^2+4{,}5^2+\cdots+7^2}{17} = \frac{671{,}5}{17}$
$Var(X) = m(X^2) - m(X)^2 = \frac{671{,}5}{17} - \left(\frac{99}{17}\right)^2 \simeq 5{,}59$
Écart-type : $s(X) = \sqrt{Var(X)} \simeq 2{,}36$

moyenne : $m(Y) = \frac{\sum x_i}{n} = \frac{9{,}5+4+5{,}5+\cdots+8}{17} = \frac{111}{17} \simeq 6{,}53$
$m(Y^2) = \frac{\sum x_i^2}{n} = \frac{9{,}5^2+4^2+5{,}5^2+\cdots+8^2}{17} = \frac{837{,}5}{17}$
$Var(Y) = m(Y^2) - m(Y)^2 = \frac{837{,}5}{17} - \left(\frac{111}{17}\right)^2 \simeq 6{,}631$
Écart-type : $s(Y) = \sqrt{Var(Y)} \simeq 2{,}58$

$m(X\,Y) = \frac{\sum x_i\,y_i}{n} = \frac{10\times 9{,}5+5\times 4+\ldots+7\times 8}{17} = \frac{728{,}5}{17} \simeq 42{,}853$
$Cov(X,Y) = m(X\,Y) - m(X)\,m(Y) = \frac{728{,}5}{17} - \frac{99}{17}\frac{111}{17} \simeq 4{,}829$
$r(X,Y) = \frac{Cov(X,Y)}{s(X)\,s(Y)} = \frac{4{,}829}{2{,}36\times 2{,}58} \simeq 0{,}793\,1$

4. On souhaite désormais estimer à quelle note s'attendre avec la méthode écrite pour un enfant ayant obtenu une note $x = 9$ avec la méthode verbale :
   (a) Quelle droite de régression peut-on utiliser pour répondre à cette question ? Donner son équation.
   La droite qui est pertinente pour cette question est $D_{Y|X}$.
   on pose $a = \frac{Cov(X,Y)}{Var(X)} \simeq \frac{4{,}829}{5{,}59} \simeq 0{,}8639$ et $b = m(Y) - a\,m(X) \simeq 6{,}529 - 0{,}863\,9 \times 5{,}824 \simeq 1{,}498$

D'où l'équation de la droite $D_{Y|X}$: $Y = 0{,}863\,9\,X + 1{,}498$

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

Donc pour $x = 9$, on s'attend à $y = 0{,}863\,9 \times 9 + 1{,}498 \simeq 9{,}273$.

Chapitre 3: probabilités

Exercice 33 : Situations d'emploi

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

Nom	Mme Besson	Mme Thomas	Mr Ferreira	Mr Alexandre	Mr Martin
Situation	emploi	chômage	emploi	emploi	retraite

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

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

Exercice 34 : Factorielles & coefficients binomiaux

1. Simplifiez les nombres suivants : $\dfrac{20!}{18!}$, $\dfrac{26!}{22! \times 4!}$.
2. Écrivez à l'aide de deux factorielles l'expression $7 \times 8 \times 9 \times 10$.
3. Calculez les nombres suivants : $\dbinom{26}{4}$, $\dbinom{13}{5}$, $\dbinom{13}{8}$, $\dbinom{11}{0}$, $\dbinom{15}{1}$.

Exercice 35 : Croyance religieuse

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

   Argumentez le fait que ces probabilités correspondent à une loi binomiale, dont vous donnerez les paramètres.
2. Au sein d'une famille de 5 personnes, on constate que 2 personnes ont une croyance religieuse (les 3 autres personnes sont donc athées). On choisit au hasard 2 personnes au sein de cette famille, et on note X le nombre de personnes athées parmi ces deux personnes choisies au hasard. Calculez la loi de X.
3. Au sein d'une ville de 2000 personnes, on constate que 40% (c'est-à-dire 800 personnes) ont une croyance religieuse. On choisit au hasard 2 personnes au sein de cette ville, et on note X le nombre de personnes athées parmi ces deux personnes choisies au hasard. Calculez la loi de X.
4. Comparez entre elles les lois obtenues dans ces trois situations.

Exercice 36 : Ressources humaines

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

On choisit, au sein de l'entreprise, 21 employés au hasard, et on suppose que les effectifs de l'entreprise sont assez grands pour que cet échantillon puisse être considéré comme un tirage "avec remise".

1. Calculez la probabilité que l'échantillon contienne exactement 4 employés notés B.
2. Quelle est la probabilité que l'échantillon contienne au moins 4 individus notés B ?
3. Quelle est la probabilité que l'échantillon contienne au moins 19 individus notés B ou C ?
4. Calculez la probabilité que l'échantillon contienne moins que 5 personnes notées A.
5. Quel est le nombre moyen d'individus notés C au sein d'un échantillon aléatoire de 21 employés ?

Exercice 37 : Analyses de sang

Sur recommandation de son médecin, Mr Georges effectue un bilan sanguin : il s'agit de doser les concentrations sanguines de 40 indicateurs (parmi ces 40 indicateurs, certains sont des hormones, d'autres des vitamines, d'autres encore sont des composants comme les globules blancs ou les plaquettes, etc).

Quand Mr Georges reçoit les résultats, il constate pour chaucun des 40 indicateurs, la concentration mesurée est indiquée, et comparée à une fourchette de valeurs. Pour chaque indicateur, la fourchette est calculée pour que 95% de la population soit dans la fourchette.

Il est inquiet de voir qu'il y a deux indicateurs pour lesquels il est en dehors de la fourchette, et se précipite chez son médecin pour savoir de quelle grave maladie il est atteint.

À sa plus grande surprise, le médecin lui répond « *C'est tout à fait normal d'être en dehors de la fourchette. À vue de nez, je pense que presque les trois quarts de mes patients sont en dehors de la fourchette pour au moins une des concentrations. Et il doit y avoir environ la moitié de mes patients qui sont en dehors pour au moins deux concentrations. Donc ne vous inquiétez pas, votre situation est parfaitement normale ; rentrez chez vous, je suis sûr que vous allez très bien* ».

1. Si on suppose que pour chaque dosage, il y a bien 95% de chance d'être « *dans la fourchette* », et que les dosages sont indépendents, alors quelle est la loi du nombre de dosages pour lesquelles un·e patient·e est « *en dehors de la fourchette* » ?
2. En moyenne, pour combien de dosages chaque patient·e est-il/elle *en dehors de la fourchette* ?
3. Calculer la probabilité d'avoir au moins un dosage en dehors de la fourchette, ainsi que la probabilité d'en avoir au moins deux. Cela correspond-il au ressenti dont fait preuve le médecin ?
4. Trouver les valeurs de k telles que $\mathbb{P}[X \geqslant k] \leqslant 0{,}05$.

On se limitera à k entier.

Remarque : cet exemple sensibilise à un problème méthodologique actuel (appelé *problème de comparaison multiple*). Plus on cherche à tirer de multiples conclusions à partir de données (en faisant de multiples *tests statistiques*, que vous verrez l'an prochain), et plus il y a de chances (si l'on ne prends pas les précautions nécessaires) de croire à tord que quelque-chose d'inattendu a lieu (et de conclure à tord qu'on a trouvé un nouveau resultat scientifique).

Exercice 38 : Images mentales

Dans une expérience sur les images mentales, on demande à un échantillon de 230 enfants d'apprendre une liste de 40 mots en leur montrant, pour chaque mot, un dessin illustrant ce mot afin de faciliter leur apprentissage. On désigne par X le nombre de mots dont se souvient chaque enfant le jour suivant, et on regroupe par classes les données :

Nb de mots X	[5 ; 10[	[10 ; 15[	[15 ; 20[	[20 ; 25[	[25 ; 30[	[30 ; 35[	[35 ; 40[
Effectif	1	11	12	145	54	6	1

1. Calculez 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 de CM2, une moyenne $\mu \simeq 9{,}58$ et un écart-type $\sigma \simeq 3{,}11$.

Dans la suite, on effectue des calculs en supposant que X suit la loi normale $\mathcal{N}(9{,}58; 3{,}11)$.

1. Calculez la proportion théorique des scores supérieurs à 9,97.
2. Trouvez une valeur approchée du nombre a tel que $\mathbb{P}[X < a] = 0{,}95$.
3. Déterminez les quartiles de la variable X.
4. Quel est le score minimal des 20 % de ceux qui ont les scores D.S. les plus élevés ?
5. Quel est le score maximal des 35 % de ceux qui ont les scores D.S. les moins élevés ?
6. Déterminer quatre intervalles différents qui ont tous une probabilité 0,6.

Exercice 40 : Taux d'échec à un concours

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.

1. On prend $n = 18$. Calculez $p_k = \mathbb{P}[S_{18} \leqslant k]$ pour $k = 0$, 1, 2 et 3.
2. Pour $n = 150$, on souhaite calculer $\mathbb{P}[30 \leqslant S_{150} \leqslant 47]$:
   a) Calculez $\mathbb{P}[30 \leqslant S_{150} \leqslant 47]$ avec la calculatrice.
   b) Justifiez qu'on peut approcher la loi de S_{150} par une loi normale. Calculez $\mathbb{P}[30 \leqslant S_{150} \leqslant 47]$ de manière approchée. (On fera le calcul avec et sans correction de continuité, pour juger de l'intérêt de cette correction.)
3. On considère désormais un échantillon de taille $n = 11\,000$.
   (a) Par quelle loi normale peut-on approcher la variable S_{11000} ? Calculez ainsi (une valeur approchée de) la probabilité $\mathbb{P}[2\,669 \leqslant S_{11000} \leqslant 2\,831]$.
   (b) Par quelle loi normale peut-on approcher la proportion $P_{11000} = \frac{S_{11000}}{11\,000}$? Calculez ainsi (une valeur approchée de) la probabilité $\mathbb{P}[0{,}242\,6 \leqslant P_{11000} \leqslant 0{,}257\,4]$.
   (c) Trouvez une valeur approchée de v telle que $\mathbb{P}[S_{11000} < v] = 0{,}03$.
   (d) Determiner un intervalle $[a;\ b]$ tel que $\mathbb{P}[S_{11000} < a] \simeq 0{,}05$ et $\mathbb{P}[a \leqslant S_{11000} \leqslant b] \simeq 0{,}8$.

Exercice 41 : Loi binomiale

corrigé en ligne

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

1) Calculez n et p.

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

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

2) Calculez les probabilités $\mathbb{P}[8 \leqslant X \leqslant 11]$ et $\mathbb{P}[X > 13]$.

- $$\begin{aligned} \mathbb{P}[8 \leqslant X \leqslant 11] &= \mathbb{P}[X=8] + \mathbb{P}[X=9] + \mathbb{P}[X=10] + \mathbb{P}[X=11] \\ &= \binom{16}{8} \times 0{,}25^8\,(1-0{,}25)^8 + \cdots + \binom{16}{11} \times 0{,}25^{11}\,(1-0{,}25)^5 \\ &\simeq 0{,}019\,66 + 0{,}005\,83 + 0{,}001\,36 + 0{,}000\,25 \\ &= 0{,}027\,1 \end{aligned}$$

Remarque : Avec la calculatrice, on obtient directement $\mathbb{P}[8 \leqslant X \leqslant 11] = \mathbb{P}[X \leqslant 11] - \mathbb{P}[X \leqslant 7] \simeq 0{,}999\,96 - 0{,}972\,87 = 0{,}027\,09$

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

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

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

On calcule successivement :
$\mathbb{P}[X \geqslant 16] \simeq 2{,}328 \times 10^{-10}$
$\mathbb{P}[X \geqslant 15] \simeq 1{,}141 \times 10^{-08}$
$\mathbb{P}[X \geqslant 14] \simeq 2{,}629 \times 10^{-07}$
$\mathbb{P}[X \geqslant 13] \simeq 3{,}783 \times 10^{-06}$
$\mathbb{P}[X \geqslant 12] \simeq 3{,}811 \times 10^{-05}$
$\mathbb{P}[X \geqslant 11] \simeq 0{,}0002852$
$\mathbb{P}[X \geqslant 10] \simeq 0{,}001644$
$\mathbb{P}[X \geqslant 9] \simeq 0{,}00747$
$\mathbb{P}[X \geqslant 8] \simeq 0{,}02713$
$\mathbb{P}[X \geqslant 7] \simeq 0{,}07956$
Donc les valeurs de k qui conviennent sont lorsque $k > 7$.
Remarque : Un tel calcul sera utile en deuxième année pour savoir dans quels cas on peut (ou pas) conclure comme dans l'exercice 44.

Exercice 42 : Groupe d'ami·e·s.

corrigé en ligne

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

1. On choisit au hasard le nom d'une de ces trois personnes, puis à nouveau le nom d'une de ces trois personnes au hasard (ce peut être la même personne – ou pas).
   (a) Listez les neuf possibilités pour ces deux noms choisis au hasard.
   En notant par exemple “AB” pour “Alice puis Bernard”, les neuf possibilités sont :
   AA, AB, AC, BA, BB, BC, CA, CB et CC.
   (b) Parmi ces possibilités, combien comptent uniquement des personne en situation d'emploi ?
   Il y en a 4 : AA, AB, BA et BB.
   (c) Quelle est la proportion, parmi ces possibilités, qui comptent une personne sans emploi, et une qui exerce un emploi ?
   Cela correspond aux 4 possibilités AC, BC, CA et CB.
   Donc à la proportion $\frac{4}{9} \simeq 0{,}444\,4$.
   (d) Pour un tel choix aléatoire de deux noms, on note X le nombre de personnes choisies exerçant un emploi. Quelle est la loi de variable X ?On réalise deux fois de suite l'expérience “choisir un des trois noms et constater si c'est une personne qui exerce un emploi”, de manière indépendante. À chaque fois il y a deux chances sur trois de choisir quelqu'un qui exercice un emploi, donc X suit la loi $\mathcal{B}\left(2; \frac{2}{3}\right)$.

(e) Retrouvez le résultat de la question c), en utilisant cette loi.
$\mathbb{P}[X=1]=\binom{2}{1}(2/3)^1(1/3)^1\simeq 0{,}444.$

(f) Déterminez de même $\mathbb{P}[X=0]$ et $\mathbb{P}[X=2]$.
- $\mathbb{P}[X=0]=\binom{2}{0}(2/3)^0(1/3)^2\simeq 0{,}111.$
- $\mathbb{P}[X=2]=\binom{2}{2}(2/3)^2(1/3)^0\simeq 0{,}444.$

2. On convient d'une autre manière de choisir deux noms : on choisit un premier nom au hasard, puis on choisit le second nom parmi les deux restant. Reprenez, avec cette nouvelle façon de choisir les deux noms, les questions 1a, 1b et 1c.
   (a) On a cette fois-ci 6 possibilités : AB, AC, BA, BC, CA et CB.
   (b) Il y en a 2 : AB et BA.
   (c) Cela correspond aux 4 possibilités AC, BC, CA et CB.
   Donc à la proportion $\frac{4}{6}\simeq 0{,}666\,7$.

Exercice 43 : Collection de dessins

corrigé en ligne

Mme Laine est pédiatre et utilise fréquemment de petites illustrations pour stimuler les enfants. Elle a acheté à cet effet 5 cartes illustrées qu'elle stocke dans un tiroir.

1. Elle mélange les cartes, et constate qu'elle sont dans un certain ordre à l'issue du mélange. Quelle était la probabilité qu'à l'issue du mélange, elles se retrouvent précisément dans cet ordre là ?
   Il y a $5!=5\times4\times3\times2=120$ résultats possibles à l'issue du mélange. La probabilité est donc $\frac{1}{120}\simeq 0{,}008\,333$.
2. Si elle remélange les cartes. Calculez les probabilités des évènements suivants :
   (a) la carte qui était en haut du tas soit à nouveau en haut du tas ?
   Il y a $4!=4\times3\times2=24$ 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 cartes qui étaient en haut du tas soient à nouveau en haut du tas, et dans le même ordre ?
   Il y a $3!=3\times2\times1=6$ mélanges qui conservent les deux premières carte (car cela correspond à ne mélanger que les 3 autres cartes). La probabilité est donc $\frac{6}{120}=0{,}05$.
   (c) les deux cartes 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 a 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 cartes qui étaient en haut du tas soient à nouveau en haut du tas, dans n'importe quel ordre ?
   Il suffit d'additionner les probabilités des deux dernières questions. On obtient $\frac{6}{120}+\frac{6}{120}=0{,}1$.
   (e) l'ensemble des cartes soient exactement dans le même ordre que lors du premier mélange ?
   $\frac{1}{120}\simeq 0{,}008\,333$.
3. Répondez au mêmes questions en supposant qu'il y ait cette fois ci 35 cartes.
   1. $\frac{1}{35!}=\frac{1}{35\times34\times\ldots\times1}=9{,}677\,592\,958\,631\,89\times10^{-41}$
   2. (a) $\frac{34!}{35!}=\frac{34\times33\times\ldots\times1}{35\times34\times\ldots\times1}=\frac{1}{35}\simeq 0{,}028\,57$
      (b) $\frac{33!}{35!}=\frac{33\times32\times\ldots\times1}{35\times34\times\ldots\times1}=\frac{1}{35\times34}\simeq 0{,}000\,840\,34$
      (c) $\frac{1}{35\times34}\simeq 0{,}000\,840\,34$
      (d) $\frac{1}{35\times34}+\frac{1}{35\times34}\simeq 0{,}001\,680\,7$
      (e) $\frac{1}{35!}=\frac{1}{35\times34\times\ldots\times1}=9{,}677\,592\,958\,631\,89\times10^{-41}$
4. Peut-on conclure des probabilités calculées que certains évènement 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 35 cartes).

Dans la question 1 on ne peut rien en déduire : chaque résultat à l'issue du mélange a une très faible probabilité, et iul est complétement artificiel de choisir de calculer la probilité du tirage qui a eu lieu. Quoi qu'il arrive le résultat sur lequel on tombe sera forcément un parmi des milliards (de milliards), donc un résultat qui avait extrêmement peu de chances de tomber, mais il faut bien tomber sur l'un de ces résultats.

En revanche, dans la question 2e, on peut vraiment déduire quelque-chose : quand on remélange un paquet de cartes il est très improbable (pour 5 cartes) et même quasiment impossible (pour 35 cartes) de reproduire le même ordre qu'initialement.

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

Exercice 44 : Thérapie contre la dépression

corrigé en ligne

Un psychologue a développé une nouvelle thérapie contre la dépression, et il affirme que cette thérapie permet la rémission de 60% des patients. En interrogeant un échantillon aléatoire de 28 patients, on constate 9 rémissions.

1) S'il y avait 60% de rémission parmi l'ensemble des patients, quelle loi suivrait le nombre de rémissions au sein d'un échantillon de 28 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 28 patients, et si l'on suppose qu'il y ait 60% de rémission parmi l'ensemble des patients, alors X suit la loi $\mathcal{B}(28; 0{,}6)$.

   Le nombre moyen de rémissions au sein d'un tel échantillon serait alors $m(X) = 28 \times 0{,}6 = 16{,}8$.

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

   L'énoncé ne l'indique pas, mais on peut penser que déjà rien qu'en France, il y a plusieurs millions de personnes qui souffrent de dépression. En comparaison l'échantillon de 28 personnes est beaucoup plus petit (plus de dix fois plus petit), donc d'après le cours, les probabilités sont celles d'un tirage avec remise.

3) Sous cette même hypothèse, quelle serait la probabilité d'avoir au maximum 9 rémissions ?

   On aurait alors
   $\mathbb{P}[X \leqslant 9] \simeq 0{,}002\,67$

4) Conclure : vous semble t-il vraisemblable qu'il y ait, comme l'affirme ce psychologue, 60% de rémissions parmi l'ensemble des patients ?

   On a obtenu que s'il y avait bien 60% de rémission parmi l'ensemble des patients (comme l'affirme ce psychologue), alors il y aurait très peu de chances (environ 0,3% de chances) d'avoir un échantillon où il n'y ait pas plus de 9 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 60% (à moins par exemple que notre échantillon soit biaisé à cause de la façon dont il a été établi).

Exercice 45 : Nombre de filles et de garçons

corrigé en ligne

Dans une classe, il y a 11 filles et 15 garçons. On choisit au hasard 5 élèves distincts dans cette classe.

1. Quel est le nombre de choix possibles ?
   On choisit 5 personnes parmi l'ensemble des $11 + 15 = 26$ élèves. Le nombre de choix possibles est donc
$$\begin{aligned}\binom{26}{5} &= \frac{26!}{5!\times 21!}\\ &= \frac{26\times 25\times\ldots\times 1}{(5\times 4\times 3\times 2)\times(21\times 20\times\ldots\times 1)}\\ &= (26\times 25\times 24\times 23\times 22)\times\frac{1}{5\times 4\times 3\times 2}\\ &= \frac{26\times 5\times 5\times 24\times 23\times 22}{5\times 4\times 3\times 2}\\ &= \frac{26\times 5\times 24\times 23\times 22}{4\times 3\times 2}\\ &= \frac{26\times 5\times 4\times 6\times 23\times 22}{4\times 3\times 2}\\ &= \frac{26\times 5\times 6\times 23\times 22}{3\times 2}\\ &= \frac{26\times 5\times 3\times 2\times 23\times 22}{3\times 2}\\ &= 26\times 5\times 23\times 22\\ &= 65780\end{aligned}$$
   .

2. Quel est le nombre de choix ne comportant que des garçons ?
   Ces choix correspondent à choisir 5 élèves parmi les 15 garçons. Le nombre de choix correspondant est donc
$$\begin{aligned}\binom{15}{5} &= \frac{15!}{5!\times 10!}\\ &= \frac{15\times 14\times\ldots\times 1}{(5\times 4\times 3\times 2)\times(10\times 9\times\ldots\times 1)}\\ &= (15\times 14\times 13\times 12\times 11)\times\frac{1}{5\times 4\times 3\times 2}\\ &= \frac{5\times 3\times 14\times 13\times 12\times 11}{5\times 4\times 3\times 2}\\ &= \frac{14\times 13\times 12\times 11}{4\times 2}\\ &= \frac{14\times 13\times 4\times 3\times 11}{4\times 2}\\ &= \frac{14\times 13\times 3\times 11}{2}\\ &= \frac{2\times 7\times 13\times 3\times 11}{2}\\ &= 7\times 13\times 3\times 11\\ &= 3003\end{aligned}$$
   .

3. Quelle est la proportion de choix ne comportant que des filles ?
   Ces choix correspondent à choisir 5 élèves parmi les 11 filles. Le nombre de choix correspondant est donc
$$\begin{aligned}\binom{11}{5} &= \frac{11!}{5!\times 6!}\\ &= \frac{11\times 10\times\ldots\times 1}{(5\times 4\times 3\times 2)\times(6\times 5\times\ldots\times 1)}\\ &= (11\times 10\times 9\times 8\times 7)\times\frac{1}{5\times 4\times 3\times 2}\\ &= \frac{11\times 5\times 2\times 9\times 8\times 7}{5\times 4\times 3\times 2}\\ &= \frac{11\times 9\times 8\times 7}{4\times 3}\\ &= \frac{11\times 9\times 4\times 2\times 7}{4\times 3}\\ &= \frac{11\times 9\times 2\times 7}{3}\\ &= \frac{11\times 3\times 3\times 2\times 7}{3}\\ &= 11\times 3\times 2\times 7\\ &= 462\end{aligned}$$
   Cela correspond donc à la probabilité $\frac{462}{65\,780} \simeq 0{,}007\,023$.

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

   Les cas correspondants sont formés d'une part d'un groupe de 4 filles (pour lequel il y a $\binom{11}{4}$ choix possibles), et pour chaque choix de ce groupe de filles, on peut y ajouter n'importe lequel des garçons (pour lequel il y a $\binom{15}{1}$ choix possibles). En conséquence, le nombre de cas qui correspondent est
$$\begin{aligned}\binom{11}{4}\times\binom{15}{1} &= \frac{11!}{4!\times 7!}\times\frac{15!}{1!\times 14!}\\ &= \frac{11\times 10\times\ldots\times 1}{(4\times 3\times 2)\times(7\times 6\times\ldots\times 1)}\times\frac{15\times 14\times\ldots\times 1}{14\times 13\times\ldots\times 1}\\ &= (11\times 10\times 9\times 8)\times\frac{1}{4\times 3\times 2}\times 15\\ &= \frac{11\times 10\times 9\times 4\times 2}{4\times 3\times 2}\times 15\\ &= \frac{11\times 10\times 9}{3}\times 15\\ &= \frac{11\times 10\times 3\times 3}{3}\times 15\\ &= 11\times 10\times 3\times 15\\ &= 330\times 15\\ &= 4950\end{aligned}$$
   D'où la probabilité $\frac{4\,950}{65\,780} \simeq 0{,}075\,25$.

5. Quelle est donc la probabilité d'avoir choisi au plus un garçon ?
   On additionne, d'une part le nombre de choix comportant 0 garçons (calculé à la question 3) et d'autre part le nombre de choix

comportant exactement 1 garçon (calculé à la question 4) : on obtient $462 + 4\,950 = 5\,412$. La probabilité d'avoir choisi au plus un garçon est donc $\frac{5\,412}{65\,780} \simeq 0{,}082\,27$.

Exercice 46 : Indice de masse corporelle

corrigé en ligne

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

- **IMC$<$ 16,5** dénutrition ou famine
- **16,5 $\leqslant$ IMC$<$ 18,5** maigreur
- **18,5 $\leqslant$ IMC$<$ 25** corpulence normale
- **25 $\leqslant$ IMC$<$ 30** surpoids
- **30 $\leqslant$ IMC$<$ 35** obésité modérée
- **35 $\leqslant$ IMC$<$ 40** obésité sévère
- **IMC $\geqslant$40** obésité morbide

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

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

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

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

1. Déterminez la probabilité (« *fréquence théorique* ») qu'aurait chaque classe sous l'hypothèse de loi normale $(\mathcal{H})$.
   Si X suivait la loi $\mathcal{N}(25; 5)$, on obtiendrait les probabilités suivantes :
   - $\mathbb{P}\,[X < 18{,}5] \simeq 0{,}096\,8$
   - $\mathbb{P}\,[18{,}5 \leqslant X < 25] \simeq 0{,}403\,2$
   - $\mathbb{P}\,[25 \leqslant X < 30] \simeq 0{,}341\,3$
   - $\mathbb{P}\,[30 \leqslant X < 35] \simeq 0{,}135\,9$
   - $\mathbb{P}\,[35 \leqslant X < 40] \simeq 0{,}021\,4$
   - $\mathbb{P}\,[X \geqslant 40] \simeq 0{,}001\,35$

   D'où les fréquences théoriques suivantes :

IMC	$[-\infty\,;18{,}5[$	$[18{,}5\,;25[$	$[25\,;30[$	$[30\,;35[$	$[35\,;40[$	$[40\,;+\infty[$
fréquence théorique	0,096 8	0,403 2	0,341 3	0,135 9	0,021 4	0,001 35

2. Justifiez que, sous l'hypothèse $(\mathcal{H})$, la proportion *théorique* d'obésité (c'est à dire $\mathbb{P}[X \geqslant 30]$, en regroupant les obésités modérées, sévères et morbides) serait d'environ 15,9%.
   La proportion théorique d'obésité serait la somme des fréquences théoriques des trois dernières classes, c'est à dire $\mathbb{P}[X \geqslant 30] \simeq 0{,}135\,9 + 0{,}021\,4 + 0{,}001\,35 \simeq 0{,}159$.
3. Quelle est, en fait, la proportion des sujets obèses observée dans l'échantillon épidémiologique de l'INSERM ?
   La proportion d'obésité observée dans l'échantillon de l'INSERM est la somme des fréquences des trois dernières classes, c'est à dire $10{,}7\% + 3{,}1\% + 1{,}2\% \simeq 15\%$.
4. On cherche désormais à comparer les résultats des questions 2 et 3 : pour cela, on va évaluer si la proportion $\mathbb{P}_r[X \geqslant 30]$ observée dans l'échantillon de l'INSERM s'accorde avec la probabilité $\mathbb{P}[X \geqslant 30]$ obtenue sous l'hypothèse $(\mathcal{H})$.
   (a) Si on suppose qu'un·e français·e choisi·e au hasard a une probabilité de 15,9% d'être en situation d'obésité, quelle est la loi du nombre de personnes en situation d'obésité dans un échantillon de 25 714 français ? Le fait que l'échantillon soit *avec* ou *sans remise* a-t-il la moindre importance ?
   On choisit un échantillon aléatoire de taille 25 714 français parmi une population constituée de tous les français adultes (environ 50 000 000). La taille de cette population est plus de 10 fois supérieure à cette de l'échantillon ($50\,000\,000 > 10 \times 25\,714$) donc le fait d'être “avec” ou “sans remise” n'a aucun impact. On peut donc utiliser la loi binomiale $\mathcal{B}(25714; 0{,}159)$.

(b) Cette loi peut-elle être approximée par une loi normale, et si oui laquelle ?
$n\,p(1-p) = 25\,714 \times 0{,}159\,(1-0{,}159) \simeq 3\,438{,}45 > 1\,000$, donc on peut approcher
$$\mathcal{B}(25714\ ;\ 0{,}159) \approx \mathcal{N}(25\,714 \times 0{,}159\ ;\ \sqrt{25\,714 \times 0{,}159\,(1-0{,}159)}) \simeq \mathcal{N}(4088{,}526\ ;\ 58{,}638)$$
, et il n'y a pas besoin de faire de correction de continuité .

(c) On considère toujours un échantillon aléatoire de 25 714 français·es, mais cette fois-ci on considère comme variable aléatoire la **proportion** de personnes, au sein de l'échantillon, qui sont en situation d'obésité. Sa loi peut-elle aussi être approximée par une loi normale, et si oui laquelle ?
$n\,p(1-p) = 25\,714 \times 0{,}159\,(1-0{,}159) \simeq 3\,438{,}45 > 1\,000$, donc on peut approcher la proportion par $\mathcal{N}(0{,}159\ ;\ \sqrt{\frac{0{,}159\,(1-0{,}159)}{25\,714}}) \simeq \mathcal{N}(0{,}159\ ;\ 0{,}00228)$, et il n'y a pas besoin de faire de correction de continuité .

(d) En utilisant cette approximation par une loi normale, calculer la probabilité qu'un échantillon aléatoire de 25 714 français·es contienne au maximum 15% de personnes en situation d'obésité.
Comme cette proportion P suit la loi normale trouvée dans la question (c), on doit avoir $\mathbb{P}[P \leqslant 0{,}15] \simeq 3{,}951 \times 10^{-05}$

(e) En utilisant le résultat de la question (d) et en faisant appel à l'étude épidémiologique de l'INSERM, que pouvez-vous en conclure au sujet de l'hypothèse $(\mathcal{H})$?
Si X avait vraiment suivi la loi $\mathcal{N}(25;5)$, alors il n'y aurait eu quasiment aucune chance qu'un échantillon aléatoire de 25714 français ait au plus 15% de personnes en situation d'obésité. Or l'échantillon aléatoire de l'étude épidémiologique de l'INSERM a donné 15% de personnes en situation d'obésité. Donc, en toute vraisemblance, l'IMC des français ne suit pas la loi $\mathcal{N}(25;5)$.

Exercice 47 : Loi normale générique

corrigé en ligne

1. Déterminez les probabilités des évènements suivants pour une variable aléatoire Z suivant la loi normale centrée réduite :
   (a) Z est supérieure ou égale à 1,3 ;
   $\mathbb{P}[Z \geqslant 1{,}3] \simeq 0{,}096\,8$
   (b) Z est comprise entre -1 et 1 ;
   $\mathbb{P}[-1 \leqslant Z \leqslant 1] \simeq 0{,}682\,7$
   (c) Z est comprise entre -2 et 2 ;
   $\mathbb{P}[-2 \leqslant Z \leqslant 2] \simeq 0{,}954\,5$
   (d) Z est comprise entre -3 et 3.
   $\mathbb{P}[-3 \leqslant Z \leqslant 3] \simeq 0{,}997\,3$
2. Déterminez les probabilités des évènements suivants pour une variable aléatoire X suivant une loi normale de moyenne $\mu = 92$ et d'écart-type $\sigma = 9$ (c'est-à-dire $X \sim \mathcal{N}(92;9)$) :
   (a) X est supérieure ou égale à $\mu + 1{,}3\,\sigma$;
   On note tout d'abord que $\mu + 1{,}3\,\sigma = 92 + 1{,}3 \times 9 = 103{,}7$.
   $\mathbb{P}[X \geqslant 103{,}7] \simeq 0{,}096\,8$
   (b) X est comprise entre $\mu - \sigma$ et $\mu + \sigma$;
   On note tout d'abord que $\mu - \sigma = 92 - 9 = 83$ et $\mu + \sigma = 92 + 9 = 101$.
   $\mathbb{P}[83 \leqslant X \leqslant 101] \simeq 0{,}682\,7$
   (c) X est comprise entre $\mu - 2\,\sigma$ et $\mu + 2\,\sigma$;
   On note tout d'abord que $\mu - 2\,\sigma = 92 - 2 \times 9 = 74$ et $\mu + 2\,\sigma = 92 + 2 \times 9 = 110$.
   $\mathbb{P}[74 \leqslant X \leqslant 110] \simeq 0{,}954\,5$
   (d) X est comprise entre $\mu - 3\,\sigma$ et $\mu + 3\,\sigma$.
   On note tout d'abord que $\mu - 3\,\sigma = 92 - 3 \times 9 = 65$ et $\mu + 3\,\sigma = 92 + 3 \times 9 = 119$.
   $\mathbb{P}[65 \leqslant X \leqslant 119] \simeq 0{,}997\,3$

3. Reprenez la question 2 pour une variable aléatoire X suivant la loi normale $\mathcal{N}(18{,}06; 3{,}93)$. Qu'observez-vous ?
   (a) On note tout d'abord que $\mu + 1{,}3\,\sigma = 18{,}06 + 1{,}3 \times 3{,}93 = 23{,}169$.
   $\mathbb{P}\,[X \geqslant 23{,}169] \simeq 0{,}096\,8$
   (b) On note tout d'abord que $\mu - \sigma = 18{,}06 - 3{,}93 = 14{,}13$ et $\mu + \sigma = 18{,}06 + 3{,}93 = 21{,}99$.
   $\mathbb{P}\,[14{,}13 \leqslant X \leqslant 21{,}99] \simeq 0{,}682\,7$
   (c) On note tout d'abord que $\mu - 2\,\sigma = 18{,}06 - 2 \times 3{,}93 = 10{,}2$ et $\mu + 2\,\sigma = 18{,}06 + 2 \times 3{,}93 = 25{,}92$.
   $\mathbb{P}\,[10{,}2 \leqslant X \leqslant 25{,}92] \simeq 0{,}954\,5$
   (d) On note tout d'abord que $\mu - 3\,\sigma = 18{,}06 - 3 \times 3{,}93 = 6{,}27$ et $\mu + 3\,\sigma = 18{,}06 + 3 \times 3{,}93 = 29{,}85$.
   $\mathbb{P}\,[6{,}27 \leqslant X \leqslant 29{,}85] \simeq 0{,}997\,3$

   On observe au travers de nos réponses aux questions 1, 2 et 3 que les résultats restent les mêmes quand on change μ et σ (éventuellement à quelques imprécisions près liées au fait qu'on a arrondi des valeurs dans les calculs intermédiares).

   Cela découle d'une propriété rapelée dans le formulaire, à savoir : « *Si* $Z \sim \mathcal{N}(0;\ 1)$*, alors* $\mu + Z\,\sigma \sim \mathcal{N}(\mu; \sigma)$. »

   Explication : par exemple la question 1c indique que (si $Z \sim \mathcal{N}(0;\ 1)$), $\mathbb{P}[-2 \leqslant Z \leqslant 2] \simeq 0{,}954\,5$. Mais la condition « $-2 \leqslant Z \leqslant 2$ » revient au même que $\mu - 2\sigma \leqslant \mu + Z\,\sigma \leqslant \mu + 2\sigma$ où $\mu + Z\,\sigma$ est la variable X qui suit la loi $\mathcal{N}(\mu;\ \sigma)$. Il en découle que $\mathbb{P}[\mu - 2\sigma \leqslant X \leqslant \mu + 2\sigma] = \mathbb{P}[-2 \leqslant Z \leqslant 2] \simeq 0{,}954\,5$, d'où le resultat des c) des questions 2 et 3.

   Pour les autres questions (par exemple $\mathbb{P}[\mu - 3\,\sigma \leqslant X \leqslant \mu + 3\sigma]$) le même raisonnement explique aussi que l'on ait eu trois fois la même réponse (en questions 1, 2 et 3). .

4. Déterminer un intervalle $[u;\ v]$ tel que $\mathbb{P}[X \leqslant u] \simeq 0{,}1$ et que $\mathbb{P}[u \leqslant X \leqslant v] \simeq 0{,}85$. Vous exprimer la réponse en fonction de μ et σ.

   Indication : vous pourrez commencer par traiter le cas d'une loi normale centrée réduite.

   Dans le cas où $Z \sim \mathcal{N}(0; 1)$, $\mathbb{P}[X \leqslant u] \simeq 0{,}1$ indique (d'après la calculette) que $u \simeq -1{,}282$. De plus $\mathbb{P}[X \leqslant v] \simeq 0{,}1 + 0{,}85 = 0{,}95$ donne $v \simeq 1{,}645$. Donc on obtient l'intervale $[-1{,}282; 1{,}645]$.

   Si $X \sim \mathcal{N}(\mu; \nu)$, on peut l'écrire comme $X = \mu + Z\sigma$ et la condition « $-1{,}282 \leqslant Z \leqslant 1{,}645$ » se réécrit comme

   $$\text{« } \mu - 1{,}282\sigma \leqslant \underbrace{\mu + Z\sigma}_{X} \leqslant \mu + 1{,}645\sigma \text{ ».}$$

   Donc l'intervalle est $[\mu - 1{,}282\sigma; \mu + 1{,}645\sigma]$.

Exercice 48 : Diagnostic de la dépression

corrigé en ligne

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

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

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

$\mathbb{P}\,[X \geqslant 70] \simeq 0{,}022\,75$

Donc parmi un échantillon de $10\,000$ personnes, on s'attend à trouver environ $10\,000 \times 0{,}022\,75 = 227{,}5$ personnes pathologiquement dépressives.

Exercice 49 : Épreuve d'organisation

corrigé en ligne

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

1. Lorsque $n = 26$, précisez la loi de probabilité de X, puis calculez $\mathbb{P}[15 \leqslant X \leqslant 19]$.
   On a $X \sim \mathcal{B}(26; 0{,}65)$.
   Donc on trouve $\mathbb{P}[15 \leqslant X \leqslant 19] = \mathbb{P}[X \leqslant 19] - \mathbb{P}[X \leqslant 14] \simeq 0{,}858\,4 - 0{,}161\,6 = 0{,}696\,8$ en utilisant la calculatrice.
2. On considère désormais que $n = 260$.
   (a) Justifiez qu'on peut approcher la loi de X par une loi normale, préciser laquelle, et indiquer s'il est nécessaire de faire une correction de continuité.
   $X \sim \mathcal{B}(260; 0{,}65)$.
   $n\,p(1-p) = 260 \times 0{,}65\,(1-0{,}65) = 59{,}15 > 10$, donc on peut approcher $\mathcal{B}(260\ ;\ 0{,}65) \approx \mathcal{N}(260 \times 0{,}65\ ;\ \sqrt{260 \times 0{,}65\,(1-0{,}65)})$, $\simeq \mathcal{N}(169\ ;\ 7{,}691)$
   mais il est nécessaire de faire une correction de continuité.
   (b) Déduisez de la question précédente des valeurs approchées de $\mathbb{P}[X \geqslant 156]$ et $\mathbb{P}[150 \leqslant X \leqslant 190]$.
   Avec cette approximation par une loi normale et en faisant la correction de continuité, on obtient :
   $\mathbb{P}[X \geqslant 156] = \mathbb{P}[X \geqslant 155{,}5] \simeq 0{,}960\,4$
   $\mathbb{P}[150 \leqslant X \leqslant 190] = \mathbb{P}[149{,}5 \leqslant X \leqslant 190{,}5] \simeq 0{,}991\,79$

Exercice 50 : Pile ou face corrigé en ligne

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

1. 100 fois.
   Si on note X le nombre de pile, on a $X \sim \mathcal{B}(100; 0{,}5)$.
   On observe que $n\,p(1-p) = 100 \times 0{,}5\,(1-0{,}5) = 25 > 10$, donc on peut approcher $\mathcal{B}(100\ ;\ 0{,}5) \approx \mathcal{N}(100 \times 0{,}5\ ;\ \sqrt{100 \times 0{,}5\,(1-0{,}5)})$, $= \mathcal{N}(50\ ;\ 5)$ mais
   il est nécessaire de faire une correction de continuité.
   Avoir une proportion de pile entre 0,48 et 0,52 signifierait avoir $48 \leqslant X \leqslant 52$. La probabilité demandée vaut donc (avec la correction de continuité) :
   $\mathbb{P}[48 \leqslant X \leqslant 52] = \mathbb{P}[47{,}5 \leqslant X \leqslant 52{,}5] \simeq 0{,}383$
2. 1 000 fois.
   on a $X \sim \mathcal{B}(1000; 0{,}5)$.
   On observe que $n\,p(1-p) = 1\,000 \times 0{,}5\,(1-0{,}5) = 250 > 10$, donc on peut approcher $\mathcal{B}(1000\ ;\ 0{,}5) \approx \mathcal{N}(1\,000 \times 0{,}5\ ;\ \sqrt{1\,000 \times 0{,}5\,(1-0{,}5)})$, $\simeq \mathcal{N}(500\ ;\ 15{,}811)$
   mais il est nécessaire de faire une correction de continuité.
   Avoir une proportion de pile entre 0,48 et 0,52 signifierait avoir $480 \leqslant X \leqslant 520$. La probabilité demandée vaut donc (avec la correction de continuité) :
   $\mathbb{P}[480 \leqslant X \leqslant 520] = \mathbb{P}[479{,}5 \leqslant X \leqslant 520{,}5] \simeq 0{,}805$
3. 2 000 fois.
   on a $X \sim \mathcal{B}(2000; 0{,}5)$.
   On observe que $n\,p(1-p) = 2\,000 \times 0{,}5\,(1-0{,}5) = 500 > 10$, donc on peut approcher $\mathcal{B}(2000\ ;\ 0{,}5) \approx \mathcal{N}(2\,000 \times 0{,}5\ ;\ \sqrt{2\,000 \times 0{,}5\,(1-0{,}5)})$, $\simeq \mathcal{N}(1000\ ;\ 22{,}361)$
   mais il est nécessaire de faire une correction de continuité.
   Avoir une proportion de pile entre 0,48 et 0,52 signifierait avoir $960 \leqslant X \leqslant 1\,040$. La probabilité demandée vaut donc (avec la correction de continuité) :
   $\mathbb{P}[960 \leqslant X \leqslant 1\,040] = \mathbb{P}[959{,}5 \leqslant X \leqslant 1\,040{,}5] \simeq 0{,}929\,9$

4. 10 000 fois.
   On considère désormais la proportion $p = \frac{X}{n}$.
   On observe que $n\,p(1-p) = 10\,000 \times 0{,}5\ (1-0{,}5) = 2\,500 > 1\,000$, donc on peut approcher la proportion par $\mathcal{N}(0{,}5\ ;\ \sqrt{\frac{0{,}5\,(1-0{,}5)}{10\,000}}) = \mathcal{N}(0{,}5\ ;\ 0{,}005)$, et il n'y a pas besoin de faire de correction de continuité . La probabilité demandée vaut donc :
   $\mathbb{P}\,[0{,}48 \leqslant P \leqslant 0{,}52] \simeq 0{,}999\,936\,7$

Exercice 51 : Activité d'un neurone

corrigé en ligne

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 125 et d'écart-type 20.

1. Quelle est la probabilité pour qu'un neurone émette plus de 150 PA en 1 min s'il n'est pas activé ?
   On note X le nombre de PA émis en une minute. Si le neurone n'est pas activé on a $X \sim \mathcal{N}(125; 20)$. Donc on a $\mathbb{P}\,[X > 150] \simeq 0{,}105\,6$ (d'après la calculatrice).
2. Quelle est la probabilité pour qu'un neurone émette moins de 100 PA en 1 min s'il n'est pas activé ?
   On a $\mathbb{P}\,[X < 100] \simeq 0{,}105\,6$ (d'après la calculatrice).
3. Quelle est la probabilité pour qu'un neurone émette entre 110 et 140 PA en 1 min s'il n'est pas activé ?
   On a $\mathbb{P}\,[110 \leqslant X < 140] \simeq 0{,}546\,7$ (d'après la calculatrice).
4. On enregistre l'activité d'un neurone durant une minute, et on observe que ce neurone a émis plus de 230 PA. Selon vous, est-il plus vraisemblable que ce neurone était activé ou au repos lors de l'enregistrement ? Justifiez votre réponse.
   On a $\mathbb{P}\,[X \geqslant 230] \simeq 7{,}605 \times 10^{-08}$ (d'après la calculatrice).
   Si ce neurone avait été au repos, il aurait donc été très improbable d'observer plus de 230 PA durant l'enregistrement. Ainsi, très vraisemblablement, le neurone était actif durant l'enregistrement.

Chapitre 4: Estimation

Exercice 52 : Efficacité d'un traitement

Un fabricant de médicaments affirme qu'en 5 jours, 35% des malades qui sont traités par l'un de ses produits guérissent de leur maladie.

Pour vérifier cette affirmation du fabricant, on administre ce produit à 800 malades. On constate alors que, au sein de cet échantillon de 800 patients traités, il y en a 399 qui guérissent en 5 jours.

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

Exercice 53 : 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 1001 personnes interrogées, 495 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 1158 personnes interrogées, il y en a alors 609 qui sont satisfaites.

1. On désigne par p_1 et p_2 les proportions de satisfaits sur l'ensemble de la clientèle de l'entreprise, avant et après l'amélioration apportée. Donnez une estimation de p_1 et p_2 par intervalle de confiance en prenant comme confiance $c = 0{,}96$. Peut-on dire que la proportion de satisfaits pour l'ensemble de la clientèle a augmenté ?
2. Lors du deuxième sondage, quelle aurait dû être la taille de l'échantillon pour avoir une estimation de p_2 à 2% près avec une confiance de 0,99 ?

Exercice 54 : Résistance à la persuasion

On émet l'hypothèse que la résistance à la persuasion passe par une réaction active des sujets qui développent intérieurement des contre-arguments. Pour tester cette hypothèse on considère l'opinion qu'ont les étudiants sur la coopération entre étudiants et enseignants pour établir les programmes. Cette opinion est mesurée par un questionnaire pour lequel, sur l'ensemble des étudiants, la moyenne est 13,2 et l'écart type est 1,8.

1. On soumet dans un premier temps un échantillon de 60 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	[7 ; 9[	[9 ; 11[	[11 ; 13[	[13 ; 15[	[15 ; 17[
Effectif	2	16	20	16	6

   a) Calculer la moyenne expérimentale $m_e(X)$ et l'écart type expérimental $s_e(X)$ de l'échantillon.
   b) Estimer pour l'ensemble des étudiants la note moyenne après avoir entendu un tel argumentaire (on l'appelle $\mu(X)$). On déterminera un intervalle de confiance, avec la confiance $c = 0{,}95$.
   c) Peut-on dire avec un risque d'erreur de 5% qu'en moyenne, la note (traduisant l'opinion en faveur d'une coopération entre étudiants et enseignants) a diminué après l'argumentaire ? Justifiez votre réponse.
   d) Quelle devrait être la taille minimale de l'échantillon à prélever pour estimer la moyenne $\mu(X)$ à 0,5 points près avec une confiance de 0,99 ?

2. On considère désormais un échantillon de 27 étudiants à qui on demande, pendant qu'ils écoutent l'argumentaire, de réaliser des « opérations » arithmétiques (réciter mentalement la table de multiplication par 8). La moyenne et l'écart type expérimentaux sont alors $m_e = 9{,}09$ et $s_e = 3{,}9$.
   a) Avec une confiance de 0,95 déterminer la note moyenne de l'ensemble des étudiants s'ils écoutaient l'argumentaire en récitant des tables de multiplication. *(On pourra supposer que les notes suivent une loi normale.)*
   b) Comparer cette moyenne à celle qu'ils auraient sans réciter les tables de multiplication (en étant donc en mesure de réfléchir à des contre-arguments). Ces résultats sont-ils compatibles avec l'hypothèse théorique que l'on souhaitait tester ?
   c) Donner une estimation de l'écart type des notes, dans le deuxième cas (en récitant les tables de multiplication) avec un risque d'erreur de 5%.

Exercice 55 : Motivation lors d'un test

Une psychologue qui étudie la prise de décisions a demandé à 29 enfants de résoudre le plus grand nombre possible de problèmes en 30 minutes. Elle a expliqué à un groupe de 15 enfants qu'elle voulait tester leur aptitude innée à résoudre des problèmes, et aux autre enfants (14 enfants) qu'il ne s'agissait que d'une tâche destinée à les occuper. Les résultats obtenus sont les suivants :

Aptitude innée	21	20	25	33	30	26	30	25	34	26	37	32	27	24	25
Occupation	10	8	25	15	16	28	15	4	20	13	17	19	20	12	

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

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

Exercice 56 : Pile ou face

corrigé en ligne

Vous jouez à « pile » ou « face » avec une pièce de monnaie. Vous la lancez 1000 fois et obtenez 548 fois pile.

1. Estimez, avec la confiance 90%, la probabilité qu'a cette pièce de tomber sur pile quand on la lance.
   On a $n\,p_e(1-p_e) = 1\,000\times\frac{548}{1\,000}\left(1-\frac{548}{1\,000}\right) = 247{,}696 > 10$, donc on peut utiliser la procédure du formulaire pour estimer la proportion p.
   Le formulaire indique que pour la confiance 0,9 on a $z_\alpha \simeq 1{,}645$ d'où $a_\alpha = z_\alpha\sqrt{\frac{p_e(1-p_e)}{n}} = 1{,}645\times\sqrt{\frac{\frac{548}{1000}(1-\frac{548}{1000})}{1000}} \simeq 0{,}026$ et $p_e - a_\alpha \simeq \frac{548}{1\,000} - 0{,}026 = 0{,}522$ et $p_e + a_\alpha \simeq \frac{548}{1\,000} + 0{,}026 = 0{,}574$.
   On estime donc que p est dans l'intervalle $[0{,}522; 0{,}574]$ avec la confiance $c = 0{,}9$.
2. Conclure : pouvez-vous affirmer (avec la confiance 90%) que cette pièce soit biaisée ?
   On peut effectivement affirmer, avec la confiance 90%, que la pièce est biaisée (on a vu qu'elle a au moins 52,2% de chances de faire pile).

Exercice 57 : Taux de cholesterol

corrigé en ligne

On étudie les taux de cholestérol dans le sang chez les femmes et chez les hommes de plus de 50 ans. On suppose qu'ils suivent des lois normales.

Dans un échantillon de 16 femmes de plus de 50 ans on a mesuré un taux moyen de 197,1 mg/dL avec un écart type de 33,9 mg/dL.

Dans un échantillon de 21 hommes de plus de 50 ans on a mesuré un taux moyen de 172,9 mg/dL avec un écart type de 57,5 mg/dL.

1. Donner des estimations des écarts types des deux populations. *On demande de déterminer un intervalle de confiance, avec la confiance $c = 98\%$.*

   Chez les femmes On lit dans la table du χ^2 à 15 degrés de liberté la valeur $x_1 = 5{,}229$ correspondant à $q = 0{,}01$ et la valeur $x_2 = 30{,}58$ correspondant à $p = 0{,}01$
   On obtient alors $s_e\sqrt{\frac{n}{x_2}} \simeq 33{,}9\sqrt{\frac{16}{30{,}58}} \simeq 24{,}5$ et $s_e\sqrt{\frac{n}{x_1}} \simeq 33{,}9\sqrt{\frac{16}{5{,}229}} \simeq 59{,}3$
   On estime donc que σ est dans l'intervalle $[24{,}5; 59{,}3]$ avec la confiance $c = 0{,}98$.

   Chez les hommes On lit dans la table du χ^2 à 20 degrés de liberté la valeur $x_1 = 8{,}26$ correspondant à $q = 0{,}01$ et la valeur $x_2 = 37{,}57$ correspondant à $p = 0{,}01$
   On obtient alors $s_e\sqrt{\frac{n}{x_2}} \simeq 57{,}5\sqrt{\frac{21}{37{,}57}} \simeq 43$ et $s_e\sqrt{\frac{n}{x_1}} \simeq 57{,}5\sqrt{\frac{21}{8{,}26}} \simeq 91{,}7$
   On estime donc que σ est dans l'intervalle $[43; 91{,}7]$ avec la confiance $c = 0{,}98$.
2. En utilisant les estimations par intervalle de confiance à 98%, peut-on conclure que les hommes de plus de 50 ans ont en moyenne moins de cholestérol que les femmes de plus de 50 ans ?
   On commence par estimer les moyennes :

   Chez les femmes Comme $n = 16 \leqslant 30$, on cherche t_α à partir de la table inverse de Student avec $p = \frac{\alpha}{2} = 0{,}01$ et $n-1 = 15$ degrés de liberté (ddl)
   On lit $t_\alpha \simeq 2{,}6025$ d'où $a_\alpha = t_\alpha\frac{s_e}{\sqrt{n-1}} \simeq 2{,}6025 \times \frac{33{,}9}{\sqrt{16-1}} \simeq 22{,}78$.
   On estime donc que μ est dans l'intervalle $[197{,}1 - 22{,}78\,;\, 197{,}1 + 22{,}78] = [174{,}32\,;\, 219{,}88]$ avec la confiance $c = 0{,}98$

   Chez les hommes Comme $n = 21 \leqslant 30$, on cherche t_α à partir de la table inverse de Student avec $p = \frac{\alpha}{2} = 0{,}01$ et $n-1 = 20$ degrés de liberté (ddl)
   On lit $t_\alpha \simeq 2{,}528$ d'où $a_\alpha = t_\alpha\frac{s_e}{\sqrt{n-1}} \simeq 2{,}528\times\frac{57{,}5}{\sqrt{21-1}} \simeq 32{,}503$.
   On estime donc que μ est dans l'intervalle $[172{,}9 -$

32,503 ; 172,9 + 32,503] $\simeq$ [140,4 ; 205,4] avec la confiance $c = 0{,}98$

Les intervalles se chevauchent, donc on ne peut pas conclure (avec la confiance $c = 98\%$) que les hommes de plus de 50 ans aient en moyenne plus de cholesterol que les femmes de plus de 50 ans.

Exercice 58 : Épreuve de dictée

corrigé en ligne

Le tableau suivant représente les résultats obtenus par un échantillon de 13 enfants de CM2 dans deux épreuves de dictée préparée : une liste de 10 mots (épreuve $\mathcal{E}_1$) et un texte (épreuve $\mathcal{E}_2$).

Élève	A	B	C	D	E	F	G	H	I	J	K	L	M
Épreuve 1	16,7	20	17,3	14,6	15,7	18,1	13,5	14,3	18,3	15,3	20	12,3	14,1
Épreuve 2	12,5	15,1	11,9	9,6	8,7	12,1	8,3	10,5	11,5	10,1	13,6	6,1	10

1. Peut-on affirmer, avec un risque d'erreur de 10%, qu'une des deux épreuves donne, pour l'ensemble des enfants de CM2, une note moyenne plus élevée que l'autre épreuve ?

   (a) On commence par calculer les moyennes et ecart types expérimentaux. En notant X les resultats de l'épreuve 1 et Y les resultats de l'épreuve 2, on obtient
   $m(X) = 16{,}17$ $s(X) = 2{,}36$.
   $m(Y) = 10{,}77$ $s(Y) = 2{,}29$.

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

   Épreuve 1 Comme $n = 13 \leqslant 30$, on cherche t_α à partir de la table inverse de Student avec $p = \frac{\alpha}{2} = 0{,}05$ et $n - 1 = 12$ degrés de liberté (ddl)
   On lit $t_\alpha \simeq 1{,}7823$ d'où $a_\alpha = t_\alpha \frac{s_e}{\sqrt{n-1}} \simeq 1{,}7823 \times \frac{2{,}36}{\sqrt{13-1}} \simeq 1{,}214\,2$.
   On estime donc que $\mu(X)$ est dans l'intervalle $[16{,}17 - 1{,}214\,2\,;\, 16{,}17 + 1{,}214\,2] \simeq [14{,}96\,;\, 17{,}38]$ avec la confiance $c = 0{,}9$

   Épreuve 2 Comme $n = 13 \leqslant 30$, on cherche t_α à partir de la table inverse de Student avec $p = \frac{\alpha}{2} = 0{,}05$ et $n - 1 = 12$ degrés de liberté (ddl)
   On lit $t_\alpha \simeq 1{,}7823$ d'où $a_\alpha = t_\alpha \frac{s_e}{\sqrt{n-1}} \simeq 1{,}7823 \times \frac{2{,}29}{\sqrt{13-1}} \simeq 1{,}178\,2$.
   On estime donc que $\mu(Y)$ est dans l'intervalle $[10{,}77 - 1{,}178\,2\,;\, 10{,}77 + 1{,}178\,2] \simeq [9{,}59\,;\, 11{,}95]$ avec la confiance $c = 0{,}9$

   (c) : Ces intervalles indiquent que $\mu(X) > \mu(Y)$ (en effet $\mu(X) > 14{,}5$ et $\mu(Y) < 12$). Donc on peut dire (avec la confiance 95%) que les résultats moyens diffèrent significativement d'une épreuve à l'autre, et plus précisemment que les résultats sont meilleurs pour la liste de mots que pour le texte.

Remarque : Comme l'echantillon est aparié, on aurait aussi pu calculer les différences entre les deux notes pour chaque individu, puis estimer la moyenne de la différence (afin de determiner 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 12 degrés de liberté la valeur $x_1 = 4{,}404$ correspondant à $q = 0{,}025$ et la valeur $x_2 = 23{,}34$ correspondant à $p = 0{,}025$
On obtient alors $s_e\sqrt{\frac{n}{x_2}} \simeq 2{,}36\sqrt{\frac{13}{23{,}34}} \simeq 1{,}76$ et $s_e\sqrt{\frac{n}{x_1}} \simeq 2{,}36\sqrt{\frac{13}{4{,}404}} \simeq 4{,}05$
On estime donc que $\sigma(X)$ est dans l'intervalle $[1{,}76; 4{,}05]$ avec la confiance $c = 0{,}95$.

Épreuve 2 On lit dans la table du χ^2 à 12 degrés de liberté la valeur $x_1 = 4{,}404$ correspondant à $q = 0{,}025$ et la valeur $x_2 = 23{,}34$ correspondant à $p = 0{,}025$
On obtient alors $s_e\sqrt{\frac{n}{x_2}} \simeq 2{,}29\sqrt{\frac{13}{23{,}34}} \simeq 1{,}71$ et $s_e\sqrt{\frac{n}{x_1}} \simeq$

$2{,}29\sqrt{\frac{13}{4{,}404}} \simeq 3{,}93$

On estime donc que $\sigma(Y)$ est dans l'intervalle $[1{,}71; 3{,}93]$ avec la confiance $c = 0{,}95$.

Exercice 59 : Boulimie et emploi corrigé en ligne

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

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

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

   On choisit au hasard un échantillon de 25000 français. Quelle est la probabilité d'avoir, au sein de cet échantillon, entre 7,6% et 8% de chômeurs ?

   Justifiez bien les éventuelles approximations que vous seriez amené à faire

   - L'échantillon de 25000 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 $X \sim \mathcal{B}(25000\,;\,0{,}078)$.
   - $n\,p(1-p) = 25\,000 \times 0{,}078\,(1-0{,}078) = 1\,797{,}9 > 1\,000$, donc on peut approcher la proportion par $\mathcal{N}(0{,}078\;;\;\sqrt{\frac{0{,}078\,(1-0{,}078)}{25\,000}}) \simeq \mathcal{N}(0{,}078\;;\;0{,}001696)$, et il n'y a pas besoin de faire de correction de continuité .
   - On designe par P cette proportion et on obtient : $\mathbb{P}\,[0{,}076 \leqslant P \leqslant 0{,}08] \simeq 0{,}762$

2. Sur un échantillon de 25000 patients atteints de boulimie, le psychologue a constaté que 8,132% d'entre eux étaient au chômage.

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

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

   Remarques : ces 8,132% de l'échantillon correspondent à $25\,000 \times 0{,}081\,32 = 2\,033$ personnes.

   On a $n\,p_e(1-p_e) = 25\,000 \times \frac{2\,033}{25\,000}\left(1 - \frac{2\,033}{25\,000}\right) \simeq 1\,867{,}676 > 10$, donc on peut utiliser la procédure du formulaire pour estimer la proportion p.

   Le formulaire indique que pour la confiance 0,9 on a $z_\alpha \simeq 1{,}645$

   d'où $a_\alpha = z_\alpha\sqrt{\frac{p_e(1-p_e)}{n}} = 1{,}645 \times \sqrt{\frac{\frac{2033}{25000}(1-\frac{2033}{25000})}{25000}} \simeq 0{,}00284$ et $p_e - a_\alpha \simeq \frac{2\,033}{25\,000} - 0{,}002\,84 = 0{,}078\,48$ et $p_e + a_\alpha \simeq \frac{2\,033}{25\,000} + 0{,}002\,84 = 0{,}084\,16$.

   On estime donc que p est dans l'intervalle $[0{,}07848; 0{,}08416]$ avec la confiance $c = 0{,}9$.

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

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

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

   On a $n\,p_e(1-p_e) = 25\,000 \times \frac{2\,033}{25\,000}\left(1 - \frac{2\,033}{25\,000}\right) \simeq 1\,867{,}676 > 10$, donc on peut utiliser la procédure du formulaire pour estimer la proportion p.

   Le formulaire indique que pour la confiance 0,98 on a $z_\alpha \simeq 2{,}326$

d'où $a_\alpha = z_\alpha\sqrt{\frac{p_e(1-p_e)}{n}} = 2{,}326 \times \sqrt{\frac{\frac{2033}{25000}(1-\frac{2033}{25000})}{25000}} \simeq 0{,}00402$ et $p_e - a_\alpha \simeq \frac{2\,033}{25\,000} - 0{,}004\,02 = 0{,}077\,3$ et $p_e + a_\alpha \simeq \frac{2\,033}{25\,000} + 0{,}004\,02 = 0{,}085\,34$.
On estime donc que p est dans l'intervalle $[0{,}0773; 0{,}08534]$ avec la confiance $c = 0{,}98$.
Cette fois-ci, la valeur 7,8% se situe à l'intéreur 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 chomage.

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

(c) Pour pouvoir estimer à 0,01% près le taux de chômage des personnes boulimiques avec une confiance de 98%, quelle taille d'échantillon faudrait il considérer ?
Pour avoir une précision 0,0001 avec la confiance 0,98, il faut $n > z_\alpha^2 \frac{p_e(1-p_e)}{h^2} \simeq 2{,}326^2 \frac{0{,}08132(1-0{,}08132)}{0{,}0001^2} \simeq 40418580$

Exercice 60 : Perception des couleurs

corrigé en ligne

Un chercheur, Mr Langlois, se demande si la perception des couleurs diffère en fonction du sexe des individus.

1. On étudie tout d'abord la prévalence du Daltonisme.
On considère d'une part un échantillon de 3 000 hommes, parmi lesquels 224 sont daltoniens, et d'autre part un échantillon de 3 000 femmes, parmi lesquelles 11 sont daltoniennes.
Peut on déduire, avec la confiance $c = 95\%$, que la prévalence du daltonisme soit plus importante chez l'homme que chez la femme ?
(a) On détermine tout d'abord des intervalles de confiance pour la proportion de daltioniens parmi les hommes et les femmes :
- **parmi les hommes :**
On a $n\,p_e(1-p_e) = 3\,000 \times \frac{224}{3\,000}\left(1 - \frac{224}{3\,000}\right) \simeq 207{,}275 > 10$, donc on peut utiliser la procédure du formulaire pour estimer la proportion p.
Le formulaire indique que pour la confiance 0,95 on a $z_\alpha \simeq 1{,}96$
d'où $a_\alpha = z_\alpha\sqrt{\frac{p_e(1-p_e)}{n}} = 1{,}96 \times \sqrt{\frac{\frac{224}{3000}(1-\frac{224}{3000})}{3000}} \simeq 0{,}00941$ et $p_e - a_\alpha \simeq \frac{224}{3\,000} - 0{,}009\,41 \simeq 0{,}065\,26$ et $p_e + a_\alpha \simeq \frac{224}{3\,000} + 0{,}009\,41 \simeq 0{,}084\,08$.
On estime donc que p est dans l'intervalle $[0{,}06526; 0{,}08408]$ avec la confiance $c = 0{,}95$.
- **parmi les femmes :**
On a $n\,p_e(1-p_e) = 3\,000 \times \frac{11}{3\,000}\left(1 - \frac{11}{3\,000}\right) \simeq 10{,}96 > 10$, donc on peut utiliser la procédure du formulaire pour estimer la proportion p.
Le formulaire indique que pour la confiance 0,95 on a $z_\alpha \simeq 1{,}96$
d'où $a_\alpha = z_\alpha\sqrt{\frac{p_e(1-p_e)}{n}} = 1{,}96 \times \sqrt{\frac{\frac{11}{3000}(1-\frac{11}{3000})}{3000}} \simeq 0{,}00216$ et $p_e - a_\alpha \simeq \frac{11}{3\,000} - 0{,}002\,16 \simeq 0{,}001\,507$ et $p_e + a_\alpha \simeq \frac{11}{3\,000} + 0{,}002\,16 \simeq 0{,}005\,827$.
On estime donc que p est dans l'intervalle $[0{,}00151; 0{,}00583]$ avec la confiance $c = 0{,}95$.

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

2. Le chercheur pense que, bien au delà de la question du daltonisme, les filles sont bien plus douées que les garçons pour détecter d'infimes nuances de couleurs. Il teste cette hypothèse sur des enfants de 8 ans, auxquels il fait passer un test standardisé de perception des couleurs. Sur un échantillon de 12 garçons, il obtient les

données suivantes :

Score	1	12	15	13	19	14	15	2	11	10	10	12

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

- On détermine tout d'abord la moyenne et l'écart-type au sein de cet échantillon :
  moyenne : $m(X) = \frac{\sum x_i}{n} = \frac{1+12+15+\cdots+12}{12} = \frac{134}{12} \simeq 11{,}17$
  $m(X^2) = \frac{\sum x_i^2}{n} = \frac{1^2+12^2+15^2+\cdots+12^2}{12} = \frac{1\,790}{12}$
  $Var(X) = m(X^2) - m(X)^2 = \frac{1\,790}{12} - \left(\frac{134}{12}\right)^2 \simeq 24{,}47$
  Écart-type : $s(X) = \sqrt{Var(X)} \simeq 4{,}95$
- Comme $n = 12 \leqslant 30$, on cherche t_α à partir de la table inverse de Student avec $p = \frac{\alpha}{2} = 0{,}005$ et $n-1 = 11$ degrés de liberté (ddl)
  On lit $t_\alpha \simeq 3{,}1058$ d'où $a_\alpha = t_\alpha \frac{s_e}{\sqrt{n-1}} \simeq 3{,}1058 \times \frac{4{,}95}{\sqrt{12-1}} \simeq 4{,}635\,3$.
  On estime donc que μ est dans l'intervalle $[11{,}17 - 4{,}635\,3\,;\,11{,}17 + 4{,}635\,3] \simeq [6{,}53\,;\,15{,}81]$ avec la confiance $c = 0{,}99$

3. Sur un échantillon de 37 filles auxquelles le chercheur fait passer le test, le score moyen est de 14,62 avec un écart-type de 2,45.

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

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

   - Comme n= 37 >30, le formulaire indique que pour la confiance 0,99 on a $z_\alpha \simeq 2{,}576$ d'où $a_\alpha = z_\alpha \frac{s_e}{\sqrt{n-1}} \simeq 2{,}576 \times \frac{2{,}45}{\sqrt{37-1}} \simeq 1{,}052$
     On estime donc que μ est dans l'intervalle $[14{,}62 - 1{,}052\,;\,14{,}62 + 1{,}052] \simeq [13{,}57\,;\,15{,}67]$, avec la confiance $c = 0{,}99$.
   - Les intervalles se chevauchent, donc ces échantillons ne permettent pas d'affirmer, avec la confiance 99%, que les filles aient en moyenne une meilleure perception des couleurs que les garçons.

4. Estimez de même les écart-types des garçons et des filles, avec la confiance 99%. Peut on conclure qu'ils diffèrent selon le sexe ?
   (a) **Estimation de l'écart-type des garçons**
   On lit dans la table du χ^2 à 11 degrés de liberté la valeur $x_1 = 2{,}603$ correspondant à $q = 0{,}005$ et la valeur $x_2 = 26{,}76$ correspondant à $p = 0{,}005$
   On obtient alors $s_e\sqrt{\frac{n}{x_2}} \simeq 4{,}95\sqrt{\frac{12}{26{,}76}} \simeq 3{,}31$ et $s_e\sqrt{\frac{n}{x_1}} \simeq 4{,}95\sqrt{\frac{12}{2{,}603}} \simeq 10{,}6$
   On estime donc que σ est dans l'intervalle $[3{,}31; 10{,}6]$ avec la confiance $c = 0{,}99$.
   (b) **Estimation de l'écart-type des filles**
   On lit dans la table du χ^2 à 36 degrés de liberté la valeur $x_1 = 17{,}89$ correspondant à $q = 0{,}005$ et la valeur $x_2 = 61{,}58$ correspondant à $p = 0{,}005$
   On obtient alors $s_e\sqrt{\frac{n}{x_2}} \simeq 2{,}45\sqrt{\frac{37}{61{,}58}} \simeq 1{,}9$ et $s_e\sqrt{\frac{n}{x_1}} \simeq 2{,}45\sqrt{\frac{37}{17{,}89}} \simeq 3{,}52$
   On estime donc que σ est dans l'intervalle $[1{,}9; 3{,}52]$ avec la confiance $c = 0{,}99$.
   (c) Les intervalles de confiance se chevauchent, donc on ne peut pas affirmer, avec la confiance 99%, que les écart-types diffèrent.

5. Quelle taille d'échantillon aurait-il fallu considérer pour estimer le score moyen des garçons à 0,2 point près, avec une confiance de 99% ?
   Pour avoir une précision 0,2 avec la confiance 0,99, il faut $n > z_\alpha^2 \frac{(s_e)^2}{h^2} \simeq 2{,}576^2 \frac{4{,}95^2}{0{,}2^2} \simeq 4065$

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