Exercices de TD de statistiques

Rappels

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

Exercice 1 : Fractions

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

1. Grâce à cette fonctionnalité de la calculatrice, simplifiez chacune des fractions suivantes : $\frac{-143}{-104}$, $\frac{-10\times3}{-10}$, $\frac{-13-9}{4\times2}$, $\frac{14\times2}{67-4}$, $\frac{-7-48}{126-49}$, $\frac{80-35}{-27-15}$.

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

   ```
   -143/-104▸Frac
                11/8
   (-10×3)/-10▸Frac
                   3
   (-13-9)/(4×2)▸Frac
               -11/4
   (14×2)/(67-4)▸Frac
                 4/9
   (-7-48)/(126-49)▸Frac
                -5/7
   (80-35)/(-27-15)▸Frac
              -15/14
   ```

   $\frac{-143}{-104}=\frac{11}{8}$, $\frac{-10\times3}{-10}=3$,

   $\frac{-13-9}{4\times2}=-\frac{11}{4}$, $\frac{14\times2}{67-4}=\frac{4}{9}$,

   $\frac{-7-48}{126-49}=-\frac{5}{7}$, $\frac{80-35}{-27-15}=-\frac{15}{14}$.

   Remarque : La capture d'écran correspond à une TI. Sur une casio, si l'on veut par exemple calculer $\frac{-13-9}{4\times2}$, on tappe **(-13-9)÷(4*2)** puis si le résultat affiché est un chiffre à virgule on appuie sur F↔D pour en faire une fraction. Au lieu d'afficher $\frac{-11}{4}$, certaines calculatrices affichent alors $-2\,\frac{3}{4}$ (ces calculatrices considèrent que cela signifie $-\left(2+\frac{3}{4}\right)$). 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) :

   $\frac{-143}{-104}=1{,}375$, $\frac{-10\times3}{-10}=3$, $\frac{-13-9}{4\times2}=-2{,}75$, $\frac{14\times2}{67-4}\simeq0{,}444$,

   $\frac{-7-48}{126-49}\simeq-0{,}714$, $\frac{80-35}{-27-15}\simeq-1{,}071$.

   ```
   -143/-104
            1.375
   (-10×3)/-10
                3
   (-13-9)/(4×2)
            -2.75
   (14×2)/(67-4)
     0.4444444444
   (-7-48)/(126-49)
    -0.7142857143
   (80-35)/(-27-15)
     -1.071428571
   ```

   L'ordre croissant de ces nombres est :
   $-2{,}75<-1{,}071<-0{,}714<0{,}444<1{,}375<3$
   Donc la réponse est :
   $\frac{-13-9}{4\times2}<\frac{80-35}{-27-15}<\frac{-7-48}{126-49}<\frac{14\times2}{67-4}<\frac{-143}{-104}<\frac{-10\times3}{-10}$

Exercice 2 : Utilisation de la calculatrice

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 calculette et vaut -1.

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

2.
```
1/(2020×2021)
  2.449527486E-7
```
(a) $\frac{1}{2020\times2021} \simeq 0{,}000000245$
(b) $\frac{1}{2020\times2021} \simeq 24495274{,}86$
(c) $\frac{1}{2020\times2021} \simeq 2{,}45$

La bonne réponse est la réponse (a). En effet `2.449527486E-7` désigne la quantité $2{,}449527486 \times 10^{-7}$, qui vaut 0,0000002449527486.

3.
```
-8×44→X
            -352
X²
          123904
```
(a) $-8 \times 44^2 = 123904$
(b) $-352^2 = 123904$
(c) $(-8 \times 44)^2 = 123904$

La bonne réponse est la réponse (c). En effet dans `X²`, `X` vaut -8×44, (calculé à la ligne précédente), de sorte qu'on a calculé $(-8 \times 44)^2$.

4.
```
16-25
              -9
Ans×2021
          -18189
```
(a) $16 - 25 \times 2021 = -18189$
(b) $9 \times 2021 = -18189$
(c) $(16 - 25) \times 2021 = -18189$

La bonne réponse est la réponse (c). En effet dans `Ans×2021`, `Ans` vaut $16 - 25$, (calculé à la ligne précédente), de sorte qu'on a calculé $(16 - 25) \times 2\,021$.

5.
```
{4,2,-1}→List 1
        {4,2,-1}
Sum((1+List 1)²)
              34
```
(a) $1 + (4 + 2 - 1)^2 = 34$
(b) $1 + 4^2 + 2^2 - 1^2 = 34$
(c) $(1 + 4)^2 + (1 + 2)^2 + (1 - 1)^2 = 34$

La bonne réponse est la réponse (c). En effet `(1+List 1)²` désigne la liste dont les éléments sont $(1+4)^2$, $(1+2)^2$ et $(1-1)^2$. En conséquence, leur somme `Sum((1+List 1)²)` désigne $(1+4)^2+(1+2)^2+(1-1)^2$,

Exercice 3 : Puissances de dix

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

$47 \div 100$	$0{,}47 \times 10^2$	$18{,}885\,392\,4 \times 10^{-5}$
$1{,}048\,4 \times 10^6$	$8 \times 100\,000\,000$	$111 \times 10\,000\,000$
$19\,820 \div 100\,000$	$11{,}283\,133 \times 10^{-6}$	8×100

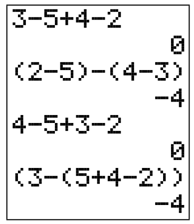
- $47 \div 100 = 0{,}47$ (que l'on peut aussi écrire $4{,}7 \times 10^{-1}$)
- $0{,}47 \times 10^2 = 47$ (que l'on peut aussi écrire $4{,}7 \times 10^1$)
- $18{,}885\,392\,4 \times 10^{-5} = 0{,}000188853924$ (que l'on peut aussi écrire $1{,}88853924 \times 10^{-4}$)
- $1{,}048\,4 \times 10^6 = 1048400$ (que l'on peut aussi écrire $1{,}0484 \times 10^6$)
- $8 \times 100\,000\,000 = 800000000$ (que l'on peut aussi écrire 8×10^8)
- $111 \times 10\,000\,000 = 1110000000$ (que l'on peut aussi écrire $1{,}11 \times 10^9$)
- $19\,820 \div 100\,000 = 0{,}1982$ (que l'on peut aussi écrire $1{,}982 \times 10^{-1}$)
- $11{,}283\,133 \times 10^{-6} = 0{,}000011283133$ (que l'on peut aussi écrire $1{,}1283133 \times 10^{-5}$)
- $8 \times 100 = 800$ (que l'on peut aussi écrire 8×10^2)

Exercice 4 : Règles de calcul

1. On considère les expressions suivantes :
   - $3-5+4-2$
   - $4-5+3-2$
   - $(2-5)-(4-3)$
   - $(3-(5+4-2))$

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

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



   - $3-5+4-2=0$
   - $(2-5)-(4-3)=-4$
   - $4-5+3-2=0$
   - $(3-(5+4-2))=-4$

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

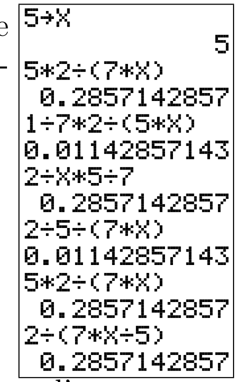
   Remarque : L'égalité de ces expression se comprend à partir des règles de parenthésage et de signe expliquées dans les "Rappels" au début des notes de cours.

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

   (a) Les expressions suivantes
   - $5\frac{2}{7\times x}$
   - $\frac{2\div 5}{7\times x}$
   - $\frac{1}{7}\times\frac{2}{5\times x}$
   - $\frac{5\times 2}{7\times x}$
   - $2\div x\times 5\div 7$
   - $\frac{2}{7\times x\div 5}$

   lorsque $x=5$.

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



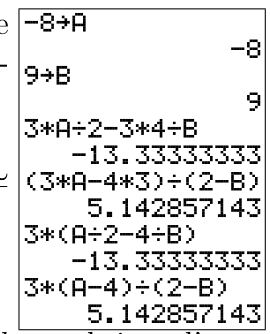
   - $5\frac{2}{7\times x}\simeq 0{,}286$
   - $\frac{1}{7}\times\frac{2}{5\times x}\simeq 0{,}011$
   - $2\div x\times 5\div 7\simeq 0{,}286$
   - $\frac{2\div 5}{7\times x}\simeq 0{,}011$
   - $\frac{5\times 2}{7\times x}\simeq 0{,}286$
   - $\frac{2}{7\times x\div 5}\simeq 0{,}286$

   Ainsi, lorsque $x=5$, on obtient d'une part que les expression $5\frac{2}{7\times x}$, $2\div x\times 5\div 7$, $\frac{5\times 2}{7\times x}$ et $\frac{2}{7\times x\div 5}$ sont égales, d'autre part que les expression $\frac{1}{7}\times\frac{2}{5\times x}$ et $\frac{2\div 5}{7\times x}$ sont égales.

   (b) Les expressions
   - $\frac{3\times a}{2}-\frac{3\times 4}{b}$
   - $3\left(\frac{a}{2}-\frac{4}{b}\right)$
   - $(3\times a-4\times 3)\div(2-b)$
   - $3\frac{(a-4)}{(2-b)}$

   où $a=-8$ et $b=9$.

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

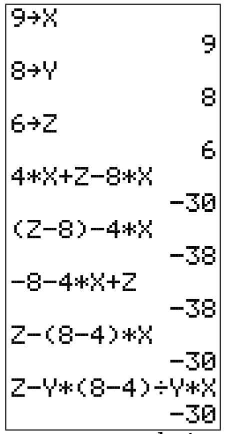


   - $\frac{3\times a}{2}-\frac{3\times 4}{b}\simeq -13{,}333$
   - $(3\times a-4\times 3)\div(2-b)\simeq 5{,}143$
   - $3\left(\frac{a}{2}-\frac{4}{b}\right)\simeq -13{,}333$
   - $3\frac{(a-4)}{(2-b)}\simeq 5{,}143$

   Ainsi, pour ces valeurs de a et b, on obtient d'une part que les expression $(3\times a-4\times 3)\div(2-b)$ et $3\frac{(a-4)}{(2-b)}$ sont égales, d'autre part que les expression $\frac{3\times a}{2}-\frac{3\times 4}{b}$ et $3\left(\frac{a}{2}-\frac{4}{b}\right)$ sont égales.

   (c) Lorsque $x=9$, $y=8$ et $z=6$:
   - $4x+z-8x$
   - $-8-4x+z$
   - $z-y\times\frac{8-4}{y}x$

• $(z-8)-4x$ • $z-(8-4)x$

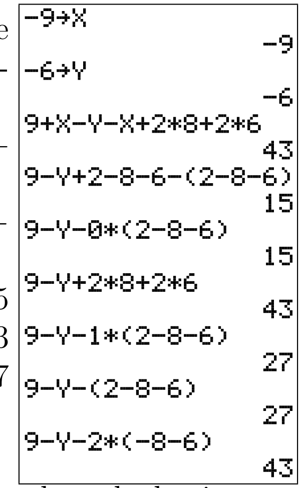


Comme en témoigne la capture d'écran ci-contre, on a les valeurs ci-dessous :

- $4x+z-8x=-30$
- $(z-8)-4x=-38$
- $-8-4x+z=-38$
- $z-(8-4)x=-30$
- $z-y\times\frac{8-4}{y}x=-30$

Ainsi, pour ces valeurs de x, y et z, on obtient d'une part que les expression $4x+z-8x$, $z-(8-4)x$ et $z-y\times\frac{8-4}{y}x$ sont égales, d'autre part que les expression $(z-8)-4x$ et $-8-4x+z$ sont égales.

(d) Lorsque $x_1=-9$ et $x_2=-6$:
• $9+x_1-x_2-x_1+2\times 8+2\times 6$
• $9-x_2+2-8-6-(2-8-6)$
• $9-x_2-0(2-8-6)$
• $9-x_2+2\times 8+2\times 6$
• $9-x_2-1(2-8-6)$
• $9-x_2-(2-8-6)$
• $9-x_2-2(-8-6)$



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

- $9+x_1-x_2-x_1+2\times 8+2\times 6=43$
- $9-x_2+2-8-6-(2-8-6)=15$
- $9-x_2-0(2-8-6)=15$
- $9-x_2+2\times 8+2\times 6=43$
- $9-x_2-1(2-8-6)=27$
- $9-x_2-(2-8-6)=27$
- $9-x_2-2(-8-6)=43$

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 expression $9+x_1-x_2-x_1+2\times 8+2\times 6$, $9-x_2+2\times 8+2\times 6$ et $9-x_2-2(-8-6)$ sont égales, d'autre part que les expression $9-x_2-1(2-8-6)$ et $9-x_2-(2-8-6)$ sont égales, et enfin que les expression $9-x_2+2-8-6-(2-8-6)$ et $9-x_2-0(2-8-6)$ sont égales.

(e) Lorsque $a_1=5$, $a_2=-8$, $x_1=-8$ et $x_2=-2$:
• $x_2+5\times a_2+6^2\times x_1+a_1$
• $x_2\times a_2+5\times a_2+6^2(x_1+a_1)$
• $(x_2+5)\times a_2+(x_1+a_1)\times 6^2$
• $(x_2+(5\times a_2))+6^2(x_1+a_1)$
• $x_2+5\times a_2+(6^2\times x_1+(a_1))$
• $5\times a_2+(x_2+(x_1+a_1)\times 6\times 6)$

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

- $x_2+5\times a_2+6^2\times x_1+a_1 = -325$
- $x_2\times a_2+5\times a_2+6^2(x_1+a_1) = -132$
- $(x_2+5)\times a_2+(x_1+a_1)\times 6^2 = -132$
- $(x_2+(5\times a_2))+6^2(x_1+a_1) = -150$
- $x_2+5\times a_2+(6^2\times x_1+(a_1)) = -325$
- $5\times a_2+(x_2+(x_1+a_1)\times 6\times 6) = -150$

```
5→A
                    5
-8→B
                   -8
-8→X
                   -8
-2→Y
                   -2
Y+5*B+6²*X+A
                 -325
Y*B+5*B+6²*(X+A)
                 -132
(Y+5)*B+(X+A)*6²
                 -132
(Y+(5*B))+6²*(X+A)
                 -150
Y+5*B+(6²*X+(A))
                 -325
5*B+(Y+(X+A)*6*6)
                 -150
```

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 expression $(x_2+(5\times a_2))+6^2(x_1+a_1)$ et $5\times a_2+(x_2+(x_1+a_1)\times 6\times 6)$ sont égales, d'autre part que les expression $x_2\times a_2+5\times a_2+6^2(x_1+a_1)$ et $(x_2+5)\times a_2+(x_1+a_1)\times 6^2$ sont égales, et enfin que les expression $x_2+5\times a_2+6^2\times x_1+a_1$ et $x_2+5\times a_2+(6^2\times x_1+(a_1))$ sont égales.

Compléments

Exercice 5 : Précision des calculs

Maeva, Clara, Jeanne, Erwan et Hedi étudient en première année de psychologie à l'université de Bourgogne.

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

$m(X) = \frac{1{,}59+1{,}71+1{,}63}{3}$

$m(X^2) = \frac{1{,}59^2+1{,}71^2+1{,}63^2}{3}$

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

$s = \sqrt{Var}$

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

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

Copie de Maeva

$m(X) = \frac{1{,}59+1{,}71+1{,}63}{3} \simeq 1{,}64$

$m(X^2) = \frac{1{,}59^2+1{,}71^2+1{,}63^2}{3} \simeq 2{,}70$

$Var \simeq 2{,}7 - 1{,}64^2 \simeq 0{,}01$

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

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

Calculatrice de Maeva

```
(1.59+1.71+1.63)÷3
              1.643333333
(1.59²+1.71²+1.63²)÷3
              2.703033333
2.7-1.64²
                   0.0104
√(0.01)
                      0.1
```

Copie de Clara

$m(X) = \frac{1{,}59+1{,}71+1{,}63}{3} = \frac{4{,}93}{3}$

$m(X^2) = \frac{1{,}59^2+1{,}71^2+1{,}63^2}{3} = \frac{8{,}109\,1}{3}$

$Var \simeq \frac{8{,}109\,1}{3} - \left(\frac{4{,}93}{3}\right)^2 \simeq 0{,}0025$

$s = \sqrt{Var} \simeq 0{,}05$ mètre.

Donc l'écart-type est d'environ 5 cm.

Calculatrice de Clara

```
1.59+1.71+1.63
                     4.93
1.59²+1.71²+1.63²
                   8.1091
8.1091÷3-(4.93÷3)²
           0.002488888889
√(Ans)
            0.04988876516
```

Copie de Jeanne

$m(X) = \frac{1{,}59+1{,}71+1{,}63}{3} \simeq 1{,}64$
$m(X^2) = \frac{1{,}59^2+1{,}71^2+1{,}63^2}{3} \simeq 2{,}70$
$Var = m(X^2) - (m(X))^2 \simeq 0{,}0025$
$s = \sqrt{Var} \simeq 0{,}05$ mètre.
Donc l'écart-type est d'environ 5 cm.

Calculatrice de Jeanne

```
(1.59+1.71+1.63)÷3→M
                 1.643333333
(1.59²+1.71²+1.63²)÷3→N
                 2.703033333
N-M²→V
              0.002488888889
√(V)
               0.04988876516
```

Copie d'Erwan

$m(X) = \frac{1{,}59+1{,}71+1{,}63}{3} \simeq 1{,}64$
$m(X^2) = \frac{1{,}59^2+1{,}71^2+1{,}63^2}{3} \simeq 2{,}70$
$Var = m(X^2) - (m(X))^2 \simeq 0{,}0025$
$s = \sqrt{Var} \simeq 0{,}05$ mètre.
Donc l'écart-type est d'environ 5 cm.

Calculatrice d'Erwan

```
{1.59,1.71,1.63}→List 1
           {1.59,1.71,1.63}
Sum (List 1)÷3→M
                 1.643333333
Sum ((List 1)²)÷3→N
                 2.703033333
N-M²→V
              0.002488888889
√(V)
               0.04988876516
```

Copie de Hedi

$s = \sqrt{\frac{1{,}59^2+1{,}71^2+1{,}63^2}{3} - \left(\frac{1{,}59+1{,}71+1{,}63}{3}\right)^2}$
$\simeq 0{,}05$ mètre.
Donc l'écart-type est d'environ 5 cm.

Calculatrice de Hedi

```
√((1.59²+1.71²+1.63²)÷3
-((1.59+1.71+1.63)÷3)²)
               0.04988876516
```

1. Parmi ces différentes réponses, laquelle/lesquelles vous semble(nt) correcte(s) ?
   Expliquez pourquoi l'un·e des étudiant·e·s a obtenue une réponse erronée ?
2. On demande ensuite de calculer l'écart type de la taille d'un échantillon de 5 hommes. Comme l'échantillon n'est plus le même, il faut cette fois-ci calculer
   $m(X) = \frac{1{,}74+1{,}82+1{,}89+1{,}73+1{,}88}{5}$
   $m(X^2) = \frac{1{,}74^2+1{,}82^2+1{,}89^2+1{,}73^2+1{,}88^2}{5}$
   $Var = m(X^2) - (m(X))^2$
   $s = \sqrt{Var}$
   Rédigez ce calcul d'une façon qui empêche des erreurs arrondis (vous pourrez reprendre la rédaction d'un·e des étudiant·e·s ayant trouvé la bonne réponse).

Exercice 6 : Utilisation du symbole $\sum$

1. Calculer, sans utiliser la calculatrice, chacune des sommes suivantes : $\sum_{n=1}^{5} n$, $\sum_{n=-5}^{5} n$, $\sum_{i=-3}^{3} i^2$.
2. Écrivez avec le symbole $\sum$ la somme suivante $\frac{6^2}{6+9} + \frac{7^2}{7+9} + \frac{8^2}{8+9} + \frac{9^2}{9+9} + \frac{10^2}{10+9} + \frac{11^2}{11+9} + \frac{12^2}{12+9} + \frac{13^2}{13+9}$.
3. On pose $x_1 = 6$, $x_2 = 7$, $x_3 = 8$, $x_4 = 9$, $x_5 = 10$, $n_1 = 6$, $n_2 = 5$, $n_3 = 5$, $n_4 = 1$ et $n_5 = 9$.
   (a) Calculer $\sum_{i=1}^{5} n_i (x_i)^2$ sans utiliser de liste sur la calculatrice.
   (b) Calculer à nouveau $\sum_{i=1}^{5} n_i (x_i)^2$, en utilisant cette fois-ci des listes sur la calculatrice.

Chapitre 1 : Statistiques descriptives univariées

Exercice 7 : Type de variables

1. Mme Giraud recueille des données sur
   (a) le nombre de rendez-vous médicaux que différents patients ont pris au cours de l'année 2020
   (b) la popularité des différents footballeurs de l'équipe de France
   (c) les noms des rues de Dijon
   (d) le taux de testostérone parmi les patients souffrant de troubles cognitifs

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

2. De même quel est le type de variable lorsqu'on étudie :
   (a) la couleur des voitures stationnées dans le campus
   (b) la pluie tombée à Dijon au cours des différents mois de l'année
   (c) le degré de satisfaction des usagers des transports en commun
   (d) les prénoms des enfants nés en 2020
   (e) le temps hebdomadaire passé devant la télévision par des enfants

Exercice 8 : Taille des ménages

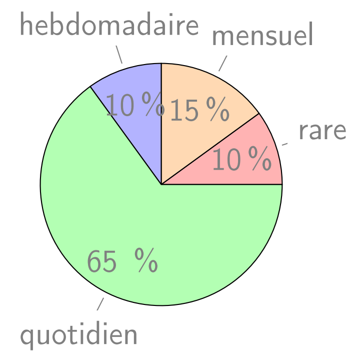
On appelle T la variable statistique indiquant le nombre de personnes d'un ménage. Au sein d'une certaine ville, on obtient les données suivantes :

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

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

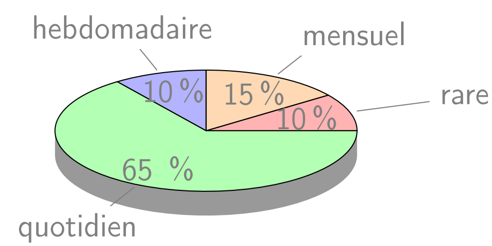
Exercice 9 : Moyens de transport

1. **Utilisation des transports en commun**
   (a) Un étudiant en sociologie interroge 20 personnes de son entourage sur leur utilisation des transports en commun. Il réalise le graphique ci dessous après les avoir regroupés en quatre groupes :
   rare : ceux qui utilisent les transports en communs moins d'une fois par mois
   mensuel : au moins une fois par mois, mais moins d'une fois par semaine
   hebdomadaire : entre une et cinq fois par semaine
   quotidien : au moins 5 fois par semaine

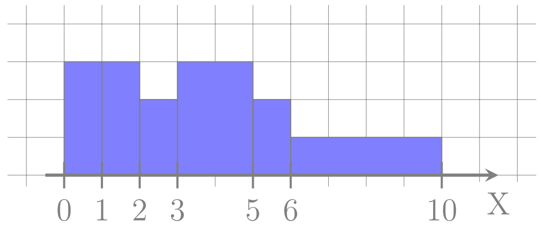


   i. Déterminer les effectifs des différentes modalités
   ii. Quelle est, au sein de cet échantillon, la proportion d'individus qui utilisent les transports en commun moins d'une fois par semaine ?

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



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



(a) Déterminer les fréquences des différentes classes

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

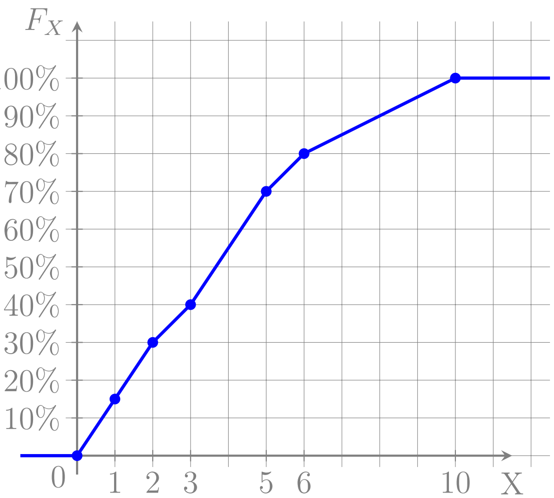
(b) Commentez : comparer les fréquences de la première et de la dernière classe, ainsi que la hauteur des rectangles correspondants.

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

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

   (b) Déterminer graphiquement les proportions $\mathbb{P}_r[X \leqslant 4]$, $\mathbb{P}_r[X \geqslant 7]$ et $\mathbb{P}_r[4 \leqslant X \leqslant 7]$ (de manière approchée).

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



Exercice 10 : “Partenariat régional économique global”

On considère dans cet exercice les pays signataires du *Partenariat régional économique global*, un accord de libre échange dans le pacifique. Les pays concernés sont les pays ci-dessous :

Australie Monarchie 25,7 M.hab.	Birmanie République 56,6 M.hab.	Brunei Monarchie 0,5 M.hab.	Cambodge Monarchie 16,9 M.hab.
Chine République 1 394 M.hab.	Corée du Sud République 51,7 M.hab.	Indonésie République 267 M.hab.	Japon Monarchie 125,5 M.hab.
Laos République 7,4 M.hab.	Malaisie Monarchie 32,7 M.hab.	Nouvelle-Zélande Monarchie 4,9 M.hab.	Philippines République 109,2 M.hab.
Singapour République 6,2 M.hab.	Thaïlande Monarchie 69 M.hab.	Viêt Nam République 98,7 M.hab.	

1. Quelle est la proportion de monarchies parmi ces pays ?
2. Parmi l'ensemble des habitants de ces états, quelle proportion vit dans une république ?

3. Parmi ces pays, quelle est la population moyenne des républiques et celle des monarchies ? Calculer aussi les médianes et écarts type.

Exercice 11 : Développement psychomoteur

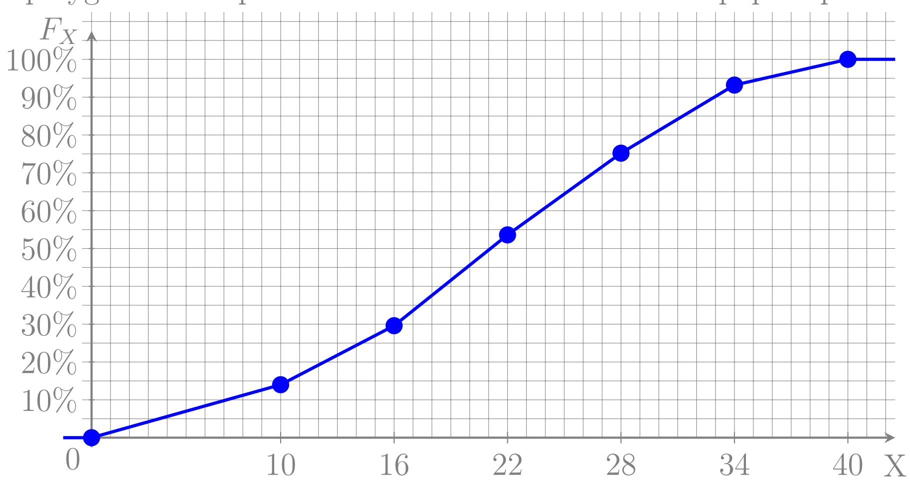
Les données ci-dessous décrivent, pour un ensemble de 160 bébés âgés de deux ans, la variable statistique X représentant le score de développement psychomoteur (SDP) :

SDP	[54 ; 64[	[64 ; 74[	[74 ; 84[	[84 ; 94[	[94 ; 104[	[104 ; 114[	[114 ; 124[	[124 ; 134[
Effectif	1	13	26	49	36	23	8	4

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

Exercice 12 : Méthodes d'apprentissage

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



1. Déterminer graphiquement la médiane et les quartiles de la variable statistique X.
2. (a) Donner une approximation de $\mathbb{P}_r[X < 25]$.
   (b) Quelle est environ la proportion d'enfants dont la note est supérieure à 18 ?
   (c) Combien vaut environ $\mathbb{P}_r[20 \leqslant X < 30]$?
3. Plus précisément, les notes recueillies sont données ci-dessous :

X	[0 ; 10[	[10 ; 16[	[16 ; 22[	[22 ; 28[	[28 ; 34[	[34 ; 40[
Effectif	35	39	60	54	45	17

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

Exercice 13 :

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

Exercice 14 : Type de variables

1. Mr Riviere recueille des données sur
   (a) la sensibilité aux questions écologiques des habitants d'un même quartier
   (b) le temps mis par des rats pour sortir d'un labyrinthe
   (c) les langues officielles des différents pays européens
   (d) le nombre de personnes des ménages Français

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

Cas	Population	Variable	Nature
(a)	les habitants de ce quartier	la sensibilité aux questions écologiques	qualitative ordinale
(b)	des rats	temps de sortie du labyrinthe	quantitative continue
(c)	l'ensemble des pays européens	langue officielle	qualitative nominale
(d)	les ménages Français	le nombre de personnes	quantitative discrète

2. De même quel est le type de variable lorsqu'on étudie :
   (a) les marques des téléphones vendus en 2020
   qualitative nominale
   (b) le nombre de crises d'épilepsie de différents patients au cours du mois d'octobre 2020
   quantitative discrète
   (c) le nombre moyen d'enfants par femme dans chaque pays européen
   quantitative continue
   (d) le nombre d'enfants de chaque femme dijonnaise
   quantitative discrète
   (e) le poids des appareils photos vendus en 2020
   quantitative continue
   (f) le nombre de pages des livres d'une bibliothèque
   quantitative discrète
   (g) les numéros de téléphones figurant dans les pages blanches dijonnaises
   qualitative nominale
   (h) la douleur ressentie par différents patients atteints de la même pathologie
   qualitative ordinale
   (i) les numéros de sécurité sociale des enseignants de l'université de Bourgogne
   qualitative nominale
   (j) le nombre de pièces des appartements en location à Dijon
   quantitative discrète

Exercice 15 : Questionnaire

Un chercheur établit un questionnaire visant à évaluer les compétences logiques de jeunes adolescents. Il fait remplir ce questionnaire à 13 sujets, et pour chacun d'eux il note dans un tableau le symbole « ✓ » si la réponse donnée est correcte, « 0 » si la réponse donnée est fausse, et a laissé la case vide si le sujet n'a pas répondu. Il obtient le tableau suivant :

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

1. Quelle a été le taux de réponse à la question 7 ?
   En tout, 7 personnes (parmi les 13 interrogés) ont répondu à la question 7. Le taux de réponse à cette question est donc $\frac{7}{13} \simeq 0{,}538$.

2. Quelle proportion de sujets a répondu correctement à la question 3 ?
   Comme 5 personnes ont répondu correctement à la question 3, cette proportion est $\frac{5}{13} \simeq 0{,}385$.
3. Parmi les individus ayant répondu à la question 8, quelle proportion à donné la bonne réponse ?
   Parmi les 9 personnes qui ont répondu à la question 8, il y en a 7 qui ont répondu correctement, soit une proportion de $\frac{7}{9} \simeq 0{,}778$.
4. Le chercheur décide de résumer par une note les réponses données au questionnaire. Il attribue un point par bonne réponse (et zéro point par réponse erronée ou manquante). Calculer les notes (sur 10) attribuée à chaque jeune interrogé.

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

5. Calculer la moyenne, l'écart type et la médiane de ces notes.
   moyenne : $m(X) = \frac{\sum x_i}{n} = \frac{4+3+4+\cdots+5}{13} = \frac{56}{13} \simeq 4{,}31$
   $m(X^2) = \frac{\sum x_i^2}{n} = \frac{4^2+3^2+4^2+\cdots+5^2}{13} = \frac{282}{13}$
   $Var(X) = m(X^2) - m(X)^2 = \frac{282}{13} - \left(\frac{56}{13}\right)^2 \simeq 3{,}14$
   Écart-type : $s(X) = \sqrt{Var(X)} \simeq 1{,}77$
   La médiane est la valeur numéro $\frac{13+1}{2} = 7$ (en ordonnant par ordre croissant). C'est donc 4.

Exercice 16 : Regroupement en classes

Voici les notes obtenues par un groupe d'élèves lors d'un contrôle sur 10 points :
9,5 ; 7 ; 6 ; 7 ; 7 ; 3,5 ; 4 ; 6 ; 7,5 ; 1 ; 8 ; 6 ; 4 ; 9,5 ; 8 ; 4 ; 10 ; 3 ; 3 ; 5 ; 6,5 ; 7,5 ; 4,5 ; 3 ; 6 ; 5 ; 2 ; 3 ; 8 et 7

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

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

2. Quelle est la taille de l'échantillon ?
   La taille de l'échantillon est $n = 30$.
3. Donner la moyenne, l'écart type et la médiane.
   moyenne : $m(X) = \frac{\sum_i x_i n_i}{n} = \frac{1\times1+2\times1+3\times4+\cdots+10\times1}{30} = \frac{171{,}5}{30} \simeq 5{,}72$
   $m(X^2) = \frac{\sum_i x_i^2 n_i}{n} = \frac{1^2\times1+2^2\times1+3^2\times4+\cdots+10^2\times1}{30} = \frac{1\,138{,}75}{30}$
   $Var(X) = m(X^2) - m(X)^2 = \frac{1\,138{,}75}{30} - \left(\frac{171{,}5}{30}\right)^2 \simeq 5{,}28$
   Écart-type : $s(X) = \sqrt{Var(X)} \simeq 2{,}3$
   La médiane est la valeur numéro $\frac{30+1}{2} = 15{,}5$ (en ordonnant par ordre croissant). C'est donc 6.
4. Transformer les données en les rangeant en classes d'amplitude 1,5. Calculer les fréquences et les fréquences cumulées des classes.

classes	[0,000000000000000 ; 1,5[	[1,5 ; 3[	[3 ; 4,5[	[4,5 ; 6[	[6 ; 7,5[	[7,5 ; 9[	[9 ; 10,5[
effectifs	1	1	8	3	9	5	3
fréquences	0,033	0,033	0,267	0,1	0,3	0,167	0,1
fréquences cumulées	0,033	0,066	0,333	0,433	0,733	0,9	1

5. Calculer la moyenne, l'écart type et la médiane des notes regroupées en classes. Que remarque-t-on ?
   moyenne :
   $$m(X) = \frac{\sum c_i n_i}{n} = \frac{0{,}75\times1+2{,}25\times1+3{,}75\times8+\cdots+9{,}75\times3}{30} = \frac{180}{30} = 6$$
   $$m(X^2) = \frac{\sum c_i^2 n_i}{n} = \frac{0{,}75^2\times1+2{,}25^2\times1+3{,}75^2\times8+\cdots+9{,}75^2\times3}{30} = \frac{1\,236{,}375}{30}$$
   $$Var(X) = m(X^2) - m(X)^2 = \frac{1\,236{,}375}{30} - \left(\frac{180}{30}\right)^2 \simeq 5{,}21$$
   Écart-type : $s(X) = \sqrt{Var(X)} \simeq 2{,}28$
   Classe de la médiane : [6 ; 7,5[
   Méd $\simeq a_i + \frac{a_{i+1}-a_i}{F_X(a_{i+1})-F_X(a_i)}(0{,}5 - F_X(a_i)) \simeq 6 + \frac{7{,}5-6}{0{,}733-0{,}433}(0{,}5 - 0{,}433) \simeq 6{,}33$
   Ce regroupement en classe donne lieu à des erreurs de l'ordre de 0,25 points (dans le cas de la moyenne), ce qui est de l'ordre de

4% de la note.

Exercice 17 : Âge et performances mémorielles

On considère un échantillon de 16 personnes de 35 à 80 ans, auxquels on attribue une note indiquant leurs performances mémorielles. On note leur âge X et leur performances mémorielles Y.

Les données mesurées sont les suivantes :

X	80	56	67	72	67	51	80	42	56	47	53	37	48	46	44	35
Y	22	38	33	36	22	47	30	45	48	55	38	81	46	33	51	60

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

• Âge X :

moyenne : $m(X) = \frac{\sum x_i}{n} = \frac{80+56+67+\cdots+35}{16} = \frac{881}{16} \simeq 55{,}06\,\text{ans}$

$m(X^2) = \frac{\sum x_i^2}{n} = \frac{80^2+56^2+67^2+\cdots+35^2}{16} = \frac{51\,567}{16}$

$Var(X) = m(X^2) - m(X)^2 = \frac{51\,567}{16} - \left(\frac{881}{16}\right)^2 \simeq 191{,}06$

Écart-type : $s(X) = \sqrt{Var(X)} \simeq 13{,}82\,\text{ans}$

• Performances mémorielles Y :

moyenne : $m(Y) = \frac{\sum x_i}{n} = \frac{22+38+33+\cdots+60}{16} = \frac{685}{16} \simeq 42{,}81$

$m(Y^2) = \frac{\sum x_i^2}{n} = \frac{22^2+38^2+33^2+\cdots+60^2}{16} = \frac{32\,671}{16}$

$Var(Y) = m(Y^2) - m(Y)^2 = \frac{32\,671}{16} - \left(\frac{685}{16}\right)^2 \simeq 209{,}03$

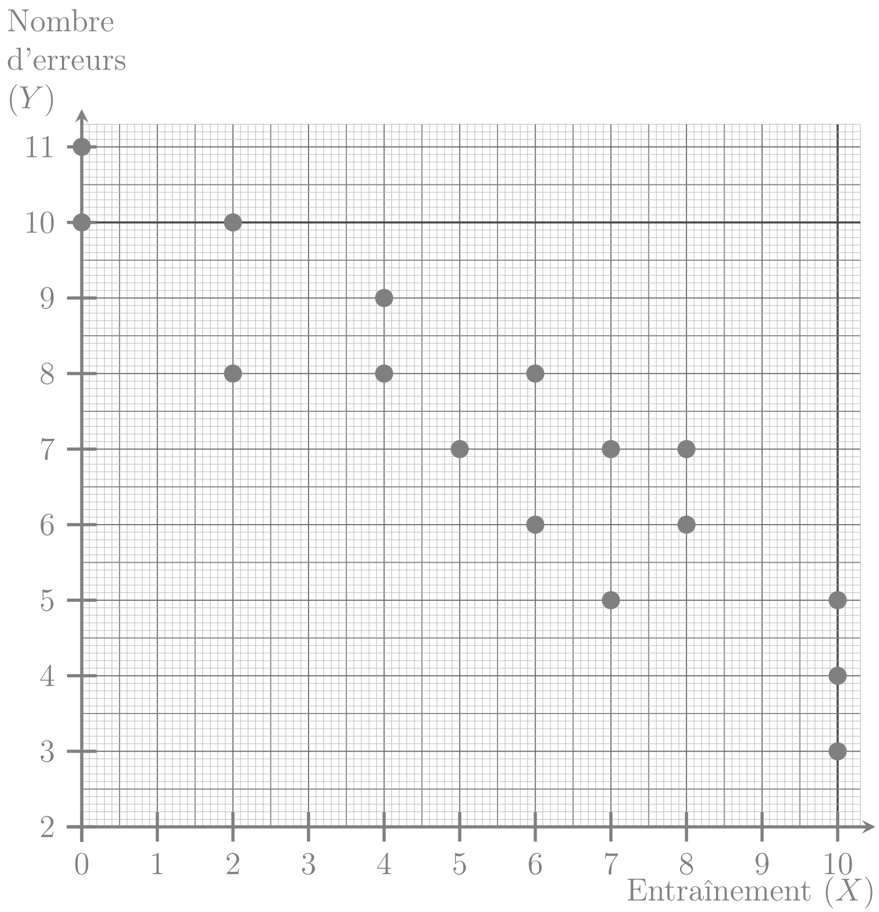
Écart-type : $s(Y) = \sqrt{Var(Y)} \simeq 14{,}46$

Chapitre 2 : Statistiques descriptives bivariées

Exercice 18 : Entraînement à un exercice de logique

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

Pour un échantillon de 16 enfants, on considère le nombre Y d'erreur commises à un test de logique, tandis que le nombre d'exercices similaires qu'ils ont déjà effectués auparavant est noté X. On regroupe ces résultats sous la forme du nuage de points suivant (effectué sur un papier "millimétré") :



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

Entraînement X																
Nombre d'erreurs Y																

2. Calculer le coefficient de corrélation des rangs de Spearman. Interpréter la valeur de ce coefficient.
3. Calculer les moyennes et écarts type des variables X et Y, puis leur coefficient de corrélation linéaire. Interpréter la valeur de ce coefficient.
4. Si un enfant a déjà fait 2 exercices de logique, alors combien estimeriez qu'il fera d'erreurs si on lui fait à nouveau passer un test similaire ?
5. Un enfant a commis 7 erreurs. Combien de fois estimeriez vous qu'il avait déjà fait un test similaire pour s'entraîner ?

Exercice 19 : Stress et temps de réponse

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

Sujet	A	B	C	D	E	F	G	H	I	J
niveau de stress X	3,47	3,03	3,45	7,9	2,86	8,48	3,04	3,03	7,82	2,97
temps de réponse Y	479	502	486	668	490	686	495	471	687	507

1. Déterminer le coefficient de corrélation des rangs de Spearman. Ces variables sont-elles très corrélées ?
2. Déterminer le coefficient de corrélation linéaire de ces deux variables. Ces variables sont-elles très corrélées ?

3. Suite à une perte d'échantillon, on ne connaît plus le niveau de stress d'un rat, dont on a mesuré qu'il met 686ms à réagir aux stimuli visuels. Déterminer environ son niveau de stress.
4. Tracer le nuage de point des variables X et Y.
5. Commenter vos réponses aux questions précédentes en vous appuyant sur le nuage de points.

Exercice 20 : Revenus et espérance de vie

On étudie le revenu par habitant X (en milliers de "dollars internationaux" par an) et l'espérance de vie Y dans 12 pays différents :

pays	revenu/hab. X	espérance de vie Y
Afghanistan	1,9	60,5
Burundi	0,8	59,6
Chili	24,6	80,5
Djibouti	3,6	63,5
Espagne	38,2	82,8
Finlande	44	81,1
Grenade	14,8	73,6
Honduras	5,5	74,6
Iran	20	75,5
Jordanie	12,5	74,1
Koweït	69,7	74,7
Laos	7,4	65,7

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

Exercice 21 : Extraversion

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

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

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

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

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

Exercice 22 : Reconnaissance de forme

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

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

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

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

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

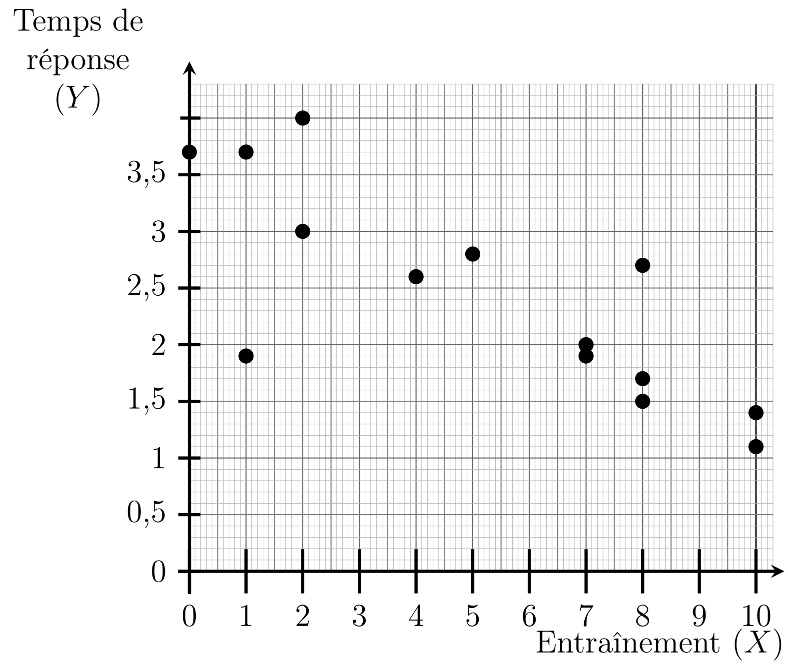
le coefficient de corrélation des rangs de Spearman est donc
$1-\left(6\times\frac{132{,}25+9+100+42{,}25+90{,}25+\cdots+132{,}25}{14\left(14^2-1\right)}\right)\simeq -0{,}766$
Cela montre un lien entre les deux variables, où Y diminue quand X augmente.

2. Calculer les moyennes et écarts type des variables X et Y, puis leur coefficient de corrélation linéaire. Interpréter la valeur de ce coefficient.
moyenne : $m(X)=\frac{\sum x_i}{n}=\frac{0+1+1+\cdots+10}{14}=\frac{73}{14}\simeq 5{,}21$
$m(X^2)=\frac{\sum x_i^2}{n}=\frac{0^2+1^2+1^2+\cdots+10^2}{14}=\frac{541}{14}$
$Var(X)=m(X^2)-m(X)^2=\frac{541}{14}-\left(\frac{73}{14}\right)^2\simeq 11{,}45$
Écart-type : $s(X)=\sqrt{Var(X)}\simeq 3{,}38$

moyenne : $m(Y)=\frac{\sum x_i}{n}=\frac{3{,}7+1{,}9+3{,}7+\cdots+1{,}4}{14}=\frac{34}{14}\simeq 2{,}43$
$m(Y^2)=\frac{\sum x_i^2}{n}=\frac{3{,}7^2+1{,}9^2+3{,}7^2+\cdots+1{,}4^2}{14}=\frac{93{,}8}{14}$
$Var(Y)=m(Y^2)-m(Y)^2=\frac{93{,}8}{14}-\left(\frac{34}{14}\right)^2\simeq 0{,}802$
Écart-type : $s(Y)=\sqrt{Var(Y)}\simeq 0{,}9$

$m(X\,Y)=\frac{\sum x_i\,y_i}{n}=\frac{0\times 3{,}7+1{,}9+\ldots+10\times 1{,}4}{14}=\frac{143{,}5}{14}=10{,}25$
$Cov(X,Y)=m(X\,Y)-m(X)\,m(Y)=\frac{143{,}5}{14}-\frac{73}{14}\frac{34}{14}\simeq -2{,}413$
$r(X,Y)=\frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}=\frac{-2{,}413}{\sqrt{11{,}454\times 0{,}802}}\simeq -0{,}796$
Cela montre un lien linéaire entre les deux variables, où Y diminue quand X augmente.

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



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

4. Si un enfant a déjà fait 6 exercices de reconnaissance de forme, alors à combien estimeriez vous son temps de réponse si on lui fait à nouveau passer ce test ?
Pour répondre à cette question, on utilise la droite $D_{Y|X}$ (qui fait sens car il y a une forte corrélation linéaire) :
on pose $a=\frac{Cov(X,Y)}{Var(X)}\simeq\frac{-2{,}413}{11{,}45}\simeq -0{,}21$ et $b=m(Y)-a\,m(X)\simeq 2{,}429-(-0{,}211)\times 5{,}214\simeq 3{,}53$
D'où l'équation de la droite $D_{Y|X}$: $Y=-0{,}21\,X+3{,}53$
Donc pour $x=6$, on s'attend à $y=-0{,}21\times 6+3{,}53=2{,}27$.

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

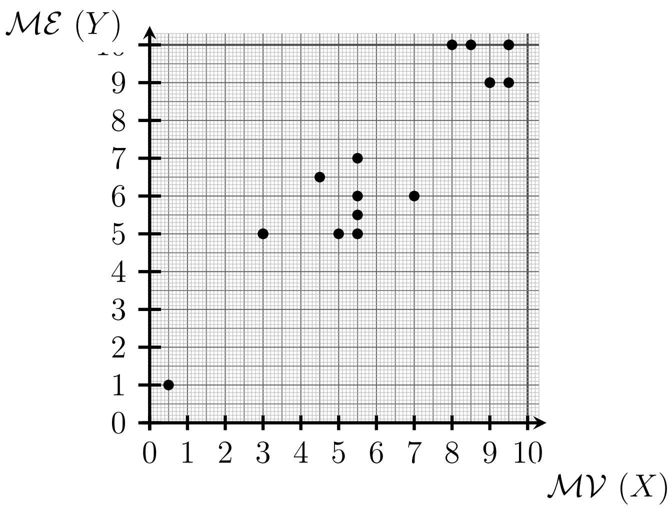
Pour répondre à cette question, on utilise la droite $D_{X|Y}$:
on pose $a' = \frac{Cov(X,Y)}{Var(Y)} \simeq \frac{-2{,}41}{0{,}8} \simeq -3{,}01$ et $b' = m(X) - a'\,m(Y) \simeq 5{,}21 - (-3{,}01) \times 2{,}43 \simeq 12{,}52$
D'où l'équation de la droite $D_{X|Y}$: $X = -3{,}01\,Y + 12{,}52$
Donc pour $y = 1{,}6$, on s'attend à $x = -3{,}01 \times 1{,}6 + 12{,}52 = 7{,}704$.

Exercice 23 : Résolution de problèmes mathématiques

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

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

1. Dessiner le nuage statistique de ces variables.



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

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

le coefficient de corrélation des rangs de Spearman est donc
$1 - \left(6 \times \frac{6{,}25+0{,}0+2{,}25+1{,}0+0{,}25+\cdots+16}{18\left(18^2-1\right)}\right) \simeq 0{,}852$

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

moyenne : $m(X) = \frac{\sum x_i}{n} = \frac{7+0{,}5+9+\cdots+5{,}5}{18} = \frac{113{,}5}{18} \simeq 6{,}31$
$m(X^2) = \frac{\sum x_i^2}{n} = \frac{7^2+0{,}5^2+9^2+\cdots+5{,}5^2}{18} = \frac{832{,}75}{18}$
$Var(X) = m(X^2) - m(X)^2 = \frac{832{,}75}{18} - \left(\frac{113{,}5}{18}\right)^2 \simeq 6{,}5$
Écart-type : $s(X) = \sqrt{Var(X)} \simeq 2{,}55$

moyenne : $m(Y) = \frac{\sum x_i}{n} = \frac{6+1+9+\cdots+5}{18} = \frac{124}{18} \simeq 6{,}89$
$m(Y^2) = \frac{\sum x_i^2}{n} = \frac{6^2+1^2+9^2+\cdots+5^2}{18} = \frac{962{,}5}{18}$
$Var(Y) = m(Y^2) - m(Y)^2 = \frac{962{,}5}{18} - \left(\frac{124}{18}\right)^2 \simeq 6{,}02$
Écart-type : $s(Y) = \sqrt{Var(Y)} \simeq 2{,}45$

$m(X\,Y) = \frac{\sum x_i\,y_i}{n} = \frac{7\times 6+\frac{1}{2}+\ldots+5{,}5\times 5}{18} = \frac{886}{18} \simeq 49{,}222$
$Cov(X,Y) = m(X\,Y) - m(X)\,m(Y) = \frac{886}{18} - \frac{113{,}5}{18}\frac{124}{18} \simeq 5{,}784$
$r(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} = \frac{5{,}784}{\sqrt{6{,}504\times 6{,}015}} \simeq 0{,}925$

4. On souhaite désormais estimer à quelle note s'attendre avec la méthode écrite pour un enfant ayant obtenu une note $x = 3{,}5$ avec la méthode verbale :

   (a) Quelle droite de régression peut-on utiliser pour répondre à cette question ? Donner son équation.
   La droite qui est pertinente pour cette question est $D_{Y|X}$.
   on pose $a = \frac{Cov(X,Y)}{Var(X)} \simeq \frac{5{,}784}{6{,}5} \simeq 0{,}89$ et $b = m(Y) - a\,m(X) \simeq 6{,}889 - 0{,}89 \times 6{,}306 \simeq 1{,}277$

D'où l'équation de la droite $D_{Y|X} : Y = 0{,}89\,X + 1{,}277$

(b) Si un enfant a obtenu une note $x = 3{,}5$, donner une estimation de sa note y.

Donc pour $x = 3{,}5$, on s'attend à $y = 0{,}89 \times 3{,}5 + 1{,}277 = 4{,}392$.

Exercice 24 : Groupe d'étudiants

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

Anais
de très bonne humeur
1 frère/soeur
1m69
bac : 14,6/20

Nicolas
de très bonne humeur
4 frères/soeurs
1m82
bac : 14,26/20

Catherine
de très bonne humeur
0 frère/soeur
1m65
bac : 11,07/20

Olivier
de relativement bonne humeur
2 frères/soeurs
1m74
bac : 15,97/20

Enzo
de très bonne humeur
3 frères/soeurs
1m91
bac : 12,66/20

Loc
de très bonne humeur
2 frères/soeurs
1m74
bac : 11,14/20

Isabelle
de très bonne humeur
1 frère/soeur
1m62
bac : 15,44/20

Bernard
de mauvaise humeur
0 frère/soeur
1m78
bac : 11,89/20

Rmi
de mauvaise humeur
1 frère/soeur
1m81
bac : 13,23/20

Mlanie
de bonne humeur
4 frères/soeurs
1m60
bac : 13,07/20

Corinne
de très bonne humeur
1 frère/soeur
1m69
bac : 11,02/20

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

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

moyenne : $m(X) = \frac{\sum x_i}{n} = \frac{1{,}69+1{,}82+1{,}65+\cdots+1{,}69}{11} = \frac{19{,}05}{11} \simeq 1{,}73$

$m(X^2) = \frac{\sum x_i^2}{n} = \frac{1{,}69^2+1{,}82^2+1{,}65^2+\cdots+1{,}69^2}{11} = \frac{33{,}079\,3}{11}$

$Var(X) = m(X^2) - m(X)^2 = \frac{33{,}079\,3}{11} - \left(\frac{19{,}05}{11}\right)^2 \simeq 0{,}008$

Écart-type : $s(X) = \sqrt{Var(X)} \simeq 0{,}09$

moyenne : $m(Y) = \frac{\sum x_i}{n} = \frac{14{,}6+14{,}26+11{,}07+\cdots+11{,}02}{11} = \frac{144{,}35}{11} \simeq 13{,}12$

$m(Y^2) = \frac{\sum x_i^2}{n} = \frac{14{,}6^2+14{,}26^2+11{,}07^2+\cdots+11{,}02^2}{11} = \frac{1\,925{,}532\,5}{11}$

$Var(Y) = m(Y^2) - m(Y)^2 = \frac{1\,925{,}532\,5}{11} - \left(\frac{144{,}35}{11}\right)^2 \simeq 2{,}84$

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

$m(X\,Y) = \frac{\sum x_i\,y_i}{n} = \frac{1{,}69\times14{,}6+1{,}82\times14{,}26+\ldots+1{,}69\times11{,}02}{11} = \frac{249{,}903\,8}{11} \simeq 22{,}719$

$Cov(X,Y) = m(X\,Y) - m(X)\,m(Y) = \frac{249{,}903\,8}{11} - \frac{19{,}05}{11}\frac{144{,}35}{11} \simeq -0{,}008$

$r(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} = \frac{-0{,}008}{\sqrt{0{,}008\times2{,}842}} \simeq -0{,}053$

Exercice 25 : Thérapie pour réduire l'angoisse

Pour mettre en évidence l'efficacité d'une thérapie visant à réduire l'angoisse de personnes victimes d'agressions, nous avons 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	33	31	39	31	23	24	37	28	23	14	31	37	23	38	19	32
Après	21	14	29	28	13	16	28	21	25	8	24	19	11	38	16	27

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

Calcul du coefficient de corrélation linéaire

moyenne : $m(X) = \frac{\sum x_i}{n} = \frac{33+31+39+\cdots+32}{16} = \frac{463}{16} \simeq 28{,}94$

$m(X^2) = \frac{\sum x_i^2}{n} = \frac{33^2+31^2+39^2+\cdots+32^2}{16} = \frac{14\,203}{16}$

$Var(X) = m(X^2) - m(X)^2 = \frac{14\,203}{16} - \left(\frac{463}{16}\right)^2 \simeq 50{,}31$

Écart-type : $s(X) = \sqrt{Var(X)} \simeq 7{,}09$

moyenne : $m(Y) = \frac{\sum x_i}{n} = \frac{21+14+29+\cdots+27}{16} = \frac{338}{16} \simeq 21{,}12$

$m(Y^2) = \frac{\sum x_i^2}{n} = \frac{21^2+14^2+29^2+\cdots+27^2}{16} = \frac{8\,088}{16}$

$Var(Y) = m(Y^2) - m(Y)^2 = \frac{8\,088}{16} - \left(\frac{338}{16}\right)^2 \simeq 59{,}23$

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

$m(X\,Y) = \frac{\sum x_i\,y_i}{n} = \frac{33\times21+31\times14+\ldots+32\times27}{16} = \frac{10\,432}{16} = 652$

$Cov(X,Y) = m(X\,Y) - m(X)\,m(Y) = \frac{10\,432}{16} - \frac{463}{16}\frac{338}{16} \simeq 40{,}695$

$r(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} = \frac{40{,}695}{\sqrt{50{,}309\times59{,}234}} \simeq 0{,}745$

Coefficient de spearman

sujet	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
X	33	31	39	31	23	24	37	28	23	14	31	37	23	38	19	32
Y	21	14	29	28	13	16	28	21	25	8	24	19	11	38	16	27
rang X'	12	9	16	9	4	6	13,5	7	4	1,0	9	13,5	4	15	2	11
rang Y'	8,5	4	15	13,5	3	5,5	13,5	8,5	11	1,0	10	7	2	16	5,5	12
$(X'-Y')^2$	12,25	25	1,0	20,25	1,0	0,25	0,0	2,25	49	0,0	1,0	42,25	4	1,0	12,25	1,0

le coefficient de corrélation des rangs de Spearman est donc $1 - \left(6 \times \frac{12{,}25+25+1{,}0+20{,}25+1{,}0+\cdots+1{,}0}{16\left(16^2-1\right)}\right) \simeq 0{,}746$

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

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

on pose $a = \frac{Cov(X,Y)}{Var(X)} \simeq \frac{40{,}695}{50{,}31} \simeq 0{,}809$ et $b = m(Y) - a\,m(X) \simeq 21{,}125 - 0{,}809 \times 28{,}938 \simeq -2{,}286$

D'où l'équation de la droite $D_{Y|X}$: $Y = 0{,}809\,X - 2{,}286$

Chapitre 3 : probabilités

Exercice 26 : Emploi

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

Nom	Mme Bernard	Mr Martin	Mme Allard	Mr Roche	Mme Lemoine
Situation	chômage	retraite	retraite	emploi	emploi

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

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

Exercice 27 : Factorielle, coefficients binomiaux

1. Simplifier les quantités $\dfrac{17!}{15!}$, $\dfrac{18!}{16! \times 2!}$.
2. Écrire à l'aide de deux factorielles l'expression ci-contre : $6 \times 7 \times 8$.
3. Calculer les nombres suivants : $\binom{18}{2}$, $\binom{12}{3}$, $\binom{12}{9}$, $\binom{9}{0}$, $\binom{11}{1}$.

Exercice 28 : Croyance religieuse

1. Cyril et Valérie 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.
   Argumenter que ces probabilités correspondent à une loi binômiale, 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.
   Calculer 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.
   Calculer la loi de X.
4. Comparer entre elles les lois obtenues dans ces trois situations.

Exercice 29 : Loi binomiale

On considère une loi binômiale $X \sim \mathcal{B}(n,p)$ de moyenne $m(X) = 4{,}8$ et de variance $\mathrm{Var}(X) = 2{,}88$.

1) Calculer n et p.
2) Calculer les probabilités $\mathbb{P}[4 \leqslant X \leqslant 6]$ et $\mathbb{P}[X > 9]$.
3) Trouver toutes les valeurs de k telles que $\mathbb{P}[X \geqslant k] \leqslant 0{,}05$.

Exercice 30 : Troubles de l'humeur et emploi

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

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) Lister les neuf possibilités pour ces deux noms choisis au hasard.
   En notant par exemple "AB" pour "Alice puis Bernard", les neuf possibilités sont :
   AA, AB, AC, BA, BB, BC, CA, CB et CC.
   (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 $\frac{4}{9} \simeq 0{,}444$.
   (d) Pour un tel choix aléatoire de deux noms, on note X le nombre de noms choisis qui correspondent à une personne exerçant un emploi. Quelle est la loi de variable X ?On réalise deux fois de suite l'expérience "choisir un des trois noms et constater si c'est une personne qui exerce un emploi", de manière indépendante. À chaque fois il y a deux chances sur trois de choisir quelqu'un qui exercice un emploi, donc X suit la loi $\mathcal{B}\left(2,\frac{2}{3}\right)$.
   (e) Retrouver le résultat de la question c), en utilisant cette loi.
   $\mathbb{P}[X=1] = \begin{pmatrix}2\\1\end{pmatrix} (2/3)^1 (1/3)^1 \simeq 0{,}444$.
   (f) Déterminer de même $\mathbb{P}[X=0]$ et $\mathbb{P}[X=2]$.
   - $\mathbb{P}[X=0] = \binom{2}{0} (2/3)^0 (1/3)^2 \simeq 0{,}111$.
   - $\mathbb{P}[X=2] = \binom{2}{2} (2/3)^2 (1/3)^0 \simeq 0{,}444$.
2. On décide d'une autre façon de choisir deux noms : on choisit un premier nom au hasard, puis on choisit le second nom en demandant qu'ils soit différent du premier. Reprendre, avec cette nouvelle façon de choisir les deux noms, les question 1a, 1b et 1c. On prendra garde qu'en question 1a le nombre de cas ne sera plus égal à neuf.
   (a) AB, AC, BA, BC, CA et CB.
   (b) Il y en a 2 : AB et BA.
   (c) Cela correspond aux 4 possibilités AC, BC, CA et CB.
   Donc à la proportion $\frac{4}{6} \simeq 0{,}667$.

Exercice 31 : Collection de dessins

Mme Chauvet 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 dans quelle ordre elles se retrouvent à l'issue du mélange. Quelle était la probabilité qu'à l'issue du mélange, elles se retrouvent précisément dans cet ordre ?
   Il y a $5! = 5 \times 4 \times 3 \times 2 = 120$ résultats possibles à l'issue du mélange. La probabilité est donc $\frac{1}{120} \simeq 0{,}008$.
2. Si elle remélange les cartes, quelle est la probabilité qu'après le mélange
   (a) la carte qui était en haut du tas soit à nouveau en haut du tas ?
   Il y a $4! = 4 \times 3 \times 2 = 24$ mélanges qui conservent la première carte (car cela correspond à ne mélanger que les 4 autres cartes). La probabilité est donc $\frac{24}{120} = 0{,}2$.

(b) les deux cartes cartes qui étaient en haut du tas soient à nouveau en haut du tas, et dans le même ordre ?
Il y a $3! = 3 \times 2 \times 1 = 6$ mélanges qui conservent les deux premières 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 qu'avant ?
$\frac{1}{120} \simeq 0{,}008$.

3. Répondre au mêmes questions en supposant qu'il y ait cette fois ci 35 cartes.
   1. $\frac{1}{35!} = \frac{1}{35\times34\times\ldots\times1} \simeq 9{,}7 \times 10^{-41}$
   2. (a) $\frac{34!}{35!} = \frac{34\times33\times\ldots\times1}{35\times34\times\ldots\times1} = \frac{1}{35} \simeq 0{,}029$
      (b) $\frac{33!}{35!} = \frac{33\times32\times\ldots\times1}{35\times34\times\ldots\times1} = \frac{1}{35\times34} \simeq 0{,}000\,8$
      (c) $\frac{1}{35\times34} \simeq 0{,}000\,8$
      (d) $\frac{1}{35\times34} + \frac{1}{35\times34} \simeq 0{,}001\,7$
      (e) $\frac{1}{35!} = \frac{1}{35\times34\times\ldots\times1} \simeq 9{,}7 \times 10^{-41}$
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. Le résultat sur lequel on tombe sera forcément un parmi des milliards (de milliards), il avait extrêmement peu de chances de tomber sur celui-là en particulier, mais il faut tomber sur l'un de ces résultats.

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

Exercice 32 : Ressources humaines

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

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

1. Calculer la probabilité que l'échantillon contienne exactement 3 employés notés A.
   Notons X le nombre de personnes notées A dans l'échantillon. Alors $X \sim \mathcal{B}(25; 0{,}2)$, car on a choisi 25 employés (avec remise) parmi une population dont 20% ont la note A.
   $$\begin{aligned}\mathbb{P}[X=3] &= \binom{25}{3}(0{,}2)^3(1-0{,}2)^{25-3} \\ &\simeq 0{,}1358\end{aligned}$$
2. Quelle est la probabilité que l'échantillon contienne au moins 3 individus notés A ?
   $\mathbb{P}[X \geqslant 3] = 1 - \mathbb{P}[X \leqslant 2] \simeq 1 - 0{,}098 \simeq 0{,}902$

3. Quelle probabilité y a-t-il que l'échantillon contienne au moins 12 individus notés A ou B ?
   On procède de même en notant cette fois ci Y le nombre d'individus notés A ou B dans l'échantillon. On a $Y \sim \mathcal{B}(25; 0{,}5)$.
   $\mathbb{P}[Y \geqslant 12] = 1 - \mathbb{P}[Y \leqslant 11] \simeq 1 - 0{,}35 \simeq 0{,}65$
   Remarque : le moyen le plus rapide d'obtenir le résultat est de calculer $\mathbb{P}[Y \leqslant 11] \simeq 0{,}345$ avec la calculette (voir explications dans le formulaire), puis de déduire $\mathbb{P}[Y \geqslant 12] = 1 - \mathbb{P}[Y \leqslant 11] \simeq 0{,}655$.
4. Calculer la probabilité que l'échantillon contienne moins que 8 personnes notées C.
   On procède de même en notant cette fois ci Z le nombre d'individus notés C dans l'échantillon. On a $Z \sim \mathcal{B}(25; 0{,}5)$.
   $\mathbb{P}[Z < 8] \simeq 0{,}021\,6$
5. Quel est le nombre moyen d'individus notés B au sein d'un échantillon aléatoire de 25 employés ?
   Si on note W le nombre de personnes notées B, alors $W \sim \mathcal{B}(25; 0{,}3)$, donc $m(W) = 25 \times 0{,}3 = 7{,}5$.

Exercice 33 : Thérapie contre la dépression

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

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

1) S'il y avait 70% de rémission parmi l'ensemble des patients, quelle loi suivrait le nombre de rémissions au sein d'un échantillon de 20 patients choisis au hasard avec remise ? Quel serait le nombre moyen de rémissions au sein d'un tel échantillon ?
   Si l'on désigne par X le nombre de rémissions dans un échantillon aléatoire de 20 patients, et si l'on suppose qu'il y ait 70% de rémission parmi l'ensemble des patients, alors X suit la loi $\mathcal{B}(20; 0{,}7)$.
   Le nombre moyen de rémissions au sein d'un tel échantillon serait alors $m(X) = 20 \times 0{,}7 = 14$.
2) Sous cette hypothèse, quelle serait la probabilité d'avoir au maximum 8 rémissions ?
   $\mathbb{P}[X \leqslant 8] \simeq 0{,}005\,1$
   Remarque : Encore une fois, la calculette permet de le calculer directement (cf formulaire).
3) Conclure : vous semble t-il vraisemblable qu'il y ait, comme l'affirme ce psychologue, 70% de rémissions parmi l'ensemble des patients ?
   On a obtenu que s'il y avait bien 70% de rémission parmi l'ensemble des patients (comme l'affirme ce psychologue), alors il y aurait très peu de chances (environ 1% de chances) d'avoir un échantillon où il n'y ait pas plus de 8 rémissions. Le fait que nous soyons tombé sur un tel échantillon donne une forte présomption que le taux de rémissions n'est en fait pas de 70% (à moins par exemple que notre échantillon soit biaisé à cause de la façon dont il a été établi).

Exercice 34 : Nombre de filles et de garçons

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

1. Quel est le nombre de choix possibles ?
   On choisit 4 personnes parmi l'ensemble des $16 + 12 = 28$ élèves. Le nombre de choix possibles est donc
   $$\binom{28}{4} = \frac{28!}{4!\times 24!} = \frac{28\times 27\times\ldots\times 1}{(4\times 3\times 2)\times(24\times 23\times\ldots\times 1)} = (28\times 27\times 26\times 25)\times\frac{1}{4\times 3\times 2} = \frac{4\times 7\times 27\times 26\times 25}{4\times 3\times 2} = \frac{7\times 27\times 26\times 25}{3\times 2} = \frac{7\times 3\times 9\times 26\times 25}{3\times 2} = \frac{7\times 9\times 26\times 25}{2} = \frac{7\times 9\times 2\times 13\times 25}{2} = 7\times 9\times 13\times 25 = 20475$$
2. Quel est le nombre de choix ne comportant que des garçons ?
   Ces choix correspondent à choisir 4 élèves parmi les 12 garçons. Le nombre de choix correspondant est donc
   $$\binom{12}{4} = \frac{12!}{4!\times 8!} = \frac{12\times 11\times\ldots\times 1}{(4\times 3\times 2)\times(8\times 7\times\ldots\times 1)} = (12\times 11\times 10\times 9)\times\frac{1}{4\times 3\times 2} = \frac{4\times 3\times 11\times 10\times 9}{4\times 3\times 2} = \frac{11\times 10\times 9}{2} = \frac{11\times 2\times 5\times 9}{2} = 11\times 5\times 9 = 495$$

3. Quelle est la proportion de choix ne comportant que des filles ?
   Ces choix correspondent à choisir 4 élèves parmi les 16 filles. Le nombre de choix correspondant est donc
   $$\begin{aligned}\binom{16}{4} &= \frac{16!}{4!\times 12!} = \frac{16\times 15\times\ldots\times 1}{(4\times 3\times 2)\times(12\times 11\times\ldots\times 1)}\\ &= (16\times 15\times 14\times 13)\times\frac{1}{4\times 3\times 2} = \frac{4\times 4\times 15\times 14\times 13}{4\times 3\times 2}\\ &= \frac{4\times 15\times 14\times 13}{3\times 2} = \frac{4\times 3\times 5\times 14\times 13}{3\times 2}\\ &= \frac{4\times 5\times 14\times 13}{2} = \frac{2\times 2\times 5\times 14\times 13}{2}\\ &= 2\times 5\times 14\times 13 = 1820\end{aligned}$$
   Cela correspond donc à la probabilité $\frac{1\,820}{20\,475} \simeq 0{,}089$.
4. Quelle est la probabilité, en choisissant ainsi un échantillon de 4 élèves au hasard, que cet échantillon comporte 3 filles et 1 garçon ?

   Les cas correspondant sont formés d'une part d'un groupe de 3 filles (pour lequel il y a $\binom{16}{3}$ choix possibles), et pour chaque choix de ce groupe de filles, on peut y ajouter n'importe lequel des garçons (pour lequel il y a $\binom{14}{1}$ choix possibles). En conséquence, le nombre de cas qui correspondent est :
   $$\begin{aligned}\binom{16}{3}\times\binom{12}{1} &= \frac{16!}{3!\times 13!}\times\frac{12!}{1!\times 11!}\\ &= \frac{16\times 15\times\ldots\times 1}{(3\times 2\times 1)\times(13\times 12\times\ldots\times 1)}\times\frac{12\times 11\times\ldots\times 1}{11\times 10\times\ldots\times 1}\\ &= (16\times 15\times 14)\times\frac{1}{3\times 2\times 1}\times 12\\ &= \frac{16\times 3\times 5\times 14}{3\times 2}\times 12\\ &= \frac{16\times 5\times 14}{2}\times 12\\ &= \frac{2\times 8\times 5\times 14}{2}\times 12\\ &= 8\times 5\times 14\times 12\\ &= 560\times 12\\ &= 6720\end{aligned}$$
   D'où la probabilité $\frac{6\,720}{20\,475} \simeq 0{,}328$.
5. Quel est le nombre de choix comportant au plus un garçon ?
   On additionne, d'une part le nombre de choix comportant 0 garçons (calculé en question 3) et d'autre part le nombre de choix comportant exactement 1 garçon (calculé en question 4). On obtient donc $1\,820 + 6\,720 = 8\,540$.

Chapitre 4 : Loi Normale

Exercice 35 : Loi normale centrée réduite

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

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

1. Pour cette question, on pose $z_1 = 1{,}05$; $z_2 = 1{,}92$ et $z_3 = 0{,}588$.
   (a) Déterminer $F(z_1)$, $F(z_2)$ et $F(z_3)$. Chacune de ces probabilités peut être obtenue à partir de la table du formulaire, ou directement en utilisant les fonctions spécifiques de la calculatrice. Assurez-vous de savoir utiliser chacune de ces deux méthodes.
   (b) Calculer : $\mathbb{P}[z_1 < Z \leqslant z_2]$, $\mathbb{P}[-z_2 < Z < -z_1]$, $\mathbb{P}[z_1 \leqslant Z < 5]$, $\mathbb{P}[-z_2 \leqslant Z \leqslant z_1]$, $\mathbb{P}[-z_1 < Z < z_1]$, $\mathbb{P}[Z > z_2]$ et $\mathbb{P}[Z < z_1]$. On peut calculer ces probabilités à partir des valeurs obtenues dans la question a) ou en utilisant directement les fonctions spécifiques de la calculatrice. Assurez-vous de savoir utiliser chacune de ces deux méthodes.
2. (a) Déterminer des valeurs approchées de z_4, z_5, et z_6 tels que $F(z_4) = 0{,}7967$; $F(z_5) = 0{,}0256$ et $F(z_6) = 0{,}8782$.
   (b) Déterminer des valeurs approchées de z_7, z_8, et z_9 tels que $\mathbb{P}[Z \leqslant z_7] = 0{,}99$; $\mathbb{P}[Z > z_8] = 0{,}9515$; $\mathbb{P}[Z < z_9] = 0{,}9345$. On peut le faire à partir de la table du formulaire, ou bien obtenir une réponse directe de la calculette. Assurez-vous de savoir utiliser chacune de ces deux méthodes.

Exercice 36 : Images mentales

Dans une expérience sur les images mentales, on demande à un échantillon de 337 enfants d'apprendre une liste de 40 mots en leur montrant, pour chaque mot, un dessin de l'objet correspondant au mot. On désigne par X le nombre dont se souvient chaque enfant un jour après. On obtient :

Nb de mots X	$[5\,;10[$	$[10\,;15[$	$[15\,;20[$	$[20\,;25[$	$[25\,;30[$	$[30\,;35[$	$[35\,;40[$
Effectif	7	70	114	115	28	1	2

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

Exercice 37 : Désirabilité sociale

On étudie les scores (notés X) obtenus par des enfants sur une échelle de désirabilité sociale (D.S.). Une étude a montré que pour cette échelle, les scores ont, sur l'ensemble des enfants de CM2, une moyenne $\mu \simeq 10{,}42$ et un écart-type $\sigma \simeq 2{,}18$.

Dans la suite, on effectue des calculs en supposant que X suit la loi normale $\mathcal{N}(10{,}42; 2{,}18)$.

1. Calculer la proportion théorique des scores supérieurs à 10,97.
2. Trouver la valeur de a telle que $\mathbb{P}[X < a] \simeq 0{,}95$.
3. Trouver les quartiles de la variable X.
4. Quel est le score minimal des 20 % de ceux qui ont les scores D.S. les plus élevés ?
5. Quel est le score maximal des 35 % de ceux qui ont les scores D.S. les moins élevés ?

Exercice 38 :

Lors d'un concours le pourcentage d'échecs est de 55%. 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 $n = 16$. Calculer $p_k = \mathbb{P}[S_{16} \leqslant k]$ pour $k = 0; 1; 2; 3$.
2. Pour $n = 100$, on souhaite calculer $\mathbb{P}[50 \leqslant S_{100} \leqslant 60]$:
   a) Calculer $\mathbb{P}[50 \leqslant S_{100} \leqslant 60]$ avec la calculette.
   b) Argumenter que l'on peut approcher la loi de S_{100} par une loi normale, et calculer $\mathbb{P}[50 \leqslant S_{100} \leqslant 60]$ de manière approchée.
   On vérifiera que la précision de l'approximation est bien meilleure si on fait la *correction de continuité*.
3. Pour $n = 125$, déterminer de même $\mathbb{P}[59 \leqslant S_{125} \leqslant 79]$ (on pourra utiliser l'approximation par une loi normale en faisant une correction de continuité).
4. Pour $n = 125$, trouver une valeur de a telle que $\mathbb{P}[S_{125} < a] \simeq 0{,}03$ (pour simplifier les calculs, on pourra omettre la correction de continuité dans cette question).

Exercice 39 : Indice de masse corporelle dans la population française

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

X< 16,5 dénutrition ou famine
16,5 ⩽ X< 18,5 maigreur
18,5 ⩽ X< 25 corpulence normale
25 ⩽ X< 30 surpoids
30 ⩽ X< 35 obésité modérée
35 ⩽ X< 40 obésité sévère
X ⩾40 obésité morbide

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

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

1. Déterminer la probabilité (« *fréquence théorique* ») qu'aurait chaque classe si X suivait la loi normale $\mathcal{N}(25; 5)$.
   Si X suit le loi $\mathcal{N}(25; 5)$, on obtient les probabilités suivantes :

- $\mathbb{P}[X \leqslant 18{,}5] = \mathbb{P}[\frac{X-25}{5} < \frac{18{,}5-25}{5}] \simeq \mathbb{P}[Z < -1{,}3]$
$\simeq F(-1{,}3) \simeq 1 - 0{,}903\,2$
$\simeq 0{,}0968$
- $\mathbb{P}[18{,}5 \leqslant X \leqslant 25] = \mathbb{P}[\frac{18{,}5-25}{5} < \frac{X-25}{5} < \frac{25-25}{5}] = \mathbb{P}[-1{,}3 < Z < 0{,}0]$
$= F(0{,}0) - F(-1{,}3) \simeq 0{,}5000 - (1 - 0{,}903\,2)$
$\simeq 0{,}5 - 0{,}096\,8 \simeq 0{,}4032$
- $\mathbb{P}[25 \leqslant X \leqslant 30] = \mathbb{P}[\frac{25-25}{5} < \frac{X-25}{5} < \frac{30-25}{5}] = \mathbb{P}[0{,}0 < Z < 1{,}0]$
$= F(1{,}0) - F(0{,}0) \simeq 0{,}841\,3 - 0{,}5$
$\simeq 0{,}3413$
- $\mathbb{P}[30 \leqslant X \leqslant 35] = \mathbb{P}[\frac{30-25}{5} < \frac{X-25}{5} < \frac{35-25}{5}] = \mathbb{P}[1{,}0 < Z < 2{,}0]$
$= F(2) - F(1{,}0) \simeq 0{,}977\,2 - 0{,}841\,3$
$\simeq 0{,}1359$
- $\mathbb{P}[35 \leqslant X \leqslant 40] = \mathbb{P}[\frac{35-25}{5} < \frac{X-25}{5} < \frac{40-25}{5}] = \mathbb{P}[2{,}0 < Z < 3{,}0]$
$= F(3) - F(2) \simeq 0{,}998\,7 - 0{,}977\,2$
$\simeq 0{,}0215$
- $\mathbb{P}[X \geqslant 40] = \mathbb{P}[\frac{40-25}{5} < \frac{X-25}{5}] \simeq \mathbb{P}[3{,}0 < Z]$
$\simeq 1 - F(3) \simeq 1{,}0 - 0{,}998\,7$
$\simeq 0{,}0013$

D'où les fréquences théoriques suivantes :

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

2. En particulier, si X suit la loi normale $\mathcal{N}(25; 5)$, quelle est la proportion (*théorique*) d'obésité sévère ou morbide (c'est à dire $\mathbb{P}[X \geqslant 35]$) ?
En particulier la proportion théorique d'obésité sévère ou morbide serait la somme des fréquences théoriques des deux dernières classes, c'est à dire $\mathbb{P}[X \geqslant 35] \simeq 0{,}021\,5 + 0{,}001\,3 \simeq 0{,}023$.
3. Au contraire, quelle est dans l'échantillon de cette étude épidémiologique, la proportion d'obésité sévère ou morbide (c'est à dire $\mathbb{P}_r[X \geqslant 35]$) ?
La proportion d'obésité sévère ou morbide est la somme des fréquences des deux dernières classes, c'est à dire $3.1 + 1.2 \simeq 4.3\%$.
4. On cherche désormais à comparer les résultats des questions 2 et 3 : on souhaite déterminer si la proportion $\mathbb{P}_r[X \geqslant 35]$ est compatible avec la probabilité $\mathbb{P}[X \geqslant 35]$ obtenue sous l'hypothèse d'une loi normale.
   (a) Si on suppose qu'une personne choisie au hasard a une probabilité de 2,3% d'être en obésité sévère ou morbide, quelle est la loi du nombre de personnes en obésité sévère ou morbide dans un échantillon de 25 714 français ? Le fait que l'échantillon soit *avec* ou *sans remise* a-t-il la moindre importance ?
   On choisit un échantillon aléatoire de taille 25 714 français parmi une population consitiuée 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 $\mathcal{B}(25714; 0{,}023)$.
   (b) Cette loi peut-elle être approximée par une loi normale, et si oui laquelle ?
   On a $n = 25714 > 30$, et $n\,p \simeq 591 > 5$, et $n(1-p) \simeq 25120 > 5$ donc on peut approximer la loi de ce nombre de personne par $\mathcal{N}(591; 24)$ où $24 \simeq \sqrt{n\,p(1-p)}$
   (c) En utilisant cette approximation par une loi normale, calculer la probabilité qu'un échantillon aléatoire de 25 714 personnes contienne au moins 4,3% de personnes en obésité sévère ou morbide.
   On note X le nombre de personnes en obésité sévère ou morbide dans un échantillon aléatoire de 25714 français. Une proportion de 4,3% des personnes correspondrait à $25\,714 \times 0{,}043 \simeq 1\,105{,}7$ personnes, donc on calcule $\mathbb{P}[X \geqslant 1106]$:
   $\mathbb{P}[X \geqslant 1106] = \mathbb{P}[X \geqslant 1105{,}5] = \mathbb{P}[\frac{1105{,}5-591{,}42}{24{,}038} < \frac{X-591{,}42}{24{,}038}]$
   $\simeq \mathbb{P}[21{,}3861 < Z] \simeq 1 - F(21{,}3861)$
   $\simeq 1{,}0 - 1 \simeq 0$
   (d) Comparer avec l'échantillon de l'étude épidémiologique mentionnée en début d'exercice.

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

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

Exercice 40 : Loi normale avec moyenne et écart type

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

   (a) supérieure ou égale à 1,5 ?
   $$\begin{aligned}\mathbb{P}[Z \geqslant 1{,}5] &\simeq \mathbb{P}[1{,}5 < Z] \simeq 1 - F(1{,}5)\\ &\simeq 1{,}0 - 0{,}933\,2 \simeq 0{,}0668\end{aligned}$$

   (b) comprise entre -1 et 1 ?
   $$\begin{aligned}\mathbb{P}[-1 \leqslant Z \leqslant 1] &= F(1{,}0) - F(-1) \simeq 0{,}8413 - (1 - 0{,}841\,3)\\ &\simeq 0{,}841\,3 - 0{,}158\,7 \simeq 0{,}6826\end{aligned}$$

   (c) comprise entre -2 et 2 ?
   $$\begin{aligned}\mathbb{P}[-2 \leqslant Z \leqslant 2] &= F(2) - F(-2) \simeq 0{,}9772 - (1 - 0{,}977\,2)\\ &\simeq 0{,}977\,2 - 0{,}022\,8 \simeq 0{,}9544\end{aligned}$$

   (d) comprise entre -3 et 3 ?
   $$\begin{aligned}\mathbb{P}[-3 \leqslant Z \leqslant 3] &= F(3) - F(-3) \simeq 0{,}9987 - (1 - 0{,}998\,7)\\ &\simeq 0{,}998\,7 - 0{,}001\,3 \simeq 0{,}9974\end{aligned}$$

2. Pour une loi normale de moyenne $\mu = 123$ et d'écart-type $\sigma = 9$, (c'est-à-dire $X \rightsquigarrow \mathcal{N}(123;9)$), quelle est la probabilité qu'une observation soit :

   (a) supérieure ou égale à $\mu + 1{,}5\,\sigma$?
   On note tout d'abord que $\mu + 1{,}5\,\sigma = 123 + 1{,}5 \times 9 = 136{,}5$.
   $$\begin{aligned}\mathbb{P}[X \geqslant 136{,}5] &= \mathbb{P}[\tfrac{136{,}5-123}{9} < \tfrac{X-123}{9}] \simeq \mathbb{P}[1{,}5 < Z]\\ &\simeq 1 - F(1{,}5) \simeq 1{,}0 - 0{,}933\,2\\ &\simeq 0{,}0668\end{aligned}$$

   (b) comprise entre $\mu - \sigma$ et $\mu + \sigma$?
   On note tout d'abord que $\mu - \sigma = 123 - 9 = 114$ et $\mu + \sigma = 123 + 9 = 132$.
   $$\begin{aligned}\mathbb{P}[114 \leqslant X \leqslant 132] &= \mathbb{P}[\tfrac{114-123}{9} < \tfrac{X-123}{9} < \tfrac{132-123}{9}]\\ &= \mathbb{P}[-1{,}0 < Z < 1{,}0]\\ &= F(1{,}0) - F(-1)\\ &\simeq 0{,}8413 - (1 - 0{,}841\,3)\\ &\simeq 0{,}841\,3 - 0{,}158\,7\\ &\simeq 0{,}6826\end{aligned}$$

   (c) comprise entre $\mu - 2\,\sigma$ et $\mu + 2\,\sigma$?
   On note tout d'abord que $\mu - 2\,\sigma = 123 - 2 \times 9 = 105$ et $\mu + 2\,\sigma = 123 + 2 \times 9 = 141$.
   $$\begin{aligned}\mathbb{P}[105 \leqslant X \leqslant 141] &= \mathbb{P}[\tfrac{105-123}{9} < \tfrac{X-123}{9} < \tfrac{141-123}{9}]\\ &= \mathbb{P}[-2{,}0 < Z < 2{,}0]\\ &= F(2) - F(-2)\\ &\simeq 0{,}9772 - (1 - 0{,}977\,2)\\ &\simeq 0{,}977\,2 - 0{,}022\,8\\ &\simeq 0{,}9544\end{aligned}$$

   (d) comprise entre $\mu - 3\,\sigma$ et $\mu + 3\,\sigma$?
   On note tout d'abord que $\mu - 3\,\sigma = 123 - 3 \times 9 = 96$ et $\mu + 3\,\sigma = 123 + 3 \times 9 = 150$.
   $$\begin{aligned}\mathbb{P}[96 \leqslant X \leqslant 150] &= \mathbb{P}[\tfrac{96-123}{9} < \tfrac{X-123}{9} < \tfrac{150-123}{9}]\\ &= \mathbb{P}[-3{,}0 < Z < 3{,}0]\\ &= F(3) - F(-3)\\ &\simeq 0{,}9987 - (1 - 0{,}998\,7)\\ &\simeq 0{,}998\,7 - 0{,}001\,3\\ &\simeq 0{,}9974\end{aligned}$$

3. Reprendre les questions précédentes avec la loi normale

$\mathcal{N}(12{,}53; 1{,}34)$. Que remarque-t-on ?

(a) On note tout d'abord que $\mu + 1{,}5\,\sigma = 12{,}53 + 1{,}5 \times 1{,}34 = 14{,}54$.

$$\begin{aligned}\mathbb{P}[X \geqslant 14{,}54] &= \mathbb{P}[\tfrac{14{,}54-12{,}53}{1{,}34} < \tfrac{X-12{,}53}{1{,}34}] \simeq \mathbb{P}[1{,}5 < Z]\\ &\simeq 1 - F(1{,}5) \simeq 1{,}0 - 0{,}933\,2\\ &\simeq 0{,}0668\end{aligned}$$

(b) On note tout d'abord que $\mu - \sigma = 12{,}53 - 1{,}34 = 11{,}19$ et $\mu + \sigma = 12{,}53 + 1{,}34 = 13{,}87$.

$$\begin{aligned}\mathbb{P}[11{,}19 \leqslant X \leqslant 13{,}87] &= \mathbb{P}[\tfrac{11{,}19-12{,}53}{1{,}34} < \tfrac{X-12{,}53}{1{,}34} < \tfrac{13{,}87-12{,}53}{1{,}34}]\\ &= \mathbb{P}[-1{,}0 < Z < 1{,}0]\\ &= F(1{,}0) - F(-1)\\ &\simeq 0{,}8413 - (1 - 0{,}841\,3)\\ &\simeq 0{,}841\,3 - 0{,}158\,7\\ &\simeq 0{,}6826\end{aligned}$$

(c) On note tout d'abord que $\mu - 2\,\sigma = 12{,}53 - 2 \times 1{,}34 = 9{,}85$ et $\mu + 2\,\sigma = 12{,}53 + 2 \times 1{,}34 = 15{,}21$.

$$\begin{aligned}\mathbb{P}[9{,}85 \leqslant X \leqslant 15{,}21] &= \mathbb{P}[\tfrac{9{,}85-12{,}53}{1{,}34} < \tfrac{X-12{,}53}{1{,}34} < \tfrac{15{,}21-12{,}53}{1{,}34}]\\ &= \mathbb{P}[-2{,}0 < Z < 2{,}0]\\ &= F(2) - F(-2)\\ &\simeq 0{,}9772 - (1 - 0{,}977\,2)\\ &\simeq 0{,}977\,2 - 0{,}022\,8\\ &\simeq 0{,}9544\end{aligned}$$

(d) On note tout d'abord que $\mu - 3\,\sigma = 12{,}53 - 3 \times 1{,}34 \simeq 8{,}51$ et $\mu + 3\,\sigma = 12{,}53 + 3 \times 1{,}34 = 16{,}55$.

$$\begin{aligned}\mathbb{P}[8{,}51 \leqslant X \leqslant 16{,}55] &= \mathbb{P}[\tfrac{8{,}51-12{,}53}{1{,}34} < \tfrac{X-12{,}53}{1{,}34} < \tfrac{16{,}55-12{,}53}{1{,}34}]\\ &\simeq \mathbb{P}[-3{,}0 < Z < 3{,}0]\\ &\simeq F(3) - F(-3)\\ &\simeq 0{,}9987 - (1 - 0{,}998\,7)\\ &\simeq 0{,}998\,7 - 0{,}001\,3\\ &\simeq 0{,}9974\end{aligned}$$

On remarque que ces résultats sont indépendants de σ et μ.

Exercice 41 :

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

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

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

$$\begin{aligned}\mathbb{P}[X \geqslant 70] &= \mathbb{P}[\tfrac{70-50}{10} < \tfrac{X-50}{10}] \simeq \mathbb{P}[2{,}0 < Z]\\ &\simeq 1 - F(2) \simeq 1{,}0 - 0{,}977\,2 \simeq 0{,}0228\end{aligned}$$

Donc parmi un échantillon de 50000 personnes, on s'attend à trouver environ $50\,000 \times 0{,}022\,8 = 1\,140$ personnes pathologiquement dépressives.

Exercice 42 :

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

1. On prend $n = 28$. Préciser la loi de probabilité de X et calculer la probabilité : $\mathbb{P}[15 \leqslant X \leqslant 18]$.
   $X \sim \mathcal{B}(28; 0{,}6)$.
   $\mathbb{P}[15 \leqslant X \leqslant 18] = \mathbb{P}[X \leqslant 18] - \mathbb{P}[X \leqslant 14] \simeq 0{,}741 - 0{,}187 \simeq 0{,}554$

2. On prend $n = 280$.

(a) On peut approcher la loi de X par une loi normale. Justifier et préciser laquelle. Dans la suite, si vous utilisez cette approximation, il faudra faire une correction de continuité.
$X \sim \mathcal{B}(280; 0{,}6)$.
On a $n = 280 > 30$, et $n\,p = 168 > 5$, et $n(1-p) = 112 > 5$ donc on peut approximer X par $\mathcal{N}(168; 8{,}2)$ où $8{,}2 \simeq \sqrt{n\,p(1-p)}$

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

$$\begin{aligned}\mathbb{P}[X \geqslant 180] &= \mathbb{P}[X \geqslant 179{,}5] = \mathbb{P}[\tfrac{179{,}5-168}{8{,}2} < \tfrac{X-168}{8{,}2}] \\ &\simeq \mathbb{P}[1{,}4 < Z] \simeq 1 - F(1{,}4) \\ &\simeq 1{,}0 - 0{,}919\,2 \simeq 0{,}08\end{aligned}$$

$$\begin{aligned}\mathbb{P}[150 \leqslant X \leqslant 185] &= \mathbb{P}[149{,}5 \leqslant X \leqslant 185{,}5] \\ &= \mathbb{P}[\tfrac{149{,}5-168}{8{,}2} < \tfrac{X-168}{8{,}2} < \tfrac{185{,}5-168}{8{,}2}] \\ &\simeq \mathbb{P}[-2{,}26 < Z < 2{,}13] \\ &\simeq F(2{,}13) - F(-2{,}26) \\ &\simeq 0{,}9834 - (1 - 0{,}988\,1) \\ &\simeq 0{,}983\,4 - 0{,}011\,9 \\ &\simeq 0{,}97\end{aligned}$$

Exercice 43 : Pile ou face Vous jouez à « pile » ou « face » avec une pièce de monnaie équilibrée. Avec quelle probabilité la fréquence de « pile » est-elle comprise entre 0,48 et 0,52 si vous lancez la pièce :

i) 100 fois.
Si on note X le nombre de pile, on a $X \sim \mathcal{B}(100; 0{,}5)$.
On a $n = 100 > 30$, et $n\,p = 50 > 5$, et $n(1-p) = 50 > 5$ donc on peut approximer X par $\mathcal{N}(50; 5)$ où $5 = \sqrt{n\,p(1-p)}$
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) :

$$\begin{aligned}\mathbb{P}[48 \leqslant X \leqslant 52] &= \mathbb{P}[47{,}5 \leqslant X \leqslant 52{,}5] \\ &= \mathbb{P}[\tfrac{47{,}5-50}{5} < \tfrac{X-50}{5} < \tfrac{52{,}5-50}{5}] \\ &= \mathbb{P}[-0{,}5 < Z < 0{,}5] \\ &= F(0{,}5) - F(-0{,}5) \\ &\simeq 0{,}6915 - (1 - 0{,}691\,5) \\ &\simeq 0{,}691\,5 - 0{,}308\,5 \\ &\simeq 0{,}38\end{aligned}$$

ii) 1000 fois.
On a $X \sim \mathcal{B}(1000; 0{,}5)$.
On a $n = 1000 > 30$, et $n\,p = 500 > 5$, et $n(1-p) = 500 > 5$ donc on peut approximer X par $\mathcal{N}(500; 15{,}8)$ où $15{,}8 \simeq \sqrt{n\,p(1-p)}$
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) :

$$\begin{aligned}\mathbb{P}[480 \leqslant X \leqslant 520] &= \mathbb{P}[479{,}5 \leqslant X \leqslant 520{,}5] \\ &= \mathbb{P}[\tfrac{479{,}5-500}{15{,}8} < \tfrac{X-500}{15{,}8} < \tfrac{520{,}5-500}{15{,}8}] \\ &\simeq \mathbb{P}[-1{,}3 < Z < 1{,}3] \\ &\simeq F(1{,}3) - F(-1{,}3) \\ &\simeq 0{,}9032 - (1 - 0{,}903\,2) \\ &\simeq 0{,}903\,2 - 0{,}096\,8 \\ &\simeq 0{,}81\end{aligned}$$

iii) 2000 fois.
On a $X \sim \mathcal{B}(2000; 0{,}5)$.
On a $n = 2000 > 30$, et $n\,p = 1000 > 5$, et $n(1-p) =$

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

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

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

$$\begin{aligned}\mathbb{P}[960 \leqslant X \leqslant 1040] &= \mathbb{P}[959{,}5 \leqslant X \leqslant 1040{,}5] \\ &= \mathbb{P}[\tfrac{959{,}5-1000}{22{,}4} < \tfrac{X-1000}{22{,}4} < \tfrac{1040{,}5-1000}{22{,}4}] \\ &\simeq \mathbb{P}[-1{,}81 < Z < 1{,}81] \\ &\simeq F(1{,}81) - F(-1{,}81) \\ &\simeq 0{,}9649 - (1 - 0{,}964\,9) \\ &\simeq 0{,}964\,9 - 0{,}035\,1 \\ &\simeq 0{,}93\end{aligned}$$

iv) 10000 fois.

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

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

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

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

$$\begin{aligned}\mathbb{P}[4800 \leqslant X \leqslant 5200] &= \mathbb{P}[4799{,}5 \leqslant X \leqslant 5200{,}5] \\ &= \mathbb{P}[\tfrac{4799{,}5-5000}{50} < \tfrac{X-5000}{50} < \tfrac{5200{,}5-5000}{50}] \\ &= \mathbb{P}[-4{,}01 < Z < 4{,}01] \\ &= F(4{,}01) - F(-4{,}01) \\ &\simeq 1{,}0000 - (1 - 1) \\ &\simeq 1{,}0 \\ &\simeq 1\end{aligned}$$

Exercice 44 :

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

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

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

   $$\begin{aligned}\mathbb{P}[X \geqslant 110] &= \mathbb{P}[\tfrac{110-100}{10} < \tfrac{X-100}{10}] \simeq \mathbb{P}[1{,}0 < Z] \\ &\simeq 1 - F(1{,}0) \simeq 1{,}0 - 0{,}841\,3 \simeq 0{,}16\end{aligned}$$

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

   $$\begin{aligned}\mathbb{P}[X \leqslant 80] &= \mathbb{P}[\tfrac{X-100}{10} < \tfrac{80-100}{10}] \simeq \mathbb{P}[Z < -2{,}0] \\ &\simeq F(-2) \simeq 1 - 0{,}977\,2 \simeq 0{,}02\end{aligned}$$

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

   $$\begin{aligned}\mathbb{P}[90 \leqslant X \leqslant 110] &= \mathbb{P}[\tfrac{90-100}{10} < \tfrac{X-100}{10} < \tfrac{110-100}{10}] \\ &= \mathbb{P}[-1{,}0 < Z < 1{,}0] \\ &= F(1{,}0) - F(-1) \\ &\simeq 0{,}8413 - (1 - 0{,}841\,3) \\ &\simeq 0{,}841\,3 - 0{,}158\,7 \\ &\simeq 0{,}68\end{aligned}$$

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

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

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

Chapitre 5 : Estimation

Exercice 45 : Pile ou face

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

1. Estimez, avec la confiance 99%, la probabilité qu'a cette pièce de tomber sur pile quand on la lance.
   On a $n = 800 > 30$, et $p_e = \frac{443}{800} \simeq 0{,}554$, d'où $n\,p_e = 443 > 5$, et $n(1-p_e) = 357 > 5$ donc on peut utiliser la procédure du formulaire pour estimer la proportion p.
   On a $F(2{,}576) \simeq 0{,}995$ d'où $z_\alpha \simeq 2{,}576$
   d'où $a_\alpha = z_\alpha\sqrt{\frac{p_e(1-p_e)}{n}} \simeq 0{,}045$.
   On estime donc que p est dans l'intervalle $[0{,}509; 0{,}599]$ avec la confiance $c = 0{,}99$.
2. Conclure : pouvez-vous affirmer (avec la confiance 99%) que cette pièce soit biaisée ?
   On peut effectivement affirmer, avec la confiance 99%, que la pièce est biaisée (on a vu qu'elle a au moins 50,9% de chances de faire pile).

Exercice 46 : Efficacité d'un traitement

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

Pour vérifier cette affirmation du fabricant, on l'administre à 800 malades, et on constate que parmi cet échantillon de 800 patient traités, il y en a 447 qui guérissent en 10 jours.

1. Donner une estimation de la vraie proportion p de guérisons (pour l'ensemble des patients que l'on traiterait avec ce médicament) avec une confiance de 95%.
   On a $n = 800 > 30$, et $p_e = \frac{447}{800} \simeq 0{,}559$, d'où $n\,p_e = 447 > 5$, et $n(1-p_e) = 353 > 5$ donc on peut utiliser la procédure du formulaire pour estimer la proportion p.
   On a $F(1{,}96) \simeq 0{,}975$ d'où $z_\alpha \simeq 1{,}96$
   d'où $a_\alpha = z_\alpha\sqrt{\frac{p_e(1-p_e)}{n}} \simeq 0{,}0344$.
   On estime donc que p est dans l'intervalle $[0{,}5243; 0{,}5931]$ avec la confiance $c = 0{,}95$.
2. Que peut-on conclure par rapport à l'affirmation du fabriquant ?
   Donc avec la confiance 95%, on peut affirmer que le taux est bien supérieur à ce qu'affirme le fabriquant : il est d'au moins 52%.
3. On veut réduire la marge de l'estimation à 1% tout en atteignant une confiance de 98%. Quelle doit alors être la taille minimale de l'échantillon à considérer ?
   Pour avoir une précision 0,01 avec la confiance 0,98, il faut $n > z_\alpha^2\frac{p_e(1-p_e)}{h^2} \simeq 2{,}326^2\frac{0{,}559(1-0{,}559)}{0{,}01^2} \simeq 13337$

Exercice 47 : Satisfaction de clients

Afin de répondre au mieux au désir de sa clientèle, une entreprise réalise un sondage pour connaître l'avis des consommateurs sur un produit qu'elle fabrique. Sur 1218 personnes interrogées, 618 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 905 personnes interrogées, il y en a alors 482 qui sont satisfaites.

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

   Estimation de p_1 : On a $n = 1218 > 30$, et $p_e = \frac{618}{1\,218} \simeq 0{,}507$, d'où $n\,p_e = 618 > 5$, et $n(1-p_e) = 600 > 5$ donc on peut utiliser la procédure du formulaire pour estimer la proportion p_1.
   On a $F(1{,}645) \simeq 0{,}95$ d'où $z_\alpha \simeq 1{,}645$
   d'où $a_\alpha = z_\alpha\sqrt{\frac{p_e(1-p_e)}{n}} \simeq 0{,}0236$.
   On estime donc que p_1 est dans l'intervalle $[0{,}4838; 0{,}531]$ avec la confiance $c = 0{,}9$.

Estimation de p_2 : On a $n = 905 > 30$, et $p_e = \frac{482}{905} \simeq 0{,}533$, d'où $n\,p_e = 482 > 5$, et $n(1-p_e) = 423 > 5$ donc on peut utiliser la procédure du formulaire pour estimer la proportion p_2.
On a $F(1{,}645) \simeq 0{,}95$ d'où $z_\alpha \simeq 1{,}645$
d'où $a_\alpha = z_\alpha\sqrt{\frac{p_e(1-p_e)}{n}} \simeq 0{,}0273$.
On estime donc que p_2 est dans l'intervalle $[0{,}5053; 0{,}5599]$ avec la confiance $c = 0{,}9$.

conclusion Comme les intervales se chevauchent on ne peut pas savoir la quelle des proportions est plus grande que l'autre, donc ces intervalles ne permettent pas de conclure que la proportion de clients satisfaits ait augmenté. (Par exemple il serait tout à fait possible d'avoir $p_1 = 0{,}53$ et $p_2 = 0{,}51$ donc $p_1 > p_2$, mais il serait possible aussi d'avoir $p_1 = 0{,}51$ et $p_2 = 0{,}53$ donc $p_1 < p_2$.)

2. Lors du deuxième sondage, quelle aurait du être la taille de l'échantillon pour avoir une estimation de p_2 à 1% près avec une confiance de 0,99 ?
   Pour avoir une précision 0,01 avec la confiance 0,99, il faut $n > z_\alpha^2\frac{p_e(1-p_e)}{h^2} \simeq 2{,}576^2\frac{0{,}533(1-0{,}533)}{0{,}01^2} \simeq 16517$

Exercice 48 : Résistance à la persuasion

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

1. On soumet dans un premier temps un échantillon de 50 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[	[17 ; 19[
Effectif	5	6	15	17	4	3

   a) Calculer la moyenne expérimentale $m_e(X)$ et l'écart type expérimental $s_e(X)$ de l'échantillon.
   moyenne : $m_e(X) = \frac{\sum c_i\,n_i}{n} = \frac{8\times5+10\times6+12\times15+\cdots+18\times3}{50} = \frac{636}{50} = 12{,}72$
   $m_e(X^2) = \frac{\sum c_i^2\,n_i}{n} = \frac{8^2\times5+10^2\times6+12^2\times15+\cdots+18^2\times3}{50} = \frac{8\,408}{50}$
   $Var_e(X) = m_e(X^2) - m_e(X)^2 = \frac{8\,408}{50} - \left(\frac{636}{50}\right)^2 \simeq 6{,}362$
   Écart-type : $s_e(X) = \sqrt{Var_e(X)} \simeq 2{,}522$

   b) Estimer pour l'ensemble des étudiants la note moyenne après avoir entendu un tel argumentaire (on l'appelle $\mu(X)$). On déterminera un intervalle de confiance, avec la confiance $c = 0{,}95$.
   Comme n= 50 >30, on cherche z_α tel que $F(z_\alpha) = \frac{0{,}95+1}{2} = 0{,}975$
   On a $F(1{,}96) \simeq 0{,}975$ d'où $z_\alpha \simeq 1{,}96$ d'où $a_\alpha = z_\alpha\frac{s_e}{\sqrt{n-1}} \simeq 0{,}7062$
   On estime donc que $\mu(X)$ est dans l'intervalle $[12{,}72 - 0{,}706\,2\,;\,12{,}72 + 0{,}706\,2] \simeq [12{,}01\,;\,13{,}43]$, avec la confiance $c = 0{,}95$.

   c) Peut-on dire avec un risque d'erreur de 5% qu'en moyenne, la note (traduisant l'opinion en faveur d'une coopération entre étudiants et enseignants) a diminué après l'argumentaire ? Justifiez votre réponse.
   Toutes les valeurs de cet intervalle de confiance sont plus petites que 13,7, donc on peut déduire avec la confiance 95% que la note moyenne a baissé.

   d) Quelle devrait être la taille minimale de l'échantillon à prélever pour estimer la moyenne $\mu(X)$ à 0,5 points près avec une confiance de 0,99 ?

Pour avoir une précision 0,5 avec la confiance 0,99, il faut $n > z_\alpha^2 \frac{(s_e)^2}{h^2} \simeq 2{,}576^2 \frac{2{,}522^2}{0{,}5^2} \simeq 169$

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

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

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

   On lit $t_\alpha \simeq 2{,}0595$ d'où $a_\alpha = t_\alpha \frac{s_e}{\sqrt{n-1}} \simeq 1{,}73$.

   On estime donc que $\mu(Y)$ est dans l'intervalle $[9 - 1{,}73\,;\, 9 + 1{,}73] = [7{,}27\,;\, 10{,}73]$ avec la confiance $c = 0{,}95$

   b) Comparer cette moyenne à celle qu'ils auraient sans réciter les tables de multiplication (en étant donc en mesure de réfléchir à des contre-arguments). Ces résultats sont-ils compatibles avec l'hypothèse théorique que l'on souhaitait tester ?

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

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

   c) Donner une estimation de l'écart type des notes, dans le deuxième cas (en récitant les tables de multiplication) avec un risque d'erreur de 5%.

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

   On obtient alors $s_e\sqrt{\frac{n}{x_2}} \simeq 4{,}2\sqrt{\frac{26}{40{,}65}} \simeq 3{,}36$ et $s_e\sqrt{\frac{n}{x_1}} \simeq 4{,}2\sqrt{\frac{26}{13{,}12}} \simeq 5{,}91$

   On estime donc que $\sigma(Y)$ est dans l'intervalle $[3{,}36; 5{,}91]$ avec la confiance $c = 0{,}95$.

Exercice 49 : Motivation lors d'un test

Un expérimentateur qui étudie la prise de décisions a demandé à 30 enfants de résoudre le plus grand nombre possible de problèmes en 30 minutes. Il a expliqué à un groupe de 16 enfants qu'il voulait tester leur aptitude innée à résoudre des problèmes, et 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	36	22	34	31	19	20	33	27	35	34	28	32	28	26	24	29
Occupation	19	18	24	19	18	20	18	20	22	23	13	20	19	22		

On suppose que les résultats varient normalement.

1. Calculer les moyennes et les écarts types des deux échantillons.

   On note X les notes obtenues en affirmant qu'on teste leurs aptitudes inées, et Y les notes si on affirme qu'on veut les ocuper.

   $m(X) \simeq 28{,}625$ $\quad s(X) \simeq 5{,}195.$

   $m(Y) \simeq 19{,}643$ $\quad s(Y) \simeq 2{,}608.$

2. Donner une estimation par intervalle de confiance des moyennes et des écarts types associés aux deux conditions d'expérience. *On fera cette estimation avec une confiance de* 95%.

a) Si on leur dit qu'on teste une **aptitude innée** :
Comme $n = 16 \leqslant 30$, on cherche t_α à partir de la table inverse de Student avec $p = \frac{\alpha}{2} = 0{,}025$ et $n-1 = 15$ degrés de liberté (ddl)
On lit $t_\alpha \simeq 2{,}1314$ d'où $a_\alpha = t_\alpha \frac{s_e}{\sqrt{n-1}} \simeq 2{,}858\,9$.
On estime donc que $\mu(X)$ est dans l'intervalle $[28{,}625 - 2{,}858\,9\,;\, 28{,}625 + 2{,}858\,9] \simeq [25{,}77\,;\, 31{,}48]$ avec la confiance $c = 0{,}95$

On lit dans la table du χ^2 à 15 degrés de liberté la valeur $x_1 = 6{,}262$ correspondant à $p = 0{,}025$ et la valeur $x_2 = 27{,}49$ correspondant à $q = 0{,}025$
On obtient alors $s_e\sqrt{\frac{n}{x_2}} \simeq 5{,}195\sqrt{\frac{16}{27{,}49}} \simeq 3{,}96$ et $s_e\sqrt{\frac{n}{x_1}} \simeq 5{,}195\sqrt{\frac{16}{6{,}262}} \simeq 8{,}3$
On estime donc que $\sigma(X)$ est dans l'intervalle $[3{,}96; 8{,}3]$ avec la confiance $c = 0{,}95$.

b) Si on leur dit juste qu'on les **occupe** :
Comme $n = 14 \leqslant 30$, on cherche t_α à partir de la table inverse de Student avec $p = \frac{\alpha}{2} = 0{,}025$ et $n-1 = 13$ degrés de liberté (ddl)
On lit $t_\alpha \simeq 2{,}1604$ d'où $a_\alpha = t_\alpha \frac{s_e}{\sqrt{n-1}} \simeq 1{,}562\,7$.
On estime donc que $\mu(Y)$ est dans l'intervalle $[19{,}643 - 1{,}562\,7\,;\, 19{,}643 + 1{,}562\,7] \simeq [18{,}08\,;\, 21{,}21]$ avec la confiance $c = 0{,}95$

On lit dans la table du χ^2 à 13 degrés de liberté la valeur $x_1 = 5{,}009$ correspondant à $p = 0{,}025$ et la valeur $x_2 = 24{,}74$ correspondant à $q = 0{,}025$
On obtient alors $s_e\sqrt{\frac{n}{x_2}} \simeq 2{,}608\sqrt{\frac{14}{24{,}74}} \simeq 1{,}96$ et $s_e\sqrt{\frac{n}{x_1}} \simeq 2{,}608\sqrt{\frac{14}{5{,}009}} \simeq 4{,}36$
On estime donc que $\sigma(X)$ est dans l'intervalle $[1{,}96; 4{,}36]$ avec la confiance $c = 0{,}95$.

3. Que peut-on conclure au vu de ces résultats ?
Ces intervalles montrent que $\mu(X) > \mu(Y)$ (en effet $\mu(X) > 25$ alors que $\mu(Y) < 22$). Cela signifie qu'avec la confiance 95%, on peut affirmer qu'en moyenne les enfants y arivent mieux si on leur dit qu'on veut les tester que si on leur dit qu'on veut les occuper.

Exercice 50 : Taux de cholesterol

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

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

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

1. Donner des estimations des écarts types des deux populations.
On choit par exemple de faire des estimations avec la confiance 95% :

femmes On lit dans la table du χ^2 à 21 degrés de liberté la valeur $x_1 = 10{,}28$ correspondant à $p = 0{,}025$ et la valeur $x_2 = 35{,}48$ correspondant à $q = 0{,}025$
On obtient alors $s_e\sqrt{\frac{n}{x_2}} \simeq 36{,}4\sqrt{\frac{22}{35{,}48}} \simeq 28{,}7$ et $s_e\sqrt{\frac{n}{x_1}} \simeq 36{,}4\sqrt{\frac{22}{10{,}28}} \simeq 53{,}2$
On estime donc que σ est dans l'intervalle $[28{,}7; 53{,}2]$ avec la confiance $c = 0{,}95$.

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

On obtient alors $s_e\sqrt{\frac{n}{x_2}} \simeq 45{,}2\sqrt{\frac{18}{30{,}19}} \simeq 34{,}9$ et $s_e\sqrt{\frac{n}{x_1}} \simeq 45{,}2\sqrt{\frac{18}{7{,}564}} \simeq 69{,}7$

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

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

On commence par estimer les moyennes :

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

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

On estime donc que μ est dans l'intervalle $[185{,}7 - 16{,}519\,;\, 185{,}7 + 16{,}519] \simeq [169{,}18\,;\, 202{,}22]$ avec la confiance $c = 0{,}95$

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

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

On estime donc que μ est dans l'intervalle $[192{,}4 - 23{,}129\,;\, 192{,}4 + 23{,}129] \simeq [169{,}27\,;\, 215{,}53]$ avec la confiance $c = 0{,}95$

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

Exercice 51 : Épreuve de dictée

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

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

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

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

$m(X) = 15{,}933$ $s(X) = 1{,}974$.

$m(Y) = 12{,}067$ $s(Y) = 1{,}389$.

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

Épreuve 1 Comme $n = 15 \leqslant 30$, on cherche t_α à partir de la table inverse de Student avec $p = \frac{\alpha}{2} = 0{,}025$ et $n - 1 = 14$ degrés de liberté (ddl)

On lit $t_\alpha \simeq 2{,}1448$ d'où $a_\alpha = t_\alpha \frac{s_e}{\sqrt{n-1}} \simeq 1{,}131\,5$.

On estime donc que $\mu(X)$ est dans l'intervalle $[15{,}933 - 1{,}131\,5\,;\, 15{,}933 + 1{,}131\,5] \simeq [14{,}8\,;\, 17{,}06]$ avec la confiance $c = 0{,}95$

Épreuve 2 Comme $n = 15 \leqslant 30$, on cherche t_α à partir de la table inverse de Student avec $p = \frac{\alpha}{2} = 0{,}025$ et $n - 1 = 14$ degrés de liberté (ddl)

On lit $t_\alpha \simeq 2{,}1448$ d'où $a_\alpha = t_\alpha \frac{s_e}{\sqrt{n-1}} \simeq 0{,}796\,21$.

On estime donc que $\mu(Y)$ est dans l'intervalle $[12{,}067 - 0{,}796\,21\,;\, 12{,}067 + 0{,}796\,21] \simeq [11{,}27\,;\, 12{,}86]$ avec la confiance $c = 0{,}95$

(c) : Ces intervalles indiquent que $\mu(X) > \mu(Y)$ (en effet $\mu(X) > 14$ et $\mu(Y) < 13$). Donc on peut dire (avec la confiance 95%) que les résultats moyens diffèrent significativement d'une épreuve à l'autre, et plus 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 14 degrés de liberté la valeur $x_1 = 5{,}629$ correspondant à $p = 0{,}025$ et la valeur $x_2 = 26{,}12$ correspondant à $q = 0{,}025$
On obtient alors $s_e\sqrt{\frac{n}{x_2}} \simeq 1{,}974\sqrt{\frac{15}{26{,}12}} \simeq 1{,}5$ et $s_e\sqrt{\frac{n}{x_1}} \simeq 1{,}974\sqrt{\frac{15}{5{,}629}} \simeq 3{,}22$
On estime donc que $\sigma(X)$ est dans l'intervalle $[1{,}5; 3{,}22]$ avec la confiance $c = 0{,}95$.

Épreuve 2 On lit dans la table du χ^2 à 14 degrés de liberté la valeur $x_1 = 5{,}629$ correspondant à $p = 0{,}025$ et la valeur $x_2 = 26{,}12$ correspondant à $q = 0{,}025$
On obtient alors $s_e\sqrt{\frac{n}{x_2}} \simeq 1{,}389\sqrt{\frac{15}{26{,}12}} \simeq 1{,}05$ et $s_e\sqrt{\frac{n}{x_1}} \simeq 1{,}389\sqrt{\frac{15}{5{,}629}} \simeq 2{,}27$
On estime donc que $\sigma(Y)$ est dans l'intervalle $[1{,}05; 2{,}27]$ avec la confiance $c = 0{,}95$.

Exercice 52 :

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

1. En France, le taux de chômage est de 10% (selon les données de l'INSEE).
   On choisit au hasard un échantillon de 117 français. Quelle est la probabilité d'avoir, au sein de cet échantillon, entre 9% et 11% de chômeurs ?
   Justifiez bien les éventuelles approximations que vous seriez amené à faire
   - L'échantillon de 117 personnes est choisi parmi une population de dizaines de millions de français adultes, donc $\frac{n}{N} < \frac{1}{10}$ et on peut utiliser la loi binomiale : si on note X le nombre chômeurs d'un tel échantillon aléatoire, alors $X \sim \mathcal{B}(117; 0{,}1)$.
   - On a $n = 117 > 30$, et $n\,p \simeq 11{,}7 > 5$, et $n(1-p) \simeq 105 > 5$ donc on peut approximer X par $\mathcal{N}(11{,}7; 3{,}24)$ où $3{,}24 \simeq \sqrt{n\,p(1-p)}$
   - Une proportion entre 0,09 et 0,11 signifie que X est entre $117 \times 0{,}09 = 10{,}53$ et $117 \times 0{,}11 = 12{,}87$. (Donc entre 11 et 12 comme X est entier.) On calcule donc :

$$\begin{aligned}\mathbb{P}[11 \leqslant X \leqslant 12] &= \mathbb{P}[10{,}5 \leqslant X \leqslant 12{,}5]\\ &= \mathbb{P}[\tfrac{10{,}5-11{,}7}{3{,}24} < \tfrac{X-11{,}7}{3{,}24} < \tfrac{12{,}5-11{,}7}{3{,}24}]\\ &\simeq \mathbb{P}[-0{,}37 < Z < 0{,}25]\\ &\simeq F(0{,}25) - F(-0{,}37)\\ &\simeq 0{,}5987 - (1 - 0{,}644\,3)\\ &\simeq 0{,}598\,7 - 0{,}355\,7\\ &\simeq 0{,}24\end{aligned}$$

2. Sur un échantillon de 117 patients atteints de boulimie, le psychologue a constaté que 5,98% d'entre eux étaient au chômage.
   (a) Avec la confiance $c = 90\%$, déterminer un intervalle de confiance pour le taux de chômage des personnes boulimiques.
   Peut on en déduire (avec la confiance 90%) que le taux de chômage des personnes boulimiques est différent de celui du reste de la population ?
   On a $n = 117 > 30$, et $p_e = \frac{7}{117} \simeq 0{,}06$, d'où $n\,p_e = 7 > 5$,

et $n(1-p_e) = 110 > 5$ donc on peut utiliser la procédure du formulaire pour estimer la proportion p.
On a $F(1{,}645) \simeq 0{,}95$ d'où $z_\alpha \simeq 1{,}645$
d'où $a_\alpha = z_\alpha \sqrt{\frac{p_e(1-p_e)}{n}} \simeq 0{,}0361$.
On estime donc que p est dans l'intervalle $[0{,}0237; 0{,}0959]$ avec la confiance $c = 0{,}9$.
Toutes les valeurs de cet intervalle sont inférieures à 10%, donc on peut affirmer, avec la confiance 90%, que le taux de chomage des personnes boulimiques est plus faible que parmi le reste de la population.

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

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

On a $n = 117 > 30$, et $p_e = \frac{7}{117} \simeq 0{,}06$, d'où $n\,p_e = 7 > 5$, et $n(1-p_e) = 110 > 5$ donc on peut utiliser la procédure du formulaire pour estimer la proportion p.
On a $F(2{,}326) \simeq 0{,}99$ d'où $z_\alpha \simeq 2{,}326$
d'où $a_\alpha = z_\alpha \sqrt{\frac{p_e(1-p_e)}{n}} \simeq 0{,}051$.
On estime donc que p est dans l'intervalle $[0{,}0088; 0{,}1108]$ avec la confiance $c = 0{,}98$.
Cette fois-ci, la valeur 10% se situe à l'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 1% près le taux de chômage des personnes boulimiques avec une confiance 98%, quelle taille d'échantillon faudrait il considérer ?

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

Exercice 53 : Une chercheuse, Mme Lefebvre, se demande si la perception des couleurs diffère en fonction du sexe des individus.

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

   On considère d'une part un échantillon de 1 000 femmes, parmi lesquelles 7 sont daltoniennes, et d'autre part un échantillon de 1 000 hommes, parmi lesquels 94 sont daltoniens.

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

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

   - **parmi les femmes :**
     On a $n = 1000 > 30$, et $p_e = \frac{7}{1\,000} = 0{,}007$, d'où $n\,p_e = 7 > 5$, et $n(1-p_e) = 993 > 5$ donc on peut utiliser la procédure du formulaire pour estimer la proportion p.
     On a $F(2{,}576) \simeq 0{,}995$ d'où $z_\alpha \simeq 2{,}576$
     d'où $a_\alpha = z_\alpha \sqrt{\frac{p_e(1-p_e)}{n}} \simeq 0{,}00679$.
     On estime donc que p est dans l'intervalle $[0{,}00021; 0{,}01379]$ avec la confiance $c = 0{,}99$.
   - **parmi les hommes :**
     On a $n = 1000 > 30$, et $p_e = \frac{94}{1\,000} = 0{,}094$, d'où $n\,p_e = 94 > 5$, et $n(1-p_e) = 906 > 5$ donc on peut utiliser la procédure du formulaire pour estimer la proportion p.
     On a $F(2{,}576) \simeq 0{,}995$ d'où $z_\alpha \simeq 2{,}576$
     d'où $a_\alpha = z_\alpha \sqrt{\frac{p_e(1-p_e)}{n}} \simeq 0{,}02377$.
     On estime donc que p est dans l'intervalle $[0{,}07023; 0{,}11777]$ avec la confiance $c = 0{,}99$.

(b) On remarque que ces intervalles ne se chevanchent pas et indiquent une plus grande proportion chez les hommes. On peut donc affirmer, avec la confiance $c = 99\%$, que la prévalence du daltonisme est plus importante chez l'homme que chez la femme.

2. La chercheuse pense que, bien au delà de la question du daltonisme, les filles sont bien plus douées que les garçons pour détecter d'infimes nuances de couleurs. Elle teste cette hypothèse sur des enfants de 7 ans, auxquels elle fait passer un test standardisé de perception des couleurs. Sur un échantillon de 15 filles, elle obtient les données suivantes :

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

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

- On détermine tout d'abord la moyenne et l'écart-type au sein de cet échantillon :
  moyenne : $m(X) = \frac{\sum x_i}{n} = \frac{7+6+8+\cdots+7}{15} = \frac{84}{15} = 5{,}6$
  $m(X^2) = \frac{\sum x_i^2}{n} = \frac{7^2+6^2+8^2+\cdots+7^2}{15} = \frac{534}{15}$
  $Var(X) = m(X^2) - m(X)^2 = \frac{534}{15} - \left(\frac{84}{15}\right)^2 = 4{,}24$
  Écart-type : $s(X) = \sqrt{Var(X)} \simeq 2{,}06$
- Comme $n = 15 \leqslant 30$, on cherche t_α à partir de la table inverse de Student avec $p = \frac{\alpha}{2} = 0{,}01$ et $n-1 = 14$ degrés de liberté (ddl)
  On lit $t_\alpha \simeq 2{,}6245$ d'où $a_\alpha = t_\alpha \frac{s_e}{\sqrt{n-1}} \simeq 1{,}444\,9$.
  On estime donc que μ est dans l'intervalle $[5{,}6 - 1{,}444\,9\,;\ 5{,}6 + 1{,}444\,9] \simeq [4{,}16\,;\ 7{,}04]$ avec la confiance $c = 0{,}98$

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

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

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

- Comme n= 40 >30, on cherche z_α tel que $F(z_\alpha) = \frac{0{,}98+1}{2} = 0{,}99$
  On a $F(2{,}326) \simeq 0{,}99$ d'où $z_\alpha \simeq 2{,}326$ d'où $a_\alpha = z_\alpha \frac{s_e}{\sqrt{n-1}} \simeq 0{,}6444$
  On estime donc que μ est dans l'intervalle $[4{,}27 - 0{,}644\,4\,;\ 4{,}27 + 0{,}644\,4] \simeq [3{,}63\,;\ 4{,}91]$, avec la confiance $c = 0{,}98$.
- Les intervalles se chevauchent, donc ces échantillons ne permettent pas d'affirmer, avec la confiance 98%, que les filles aient en moyenne une meilleure perception des couleurs que les garçons.

4. Estimez de même les écart-types des garçons et des filles, avec la confiance 98%. Peut on conclure qu'ils diffèrent selon le sexe ?

(a) **Estimation de l'écart-type des filles**
On lit dans la table du χ^2 à 14 degrés de liberté la valeur $x_1 = 4{,}66$ correspondant à $p = 0{,}01$ et la valeur $x_2 = 29{,}14$ correspondant à $q = 0{,}01$
On obtient alors $s_e\sqrt{\frac{n}{x_2}} \simeq 2{,}06\sqrt{\frac{15}{29{,}14}} \simeq 1{,}48$ et $s_e\sqrt{\frac{n}{x_1}} \simeq 2{,}06\sqrt{\frac{15}{4{,}66}} \simeq 3{,}7$
On estime donc que σ est dans l'intervalle $[1{,}48; 3{,}7]$ avec la confiance $c = 0{,}98$.

(b) **Estimation de l'écart-type des garçons**
On lit dans la table du χ^2 à 39 degrés de liberté la valeur $x_1 = 21{,}43$ correspondant à $p = 0{,}01$ et la valeur $x_2 = 62{,}43$ correspondant à $q = 0{,}01$
On obtient alors $s_e\sqrt{\frac{n}{x_2}} \simeq 1{,}73\sqrt{\frac{40}{62{,}43}} \simeq 1{,}38$ et $s_e\sqrt{\frac{n}{x_1}} \simeq 1{,}73\sqrt{\frac{40}{21{,}43}} \simeq 2{,}36$
On estime donc que σ est dans l'intervalle $[1{,}38; 2{,}36]$ avec la confiance $c = 0{,}98$.

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

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

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