Parcours différenciés : Statistiques

Objectif 1 : savoir faire les exercices ✧, tenter les exercices ✧✧.
Objectif 2 : savoir faire les exercices ✧, les exercices ✧✧, tenter les exercices ✧✧✧.
Objectif 3 : savoir faire les exercices ✧ (si possible mentalement), les exercices ✧✧ et les exercices ✧✧✧ et prendre des initiatives.
savoir faire : travail autonome avec des stratégies d'auto-correction.
tenter : travail de recherche, précision (par écrit) des pistes engagées, réflexion sur les résultats éventuellement établis.
prendre des initiatives : étendre l'exercice à une réflexion personnelle pour prolonger le travail réalisé (recherches documentaires, se poser des questions et y répondre, trouver d'autres solutions pour une même question).

Exercice 1 ✧

Soit la série statistique discrète de variable x prenant les valeurs suivantes :

10	6	8	15	14
13	16	14	20	11
9	12	4	13	10
14	11	8	17	15

1. Calculer la moyenne $\overline{x}$ de la série
2. Déterminer la médiane M de la série (ou le quartile Q_2).
3. Déterminer le premier Q_1 et le troisième quartile Q_3 de la série, en déduire l'écart-inter-quartile.
4. Calculer l'étendue de la série.
5. Avec la calculatrice, déterminer l'écart-type σ de la série.

Exercice 2 ✧

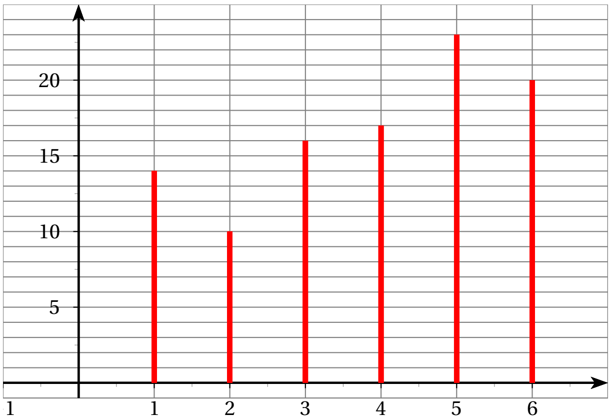
Soit la série statistique discrète de variable x prenant les valeurs suivantes :

x	0	1	2	3	4	5	6	7	8	9	10
effectifs	2	5	4	8	12	10	15	11	12	8	6

1. Calculer la moyenne $\overline{x}$ de la série
2. Déterminer la médiane M de la série (ou le quartile Q_2).
3. Déterminer le premier Q_1 et le troisième quartile Q_3 de la série, en déduire l'écart-inter-quartile.
4. Calculer l'étendue de la série.
5. Avec la calculatrice, déterminer l'écart-type σ de la série.

☞**Exercice 3** ✧

Soit la série statistique discrète de variable x prenant les valeurs suivantes, cette série décrit 100 lancers d'un dé à six faces :



1. Calculer la moyenne $\overline{x}$ de la série. Placer la moyenne sur le graphique.
2. Déterminer la médiane M de la série (ou le quartile Q_2). Placer la médiane sur le graphique.
3. Déterminer le premier Q_1 et le troisième quartile Q_3 de la série, en déduire l'écart-interquartile. Placer les quartiles sur le graphique.
4. Calculer l'étendue de la série.
5. Avec la calculatrice, déterminer l'écart-type σ de la série.
6. À l'aide des critères statistiques repérés, choisir deux pour justifier que le dé est peut-être truqué.

☞**Exercice 4** ✧

Soit la série statistique discrète de variable x prenant les valeurs suivantes :

x	0	1	2	3	4	5	6	7	8	9	10
fréquences	6%	12%	14%	8%	12%	6%	12%	7%	10%	6%	7%

1. Calculer la moyenne $\overline{x}$ de la série
2. Déterminer le premier Q_1, le deuxième quartile Q_2 et le troisième quartile Q_3 de la série, en déduire l'écart-inter-quartile.
3. Calculer l'étendue de la série.
4. Avec la calculatrice, déterminer l'écart-type σ de la série.

☞**Exercice 5** ✧

Un jardinier a deux lots de bulbes de tulipes A et B de provenance différentes. Il a pesé un à un tous les bulbes.

Pour le lot A on donne les indicateurs suivants :

- la masse varie entre 10 et 100 g,
- la masse médiane est 55 g, le premier quartile est 40 g, le troisième quartile est 70 g.

1. À partir des résultats des masses en grammes des bulbes du lot A.
   (a) Estimer le pourcentage de bulbes dont la masse est supérieure ou égale à 40 g.
   (b) Donner l'intervalle interquartile et donner une interprétation de ce résultat en terme de pourcentage.
2. Pour le lot B, voici le tableau des effectifs :

masse	20	25	30	35	40	45	50	55	60
nombre de bulbes	10	14	22	25	18	12	8	6	5

   (a) Déterminer la masse moyenne au gramme près des bulbes du lot B.
   (b) Déterminer la médiane, les premier et troisième quartiles.
   (c) Quel est le pourcentage de bulbes dont la masse est strictement comprise entre 25 et 55 g ?
3. Lequel des deux lots semble le mieux calibré ? Justifier votre réponse.

☞**Exercice 6** ✧✧

Un directeur de supermarché décide d'étudier le temps d'attente aux caisses de son établissement pour ajuster le nombre de caisses ouvertes à la demande. Pour cela, il interroge le lundi et le vendredi cent clients et note les temps d'attente approximatifs en minutes entières.

PARTIE A : Étude de l'échantillon du lundi

Le lundi, il obtient la répartition suivante :

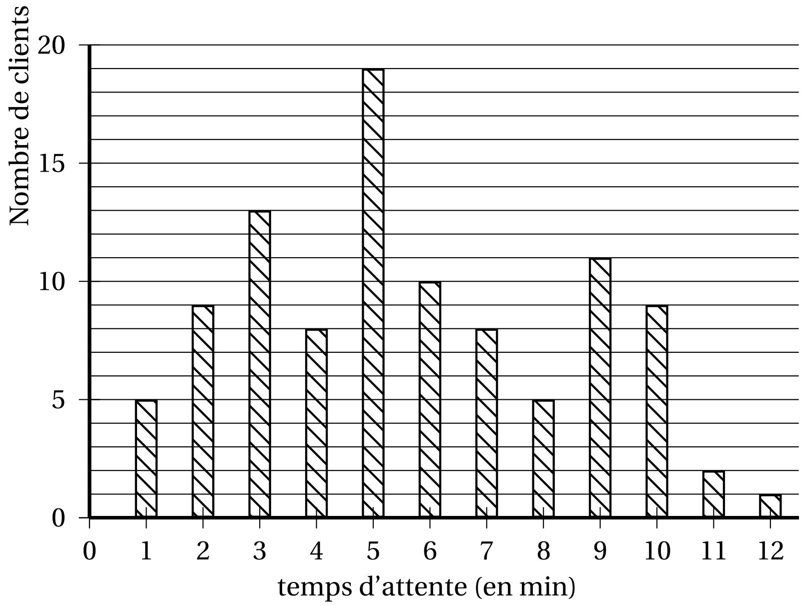
Temps d'attente en caisse (en min)	1	2	3	4	5	6	7	8	9	10
Nombre de clients	14	13	23	9	14	8	12	4	1	2

1. Calculer le temps moyen d'attente aux caisses du supermarché pour l'échantillon étudié.
2. Déterminer la médiane et les quartiles de la série statistique des temps d'attente.
3. (a) Son adjoint souhaite ouvrir une caisse supplémentaire si plus de 15 % des clients attendent 7 min ou plus en caisse. Doit-il ouvrir une nouvelle caisse le lundi ?
   (On justifiera la réponse).
   (b) Le directeur décide d'ouvrir une caisse supplémentaire si le temps moyen d'attente aux caisses dépasse 5 min. Doit-il ouvrir une nouvelle caisse le lundi ?
   (On justifiera la réponse).

PARTIE B : Étude de l'échantillon du vendredi

Le directeur décide de comparer les temps d'attente en début et en fin de semaine. Il a donc relevé le vendredi les temps d'attente aux caisses d'un échantillon de cent clients et obtient les résultats résumés dans le diagramme donné ci-dessous :

Temps d'attente le vendredi



1. Calculer le temps moyen d'attente aux caisses du supermarché le vendredi pour l'échantillon étudié (arrondi au dixième).

2. Dans un questionnaire, les clients qualifient d'acceptable un temps d'attente compris entre 2 et 6 minutes inclus. Pour chacune des affirmations suivantes, dire si elle est vraie ou fausse en justifiant la réponse.

   Affirmation A :
   Le vendredi, la moitié des clients attendent cinq min ou plus de cinq min en caisse.

   Affirmation B :
   Le vendredi, un quart des clients attend au plus trois minutes en caisse.

   Affirmation C :
   Il y a autant de clients qui trouvent le temps d'attente acceptable le lundi que le vendredi

☞**Exercice 7** ✧✧✧

Le tableau ci-dessous donne le nombre de naissances (en milliers) par an en France métropolitaine entre 1901 et 1920.

Année	1901	1902	1903	1904	1905	1906	1907	1908	1909	1910
Nombre de naissances en milliers	917,1	904,4	884,5	877,1	865,6	864,7	829,6	849,0	824,7	828,1
Année	1911	1912	1913	1914	1915	1916	1917	1918	1919	1920
Nombre de naissances en milliers	793,5	801,6	795,9	757,9	483,0	384,7	412,7	472,8	507,0	838,1

1. (a) Donner le nombre moyen de naissances par an en France métropolitaine entre 1901 et 1920. Arrondir la réponse à la centaine.

   (b) Donner la médiane, les premier et troisième quartiles de cette série statistique.

   (a) On donne les indicateurs statistiques de la série du nombre annuel de naissances x (en milliers) entre 1981 et 2000 :

   - $x \in [720\,;\,805]$
   - la médiane est 765, le premier quartile est 740 et le troisième quartile est 770.

2. Les phrases suivantes sont-elles vraies ou fausses ? Justifier la réponse.

   (a) Entre 1981 et 2000, le nombre annuel de naissances est supérieur à 760 000 pendant plus de 16 ans.

   (b) L'étendue du nombre annuel de naissances est plus de 5 fois plus élevée entre 1901 et 1920 qu'entre 1981 et 2000.

3. Quel contexte historique pourrait justifier la différence d'étendue entre les deux séries ?

Correction

Exercice 1

Soit la série statistique discrète de variable x prenant les valeurs suivantes :

4	6	8	8	9
10	10	11	11	12
13	13	14	14	14
15	15	16	17	20

1. $\overline{x} = \dfrac{4+6+8\times 2+9+10\times 2+11\times 2+12+13\times 2+14\times 3+15\times 2+16+17+20}{20} = 12$
   On peut utiliser la linarité de la moyenne pour calculer cette moyenne de tête en estimant la moyenne à 12 le tableau présente la variable $X = x-12$:

-8	-6	-4	-4	-3
-2	-2	-1	-1	0
1	1	2	2	2
3	3	4	5	8

   $\overline{X} = \dfrac{-8-6-4\times 2-3-2\times 2-1\times 2+0+1\times 2+2\times 3+3\times 2+4+5+8}{20} = \dfrac{0}{20}$
   $\vec{x} = \overline{X+12} = \overline{X}+12 = 0+12 = 12.$

2. On range les valeurs dans l'ordre croissant.
   l'effectif total est 20 :
   - La médiane est la moyenne de la 10^e et de la 11^e valeur de la série, soit $\dfrac{12+13}{2} = 12,5$.
   - Le deuxième quartile est la 10^e valeur de la série : $Q_2 = 12$.

3. L'effectif total est 20 :
   - $\dfrac{20}{4} = 5$. Le premier quartile est la 5^e valeur de la série. $Q_1 = 9$.
   - $\dfrac{20\times 3}{4} = 15$. Le troisième quartile est la 15^e valeur de la série. $Q_3 = 14$.
     L'écart-inter-quartile est $14-9 = 5$.

4. L'étendue de la série est $20-4 = 16$.

5. $\sigma = \sqrt{\dfrac{(4-12)^2+(6-12)^2+2\times(8-12)^2+\ldots+2\times(15-12)^2+(16-12)^2+(17-12)^2+(20-12)^2}{20}} \simeq 3,79$

☞**Exercice 2** ✧

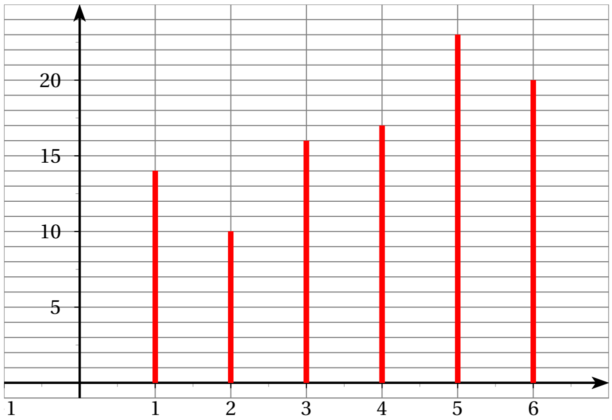
Soit la série statistique discrète de variable x prenant les valeurs suivantes :

x	0	1	2	3	4	5	6	7	8	9	10
effectifs	2	5	4	8	12	10	15	11	12	8	6

1. $\overline{x} = \dfrac{0\times 2+1\times 5+\dots 9\times 8+10\times 6}{2+5+\dots 8+6} \dfrac{530}{93} \simeq 5,7$
2. L'effectif total 93 est impair.
   - La médiane est la 47^e valeur de la série, $M=6$
   - Le deuxième quartile est la 47^e valeur de la série $Q_2 = M = 6$.
3. - $\dfrac{93}{4} = 23,25$. Le premier quartile est la 24^e valeur de la série, $Q_1 = 4$.
   - $\dfrac{93\times 3}{4} = 69,75$. Le troisième quartile est la 70^e valeur de la série, $Q_3 = 8$.
   - L'écart-inter-quartile est $8-4=4$.
4. L'étendue de la série est $10-0=10$.
5. $\sigma = \sqrt{\dfrac{2\times\left(0-\dfrac{530}{93}\right)^2+5\times\left(1-\dfrac{530}{93}\right)^2+\dots+8\times\left(9-\dfrac{530}{93}\right)^2+6\times\left(10-\dfrac{530}{93}\right)^2}{93}} \simeq 2,56.$

☞**Exercice 3** ✧

Soit la série statistique discrète de variable x prenant les valeurs suivantes, cette série décrit 100 lancers d'un dé à six faces :



1. $\overline{x} = \dfrac{1\times 14 + 2\times 10 + 3\times 16 + 4\times 17 + 5\times 23 + 6\times 20}{100} = 3,85$

2.
   - La médiane est la moyenne de la 50^e et de la 51^e valeur de la série, $\dfrac{4+4}{2} = 4$
   - Le deuxième quartile est la 50^e valeur de la série, $Q_2 = 4$.

3.
   - $\dfrac{100}{4} = 25$, le premier quartile est la 25^e valeur de la série, $Q_1 = 3$
   - $\dfrac{100\times 3}{4} = 25$, le premier quartile est la 75^e valeur de la série, $Q_3 = 5$
   - L'écart-inter-quartile est $5 - 3 = 2$

4. l'étendue de la série est $6 - 1 = 5$.

5. $\sigma = \sqrt{\dfrac{14\times(1-3,85)^2 + 10\times(2-3,85)^2 + \ldots + 23\times(5-3,85)^2 + 20\times(6-3,85)^2}{100}} \simeq 1,68.$

6. Si on considère le tableau des probabilités élémentaires d'un lancer de dé :

1	2	3	4	5	6
$\frac{1}{6} \simeq 0,17$	$\frac{1}{6} \simeq 0,17$	$\frac{1}{6} \simeq 0,17$	$\frac{1}{6} \simeq 0,17$	$\frac{1}{6} \simeq 0,17$	$\frac{1}{6} \simeq 0,17$

La moyenne de la série théorique est $3,5$, la médiane est $3,5$, le premier quartile est 2 le troisième quartile est 5.
Les critères statistiques de l'étude du dé truqué sont un peu plus élevés pour la moyenne, le premier quartile et la médiane. Les différences ne sont peut-être pas assez significative pour penser que le dé est truqué, il faudrait étudier davantage de lancers pour affirmer le défaut du dé.
S'il est truqué alors il semblerait que ce soit en faveur des numéros supérieurs à 3.

☞**Exercice 4** ✧

Soit la série statistique discrète de variable x prenant les valeurs suivantes :

x	0	1	2	3	4	5	6	7	8	9	10
fréquences	6%	12%	14%	8%	12%	6%	12%	7%	10%	6%	7%
fcc	6%	18%	32%	40%	52%	58%	70%	77%	87%	93%	100%

ffc : fréquences cumulées croissantes.

1. $\overline{x} = 0 \times 6\% + 1 \times 12\% + \ldots + 9 \times 6\% + 10 \times 7\% = 4,67$

2. - Le premier quartile Q_1 est la plus petite valeur de la série telle qu'au moins 25% des valeurs soient inférieures ou égales Q_1 : $Q_1 = 2$
   - Le deuxième quartile Q_2 (assimilé à la médiane) est la plus petite valeur de la série telle qu'au moins 50% des valeurs soient inférieures ou égales Q_2 : $Q_2 = 4$
   - Le premier quartile Q_3 est la plus petite valeur de la série telle qu'au moins 75% des valeurs soient inférieures ou égales Q_3 : $Q_3 = 7$
   -

3. L'étendue de la série est $10 - 0 = 10$.

4. $\sigma = \sqrt{0,06 \times (0-4,67)^2 + 0,12 \times (1-4,67)^2 + \ldots + 0,06 \times (9-4,67)^2 + 0,07 \times (10-4,67)^2} \simeq 3,00$

☞**Exercice 5** ✧

Un jardinier a deux lots de bulbes de tulipes A et B de provenance différentes. Il a pesé un à un tous les bulbes.

Pour le lot A on donne les indicateurs suivants :

- la masse varie entre 10 et 100 g,
- la masse médiane est 55 g, le premier quartile est 40 g, le troisième quartile est 70 g.

1. À partir des résultats des masses en grammes des bulbes du lot A.
   (a) au moins 25% des bulbes du lot A ont une masse inférieure ou égale à 40 g.
   (b) L'intervalle inter-quartile est $[40\ ;\ 70]$. Environ 50% des bulbes du lot A ont une masse comprise entre 40 g et 70 g.
2. Pour le lot B, voici le tableau des effectifs :

masse	20	25	30	35	40	45	50	55	60
nombre de bulbes	10	14	22	25	18	12	8	6	5

   (a) $\overline{x} = \dfrac{20 \times 10 + 20 \times 14 + .. + 55 \times 6 + 60 \times 5}{120} \simeq 36,5$.
   La masse moyenne des bulbes du lot B est $36,5$ g.
   (b)
   - La médiane est la moyenne entre la 60^e valeur et la 61^e valeur de la série : $M = \dfrac{35+35}{2} = 35$.
   - $\dfrac{120}{4} = 30$, Q_1 est la 30^e valeur de la série, $Q_1 = 30$.
   - $\dfrac{120 \times 3}{4} = 90$, Q_3 est la 90^e valeur de la série, $Q_3 = 45$.

   (c) $\dfrac{14+22+25+18+12+8+6}{120} \simeq 87,5\%$ ou $100\% - \dfrac{10+5}{120} = 87,5\%$. 87,5% des bulbes du lot B ont une masse comprise entre 25 g et 55 g.
3. Les masses des bulbes du lot A sont plus élevées (la médiane et les quartiles sont plus élevés), mais la répartition est moins homogène que celle des masses des bulbes du lot B, l'écart-inter-quartile et l'étendue sont plus faible pour les masses des bulbes le lot B.

☞**Exercice 6** ✧✧

PARTIE A : Étude de l'échantillon du lundi

Le lundi, il obtient la répartition suivante complétée par les effectifs cumulés croissants:

Temps d'attente en caisse (en min)	1	2	3	4	5	6	7	8	9	10
Nombre de clients	14	13	23	9	14	8	12	4	1	2
eff cum croissant	14	27	50	59	73	81	93	97	98	100

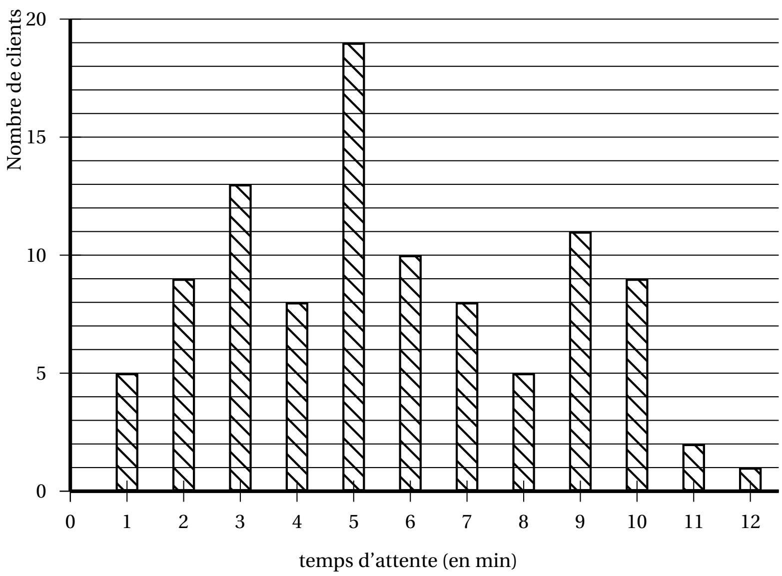
1. $\overline{x} = \frac{1\times14+2\times13+\cdots+9\times1+10\times2}{14+13+\cdots+1+2} = \frac{408}{100} = 4{,}08$ Le temps moyen d'attente aux caisses est d'environ $\boxed{4\text{ min}}$.
2.
   - la médiane est une valeur qui partage la série en deux parties de même effectif. Il y a 100 valeurs, nous prendrons pour médiane le centre de l'intervalle médian. La 50^e valeur est 3, la 51^e 4 donc $M_e = 3{,}5$
   - Le premier quartile Q_1 est la valeur du caractère dont le rang est le plus petit entier supérieur ou égal à $\frac{N}{4}$. $\frac{N}{4} = \frac{100}{4} = 25$. La 25^e valeur est 2 d'où $Q_1 = 2$.
   - Le troisième quartile Q_3 est la valeur du caractère dont le rang est le plus petit entier supérieur ou égal à $\frac{3N}{4}$. $\frac{3N}{4} = \frac{3\times100}{4} = 75$. La 75^e valeur est 6 d'où $\boxed{Q_3 = 6}$.

   (a) Son adjoint souhaite ouvrir une caisse supplémentaire si plus de 15 % des clients attendent 7 min ou plus en caisse.
   Les personnes qui attendent au moins 7 minutes sont au nombre de 12+4+1+2= 19. 19 % attendent 7 min ou plus, il doit ouvrir une nouvelle caisse le lundi.

   (b) Le directeur décide d'ouvrir une caisse supplémentaire si le temps moyen d'attente aux caisses dépasse 5 min.
   À la question 1., on a calculé que le temps d'attente moyen était de 4 minutes environ. Il ne doit donc pas ouvrir une nouvelle caisse le lundi.

PARTIE B : Étude de l'échantillon du vendredi

Temps d'attente le vendredi



1. $\overline{x} = \frac{1\times5+2\times9+\cdots+11\times2+12\times1}{100} \approx 5{,}68$. Le temps moyen d'attente le vendredi est de 5,7 min à 0,1 min près
2. **Affirmation A :**

   Le vendredi, la moitié des clients attendent cinq min ou plus de cinq min en caisse.

VRAI la médiane de la série est 5, donc la moitié des personnes attendent cinq minutes ou plus;

Affirmation B :

Le vendredi, un quart des clients attend au plus trois minutes en caisse.

VRAI Le premier quartile est 3, par conséquent 25% des valeurs lui sont inférieures ou égales.

Affirmation C :

Il y a autant de clients qui trouvent le temps d'attente acceptable le lundi que le vendredi.

FAUX Le lundi, ils sont 13 + 23 + 9+ 14 + 8 = 67

Le vendredi, ils sont 9 + 13 + 8 + 19 + 10 = 59

Tableau de la série du vendredi

Temps d'attente en caisse (en min)	1	2	3	4	5	6	7	8	9	10	11	12
Nombre de clients	5	9	**13**	8	**19**	10	**8**	**5**	**11**	9	2	1