CORRIGE DES EXERCICES : Statistique descriptive univariée

Exercice 1

$\mathcal{P}$={patients hospitalisés pour un état dépressif majeur} de taille (effectif total) N =12

1) a. X= efficacité du traitement, variable qualitative dichotomique (deux modalités) définie sur l'échelle descriptive E={x_1, x_2}={efficace, inefficace} = {oui, non} ou par les deux modalités : x_1 = efficace, x_2 = inefficace

b. tableau des distributions d'effectifs (n_i, i=1,2) et de proportions (p_i, i=1,2)

efficacité	effectif n_i	proportion p_i
x_1 = efficace	9	9/12=0,75=75%
x_2 = inefficace	3	3/12=0,25=25%
total	12	1,00=100%

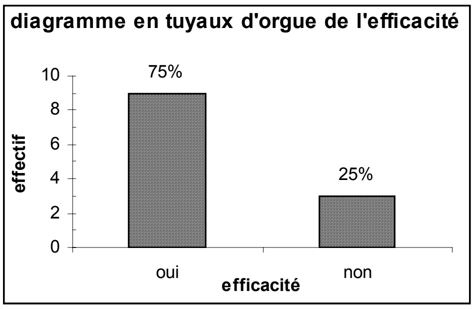
$n_1 + n_2 = N$

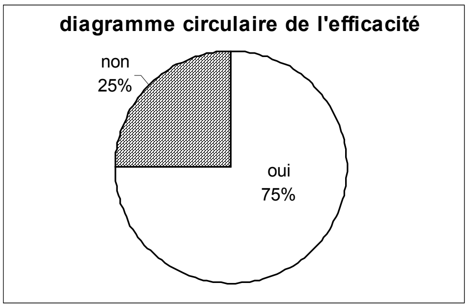
$p_1 + p_2 = 1$ on note : $p_1 = p$ et $p_2 = 1-p$

➨ la proportion d'efficacité du traitement (notée p) est égale à 75% dans la population des 12 patients hospitalisés pour un état dépressif majeur, et la proportion d'inefficacité du traitement est donc égale à 25% ($1-p$) dans la population étudiée.

c. Le mode est la modalité d'effectif (ou de proportion) maximum : le mode est donc x_1 = efficace (pour 75% des individus de la population).

d. Plusieurs types de représentations graphiques possibles : diagramme en tuyaux d'orgue, diagramme circulaire (en secteurs, camembert), diagramme rectangulaire, soit des effectifs (car l'effectif total est faible) soit des proportions.





2) a. X = score à l'inventaire psychiatrique standardisé (PSE), variable quantitative discrète définie sur l'échelle descriptive E={x_1, x_2, x_3, x_4}={4,5,6,7} : x_1 = 4, x_2 = 5, x_3 = 6 et x_4 = 7 (4 valeurs numériques (entières) représentant des points de score).

b. Tableau de la distribution d'effectifs (n_i, i=1,…4)

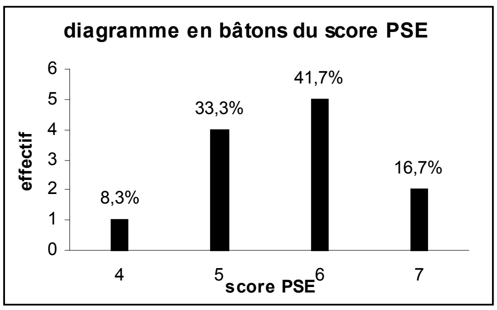
score PSE	effectif n_i
x_1 = 4	1
x_2 = 5	4
x_3 = 6	5
x_4 = 7	2
total	Σn_i = 12

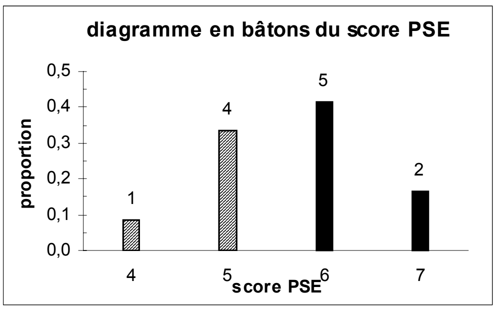
c. Le mode est la modalité d'effectif maximum : le mode est donc x_3 = 6 points de score (pour 5 individus de la population).

d. On représente par un diagramme en bâtons, soit la distribution d'effectifs (car l'effectif total est faible) soit celle des proportions.

tableau des distributions d'effectifs (n_i, i=1,…4) et de proportions (p_i, i=1,…4)

score PSE	effectif n_i	proportion p_i
x_1 = 4	1	1/12=0,083=8,3%
x_2 = 5	4	4/12=0,333=33,3%
x_3 = 6	5	5/12=0,417=41,7%
x_4 = 7	2	2/12=0,167=16,7%
total	12	Σp_i = 1,000=100%





e. La proportion de patients ayant un score PSE au plus égal à 5 (bâtons hachurés sur le diagramme des proportions) s'écrit :

$$P(X \le 5) = P(X = 4) + P(X = 5) \begin{cases} = \dfrac{1}{12} + \dfrac{4}{12} = \dfrac{5}{12} = 0{,}417 \\ \approx 0{,}083 + 0{,}333 = 0{,}416 \end{cases}$$ (c'est aussi la proportion cumulée F_2).

En utilisant la fonction de répartition F de la variable X, on note cette proportion P(X≤5)=F(5).

La proportion de patients ayant un score PSE au plus égal à 6 (bâtons noirs sur le diagramme des proportions) s'écrit

$$P(X \le 6) = P(X = 4) + P(X = 5) + P(X = 6) \begin{cases} = \dfrac{1}{12} + \dfrac{4}{12} + \dfrac{5}{12} = \dfrac{10}{12} = 0{,}833 \\ \approx 0{,}083 + 0{,}333 + 0{,}417 = 0{,}833 \end{cases}$$ (c'est aussi la proportion cumulée F_3).

En utilisant la fonction de répartition F de la variable X, on note cette proportion P(X≤6)=F(6).

f. La proportion de patients ayant un score PSE au moins égal à 6 s'écrit :

$$P(X \ge 6) = P(X = 6) + P(X = 7) \begin{cases} = \dfrac{5}{12} + \dfrac{2}{12} = \dfrac{7}{12} = 0{,}583 \\ \approx 0{,}417 + 0{,}167 = 0{,}584 \end{cases}$$

En utilisant la fonction de répartition F de la variable X, on note cette proportion P(X≥6) = 1−P(X<6) = 1−P(X≤5) = 1−F(5).

g. Moyenne de X dans $\mathcal{P}$ = moyenne du score PSE dans $\mathcal{P}$ = score PSE moyen dans $\mathcal{P}$ = μ

variance de X dans $\mathcal{P}$ = variance du score PSE dans $\mathcal{P}$ = σ^2

écart-type de X dans $\mathcal{P}$ = écart-type du score PSE dans $\mathcal{P}$ = σ

score PSE	effectif n_i	$n_i x_i$	x_i^2	$n_i x_i^2$
$x_1 = 4$	1	4	16	16
$x_2 = 5$	4	20	25	100
$x_3 = 6$	5	30	36	180
$x_4 = 7$	2	14	49	98
total	12	$\Sigma n_i x_i = 68$		$\Sigma n_i x_i^2 = 394$

donc $\mu = \dfrac{\sum_{i=1}^{4} n_i x_i}{N} = \dfrac{68}{12} = 5{,}667 \approx 5{,}7$ $\sigma^2 = \dfrac{\sum_{i=1}^{4} n_i x_i^2}{N} - \mu^2 = \dfrac{394}{12} - 5{,}7^2 = 32{,}833 - 32{,}111 = 0{,}7222 \approx 0{,}72$ et

$\sigma = \sqrt{\sigma^2} = \sqrt{0{,}72} = 0{,}8498 \approx 0{,}85$

➨ dans la population des 12 patients hospitalisés pour un état dépressif majeur : le score PSE moyen est de 5,7 points, la variance du score PSE est de 0,72 et l'écart-type du score PSE est de 0,85 point.

Exercice 2

1) $\mathcal{P}$={sujets} de taille (effectif total) N=12+19+0+13+6=50

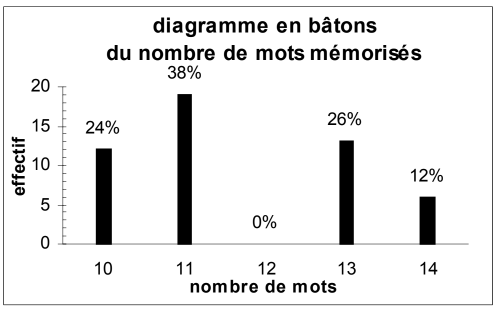
2) X = nombre de mots mémorisés, définie sur l'échelle descriptive E={x_1, x_2, x_3, x_4, x_5}={10, 11, 12, 13, 14} : x_1 = 10, x_2 = 11, x_3 = 12, x_4 = 13 et x_5 = 14 (5 valeurs numériques (entières)), c'est donc une variable quantitative discrète.

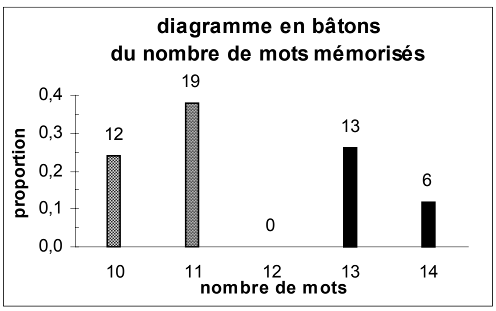
3) Tableau de la distribution de proportions (p_i, i=1,…5)

nombre de mots	effectif n_i	proportion p_i
$x_1 = 10$	12	12/50=0,24=24%
$x_2 = 11$	19	19/50=0,38=38%
$x_3 = 12$	0	0/50=0=0%
$x_4 = 13$	13	13/50=0,26=26%
$x_5 = 14$	6	13/50=0,12=12%
total	$\Sigma n_i = 50$	Σp_i = 1,000=100%

4) Le mode est la modalité d'effectif (ou de proportion) maximum : le mode est donc x_2 = 11 mots mémorisés (pour 19 individus de la population, soit 38%).

5) On représente par un diagramme en bâtons, soit la distribution d'effectifs (car l'effectif total est faible) soit celle des proportions.





6) La proportion de patients ayant mémorisé au plus 11 mots (bâtons hachurés sur le diagramme des proportions) s'écrit :

$$P(X \le 11) = P(X = 10) + P(X = 11) = \begin{cases} \frac{12}{50} + \frac{19}{50} = \frac{31}{50} = 0{,}62 \\ 0{,}24 + 0{,}38 = 0{,}62 \end{cases}$$ (c'est aussi la proportion cumulée F_2).

En utilisant la fonction de répartition F de la variable X, on note cette proportion P(X≤11)=F(11).
La proportion de patients ayant mémorisé au moins 12 mots (bâtons noirs sur le diagramme des proportions) s'écrit :

$$P(X \ge 12) = P(X = 12) + P(X = 13) + P(X = 14) = \begin{cases} \frac{0}{50} + \frac{13}{50} + \frac{6}{50} = \frac{19}{50} = 0{,}38 \\ 0 + 0{,}26 + 0{,}12 = 0{,}38 \end{cases}$$

En utilisant la fonction de répartition F de la variable X, on note cette proportion P(X≥12)=1−P(X<12)= 1−P(X≤11) =1−F(11).

7) Moyenne de X dans $\mathcal{P}$ = nombre moyen de mots mémorisés dans $\mathcal{P}$ = μ
variance de X dans $\mathcal{P}$ = variance du nombre de mots mémorisés dans $\mathcal{P}$ = σ^2
écart-type de X dans $\mathcal{P}$ = écart-type du nombre de mots mémorisés dans $\mathcal{P}$ = σ

nombre de mots	effectif n_i	$n_i\,x_i$	x_i^2	$n_i\,x_i^2$
$x_1 = 10$	12	120	100	1200
$x_2 = 11$	19	209	121	2299
$x_3 = 12$	0	0	144	0
$x_4 = 13$	13	169	169	2197
$x_5 = 14$	6	84	196	1176
total	50	$\Sigma n_i x_i = 582$		$\Sigma n_i x_i^2 = 6872$

donc $\mu = \frac{582}{50} = 11{,}64 \approx 11{,}6$ $\sigma^2 = \frac{6872}{50} - 11{,}6^2 = 137{,}44 - 135{,}49 = 1{,}9504 \approx 1{,}95$ et $\sigma = \sqrt{1{,}95} = 1{,}3966 \approx 1{,}4$

➨ les 50 sujets mémorisent en moyenne 11,6 mots, la variance du nombre de mots mémorisés est de 1,95 et l'écart-type du nombre de mots mémorisés est de 1,4 mot.

<u>Exercice 3</u>

$\mathcal{P}$={adolescents habitant l'état de Caroline du Sud un an après l'ouragan Hugo} de taille (effectif total) N =1 264

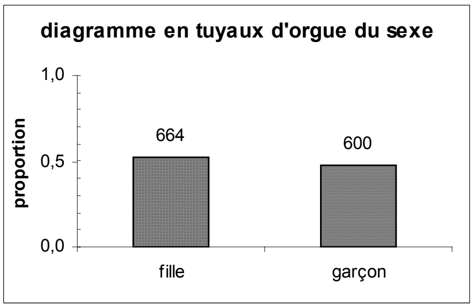
1) a. X= sexe, variable qualitative dichotomique définie sur l'échelle descriptive E={x_1, x_2}={fille, garçon} ou deux modalités : x_1 = fille, x_2 = garçon
tableau des distributions d'effectifs (n_i, i=1,2) et de proportions (p_i, i=1,2) car le nombre de garçons $n_2=N-n_1$

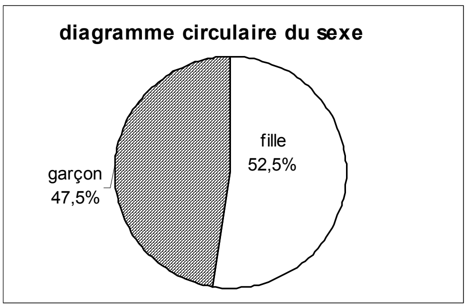
sexe	effectif n_i	proportion p_i
x_1 = fille	n_1 = 664	664/1264=0,525=52,5%
x_2 = garçon	n_2 = 1264−664=600	600/1264=0,475=47,5%
total	N = 1264	1,00=100%

$n_1 + n_2 = N$
$p_1 + p_2 = 1$ on note : $p_1 = p$ et $p_2 = 1-p$

➨ dans la population des 1264 adolescents étudiée, la proportion de filles (notée p) est égale à 52,5% et donc la proportion de garçons est égale à 47,5% (1−p).

b. Plusieurs types de représentations graphiques possibles de la distribution de proportions (car l'effectif total est élevé) : diagramme en tuyaux d'orgue, diagramme circulaire (en secteurs, camembert), diagramme rectangulaire.





2) a. X= revivre l'événement, variable qualitative dichotomique définie sur l'échelle descriptive E={x_1, x_2}={oui, non} ou deux modalités : x_1 = revivre l'événement, x_2 = ne pas revivre l'événement
tableau des distributions d'effectifs (n_i, i=1,2) et de proportions (p_i, i=1,2) car :
le nombre d'adolescents qui revivent l'événement : n_1=161+82= 243
le nombre d'adolescents qui ne revivent pas l'événement : $n_2=N-n_1$=1264−243=1021

revivre l'événement	effectif n_i	proportion p_i
x_1 = oui	243	243/1264=0,192=19,2%
x_2 = non	1264−243=1021	1021/1264=0,808=80,8%
total	1264	1,00=100%

$n_1 + n_2 = N$
$p_1 + p_2 = 1$ d'où : $p = p_1$ et $1-p = p_2$

le mode est la modalité x_2 = ne pas revivre l'événement

➨ dans la population des 1264 adolescents étudiée, la proportion d'adolescents qui revivent l'événement (notée p) est égale à 19,2% et la proportion de ceux qui ne revivent pas l'événement est donc égale à 80,8% (1-p).

b. $\mathcal{P}$={adolescentes (filles) habitant l'état de Caroline du Sud un an après l'ouragan Hugo} de taille $N = 664$
X= revivre l'événement, variable qualitative dichotomique définie sur l'échelle descriptive E={x_1, x_2}={oui, non}
tableau des distributions d'effectifs (n_i, i=1,2) et de proportions (p_i, i=1,2) car :
le nombre d'adolescentes qui revivent l'événement : n_1=161
le nombre d'adolescentes qui ne revivent pas l'événement : $n_2=N-n_1$=664−161=503

revivre l'événement	effectif n_i	proportion p_i
x_1 = oui	161	161/664=0,242=24,2%
x_2 = non	664−161=503	503/664=0,758=75,8%
total	664	1,00=100%

c. $\mathcal{P}$={adolescents (garçons) habitant l'état de Caroline du Sud un an après l'ouragan Hugo} de taille $N = 600$
X= revivre l'événement, variable qualitative dichotomique définie sur l'échelle descriptive E={x_1, x_2}={oui, non}
tableau des distributions d'effectifs (n_i, i=1,2) et de proportions (p_i, i=1,2) car :
le nombre d'adolescentes qui revivent l'événement : n_1=82
le nombre d'adolescentes qui ne revivent pas l'événement : $n_2=N-n_1$=600−82=518

revivre l'événement	effectif n_i	proportion p_i
x_1 = oui	82	82/600=0,137=13,7%
x_2 = non	600−82=518	518/600=0,863=86,3%
total	600	1,00=100%

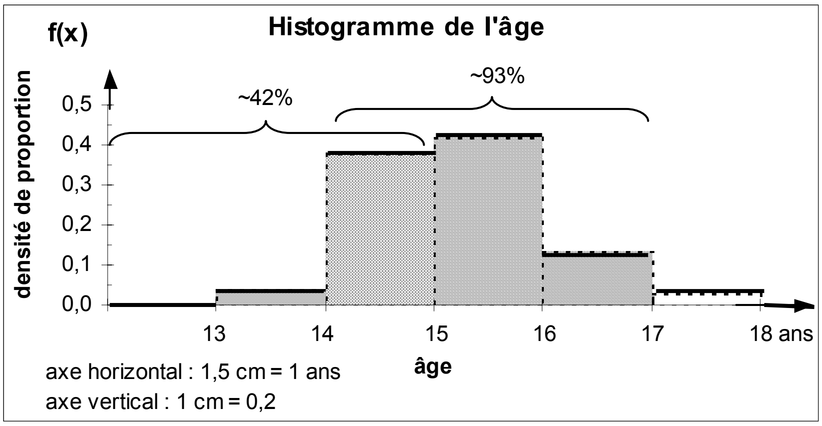
Dans les deux sous-populations filles ou garçons, le mode est le même : "ne pas revivre l'événement", mais la proportion correspondante est plus élevée chez les garçons (86,3%) que chez les filles (75,8%) : les filles revivent plus souvent l'événement (24,2%) que les garçons (13,7%). La proportion de filles qui revivent l'événement est égale à 24,2%, celle des garçons est égale à 13,7%.

3) a. X = âge, définie sur l'échelle descriptive E=]13; 18] (valeurs numériques continues), c'est donc une variable quantitative continue.

b. Tableau de la distribution de proportions de l'âge associée au découpage en 5 intervalles (p_i, i=1,…5)

i	âge	effectif n_i	proportion p_i
1	]13; 14]	49	0,039=3,9%
2	]14; 15]	479	0,379=37,9%
3	]15; 16]	529	0,418=41,8%
4	]16; 17]	169	0,134=13,4%
5	]17; 18]	38	0,03=3%
	total	1264	1,000=100%

c. La densité de proportion pour chaque intervalle est égale à : $f(x) = p_i / l_i$ où l_i est la longueur du i$^{\text{ème}}$ intervalle. Ici tous les intervalles sont de longueur 1 an donc $f(x) = p_i$.
La représentation graphique de la densité de proportion est donnée par l'histogramme associé à ce découpage.



d. L'intervalle modal est l'intervalle de densité maximale : c'est donc le 3$^{\text{ème}}$ intervalle]15; 16].

e. La proportion d'adolescents de moins de 15 ans (aire hachurée ▨ sur l'histogramme) s'écrit :

$$P(X \leq 15) = P(13 < X \leq 14) + P(14 < X \leq 15) \begin{cases} = \dfrac{49}{1264} + \dfrac{479}{1264} = \dfrac{528}{1264} \approx 0{,}418 \approx 0{,}42 \\ \approx 0{,}039 + 0{,}379 = 0{,}418 \approx 0{,}42 \end{cases}$$

f. La proportion d'adolescents de plus de 15 ans s'écrit :

$$P(X > 15) = \begin{cases} P(15 < X \leq 16) + P(16 < X \leq 17) + P(17 < X \leq 18) \begin{cases} = \dfrac{529}{1264} + \dfrac{169}{1264} + \dfrac{38}{1264} = \dfrac{736}{1264} \approx 0{,}582 \approx 0{,}58 \\ \approx 0{,}418 + 0{,}134 + 0{,}03 = 0{,}582 \approx 0{,}58 \end{cases} \\ 1 - P(X \leq 15) = 1 - 0{,}418 = 0{,}582 \end{cases}$$

g. La proportion d'adolescents âgés de 14 à 17 ans (aire hachurée ▧ sur l'histogramme) s'écrit :

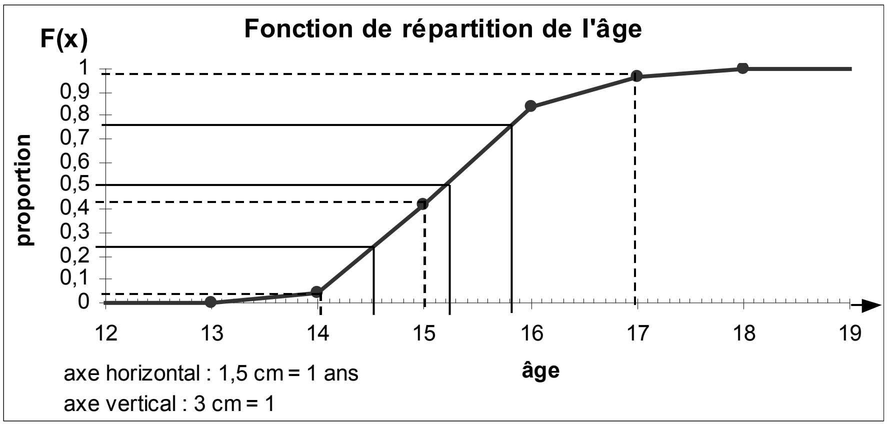
$$P(14 < X \le 17) = P(14 < X \le 15) + P(15 < X \le 16) + P(16 < X \le 17) = \begin{cases} \frac{479}{1264} + \frac{529}{1264} + \frac{169}{1264} = \frac{1177}{1264} \approx 0{,}931 \approx 0{,}93 \\ 0{,}379 + 0{,}418 + 0{,}134 = 0{,}931 \approx 0{,}93 \end{cases}$$

h. Tableau de la distribution de proportions cumulées de l'âge, associée au découpage en 5 intervalles (F_i, i=1,…5) : pour chaque intervalle la proportion cumulée correspondante $F_i = N_i/N$ où N_i est l'effectif cumulé $N_i=n_1+n_2+\ldots+n_i$ ou $F_i=p_1+p_2+\ldots+p_i$

i	âge	effectif n_i	proportion p_i	effectif cumulé N_i	proportion cumulée F_i
1	]13; 14]	49	0,039=3,9%	$N_1=n_1=49$	$F_1=p_1=0{,}039$
2	]14; 15]	479	0,379=37,9%	$N_2=n_1+n_2=49+479=528$	$F_2=p_1+p_2=0{,}039+0{,}379=0{,}418$
3	]15; 16]	529	0,418=41,8%	$N_3=n_1+n_2+n_3=N_2+n_3=528+529=1057$	$F_3=p_1+p_2+p_3=0{,}418+0{,}418=0{,}836$
4	]16; 17]	169	0,134=13,4%	$N_4=n_1+n_2+n_3+n_4=N_3+n_4=1057+169=1226$	$F_4=p_1+p_2+p_3+p_4=0{,}836+0{,}134=0{,}97$
5	]17; 18]	38	0,03=3%	$N_5=n_1+n_2+n_3+n_4+n_5=N=1264$	$F_5=p_1+p_2+p_3+p_4+p_5=1$
	total	1264	1,000=100%		

i. La fonction de répartition de l'âge associée au découpage en 5 intervalles, est définie par $F(x) = P(X\le x)$ donc :

$$\begin{cases} F(x)=0 \text{ pour tout } x \le 13 \\ F(14)= P(X\le14)=F_1=0{,}039 \\ F(15)= P(13<X\le14)+P(14<X\le15)=p_1+p_2=F_2=0{,}418 \\ F(16)= P(13<X\le14)+P(14<X\le15)+P(15<X\le16)=p_1+p_2+p_3=F_3=0{,}836 \\ F(17)= P(13<X\le14)+P(14<X\le15)+P(15<X\le16)+P(16<X\le17)=p_1+p_2+p_3+p_4=F_4=0{,}97 \\ F(18)= P(13<X\le14)+P(14<X\le15)+P(15<X\le16)+P(16<X\le17)+P(17<X\le18)=p_1+p_2+p_3+p_4+p_5=F_5=1 \\ F(x)=1 \text{ pour tout } x > 18 \end{cases}$$



j. La proportion d'adolescents de moins de 15 ans s'exprime comme la proportion cumulée F_2. En utilisant la fonction de répartition F de la variable X, on note cette proportion $P(X\le15)=F(15)=0{,}418$

La proportion d'adolescents de plus de 15 ans s'écrit en utilisant la fonction de répartition F de la variable X :
$P(X>15)=1-P(X\le15)=1-F(15)=1-0{,}418=0{,}582$

La proportion d'adolescents âgés de 14 à 17 ans s'écrit en utilisant la fonction de répartition F de la variable X :
$P(14 < X \le 17) = P(X \le 17) - P(X \le 14) = F(17) - F(14)$. On calcule séparément F(17) puis F(14) :

$$F(17) = P(X \le 17) = P(13 < X \le 14) + P(14 < X \le 15) + P(15 < X \le 16) + P(16 < X \le 17) = F_4 \approx 0{,}9699$$

$$F(14) = P(X \le 14) = P(13 < X \le 14) = F_1 = 0{,}039$$

donc $P(14 < X \le 17) = F(17) - F(14) = 0{,}9699 - 0{,}0388 = 0{,}9311 \approx 0{,}93$

k. Les trois quartiles de l'âge sont représentés en abscisse sur le graphique de la fonction de répartition :

la médiane (second quartile) notée $Q_{0,5}$ est telle que $F(Q_{0,5})=0{,}5$ donc graphiquement on déduit que $Q_{0,5} \approx 15{,}2$ ans

le premier quartile noté $Q_{0,25}$ est tel que $F(Q_{0,25})=0{,}25$ donc graphiquement on déduit que $Q_{0,25} \approx 14{,}5$ ans

le troisième quartile noté $Q_{0,75}$ est tel que $F(Q_{0,75})=0{,}75$ donc graphiquement on déduit que $Q_{0,75} \approx 15{,}8$ ans

➨ 50% des adolescents de la population ont moins de 15,2 ans (50% ont plus de 15,2 ans), 25% des adolescents ont moins de 14,5 ans (75% ont plus de 14,5 ans) et 75% des adolescents ont moins de 15,8 ans (25% ont plus de 15,8 ans).

l. Moyenne de X dans $\mathcal{P}$ = moyenne de l'âge dans $\mathcal{P}$ = âge moyen dans $\mathcal{P}$ =μ

variance de X dans $\mathcal{P}$ = variance du nombre de mots mémorisés dans $\mathcal{P}$ = σ^2

écart-type de X dans $\mathcal{P}$ = écart-type du nombre de mots mémorisés dans $\mathcal{P}$ =σ

i	âge	effectif n_i	milieu x_i	$n_i\, x_i$	x_i^2	$n_i\, x_i^2$
1	]13; 14]	49	13,5	661,5	182,25	8930,25
2	]14; 15]	479	14,5	6945,5	210,25	100709,75
3	]15; 16]	529	15,5	8199,5	240,25	127092,25
4	]16; 17]	169	16,5	2788,5	272,25	46010,25
5	]17; 18]	38	17,5	665,0	306,25	11637,5
	total	1264		$\Sigma n_i\, x_i=19260{,}0$		$\Sigma n_i\, x_i^2=294380{,}0$

donc $\mu = \dfrac{19260}{1264} = 15{,}24 \approx 15{,}2$ $\sigma^2 = \dfrac{294380}{1264} - 15{,}2^2 = 232{,}896 - 232{,}177 = 0{,}719 \approx 0{,}72$ et

$\sigma = \sqrt{0{,}72} = 0{,}8479 \approx 0{,}85$

➡ les 1264 adolescents ont en moyenne 15,2 ans, la variance de l'âge est de 0,72 et l'écart-type de l'âge est de 0,85 an.

Exercice 4

$\mathcal{P}$={sujets fumeurs} de taille (effectif total) N =60

1) Le diagramme est appelé diagramme tige et feuille (stem and leaf). La valeur minimale est 3, la valeur maximale 59.

2) X= durée de tabagisme, définie sur l'échelle descriptive E=]0; 60] (valeurs numériques continues), c'est donc une variable quantitative continue.

3) Moyenne de X dans $\mathcal{P}$ = moyenne de la durée de tabagisme dans $\mathcal{P}$ = durée moyenne de tabagisme dans $\mathcal{P}$ = μ

variance de X dans $\mathcal{P}$ = variance de la durée de tabagisme dans $\mathcal{P}$ = σ^2

écart-type de X dans $\mathcal{P}$ = écart-type de la durée de tabagisme dans $\mathcal{P}$ = σ

$$\sum_{i=1}^{N} x_i = 1382 \text{ d'où } \mu = \frac{\sum_{i=1}^{N} x_i}{N} = \frac{1382}{60} = 23{,}03 \approx 23$$

$$\sum_{i=1}^{N} x_i^2 = 41528 \text{ d'où } \sigma^2 = \frac{\sum_{i=1}^{N} x_i^2}{N} - \mu^2 = \frac{41528}{60} - 23^2 = 692{,}13 - 530{,}53 = 161{,}6 \text{ et } \sigma = \sqrt{\sigma^2} = \sqrt{161{,}6} = 12{,}7$$

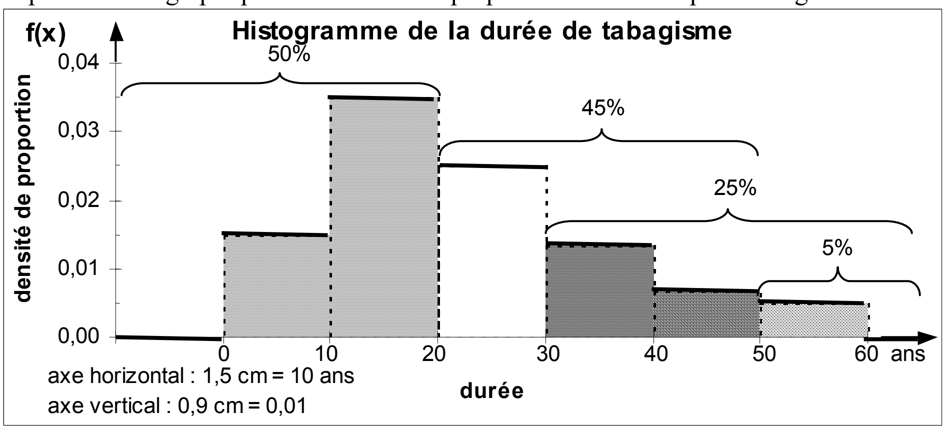
➡ les 60 fumeurs fument en moyenne depuis 23 ans, la variance de la durée de tabagisme est de 161,6 et l'écart-type de la durée de tabagisme est de 12,7 ans.

4) a. Tableau des distributions d'effectifs (n_i, i=1,…6) et de proportions (p_i, i=1,…6) de la durée de tabagisme, associées au découpage en 6 intervalles de longueur 10 ans.

i	durée de tabagisme	effectif n_i	proportion p_i
1	]0; 10]	9	0,15=15%
2	]10; 20]	21	0,35=35%
3	]20; 30]	15	0,25=25%
4	]30; 40]	8	0,133=13,3%
5	]40; 50]	4	0,067=6,7%
6	]50; 60]	3	0,05=5%
	total	60	1,000=100%

b. La densité de proportion pour x appartenant au i$^{\text{ème}}$ l'intervalle est égale à : $f(x) = p_i / l_i$ où l_i est la longueur du i$^{\text{ème}}$ intervalle. Ici tous les intervalles sont de longueur 10 ans donc $f(x) = p_i / 10$.

La représentation graphique de la densité de proportion est donnée par l'histogramme associé à ce découpage.



c. L'intervalle modal est l'intervalle de densité maximale : c'est donc le 2$^{\text{ème}}$ intervalle]10; 20].

5) a. La proportion de sujets ayant fumé pendant moins de 20 ans (aire hachurée sur l'histogramme) s'écrit :

$$P(X \le 20) = P(0 < X \le 10) + P(10 < X \le 20) = \begin{cases} \dfrac{9}{60} + \dfrac{21}{60} = \dfrac{30}{60} = 0{,}5 \\ 0{,}15 + 0{,}35 = 0{,}5 \end{cases}$$ (c'est aussi la proportion cumulée F_2).

En utilisant la fonction de répartition F de la variable X, on note cette proportion P(X≤20)=F(20).

b. La proportion de sujets ayant fumé pendant plus de 20 ans s'écrit :

$$P(X > 20) = \begin{cases} P(20 < X \le 30) + P(30 < X \le 40) + P(40 < X \le 50) + P(50 < X \le 60) = \begin{cases} \dfrac{15}{60} + \dfrac{8}{60} + \dfrac{4}{60} + \dfrac{3}{60} = \dfrac{30}{60} = 0{,}5 \\ 0{,}25 + 0{,}133 + 0{,}067 + 0{,}05 = 0{,}5 \end{cases} \\ 1 - P(X \le 20) = 1 - 0{,}5 = 0{,}5 \end{cases}$$

En utilisant la fonction de répartition F de la variable X, on note cette proportion P(X>20)=1−P(X≤20)=1−F(20).

c. La proportion de sujets ayant fumé pendant plus de 30 ans (aire hachurée sur l'histogramme) s'écrit :

$$P(X>30)=P(30<X\le 40)+P(40<X\le 50)+P(50<X\le 60)=\begin{cases}\dfrac{8}{60}+\dfrac{4}{60}+\dfrac{3}{60}=\dfrac{15}{60}=0{,}25\\ 0{,}133+0{,}067+0{,}05=0{,}25\end{cases}$$

En utilisant la fonction de répartition F de la variable X, on note cette proportion P(X>30)=1−P(X≤30)=1−F(30).

La proportion de sujets ayant fumé pendant plus de 50 ans (aire hachurée sur l'histogramme) s'écrit :

$$P(X>50)=P(50<X\le 60)=\frac{3}{60}=0{,}05$$

En utilisant la fonction de répartition F de la variable X, on note cette proportion P(X>50)=1−P(X≤50)=1−F(50).

d. La proportion de sujets ayant fumé pendant plus de 20 ans mais moins de 50 ans s'écrit :

$$P(20<X\le 50)=\begin{cases}P(20<X\le 30)+P(30<X\le 40)+P(40<X\le 50)=\begin{cases}\dfrac{15}{60}+\dfrac{8}{60}+\dfrac{4}{60}=\dfrac{27}{60}=0{,}45\\ 0{,}25+0{,}133+0{,}067=0{,}45\end{cases}\\ \\ P(X\le 50)-P(X\le 20)=F(50)-F(20)\end{cases}$$

où F est la fonction de répartition de la variable X.

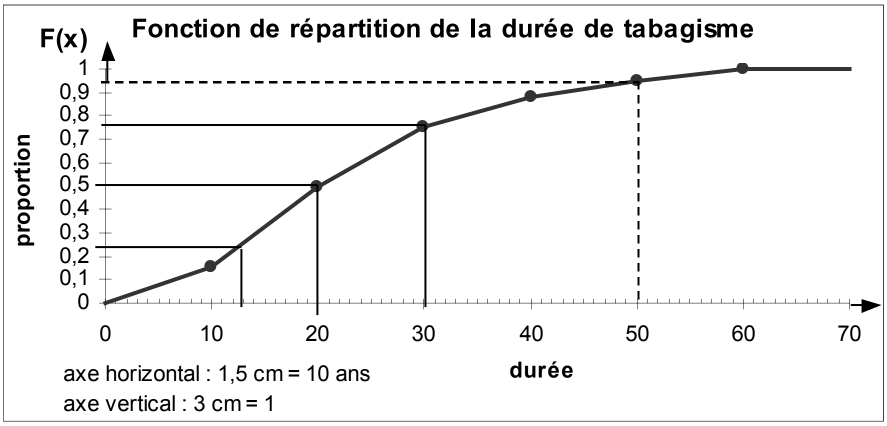
6) a. La fonction de répartition de la durée de tabagisme est définie par $F(x) = P(X\le x)$ donc :

$F(x)=0$ pour tout $x \le 0$

$F(x)=1$ pour tout $x > 60$

et à partir de la distribution des proportions cumulées associé au découpage en 6 intervalles (F_i, i=1,…6) : $F(x)=F_i$ si x est la borne supérieure du i[ème] intervalle

i	durée	effectif n_i	proportion p_i	effectif cumulé N_i	proportion cumulée F_i
1	]0; 10]	9	0,15=15%	$N_1=n_1=9$	$F_1=p_1=0{,}15=F(10)$
2	]10; 20]	21	0,35=35%	$N_2=n_1+n_2=9+21=30$	$F_2=30/60=0{,}5=F(20)$
3	]20; 30]	15	0,25=25%	$N_3=n_1+n_2+n_3=N_2+n_3=30+15=45$	$F_3=45/60=0{,}75=F(30)$
4	]30; 40]	8	0,133=13,3%	$N_4=n_1+n_2+n_3+n_4=N_3+n_4=45+8=53$	$F_4=53/60=0{,}883=F(40)$
5	]40; 50]	4	0,067=6,7%	$N_5=n_1+n_2+n_3+n_4+n_5= N_4+n_5=53+4=57$	$F_5=57/60=0{,}95=F(50)$
6	]50; 60]	3	0,05=5%	$N_6=n_1+n_2+n_3+n_4+n_5+n_6= N=60$	$F_6=60/60=1=F(60)$
	total	60	1,000=100%		



b. Puisque F(20)=0,5 : 50% des sujets ont fumé pendant moins de 20 ans, la médiane de la durée de tabagisme est de 20 ans, $Q_{0,5}=20$

puisque F(30)=0,75 : 75% des sujets ont fumé pendant moins de 30 ans, le troisième quartile de la durée de tabagisme est de 30 ans, $Q_{0,75}=30$

puisque F(50)=0,95 : 95% des sujets ont fumé pendant moins de 50 ans, le quantile d'ordre 0,95 (95%) ou 95[ème] percentile de la durée de tabagisme est de 50 ans, $Q_{0,95}=50$

c. Les trois quartiles de l'âge sont représentés en abscisse sur le graphique de la fonction de répartition :

la médiane (second quartile) $Q_{0,5}$ = 20 ans, le troisième quartile $Q_{0,75} \approx$ 30 ans

le premier quartile noté $Q_{0,25}$ est telle que $F(Q_{0,25})=0{,}25$ donc graphiquement on déduit que $Q_{0,25} \approx 13$ ans

➨ 50% des sujets ont fumé pendant moins de 20 ans (50% pendant plus de 20 ans), 25% des sujets ont fumé pendant moins de 12,5 ans (75% pendant plus de 12,5 ans) et 75% des sujets ont fumé pendant moins de 30 ans (25% pendant plus de 30 ans).

7) Sur le diagramme tige et feuille, un quantile est déterminé en comptant (à partir de la valeur minimale) le nombre de valeurs correspondant à son ordre.

a. La médiane, quantile d'ordre 0,5 (50%) correspond donc à la moitié de la taille de la population c'est à dire $N/2=30$: elle est donc égale à n'importe quelle valeur comprise entre la 30[ème] et la 31[ème] valeur c'est à dire entre 20 et 21; par convention on prendra le milieu, donc $Q_{0,5} \approx 20{,}5$ ans.

le premier quartile, quantile d'ordre 0,25 (25%) correspond au quart de la taille de la population c'est à dire $N/4=15$: il est donc égal à n'importe quelle valeur comprise entre la $15^{ème}$ et la $16^{ème}$ valeur c'est à dire 14 ; donc $Q_{0,25} = 14$ ans.
le troisième quartile, quantile d'ordre 0,75 (75%) correspond aux trois quarts de la taille de la population c'est à dire $3N/4=45$: il est donc égal à n'importe quelle valeur comprise entre la $45^{ème}$ et la $46^{ème}$ valeur c'est à dire entre 29 et 31; par convention on prendra le milieu, donc $Q_{0,75} \approx 30$ ans.
On retrouve des quartiles très similaires à ceux déduits à la question 6.

b. Le 1^{er} décile, quantile d'ordre 0,1 (10%) correspond au dixième de la taille de la population c'est à dire $N/10=6$: il est donc égal à n'importe quelle valeur comprise entre la $6^{ème}$ et la $7^{ème}$ valeur c'est à dire entre 9 et 10 ; par convention on prendra le milieu, donc $Q_{0,1} \approx 9,5$ ans.
le $9^{ème}$ décile, quantile d'ordre 0,9 (90%) correspond aux neuf dixièmes de la taille de la population c'est à dire $N\times0,9=6\times9=54$: il est donc égal à n'importe quelle valeur comprise entre la $54^{ème}$ et la $55^{ème}$ valeur c'est à dire entre 42 et 45 ; par convention on prendra le milieu, donc $Q_{0,9} \approx 43,5$ ans.
L'intervalle de variation à 80% (au risque 20%) de la durée de tabagisme est défini par les deux quantiles d'ordre 0,1 et 0,9 : $I_{80\%} = [Q_{0,1} ; Q_{0,9}]$ puisque $P(X \in I_{80\%}) = P(Q_{0,1} \leq X \leq Q_{0,9})=80\%$ et $P(X \notin I_{80\%}) = 20\%$
donc $I_{80\%} = [Q_{0,1} ; Q_{0,9}] = [9,5 ; 43,5]$

➨ 10% des sujets ont fumé pendant moins de 9,5 ans (90% pendant plus de 9,5 ans), 90% des sujets ont fumé pendant moins de 43,5 ans (10% pendant plus de 43,5 ans) et 80% des sujets ont fumé pendant une durée comprise entre 9,5 et 43,5 ans.

c. Le $5^{ème}$ percentile, quantile d'ordre 0,05 (5%) correspond à 5% de la taille de la population c'est à dire $N\times0,05=3$: il est donc égal à n'importe quelle valeur comprise entre la $3^{ème}$ et la $4^{ème}$ valeur c'est à dire 5 ; donc $Q_{0,05} = 5$ ans.
le $95^{ème}$ percentile, quantile d'ordre 0,95 (95%) correspond à 95% de la taille de la population c'est à dire $N\times0,95=60\times0,95=57$: il est donc égal à n'importe quelle valeur comprise entre la $57^{ème}$ et la $58^{ème}$ valeur c'est à dire entre 47 et 51 ; par convention on prendra le milieu, donc $Q_{0,9} \approx 49$ ans.
L'intervalle de variation à 90% (au risque 10%) de la durée de tabagisme est défini par les deux quantiles d'ordre 0,05 et 0,95 : $I_{90\%} = [Q_{0,05} ; Q_{0,95}]$ puisque $P(X \in I_{90\%}) = P(Q_{0,05} \leq X \leq Q_{0,95}) = 90\%$ et $P(X \notin I_{90\%}) = 10\%$
donc $I_{90\%} = [Q_{0,05} ; Q_{0,95}] = [5 ; 49]$

➨ 5% des sujets ont fumé pendant moins de 5 ans (95% pendant plus de 5 ans), 95% des sujets ont fumé pendant moins de 49 ans (5% pendant plus de 49 ans) et 90% des sujets ont fumé pendant une durée comprise entre 5 et 49 ans.

8) Moyenne et variance de X sont calculées en faisant l'approximation que les n_i individus du $i^{ème}$ intervalle ont pour valeur le milieu de l'intervalle noté x_i :

durée de tabagisme	milieu x_i	effectif n_i	$n_i\, x_i$	x_i^2	$n_i\, x_i^2$
]0; 10]	5	9	45	25	225
]10; 20]	15	21	315	225	4725
]20; 30]	25	15	375	625	9375
]30; 40]	35	8	280	1225	9800
]40; 50]	45	4	180	2025	8100
]50; 60]	55	3	165	3025	9075
total		60	$\Sigma n_i x_i=1360$		$\Sigma n_i x_i^2=41300$

a. Moyenne de X dans $\mathcal{P}$ = moyenne de la durée de tabagisme dans $\mathcal{P}$ = durée moyenne de tabagisme dans $\mathcal{P}$ = μ

donc $\mu = \frac{1360}{60} = 22,667 \approx 22,7$

➨ les 60 fumeurs ont fumé en moyenne pendant 22,7 ans.

b. Variance de X dans $\mathcal{P}$ = variance de la durée de tabagisme dans $\mathcal{P}$ = σ^2
écart-type de X dans $\mathcal{P}$ = écart-type de la durée de tabagisme dans $\mathcal{P}$ =σ

donc $\sigma^2 = \frac{41300}{60} - 22,7^2 = 688,33 - 513,78 = 174,56$ et $\sigma = \sqrt{174,56} = 13,2119 \approx 13,2$

➨ l'écart-type de la durée de tabagisme dans la population étudiée des 60 sujets fumeurs est de 13,2 ans.

c. Les résultats obtenus avec l'approximation précédente sur des données regroupées sont proches de ceux trouvés à la question 3 calculés sur les données individuelles.

9) Cinq ans plus tard tous les sujets auront fumé 5 ans de plus donc :
a. la moyenne sera de 23+5=28 ans
b. les écarts à la moyenne restant inchangés, la variance et l'écart-type ne seront pas modifiés
c. la médiane sera de 20+5=25 ans

10) Les 30 sujets supplémentaires fument depuis 60 ans donc :
a. les 30 sujets supplémentaires fument en moyenne depuis 60 ans : la moyenne sera de $\frac{60\times23+30\times60}{60+30} = \frac{1380+1800}{90} = \frac{3180}{90} = 35,333 \approx 35,3$ ans
b. pour une population de taille 90, la médiane se trouve entre la $45^{ème}$ et la $46^{ème}$ valeur soit entre 29 et 31 ; par convention on prendra le milieu, donc la durée médiane sera de 30 ans.

Exercice 5

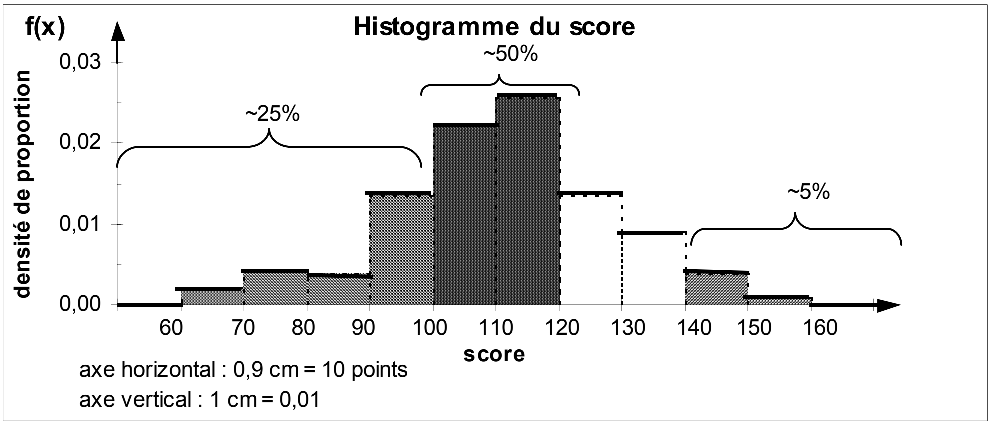
1) $\mathcal{P}$={enfants de 18 mois} de taille (effectif total) N =5+10+9+31+51+59+31+21+9+2=228

2) X= score sur l'échelle de Bayley, définie sur l'échelle descriptive E=]60; 160] (valeurs numériques), c'est donc une variable quantitative.

a. Tableau des distributions d'effectifs (n_i, i=1,…10) et de proportions (p_i, i=1,…10) du score, associées au découpage en 10 intervalles de longueur 10 points.

i	score	effectif n_i	proportion p_i
1	]60; 70]	5	0,022=2,2%
2	]70; 80]	10	0,044=4,4%
3	]80; 90]	9	0,039=3,9%
4	]90; 100]	31	0,136=13,6%
5	]100; 110]	51	0,224=22,4%
6	]110; 120]	59	0,259=25,9%
7	]120; 130]	31	0,136=13,6%
8	]130; 140]	21	0,092=9,2%
9	]140; 150]	9	0,039=3,9%
10	]150; 160]	2	0,009=0,9%
	total	228	1,000=100%

b. L'histogramme est la représentation graphique de la densité de proportion f de la variable X associée à ce découpage. Pour x appartenant au $i^{\text{ème}}$ intervalle la densité est égale à : f(x) = p_i/l_i où l_i est la longueur du $i^{\text{ème}}$ intervalle. Ici tous les intervalles sont de longueur 10 ans donc f(x) = $p_i/10$.



c. L'intervalle modal est l'intervalle de densité maximale : c'est donc le $6^{\text{ème}}$ intervalle]110; 120].

d. La proportion d'enfants ayant un score inférieur à 100 (aire hachurée sur l'histogramme) s'écrit :

$$P(X \le 100) = P(60 < X \le 70) + P(70 < X \le 80) + P(80 < X \le 90) + P(90 < X \le 100)$$

$$= \begin{cases} \frac{5}{228} + \frac{10}{228} + \frac{9}{228} + \frac{31}{228} = \frac{55}{228} \approx 0{,}241 \\ 0{,}022 + 0{,}044 + 0{,}039 + 0{,}136 \approx 0{,}241 \end{cases}$$

(c'est aussi la proportion cumulée F_4).

La proportion d'enfants ayant un score inférieur à 120 s'écrit :

$$P(X \le 120) = P(60 < X \le 70) + P(70 < X \le 80) + P(80 < X \le 90) + P(90 < X \le 100) + P(100 < X \le 110) + P(110 < X \le 120)$$

$$= \begin{cases} \frac{5}{228} + \frac{10}{228} + \frac{9}{228} + \frac{31}{228} + \frac{51}{228} + \frac{59}{228} = \frac{165}{228} \approx 0{,}724 \\ 0{,}022 + 0{,}044 + 0{,}039 + 0{,}136 + 0{,}224 + 0{,}259 \approx 0{,}724 \end{cases}$$

(c'est aussi la proportion cumulée F_6).

La proportion d'enfants ayant un score compris entre 100 et 120 (aire hachurée sur l'histogramme) s'écrit :

$$P(100 < X \le 120) = \begin{cases} P(100 < X \le 110) + P(110 < X \le 120) = \begin{cases} \frac{51}{228} + \frac{59}{228} = \frac{110}{228} \approx 0{,}482 \\ 0{,}224 + 0{,}259 \approx 0{,}483 \end{cases} \\ P(X \le 120) - P(X \le 100) \approx 0{,}724 - 0{,}241 = 0{,}483 \end{cases}$$

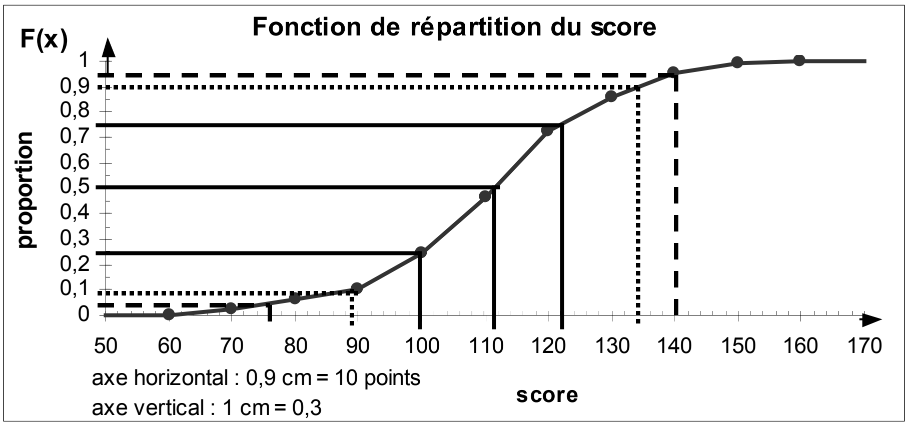
La proportion d'enfants ayant un score supérieur à 140 (aire hachurée sur l'histogramme) s'écrit :

$$P(X > 140) = \begin{cases} P(140 < X \le 150) + P(150 < X \le 160) = \begin{cases} \frac{9}{228} + \frac{2}{228} = \frac{11}{228} \approx 0{,}048 \\ 0{,}039 + 0{,}009 = 0{,}048 \end{cases} \\ 1 - P(X \le 140) = 1 - F_8 \end{cases}$$

3) a. Tableau de la distribution de proportions cumulées du score associée au découpage en 10 intervalles (F_i, i=1,…10) :
pour chaque intervalle la proportion cumulée correspondante : $F_i = N_i/N$ où N_i est l'effectif cumulé $N_i=n_1+n_2+\ldots+n_i$ ou $F_i=p_1+p_2+\ldots+p_i$

i	score	effectif n_i	proportion p_i	effectif cumulé N_i	proportion cumulée F_i
1	]60; 70]	5	0,022=2,2%	$N_1=n_1=5$	$F_1=p_1=0{,}022$
2	]70; 80]	10	0,044=4,4%	$N_2=5+10=15$	$F_2=0{,}066$
3	]80; 90]	9	0,039=3,9%	$N_3=15+9=24$	$F_3=0{,}105$
4	]90; 100]	31	0,136=13,6%	$N_4=24+31=55$	$F_4=0{,}241$
5	]100; 110]	51	0,224=22,4%	$N_5=55+51=106$	$F_5=0{,}465$
6	]110; 120]	59	0,259=25,9%	$N_6=106+59=165$	$F_6=0{,}724$
7	]120; 130]	31	0,136=13,6%	$N_7=165+31=196$	$F_7=0{,}86$
8	]130; 140]	21	0,092=9,2%	$N_8=196+21=217$	$F_8=0{,}952$
9	]140; 150]	9	0,039=3,9%	$N_9=217+9=226$	$F_9=0{,}991$
10	]150; 160]	2	0,009=0,9%	$N_{10}=226+2=228=N$	$F_{10}=1$
	total	228	1,000=100%		

b. La fonction de répartition du score est définie par F(x) = P(X≤x) donc :
F(x)=0 pour tout $x \leq 60$
F(x)=1 pour tout $x > 160$
et à partir de la distribution des proportions cumulées associé au découpage en 10 intervalles (F_i, i=1,…10) : si x est la borne supérieure du i$^{\text{ème}}$ intervalle F(x)= F_i



c. La proportion d'enfants ayant un score inférieur à 100 s'exprime en fonction de la fonction de répartition F de la variable X : P(X≤100)=F(100)= F_4=0,241
La proportion d'enfants ayant un score inférieur à 120 s'écrit : P(X≤120)=F(120) =F_6=0,724
La proportion d'enfants ayant un score compris entre 100 et 120 s'écrit :
$$P(100 < X \leq 120) = P(X \leq 120) - P(X \leq 100) = F(120) - F(100) \approx 0{,}724 - 0{,}241 = 0{,}483$$
La proportion d'enfants ayant un score supérieur à 140 s'écrit :
$$P(X > 140) = 1 - P(X \leq 140) = 1 - F(140) = 1 - F_8 \approx 1 - 0{,}952 = 0{,}048$$

4) a. Les trois quartiles de l'âge sont représentés en abscisse sur le graphique de la fonction de répartition :
la médiane (second quartile) notée $Q_{0,5}$ est telle que F($Q_{0,5}$)=0,5 donc graphiquement on déduit que $Q_{0,5} \approx 111$,
le premier quartile noté $Q_{0,25}$ est tel que F($Q_{0,25}$)=0,25 donc graphiquement on déduit que $Q_{0,25} \approx 100$
le troisième quartile noté $Q_{0,75}$ est tel que F($Q_{0,75}$)=0,75 donc graphiquement on déduit que $Q_{0,75} \approx 122$
➨ 50% des enfants ont un score inférieur à 112 points (50% ont un score supérieur à 112 points), 25% des enfants ont un score inférieur à 100 points (75% ont un score supérieur à 100 points) et 75% des enfants ont un score inférieur à 120 points (25% ont un score supérieur à 120 points).

b. Puisque F(140)=0,952 : environ 95% des enfants ont un score inférieur à 140 points, le quantile d'ordre 0,95 (95%) ou 95$^{\text{ème}}$ percentile du score est de 140 points, $Q_{0,95} \approx 140$

c. L'intervalle de variation à 90% (au risque 10%) du score est défini par les deux quantiles d'ordre 0,05 et 0,95 : $I_{90\%} = [Q_{0,05} ; Q_{0,95}]$ puisque $P(X \in I_{90\%}) = P(Q_{0,05} \leq X \leq Q_{0,95}) = 90\%$ et $P(X \notin I_{90\%}) = 10\%$
en abscisse sur le graphique de la fonction de répartition, le 5$^{\text{ème}}$ percentile noté $Q_{0,05}$ est tel que F($Q_{0,05}$)=0,05 donc graphiquement on déduit que $Q_{0,05} \approx 76$, donc $I_{90\%} = [Q_{0,05} ; Q_{0,95}] \approx [76 ; 140]$
➨ 5% des enfants ont un score inférieur à 76 points (95% ont un score supérieur à 76 points), 95% des enfants ont un score inférieur à 140 points (5% ont un score supérieur à 140 points) et 90% des enfants ont un score compris entre 76 et 140 points.
l'intervalle de variation au risque 20% (à 80%) du score est défini par les deux quantiles d'ordre 0,1 et 0,9 :
$I_{80\%} = [Q_{0,1} ; Q_{0,9}]$ puisque $P(X \in I_{80\%}) = P(Q_{0,1} \leq X \leq Q_{0,9}) = 80\%$ et $P(X \notin I_{80\%}) = 20\%$
en abscisse sur le graphique de la fonction de répartition, le 1$^{\text{er}}$ décile noté $Q_{0,1}$ est tel que F($Q_{0,1}$)=0,1 donc graphiquement on déduit que $Q_{0,1} \approx 89$, et le 9$^{\text{ème}}$ décile noté $Q_{0,9}$ est tel que F($Q_{0,9}$)=0,9 donc graphiquement on déduit que $Q_{0,9} \approx 134$,donc $I_{80\%} = [Q_{0,1} ; Q_{0,9}] \approx [89 ; 134]$

➨ 10% des enfants ont un score inférieur à 90 points (90% ont un score supérieur à 90 points), 90% des enfants ont un score inférieur à 132 points (10% ont un score supérieur à 132 points) et 80% des enfants ont un score compris entre 90 et 132 points.

5) Moyenne et variance de X sont calculées en faisant l'approximation que les n_i individus du i[ème] intervalle du découpage du score ont pour valeur x_i le milieu de l'intervalle :

i	score	milieu x_i	effectif n_i	$n_i\,x_i$	x_i^2	$n_i\,x_i^2$
1	]60; 70]	65	5	325	4 225	21 125
2	]70; 80]	75	10	750	5 625	56 250
3	]80; 90]	85	9	765	7 225	65 025
4	]90; 100]	95	31	2 945	9 025	279 775
5	]100; 110]	105	51	5 355	11 025	562 275
6	]110; 120]	115	59	6 785	13 225	780 275
7	]120; 130]	125	31	3 875	15 625	484 375
8	]130; 140]	135	21	2 835	18 225	382 725
9	]140; 150]	145	9	1 305	21 025	189 225
10	]150; 160]	155	2	310	24 025	48 050
	total		228	$\Sigma n_i x_i$=25 250		$\Sigma n_i x_i^2$=2 869 100

moyenne de X dans $\mathcal{P}$ = moyenne du score dans $\mathcal{P}$ = score moyen dans $\mathcal{P}$ = μ

variance de X dans $\mathcal{P}$ = variance du score dans $\mathcal{P}$ =σ^2

écart-type de X dans $\mathcal{P}$ = écart-type du score dans $\mathcal{P}$ =σ

donc $\mu = \frac{25250}{228} = 110{,}7456 \approx 110{,}7$ $\sigma^2 = \frac{2869100}{228} - 110{,}7^2 = 12583{,}77 - 12264{,}59 = 319{,}18 \approx 319{,}2$ et

$\sigma = \sqrt{319{,}2} = 17{,}866 \approx 17{,}9$

➨ les 228 enfant de 18 mois ont un score moyen de 110,7 points, la variance du score dans cette population est de 319,2 et l'écart-type du score est de 17,9 points.

6) Les 16 enfants supplémentaires fument depuis 60 ans :

a. les 16 enfants supplémentaires ont un score moyen de 30 : la moyenne pour la population des 228+16=244 enfants sera de $\frac{228 \times 110{,}7456 + 16 \times 30}{228+16} = \frac{25250+480}{244} = \frac{25730}{244} = 105{,}45 \approx 105{,}5$ points

b. en rajoutant 16 aux effectifs cumulés calculés à la question 3.a, N_5=106+16=122 correspond à l'effectif cumulé pour la valeur 110 : F(110)=122/244=0,5 donc la médiane est de 110 points pour la population des 244 enfants de 18 mois.