Université M. MAMMERI (Tizi-Ouzou) Année Universitaire 2023/2024
Fac : Sciences, Dpt. : **Maths/Info.** **Probabilités/Statistique**
Licence (**1**ère Année, Section A) Série N° **3** (**S**tatistique **D**escriptive)

Exercice 1 : Indiquer de quels types sont les variables statistiques suivantes :
(**1**) l'état civil d'un habitant ; (**2**) la *taille* d'un étudiant ; (**3**) le *nombre de page*s d'un support de cours ; (**4**) la *pluviométrie* dans une région donnée ; (**5**) le *poids* d'un nouveau-né ; (**6**) le *degré de qualification* du personnel d'une entreprise ; (**7**) la *nationalité* d'un résident ; (**8**) le *groupe sanguin* d'un candidat au Bac ; (**9**) le *nombre d'enfants* à charge d'un employé ; (**10**) la *commune* de résidence.
On **justifiera** les réponses, en *donnant* dans chaque cas, quelques *modalités/ classes* possibles.

Exercice 2 : Une enquête menée dans une région montagneuse sur la répartition des groupes sanguins a donné les résultats suivants **:**

Groupe Sanguin	**A**	**B**	**AB**	**O**
Effectif	**40**	**43**	**12**	**5**

1. Décrire la population et l'échantillon étudiés.
2. Quel caractère $\boldsymbol{X}$ est étudié ? de quelle nature est-il ?
3. Le représenter par le graphique adéquat.

Exercice 3 : Le gérant d'une librairie a relevé pour un livre particulier (qui connaît une très forte popularité) le nombre d'exemplaires vendus par jour. Son relevé a porté sur les ventes des mois de Juillet et Août, ce qui correspond à cinquante-deux (**52**) jours de vente. Le relevé des observations a donné **:**

7, 13, 8, 10, 9, 12, 10, 8, 9, 10, 13, 14, 7, 11, 9, 11, 12, 11, 12, 13, 14, 11, 8, 10, 12, 10,
8, 14, 7, 13, 12, 13, 11, 9, 11, 12, 13, 12, 11, 14, 8, 14, 9, 9, 14, 13, 11, 10, 12, 9, 13, 11.

1. Définir la variable statistique étudiée ? quelle est sa nature ? décrire ses modalités.
2. Déterminer le tableau statistique et le compléter par les calculs ultérieurs (nécessaires à la détermination des différents paramètres).
3. Tracer le diagramme associé à la variable $\boldsymbol{X}$**.** Déduire les polygones des effectifs et des fréquences**.**
4. Déterminer le mode, la médiane, la moyenne, la variance et l'écart type**.**
5. Calculer les quartiles **;** les interpréter.

Exercice 4 : On a recensé dans quatre**-**vingt**-**deux (**82**) hôpitaux d'un pays, les cas d'une maladie rare et inconnue. On a recueilli les résultats dans le tableau suivant **:**

$\boldsymbol{X_i}$	0	1	2	3	4	5
$\boldsymbol{n_i}$	4	10	28	22	12	6

1. Définir le caractère $\boldsymbol{X}$ étudié, de quelle nature est-il ?
2. Que représentent *exactement* $\boldsymbol{X_i}$ et $\boldsymbol{n_i}$?
3. Calculer le mode et la médiane du caractère **;** retrouver la médiane graphiquement.
4. Calculer les premier et troisième quartiles du caractère. Interpréter.
5. Calculer la variance et l'écart type de $\boldsymbol{X}$ (présenter les calculs dans un tableau).

Exercice 5 : La distribution suivante représente la taille (en centimètres) de quarante (40) étudiants :

165	175	180	152	161	152	155	158
154	173	178	179	160	178	180	167
152	172	177	180	175	154	157	159
169	169	170	167	153	169	171	172
172	158	162	157	158	163	168	164

1. Définir la variable statistique étudiée. De quelle nature est-elle ?
2. En utilisant la règle de STURGES, ranger ces observations dans un tableau statistique.
3. Tracer le diagramme associé à cette distribution.
4. Calculer le mode et la médiane du caractère ; retrouver la médiane graphiquement.
5. Calculer les premier et troisième quartiles du caractère. Interpréter.
6. Calculer la variance et l'écart type de $\boldsymbol{X}$ (présenter tous les calculs dans un tableau).

Exercice 6 : Les observations relatives à la variable « poids » obtenues sur un échantillon de soixante (60) étudiants sont transcrites dans le tableau suivant :

Classes	[50,60[	[60,70[	[70,80[	[80,90[	[90,100]
Effectifs	7	25	21	2	5

1. Définir *rigoureusement* le caractère étudié et donner sa nature.
2. Dresser le tableau des fréquences relatives et des fréquences cumulées.
3. Calculer : le mode, la médiane, les quartiles, l'intervalle interquartile, la moyenne arithmétique, la variance ainsi que l'écart type.

Exercice 7 (C) : Un supermarché reçoit cent (100) caissettes comprenant chacune neuf (9) pêches. La distribution du nombre de pêches abîmées par caissette est donnée dans le tableau suivant :

N^{bre} de pêches abîmées	0	1	2	3	4	5	6	7	8	9
N^{bre} de caissettes	12	29	31	17	5	1	2	2	0	1

1. Définir la variable $\boldsymbol{X}$ étudiée, de quelle nature est-elle ?
2. Dresser le tableau des fréquences relatives et des fréquences cumulées.
3. Quel est le nombre de caissettes ayant un nombre de pêches abîmées au moins égal à 3 ?
4. Quelle est la proportion de caissettes dont le nombre de pêches abîmées est au plus de 4 ?
5. Calculer : le mode, les différents quartiles, l'étendue du nombre de pêches abimées, la moyenne arithmétique, l'écart absolu moyen et l'écart type.

Exercice 8 (C) : Dans une ferme, on a pesé les œufs qui ont été pondus à une date déterminée (les masses sont exprimées en grammes) :

Masse de l'œuf	[28; 38[	[38 ; 48[	[48 ; 53[	[53 ; 58[	[58 ; 63[	[63 ; 73[	[73 ; 83[
N^{bre} d'œufs	3	51	74	112	92	62	6

1. Quelle est la variable $\boldsymbol{X}$ étudiée ? Quelle est sa nature ?
2. Donner la représentation graphique de la distribution.
3. Déterminer le mode, la médiane et la moyenne arithmétique.
4. Calculer les 1er et 3ème quartiles, la variance et l'écart type de la variable $\boldsymbol{X}$.

Exercice 9 (C) : Expliquer comment trouver les quartiles d'une série statistique ordonnée $\boldsymbol{x}_{(1)}, \dots, \boldsymbol{x}_{(1)}$ d'une variable X quantitative discrète. Appliquer aux cas $n = 18$ et $n = 27$.

Corrigé de la série 3 (Statistique Descriptive)

Ex. 1 : Nature des caractères :

(**1**) l'état civil (matrimonial) d'un habitant : Qualitatif nominal (célibataire, marié, divorcé, veuf)

(**2**) la *taille* d'un étudiant : Quantitatif continu (classes : [130, 140[, …,, [180, 190[, en cm)

(**3**) le *nombre de pages* d'un support de cours : Quantitatif discret (entiers entre 15/20 et 50/60) s'il s'agit de notes de cours, ce serait proche de la vingtaine, si c'est un cours détaillé-par exple avec des graphiques/schémas, ce serait la cinquantaine ou mm plus !

(**4**) la *pluviométrie* dans une région donnée : quantitatif continu (elle peut varier de qlq mm à plus de 100 mm)

(**5**) le *poids* d'un nouveau-né : quantitatif continu (varie de 2.8 kg à 5 kg)

(**6**) le *degré de qualification* du personnel d'une entreprise : qualitatif ordinal (on peut avoir à classer les modalités selon le degré de compétence)

(**7**) la *nationalité* d'un résident ; qualitatif nominal

(**8**) le *groupe sanguin* d'un candidat au Bac ; qualitatif nominal (A, B, AB, O)

(**9**) le *nombre d'enfants* à charge d'un employé ; quantitatif discret (varie de 0 à 8)

(**10**) la *commune* de résidence ; qualitatif nominal (il n'y a pas d'ordre sur les modalités)

Ex. 2 :

1. La population : les habitants d'une région montagneuse, l'échantillon : un groupe de 100 personnes.
2. Le caractère $\boldsymbol{X}$ étudié : Le groupe sanguin d'une personne, il est qualitatif nominal.
3. Le graphique adéquat est un diagramme : le 1er *en barres*, constitué de 4 rectangles dont la surface de chacun est proportionnelle à son effectif. Elle s'obtient par une simple règle de trois ! c'est une taille d'échantillon standard qui donne 40%, 43, 12 et 5% ; le 2nd est un diagramme *circulaire*, il a pour angles des secteurs, respectivement : 144°, 155°, 43° et 18°.

Ex. 3 :

7, 13, 8, 10, 9, 12, 10, 8, 9, 10, 13, 14, 7, 11, 9, 11, 12, 11, 12, 13, 14, 11, 8, 10, 12, 10,
8, 14, 7, 13, 12, 13, 11, 9, 11, 12, 13, 12, 11, 14, 8, 14, 9, 9, 14, 13, 11, 10, 12, 9, 13, 11.

1. La variable statistique étudiée : Nombre d'exemplaires vendus en un jour ; il est quantitatif discret ; ses modalités sont les entiers entre 7 et 14 (8 modalités).

2. Le tableau statistique (complet) : $\boldsymbol{X_i}$ désigne le nombre de livres et $\boldsymbol{n_i}$ le nombre de jours durant lesquels il s'est vendu $\boldsymbol{X_i}$ livres.

X_i	7	8	9	10	11	12	13	14	Total
n_i	3	5	7	6	9	8	8	6	52
f_i	0.058	0.096	0.135	0.115	0.173	0.154	0.154	0.115	1
n_i^c	3	8	15	21	30	38	46	52	
n_iX_i	21	40	63	60	99	96	104	84	567 (10.90)
$n_iX_i^2$	147	320	567	600	1089	1152	1352	1176	6403 (123.135)

3. Tracer le diagramme associé à la variable $\boldsymbol{X}$ **:** c'est un diagramme en bâtons **;** duquel on tire le polygone des effectifs et celui des fréquences.

4. Détermination des différents **paramètres :** le mode est : $\boldsymbol{M_0 = 11\ l/j}$, la médiane vaut : $Med = 11\ l/j$; la **moyenne** arithmétique $\overline{\boldsymbol{X}} = \boldsymbol{10.9}\ livres/jour$ (cela signifie que sur une période de 10 jours, il s'est vendu 109 livres en moyenne).

La **variance** $Var(X) = \frac{1}{n}\sum_{i=1}^{6} \boldsymbol{n_i X_i^2} - \overline{\boldsymbol{X}}^2 = 123.135 - 118.81 = 4.325\ (l/j)^2$ et **l'écart type** est donné par $\boldsymbol{\sigma_X = 2.08}\ l/j$.

5. Calcul des **quartiles :** $Q_1 = \frac{x_{(13)}+x_{(14)}}{2} = \mathbf{9}\ l/j$, $Q_3 = \frac{x_{(39)}+x_{(40)}}{2} = \mathbf{13}\ l/j$. Ceci signifie qu'il y a 25% de jours (les 13 premiers jours) où il s'est vendu *au plus* 9 livres, et aussi 13 jours (les derniers) où il s'est vendu *au moins* 13 livres (13 ou 14 livres).

Ex. 4 : On a les résultats du tableau suivant :

X_i	0	1	2	3	4	5	Total
n_i	4	10	28	22	12	6	82
f_i	0.050	0.122	0.341	0.268	0.146	0.073	1
n_i^c	4	14	42	64	76	82	
n_iX_i	0	10	56	66	48	30	210 (2.56)
$n_iX_i^2$	0	10	112	198	192	150	662 (8.073)

1. Le caractère X : le nombre de cas d'une maladie rare dans ***un*** hôpital ; il est *quantitatif discret*.
2. X_i représente le nombre de cas et n_i désigne le nombre d'hôpitaux correspondants (ie dans lesquels on a dénombré X_i cas).
3. Le **mode** est la modalité de plus grand effectif, $M_0 = 2\ c/h$. ; la **médiane** est donnée par $Med = \frac{x_{(41)}+x_{(42)}}{2} = 2\ c/h$; ***graphiquement***, c'est l'abscisse du point d'ordonnée $n/2$ sur le diagramme (en escalier) des effectifs cumulés.
4. Calcul des premier et troisième **quartiles :** $Q_1 = x_{(21)} = 2\ cas$; $Q_3 = x_{(62)} = 3\ cas$.
   Interprétation **:** on trouve 25% d'hôpitaux (soit **20**) dans lesquels il y a ***au plus** 2 cas* de cette maladie, et 25 % (**20**) dans lesquels il y a ***au moins** 3 cas*.
5. Calcul de la **variance** et de **l'écart type :** la **moyenne** est donnée par $\overline{X} = 210/82 = \mathbf{2.56}$ **cas** (ceci signifie que dans **100** hôpitaux, on dénombre, en moyenne **256** cas). La **variance** est donnée par : $Var(X) = \frac{1}{n}\sum_{i=1}^{6} n_iX_i^2 - \overline{X}^2 = 8.0732 - (2.56)^2 = \mathbf{1.52}\ (c/h)^2$, d'où : $\sigma_X = \mathbf{1.23}\ c/h$.

Ex. 5 :

1. La variable statistique : X=taille d'un étudiant (en cm) ; elle est *quantitative **continue***.
2. D'après la règle de STURGES, le nombre de classes vaut $k \cong 1 + 3.3log_{10}(n) \cong 6$; d'où le tableau statistique suivant (complété par les calculs ultérieurs) :

Taille (classes)	[152;157[	[157;162[	[162;167[	[167;172[	[172;177[	[177;182]	Total
Nbre d'étudiants	7	8	4	8	6	7	40
f_i	0.175	0.2	0.1	0.2	0.15	0.175	1
n_i^c	7	15	19	27	33	40	
Centres C_i	154.5	159.5	164.5	169.5	174.5	179.5	
n_iC_i	1081.5	1276	658	1356	1047	1256.5	6675
$n_iC_i^2$	167091.75	203522	108241	229842	182701.5	225541.75	1116940
C_i'	-15	-10	-5	0	5	10	
$n_i.C_i'$	-105	-80	-20	0	30	70	-105
$n_i.(C_i')^2$	1575	800	100	0	150	700	3325

3. Le diagramme associé est un *histogramme*, duquel on déduit les polygones des effectifs et celui des fréquences.

4. Les ***modes***: $\boldsymbol{M_0} = \boldsymbol{l_{inf}} + \left(\frac{\Delta_1}{\Delta_1+\Delta_2}\right)\boldsymbol{a} = 157 + \left(\frac{1}{1+4}\right)5 = \mathbf{158}\ cm$ et $M_0' = 167 + \frac{10}{3} = \boxed{170.33}$; (on fera remarquer que $\boldsymbol{\Delta_1}$ (resp $\boldsymbol{\Delta_2}$) est la différence entre $\boldsymbol{n}_{\max}$ et l'effectif de la classe précédente (resp suivante) ; la ***médiane*** : $\boldsymbol{Med} = \boldsymbol{l_{inf}} + \frac{\boldsymbol{a}}{\boldsymbol{n_0}}\left(\frac{\boldsymbol{n}}{2} - \boldsymbol{n_{inf}^c}\right) = \mathbf{167} + \frac{\mathbf{5}}{\mathbf{8}}(\mathbf{20} - \mathbf{19}) = \boxed{\mathbf{167.625}}$ cm($\cong \mathbf{1.68}\boldsymbol{m}$) ; ***graphiquement,*** c'est l'abscisse du point d'ordonnée $n/2$ sur le polygone des effectifs cumulés.

5. Calcul des ***quartiles*** :

$$\boldsymbol{Q_1} = \boldsymbol{l_{inf}^{(1)}} + \frac{\boldsymbol{a}}{\boldsymbol{n_0^{(1)}}}\left(\frac{\boldsymbol{n}}{\mathbf{4}} - \boldsymbol{n_{inf}^{c,(1)}}\right) = \mathbf{157} + \frac{\mathbf{5}}{\mathbf{8}}(\mathbf{10} - \mathbf{7}) = \mathbf{158.875}\ \boxed{(\cong \mathbf{159})\ \boldsymbol{cm}}\ ;$$

$$\boldsymbol{Q_3} = \boldsymbol{l_{inf}^{(3)}} + \frac{\boldsymbol{a}}{\boldsymbol{n_0^{(3)}}}\left(\frac{\mathbf{3}\boldsymbol{n}}{\mathbf{4}} - \boldsymbol{n_{inf}^{c,(3)}}\right) = \mathbf{172} + \frac{\mathbf{5}}{\mathbf{6}}(\mathbf{30} - \mathbf{27}) = \boxed{\mathbf{174.5}\ \boldsymbol{cm}}.$$

<u>Interprétation</u> : il y a 25% d'étudiants (ie **10**) qui ont une taille *inférieure* à $\mathbf{1m59}$, et autant (soit 10) qui *dépassent* $\mathbf{1m745}$.

6. Calcul de la **variance** et de l'**écart type** de $\boldsymbol{X}$ **:** la *moyenne* vaut $\overline{\boldsymbol{X}} = \frac{\mathbf{6675}}{\mathbf{40}} = \mathbf{166.875} \cong \mathbf{167}\boldsymbol{cm},$ et alors $\boldsymbol{Var}(\boldsymbol{X}) = \frac{\sum \boldsymbol{n_i}.\boldsymbol{C_i^2}}{\boldsymbol{n}} - \overline{X}^2 = \frac{\mathbf{1116940}}{\mathbf{40}} - (\mathbf{166.875})^{\mathbf{2}} = 27923.5 - 27847.27 = \mathbf{76.23}\ cm^2$. Cela donne un écart type de **:** $\boldsymbol{\sigma_X} = \sqrt{\mathbf{76.23}} = \boxed{\mathbf{8.73}\ \boldsymbol{cm}}$.
Rque : Soit $\boldsymbol{C_i'} = \boldsymbol{C_i} - M_0'$ ($M_0' = \mathbf{169.5}$). Les $\boldsymbol{C_i'}$ sont donnés dans le tableau ; soit $\boldsymbol{X'} = \boldsymbol{X} - M_0'$; on voit que la moyenne de $\boldsymbol{X}$ est $\overline{\boldsymbol{X}} = \overline{\boldsymbol{X'}} + M_0'$ soit : $\overline{\boldsymbol{X}}$ =(-105/40)+169.5 =<u>166.875</u> ; ainsi, la variance est donnée par : Var($\boldsymbol{X'}$)$= \frac{\sum \boldsymbol{n_i}.\boldsymbol{C_i'^2}}{\boldsymbol{n}} - \overline{\boldsymbol{X'}}^2 = \frac{\mathbf{3325}}{\mathbf{40}} - (-\mathbf{2.625})^{\mathbf{2}} = \mathbf{83.125} - \mathbf{6.89} = \mathbf{76.235} : \boldsymbol{Var}(\boldsymbol{X})$. Cette méthode de calcul s'appelle méthode du *changement d'origine* ; elle est très utile quand on manipule de grands nombres : on les ramène tous autour de 0 en leur retranchant la même valeur, (ce qui a donné d'ailleurs une moyenne « décalée » ici négative).

Ex. 6 :

Classes C_i	[50,60[	[60,70[	[70,80[	[80,90[	[90,100]	Total
Effectifs ($\boldsymbol{n_i}$)	7	25	21	2	5	60
Efctfs cumulés ($\boldsymbol{n_i^c}$)	7	32	53	55	60	/
Fréquences $\boldsymbol{f_i}$	0.117	0.417	0.35	0.033	0.083	1
Fréq. cumulées ($\boldsymbol{f_i^c}$)	0.117	0.534	0.884	0.917	1	/
$\boldsymbol{C_i}$	55	65	75	85	95	/
$\boldsymbol{n_i}.\boldsymbol{C_i}$	385	1625	1575	170	475	4230 (70.5)
$\boldsymbol{n_i}.\boldsymbol{C_i}^2$	21175	105625	118125	14450	45125	304500 (5075)

1. La *variable statistique* $\boldsymbol{X}$ étudiée **: le poids d'un** étudiant (en kg) ; elle est de nature *quantitative continue*.
2. Tracé de l'histogramme + polygones (des effectifs et des fréquences). **Voir figure.**
3. Le ***mode*** : $\boldsymbol{M_0} = l_{inf} + \left(\frac{\Delta_1}{\Delta_1+\Delta_2}\right)a = 60 + \left(\frac{18}{18+4}\right)10 = \boxed{\mathbf{68.18}}\ kg$;

   La ***médiane*** se calcule par : $\boldsymbol{Med} = \boldsymbol{l_{inf}} + \frac{\boldsymbol{a}}{\boldsymbol{n_0}}\left(\frac{\boldsymbol{n}}{\mathbf{2}} - \boldsymbol{n_{inf}^c}\right) = \mathbf{60} + \frac{\mathbf{10}}{\mathbf{25}}(\mathbf{30} - \mathbf{7}) = \boxed{\mathbf{69.2}\ \boldsymbol{kg}}$.

   La **moyenne** arithmétique vaut : $\overline{\boldsymbol{X}} = \frac{\sum \boldsymbol{n_i}.\boldsymbol{C_i}}{\boldsymbol{n}} = \frac{\mathbf{4230}}{\mathbf{60}} = \boxed{\mathbf{70.5}}\ \boldsymbol{kg}.$

De même, la *variance* : $\boldsymbol{Var(X)} = \frac{\sum n_i.c_i^2}{n} - \bar{X}^2 = \frac{304500}{60} - (\mathbf{70.5})^2 = \boxed{\mathbf{104.75}}\,\boldsymbol{kg^2}$; l'*écart type* s'en déduit : $\boldsymbol{\sigma_X} = \sqrt{\boldsymbol{Var(X)}} = \boxed{\mathbf{10.235}\ \boldsymbol{kg}}$.

4. Le calcul des *quartiles* se fait de la « même » manière que la médiane :

$$\boldsymbol{Q_1 = l_{inf}^{(1)} + \frac{a}{n_0^{(1)}}\left(\frac{n}{4} - n_{inf}^{c,(1)}\right) = 60 + \frac{10}{25}(15 - 7) = \boxed{63.2}\ kg}\ ;$$

$$\boldsymbol{Q_3 = l_{inf}^{(3)} + \frac{a}{n_0^{(3)}}\left(\frac{3n}{4} - n_{inf}^{c,(3)}\right) = 70 + \frac{10}{21}(45 - 32) = \boxed{76.2}\ kg}.$$

L'*interprétation est : il y a 25% d'étudiants (soit **15**) qui ont un poids inférieure à **63.2** kg, il y en a autant (**15** étudiants) dont le poids dépasse **76.2** kg et il reste **50%** (soit **30** individus) dont le poids est compris entre ces deux valeurs (des 1er et 3ème quartiles), ie dans l'intervalle (63.2-76.2).*

Exercice 7 :

a) X = Nombre de pêches abîmées dans une caissette. X est *quantitative discrète*.

b) X présente dix (10) modalités : 0, 1, …, 9. On complète le tableau statistique par les calculs dont on aura besoin par la suite pour l'évaluation des différents paramètres. On obtient :

Pêch abm (X_j)	0	1	2	3	4	5	6	7	8	9	Total
Nbe csst (n_j)	12	29	31	17	5	1	2	2	0	1	$(\boldsymbol{n})$**100**
Eff Cumulés	**12**	**41**	**72**	**89**	**94**	**95**	**97**	**99**	**99**	**100**	/
Eff Cum ↓	**100**	**88**	**59**	**28**	**11**	**6**	**5**	**3**	**1**	**1**	/
Fréquences	**0.12**	**0.29**	**0.31**	**0.17**	**0.05**	**0.01**	**0.02**	**0.02**	**0**	**0.01**	/
Fréq. Cumulées	**0.12**	**0.41**	**0.72**	**0.89**	**0.94**	**0.95**	**0.97**	**0.99**	**0.99**	**1.00**	/
$n_j.X_j$	0	29	62	51	20	5	12	14	0	9	202
$n_j.X_j^2$	0	29	124	153	80	25	72	98	0	81	662
$\boldsymbol{X_j - \bar{X}}$	-2.02	-1.02	-0.02	0.98	1.98	2.98	3.98	4.98	5.98	6.98	
$\boldsymbol{n_j \times \lvert X_j - \bar{X}\rvert}$	24.24	29.58	0.62	16.66	9.9	2.98	7.96	9.96	0	6.98	108.88

c) La proportion de caissettes ayant un nombre de pêches abîmées au moins égal à 3 est donnée par $100 \times (\frac{n-72}{n})\% = \mathbf{28\%}$ **;** et la proportion de caissettes dont le nombre de pêches abîmées est au plus de 4 vaut **94**%.

d) Le ***mode*** est donné par $\boldsymbol{M_0 = 2pa/c}$ et la ***médiane*** vaut : $Med = \frac{x_{(50)}+x_{(51)}}{2} = \mathbf{2}\ \boldsymbol{pa/c}$.

e) La ***moyenne*** se calcule par : $\boldsymbol{\bar{X}} = \frac{1}{n}\sum_{i=1}^{10} \boldsymbol{n_i X_i} = \frac{202}{100} = \mathbf{2.02}\ \boldsymbol{pa/c}$: ceci signifie que, en moyenne, on dénombre deux-cent deux (**202**) *pêches abîmées dans 100 caissettes (cad sur 900, ce qui constitue 22.44%, qui est un pourcentage non négligeable)*. La variance est donnée par : $Var(\boldsymbol{X}) = (\frac{1}{n}\sum_{i=1}^{10} \boldsymbol{n_i X_i^2}) - \boldsymbol{\bar{X}^2} = 6.62 - (2.02)^2 = \mathbf{2.5396}\ (\boldsymbol{pa/c})^2$; et l'**écart type** vaut $\boldsymbol{\sigma_X} = \mathbf{1.6}\ \boldsymbol{pa/c}$.

f) On a aussi : $Q_1 = \frac{x_{(25)}+x_{(26)}}{2} = \mathbf{1}\ \boldsymbol{pa/c}$; de même : $Q_3 = \frac{x_{(75)}+x_{(76)}}{2} = \mathbf{3}\boldsymbol{pa/c}$: il y a 25 caissettes (25%) avec *au plus* 1 ***pa*** et 25 caissettes (25%) avec *au moins* 3 ***pa.***

g) L'écart absolu moyen est donné par : $\boldsymbol{Eam} = \frac{1}{n}\sum_{i=1}^{10} \boldsymbol{n_i \lvert X_i - \bar{X}\rvert} = \mathbf{1.09}$ **pa/c.**

Exercice 8 (C) :

Classes C_i	[28; 38[	[38 ; 48[	[48 ; 53[	[53 ; 58[	[58 ; 63[	[63 ; 73[	[73 ; 83[	Total
Effectifs (n_i)	3	51	74	112	92	62	6	400
Eff Crgés	1.5	25.5	74	112	92	31	3	/
Eff cum (n_i^c)	3	54	128	240	332	394	400	/
Frégnces f_i	0.0075	0.1275	0.185	0.28	0.23	0.155	0.015	1
Fréq. cum (f_i^c)	0.0075	0.1350	0.320	0.60	0.83	0.985	1.000	/
C_i	33	43	50.5	55.5	61.5	68	78	/
$n_i.C_i$	99	2193	3737	6216	5658	4216	468	22587 (56.4675)
$n_i.C_i{}^2$	3267	94299	188718.5	344988	347967	286688	36504	1302431.5 (3256.08)

1. La *variable statistique* X étudiée : **le poids d'un** œuf (en g) ; elle est de nature *quantitative continue*.

2. Tracé de l'histogramme + polygones (des effectifs et des fréquences). Voir **figure.** Ici, il faut considérer les différentes amplitudes de classes, il y en a deux (5 et 10), et donc on aura des effectifs corrigés : une amplitude de base à choisir étant 5, certaines classes ont 2 unités d'amplitude ! et donc leur effectif sera divisé par 2.

3. Le ***mode*** : $M_0 = l_{inf} + \left(\frac{\Delta_1}{\Delta_1+\Delta_2}\right) a_{mo} = 53 + \left(\frac{38}{38+20}\right) 5 = \boxed{56.28\ g}$;

   La ***médiane*** se calcule par : $\boldsymbol{Med = l_{inf} + \frac{a}{n_0}(\frac{n}{2} - n_{inf}^c) = 53 + \frac{5}{112}(200 - 128)} = \boxed{66.21\ g}$.

   La **moyenne** arithmétique vaut : $\overline{X} = \frac{\sum n_i.C_i}{n} = \frac{22587}{400} = \boxed{56.47\ g}$.

4. Le calcul des *quartiles* se fait de la « même » manière que la médiane :

$$\boldsymbol{Q_1 = l_{inf}^{(1)} + \frac{a}{n_0^{(1)}}\left(\frac{n}{4} - n_{inf}^{c,(1)}\right) = 48 + \frac{5}{74}(100 - 54)} = \boxed{51.11\ g}\ ;$$

$$\boldsymbol{Q_3 = l_{inf}^{(3)} + \frac{a}{n_0^{(3)}}\left(\frac{3n}{4} - n_{inf}^{c,(3)}\right) = 58 + \frac{5}{92}(300 - 240)} = \boxed{61.26\ g}.$$

I*nterprétation* *: il y a –dans l'échantillon, 25% d'œufs (soit* ***100****) qui ont un poids inférieur à* ***51.11*** *g, il y en a autant (****100****) dont le poids dépasse* ***61.26*** *g et il reste* ***50%*** *(soit* ***200*** *œufs) dont le poids est compris entre ces deux valeurs (des 1er et 3ème quartiles), ie dans l'intervalle (51.11-61.26) g. cet intervalle est de longueur 10.15 g (=1.2* $\boldsymbol{\sigma}$*).*

La *variance* : $\boldsymbol{Var(X)} = \frac{\sum n_i.C_i{}^2}{n} - \overline{X}^2 = \frac{1302431.5}{400} - (56.47)^2 = \boxed{67.22\ g^2}$; l'*écart type* s'en déduit : $\boldsymbol{\sigma_X} = \sqrt{\boldsymbol{Var(X)}} = \boxed{8.2\ g}$.

Exercice 9 : (Complément)

Calcul des **quartiles** d'une distribution (les 3 **paramètres** qui **divisent** celle-ci **en 4 parties** de même effectif **: 25%** chacune) dans le *cas discret* **:** (***Rappel*** **:** $\boldsymbol{Q_2 = Med}$).

Soit r le reste de la division de n par 4 ($r = 0,1,2$ ou 3)

1) $n = 4p$ ($\boldsymbol{r = 0}$) : $Q_1 = \frac{x_{(p)}+x_{(p+1)}}{2}$, $Q_2 = \frac{x_{(2p)}+x_{(2p+1)}}{2}$ et $Q_3 = \frac{x_{(3p)}+x_{(3p+1)}}{2}$;

2) $n = 4p + 1$ ($r = 1$) : $Q_1 = \frac{x_{(p)}+x_{(p+1)}}{2}$, $Q_2 = \boldsymbol{x_{(2p+1)}}$, $Q_3 = \frac{x_{(3p+1)}+x_{(3p+2)}}{2}$;

3) $n = 4p + 2[= 2\boldsymbol{k}]$ ($r = 2$) **:** $Q_1 = \boldsymbol{x_{(p+1)}}$, $Q_2 = \frac{x_{(2p+1)}+x_{(2p+2)}}{2}$ et $Q_3 = \boldsymbol{x_{(3p+2)}}$;

4) $n = 4p + 3[= 2\boldsymbol{k} + 1]$ ($r = 3$) **:** $Q_1 = \boldsymbol{x_{(p+1)}}$, $Q_2 = \boldsymbol{x_{(2p+2)}}$ et $Q_3 = \boldsymbol{x_{(3p+3)}}$.

Remarque 1 : Dans ce dernier cas, on peut écrire **:** $\boxed{\boldsymbol{Q_i = x_{(i\times(p+1))}}}$ ($\boldsymbol{i = 1, 2, 3}$).

Remarque 2 (Mnémotechnique) :

Des 3 quartiles Q_1, Q_2 et Q_3, le nombre de **ceux qui sont,** dans chaque cas**, égaux à des observations (**dont il faut déterminer le rang**) est** le même que **le reste** $\boldsymbol{r}$ **de la division de** $\boldsymbol{n}$ **par 4.**

Ainsi **:**

- **quand** $\boxed{r = 0}$, il n'y a ***AUCUN*** des 3 qui soit une observation **;**
- **pour** $\boxed{\boldsymbol{r = 1}}$**, la médiane** est la **SEULE** qui est une observation **(de rang** $\boldsymbol{2p + 1}$**) ;**
- **si** $\boxed{\boldsymbol{r = 2}}$, $\boldsymbol{Q_1}$ **ET** $\boldsymbol{Q_3}$ sont des observations (*respectivement* **de rangs :** $\boldsymbol{p + 1}$ et $\boldsymbol{3p + 2}$**) ;**
- **et pour** $\boxed{\boldsymbol{r = 3}}$, **TOUS** les **3** sont des observations **(de rangs** $\boldsymbol{i(p + 1), i = 1, 2, 3}$).

Application **:**

a) $\boldsymbol{n = 18} = 4 \times \boldsymbol{4} + \boldsymbol{2}$ **:** $Q_1 = \boldsymbol{x_{(5)}}$, $Q_2 = \frac{x_{(9)}+x_{(10)}}{2}$ et $Q_3 = \boldsymbol{x_{(14)}}$;

b) $\boldsymbol{n = 27} = 6 \times \boldsymbol{4} + \boldsymbol{3}$ **:** $Q_1 = \boldsymbol{x_{(7)}}$, $Q_2 = \boldsymbol{x_{(14)}}$ et $Q_3 = \boldsymbol{x_{(21)}}$.

Deux autres cas **:**

c) $\boldsymbol{n = 1001} = 250 \times \boldsymbol{4} + \boldsymbol{1}$ **:** $Q_1 = \frac{x_{(250)}+x_{(251)}}{2}$, $Q_2 = \boldsymbol{x_{(501)}}$, $Q_3 = \frac{x_{(751)}+x_{(752)}}{2}$.

d) $\boldsymbol{n = 2022} = 500 \times \boldsymbol{4} + \boldsymbol{2}$ **:** $Q_1 = \boldsymbol{x_{(501)}}$, $Q_2 = \frac{x_{(1011)}+x_{(1012)}}{2}$, $Q_3 = \boldsymbol{x_{(1502)}}$.