

Solutions des exercices de la série n°2

N.B. les calculs et les diagrammes sont faits à l'aide de l'outil Excel de Microsoft

Exercice 1 : Dans une enquête effectuée par un institut de statistique, on a obtenu la répartition de l'âge des entreprises informatiques en Québec, qui ont une activité de développement d'applications et de logiciels spécialisés, les données sont regroupées dans le tableau suivant :

Répartition par classes	Nombre d'entreprises
Moins de 3 ans	84
De 3 à moins de 5 ans	37
De 5 à moins de 10 ans	39
10 ans et plus	55

1. Est-ce qu'on peut répondre à la question suivante : En combien d'année en moyenne les entreprises actives dans le secteur du développement d'applications et de logiciels spécialisés existent ?
2. Quelle est la valeur (en nombre d'années) à laquelle 50% des entreprises ont une existence inférieure ?

Solution de l'exercice 1

1. Il est difficile, peut être impossible, de répondre à cette question à cause de la dernière classe qui n'est pas bornée, cette classe doit être représenté par une valeur centrale, rien nous informe sur la limite supérieur. Il faudra donc éviter le calcul de la moyenne et chercher une autre mesure de tendance centrale.
2. On doit calculer la médiane, on va utiliser la formule suivante :

$$Me = b_{inf} + \left(\frac{\frac{n}{2} - N_{i-1}}{n_i} \right) c$$

Avant cela, on doit compléter le tableau-synthèse avec l'effectif cumulé N_i :

Répartition par classes	Nombre d'entreprises n_i	N_i
Moins de 3 ans	84	84
De 3 à moins de 5 ans	37	121
De 5 à moins de 10 ans	39	160
10 ans et plus	55	215
<i>Total</i>	215	

Si les données sont triées dans l'ordre croissant, on calcule le rang de la valeur médiane :

$$\frac{n}{2} = \frac{215}{2} = 107.5$$

La valeur médiane dont le rang est 107.5 se trouve dans la classe médiane [3, 5[. Les données à utiliser dans la formule sont : $b_{inf} = 3$; $N_{i-1} = 84$; $n_i = 37$ et $c = 2$. Le calcul nous donne le résultat suivant :

$$Me = 3 + \left(\frac{107.5 - 84}{37} \right) * 2 = 4.27$$

Conclusion : 50% des entreprises ont une existence inférieure à quatre années et trois mois.

Exercice 2 : On considère la distribution des entreprises du secteur Bâtiments et Travaux Publics, selon leurs chiffres d'affaires en millions de dirhams (MDh) :

C.A.	Fréquences en %
[2 , 4 [	15
[4 , 6 [	25
[6 , 12 [	48
[12 , 20 [	12

1. Déterminer le mode, la moyenne et la médiane de cette distribution, puis interpréter les résultats.
2. Tracer la boîte à moustache de la série statistique.
3. Calculer l'écart-type de la série statistique.
4. Calculer le coefficient de Fisher d'asymétrie et celui d'aplatissement.

Solution de l'exercice 2

1. Pour connaître la classe modale, il faudra éliminer l'effet de l'amplitude de la classe, on voit bien que les amplitudes sont différentes : 2, 2, 6 puis 8.

Pour cela on peut chercher la classe qui a le rapport $f_i/(b_i - a_i)$ le plus élevé, avec f_i la fréquence de la classe $[a_i, b_i[$. On peut de la même façon comparer les $n_i/(b_i - a_i)$ avec n_i l'effectif, la classe modale correspond au rapport le plus élevé.

C.A. $[a_i, b_i[$	Amplitude $b_i - a_i$	Fréquences en % (100 f_i)	$f_i/(b_i - a_i)$
[2 , 4 [	2	15	0.075
[4 , 6 [	2	25	0.125
[6 , 12 [	6	48	0.080
[12 , 20 [	8	12	0.015

La classe modale correspond à l'intervalle [4 , 6 [; on prendra comme valeur modale le centre de classe

$$Mo = \frac{4 + 6}{2} = 5$$

Pour calculer la moyenne et la médiane, on doit compléter le tableau-synthèse :

C.A.	Centre m_i	Fréquences en % (100 f_i)	Fréquences cumulées F_i	$f_i m_i$
[2 , 4 [	3	15	0.15	0.45
[4 , 6 [	5	25	0.40	1.25
[6 , 12 [	9	48	0.88	4.32
[12 , 20 [	16	12	1.00	1.92
Total		100		7.94

- La moyenne sera calculée sur la base des fréquences (poids de chaque classe) et des moyennes des classes :

$$\mu = \sum_{i=1}^4 f_i m_i = 7.94.$$

- Pour chercher la classe médiane, on doit cumuler au moins 50%=0.5 des données : la classe médiane correspond à l'intervalle [6, 12 [. Le calcul de la valeur médiane sera fait à partir de la formule utilisant les fréquences au lieu des effectifs :

$$Me = b_{inf} + \left(\frac{0.5 - F_{i-1}}{f_i} \right) * c$$

On a les données suivantes à utiliser dans le calcul :

$$b_{inf} = 6 ; F_{i-1} = 0.40 ; f_i = 0.48 \text{ et } c = 6$$

Le calcul donne :

$$Me = 6 + \left(\frac{0.5 - 0.4}{0.48} \right) * 6 = 7.25.$$

- Interprétation des résultats :

- ↳ Le chiffre d'affaire le plus fréquent observé dans le secteur Bâtiments et Travaux Publics (BTP) est de 5 000 000 DH,
- ↳ Tandis que le chiffre d'affaire moyen des entreprises du secteur (BTP) s'élève à 7 940 000 DH.
- ↳ On constate de même que 50% des entreprises du secteur BTP ont un chiffre d'affaire supérieur à 7 250 000 DH.

2. Construction d'une boîte à moustache (diagramme en boîte ou « box plot ») :

Tout d'abord, on doit calculer les quartiles Q_1 et Q_3 ; le deuxième quartile Q_2 est déjà calculé :

$$Q_2 = Me = 7.25$$

$$Q_1 = b_{inf} + \left(\frac{0.25 - F_{i-1}}{f_i} \right) * c$$

avec $b_{inf} = 4 ; F_{i-1} = 0.15 ; f_i = 0.25$ et $c = 2$

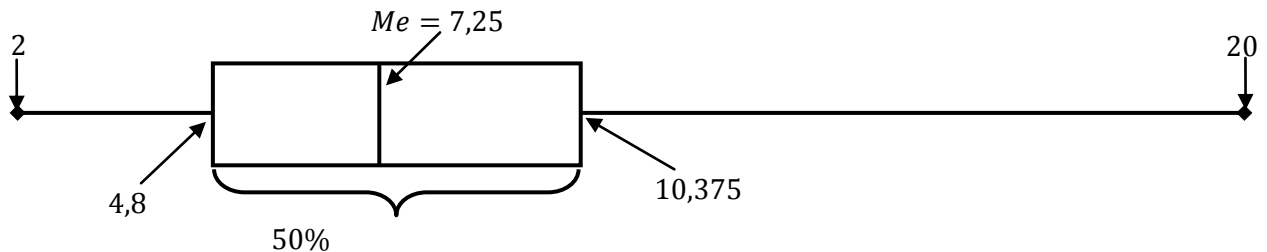
$$Q_1 = 4 + \left(\frac{0.25 - 0.15}{0.25} \right) * 2 = 4.8$$

Pour Q_3 :

$$Q_3 = b_{inf} + \left(\frac{0.75 - F_{i-1}}{f_i} \right) * c$$

avec $b_{inf} = 6$; $F_{i-1} = 0.40$; $f_i = 0.48$ et $c = 6$

$$Q_3 = 6 + \left(\frac{0.75 - 0.40}{0.48} \right) * 6 = 10.375$$



Il semblerait que la distribution est dissymétrique (asymétrique), pour se convaincre on va calculer le coefficient d'asymétrie de Fisher (question 4).

Remarque : si on cherchait dans la série des données des valeurs aberrantes, alors toute valeur aberrante doit être

- Supérieure à $Q_3 + 1.5 * (Q_3 - Q_1) = 18.7375$
- Ou inférieure à $Q_1 - 1.5 * (Q_3 - Q_1) = -3.5625$

On a pas ici la série des données, mais toute valeur observée dans l'intervalle $[18.74; 20[$ sera considérée comme valeur aberrante.

3. Calcul de l'écart-type :

On calcule en premier la variance, on utilisera la formule suivante (formule dite de Koenig) :

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^k n_i m_i^2 - \mu^2 = \sum_{i=1}^k f_i m_i^2 - \mu^2$$

On réutilise le tableau-synthèse :

C.A.	Centre m_i	Fréquences en % ($100 f_i$)	$f_i m_i^2$
[2 , 4 [	3	15	1.35
[4 , 6 [	5	25	6.25
[6 , 12 [	9	48	38.88
[12 , 20 [	16	12	30.72
Total		100	77.2

$$\sigma^2 = 77.2 - 7.94^2 = 14.16$$

L'écart-type de la série statistique est égale à :

$$\sigma = 3.76$$

Il correspond à un écart de 3 760 000 DH.

4. Calcul du coefficient de Fisher

$$m_h^* = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^h \cong \frac{1}{n} \sum_{j=1}^K n_j (m_j - \bar{x})^h = \sum_{j=1}^K f_j (m_j - \bar{x})^h$$

$$\bar{x} = 7.94$$

- Coefficient d'asymétrie de Fisher (ou skewness) est

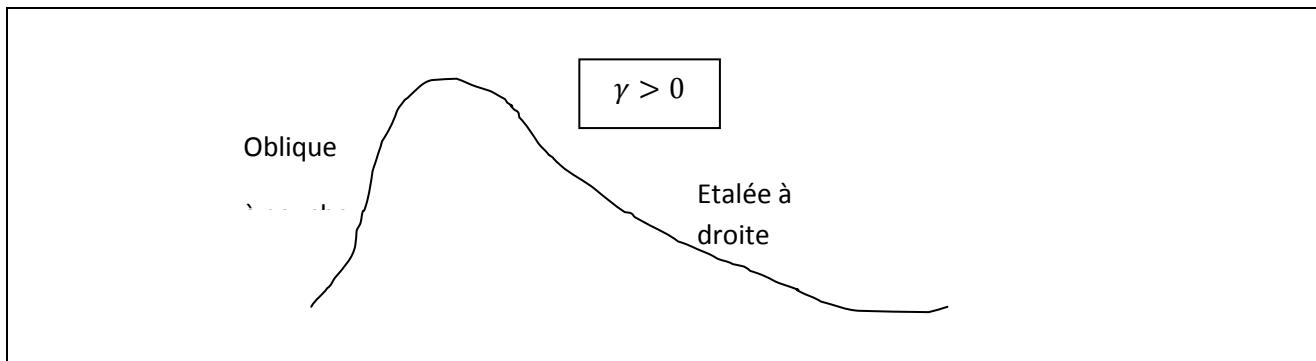
$$\gamma = \frac{m_3^*}{\sigma^3}$$

C.A.	Centre m_i	Fréquences en % (100 f_i)	$m_i - \mu$	$f_i (m_i - \mu)$	$f_i (m_i - \mu)^2$	$f_i (m_i - \mu)^3$	$f_i (m_i - \mu)^4$
[2 , 4 [	3	15	-4.94	-74.10	366.05	-1808.31	8933.04
[4 , 6 [	5	25	-2.94	-73.50	216.09	-635.30	1867.80
[6 , 12 [	9	48	1.06	50.88	53.93	57.17	60.60
[12 , 20 [	16	12	8.06	96.72	779.56	6283.28	50643.23
Total		100		0	1415.64	3896.84	61504.66

$$\sigma^2 = m_2^* = \frac{1415.64}{100} = 14.16$$

$$m_3^* = 38.97 \text{ et } m_4^* = 615.05 ;$$

$$\gamma = \frac{38.97}{53.28} = 0.73 > 0$$



La courbe de la distribution du chiffre d'affaire est non symétrique, elle est oblique à gauche et s'étale à droite.

- Coefficient d'aplatissement de Fisher (ou Kurtosis) est

$$\Gamma = \frac{m_4^*}{\sigma^4} - 3$$

$$\Gamma = \frac{615.05}{200.40} - 3 = 0.069 > 0$$

La courbe de la distribution est moins aplatie (courbe leptokurtique)

Exercice 3 : Une enquête sur la consommation annuelle d'électricité a été effectuée sur une population de 2600 ménages. Les résultats de l'enquête figurent dans le tableau suivant :

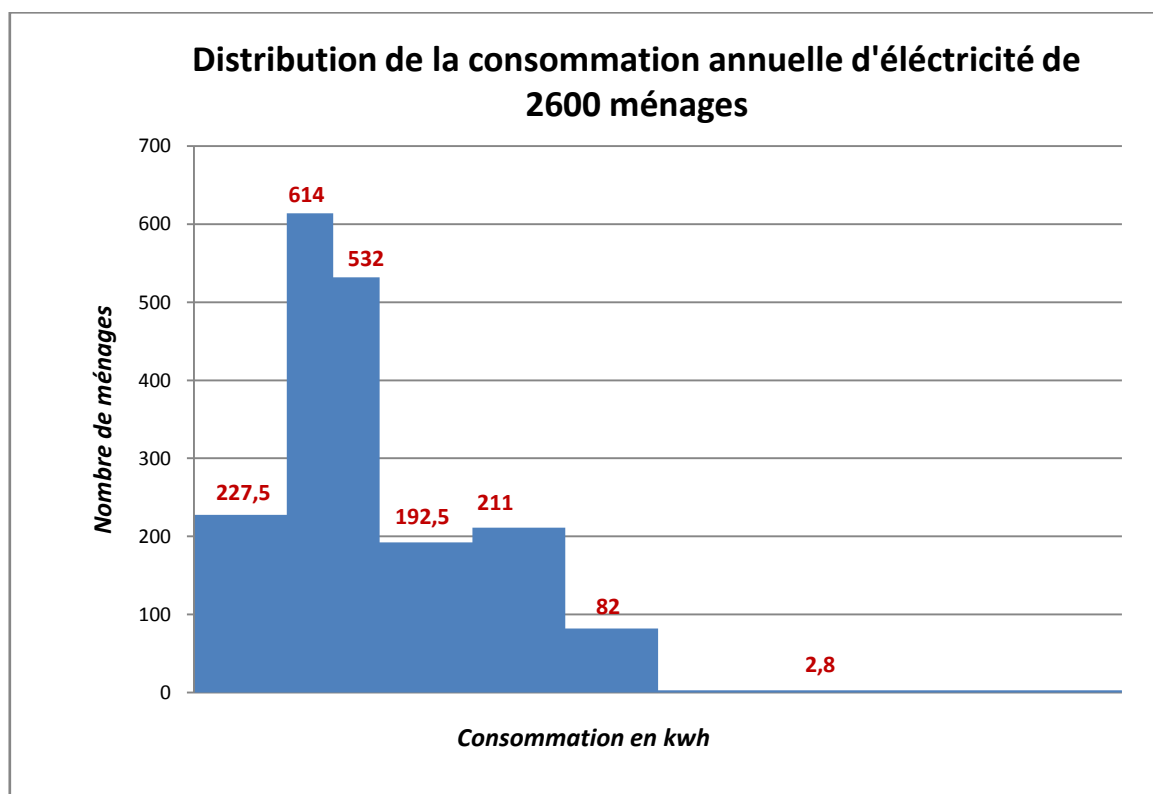
Consommation annuelle en kilowattheure(kwh)	Nombre de ménages
[0 , 200 [	455
[200 , 300 [	614
[300 , 400 [	532
[400 , 600 [	385
[600 , 800 [	422
[800 , 1000[	164
[1000 , 2000[	28

1. Construire l'histogramme des consommations annuelles.
2. Déterminer le mode et la médiane de cette distribution statistique.
3. Calculer la consommation annuelle moyenne et l'écart-type des consommations.

Solution de l'exercice 3

1. Dans le tableau ci-dessus, on a des amplitudes de classes différentes, 200, 100, 100, 200, 200, 200, 1000. On remarquera que le plus grand commun diviseur de ces amplitudes est 100.

On corrigera par exemple le premier effectif 455 en le divisant par 2, puisque l'amplitude de la première classe est égal à $200=2*100$. La même chose sera faite avec la 4^{ème}, 5^{ème} et 6^{ème} classe. Pour la dernière classe on divisera l'effectif 28 par 10 puisque l'amplitude de la classe qui lui correspond est égal à $1000=10*100$. On obtient ainsi l'histogramme suivant :



2. La classe modale est [200 , 300 [; elle correspond au rapport $(n_i/\text{amplitude})$ le plus élevé ; la valeur modale qu'on prendra est le centre de la classe :

$$Mo = (200 + 300)/2 = 250 \text{ kwh}$$

Consommation annuelle (kwh)	<i>amplitude</i>	Nombre de ménages n_i	$n_i/\text{amplitude}$
[0 , 200 [	200	455	2.275
[200 , 300 [	100	614	6.140
[300 , 400 [	100	532	5.320
[400 , 600 [	200	385	1.925
[600 , 800 [	200	422	2.110
[800 , 1000[	200	164	0.820
[1000 , 2000[	1000	28	0.028

Pour le calcul de la médiane, on va calculer les effectifs cumulés pour connaître la classe médiane.

On cherche la valeur de rang $n/2 = 1300$

Consommation annuelle (kwh)	Nombre de ménages n_i	N_i
[0 , 200 [	455	455
[200 , 300 [	614	1069
[300 , 400 [	532	1601
[400 , 600 [	385	1986
[600 , 800 [	422	2408
[800 , 1000[	164	2572
[1000 , 2000[	28	2600

La classe médiane est [300 , 400 [; la valeur médiane est la suivante :

$$Me = 300 + \left(\frac{1300 - 1069}{532} \right) * 100 = 343.42 \text{ kwh}$$

3. Calculons la consommation annuelle moyenne et l'écart-type des consommations.

On complète le tableau synthèse :

Consommation annuelle (kwh)	Centre m_i	Nombre de ménages n_i	$n_i m_i$ (masse de la classe)	$n_i m_i^2$
[0 , 200 [	100	455	45500	4550000
[200 , 300 [	250	614	153500	38375000
[300 , 400 [	350	532	186200	65170000
[400 , 600 [	500	385	192500	96250000
[600 , 800 [	700	422	295400	206780000
[800 , 1000[	900	164	147600	132840000
[1000 , 2000[	1500	28	42000	63000000
Total		2600	1062700	606965000

La moyenne μ à calculer est approchée par la formule approximative :

$$\mu = \frac{1}{2600} \sum_{i=1}^{2600} x_i \cong \frac{1}{2600} \sum_{j=1}^7 n_j m_j$$

Calcul numérique :

$$\mu = \frac{1062700}{2600} = 408.73 \text{ kwh}$$

Pour la variance, on utilise la formule (de Koenig) suivante :

$$\sigma^2 = \frac{1}{N} \sum_{j=1}^k n_j m_j^2 - \mu^2$$

On fait le calcul :

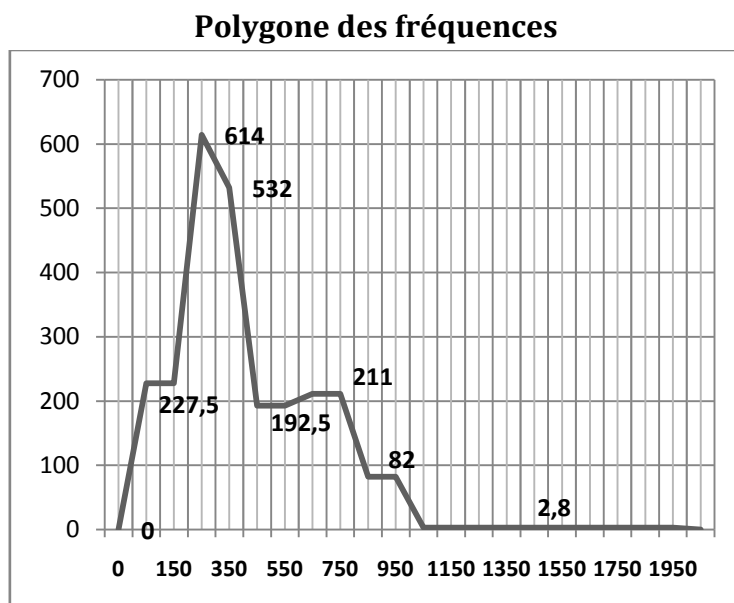
$$\sigma^2 = \frac{606965000}{2600} - 408.73^2 = 233448.08 - 167060.84 = 66387.24$$

Et l'écart-type sera : $\sigma = \sqrt{66387.24} = 257.66 \text{ kwh}$

Interprétation des résultats :

- ↘ La consommation la plus fréquente observée s'élève à 250 kwh , une telle consommation est observée chez presque un quart de la population (24%).
- ↘ La consommation annuelle moyenne des ménages en électricité est 408.73 kwh
- ↘ Tandis que 50% de ces ménages consomment moins de 343.42 kwh par an.

Remarque : la distribution de la consommation de l'électricité n'est pas symétrique, en effet on a $\mu > Me > Mo$, ce qui est très visible sur l'histogramme ou le polygone des fréquences. (avec une distribution oblique à gauche et un étalement à droite)



Exercice 4 : Le tableau suivant donne la répartition des salaires mensuels moyens en milliers de dirhams (kDH), ainsi que leurs écart-types, pour les trois catégories socio-professionnelle d'une usine :

Catégories	pourcentage %	Salaire moyen en kDH	Ecart-type
Cadres	10	8.2	1.6
Employés	30	4.2	0.9
Ouvriers	60	2.8	1.3

1. Rappeler la formule de calcul du salaire mensuel moyen puis calculer le.
2. Comparer les dispersions des salaires des trois catégories.

Solution de l'exercice 4

1. La moyenne à utiliser ici est la moyenne pondérée, sa formule est donnée par l'équation suivante :

$$\bar{x}_p = \frac{\sum_{i=1}^3 p_i x_i}{\sum_{i=1}^3 p_i}$$

Dans cet exercice les quantités p_i sont des pourcentages et $\sum_{i=1}^3 p_i = 100\% = 1$, on écrit ainsi :

$$\bar{x}_p = \sum_{i=1}^3 p_i x_i$$

Calcul numérique :

$$\begin{aligned}\bar{x}_p &= 10\% * 8.2 + 30\% * 4.2 + 60\% * 2.8 \\ &= 0.82 + 2.52 + 1.68 = 3.76 \text{ kDH}\end{aligned}$$

Le salaire mensuel moyen dans l'usine est égal à 3760 DH

2. Comparaison des trois écart-types :

On utilisera pour cela le coefficient de variation suivant,

$$V = \left| \frac{\sigma}{\mu} \right| * 100$$

Catégories	Salaire moyen en kDH μ_k	Ecart-type en kDH σ_k	Coefficient de variation $\left \frac{\mu_k}{\sigma_k} \right * 100$
Cadres	8.2	1.6	19.51%
Employés	4.2	0.9	21.43%
Ouvriers	2.8	1.3	46.43%

Il y'a une **forte disparité (différence) des salaires** dans la classe ouvrière par rapport aux deux autres classes.

Exercice 5 : Le vérificateur interne d'une entreprise a relevé le délai de recouvrement (en jours) des comptes auprès de 100 clients de l'entreprise. Les données sont dans le tableau suivant, triées dans l'ordre croissant :

40	40	42	42	46	46	47	47	47	47	48	49	49	49	50	52
52	52	52	53	53	54	57	57	59	61	61	61	63	66	67	68
68	69	70	75	79	79	80	80	81	86	88	88	88	89	90	91
91	92	94	98	98	99	99	100	101	101	102	103	103	104	104	104
106	109	109	111	111	112	113	113	114	114	115	118	118	119	119	120
121	122	123	127	127	128	129	130	131	132	132	132	133	133	134	135
138	138	139	139												

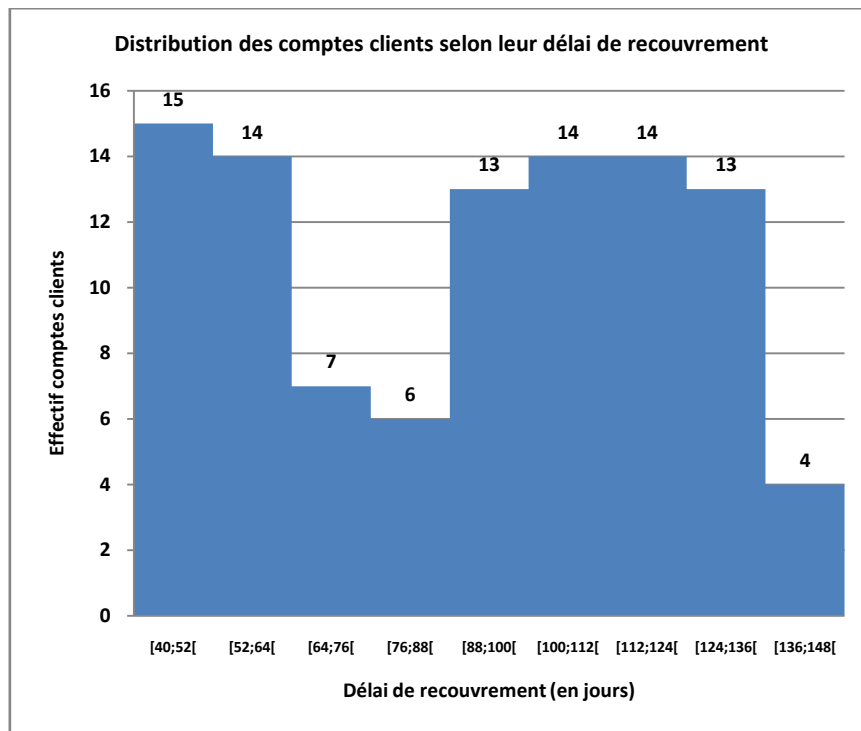
1. Construire un tableau synthèse regroupant les données, prenez une amplitude égale à 12.
2. Tracer l'histogramme correspondant et la courbe cumulative des pourcentages.
3. Déterminer le délai moyen de recouvrement des comptes, la variance et l'écart-type.
4. Déterminer la valeur médiane du délai de recouvrement. Que représente cette mesure de tendance centrale ?
5. En utilisant les données, quelle proportion de comptes clients ont un délai de recouvrement entre $\bar{x} - s$ et $\bar{x} + s$?
6. En utilisant la série de données, déterminer les valeurs du premier quartile Q_1 et du troisième quartile Q_3 .
7. Existe-t-il des valeurs aberrantes dans la série de données ?

Solution de l'exercice 5

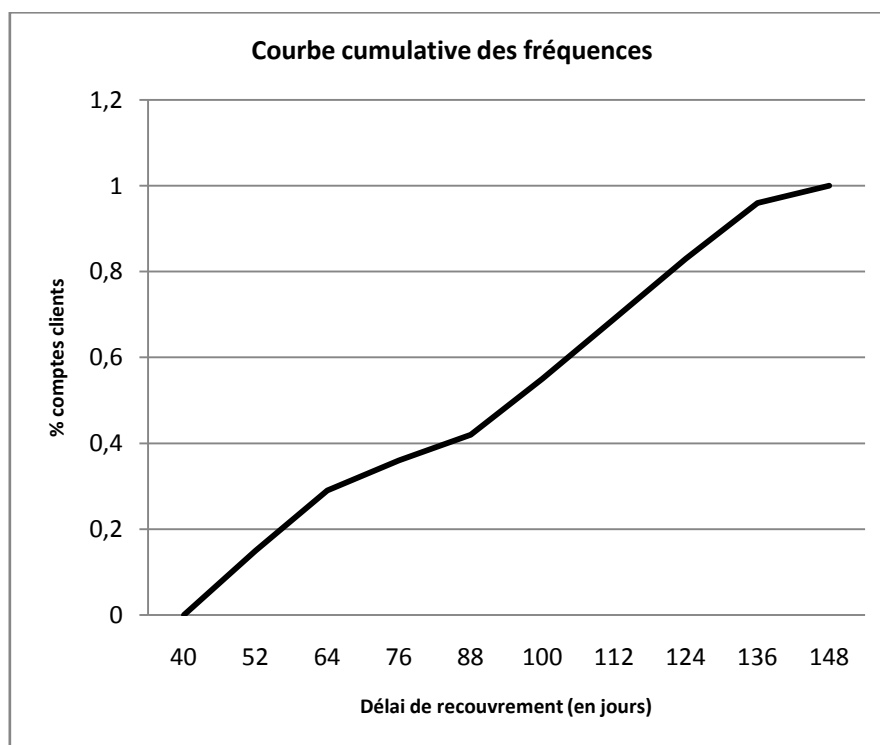
1. Le tableau-synthèse est le suivant,

Classes	n_i	N_i	f_i	F_i
[40, 52[	15	15	0.15	0.15
[52, 64[	14	29	0.14	0.29
[64, 76[	7	36	0.07	0.36
[76, 88[	6	42	0.06	0.42
[88, 100[	13	55	0.13	0.55
[100, 112[	14	69	0.14	0.69
[112, 124[	14	83	0.14	0.83
[124, 136[	13	96	0.13	0.96
[136, 148[	4	100	0.04	1.00
Total	100		1.00	

2. Histogramme des effectifs :



La courbe cumulative des fréquences :



3. Pour calculer le délai moyen de recouvrement, on complète le tableau-synthèse :

Classes	n_i	N_i	Centre m_i	f_i	F_i	$f_i m_i$	$f_i m_i^2$
[40;52[	15	15	46	0.15	0.15	6.9	317.4
[52;64[	14	29	58	0.14	0.29	8.12	470.96
[64;76[	7	36	70	0.07	0.36	4.9	343
[76;88[	6	42	82	0.06	0.42	4.92	403.44
[88;100[	13	55	94	0.13	0.55	12.22	1148.7
[100;112[	14	69	106	0.14	0.69	14.84	1573
[112;124[	14	83	118	0.14	0.83	16.52	1949.4
[124;136[	13	96	130	0.13	0.96	16.9	2197
[136;148[	4	100	142	0.04	1.00	5.68	806.56
<i>Total</i>	100					91	9209.44

Le délai moyen de recouvrement des comptes est égale à

$$\bar{x}_n = \sum_{i=1}^9 f_i m_i = 91 \text{ jours.}$$

La variance est égale à

$$\begin{aligned}
 s^2 &= \frac{1}{n} \sum_{i=1}^9 n_i m_i^2 - \bar{x}^2 = \sum_{i=1}^9 f_i m_i^2 - \bar{x}^2 \\
 &= 9209.44 - 91^2 = 928.44
 \end{aligned}$$

L'écart-type du délai de recouvrement des comptes est

$$s = \sqrt{928.44} = 30.47 \text{ jours}$$

4. La classe médiane est [88, 100[; le calcul de la valeur médiane nous donne

$$Me = 88 + \left(\frac{0.5 - 0.42}{0.13} \right) * 12 = 95.385 \text{ jours}$$

Cette valeur va nous permettre de positionner les clients, il ya 50% des clients qui ont un délai de recouvrement de plus de 95.4 jours.

5. En utilisant directement les données, on veut retrouver le pourcentage des clients qui ont un délai de recouvrement situé dans l'intervalle suivant :

$$[\bar{x} - s ; \bar{x} + s] = [91 - 30.47 ; 91 + 30.47] = [60.53 ; 121.47]$$

On compte un nombre de 56 parmi 100, càd on a un pourcentage de 56% de clients qui ont un délai de recouvrement entre 61 jours et 121 jours.

Remarque : si on cherche la proportion de comptes clients qui ont un délai de recouvrement entre $[\bar{x} - 2s ; \bar{x} + 2s] = [30.06 ; 151.94]$; on remarque déjà que tous les recouvrements observés sont dans l'intervalle $[30.06 ; 151.94]$ c.à.d. que la proportion est de 100%.

6. En utilisant la série des données on peut fournir les quartiles suivants :

$$Q_1 = q_{25\%} = \frac{x_{25} + x_{26}}{2} = \frac{59 + 61}{2} = 60 \text{ jours}$$

$$Q_3 = q_{75\%} = \frac{x_{75} + x_{76}}{2} = \frac{123 + 127}{2} = 125 \text{ jours}$$

7. On calcul la distance :

$$D = 1.5 * (Q_3 - Q_1) = 97.5$$

L'intervalle interquartile = $[Q_1, Q_3] = [60, 125]$

Une valeur est aberrante si elle s'éloigne de la valeur 125 d'au moins 97.5, c.à.d. supérieure à $125+97.5=222.5$ ou si elle s'éloigne dans le sens inverse de la valeur 60 d'au moins 97.5, c.à.d. qu'elle doit être inférieure à -37.5 ; aucune valeur ne satisfait à ces conditions d'où la série des données ne contient aucune valeur aberrante.

Délai moyen de recouvrement des comptes clients

Le délai moyen de recouvrement des comptes clients mesure le nombre moyen de jours qu'il faut aux clients pour payer leurs comptes.

Le résultat permet aussi de déterminer si les modalités de crédit dont se sert l'entreprise sont réalistes.

Pour déterminer si le délai moyen de recouvrement des comptes clients est adéquat, il suffit de le comparer aux modalités de crédit qu'on offre aux clients. Par exemple, si on donne aux clients 30 jours pour payer leur facture et que le délai moyen de recouvrement des comptes clients est de 45 jours, il y a un problème; en revanche, si le délai moyen est inférieur à 30 jours, c'est une bonne nouvelle.

Extrait de <https://www.bdc.ca/fr/articles-outils/boite-outils-entrepreneur/evaluation>

[Retour](#)