

PLAN:

Cours	2
1 Un peu de vocabulaire	2
2 DIAGRAMMES A BATONS OU HISTOGRAMMES	3
2.1 Quelles différences voyez-vous entre les deux graphiques ci-dessous ?	3
2.2 Le cas des classes d'amplitudes différentes	3
3. MODE, ETENDUE	3
3. MODE, ETENDUE	4
3.1 Définitions	4
3.2 Dans les exercices :	4
3.2.1 Données discrètes 9, 11, 8, 10, 13, 12, 10, 11, 10	4
3.2.2 Données réparties par classes	4
4. MEDIANE, QUARTILES, DECILES	5
4.1 Définitions	5
4.2 Dans les exercices :	6
a) Données discrètes "en vrac" 21, 25, 28, 30, 27, 24, 31, 21, 28, 30, 25, 28, 26, 25	6
b) Tableau d'effectifs	6
c) Données réparties par classes	6
4.3 Diagrammes en boîtes	7
5 MOYENNE, ECART TYPE, DONNEES GAUSSIENNES	7
5 MOYENNE, ECART TYPE, DONNEES GAUSSIENNES	8
5.1 Définitions	8
5.2 Dans les exercices :	8
a) Tableau des fréquences	8
b) Données réparties en classes	8
5.3 Propriétés	9
5.3.1 Addition ou Multiplication de toutes les données par un même nombre :	9
5.3.2 Moyennes partielles	9
5.4 Moyenne et médiane	9
5.5 Données Gaussiennes	9
5.5 Données Gaussiennes	10
6 QUELS INDICATEURS STATISTIQUES UTILISER ?	10
Exercices	11

1 UN PEU DE VOCABULAIRE

Toute étude statistique s'appuie sur des données. Dans le cas où ces données sont numériques (99% des cas), on distingue les données discrètes (qui prennent un nombre fini de valeurs : par ex, le nombre de voitures par famille en France) des données continues (qui prennent des valeurs quelconques : par ex, la taille des animaux d'un zoo).

- Dans le cas d'une série discrète, le nombre de fois où l'on retrouve la même valeur s'appelle l'effectif de cette valeur. Si cet effectif est exprimé en pourcentage, on parle alors de fréquence de cette valeur. (cf 17 p82)
- Dans le cas d'une série continue, on répartit souvent les données par classes. (cf 13 p82)

Dans les exercices, les données se présenteront donc ainsi :

données numériques	discrètes	données "en vrac"
		tableau des effectifs ou des fréquences
	continues	données "en vrac"
		données réparties par classes

Le but des statistiques est d'analyser les données dont on dispose :

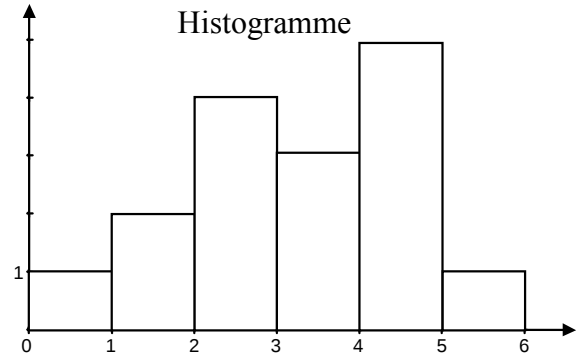
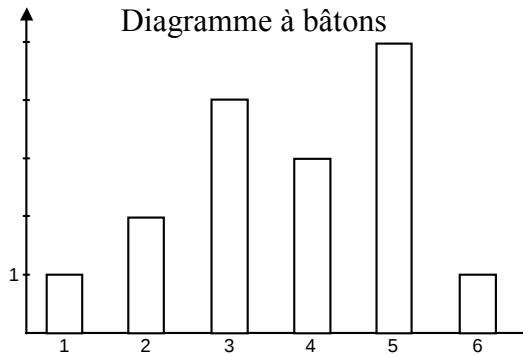
- Pour cela, on peut s'aider d'un graphique : Nous verrons notamment cette année les diagrammes à bâtons, les histogrammes et les diagrammes en boîtes (ou à moustaches).
- On peut aussi chercher à déterminer la moyenne ou la médiane de la série. De tels nombres permettent notamment de comparer plusieurs séries entre elles. On les appelle indicateurs statistiques ou paramètres statistiques. On distingue les indicateurs de position (qui proposent une valeur "centrale" de la série) et les indicateurs de dispersion (qui indiquent si la série est très regroupée autour de son "centre" ou non).

Nous étudierons cette année les indicateurs statistiques suivants :

Indicateurs de position :		Indicateurs de dispersion :
mode, classe modale		étendue
médiane, classe médiane	quartiles, déciles	écart interquartile
moyenne		écart type

2 DIAGRAMMES A BATONS OU HISTOGRAMMES

2.1 Quelles différences voyez-vous entre les deux graphiques ci-dessous ?

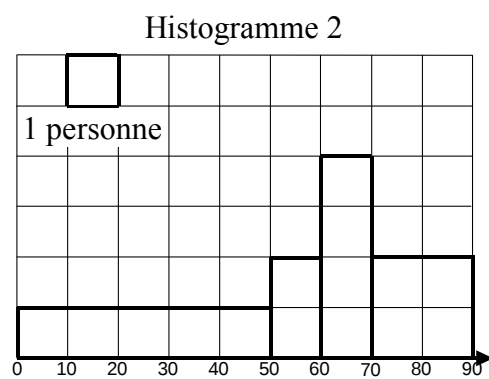
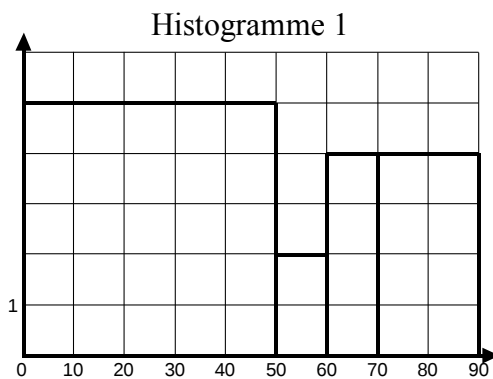


- Dans le diagramme à bâtons, l'axe des abscisses n'est pas gradué et la largeur des bâtons ne signifie rien.
- Dans l'histogramme, l'axe des abscisses est gradué et les bâtons sont donc "collés" les uns aux autres. L'histogramme est donc surtout utilisé pour représenter graphiquement des séries continues où les données ont été réparties en classes.
- Attention, Excel appelle histogramme les diagrammes à bâtons et ne sais pas faire de vrais histogrammes !

2.2 Le cas des classes d'amplitudes différentes

Pour représenter la série ci-contre, quel est le graphique le plus équitable ?

poids (Kg)	[0 ; 50[	[50 ; 60[	[60 ; 70[	[70 ; 90]
nbre de personnes	5	2	4	4



- L'histogramme 1 est inadapté car il laisse entendre que la majorité des gens pèsent moins de 50 kg !
- L'histogramme 2 est équitable car on a pondéré la hauteur de chaque bâton en tenant compte de l'amplitude de la classe. Pour construire ce deuxième histogramme, on réalise le tableau ci-dessous :

classe	[0 ; 50[	[50 ; 60[	[60 ; 70[	[70 ; 90]
effectif	5	2	4	4
amplitude	5	1	1	2
effectif/amplitude	1	2	4	2

Remarques :

- Dans l'histogramme 1, c'est la hauteur des bâtons qui permet de lire l'effectif. Dans l'histogramme 2, c'est l'aire des bâtons qui permet de lire l'effectif.
- Dans l'histogramme 2, nous n'avons pas tracé l'axe des ordonnées, car il aurait fallu le graduer en nombre de personnes par kilo ! Par contre, pour permettre la lecture du graphique, nous avons indiqué en légende la signification de l'unité d'aire.
- Dans les exercices, quand les classes ont toutes la même amplitude, on fait un histogramme de type 1, quand les classes ont des amplitudes différentes, on fait un histogramme de type 2.

3. MODE, ETENDUE

3.1 Définitions

Si les données d'une série sont discrètes, le mode est la ou les valeurs qui ont le plus grand effectif.

Si les données ont été réparties en classes, on parle alors plutôt de classe modale.

L'étendue d'une série est la différence entre la plus grande valeur et la plus petite.

3.2 Dans les exercices :

3.2.1 Données discrètes 9, 11, 8, 10, 13, 12, 10, 11, 10

Faisons le tableau des effectifs :

valeur	8	9	10	11	12	13
effectif	1	1	3	2	1	1

- Le mode est la valeur qui a le plus gros effectif, c'est-à-dire 10
- $13 - 8 = 5$ donc l'étendue de cette série est 5

Remarque :

Ici, vu le petit nombre de données, faire un tableau des effectifs est un peu artificiel. Par contre, dès que l'on travaille sur un nombre important de données, il devient vite très utile pour mettre en évidence le mode et l'étendue de la série.

3.2.2 Données réparties par classes

classe	[0 ; 5[	[5 ; 10[	[10 ; 15[	[15 ; 20]
effectif	0	5	14	2

- La classe modale est la classe qui a le plus gros effectif, c'est à dire la classe [10 ; 15[
- $20 - 5 = 15$ donc l'étendue de cette série est inférieure ou égale à 15

Remarque :

Par simplification, on dira souvent que l'étendue est 15 mais c'est un abus de langage ! En effet, dans le tableau des données ci dessus, rien ne permet d'affirmer que les valeurs extrêmes sont 5 et 20 !

4. MEDIANE, QUARTILES, DECILES

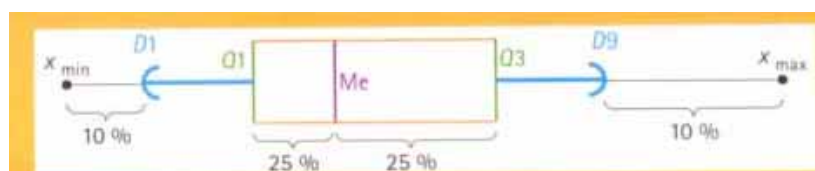
4.1 Définitions

Soit une série rangée par ordre croissant. Appelons n l'effectif total de la série.

Définitions	Pour déterminer le rang
La <u>médiane</u> • C'est la valeur "centrale" de la série. On dit qu'elle partage la série en deux moitiés	• si n est impair : la médiane est la valeur de rang $(n+1)/2$ • si n est pair : nous prendrons la moyenne des deux valeurs qui sont au centre de la série, c'est à dire dont les rangs entourent le nombre $n/2$
Les <u>quartiles</u> (partagent la série en 4 : il y en a donc 3) • Le 1^{er} quartile Q1 est la plus petite valeur telle que 25% des données lui soit inférieures ou égales. • Le 3^{ème} quartile Q3 est la plus petite valeur telle que 75% des données lui soit inférieures ou égales.	• Q1 est la valeur dont le rang est le premier entier supérieur ou égal à $n/4$ • Q3 est la valeur dont le rang est le premier entier supérieur ou égal à $n*3/4$
Les <u>déciles</u> (partagent la série en 10 : il y en a donc 9) • Le 1^{er} décile D1 est la plus petite valeur telle que 10% des données lui soit inférieures ou égales. • Le 9^{ème} décile D9 est la plus petite valeur telle que 90% des données lui soit inférieures ou égales.	• D1 est la valeur dont le rang est le premier entier supérieur ou égal à $n/10$ • D9 est la valeur dont le rang est le premier entier supérieur ou égal à $n*9/10$

Remarques :

- Les trois nombres Q1, méd, Q3 partagent la série en 4 parts égales (à une unité près)
- Si les données ont été réparties en classes, on ne peut déterminer la médiane exacte. En revanche, on appellera classe médiane, la classe qui la contient (et permet donc d'en donner un encadrement).
- L'intervalle [Q1 ; Q3] s'appelle l'intervalle interquartile.
- Le nombre Q3 – Q1 s'appelle l'écart interquartile.
- Quand il y a beaucoup de valeurs, on peut dire que les quartiles et la médiane séparent la série en quarts (25%) et les déciles en dixièmes (10%). On peut alors représenter la série par un diagramme en boîte, ou à moustache.



4.2 Dans les exercices :

a) Données discrètes "en vrac" 21, 25, 28, 30, 27, 24, 31, 21, 28, 30, 25, 28, 26, 25

Ordonnons la série par ordre croissant :

21, 21, 24, 25, 25, 25, 26, 27, 28, 28, 28, 30, 30, 31

Il y a 14 termes :

- $\frac{14+1}{2} = 7,5$. La médiane est donc la demi somme des 7^{ème} et 8^{ème} termes : méd = $(26 + 27) / 2 = 26,5$
- $\frac{14}{4} = 3,5$. Le 1^{er} quartile est donc le 4^{ème} terme : Q1 = 25
- $\frac{3 \times 14}{4} = 10,5$. Le 3^{ème} quartile est donc le 11^{ème} terme : Q3 = 28
- $8 - 5 = 3$. L'écart interquartile est donc $28 - 25 = 3$

b) Tableau d'effectifs

valeur	1	2	3	4	5	6
effectif	6	11	25	19	15	5
effectif cumulé	6	17	42	61	76	81

! Bien interpréter la dernière ligne !
La valeur 3 va du rang 18 au rang 42

L'effectif total est de

- $\frac{81+1}{2} = 41$. La médiane est donc le 41^{ème} terme : méd = 3
- $\frac{81}{10} = 8,1$. Le 1^{er} décile est donc le 9^{ème} terme : D1 = 2
- $\frac{81}{4} = 20,25$. Le 1^{er} quartile est donc le 21^{ème} terme : Q1 = 3
- $\frac{3 \times 81}{4} = 60,75$. Le 3^{ème} quartile est donc le 61^{ème} terme : Q3 = 4
- $\frac{9 \times 81}{10} = 72,9$. Le 9^{ème} décile est donc le 73^{ème} terme : D9 = 5

c) Données réparties par classes

classe	[0 ; 2[	[2 ; 4[	[4 ; 6[	[6 ; 8]
fréquence	10%	38%	45%	7%
fréquence cumulée	10	48	93	100

! Bien interpréter ce tableau !
48% des valeurs sont comprises
Entre [0 ; 4[
93% des valeurs sont < 6

48% des valeurs sont < 4

Et 93% des valeurs sont < 6

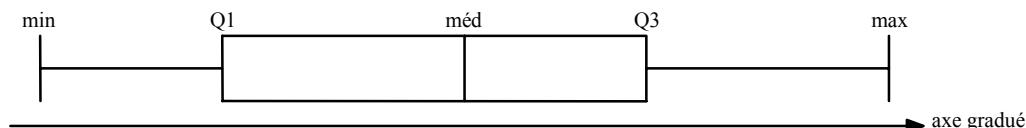
La classe médiane est donc la classe [4 ; 6[

On peut donc en déduire l'encadrement suivant $5 < \text{méd} < 6$

p80: 3, 4
1L-exo-statistiques.doc : I, J, K, L
vidéo-projecteur : 1L-cmp-quartiles.xls

4.3 Diagrammes en boîtes

Le diagramme en boîte d'une série à l'allure suivante :



Remarques :

- Lorsque la série est trop importante, que l'on ne connaît pas les valeurs extrêmes ou qu'on les considère comme non significatives, on raccourci souvent les moustaches aux déciles D1 et D9.
- La boîte centrale représente l'intervalle interquartile et contient donc la moitié des données.
- Vous devez légender votre diagramme (min, max, nom de la série) et graduer l'axe.
- On emploie surtout ce type de diagramme pour comparer plusieurs séries entre elles.
- Ces diagrammes ont reçu beaucoup de noms différents : boîtes à pattes, diagrammes à moustaches,...

Ex :

Deux classes de 1L comparent leurs résultats du trimestre et déclarent : "nos classes ont le même profil puisque dans les deux cas la médiane des résultats est 10". Qu'en pensez-vous ?

notes	5	6	7	8	9	10	11	12	13	14	15	16
effectifs 1L1	0	3	4	4	5	7	3	4	2	1	0	0
effectifs 1L2	2	4	3	3	3	4	3	2	2	3	1	2

- 1) Vérifier que les deux médianes valent 10 et déterminer les quartiles de chaque série
- 2) Tracer côte à côte les diagrammes en boîtes de ces deux séries.

Pour la 1L1 : L'effectif total est $3+4+4+\dots+1 = 33$

$$\frac{33+1}{2} = 17 \text{ donc la médiane est le } 17^{\text{ème}} \text{ terme de la série : Méd} = 10$$

$$\frac{33}{4} = 8,25 \text{ donc le } 1^{\text{er}} \text{ quartile est le } 9^{\text{ème}} \text{ terme de la série : } Q1 = 8$$

$$\frac{3 \times 33}{4} = 24,75 \text{ donc le } 3^{\text{ème}} \text{ quartile est le } 25^{\text{ème}} \text{ terme de la série : } Q3 = 11$$

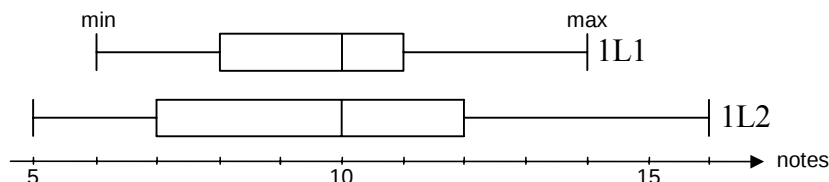
Pour la 1L2 : L'effectif total est $2+4+3+\dots+2 = 32$

$$\frac{32+1}{2} = 16,5 \text{ donc la médiane est la moyenne des } 16^{\text{ème}} \text{ et } 17^{\text{ème}} \text{ terme de la série : Méd} = \frac{10+10}{2} = 10$$

$$\frac{32}{4} = 8 \text{ donc le } 1^{\text{er}} \text{ quartile est le } 8^{\text{ème}} \text{ terme de la série : } Q1 = 7$$

$$\frac{3 \times 32}{4} = 24 \text{ donc le } 3^{\text{ème}} \text{ quartile est le } 24^{\text{ème}} \text{ terme de la série : } Q3 = 12$$

Diagrammes en boîtes :



Bilan : Le graphique ci-dessus met bien en évidence que l'écart interquartile et l'étendue sont plus resserrés en 1L1 qu'en 1L2 donc les élèves de 1L1 ont globalement un niveau plus homogène que ceux de 1L2.

p81: 7
p82: 12
1L-exo-statistiques.doc : M, N, O, P
vidéo-projecteur : 1L-outil-boites-a-moustaches.xls

5 MOYENNE, ECART TYPE, DONNEES GAUSSIENNES

5.1 Définitions

AP sur l'écart type puis reprendre les données, les présenter dans un tableau d'effectif et en déduire les 2 formules ci-dessous

Soit la série statistique ci-contre :

valeurs	x_1	x_2	...	x_p
effectifs	n_1	n_2	...	n_p

La moyenne est : $\bar{x} = \frac{n_1x_1 + n_2x_2 + \dots + n_px_p}{n_1 + n_2 + \dots + n_p}$

L'écart type est : $\sigma = \sqrt{\frac{n_1(x_1 - \bar{x})^2 + n_2(x_2 - \bar{x})^2 + \dots + n_p(x_p - \bar{x})^2}{n_1 + n_2 + \dots + n_p}}$

Remarques :

- L'écart type mesure la dispersion de la série autour de sa moyenne.
- Vous entendrez aussi parler de variance de la série. Il s'agit en fait de σ^2

$$V = \frac{n_1(x_1 - \bar{x})^2 + n_2(x_2 - \bar{x})^2 + \dots + n_p(x_p - \bar{x})^2}{n_1 + n_2 + \dots + n_p}$$

L'avantage de l'écart type sur la variance est qu'il s'exprime, comme la moyenne, dans la même unité que les données.

- Dans le cas de données regroupées en classes, on ne peut calculer la valeur exacte de la moyenne ou de l'écart type. On peut toutefois en déterminer une bonne approximation en remplaçant chaque classe par son milieu dans les formules ci-dessus.

5.2 Dans les exercices :

a) Tableau des fréquences

valeurs	12	13	14	15	16
fréquences	0,05	0,17	0,43	0,30	0,05

$$\bar{x} = (0,05*12+0,17*13+0,43*14+0,3*15+0,05*16)/1=14,13$$

$$\sigma = \text{RACINE}((0,05*(12-14,13)^2+0,17*(13-14,13)^2+0,43*(14-14,13)^2+0,3*(15-14,13)^2+0,05*(16-14,13)^2)/1)=0.9236$$

b) Données réparties en classes

classes	[0 ; 5[	[5 ; 10[	[10 ; 15[	[15 ; 20[
effectifs	7	12	14	2

Remplaçons chaque classe par son milieu :

$$\bar{x} \approx (7*2,5+12*7,5+14*12,5+2*17,5)/(7+12+14+2)=9$$

$$\sigma \approx \text{RACINE}((7*(2,5-9)^2+12*(7,5-9)^2+14*(12,5-9)^2+2*(17,5-9)^2)/(7+12+14+2))=4,27$$

p82: 13, 14
1L-exo-statistiques.doc : Q, R, S
Salle info : 1L-cmp-moyenne-ecart-type.xls

5.3 Propriétés

5.3.1 Addition ou Multiplication de toutes les données par un même nombre :

Ex Soit la série : 10, 12, 14.

$$\bar{x} = \frac{(10+12+14)}{3} = 12 \quad \text{et} \quad \sigma = \sqrt{\frac{(10-12)^2 + (12-12)^2 + (14-12)^2}{3}} = 1,63$$

Ajoutons 2 : la nouvelle série est : 12, 14, 16.

$$\bar{x} = \frac{(12+14+16)}{3} = 14 \quad \text{et} \quad \sigma = \sqrt{\frac{(12-14)^2 + (14-14)^2 + (16-14)^2}{3}} = 1,63$$

Divisons par 2 : la nouvelle série est : 6, 7, 8.

$$\bar{x} = \frac{(6+7+8)}{3} = 7 \quad \text{et} \quad \sigma = \sqrt{\frac{(6-7)^2 + (7-7)^2 + (8-7)^2}{3}} = 0,81$$

Cas général : Soit α un réel quelconque :

- Si l'on ajoute α à toutes les données, la moyenne augmente d' α
l'écart type ne change pas
- Si on multiplie toutes les données par α , la moyenne est multipliée par α
l'écart type est multipliée par α

5.3.2 Moyennes partielles

Ex : Sur les 5 premières interros, Paul a eu 12,5 de moyenne. Il vient d'avoir 15,5 à la 6^{ème} interro.

Les notes ayant toutes le même coefficient, quelle est sa nouvelle moyenne ?

La somme des notes des 5 premières interros est : $12,5 \times 5$

La somme des notes des 6 interros est donc : $12,5 \times 5 + 15,5$

La nouvelle moyenne est donc : $\bar{x} = \frac{12,5 \times 5 + 15,5}{6} = 13$

Cas général : Si on réunit deux groupes disjoints ayant respectivement pour moyennes et effectifs, $\bar{x}_1$ et n_1 d'une part, $\bar{x}_2$ et n_2 d'autre part, la moyenne de l'ensemble sera alors :

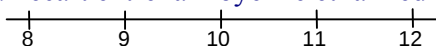
$$\bar{x} = \frac{n_1 \times \bar{x}_1 + n_2 \times \bar{x}_2}{n_1 + n_2}$$

5.4 Moyenne et médiane

- Quand on modifie les valeurs extrêmes d'une série, la moyenne change contrairement à la médiane qui ne change pas. On dit que la moyenne est "sensible aux valeurs extrêmes".
Il arrive que certaines de ces valeurs extrêmes soient douteuses ou influent de façon exagérée sur la moyenne. On peut alors, soit calculer une moyenne élaguée (c'est à dire recalculer la moyenne sans ces valeurs gênantes), soit utiliser la médiane.

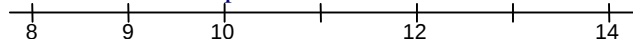
- Comment interpréter un écart entre la moyenne et la médiane ?

Soit la série suivante :



Ici la moyenne et la médiane sont identiques : la série est bien "centrée".

Soit la nouvelle série :



Ici la moyenne est plus importante que la médiane : la série est plus "étalée vers la droite".

p82: 16, 17

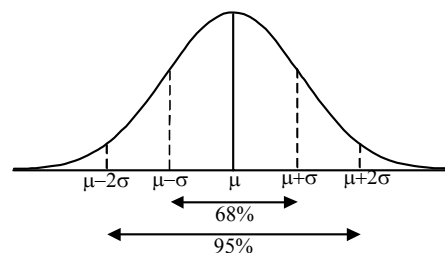
1L-exo-statistiques.doc : T, U, V

5.5 Données Gaussiennes

Dans de très nombreuses situations (issues de la biologie, géographie, sociologie, économie...) les données se présentent graphiquement sous la forme de courbes "en cloche" dites de Gauss.

Le comportement de ces séries est modélisable par une loi mathématique appelé loi normale ou loi de Gauss qui donne une grande importance à la moyenne μ et à l'écart type σ :

- Ces séries sont à peu près symétriques autour de μ
- Environ 68% des données sont dans l'intervalle $[\mu - \sigma ; \mu + \sigma]$
- Environ 95% des données sont dans l'intervalle $[\mu - 2\sigma ; \mu + 2\sigma]$
- Environ 99% des données sont dans l'intervalle $[\mu - 3\sigma ; \mu + 3\sigma]$



Les intervalles ci-dessus sont appelés plages de normalité pour les niveaux de confiance 0,68 ; 0,95 ; 0,99

Remarque : les observations ci-dessus n'ont aucun sens pour :

Les séries qui traduisent des phénomènes non gaussiens

Les séries gaussiennes pour lesquelles l'échantillon est trop petit.

6 QUELS INDICATEURS STATISTIQUES UTILISER ?

Dans la pratique :

- On utilise très peu le mode et l'étendue (faciles à déterminer mais simplistes !)
- On utilise la médiane, quartiles, déciles et écart interquartile surtout pour les séries à grands effectifs (pas de calculs, il suffit d'ordonner la série ; peu sensible aux valeurs douteuses)
- On utilise souvent la moyenne et l'écart type pour des séries de tailles intermédiaires ou des séries gaussiennes (la moyenne reste l'indicateur le plus intuitif ; intérêt des plages de normalité)

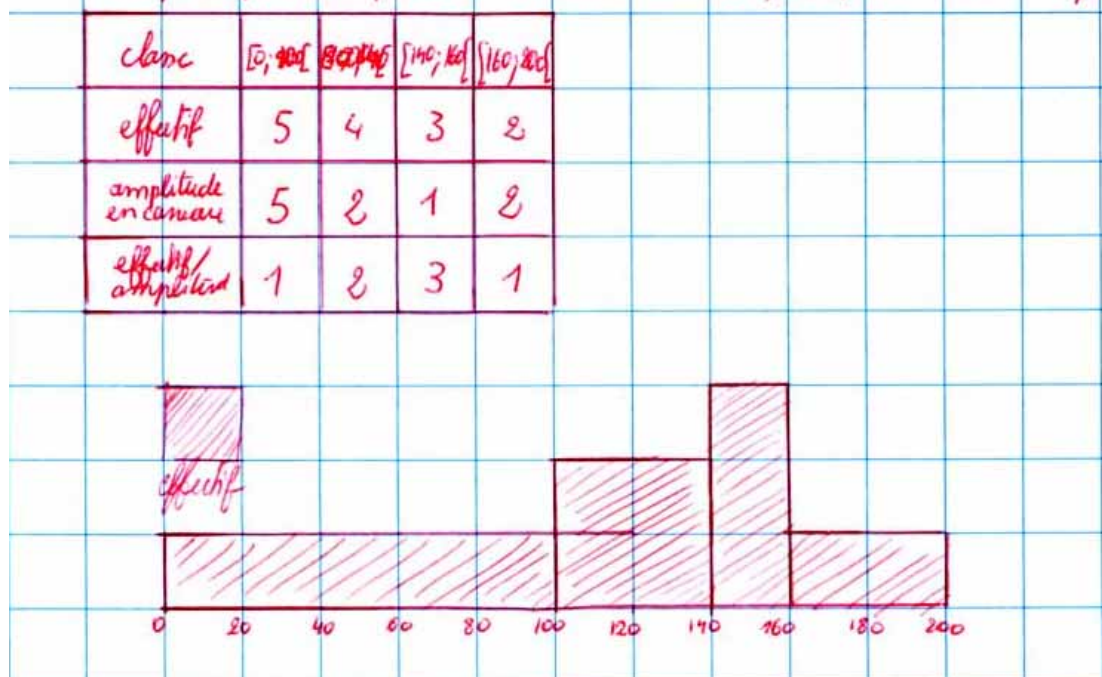
p81: 10
p82: 15
p83: 19
1L-exo-statistiques.doc : W, X, Y, Z
Salle info : 1L-cmp-donnees-gaussiennes.xls

EXERCICES

A) Tracer l'histogramme de la série ci-dessous :

classe	$[0;100[$	$[100;140[$	$[140;160[$	$[160;200[$
effectif	5	4	3	2

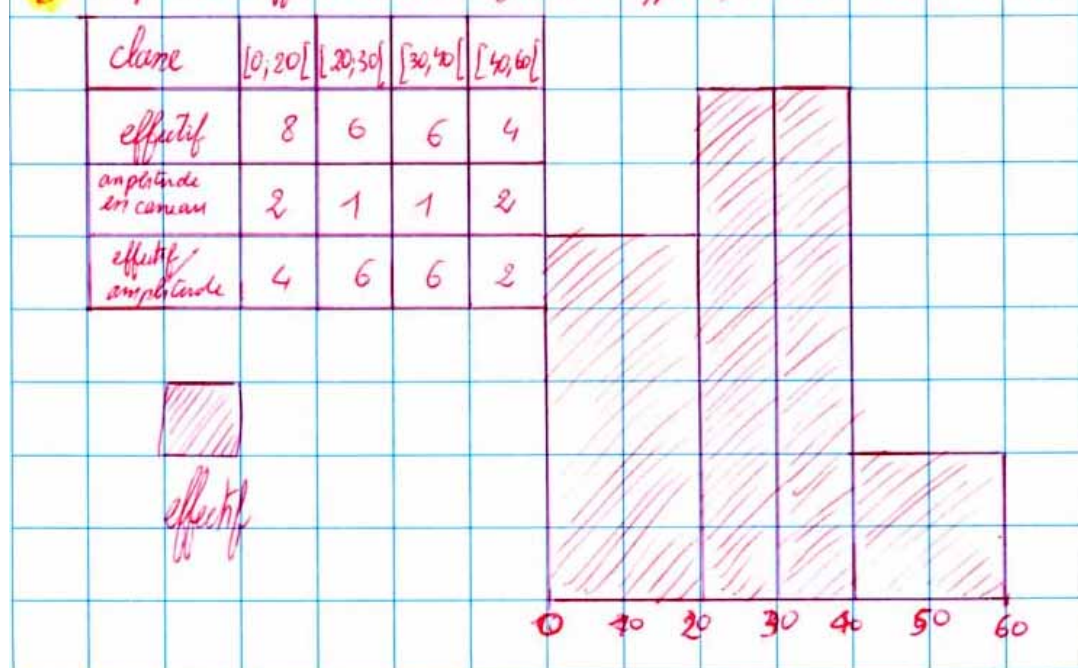
A les classes n'ont pas toutes la même amplitude donc histogramme type 2
la plus petite amplitude commune est 20, on prendra 1 carreau pour



B) Tracer l'histogramme de la série ci-dessous :

classe	$[0;20[$	$[20;30[$	$[30;40[$	$[40;60[$
effectif	8	6	6	4

B amplitudes différentes $\Rightarrow$ histogramme de type 2, On choisit 1 carreau = 10

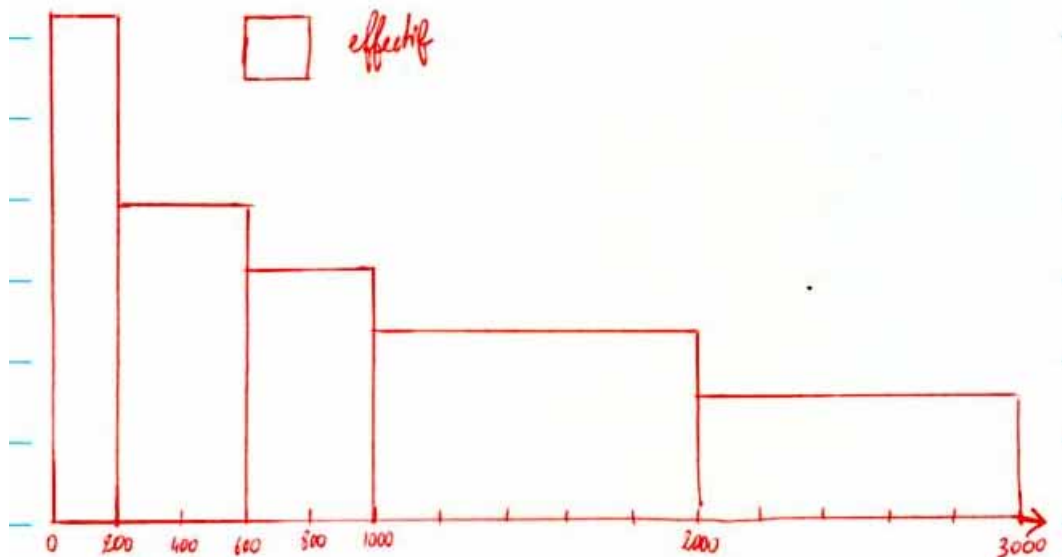


C) Tracer l'histogramme de la série ci-dessous :

classe	$[0;200[$	$[200;600[$	$[600;1000[$	$[1000;2000[$	$[2000;3000[$
effectif	8	10	8	15	10

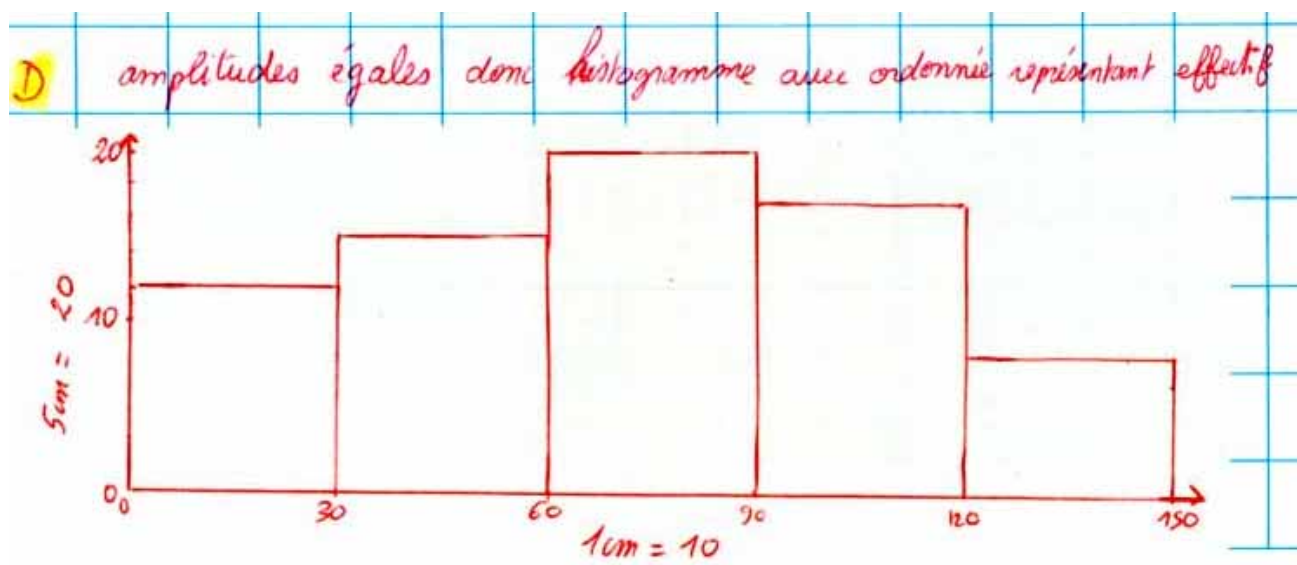
C amplitudes différentes 1 classe = 200

classe	$[0;200[$	$[200;600[$	$[600;1000[$	$[1000;2000[$	$[2000;3000[$
effectif	8	10	8	15	10
amplitude en classes	1	2	2	5	5
effectif / amplitude	8	5	4	3	2



D) Tracer l'histogramme de la série ci-dessous :

classe	$[0;30[$	$[30;60[$	$[60;90[$	$[90;120[$	$[120;150[$
effectif	12	15	20	17	8



E) Tracer l'histogramme de la série ci-dessous :

classe	$[150;155[$	$[155;160[$	$[160;165[$	$[165;170[$
fréquence	0,1	0,5	0,3	0,1



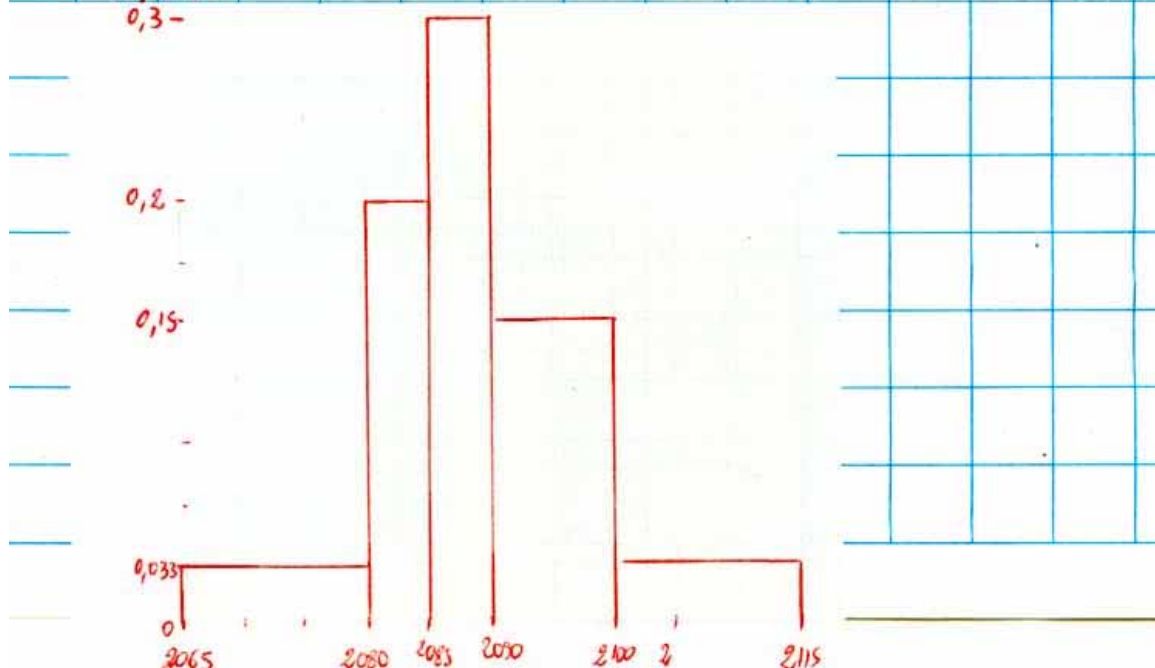
F) Tracer l'histogramme de la série ci-dessous :

classe	[2065;2080[	[2080;2085[	[2085;2090[	[2090;2100[	[2100;2115[
fréquence	0,1	0,2	0,3	0,3	0,1

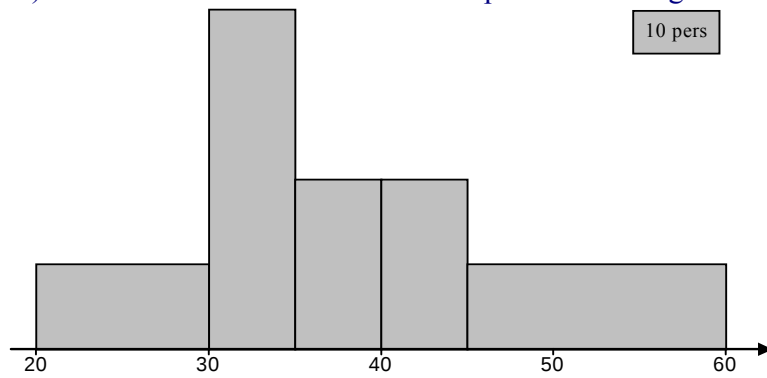
F classes d'amplitudes différentes $\Rightarrow$ axes représentatifs de la fréquence

On choisit 1 caseau pour 5 en amplitude

classe	[2065; 2080[	[2080; 2085[	[2085; 2090[	[2090; 2100[	[2100; 2115[
fréquence	0,1	0,2	0,3	0,3	0,1
amplitude en caseau	3	1	1	2	3
fréquence / amplitude	0,0333	0,2	0,3	0,15	0,033



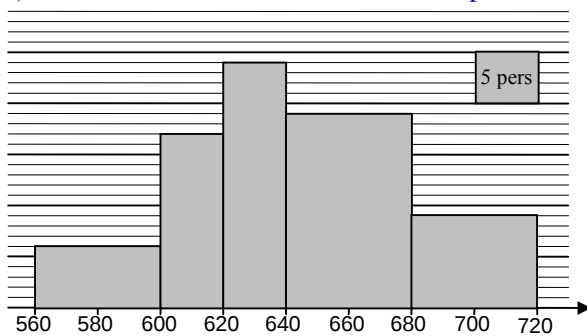
G) Retrouver le tableau des données à partir de l'histogramme ci-dessous



Correction :

classes	$[20 ; 30[$	$[30 ; 35[$	$[35 ; 45[$	$[45 ; 60[$
Distances verticales = effectif / largeur intervalle	20	80	40	20
Distances horizontales =	10	5	10	15
Effectif en nombre de personnes	$20 \cdot 10 = 200$	$80 \cdot 5 = 400$	$40 \cdot 10 = 400$	$20 \cdot 15 = 300$

H) Retrouver le tableau des données à partir de l'histogramme ci-dessous



H	classe	$[560 ; 600[$	$[600 ; 620[$	$[620 ; 640[$	$[640 ; 680[$	$[680 ; 720[$
	effectif en personnes	6	17	24	19	9

I) Déterminer la médiane, les 1^{er} et 3^{ème} quartiles, les 1^{er} et 9^{ème} déciles de la série ci-dessous :
Tailles en cm d'un groupe d'enfants de 5 à 7 ans :

104	107	107	107	108	108	109	110	111	111	112	112	112	112	113	113	114	114	114	114	115	115
115	115	115	116	116	117	117	117	118	118	118	119	119	120	120	120	121	121	122	123	123	125

Correction :

Commençons par ranger les valeurs :

Taille	104	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	125
effectif	1	3	2	1	1	2	4	2	4	5	2	3	3	2	3	2	1	2	1
rang	1	2 à 4	5 à 6	7	8	9 à 10	11 à 14	15 à 16	17 à 20	21 à 25	26 à 28	28 à 30	31 à 33	34 à 35	36 à 38	39 à 40	41	42 à 43	44

Il y a n=44 termes

- 1^{er} quartile $1/4 \times 44 = 11$ soit la 11^{ème} valeur $Q1 = 112$ cm
- 3^{ème} quartile $3/4 \times 44 = 33$ soit la 33^{ème} valeur $Q3 = 118$ cm
- 1^{er} décile $1/10 \times 44 = 4.4$ soit la 5^{ème} valeur $D1 = 108$ cm
- 9^{ème} décile $9/10 \times 44 = 39.6$ soit la 40^{ème} valeur $D9 = 121$ cm

J) 1) Déterminer la médiane, Q1, Q3, D1 et D9 pour les 100 nombres aléatoires de 1 à 40 ci-dessous.

1	2	5	7	9	11	13	15	18	21	22	23	26	27	29	30	31	32	36	37
1	3	5	8	9	11	14	15	19	22	22	23	27	28	29	30	31	33	36	38
1	4	5	8	10	12	14	16	19	22	22	24	27	28	30	30	31	34	37	38
1	4	6	8	10	12	14	16	19	22	23	25	27	29	30	31	32	35	37	39
2	5	6	9	11	12	14	17	21	22	23	26	27	29	30	31	32	35	37	40

Correction :

Commençons par ranger les valeurs :

Nombre	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
effectif	4	2	1	2	4	2	1	3	3	2	3	3	1	4	2	2	1	1	3	0
rang	1 à 4	5 à 6	7	8 à 9	10 à 13	14 à 15	16	17 à 19	20 à 22	23 à 24	25 à 27	28 à 30	31	32 à 35	36 à 37	38 à 39	40	41	42 à 44	

Nombre	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
effectif	2	7	4	1	1	2	5	2	4	6	5	3	1	1	2	2	4	2	1	1
rang	45 à 46	47 à 53	54 à 57	58	59	60 à 61	62 à 66	67 à 68	69 à 72	73 à 78	79 à 83	84 à 86	87	88	89 à 90	91 à 92	93 à 96	97 à 98	99	100

Il y a n=100 termes

- 1^{er} quartile $1/4 \times 100 = 25$ soit la 25^{ème} valeur $Q1 = 11$
- 3^{ème} quartile $3/4 \times 100 = 75$ soit la 75^{ème} valeur $Q3 = 30$
- 1^{er} décile $1/10 \times 100 = 10$ soit la 10^{ème} valeur $D1 = 5$
- 9^{ème} décile $9/10 \times 100 = 90$ soit la 90^{ème} valeur $D9 = 35$

2) On ajoute un nouveau 40 comme 101^{ème} nombre. Parmi les paramètres ci-dessus, quels sont ceux qui sont modifiés ?

Alors n=101 termes

- 1^{er} quartile $1/4 \times 101 = 25.25$ soit la 26^{ème} valeur $Q1 = 11$ identique à 1)
- 3^{ème} quartile $3/4 \times 101 = 75.75$ soit la 76^{ème} valeur $Q3 = 30$ identique à 1)
- 1^{er} décile $1/10 \times 101 = 10.1$ soit la 11^{ème} valeur $D1 = 5$ identique à 1)
- 9^{ème} décile $9/10 \times 101 = 90.9$ soit la 91^{ème} valeur $D9 = 36$

K) L'indice du coût de la construction est donné par le tableau suivant pour la période s'étendant du 2^{ème} trimestre 1997 au 4^{ème} trimestre 1999 :

Année	1997	1997	1997	1998	1998	1998	1998	1999	1999	1999	1999
Trimestre	2	3	4	1	2	3	4	1	2	3	4
Indice	1060	1067	1068	1058	1058	1057	1074	1071	1074	1080	1065

Calculer la médiane, les quartiles, l'écart interquartile et l'étendue de cette série.

Commençons par ranger les valeurs d'indice par ordre croissant

Indices	1058	1057	1060	1065	1067	1068	1071	1074	1080
effectif	2	1	1	1	1	1	1	2	1
rang	1 à 2	3	4	5	6	7	8	9 à 10	11

Il y a n=11 termes

- Médiane $(11+1)/2=6$ soit la valeur de rang 6 $M=1067$
- 1^{er} quartile $1/4*11=2.75$ soit la valeur de rang 3 $Q1=1057$
- 3^{ème} quartile $3/4*11=8.25$ soit la valeur de rang 9 $Q3=1074$
- écart interquartile $Q3 - Q1 = 1074 - 1067 = 7$
- étendue $\text{Max} - \text{Min} = 1080 - 1058 = 22$

L) On a relevé les valeurs de l'indice des prix à la consommation pour l'ensemble des ménages en France (hors tabac) de janvier 1998 à septembre 2000 (base 100 en 1998). On a trouvé :

99,5-99,8-100-100,2-100,2-100,3-100-100-100-99,9-100-99,6-99,9-100,3-100,6-100,6-100,6-100,3-100,5-100,6-100,7-100,7-101,2-101,1-101,2-101,7-101,7-101,9-102,2-102,2-102,7.

Déterminer la médiane, les quartiles, le 1^{er} et le 9^{ème} décile de cette série statistique.

Commençons par ranger les valeurs :

Nombre	99.5	99.6	99.8	99.9	100	100. 2	100. 3	100. 5	100.6	100. 7	101. 1	101. 2	101. 7	101. 9	102	102. 2	102. 7
effectif	1	1	1	2	5	2	3	1	4	2	1	3	2	1	1	2	1
rang	1	2	3	4 à 5	6 à 10	11 à 12	13 à 15	16	17 à 20	21 à 22	23	24 à 25	26 à 27	28	29	30 à 31	32

Il y 32 termes

- Médiane $(32)/2=16$ soit la 16^{ème} valeur $M=100.5$
- 1^{er} quartile $1/4*32=8$ soit la 8^{ème} valeur $Q1=100$
- 3^{ème} quartile $3/4*32=24$ soit la 24^{ème} valeur $Q3=101.2$
- 1^{er} décile $1/10*32=3.2$ soit la 4^{ème} valeur $D1=99.9$
- 9^{ème} décile $9/10*32=28.8$ soit la 29^{ème} valeur $D9=102$

M) Nombre de battements du cœur par minutes : Déterminer la médiane, les 1^{er} et 3^{ème} quartiles et les 1^{er} et 9^{ème} déciles de la série ci-dessous puis faire un diagramme en boîtes.

25	66	78	82	86	93	96	99	102	105	110	112	122	129	135	152
28	73	78	83	87	94	97	100	103	106	110	113	122	130	135	154
53	73	78	83	87	95	97	100	103	106	110	114	126	130	136	169
54	74	79	84	89	95	97	101	104	107	111	116	126	130	140	175
55	76	80	85	90	95	97	101	104	107	111	118	126	134	140	176
58	76	81	85	91	96	98	101	104	109	112	119	126	134	140	188
59	78	82	86	92	96	99	102	105	110	112	120	127	135	145	217

Commençons par ranger les valeurs :

Nombre	25	28	53	54	55	58	59	66	73	74	76	78	79	80	81	82	83	85	86	87
effectif	1	1	1	1	1	1	1	1	2	1	2	4	1	1	1	2	2	2	2	2
rang	1	2	3	4	5	6	7	8	9 à 10	11	12 à 13	14 à 17	18	19	20	21 à 22	23 à 24	26 à 27	28 à 29	30 à 31

Nombre	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	109
effectif	1	1	1	1	1	1	3	3	4	1	2	2	3	2	2	2	2	2	2	1
rang	32	33	34	35	36	37	38 à 40	41 à 43	44 à 47	48	49 à 50	51 à 52	53 à 55	56 à 57	58 à 59	60 à 62	63 à 64	65 à 66	67 à 68	69

Nombre	110	111	112	113	114	116	118	119	120	122	126	127	129	130	134	135	136	140	145	152
effectif	4	2	2	1	1	1	1	1	1	2	4	1	1	2	2	3	1	3	1	1
rang	70 à 73	74 à 75	76 à 78	79	80	81	82	83	84	85 à 86	87 à 90	91	92	93 à 94	96 à 97	98 à 100	101	102 à 104	105	106

Nombre	154	169	175	176	178	217
effectif	1	1	1	1	1	1
rang	107	108	109	110	188	112

Il y 112 termes

- Médiane $(112)/2=56$ soit la 56^{ème} valeur
- 1^{er} quartile $1/4*112=28$ soit la 28^{ème} valeur
- 3^{ème} quartile $3/4*112=84$ soit la 84^{ème} valeur
- 1^{er} décile $1/10*112=11.2$ soit la 12^{ème} valeur
- 9^{ème} décile $9/10*112=100.8$ soit la 101^{ème} valeur

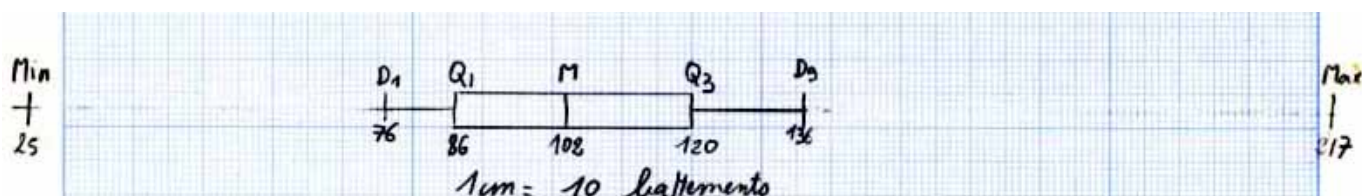
M=102 battements par minutes

Q1=86 battements par minutes

Q3=120 battements par minutes

D1=76 battements par minutes

D9=136 battements par minutes



N) On a relevé dans 2 tableaux les âges de décès de 90 hommes et 88 femmes

Hommes														
25	65	79	59	54	77	72	33	74	68	66	74	71	64	87
66	77	84	84	79	77	59	60	81	44	65	83	32	66	82
74	78	76	76	54	59	56	78	84	62	48	78	41	60	40
79	70	72	34	37	59	52	68	65	88	54	59	82	76	76
75	61	73	76	52	68	87	62	72	52	71	78	80	80	58
86	75	83	63	63	80	59	84	64	64	78	78	63	75	63

Femmes														
94	82	63	56	90	88	72	76	88	89	94	67	82	82	62
97	88	68	89	91	56	85	77	68	81	88	46	73	97	78
83	75	91	80	95	80	76	63	98	73	66	90	71	73	66
55	53	95	57	69	62	69	80	90	82	73	78	91	88	87
73	80	87	87	84	64	78	77	64	54	83	89	85	86	74
44	84	97	54	76	87	59	80	72	90	90	70	87		

Tracer les diagrammes en boîtes de ces deux séries. Que peut-t-on en conclure ?

Correction :

Ages	25	32	33	34	37	40	41	44	48	52	54	56	58	59	60	61	62	63	64	65
effectif	1	1	1	1	1	1	1	1	1	3	3	1	1	6	2	1	2	4	4	2
rang	1	2	3	4	5	6	7	8	9	10 à 12	13 à 15	16	17	18 à 23	24 à 25	26	27 à 28	29 à 32	33 à 35	36 à 38

Age	66	68	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	86	87	88
effectif	3	3	1	2	3	1	3	3	5	3	5	3	3	1	2	2	3	1	2	1
rang	39 à 41	42 à 44	45	46 à 47	48 à 49	51	52 à 54	55 à 57	58 à 62	63 à 65	67 à 71	72 à 74	75 à 77	78	79 à 80	81 à 82	84 à 86	87	88 à 89	90

Il y a n=90 termes à

- Médiane $(90)/2=45$ soit la 45^{ème} valeur $M=70$ ans
- 1^{er} quartile $1/4*90=22.5$ soit la 23^{ème} valeur $Q1=59$ ans
- 3^{ème} quartile $3/4*90=67.5$ soit la 68^{ème} valeur $Q3=78$ ans
- 1^{er} décile $1/10*90=9$ soit la 9^{ème} valeur $D1=48$ ans
- 9^{ème} décile $9/10*90=81$ soit la 81^{ème} valeur $D9=81$ ans

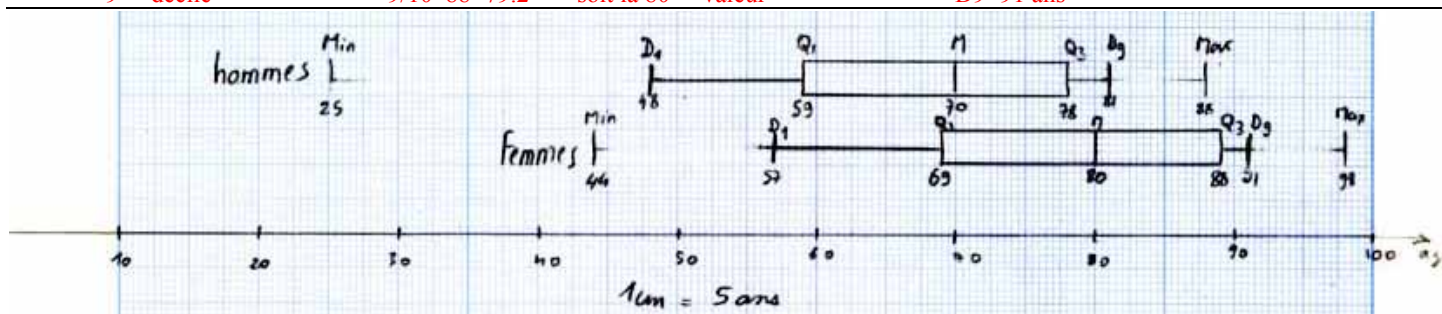
Femmes

Ages	44	46	53	54	55	56	57	59	62	63	64	66	67	68	69	70	71	72	73	74
effectif	1	1	2	2	1	2	1	1	2	2	2	2	1	2	2	1	1	2	5	1
rang	1	2	3	4 à 5	6	7 à 8	9	10	11 à 12	13 à 14	15 à 16	17 à 18	19	20 à 21	22 à 23	24	25	26 à 27	28 à 32	33

Age	75	76	77	78	80	81	82	83	84	85	86	87	88	89	90	91	94	95	97	98
effectif	1	3	2	3	5	1	4	2	2	2	1	5	5	3	5	3	2	2	3	1
rang	34	35 à 37	38 à 39	40 à 42	43 à 47	48	49 à 52	53 à 54	55 à 56	57 à 58	59	60 à 64	65 à 69	70 à 72	73 à 77	78 à 80	81 à 82	83 à 84	85 à 87	88

Il y a n=88 termes

- Médiane $(88)/2=44$ soit la 44^{ème} valeur $M=80$ ans
- 1^{er} quartile $1/4*88=22$ soit la 22^{ème} valeur $Q1=69$ ans
- 3^{ème} quartile $3/4*88=66$ soit la 66^{ème} valeur $Q3=88$ ans
- 1^{er} décile $1/10*88=8.8$ soit la 9^{ème} valeur $D1=57$ ans
- 9^{ème} décile $9/10*88=79.2$ soit la 80^{ème} valeur $D9=91$ ans



O) On compare les températures maximales moyennes (en °C) de chaque mois de l'année pour deux communes de Haute-Savoie situées à 1000m d'altitude : Chamonix et La Clusaz. (Atlas climatique de la Haute-Savoie, 1991)

Mois	1	2	3	4	5	6	7	8	9	10	11	12
Chamonix	1,5	4	7,5	12	15,5	20	23	22	19	14	6,5	2
La Clusaz	2,5	3,5	6	9,5	14	17	20,5	20,0	17	13	7	3,5

1) Tracer les diagrammes en boîte de ces deux séries (sans les déciles).

2) Interpréter les différences constatées.

Correction :

Chamonix

valeur	1,5	2	4	6,5	7,5	12	14	15,5	19	20	22	23
effectif	1	1	1	1	1	1	1	1	1	1	1	1
rang	1	2	3	4	5	6	7	8	9	10	11	12

Il y a n=12 termes

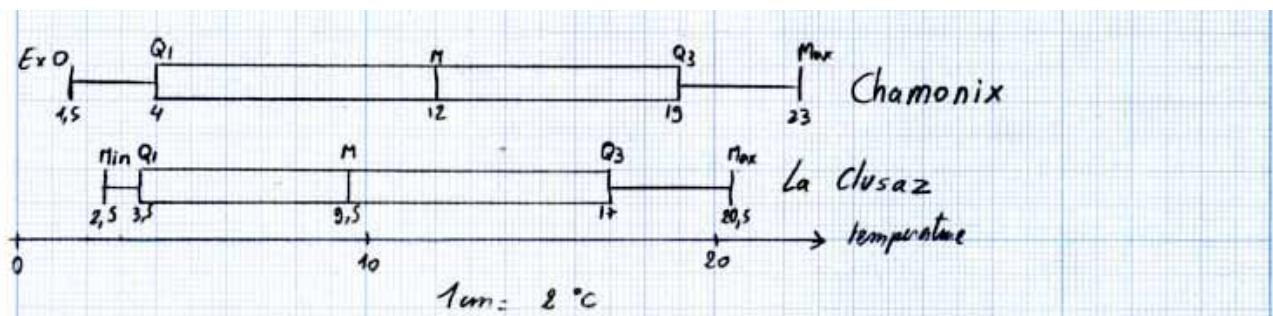
- Médiane $(12)/2=6$ soit la 6^{ème} valeur $M=12\text{ °C}$
- 1^{er} quartile $1/4*12=3$ soit la 3^{ème} valeur $Q1=4\text{ °C}$
- 3^{ème} quartile $3/4*12=9$ soit la 9^{ème} valeur $Q3=19\text{ °C}$

La Clusaz

valeur	2,5	3,5	6	7	9,5	13	14	17	20	20,5
effectif	1	2	1	1	1	1	1	2	1	1
rang	1	2 à 3	4	5	6	7	8	9 à 10	11	12

Il y a n=12 termes

- Médiane $(12)/2=6$ soit la 6^{ème} valeur $M=9,5\text{ °C}$
- 1^{er} quartile $1/4*12=3$ soit la 3^{ème} valeur $Q1=3,5\text{ °C}$
- 3^{ème} quartile $3/4*12=9$ soit la 9^{ème} valeur $Q3=17\text{ °C}$



Le climat est plus froid et moins contrasté à La Clusaz

P) Les tableaux suivants présentent la répartition de la population française âgée de moins de 18 ans au 1^{er} janvier 2000, selon l'année de naissance (en milliers d'individus, les valeurs étant arrondies).

Né en	1999	1998	1997	1996	1995	1994	1993	1992	1991	1990	1989	1988	1987	1986	1985	1984	1983	1982
Garçons	376	371	367	371	369	361	362	379	388	391	391	396	396	402	397	391	387	414
Filles	358	354	350	353	354	345	346	361	370	372	375	378	379	383	379	375	371	396

- 1) Déterminer la médiane, les quartiles, le premier et le neuvième déciles de ces deux séries.
- 2) Tracer les diagrammes en boîte à moustaches de ces deux séries dans un même repère. Commenter.

Correction :

Garçons

Né en	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999
Effectif	414	387	391	397	402	396	396	391	391	388	379	362	361	369	371	367	371	376
rang de	1	415	802	1193	1590	1992	2388	2784	3175	3566	3954	4333	4695	5056	5425	5796	6163	6534
à	414	801	1192	1589	1991	2387	2783	3174	3565	3953	4332	4694	5055	5424	5795	6162	6533	6909

Il y a n=6909 termes

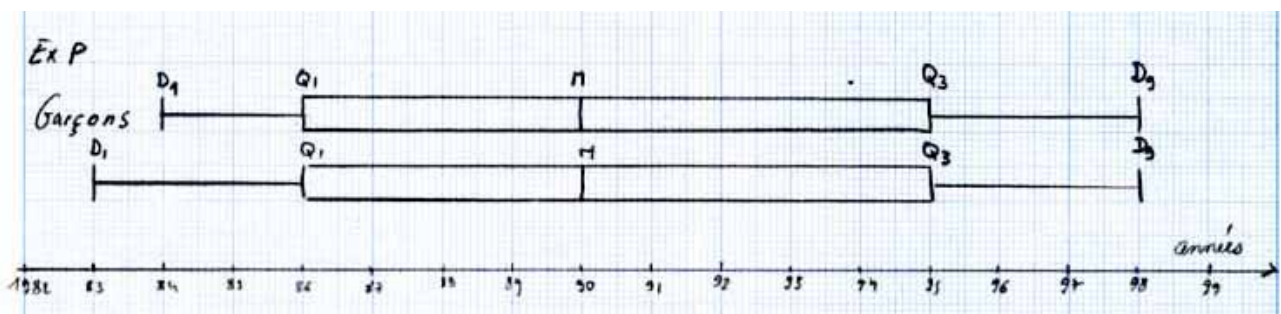
- Médiane $(6909+1)/2=3455$ soit la 3455^{ème} valeur $M=1990$
- 1^{er} quartile $1/4*6909=1727.25$ soit la 1728^{ème} valeur $Q1=1986$
- 3^{ème} quartile $3/4*6909=5181.75$ soit la 5182^{ème} valeur $Q3=1995$
- 1^{er} décile $1/10*6909=690.9$ soit la 691^{ème} valeur $D1=1984$
- 9^{ème} décile $9/10*6909=6218.1$ soit la 6219^{ème} valeur $D9=1998$

Filles

Née en	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999
Effectif	396	371	375	379	383	379	378	375	372	370	361	346	345	354	353	350	354	358
rang de	1	397	768	1143	1522	1905	2284	2662	3037	3409	3779	4140	4486	4831	5185	5538	5888	6242
à	396	767	1142	1521	1904	2283	2661	3036	3408	3778	4139	4485	4830	5184	5537	5887	6241	6599

Il y a n=6599 termes

- Médiane $(6599+1)/2=3300$ soit la 3300^{ème} valeur $M=1990$
- 1^{er} quartile $1/4*6599=1649.75$ soit la 1650^{ème} valeur $Q1=1986$
- 3^{ème} quartile $3/4*6599=4949.25$ soit la 4950^{ème} valeur $Q3=1995$
- 1^{er} décile $1/10*6599=659.9$ soit la 660^{ème} valeur $D1=1983$ différent des garçons
- 9^{ème} décile $9/10*6599=5939.1$ soit la 5940^{ème} valeur $D9=1998$



Q) Déterminer la médiane et les 1^{er} et 3^{ème} quartiles puis la moyenne et l'écart type de la série ci-dessous :
Poids en Kg d'un groupe d'enfants :

15	17	18	18	19	19	19	20	20	21	21	22	23	23	24
15	17	18	18	19	19	20	20	20	21	22	23	23	24	25
17	17	18	19	19	19	20	20	21	21	22	23	23	24	26

Correction :

valeur	15	17	18	19	20	21	22	23	24	25	26
effectif	2	4	5	8	7	5	3	6	3	1	1
rang	1 à 2	3 à 6	7 à 11	12 à 19	20 à 26	27 à 31	32 à 34	35 à 40	41 à 43	44	45

Il y a n=45 termes

- 1^{er} quartile $1/4 \times 45 = 11.25$ soit la 12^{ème} valeur $Q_1 = 19$ kg
- 3^{ème} quartile $3/4 \times 45 = 33.75$ soit la 34^{ème} valeur $Q_3 = 22$ kg
- Moyenne $(2 \times 15 + 4 \times 17 + 5 \times 18 + 8 \times 19 + 7 \times 20 + 5 \times 21 + 3 \times 22 + 6 \times 23 + 3 \times 24 + 1 \times 25 + 1 \times 26) / 45 = 20.3$ kg
- écart type $= \text{RACINE}(2 \times (15-20,3)^2 + 4 \times (17-20,3)^2 + 5 \times (18-20,3)^2 + 8 \times (19-20,3)^2 + 7 \times (20-20,3)^2 + 5 \times (21-20,3)^2 + 3 \times (22-20,3)^2 + 6 \times (23-20,3)^2 + 3 \times (24-20,3)^2 + 1 \times (25-20,3)^2 + 1 \times (26-20,3)^2) / 45 = 0.38$ kg

R) Deux tireurs s'entraînent au tir à la cible. Ils ont noté leurs résultats en points obtenues au bout de 30 tirs :

Points	50	30	20	10	0
Tireur A	8	9	8	4	1
Tireur B	6	16	3	3	2

- 1) Calculer la moyenne et l'écart-type pour les séries des résultats des deux tireurs.
- 2) Comparer ces résultats. Lequel est le plus régulier ?

Correction

1)

Tireur A

- Moyenne $= (8 \times 50 + 9 \times 30 + 8 \times 20 + 4 \times 10 + 1 \times 0) / (8 + 9 + 8 + 4 + 1) = 29$
- Ecart type $\text{RACINE}((8 \times (50-29)^2 + 9 \times (30-29)^2 + 8 \times (20-29)^2 + 4 \times (10-29)^2 + 1 \times (0-29)^2) / (8 + 9 + 8 + 4 + 1)) = 14.6$

Tireur B

- Moyenne $= (6 \times 50 + 16 \times 30 + 3 \times 20 + 3 \times 10 + 2 \times 0) / (6 + 16 + 3 + 3 + 2) = 29$
- Ecart type $= \text{RACINE}(6 \times (50-29)^2 + 16 \times (30-29)^2 + 3 \times (20-29)^2 + 3 \times (10-29)^2 + 2 \times (0-29)^2) / (6 + 16 + 3 + 3 + 2) = 2.5$

2)

Ecart type du tireur B < Ecart type du tireur A donc le tireur B est plus régulier que le tireur A.

S) On a mesuré la fluorescence de la chlorophylle α (en millivolts) dans un océan. Calculer la moyenne et l'écart-type de cette série :

Fluorescence	[15;20[	[20;25[	[25;30[	[30;35[	[35;40[	[40;45[
Effectif	8	13	21	20	14	4

Correction

On prend le centre de chaque intervalle

- Moyenne $= (8 \times 17,5 + 13 \times 22,5 + 21 \times 27,5 + 20 \times 32,5 + 14 \times 37,5 + 4 \times 42,5) / (8 + 13 + 21 + 20 + 14 + 4) = 29,4375$
- Ecart type $= \text{RACINE}((8 \times (17,5-29,4375)^2 + 13 \times (22,5-29,4375)^2 + 21 \times (27,5-29,4375)^2 + 20 \times (32,5-29,4375)^2 + 14 \times (37,5-29,4375)^2 + 4 \times (42,5-29,4375)^2) / (8 + 13 + 21 + 20 + 14 + 4)) = 6,7312$

T) Un élève a obtenu les notes suivantes à une série de devoirs de français : 8, 12, 10 et 14

1) Calculer la moyenne et l'écart-type de cette série de note.

2) Son professeur décide d'augmenter toutes les notes d'un point. Que deviennent la moyenne et l'écart-type de la nouvelle série de notes ?

3) Même question s'il décide plutôt d'augmenter toutes les notes de 10%

Correction

1)

- Moyenne $= (8+12+10+14)/4 = 11$
- Ecart type $= \text{RACINE}((8-11)^2 + (12-11)^2 + (10-11)^2 + (14-11)^2)/4 = 2,23$

2) avec 9, 13, 11 et 15

- Moyenne $= (9+13+11+15)/4 = 12$ en ajoutant 1 aux valeurs, la moyenne est elle aussi augmentée de 1
- Ecart type $= \text{RACINE}((9-12)^2 + (13-12)^2 + (11-12)^2 + (15-12)^2)/4 = 2,23$ en ajoutant 1 aux valeurs, l'écart type ne change pas

3) avec $8 \times 1,1 = 8,8$; $13 \times 1,1 = 14,3$; $11 \times 1,1 = 12,1$ et $15 \times 1,1 = 16,5$

- Moyenne $= (8,8+14,3+12,1+16,5)/4 = 12,925$ en multipliant les valeurs, la moyenne est changée
- Ecart type $= \text{RACINE}((8,8-12,925)^2 + (14,3-12,925)^2 + (12,1-12,925)^2 + (16,5-12,925)^2)/4 = 2,844$ en multipliant les valeurs, l'écart type est changé

U) Le tableau ci-dessous donne le coût moyen et le coût médian (en millions de francs) des films de long métrage d'initiative française de 1986 à 1998 :

	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998
Coût moyen	12,5	14,4	18,3	21,0	21,3	23,7	25,9	22,5	26,1	28,1	24,3	31,3	28,6
Coût médian	10,4	12,7	13,5	15,1	15,7	18,5	19,0	17,5	18,0	20,7	17,3	18,6	17,5

1) Comment expliquer le fait que le coût moyen soit toujours supérieur au coût médian ?

2) Interpréter l'accroissement de la différence entre ces 2 coûts à partir de l'année 1994.

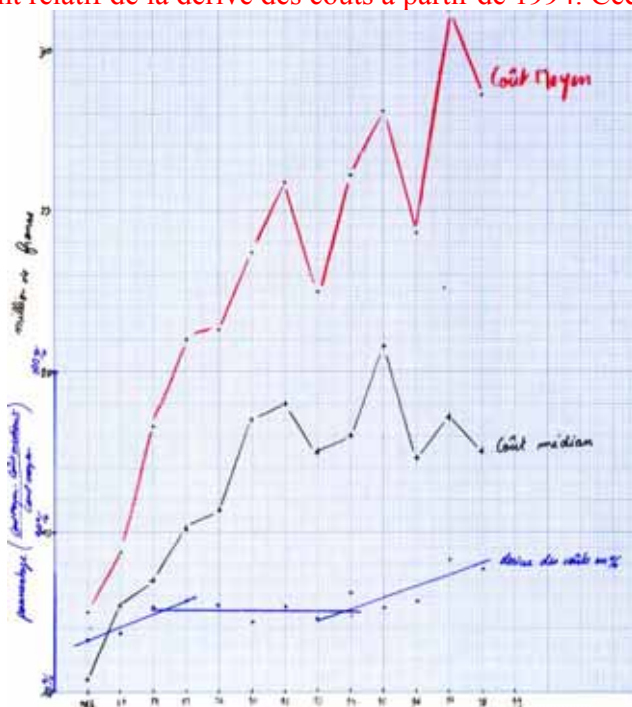
Correction

1)

Les couts moyens sont toujours supérieurs au coût médian car ces séries de couts sont étalées vers la droite

2)

On constate un accroissement relatif de la dérive des couts à partir de 1994. Ceci est un instrument de gestion



V) Étude des températures moyennes au Canada : Le tableau ci-dessous donne les écarts des températures moyennes annuelles par rapport à une température de référence choisie.

Année	1951	1952	1953	1954	1955	1956	1957	1958	1959	1960	1961	1962	1963	1964
Ecart	-0.2	0.9	1	0.2	0.1	-0.5	0	0.8	-0.3	0.6	-0.1	0.1	0.3	-0.5
Année	1965	1966	1967	1968	1969	1970	1971	1972	1973	1974	1975	1976	1977	1978
Ecart	-0.6	-0.2	-0.3	0.2	0.4	-0.2	0	-1.8	0.7	0.8	-0.2	0.1	0.9	-0.3
Année	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992
Ecart	-0.1	0.4	1.9	-0.9	0.2	0.3	0	0.2	1.4	0.8	-0.2	0	0.5	-0.1
Année	1993	1994	1995	1996	1997	1998	1999	2000						
Ecart	0.5	0.4	0.7	0.1	0.4	2.4	1.8	1						

- 1) a) Déterminer la médiane, la moyenne et les quartiles de la série. Comparer médiane et moyenne. Commenter.
b) Représenter la série par un diagramme en boîte. Est-il symétrique ? Proposer une explication.
c) Déterminer les déciles D1 et D9. Représenter la série par un diagramme en boîte limité par ces déciles. Est-il symétrique ? Commenter.
- 2) a) Déterminer la médiane, les quartiles et les déciles D1 et D9 de la série la série des 20 dernières années.
b) Représenter sur le même graphique les diagrammes en boîtes des deux séries (choisir d'utiliser ou non les déciles).

Correction

1) a)

Commençons par ranger les valeurs des écarts par ordre croissant

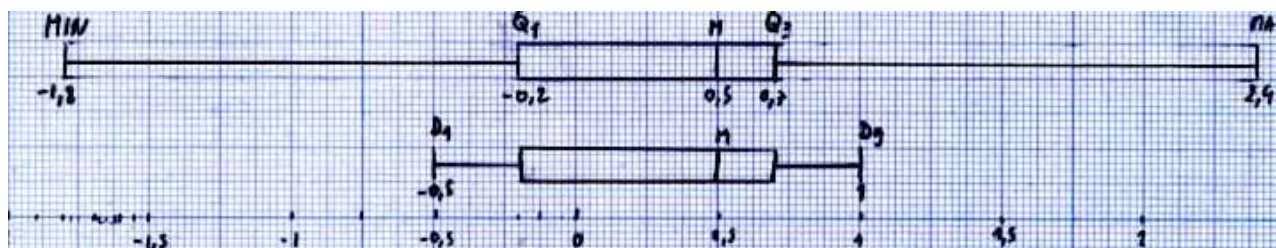
Ecart	1,8	0,9	0,6	0,5	0,3	0,2	0,1	0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1	1,4	1,8	1,9	2,4
effectif	1	1	1	2	3	5	3	4	4	4	2	4	2	1	2	3	2	2	1	1	1	1
rang debut	1	2	3	4	6	9	14	17	21	25	29	31	35	37	38	40	43	45	47	48	49	50
rang de fin	1	2	3	5	8	13	16	20	24	28	30	34	36	37	39	42	44	46	47	48	49	50

Il y a 50 termes

- La médiane est le 25^{ème} terme soit 0,5 °C
- La moyenne est la somme de tous les écarts (13,6) divisé par le nombre de terme (50) soit 0,272 °C
- Cette série n'est pas bien centrée, la moyenne étant plus faible que la médiane, elle est étalée vers la gauche

b)

- 1^{er} quartile $1/4 \times 50 = 12,5$ soit la 13^{ème} valeur $Q1 = -0,2$
- 3^{ème} quartile $3/4 \times 50 = 37,5$ soit la 38^{ème} valeur $Q3 = 0,7$



Le diagramme en boîte n'est pas symétrique, comme pressenti avec la moyenne, les valeurs sont à gauche de la médiane.

c)

- 1^{er} décile $1/10 \times 50 = 5$ soit la 5^{ème} valeur $D1 = -0,5$
- 9^{ème} décile $9/10 \times 50 = 45$ soit la 45^{ème} valeur $D9 = 1$

Le diagramme en boîte ci-dessus, en dessous n'est alors moins dissymétrique car plus resserré.

2) a)

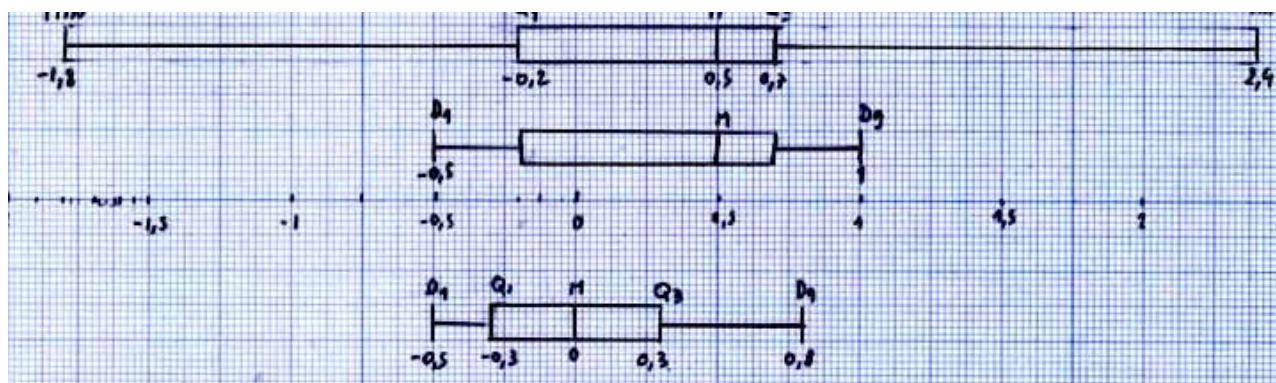
Commençons par ranger les valeurs des écarts par ordre croissant

Ecart	-0,6	-0,5	-0,3	-0,2	-0,1	0	0,1	0,2	0,3	0,4	0,6	0,8	0,9	1
effectif	1	2	2	3	1	1	2	2	1	1	1	1	1	1
rang debut	1	2	4	6	9	10	11	13	15	16	17	18	19	20
rang de fin	1	3	5	8	9	10	12	14	15	16	17	18	19	20

Il y a 20 termes.

- La médiane est le 10^{ème} terme soit 0 °C
- 1^{er} quartile $1/4 * 20 = 5$ soit la 5^{ème} valeur $Q1 = -0,3$
- 3^{ème} quartile $3/4 * 20 = 15$ soit la 15^{ème} valeur $Q3 = 0,3$
- 1^{er} décile $1/10 * 20 = 2$ soit la 2^{ème} valeur $D1 = -0,5$
- 9^{ème} décile $9/10 * 20 = 18$ soit la 18^{ème} valeur $D9 = 0,8$

b)



On voit sur le diagramme du dessous représentant les 20 premières années qu'il a fait plus froid de 1951 à 1970 que de 1951 à 2000, comme si le climat se réchauffait...

W) Une communauté souhaite limiter la longueur des conversations téléphoniques. Elle décide d'envoyer un signal aux 15% des appels les plus longs. On cherche la durée après laquelle un signal doit être envoyé.

Une étude montre que la durée d'appel suit approximativement une loi normale de moyenne $\mu = 8'30''$ et d'écart-type $\sigma = 2'15''$.

1) Déterminer les intervalles de centre μ qui contiennent 99%, 95% et 68% des appels.

2) Après quelle durée de conversation faut-il envoyer le signal ?

Correction

1)

Moyenne μ et à l'écart type σ :

- Environ 99% des données sont dans l'intervalle $[\mu - 3\sigma ; \mu + 3\sigma]$

$$\mu - 3\sigma = 8'30'' - 3*(2'15'') = 8*60+30-3*(2*60+15) = 105'' = 1'45''$$

$$\mu + 3\sigma = 8'30'' + 3*(2'15'') = 8*60+30+3*(2*60+15) = 915 = 15'15''$$

Intervalle $[1'45'' ; 15'15'']$

- Environ 95% des données sont dans l'intervalle $[\mu - 2\sigma ; \mu + 2\sigma]$

$$\mu - 2\sigma = 8'30'' - 2*(2'15'') = 8*60+30-2*(2*60+15) = 240 = 4'$$

$$\mu + 2\sigma = 8'30'' + 2*(2'15'') = 8*60+30+2*(2*60+15) = 780 = 13'$$

Intervalle $[4' ; 13']$

- Environ 68% des données sont dans l'intervalle $[\mu - \sigma ; \mu + \sigma]$

$$\mu - \sigma = 8'30'' - (2'15'') = 8*60+30 - (2*60+15) = 375 = 6'20''$$

$$\mu + \sigma = 8'30'' + (2'15'') = 8*60+30 + (2*60+15) = 645 = 10'45''$$

Intervalle $[6'20'' ; 10'45'']$

2)

15% c'est proche de 16% Il faut donc envoyer le signal au bout de 10'45''

X) Lors de la fabrication d'un lot de fromages de chèvres, on a relevé la masse des fromages fabriqués :

x_i : masse (en g)	[80;85[	[85;90[	[90;95[	[95;100[	[100;105[	[105;110[	[110;115[
n_i : effectifs	5	9	14	18	25	16	7

Dans une production de ce type, tous les fromages ne sont pas commercialisés.

Les fromages dont la masse est comprise dans l'intervalle $[\bar{x} - \sigma ; \bar{x} + \sigma]$ sont commercialisés au prix courant.

Les fromages dont la masse est comprise au-delà de l'intervalle $[\bar{x} - 2\sigma ; \bar{x} + 2\sigma]$ ne sont pas commercialisés.

Les autres fromages sont commercialisés au rabais.

1) Peut-on déterminer le nombre de fromages qui seront commercialisés au prix courant ?

Peut-on donner un encadrement de ce nombre ?

2) Donner un encadrement du nombre de fromages non commercialisés ?

3) Donner un encadrement du nombre de fromages commercialisés au rabais.

Correction

1)

Calculons la moyenne $\bar{x}$ et à l'écart type σ :

- $\bar{x} = (5 \cdot (82,5) + 9 \cdot (87,5) + 14 \cdot (92,5) + 18 \cdot (97,5) + 25 \cdot (102,5) + 16 \cdot (107,5) + 7 \cdot (112,5)) / (5 + 9 + 14 + 18 + 25 + 16 + 7) \approx 99$ g
- $\sigma = \text{RACINE}((5 \cdot (82,5 - 99)^2 + 9 \cdot (87,5 - 99)^2 + 14 \cdot (92,5 - 99)^2 + 18 \cdot (97,5 - 99)^2 + 25 \cdot (102,5 - 99)^2 + 16 \cdot (107,5 - 99)^2 + 7 \cdot (112,5 - 99)^2) / (5 + 9 + 14 + 18 + 25 + 16 + 7)) \approx 8$ g
- L'intervalle $[\bar{x} - \sigma ; \bar{x} + \sigma] = [99 - 8 ; 99 + 8] = [91 ; 104]$ est contenu dans les 3 intervalles $[90;95[$, $[95;100[$ et $[100;105[$ donc le nombre de fromages qui seront commercialisés au prix courant sera inférieur à la somme des effectifs de ces 3 intervalles, soit $14 + 18 + 25 = 40$, et plus grand que l'effectif de l'intervalle du milieu $[95;100[$, soit 18.

Si on note « FCourant » le nombre de fromages qui seront commercialisés au prix courant

$$18 < \text{FCourant} < 40$$

2)

- Les fromages non commercialisés sont en dehors de l'intervalle $[\bar{x} - 2\sigma ; \bar{x} + 2\sigma] = [99 - 16 ; 99 + 16] = [83 ; 115]$. Cet intervalle n'exclut aucun des intervalles de masse, donc il se peut qu'aucun fromage ne soit pas commercialisé. Au maximum, 5 peuvent être non commercialisés.

Si on note « FnonCommercialisés » le nombre de fromages non commercialisés

$$0 < \text{FnonCommercialisés} < 5$$

3)

Si on note « FCommercialisésRabais » le nombre de fromages commercialisés au rabais.

$$\text{FCourant} + \text{FnonCommercialisés} + \text{FCommercialisésRabais} = 94 \text{ fromages au total}$$

$$\text{FCommercialisésRabais} = 94 - \text{FCourant} - \text{FnonCommercialisés}$$

- Envisageons les cas extrêmes :
 - o On aura un minimum de fromages commercialisés au rabais quand il y a un maximum de fromages non commercialisés, soit 5, et un maximum de fromages commercialisés au prix courant, soit 40. Donc un minimum de $94 - 40 - 5 = 45$
 - o On aura un maximum de fromages commercialisés au rabais quand il y a un minimum de fromages non commercialisés, soit 0, et un minimum de fromages commercialisés au prix courant, soit 18. Donc un maximum de $94 - 18 - 0 = 76$

$$45 < \text{FCommercialisésRabais} < 76$$

Y) Analyses sanguines (on admettra que les variables biologiques sont gaussiennes et que les plages de normalité indiquées le sont avec niveau de confiance de 95%)

1) Les triglycérides : Les valeurs normales sont entre 0,5 et 1,5 g/L. Calculer la moyenne μ et l'écart type σ .

2) Le cholestérol : Les valeurs normales sont entre 1,5 et 2,4 g/L. Calculer la moyenne μ et l'écart type σ .

Quel intervalle de centre μ contient 99% des cas ?

3) La glycémie : La valeur moyenne est de 0,97 g/L, l'écart type est de 0,09 g/L.

a) Quelle est la plage de normalité ?

b) Pour quel pourcentage de la population la glycémie est-elle supérieure à 1,15 g/L.

Correction

1) triglycérides

Environ 95% des données sont dans l'intervalle $[\mu - 2\sigma ; \mu + 2\sigma]$ donc $\mu - 2\sigma = 0,5$ g/L et $\mu + 2\sigma = 1,5$ g/L

$$0,5 + 1,5 = \mu - 2\sigma + \mu + 2\sigma = 2\mu$$

$$\mu = 2/2 = 1 \text{ g/L}$$

$$1,5 - 0,5 = \mu + 2\sigma - (\mu - 2\sigma) = 4\sigma$$

$$\sigma = 1/4 = 0,25 \text{ g/L}$$

2) cholestérol

Environ 95% des données sont dans l'intervalle $[\mu - 2\sigma ; \mu + 2\sigma]$ donc $\mu - 2\sigma = 1,5$ g/L et $\mu + 2\sigma = 2,4$ g/L

$$1,5 + 2,4 = \mu - 2\sigma + \mu + 2\sigma = 2\mu$$

$$\mu = 3,9/2 = 1,95 \text{ g/L}$$

$$2,4 - 1,5 = \mu + 2\sigma - (\mu - 2\sigma) = 4\sigma$$

$$\sigma = 0,9/4 = 0,225 \text{ g/L}$$

Environ 99% des données sont dans l'intervalle $[\mu - 3\sigma ; \mu + 3\sigma]$

$$\mu - 3\sigma = 1,95 - 3 \cdot 0,225 = 1,275 \text{ g/L}$$

$$\mu + 3\sigma = 1,95 + 3 \cdot 0,225 = 2,625 \text{ g/L}$$

Donc 99% des données sont dans l'intervalle $[1,275 ; 2,625]$

3) glycémie

a) La plage de normalité c'est l'intervalle contenant 95% des données soit $[\mu - 2\sigma ; \mu + 2\sigma]$

$$\mu - 2\sigma = 0,97 - 2 \cdot 0,09 = 0,79 \text{ g/L}$$

$$\mu + 2\sigma = 0,97 + 2 \cdot 0,09 = 1,15 \text{ g/L}$$

La plage de normalité c'est l'intervalle $[0,79 ; 1,15]$

b) 5% de la population ne sont pas normaux c'est à dire avec une glycémie inférieure à 0,79 g/L ou supérieure à 1,15 g/L. Ceci selon une distribution symétrique par rapport à la moyenne, donc la moitié de 5%, soit 2 % de la population, a une glycémie supérieure à 1,15 g/L.

Z) 1) Une association de consommateurs contrôle le poids d'un lot de barres chocolatées. Par contrat, le fabricant s'engage à un poids moyen de 50 g, avec 97,5% des poids supérieurs à 48g : Quel doit être l'écart type ?

2) Une barre choisie au hasard pèse 49g. Peut-on accuser le fabricant de tricherie ?

3) Un échantillon de 100 barres à un poids de 4,970 kg. On sait que si le fabricant dit vrai, les poids des échantillons ont une moyenne de 5 kg et un écart type de 10 g. Que peut-on dire ?

Correction

1)

La distribution normale est symétrique par rapport à la moyenne de 50 g donc 97,5% des poids supérieurs à 48g implique 97,5% des poids inférieurs à 52 g. Autrement dit 95% des données sont dans l'intervalle $[48 ; 52]$ qui correspond à $[\mu - 2\sigma ; \mu + 2\sigma]$.

$$\mu = 50$$

$$50 - 2\sigma = 48$$

$$\sigma = 1 \text{ g} = \text{écart type}$$

2)

Non un échantillon n'est pas assez pour être représentatif.

3)

En supposant que la distribution est normale avec une moyenne de 5 kg et un écart type de 10 g on doit avoir 95% des données soit $[\mu - 2\sigma ; \mu + 2\sigma] = [5000 - 2 \cdot 10 ; 5000 + 2 \cdot 10] = [4980 ; 5020]$

Or un nombre de 100 échantillons peut être considéré comme représentatif et l'on mesure 4970 g donc en dehors des 95% des données. On peut donc raisonnablement penser que le fabriquant triche sur sa marchandise.