Statistiques

Ch. n°2 ; 1 L ; Année scolaire 2007/2008

CHAPITRE 2 : PAGE 69 - 94 ; Le Lundi 7 Janvier 2008

PARAMÈTRES DE POSITION : MOYENNE , PARAMÈTRES DE DISPERSION : ECART-TYPE

La moyenne de la série statistique donnée par le tableau ci-dessous est définie par :

Valeur prise par le caractère	x_1	x_2	...	x_p	Total
Effectif	n_1	n_2	...	n_p	N

est définie par :

$$\bar{x}=\frac{n_1x_1+n_2x_2+...+n_px_p}{n_1+n_2+...+n_p}=\frac{n_1x_1+n_2x_2+...+n_px_p}{N}$$

où $\bar{x}$ est la moyenne de la série statistique
et N la taille de l'échantillon.

La variance de la série statistique donnée par le tableau ci-dessus est définie par :

$$V=\frac{n_1x_1^2+n_2x_2^2+...+n_px_p^2}{n_1+n_2+...+n_p}-\left(\bar{x}\right)^2$$

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

où V est la variance de la série statistique.

L'écart-type est noté σ : $\sigma=\sqrt{V}$.

EXERCICE N°1 :

	Sep	Oct	Nov	Dec	Jan	Fev	Mar	Avr	Mai	Jun	Total	Moyenne
Arnaud	7	8	12	15	15	15	14	12	14	8		
Bertrand	12	11	12	12	13	12	12	13	11	12		
Céline	6	0	14	18	20	14	16	0	18	14		
Mickael	8	9	16	11	13	14	15	13	14	7		

Les quatres moyennes sont égales. Cependant, la répartition des notes n'est vraiment pas la même ! Pour Bertrand les notes sont regroupées autour de la moyenne ; pour Céline elle sont dispersées. Pour traduire cette différence de comportement on peut calculer l'étendue de la série statistique , mais on va utiliser principalement : l'intervalle interquartile et l'écart-type.

① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨ ⑩

PARAMÈTRES DE POSITION : MÉDIANE

La médiane M sépare une série statistique en deux sous-séries de même effectif, l'une contenant les plus petites valeurs, l'autre les plus grandes.

Pour une série statistique (x_1 ; x_2 ; x_3 ; ... : x_n) le nombre M est obtenu de la façon suivante :

① – On range d'abord les valeurs prises par le caractère par ordre croissant :

la série (x_1 ; x_2 ; x_3 ; ... : x_n) devient la suite (a_1 , a_2 , a_3 , ... : a_n) chacune des valeurs figure un nombre de fois égal à son effectif, ainsi les termes de la suite vérifient $a_1 \leq a_2 \leq a_3 \leq \ ... \leq a_n$;

② – deux cas se présentent suivant que n est pair ou impair :

☞ si **n est impair** (ici 7 termes) , M est le nombre de cette suite situé "au milieu" c'est à dire entre la sous-série (a_1 , a_2 , a_3) et la sous-série (a_5 , a_6 , a_7) ; donc la médiane M est = a_4

a_1 a_2 a_3 a_4 a_5 a_6 a_7

☞ si **n est pair** , (ici 6 termes) , M est le nombre de cette suite situé "au milieu" c'est à dire entre la sous-série (a_1 , a_2 , a_3) et la sous-série (a_4 , a_5 , a_6) ;
Donc la médiane M est calculée au centre de l'intervalle [a_3 ; a_4] c'est à dire [$a_{n/2}$; $a_{1+n/2}$] (intervalle médian) ;

M = milieu de l'intervalle [a_3 ; a_4]

a_1 a_2 a_3 **M** a_4 a_5 a_6

Statistiques

EXERCICE N°2 :

Le comparatif des différents prix de vente, en euros, d'un article donne le résultat suivant, calculer dans chacun des cas la médiane

① – Exemple n°1 :

Prix de l'article en euros	50	45	30	60	70	40
Effectif correspondant	3	2	2	4	4	2

② – Exemple n°2 :

Prix de l'article en euros	50	45	30	60	70
Effectif correspondant	7	7	1	2	3

On commence par ranger d'abord les valeurs prises par le caractère par ordre croissant, chacune figurant un nombre de fois égal à son effectif.

PARAMÈTRES DE DISPERSION : QUARTILES

Une médiane M sépare une série statistique en deux sous-séries de même effectif, l'une contenant les plus petites valeurs, l'autre les plus grandes.

Les quartiles sont les médianes de ces sous-séries. Le premier quartile noté Q_1 est la médiane de la sous-série inférieure , le troisième noté Q_3 est la médiane de la sous-série supérieure

☞ **Le nombre Q_3 - Q_1 est appelé écart interquartile.**
☞ **L'intervalle] Q_3 - Q_1 [est appelé intervalle interquartile.**

REMARQUES :

REMARQUE N°1 : *Les nombres Q_1 , M , Q_3 permettent de partager la série en quatres sous-séries contenant le même effectif : donc 25% de l'effectif pour chaque sous-série.*

REMARQUE N°2 : **Notion de déciles** : *ces nombres notés Q_1 , Q_2 , ... , M , ... , Q_8 , Q_9 permettent de partager la série en dix sous-séries contenant le même effectif : donc 10 % de l'effectif pour chaque sous-série.*

EXERCICE N°2 :

Le comparatif des différents prix de vente, en euros, d'un article donne le résultat suivant, calculer dans chacun des cas le premier et le troisième quartile :

① – Exemple n°1 :

Prix de l'article en euros	50	45	30	60	70	40
Effectif correspondant	3	2	2	3	4	2

② – Exemple n°2 :

Prix de l'article en euros	50	45	30	60	70
Effectif correspondant	7	7	1	2	3

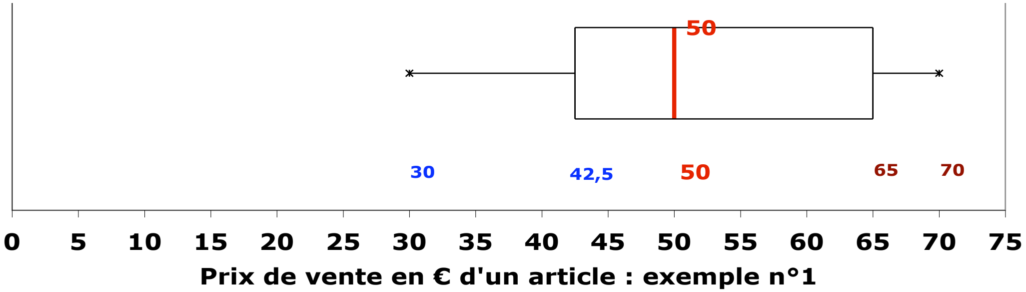
On commence par ranger d'abord les valeurs prises par le caractère par ordre croissant, chacune figurant un nombre de fois égal à son effectif.

PARAMÈTRES DE DISPERSION : DIAGRAMME EN BOÎTES

On peut représenter graphiquement les valeurs Q_1 , M , Q_3 .

Ce graphique peut se construire aussi verticalement.

Les valeurs Q_1 , Q_3 correspondent aux côtés verticaux délimitant la boîte , M au côté vertical intérieur à la boîte.



EXERCICE N°2 :

Dessiner le diagramme en boites pour l'exemple n°2.

Statistiques

DONNÉES GAUSSIENNES :

Voici deux graphiques.

Les centres des côtés supérieurs des rectangles sont sur une courbeen “cloche”.

Il en est ainsi dans de nombreuses situations : données biologiques, données industrielles ...

De telles courbes sont appelées : **COURBES DE GAUSS.**

DE TELLES SÉRIES ONT LES PROPRIÉTÉS SUIVANTES :

☞ Leur répartition est à peu près symétrique autour de la moyenne μ.
☞ Environ 95% des données sont dans l'intervalle $[\ \mu - 2\sigma\ ;\ \mu + 2\sigma\ ]$; où σ désigne l'écart-type.
☞ Environ 99% des données sont dans l'intervalle $[\ \mu - 3\sigma\ ;\ \mu + 3\sigma\ ]$.

☞ L'intervalle $[\ \mu - 2\sigma\ ;\ \mu + 2\sigma\ ]$ est appelé plage de normalité à 95 %.
☞ L'intervalle $[\ \mu - 3\sigma\ ;\ \mu + 3\sigma\ ]$ est appelé plage de normalité à 99 %.

NOTION DE PRÉVISION PROBABILISTE :

☞ On peut interpréter la plage de normalité à 95 % de la manière suivante : si on observe une nouvelle donnée x , elle a 95 chances sur 100 d'être dans l'intervalle $[\ \mu - 2\sigma\ ;\ \mu + 2\sigma\ ]$.

EXEMPLE :

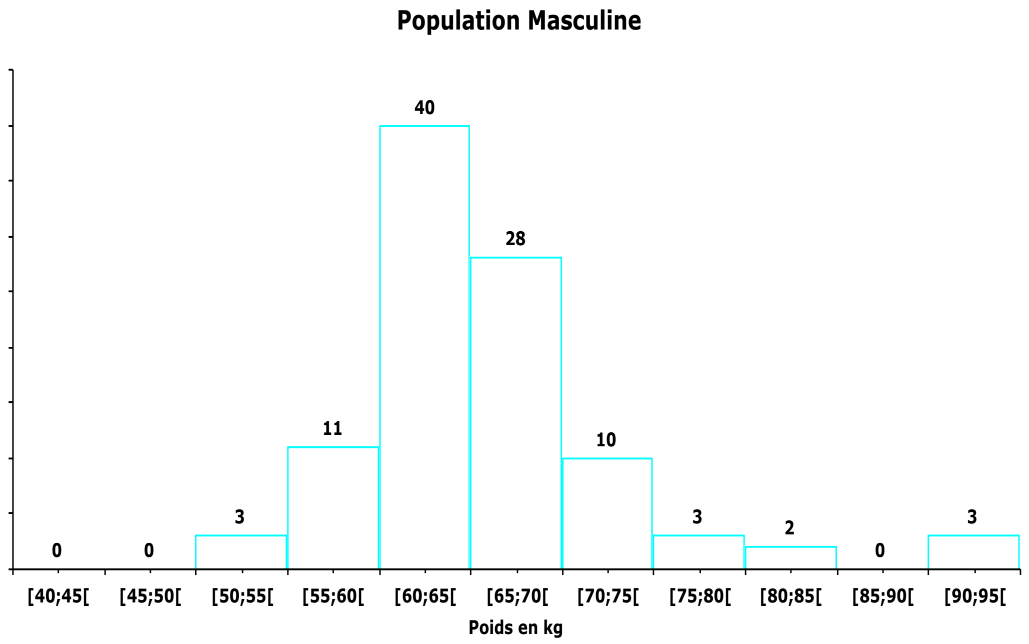
Une série de données gaussiennes admet comme plage de normalité à 95 % de l'intervalle [9 ; 12]. On peut affirmer μ que sa moyenne est égale à 10,5 (milieu de l'intervalle) et que l'écart-type σ est tel que $2\sigma = 1,5$; donc $\sigma = 0,75$.

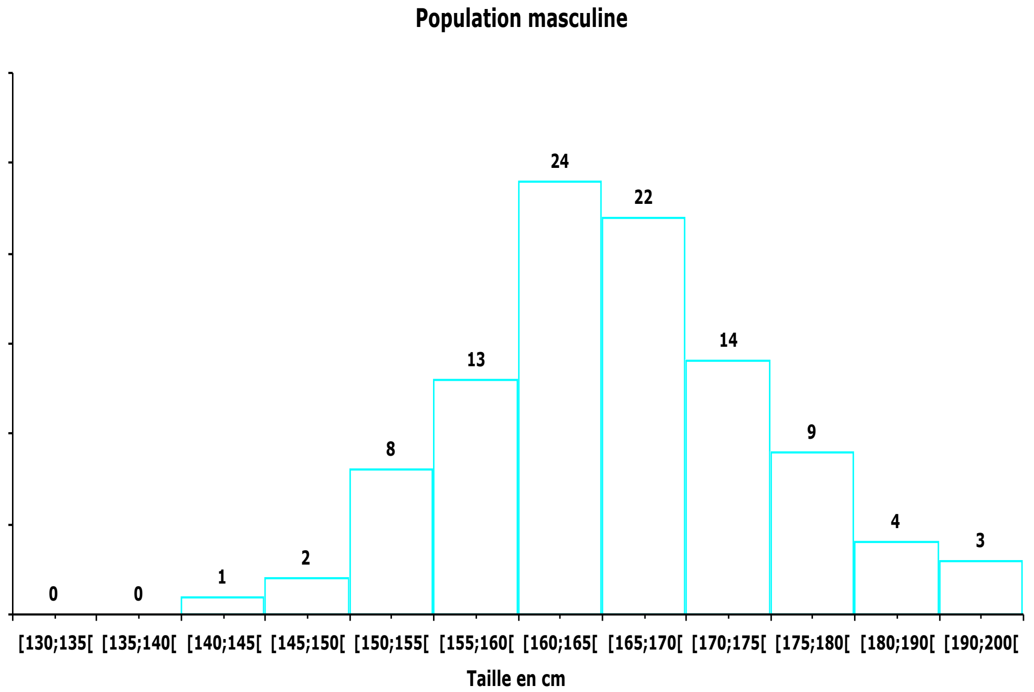
EXERCICE N°3 :

La plage de normalité à 95 % d'une série de données gaussiennes est l'intervalle [1,72 ; 1,78]. Quelle est sa moyenne ? Quelle est son écart-type ?

EXERCICE N°4 :

La plage de normalité à 99 % d'une série de données gaussiennes est l'intervalle [0,97 ; 1,03]. Quelle est sa moyenne ? Quelle est son écart-type ?





EXERCICE N°5 :

Une étude portant sur la taille et le le poids des élèves d'une classe de collège données résultats suivants ;

Population masculine :
Taille en cm : $Q_0 = 144$; $Q_1 = 160$; $Q_2 = 165$; $Q_3 = 170$; $Q_4 = 198$;
Population féminine :
Taille en cm : $Q_0 = 132$; $Q_1 = 145$; $Q_2 = 150$; $Q_3 = 155$; $Q_4 = 168$;

☞ Calculer l'écart interquartile = $Q_3 - Q_1$ pour chacune des séries statistiques;

☞ Représenter les diagrammes en boites sur un même graphique : vous choisirez vous meme l'échelle permettant de présenter une graduation comprise entre 130 et 200 ;

Population masculine :
Poids en kg : $Q_0 = 53$; $Q_1 = 61$; $Q_2 = 64$; $Q_3 = 67{,}25$; $Q_4 = 93$;
Population féminine :
Poids en kg : $Q_0 = 44$; $Q_1 = 50$; $Q_2 = 53{,}5$; $Q_3 = 56{,}25$; $Q_4 = 68$;

☞ Calculer l'écart interquartile = $Q_3 - Q_1$ pour chacune des séries statistiques;

☞ Représenter les diagrammes en boites sur un même graphique : vous choisirez vous même l'échelle permettant de présenter une graduation comprise entre 40 et 95 ;

EXERCICE N°6 :

Durée d'un match de basket-ball :

Les relevés des durées en minutes d'un match de basket-ball ont été disposés dans un tableau :

138	142	113	126	135
142	159	157	140	157
121	128	142	164	155
139	143	158	140	118
142	146	123	130	137

☞ Calculer les valeurs des quartiles et l'écart interquartile = $Q_3 - Q_1$; Représenter le diagramme en boite ,

EXERCICE N°7 :

Nombre de voitures volées journalièrement dans une ville durant une période de 12 jours :

6	3	7	11	5
3	8	7	2	6
9	13			

☞ Calculer les valeurs des quartiles et l'écart interquartile = $Q_3 - Q_1$; Représenter le diagramme en boite ,

EXERCICE N°8 :

On dispose dans un tableau le nombre de serveurs mis au point dans une société de service informatique :

24	32	27	23	33	33	29	25	23	28
21	26	31	20	27	33	27	23	28	29
31	35	34	22	26	28	23	35	31	27

☞ Calculer les valeurs des quartiles et l'écart interquartile = $Q_3 - Q_1$; Représenter le diagramme en boite ,

EXERCICE N°9 :

Voici le nombre de voitures neuves vendus par un concessionnaire durant une période de 20 jours :

8	5	12	3	9	10	6	3	8	8	
4	6	10	11	7	7	3	5	9	11	

☞ Calculer les valeurs des quartiles et l'écart interquartile = $Q_3 - Q_1$; Représenter le diagramme en boite ,