Première 1 L - Année Scolaire 2007-2008
Chapitre n°2 : Statistiques page 70 - 94
Programme d'étude :

<u>Avant-Propos:</u>
Ce chapitre résume les savoirs de base dans le domaine des statistiques : population, caractère étudié, caractère quantitatif, moyenne, écart-type, étendue d'une série, médiane, quartiles, déciles, répartition par classes, effectifs, fréquences, histogrammes, diagramme en bâtons, nuages de points .

Contenu :

Les techniques de base sont :

Etre capable d'identifier la série statistique : reconnaître la population par ses individus, et reconnaître le caractère étudié sur cette population ; conseil : ne pas confondre les valeurs prises par le caractère et l'effectif correspondant à la valeur prise par le caractère ; cette série peut être présentée différemment : tableau de valeurs , nuage de points, histogrammes.

Etre capable de repérer les valeurs correspondant à la médiane, aux quartiles d'une série statistique ; conseil : ranger les valeurs prises par le caractère par ordre croissant.

Etre capable de calculer la moyenne et l'écart-type d'une série statistique ;

Dans le livre le cours est bien présenté de la page 72 à la page 77.

Progression :

Leçon n°1 et TP salle informatique : ex n°1 ; moyenne et écart-type ;

Leçon n°2 : page 72 - 76 du livre ; moyenne et écart-type ; médiane, quartile, diagramme en boîtes y compris exercices du livre

Exercice n° 1 & 2 : Médiane, quartile, diagramme en boîtes ;

Leçon n°3 : page 77 ; données gaussiennes ;

Exercices de la feuille de travaux dirigés.

L'essentiel du cours, page 72 - 77 :

;

Les exercices d'entraînement :

D'après la feuille de travaux dirigés ; autres à venir d'après sujets de Baccalauréat

Devoir maison :

Exclusion du cours :

Fait à Nantes le dimanche 30 décembre 2007 21:38:30

Statistique

1 Paramètres de dispersion : introduction

• **Un exemple**

Dans deux groupes, l'un de dix élèves et l'autre de huit élèves, les notes obtenues à un même devoir de Mathématiques sont les suivantes :

Premier groupe : 1 ; 1 ; 1 ; 2 ; 3 ; 17 ; 20 ; 20 ; 20 ; 20 .

Deuxième groupe : 8 ; 10 ; 10 ; 11 ; 11 ; 11 ; 11 ; 12 .

Nous allons présenter ces renseignements dans deux tableaux d'effectifs :

a) Complétez les deux tableaux d'effectifs suivants.

Première série

note x_i	1	2	3	17	20
effectif n_i	3				

Deuxième série

note x_i	8	10	11	12
effectif n_i	1			

b) Calculez la moyenne des notes obtenues dans chacun des deux groupes.

. . . .

c) Calculez l'étendue de chacune de ces deux séries de notes, c'est-à-dire la différence entre la plus grande valeur et la plus petite valeur.

. . . .

COMMENTAIRE : Les deux moyennes sont donc égales. Cependant, la répartition des notes n'est vraiment pas la même ! Pour le deuxième groupe, les notes sont regroupées autour de la moyenne. Dans le premier groupe, au contraire, elles sont très dispersées. Pour traduire cette différence de comportement, on utilise usuellement en statistique des paramètres de dispersion tels que l'étendue, vue en Seconde. Nous allons introduire cette année deux autres paramètres de dispersion : l'intervalle interquartile et l'écart-type.

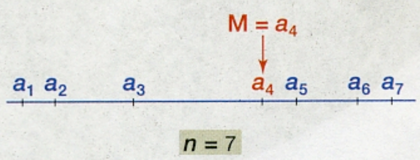
2 Médiane

Définition 1

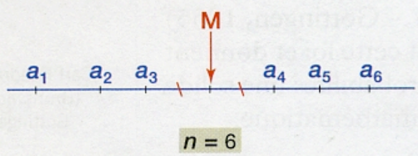
On appelle **médiane** de la série statistique $(x_1, x_2, \ldots, x_n)$ le nombre M obtenu de la façon suivante :

• on range d'abord les valeurs du caractère par ordre croissant, chacune figurant un nombre de fois égal à son effectif. On obtient la suite $a_1 \leqslant a_2 \leqslant \ldots \leqslant a_n$;

• si n **est impair**, M est le nombre a_i de cette suite situé « au milieu », c'est-à-dire tel que $i = \dfrac{n+1}{2}$;



• si n **est pair**, M est le nombre situé au centre de l'intervalle $\left[a_{\frac{n}{2}} ; a_{\frac{n}{2}+1}\right]$ (intervalle médian).



$[a_3 ; a_4]$ est l'intervalle médian.

Exemples :

1.

valeur du caractère	50	45	30	60	61
effectif	2	3	2	2	1

Commençons par ranger les valeurs du caractère par ordre croissant, **chacune figurant un nombre de fois égal à son effectif** : 30 ; 30 ; 45 ; 45 ; 45 ; 50 ; 50 ; 60 ; 60 ; 61 .

Ici n est pair ($n = 10$) ; donc la médiane est le milieu de l'intervalle médian, c'est-à-dire l'intervalle délimité par la 5^e^ et la 6^e^ valeur, c'est-à-dire [45 ; 50] donc $M = 47{,}5$.

2.

valeur du caractère	50	45	30	60	61
effectif	2	3	2	2	2

Commençons par ranger les valeurs du caractère par ordre croissant, **chacune figurant un nombre de fois égal à son effectif** : 30 ; 30 ; 45 ; 45 ; 45 ; 50 ; 50 ; 60 ; 60 ; 61 ; 61 .

Ici n est impair ($n = 11$) ; donc la médiane est le 6^e^ terme $(i = \dfrac{11+1}{2})$; donc $M = 50$.

3 Quartiles

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.

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 quartile noté Q_3 est la médiane de la sous-série supérieure.

Note : La médiane est parfois appelée deuxième quartile.

Définition 2

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

Remarque 1. Les nombres Q_1 , M et Q_3 permettent de couper la population étudiée en quatre groupes contenant chacun le même nombre d'éléments.

Remarque 2. Notion de déciles. On peut définir de manière analogue la notion de déciles : ce sont des nombres qui permettent de couper la population étudiée en dix groupes contenant chacun le même nombre d'éléments.
Le premier décile et le neuvième décile sont les plus fréquemment utilisés.

Exemples :

1.

valeur du caractère	50	45	30	60	61
effectif	2	3	2	2	2

• Commençons par ranger les valeurs du caractère par ordre croissant, **chacune figurant un nombre de fois égal à son effectif** : 30 ; 30 ; 45 ; 45 ; 45 ; 50 ; 50 ; 60 ; 60 ; 61 ; 61 . Ici n est impair ($n = 11$), donc la médiane M est la 6^e^ valeur, c'est-à-dire 50 . Donc $M = 50$.

• On a donc $\underbrace{30 ;\ 30 ;\ 45 ;\ 45 ;\ 45}_{\text{série inférieure}} ;\ \underset{\substack{\uparrow \\ M}}{50} ,\ \underbrace{50 ;\ 60 ;\ 60 ;\ 61 ;\ 61}_{\text{série supérieure}}$.

• Q_1 est la médiane de la série inférieure (30 ; 30 ; 45 ; 45 ; 45), donc $Q_1 = 45$.

• Q_3 est la médiane de la série supérieure (50 ; 60 ; 60 ; 61 ; 61), donc $Q_3 = 60$.

2.

valeur du caractère	50	45	30	60	70
effectif	2	5	3	2	2

• Commençons par ranger les valeurs du caractère par ordre croissant, **chacune figurant un nombre de fois égal à son effectif** .

Ici n est pair ($n = 14$), donc M est le centre de l'intervalle d'extrémités la 7ᵉ valeur et la 8ᵉ valeur, donc **M = 45** .

$$\underbrace{30 ;\ 30 ;\ 30 ;\ 45 ;\ 45 ;\ 45 ;\ 45}_{\text{série inférieure}} \underset{\substack{\uparrow \\ M}}{;}\ \underbrace{45 ;\ 50 ;\ 50 ;\ 60 ;\ 60 ;\ 70 ;\ 70}_{\text{série supérieure}} .$$

• Q_1 est la médiane de la série inférieure (30 ; 30 ; 30 ; 45 ; 45 ; 45 ; 45), donc $\mathbf{Q_1 = 45}$.

• Q_3 est la médiane de la série supérieure (45 ; 50 ; 50 ; 60 ; 60 ; 70 ; 70), donc $\mathbf{Q_3 = 60}$.

Remarque : Dans le cas où n est impair (voir exemple 1), la médiane n'appartient ni à la série inférieure, ni à la série supérieure.

APPLICATIONS

Pour chacune des séries statistiques suivantes, trouvez la médiane et les deux quartiles Q_1 et Q_3 .

Exo 1

valeur du caractère	10	15	12	14	20
effectif	2	5	7	3	3

....

....

....

....

Exo 2

valeur du caractère	10	14	12	20	15
effectif	2	3	4	5	3

....

....

....

....

5 Variance – Écart-type

Définition 4

La **variance V** de la série statistique donnée par le tableau ci-contre est définie par :

valeur	x_1	x_2	…	x_p	TOTAL
effectif	n_1	n_2	…	n_p	N

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

où $\overline{x}$ désigne la moyenne de cette série.
L'**écart-type σ** est la racine carrée de la variance : $\sigma = \sqrt{V}$.

Note : On dit que la variance et l'écart type sont des paramètres de dispersion (voir le paragraphe 1, page 72).

Exemple : Considérons les deux séries de notes suivantes, présentées en **1** .

Première série

note x_i	1	2	3	17	20
effectif n_i	3	1	1	1	4

Deuxième série

note x_i	8	10	11	12
effectif n_i	1	2	4	1

Chacune de ces séries a pour moyenne 10,5 .

Notons V_1 et V_2 les variances respectives de chacune de ces séries et σ_1 , σ_2 leurs écarts-types.

$$V_1 = \frac{(1-10{,}5)^2 \times 3 + (2-10{,}5)^2 \times 1 + (3-10{,}5)^2 \times 1 + (17-10{,}5)^2 \times 1 + (20-10{,}5)^2 \times 4}{10} = 80{,}25$$

$$V_2 = \frac{(8-10{,}5)^2 \times 1 + (10-10{,}5)^2 \times 2 + (11-10{,}5)^2 \times 4 + (12-10{,}5)^2 \times 1}{10} = 1{,}25$$

donc $\sigma_1 \approx 8{,}96$ et $\sigma_2 \approx 1{,}12$.

Note : L'écart-type de la première série est nettement supérieur à celui de la seconde. Ce résultat était prévisible *a priori*, compte tenu de la plus grande dispersion des notes de la première série.

APPLICATION

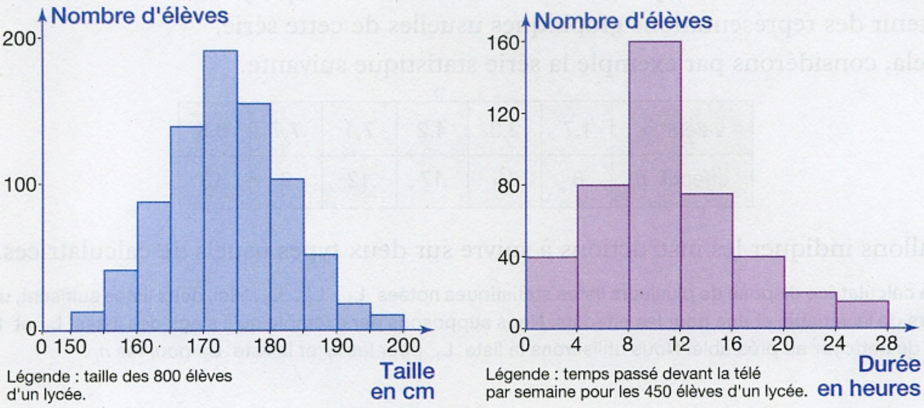
Exo 4 Trouvez la moyenne, la variance et l'écart-type de la série statistique suivante :

x_i	7	13	18
n_i	5	3	2

. . . .
. . . .
. . . .

6 Données gaussiennes

Voici deux graphiques.



Les données sont de natures différentes. Cependant, les graphiques ont la même allure : les centres des côtés supérieurs des rectangles sont sur des courbes en « cloche ».

Note : Il en est ainsi dans de nombreuses situations : données biologiques, données industrielles… De telles courbes sont appelées **courbes de Gauss**.

Propriétés

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 $[\mu - 2\sigma\ ,\ \mu + 2\sigma]$, σ désignant l'écart-type.
- Environ 99 % des données sont dans l'intervalle $[\mu - 3\sigma\ ,\ \mu + 3\sigma]$.

Définition 5

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 %** .

Note : 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 % l'intervalle $[9\ ;\ 12]$. On peut alors affirmer que sa moyenne μ est égale à 10,5 (milieu de l'intervalle) et que l'écart-type est tel que $2\sigma = 1{,}5$; donc $\sigma = 0{,}75$.

APPLICATIONS

Exo 5 La plage de normalité à 95 % d'une série de données gaussiennes est [1,72 ; 1,78].
Quelle est sa moyenne ? Quel est son écart-type ?

. . . .

Exo 6 La plage de normalité à 99 % d'une série de données gaussiennes est [0,97 ; 1,03].
Quelle est sa moyenne ? Quel est son écart-type ?

. . . .

Première 1 L - Année Scolaire 2007-2008
Chapitre n°2 : Statistiques page 70 - 94
Programme d'étude :