

*Fiches à coller dans le cahier, à lire dans le détail au fur et à mesure du cours.
Les exercices et les ajouts du professeur au tableau sont à reporter avec soin dans le cahier avec une référence systématique aux paragraphes concernés.*

STATISTIQUE

Table des matières

I) Représentation de séries statistiques	1
II) Rappels de seconde sur les séries statistiques	1
1. Médiane et étendue	2
2. Moyenne	2
3. Quartiles	2
III) Nouveaux outils sur une série statistique	3
1. Écart interquartile	3
2. Diagramme en boîte	3
3. Variance et écart-type	3
4. Classement des paramètres d'une série statistique	4
IV) Effet de structure	4
V) Utilisation de la calculatrice pour une série statistique	4

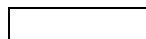


I) Représentation de séries statistiques

partir d'un tableau d'effectifs à double entrée donnant la répartition d'une population suivant deux caractères compatibles, on peut établir trois autres tableaux :

À

- le tableau des fréquences, chaque effectif y est divisé par l'effectif total ;
- le tableau des fréquences en lignes, chaque effectif y est divisé par l'effectif total de sa ligne ;
- le tableau des fréquences en colonnes, chaque effectif y est divisé par l'effectif total de sa colonne.



II) Rappels de seconde sur les séries statistiques

Données chiffrées pour les exemples : on considère la série statistique suivante sur la taille en millimètres d'une centaine d'escargots.

Caractères	10	20	15	5	25	30
Effectifs	12	15	8	20	30	15

1. Médiane et étendue

- La médiane est une valeur qui découpe la liste ordonnée en deux listes ayant pratiquement le même effectif.
- L'étendue d'une série statistique est la différence entre la plus grande valeur et la plus petite.

Exercice

Calculer ces paramètres sur la colonie d'escargots.

(Se reporter à la dernière section de ce chapitre pour l'utilisation de la calculatrice).

2. Moyenne

Une **série statistique** est un objet issu de l'étude d'un certain **population** : cette population est donc composée d'**individus**, individus dont on va étudier un certain **caractère** qui apparaîtra selon une certaine fréquence.

On note en général (x_i, n_i) une série statistique, x_i étant le caractère étudié et n_i l'**effectif** associé.

Notons $N = n_1 + n_2 + \dots + n_p$ (on peut aussi écrire $N = \sum_{i=1}^p n_i$) l'effectif de la population totale.

On rappelle les notions et les notations suivantes :

- La fréquence f_i du caractère x_i se déduit de son effectif n_i à l'aide de l'effectif N de la population totale :

$$f_i = \frac{n_i}{N}.$$

- La moyenne s'obtient par la formule

$$\bar{x} = f_1x_1 + f_2x_2 + \dots + f_px_p = \sum_{i=1}^p f_ix_i,$$

ou encore

$$\bar{x} = \frac{n_1x_1 + n_2x_2 + \dots + n_px_p}{N} = \frac{1}{N} \sum_{i=1}^p n_ix_i.$$

Exercice

Établir le tableau des fréquences de la colonie d'escargots et calculer la taille moyenne d'un escargot de la colonie.

3. Quartiles

Les quartiles Q_1 , Me et Q_3 d'une série statistique sont trois valeurs de la série qui découpent la liste ordonnée en quatre listes ayant sensiblement le même effectif.

Ci-dessous une définition plus formelle.

Définition

Soit une série statistique **ordonnée** contenant N valeurs : $x_1 \leq x_2 \leq x_3 \leq \dots \leq x_N$.

Le **premier quartile** d'une série statistique est la plus petite valeur x_i de la série, supérieure ou égale à au moins 25% des valeurs de la série. On note Q_1 cette valeur.

Le **troisième quartile** d'une série statistique est la plus petite valeur x_i de la série, supérieure ou égale à au moins 75% des valeurs de la série. On note Q_3 cette valeur.

(En pratique, Q_1 est la valeur de la série dont le rang est le plus petit entier supérieur ou égal à $\frac{N}{4}$, et Q_3 la valeur dont le rang est le plus petit entier supérieur ou égal à $\frac{3N}{4}$).

L'idée fondamentale à retenir est que pour déterminer ces caractéristiques de dispersion, **les caractères x_i doivent être classés dans l'ordre croissant**. On peut alors parler du premier, du deuxième, etc.

Exercices

III) Nouveaux outils sur une série statistique

1. Écart interquartile

Définition

Soit une série statistique **ordonnée** contenant N valeurs : $x_1 \leq x_2 \leq x_3 \leq \dots \leq x_N$.

On appelle **intervalle interquartile** l'intervalle $[Q_1; Q_3]$ et **écart interquartile** le nombre $Q_3 - Q_1$.

Plus l'écart interquartile est réduit, moins les valeurs de la série sont dispersées : on dit que l'écart interquartile est une caractéristique de dispersion.

L'intervalle interquartile contient au moins 50% des valeurs de la série.

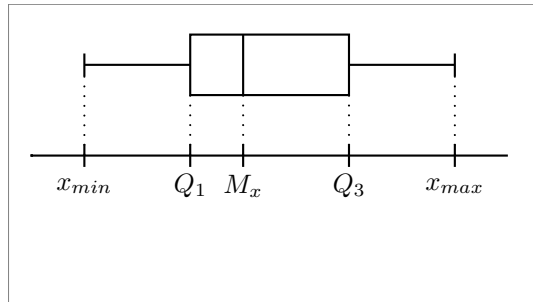
Exercice

Calculer ces paramètres sur la colonie d'escargots.

2. Diagramme en boîte

Définition

On appelle **diagramme de Tukey** ou **diagramme en boîte** ou **boîte à moustaches** ou encore **boîte à pattes** un diagramme comme ci-dessous sur lequel sont indiquées les caractéristiques suivantes : minimum, premier quartile, médiane, troisième quartile et maximum.



La boîte à moustaches, une traduction de « Box & Whiskers Plot », fut inventée en 1977 par le statisticien américain John Wilder Tukey (1915-2000) pour représenter schématiquement une distribution ou d'en comparer plusieurs entre elles.

À la calculatrice :

3. Variance et écart-type

Nous avons vu l'écart interquartile qui est une mesure de la dispersion autour de la médiane, l'écart-type est une autre mesure de la dispersion, elle autour de la moyenne.

Sa formule, compliquée, est donnée par la définition suivante. Nous ne l'utiliserons pas car la calculatrice et le tableur calculent automatiquement ces valeurs à partir d'une la série statistique.

Définition

Variance et écart-type

En notant \bar{x} la moyenne d'une série statistique (x_i, n_i) , la variance V est définie par

$$V = \frac{1}{N} \sum_{i=1}^p n_i \times (x_i - \bar{x})^2 \quad \text{ou} \quad V = \frac{1}{N} \sum_{i=1}^p n_i x_i^2 - \bar{x}^2$$

Son écart-type est alors la racine carrée de la variance : $s = \sqrt{V}$.

4. Classement des paramètres d'une série statistique

Le choix d'un résumé d'une série statistique par ses paramètres n'est pas des compétences du mathématicien, ce sont celles des statisticiens, des économistes. . . suivant ce qu'ils veulent en faire. En tous cas, une étude statistique est accompagnée de commentaires qui justifient la méthode employée et les choix faits. On peut cependant indiquer les résumés possibles suivants.

- Le couple (médiane ; étendue).
- Le couple (moyenne ; étendue).
Ces deux couples sont simples à obtenir mais ils ne permettent pas de positionner le maximum et le minimum de la série. De plus l'étendue est un caractère de dispersion très grossier car sensible aux valeurs extrêmes.
- Le couple (médiane ; intervalle interquartile).
Il est insensible aux valeurs extrêmes.
- L'ensemble {minimum, premier quartile, médiane, troisième quartile, maximum}.
Il permet de construire un diagramme en boîte et donc de mieux visualiser le comportement d'une série (notamment sa dispersion) et de comparer des séries. Il présente un inconvénient : la connaissance de ces paramètres pour deux séries ne permet pas de calculer les paramètres du regroupement des deux séries.
- Le couple (moyenne, écart-type).
Ce couple permet de faire des calculs sur le regroupement de plusieurs séries statistiques.

IV) Effet de structure

Exercice

voir le document « 1stmg_stat_structure.pdf »

V) Utilisation de la calculatrice pour une série statistique

Voici le détail des manipulations à effectuer pour obtenir les paramètres statistiques (et le diagramme en boîte même si celle-ci y est peu exploitable) d'une série statistique à une variable x ; chaque x_i ayant un effectif n_i .

Pour les « casio »**Pour les « TI »**

Entrée de la série : Sélectionner le menu 2 et entrer dans la colonne LIST1 les valeurs de la série, puis dans la colonne LIST2 les effectifs correspondants.

Obtention des paramètres :

- Appuyer sur F2(CALC), puis sur F6(SET) (ou F4 sur la graph25).
- Sur la ligne 1VARXLIST, indiquer LIST1 avec les touches de fonctions; sur la ligne 1VARFREQ, indiquer LIST2. Terminer en appuyant sur QUIT.
- En appuyant sur la touche de fonction correspondant à 1VAR, on obtient les paramètres de la série : \bar{x} (moyenne), x_{σ_n} (écart type), n (effectif total), Q_1 (premier quartile), ...

Tracé du diagramme en boîte :

- Dans le menu 2 (STAT), sélectionner le menu GRAPH.
- Sélectionner le menu SET (touche F6 ou F4 deux fois sur la graph25).
- Sur la ligne G-Type, choisir l'option BOX (en appuyant éventuellement sur F4).
- Sur la ligne 1VARXLIST, indiquer LIST1 avec les touches de fonctions; sur la ligne 1VARFREQ, indiquer LIST2. Terminer en appuyant sur QUIT.
- Appuyer sur F1 (GRAPH1) pour obtenir le diagramme en boîte.

Entrée de la série : Appuyer sur la touche STAT, puis sur 1:EDIT. Dans la colonne L1, saisir les valeurs de la série et dans la colonne L2 les effectifs correspondants. Appuyer à nouveau sur STAT.

Obtention des paramètres :

- Sélectionner l'onglet CALC (avec la flèche droite) et appuyer sur la touche 1:1-Var Stats. Appuyer sur 2ND, puis 1 pour afficher L1, puis ,2ND 2 (ne pas oublier la « , »)
- Appuyer sur ENTER pour obtenir les paramètres : \bar{x} (moyenne), σ_x (écart-type), n (effectif total), Q_1 (premier quartile), ...

Tracé du diagramme en boîte :

- Sélectionner le menu STATPLOT en appuyant sur 2ND et Y=.
- Appuyer sur 1 et sélectionner l'option ON.
- Sur la ligne Type, sélectionner le diagramme en boîte (cinquième type de graphique).
- Sur la ligne XList, choisir L1 (en appuyant sur 2ND puis 1).
- Sur la ligne Freq, choisir L2
- Dans le menu WINDOW, indiquer comme Xmin un nombre inférieur à la plus petite valeur de la série, et comme Xmax, un nombre supérieur à la valeur maximale de la série.
- Appuyer sur la touche GRAPH.

Séance de calcul mental.