Statistiques

Ce chapitre de seconde sur les séries statistiques vont introduire (ou ré-introduire) toutes les notions que vous utiliserez dans les années qui viennent pour préparer votre Baccalauréat.
Le chapitre sur les statistiques comportent beaucoup de formules, donc beaucoup de lettres. Pas de panique, essayez juste de comprendre le sens des formules, se sera bien plus simple ensuite pour les retenir. On y va ?

www.mathsbook.fr

I - Notions de base

Dans cette première partie, nous allons (re)voir les notions de base de statistiques. Parmi elles : les effectifs, les fréquences, la médiane, la moyenne... Je suis sûr que vous avez déjà rencontrer ces notions au collège.

1 - Vocabulaire de base

Dans cette section, je vais vous définir les notions de bases, mais alors vraiment de base, sur les séries statistiques. On commence légèrement donc.

Premièrement, qu'est-ce que la statistique ? La **statistique** est tout simplement l'étude d'une **population** composée d'**individus**.

Ensuite, le **caractère** : c'est l'aspect que l'on observe sur les individus. Il peut être qualitatif, quantitatif discret ou quantitatif continu.

Qu'est-ce que cela veut dire "discret" et "continu" ? Et le reste d'ailleurs ?

Je m'explique de suite.

Caractère qualitatif : Si l'on fait, par exemple, une étude statistique sur le mois de naissance d'une population, on parle de caractère qualitatif car on ne parle pas de valeurs numériques. En effet, les mois de l'année ne sont pas des valeurs numériques.

Caractère quantitatif : Si on fait au contraire une étude statistique sur l'âge d'une population, alors là (se sont des valeurs numériques) on parle de caractère quantitatif. On distingue deux caractères quantitatifs distincts :
- Discrète : 16 ans, 17 ans, 18 ans, etc.
- Continue : se sont tout simplement les intervalles : $[15; 20[$, $[20; 25[$, $[25; 30[$, etc.

2 - Effectifs

Plusieurs définitions sur les effectifs.

Définition des effectifs : L'**effectif** de la valeur x_i est le nombre d'individus de la population ayant cette valeur ou appartenant à cette classe : on le note n_i.

L'**effectif total** N est la somme de tous les effectifs : $N = n_1 + n_2 + ... + n_k$.

En rangeant les valeurs du caractère dans l'ordre croissant, on peut calculer l'**effectif cumulé croissant** en faisant la somme des effectifs de cette valeur et de tous ceux qui la précèdent.

Je donne un bon exemple pour vous expliquer ces trois définitions.

Exemple : Dans une classe de 20 élèves de seconde, voici les notes obtenues au dernier contrôle de maths :

Notes	10	14	12	15	7	8	10	11	12	18	2	4	12	13	14	15	19	11	9

On va calculer les effectifs et les effectifs cumulés.

Premièrement, les effectifs : combien d'élèves ont eut 10 ? 2 élève, ok. Combien d'élèves ont eut 12 ? 3 élèves, ok. On continu ainsi et on forme le tableau suivant :

Notes	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Effectifs	1	0	1	0	1	0	0	1	1	1	2	2	3	1	2	2	0	0	1	1	0

Facile non ? Les effectifs cumulés maintenant. On fait la somme des effectifs de la note + la somme de des effectifs de toutes les notes qui la précédent. Ce qui nous donne :

Notes	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Effectifs	1	0	1	0	1	0	0	1	1	1	2	2	3	1	2	2	0	0	1	1	0
Effectifs cumulés	1	1	2	2	3	3	3	4	5	6	8	10	13	14	16	18	18	18	19	20	20

Et voilà.

Remarque : Pour vérifier qu'on ne sait pas trompé dans le calcul des effectifs cumulés, on vérifie bien que le dernier effectif cumulé correspond bien au nombre d'individus.
Ici, on retrouve bien 20, le nombre d'élève de cette classe de seconde.

3 - Fréquences

Passons aux fréquences maintenant.

Définition des fréquences : la **fréquence** d'une valeur est le quotient de l'effectif de la valeur par l'effectif total.

En rangeant les valeurs du caractère dans l'ordre croissant, on peut calculer les **fréquences cumulées croissantes** en faisant la somme des fréquences de cette valeur et de tous ceux qui la précèdent.

Pour les fréquences cumulées croissantes, c'est un peu le même principe que pour les effectifs cumulée croissants.

Petite remarque : Les fréquences sont comprises entre 0 et 1.

Exemple : On reprends l'exemple précédent et on applique tout simplement la formule des fréquences pour les calculer.

Notes	0	1	2	3	4	5	6	7	8	9	10
Effectifs	1	0	1	0	1	0	0	1	1	1	2
Effectifs cumulés	1	1	2	2	3	3	3	4	5	6	8
Fréquence	0,05	0	0,05	0	0,05	0	0	0,05	0,05	0,05	0,1
Fréquence cumulées	0,05	0,05	0,1	0,1	0,15	0,15	0,15	0,2	0,25	0,3	0,4

Et la suite :

Notes	11	12	13	14	15	16	17	18	19	20
Effectifs	2	3	1	2	2	0	0	1	1	0
Effectifs cumulés	10	13	14	16	18	18	18	19	20	20
Fréquence	0,1	0,15	0,05	0,1	0,1	0	0	0,05	0,05	0
Fréquence cumulées	0,1	0,25	0,3	0,4	0,5	0,5	0,5	0,55	0,6	0,6

Remarque : Pareil, pour vérifier qu'on ne sait pas trompé dans le calcul des fréquences cumulés, on vérifie bien que la dernière fréquence cumulés vaut bien 1.
Ici, on retrouve bien 1, c'est bon.

4 - Médiane

On continue avec la définition de la médiane.

Définition de la médiane : C'est la valeur du caractère qui permet de partager la population N en deux groupes de même effectifs. On distingue deux cas : celui d'un caractère quantitatif discret et celui d'un caractère quantitatif continu.

Cas d'un caractère quantitatif discret :
- Si N est impair : la médiane est la valeur du caractère observé au rang $\frac{N+1}{2}$.
- Si N est pair : la médiane n'est pas définie, mais on convient de prendre pour médiane la moyenne des caractères observés au rang $\frac{N}{2}$ et $\frac{N}{2}+1$.

Cas d'un caractère quantitatif continu : on construit la courbe des fréquences cumulées et la médiane est l'antécédent de $0,5$.

Je vais vous donner un exemple simple du cas d'un caractère quantitatif discret.
Exemple : Les notes d'un élève de première sont les suivantes : 3, 5, 12, 14 et 18.
On dénombre cinq notes distinctes, donc un nombre impair de notes.
La médiane est donc la valeur du rang 3. En effet, on applique bêtement la formule précédente :

$$\frac{5+1}{2} = 3$$

D'où : la médiane est 12.

Maintenant, si l'on rajoute la note de 15 à l'élève. On aurait donc les notes suivantes : 3, 5, 12, 14, 15 et 18.
La on est dans le cas d'un nombre de notes pair. On va prendre la moyenne des rang $\frac{N}{2}$, soit 12, et $\frac{N}{2}+1$, soit 14. Ce qui nous donne :

$$\frac{12+14}{2} = \frac{26}{2} = 13$$

La médiane est donc 13.

5 - Moyenne arithmétique pondérée

Bon, je sais que l'expression "moyenne arithmétique pondérée" peut faire peur au premier abord. Mais il n'en est rien en réalité. Vous allez voir.

Définition de la moyenne arithmétique pondérée : La moyenne arithmétique pondérée, que l'on note $\bar{x}$, est donnée par la formule suivante :

$$\bar{x} = \frac{n_1x_1 + n_2x_2 + ... + n_kx_k}{n_1 + n_2 + ... + n_k} = \frac{1}{N} \times \sum_{i=1}^{i=k} n_ix_i$$

Avec $N = n_1 + n_2 + ... + n_k$ et n_i l'effectif de la valeur x_i.

6 - Exemples

Bon, maintenant on va s'exercer un peu sur des exemples pour bien clarifier toutes les notions que l'on vient d'aborder.

Voici donc deux exemples complets à savoir faire et refaire.

Exemple : Etude d'une série statistique à caractère discret :

Dans une classe de 25 élèves de première, les résultats à un contrôle de mathématiques sont les suivants :

7 ; 9 ; 15 ; 11 ; 10 ; 10 ; 16 ; 7 ; 8 ; 14 ; 15 ; 9 ; 10 ; 10 ; 14 ; 15 ; 18 ; 12 ; 8 ; 14 ; 8 ; 8 ; 10 ; 11 ; 15.

Alors, déjà, quelle est la population, le caractère et les valeurs prises par ce dernier ?

...

Eh bien, allez-y ? Vous connaissez la réponse, j'en suis sûr !

Bon, je vous aide.
La population est l'ensemble des contrôles de mathématiques.
Le caractère étudié est la note obtenue par chaque élève de première de cette classe.
Les valeurs prises par le caractères sont les entiers compris entre 7 et 18 (les valeurs des notes quoi).

On va résumer les notes dans l'ordre croissante, l'effectif, l'effectif cumulé et la fréquence dans un tableau :

Note	7	8	9	10	11	12	14	15	16	18
Effectif	2	4	2	5	2	1	3	4	1	1
Effectif cumulé	2	6	8	13	15	16	19	23	24	25
Fréquence	0,08	0,16	0,08	0,2	0,08	0,04	0,12	0,16	0,04	0,04

Normalement, si vous avez bien compris et bien appris toutes les formules précédentes, vous saurez sans aucun problème retrouver toutes les valeurs de ce tableau.

Je l'explique un peu quand même.
La première ligne correspond aux notes des élèves au contrôle de maths. Ca, pas de problème je pense.
La deuxième ligne correspond au nombre de chacune des notes. Par exemple, 2 personnes ont obtenu 7 au contrôle, 4 ont eut 8, etc.
La troisième ligne, c'est la même chose, sauf qu'on compte cette fois-ci combien de personne au eut la note ou moins, soit : 8 personnes ont eut 9 ou moins, etc. On retombe bien sur le nombre total d'élèves, à savoir 25, à la fin.
La dernière ligne, c'est la fréquence. Vous avez la formule un peu plus haut. Pas besoin de réexpliquer.

Calculons maintenant l'étendue, le mode et la médiane.

Calcul de l'étendue : Je vous rappelle que l'étendue est la différence entre la valeur maximale et la valeur minimale, soit ici 11 :

$$18 - 7 = 11$$

Calcul du mode : C'est la valeur qui correspond au plus grand effectif, c'est-à-dire ici la note qui a été obtenue par le plus d'élève. Il s'agit de... 10 ! Oui, 10, obtenue par cinq élèves.
Calcul de la médiane : On a un nombre impair de notes, donc on applique la formule suivante :

$$\frac{25+1}{2} = \frac{26}{2} = 13$$

La médiane est donc la note obtenue par le 13^{me} élève. C'est là que va nous service la ligne des effectifs cumulés.
On lit aisément que le 13^{me} élève a eut 10 à son contrôle de maths, la médiane est donc ici de 10.

Exemple : Etude d'une série statistique à caractère continu :

Dans un lycée, nous avons relevé la taille des élèves et les avons regroupées dans le tableau suivant :

Taille en cm	$[150;160[$	$[160;165[$	$[165;170[$	$[170;175[$	$[175;180[$	$[180;200[$
Effectif	34	42	80	59	50	35

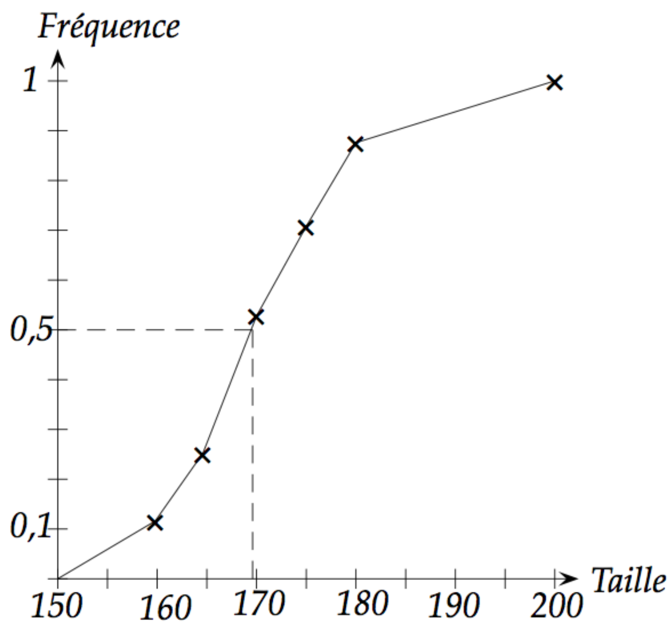
On va calculer, ensemble (oui, je ne vous lâche pas, ne vous inquietez pas) :
- L'étendue,
- La classe modale,
- Le mode,
- La médiane,
- La moyenne.

Alors, pas de temps à perdre, on y va de suite. Je ne rappelle pas à chaque fois les formules pour gagner du temps.

Calcul de l'étendue : $200 - 150 = 50$.
Calcul de la classe modale : $[165; 170[$.
Calcul du mode : C'est le centre de la classe modale, soit : $167,5$.
Calcul de la médiane : Rappelons simplement que dans une série statistique à caractère continu, la médiane est la valeur qui correspond à une fréquence de 0,5. Vous avez compris ce que cela veut dire ? On est obligé de calculer les fréquences oui. Allons-y. Je les ai regroupé dans le tableau suivant :

Taille en cm	$[150; 160[$	$[160; 165[$	$[165; 170[$	$[170; 175[$	$[175; 180[$	$[180; 200[$
Effectif	34	42	80	59	50	35
Fréquence	0,11	0,14	0,27	0,20	0,16	0,12
Fréquence cumulée	0,11	0,25	0,52	0,72	0,88	1

Puis on construit la courbe des fréquences cumulées.



Après lecture graphique, on détermine facilement la médiane qui vaut 169cm.
Calcul de la moyenne : on termine par le plus simple :

$$\frac{34 \times 155 + 42 \times 162,5 + 80 \times 167,5 + 59 \times 172,5 + 50 \times 177,5 + 35 \times 190}{34 + 42 + 80 + 59 + 50 + 35} = \frac{51197,5}{300} = 170,66$$

La moyenne est donc de 170,66cm.

II - Propriétés de la moyenne

1 - Calculs de moyennes à partir de n sous-groupes

Une à connaître sur la moyenne d'une série statistique.

Propriété de la moyenne : Soit une série statistique $(x_1, x_2, ..., x_n)$ de N éléments, réparties en p sous groupes ayant respectivement $n_1, n_2, ..., n_p$ éléments (tels que $n_1 + n_2 + ... + n_p = N$) et de moyennes respectives $\bar{y_1}$, $\bar{y_2}$, ..., $\bar{y_p}$.
Alors la série statistique a pour **moyenne** :

$$m = \frac{n_1\bar{y_1} + n_2\bar{y_2} + ... + n_p\bar{y_p}}{N}$$

Oulala, je n'ai rien compris ! Des explications s'il-vous-plaît ?

Je vous explique tout sur un exemple.

Exemple : Dans un lycée, il y a quatre classes de première contenant respectivement 25, 26, 30 et 29 élèves.
A un contrôle commun de mathématiques, les moyennes de ces classes sont respectivement 10,5 ; 11,2 ; 9,8 et 10,3.
On va calculer la moyenne au contrôle pour l'ensemble des classes de première. Vous allez voir, c'est plus simple que cela en a l'air.
On utilise la formule précédent et c'est tout :

$$m = \frac{25 \times 10,5 + 26 \times 11,2 + 30 \times 9,8 + 29 \times 10,3}{25 + 26 + 30 + 29} = \frac{1146,4}{110} = 10,42$$

Donc, la moyenne au contrôle commun de maths est de 10,42. Simple, non ?

2 - Calculs de moyennes à partir des fréquences

Pour vous expliqué cette partie sur le calcul de moyennes à partir des fréquences, je vais tout simplement vous donner un exemple qui résume tout à lui seul. Oui, parfois un exemple clair et précis vaut mieux qu'un long discours.

Exemple : Dans une ville, 30% de la population n'ont aucun animal domestique, 40% en ont un, 15% en ont deux, 10% en ont trois et 5% en ont quatre.
Quel est le nombre moyen d'animal domestique dans cette ville ?

Voici le calcul à effectuer :

$$m = 0 \times 0,3 + 1 \times 0,4 + 2 \times 0,15 + 3 \times 0,1 + 4 \times 0,05 = 1,2$$

La moyenne est donc de 1,2.
Vous avez saisi l'idée ? En fait on fait la somme des produit entre la valeur et la fréquence de cette valeur.

III - Quartile et décile

Deux autres notions que je vous définis tout de suite avant de les expliquer dans un exemple.

Définitions : Soit une série statistique $(x_1, x_2, ..., x_n)$ de taille n.

- Les **quartiles** partagent la série statistique en quatre parties de même effectif.
  Le premier quartile Q_1 est la plus petite valeur de la série telle qu'au moins 25% des données soient inférieures ou égales à Q_1.
  Le troisième quartile Q_3 est la plus petite valeur de la série telle qu'au moins 75% des données soient inférieures ou égales à Q_3.
- L'**intervalle interquartile** est l'intervalle $[Q_1; Q_3]$.
- L'**écart interquartile** est $Q_3 - Q_1$.
- Les **déciles** partagent la série en 10 parties de même effectif.

Remarques : Deux remarques intéressantes.
- La médiane est le deuxième quartile.
- L'écart interquartile mesure la dispersion des valeurs autour de la médiane, c'est-à-dire la dispersion des 50% des valeurs autour de la médiane.

Et on explique tout ça dans un exemple.

Exemple : Dans une classe de 21 élèves, les notes rangées dans l'ordre croissant sont les suivantes :

Rang	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
Note	1	4	4	5	7	7	8	9	10	11	11	11	11	11	13	14	14	15	16	16	19

Quel est le premier et le troisième quartile ?
Pour le premier quartile, il suffit de faire :

$$21 \times 0,25 = 5,25$$

Donc, le premier quartile est la 6^{me} valeur, soit :

$$Q_1 = 7$$

Et pour le troisième quartile :

$$21 \times 0,75 = 15,75$$

Donc, le premier quartile est la 16^{me} valeur, soit :

$$Q_3 = 14$$

D'autres exemples sont données dans la section suivantes.

Je vous conseille de vous entraîner sur tout ce qu'on vient d'apprendre sur les exercices.