FICHE n°11
Outils statistiques

INTRODUCTION sur les STATISTIQUES
Depuis quand ? Pourquoi ? Comment ? Utilisations actuelles ? Et en 1ère ES ?

Depuis quand ? *Les statistiques dans le temps…*

- Les premiers relevés d'hommes et de bien ont eu lieu ***vers 3000 ans avant J.-C. en Mésopotamie*** ;
- ***L'Egypte des pharaons*** organisait régulièrement des recensements notamment pour les impôts ;
- ***Tycho Brahe*** (1546-1601), astronome danois, utilise la moyenne arithmétique pour réduire les erreurs d'observations ;
- Au XVIIIe et XIXe siècle se développe la ***théorie des erreurs*** ;
- A partir du XXe siècle, les ***ordinateurs*** ont donné une place primordiale aux statistiques car ils permettent de faire rapidement de ***nombreuses simulations***.

Pourquoi ? *Connaissance du passé, connaissance du futur…*

- Assiste-t-on à un ***réchauffement de la planète*** ?
- Faut-il encore ***vacciner*** les enfants contre la rougeole ?
- Une ***pièce*** qui retombe 650 fois sur pile en mille lancers est-elle ***déséquilibrée ?***
- Comment faire pour ***être « sûr »*** que dans un lot de 1000 piles électriques vendues, il y en a au moins 980 qui fonctionnent correctement ?
- ***Météorologie*** : fera-t-il beau dimanche ?

Toutes ces questions ont un point commun : elles sont du **domaine des statistiques**.

Comment ? *Les deux points de vue de la statistique…*

- ***Les recensements*** : ils donnent une image précise de ce que l'on désire observer mais pose des problèmes techniques évidents pour le recueil de données qui sont souvent d'un trop grand nombre…
- ***Les sondages sur des échantillons*** : on effectue un recensement sur une partie seulement de la population à étudier et ces sondages présentent donc une incertitude qu'il faut tenter de minimiser…

Utilités actuelles ? *Les statistiques dans le monde contemporain*

- ***Trouver et décrire une relation***
  - ✓ on établit le risque cardio-vasculaire lié au tabac en étudiant le pourcentage de fumeurs chez les cardiaques et le pourcentage de cardiaques chez les fumeurs et les non-fumeurs ;
- ***Prendre une décision*** :
  - ✓ On fait évoluer des semences de céréales par croisements successifs en comparant divers échantillons correspondants à des techniques ou avancées scientifiques différentes ;
  - ✓ On contrôle la qualité de fabrication et de fiabilité dans l'industrie ;
  - ✓ On détermine l'efficacité d'un médicament ou non ;
- ***Prévoir et planifier*** :
  - ✓ Planifier un budget prévisionnel (commune, collectivités, gouvernement, …)
  - ✓ Planifier des évolutions socio-professionnelles (négociations syndicales, inter-gouvernementales…)

Et en 1ère ES ? *Les outils statistiques*

L'objectif du programme de 1ère ES est de se munir d'outils statistiques (moyenne, écart-type, médiane, écart interquartile, diagramme en boîte, …) afin de pouvoir analyser une série statistique ou mener une comparaison critique et pertinente de deux séries statistiques…

I. *Le couple « Moyenne ; Ecart-type »*

<u>*Définition*</u> Une série statistique étant donnée par le tableau :

Valeur	x_1	x_2	…	x_p	*total*
Effectif	n_1	n_2	…	n_p	N

☞ La ***variance*** de cette série est la moyenne des carrés des écarts entre chaque valeur x_i et la moyenne arithmétique $\overline{x}$ de cette série.

On peut écrire : $$\mathbf{V} = \frac{n_1(x_1 - \overline{x})^2 + n_2(x_2 - \overline{x})^2 + n_3(x_3 - \overline{x})^2 + \ldots + n_p(x_p - \overline{x})^2}{N}$$

☞ L'***écart-type*** **s** d'une série statistique est la racine carrée de sa variance : $\mathbf{s} = \sqrt{\mathbf{V}}$.

<u>*Interprétation*</u> De part sa définition, la variance et l'écart-type permettent de ***mesurer la dispersion*** des valeurs ***autour de sa valeur moyenne***.

Autrement dit, elles permettent aussi de dire si une série de données peut-être représentée ou non par sa moyenne…

EXERCICE TYPE 1 **Comparer la dispersion de deux séries discrètes**

Voici les notes de Laura et Clément au 1er trimestre…

Clément	11	12	11	10
Laura	3	19	18	4

1. Calculer la moyenne et l'écart-type de ces deux séries.
2. Comparer ces indicateurs statistiques et commenter…

<u>*Solution*</u>

Clément

1. Calculons les ***moyennes*** : $\overline{x} = \frac{10 + 11\times2 + 12}{4} = \frac{44}{4} = 11$

Calculons les ***variances*** : $\mathbf{V} = \frac{(10-11)^2 + 2\times(11-11)^2 + (12-11)^2}{4} = \frac{0^2 + 1^2 + 0^2 + (-1)^2}{4} = \frac{2}{4} = 0{,}5$

Calculons les ***écarts-type*** : $\mathbf{s} = \sqrt{\mathbf{V}} = \sqrt{0{,}5} \approx 0{,}71$

Laura

$\overline{x} = \frac{3 + 19 + 18 + 4}{4} = \frac{44}{4} = 11$

$\mathbf{V} = \frac{(3-11)^2 + (19-11)^2 + (18-11)^2 + (4-11)^2}{4} = \frac{(-8)^2 + 8^2 + 7^2 + (-7)^2}{4} = 14{,}12$

$\mathbf{s} = \sqrt{\mathbf{V}} = \sqrt{14{,}125} \approx 3{,}76$

2. Les deux séries de notes ont les mêmes moyennes et médianes mais l'écart type est bien plus important pour les notes de Laura : la dispersion des notes de Laura est effectivement bien plus importante que celle de Clément.

EXERCICE TYPE 2 **Déterminer l'écart-type d'une série continue**

Voici les tailles obtenues lors du recensement des 20 habitants de Smalltown.

Taille (en m)	[1,30 ; 1,50[	[1,50 ; 1,60[	[1,60 ; 1,80[
Effectifs	6	10	4
Centre	1,40	1,55	1,70

Déterminer la moyenne et l'écart-type de ce recensement.

<u>*Solution*</u> Pour une série continue, il faut d'abord déterminer les centres de chaque classe (voir tableau).

✓ Calculons la moyenne : $\overline{x} = \frac{1{,}40\times6 + 1{,}55\times10 + 1{,}70\times4}{20} = \frac{30{,}7}{20} = \mathbf{1{,}535}$

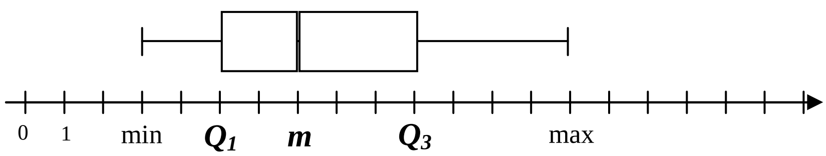
✓ Calculons la variance : $\mathrm{V} = \frac{(1{,}40-1{,}535)^2\times6 + (1{,}55-1{,}535)^2\times10 + (1{,}70-1{,}535)^2\times4}{20} = \frac{0{,}2205}{20} \approx 0{,}011$

L'écart-type est donc : $\mathbf{s} = \sqrt{\mathbf{V}} = \sqrt{0{,}011025} = \mathbf{0{,}105}$

II. *Le couple « Médiane ; écart interquartile » ; diagramme en boîte*

<u>Définitions</u>

- ☞ La ***médiane m*** d'une série est une valeur telle qu'il y ait autant de valeurs inférieures ou égales que de valeurs supérieures ou égales.
- ☞ Le ***1^er^ quartile*** est la plus petite valeur Q_1 de la série telle qu'au moins 25 % des valeurs prises sont inférieures ou égales à Q_1.
- ☞ Le ***3^ème^ quartile*** est la plus petite valeur Q_3 de la série telle qu'au moins 75 % des valeurs prises sont inférieures ou égales à Q_3.
- ☞ L'***écart interquartile*** est la différence $Q_3 - Q_1$.
- ☞ L'***intervalle interquartile*** est $[Q_1 ; Q_3]$.
- ☞ Le ***diagramme en boîte*** est un diagramme regroupant la médiane, le 1^er^ quartile et le 3^e^ quartile.



<u>En pratique</u> Il convient d'***ordonner*** la série ou de réaliser un ***tableau des effectifs cumulés croissants***.

EXERCICE TYPE 3 **Comparer deux séries par leur diagramme en boîte…**

Un même devoir commun de mathématiques a été réalisé dans deux lycées différents avec les résultats suivants :

Notes	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Lycée J	0	0	0	8	8	13	28	27	25	9	3	5	3	0	1	0
Lycée F	1	3	4	7	9	14	19	23	22	7	5	7	6	5	1	2

Réaliser les diagrammes en boîtes de chaque série et les comparer.

<u>Solution</u> ¤ <u>*Etude des résultats du lycée J*</u>

Tableau des effectifs cumulés :

Notes	6	7	8	9	10	11	12	13	14	15	17
E.C.C.	8	16	29	57	84	109	118	121	126	129	130

L'effectif total est 130.

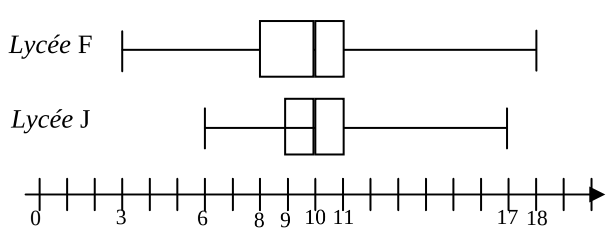
- ❖ 130 ÷ 2 = 65. La médiane est donc comprise entre la valeur de rang 65 et celle de rang 66.
  Le tableau des effectifs cumulés croissants montre que 9 est la 57^ème^ valeur puis que les valeurs sont des 10 de la 58^ème^ à la 80^ème^ valeur… ***La médiane est donc m = 10***.
- ❖ 130 ÷ 4 = 32,5 Le 1^er^ quartile est donc la valeur de rang 33, d'où $\boldsymbol{Q_1 = 9}$.
- ❖ 3× 130 ÷ 4 = 97,5 Le 3^ème^ quartile est donc la valeur de rang 98, d'où $\boldsymbol{Q_3 = 11}$.

¤ <u>*Etude des résultats du lycée F*</u>

De la même manière que pour le lycée J, on peut déterminer et vérifier les résultats suivants obtenus avec un tableur ou avec une calculatrice pour le lycée F : ***m* = 10** ; $\boldsymbol{Q_1 = 8}$; $\boldsymbol{Q_3 = 11}$

¤ <u>*Construction des diagrammes en boîte*</u>

A l'aide des indicateurs ci-dessus déterminés et après avoir repéré les valeurs minimales et maximales pour chaque série, on obtient les diagrammes en boite suivants :



¤ <u>*Commentaires*</u>

Ces deux séries ont la même médiane, mais la série F est beaucoup plus dispersée. Cela signifie que le niveau des élèves du lycée J est beaucoup plus homogène. Les diagrammes en boîte permettent d'avoir une vision très simple de la répartition des notes des élèves de chaque lycée…