Terminale -Maths Complémentaires – Thème 06

RÉPARTITION DES RICHESSES, INÉGALITÉS

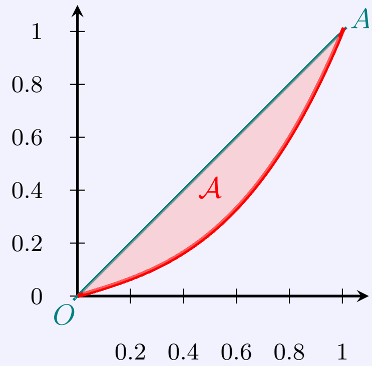


Table des matières

I Statistiques

1) Vocabulaire de base

• **Série Statistique** : une série statistique est un ensemble d'observations collectées.

• **Population** : c'est l'ensemble sur lequel porte une étude statistique. Si elle est trop grande, on peut alors décider de ne s'intéresser qu'à un échantillon de population.

• **Individu** : c'est un élément de la population.

• **Caractère** : c'est ce qu'on observe chez l'individu.

• **Modalités** : ce sont les différentes valeurs prises par le caractère.

• **Série statistique quantitative ou qualitative** : une série statistique est dite **quantitative** quand les modalités sont des nombres (nombre de frères et sœurs, dimensions d'une piège, âges, notes...) et **qualitative** sinon (candidat pour lequel un individu à l'intention de voter, couleurs des yeux...)
Dans le cas d'une série quantitative, celle-ci est dite **discrète** si les modalités sont limitées à un ensemble fini de valeurs (exemple : le nombre de frères et sœurs ne peut être qu'un élément de l'ensemble $\{0; 1; 2; ...; 100\}$) et **continue** si les modalités peuvent prendre n'importe quelle valeur dans un intervalle (exemple : taille d'un individu, température...)

• **Effectif d'une valeur** : c'est le nombre de fois que la valeur d'un caractère (la « modalité ») revient dans la série.

• **Fréquence d'une valeur** : c'est l'effectif de la modalité divisé par l'effectif total : elle est comprise entre 0 et 1. Elle peut également être exprimée sous la forme d'un pourcentage (et est alors comprise entre 0% et 100%).

• **Classe de valeurs** : S'il y a trop de valeurs différentes, elles sont rangées par classe (intervalle), l'effectif de la classe étant alors le nombre de modalités appartenant à cet intervalle.

2) Mesures centrales

Les mesures centrales visent à résumer la série par une seule valeur qu'on espère représentative de toutes les valeurs de la série. On en connaît trois : le **M**ode, la **M**oyenne et la **M**édiane :

a) Le mode

DÉFINITION

Le **mode** d'une série statistique est la donnée la plus fréquente de la série (celle ayant le plus grand effectif).

EXEMPLE

Soit la série de valeurs suivantes : 1 ; 3 ; 4 ; 4 ; 2 ; 2 ; 1 ; 3 ; 4 ; 1 ; 4. Alors le mode est : ....

REMARQUES

- S'il y a plusieurs données arrivant à égalité, il y a plusieurs modes.
- Si les données sont rangées en classe (intervalle), on parle de classe modale.
- Le mode est défini aussi bien pour les séries quantitatives que qualitatives.
- Le mode est un résumé sommaire d'une série qui fournit un type d'information assez limité. Il pourra intéresser un publicitaire.

b) La moyenne

DÉFINITION

La **moyenne arithmétique** d'une série statistique quantitative $S = \{x_1; x_2; ...; x_n\}$ est le nombre, souvent noté $\overline{x}$, défini par :

$$\overline{x} = \frac{x_1 + x_2 + ... + x_n}{n}$$

REMARQUE

Cas d'une série où chaque modalité à un effectif précis :

Valeur x_i	x_1	x_2	x_3	...	x_p
Effectif n_i	n_1	n_2	n_3	...	n_p

Alors :

$$\boxed{\overline{x} = \frac{n_1x_1 + n_2x_2 + ... + n_px_p}{n_1 + n_2 + ... + n_p}}$$ où $n_1 + n_2 + ... + n_p$ est l'effectif total.

REMARQUES

- La moyenne a des avantages calculatoires : si l'on connaît les moyennes et les effectifs de deux séries (ou de deux sous-séries), alors on peut obtenir la moyenne de la série constituée de l'agrégation de ces deux séries.
- La moyenne a le défaut d'être très sensible aux valeurs extrêmes.

c) La médiane

DÉFINITION

On appelle **médiane** d'une série statistique quantitative tout nombre Me tel que :
- La moitié au moins des valeurs de la série est inférieure ou égale à Me.
- La moitié au moins des valeurs de la série est supérieure ou égale à Me.

REMARQUES

- La médiane a l'avantage de ne pas être influencée par les valeurs extrêmes.
- La médiane n'a aucun avantage pratique dans les calculs, puisque pour connaître la médiane d'une série constituée de l'agrégation de deux séries, il faut nécessairement réordonner la nouvelle série pour trouver la médiane, qui n'aura alors pas de lien avec les médianes des deux séries initiales.

3) Mesures de dispersion

Les mesures de dispersion d'une série étudiées sont au nombre de quatre : L' **É**tendue, l'**É**cart inter-quartile, l'**É**cart inter-décile et l'**É**cart-type.

a) L'étendue

DÉFINITION

Les valeurs extrêmes d'une série sont ses valeurs minimale et maximale et l'étendue est la différence entre les valeurs extrêmes de la série.

EXEMPLE

Déterminer les valeurs extrêmes et l'étendue de la série : 2 ; 5 ; 0,5 ; 4 ; 3 ; 2.
Les valeurs extrêmes sont ...... et ...... donc l'étendue vaut : ....

b) L'écart inter-quartile

DÉFINITION

Soit S une série statistique quantitative.

On appelle **premier quartile**, noté Q_1, toute valeur de la série S telle que :
- au moins 25% des valeurs de la série ont une valeur inférieure ou égale à Q_1.
- au moins 75% des valeurs de la série ont une valeur supérieure ou égale à Q_1.

On appelle **deuxième quartile** (ou **médiane**), noté **Me**, toute valeur de la série S telle que :
- au moins 50% des valeurs de la série ont une valeur inférieure ou égale à **Me**.
- au moins 50% des valeurs de la série ont une valeur supérieure ou égale à **Me**.

On appelle **troisième quartile**, noté Q_3, toute valeur de la série S telle que :
- au moins 75% des valeurs de la série ont une valeur inférieure ou égale à Q_3.
- au moins 25% des valeurs de la série ont une valeur supérieure ou égale à Q_3.

DÉFINITION

L'écart inter-quartile est la différence $Q_3 - Q_1$.
L'intervalle $[Q_1\,;Q_3]$ est appelé **l'intervalle inter-quartile** ; il contient au moins 50% des valeurs de la série.

c) L'écart-interdécile

DÉFINITION

Soit S une série statistique quantitative.

On appelle **premier décile**, noté D_1, toute valeur de la série S telle que :
- au moins 10% des valeurs de la série ont une valeur inférieure ou égale à D_1.
- au moins 90% des valeurs de la série ont une valeur supérieure ou égale à D_1.

On appelle **neuvième décile**, noté D_9, toute valeur de la série S telle que :
- au moins 90% des valeurs de la série ont une valeur inférieure ou égale à D_9.
- au moins 10% des valeurs de la série ont une valeur supérieure ou égale à D_9.

DÉFINITION

L'écart inter-décile est la différence $D_9 - D_1$.
L'intervalle $[D_1\,;D_9]$ est appelé **l'intervalle inter-décile** ; il contient au moins 80% des valeurs de la série.

d) L'écart-type

DÉFINITION

Soit S une série statistique quantitative comportant n données : $S = \{x_1; x_2; \ldots; x_n\}$ et de moyenne $\overline{x}$.
On appelle :
Variance de S le réel $V = \dfrac{(\overline{x}-x_1)^2 + (\overline{x}-x_2)^2 + \ldots + (\overline{x}-x_n)^2}{n}$.
Ecart-type de S le réel $\sigma = \sqrt{V}$.

EXEMPLE

Soit la série de notes : 7 7 7,5 9 9,5 11 13 14,5 15.
Calculer la moyenne, la variance et l'écart-type de cette série de notes.

REMARQUE

Cas d'une série où chaque modalité à un effectif précis :

Valeur x_i	x_1	x_2	x_3	...	x_p
Effectif n_i	n_1	n_2	n_3	...	n_p

Alors :

$$\overline{x} = \frac{n_1x_1 + n_2x_2 + \ldots + n_px_p}{n_1 + n_2 + \ldots + n_p} \qquad \text{et} \qquad V = \frac{n_1(\overline{x}-x_1)^2 + n_2(\overline{x}-x_2)^2 + \ldots + n_p(\overline{x}-x_p)^2}{n_1 + n_2 + \ldots + n_p}$$

REMARQUES

- La **variance** est la moyenne des carrés des écarts à la moyenne. Elle mesure donc la dispersion des valeurs autour de la moyenne. Elle n'est pas très parlante car elle s'exprime dans le carré de l'unité du caractère.
- L' **écart-type** a l'avantage de s'exprimer dans la même unité que le caractère.
- L'**écart-type** permet de comparer la dispersion de deux séries, quand l'ordre de grandeur des données des deux séries est le même. Contrairement à l'écart inter-quartile, il tient compte de l'ensemble de la population.

REMARQUE

Autre formule pour la variance :

$$V = \frac{n_1x_1^2 + n_2x_2^2 + \ldots + n_px_p^2}{n_1 + n_2 + \ldots + n_p} - \overline{x}^2$$

II Statistiques en analyse

1) Valeur moyenne d'une fonction continue

On rappelle la définition de la valeur moyenne vue au chapitre précédent :

DÉFINITION

Soit f une fonction continue sur un intervalle $[a\,;b]$ (avec a et b des réels et $a < b$).
On appelle valeur moyenne de f sur $[a\,;b]$ le réel :

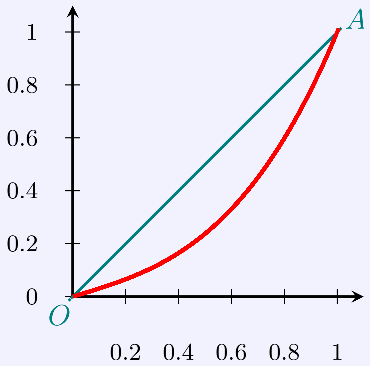
$$\mu = \frac{1}{b-a}\int_a^b f(x)\,\mathrm{d}x$$

REMARQUE

On peut interpréter de façon similaire la valeur moyenne μ pour une fonction f à la moyenne $\overline{x}$ d'une série statistique.

2) Courbe de Lorenz et indice de Gini

DÉFINITION



Une **courbe de Lorenz** est une représentation graphique qui sert à évaluer les inégalités de répartition d'une grandeur dans une population donnée.

Pour la construire, on place en abscisse le pourcentage cumulé de la population, et en ordonnée le pourcentage cumulé de la grandeur.

REMARQUE

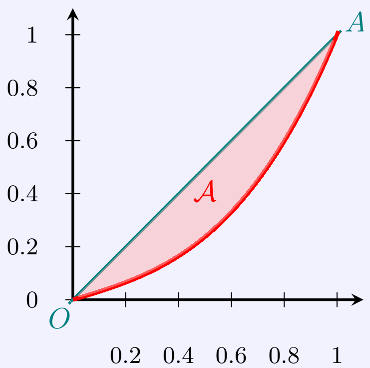
Une courbe de Lorenz vérifie les propriétés suivantes :
- Elle passe par les points de coordonnées $(0;0)$ et $(1;1)$.
- Elle représente une fonction croissante et convexe sur $[0;1]$.
- Elle est située en-dessous de la droite d'équation $y = x$ (conséquence immédiate des deux propriétés précédentes).

EXEMPLE

Démontrer que la fonction f définie sur $[0;1]$ par $f(x) = 0{,}2x^2 + 0{,}8x$ est représentée par une courbe de Lorenz (On montre que f vérifie les deux premières propriétés ci-dessus).

Interprétation de la courbe de Lorenz :
Plus la courbe de Lorenz est proche de la droite d'équation $y = x$, plus la répartition est égalitaire.
Plus la courbe de Lorenz s'en éloigne, plus les grandeurs observées se concentrent chez peu de personnes.

DÉFINITION



Soit $\mathcal{C}$ une courbe de Lorenz, dans un repère orthogonal, et soient $O(0;0)$ et $A(1;1)$ deux points.

L'indice de Gini, noté G, est égal au double de l'aire $\mathcal{A}$ de la partie délimitée par la courbe $\mathcal{C}$ de Lorenz et le segment $[OA]$.

Autrement dit, $G = 2 \times \mathcal{A}$.

Interprétation de l'indice de Gini :
Plus l'indice de Gini est proche de 0, plus la répartition est équitable.