

Statistiques à une variable

I Généralités

Définition : c'est un relevé de valeurs, effectué sur une population, relatif à une caractère (ex. marque de voiture, taille, âge, etc.), souvent donné sous la forme d'un tableau.

Un caractère non mesurable est qualitatif (ex. : couleur)

Un caractère mesurable est quantitatif. Il est :

- continu s'il peut prendre toutes les valeurs d'un intervalle (on parle de répartition en classe)
- discret s'il ne prend que des valeurs ponctuelles (pointure).

Quand le caractère est quantitatif il prend différentes valeurs, on l'appelle alors variable (x_i). A chaque valeur de la variable est associé un effectif (n_i). Pour les répartitions en classe, on donne aux x_i les valeurs centrales des classes.

II Les indicateurs de tendance centrale des séries quantitatives

1/ Mode, classe modale

Le mode d'une série statistique quantitative est la valeur du caractère qui à le plus grand effectif.

Exemples :

Age	Effectif

Taille	Effectif

2/ La moyenne

La moyenne d'une série statistique, notée \bar{x} , est obtenue par :

$$\bar{x} = \frac{\sum_{i=1}^p n_i \cdot x_i}{N} = \frac{n_1 \cdot x_1 + n_2 \cdot x_2 + \dots + n_p \cdot x_p}{N}$$

Avec x_i : valeur du caractère (ou centre de la classe)
 n_i : effectif
 N : effectif total
 p : nombre de classes

3/ La médiane : Me

Définition : C'est la valeur de la variable (du caractère) permettant de partager la population en deux groupes d'effectifs le plus voisins possibles : ceux ayant une valeur inférieure à la médiane et ceux ayant une valeur supérieure.

Détermination : par classement pour des petits effectifs, par recherche du mi effectif dans les ECC pour les grands effectifs ou par outils de calculs (TIC).

Remarque : pour les répartitions en classe, on prendra la valeur centrale de la classe médiane.

III Les indicateurs de dispersion des séries quantitatives

1/ L'étendue

L'étendue est la différence entre les deux valeurs extrêmes du caractère

2/ L'écart-type

L'écart type (noté σ , sigma) est la racine carrée de la moyenne des écarts de chaque valeur à la moyenne.

Formules :

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

$$\sigma = \sqrt{V} .$$

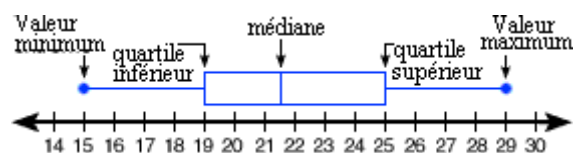
3/ L'écart interquartile

Définition : En statistique un quartile est chacune des 3 valeurs (noté Q_1 , Q_2 , Q_3) qui divisent les données triées en 4 parts égales, de sorte que chaque partie représente 1/4 de l'échantillon de population.

Remarque : $Q_2 = Me$

Définition : L'écart interquartile est la différence entre le dernier et le premier quartile : $Q_3 - Q_1$

Diagramme en boîte (ou boîte à moustache) : Les diagrammes à boîte et moustaches sont idéals pour comparer des distributions, parce qu'ils font apparaître immédiatement un indicateur de tendance centrale (la médiane), un indicateur de dispersion (l'écart interquartile) et l'étendue.



IV Probabilité

1/ Fluctuation d'échantillonnage

Si l'on répète n fois une expérience aléatoire, l'ensemble des résultats collectés s'appelle un échantillon de taille n .

Exemples :

On lance 100 fois un dé cubique équilibré. On obtient un échantillon de 100 nombres compris entre 1 et 6.

On observe la durée de 300 appels téléphoniques d'une hotline pour savoir si la durée est inférieure à 30 minutes. On obtient un échantillon de 300...

La fréquence sur plusieurs échantillons de même taille prends des valeurs différentes. Ce phénomène s'appelle la fluctuation d'échantillons.

2/ Distribution d'échantillonnage d'une fréquence

En déterminant la fréquence du caractère étudié dans N échantillons de même tailles, on obtient une série statistique des N fréquences. Cette série constitue une distribution d'échantillonnage d'une fréquence.

Exemple :

On s'intéresse à l'apparition de la face 6 dans 5 échantillons de 200 lancers d'un dé équilibré. Le tableau suivant donne les résultats des observations.

Echantillon : i	1	2	3	4	5	Moyenne
Nombre d'apparition de la face 6	30	32	38	34	26	
Fréquence de l'échantillon : f_i						

Propriété : Stabilisation de la fréquence moyenne.

Lorsque la taille des échantillons (n) augmente la fréquence moyenne se stabilise vers, un nombre noté p , appelé fréquence de la population.

3/ Intervalle de fluctuation

L'intervalle de fluctuation au seuil de 95% est : $\left[p - \frac{1}{\sqrt{n}} ; p + \frac{1}{\sqrt{n}} \right]$

Où n est la taille des échantillons et p la fréquence moyenne.

Lorsque $n > 30$, $np > 5$ et $n(1-p) < 10$, alors 95% des échantillons ont une fréquence du caractère étudié comprise dans cet intervalle.