RAPPEL SUR LA STATISTIQUE DESCRIPTIVE

Notions de bases

Statistiques : Ce sont des collections de nombres présentées sous forme de tableaux ou de graphes groupant des observations relatives à un phénomène considéré.

Population et Echantillon statistique : L'ensemble des observations possibles sur la variable étudiée est appelée **population statistique.**

Comme il n'est pas toujours possible d'étudier toute la population, l'étude statistique portera alors sur une partie de celle ci. Cette partie de population est appelée **échantillon statistique** (à ne pas confondre avec échantillon géologique).

Variable statistique : On appelle **variable statistique** le caractère sur lequel porte l'étude. Ce caractère peut être qualitatif : la couleur d'un minéral, ou quantitatif : les teneurs chimiques d'une série d'échantillons géologiques.

Cette variabilité est dite continue si les valeurs sont très proches les unes des autres. S'il existe un certain intervalle constant ou non entre les valeurs, comme par exemple les teneurs obtenues par analyse spectrale semi - quantitative, alors la variable est dite discrète.

Effectif : L'effectif de la valeur xi est le nombre d'individus de la population ayant cette

valeur ou appartenant à cette classe : on le note ni.

L'effectif total N est la somme de tous les effectifs : N = n1 + n2 + ... + nk.

Série statistique : L'ensemble des données de la variable est appelé série statistique. Si les valeurs sont classées par ordre croissant, la série statistique est dite ordonnée et la différence entre la plus petite et la plus grande valeur est appelée **étendue de la série.**

Classes et intervalle de classes : Quand les valeurs d'une série statistique sont très proches les unes des autres, elles sont alors regroupées en classes dont l'amplitude est constante. Cette amplitude est appelée intervalle de classe.

Il n'existe pas de méthodes universelles pour le choix de l'intervalle de classe. Généralement on prend n = √N où n est le nombre de classes, et N l'effectif de l'échantillon statistique.

L'amplitude des classes k est égale à l'étendue divisée par le nombre de classes.

$$K = \frac{X\,\max - X\,\min}{n}$$

Xmin et Xmax sont respectivement la plus petite et la plus grande valeur de la série statistique.

Centre de classe et variable aléatoire : La valeur qui correspond à la demie - somme des valeurs extrêmes de la classe est appelée centre de classe. Elle est notée Xi et appelée variable aléatoire.

Fréquence relative et fréquence cumulée : 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 l'effectif cumulé croissants.

1-ANALYSE MONOVARIEE
1.1. PARAMETRE STATISTIQUE

1.1.1. PARAMETRES DE POSITION CENTRALE

Ces paramètres permettent de quantifier la tendance centrale des valeurs d'une série statistique. Les principaux paramètres de tendance centrale sont le mode, la médiane et la moyenne arithmétique.

Le mode : Désigné généralement par Mo, il est défini comme étant la valeur de la variable aléatoire qui a l'effectif le plus élevé. Une série statistique peut être uni ou multimodale.

Le nombre de modes d'une série statistique renseigne sur l'homogénéité ou l'hétérogénéité de l'échantillon ou population statistique. Cependant dans le cas de classement de la série statistique, le nombre de modes peut être fonction du nombre de classes et l'intervalle de classe.

La médiane : valeur de la variable telle qu'une moitié des valeurs lui soit supérieure ou égale et l'autre moitié des valeurs lui soit inférieure ou égale. Deux cas apparaissent suivant la parité de n.

n impair : $\text{médiane} = x_{\frac{n+1}{2}}$

n pair : $\text{médiane} = \frac{1}{2} \cdot \left(x_{\frac{n}{2}} + x_{\frac{n}{2}+1} \right)$

La moyenne arithmétique :

Rappelons les principales propriétés des sommes algébriques :

1 - $\sum ax_i = a\sum x_i$

2 - $\sum (ax_i + b) = \sum ax_i + \sum b$

3 - $\sum (x_i + y_i) = \sum x_i + \sum y_i$

Si N est l'effectif de l'échantillon statistique, n' le nombre de classes et n_i l'effectif de la ième classe alors on peut écrire :

$$N = n_1 + n_2 + .. + n_i + .. + n_{n'}$$

La moyenne arithmétique désignée souvent par M ou $\bar{x}$ est égale à la somme de la série statistique divisée par l'effectif N.

Si X est une variable continue alors :

$$\overline{X} = M = \sum_{i=1}^{N} \frac{X_i}{N}$$

Si X est une variable discrète et si X_i représente le centre de la classe i et n_i/N; son effectif f_i alors :

$$\overline{X} = M = \sum_{i=1}^{k} f_i . X_i$$

La moyenne arithmétique joue le rôle d'un certain milieu par rapport aux extrêmes. Elle est l'analogue d'un centre de gravité.

1.1.2. PARAMETRES DE DISPERSION

Les paramètres de dispersion permettent de quantifier la dispersion des valeurs de la série statistique. Les principaux paramètres sont l'étendue, les quartiles, la variance, l'écart type et le coefficient de variation.

L'étendue : c'est la différence entre les valeurs extrêmes de la série statistique ordonnée.

Quartile : valeurs Q1. Q2 et Q3 de la grandeur mesurée qui partagent la série statistique en 4 parties d'effectifs à peu près identiques.
Q1

Q2 est la médiane.

On calcule $\left(\frac{n+1}{4}\right)$ pour le rang de Q1 et $3\left(\frac{n+1}{4}\right)$ pour le rang de Q3.

Si ces grandeurs ne sont pas des entiers, les quartiles ne sont donc pas des valeurs de la distribution. On réalise alors une interpolation.

- **Etendue interquartile :** intervalle contenant la moitié de la population autour de la médiane c-a-d Q3 – Q1

- **Etendue R :** **R = xmax - xmin**

Variance : elle est désignée par S^2. Dans le cas de variable continue, elle est égale à :

$$S^2 = \frac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{N} = \sum_{i=1}^{n}\frac{x_i^2}{N} - \bar{x}^2$$

N - l'effectif total; $\bar{x}$ - moyenne arithmétique et x_i - variable

dans le cas de variable discrète alors :

$$S^2 = \sum_{i=1}^{n'}[f_i(X_i - \bar{x})^2]$$

n' - nombre de classe, X_i - centre de classe, f_i - fréquence relative

de la classe i

Le calcul de la variance suppose déjà connu la moyenne arithmétique. La variance d'un échantillon statistique est généralement désigné par S^2 et celle de toute la population par σ^2.

Ecart type S ou σ : C'est la racine carrée de la variance

Coefficient de variation : il représente une sorte d'écart-type relatif pour comparer les dispersions indépendamment des valeurs de la variable. Il s'exprime souvent en pourcentage.

$$CV = \left(\frac{\text{écart type}}{\text{Moyenn}}\right) \quad 100$$

1.1.3. REPRESENTATION GRAPHIQUE

Il existe plusieurs types de graphes; cependant les graphes les plus utilisés sont ceux en bâtons, en rectangles et les polygones de fréquences. Le graphe formé par des rectangles ayant pour largeurs les classes et pour hauteur une longueur proportionnelle à l'effectif correspondant n_i, est appelé **histogramme**. On obtient un histogramme en plaçant les centres de classes par ordre croissant sur l'axe des abscisses et les fréquences relatives sur l'axe des ordonnées (Fig.1).