

La variation génétique et la génétique des populations

1) L'étude quantitative de la variation : la biométrie

I) Les types de caractères des variables :

Dans une population les individus sont différents par plusieurs caractères :

Caractère qualitatif :

Quand la réponse donnée est un mot ou une expression (la langue maternelle, le sexe, la couleur des yeux, etc.).

Caractère quantitatif

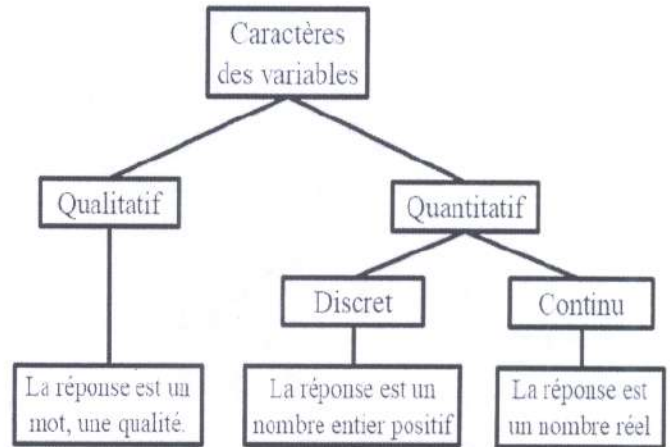
Quand la réponse donnée est un nombre. Selon la nature de ce nombre, cette variable sera discrète ou continue.

* Caractère quantitatif discret

Quand la réponse donnée est un nombre naturel (nb d'enfants dans une famille, nb d'amis, nb de maisons, etc.).

* Caractère quantitatif continu

Quand la réponse donnée peut prendre n'importe quelle valeur dans un intervalle donné appelé domaine de variation et défini par les valeurs X_{min} et X_{max} . (la taille d'une personne, le temps, etc.).



II) Représentation graphique des données

Après avoir obtenu un échantillon ou dénombré une population, on dispose le plus souvent de données numériques brutes présentées sous la forme d'une série de valeurs. Présentées ainsi, ces données sont rarement « parlantes » et il est nécessaire de dresser une représentation graphique afin de faire ressortir une partie de l'information. Suivant le type de variable, le mode de représentation graphique va être différent.

Variable quantitative discontinue :

Ce type de variable est associée généralement à un diagramme en bâtons où l'axe horizontal des abscisses porte les valeurs prises par la variable (x_i) tandis que l'axe vertical des ordonnées porte l'effectif absolu (n_i) observé.

Variable quantitative continue :

Dans le cas de ces variables, suivant la grandeur mesurée et la sensibilité de la méthode utilisée, il est fréquent d'obtenir autant de valeurs différentes que de données si bien que la représentation graphique n'a dans ces conditions aucun sens. On est donc généralement amené à regrouper les données en classes.

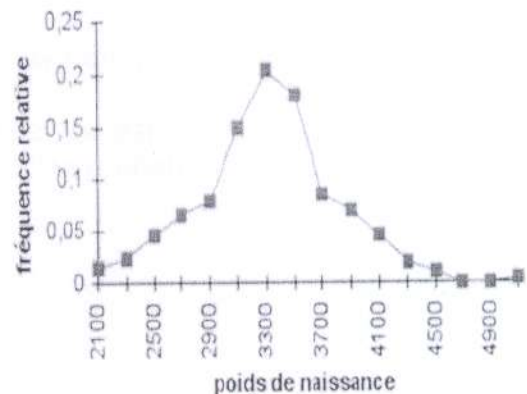
A la différence des variables discontinues, une classe donnée ne contient pas une seule valeur mais une infinité de valeurs possibles sur un intervalle défini (appelé intervalle de classe). Cet intervalle permet de définir également une amplitude de classe (différence entre les valeurs supérieure et inférieure de la classe). La valeur centrale de la classe est appelée centre de classe.

Pour présenter graphiquement une série statistique à variation continue on utilise un histogramme : C'est un ensemble de rectangles accolés ayant les caractéristiques suivantes :

1. la base de chaque rectangle correspond à l'amplitude d'une classe. Généralement toutes les classes d'une série statistique ont même amplitude.
2. la hauteur du rectangle est égale soit à l'effectif absolu (ou fréquence absolue) de la classe, soit à la fréquence relative (correspondant au rapport n_i/n où n_i est l'effectif absolu de la classe i et n l'effectif total de la série statistique). La surface de chacun des rectangles, si l'amplitude de classe est constante est alors proportionnelle à l'effectif de la classe

Le polygone des fréquences représente le deuxième type de représentation graphique utilisée dans l'étude de la variation continue

Le polygone des fréquences est représenté en joignant les milieux des cotés supérieurs des rectangles dans un histogramme. C'est une ligne brisée dont les extrémités rejoignent l'axe des abscisses.



Conclusion

La représentation graphique d'une série statistique est une étape très importante dans l'analyse d'un problème statistique car elle donne une information sur la forme de la distribution observée. Pour un échantillon donné, l'analyse graphique de la distribution, couplée à une étude statistique plus approfondie fournit une indication sur la distribution de la variable aléatoire dans la population d'où l'on a extrait l'échantillon.

III) Paramètres de position et paramètres de dispersion

a) Paramètres de position :

1) Le mode : Le mode ou valeur dominante est la valeur la plus fréquente d'une distribution. Il se calcule toujours à partir d'un dénombrement des modalités du caractère. Comme pour le tableau de dénombrement, il faut distinguer le cas des caractères discrets et des caractères continus.

Caractère quantitatif discontinu : Pour un caractère quantitatif discontinu, le mode est la modalité qui a la fréquence la plus élevée (ou l'effectif le plus élevé).

Caractère quantitatif continu : Le mode est le centre de la classe modale, c'est à dire la classe qui a la fréquence moyenne la plus élevée.

NB : Le mode correspond au sommet de l'histogramme. Lorsque celui-ci présente deux pics séparés par un creux, on dit que la distribution est bimodale. On distingue alors un mode principal et un mode secondaire. Plus rarement, on peut trouver des distributions ayant trois modes et plus.

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

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

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

b) Paramètres de dispersion

- Ecart-type

Définition : l'écart-type est la racine carrée de la variance, elle-même définie comme la moyenne du carré des écarts à la moyenne :

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

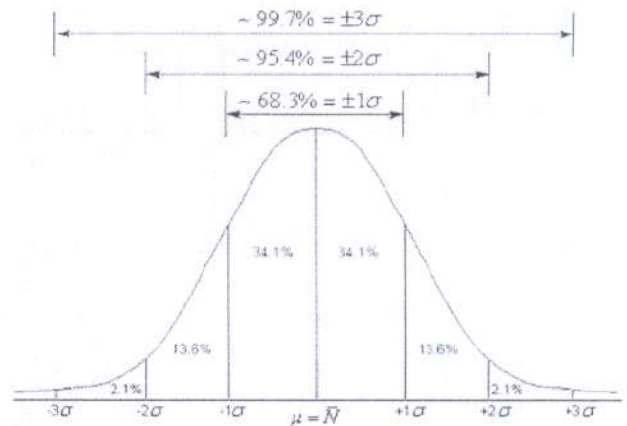
Formules de la variance et de l'écart-type :

Les utilisateurs de l'écart-type doivent être conscient que l'utilisation de celui-ci n'est pleinement justifié que dans le cas où la distribution des valeurs de la distribution observé est de forme gaussienne ou, à tout le moins, symétrique et unimodale.

$$\text{écart - type : } \sigma_x = \sqrt{\text{Variance}} = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2}$$

En effet, dans le cas où la forme de la distribution observée est de type gaussien (mais dans ce cas seulement !) l'écart-type peut revêtir une signification probabiliste et servir à définir des intervalles de confiance autour de la moyenne.

intervalle	% des valeurs d'une distribution gaussienne
moyenne + ou - 1 σ_x	68.3%
moyenne + ou - 2 σ_x	95.5%
moyenne + ou - 3 σ_x	99.7%



VI la sélection artificielle :

La sélection artificielle est un procédé qui consiste à croiser volontairement les organismes qui disposent de caractères (couleur, goût, productivité...) que l'on désire perpétuer.

C'est par ce procédé que l'agriculture a créé les espèces végétales et animales qui sont cultivées et élevées actuellement pour produire de la nourriture et d'autres produits (textiles, arômes...). C'est aussi de cette façon que les animaux de compagnie et les plantes horticoles ont été créés.

La sélection artificielle est efficace dans une population hétérogène . c'est une population qui renferme plus d'une lignée pure (les populations hétérogènes présentent le plus souvent plus qu'un mode pour le caractère étudié.)

L'importance de la sélection réside dans l'obtention de lignées pures dont les individus se caractérisent par un phénotype quasi-identique et lorsqu'on effectue des croisements entre les individus portant ce phénotype celui-ci apparait de manière identique à chaque génération : la sélection devient inefficace.