République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

UNIVERSITÉ MOHAMED KHIDER, BISKRA

FACULTÉ des SCIENCES EXACTES et des SCIENCES de la NATURE et de la VIE
DÉPARTEMENT DE Biologie

Chapitre 03 :

Le 20/11/2021

Par
Prof : CHALA ADEL

Mathémathiques-Statistiques

2021-2022

Je dédie ce travail............

A mes parents ils m'ont tous,
avec leurs moyens, soutenu et donné
la force d'aller toujours
plus loin.
A ma chère femme Houda.
A l'esprit du professeur Bahlali Seid

Table des matières

Chapitre 1

Statistique descriptive

Lorsqu'on veut étudier les données relatives aux caractéristiques d'un ensemble d'individus ou d'objets, il est difficile d'observer toutes les données lorsque leurs nombres sont élevés. Au lieu d'examiner l'ensemble qu'on appelle la population, on examine un nombre restreint qu'on appelle échantillon, -pour être représentatif- l'échantillon doit être pris au hasard (une population peut-être finie ou infinie).

1.1 Notions des bases

Définition 1 ***Population :*** *C'est l'ensemble sur lequel porte l'étude statistique.*

Définition 2 ***Individus :*** *Les éléments de cet ensemble.*

Définition 3 ***Echantillon :*** *Est un sous-ensemble de la population.*

Définition 4 ***Caractère :*** *C'est le trait (ou propriété) choisi pour l'étude statistique.*

Définition 5 ***Modalités :*** *Les différentes positions que peut prendre un caractère. Usage on numéroté les modalités de 1 à k, la modalité numéro i est notée* C_i.

Définition 6 ***Effectifs :*** *Lorsque la population est répartie sur les différentes modalités nous obtenons pour chacune d'elles un nombre, c'est le nombre des individus ayant cette modalité. On note habituellement* n_i *l'effectif correspondant à la modalité* C_i *: ; les fréquences absolu..*

Définition 7 ***Fréquence relative :*** *Par définition c'est le rapport entre* n_i *et* N*, où* N *est la somme totale des individus.*

Nous allons ainsi adopter les définitions suivantes :

Définition 8 *Un caractère est dit quantitatif quand ses différentes modalités sont mesurables par des nombres qui en indiquent l'intensité.*

Définition 9 *Un caractère est dit qualitatif quand ses différentes modalités ne peuvent être désignées que par leurs qualités.*

Définition 10 *Une variable statistique est dite discrète lorsque ses modalités ne peuvent être que des nombres isolés.*

Définition 11 *Une variable statistique est dite continue quand elle peut prendre n'importe quelle valeur dans un intervalle donné.*

1.1.1 Paramètres des positions centrales

Définition 12 ***Le mode*** *: c'est la valeur la plus fréquente.*

Définition 13 ***La médiane*** *: C'est la valeur de la variable statistique qui partage la population en deux populations d'effectifs égaux.*

Définition 14 ***Les quartiles :*** *Comme on a définit la médiane, alors on peut définir des paramètres qui la répartissent en quarts.*

Définition 15 ***La moyenne arithmétique*** *: est égale par définition* $\overline{X}_n = \frac{1}{n}\sum_1^n X_i$.

1.1.2 Paramètres Caractéristiques de dispersion

Définition 16 ***L'étendue :*** *C'est la longueur de l'intervalle sur lequel se disperse la variable.*

Définition 17 ***L'écart-interquartiles*** *: C'est la différence entre les deux quartiles* Q_1 *et* Q_3.

Définition 18 ***La variance :*** *C'est la caractéristique qui est réellement utilisée pour mesurer la dispersion :*

1.1.3 Effectifs et fréquence cumulés croissantes

Définition 19 *Quand les modalités ou les classes d'un caractère sont rangés dans l'ordre croissante, alors l'effectif cumulés croissantes d'une valeur s'obtient en ajoutant à chaque effectif les effectifs des valeurs qui la précédent, d'où*

$$N_j \nearrow = \sum_{i=1}^{j} n_i.$$

Remarque 20 *L'effectif cumulés décroissantes d'une valeur s'obtient en dimunant à chaque effectif les effectifs des valeurs qui la précédent, d'où*

$$N_j \searrow = 1 - \sum_{i=1}^{j} n_i.$$

Définition 21 *La fréquence cumulés croissantes d'une valeur s'obtient en ajoutant à chaque fréquence les fréquences des valeurs qui la précédent.*

$$F_j = \sum_{i=1}^{j} f_i.$$

Remarque 22 *La fréquenc cumulés décroissantes d'une valeur s'obtient en dimunant à chaque fréquence les fréquences des valeurs qui la précédent, d'où*

$$F_j \searrow = 1 - \sum_{i=1}^{j} f_i.$$

1.2 Représentation graphique

1.2.1 Diagramme en bâtons

Définition 23 *Lorsque le caractère étudié est quantitatif discret, on peut représenter la série statistique par un diagramme en bâtons, dont la hauteur de chaque bâton est alors l'effectif (la fréquence) associe à chaque valeur.*

1.2.2 Courbe de fonction de répartition

Définition 24 *Sut l'axe d'abscices, on trouve les valeurs de caractère, et sur l'axe des ordonnées on trouve les valeurs correspondantes des effectifs cumulés croissantes. D'où la courbe est représentée comme une fonction en escalier.*

1.2.3 Diagramme circulaire

Définition 25 *La mesure de chaque secteur angulaire est proportionnelle à chaque effectif associe. L'angle de chaque modalité se calcul par la formule suivante*

$$\alpha^{\circ} = \frac{\text{Effectif de modalité}}{\text{Effectif total}} \times 360^{\circ}.$$

1.3 série Statistique discrèt

Lorsque le caractère ne prend que des valeurs (ou modalités) numériques, il est quantitatif. Le caractère quantitatif est dit discrèt s'il ne peut prendre que des valeurs isolées.

1.3.1 Paramètres caractéristiques de tendance centrale

Ces valeurs permettent de savoir autour de quelles valeurs se situent les valeurs d'une variable statistique (Caractère statistique).

Définition 26 ***Mode*** *: C'est la modalité la plus fréquente, ou bien c'est la modalité qui possède le plus grand effectif.*

Remarque 27 *On peut établir la valeur de la mode graphiquement, en utilisant le diagramme en bâtons, par observé le bâton le plus grand.*

Définition 28 ***Médiane*** *C'est la modalité qui partage l'échantillon en deux parties d'effectifs égaux, contenant le même nombre d'individus, et on peut calculer par la formule suivante*

$$Me = N^{-1} \nearrow \left(\frac{N}{4}\right) = F^{-1} \nearrow (0,25).$$

Remarque 29 *On peut établir la valeur de la médiane graphiquement, en utilisant la courbe de la fonction de répartition, par faisant l'intersection de la courbe de* $N_j \nearrow$ $(F_j \nearrow)$ *avec la courbe de* $N_j \searrow$ $(F_j \searrow)$. *ou bien en faisant la projection de la valeur* $(\frac{N}{2})$ *dans l'axe des cordonnées sur l'axe des abscices par la fonction de répartition.*

Définition 30 ***Les quartiles*** *Sont les modalités qui partages l'échantillon en quarte parties d'effectifs égaux, contenant le même nombre d'individies.*

Définition 31 ***Pour*** Q_1 *C'est le premièr quartile, c'est lui qui prend la valeur de modalité qui donne la situation de la première 25% de la population, et on peut calculer par la formule suivante*

$$Q_1 = N^{-1} \nearrow \left(\frac{N}{4}\right) = F^{-1} \nearrow (0,25)\,.$$

Définition 32 ***Pour*** Q_2 *C'est le deuxième quartile, c'est lui qui prend la valeur de modalité qui donne la situation de la deuxième 25% de la population, et on peut calculer par la formule suivante*

$$Q_2 = N^{-1} \nearrow \left(\frac{N}{2}\right) = F^{-1} \nearrow (0,50)\,.$$

Définition 33 ***Pour*** Q_3 *C'est le troisième quartile, c'est lui qui prend la valeur de modalité qui donne la situation de la derinère 25% de la population, et on peut calculer par la formule suivante*

$$Q_3 = N^{-1} \nearrow \left(\frac{3N}{4}\right) = F^{-1} \nearrow (0,75)\,,$$

où $N^{-1} \nearrow$ *(*$F^{-1} \nearrow$*) represente la projection verticale de la valeur qui dans les parenthèses sur la colonne des modalités.*

Remarque 34 *On peut établir la valeur du premièr quartile (*Q_1*) graphiquement, en utilisant la courbe de la fonction de répartition, en faisant l la projection de la valeur (*$\frac{N}{4}$*) dans l'axe des cordonnées sur l'axe des abscices par la fonction de répartition. On peut établir la valeur du troisième quartile (*Q_3*) graphiquement, en utilisant la courbe de la fonction de répartition, en faisant l la projection de la valeur (*$\frac{3N}{4}$*) dans l'axe des cordonnées sur l'axe des abscices par la fonction de répartition.*

Définition 35 ***La moyenne*** *C'est la première valeur qui représente la caractérisation de la population, et on peut obtenir en utilisant la formule suivant*

$$\begin{aligned}\overline{x} &= \frac{1}{N}\sum_{i=1}^{N} x_i = \frac{1}{N}\sum_{i=1}^{k} n_i x_i \\ &= \sum_{i=1}^{k}\left(\frac{1}{N}n_i\right)x_i = \sum_{i=1}^{k} f_i x_i,\end{aligned}$$

avec N *c'est la somme total de la population, et* k *c'est le nombre des modalités.*

1.3.2 Paramètres caractéristiques de dispersion

Définition 36 ***Etendu*** *La longueur de la série statistique, et donné par la formule*

$$E = X_{\max} - X_{\min}.$$

Définition 37 ***Ecart-interquartiles*** *La longueur de la classe interquartile* $[Q_1, Q_3]$, *et donné par la formule*

$$I_Q = Q_3 - Q_1.$$

Définition 38 ***La variance*** *C'est la meilleur paramètre qui nous donne une vision générale sur la population, et donne par la formule*

$$\begin{aligned}
Var(X) &= \frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})^2 \\
&= \frac{1}{N}\sum_{i=1}^{k} n_i (x_i - \overline{x})^2 = \sum_{i=1}^{k}\left(\frac{1}{N}n_i\right)(x_i - \overline{x})^2 \\
&= \sum_{i=1}^{k} f_i (x_i - \overline{x})^2 = \sum_{i=1}^{k} f_i \left(x_i^2 + \overline{x}^2 - 2x_i\overline{x}\right) \\
&= \sum_{i=1}^{k} f_i \left(x_i^2\right) + \sum_{i=1}^{k} f_i \left(\overline{x}^2\right) - \sum_{i=1}^{k} f_i (2x_i\overline{x}) \\
&= \sum_{i=1}^{k} f_i \left(x_i^2\right) + \left(\overline{x}^2\right)\sum_{i=1}^{k} f_i - 2\overline{x}\sum_{i=1}^{k} f_i (x_i) \\
&= \sum_{i=1}^{k} f_i \left(x_i^2\right) + \left(\overline{x}^2\right) 1 - 2\overline{x}\overline{x} \\
&= \sum_{i=1}^{k} f_i \left(x_i^2\right) + \left(\overline{x}^2\right) - 2\overline{x}^2 = \sum_{i=1}^{k} f_i \left(x_i^2\right) - \left(\overline{x}^2\right).
\end{aligned}$$

Définition 39 ***Ecart-type*** *C'est le paramètre qui nous permet de savoir les relations entre les individus, et aussi de savoir le taux d'homogénéité, ainsi que le calcul du coefficient pour la forme, et on peut le calculer par la formule*

$$\sigma_X = \sqrt{Var(X)}.$$

Remarque 40 *La variance et l'écart-type sont des nombres positives ou nuls.*

Remarque 41 *Une variance nulle signifie que toutes les valeurs de la série sont égales à sa moyenne.*

Remarque 42 *Plus la variance d'une série est grange plus que cette série est dispersé autour de sa moyenne, et vice versa.*

Exemple 43 *Une étude sur le nombre d'enfants d'un échantillon de 50 ménages, noud donne*

Nombre d'enfants	*0*	*1*	*2*	*3*	*4*
Nombre des ménages	*20*	*15*	*10*	*5*	*0*

1) Quelle est la variable étudie ?.
2) Quelle est sa nature ?
3) Quelle échantillon étudie ?
4) Quelle sont les modalités ?
5) Calculer les paramètres de la position centrale.
6) Calculer les paramètres de dispersion.

Solution 44 *1) La variable édudie Le nombre d'enfants par ménage.*

2) Sa nature Quantitatif (mesurable) discret (elle ne prend que des valeurs isolées).

3) Echantillon étudie Les ménages, avec une taille de 50 ménages.

4) Les modalités Sont les valeurs qui peuvent prendre par le caractère, alors

$$Modalités = \{0, 1, 2, 3, 4\}.$$

5) Les paramètres de la position centrale.

Tout d'abord, on donne le tableau statistique, qui résume tous les données

X_i	n_i	f_i	$N\nearrow(ECC)$	$F\nearrow(FCC)$	f_ix_i	$f_ix_i^2=(f_ix_i)\,x_i$
0	*20*	$\frac{20}{50}=0,400$	*20*	*0,400*	*00*	*00*
1	*15*	$\frac{15}{50}=0,300$	*35*	*0,700*	*0,300*	*0,300*
2	*10*	$\frac{10}{50}=0,200$	*45*	*0,900*	*0,400*	*0,800*
3	*5*	$\frac{5}{50}=0,100$	*50*	*1,000*	*0,300*	*0,900*
4	*0*	$\frac{0}{50}=0$	*50*	*1,000*	*0,000*	*00*
Total	*50*	$\frac{20}{50}=1$	*/ / /*	*/ / / /*	*1*	*2*

Mode *: C'est la modalité la plus fréquente, ou bien c'est la modalité qui possède le plus grand effectif.*

$$Mo = 00.$$

Médiane *C'est la modalité qui partage l'échantillon en deux parties d'effectifs égaux, contenant le même nombre d'individus, et on peut calculer par la formule suivante*

$$Me = N^{-1} \nearrow \left(\frac{N}{4}\right) = F^{-1} \nearrow (0,25) = 1.$$

Les quartiles *Sont les modalités qui partagent l'échantillon en quarte parties d'effectifs égaux, contenant le même nombre d'individies.*

Pour Q_1 *C'est le premièr quartile, c'est lui qui prend la valeur de modalité qui donne la situation de la première 25% de la population, et on peut calculer par la formule suivante*

$$Q_1 = N^{-1} \nearrow \left(\frac{N}{4}\right) = F^{-1} \nearrow (0,25) = 00.$$

Pour Q_2 *C'est le deuxième quartile, c'est lui qui prend la valeur de modalité qui donne la situation de la deuxième 25% de la population, et on peut calculer par la formule suivante*

$$Q_2 = N^{-1} \nearrow \left(\frac{N}{2}\right) = F^{-1} \nearrow (0,50) = 1.$$

Pour Q_3 *C'est la troisième quartile, c'est lui qui prend la valeur de modalité qui donne la situation de la derinère 25% de la population, et on peut calculer par la formule suivante*

$$Q_3 = N^{-1} \nearrow \left(\frac{3N}{4}\right) = F^{-1} \nearrow (0,75) = 2.$$

La moyenne *C'est la première valeur qui qui represente le caracterisation de la population, et on peut obtenir en utilisant la formule suivant*

$$\overline{x} = \sum_{i=1}^{k} f_i x_i, = 1.$$

6) Les paramètres de dispersion.

Etendu *La longueur de la série statistique, et donné par la formule*

$$E = X_{\max} - X_{\min} = 4 - 0 = 4.$$

Ecart-interquartiles *La longuer de la classe interquartile* $[Q_1, Q_3]$, *et donné par la formule*

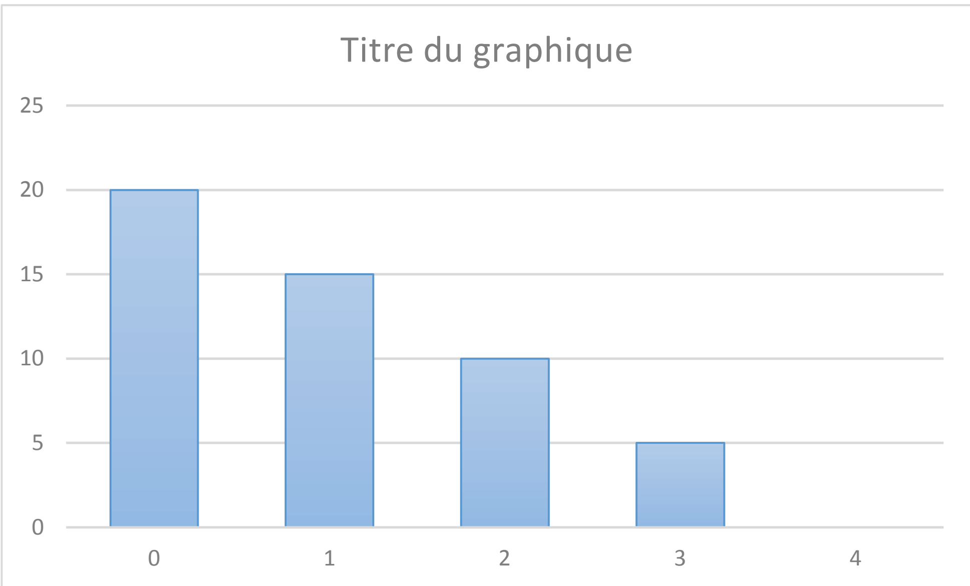
$$I_Q = Q_3 - Q_1 = 2 - 0 = 2.$$

La variance *C'est la meilleur paramatre qui nous donne une vesion generale sur la population, et donne par la formule*

$$\begin{aligned} Var(X) &= \frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})^2 \\ &= \sum_{i=1}^{k} f_i\left(x_i^2\right) - \left(\overline{x}^2\right) \\ &= 2 - (1)^2 = 1. \end{aligned}$$

Ecart-type *C'est le paramètre qui nous permet de savoir les relation entre les individus, et aussi de savoir le taux d'homogenèité, ainsi que le calcul de la coefficient pour la forme, et on peut le calculer par la formule*

$$\sigma_X = \sqrt{Var(X)} = \sqrt{1} = 1.$$



1.4 Série statistique continue

La série statistique continu est une méthode de regroupement des données sous forme des classes, mais malheureusement, le problème qui se pose ici c'est comment peut on regrouper les données sous des classes, pour cela on introduit deux méthodes majeures qui nous permettent de donner les nombres des classes qui necéssite pour cette regroupment.

Méthode de Yule

Il nous permet de calculer le nombre des classe, pour cela on utilise la formule suivante

$$\text{Nombre des classes}=1+3{,}3\log N,$$

où N c'est la taille totale d'échantillon.

Méthode de Sturge

Il nous permet de calculer le nombre des classe, pour cela on utilise la formule suivante

$$\text{Nombre des classes}=2{,}5\sqrt[4]{N},$$

où N c'est la taille totale d'échantillon.

Remarque 45 *Souvent les deux méthodes donnent la même nombre.*

Longueur de chaque classe Lorsque le nombre des classes existe, alors il est facile de savoir la longueur de chaque classe (les longueurs généralement sont égaux), pour cela on utilise la formule suivante

$$\text{longueur de chaque classe} = \frac{X_{\max} - X_{\min}}{\text{Nombre des classes}} = L.$$

1.4.1 Représentations graphiques

Pour les données statistiques forment une série statistique continues, on peut le représenter sous :

1) Histogramme : c'est une ensemble de rectangle, chaque rectangle est associe à une classe et a une surface proportionnelle à l'effectifs (ou fréquence) des classes.

2) Courbe de fonction de répartition : On place les points correspondantes aux extrimités de chaque classe avec ECC ou bien avec FCC, sur le graphe.

1.4.2 Paramètres caractéristiques de tendance centrale

Ces valeurs permettent de savoir autour de quelles valeurs se situent les valeurs d'une variable statistique continue.

<u>Mode</u> : C'est la classe la plus fréquente, ou bien c'est la classe qui possède le plus grand effectif, mais ici on peut voir seulent que des classes, alors pour calculer le mode qui se trouve dans la classe modale $[a, b]$, on utilise la formule

$$Mo = a + (b - a) \frac{\triangle I}{\triangle I + \triangle S},$$

où a : c'est la valeur minimale de la classe modale.

b : c'est la valeur maximale de la classe modale.

$\triangle I$: c'est la difference entre l'effectif de la classe modale et l'effectif de la classe qui précède.

$\triangle S$: c'est la difference entre l'effectif de la classe modale et l'effectif de la classe qui suivre.

<u>Médiane</u> C'est la modalité qui partage l'échantillon en deux parties d'effectifs égaux, contenant le même nombre d'individies, et on peut calculer par la formule d'interpolation linéaire, en appliquant le rôle de Thalès.

On remarque tout d'abord et d'aprés la figure ci dessous que si $0,5 \in [F(x), F(y)[$, alors $Me \in [x, y[$, donc on peut écrire

$$\frac{Me - x}{y - x} = \frac{0,5 - F(x)}{F(y) - F(x)}.$$

Alors

$$Me = x + (y - x) \frac{0,5 - F(x)}{F(y) - F(x)}.$$

Ou bien par ECC

$$Me = x + (y - x) \frac{\frac{N}{2} - N(x)}{N(y) - N(x)}.$$

<u>Pour Q_1</u> C'est la modalité qui se trouve entre le premier quartile et deuxième quartile, et on peut calculer par la formule d'interpolation linéaire, en appliquant le rôle de Thalès.

On remarque tout d'abord et d'aprés la figure ci dessous que si $0,25 \in [F(x), F(y)[$, alors $Q_1 \in [x, y[$, donc on peut écrire

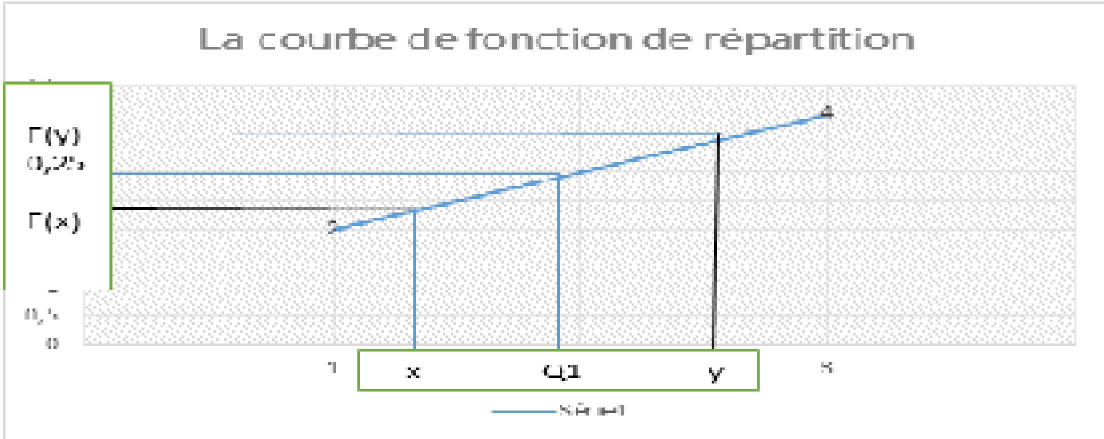
$$\frac{Q_1 - x}{y - x} = \frac{0,25 - F(x)}{F(y) - F(x)}.$$

Alors

$$Q_1 = x + (y - x) \frac{0,25 - F(x)}{F(y) - F(x)}.$$

Ou bien par ECC

$$Q_1 = x + (y - x)\,\frac{\frac{N}{4} - N(x)}{N(y) - N(x)}.$$



<u>Pour Q_3</u> C'est la modalité qui se trouve entre le troisième quartile et le quatrième quartile, et on peut calculer par la formule d'interpolation linéaire, en appliquant le rôle de Thalès.

On remarque tout d'abord et d'aprés la figure ci dessous que si $0,75 \in [F(x), F(y)[$, alors $Q_3 \in [x, y[$, donc on peut écrire

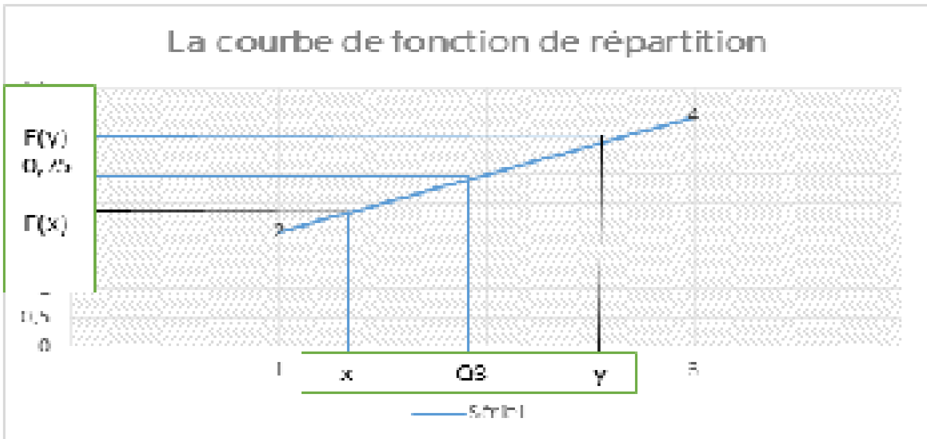
$$\frac{Q_3 - x}{y - x} = \frac{0,75 - F(x)}{F(y) - F(x)}.$$

Alors

$$Q_3 = x + (y - x)\,\frac{0,75 - F(x)}{F(y) - F(x)}.$$

Ou bien par ECC

$$Q_3 = x + (y - x)\,\frac{\frac{3N}{4} - N(x)}{N(y) - N(x)}.$$



La moyenne C'est la première valeur qui qui represente le caracterisation de la population, et on peut obtenir en utilisant la formule suivant

$$\overline{x} = \sum_{i=1}^{k} f_i c_i, ,$$

où $c_i = \frac{b_i + a_i}{2}$, où b_i et a_i sont designeent les bornes sup et inf des classes.

1.4.3 Paramètres caractéristiques de dispersion

Définition 46 ***Etendu*** *La longueur de la série statistique, et donné par la formule*

$$E = X_{\max} - X_{\min}.$$

Définition 47 ***Ecart-interquartiles*** *La longuer de la classe interquartile* $[Q_1, Q_3]$*, et donné par la formule*

$$I_Q = Q_3 - Q_1.$$

Définition 48 ***La variance****C'est la meilleur paramatre qui nous donne une vesion generale sur la population, et donne par la formule*

$$\begin{aligned}
Var(X) &= \frac{1}{N} \sum_{i=1}^{N} (c_i - \overline{x})^2 \\
&= \frac{1}{N} \sum_{i=1}^{k} n_i (c_i - \overline{x})^2 = \sum_{i=1}^{k} \left(\frac{1}{N} n_i \right) (c_i - \overline{x})^2 \\
&= \sum_{i=1}^{k} f_i (c_i - \overline{x})^2 = \sum_{i=1}^{k} f_i \left(c_i^2 + \overline{x}^2 - 2c_i \overline{x} \right) \\
&= \sum_{i=1}^{k} f_i \left(c_i^2 \right) + \sum_{i=1}^{k} f_i \left(\overline{x}^2 \right) - \sum_{i=1}^{k} f_i (2c_i \overline{x}) \\
&= \sum_{i=1}^{k} f_i \left(c_i^2 \right) + \left(\overline{x}^2 \right) \sum_{i=1}^{k} f_i - 2\overline{x} \sum_{i=1}^{k} f_i (c_i) \\
&= \sum_{i=1}^{k} f_i \left(c_i^2 \right) + \left(\overline{x}^2 \right) 1 - 2\overline{x}\overline{x} \\
&= \sum_{i=1}^{k} f_i \left(c_i^2 \right) + \left(\overline{x}^2 \right) - 2\overline{x}^2 = \sum_{i=1}^{k} f_i \left(c_i^2 \right) - \left(\overline{x}^2 \right).
\end{aligned}$$

Définition 49 ***Ecart-type*** *C'est le paramètre qui nous permet de savoir les relation entre les individus, et aussi de savoir le taux d'homogenèité, ainsi que le calcul de la coefficient pour la forme, et on peut le calculer par la formule*

$$\sigma_X = \sqrt{Var\left(X\right)}.$$

Remarque 50 *La variace et l'écart-type sont des nombres positives ou nuls.*

Remarque 51 *Une variance nulle signifie que toutes les valeurs de la série sont égales à sa moyenne.*

Remarque 52 *Plus la variance d'une série est grange plus que cette série est dispersé autour de sa moyenne, et vice vers sa.*

1.5 Paramètre de la forme

Une série à une distribution symétriques si ses valeurs sont également dispersés de la valeurs centrales, c'est à dire si le graphe de la distribution (Histogramme, ou bien diagramme en bâtons), admet une axe de symétrie.

Dans une distribution parfaitement symétrique

$$\overline{x} = Me = Mo.$$

Pour cela, on peut indéquer deux coefficients majeurs pour savoir la symétrie.

Coefficient d'asymétrie de Pearson On peut calculer en appliquant la formule

$$\delta = \frac{\overline{x} - Me}{\sigma_X}.$$

On a

$$-1 \leq \delta \leq +1.$$

Si $\delta = 0$, alors la symétrie est parfaite.

Si $\delta < 0$, alors la série étalée à gauche.

Si $\delta > 0$, alors la série étalée à droite.

Coefficient d'asymétrie de Yule On peut calculer en appliquant la formule

$$q = \frac{Q_3 + Q_1 - 2Me}{I_Q} = \frac{(Q_3 - Me) + (Q_1 - Me)}{I_Q}.$$

On a

$$-1 \leq q \leq +1.$$

Si $q = 0$, alors la symétrie est parfaite.

Si $q < 0$, alors la série étalée à gauche.

Si $q > 0$, alors la série étalée à droite.

1.6 Paramètre d'homogénéité

Coefficient da la variation Ce coefficient permet d'apprecier la représentativité de la moyenne par rapport à l'ensemble des observation. Il donne une bonne idée du dégré d'homogénéité d'une série, il faut qu'il soit le plus faible possible (<15% en pratique), et on peut l'obtenir en appliquant la formule suivante

$$CV = \frac{\sigma}{\overline{x}} \times 100.$$

Exemple 53 *On choisit les notes des étudiants en L1 biologie dans module Mathématiques,en 2019, comme suite*

2	*0*	*2*	*2*	*3*	*3*	*4*
1	*5*	*5*	*5*	*6*	*6*	*6*
6	*7*	*7*	*8*	*8*	*8*	*8*
8	*8*	*8*	*9*	*9*	*9*	*8*
5	*5*	*5*	*6*	*7*	*8*	*9*
10	*10*	*10*	*11*	*11*	*11*	*11*
14	*14*	*14*	*10*	*15*	*16*	*19*
20						

1) Quelle est la variable édudie ?.

2) Quelle est sa nature ?

3) Quelle echantillon étudie ?

4) Quelle sont les modalités ?

5) Calculer les paramètres de la position centrale.

6) Calculer les paramètres de dispersion.

7) Etudier la symétrie de la série.

8) Calculer le coefficient de la variation et étudier l'homogénéité de la série.

Solution 54 *1) La variable édudie : c'est les notes.*

2) Sa nature : quantitatif continu.

3) Echantillon étudie : c'est la section A

4) Les modalités : sont les classes, mais avant de déterminer les modalités il faut d'abord savoir le nombre des classes, ainsi que la longueur de chaque classe, pour cela on utilise la méthode de Sturge

$$\text{Le nombre des classes} = 2,5\sqrt[4]{50} = 6,64 \approx 7 \text{ classes.}$$

De plus

$$L = \frac{X_{\max} - X_{\min}}{\text{Nombre des classes}} == \frac{20-0}{7} = 3$$

Alors le tableau statistique c'est

Classe	*Eff*	*Centres*	*Freq*	*EEC*	*FCC*	f_ic_i	$f_ic_i^2 = (f_ic_i)\,c_i$
$[0,3[$	5	1,5	$\frac{5}{50} = 0,100$	5	0,100	0,150	0,225
$[3,6[$	9	4,5	$\frac{9}{50} = 0,180$	14	0,280	0,810	3,645
$[6,9[$	17	7,5	$\frac{17}{50} = 0,340$	31	0,620	2,550	19,125
$[9,12[$	12	10,5	$\frac{12}{50} = 0,240$	43	0,860	2,520	26,46
$[12,15[$	3	13,5	$\frac{3}{50} = 0,060$	46	0,920	0,810	10,935
$[15,18[$	2	17,5	$\frac{2}{50} = 0,040$	48	0,960	0,700	12,25
$[18,21[$	2	19,5	$\frac{2}{50} = 0,040$	50	1	0,780	15,21
Total	50	////	$\frac{50}{50} = 1$	///		8,320	87,850

Mode *: C'est la classe la plus fréquente, ou bien c'est la classe qui possède le plus grand effectif, alors pour calculer le mode qui se trouve dans la classe modale* $[a,b]$ *c'est a dire* $Mo \in [6,9]$, *on utilise la formule*

$$Mo = a + (b-a)\frac{\triangle I}{\triangle I + \triangle S},$$

où $a = 6.$

$b = 9.$

$\triangle I = 17 - 9 = 8.$

$\triangle S = 17 - 12 = 5$, *alors application numérique nous donne*

$$\begin{aligned} Mo &= a + (b-a)\frac{\triangle I}{\triangle I + \triangle S} \\ &= 6 + (9-6)\frac{8}{8+5} = 7,846. \end{aligned}$$

Médiane *On remarque tout d'abord que si* $0,5 \in [0,280, 0,620[$, *alors* $Me \in [6,9[$, *donc on peut écrire*

$$\begin{aligned} Me &= x + (y - x)\frac{0,5 - F(x)}{F(y) - F(x)} \\ &= 6 + (9 - 6)\frac{0,5 - 0,280}{0,620 - 0,280} \\ &= 7,941. \end{aligned}$$

Ou bien par ECC, c'est a dire si $\frac{N}{2} = \frac{50}{2} = 25 \in [14 - 31[$, *alors* $Me \in [6,9[$, *donc on peut écrire*

$$\begin{aligned} Me &= 6 + (9 - 6)\frac{25 - 14}{31 - 14} \\ &= 7,941. \end{aligned}$$

Les quartiles

Pour Q_1 *On remarque tout d'abord que si* $0,25 \in [0,100 - 0,280[$, *alors* $Q_1 \in [3,6[$, *donc on peut écrire*

$$\begin{aligned} Q_1 &= x + (y - x)\frac{0,25 - F(x)}{F(y) - F(x)} \\ &= 3 + (6 - 3)\frac{0,25 - 0,100}{0,280 - 0,100} \\ &= 5,499. \end{aligned}$$

Ou bien par ECC, c'est a dire si $\frac{N}{4} = \frac{50}{4} = 12,5 \in [4 - 14[$, *alors* $Q_1 \in [3,6[$, *donc on peut écrire*

$$\begin{aligned} Q_1 &= 3 + (6 - 3)\frac{12,5 - 5}{14 - 5} \\ &= 5,499. \end{aligned}$$

Pour Q_3 *On remarque tout d'abord que si* $0,75 \in [0,62 - 0,860[$, *alors* $Q_3 \in [9 - 12[$, *donc on peut écrire*

$$\begin{aligned} Q_3 &= x + (y - x)\frac{0,75 - F(x)}{F(y) - F(x)} \\ &= 9 + (12 - 9)\frac{0,75 - 0,620}{0,860 - 0,620} \\ &= 10,625. \end{aligned}$$

Ou bien par ECC, c'est a dire si $\frac{3N}{4} = \frac{150}{4} = 37,5 \in [31 - 43[$, *alors* $Q_3 \in [9 - 12[$, *donc on peut écrire*

$$\begin{aligned} Q_3 &= 9 + (12 - 9)\,\frac{37,5 - 31}{43 - 31} \\ &= 10,625. \end{aligned}$$

La moyenne

$$\overline{x} = \sum_{i=1}^{7} f_i c_i = 8,320.$$

5) Les paramètres de la position centrale.

Etendu *C'est la longueur de la série statistique*

$$\begin{aligned} E &= X_{\max} - X_{\min} \\ &= 20 - 0 = 20. \end{aligned}$$

Ecart-interquartiles*C'est la longueur de la classe interquartiles*

$$I_Q = Q_3 - Q_1 = 10,625 - 5,499 = 5,126.$$

Variance

$$Var\,(X) = \sum_{i=1}^{7} f_i c_i^2 - \overline{x}^2 = 87,850 - (8,320)^2 = 18.6276.$$

Ecart-type

$$\sigma_X = \sqrt{Var\,(X)} = \sqrt{18.6276} = 4.315.$$

7) ***Coefficient d'asymétrie de Pearson*** *On peut calculer en appliquant la formule*

$$\begin{aligned} \delta &= \frac{\overline{x} - Me}{\sigma_X} = \frac{8,320 - 7,941}{4.315} \\ &= 0,087 > 0. \end{aligned}$$

Alors la série étalée à droite.

Coefficient d'asymétrie de Yule *On peut calculer en appliquant la formule*

$$\begin{aligned} q &= \frac{Q_3 + Q_1 - 2Me}{I_Q} = \frac{10,625 + 5,499 - 2\,(7,941)}{5,126} \\ &= 0,047 > 0. \end{aligned}$$

Alors la série étalée à droite.

8) ***Coefficient da la variation*** *On peut l'obtenir en appliquant la formule suivante*

$$CV = \frac{\sigma}{\overline{x}} \times 100 = \frac{4.315}{8,320} \times 100 = 51,862\% > 15\%.$$

Alors la série n'est pas homogène.