Module de Biostatistiques,

Niveau Mster I, Biodiversite et environnement

Chapitre 1 : Statistique descriptive ;

Chapitre 2 : Statistique bivariée ;

Chapitre 3 : Loi normale ;

Chapitre 4 : Tests d’hypothese.

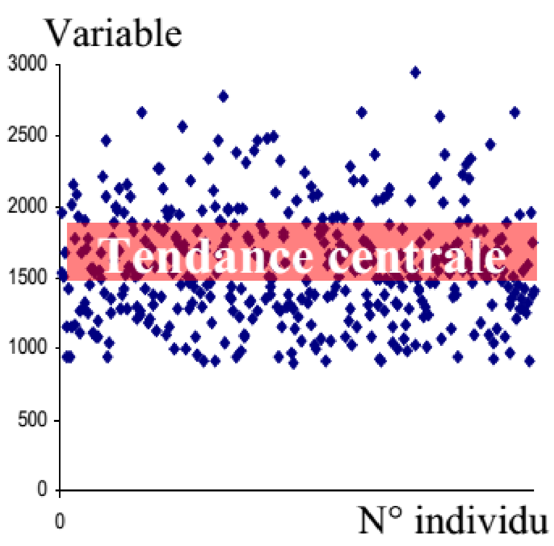
Chapitre 1 : statistiques descriptives

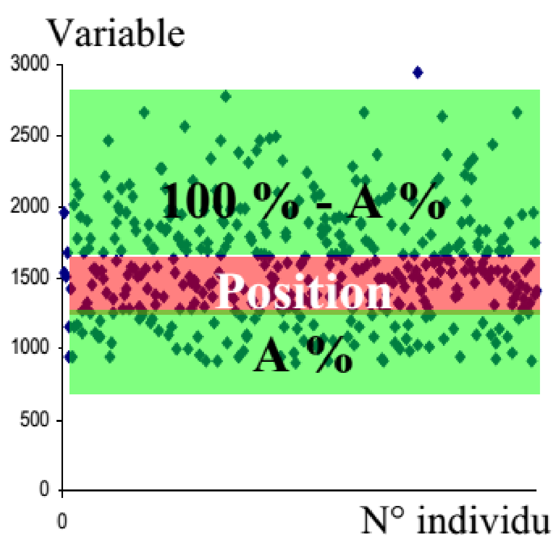
Rappel sur les statistiques descriptives (Voir Cous de M. Bouamra 2me année

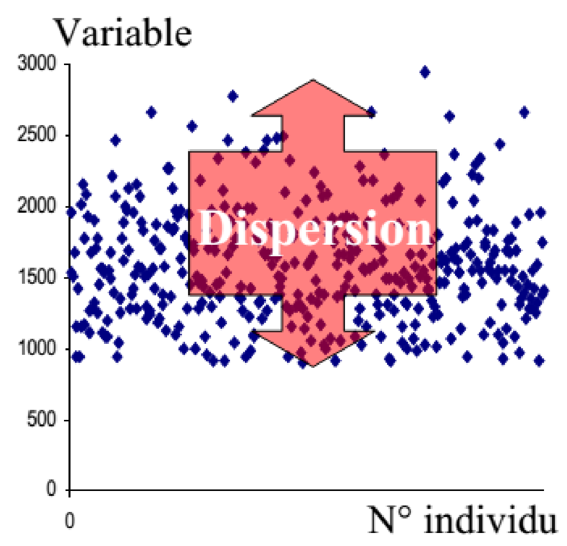
Suite des statistiques descriptives

Paramètres Statistiques

Les paramètres statistiques ne concernent que les variables quantitatives



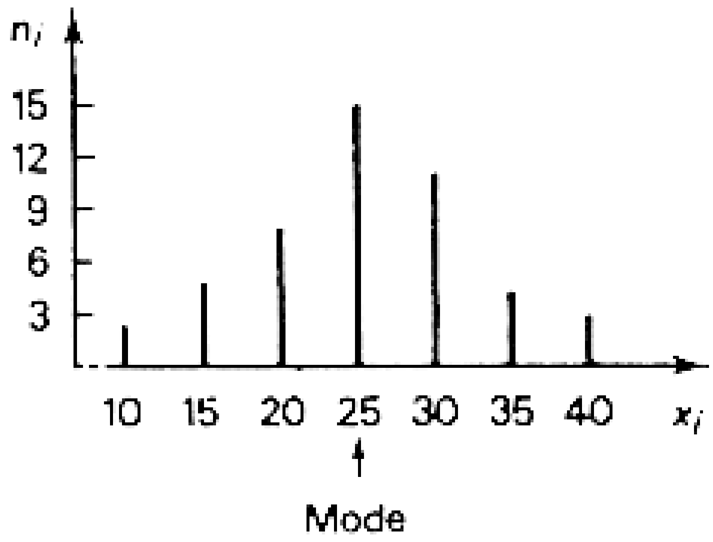


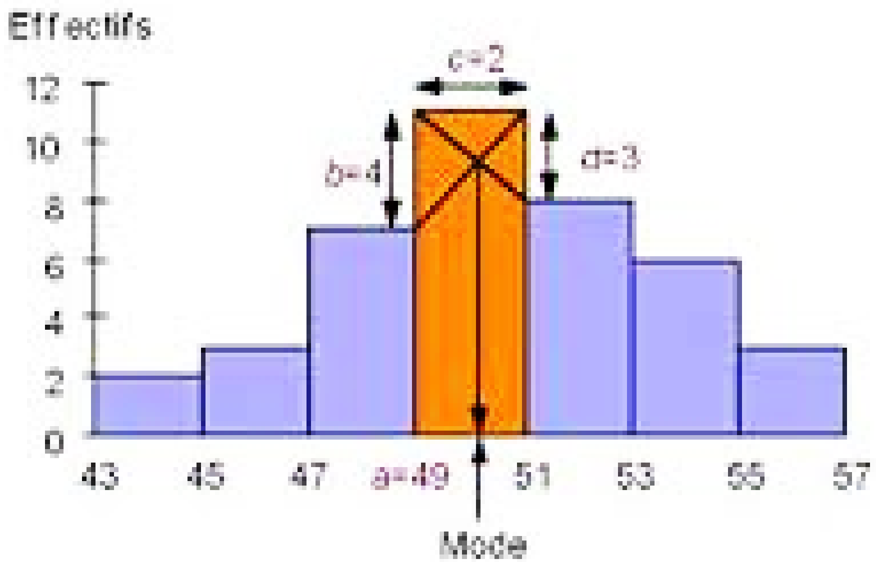


1- Paramètres de tendance centrale

1,1, Le mode

Une distribution est unimodale si elle présente un seul maximum, et la lecture s'exécute sur le diagramme en bâtons ou l'histogramme comme ci-après.
Le mode correspond a la valeur la plus frequente.





1,2, La médiane

Les valeurs observées doivent être rangées par ordre croissant.
La **médiane M** est la valeur du milieu de la série d'observations, c.à.d. elle divise la serie en deux séries égales.

Nombre impair d'observations

$$\underbrace{3\quad 4\quad 4\quad 5}_{\text{4 valeurs}}\quad \underset{\uparrow\atop \mathbf{M}}{6}\quad \underbrace{8\quad 8\quad 9\quad 10}_{\text{4 valeurs}}$$

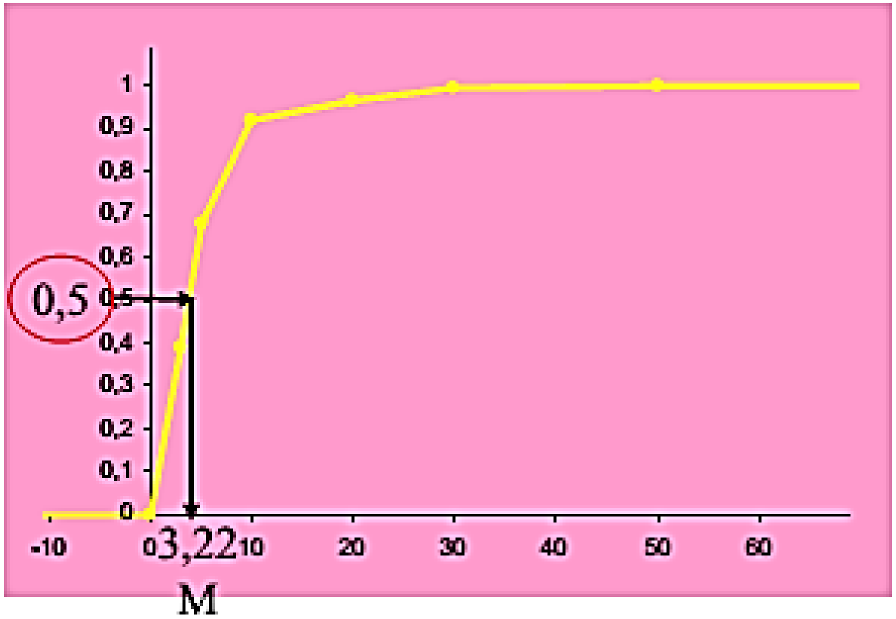
Nombre pair d'observations

$$\underbrace{3\quad 4\quad 4\quad 5}_{\text{4 valeurs}}\quad \underbrace{6\quad 8\quad 8\quad 9}_{\text{4 valeurs}}$$

Intervalle médian

M = milieu = 5,5

LA MEDIANE à partir d'une distribution continue



1,3, La moyenne arithmétique

La moyenne arithmétique d'une série statistiques est la somme des valeurs divisée par le nombre de valeurs

$$\bar{x} = \frac{\sum_{i=0}^{k} ni\ xi}{n}$$

K= 0;1;2; 5;4;5;6
n=150

$$\bar{x} = \frac{(11*0)+(1*22)+\cdots+(6*2)}{150} = 2{,}5$$

Moyenne de plusieures populations :

$$\bar{\bar{x}} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n} = \sum_{i=1}^{k} \frac{n_i\bar{x}_i}{n}$$

Moyenne globale = moyenne des moyennes

Moyenne géométrique

$$G = \sqrt[n]{x_1^{n_1} x_2^{n_2} \ldots\ldots x_k^{n_k}}$$

Utilisée dans le cas de phénomènes multiplicatifs (taux de croissance moyen)

Moyenne harmonique

$$H = \frac{n}{\sum_{i=1}^{k} \frac{n_i}{x_i}}$$

Utilisée dans le cas où l'on combine 2 variables sous forme de rapport (pièces/heure, km/litre,…)

2, Paramètres de position

On appelle fractiles ou quantiles d'ordre k les (k-1) valeurs qui divisent les observations en k parties d'effectifs égaux.

1 médiane M qui divise les observations en 2 parties égales

3 quartiles Q_1, Q_2, Q_3 qui divisent les observations en 4 parties égales

9 déciles D_1, D_2, …, D_9 qui divisent les observations en 10 parties égales

99 centiles C_1, C_2, …, C_{99} qui divisent les observations en 100 parties égales

3, Paramètres de dispersion

3,1, Intervalle interquartile

Les étapes :

1- Classement des données par ordre croissant ;
2- Trouver la médiane ;
3- Trouver la médiane des deux nouvelles séries :
La Me de la 1ere série ------------- 1er quartile **(Q1)**
La Me de la 2ere série ------------- 2e quartile **(Q3)**
4- Calculer de IQ par : **IQ = Q3 – Q1**,

Exemple 1

4, 13, 17, 7, 1, 3, 9, 14, 12, 20, 16, 15, 11, 6, 5

Exemple 2

4, 13, 4, 13, 17, 7, 15, 7, 16, 9, 6, 7, 1, 3, 19, 14, 1, 1, 12, 11, 20, 16, 15, 11, 6, 11,

Remarque

On peut rapporter l'IQ à l'intervalle de variation

$$\frac{IQ}{Intervalle\ de\ variation} * 100$$

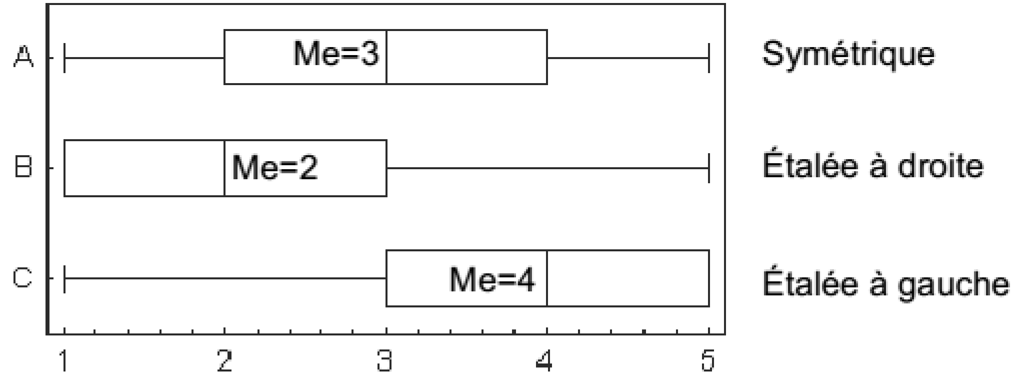
Représentation graphique

Exemple

A = { 1, 1, 2, 2, 2, 2, 3, 3, 3,3,3, 4, 4, 4, 4, 5, 5 }

B = { 1, 1, 1, 1, 1, 1, 1, 1, 1,1, 2, 2, 2, 2, 2, 2, 2, 2, 3, 3, 3, 3,3 , 3, 4, 4, 4, 4, 5, 5 }

C = {1, 1, 2, 2, 2, 2, 3, 3, 3,3, 3, 3, 4, 4, 4, 4, 4, 4, 4, 4, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5 }



Boite à moustache

3,2, Variance

Soit une série de valeurs d'une variable X : $\{x_1, x_2, \ldots, x_k\}$. Soit les effectifs associés : $\{n_1, n_2, \ldots, n_k\}$. La variance de cette série s'écrit :

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{k} n_i (x_i - \bar{x})^2$$, si l'effectif considéré est celui d'une population. (1)

$$\sigma^2 = \frac{1}{n-1}\sum_{i=1}^{k} n_i (x_i - \bar{x})^2$$, si l'effectif considéré est celui d'un échantillon. (2)

Comment faire la variance de plusieurs populations ?

$$V = \frac{1}{n}\sum_{i=1}^{k} n_i V_i + \frac{1}{n}\sum_{i=1}^{k} n_i \left(\bar{x}_i - \bar{\bar{x}}\right)^2$$

Variance globale = Moyenne des variances + Variance des moyenne

3,3, Écart-type

L'écart-type est égal à la racine carrée de la variance :

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{k} n_i \left(x_i - \bar{x}\right)^2} = \sqrt{\frac{1}{n}\sum_{i=1}^{k} n_i x_i^2 - \bar{x}^2}$$

3,4, Coefficient de variation

$$CV = \left(\frac{\sigma}{\bar{x}}\right) \times 100$$

3,5, Etendue

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

Merci de votre attention

Prochain cours: statistique bivariée

Université Mohand Akli Ouelhadj Bouira
Faculté des Sciences de la Nature et de la Vie et des Sciences de la Terre
Département de Biologie

Exercice 1.

On s'intéresse à l'effet d'une faible dose faible de cambendazole sur les infections des lapins par *Trichinella Spiralis*. 20 lapins ont été infectés par un même nombre (916) de larves de Trichinella. Les lapins ont reçu du cambendazole, à raison du 20 mg par Kg, 60 heures après l'infection. Au bout d'une semaine, tous les lapins ont été sacrifiés et les nombres suivants de vers adultes ont été retrouvés dans les intestins :

80 75 80 69 75 80 96 102 80 96 102 80 96 75 80 96 75 80 69 80

1. Etablir le tableau de distribution statistique
2. Représenter graphiquement la distribution du nombre de vers adultes, en déduisant le mode. Représenter sur le même graphe le polygone des effectifs.
3. Calculer : la moyenne, le mode, la médiane, l'étendue, la variance et l'écart-type.

Exercice 2.

Une étude réalisée dans 50 champs d'orange a donné les résultats suivants concernant le nombre de plantes souffrantes des pucerons.

30	30	38	50	64	42	70	60	64	42
74	60	64	64	50	64	42	64	50	42
60	42	64	50	64	50	60	50	38	64
64	70	74	64	64	60	50	74	74	70
64	30	42	74	74	70	64	60	30	50

1. Donner le tableau statistique et la représentation graphique
2. Représenter graphiquement les fréquences cumulées croissantes et décroissantes
3. Calculer le mode, la médiane, les quartiles, la moyenne, la variance, l'écart type ainsi que le coefficient de variation.

Exercice 3.

Un chercheur veut étudier la valeur biologique d'une protéine sur un groupe de 50 souris. Il a introduit 50g de cette protéine dans leur alimentation pendant un mois puis il a constaté des prises de poids des souris comme suit :

0.175 0.190 0.213 0.233 0.244 0.250 0.268 0.273 0.298 0.334 0.348 0.350 0.356 0.359 0.370 0.390 0.416 0.424 0.434 0.446 0.447 0.452 0.460 0.472 0.478 0.493 0.510 0.518 0.524 0.537 0.539 0.543 0.550 0.580 0.599 0.616 0.628 0.637 0.644 0.656 0.667 0.677 0.680 0.690 0.721 0.736 0.742 0.780 0.798 0.812.

1. Quelles sont les différentes notions statistiques exprimées dans cette série
2. Quel est le type du caractère étudié
3. Déterminer le tableau de la distribution statistique et tracer le graphe correspondant
4. Calculer la moyenne, le mode et l'écart-type de cette série.

Université Mohand Akli Ouelhadj Bouira
Faculté des Sciences de la Nature et de la Vie et des Sciences de la Terre
Département de Biologie

Exercice 4.

Des étudiants du Master Biochimie Appliquée étudient les œufs d'une ferme et les facteurs influençant leurs durées de vie. Deux études ont étés proposées : observer le comportement du poids des œufs et étudier l'effet de la température sur leurs durées de vie.

I. Pour les aider, on a pesé un échantillon de 400 œufs de la ferme (les masses des œufs sont exprimées en grammes). Les résultats sont résumés dans le tableau suivant :

Masse de l'œuf (X)	[25 35[	[35 45[	[45 55[	[55 60[	[60 70[	[70 80[
Nombre d'œufs	3	51	186	92	62	6

1. Identifier l'individu, le caractère statistique ainsi que sa nature
2. Représenter la série et calculer le poids le plus fréquent
3. Représenter la courbe cumulative croissante puis calculer le poids médian
4. Calculer le poids moyen de cette série et son écart-type
5. On admet que le prix de vente d'un œuf est une variable statistique $Y = 0.08\,X + 5$, X étant la masse de l'œuf. Déduire le prix moyen de vente de l'œuf.