UNIVERSITE D'ALGER /FACULTE DE MEDECINE /
DEPARTEMENT DE MEDECINE « ZIANIA »

MODULE D'EPIDEMIOLOGIE

INTRODUCTION AUX STATISTIQUES « STATISTIQUES DESCRIPTIVES »

6ème ANNEE MEDECINE/A6

ANNEE UNIVERSITAIRE: 2021-2022

Cours rédigé et présenté par Dr.BENELFEKIR.H
Maitre assistante en Epidémiologie et Médecine Préventive/CHU Hussein Dey

OBJECTIFS DU COURS

L'étudiant en 6ème année de médecine doit savoir:

1) définir les différents termes constituant le vocabulaire élémentaire de la statistique.
2) déterminer le type, la nature et les modalités d'une variable ou d'un caractère.
3) construire le tableau de distribution des fréquences selon la nature de la variable
4) calculer les différentes fréquences constituant ce tableau.
5) représenter graphiquement les données selon la nature de la variable.
6) calculer les différents paramètres de réduction.
7) interpréter la relation entre les paramètres de tendance centrale et ceux de dispersion.

PLAN DU COURS

I. Définition préliminaires:
   A) Population , Echantillon
   B) Variable ou caractère
   C) Fréquences: absolue, relative et cumulée
   D) Rapports, propositions, taux et indices

II. Mise en ordre des données: tableau de fréquence

III. Représentation graphique des données

IV. Paramètres de réduction
   A) Paramètres de tendance centrale
   B) Paramètres de dispersion
   C) Coefficient de variation
   D) Quantiles

I. Définitions de certains termes : vocabulaire élémentaire de la statistiques

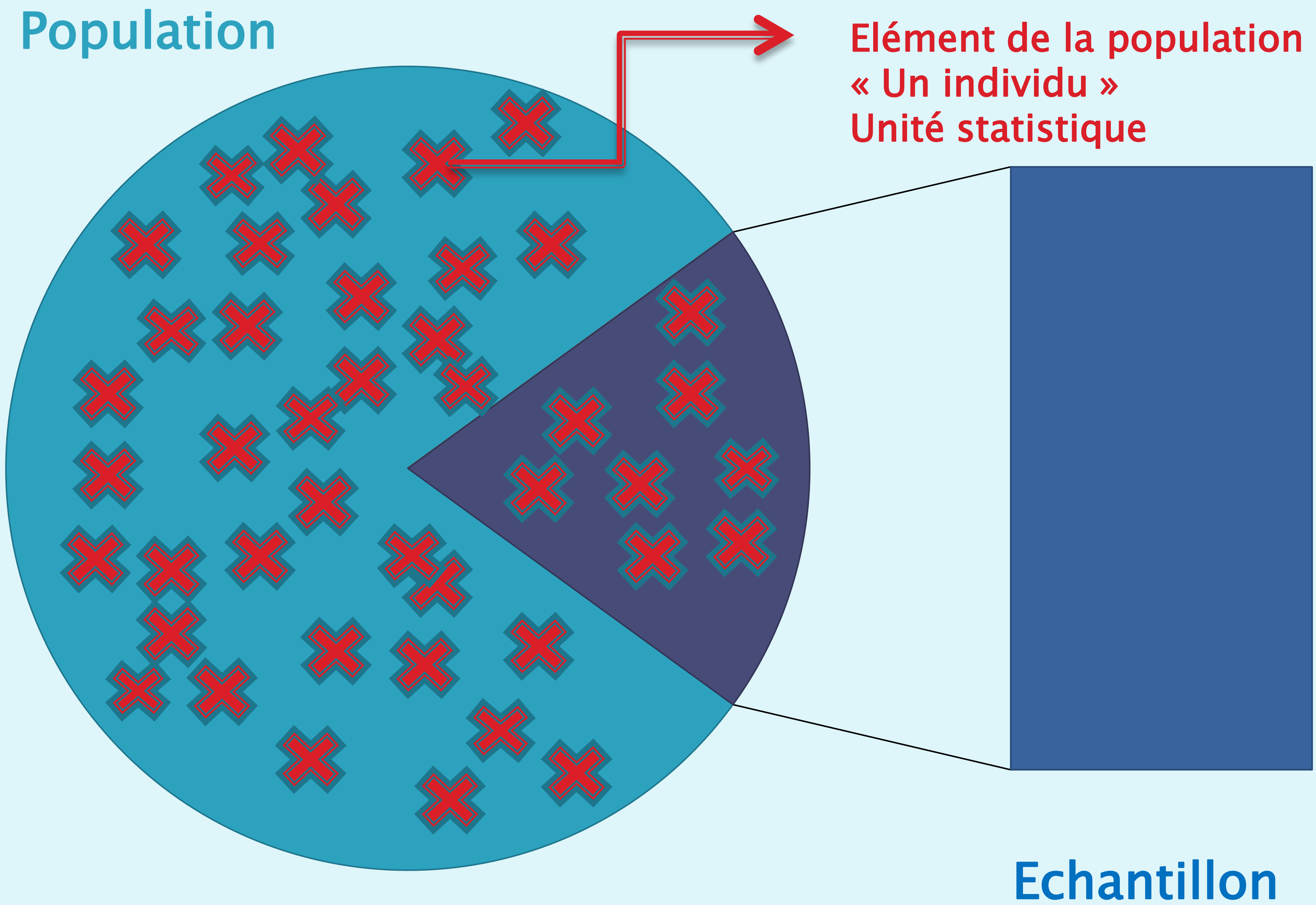
1) Population , Echantillon:

1-1) Population: est un ensemble d'éléments sur les quels va porter l'étude statistique. Chaque élément de la population est appelé **unité statistique** ou **individu.**

- **Exemple de population statistique:** le personnel enseignant de la faculté de médecine d'Alger.
- **Unité statistique**: l'enseignant de la faculté de médecine.

1-2) Echantillon : est sous ensemble d'une population. C'est une partie restreinte de la population.

- L'opération de choix de l'échantillon est appelée **échantillonnage** ou **sondage.**
- Le nombre d'éléments constituant l'échantillon est appelé t**aille de l'échantillon, noté N/n.**
- **Exemple d'échantillon**: un groupe du personnel enseignant de la faculté de médecine d'Alger(ex: enseignant du module d'épidémiologie)


Population
Elément de la population
« Un individu »
Unité statistique
Echantillon

2) Variable ou caractère, Modalités:

2-1) Caractère / variable: est une propriété possédée par les unités statistiques, permettant de les décrire et les distinguer les uns des autres.

Exemples de caractère: poids, taille, âge, nationalité, maladie etc.

2-2) Modalités: sont les différentes situations susceptibles d'être prises par le caractère ou la variable. Un caractère peut présenter deux ou plusieurs modalités. L'ensemble des modalités observées sur les unités statistiques constitue les **données** ou **observations**.

Exemples:

- sexe: masculin , féminin;
- Maladie: malade , non malade;
- Exposition : (exposé , non exposé;
- Diarrhée: diarrhée présente, diarrhée absente;
- Groupe sanguin dans le système ABO : O, A, B, AB.

2-3) Nature de variables: les caractères sont classés en deux catégories: les caractères qualitatifs et les caractères quantitatifs

2-3-1) Variable qualitative: c'est un caractère dont les modalités s'expriment par des qualités et non plus par des valeurs numériques ou mesurables.

- Les caractères présentant deux modalités sont appelés: **caractères dichotomiques ou binaires.**

Exemple: sexe: masculin , féminin

- Les variables comportant plus de deux modalités sont appelées **multichotomiques**: on distingue:
  - ✓ **les variables qualitatives ordinales**: ce sont des variables multichotomiques dont les modalités sont soumises à un ordre ou à une progression naturelle.

  Exemple: stade de la maladie de Hodgking : stades I à IV .
  - ✓ **les variables qualitatives nominales:** ce sont des variables multichotomiques dont les modalités ne sont pas soumises à un ordre

  Exemples: groupe sanguin dans le système ABO (O, A, B, AB) ; statut matrimonial (célibataire, marié, divorcé, veuf).

2-3-2) Variable quantitative: c'est un caractère dont les modalités s'expriment par des valeurs numériques ou mesurables; on distingue :

- **la variable quantitative continue** :a des modalités en nombre infini qui se situent à un point quelconque d'une échelle numérique.

Exemples: Glycémie, heure du jour, âge , taille, tension artérielle, taux d'urée sanguine, taux de cholestérol

- **la variable quantitative discontinue ou discrète:** a des modalités qui s'expriment par des nombres entiers;

Exemples: nombre d'enfants par famille, fréquence cardiaque , nombre de globules par unité de volume sanguin, nombre d'accidents de la route recensés, etc.

3) Fréquences: absolue, relative et cumulée:

3-1) Fréquences simples:

3-1-1) Fréquence absolue (effectif) « ni »: est le nombre d'individus correspondant à une modalité donnée d'une variable.

Exemple: distribution de 50 malades selon le sexe. Parmi ces 50 malades, 15 sont de sexe masculin et 35 de sexe féminin. Les effectifs correspondant à chacune des deux modalités sont 15 et 35.

3-1-2) Fréquence relative (pourcentage) « fi »: est le rapport entre l'effectif d'une modalité de variable et l'effectif total de la série sur laquelle cette variable est mesurée. Le numérateur fait obligatoirement partie du dénominateur. La fréquence relative s'exprime généralement en pourcentage .

Exemple: dans l'exemple précédent:

- FR des sujets de sexe masculin est: 15/50=0.30= 30%
- FR des sujets de sexe féminin est : 35/50 = 0.70 = 70.0%.
- La somme des fréquences relatives est égale à 1 ou 100%.

3-2) Effectifs et fréquences cumulés:

3-2-1) Effectif cumulé « nic »: peut concerner deux ou plusieurs modalités complémentaires et incompatibles d'une même variable.

Exemple: distribution de 50 malades selon le sexe : les cumulés des deux modalités : masculin « ni1=15 » et féminin « ni2=35 » est la somme des effectifs simples (N=50).

3-2-2) Fréquence relative cumulée (Probabilité cumulée)«fic» : qui peut concerner deux ou plusieurs modalités complémentaires et incompatibles d'une même variable. la probabilité cumulée s'obtient par la somme de fréquences relatives simples.

Exemple : la probabilité pour un malade d'être, soit un sujet de sexe masculin, soit un sujet de sexe féminin est évidemment de 100%.

4) Rapports: ratio, propositions, taux et indices:

4-1) Ratio:
est une expression de la relation entre deux entités ou encore le rapport des effectifs de deux modalités d'une même variable. Il est de la forme suivante: x : y ou x / y * k

- Le sex-ratio est le rapport numérique des sexes :

  SR= (effectif de sexe masculin) / (effectif de sexe féminin)

Ex1: on a trouvé que dans un échantillon il y'avait 4000 hommes et 2000 femmes. Le rapport homme: femmes est de 2:1 (2/1). Quant au rapport femmes : hommes est de 0,5 :1 avec k=1

Ex2: rapport des sexes à la naissance: si le sex-ratio est égal à 1.06, cela signifie que pour 100 naissances féminines, il y a 106 naissances masculines.

4-2) Proportion:
est un rapport dont le numérateur fait partie du dénominateur, tous deux étant mesurés simultanément. Elle est de la forme: x / x+y * 100 . Une proportion s'exprime en pourcentage

Exemple: la proportion d'hommes dans l'échantillon est: [4000/(2000+4000)] * 100 = 66,6%

4-3) Taux: mesure la probabilité de survenue d'un évènement donné au cours du temps. C'est une expression de la forme: $x / y * k$

Il doit toujours s'exprimer en fonction d'une certaine unité de temps, pour un lieu géographique donné et pour un groupe de personnes bien défini :

- Le numérateur est un nombre d'évènements (décès, maladie, handicap) survenus au cours d'une certaine période.
- Le dénominateur représente la population exposée au risque de survenue de cet événement pendant cette période.
- La constante k est une puissance de 10. On choisira k de telle sorte que le taux comporte un à deux chiffres avant la virgule.

Exemple: soit un taux de 0,00036/an, la constante k = 10 000 donc le taux est 3,6 pour 10 000/an

4-4) Indice: est un paramètre de mesure servant à estimer un taux dans le cas où le dénominateur de ce dernier ne peut être mesuré.

- Il peut être le rapport de deux effectifs relatifs aux données de deux variables quantitatives discrètes, il est notamment utilisé en économie de la santé comme indicateur de fonctionnement

Exemple: nombre de lits d'hôpital/nombre de médecins d'hôpital

II. Mise en ordre des données

- La mise en ordre des données est l'étape qui suit immédiatement celle du recueil des valeurs de la (ou des) variable(s) étudiée(s).
- Elle consiste à dresser un tableau qui fait correspondre aux valeurs ou qualités de la variable prise en considération, le nombre d'individus présentant effectivement ou ces qualités.
- Le tableau s'appelle tableau d'effectifs ou de distribution de fréquences.

Tableau de distribution des fréquences:

Variable	Fréquence absolue / effectif/ ni	Fréquence relative %/fi
M1	ni1	fi1=ni1/N
M2	ni2	fi2=ni2/N
M3	ni3	fi3=ni3/N
Total	N =ni1+ni2+ni3	100% =fi1+fi2+fi3

Tableau de distribution des fréquences: ffectif, fréquence relative, effectif cumulé et fréquence cumulée

Variable	Fréquence absolue / effectif/ ni	Effectif cumulé effcum	Fréquence relative %/fi	Fréquence relative cumulée/ ficum
M1	ni1	ni1	fi1=ni1/N	fi1
M2	ni2	ni1+ni2	fi2=ni2/N	fi1+fi2
M3	ni2	ni1+ni2+ni3 =N	fi3=ni3/N	fi1+fi2+fi3 =100%
Total	N =ni1+ni2+ ni3	//	100% =fi1+fi2+ fi3	//

Tableau de fréquence pour variable qualitative: Répartition de 120 malades selon le groupe sanguin dans le système ABO:

Groupe sanguin	Fréquence absolue / effectif/ ni	Fréquence relative %/fi	Fréquence relative cumulée/ ficum
O	50	41,7%	41,7%
A	40	33,3%	75,0%
B	17	14,2%	89,2%
AB	13	10,8%	100%
Total	120	100%	

Tableau de fréquence pour variable quantitative discrète: distribution du nombre d'épisodes grippal parmi 19 personnes :

Nombre d'épisodes	Fréquence absolue / effectif/ ni	Effectif cumulé nicum	Fréquence relative %/fi	Fréquence relative cumulée/ ficum
0	3	3	15,8%	15,8%
1	7	10	36,8%	52,6%
2	6	16	31,6%	84,2%
3	2	18	10,5%	94,7%
4	1	19	05,3%	100%
Total	19		100%	

Tableau de fréquence pour variable quantitative continue: répartition du poids de 19 étudiants:

Poids	Centre de classe/ xi	Fréquence absolue / effectif/ ni	Fréquence relative %/fi
[50-55[	52,5	1	5,3%
[55-60[	57,5	2	10,5%
[60-65[	62,5	5	26,3%
[65-70[	67,5	4	21,2%
[70-75[	72,5	3	15,8%
[75-80[	77,5	3	15,8%
[80-85[	82,5	1	05,3%
Total		19	100%

Ex: soit le série du poids de 19 étudiants en Kg:

76,340; 64,990; 83,450; 69,120; 59,990; 66,330; 52,990; 60,400; 79,650; 61,820; 70,560; 68,280; 64,100; 61,820; 70,130; 57,740; 72,880; 76,360; 65,450.

- Avant de déterminer les classes et leurs effectifs, on doit classer les données par ordre croissant:

  52,990; 57,740; 59,990; 60,400; 61,820; 61,820; 64,100; 64,990; 65,450; 66,330; 68,280;69,120; 70,130; 70,560; 72,880;76,340; 76,360; 79,650; 83,450.

- Amplitude de classe: Etendu/Nombre de classe
- Nombre de classe: $P=\sqrt{N}$ ou
  - La règle de *STURGE* : $P= 1+ (3,3 \log N)$
  - La règle de *YULE* : $P= 2,5\sqrt[4]{N}$

$P= 1+ (3,3 \log 19) =5,22\approx 6$classes

- L'étendu (ou marge) est la différence entre la valeur la plus grande et la valeur la plus faible de la série:

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

E= 83,45 – 52,990= 30,460 kg

- Donc l'amplitude de classe est: a=30,46 /6= 5

- Donc la classe sera déterminée comme suit:

$L_S= L_I + a$, $[X_{LI}-X_{LS}[$

ainsi la première classe est: 52,990+5=57,990

[52,990–57,990[

- Nombre de classes= **étendu/amplitude**
- L'amplitude de classe est la différence entre les limites inferieure et supérieure de la classe.

$$a = L_s - L_I$$

Pour la classe 1: **a= 55 - 50= 5**

Donc le nombre de classe est:

P=30,46 / 5= 6 ,1≈ 7classes

- Le centre de classe est la valeur située au milieu de la classe, il est égal à :

$$x_i = (L_{inf} + L_{sup}) / 2$$

Le centre de classe de la première classe est ainsi égal à :

xi= 50 + 55/2= 52,5

III. Représentation graphique

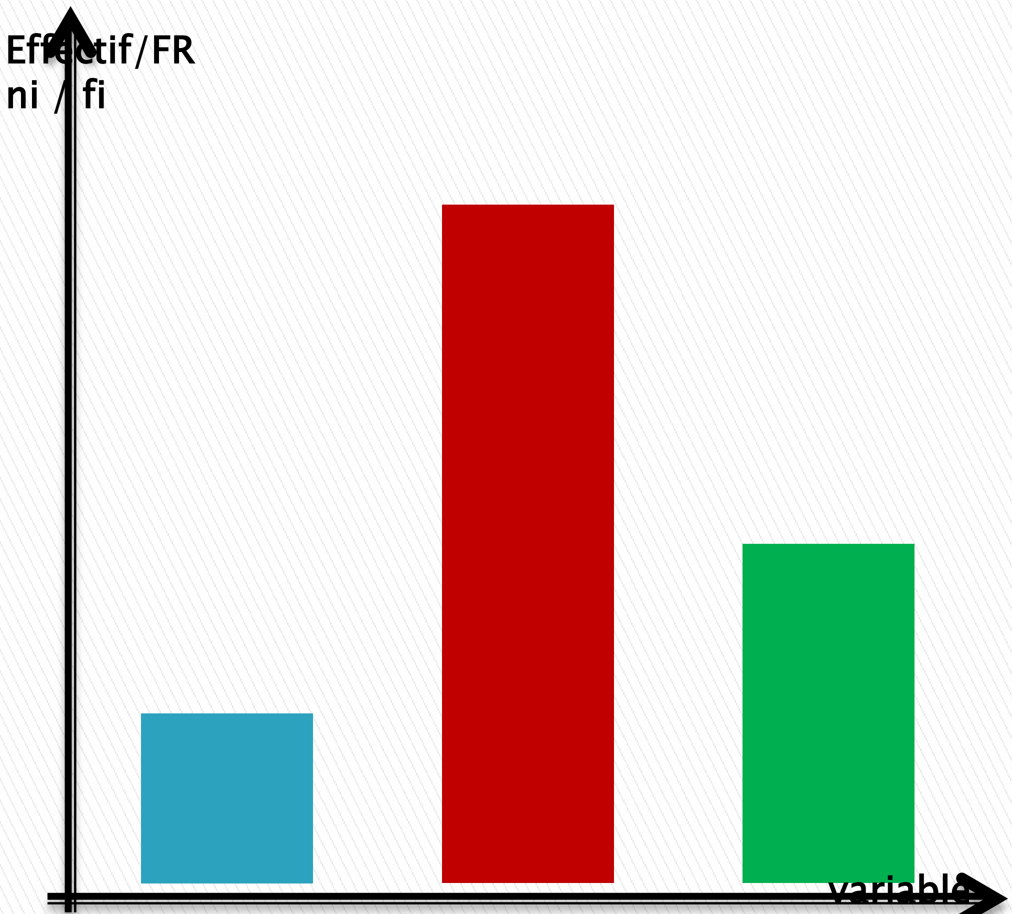
- Elle permet de saisir rapidement et facilement les grands traits d'une distribution.

- Il existe deux types de graphiques: les graphiques à échelles semi–logarithmiques et les graphiques à échelles **arithmétiques**.

- La représentation graphique va dépendre de la nature de la variable.

1) Représentations graphiques pour la variable qualitative:

Graphe en tuyau d'orgue (en barres)	Graphe en secteur (camembert)



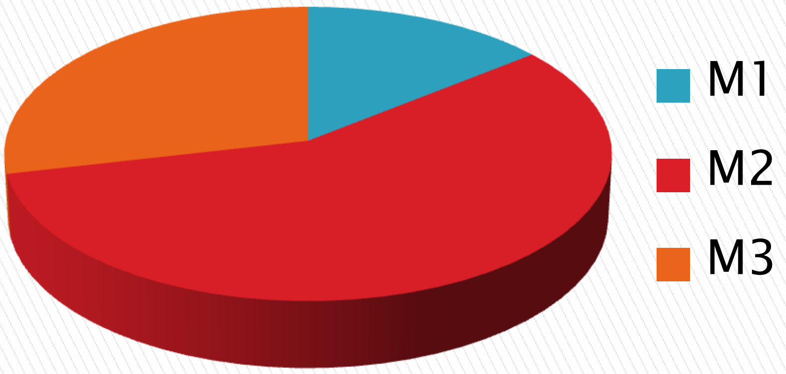


Diagramme en barres: Répartition de 120 malades selon le groupe sanguin dans le système ABO:

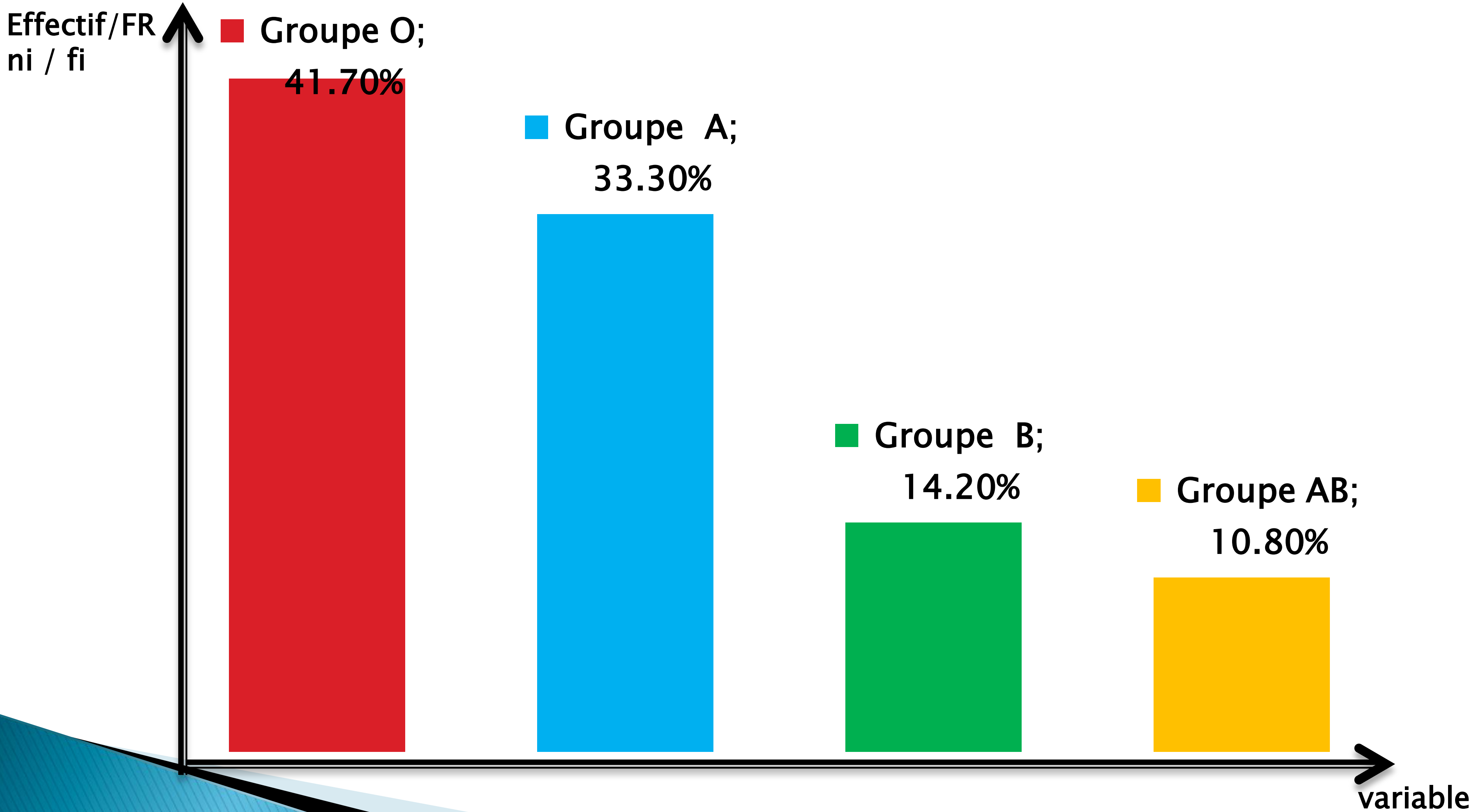
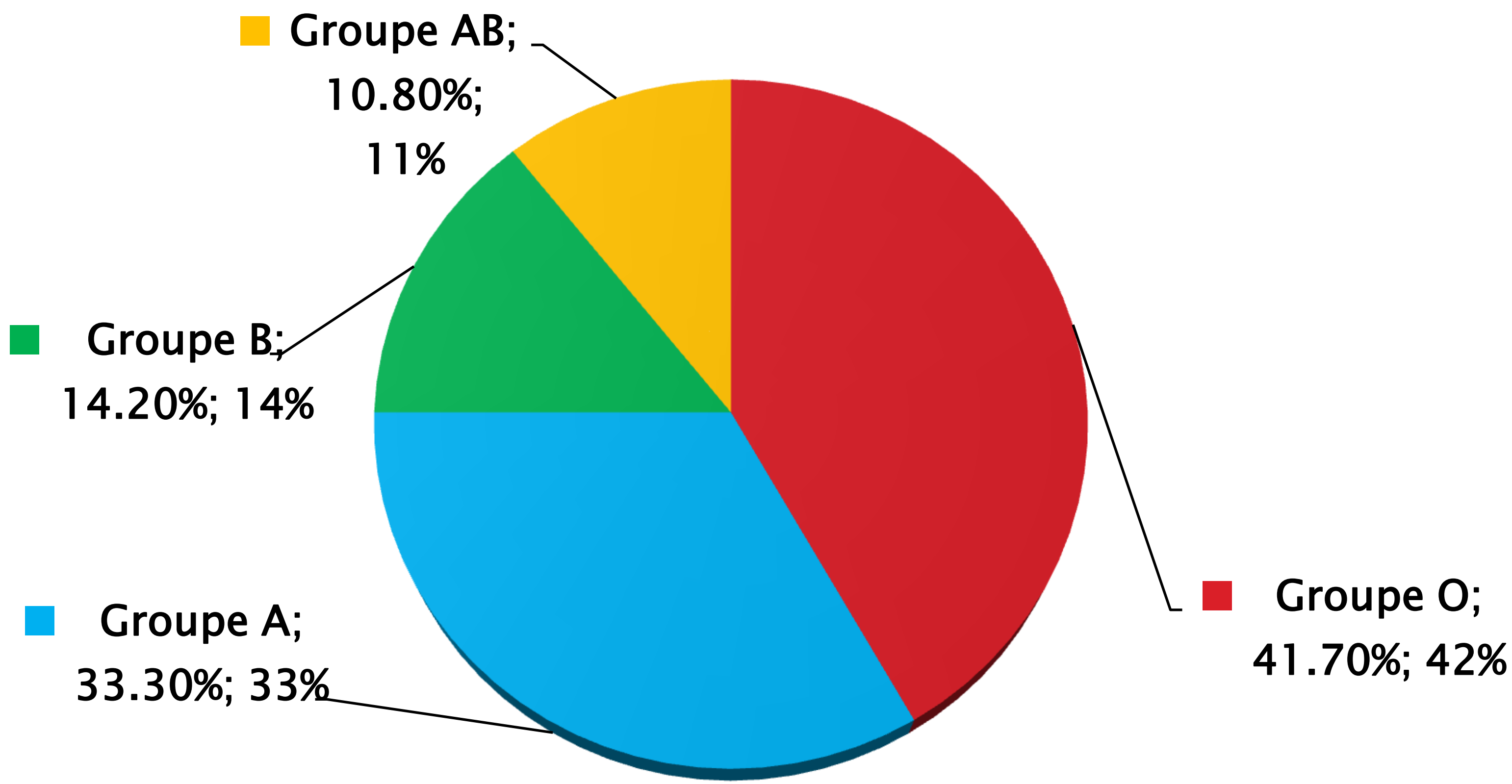
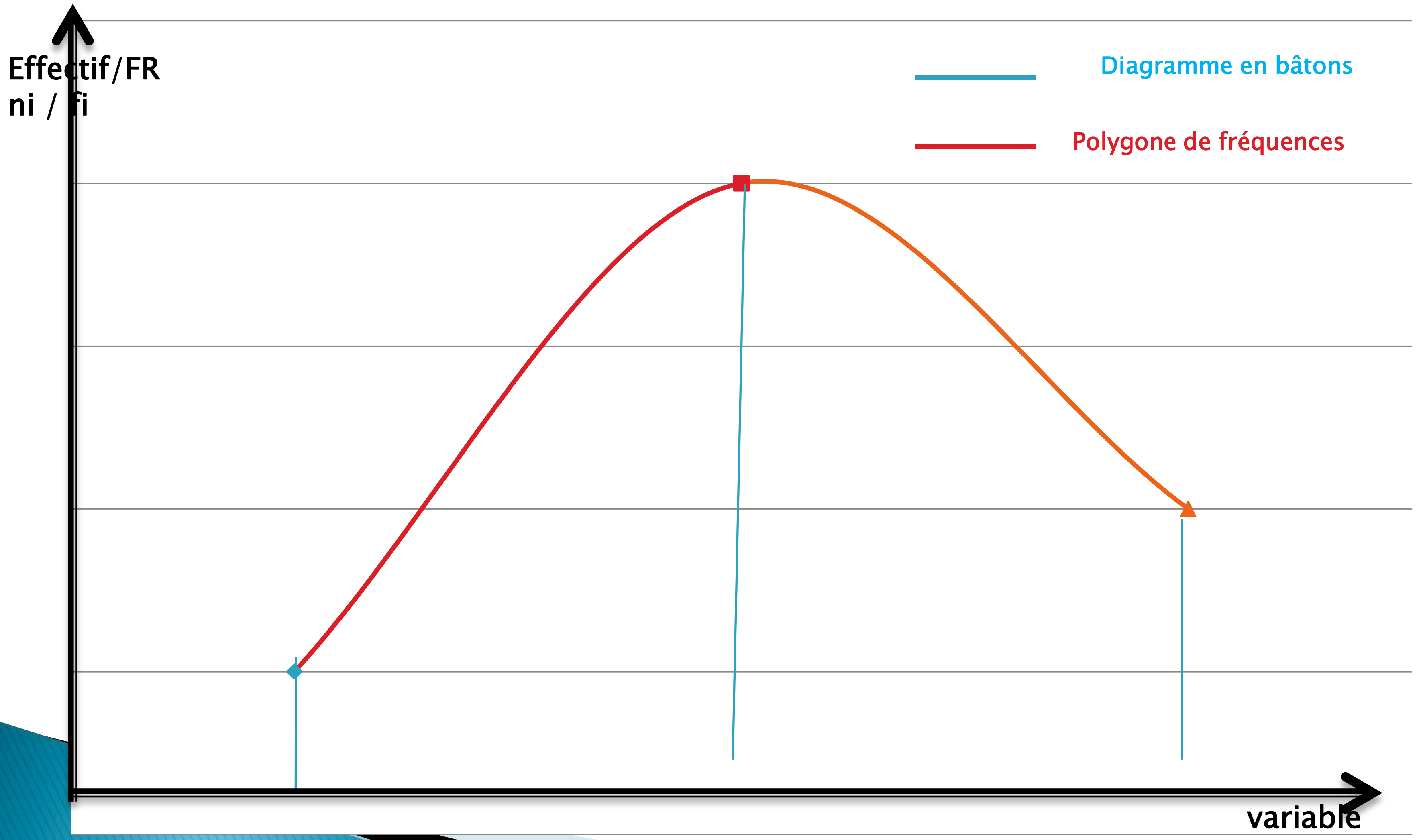


Diagramme en secteur: Répartition de 120 malades selon le groupe sanguin dans le système ABO:

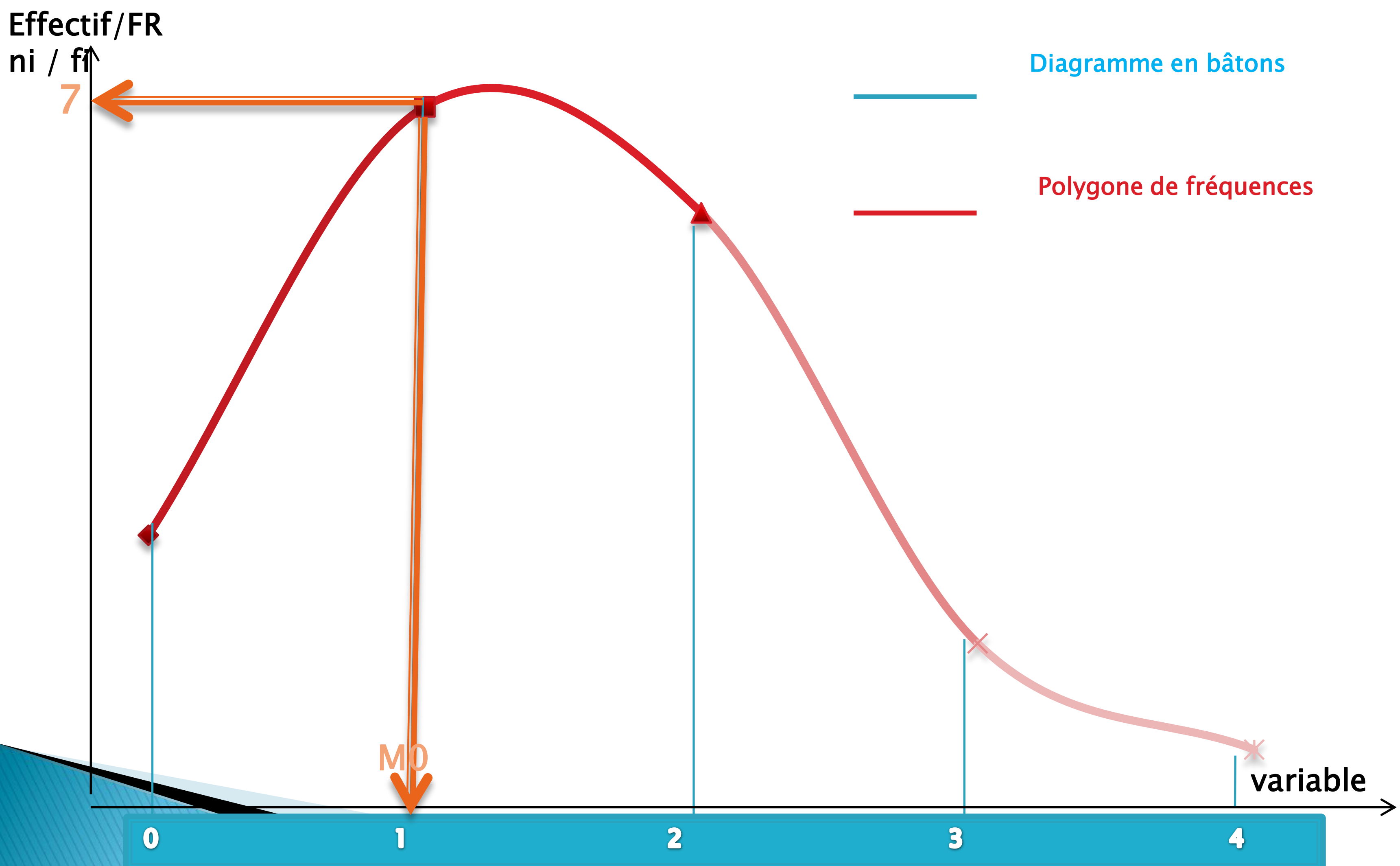


2) Représentation graphique pour la variable quantitative discontinue (discrète):

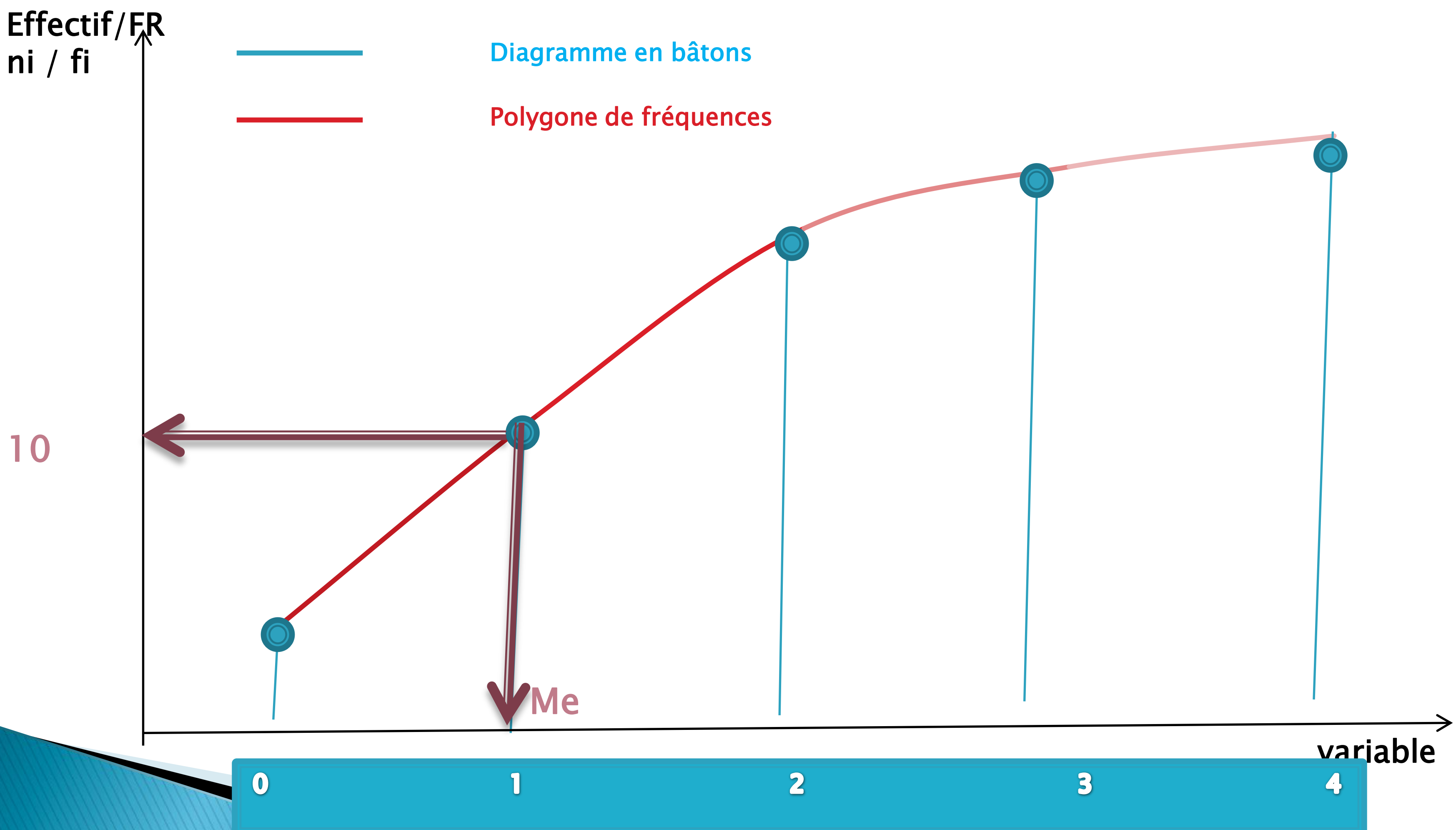
Graphes en bâtons et polygone de fréquences



Diagrammes en bâtons et en polygone de fréquences : distribution du nombre d'épisodes grippal parmi 19 personnes :



Diagrammes en bâtons et en polygone de fréquences des effectifs cumulés: distribution du nombre d'épisodes grippal parmi 19 personnes :



3) Représentation graphique pour la variable quantitative continue:

Graphe en histogramme des fréquences

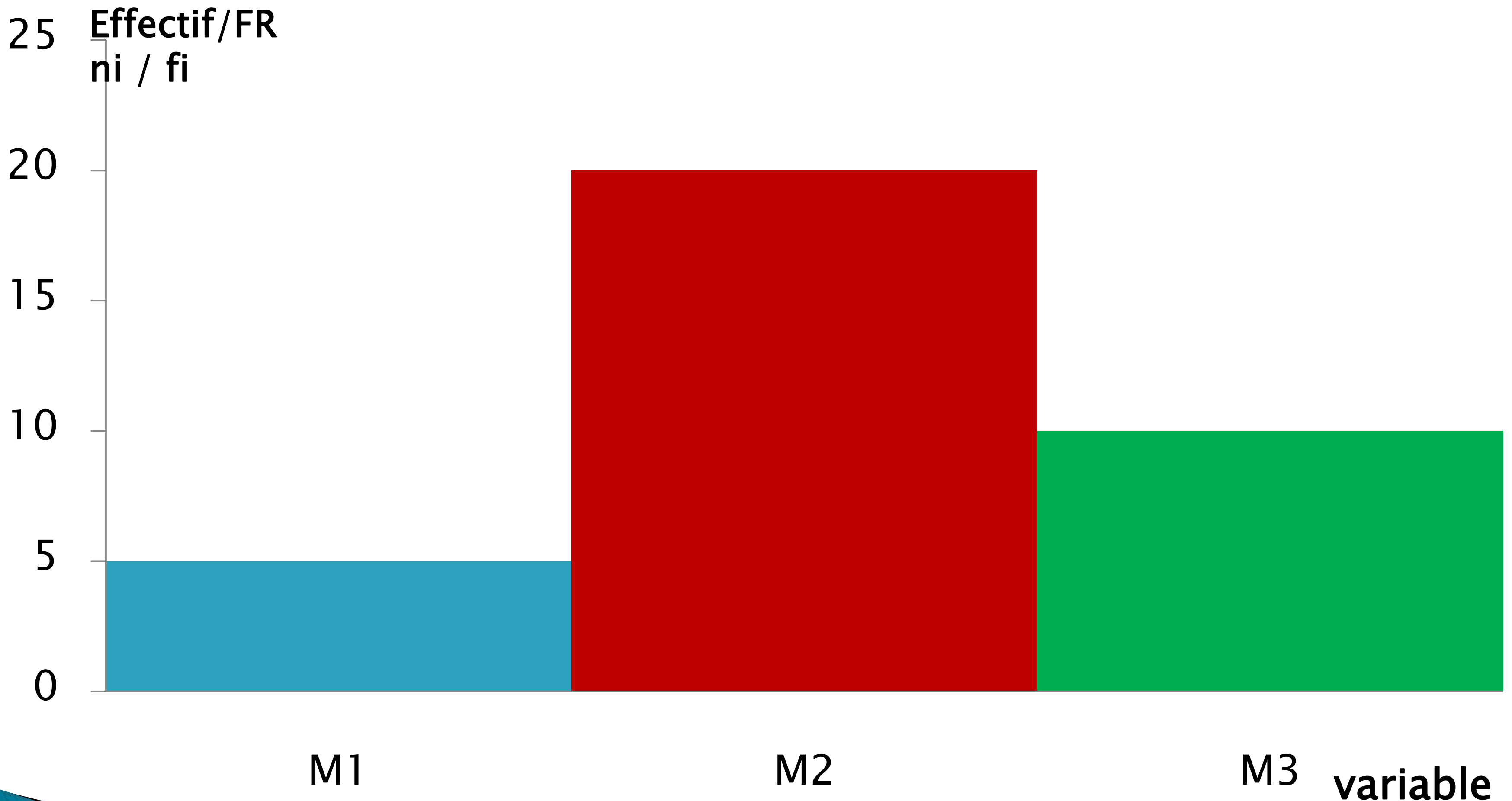
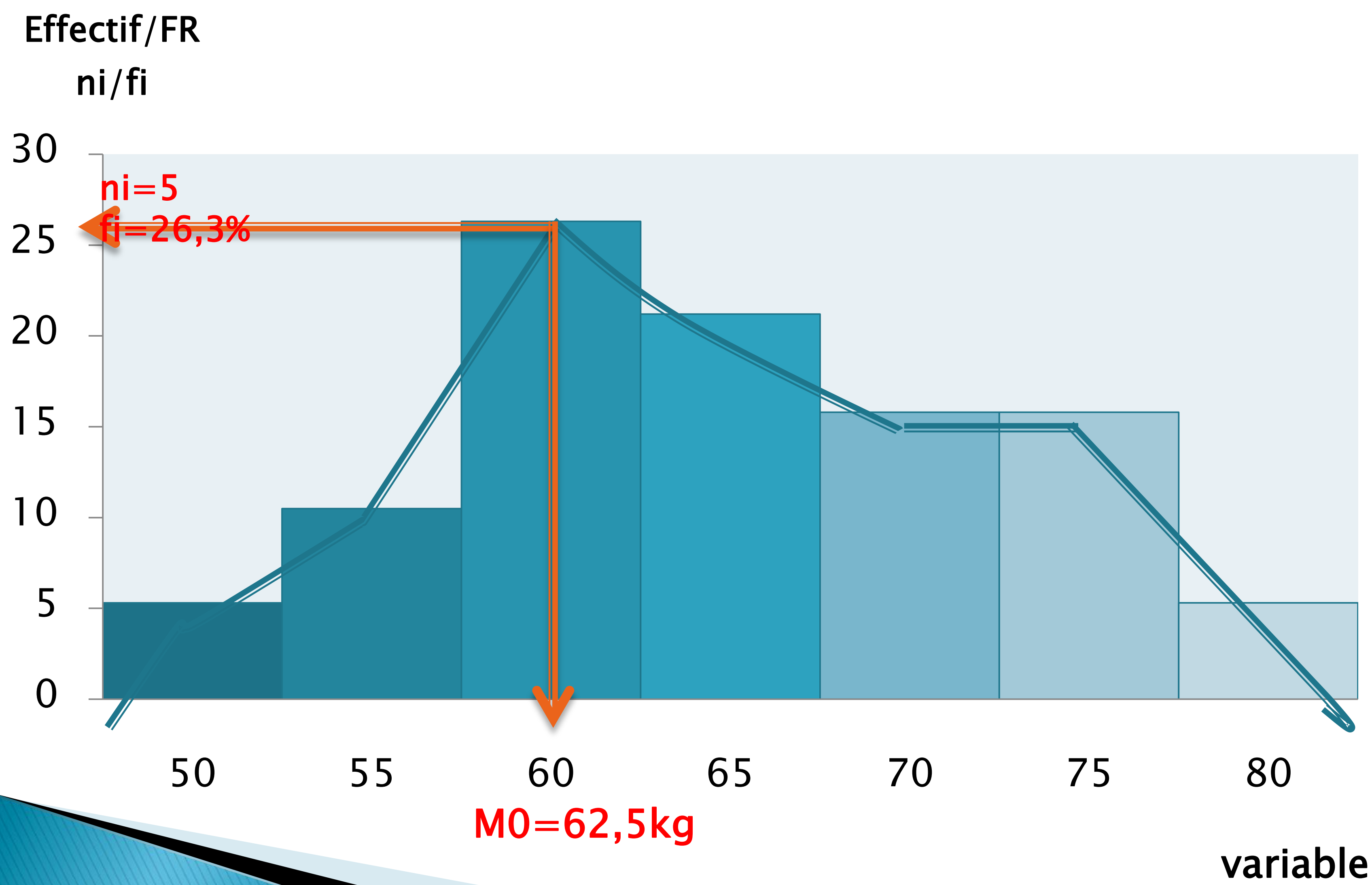


Diagramme en histogramme des fréquences : répartition du poids de 19 étudiants:



IV. Paramètres de réduction

1) Paramètres de tendance centrale :

1-1) Moyenne arithmétique « m »:

Il s'agit d'une valeur centrale autour de laquelle gravitent toutes les autres valeurs.

$$m = \Sigma\ xi\ /\ N \text{ ou } m = \Sigma\ nixi\ /\ N$$

Ex1: soit la série suivante: fréquence cardiaque de 10 étudiants: **59, 72, 58, 65, 77, 83, 72, 77, 62, 62.**

la moyenne des fréquences cardiaques de 10 étudiants est :**m= 68,7 batt/min**

Ex2: soit la série suivante: 3,3,3,3,4,4,5,5,5,5,7,9

La moyenne pondérée est:

$$m = [(4x3)+(2x4)+(4x5)+(1x7)+(1x9)]/12=4,7$$

Ex3: calcul de la moyenne du poids de 19 étudiants sur la base des données groupées en classes s'effectue ainsi:

Poids (Kg)	Centre de la classe (Xi)	Effectifs (ni)	Effectifs cumulés nicu	xini
[50–55[	52,5	1	1	52,5
[55–60[	57,5	2	3	115
[60–65[	62,5	5	8	312,5
[65–70[	67,5	4	12	270
[70–75[	72,5	3	15	217,5
[75–80[	77,5	3	18	232,5
[80–85[	82,5	1	19	82,5
Total		19		1282,5

m=(52,5x1+57,5x2+62,5x5+67,5x4+72,5x3+77,5x3+82,5x1)/ 19= 1282,5/19= 67,5Kg

1-2) Mode « M_0 »:

- Le mode ou la valeur modale est **la valeur de la variable de fréquence maximum**. C'est la valeur dominante qui correspond au plus grand effectif.
- La distribution peut être **uni modale** (un seule mode); **bimodale (**deux valeurs distinctes correspondent aux deux grands effectifs égaux); ou **multimodale** (plusieurs modes différents)
- Lorsque les valeurs de la variable sont répartis en classes, la classe modale est celle pour laquelle la fréquence est maximale. On peut attribuer au mode le centre de la classe modale.

Ex 1: la série: {4,5,8,9,10,12,15} n'a pas de mode.

Ex2: dans la série de fréquence cardiaque de 10 étudiants:
59, 72, 58, 65, 77, 83, 72, 77, 62, 62; la distribution et multimodale: **trois modes: 72,77 et 62**

Ex3: dans l'exemple du nombre d'épisodes de grippe parmi 19 personnes: **le mode est égal à 1 (ni=7)**

Ex4: dans l'exemple du poids de 19 étudiants:
-**la classe modale est : [60-65[**
- **le mode est le centre de la classe modale: Mo= 62,5 Kg (ni=5)**

2-3) Médiane « Me »:

C'est la valeur de la variable qui sépare la série en deux: 50% des observations sont inférieurs à Me et 50% des observations lui sont supérieurs.

❖ **Pour la série statistique non classée:**

- Si le nombre des observations « N » est **impair**, le rang ou la position de la médiane est déterminé par la formule:

$(N + 1)/2 \rightarrow X_{(2K+1/2)}$ avec $N=2K$: $Me = X_{(2K+1/2)}$

Ex: soit une série de 5 poids de nouveau-nés en grammes : 13150; 3200; **3500;** 3510; 3720. $X_{2K+1/2}$ $(2K+1/2=3)=3500$

La médiane est: Me= 3500g

- Si le nombre des observations « N » est **pair**, la médiane correspondant aux rangs: **$(N/2) \rightarrow X_k$ et $(N/2 + 1) \rightarrow X_{k+1}$ avec $N= 2K$:** $Me = X_k + X_{k+1} / 2$

Ex: soit une série de 6 poids de nouveau-nés en grammes « g »: 13150; 3200; **3500; 3510**; 3720; 3800.

Donc : **X_k $(k=3)=3500$ et X_{k+1} $(k+1=4)=3510$**

La médiane est: Me= (3500+3510)/2=3505g

- **Pour la série statistique classée (groupée):**
- on détermine la classe médiane par l'utilisation des effectifs cumulés ,correspond à **n+1/2** ou des fréquences cumulées correspond à **50%.**
- on utilise la formule suivante:

Me= x + [(N/2 – S)/n]*a

X: limite inferieure de la classe médiane

N: taille globale de l'échantillon

S: effectif cumulé de la classe inférieure à la classe médiane

n: effectif de la classe médiane

a: amplitude de la classe médiane

Ex: on considère la série du poids de 19 étudiants:

- **La classe médiane est à 19+1/2=10 est:[65-70[; n=4**
- **Me= 65 + [(19/2 – 8)/4] * 5= 66,9 Kg**

2) Paramètres de dispersion:

Les caractéristiques de tendance centrale et de position sont insuffisantes pour caractériser complètement une série statistique, elle doit être accompagnée d'un paramètre de dispersion qui renseigne sur l'étalement de la série autour de la valeur de cette moyenne.

Ex: soient deux observations (séries statistiques) suivantes:

1) **série1**: 15,20,25,30,35. **m_1=25**; S^2_1=62,5; S_1=7,9; **cv=31%**
2) **Série2**: 5,15,25,35,4. **m_2=25**; S^2_2=250; S_2=15,8; **cv=63%**

Commentaire: les deux séries ont la même moyenne, mais elles ne se ressemblent pas. On dit que la deuxième série est plus dispersée

- Les plus connus de ces paramètres sont la variance et l'écart-type. La variance est simplement le carré de l'écart-type.

2-1) Variance (S^2):

- En cas de données non groupées, la variance s'écrit: $S^2 = [\Sigma (x - m)^2]/N-1$
- En cas de données groupées, la variance s'écrit:

$S^2 = [\Sigma\ ni(xi - m)^2]/N-1$ / $S^2 = (\Sigma nixi^2/N) - m^2)$

2-2) Ecart-type: $S = \sqrt{S^2}$

Exemples:

Ex1: en cas de données non groupées; considérons les deux séries précédentes:

- **Série1**: $S^2 = [(15-25)^2+(20-25)^2+(25-25)^2+(30-25)^2+(35-25)^2]/(5-1) = 62{,}5$. $S = \sqrt{62{,}5} = 7{,}9$
- **Série2**: $S^2 = [(5-25)^2+(15-25)^2+(25-25)^2+(35-25)^2+(45-25)^2]/(5-1) = 250$. $S = \sqrt{250} = 15{,}8$

EX2: en cas de données groupées; considérons le poids de 19 étudiants:

- Le calcul de la variance et l'écart-type du poids de 19 étudiants s'effectue ainsi: **m= 67,5 kg**

Poids (Kg)	Centre de la classe (Xi)	Effectifs (ni)	xini	nixi²
[50-55[	52,5	1	52,5	2756,25
[55-60[	57,5	2	115	6612,5
[60-65[	62,5	5	312,5	19531,25
[65-70[	67,5	4	270	18225
[70-75[	72,5	3	217,5	15768,75
[75-80[	77,5	3	232,5	18018,75
[80-85[	82,5	1	82,5	6806,25
Total		19	1282,5	87718,75

$S^2 = [1(52,5-67,5)^2+2(57,5-67,5)^2+5(62,5-67,5)^2+4(67,5-67,5)^2+3(72,5-67,5)^2 +3(77,5-67,5)^2+1(82,5-67,5)^2]/ (19-1)= 63,9\ kg^2$

$S= \sqrt{63,9}=8$ Kg

2-3) Coefficient de variation:

- Le coefficient de variation (CV) exprime l'écart type en fonction de la moyenne :

$$CV = s/m.$$

- Pour la distribution du poids de 19 étudiants,

CV = 8,0/67.5 = 11,85%.

- La moyenne est représentative des valeurs de la série.
- De façon générale :
  - **0 < CV <10%** : représentativité bonne de la moyenne
  - **10% < CV < 25%** : représentativité acceptable de la moyenne
  - **CV > 25%** : représentativité médiocre de la moyenne

- Si l’écart type quantifie la variabilité de la distribution autour de la moyenne, le CV exprime, sans unité, le degré de variabilité relative, ou, en d’autres termes, le degré de dispersion en fonction de la valeur moyenne.
- Le CV d’une distribution avec une moyenne de 10 et un écart type de 5 est : **5/10 = 1/2 (50%).**
  - Cette distribution a une variabilité plus importante que celle d’une distribution de même écart type mais avec une moyenne de 100 : **5/100 (5%).**
- Les coefficients de variation sont directement comparables. Lorsque deux distributions ont des moyennes différentes, la comparaison de leurs CV est plus instructive que celle de leurs écarts types (ou de leurs variances) respectifs.

3) Quantiles:

- Les quantiles ou statistiques de position sont des paramètres qui occupent un certain rang dans la série des valeurs d’une série statistique, valeurs classées selon un ordre croissant ou décroissant.
- Les quantiles se subdivisent en:
  - ✓ **quartiles**
  - ✓ **déciles**
  - ✓ **percentiles**

3-1)Quartiles:

- Les quartiles divisent la série statistique **en quatre parties égales** comprenant le même nombre de sujets.
- cela lorsqu’on utilise les pourcentages cumulés.

- L’écart interquartile **(Q3 – Q1 = Interquartile Range)** peut être utilisé comme indicateur de dispersion, il correspond à la moitié des valeurs situées dans la partie centrale de la série.

Rang du quantile	• Rang de la valeur du quantile
1er quantile « Q1 »	• 25ème sujet sur 100
2ème quantile « Q2 »	• 50ème sujet sur 100 • Correspond à la médiane
3ème quantile « Q3 »	• 75ème sujet sur 100

❑ Exemple du poids de 19 étudiants:

Rang du quantile	Q1	Q2	Q3
Rang de la valeur	4.75ème valeur (19/4)	9,5ème valeur (19*2/4)	14,25ème valeur (19*3/4)
Classe de la valeur	60–65	65–70	70–75
Valeur	62,5 kg	67,5 kg	72,5 kg

L'écart interquartile : 72,5 kg – 62,5kg= 10 kg

3-2) Déciles:

- Les déciles sont au nombre de 9.
- Ce sont les valeurs de la variable qui partagent la série statistique **en 10 parties** comprenant chacune 1/10ème de l'effectif total.

Rang du décile

- Rang de la valeur du décile

1er décile « D1 »

- 10ème sujet sur 100

2ème décile « D2 »

- 20ème sujet sur 100

5ème décile « D5 »

- 50ème sujet sur 100
- Correspond au deuxième quartile
- Correspond à la médiane

❑ Exemple du poids de 19 étudiants:

Rang du décile	D1	D2	D5	D9
Rang de la valeur	2ème valeur 19/10 = 1,9	4ème valeur 19*2/10=3.8	10ème valeur 19*5/10=9,5	17ème valeur 19*9/10 = 17,1
Classe de la valeur	55-60	60-65	65-70	75-80
Valeur	57,5 kg.	62,5	67,5 kg	77,5 kg

Les classes «55-60» et «75-80» auxquelles appartiennent D1 et D9 contiennent respectivement les valeurs brutes :/ 57,740; 59,990; 60,400; 61,820; 61,820; 64,100; 64,990; 65,450; 66,330; 68,280;69,120; 70,130; 70,560; 72,880;76,340; 76,360; 79,650/

80% des valeurs de la série devraient se situer entre D1 et D9.

3-3) Percentiles:

- Les percentiles ou centiles, au nombre de 99, sont les valeurs de la variable qui divisent la série statistique **en 100 parties** contenant chacune 1/100ème de l'effectif global.
- De façon générale, les percentiles sont utilisés lorsque le nombre de valeurs de la série statistique est supérieur à 1000.
- Le 42ème percentile, par exemple, est la note du 420ème sujet sur 1000.
- Certains percentiles se confondent avec des quantiles déjà vus. Le dixième percentile, par exemple, se confond avec le premier décile.
- Les percentiles sont aussi fréquemment utilisés en biométrie. Comme les déciles, ils permettent de situer la position d'un sujet quelconque par rapport aux autres sujets de la série.

- <u>Ex</u>:

  Par exemple, dans un percentilage du poids, la mesure d’un sujet se situe dans le 5ème percentile, cela signifie que, sur 1000 sujets, il y en aura seulement 40 avec un poids inférieur, mais 950 avec un poids supérieur. Les 10 restants appartiennent évidemment au 5ème percentile.

- NB: Les paramètres de tendance centrale, les paramètres de dispersion ainsi que les quantiles sont déterminés, de nos jours, exclusivement par des moyens informatiques.

- **Logiciels :**

  Epiinfo, SPSS, Epidata, statistica , TGV Biosta, EpiNut, etc.